



Università degli Studi di Roma Torvergata

FACOLTÀ DI LETTERE E FILOSOFIA

Corso di Laurea in Scitur

TESI IN INFORMATICA E RAPPRESENTAZIONE DELLA
CONOSCENZA

**Essere Opinion Leader credibili per promuovere il proprio
prodotto turistico sul Web: Proposta di un Sistema per il
Rilevamento Automatico di Cheaters su Siti di Commenti
su Prodotti Turistici**

RELATORE:

Chiar.mo Fabio Massimo ZANZOTTO

LAUREANDA:

Luigia URSI

ANNO ACCADEMICO 2010-2011

*...alla forza della natura,
perché essa ci spinge a reg-
gerci in piedi da soli, per
quanto disperati possiamo
essere.*

Il computer non è una macchina
intelligente che aiuta le persone
stupide, anzi è una macchina
stupida che funziona solo nelle
mani delle persone intelligenti.

U. ECO

1	Introduzione	1
2	Il mondo di TripAdvisor: tra il vero e il falso	10
2.1	Breve storia di TripAdvisor	10
2.2	Un esempio di recensione falsa rilasciata dalle strutture alberghiere su “TripAdvisor”	16
2.3	Proposta di risoluzione del problema alle aziende alberghiere	17
3	Stato dell’arte	20
3.1	Teorie sugli Opinion Leaders	20
3.2	Opinion Mining automatico and Sentiment Analysis	23
3.3	Classificazione automatica	28
3.4	La categorizzazione	32
4	Raccolta ed analisi delle recensioni	39

4.1	Metodi utilizzati al fine di ricercare le recensioni “false” su TripAdvisor	39
4.2	Metodi di raccolta delle recensioni “false”	42
4.3	Raccolta dei tre casi di recensioni “false” accorte su TripAdvisor	45
4.3.1	Primo caso di recensione “falsa”	45
4.3.2	Secondo caso di recensione “falsa”	47
4.3.3	Terzo caso di recensione “falsa”	48
5	Proposta di metodo: diventare un Opinion Leader cre- dibile su TripAdvisor	51
5.1	Consigli utili per far salire la reputazione di un albergo	51
5.2	Esempio di una corretta recensione “falsa”	54
6	Conclusioni	56

Elenco delle figure

2.1	Immagine di TripAdvisor.	10
2.2	Caratteristiche TripAdvisor.	12
2.3	Punti forti e punti deboli di TripAdvisor.	13
2.4	Primo esempio di "falsa" recensione.	16
3.1	Gli Opinion Leaders.	23
3.2	SentiWordNet.	27
3.3	Opinion Mining automatico.	28
3.4	Esperimenti di Bruner.	33
3.5	Classificazione automatica.	38
4.1	Metodi di ricerca di recensioni "false".	39
4.2	Unione recensioni su Word.	43
4.3	I tre casi di recensione "falsa".	43
4.4	Primo gestore, primo caso.	45
4.5	Secondo gestore, primo caso.	45

ELENCO DELLE FIGURE

VI

4.6	Primo gestore, secondo caso.	47
4.7	Secondo gestore, secondo caso.	47
4.8	Primo gestore, terzo caso.	49
4.9	Secondo gestore, terzo caso.	49
5.1	Proposta di scrittura di "falsa" recensione.	53
5.2	Esempio di scrittura di "falsa" recensione.	55

Quello di cui ci occuperemo in questa tesi è di individuare su “TripAdvisor” nonché il network di viaggiatori che si scambiano foto, video, e consigli utili per pianificare al meglio la loro vacanza, le recensioni che hanno suscitato forti dubbi sulla loro veridicità e onestà. In particolare il nostro intento sarà soprattutto ravvisare quelle strutture ricettive, con un basso livello di gradimento da parte degli utenti, che vogliono far salire la propria reputazione da 1 a 5 stelle in poche ore, modificando opinioni negative.

Così andremo ad esplorare ogni recensione e dimostreremo quanto qui sopra citato attraverso degli esempi attendibili sulla base di alcune caratteristiche tra cui per esempio l’uso di aggettivi superlativi, senza mai scendere in particolari descrittivi o di parole come "io" e "me" come a sottolineare la propria credibilità e il ripetersi del nome dell’albergo nella recensione.

Dopo un’analisi e un’accurata esplorazione di questa piattaforma

dovremmo essere in grado di fungere da guida esaustiva alle aziende alberghiere per dare loro una metodologia su come pubblicare una recensione (su TripAdvisor) su un ristorante o un albergo per risollevere le opinioni sulla struttura stessa. Diverse agenzie estere si occupano di servizi di questo tipo e il nostro obiettivo vorrà essere anche dimostrazione di come il sito TripAdvisor , se inizialmente poteva sembrare un successo senza precedenti, adesso inizia a perdere quasi la sua importanza perché lasciare recensioni su TripAdvisor è semplice e soprattutto non c'è un controllo accurato sulla veridicità delle informazioni pubblicate. Per arrivare a parlare di TripAdvisor è bene premettere l'evoluzione dei consumatori con internet. Internet ha un notevole impatto sulle strategie che il consumatore può adottare per risolvere la problematicità relativa all'acquisto, che genera un fabbisogno informativo che la rete contribuisce a soddisfare ospitando:

- un gran numero di fonti informative relative ai prodotti e al loro utilizzo;
- contesti virtuali;
- strumenti e siti che assistono il consumatore nel processo di acquisto e agevolando l'analisi comparativa delle alternative.

L'aspettativa del consumatore è di poter esser supportato non solo nelle fasi di ricerca e selezione delle informazioni ma anche in quella di valutazione. Ciò si realizza attraverso agenti elettronici e siti collettivi

come quelli del consumerismo e le comunità virtuali di consumatori, che si raffigurano come spazi aperti per la ricerca e lo scambio di informazioni, la condivisione di esperienze, la sperimentazione e la produzione di conoscenze portando così l'apprendimento derivante dall'esperienza d'uso dei prodotti.

Un costrutto utile per lo studio del comportamento in internet è quello di set mentale (mindset).

La collaborazione e la comunità sono importanti caratteristiche dello sviluppo del Web 2.0, e sono caratteristiche chiave dei servizi di comunicazione come i siti di social network, virtual reality e comunità online.

Grazie a questi social network si possono esaminare gli effetti dei servizi del settore commerciale, cioè, l'impatto del rapporto che viene a crearsi tra azienda e cliente, perché essi tramite il meccanismo denominato word of mouth branding, ovvero passaparola, possono modificare le idee della gente.

Le persone fanno affidamento sulle famiglie, gli amici, e gli altri nel social network. La ricerca indica anche che le persone sembrano fidarsi di pareri apparentemente disinteressati di persone al di fuori del loro immediato social network, come ad esempio riviste online. Più precisamente il passaparola (WOM, in marketing voce del consumatore) è il processo di trasmissione delle informazioni da persona a persona e costituisce una forma particolare di impegno nella fase successiva

all'acquisto. In situazioni commerciali, il WOM può modificare gli atteggiamenti, le opinioni, o le relazioni da parte dei consumatori su aziende, prodotti o servizi con altre persone.

Il potenziale di voce dipende dal numero di clienti che non hanno scelto, o non ancora, l'opzione uscita, ovvero il cliente cessa di comprare i prodotti dell'azienda.

La voce comporta in genere dei costi per l'impresa, ma se il potenziale di voce si traduce in voce effettiva ed il management ha sviluppato un'adeguata capacità di ascolto e risposta, la voce innesca il processo di recupero dove il management può porre rimedio all'errore commesso.

Quanto detto ci serve per comprendere perché TripAdvisor o altri siti della stessa portata tra cui Venere.com o Booking Blog e così via sono diventati così importanti e perché serve anche per le aziende alberghiere (nel nostro caso) tenere sotto controllo queste piattaforme e controllare sempre la propria reputazione, il proprio andamento, cosa la gente pensa dei loro prodotti.

Proprio per questo, dunque, poiché le opinioni stanno diventando sempre più oggetto di studio da parte delle aziende per scoprire cosa la gente pensa dei loro prodotti, andremo ad argomentare l'"opinion mining and sentiment analysis" nonché i metodi di analisi di tali opinioni con le varie classificazioni e categorizzazioni. Il lavoro di tesi si propone di studiare quindi la reputazione in ambienti Web 2.0 e si pone

come obiettivo quello di dare delle metodologie di analisi che aiutino a trarre delle conclusioni, il più possibile dettagliate, sull'andamento del sentiment riguardante un brand o prodotto, nel nostro caso di un albergo.

Lo studio riportato in questo elaborato vuole essere sia un contributo alla ricerca accademica del settore, attraverso gli algoritmi implementati e le idee innovative proposte per ottenere un risultato, che un aiuto alle aziende che intendono valutare l'immagine del marchio e dei propri prodotti online, arricchendo la Business Intelligence con una nuova fonte di informazione(Lorenzo Radice and Saverio Bruno,2009).

Oggi giorno tutte le aziende che sviluppano business sanno che la capacità di monitorare in modo efficiente le opinioni sui prodotti propri e dei competitor è un elemento portante per affermare il proprio brand e acquisire un vantaggio competitivo apprezzabile. Non è sempre facile capire ciò che i clienti desiderano, né tanto meno rispondere in modo esauriente e puntuale alle loro richieste, che siano esse esplicite o implicite.

Grazie alla diffusione della comunicazione mediata da Internet, i clienti di un'azienda, di un servizio o di un prodotto trovano più facilmente recensioni e prove ospitate da siti istituzionali (siti di riviste o portali tematici). Inoltre si diffondono sempre più nuovi spazi dove esprimere le proprie opinioni indirette e chiedere pareri: questa condivisione di informazioni negli ambienti tematici dedicati (new-

sgroup, forum di discussione e mailing list) fornisce al cliente un potere decisionale straordinario.

Attualmente esistono in Italia:

- Centinaia di portali tematici e siti di pubblicazioni;
- decine di migliaia di mailing list;
- migliaia di forum di discussione ospitati da siti Web.

Questi punti di aggregazione sono dedicati agli argomenti più disparati e, per un'azienda, è semplice individuare quelli di maggior interesse per le proprie attività di analisi. Se da un lato, questi forum sono fondamentali per la qualità delle informazioni che forniscono sul mood dei clienti, dall'altro possono costituire una minaccia nascosta nel caso non vengano sottoposti a costante monitoraggio(Mario Gerosa, Roberta Milano, 2009).

Un'analisi di questo corpus con l'approccio manuale tradizionale comporterebbe il coinvolgimento di numerosi operatori con un conseguente aumento dei costi del servizio: stimando che il tempo medio di analisi di un documento sia di 2 minuti, i 140.000 documenti possono essere esaminati in circa 100 giorni uomo.

Se invece tutti i dati sulle opinioni dei clienti potessero essere collezionati e immagazzinati in database appropriati, si avrebbe a disposizione una fonte di informazioni ideale per applicazioni di data mining.

Affinché questa opportunità possa diventare reale, è necessario compiere un altro passo: la fonte di informazione testuale deve essere trasformata in informazione strutturata e inserita nei record di un database.

La soluzione si avvale di un sistema automatico per la raccolta e l'analisi dei documenti grazie al quale i messaggi vengono analizzati e classificati automaticamente dal software, in base al tipo di argomento trattato.

Attraverso l'uso del Text Mining, che integra due metodologie tradizionalmente isolate, il Natural Language Processing (NLP) e l'Information Management (IM), il sistema è in grado di identificare le parole, le frasi, le relazioni e persino “le emozioni” presenti all'interno dei testi delle e-mail.

I messaggi sono inoltre filtrati, categorizzati, e raggruppati secondo tassonomie predefinite. Il software viene inizialmente “educato”, ovvero supportato delle informazioni sul contesto di attività e su dove effettuare la raccolta delle opinioni, e in seguito opererà per fornire analisi in tempo reale o periodiche: la tecnologia permette di raggiungere un grado di precisione nell'interpretazione semantica e della categorizzazione di circa il 95%.

Dai risultati dell'analisi si otterranno strumenti in grado di delineare le particolari esigenze del cliente o i problemi inerenti a un determinato prodotto o servizio. Il tutto sotto forma di:

- Report
- Cruscotti
- Grafici e mappe.

Nella prima parte andremo a presentare TripAdvisor spiegandone in breve la sua storia e l'importanza che sta avendo ma sta anche perdendo dato le recensioni "false" che vi si trovano sempre più e quindi la sua perdita di credibilità, e in questa breve descrizione posteremo un esempio di recensione "falsa" proveniente da un gestore di un albergo o chi per lui rilascia, forse anche sotto pagamento, queste recensioni facendone comprendere l'ingenuità con cui sono state scritte. Parliamo d'ingenuità poichè dopo un'analisi accurata abbiamo supposto di essere di fronte a una "falsa" recensione.

Nel Cap.3 si passerà ad argomentare la Sentiment Analysis, Opinion Mining automatico con la sua classificazione e categorizzazione automatiche per far comprendere il perchè sono aree di ricerca interessate al fine della nostra ricerca e parleremo sempre nel medesimo capitolo degli Opinion Leaders, chi sono e il motivo per cui sono importanti circa il nostro fulcro del discorso nonché TripAdvisor . Nella terza parte viene applicato il lavoro svolto ai micro post estrapolati dal sito internet "TripAdvisor" cercando di capire meglio gli atteggiamenti delle persone con tre esempi di portata differente l'uni dagli altri.

Dopo uno studio del sito TripAdvisor, della ricerca di 50 recensioni “false” (anche se riportate come esempio solo tre) e dopo aver compreso il perché li abbiamo individuati, qual è stata l’ingenuità con cui sono stati rilasciati, dovremmo essere in grado di fungere da guida alle aziende alberghiere per dare loro una metodologia sul miglior modo per rilasciare una recensione al fine di diventare un opinion leader credibile e far salire la propria reputazione.

Il mondo di TripAdvisor: tra il vero e il falso

2.1 Breve storia di TripAdvisor



Figura 2.1: Immagine di TripAdvisor.

L'idea di costituire TripAdvisor nasce nel 2000 quando Steve Kaufer, fondatore e attuale direttore, decise di creare un sito che desse la

possibilità, a coloro che avevano soggiornato in un hotel, di scrivere una recensione a riguardo; grazie al contributo disinteressato degli utenti, si sarebbe costituita una sorta di guida ai migliori alberghi delle varie città e località turistiche. Con il passare del tempo il sito è diventato sempre di più un network di viaggiatori che si scambiano foto, video e consigli utili per pianificare la migliore vacanza possibile.

Così, grazie all'entusiasmo degli iscritti, TripAdvisor è diventato in pochi anni la più grande comunità di viaggiatori on line, con oltre quindici milioni di recensioni di hotel e venticinque milioni di visitatori unici al mese¹.

Inoltre TripAdvisor dispone di oltre quindici milioni di recensioni e opinioni provenienti da veri viaggiatori di tutto il mondo riguardo oltre 33.000 città, 230.000 hotel, 76.000 attrazioni turistiche e 435.000 ristoranti. Gli utenti del sito hanno inoltre messo a disposizione oltre un milione di fotografie riguardo 70.000 hotel.

La vacanza rappresenta per la maggior parte dei turisti il mezzo per vivere un'esperienza unica, in grado di ricreare un benessere psico-fisico spesso deteriorato dal lavoro e dalla routine quotidiana; di conseguenza, più che per l'acquisto di altri beni di consumo, se le aspettative del viaggiatore vengono disattese, l'insoddisfazione dell'individuo si ripercuoterà sulla vita di tutti i giorni. In base a quanto appena affermato è chiara la grande rilevanza delle opinioni di persone

¹Dati relativi all'intero network (TripAdvisor Media Network) pubblicati sul sito www.tripadvisor.it/pages/about_us.html.

che hanno effettuato la stessa tipologia di vacanza che l'individuo si appresta a fare. Il turista infatti, ormai abituato alle foto e le descrizioni dei cataloghi dei tour operator e dei siti internet dei vari operatori turistici, che tutto fanno tranne che descrivere la realtà, considera preziosi ed attendibili i consigli disinteressati di altri consumatori come lui (Ginny Mc Grath and Steve Keenan, 2007).

Il sito presenta molte caratteristiche tra cui (vedere Fig.2.2).

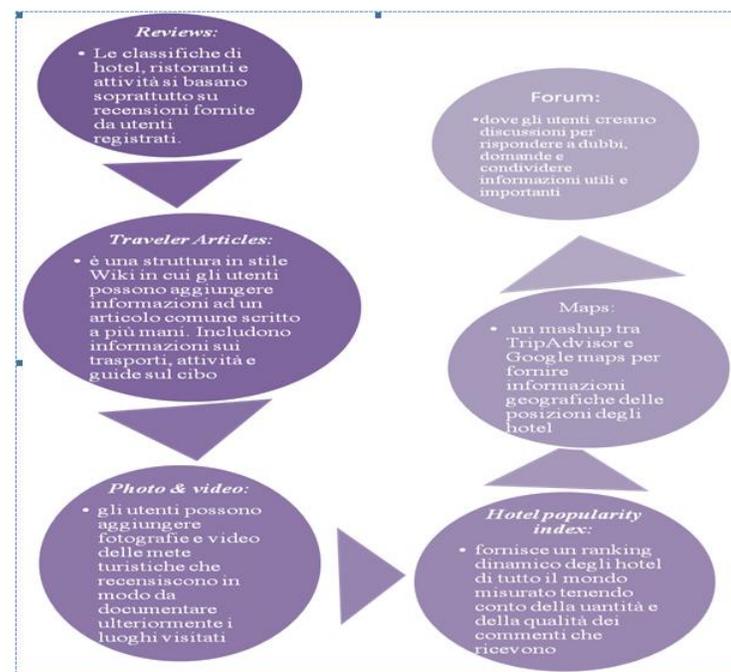


Figura 2.2: Caratteristiche TripAdvisor.

Quando si tratta di viaggiare, forse più di qualunque altro settore, le recensioni sono fondamentali. I viaggiatori prendono sempre più spesso decisioni su dove alloggiare, dove mangiare e cose da fare sulla base di raccomandazioni da amici in rete.

TripAdvisor ha recentemente svolto un'indagine coinvolgendo 2200 viaggiatori per comprendere quali sono le principali caratteristiche, siano esse positive o negative, che determinano la qualità di un hotel. Dalla ricerca è emerso che i fattori che determinano un buon giudizio sono la location e i letti comodi, rispettivamente per il 30% e il 29% degli intervistati, mentre il 24% considera essenziali la qualità dello staff e del servizio offerto dall'albergo. Risulta poi fondamentale la pulizia dell'hotel; infatti il 54% dei viaggiatori considera pessimo un albergo con camere sporche. Le altre cause che determinano un giudizio negativo sono gli ospiti rumorosi e uno staff scadente (<http://www.tripadvisor.it/>).



Figura 2.3: Punti forti e punti deboli di TripAdvisor.

Navigando nel sito si può notare come il 99% delle recensioni presenti sono relative al lavoro svolto dallo staff dell'albergo; per incentivare i commenti da parte dei clienti, risulta quindi di fondamentale importanza che il personale, oltre a tenere un comportamento professionale corretto, instauri un rapporto confidenziale con il turista.

Per dare un'idea di quanto TripAdvisor sia divenuto una componente fondamentale nel destino di una struttura ricettiva, basta dire che circa il 30% dei turisti prima di prenotare una camera d'albergo, verifica le eventuali recensioni lasciate da altri utenti. Data la grande influenza che le recensioni scritte dagli utenti di TripAdvisor hanno sui turisti, molti operatori del settore hanno suscitato forti dubbi sulla veridicità e l'onestà dei commenti che si trovano sul sito.

Se però inizialmente questa piattaforma poteva sembrare un successo senza precedenti, adesso inizia a perdere quasi la sua importanza perché lasciare recensioni su TripAdvisor è semplice e soprattutto non c'è un controllo accurato sulla veridicità delle informazioni pubblicate. Quindi da anni ne sta venendo fuori che:

1. "Ospiti" che non hanno mai soggiornato in un hotel possono influenzarne la classifica su TripAdvisor lasciando una finta recensione positiva o negativa;
2. Strutture ricettive con un basso livello di gradimento da parte degli utenti, possono far salire la propria reputazione da 1 a 5 stelle in poche ore, scrivendo finte recensioni positive.

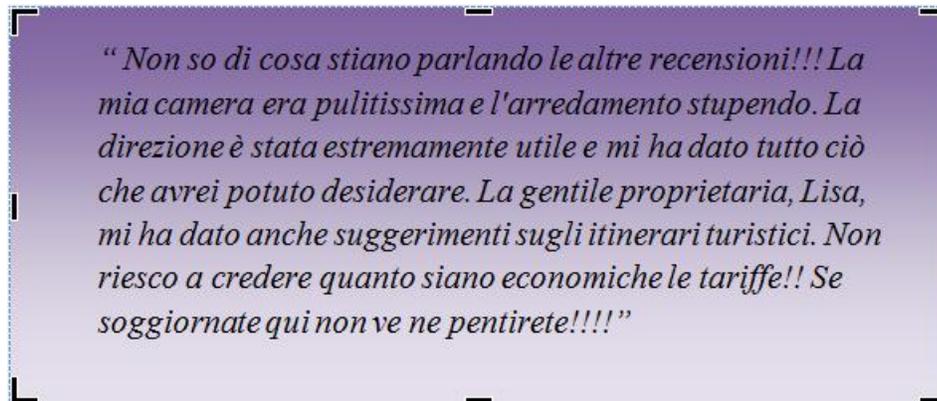
3. Molti albergatori cercano di danneggiare la reputazione della concorrenza, per cui si innesca una sorta di competizione che genera la creazione di recensioni false per mantenere alto il livello della propria attività.

Secondo alcune statistiche si stima che addirittura un terzo delle recensioni su TripAdvisor sono false e questo risulta essere dannoso per gli alberghi che stanno cercando sempre di più di tutelarsi, anche a seguito di ricatti provenienti dai turisti, e per questi ultimi che devono prenotare ritrovandosi di fronte a false recensioni positive.

Per i siti di recensioni, non solo TripAdvisor, esiste la cattiva abitudine da parte di alcune aziende di pagare gli utenti affinché questi vadano ad elargire parole positive nei confronti del proprio Brand. Oltre alle aziende, ci sono delle vere e proprie società organizzate che reclutano persone che svolgano questo “macabro lavoro”. Ad esempio Craigslist una vera e propria oasi di spam, dove le persone si organizzano per la costruzione di opinioni false. Dopo quanto affermato nel punto 2.2 ne mostreremo un esempio di recensione “falsa” estrapolata direttamente da TripAdvisor.

2.2 Un esempio di recensione falsa rilasciata dalle strutture alberghiere su “TripAdvisor”

Come accennato nell’introduzione e nel punto 2.1 sempre più la piattaforma “TripAdvisor” sta perdendo di credibilità per il rilascio di recensioni “false” sia da parte degli utenti che da parte delle strutture ricettive. Eccone un esempio di “ recensione falsa” rilasciata da quest’ultime spiegandone anche l’ingenuità.



“ Non so di cosa stiano parlando le altre recensioni!!! La mia camera era pulitissima e l'arredamento stupendo. La direzione è stata estremamente utile e mi ha dato tutto ciò che avrei potuto desiderare. La gentile proprietaria, Lisa, mi ha dato anche suggerimenti sugli itinerari turistici. Non riesco a credere quanto siano economiche le tariffe!! Se soggiornate qui non ve ne pentirete!!!!”

Figura 2.4: Primo esempio di "falsa" recensione.

Andando ad analizzare la Fig.2.4, si può notare che la prima frase ne è la prova cerca di quanto non sia attendibile poiché, premettendo che questa recensione è stata rilasciata in seguito a molte recensioni negative da parte di molti utenti, non si può opporre completamente alle altre arrivando ad affermare totalmente il contrario. E’ quasi impossibile che tutti affermano una cosa e una sola persona dichiara totalmente il contrario. Potrebbe non essere d’accordo in parte, ma

su tutto è improponibile. Inoltre molti sono i punti esclamativi per enfatizzare e solo l'albergatore vorrebbe arrivare a mettere in maniera così enfatizzata ciò di cui sta parlando, anche perché dopo diverse indagini su questo settore ne è venuto fuori che solo riguardo ai commenti negativi ci si presta a mettere molti punti esclamativi derivanti dall'insoddisfazione dell'utente magari adirato in quel momento; andando avanti, troviamo il nome della proprietaria dell'albergo ed è un segno per voler essere a tutti i costi credibili nel specificare un particolare non importante; quasi nessun turista è a conoscenza dei nomi dei proprietari o gestori dell'albergo almeno che non si tratti di clienti abituali; infine, l'ultima frase conferma quanto detto in quanto si vuole a tutti i costi vendere un qualcosa e chi se non il gestore dell'albergo?

Nei prossimi capitoli andremo a postare altri esempi con tre casi differenti ed analizzarli.

2.3 Proposta di risoluzione del problema alle aziende alberghiere

Nel Cap.1 abbiamo dato un'idea del mondo di TripAdvisor, la sua importanza e la sua pecca per l'avvento sempre più di recensioni false e il nostro scopo è quello di andarle ad individuare sul sito dimostrandone anche l'ingenuità con cui un albergatore, o chi per lui, le rilascia

postando un esempio nel Cap.1 e altri tre nel Cap.4² con le varie argomentazioni passo per passo per poi arrivare al nostro fine, dopo uno studio approfondito, ovvero proporre alle aziende alberghiere il miglior modo di rilasciare recensioni per far salire la propria reputazione senza essere indirizzabili o scoperti. Ma prima di passare a questo è bene argomentare la parte che precede il rilascio di recensioni false. Come vengono definiti gli albergatori o chi per loro rilascia recensioni false su TripAdvisor? Perché si apprestano a rilasciare queste recensioni false e quando decidono di farlo? Ecco che nel capitolo III introdurremo gli Opinion Leaders i quali, nel nostro caso di ricerca, assumono il ruolo di albergatori cioè chi vuole veicolare tutte le altre recensioni dato la tanta negatività ricevuta e la posizione molto bassa su TripAdvisor³. Andando a far comprendere meglio, gli albergatori su TripAdvisor indossano le vesti di Opinion Leaders, o tentano almeno di farlo e noi proporremo il miglior modo per divenirlo, dato le tante recensioni negative ricevute che hanno fatto sì, a parte dare una brutta immagine dell'albergo, che quest'ultimo fosse classificato in basse posizioni su TripAdvisor. Questi albergatori decidono, inoltre, di affrettarsi a rilasciare queste recensioni dopo aver analizzato le varie opinioni ricevute cioè attraverso l' "opinion mining automatico and sentiment analysis", un'analisi quantitativa e qualitativa delle opinioni computerizzata con

²La nostra ricerca si è basata sul ritrovamento di più di 50 casi di recensioni false se pur postandone solo 3 nel nostro elaborato tesi.

³La posizione di un albergo su TripAdvisor non dipende dalle stelle ricevute ma dall'andamento positivo o negativo proveniente dalle recensioni degli utenti.

le varie classificazione che permette di dare in poco tempo alle strutture alberghiere, nel nostro caso, l'andamento della struttura stessa classificando e categorizzando le recensioni in positive o negative. Gli albergatori che vogliono diventare a tutti i costi degli Opinion Leaders sono coloro che in seguito a questa analisi automatica delle recensioni, ne riceve un andamento prettamente negativo e tenta di risolvere questo problema rilasciando appunto false recensioni. Ma molti albergatori non sono in grado di farlo per l'ingenuità e perdono subito di credibilità e il nostro scopo è insegnare loro a rilasciare su TripAdvisor una recensione che, pur non essendo veritiera, riesca a far salire la reputazione della struttura, la classifica sul sito TripAdvisor e a veicolare e influenzare le altre opinioni.

3.1 Teorie sugli Opinion Leaders

Al giorno d'oggi, l'attenzione si sta spostando dall'analisi del singolo utente/azienda verso l'analisi di gruppi di utenti e sulle interconnessioni che hanno tra loro. La persona adesso viene analizzata in correlazione alla rete sociale che si è creata e dentro cui opera. Infatti le stesse connessioni che una persona stabilisce nella vita di tutti i giorni, basate sulla sua indole, sul proprio carattere e sulla gestione dei rapporti personali, vengono replicate anche nel mondo delle reti sociali su Internet. La grossa differenza è che sul Web tutte queste informazioni sono mappabili, convertibili in bit e poi misurabili.

Diventa quindi possibile catalogare gli utenti in base ai loro comportamenti. Si possono così distinguere i cosiddetti opinion leaders, personalità in grado di guidare gli altri con personalità non dotate di iniziativa che spesso emulano gli altri. L'individuazione di un lea-

der in una rete comportamentale ha trovato grande applicazione nel marketing, con l'obiettivo di analizzare il passaparola delle persone e l'impatto che esso poteva avere nello sviluppo di nuovi trend sul mercato. Quindi potremmo definire gli opinion leaders come coloro che hanno una grande capacità di veicolare i loro messaggi attraverso diversi mezzi di comunicazione tra cui e-mail, newsgroup, bacheche elettroniche e così via. Risultano essere estremamente influenti online, ovviamente, se si tratta di specialisti che diventano presto fonti attendibili e facilmente consultabili(Quart, 2003).

In tutti questi casi gli OL sono un forte punto d'interesse per il resto del mercato, che vede in loro delle guide con cui confrontarsi, per regolare i propri rapporti sociali ed ottenere informazioni utili.

Le opinioni del 10% dei consumatori possono guidare emotivamente i giudizi del rimanente 90% della popolazione, che si fida del parere degli Opinion Leader e ne segue l'esempio. Individuando le giuste persone influenti è quindi possibile trovare un importante veicolo di contagio per il messaggio dell'azienda: questi personaggi diffondono spontaneamente ogni prodotto di successo, parlandone positivamente con gli amici o semplicemente utilizzandolo personalmente, convincendo la propria rete sociale della qualità dell'offerta. I leader d'opinione ottengono la fiducia e l'ammirazione degli altri grazie alla loro cultura in merito all'argomento specifico di competenza.

È possibile ritrovare i concetti fondamentali di questo discorso nella

classica teoria di marketing denominata come “two step flow of communication”. Questo principio afferma che ogni messaggio diffuso dai mass media raggiunge in primo luogo gli Opinion Leader, che in seguito passano le informazioni alla propria rete d’influenza sociale (Elihu Katz, 2010).

La strategia del Product Opinion Leader Placement tenta però di pubblicizzare i brand scavalcando ogni tipo di media (TV, Radio, Cinema, Giornali, Videogiochi...), per raggiungere direttamente gli Opinion Leader inserendo il prodotto nelle loro vite.

Il brand posiziona la merce nelle mani di un opinion leader, per stimolare l’attenzione del rispettivo gruppo sociale. L’uso del prodotto e gli eventuali apprezzamenti da parte di questi personaggi influenti, suggestionano positivamente il resto del mercato, promuovendo al meglio qualsiasi tipo d’offerta. Dopo aver definito nel miglior modo chi sono gli opinion leader, possiamo affermare che sono nella piattaforma TripAdvisor importanti poiché il loro tentativo è quello di far salire la reputazione di un albergo da 1 a 5 stelle in poche ore, rilasciando finte recensioni tentando di veicolare altri messaggi. In sintesi (vedere Fig.3.1):

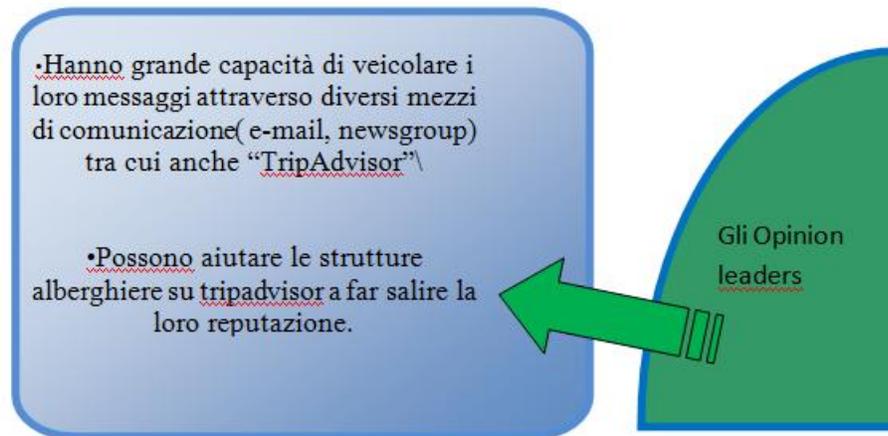


Figura 3.1: Gli Opinion Leaders.

3.2 Opinion Mining automatico and Sentiment Analysis

Ogni giorno in tutto il mondo le persone si scambiano miliardi di opinioni e di pensieri sotto forma di messaggi, ma anche di contenuti audiovisivi, aumentando a dismisura il flusso di dati che corre sulle autostrade informatiche.

Queste informazioni sono oggetto di studio e di ricerca, per diverse finalità, sia scientifiche, sia di mercato, ponendo ai regolatori politici delle scelte da compiere, ma anche la consapevolezza che in tutto questo universo di bit in movimento si nascondono delle grandi opportunità di sviluppo..

L'interesse che i singoli utenti mostrano per le opinioni online su

prodotti e servizi, e la tale potenziale influenza che le opinioni esercitano appunto, è un motivo per cui i venditori di questi oggetti stanno prestando sempre più attenzione.

Infatti le grandi aziende, oltre alle esigenze di pubblicizzare i propri prodotti, hanno anche la necessità di tenere sotto controllo la reputazione del loro brand. Più cresce la dimensione dell'azienda più la sua reputazione deve essere misurata su un insieme di utenti sempre più grande, partendo dall'artigiano che si preoccupa dell'opinione dei clienti abituali fino a raggiungere le dimensioni delle multinazionali che devono mantenere una reputazione a livello globale.

Più cresce il numero di utenti da monitorare, più si rende necessario sviluppare dei sistemi automatici che calcolano il sentimento degli utenti e forniscono degli indicatori ai manager che si occupano di marketing.

Quindi le grandi aziende stanno sempre arrivando di più a capire, che queste voci dei consumatori possono esercitare un'influenza enorme nella formazione delle opinioni di altri consumatori, in ultima analisi, la loro lealtà di marca, le loro decisioni di acquisto, e la loro difesa propria del marchio. Ora infatti le aziende possono rispondere alle intuizioni del consumatore che essi generano attraverso il monitoraggio dei social media e dell'analisi modificando i loro messaggi di commercializzazione, posizionamento del brand, sviluppo del prodotto, e altre attività di conseguenza.

Con gli strumenti attuali siamo oggi in grado di monitorare queste reti sociali attraverso i metodi di analisi dei dati empirici che sono oggi in circolazione, anche conosciute come “opinion mining”, cioè l’analisi quantitativa e qualitativa delle percezioni degli utenti e dei consumatori, a partire da grandi quantità di dati e grazie all’utilizzo di modelli matematici, che potranno poi avere un impiego industriale o culturale. Sentiment Analysis, Sentiment Classification, Opinion Extraction sono altri nomi usati per identificare questa disciplina in letteratura.

Il primo livello di misurazione è di tipo quantitativo, il passo successivo è capire i toni delle conversazioni. Se ne parla bene o male? È in questo caso che entrano in azione le tecniche di sentiment analysis (B. Pang and L.Lee,2008).

Le aziende che lavorano sul web e sulle reti sociali stanno cercando il modo più veloce per arrivare a risultati concreti su cui ragionare, perché il marketing relazionale fino ad ora ha dato solo delle indicazioni di massima, delle suggestioni più che delle vere e proprie best practices su cui poter contare. Non è semplice raccogliere milioni di dati ogni giorno ed analizzarli in tempi commercialmente sfruttabili poiché parliamo di 50 milioni di opinioni per utente al giorno solo in Italia e di cui si sono riusciti a raccoglierne ben 29 milioni.

La Sentiment Analysis è la capacità di decifrare le opinioni contenute in un testo scritto o parlato e da la possibilità di estrarre, tramite

i computer, queste opinioni da qualunque Social Media in modo automatico utilizzando elementi di machine learning come l'analisi latente semantica. E' un nuovo approccio che determina l'atteggiamento di un oratore o uno scrittore rispetto a qualche argomento, ed il compito fondamentale è la classificazione della polarità di un'espressione lessicale, la quale può avere polarità neutra, positiva o negativa. Con questo sistema, si è in grado di identificare automaticamente la polarità delle espressioni e quindi si possono prendere delle misure di gestione della conversazione in Rete che sono dette "social customer relationship management", un'evoluzione del "customer relationship management" utilizzato da anni per la gestione del cliente (N. Jindal and B. Liu,2008).

Il problema di questo nuovo studio è che i fattori culturali, linguistici e le sfumature nei differenti contesti rendono estremamente difficile la trasformazione di una stringa di testo scritto in un semplice sentimento pro o contro.

Tutto questo processo può essere semplice come l'apprendimento di una lista di parole positive e negative, o complicato come condurre in profondità l'analisi dei dati al fine di comprendere la grammatica e la struttura della frase usata (N. Jindal and B. Liu,2007).

Le analisi vengono principalmente condotte mediante tecniche proprie da altre discipline quali l' Information Retrieval e il Natural Language Processing.

Quello che interessa al fine di questa analisi è rilevare il sentimento con cui si esprime un’opinione , se questa è positiva , negativa o obiettiva e quanto l’opinione positiva o negativa sia forte (Debolmente, Fortemente, Mediamente). Ciò lo si può ottenere attraverso diversi metodi tra cui per esempio “SentiWordNet”.

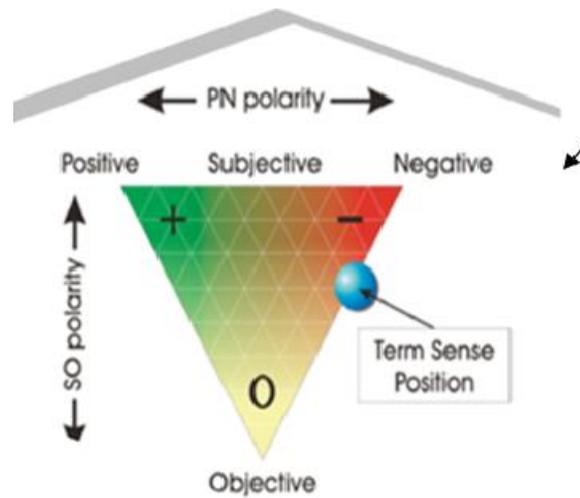


Figura 3.2: SentiWordNet.

Una volta pronto il set di dati che interessano ,quindi eliminando da un testo HTML i caratteri non necessari all’analisi, si possono applicare le procedure tipiche del Data Mining, quali la classificazione, la clustering o l’analisi delle associazioni. Le teorie dell’ “opinion mining” sono state oggetto di studio facendo sempre riferimento alla piattaforma “tripadvisor” poiché per le aziende, nel nostro caso le aziende alberghiere, da un punto di vista scientifico e applicativo, è molto interessante poter analizzare le opinioni e misurare la propria

reputazione in poco tempo (A.Esuli and F.Sebastiani,2006). In sintesi (vedere Fig.3.3):



Figura 3.3: Opinion Mining automatico.

3.3 Classificazione automatica

La Sentiment Analysis, come spiegato nel capitolo precedente, riguarda la possibilità di reperire e trattare computazionalmente le opinioni e i sentimenti espressi dalle persone sui media digitali (ad esempio blog) a proposito di determinati argomenti. Le opinioni, una volta reperiti i documenti nelle quali sono contenute, possono essere classificate in due o più classi, ciascuna classe indicante l'intensità o la polarità (positiva o negativa) dei sentimenti espressi, e una volta aggregate possono concorrere a fornire il livello di un gradimento nella rete di un film o prodotto, e non solo. La precisa classificazione delle opinioni è stata

fin dal principio uno dei principali argomenti di discussione in ambito accademico. A tal fine la Sentiment Analysis viene trattata come un problema di classificazione di testi secondo due livelli di dettaglio. Il primo livello di classificazione consiste nell'assegnare una categoria positiva o negativa a un testo che esprime opinioni, la cosiddetta classificazione a livello di documento. Il secondo livello consiste invece nella classificazione di frasi dapprima in soggettive e oggettive e poi, per le soggettive, in positive, negative o neutrali. Si parla in questo caso di classificazione a livello di frase.

La classificazione a livello di documento o di frase fornisce un'indicazione di massima sul giudizio degli utenti su un particolare aspetto di interesse e non riesce ad esprimere informazioni dettagliate. Una frase con polarità negativa non è necessariamente indice di una visione negativa globale. Ad esempio sebbene la frase “ la qualità dello schermo del telefono non è buona” abbia una polarità negativa, questa non dovrebbe essere associata con il cellulare, ma solamente con lo schermo di questo. Il problema può essere scomposto in due frasi: identificare l'oggetto del discorso (qualità dello schermo) e determinare la polarità dell'opinione su quella caratteristica. Inoltre una frase comparativa esprime una relazione tra uno o più soggetti, generalmente con aggettivi o avverbi comparativi e superlativi. Necessitano di essere trattate diversamente rispetto alle altre frasi in quanto la loro semantica e la loro forma sintattica sono estremamente differenti.

Prima di effettuare la classificazione dei “sentimenti”, dobbiamo individuarli all’interno dei tweets, che in genere sono emozioni oppure giudizi o idee, dei quali tramite un’accurata valutazione si verifica se c’è la presenza di qualcosa di negativo o positivo, prestando particolare attenzione anche ai piccoli particolari del lessico. Lo studio delle emozioni può essere svolto attraverso la verifica d’influenza delle emozioni di colui che scrive tramite la scelta delle parole scelte e/o degli elementi linguistici e/o dall’impostazione del testo (Pang, B., Ledd, L. and Vaithyanathan, S. Thumbs up, 2002).

Oltre a questi due modi per effettuare l’analisi, ce ne sono altri ma che non fanno strettamente parte di quelli sopra citati, e sono, per esempio, l’uso dell’ironia e quello di un linguaggio figurativo. Però è opportuno sottolineare che le maggiori tecniche di Sentiment Analysis non si basano su emozioni reali ma sulla valutazione, apprezzamento o giudizio perché si concentrano su recensioni fatte su film, prodotti, automobili, dove le valutazioni sono prive di “sentimentalismi” proprio perché sostanzialmente l’utente che ha una conoscenza dell’oggetto preso in esame, da una propria valutazione in maniera distaccata, dando ai semplici lettori e/o a coloro che sono alla ricerca di confronti tra prodotti nella fase di pre acquisto l’opportunità di farsi una propria idea del medesimo oggetto.

Importante al fine della classificazione è il vocabolario anche se il maggior problema spesso si concentra proprio sulla creazione di

quest'ultimo, il quale è un elenco di termini classificati secondo il sentimento positivo o negativo che esprimono.

In questo lavoro i dati sono strutturati in un data frame composto da sei colonne dove si ha nella prima la distinzione a seconda se la parola è debolmente soggettiva o fortemente soggettiva.

Nella seconda colonna sono presenti le parole del dizionario inglese, mentre nella terza le corrispondenti e svariate traduzioni in italiano fino ad arrivare alle penultima dove vi è il tipo di parola presente cioè se è un aggettivo, nome, verbo, ed infine nell'ultima si ha la corrispettiva polarità di ogni singola parola (Positive, Neutral, Negative e Negation, dove queste ultime sono le negazioni forti che invertono la polarità o la intensificano).

Il problema principale è l'assegnare la polarità alle parole perché alcune di essere possono essere tutte e tre i "sentimenti" (positive, negative o neutral), a seconda della loro sistemazione nella frase (Bing Liu, 2011).

Il problema dei post è che gli utenti li scrivono aggiungendo faccine attraverso le quali vengono espressi dei sentimenti e queste non appartenendo al vocabolario devono essere studiate in qualche maniera, e per poter effettuare lo studio al meglio è stata ristrutturata la funzione "pulisci tweets". Inoltre un problema generale è che la classificazione è costosa. Ad esempio l'analisi delle co-occorrenze di aggettivi e verbi al fine di stabilirne un orientamento concorde o dissimile è di fatto in-

trattabile, in quanto correrebbe addestrare un classificatore contenente centinaia di migliaia di features (ChengXiangZhai, 2009). Un'alternativa più efficiente alla classificazione completa è di considerare per ogni documento solo quei termini che coincidono con quelli utilizzati nella interrogazione formulata dall'utente e cercare di stabilire se essi cadono nel raggio di azione di aggettivi e verbi che esprimono sentimenti. In generale, per accelerare il reperimento e la classificazione di opinioni, si devono adottare tecniche di riduzione della dimensione del dizionario che permettono di pesare e ordinare efficacemente i documenti per contenuto e qualità di opinione. Così, quindi, attraverso la classificazione automatica si permette di classificare e organizzare le opinioni secondo criteri predefiniti in positive e negative, e una volta che queste vengono aggregate, si può concorrere a far capire il livello di gradimento alle aziende alberghiere.

3.4 La categorizzazione

La creazione di categorie rende equivalenti cose diverse, consente di raggruppare gli oggetti e gli eventi in classi e di rispondere ad essi in funzione della loro appartenenza ad una data classe piuttosto che della loro unicità.

Il libro *A study of thinking* di Bruner¹, Goodnow² e Austin (1956) (*Jerome Seymour Bruner, 1986*) costituisce una pietra miliare nello studio dei processi cognitivi. In questo libro sono riportati una serie di esperimenti sull'acquisizione di concetti.

Tutti gli individui fanno uso di concetti. Raramente gli eventi vengono concepiti nella loro unicità. Piuttosto, ciascun evento viene percepito come il rappresentante di un categoria. Nei loro esperimenti Bruner et al. hanno usato come stimoli le carte riprodotte nella Fig.3.4.

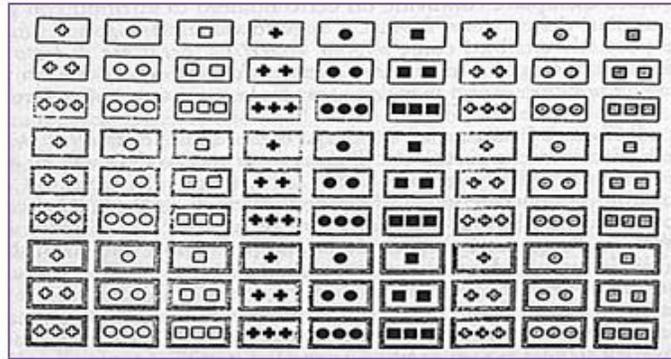


Figura 3.4: Esperimenti di Bruner.

Se una di queste carte costituisce un esempio di un dato concetto, essa viene detta un caso positivo. Un semplice concetto congiuntivo, per esempio, può essere definito in base all'attributo nero e all'attributo quadrato. Ciascuna carta con questi attributi è un caso positivo del concetto.

¹Jerome Seymour Bruner (New York, 1 ottobre 1915) è uno psicologo statunitense che ha contribuito allo sviluppo della psicologia cognitiva e la psicologia culturale nel campo della psicologia dell'educazione.

²Jacqueline Jarrett Goodnow (nato nel 1924) "ha avuto una lunga e brillante carriera.

Questo significa che tutte le carte nella sesta colonna della figura sono dei casi positivi di questo concetto. Se una carta non contiene gli attributi giusti allora essa viene detta caso negativo. Per esempio, la carta nella prima riga della prima colonna della figura è un caso negativo. In questo esempio, il numero delle figure su una carta è irrilevante.

Lo scopo della categorizzazione del testo è la classificazione dei documenti, nel caso in esame post, in un numero fissato e predefinito di categorie. In generale ogni documento può appartenere ad una categoria, a più categorie o a nessuna categoria. Il problema della categorizzazione viene generalmente visto come un caso particolare del problema più generale dell'apprendimento, ovvero sull'idea che l'esperienza possa migliorare la capacità di un agente automatico di agire in futuro. In letteratura, il classificatore Naïve Bayes³ è uno dei metodi più usati per la classificazione del testo. Tuttavia non è l'unico approccio, infatti, recentemente è emerso un differente orientamento che fa riferimento alle Support Vector Machine (SVM)⁴, oppure è stata impiegata una tecnica più vecchia chiamata TF-IDF3.

Una grossa barriera all'adozione di questi algoritmi è quella di richiedere un elevato numero, spesso proibitivo, di dati etichettati per

³Un classificatore bayesiano semplificato con un modello di probabilità sottostante che fa l'ipotesi di indipendenza delle feature, ovvero assume che la presenza o l'assenza di una particolare feature in un documento testuale non è correlata alla presenza o assenza di altre feature.

⁴Le macchine a vettori di supporto (SVM, dall'inglese Support Vector Machines), o macchine kernel, sono un insieme di metodi di apprendimento supervisionato per la regressione e la classificazione di pattern.

poter essere addestrati ed ottenere risultati accurati. L'etichettatura dei dati è un compito che deve essere tipicamente eseguito da un essere umano, risultando quindi un processo lungo e costoso. Il bisogno di grandi quantità di dati per ottenere una buona accuratezza e la difficoltà di ottenere dei dati etichettati ha portato, durante lo sviluppo del componente da inserire nell'architettura dello strumento di Sentiment Analysis, alla ricerca di un metodo alternativo che non richiedesse dati già etichettati. Inoltre, le categorie sono definite da una serie di attributi, tutti necessari e sufficienti per definire l'appartenenza di un elemento alla categoria. Nella categorizzazione i modelli di rappresentazione categoriale in psicologia sociale cognitiva sono:

- il modello del prototipo;
- il modello degli esemplari;
- il modello associativo.

Nel modello per prototipi le categorie non sono definite da criteri necessari e sufficienti ma presentano dei margini sfuocati e non sempre è chiaro se un esemplare appartenga ad una o ad un'altra categoria. I membri di una categoria hanno diverso status ovvero i membri più tipici sono i più rappresentativi della categoria, membri meno tipici sono periferici.

e.g. categoria: uccelli = esemplari più tipici
--

passeri = esemplari meno tipici (pinguini)

Il membro che meglio rappresenta le informazioni categoriali è detto prototipo, ossia è una rappresentazione astratta dell'esemplare che meglio personifica la rappresentazione categoriale.

Il prototipo svolge un ruolo fondamentale nel processo di classificazione di un oggetto ed il criterio di somiglianza oggetto-prototipo è alla base dell'inclusione-esclusione categoriale. Per quanto riguarda, invece, il modello per esemplari le categorie sono costituite da un insieme di esemplari di cui l'osservatore ha avuto nel passato un'esperienza (diretta vs indiretta).

Ogni qualvolta incontriamo un esemplare e lo dobbiamo classificare, verifichiamo in che misura questi richiama o ricorda una serie di esemplari che appartengono ad una categoria(Zanzotto, Fabio Massimo, Marco Pennacchiotti, and Kostas Tsiou-tsioulis. 2011)

e.g. Vedi passeggiare sulla spiaggia di Rimini un ragazzo con sandali e calze bianche, la sua carnagione è bianca ed ha i capelli biondi. Questo stimolo ti attiva un set di esemplari simili che sono riuniti assieme nella categoria dei turisti tedeschi.

Le valutazioni a proposito di persone sconosciute sono legate a delle somiglianze percepite tra la persona sconosciuta e gli esemplari catego-

riali e tale somiglianza è spesso basata su elementi irrilevanti o criteri del tutto ininfluenti.

Infine vi è il modello associativo nel quale le categorie sono organizzate in memoria sotto forma di noti concettuali e i nodi sono associati ad alcuni attributi e ad altri nodi.

La categorizzazione è analoga alla classificazione e l'analisi o sentiment analysis è di tipo strutturato ovvero suddivisa in classi di indagini(le recensioni) che rappresentano la realtà che si intende misurare. Il sistema, così, automaticamente si occupa di cercare 24 ore su 24 l'emergere di nuovi commenti o recensioni e di categorizzarli in una delle classi predefinite. Ogni categoria è associata ad una lista di concetti rilevanti che verranno ricercati e analizzati su tutto il web (nel nostro caso TripAdvisor).

La categorizzazione, dunque, serve per classificare automaticamente i documenti contestualizzandoli a un determinato dominio (es. cinema, musica, sport, etc.).

Grazie agli algoritmi di apprendimento automatico utilizzati per la classificazione, il Categorizzatore adatta dinamicamente le categorie in base alle esigenze del cliente(Jerome Seymour Bruner, Jacqueline J. Goodnow, George A. Austin, 1986).

Il modulo di categorizzazione permette di velocizzare il recupero delle informazioni, grazie alla possibilità di assegnare automaticamente e con alta precisione una o più categorie in funzione delle tematiche

trattate o tramite criteri personalizzabili. In sintesi (vedere Fig.3.5).

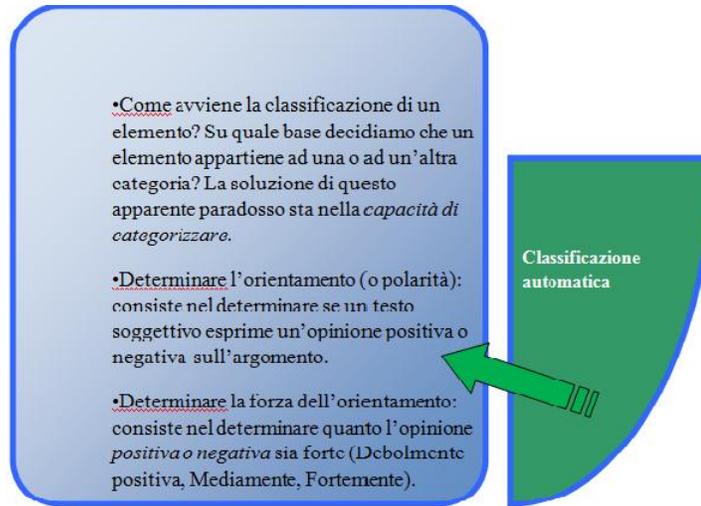


Figura 3.5: Classificazione automatica.

Raccolta ed analisi delle recensioni

4.1 Metodi utilizzati al fine di ricercare le recensioni "false" su TripAdvisor

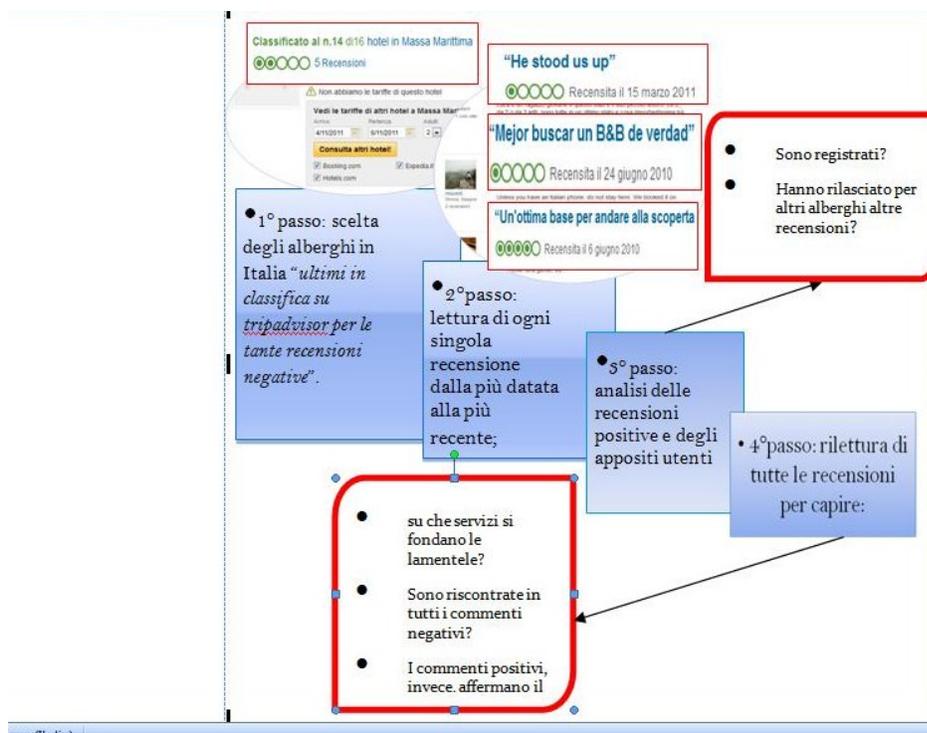


Figura 4.1: Metodi di ricerca di recensioni "false".

Come si può notare nella Fig.4.1, il primo passo al fine di individuare le recensioni ritenute “false” è stato quello di navigare appunto su “ www.Tripadvisor.com”.

Qui ci siamo soffermati su tutti gli alberghi che il sito mette a disposizione a partire dagli alberghi con le mete di viaggio più popolari a centro o nord Italia per finire con gli alberghi con le mete principali del turismo e degli affari.

Per questa analisi abbiamo preso in considerazione soltanto gli alberghi presenti in Italia anche se il sito comprende alberghi di tutta Europa.

Una volta scelta la città e cliccatoci su, il passo successivo è stato quello di cercare soprattutto quegli alberghi, ultimi in classifica su tripadvisor dato le recensioni negative ricevute, indipendentemente dal numero di stelle di un albergo. Questo perché il nostro intento è di andare a tastare e verificare la veridicità delle poche recensioni positive di un albergo laddove ve ne siano molte negative.

Successivamente ci siamo soffermati nella lettura di ogni singola recensione dalla più datata alla più recente e su un’analisi approfondita qualora captassimo inquadramenti non veritieri.

In tal caso per affermare le nostre ipotesi o le nostre supposizioni, siamo andati più a fondo nella ricerca soprattutto per gli utenti che rilasciavano opinioni positive andando a verificare, in primis, se avevano schede che li identificasse con foto e altri dati e se erano utenti

che avevano già in precedenza contribuito a dare recensioni su altri alberghi o se per loro, invece, era la prima volta. Un dato molto importante perché se un utente ha rilasciato molte recensioni per alberghi diversi è più attendibile rispetto ad uno che è la prima volta che ne scrive una, in quanto dimostra che è solito a viaggiare e in realtà il suo è stato un commento oggettivo poiché a differenza di tutti gli altri utenti, ha trovato l'albergo ottimale. Ma quanto detto non basta per essere sicuri che sia una falsa recensione. Infatti il successivo passo per accertarci è stato quello di leggere attentamente tutti i commenti, positivi e negativi, capire su che servizi si fondavano le lamentele, se quest'ultime erano state riscontrate in tutti i commenti negativi e se quei commenti positivi, invece, affermavano totalmente il contrario.

Inoltre, non di grandissima rilevanza, ma comunque utile al fine di essere ancora più sicuri sui nostri dubbi, abbiamo tenuto conto dei “voti utili” assegnati a ciascuna recensione, e per voti utili s'intende il numero di persone che mettono un semplice “mi piace” qualora fossero d'accordo con la recensione appena letta.

Esempio: una recensione con 7 voti utili vuol dire che a 7 persone tale recensione è piaciuta. Quindi si può affermare che se una recensione positiva ne aveva 0 di voti utili non poteva essere molto credibile in quanto nessuno affermava tale tesi.

Per finire, ci siamo anche basati sulle date nelle quali veniva rilascia-

ta una recensione per diversi motivi, per esempio sul tempo trascorso tra una recensione falsa e una negativa. Se nel frattempo era passato un anno, non lo abbiamo preso in considerazione perché la gestione poteva essere cambiata e i servizi migliorati.

Dopo tutte queste considerazioni siamo arrivati a individuare le recensioni falsi le quali porremo con degli esempi nel capitolo successivo. In particolare abbiamo individuato tre tipi di casi differenti.

4.2 Metodi di raccolta delle recensioni “false”

- Al fine del nostro scopo, la nostra ricerca ha avuto come obiettivo quello di ricercare e raccogliere 50 casi in cui vi erano recensioni ritenute false.
- Su una pagina word (vedere Fig.4.2) abbiamo salvato a parte ogni caso trovato inserendo in ordine di datazione sia le recensioni false sia le recensioni vere ovvero i vari commenti negativi. Sono state inserite anche recensioni in lingua inglese in ultime sulla pagina word.
- Dei 50 casi trovati, per porre come esempio qui sull’elaborato tesi, abbiamo scelto solamente tre casi che mostreremo e analizzeremo nel capitolo a seguire. Di questi tre casi, ognuno si presenta come un caso individuale e completamente differente, per le sue caratteristiche, dagli altri due casi.

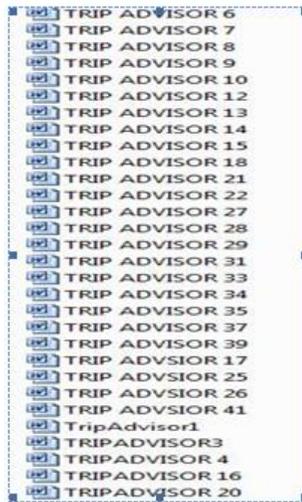


Figura 4.2: Unione recensioni su Word.

1° caso:

- recensione *falsa* proveniente da utenti diversi nella registrazione su tripadvisor, ma facenti parte della stessa persona nonché il gestore.

2° caso:

- recensione *falsa* proveniente sempre dallo stesso utente riconosciuto poiché registrato con lo stesso nome.

• "Fly68 ha 1 recensioni in più"

3° caso:

- recensione *falsa* proveniente da uno stesso utente ma non registrato

• "un viaggiatore di Venere.com"

Figura 4.3: I tre casi di recensione "falsa".

Il primo caso, come si può notare dalla Fig.4.3, riguarda una recensione falsa proveniente da diversi utenti nella registrazione su TripAdvisor, ma date le caratteristiche nonché le varie espressioni

uguali riscontrate in ogni recensione se non identiche se pur da utenti differenti, abbiamo potuto affermare che eravamo di fronte ad un caso di recensione non veritiera.

Il secondo caso, invece, riguarda uno stesso utente poiché registrato e rintracciabile, che in anni differenti, in seguito a tanti commenti negativi, ha rilasciato recensioni positive per lo stesso albergo.

Per il terzo caso siamo di fronte ad un utente non registrato se non con il sole nominativo “un viaggiatore di Venere.com” il quale in maniera successiva e anche con il ripetersi nello stesso giorno, rilasciava recensioni positive con le stesse e identiche espressioni poiché questo albergo riceveva solo ed esclusivamente commenti negativi.

- Infine, dei tre casi posti come esempio, abbiamo messo in evidenza, per far comprendere quello in cui al punto precedente abbiamo affermato con dimostrazioni e varie argomentazioni, solo le recensioni false rilasciando, ovvero non riportando, i vari commenti negativi poiché ricordiamo che le recensioni false si sono riscontrate solo per quegli alberghi con maggiori recensioni negative.

4.3 Raccolta dei tre casi di recensioni “false” accorte su TripAdvisor

4.3.1 Primo caso di recensione “falsa”



Figura 4.4: Primo gestore, primo caso.



Figura 4.5: Secondo gestore, primo caso.

Come già accennato (vedi Par.4.2), questo primo caso riguarda utenti diversi nella registrazione, ma per le caratteristiche che andremo a

spiegare subito dopo le Fig.4.4 e 4.5, risultano appartenere allo stesso utente nonché il gestore dell'albergo.

Nelle Fig.4.4 e 4.5 abbiamo evidenziato le espressioni quasi identiche e andando ad analizzare possiamo vedere che inizialmente si tende a specificare con chi si è andato in vacanza (caratteristica tipica che avviene nelle recensioni non vere); si ripete il nome dell'albergo nonché "borgo pan dei mucini" (altra caratteristica tipica) ; in entrambi le recensioni troviamo questa affermazione: " le camere dove abbiamo alloggiato/siamo stati erano pulite, efficienti e con tutti i confort; si fa riferimento alla posizione dell'albergo ritenendola fantastica; altre identiche espressioni: buonissimo/ottimo staff di animazione, torneremo volentieri, complimenti, le piscine molto pulite e curate da personale serio/ personale professionale, la cucina molto buona/ottima. Inoltre, come si è evidenziato, alla fine della recensione si è assegnato un voto 5 su 5 per tutte le caratteristiche dell'albergo. Infine, nessun utente ha ritenuto utile, ovvero è risultato d'accordo con la recensione sottoscritta se non solo nella prima recensione (ma non rilevante in ogni caso poiché chiunque legge la recensione può cliccare su "mi piace").

Altra caratteristica, inoltre che può confermare la nostra tesi è la data del rilascio di recensioni. Notando le Fig.4.4 e 4.5 si nota che sono state rilasciate con un giorno solo di distacco l'una dall'altra e si vuole ricordare che sono state le uniche recensioni positive dopo una lunga serie di commenti negativi.

4.3.2 Secondo caso di recensione “falsa”

Questo è il secondo caso, come accennato nel Par.4.2, di recensione “falsa” in cui vi sono due recensioni rilasciate da uno stesso utente registrato. Riportando l’esempio del punto 4.2, si noterà nella Fig.4.6 “Fly68 ha 1 recensioni in più”.



Figura 4.6: Primo gestore, secondo caso.

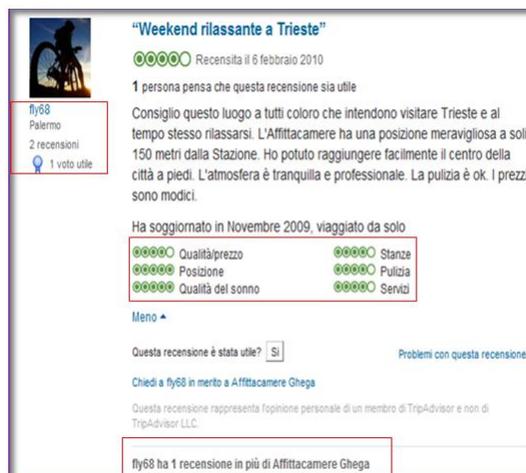


Figura 4.7: Secondo gestore, secondo caso.

In questo caso non possiamo affermare la presenza di espressioni identiche come nelle Fig.4.4 e 4.5. Diverse sono le caratteristiche che si vanno a citare in entrambi le recensioni, ma come abbiamo evidenziato questo utente registrato ha rilasciato per lo stesso albergo due recensioni senza neanche averlo citato. Può capitare che un utente vada nello stesso albergo in anni diversi e rilasci anche successivamente alla vacanza recensioni. Abbiamo, nella nostra ricerca, individuato casi come quello appena citato, ma ricordiamo che questa recensione era per questo albergo l'unica ad essere positiva. La sola ipotesi, dopo una serie di analisi in questo caso, è che l'utente era inesperto della piattaforma TripAdvisor e non era quindi a conoscenza che sotto la recensione risultava scritto(vedi Fig.4.6) che aveva rilasciato altre recensioni e si è limitato a scrivere le recensioni cambiando espressioni e anche i voti sulle caratteristiche dell'albergo.

4.3.3 Terzo caso di recensione “falsa”

Questo è il terzo caso di recensioni falsa, come accennato nel punto 4.3, la quale ha come caratteristica la presenza di utenti non registrati ma aventi il solo nominativo “ un viaggiatore di venire, com.” ma facenti parte della stessa persona nonché il gestore date le stesse ed identiche espressioni che andremo nelle Fig.4.8 e 4.9 ad analizzare.

“Tornerò in questo B&B...”

★★★★★ Recensita il 6 dicembre 2010

Un viaggiatore di Venere.com nullnullitaly

Tornerò in questo B&B. Vicinissimo alla Metrò e a San Pietro, interessantissima la visita a vicinissimi Musei Vatican, ma andateci presto che altrimenti non riuscite a vederli tutti. E poi c'è un mercato rionale dove c'è di tutto dalla frutta, al forno con pizza e panini caldi, SQUISITI!, ai vestiti. Grazie Antonio per i preziosi suggerimenti su dove andare e cosa vedere.

Consiglio l'Albatros B&B a tutti coloro che vogliono trascorrere una vacanza nel cuore di Roma

★★★★★ Qualità/prezzo ★★★★★ Stanze
 ★★★★★ Pulizia
 ★★★★★ Servizi

Figura 4.8: Primo gestore, terzo caso.

“OTTIMO !! Ci siamo trovati...”

★★★★★ Recensita il 20 marzo 2011

Un viaggiatore di Venere.com nullnullitaly

OTTIMO !! Ci siamo trovati veramente bene. Roma è bellissima e da qui potevamo girarla facilmente, i Musei Vatican e San Pietro a un paio di minuti a piedi, poi con la metropolitana a soli 200 metri raggiungevamo le mete più lontane come la Galleria Borghese o il Quartiere dell'EUR. Veramente la disponibilità e la discrezione degli addetti merita una lode. Non ci è mancato nulla. Antonio gentilissimo, ci ha regalato anche il biglietto della metro e una mappa, è tanti preziosi consigli, tipo cosa vedere e dove mangiar bene e spender poco (un'ottima cena a 14 euro !!). GRAZIE ANTONIO

TORNEREMO AL B&B ALBATROS e lo consigliamo a tutti coloro che cercano un angolo a Roma Centro Storico dove trascorrere le vacanze alla scoperta di questa affascinante città. Ciao a tutti

★★★★★ Qualità/prezzo ★★★★★ Stanze
 ★★★★★ Pulizia
 ★★★★★ Servizi

Figura 4.9: Secondo gestore, terzo caso.

In questo caso quindi, come si può notare nelle Fig.4.8 e 4.9, siamo di fronte a due recensioni con utenti non registrati, ma lo stesso nominativo e espressioni identiche. Andando ad analizzare notiamo che si specifica che l'albergo è situato

vicino ai Musei Vaticani, si ripete il nome del proprietario dell'albergo e andando a ringraziare quest'ultimo, si utilizza la stessa espressione nonché: "grazie Antonio/Antonio gentilissimo per i tanti consigli utili tipo cosa vedere, su dove andare". Altra ugual espressione: "tornerò in questo B&B Albatros e lo consiglio a tutti coloro che cercano un angolo a Roma Centro Storico/nel cuore di Roma". Si nota anche che per entrambi il voto assegnato è stato 5 su 5 per le caratteristiche dell'albergo. Inoltre, ci teniamo a precisare che di queste recensioni con questo nominativo ve ne erano molte, una dopo l'altra, sempre con tali espressioni e che questo albergo aveva solo queste recensioni positive. Il restante, invece, era estremamente negativo.

Proposta di metodo: diventare un Opinion Leader credibile su TripAdvisor

5.1 Consigli utili per far salire la reputazione di un albergo

Nel Cap.4 abbiamo dimostrato le varie recensioni circolanti su TripAdvisor e l'ingenuità con cui un albergatore tende a rilasciarle poiché sono facilmente individuabili.

La nostra proposta così, dopo un'analisi della piattaforma TripAdvisor, vuole essere un aiuto agli albergatori al fine di consigliare loro il miglior modo di rilasciare recensioni per far salire la propria reputazione senza essere riconosciuti. Premettiamo però che di sicuro non c'è da imparare come scrivere una recensione falsa affinché sembri autentica, in modo da poter così affossare una recensione negativa con 10/100 recensioni positive fasulle. Bisogna imparare a gestire una recensione negativa e rispondere in modo appropriato ad essa. Sicuramente non

sempre il cliente ha ragione e quindi in alcuni casi particolari dove diventa difficile gestire la situazione può anche essere opportuno cercare di far eliminare una recensione scritta impropriamente che reca particolari danni d'immagine al vostro brand.

Vogliamo essere originali e pragmatici e soffermarci su un aspetto diverso da “è etico o meno”, il problema è che scrivere recensioni false su TripAdvisor è in realtà di scarsa utilità. Il punteggio della classifica di TripAdvisor dipende in gran parte dal numero e dalla frequenza delle recensioni, quindi un paio di recensioni false non spostano né la classifica degli hotel né l'opinione dei navigatori in caso siano presenti diverse recensioni negative.

Ma se proprio la recensione falsa la vogliamo scrivere, vediamo almeno di farlo correttamente.

Dopo questa premessa, ci limitiamo a consigliare alle aziende alberghiere come bisognerebbe scrivere una “falsa” recensione, cosa bisognerebbe omettere o inserire in una recensione al fine di essere credibili.

Alcune indicazioni importanti per aggiungere un po' di credibilità alla nostra recensione sono le seguenti.

Andando ad analizzare il punto 1 della Fig.5.1, il consiglio è di non riempire la recensione di punti esclamativi poiché è difficile che qualcun altro se non il proprietario dell'albergo voglia enfatizzare così tanto la positività della sua esperienza o al massimo questo accade

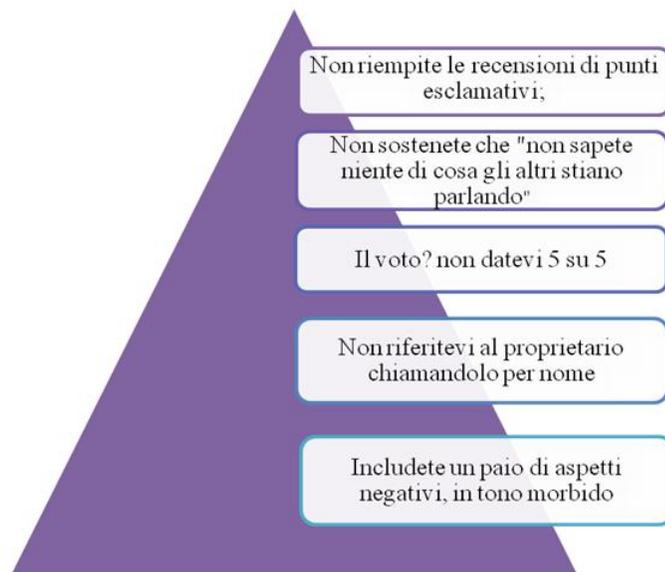


Figura 5.1: Proposta di scrittura di "falsa" recensione.

quando la recensione è negativa e si riscontra una totale insoddisfazione nella propria esperienza. Al punto 2 si consiglia di non iniziare una recensione con frasi di questo tipo (vedi Fig.5.1) poiché se le recensioni dell' hotel sono quasi tutte negative, non servirà a niente scriverne una falsa, tanto meno infastidire i lettori con frasi come "non so di cosa gli altri stiano parlando". A riguardo è impossibile andare totalmente ad opporsi alle altre recensioni negative.

In merito al punto 3 si consiglia di non dare 5 su 5, meglio un bel 4 su 5, lo fa sembrare un ottimo voto da una persona esigente che non regala il massimo dei voti in modo acritico. Sono rare le volte in cui si trova la perfezione in un albergo soprattutto se si tratta di una rara recensione positiva rilasciata per un albergo. In riferimento al

punto 4, non riferitevi al proprietario chiamandolo per nome, questo è un segnale anche troppo palese che chi scrive è andato a cercar gloria personale più che a descrivere un hotel che non è suo. E' improbabile, tranne in certi casi, che gli ospiti facciano menzione di un nome in particolare nel sottolineare la bravura dello "staff". Per concludere con il punto 5, si consiglia di includere un paio di aspetti negativi se pur in tono morbido. Questa caratteristica è molto importante al fine di essere credibili.

Il problema non è tanto che la recensione suona "falsa". La vera questione è che è troppo positiva per dare realmente qualche tipo di indicazione; nella migliore delle ipotesi è completamente inutile. Le opposizioni, gli aspetti negativi, servono a farsi un'idea più chiara degli aspetti positivi. Se tutto è assolutamente positivo, non ricaviamo dalla recensione nessun reale punto di riferimento.

L'hotel che "va bene per tutti" ed è "perfetto" non interessa a nessuno e non esiste; provate a focalizzare in dettaglio i particolari positivi, ed enfatizzateli mostrando anche qualche piccola "pecca" dell'hotel.

5.2 Esempio di una corretta recensione "falsa"

Dopo aver consigliato il modo migliore per rilasciare una falsa recensione al fine di far salire la propria reputazione nel Par.5.1, andremo in questo punto a porre un esempio di come dovrebbe essere scritta

una falsa recensione per ottenere tale risultato.

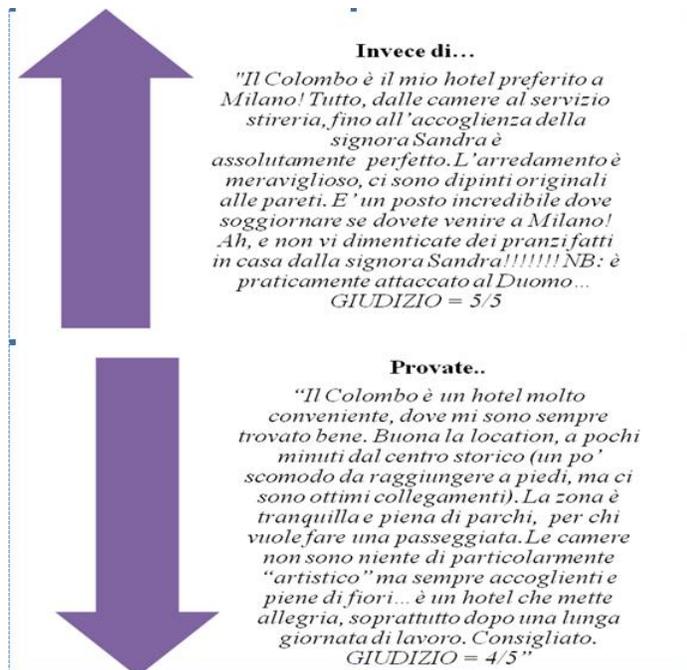


Figura 5.2: Esempio di scrittura di "falsa" recensione.

In questo esempio sono stati riportati i consigli dati al punto 5.1 nonché l'inserimento di alcuni commenti negativi, il punteggio 4/5 invece di 5/5 e senza troppo enfatizzare ogni commento rilasciato.

La nostra ricerca si è basata e attorniata a quel che è il mondo TripAdvisor e dalle tante introduzioni in questo elaborato tesi si sarà compreso che si tratta di una piattaforma la quale, da quando i consumatori hanno cambiato il loro rapporto con internet ovvero sempre più si apprestano a fare i loro acquisti su internet basandosi sulle varie recensioni provenienti da terzi, da sconosciuti, è divenuta molto importante e per i consumatori i quali si basano prima del loro acquisto delle recensioni su piattaforme di questa portata insieme a Booking Blog o Venere.com. e per le aziende che hanno sotto controllo il loro andamento e la loro reputazione.

Proprio perché è diventato quindi così importante TripAdvisor, le aziende sono arrivate a capire che gli utenti potevano lasciare recensioni negative se insoddisfatti dalla loro vacanza intaccando la loro immagine e influenzando possibili futuri acquisti da altri utenti. Insieme all'evoluzione di TripAdvisor, consumatori e internet, le aziende

alberghiere si sono evoluti nell'utilizzare un'analisi automatica di queste recensioni così detta "sentiment analysis e opinion mining" per poter in minor tempo avere sotto controllo il proprio andamento, cosa la gente pensa della propria struttura attraverso una classificazione delle opinioni in positive e negative sempre in maniera computerizzata.

Non tutte le strutture purtroppo sono oneste e offrono dei buoni servizi quanto promesso ai propri clienti e può capitare che un cliente, dunque, sia soddisfatto o meno.

Ora, una volta che un'azienda arriva dunque a conoscenza attraverso la sentiment analysis del proprio andazzo, ci sono due conseguenze: una che l'azienda alberghiera sia soddisfatta e resta sul proprio processo; l'altra, invece, che un'azienda data l'insoddisfazione della propria reputazione e quindi di conseguenza la sua classifica ultima tra tutti gli alberghi del luogo in cui è situata, si appresta a modificare tale situazione.

Quello che abbiamo notato durante una ricerca sul sito TripAdvisor è che molti albergatori vogliono modificare la propria situazione e migliorarla attraverso il rilascio di recensioni false e come avete potuto notare attraverso gli esempi che abbiamo postato nel Cap.4, è vera la nostra ipotesi.

Il nostro studio si è fondato sulla ricerca di più di 50 casi di recensioni false anche se pur postandone 3 come esempio. Dopo una lunga analisi siamo stati in grado, scopo della tesi, di aiutare e consigliare

le aziende alberghiere a scrivere meglio una recensione per sollevare le opinioni della stessa come in esempio nel Cap.5.

Bibliografia

- [1] Alissa Quart, 2003 – I giovani e le ossessioni del marchio, Sperling & Kupfer editor.
- [2] A. Esuli and F. Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In Proceedings of LREC, volume 6, pages 417-422. Citeseer, 2006.
- [3] Ginny Mc Grath and Steve Keenan. We're clean' pledges TripAdvisor. The Sunday Times. May 2007.
- [4] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2): 1-135, 2008.
- [5] N. Jindal and B. Liu., Opinion spam and analysis. In Proceedings of the International conference on Web

- search and web data mining, pages 219-230. ACM, 2008.
- [6] N. Jindal and B. Liu. Review spam detection. In Proceedings of the 16th International conference on World Wide Web, page 1190. ACM, 2007.
- [7] Zanzotto, Fabio Massimo, Marco Pennacchiotti, and Kostas Tsiou-tsiouliklis. 2011. Linguistic redundancy in twitter. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 659–669, Edinburgh, Scotland, UK., July. Association for Computational Linguistics. 12.
- [8] Mario Gerosa, Roberta Milano, 2009- Viaggi in Rete. Dal nuovo marketing turistico ai viaggi nei mondi virtuali. Franco Angeli Editore.
- [9] Lorenzo Radice and Saverio Bruno. Analisi della reputazione in ambienti Web 2.0: dallo studio alla realizzazione di un tool per il Sentiment Analysis. Master's tesi, Politecnico di Milano, Italy, December 2009.
- [10] Pang, B., Ledd, L. and Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), 2002.

- [11] Jerome Seymour Bruner, Jacqueline J. Goodnow, George A. Austin, 1986- A study of thinking, Social science classics series. Transaction Publishers Editor.
- [12] Elihu Katz, 2010- The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. Ardent Media Editor.
- [13] Bing Liu, 2011- Web Data Mining: Exploring Hyperlinks, Content, and Usage Data. Data-Centric Systems and Applications Series. Springer Editor.
- [14] ChengXiangZhai, 2009- Statistical language models for information re-trieval. Synthesis lectures on human language technologies. Morgan & Claypool Publisher.
- [15] TripAdvisor- <http://www.tripadvisor.it/>

Ringraziamenti

Per questo lavoro svolto vorrei ringraziare il professore Fabio Massimo Zanzotto per la possibilità data di collaborare insieme in questa tesi scegliendo un argomento utile, attuale e importante nel mio campo di studio il quale mi ha tanto entusiasmato e per i suoi insegnamenti informatici date le mie carenze qualitative informatiche.

Vorrei anche ringraziare la mia famiglia per avermi dato la possibilità di essere lontana da loro e di accettare ogni mia decisione. E un ringraziamento in particolare va a Francesco che, se pur a distanza, è riuscito a sopportarmi per tutti i miei momenti di “crisi esistenziale” con tanta pazienza, che mi ha anche supportata molto nell’informatica e che semplicemente per la sua esistenza e per il suo sapermi far ridere, da quando è entrato a far parte della mia vita, è riuscito a darmi una carica e una positività indescrivibili. Ultimi ringraziamenti vanno a quelle poche, pochissime amiche che togliendo l’invidia mi hanno sostenuto a raggiungere questo mio grande traguardo in tutta la mia

carriera universitaria e in particolare a Selene e Ariana.