Spam Classification Using MOEA/D

Rand Ahmad Atta*, Soukaena H. Hashem** and Ekhlas Khalaf Gbashi***
Department of Computer science, University of Technology, Baghdad-Iraq.
Corresponding author: * randahmad_at@yahoo.com ** Soukaena.hassan@yahoo.com &
*** 110026@uotechnology.edu.iq

Abstract

In mathematics, it's very easy to find the maximum point or minimum point of a function or a set of functions, but it's difficult to find a set of function simultaneously in the real world due to the different kinds of mathematical relationships between objective functions. So the multi objective optimization algorithm has the ability to deal with a many objectives instead of one objective, because of the difficulties in the classical methods of multi objectives optimization, the evolutionary algorithm (EA) is effective to eliminate these difficulties, in order to apply the evolutionary algorithms to improve the multi-objective optimization algorithm, the multi - objective evolutionary algorithm based on decomposition is one of the algorithms that solve multi objective optimization problems. This paper aims to enhance the e-mail spam filtering by using multi - objective evolutionary algorithm for classifying the e-mail messages to spam or non-spam in high accuracy. The first step in the proposal is applying normalization. The second step is applying feature selection which is implemented to choose the best features. Finally, implement multi - objective evolutionary algorithm based on decomposition. The evaluation of the performance of model by using testing databases from the spam database. The model depended accuracy as a criterion to evaluate model performance. The experimental results showed that the proposed system provides good accuracy in the experiment 1 (91%), very good accuracy in the experiment 2 (92%) and excellent accuracy in the experience 3 (98%). [DOI: 10.22401/ANJS.21.4.14]

Keywords: Spam database, Multi objective optimization, Classification.

1. Introduction

Email spam is the most common form of spam. The aim of these spam message is deceiving recipients by disclosure important information such as bank account numbers and credit card details that causing losses billions of dollars to companies, where used spam filters to control the email spam [1]. Spam filtering system trained on spam email dataset, the feature selection is an effective strategy in the preparation and processing the dataset in order to construct a model with good accounting efficiency [2]. Many of machine learning algorithms can be used to build the spam filtering systems such as multi objective evaluation algorithm (MOEA). At the current time, evolutionary techniques for multi-objective optimization to search for a set of global trade-off solutions gaining great attentions from the researchers due to their robustness and effectiveness, and it's perfectly suited to multi-objective optimization problems due to the strong features of this algorithm, the evolutionary search global optimization algorithm has recently been growing in popularity, the evolutionary algorithm (EA) has become a powerful tool to solve the multi-objective optimization problems unlike other search methods, also the evolutionary algorithm not need gradient information. The multi objective evaluation algorithm (MOEA) is considered a solution in the fields related with multi objective problems conflicting (minimized or maximized), where it's give a number of solutions instead of one solution, where using multi objective evaluation algorithm based on decomposition (MOEA/D) to solve the optimization problems [3,4]. The proposal will exploit information gain (IG) as one of feature selection method and MOEA/D to enhance email spam filtering system by increasing the accuracy.

2. Related Work

López-Herrera, A. G. et.al in 2008, discuss how ascertain from set of filtering rules with various profiles. They proposed using the non-dominated sorting genetic algorithm (NSGA-II) it is built on the basis of MOEA. Where the

model was applied on corpus PU1 dataset and it has been tested using a public spam dataset, through experimental results which represent average number of various rules, can query that NSGA-II achieves number of rules begin from very strong rules to weak rules, and NSGA-II provides a flexible way to set a filtering rule profile in a spam filtering system [1].

Vu, M. T. et.al in 2014, discussed how to solve the problem of Vietnamese anti-spam system by proposed a multi-objective for generating group of trade-off solutions by using spam assassin and by taking both detection rate (DR) and false alarm rate (FAR) are two objectives. By three scenarios with various numbers of spam assassin rules select set of Vietnamese spam data. Conducted the experiments on Vietnamese spam data set and recorded result of each scenario then compare with single objective optimization (SOO) performance. The results of new approach were more efficient and created a group of ready to use rule, which supported various levels of the trade-off between detection rate (DR) and false alarm rate (FAR) [5].

Kadhim, J. N. 2015, the proposal provides work to achieve a parallel between the goals of the system of summarizing texts. Where the model based on sentence extraction, and redirected into more semantic measure reflects individually. There are two optimization models are content diversity and content coverage. The two models are defined and coupled as single-objective optimization or a multi-objective optimization (MOO) problem. In order to harness its power has been added each of proposed (heuristic local repair operators and perturbation) to evolutionary algorithm. and applied the models using document sets supplied by Document Understanding Conference 2002 (DUC2002). Then compared with other methods. And using Recall-Oriented Understudy for Gisting Evaluation (ROUGE) to measure performance of the proposed work. Results obtained is positive effectiveness to the proposed models [4].

Ruano-Ordás, D. et.al in 2017, discuss how treat both the false positive rate and false negative error rates, and treat email classification time, in order to presents evaluation multi objective algorithm formulation for the anti-spam filtering problem, and compared them with single objective optimization (SOO) formulations found in the literature. The study is conducted by using both the spam assassin email dataset and the (Wire brush4SPAM) framework anti-spam filtering. Where the non-dominated sorting genetic algorithm (NSGA-II) was applied in order to confirmation and proof depending this novel approach to the anti-spam filtering problem. The comparison showed the best results were for NSGA-II. And by the results obtained from the experiments allows the decision maker to chooses between a group of optimal and (classification quality and efficiency) with respect to flexibility filter formation [6].

3. Feature Selection

Variable selection also known as (feature selection) is the process of select attribute in terms of the problem of a target learning. the aim of feature selection is to delete excessive and not relevant attribute [7]. Benefits of feature selection using only a subset of all features ability to build faster and simpler models, and by concentration on a selected subset of features can gaining a best understanding of the processes by described the data [8]. Information gain is one of evaluations that using in the used attribute evaluations, and it is another key concept of information theory [7]. Entropy measures is the foundation in the information gain attribute ranking methods. Which distinguish the purity of examples in set of an arbitrary [7].

The entropy of A is:

$$H(A) = -\sum_{y \epsilon Y} P(A) \log_2 P(A) \quad \text{Equation} \ldots\ldots\ldots (1)$$

Where: P(A) = probability function for the random variable A.

There is a relationship between features "A" and "B" is:

$$H(A|B) = -\sum_{b \epsilon B} P(B) \sum_{a \epsilon A} P(A|B) \log_2 P(A|B) \quad \text{Equation} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where: P (A|B) = probability of "A" given "B".

Given the entropy as a criterion of not purity in a training set S, this measure is

known as information gain. Is measure invert additional information about "A" provided by "B" it is given by.

$IG = H(A) - H(A|B) =$
$H(B) - H(B|A)$ Equation .....(3)

4. Evolutionary Algorithms

The simulated was genetics as a process to learning how to optimizing the fitness of types and to adapt with the environment. Where evolution algorithms can tasks learning with the ability to evolve or perform to tasks optimization [9]. Here are four paradigms using as the basis for many of the activity of the field: (A) genetic algorithms (B) genetic programming (C) evolutionary strategies (D) evolutionary programming. Genetic algorithm based on the rules of genetics and natural selection this reason is working on a solid and powerful search procedures [10]. The Genetic algorithm process is discussed through the GA cycle as in Fig.(1) [10].

The first step of GA is initializing a population, the population of solutions have been encoded into strings, this strings depending on the nature the problem so the string could be (bits, characters or numbers). The second step is evaluation each chromosome. The third step is created new population by applying (selection, crossover, mutation). Finally, when solutions are matched with some value of optimal or get to the exit condition the process is stop [11].

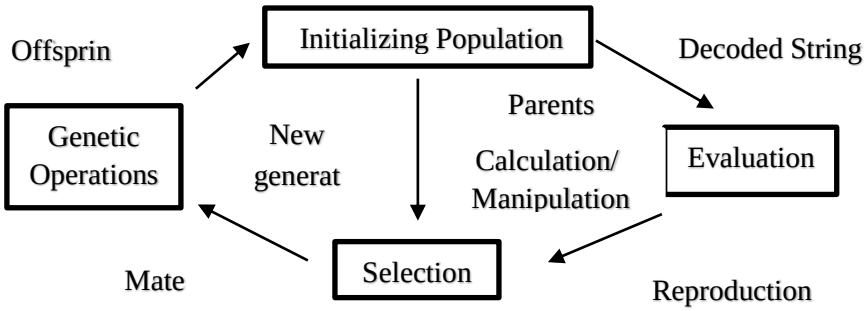


Fig.(1): Genetic Algorithm Cycle.

5. Multi-Objective Optimization Problem

A multi-objective optimization problem (MOOP) deal with a number of objective functions which are to be maximized or minimized [12]. (MOOP) can follows [13]:

$Minimize\ or\ Maximize\ F(X) =$
$(f_1(x), f_2(x), \dots, f_m(x))^T$ Equation .....(4)
subject to $x \in \Omega$

where: x = is dimensional decision variable.
m = is number of objective functions
Ω = is contains all possible x that can be used to satisfy an evaluation of F(x).

Through the conflicting objectives functions can produced set of optimal solutions, and known as Pareto-Optimal solutions, these optimal solutions can't say no one better than the other with respect to all objective functions and the reason for there are many optimality solutions [4,14]. Also Pareto optimal set known as solution to multi objective optimization is represent collection of points, these solutions are also termed non-inferior, admissible, or efficient solutions [17].

PS = {x ∈ Ω|x is a Pareto optimal solution} [15]

The non-dominated vectors when plotted in objective space known as the Pareto front [17].

PF = {F(x) = (f1(x), ..., fk(x)|x ∈ PS} [15].

A concept of domination by comparing between two solutions using in multi objective optimization, the solution known as non-dominated solution is when solution is not

dominated by any other solutions of the multi-objective optimization problem, through that can find a set of non-dominated solutions and a set of dominated solutions. Fig.(2) illustrates the concept of Pareto for sample two minimize objectives problems [14].

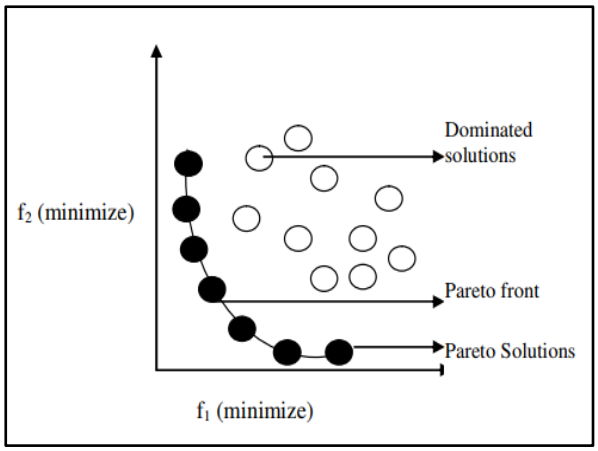


Fig.(2): Concept of Pareto.

6. Multi-Objective Evolutionary Algorithm

Classical search and optimization methods are difficult to extend to the multi objective due to their basic design prevents the consideration of multiple result. In contrast, (EA)s are good to deal with these situations [13,14]. The idea of MOEA/D, is decomposes the multi-objective problem into optimization sub problems. By developing a set of solutions these sub-problems are resolved simultaneously, the population is made up of the best solutions that have been found at each generation, depending on the distances between the aggregation coefficients on the vector, the neighbor relationships are defined among these sub-problems, the optimal Solutions to two neighboring sub-problems should be very similar. In MOEA/D, used the information only from the neighboring sub-problems in order to optimize each sub-problem [5, 18]. There are many approaches using to decomposing an MOP into sub-problems:

- Weighted Sum.
- Tchebycheff.
- Boundary Intersection.

The most widely used is Tchebycheff approach is one of among decomposition approaches, and as one of many approaches used to decompose a multi objective optimization into a set of scalar optimization problem [16]. The definition of Tchebycheff approach is:

$$minimize\ g^{tc}(x|\lambda, z^*) = \max_{1\le i\ge m}\ \{\lambda_i|f_i(x) - z_i^*|\} \quad \text{Equation} \ldots\ldots (5)$$
$$subject\ to\ x\epsilon\ \Omega$$

where: x = is variables to be optimized.

z^* = is the reference point.

λ = is a weight vector.

For more explanation, see algorithm (1).

Algorithm (1): The MOEA/D

Input:

- MOP: $\boldsymbol{Maximize}$/**Minimize**. $\boldsymbol{MOEA} = [\boldsymbol{F_{acc}(I)}, \boldsymbol{F_{fp\ rate}(I)}]$.
- N: total of the sub-problems in MOEA.
- p_c : probability of crossover and p_m : probability of mutation.
- T: is size of neighborhood
- $\text{gen}_{\max}$: the large number of generations

Output: external population(EP).

Begin:

1. Initialization

1.1 EP = ∅, initialize of Z°=$\{Z^\circ_1 = 0, Z^\circ_2 = 100\}$, generate initial population $P_0 = \{P_1, \ldots, P_N\}$.

1.2 find weight: $\lambda^1 = i/N$, $\lambda^2 = {N - i}/{N}$. And calculation the (ED) between two weight vectors and the T closest weight vectors to each weight vector, store in B(ι), ∀ι =1,…, N, set B(ι)=$\{\iota_1, \ldots, \iota_T\}$, where$\lambda^{i1}$ … λ^{iT}, the (T) closest to λ^i.

2. Update For i= 1, …, N

2.1 Evaluate P_0. And select operator two indexes V, Q from [i], applying (crossover, mutation) to create child($\overline{p}$) and evaluate the child($\overline{p}$).

2.2 Update to Z°: if (Z°_i< $\boldsymbol{MOEA}(\overline{p})$) then Z°_i<= $\boldsymbol{MOEA}(\overline{p})$.

2.3 Update of Neighboring : if ($G^{te}(\bar{p}|\lambda^j, Z^\circ) \le G^{te}(P^j|\lambda^j, Z^\circ)$) Then $P^j = \bar{p}$, $\boldsymbol{MOEA}(p)$= $\boldsymbol{MOEA}(\overline{p})$.

2.4 Update of external population(EP): **A.** delete from EP all the vectors dominated by $\boldsymbol{MOEA}(\overline{x})$. **B.** the $\boldsymbol{MOEA}(\overline{p})$ insert to EP if no vector in EP is dominate.

3. Condition to stop

If ($\text{gen}_{\text{level}} >= \text{gen}_{\max}$) *then stop*, Output external population (EP).

Else, $\text{gen}_{\text{level}} = \text{gen}_{\text{level}} + 1$ and go to step 2.

End if

END

7. The Evaluation Measures of Classification

The models have been experimented and evaluation. The evaluation measures are defined from a matrix which has only two classes (spam and not spam) [19].

	True Class	
Texting Class	Spam	Not spam
Spam	True positives (TP)	False negatives (FP)
Not spam	False positives (FN)	True negatives (TN)

There are Different evaluations measures to evaluate the performance for this system: [20].

1. **Accuracy:** measures the correct predictions over total of instances evaluated. $\frac{tp+tn}{tp+tn+fp+fn}$
2. **Error Rate:** measures the incorrect predictions over total evaluated. $\frac{fp+fn}{tp+tn+fn+fp}$
3. **Precision:** measures correctly predicted over the total patterns in a positive class. $\frac{tp}{tp+fp}$

Recall: measures positive patterns that are correctly classified. $\frac{tp}{tp+tn}$

8. The Proposed Classification Based MOEA/D

There are three main components of system: normalization: apply normalization on spam data base to uniform the variants frequencies of words over the datasets. Feature selection: to choose the best features that improve performance that contribute to raising the rate of accuracy of the model, where using the information gain is one of the ways of feature selection used in the proposal. And then applied MOEA/D, as show in Fig.(3). The following sections will explain each phase in details.

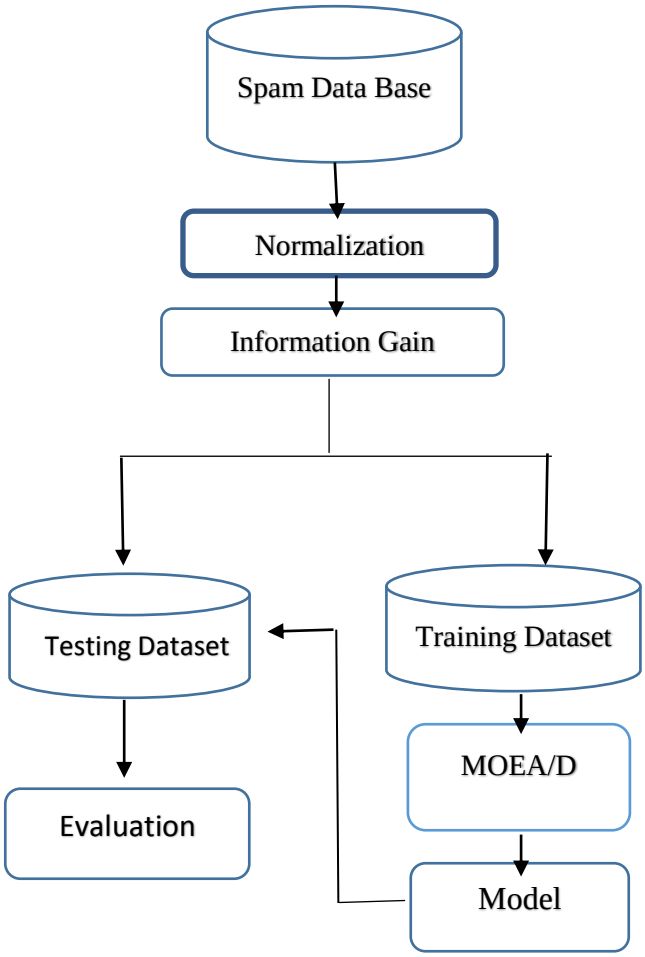
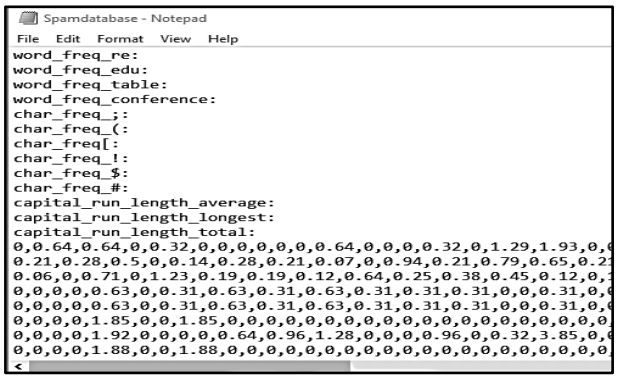


Fig.(3): Scheme for Proposed System.

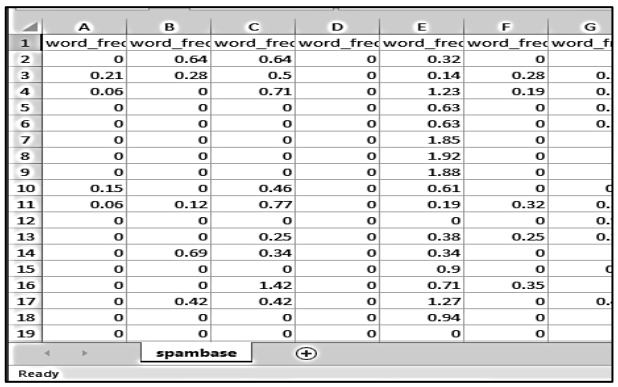
8.1 Applying Normalization

The database used in the proposal is consists of words and symbols and real number values that separated by a comma as Fig.(4.a). Where each of words and symbols are represented as features, the real number values are represented as features values. Spam base text file data will be converted into excel database by using Microsoft office where each row in excel database as Fig.(4.b). Values of features should be uniform between 1 and 0, because each feature has values with a different range. So applying normalization on dataset. By using equation (6). As in figure (4.c), (4.d).

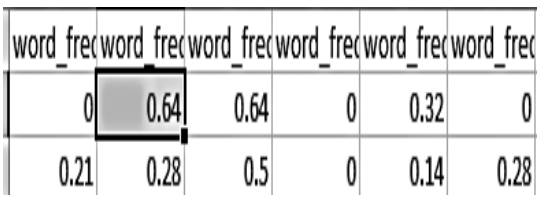
$$value_{new} = \frac{value^{old} - value^{min}}{value^{mix} - value^{min}}$$ Equation .....(6)



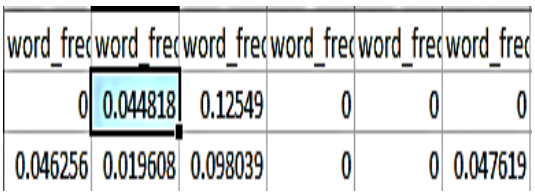
(a)



(b)



(c)



(d)

Fig.(4): a) Spambase Text File, b) Spambase Excel database after Conversion, c) Spambase Dataset Before Normalization, and d) Spambase Dataset After Normalization.

8.2 Feature Selection

Information gain is one of the methods for feature selection that used in the proposal to choose the best feature the following algorithm shows the process of feature selection.

Algorithm (2): Feature selection

Input: spam dataset after normalization.

Output: information gain for each feature.

Begin:

Step 1: X= total No. of email in spam DB.

Step 2: For each class in DB.

A. Pro(Spam)= No. of spam /X

B. Pro (Not Spam) = No. of Not Spam/ X

C. Find the entropy of class by using equation (1)

$$Entropy(class) = -\sum pro(spam) * \log_2 pro(spam) + pro(Not\ Spam) * \log_2 pro(Not\ Spam)$$

End For

Step 3: For each Feature in DB.

For each value in feature.

A. Pro (value) in DB.

B. Pro (value) with two class type.

C. Find the entropy for each value with two class type by using equation (2).

$$Entropy(value|class) = pro(value) * -\sum pro(value|spam) * \log_2 pro(value|spam) + pro(value|Not\ Spam) \log_2 pro(value|N$$

End For

End For

Step 4: Find information gain by using equation (3).

$$IG = Entropy(class) - Entropy(value|calss)$$

End

Then arrangement the features from high IG to lower in an excel database to be used in next step. In next step is preparation training dataset and testing dataset by divided spam database into two part. The training database contains 3000 records and the testing database includes 1601 records.

8.3 Multi-Objective Evolutionary Algorithm Based on Decomposition

The goal of MOEA is to find the best trade-off between two objectives function (accuracy and false alarm rate), the first

objective should be maximized and the second objective should be minimized. And because of a contradiction between objectives function it is very difficult to find single solution that optimizes all objectives function simultaneously. Therefore, multi-objective evolutionary algorithm (MOEA) give the set of solutions that are not dominated by any of the other solutions known as the Pareto optimal set. Algorithm (1) in section 6 is adopted to solve the optimization problem of email spam filtering.

$$Maximize\ MOEA = \left[F_{acc}(I), F_{fp\ rate}(I)\right] \quad \text{Equation } \ldots\ldots\ldots\ldots (7)$$

Where: $F_{acc}(I) = \frac{TP+TN}{TP+TN+FP+FN} *100$

$F_{fp\ rate}(I) = \frac{FP}{FP+TN} *100$

Considering N that denotes the number of sub-problems, $\lambda^1 \ldots \lambda^N$ that denotes the weight vectors for each sub-problem, $Z^{\circ}=\{Z^{\circ}_1, Z^{\circ}_2\}$ that denotes a reference point to the two contradictory objective functions, T is size of neighborhood and external population (EP) using to store non-dominated solutions, the Tchebycheff approach is used to decomposed the Pareto Front into scalar optimization sub-problems. Representation of each individual is as a vector with having fixed-length. Each individual calculated the distance between two vector weight by Euclidean Distances (ED) and store the closest distance in B[i] as index, the number of individual that are stored in B[i] depends on T, then evaluated each population by using equation (7). And at each generation of MOEA/D the selection operator will select two parents randomly from B[i], then both crossover and mutation operator were applied to create new generation, and evaluated each generation by using equation (7). After that update the reference point Z^*, we used two objective function so we will have two reference point (Z_1,Z_2). And update the neighbors using the Tchebycheff approach, finally, update EP to determine the non-dominant solutions.

- The objective function evaluation used in classifier the email spam filtering and evaluation for each solution in this proposed.

The gene in chromosome was (1, 0), if number of gene type (1) is greater than (K), then the email is spam. Otherwise not spam. where (K) is number using for the classifier email spam filtering to spam or not, through an experiments were carried out on spam database, we found the value of minimum for (K) was equal 3 for the classifier email spam filtering. Then comparison between current class and previous class to calculate each of the (true positive, true negative, false positive, false negative) and evaluation the solution by using equation (7). The model was trained with all 57 features, then training will be done with best 40 features,

9. The Experimental Results

The experiments of MOEA/D model are depended on the features and (crossover and mutation) on position of bit in the chromosome, to show how effect (features and (crossover and mutation on index of position of bit)) on results of accuracy. Where since some features prevented results from reaching higher efficiency, the experiments start with all (57) features, with best (40) features, then with best (20) features. And the different in the index of position of bit has an effect in obtaining different results in accuracy. And show which of the experiments give highest of accuracy. In experimented, we used 3000 record as training dataset, and used 1601 record as testing dataset to evaluated the performance, to get The parameters used in model for all experiments ($P_C = 0.7$,$P_m = 0.001, N = 100, K = 3$). Note: The CP: is index of position of bit, that applying crossover on position of bit. And MP: is index of position of bit, that applies mutation on position of bit.

Experiment 1: The experiment done with all 57 feature and with different indexes of positions of bits in order to applying (crossover and mutation) on her.

Table (1)
Result of Experiment 1.

NO. EXP.	CP	MP	Accuracy
1	36	26	91%
2	35	25	89%
3	45	30	86%

Experiment 2: The experiment done with best 40 feature and with different indexes of positions of bits in order to applying (crossover and mutation) on her.

Table (2)
Result of Experiment 2

NO. EXP.	CP	MP	Accuracy
1	25	18	92%
2	30	15	90%
3	25	16	89%

Experiment 3: The experiment done with best 20 feature and with different indexes of positions of bits in order to applying (crossover and mutation) on her.

Table (3)
Result of Experiment 3.

NO. EXP.	CP	MP	Accuracy
1	25	18	92%
2	30	15	90%
3	25	16	89%

1. In the index of position of bit (45) and in the index of position of bit (30) with all (57) features giving the minimum accuracy is 86%. In the index of position of bit (36) and in the index of position of bit (26) with all (57) features giving the maximum accuracy is 91%.
2. In the index of position of bit (25) and in the index of position of bit (16) with best (40) features giving the minimum accuracy is 89%. In the index of position of bit (25) and in the index of position of bit (18) with best (40) features giving the maximum accuracy is 92%.
3. In the index of position of bit (13) and in the index of position of bit (4) with best (20) features giving the minimum accuracy is 90%. In the index of position of bit (11) and in the index of position of bit (3) with best (20) features giving the maximum accuracy is 98%. In figure (5) shows classification results of highest of accuracy for three experiment of MOEA/D.

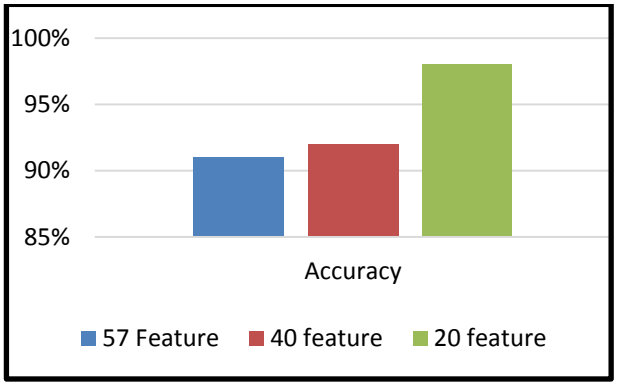


Fig.(5): Classification Results of Highest of Accuracy.

10. Comparison with Related Works

The Table (4) lists of the related work, which are collected through the research survey. The Table (4) displays the used algorithms and results.

Table (4)
Comparison with Related Works.

NO.	Researcher name and year	Algorithms	Training Dataset	Problem Solved	Results
1	López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F., 2008	Multi objective Evolutionary Algorithm (NSGA-Il)	Spam dataset (PU1)	Spam filtering problem by ascertain from set of filtering rules with various profiles were solved	Average number of different rules: A= 270,2 B =328,3 C =232,2
2	Vu, M. T., Tran, Q. A., Ha, Q. M., & Bui, L. T., 2014	Multi objective Evolutionary Algorithm (NSGA-II)	Spam Assassin	solve the problem of Vietnamese spam detection	with 30 rules the FAR, SDR: {(74.03%, 7.79%); (74.46%, 8.66%); (72.29%, 6.93%)}. with 50 rules the FAR, SDR: {(83.98%, 9.96%); (83.55%, 8.66%); (82.68%, 7.36%)}. with 100 rules the FAR, SDR: {(83.55%, 8.23%); (81.39%, 6.06%); (82.25%, 6.93%)}.
3	Ruano-Ordás, D., Basto-Fernandes, V., Yevseyeva, I., & Méndez, J. R., 2017,	Multi objective Evolutionary Algorithm (NSGA-II)	spam assassin and the (Wire brush4SPAM)	Spam filtering problem were solved by confirmation and proof depending on MOEA to the anti-spam filtering problem	NSGA-II closest to origin (FP=487, FN= 7 execution time=425,036) NSGA-II best on FP (FP=543, FN= 0, Execution Time= 435,66). NSGA-II best on FN (FP=459, FN=22 Execution Time=1971,072) NSGA-II best on Execution Time (FP=536, FN= 3, Execution Time =387,136)
4	Kadhim, J. N., 2015	SOEA algorithm, MOEA/D algorithm	DUC 2002	the problem of automatic text summarization were solve by create a summarized text	(Rouge-2, Rouge-L) SOEA1 (0.25437, 0.48314) SOEA2 (0.27889, 0.49412) MOEA1(0.46578, 0.60105) MOEA2(0.47412, 0.61742)
5	Atta A.R., 2018.	MOEA/D Algorithm	Spam email	Email Spam filtering problem were solved by MOEA/D	The accuracy in the experiment 1 (91%), in the experiment 2 (92%) and in the experience 3 (93%).

11. Conclusion

In this paper, the proposed system adopt the MOEA/D to classify email spam filter into both spam and non-spam, in order to enhance e-mail spam filtering. The proposed system included three phases:

1. First phase: To uniform the values of database between [0, 1] by applying normalization.
2. Second phase: The suggested system used IG to choose important features from spam database.
3. Third phase: Training and testing the model.

In training: Implementing the MOEA/D on training dataset. In the first: the model was trained with all 57 features and with different indexes of positions of bits. And in the second: the model was trained with best 40 features and with different indexes of positions of bits. Finally, the model was trained with best 20 features and with different indexes of positions of bits.

In testing: Evaluation the performance will be done for each (all 57 feature, best 40 feature, best 20 feature). In order to get result of high of accuracy from each part as shown in figure (5).

The different in results of accuracy in each experiment are due to some features prevented results from reaching higher efficiency, and also (crossover and mutation) on position of bit in the chromosome showed a significant impact on the accuracy of the proposed system. as in Table (1), Table (2) and Table (3). Despite this the model of system is succeeded in classifying email spam filtering

into (spam email and non-spam email). With better accuracy in the experiment 1 (91%), very high accuracy in the experiment 2 (92%) and excellent accuracy in the experience 3 (98%).

References

[1] López-Herrera, A. G., E. Herrera-Viedma, and F. Herrera. "A multi objective evolutionary algorithm for spam e-mail filtering." Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on. Vol. 1., 366-371, 2008.

[2] Lewis, David D. "Feature selection and feature extraction for text categorization." Proceedings of the workshop on Speech and Natural Language. Association for Computational Linguistics. 212-217, 1992.

[3] Deb, K. "An investigation of niche and species formation in genetic function optimization". ICGA'89, 42-50, 1989.

[4] Vu, M. T., Tran, Q. A., Ha, Q. M., & Bui, L. T. "A multi-objective approach for vietnamese spam detection". In Knowledge and Systems Engineering. Springer, Cham, 2014.

[5] Ruano-Ordás, D., Basto-Fernandes, V., Yevseyeva, I., & Méndez, J. R. "Evolutionary Multi-Objective Scheduling for Anti-Spam Filtering Throughput Optimization". In International Conference on Hybrid Artificial Intelligence Systems. Springer, Cham, 2017.

[6] Kadhim, J. N. "Evolutionary Based Extractive Multi Document Text Summarization", 2015.

[7] Cover, T. M., & Thomas, J. A. "Elements of information theory". John Wiley & Sons, 2012.

[8] Guyon, I., & Elisseeff, A. "An introduction to variable and feature selection". Journal of machine learning research, 3(Mar).1157-1182, 2003.

[9] Yu, X., & Gen, M. "Introduction to evolutionary algorithms". Springer Science & Business Media, 2010.

[10] Sivanandam, S. N., & Deepa, S. N. "Genetic algorithm optimization problems". In Introduction to Genetic Algorithms. Springer, Berlin, Heidelberg.165-209, 2008.

[11] Aissa, N. B., & Guerroumi, M. "A genetic clustering technique for Anomaly-based Intrusion Detection Systems". In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACIS International Conference on (pp. 1-6),2015.

[12] Kalyanmoy, D. "Multi objective optimization using evolutionary algorithms". John Wiley and Sons, 2001.

[13] Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. "A Survey of Multi objective Evolutionary Algorithms for Data Mining: Part I". IEEE Trans. Evolutionary Computation, 18(1), 4-19, 2014.

[14] Jeyanthy, P. A., & Devaraj, D. "Multi-objective genetic algorithm for reactive power optimization inclusing voltage stability". International Journal of Engineering Science and Technology, 2(7), 2715-2729, 2010.

[15] Martínez, S. Z., & Coello, C. A. C. "Use of gradient-free mathematical programming techniques to improve the performance of multi-objective evolutionary algorithms". Science, 2013.

[16] Qi, Y., Ma, X., Liu, F., Jiao, L., Sun, J., & Wu, J. "MOEA/D with adaptive weight adjustment". Evolutionary computation, 22(2), 231-264, 2014.

[17] Srinivas, N., & Deb, K. "Multi objective optimization using non dominated sorting in genetic algorithms". Evolutionary computation, 2(3), 221-248, 1994.

[18] Zhang, Q., & Li, H. "MOEA/D: A multiobjective evolutionary algorithm based on decomposition". IEEE Transactions on evolutionary computation, 11(6), 712-731, 2007.

[19] Costa, E., Lorena, A., Carvalho, A. C. P. L. F., & Freitas, A. "A review of performance evaluation measures for hierarchical classifiers". In Evaluation methods for machine learning II: Papers from the AAAI-2007 workshop, 2007.

[20] Hossin, M., & Sulaiman, M. N. "A review on evaluation metrics for data classification evaluations". International Journal of Data Mining & Knowledge Management Process, 5(2), 1, 2015.