

Una Strategia di Ranking di Attivita` Commerciali Basata su User Generated Content

Stefania Marrara¹, Gabriella Pasi¹, and Simone Pellegrini¹

University of Milano - Bicocca, DISCo
Viale Sarca 336, 20126, Milano
{stefania.marrara, pasi}@disco.unimib.it

Sommario Questo articolo presenta una strategia di ranking di attività commerciali basata su tre criteri: la “topicalità” dell’attività rispetto alla query dell’utente, il rapporto tra recensioni negative/positive e la popolarità forniti all’attività direttamente dagli utenti. Il prototipo sviluppato su Yelp e testato sullo Yelp Academic Dataset ha dato risultati preliminari interessanti.

Keywords: User Generated Content, Yelp, ranking

1 Introduzione

Nell’ultimo decennio i siti di recensioni di attività commerciali hanno dimostrato di avere un incredibile impatto commerciale nei riguardi delle attività da loro indicizzate. Essere o non essere indicizzati in questi siti, e la posizione in cui la propria attività viene restituita nella lista dei risultati a seguito di una query, può decisamente cambiare la fortuna commerciale di un’attività rispetto al proprio mercato di riferimento (vedi [2]). Purtroppo questi siti hanno dimostrato di possedere una certa vulnerabilità rispetto ai criteri con cui viene impostato l’ordinamento delle attività reperite a seguito di una query [4]. In questo contesto, l’aspetto innovativo della strategia di ranking qui presentata consiste nell’utilizzo di un insieme di criteri (alcuni tradizionali, altri qui definiti) per la stima della rilevanza di un’attività commerciale rispetto alla query formulata da un utente. Data la particolare tipologia di “oggetti” informativi da reperire, la definizione di criteri di rilevanza che possano ben valorizzare le caratteristiche di tali oggetti e tenere conto delle informazioni generate dagli utenti in relazione ad essi costituisce un interessante problema.

Pertanto è stata definita una strategia per la stima di rilevanza delle attività rispetto alle necessità informative dell’utente (strategia di ranking) che combina la “topicalità” (stima della pertinenza rispetto alle parole chiave della query) dell’attività, il rapporto tra recensioni negative/positive e la popolarità forniti all’attività direttamente dagli utenti. La classificazione delle recensioni di ogni attività in positive/negative è stata effettuata sulla base del loro contenuto usando SentiStrength¹. L’approccio definito ha portato alla progettazione ed allo sviluppo di un prototipo basato sulla collezione costituita dallo Yelp Academic Dataset, mediante il quale sono state effettuate delle valutazioni preliminari che hanno dato risultati interessanti, presentati nella sezione 3.

¹ <http://sentistrength.wlv.ac.uk/>

2 La strategia di ranking proposta

In questa sezione viene formalizzata la strategia di ranking di attività commerciali che, data una query, combina tre diversi valori per ciascuna attività reperita:

- il valore (TR) prodotto dalla valutazione del matching tra query e descrizione dell'attività, calcolata sulla base di un modello di Information Retrieval (ad esempio il *vector space model* [3]),
- il *valore basato sulle recensioni (RS)* che tiene in considerazione l'opinione dell'utente così come è espressa dal contenuto delle recensioni stesse,
- un *punteggio di popolarità (PS)* assegnato direttamente dagli utenti (ad esempio attraverso il meccanismo delle stelle in Yelp).

La strategia di ranking qui proposta si basa su un ordine di priorità tra i tre criteri coinvolti nella stima di rilevanza di ciascuna attività commerciale: 1) non vogliamo che, nella lista dei risultati di una query, un'attività non rilevante rispetto alla query stessa abbia una posizione migliore rispetto ad una rilevante solo perchè più popolare, 2) il valore calcolato analizzando la polarità delle recensioni descrive in modo più significativo il reale sentimento degli utenti verso una data attività commerciale, rispetto al valore derivato dal punteggio assegnato dagli utenti in base a criteri non ben definiti.

Per queste ragioni si è scelto di aggregare i criteri per calcolare la stima di rilevanza utilizzando un *operatore di aggregazione con priorità* F_p [5]. Questo operatore combina linearmente diversi criteri ove: 1) per ciascuna attività commerciale a , il peso del criterio più importante C_1 ha sempre valore 1; 2) i pesi degli altri criteri C_i , $i = \{2, 3\}$ sono $\lambda_i = \lambda_{i-1} * C_{i-1}(a)$ dove $C_{i-1}(a)$ è il grado di soddisfazione del criterio C_{i-1} per l'attività a , mentre λ_{i-1} è il peso del criterio C_{i-1} . Pertanto $\lambda_i \in [0, 1]$. In questo modo, dati due criteri C_i e C_{i-1} , più è alto il grado di soddisfazione di C_{i-1} , più il criterio C_i influisce sul calcolo del valore complessivo.

L'operatore di aggregazione con priorità usato per calcolare il valore di rilevanza complessivo è definito come $F_p : [0, 1]^n \rightarrow [0, n]$, ed è tale che data un'attività a

$$F_p(C_1(a), \dots, C_n(a)) = \sum_{i=1}^n \lambda_i * C_i(a) \quad (1)$$

Siano a un'attività commerciale, A l'insieme delle attività indicizzate dal sito e q una query dell'utente. Siano $NP \in \mathbb{N}$ il numero di recensioni positive dell'attività a , $NN \in \mathbb{N}$ il numero di recensioni negative di a , $stars \in \{0, 0.5, 1, \dots, 4.5, 5\}$ il valore di popolarità di a assegnato direttamente dagli utenti. Per ogni attività Yelp assegna un valore di popolarità calcolato come media dei punteggi assegnati direttamente dagli utenti durante la recensione dell'attività commerciale. Questo valore $avstars$ ha di nuovo valori in $\{0, 0.5, 1, \dots, 4.5, 5\}$. $TR \in \mathbb{R}^+$ è il valore di topicalità del documento d rispetto alla query q .

Il valore basato sulla polarità delle recensioni RS è calcolato attraverso una funzione non simmetrica $RS(NP, NN)$ definita in modo da promuovere le atti-

vità commerciali con un alto numero di recensioni positive rispetto a quelle con un alto numero di recensioni negative.

$RS : \mathbb{N} \times \mathbb{N} \rightarrow [0, 1]$ ha la seguente definizione

$$RS(NP, NN) = \begin{cases} 0 & \text{se } NP = 0 \text{ oppure } NP < NN \\ 1 - \frac{1}{NP+1}, & \text{se } NN = 0 \\ 1 - \frac{NN}{NP}, & \text{se } NP \geq NN \text{ e } NN > 0; \end{cases} \quad (2)$$

Si noti che il valore così calcolato dipende dal rapporto tra il numero di recensioni negative e positive, non dal loro numero complessivo. In questo modo si evita il noto problema del *cold start* in cui una nuova attività commerciale, priva di recensioni, difficilmente apparirà nelle prime posizioni della lista dei risultati di una query e avrà pertanto la possibilità di essere recensita.

Il valore di popolarità è calcolato mediante una funzione PS normalizzata nell'intervallo $[0, 1]$ e definita come $PS : \{0, 0.5, 1, \dots, 4.5, 5\} \rightarrow [0, 1]: PS(avstars) = avstars/5$

Sia $TR_{max} \in \mathbb{R}^+$ il valore massimo di TR di tutte le attività commerciali reperite rispetto alla query q nella collezione A .

Il valore complessivo di stima della rilevanza qui definito è calcolato dalla funzione $F_p(TR(a), RS(a), PS(a))$ definita come $F_p : [0, 1]^3 \rightarrow [0, 3]$

$$F_p(TR(a), RS(a), PS(a)) = \lambda_{TR} \frac{TR(a)}{TR_{max}} + \lambda_{RS} RS(a) + \lambda_{PS} PS(a) \quad (3)$$

dove $\lambda_{TR} = 1$, $\lambda_{RS} = \lambda_{TR} \frac{TR(a)}{TR_{max}}$, e $\lambda_{PS} = \lambda_{RS} RS(a)$.

Grazie a questa funzione la posizione di un'attività commerciale nella lista dei risultati risulta premiata o penalizzata dai due criteri derivanti dalla comunità di utenti, senza però che un'attività non pertinente rispetto alla query risulti in posizione migliore rispetto ad una pertinente solo a causa della popolarità.

3 Implementazione

Al momento della stesura di questo lavoro non è stata ancora effettuata una valutazione esaustiva del prototipo, ma sono state fatte alcune prove d'uso utilizzando lo *Yelp Academic Dataset*. Il prototipo è stato sviluppato utilizzando Lucene 5.2.1, JsonSimple 1.1.1 e SentiStrength 2. Ogni query è stata eseguita due volte: nel primo caso la lista restituita è ordinata usando l'approccio proposto, nell'altra invece viene utilizzato l'ordinamento basato sul vector space model fornito da Lucene. Sono state definite sette query nelle categorie "hotel ristoranti" e "salute, cura della persona"; si è proceduto quindi ad una valutazione manuale delle precedenti categorie contenute nello *Yelp Academic Dataset* per costruire un piccolo benchmark. Gli utenti hanno valutato le attività, contenute nelle due categorie prescelte, classificandole in modo binario in *interessanti/non interessanti* rispetto alle query espresse in linguaggio naturale. In Tabella 1 sono mostrati i valori di precisione e di Normalized Discounted Cumulative Gain [1] ottenuti considerando le prime cinque ($P@5, NDCG@5$) e le

Query	La nostra strategia				Approccio VSM			
	P@5	NDCG@5	P@10	NDCG@10	P@5	NDCG@5	P@10	NDCG@10
Q1	0.8	0.82	0.7	0.76	0.7	0.82	0.9	0.88
Q2	0.4	0.36	0.3	0.48	0.6	0.66	0.3	0.66
Q3	0.8	0.88	0.5	0.87	0.5	0.54	0.5	0.67
Q4	0.4	0.42	0.5	0.58	0.2	0.28	0.3	0.36
Q5	1	1	0.8	0.88	0.8	0.7	0.6	0.8
Q6	0.6	0.73	0.3	0.73	0.5	0.6	0.4	0.8
Q7	0.6	0.5	0.6	0.81	0.6	0.84	0.4	0.96

Tabella 1. Calcolo di P@5, NDCG@5, P@10 e NDCG@10 considerando l’approccio proposto o il vector space model (vsm)

prime 10 ($P@10$, $NDCG@10$) posizioni della lista dei risultati ottenuti grazie alle due diverse strategie di ranking a confronto. Come evidente in tabella i risultati ottenuti sono incoraggianti. Da notare ad esempio i valori della query 2 che chiede di trovare attività di massaggi orientali: in questo caso un basso gradimento da parte degli utenti di tutte le attività gestite da cinesi ha abbassato il ranking delle attività pertinenti rispetto al ranking generato dalla stima della sola topicalità. La stessa situazione si è verificata con la query 7 che chiede di trovare autonoleggi economici: la bassissima reputazione sociale ha penalizzato il ranking delle attività valutate più pertinenti dalla stima di topicalità.

4 Conclusioni

In questo lavoro abbiamo presentato una funzione per il calcolo del ranking di una lista di risultati per siti di attività commerciali con recensioni che tiene conto sia della stima di rilevanza basata sul concetto di topicalità che di fattori che esprimono la popolarità dell’attività stessa in una comunità di utenti.

Come caso di studio abbiamo usato la collezione Yelp Academic Dataset ed il prototipo sviluppato ha ottenuto risultati incoraggianti. Pertanto si procederà ad una fase di valutazione completa con un benchmark adatto. Una possibile scelta da valutare è il benchmark offerto dalla TREC Contextual Suggestion Track, che però è orientato al problema delle personalizzazione.

Riferimenti bibliografici

1. Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, October 2002.
2. Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. *Com (September 16, 2011). Harvard Business School NOM Unit Working Paper*, (12-016), 2011.
3. Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.
4. Holley Simmons. How some yelpers are holding restaurants hostage. *The Washington Post*, 2015. <http://tinyurl.com/pc578ea>.
5. Ronald R. Yager. Prioritized aggregation operators. *International Journal of Approximate Reasoning*, 48(1):263 – 274, 2008. Special Section: Perception Based Data Mining and Decision Support Systems.