© UNIVERSAL RESEARCH REPORTS | REFEREED | PEER REVIEWED
ISSN : 2348 - 5612 | Volume : 05 , Issue : 05 | April – June 2018

Enhance the Efficiency of Query by Appling Genetic Algorithm Using Weighted Cosine Matching Function

Manoj Chahal
Master of Technology (Computer Science and Engineering) Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India[1]

Abstract: - Query Optimization is a technique which refines the search space and increase the relevance of retrieved information. It help user to extract required information quickly and effectively. There are various methods which are used to optimize user query. In this paper Genetic Algorithm and weighted matching function is used for query Optimization. Genetic Algorithm is used for optimization. It explores and exploits user search space. It inspired by biological optimization algorithm. Matching Function is a technique which is used to measure the degree of similarity between query and document. By using GA and matching function efficiency of user query is enhance.

ISSN : 2348-5612 © URR
9 770234 856124

Keywords: Query Optimization, Genetic Algorithm, Weighted TF-IDF Cosine Matching Function, Evolutionary Computation.

I. INTRODUCTION

User Query is used to retrieve relevant information from large search space. It explores the whole space and finds relevant information according to user requirement. The search can be further refined by applying optimization technique. Optimization increases the efficiency of query and retrieves more important information. In order to optimize query genetic Algorithm is applied. In this paper Genetic Algorithm and Weighted Cosine matching Function is applied to get Optimized query.

Genetic Algorithm was inspired on Darwin's theory of evolution. It is adaptive heuristic search algorithm based on the evolutionary idea of natural selection and genetics. Genetic algorithm has been widely studied experimented and applied in many field in the engineering world. Genetic algorithm Operation can be used to generate new and better generation. As shown in flow chart the Genetic Algorithm step includes: -

Initially query and document is used to generate initial population which is given as input to genetic algorithm. Evaluating fitness value is used to measure the fitness value of given Query with documents and selection operation is applied on the entire population or chromosome with the help of roulette wheel. Based on the fitness value optimization criteria are tested if optimization is achieved then process is terminated.

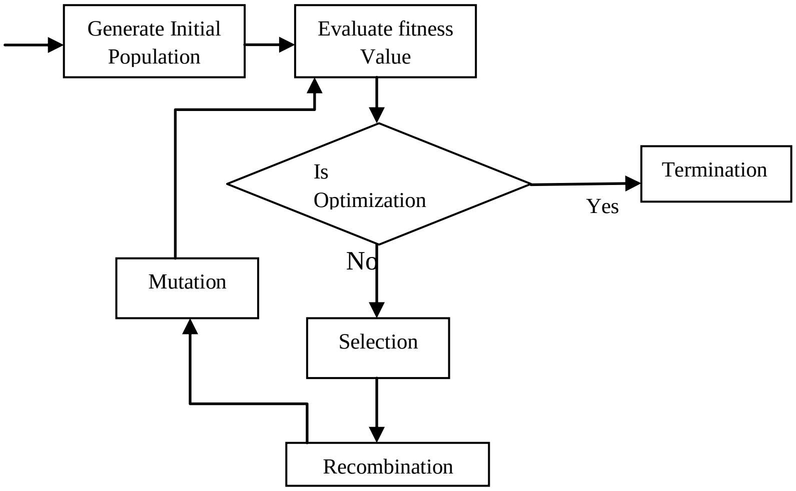


Fig 1.1 flow chart for genetic algorithm

If optimization criteria are not achieved then basic operation of genetic algorithm selection, Recombination and Mutation is applied. This basic operation generates new better population then the last input population. Crossover is responsible to search different area of search space and mutation is responsible to keep diversity in the population. This process is repeated until we get optimized value for the given population. If no accepted solution are found the genetic algorithm may be restarted or a fresh search is initiated.

Similarity function is used to measure the degree of similarity between query and document. It measures the optimization criteria between query and document. In this paper weighted similarity measure is used to find the similarity between user query and document. This weighted similarity measure combine with genetic algorithm to increase the efficiency of user query.

II. Literature Survey

There are several studies that used Genetic Algorithm for query optimization and measuring similarity between query and documents.

Miss Komal R.Hole et al.[1] described Canonical genetic algorithm and task of image preprocessing. It also introduced various approaches based on genetic algorithm to get image with good and natural construct and also explained image enhancement and image segmentation using genetic algorithm. Abdelmgeid A. Aly [2] described genetic algorithm to modify user queries based on relevance judgements. It also conducted experiment to reterive relevance document on three well known document collection: CTST , NLP and CACM. NIRPAL P.B and Kale K.V [3] described genetic algorithm for software reliability and test coverage to increase the software efficiency. They used genetic algorithm to generate test case for selected path automatically. L.M.Q Abualigah et al. [4] described the use of genetic algorithm in information retrival system and also explained the method to enhance the efficiency of IRS by applying genetic algorithm.

Sergiy D.Pogorilyy et al. [5] described the use of genetic algorithm in networking. It explained how to optimize network performance by applying genetic algorithm. Richa Garg et al. [6] described the effect of hybridization of local search with replacement operators on the performance of genetc algorithm. It also discussed with the helpof graph that hybrid algorithm is converging toward optimal more quickly than conventional algorithm. Tarek A.El-Mihoub et al. [7] described different forms of integration between genetic algorithm and other search and optimization techniques. It also examined various issue that are taken into consideration when desigining a hybrid genetic algorithm. Cristina Lopez Pujalte, Felix de Moya Anegon et al [8] described various order based fitness functions than evaluate efficiency of genetic algorithm using this fitness function for relevance feedback. Vaibhav Chaudhary, Pushpa Rani Suri [9] discussed the impact of optimization using genetic algorithm and share genetic algorithm on multimodal image registration by considering mutual information concept. Chengjun Liu [10] proposed that popular whitened cosine similarity measure is related to the Bayes decision rule under specification assumptions and presented two new similarity measures first PRM whitened cosine similarity measure and second within class whitened cosine similarity measure. Philomina Simon and S. Siva Sathya [11] described a general frame work of information retrieval system. The applicability of genetic algorithm was discussed in different areas of information retrieval such as genetic mining, query optimization, document clustering, and query optimization etc. Siti Nurkhadijah Aishah Ibrahim et al. [12] presented a model of hybrid GA-Particle Swarm Optimization (HGAPSO) based query optimization for Web information retrieval. The keywords are used to produce new keywords that are related to the user search.

III. PROCESS OF EXPERIMENT

- User put query on search engine
- Search engine give result for user query then extract top ten documents.
- Find the Weighted term frequency for each document using Text Analyzer.
- This Weighted TF and IDF is used to create initial chromosome or population. This is used to give input to the Genetic Algorithm.
- Weighted Cosine similarity function is used for calculating fitness value.
- Apply basic operation of Genetic Algorithm.
- Repeat process until optimization criteria achieved.

IV. Algorithm for Calculating Weighted Cosine Similarity Function

Step 1: Calculate Term Frequency for each document.

Step 2: Normalize Term Frequency for each document.

TF (Di) = TF (Di) / max TF (Di)

Step 3: Calculate Inverse Document Frequency for each term in document.

IDF (Di) = log (N/Ni) +1

N=Total No. of document

Ni= Total no. of documents in which term Ti occur

Step 4: Calculate TF * IDF

Step 5: Apply Cosine Matching Function to calculate degree of similarity between documents.

Cos (Q, D) = |Q * D| /|Q| |D|

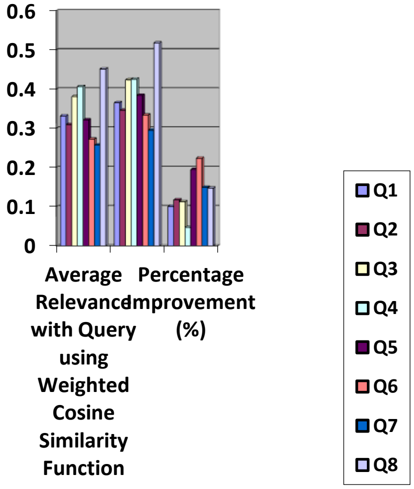
Step 6: Exit

V. RESULT

In this experiment Weighted Cosine Similarity Function is used to calculate the similarity between document and query. Genetic Algorithm is used to find new keyword. This is used to increase the efficiency of query. In order to conduct experiment we take crossover (Pc = 0.6) and mutation (Pm = 0.010).

User Query	Average Relevance with Query using Weighted Cosine Similarity Function	Average Relevance with Optimized Query using Weighted Cosine Similarity Function	Percentage Improvement (%)
Q1	0.331742	0.365271	10.10%
Q2	0.309661	0.346063	11.75%
Q3	0.381066	0.423947	11.25%
Q4	0.406518	0.425756	4.73%
Q5	0.321451	0.384196	19.5%
Q6	0.272714	0.333741	22.37%
Q7	0.257141	0.295548	14.93%
Q8	0.451531	0.518005	14.72%

Table 1.1 percentage improvement after query optimization



VI. CONCLUSION

Tremendous amount of information separated all over the digital world. In order to extract relevant information user query is optimized. If user query is not efficient then relevant information cannot retrieved. In this paper Genetic Algorithm is used with Weighted Cosine Similarity Function to find the keyword which after adding to user query increases their efficiency of retrieving relevant information. Table 1.1 show percentage improvement in retrieving relevant information after optimizing user query.

VI REFERENCES

[1] Miss Komal R. Hole et al. , "Application of Genetic Algorithm for Image Enhancement and Segmentation" , International Journal of Advanced Research in Computer Enginnering and Technolog, volume 2 , Issue 4 ,pp 1342-1346 , April 2013.

[2] Abdelmgeid A.Aly, "Applying genetic Algorithm in Query Improvement Problem",International Journal Information Technology and knowledge, vol 1,pp 309-316, 2007.

[3] Nirpal P.B. and Kale K.V , Genetic Algorithm Based Software Tesing Specifically Structured Testing for Software Reliability Enhancement" , International Journal of

computational Intelligence Techniques, ISSN 0976-0466 , Vol 3 , Issue 1 , pp 60-64 , April 2012.

4] Laith Mohammad Qasim Abualigah et al., “Applying Genetic Algorithms to Information Retrieval using Vector Space Model” International Conference of Computer Science , Engineering and Applications , Vol 5 , No.1, Feb 2015.

[5] Sergiy D.Pogorilyy et al., “Genetic Algorithm For Network Performance Optimization”, Proceedings of IAM, Vol 1, N.2, pp 121-128.

[6] Richa Garg and Saurabh Mittal , Effect of Local Search on the Performance of Genetic Algorithm ” , International Journal of Emerging Research in Management and Technology , ISSN 2278-9359 , Volume 3 , Issue 6 ,pp 41-45 , June 2014.

[7] Tarek A. El-Mihoub et al., “ Hybrid Genetic Algorithm : A Review”, Engineering Letters , 13:2 , EL_13_2_11 ,Advance online Publication : 4 August 2006.

[8] Cristina Lopez Pujalte, Felix de Moya Anegon et al. “Order Based Fitness Function for Genetic Algorithms Applied to Relevance Feedback “, Journal of the American Society for Information Science and Technology, January 2003.

[9] Vaibhav Chaudhary, Dr. Pushpa Rani Suri ,” Genetic Algorithm v/s Share Genetic Algorithm with Roulette Wheel Selection method for Registration of Multimodal Images”, International Journal of Engineering Research and Application, August 2012.

[10] Chengjun Liu, “The bayes decision rule induced similarity measures”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1086-1090, 2007.

[11] P.Simon, and S.S. Sathya, “Genetic algorithm for information retrieval”, International Conference on Intelligent Agent & Multi-Agent Systems (IAMA), ISBN: 978-1-4244-4710-7, pp. 1 – 6, 2009.

[12] Nurkhadijah Aishah Ibrahim, Ali Selamat, Mohd Hafiz Selamat, “Query optimization in relevance feedback using hybrid GA-PSO for effective web information retrieval”, IEEE Transaction DOI 10.1109, pp. 91-96, 2009.

.