# Intelligent Discovery of Novel Classification Rules Based on Genetic Algorithms

Fahd Nasir A. Alwesabi[1], Ahmed Sultan Al-Hegami[2]
[1]University of Science and Technology
Computer Science & Information Technology Department
Sana'a University
Yemen

[2] Assistant Professor of Artificial Intelligence and Intelligent Information Systems
Sana'a University
Yemen
f.alwsabi@ust.edu.ye, aalhegami@suye.ac

***ABSTRACT:*** *Data mining technique deals with the problem of discovering novel and interesting knowledge from huge amount of data. This problem is often performed heuristically when the extraction of patterns is difficult using standard query mechanisms or classical statistical methods. Data mining researchers have studied subjective measures of interestingness to reduce the volume of discovered knowledge to ultimately improve the overall efficiency of KDD process. In this study, we pushed the novelty measure into a genetic algorithm to form constraints to the algorithm to discover only novel and hence interesting patterns. The proposed approach has a flexible chromosome encoding technique that uses Bayesian theorem where each chromosome corresponds to a classification rule. The proposed approach makes use of a hybrid approach that uses objective and subjective measures to quantify novelty of rules during the discovery process in terms of their deviations from the known rules. We experiment the proposed framework with some public dataset and tested using real life applications. The experimental results are quite promising.*

**Keywords**: Data mining, KDD, Classification, Genetic algorithm, Interestingness, Rule Discovery, Novelty Measures

**Received:** 1 March 2010, Revised 29 March 2010, Accepted 3 April 2010



## 1. Introduction

Most recently, information collection has become easier, but retrieve relevant pieces process of it required effort and has become significantly greater, especially in a huge databases. As a consequence, there has been a growing interest in powerful tools capable of facilitating the discovering of interesting and useful information within such a huge amount of data.

The process of knowledge discovery brings together the latest research in statistics, databases, machine learning and artificial intelligence. The core of this process is the application of machine learning-based algorithms to databases.

There are two basic ways for performing data mining process are the supervised and the unsupervised learning. In addition there are several techniques for performing data mining such as neural networks and genetic algorithms, etc...

There are many challenges that face data mining techniques, One of the challenges is that most of the discovered patterns and knowledge are not interesting and useful, and the data mining algorithms may discover huge amount of patterns which requires mining of mined patterns.

There are two main measures that have been proposed to quantity the patterns interestingness, namely objective and subjective measures. Objective measure is based on statistics and structures of rules and patterns such as confidence and support. On the other hand, yjr dubjective measure is based on user's belief about the field such as unexpectedness, novelty, and actionability. Data mining researchers have studied subjective measures of interestingness to reduce the volume of discovered rules to mainly improve the overall efficiency of KDD process.

In this paper, we study novelty of the discovered rules as subjective and objective measures of interestingness. We propose an approach based on genetic algorithms (GAs) to perform the task of classification and to find not only accurate knowledge but also comprehensible and novel knowledge. The proposed approach extends the work proposed by [4], which quantity the novelty in both objective and subjective aspects. Our work push the novelty criterion presented in [4] into a genetic algorithm to form a constraints to the algorithm to discover only novel and hence interesting patterns.

**2. Problem Statement**

The core problem for study in this research is how to develop models using data mining techniques at two different times $T_{t_i}$ and $T_{t_{i+1}}$ , or develop based on the normal approaches, which is usually at different two points of time, where the discovered may be repeated, conformed, generalized, and specialized knowledge and hence not interesting.

Our proposed approach discovers patterns that are interesting, smaller in size and generate novel rules with acceptable degree of accuracy. This is a useful feature to reduce the volume of data on escalating over the time and hence the user background knowledge is monotonically augmented.

In addition, the approach guarantees that a chromosome is pruned during constructing phase and will certainly be pruned as the model is built first and then pruned using novelty criterion. This strategy saves time and effort required to build the model.

**3. Related Work**

There are many proposals that studied the novelty and interestingness rules proposed in the literatures. We briefly review some of them below.

In [1], authors propose an incremental association rules based on data mining algorithm that integrates shocking interestingness criterion during the process of building the model. A new interesting measure called shocking measure is introduced.

There are many features of the proposed approach. One of the main this features is to confine the user background knowledge, which is monotonically augmented. The incremental model that reflects the changing data and the user beliefs is attractive in order to make the over all KDD process more effective and efficient.

In [2], author presents an algorithm based on attribute information gain which can combine between the subjective and objective evaluation method together to discover interesting classification rules. The algorithm allows the users themselves to set the weight of each attribute's information gain, and the weights can reflect they prefer; with different weights the algorithm can discover different interesting classification rules.

The main disadvantage is that the accuracy of the interesting rules is not near to 1, and it costs too much time because the authors have to run the algorithm for each class especially for the large database.

In [3, 4], the quantification of novelty is performed objectively and user participation is sought for categorization of rules (as novel, unexpected, generalized, specialized, conformed) based on novelty measure, but the approaches used do not mine and extract new rules.

In [5], a new efficient type of genetic algorithm (GA) called uniform two-level GA is proposed as a search strategy to discover actually interesting, high-level prediction rules, Although the process of generalized rule induction requires a lot of computations, which is usually not satisfied with the normal algorithms, it was demonstrated that this method has coped with the problems of GAs such as divergence of genetic search process and remaining stuck on local solution of genetic search, and rapidly found interesting rules.

In [6], the authors proposed a framework to quantify the novelty in terms of computing the deviation of currently discovered knowledge with respect to domain knowledge and previously discovered knowledge. The approach presented is intuitive in nature and lays more emphasis on user involvement in quantification process by way of parameter specification. In the present work, the quantification is performed objectively and user involvement is sought in categorization of rules based on novelty index.

In [7], the novelty is estimated based on the lexical knowledge in WordNet. The proposed approach defines a measure of semantic distance between two words in WordNet and determined by the length of the shortest path between the two words (wi,wj). The novelty is defined as the average of this distance across all pairs of the words (wi,wj), where wi is a word in the antecedent and wj is a word in the consequent.

In [8, 9, 10, 11], generally, all these studies construct a model of training set that is selected to contain no examples of the important (i.e. novel) class. Subsequently, the mechanisms detect the deviation from this model by some way.

In [12], the use of GAs for rule discovery in the application of data mining has been studied in some another studies, These algorithms are based on the Michigan approach in a way that each rule is encoded in a chromosome and the rule set is represented by the entire population.

In [13], authors discussed objective measures of interestingness rule and the degree is calculated by an information theoretical measure. Briefly, while some of the researchers discuss objective measures, the others propose subjective measures to assess interestingness rule.

In [14], authors partitioned interestingness measures into objective and subjective measures, and further partition subjective measures into those that capture unexpectedness and those that capture actionability.

The paper [15], build classifier system that have modified the individual encoding method to use non-binary representation, do not encode the consequents of rules into the individuals and use extended version of crossover and mutation operators suitable to their representations. Further they do not allow rules to be invoked as a result of the invocation of other rules, and define fitness functions in terms of some measures of classification performance.

Our proposed approach will use some techniques for data mining and evaluate rule interestingness in both objective and subjective Manners during mining process based on novelty measure.

**4. Classification and Genetic Algorithms**

The classification process represents one of the most studied in data mining. In essence, the problem consists of assigning records to one out of a small set of pre-defined classes, by discovering some relationship between attributes. Each record consists of a set of predicting attributes and class value attribute to be predicted [25].

A data-mining algorithm is applied to a set of training data set namely examples, with a known class, to discover rules detecting some relationship between the predicting attributes and the goal attribute. This relationship is then used to predict the class value attribute of examples whose class is unknown.

The discovered knowledge from classification process is usually represented in the form of IF-THEN prediction rules, which has the advantage of being a high-level, symbolic knowledge representation, contributing to the comprehensibility of the discovered knowledge.

The discovered rules can be evaluated according to several criteria, such as the degree of confidence in the prediction, classification accuracy rate on unknown-class examples, comprehensibility, etc. We emphasize that this latter is a crucial criterion in the context of data mining.

Genetic Algorithm (GAs) is a search method that has been widely used in applications where the size of the search space is very large. In essence, GAs are “search algorithms based on the mechanics of natural selection and natural genetics” [25]. GAs are based on the principle of survival of the fittest, where the fittest individuals are selected to produce genes namely offspring for the next generation. In the context of search, individuals are candidate solutions to a given search problem. Hence, reproduction of the fittest individuals means reproduction of the best current candidate solutions. Genetic operators such as selection, crossover and mutation generate genes from the fittest individuals.

One of the main advantages of GAs over “traditional” search methods is that the former performs a kind of global search

using a population of individuals, rather than performing a local, hill-climbing search. Global search methods are less likely to get trapped into local maxima, in comparison with local search methods.

It is interesting to overall note that, the knowledge discovery paradigm most used in data mining is still rule induction. Most of the algorithms in this paradigm perform a kind of local search.

**5. The SUFA-GA and Novelty Measure**

The GA used in this work to incorporate novelty criterion into GA, which is a novelty measure developed by [5].
The next subsections describe several aspects of the proposed algorithm, namely Encoding and Settings the parameters, compute the probability of belongs each class to dataset, population process, Crossover and mutation process, and Novelty as a Fitness measure.

**5.1 The SUFA-GA Framework**

The general scheme of proposed approach is illustrated in Figure (1)**.**

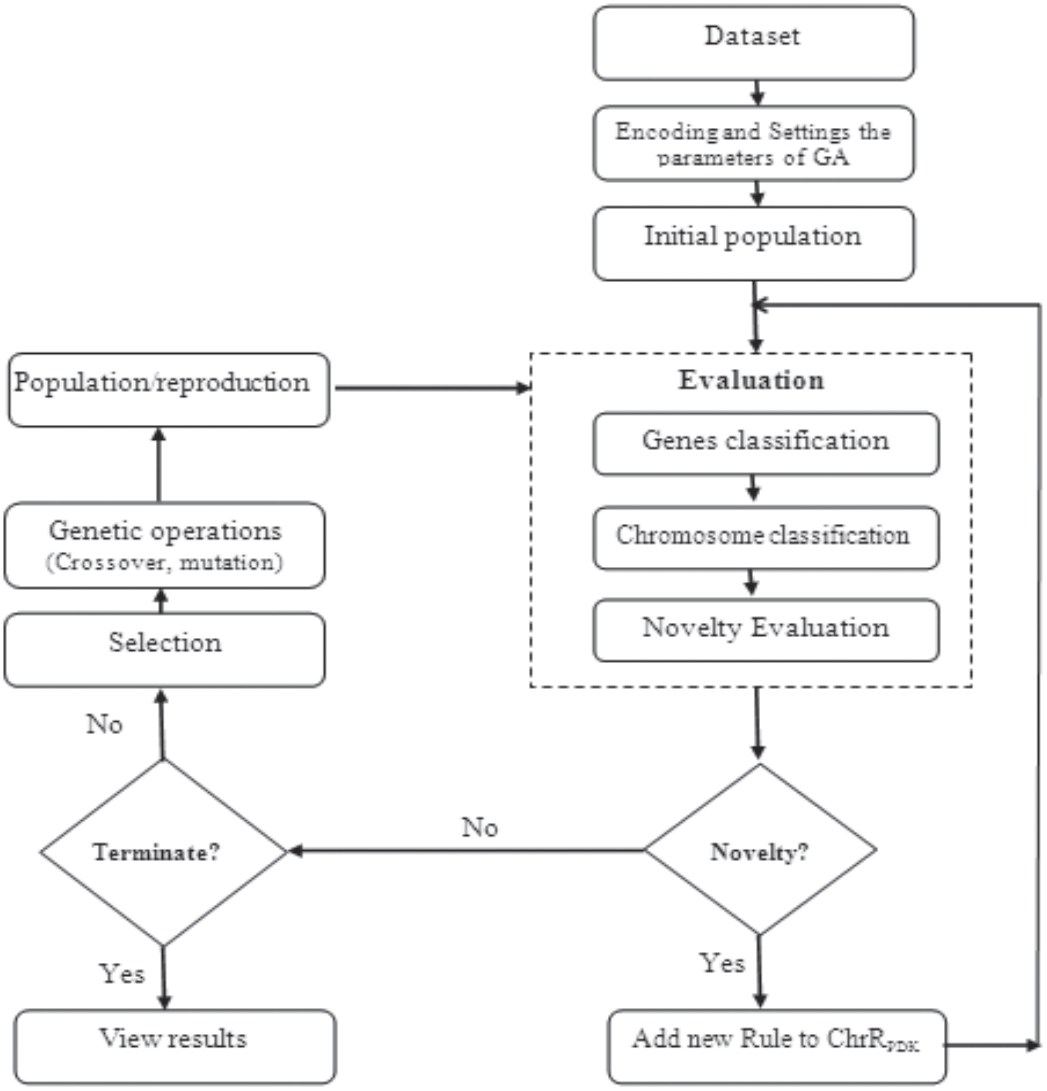


Figure 1. The SUFA-GA Framework

**5.2 Encoding and Settings the parameters**

To encoding and setting the SUFA-GA parameters, Let **n** be the number of predicted attributes in the dataset being mined. Then a chromosome is collected of **n** genes, where each gene corresponds to a condition containing one attribute. (*gene**1*** represent *Att. **1**,., gene **n*** represent *Att. **n***).

Each gene is partitioned into **two** fields relational operator (*Oi* ), value (*Vi* ), and some other fields based on (k) that represent the number of classes(goal Att.) in the dataset, we represent it's as ($W_1$ to $W_k$) as shown in figure (2).
A chromosome corresponds to the entire IF part of the rule and each gene corresponds to one condition in this IF part, the chromosomes do not involve the class predicted by a rule as shown in figure (3).

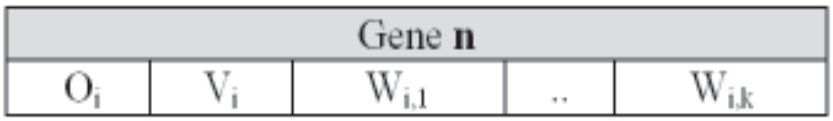


Figure 2. Gene representation


Figure 3. Chromosome representation

*Where,*

- ***n*** = number of predicted attributes in the data being mined.
- **K** = number of classes in the dataset.
- ***$O_i$*** **=** The relational operator field is indicates the relational operator used in the $i^{th}$ condition and based on Attribute type as:
  - IF attribute ($G_{Ai}$) is ***nominal/categorical,*** this field can involve the operators ***(= , ≠)***
  - IF attribute ($G_{Ai}$) is ***continuous,*** this field can involve the operators ***( < , ≤ , > , ≥)***.
- **$V_i$** = The Attribute value, This field can involves one of the values belonging to the domain of the attribute.
- **$W_{i,k}$** = The weight of $G_{Ai}$ in every class in range[0 , 1], this field is partitioned into k sub-fields, each sub-field can be involve value at range [0,1] to indicate the percentage of belonging this Att. (gene) into a every class.
- **$N_{DI}$ =** The degree of novelty field is indicates the fitness value for each chromosome in the populations.

So, before encoding process the proposed approach should be determine the some important parameters as **n** , **k, $G_L$ , $C_L$ , $R_L$.**

*where*:

- **n** = number of attributes that will be use it for mining.
- **k** = number of classes in the dataset.
- **$G_L$** = The gene length = dynamic, based on ***k*** (number of classes in the dataset).
- **$C_L$** = The chromosomes length = Fixed = [ n * (2 + k) + 2]
- **$R_L$** = The Rule length = Fixed = [n + 1].

- $\mathbf{P_S}$ = The Population Size.
- $\mathbf{M_C}$ = The Maximum Cycles.
- $\mathbf{C_R}$ = The Cross Over Rate.
- $\mathbf{M_R}$ = The Mutation Rate.
- $\mathbf{T_S}$ = The threshold value.

**5.3 Bayesian Approach to Determine Classes**

In this process we compute the percentage of each class in the dataset using the Bayesian theorem as shown in equation (1) [16, 18]:

$$P(X|Ci) = \frac{P(X|Ci)P(Ci)}{P(X)} \quad (1)$$

*Where,*

- P(X) = The number of training tuples in D.
  - D = is a training set of tuples and their associated class labels.
- P(Ci) = The number of training tuples of class Ci in D.

**5.4 Population Process**

The process of generating new populations is based on the number of maximum cycles of GA.

Programmatically we represent a chromosome as two-dimension array, where the rows represent the chromosomes in populations, and column represent the values of gene attributes for each chromosome in populations.

Figure (8) demonstrates how the population process is working for generating and predicting the Populations in the proposed approach, and the following steps are used:

1. Determine the number of population.
2. Create a chromosome array based on the number of populations.
3. Take the chromosome from population.
4. Set the random values from dataset for $O_i$ , $V_i$ fields for the $j^{th}$ gene.
5. Computation the probability percentage of belongs the $j^{th}$ gene on the current chromosome to each classes $\mathbf{P(G_{Wj})}$, computation it based on attribute type as the following:

- IF $\mathbf{G_{Ak}}$ is **categorical**, in this case we should use the following Equation (2):

$$P(Gwj|Ci) = \frac{P(Gvi)}{P(Ci)} \quad (2)$$

*where*:

- **P(Ci)** = The number of training records of class *Ci* in D.
- $\mathbf{P(G_{Vj})}$ = The number of records of class *Ci* in D having the value $G_{Vj}$ for $G_{Ak}$
- $\mathbf{G_{Ak}}$ = The $k^{th}$ gene attribute.

- IF $G_{Ak}$ is **continuous-valued**, in this case we need to do a bit more work, but the computation process is pretty straightforward. A continuous-valued attribute is typically assumed to have a Gaussian distribution with a mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$ of the training data.

So, to normalize the training data, it is necessary first to compute the sample mean, and the standard deviation of each feature in this dataset, and then normalize the data using the Equation (3):

$$P(GWj \mid Ci) = \frac{Gvi - \mu}{\varepsilon} \quad (3)$$

6. Repeat steps (4, 5) for each gene, exception the last two indexes of chromosome.

7. Compute the probability percentage of belongs $i^{th}$ chromosome to which class using the Equation (4):

$$P(X|Ci) = \prod_{k=1}^{n} p(Gwj|Ci)$$

$$= [\, P(G_{W1} \mid C_i) * P(G_{W2} \mid C_i) * \ldots \; P(G_{WK} \mid C_i) \,] * P(C_i) \quad (4)$$

8. Predict the consequent class label of X, P(X|Ci)P(Ci) is evaluated for each class Ci and take which has the maximum value[16]. *if and only if*

$$Ci = \begin{cases} Ci, if\ P(Ci) > P(Ci+1) & , \\ Ci+1, if\ P(Ci) < P(Ci+1) & , \\ \theta, if\ P(Ci) = P(Ci+1) & , \end{cases} \quad (5)$$

In case Ci =θ, the rule is called a dummy rule. Such rules may also result if none of the patterns match the consequent class.

9. Put the maximum value of class, and the number or name of class that all gene values belong to it in the last position – 1 of chromosome.

10. Compute the degree of novelty, and put the result in the last position of chromosome.

11. Repeat steps (4, 5, 6, 7, 8, 9, 10) for each chromosome in population.

**5.5 Crossover and Mutation Process**

The chromosomes which selected by selection process habituation to dealing with it as new population, then crossover occurs to exchange information between randomly selected parent chromosomes by recombining parts of their genetic materials.

We have used stochastic tournament selection with tournament size 3 and two-point crossover, with crossover probability = 100%, and two points(genes) are selected at random in range (1, n), and the two middle parts in the two chromosomes are interchanged.

*Note*: The proposed approach should set weights by zero value for each class in all genes before execute the crossover operation between two chromosomes.

When two point-crossovers occur randomly between Pc1 and Pc2, then assuming that, the two children are formulated as:

$\mathbf{C_{C1}}$ = { Gene X1 , ***Gene Y2*** , Gene X3 , Gene X4 , ***Gene Y5***}
And
$\mathbf{C_{C1}}$ = {Gene Y1, ***Gene X2***, Gene Y3 , Gene Y4 , ***Gene X5***}

*Where,*

$\mathbf{P_{C1}}$ = parent chromosome 1
$\mathbf{P_{C2}}$ = parent chromosome 2
$\mathbf{C_{C1}}$ = Child chromosome 1
$\mathbf{C_{C1}}$ = Child chromosome 2

We also used an selective reproduction strategy, where the best individual of each generation was passed unaltered to the next generation.

We developed two mutation operators adapted for our genome representation, namely operator mutation and value mutation. Each of these operators occurs on a different field of a gene. We used mutation rates of 50% for each kind of mutation.

- **The operator mutation** modifies the relational operator currently being used in a condition of the rule, by replacing it with another one, randomly generated among the valid operator's based on type of attribute is categorical or continuous as:
  - IF $G_A$ is categorical then operator mutation modifies between "=" and "≠" operators.
  - IF $G_A$ is continuous then operator mutation modifies between "<" , "≤" , ">" and "≥" operators.
- **The value mutation** modifies the contents of the field value, by replacing the current value with another one randomly generated. There are two possible cases:
  - IF the attribute type is categorical, the mutation process occurs to replace the current field value with another value belonging to the domain of the attribute.
  - IF the attribute type is a continuous, the mutation produces a small number related by limit percentage of value which mutate (for example 20%), then added to or subtracted from the current field value. This is implemented in such a way that the lower and upper bounds of the domain of the attribute are never exceeded.

**5.6 Novelty as a Fitness Measure**

The SUFA-GA approach evaluates the degree of novelty for each rule during the mining process by fitness function as:

1. Look for in $\mathbf{ChrR_{PDK}}$ about rule that has highest corresponding at conjuncts (genes) to new rule in $\mathbf{ChrR_{CDK}}$ .
2. Compute the grade of novelty for new rule using the equation(6):

$$\Psi(S1,S2) = \frac{\{|S1|+|S2|-2+k)+\sum_{i=1}^{k}\delta(G_1^i,G_2^i}{|S1|+|S2|} \quad [4,6] \qquad (6)$$

*Where,*

- $\mathbf{S_1}$ , $\mathbf{S_2}$ = are two gene sets(chromosome) with cardinalities |S1| and |S2| respectively.
- K = the pairs of compatible genes between S1 and S2.
- $G^i_1$, $G^i_2$ are the $i^{th}$ pair of compatible genes.

- IF (Ψ(S1,S2) > threshold) **ChrR**(i), then is novel, add **ChrR**(i) from $\mathbf{ChrR_{CDK}}$ to $\mathbf{ChrR_{PDK}}$.

**5.7 The SUFA-GA Algorithm**

```
Begin
  Compute the probability of classes into the training set
  Initialize the population P(n)
  While the termination condition is not met do
    begin
        // Fitness function
        For each ChrRi (Ai → Ci) in ChrR_CDK
        begin
            Find ChrRj (Aj → C)j from ChrR_PDK , such
            that Ø (Ai , Aj) is minimum.
            Compute Ø (Ci , Cj) //novelty measure
            IF ChrRi is Novel, add to ChrR_PDK. [4, 6]
      end
```

*Select the best chromosomes from population p(n) based on fitness value.*
*Execute crossover and mutation process for produce the second genes(offspring).*
*Replace, based on fitness, candidates of p(n), with these offspring.*
*end*
*End*

**6. Experimental Study**

The proposed approach is implemented and tested using public dataset available in [26]. Since, there are no other approaches available, which pushed incorporating novelty during mining process; we could not perform any comparison against our approach.

We considered these datasets as evolving with time, and partitioned them into 2 increments: D1, and D2 mined at times $T_1$, and $T_2$ respectively. We took each of these partitions to be equal for purpose of generating rules.

The following experiment was conducted to show the effectiveness of the framework using iris dataset and based on the following parameters:

- **n** (number of attributes) = 4.
- **k** (number of classes) = 3.
- $\mathbf{G_L}$ (Gene length) = 5.
- $\mathbf{C_L}$ (chromosomes length) = [ n * (2 + k) + 2] = 22.
- $\mathbf{P_S}$ (The Population Size) = 4
- $\mathbf{M_C}$ (The Maximum Cycles) = 40 (20 at $T_1$ and 20 at $T_2$)
- $\mathbf{C_R}$ (The Cross Over Rate) = 100%
- $\mathbf{M_R}$ (The Mutation Rate) = 50%
- $\mathbf{T_S}$ (The Threshold value) = 0.5

For simplicity, we state running the proposed algorithm by 4 individuals as an initial population. Also we assume that the algorithm will look for a solution in 40 generation.

The dataset was randomly portioned into two parts, with 2/3 of the records used for training and 1/3 of the records used for testing the accuracy of the discovered rules.

The SUFA-GA approach discovered 7 novel rules only during 40 generation, two rules belong to the Iris-setosa class, two rules belong to the Iris-virginica class, and three rules belong to the Iris-versicolor class.

Table (1) presents the final 7 rules discovered by the proposed approach. For each class, the proposed approach was run two times, varying the random seed used to generate the initial population with size equal to 4.

The best rule of the two runs, according to its fitness value measured on the training set and degree of novelty, was selected as the rule predicting that class (this is the rule shown in Table 1). Once the 6 rules were discovered they were evaluated on a separate test set to compute the accuracy of model.

**Note** that the set of rules used for classification are shown in Table 1 also includes a default rule, that represent a rule with no conditions which is automatically applied when no other rule has its conditions satisfied by the example to be classified. The rule number(1) in table(1) simply predicts the Iris-setosa class.

For each rule in Table (1) the third column shows two values, the first one indicates to the degree of novelty for the antecedent of the rule, and the second one indicates to the degree of novelty for the consequent of the rule (class value), namely the degree of novelty of the rule computed by equation 6 in the training set and in the test set, respectively.

| Rule | Degree of novelty |
|---|---|
| IF Sepal length= 5 ^ Sepal width= 3.2 ^ Petal length= 1.2 ^ Petal width= 0.2 → **Iris- setosa** | [1,1] |
| IF Sepal length= 6.7 ^ Sepal width= 3 ^ Petal length= 5.2 ^ Petal width= 2.3 → **Iris- virginica** | [1,1] |
| IF Sepal length= 5.5 ^ Sepal width= 2.4 ^ Petal length= 3.8 ^ Petal width= 1.1 → **Iris-versicolor** | [1,1] |
| IF Sepal length= 6.7 ^ Sepal width= 3 ^ Petal length= 1.2 ^ Petal width= 0.2 → **Iris- versicolor** | [0.5,1] |
| IF Sepal length= 5.4 ^ Sepal width= 3 ^ Petal length= 1.8 ^ Petal width= 0.1 → **Iris- setosa** | [0.5,1] |
| IF Sepal length= 7.7 ^ Sepal width= 3.8 ^ Petal length= 5.4 ^ Petal width= 2.2 → **Iris- virginica** | [0.75,1] |
| IF Sepal length= 6.7 ^ Sepal width= 3 ^ Petal length= 1.5 ^ Petal width= 0.1 → **Iris- versicolor** | [0.5,1] |

Table 1. Discovered Rule Set for the Iris Dataset

This time, however, the rules discovered from the training set did not generalize so well for examples of the test set. This seems to be due to the fact that this is a considerably more difficult classification problem.

Anyway, it is interesting to evaluate the performance of the set of discovered rules as a whole, by measuring the accuracy rate and novelty measure. The accuracy rate of the discovered rule set was 74% (out of 50 examples, 36 were correctly classified).

**7. Conclusion**

In this research, we proposed a strategy for rule set reduction based on the Novelty measure of the rule. We pushed the novelty criterion into a genetic algorithm to form a constraint to the algorithm to discover only novel and hence interesting patterns.

Novelty index of a newly discovered rule is the quantification of its deviation with respect to the known rule set. User subjectivity is captured by specification of threshold(s) for rule categorization.

The framework is experimented and evaluated using real-life datasets and results have been presented. The generated rules were categorized as conforming, generalized/specialized, unexpected and novel rules.

Future work should consist of more experiments with other datasets, as well as more elaborated experiments to optimize several parameters of the algorithm, such as mutation rates, the Limit threshold for the weight field, etc.

(Recall that the results of this paper were achieved without any serious attempt to optimize the parameters of the GA.) In addition, we will implement our proposed approach as a program in the near future.

**References**

[1] Yafi, E., Al-Hegami, A. S., Alam, M. A., Biswas, R.(2009). Incremental Mining of Shocking Association Patterns, *In:* Proceedings of World Academy of Science, Engineering and Technology, V. 37, January 2009, Dubai, UAE.

[2] Zhou, Y., Xia, S., Gong, D.,Youwen, L (2007). Knowledge Discovery of interesting Classification Rules Based on Adaptive Genetic Algorithm, ISKE.

[3] Bhatnagar, V., Al-Hegami, A. S., Kumar, N (2005). A hybrid approach for Quantification of Novelty in Rule Discovery, *In:* Proceedings of International Conference on Artificial Learning and Data Mining (ALDM'05), Turkey, Feb. 25-27, 2005, p 39-42.

[4] Bhatnagar, V., Al-Hegami, A. S., Kumar, N. (2005). Novelty as a Measure of Interestingness in Knowledge Discovery, *In:*

*International Journal of Information Technology*, 2 (1).

[5] Alatas, B., Arslan, A.(2005). Mining of Interesting Prediction Rules with Uniform Two-Level Genetic Algorithm, *International Journal of Computational Intelligence*, Fall.

[6] Al-Hegami, A. S. Bhatnagar, V.and Kumar, N. (2004). Novelty Framework for Knowledge Discovery in Databases, *In:* Proceedings of the 6th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2004), Zaragoza, Spain, 2004, p 48-55

[7] Basu, S., Mooney, R. J., Pasupuleti, K.,V., Ghosh, J.(2001).Using Lexical Knowledge to Evaluate the Novelty of Rules Mined from Text , *In:* Proceedings of the NAACL workshop and other Lexical Resources: Applications, Extensions and Customizations.

[8] Marsland, S. (2001). On-Line Novelty Detection Through Self-Organization, with Application to Robotics , Ph.D. Thesis, Department of Computer Science, University of Manchester.

[9] Japkowicz , N., Myers, C., Gluck, M (1995). A Novelty Detection Approach to Classification, *In*: Proceedings of the 14th International Joint Conference on Artificial Intelligence.

[10] Roberts, S., Tarassenko, L (1994). A Probabilistic Resource Allocation Network for Novelty Detection, *In*: Neural Computation, 6 (2).

[11] Ypma, A., Duin, R (1997).Novelty Detection Using Self-Organizing Maps, *In:* Progress in Connectionist-Based Information Systems. Volume 2.

[12] Choenni, S.(2000). Design and Implementation of a Genetic-Based Algorithm for Data Mining, *In:* Proc. of the 26th Int'l Conf. on Very Large Data Bases, Cairo, Egypt, 2000, p. 33-42.

[13] Suzuki, E., Kodratoff, Y.(1998). Discovery of surprising exception rules based on intensity of implication, *In*: Proc. 2nd European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-98), Lecture Notes in Artificial Intelligence, 1510, p. 10-18,.

[14] Silbershatz, A., Tuzhilin, A. (1996). What makes patterns interesting in knowledge discovery systems, *IEEE Transactions on Knowledge and Data Engineering* 8 (6) 970–974.

[15] McAulay, A.D., Oh, J.C. (1994). Improving Learning of Genetic Rule-Based Classifier Systems, *IEEE Trans. on Systems, Man, and* Cybernetics, 24 (1) 152-159.

[16] Kantardzic, J. B. Data Mining—Concepts, Models, Methods, and Algorithms, IEEE Computer Society.

[17] Han, J., Kamber, M. (2006). Data Mining: Concepts and Techniques, University of Illinois at Urbana-Champaign, Second Edition.

[18] Gyorodi, R. Data Mining & Knowledge Discovery.

[19] Goldberg , D.E. (1989). Addison-Wesley, Genetic Algorithms in Search, Optimisation and Machine Learning, 1989 rec.

[20] Mitchell, M. (1996). Introduction to Genetic Algorithms, MIT Press.

[21] Liu, B., Hsu,W.(1996). Post Analysis of Learned Rules, *In:* Proceedings of the 13 th National Conference on AI (AAAI'96).

[22] Murphy, P. M., Aha, D.W.(1994). UCI Repository of Machine Learning databases, [http://www.ics.uci.edu /~mlearn/MLRepository.html]. Irvine, CA: University of California, Departament of Information and Computer Science.

[22] Niblett, P. C. (1987). T. Induction in Noisy Domains. *In*: Progress in Machine Learning (from the Proceedings of the 2nd European Working Session on Learning), p.11-30, Sigma Press.

[23] Demiroz, G., Govenir, H.A (1998). Ilter N. Learning differential diagnosis of erythemato squamous diseases using voting feature. Artificial Intelligence in Medicine, v. 13, p. 147-165.

[24] Padmanabhan, B., Tuzhilin, A. (1997). Unexpectedness as a Measure of Interestingness in Knowledge Discovery, Working paper # IS-97-. Dept. of Information Systems, Stern School of Business, NYU.

[25] Fidelis, M. V., Lopes, H. S., Freitas, A. A.Discovering Comprehensible Classification Rules with a Genetic Algorithm.

[26] http://kdd.ics.uci.edu/

[27] http://www.comp.nus.edu.sg/~dm2/p_download.html

[28] http://www.kdnuggets.com