LDA2Vec 항목 모델을 기반으로 한 협업 필터링 권장 알고리즘

장 흠[O], Scott Uk-Jin Lee[*]

[O]한양대학교 컴퓨터공학과,

[*]한양대학교 컴퓨터공학과

e-mail: {zhangxin123[O], scottlee[*]}@hanyang.ac.kr

Collaborative Filtering Recommendation Algorithm Based on LDA2Vec Topic Model

Zhang Xin[O], Scott Uk-Jin Lee[*]

[O]Dept. of Computer Science and Engineering, Hanyang University,

[*]Dept. of Computer Science and Engineering, Hanyang University

● 요 약 ●

In this paper, we propose a collaborative filtering recommendation algorithm based on the LDA2Vec topic model. By extracting and analyzing the article's content, calculate their semantic similarity then combine the traditional collaborative filtering algorithm to recommend. This approach may promote the system's recommend accuracy.

키워드: 협업 필터링(collaborative filtering), 추천 알고리즘(recommendation algorithm), LDA2Vec

I. Introduction

With the rise of technology and mobile networks, there are more and more articles we can access, how to obtain a useful article from a large number of articles has become a major demand for modern data processing. Traditional collaborative filtering(CF) recommendation algo- rithm[1] calculates the users who have the same interests and recommends them items that similar users have liked. However, this method only takes into account the user's scoring data, which is subject to the sparsity problem and does not take into account the information hidden in the text.

To address these problems, we propose a hybrid recommend method that combines CF algorithm with the LDA2Vec[2] topic model. LDA2Vec can effectively extract an article's topic. Then compute these topics' semantic similarity. Fuse the topic semantic similarity with user similarity can get better accuracy.

II. Related Works

There are various topic models used in the recommendation system. The Latent Dirichlet Allocation(LDA) topic model[3] is the most popular one. It can infer the distribution of topics

from an article in the form of a probability distribution. After analyzing their topic distribution, these articles can be clustered or classified. But the LDA model doesn't establish local word relationships like word2vec. LDA2Vec[2] is a new topic model that combines the advantages of word2vec and LDA. Therefore, applying the LDA2Vec model to the recommendation system can improve the system's accuracy.
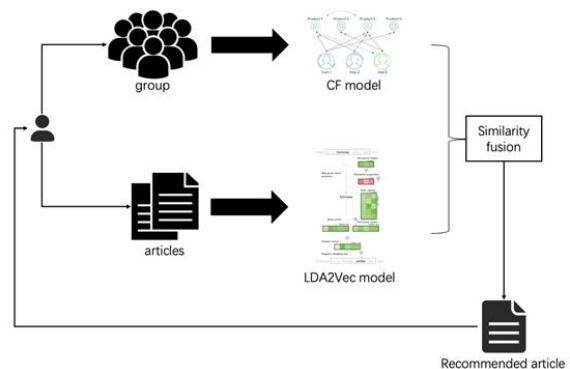

Fig. 1. System Architecture

III. The Proposed Scheme

The architecture of the overall recommendation system is shown in Fig. 1.

1. CF model to compute user similarity

In this architecture, we use user-based CF. This method first uses statistical techniques to find neighbors who have the same preferences as the target user. Then generate recommenda- tions to the target user based on the prefer- ences of the target user's neighbors. For example, given user u and user v, make N(u) is the set of items for user u has positive feedback, and N(v) is the set of items for user v has positive feedback. Then calculate similarity between u and v by using Jaccard formula:

$$sim_{uv} = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

2. LDA2Vec model to compute semantic similarity

Word and document topics can be jointly trained in the LDA2Vec model. In the word level, the word vectors are obtained through word2vec training. In the document level, the LDA[3] model is used to map the latent semantic space to form the document topic vector. Then the LDA2Vec can integrate the word vector and the document topic vector into the same space to train. Finally, the article's semantic vector can be obtained. In the semantic vector space, the semantic similarity between articles can be easy to calculate.

3. Similarity fusion

Fuse the user similarity and semantic similarity with the control factor α. We assume that its value is from 0 to 1, and then the improve similarity is expressed by:

$$sim_{total} = sim_{user} + (1 - \alpha)sim_{article}$$

Finally according to the fusion similarity. Select the article with the highest similarity to recommend to the target user.

IV. Conclusions

The proposed collaborative filtering algorithm based on the LDA2Vec model in this paper can improve the accuracy of the recommendation algorithm. Compared with the traditional collaborative filtering algorithm by adding article content analysis as side information. However, there are still many challenges in the recommen -dation system, such as the cold start problem and how to incorporate the time series into the recommendation system. Therefore, we will continue our work to develop more algorithms to solve these challenges.

REFERENCES

[1] Schafer J.B., Frankowski D., Herlocker J., Sen S. "Collaborative filtering recommender systems." The adaptive web. Springer, Berlin, Heidelberg, 2007. 291-324.

[2] Moody, Christopher E.. "Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec.".

[3] Wang, Chong and David M. Blei. "Collaborative topic modeling for recommending scientific articles." KDD (2011).