Timely Tip Selection for Foursquare Recommendations

Max Sklar
Foursquare Labs
568 Broadway, 10th Floor
New York, NY
max@foursquare.com

Kristian J. Concepcion
Foursquare Labs
568 Broadway, 10th Floor
New York, NY
kjc@foursquare.com

ABSTRACT

This poster summarizes the techniques we use to serve Foursquare tips for a given venue and more specifically the strategies employed for choosing timely and seasonal tips.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering; G.3 [**Probability and Statistics**]: Time series analysis; I.2.7 [**Natural Language Processing**]: Text analysis; I.5.1 [**Pattern Recognition**]: Models–Statistical

Keywords

bhattacharyya coefficient, context-aware recommenders, foursquare, machine learning, natural language processing, text classification

1. INTRODUCTION

Foursquare is a location-based recommendation engine. A primary action for users is to write a tip, which is a short public note attached to a venue, often a review or suggestion. Any given venue is likely to have many tips attached to it, which vary in quality and relevance. With the recent focus on search and discovery as well as passive location awareness, we have developed a number of heuristics in order to serve the right tips to the right people at the right time.

2. TIP SELECTION COMPONENTS

Language Identification: In order to avoid serving languages that a user does not understand, a language classifier on Foursquare tips was built using an ensemble of open source and home-grown solutions.

Global quality: We created a hand-labelled training set of high and low quality tips based off of a strict set of quality guidelines. Raw scores from various statistical classifiers that were trained to identify specific traits such as sentiment or spam were used as features to train a quality model.

Personalization: We developed a number of signals which take into account the user's tastes and social connections.

Timeliness and Seasonality: For any given date and time, a tip is analyzed in order to determine whether it is appropriate for a particular time of week or time of year. In this poster, we go into more detail on the system for analyzing this component.

RecSys 2014 Poster Proceedings October 6-10, 2014, Foster City, Silicon Valley, USA

Table 1: Comparing Food Items with the Highest Bhattacharyya Similarity to Lunch

English Phrase	Score	Thai	Translated	Score
salad sandwich	0.943	ก๋วยเตี๋ยว	noodles	0.884
turkey sandwich	0.929	ชา เขียว	green tea	0.845
cuban sandwich	0.918	กาแฟ	coffee	0.832
panini	0.913	ขา หมู	pig's feet	0.827

3. TIMELINESS OF PHRASES AND TIPS

Through our Swarm app, users check in to share their location and leave a short update for their friends called a shout. In order to find phrases which are time-sensitive, we looked at shouts instead of tips because they were more specific to what users were doing at any particular time.

Our model for phrase popularity over the course of the week mirrors our model for venue popularity[4]. For each supported language, we divided the week into 168 hour buckets. We then counted the number of times each phrase was used in a given bucket. We also counted the total number of shouts in each bucket to produce a baseline distribution.

The *Bhattacharyya coefficient*[1] is a metric for comparing the similarity between two probability distributions. Given two phrase distributions P and Q, we define the similarity to be

$$S(P,Q) = \sum_{w \in W} \sqrt{P(w)Q(w)}$$

where W is the set of all 168 weekhour buckets.

For example, the Bhattacharyya coefficient between any phrase and the word "lunch" provides a measure of how appropriate that phrase is for lunch time. The food items which rank most highly in this metric for English and Thai give interesting insights into the lunch habits of different language groups (Table 1).

Furthermore, the Bhattacharyya coefficient between any phrase and the baseline distribution measures the time sensitivity of that phrase. We extracted all the phrases that meet a certain threshold for time sensitivity. Then, each phrase-bucket was assigned a timeliness score which is the log-ratio of the phrase probability and the baseline probability.

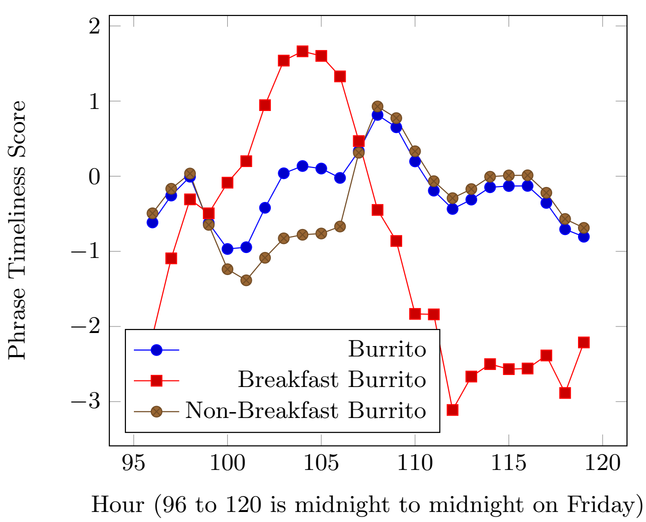
We defined $C(p)$ to be the total number of times phrase p appears in the corpus, and $C(p_w)$ to be the total number of times p appears in weekhour w. Finally, α is a 168-dimentional Dirichlet smoothing constant on phrase count data[5] and b is defined as a phrase to correspond with the baseline counts. The timeliness score for a phrase at weekhour w is computed as follows.

$$T(p, w) = \ln\left(\frac{C(p_w) + \alpha_w}{C(p) + \sum_i \alpha_i} \div \frac{C(b_w) + \alpha_w}{C(b) + \sum_i \alpha_i}\right)$$

The timeliness score for a tip is the sum of the scores of its phrases. For example, at Veselka (a popular Ukrainian restaurant in New York's East Village), a user wrote "They're open 24/7 - turn up after your night out and partake of the pierogis with applesauce." The terms "24/7", "turn up", "night out", and "pierogi" all meet the Bhattacharyya threshold. Their respective scores for Sunday night at midnight are 1.2, 0.6, 0.3, and -0.6. These sum to 1.5 which is positive and indicates that this tip is timely on Sunday night.

We supplemented our shout counts with the English Wordnet[2] food corpus and our English menu database. This allowed us to associate entries in the Wordnet corpus with specific meals (breakfast, lunch, dinner, dessert, and late night). For phrases in the Wordnet food corpus with insufficient shout data, we replaced the distribution with that of the matching abstract mealtimes.

One problem we encountered was with non-compositional compound phrases. The timeliness of "burrito" is very different from that of "breakfast burrito", but because the burrito data included all mentions of breakfast burrito as well, its timeliness score was dampened. To counteract this problem, we merged phrases in our training data that appeared more frequently together so that they would be considered as completely separate entities from their constituent tokens. In terms of burritos, this meant that all mentions of breakfast burrito were counted as one term, and all mentions of burrito not following breakfast were considered as an entirely separate term.



3.1 Evaluation

We evaluated the timeliness score on a hand-labelled set of 825 tips, each with four abstract meal times: breakfast, lunch, dinner, and late night. For each tip and time period, we applied the label of timely, neutral, or untimely. We then compared those labels to our timeliness scores, the result of which satisfied us for using the feature in the product.

The timeliness score serves two purposes: detecting specifically timely tips, and disqualifying untimely tips. With our chosen threshold, we achieved 71.3% precision and 74.5% recall for timely tips against our hand labelled set. Untimely scoring used a different threshold and achieved 74.7% precision and 67.0% recall.

4. EXTENSION TO SEASONALITY

The ability to detect and exploit seasonality is an important feature for search and recommendation systems[3]. There was not enough data to create 365 day-buckets so instead we chose to create buckets based off of weeks. Unfortunately, in the unix calendar utility, many popular holidays crucial to seasonality fall in different buckets each year. To ameliorate this, we forced every month into a 4 week model, with the last week of the month subsuming all extra days beyond the 28th. The last week of each month was then normalized to account for the extra days before the Bhattacharyya coefficients were calculated.

Another issue was caused by phrases that were seasonal in only one year. Very popular movies caused us to associate "James Bond" with mid-November and "Star Trek" with June. We solved this problem by looking at data for each year individually and flagged outliers. Once flagged, we smoothed the counts to bring the offending year more in line with the rest of the data.

4.1 Future Work

Some terms follow a different seasonal pattern depending on geographic region and performance would be improved by geo-fencing phrase distributions by region. For example, the term "fireworks" was found to be incredibly timely during the first week of July for American Independance Day, but there is also a smaller spike in the first week of November for Guy Fawkes Day in Great Britain. Another example was the term "Rangers" being timely in the summer and the winter. The Texas Rangers (a baseball team that plays during the summer) was being conflated with the New York Rangers (a hockey team that plays during the winter).

Geo-fencing by climate zone as opposed to national borders or metropolitan areas would improve results for weather-related phrases such as "outdoor seating" "hot soup", and "air conditioning".

5. REFERENCES

[1] Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. Sankhyā: The Indian Journal of Statistics, 401-406.

[2] Miller, George A. (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

[3] Shokouhi, M. (2011, July). Detecting seasonal queries by time-series analysis. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 1171-1172). ACM.

[4] Sklar, M., Shaw, B., & Hogue, A. (2012, September). Recommending interesting events in real-time with foursquare check-ins. In Proceedings of the sixth ACM conference on Recommender systems (pp. 311-312). ACM.

[5] Sklar, M. (2014). Fast MLE Computation for the Dirichlet Multinomial. arXiv preprint arXiv:1405.0099.