AN EFFECTIVE COMBINATION ALGORITHM TO SOLVE THE PROBLEM OF FEATURE SELECTION USING COMBINATION OF GENETIC ALGORITHM AND CHEMICAL REACTIONS ALGORITHM

Mohammad Saraei, Rahil Analouei, Parvaneh Mansouri

Department. Computer
Azad University of Arak
Arak, Iran

Department. Computer
Azad University of Arak
Arak, Iran

Department. Computer
Azad University of Arak
Arak, Iran

Phone Number: +98-935-6483183

*Corresponding Author's E-mail: mohammadsaraei@gmail.com

Abstract

Feature selection (FS) is a subject in the field of machine learning and statistical pattern recognition. It is a highly important procedure in some processes, such as classification, which contain many features mostly neither usable nor much informative. Although maintaining these features is not problematic in terms of information, it increases computational burden for desired application. It also leads to the storage of useless information, along with useful data. Different methods and algorithms are proposed to tackle this issue. This study used a combination of chemical reaction optimization algorithms and a genetic algorithm through modeling a multi-layer perceptron (MLP) neural network, as of the newest and strongest evolutionary optimization methods introduced in 2010. At the end, the proposed method was compared to available standard dataset. In most cases, results indicated high efficiency of the proposed algorithm. This algorithm was based on two criteria at the same time: accuracy of classification and selection of a small subset of features as prominent ones.

Keywords: feature selection, Classification, Optimization, Genetic Algorithm, Chemical Reaction Algorithm

1. Introduction

Deceleration of classification systems, reduction of efficiency reduction, and escalation of fabrication cost of such systems are among many consequences of FS. Feature selection is an important issue for data analysis in retrieval processes, pattern classification systems, and data mining applications. This process reduces the number of features by removing noisy, as well as irrelevant and iterative features. The feature selection methods are used to reduce computational time, improve performance prediction, and enhance understanding of data in machine learning and pattern recognition programs. As a result, FS refers to a set with minimum number of features that includes necessary and adequate information for the given objective. The feature selection process is
This article intended to provide a new and optimized application for meta-heuristic algorithms in datasets with many useless features to reduce the operation time through reducing the number of such features. This study used a combination of a genetic and a chemical reaction optimization algorithm, called ‘chemical reaction optimization,’ as of the newest and strongest evolutionary optimization methods introduced in 2010. At the end, the proposed method was analyzed and the majority of results indicated its high efficiency based on two criteria at the same time: accuracy of classification, and selection of a small set of prominent features. As a result, this article developed a model for solving feature selection problems, using a MLP neural network and combination of genetic and chemical reaction algorithms. It is worth noting that a standard dataset was used to compare the proposed method with other methods. Due to the nature of the proposed algorithm, a new solution generation mechanism was designed to reduce operation time. This article addressed the history of the most important works in this area and presented details of the [proposed] algebraic model in Sections 2 and 3, respectively. Section 4 introduced the methodology and rational of the given algorithm. The proposed algorithm was explained in Section 5. Simulation and final conclusion were dealt with in Sections 6 and 7, respectively. The last part of the article included acknowledgment and references.

2. Literature review

Feature selection is a well-known issue in machine learning, aiming to select a small set of features. The resulting class distribution, given only the values for the selected features, should be as close as possible to the original class distribution given all feature values [9]. The feature set can be relevant, irrelevant, or deprived. The FS process should look for only relevant features to minimize feature space size, reduce computational and economic costs, and improve prediction accuracy in classification. All FS algorithms have two key components: evaluation strategies and search strategies. Given the evaluation strategies, feature selection algorithms can be classified into two types: filter-based and wrapper-based methods. The feature selection is a discrete optimization problem. It is worth noting that the search space in the FS includes all features of such sets.

\[
\sum_{L=1}^{N} \binom{N}{L} = \binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = 2^N - 1
\]  

(1)

Where, N stands for the number of main features and L is the length of a set of current features.

Each FS sample includes some independent variables, which are directly measurable. They are fed into a system to produce one or more dependent variables. The feature selection aims to select some of these independent variables (features or inputs), that a logical relationship between independent and dependent variables is maintained and the number of independent variables (target or outcome) is minimized. Different FS methods intend to select the best subsets from 2N candidate subset. In general, FS methods are categorized into different groups based on the search type. In some FS methods, the whole possible space is searched. In other methods, which could be
heuristic search and/or random search, a degree of efficiency is lost for the sake of achieving a smaller search space. To have better categorization of FS methods, the FS process in all methods is divided into following parts:

**Generation Procedure:** this procedure selects a candidate set for the given method.

**Evaluation Function:** it evaluates the given subset, based on the given method, and returns a number as the fitness value of the method. Different methods seek out subsets that optimize this value.

**Termination Criterion:** it decides the termination time.

**Validation Procedure:** It decides whether the selected subset is valid.

According to subset search strategies, the FS algorithm can be divided into comprehensive, random, and heuristic strategies. In the comprehensive strategy, all subsets are evaluated and the most prominent feature is selected. In the random search, the search process is conducted randomly. In the heuristic search, the feature space is searched via specific strategies, such as successive selection. The large size of feature space is a classification issue. The majority of features are irrelevant and redundant and have negative effects on the efficiency of classifiers. As a result, feature selection aims to reduce the feature space and enhance classification efficiency. Many different feature selection methods and algorithms have been used [by different researchers]. For example, the genetic algorithm has been applied to solve feature selection problem [13, 14]; the HPSO-LS has been developed for selecting the less-dependent feature set and also the superior set through local search strategy inside the particle swarm optimization [11]; a feature selection method was proposed based on the distance and exclusive information [5]; the feature selection problem was solved using bat and firefly algorithms [3, 10]; and the uncorrelated linear discriminant analysis (ULDL) was used to develop the ULDA/QR and ULDA/GSVD, which significantly improved optimization [8]. Some classical feature size reduction methods have used the heuristic and filter-based functions. The main problem of such methods is their computational complexity. To tackle this problem, a new method was developed and the accuracy of results was empirically examined by introducing it to a standard data set [6]. In a study, the particle swarm algorithm was used to solve this problem [16]. Another study applied feature selection and extraction methods to improve text classification performance [7]. In an experimental evaluation, two methods were developed to solve the problem given the range of feature selection [12]. One of them was genetic algorithm-based heuristic method, and the other one was based on the harmony search. In [4, 15], a fuzzy method was employed to solve this problem. A study investigated and compared available methods and algorithms to solve the feature selection problem [1].

### 3. Main methodology

#### 3.1. Basic Concept of Chemical Reaction Algorithm

This algorithm, developed by Albert in 2010, has been inspired by molecular function in chemical reactions to achieve minimal energy. As we know, particles naturally tend to decrease their internal energy. It is assumed that the molecules are in a closed system. As a result, they either collide with each other or the system’s wall. This algorithm is inspired by molecular processes and reactions. Each molecule is a set of atoms and traits, such as their number. We presented each molecule with a profile to show its features. These features are:

- **W:** Indicates the structure of each molecule or the efficient solution.
- **PE:** a potential energy (PE) is defined for each molecule, indicating its competence.
- **KE:** a kinetic energy (KE) is defined for each molecule.
numHit: a numHit feature is defined for each molecule, indicating the number of its collision.

minHit: minHit feature is defined for each molecule, indicating the minimum number of its collision.

localMin: a localMin feature is defined for each molecule to keep the minimum solution value observed currently by this molecule. The chemical reaction optimization (CRO) is an evolutionary algorithm inspired by modeling molecules, molecular energy, and intermolecular collisions.

The CRO phases are as follows:

1) Giving value to the CRO parameters
2) Producing initial population

To this end, a structure or profile is defined for each molecule, specifying its features. At this stage, the main loop of the algorithm is presented as follows:

Following phases are repeated until FE is less than FELimit:

A random number in the range between 0-1 is generated. If this number is higher than Molecoll or if population size is 1, the algorithm moves to the second phase; otherwise, it goes to the third phase.

3) The decomposition or on-wall collision is applied to the molecule as follows:

The decomposition or on-wall collision is applied to the molecule as follows:

A molecule is randomly selected from a population.

If the combination condition is fulfilled, it is applied on two molecules to produce two new molecules. If the combination operation succeeded, the initial two molecules are replaced with the new molecule.

If the combination condition is not fulfilled, the intermolecular collision is applied to upgrade the given two molecules. It then moves to phase 4.

4) At this phase, the whole population is searched and the molecule with PE<min global is selected as the optimal solution. Then the algorithm moves to phase 1.

The obtained results are presented after termination of the main loop.
Other features of chemical reaction optimization:

1) The CRO can be programmed in an object-oriented programming language, in which a class represents a molecule.

2) The CRO can be easily run in parallel, as the population size does not need to be synchronized between computing units. proposed algorithm

The initial population can be generated randomly by each algorithm. In this article, the initial population was generated using the CRO algorithm. Then, it was changed into a binary problem to be used in a binary genetic algorithm. This was because the feature selection is a binary problem. The generated initial population was organized and divided into the strong and weak parts based on the population member cost. The strong and weak groups of the population were assigned to the CRO and GA, respectively. This process was repeated for 10 times. The removal of weak groups and use of strong groups manifest an elitism system. In this problem, each molecule is a logical solution to the problem (each chromosome in GA). As it was said above, the neural network was used for modeling the problem. In other words, it was recalled for solutions and introduced to the cost function after computing the neural network error or modeling error with Equation 2 in order to compute its cost according to Equation 3. In fact, the neural network was used for the cost function of the algorithm. The Levenberg-Marquardt algorithm was used in the neural network, in which 70%, 15%, and 15% of data were training, validation, and test data, respectively. For feature selection, the combination of genetic algorithm and chemical reaction algorithm was used to increase the rate and accuracy of the process. The combination of methods produced better result than each other method used separately.

3.2. Phases of Proposed Algorithm

1) The initial population is continuously and randomly generated by the chemical reaction algorithm.
2) The generated population is organized and divided into the strong and weak groups.

3) The strong part is introduced to the CRO for improvement (the CRO is a continuous algorithm and the feature selection problem has a binary coding).

4) The weak group is introduced to the genetic algorithm for improvement.

5) The recursive populations from both algorithms are combined, and the generated population is organized based on the cost function.

6) The information is introduced to the neural network for evaluation of fitness and calculation of the cost function.

7) The weak group of the population is removed (using the elitism system).

8) The remaining group is divided into the weaker and stronger groups.

9) Again, the weak group is introduced to the genetic algorithm and the strong group to the CRO.

If the termination criterion (reaching the maximum possible repetitions) is not fulfilled, the algorithm is moved to phase 2; otherwise, the current solution is presented as the output.

3.3. Cost Function

In this article, each population member is introduced to the neural network for three times and the mean classification error is obtained. The reason behind using error rate of three-time recalling of the neural network was to reduce the degree of randomization of the neural network's solution. The mean MSE was placed in Equation 3, and then the solution cost or fitness was obtained. It is worth noting that following parameters were considered for modeling and computing the neural network error:

\[ Z = \text{objective function}, \quad \text{MSE} = \text{neural network error}, \quad B = \text{cost of using feature}, \quad w = \text{importance of number of features}, \quad \text{and } NF = \text{number of features}. \]

\[ Z = \text{MSE} + w.n_f \]  
\[ Z = \text{MSE} (1 + \beta.n_f) \]

**Figure 3:** The proposed algorithm's flowchart
4. Simulation and computational result

In the current study, we introduced a new hybrid algorithm, ‘CROGA,’ using MLP, CRO, and GA algorithms to be used for feature selection problems, which are amongst the NP-Hard problems. Results indicated the accuracy of the proposed method in selecting better features and avoiding local minima. Although the proposed method produced good results, it took a long time to select the best subset. Therefore, this algorithm is recommended for important problems, which require highly accurate results. In this article, an effective method for solving feature selection problems was employed, using MATLAB R2013a on a desktop with Intel Core-i3 processor and 4GB RAM, running on Windows 7. To solve the feature selection problem, the standard data proposed by MATLAB was tested with the proposed algorithm. Then, the proposed method was compared to the genetic and chemical reaction algorithms. In the given problem, each dataset was repeated for 20 times. Simulation results indicated that the proposed algorithm was highly flexible and simple for implementation and can be regarded as a strong optimization algorithm. The aim was to reduce the number of features. In this regard, the proposed method could achieve the desirable solution with few features. The chemical reaction algorithm has better performance than its counterparts and can be regarded as a suitable solution to difficult hybrid and large-sized problems. For large-sized datasets, the genetic algorithm alone cannot find the best solution; whereas, the proposed algorithm searches the search space until the desirable feature is obtained. The adjustable parameters for implementing the algorithms are presented in Tables 1-2.

Table 1: Genetic algorithm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of generation</td>
<td>20</td>
</tr>
<tr>
<td>Population size</td>
<td>10</td>
</tr>
<tr>
<td>Crossover Coefficient</td>
<td>0.8</td>
</tr>
<tr>
<td>Mutation Percentage</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 2: Chemical reaction algorithm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Generation</td>
<td>20</td>
</tr>
<tr>
<td>Population size</td>
<td>10</td>
</tr>
<tr>
<td>molecule</td>
<td>0.2</td>
</tr>
<tr>
<td>buffer</td>
<td>1e6</td>
</tr>
</tbody>
</table>

In addition, the number of considered features and samples for each dataset is presented in Table 3. The performance of the genetic, chemical reaction, and proposed algorithms corresponding to the standard sample with 10 repetitions, on average, was compared. The samples used in the current study were body fat, house pricing, and chemical (Table 4). In this table, the best value, mean value, worst value, standard deviation, and AVG time in 10 repetitions are presented in the Action Column.
In addition to this table, Tables 3 and 5 show that the proposed algorithm used how many of the given features to produce the desirable result as compared to algorithms in the current study.

5. Second and following pages

With rapid growth of information in the world, instruments used in different fields of knowledge should be proportionally developed to be able to respond to this growth. Traditionally, data is analytically converted into knowledge. In many applications, these traditional data analysis techniques are slow, costly, and highly cognitive.

Table 3: Data specification

<table>
<thead>
<tr>
<th>Dataset</th>
<th>N of Feather</th>
<th>N of Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body fat</td>
<td>13</td>
<td>252</td>
</tr>
<tr>
<td>House pricing</td>
<td>13</td>
<td>506</td>
</tr>
<tr>
<td>Chemical</td>
<td>8</td>
<td>498</td>
</tr>
</tbody>
</table>

Table 4: The number of features is the best solution for each dataset by the algorithm

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Action with Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Action</td>
</tr>
<tr>
<td>Body fat</td>
<td>Best</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Worst</td>
</tr>
<tr>
<td></td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AVG Time</td>
</tr>
<tr>
<td>House pricing</td>
<td>Best</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Worst</td>
</tr>
<tr>
<td></td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AVG Time</td>
</tr>
<tr>
<td>Chemical</td>
<td>Best</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Worst</td>
</tr>
<tr>
<td></td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AVG Time</td>
</tr>
</tbody>
</table>
Table 5: The number of features is the best solution for each dataset by the algorithm.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GA</th>
<th>CRO</th>
<th>CRO-GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body fat(13)</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>House pricing(13)</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>chemical(8)</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

In fact, such data analysis methods become gradually useless in many areas with rapid growth of information. As a result, the use automated methods is felt. Knowledge discovery is the process of identifying valid, innovative, potential, useful, and understandable projects in data largely made of preprocessed data. Feature selection is an important step in data preprocessing. The feature selection cornerstone is the removal of a set of input features that provide few information or lack data useable for prediction. In addition to increasing the accuracy and efficiency of classification, the feature selection leads to better understanding of results and often development of a model with greater generalizability to new data. With reducing the number of features, dataset becomes smaller and a better representative of the overall data. Feature selection is an optimization problem, which belongs to the hard-NP category. The meta-heuristic algorithms have proven powerful in solving optimization and hard-NP problems. After investigating new meta-heuristic algorithms, inspired by the nature and community, this thesis proposed the combination of genetic algorithm and chemical reaction algorithm for solving feature selection problems. Comparison of the proposed algorithm with some genetic and chemical reaction algorithms showed the accuracy and efficiency of the

![Figure 4: Compare the cost of algorithms for Body Fat dataset](image-url)

![Figure 5: Compare the cost of algorithms for House pricing dataset](image-url)
6. CONCLUSION

Feature selection is an important and fundamental stage of pattern recognition, machine learning, and data mining. It seems that the algorithm combination approach can be applied to other local search methods (ant colony optimization, tabu search, etc.) and its results can be compared to findings of the current study. The application of the proposed hybrid algorithm can be examined in specific problems, such as the use of feature selection in identification of risk factors of the collision insurance.

7. ACKNOWLEDGEMENTS

Hereby, all people who contributed to the study by their help and guidance are thanked. The objective behind writing this article was to solve a related research-applied problem in the future.

8. References


