International Journal of Current Research and Modern Education (IJCRME)
ISSN (Online): 2455 – 5428 & Impact Factor: 3.165
Special Issue, NCFTCCPS - 2016

A MOVIE RATING APPROACH BASED ON OPINION MINING

C. Arivazhagan*, P. Govindan, M. Manikandan**
& K. R. Praveen****

* Assistant Professor, Department of Computer Science & Engineering, Jayam College Of Engineering & Technology, Tamilnadu
** Student, Department of Computer Science & Engineering, Jayam College of Engineering & Technology, Tamilnadu

Abstract:

With the tremendous growth of information available on the Web, there is a need of classifying it for the ease of use and for the brisk accessibility. The classified data can be used for making recommendations. This paper discusses and compares three classifiers applied on movie data-set using WEKA 3.7 data mining tool. The data is classified into five different classes namely: bad, ok, average, good and excellent. A discussion about true positive rate, false positive rate, precision, and recall based on confusion matrix for each class is carried out. Subsequently the boundary visualization of data is captured inform of a graph, and a meaningful comparison between Zero R rule, Naïve Bayes classifier and J48 tree is done by experimentation and analysis. In this paper, an attempt has been made to analyze the best classifier for movie data based on users' ratings and then the classification is used for making the recommendations for users.

Introduction:

Computational process of discovering patterns in large datasets. Involving methods at the intersection of Artificial Intelligence, Machine learning Statistics, database systems. The overall goal of the data mining is to extract information from a data set, then transform it into an understandable structure for further use. The WEKA 3.7 data mining tool for making classification of data has been used for this purpose. In the different classification methods in carried out, by analyzing various properties like true positive rate, false positive rate, precision, recall, on movie data. This analysis and its results prove that J48 tree classifier outperforms Zero R rule and Naïve Bayes classifier.

Existing System:

The system which classifying the data, The system discusses and compares three classifiers applied on movie data set using WEKA 3.7 data mining tool, Zero R rule, Naive Bayes classifier, J48 algorithm.

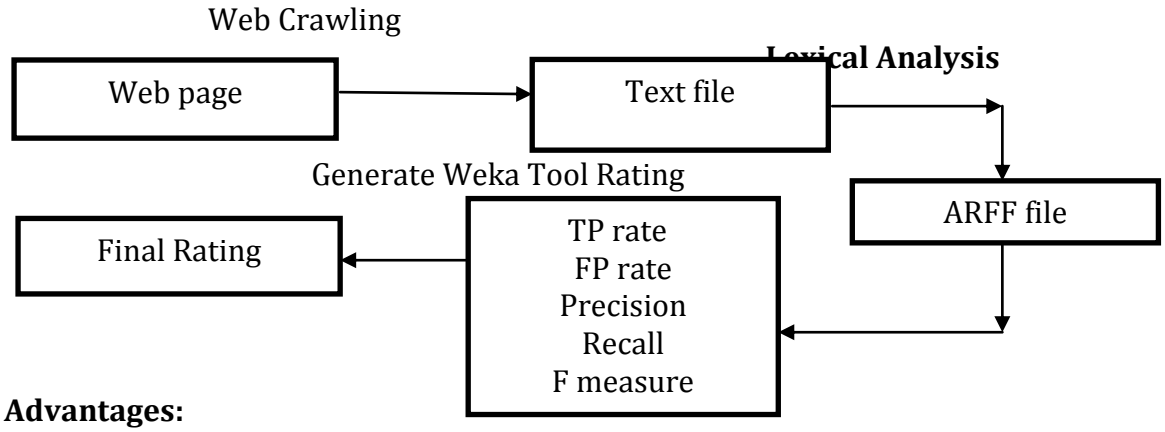
Limitations:

The input data's are existed data sets, there is no way to create dataset from opinions. The classification can be done only on movie data but not on other data sets from different social websites.

Proposed System:

Create ARFF (dataset) file through the web crawling in php. Extracting the opinions by using lexical or semantic analyses, rating generation through j48 classification algorithm. The quality of recommendations can be improved by applying other criteria along with collaborative filtering and similarity of content.

System Design:



Advantages:

No need to create dataset separately, classifies the data using different algorithms. Creates the rating with the classification, Produce the more accurate rating.

Modules:

Web Crawling Module:

Systematically browses the World Wide Web, typically for the purpose of Web indexing. The number of pages on the internet is extremely large, even the largest crawler's fall short of making a complete index, crawlers can validate hyperlinks and HTML code.

Lexical Analysis:

The process of converting a sequence of characters into a sequence of tokens, a program that performs lexical analysis may be called a lexer or tokenizer. A lexer is generally combined with a parser, which together analyze the syntax of programming languages & web pages.

Generation of ARFF:

The ARFF file means (Attribute Relation File Format), an ARFF file is an ASCII text file that describes a list of instances sharing a set of attributes. There three keywords used for generate ARFF @RELATION, @ ATTRIBUTE, @DATA.

Conclusion:

Generate the dataset (ARFF) from the web page using web crawling & lexical analysis.The use of three classifiers is conducted in WEKA for classifying movie dataset. The outputs of the classifier are used for generating recommendations & rating.

References:

1. R. Hooda, K. Singh and S. Dhawan, "A Study of Recommender System on Social Networks and Content-based Web Systems," Proceedings of International Journal of Computer Applicatios, vol. 97(4), pp. 23-28, 2014.
2. Jyoti, S. Dhawan and K. Singh, "Critical Investigation of Recommendation Systems for Online Shopping and Social Websites," Proceedings of 2nd National Conference on Converging Technologies Beyond 2020, pp.331-334, 2014.
3. C. Rana and S. K. Jain, "Building a Book Recommender System using Time Based Content Filtering," WSEAS Transactions on Computers, Vol. 11 (2), pp. 27-33, 2012.
4. G Nandi, A Das," A Survey on Using Data Mining Techniques for Online Social Network Analysis", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.