

Artículo de investigación

## Feature selection on large-scale issues using clustering and meta-algorithms

Característica de selección en campos de escala grande utilizando clústeres y meta-algoritmos  
Seleção de recursos em problemas de larga escala usando clustering e meta-algoritmos

Recibido: 20 de abril de 2018. Aceptado: 10 de mayo de 2018

Escrito por:

**Fardin Akhlaghian Tab**

(Corresponding Author)

**Shabnam Amiri**

English Language Department, Faculty of Communication and Modern Languages  
Universiti Putra Malaysia (UPM)  
Emails:nimechie@gmail.com

### Abstract

Selection of appropriate input features in the increase of the efficiency of data mining algorithms has a direct and significant effect. More precisely, this extraction of knowledge from problem data is facilitated by three things: reducing data volumes, eliminating duplicate features, and eliminating unrelated features. Given this necessity, extensive research has been carried out in recent years with a variety of trends (statistical, algorithmic, and learning) in this regard. In the meantime, hyper-algorithms such as genetic algorithms have been considered by many researchers. In this research, we have tried to achieve more efficiency by combining clustering and genetic algorithms and reducing computational time.

In this regard, a new representation of the genetic algorithm corresponding to this problem is presented and its operators are appropriately defined. Also, for efficient use of clustering in this study, it was necessary to provide a relatively new algorithm for rapid clustering. To validate the proposed methods and determine their efficiency in solving real problems, several experiments have been carried out on standard data. In the next step, analyzing the proposed methods, we compared the results of the experiments with various algorithms reported in valid and new articles. These comparisons have shown improvements in the efficiency of proposed methods in terms of the accuracy of categorization and feature reduction compared to competing methods. According to the analysis, this improvement was due to the positive effect

### Resumen

La selección de las características de entrada apropiadas en el aumento de la eficiencia de los algoritmos de minería de datos tiene un efecto directo y significativo. Más precisamente, esta extracción de conocimiento de los datos de problemas se ve facilitada por tres factores: la reducción de volúmenes de datos, la eliminación de características duplicadas y la eliminación de características no relacionadas. Dada esta necesidad, se ha llevado a cabo una extensa investigación en los últimos años con una variedad de tendencias (estadística, algorítmica y de aprendizaje) en este sentido. Mientras tanto, hiper-algoritmos tales como algoritmos genéticos han sido considerados por muchos investigadores. En esta investigación, hemos intentado lograr una mayor eficiencia combinando clustering y algoritmos genéticos y reduciendo el tiempo de computación.

En este sentido, se presenta una nueva representación del algoritmo genético correspondiente a este problema y sus operadores se definen de manera apropiada. Además, para un uso eficiente de la agrupación en este estudio, fue necesario proporcionar un algoritmo relativamente nuevo para la agrupación rápida. Para validar los métodos propuestos y determinar su eficacia en la resolución de problemas reales, se han llevado a cabo varios experimentos con datos estándar. En el siguiente paso, al analizar los métodos propuestos, comparamos los resultados de los experimentos con varios algoritmos informados en artículos válidos y nuevos. Estas

of clustering in a faster search of the problem space by the genetic algorithm and adapted display.

**Keywords:** Feature selection, clustering and meta-algorithms.

comparaciones han mostrado mejoras en la eficiencia de los métodos propuestos en términos de la precisión de la categorización y la reducción de características en comparación con los métodos de la competencia. Según el análisis, esta mejora se debió al efecto positivo de la agrupación en una búsqueda más rápida del espacio problemático mediante el algoritmo genético y la visualización adaptada.

**Palabras claves:** Selección de funciones, clustering y meta-algoritmos.

## Resumo

A seleção de recursos de entrada apropriados no aumento da eficiência dos algoritmos de mineração de dados tem um efeito direto e significativo. Mais precisamente, essa extração de conhecimento dos dados de problemas é facilitada por três coisas: redução de volumes de dados, eliminação de recursos duplicados e eliminação de recursos não relacionados. Dada essa necessidade, uma extensa pesquisa foi realizada nos últimos anos com uma variedade de tendências (estatística, algorítmica e aprendizagem) a esse respeito. Nesse meio tempo, hiper-algoritmos como algoritmos genéticos têm sido considerados por muitos pesquisadores. Nesta pesquisa, tentamos obter mais eficiência combinando algoritmos de agrupamento e genéticos e reduzindo o tempo computacional.

A este respeito, uma nova representação do algoritmo genético correspondente a este problema é apresentada e seus operadores são apropriadamente definidos. Além disso, para o uso eficiente de clustering neste estudo, foi necessário fornecer um algoritmo relativamente novo para clustering rápido. Para validar os métodos propostos e determinar sua eficiência na solução de problemas reais, vários experimentos foram realizados em dados padrão. Na próxima etapa, analisando os métodos propostos, comparamos os resultados dos experimentos com vários algoritmos reportados em artigos novos e válidos. Essas comparações mostraram melhorias na eficiência dos métodos propostos em termos de precisão de categorização e redução de recursos em comparação com os métodos concorrentes. De acordo com a análise, essa melhoria deveu-se ao efeito positivo do agrupamento em uma busca mais rápida do espaço do problema pelo algoritmo genético e exibição adaptada.

**Palavras-chave:** Selección de funciones, clustering y meta-algoritmos.

## Introduction

A property is a specified value with the meaning of a specified source and is recorded by input (sensing) to help extract the model of a system. For example, in an image that is a visual model of a landscape, each pixel is a function. Of course, a feature can be the result of processing on one or more features. The goal of the machine learning algorithm is to select features that describe well the input data, while lacking additional features and noise in such a way that predictive power is not reduced. (Guyon, Andr, # 233, & Elisseeff, 2003) Image processing, text categorization and gene sequence analysis (Chuang, Chang, Tu, & Yang, 2008; Guyon et al., 2003; Guyon, Weston, Barnhill, & Vapnik, 2002) are among the major applications of choice Because these issues have

a large number of input features, with a high percentage of these features being removed, and the hand-picking of manually-related features is almost impossible for them.

In the mentioned issues, there are a lot of both types of features, extra and unrelated (noise). For this reason, extensive efforts have been made in the subject literature to reduce these features. (Chuang et al., 2008; Guyon et al., 2003) In this case, the choice of the feature is, in fact, the choice of the best subset. If we assume that  $\mathbf{N}$  has properties and we want to find the best subset, we must examine  $2^N$  sets. Since the characteristics of the problems we face are several thousand, so it is not possible for them to

reach a precise solution at an acceptable time; consequently, the use of inaccurate methods based on statistical methods (Guyon et al., 2003) (Blum & Langley, 1997) (Herv et al., 2003), Macroeconomic Algorithms (Alexandridis, Patrinos, Sarimveis, & Tsekouras, 2005), machine clustering and learning It seems necessary.

### Research objectives

With the benefit of the prior knowledge of this field found in library resources and the idea that will be mentioned later, we will present a new hybrid algorithm for the selection of related features that, as far as possible, we will Close the following goals:

- Reach classification accuracy comparable to existing methods
- The proposed algorithm is efficient in terms of runtime and memory usage
- To acquire knowledge and innovate in the field of feature selection
- Solve the problem of existing practices in this field and obtain new results

In fact, in this thesis three methods are proposed for solving the problem. In fact, these methods are complementary and ultimately provided to enhance the productivity of previous methods. In all of these methods, clustering is used as the basis and the initial stage of work. So, we have used a similarity criterion to remove features and recognize their type.

### Research hypotheses

As we did in clustering data mining operations, we also expected that by applying a new method in the clustering of the feature space, we could highlight the efficiency of the feature selection algorithms based on the genetic algorithm. Take it for this reason, in the first phase, we first combined a new and efficient clustering algorithm with the appropriate genetic algorithm, and in the next step we resolved some of the disadvantages of the clustering algorithm, and finally, in the third method, the results of these two work We combined to achieve more favorable results.

**-Select feature and build features.** The features can be divided into three categories, related, unrelated, and extra. Related features are also useful (Blum & Langley, 1997; Guyon et al., 2003; Kohavi & John, 1997). In fact, there are

features that increase predictive performance (Guyon et al., 2003).

Unrelated features are features that reduce predictive performance. Removing these features will result in improved performance (accuracy, accuracy, and stability) and efficiency (lower computational cost) of the categorization. Additional features are referred to as features that are roughly dependent on the value of one or more other attributes. In other words, having independent properties, the associated attributes do not have additional information, so eliminating these features will not reduce the efficiency of the algorithm. Since removing these features can reduce the search time, it can be said that deleting them will increase the efficiency of the algorithm. In fact, by eliminating these features, we expect the learner model (categorizing / approximating) to be trained with less evaluation, which can reduce the training time of the classifier.

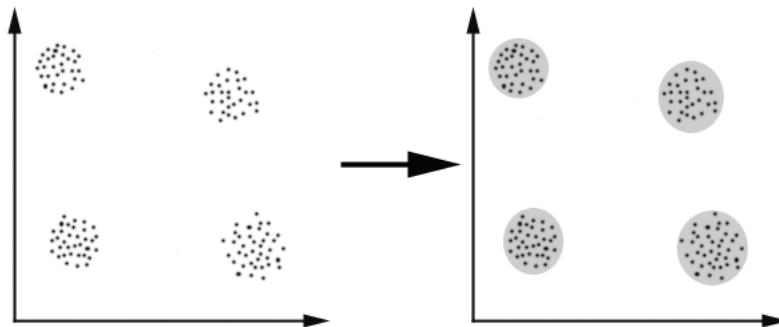
Choosing the property is actually using methods such as selecting the optimal subset, which reduces the number of features that are available for the desired problem. But in extracting the attribute, linear transformations are used to change the data space. Different methods differ in the definition of the transfer function (transmission matrix). Two main element analysis methods (PCA) and linear separator analysis (LDA) are the most common linear transformation methods.

There are four basic essays in this field, all of which are theoretically close to the problem. The first two papers use the clustering technique to create a feature vector (Bekkerman, El-Yaniv, Tishby, & Winter, 2003; Dhillon, Mallela, & Kumar, 2003) Another paper uses matrix completion (Globerson & Tishby, 2003), and the latter uses a new method to create new features. (Torkkola, 2003)

**-Build attributes through clustering.** In fact, clustering is one of the uncontrolled learning areas, by which examples are divided into groups whose members are similar to each other, which is referred to as clusters. Different criteria can be used to determine the similarity. For example, if the distance criterion is used for clustering, this clustering method is called distance-based clustering. For example, in Figure 1, the input examples on the left-hand side are divided into four spikes in the right-hand side. In this example, each instance belongs to one of the clusters. As an example, consider the form (1); in this figure,

each point of a vehicle (object) that shows the characteristics of maximum speed and weight. Also, each cluster is represented by an ellipse, and the phrase next to each ellipse indicates the

label of that cluster. The axes of the coordinate system, in which the samples are shown, are called space features.



As shown in the figure, in this example vehicles are divided into three general clusters. For each of these clusters, one representative is considered, for example, the average vehicle cluster is considered and introduced.

### Combined approach

In a combination of several approaches, the approach is used in feature selection. One of the most common hybrid systems is the use of the filter approach in the first step. In other words, since the initial number of features is usually very high, the filter approach can be very expensive to eliminate a large number of unrelated features. Then, in the second step, the hood approach is used to select the appropriate subset of the attributes. At this stage, since the number of features has already been reduced by the filter approach, the use of the cover can be acceptable. Though in the thematic literature, this combination of maximal contributions is allocated, but combining the filtering of the filter with the merged approach is equally valuable and valuable (Unler, Murat, & Chinnam, 2011)

In (Chun-Kai & Hong, 2005), cross-information with colony ants has been used to select the attribute. In this paper, firstly, the number of input characteristics was reduced using the mutual information statistics, and then using the anion clone algorithm, a typical number of attributes were selected by displaying the graph (the full graph).

In Qinbao et al., 2013, whose algorithm is presented in Fig. 2, the first step is the removal of unrelated features. The article then uses clustering to characterize similar features. Then, with a forward search method, a certain number of centers of these clusters are selected and returned as the final property set. In Mesleh & Kanaan (2008), an anion colony has been used along with statistics and backup vector machine categorization. In Jinjie, Yunze, & Xiaoming, 2006, interactive information was used before the genetic algorithm was used to select the feature. In Unler et al., 2011, mutual information and cumulative particle optimization algorithm and backup vector machine have been used.

### **-Combine Fast Clustering Algorithm with Genetic Algorithms to select the most suitable features.**

This algorithm offers a new combination of filtering, clustering, and genetic algorithms to select the most effective features. Many of the previous algorithms focus only on a bunch of removable features (unrelated and duplicate), while the algorithm is designed in a way that, in addition to removing unrelated features, a certain feature in the removal of the feature it has repetitions. The clustering used, as claimed in Qinbao et al., 2013, can greatly outline and eliminate similar features. Of course, in Qinbao et al., 2013), a completely exploratory method is used to extract the feature vector from clusters. While this algorithm has tried to use the genetic algorithm as one of the most widely used meta-algorithms for this work. It does, in fact, enable our approach to take into account when choosing features.

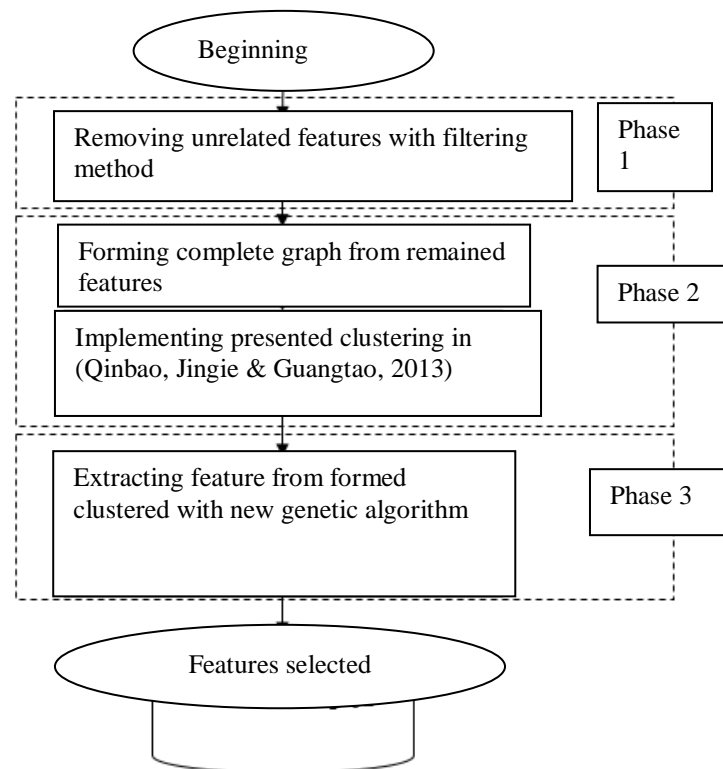


Figure 2- The general schema of the proposed algorithm

**-Unrelated feature filtering.** At this stage, features whose dependence on the target object (class) is less than a certain value is deleted. The dependency criterion used is symmetric similarity (Qinbao et al., 2013). This criterion is calculated on the basis of mutual information. The mutual information calculated from equation (2) gives the degree of similarity of the distribution of the dependent property to the independent property. The symmetric similarity is a normalized mutual information that is divided into two entropy sets. How to calculate it in relation (3).

$$SU(X, Y) = \frac{2 \times IG(X|Y)}{H(X) + H(Y)} \quad (1)$$

**-Characteristic clustering using the Fast method (Qinbao et al., 2013).** To implement this method, symmetric dependence was firstly calculated between all input characteristics. Then a complete graph of input features was constructed in which the input characteristics of the graph nodes and the two-headed interconnected dependence were considered as the weights of the edges. The main step of creating the spanning tree is the minimum of this graph. In other words, using a primitive algorithm (as explained in the reference article), a minimal spanning tree was created. The minimum spanning tree refers to a tree that includes all nodes of the connected graph, provided that the total weight of its edges is minimized among all possible spanning trees. Although there may be more than one minimum spanning tree for a graph, it is enough to have one of them for this application.

**-Extraction of features with genetic algorithm.** In this phase, we try to use the genetic algorithm to extract the characteristic of clusters. Of course, according to experiments conducted using the Fast Clustering (Qinbao et al., 2013), some of the clusters can be very large and, as shown in Fig. 3, all the features inside these clusters are not necessarily dependent. In other words, since it is assumed in that article that all the properties inside a cluster are dependent on each cluster, only a feature is selected, while this may not be the case. It is also assumed in that article that there is useful information for all clusters in the cluster, while this may also not be true. It can be said that one of the disadvantages of this clustering algorithm (Qinbao et al., 2013) is that it shows that the use of greedy and local methods, such as those used in the reference article itself, is not efficient enough. To solve this problem, in this paper, the genetic

algorithm has been used in a way that can extract more than one attribute from a cluster, and this algorithm is able to delete clusters that do not have useful information, and do not choose a representative of it.

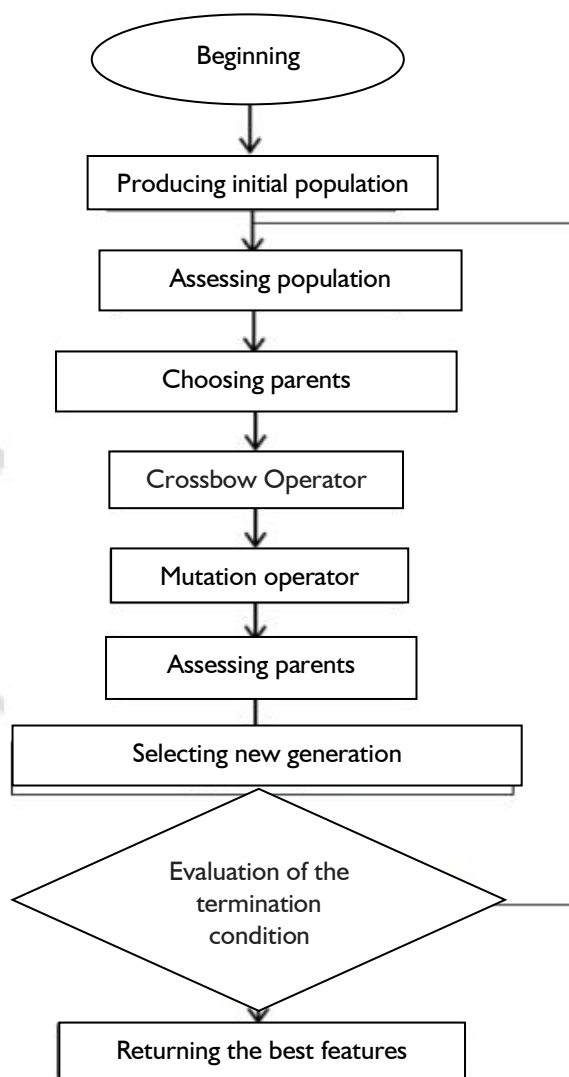


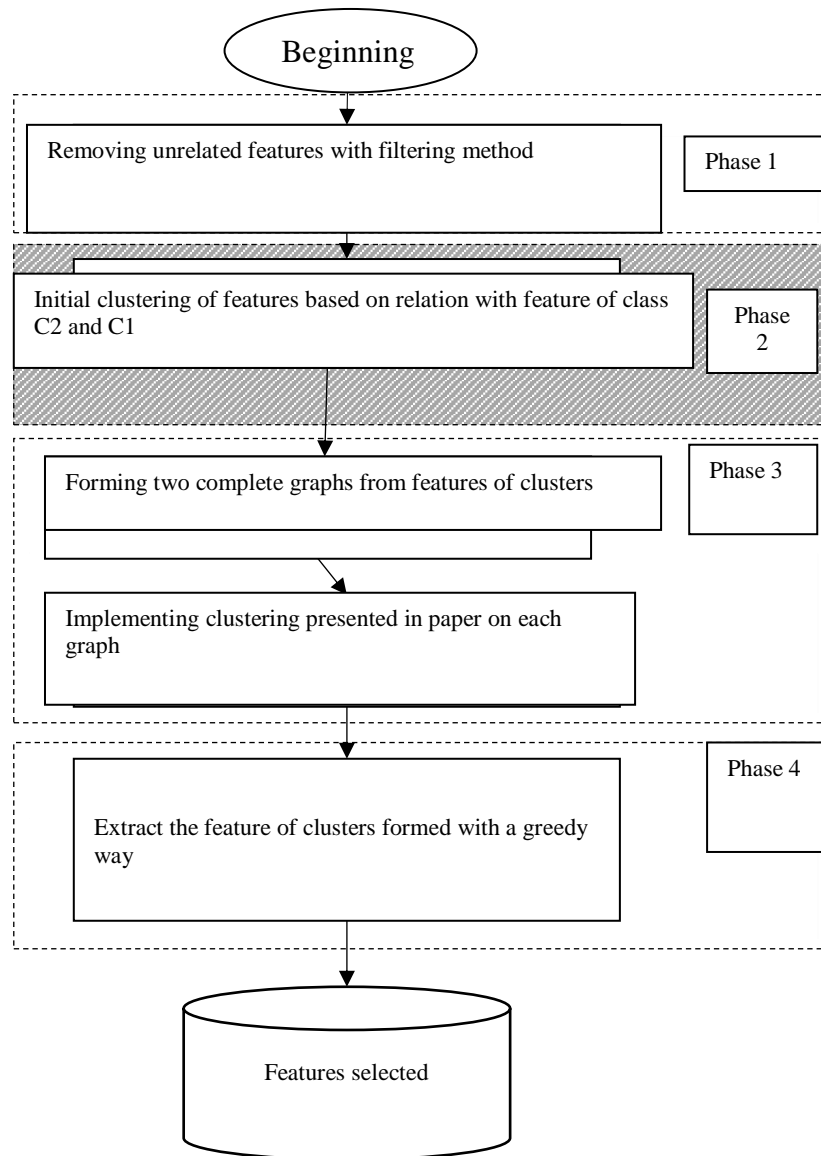
Figure 3- schema of the genetic algorithm used in this thesis

### Correction of Fast Clustering Algorithm (Qinbao et al., 2013)

The proposed approach attempts to focus on the defects of the clustering algorithm used in the previous algorithm. To make the effect of the corrections made meaningful and invoked, the rest of the terms of the article have been kept constant. In other words, how to select features after the clustering process is assumed exactly as the method presented in the paper.

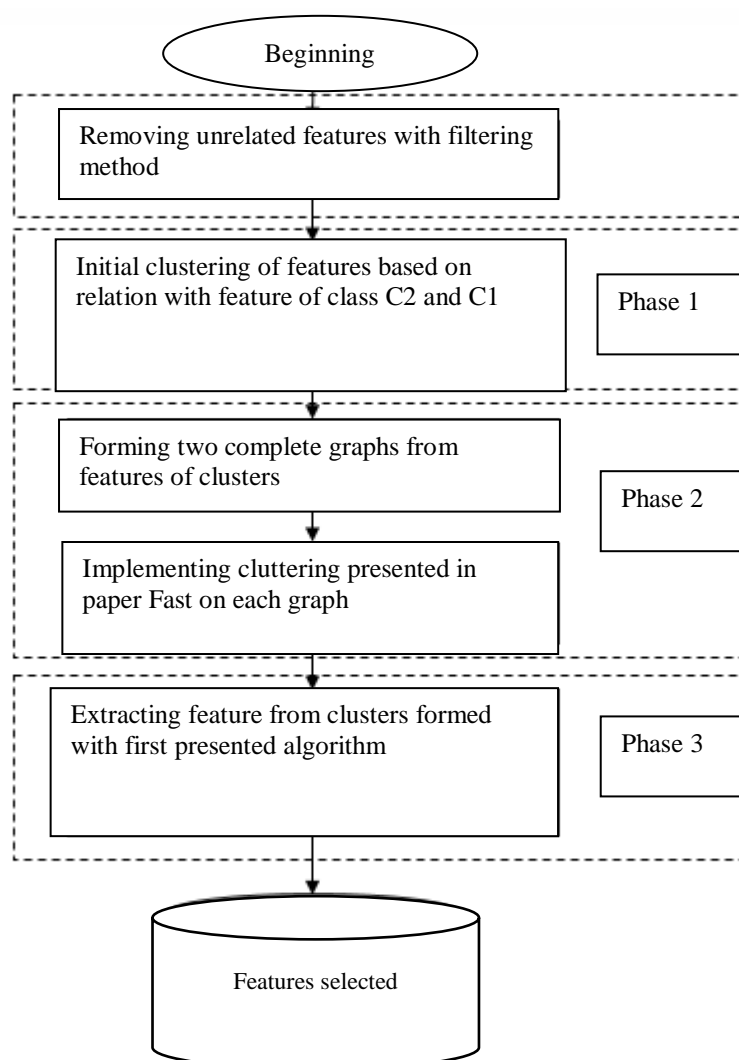
**-Second presented algorithm.**

Figure 4- The stages of the proposed second algorithm



In this figure, it is seen that the steps of the first algorithm shown in Fig. 4 have been added to a phase preceding the formation of the entire features graph. This step actually delivers input features into two separate clusters to the third phase. In the third phase of this algorithm, for each cluster, a complete graph is formed, and then from each graph a minimum spanning tree is obtained. Then the clustering algorithm described in the next section is executed on each tree and the clusters are obtained. To come Finally, in the fourth phase, a representative of each cluster is extracted. This step is carried out in the same manner as the greedy, as described in Qinbao et al., 2013.

**-Third algorithm; modified clustering algorithm + adapted genetic algorithm.**



### Standard dataset

In this research, a large number of standardized dataset categorization has been used for analyzing and evaluating proposed algorithms. These datasets have been used in other authoritative articles that allow us to compare the methods presented with other articles. Table 1 summarizes the specifications of the data used. These datasets can be downloaded from the UCI site.

Table 1- Specifications of the data used

Field of applying	Number of class	Number of samples	Number of input features	Dataset
Dietary	3	178	13	Wine
Medical	2	569	30	WDBC
Signal	2	351	34	Ionosphere
Signal	2	208	60	Sonar
Text	2	3196	37	chess
Image, face	10	2000	77	Mfeat-fourier
Text	2	9822	86	Coil2000
Microarray	2	1391	232	elephant
Microarray	13	452	280	arrhythmia
Text	17	2463	2001	Fbis.wc



**-Set parameters.** The determination of the parameters of the meta-meta-algorithms (inaccurate) is very important and has a huge impact on the efficiency of the algorithm. In this thesis, there are various parameters based on the methods used and developed. At first, the filter threshold parameter is to be determined, the large amount for this parameter eliminates the characteristics of the attribute, while the very small amount for it causes the remaining unrelated features to be greater in addition to the features. In this study, by measuring the amount and interval of correlation of the input characteristics to the class property in different datatypes, the value of 0.05 for this parameter has been selected.

Since the first and third methods are based on the genetic algorithm, the following table describes the parameters of the two methods, which include mutation rate, merger rate, population size and number of generations.

To achieve these parameters, sensitivity analysis has been used by changing a parameter. In other words, in each experiment, one or more data is randomly selected and implemented with a specific stenting algorithm. By observing the results, it was detected which parameter and how to change in the next experiment. In this way, the rest of the parameters were considered constant.

Table 2- The values of the parameters used in the proposed first and third algorithms

Third presented method	Fist presented method	
100	100	Number of chromosomes
50	100	Number of repetition
0.02	0.05	Probability of mutation
0.9	0.9	Probability of merge

### Analysis of proposed methods

In this section, the features of the proposed methods have been analyzed. Initially, the results of testing these methods on standard data are presented in the same way as the test described. The results are presented in the table and it is shown that the two criteria of classification accuracy and the number of output characteristics of the algorithm are shown. These two criteria have actually been reported in many valid algorithms, which makes it easier for us to compare proposed algorithms with the methods in the thematic literature. In this table, the results of each algorithm appear on any data. These data have shown that the proposed algorithms have had satisfactory results. This table clearly shows that the first algorithm (genetic algorithm) is better than the second algorithm in most cases.

The reason for this performance is, in fact, the fact that the greedy method is not a good way to choose the representative of the clusters. That is, while the second algorithm has to conceptually make clustering better, it has not been able to produce good results due to its greedy selection. Despite these justifications, it is not surprising that the third algorithm has better results than both algorithms. This algorithm uses the advantages of both previous proposed methods. It has therefore shown better results. In other words, this method works in the clustering process as a second method, which is better than the first method, and also benefits from the genetic algorithm. In addition to these criteria, further factors are considered to further illustrate the effectiveness of the proposed algorithms to solve the problems.

Table 3- Categorical accuracy and percent reduction of the feature of the proposed methods on the various data sets

Third method		Second method		First method		Data
Accuracy of grouping	Number of features selected	Accuracy of grouping	Number of features selected	Accuracy of grouping	Number of features selected	
97.00	6	95.90	7	95.51	6	Wine

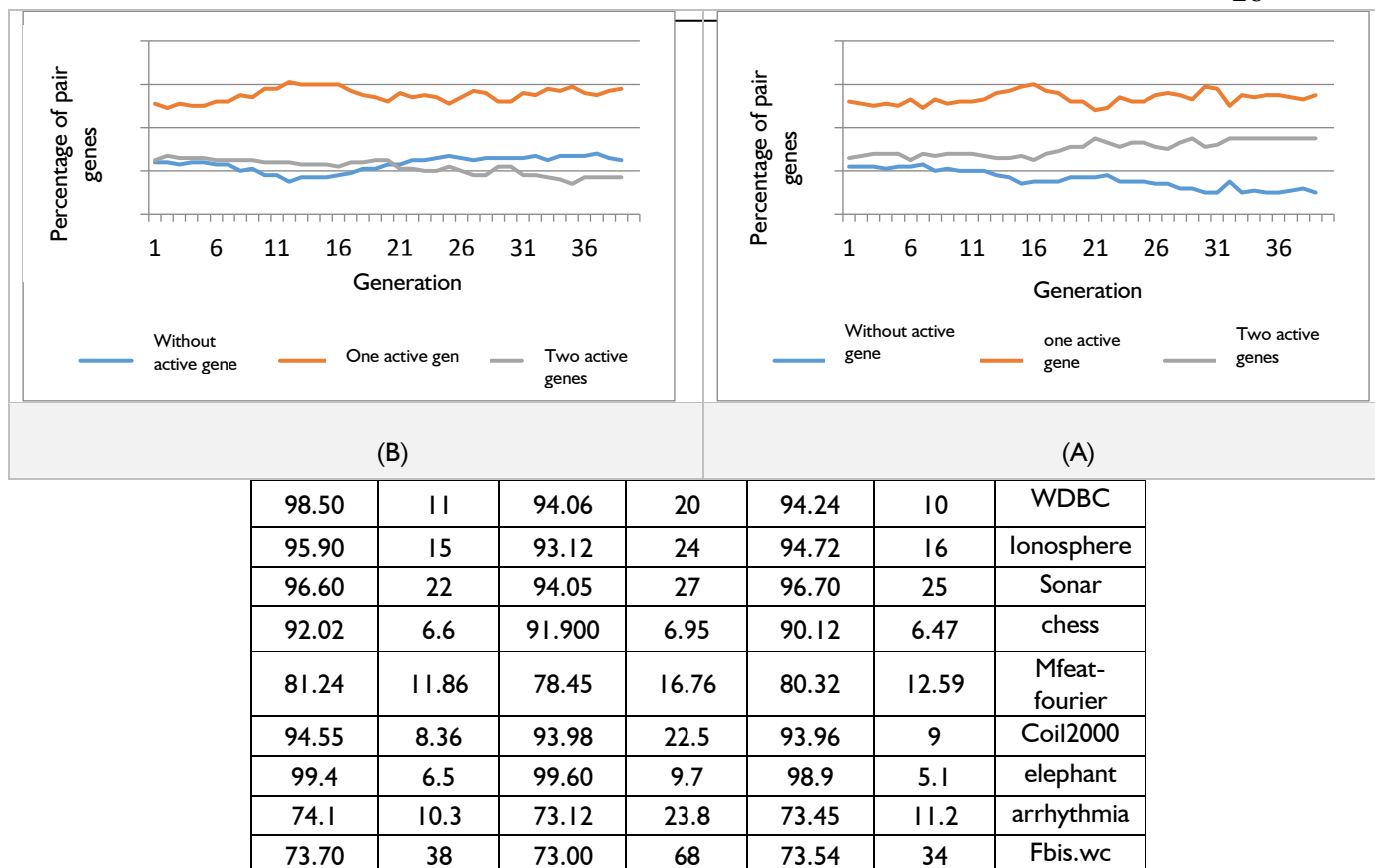


Figure 6- Percentage of pair of genes with zero, one and two active genes in two independent implementations

**-The effect of the presentation on the selection of features in the first proposed method.** In the first approach, it has been claimed that this algorithm chooses the cluster representative better than the Fast algorithm. Because in that algorithm, only one feature is selected from each cluster. While there can be more than one useful feature in a cluster. However, clusters may not have any proper representatives. To measure the performance of this algorithm, a new benchmark has been used. As mentioned above, in this algorithm, for each cluster, one pair of genes is considered to indicate that if the gene value is zero, it indicates the inactivation of the gene and otherwise represents the selection of a property. (Active gene) In fact, at the beginning of the initial population formation, the number of pairs of features with zero, one or two active genes is counted.

This chart is plotted to run the algorithm on the first data. These two charts follow two independent algorithms. By observing these two charts, it is clear that in both of these, the pair of genes that contain an active gene has always dominated two other statistics. However, if pairs of genes active with two active genes and no active genes are precise, it is known that in one of them, the pair of genes dominates with two active genes, whereas in the latter it is less than the number of pairs of genes without The gene is active.

**-The effect of the corrections given on the greedy Fast algorithm.** In this section, the effect of the corrections proposed in the second proposed method on Fast algorithm is analyzed and analyzed. For this purpose, we compare the mean and standard deviation of the number of characteristics of clusters derived from the two algorithms in the following table. As mentioned earlier, in some datasets, the Fast algorithm leads to very large or very small clusters. This is while the modifications made to this algorithm largely cover this shape.

**-The average and standard deviation of the number of produced clusters**

Second presented method		Method Fast		Data
SD	Average	SD	Average	
0.63	1.85	0.69	1.62	Wine
0.48	1.37	0.96	1.19	WDBC
0.48	1.37	0.17	1.03	Ionosphere
0.43	1.75	4.90	1.90	Sonar
0.50	1.59	1.68	1.43	Average

**-Compare proposed methods with other feature selection methods.** In addition to analyzing the data from the proposed algorithms, it is necessary to compare their performance with other methods in this field in order to determine their efficiency and effectiveness against these methods. For this purpose, methods have been chosen to compare the data outcomes. These methods are chosen to be appropriate in addition to the variety of search approaches in terms of being up to date and new. In this regard, the selected methods fall into two categories of filtering methods and methods of dressing.

The following table shows that the number of features selected by method 1 and method 2 is close to the results of rival algorithms. But the third method in most cases has returned fewer features. This is while the accuracy of the classification returned from these algorithms is quite mundane in competing ways. Particularly in the case of the third method, the results are, in many cases, better than rival methods. In order to make the comparison easier, in Table 4-6, each of the methods is scored. This rating is determined by the accuracy of the category. How to calculate this score is that for any data, the algorithm with the highest accuracy of the category has the highest score (integer) given, and the score of the rest of the algorithms is calculated in the same way.

Table 5- compares the accuracy of the classification and the number of selected features of the proposed methods in other ways

Dataset	SGA	HGA 1	HGA 2	HGA 3	BPS O	FCBF	ReliefF	CFS	Consist	Fast	Meth od 1	Meth od 2	Meth od 3
Wine(13) #F	95.5 5	95.51 5	95.51 5	95.51 5	98.60 7	--	--	--	--	96.64 8	95.51 6	95.90 7	97.00 6
WDBC(30) #F	93.9 12	94.06 12	94.06 12	94.06 12	99.70 8	--	--	--	--	93.91 26	94.24 10	94.06 20	98.50 11
Ionosphere(34)) #F	94.3 14	94.93 14	95.56 14	95.21 14	94.30 10.5	--	--	--	--	92.9 33	94.72 16	93.12 24	95.90 15
Sonar(60) #F	95.4 24	95.86 24	95.96 24	96.34 24	94.47 31	--	--	--	--	94.45 33	96.70 25	94.05 27	96.60 22
Chess(37) #F	--	--	--	--	--	92.12 8	88.56 23	90.4 4	89.50 30.0	92.92 6.00	90.12 6.47	91.90 6.95	92.02 6.6
Mfeat- #F	--	--	--	--	--	78.57 38	76.32 76	79.3 19	76.72 12	80.05 15	80.32 12.59	78.45 16.76	81.24 11.86
Coil2000(86) #F	--	--	--	--	--	93.53 7.1	76.96 44	92.5 10.2	84.64 32.7	94.04 7.1	93.96 9	93.98 22.5	94.55 8.36
Elephant(232) #F	--	--	--	--	--	67.96 9	87.54 14	85.9 13	99.94 2	99.47 11.2	98.9 5.1	99.60 9.7	99.4 6.5
Arrhythmia(280) #F	--	--	--	--	--	65.98 13	64.53 140	69.6 26	69.24 25	73.01 28.2	73.45 11.2	73.12 23.8	74.1 10.3

Fbis.wc(2001) #F	--	--	--	--	--	49.79 29	38.75 19	61.2 46	39.20 35	74.04 56	73.54 34	73.00 68	73.70 38
---------------------	----	----	----	----	----	-------------	-------------	------------	-------------	-------------	-------------	-------------	-------------

In order to compare the second category of the data, looking at the sum of the specified rates, the proposed method first returns from the four methods of filtering better results. The second method, which has no better results than this method, has got better results with a very small difference with the first method. But the results obtained from these two methods could not be better than the results reported by the Fast method. But the results of the third method are far better and have been able to excel in this way.

**Table 6: Comparison of methods based on classification accuracy and use of ranking scores**

Dataset	SGA	HGA1	HGA2	HGA3	BPSO	FCBF	ReliefF	CFS	Consist	Fast	Method	Method	Method	maxi
Wine(13)	5	5	5	5	9	--	--	--	--	7	5	6	8	9
WDBC(30)	5	6	6	6	9	--	--	--	--	4	7	6	8	9
Ionosphere(34)	4	6	8	7	4	--	--	--	--	2	5	3	9	9
Sonar(60)	4	5	6	7	3	--	--	--	--	2	9	1	8	9
<i>Total Sum</i>	<i>18</i>	<i>22</i>	<i>25</i>	<i>25</i>	<i>25</i>					<i>15</i>	<i>26</i>	<i>16</i>	<i>33</i>	<i>36</i>
Chess(37)	--	--	--	--	--	7	1	4	2	8	3	5	6	8
Mfeat-fourier(77)	--	--	--	--	--	4	1	5	2	6	7	3	8	8
Coil2000(86)	--	--	--	--	--	3	4	8	2	6	4	5	7	8
Elephant(232)	--	--	--	--	--	1	3	2	8	5	4	7	6	8
Arrhythmia(280)	--	--	--	--	--	2	1	4	3	5	8	6	7	8
Fbis.wc(2001)	--	--	--	--	--	3	1	4	2	8	6	5	7	8
<i>Total Sum</i>						<i>20</i>	<i>11</i>	<i>27</i>	<i>19</i>	<i>38</i>	<i>32</i>	<i>31</i>	<i>41</i>	<i>48</i>

The data from the experiments and the computational results presented in Tables 6 and 5, respectively, show the efficiency of the proposed algorithms on a range of standardized property selection databases. Table 4-6 simplifies the analysis of these results. In a nutshell, it can be assumed that, despite the fact that the number of features selected has not been greater than the rival methods and, in many cases, has been less. The accuracy of the categorization category has, in many cases, been greater than other methods.

Table 7- Comparing the methods based on the accuracy of the classification and the use of ranking scores

Dataset	SGA	HGA1	HGA2	HGA3	BPSO	FCBF	ReliefF	CFS	Consist	Fast	Meth od	Meth od	Meth od	maxi
Wine(13)	5	5	5	5	<b>9</b>	--	--	--	--	7	5	6	8	9
WDBC(30)	5	6	6	6	<b>9</b>	--	--	--	--	4	7	6	8	9
Ionosphere(34)	4	6	8	7	4	--	--	--	--	2	5	3	<b>9</b>	9
Sonar(60)	4	5	6	7	3	--	--	--	--	2	<b>9</b>	1	8	9
<i>Total Sum</i>	<i>18</i>	<i>22</i>	<i>25</i>	<i>25</i>	<i>25</i>					<i>15</i>	<i>26</i>	<i>16</i>	<i>33</i>	<i>36</i>
Chess(37)	--	--	--	--	--	7	1	4	2	<b>8</b>	3	5	6	8
Mfeat- fourier(77)	--	--	--	--	--	4	1	5	2	6	7	3	<b>8</b>	8
Coil2000(86)	--	--	--	--	--	3	4	<b>8</b>	2	6	4	5	7	8
Elephant(232)	--	--	--	--	--	1	3	2	<b>8</b>	5	4	7	6	8
Arrhythmia(280)	--	--	--	--	--	2	1	4	3	5	<b>8</b>	6	7	8
Fbis.wc(2001)	--	--	--	--	--	3	1	4	2	<b>8</b>	6	5	7	8
<i>Total Sum</i>						<i>20</i>	<i>11</i>	<i>27</i>	<i>19</i>	<i>38</i>	<i>32</i>	<i>31</i>	<i>41</i>	<i>48</i>

The data from the experiments and the computational results presented in Tables 7 and 6, respectively, show the efficiency of the proposed algorithms on a range of standardized property selection databases. Table 4-6 simplifies the analysis of these results. In a nutshell, it can be assumed that, despite the fact that the number of selected features is not more than rival methods.

## Conclusion

The proposed methods have been analyzed in terms of computational results from experimental experiments. Initially, the specifications of the data used in this study were stated. In this section, we tried to use data of a variety of applications and a relatively large variety of complexity. Another benefit of these data is that they are used in a variety of articles. In the following, the methods for choosing the new feature as well as the known methods in this area are selected for comparison, and their specifications and references refer to each one. The parameters used in genetic algorithms have been developed and the computational results of all three methods are presented in the presented tables. The reported results for

each data are the accuracy of the category and the number of features extracted by the algorithms. In the next section, the proposed algorithms have been analyzed for the benefit that we expect.

In other words, the proposed algorithm was analyzed in terms of the ratio of zero-cluster, one or two attributes, and it was concluded that the problem raised by the greedy method of the Fast article is serious. Because the ratio of the number of clusters obtained with only one attribute is equal to half the whole cluster. In the same section, the improvements to the Fast algorithm, which are actually the proposed second algorithm of this study, were analyzed and analyzed. The results of this section also showed that the creation of two clusters is better than a general cluster, and also the principle of removing the new edge is better than the rule described in the original article. The proposed third algorithm is also discussed in this section and its merits are specified. To evaluate proposed algorithms, we compared their results with new and valid algorithms in Table 4-5.

This comparison enabled us to make sure that the performance of the proposed algorithms (especially the third algorithm) is superior to the existing methods than the existing ones. To make the comparison easier in Table 7 for each data, we gave each of the algorithms a rating. So that the algorithm that returns the most accurate classification for that data has the highest score in that row. Thus, comparison of algorithms based on the accuracy of the classification in this table is much easier. The results obtained from these algorithms show that the rate of selection of the attribute is as high as the other algorithms and, in some cases, we have also reduced the significant feature. However, the accuracy of the classification of developed algorithms indicates improvement of results. In particular, the proposed third algorithm returns great results.

## References

- Alexandridis, A., Patrinos, P., Sarimveis, H., & Tsekouras, G, (2005), A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models. *Chemometrics and Intelligent Laboratory Systems*, 75(2), 149-162. doi: <http://dx.doi.org/10.1016/j.chemolab.2004.06.004>
- Bekkerman, R., El-Yaniv, R., Tishby, N., & Winter, Y. (2003). Distributional word clusters vs. words for text categorization. *J. Mach. Learn. Res.*, 3, 1183-1208 .
- Blum, A. L., & Langley, P, (1997), Selection of relevant features and examples in machine learning. *Artif. Intell.*, 97(1-2), 245-271. doi: 10.1016/s0004-3702(97)00063-5
- Chuang, L.-Y., Chang, H.-W., Tu, C.-J., & Yang, C.-H, (2008), Improved binary PSO for feature selection using gene expression data. *Comput. Biol. Chem.*, 32(1), 29-38. doi: 10.1016/j.compbiolchem.2007.09.005
- Dhillon, I. S., Mallela, S., & Kumar, R. (2003). A divisive information theoretic feature clustering algorithm for text classification. *J. Mach. Learn. Res.*, 3, 1265-1287 .
- Guyon, I., Andr, #233, & Elisseeff, (2003), An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3, 1182-157 .
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V, (2002), Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.*, 46(1-3), 389-422. doi: 10.1023/a:1012487302797
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273-32 .doi: [http://dx.doi.org/10.1016/S0004-3702\(97\)00043-X](http://dx.doi.org/10.1016/S0004-3702(97)00043-X)
- Torkkola, K, (2003), Feature extraction by non parametric mutual information maximization. *J. Mach. Learn. Res.*, 3, 1415-1438 .