BJUT at TREC 2015 Contextual Suggestion Track

Weitong Chen[1,2,3], **Hanchen Li**[1,2,3], **Zhen Yang**[1,2,3,*]
1. College of Computer Science, Beijing University of Technology, Beijing 100124, China
2. Beijing Key Laboratory of Trusted Computing, Beijing 100124, China
3. National Engineering Laboratory for CTISCP, Beijing 100124, China
*yangzhen@bjut.edu.cn

Abstract

In this paper we described our efforts for TREC contextual suggestion task. Our goal of this year is to evaluate the effectiveness of: (1) predict user preferences of each scenic spot based on non-negtive matrix factorization, (2) automatic summarization method that leverages the information from multiple resources to generate the description for each candidate scenic spots; and (3) hybrid recommendation method that combing a variety of factors to construct a system of hybrid recommendation system. Finally, we conduct extensive experiments to evaluate the proposed framework on TREC 2015 Contextual Suggestion data set, and, as would be expected, the results demonstrate its generality and superior performance.

Introduction

In this year Contextual Suggestion (CS) Track, we main aims are two folds: (1) combing a variety of factors which are crawled from open-web to construct a system of hybrid recommendation system(Albadvi and Shahbazi 2009)(Sobecki *et al.* 2006). (2) Explore a new description generation method which combines multiple aspects of information. Information recommendation is always a dilemma(Tang *et al.* 2013)(Yokoya *et al.* 2012). It's a contradiction by generality and individuality. Recommend items need to make a compromise between popularity and user's personalized interest. First, the higher popularity of items tend that each user will like it, but it can't reflect users personalized interest. At the same time, recommending according to user's personalized needs the data describes the user's interest accurately. The data about spots crawled from open-web has sparseness problem, and it is difficult to truly reflect the personal interest of each user and reflects more of the spots' popularity.

In this sense, we crawled a variety of indirect information of scenic spots from the open-web such as: attractions, spots rank, reviews of spots, etc. using this information to reflect the quality of spots. Through analysis user profiles, we can get the interest preference of each user to each Category, and use spots in Example as the training dataset to train the SVM classifier for each user interest(Xu and Araki 2006). Then, we use classifier to get the judgments about like or dislike for each user-spots pairs. Finally, we use the information crawled from website as the reflecting of spots' popularity, while use the user's interest which is analyzed from profiles as the reflecting of user's personalized interest. In the recommendation algorithm module, we combine the spots popularity and user personalized interest to generate two recommendation algorithms, eventually get BJUTa and BJUTb as two submitted results.

Our Method

Hybrid Recommendation Based on Open-web Information

Figure 1 shows our system framework. It mainly consists of three parts: (1) Useful information gathering, (2) Examples labeling, (3) Profile Modeling and Interest classification, (4) Recommendation algorithm, (5) Description generation, (6) Results generation and checking. Figure 2 shows the legend of Figure 1.

- Useful information gathering component mainly crawls everything that we need to rank the candidate scenic spots.
- Examples labeling component determine the scenic spots' category in Examples through searching the internet and a small part of the manual scenic spots.
- Profile Modeling and Interest classification component mainly consists of two parts: (1) Modeling user profiles; (2) user-spots Interest classification. Statistical method is used to identify each user preferences for each category of spots. User-spots Interest classification use the spots in Examples as the training sample to train to the SVM classifier, and use it to classify the spots into two class, user like and user dislike.
- Recommendation algorithm component mainly consists of two parts: (1) for each user - context pair choose 50 candidate recommendation spots. (2) Sort the 50 candidate recommendation spots for each user - context pair.
- Description generation component mainly utilizes multiresource information to generate spot's brief automatically. We also describe this part in details later this paper.
- Results generation and checking component get the recommend spots and spots briefly together, and use the official script to check the results and submit results to TREC.

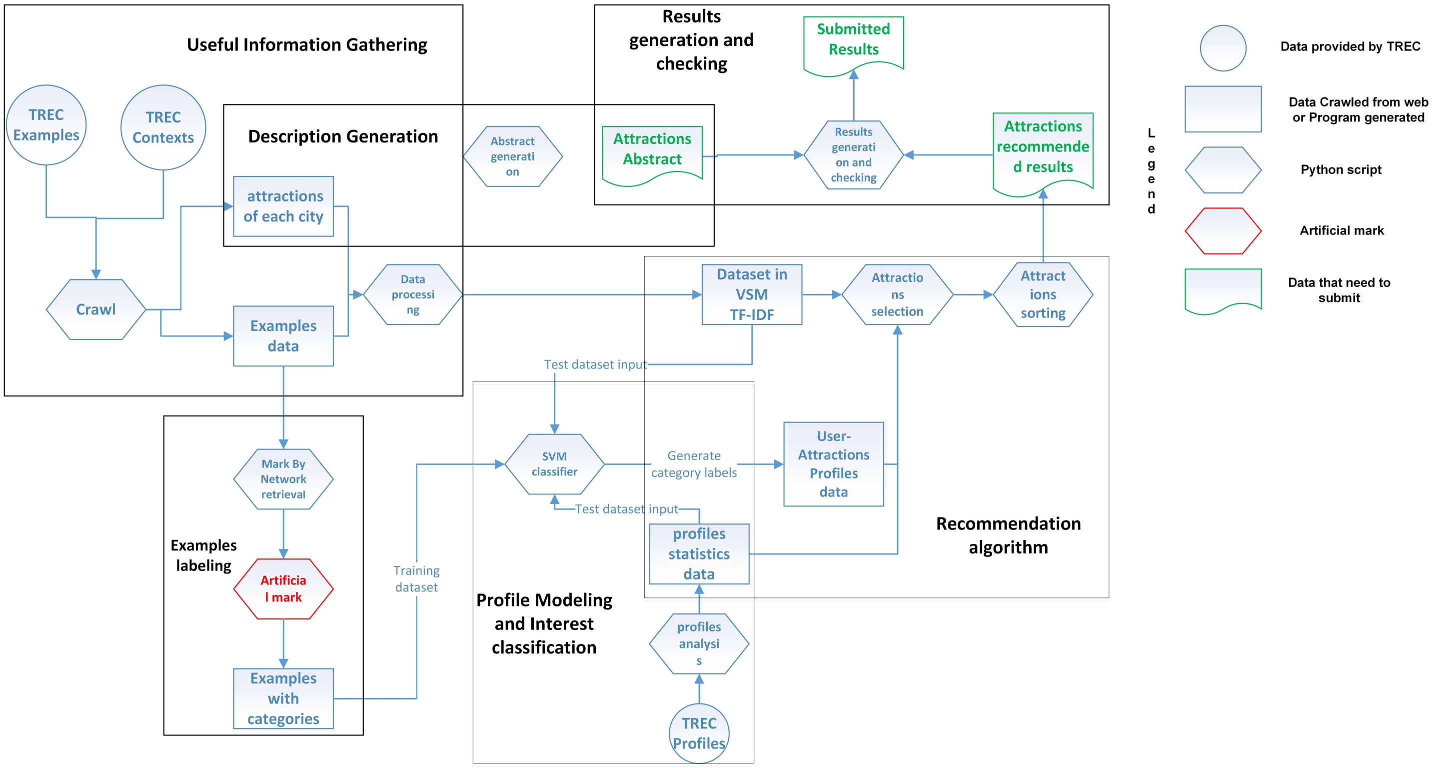

Figure 1: **The framework of BJUTa.**

Recommendation Based on D-CNMF

Figure 2 shows our system framework. It mainly consists of three parts:

- Get data matrix, statistics trec data and network data, building the matrix.
- Recommendation algorithm D-CNMF.
- Get the result, sorting by spot's score for user from new matrix.

At last, combining the two algorithm and we will get the final submitted results, BJUTa and BJUTb.

Conclusion and Discussion

In TREC 2015 Contextual Suggestion Track, we submitted two runs. Both of them use the description information of candidate spots and user interest information to select and sort the candidate spots. Description information of candidate spots include: spot's category and web information. User interest information includes: probability of user interest in each category and user favorite label of each spots. We use these indicators to make recommendation algorithm. Build matrix and using D-CNMF to get the result of BJUTa, while the spots category, web information, and probability of user interest in each category and user favorite label of each spots are used to get the result of BJUTb. Due to the open-web data sparseness problem, our recommendation algorithm does not depend on the similarity between two spots, but using a variety of indirect description of scenic spot from the open - the web which reflect the quality of spots and user profile which reflect the user interest to select and sort the candidate spots. We use a variety of information on the open-web with whole sentence extraction method to generate spots brief automatically.

The performances of our submitted runs BJUTb are in general better than the median performance. Some of the results are even best results, indicating the effectiveness of our proposed method.

References

Amir Albadvi and Mohammad Shahbazi. A hybrid recommendation technique based on product category attributes. *Expert Systems with Applications*, 36(9):11480–11488, 2009.

Janusz Sobecki, Emilia Babiak, and M Słanina. Application of hybrid recommendation in web-based cooking assistant. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 797–804. Springer, 2006.

Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. Exploiting homophily effect for trust prediction. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 53–62. ACM, 2013.

Jin An Xu and Kenji Araki. A svm-based personal recommendation system for tv programs. In *Multi-Media Modelling Conference Proceedings, 2006 12th International*, pages 4–pp. IEEE, 2006.

Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled

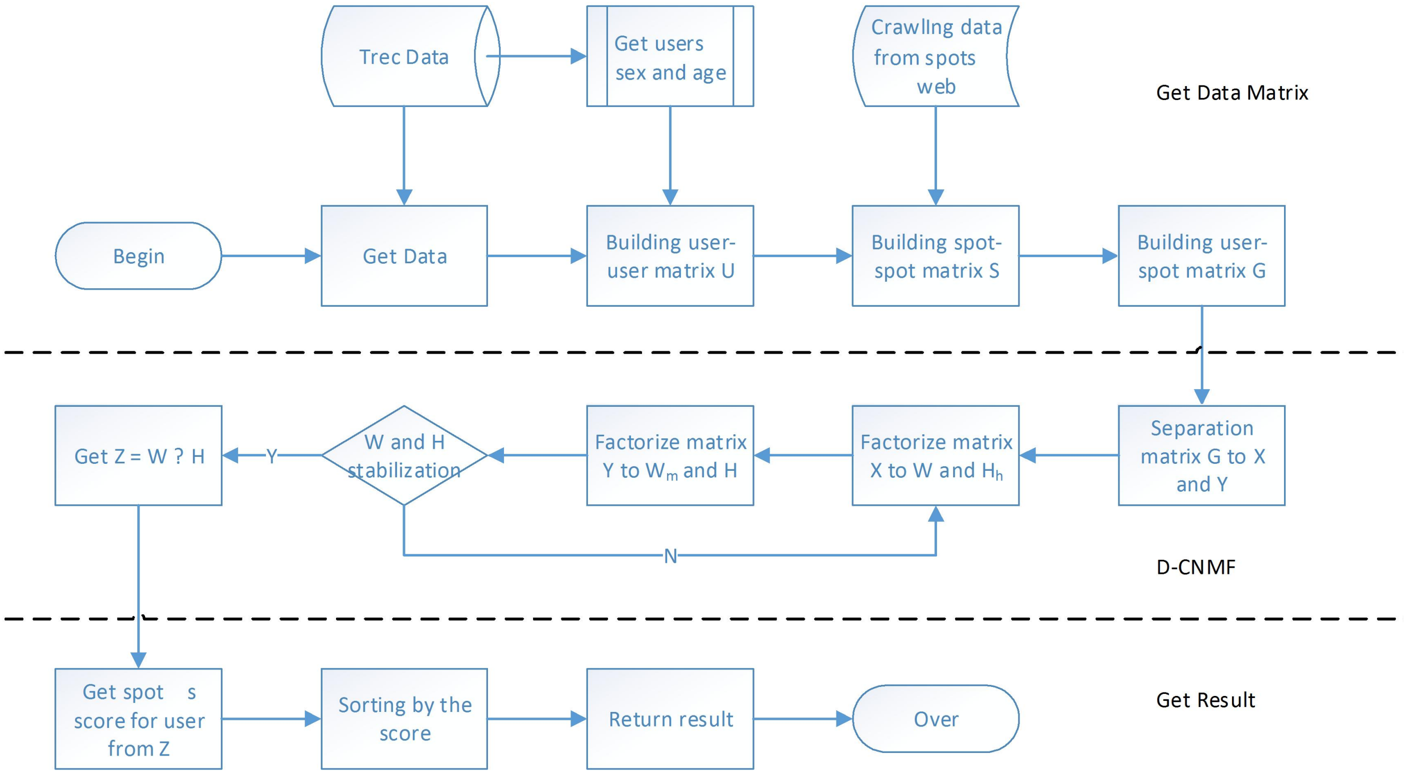

Figure 2: **The framework of BJUTb.**

nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(2):528–537, 2012.