Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 10, No. 1, 2022, 15-24.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Question Classification in Question Answering System using Combination of Ensemble Classification and Feature Selection

Shahram Golzari[1*], Farideh Sanei[2], Mahmoud Reza Saybani [3], Abbas Harifi[1] and Mojtaba Basir[2]

1. Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, Iran and Deep Learning Research Group, University of Hormozgan, Bandar Abbas, Iran.
2. Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, Iran.
3. Markaz-e Elmi Karbordi Bandar Abbas 1, University of Applied Science and Technology, Farahani Boulevard, Bandar Abbas, 79199_33153, Iran.

Article Info

Article History:
Received 02 September 2020
Revised 07 July 2021
Accepted 13 August 2021

DOI:10.22044/JADM.2021.10016.2142

Keywords:
Question Answering System, Question Classification, Feature Selection, Ensemble Classification, Gravitational Search Algorithm, Differential Evolution Algorithm.

**Corresponding author: golzari@hormozgan.ac.ir (Sh. Golzari).*

Abstract

A Question Answering System (QAS) is a special form of information retrieval that consists of three parts: question processing, information retrieval, and answer selection. Determining the type of question is the most important part of QAS as it affects the other following parts. In this work, we use the effective features and ensemble classification in order to improve the QAS performance by increasing the accuracy of question type identification. We use the Gravitational Search Algorithm (GSA) in order to select the features and perform the ensemble classification. The proposed system is extensively tested on different datasets using four types of experiments: (1) neither feature selection nor ensemble classification, (2) feature selection without ensemble classification, (3) ensemble classification without feature selection, and (4) feature selection with ensemble classification. These four kinds of experiments are carried out under the Differential Evolution (DE) algorithm and GSA. The experimental results obtained show that the proposed method outperforms compared to the state-of-the-art methods in the previous research works.

1. Introduction

With the increase in the amount of information on the web, the search engines are required to be more intelligent than ever before. In many applications, the user only requires a specific part of the information instead of a lot of documents. Therefore, it is often preferable to provide a short and brief answer for the user. The goal of a question answering system (QAS) is to provide an accurate information in response to a question. Similar to a human, a QAS should be able to answer a question written in the natural language [1].

QAS can be considered as another step of data retrieval that allows the users to ask questions in the natural language and receive short answers [2]. QAS is more complex than other types of Information Retrieval (IR) methods such as document retrieval due to the need for the Natural Language Processing (NLP) techniques [3]. This system can return the correct answer in a relatively short time [4]. Therefore, different studies have been performed presented for this purpose that have demonstrated the efficiency of QAS [4, 5, and 6].

When a user requests for information from the IR system using a question, instead of keywords, the answered results may be different from the intended purpose. In order to address this challenge, QASs have been developed, which can provide appropriate answers to the natural language questions. The main parts of these

systems are question processing and the identification of question type [6].
A QAS is actually a type of IR system, where the responses are processed and evaluated in order to provide the user with short answers rather than returning a large set of responses to the user that is common in the IR systems. The short answers allow the user to receive the expected answers faster. The search space of the questions is a collection of documents that can be stored either in a database or on the information networks such as the Internet [7].
The question analysis, search, and answer selection are three important parts of a QAS. Question classification and formulation are two important components of question processing. In order to extract the answer from a large number of documents and texts, the system must first know what it is looking for. This task is performed by assigning a question to certain predefined classes. Question classification is an important part that has a direct impact on the efficiency of QAS. Question formulation aims to provide answers to the natural language questions. The IR section retrieves documents related to the user's question, in which the question is converted into a specific form, and the relevant documents are extracted from the available sources. Answer processing consists of two main parts: answer extraction and answer validation. In the extraction section, the response extraction algorithms are performed in order to extract response candidates from those documents recovered by a search engine in the response extraction unit. Once the response candidates are retrieved, they are validated using the filtering and ranking techniques [8].
The main challenge of QAS is the low performance of classification, particularly question classification [9, 10, and 11]. In this work, ensemble classification is used in order to increase the efficiency of classification. Aburomman et al. [12] have combined the two algorithms of k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) based on the Particle Swarm Optimization (PSO) algorithm. The authors also used LUS to adjust the parameters of the PSO algorithm. Syarif et al. [13] have used the bagging, boosting, and stacking techniques in order to increase accuracy and reduce the rate of positive errors. The four classification techniques of Naïve Bayes, J48, JRIP rule inference, and nearest neighbor were used in all the three methods. Bahri et al. [14] have used the Greedy-Boost method for composite classification. They compared their proposed system with the AdaBoost and C4.5 decision trees based on the accuracy and recall measures. The ensemble classification was used in the text processing phase [15, 16], in which the evolutionary techniques were used to identify the name entities.
Computer scientists consider Feature selection, as a technique to improve the performance of the classification methods. Due to enormous impact that matrix’s dimensions have on the performance of processing on it, applying reduction in the number of features through choosing the best subset of all features will affect the performance of the algorithms [17]. Therefore, selecting a set of appropriate features for building strong learning models is a widely-used technique in machine learning concerning problem optimization. Feature selection is also known as variable selection, feature reduction, and selection of variables set. Feature selection problem can be addressed with the help of a number of single-objective optimization methods such as genetic algorithms. A single classification alone cannot work well for all types of classes, and it optimizes only one aspect of quality. Ensemble classification improves the performance and quality measures by assigning the right weight to each classifier. Majority voting and weighting are the most significant techniques in ensemble classification to combine the output of several classifiers [18]. The problems of feature selection and ensemble classification can be considered as the optimization problems aiming to look for the optimal set of answers. Evolutionary algorithms are a state-of-the-art and efficient strategy for finding near-optimal solutions. These algorithms encode the problem in terms of solution(s) to be evolved to improve its quality. [18].
Gravitational Search Algorithm (GSA) is an optimization method inspired by the Newton’s law of universal gravitation. According to this law, each object identifies the location of other objects through the law of gravity between the planets. Therefore, gravity can be used as a tool for information exchange. The position of each agent presents a candidate solution for the problem, while the agent's mass is assigned using an objective function [19]. In the case of GSA, path planning is calculated based on the force received from other planets. Also GSA is memory less so that only the current position of the planets contributes to the process of update. In this algorithm, the gravitational force is considered suitable according to its fitness value.
In this work we use the GSA-based feature selection and ensemble classification in order to

improve classification accuracy when finding appropriate answers to the questions in the question answering systems. Consequently, GSA is used to select appropriate features and during performing classification. The simulation results confirm that the proposed method increases the classification accuracy in comparison to the lack of those methods.

The rest of this paper is organized as what follows. Section 2 provides the related works about question classification, feature selection, and ensemble classification. In Section 3, a novel QAS is introduced by applying feature selection and ensemble classification. Section 4 presents the experimental results and discussions. Finally, Section 5 concludes the paper.

2. Related Works

In the following, the studies presented in the sections of question classification, feature selection and ensemble classification are described.

2.1. Question Classification

The first manual system was introduced by Hermjakob [4] in 2001. A manual rules-based QAS was designed in order to identify the type of response. Although the rules may be very precise, they are time-consuming, tedious, and non-upgradeable. On the other hand, automatic classification has been developed to a variety of new questions and to classify the questions with a good accuracy. In these methods, the machine learning algorithms and language modeling are used. Hacioglu and Ward [20] have used SVM to classify the question. They used named entity and n-grams for the feature extraction, and obtained an accuracy of 82%. In another model, they used just n-grams as a feature extraction method and SVM as the classifier. They were able to achieve an accuracy of 80.2%. Zhang, D. and Lee, W.S. [5] have used Bag-of-ngrams to extract the features, and have used the following classifiers to classify the data: neural network, naïve bayes, decision trees, SNoW, and SVM. They obtained the accuracies of 79.8%, 83.2%, 84.2%, 86.6%, and 87.4%, respectively.

Li and Roth [21] have used a lexical network to classify the question. They showed that the use of the lexical network can yield a better and more acceptable performance compared to the syntactic features. They used the accuracy criterion in order to evaluate the system that achieved an accuracy of 84.2%. Yahya and Osman [22] have used the Bag-of-words model instead of the lexical and semantic features. They used an SVM algorithm in order to classify the questions. Different cores were considered for the SVM, and their results demonstrated that the linear core achieved the best performance.

Wang et al. [23] have used a word sequence method to classify the question using the SVM algorithm. The sequence of words was used to distinguish between the Chinese letters. Since there are many similarities between the Chinese words, it is not possible to easily distinguish between the words. The authors adopted the HowNet semantic lexical network. Sixty hundred Chinese questions were used for classification with two types of classes, including coarse-grained with six classes and fine-grained with 59 classes. The results obtained showed that the proposed method could achieve an acceptable performance using the unlimited domain. The accuracy rates for the coarse-grained and fine-grained classification were 91% and 83.67%, respectively. Blunsom et al. [24] have used lexical and syntactic entity for feature extraction and Max Entropy for classification. They obtained an accuracy of 86%. Ray et al. [25] have used words, semantic information, and named entity in order to extract the features, and obtained an accuracy of 91%.

Li et al. [26] have used the SNoW classifier, and for extracting the features, they used Words, POS, named entities, chunks, head chunks, and semantically related words. They were able to achieve an accuracy of 91%. Huang et al. [27] have used head word, wh-words, and semantic information to extract the features, and for the classification, they used SVM and Maximum Entropy. The performance of both classifiers was almost equal. They achieved a classification accuracy of 89%. Mohd and Hashmy [28] have proposed a knowledge-based semantic kernel that uses WordNet to compute semantic relatedness between the sentences and to overcome the bag of words drawbacks. The experiments using the UIUC dataset show that the SVM model using the SR Kernel achieved an accuracy of 91.9%.

2.2. Feature Selection

Anjomshoaa et al. [29] have used progressive selection and genetic algorithm for an effective feature selection in email spam. Once the preprocessing operation is performed, the data is fed into the feature selection algorithms. Each feature selection algorithm consists of four steps: production function, evaluation function, condition of termination, and credit determination function. Once the steps are performed, the selected features are fed into the three algorithms

of k-NN, SVM, and multi-layer neural network. Then the result accuracy of each class was calculated and used in order to evaluate the system. The results obtained were 93.79%, 97.02%, and 97.67% for k-NN, SVM, and multi-layer neural network, respectively.
Ganji et al. [30] have used an imperialist competitive algorithm to select the effective features. In addition to selecting the features, the authors optimized the SVM parameters. In the imperialist competitive algorithm, the countries are the answers' symbol to the problem, and the answers are improved during the algorithm being performed. Each time the imperialist competitive algorithm is repeated, the C and Gamma parameters are set for SVM alongside selecting the best features. The radial kernel function was used as the kernel of SVM. In this work, accuracy, precision, and recall criteria were used in order to evaluate the system. The values obtained from these criteria were 94.5%, 91.15%, and 97.7%, respectively.

2.3. Ensemble Classification

In this section, the studies related to ensemble classification are described. Ghanbari et al. [31] combined the neural network model and k-NN using a threshold. In their method, the feature extracted from a new sample is fed to the k-NN model. If the output of k-NN is greater than a threshold, the class of the new sample is determined through the nearest neighbor classification; otherwise, the sample is classified using a neural network. The experimental results showed that the use of threshold was able to increase the efficiency of the method compared to the use of only one classifier.
Kumar Sikdar et al. [32] have used a multi-objective Differential Evolution (DE) algorithm for feature selection and ensemble classification. In the feature selection phase, the number of features and the F-measure were considered as two objective functions. In this work CRF was used for classification. Once the optimal population is generated, it is considered as the base classifier so that these classifiers are used by the DE algorithm for ensemble classification. In the ensemble classification phase, the two criteria of accuracy and recall were used. In order to find the final class, the F-measure of each classifier is first multiplied by the weight that is assigned to that class by the DE algorithm. Then the amount of results for similar classes is accumulated. Finally, the class with the highest weight is considered as the final class. The study used accuracy, recall, and F-measure in order to evaluate the system. The results obtained from these measures were 85.66%, 90.67%, and 88%, respectively.
According to [32], in which the multi-objective Differential Evolution (DE) algorithm was performed for feature selection and ensemble classification, we knew that the combination of Feature Selection and Ensemble Classification resulted in a proper performance. In addition, GSA has advantages over the DE algorithm. The DE algorithm has some weaknesses as it depends strongly on differential vectors for producing a new population, and the construction of these vectors requires a lot of time and accuracy. On the other hand, GSA as a strong evolutionary algorithm uses the gravitational force between the objects to produce a new route. Thus, GSA is more efficient than the DE algorithm, and we intend to use this quality as an added advantage to our study. We hypothesize that due to the superiority of the GSA algorithm compared to the DE algorithm, we would witness performance enhancement if we use the GSA algorithm for feature selection and ensemble classification, while carrying out question classification in the question answering system.

3. Proposed Method

In this section, a new classification method is presented. The proposed method consists of two stages: feature selection and ensemble classification. Figure 1 shows a general framework of the proposed method.

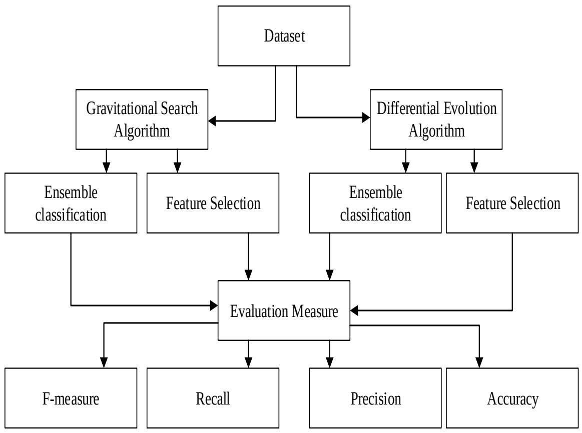

Figure 1. A general framework of the research plan.

As shown in Figure 1, a two-objective GSA and a two-objective DE are used for feature selection and ensemble classification. Also the evaluation measures of accuracy, precision, recall, and f-measure are used to compare the proposed method with the other methods. In the feature selection stage, the number of features and the F-measure are used as the objectives. Decision tree and SVM are used for the ensemble classification. Since the DE algorithm relies on differential vectors for producing a new population, the construction of these vectors requires a lot of time and accuracy. On the other hand, the GSA algorithm uses the gravitational force between the objects to produce a new route; therefore, it is more efficient than the DE algorithm. This work takes the advantages of GSA for feature selection and ensemble classification. Different phases of the proposed method are described in the following sub-sections.

3.1. Feature Selection based on GSA

The GSA used in the method is a continuous version of the algorithm. In order to perform feature selection, discretization should be applied. Equation (1) is used to discretize the GSA algorithm:

$$X = \begin{cases} 1(x \geq 0.5) \\ 0(otherwise) \end{cases} \quad (1)$$

In the above equation, X is the position of objects. Figure 2 shows the framework of feature selection using GSA.

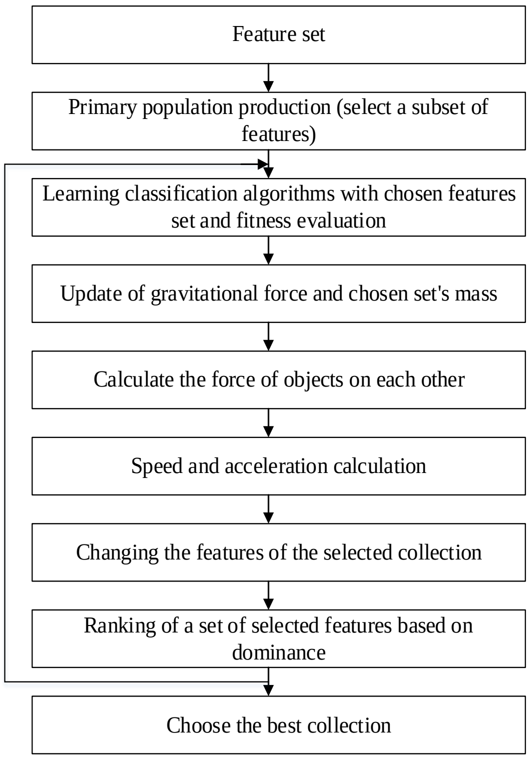

Figure 2. Feature selection steps.

From Figure 2, first, the initial population is created in the form of vectors with random numbers between 0 and 1. Then the discretization operation should be performed on the gene values; at this stage, the selected features are determined. Then each member of the population should be evaluated based on the objectives of the problem, and the closer it is to the desired goal, the better fitness belongs to that member. At this stage, finding the appropriate fitness function, which can be single or multi-objective, is an essential task. Due to the use of several objective functions, calculation of the multi-objective fitness function is more complex, and at the same time, more efficient than that of a single-objective function. For the purpose of efficiency enhancement, the multi-objective GSA is used for optimizing the objectives of this research work. The following objectives are used:

- F-measure: GSA tries to maximize this objective.
- Number of selected features: GSA tries to minimize this goal.

As mentioned earlier, in this research work the two classification algorithms of decision tree and SVM are used. After training these classifiers using the training data, the F-measure of each model is calculated the using validation dataset. Then the classifier with the best F-measure is selected. The second objective function is equal to the number of features selected by GSA. Once the fitness of each member is calculated, the parameters of GSA such as the gravitational force of objects, velocity, acceleration, and position of objects are updated. Then the elitist operation is performed. In order to perform the elite operations, the members of the population should be organized. Since we use a multi-objective GSA, the elitist operation is performed with two changes in the single-objective GSA. These changes include the non-dominated sorting of the population members based on the superiority criteria and diversity of answers in the population.

3.2. Ensemble Classification based on GSA

Figure 3 shows the framework of the GSA-based ensemble classification. In the GSA-based ensemble classification, first, the number of classifiers for performing the ensemble classification is multiplied by the number of classes in the dataset. The resulting value determines the number of variables in the problem. The initial population, which consists of the weights assigned to each class of each classifier, is created. The purpose of this idea is to determine the appropriate weight of classifiers in accordance with the detection accuracy of each class. In order to obtain the fitness value of each member, the class of each new sample is first specified using both classifiers distinctly.

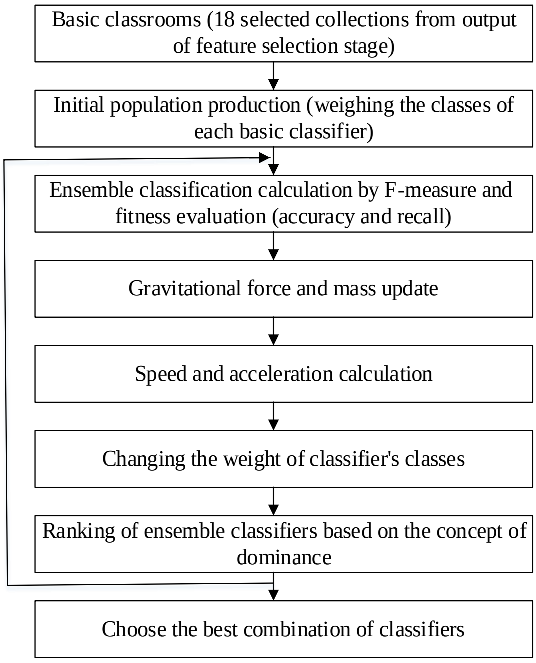

Figure 3. Ensemble classification steps.

Then, the determined weight of member is multiplied by the F-measure of that classifier. Finally, the values are accumulated for similar classes. The class with the highest value is assigned to the new sample. Table 1 presents an example to provide more description. Suppose that there are three classifiers and two classes; the number of problem variables is equal to six (i.e., 3 * 2).

Table 1: An example of values for GSA-based ensemble classification.

Classifier	Weight for first class	Weight for second class	F-measure (%)
First	0.9	0.3	0.98
Second	0.7	0.2	0.96
Third	0.8	0.3	0.90

The classifiers are derived from the feature selection phase, and the weights are assigned to the classes of each classifier through GSA. As mentioned in Section 3.1, the training and validation datasets are used for feature selection. The validation dataset is also used at this stage for the ensemble classification. Now we intend to identify the class of a new sample. Assume that the first classifier detects the class of the sample as the first class, and both the second and third classifiers categorize the sample into the second class. According to Table 1 and above relation (i.e. multiplication of F-measure by the weight of class), the steps are as what follows.

The first classifier (first class)
$0.98 \times 0.9 = 0.882$

The second classifier (second class)
$0.96 \times 0.2 = 0.192$

The third classifier (second class)
$0.9 \times 0.3 = 0.27$

The values of the same classes are then accumulated. Since there are two classifiers that have been classified in the second class, the two values are added, and the result is equal to:
0.27 + 0.192 = 0.462

We now compare the values of the two classes, which are 0.882 and 0.462 for the first and second classes, respectively. Since the value of the first class is greater than the second one, the first class is considered as the class of the desired sample. After performing the ensemble classification, the fitness value of the population is calculated using the two-objective GSA with objectives of maximizing accuracy and recall. Once the fitness value of the initial population is calculated, the parameters of GSA are updated in such a way that the gravitational force of the objects, velocity, and acceleration are updated in order. The weights assigned to the classes are changed based on the updated velocity. At the end, the elitist operation is performed according to the non-dominated sorting and crowding distance. The number of iterations specified by the user (i.e. trial and error) is considered as the stopping criteria in GSA. When GSA is completed, the best ensemble classification is selected. The F-measure is used to select the best combination. In other words, a combination with a higher F-measure is considered as the best combination.

3.3. Datasets for QAS

This section describes the UIUC dataset, which consists of 5,452 questions for training and 500 questions for evaluation. This dataset consists of four sources [33]: (1) USC English Questions, (2) Questions belonging to TREC 8 and TREC 9, (3) Questions that are presented manually, and (4) TREC 10 questions that are used for the testing phase.

3.4. Question Classification using Proposed Method

In order to classify the question, the feature vector is first extracted according to the technique proposed in [34], and the dataset is divided into three categories including the training data, validation data, and testing data. The amount of data assigned to each category is 50%, 20%, and 30%, respectively. After dividing the data, the training and validation datasets are used to extract the effective features by the GSA-based feature selection algorithm. Afterwards, the features are given to the ensemble classification algorithm that uses GSA for optimization. In our method, the CART decision tree and the linear function of SVM are used as the basic classifiers. After finding the appropriate combination of classes, the test dataset is used to evaluate the system. Also in order to adjust the parameters of GSA, the values of 50, 20, and 100 are defined for the maximum iteration, the gravitational coefficient, and the initial gravitational coefficient, respectively.

3.5. Evaluation Parameters

In order to evaluate the efficiency of the proposed method, the measures of accuracy, recall, precision, and the F-measure are used. The accuracy is obtained using (2).

$$Accuracy = \frac{m}{n} \times 100 \tag{2}$$

Where m is the number of questions that the method can classify correctly and n is the total number of questions [35]. The recall, precision, and F-measure are also calculated using the following equations [36]:

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{3}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{4}$$

$$F - Measure = \frac{2 \times \mathrm{Pr}ecision \times \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} \tag{5}$$

In the above equations, True Positive (TP) is the number of questions that are correctly identified, False Positive (FP) represents the number of negative samples classified as positive, True Negative (TN) refers to the number of negative questions that are correctly classified as negative, and False Negative (FN) is the number of positive questions that are mistakenly classified as negative.

4. Experimental Results

In this section, the results obtained by four approaches are shown. These approaches include: (1) neither feature selection nor ensemble classification, (2) feature selection without ensemble classification, (3) ensemble classification without feature selection, and (4) feature selection with ensemble classification. All experiments were carried out on the UIUC datasets.

4.1. Neither Feature Selection nor Ensemble Classification

Table 2 reports the results obtained by the experiment of neither feature selection nor ensemble classification.

Table 2. Neither feature selection nor ensemble classification.

Classifier	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
Decision Tree (DT)	77.20	76.40	80.16	77.55
SVM	87.40	90.10	88.93	98.51

This table has almost minimum values among other tables, which will be reported in this section, from which we can conclude that the proposed methods have an appropriate impact on the question classification.

4.2. Feature Selection without Ensemble Classification

Since both the GSA and the DE algorithms can be performed for feature selection, experiments were carried out for both algorithms. Table 3 shows the experimental results of feature selection without the ensemble classification strategy when applying DE and when applying GSA algorithms.

The values in Table 3 shows that feature selection has improved the performance in most of proposed comparative scales as these measures are increased in their values. This is due to the role that selecting better features plays in classification improvement. Moreover, GSA has shown relatively better results than the DE algorithm.

Table3. Feature selection without ensemble classification using DE and GSA algorithms.

Method	Features No.	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DE + DT	119-227	79.2	83.21	81.89	82.55
DE + SVM	162-227	91	92.74	91.85	92.29
GSA + DT	136-227	79.2	83.11	81.45	81.78
GSA + SVM	142-227	91	93.06	92.11	92.48

4.3. Ensemble Classification without Feature Selection

Like the experiments carried out in the previous approach (described in Section 4.2), we used both DE and GSA in this experiment in order to evaluate the efficiency of the proposed method. The evaluation results of ensemble classification without feature selection when applying DE and when applying the GSA algorithms are reported in Table 4.

Table 4. Ensemble classification without feature selection using DE and GSA algorithms.

Algorithm	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
DE	91.4	89.13	92.45	90.76
GSA	91.8	91.69	92.81	92.25

In this method, we witnessed more significant improvement in the evaluation parameters than Table 2, especially when using the GSA algorithm.

Compared with Table 3, applying the ensemble strategy improved the results significantly in comparison to using only decision tree; also compared to SVM, the results obtained relatively improved. This is because ensemble classification uses the strong points of involved classifiers for improving performance in comparison with using each classifier alone. In this table, again, there are better results for the GSA algorithm than the DE algorithm.

4.4. Feature Selection with Ensemble Classification

Since the feature selection with ensemble classification approach uses evolutionary algorithms (i.e. GSA and DE) for both the feature selection and ensemble classification, experiments were performed using both the GSA and the DE algorithms. The results obtained for these algorithms are reported in Table 5.

Table 5. Feature selection with ensemble classification approach using DE and GSA algorithms.

Algorithm	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
DE	89.80	91.29	90.79	91.4
GSA	91.80	93.06	92.11	92.58

Compared to Table 3, the results obtained especially with the GSA algorithm, are significantly better than decision tree and relatively better than SVM. The same is true for the comparisons with Table 2. Also it has better results than Table 4. This strategy has led to, on average, better results than the three previous tables, as it takes the advantages of selecting suitable features as well as using a proper combination of two strong classifications. From Table 5, it can be concluded that GSA achieved more acceptable efficiency than the DE algorithm. The reason can stem from the fact that GSA can produce a high-quality new generation compared to the DE algorithm.

4.5. Comparison with Previous Studies

In this section, we make a comparison of the proposed method with other approaches regarding the question classification.

Table 6. Comparison of proposed method with other methods in terms of accuracy.

Dataset	Research	Accuracy (%)
UIUC Question and Answering	Zhang et al. [5]	87.4
	Hacioglu et al. [20]	82.0
	Li and Roth [21]	84.2
	Yahya et al. [22]	87.4
	Wang et al. [23]	91
	Blunsom et al. [24]	86.0
	Ray et al. [25]	91.0
	Li [26]	91.0
	Huang et al. [27]	89.0
	Mohd and Hashmy.[28]	91.9
	Proposed method	91.8

Table 6 shows the experimental results in terms of accuracy. According to Table 6, it can be seen that the proposed method, in most cases, has a more acceptable efficiency due to the use of feature selection and combined classification approaches simultaneously. This is because appropriate features are selected and the capabilities of classifiers are used properly.

5. Conclusions

In this work, we proposed a new method for QAS using the feature selection and ensemble classification with the help of GSA. The proposed method aims to find the question class of the user. The method tries to provide an accurate question classification to affect positively the other stages of QAS. In this method, the lexical and syntactic features of questions are first extracted to identify the question class, and the feature vector is constructed using the extracted features. Then, GSA is used to select the features that have a significant impact on data classification. Since each classification algorithm has special strengths and weaknesses, the proposed method uses the strengths of the classification algorithms to enhance efficiency and to reduce the weaknesses. In this regard, the ensemble classification was applied with the use of the decision tree and SVM.

References

[1] M. Ramprasath and S. Hariharan, "A survey on question answering system," *International Journal of Research and Reviews in Information Sciences*, vol. 2, no. 1, pp. 171-179, 2012.

[2] P. Shafiei and M. Parsa, "Designing question answering system using genetic algorithm," in *The 4th Data Mining Conference of Sharif university of Tehran, Tehran,* 2010 (in Persian).

[3] M. R. Kangavari, S. Ghandchi, and M. Golpour, "A new model for question answering systems," *World Academy of Science, Engineering and Technology*, vol. 42, pp. 506-513, 2008.

[4] U. Hermjakob, "Parsing and question classification for question answering," in *Proceedings of the ACL 2001 workshop on open-domain question answering*, 2001, pp. 1-6.

[5] D. Zhang and W. S. Lee, "Question classification using support vector machines," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, 2003, pp. 26-32.

[6] A. Tahri and O. Tibermacine, "DBPedia based factoid question answering system," *International Journal of Web and Semantic Technology*, vol. 4, no. 3, pp. 23, 2013.

[7] M. Shamsfard and M. A. Yarmohammadi, "A semantic approach to extract the final answer in SBUQA question answering system," *International Journal of Digital Content Technology and its Applications*, vol. 4, no. 7, pp. 165-176, 2010.

[8] O. Kolomiyets and M.F. Moens, "A survey on question answering technology from an information retrieval perspective," *Information Sciences*, vol. 181, no. 24, pp. 5412-5434, 2011.

[9] A. Mollaei, S. Rahati-Quchani, and A. Estaji, "Persian Question Classification Accuracy Improvement by Using Question Informer based on Conditional Random Fields Model," in *the 2nd International eConference on Computer and Knowledge Engineering (ICCKE)*, 2012, pp. 295-300.

[10] B. Liu, Z. Hao, X. Yang, and X. Lin, "Chinese question classification with support vector machine," *IJCSNS International Journal of Computer Science and Network Security*, vol. 6, pp. 231-240, 2006.

[11] W. Zhang, J. Chen, and Y. Niu, "Research on Chinese question classification based on hownet and dependency parsing," in *2011 3rd International Workshop on Intelligent Systems and Applications*, 2011, pp. 1-4.

[12] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360-372, 2016.

[13] I. Syarif, E. Zaluska, A. Prugel-Bennett, and G. Wills, "Application of bagging, boosting and stacking to intrusion detection," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 2012, pp. 593-602.

[14] E. Bahri, N. Harbi, and H. N. Huu, "Approach based ensemble methods for better and faster intrusion detection," in *Computational Intelligence in Security for Information Systems*, 2011, pp. 17-24.

[15] A. Ekbal and S. Saha, "Weighted vote based classifier ensemble selection using genetic algorithm for named entity recognition," in *International Conference on Application of Natural Language to Information Systems*, 2010, pp. 256-267.

[16] A. Ekbal and S. Saha, "Weighted vote-based classifier ensemble for named entity recognition: A genetic algorithm-based approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 10, no. 2, pp 1-37, 2011.

[17] Z. Shojaee, S. A. Shahzadeh Fazeli, E. Abbasi, and F. Adibnia. "Feature Selection based on Particle Swarm Optimization and Mutual Information." *Journal of AI and Data Mining*, vol. 9, no. 1, pp.39-44, 2021.

[18] K. H. Nabavi and A. Kabir, "Ensemble classification: creating diversity and combination rules," *Scientific Journal of Iran Computer Association*, vol. 3, no. 3, 2005 (in Persian).

[19] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information sciences*, vol. 179, no. 13, pp. 2232-2248, 2009.

[20] K. Hacioglu and W. Ward, "Question classification with support vector machines and error correcting codes," in *Companion Volume of the Proceedings of HLT-NAACL 2003-Short Papers*, 2003, pp. 28-30.

[21] X. Li and D. Roth, "Learning question classifiers: the role of semantic information," *Natural Language Engineering*, vol. 12, no. 3, pp. 229-249, 2006.

[22] A. A. Yahya and A. Osman, "Automatic classification of questions into Bloom's cognitive levels using support vector machines," in *The International Arab Conference on Information Technology, Naif*

Arab University for Security Science (NAUSS), 2011, pp. 1-6.

[23] L. Wang, H. Zhang, D. Wang, and J. Huang, "Chinese question classification based on semantic gram and SVM," in *2009 International Forum on Computer Science-Technology and Applications,* 2009,
vol. 1, pp. 432-435.

[24] P. Blunsom, K. Koicik, J. R. Curran, "Question classification with log-linear models," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval,* 2006, pp. 615-6616.

[25] S. K. Ray, S. Singh, and B. P. Joshi, "A semantic approach for question classification using WorldNet and Wikipedia," *Pattern Recognition letters,* vol. 31, no. 13, pp. 1935-1943, 2010.

[26] X. Li and D. Roth, "Learning question classifiers," in *COLING 2002: The 19th International Conference on Computational Linguistics,* 2002.

[27] Z. Huang, M. Thint, and Z. Qin, "Question classification using head words and their hypernyms," in *Proceedings of the 2008 Conference on empirical methods in natural language processing,* 2008, pp. 927-936.

[28] M. Mohd and R. Hashmy, "Question classification using a knowledge-based semantic kernel," in *Soft Computing: Theories and Applications,* 2018: springer, pp. 599–606.

[29] A. Anjomshoaa, H. Nezamabadi-Pour, and H. Sanatnama, "Filtering of spam emails from valid emails using two methods of feature selection," in *The First Iranian Conference on Pattern Recognition and Image Analysis, Birjand,* 2012 (in Persian).

[30] M. Ganji, A. Yousef, N. Charkari, and M. Saniee-Abadeh, "Predicting protein interaction based on colonial competition algorithm," in *7th International Conference on Information Technology and Knowledge,* no. 10, 2015 (in Persian).

[31] Y. Ghanbari, A. Harifi, and S. Ghanbari, "Recognition of Persian Manuscript Numbers Using Gabor Filter, Principal Component Analysis and Ensemble Classification," in *2nd International Conference on Electrical Engineering,* no. 10, 2015 (in Persian).

[32] U. K. Sikdar, A. Ekbal, and S. Saha, "MODE: Multiobjective Differential Evolution for feature selection and classifier ensemble," *Soft Computing,* vol. 19, no. 12, pp. 3529-3549, 2015.

[33] B. Loni, "Enhanced Question Classification with Optimal Combination of Features", M.S. thesis, Dept. Media and Knowledge Eng., Delft Univ., Delft, NL, 2011.

[34] Y. Ghanbari, Sh. Golzari, and S. Doraisamy, "Feature Selection for questions classification using cuckoo evolutionary algorithm," in *the Proceedings of Second International Conference on Technology, Communication and Knowledge, ICTCK2015,* 2015 (in Persian).

[35] S. Bakhtiar-Ghale-Taki, M. H. Dezfoulian, and M. Mansourizadeh, "Provide a combined method for scoring answers in question answering systems, " in *7th International Conference of Information Technology and Knowledge, Hamadan,* 2013, pp. 5-30. (in Persian).

[36] Z. Huang, M. Thint, and Z. Qin, "Question classification using head words and their hypernyms," in *Proceedings of the 2008 Conference on empirical methods in natural language processing*, 2008, pp. 927-936.

طبقه‌بندی پرسش در سیستم پرسش و پاسخ با استفاده از طبقه بندی ترکیبی و انتخاب ویژگی

شهرام گلزاری[۱،*]، فریده صانعی[۲]، محمودرضا سایبانی[۳]، عباس حریفی[۱] و مجتبی بصیر[۲]

[۱] گروه مهندسی برق و کامپیوتر، دانشگاه هرمزگان، بندرعباس، ایران و هسته پژوهشی یادگیری عمیق، دانشگاه هرمزگان، بندرعباس، ایران.

[۲] گروه مهندسی برق و کامپیوتر، دانشگاه هرمزگان، بندرعباس، ایران.

[۳] مرکز علمی کاربردی بندرعباس ، دانشگاه علمی کاربردی، بلوار فراهانی، بندرعباس، ۳۳۱۵۳-۷۹۱۹۹، ایران.

ارسال ۲۰۲۰/۰۹/۰۲؛ بازنگری ۲۰۲۱/۰۷/۰۷؛ پذیرش ۲۰۲۱/۰۸/۱۳

چکیده:

سیستم پرسش و پاسخ شکل خاصی از بازیابی اطلاعات است که از سه بخش پردازش پرسش، بازیابی اطلاعات و انتخاب پاسخ تشکیل می‌شود .مهم‌ترین بخش سیستم پرسش و پاسخ را می‌توان تعیین نوع پرسش دانست زیرا بر روی بخش‌های بعدی سیستم تاثیر می‌گذارد. در این تحقیق، از ویژگی‌های مؤثر و طبقه‌بندی ترکیبی استفاده‌شده‌است تا از طریق افزایش دقت تشخیص نوع پرسش‌ها، کارایی سیستم پرسش‌وپاسخ ارتقا داده شود. همچنین برای انتخاب ویژگی‌ها و انجام طبقه‌بندی ترکیبی از الگوریتم جستجوی گرانشی استفاده‌شده‌است. روش پیشنهادی بر روی مجموعه داده‌های مختلف از طریق آزمایش‌های متنوع شامل چهار روش عدم انتخاب ویژگی و عدم استفاده از طبقه‌بندی ترکیبی، روش انتخاب ویژگی و عدم استفاده از طبقه‌بندی ترکیبی، روش عدم انتخاب ویژگی و استفاده از طبقه‌بندی ترکیبی و همچنین روش انتخاب ویژگی و استفاده از طبقه‌بندی ترکیبی مورد ارزیابی قرار گرفته است. این آزمایش‌ها با استفاده از دو الگوریتم تفاضل تکاملی و الگوریتم جستجوی گرانشی انجام شده‌اند. نتایج حاصل از آزمایش‌ها نشان می‌دهد که روش پیشنهادی عملکرد خوبی در مقایسه با روش‌های پیشرفته‌ی مورد استفاده در پژوهش‌های پیشین دارد.

کلمات کلیدی: سیستم پرسش‌وپاسخ، طبقه‌بندی پرسش، انتخاب ویژگی، طبقه‌بندی ترکیبی، الگوریتم جستجوی گرانشی، الگوریتم تکاملی تفاضلی.