Study of People's Eat-out Behaviour using Natural Language Processing on Tweets

Abul Hasan Fahad, Syed Anjoom Iqbal, Mohamed Hassan Abdalla
{ahfahad, saiqbal, mhabdall}@uwaterloo.ca
Group 36, ECE 657A, Winter 2019, University of Waterloo, Ontario Canada

1. Project Description

This is an "application-oriented" project where the team will use Natural Language Processing (NLP) and Machine Learning (ML) techniques to extract knowledge and make prediction on people's eat-out behaviour by analyzing their tweets. Questions that will be attempted to answer in this project are below.

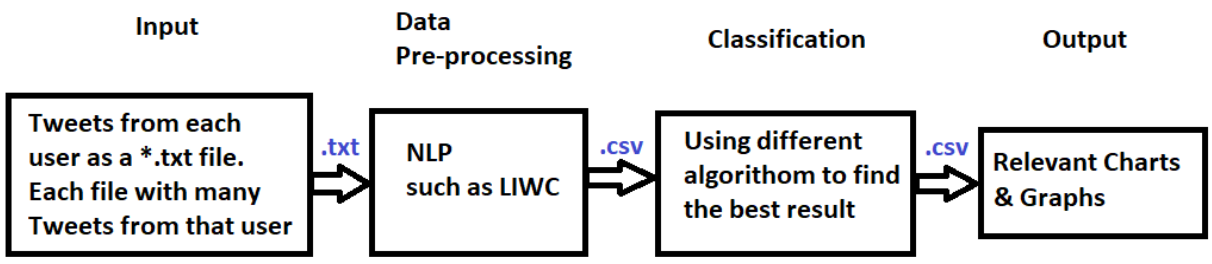
1. Around what time of the day, is typically a user online on Twitter portal and around what time does that user usually tweet related to eating-out? From these answers, we can develop knowledge around what time of the day a person is usually going to or ordering from a restaurant. We can also know if someone is going for breakfast, lunch or dinner. If someone is enthusiastic enough to tweet about any meal of a restaurant, we can promote relevant restaurant ads when that person is on the twitter portal. This is personalized and targeted marketing. For simplicity, we are assuming that people tweet about eating-out when they are making a purchase from a restaurant and not few hours after the meal. We are also assuming all the Tweets are from people who are living in a specific city and not all over the world.

2. How is the price sensitivity of a user? What type of restaurants does a user prefer to go to in terms of price? We can classify the restaurants as cheap, moderate, expensive and very expensive. By performing Natural Language Processing (NLP) and Linguistic Inquiry and Word Count (LIWC) we can correlate a user with different classes of restaurants. Then we can promote the ads of different classes of restaurants to the right market with appropriate price sensitivity. That will increase is probability of generating a revenue from an ads.

2. Review of Academic Papers

Twitter is a major social-media platform of communication among users on the web. This platform allows users to express their opinions and share information with others via tweets (micro-blogs). In addition, location-based social networking sites such as Foursquare have become popular, enabling users to publish their visited places through check-ins [1]. A Foursquare check-in consists of latitude, longitude, the name and the category of the venue, and the time of the check-in. Accordingly, Social media researchers are being able to find many interesting insights, such as- personality, value, and preferences of users by analyzing the texts of these tweets and check-ins [2,3,4]. Drawing inspiration from these works, authors in [5] collected tweets with Foursquare check-ins of 731 Twitter users. They first figured out each user's pattern of visiting four different categories of restaurants, and then built a model that correlates between user's word use in tweets and visiting frequency in different categories of restaurants. Finally, they proposed a prediction model, which can predict of a person's eat-out preference by analyzing the tweets with Foursquare check-ins. In this project, the team will try to reproduce the works of this paper [5].

3. Sketch of planned approach



1. Performing linguistic analysis: We will compute LIWC (Linguistic Inquiry and Word Count) categories of words from users' tweets
2. Computation of visiting different categories of restaurants: We will compute visiting frequencies of a user to different categories (cheap, moderate, expensive) of restaurants
3. Temporal Eat-out Behaviour Analysis: We will create time-series profile of each user based on restaurant visit
4. Prediction: We will build prediction model to predict a user's eat-out behaviour from the analysis of tweets

4. Expected Challenges & Difficulties

1. Lack of experience with Natural Language Processing techniques
2. Lack of experience with appropriate classification algorithm selection

5. Description of Dataset

The Dataset was acquired from the authors of [5]. The dataset preparation method is outlined below:

1. Authors used Twitter advanced search technique to find users whose English tweets contain Foursquare links. If a user uses Foursquare links, the check-ins of her tweets usually contain keywords such as "4sq", and "Foursquare". To find out required Twitter users, they searched with these two keywords through advanced search technique.
2. After selecting Twitter id of users who regularly tweets using Foursquare links, authors collected their tweets through http://greptweet.com/. They collected tweets of 731 users. Then they used HTML parsing to gather information about the place or restaurant found in the link.
3. Finally the input dataset consists of 731 *.txt files. File name is the Twitter handle of that user. Each file has hundreds of Tweets from that user with a timestamp.

6. References

1. Henriette Cramer , Mattias Rost , Lars Erik Holmquist, Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare, Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, August 30-September 02, 2011, Stockholm, Sweden
2. Jilin Chen , Gary Hsieh , Jalal U. Mahmud , Jeffrey Nichols, Understanding individuals' personal values from social media word use, Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, February 15-19, 2014, Baltimore, Maryland, USA
3. Kenneth Joseph , Chun How Tan , Kathleen M. Carley, Beyond "local", "categories" and "friends": clustering foursquare users with latent "topics", Proceedings of the 2012 ACM Conference on Ubiquitous Computing, September 05-08, 2012, Pittsburgh, Pennsylvania
4. Hamshaw RJT, Barnett J, Lucas JS. Tweeting and Eating: The Effect of Links and Likes on Food-Hypersensitive Consumers' Perceptions of Tweets. Front Public Health. 2018;6:118. Published 2018 Apr 23.
5. Md. Mahabur Rahman , Md Taksir Hasan Majumder , Md Saddam Hossain Mukta , Mohammed Eunus Ali , Jalal Mahmud, Can we predict eat-out preference of a person from tweets?, Proceedings of the 8th ACM Conference on Web Science, May 22-25, 2016, Hannover, Germany