



*Dipartimento di **Impresa & Management** Cattedra **Customer Intelligence e Logiche di Analisi dei Big Data***

Explicit factor model:

l'unione dei sistemi di raccomandazione con la sentiment analysis

Relatore

Prof. Giuseppe F. Italiano

Candidato

Marco Basile

Matricola 680111

Correlatore

Prof. Luigi Laura

ANNO ACCADEMICO 2017/2018

Indice

| | |
|---|-----------|
| Introduzione | 3 |
| 1. Sistemi di raccomandazione | 10 |
| 1.1 <i>Metodi non personalizzati</i> | 13 |
| 1.2 <i>Metodi content-based</i> | 13 |
| 1.3 <i>Metodi collaborativi</i> | 15 |
| 1.3.1 <i>Metodi user-to-user</i> | 16 |
| 1.3.2 <i>Metodi item-to-item</i> | 18 |
| 1.3.3 <i>Criteri per la scelta tra metodo user-to-user o item-to-item</i> | 19 |
| 1.3.4 <i>Vantaggi e limiti dei metodi user-to-user e item-to-item</i> | 20 |
| 1.3.5 <i>Latent factor model e matrix factorization</i> | 21 |
| 1.4 <i>Metodi ibridi</i> | 22 |
| 1.5 <i>Benefici dei sistemi di raccomandazione</i> | 24 |
| 2. Sentiment analysis | 27 |
| 2.1 <i>Differenti tipologie di opinioni</i> | 29 |
| 2.2 <i>Problemi della Sentiment Analysis</i> | 30 |
| 2.3 <i>Tecniche di pre-processing</i> | 31 |
| 2.4 <i>Creazione di un Sentiment Lexicon</i> | 32 |
| 2.5 <i>Sentiment analysis a livello di documento</i> | 34 |
| 2.5.1 <i>Classificazione del sentiment con apprendimento supervised</i> | 34 |
| 2.5.2 <i>Classificazione del sentiment con apprendimento unsupervised</i> | 35 |
| 2.6 <i>Sentiment analysis a livello di frase</i> | 36 |
| 2.6.1 <i>Classificazione della soggettività delle frasi</i> | 37 |
| 2.6.2 <i>Classificazione del sentiment delle frasi</i> | 37 |
| 2.6.3 <i>Problemi di classificazione a livello di frase</i> | 38 |
| 2.7 <i>Sentiment analysis a livello di aspetto</i> | 39 |
| 2.7.1 <i>Estrazione degli aspetti</i> | 40 |
| 2.7.1.1 <i>Particolarità nell'estrazione degli aspetti</i> | 42 |
| 2.7.2 <i>Raggruppamento degli aspetti in categorie</i> | 43 |
| 2.7.3 <i>Classificazione del sentiment degli aspetti</i> | 44 |

| | | |
|-----------|--|-----------|
| 2.7.4 | <i>Regole di base delle opinioni</i> | 46 |
| 2.7.5 | <i>Estrazione di entità, opinion holder e tempo</i> | 48 |
| 3. | Explicit Factor Models | 50 |
| 3.1 | <i>Costruzione del sentiment lexicon</i> | 52 |
| 3.2 | <i>Mappatura delle coppie aspetto-opinione</i> | 54 |
| 3.3 | <i>Matrice di attenzione utente-aspetto</i> | 56 |
| 3.4 | <i>Matrice di qualità prodotto-aspetto</i> | 56 |
| 3.5 | <i>Integrazione degli aspetti impliciti ed espliciti</i> | 57 |
| 3.6 | <i>Raccomandazioni personalizzate</i> | 58 |
| 3.7 | <i>Experimental setup</i> | 59 |
| 3.7.1 | <i>Esperimenti offline</i> | 60 |
| 3.7.2 | <i>Esperimenti online</i> | 62 |
| 3.8 | <i>Vantaggi per i consumatori e le imprese</i> | 63 |
| 4. | Questionario e risultati | 66 |
| 4.1 | <i>Struttura del questionario</i> | 66 |
| 4.2 | <i>Risultati del questionario</i> | 70 |
| 4.2.1 | <i>Prima sezione: informazioni generali</i> | 70 |
| 4.2.2 | <i>Seconda sezione: informazioni sulle modalità di scrittura delle recensioni</i> | 73 |
| 4.2.3 | <i>Terza sezione: proposta di un explicit factor model per siti di e-commerce</i> | 76 |
| 4.2.4 | <i>Quarta sezione: proposta di un explicit factor model per siti di recensioni di attività ristorative</i> | 78 |
| 4.2.5 | <i>Quinta sezione: proposta di un explicit factor model per siti di organizzazione di viaggi online</i> | 80 |
| 4.3 | <i>Analisi dei risultati</i> | 83 |
| | Conclusioni | 92 |
| | Bibliografia e sitografia | 95 |

Introduzione

“Information is the oil of the 21st century, and analytics is the combustion engine.”

Questa frase è di Peter Sondergaard, vicepresidente esecutivo e membro del comitato operativo di Gartner, una società per azioni multinazionale leader mondiale nella consulenza strategica, ricerca e analisi nel campo dell'Information Technology con oltre 60.000 clienti nel mondo¹.

Questa citazione rappresenta in pieno il mondo in cui ci troviamo al momento, dove tutto gira intorno ai dati ed alle informazioni in essi contenuti ed estraibili tramite l'analisi.

Stando ai dati raccolti ed analizzati da WeAreSocial ed Hootsuite, in Italia, al gennaio 2018, il 73% della popolazione, 43,31 milioni di persone su un totale di 59,33 milioni, utilizza internet, un dato cresciuto del 10% rispetto allo stesso mese dell'anno precedente² (Immagine 1).

(Immagine 1)



Questo dato fornisce ancor maggior valore alla frase sopracitata se si pensa che per ogni azione compiuta online, un utente lascia un *social footprint*, un'impronta digitale, come ad esempio un “mi piace” su Facebook, un tweet, un *upload* su Youtube, una recensione su TripAdvisor, e così via. Tutte queste impronte possono poi essere unite in *lifestreams*, flussi di *social footprint* appunto, ordinati su base cronologica.

¹ <https://it.wikipedia.org/wiki/Gartner>

² <https://www.slideshare.net/wearesocial/digital-in-2018-in-southern-europe-part-1-west-86864268>

Peter Sondergaard ha quindi ragione nel dire che le informazioni sono la benzina del ventunesimo secolo, infatti se in passato spesso le aziende avevano il problema della scarsità di informazioni e di dati, oggi hanno il problema opposto, esiste una grande mole di dati difficili da gestire ed analizzare.

Da qui nasce il concetto di *Big Data*, una grande mole di dati sia strutturati che non che necessitano però di essere analizzati per poi poter essere sfruttati al meglio.

Per quanto osservato in precedenza, i *Big Data*, quindi, sono una diretta conseguenza delle nuove abitudini digital che pervadono la vita quotidiana, sia a livello individuale che aziendale, riempiendola di accessi verso siti web, di “mi piace”, di tweet, di commenti, di condivisioni, dati digitali che vengono salvati e immagazzinati per poter essere utilizzati.

Per descrivere nel dettaglio i *Big Data* si può far riferimento al modello delle “4V” ideato da IBM³ (Immagine 2):

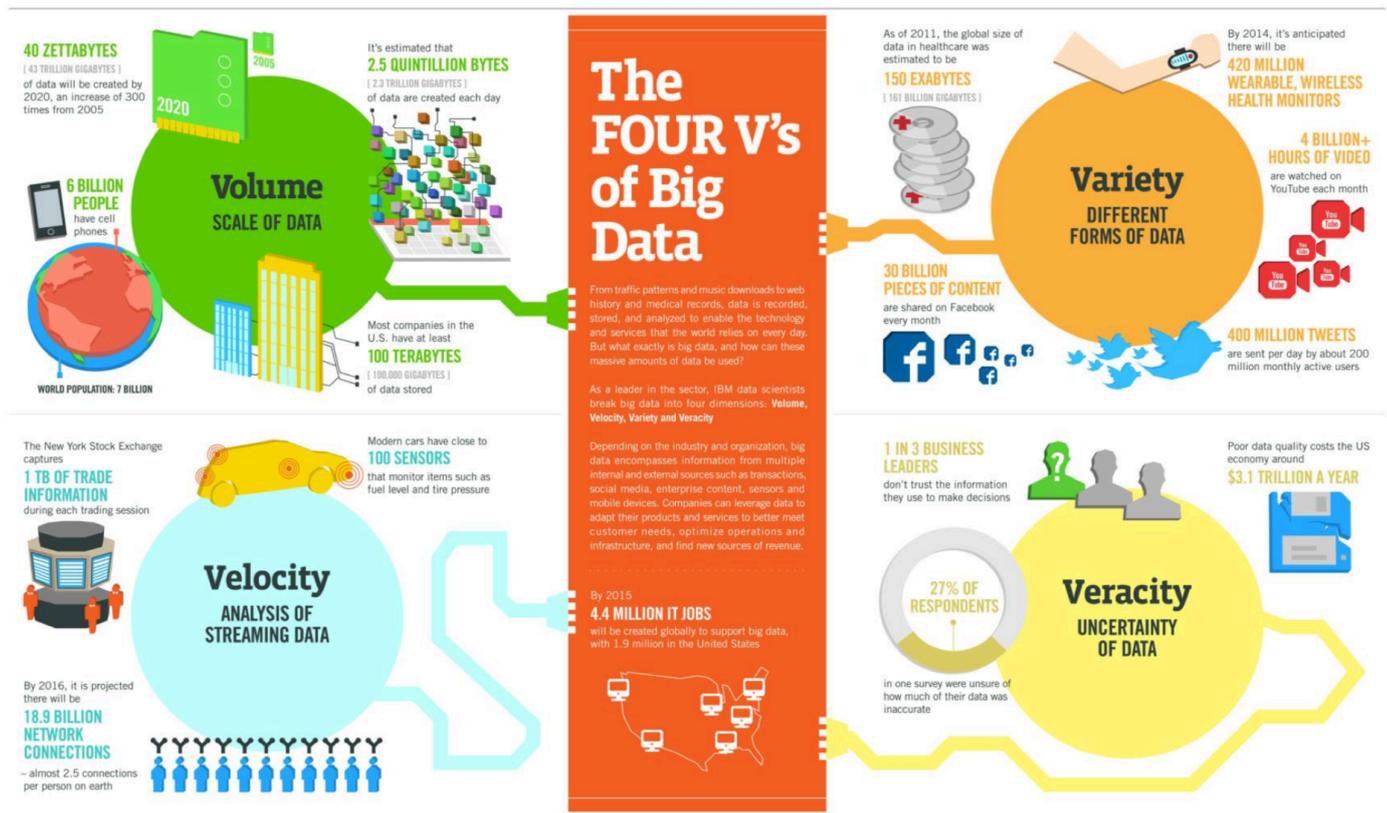
- **Volume:** il volume di dati, cioè la possibilità di ottenere, immagazzinare e accedere a grandi volumi di dati, così elevati da non poter essere gestiti e controllati da database tradizionali.
- **Velocità:** si riferisce alla velocità con cui i dati sono a disposizione delle aziende e con cui vengono generati e trattati. L’analisi delle informazioni deve essere effettuata il più rapidamente possibile se non in tempo reale poiché spesso il valore generato dall’analisi stessa dipende dalla velocità con cui si ricavano informazioni, soprattutto nei casi in cui diventano obsolete in tempi molto rapidi.
- **Varietà:** si riferisce alla varietà dei dati a disposizione. In passato l’analisi dei dati poteva essere effettuata solo su dati strutturati in tabelle o database relazionali, all’interno dei quali non si potevano inserire dati di formati diversi come foto, conversazioni, dati audio e video, documenti, ecc. Con il termine *Big Data*, invece, si fa riferimento a dati provenienti da fonti diverse e di tipologia diversa, in particolare è possibile unire nello stesso dataset dati strutturati e non strutturati. Questi ultimi hanno ormai raggiunto un’importanza determinante rappresentando l’80% dei dati prodotti nel mondo⁴ divenendo quelli di maggior interesse per le aziende.
- **Veridicità:** i *Big Data* sono una fonte di valore per chi li utilizza. Oltre ad essere utili per analisi di mercato, possono essere utilizzati anche per migliorare le procedure interne all’azienda, per esempio per misurare performance, per ridurre tempi di produzione o migliorare le attività di marketing e dello stesso prodotto o servizio offerto. Ed ovviamente chi non considera l’ipotesi di utilizzarli rischia seriamente di rimanere indietro nella competizione e di essere estromesso dal mercato.

Ma perché i dati possano essere considerati di valore, c’è bisogno che sia dei dati veri, e questo potrebbe essere un problema in un mondo dove sono sempre più le informazioni fasulle sul web, basti pensare al problema delle *fake news* nel mondo dell’informazione e quello delle *fake reviews* in quello dei feedback dai clienti ad esempio.

³ <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

⁴ Nadkarni A., Yezhkova N. 2014

(Immagine 2)



Collegati ai *Big Data* ci sono alcuni limiti che ne rallentano l'espansione. Il primo è quello pocanzi descritto, ovvero la veridicità delle informazioni, anche se è la qualità in generale ad essere importante, poiché i *Big Data*, per poter essere considerati affidabili e quindi per poter rappresentare uno strumento decisionale e di analisi veramente utile, devono possedere alcune caratteristiche precise: devono possedere dati completi, devono essere il più aggiornati possibile, devono essere coerenti tra loro e tecnicamente conformi alle regole su cui si basano i *Big Data* e devono essere quanto più accurati possibile.

Ovviamente la raccolta dei dati e, di conseguenza, la qualità dei dati ricavati sono collegati all'obiettivo dell'analisi che si vuole compiere. Per esempio se si vuole analizzare il "*click-stream*" di un sito, cioè il percorso che l'utente compie all'interno di quest'ultimo, la qualità dei dati è una dimensione poco importante, mentre sarà importante la capacità di individuare i software che eseguono operazioni automatiche sui siti, in maniera tale da separare gli utenti effettivi da quelli "virtuali".

Altro problema, e limite, collegato con i *Big Data*, è quello della privacy. È proprio di questo recente periodo il polverone alzato dal caso Cambridge Analytica su questo tema. Tutto nasce con un'intervista da parte di Christopher Wylie al The Guardian. Il giovane è un ex dipendente della Cambridge Analytica, società che ha come scopo quello di raccogliere dai social network un'enorme quantità di dati sui loro utenti e successivamente fornirli ad aziende, idealmente per ricerche di mercato. Questa società, secondo le indagini effettuate, avrebbe poi a sua volta fornito le informazioni ottenute al governo russo per supportare online la propaganda a favore del presidente USA Donald Trump, eletto nel novembre del 2016. In pratica, sembra che

in base a questi dati siano stati creati diversi account falsi per diffondere post online e *fake news* contro Hillary Clinton.

Probabilmente questo è solo uno dei tanti casi di uso inappropriato dei dati degli utenti da parte di aziende, questi ultimi però, dal canto loro, hanno anche la colpa di fornire dati in maniera un po' troppo "leggera". Tuttavia ciò non toglie che, in un mondo ormai basato sui dati, sia necessario un costante aggiornamento delle leggi in materia di privacy, in modo tale da evitare che gli utenti possano essere danneggiati direttamente o indirettamente.

Se le informazioni sono la benzina, serve qualcosa che sia in grado di utilizzarla, un "*combustion engine*", che non è altro che la capacità di analizzare la mole immensa di dati a disposizione. Infatti senza capacità di analisi ed interpretazione, il potenziale informativo di un volume di dati si annulla, quindi sarà necessario sviluppare un'infrastruttura fisica e degli algoritmi appositi per poter utilizzare efficacemente i dati raccolti⁵.

Due strumenti che si sono rivelati molto utili per tale fine per le aziende, sono i sistemi di raccomandazione e la *sentiment analysis*.

Quest'ultima rappresenta uno degli orizzonti più importanti, in termini di volume e di rilevanza delle informazioni ottenibili. Attualmente le imprese sono più o meno dotate di sistemi e metodi per analizzare i dati non testuali, lo stesso non si può però dire per quanto riguarda i dati testuali. Ovviamente questo ritardo è comprensibile dato che si sono prima sviluppati gli strumenti per analizzare i dati che sono in forma strutturata e numerica, oltretutto il valore dei dati testuali ha acquisito vera importanza solo negli ultimissimi anni grazie, in particolar modo, alla crescita esponenziale di utilizzo degli smartphone ed all'entrata massiccia dei social networks nella vita quotidiana della maggior parte della popolazione.

Uno degli obiettivi delle imprese, oggi, risiede proprio nel riuscire ad interpretare ed estrarre da questa grandissima mole di dati, generati ogni giorno, informazioni utili per le proprie attività, si pensi, per esempio, al vantaggio di poter capire se un prodotto o servizio è apprezzato o meno dal mercato, quali sono gli aspetti meno graditi e quindi da modificare di un prodotto, alla possibilità di valutare in maniera maggiormente precisa la *brand reputation* e così via. In generale tutti i settori possono trarre un vantaggio dall'analisi dei dati di testo.

In pratica, la *Sentiment Analysis* è l'area di ricerca che studia opinioni, sentimenti, valutazioni, attitudini ed emozioni delle persone riguardo entità ed i loro aspetti espresse attraverso un testo scritto.

Questa tecnica conferisce alle aziende la possibilità di ottenere informazioni generate direttamente dai consumatori, senza che siano influenzati in alcun modo dall'impresa, quindi informazioni utili anche verso altre funzioni aziendali oltre a quella di marketing.

I dettagli di tale tecnica fondamentale per le imprese saranno discussi nel capitolo due.

⁵ Pasini P., Perego A. 2012

L'altro strumento sopracitato, utile per l'estrazione di informazione dai dati, sono i sistemi di raccomandazione.

Questi ultimi nascono con lo scopo di fornire agli utenti dei suggerimenti personalizzati riguardo un set di oggetti, materiali o immateriali, basando le raccomandazioni sui dati a loro disposizione riguardo gli user e gli oggetti stessi.

I sistemi di raccomandazione salirono alla ribalta per la prima volta nel 2006, quando Netflix mise in palio 1 milione di dollari, che sarebbe stato consegnato a chiunque fosse riuscito a migliorare le performance del loro algoritmo di raccomandazione di almeno il 10%.

Netflix fu fondata nel 1997 e iniziò il suo business noleggiando film ai suoi clienti inviandoli tramite mail, il consumatore poteva tenerlo per tutto il tempo che desiderava, ma non poteva noleggiarne un altro senza restituire prima il precedente. In un secondo momento l'azienda è passata all'utilizzo di internet ed allo streaming online. L'importanza e la centralità per il business di Netflix del sistema di raccomandazione fu palesata proprio nel 2006 con il premio di 1 milione di dollari messo in palio nel "Netflix prize", il concorso a premi sopracitato. Ciò che più interessa al colosso americano è riuscire a prevedere i gusti del cliente ed aumentare la sua soddisfazione tramite consigli personalizzati.

Netflix, però, non ha sfruttato i dati a sua disposizione soltanto attraverso i sistemi di raccomandazione, ma li ha utilizzati, e li utilizza ancora, per progettare e creare film e serie TV. Un esempio di ciò è la serie "House of Cards", prodotta da Netflix appunto, e nata proprio grazie all'analisi dei dati a sua disposizione. In particolare, partendo da una serie TV omonima degli anni '90, l'azienda americana ha inserito come protagonista un attore particolarmente amato dal pubblico a quei tempi ed un regista altrettanto stimato dai consumatori. In questo modo Netflix è riuscito a creare una serie TV dal successo mondiale, sottolineando una volta di più l'importanza e la centralità dei dati e le informazioni in essi contenuti.

Ma l'utilizzo dei sistemi di raccomandazione nel settore della distribuzione via internet di film, serie televisive e altri contenuti di intrattenimento è solo uno dei tanti. Ormai quasi tutte le aziende operanti in internet li utilizza, per Amazon.com, ad esempio, è uno strumento centrale grazie alla sua capacità di reindirizzare gli utenti su diversi oggetti, anche meno conosciuti e richiesti, riuscendo così a sfruttare al meglio il *long tail concept*.

In particolare i sistemi di raccomandazioni si basano principalmente su tre tipi di approcci differenti:

- quello collaborativo: che crea suggerimenti utilizzando le similarità tra gli utenti, partendo dal presupposto che utenti simili abbiano gusti simili;
- quello basato sui contenuti: che utilizza le informazioni che il sistema ha sui diversi oggetti in catalogo per fornire dei consigli personalizzati;
- e quello ibrido: il quale si ottiene dalla combinazione dei due approcci precedenti.

Nel capitolo 1 saranno analizzati in dettaglio i sistemi di raccomandazione.

Obiettivo di questa tesi è quello di analizzare e testare l'applicabilità in Italia e ipotizzare e testare, tramite analisi empirica, nuove possibili applicazioni, da me ipotizzate, del modello chiamato "*Explicit factor model*", ideato da Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma, che unisce insieme tra loro sistemi di raccomandazione ed analisi testuale, in un approccio innovativo e non ancora presente sul mercato, al fine di migliorare la qualità delle raccomandazioni mostrate agli utenti, fornire delle spiegazioni sul perché di tali consigli e per suggerire delle dis-raccomandazioni riguardo entità da non acquistare poiché non adatte al consumatore che le sta valutando, tutto ciò per creare valore per le due parti in gioco, l'azienda ed il consumatore. L'idea è quella di applicare la *sentiment analysis* alle recensioni scritte dagli utenti, questo perché sono un contenuto sempre più presente online e sempre più centrale nelle scelte di acquisto dei consumatori, come dimostrato da diversi studi che saranno riportati all'interno di questa tesi. Ed è proprio la centralità nella vita dei consumatori di questi *user generated content* che rende questo modello così innovativo ed unico, capace di agevolare le scelte degli utenti garantendogli anche una miglior soddisfazione e, di conseguenza, di migliorare le performance dell'azienda, fornendo a quest'ultima una miglior capacità persuasiva e clienti più soddisfatti e fedeli.

Ad oggi, infatti, ci troviamo nell'era del web 2.0, caratterizzato da una forte partecipazione degli utenti, che da statici che erano nell'internet dei contenuti, il web 1.0, dove i siti internet erano destinati alla sola consultazione di documenti in grado di generare un rapporto unidirezionale con l'utente che era passivo, diventano attivi e partecipano alla creazione di contenuti. Questa fase è caratterizzata dall'annullamento di barriere geografiche tra le persone e c'è una facilità assoluta di interazione tra utenti.

Tutto ciò ci porta all'e-WOM, *electronic word-of-mouth*, il passaparola su internet, del quale le recensioni online sono l'espressione più chiara. In particolare nel momento in cui non è possibile verificare le caratteristiche di un prodotto prima del suo acquisto, l'e-WOM diventa un passaggio inevitabile nel processo di scelta pre-acquisto da parte del consumatore.

All'interno dei diversi siti, egli compie azioni complementari e contrarie: suggerisce e ascolta i consigli di terzi, condivide foto e scarica video, scrive una recensione su un ristorante e legge attentamente un giudizio lasciato da un utente sconosciuto.

In modo volontario e propositivo, gli utenti hanno cominciato in questo modo ad inserire le proprie valutazioni su strutture ricettive prima, su territori e altri servizi poi, facendo di questi siti un punto di riferimento per gli altri utenti web, ma anche per le imprese, le quali raccolgono tutti i suggerimenti online per ottimizzare la propria offerta turistica⁶.

Il consumatore finale si crea aspettative e progetti basandosi sui racconti condivisi in rete, le recensioni degli utenti, infatti, vengono considerate come una fonte molto più affidabile rispetto a quelle di carattere

⁶ Sfodera, 2011

commerciale perché ritenute veritieri, facili da usare e più affidabili e di conseguenza, in grado di influenzare le scelte di altri⁷.

Nella maggior parte dei portali turistici dove è possibile pubblicare una recensione su un dato bene/servizio, la descrizione dell'esperienza personale nella struttura, è accompagnato da un punteggio numerico: da qui scaturisce una classifica che ordina gli esercizi pubblici recensiti all'interno di una categoria.

Da evidenziare è la duplice natura del passaparola: può, infatti, esistere quello positivo, che aumenta la percezione della qualità del prodotto/servizio in essere, ma allo stesso tempo può essercene uno negativo che secondo la gravità dell'insoddisfazione, può ledere l'immagine dello stesso prodotto/servizio, scoraggiando i consumatori nell'acquisto.

E proprio da questa centralità delle recensioni nel processo di acquisto di prodotti e servizi online che nasce l'idea di una fusione tra sistemi di raccomandazione e *sentiment analysis* che sarà spiegata e testata all'interno di questa tesi.

⁷ Gretzel e Yoo, 2008

1 Sistemi di raccomandazione

L'aspetto e la crescita dei mercati online hanno avuto un impatto considerevole sulle abitudini dei consumatori, fornendo loro l'accesso ad una maggiore varietà di prodotti e informazioni su questi beni. Mentre questa libertà di acquisto ha reso il commercio online un'industria multimiliardaria, ha reso anche più difficile per i consumatori selezionare i prodotti che meglio si adattano alle loro esigenze. Una delle principali soluzioni proposte per questo problema di sovraccarico di informazioni sono appunto i sistemi di raccomandazione, che forniscono suggerimenti automatici e personalizzati di prodotti ai consumatori. Tali sistemi sono stati utilizzati in un'ampia varietà di applicazioni, come la raccomandazione di libri, musica, film, notizie e pagine web.

I sistemi di raccomandazione hanno il compito di fornire agli utenti dei suggerimenti personalizzati riguardo un set di oggetti, basandosi sui dati a loro disposizione riguardo gli user e gli oggetti stessi.

Le raccomandazioni personalizzate aiutano gli utenti a scoprire prodotti interessanti per loro che altrimenti non avrebbero scoperto, è come se tali sistemi siano dei personal shopper che conoscono non soltanto il prodotto venduto, ma anche il cliente.

Lo sviluppo di tali sistemi nasce da una semplice osservazione, le persone molto spesso fanno affidamento sui consigli di altri individui quando devono prendere delle decisioni nella loro vita quotidiana, come ad esempio che libro leggere, che film guardare o in che ristorante cenare.

Cercando di imitare queste dinamiche, i primi sistemi di raccomandazione utilizzavano degli algoritmi per creare dei suggerimenti per un utente attivo, ovvero in cerca di suggerimenti, che riguardavano articoli che users con gusti simili a quest'ultimo avevano apprezzato. Questo approccio è quello utilizzato dai sistemi collaborativi e il suo fondamento logico è che se l'utente attivo ha concordato in passato con alcuni utenti, allora anche le altre raccomandazioni provenienti da questi utenti simili dovrebbero essere rilevanti e di interesse per lui.

Di solito le raccomandazioni vengono mostrate come classifiche di oggetti, create utilizzando le preferenze degli utenti, che possono essere esplicite, come la valutazione di un film, o implicite e quindi derivanti dalle loro attività, come ad es. la consultazione della pagina di un determinato oggetto su un sito di e-commerce.

Tali sistemi di raccomandazione svolgono un ruolo importante in siti Internet altamente qualificati come Amazon.com, YouTube, Yahoo, Tripadvisor e Netflix, basti pensare che quest'ultimo, nel 2006, ha messo in palio 1 milione di dollari che sarebbe andato a chiunque avesse migliorato l'algoritmo di raccomandazione del 10%. La stessa azienda, nel dicembre del 2015, ha pubblicato una ricerca⁸ dove evidenziava come l'80% delle ore di streaming sul sito era dovuto ai sistemi di raccomandazione, e solo il 20 era collegato a ricerche autonome degli utenti. L'importanza delle raccomandazioni è stata analizzata anche in una ricerca⁹ che ha

⁸ *Carlos A. Gomez-Uribe, Neil Hunt 2015*

⁹ *Renjie Zhou, Samamon Khemmarat, Lixin Gao 2010*

mostrato come su YouTube la fonte principale di visualizzazioni per la maggior parte dei video sia la raccomandazione di video correlati, causa di circa il 30% del totale delle *views*.

Una definizione formale del problema delle raccomandazioni è: sia C un insieme di utenti, e sia S l'insieme dei possibili oggetti che possono essere raccomandati, insieme che può essere di enormi dimensioni, sia u una funzione di utilità che misura quanto un determinato oggetto s può essere utile all'utente c , allora il problema di un sistema di raccomandazione è quello di scegliere per ogni utente $c \in C$ l'oggetto $s' \in S$ tale che massimizzi la funzione di utilità:

$$\forall c \in C, \quad s' = \arg \max_{s \in S} u(c, s)$$

In generale, i dati dei sistemi di raccomandazione sono di tre tipi: dati sugli articoli, sugli utenti e sulle transazioni, ovvero relazioni tra articoli ed utenti.

Gli articoli sono ciò che viene consigliato all'utente, e sono caratterizzati dalla loro complessità e dal valore o utilità, che può essere positiva se l'articolo ha valore per l'utente o sarà negativa se non gli è utile.

Va tenuto conto che a tali oggetti è collegato un costo, che sarà la somma del costo cognitivo e di quello monetario eventualmente pagato dall'utente. Se un elemento selezionato è rilevante per quest'ultimo, il costo è dominato dal vantaggio di aver acquisito un articolo utile, mentre se l'articolo non è pertinente, il valore netto di tale articolo per l'utente è negativo.

Inoltre i sistemi di raccomandazione utilizzano le proprietà e le caratteristiche degli oggetti per calcolare che utilità potrebbero avere per un dato utente. Ad es. per un film il regista, gli attori ed il genere, sono valutati come caratteristiche e sono utilizzati per generare dei suggerimenti.

Per quanto riguarda gli utenti, questi possono avere obiettivi e caratteristiche molto differenti, e per generare le raccomandazioni, i sistemi utilizzano le informazioni che hanno a disposizione su di loro. Ovviamente, le informazioni da utilizzare, dipenderanno dal tipo di sistema di raccomandazione che si vuole utilizzare, ad es. nel *collaborative filtering*, gli utenti vengono visti come una semplice lista contenente le valutazioni da loro fornite per alcuni articoli, in un *recommendation system* demografico, vengono utilizzati attributi socio-demografici come età, sesso, professione e istruzione dell'utente.

Gli utenti possono anche essere valutati in base ai loro comportamenti, ad es. in che modo visitano un sito o come ricercano informazioni per un viaggio.

Inoltre, i dati che il sistema ha sugli user, possono includere informazioni riguardo le relazioni che si instaurano tra utenti e, queste, potrebbero essere utilizzate per consigliare gli articoli più adatti.

Ultima tipologia di dati sono le transazioni, interazioni tra utente ed articolo che vengono registrate ed utilizzate per generare delle raccomandazioni. Le valutazioni dei consumatori su un determinato oggetto sono la forma più popolare di transazioni. Le valutazioni possono assumere forme diverse¹⁰: valutazioni numeriche, per esempio da 1 a 5, valutazioni ordinali, come ad es. "fortemente d'accordo, d'accordo, neutrali, in disaccordo, fortemente in disaccordo" in cui l'utente è invitato a selezionare il termine che meglio indica la sua opinione su un oggetto e valutazioni binarie, dove all'utente viene chiesto soltanto se un articolo è buono o meno.

Nelle transazioni che raccolgono valutazioni implicite, il sistema mira a dedurre l'opinione degli utenti in base alle azioni dell'utente. Ad esempio, se un utente inserisce la parola chiave "Calcio" su Amazon.com, vedrà restituita una lunga lista di libri, l'utente può fare clic su un determinato libro nell'elenco per ricevere ulteriori informazioni, ed a questo punto, il sistema può dedurre che l'utente è leggermente interessato a quel libro.

Nei sistemi conversazionali, cioè i sistemi che supportano un processo interattivo, il modello di transazione è più raffinato. In questi sistemi, le richieste degli utenti si alternano alle azioni di sistema, l'utente, cioè, può richiedere una raccomandazione e il sistema può produrre un elenco di suggerimenti, ma può anche richiedere ulteriori input per fornire all'utente risultati migliori. In questo modello di transazione, il sistema raccoglie le varie richieste e risposte e può imparare a modificare la propria strategia di interazione osservando il risultato del processo di raccomandazione¹¹.

Questi sono gli elementi comuni a tutti i sistemi di raccomandazione, nei paragrafi successivi verranno analizzati i diversi tipi di raccomandazioni possibili: metodi non personalizzati, metodi *content based*, metodi collaborativi e metodi ibridi.

¹⁰ Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S. 2007

¹¹ Mahmood, T., Ricci, F. 2009

1.1 Metodi non personalizzati

Gli algoritmi non personalizzati sono quelli che non generano delle raccomandazioni ad hoc per ogni utente, quindi sono molto semplici e solitamente sono utilizzati quando le informazioni che si hanno a disposizione sull'utente non sono sufficientemente numerose.

I metodi non personalizzati più popolari sono due, uno che consiglia agli utenti gli articoli più popolari senza tener conto delle valutazioni degli altri utenti e l'altro che invece base le raccomandazioni sulla media di tutte le valutazioni espresse dagli utenti.

Il primo, non tenendo conto delle valutazioni, potrebbe anche consigliare un articolo con un numero elevato di recensioni negative, ed allo stesso tempo, consigliando gli articoli più popolari, potrebbe risultare poco utile poiché potrebbe consigliare degli oggetti che l'utente già conosce. Ulteriore aspetto negativo, è che questo metodo, nel caso di grandi cataloghi di prodotti, accrescerà il problema della long tail, ovvero quella percentuale di prodotti meno richiesti che non diventerà mai popolare.

Il secondo metodo è più complicato del primo poiché non calcola solo il numero delle valutazioni, ma ne effettua anche la media, ma risulta anche essere più utile ed efficace poiché riesce a mostrare all'utente il sentimento generale intorno ad un determinato articolo.

Tali metodi non personalizzati possono essere utili quando si hanno poche informazioni a disposizione, ma i loro effetti sono nettamente meno positivi rispetto a quelli generati dai metodi personalizzati che analizzeremo di seguito.

1.2 Metodi content-based

In questi metodi, il sistema impara a consigliare articoli simili a quelli che in passato l'utente ha acquistato ed apprezzato. La somiglianza degli oggetti è calcolata in base alle caratteristiche degli articoli. In particolare, l'utilità $u(c,s)$ dell'articolo s per l'utente c , è stimata basandosi sulle utilità assegnate dall'utente c all'oggetto $s_i \in S$ che è simile all'item s . I suggerimenti basati sui contenuti, quindi, si basano sul presupposto che articoli con caratteristiche oggettive o attributi simili, siano valutati allo stesso modo.

Ad esempio, nel caso di raccomandazioni di film, per mostrare dei suggerimenti all'utente c , il sistema di raccomandazione *content based* cerca di comprendere le caratteristiche in comune tra i film che l'utente ha votato in maniera positiva nel passato (attori specifici, registi, generi, ecc.) e solo i film che hanno il più alto grado di similarità con le preferenze dell'utente saranno raccomandati.

In altre parole, un approccio basato sui contenuti crea un profilo dell'utente in base alle caratteristiche presenti negli articoli che l'utente ha valutato positivamente¹².

Questi metodi *content-based*, richiedono delle tecniche adeguate per la costruzione dei profili degli utenti e degli articoli per confrontarli tra loro. In particolare si susseguono tra loro tre passaggi:

- analisi dei contenuti a disposizione: le descrizioni degli articoli provenienti da diverse fonti sono elaborate in questa fase per estrarre le informazioni dal testo e per produrre una rappresentazione strutturata degli articoli e delle loro caratteristiche da poter utilizzare nelle fasi successive;
- analisi dei dati rappresentativi delle preferenze dell'utente e costruzione del profilo dell'utente: per costruire e aggiornare il profilo dell'utente attivo, per il quale devono essere fornite raccomandazioni, le sue reazioni agli elementi vengono raccolte e registrate. Solitamente si distingue tra due tipi di reazioni o feedback, positive, se l'articolo è apprezzato dall'utente, e negative se viceversa non lo è. È possibile adottare due tecniche diverse per registrare i feedback, una chiamata feedback esplicito, quando il sistema chiede all'utente di valutare esplicitamente gli articoli, e la tecnica chiamata feedback implicito, quando non vi è coinvolgimento attivo dell'utente poiché il feedback deriva dal monitoraggio e dall'analisi delle attività dell'utente. Esistono tre approcci principali per ottenere un feedback esplicito: attraverso una valutazione binaria del tipo *like/dislike*, valutazioni su una scala numerica (da 1 a 5 ad esempio) e commenti testuali degli utenti. I metodi di feedback implicito si basano invece sull'assegnazione di un punteggio a specifiche azioni dell'utente inerenti ad un articolo, come il salvataggio, l'eliminazione, la stampa, ecc. il feedback esplicito ha il vantaggio della semplicità, quello implicito ha come vantaggio principale quello di non dover chiedere un coinvolgimento diretto all'utente;
- ultima fase è quella in cui vengono sfruttate entrambe le informazioni ricavate nei passaggi precedenti per raccomandare un determinato articolo ad un utente potenzialmente interessato. Solitamente il sistema di raccomandazione implementa alcune strategie per classificare gli articoli potenzialmente interessanti in base alla pertinenza rispetto al profilo utente, e gli articoli con il punteggio più alto sono inclusi in un elenco di raccomandazioni che viene presentato all'utente. I gusti degli utenti di solito cambiano nel tempo, pertanto le nuove informazioni su di loro devono essere mantenute e utilizzate per aggiornare automaticamente i profili. Quindi i feedback degli utenti sulle raccomandazioni ricevute vengono raccolti ed il processo di apprendimento viene nuovamente eseguito così che il profilo tenga conto degli interessi aggiornati degli utenti vista la natura dinamica delle loro preferenze.

¹² Burke 2002

L'utilizzo di un sistema di questo tipo ha alcuni vantaggi come:

- indipendenza dagli altri utenti, poiché per costruire il profilo di un determinato utente servono solo le informazioni su quest'ultimo e non anche su altri utenti (come avviene ad es. nei metodi collaborativi);
- trasparenza, all'utente possono essere mostrate esplicitamente le caratteristiche dell'articolo che hanno causato la sua presenza nell'elenco delle raccomandazioni;
- possono essere consigliati anche articoli nuovi che ancora non hanno ricevuto valutazioni da nessuno, cosa che non può essere fatta se si utilizza un sistema basato su metodi collaborativi.

Oltre a questi vantaggi però, ci sono anche alcune limitazioni che riguardano tale metodo:

- analisi degli articoli limitata: è difficile per i sistemi riuscire ad identificare tutte le caratteristiche degli oggetti nel catalogo, e nessun sistema di raccomandazione basato sul contenuto può fornire suggerimenti adeguati se non ci sono informazioni sufficienti su tali caratteristiche per dividere gli articoli che l'utente potrebbe apprezzare da quelli che non gradirebbe;
- eccessiva specializzazione: tali sistemi non possono in alcun modo suggerire qualcosa di nuovo ed inaspettato agli utenti, poiché consiglieranno solo articoli simili a quelli già valutati;
- nuovi utenti, i sistemi hanno bisogno di raccogliere sufficienti valutazioni prima di poter costruire un profilo utente adeguato e fornire raccomandazioni precise. Pertanto, quando sono disponibili poche valutazioni, come per un nuovo utente, il sistema non sarà in grado di fornire raccomandazioni affidabili.

Queste sono le caratteristiche principali dei sistemi di raccomandazione che utilizzano dei metodi basati sui contenuti, nel paragrafo successivo saranno analizzati i metodi collaborativi.

1.3 Metodi collaborativi

Come accennato in precedenza, i metodi collaborativi nascono dalla volontà di consigliare articoli all'utente attivo basandosi sui "consigli" degli altri utenti, come avviene offline con il passaparola. I sistemi di raccomandazione basati su questi metodi ovviamente sono superiori rispetto al semplice passaparola offline poiché utilizzano migliaia se non milioni di opinioni di utenti per consigliare all'utente attivo gli articoli che sono stati maggiormente apprezzati da utenti a lui simili.

L'idea chiave è che la valutazione dell'utente u per un nuovo elemento è probabile che sia simile a quella di un altro utente v , se u e v hanno valutato altri articoli in modo simile. Allo stesso modo, è probabile che valuti due articoli i e j in modo simile, se altri utenti hanno dato valutazioni simili a questi due elementi.

Più formalmente, l'utilità $u(c,s)$ dell'articolo s per l'utente c è stimata basandosi sull'utilità $u(j,s)$ assegnata all'articolo s dall'utente $j \in C$ che risulta essere simile all'utente attivo c .

Anche in questo caso le raccomandazioni possono essere basate su feedback sia impliciti che espliciti. Il feedback implicito, come suggerisce il termine, è il feedback che viene acquisito automaticamente senza uno sforzo dell'utente. Esempi di questo sono i clic effettuati da un utente, la quantità di tempo che un utente rimane su una pagina, l'atto di acquistare un prodotto, ecc. Il feedback esplicito è un feedback che viene dato coscientemente ed esplicitamente dall'utente. Questo feedback è solitamente sotto forma di valutazioni, ma può anche contenere recensioni e forme simili.

La maggior parte dei sistemi di raccomandazione basati sulla collaborazione utilizzano entrambe le forme di feedback per prevedere se un utente sarà interessato a un articolo e classificano i diversi articoli in base alla loro valutazione prevista.

Gli algoritmi di filtraggio collaborativo sono tipicamente suddivisi in *user-to-user* e *item-to-item collaborative filtering*.

1.3.1 Metodi user-to-user

In questi metodi, il sistema, come prima fase, ricerca gli utenti più simili all'utente attivo creando un insieme di user simili tra loro, successivamente valuta i rating di tali utenti in relazione ad un articolo non valutato dall'utente attivo ed infine crea una lista di possibili raccomandazioni da mostrargli.

Una delle misure più comuni per calcolare la similitudine tra due utenti a e b , è il coefficiente di correlazione di Pearson definito così:

$$w_{a,b} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{b,i} - \bar{r}_b)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \sum_{i \in I} (r_{b,i} - \bar{r}_b)^2}}$$

dove I è l'insieme degli articoli che sono stati valutati da entrambi gli user, $r_{a,i}$ è il rating dato dall'user a all'articolo i e \bar{r}_a è la valutazione media data dall'utente a .

I sistemi di raccomandazione inoltre sono in grado di ipotizzare la valutazione che l'utente attivo darà ad un determinato item, tali previsioni sono generalmente calcolate come segue:

$$p_{ai} = \bar{r}_a + \frac{\sum_{b \in K} (r_{b,i} - \bar{r}_b) (w_{a,b})}{\sum_{b \in K} w_{a,b}}$$

dove $p_{a,i}$ è la previsione di valutazione dell'utente a rispetto all'articolo i, \bar{r}_a è il rating medio dato dall'user a, K è l'insieme degli utenti simili tra loro, $r_{b,i}$ è il rating dato dall'user b all'item i, e $w_{a,b}$ è il coefficiente di similitudine tra gli utenti a e b.

Un esempio pratico di quanto appena descritto può essere il seguente:

| | Forrest Gump | 300 | Inception | Titanic | Valutazione media |
|----------|-----------------|-----|-----------|---------|----------------------|
| Federico | 5 | | 4 | 3 | 4,00 |
| Arianna | 3 | 5 | 2 | 5 | 3,75 |
| Marco | | 3 | 5 | 4 | 4,00 |
| Carlo | 4 | 5 | 4 | | 4,33 |
| Eva | 4 | 4 | | 3 | 3,67 |

Questa tabella mostra le votazioni assegnate a 4 film da 5 utenti differenti e la valutazione media per ogni utente. Partendo da questi dati è poi possibile calcolare la similitudine tra i diversi utenti che risulta essere la seguente:

| | Federico | Arianna | Marco | Carlo | Eva |
|----------|----------|---------|-------|-------|-------|
| Federico | 1 | -0,63 | 0,00 | -0,43 | 1,30 |
| Arianna | -0,63 | 1 | -0,94 | 0,91 | -0,47 |
| Marco | 0,00 | -0,94 | 1 | -1,30 | -0,43 |
| Carlo | -0,43 | 0,91 | -1,30 | 1 | 0,25 |
| Eva | 1,30 | -0,37 | -0,43 | 0,25 | 1 |

Dove più i valori sono maggiori di zero e più gli utenti sono simili tra loro e viceversa più i valori sono negativi e minore sarà la similitudine tra gli user. Quindi, ad esempio, Eva e Federico sono molto simili tra loro, mentre Marco e Carlo hanno gusti molto differenti.

Partendo da questi dati, il sistema può provare e predire la valutazione di un utente su di un film che ancora non ha guardato utilizzando la formula mostrata in precedenza.

Quindi se ad esempio si vuole predire la valutazione di Eva per il film Inception:

$$3,67 + \frac{[1,30(4 - 4) - 0,37(2 - 3,75) - 0,43(5 - 4) + 0,25(4 - 4,33)]}{(1,3 - 0,37 - 0,43 + 0,25)}$$

ed il risultato sarebbe 3,85, simile alla valutazione data da Federico, utente con il quale Eva ha una similarità pari a 1,30.

Questo approccio *user-to-user* fornisce delle raccomandazioni di qualità ma uno dei problemi è la sua lentezza nel caso in cui ci siano numerosi utenti. L'approccio *item-to-item* invece ha un tempo di risposta minore ed offre delle raccomandazioni di altro livello.

1.3.2 Metodi item-to-item

Gli algoritmi item-to-item creano correlazioni tra coppie di articoli e quindi generano raccomandazioni trovando elementi simili al set di elementi che l'utente attivo ha apprezzato (implicitamente o esplicitamente). Questo approccio non va confuso con i metodi *content-based* in quanto le similitudini tra gli oggetti vengono estrapolate dalle interazioni passate tra gli utenti ed il sistema. In altre parole, se l'approccio *user-to-user* si basa su utenti simili che apprezzano elementi simili, l'approccio *item-to-item* si basa sugli utenti che apprezzano articoli simili agli articoli che hanno preferito.

Quindi, in questo caso, se si vuole calcolare la valutazione che l'utente a darà all'articolo i la formula sarà la seguente:

$$p_{ai} = \frac{\sum_{j \in K} (r_{a,j})(w_{i,j})}{\sum_{j \in K} (w_{i,j})}$$

dove $p_{a,i}$ è la previsione di valutazione dell'utente a rispetto all'articolo i, K è l'insieme degli articoli simili tra loro, $r_{a,i}$ è il rating dato dall'user a all'item j, e $w_{i,j}$ è il coefficiente di similitudine tra gli item i e j.

Quindi, in entrambi i metodi, quello *user-to-user* e *item-to-item*, l'idea è quella di generare raccomandazioni in base alle preferenze degli utenti con gusti simili all'utente corrente, cioè ignorare le caratteristiche dell'articolo e sfruttare le preferenze collettive, e in tali approcci, due elementi sono simili se diversi utenti del sistema hanno valutato questi articoli in modo simile.

1.3.3 Criteri per la scelta tra metodo *user-to-user* o *item-to-item*

Quando bisogna scegliere tra quale dei due approcci utilizzare, se quello *user-to-user* o quello *item-to-item*, bisogna tener presenti 5 criteri:

- **precisione:** l'accuratezza dei metodi di raccomandazione dipende principalmente dal rapporto tra il numero di utenti e di articoli all'interno del sistema. Come detto, la somiglianza tra gli utenti nell'approccio *user-to-user* viene calcolata confrontando le valutazioni effettuate dagli utenti sugli stessi articoli, nel metodo *item-to-item*, invece, la somiglianza tra due articoli è calcolata confrontando le valutazioni di uno stesso utente su questi articoli. Quindi, nei casi in cui il numero degli utenti è molto maggiore rispetto a quello degli articoli, come accade nei maggiori siti di e-commerce come Amazon.com, i metodi *item-to-item* sono da preferire poiché effettuano delle raccomandazioni più accurate, allo stesso modo, nel caso in cui il sistema abbia un numero inferiore di utenti rispetto agli articoli, i metodi *user-to-user* saranno da preferire;
- **efficienza:** anche la memoria e l'efficienza computazionale dei sistemi di raccomandazione dipendono dal rapporto tra il numero di utenti e articoli. Pertanto, quando il numero di utenti supera il numero di articoli, come spesso accade, gli approcci basati sugli item richiedono molta meno memoria e tempo per calcolare la similarità rispetto a quelli basati su utenti. In pratica, calcolare la similitudine tra gli articoli è molto meno costoso poiché gli utenti valutano solo alcuni degli articoli disponibili. Di conseguenza, è necessario memorizzare solo i coefficienti di similarità diversi da zero, che sono spesso molto inferiore rispetto al numero di coppie di utenti;
- **stabilità:** la scelta tra un approccio piuttosto che un altro dipende anche dalla frequenza e dalla quantità di cambiamenti negli utenti e negli elementi nel sistema. Se l'elenco di articoli disponibili è abbastanza statico rispetto a quello degli utenti, il metodo *item-to-item* potrebbe essere preferibile in quanto la similitudine degli articoli potrebbe essere calcolata a intervalli di tempo non frequenti. Al contrario, nelle applicazioni in cui la lista degli articoli disponibili è in continua evoluzione, ad es. un *recommendation system* di articoli online, i metodi basati sugli utenti potrebbero rivelarsi più stabili;
- **giustificabilità:** un vantaggio dei metodi basati sugli articoli è che possono facilmente giustificare una raccomandazione, l'elenco degli elementi simili usati per la raccomandazione, così come i coefficienti di similarità, possono essere mostrati all'utente come spiegazione del consiglio fornitogli;
- **serendipità:** nei metodi *item-to-item*, la valutazione prevista per un articolo si basa sul rating assegnato ad oggetti simili. Di conseguenza, i sistemi di raccomandazione che utilizzano questo approccio, tenderanno a raccomandare ad un utente elementi correlati a quelli da lui solitamente apprezzati. Ad esempio, nel caso di una raccomandazione di un film, è probabile che i film con lo stesso genere, attori o regista, di quelli votati positivamente dall'utente, gli siano raccomandati. Anche se questo può portare

a raccomandazioni sicure, non aiuta però l'utente a scoprire diversi tipi di articoli che gli potrebbero piacere. Dal momento che funzionano con la similarità degli utenti, gli approcci basati sugli user hanno maggiori probabilità di fare raccomandazioni meno banali. Questo è particolarmente vero se la raccomandazione è fatta con un piccolo numero di utenti simili. Ad esempio, un utente A che ha guardato solo commedie può essere molto simile a un utente B solo dalle valutazioni fatte su film di questo genere, tuttavia, se B è appassionato di un film di un genere diverso, questo film può essere raccomandato ad A grazie alla sua somiglianza con B dovuta alle commedie.

Quindi, in generale, nei tipici sistemi di raccomandazione sui siti di e-commerce, dove il numero di utenti supera di gran lunga il numero di articoli disponibili, gli approcci basati sugli articoli sono in genere preferiti poiché forniscono consigli più accurati, sono più efficienti dal punto di vista computazionale e richiedono aggiornamenti meno frequenti. D'altra parte, i metodi basati sugli utenti, forniscono raccomandazioni più originali, che possono portare gli utenti ad un'esperienza più soddisfacente.

1.3.4 Vantaggi e limiti dei metodi user-to-user e item-to-item

I principali vantaggi collegati a questi due metodi sono quattro:

- semplicità: sono dei metodi intuitivi e facili da implementare, nella loro forma più semplice solo un parametro, il numero di utenti simili da utilizzare nella previsione, necessita la messa a punto;
- giustificabilità: questi due approcci forniscono una spiegazione per le raccomandazioni che vengono fornite, aiutando l'utente attivo a comprendere meglio questi consigli e a valutarne la pertinenza;
- efficienza: questi metodi non richiedono costose fasi di training, la fase della raccomandazione, che è quella che richiede maggior tempo, può essere migliorata calcolando offline i gruppi di utenti simili tra loro in modo tale da rendere le raccomandazioni quasi istantanee, ed infine la memorizzazione di questi gruppi di user richiede pochissima memoria rendendo tali metodi utili anche in caso di elevati numeri di user ed articoli;
- stabilità: un'altra utile proprietà dei sistemi di raccomandazione basati su questi approcci è che sono poco influenzati dall'aggiunta costante di utenti, articoli e valutazioni. Ad esempio, una volta che le similitudini degli articoli sono state calcolate, un sistema item-to-item può facilmente dare consigli ai nuovi utenti, senza dover ri-addestrare il sistema. Inoltre, dopo aver inserito alcune valutazioni per un nuovo articolo, è necessario calcolare solo le somiglianze tra questo elemento e quelli già presenti nel sistema;

- raccomandazioni meno banali: tali sistemi permettono di superare uno dei problemi dei sistemi *content-based*, quello della eccessiva specializzazione, poiché usano le valutazioni degli utenti e quindi possono raccomandare degli articoli differenti rispetto a quelli apprezzati in passato dall'utente attivo.

Ad ogni modo però anche questi sistemi hanno delle limitazioni:

- nuovi utenti: questa è una limitazione che affligge anche i sistemi *content-based*, ed è collegata con il fatto che per fornire delle raccomandazioni accurate, il sistema deve prima apprendere le preferenze dell'utente espresse attraverso le sue valutazioni;
- nuovi articoli: i metodi collaborativi si basano sulle preferenze degli utenti, quindi se un item non riceve voti sufficienti il sistema non sarà in grado di raccomandarlo;
- *sparsity* (scarsa densità): spesso nei sistemi di raccomandazione ci sono molti oggetti che vengono votati solo da pochi utenti, ad esempio ci sono alcuni libri o film che vengono votati raramente. Inoltre, per gli utenti che hanno dei gusti insoliti rispetto a quelli degli altri, può essere complicato trovare dei membri che risultano essere simili, e quindi le raccomandazioni per questi utenti risulteranno essere poco accurate. Per cercare di risolvere questo problema si possono prendere in considerazione le informazioni del profilo degli utenti. In questo modo due utenti saranno simili sia se hanno gli stessi gusti, sia se i due profili si assomigliano, ad esempio hanno più o meno la stessa età, vivono in zone vicine ecc. Questo metodo differente di sistema collaborativo prende il nome di *demographic filtering*¹³.

Se per il problema della scarsa densità una soluzione potrebbe essere quello dell'utilizzo del *demographic filtering*, per le prime due limitazioni una soluzione potrebbe essere quella di utilizzare i sistemi ibridi, che utilizzano contemporaneamente sia i metodi collaborativi che quelli basati sul contenuto. Tali metodi saranno spiegati nel paragrafo successivo.

1.3.5 Latent factor model e matrix factorization

Altro metodo appartenente ai sistemi di raccomandazioni basati sul filtraggio collaborativo è il *latent factor model*. Questo è un approccio alternativo che cerca di spiegare le valutazioni degli utenti attraverso la ricerca di fattori latenti utili per costruire i profili di user ed elementi. Ad es. per i film, tali fattori latenti potrebbero essere il genere, thriller o commedia, la quantità di azione, se è più adatto agli uomini, alle donne o ai

¹³ M.Pazzani 1999

bambini, lo sviluppo dei personaggi ecc. Le applicazioni di maggior successo del *latent factor model* si basano sulla tecnica chiamata *matrix factorization*. Quest'ultima ci consente di scoprire i fattori latenti nascosti che sono alla base delle valutazioni date dagli utenti.

In particolare il *matrix factorization model* mappa sia gli utenti che gli articoli in uno spazio congiunto di dimensionalità f e le interazioni utente-item sono interne a questo spazio. Ogni articolo i è associato ad un vettore $q_i \in \mathbb{R}^f$ e ogni utente u è associato ad un vettore $p_u \in \mathbb{R}^f$. Per un determinato elemento i , gli elementi di q_i misurano la misura in cui l'elemento possiede quei fattori, positivi o negativi. Per un determinato utente u , gli elementi di p_u misurano la misura in cui l'utente è interessato agli elementi che hanno determinati fattori. Infine, il prodotto tra i due vettori, mostra l'interazione che si ha tra l'utente u e l'elemento i e l'interesse generale dell'utente per le caratteristiche dell'oggetto, e questo aiuta a stimare la valutazione dell'utente per l'articolo i ¹⁴.

Ad es., supponiamo di avere 5 utenti e 6 film, e la valutazione è di numeri interi tra 1 e 5, il compito della *matrix factorization* è quello di prevedere le valutazioni mancanti. L'idea alla base di questa tecnica è che ci sono caratteristiche nascoste negli elementi che determinano come un utente valuta un film. Ad es., un utente che dà a due film con Di Caprio una valutazione elevata, potrebbe significare che una caratteristica nascosta potrebbe essere che apprezzi Di Caprio come attore.

Se si riescono a scoprire questi fattori latenti, si può prevedere una valutazione di un film, non ancora votato da un utente.

1.4 Metodi ibridi

Il fatto che tutti gli algoritmi utilizzati nei sistemi di raccomandazione hanno i loro punti di forza e di debolezza, ha portato a molti tipi di combinazioni per combattere i punti deboli e sfruttare quelli di forza di ognuno di loro. Quindi, oltre a coprire le debolezze, gli approcci vengono combinati per essere in grado di sfruttare i punti di forza dei diversi metodi a disposizione, ad es. si possono sfruttare congiuntamente sia gli attributi dell'articolo che l'utente ama (filtraggio basato sul contenuto), sia le preferenze della comunità (filtraggio collaborativo).

Un sistema di raccomandazione ibrido, quindi, è qualsiasi sistema di raccomandazione che combini insieme più tecniche di raccomandazione per produrre i suoi risultati¹⁵.

In particolare, i differenti modi di combinare i metodi collaborativi e *content-based* in un sistema di raccomandazione ibrido possono essere quattro:

¹⁴ Yehuda Koren 2009

¹⁵ Burke 2007

- implementare i metodi collaborativi e *content-based* separatamente e combinare le loro previsioni: questo approccio porta ad una scelta tra due possibili modi di proseguire dopo l'implementazione dei due sistemi, o si possono combinare i risultati dei due sistemi in un singolo risultato finale ad esempio attraverso una combinazione lineare dei due voti predetti, o si può scegliere il *recommendation system* che è in grado di fornire la migliore raccomandazione, basandosi su delle metriche di qualità delle raccomandazioni¹⁶. Un esempio per il primo caso potrebbe essere che il metodo collaborativo fornisce una stima della valutazione dell'utente attivo pari a 3,5, mentre quello *content-based* ne fornisce una pari a 4, allora si possono combinare questi due voti facendone una media, ed ottenendo una stima della valutazione pari a 3,75. Esempio per il secondo metodo invece è il sistema DailyLearner¹⁷, il quale seleziona il sistema di raccomandazione che fornisce i consigli con il livello di confidenza più elevato;
- incorporare alcune caratteristiche dei sistemi *content-based* in un metodo collaborativo: in questo caso si ottiene un sistema con le tradizionali tecniche dei sistemi collaborativi che però mantiene i profili degli utenti utilizzati nei sistemi basati sul contenuto. In questo modo per calcolare la somiglianza tra due utenti vengono usati non solo gli oggetti che sono stati votati da entrambi, ma anche i loro profili. Con questo metodo si riesce ad evitare il problema della scarsità dei voti, ed inoltre all'utente possono essere raccomandati non solo gli oggetti altamente apprezzati dagli utenti simili, ma anche quegli oggetti che si abbinano al profilo dell'utente;
- incorporare alcune caratteristiche dei sistemi collaborativi in un metodo *content-based*: in questo caso l'approccio più utilizzato è quello di far ricorso a delle tecniche di *dimensionality reduction* su un gruppo di profili utente generati con tecniche *content-based*. Ad es. partendo da un insieme di profili utenti, che contengono molte caratteristiche e sono perciò multidimensionali, questi vengono rappresentati come dei vettori di dimensione inferiore sui quali poi si applicano le tecniche tipiche di un sistema *content-based*;
- costruire un modello unificato di raccomandazione: qui vengono unificate le caratteristiche dei due metodi in un modello generale. Un esempio¹⁸ di questi sistemi utilizza le informazioni degli utenti e degli oggetti in un unico modello statistico per stimare i voti sconosciuti.

In conclusione, quindi, si può dire che i metodi ibridi presentano in parte tutti i vantaggi dei sistemi *content-based* e collaborativi, anche se ogni vantaggio è limitato poiché inserito in un approccio non puro. Allo stesso modo si può dire che il limite principale sia il fatto che i vantaggi derivanti dall'uno e dall'altro approccio siano limitati da questo utilizzo congiunto.

¹⁶ Adomavicius, Tuzhilin 2005

¹⁷ Billsus, Pazzani 2000

¹⁸ Ansari, Essegaier, Kohli 2000

1.5 Benefici dei sistemi di raccomandazione

I sistemi di raccomandazione possono essere considerati vantaggiosi sia per i service provider che per gli utenti.

In particolare ci sono diversi motivi per cui un venditore può scegliere di utilizzare tali sistemi:

- aumentare il numero di articoli venduti attraverso il miglioramento del tasso di conversione, ovvero il numero di utenti che, visitano il sito web, accettano la raccomandazione ed acquistano l'articolo;
- vendere articoli differenti e non solo i più popolari: ad es. un sito di e-commerce ha interesse a vendere tutti i suoi prodotti e non solo quelli più richiesti, e grazie ai sistemi di raccomandazione questo è possibile consigliando ad un utente potenzialmente interessato quel determinato prodotto meno famoso. Questa è anche alla base del *long tail concept*, secondo il quale i ricavi vengono ottenuti non solo con la vendita di molte unità di pochi oggetti (i best seller), ma anche vendendo pochissime unità di tantissimi oggetti diversi, quindi, un sito web, che ha meno costi di stoccaggio rispetto ad un negozio fisico, può avere anche i prodotti meno richiesti nel catalogo ed aumentarne anche il prezzo, vista la difficoltà di reperirli offline, e può venderli semplicemente consigliandoli ai potenziali utenti interessati grazie ai *reccomandation systems*;
- aumentare la soddisfazione dei clienti: attraverso l'offerta di un'esperienza, sul sito o sull'applicazione, *user friendly*, grazie a delle raccomandazioni interessanti e pertinenti che portino l'utente e riutilizzare il sistema ed accettarne i consigli;
- aumentare la fedeltà dell'utente attraverso la costruzione di uno store online completamente personalizzato, che faccia sentire l'user un visitatore prezioso. Inoltre, se la sua fedeltà aumenta, migliorano anche le raccomandazioni poiché il sistema ha più dati a disposizione per soddisfare le preferenze dell'utente;
- aumentare il cross-sell: i sistemi di raccomandazione sono particolarmente adatti a suggerire prodotti aggiuntivi o correlati ai clienti, e se le raccomandazioni sono buone, il risultato dovrebbe essere un aumento delle vendite.

Queste sono alcune importanti motivazione che spiegano il perché i service provider utilizzano i *recommendation systems*, ma anche gli utenti potrebbero desiderare tali sistemi per riuscire a migliorare la propria soddisfazione, quindi tali sistemi dovranno essere in grado di bilanciare le esigenze di questi due *player* offrendo un servizio che sia prezioso per entrambi.

I sistemi di raccomandazione sono per natura incentrati sull'utente, poiché il valore del cliente è il motore dell'azienda e forniscono servizi a valore aggiunto ai consumatori.

I principali vantaggi di tali sistemi per i consumatori sono:

- una maggior sicurezza nel momento in cui acquistano un oggetto o compiono altre scelte online, in questo senso hanno anche una funzione persuasiva nel processo decisionale del consumatore;
- il vantaggio principale delle raccomandazioni è quello di ridurre i costi di transazione, i quali si hanno perché il processo decisionale è soggetto a un limitato potere di elaborazione cognitiva e non tutti gli scenari possibili possono essere considerati, in parte a causa della razionalità limitata. Nel processo che porta all'acquisto di un prodotto, un utente deve passare attraverso un processo di transazione che inizia con una ricerca di elementi pertinenti ed è seguito da una fase di confronto con altri articoli (ad es. confronti dei prezzi e valutazioni della qualità degli articoli) e domande relative all'ordine, alla consegna e ai servizi post-vendita, a seconda della natura dell'articolo. Tale processo quindi porta incertezze sul prodotto, e tale sensazione si riferisce alla possibilità che un consumatore scopra, dopo aver consumato il prodotto, ad esempio leggendo un libro, che ciò che ha ricevuto è diverso da quello che ha percepito al momento dell'ordine, in generale l'incertezza si riferisce ai costi connessi a esiti imprevisti dovuti all'asimmetria informativa tra venditore e compratore.

Elevata incertezza significa alti costi di transazione, e l'inclinazione naturale del consumatore è giudicare la qualità intrinseca del prodotto sulla base di tutte le informazioni disponibili e scegliere l'articolo con il più basso costo di transazione, cioè con la più bassa incertezza.

La teoria della riduzione dell'incertezza, infatti, afferma che di fronte a quest'ultima, il consumatore tenderà di ridurla per eliminare o almeno mitigare i rischi e massimizzare il valore del risultato.

I sistemi di raccomandazione sono in grado di ridurre i costi di transazione¹⁹, possono aiutarci a localizzare in modo efficiente gli articoli interessanti, riducendo così il numero di articoli da sfogliare e possono ridurre l'incertezza sui prodotti, ad es. mostrando le recensioni dei clienti, riducendo anche il tempo complessivo necessario per prendere una decisione²⁰. In altre parole, i sistemi di raccomandazione ci assistono in compiti impegnativi dal punto di vista cognitivo.

Di conseguenza, è stato dimostrato che i sistemi di raccomandazione migliorano la qualità delle decisioni²¹;

- altro valore aggiunto offerto all'utente dai sistemi di raccomandazione è la personalizzazione. I siti web si adattano a ciascun consumatore offrendogli un'esperienza di acquisto completamente personalizzata. Un esempio ne è Amazon.com, dove è praticamente impossibile che due utenti vedano esattamente gli stessi contenuti nella pagina iniziale, poiché è costruita basandosi sulle informazioni che il sistema ha sull'user (acquisti effettuati, pagine visitate ecc...).

¹⁹ Hu, Liu e Zhang, 2008; Pathak et al., 2010

²⁰ Schafer, Konstan, & Riedl, 2001; Xiao & Benbasat, 2007; Hu, Liu, & Zhang, 2008; Campochiaro et al., 2009; Pathak et al., 2010

²¹ Xiao e Benbasat, 2007; Pathak et al., 2010

La personalizzazione, ovviamente, dipende dai dati a disposizione del sistema, e quanto più le informazioni sull'utente sono dettagliate, tanto più lo possono essere le raccomandazioni.

Uno dei problemi che riguarda tale valore aggiunto è la *sparsity* dei dati, uno stato in cui non c'è abbastanza sovrapposizione tra le valutazioni degli utenti, e quindi è difficile trovare similitudini tra gli utenti stessi. Prima dell'esistenza di un numero sufficiente di dati per attivare la personalizzazione, molti sistemi commerciali utilizzano delle raccomandazioni non personalizzate, come ad esempio approcci contestuali, demografici e articoli più popolari.

Altro aspetto da considerare quando si parla di personalizzazione è che il tipo di utente può influenzare la soddisfazione riguardo le raccomandazioni, in quanto soprattutto i principianti e gli utenti che aspirano al miglior risultato possibile, traggono più beneficio dai consigli non personalizzati che mostrano semplicemente gli articoli più popolari²².

Allo stesso modo nei siti in cui la maggior parte degli articoli hanno approssimativamente la stessa qualità un suggerimento che mostra semplicemente gli elementi più popolari potrebbe funzionare bene²³, inoltre se tutti gli utenti sono in generale d'accordo sulla qualità degli articoli, le raccomandazioni personalizzate possono aggiungere poco valore²⁴.

Inoltre si è visto come il grado di personalizzazione percepita influenzi direttamente la fiducia cognitiva ed emotiva e che la fiducia emotiva media l'impatto della fiducia cognitiva sull'intenzione di adottare il sistema di raccomandazione²⁵. Quindi, la personalizzazione può anche influire sul fatto che l'utente utilizzi o meno il sistema e si fidi dei suoi suggerimenti.

²² Knijnenburg, Reijmer e Willemsen 2011

²³ Svensson, Höök e Cöster 2005

²⁴ Recker, Walker e Lawless 2003

²⁵ Komiak e Benbasat, 2006

2 Sentiment analysis

La *Sentiment Analysis* è l'area di ricerca che studia opinioni, sentimenti, valutazioni, attitudini ed emozioni delle persone riguardo entità ed i loro aspetti espresse attraverso un testo scritto²⁶.

Essendo uno strumento di raccolta di informazioni, è inevitabilmente una leva di quella funzione aziendale che, come definiva Philip Kotler, “individua esigenze e desideri insoddisfatti, definisce e misura la loro rilevanza e redditività potenziale, determina quali mercati obiettivo l'organizzazione può servire meglio, decide in merito ai prodotti, servizi e programmi appropriati per servire i mercati individuali e fa appello a tutte le componenti dell'organizzazione per assistere il cliente e servirlo” e che generalmente è chiamata marketing.

Tale tecnica conferisce alle aziende la possibilità di ottenere informazioni generate direttamente dai consumatori, senza che siano influenzati in alcun modo dall'impresa, quindi informazioni utili anche verso altre funzioni aziendali oltre a quella di marketing.

Vista la costante crescita di internet ed in generale di modelli che pongono l'utente al centro della creazione dei contenuti presenti all'interno della rete, si riscontra un sempre maggiore interesse da parte di aziende e organizzazioni riguardo la possibilità di utilizzare questo tipo di dati al fine di ideare, progettare, e monitorare le proprie strategie di marketing. In una situazione del genere, in cui gli user sempre più spesso si rivolgono alla rete per avere informazioni riguardo un prodotto, le sue caratteristiche e le opinioni degli acquirenti, molte aziende si trovano davanti alla necessità di analizzare la *customer satisfaction*, espressa in forma scritta attraverso la rete.

Per il SAS Institute i maggiori benefici per i business portati dalla *sentiment analysis* sono²⁷: la capacità di monitoraggio delle opinioni in tempo reale, la possibilità di ottenere feedback e nuove idee, il poter avere sempre sotto controllo la propria posizione e quella dei concorrenti all'interno dell'ambiente competitivo e la possibilità di quantificare le percezioni del mercato.

Volendo rappresentare il problema della *sentiment analysis* in una maniera più formale, potremmo dire che l'entità e , che può essere un prodotto, un servizio, una persona, un evento o una organizzazione, è associata ad una coppia (T, W) dove T rappresenta i componenti di e , e W è l'insieme delle caratteristiche o attributi di e . Prendiamo come esempio un cellulare che è un'entità, ed è formato da un insieme di componenti come batteria, schermo, casse ecc., e possiede un insieme di caratteristiche come la risoluzione, la qualità delle casse, il peso ecc. Partendo da ciò, un'opinione, un sentimento, può essere definita come una quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ ²⁸, dove:

- e_i rappresenta l'entità sulla quale è stato espresso un parere;

²⁶ Liu 2012

²⁷ http://www.sas.com/en_us/software/analytics/sentiment-analysis.html

²⁸ Liu 2012

- a_{ij} è un aspetto dell'entità sul quale viene espressa l'opinione, se quest'ultima riguarda l'entità nel suo complesso allora l'aspetto sarà generale;
- s_{ijkl} rappresenta l'opinione riguardo l'aspetto a_{ij} dell'entità e_i e può essere positiva, negativa o neutrale e può essere espressa in modi differenti;
- h_k è colui che esprime l'opinione;
- t_i ci dice quando h_k ha espresso l'opinione.

In merito a questa definizione si devono tener presenti cinque punti essenziali:

- ogni parte della quintupla deve essere collegata con le altre, ogni discrepanza è un errore che non rende possibile l'estrazione del sentimento, quindi l'opinione s_{ijkl} deve essere fornita da colui che esprime l'opinione (h_k) riguardo l'aspetto a_{ij} dell'entità e_i al tempo t_i ;
- ogni parte della quintupla è essenziale, la mancanza di una di queste è un problema per la *sentiment analysis*, ad es. la mancanza della componente temporale rende impossibile l'analisi dell'entità in base al tempo, il quale è essenziale poiché una opinione di tre anni fa non ha lo stesso peso di una espressa una settimana fa;
- la definizione dell'opinione come una quintupla è una semplificazione che copra la maggior parte, ma non tutte le possibili sfaccettature del significato semantico di un sentimento, poiché quest'ultimo a volte può essere espresso in maniera complicata. Ad esempio, la frase “questa macchina è troppo piccola per le persone alte” ci dice che la macchina non è troppo piccola per chiunque, ma solo per le persone alte, che in questo caso rappresentano il contesto dell'opinione. Quindi, in definitiva, questa definizione è semplificata poiché se fosse troppo complessa renderebbe il problema *della sentiment analysis* troppo complicato da risolvere;
- questa definizione consente di trasformare dei dati non strutturati, come quelli testuali, in dati strutturati sui quali è possibile effettuare diverse analisi sia qualitative che quantitative;
- l'opinione definita da questa quintupla è solo una tipologia di opinione possibile, un altro tipo è quello delle opinioni comparative.

In generale, la *sentiment analysis* può essere applicata su tre livelli differenti:

- livello di documento: in questo caso l'obiettivo è valutare se il sentimento di un intero documento è positivo, negativo o neutrale. Ad esempio se si tratta di una recensione di un prodotto si valuta com'è l'opinione nei confronti di quest'ultimo. Questo livello di analisi, però, presuppone che all'interno di un documento si discuta soltanto di una entità, e non può essere utilizzato nel caso in cui ce ne siano di più in un singolo testo;

- livello di frase: a questo livello il compito è quello di determinare l'opinione di ogni singola frase presente in un documento. Questo livello di analisi è strettamente correlato alla classificazione della soggettività²⁹, che distingue le frasi in oggettive, che esprimono informazioni fattuali, e soggettive, che esprimono delle opinioni soggettive;
- livello di entità e aspetto: l'obiettivo di questo livello di analisi è scoprire i sentimenti sulle entità e/o sui loro aspetti. Ad esempio, la frase "La qualità delle chiamate dell'iPhone è buona, ma la durata della batteria è breve" valuta due aspetti, la qualità della chiamata e la durata della batteria dell'iPhone che è l'entità. Il sentimento sulla qualità delle chiamate del cellulare è positivo, ma il sentimento sulla durata della batteria è negativo. Basandosi su questo livello di analisi, è possibile produrre un riassunto strutturato delle opinioni sulle entità e sui loro aspetti, che trasforma il testo non strutturato in dati strutturati che possono essere utilizzati per tutti i tipi di analisi qualitative e quantitative.

2.1 Differenti tipologie di opinioni

Come accennato in precedenza ci sono diverse tipologie di opinioni, opinioni regolari, opinioni comparative, opinioni esplicite e opinioni implicite.

Le opinioni regolari a loro volta possono essere divise in opinioni dirette, se vi è un parere espresso direttamente su una entità o un suo aspetto, e opinioni indirette, caso in cui il sentimento è espresso indirettamente sull'entità o su un aspetto della stessa, un esempio potrebbe essere la frase "dopo che ho preso quel farmaco mi è venuto mal di stomaco", qui viene descritto un effetto indesiderato del farmaco sullo "stomaco", che indirettamente dà un'opinione negativa sul farmaco. In questo esempio, l'entità è il farmaco e l'aspetto è l'effetto sullo stomaco.

Le opinioni comparative, invece, esprimono una relazione di similarità o di differenza tra due o più entità e/o una preferenza di colui che esprime l'opinione basata su alcuni aspetti condivisi delle entità. Solitamente tale tipologia di opinione è espressa utilizzando la forma comparativa o superlativa di un aggettivo o un avverbio. In particolare si può distinguere tra opinioni comparative che esprimono un confronto classificando le entità, ad es. l'entità A è migliore della B, A e B sono uguali o l'entità A è la migliore di tutte, o che lo esprimono ma senza stipulare una classifica, ad es. l'entità A è simile o diversa dall'entità B in base ad alcuni aspetti condivisi, l'entità A ha l'aspetto a_1 mentre l'entità B ha l'aspetto a_2 o l'entità A ha l'aspetto a_1 che l'entità B non ha.

Altra distinzione va fatta tra opinioni esplicite ed implicite. Le prime sono delle dichiarazioni soggettive che forniscono una opinione regolare o comparativa, ad es. "La Coca-Cola è ottima" o "La coca-Cola ha un sapore

²⁹ *Wiebe, Bruce, O'Hara 1999*

migliore della Pepsi”, le seconde sono delle dichiarazioni oggettive che forniscono una opinione regolare o comparativa, esempi possono essere “questa auto consuma molta benzina” o “la durata della batteria dei telefoni Samsung è maggiore rispetto ai telefoni Apple”. Ovviamente le opinioni esplicite sono più facili da individuare e classificare rispetto a quelle implicite.

2.2 Problemi della Sentiment Analysis

Ovviamente gli indicatori principali del sentimento in una frase sono le parole d’opinione. Gli elenchi che comprendono tali parole sono chiamati *sentiment lexicon*, i quali sono necessari ma non sufficienti per effettuare una *sentiment analysis*, questo perché ci sono alcuni problemi ad essi collegati:

- il primo è dovuto al fatto che alcune parole possono avere significati opposti a seconda del contesto in cui si trovano, quindi non è facile generalizzare il problema, anzi spesso è necessario avere differenti lessici di sentimento per diversi ambiti di riferimento;
- altro problema è che in alcune tipologie di frasi, come quelle interrogative e quelle condizionali, la presenza di parole d’opinione non comporta sempre l’espressione di un sentimento, ad es. nella frase “sapresti dirmi quale cellulare Samsung è buono?” c’è la parola “buono”, una parola che esprime sentimento positivo solitamente, ma non in questo caso;
- problema opposto si ha quando ci sono alcune frasi che non contengono delle parole d’opinione ma esprimono comunque un sentimento positivo o negativo, un esempio sono le frasi che esprimono un fatto oggettivo come la seguente “questa macchina consuma molta benzina”, in questo caso non ci sono delle parole che indicano un sentimento, ma l’opinione nei confronti della macchina è palesemente negativa;
- altre difficoltà sono dovute all’uso di espressioni non formali (slang) e non appartenenti al vocabolario proprio di una lingua. Questi termini spesso vengono utilizzati per esprimere una particolare opinione o un certo stato d’animo;
- le negazioni giocano un ruolo importante nella *sentiment analysis*, frasi come "Mi piace questo film" e "Non mi piace questo film" sono molto simili, ma a causa di un termine di negazione devono essere classificate in classi opposte, tuttavia, a causa della similitudine nelle frasi e specialmente nella parola positiva "mi piace", non è semplice che ciò avvenga. Un modo possibile per gestire questa situazione è convertire il sentimento della frase se vi è la presenza di un termine di negazione. Questo metodo potrebbe sembrare abbastanza facile, ma non tutti i termini di negazione invertono la polarità della frase, ad es. nella frase "Non c'è da stupirsi se questo è considerato uno dei migliori film", non è corretto allegare "non" alla parola "migliore" e quindi invertire il sentimento;

- un'altra difficoltà collegata alla negazione è che spesso può essere espressa in modi che sono difficili da rilevare, ad esempio tramite ironia e sarcasmo. Per le persone, rilevare l'ironia può essere abbastanza difficile, ma per le macchine lo è ancora di più;
- ultimo problema è quello dello spam, possiamo utilizzare tutte le tecniche possibili per migliorare la *sentiment analysis*, ma non potremo mai sapere se l'opinione rilevata è sincera o è frutto di speculazioni personali o economiche.

2.3 Tecniche di pre-processing

Il linguaggio spesso possiede strutture e codici complessissimi ma, per fortuna, per completare un'analisi testuale efficace non è necessario riuscire ad interpretarlo e decodificarlo in tutta la sua complessità, infatti quando un testo scritto deve essere analizzato è necessario svolgere alcuni passaggi fondamentali di pre-processing allo scopo di ridurre il numero di termini presenti all'interno del testo stesso.

Quindi, per prima cosa, nel processo di analisi di un testo, bisogna cercare di ridurre il testo in un dato quantitativo, tale da poter essere trattato da un modello statistico. La fase preliminare del processo sarà eliminare tutte quelle parole e simboli utili più alla forma del linguaggio che alla sua sostanza e quindi non importanti sotto il punto di vista informativo. Questo andrà fatto per ciascun documento in modo tale da formare un insieme di testi, corpus, facilmente interpretabile dall'algoritmo e che non abbia problemi di *overfitting*, tipico di modelli che si adattano eccessivamente ai dati osservati usando un numero eccessivo di parametri e perdendo la possibilità di essere utilizzati anche in altre analisi.

Questa fase è essenziale e necessaria per tutti i processi di analisi, indipendentemente da metodi di classificazione o algoritmi utilizzati.

In altre parole si migliora l'efficacia del modello con alcuni accorgimenti come: l'uniformazione del testo in minuscolo o maiuscolo, la rimozione della punteggiatura, rimozione delle *stopwords*, ovvero quelle parole che si ripetono spesso in una determinata lingua, come articoli pronomi congiunzioni ecc. ma che non hanno un significato semantico e che quindi non servono allo scopo della *sentiment analysis*, e lo *stemming* ovvero l'azione di riportare ogni parola alla sua radice attraverso l'eliminazione dei suffissi, ad es. le parole come "andai", "andiamo", "andaste" vengono tutte sostituite con "and". Tale tecnica viene eseguita solitamente tramite programmi basati sui dizionari, essi infatti conoscono le parole di una lingua e sono in grado di ricavarne la radice. Altre metodologie per l'applicazione dello *stemming* utilizzano dizionari dei sinonimi e contrari per cercare di non mantenere troppe parole con lo stesso significato.

Il passo successivo è quello di verificare la frequenza dei termini all'interno del nostro corpus. In particolare è possibile creare una matrice dove le righe corrispondono ai documenti, ad es. dei tweet, e le colonne corrispondono alle parole in questi documenti e i valori della matrice sono dati dalla frequenza di quelle parole in quei documenti. Successivamente si dovranno andare ad identificare i termini che appaiono con meno

frequenza ed eliminare quelli che hanno una frequenza inferiore ad una determinata soglia, ad es. impostare una soglia dello 0.995 significa che si vogliono tenere solo i termini che compaiono in almeno lo 0.5% di tutti i testi. La rimozione di tali termini meno frequenti viene fatta poiché avere molti termini poco utili comporta due inefficienze principali: inefficienza computazionale, avere più termini significa avere più variabili indipendenti richiedendo così più risorse computazionali per il calcolo del sentimento, e una scarsa qualità del modello poiché il rapporto tra numero di variabili indipendenti e numero di osservazioni impatta direttamente la capacità del modello di generalizzare.

2.4 Creazione di un sentiment lexicon

Come accennato in precedenza il fulcro della *sentiment analysis* sono le parole e le frasi che esprimono sentimenti positivi o negativi. Tali parole possono essere o di tipo base o comparative, le prime sono ad esempio “buono”, “brutto”, “bello” ecc., quelle del secondo tipo sono usate per esprimere opinioni comparative e superlative, esempi sono “migliore”, “peggiore” ecc. sono stati individuati tre modi possibili per creare un *sentiment lexicon*, insieme di parole e frasi di opinione, utilizzabile poi per la *sentiment analysis* nei tre livelli che saranno esposti di seguito: un approccio manuale, uno basato sul dizionario e uno basato sul corpus. Quello manuale è particolarmente complicato da applicare e viene quasi sempre utilizzato insieme con uno degli altri due approcci automatici che possono commettere degli errori.

Un approccio basato sul dizionario, semplice da utilizzare, prevede la raccolta manuale di alcune parole d’opinione con orientamenti positivi o negativi dal corpus, e successivamente l’algoritmo aumenta questo set cercando su un qualsiasi dizionario online i loro sinonimi e contrari.

Il vantaggio dell’utilizzo di un approccio basato sul dizionario è che si possono trovare facilmente e rapidamente un gran numero di parole d’opinione sia positive che negative, anche se l’elenco risultante può avere molti errori che però è possibile ridurre tramite un controllo manuale. Il principale svantaggio è che gli orientamenti del *sentiment* delle parole raccolte in questo modo sono generali e indipendenti dal contesto, e molte parole di sentimento hanno orientamenti che dipendono dal contesto, ad esempio, per una cassa, se è silenziosa, di solito è negativo, mentre per un’auto, se è silenziosa, è positivo. L’approccio basato sul corpus può aiutare a risolvere questo problema.

L’approccio basato sul corpus parte da una lista di parole d’opinione, spesso generali, e ne scopre altre da un corpus. Una tecnica³⁰ potrebbe essere quella di sfruttare un insieme di regole linguistiche e convenzioni sulle congiunzioni per identificare le parole d’opinione nel corpus. Una delle regole riguarda la congiunzione “e”,

³⁰ Hazivassiloglou, McKeown 1997

la quale indica che gli aggettivi da lei congiunti solitamente hanno lo stesso orientamento, ad esempio, nella frase "Questa macchina è bella e spaziosa", se "bello" è positivo, si può dedurre che "spazioso" è anche positivo, questo perché le persone di solito esprimono lo stesso sentimento su entrambi i lati di una congiunzione. Tale ragionamento può essere utilizzato anche per altre congiunzioni come, o, ma, o-o, ed è chiamato coerenza sentimentale.

Estendendo questa idea alle frasi vicine, e non solo quindi alle parole nella stessa frase, si può avere la coerenza inter-sentenziale³¹ e si applica lo stesso ragionamento espresso in precedenza. Quindi le frasi congiunte con la congiunzione "e" esprimono lo stesso sentimento, mentre quelle che presentano congiunzioni come "ma" e "tuttavia", prevedono un cambiamento di sentimento da una frase all'altra.

Va notato però che non basta trovare le parole d'opinione con i loro orientamenti, poiché ci sono degli aggettivi che hanno un significato differente anche all'interno dello stesso contesto, ad es, l'aggettivo silenzioso è positivo nella frase "questa macchina è silenziosa", ma ha un'accezione negativa nella frase "lo stereo di questa auto è silenzioso". Per ovviare a questo problema si può guardare alla coppia³² (aspetto, parola d'opinione), quindi si vanno a considerare la parola d'opinione, il suo orientamento e anche l'aspetto a cui si riferiscono. Anche in questo caso però sono valide le regole relative alle congiunzioni all'interno della frase. Infine, l'utilizzo dell'approccio basato sul corpus per identificare tutte le parole di opinione, non è efficace quanto l'approccio basato sul dizionario, perché è difficile preparare un corpus enorme che comprenda tutte le parole di una data lingua. Tuttavia, come accennato in precedenza, questo approccio ha un grande vantaggio che l'approccio basato sul dizionario non ha, può aiutare a trovare il contesto e le parole di opinione specifiche di quest'ultimo.

Va comunque tenuto presente che compilare un *sentiment lexicon*, dipendente dal contesto o meno, è diverso dal determinare se una parola o una frase in realtà esprime un'opinione e quale sia il suo orientamento in una particolare frase. Solo perché una parola o una frase è elencata in un lessico di opinione non significa che in realtà stia esprimendo un'opinione in una frase, ad esempio, nella frase "Sto cercando una buona assicurazione sanitaria", "buono" non esprime un'opinione positiva o negativa su alcuna assicurazione in particolare.

³¹ Kanayama, Nasukawa 2006

³² Liu, Yu 2008

2.5 Sentiment analysis a livello di documento

Questo tipo di analisi è chiamata così poiché considera l'intero documento, come ad esempio una recensione di un prodotto, come una fonte di informazione riguardo l'opinione dell'individuo su una determinata entità. L'obiettivo è quindi quello di determinare il *sentiment* complessivo s di chi esprime l'opinione riguardo l'entità e dato un determinato documento d . Per far sì che questo compito abbia senso si deve tener presente che il documento d deve esprimere un'opinione su una singola entità e espressa da un singolo portatore di opinione h . Questo perché se in un singolo documento vengono valutate più entità, l'individuo potrebbe avere una opinione differente per ognuna di queste e quindi non avrebbe senso calcolare un sentimento a livello di documento, allo stesso modo se ci sono più portatori di opinione, questi potrebbero avere dei pareri diversi e quindi un sentimento generale a livello di documento non può essere calcolato.

In particolare se s ha un valore categorico del tipo "positivo" o "negativo", allora si avrà un problema di classificazione ovvero, dato un insieme di classi cerchiamo di determinare a quale classe un dato oggetto appartiene, se invece ha un valore numerico o ordinale con un range ad es. da 1 a 5, il problema diventa di regressione.

Di seguito saranno analizzate le classificazioni attraverso metodi di apprendimento supervisionato e non supervisionato.

2.5.1 Classificazione del sentiment con apprendimento supervised

In questa tipologia di analisi il sistema riceve in una fase iniziale di training, un insieme di dati già classificati ed etichettati. Da questo acquisisce conoscenza ed esperienza per classificare i dati successivi. La classificazione del *sentiment* è solitamente formulata come un problema di classificazione a due classi, positivo e negativo. I dati di addestramento e test utilizzati sono solitamente delle recensioni di prodotti. Dato che queste hanno dei punteggi di valutazione assegnati, ad esempio 1-5 stelle, le classi positive e negative sono determinate utilizzando queste ultime, una recensione con 4 o 5 stelle è considerata una recensione positiva e una recensione con 1 o 2 stelle è considerata una recensione negativa, ma è possibile utilizzare la classe neutrale, ad esempio, assegnando tutte le recensioni a 3 stelle alla classe neutra.

Il compito principale della classificazione è quello di progettare un insieme efficace di tecniche per analizzare in maniera corretta l'opinione espressa, alcune di queste tecniche sono:

- termini e la loro frequenza: in questo caso ogni singola parola viene presa in esame e viene conteggiata la sua frequenza, in alcuni casi è presa in esame anche la posizione delle parole;
- parte del discorso: è stato riscontrato che gli aggettivi sono i principali portatori di opinione, e quindi sono trattati come parti speciali del testo;
- parole di opinioni e frasi: le parole di opinione sono termini che sono comunemente usati per esprimere sentimenti positivi o negativi. Sebbene molte parole di opinione siano aggettivi e avverbi, anche i nomi e i verbi (ad es. odio e simili) possono esserlo. Oltre alle singole parole, ci sono anche frasi e idiomi di opinione;
- negazioni: le negazioni sono importanti perché la loro presenza inverte spesso l'orientamento delle opinioni. Tuttavia tali parole devono essere gestite con cura perché non tutte le volte che sono presenti invertono il senso della frase.

2.5.2 Classificazione del sentiment con apprendimento unsupervised

In questo caso il sistema riceve, nella fase di training, una serie di dati non etichettati che verranno classificati sulla base di caratteristiche comuni. Al contrario dell'apprendimento supervisionato, le classi non sono note a priori ma devono essere apprese automaticamente.

Ovviamente anche qui un ruolo centrale è ricoperto dalle frasi e dalle parole d'opinione. In particolare ci sono due metodi principali che utilizzano una classificazione del *sentiment* con apprendimento *unsupervised*, il primo³³ si divide in tre step principali:

- nel primo passaggio vengono analizzate le frasi che contengono aggettivi o avverbi, poiché sono le parole che solitamente esprimono un'opinione, anche se estrarre un aggettivo o un avverbio in maniera isolata non è abbastanza poiché questi potrebbero avere un significato differente a seconda del contesto. Quindi viene estratta una coppia di parole, un aggettivo o un avverbio e una parola che aiuti a capire il contesto;
- nella seconda fase viene stimato l'orientamento semantico della frase estratta attraverso l'indice PMI (*pointwise mutual information*);
- nella terza ed ultima fase viene calcolata la media di tutti gli orientamenti semantici stimati all'interno del documento, se tale valore medio è positivo allora il documento sarà valutato come positivo, viceversa sarà valutato come negativo.

³³ Turney 2002

L'altro approccio *unsupervised* è il metodo basato sul lessico, che usa un dizionario di parole e frasi di sentimento con i loro orientamenti e la loro forza e li utilizza per calcolare un punteggio di sentimento per ogni documento.

Riassumendo, possiamo osservare come il principale vantaggio della classificazione del *sentiment* a livello di documento sia che fornisce un'opinione generale su un'entità, un argomento o un evento. I principali difetti sono che non fornisce dettagli su ciò che le persone hanno gradito e/o che non hanno gradito e non è facilmente applicabile ai documenti che non siano delle recensioni, ad esempio forum e blog, perché solitamente questi documenti valutano più entità contemporaneamente confrontandole.

2.6 Sentiment analysis a livello di frase

Come discusso nel paragrafo precedente, la classificazione del *sentiment* a livello di documento potrebbe essere troppo grezza per la maggior parte delle applicazioni. Tuttavia, non vi è alcuna differenza fondamentale tra classificazioni a livello di documento e frase perché le frasi sono solo brevi documenti, e spesso anche le frasi contengono molteplici opinioni, ad es. "ho comprato un telefono Samsung due settimane fa. All'inizio tutto andava bene. La durata della batteria era buona e la fotocamera era ottima. Poi ieri ha smesso di funzionare", in questo caso la prima frase non esprime alcuna opinione in quanto afferma semplicemente un dato di fatto, tutte le altre frasi invece esprimono sentimenti.

In generale la definizione del problema di classificazione a livello di frase è: data una frase x , bisogna determinare se x esprime un'opinione positiva, negativa, neutra o non esprime opinioni.

La definizione della quintupla (e, a, s, h, t) in questo caso non è utilizzata poiché la classificazione a livello di frase è un passaggio intermedio, dato che nella maggior parte delle applicazioni è necessario conoscere gli obiettivi dell'opinione, e sapere solo che una frase esprime un'opinione positiva o negativa, ma non quali entità/aspetti l'opinione riguarda, è di uso limitato. Tuttavia, la classificazione a livello di frase può essere utile perché in molti casi, se sappiamo quali entità e/o aspetti dell'entità sono discussi in una frase, questo passaggio può aiutare a determinare se le opinioni sulle entità e i loro aspetti sono positive o negative.

In particolare, la classificazione a livello di frase può essere risolta come due problemi di classificazione differenti, il primo problema, o primo step, consiste nella classificazione della frase come soggettiva o oggettiva, quindi si va a vedere se contiene o meno una opinione. Il secondo problema classifica le frasi soggettive o d'opinione in positive o negative. Spesso le frasi oggettive sono considerate come prive d'opinione, ma, come accennato in precedenza, spesso non è così, nella frase presa ad esempio in precedenza, la frase "poi ieri ha smesso di funzionare", è oggettiva ma esprime palesemente un sentimento negativo. Quindi sarebbe più consono, nel primo step, classificare le frasi come d'opinione o meno piuttosto che come soggettive o oggettive.

Di seguito saranno analizzati i due step.

2.6.1 Classificazione della soggettività delle frasi

La classificazione di soggettività classifica le frasi in due classi, soggettive e oggettive, queste ultime esprimono dei dati di fatto, mentre una frase soggettiva di solito dà opinioni personali. In particolare, le frasi soggettive possono esprimere molti tipi di informazioni, ad es. opinioni, valutazioni, emozioni, giudizi, accuse, ecc, alcune di loro indicano sentimenti positivi o negativi e altre no.

La maggior parte degli approcci esistenti alla classificazione della soggettività sono basati sull'apprendimento *supervised*.

Una classificazione completa e conforme ai discorsi fatti in precedenza è quella che divide la soggettività in quattro classi³⁴: S, OO, O e SN, dove S indica una frase soggettiva e valutativa il cui sentimento può essere positivo o negativo, OO indica un'opinione positiva o negativa implicita in una frase oggettiva o segmento di frase, O significa oggettiva senza opinione e SN significa soggettiva ma non valutativa, ovvero nessun sentimento positivo o negativo è espresso. Questa classificazione è più conforme a quanto accennato anche in precedenza, dato che una frase soggettiva potrebbe non essere espressiva di un'opinione con un sentimento positivo o negativo e una frase oggettiva può anche implicare un sentimento.

2.6.2 Classificazione del sentiment delle frasi

Dopo il primo step, va poi determinato se una frase esprime un'opinione positiva o negativa. L'apprendimento *supervised* può essere applicato proprio come nella classificazione del *sentiment* a livello di documento.

Assunzione primaria nella classificazione del sentimento a livello di frase è che una frase deve esprimere un singolo sentimento di un singolo individuo.

Questa ipotesi è giusta per frasi semplici, ad esempio "La qualità dell'immagine di questa fotocamera è ottima", tuttavia, per frasi composte e complesse, una singola frase può esprimere più di un sentimento, ad esempio, la frase "La qualità delle immagini di questa fotocamera è ottima, così come la durata della batteria, ma il mirino è troppo piccolo ", esprime sia sentimenti positivi che negativi. Per "qualità dell'immagine" e "durata della batteria", la frase è positiva, ma per "mirino" è negativa.

Come detto, anche in questa analisi possono essere utilizzate le tecniche viste per la classificazione del *sentiment* a livello di documento, ma un'altra possibile tecnica è quella che comporta il calcolo dell'orientamento del sentimento di una frase attraverso la somma dei punteggi di orientamento di tutte le

³⁴ Benamara et al. 2011

parole d'opinione presenti nella frase, in particolare ad una parola positiva va assegnato il punteggio di +1 e ad una parola negativa il punteggio di -1, in questa situazione vengono considerate anche le parole di negazione e le parole che comportano un'inversione d'opinione nella frase come ad esempio la congiunzione "ma".

2.6.3 Problemi di classificazione a livello di frase

Gli approcci di classificazione menzionati sopra non tengono conto della presenza di alcune frasi particolari che esprimono opinioni in maniera differente e quindi richiedono un trattamento particolare. Queste frasi sono:

- le frasi condizionali: esprimono delle situazioni ipotetiche con le relative possibili conseguenze. Tali frasi solitamente contengono due proposizioni, una di condizione ed una conseguente, e dipendono l'una dall'altra. La loro relazione ha un impatto significativo sul fatto che la frase esprima un sentimento positivo o negativo. Quindi la sola presenza delle parole d'opinione non può determinare se una determinata frase esprima un sentimento o meno, ad esempio "Se troverò un'auto affidabile, la comprerò" e "Se il tuo telefono Samsung non va bene, compra questo telefono Apple ". La prima frase non esprime alcun sentimento nei confronti di una particolare vettura, sebbene "affidabile" sia una parola positiva, ma la seconda frase è positiva per il telefono Apple e non esprime un'opinione sul telefono Samsung. Quindi, un metodo per determinare l'opinione in frasi non condizionali non andrà bene per frasi condizionali;
- le frasi interrogative: queste sono difficili da analizzare perché in alcuni casi possono contenere una parola d'opinione pur non esprimendo nessun sentimento, né positivo né negativo. Due esempi possono essere i seguenti: "Qualcuno sa dirmi dove posso trovare un buon telefono Samsung?" e "Qualcuno può dirmi come risolvere questo problema con questo pessimo telefono Samsung?". Nella prima frase, pur essendoci la parola "buon", non viene espresso alcun sentimento, nella seconda, invece, vi è chiaramente l'espressione di un'opinione negativa nei confronti del telefono Samsung;
- le frasi sarcastiche: sono delle frasi particolari in cui solitamente lo scrittore scrive il contrario di ciò che intende realmente. Nel contesto di *sentiment analysis*, quindi, solitamente, chi scrive qualcosa di positivo vuole intendere un'opinione negativa e viceversa. Questa tipologia di frasi è poco utilizzata nelle recensioni di prodotti o altro, mentre è maggiormente presente nei forum e nei blog, soprattutto quelli che parlano di politica. Nel corso del tempo sono state sviluppate diverse tecniche per individuare le frasi sarcastiche, ad esempio spesso le frasi sarcastiche sono presenti in documenti che presentano altre frasi dello stesso tipo, o è possibile osservare la punteggiatura e quindi guardare al numero di punti interrogativi, esclamativi, di virgolette e di parole con tutte lettere maiuscole

utilizzato³⁵, altro metodo utilizzabile è quelli di guardare all'utilizzo di espressioni informali come "ah", "oh", "yeah" ecc³⁶.

Per concludere, va detto che nonostante questa tipologia di analisi sia migliore rispetto a quella a livello di documento, presenta comunque alcune lacune importanti che la inquadrano come un passaggio intermedio. In particolare i principali limiti di tale approccio sono che non si saprà a quali aspetti dell'entità l'opinione positiva o negativa si riferisca, così come avveniva per la *sentiment analysis* a livello di documento, ed anche se si sapesse l'oggetto dell'opinione, comunque non sarebbe sufficiente dato che alcune frasi presentano opinioni su molteplici entità o aspetti, o ancora le frasi potrebbero avere un tono generale positivo o negativo ma comunque esprimere un'opinione opposta su alcuni aspetti, ed infine questo tipo di analisi non può essere utilizzata per le frasi comparative poiché per questa tipologia di proposizioni non ha senso una classificazione in positiva, negativa o neutra.

2.7 Sentiment analysis a livello di aspetto

Come abbiamo discusso nei due paragrafi precedenti, la classificazione del *sentiment* a livello di documento o di frase è spesso insufficiente poiché spesso non si identificano i veri obiettivi dell'opinione espressa. Anche supponendo che ogni individuo valuti un'unica entità, un documento che esprime un'opinione positiva sull'entità, non significa che l'autore abbia opinioni positive su tutti gli aspetti dell'entità. Allo stesso modo, un documento di opinione negativo non significa che l'autore abbia un'opinione negativa su tutto. Per un'analisi più completa, è necessario scoprire gli aspetti dell'entità e determinare se il sentimento è positivo o negativo su ciascuno di questi.

Per estrarre tali dettagli, si applica la *sentiment analysis* a livello di aspetto appunto, il che significa che l'obiettivo è scoprire ogni quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ in un determinato documento d .

Per raggiungere questo obiettivo, devono essere eseguiti due step in particolare:

- estrazione dell'aspetto: questo compito estrae dalla frase gli aspetti che sono stati valutati. Ad esempio, nella frase "La fotocamera di questo telefono è sorprendente", l'aspetto dell'entità "questo telefono" è "fotocamera". Si noti che "questo telefono" non è più il fulcro dell'analisi perché la valutazione non riguarda il cellulare nel suo complesso, ma solo la sua fotocamera. Va comunque tenuto presente che quando si parla di un aspetto, bisogna sapere a quale entità appartiene;

³⁵ Tsur, Davidov e Rappoport 2010

³⁶ González-Ibáñez, Muresan e Wacholder 2011

- classificazione del sentiment degli aspetti: questo compito consiste nel determinare se le opinioni su diversi aspetti sono positive, negative o neutre. Ad esempio nell'esempio sopra riportato, l'opinione sull'aspetto "fotocamera" è positiva.

2.7.1 Estrazione degli aspetti

Per effettuare l'analisi a livello di aspetto. Il primo passo da dover compiere è quello di estrarre l'aspetto di cui si parla nella frase e sul quale l'opinione viene espressa. Per fare ciò ci sono quattro approcci principali possibili: estrazione basata sui sostantivi frequenti o sulle frasi nominali, estrazione attraverso l'individuazione delle relazioni tra l'opinione e l'oggetto di quest'ultima, estrazione utilizzando l'apprendimento *supervised* ed estrazione attraverso il metodo *topic models*.

Sul web esistono due tipologie principali di recensioni: la prima consiste in un'analisi dei pro e dei contro e successivamente una recensione dettagliata, la seconda tipologia invece prevede un formato libero dove il recensore scrive liberamente senza seguire una struttura definita. Il primo tipo di recensione è una tipologia particolare del secondo tipo, ed è anche più semplice da analizzare poiché solitamente tali testi sono formati da brevi frasi o segmenti di frase che sono divise da punteggiatura o congiunzioni, e ogni frase si riferisce ad un solo aspetto.

Il primo approccio possibile, ovvero quello che basa l'estrazione sui sostantivi frequenti o sulle frasi nominali, trova gli aspetti espliciti che sono presenti in un gran numero di recensioni in un determinato contesto. In particolare viene monitorato il numero di volte che un determinato sostantivo appare nelle recensioni, e vengono mantenuti solo quelli che appaiono frequentemente nel contesto considerato. Le ragioni alla base di questo metodo sono da collegare al fatto che quando gli utenti recensiscono degli aspetti di una determinata entità, solitamente utilizzano dei termini comuni e quindi solitamente gli aspetti menzionati più volte sono importanti in quel determinato contesto. Per lo stesso motivo, i termini meno frequenti, o non sono degli aspetti o sono aspetti di minor importanza.

Dato che le opinioni hanno un oggetto a cui si riferiscono, queste due parti delle frasi sono correlate tra loro ovviamente, e questa loro relazione può essere sfruttata per estrarre gli aspetti di una determinata entità. Alla base di questo secondo approccio per l'estrazione degli aspetti vi è il fatto che una stessa parola d'opinione può essere utilizzata per diversi aspetti, e se una frase non ne ha uno che risulta essere frequente nel contesto considerato ma ha una parola d'opinione, il sostantivo più vicino a quest'ultima viene estratto, quindi questo metodo può essere utile per estrarre gli aspetti poco frequenti. Ad esempio se si considera la frase "questo software è ottimo" e si sa che "ottimo" è una parola d'opinione, allora l'aspetto "software" verrà estratto.

Il terzo approccio vede l'utilizzo di un apprendimento *supervised*, questo perché l'estrazione degli aspetti può essere vista come un caso speciale del problema generale di estrazione delle informazioni e molti algoritmi basati sull'apprendimento *supervised* sono utilizzati per questo scopo. I metodi migliori sono quelli basati sull'apprendimento sequenziale, e poiché si tratta di tecniche *supervised*, hanno bisogno di dati etichettati manualmente per l'addestramento, cioè, è necessario annotare manualmente gli aspetti e i non-aspetti in un corpus.

Ultimo approccio è quello del *topic models*, in particolare questi modelli sono utilizzati per estrarre argomenti da un insieme di testi. Sono metodi di apprendimento *unsupervised* che presuppongono che ogni documento sia costituito da una combinazione di argomenti e che ogni argomento sia una distribuzione di probabilità sulle parole.

Gli argomenti in questi modelli sono gli aspetti nella *sentiment analysis*, quindi il *topic model* può essere utilizzato per estrarre gli aspetti. Va però tenuto presente che gli argomenti possono riguardare sia le parole d'opinione sia le parole di aspetto, e per la *sentiment analysis* devono essere separate tra loro. Inoltre tali modelli oltre ad estrarre gli aspetti, li raggruppano in gruppi di sinonimi.

Tali modelli quindi non sono semplici da interpretare come quelli precedenti, ma hanno il vantaggio di raggruppare automaticamente sotto lo stesso aspetto le parole diverse che esprimono lo stesso aspetto o aspetti correlati tra loro.

Sebbene i *topic model* siano dei modelli basati sulle inferenze probabilistiche, presentano alcuni punti deboli che ne limitano l'uso pratico nelle applicazioni di *sentiment analysis*. Il problema principale è che hanno bisogno di un grande volume di dati e molte messe a punto per poter ottenere dei buoni risultati.

Inoltre non è difficile per tale approccio trovare argomenti o aspetti molto generali e frequenti in una vasta raccolta di documenti, ma non è facile trovare quegli aspetti frequenti in un dato contesto ma globalmente non molto frequenti. Tali aspetti localmente frequenti sono spesso i più utili per la *sentiment analysis* perché sono probabilmente più rilevanti per le specifiche entità a cui l'utente è interessato. Gli aspetti molto generali e frequenti possono anche essere facilmente trovati con i metodi discussi in precedenza, e questi metodi possono trovare anche aspetti meno frequenti senza la necessità di una grande quantità di dati. In pratica quindi, questo approccio potrebbe non essere abbastanza specifico per diverse applicazioni di *sentiment analysis*.

2.7.1.1 Particolarità nell'estrazione degli aspetti

Nel compito di estrazione degli aspetti, va tenuto conto del fatto che non ci sono soltanto gli aspetti espliciti, ovvero quelli espressi con parole o frasi all'interno di una frase che esprime un'opinione su un'entità, ma ci sono anche espressioni che si riferiscono ad aspetti non esplicitamente citati e perciò chiamati aspetti impliciti. Gli aggettivi e gli avverbi sono le espressioni più comuni per esprimere un aspetto implicito, perché la maggior parte degli aggettivi descrive alcuni attributi o proprietà specifici delle entità, ad esempio "costoso" è un termine che descrive il "prezzo" e "bello" descrive l'aspetto esteriore, ma gli aspetti impliciti possono anche essere rappresentati da verbi. In generale, le espressioni che si riferiscono ad un aspetto implicito possono essere molto complesse, ad esempio "Questa fotocamera non entra facilmente in una tasca", "entra in una tasca" indica l'aspetto dimensione dell'entità "fotocamera".

Altra particolarità da dover tenere in considerazione quando si parla di estrazione degli aspetti, è che in molti casi l'utilizzo delle risorse da parte di un'entità risulta essere un aspetto importante nell'ambito della *sentiment analysis*. Ad esempio nella frase "questa macchina consuma molta benzina", il consumo della benzina, quindi di una risorsa, è un aspetto negativo dell'entità "macchina", poiché non è desiderabile consumare molto carburante. In questo esempio non vi è una parola o frase d'opinione, ma comunque il *sentiment* è palesemente negativo, quindi andare ad individuare parole e frasi che si riferiscono alle risorse è fondamentale per la *sentiment analysis*. Un metodo³⁷ utilizzabile per l'estrazione dei termini di risorsa parte dall'osservazione che il sentimento o l'opinione espressi in una frase sull'utilizzo delle risorse vedono spesso la presenza della seguente tripla (verbo, quantificatore, sostantivo o frase nominale). Prendendo l'esempio precedente "questa macchina consuma molta benzina", "consuma" è il verbo principale, "molta" è il quantificatore e "benzina" è il termine che si riferisce alla risorsa. Quindi in particolare questo metodo utilizza tali tripli per identificare i termini di risorsa in un corpus.

³⁷ Zhang e Liu 2011

2.7.2 Raggruppamento degli aspetti in categorie

Dopo l'estrazione, le parole e le frasi che indicano un determinato aspetto devono essere raggruppate in categorie dove ogni categoria rappresenta un aspetto unico. Le persone sono solite utilizzare parole e frasi diverse per descrivere lo stesso aspetto. Ad esempio, "qualità della chiamata" e "qualità della voce" si riferiscono allo stesso aspetto per i telefoni, e raggruppare tali espressioni che indicano uno stesso aspetto è fondamentale per la *sentiment analysis*. Sebbene i dizionari possano aiutare in una certa misura, non sono di certo sufficienti perché molti sinonimi dipendono dal contesto. Ad esempio, "film" e "immagine" possono essere sinonimi nelle recensioni dei film, ma non sono sinonimi nelle recensioni delle fotocamere, perché "immagine" è più probabile che sia sinonimo di "foto" mentre "film" di "video". Inoltre molte espressioni di aspetto sono frasi di più parole che non possono essere facilmente gestite con i dizionari. È anche importante notare che molte di queste espressioni che descrivono lo stesso aspetto non sono sinonimi in generale o specifici del contesto, ad esempio, "costoso" e "conveniente" possono entrambi indicare l'aspetto prezzo dell'entità, ma non sono sinonimi l'uno dell'altro, ma anzi contrari, o sinonimi di prezzo.

Un approccio³⁸ possibile a tale problema è un metodo *semi-supervised* che come primo step effettua l'etichettatura di pochi aspetti per ogni categoria, e successivamente assegna i restanti aspetti alle varie categorie utilizzando sia esempi etichettati che non. Tale modello utilizza anche due accorgimenti per migliorare il raggruppamento degli aspetti, il primo è considerare che le espressioni di aspetto che condividono alcune parole comuni appartengono probabilmente allo stesso gruppo, ad esempio "durata della batteria" e "alimentazione della batteria" e il secondo si basa sul fatto che le espressioni di aspetto che sono sinonimi in un dizionario sono probabilmente appartenenti allo stesso gruppo.

Inoltre, come detto in precedenza, i *topic model* eseguono sia la scoperta ed estrazione di espressioni di aspetto che la loro categorizzazione raggruppandoli in una raccolta di documenti con un metodo *unsupervised*.

Infine, un altro possibile approccio³⁹ per il raggruppamento di aspetti in categorie, vede l'aggiunta ai *topic model* di due vincoli che migliorano la categorizzazione. Il primo dice che se due espressioni di aspetto a_i e a_j condividono una o più parole allora faranno parte della stessa categoria, mentre il secondo afferma che se due espressioni di aspetto a_i e a_j si trovano nella stessa frase allora non dovranno far parte della stessa categoria, questo perché le persone solitamente non ripetono lo stesso aspetto più volte nella stessa frase, ad esempio nella frase "Mi piacciono la qualità dell'immagine, la durata della batteria e lo zoom di questa fotocamera" abbiamo una entità "fotocamera" e tre aspetti differenti che non dovranno essere raggruppati nella stessa categoria: "qualità dell'immagine", "durata della batteria" e "zoom".

³⁸ Zhai et al. 2010

³⁹ Zhai et al. 2011

2.7.3 Classificazione del sentiment degli aspetti

Per andare ad individuare l'opinione riguardo un determinato aspetto ci sono due approcci principali: quello di apprendimento *supervised* e quello basato sul lessico.

Per il primo approccio sono applicabili i metodi visti in precedenza per le altre tipologie di classificazione, qui il problema chiave è come determinare l'ambito di ciascuna espressione di sentimento, cioè se copre l'aspetto di interesse all'interno della frase. Tale forma di apprendimento tuttavia dipende dai dati che si hanno a disposizione per l'allenamento, e solitamente un modello addestrato su dei dati appartenenti ad un certo contesto, si comporta in maniera non precisa in altri contesti. Per questo motivo tali approcci funzionano meglio nella *sentiment analysis* a livello di documento, poiché questi sono più lunghi e contengono più aspetti rispetto alle singole frasi.

L'approccio basato sul lessico, invece, funziona bene in un gran numero di contesti differenti. Tali metodi in genere utilizzano un *sentiment lexicon*, che contiene un elenco di parole sentimentali, frasi e espressioni idiomatiche, espressioni composte, considerano i cambi di opinione dovuti alle negazioni o ad avverbi che indicano il cambio di opinione nella frase, e molti altri costrutti che possono influenzare i sentimenti.

Un esempio⁴⁰ di tale approccio è il seguente, ed ha quattro passaggi:

- identificare le parole e le frasi di sentimento: per ogni frase che contiene uno o più aspetti, in questo passaggio vengono identificate tutte le parole e le frasi d'opinione nella frase. Ad ogni parola positiva viene assegnato il punteggio di sentimento di +1 e ad ogni parola negativa viene assegnato il punteggio di sentimento di -1. Ad esempio, nella frase: "La cassa di questo telefono non è buona, ma la durata della batteria è lunga." dopo questo step, la frase diventa "La cassa di questo telefono non è buona [+1], ma la durata della batteria è lunga" perché "buona" è una parola positiva. Inoltre "lunga" qui non è una parola d'opinione, in quanto non indica un sentimento positivo o negativo di per sé in generale, ma si può inferire il suo sentimento in questo contesto. Infatti, "lunga" può essere considerato come una parola di sentimento dipendente dal contesto;
- applicare i modificatori di opinione: tali modificatori del *sentiment* sono parole o frasi che modificano l'orientamento dell'opinione. Le parole di negazione come no, non, mai ecc. sono i modificatori più comuni. Questo passaggio trasforma la nostra frase precedente in "La cassa di questo telefono non è buona [-1], ma la durata della batteria è lunga" a causa della presenza del modificatore d'opinione "non". Va tenuto presente che non tutti i modificatori cambiano l'orientamento del sentimento, ad esempio "non solo ... ma anche" è una frase che non inverte l'opinione;

⁴⁰ Ding, Liu e Yu 2008

- gestire la congiunzione “ma”: è una congiunzione avversativa che esprime contrapposizione tra due elementi di una proposizione o fra due proposizioni coordinate⁴¹. Le frasi che contengono tale congiunzione vengono gestite con tale regola: l’orientamento del sentimento della frase antecedente a tale parola e dopo tale parola sono opposti tra loro se non è possibile determinare l’orientamento dell’opinione di una delle due frasi. Dopo questo passaggio, la frase dell’esempio precedente diventa: "La cassa di questo telefono non è buona [-1], ma la durata della batteria è lunga [+1]" a causa di "ma", viene aggiunto il +1 alla fine della frase successiva a tale congiunzione, e quindi si può dedurre che "lunga" ha accezione positiva per l’aspetto "batteria". Oltre a questa congiunzione, anche frasi come "ad eccezione di" e "tranne che" vengono gestite allo stesso modo. Anche qui va tenuto conto che non in tutti i casi in cui vi è tale congiunzione ci sarà un cambiamento d’opinione, ad esempio “questa macchina è ottima ma quell’altra è migliore”;
- unione delle singole opinioni: questo passaggio applica una sommatoria ai punteggi di opinione assegnati ad ogni aspetto per determinare l’orientamento finale del *sentiment* su ciascun aspetto della frase.

Formalmente: si ha una frase s che contiene un insieme di aspetti $\{a_1, \dots, a_m\}$ e un insieme di parole o frasi di sentimento $\{sw_1, \dots, sw_n\}$ con i loro punteggi di sentimento ottenuti dai passaggi dall’1 al 3. L’orientamento del *sentiment* per ogni aspetto a_i in s è determinato dalla seguente funzione:

$$score(a_i, s) = \sum \frac{sw_j \cdot so}{dist(sw_j, a_i)}$$

dove sw_j è una parola o frase d’opinione in s , $dist(sw_j, a_i)$ è la distanza tra l’aspetto a_i e la parola o frase d’opinione sw_j in s . $sw_j \cdot so$ è il punteggio di *sentiment* di sw_j . Tale forma assegna minor peso alle parole o frasi d’opinione che sono distanti dall’aspetto a_i . Alla fine se tale punteggio risulta essere positivo allora l’opinione sull’aspetto a_i nella frase s è positiva, se il punteggio finale è negativo è negativa, altrimenti sarà neutra.

Questo è un approccio semplice alla classificazione delle opinioni riguardo un aspetto, e può essere migliorato in diversi modo come ad esempio integrando tale metodo basato sul lessico con un approccio *supervised*.

Come accennato in precedenza, oltre alle parole e frasi d’opinione, ci sono altre tipologie di espressioni che possono trasmettere o implicare dei sentimenti. Di seguito ne saranno indicate alcune che sono chiamate regole di base delle opinioni⁴².

⁴¹ <http://dizionari.repubblica.it/Italiano/M/ma.php>

⁴² Liu, 2010

2.7.4 Regole di base delle opinioni

Una regola d'opinione esprime un concetto che implica un sentimento positivo o negativo. Può essere semplice come le singole parole d'opinione o possono essere espressioni composte che possono richiedere la conoscenza del senso comune o del contesto per determinare l'orientamento dei sentimenti. Un modo per analizzare queste regole è usare l'idea della semantica compositiva⁴³, la quale afferma che il significato di un'espressione composta è una funzione del significato dei suoi costituenti e delle regole di sintassi con cui sono combinati. Tali regole possono essere così rappresentate:

1. POSITIVO ::= P
2. PO
3. modificatore_sentimento N
4. modificatore_sentimento NE
5. NEGATIVO ::= N
6. NE
7. modificatore_sentimento P
8. modificatore_sentimento PO

In particolare P e PO rappresentano due tipologie di espressioni che esprimono opinioni positive, le prime sono delle semplici parole o frasi, mentre PO sono delle espressioni positive composte da più espressioni. Allo stesso modo N e NE invece rappresentano due tipi di espressioni che esprimono opinioni negative. "modificatore_sentimento N" e "modificatore_sentimento NE" rappresentano rispettivamente la negazione di N e NE, e "modificatore_sentimento P" e "modificatore_sentimento PO" rappresentano rispettivamente la negazione di P e PO.

Come accennato in precedenza, i modificatori di sentimento sono le parole di negazione come non, mai, nessuno ecc., alcuni verbi, alcuni avverbi, alcune particolari parole che hanno un senso negativo nel linguaggio comune ed anche il sarcasmo.

Per quanto riguarda le espressioni P, PO, N e NE, possono essere raggruppate in sei categorie concettuali in base alle loro caratteristiche:

- parole o frasi d'opinione: questa è la categoria più semplice e più utilizzata, e tali espressioni esprimono solo espressioni o positive o negative su determinati aspetti. Fanno parte di tale categoria P ed N;

9. P ::= una_parola_o_frase_positiva
10. N ::= una_parola_o_frase_negativa

⁴³ Dowty, Wall and Peters, 1981

- diminuzione o aumento della quantità di una determinata entità: tali regole funzionano in maniera simile ai modificatori di sentimento, in particolare esprimono che la diminuzione o l'aumento della quantità associata a un'entità possa cambiare l'orientamento del sentimento. Ad esempio, nella frase "Questo farmaco ha ridotto significativamente il mio dolore", "il dolore" è una parola negativa, e la riduzione del "dolore" indica un effetto positivo del farmaco, quindi, la diminuzione del dolore implica un'opinione positiva sul farmaco. Il concetto di diminuzione si estende anche alla rimozione e alla scomparsa, ad esempio "il mio dolore è scomparso dopo aver assunto il farmaco";

11. PO ::= meno_o_diminuzione N
12. più_o_aumenta P
13. NE ::= meno_o_diminuzione P
14. più_o_aumenta N

- quantità alta, bassa, aumentata e ridotta di un potenziale elemento positivo o negativo: per alcune entità un piccolo valore/quantità è negativo e un grande valore/quantità è positivo, ad esempio, "La durata della batteria è breve" e "La durata della batteria è lunga". Tali espressioni sono chiamate potenziali elementi positivi (PPI), in questo caso "la durata della batteria" è un potenziale elemento positivo. Per altre entità, un piccolo valore/quantità è positivo e un grande valore/quantità è negativo, ad esempio "Questo telefono costa molto" e "Sony ha ridotto il prezzo della fotocamera". Tali elementi sono chiamati elementi potenzialmente negativi (NPI), in questi esempi "costo" e "prezzo" sono elementi potenziali negativi.

Sia gli elementi potenziali positivi che quelli negativi non implicano alcuna opinione, ma quando sono modificati dagli aggettivi di quantità o dalle parole o dalle frasi di cambiamento di quantità, possono implicare un orientamento dell'opinione positivo o negativo;

15. PO ::= quantità_bassa_o_ridotta_di NPI
16. quantità_alta_o_incrementata di PPI
17. NE ::= quantità_bassa_o_ridotta_di PPI
18. quantità_alta_o_incrementata di NPI
19. NPI ::= potenziale_elemento_negativo
20. PPI ::= potenziale_elemento_positivo

- fatto desiderabile o indesiderabile: tutte le regole appena indicate contengono soprattutto delle espressioni soggettive, ma anche le espressioni oggettive possono implicare sentimenti positivi o negativi in quanto possono descrivere delle situazioni desiderabili o meno. Tali frasi spesso non contengono parole d'opinione;

21. P ::= fatto_desiderabile

22. N ::= fatto_nondesiderabile

- deviazione dalla norma o da un range desiderato: in alcuni casi il valore di un'entità ha un intervallo o una norma desiderati, e se il valore si discosta dall'intervallo normale, è negativo, ad es. "dopo aver assunto il farmaco, la mia pressione sanguigna è passata a 190". Tali frasi sono spesso delle frasi oggettive;

23. P ::= dentro_dal_range_desiderato

24. N ::= fuori_dal_range_desiderato

- Produrre e consumare risorse e rifiuti: se un'entità produce una grande quantità di risorse, è solitamente è positiva. Se consuma una grande quantità di risorse, è invece negativa, ad esempio, l'elettricità è una risorsa e la frase "Questo computer utilizza molta elettricità" dà un'opinione negativa sul computer stesso. Allo stesso modo, se un'entità produce una grande quantità di rifiuti, è negativa, e se consuma una grande quantità di rifiuti, è positiva.

25. P ::= produce_molte_o_più_risorse

26. non_produce,_o_pochi_o_meno_rifiuti

27. non_consumo,_o_poche_o_meno_risorse

28. consuma_molti_o_più_rifiuti

29. N ::= non_produce,_o_pochi_o_meno_risorse

30. produce_molto_o_più_rifiuti

31. consuma_molto_o_più_risorse

32. non_consumo,_o_poche_o_meno_rifiuti

2.7.5 Estrazione di entità, opinion holder e tempo

In alcune applicazioni della *sentiment analysis* è utile identificare ed estrarre l'entità, chi esprime l'opinione e il momento in cui viene espressa l'opinione.

Nel caso dei social media, gli *opinion holder* sono spesso gli autori dei messaggi scritti, e la loro identità digitale è nota anche se a volte la loro identità nel mondo reale potrebbe essere sconosciuta. Anche la data e l'ora in cui viene inviata un'opinione sono note e visualizzabili sulla pagina web, e quindi la loro estrazione risulta essere particolarmente semplice.

Per quanto riguarda l'estrazione dell'entità, la situazione cambia e si deve tener presente che l'utente spesso desidera trovare delle opinioni su entità concorrenti tra loro, come modelli di prodotto o brand, per poterli confrontare. Altra questione da tenere in considerazione è che spesso gli utenti scrivono il nome dei brand o dei modelli in maniera diversa ed è quindi importante che il sistema li individui automaticamente da un dato

corpus. Il requisito fondamentale di tale estrazione è che le entità individuate dal sistema devono essere dello stesso tipo delle entità ricercate dall'utente ad es. marche e modelli di cellulari.

Un modo⁴⁴ possibile di affrontare questo problema è quello di trattarlo come un problema d'espansione di un insieme, e formalmente può essere così descritto: dato un insieme Q di entità di una particolare classe C , e un insieme D di entità possibili, si deve determinare quale delle entità in D appartengono a C . Cioè creiamo la classe C basata sull'insieme di esempi Q , in pratica si inseriscono le entità in D in base alla loro probabilità di appartenenza a C .

Per quanto riguarda l'estrazione di colui che esprime l'opinione, al di fuori del contesto dei social media, si ha un possibile approccio⁴⁵ che genera prima tutti i possibili candidati in una frase, analizzando tutte le frasi nominali, comprese le frasi con nomi comuni, le entità nominate e i pronomi e successivamente analizza la frase e classifica tutti i candidati scegliendo quello con il punteggio più alto come portatore dell'opinione nella frase.

⁴⁴ *Li et al., 2010*

⁴⁵ *Kim and Hovy 2006*

3 Explicit Factor Models

I sistemi di raccomandazione hanno vissuto una crescita esponenziale negli ultimi anni, diventando parte centrale ed indispensabile per alcuni service provider leader nel mondo online come Amazon.com, Netflix e TripAdvisor.

Negli ultimi anni i *Latent Factor Model*, ed in particolare la tecnica *Matrix Factorization*, sono stati studiati con molta attenzione dalla comunità di ricerca e dall'industria grazie alla loro capacità di fornire delle raccomandazioni precise su diversi tipologie di dati.

Tuttavia, i *recommendation systems* basati su questi algoritmi incontrano alcuni problemi importanti nelle applicazioni pratiche. Innanzitutto, è difficile sapere in che modo gli utenti compongono il loro giudizio sui vari aspetti di un'entità, il che rende difficile formulare delle raccomandazioni in base alle esigenze ed ai gusti specifici degli utenti, ed in secondo luogo, solitamente è difficile dare spiegazioni sul motivo per cui un articolo viene raccomandato, e ancora più complicato è spiegare perché un'entità non è raccomandata. Ed inoltre va aggiunto che la mancanza di spiegazioni indebolisce la capacità di persuadere gli utenti e di aiutarli a prendere decisioni di scelta ed acquisto migliori⁴⁶.

Le spiegazioni nei sistemi di raccomandazione potrebbero essere molto utili dato che, spiegando come funziona il sistema e/o perché un prodotto viene raccomandato, il sistema diventa più trasparente e dà la possibilità agli utenti di dire quando esso sbaglia. Inoltre, la maggior trasparenza, aumenta la confidenza e la fiducia degli utenti, aiuta gli user a prendere decisioni migliori e più rapide, migliorando così l'efficacia e l'efficienza, riesce più facilmente a convincere gli utenti a provare o acquistare un prodotto ed aumenta la loro soddisfazione riguardo l'utilizzo dei sistemi di raccomandazione rendendo più probabile una loro fedeltà futura.

A questo punto un problema potrebbe essere quello di dover scegliere tra un sistema di raccomandazione con un'elevata accuratezza, come ad esempio i *Latent Factor Model*, o uno che sia maggiormente trasparente e comprensibile rinunciando ad una maggior precisione nelle previsioni. Ma, fortunatamente, lo sviluppo della *sentiment analysis* e la crescente popolarità delle recensioni testuali degli utenti online potrebbero evitare ai service provider una scelta di questo tipo.

In particolare, la maggior parte dei siti di e-commerce e di servizi, come Amazon.com e TripAdvisor, consentono agli utenti di scrivere recensioni a testo libero congiuntamente ad una valutazione numerica. Come detto, le recensioni contengono informazioni dettagliate sulle opinioni, sugli atteggiamenti e sulle preferenze degli utenti nei confronti degli aspetti del prodotto che possono aiutare a creare nuovi modelli per raccomandazioni maggiormente trasparenti e spiegabili.

Inoltre, utenti diversi potrebbero essere interessati a diversi aspetti del prodotto, ad esempio c'è chi potrebbe essere maggiormente interessato alle dimensioni dello schermo di un cellulare, e chi invece alla durata della

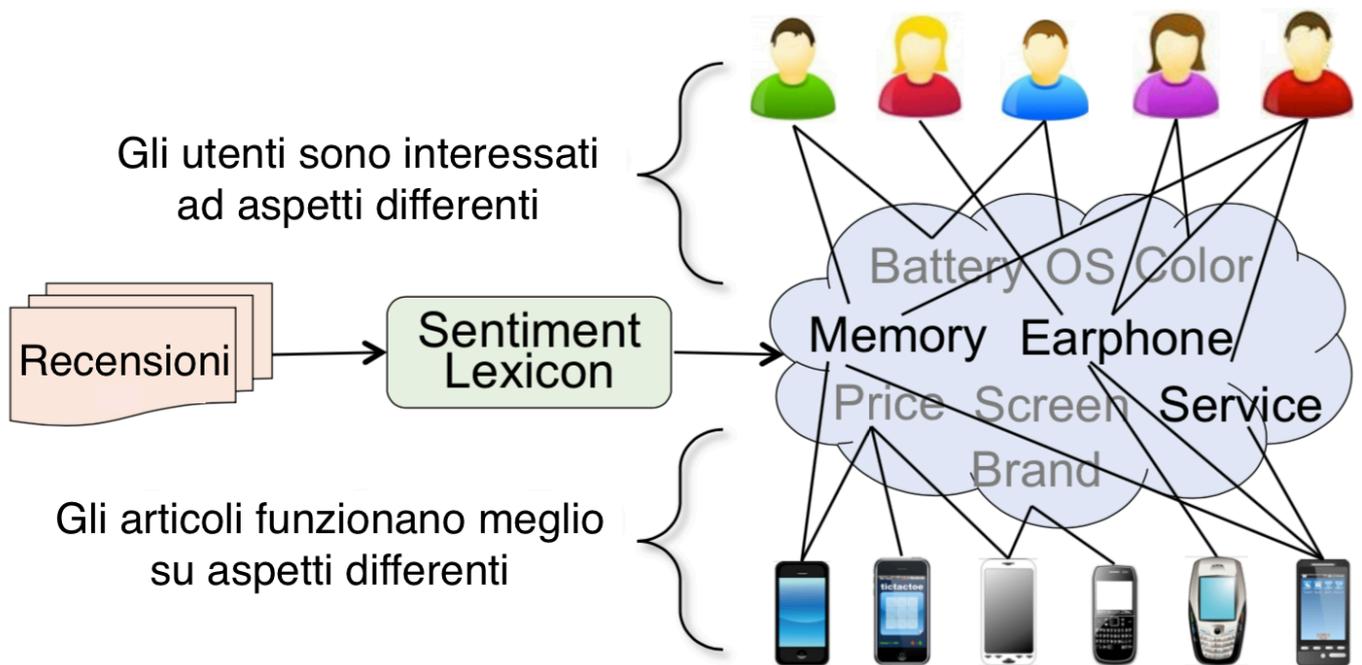
⁴⁶ Vig, Sen and Riedel 2009

batteria, anche se entrambi potrebbero aver assegnato la stessa valutazione numerica al prodotto. Quindi estrarre le caratteristiche del prodotto e le relative opinioni degli utenti dalle recensioni, non solo aiuta a capire le diverse preferenze degli utenti e a formulare raccomandazioni migliori, ma aiuta anche a capire perché e come un determinato articolo dovrebbe o meno essere raccomandato. In questo modo, non solo si potrebbero raccomandare agli utenti determinati prodotti, ma si potrebbe anche spiegarli il perché non dovrebbero acquistare una data entità.

In questo capitolo sarà analizzato un modello proposto da Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma chiamato *Explicit Factor Model* che ha l'obiettivo di ottenere sia un'elevata precisione sia un'elevata trasparenza e comprensibilità attraverso delle raccomandazioni frutto di una *sentiment analysis* sulle recensioni degli utenti utilizzata per individuare le caratteristiche ricercate da ogni user e quelle possedute dai prodotti nel catalogo, ad esempio se il sistema trova che un utente potrebbe essere particolarmente interessato alla memoria, agli auricolari ed alla batteria di un cellulare, allora gli raccomanderà un prodotto che funziona particolarmente bene su questi tre aspetti.

Nei paragrafi successivi sarà analizzato nel dettaglio tale modello.

(Immagine 3)



3.1 Costruzione del sentiment lexicon

Come detto nel capitolo precedente, nella maggior parte delle applicazioni di analisi del *sentiment*, il *sentiment lexicon* svolge un ruolo centrale. Tuttavia, non ne esiste uno universalmente utilizzabile, poiché la polarità delle parole è sensibile al contesto dell'argomento. Ancor peggio, nello stesso dominio la stessa parola può indicare un sentimento differente rispetto a diversi aspetti, ad esempio, in una recensione di un computer portatile, "grande" potrebbe essere negativo per l'aspetto batteria, pur essendo positivo per l'aspetto schermo. L'obiettivo è quindi di creare un *sentiment lexicon* che non sia solo dipendente dal contesto, ma che tenga conto anche dei diversi aspetti considerati all'interno dello stesso dominio.

In questo modello, la creazione del *sentiment lexicon*⁴⁷ viene effettuata combinando quattro tipi di vincoli, rispettivamente:

- vincoli che derivano da *sentiment lexicon* generali, ovvero che contengono parole che sono quasi sempre positive o negative in qualsiasi contesto, come "eccellente" e "cattivo";
- vincoli basati sulla valutazione complessiva del *sentiment* in tutto il documento. In molti casi, ogni recensione viene fornita insieme con una valutazione complessiva dell'opinione da parte dell'utente, come nelle recensioni su TripAdvisor e Amazon. Questo tipo di valutazione complessiva è abbondante sul Web, ad esempio, ci sono oltre 40 milioni di recensioni relative ai viaggi su TripAdvisor. La logica alla base di questo vincolo, è che la valutazione complessiva trasmette alcune informazioni sul sentimento espresso nel testo, ad esempio, è improbabile che un utente utilizzi solo parole negative nel testo dando poi una valutazione complessiva di 5 stelle;
- vincoli basati su di un dizionario dei sinonimi e contrari. Ad esempio, potremmo non sapere se l'aggettivo "largo" è positivo o negativo per l'aspetto dello schermo in una recensione su un computer portatile, ma sappiamo che dovrebbe avere un significato simile a "grande" e molto diverso da "piccolo". Quindi se abbiamo informazioni riguardo la polarità di "grande" o di "piccolo", possiamo dedurre facilmente la polarità di "largo";
- vincoli basati sull'euristica linguistica: regola della "e": le frasi che sono connesse con la congiunzione "e" solitamente esprimono la stessa polarità di sentimento. Ad esempio, "la batteria dura a lungo e le dimensioni dello schermo sono grandi" significa che "lungo" per "batteria" e "grande" per "dimensione schermo" hanno la stessa polarità.

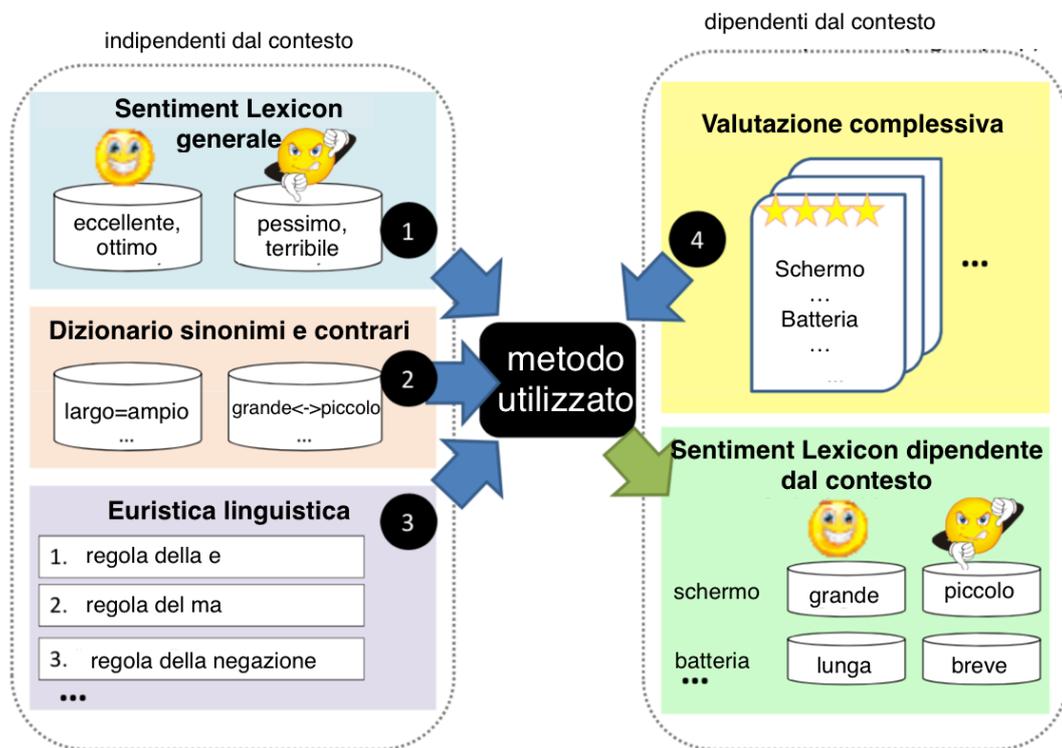
Regola del "ma": le frasi che sono connesse con "ma" solitamente esprimono una polarità di sentimento opposto. Ad esempio, "la batteria dura a lungo ma la dimensione dello schermo è minuscola" indica che "lungo" per "batteria" e "minuscolo" per "dimensioni dello schermo" sono della polarità opposta.

⁴⁷ Y. Lu, M. Castellanos, U. Dayal, and C. Zhai 2011

Regola della negazione: parole di negazione come "no", "non" e "mai" invertono il sentimento della parola di opinione nella stessa frase. Ad esempio, "non felice" dovrebbe avere sentimento positivo come "felice", ma invece ha sentimento negativo grazie alla presenza del "non".

In particolare tale creazione del *sentiment lexicon* L, che tiene conto del contesto, prevede tre fasi: la prima prevede l'estrazione del set degli aspetti F, successivamente vengono estratte le parole d'opinione O che vengono abbinate agli aspetti a cui si riferiscono creando così la coppia (F,O) ed infine vi è l'assegnazione della polarità del *sentiment* S ad ogni coppia aspetto-opinione portando alla tripletta finale (F,O,S) con S che può avere un valore di +1 o -1.

(Immagine 4)



3.2 Mappatura delle coppie aspetto-opinione

Una volta creato il *sentiment lexicon* L, bisogna andare ad individuare le coppie aspetto-opinione (F,S'), dove S' rappresenta l'opinione dell'*opinion holder* sull'aspetto dell'entità menzionata nella recensione.

In particolare il processo di estrazione di questi due elementi è suddiviso in tre passaggi⁴⁸: estrazione, unione e filtraggio. Nella prima fase ogni recensione viene scomposta in un elenco di parole che indicano un aspetto, un'opinione e negazioni, e vengono classificate in base all'ordine in cui appaiono nella recensione. Nella seconda fase, saranno create le coppie di aspetto-opinione in base al contesto e al *sentiment lexicon*. Inoltre, in questo step, viene definita anche la polarità di sentimento di ciascuna coppia. Quindi, nell'ultima fase, viene valutata l'utilizzabilità di ogni coppia estratta.

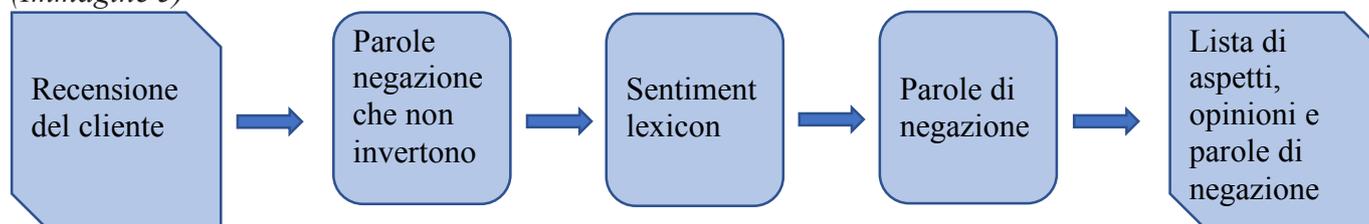
Nel primo step, molte tecniche di estrazione potrebbero dare una ridondanza nei risultati, ad esempio, per la recensione "L'immagine televisiva è chiara", ci sono due coppie di aspetto-opinione che possono essere ricavate: "immagine-chiara" e "immagine televisiva-chiara".

Per risolvere tale problema della ridondanza, si può far sì che vengano estratte le parole di aspetto e quelle di opinione più lunghe possibili nel caso in cui sia possibile estrarre più termini di aspetto o parole di opinione. Utilizzando tale metodo, nell'esempio precedente, verrebbe estratta soltanto la coppia "immagine televisiva-chiara".

Individuare le parole di negazione presenti nelle recensioni è un compito essenziale dato che tali termini potrebbero invertire la polarità del sentimento di una recensione. Quindi si dovrà creare una lista di parole di negazione che modificano il senso della recensione, senza però dimenticare che alcune di queste parole non hanno lo stesso effetto, quindi si dovrà creare anche un insieme di espressioni che, se pur contenente avverbi di negazione, non hanno un significato negativo.

A questo punto si hanno tre risorse a disposizione, il *sentiment lexicon*, la lista delle parole di negazione, e la lista di parole di negazione che non invertono la polarità del sentimento, le quali sono utilizzate per creare una lista di parole di aspetto, di opinione e di negazione da utilizzare nella fase successiva.

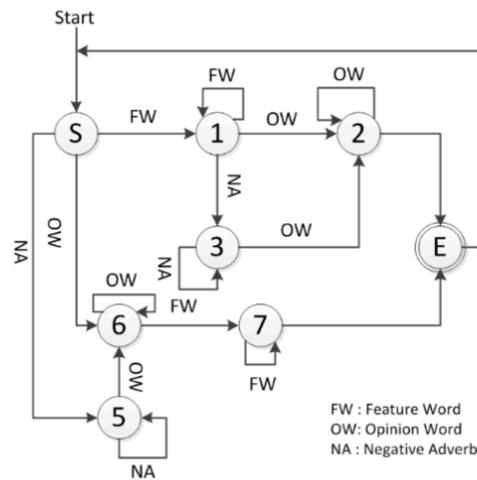
(Immagine 5)



⁴⁸ Yunzhi Tan, Yongfeng Zhang, Min Zhang, Yiqun Liu, Shaoping Ma 2013

Nella seconda fase si utilizza una finite state machine per definire se un aspetto ed una parola d'opinione sono una coppia basandosi sul contesto. L'input di tale metodo è la lista delle parole ottenuto nel primo step. In particolare, la finite state machine può essere rappresentata in questo modo:

(Immagine 6)



Tale modello parte da due presupposti: le parole di negazione appaiono solo davanti alle parole di opinione o ad altre parole di negazione e quanto un utente scrive una recensione su più di un aspetto posizionerà le parole d'opinione o tutte davanti agli aspetti o tutte dopo.

Guardando l'immagine sopra, il punto di partenza della macchina è la S, e successivamente si trasforma in base alla tipologia di parola che incontra fino ad arrivare allo stato finale E dove si avranno una o più coppie di parole.

Un esempio che mostra il funzionamento di tale modello è: “Il colore è luminoso ma il suono non è buono”: in questa recensione abbiamo le parole di aspetto “colore” e “suono”, le parole di opinione “luminoso” e “buono” e la parola di negazione non, quindi il processo è: S → 1 → 2 → E → 1 → 3 → 2 → E.

Dopo questi due step comunque ci potrebbero essere degli errori nelle coppie di parole estratte, per questo motivo vi è l'ultimo passaggio dove si verifica la ammissibilità delle coppie basandosi su 3 regole:

- ordine della parola d'opinione e della parola di aspetto: ci sono molte parole d'opinione che possono apparire sia prima che dopo la parola di aspetto, ma ce ne sono altre che appaiono solitamente o solo prima o solo dopo;
- distanza tra le parole d'opinione e quelle di aspetto: solitamente queste due parole sono vicine all'interno della recensione;
- probabilità che una parola d'opinione ed una di aspetto siano una coppia: questa probabilità viene calcolata nel momento in cui si costruisce il *sentiment lexicon*.

3.3 Matrice di attenzione utente-aspetto

In questo modello si assume che utenti diversi sono interessati a diversi aspetti e che tendono a recensire quelli che per loro sono più importanti. Quindi è utile costruire una matrice di attenzione utente-aspetto X in cui ogni elemento misura quanto ad un utente interessa un determinato aspetto di un prodotto.

In particolare, sia $F = \{F_1, F_2, \dots, F_p\}$ l'insieme di caratteristiche/aspetti espliciti del prodotto, e sia $U = \{u_1, u_2, \dots, u_m\}$ l'insieme che denota gli m utenti, per generare la matrice, si considerano tutte le recensioni di testo scritte da un utente u_i , e vengono estratte tutte le coppie (F, S') . Ipotizzando che la funzione F_j sia menzionata dall'utente u_i per t_{ij} volte, ogni elemento nella matrice di attenzione X è definito come segue:

$$X_{ij} = \begin{cases} 0, & \text{se l'utente } u_i \text{ non menziona l'aspetto } F_j \\ 1 + (N - 1) \left(\frac{2}{1 + e^{-t_{ij}}} - 1 \right), & \text{se è menzionato} \end{cases}$$

N è solitamente uguale a 5 poiché in molti siti le recensioni possono essere accompagnate da un punteggio che va da 1 a 5 stelle, come avviene su Amazon.com e TripAdvisor.

3.4 Matrice di qualità prodotto-aspetto

Altra matrice importante per tale modello è la matrice di qualità prodotto-aspetto Y , dove ogni elemento misura la qualità di un prodotto per un determinato aspetto corrispondente.

Sia $P = \{p_1, p_2, \dots, p_n\}$ l'insieme degli n prodotti. Per ciascuno degli elementi p_i , si utilizzano tutte le recensioni corrispondenti e vengono estratte tutte le coppie (F, S') corrispondenti. Supponendo che l'aspetto F_j sia menzionato per k volte riguardo il prodotto p_i , e che la media del sentimento riguardo la caratteristica F_j nelle quelle k menzioni sia s_{ij} , la misura del prodotto-aspetto Y_{ij} è calcolata così:

$$Y_{ij} = \begin{cases} 0, & \text{se il prodotto } p_i \text{ non è recensito sull'aspetto } F_j \\ 1 + \frac{N-1}{1 + e^{-k s_{ij}}}, & \text{se lo è} \end{cases}$$

Questa misura cattura sia l'orientamento del sentimento (attraverso s_{ij}) sia la popolarità (attraverso k) dell'aspetto F_j per il prodotto p_i .

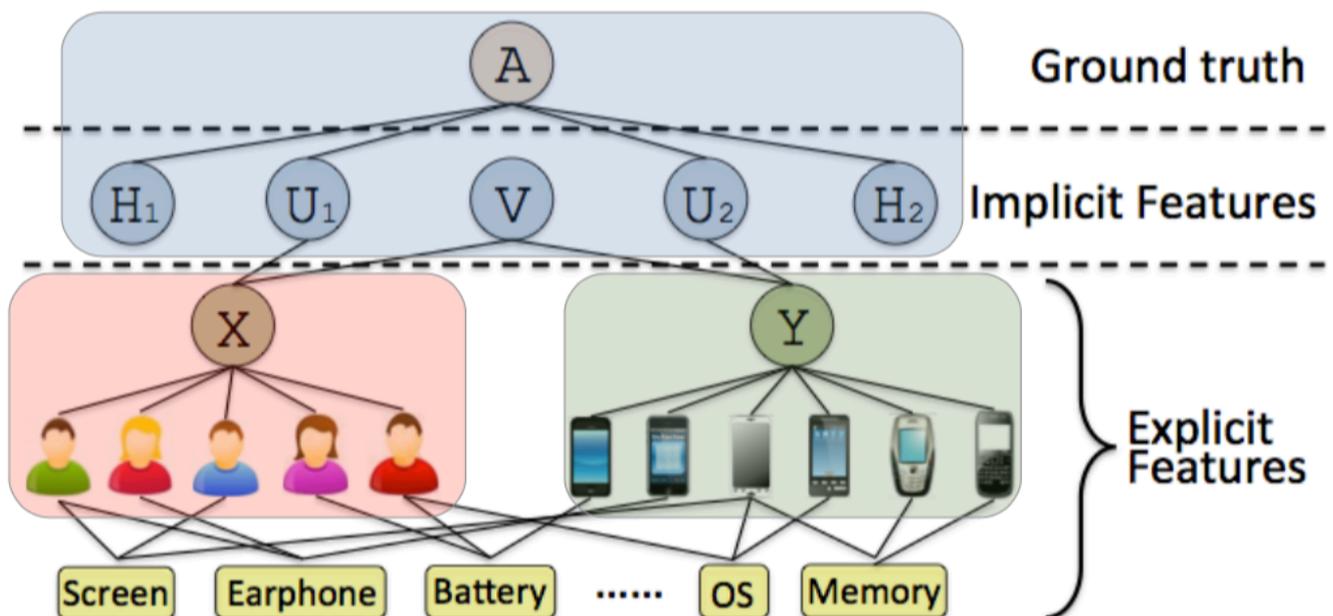
3.5 Integrazione degli aspetti impliciti ed espliciti

Gli elementi diversi da zero nelle matrici X e Y sopra menzionate indicano le relazioni osservate tra utenti, oggetti ed aspetti espliciti. Questi dati vanno poi integrati in un modello di fattorizzazione per ottenere delle raccomandazioni che siano accurate e spiegabili. In particolare, la costruzione di tale modello di fattorizzazione, che utilizza la matrice di attenzione e quella di qualità, si sostanzia nello stimare le valutazioni probabili degli utenti, aspetti ed oggetti basandosi sulle relazioni osservate tra gli utenti e gli aspetti e tra gli oggetti e gli aspetti.

Assumendo che la valutazione complessiva di un utente relativa ad un prodotto è basata sull'opinione complessiva che lui stesso ha su diversi aspetti dell'articolo, considerare solo gli aspetti espliciti (U_1 e U_2), che catturano la qualità dell'articolo e l'attenzione degli utenti, non è sufficiente per spiegare il punteggio assegnato dall'user, e quindi vanno considerati anche i fattori nascosti (*latent factor*, H_1 e H_2) utili nella creazione della matrice di classificazione generale A . Tali relazioni tra gli aspetti del prodotto, le matrici X e Y ed i fattori nascosti sono illustrate nell'immagine sottostante (Immagine 7).

Si noti come, qualora non vengano utilizzati i fattori espliciti, la matrice di valutazione utente-articolo A sarebbe un modello tradizionale di *latent factorization* (visto nel paragrafo 1.3.5).

(Immagine 7)



3.6 Raccomandazioni personalizzate

Una volta trovata la soluzione ottima del modello di fattorizzazione, è possibile stimare qualsiasi elemento mancante nella matrice di attenzione utente-aspetto X , in quella di qualità oggetto-aspetto Y ed in quella utente-oggetto A .

Partendo da questo presupposto, è possibile generare delle raccomandazioni personalizzate Top-k e fornire delle spiegazioni a livello di aspetto.

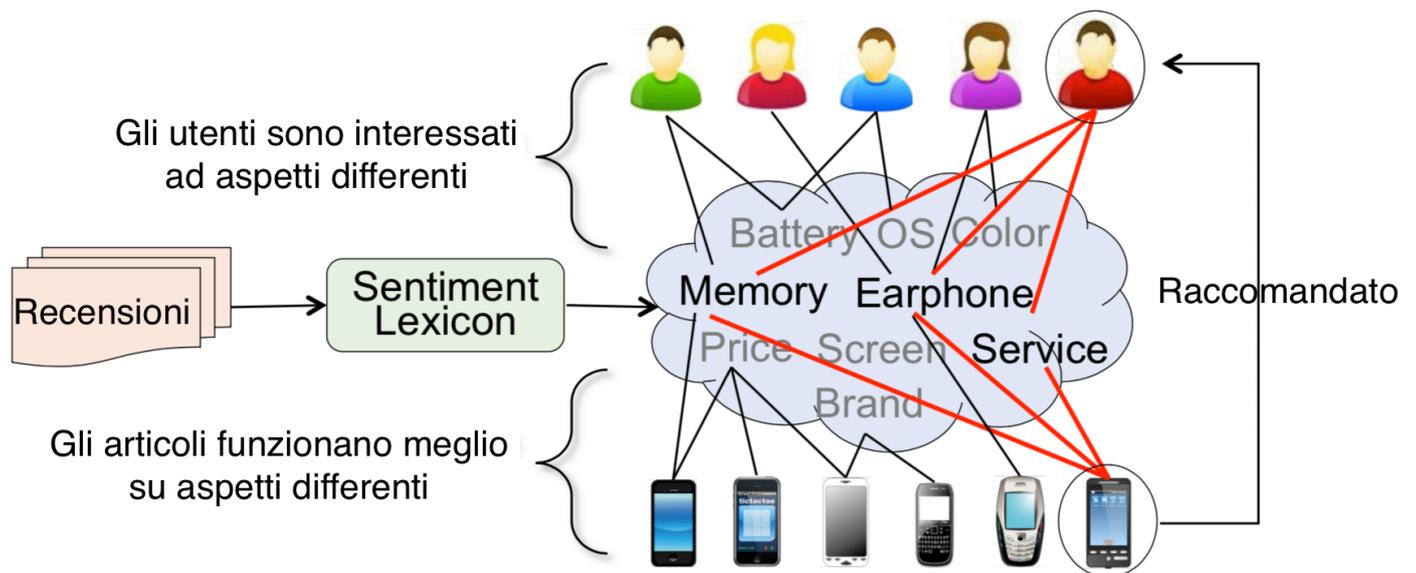
La prima tipologia di raccomandazioni, Top-k, parte dal presupposto che gli utenti, quando devono effettuare un acquisto, basano la loro valutazione su alcuni aspetti particolari del prodotto, e non su tutte le caratteristiche dello stesso. Quindi, in questo metodo, quando viene generato l'elenco delle raccomandazioni personalizzate, si tiene conto dei k aspetti più importanti per un utente. In particolare il punteggio assegnato ad ogni oggetto, che viene utilizzato per creare la classifica delle raccomandazioni, si basa sul punteggio di somiglianza utente-articolo basato sui k aspetti più importanti del prodotto che sono di interesse dell'utente.

Come già detto in precedenza, uno dei limiti dei modelli di fattorizzazione è che sono difficili da comprendere ed è complicato capire perché un determinato articolo viene o meno consigliato. Il vantaggio principale dell'*explicit factor model*, invece, è proprio la capacità di individuare quegli aspetti che fanno sì che un oggetto sia in cima alle raccomandazioni e quindi generare dei consigli con spiegazioni per l'utente. Oltre a raccomandare i prodotti, questo modello ha uno scenario di utilizzo non comune, sconsigliare un prodotto che il sistema considera non adatto all'acquisto quando un utente lo visualizza. Questo appena descritto potrebbe essere un utilizzo importante per un sistema di raccomandazione dato che spiegare perché il sistema pensa che l'utente non debba acquistare un prodotto potrebbe far crescere la fiducia da parte dell'utente e aiutarlo a prendere una decisione d'acquisto più consapevole.

Ci potrebbero essere diversi modi per creare delle raccomandazioni esplicite, ad esempio potrebbe apparire una di queste due frasi quando l'utente si trova sulla pagina di un prodotto: “Potresti essere interessato all'aspetto x , sul quale questo prodotto ha buone performance” o “Potresti essere interessato all'aspetto x , sul quale questo prodotto non ha buone performance”. In particolare i vantaggi delle raccomandazioni esplicite sono principalmente 5:

- intuitività: il sistema agli occhi dell'utente è più trasparente e facile da capire;
- efficacia: vi è un aumento di confidenza e fiducia nel sistema degli utenti che sono in grado di prendere delle decisioni migliori;
- efficienza: gli utenti sono in grado di prendere delle decisioni in maniera più rapida
- persuasività: gli utenti sono maggiormente orientati alla prova ed all'acquisto
- soddisfazione: vi è un aumento di soddisfazione degli utenti poiché aumenta anche la facilità di utilizzo.

(Immagine 8)



3.7 Experimental setup

In questa sezione saranno mostrati gli esperimenti fatti sia offline che online per testare l'*Explicit factor model*. La prima tipologia di esperimenti, quelli offline, sono serviti per valutare l'accuratezza delle raccomandazioni, ovvero la capacità di predire le valutazioni e di effettuare delle raccomandazioni Top-k di un certo livello. Gli esperimenti online, invece, avevano il compito di valutare l'efficacia delle spiegazioni date agli utenti concentrandosi sulla loro persuasività.

In questi esperimenti sono utilizzati due set di dati: uno contenente le recensioni degli utenti sulle attività commerciali situate nella città di Phoenix negli Stati Uniti su Yelp e il set di dati Dianping, costituito dalle recensioni sui ristoranti situati in diverse città principali della Cina. In particolare però il set di dati di Yelp include molti utenti che hanno fatto pochissime recensioni, infatti il 49% di loro ha effettuato solo una recensione, il che renderebbe difficile valutare le prestazioni della raccomandazione top-K. Di conseguenza sono stati selezionati solo gli utenti con 10 o più recensioni.

3.7.1 Esperimenti offline

La prima analisi è stata effettuata sulla costruzione del *sentiment lexicon*, ed i risultati sono stati positivi dato che la precisione nell'individuazione delle parole di opinione, di aspetto e dei sentimenti non è mai risultata essere inferiore all'89%. In particolare si è visto come il *sentiment lexicon* agisse bene in diversi contesti, riuscendo ad individuare i sentimenti che dipendevano dalle parole che indicavano aspetti e quelli che dipendevano dalle parole di opinione, ad esempio se la coppia (F, O) era (prezzo, alto) il sentimento riportato era negativo, mentre se si aveva (qualità del servizio, alta) era ovviamente positivo.

Per verificare l'ipotesi che gli utenti si concentrano su diversi aspetti, è stata costruita la matrice di attenzione utente-aspetto X sul set di dati di Yelp, ed è stata condotta una clusterizzazione per assegnare gli utenti in 5 cluster, ed ognuno di questi conteneva 5 aspetti dei ristoranti ordinati in modo decrescente in base alla loro frequenza di apparizione nelle recensioni. In questo modo si è visto come gli utenti di ciascun cluster erano più interessati ad un determinato sottoinsieme di caratteristiche, ed ogni sottoinsieme era riconducibile ad un dato aspetto del prodotto. Un esempio di questa clusterizzazione è la tabella qui sotto:

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|------------|-----------|------------|-----------|-----------|
| Cibo | Servizio | Posto | Prezzo | Birra |
| Pranzo | Menu | Luogo | Ordini | Drink |
| Servizio | Cibo | Location | Cibo | Bar |
| Pasto | Ordini | Ristorante | Servizio | Ordini |
| Esperienza | Staff | Bar | Menu | Vino |

Una seconda analisi è stata effettuata sul numero di fattori espliciti e fattori latenti da dover utilizzare. Questi due elementi sono importanti perché acquisiscono due diversi tipi di informazioni nel processo di fattorizzazione per la predizione delle valutazioni. Gli esperimenti sono stati effettuati con un numero totale di fattori pari a 100, e si è notato che quando il modello utilizzava tra il 30 e l'80% di fattori espliciti riusciva ad avere delle ottime prestazioni in termini di RMSE (radice dell'errore quadratico medio). Ciò evidenzia come, sebbene l'incorporazione di fattori espliciti migliori la precisione della previsione rispetto ai modelli che non li utilizzano, è necessario utilizzare una moderata quantità di fattori latenti per garantire flessibilità al modello, poiché i fattori espliciti potrebbero non acquisire completamente tutti i criteri che portano l'utente ad una determinata valutazione. Dagli esperimenti è comunque emerso che, finché il numero dei fattori latenti non è sufficiente per catturare le motivazioni nascoste degli utenti, l'utilizzo dei fattori espliciti migliora la previsione delle valutazioni.

Altra analisi ha avuto ad oggetto il numero di caratteristiche più importanti per gli utenti da dover utilizzare per effettuare delle raccomandazioni di tipo *top-k*. In particolare questo esperimento conferma l'ipotesi che

utilizzare troppe caratteristiche introdurrebbe rumore nella procedura di raccomandazione intaccandone l'efficacia.

Per analizzare ulteriormente le relazioni tra le caratteristiche e gli utenti, e capire l'impatto e l'importanza delle caratteristiche esplicite, calcoliamo la copertura media delle caratteristiche più importanti in termini di frequenza di apparizione rispetto a tutte le caratteristiche nelle recensioni di un utente.

In particolare si nota come un piccolo numero di caratteristiche più importanti per gli utenti domina la copertura della frequenza dei termini nelle recensioni degli utenti. Circa 24 delle 96 caratteristiche coprono fino all'80% del totale delle apparizioni delle caratteristiche su Yelp. Ciò implica che gli utenti di solito sono più interessati a poche caratteristiche più importanti, il che suggerisce di utilizzare le migliori k caratteristiche per le raccomandazioni, perché l'utilizzo di più caratteristiche potrebbe determinare effetti negativi dato che potrebbero essere irrilevanti per le preferenze degli utenti.

Tuttavia, sono necessarie 46 caratteristiche per ottenere la stessa copertura dell'80% sul set di dati Dianping, dato che è quasi il doppio di quanto riscontrato sul *dataset* Yelp.

Per capirne il motivo, le varie caratteristiche sono state raggruppate in gruppi di sinonimi e si è visto che diverse parole sono utilizzate per descrivere lo stesso aspetto degli oggetti. Ad esempio, le caratteristiche prezzo e costo sono raggruppate in un singolo cluster. I risultati hanno mostrato che in media ci sono più sinonimi nella lingua cinese, e l'esistenza di sinonimi divide lo spazio delle caratteristiche esplicite. Ad esempio, mentre ci sono due caratteristiche "prezzo" e "costo" raggruppati in un solo cluster su Yelp, il cluster semanticamente corrispondente su Dianping contiene quattro parole. Ciò suggerisce che il numero ottimale di caratteristiche da utilizzare potrebbe essere diverso per le diverse lingue. Le prime 15 caratteristiche su Yelp sono raggruppate in 7 cluster, come mostrato nella tabella sottostante in cui le prime 15 caratteristiche sono in grassetto ed i cluster sono classificati in base alla frequenza totale di apparizione dei termini presenti al loro interno.

| | | | |
|---|---|---|---|
| 1 | Posto, ristorante, location , zona, tragitto | 2 | Cibo, menu, pranzo, pizza , cena |
| 3 | Servizio, tempo, staff , ordini | 4 | Esperienza , qualità |
| 5 | Stanza, atmosfera , arredamento | 6 | Prezzo , costo |
| 7 | Birra , vino, drink, acqua, caffè | | |

3.7.2 Esperimenti online

Tale tipologia di esperimenti è stata condotta con utenti di un sito di e-commerce reale con lo scopo di studiare gli effetti delle raccomandazioni contenenti delle spiegazioni sulla disponibilità degli utenti a seguire tali consigli (persuasività delle raccomandazioni).

In particolare è stato condotto un test A/B su un popolare sito web che vanta più di 100 milioni di utenti. L'obiettivo era quello di consigliare i cellulari migliori per ogni utente.

La schermata presentava 4 cellulari sul lato destro dello schermo e, quando l'utente passava con l'indicatore del mouse su uno di questi, appariva un indicatore che lo consigliava o sconsigliava (come nell'immagine esempio qui sotto) mostrando anche delle coppie di parole aspetto-opinione colorate diversamente a seconda che l'opinione fosse positiva o negativa e di dimensione proporzionale alla frequenza di apparizione di quei termini nelle recensioni dell'oggetto in questione.

(Immagine 9)



Per analizzare gli effetti di tali raccomandazioni, ogni utente è stato assegnato in maniera casuale ad uno di questi tre gruppi: utenti A (gruppo sperimentale) ricevono le spiegazioni a livello di aspetto sull'oggetto preso in considerazione, gli utenti B (gruppo di confronto) ricevono le spiegazioni di tipo "gli utenti hanno visto anche" e utenti C (gruppo di controllo) non ricevono spiegazioni. I risultati ottenuti da questo test sono i seguenti:

| Gruppo di utenti | A | | B | | C | |
|------------------|-----------------|-------|-----------------|-------|-----------------|-------|
| Dati registrati | Visualizzazioni | Click | Visualizzazioni | Click | Visualizzazioni | Click |
| | 15933 | 691 | 11483 | 370 | 17265 | 552 |
| CTR | 4,34% | | 3,22% | | 3,20% | |

Come si vede, il *Click Through Rate* (CTR), calcolato come rapporto tra le visualizzazioni ed i click, del gruppo sperimentale A è maggiore rispetto a quello degli altri due gruppi, dimostrando che le spiegazioni a livello di aspetto fornite dall'EFM sono più efficaci nel persuadere gli utenti ad esaminare più attentamente gli articoli consigliati.

Per studiare come le “dis-raccomandazioni” influenzano gli utenti, invece, sono stati utilizzati solo due gruppi di utenti, il gruppo A, quello sperimentale, dove gli utenti potevano vedere le spiegazioni quando passavano l'indicatore del mouse sull'oggetto, e quello B dove gli utenti non ricevevano alcuna spiegazione dato che non ci sono sistemi che forniscono delle spiegazioni per una dis-raccomandazione.

I dati raccolti sono stati: 53.372 visualizzazioni di 1.328 articoli, che includono 20.735 visualizzazioni da 582 utenti del gruppo A e 32.637 visualizzazioni da 733 utenti del gruppo B. Gli indicatori utilizzati per confrontare i due gruppi di utenti erano: percentuale di utenti che avevano aggiunto al carrello un determinato prodotto e le percentuali di utenti che sono d'accordo con la raccomandazione o la dis-raccomandazione.

Anche in questo caso i dati del gruppo sperimentale A sono migliori rispetto a quelli del gruppo di controllo B, e ciò dimostra che la presentazione delle spiegazioni migliora la capacità persuasiva delle raccomandazioni e allo stesso tempo migliora la conversione, cioè l'aggiunta al carrello, negli scenari di acquisto online reali.

Quindi si può dire che le spiegazioni aiutano a persuadere un utente ad aggiungere un prodotto consigliato al carrello o ad ignorarne uno non consigliato.

3.8 Vantaggi per i consumatori e le imprese

Questo modello, con tutte le sue possibili applicazioni e varianti, rappresenterebbe una svolta nel modo di analizzare i dati ed utilizzare le informazioni in essi contenute, garantendo un importante vantaggio competitivo all'azienda che per prima lo inserisse nel suo business. In particolare, i principali vantaggi di utilizzare un sistema di raccomandazione che dia dei consigli spiegando il perché sono: il sistema diventa più trasparente e ha il potenziale di consentire agli utenti di dire quando il sistema sbaglia, aumenta la confidenza e la fiducia degli utenti nel sistema, aiuta gli utenti a prendere decisioni migliori (maggior efficacia) e più rapide (maggior efficienza), convince gli utenti a provare o acquistare (maggior persuasività) e aumenta la soddisfazione di questi ultimi.

Utilizzando questo metodo, inoltre, è possibile creare dei profili di utenti e di attività, riuscendo ad individuare quegli elementi che sono particolarmente importanti per un utente in un determinato contesto e riuscendo anche ad estrapolare i prodotti/servizi particolarmente apprezzati dai consumatori di una determinata attività. Sfruttando questi dati, ad esempio, è possibile quindi consigliare ad un determinato utente, in base ai suoi gusti ed al ristorante scelto, che tipo di piatto ordinare, andando ad influenzare la sua esperienza e la sua soddisfazione. Ad esempio si pensi ad un utente che si trova a dover scegliere tra il dover ordinare una pizza o un panino in un ristorante, a seconda della scelta che farà, probabilmente darà valutazioni diverse all'attività in questione, pertanto, l'esperienza dell'utente con una particolare impresa può essere migliorata raccomandando alcuni aspetti particolari, come ordinare la pizza in quel ristorante.

Ovviamente non tutti gli aspetti di un'esperienza possono essere selezionati dall'utente al fine di migliorare la sua soddisfazione, ad esempio, nel caso di un film, aspetti come la trama del film o gli attori sono al di fuori del controllo dell'utente.

Stesso discorso può essere fatto dal lato delle imprese. È possibile, grazie a questo metodo, consigliare ai gestori di una data attività determinate azioni da compiere nei confronti di un certo utente per far sì che la sua esperienza sia migliore e sia personalizzata, ad esempio si potrebbe suggerire ai gestori di una SPA di offrire un determinato drink che il sistema ha stimato che sarebbe apprezzato dall'utente, migliorando così l'esperienza di quest'ultimo nel centro benessere.

In particolare, per poter fornire questa tipologia di consigli ad utenti e gestori, andrebbero selezionati manualmente due gruppi di aspetti tra tutti quelli di un determinato contesto: quelli su cui l'utente ha il controllo e quelli su cui la gestione dello stabilimento ha il controllo. Ad esempio, l'aspetto "SPA" in un hotel è sotto il controllo dell'utente, perché è lui che può decidere se usarla o meno durante la sua permanenza in quella struttura. Ovviamente, all'interno di questi due gruppi, vanno identificati gli aspetti più importanti che si possono raccomandare all'utente. Queste raccomandazioni possono essere positive, suggerimento di sperimentare un aspetto, o negative, suggerimento di evitare un aspetto.

Ad esempio, se il sistema ha identificato l'aspetto "pizza" come ad alto impatto positivo sulle valutazioni di un ristorante, raccomandando questo locale sarà suggerito all'utente di ordinare la pizza. Allo stesso modo, se l'aspetto "torta" ha un forte impatto negativo sul rating, si potrebbe comunque raccomandare di visitare quel ristorante all'utente con il suggerimento però di non ordinare le torte lì.

In particolare quando si parla di esperienza ci si riferisce al qui ed ora riguardo le percezioni, le emozioni, i pensieri quando si incontrano e testano prodotti/servizi e brand nelle attività di consumo ma anche al ricordo di quell'esperienza. Per molti prodotti/servizi la memoria di quell'esperienza vissuta è altrettanto importante e, per alcune di queste esperienze, la memoria è addirittura più importante.

Quindi, riuscire a migliorare l'esperienza di consumo di un consumatore porta dei vantaggi notevoli all'impresa oltre che al consumatore, il quale, ricordando l'esperienza positiva, probabilmente la condividerà

su siti di recensioni appositi andando così ad aumentare i dati a disposizione del sistema per fornire delle raccomandazioni personalizzate sempre migliori.

Altra possibile applicazione di tale modello potrebbe essere quella di aiutare gli utenti nella scelta della meta o di un alloggio per un determinato viaggio. Attualmente sono moltissimi i blog ed i siti che hanno ad oggetto i viaggi, sono moltissimi gli utenti che raccontano le loro esperienze, ed ancor di più gli utenti che cercano informazioni per programmare un viaggio. Secondo gli ultimi dati Istat, infatti, nel 2016 gli italiani hanno effettuato 66,5 milioni di viaggi con pernottamento. Di questi spostamenti di lavoro o piacere, in Italia o verso l'estero, circa 13 milioni (il 19% del totale) sono stati prenotati direttamente via Internet⁴⁹. Come mostrato da questi dati, quindi, il settore del turismo online è sempre in aumento e va quindi sfruttato al meglio, e poter consigliare ad un determinato utente X, alla ricerca di una meta rilassante, una meta Y che ha ricevuto numerose recensioni positive sull'aspetto "relax", potrebbe fare la differenza e consentire a quel sito web di catturare l'attenzione dell'user ed assicurarsi la vendita di un pacchetto di viaggio, questione molto difficile come mostrato da un'indagine di PhoCusWright che mostra come i consumatori usino in media 6,5 siti e dispositivi per effettuare le ricerche preliminari alla prenotazione⁵⁰. Quindi portare queste persone su un sito è difficile, e quando ci arrivano, bisogna essere in grado di trattenerli e convertirli in acquirenti riuscendo a consigliare loro la meta e l'alloggio migliori possibile.

In questo capitolo è stato analizzato *l'Explicit Factor Model* e i suoi possibili vantaggi, successivamente sarà presentato il questionario che sarà utilizzato per estrapolare dei dati utili per verificare l'ammissibilità e l'utilizzabilità di tale modello valutando le abitudini dei consumatori italiani.

⁴⁹

http://www.repubblica.it/economia/rapporti/paesedigitale/trend/2017/04/28/news/turismo_dalle_prenotazioni_online_un_indotto_da_oltre_2_miliardi-164087252/

⁵⁰ <https://www.insidemarketing.it/turismo-il-65-delle-prenotazioni-di-viaggi-e-vacanze-avviene-online/>

4 Questionario e risultati

In questo capitolo sarà presentato il questionario che ho creato e somministrato a dei consumatori italiani raccogliendo 140 risposte totali.

Obiettivo di tale questionario era andare a raccogliere dati riguardo le modalità con cui i consumatori italiani si approcciano alle recensioni online, riguardo l'applicabilità dell'*Explicit Factor Model* in Italia, dato che il modello è stato testato su database americani e cinesi, e riguardo il grado di interesse per alcune applicazioni da me ipotizzate di tale modello che, se applicato, garantirebbe un valore aggiunto non indifferente ad ogni azienda operante online data la centralità delle recensioni nelle scelte di acquisto di ogni consumatore e visto che sarebbe una tecnologia innovativa e non ancora utilizzata da nessun sito web.

4.1 Struttura del questionario

Il questionario presenta una struttura divisa in 5 sezioni:

1. informazioni generali
2. informazioni sulle modalità di scrittura delle recensioni
3. proposta di un *Explicit Factor Model* per siti di e-commerce
4. proposta di un *Explicit Factor Model* per siti di recensioni di attività ristorative
5. proposta di un *Explicit Factor Model* per siti di organizzazione di viaggi online

La prima sezione comprende le domande su età, sesso e sull'aver mai scritto o meno una recensione.

(Immagini 10-11)

The image shows a dark-themed questionnaire interface. On the left, there are two sections: 'Sesso' (Gender) with options 'Maschio' (Male) and 'Femmina' (Female), and 'Età' (Age) with options '<18', '18-25', '26-35', '35-45', and '45+'. On the right, there is a question: 'Hai mai scritto una recensione su un qualunque sito web (e-commerce, hotel, viaggi, ristoranti...)?' (Have you ever written a review on any website (e-commerce, hotel, travel, restaurants...)?). Below the question are two radio button options: 'si' (yes) and 'no'.

La seconda, invece, appariva solo a coloro avessero risposto positivamente alla terza domanda, dato che riguardava le modalità con cui solitamente viene scritta una recensione. In particolare le domande di questa categoria sono sui prodotti o servizi sui quali il rispondente aveva effettuato una recensione (Immagine 12), con possibilità di scegliere anche più di una sola risposta, e sugli aspetti sui quali si concentrava quando ne scriveva una, aspetti più importanti per lui (Immagine 13), aspetti che si sono rivelati maggiormente positivi o quelli maggiormente negativi (Immagini 14-15).

(Immagini 12-13)

| Su quali prodotti/servizi hai scritto delle recensioni (sono ammesse più risposte)? | Indica il tuo grado di accordo o disaccordo con le seguenti affermazioni |
|---|--|
| Ristoranti/bar | quando scrivo una recensione mi concentro sugli aspetti del prodotto/servizio più importanti per me (1 fortemente in disaccordo – 5 fortemente d'accordo). |
| Hotel | 1 |
| Viaggi | 2 |
| Articoli e-commerce | 3 |
| Altro | 4 |
| | 5 |

(Immagini 14-15)

| quando scrivo una recensione mi concentro sugli aspetti del prodotto/servizio che si sono rivelati particolarmente positivi (1 fortemente in disaccordo – 5 fortemente d'accordo). | quando scrivo una recensione mi concentro sugli aspetti del prodotto/servizio che si sono rivelati particolarmente negativi (1 fortemente in disaccordo – 5 fortemente d'accordo). |
|--|--|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |

Nella terza sezione veniva chiesto al rispondente di immaginare un sito di e-commerce che utilizzi le recensioni sue e degli altri utenti per creare dei suggerimenti per gli acquisti, e di mostrare l'apprezzamento o meno, su una scala da 1 a 5 con valori in ordine crescente di opinione favorevole, per le funzioni aggiuntive che porterebbe questo metodo ai sistemi di raccomandazione.

(Immagini 16-17)

| | |
|---|---|
| Immagina un sito web di e-commerce (ad es. Amazon) | Quanto ti piacerebbe ricevere dei consigli dal sito web sul perché NON acquistare un determinato prodotto (ad es. "questo cellulare non è adatto a te poiché ha una pessima batteria" dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)? (1 per niente – 5 molto) |
| Quanto ti piacerebbe ricevere delle raccomandazioni personalizzate con delle spiegazioni sul perché dovresti acquistare un determinato prodotto (ad es. "questo cellulare potrebbe interessarti poiché ha un'ottima batteria" dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)? (1 per niente – 5 molto) | |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |

Stessa struttura presentava la quarta sezione, con la differenza che si chiedeva al rispondente di immaginare un sito di recensioni per attività ristorative online, che utilizzasse le recensioni per consigliare ristoranti, bar e cosa ordinare o meno al loro interno.

(Immagini 18-19)

| | |
|---|--|
| Immagina un sito web di recensioni su ristoranti e bar (ad es. TripAdvisor) | Quanto ti piacerebbe ricevere dei consigli su cosa ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in questo ristorante fanno un'ottima pizza")? |
| Quanto ti piacerebbe ricevere dei consigli su quali ristoranti/bar visitare in base alle tue recensioni passate su altre attività (ad es. "questo ristorante potrebbe piacerti perché ha ottima carne" dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto carne)? (1 per niente – 5 molto) | |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |

Quanto ti piacerebbe ricevere dei consigli su cosa NON ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in questo ristorante la pizza non ha un'ottima valutazione")?

| |
|---|
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |

Infine nell'ultima sezione si ha la stessa idea di base, ma questa volta si considera un sito di prenotazione di viaggi online in grado di fornire raccomandazioni riguardo le strutture alberghiere o le mete da visitare una volta analizzati i gusti degli utenti.

(Immagine 20-21)

| Immagina un sito web di prenotazioni di viaggi online (ad es. Booking) | Quanto ti piacerebbe ricevere dei consigli su che paese/città visitare nel tuo prossimo viaggio basati sulle tue recensioni passate su altre mete (ad es. consiglia una meta con un punteggio alto in divertimento poiché in una recensione passata hai criticato tale aspetto di un'altra meta) (1 per niente – 5 molto)? |
|--|--|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |

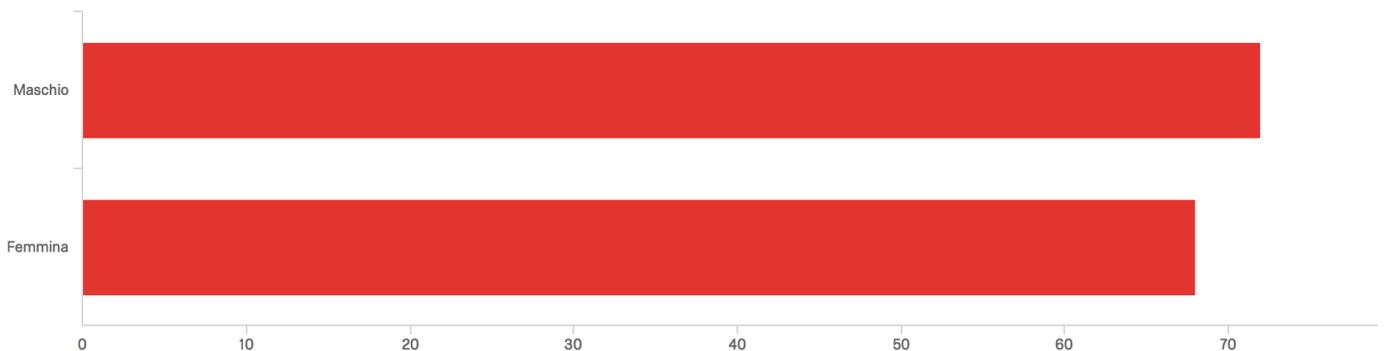
4.2 Risultati del questionario

In questo paragrafo saranno mostrati i risultati ottenuti nel questionario sottoposto ai rispondenti. In particolare saranno esposti divisi nelle 5 sezioni sopracitate.

4.2.1 Prima sezione: informazioni generali

Come detto in precedenza, i rispondenti totali sono stati 140, di cui 72 uomini e 68 donne, rispettivamente il 51% ed il 49% del totale, mostrando così un sostanziale equilibrio tra i due generi.

(Immagine 22)

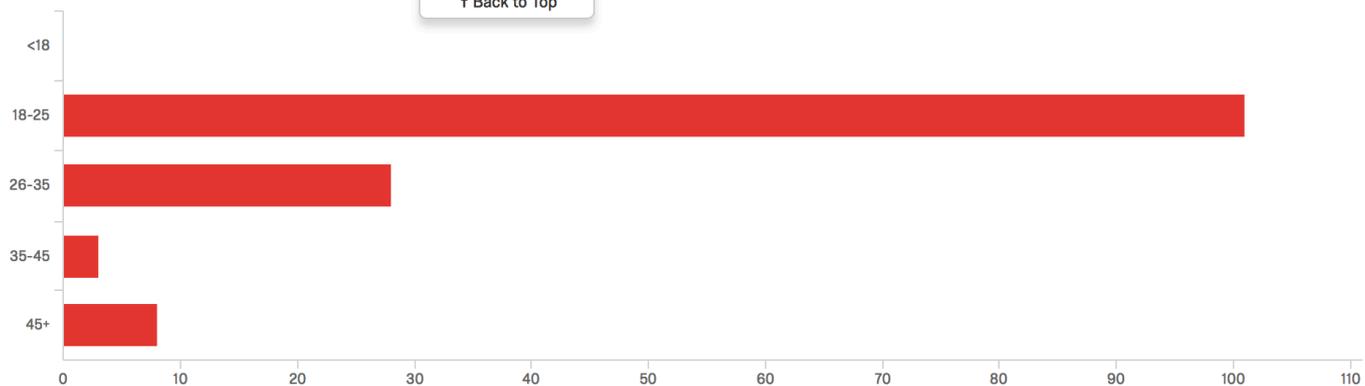


| # | Field | Choice Count |
|---|---------|--------------|
| 1 | Maschio | 51.43% 72 |
| 2 | Femmina | 48.57% 68 |
| | | 140 |

Questo questionario è stato diffuso tramite Facebook, all'interno di gruppi principalmente studenteschi, e tramite WhatsApp, e il riflesso di queste scelte si ha nelle risposte ottenute nella seconda domanda, quella sull'età dei rispondenti. In particolare, la maggior parte di loro, nello specifico il 72% del totale, ha tra i 18 ed i 25 anni. Nell'immagine di seguito i dati completi.

(Immagine 23)

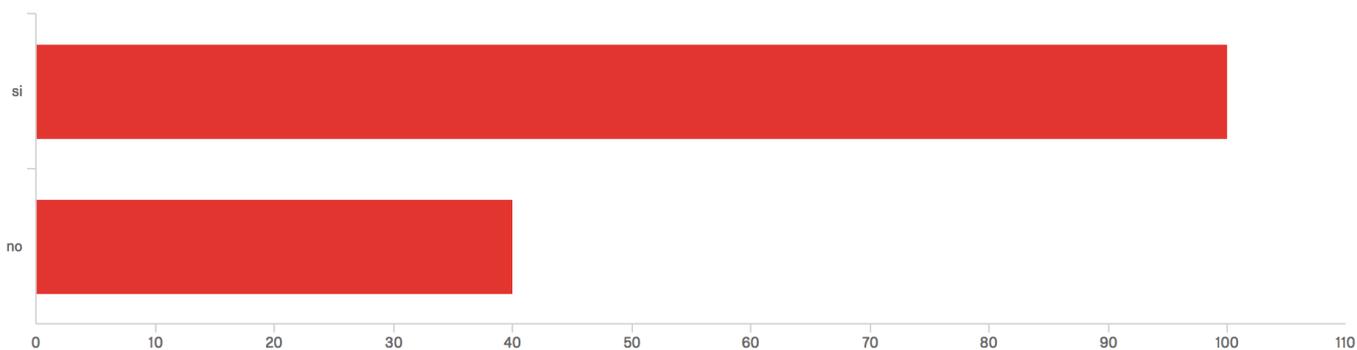
↑ Back to Top



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | <18 | 0.00% 0 |
| 2 | 18-25 | 72.14% 101 |
| 3 | 26-35 | 20.00% 28 |
| 4 | 35-45 | 2.14% 3 |
| 5 | 45+ | 5.71% 8 |

La terza domanda è quella che permette di effettuare una scrematura per le successive quattro, dato che è quella che chiede di indicare se si è mai effettuata una recensione su un qualunque prodotto o una qualsiasi attività. Anche in questo caso si ha una risposta che sovrasta l'altra, ovvero il 71% dei rispondenti ha effettuato almeno una recensione, in particolare 100 rispondenti su 140 totali.

(Immagine 24)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | si | 71.43% 100 |
| 2 | no | 28.57% 40 |
| | | 140 |

Analizzando più nello specifico i dati ottenuti da queste ultime due domande, si nota come l'età influenzi la propensione a scrivere una recensione o meno. In particolare 26 dei rispondenti che hanno tra i 18 ed i 25 non ha mai effettuato una recensione, in percentuale il 25% del totale. Passando alle fasce d'età successive, con l'età cresce anche la percentuale di utenti che non ha mai scritto una recensione di nessun tipo, nello specifico, il 29% dei rispondenti che hanno tra i 26 ed i 35 anni, il 33% di quelli che hanno tra i 35 ed i 45 anni e infine il 62% di utenti che hanno più di 45 anni non ne hanno mai scritta una.

(Immagine 25)

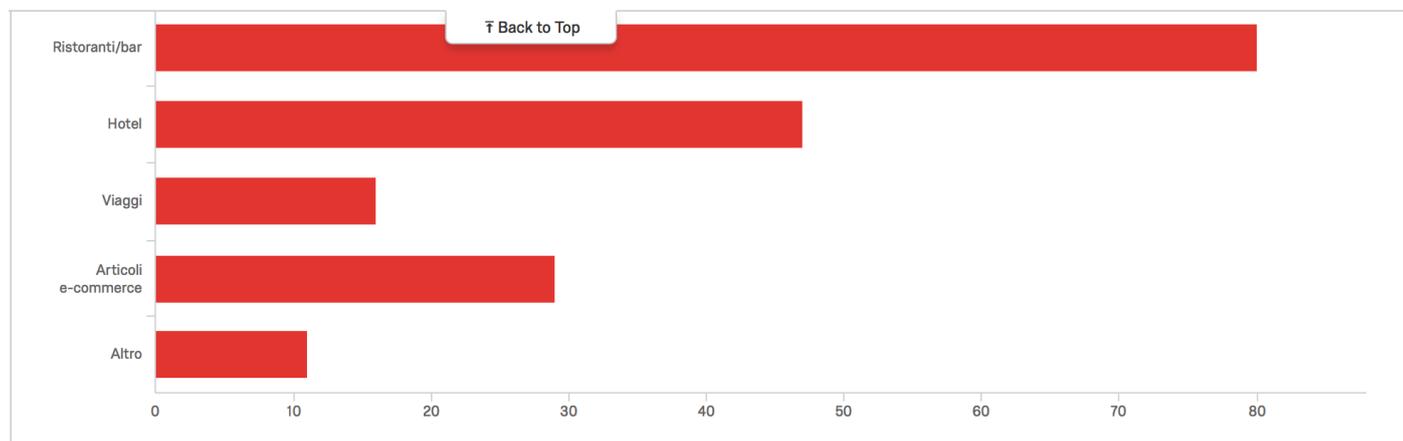
| | | Hai mai scritto una recensione su un qualunque sito web (e-commerce, hotel, viaggi, ristoranti...)? | | Total |
|-----|-------|---|----|-------|
| | | si | no | |
| Età | <18 | 0 | 0 | 0 |
| | 18-25 | 75 | 26 | 101 |
| | 26-35 | 20 | 8 | 28 |
| | 35-45 | 2 | 1 | 3 |
| | 45+ | 3 | 5 | 8 |
| | Total | 100 | 40 | 140 |

Questa analisi ci mostra come quindi i dati siano stati raccolti, principalmente, tra gli utenti appartenenti alla fascia d'età che effettua, in percentuale, più recensioni online.

4.2.2 Seconda sezione: informazioni sulle modalità di scrittura delle recensioni

La domanda successiva, apparsa solo ai 100 utenti che hanno risposto positivamente al terzo quesito, ha mostrato che la maggior parte delle recensioni, il 44% del totale, viene effettuata su ristoranti e bar, il 26% sugli hotel, il 16% su articoli di e-commerce e rispettivamente il 9% ed il 6% su viaggi ed altre categorie non comprese in queste 4.

(Immagine 26)



| # | Field | Choice Count |
|---|---------------------|--------------|
| 1 | Ristoranti/bar | 43.72% 80 |
| 2 | Hotel | 25.68% 47 |
| 3 | Viaggi | 8.74% 16 |
| 4 | Articoli e-commerce | 15.85% 29 |
| 5 | Altro | 6.01% 11 |

Anche le successive tre domande sono state mostrate solo ai 100 utenti che hanno risposto “sì” alla terza domanda, in quanto sono dei quesiti che puntano ad identificare cosa spinge gli utenti a recensire e quali sono gli aspetti sui quali si concentrano.

In particolare la prima domanda chiedeva, da 1 a 5, quanto, nel momento di scrivere una recensione, la concentrazione fosse rivolta sugli aspetti più importanti per loro, la seconda e la terza, invece, domandavano all’utente quanto si concentrasse sugli aspetti maggiormente positivi o su quelli maggiormente negativi dell’esperienza vissuta.

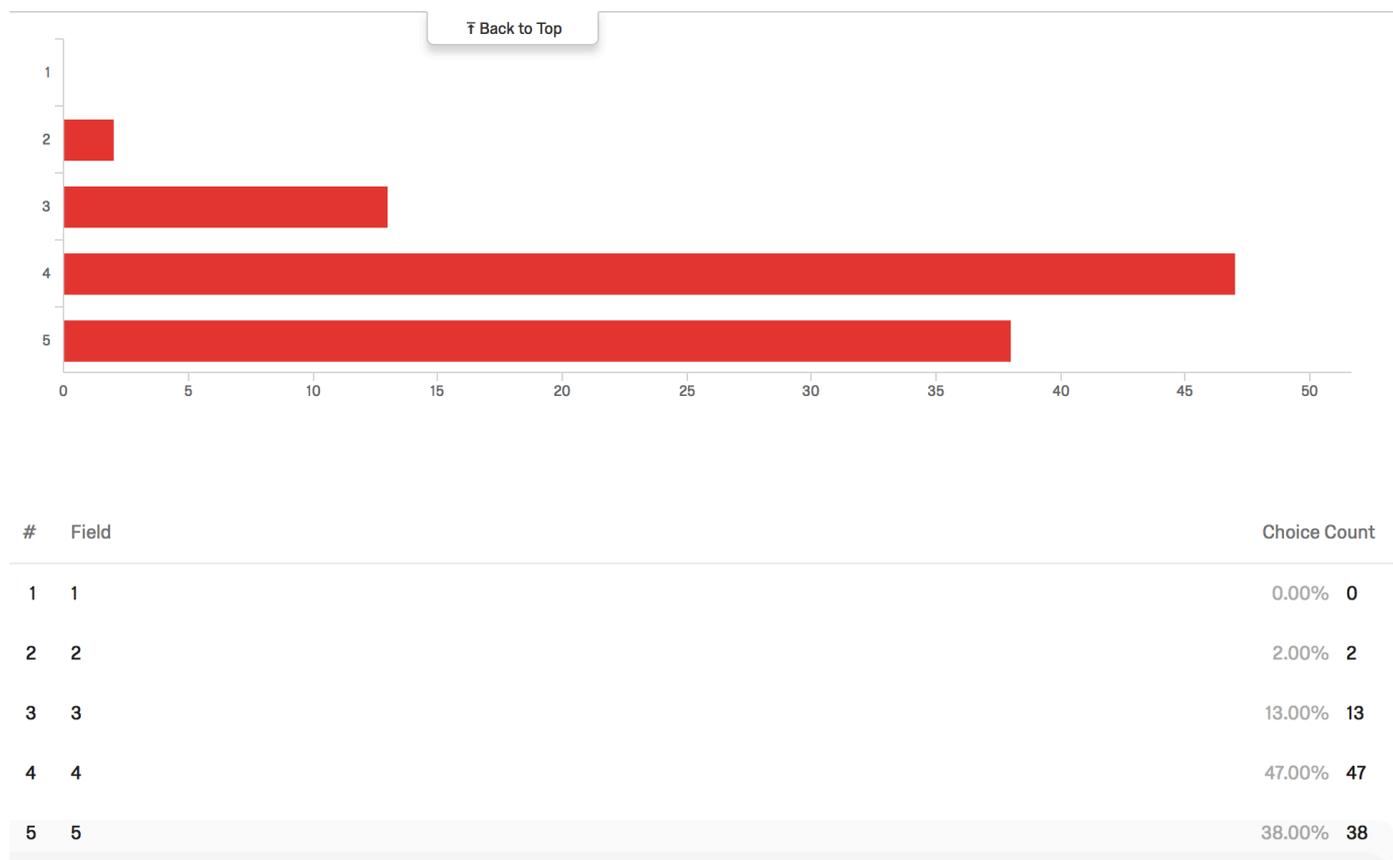
I risultati hanno mostrato come l’attenzione principale sia sugli aspetti maggiormente importanti per l’utente, visto che la prima domanda (Immagine 27) ha raggiunto un totale di 85 rispondenti che si sono detti “d’accordo”, valore di 4, e “fortemente d’accordo”, valore di 5.

Gli altri due quesiti, invece, hanno raccolto, rispettivamente, 64 e 68 “d’accordo” e “molto d’accordo”. Va notato, però, che la terza domanda (Immagine 29), quella che chiedeva ai rispondenti se si concentravano sugli

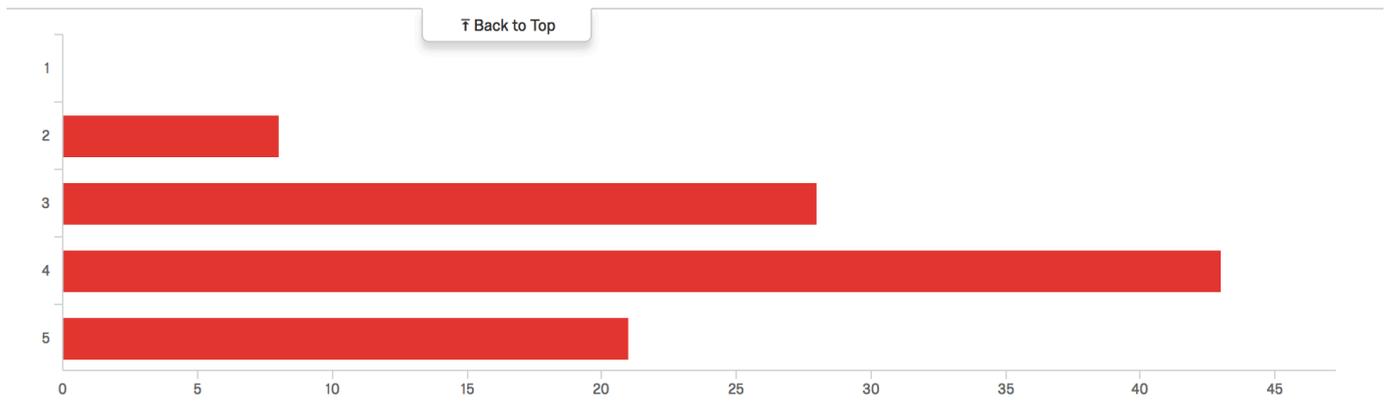
aspetti che si erano rivelati particolarmente negativi nella loro esperienza, ha ottenuto, rispetto alle altre due, un maggior numero di “fortemente d’accordo” rispetto a quello di “d’accordo”, e questo ci mostra come spesso sia proprio l’esperienza negativa a portare il consumatore a scrivere una recensione.

Nelle immagini di seguito i dati completi.

(Immagine 27)

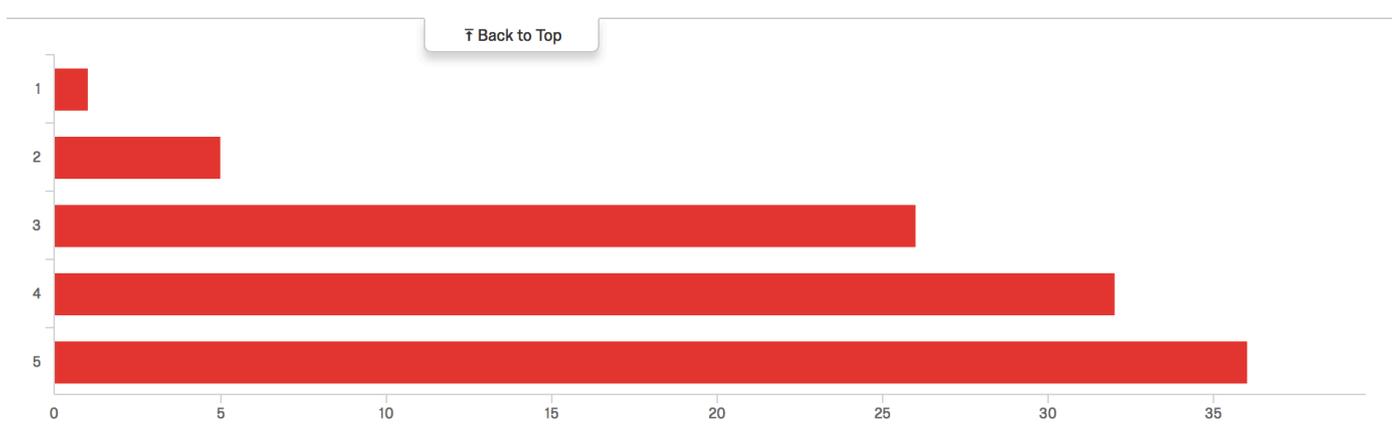


(Immagine 28)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 0.00% 0 |
| 2 | 2 | 8.00% 8 |
| 3 | 3 | 28.00% 28 |
| 4 | 4 | 43.00% 43 |
| 5 | 5 | 21.00% 21 |

(Immagine 29)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 1.00% 1 |
| 2 | 2 | 5.00% 5 |
| 3 | 3 | 26.00% 26 |
| 4 | 4 | 32.00% 32 |
| 5 | 5 | 36.00% 36 |

4.2.3 Terza sezione: proposta di un explicit factor model per siti di e-commerce

Le due domande successive sono quelle facenti parte della terza sezione, dove in particolare veniva chiesto ai rispondenti di immaginare un sito di e-commerce, tipo Amazon.com, dove venissero fornite delle raccomandazioni con delle spiegazioni sul perché bisognerebbe acquistare o meno un determinato prodotto. Queste due domande hanno lo scopo di verificare l'applicabilità dell'idea esposta da Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma nel loro paper.

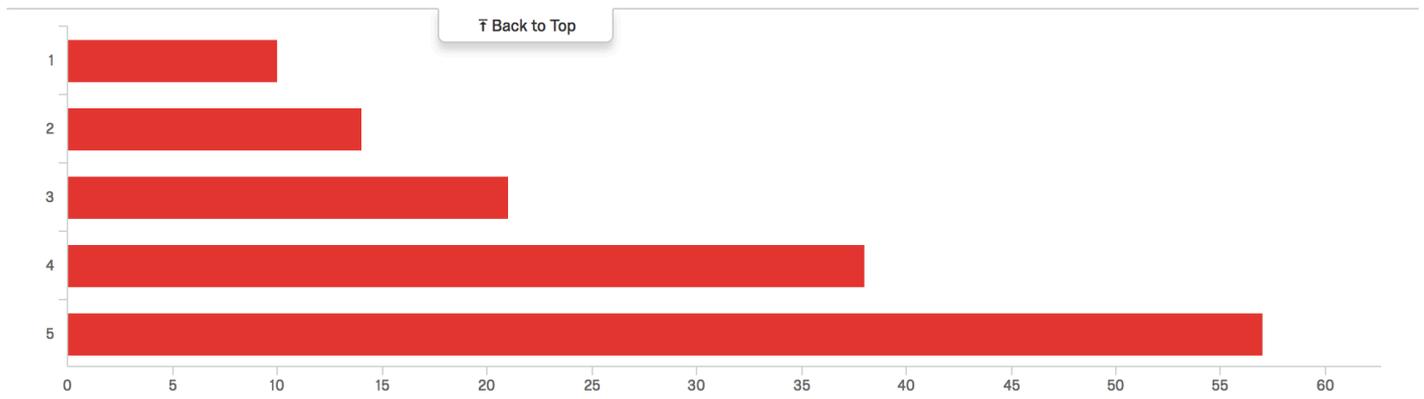
I risultati si sono rivelati particolarmente positivi, in particolare la prima domanda recita: “Quanto ti piacerebbe ricevere delle raccomandazioni personalizzate con delle spiegazioni sul perché dovresti acquistare un determinato prodotto (ad es. “questo cellulare potrebbe interessarti poiché ha un'ottima batteria” dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)?”, e la seconda: “Quanto ti piacerebbe ricevere dei consigli dal sito web sul perché NON acquistare un determinato prodotto (ad es. “questo cellulare non è adatto a te poiché ha una pessima batteria” dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)?”.

Come detto entrambe hanno ricevuto riscontri positivi, ma la seconda (Immagine 31) ha ottenuto addirittura il 60% di voti uguali a 5, il valore massimo attribuibile, per un totale di ben 84 voti su 140, percentuale che sale al 77% se si considerano anche i voti ricevuti dal quarto valore, di un 10% maggiore rispetto alla somma dei voti, degli stessi due valori, ottenuta dalla prima domanda.

Questo ci mostra come i consumatori evidentemente gradirebbero particolarmente un sistema che posso fargli risparmiare tempo e denaro, ciò con ovvi vantaggi anche per l'azienda che probabilmente riuscirebbe a fidelizzarli aumentando così i ricavi.

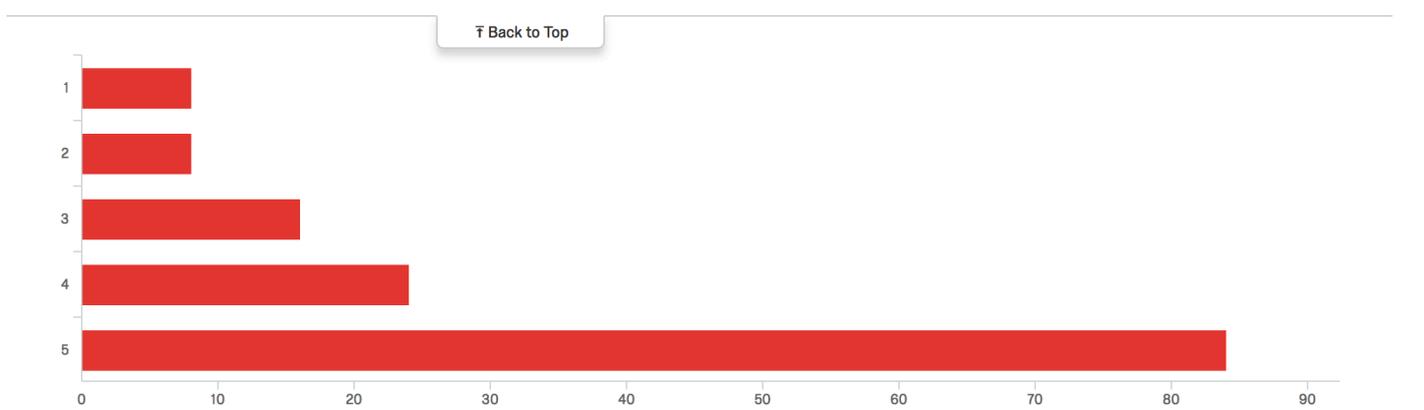
Di seguito le immagini che mostrano i dati nel dettaglio.

(Immagine 30)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 7.14% 10 |
| 2 | 2 | 10.00% 14 |
| 3 | 3 | 15.00% 21 |
| 4 | 4 | 27.14% 38 |
| 5 | 5 | 40.71% 57 |

(Immagine 31)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 5.71% 8 |
| 2 | 2 | 5.71% 8 |
| 3 | 3 | 11.43% 16 |
| 4 | 4 | 17.14% 24 |
| 5 | 5 | 60.00% 84 |

4.2.4 Quarta sezione: proposta di un explicit factor model per siti di recensioni di attività ristorative

Le successive tre domande sono quelle appartenenti alla quarta e penultima sezione, quella che ha ad oggetto ristoranti e bar. La struttura è simile a quella delle due domande precedenti, ovvero viene chiesto al rispondente di immaginare un sito web di recensioni di attività ristorative come TripAdvisor, e le tre domande sono le seguenti: “Quanto ti piacerebbe ricevere dei consigli su quali ristoranti/bar visitare in base alle tue recensioni passate su altre attività (ad es. “questo ristorante potrebbe piacerti perché ha ottima carne” dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto carne)?”, “Quanto ti piacerebbe ricevere dei consigli su cosa ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar (“in questo ristorante fanno un’ottima pizza”)?” e “Quanto ti piacerebbe ricevere dei consigli su cosa NON ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar (“in questo ristorante la pizza non ha un’ottima valutazione”)?”.

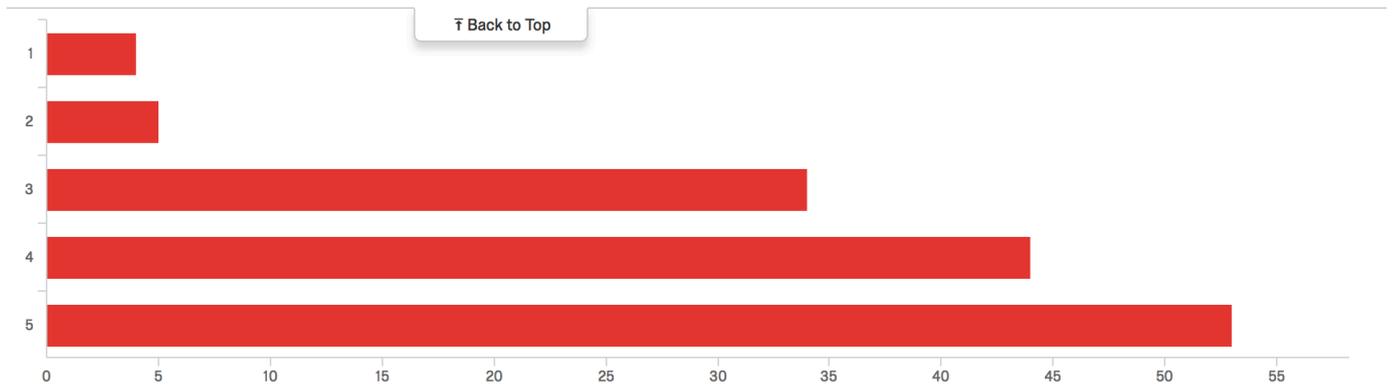
In particolare questo non è un possibile utilizzo dell’*Explicit Factor Model* esposto nel *paper*, ma è una sua possibile applicazione da me ipotizzata.

Anche in questa sezione i risultati sono positivi, anche se meno rispetto a quelli della precedente, in particolare le tre domande hanno ottenuto rispettivamente le percentuali di 68%, 65% e 62% di votazioni per il quarto ed il quinto valore, i due maggiori.

Un altro dato balza agli occhi, la percentuale di voti registrati per il valore 3, quello intermedio, che è rappresentazione di una certa indifferenza al servizio potenzialmente offerto. In particolare le percentuali sono rispettivamente del 24%, 23% e 26%, per le tre domande in questione, valori discretamente maggiori rispetto a quelli del 15% ed 11% ricevuti dai due quesiti precedenti.

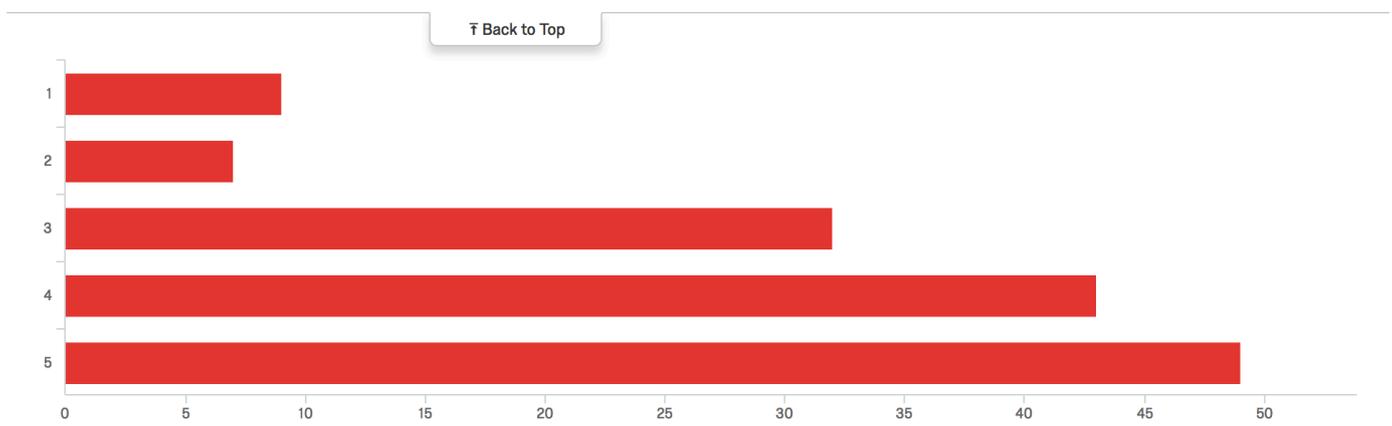
Di seguito le immagini con tutti i voti ottenuti da ogni valore.

(Immagine 32)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 2.86% 4 |
| 2 | 2 | 3.57% 5 |
| 3 | 3 | 24.29% 34 |
| 4 | 4 | 31.43% 44 |
| 5 | 5 | 37.86% 53 |

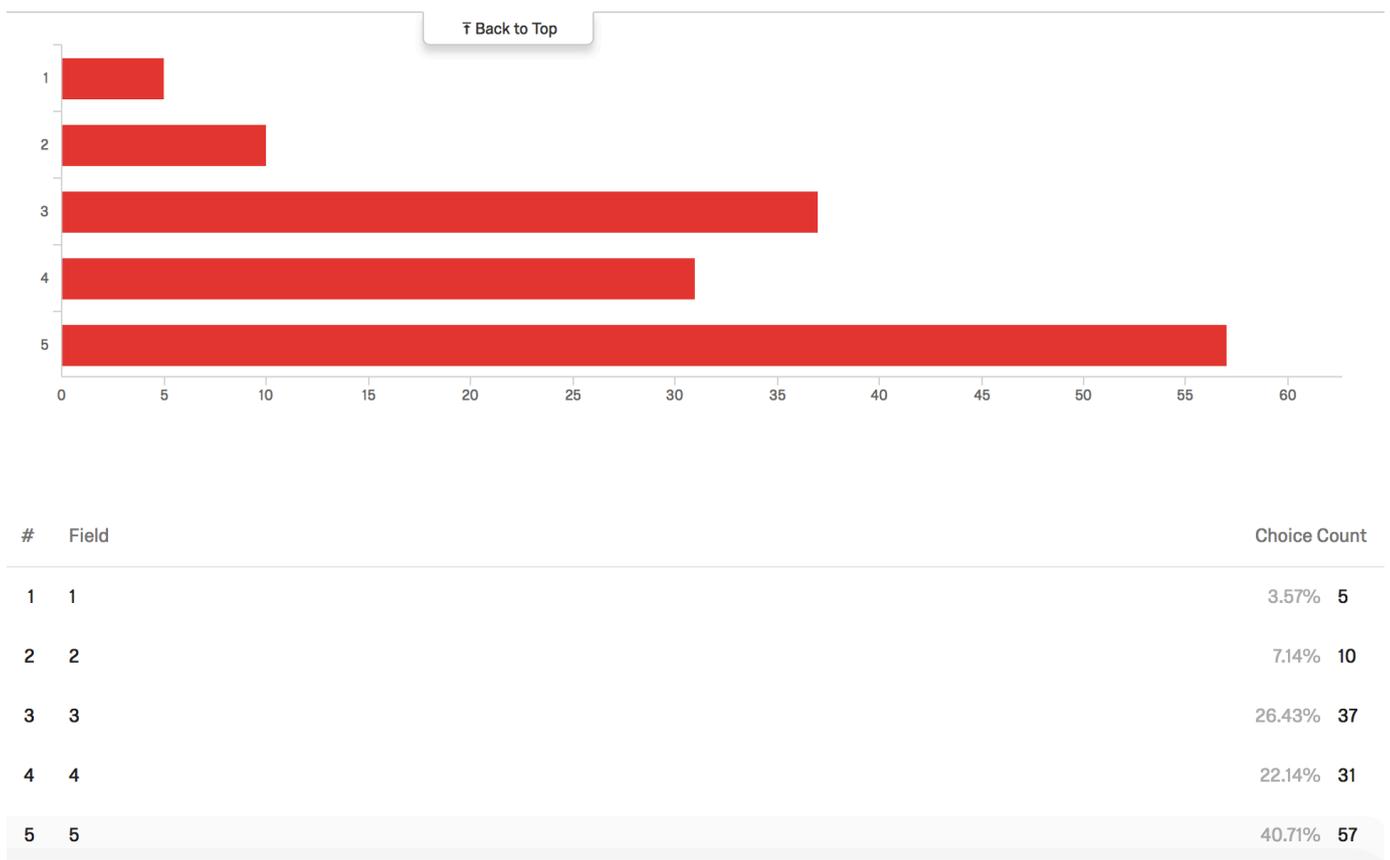
(Immagine 33)



∨ ^

| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 6.43% 9 |
| 2 | 2 | 5.00% 7 |
| 3 | 3 | 22.86% 32 |
| 4 | 4 | 30.71% 43 |
| 5 | 5 | 35.00% 49 |

(Immagine 34)



4.2.5 Quinta sezione: proposta di un explicit factor model per siti di organizzazione di viaggi online

Le ultime due domande del questionario sono quelle che hanno ad oggetto un sistema di raccomandazione per hotel e mete di viaggio.

In particolare ai rispondenti viene chiesto di immaginare un sito di prenotazioni online, come ad esempio Booking, e le due domande sono: “Quanto ti piacerebbe ricevere dei consigli su che hotel prenotare in base alle tue recensioni passate su altre strutture (ad es. consiglia un hotel con un alto punteggio in pulizia poiché in una recensione passata hai criticato tale aspetto di un’altra struttura)?” e “Quanto ti piacerebbe ricevere dei consigli su che paese/città visitare nel tuo prossimo viaggio basati sulle tue recensioni passate su altre mete (ad es. consiglia una meta con un punteggio alto in divertimento poiché in una recensione passata hai criticato tale aspetto di un’altra meta)?”.

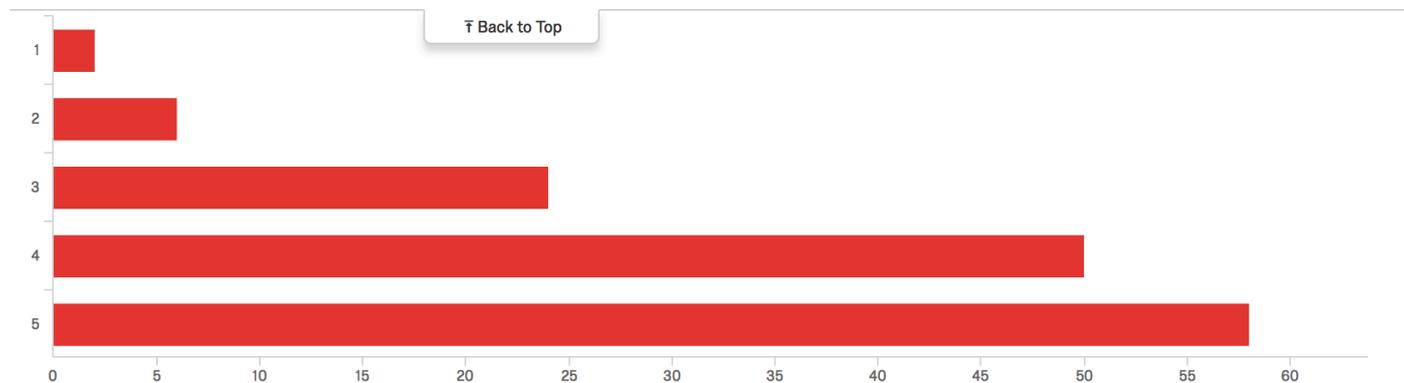
Anche questa sezione, come quella precedente, non è frutto di un’idea avuta dagli scrittori del *paper*, ma è un mio pensiero di possibile applicazione dell’*Explicit Factor Model*.

I risultati sono positivi, così come nelle due sottosezioni precedenti, in particolare le percentuali di voti ottenuti dai due valori maggiori, 4 e 5, sono: 77% e 65%, rispettivamente per la prima (Immagine 35) e seconda domanda (Immagine 36). In particolare, il valore di 77% eguaglia il valore massimo ottenuto in questo

questionario dalla domanda che ipotizza un sistema di raccomandazione che dia dei consigli sul perché non bisognerebbe acquistare un dato prodotto.

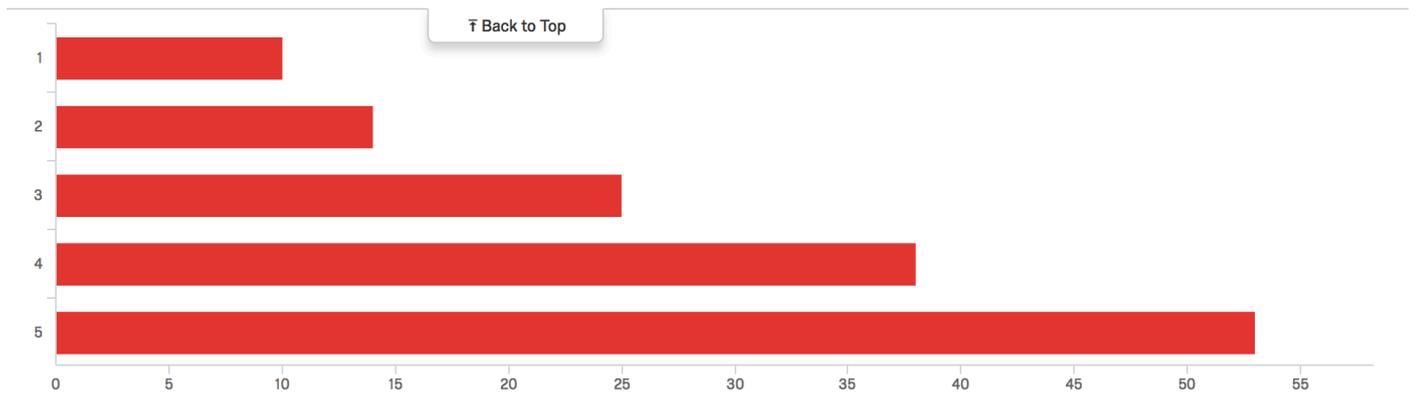
A seguire le immagini contenenti tutti i voti registrati nel dettaglio.

(Immagine 35)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 1.43% 2 |
| 2 | 2 | 4.29% 6 |
| 3 | 3 | 17.14% 24 |
| 4 | 4 | 35.71% 50 |
| 5 | 5 | 41.43% 58 |

(Immagine 36)



| # | Field | Choice Count |
|---|-------|--------------|
| 1 | 1 | 7.14% 10 |
| 2 | 2 | 10.00% 14 |
| 3 | 3 | 17.86% 25 |
| 4 | 4 | 27.14% 38 |
| 5 | 5 | 37.86% 53 |

Analizzando più nello specifico questa sottosezione, andando ad incrociare i dati ottenuti dalla domanda “Su quali prodotti/servizi hai scritto delle recensioni” e dal primo quesito di questa parte di sondaggio, si nota come tra i 47 voti ricevuti dalla risposta “hotel” ed i 16 ottenuti da quella “viaggi” solo un rispondente abbia dato un voto inferiore a tre, e, rispettivamente, ben l’85% ed il 94% ha dato un voto tra 4 e 5.

In particolare queste percentuali sono le più alte, e ciò ci mostra come chi ha scritto recensioni su viaggi e hotel gradisca particolarmente l’idea di poter essere aiutato nella scelta di un hotel che incontri i propri gusti, analizzandoli attraverso le recensioni da loro lasciate su altre strutture.

(Immagine 37)

| | | Quanto ti piacerebbe ricevere dei consigli su che hotel prenotare in base alle tue recensioni pas... | | | | | Total |
|---|---------------------|--|---|----|----|----|-------|
| | | 1 | 2 | 3 | 4 | 5 | |
| Su quali prodotti/servizi hai scritto delle recensioni (sono ammesse più risposte)? | Ristoranti/bar | 0 | 4 | 10 | 35 | 31 | 80 |
| | Hotel | 0 | 1 | 6 | 20 | 20 | 47 |
| | Viaggi | 0 | 0 | 1 | 9 | 6 | 16 |
| | Articoli e-commerce | 0 | 2 | 8 | 9 | 10 | 29 |
| | Altro | 0 | 1 | 4 | 1 | 5 | 11 |
| | Total | 0 | 5 | 16 | 37 | 42 | 100 |

Stesso discorso può essere fatto incrociando i dati sulla tipologia di recensioni lasciate dagli utenti e la seconda domanda di questa sezione. Anche qui i rispondenti che hanno dichiarato di aver effettuato recensioni su hotel

e viaggi hanno apprezzato maggiormente, rispetto agli altri, la possibilità di ricevere dei consigli su che meta visitare. In particolare 35 degli utenti che hanno effettuato recensioni sugli hotel hanno dato un valore tra 4 e 5 a questa idea, su 47 rispondenti totali, per una percentuale del 74%.

(Immagine 38)

| | | Quanto ti piacerebbe ricevere dei consigli su che paese/città visitare nel tuo prossimo viaggio b... | | | | | Total |
|---|---------------------|--|----|----|----|----|-------|
| | | 1 | 2 | 3 | 4 | 5 | |
| Su quali prodotti/servizi hai scritto delle recensioni (sono ammesse più risposte)? | Ristoranti/bar | 6 | 6 | 10 | 27 | 31 | 80 |
| | Hotel | 0 | 6 | 6 | 17 | 18 | 47 |
| | Viaggi | 1 | 3 | 1 | 6 | 5 | 16 |
| | Articoli e-commerce | 1 | 4 | 5 | 10 | 9 | 29 |
| | Altro | 0 | 3 | 3 | 1 | 4 | 11 |
| | Total | 6 | 10 | 17 | 30 | 37 | 100 |

4.3 Analisi dei risultati

Compito di questo questionario era andare a ricercare dei dati riguardo i comportamenti degli utenti online, in particolare nell'ambito della scrittura di recensioni sul web, l'applicabilità di un modello che vedesse l'unione tra analisi testuale e sistemi di raccomandazioni per migliorare le prestazioni di questi ultimi e la soddisfazione dei consumatori e l'apprezzamento, da parte degli utenti, di applicazioni di tale modello da me ipotizzate. Per raggiungere questi scopi, il questionario è stato diviso in sezioni, ognuna avente un punto centrale al quale cercare delle risposte.

La prima richiedeva soltanto dei dati di carattere generale come età e sesso. La terza domanda di questa prima sottosezione, però, fornisce già un'indicazione per il nostro studio, in particolare ci mostra come ben il 71% dei rispondenti abbia effettuato almeno una recensione online. Tale dato è già indicativo, poiché ci dice che la maggioranza degli utenti online ha scritto un suo parere su un prodotto o un servizio rendendo possibile l'applicabilità dell'*explicit factor model* mostrato nel capitolo tre, dato che si basa sulla costruzione di profili di aziende e utenti partendo proprio dalle recensioni scritte da questi ultimi.

La domanda successiva, facente parte della seconda sezione, ci mostra come la maggior parte delle recensioni venga fatta su ristoranti e bar, al secondo posto troviamo le recensioni sugli hotel, al terzo quello sugli articoli di e-commerce, al quarto quelle sui viaggi ed all'ultimo posto le recensioni su "altro".

Le successive tre domande, così come quella appena analizzata, sono state mostrate solo a coloro che hanno risposto "sì" alla domanda "Hai mai scritto una recensione su un qualunque sito web (e-commerce, hotel, viaggi, ristoranti...)?", e sono servite per valutare l'ammissibilità dell'*explicit factor model*, in particolare riguardo alcuni aspetti che nel *paper* di Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma erano dati per scontati.

Oggetto di questi quesiti era capire le abitudini dei consumatori riguardo la scrittura delle recensioni, nel dettaglio l'intento era individuare cosa spingesse gli user a recensire e su quali aspetti si concentrano maggiormente quando lo fanno. I dati raccolti vanno nella direzione di quanto ipotizzato dagli autori del *paper*, dato che tutte e tre le affermazioni contenute nei quesiti hanno ottenuto la maggioranza di risposte “d'accordo” e “molto d'accordo”, rispettivamente con l'84%, il 64% ed il 68% dei 100 voti totali.

Come risalta da questi risultati, gli utenti incentrano principalmente le loro recensioni sugli aspetti per loro più importanti, rendendo applicabile il modello esposto nel capitolo tre, che, come detto, punta a costruire profili di utenti basati sulle loro recensioni ed in particolare sugli aspetti dei prodotti e servizi citati in queste ultime. In altre parole, se un consumatore *x* nelle recensioni di diversi hotel si concentra sull'aspetto “pulizia”, è probabile che questo sia per lui molto importante. Quindi, se in futuro il sistema genera delle raccomandazioni per l'utente in questione basandosi sui punteggi ottenuti dalle altre strutture sul medesimo aspetto, andando a consigliargli quelle con i punteggi più elevati in “pulizia”, è probabile che la raccomandazione venga seguita e che il consumatore viva un'esperienza per lui soddisfacente, migliorando la *customer experience* e creando valore anche per l'azienda.

Analizzando nel dettaglio i dati ottenuti da queste tre domande, si nota come la terza domanda, “Quando scrivo una recensione mi concentro sugli aspetti del prodotto/servizio che si sono rivelati particolarmente negativi (1 fortemente in disaccordo – 5 fortemente d'accordo)”, ha ottenuto, rispetto alle altre due, un maggior numero di “fortemente d'accordo” rispetto a quello di “d'accordo”, e questo ci mostra come spesso sia proprio l'esperienza negativa a portare il consumatore a scrivere una recensione, probabilmente mosso dal desiderio di “vendetta” nei confronti dell'azienda o dalla voglia di mettere in guardia altri possibili consumatori da un'esperienza negativa. Anche questo dato avvalorava l'idea di come l'*explicit factor model* sia applicabile con successo, poiché si riuscirebbe a costruire un profilo di azienda con gli aspetti particolarmente positivi o negativi dei prodotti o servizi da poter consigliare o sconsigliare agli utenti nel momento della creazione di raccomandazioni.

Completata l'analisi dell'aspetto riguardante le basi dell'*explicit factor model*, ovvero i dati ottenuti dalle recensioni degli utenti con cui creare poi un sistema di raccomandazione, nel questionario si passa alla sezione successiva dove il focus sono le raccomandazioni spiegate e le così dette “dis-raccomandazioni”, ovvero i consigli sul perché un determinato prodotto non è adatto all'user che lo sta osservando. Entrambe queste tipologie di consigli non esistono in giro sul web e quindi rappresenterebbero un'interessante novità da poter inserire nei siti di e-commerce.

Proprio sulla base di quanto appena detto, nel questionario agli utenti viene chiesto di immaginare un sito di vendita di prodotti online, come Amazon.com ad esempio, e successivamente di rispondere a due domande con focus proprio sulle raccomandazioni spiegate e le dis-raccomandazioni. I due quesiti sono i seguenti: “Quanto ti piacerebbe ricevere delle raccomandazioni personalizzate con delle spiegazioni sul perché dovresti acquistare un determinato prodotto (ad es. “questo cellulare potrebbe interessarti poiché ha un'ottima batteria”

dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)?" e "Quanto ti piacerebbe ricevere dei consigli dal sito web sul perché NON acquistare un determinato prodotto (ad es. "questo cellulare non è adatto a te poiché ha una pessima batteria" dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto batteria)?".

Entrambe hanno ricevuto degli ottimi risultati, ottenendo, rispettivamente, il 66% ed il 77% di voti tra il 4 ed il 5, su una scala dove 1 sta per "per niente" e 5 sta per "moltissimo". Salta all'occhio il risultato della seconda domanda, che ha raggiunto addirittura il 60% di voti uguali a 5 ("moltissimo"), per un totale 84 voti su 140, risultato non raggiunto da nessuna altra domanda del questionario. Questo indica come la possibilità di ricevere delle dis-raccomandazioni sia una funzionalità che sarebbe particolarmente apprezzata dagli utenti, i quali sempre più utilizzano le recensioni per effettuare una scelta e sempre più acquistano su siti di e-commerce, così come mostra una ricerca⁵¹ riportata sul sito di vpnMentor, della quale riporto tre grafici qui sotto. Inoltre il forte consenso ricevuto dalla seconda domanda, ha un legame anche con il quesito precedente riguardo le abitudini degli utenti sulle recensioni online, dal quale era emerso un elevato numero di utenti che si concentrava sugli aspetti particolarmente negativi dell'esperienza. Il legame sta nel fatto che i consumatori sono particolarmente colpiti da un'esperienza negativa, tanto da sentire spesso il bisogno di condividerla con gli altri, e proprio per questo motivo, gradirebbero molto un sistema che glielne possa evitare. E, come detto in precedenza, un sistema di questo tipo migliorerebbe la fiducia che gli user hanno nei confronti dell'azienda, instaurando così con loro un rapporto soddisfacente e duraturo nel tempo.

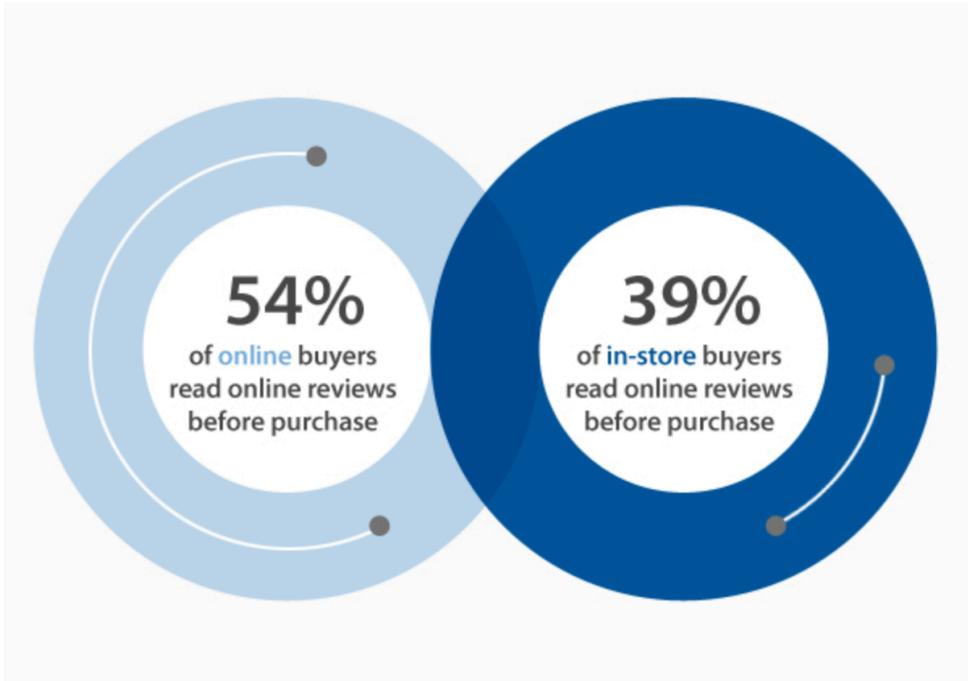
I dati raccolti da queste due domande, e quelli rappresentati nella ricerca, sono esplicativi di come l'idea portata avanti nel paper, ed in questa tesi, di un sistema di raccomandazione basato sull'analisi testuale delle recensioni online per creare delle raccomandazioni con spiegazioni incorporate e dei consigli su quali prodotti non acquistare e perché, su un sito di e-commerce, sia applicabile e riceverebbe un riscontro probabilmente molto positivo dai consumatori.

⁵¹ <https://it.vpnmentor.com/blog/vital-internet-trends/>

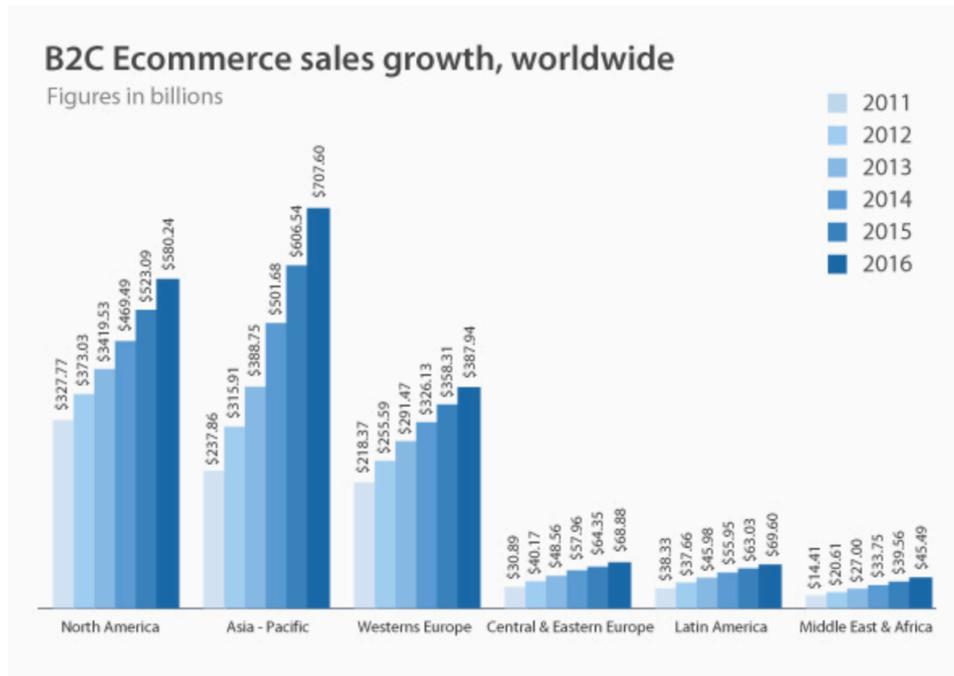
(Immagine 39)



(Immagine 40)



(Immagine 41)



Il primo grafico (Immagine 39) tra quelli estrapolati dalla ricerca mostrata sul sito vpnMentor sopra citata, mostra come la maggioranza degli utenti intervistati, il 55% per l'esattezza, trovi le recensioni molto utili per effettuare una scelta d'acquisto, e solo il 23% non le trova utili. Il secondo (Immagine 40) rafforza ancor di più l'importanza delle recensioni online, riportando come il 54%, la maggioranza, ed il 39% dei consumatori online ed offline, rispettivamente, leggano e si basino sui commenti degli altri utenti prima di effettuare un acquisto. Questi dati indicano il potenziale enorme di questo modello, unico e non ancora utilizzato da nessun *player* del mercato, che faciliterebbe di molto le scelte dei consumatori, analizzando al posto loro tutte le recensioni disponibili su di un prodotto o servizio, fornendo un vantaggio competitivo non indifferente all'impresa utilizzatrice.

Infine l'ultimo grafico (Immagine 41) è esplicativo della fortissima crescita dell'e-commerce negli ultimi anni, in particolare in Europa si è passati dai circa 242 miliardi di dollari di vendite online nel 2011, ai circa 457 miliardi di dollari del 2016, con una crescita quasi raddoppiata in appena 5 anni. Ciò è esplicativo di come quello dell'e-commerce sia un mondo in fortissima crescita, e quindi riuscire a trovare un modo per migliorare l'esperienza sia d'acquisto che post acquisto del consumatore potrebbe fare la differenza per diversi business operanti nel settore, riuscendo a garantirgli un vantaggio sui competitor.

Le tre domande facenti parte della quarta sezione, avevano lo scopo di testare l'applicabilità dell'*explicit factor model* nell'ambito della ristorazione. Il possibile utilizzo di tale modello in questo campo non è un'idea nata dagli autori del *paper*, ma è venuta a me personalmente. L'idea di fondo è quella di utilizzare la *sentiment analysis* per andare ad individuare, all'interno delle recensioni, ciò che più interessa agli utenti nel mondo dei ristoranti e dei bar, in che cibo, bevanda o servizio eccelle una determinata attività e cosa invece un cliente

dovrebbe evitare nella sua esperienza in un dato locale. Per fare un esempio pratico, se l'utente x effettua diverse recensioni citando l'aspetto "birra", il sistema dedurrà che lui è particolarmente interessato a tale bevanda e, quindi, gli fornirà una lista di consigli dei migliori bar o ristoranti che servono un'ottima birra. Ovviamente, per selezionare tali attività, dovrà essere effettuata un'analisi testuale delle recensioni scritte sui diversi ristoranti e bar per poter individuare quali abbiano le recensioni migliori sull'aspetto "birra". Allo stesso modo, sarà possibile creare una lista di cibi, bevande o servizi che sarebbe meglio evitare in una certa attività e ciò potrebbe essere esplicitato all'utente che si informa su quest'ultima per provare ad evitargli un'esperienza negativa.

Quindi, in altre parole, l'idea è quella di creare dei profili per gli utenti e per le diverse attività attraverso la *sentiment analysis* utilizzata sulle recensioni degli utenti, in modo tale da poter fornire delle raccomandazioni o dis-raccomandazioni giustificate.

In particolare le tre domande erano le seguenti: "Quanto ti piacerebbe ricevere dei consigli su quali ristoranti/bar visitare in base alle tue recensioni passate su altre attività (ad es. "questo ristorante potrebbe piacerti perché ha ottima carne" dato che nelle tue recensioni passate hai mostrato attenzione verso l'aspetto carne)?", "Quanto ti piacerebbe ricevere dei consigli su cosa ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in questo ristorante fanno un'ottima pizza")?" e "Quanto ti piacerebbe ricevere dei consigli su cosa NON ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in questo ristorante la pizza non ha un'ottima valutazione")?".

Andando ad analizzare i risultati ottenuti, si nota che anche in questo caso sono positivi e che l'idea è stata apprezzata. In particolare le percentuali degli utenti che hanno votato il valore 4 ed il 5 sono, rispettivamente per le tre domande, le seguenti: 68%, 65% e 62%.

Come si nota queste percentuali sono inferiori a quelle riscontrate nelle due domande precedenti che ipotizzavano un utilizzo del modello nei siti di e-commerce e, come si vedrà in seguito, sono inferiori, fatta eccezione per la percentuale della prima domanda, anche a quelle delle due domande successive.

Anche solo questo dato potrebbe bastare per capire che questa proposta sia meno allettante agli occhi degli utenti italiani rispetto alla precedente, ma un altro dato ci conferma questa affermazione, la percentuale di voti registrati per il valore 3, quello intermedio. In particolare le percentuali ottenute da questo valore sono rispettivamente del 24%, 23% e 26%, per le tre domande in questione, valori abbastanza maggiori rispetto a quelli del 15% ed 11% ricevuti dai due quesiti precedenti ed a quelli del 17% ottenuti dalle ultime due domande del questionario. Il valore 3, essendo quello intermedio, esprime una certa indifferenza del rispondente alla domanda posta, ed il fatto che le percentuali di questa risposta siano così elevate, rispetto a quelle delle altre domande, ci mostra come probabilmente questo sia un servizio che attira meno gli utenti rispetto agli altri proposti. Una possibile spiegazione a ciò potrebbe essere che il gusto è un qualcosa di prettamente personale, e quindi, secondo i rispondenti, delle raccomandazioni in tal senso potrebbero essere poco utili ed addirittura indurre in errore.

Come per la sezione precedente, anche in questa, le domande poste ai rispondenti avevano lo scopo di verificare il grado di apprezzamento di un'idea mia e non di una nata dal *paper* ed i suoi autori.

L'obiettivo era quello di testare il grado di apprezzamento degli utenti nei confronti di un ipotetico sistema di raccomandazione che utilizzi le recensioni testuali per fornire delle raccomandazioni personalizzate e spiegate riguardo possibili hotel e mete dove poter trascorrere una vacanza.

Nello specifico le due domande in questione erano le seguenti: “Quanto ti piacerebbe ricevere dei consigli su che hotel prenotare in base alle tue recensioni passate su altre strutture (ad es. consiglia un hotel con un alto punteggio in pulizia poiché in una recensione passata hai criticato tale aspetto di un'altra struttura)?” e “Quanto ti piacerebbe ricevere dei consigli su che paese/città visitare nel tuo prossimo viaggio basati sulle tue recensioni passate su altre mete (ad es. consiglia una meta con un punteggio alto in divertimento poiché in una recensione passata hai criticato tale aspetto di un'altra meta)?”.

Il fulcro di tale idea è sempre il medesimo, costruire dei profili di utenti, strutture turistiche e mete di viaggio per fornire delle raccomandazioni personalizzate e che soddisfino a pieno il consumatore. Un esempio pratico per la prima domanda potrebbe essere il seguente, un utente x che in soggiorno passato ha avuto una brutta esperienza con la posizione del suo hotel, tanto da criticarla in una recensione, e che si vede mostrare, come suggerimento dal sito, una struttura che invece ha un elevato punteggio su tale aspetto.

Stesso funzionamento di fondo avrebbero le raccomandazioni sulle mete di viaggio. Sono numerosissimi i blog dove viaggiatori o aspiranti tali possono scrivere e leggere recensioni sui luoghi turistici, e sfruttare questi dati testuali potrebbe essere una svolta per i siti di recensioni come TripAdvisor o Booking, visti anche i dati Istat, già citati in precedenza, che mostrano come nel 2016 gli italiani abbiano effettuato 66,5 milioni di viaggi con pernottamento, e di questi spostamenti di lavoro o piacere, in Italia o verso l'estero, circa 13 milioni (il 19% del totale) sono stati prenotati direttamente via Internet. Se a questi dati molto confortanti per i siti di prenotazione di viaggi online, si aggiungono quelli derivanti da un'indagine fatta da PhoCusWright, che mostra come i consumatori usino in media 6,5 siti e dispositivi per effettuare le ricerche preliminari alla prenotazione, si può notare come, se pur sia un settore in crescita, per tali siti sia fondamentale riuscire a convincere il potenziale consumatore a fermare la sua ricerca e prenotare presso loro.

Un modo per riuscire in questo scopo potrebbe essere appunto quello di fornire delle raccomandazioni spiegate su possibili mete ed hotel da poter visitare in base alle recensioni da loro scritte in passato, ad esempio il sito potrebbe consigliare all'utente una meta che ha avuto ottime recensioni sul cibo, aspetto da lui in precedenza criticato riguardo un'altra località visitata.

Anche i risultati di questa quinta ed ultima sottosezione hanno mostrato come *l'explicit factor model* possa essere utilizzato e soprattutto che sarebbe apprezzato dagli utenti. Andando ad analizzare i dati nel dettaglio, si vede come la prima domanda risulti essere quella che ha ricevuto una percentuale maggiore, rispetto a tutte le altre domande del questionario, di votazioni per i due valori maggiori, il 4 ed il 5. In particolare tale percentuale è del 77% rispetto al totale dei voti, numero eguagliato dalla domanda della terza sottosezione

dove veniva chiesto agli utenti di esprimere il grado di apprezzamento per un sito web che mostrasse loro il perché non avrebbero dovuto completare un determinato acquisto.

Questo dato ci porta ad una delle considerazioni finali, ovvero il fatto che un sistema di raccomandazione che dia delle dis-raccomandazioni sugli acquisti sui siti di e-commerce, ed uno che invece ne dia altri su che hotel scegliere, sarebbero particolarmente apprezzati dagli utenti e la loro introduzione porterebbe un vantaggio sia per questi ultimi che per le aziende.

Per quanto riguarda la seconda domanda, anche lei ha ottenuto un risultato positivo, nello specifico il 65% dei voti è stato distribuito tra i due valori maggiori, il 4 ed il 5.

Andando poi ad incrociare tra loro i dati relativi alla tipologia di recensioni lasciate in passato, e quelli derivanti da queste ultime due domande, appare palese che coloro abbiamo effettuato delle recensioni su viaggi e/o hotel siano maggiormente d'accordo con l'idea di inserire questo nuovo modello di raccomandazioni all'interno di siti del settore turistico.

Osservando la prima delle due tabelle, mostrate nel paragrafo precedente, quella che incrocia i dati ottenuti dalla domanda "Su quali prodotti/servizi hai scritto delle recensioni" e dal primo quesito di questa sezione, subito si nota come solo un rispondente sui 47 voti ricevuti dalla risposta "hotel" ed i 16 ottenuti da quella "viaggi" abbia dato un voto inferiore a tre, e, rispettivamente, ben l'85% ed il 94% ha dato un voto tra 4 e 5. Considerando che questi due voti sono quelli considerabili positivi, ed il valore 3 è considerabile come indifferenza all'argomento mostrato, solo l'1,6% ha esplicitato un parere negativo nei confronti di questa tipologia di raccomandazioni per il mondo degli hotel.

Inoltre, se tali dati si confrontano con le percentuali di voti pari a 4 e 5 ricevuti da questa penultima domanda del questionario dagli utenti che non hanno effettuato recensioni su hotel e/o viaggi, si vede che sono tutte più basse rispetto all'85% ed il 94%, suggerendo come, chi ha effettuato delle recensioni su tali tipologie di servizi sia maggiormente propenso a lasciarsi aiutare nella scelta di un hotel che possa incontrare i suoi gusti ed evitarli delle cattive sorprese.

Infine, andando ad analizzare la seconda tabella prima menzionata, quella che intreccia i dati sulla tipologia di recensioni effettuate e quelli relativo all'ultima domanda del questionario, si hanno gli stessi risultati osservati con la tabella precedente. In particolare, anche in questo caso, i rispondenti che hanno dichiarato di aver effettuato recensioni su hotel e viaggi hanno apprezzato maggiormente, rispetto agli altri, la possibilità di ricevere dei consigli su che meta visitare nei loro prossimi viaggi. Nello specifico 35 degli utenti che hanno effettuato recensioni sugli hotel hanno dato un valore tra il 4 ed il 5 a questa idea, per una percentuale del 74% rispetto ai 47 rispondenti totali.

Una delle spiegazioni a tali dati di fatto potrebbe essere che chi non ha effettuato delle recensioni su hotel e/o viaggi, sia meno abituato a viaggiare o meno propenso ad organizzare viaggi da solo, rivolgendosi probabilmente ad agenzie di viaggio, e quindi possa fare maggior fatica a capire i vantaggi che si avrebbero da delle raccomandazioni di questo tipo.

Per concludere, tale questionario è stato utile per analizzare i comportamenti degli utenti quando si tratta di effettuare delle recensioni e per testare l'applicabilità e il livello di consenso che potrebbe riscuotere l'*explicit factor model* in Italia. I risultati ci hanno mostrato come gli ambiti in cui tale tipologia di sistema di raccomandazione sarebbe probabilmente maggiormente apprezzato siano quelli dell'e-commerce e del turismo. In particolare però, le funzionalità che hanno avuto il maggior riscontro positivo sono quelle relative alle dis-raccomandazioni per i prodotti di e-commerce e alle raccomandazioni riguardo l'hotel da dover prenotare.

Per quanto riguarda la prima, probabilmente sarebbe particolarmente apprezzata perché aiuterebbe nel non fare acquisti azzardati che potrebbero rivelarsi non adatti ai gusti ed alle funzionalità ricercate dagli utenti, facendogli risparmiare tempo e denaro, con risvolti in termini di fiducia, fedeltà e *brand equity*. La seconda funzionalità, invece, consentirebbe di ridurre il tempo per la ricerca di un hotel che sia adatto ai gusti ed alle esigenze degli user, evitandogli delle esperienze negative in strutture con lacune sugli aspetti più importanti per loro. Dal lato dell'azienda, questa possibile funzione, probabilmente permetterebbe non solo un aumento in termini di fiducia, fedeltà e *brand equity*, ma anche in termini di prezzo che gli utenti sarebbero disposti a pagare, perché si tratterebbe di hotel che incontrano perfettamente i loro gusti e che non dovrebbe nascondere sgradite sorprese.

La quarta sottosezione, invece, è quella che ha ottenuto i riscontri meno positivi, ed in particolare la domanda "Quanto ti piacerebbe ricevere dei consigli su cosa NON ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in questo ristorante la pizza non ha un'ottima valutazione")?", è quella che ha ottenuto, tra tutte quelle del questionario, la percentuale minore di voti pari a 4 e 5, per un totale del 62%.

Una spiegazione a tale fenomeno potrebbe essere il fatto che il mondo del cibo sia caratterizzato da una forte soggettività, e ciò che piace ad un utente potrebbe non piacere ad un altro, e quindi probabilmente potrebbero non essere delle raccomandazioni o dis-raccomandazioni corrispondenti alla realtà dei fatti.

Tale soggettività però non si ha con ogni aspetto di un ristorante o bar, visto che aspetti importanti sono anche la location, l'atmosfera e la cordialità dei camerieri, per i quali un sistema di questo tipo potrebbe essere applicato e forse accolto con maggior entusiasmo anche dai potenziali consumatori.

Infine, quindi, si può affermare che l'*explicit factor model*, e le sue raccomandazioni e dis-raccomandazioni basate sull'analisi testuale delle recensioni utile per creare profili di utenti, con i relativi aspetti più importanti, ed aziende, con quelli maggiormente positivi e maggiormente negativi, sia applicabile, con buoni risultati previsti, nel mondo dell'e-commerce ed in quello degli hotel e dei viaggi in generale. Andrebbe invece ancora affinata la sua utilizzabilità nel mondo della ristorazione.

Conclusioni

Obiettivo di questa tesi era analizzare, testare e ipotizzare e testare, tramite analisi empirica, nuove possibili applicazioni, da me pensate, per l' "Explicit Factor Model", un modello ideato e descritto da Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma all'interno del *paper* "Explicit Factor Models for explainable recommendation based on phrase-level sentiment analysis", che unisce insieme sistemi di raccomandazione ed analisi testuale in un approccio innovativo, non ancora utilizzato da nessun sito web, che ha l'obiettivo di migliorare la qualità dei consigli mostrati agli utenti, dare delle spiegazioni sul perché di tali consigli e fornire delle dis-raccomandazioni riguardo entità da non acquistare poiché non adatte al consumatore che le sta valutando.

L'idea di un sistema di raccomandazione unito con la *sentiment analysis* è cresciuta in me dopo aver cercato informazioni su questi due strumenti, che ritengo essenziali in un mondo dominato dai dati e dalle informazioni in essi contenute, e dopo aver pensato alla centralità delle recensioni online nelle scelte di acquisto della quasi totalità dei consumatori.

È un'idea rivoluzionaria, perché se è vero che la maggior parte delle aziende operanti online, come Amazon.com e Netflix ad esempio, hanno alla base del loro business i sistemi di raccomandazione, e se è vero che l'analisi testuale è utilizzata già per diversi motivi dalle imprese, è anche vero che ancora nessuna di queste utilizza un modello come questo. E proprio questa sua assenza dal mercato attuale potrebbe fornire alle aziende un vantaggio importante, sia in termini di miglior servizio offerto sia di miglior *brand image* percepita dai consumatori.

Come riscontrato dagli esperimenti da loro effettuati, e riportati all'interno del *paper*, l'*Explicit Factor Model* riesce ad ottenere dei risultati migliori in termini di accuratezza predittiva rispetto ad altri algoritmi di raccomandazione e fornisce agli utenti un qualcosa che fino ad oggi non si è ancora visto, una spiegazione ai consigli presentati e delle dis-raccomandazioni, ovvero dei consigli sul perché un determinato oggetto non andrebbe comprato.

Appurata la migliore accuratezza predittiva del modello rispetto ad altri sistemi di raccomandazione, un'altra analisi altrettanto importante è stata compiuta. I ricercatori hanno effettuato degli esperimenti online per verificare in che modo le spiegazioni delle raccomandazioni e le dis-raccomandazioni impattassero sui comportamenti dei consumatori. I risultati hanno mostrato un *click through rate* (CTR) di circa 1,1% maggiore ottenuto dai consigli contenenti delle spiegazioni sul perché un dato oggetto fosse stato appunto raccomandato rispetto alle raccomandazioni senza un chiarimento, un dato particolarmente positivo. Per quanto riguarda le dis-raccomandazioni, anche in questo caso i risultati sono stati positivi con una forte capacità persuasiva riscontrata nei consigli sul perché non acquistare un dato oggetto.

Accertati e riportati i vantaggi di tale modello per consumatori ed imprese, in particolare il sistema diventa più trasparente e ha il potenziale per consentire agli utenti di dire quando il sistema sbaglia, aumenta la confidenza e la fiducia degli utenti nel sistema, aiuta gli utenti a prendere decisioni migliori (maggior efficacia)

e più rapide (maggior efficienza), convince gli utenti a provare o acquistare (maggior persuasività) e aumenta la soddisfazione di questi ultimi, il mio apporto è stato quello di creare un questionario per verificare l'applicabilità di tale modello in Italia, dato che i ricercatori avevano effettuato la loro analisi su due database, uno americano ed uno cinese, e testare altre possibili applicazioni in campi diversi dall'e-commerce, sul quale era incentrata l'idea di Yongfeng Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma.

La mia idea era quella di applicare il modello delle raccomandazioni spiegate e delle dis-raccomandazioni ai siti di recensioni online di attività ristorative come ristoranti e bar ed ai siti per il recupero di informazioni e prenotazione dei viaggi.

Il questionario l'ho strutturato dividendolo in 5 parti per ottenere delle informazioni chiare per ogni argomento di studio.

A seguito della prima sezione, che chiedeva informazioni generali agli utenti, ce ne era una seconda dalla quale ho ottenuto le prime risposte positive. Obiettivo era quello di valutare l'applicabilità del modello analizzando il comportamento degli utenti nell'effettuare le recensioni. È emerso come questi ultimi si concentrino principalmente sugli aspetti più importanti per loro, e ciò evidenzia l'applicabilità del modello che mira appunto ad identificare gli aspetti degli oggetti/servizi centrali per i consumatori per costruire un loro profilo e fornire delle raccomandazioni di conseguenza.

Se questa seconda sezione analizzava l'applicabilità del modello, le successive tre avevano come focus la proposta di raccomandazioni spiegate e dis-raccomandazioni, rispettivamente, nei settori dell'e-commerce, della ristorazione e dei viaggi.

In particolare i risultati ci hanno mostrato come gli ambiti in cui tale tipologia di sistema di raccomandazione sarebbe probabilmente maggiormente apprezzato siano quelli dell'e-commerce e del turismo. Analizzando nel dettaglio i dati, però, le funzionalità che hanno avuto il maggior riscontro positivo sono quelle relative alle dis-raccomandazioni per i prodotti di e-commerce e alle raccomandazioni riguardo l'hotel da dover prenotare. Probabilmente sarebbero particolarmente apprezzate perché da un lato, la prima funzionalità, aiuterebbe l'utente a non fare acquisti azzardati che potrebbero rivelarsi non adatti ai gusti ed alle funzionalità ricercate, facendogli risparmiare tempo e denaro, la seconda, consentirebbe di ridurre il tempo per la ricerca di un hotel che sia adatto ai gusti ed alle esigenze degli user, evitandogli delle esperienze negative in strutture con lacune sugli aspetti più importanti per loro.

Dal lato dell'azienda, queste possibili funzioni, probabilmente permetterebbe non solo un aumento in termini di fiducia, fedeltà e *brand equity*, ma anche in termini di prezzo che gli utenti sarebbero disposti a pagare, nel caso dei siti di prenotazione di viaggi, perché si tratterebbe di hotel che incontrano perfettamente i loro gusti e che non dovrebbe nascondere sgradevoli sorprese.

Se queste due possibili funzioni hanno ottenuto i maggiori consensi tra tutte, la quarta sezione, invece, quella avente ad oggetto il nuovo modello applicato all'interno del mondo della ristorazione, è quella che ha ottenuto i riscontri meno positivi, ed in particolare la domanda "Quanto ti piacerebbe ricevere dei consigli su cosa NON ordinare in un ristorante/bar in base alle recensioni degli altri utenti su questo particolare ristorante/bar ("in

questo ristorante la pizza non ha un'ottima valutazione"?)", è quella che ha ottenuto la percentuale minore di voti pari a 4 e 5, per un totale del 62%.

Una spiegazione a tale fenomeno potrebbe essere il fatto che i gusti siano un qualcosa di particolarmente soggettivo, e ciò che piace ad un utente potrebbe non piacere ad un altro, e quindi probabilmente potrebbero non essere delle raccomandazioni o dis-raccomandazioni di qualità.

Tale soggettività però non si ha con ogni aspetto di un ristorante o bar, visto che aspetti importanti sono anche la location, l'atmosfera o la cordialità dei camerieri ad esempio, e per questa tipologia di aspetti un sistema di questo tipo potrebbe essere applicato e forse accolto con maggior entusiasmo anche dai potenziali consumatori.

Una volta analizzato il modello e le sue potenzialità, ora è giusto valutarne i limiti. Quello principale non è collegato con la struttura e i processi dell'*Explicit Factor Model*, ma bensì è collegato a ciò che utilizza quest'ultimo, ovvero le recensioni. Purtroppo il fenomeno delle *fake reviews* è quanto mai attuale, e sta crescendo sempre più un vero e proprio business basato sulle recensioni fasulle per oggetti e servizi offerti.

Secondo uno studio effettuato in Francia⁵², circa una recensione su tre risulterebbe essere falsa, questo ci mostra quanto effettivamente sia grande il problema.

E, data la centralità delle recensioni per le scelte dei consumatori e la loro capacità di spostare le sorti di un business, negli anni i diversi siti che permettono di commentare e valutare attività o prodotti stanno cercando di correggere alcuni problemi pur provando a non togliere la possibilità di scriverne una a nessuno. Ad esempio da diversi anni Amazon pubblica una classifica dei recensori, così come fa anche TripAdvisor che aggiunge anche un ulteriore strumento ai proprietari delle attività recensite, ovvero la possibilità di segnalare una determinata recensione ritenuta fasulla, per richiedere un controllo accurato a TripAdvisor stesso.

Nonostante questo grosso problema riguardo le recensioni online, queste ultime restano comunque al centro dei processi di acquisto dei consumatori, qualunque sia il prodotto o servizio che si sta andando a comprare, e quindi, perché non integrarle nei sistemi di raccomandazione per migliorare il servizio offerto e rendere la vita dei consumatori più semplice?

D'altronde si sa, "il cliente ha sempre ragione".

⁵² <https://www.economie.gouv.fr/dgccrf/faux-avis-consommateurs-sur-plateformes-numeriques>

Bibliografia & Sitografia

A. Ansari, S. Essegaier, R. Kohli. Internet recommendation systems. JMR, 2000.

A. Jain, V. Jain, N. Kapoor. A literature survey on recommendation system based on a sentimental analysis. ACII, 2013.

A. K. A. Hassan, A. B. Aldeen Abdulwahhab. Reviews Sentiment analysis for collaborative recommender system. 2017.

A. S. Nair, K. Sreelakshmi. Movie recommendation system using sentiment analysis. 2017.

A. Sharma, J. M. Hoffman, D. J. Watts. Estimating the casual impact of recommendation systems from observational data. EC, 2015.

B. Liu, P. S. Yu, X. Ding. A holistic lexicon-based approach to opinion mining. WSDM, 2008.

B. Liu. Sentiment Analysis and opinion mining. Morgan & Claypool Publishers, 2012.

B. P. Knijnenburg, N. J. M. Reijmer, M. C. Willemsen. Each to his own: how different users call for different interaction methods in recommender systems. ACM, 2011.

B. Pang, L. Lee. Opinion mining and entiment analysis. NOW, 2008.

B. Smith, G. Linden. Two decades of recommender systems ad Amazon.com. IEEE, 2017.

- B. Xiao, I. Benbasat. E-commerce product recommendation agents: use, characteristics, and impact. 2007
- Burke. Hybrid recommender systems: survey and experiments. 2002.
- C. A. Gomez-Uribe, N. Hunt. The Netflix Recommender system: algorithms, business value and innovation. ACM, 2015.
- Campochiaro et al. Do Metrics Make Recommender Algorithms? AINA, 2009.
- D. R. Dowty, R. E. Wall, S. Peters. Introduction to montague Semantics. 1981.
- D.Billsus,M. J. Pazzani. User modeling for adaptive news access. 2000
- Digital in 2018 in souther Europe. <https://www.slideshare.net/wearesocial/digital-in-2018-in-southern-europe-part-1-west-86864268>
- F. Benamara, B. Chardon, Y. Mathieu, V. Popescu. Towards context-based subjectivity analysis. AFNLP, 2011.
- F. Ricci, L. Rokach, B. Shapira, P. B. Kantor. Recommender systems handbook. LLC, 2011.
- F. Sfodera. Turismi, destinazioni e internet. La rilevazione della consumer experience nei portali turistici. 2011.

G. Adomavicius, A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE, 2005.

Gli andamenti di Internet, le statistiche & i dati certi negli U.S.A. e nel Mondo nel 2018
<https://it.vpnmentor.com/blog/vital-internet-trends/>

González-Ibáñez, Muresan e Wacholder. Identifying sarcasm in twitter: a closer look. 2011.

H. Kanayama, T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. EMNLP, 2006.

H. W. Kim, K. Han, M. Y. Yi, J. Cho, J. Hong. MovieMine: personalized movie content search by utilizing user comments. IEEE, 2012.

<http://dizionari.repubblica.it/Italiano/M/ma.php>

http://www.repubblica.it/economia/rapporti/paesedigitale/trend/2017/04/28/news/turismo_dalle_prenotazioni_online_un_indotto_da_oltre_2_miliardi-164087252/

J. B. Schafer, J. A. Konstan, J. Riedl. E-commerce recommendation applications. 2001.

J. Grau. Personalized product recommendations: predict shoppers' needs. eMarketer, 2009.

J. M. Wiebe, R. F. Bruce, T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. ACL, 1999.

J. Vig, S. Sen, J. Riedel. Tagsplanations: explaining recommendations using tags. IUI, 2009.

J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen. Collaborative filtering recommender systems. 2007.

K. Bauman, B. Liu, A. Tuzhilin. Aspect based recommendations: recommending items with the most valuable aspect based on user reviews. KDD, 2017.

K. Ravi, V. Ravi. A survey on opinion mining and sentiment analysis: tasks, approaches and application. 2015.

L. Chen, G. Chen, F. Wang. Recommender systems based on user reviews: the state of the art. 2015.

Latent factor models for collaborative filtering. <https://www.slideshare.net/sscdotopen/latent-factor-models-for-collaborative-filtering>

Les faux avis de consommateurs sur les plateformes numériques. <https://www.economie.gouv.fr/dgccrf/faux-avis-consommateurs-sur-plateformes-numeriques>

M. J. Pazzani. A framework for collaborative, content-based and demographic filtering. 1999.

M. M. Recker, A. Walker, K. Lawless. What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. K. Instructional Science, 2003.

Matrix factorization techniques for recommender systems. <https://www.slideshare.net/studentalei/matrix-factorization-techniques-for-recommender-systems>

Nadkarni, N. Yezhkova. Structured versus unstructured Data. IDC Corporate, 2014.

O. Tsur, D. Davidov, A. Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. 2010.

O. Tsur, D. Davidov, A. Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. 2010.

P. Pasini, A. Perego. Big Data: nuove fonti di conoscenza aziendale e nuovi modelli di management. SDA Bocconi school management, 2012.

R. Burke. Hybrid Web Recommender Systems. 2007.

R. Garfinkel, R. Gopal, B. Pathak, R. Venkatesan, F. Yin. Empirical analysis of the business value of recommender systems. 2006.

R. Zhou, S. Khemmart, L. Gao. The impact of YouTube recommendation system on video views. IMC, 2010.

S. Kim, E. Hovy. Extracting opinions, opinion holders, and topics expressed in online news media text. 2006.

S. Y. X. Komiak, I. Benbasat. The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. 2006.

SAS Sentiment Analysis. Align with your customers' desires.
http://www.sas.com/en_us/software/analytics/sentiment-analysis.html

Sentiment analysis: cos'è e perché serve alle aziende. <https://www.insidemarketing.it/glossario-marketing-comunicazione/sentiment-analysis/>

T. Mahmood, F. Ricci. Improving recommender systems with adaptive conversational strategies. ACM, 2009.

The Four V's of Big Data. <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

Turismo: il 65% delle prenotazioni di viaggi e vacanze avviene online.
<https://www.insidemarketing.it/turismo-il-65-delle-prenotazioni-di-viaggi-e-vacanze-avviene-online/>

Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. ACL, 2002.

U. Gretzel, K. H. Yoo. Use and impact of online travel reviews. 2008.

V. Hazivassiloglou, K. R. McKeown. Predicting the semantic orientation of adjectives. ACL, 1997.

Y. Koren, R. Bell, C. Volinsky. Matrix factorization techniques for recommender systems. IEEE, 2009.

Y. Lu, M. Castellanos, U. Dayal, C. Zhai. Automatic Construction of a Context-Aware Sentiment Lexicon: an optimization approach. WWW, 2011.

Y. Lu, M. Castellanos, U. Dayal, C. Zhai. Automatic construction of a context-aware sentiment lexicon: an optimization approach. WWW, 2011.

Y. Tan, Y. Zhang, M. Zhang, Y. Liu, S. Ma. A unified framework for emotional elements extraction based on finite state matching machine. CCIS, 2013.

Y. Zhang, G. Lai, M. Zhang, Y. Liu e S. Ma. Explicit Factor Models for explainable recommendation based on phrase-level sentiment analysis. SIGIR, 2014.

Y. Zhang, R. Liu, A. Li. A novel approach to recommendation system based on aspect-level sentiment analysis. NCEECE, 2015.

Y. Zhang. Incorporating Phrase-level Sentiment Analysis on textual reviews foe personalized recommendation. WSDM, 2015.

Z. Zhai, B. Kang, H. Xu, P. Jia. Exploiting effective features for chinese sentiment classification. 2011.

Z. Zhai, B. Liu, H. Xu, P. Jia. Grouping product features using semi-supervised learning with soft-constraints. COLING, 2010.

Riassunto

Nel mondo moderno tutto gira intorno ai dati ed alle informazioni in essi contenute estraibili tramite l'analisi. Stando all'analisi effettuata da WeAreSocial ed Hootsuite, in Italia, al gennaio 2018, il 73% della popolazione, 43,31 milioni di persone su un totale di 59,33 milioni, utilizza internet, un dato cresciuto del 10% rispetto allo stesso mese dell'anno precedente, e si pensi che per ogni azione compiuta online, un utente lascia un *social footprint*, un'impronta digitale che viene registrata e non sarà mai cancellata.

Da qui nasce il concetto di *Big Data*, una grande mole di dati sia strutturati che non che necessitano però di essere analizzati per poter essere sfruttati al meglio.

Per quanto detto in precedenza, i *Big Data*, quindi, sono una diretta conseguenza delle nuove abitudini digitali che pervadono la vita quotidiana, sia a livello individuale che aziendale, riempiendola di accessi verso siti web, di "mi piace", di tweet, di commenti, di condivisioni, dati digitali che vengono salvati e immagazzinati per poter essere utili.

Per descrivere nel dettaglio i *Big Data* si può far riferimento al modello delle "4V" ideato da IBM⁵³:

- **Volume:** il volume di dati, cioè la possibilità di ottenere, immagazzinare e accedere a grandi volumi di dati, così elevati da non poter essere gestiti e controllati da database tradizionali.
- **Velocità:** si riferisce alla velocità con cui i dati sono a disposizione delle aziende e con cui vengono generati e trattati.
- **Varietà:** si riferisce alla varietà dei dati a disposizione. In passato l'analisi dei dati poteva essere effettuata solo su dati strutturati in tabelle o database relazionali, all'interno dei quali non si potevano inserire dati di formati diversi come foto, conversazioni, dati audio e video, documenti, ecc. Con il termine *Big Data*, invece, si fa riferimento a dati provenienti da fonti diverse e di tipologia diversa.
- **Veridicità:** perché i dati possano essere considerati di valore, c'è bisogno che sia dei dati veri, e questo potrebbe essere un problema in un mondo dove sono sempre più le informazioni fasulle sul web, basti pensare al problema delle *fake news* nel mondo dell'informazione e quello delle *fake reviews* in quello dei feedback dai clienti ad esempio.

Collegati ai *Big Data* ci sono alcuni limiti che ne rallentano l'espansione. Il primo è quello pocanzi descritto, ovvero la veridicità delle informazioni, anche se è la qualità in generale ad essere importante, poiché i *Big Data*, per poter essere considerati affidabili e quindi per poter rappresentare uno strumento decisionale e di analisi veramente utile, devono possedere alcune caratteristiche precise: devono possedere dati completi, devono essere il più aggiornati possibile, devono essere coerenti tra loro e tecnicamente conformi alle regole su cui si basano i *Big Data* e devono essere quanto più accurati possibile.

⁵³ <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

Altro problema, e limite, collegato ai *Big Data*, è quello della privacy. È proprio di questo recente periodo il polverone alzato dal caso Cambridge Analytica.

Senza capacità di analisi ed interpretazione, però, il potenziale informativo di un volume di dati si annulla, quindi è necessario sviluppare un'infrastruttura fisica e degli algoritmi appositi per poter utilizzare efficacemente i dati raccolti⁵⁴. Due strumenti che si sono rivelati molto utili per tale fine per le aziende, sono i sistemi di raccomandazione e la *sentiment analysis*.

Quest'ultima è l'area di ricerca che studia opinioni, sentimenti, valutazioni, attitudini ed emozioni delle persone riguardo entità ed i loro aspetti espresse attraverso un testo scritto.

Questa tecnica conferisce alle aziende la possibilità di ottenere informazioni generate direttamente dai consumatori, senza che siano influenzati in alcun modo dall'impresa, quindi informazioni utili anche verso altre funzioni aziendali oltre a quella di marketing.

In generale, la *sentiment analysis* può essere applicata su tre livelli differenti:

- livello di documento: in questo caso l'obiettivo è valutare se il sentimento di un intero documento è positivo, negativo o neutrale. Questo livello di analisi, però, presuppone che all'interno di un documento si discuta soltanto di una entità, e non può essere utilizzato nel caso in cui ce ne siano di più in un singolo testo;
- livello di frase: a questo livello il compito è quello di determinare l'opinione di ogni singola frase presente in un documento. Questo livello di analisi è strettamente correlato alla classificazione della soggettività⁵⁵, che distingue le frasi in oggettive, che esprimono informazioni fattuali, e soggettive, che esprimono delle opinioni soggettive;
- livello di entità e aspetto: l'obiettivo di questo livello di analisi è scoprire i sentimenti sulle entità e/o sui loro aspetti. Basandosi su questo livello di analisi, è possibile produrre un riassunto strutturato delle opinioni sulle entità e sui loro aspetti, che trasforma il testo non strutturato in dati strutturati che possono essere utilizzati per tutti i tipi di analisi qualitative e quantitative.

Come si può facilmente immaginare, gli indicatori principali del sentimento in una frase sono le parole d'opinione, e gli elenchi che comprendono tali parole sono chiamati *sentiment lexicon*, i quali sono necessari ma non sufficienti per effettuare una *sentiment analysis*.

Prima di poter creare un *sentiment lexicon*, però, sono necessarie alcune azioni preliminari.

Il linguaggio spesso possiede strutture e codici complessissimi ma, per fortuna, per completare un'analisi testuale efficace non è necessario riuscire ad interpretarlo e decodificarlo in tutta la sua complessità, infatti quando un testo scritto deve essere analizzato è necessario svolgere alcuni passaggi fondamentali di pre-processing allo scopo di ridurre il numero di termini presenti all'interno del testo stesso.

⁵⁴ Pasini P., Perego A. 2012

⁵⁵ Wiebe, Bruce, O'Hara 1999

Quindi, per prima cosa, nel processo di analisi di un testo, bisogna cercare di ridurre il testo in un dato quantitativo, tale da poter essere trattato da un modello statistico.

Come accennato in precedenza il fulcro della *sentiment analysis* sono le parole e le frasi che esprimono sentimenti positivi o negativi. Tali parole possono essere o di tipo base o comparative, le prime sono ad esempio “buono”, “brutto”, “bello” ecc., quelle del secondo tipo sono usate per esprimere opinioni comparative e superlative, esempi sono “migliore”, “peggiore” ecc. sono stati individuati tre modi possibili per creare un *sentiment lexicon*, insieme di parole e frasi di opinione, utilizzabile poi per la *sentiment analysis*, uno manuale, uno basato sul dizionario e uno basato sul corpus. Quello manuale è particolarmente complicato da applicare e viene quasi sempre utilizzato insieme con uno degli altri due approcci automatici che possono commettere degli errori.

Un approccio basato sul dizionario, semplice da utilizzare, prevede la raccolta manuale di alcune parole d’opinione con orientamenti positivi o negativi dal corpus, e successivamente l’algoritmo aumenta questo set cercando su un qualsiasi dizionario online i loro sinonimi e contrari.

L’approccio basato sul corpus parte da una lista di parole d’opinione, spesso generali, e ne scopre altre da un corpus. Una tecnica⁵⁶ potrebbe essere quella di sfruttare un insieme di regole linguistiche e convenzioni sulle congiunzioni per identificare le parole d’opinione nel corpus. Una delle regole riguarda la congiunzione “e”, la quale indica che gli aggettivi da lei congiunti solitamente hanno lo stesso orientamento, ad esempio, nella frase "Questa macchina è bella e spaziosa", se "bello" è positivo, si può dedurre che "spazioso" è anche positivo, questo perché le persone di solito esprimono lo stesso sentimento su entrambi i lati di una congiunzione. Tale ragionamento può essere utilizzato anche per altre congiunzioni come, o, ma, o-o, ed è chiamato coerenza sentimentale.

Va notato però che non basta trovare le parole d’opinione con i loro orientamenti, poiché ci sono degli aggettivi che hanno un significato differente anche all’interno dello stesso contesto, ad es, l’aggettivo silenzioso è positivo nella frase “questa macchina è silenziosa”, ma ha un’accezione negativa nella frase “lo stereo di questa auto è silenzioso”. Per ovviare a questo problema si può guardare alla coppia⁵⁷ (aspetto, parola d’opinione), quindi si vanno a considerare la parola d’opinione, il suo orientamento e anche l’aspetto a cui si riferiscono. Anche in questo caso però sono valide le regole relative alle congiunzioni all’interno della frase. Va comunque tenuto presente che compilare un *sentiment lexicon*, dipendente dal contesto o meno, è diverso dal determinare se una parola o una frase in realtà esprime un’opinione e quale sia il suo orientamento in una particolare frase. Solo perché una parola o una frase è elencata in un lessico di opinione non significa che in realtà stia esprimendo un’opinione in una frase, ad esempio, nella frase "Sto cercando una buona assicurazione sanitaria", "buono" non esprime un’opinione positiva o negativa su alcuna assicurazione in particolare.

L’altro strumento sopracitato, utile per l’estrazione di informazione dai dati, sono i sistemi di raccomandazione.

⁵⁶ Hazivassiloglou, McKeown 1997

⁵⁷ Liu, Yu 2008

Questi ultimi nascono con lo scopo di fornire agli utenti dei suggerimenti personalizzati riguardo un set di oggetti, materiali o immateriali, basando le raccomandazioni sui dati a loro disposizione riguardo gli user e gli oggetti stessi.

Di solito le raccomandazioni vengono mostrate come classifiche di oggetti, create utilizzando le preferenze degli utenti, che possono essere esplicite, come la valutazione di un film, o implicite e quindi derivanti dalle loro attività, come ad es. la consultazione della pagina di un determinato oggetto su un sito di e-commerce.

Tali sistemi di raccomandazione svolgono un ruolo importante in siti Internet altamente qualificati come Amazon.com, YouTube, Yahoo, Tripadvisor e Netflix, basti pensare che quest'ultimo, nel 2006, ha messo in palio 1 milione di dollari che sarebbe andato a chiunque avesse migliorato l'algoritmo di raccomandazione del 10%.

In particolare i sistemi di raccomandazioni si basano principalmente su tre tipi di approcci differenti: quello collaborativo, quello basato sui contenuti e quello ibrido.

Il primo nasce dalla volontà di consigliare articoli all'utente attivo basandosi sui "consigli" degli altri utenti, come avviene offline con il passaparola. I sistemi di raccomandazione basati su questi metodi ovviamente sono superiori rispetto al semplice passaparola offline poiché utilizzano migliaia se non milioni di opinioni di utenti per consigliare all'utente attivo gli articoli che sono stati maggiormente apprezzati da utenti a lui simili. L'idea chiave è che la valutazione dell'utente u per un nuovo elemento è probabile che sia simile a quella di un altro utente v , se u e v hanno valutato altri articoli in modo simile. Allo stesso modo, è probabile che valuti due articoli i e j in modo simile, se altri utenti hanno dato valutazioni simili a questi due elementi. Più formalmente, l'utilità $u(c,s)$ dell'articolo s per l'utente c è stimata basandosi sull'utilità $u(j,s)$ assegnata all'articolo s dall'utente $j \in C$ che risulta essere simile all'utente attivo c .

Gli algoritmi di filtraggio collaborativo sono tipicamente suddivisi in *user-to-user* e *item-to-item collaborative filtering*. In quello *user-to-user*, il sistema, come prima fase, ricerca gli utenti più simili all'utente attivo creando un insieme di user simili tra loro, successivamente valuta i rating di tali utenti in relazione ad un articolo non valutato dall'utente attivo ed infine crea una lista di possibili raccomandazioni da mostrargli.

Gli algoritmi *item-to-item* creano correlazioni tra coppie di articoli e quindi generano raccomandazioni trovando elementi simili al set di elementi che l'utente attivo ha apprezzato (implicitamente o esplicitamente). Questo approccio non va confuso con i metodi *content-based* in quanto le similitudini tra gli oggetti vengono estrapolate dalle interazioni passate tra gli utenti ed il sistema. In altre parole, se l'approccio *user-to-user* si basa su utenti simili che apprezzano elementi simili, l'approccio *item-to-item* si basa sugli utenti che apprezzano articoli simili agli articoli che hanno preferito.

Quindi, in entrambi i metodi, quello *user-to-user* e *item-to-item*, l'idea è quella di generare raccomandazioni in base alle preferenze degli utenti con gusti simili all'utente corrente, cioè ignorare le caratteristiche dell'articolo e sfruttare le preferenze collettive, e in tali approcci, due elementi sono simili se diversi utenti del sistema hanno valutato questi articoli in modo simile.

Altro metodo appartenente ai sistemi di raccomandazioni basati sul filtraggio collaborativo è il *latent factor model*. Questo è un approccio alternativo che cerca di spiegare le valutazioni degli utenti attraverso la ricerca di fattori latenti utili per costruire i profili di user ed elementi. Ad es. per i film, tali fattori latenti potrebbero essere il genere, thriller o commedia, la quantità di azione, se è più adatto agli uomini, alle donne o ai bambini, lo sviluppo dei personaggi ecc. Le applicazioni di maggior successo del *latent factor model* si basano sulla tecnica chiamata *matrix factorization*. Quest'ultima ci consente di scoprire i fattori latenti nascosti che sono alla base delle valutazioni date dagli utenti. Ad es., supponiamo di avere 5 utenti e 6 film, e la valutazione è di numeri interi tra 1 e 5, il compito della *matrix factorization* è quello di prevedere le valutazioni mancanti. L'idea alla base di questa tecnica è che ci sono caratteristiche nascoste negli elementi che determinano come un utente valuta un film. Ad es., un utente che dà a due film con Di Caprio una valutazione elevata, potrebbe significare che una caratteristica nascosta potrebbe essere che apprezzi Di Caprio come attore.

Se si riescono a scoprire questi fattori latenti, si può prevedere una valutazione di un film, non ancora votato da un utente.

Nel metodo basato sui contenuti, invece, il sistema impara a consigliare articoli simili a quelli che in passato l'utente ha acquistato ed apprezzato. La somiglianza degli oggetti è calcolata in base alle caratteristiche degli articoli. In particolare, l'utilità $u(c,s)$ dell'articolo s per l'utente c , è stimata basandosi sulle utilità assegnate dall'utente c all'oggetto $s_i \in S$ che è simile all'item s . I suggerimenti basati sui contenuti, quindi, si basano sul presupposto che articoli con caratteristiche oggettive o attributi simili, siano valutati allo stesso modo.

Ad esempio, nel caso di raccomandazioni di film, per mostrare dei suggerimenti all'utente c , il sistema di raccomandazione *content based* cerca di comprendere le caratteristiche in comune tra i film che l'utente ha votato in maniera positiva nel passato (attori specifici, registi, generi, ecc.) e solo i film che hanno il più alto grado di similarità con le preferenze dell'utente saranno raccomandati. In altre parole, un approccio basato sui contenuti crea un profilo dell'utente in base alle caratteristiche presenti negli articoli che l'utente ha valutato positivamente⁵⁸.

Il fatto che tutti gli algoritmi utilizzati nei sistemi di raccomandazione abbiano i loro punti di forza e di debolezza, ha portato a molti tipi di combinazioni per combattere i punti deboli e sfruttare quelli di forza di ognuno di loro. Quindi, oltre a coprire le debolezze, gli approcci vengono combinati per essere in grado di sfruttare i punti di forza dei diversi metodi a disposizione, ad es. si possono sfruttare congiuntamente sia gli attributi dell'articolo che l'utente ama (filtraggio basato sul contenuto), sia le preferenze della comunità (filtraggio collaborativo).

Un sistema di raccomandazione ibrido, quindi, è qualsiasi sistema di raccomandazione che combini insieme più tecniche di raccomandazione per produrre i suoi risultati⁵⁹.

Obiettivo di questa tesi è analizzare, testare ed ipotizzare e testare, tramite analisi empirica, nuove possibili applicazioni, da me pensate, per l'"Explicit Factor Model", un modello ideato e descritto da Yongfeng

⁵⁸ Burke 2002

⁵⁹ Burke 2007

Zhang, Guokun Lai, Min Zhang, Yiqun Liu e Shaoping Ma all'interno del paper "*Explicit Factor Models for explainable recommendation based on phrase-level sentiment analysis*", che unisce insieme sistemi di raccomandazione ed analisi testuale in un approccio innovativo, non ancora utilizzato da nessun sito web, che ha l'obiettivo di migliorare la qualità dei consigli mostrati agli utenti, dare delle spiegazioni sul perché di tali consigli e fornire delle dis-raccomandazioni riguardo entità da non acquistare poiché non adatte al consumatore che le sta valutando. L'idea di un sistema di raccomandazione unito con la *sentiment analysis* è cresciuta in me dopo aver cercato informazioni su questi due strumenti, che ritengo essenziali in un mondo dominato dai dati e dalle informazioni in essi contenute, e dopo aver pensato alla centralità delle recensioni online nelle scelte di acquisto della quasi totalità dei consumatori.

È un'idea rivoluzionaria, perché se è vero che la maggior parte delle aziende operanti online, come Amazon.com e Netflix ad esempio, hanno alla base del loro business i sistemi di raccomandazione, e se è vero che l'analisi testuale è utilizzata già per diversi motivi dalle imprese, è anche vero che ancora nessuna di queste utilizza un modello come questo. E proprio questa sua assenza dal mercato attuale potrebbe fornire alle aziende un vantaggio importante, sia in termini di miglior servizio offerto, con maggiore *customer satisfaction*, sia di miglior *brand image* percepita dai consumatori.

Negli ultimi anni i *Latent Factor Model*, ed in particolare la tecnica *Matrix Factorization*, sono stati studiati con molta attenzione dalla comunità di ricerca e dall'industria grazie alla loro capacità di fornire delle raccomandazioni precise su diversi tipologie di dati.

Tuttavia, i *recommendation systems* basati su questi algoritmi incontrano alcuni problemi importanti nelle applicazioni pratiche. Innanzitutto, è difficile sapere in che modo gli utenti compongono il loro giudizio sui vari aspetti di un'entità, il che rende difficile formulare delle raccomandazioni in base alle esigenze ed ai gusti specifici degli utenti, ed in secondo luogo, solitamente è difficile dare spiegazioni sul motivo per cui un articolo viene raccomandato, e ancora più complicato è spiegare perché un'entità non è raccomandata. Ed inoltre va aggiunto che la mancanza di spiegazioni indebolisce la capacità di persuadere gli utenti e di aiutarli a prendere decisioni di scelta ed acquisto migliori⁶⁰.

Le spiegazioni nei sistemi di raccomandazione potrebbero essere molto utili dato che, spiegando come funziona il sistema e/o perché un prodotto viene raccomandato, il sistema diventa più trasparente e dà la possibilità agli utenti di dire quando esso sbaglia. Inoltre, la maggior trasparenza, aumenta la confidenza e la fiducia degli utenti, aiuta gli user a prendere decisioni migliori e più rapide, migliorando così l'efficacia e l'efficienza, riesce più facilmente a convincere gli utenti a provare o acquistare un prodotto ed aumenta la loro soddisfazione riguardo l'utilizzo dei sistemi di raccomandazione rendendo più probabile una loro fedeltà futura.

A questo punto un problema potrebbe essere quello di dover scegliere tra un sistema di raccomandazione con un'elevata accuratezza, come ad esempio i *Latent Factor Model*, o uno che sia maggiormente trasparente e comprensibile rinunciando ad una maggior precisione nelle previsioni. Ma, fortunatamente, lo sviluppo della

⁶⁰ Vig, Sen and Riedel 2009

sentiment analysis e la crescente popolarità delle recensioni testuali degli utenti online potrebbero evitare ai service provider una scelta di questo tipo.

In particolare, la maggior parte dei siti di e-commerce e di servizi, come Amazon.com e TripAdvisor, consentono agli utenti di scrivere recensioni a testo libero congiuntamente ad una valutazione numerica. Come detto, le recensioni contengono informazioni dettagliate sulle opinioni, sugli atteggiamenti e sulle preferenze degli utenti nei confronti degli aspetti del prodotto che possono aiutare a creare nuovi modelli per raccomandazioni maggiormente trasparenti e spiegabili.

Inoltre, utenti diversi potrebbero essere interessati a diversi aspetti del prodotto, ad esempio c'è chi potrebbe essere maggiormente interessato alle dimensioni dello schermo di un cellulare, e chi invece alla durata della batteria, anche se entrambi potrebbero aver assegnato la stessa valutazione numerica al prodotto. Quindi estrarre le caratteristiche del prodotto e le relative opinioni degli utenti dalle recensioni, non solo aiuta a capire le diverse preferenze degli utenti e a formulare raccomandazioni migliori, ma aiuta anche a capire perché e come un determinato articolo dovrebbe o meno essere raccomandato. In questo modo, non solo si potrebbero raccomandare agli utenti determinati prodotti, ma si potrebbe anche spiegarli il perché non dovrebbero acquistare una data entità.

L'*Explicit Factor Model* ha l'obiettivo di ottenere sia un'elevata precisione sia un'elevata trasparenza e comprensibilità attraverso delle raccomandazioni frutto di una *sentiment analysis* sulle recensioni degli utenti utilizzata per individuare le caratteristiche ricercate da ogni user e quelle possedute dai prodotti nel catalogo, ad esempio se il sistema trova che un utente potrebbe essere particolarmente interessato alla memoria, agli auricolari ed alla batteria di un cellulare, allora gli raccomanderà un prodotto che funziona particolarmente bene su questi tre aspetti.

Come detto in precedenza, nella maggior parte delle applicazioni di analisi del *sentiment*, il *sentiment lexicon* svolge un ruolo centrale. Tuttavia, non ne esiste uno universalmente utilizzabile, poiché la polarità delle parole è sensibile al contesto dell'argomento. Ancor peggio, nello stesso dominio la stessa parola può indicare un sentimento differente rispetto a diversi aspetti, ad esempio, in una recensione di un computer portatile, "grande" potrebbe essere negativo per l'aspetto batteria, pur essendo positivo per l'aspetto schermo. L'obiettivo è quindi di creare un *sentiment lexicon* che non sia solo dipendente dal contesto, ma che tenga conto anche dei diversi aspetti considerati all'interno dello stesso dominio.

In questo modello, la creazione del *sentiment lexicon*⁶¹ viene effettuata combinando quattro tipi di vincoli, rispettivamente:

- vincoli che derivano da *sentiment lexicon* generali, ovvero che contengono parole che sono quasi sempre positive o negative in qualsiasi contesto, come "eccellente" e "cattivo";

⁶¹ Y. Lu, M. Castellanos, U. Dayal, and C. Zhai 2011

- vincoli basati sulla valutazione complessiva del *sentiment* in tutto il documento. In molti casi, ogni recensione viene fornita insieme con una valutazione complessiva dell'opinione da parte dell'utente, come nelle recensioni su TripAdvisor e Amazon. Questo tipo di valutazione complessiva è abbondante sul Web, ad esempio, ci sono oltre 40 milioni di recensioni relative ai viaggi su TripAdvisor. La logica alla base di questo vincolo, è che la valutazione complessiva trasmette alcune informazioni sul sentimento espresso nel testo, ad esempio, è improbabile che un utente utilizzi solo parole negative nel testo dando poi una valutazione complessiva di 5 stelle;
- vincoli basati su di un dizionario dei sinonimi e contrari. Ad esempio, potremmo non sapere se l'aggettivo "largo" è positivo o negativo per l'aspetto dello schermo in una recensione su un computer portatile, ma sappiamo che dovrebbe avere un significato simile a "grande" e molto diverso da "piccolo". Quindi se abbiamo informazioni riguardo la polarità di "grande" o di "piccolo", possiamo dedurre facilmente la polarità di "largo";
- vincoli basati sull'euristica linguistica: regola della "e": le frasi che sono connesse con la congiunzione "e" solitamente esprimono la stessa polarità di sentimento. Ad esempio, "la batteria dura a lungo e le dimensioni dello schermo sono grandi" significa che "lungo" per "batteria" e "grande" per "dimensione schermo" hanno la stessa polarità.

Regola del "ma": le frasi che sono connesse con "ma" solitamente esprimono una polarità di sentimento opposto. Ad esempio, "la batteria dura a lungo ma la dimensione dello schermo è minuscola" indica che "lungo" per "batteria" e "minuscolo" per "dimensioni dello schermo" sono della polarità opposta.

Regola della negazione: parole di negazione come "no", "non" e "mai" invertono il sentimento della parola di opinione nella stessa frase. Ad esempio, "non felice" dovrebbe avere sentimento positivo come "felice", ma invece ha sentimento negativo grazie alla presenza del "non".

In particolare tale creazione del *sentiment lexicon* L , che tiene conto del contesto, prevede tre fasi: la prima prevede l'estrazione del set degli aspetti F , successivamente vengono estratte le parole d'opinione O che vengono abbinate agli aspetti a cui si riferiscono creando così la coppia (F,O) ed infine vi è l'assegnazione della polarità del *sentiment* S ad ogni coppia aspetto-opinione portando alla tripletta finale (F,O,S) con S che può avere un valore di $+1$ o -1 .

Una volta creato il *sentiment lexicon* L , bisogna andare ad individuare le coppie aspetto-opinione (F,S') , dove S' rappresenta l'opinione dell'*opinion holder* sull'aspetto dell'entità menzionata nella recensione.

In particolare il processo di estrazione di questi due elementi è suddiviso in tre passaggi⁶²: estrazione, unione e filtraggio. Nella prima fase ogni recensione viene scomposta in un elenco di parole che indicano un aspetto, un'opinione e negazioni, e vengono classificate in base all'ordine in cui appaiono nella recensione. Nella

⁶² Yunzhi Tan, Yongfeng Zhang, Min Zhang, Yiqun Liu, Shaoping Ma 2013

seconda fase, saranno create le coppie di aspetto-opinione in base al contesto e al *sentiment lexicon*. Inoltre, in questo step, viene definita anche la polarità di sentimento di ciascuna coppia. Quindi, nell'ultima fase, viene valutata l'utilizzabilità di ogni coppia estratta.

In questo modello si assume che utenti diversi siano interessati a diversi aspetti e che tendono a recensire quelli che per loro sono più importanti. Quindi è utile costruire una matrice di attenzione utente-aspetto X in cui ogni elemento misura quanto ad un utente interessa un determinato aspetto di un prodotto.

In particolare, sia $F = \{F_1, F_2, \dots, F_p\}$ l'insieme di caratteristiche/aspetti espliciti del prodotto, e sia $U = \{u_1, u_2, \dots, u_m\}$ l'insieme che denota gli m utenti, per generare la matrice, si considerano tutte le recensioni di testo scritte da un utente u_i , e vengono estratte tutte le coppie (F, S) . Ipotizzando che la funzione F_j sia menzionata dall'utente u_i per t_{ij} volte, ogni elemento nella matrice di attenzione X è definito come segue:

$$X_{ij} = \begin{cases} 0, & \text{se l'utente } u_i \text{ non menziona l'aspetto } F_j \\ 1 + (N - 1) \left(\frac{2}{1 + e^{-t_{ij}}} - 1 \right), & \text{se è menzionato} \end{cases}$$

N è solitamente uguale a 5 poiché in molti siti le recensioni possono essere accompagnate da un punteggio che va da 1 a 5 stelle, come avviene su Amazon.com e TripAdvisor.

Altra matrice importante per tale modello è la matrice di qualità prodotto-aspetto Y , dove ogni elemento misura la qualità di un prodotto per un determinato aspetto corrispondente.

Sia $P = \{p_1, p_2, \dots, p_n\}$ l'insieme degli n prodotti. Per ciascuno degli elementi p_i , si utilizzano tutte le recensioni corrispondenti e vengono estratte tutte le coppie (F, S) corrispondenti. Supponendo che l'aspetto F_j sia menzionato per k volte riguardo il prodotto p_i , e che la media del sentimento riguardo la caratteristica F_j nelle quelle k menzioni sia s_{ij} , la misura del prodotto-aspetto Y_{ij} è calcolata così:

$$Y_{ij} = \begin{cases} 0, & \text{se il prodotto } p_i \text{ non è recensito sull'aspetto } F_j \\ 1 + \frac{N-1}{1 + e^{-k s_{ij}}}, & \text{se lo è} \end{cases}$$

Questa misura cattura sia l'orientamento del sentimento (attraverso s_{ij}) sia la popolarità (attraverso k) dell'aspetto F_j per il prodotto p_i .

Gli elementi diversi da zero nelle matrici X e Y sopra menzionate indicano le relazioni osservate tra utenti, oggetti ed aspetti espliciti. Questi dati vanno poi integrati in un modello di fattorizzazione per ottenere delle raccomandazioni che siano accurate e spiegabili. In particolare, la costruzione di tale modello di fattorizzazione, che utilizza la matrice di attenzione e quella di qualità, si sostanzia nello stimare le valutazioni probabili degli utenti, aspetti ed oggetti basandosi sulle relazioni osservate tra gli utenti e gli aspetti e tra gli oggetti e gli aspetti.

Assumendo che la valutazione complessiva di un utente relativa ad un prodotto è basata sull'opinione complessiva che lui stesso ha su diversi aspetti dell'articolo, considerare solo gli aspetti espliciti (U_1 e U_2), che

catturano la qualità dell'articolo e l'attenzione degli utenti, non è sufficiente per spiegare il punteggio assegnato dall'utente, e quindi vanno considerati anche i fattori nascosti (*latent factor*, H_1 e H_2) utili nella creazione della matrice di classificazione generale A .

Una volta trovata la soluzione ottima del modello di fattorizzazione, è possibile stimare qualsiasi elemento mancante nella matrice di attenzione utente-aspetto X , in quella di qualità oggetto-aspetto Y ed in quella utente-oggetto A .

Partendo da questo presupposto, è possibile generare delle raccomandazioni personalizzate Top-k e fornire delle spiegazioni a livello di aspetto.

La prima tipologia di raccomandazioni, Top-k, parte dal presupposto che gli utenti, quando devono effettuare un acquisto, basano la loro valutazione su alcuni aspetti particolari del prodotto, e non su tutte le caratteristiche dello stesso. Quindi, in questo metodo, quando viene generato l'elenco delle raccomandazioni personalizzate, si tiene conto dei k aspetti più importanti per un utente. In particolare il punteggio assegnato ad ogni oggetto, che viene utilizzato per creare la classifica delle raccomandazioni, si basa sul punteggio di somiglianza utente-articolo basato sui k aspetti più importanti del prodotto che sono di interesse dell'utente.

Ci potrebbero poi essere diversi modi per creare delle raccomandazioni esplicite, ad esempio potrebbe apparire una di queste due frasi quando l'utente si trova sulla pagina di un prodotto: "Potresti essere interessato all'aspetto x , sul quale questo prodotto ha buone performance" o "Potresti essere interessato all'aspetto x , sul quale questo prodotto non ha buone performance".

In particolare i vantaggi delle raccomandazioni esplicite sono principalmente 5:

- intuitività: il sistema agli occhi dell'utente è più trasparente e facile da capire;
- efficacia: vi è un aumento di confidenza e fiducia nel sistema degli utenti che sono in grado di prendere delle decisioni migliori;
- efficienza: gli utenti sono in grado di prendere delle decisioni in maniera più rapida
- persuasività: gli utenti sono maggiormente orientati alla prova ed all'acquisto
- soddisfazione: vi è un aumento di soddisfazione degli utenti poiché aumenta anche la facilità di utilizzo.

All'interno del paper, per testare il modello, sono state condotte due tipologie di esperimenti.

La prima, quella offline, è servita per valutare l'accuratezza delle raccomandazioni, ovvero la capacità di predire le valutazioni e di effettuare delle raccomandazioni Top-k di un certo livello. Gli esperimenti online, invece, avevano il compito di valutare l'efficacia delle spiegazioni date agli utenti concentrandosi sulla loro persuasività.

Negli esperimenti offline sono stati utilizzati due set di dati: uno contenente le recensioni degli utenti sulle attività commerciali situate nella città di Phoenix negli Stati Uniti su Yelp e il set di dati Dianping, costituito dalle recensioni sui ristoranti situati in diverse città principali della Cina. In particolare però il set di dati di

Yelp include molti utenti che hanno fatto pochissime recensioni, infatti il 49% di loro ha effettuato solo una recensione, il che renderebbe difficile valutare le prestazioni della raccomandazione top-K. Di conseguenza sono stati selezionati solo gli utenti con 10 o più recensioni.

La prima analisi è stata effettuata sulla costruzione del *sentiment lexicon*, ed i risultati sono stati positivi dato che la precisione nell'individuazione delle parole di opinione, di aspetto e dei sentimenti non è mai risultata essere inferiore all'89%. In particolare si è visto come il *sentiment lexicon* agisse bene in diversi contesti, riuscendo ad individuare i sentimenti che dipendevano dalle parole che indicavano aspetti e quelli che dipendevano dalle parole di opinione, ad esempio se la coppia (F, O) era (prezzo, alto) il sentimento riportato era negativo, mentre se si aveva (qualità del servizio, alta) era ovviamente positivo.

Per verificare l'ipotesi che gli utenti si concentrano su diversi aspetti, è stata costruita la matrice di attenzione utente-aspetto X sul set di dati di Yelp, ed è stata condotta una clusterizzazione per assegnare gli utenti in 5 cluster, ed ognuno di questi conteneva 5 aspetti dei ristoranti ordinati in modo decrescente in base alla loro frequenza di apparizione nelle recensioni. In questo modo si è visto come gli utenti di ciascun cluster erano più interessati ad un determinato sottoinsieme di caratteristiche, ed ogni sottoinsieme era riconducibile ad un dato aspetto del prodotto.

Una seconda analisi è stata effettuata sul numero di fattori espliciti e fattori latenti da dover utilizzare. Questi due elementi sono importanti perché acquisiscono due diversi tipi di informazioni nel processo di fattorizzazione per la predizione delle valutazioni. Gli esperimenti sono stati effettuati con un numero totale di fattori pari a 100, e si è notato che quando il modello utilizzava tra il 30 e l'80% di fattori espliciti riusciva ad avere delle ottime prestazioni in termini di RMSE (radice dell'errore quadratico medio). Ciò evidenzia come, sebbene l'incorporazione di fattori espliciti migliori la precisione della previsione rispetto ai modelli che non li utilizzano, è necessario utilizzare una moderata quantità di fattori latenti per garantire flessibilità al modello, poiché i fattori espliciti potrebbero non acquisire completamente tutti i criteri che portano l'utente ad una determinata valutazione.

Altra analisi ha avuto ad oggetto il numero di caratteristiche più importanti per gli utenti da dover utilizzare per effettuare delle raccomandazioni di tipo *top-k*. In particolare questo esperimento conferma l'ipotesi che utilizzare troppe caratteristiche introdurrebbe rumore nella procedura di raccomandazione intaccandone l'efficacia.

Per analizzare ulteriormente le relazioni tra le caratteristiche e gli utenti, e capire l'impatto e l'importanza delle caratteristiche esplicite, calcoliamo la copertura media delle caratteristiche più importanti in termini di frequenza di apparizione rispetto a tutte le caratteristiche nelle recensioni di un utente.

In particolare si nota come un piccolo numero di caratteristiche più importanti per gli utenti domina la copertura della frequenza dei termini nelle recensioni degli utenti.

La seconda tipologia di esperimenti, quella online, è stata condotta un sito di e-commerce reale con lo scopo di studiare gli effetti delle raccomandazioni contenenti delle spiegazioni sulla disponibilità degli utenti a seguire tali consigli (persuasività delle raccomandazioni).

In particolare è stato condotto un test A/B su un popolare sito web che vanta più di 100 milioni di utenti. L'obiettivo era quello di consigliare i cellulari migliori per ogni utente.

La schermata presentava 4 cellulari sul lato destro dello schermo e, quando l'utente passava con l'indicatore del mouse su uno di questi, appariva un indicatore che lo consigliava o sconsigliava, mostrando anche delle coppie di parole aspetto-opinione colorate diversamente a seconda che l'opinione fosse positiva o negativa e di dimensione proporzionale alla frequenza di apparizione di quei termini nelle recensioni dell'oggetto in questione.

Per analizzare gli effetti di tali raccomandazioni, ogni utente è stato assegnato in maniera casuale ad uno di questi tre gruppi: utenti A (gruppo sperimentale) ricevono le spiegazioni a livello di aspetto sull'oggetto preso in considerazione, gli utenti B (gruppo di confronto) ricevono le spiegazioni di tipo "gli utenti hanno visto anche" e utenti C (gruppo di controllo) non ricevono spiegazioni

Il *Click Through Rate* (CTR), calcolato come rapporto tra le visualizzazioni ed i click, del gruppo sperimentale A è maggiore rispetto a quello degli altri due gruppi, dimostrando che le spiegazioni a livello di aspetto fornite dall'EFM sono più efficaci nel persuadere gli utenti ad esaminare più attentamente gli articoli consigliati.

Per studiare come le "dis-raccomandazioni" influenzano gli utenti, invece, sono stati utilizzati solo due gruppi di utenti, il gruppo A, quello sperimentale, dove gli utenti potevano vedere le spiegazioni quando passavano l'indicatore del mouse sull'oggetto, e quello B dove gli utenti non ricevevano alcuna spiegazione dato che non ci sono sistemi che forniscono delle spiegazioni per una dis-raccomandazione.

Gli indicatori utilizzati per confrontare i due gruppi di utenti erano: percentuale di utenti che avevano aggiunto al carrello un determinato prodotto e le percentuali di utenti che sono d'accordo con la raccomandazione o la dis-raccomandazione.

Anche in questo caso i dati del gruppo sperimentale A sono migliori rispetto a quelli del gruppo di controllo B, e ciò dimostra che la presentazione delle spiegazioni migliora la capacità persuasiva delle raccomandazioni e allo stesso tempo migliora la conversione, cioè l'aggiunta al carrello, negli scenari di acquisto online reali.

Quindi si può dire che le spiegazioni aiutano a persuadere un utente ad aggiungere un prodotto consigliato al carrello o ad ignorarne uno non consigliato.

Il questionario da me ideato e somministrato ad un campione di 140 persone, aveva l'obiettivo di andare a raccogliere dati riguardo le modalità con cui i consumatori italiani si avvicinano alle recensioni online e riguardo l'applicabilità dell'*Explicit Factor Model* nel mercato nostrano, dato che i test effettuati dagli ideatori di tale modello sono stati effettuati sul mercato cinese ed americano.

Il questionario presenta una struttura divisa in 5 sezioni:

6. informazioni generali
7. informazioni sulle modalità di scrittura delle recensioni
8. proposta di un *Explicit Factor Model* per siti di e-commerce
9. proposta di un *Explicit Factor Model* per siti di recensioni di attività ristorative

10. proposta di un *Explicit Factor Model* per siti di organizzazione di viaggi online

La prima sezione comprendeva le domande su età, sesso e sull'aver mai scritto o meno una recensione. La terza domanda di questa prima sezione fornisce già un'indicazione per lo studio, in particolare mostra come ben il 71% dei rispondenti abbia effettuato almeno una recensione online. Tale dato è già indicativo, poiché ci dice che la maggioranza degli utenti online ha scritto un suo parere su un prodotto o un servizio rendendo possibile l'applicabilità dell'*Explicit Factor Model*, dato che si basa sulla costruzione di profili utenti partendo proprio dalle recensioni scritte da loro stessi.

La seconda sezione, invece, appariva solo a coloro avessero risposto positivamente alla terza domanda, dato che riguardava le modalità con cui solitamente viene scritta una recensione. In particolare le domande di questa categoria erano sui prodotti o servizi sui quali il rispondente aveva effettuato una recensione, con possibilità di scegliere anche più di una sola risposta, e sugli aspetti sui quali si concentrava quando ne scriveva una, aspetti più importanti per lui, aspetti che si sono rivelati maggiormente positivi o quelli maggiormente negativi. La prima domanda di questa sezione ha mostrato come la maggior parte delle recensioni venga fatta su ristoranti e bar, al secondo posto le recensioni sugli hotel, al terzo quello sugli articoli di e-commerce, al quarto quelle sui viaggi ed all'ultimo posto le recensioni su "altro".

Oggetto delle domande successive della sezione, era capire le abitudini dei consumatori riguardo la scrittura delle recensioni, nel dettaglio l'intento era individuare cosa spingesse gli user a recensire e su quali aspetti si concentrano maggiormente quando lo fanno. Come risalta dai risultati, gli utenti incentrano principalmente le loro recensioni sugli aspetti per loro più importanti, rendendo applicabile il modello, che, come detto, punta a costruire profili di utenti basati sulle loro recensioni ed in particolare sugli aspetti dei prodotti e servizi citati in queste ultime. In altre parole, se un consumatore x nelle recensioni di diversi hotel si concentra sull'aspetto "pulizia", è probabile che questo sia per lui molto importante.

Nella terza sezione veniva chiesto al rispondente di immaginare un sito di e-commerce che utilizzi le recensioni sue e degli altri utenti per creare dei suggerimenti per gli acquisti.

Entrambe le domande di questa parte di questionario hanno ricevuto degli ottimi risultati, ottenendo, rispettivamente, il 66% ed il 77% di voti tra il 4 ed il 5, su una scala dove 1 sta per "per niente" e 5 sta per "moltissimo". Salta all'occhio il risultato della seconda domanda, che ha raggiunto addirittura il 60% di voti uguali a 5 ("moltissimo"), per un totale 84 voti su 140, risultato non raggiunto da nessuna altra domanda del questionario. Questo indica come la possibilità di ricevere delle dis-raccomandazioni sia una funzionalità che sarebbe particolarmente apprezzata dagli utenti, i quali sempre più utilizzano le recensioni per effettuare una scelta e sempre più acquistano su siti di e-commerce, così come mostra una ricerca⁶³ riportata sul sito di vpnMentor.

⁶³ <https://it.vpnmentor.com/blog/vital-internet-trends/>

Le tre domande facenti parte della quarta sezione, avevano lo scopo di testare l'applicabilità dell'*Explicit Factor Model* nell'ambito della ristorazione. Il possibile utilizzo di tale modello in questo campo non è un'idea nata dagli autori del *paper*, ma è venuta a me personalmente.

L'idea di fondo è quella di utilizzare la *sentiment analysis* per andare ad individuare, all'interno delle recensioni, ciò che più interessa agli utenti nel mondo dei ristoranti e dei bar, in che cibo, bevanda o servizio eccelle una determinata attività e cosa invece un cliente dovrebbe evitare nella sua esperienza in un dato locale. Andando ad analizzare i risultati ottenuti, si nota che anche in questo caso sono positivi e che l'idea è stata apprezzata. In particolare le percentuali degli utenti che hanno votato il valore 4 ed il 5 sono, rispettivamente per le tre domande, le seguenti: 68%, 65% e 62%.

Come si nota queste percentuali sono inferiori a quelle riscontrate nelle due domande precedenti, e come si vedrà in seguito, sono inferiori, fatta eccezione per la percentuale della prima domanda, anche a quelle delle due domande successive.

Anche solo questo dato potrebbe bastare per capire che questa proposta sia meno allettante agli occhi degli utenti italiani rispetto alla precedente, ma un altro dato ci conferma questa affermazione, la percentuale di voti registrati per il valore 3, quello intermedio. In particolare le percentuali ottenute da questo valore sono rispettivamente del 24%, 23% e 26%, per le tre domande in questione, valori abbastanza maggiori rispetto a quelli del 15% ed 11% ricevuti dai due quesiti precedenti ed a quelli del 17% ottenuti dalle ultime due domande del questionario. Il valore 3, essendo quello intermedio, esprime una certa indifferenza del rispondente alla domanda posta, ed il fatto che le percentuali di questa risposta siano così elevate, rispetto a quelle delle altre domande, ci mostra come probabilmente questo sia un servizio che attira meno gli utenti rispetto agli altri proposti. Una possibile spiegazione a ciò potrebbe essere che il gusto è un qualcosa di prettamente personale, e quindi, secondo i rispondenti, delle raccomandazioni in tal senso potrebbero essere poco utili ed addirittura indurre in errore.

Come per la sezione precedente, anche nell'ultima, le domande poste ai rispondenti avevano lo scopo di verificare il grado di apprezzamento di un'idea mia e non nata dagli autori del *paper*.

L'obiettivo era quello di testare il grado di apprezzamento degli utenti nei confronti di un ipotetico sistema di raccomandazione che utilizzi le recensioni testuali per fornire delle raccomandazioni personalizzate e spiegate riguardo possibili hotel e mete dove poter trascorrere una vacanza.

Anche i risultati di questa quinta ed ultima sottosezione hanno mostrato come l'*Explicit Factor Model* possa essere utilizzato e soprattutto che sarebbe apprezzato dagli utenti. Andando ad analizzare i dati nel dettaglio, si vede come la prima domanda risulti essere quella che ha ricevuto una percentuale maggiore, rispetto a tutte le altre domande del questionario, di votazioni per i due valori maggiori, il 4 ed il 5. In particolare tale percentuale è del 77% rispetto al totale dei voti, numero eguagliato dalla domanda della terza sezione dove veniva chiesto agli utenti di esprimere il grado di apprezzamento per un sito web che mostrasse loro il perché non avrebbero dovuto completare un determinato acquisto.

Per quanto riguarda la seconda domanda, anche lei ha ottenuto un risultato positivo, nello specifico il 65% dei voti è stato distribuito tra i due valori maggiori, il 4 ed il 5.

Per concludere, i risultati ci hanno mostrato come gli ambiti in cui tale tipologia di sistema di raccomandazione sarebbe probabilmente maggiormente apprezzato siano quelli dell'e-commerce e del turismo, quest'ultimo da me ipotizzato. In particolare però, le funzionalità che hanno avuto il maggior riscontro positivo sono quelle relative alle dis-raccomandazioni per i prodotti di e-commerce e alle raccomandazioni riguardo l'hotel da dover prenotare.

Per quanto riguarda la prima, probabilmente sarebbe particolarmente apprezzata perché aiuterebbe nel non fare acquisti azzardati che potrebbero rivelarsi non adatti ai gusti ed alle funzionalità ricercate dagli utenti, facendogli risparmiare tempo e denaro, con risvolti positivi in termini di fiducia, fedeltà e *brand equity* per l'azienda. La seconda funzionalità, invece, consentirebbe di ridurre il tempo per la ricerca di un hotel che sia adatto ai gusti ed alle esigenze degli user, evitandogli delle esperienze negative in strutture con lacune sugli aspetti più importanti per loro. Dal lato dell'azienda, questa possibile funzione, probabilmente permetterebbe non solo un aumento in termini di fiducia, fedeltà e *brand equity*, ma anche in termini di prezzo che gli utenti sarebbero disposti a pagare, perché si tratterebbe di hotel che incontrano perfettamente i loro gusti e che non dovrebbe nascondere sgradite sorprese.

Infine, quindi, si può affermare che l'*Explicit Factor Model*, e le sue raccomandazioni e dis-raccomandazioni basate sull'analisi testuale delle recensioni utile per creare profili di utenti, con i relativi aspetti più importanti, ed aziende, con quelli maggiormente positivi e maggiormente negativi, sia applicabile, con buoni risultati previsti, nel mondo dell'e-commerce ed in quello degli hotel e dei viaggi in generale. Andrebbe invece ancora studiata la sua utilizzabilità nel mondo della ristorazione.