Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017)
IEEE Xplore Compliant - Part Number:CFP17M19-ART, ISBN:978-1-5386-1959-9

# Credible User-Review Incorporated Collaborative Filtering for Video Recommendation System

Anbazhagan Mahadevan
Department of Computer Science and Engineering
Saranathan College of Engineering
Tiruchirappalli, India
anbazhagan-cse@saranathan.ac.in

Michael Arock
Department of Computer Applications
National Institute of Technology - Trichy
Tiruchirappalli, India
michael@nitt.edu

***Abstract*—A system that recommends an item to a user that he/she is likely to be interested in is said to be a recommender system. Collaborative Filtering(CF) is a technique used to implement recommender systems. CF uses numeric ratings given by users to find the nearest neighbors to the target user and generates recommendations. An upgraded collaborative filtering algorithm that uses credible user-reviews to generate accurate recommendations is proposed in this paper. While in the earlier CF approaches, mere numerical ratings are used for making recommendations, but overall ratings alone cannot properly reflect user's opinion about an item. Another deficiency associated with the existing rating based CF approaches is sparseness in rating database. Data sparsity problem can be got rid of, only if we have an alternate way of filling up the empty ratings. The proposed approach tries to do this by inferring numeric ratings from text reviews. Rating inference involves a sentiment analysis problem of finding the sentiment orientation and strength of opinion words expressed in user-reviews. Some existing recommender systems have already incorporated user-reviews for making better recommendations, but they did not take into account the credibility of those reviews. The proposed CF approach grades the credibility of user-reviews by considering the factors such as reputation of the reviewer and quality of the contents in review. Experimentation of the proposed framework is done and results are validated**

***Keywords—Recommender Systems; Collaborative Filtering; Sentiment Analysis; Sentiment Orientation; Opinion Dictionary; User-Item Matrix; Mean Absolute Error;***

## I. INTRODUCTION

Recommender Systems are systematizing the process of giving word-of-mouth recommendations which allows people to share their opinions about items(e.g., Videos, Gadgets, Books, and many more) among their friends to help them go for the relevant items they are likely to find interesting. Collaborative Filtering (CF) is an important and noticeable technique in recommender systems. It uses a database of user preferences(numerical ratings) to provide personalized recommendations to users.

CF based Recommender Systems are categorized into two types based on how user opinions about items are collected for input: Implicit feedback and Explicit feedback. In implicit feedback CF, implicit votes which are captured via users' interaction with the system are the main source of user preferences, e.g. purchase history, logs. Explicit feedback CF uses explicit information provided by users on their preferences; generally numeric ratings are the main form of explicit feedback [1].

Recommender Systems using typical Collaborative Filtering technique are relying on numeric ratings given by users as their only basis of user opinion about items. But, mere numerical ratings cannot fully reflect a user's actual preferences and relying on numerical ratings alone is difficult. For instance, two users in a recommender system, $U_1$ and $U_2$, both buy the same model smartphone and give it a high rating. However, the reasons for their favorable ratings may be different. $U_1$ may like this phone because its camera resolution is as excellent as he expected, while $U_2$ expresses interest for the phone due to its battery life. Thus, CF systems may not be able to find the nearest neighbor to the target user based merely on their numeric ratings given on items, resulting in possibly poor recommendations. Likewise, items with a lot of missing ratings may result in the data sparsity problem in rating information database. Consequently, the performance of CF may get degraded and recommendations made by such system will be less accurate.

Apart from the traditional techniques to generate recommendations for users; nowadays modern recommender systems are striving to ponder additional information provided by the target user or other users about the item. Ever since Web 2.0 had come into existence, users have been expressing their opinions about items on the Internet using free-text reviews. Such consumer reviews are the powerful source of information to a recommender system to provide nuanced, more thorough, and trustworthy user opinion information. Researches on mining user preferences from reviews, problem known as sentiment analysis or opinion mining, are becoming progressively widespread in the text mining. But, little research attention has been paid to the integration Sentiment Analysis and CF. The process of giving word-of-mouth recommendation can be easily augmented by the combination of use-reviews and numerical ratings.

There have been many approaches introduced to leverage user opinion information inferred from user-reviews for the purpose of making good recommendations. Nevertheless, those approaches are model-based, failed to directly integrate user reviews into collaborative filtering, and not paid attention to the credibility of user-reviews. The proposed user-review incorporated CF approach, firstly weighs the credibility of user reviews, extracts user preferences/opinions communicated in the credible user reviews and directly integrates those inferred preferences into CF, striving to find appropriate neighbors and generate better recommendations.

The proposed framework aims to integrate opinion mining/sentiment analysis and CF. An inference approach to infer rating from reviews is incorporated into the proposed framework in order to extract user preferences expressed in the user-reviews and to input them to CF algorithms. There are three-fold contributions of this approach. First, the credibility of user-reviews is gauged before considering it for recommendation process so that the effect of shilling attack gets reduced to a descent extent. Secondly, user-reviews are used as a supplementary basis of user preferences to resolve the familiar data sparsity problem in CF. Thirdly, it allows existing CF algorithms to be used in domains where preferences cannot be conveyed as numeric ratings.

The remaining parts of this paper are organized as follows. Section II describes related work on user review incorporated CF. Section III outlines the proposed framework for gauging the credibility of text reviews and inferring user preferences from credible reviews for generating good recommendations. Section IV covers a discussion on experimental setup and results, and finally, Section V concludes this article by giving an outline of the on-going and future work.

## II. RELATED WORK

The work proposed here uses opinion mining or sentiment analysis at the phrase/expression level. The proposed framework aims at extracting user opinion expressed in text reviews and makes use of them in existing CF algorithms. This was the most prominent work in many previous researches [2, 3, 4, 5]. The text reviews are analyzed to consider the aspects discussed by users and their associated opinions. This offers two different visualizations: one for businesses and one for end-users [2]. User reviews are used to produce descriptions of items. An agreement of users about items' features is denoted by those descriptions. The accuracy of recommendations is improved by using agreement of users' opinions about different aspects of the content [3]. The topic probability distribution of users' review is inferred by using Latent Dirichlet Allocation (LDA) and then two aggregation methods are used to discover users' topic of interest and item features' topic profile. Users' interesting topic and items' topic profile are used to find nearest neighbors to target user and target item in order to generate perfect recommendations [4]. User preferences expressed in text reviews known as opinion words are extracted from those reviews by using sentiment analysis. Then the extracted opinion words are mapped onto corresponding numeric ratings that can be used as inputs to the existing CF algorithms for the purpose of making accurate recommendations [5]. User preferences on individual item features are extracted from text reviews and a new weighting scheme complying with TF-IDF is used to find the priority of item features to influence users' overall opinion about different items which allows us to generate accurate recommendation [9]. Gauging information reliability and enhancing consumer trust tends to help realizing the benefits of sharing information with friends online. Present days rating systems not only allow consumers to read and write review on products, but also to check the credibility of the review and the reviewer [6]. There are two approaches available to infer numerical rating from text reviews based on the related works discussed in this section. The first approach tries to associate a polarity to opinions expressed in reviews as a classification problem. The second approach assigns scores to opinions words expressed in reviews. Though all these related works have used user-reviews for making recommendations, they failed to consider credibility of reviews. The proposed works does both the incorporation of text reviews and grading their credibility before using them.

## III. THE PROPOSED APPROACH

The framework proposed here has three components contained in it. The first component gauges the credibility of reviews and reviewers based on certain factors. The second component is for analyzing user-reviews and inferring numeric ratings from those reviews, and the last one is for doing collaborative filtering which makes predictions for item recommendations based on numeric ratings given by the user and the ratings inferred from user-reviews. An outline of the proposed framework is depicted in Fig. 1 below

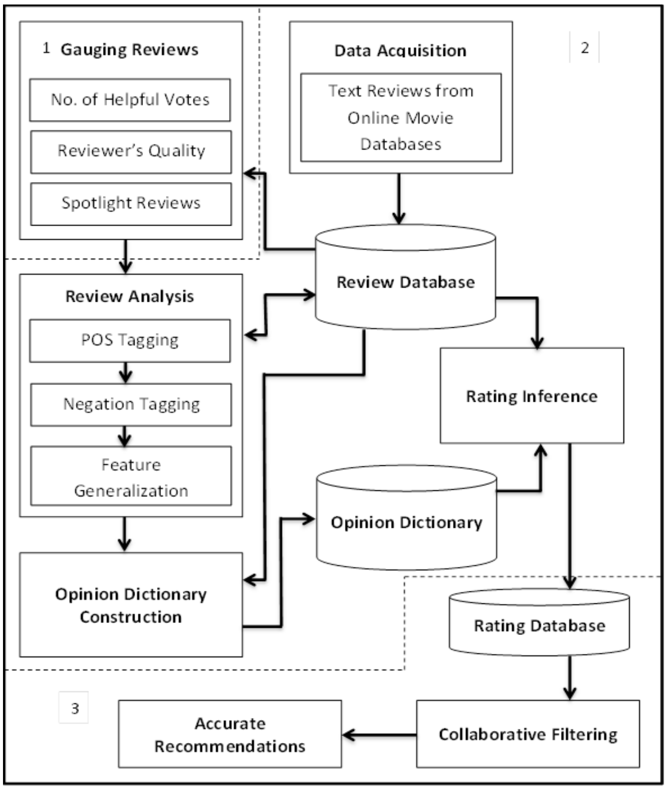


Figure 1. Outline of the proposed framework

The rating inference process involves five important steps, they are data acquisition, gauging the credibility of reviews, user-review analysis, constructing a dictionary of opinion words and inferring numeric rating out of reviews. Inferred numeric ratings from credible user-reviews can be used as input to existing CF algorithms in order to make better recommendations.

### *A. Data Acquistion*

Data acquisition comprises gathering and pre-processing user-reviews for the later analysis. Data pre-processing turns raw data into an understandable format. Depending on the sources of data gathered different pre-processing steps may be required to be applied. User reviews are likely to be semi-structured documents, having an unstructured text body and certain structured headers. For example, a video review on Amazon Instant Video has structured headers with a reviewer's identity and date of review, and unstructured blocks of user's comments written in natural language, on the movie being reviewed. Next to text reviews, a provision is available to the readers to tell whether the review is helpful or not.

### *B. Gauging Reviews*

Credibility of a review involves two things namely reliability of the review and consumer (reviewer) trust [6]. Credibility of a review is based on the reputation of the reviewer or quality of the review. The quality of the review can be measured by how much helpful readers found the review. Second, we consider the reputation of the reviewer. Top reviewers are being identified by Amazon, and these standard reviewers can create a great impact in product sales. Third, we look into the influence of spotlight reviews on product sales. These spotlight reviews are separated from the other reviews and are displayed on the product page before any other reviews, so they can more easily get into customers' sight than other reviews.

### *C. Review Analysis*

User-reviews are thoroughly analyzed to uncover and extract useful information given by the users. Some important steps in review analysis are:

*1) Part-of-speech(POS) tagging:* Information regarding product features are usually expressed as nouns or noun phrases in reviews, whereas opinions are generally expressed as adjectives or verbs. Therefore, POS tagging is necessary to extract those information from reviews. Stanford POS tagger is used in the proposed work to carry out POS tagging.

*2) Negation tagging:* Negation tagging helps to identify those words which would have negation effects on other words and reflecting the effects of those words while determining the Sentiment Orientation of reviews.

*3) Feature generalization:* Feature generalization aims to generalize item features that are too specific. For example, a sentence "Jurassic Park - III is adventure and science fiction.", in which the name of the movie being reviewed is "Jurassic Park - III". The sentence is generalized to " MOVIE is adventure and science fiction".

### *D. Constructing Opinion Dictionary*

Opinion dictionary comprises opinion words, their determined sentiment orientations(SO) and the strengths of their SO [5]. Opinion dictionary is used to compute the overall SO of a text review. The SO and strengths of opinion words are determined based on how likely a given opinion word is to be a positive or a negative sentiment. Sometimes SO of a word need not be similar to SO of its synonym.

The proposed framework uses a relativefrequency-based method to construct an opinion dictionary. In relativefrequency-based method, the strength of a word with respect to a particular sentiment class is calculated as the relative frequency of its existence in that class [5]:

$$OS(w_i, c) = \frac{F(w_i,c)}{\sum_{c_j \in C} F(w_i,c_j)} \quad (1)$$

Where $w_i$ is an opinion word expressed in the review. $OS(w_i,c)$ is the strength of $w_i$ with respect to a certain sentiment class *c*. *C* is a set of sentiment classes used to compute the relative frequencies of $w_i$. *c* and $c_j$ are elements of sentiment class *C*. $F(w_i, c)$ indicates relative frequency of $w_i$ with respect to class *c*.

### *E. Numeric Rating Inference*

This component tries to determine the overall sentiment expressed by a user in a review. The aggregation of strengths of all opinion words with respect to a given sentiment class gives the numeric rating corresponding to the user-review.

### *F. The Proposed Algorithm*

**Algorithm:** User-Review Incorporated Collaborative Filtering
**Input:** User-Item matrix R, User-Reviews
**Const n:** Maximum number of users in *N(u)*, the neighbors of user *u*
**Output:** Recommendation for a video *v* that the active user *u* is likely to watch

```
FOR each user review t_uv DO
    Measure the credibility based on content quality and
    reviewer's reputation
    IF the review t_uv is proved to be credible
        Analyze the review and infer the corresponding
        rating
        Add this rating r_uv to newly constructed user-item
        matrix R'(U X V)
    ELSE
        Ignore the review
    END IF
END FOR
FOR each user u DO
    Set N(u) to the n users most similar to user u
    FOR each video v that user u has not rated DO
```

Calculate the Weighted Combination of ratings given to video *v* by neighbors *N(u)*
END FOR
END FOR
Recommend to user *u* the video *v* with the highest predicted rating $p_{uv}$
END

This algorithm takes as input text reviews given by users on various videos. The original user-item matrix *R* is considered only at the end for verifying the correctness of this algorithm. A new user-item matrix *R'* based on ratings inferred from reviews is constructed. Matrix *R'* is used to find the similar users to the target user *u* so as to generate better recommendations. More importantly not all user reviews are taken for analysis, only those reviews which are found to be credible are considered for rating inference. Grading the credibility of review is carefully done based on factors such as number of helpful votes received by the user-review, reputation of the reviewer, and whether the review is taking place in product's page(spotlight reviews).

## IV. Experiment

This section is for describing the dataset used, and reporting the results of experiments carried out.

### A. Data Set

We used Amazon Instant Video data for the experimental purpose[6]. This Dataset has a total of 37126 reviews given by 5131 reviewers for 1686 videos. As an initial step in analyzing reviews we have taken only those reviews which had received atleast one helpful vote. As a result only 10106 reviews were directly taken for further analysis and rating inference, for the remaining 27020, the reputation of reviewers and quality of their contents were assessed and 52% of those reviews were also considered for inclusion in rating inference.

Each user-review comprises a number of headings and a text body. The headings include reviewerID, asin(video ID), reviewerName, helpful, which is number of helpful votes received by the review, overall, which is a numeric rating given by the user ranging from 1 (poor) to 5 (excellent), summary, which is one line summary of user's opinion about the product, unixReviewTime and reviewTime, which gives date and time of the review. The text body is the textual description of user's opinion about the video.

### B. Experimental Setup

We have experimented two algorithms on the dataset mentioned just above: User based Collaborative Filtering(UCF) which is nothing but the existing CF algorithm, and User Review Incorporated Collaborative Filtering(URICF) which is the algorithm discussed in section III. The original user ratings available in the dataset and ratings inferred from the text reviews are fed into UCF and URICF algorithms, respectively. Then URICF with the basic recommendation algorithm UCF are compared. For the experimental purpose, the dataset is divided into two parts namely a training data and test data. The variable α is used to represent the ratio of training data. In input dataset, 80% of the data are chosen as training data(α = 0.8) and the remaining 20% as test data. The accuracy of recommender systems is evaluated by using Mean Absolute Error (MAE) [1]. Suppose that there N actual and predicted rating pairs ($r_{u,v}$, $p_{u,v}$), where *u* means a user and *v* refers to a video. Mean Absolute Error is expressed as follows.

$$MAE = \frac{1}{N}\left|\sum_{v=1}^{N}\left(r_{u,v} - p_{u,v}\right)\right| \qquad (2),$$

where $r_{u,v}$ is the actual numeric rating given by user *u* on video *v* and $p_{u,v}$ is the predicted rating for the target user *u* on video *v*.

### C. Evaluation of the Proposed Framework

This experimental study tires to discuss the effects of certain key parameters such as similarity measure and neighbor size on the prediction accuracy. The UCF and URICF algorithms are experimented and prediction accuracy of both the algorithms is evaluated by using MAE.

An experiment was done to measure the MAE when the neighbor size is changing from 5 to 80 with the value of the ratio of training set fixed at 0.8. It can be clearly observed that both the curves are having a similar trend and the proposed URICF performs slightly better than the existing UCF. From the results it is very clear that for the Neighbor size from 5 to 35 both the algorithms generate better recommendations then the curves start rising. Hence, the optimal Neighbor size is found to be 35.

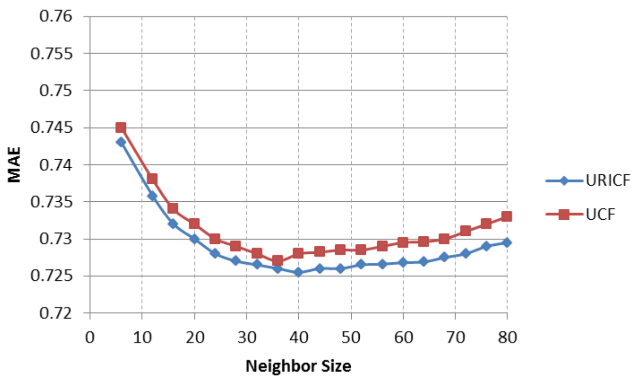


Figure 2. Neighbor Size Vs MAE

The impact of two similarity measures such as Cosine Similarity, and Pearson Correlation(PCC) is analyzed next. For the neighbor size of 20 and the training data ratio of 80%, the algorithms behave as in Fig. 3. It is very clear from the results that for both UCF and URICF algorithms, Pearson Correlation has a noticeable advantage and URICF slightly outperforms over UCF. The results of two experiments proved a fact that the numeric ratings inferred from user-reviews are better than overall numeric ratings given by users, where there is a need better recommendations.

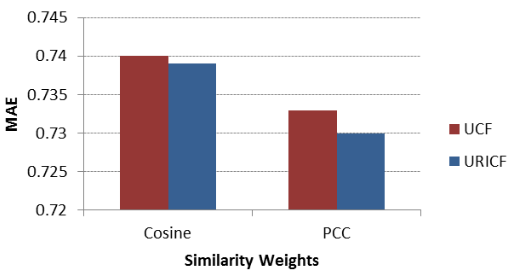


Figure 3. Similarity Weights Vs MAE

## V. Conclusion And Future Research

This article proposes an upgraded CF approach to integrate Sentiment Analysis and existing Collaborative Filtering. Such approach tries to infer numerical ratings from user opinions expressed in unstructured text reviews. This article also tries to take into account the quality of reviews and the trustworthiness of reviewers which helps the readers to check the validity of the opinions expressed in the review, before taking them for further analysis. A framework is proposed at first place in order to exhibit the behavior of URICF, i.e., data collection, gauging the credibility of reviews, review analysis, and collaborative filtering. The purpose of doing review analysis is to find the Sentiment Orientation of opinion words expressed in text reviews and strength of the SO. Finally, in order to validate the proposed framework, an experiment is done to compare the performance of existing UCF and proposed URICF algorithms. From the result it is clearly understood that user reviews are having high impact on recommendation process.

The future work is to incorporate feature engineering into our present work. Instead of inferring single numerical rating which would reflect the whole text review, it would be better, if we had individual ratings for all product features. Feature engineering will lead us to a more accurate recommendation process.

## References

[1] Gediminas Adomavicius, and Alexander Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, June 2005, pp. 734-749

[2] Dimitris Proios, Magdalini Eirinaki, Iraklis Varlamis, "TipMe: Personalized advertising and aspect-based opinion mining for users and businesses," IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining , pp. 1489-1494, 2015

[3] Rafael Martins D'Addio, Marcelo Garcia Manzato, "A Collaborative Filtering Approach based on User's Reviews," Brazilian Conference on Intelligent Systems, pp. 204-209, 2014

[4] Huiming Wang, Nianlong Luo, "Collaborative filtering enhanced by user free-text reviews topic modelling," Iinternational conference on information and communications technologies, pp. 1–5, 2014

[5] CaneWing-ki Leung, Stephen Chi-fai Chan, Fu-lai Chung, "Integrating Collaborative Filtering and Sentiment Analysis: A Rating Inference Approach," The ECAI 2006 Workshop on Recommender Systems, pp. 62-66, 2006

[6] Samita Dhanasobhon, Pei-Yu Chen, Michael Smith, Pei-yu Chen, "An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.com," International Conference on Information Systems, pp. 1 – 17, 2007

[7] Niklas Jakob, Stefan Hagen Weber, Mark-Christoph Müller, Iryna Gurevych, "Beyond the Stars: Exploiting Free-Text User Reviews to Improve the Accuracy of Movie Recommendations," international CIKM workshop on Topic-sentiment analysis for mass opinion, pp. 57-64, 2009

[8] Jingjing Cao, Wenfeng Li, "Sentimental Feature based Collaborative Filtering Recommendation", IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 463-464, 2017

[9] Zhenxue Zhang, Dongsong Zhang, Jianwei Lai, "urCF: User Review Enhanced Collaborative Filtering", Twentieth Americas Conference on Information Systems,pp. 1-14, 2014

[10] R. He, J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering", WWW, 2016