International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 9 (2015) pp. 22201-22210
© Research India Publications
http://www.ripublication.com

Automated Recommendation System With Feedback Analysis

Kavin Kumar.V, Rachamalla Rahul Reddy, Rohit Balasubramanian, Sridhar.M, Sridharan.K
Dr. D. Venkataraman (corresponding author)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham (University), India
kavinkumar6494@gmail.com,
sunnyracha14@gmail.com,
b.rohit1993@gmail.com,
kibbus@gmail.com,
sridharank94@gmail.com.
d_venkat@cb.amrita.edu

Abstract

In today's world with the increasing e-commerce and online shopping involved recommendation systems have become a major part of decision making. In domains such as automobiles there are many websites but most of them are not having enhanced recommendation systems to enable easy decision making. Thus we have taken the onus of building a dataset with multiple parameters based on a survey of the communities needs and created a recommendation system using user based and item based collaborative filtering. To take into account the vast majority of people and their opinions we have added internal and external feedback analysis. Feedback analysis is the classification of textual data (comments) and analyzing the sentiment derived from it. We have proposed it at two levels external that is gathering comments from public platforms social media and automobile websites and internal i.e. the feedback taken from users who have been recommended items. We have developed the prototype for the proposed architecture and preliminary evaluation has been done.

Keywords: Collaborative Filtering, Multiple Parameter, Feedback analysis, Sentiment

Introduction

The internet is filled with a number of websites for rating of automobiles. Most of these websites offer ratings based on the critic centric overall reviews and feedbacks that are unidirectional in nature. A new idea of recommendations along with the ratings can be used to amplify the process. This not only takes into account the ratings generated by critics but also the likes and dislikes given by the user. This in turn reduces the ambiguity prevailing in the minds of the customers.

A recommendation system helps the customers or end users in suggesting relevant products out of a large number of data by analyzing the users’ behavior. Collaborative filtering techniques such as user based and item based are some of the most widely used methods in recommendation. But there is a scope for improving the collaborative filtering techniques further especially in domains like automobile recommendation. In such areas, the dataset generated through multiple features or components improves the recommendation. A side-by-side feedback system which considers user reviews from different sources when integrated with collaborative algorithms produce better results. The paper is organized as follows. Section 2 briefly describes about the related papers. Section 3 specifies the proposed system followed by experiment and results in Section 4. Finally, Section 5 concludes the work.

Related Works

Many authors presented different recommendation systems and techniques to improve the suggestions or recommendations for users. This section comprises a brief mention about some methods proposed by different authors.

Thangavel et al. [1] discussed the recommendation system by customizing collaborative filtering and clustering using hadoop, a distributed environment. Though the speed up and efficiency is increased the scope for increasing accuracy is still left.

Nachiketa et al. [2] presented mixture model based algorithm to consider multiple aspects of items. The use of multiple components leads to better recommendation. But psychometric literature suggests that rating particular features of a subject is difficult. Though the algorithm is able to retrieve relevant items it's not an appropriate choice when training data is not sparse.

Yuanchun et al. [3] employed an associative classification method. It lays importance on customer's satisfaction. The model tries to maximize the customer's contentment by mining multi-class information and considers the user's profile to tap the after-sale purchasing satisfaction. But it doesn't describe features or preferences to improve the effectiveness of recommendation as customers often do not have clear idea about their needs.

Vivek et al. [4] proposed an approach with combined content-based recommendation and sentiment analysis. The method considered only two labels positive and negative for sentiment classification which results in elimination of many recommendations when coupled with content-based filtering. The user reviews for sentiment analysis are also sparse.

Zhuo et al. [5], proposed a recommendation system to recommend top-k ranked suggestions. The algorithm relies on parameters modelled specific to a user community.

Chong et al [13], proposed an algorithm to recommend scientific articles for users of an online community. The algorithm combines the merits of only the traditional collaborative filtering and probabilistic top modelling.

Our paper in contrast to other works proposes a framework for internal and external feedback analysis. The internal feedbacks are drawn from the users to whom the items are recommended and external feedbacks comprises the pool of comments from social media like twitter and other sources. Thus, it ensures a balance and generalizes the system rather than narrowing the reviews to internal feedbacks.

To address the issue of preference we also considered multicomponent ratings which are later dimensionally reduced using Principal Component Analysis (PCA). The reduced dimension or the principal component with the most variance is processed using collaborative filtering algorithms such as user based and item based. Hence, our work tries to improve the recommendation by combining the collaborative filtering algorithms with a broad feedback system.

Proposed System:

In this paper we propose a multi-criteria recommendation system using User Based and Item Based Collaborative Filtering techniques with external as well as internally mentioned feedback analysis. The proposed approach can be classified into following steps:

1. Data Collection
2. Feature Extraction
3. User Based Collaborative Filtering
4. Item Based Collaborative Filtering
5. Feedback Analysis

The Architecture of the System has been outlined in Fig 1.

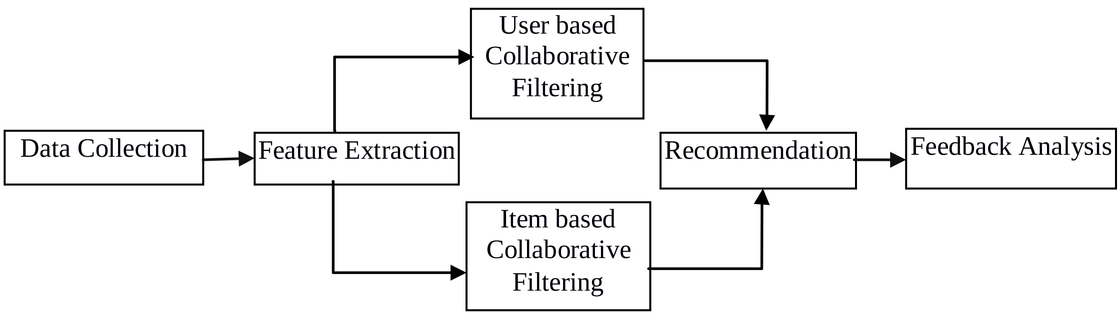

Figure 1: System Architecture

Data Collection

Due to the sparse availability of dataset on various features of automobiles, we made a survey of user requirements and the attributes they look into when they buy an automobile and were able to determine that the general public or user community's interest can be classified into five below mentioned parameters.

1) Design 2) Power 3) Maintenance 4) Safety 5) Comfort

All parameters mentioned above are intangible but quantifiable attributes which are derived from user experience and views of a product.

Feature Extraction

In this step we considered Dimensionality Reduction as accuracy of collaborative filtering algorithms do not scale very well with increasing parameters. Thus there arises a need to reduce or condense the data with minimal loss of information. Dimensionality Reduction is the process of choosing a mathematical representation within which we can describe the variance within your data, thereby retaining the relevant information, while reducing the amount of information necessary to represent it.

The general approaches used for Dimensionality Reduction are

1. Feature Selection
2. Feature Extraction

Feature Selection is manually hand selecting features or automating attributes that are to be dropped based on machine learning techniques that are highly discriminative. In this method many attributes are dropped from the original set and thus the new data set will be subset of previous input data.

Feature Extraction is a process involving generation of new features that are composites of existing features. In this approach no attribute is dropped which reduces loss of information.

In the proposed approach we used Feature Extraction to decrease the dimensionality. There are various Feature Extraction algorithms like Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Singular Value Decomposition (SVD). PCA and SVD are the most widely used techniques, both the algorithms are very similar and can almost be used interchangeably. In this approach we use PCA algorithm to reduce the parameters as there is no classification involved (no class label). The steps involved in PCA algorithm are illustrated in Fig 2.

We model the dataset with columns representing the parameters and the rows representing the user-item pair. The mean vector is calculated (mean of every dimension) followed by similarity measure by calculating covariance matrix. The eigenvector and the eigenvalue of the matrix are then found and using the eigenvector corresponding to the largest eigenvalue i.e the principal component is determined. The composite attribute is calculated using equation (1).

$$\mathbf{Y = W^T X D} \tag{1}$$

where W represents the k-highest eigenvectors and D is an n x 1 matrix representing the sample. In our case the value of k is 1 as we rescaled to a single feature and the value of n is 5 derived from 5 initial attributes.

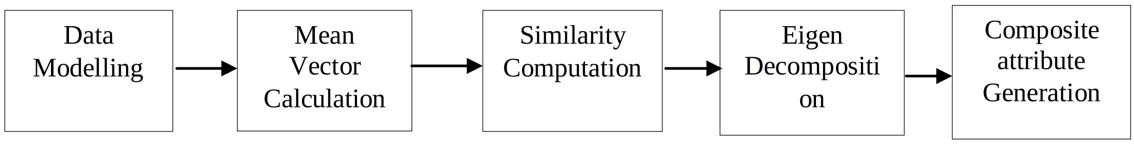

Figure 2: PCA Flowchart

User Based Collaborative Filtering

We have created a dataset and modelled it to use Collaborative Filtering (CF) algorithm, here we propose to use Pearson Correlation algorithm which is implemented using Mahout. In this we give the previously reduced dataset i.e. the dataset that consists of the Userid, Itemid and the composite rating as input to find similarity.

The dataset is modelled with row and column representing users and their composite ratings respectively. The modelled data is given as input and Pearson Coefficient i.e. the similarity index is calculated using Mahout and then the nearest neighbors are determined. This is used to find similar users. The steps of User Based CF are as shown in Fig 3.

Figure 3: Flowchart for User based Collaborative Filtering

Item Based Collaborative Filtering

We once again model the dataset as mentioned in the previous step and perform item based collaborative filtering using mahout. In this approach CF is done based on users past ratings, an item to item similarity matrix is computed by using the existing data model and when any user rates an item then the top n similar items derived from this matrix is recommended. The Steps of Item Based CF are as shown in Fig 4.

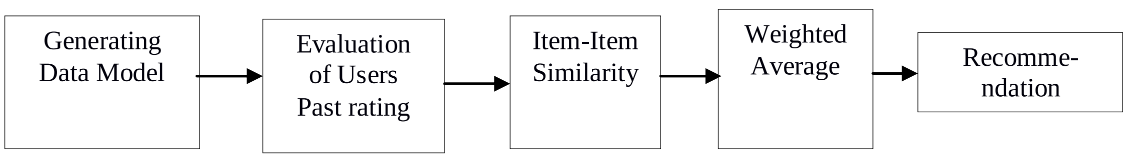

Figure 4: Flow chart for item based Collaborative Filtering

Combined Recommendation of User based and Item based Collaborative Filtering

We use a combined recommendation i.e. outputs from both user as well as item based CF are used to provide recommendations. In many scenarios, the system is unable to find similar users due to sparseness of users, in these situations item based recommendation gives better results. Thus whenever there is a lack of similar users to get ‘N’ recommendation, the item based recommendation is used to fill the niche. The flowchart is shown in Fig 5.

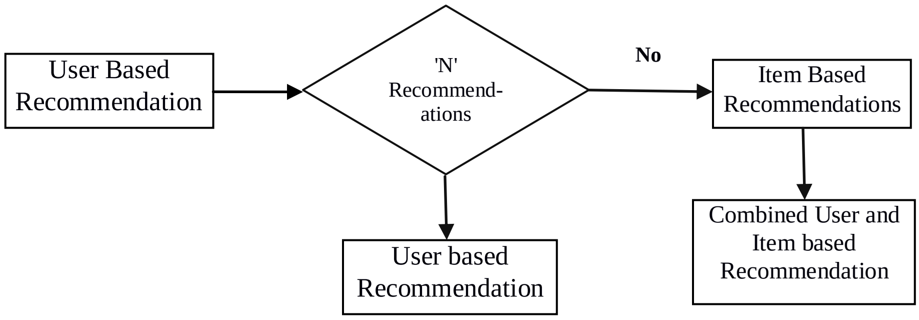

Figure 5: Flowchart for Combined User and Item based Collaborative Filtering

Feedback Analysis

In this part we propose a framework which can be used to integrate textual feedback internal (feedback from users on our recommendations) and external (comments in various web forums) in order to improve the recommendation systems. The Social media is also a large unexplored field of opinions which when properly used can help derive better recommendations. Here we are scraping comments from Social media as well as other automobile review web sites followed by stemming, stop word removal and opinion word extraction. The architecture is as shown in Fig 6.

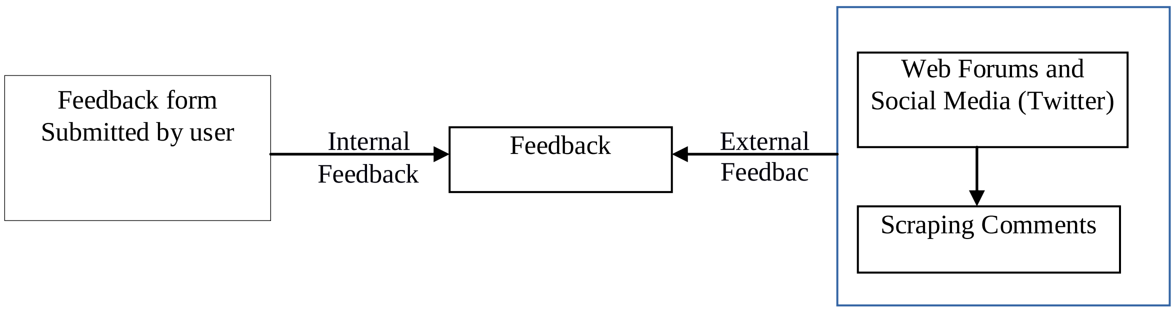

Figure 6: Feedback System Architecture

In order to process the reviews the first step is to convert all words to its root form and then remove stop words like ‘the’, ‘is’, ‘and’ etc. Then all opinion words are extracted and weighted depending on how positive or negative they are. This step also involves identification of negation words and parts of speech tagging which are used to identify opinion. The opinion is then used to classify the comment as very negative, negative, neutral, positive or very positive. This is fed back to the system which alters the weightage of recommendations based on the opinion rating.

Experiment and Results

Dataset
We have built our own dataset consisting of the previously mentioned parameters, it consists 1250 ratings from 50 users across 40 cars. Due to the sparse availability of any dataset in the domain and as the parameters being considered having not been used elsewhere we have generated a dataset by conducting an online survey through a blog. The sample dataset is shown in Table 1.

PCA algorithm
In this transformation, the principal component with maximum variance is used to generate the composite feature as shown in Table 2. Thus, translating variable space into optimal factor space decreases the overhead of recommendation system when compared to a dataset with multiple components.

User Based Recommendation
For the mentioned sample input in Table 3 the recommendations derived using user based CF are items 9,7,10,8,4 in the order of decreasing preference values. The threshold for user similarity is locked at 0.9 with top 'N' recommendation as 5. The output of user based collaborative filtering for user 33 is show in Table 4.

Item Based Recommendation
For the sample input in Table 3 the similar items derived using Item Based CF are as shown in Table 5.

Feedback Analysis
The relative URLs of various websites and blogs containing reviews of automobile was first extracted from twitter using Twitter API. The required comments is scrapped from the absolute URLs derived from the relative URLs. Then these comments are classified and opinion is extracted using sentiment analysis. The classification of comments and the output from sentiment analysis are shown in Table 6.

Table 1: Sample Dataset with multiple parameters

User Id	Username	Item Id	Item	Power	Design	Safety	Comfort	Maintenance
1	Anandraja	1	Nissan Sunny	4	2.5	3.5	5	5
2	Hariharan	10	Hyundai i20	4	4	5	5	4.5
3	Soundappan	1	Nissan Sunny	3.5	3.5	4	4	3.5
...	...	...	...	...	...	...	...	...

31	Venkatesh	20	Ford Endeavor	3.5	3.5	3.5	3	3
32	Kavin	10	Hyundai i20	3.5	3	2	4	4
32	Kavin	22	Nissan Terrano	4	3.5	3	3.5	4

Table 2: Sample Output of Feature Extraction

User ID	Item ID	Feature
1	2	2.72604
17	13	0.75602
…	…	…
23	19	1.92082
30	12	2.11839

Table 3: Input For Collaborative Filtering Considering User 33

User ID	Item ID	Feature
…	…	…
33	1	2.56238
33	2	3.77677
33	3	1.23851

Table 4: User 33's Recommended Items

Item ID	Preference Value
9	3.243069
7	2.4581885
10	2.3153925
8	2.077144
4	1.5103947

Table 5: Similar items of user 33's rated items

Item ID	Recommended Item ID	Item Similarity
1	3,2,5	1.0,1.0,0.94 (respectively)
2	1,5,7	1.0,0.94,0.9 (respectively)
3	2,1	1.0,1.0

Table 6: Comments and sentiment comparison for feedbacks

Comment	Sentiment
Tata Nano is worst in maintenance	Very Negative
We went for a long drive in Polo	Neutral
I enjoyed driving my new Nissan Sunny. Loved it.	Positive
Verna is truly best in its class	Very Positive

Metrics

The metrics used for recommendation systems are of 2 types Statistical Accuracy metrics and Decision-Support Accuracy metrics. In our paper we have used Statistical Accuracy Metrics using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for evaluating our result.

$$\textbf{MAE}= \frac{\sum_{i=1}^{n}|p_i-q_i|}{n} \quad \textbf{(2)}$$

where p =predicted ratings, q= actual ratings

The lower the MAE the more accurate the result.

Conclusion and Future Work

Recommendation Systems help people in taking decisions and makes their work simple, especially in today's world where e-commerce is growing at such an alarming rate such systems become invaluable. We in this paper have done multi-criteria rating taking five broad parameters and condensed to arrive at a single composite attribute. We used the composite feature to generate recommendations based on combined user and item based collaborative filtering algorithms. The enhanced recommendations is possible with added external and internal feedback analysis.

As the dataset grows and becomes larger there is a scope to improve the feedback analysis further. Hence as a part of our future work we intend to consider the demographic and psychological aspects and behaviors that drive the customers to choose the products of their choice. There arises scenarios where comments are more technical which are difficult to classify as positive or negative using corpus method, but these sentences might convey opinions which are invaluable. Further investigation is required to correct such situations and extract the appropriate sentiment.

References

[1] Thangavel Senthil Kumar, Swati Pandey, ”Customization of Recommendation System Using Collaborative Filtering Algorithm on Cloud Using Mahout”, Advances in Intelligent Systems and Computing, Vol. No.-321, pp 1-10, 2015.

[2] Nachiketa Sahoo, Ramayya Krishnan, George Duncan, Jamie Callan, ”Information Systems Research, Vol. No.-23,pp. 231–246,2012.

[3] Yuanchun Jiang,Jennifer Shang,Yezheng Liu, ”Maximizing customer satisfaction through an online recommendation system: A novel associative classification model”, Decision Support Systems, Vol. No.-48,pp 470-479,2010.

[4] Vivek Kumar Singh, Mousumi Mukherjee, Ghanshyam Kumar Mehta, ”Combining a Content Filtering Heuristic and Sentiment Analysis for Movie Recommendations”, Communications in Computer and Information Science, Vol. No.-157, pp 659-664, 2011.

[5] Zhuo Chen, Yong Feng, Heng Li, "A Novel Top-K Automobiles Probabilistic Recommendation Model Using User Preference and User Community", Proc. IEEE 11th International Conference on e-Business Engineering (ICEBE), pp. 105-111, 2014.

[6] Fengkun Liu, Hong Joo Lee,”Use of social network information to enhance collaborative filtering performance”, Expert Systems with Applications, Vol. No.-37, pp 4772- 4778, 2010.

[7] Xiaoyuan Su, Taghi M. Khoshgoftaar,”A survey of collaborative filtering techniques”, Advances in Artificial Intelligence, Vol.No.- 2009, Article No. 4,2009.

[8] Paolo Cremonesi, Yehuda Koren, Roberto Turrin,”Performance of recommender algorithms on top-n recommendation tasks”, Proc. of the fourth ACM conference on Recommender systems (RecSys '10), pp 39-46, 2010.

[9] Joseph A. Konstan, John Riedl, ”Recommender systems: from algorithms to user experience”, User Modelling and User-Adapted Interaction, Vol. No.-22,pp 101 -123,2012.

[10] Deuk Hee Park, Hyea Kyeong Kim, ll Young Choi, Jae Kyeong Kim,” A literature review and classification of recommender systems research”, Expert Systems with Applications, Vol. No.-39 ,pp 10059-10072,2012.

[11] Qing Li, Jia Wang, Yuanzhu Peter Chen, Zhangxi Lin, ”User comments for news recommendation in forum-based social media”, Information Sciences, Vol. No.- 180,pp 4929-4939,2010.

[12] Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang,” Recommender systems”, Physics Reports, Vol. No.-519,pp 1-49,2012.

[13] Chong Wang, David M. Blei “Collaborative topic modeling for recommending scientific articles”, Proc. 17TH international ACM SIGKDD international conference on Knowledge discovery and data mining, pp 448-456, 2011.