Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01247-9

ORIGINAL RESEARCH

A hybrid evolutionary algorithm based automatic query expansion for enhancing document retrieval system

Dilip Kumar Sharma[1,2] · Rajendra Pamula[1] · D. S. Chauhan[2]

Received: 13 August 2018 / Accepted: 9 February 2019

Abstract
Nowadays, searching the relevant documents from a large dataset becomes a big challenge. Automatic query expansion is one of the techniques, which addresses this problem by refining the query. A new query expansion approach using cuckoo search and accelerated particle swarm optimization technique is proposed in this paper. The proposed approach mainly focused to find the most relevant expanded query rather than suitable expansion terms. In this paper, Fuzzy logic is also employed, which improves the performance of accelerated particle swarm optimization by controlling various parameters. We have compared the proposed approach with other existing and recently developed automatic query expansion approaches on various evaluating parameters such as average recall, average precision, Mean-Average Precision, F-measure and precision-recall graph. We have evaluated the performance of all approaches on *three* datasets *CISI, CACM* and *TREC-3*. The results obtained for all three datasets depict that the proposed approach gets better results in comparison to other automatic query expansion approaches.

Keywords Document retrieval · Accelerated particle swarm optimization · Cuckoo search · Fuzzy logic · Precision · Recall and F-measure

1 Introduction

Query expansion technique is one of the efficient approaches to improve the performance and reliability of document retrieval system. It gives more suitable query for users in comparison to initial or original queries by adding one or more expansion keywords (Carpineto et al. 2012). Nowadays, query expansion is being used in various applications like question answering system (Park and Croft 2015), multimedia information retrieval (Li et al. 2016) and information filtering (Leturia et al. 2013); and also applied to different domains. Sports (Kabary and Schuldt 2014), Healthcare (Maurer et al. 2015), Medical (Oh and Jung 2015) and e-commerce (Saraiva et al. 2016) are some of them. Query expansion is of three types: manual query expansion, interactive query expansion and automatic query expansion. In manual query expansion, user adds terms in query, which gives good results but takes a lot of time. In interactive query expansion approach, system identifies a pool of terms then user selects most suitable term for query expansion. This process also not time efficient. *Automatic Query Expansion (AQE)* approach finds the most suitable terms automatically.

Researchers have proposed many approaches in literature related to *AQE*. Conventional document retrieval extracts the relevant documents for any query. But in *AQE*, first top relevant documents are extracted against the initial user query, then all unique terms are selected from these documents and finally suitable terms are determined to expand the query. In the last few years, various automatic query expansion approaches have been implemented to improve document retrieval system. These approaches are based on different methods and soft computing techniques to find the best terms for query expansion. However, these approaches are not much efficient to improve the document retrieval performance significantly and return irrelevant information too. The main reason is that all of these approaches were

✉ Dilip Kumar Sharma
dilip.sharma@gla.ac.in

Rajendra Pamula
rajendrapamula@gmail.com

D. S. Chauhan
pdschauhan@gmail.com

[1] IIT(ISM), Dhanbad, India

[2] GLA University, Mathura, India

Published online: 25 February 2019

following traditional way of query modeling. To overcome this problem, a new way of query modeling is proposed for automatic query expansion. The proposed approach mainly focused to generate a suitable query after expansion instead of finding expansion terms. A candidate query pool (a set of possible queries) is constructed in this approach to select the best expanded query. This candidate query pool contains all the possible combinations of unique terms those are selected from top retrieved documents.

It is very difficult to find best query from this candidate query pool using any conventional method as it contains a large number of potential expanded queries. This problem becomes more complex for large datasets. Therefore, hybridization of soft computing techniques is used to solve this problem. The used techniques are cuckoo search, accelerated particle swarm optimization (PSO) and fuzzy logic. The performance of proposed approach is tested on three benchmark datasets CISI, CACM and TREC-3. The results are compared with *original query*, query expansion approach proposed by Singh et al. (2017a) and Ramalingam et al's. *AQE* approach (Ramalingam and Dhandapani 2014). Three benchmark datasets *CACM, CISI* and TREC-3 are used for the experiments.

The organization of this paper is as follows: In Sect. 2, related research already done to automatic query expansion is described. Section 3 discusses preliminaries and theoretical foundation of query expansion. The proposed approach is described in Sect. 4. The experimental results and analysis are presented in Sect. 5. Finally, conclusion is drawn in Sect. 6.

2 Related work on AQE

AQE has made a big impact to make effective query and to retrieve relevant documents from large datasets. It also helps users to express their needs precisely. In last few decades, researchers analyzed various aspects of *AQE* and have done research in several domains. This section explores related work in the field of *AQE*. Initially work in *AQE* was done at corpus level leading to local and global query expansion.

The first work was reported by Van Rijsbergen in 1979. The proposed work was based on relevance feedback. Yang and Korfaghe (1994) used real coded genetic algorithm (GA) with random mutation and two-point crossover operators for improving the performance of query expansion. Sanchez et al. proposed GA based query expansion approach using user relevance feedback. GA was used to determine weights of all possible expanded terms for Boolean queries (Sanchez et al. 1995). They tested their approach on patent dataset consisting 479 documents. Robertson and Willet (Robertson and Willet 1996) used evolutionary algorithm to identify the upper bound of relevance feedback for automatic query expansion technique in vector space model based document retrieval systems. They compared their results with Robertson et al.'s retrospective relevance weighting technique (Robertson et al. 1976). The results were satisfactory.

In recent years, pseudo relevance feedback (PRF) based *AQE* is used widely and improved query expansion performance and retrieval processes. *PRF* is a type of local query expansion technique. However, there are a lot of limitations in *PRF* based *QE* in term of accuracy and computational complexity. To overcome these limitations, some other techniques were used with *PRF* i.e., semantic filtering such as WordNet etc. (Gupta and Saini 2017). However, it is also reported in literature that WordNet alone does not improve query expansion to large extent. Therefore, different variants were also introduced in recent years (Gupta and Saini 2017). The use of concept and context of queries and documents is another way to enhance *PRF* based query expansion. Later on, some researchers also used soft computing technique to improve the performance of query expansion.

A new query expansion technique based on co-occurrence was proposed by Xu and Croft (1996) for improving document retrieval and tested on *CACM, CISI* and *TREC-3* datasets. This approach successfully enhanced the performance of the system. Two different query expansion approaches using local collocation and global collocation were proposed in (Vechtomova et al. 2003). These approaches were based on long span collocates. A new semantic similarity based query expansion approach using clusters was proposed to overcome the limitation of ambiguous and short queries (Barathi and Valli 2013). This approach constructed various clusters of documents those are retrieved by the original query, and each cluster is ranked according to the content similarity with the query. At last, this approach was suggesting terms from these ranked clusters to disambiguate the query.

Gong et al. (2006) developed a new approach for expanding the queries which was based on WordNet lexical chains. They used synonym and hypernym/hyponymy relations in WordNet. They used lexical chains and relations as expansion rules. This approach improved query performance dramatically. Bendersky et al. (2012) developed a new method for term weighting-reweighting to enhance the performance of document retrieval. They used Genetic Algorithm (GA) to give the weights to user's query vector. The proposed approach was based on relevance feedback given by the user. Cooper et al. (1998) proposed a graphical user interface for users, which had graphical relations among different items. A novel term weighting based query expansion approach was proposed by Horng et al. (2000). They used GA to determine the weights for query terms. Further, they used these weights to find the closest query vector to the optimal one. Chen et al. (2001) implemented a new automatic query expansion technique using association rules. They computed the similarity among terms and constructed a tree of these terms. Kim et al. (2001) also proposed a novel approach for query expansion which was based on term co-occurrence similarity. A new query expansion method was proposed by Billerbeck et al. (2003). The method was based on associated queries. Chang et al. (2007) proposed the query expansion approach using fuzzy rules. The results were satisfactorily.

Ben and Ounis (2003) framed few association rules for mining query expansion terms and proposed a filtering technique to remove duplicate rules. Gao et al. (2010) proposed query expansion approach for web documents. Lin et al. (2008) proposed a new mining technique to find out suitable terms for query expansion. Chang and Chen (2006) proposed a novel query expansion approach using weighting and re-weighting methods to enhance the performance of document retrieval system. Grootjen et al. (2006) developed a hybrid query expansion approach which projected the results coming from initial query to global information. Lin et al. (2006) proposed a novel automatic query expansion approach to find out most suitable query terms. The proposed was based on user relevance feedback. Chang et al. (2007) framed fuzzy rules for user relevance based query expansion approach for document retrieval.

Nowacka et al. (2008) implemented an AQE approach based on fuzzy logic to enhance the performance of document retrieval process. Fattahi et al. (2008) developed a new query expansion using domain specific topical and non-topical terms. Piotr Wasilewski (2011) proposed a new query expansion approach using semantic modeling of query. Carlos and Maguitman (2009) proposed an approach to learn terms which actually helped to bridge the terminology gap existing between initial query and documents those are relevant. Tayal et al. (2012) used fuzzy logic to give weights to each query term using fuzzy logic. Latiri et al. (2012) developed a query expansion approach based on association rule mining.

Rivas et al. (2014) applied developed query expansion technique in biomedical document retrieval system. They combined text preprocessing with query expansion approach to improve the performance of document retrieval. They used one of the part of MEDLINE dataset, called Cystic Fibrosis for all the experiments. Wu et al. (2018) presented three noise control approaches in query expansion. These approaches were based on Indri search engine. They tested the performance on two benchmark datasets: OHSUMED and TREC. They found satisfactory results. Gupta et al. (2018) applied cuckoo search for filtering in ultrasound images. The developed cuckoo search filter was based on non-local means filter and 2D finite impulse response filter. Eniready et al. (2015) also used cuckoo search and particle swarm optimization in image retrieval. They used Haar wavelet for compressing the images and then extracted features from them. Khennak and Drias (2017) used accelerated particle swarm for query expansion for MEDLINE dataset. Gupta et al. (2015) proposed a new ranking function for enhancing document retrieval using fuzzy logic. Singh and Sharan (2018) proposed a combined approach for selecting most suitable query expansion terms. They found much superior results. Lee and Bang (2018) presented a new image search method to extract the features of an image. They used a combined invariant features and color description to retrieve specific images using query-by-example. Saeedeh et al. (2012) proposed a new information retrieval system, which took the inputs from user provided relevance feedback.

Singh and Sharan (2016) combined various term selection based query expansion approaches to improve document retrieval performance. They also used Word2vec for selecting query expansion terms semantically. They obtained satisfactorily results. Singh and Sharan (2015) proposed PRF and corpus-based term co-occurrence approach to find suitable terms for query expansion. They tested their approach on two datasets: FIRE and TREC-3. Singh et al. (2017a) developed a novel query expansion approach using fuzzy logic. Authors obtained better recall and precision for the proposed *AQE* approach. Singh and Sharan (2017b) implemented an automatic query expansion model for retrieving relevant documents using fuzzy logic. They used multiple terms selection methods to determine suitable expansion terms. They combined the weights of each term obtained from multiple term selection methods using fuzzy rules and determined a final weight of the same. Bhatnagar and Pareek (2015) proposed a new query expansion approach based on genetic algorithm to find the suitable terms for query expansion. They have shown improvement in results. Khennak et al. (2016) presented a bat algorithm based automatic query expansion approach for document retrieval. They used bat algorithm to select appropriate terms for query expansion. Qian et al. (2017) proposed a soft computing technique based approach for the tipping paper permeability measurement in tobacco factory. Authors focused on the structure and parameters optimization of belief rule base (BRB) by taking the referential values of the antecedent attributes and the utilities of the consequents into account to improve the input–output modeling ability of BRB. Zhang et al. (2016) used PSO for solving complex functions. They used ontology method to improve the performance of PSO. The proposed algorithm included semantic roles and concepts to update crucial parameters based on the cooperation framework. Ibrahim et al. (2018) presented a hybrid optimization method for the FS problem. They combined the slap swarm algorithm (SSA) with the particle swarm optimization.

The novelty of the proposed work is in handling the problem of AQE for document retrieval. In this paper, this problem is considered as an optimization problem and cuckoo search with accelerated swarm intelligence algorithm using fuzzy logic is being used in this paper to solve this problem.

3 Preliminaries and theoretical foundation of query expansion

This section describes the issues with query expansion approaches and the necessity of using soft computing techniques in *AQE* approaches. This section also discusses used soft computing techniques.

3.1 The complex issue of query expansion

As mentioned above the automatic query expansion approach always helps to improve the performance of document retrieval system by suggesting new suitable terms. There are various automatic query expansion approaches have been implemented and developed in literature. These approaches are based on various methods such as term weighting and reweighting method, term selection method, pseudo relevant feedback method and etc. Pseudo relevant feedback tells about how to choose document from the list of ranked documents. This method says that the top retrieved documents are more relevant to query in comparison to other documents (Carpineto 2012). Figure 1 presents a flow chart for a standard document retrieval process. This process starts by giving a query to retrieval system. Then this initial query extracts some documents. An inverted index is created and processed for each term of the original query. The main issue is identification of suitable terms for query expansion.

3.2 The necessity of soft computing techniques for query expansion

As finding the suitable terms for AQE is a combinatorial optimization problem, many algorithms have been developed to address this complex problem in literature. Soft computing techniques strengthen AQE approaches to determine weights of terms (Gupta and Saini 2017). The strength of soft computing techniques is in their capability to deal with such type of problems by having little knowledge of the search space.

4 Proposed AQE approach

In this section, a novel hybrid automatic query expansion approach is described to select the suitable terms for expanding query. The architecture of proposed approach is demonstrated by Fig. 2. This figure shows that after obtaining the ranked list of documents, a term pool is formed and the proposed approach is applied to determine the best candidate terms. The overall procedure follows following steps:

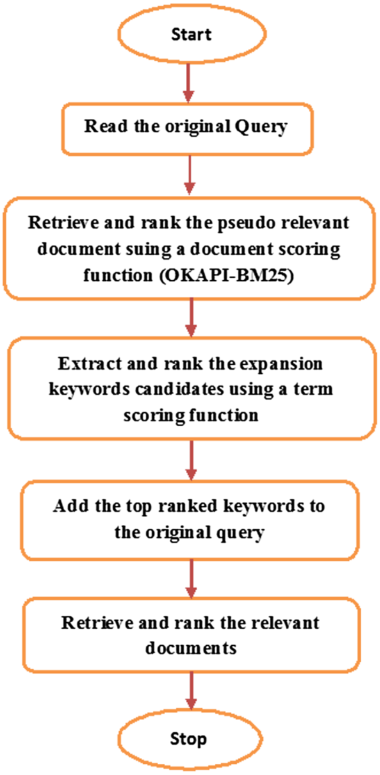

Start
Read the original Query
Retrieve and rank the pseudo relevant document suing a document scoring function (OKAPI-BM25)
Extract and rank the expansion keywords candidates using a term scoring function
Add the top ranked keywords to the original query
Retrieve and rank the relevant documents
Stop

Fig. 1 Standard query expansion based retrieval system

1. Retrieve pseudo-relevant documents using Okapi-BM25 ranking function.
2. Prepare candidate term pool which consists of all unique terms from pseudo-relevant documents.
3. Apply Cuckoo search to determine best suitable terms for query expansion.
4. Enhance Cuckoo search using new fuzzy logic based accelerated particle swarm optimization.
5. Add top selected terms in original query

The modeling of proposed approach, fitness function, encoding of the solution and the parameter setting are discussed in the following subsections:

4.1 Encoding of Solution

The first step of the proposed approach is to represent a suitable candidate solution by using cuckoo search algorithm. The entire candidate solution is represented by generating very first population of n host nests (eggs) x_i ($i = 1,2,3\ldots n$). In proposed approach, each solution of the population has two parts: First part is a set of all terms of the original query. Second part contains all other terms of candidate term pool. The main problem

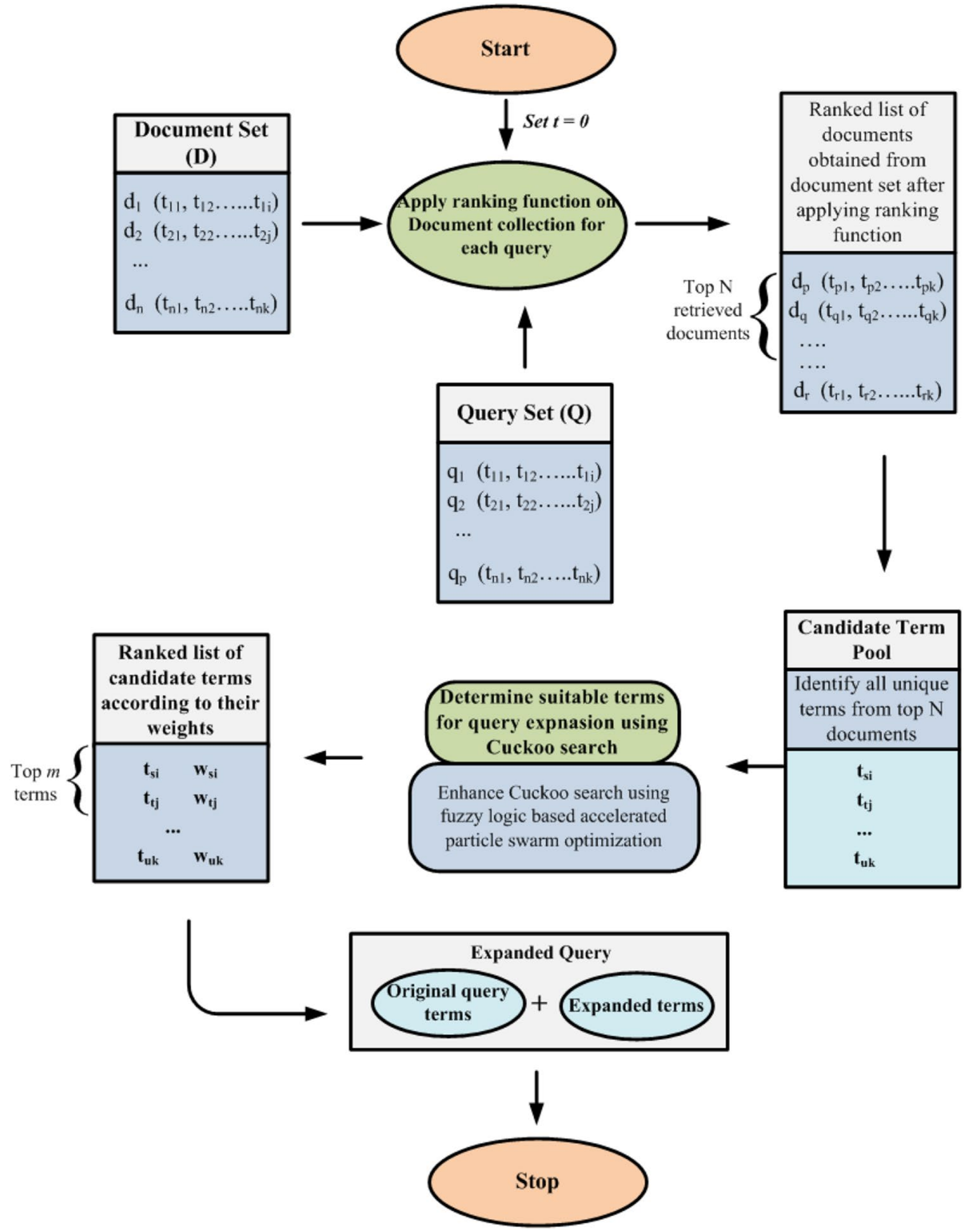

Start
Set t = 0
Document Set (D)
d_1 $(t_{11}, t_{12}……t_{1i})$
d_2 $(t_{21}, t_{22}……t_{2j})$
…
d_n $(t_{n1}, t_{n2}…..t_{nk})$
Apply ranking function on Document collection for each query
Ranked list of documents obtained from document set after applying ranking function
Top N retrieved documents
d_p $(t_{p1}, t_{p2}…..t_{pk})$
d_q $(t_{q1}, t_{q2}…..t_{qk})$
….
….
d_r $(t_{r1}, t_{r2}…..t_{rk})$
Query Set (Q)
q_1 $(t_{11}, t_{12}……t_{1i})$
q_2 $(t_{21}, t_{22}……t_{2j})$
…
q_p $(t_{n1}, t_{n2}…..t_{nk})$
Candidate Term Pool
Identify all unique terms from top N documents
t_{si}
t_{tj}
…
t_{uk}
Determine suitable terms for query expnasion using Cuckoo search
Enhance Cuckoo search using fuzzy logic based accelerated particle swarm optimization
Ranked list of candidate terms according to their weights
Top m terms
t_{si} w_{si}
t_{tj} w_{tj}
…
t_{uk} w_{uk}
Expanded Query
Original query terms
+
Expanded terms
Stop

Fig. 2 Architecture for the proposed query expansion approach

in determining optimized solution is variable length of the original query and expanded query. Therefore, the length of expanded query is fixed. Therefore, the solution is represented as (1).

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus Levy(s, \lambda), \quad (1)$$

where, $x^{(t+1)}$ is a new solution for a cuckoo i. α denotes the step size and it depends on the problem. Levy(s, γ) is *Levy* flight, which is defined as (2).

$$Levy(s, \lambda) \,:\, s^{-\lambda}, (1 < \lambda \le 3) \quad (2)$$

which has an infinite variance with an infinite mean (Gao et al. 2010). Here, s represents the step size drawn from a Levy's distribution.

4.2 Population initialization

The first population is initialized i.e., initial nest with n eggs. Each egg in initial nest is represented by a |Q|-dimension vector.

4.3 Fitness function used in proposed approach

The main objective of a fitness function is to determine the solution quality. In this work, a solution can be defined in terms of an expanded query which contains new terms. Therefore, it is obvious to use inverted indexes to determine its performance and then after the relevance score of each document is computed for expanded query. Then, the best solution is taken according to the fitness value i.e., expanded query in our case. The fitness function used in present work can be given as (3).

$$f(\dot{Q}) = \max(ScoreBM25(d1, Q + \dot{Q}), \ldots\ldots.., ScoreBM25(d|R|, Q + \dot{Q})), \tag{3}$$

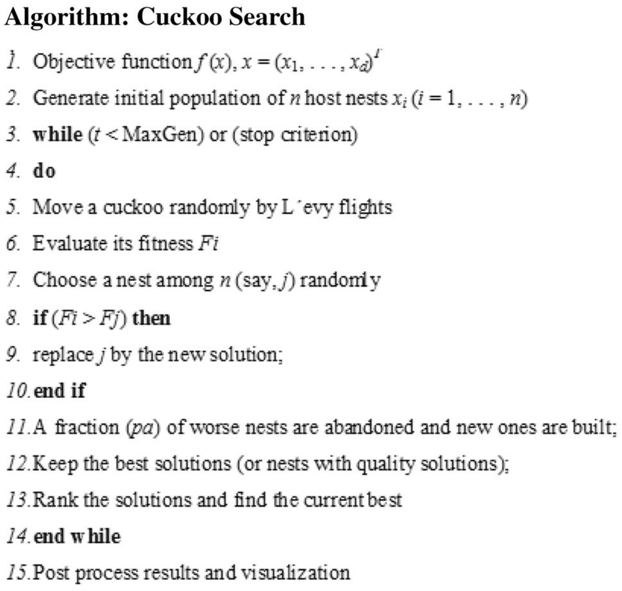
where, $\text{Score}_{\text{BM25}}$(d,Q) is the similarity value of document d against a query Q. This score is computed using Okapi-BM25 ranking function.

This fitness function depends on the inverted index of each expansion term $`qi \in `Q$, and determines the relevancy score of documents.

4.4 Update cuckoo search algorithm using fuzzy based accelerated PSO

The parameters of Cuckoo Search (*CS*) play very important role in generating the optimal solution. The performance of Cuckoo Search basically depends on the values of few parameters i.e. Cuckoo Search parameters. In this work, the problem of optimizing *CS* parameters is considered as a meta-optimization problem and fuzzy logic based accelerated *PSO* is applied to compute the optimal values for these *CS* parameters.

The pseudo-code for Cuckoo Search algorithm is given by Tuba et al. in (2011) as follows:



Algorithm: Cuckoo Search

```
Objective function $f(x)$, $x = (x_1, \ldots, x_d)^T$
Generate initial population of $n$ host nests $x_i$ ($i = 1, \ldots, n$)
while ($t$ < MaxGen) or (stop criterion)
do
Move a cuckoo randomly by L´evy flights
Evaluate its fitness $Fi$
Choose a nest among $n$ (say, $j$) randomly
if ($Fi > Fj$) then
replace $j$ by the new solution;
end if
A fraction ($pa$) of worse nests are abandoned and new ones are built;
Keep the best solutions (or nests with quality solutions);
Rank the solutions and find the current best
end while
Post process results and visualization
```

The parameters p_d, λ and α in cuckoo search are used to find global solutions and local solutions as well. The parameters p_d and α adjust the convergence rate of algorithm. The performance of this algorithm depends on the values of p_d and α. In this work, fuzzy based accelerated PSO is used to find optimal values of p_d and α.

4.5 Accelerated PSO

The standard particle swarm optimization technique updates the velocity of particles by using two parameters: current global best (x_*) and individual best position (p_i). The significance of individual best position is for diversifications of swarms. However, randomness may be used for this diversification. Therefore, this parameter may be avoided to consider. But, on the other hand, global best is used for convergence of the algorithm, which must be considered for optimal solutions. The velocity v_{t+1i} in accelerated PSO is given by (4):

$$v_{t+1i} = v_t I + \beta(x_* - x_{ti}) + \alpha. \tag{4}$$

The accelerated PSO updates particle positions by using following expression:

$$x_{t+1}I = x_t I + \beta(x_* - x_{ti}) + \alpha. \tag{5}$$

4.6 Fuzzy logic based accelerated PSO

Accelerated PSO is an evolutionary optimization technique, which is nonlinear and dynamic in nature. This technique contains many parameters like acceleration coefficients and inertia. These parameters affect performance of *PSO* directly and control the local and global optimum points.

Few efforts have been made to improve the efficiency of PSO (Suganthan 1999). Suganthan decreased the values of c_1 and c_2 linearly with time and tested the approach. He also fixed the values of c_1 and c_2 at 2 and tested the approach. He observed that the variation in the values of c_1 and c_2 enhance the performance and generate better solutions. It is desirable in population based optimization techniques to explore the entire search space during the early stages of optimization. It is very important to get converged at later stages to get the optimum solution. Therefore, we have used fuzzy logic with accelerated *PSO* to improve its performance and to make it dynamic in nature by controlling its coefficients.

In proposed approach, the values of c_1 and c_2 are being changed adaptively using fuzzy logic. We have observed an improvement in PSO when the value of c_1 is varying from 2.5 to 0.5 and the value of c_2 is varying from 0.5 to 2.5. The variation of these acceleration coefficients depend on *gbest* (global best solution), *UN* (the number of generations for which the value of *gbest* is not changed) and other variables.

The variation of these *two* parameters depends on the following facts:

- If the value of *gbest* is not being changed significantly over the generations then the values of c_1 and c_2 must be changed so that gbest may get some improvement.
- The current value of c_1 and c_2 also affect the variation in acceleration coefficients (δc_1 and δc_2).
- The values of δc_1 and δc_2 also depend on the value of *gbest*.

In this approach, we have used *four* input variables: *gbest, UN,* c_1 and c_2. These input variables are transformed into fuzzy set from crisp values using fuzzification process. This fuzzification process maps the inputs into fuzzy set using the membership functions and linguistic terms. We have used *five* membership functions in this work to represent each input variable as demonstrated in Fig. 3. The linguistic terms to represent these membership functions are Very High (VH), High (H), Medium (M), Low (L) and Very Low (VL). In the proposed approach, output variables (variation in c_1 and c_2) are also presented by *five* linguistic terms such as Very High (VH), High (H), Medium (M), Low (L) and Very Low (VL) as shown in Fig. 4.

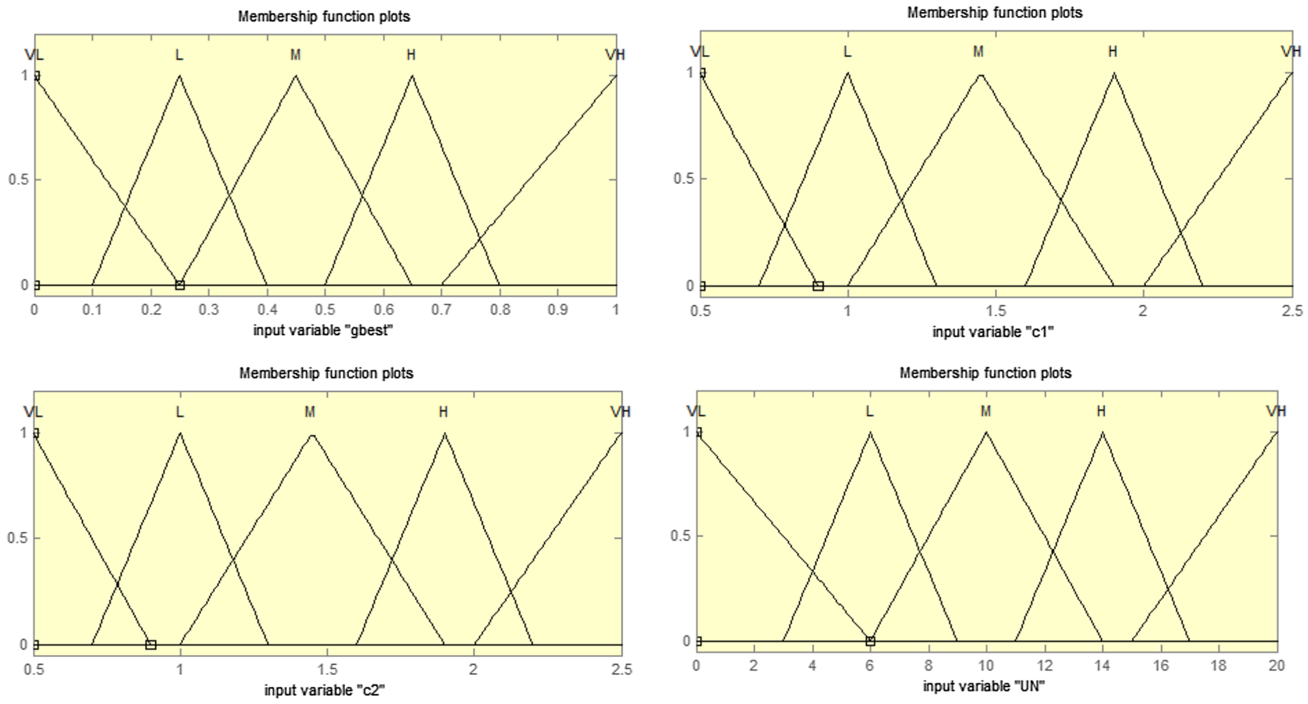


Fig. 3 Membership functions and linguistics terms for input variables

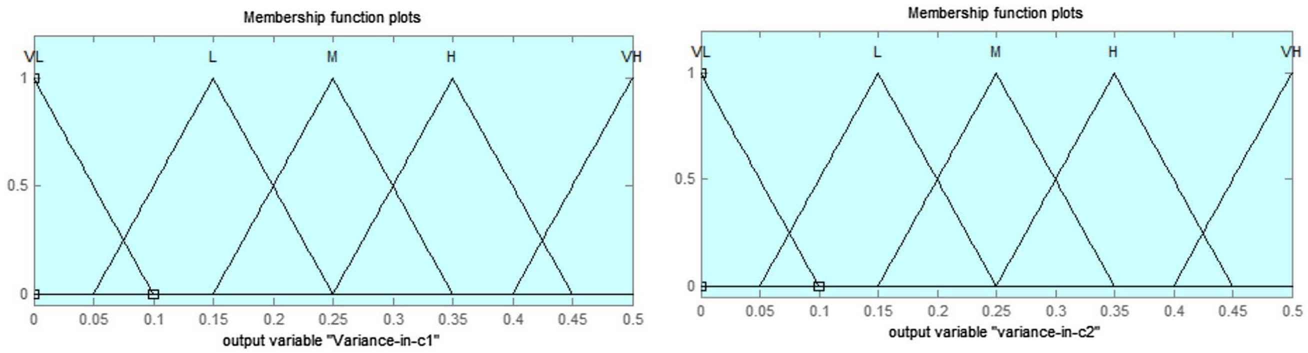


Fig. 4 Membership functions and linguistics terms for output variables

4.7 Framing fuzzy rules for fuzzy based accelerated PSO

The fuzzy rules for *fuzzy based accelerated PSO* are designed from the knowledge of domain. Table 1 presents the used knowledge of domain to frame these fuzzy rules and the examples.

The framed fuzzy rules are used to infer the values of variation in c_1 (δc_1) and variation in c_2 (δc_2). We have used *Mamdani-type* fuzzy rule frame and have evaluated the conditional statements. There is no effect of the order of fuzzy rules on results. All framed fuzzy rules carry equal weights in the proposed approach. To combine membership values for each active fuzzy rule, AND operator is used. We fix the range [0, 0.5] for δc_1 and δc_2. The values of c_1 and c_2 are computed for next iteration using following equations:

$$c_1(next) \ = c_1(current) \ - \ \delta c_1, \tag{6}$$

$$c_2(next) \ = c_2(current) \ - \ \delta c_2. \tag{7}$$

Figure 5 explains the whole fuzzy inference process by showing all fuzzy rules. This figure clearly shows that total 24 fuzzy rules are designed in this work, which is represented by 24 rows. Each row consists of *six* small rectangular plots. Each column in the figure presents a variable. We have total six variables (four input variables and two output variables) in proposed approach. In the last row, the plots in fifth and sixth columns show the aggregated weights of δc_1

Table 1 Knowledge base and examples of framed fuzzy rules in proposed approach

S. no.	Knowledge base	Examples of fuzzy rules
1	If Normalized gbest is low, UN is low and acceleration coefficients (c_1; and c_2) are also low then variation in c_1 (Δc_1) and variation in c_2 (Δc_2) are likely to be *medium*	If Normalized gbest is Low and *UN is Low* and c_1, c_2 are Low then Δc_1 and Δc_2 *are Medium*
2	If *Normalized gbest* is *low*: and *UN* is *low* and acceleration coefficients (c_1; and c_2) are High then variation in c_1 (Δc_1) and variation in c_2 (Δc_2) are likely to be *high*	If *Normalized gbest* is *Low* and *UN* is *Low*; and c_1, c_2 are High then Δc_1; and Δc_2 are *High*

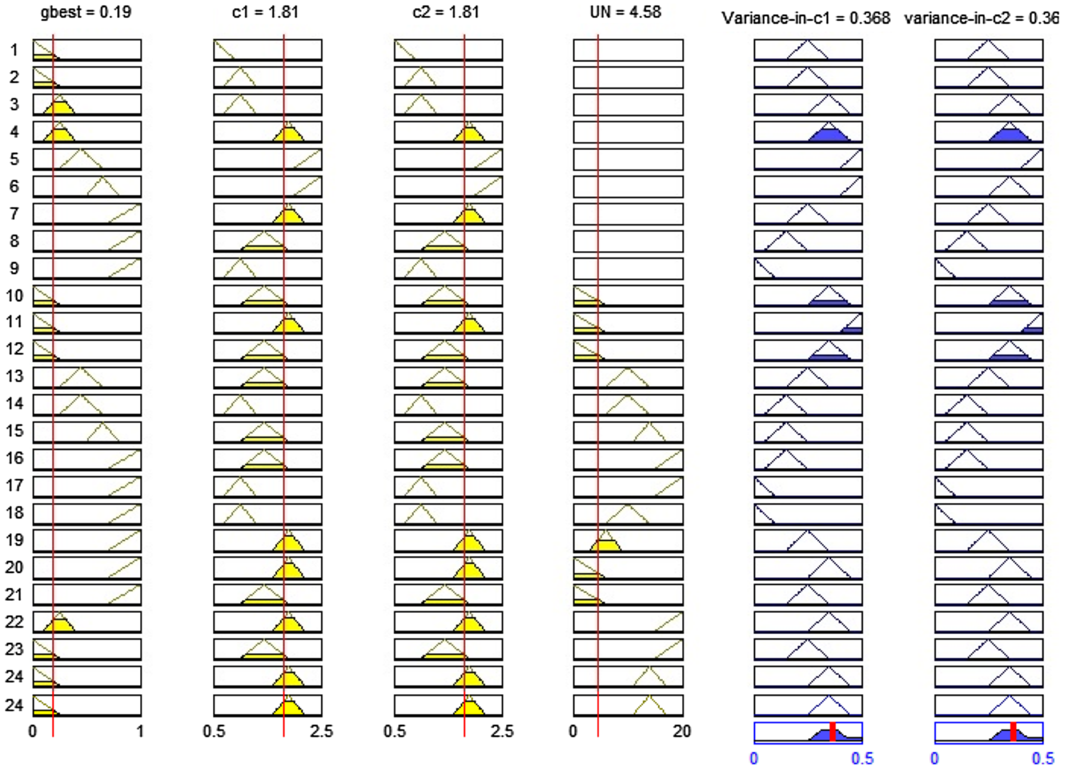


Fig. 5 Rule view diagram

and δc_2. This diagram may also be used to see active and passive rules.

We have plotted *three* dimensional surfaces diagrams to show the dependencies of outputs (i.e. δc_1 and δc_2) on input variables (i.e. *gbest*, c_1, c_2 and *UN*) as shown in Fig. 6. This figure shows that δc_1 decreases when the values of *UN* and *gbest* increases. Similar type of analysis can be done for other three figures. We have used *centroid method as* a defuzzification method to obtain a crisp value from output fuzzy set. The centroid method can be expressed mathematically as (8).

$$Y = \int_Y \sum_{i=1} Y \cdot \mu_{Bi}(y)dy \bigg/ \int_Y \sum_{i=1} \mu_{Bi}(y)dy, \tag{8}$$

where the input is a fuzzy set $\mu_{Bi}(y)$ for the defuzzification process and Y denotes the output, which is actually a crisp valued number.

5 Experimental results and discussion

We have used three benchmark and widely used datasets *TREC-3, CISI* and *CACM* to perform all the experiments. We have selected total *fifty* queries randomly from these datasets to test the performance of proposed *AQE* approach and to compare with other similar type of approaches. The statistics of *TREC-3, CISI* and *CACM* datasets are tabulated in Table 2. We have analyzed the results in following *three* ways:

1. Query specific analysis.
2. Overall performance analysis.
3. Statistical analysis.

In this paper, we have compare the results of proposed *AQE* approach with *original query;* Singh and Sharan (2017a) *QE* approach and hybrid soft computing based approach given by Ramalingam and Dhandapani (2014).

Table 2 Statistics of CACM, CISI and TREC-3 Datasets

Collection	Number of documents	Number of queries	Average age of Length documents	Average length of queries
CACM	3204	64	24.52	10.80
CISI	1460	112	46.55	28.29
TREC-3	6,72,611	50	237.56	5.83

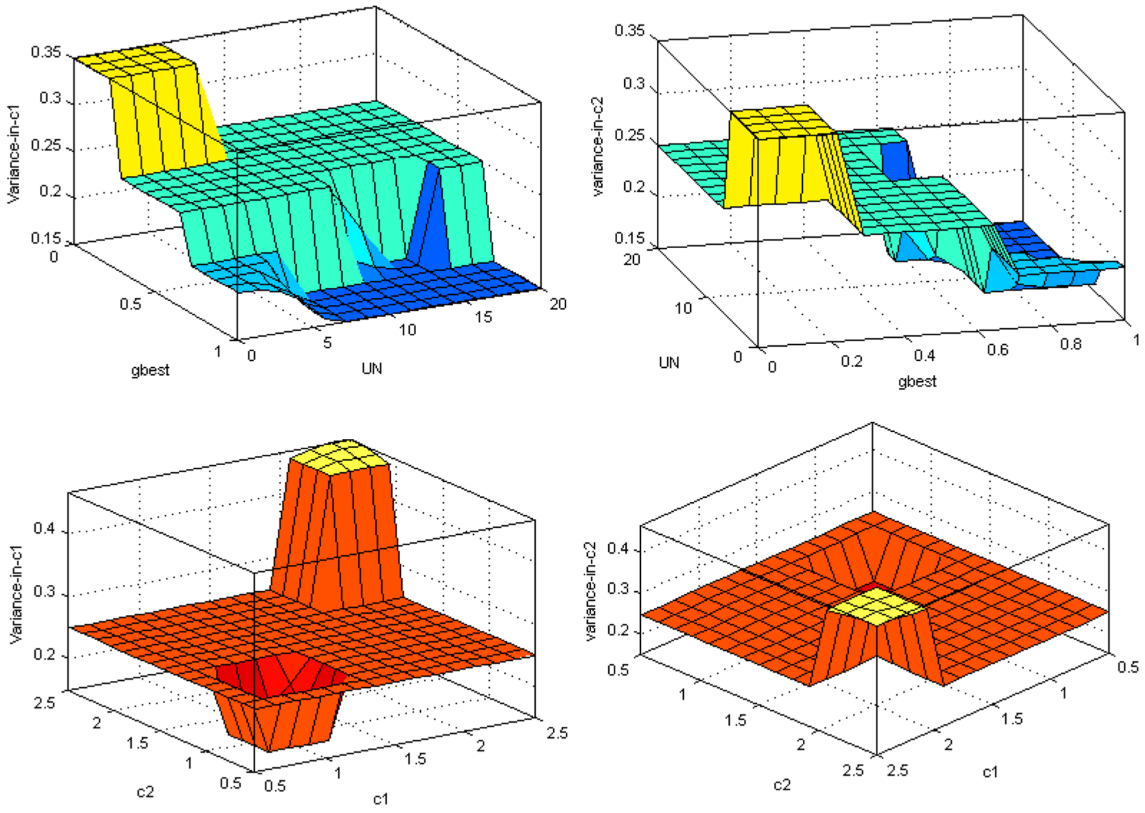


Fig. 6 Fuzzy Inference surface view diagram

Table 3 Parameters used in the proposed approach

Parameter type	Parameter name	Value
Cuckoo search	Levy step size α	0–1.5
	λ	1–3
APSO	C_1	0.5–2.5
	C_2	2.5–0.5
	w	0.9
	No. of iteration	100

The following values (tabulated in Table 3) of different parameters of cuckoo search and accelerated PSO are used to perform experiments.

5.1 Query wise retrieval effectiveness

In this paper, we have computed *F-Measure* to test the performance of each query. The results are analyzed and compared with the initial user query i.e. *original query*, Singh and Sharan (2017a) and Ramalingam et al. (2014) *AQE* approaches. We have determined *F-measure* value at three cut-offs. These cut-offs are top 10, *top* 30 and *top* 50 documents for all three standard datasets. The results for these three cut-offs show the consistency of the proposed approach. We have obtained better F-measure values using proposed approach in comparison to other approaches for both the datasets as shown in Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15.

Figure 7 shows the comparison of F-measure obtained by different query expansion approaches along with proposed approach at top 10 documents cut-off for *CACM* dataset. It can be observed from the figure easily that the proposed *AQE* approach is outperforming other approaches for all randomly selected 50 queries. The proposed *AQE* approach achieves better *F-measure values* as compared to *Singh and Sharan approach* for 45 queries out of 50. However, *F-measure* are equal for three queries of both approaches. Figure 8 depicts the results in terms of *F-measure* for top 10 documents cut-off for *CISI*. This diagram clearly depicts that the

Fig. 7 Comparison of *F-measure* at top 10 documents cut-off for *CACM* dataset

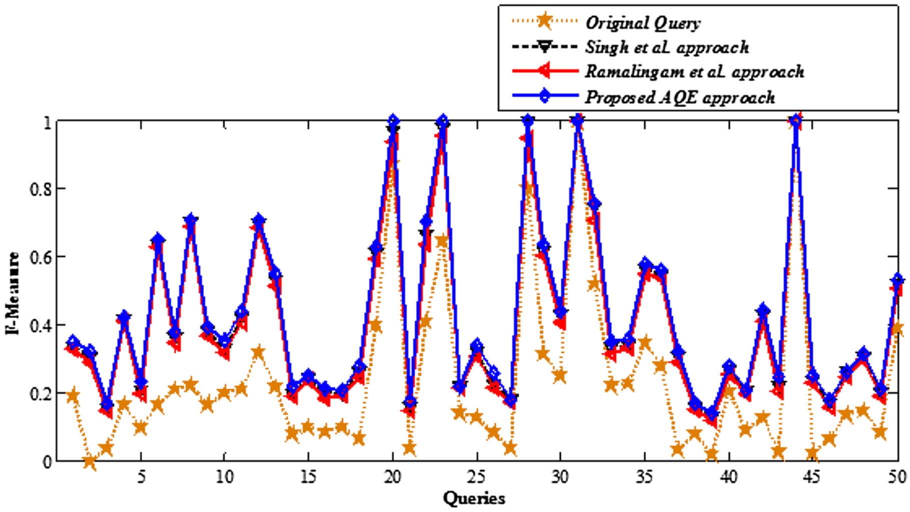


Fig. 8 Comparison of *F-measure* at top 10 documents cut-off for *CISI* dataset

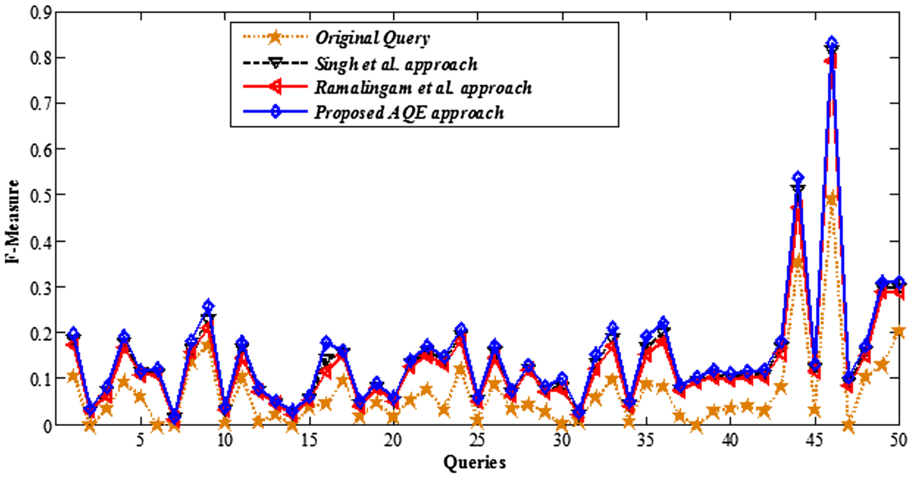


Fig. 9 Comparison of *F-measure* at top 10 documents cut-off for *TREC-3* dataset

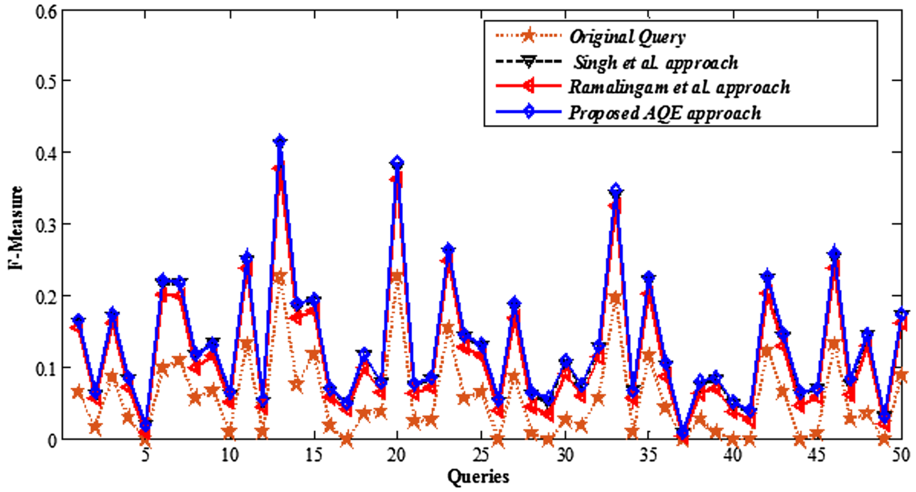


Fig. 10 Comparison of *F-measure* at top 30 documents cut-off for *CACM* dataset

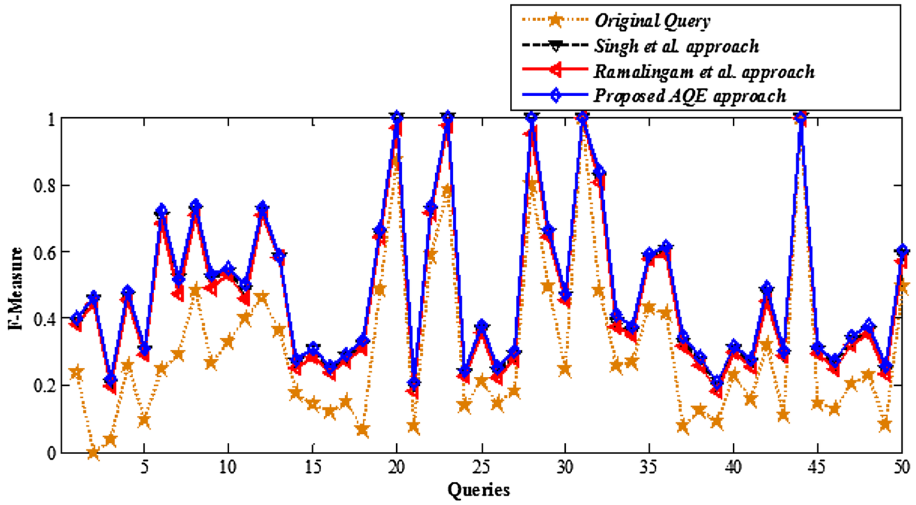
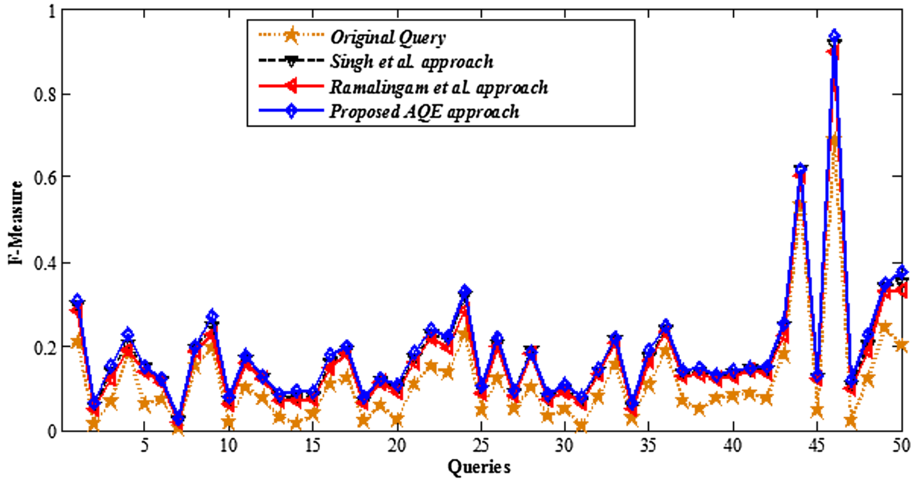


Fig. 11 Comparison of *F-measure* at top 30 documents cut-off for *CISI* dataset



higher values of *F-measure* are obtained by proposed AQE approach as compared to *Ramalingam et al*. and *Singh et al. AQE* approaches for 48 queries and 45 queries respectively.

Figure 9 demonstrates the results for *TREC-3* standard dataset at top 10 documents cut-off. This figure clearly shows

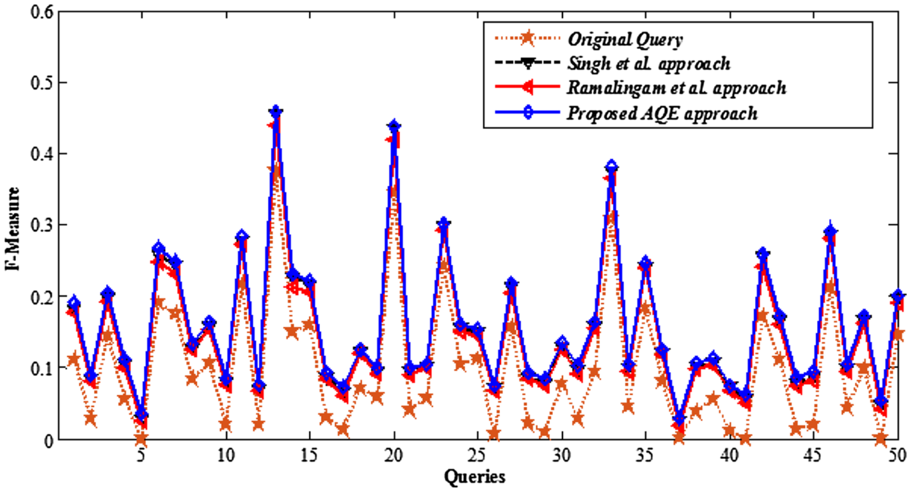


Fig. 12 Comparison of *F-measure* at top 30 documents cut-off for *TREC-3* dataset

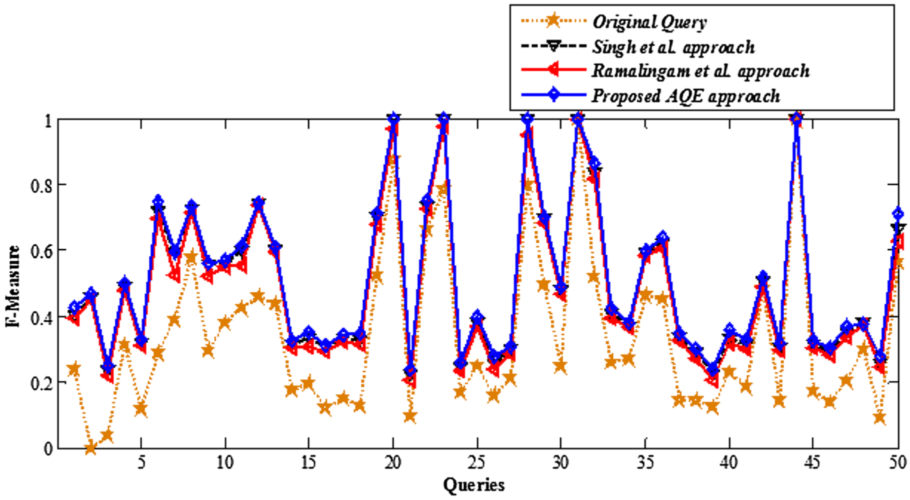


Fig. 13 Comparison of *F-measure* at top 50 documents cut-off for *CACM* dataset

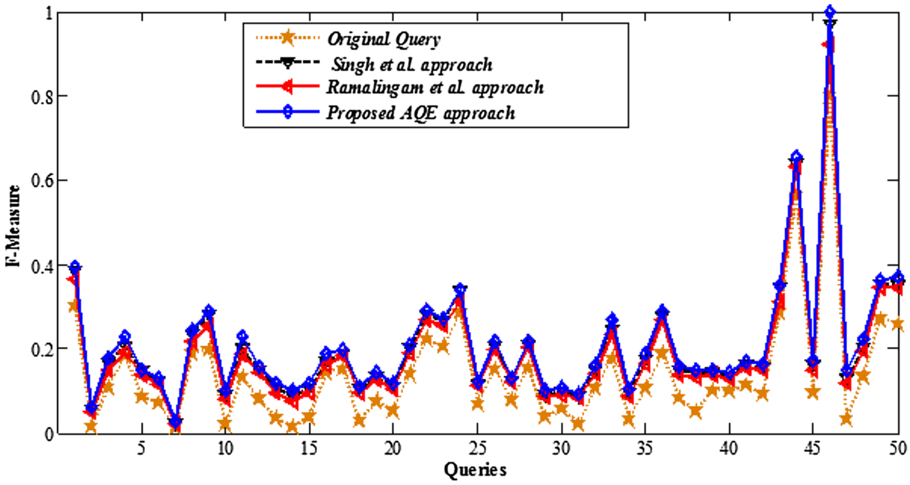


Fig. 14 Comparison of *F-measure* at top 50 documents cut-off for *CISI* dataset

that F-measure value is obtained by proposed approach is better for all the selected queries.

Figures 10, 11 and 12 demonstrate the results for top 30 documents for TREC-3, CISI and CACM. These figures show that the proposed *AQE* approach gives better

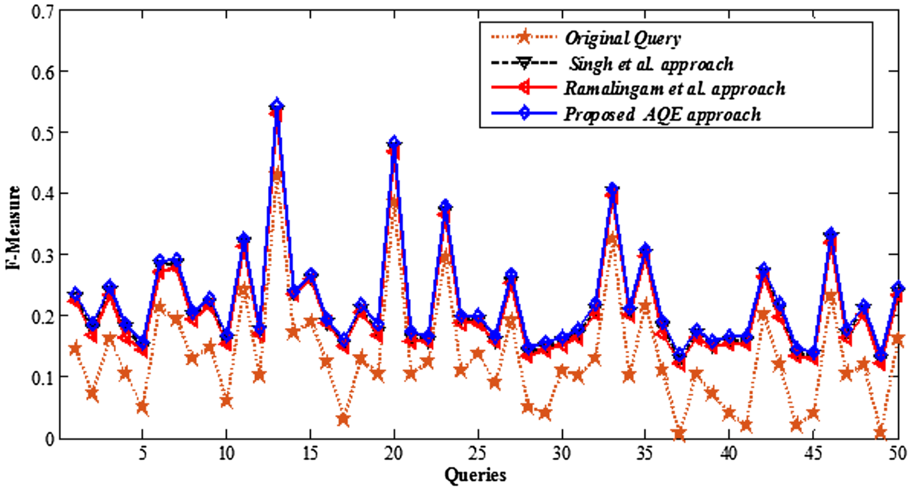


Fig. 15 Comparison of *F-measure* at top 50 documents cut-off for *TREC-3* dataset

F-measure for 48 queries, 44 queries and 40 queries in case of *TREC-3, CISI* and *CACM* datasets respectively as compared to Ramalingam et al. and Singh et al. *query expansion approach*es. Similarly, Figs. 13, 14 and 15 also show that the superiority of proposed *AQE* approach over other query expansion approaches for top 50 retrieved documents cut-off in case of *CACM, CISI* and *TREC-3* datasets.

The performance improvement in *precision values* for each individual query for proposed AQE approach is also analyzed. Singh et al. and Ramalingam et al. *AQE approaches* are taken as baseline approaches. Figures 16, 17, 18, 19, 20 and 21, show the range of *precision* variations using bars of proposed *AQE* approach over Singh et al. and Ramalingam et al. approaches. The precision values are taken for top 50 documents cut-off. Figure 16 depicts that there is improvement in results obtained by the proposed approach for 41 queries and decrement in nine queries in comparison to Singh et al. approach for *CACM* dataset. Figure 17 also shows the improvement in results obtained by the proposed approach for 43 queries and decrement for *seven* queries only in comparison with Singh et al. *approach* in case of *CISI* dataset. Figure 18 shows the performance variation for *TREC-3* dataset. It depicts that the proposed approach gives better results for 48 queries as compared to Singh et al. approach. Figures 19, 20 and 21 demonstrates that proposed *AQE* increases the performance for most of the queries for all three standard datasets i.e. *TREC-3, CISI* and *CACM* as compared to Ramalingam et al. approach.

5.2 Overall retrieval effectiveness

We have also check the overall effectiveness of all the approaches in terms of *two* evaluating parameters: *MAP* and *P@rank*. This analysis is done for *TREC-3, CISI* and *CACM* datasets. The comparative study on *MAP* values of proposed *AQE* approach with initial user query, Singh et al. *AQE* approach and Ramalingam et al. approach is shown in Table 4. This table clearly presents that the proposed *AQE* approach gives higher values of *MAP* as compared to

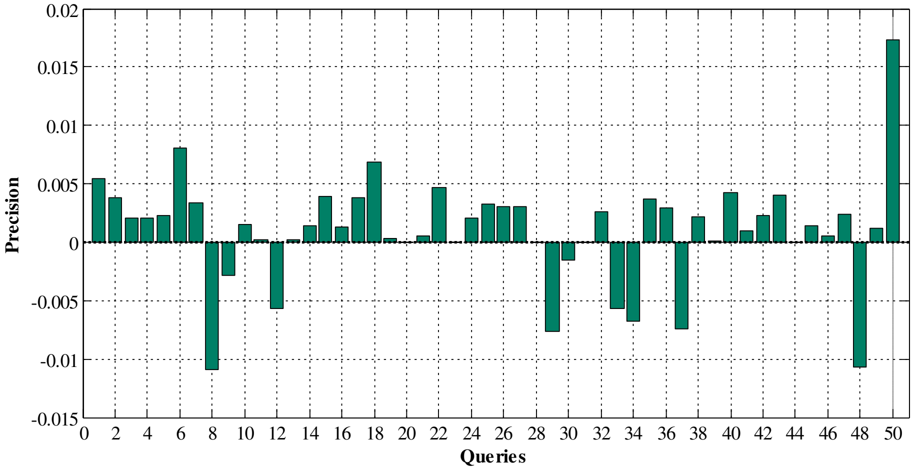


Fig. 16 Query-wise variation in precision-value of proposed AQE approach with respect to Singh et al. approach for *CACM*

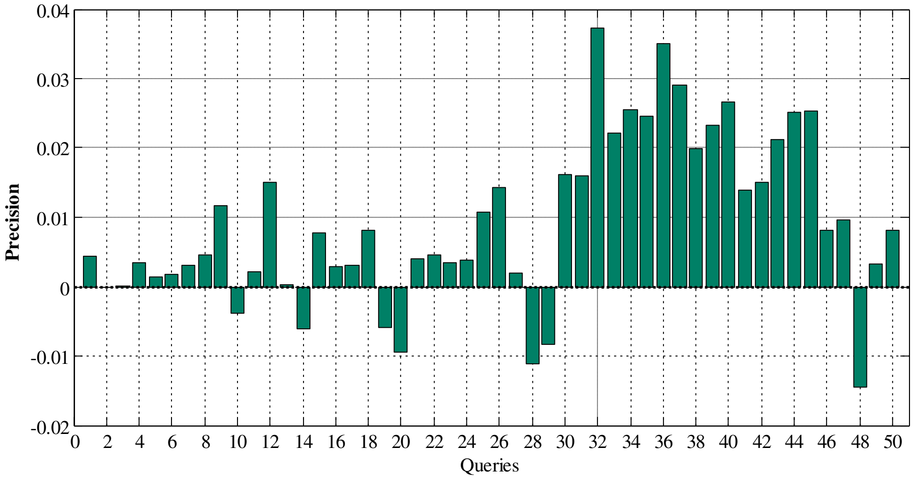


Fig. 17 Query-wise variation in precision-value of proposed AQE approach with respect to Singh et al. approach for *CISI*

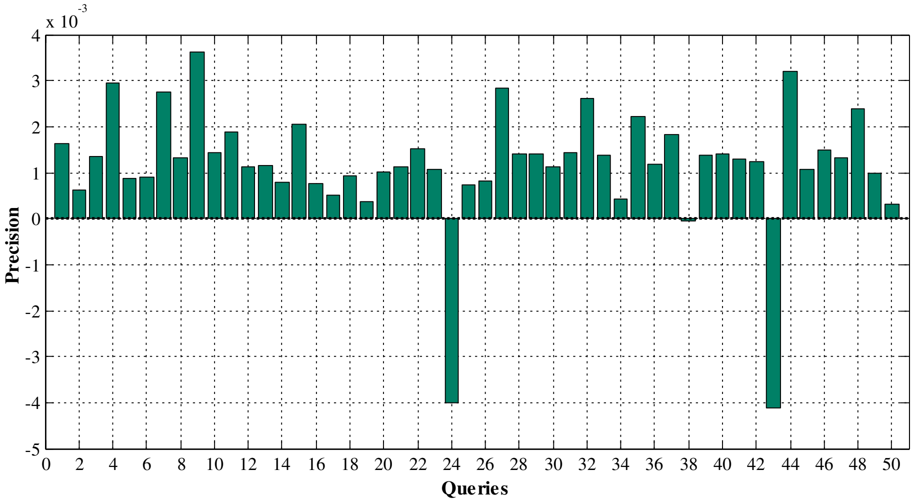


Fig. 18 Query-wise variation in precision-value of proposed AQE approach with respect to Singh et al. approach for *TREC-3*

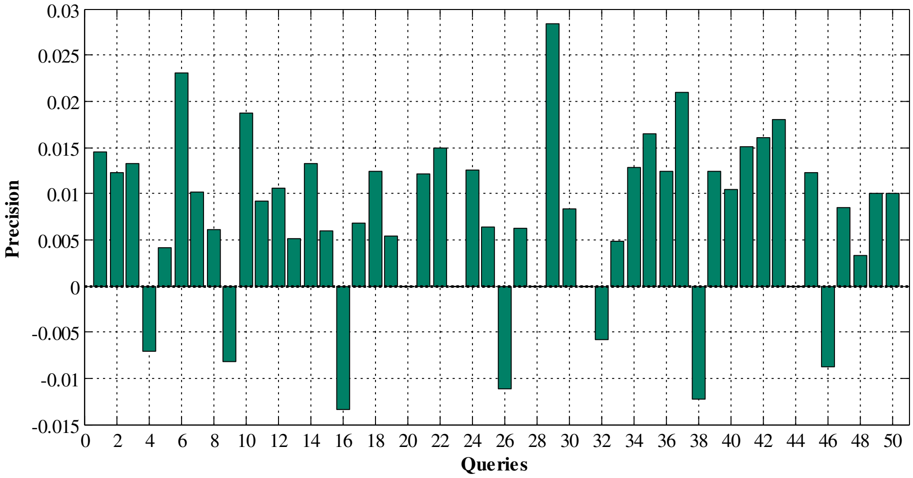


Fig. 19 Query-wise variation in precision-value of proposed AQE approach with respect to Ramalingam et al. approach for *CACM*

initial user query, Ramalingam et al. and Singh et al. *AQE approaches*.

We have also computed *P@rank* to test the overall performance of all *AQE* approaches and comparative results are tabulated in Tables 5, 6 and 7. These tables clearly show that the proposed automatic query expansion approach gives better *P@rank* values as compared with other *AQE* approaches for all used datasets. The

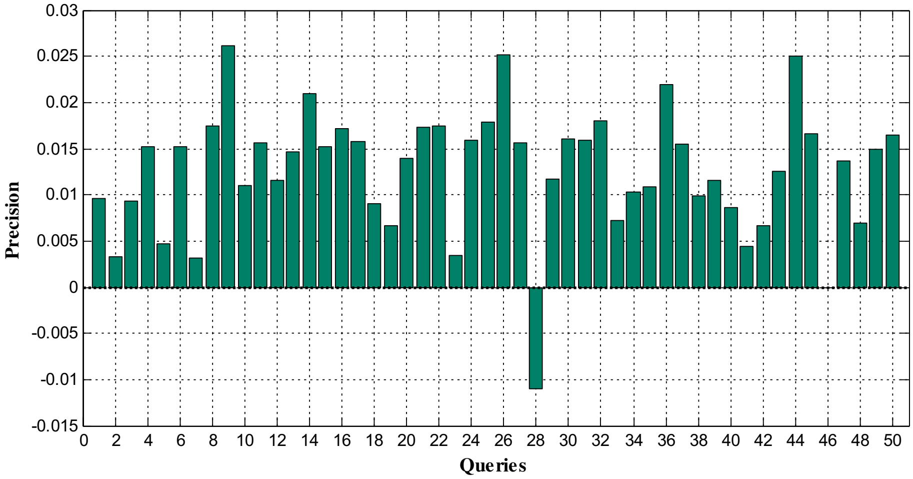


Fig. 20 Query-wise variation in precision-value of proposed AQE approach with respect to Ramalingam et al. approach for *CISI*

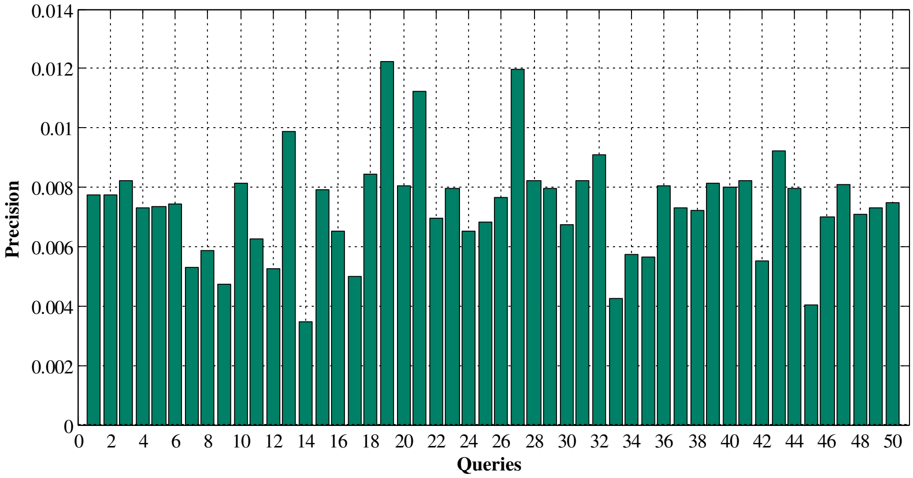


Fig. 21 Query-wise variation in precision-value of proposed AQE approach with respect to Ramalingam et al. approach for *TREC-3*

Table 4 *MAP* value comparison of all AQE approaches along with *proposed approach*

Dataset	Original query	Ramalingam et al. approach	Singh et al. approach	Proposed AQE approach
CACM	0.1873	0.2767	0.2791	0.2316
CISI	0.1586	0232464	0.2467	0.2534
TREC-3	0.1957	0.2357	0.2889	0.2908

Table 5 *P@rank* value comparison of all AQE approaches along with *proposed approach* for *CACM*

	Original query	Ramalingam et al. approach	Singh et al. approach	Proposed AQE approach
P@5	0.4042	0.7655	0.7750	0.7881
P@10	0.3585	0.7111	0.7219	0.7353
P@15	0.3151	0.6769	0.6807	0.6925
P@20	0.2917	0.6356	0.6468	0.6598
P@30	0.2476	0.5840	0.5942	0.6093
P@50	0.1610	0.4944	0.5037	0.5133

Table 6 *P@rank* value comparison of all AQE approaches along with *proposed approach* for *CISI*

	Original query	Ramalingam et al. approach	Singh et al. approach	Proposed AQE approach
P@5	0.3648	0.6428	0.6683	0.6722
P@10	0.3202	0.6029	0.6248	0.6278
P@15	0.2361	0.5479	0.5667	0.5715
P@20	0.2536	0.5116	0.5289	0.5343
P@30	0.1983	0.4626	0.4782	0.4798
P@50	0.1123	0.3781	0.3938	0.4012

consistency of the proposed approach is also checked using *Precision–Recall* graphs. Figures 22, 23 and 24 show that proposed *AQE* approach outperforms initial user query, Ramalingam et al. *approach* and Singh et al. *approach* at all levels of *recall* for *TREC-3, CISI* and *CACM* datasets.

Above mentioned results clearly depict that the proposed automatic query expansion approach gets better performance in comparison to *Original user query*, Ramalingam et al.

Table 7 *P@rank* value comparison of all AQE approaches along with *proposed approach* for *TREC-3*

	Original query	Ramalingam et al. approach	Singh et al. approach	Proposed AQE approach
P@5	0.5087	0.7051	0.7092	0.7120
P@10	0.4783	0.6597	0.6641	0.6678
P@15	0.4527	0.6289	0.6334	0.6351
P@20	0.4431	0.6071	0.6130	0.6182
P@30	0.4073	0.5662	0.5759	0.5798
P@50	0.2905	0.4607	0.4646	0.4662

approach and Singh et al. *approach. The proposed* approach gives better results because of the variations in cuckoo search parameters and accelerated *PSO* parameters throughout the generations. This variation is shown in Figs. 25 and 26 for two standard datasets i.e. *CISI* and *CACM* datasets. We are controlling acceleration coefficients in a *PSO* using fuzzy. The parameters lambda and levy step size of cuckoo search are controlled by accelerated *PSO*.

As our approach is based on cuckoo search and accelerated PSO and these are evolutionary algorithms. Therefore, the time complexity is more in comparison to other approaches. But we have tried to minimize it using fuzzy logic, which increases the convergence rate and makes accelerated PSO adaptive in nature.

5.3 Statistical analysis

We have performed *paired t test* to check the authenticity and reliability of the proposed *AQE* approach. Table 8 clearly illustrates that superiority of proposed *AQE* approach over other *AQE* approaches. It shows that the results obtained from the proposed approach are statistically significant at α = 0.05. As *p value* is *0.0011, 0.0019* and *0.0003* against original query for *TREC-3, CISI* and *CACM* respectively. The *p value* is *0.0117* for *CACM, 0.0192* for *CISI* and *0.0156* for *TREC-3* against *Ramalingam et al. approach*. The proposed *AQE* approach is also significantly different from *Singh et al. approach* at *p value = 0.0056* for *TREC-3, p value = 0.0307* for *CACM* and *p-value = 0.0392* for *CISI.*

Fig. 22 *Precision-Recall* graph for *CACM* dataset

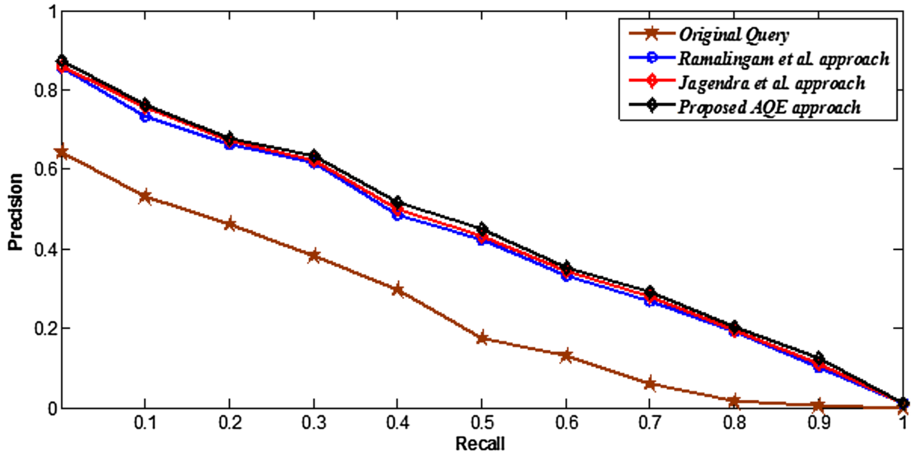


Fig. 23 *Precision-Recall* graph for *CISI* dataset

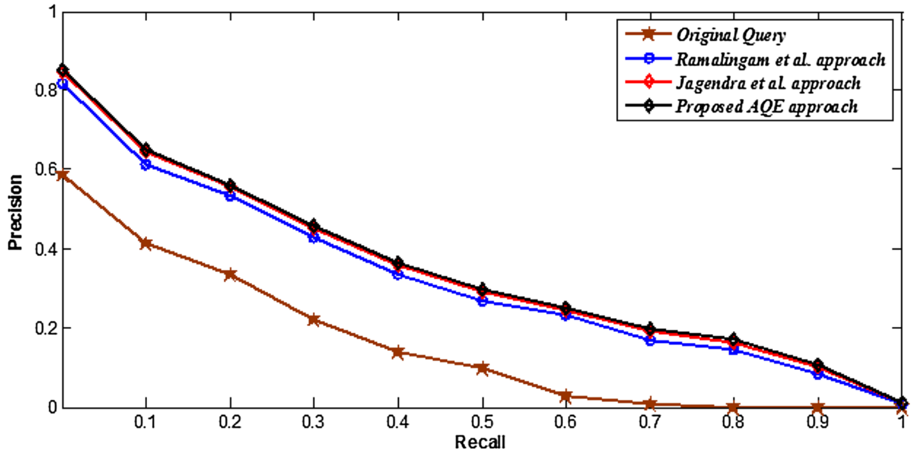
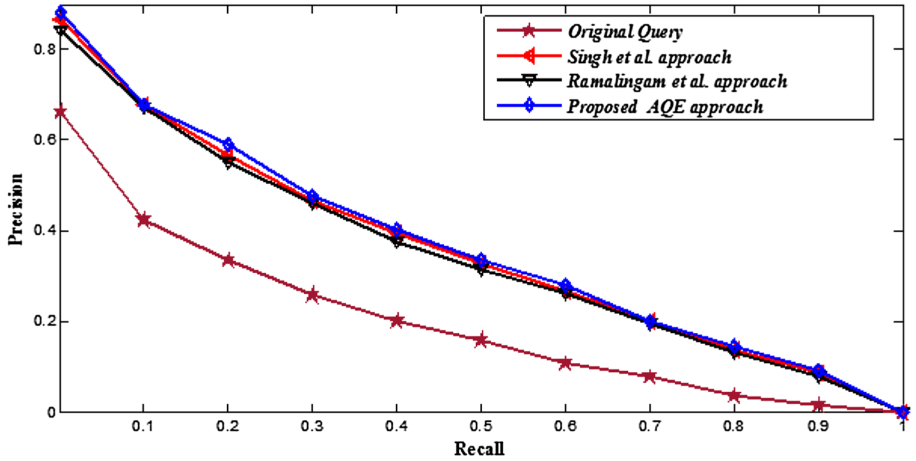


Fig. 24 *Precision-Recall* graph for *TREC-3* dataset



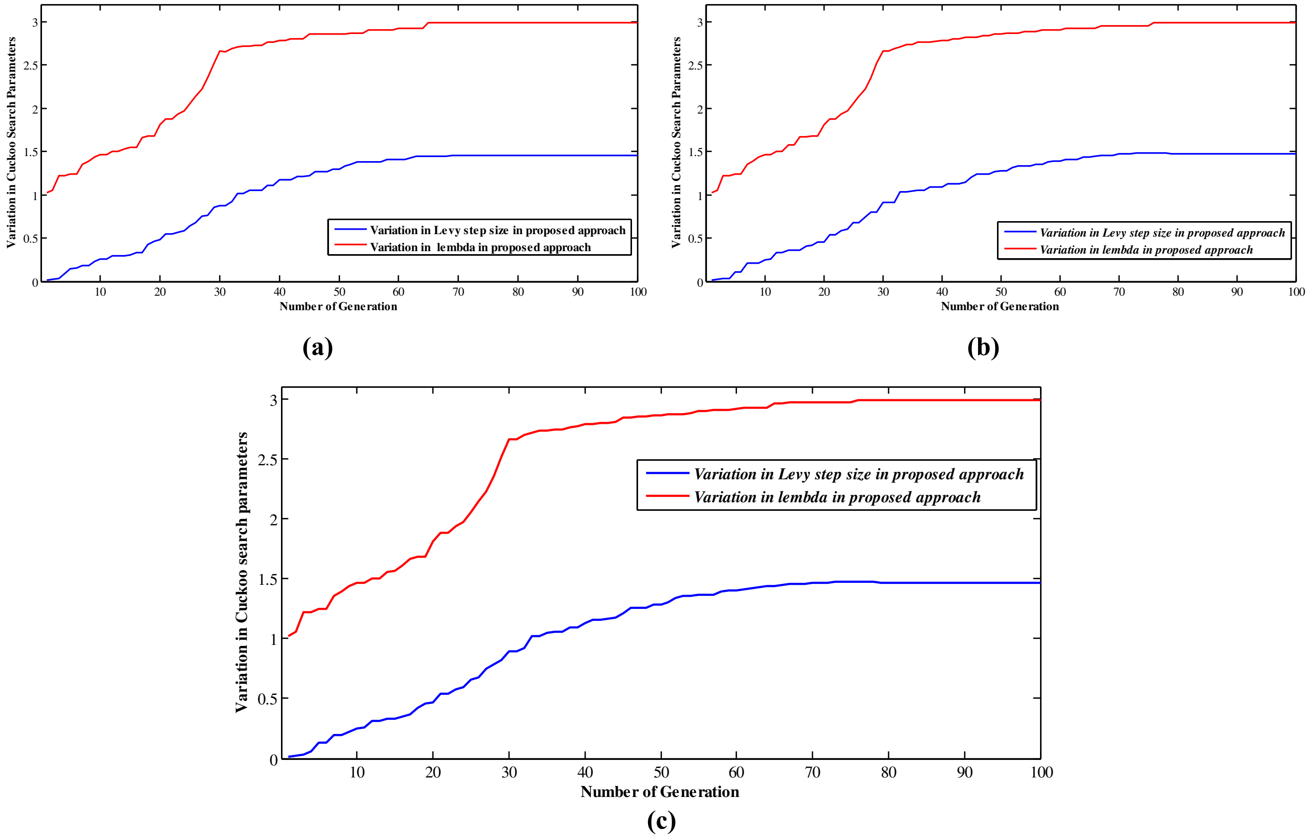
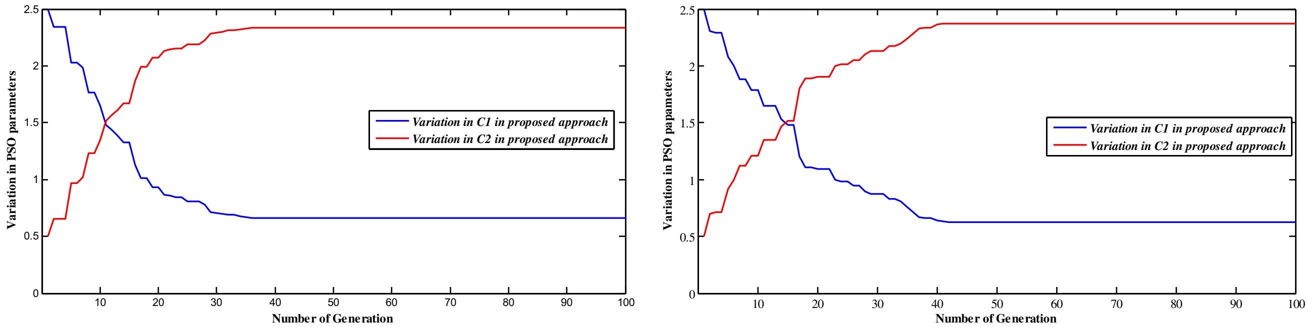


Fig. 25 Variation in cuckoo search parameters over generations for **a** *CACM*, **b** *CISI* and **c** *TREC-3*

6 Conclusion

A new automatic query expansion approach using hybrid evolutionary techniques is proposed and developed in this work. The proposed *AQE* approach overcomes limitations existed in conventional query expansion approaches by modeling a query in a new way. This proposed approach mainly focused on generating suitable expanded queries rather than finding terms for expansion. In this approach, the best suitable expanded query is identified from a set of all possible expanded queries i.e. candidate query pool. This candidate query pool consists of all possible combinations of terms those are selected from top retrieved documents. Cuckoo search and accelerated *PSO* techniques are used



(a) (b)

Fig. 26 Variation in accelerated PSO parameters over generations for **a** *CACM* and **b** *CISI*

Table 8 Results for the *Paired t-test*

	Datasst	Propossd AGE approach		
		h value	*p* values	*CI*
Original query	*CACM*	1	0.0003	[−0.1724, −0.0431]
	CISI	1	0.0019	[−0.0910, −0.0345]
	TREC-3	1	0.0011	[−0.0925, −0.0393]
Ramalingam st al. approach	*CACM*	1	0.0117	[−0.0472, −0.0203]
	CISI	1	0.0192	[−0.0267, −0.0093]
	TREC-3	1	0.0156	[−0.0315, −0.0143]
Singh et al. approach	*CACM*	1	0.0307	[−0.0112, −0.0053]
	C1S1	1	0.0392	[−0.0061, −0.0010]
	TREC-3	1	0.0328	[−0.0031, −0.0012]

to find best suitable query from query pool. Fuzzy logic is used to improve the performance of accelerated *PSO*. The proposed approach is tested on three widely accepted and benchmark datasets such as *TREC-3, CISI* and *CACM*. The results obtained from experiments confirm that proposed approach performs better as compared to Singh et al. query expansion approach and Ramalingam et al. approach in terms of recall value, precision value, MAP and F-measure values. A *paired t test* analysis is also performed to validate the obtained results. This analysis shows that the proposed *AQE* approach enhances the efficiency significantly. In future, we will try other evolutionary algorithms to enhance the performance of both document retrieval and computational efficiency. We may also use other datasets for testing the performance of our approach.

Acknowledgements We are very thankful to anonymous reviewers for their valuable suggestions. We are also thankful to Dr. Ashish Saini and Dr. Yogesh Gupta to provide their support to access datasets for experiments.

References

Barathi M, Valli S (2013) Query disambiguation using clustering and concept based semantic web search for efficient information retrieval. Life Sci. J. 10(2):147–155. https://doi.org/10.7537/marslsj100213.23

Ben HE, Ounis I (2003) A study of parameter tuning for term frequency normalization. In: Proceedings of the twelfth international conference on information and knowledge management. ACM Press, New Orleans, pp 10–16. https://doi.org/10.1145/956863.956867

Bendersky M, Metzler D, Croft BW (2012) Effective query expansion with multiple information sources. fifth ACM international conference on web search and data mining, USA, pp 1–10. https://doi.org/10.1145/2124295.2124349

Bhatnagar P, Pareek N (2015) Genetic algorithm-based query expansion for improved information retrieval. In: proceedings of the international conference on intelligent computing, communication and devices, pp 47–55. https://doi.org/10.1007/978-81-322-2012-1_6

Billerbeck B, Scholer F, Williams HE, Zobel J (2003) Query expansion using associated queries. In: proceedings of the 12th international conference on information and knowledge management, New Orleans, pp 2–9. https://doi.org/10.1145/956863.956866

Carlos M, Maguitman A (2009) A semi-supervised incremental algorithm to automatically formulate topical queries. Inf Sci 179:1881–1892. https://doi.org/10.1016/j.ins.2009.01.029

Carpineto C, Romano G (2012) A survey of automatic query expansion in information retrieval. ACM Computer Survey 44(1):1–50. https://doi.org/10.1145/2071389.2071390

Chang Y, Chen C (2006) A new query reweighting method for document retrieval based on genetic algorithms. IEEE Trans Evolut Comput 10(5):617–622. https://doi.org/10.1109/TEVC.2005.863130

Chang Y, Chen S, Liau C (2007) A new query expansion method for document retrieval based on the inference of fuzzy rules. J Chin Inst Eng 30(3):511–515. https://doi.org/10.1080/02533839.2007.9671279

Chen H, Yu J, Furuse K, Ohbo N (2001) Support IR query refinement by partial keyword set. In: proceedings of the second international conference on web information systems engineering, Singapore, 11, pp 245–253. https://doi.org/10.1109/WISE.2001.996485

Cooper JW, Byrd R (1998) OBIWAN—a visual interface for prompted query refinement. In: proceedings of the 31st Hawaii international conference on system sciences, Hawaii, 2, pp 277–285. https://doi.org/10.1109/HICSS.1998.651710

Cur´e OC, Maurer H, Shah NH, Le Pendu P (2015) A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest. BMC Med Inform Decision Making 15(1):1–6. https://doi.org/10.1186/1472-6947-15-S1-S8

Enireddy V, Reddi KK (2015) Improved cuckoo search with particle swarm optimization for classification of compressed images. Sadhana Indian Acad Sci 40(8):2271–2285

Fattahi R, Concepcio´n SW, Cole F (2008) An alternative approach to natural language query expansion in search engines: text analysis of non-topical terms in web documents. Inf Process Manage 44:1503–1516. https://doi.org/10.1016/j.ipm.2007.09.009

Gao Y, Zhang G, Ma J, Lu J (2010) A λ-cut and goal-programming-based algorithm for fuzzy-linear multiple-objective bilevel optimization. IEEE Trans Fuzzy Syst 18(1):1–13. https://doi.org/10.1109/TFUZZ.2009.2030329

Gong Z, Cheang C, Hou L (2006) Multi-term web query expansion using WordNet. Database and expert systems applications. Lect Notes Comput Sci 4080(388.):379 https://doi.org/10.1007/11827405_37

Grootjen FA, Weide TP (2006) Conceptual query expansion. Data Knowl Eng 56:174–193. https://doi.org/10.1016/j.datak.2005.03.006

Gupta Y, Saini A (2017) A novel Fuzzy-PSO term weighting automatic query expansion approach using semantic filtering. Knowl Based Syst 136:97–120. https://doi.org/10.1016/j.knosys.2017.09.004

Gupta Y, Saini A, Saxena AK (2015) A new fuzzy logic based ranking function for efficient information retrieval system. Expert Syst Appl 42: 1223–1234. https://doi.org/10.1016/j.eswa.2014.09.009

Gupta PK, Lal S, Kiran MS (2018) Two dimensional cuckoo search optimization algorithm based despeckling filter for the real ultrasound images. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-0891-3

Horng J, Yeh C (2000) Applying genetic algorithms to query optimization in document retrieval. Inf Process Manage 36:737–759. https://doi.org/10.1016/S0306-4573(00)00008-X

Ibrahim RA, Ewees AA, Oliva D, Elaziz MA, Lu S (2018) Improved salp swarm algorithm based on particle swarm optimization for feature selection. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-1031-9

Kabary IA, Schuldt H (2014) Enhancing sketch-based sport video retrieval by suggesting relevant motion paths. In: proceedings of the 37th international ACM SIGIR conference on research and development in information retrieval, pp 1227– 1230. https://doi.org/10.1145/2600428.2609551

Khennak I, Drias H (2016) Bat algorithm for efficient query expansion: application to MEDLINE. In: proceedings of the 4th World conference on information systems and technologies, pp 113–122. https://doi.org/10.1007/978-3-319-31232-3_11

Khennak I, Drias H (2017) An accelerated PSO for query expansion in web information retrieval: application to medical dataset. Appl Intell 47(3): 793–808. https://doi.org/10.1007/s10489-017-0924-1

Kim BM, Kim JY, Kim J (2001) Query term expansion and reweighting using term co-occurrence similarity and fuzzy inference. In: proceedings of the joint ninth IFSA world congress and 20th NAFIPS international conference, Vancouver, 2, pp 715–720. https://doi.org/10.1109/NAFIPS.2001.944690

Latiri C, Haddad H, Hamrouni T (2012) Towards an effective automatic query expansion process using an association rule mining approach. J Intell Inf Syst 39(1):209–247. https://doi.org/10.1007/s10844-011-0189-9

Lee Y, Bang S (2018) Improved image retrieval and classification with combined invariant features and color descriptor. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-0817-0

Leturia I, Gurrutxaga A, Areta N, Alegria I, Ezeiza A (2013) Morphological query expansion and language-filtering words for improving Basque web retrieval. Lang Resour Evaluat 47(2):425–448. https://doi.org/10.1007/s10579-012-9208-x

Li Q, Tian M, Liu J, Sun J (2016) An implicit relevance feedback method for CBIR with real-time eye tracking. Multimed Tools Appl 75(5):2595–2611. https://doi.org/10.1007/s11042-015-2873-1

Lin HC, Wang LH, Chen SM (2006) Query expansion for document retrieval based on fuzzy rules and user relevance feedback techniques. Expert Syst Appl 31:397–405. https://doi.org/10.1016/j.eswa.2005.09.078

Lin HC, Wang LH, Chen SM (2008) A new query expansion method for document retrieval by mining additional query terms. Inf Manag Sci 19(1):17–30

Nowacka K, Zadrozny S, Kacprzyk J (2008) A new fuzzy logic based information retrieval model. In: proceeding of IPMU'08, pp 1749–1756. http://www.gimac.uma.es/ipmu08/proceedings/papers/234-Zadrozni.pdf

Oh HS, Jung Y (2015) Cluster-based query expansion using external collections in medical information retrieval. J Biomed Inform 58:70–79. https://doi.org/10.1016/j.jbi.2015.09.017

Park JH, Croft WB (2015) Using key concepts in a translation model for retrieval. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. ACM New York, pp 927–930. https://doi.org/10.1145/2766462.2767768

Qian B, Wang Q, Hu R, Zhou Z, Yu C, Zhou Z (2017) An effective soft computing technology based on belief-rule-base and particle swarm optimization for tipping paper permeability measurement. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-017-0667-1

Ramalingam G, Dhandapani S (2014) A novel adaptive cuckoo search for optimal query plan generation. Sci World J 2014:1–7. https://doi.org/10.1155/2014/727658

Rijsbergen C (1979) Information Retrieval, 2 ed., Butterworth, Houston

Rivas A, Iglesias E, Borrajo L (2014) Study of query expansion techniques and their application in the biomedical information retrieval. Sci World J 2014:1–10. https://doi.org/10.1155/2014/132158

Robertson S, Jones S (1976) Relevance weighting of search terms. J Am Soc Inf Sci 27:129–145. https://doi.org/10.1002/asi.4630270302

Robertson A, Willet P (1996) An upperbound to the performance for ranked-output searching: optimal weighting of query terms using a genetic algorithm. J Doc 52(4):405–420. https://doi.org/10.1108/eb026973

Saeedeh MD, Siddiqi J, Zadeh Y, Rahman F (2012) Adaptive information retrieval system via modelling user behavior. J Ambient Intell Humaniz Comput 5(1):105–110. https://doi.org/10.1007/s12652-012-0138-7

Sanchez E, Miyano H, Brachet J (1995) Optimization of fuzzy queries with genetic algorithms. In: proceedings of Applications to a data base of patents in biomedical engineering, VI IFSA Congress, Sao-Paulo, Brazil, pp 293–296

Saraiva PC, Cavalcanti JM, de Moura ES, Gon˙calves MA, Torres RDS (2016) A multimodal query expansion based on genetic programming for visually-oriented e-commerce applications. Inf Process Manag 52(5):783–800. https://doi.org/10.1016/j.ipm.2016.03.001

Singh J, Sharan A (2015) Relevance Feedback Based Query Expansion Model Using Borda Count and Semantic Similarity Approach. Comput Intell Neurosci 2015(568197):1–13. https://doi.org/10.1155/2015/568197

Singh J, Sharan A (2016) Relevance Feedback-based Query Expansion Model using Ranks Combining and Word2Vec Approach. Journal of IETE Journal of Research 62(5):591–604. https://doi.org/10.1080/03772063.2015.1136575

Singh J, Sharan A (2017a) A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach. Journal Neural Computing Applications 28(9):2557–2580. https://doi.org/10.1007/s00521-016-2207-x

Singh J, Sharan A (2018) Rank fusion and semantic genetic notion based automatic query expansion model. Swarm Evolut Comput 38: 295–308. https://doi.org/10.1016/j.swevo.2017.09.007

Singh J, Sharan A, Saini M (2017b) Term co-occurrence and context window-based combined approach for query expansion with the semantic notion of terms. International Journal of Web Science 3(1):32–57. https://doi.org/10.1504/IJWS.2017.088677

Suganthan P (1999) Particle swarm optimizer with neighborhood operator. In: Proceedings of IEEE International Conference on Evolutionary Computation, 3, pp 1958–1962. https://doi.org/10.1109/CEC.1999.785514

Tayal DK, Sabharwal S, Jain A, Mittal K (2012) Intelligent query expansion for the queries including numerical terms. In: Proceedings of National Conference on Communication Technologies and its impact on Next Generation Computing CTNGC 2012, pp 35–39

Tuba M, Subotic M, Stanarevic N (2011) Modified cuckoo search algorithm for unconstrained optimization problems. In: Proceedings of the 5th European conference on European computing conference, pp 263–268

Vechtomova O, Robertson S, Jones S (2003) Query expansion with long-span collocates. Inf Retrieval 6(2):251–273. https://doi.org/10.1023/A:1023936321956

Wasilewski P (2011) Query Expansion by Semantic Modeling of Information Need. In: proceedings of international Workshop CS and P

Wu H, Li J, Kang Y (2018) Exploring noise control strategies for UMLS–based query expansion in health and biomedical information retrieval. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-018-0836-x

Xu J, Croft WB (1996) Query Expansion using Local and Global Document Analysis. ACM SIGIR conference on research and development in information retrieval, pp 4–11. https://doi.org/10.1145/243199.243202

Yang J, Korfhage R (1994) Query modifications using genetic algorithms in vector space models. International Journal of Expert Systems 7(2):165–191

Zhang C, Yang Y, Du Z, Ma C (2016) Particle swarm optimization algorithm based on ontology model to support cloud computing applications. J Ambient Intell Humaniz Comput 7(5):633–638. https://doi.org/10.1007/s12652-015-0262-2

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.