An Extreme Machine Learning Model to Validate the Performance of Web Search Engine

Althaf Ali A, Dr. R. Mahammad Shafi

Abstract--Measuring the Search engine retrieval effectiveness is employed in the small scale using data mining approaches in the past decade. At present, large exploring technique has been employed using unsupervised learning model with sort of small queries and less set of web documents. Despite of those advantages, still some challenges has been propagating on the evaluation in terms of examination time and accuracy prediction on the large scale search engine retrieval effectiveness. In this paper, a new extreme machine learning model has exploited under ensembling structure to determine the information retrieval results against evolving user queries and time drifts.

In addition, proposed model determines the association of the web document for navigational and non factoid queries through auto indexing of the data with high vagueness and uncertainty. Furthermore, periodic optimization to the model has been made on the distance estimation. The experimental results prove that extreme machine learning model produces better performance compared to the state of art approaches in the literatures in terms of precision and recall. Finally, it exhibits more consistency among other technique on the evaluation of the search engine.

Keywords--Extreme Machine Learning, Search Engine Evaluation, Performance Analysis, Accuracy, Evolving Queries

1 INTRODUCTION

Search Engine Evaluation is a complex task during which many interconnected factors has to consider determining the efficiency. It is often subjected with a set of possible actions (options/decisions) for each user queries and each result has associated possible (short-term and long-term) consequences, which are uncertain [1]. Probabilities on the extraction of the result in term of information retrieval provide quantification of evidence to handle uncertainty in the query during the query expansion and web document exploiting. Those processes can be implemented with help of machine learning algorithm.

The current research is to evaluate the search engine which employs the data mining approaches to extract the information in the web repositories. The approaches and its performance are evaluated using another set of mechanism. In this paper, an approach employed to evaluate the search engines is extreme machine learning on inclusion of ensemble model in parallel. It provides the quantitative insight about the process employed on the search engines under different circumstances [2].

- *Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore-641046, Tamilnadu, INDIA. PH- +91 9703005881. E-mail: althafa579@gmail.com*
- *Research Supervisor, Department of Computer Science, Bharathiar University, Coimbatore-641046, Tamilnadu, INDIA. PH- +91 9951622786. E-mail: rmdshafi@gmail.com*

The proposed approach provides a natural statistical analysis of the information retrieved through by incorporating an integrated summary of the available information. More specifically, it offers various advantages such as maintaining observations, statistical distributions, prior assumptions, and expert judgment on various query types. In addition, their outcomes to give the best result have been determined. It works beyond the capabilities of other approaches [3].

The remainder of the paper is organized as follows: Section 2 discusses the related works in evaluation of IR methods and its impacts against the performing evolving user queries, Section 3 briefly discusses the proposed technique in terms profiling and testing the IR method and Section 4 presents the experimental results on a number of data sets. Section 5 discusses conclusions and future work.

2 RELATED WORK

There exist many techniques to Information Retrieval Evaluation are designed and implemented efficiently. Each of these techniques follows some sort of effectiveness on the evaluation of the System among few performs nearly equivalent to the proposed model which is described as follows

2.1 An empirical analysis of software effort estimation with outlier elimination

In this technique, accurate software effort estimation has been proposed using unsupervised learning model named as KNN algorithm. It improves the estimation accuracy of software efforts which focused on effort estimation methods without any consideration of data quality. Although data quality is one of important factors to impact to the estimation accuracy, it can be compensated on investigating the influence of outlier elimination upon the accuracy of software effort estimation [4].

2.2 Use of relative code churn measures to predict system defect density

In this model, analysis of Software systems has been carried out on various segments like data evolving over time due to changes in requirements, optimization of code, fixes for security and reliability bugs. Code churn, which measures the changes made to a component over a period of time, quantifies the extent of this change. It is a technique for early prediction of system defect density using a set of relative code churn measures that relate the amount of churn to other variables such as component size and the temporal extent of churn [5].

3 Proposed Model

In this section, we model an extreme machine learning based performance exploring approach towards testing of the search engine on their information retrieval results. The proposed model incorporates in outlier estimation based on data uncertainty. Detailed description of design is as follows

3.1 Query Processing against various types

In this module, way of the representative of queries to the search operation has to be examined. Uncertain query has to be converted to certain query Retrieval Effectiveness to the search operation using query reconstruction algorithm [6]. The Fig1. represents the architecture diagram of the proposed model. The query is reconstructed in terms of misspelled works, word placement and concept of the query.

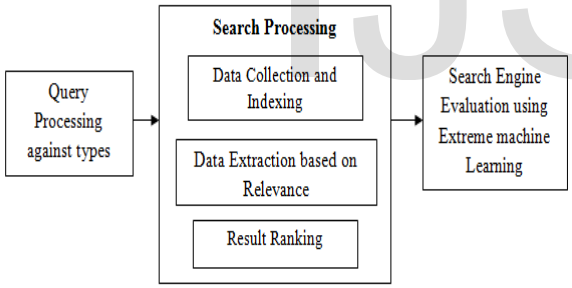

Fig1. Architecture diagram of the proposed model

3.2 Search Processing using unsupervised learning Model

The search processing of the query is carried on the web documents. Initially, data relevant to search queries are collected and it is indexed according using data mining approaches such as association rule mining or K NN algorithm. After data classification, data related to the search queries is extracted. In order to produce the result to user, extracted information is ranked using page rank algorithm [7].

3.3 Search Engine Evaluation using Extreme Machine Learning

Extreme machine learning model is employed to determine the Shortcomings of current Web search evaluation methodology. Those were identified and recommend making for future improvements based on the accuracy in retrieving results based on the user queries. It is based on random projection with several evaluation strategies and constrainst to the result set [8].

4 Experimental Analysis

In section, the experimental results of the proposed evaluation model and it is described on the two simulated web search engine with web document size above 2pb . The evaluation of the search engine has been done with two popular metric through asp.net is as follows

- **Precision**

The precision of an Information retrieval system to the uncertain query on the search engines is the proportion of results that are relevant. The performance outcome is represented in the Fig 2.

$$\text{Precision} = P = \frac{Releveant\ retrieved\ result\ set}{Overall\ Retrived\ Result\ set} \quad (1)$$

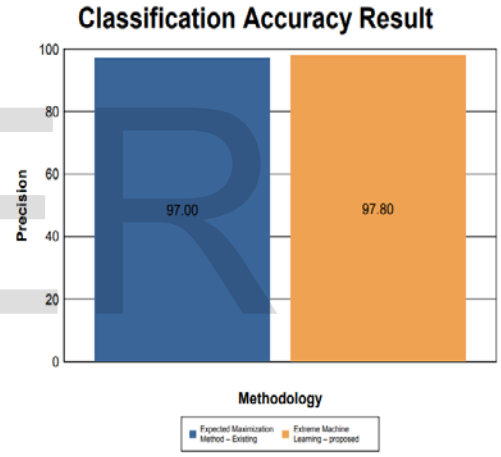

Fig 2. Performance Analysis of the Search Engine Evaluation against the Precision

- **Recall**

The Recall of an Information retrieval system for an uncertain query to the web search evaluation is the proportion of relevant results that have been retrieved and its performance outcome is represented in Fig 3. and Table 1

$$\text{Recall} = R = \frac{Releveant\ retrieved\ result\ set}{Relevant\ Result\ in\ database} \quad (2)$$

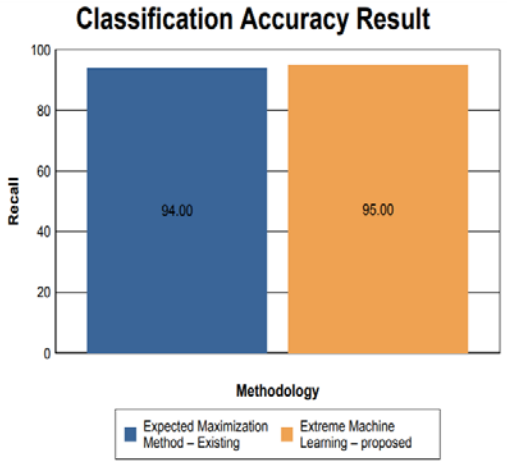

Fig 3. Performance Analysis of the Search Engine Evaluation against the Recall

The evaluation model generates the high possible evaluation results compared with other possibilities approaches in information retrieved to different class or categories of queries to the search engines.

TABLE 1
PERFORMANCE ANALYSIS OF THE SEARCH ENGINE EVALUATION

Technique	Precision	Recall
Expected Maximization Method – Existing	97	94
Extreme Machine Learning – proposed	97.8	95

5 CONCLUSION

The Extreme Machine Learning model has been designed and implemented to measure the search engine efficiency on large exploring web document and information retrieval result. It is capable of investigating the result on the various query types. The proposed model determined the processing model to the uncertain queries and evolving user queries. Periodic optimization has been done to proposed model to increase the performance of the accuracy and reduce the time of computation. It has proved to better consistency among other techniques.

REFERENCES

[1] Lempel, R.; & Moran, S. Predictive caching and prefetching of query result in search engines. In Proceedings of the 12th World Wide Web Conference (WWW2003), Budapest, Hungary, 2003.

[2] Hamid Sadeghi. "Automatic Performance Evaluation of Web search Engines using judgements of Meta search Engines", Online Information Review,ISSN:1468-4527,Emerald Publishing Limited, pp.957-971, 2011

[3] X. Benavent, A. Garcia-Serrano, R. Granados, J. Benavent, and E. de Ves, "Multimedia information retrieval based on late semantic fusion approaches: Experiments on a wikipedia image collection," IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 2009–2021, Dec 2013.

[4] Y.-S. Seo, K.-A. Yoon, and D.-H. Bae, "An empirical analysis of software effort estimation with outlier elimination," in Proc. 4th Int. Workshop Predictor Models Softw. Eng., 2008, pp. 25–32. [Online]. Available: http://doi.acm.org/10.1145/1370788.1370796

[5] N. Nagappan and B. Thomas, "Use of relative code churn measures to predict system defect density," in Proc. Int. Conf. Softw. Eng., 2005, pp.15–21.

[6] S. Chulani, B. Boehm, and B. Steece, "Bayesian analysis of empirical software engineering cost models," IEEE Trans. Softw. Eng., vol. 25, no. 4, pp. 573 –583, Jul. 1999.

[7] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. (2007, Apr.). Lessons from applying the systematic literature review process within the software engineering domain. J. Syst. Softw. [Online]. 80(4), pp. 571–583. Available: http://dx.doi.org/10.1016/j.jss.2006.07.009

[8] D. Marquez, M. Neil, and N. Fenton, "Improved reliability modeling using Bayesian networks and dynamic discretization," Rel. Eng. Syst. Saf., vol. 95, pp. 412–425, 2010.