Detecting Trending Venues Using Foursquare's Data

Stephanie Yang
Foursquare Labs
568 Broadway, 10th Floor
New York, NY
stpyang@foursquare.com

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

ABSTRACT

Foursquare is a search and discovery tool which helps users discover venues around the world. Much of the data for these recommendations come from its sister app Swarm, which is a location based social network where users can "check in" to places they visit.

Older versions of Foursquare had a strongly static component to its recommendations. For instance, the top restaurants in New York City do not vary from month to month, and venues with years of consistently strong signals will dominate search results.

In this paper we outline a new algorithm which Foursquare uses in order to discover fresh recommendations. Promoting younger venues with fewer check-ins or older venues with a recent surge of activity increases turnover in our recommendations and yields a better user experience.

Keywords

Recommender systems; Ratings; Foursquare

1. INTRODUCTION

Foursquare has a database of nine billion check-ins and 85 million public venues around the world. Using this data, the mobile app provides personalized venue recommendations to users. Core components of these recommendations are based on foot traffic data in the form of check-ins and passively generated visits from a background location service called Pilgrim [4, 7], as well as other user interactions in the form of venue feedback, tips, and photos.

There is a constant tension between consistency and freshness in Foursquare's recommendations. For example, Thomas Keller's Per Se is always at the top of the results for restaurants in New York City, but most users find value in discovering a more accessible venue like a new mom-and-pop coffee shop around the corner. Likewise, a celebrity chef moving to a new restaurant results in a flurry of activity which is not always captured well by Foursquare's long-term signals.

RecSys '16 Sept 15–19, 2016, Boston, MA, USA

Foursquare has successfully implemented short term trend detection to showcase real-time events as they happen [6]. The algorithm in this paper fills the gap between the near-instantaneous discovery of popular events, and the long-term detection of quality venues.

2. FEATURES

All of the features described below are generated by users' interaction with the Foursquare and Swarm apps and by passively generated visits from Pilgrim. Noteworthy venues inspire users to interact with their apps, and so most user activity for a venue is seen as positive.

2.1 User generated signals

Checkins and visits: The primary signals for trendiness are based on foot traffic in the form of active check-ins and passive visits. Active check-ins typically indicate better venues, since Swarm users tend to broadcast special outings more often than their day to day activities.

Saves: Foursquare users have the option of saving a venue to a list for later. This distinguishes trendy new places from average ones, because it indicates aspirations to visit.

Tips from users: Users have the option of writing tips at any venue, which are shown to other users as part of the local discovery experience [5]. Trendy venues consistently attract a larger number of tips compared to the average venues.

Tips from vetted accounts: A handful of user accounts are unusually influential. For example, some celebrities and local blogs about food maintain active Foursquare accounts with tens of thousands of followers. Tips from these accounts drive foot traffic and are a leading signal of venue trendiness.

Explicit feedback: The Foursquare app prompts it users to leave explicit ratings—like, dislike, or neutral—about the places they visit.

Photos: The excitement of visiting a noteworthy venue is often reflected by our users documenting their visit with photographs.

2.2 Trend detection

For each of the activities listed above, we calculate two statistics.

The first statistic is derived from fitting a trend line through the time series of the activity. The signal that we use is given by the equation

$$S = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}},$$

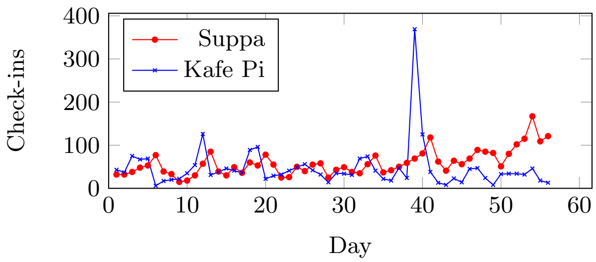

Figure 1: Number of check-ins per day for two restaurants in Istanbul. Note that check-ins exhibit a weekly cycle. The spike in activity for Kafe Pi on day 39 was due to a single event.

where $\hat{\beta}$ denotes the slope of the trend line through the time series with 56 days of data, and $\sigma_{\hat{\beta}}$ denotes the standard error of the estimate $\hat{\beta}$.

In Figure 1 we display the number of check-ins per day for two restaurants in Istanbul, Turkey. The trend lines for both time series (omitted from the figure) have similar slopes. Although the value of $\hat{\beta}$ is positive and similar for both venues, the value of $\sigma_{\hat{\beta}}$ is lower for Suppa than it is for Kaffe Pi. Hence the signal S for Suppa is larger than the corresponding signal for Kaffe Pi. In general, venues with erratic or spiky activity do not benefit from one-time events for this class of signals.

The second statistic is a decayed sum of the activity, calculated with a half life of 56 days.

$$D = \sum_d c_d e^{\lambda d},$$

where d is the number of days prior to the current day, c_d is the total amount of user activity on that day, and $\lambda = -\ln 2/56$. Note that short half lives are associated with noisier data, and long half lives lead to a lack of freshness. For example, the venue Kafe Pi in Figure 1 has a spike in activity on Day 39, which would have dominated the signal if the half life were too short. Longer half lives have more stability, and we found that very long half lives lead to a lack of freshness in our recommendations. In our research, 56 days is the best balance for both stability and freshness.

3. COMBINING THE SIGNALS

The distribution of the S-scores is roughly bell-shaped, while the distribution of the D-scores has a long tail. In order to combine the two classes of scores, we normalize each signal to a Gaussian distribution using the function

$$N = \Phi^{-1}(r),$$

where Φ is the cdf of the standard $N(0, 1)$ distribution and r is the relative rank, between 0 and 1, of the venue when compared to all other venues and sorted by a given score. We then combine the signals linearly with hand-tuned coefficients. The largest coefficients are associated with the S-score of tips left by vetted accounts—a sparse but strong signal —and the S-score of Pilgrim-generated visits. These two scores account for more than 60% of the final signal.

Figure 2: A screenshot from the Foursquare website for New York City.

4. SUMMARY AND RESULTS

The combined signal is now being used as a primary component of venue recommendations, and is showcased in the main Foursquare app and on the website in the form of weekly billboard-style "Trending This Week" lists in major metropolitan areas (Figure 2). It is also frequently covered in articles which feature best-of lists for many cities [1, 2, 3]. Weekly e-mails featuring these lists have click through rates that far exceed the industry average and drive regular in-app activity. The signal has also been integrated into Foursquare's core venue ratings algorithm resulting in greater freshness and turnover.

5. REFERENCES

[1] R. Bruner. 10 trendy Austin restaurants you need to try right now. http://www.businessinsider.com/the-hottest-restaurants-in-austin-tx-2016-3. Accessed: 2016-06-30.

[2] R. Bruner. 12 up-and-coming New York City restaurants you need to try right now. http://www.businessinsider.com.au/12-trendy-new-nyc-restaurants-to-try-now-2016-1. Accessed: 2016-06-30.

[3] R. Bruner. 15 trendy New York City restaurants you need to try right now. http://www.businessinsider.com/15-new-nyc-restaurants-to-try-now-2016-2. Accessed: 2016-06-30.

[4] A. Heath. Foursquare's location data is way more powerful than people realize. *Tech Insider*, January 2016.

[5] M. Sklar. Timely tip selection for Foursquare recommendations. In *RecSys Posters*, October 6–10 2014.

[6] M. Sklar, B. Shaw, and A. Hogue. Recommending interesting events in real-time with Foursquare check-ins. In *RecSys 2012 Poster Proceedings*, pages 311–312, September 9 2012.

[7] R. Tate. The brilliant hack that brought Foursquare back from the dead. *Wired*, December 2013.