IMPROVING THE EFFECTIVENESS OF INFORMATION RETRIEVAL SYSTEM USING ADAPTIVE GENETIC ALGORITHM

Wafa. Maitah[1], Mamoun. Al-Rababaa[2] and Ghasan. Kannan[3]

[1, 2] Department of Computer Science, Faculty of Information Technology, Al al-Bayt University, Jordan

[3] Department of Information System, Faculty of Information Technology, Amman Arab University, Jordan

ABSTRACT

Traditional Genetic Algorithm which is used in previous studies depends on fixed control parameters especially crossover and mutation probabilities, but in this research we tried to use adaptive genetic algorithm.

Genetic algorithm started to be applied in information retrieval system in order to optimize the query by genetic algorithm, a good query is a set of terms that express accurately the information need while being usable within collection corpus, the last part of this specification is critical for the matching process to be efficient, that is why most research efforts are actually put toward the query improvement.

We investigated the use of adaptive genetic algorithm (AGA) under vector space model, Extended Boolean model, and Language model in information retrieval (IR), the algorithm used crossover and mutation operators with variable probability, where a traditional genetic algorithm (GA) uses fixed values of those, and remain unchanged during execution. GA is developed to support adaptive adjustment of mutation and crossover probability; this allows faster attainment of better solutions. The paper has been tested using 242 Arabic abstracts collected from the proceedings of the Saudi Arabian National conference.

KEYWORDS

Information Retrieval, Adaptive Genetic Algorithm, Vector Space Model, Language Model, Extended Boolean Model.

1. INTRODUCTION

Information retrieval (IR) handles the representation, storage, organization, and access to information items [1]. In IR one of the main problems is to determine which documents are relevant and which are not to the user's needs. In practice, this problem is usually mentioned as a ranking problem, which aims to be solved according to the degree of relevance (similarity) between each document and the user query. [1] [2].

Genetic algorithm started to be applied in information retrieval system in order to optimize the query by a genetic algorithm, a good query is a set of terms that express accurately the information need while being usable within collection corpus, the last part of this specification is critical to make the matching process efficient.

Traditional Genetic Algorithm which is used in previous studies depends on fixed control parameters especially crossover and mutation probabilities, but in this study we try to use an adaptive genetic algorithm. In other words it depends on variable crossover and mutation probabilities so as to improve performance in an information retrieval.

According to the natural evolution process; the use of analogies of natural action led to the development of Genetic Algorithms (GAs), which has four main elements : Representation of an individuals as possible solutions , a fitness function which assigned a fitness score and that indicates how good an individual is, Reproduction method and a selection criteria selects highly fit individuals to reproduce the offspring by crossover and mutation techniques. [21], [22].

2. THE RESEARCH OBJECTIVES

In this research, we Attempt to enhance the performance of information retrieval by using an adaptive genetic algorithm which can improve the quality of query and obtain more developed queries that fit the searcher's needs. Investigate and evaluate different fitness functions. Reduce the search space which leads to saving time and reduction the number of iterations needed to generate the most optimized query. Finally obtain the best techniques to modify the query in an information retrieval system.

3. HOW GENETIC ALGORITHMS WORK

According to a natural phenomenon called "the survival of the fittest", only the fittest one survives and reproduces. The reproduction process occurred in the gene pool. Using crossover and mutation can generate a combinations of genes from previous ones [20]. Then a new gene pool is created. "Repeated selection and crossing over causes continuous evolution of the gene pool and the generation of individuals who will survive in a competitive environment" [23].

3.1. Components of Genetic Algorithm

- Representation
- Initialization
- Evaluation Function
- Parent Selection Techniques

3.2. Genetic Operators

1. Crossover: means produce two new off spring by choosing a random position in the string and exchanging the segments either to the right or to the left of this point with another string partitioned similarly. Simple Genetic Algorithm (SGA) uses one point crossover, but can use more than one point crossover [22 [24].

2. Mutation: is an arbitrary change in a situation with small probability. Sometimes it is used to prevent the algorithm from getting stuck. [23].

3.2. Genetic Parameters

1. Population size: Population size affects the efficiency of the algorithm incrementally . less population covers a small search space and may results in low performance, while larger population would give accurate results covering more space and prevent premature recovery to local solutions. The large population size needs more evaluation per generations but the convergence rate may slow down [28].

2. Probability of Crossover or crossover rate: a higher crossover rate introduces new strings more quickly into the population [28].

3. Probability of Mutation: a lower mutation rate helps to prevent any bit positions from getting stuck to single values, where as a high mutation rate results in essentially random search" [23][27].

4. USING GA WITH INFORMATION RETRIEVAL SYSTEM

A keyword equals a gene (a bit pattern),so a bit string is document's list of keywords represents individuals, and a collection of documents initially judged relevant by a user represents the initial population.

The genetic algorithm is executed in IR in the following steps:

- Encoding of a Chromosome
- Crossover.
- Mutation
- Determination of Population

5. LITERATURE REVIEW

1. Feras AL-Mashakbeh [32], In this study, author used different strategies of genetic algorithms, and the difference between the strategy and the other based on the difference between the GA operators(crossover and mutation), so grew the following strategies:

GA1: GA that used one-point Crossover and point mutation.
GA2: GA that used one point crossover operator and chromosomal mutation.
GA3: GA that used restricted Crossover operator and point mutation .
 GA4: GA that used restricted Crossover operator and chromosomal mutation
GA5: GA that used uniform Crossover operator and point mutation.
GA6: GA that used uniform Crossover operator and chromosomal mutation .
GA7: GA that used fusion operator and point mutation
GA8: GA that used fusion operator and chromosomal mutation.
GA9: GA that used dissociated operator and point mutation.
GA10: GA that used dissociated operator and chromosomal mutation.

In cosine similarity, Jaccard similarity, Dice similarity, and in Inner Product similarity the study compared different GA approaches by calculating the improvement of each approach over the traditional IR system.

The study also applied with different mutation strategies and different fitness function (Recall, Precision) on Boolean model, we noticed that GA with point mutation gave a higher improvement

than traditional IR system. The study also applied with different mutation strategies and different fitness function (Recall, Precision) on Fuzzy Set, we noticed that GA with point mutation gave a higher improvement than traditional IR system.

2. Poltak Sihombing,, Abdullah Embong,, Putra Sumari [33], implemented and compaired with the Jaccard's formulation and Dice's formulation. In the previous research, Jaccard and Dice's formulation is developed in a prototype called the Journal Browser. Each technique had been implemented in IRS using Genetic Algorithm (GA).

The goal of GA was to find a group of documents which best fits the searcher's needs. In this study was selected an evaluation function for the fitness of each chromosome based on Horng & Yeh's score [13]. This score is desinged to measure the relationship between the query with some documents in a database.

By the similarity percentage of documents, the user can choose the most relevant document from the database [33].

3. Suhail S. J. Owais, Pavel Kr¨omer, and V´aclav Sn´aˇse [34], investigated the use of Genetic algorithms in Information retrieval in the area of optimizing a Boolean query. Information retrieval effectiveness measures precision and recall used as a fitness function in their work. Other Genetic algorithms operators were used as single point crossover on Boolean logical operators, and mutation machinery was used to exchange one of the Boolean operators and, or, and xor with any other one. The aim is to retrieve most relevant documents with less number of no relevant documents with respect to user query in Information retrieval system using genetic programming [34].

The results of this study suggest that the final population composed of individuals having the same strength (quality) will have the same precision and recall values. The best individual result was randomly chosen as best.

4. Abdelmgeid A. Aly [35], presented an adaptive method using genetic algorithm to modify user's queries, based on relevance judgments. This algorithm was adapted for the three well-known documents collections (CISI, NLP and CACM). The method was shown to be applicable to large text collections, where more relevant documents were presented to users in   the genetic modification.

The algorithm showed the effects of applying GA to improve the effectiveness of queries in IR systems. [35].

This study was based on Vector Space Model (VSM) in which both documents and queries were represented as vectors; the weights were assigned to terms proposed by Salton and Buckle, and the system was evaluated by the precision and the recall formula.

We noticed that the result GA in the CISI documents collection gave a higher improvement than Classical IR system with 11.9%, in the NPL documents collection the GA gave a higher improvement than classic IR system with 11.5% as average values, and in The CACM documents collection GA gave a higher improvement than that with classic IR system 5.13%, as average values.

6. THE RESEARCH METHOD

This research performed in the following steps:

1. The corpus of 242 Arabic abstracts collected from the proceedings of the Saudi Arabian National conference used in this research.

2. Some text operations have been performed on those documents to determine documents' terms, the following procedure is used:

- Extraction of all the words from each document.
- Elimination of the stop-words
- Stemming the remaining words using the porter stemmer, this is the most commonly used.

3. Inverted file index is used in this study.

4. After determining the terms that described the documents, the weights were assigned using the formula proposed by Salton and Buckley [34]

$$a_{ij} = \frac{\left[0.5 + 0.5\frac{tf_{ij}}{\max_{tf}}\right] * \log\frac{N}{n_i}}{\sqrt{\left[0.5 + 0.5\frac{tf_{ij}}{\max_{tf}}\right]^2 \left[\log\frac{N}{n_i}\right]^2}}$$

Where aij is the weight assigned to the term t j in document Di, tfij is the number of times that term t j appears in document Di, n j is the number of documents indexed by the term t j and finally, N is the total number of documents in the database.

5. A traditional similarity approach used (Vector Space Model, Extended Boolean Model, and the Language Model) a strategy based on an inverted index file, has been used. Then, the following steps have been applied:

- For each model, each query is compared with all the documents. The results in a list giving the similarities of each query with all documents of the collection then ranked the list in decreasing order of similarity degree.
- Evaluate the retrieved document using average Recall and Precision formula.

Make a training data contain of the top 15 documents of the list with a corresponding query Automatically, the top 15 documents were retrieved as training data considered as initial population to Adaptive Genetic.

We gave some details of the characteristics of the AGA that give the best performance [as shown in Figure 4.3]; these characteristics leads the algorithm in its searching process in the following manner:

1. Representation of the chromosomes

AGA work with chromosomes using weights of terms representation, and have the same number of genes (components) as the query and the documents have terms with non-zero weights. The set of terms in the documents and the query are calculated, and the size of the chromosomes is equal to the number of terms of that set.

2. The population

AGA receives an initial population contains of the chromosomes corresponding to the relevant documents. The population is represented by terms of weight.

3. Genetic operators

this algorithm uses the one-point crossover machinary and point mutation.

4. Control parameters

The values of the control parameters crossover probability (pc) and mutation probability (pm) are variable, the fitness function is the similarity.

5. Fitness

The following functions are used to determine the fitness values [32] [3]:

Cosine

$$Sim(D,Q) = \frac{\sum_i (a_i * b_i)}{\sqrt{\sum_i a_i^2 * \sum_i b_i^2}}$$

Horng & Yeh

$$F = \frac{1}{|D|} \sum_{i=1}^{|D|} \left(r(di) \sum_{j=1}^{|D|} \frac{1}{j} \right)$$

Evaluate the retrieved document using average Recall and Precision formula
Compare effectiveness between different AGA approaches for each model
Compare effectiveness between the best AGA with the best GA approach.

7. THE PROPOSED ALGORITHM

We used an AGA that has been optimized and adapted for relevance feedback, next describe the characteristics of this AGA, chosen for having the best performance by using crossover and mutation operators with variable probabilities, where as the traditional genetic algorithm (GA) uses fixed values of those, and remains unchanged during execution. Developed GA supports adaptive adjustment of mutation and crossover probabilities; this allows faster attainment of

better solutions, and then we describe two different fitness functions, both based on the order of retrieval, which we used to guide the algorithm in the search process.

8. THE PROCESS OF THE AGA

Representation of the chromosomes: the chromosomes represented using binary representation, these chromosomes have the same number of genes (components) as there are terms with nonzero weights in the query and in the documents of the feedback. The set of different terms contained in those documents and in the query are calculated firstly , and the size of the chromosomes is equal to the number of terms in that set.

Population: Our AGA receives an initial population consisting of the chromosomes corresponding to the top 15 documents retrieved from traditional IR with respect to that query.

Selection: the selection process select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected).

Genetic operators: used one-point crossover as the crossover operator. It is defined as follows:

Given two parent chromosomes C1 = (a1. . . am) and C2= (b1 . . . bm), one generates two offspring chromosomes H1= (a1. . . ai, bi + 1, . . .,bm) and H2 = (b1, . . . , bi, ai + 1,. . . , am), where i is a random number in the interval [1, m_ 1] and m is the length of the chromosome.
Mutation: a random process is implemented in our algorithm. , In our case a real random number is generated in a given interval [0, 1], and that number is taken as the new value for the gene that has to mutate.

Control parameters: Crossover probability Pc and mutation probability Pm play an important role in GA. [26][27].

Typical initial value of P_c is in the range of 0.5»1.0. The mutation probability P_m is varied according to the generations.

The initial Pm is larger for the global search, and in some generations it is smaller for the local search.

Finally, it is larger again for avoidance of local optimum. Typical initial value of P_m is in the range 0.005»0.05.

We put forward adaptive varied values of P_c and Pm as follows [26][27]:

$$p_c = \begin{cases} p_{c1} - \dfrac{(p_{c1} - p_{c2}) * (f' - f_{avg})}{f_{max} - f_{avg}}, f' \geq f_{avg}, \\ p_{c1}, f' \prec f_{avg} \end{cases}$$

$$p_m = \begin{cases} p_{m1} - \dfrac{(p_{m1} - p_{m2}) * (f - f_{avg})}{f_{max} - f_{avg}}, f' \geq f_{avg}, \\ p_{m1}, f' \prec f_{avg} \end{cases}$$

Where, f_{max} is the maximum fitness function of current generation, f_{avg} is the average fitness function of current generation, f' is larger fitness function of the two crossover chromosomes selected, f is the fitness function of mutation chromosome selected, p_{c1}; p_{c2} is crossover probability, and p_{m1}, p_{m2} is mutation probability. The study experimental parameters include: p_{c1} = 0:9; p_{c2} = 0:6, p_{m1} = 0:1; p_{m2} = 0:001.
End Condition.

GA needs termination Condition to end the generation process. If we have no sufficient improvement in two or more consecutive generations, the number of iteration used in this study is 75 iterations.

The Fitness Functions: We ran the AGA described above with different order based fitness functions.

Fitness 1: This fitness function, due to Horng and Yeh [13], is very innovative. As well as taking into account the number of relevant and of irrelevant documents, it also takes account of the order of their appearance, because it is not the same that the relevant documents appear at the beginning or at the end of the list of retrieved documents.

Once we calculates the similarity of the query vector with all the documents, and sorts the documents into decreasing order of similarity. Finally, we calculates the fitness value of the chromosome with the following formula:

$$F = \frac{1}{|D|} \sum_{i=1}^{|D|} \left(r(di) \sum_{j=1}^{|D|} \frac{1}{j} \right)$$

Where $|D|$ is the total number of documents retrieved, and r(di) is the function that returns the relevance of document d, giving a 1 if the document is relevant and a 0 otherwise. We shall refer to this fitness function as fitness 1.

Fitness 2: cosine similarity

$$F = \frac{\sum_{k=1}^{t} (d_{ik} \bullet q_k)}{\sqrt{\sum_{k=1}^{t} d_{ik}^2 \bullet \sum_{k=1}^{t} q_k^2}}$$

Where dik is the weight of term i in document k and qk is the weight of term i in the query.

9. RESULTS AND DISCUSSION

The traditional information retrieval systems were built and implemented to handle the Arabic collection using C# NET. The following three IR systems were built and implemented:

- System that used Vector Space Model with Cosine similarity
- System that used Extended Boolean Model
- System that used Language model .

Different AGA strategies were used in this research. Those strategies are as the following:

- AGA1: AGA that use cosine similarity as fitness
- AGA2: AGA that use Horng & Yeh formula as fitness.

9.1 Applying AGA on Vector Space Model

Figure 1. shows the comparison between VSM with Cosine as fitness (AGA1) and VSM with Horng as fitness (AGA2), from figure we notice that the VSM with Horng as fitness (AGA2) represent the best strategy over VSM with Cosine as fitness (AGA1).
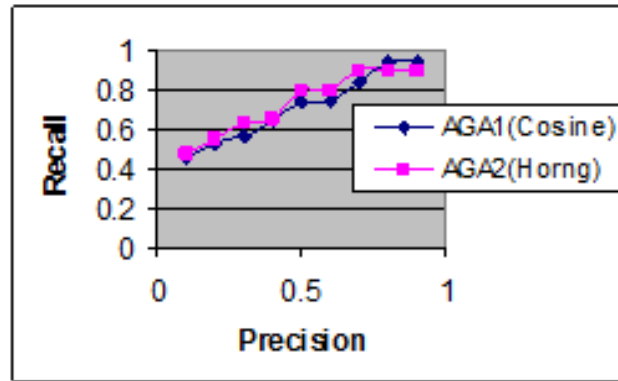

Figure1. Comparison between deferent AGA in Vector Space Model

9.2. Applying AGA on Extended Boolean Model

Figure 2. shows the comparison between EBM with Cosine as fitness (AGA1) and EBM with Horng as fitness (AGA2), from this figure we notice that the EBM with Cosine as fitness (AGA1) represent the best strategy over EBM with Horng as fitness (AGA2).
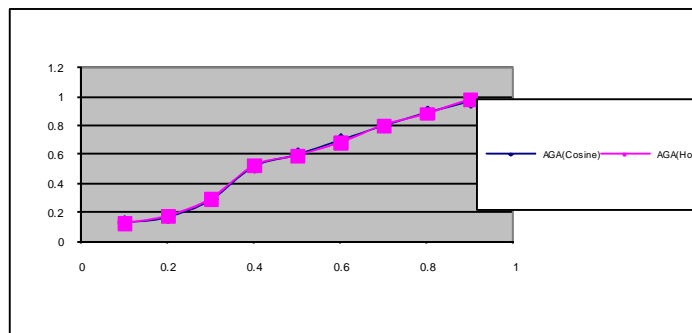

Figure2. Comparison between deferent AGA in Extended Boolean Model

9.3. Applying AGA on Language Model

Figure 3. shows the comparison between LM with Cosine as fitness (AGA1) and LM with Horng as fitness (AGA2), from this figure we notice that the LM with Horng as fitness (AGA2) represent the best strategy over LM with Cosine as fitness (AGA1).
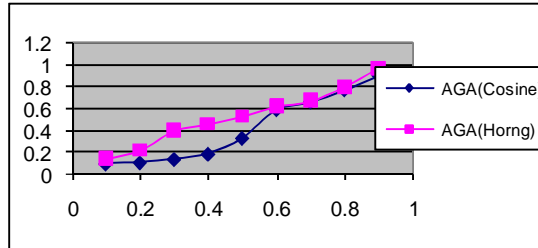

Figure 3.Comparison between deferent AGA in Language Model

9.4. AGA Using Cosine Similarity

Figure 4. shows the comparison between VSM(AGA1) ,EBM(AGA2), And LM(AGA3) with Cosine as fitness, from this figure we notice that the VSM(AGA1) represent the best strategy over EBM(AGA2), And LM(AGA3).
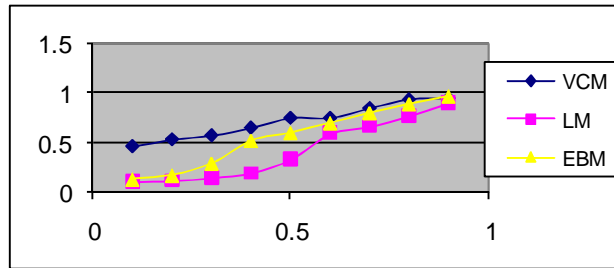

Figure4. Average Recall and Precision values for 59 Queries by applying AGAs with cosine similarity fitness

9.5. AGA using Horng & Yeh formula

Figure 5. shows the comparison between VSM(AGA1) ,EBM(AGA2), And LM(AGA3) with Horng & Yeh formula as fitness, from this figure we notice that the VSM(AGA1) represent the best strategy over EBM(AGA2), And LM(AGA3)
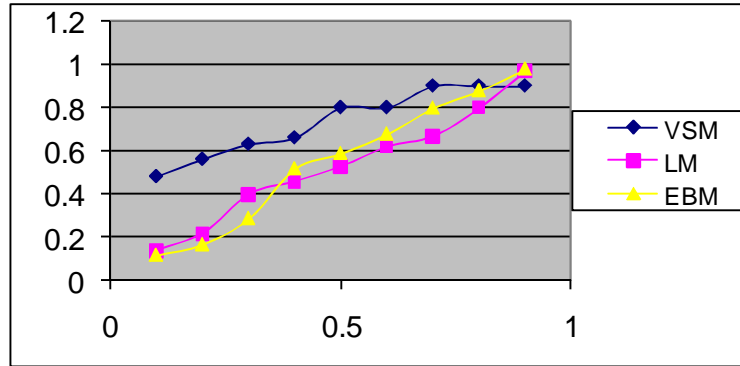
Figure5. Average Recall and Precision values for 59 Queries by applying AGAs with Horng & Yeh formula

9.6. Comparison between best AGAs strategy with traditional GAs

The results for the AGAs are shown in table 1, table 2, figure 6, and figure 7 using the average Recall and Precision relationship. From those tables and corresponding figures we notice that VSM with Horng as fitness (AGA1) compare with VSM with Cosine as fitness under traditional Genetic Algorithm (GA1) gives the highest improvement over GA1 with 2.924%., and EBM with Cosine as fitness (AGA2) compare with Boolean Model (BM) with precision as fitness under traditional Genetic Algorithm (GA1) gives the highest improvement over GA1 with 1.434%

Table1. Comparison between VSM with Horng as fitness and VSM with Cosine as fitness under traditional Genetic Algorithm

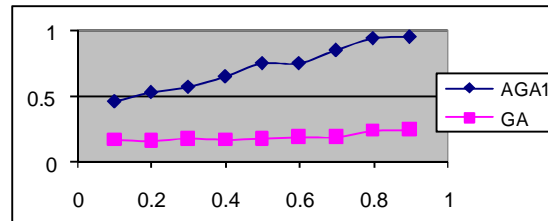| Recall	average Recall and Precision		AGA Improvement %
	GA	AGA2	
0.1	0.17	0.13	-0.307
0.2	0.17	0.17	0.00
0.3	0.18	0.29	0.611
0.4	0.19	0.52	1.736
0.5	0.2	0.6	2.00
0.6	0.23	0.7	2.043
0.7	0.24	0.8	2.33
0.8	0.26	0.89	2.423
0.9	0.27	0.97	2.592
Average	0.212	0.563	1.434

Figure6. Average Recall and Precision values for VCM Horng (AGA1 and GA)

Table2. Comparison between EBM with Cosine as fitness (AGA2) and Boolean Model (BM) with precision as fitness under traditional Genetic Algorithm
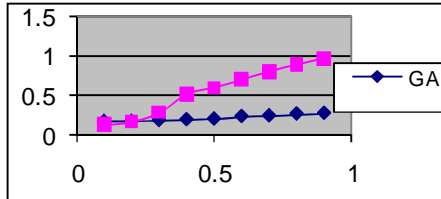

Figure7. Average Recall and Precision values for EBM with Cosine similarity (AGA2 and GA)

Recall	average Recall and Precision		AGA Improvement %
	GA	AGA1	
0.1	0.16	0.48	2.00
0.2	0.17	0.56	2.29
0.3	0.18	0.63	2.50
0.4	0.18	0.66	2.66
0.5	0.19	0.8	3.21
0.6	0.2	0.9	3.50
0.7	0.2	0.8	2.00
0.8	0.24	0.9	3.30
0.9	0.25	0.9	2.60
Average	0.196	0.736	2.924

10 CONCLUSION

The research apply Adaptive Genetic Algorithm(AGA) with different fitness functions (Cosine and Horng) and variable operators rate(crossover and mutation) on vector space model, extended model, and language model .

In vector space model, the research compares different adaptive genetic algorithm strategies by calculating evaluation using average recall formula. We noticed that the vector space model with Horng as fitness represent the best strategy over vector space model with Cosine as fitness.

In Extended Boolean Model, the research compares different adaptive genetic algorithm strategies by calculating evaluation using average recall formula. We noticed that the Extended Boolean Model with Cosine as fitness represent the best strategy over Extended Boolean Model with Horng as fitness.

In Language Model, the research compares different adaptive genetic algorithm strategies by calculating evaluation using average recall formula. We noticed that the Language Model with Horng as fitness represent the best strategy over Language model with Cosine as fitness.

The research compare between the best adaptive genetic algorithm strategies by calculating the improvement over the traditional genetic algorithm. We noticed that the vector space model with Horng as fitness compare with vector space model with Cosine as fitness under traditional Genetic Algorithm gives the highest improvement over vector space model with Cosine as fitness under traditional Genetic with 55.1%and EBM with Cosine as fitness EBM (AGA1) compare with Boolean Model with precision as fitness under traditional Genetic Algorithm BM (GA) gives the highest improvement over BM (GA) with 42.1%.

REFERENCES

[1] Christopher D. Manning ,Prabhakar Raghavan   and Hinrich Schütze , "An Introduction To Information Retrieval", Book , Cambridge University Press , February 16, 2008.

[2] Cristina Lopez-Pujalte, Vicente P. Guerrero-Bote, Felix de Moya-Anegon, "Genetic algorithms in relevance feedback: a second test and new contributions", Proceedings in Information Processing and Management 39, 2003.

[3] Andrew Trotman, "An Artificial Intelligence Approach to Information Retrieval", Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 603 – 603, 2004.

[4] Nasredine Semmar, Faïza Elkateb-Gara and Christian Fluhr, "Using a Stemmer in a Natural Language Processing system to treat Arabic for Cross-language Information Retrieval" , Proceedings of the 5th Conference On Language Engineering, pp. 1-10, 2005.

[5] Egidio Terra, Robert Warren, "Poison Pills: Harmful Relevant Documents in Feedback", Proceedings of the 14th ACM international conference on Information and knowledge, pp. 319 – 320, 2005.

[6] Nasredine Semmar, Faïza Elkateb-Gara and Christian Fluhr, "Using a Stemmer in a Natural Language Processing system to treat Arabic for Cross-language Information Retrieval" , Proceedings of the 5th Conference On Language Engineering, pp. 1-10, 2005.

[7] Graham Bennett Falk Scholer Alexandra Uitdenbogerd," A Comparative Study of Probabilistic and Language Models for Information Retrieval", Proceedings of Nineteenth Australasian Database Conference (ADC), Vol. 75, pp 1- 10, 2008

[8] Kanaan G, hanandeh E, "Evaluation of Different Information Retrieval models and Different indexing methods on Arabic Documents", PhD Thesis, ARAB Academy, 2008

[9] Ronan Cummins and Colm O'Riordan, "Determining general term weighting schemes for the Vector Space Model of Information Retrieval using Genetic Programming", Proceedings of the 15th Artificial Intelligence and Cognitive Science Conference (AICS), 2004.

[10] Ramesh Nallapati, Bruce Croft and James Allan, "Relevant Query Feedback in Statistical Language Modeling",Proceedings of the 12th international conference on Information and knowledge management, pp. 560 – 563,2003.

[11] Bangorn Klabbankoh, "Applied Genetic Algorithm in Information Retrieval ",Proceedings of the International journal of the computer, the internet and management, Vol. 7, No. 3, pp. 60-66, 1999.

[12] M. Boughanem, Chrisment, Tamine "using genetic algorithms for multimodal relevance optimisation in information retrieval", proceeding of the Journal of American Science on Information and Technology

[13] Horng & C.C Yeh, "Applying genetic algorithms to query optimisation in document retrieval", Proceedings of the Information Processing and Management, Vol 36, pp 737-759, 2000

[14] Desjardins, G. et Godin, R. "Combining Relevance Feedback and Genetic Algorithm in Internet Information Filtering Engine", Proceedings of the RIAO Content-Based Multimedia Information Access, http://www citeseerx.ist.psu.edu, 2000.

[15] Jen-Yuan Yeh, Jung-Yi Lin, Hao-Ren Ke, Wei-Pang Yang,"Learning to Rank for Information Retrieval Using Genetic Programming", Proceedings in ACM SIGIR Workshop on Learning to Rank for Information Retrieval, 2007.

[16] Ronan Cummins, Colm O'Riordan, "Using Genetic Programming for Information Retrieval: Local and Global Query Expansion", Proceedings of the 9th annual conference on Genetic and evolutionary computation, pp.2255--2255, 2007.

[17] Bangorn Klabbankoh, "Applied Genetic Algorithm in Information Retrieval ", International journal of the computer, the internet and management, Vol. 7, No. 3, pp. 60-66, 1999.

[18] Paul Ogilvie, Jamie Callan," Language Models and Structured Document Retrieval", Proceedings of the Initiative for the Evaluation of XML Retrieval Workshop, pp. 211-224, 2003.

[19] Herrera-Viedma, "An Information Retrieval Model With Ordinal Lingustic Weigted Queries Based On Two Weighting Elements" International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.2, No. 2, pp 1-11, 2001.

[20] Jianfeng Gao Jian-Yun Nie, Guangyuan Wu, Guihong Cao, "Dependence Language Model for Information Retrieval", Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp.265-278, 2004.

[21] Djoerd Hiemstra, "Using language models for information retrieval", book, 2000

[22] Xiaohua Zhou, "Semantics-based Language Models for Information Retrieval and Text Mining",phD thesis, Drexel University,2008

[23] Mike Sewell, Jagath Samarabandu, Ranga Rodrigo, and Kenneth McIsaac, "The Rank-scaled Mutation Rate for Genetic Algorithms", proceeding of International Journal of Information Technology, Vol .3 No.1, pp 32-36.2006.

[24] Shengxiang Yang, "Adaptive Crossover in Genetic Algorithms Using Statistics Mechanism", Proceedings of the 8th international conference on Artificial life PP.182-185, 2002.

[25] Nga Lam Law and K.Y. Szeto, "Adaptive Genetic Algorithm with Mutation and Crossover Matrices", Proceedings of the 20th international joint conference on Artificial intelligence, pp. 115- 120,2007.

[26] Wang Lei, Shen Tingzhi, "An Improved Adaptive Genetic Algorithm and its application to image segmentation", Proceeding of 5th International Conference on Artificial Neural Network and Genetic Algorithms, PP.112–119,2004.

[27] Mauro Annunziato, Stefano Pizzuti," Adaptive Parameterization of Evolutionary Algorithms Driven by Reproduction and Competition", Proceedings of Genetic and evolutionary computation conference, pp.1597- 1598, 2005.

[28] Eric Pellerin, Luc Pigeon, Sylvain Delisle, Self-"Adaptive Parameters in Genetic Algorithms", proceeding of Data Mining and Knowledge Discovery Conference,Vol.53, pp. 61-100, 2004

[29] Sima (Etaner) Uyar, Gulsen (Cebiroglu) Eryigit, Sanem Sariel," An Adaptive Mutation Scheme in Genetic Algorithms for Fastening the Convergence to the Optimum", proceeding of 3rd Asian Pacific International Symposium on Information Technologies APIS, pp.1257-1264 , 2005.

[30] Imtiaz Korejo, Shengxiang Yang, and ChangheLi, "A Comparative Study of Adaptive Mutation Operators for Genetic Algorithms", Metaheuristics International Conference, MIC 2009

[31] Sonja Novkovic, and Davor Sverko, "A Genetic Algorithm With Self-Generated Random Parameters", Journal of Computing and Information Technology (CIT) ,Vol.11, pp.271 -283 ,2003.

[32] Feras AL-Mashakbeh ,"Evaluate the Effectiveness of Genetic Algorithm in Information Retrieval Based on Arabic Documents ", Unpublished ph.D Thesis , Faculty of Information System and Technology , The Arab Academy for Banking and Financial Science , Amman , Jordan, 2008 .

[33] Poltak Sihombing,, Abdullah Embong,, Putra Sumari," Comparison of Document Similarity in Information Retrieval System by Different Formulation ", proceding of the 2nd IMT-GT Regional Conference on mathematics , statistics and applications University sains Malaysia, penang , 2006.

[34] Suhail S. J. Owais, Pavel Kr¨omer, and V´aclav Sn´aˇse," Query optimization by Genetic Algorithms", Proceedings of the Dateso Annual International Workshop on Databases, pp. 125–137, 2005.

[35] Abdelmgeid A. Aly, "Applying Genetic Algorithm In Query Improvement Problem" International Journal "Information Technologies and Knowledge", Vol.1, pp.309-316, 2007.

Authors

Wafa Maitah

Obtained my B.Sc of software Engineering from Hashemite University in 2005, Master of computer science from al albayt University in 2010. My research interest's areas are information retrieval, expert system, web-based applications.

Mamoun S. Al Rababaa received M.Sc degree in 1995. A Ph.D. was received in 1999 in the area of computing engineering from Ukraine.

In 2000, he joined Irbid National University as an assistant professor in Computer Science Department, faculty of IT. In 2003 he moved to Al al-Bayt University as an assistant professor in computer science department and a chairman of computer science department, faculty of IT . In 2004 a chairman of computer science and computer information systems department, faculty of IT. In 2005-2007 a chairman of information systems department, faculty of IT. In 2012, he joined Ajloun National university as associate prof,. His research interests include areas like information security (data security, computer security) , network security, internet security, viruses and artificial intelligence, expert systems and genetic algorithms. He has supervised numerous master's degrees research thesis and projects of diverse areas. Mamoun S. Al Rababaa has published many research papers in a multitude of international journals and conferences.

Ghassan G. Kanaan, was Born in Jordan, 1962. He got a Ph.D. degree in computer science field from Illinois Institute of Technology, Chicago, IL, U.S.A. 1997.

Ghassan G. Kanaan is Prof. of Computer Science, Vice dean in IT College, instructor in Computer Information Systems, in Amman Arab University. Amman, Jordan. With an extensive experience and management skills and works for a graduate research. He worked in a variety of positions in IT colleges. Ghassan had sixteen years teaching experience. He used to serve in several universities. He has had many consultancy assignments in the areas of education in Jordan. His teaching and research area are information retrieval systems and Arabic Natural Language Processing. He used to be an advisor for more than 20 PhD students.