Title: Strength in numbers? Modelling the impact of businesses on each other

In many cities, there is a small number of streets with a lot of restaurants. Being in a street like this is a double-edged sword for the individual restaurant. On one side, it is valuable because it gets them the attention of potential customers for free. On the other hand, the restaurants are competing for customers with similar needs and the offerings are not free from overlap. Our hypothesis is that this balance is a function of the size of the cluster. The bigger the cluster grows, the more customers start to search for a restaurant within this cluster, increasing the number of total customers of the cluster and emphasizing the symbiosis of the restaurants living in the cluster. At the same time, however, the number of potential customers is limited and as it nears saturation, the competitive nature of the relationship between the restaurants grows stronger.

The goal of our project is to test this hypothesis by finding clusters of restaurants in a dataset and to model how the individual restaurant is affected by being part of the cluster. Machine learning will play a crucial role in this process, from finding clusters of restaurants to model the impact on the individual restaurant. We will start by comparing clusters of different sizes but ideally we would like to find clusters where we can observe growth over time and the effect on the restaurants over this time period. If the project proves to be successful, it has the potential to offer a unique insight in the relationship between restaurants in such clusters.

To design and evaluate our model, we are planning to use the Yelp dataset. Yelp is a website where users review businesses like restaurants. The dataset contains data for more than 40’000 businesses and more than 1 million of reviews. There is a rich set of attributes for each business and there is additional data like the number of checkins that can be used to model the popularity of a place.

Resources:

http://link.springer.com/chapter/10.1007%2F978-3-642-25085-9\_76

**Ideas for proposal:**

1- Influence of businesses on each other.

2- Business clusters geographically.

3- Patterns in check-in’s and the and stars of reviews based on the time and crowdedness of place.

**Proposal:**

Title: Detecting sudden changes in a time series of reviews

Review for businesses like restaurants are not static but can change over time due to certain events. One can imagine that the cook changed and the food is much better after the change, leading to increased ratings. The change need not be solely due to the internal modifications. As an example, if the place of interest is a place selling bubble tea and a good taiwanese restaurant has opened recently in the neighborhood, the customers of the restaurant might visit the tea place after having eaten food. This synergetic effect contributes to the popularity of the tea shop. Another example in the opposite end might be that a similar restaurant opens close to a given restaurant but offers better quality for the same price, decreasing the ratings of the first restaurant.

Given a series of reviews, the goal of our project is to identify sudden changes in a time series of reviews. In this context, a review is not only a single number but a written text by customers because a good model would not only detect changes in star ratings but also in the kind of words that are used by the users to review the place. The intuition is that if for example a new place close-by offers better prices, it is likely that reviewers will start to use the word “expensive” more often in their reviews, which indicates a change of opinion of the reviewers. For the owner of a business, it is crucial to know when a sudden change happens in order to correlate it with events and react to it. We hope that the model that we are going to design for this project will be a first step towards offering such a tool to owners of a business.

We design and evaluate our model based on the Yelp dataset, which contains over 1 million reviews and for more than 40’000 businesses. Yelp is a website where users can review businesses like restaurants. The dataset is optimal for our project because reviews are not restricted to simple star ratings but consist of written comments of users.