Ranking Prediction of Search Engines

Abhilasha[1] Jhalak Jain[2] Ajay Kumar Kaushik[3]
[1,2]Student [3]Assistant Professor
[1,2,3]Department of Information Technology
[1,2,3]MAIT, GGSIPU, Delhi, India

Abstract— This project aims to do ranking prediction of various search engines available on the web. The objective of the project is to do feedback analysis of various search engines. This project deals with study of Naïve Bayes algorithm which is used in feedback analysis.
Key words: Data Mining, Feedback Analysis, Naïve Bayes Algorithm

I. INTRODUCTION

Cloud Computing is Internet-based computing, and it is shared by a configurable resources that is provided to computers and other services Examples of configurable shared resources are infrastructure, platform, and softwar. Different cloud applications may receive different levels of quality for same cloud services so that optimal service selection becomes important. Selecting the most valuable search engine is mostly important.

In today's era with the rapid growth of information on the web, makes users turn to search engines. Computer program that search database and internet sites for the documents containing keywords specified by user is know as search engines. It is a website whose primary function is providing a search engine for gathering and reporting information available on the internet. Most popular search engines include Google, Ask, Bing, Yahoo, etc.

II. AIM OF THE STUDY

Internet is gaining popularity. People are regularly using it for their convenience. Due to the availability of large amount of data many search engines are available. People tend to be more confused when it comes to choose the best. So user's comment becomes the vital data to evaluate the quality of search engines. However, for many search engines the amount of reviews is too large to be processed manually. Such reviews are very helpful to the users. This paper presents a study of ranking search engines and doing feedback analysis of user's opinion through data mining.

The aim of the study is to evaluate the performance of different search engines both in terms of accuracy and effectiveness. Search engines are ranked according to the results and the time taken by each of them to search the typed query. Feedback analysis or sentimental analysis is very important aspect in every aspect. To know whether the users are satisfied with the results or not. Feedback plays a vital role in every field whether social media, e-commerce websites or even hotels and restaurants. The customer gives their feedback to let them know about their service.

Feedback analysis helps a user to determine which site gives the best result according to the query searched for and which site does not. User feedback plays a vital role for ranking different search engines.

III. RESEARCH OBJECTIVE

Availability of different search engines makes sorting of particular information through billion of pages and displaying the relevant data the task tough for search engine. Different search engines display the result according to the query typed and according to it users have their own feedback. According to user's feedback choosing the optimal search engine is important task. Remedy for this is data mining. Data mining is the act of finding information that are statistically useful, not known previously, and extractable from data. The overall goal of data mining process is to extract information from a data set and transform it into an understandable structure for further use.

Further we will use Naïve Bayes algorithm for the process of data mining. It is useful for choosing the most optimal search engine. It is a classification technique based on Bayes' Theorem. In other words, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to construct and particularly useful for very large data sets.

The detailed feedback analysis report would be generated which would give the feedback of users in a tabular form, also the feedback comments would be segregated based on the concept of data mining. There would be 3 column for the feedback that is positive feedback, negative and neutral feedback. The user can also suggest improvements. The detailed feedback analysis report would be generated which would give the feedback of users in a tabular form, also the feedback comments would be segregated based on the concept of data mining. There would be 3 column for the feedback that is positive feedback, negative and neutral feedback. The user can also suggest improvements.

IV. METHODOLOGY

In the recent years, World Wide Web has gained popularity and become the second most widely used application of the internet. Search engines are used to extract data from www as per the users query. Since search engines consist of large collection of data it becomes difficult for users to opt for the right one. Hence there is a need for a system which is used for feedback analysis for different search engines.

Due to the availability of different search engines on internet users often get confused which website is giving better result and which website is not. In every field user opinion matters the most. Hence, feedback analysis gives user the privilege to give his feedback and get an overview which website is more suitable.

In this system, a user is asked to search a query. Based on the query different search engines such as Google ,Yahoo ,Bing etc display the result that is obtained on the

very first page. The user can go to each search engine page and see the result obtained. By clicking on each link the user goes to the respective site. Two graphs are obtained which show the time taken by each search engine to search the query and the number of results on the very first page. Now each search engines show different result. It depends on the user whether he likes it or not. The user can give his feedback. And at the end graph will be obtained depending on user feedback. The graphical representation gives a clear view about the user's opinion.

For feedback analysis, concept of data mining is used. Data mining involves the process of extracting useful information from given data. For the technique of positive and negative feedback Naïve Bayes algorithm is used. Naive Bayes is based on Bayes theorem with independence assumption between predictors. It is used to calculate posterior probability.

V. WHAT IS DATA MINING

The purpose of this paper is to select most valuable search engine through data mining. In simple terms, the process of transforming data into useful information is known as data mining. In other words, it is mining of knowledge from data. A large amount of data is available these days due to increasing computerization and digitization.

Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data. The overall goal of data mining process is to extract information from a data set and transform it into an understandable structure for further use.

The main aim of data mining is to find valuable data from large set of data. Several analysis tools of data mining (like regression, clustering, and classification) can be used for sentimental analysis task. Semantic analysis is based on extracting important data from positive or negative reviews of collected data.

The source material refers to opinions or reviews given in various search engines. The sentiments found in comments, feedback or reviews of users can be mined and use for analysis. From data mining we tend to find out if a review is positive one or negative one. If the keyword is the website is good then it is a positive review and if it is a bad website then it is bad review.

Data mining help us to extract valuable information and pattern from large amount of data sets. Data mining is very useful in the fields of science and computer. As, an application of data mining, users can learn for about the websites and go for the most optimal one. The process of data mining involves data collection, warehousing and computer processing. To do segmentation of the data and evaluation of the probability of future events, data mining uses sophisticated mathematical algorithms. Data mining is also known as data discovery or knowledge in discovery.

The steps involved in data mining process are:
- First we extract, then transform and load the data into a data warehouse
- Then the data is stored and managed in a multidimensional databases
- Then we provide data access to business analysts using application software

- Lastly we present analyzed data in easily understandable forms, such as graphs

The difficulty which can be faced in the feedback analysis is that an opinion word which is treated as positive side may be considered as negative side in another situation. Also, the degree of positivity or negativity also has a great impact on the opinions. For example good and very good cannot be treated same. Additionally a lot of words may not be mentioned in the data set which can lead to error in semantic analysis. Suppose the user has typed in the feedback useless and the word is not present in the data set then it can lead to error the software may not be able to decide whether it comes under positive or negative.

VI. HOW DOES NAIVE BAYES WORKS

In this paper of feedback analysis we are using Naïve Bayes to take the user's opinions and get an idea about the accuracy of each search engines. Naïve Bayes algorithm is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In other words, a Naive Bayes classifier works on the assumption that if a particular feature is present in a class then it is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'. Naive Bayes model is easy to build and particularly useful for very large data sets.

Naïve Bayes is easy to build and is used for large data sets. Naive Bayes is the most recommended algorithm in sentimental analysis (to identify positive and negative reviews of users mostly in social media, e-commerce websites and search engines).Naive Bayes is out perform even highly sophisticated classification methods. Naive Bayes is appealing because of its simplicity, elegance and robustness. Applications of Naïve Bayes also include test classification and spam filtering.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. Look at the equation below:

$$P(c|x)=p(x|c)p(c)/P(x)$$

Where,
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given 9*class*.
- $P(x)$ is the prior probability of *predictor*.
- $P(c|x)$ is the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).

Naïve Bayes is highly useful in the fields of real time prediction due to its high speed processing, and hence could be used for making predictions in real time. This algorithm is also well known for its use in multi class prediction. Here, we can predict probability of multiple classes of target variable.

Application of Naive Bayes classifiers includes text classification (due to better result in multi class problems and independence rule) which have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment

Analysis (in social media analysis, to identify positive and negative customer sentiments).

The combination of Naive Bayes Classifier and Collaborative Filtering is used to build a Recommendation System that widely uses the technique of machine learning and data mining to remove unseen information and see whether a user would like a given resource or not

VII. CONCLUSION

In this paper, our proposed system is to predict ranking of different search engines and to take feedback of different users. The time taken by each website and the number of results are depicted through pie charts or bar graphs. Based on the query searched for, results are display for different search engines. User are allowed to give their feedback and so feedback analysis takes place which involves the process of data mining and tells which search engine has more positive reviews and which has more negative reviews. This helps a user to get an overall idea about the accuracy and effectiveness of different search engines. Thus the user can opt for the most convenient website.

ACKNOWLEDGMENT

REFERENCES

[1] P.A. Bonatti and P. Festa, "On Optimal Service Selection," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 530-538,2005.

[2] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Ann. Conf. Uncertainty in Artificial Intelligence (UAI '98), pp. 43-52,1998.

[3] M. Deshpande and G. Karypis, "Item-Based Top-n Recommendation," ACM Trans. Information System, vol. 22, no. 1, pp. 143-177,2004.

[4] R. Jin, J.Y. Chai, and L. Si, "An Automatic Weighting Scheme for Collaborative Filtering," Proc. 27th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '04), pp. 337-344, 2004.

[5] G. Linden, B. Smith, and J. York, "Amazon. Com Recommendations: Item-to- Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76- 80, Jan. /Feb. 2003.