Data Mining Techniques for Recommending Stores to Shopping Malls

Purnima Padmanabhan
Department of Computer Science
New York University
Courant Institute
Email: pp1492@nyu.edu

John Wu
Department of Mathematics
New York University
Courant Institute
Email: jsw452@nyu.edu

Ying Zhang
Department of Computer Science
New York University
Courant Institute
Email: zy674@nyu.edu

***Abstract*—In this paper, we try to recommend stores for malls. We do this by using using various traditional recommendation techniques as well as One-Class data techniques. We compare each of these techniques and combine them together to make an ensemble of recommendation techniques.**

I. Introduction

It is of great interest for shopping mall owners to determine what store should be selected to fit an empty slot in a (target) shopping mall so that the shopping mall realizes maximum profit from this decision. We have a dataset of all the malls in the USA along with the county information of the malls [1], and all the stores of various categories within each mall. Each store falls into three groups: category (restaurant, appeals-to-parents, clothing, etc), average rating of all users and how high-ended each store in a category is (Armani, Louis Vuitton are high-ended stores in apparels, we can consider fast-food joints as low-ended stores and fine-dining is high-ended and so on). For example, a mall in a fancy neighborhood in Manhattan might want to fill an empty slot with a Michael Kors outlet, while a mall in a modest neighborhood in New Mexico might make more profit by renting out the space to a Tex-Mex food chain instead. Intuitively, the store that needs to fill in the empty slot will depend on various factors.

In this paper, we will use a series of recommendation system techniques to account these models and ultimately use these techniques to predict whether a store is suitable for a mall. The malls and stores can be seen as users and items in the context of the current recommendation system literature. We will apply One-Class User-based Collaborative Filtering (OCUCF) [2], One-Class Item-based Collaborative Filtering (OCICF)[2], Content-Based Recommender Systems [3] along with Weighted Alternating Least Squares [4] with various features to recommend stores with malls. Inherently, these models and have their own strengths and weaknesses [3]. We combined these models together using Simulated Annealing as a ensemble technique [5].

We will also discuss the performance of each these models using Root Mean-Squared Error (RMSE)and Mean Average Precision (MAP) to compare all of these models.

II. Motivation

As previously stated, it is important for mall owners to decide which stores will be appropriate for their malls. Intuitively, it would be best to recommend stores that are similar by some metric of current stores in a mall. Also, it is useful to recommend complements of stores in order to attract more customers. Techniques such as user-based collaborative filtering can utilize store and mall information to find the inter-similarity of stores and malls. This inter-similarity can be used to appropriate stores to malls. Also, techniques such content-based recommendation systems can find the similarity between stores and malls to recommend stores that are similar to a mall itself.

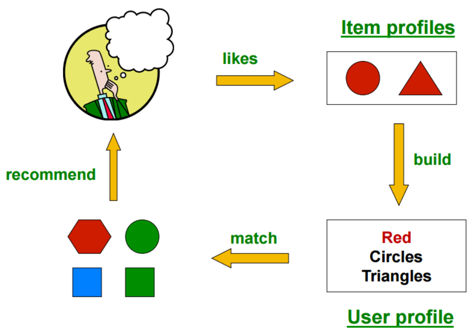
Additionally, these recommendation systems are not only useful for shopping mall managers, but it is also recommend stores for individuals as well. Individuals can be represented as malls with their favorite stores as stores contained in an arbitrary mall and their geographical location as the malls location. Using the proposed ensemble recommendation system can inform users of possible stores they have not seen before and can be used in popular websites, such as Yelp.

III. Related Work

This work is an extension of the work done by [6]. [6] used the pairwise distance between malls as comparison features for creating a user-based collaborative filter to predict the number of stores of a certain category. We extend upon this by actually predicting an appropriate store for a mall. [3] discusses the underlying basis of recommendation system. Particularly, [3] discusses how traditional recommendation systems impute the values of missing data. For our dataset, we are not dealing with this issue since our dataset is a One-Class dataset, a dataset where all entries are binary and all every entries are filled in [4]. Negative examples and missing positive examples are embedded as 0s. Our goal is to change appropriate 0s as 1s using techniques such as WLAS [4] and traditional recommendation system techniques in the context of One-Class problems. [2] A popular technique that can be used to recommend stores to malls are supervised learning for regression. In the context of real-estate, [7] uses features that have been selected to measure the housing value of the Boston suburb by using techniques such as Support Vector Machines (SVM), Partial Least Squares (PLS), and Least Squares Support Vector Machine (LSSVM). Seeing the performance of these techniques, we implemented Random Forest as baseline method to compare our ensemble technique. Using appropriate features for our recommendation models are important. [8] uses median-price indices, repeat-sales indices, hedonic indices, and stock-market-based indices to predict real estate prices. Similarly, for this project, we used demographic data [9], which consists of the racially composition of each county, as well as the industry data [1] of each county. These county information are used as features for the OCUCF model to represent the user features of the malls. Additionally, Non-Negative Matrix Factorization (NMF) has been proven useful to obtain latent topics of data points and its features as well as reducing the noise [10]. For this project, we used NMF to obtain latent representation of malls on the data-set containing the store category of malls. We used these features for the OCUCF model.

IV. Design and Implementation

Diagrams can be found on Github [11]. Here is a sample, which discusses content-based recommendation systems:



V. Result

VI. Future Work

Subsection text here.

References

[1] B. of Labor and Statistics, "Quarterly census of employment and wages," 2013.

[2] Y. Li, J. Hu, C. Zhai, and Y. Chen, "Improving one-class collaborative filtering by incorporating rich user information," in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 959–968.

[3] A. Rajaraman and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2011.

[4] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 502–511.

[5] S. Lai, Y. Liu, H. Gu, L. Xu, K. Liu, S. Xiang, J. Zhao, R. Diao, L. Xiang, H. Li *et al.*, "Hybrid recommendation models for binary user preference prediction problem." in *KDD Cup*, 2012, pp. 137–151.

[6] J. Jean, R. Lowrance, and D. Shasha, "Store-mall recommendation," http://ml2014.herokuapp.com/, accessed: 2015-02-01.

[7] J. Mu, F. Wu, and A. Zhang, "Housing value forecasting based on machine learning methods," in *Abstract and Applied Analysis*, vol. 2014, 2014.

[8] E. Ghysels, A. Plazzi, W. N. Torous, and R. I. Valkanov, "Forecasting real estate prices," *Handbook of economic forecasting*, vol. 2, 2012.

[9] U. Concensus, "Demographic informaion," 2013.

[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[11] Y. Zhang, J. Wu, and P. Padmanabhan, "Mall recommendation presentation," github.com/lily-zhangying/find_best_mall/blob/master/diagrams.pdf, accessed: 2015-04-02.