Foursquare-based venue recommendations for groups

Vivek Bhagwat
vsb2110@columbia.edu

May 7, 2012

1 Abstract

When in a new area with a group of friends, it can be difficult to figure out what restaurant, bar, or café would be the best to go to — and which place you would like the most. Foursquare has a rich set of data about venues and data about where individuals like to go. Using this individual's history and information about the venue, a simple recommendation engine can be built. While foursquare has built a recommendation engine called "explore," this system has no ability to analyze data specific to the group. However, these data are not necessarily hard to collect — it simply requires all users involved supplying access to their checkin history. Once these data are acquired, similar techniques can be used to training on a group's data as an individual's.

2 Data

Foursquare has an easy-to-use API which allows access to what venues are nearby, checkin history, and information about those venues. The data used for this recommendation engine included users' checkin history, information about the category or categories of the venues, how popular each venue is, and which venues are nearby.

However, there exist certain key limitations to this recommendation system. Due to privacy concerns, foursquare does not provide access to the checkin history of friends. Without access to friends' checkin history, the amount of data on which a system can be trained is limited. This is a key advantage of foursquare's explore function — which does have access to such data. Given access to this data, the recommendation engine could easily train on significantly more data, because the system would essentially train on the extended circles of the group. Despite this, however, the group training does prove to take into account enough data, provided all participants have checked in to many different types of venues — so as to be able to provide data on any given query that is searched.

Another limitation with the data is that foursquare data almost exclusively provides one category per venue. This severely limits the more granular information that exists about venues such as what type of bar it might be, whether a restaurant is a casual one or not, and other pertinent information such as price range. Information like this might be scattered in tips for a venue, but the entity extraction involved would render the system needlessly complex for very simple and very fast recommendations.

Another limitation with the data is that evaluating metrics on the accuracy of the recomendations is not as simple as splitting the training data into a train/test split. This is because the amount an individual "likes" a venue is not available through data foursquare provides, except again through analysis of tips that user has left, which are often unreliable sources of information. As a result, it is difficult to evaluate whether a recommendation is a "good" recommendation or not. The metric of choice therefore becomes subjective, and based on human prediction.

3 System

The foursquare venue data provides information about the location of the venue, the category or categories it falls under, and statistics about the number of checkins. For the limited purposes of this project, the OAuth token necessary to obtain data for each user was obtained manually for each of the four users who gave permission to use their checkin data anonymously. Were this project to become more of a product, this process would have to be automated (which is indeed possible, but unnecessary for these purposes).

To train the system, the checkin history for each user is first downloaded. A feature vector and weights are then created, with a simple count of each category's `shortName` function defining the weights. The very simple concept of this system is that when a member of the group has checked in at a venue (e.g. a Chinese restaurant), that venue's category is weighted slightly more heavily. This works well for an individual, as it predicts venues based on what the user has claimed to be its preferences. For groups, similar concepts apply well. Similar interests are weighed more heavily, and heavy but outlying interests are diminished (this, of course, is another benefit to the group dynamic).

Popularity is a more interesting feature in the recommendation system. In order to not weigh popularity too much, if a venue is not popular that weight is discounted by a slightly larger amount. This naive normalization is likely the weakest point of this system, especially in that over few checkins there is a large sensitivity to popularity. However, over the limited training which was performed (due to a constrained amount of data available based on the privacy concerns mentioned above), this did not prove to be an issue. The weight on popularity was similar to a few other popular features.

Once all are trained, the weight vector has been defined. The venues nearby are again acquired through use of foursquare's search API. The test phase is then simple — the very sparse feature vector (out of approximately 100 features, only 2 or 3 are ever positive) is then evaluated for each venue nearby. This is a very quick operation because it involves checking whether or not the venue is deemed "popular" (the system simply evaluates this to be true if there are more than 300 checkins at this venue) and to which category or categories it belongs. The weights for each "true" feature are summed up for each venue, and ranked in descending order based on that sum to determine rank, using the following simple formula:

$$\sum_i w_i \times f_i, \tag{1}$$

where f is a feature vector of the venue being evaluated and w is the weights vector which has been trained.

4   Analysis

Overall, it is difficult to analyze the success of the recommendations produced in any automated, mathematical way because the evaluation of what makes a good recommendation relies on subjective feedback and exact responses to a recommender. However, a more cursory, subjective analysis shows the recommendations to fit well with the data it trained on, without overfitting to particular features. In general, bars are the preferred venue overall for the four users. The data for the weights for the groups and individuals are in the appendix. When looking at these weights, it is best to keep in mind that these are all on a relative scale based on number of checkins (people with more data and history are preferred). When looking at these weights, it is clear that bars are heavily preferred, followed by a few different types of restaurants — exactly what might be expected from New York City inhabitants.

This project could be further improved by acquiring significantly more data and having access to friend's checkin history. Popularity could be better adjusted, and normalization between group members might make for different recommendations than the ones this recommendation system provided.

5   Appendix

The data provided on the next page are of the format [(weight1, feature1), (weight2, feature2), ...]

GROUP
[(58.0, ('popular', 0)), (8.0, (u'Vietnamese', 1)), (1.0, (u'Opera House', 2)), (2.0, (u'Performing Arts', 3)), (15.0, (u'Breakfast / Brunch', 4)), (6.0, (u'Chinese', 5)), (60.0, (u'Bar', 6)), (18.0, (u'Burgers', 7)), (8.0, (u'Auditorium', 8)), (8.0, (u'Park', 9)), (87.0, (u'Tech Startup', 10)), (3.0, (u'Karaoke', 11)), (9.0, (u'Ramen / Noodles', 12)), (4.0, (u'Southern / Soul', 13)), (9.0, (u'New American', 14)), (2.0, (u'South American', 15)), (20.0, (u'Coffee Shop', 16)), (18.0, (u'Japanese', 17)), (11.0, (u'Quad / Commons', 18)), (7.0, (u'Administrative Building', 19)), (3.0, (u'Other - Entertainment', 20)), (7.0, (u'Corporate / Office', 21)), (7.0, (u'Bakery', 22)), (3.0, (u'Caf\xe9', 23)), (14.0, (u'Italian', 24)), (7.0, (u'Indian', 25)), (6.0, (u'Whisky Bar', 26)), (2.0, (u'Brewery', 27)), (7.0, (u'French', 28)), (10.0, (u'Pizza', 29)), (6.0, (u'Asian', 30)), (2.0, (u'Bookstore', 31)), (4.0, (u'Art Museum', 32)), (9.0, (u'Rock Club', 33)), (2.0, (u'Technology', 34)), (5.0, (u'Math', 35)), (17.0, (u'American', 36)), (113.0, (u'Home', 37)), (5.0, (u'Diner', 38)), (114.0, (u'Residence Hall', 39)), (5.0, (u'Tacos', 40)), (43.0, (u'Street Food', 41)), (1.0, (u'Zoo / Aquarium', 42)), (1.0, (u'Caribbean', 43)), (27.0, (u'Airport', 44)), (19.0, (u'Mexican', 45)), (6.0, (u'Thai', 46)), (20.0, (u'Other - Education', 47)), (2.0, (u'Billiards', 48)), (5.0, (u'Coworking Space', 49)), (4.0, (u'Beer Garden', 50)), (3.0, (u'Plaza / Square', 51)), (3.0, (u'Football', 52)), (6.0, (u'Wine Bar', 53)), (30.0, (u'Communications', 54)), (22.0, (u'Engineering', 55)), (2.0, (u'Concert Hall', 56)), (1.0, (u'Medical', 57)), (2.0, (u'Cineplex', 58)), (3.0, (u'Sports Bar', 59)), (1.0, (u'Baseball', 60)), (2.0, (u'Stadium', 61)), (7.0, (u'Hotel', 62)), (13.0, (u'Terminal', 63)), (1.0, (u'Bowling Alley', 64)), (9.0, (u'Cocktail', 65)), (4.0, (u'Other - Food', 66)), (3.0, (u'Dive Bar', 67)), (1.0, (u'Racetrack', 68)), (2.0, (u'Spanish', 69)), (3.0, (u'Gay Bar', 70)), (5.0, (u'Pub', 71)), (1.0, (u'Beach', 72)), (13.0, (u'Grocery Store', 73)), (2.0, (u'Furniture / Home', 74)), (5.0, (u'BBQ', 75)), (3.0, (u'Nightclub', 76)), (3.0, (u'Korean', 77)), (3.0, (u'Tapas', 78)), (1.0, (u'Highway / Road', 79)), (9.0, (u'Sandwiches', 80)), (1.0, (u'Malaysian', 81)), (3.0, (u"Dentist's Office", 82)), (1.0, (u'Australian', 83)), (2.0, (u'Fried Chicken', 84)), (3.0, (u'Sushi', 85)), (1.0, (u'Greek', 86)), (1.0, (u'Ski Area', 87)), (1.0, (u'Gastropub', 88)), (2.0, (u'Gourmet', 89)), (2.0, (u'Flea Market', 90)), (1.0, (u'Cuban', 91)), (2.0, (u'Gym', 92)), (2.0, (u'Other - Nightlife', 93)), (7.0, (u'Yoga Studio', 94)), (1.0, (u'Scenic Lookout', 95)), (1.0, (u'Monument / Landmark', 96)), (1.0, (u'Hostel', 97)), (1.0, (u'Hiking Trail', 98)), (4.0, (u'Other Outdoors', 99)), (2.0, (u'Event Space', 100)), (1.0, (u'Eastern European', 101)), (4.0, (u'Apparel', 102)), (11.0, (u'Deli / Bodega', 103)), (1.0, (u'Vegetarian / Vegan', 104)), (1.0, (u'Sporting Goods', 105)), (2.0, (u'Scandinavian', 106)), (2.0, (u'Donuts', 107)), (3.0, (u'Art Gallery', 108)), (1.0, (u'Middle Eastern', 109)), (1.0, (u'Mac & Cheese', 110)), (3.0, (u'Post Office', 111)), (24.0, (u'Science', 112)), (1.0, (u'Science Museum', 113)), (1.0, (u'Ice Cream', 114)), (2.0, (u'Record Shop', 115)), (12.0, (u'Library', 116)), (1.0, (u'Shoes', 117)), (2.0, (u'Salon / Barbershop', 118)), (1.0, (u'Paper / Office', 119)), (3.0, (u'Train Station', 120)), (1.0, (u'Video Games', 121)), (2.0, (u'Department Store', 122)), (14.0, (u'Student Center', 123)), (1.0, (u'Lab', 124)), (2.0, (u'Other - Shop', 125)), (1.0, (u'Candy Store', 126)), (1.0, (u'Boutique', 127)), (2.0, (u'Classroom', 128)), (1.0, (u'Subway', 129)), (1.0, (u'Molecular Gastronomy', 130)), (1.0, (u'Other - Buildings', 131)), (1.0, (u'Sake Bar', 132)), (1.0, (u'Convenience Stores', 133)), (1.0, (u'Arts & Entertainment', 134)), (2.0, (u'Pharmacy', 135)), (27.0, (u'Arts', 136)), (18.0, (u'Academic Building', 137)), (1.0, (u'Desserts', 138)), (1.0, (u'Butcher', 139)), (1.0, (u'Laundry', 140)), (1.0, (u'Church', 141)), (1.0, (u'Bank / Financial', 142)), (1.0, (u'Cupcakes', 143)), (1.0, (u'Electronics', 144))]

INDIVIDUALS
[(20.25, ('popular', 0)), (2.0, (u'Vietnamese', 1)), (1.0, (u'Opera House', 2)), (1.0, (u'Performing Arts', 3)), (8.0, (u'Breakfast / Brunch', 4)), (1.0, (u'Chinese', 5)), (16.0, (u'Bar', 6)), (9.0, (u'Burgers', 7)), (3.0, (u'Auditorium', 8)), (3.0, (u'Park', 9)), (35.0, (u'Tech Startup', 10)), (1.0, (u'Karaoke', 11)), (2.0, (u'Ramen / Noodles', 12)), (1.0, (u'Southern / Soul', 13)), (2.0, (u'New American', 14)), (1.0, (u'South American', 15)), (3.0, (u'Coffee Shop', 16)), (3.0, (u'Japanese', 17)), (6.0, (u'Quad / Commons', 18)), (2.0, (u'Administrative Building', 19)), (2.0, (u'Other - Entertainment', 20)), (2.0, (u'Corporate / Office', 21)), (3.0, (u'Bakery', 22)), (1.0, (u'Caf\xe9', 23)), (3.0, (u'Italian', 24)), (2.0, (u'Indian', 25)), (1.0, (u'Whisky Bar', 26)), (1.0, (u'Brewery', 27)), (1.0, (u'French', 28)), (3.0, (u'Pizza', 29)), (2.0, (u'Asian', 30)), (1.0, (u'Bookstore', 31)), (1.0, (u'Art Museum', 32)), (1.0, (u'Rock Club', 33)), (1.0, (u'Technology', 34)), (1.0, (u'Math', 35)), (2.0, (u'American', 36)), (97.0, (u'Home', 37)), (1.0, (u'Diner', 38)), (2.0, (u'Residence Hall', 39)), (1.0, (u'Tacos', 40)), (2.0, (u'Street Food', 41)), (1.0, (u'Zoo / Aquarium', 42)), (1.0, (u'Caribbean', 43)), (6.0, (u'Airport', 44)), (3.0, (u'Mexican', 45)), (1.0, (u'Thai', 46)), (4.0, (u'Other - Education', 47)), (1.0, (u'Billiards', 48)), (2.0, (u'Coworking Space', 49)), (1.0, (u'Beer Garden', 50)), (1.0, (u'Plaza / Square', 51)), (1.0, (u'Football', 52)), (1.0, (u'Wine Bar', 53)), (2.0, (u'Communications', 54)), (1.0, (u'Engineering', 55)), (1.0, (u'Concert Hall', 56)), (1.0, (u'Medical', 57)), (1.0, (u'Cineplex', 58))]

[(245.25, ('popular', 0)), (2.0, (u'Asian', 1)), (1.0, (u'Performing Arts', 2)), (2.0, (u'Sports Bar', 3)), (8.0, (u'Mexican', 4)), (1.0, (u'Baseball', 5)), (41.0, (u'Bar', 6)), (2.0, (u'Stadium', 7)), (3.0, (u'Coworking Space', 8)), (8.0, (u'American', 9)), (15.0, (u'Home', 10)), (4.0, (u'Hotel', 11)), (5.0, (u'Wine Bar', 12)), (11.0, (u'Terminal', 13)), (4.0, (u'Coffee Shop', 14)), (8.0, (u'Airport', 15)), (1.0, (u'Indian', 16)), (1.0, (u'Bowling Alley', 17)), (4.0, (u'Italian', 18)), (1.0, (u'Brewery', 19)), (8.0, (u'Cocktail', 20)), (3.0, (u'Beer Garden', 21)), (2.0, (u'Street Food', 22)), (2.0, (u'Other - Food', 23)), (3.0, (u'Dive Bar', 24)), (2.0, (u'Burgers', 25)), (1.0, (u'Racetrack', 26)), (5.0, (u'Whisky Bar', 27)), (2.0, (u'Spanish', 28)), (3.0, (u'Thai', 29)), (2.0, (u'Karaoke', 30)), (3.0, (u'Gay Bar', 31)), (5.0, (u'New American', 32)), (3.0, (u'Pizza', 33)), (2.0, (u'Vietnamese', 34)), (3.0, (u'Japanese', 35)), (1.0, (u'Park', 36)), (4.0, (u'Pub', 37)), (1.0, (u'Beach', 38)), (2.0, (u'Grocery Store', 39)), (2.0, (u'Furniture / Home', 40)), (1.0, (u'South American', 41)), (44.0, (u'Tech Startup', 42)), (3.0, (u'BBQ', 43)), (4.0, (u'Chinese', 44)), (3.0, (u'Nightclub', 45)), (6.0, (u'Rock Club', 46)), (3.0, (u'Korean', 47)), (1.0, (u'Tapas', 48)), (1.0, (u'Football', 49)), (1.0, (u'Highway / Road', 50)), (5.0, (u'Sandwiches', 51)), (1.0, (u'Malaysian', 52)), (2.0, (u"Dentist's Office", 53)), (1.0, (u'Ramen / Noodles', 54)), (1.0, (u'Australian', 55)), (1.0, (u'Fried Chicken', 56)), (1.0, (u'Sushi', 57)), (1.0, (u'Cineplex', 58)), (1.0, (u'Greek', 59)), (1.0, (u'Ski Area', 60)), (1.0, (u'Gastropub', 61)), (2.0, (u'Gourmet', 62)), (1.0, (u'Diner', 63)), (2.0, (u'Flea Market', 64)), (1.0, (u'Cuban', 65)), (2.0, (u'Gym', 66)), (2.0, (u'Other - Nightlife', 67)), (7.0, (u'Yoga Studio', 68)), (1.0, (u'Plaza / Square', 69)), (1.0, (u'Scenic Lookout', 70)), (1.0, (u'Monument / Landmark', 71)), (1.0, (u'Hostel', 72)), (1.0, (u'Hiking Trail', 73)), (2.0, (u'Other Outdoors', 74)), (1.0, (u'French', 75)), (2.0, (u'Event Space', 76)), (1.0, (u'Billiards', 77)), (1.0, (u'Eastern European', 78)), (1.0, (u'Apparel', 79)), (1.0, (u'Deli / Bodega', 80)), (1.0, (u'Other - Entertainment', 81))]

[(147.75, ('popular', 0)), (1.0, (u'Vegetarian / Vegan', 1)), (10.0, (u'Street Food', 2)), (6.0, (u'Coffee Shop', 3)), (1.0, (u'Fried Chicken', 4)), (2.0, (u'Southern / Soul', 5)), (1.0, (u'New American', 6)), (1.0, (u'Other - Food', 7)), (5.0, (u'Japanese', 8)), (1.0, (u'Sporting Goods', 9)), (2.0, (u'Scandinavian', 10)), (3.0, (u'Bar', 11)), (4.0, (u'Burgers', 12)), (3.0, (u'Corporate / Office', 13)), (4.0, (u'Bakery', 14)), (2.0, (u'Caf\xe9', 15)), (5.0, (u'Italian', 16)), (1.0, (u'Diner', 17)), (1.0, (u'Vietnamese', 18)), (1.0, (u'BBQ', 19)), (2.0, (u'Park', 20)), (3.0, (u'Sandwiches', 21)), (2.0, (u'Indian', 22)), (4.0, (u'Ramen / Noodles', 23)), (2.0, (u'Donuts', 24)), (3.0, (u'Art Gallery', 25)), (3.0, (u'Tacos', 26)), (1.0, (u'Quad / Commons', 27)), (1.0, (u'Middle Eastern', 28)), (1.0, (u'Concert Hall', 29)), (6.0, (u'Mexican', 30)), (1.0, (u'Breakfast / Brunch', 31)), (2.0, (u'Rock Club', 32)), (1.0, (u'Mac & Cheese', 33)), (2.0, (u'Tapas', 34)), (2.0, (u'Other - Education', 35)), (6.0, (u'Deli / Bodega', 36)), (3.0, (u'Apparel', 37)), (5.0, (u'Tech Startup', 38)), (3.0, (u'Post Office', 39)), (12.0, (u'Residence Hall', 40)), (2.0, (u'Asian', 41)), (1.0, (u'Sports Bar', 42)), (2.0, (u'Other Outdoors', 43)), (2.0, (u'Science', 44)), (1.0, (u'Football', 45)), (2.0, (u'Sushi', 46)), (2.0, (u'Auditorium', 47)), (3.0, (u'French', 48)), (2.0, (u'Terminal', 49)), (1.0, (u'Science Museum', 50)), (1.0, (u'Ice Cream', 51)), (1.0, (u'Bookstore', 52)), (9.0, (u'Airport', 53)), (2.0, (u'Record Shop', 54)), (3.0, (u'Library', 55)), (1.0, (u'Shoes', 56)), (2.0, (u'Salon / Barbershop', 57)), (1.0, (u'Paper / Office', 58)), (3.0, (u'Train Station', 59)), (1.0, (u"Dentist's Office", 60)), (1.0, (u'Pub', 61)), (1.0, (u'Video Games', 62)), (2.0, (u'Department Store', 63)), (1.0, (u'Plaza / Square', 64)), (1.0, (u'Administrative Building', 65)), (1.0, (u'Student Center', 66)), (1.0, (u'Engineering', 67)), (1.0, (u'Lab', 68)), (1.0, (u'Other - Shop', 69)), (1.0, (u'Chinese', 70)), (1.0, (u'Candy Store', 71)), (1.0, (u'Boutique', 72)), (2.0, (u'American', 73)), (1.0, (u'Classroom', 74)), (1.0, (u'Subway', 75)), (1.0, (u'Molecular Gastronomy', 76)), (3.0, (u'Hotel', 77)), (1.0, (u'Grocery Store', 78)), (1.0, (u'Other - Buildings', 79)), (2.0, (u'Communications', 80)), (1.0, (u'Sake Bar', 81)), (1.0, (u'Convenience Stores', 82)), (1.0, (u'Cocktail', 83)), (2.0, (u'Math', 84)), (1.0, (u'Arts & Entertainment', 85)), (1.0, (u'Pharmacy', 86)), (1.0, (u'Technology', 87))]

[(284.75, ('popular', 0)), (100.0, (u'Residence Hall', 1)), (10.0, (u'Grocery Store', 2)), (29.0, (u'Street Food', 3)), (2.0, (u'Ramen / Noodles', 4)), (7.0, (u'Coffee Shop', 5)), (4.0, (u'Pizza', 6)), (3.0, (u'Vietnamese', 7)), (1.0, (u'Sandwiches', 8)), (2.0, (u'Indian', 9)), (2.0, (u'Park', 10)), (4.0, (u'Deli / Bodega', 11)), (3.0, (u'Burgers', 12)), (27.0, (u'Arts', 13)), (18.0, (u'Academic Building', 14)), (20.0, (u'Engineering', 15)), (26.0, (u'Communications', 16)), (1.0, (u'Southern / Soul', 17)), (7.0, (u'Japanese', 18)), (4.0, (u'Administrative Building', 19)), (2.0, (u'Italian', 20)), (13.0, (u'Student Center', 21)), (4.0, (u'Quad / Commons', 22)), (22.0, (u'Science', 23)), (14.0, (u'Other - Education', 24)), (3.0, (u'Art Museum', 25)), (1.0, (u'BBQ', 26)), (6.0, (u'Breakfast / Brunch', 27)), (1.0, (u'Desserts', 28)), (2.0, (u'Diner', 29)), (9.0, (u'Library', 30)), (2.0, (u'Thai', 31)), (4.0, (u'Airport', 32)), (1.0, (u'Other - Shop', 33)), (1.0, (u'Butcher', 34)), (5.0, (u'American', 35)), (3.0, (u'Auditorium', 36)), (1.0, (u'Pharmacy', 37)), (2.0, (u'Mexican', 38)), (2.0, (u'French', 39)), (1.0, (u'Laundry', 40)), (3.0, (u'Tech Startup', 41)), (1.0, (u'Home', 42)), (1.0, (u'Church', 43)), (1.0, (u'Classroom', 44)), (1.0, (u'New American', 45)), (2.0, (u'Corporate / Office', 46)), (1.0, (u'Bank / Financial', 47)), (1.0, (u'Cupcakes', 48)), (1.0, (u'Tacos', 49)), (1.0, (u'Other - Food', 50)), (1.0, (u'Electronics', 51)), (2.0, (u'Math', 52))]