The Combine of revelations of ant colony optimization algorithm and genetic algorithm based on a fuzzy logic to enquiring high dimension datasets

Sadigheh Asyaban , s_asyaban@yahoo.com
Department of Computer, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran

Navid Samimi Behbahan , navid_samimi@yahoo.com
Department of Computer, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran

Zahra Eram , zahraeram@gmail.com
Department of Mathematics, Behbahan Branch, Islamic Azad University, Behbahan, Iran

Paper Reference Nu mber: 10
Name of the Presenter: Navid Samimi Behbahan

Abstract

What always has been important for researches of the data mining field is inventing of methods to extract the knowledge of high dimension datasets. So in this paper, a new method has been recruited to dataset by combination of fuzzy logic, ACO algorithm and genetic algorithm. One of advantages of this method is decreasing of the investigated parameters. For extract of features (reduction of the dimension of datasets), two methods of Principal Component Analyse and Fisher's linear discriminant have sequentially recruited at the pre-processing step. In order to evaluate of the proposed method, some datasets from the resource of data mining of UCI which have many characteristics, were selected and were investigated. In comparison with other method, we found a promotion in the results.

Key words: Datasets, ACO, GA, Fuzzy Rules, Sequence of PCA and Fisher

1. Introduction

According to extend dataset and its complexity, nowadays there is a need to a more effective and useful implement in order to discover a useful knowledge about them searching data is a process help human in such discovery and recently is used in widespread fields.

Janecek et al. (2008) Studied related between division and selection of features and surveyed the effect of dimension reduction by analyzing main part in division. Finding explains accuracy based on PCA is depended on dataset and changes in parts which expresses a basic need to division. Finally, two datasets of email and addiction used to indicate accuracy of the proposed method.

Duangsaithong and Windeatt (2009) presented a way to reduce dataset In collection which has last of feature and a little samples and reduce division new dataset is created by omitting repeated and irrelevant data. The result is more accuracy and time for calculating for example, for lung cancer they has reduced 56 features and found 11 features.

Heyward et al. (2008) Surveyed application of searching data projects using regression of symbolic logic on different cancer’s dataset. Finding expresses pre-proceed data will improve application of a classifying algorithm, if the feature be chosen properly.

In a research presented by Duangsoithong and Windeatt (2009) they proposed a combined method of selecting feature using algorithm of non-pragmatism polarity and genetic for using of their benefits. This consists of 2 phases. Filtering phase in which omits non- pragmatism polarity feature and guides us to establish a basic group of genetic algorithm. Packing phase in which searches the extreme (goal) dataset features.

Assare et al. (2008) proposed a combined random model for classification used space and domain of sample simultaneously to increase diversity in classification, result from ovary cancer datasets. Indicate usefulness of the model in datasets with many features (many dimension)

In the paper, problem of knowledge discovery from datasets with many dimensions has taken into attention. After several pre-processing phases and using a combine of fuzzy rules and evolutionary algorithms, knowledge of input datasets is presented. The goal of this research is reducing the volume of great datasets using several pre-processing phases for selecting best feature and then classifying with fuzzy systems based on sets of fuzzy if-then rules to discover needed knowledge. The processes of using the algorithms of proposed method are show in figure

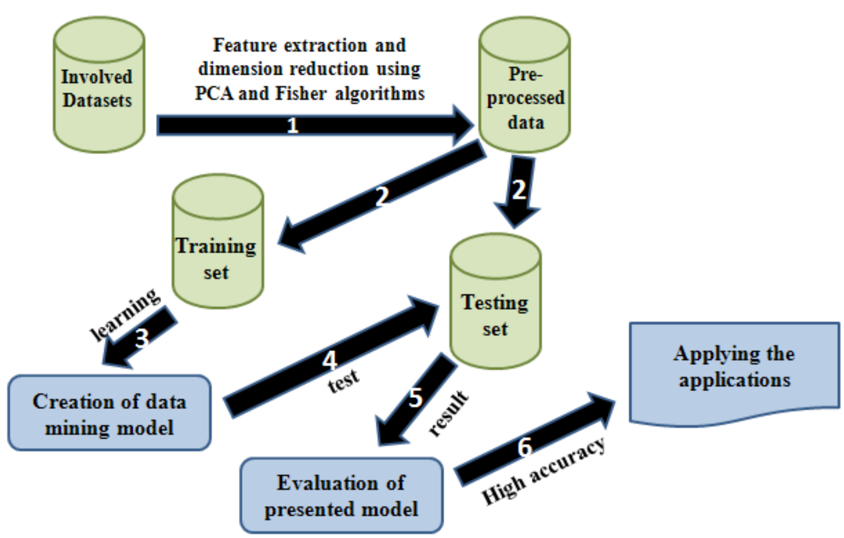

Fig 1: the processes of using algorithms of proposed method

Reducing the volume of dataset is taken into consideration because achieving knowledge performs in much less time and more accuracy. This knowledge can be a base of fuzzy logic that will improve with recovery algorithm during searching data process. For improving the ET of fuzzy logic, combining of genetic algorithm and ant colony is used. These algorithms are based on evolution and statistical concepts.

These algorithms in coordinating with each other searches set of if-then rules related to their cases to have more efficiency. Finally, presented system datasets with high-dimension from searching data store of California University is used.

2. Dimension reduce using Sequence of PCA and Fisher's Linear Discriminant Projection

The PCA, with its resulting linear Karhunen-Lőve projection, provides feature extraction and reduction that are optimal from the point of view of minimizing the reconstruction error. However, PCA does not guarantee that selecting (reduced) principal components as a feature vector will be adequate for classification (will have discriminatory power).

On the other hand, the linear Fisher transformation is designed optimally from the point of view of maximal interclass separability of projected patterns. However, we recall that the linear Fisher transformation faces problems when the within-class scatter matrix Sw becomes degenerate (noninvertible). This can happen when the number of cases (objects) in a dataset is smaller than the pattern dimension. One solution to this problem is first to transform a dataset with high dimensional patterns into a lower dimensional feature space, for example by using a KLT transform, and then to apply a discriminative linear Fisher projection to the lower-dimensional patterns. The entire projection procedure of the original high dimensional (n-dimensional) patterns **x** into the lower-dimensional (c-dimensional), more discriminative feature vectors can be decomposed into two subsequent linear transformations:

– PCA with the resulting Karhunen-Lőve projection into the m-dimensional principal component feature vectors **y.**

– A linear Fisher projection of m-dimensional principal component vectors **y** into c-dimensional Fisher discriminative feature vectors **z**.

The algorithm for the Karhunen-Lőve-Fisher transformations is as follows.

Given: A dataset *T*, containing *N* cases ($\mathrm{X}^{i}, c^{i}_{target}$) labeled by associated categorical classes (with a total of l classes)

1. Extract from dataset *T* only the pattern portion represented by the $N\times n$ matrix X with n-dimensional patterns $\boldsymbol{x}$ as rows.
2. Select a dimension m for the feature vectors containing principal components (already projected by the Karhunen- Lőve transformation), satisfying the inequalities $l \le m \le N - l$.
3. Compute, for the data in matrix X, the m×n-dimensional optimal linear Karhunen-Lőve transformation matrix $\boldsymbol{W}_{KL}$.
4. Transform (project) each original pattern **x** onto a reduced-size m-dimensional pattern **y** by using the formula $\mathbf{y} = \boldsymbol{W}_{KL}\,\mathbf{x}$ (or, for all projected patterns, by using the formula $\mathbf{Y} = \mathbf{X}\,\mathrm{W}_{\mathrm{KL}}^{\mathrm{T}}$).
5. Select a dimension c for the final reduced feature vectors $\boldsymbol{z}$ (projected by Fisher's transformation), satisfying the inequality $c+1 \le l$.
6. Compute, for the projected data in matrix **Y**, the c ×m-dimensional optimal linear Fisher Transformation matrix $\mathbf{W}_{\mathrm{F}}$.
7. Transform (project) each projected pattern $\boldsymbol{y}$ from $\boldsymbol{Y}$ into the reduced-size c-dimensional pattern **z** by using the formula $\mathbf{z} = \mathbf{W}_{\mathrm{F}}\,\mathbf{y}$ (or, for all projected patterns, by the formula $\mathbf{Z} = \mathbf{Y}\mathbf{W}_{\mathrm{TF}}$).

Result: The $m\times n$-dimensional optimal linear Karhunen-Lőve transformation matrix is $\mathbf{W}_{KL}$. The c×m dimensional optimal linear Fisher's transformation matrix is $\mathbf{W}_F$. The projected pattern matrix is $\boldsymbol{Z}$. (Krzysztof et al. 2007)

3. System based on fuzzy logic

This part expresses how to show knowledge using fuzzy rules and models classification. Then presents the formation of basic logic rules, classification and certainty of each rule and continues by show classifying a new model (sample) and surveys the use fuzzy reason. (Nakashima and Ishibuchi .2005).

3.1. Pattern classification problem

Classification of models is a matter with n dimensions C classes and m educational models ($X_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1,2,\ldots,m$). Without lasting totality of matter, each educational model's feature is normalized. It means models' space for each feature is a number in [0,1] independently. In this research, we use of if-then fuzzy rules base of classification systems.

$$\text{Rule } R_j : \text{if } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \quad \text{Then Class } C_j \text{ with } CF_j\ ,\ j=1,2,\ldots,N \tag{1}$$

R_j: *j*th if-then fuzzy rule.
$A_{j1}, \ldots, A_{jn}$: antecedent fuzzy sets on the unit interval [0, 1], like Large, Medium, Small and overcast each other.
C_j: achieved class for each rule
N: number of if-then fuzzy rules
CF_j: grade of certainty of R_j rule.
They use as basic fuzzy collections from triangle fuzzy collections.

3.2. Used fuzzy model

As show in figure 3, the used fuzzy collection is a 5 parts fuzzy collection. For coding each part, a number is taken into consideration.

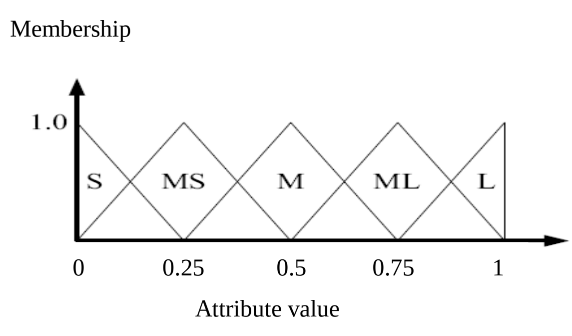

Fig 2: Proposed model to partition fuzzy sets

S (small) = 1
MS (medium small) = 2

M (medium) = 3
ML (Medium large) = 4
L (large) = 5

3.3. Establishing basic fuzzy Rules

Each rule is coded by numbers from 1 to 5. At first a collection is random established consists of 100 rules in witch each one is a branch with long of dataset's features and surveyed their recognition scales.

In many of similar works, primary rules are established independently from educational datasets. Although this way in view of time complexity is suitable but because these rules are not with good education, they have high recognition on only education datasets. So it is better to establish basic rules randomly and always try to improve them.

The way of calculating related class and grade of certainty in each if-then fuzzy rule for the best rule in fuzzy classification system according to rules is as bellow processes.

to determining the class (C_j) and grade of certainty of jth rule (CF_j), we have following steps:

Stage 1: calculating $\beta_h(R_j)$ for class h (h=1,…,C)

$$\beta_h(R_j) = \sum_{x_{p\in Class\ h}} \mu_{j1}(x_{p1}) \times \ldots \times \mu_{jn}(x_{pn}) \qquad h = 1,2,\ldots,C \tag{2}$$

Stage 2: finding class ($\hat{h}$) in witch it has maximum amount of $\beta_h(R_j)$

$$\beta_{\hat{h}}(R_j) = \max\{\beta_1(R_j), \beta_2(R_j), \ldots, \beta_C(R_j)\} \tag{3}$$

If more than one class have maximum amount, R_j rule can't exclusively assigned to C_j class. In this condition, we take it empty ($C_j = \phi$). If only one class have maximum amount, then $C_j = Class\ h$

Stage3: if only one class have maximum amount of $\beta_h(R_j)$, then grade of certainty for j rule (CF_j) will identify by formula 4.

$$CF_j = \frac{\beta_{\hat{h}}(R_j) - \bar{\beta}}{\beta_h(R_j)} \tag{4}$$

Where

$$\bar{\beta} = \frac{\sum_{h\neq\hat{h}} \beta_h(R_j)}{C-1}$$

3.4. Fuzzy argument:

With helping the way of producing a rule in previous part, we can produce randomly, N fuzzy rules then we define witch rule is related to witch class and grade of certainty for all N if-then fuzzy rules. The class of a new sample (x) is defined as below:

Stage 1: calculating $\alpha_h(x)$ for class h and *h = 1,2,…,C*

$$\alpha_h(x) = \max\{\mu_j(x) \times CF_j \mid C_j = Class\ h,\ \mathrm{j} = 1,2,\ldots,\mathrm{N}\}, h = 1,2,\ldots,C \tag{5}$$

Where

$$\mu_j(x) = \mu_{j1}(x_1) \times \ldots \times \mu_{jn}(x_n) \tag{6}$$

Stage 2: finding a class for sample (h_p^*) in witch it has maximum amount of $\alpha_h(x)$

$$\alpha_{h_p^*}(x) = \max(\alpha_1(x), \ldots, \alpha_C(x)) \tag{7}$$

If more one class had maximum amount, then sample x doesn't classify. Otherwise, class h_p^* specialize to sample x.

4. Proposed method to improve fuzzy rules

Here we use a combine of genetic algorithm and ant colony to product new rules for improving them. Using genetic algorithm, available rules will change and produce new rules. With helping of ACO we try to abscond (slip) of local extremums.

4.1. Process of corresponding ACO and classification of fuzzy if-then rules

ACO is an ultra-exploratory method that considered as an idea of searching optimization for avoids local optimum solution and converging to overall optimum solution.

Parameters of ACO algorithm for classification of rules:

*Number of ants existed in ACO: number of rules used to classification

*Ants feature: characters and parameters related to range of different datasets.

*Amount of pheromone related to selected Ant: value of evaluation function related to the Ant.

* Update of pheromone related to selected Ant: optimization of classification fuzzy if-then rule set.

*Current ACO: current rule set.

*Change of current ACO: change in current rule set.

*New ACO: new rule set.

*Calculation of fitness related to new ACO: calculation of value of evaluation function for new rule set.

*accepting of new ACO with a special probability and if degree of fitness related to the new ACO is more than that of current ACO: acceptation of new rule set if value of evaluation function of new rule set is more than that of current rule set.

In above process in first, an Ant colony is created at random and fitness of current colony is calculated using fitness of all Ants and pheromone residue from them. Value of fitness function is equivalent to number of samples is correctly classified.

Then one of the Ants is selected at random and of course with a special probability, and the ant change a number of its feature using mutation operator in GA, then update a mount of its pheromone and if fitness of colony is better, this change inform to other ant. (eynipour et al. 2009)

4.2. Overall plan of proposed algorithm steps

1) Pre-processing data
   * feature extract
   * Data normalization
   * Data fuzzy making
2) Creating a primary set of fuzzy if-then rules, class and grade of certainty of each rule.

3) Evaluating cost of current rule set using evaluation function.
4) Changing the rule set using mutation operator in GA over one rule.
5) Calculating the fitness of new rule set using evaluation function.
6) Replacing current rule set by new rule set if fitness of new rule set is more than that of current rule set, of course with a special probability.
7) Repeating the step 4 to 6 in given number.
8) Return best rule set.

4.3. Combining the GA and ACO for changing the current rule

In previous section, method of creating a primary rule set and method of evaluation the rules has described, and as we expected accuracy of the rules was not high because the rules has created at random. For increasing the accuracy of predicting the rules must use a method and antecedent part of the rule set must be changed so that achieve to maximum accuracy. For this, we used combine of GA and ACO that refereed in previous.

Input of the step is a fuzzy rule and output of that is a likely improved version of the fuzzy rule. This improvement is performed by mutation operator (of GA operators) in current fuzzy rule.

ACO is able to perform best search for best changes in during run period. Basing the combined algorithm, several changes are done in one rule in each step. Tendency of each ant to mutation (change) in *i*th antecedent of current rule R_C to A_j is the following:

$$P_a(R_C,i,A_j)=\frac{\lfloor\tau(R_C,i,A_j)\rfloor\lfloor\eta(R_C,i,A_j)\rfloor}{\sum_{u=1}^{5}[\tau(R_C,i,A_u)][\eta(R_C,i,A_u)]} \tag{8}$$

In equation 8, τ is value of pheromone and η is initiative probability that achieved using the following equations:

$$\eta(R_C,i,A_j)=\frac{N(i,A_j)}{\sum_{v=1}^{5}N(i,A_j)} \tag{10}$$

$$N(i,A_j)=\sum_{p=1}^{m}n_p(i,A_j) \tag{11}$$

$$\sum_{p=1}^{m}n_p(i,A_j)=\begin{cases}1 & \text{avg}\max_{v=1}^{5}\{\mu_{A_v}(x_{pi})\}=\mu_{A_j}(x_{pi})\\ 0 & otherwise\end{cases} \tag{12}$$

Basing the heuristic, maximum possible antecedent fuzzy sets on the unit interval [0, 1] is calculated for given antecedent part of fuzzy rule. After some change, Ant update local pheromone based equation 13.

$$\tau(R_C,i,A_j)\leftarrow\tau(R_C,i,A_j)-\rho.(\tau(R_C,i,A_j)-\Delta\tau(R_C,i,A_j)) \tag{13}$$

Where $0<\rho<1$ is parameter of local pheromone evaporation and $\Delta\tau(R_C,i,A_j)=\tau_0$ where τ_0 is primary pheromone amount. When all of the ants finish their search, overall pheromone updating is done basing equation 14.

$$\tau(R_C, i, A_j) \leftarrow (1-\alpha).\tau(R_C, i, A_j) + \alpha.\Delta\tau(R_C, i, A_j) \quad (14)$$

That in equation 15.

$$\Delta\tau(R_C, i, A_j) = \frac{fitness(R_N)}{N_{Classh}} \quad (15)$$

Where $0 < \alpha < 1$, parameter of overall pheromone evaporation and R_N imply new rule that obtained after changing the antecedent parts of current rule. Result of algorithm run, will be best change in current fuzzy rule.

5. Datasets used and results of experiments

In this research, High-dimension datasets from UCI are used (see table 1). 10 datasets used are selected at random, of course with constrain that each of them is included at least 20 features. Reason of this is to indicate importance of preprocessing step in knowledge discovery from high-dimension dataset. Correct sample classification level in each of dataset is shown in table 2.

Dataset Name	No. of Example	No. of Attributes	No. of Classes
Ionosphere	351	34	2
Waveform	5000	40	3
SPECT Heart	267	22	2
Annealing	798	38	6
Breast Cancer Wisconsin	569	32	2
Molecular Biology	106	58	2
Mushroom	8124	22	2
Large Soybean	307	35	19
Lung Cancer	32	56	3
Optical Recognition of Handwritten Digits	5620	64	10

Table 1. Datasets used in the benchmark

Dataset Name	Correct recognition (in percent)
Ionosphere	96.55
Waveform	95.63
SPECT Heart	87.2
Annealing	98.57
Breast Cancer Wisconsin	96.4
Molecular Biology	93.33
Mushroom	99.26
Large Soybean	68.66
Lung Cancer	96.33
Optical Recognition of Handwritten Digits	88.64

Table 2. Correct sample classification

6. Conclusion

A system based fuzzy logic and evolutionary algorithms with values of given parameters has presented. The slightly model has experimented on different (high dimension) datasets of data mining base UCI. 10-fold cross validation method has used for implementing the experiment. Each feature has separately normalized between 0 and 1. Influential features has extracted for increasing system accuracy using PCA and FISHER algorithms.

The system has searched optimum rule in problem state space using a mutation operator of GA and ACO, and has tried to avoid local optimum points.

Proposed method has implemented using Matlab software and returned rules with maximum correct classification.

References

Assareh, A., Moradi, M.H., Volkert, L.G., (2008), A hybrid random subspace classifier fusion approach for protein mass spectra classification, *Springer. LNCS*, vol. 4973pp. 1–11 Heidelberg..

Duangsaithong, R., and Windeatt, T., (2009), Relevance and Redundancy Analysis for Ensemble Classifiers. *Springer-Verlag Berlin Heidelberg*.

Eynipoor A., Abasidezfouli M., Yektaei M. H., (2009),Breast Cancer Using Ant Colony based Fuzzy System, *The Third Iran Data Mining Conference / IDMC*, , Sharif University of Technology, Tehran, Iran.

Heyward, J., Alvarez, S., Ruiz, C., Sullivan, M., Tseng, J. Whalen, G. (2008), Knowledge discovery in clinical performance of cancer patients, *IEEE International Conference on Bioinformatics and Biomedicine*, USA, pp. 51–58.

http://archive.ics.uci.edu/ml/machine-learning-databases

Janecek, A.G.K., Gansterer, W.N., Demel, M.A., Ecker, G.F.: (2008). On the relationship between feature selection and classification accuracy. *Journal of Machine Learning and Research*. JMLR: Workshop and Conference Proceedings 4. pp 90–105

Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski & Lukasz A. Kurgan, (2007), *Data Mining A Knowledge Discovery Approach*, New York.: Springer Science

Nakashima T., and Ishibuchi H., (2005), Using Boosting Techniques To Improve The Performance Of Fuzzy Classification Systems, In Classification And Clustering For Knowledge Discovery, *Studies In Computational Intelligence*. *Springer* , Netserlands, Vol 4. pp.146-157 .

Zachman, John A.,(1987), A Framework for Information Systems Architecture, *IBM Systems Journal*, Vol. 26, No. 3.