**REVIEW ARTICLE**

# PATTERN EXTRACTION FOR MOVIE RATING AND REVIEW SUMMARIZATION BASED ON CUSTOMER REVIEWS

## M.KEERTHANA[1], M.LOVELINPONNFELCIAH[2]

[1]M.phil. Scholars, [2]Assistant Professor

Department of Computer Science, Bishop Heber College (Autonomous), Thiruchirapalli, India.

## ABSTRACT

The Online Movie Review System provides reviews and ratings to any movie and suggest movie to the user. This online movie review system generates a common review related to a movie by using Latent-Semantic Analysis (LSA) algorithm. LSA algorithm examines 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 many users will be done and a common review will be generated. The movie feature extraction is done by various methodologies such as Latent semantic analysis (LSA) algorithm and Frequency based algorithm. The result of LSA is extended to filtering mechanism to reduce the size of review summary. To design our system by consideration of sentiment classification accuracy & system response time. The similar design can be extended to other product review domain easily.

KEY WORDS: Latent Semantic Analysis (LSA) algorithm

## INTRODUCTION

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

The movie rating score is based on sentiment-classification results. Combined movie rating information with review summary, and display result to end users with the rating and summarization information about the movie. This will reduce the overhead of any user who is commenting on any movie and will make the system more user-friendly.

## Related work

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

Wilson et al, 2009; Kobayashi et al. 2007) have used classification based methods to integrate various features. But these methods separately extract object features and opinions, which ignore the correlation among output labels, i.e. object features and opinions.

Zhuang et al. (2006) summarized movie reviews by extracting object feature keywords and opinion keywords. Object feature-opinion sets were recognized by using a dependency grammar graph. However, it used a manually explained list of keywords to recognize movie features and opinions, and thus the system capability is restricted.

Qiu et al. (2009) exploit the relations of opinions and object features by adding some linguistic rules. However, they didn't care the opinion polarity. Our framework can not only hire various

features, but also exploit the correlations among the three types of expressions, i.e. object features, positive opinions, and negative opinions, in a unified framework.

Popescu and Etzioni (2005) proposed a relaxation labeling approach to utilize linguistic rules for opinion polarity detection. However, most of these studies focus on unsupervised methods, which are hard to integrate various features.

Hu and Liu (2004) proposed a statistical approach to capture object features using association rules. They only considered adjective as opinions, and the polarities of opinions are recognized with WordNet expansion to manually selected opinion seeds.

**Proposed Technique:**

**Sentiment Classification**

Support vector machine and, random forest used for sentiment classification of product reviews. Binary classifier for classification of movie reviews into positive and negative review classes.

**1) Support Vector Machine**

Training data contains the positive and negative reviews. Data provided in training classifier uses positive and negative reviews but they do not deal with complicated data which is very hard to classify. The reviews are said tobe of low variance if they are giving only positive or only negative opinions otherwise they are having high variance.

**2) Random Forest**

Random forest algorithm is highly accurate classifier which runs efficiently on large dataset. It handles thousands of input variables without variable deletion. It is an effective classification model for estimating missing data and maintains accuracy when large proportion of the data missing.
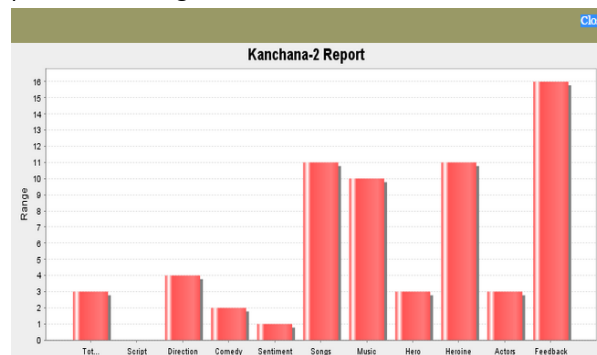
**3) Feature based Summarization**

Feature based summarization focused on the movie features on which the reviewers articulate their opinions. Identification of movie features and opinion words are both essential in feature based summarization. We propose a Latent semantic analysis (LSA) algorithm and frequency based algorithm to identify movie features. We compared these two movie-feature-identification approaches, in the experiment section.
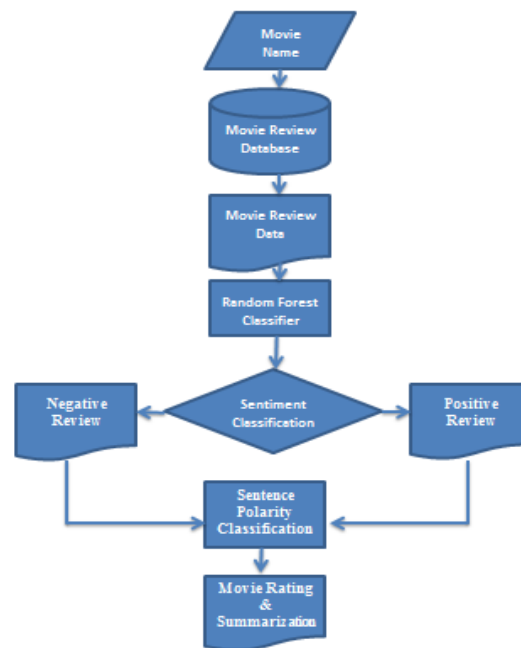


**Result and Discussion**

The result of LSA algorithm is extended to LSA based filtering mechanism is planned to employ the semantically related terms to reduce the size of review summary. LSA based purifying approach allows the users to choose the features in which they are interested. The system will generate summary of positive and negative movie review.



**CONCLUSION**

In this paper design and implement a movie-rating and review-summarization system. Sentiment classification is applied to the movie reviews and movie rating facts is based on sentiment-classification results. In feature-based summarization, product-feature identification plays a vital role, and we propose a novel approach based on LSA to identify related product features. Sentiment analysis is completed on the comments given by the

**M.KEERTHANA, M.LOVELINPONNFELCIAH**

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.

## 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]    DeukHee Park, HyeaKyeong 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, GaurangiTilak, Nan Li, "MovieGEN: A Movie Recommendation System", Computer Science Department, University of California Santa Barbara.

[5]    Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In Proceedings of the National Conference on Artificial Intelligence (pp. 755-760). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

[6]    Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

[7]    Poirier, D., Bothorel, C., De Neef, E. G., &Boullé, M. (2011). Automating opinion analysis in film reviews: the case of statistic versus linguistic approach. In Affective Computing and Sentiment Analysis (pp. 125-140). Springer Netherlands.

[8]    Jakob, N., &Gurevych, I. (2010, July). Using anaphora resolution to improve opinion target identification in movie reviews. In Proceedings of the ACL 2010 Conference Short Papers (pp. 263-268). Association for Computational Linguistics.

[9]    Freitas, L. A., & Vieira, R. (2013, May). Ontology based feature level opinion mining for portuguese reviews. In Proceedings of the 22nd international conference on World Wide Web companion (pp. 367-370). International World Wide Web Conferences Steering Committee.

[10]   Liu, C. L., Hsaio, W. H., Lee, C. H., Lu, G. C., &Jou, E. (2012). Movie rating and review summarization in mobile environment. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 42(3), 397-407.

[11]   T. Wilson, JanyceWiebe, and Paul Hoffmann 2009. Recognizing Contextual Polarity: an exploration of features for phrase-level sentiment analysis. Computational Linguistics 35(3).

[12]   N. Kobayashi, K. Inui, and Y. Matsumoto. 2007. Opinion Mining from Web documents: Extraction and Structurization. Journal of Japanese society for artificial intelligence.

[13]   L. Zhuang, Feng Jing and Xiaoyan Zhu. 2006. Movie Review Mining and Summarization. In Proceedings of CIKM.

[14]   G. Qiu, B. Liu, J. Bu and C. Chen. 2009. Expanding Domain Sentiment Lexicon through Double Propagation, International Joint Conference on Artificial Intelligence (IJCAI-09).

[15]   Popescu and O. Etzioni. 2005. Extracting Product Features and Opinions from Reviews. Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP'05), 339-346.