MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 – 6953

ONLINE MOVIE REVIEW SYSTEM

Rahul M. Sharma, Shital S. Barkul, Pankaj K. Sawane, Rahul M. Jeughale
Department of Computer Engineering
Marathwada Mitra Mandals College of Engineering
Pune, India
rahulsharma.comp@mmcoe.edu.in, shitalbarkul.comp@mmcoe.edu.in,
pankajsawane.comp@mmcoe.edu.in, rahul9619@gmail.com

Abstract: *The Online Movie Review System provides reviews and ratings to any movie and suggest movie to the user. This system generates a common review related to a movie by using Latent-Semantic Analysis (LSA) algorithm. LSA algorithm analyzes the relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. Here analysis of comments given by various users will be done and a common review will be generated. This generated review will be a simple English statement and will help user to take a correct decision while selecting any movie. Also, the system is providing a fast searching feature. Our system is going to provide annotating search results. This will be achieved by Alignment algorithm. With the help of this, user will be able to search his/her favorite movie with in no time. The Online Movie Review System also provides Recommendation option. If user wants to recommend his/her favorite movie to any other user/friend then he/she can. User recommend movie to his/her friends by sending a mail related to that movie. This mail will include factors like name of the movie, actor of the movie, producer and director of the movie, how much rating movie had got, etc. This mail will be sent with the help of JAVA MAIL API.*

Keywords*: Latent Semantic Analysis (LSA) algorithm*

1. INTRODUCTION

People’s opinion has become one of the extremely important sources for various services in ever-growing popular social networks. In particular, online opinions have turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities, and manage their reputations. In general, recommender systems are defined as the supporting systems which help users to find information, products, or services (such as books, movies, music, digital products, websites, and TV programs) by aggregating and analyzing suggestions from other users, which means reviews from various authorities, and user attributes. After viewing such reviews they take their decisions. So, such reviews must be correct and proper.

Generally, the reviews are generated in graphical format that is in star ratings. Users just have to see the ratings which are generated by analyzing the ratings given by other users to that product and have to take his/her decisions. Such ratings are easily understandable by any user. But they don't give clear idea of how the product is. They are helpful only in the scenario where if any product is excellent or very poor. The scenario where product is average, star ratings prove bit confuse for any user. They don't get clear views of what the other users think of that product.

If the reviews are in simple English statement it would be easy for any user to understand the feelings of the other users too, about the product. Also star ratings will be there for his/her help. So, the review about any product will give clear idea to any user so that he can easily take his/her decisions in such confusing scenario too.

Our system is a movie review system which will generate reviews related to the movies which are released. Unlike other systems, we are going to generate a common review by analyzing only the comments of the people (no heavy feedback). This will reduce the overhead of any user who is commenting on any movie and will make the system more user-friendly. So, the system will generate better review which will be a simple English statement. Also our system will provide star ratings too.

2. LITERATURE SURVEY

There are various IEEE papers which gives idea about the Recommender Systems. For our system we have referred papers as follows:

2.1. A Review Classification of Recommender Systems Research

The first paper we referred is a research paper. In this, author have identified 164 articles on recommender systems, which are published from 2001 to 2009 to understand the trend of recommender systems research and to provide practitioners and researchers with insight and future direction on recommender systems. Also author have implied some significant points like

- There are only 6 articles for image, movie and TV program recommendation. So, more researches are required to be studied for this.
- The approaches using social network analysis should be developed in the recommender systems as recently social network analysis has been used in the various applications.

2.2. 'MovieGen': A Movie Recommendation System

In this paper author has introduced "**MovieGEN**", an expert system for movie recommendation. SVM-based machine learning on training data and K-means cluster analysis on result sets of database inquiry constitute the key models of this system. The system takes in the users' personal information and predicts their movie preferences. Afterwards it clusters the movies and generates questions to refine the recommendation. Finally it suggests movies for the users. By the nature of the system, it is not a straightforward task to evaluate the performance since there is no right or wrong recommendation; it is just a matter of opinion. Based on informal evaluations that author carried out he got a positive response from the users. A larger data set will enable more meaningful results using this system. Additionally author would like to incorporate different machine learning and clustering algorithms and study the comparative results. Eventually he would like to implement a web based user interface that has a user database, and has the learning model tailored to each user.

2.3. A Movie Rating Approach and Application Based on Data Mining

In this work, author has presented a data mining application. He has applied the data mining to perform the moving rating in the prototypes. Many movie attributes are studied such as the producers, writers, actors, actresses, genres, subtitles etc. Based on study, it is found that the most effective attributes are genre and words used in the movies. The studied methodology is described. The database of movie information is created. This information is then extracted, cleaned and transformed to ‘Weka’ to create a decision tree. The difficult part is the extraction of words in the subtitles/short stories, selection of keywords, and keyword classifications that affect the rating. The derived model can perform about 80% accuracy with the usage of genre and word group attributes. The model is embedded in the web application that stores movie information and suggested the moving rating. Also, the approach can be further extended to consider other attributes and image processing techniques to extract contents for classification.

2.4. Privacy Risks in Recommender Systems

Recommender system users who rate items across disjoint domains face a privacy risk analogous to the one that occurs with statistical database queries. Recommender systems have become important tools in ecommerce. They combine one users ratings of products and services with ratings from other answer queries such as “Would I like X?” with predictions and suggestions. Users thus receive anonymous recommendation from people with similar tastes. While this process seems innocuous, it aggregates user preferences in ways analogous to statistical database queries, which can be exploited to identify information about the particular user. This is especially true for users with eclectic tastes who rate products across different types or domains in systems. The straddle highlights the conflict between personalization and privacy in recommender systems. While straddler’s enable serendipitous recommendations, information about the existence could be used in conjunction with other data sources to uncover identities and reveal personal details. We use graph-theoretic model to study the benefit from and risk to straddler’s.

3. SYSTEM OVERVIEW

3.1. Proposed System

Online Movie Review system is designed to overcome the drawbacks of existing system. It also provides fast searching for any movie and also viewing reviews of that movie and recommending movie to any of the friend by sending him/her a mail.
Main Aspects:

- Review Generation
- Recommendation

This system will provide fast searching as we implement Alignment algorithm. There is no issue regarding correct review generation as we are using LSA algorithm.

3.2. LSA Algorithm

LSA is a theory and method to analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. LSA can be applied to any type of count data over a discrete dyadic domain, which is so-called two-mode data.

Supposing that a collection of documents D = {d1, . . . , dn} with terms from W = {w1, . . . , wm} are given, then the system can construct a co-occurrence matrix M, where its dimension is n × m and each entry Mij denotes the number of times the term wj occurred in document di . Each document di is represented using a row vector, while each term wj is represented using a column vector.

LSA applies singular-value decomposition (SVD) to the term-document matrix M, and a low-rank approximation of the matrix M could be used to determine patterns in the relationships between the terms and concepts contained in the text,

$$\mathbf{M = U\Sigma Vt} \qquad \mathbf{(1)}$$

Where U and V are matrices with orthonormal columns (i.e., UT U = V T V = I), and Σ is a diagonal matrix whose diagonal elements are the singular values of M. The original term-document matrix could be approximated by reducing the dimensions of the term–document space, and this will allow the underlying latent relationships between terms and documents to be exploited during searching. Equation (2) shows that the reduced matrix ˜M is obtained by reducing the dimensionality, where the system truncates the singular-value matrix Σ to size k. It is this dimensionality-reduction step, i.e., the combining of surface information into a deeper abstraction, which captures the mutual implications of words and passages. Therefore, even though the original vector space is sparse, the corresponding low-dimensional space is typically not sparse. Practically, the number of dimensions retained in LSA is an empirical issue. We conducted the experiments under different dimensions in the experiment section

$$\tilde{}\mathbf{M} = \mathbf{U}\tilde{}\boldsymbol{\Sigma}\mathbf{V\,T} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V\,T} = \mathbf{M}. \qquad (\mathbf{2})$$

Algorithm 1 shows the algorithm, where the inputs include a term-document matrix, several product-feature seeds, the reduced dimensionality in SVD operation, and the number of extracted features for each seed. In Algorithm 1, lines 3 and4 are employed to perform linear algebra SVD operation on the term-document matrix, and lines 5–16 are used to compute the similarities between the seed product-feature vector and, pairwise, the other term vectors. The top ones will be collected as related product-feature terms for a specific product feature. The procedure **getTermVectorFromTermDocMatrix** is used to obtain the term-vector representation of a product feature. The seed is supposed to be one of the terms in the term-document matrix, and it is easy to obtain its corresponding document-vector representation. Meanwhile, sim in line 7 is used to store the similarities between the seed and the other terms. After sorting in descendant order, it is easy to obtain the top ones and their corresponding feature names in procedure **getTopRelatedFeatures**. When the above steps are completed, each product-feature seed can have its own semantically related term set. The advantage of this approach is that it could be applied to all the languages, it does not need any external dictionary, since LSA is language-independent, and it is based on linear algebra SVD operation.

Input: An nxm term-document matrix M, product feature seed set S, reduced dimension k, the number of extracted features for seed n.

Output: An association array F, where each key represents a product feature seed f and its corresponding value is f's related product features.

Pseudo code:

```
Start
    Initialize associated array F
    U, ~∑ ,Vt ← svd(M,K)
    ~M ←U × ~∑ × Vt
    For f Є S do
        Wf ←getTerm VectorFrom TermDocMatrix(f,~M)
        Initialize similarity list sim
        i←1
        foreachcolumn vector w of ~M do
            sim[i]←wf . w
            i←i+1
        end
        sort(sim)
        relatedFeaturelist←getTopRelatedFeatures(sim,n,~M)
        F[f]←relatedfeaturesList
    end
    return F
end
```

3.3. MOVIE REVIEW AND SUMMARIZATION FLOW

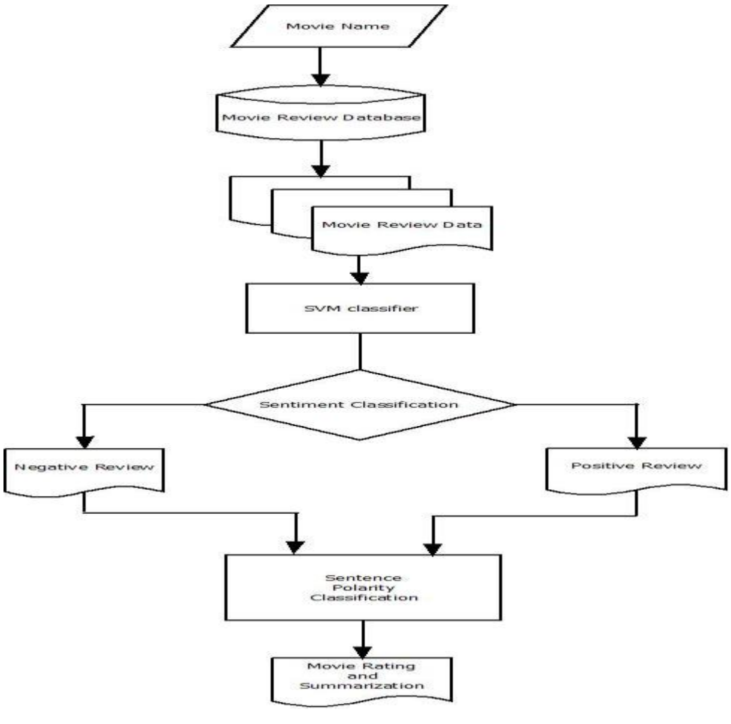

Fig.1: Movie review and summarization flow

4. CONCLUSION

In this paper, we design and implement a movie-rating and review-summarization system. Sentiment classification is applied to the movie reviews, and rating information is based on sentiment-classification results. In feature-based summarization, product-feature identification plays an essential role, and we propose a novel approach based on LSA to identify related product features. Sentiment analysis is done on the comments given by the users. The design proposed in this paper could fully utilize the Internet content to provide a new product-review summarization and rating service. The design can also be extended to other product-review domains easily.

5. FUTURE SCOPE

- Can be further extended for generating reviews related to the products in Online Shops.
- Also can be used for generating reviews for the online videos.
- Also can be used for generating reviews related to the colleges during admission process.
- Also can be used for generating reviews of the candidates in the election.

ACKNOWLEGEMENT

I feel immense pleasure in expressing my deepest sense of gratitude to my guide Mrs. Geetha Chillarge, for her encouragement and enlightened comments throughout this project work. Her appreciative suggestion always motivated me for putting most willing efforts on my study during project report. My special thanks to our head of department Prof. Harmeet Khanuja for providing co-operation. Last but not least, I express my thanks to all staff members of Computer laboratory for their enthusiasm and constant reminders to complete the work as soon as possible. I am also thankful to everyone in this college who directly or indirectly helped me in completing our project.

REFERENCES

[1] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu and Emery Jou, " Movie Rating and Review Summarization in Mobile Environment," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 3, May 2012.

[2] Yehuda Koren(Yahoo Research), Robert Bell and Chris Volinsky,(AT&T Labs—Research), "Matrix factorization techniques for recommender systems".

[3] Deuk Hee Park, Hyea Kyeong Kim, Jae Kyeong Kim, Il Young Choi and Jae Kyeong Kim ,"A Review and Classification of Recommender Systems Research" , School of Management, KyungHee University, Seoul, Korea, IPEDR vol.5 (2011).

[4] Eyrun A. Eyjolfsdottir, Gaurangi Tilak, Nan Li, "MovieGEN: A Movie Recommendation System", Computer Science Department, University of California Santa Barbara.

[5] S. Kabinsingha, S. Chindasorn, C. Chantrapornchai ,"A Movie Rating Approach and Application Based on Data Mining", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 1, July 2012 .

[6] 4) Naren Ramakrishnan, Benjamin J.Keller, and Batul J. Mirza(Virginia Tech), Ananth Y. Grama(Purdue University), George Karypis (University of Minnesota),"Privacy Risks in Recommender Systems".

[7] Rotten Tomatoes (http://www.rottentomatoes.com/)

[8] Meta Citric (http://www.metacritic.com/)

[9] Spill (http://www.spill.com/)

[10] Indy Red (http://www.indyred.com/)

[11] Movie Boozer (http://movieboozer.com/category/movie-review/)