Business Recommendation from Reviews

Hemang Bansal
Department of Computer Science,
University of Colorado, Boulder
hemag.bansal@colorado.edu

Prashil Bhimani
Department of Computer Science,
University of Colorado, Boulder
prashil.bhimani@colorado.edu

Sandesh Dhawaskar
Department of Computer Science,
University of Colorado, Boulder
sadh0344@colorado.edu

ABSTRACT

With increase in listing of businesses online and reviewing websites, it is the need of the hour to have accurate and relevant recommendation system for users to advise the users about the most relevant business based on his reviews. This increases the accuracy and efficiency of the recommendation system and enhances user satisfaction. Past reviews given by users and reviews for businesses helps in understanding user preferences and match it with the appropriate businesses the user might be interested in.

CCS CONCEPTS

• **Computer Science → Data Preprocessing, Data Mining, Natural Language Processing**

KEYWORDS

Business recommendation, recommendation system, user reviews, business reviews.

1 Motivation

In the last decade, there has been an increase of local businesses marking their online presence. According to a research[8] over 54% of small businesses had a website and another 24% were planning to build one in 2016. This boom in the number of businesses marketing online, has also promoted the growth in number of business like Yelp, Expresit, Google Reviews etc. that recommend these businesses to the user. These websites offer users a seamless collection of businesses in their area and recommend the users about the most relevant business according to his needs. In order to make the usage of such listing websites more convenient and personalized recommendations play a very important role. The better the recommendation the better are the chances that the user will view that business. With such abundance of data of user preferences and business reviews we can use data mining and natural language processing to recommend the users with the best possible recommendations.

2 Literature Review

Various researches are being done in the field of building a recommendation system. User ratings, followers, friends, reviews, behavior, likes and dislikes are few main factors being considered, when recommendation systems are built. Weighted hybrid recommendation for heterogeneous networks [1] like Social media sites, is a good example where they collect very large number of attributes like ratings, friend's interests, followers, posts, media and others to build the component based hybrid model whose components can be reused for multiple recommendation tasks instead of building a single multi-faceted data as heterogeneous network. This would reduce regenerating the computationally intensive model to build the single purpose recommendation system. Network based methods are used for generating the recommendations. Mining the text from user reviews is one of the most common techniques used in building the recommendation system. Consider the TriRank: Review-aware Explainable Recommendation by Modeling Aspects [2], where they extract the specific attributes of items using the reviews in text and build a "user-item-aspect" ternary relationship. Then they model ternary relation as a tri partite heterogeneous graph, by making recommendation task as one of the vertex ranking. And then devise a generic algorithm for ranking on tripartite graphs called Tri Rank and specialize the same for personalized recommendations. Building the recommendation system using the user behavior is also one unique way to build the recommendation system.Consider the Recommendation via User's Personality and Social Contextual [3], where they have considered, three social factors, personal interest, interpersonal interest similarity and interpersonal influence, fused into a unified personalized recommendation model based on probabilistic matrix factorization. Amazon is using item-to-item collaborative filtering to target what to advertise the user based on users likes, reviews, past purchases and items in virtual cart [7]. This work takes the consideration of user's behavior and could be developed by using the user's location as one of the factor. There are also works where in

instead of using similar attributes of one single user, different factors of user system can also be used to build a recommendation system. Work by Personalized Entity Recommendation: A Heterogeneous Information Network Approach [4] proves that a recommendation at both global and personalized levels can be efficiently generated on the fly to provide high quality personalized recommendation system results compared to other recommendation methods. There are very large numbers of recommendation system that are built using user's review. Yelp has built a multiple recommendation systems where reviews are used to give a star rating to businesses, higher the rating more the chances of business being recommended [6].There are even works done which compare the multiple recommendation systems. Comparative Recommender System Evaluation: Benchmarking Recommendation Frameworks [5], where they use three publicly available datasets to evaluate user-based and item-based nearest neighbor CF and SVD-based matrix factorization RS.

3 Proposed Work

The dataset we are using is an open source dataset provided by Yelp available at www.yelp.com/dataset. Yelp is one of the biggest local business listing website which has over 156k listed businesses and 11 million users. The dataset has data from over 12 metropolitan cities and has over 47 million reviews about various businesses. We will generate recommendations from the results we get on applying our model to the Yelp Dataset. The work is divided in 4 subtasks:

3.1 Data Preprocessing
Every dataset has some outliers and noise in it. It may be due to technical errors or incorrect user inputs. Such outliers can skew the results and hence they must be removed properly. The dataset may have reviews by some users in languages other than English. Such reviews have to be filtered out to use text mining (Accuracy can be increased by mining the meanings of the words that are used instead). The dataset may also have some noise in it. A review may have an incorrect business id or user id. The user may have incorrect id for the list of friends. The businesses might have incorrect attributes or location. Such noise and outliers need to be eliminated.

3.2 Generating Tags for Users and Business
This is an important subtasks because if the tags are useful then the recommendations will be accurate. We plan to add weights to the tags to create a ranking order. The user tags will be generated by using the following methods:
1. Past reviews written by the user for different businesses.

2. Tags generated for friends of users can also signify preferences the user might be interested in.
3. Tags for other users who have similar tags as the given user.

The tags will have weights attached to them and a positive weight implies the user will prefer businesses with similar tags and higher the weight more the likeliness.

For businesses dataset we are planning to generate tags by using the following methods:
1. Reviews by users for the particular business
2. Business attributes available in the dataset

The tags for businesses will also be weighted and positive weights will indicate that the business has positive reviews for the tag.

3.3 Mining Recommended Business for Users
After the tags are generated we can match the tags from Users and Businesses to recommend the most apt business for users. We plan to use python libraries like scikit, numpy and pandas to generate the results. We can use cosine distances or Manhattan distances to generate the similarity between user preference and business tags.

3.4 Results

We can calculate the accuracy of the predictions by using maximizing true results and minimizing false negatives. False positives are not taken into consideration as a user may have been to a place but not checked-in/ written a review about it.

4 Milestones
We have planned the following milestones for the projects

Goal	Proposed Deadline
Data preprocessing (Filtering, Outlier detection and noise removal)	10/19/2017
Gathering relevant and necessary knowledge required for the project	10/26/2017
Generating tags and preferences	11/16/2017
Recommendations & Accuracy	12/7/2017

5 EVALUATIONS

The prediction accuracy of the recommendation system will be calculated by maximizing true results and minimizing false negatives. False positives are not taken into consideration as a user may have been to a place but not checked-in or written a review about it. This will reduce the sparsity of the data. The actual results can be calculated by generating a check-in (visited) tag using the the userId and businessId i.e if a user review has a particular business ID that means he has visited the business. These actual results will be compared to the prediction results to get a measure of the accuracy.

ACKNOWLEDGMENTS

REFERENCES

[1]. Fatemeh Vahedian: Weighted hybrid recommendation for heterogeneous networks: Proceedings of the 8th ACM
Conference on Recommender systems.

[2]. Xiangnan He,Tao Chen,Min-Yen Kan, Xiao ChenTriRank: Review-aware Explainable Recommendation by Modeling Aspects: CIKM '15 Proceedings
of the 24th ACM International on Conference on Information and Knowledge Management.

[3]. He Feng, Xueming Qian: Recommendation via User's Personality and Social Contextual: CIKM '13 Proceedings of the 22nd
ACM international conference on Information & Knowledge Management.

[4]. Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu,
Bradley Sturt, Urvashi Khandelwal, Brandon Norick, Jiawei Han:Personalized Entity Recommendation: A Heterogeneous Information Network Approach:
WSDM 2014 - Proceedings of the 7th ACM International Conference on Web Search and Data Mining

[5].Alan S,Alejandro Bellogín: Comparative Recommender System Evaluation: Benchmarking Recommendation Frameworks:
RecSys '14 Proceedings of the 8th ACM Conference on Recommender systems

[6] Phil Chen and Dan Posch. Collaborative filtering on very sparse graphs a recommendation system for yelp.com, November 2012.

[7] G.D. Linden, J.A. Jacobi, and E.A. Benson, *Collaborative Recommendations Using Item-to-Item Similarity Mappings*, US Patent 6,266,649 (to Amazon.com), Patent and Trademark Office, Washington, D.C., 2001.

[8] Small Business Websites in 2016: A Survey, *https://clutch.co/web-designers/resources/small-business-websites-2016-survey*.