IJCSN International Journal of Computer Science and Network, Volume 4, Issue 3, June 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

Accurate Recommendations Using Linked Taxonomies of Subjective Assessments

[1] Vaishnavi Pakhode, [2] Advait Pakhode, [3] Nagesh Jadhav

[1] MIT College of Engineering,
Pune, Maharashtra 411038, India

[2] MIT College of Engineering,
Pune, Maharashtra 411038, India

[3] MIT College of Engineering,
Pune, Maharashtra 411038, India

Abstract - Subjective assessments like 'beautiful' and 'breathtaking' are assigned to items by users and are commonly found in reviews on many online sites. Analyzing the links between these SAs and items can help improve the recommendation accuracy. We propose a different method which links a taxonomy of items to a taxonomy of SAs to capture user's interests in detail.

Keywords - *Recommendation System, Collaborative Filtering, Subjective Assessments.*

1. Introduction

Different online retail sites like Amazon employ user generated reviews or tags of items as well as ratings. These reviews or tags consist of Subjective Assessments. These SAs provide information about the user's preferences. The ratings as well as these SAs can be used for recommendations in recommendation systems.
Subjective assessments show what the user perceives about an item. This data is very useful in understanding the user's needs. The recommendations provided by systems employing these methods can match the desires of the users.

Most commercial recommendation systems use methods based on collaborative filtering (CF). CF makes use of intuition in such a way that the user who gets the recommendation tends to have similar interests as an active user. The reviews which the user assigns to an item are also called as tags. These user generated tags are classified into the following categories. Context tags contain contextual information about the item. Subjective tags are descriptive judgments about the items. Organizational tags are about self-references and personal tasks. The content information of these items is useful in improving the recommendation accuracy.

Many studies were conducted previously to measure the use of these tags in recommendation accuracy.The results of these studies showed that the recommendation accuracy did not improve by using subjective tags. We assume that the poor accuracy of recommendations was due to sparsity, i.e. datasets used to measure the similarity of users were not adequate. This problem is called as the sparsity problem. The sparsity problem has two causes when using SAs. Different users do not assign SAs using the same items and the ratings dataset contains few items with the same SAs.

The sparsity problem due to the reasons mentioned above can be resolved by using the concept of taxonomies though not simultaneously. A previous method on taxonomies could not resolve the two sparsity problems simultaneously. This was because the taxonomies of items and SAs were based on different types of item classifications.

To solve the sparsity problem while retaining the semantics of user's item transactions, we propose a different method using the concept of linked taxonomies which links separate taxonomies of items and SAs. Our method groups the SAs assigned by users to items in SCs. Our method is based on the observation that users who assign high rating to an item and also provide an SA may also like the items from the same class and similar SAs.
For example, if user u gives a high rating to pasta restaurant Figlio and assigns it with the SA 'light' ,this user is expected to prefer pasta restaurants assigned with the SA 'non-fatty' by other users. This is because 'light' and 'non-fatty' share the same SC. Hence, our method treats user v who likes restaurant Laria as similar to u. And it treats user w who likes pasta restaurant Fratelli as dissimilar to u.

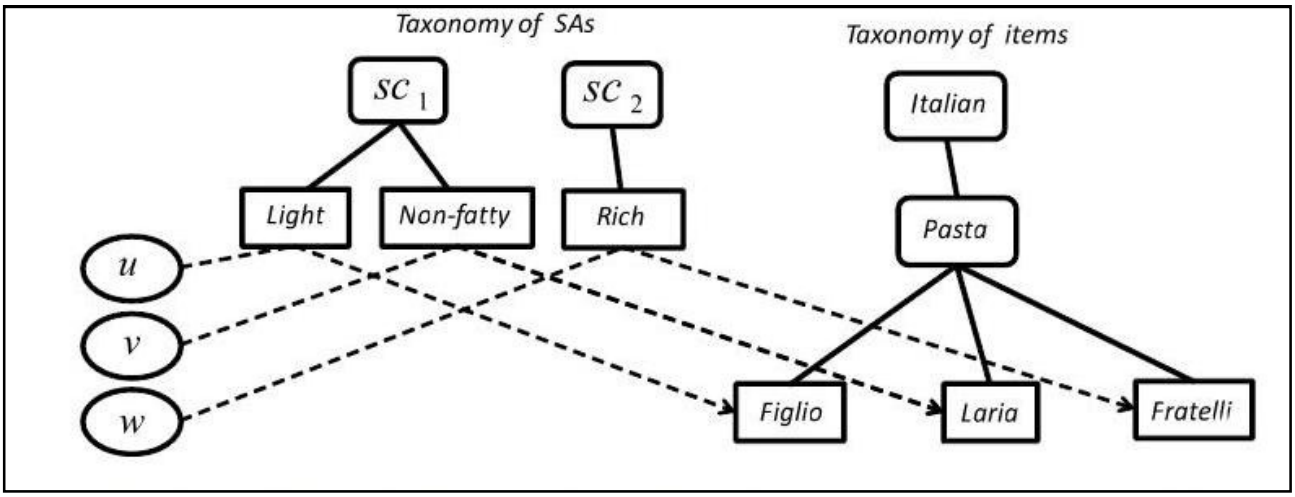

Fig 1: Linked taxonomies for modeling user's interests

This method can provide information about why items are recommended to users based on the relationship identified in the taxonomies of SAs and items as well as information related to users who assign SAs in their reviews.This paper is organized as follows. We describe previous work in the next section. We describe linked taxonomies in detail in section 3. We present our present our implementation details in section 4.Finally we conclude in section 5.

2. Previous Work

Commercial recommendation systems use methods based on collaborative filtering (CF). CF based systems work in the following manner. A user expresses his or her preferences by rating the items on the system. These ratings can be viewed as an approximate representation of the user's interest in the corresponding domain. The system matches this user's ratings against other user's ratings and finds the people with most "similar" tastes. The system recommends items that the users with similar tastes have rated highly but are not yet rated by this user.
As discussed earlier CF based methods suffer from sparsity problem.

These methods treat different words having similar meanings as different SAs. For example, the words "excellent" and "marvelous" were treated as two different SAs even though they might mean the same thing actually.
Our method which creates subjective classes for the SAs does not suffer from this problem because SAs meaning similar things are a part of the same class. Hence our method provides accurate recommendations as compared to previous methods.

3. Computing Recommendations by Linking Taxonomies of SA's and Items

3.1 Extracting SA's

For the extraction of SAs given by users to different items, we use natural language processing. We determine the phrases and compare them with patterns defined for classifying SA's. The different patterns which are used are |<adverb><adjective>|, |<adjective><noun>|,|<adjective>|,|<adverb>|,|<pronoun><adjective>|.

We can use sentiment analysis methods to extract the SAs from item reviews. In the reviews, the subject of the review is an item that has been reviewed, i.e. if we consider movie reviews, the subject is the movie. We use syntactic analysis to extract the SAs and to analyze the relationships between the SA candidates and reviewed items according to the following procedures.

(1) We model the properties of the items that are assigned SAs by the users.
(2) We analyze the relationships between these properties and the SA candidates using a syntactic analysis tool.

3.2 Creating taxonomy of SAs

A new taxonomy can be created by using Wordnet. We will create subjective classes of SAs which hold the related SAs or SAs having similar meanings.For assigning words having ambiguous meanings to SCs we use synonym nets provided on WordNet. WordNet provides

polysemous count of any word i.e. the number of synonym nets a word has. If the target SA belongs to more than two synsets we extract the most frequent synset and assign the corresponding sense of the chosen synset in the SC for the target SA.

For SAs that correspond to the pattern |<adverb><adjective>| we classify them according to the adjective in the pattern.

3.3 Recommendation of Items

Using subjective classes of SAs and taxonomies of items, users will be given accurate recommendations. The dataset which we use contains user information and ratings from 0-5 and SAs. Our approach is based on the observation that users who assign high ratings to an item and also provide SA may also like items from the same class and similar SAs. For example, suppose user 'a' gives high rating to movie 'Mission Impossible 4'and assigns it with SA 'incredible' the user is expected to prefer action movies assigned similar SAs because the SAs share the same subjective classes. Hence , our method treats user 'b' who likes the action movie 'Salt' and gives it the SA 'thrilling' as similar to' a'.

4. Implementation Details

4.1 Block Diagram

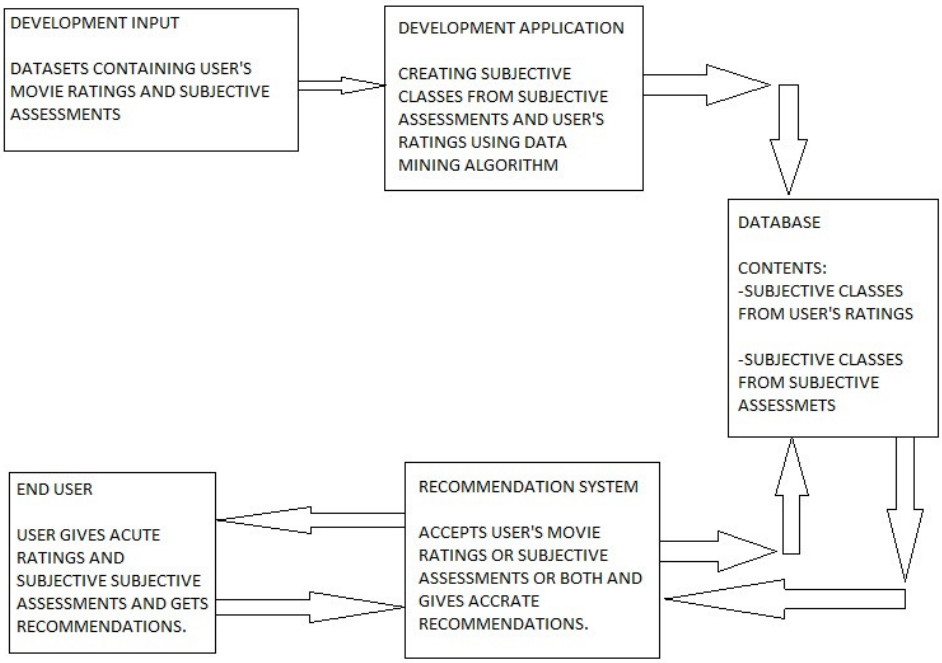

Fig 2 Block Diagram

4.2 Implementation

We used the movie dataset provided by MovieLens, a well-known online movie recommendation system, to evaluate the performance of our method. This dataset contains a huge volume of tags and ratings assigned by users to items. The maximum user rating is 5 and minimum user rating is 0.5. we used this data because this is the only publicly available data that contains both ratings as well as reviews.

First we extracted the SAs using the procedures described in section 3 which gave us around 19,000 SAs from about 100,000 tags. It should be emphasized that our method is not applicable where only rating information is available without any information on SAs.

Next we created a database which contained all the user info as well as the datasets. We also created a database for synonym nets of words using the WordNet library, which also publicly available online.
Next we created a code using Java to create a taxonomy of items and SAs and access this information to give recommendations.

Table 1: Movie Dataset

Name	Value
Range of ratings	[0.5-5]
Number of ratings	128,511
Number of ratings with SAs	19,481
Number of users	943
Number of SAs	1834
Number of item classes	20
Number of items	7301
Average number of ratings assigned to each item	17.3
Hierarchy of item taxonomy	2
Hierarchy of subjective taxonomy	3

Table 2: Examples of SCs created for and used in movie dataset

SC1	SC2	SC3	SC4	SC5
Unrealistic plot	Vivacious	Red	Exciting	Historic
Surreal violence	Racy	Gory	Thrilling	Pre historic
Unrealistic	vibrant	Red rum	Breathtaking	Ancient
Surreal drama		Bloody	Titillating	Past life
Unrealistic crap				

4.3 Accuracy

We tested our system with a number of users and found that 80% of the movies recommended were according to the user's requirements. Thus, we assume that our method is more accurate than thhe previous methods which did not use SAs.

5. Conclusion

In this study, we have implemented a novel method for the measurement of similarity of users. Our method groups the SAs assigned by the users to items in SC and the SAs/SCs reflect the classes in which they are included. Our method computes the similarities of users based on the SAs/SCs assigned to items and those assigned to item classes.

6. Future Work

Linked taxonomies could be applied to tags other than SAs like content tags or context tags because different types of semantic relationships exist for different types of tags.

References

[1] Makoto Nakatsuji and Yasuhiro Fujiwara," Linked taxonomies to capture users' subjective assessments of items to facilitate accurate collaborative filtering", ScienceDirect, Artificial Intelligence 207 (2014) 52–68.

[2] I.Cantador, I.Konstas, J.M.Jose, Categorizing social tags to improve folksonomy-based recommendations, J.WebSemant.9 (2011)1–15.

[3] Y.Zhen, W.-J.Li, D.-Y.Yeung, TagiCoFi,"tag informed collaborative filtering", Proc. Rec Sys' 09 , pp.69–76.

[4] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl "Evaluating collaborative filtering recommender systems", in ACM Trans. Inf. Syst., 22 (2004), pp. 5–53.

[5] V. Schickel-Zuber, B. Faltings, "A semantic similarity function based on hierarchical ontologies", in Proc. IJCAI'07, pp. 551–556.

[6] Y. Koren, "Collaborative filtering with temporal dynamics", in Proc. KDD'09, pp. 447–456.

[7] A. Meena, T.V. Prabhakar, "Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis", in Proc. ECIR'07, pp. 573–580.

[8] H. Kanayama, T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis", in Proc. EMNLP'06, pp. 355–363.

[9] R. Valitutti, "WordNet-Affect: an affective extension of Wordnet" ,in Proc. International Conference on Language Resources and Evaluation, pp. 1083–1086.

[10] B. Sarwar, G. Karypis, J. Konstan, J. Reidl," Item-based collaborative filtering recommendation algorithms", in Proc. WWW'01, pp. 285–295.