Personalizing Local Search with Twitter

Lu Guo
University of Texas at Austin
Department of Computer Science
guolu@cs.utexas.edu

Matthew Lease
University of Texas at Austin
School of Information
ml@ischool.utexas.edu

ABSTRACT

We propose a new ranking model for personalized local search. While local search verticals such as *Google Local* and *Yahoo! Local* incorporate physical proximity and public sentiment (reviews and ratings), their rankings reflect minimal personalization. We personalize local search by integrating Twitter social network structure and content analysis. Specifically, we infer sentiment for tweets by the user and those he follows which mention local businesses by name. We also provide a Google Android tailored interface and interaction experience for local search with Twitter integration. Evaluation of search accuracy and quality of user experience via a 25 person user study shows both improved search accuracy and anecdotal evidence of greater user satisfaction.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]

General Terms

Algorithms, Experimentation, Performance

Keywords

Local search, Twitter, Ranking algorithm, Personalization

1. INTRODUCTION

Local search [1] reflects a natural evolution of traditional off-line advertising to the Web. Increasingly people utilize online local search verticals like Google Local and Yahoo! Local to find a business category or the name of a consumer product. These verticals incorporate physical distance and social ratings and recommendations into ranking.

Without care, generic social feedback can actually reduce search accuracy. Consider the first author's own experience as a foreigner living in Texas (which inspired this work). Most people in the local area are not foreigners, and their ratings for local cuisine dominate the online reviews. However, their food preferences are markedly different, and their online ratings appear to degrade the author's search results.

To personalize local search, we incorporate Twitter's social network structure and content into our ranking algorithm. The user may have tweeted himself before about local businesses, or those he follows may have tweeted. A user re-tweeting a tweet from someone he follows indicates that the user may share similar beliefs or interests.

SIGIR 2011 Workshop on Enriching Information Retrieval (ENIR 2011)
July 24–28, 2011, Beijing, China.

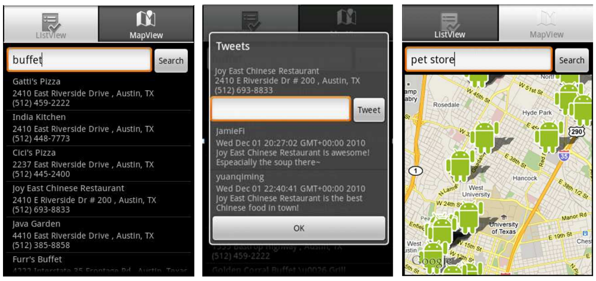


Figure 1: Screenshots from our Android Personalized Local Search application show the interface.

Interface. Figure 1 shows our personalized Local Search Google Android application, which features two primary views: *Listview* and *Mapview*. With Listview, search results are presented as an ordered list. In the Mapview, users are presented a searchable map display of results. In both cases, users can click on any result to view tweets about it written by those they follow. The application also enables users to post new tweets about search results directly through the application as they are being viewed. As users are encouraged to produce more tweets, the system can subsequently exploit these to improve personalized ranking over time.

Contributions. Our main contributions are as follows. First, we integrate personalized social media with local search to improve ranking as well as to augment the interface and interaction experince with returned results. Second, we deploy our system as useful a mobile Google Android application. Third, we report a study of 25 users which informs our understanding of local search, particularly personalization. Results shows improvement over Google Local search accuracy and anecdotal evidence of greater user satisfaction.

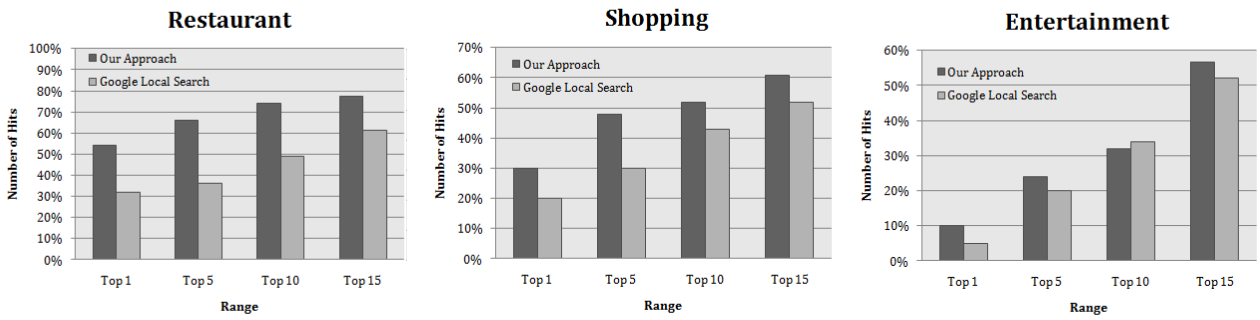
2. FEATURES AND MODEL

Our ranking algorithm utilizes four features as follows:

Baseline Ranking: $BR(b)$ from Google Local[1] orders businesses b from 1 (best) to N (worst).

Personal Preference: $PP(U, b)$ $[-1, 1]$ indicates user U's personal sentiment toward business b. For any past tweet t by U which mentions b by name, we infer positive (1) or negative (-1) polarity $p(U, t)$ via a maximum-entropy classifer[2]. We then define $PP(U, b) = \frac{1}{n}\sum_{i=1}^{n} p(U, t)$, a simple average of polarity scores over U's tweets mentioning b.

[1] http://ajax.googleapis.com/ajax/services/search/local?
[2] http://twittersentiment.appspot.com/api/classify?



(a) Example queries: "buffet","cafe", "fastfood", "seafood","pizza", ...

(b) Example queries: "drugstores", "bookstore", "jewelry","footwear","cosmetics", ...

(c) Example queries: "theater","fitness center", "club","tattoo", "haircut", ...

Figure 2: Recall of Top N results grouped into 3 query categories: restaurants, shopping, and entertainment

Following Preference: $FP(U,b)$ $[-1,1]$ is defined similarly to PP(U,b) for each person U follows.

Association Strength: $AS(U,f)$ $[0,1]$ expresses relative similarity or importance of each followed person $f \in F$ to user U. We model this as the probability of U re-tweeting tweets from f rather than another person whom U follows. Let $c(U,f)$ denote the number of f's tweets re-tweeted by U. We estimate $AS(U,f) = c(U,f)/\sum_{g \in F} c(U,g)$ via maximum likelihood, where $\sum_{f \in F} AS(U,f) = 1$.

Model. We define a linear model of three parameters ω integrating the above features to score $s(b)$ each business b:

$$s(b) = \frac{\omega_{BR}}{BR(b)} + \omega_{PP} PP(U,b) + \omega_{FP}\, \mathbf{E}_{AS(U,f)}\left[FP(f,b)\right]$$

$\mathbf{E}_{AS(U,f)}[FP(f,b)]$ denotes the expected *Following Preference* weighted over U's *Association Strength* distribution.

3. EVALUATION

Queries and assessment. We asked 25 active Twitter users to formulate a local search query for each of *dining*, *shopping*, or *entertainment* categories, totaling 75 queries. Users were 20-57 years old and spanned 4 ethnicities. 12 had recently relocated (the others were considered locals). Users followed 53-526 other Twitter users ($\mu = 193$, $\sigma = 121$). Each user's query was submitted to Google Local and the top-30 results were presented to the user. Users selected the best 15 results and manually reordered them, best to worst.

Tweets. The 50 most recent tweets were collected for each of the 25 users and those they followed via the Twitter API[3]. $203/750 \approx 27\%$ of user tweets were re-tweets. Including followings, roughly 242,000 tweets were collected (followings generated 0-8 re-tweets). All text was lower-cased.

Setup. We re-rank the top 30 results for each query from the BR model using our system. We then compare the top-15 results (only) of each system to the best 15 results selected by the user. We measure accuracy of each system at various top-N cutoffs: 1, 5, 10, and 15, comparing to the corresponding top-N results indicated by the user's ranking.

Parameters. We set ω by training the SVM on a separate set of 15 queries and tuning for accuracy@15. We find an optimal setting $\omega_{BR} = 1$, $\omega_{PP} = 1.5$ and $\omega_{FP} = 0.8$.

Results. Figure 2 shows average accuracy over all 75 queries. Our personalized ranking algorithm is seen to achieve consistently improved accuracy over Google Local.

[3]Using the twitter4j java library for the twitter API

Since our personalization method depends on Twitter coverage for the queries, we note that over the 25 users * 3 queries * 30 results $\approx$ 2250 unique businesses considered, $\sim$ 31% were mentioned by at least one tweet. The remaining $\sim$ 69% of businesses received score $s(b) = \frac{\omega_{BR}}{BR(b)}$.

Improvement over the baseline model's accuracy@1 appears most difficult due to our model's limited capacity to exchange results at higher ranks vs. those at lower ranks, i.e. difficulty demoting highly ranked baseline results. Recall that $BR(r_i)$ for the i^{th} ranked search result is $\frac{1}{i}$. Then $\Delta(\frac{\omega_{BR}}{BR(r_i)} - \frac{\omega_{BR}}{BR(r_{i+1})}) = \frac{\omega_{BR}}{i(i+1)}$, which is a decreasing function on any interval. So it is much harder to exchange the orderings with results with higher rankings than those with lower rankings via the new PP, FP, and AS features.

Some categories seem to benefit more from personalization than others. It seems restaurant queries have various sub-categories for which people with differing nationalities or ages have rather different preferences. In contrast, entertainment like theaters and fitness centers appears to be a narrower category in which user preferences are less diverse.

Open-ended feedback from users indicated that our system seemed to know more about what they wanted. Users also indicated feeling more valued by knowing a personalized rankings generated uniquely tailored to them. The main complaint we received was actually a feature request; users wanted Facebook integration as well as Twitter.

4. FUTURE WORK

Whereas the Twitter Sentiment API we used is reported to achieve 82.7% accuracy, we have now built our own classifier which combines semantic role labeling with syntactic parsing to achieve $\sim$89.5% accuracy. We will next apply this improved sentiment model in our personalized local search algorithm. We also plan to pursue approximate matching methods for recognizing business name mentions in tweets.

Acknowledgments. We appreciate the reviewers' feedback. This work was partially supported by a Google *Android for Education* award, and a John P. Commons Fellowship for the second author.

5. REFERENCES

[1] Jean-Francois Paiement, James G. Shanahan, and Remi Zajac. Crowdsourcing local search relevance. *CrowdConf 2010*, 2010.

[2] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *Technical report, Stanford Digital Library Technologies Project*, 2009.