A study of the associative pattern classifier method for multi-class processes

R. SANTIAGO-MONTERO[a], G. SERGIO VALADÉZ[b], HUMBERTO SOSSA[b],
DAVID ASAEL GUTIÉRREZ HERNÁNDEZ[a], MANUEL ORNERLAS-RODRÍGUEZ[a]
[a]*Tecnológico Nacional de México. Instituto Tecnológico de León. División de Estudios de Posgrado e Investigación. Avenida Tecnológico s/n, Industrial Julián de Obregón, 37290 León, Guanajuato, México.*
[b]*Instituto Politécnico Nacional, Centro de Investigación en Computación, Miguel Othón de Mendizábal S/N, Col. Nueva Industrial Vallejo, 07738, México D.F.*

Pattern-recognition tasks in machine vision provide solutions for industrial automation and manufacturing processes. These applications are done by extracting data from images and comparing them with well-known data stored, returning a result that helps decide whether the measurement is within a known tolerance. Pattern recognition is an artificial intelligence discipline which is focused to associate a set of features that describes an object with a class or category. Into this field, the associative memories that can be seen as a special class of neural network are used to retrieve altered binary patterns. However, in 2003 was designed the Associative Pattern Classifier (APC), which is an associative memory that is capable to extend this approach to pattern classification field. Several proposals have arisen from APC algorithm; nevertheless and in consequence of its variants, this algorithm is limited to bi-class processes. Moreover, the algorithm has a serious problem when it is configured as a hyper plane classification. The present work solves these drawbacks and it extends the algorithm to multi-class problems. An example of this application is made by using a data base provided from real measurement in the health field.

(Received March 16, 2015; accepted May 7, 2015)

Keywords: Pattern Classifier, Multi-Class, Pattern-Recognition, Associative Memories, Neural approach.

1. Introduction

Innovative solutions in the direction of industrial automation are provided by machine vision [1] and some industrial activities have been benefited from the application of it on manufacturing processes like electronics component manufacturing [2], quality textile production [3], metal product finishing [4], glass manufacturing [5], machine parts [6], printing products [7] and granite quality inspection [8], integrated circuits manufacturing [9] and many others. Machine vision technology provides a competitive advantage to industries that employ this technology improving productivity and quality management. However, there must be a way to process the information that is acquired by a machine vision; this process is then the most important part of the system, an example is shown in figure 1.

Associative Classification (AC) is considered a special case of the association rule [10] because it builds the classification models by adopting association rule discovery methods where the target attribute is considered in the rule's right hand side.

Attributes could have a finite set of possible values (nominal) or real or integer (continuous).

AC depends mainly on two important thresholds called minimum support (MinSupp) and minimum confidence (MinConf). The first one represents the frequency of the attribute value and its associated class in the training data set from the size of that data set and the second one represents the frequency of the attribute value and its related class in the training data set from the frequency of that attributes value in that training data.

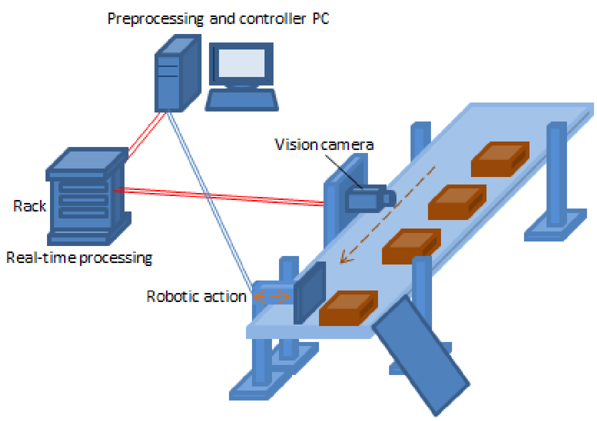

Fig. 1. An example of a vision machine.

AC integrates association mining and classification into a single system, both of them are essential to practical data mining applications [11-17]. Association mining, or pattern discovery, discover descriptive knowledge from a database, while classification focuses on building a classification model for categorizing new data. Considerable efforts have been made to integrate these two techniques into one system. A typical associative classification system is constructed in two stages: 1) discovering all the event associations (in which the frequency of occurrences is significant according to some tests); 2) generating classification rules from the association patterns to build a classifier. In the first stage, the learning target is to discover the association patterns inherent in a database (also referred to as knowledge discovery). In the second stage, the task is to select a small set of relevant association patterns discovered to construct a classifier given the predicting attribute.

Experiments reported in [11, 12, 13, 16, 18, 19] showed that associative classification systems achieve competitive classification results with traditional classification approaches.

This study presents an Associative Pattern Classifier Method for Multi-Class Processes.

2. Theoretical description of pattern classifications

A pattern can be represented as follows [20]:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{X} \subset \mathbb{R}^n \tag{1}$$

where $\mathbf{x}$ is a given pattern, $x_1, x_2, \cdots, x_n$ are the pattern features and $\mathbf{X}$ is the n-dimensional domain of the pattern features. Furthermore, patterns are naturally associated with each other forming categories or classes. Classes can be defined conceptually as [21]:

$$\{c_i \in \Omega | i = 1,2,\cdots,m\} \tag{2}$$

m is the number of classes and Ω is the set of all classes. A pattern classifier is a device that classifies patterns into different types, Fig. 1. When a set of n features of a pattern is presented to the input of the classifier, this is able to respond with an output c_i between m classes. Each output represents a class in which the pattern is placed by the classifier [22].

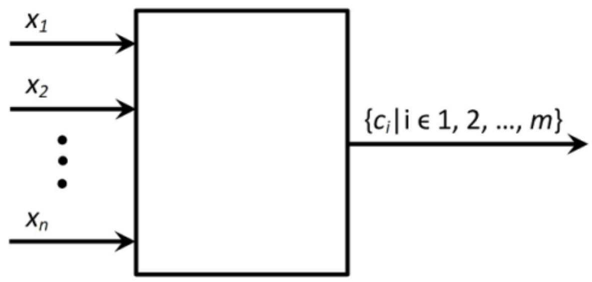

Fig. 2. Block diagram of a pattern classifier.

As it can be seen, the objective of a pattern classifier is to find a hypothesis to do a mapping between the characteristics and class spaces, normally called representative and interpretative spaces respectively, in order to reduce errors and overlapping of information [20, 21, 23].

The trouble found in order to do this mapping is the complex distribution of the majority of problems of the classified patterns.

There are two principal ways to try to solve the problem of classifying patterns. The first one is to find in the representative space a hypothesis that corresponds to the structure of the interpretation space, as it is shown in figure 3a. The second one is to find a structure in the interpretation space that corresponds to the structure in the representation space, shown in figure 3b.

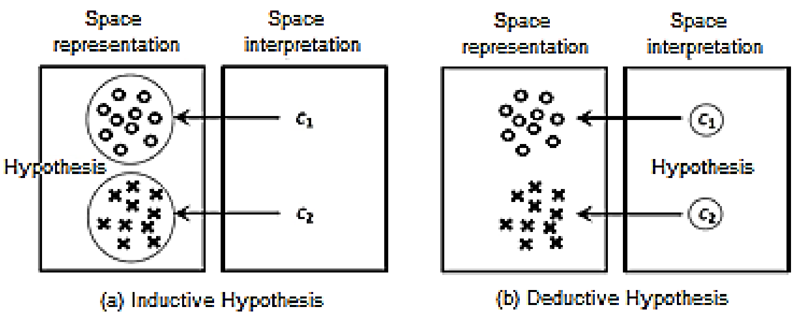

Fig. 3. Hypothesis classification

2.1 Associative classification of patterns method (ACP)

An associative memory can be viewed as a neural network of a layer that allows mapping a set of input patterns $\mathbf{x}^k$ to another set of output patterns and $\mathbf{y}^k$, such that each input pattern is associated with only one output pattern [24, 29].

Let $\mathbf{x}^k \in \mathbf{X}^n \forall k \in \{1,2,\dots,p\}$ the set of input patterns and $\mathbf{y}^k \in \mathbf{Y}^m \forall k \in \{1,2,\dots,p\}$ the output patterns where k is an index representing a particular pair of associated patterns, n and m are the dimensions of the patterns $\mathbf{x}^k$ and $\mathbf{y}^k$ respectively and p is the cardinality of the set of associations. An associative memory, M can be represented as:

$$\mathbf{x}^k \longrightarrow M \longrightarrow \mathbf{y}^k \tag{3}$$

where M is a matrix of correlation of p associations [25]. Associations make up what is known as fundamental set of associations or simply as fundamental set.

$$S = \{(\mathbf{x}^k, \mathbf{y}^k) | k = 1,2,\dots,p\} \tag{4}$$

Associative memories have two phases: the learning phase where an operational process generates the correlation matrix and a recovery phase where memory can respond to patterns of output for each input vector associated [26].

As any associative memory ACP associated an input vector (**x**) with an output vector (**y**). The set of associations are called fundamental set (S). It is so:

$$S = \{(\mathbf{x}^k, \mathbf{y}^k) | k = 1,2, \dots, p\} \quad (5)$$

where $\mathbf{x}^k \in \mathbb{R}^n$ is the set of input patterns, $\mathbf{y}^k \in \{0, 1\}^m$ is the set of output patterns, n is the dimension of $\mathbf{x}^k$, m is the dimension of $\mathbf{y}^k$ and p is the cardinality of S. in the case of ACP, the fundamental set consists of learning patterns .

The class $c \in \{1,2, \dots, m\}$, in which each input pattern $\mathbf{x}^k$ belongs is defined by :

$$y_j^k = \begin{cases} 1 & for\ j = c \\ 0 & for\ j = 1,2, \dots, c-1, c+1, \dots, m \end{cases} \quad \forall k \in \{1,2, \dots, p\} \quad (6)$$

As any associative memory, ACP associates an input vector (**x**) with an output vector (**y**). The set of associations are called fundamental set (**S**). It is so

$$S = \{(\mathbf{x}^k, \mathbf{y}^k) | k = 1,2, \dots, p\} \quad (7)$$

where $\mathbf{x}^k \in \mathbb{R}^n$ is the set of input patterns, $\mathbf{y}^k \in \{0, 1\}^m$ is the set of output patterns, n is the dimension of $\mathbf{x}^k$, m is the dimension of $\mathbf{y}^k$, and p is the cardinality of S. For the ACP the fundamental set consists of learning patterns.

The class $c \in \{1,2, \dots, m\}$ in which each input pattern $\mathbf{x}^k$ belongs is defined by:

$$y_j^k = \begin{cases} 1 & for\ j = c \\ 0 & for\ j = 1,2, \dots, c-1, c+1, \dots, m \end{cases} \quad \forall k \in \{1,2, \dots, p\} \quad (8)$$

2.2 ACP learning phase

The learning phase consists of 3 steps; the first is the construction of the border of the classifier. This border is constructed by changing space patterns to be classified. The ACP makes this change to generate a mean vector of the fundamental set

$$\bar{\mathbf{x}} = \frac{1}{p} \sum_{k=1}^{p} \mathbf{x}^k \quad (9)$$

Retrieved $\bar{\mathbf{x}}$ from the fundamental set, the translation of each pattern contained in the set S is performed (second step).

$$\mathbf{x}_t^k = \mathbf{x}^k - \bar{\mathbf{x}} \quad (10)$$

Finally, the fundamental set translated is used for the construction of the memory of the classifier by applying the Linear Association phase.

$$M = \sum_{k=1}^{p} \mathbf{y}^k \, [\mathbf{x}_t^k]^t \quad (11)$$

2.3 ACP phase recovering

In the recovery phase, associative memories recognize each input vector to respond with its associated output vector. In the case of ACP recovery phase is applied as a step test, name given to the classification process. The process comprises essentially three steps. First translation pattern used to classify the mean vector of the fundamental set is performed.

$$\mathbf{x}_t^{\omega} = \mathbf{x}^{\omega} - \bar{\mathbf{x}} \quad (12)$$

Subsequently, a dot product is applied between the translated vector and the ACP memory. The operation produces a vector of dimension $\boldsymbol{m}$, but with real values

$$\mathbf{z}^{\omega} = M\mathbf{x}_t^{\omega} \quad (13)$$

Finally, in order to know the class of the input vector, the following criteria is applied:

$$y_j^{\omega} = \begin{cases} 1 & if\ z_j^{\omega} = \bigvee_{h=1}^{p} z_h^{\omega} \\ 0 & otherwise \end{cases} \quad (14)$$

Where the maximum value of the components in the resulted vector is assigned to one and the rest of the components are assigned to zero. The class will be designated by the index where there is a one. In order to show the process the following example is shown:

Given the following set of associations that make up the fundamental set; we have

$$\mathbf{x}^1 = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix}, \mathbf{y}^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \mathbf{x}^2 = \begin{pmatrix} -4 \\ 11 \\ -8 \end{pmatrix}, \mathbf{y}^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (15)$$

In this case $p = 2, n = 3$ and $p = 2, n = 3$.

The learning phase is applied, starting with the translation of the patterns used in eq. (9).

$$\bar{\mathbf{x}} = \frac{1}{p} \sum_{k=1}^{p} \mathbf{x}^k = \frac{1}{2} \left[\begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix} + \begin{pmatrix} -4 \\ 11 \\ -8 \end{pmatrix} \right] = \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix} \quad (16)$$

The translation of the patterns is done:

$$\mathbf{x}_t^1 = \mathbf{x}^1 - \bar{\mathbf{x}} = \begin{pmatrix}6\\5\\2\end{pmatrix} - \begin{pmatrix}1\\8\\-3\end{pmatrix} = \begin{pmatrix}5\\-3\\5\end{pmatrix} \tag{17}$$

$$\mathbf{x}_t^2 = \mathbf{x}^2 - \bar{\mathbf{x}} = \begin{pmatrix}-4\\11\\-8\end{pmatrix} - \begin{pmatrix}1\\8\\-3\end{pmatrix} = \begin{pmatrix}-5\\3\\-5\end{pmatrix} \tag{18}$$

Finally, the memory of the ACP is constructed by applying eq. (11).

$$M = \sum_{k=1}^{p} \mathbf{y}^k\,[\mathbf{x}_t^k]^t = \begin{pmatrix}1\\0\end{pmatrix}(5\ \ -3\ \ 5) + \begin{pmatrix}0\\1\end{pmatrix}(-5\ \ 3\ \ -5) = \begin{pmatrix}5 & -3 & 5\\-5 & 3 & -5\end{pmatrix} \tag{19}$$

With the built memory, next step is to proceed to the trial stage; where fundamental set patterns are classified. In order to show the procedure, the first pattern of the fundamental set $= \begin{pmatrix}6\\5\\2\end{pmatrix}$ is taken. By applying translation according eq. (12), we have

$$\mathbf{x}_t^1 = \mathbf{x}^1 - \bar{\mathbf{x}} = \begin{pmatrix}6\\5\\2\end{pmatrix} - \begin{pmatrix}1\\8\\-3\end{pmatrix} = \begin{pmatrix}5\\-3\\5\end{pmatrix} \tag{20}$$

With the transferred pattern it is presented to the memory of ACP to obtain a vector of real elements.

$$\mathbf{z}^\omega = M\mathbf{x}_t^\omega = \begin{pmatrix}5 & -3 & 5\\-5 & 3 & -5\end{pmatrix}\begin{pmatrix}5\\-3\\5\end{pmatrix} = \begin{pmatrix}59\\-59\end{pmatrix} \tag{21}$$

Applying eq. (14) its associated output pattern is obtained; where the index value one will be the class to which it belongs.

$$\mathbf{y}^\omega = \begin{pmatrix}1\\0\end{pmatrix} \tag{22}$$

As it is observed, the class to which it belongs is the one and maintains the correlation with its output vector assigned at the beginning.

A pattern that belongs to the same class can be viewed as a distorted version of the values of their characteristics or traits. The following vector, $\mathbf{x} = \begin{pmatrix}4\\7\\-1\end{pmatrix}$ could be another element of class one. Applying the recovery phase it can be confirmed that the class should be the one.

Applying the translation;

$$\mathbf{x}_t^1 = \mathbf{x}^1 - \bar{\mathbf{x}} = \begin{pmatrix}4\\7\\-1\end{pmatrix} - \begin{pmatrix}1\\8\\-3\end{pmatrix} = \begin{pmatrix}3\\-1\\2\end{pmatrix} \tag{23}$$

The pattern of real elements to apply the dot product between the memory of ACP and the moved vector is obtained.

$$\mathbf{z}^\omega = M\mathbf{x}_t^\omega = \begin{pmatrix}5 & -3 & 5\\-5 & 3 & -5\end{pmatrix}\begin{pmatrix}3\\-1\\2\end{pmatrix} = \begin{pmatrix}28\\-28\end{pmatrix} \tag{24}$$

The output vector is calculated.

$$\mathbf{y}^\omega = \begin{pmatrix}1\\0\end{pmatrix} \tag{25}$$

where the index of value one indicates the class. In this case the class is the one, which confirms the hypothesis.

From this simple example we can detach the following:

The ACP is bi-class for nature. The algorithm sets the decision boundary in the middle of the space generated by the training patterns, but could not guarantee that the translation vector was ideal with more than two classes; it is not guarantee that the translation point would be at the border in the middle of the classes.

The ACP algorithm generates a matrix where each sum vectors associated class is in every line of memory. This has the direct consequence, for bi-class case, the generation of a vector system in equilibrium, where the magnitudes of the sum vectors are of equal magnitude but opposite direction.

The ACP produces distorted pattern classification, the representative vector of the class, by the direction in which is the vector sum. A vector that cannot be classified lie on the boundary that is orthogonal to the vector sum.

3. Case analysis

3.1 ACP applied for a bi-class classification

This section shows the performance of the ACP with bi-class classification problems. The databases can be consulted from the Repository of the University of California at Irvine [27]. The WEKA system is used for the comparative testing and percentage yield by applying different k-folds.

One of the databases frequently used to measure the performance of any proposed classifier is the Wisconsin Breast Cancer (WBC). The base has 699 records with nine traits or attributes; is an unbalanced data base, where 65.5% are negative cases of diagnosis and 34.5 are positive. The choice of the database is its wide range of classifiers and possibilities to prove the existence of comparative found in the literature [28].

In Table 1 the results of accuracy for data base WBC are reflected. Except for the SVM support vector machine and the ACP, were generated by applying the WEKA platform.

Table 1. Comparison of the ACP algorithm over other approaches using classification accuracy as a parameter.

Size of folds	Sorter								
	CAP	**MDC**	**NB**	**1-NN**	**2-NN**	**3-NN**	**BP**	**SVM**	**C4.5**
1%-99%	**96.39**	94.33	92.23	95.44	88.78	91.25	66.44	94.00	93.87
10%-90%	**97.16**	96.06	95.60	94.96	93.52	95.69	95.52	96.20	91.93
30%-70%	**97.14**	95.98	96.06	95.44	94.25	96.27	96.08	96.56	93.87
50%-50%	**97.31**	96.05	96.15	95.62	94.84	96.55	96.37	96.75	94.32
70%-30%	**97.31**	96.31	96.36	95.65	95.12	96.84	96.59	96.93	94.68

The percentages of accuracy realize good returns for bi-class problems of the ACP. Consequently, the algorithms derived from ACP inherit the advantages of the algorithm. However, the ACP has the following problems that also inherit their recent changes.

The first problem is obtaining the translation vector. The approach of looking at the patterns as a cloud data has undesirable implications. In Figure 4 the configuration of an AND gate is shown. The translation vector, using eq. (4), place the border over two of the elements of a class to determine $\bar{x} = \{0.5,0.5\}$; where the ACP cannot generate a decision. Therefore, the algorithm ensures not to find the border between classes with linear separability. Figure 4 illustrates this situation. It is observed that two of the element belonging to the class values 0 cannot be classified.

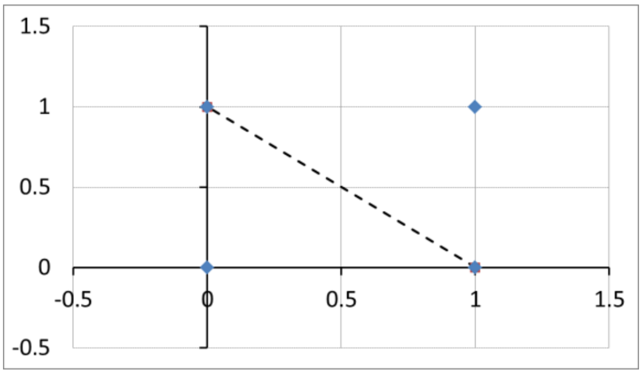

Fig. 4: Erroneous determination of the boundary of decision by the ACP for an AND gate

Another problem of the algorithm is presented in the classification of databases with more than two classes. In order to show this disadvantage, the databases iris plant and k-folds with k equal to 10 are used. The database has 3 classes (setosa, versicolor and virginica); 2 of them which have no linear separability (virginica and versicolor). ACP is applied first on the two non-linearly separable classes. Subsequently, the third class is integrated, which has linear separability to the database and the algorithm is applied again. The results are shown in Table 2.

Table 2. ACP accuracy values in bi-class and multi-class problems using the database iris plant.

Size of folds	Exactness	
	ACP over two classes (versicolor-virginica)	**ACP over 3 classes (versicolor-virginica-setosa)**
10% -90%	**86%**	**69%**

The test results show that the algorithm efficiency decays with multi-class problems; as a result of the translation vector design.

3.2 ACP extension

Designing patterns like a cloud of data and generate a partition in the center of the cloud results in disadvantages for the algorithm. The first proposed solution to improve the algorithm focuses on the translation vector calculation. For bi-class problems such as logic gates, classes will be separately and not as a cloud of data; then we can return the focus of representative patterns of class and draw the line at the midpoint of the distance between the two representatives. With this modification, the problem shown for logic gates is resolved. The calculation of the boundary decision now is given by:

$$vm = \frac{x_1 + x_2}{2} \tag{26}$$

When equation (26) is applied to the problem of the AND gate, the translation vector is now:

$$vm = \{0.2, 0.2\} \tag{27}$$

With this new result, the ACP can determine the classes correctly.

This change will also serve the ACP to address multiclass processes. By the nature of the bi-class algorithm, multi-class problems should be seen as a problem where two classes are formed: From the whole set of classes one is separated and the remaining is taken as the other class. That is:

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ where n is the number of classes present in the problem, then a hierarchical classification [22] is applied, such that bi-class processes generated at each level of a hierarchical tree with a translation vector for remaining set of classes, Figure 5 illustrates the concept.

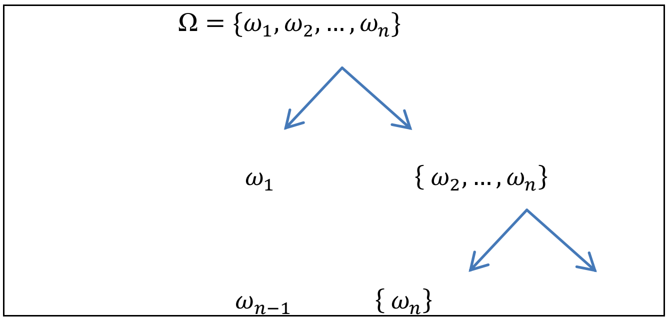

Fig. 5. Process hierarchical of classification for multi-class problems using the ACP algorithm.

4. Experimental results

ACP algorithm modifications on different databases of the UCI repository in multi-class processes were tested. In the testing were included the Iris Plant database with linearly non-separable two classes and a third linearly separable from the other two. WINE data base, which contains the characterization of 3 types of wine, Glas Identification describing six non-linearly separable glass category as part of the criminal investigation and one of the most difficult to separate by automatic means; Strakig Project is a set of descriptions of silhouettes of vehicles; finally, Image Segmentation characterization contains 7 types of scenes. Table 3 describes the characteristics of the databases used.

Table 3: Description of multiclass data bases used for testing the ACP improved.

Database	Instances	Number of attributes	Classes
Iris Plant	150	4	3
Wine	178	13	3
Glass Idenfication	214	9	6
Statlog Project	846	18	4
Image Segmentation	210	19	7

Table 4 shows the results when changes are implemented in the algorithm of ACP. The results of the original algorithm are compared to the results of the modifications. Again the process of validating the results of accuracy is performed by 10-folds. The results show that adaptation manages to improve the accuracy of the original algorithm. It is noted that the results of accuracy depends on the order in which classes are taken. Therefore, it must be performed a prior combinatorial to generate the most efficient order.

Table 4: ACP accuracy results with the proposed modifications.

Database / Yield	ACP(%)	Improved ACP (%)	Difference of exactness
Iris Plant	66	88.4	10.3
Wine	68.1	76.2	8
Glas Idenfication	50	81	31
Statlog Project	38.5	62	23
Image Segmentation	38.5	62	23

5. Conclusions

The ACP algorithm has shown its effectiveness and efficiency in bi-class classification process. The low computational work in generating memory classifier and one-shot process makes this algorithm a suitable tool in real-time processes. In this paper are shown the improvements that reduce the limitations of the original algorithm. These improvements can be translated immediately to derived versions of ACP. Simultaneously, the change in focus extends the field of application of the algorithm to multi-class processes. However, take note that the accuracy of multi-class classifier processes depends on the order in which classes are processed. Since the speed of the algorithm to generate memory and the limited number of patterns in your apprentice stage, combined with a relatively low number of classes in classification processes, the process of combinatorial might not be a constraint.

6. Acknowledgements

R. Santiago-Montero, D. A. Gutiérrez Hernández and M. Ornelas-Rodríguez would like to thank to the Instituto Tecnológico de León, Tecnológico Nacional de México and CONACYT for their support. S. Valadez and H. Sossa thank CIC-IPN and CONACYT for the economic support under the projects SIP 20151187 and 155014 to carry out this research. We all thank the comments of the reviewers to improve the paper.

References

[1] E. R. Davies, Automated Visual Inspection in Machine Vision, vol. Chapter 19, 2nd Edition ed: Academic Press, 1998, pp. 471-502.

[2] J. L. C. Sanz, D. Petkovic, IEEE Trans. on PAMI, **10**, 830 (1988).

[3] C. Bahlmann, G. Heidemann, H. Ritter, Pattern Recognition, **32**, 1049 (1999).

[4] J. W. Tucker, "Inside Beverage Can Inspection: An Application from Start to Finish," presented at Proc. of the Vision '89 Conference, 1989.

[5] A. R. Novini, “Fundamentals of Machine Vision Inspection in Metal Container Glass Manufacturing,” presented at Vision '90 Conference, 1990.
[6] J. Ker, K. Kengskool, “An Efficient Method for Inspecting Machine Parts by a Fixtureless Machine Vision System,” presented at Vision '90 Conference, 1990.
[7] T. Torres, J. M. Sebastian, R. Aracil, L. M. Jimenez, O. Reinoso, Image and Vision Computing, **16**, 947 (1998).
[8] L. Shafarenko, M. Petrou, J. Kittler, IEEE Trans. on Image Processing, **6**, 1530 (1997).
[9] H. Li, J. C. Lin, IEEE Transactions on Industry Applications, **30**, 1530 (1994).
[10] F. Thabtah, P. Cowling, Y. Peng, MCAR: Multi-class classification based on association rule approach. Proceeding of the 3rd IEEE International Conference on Computer Systems and Applications (pp. 1-7).Cairo, Egypt. (2005)
[11] G. Dong, X. Zhang, L. Wong, J. Li. CAEP:classification by aggregating emerging patterns. In Proceedings of The Second International Conference on Discovery Science (DS'99), pages 43-55, Japan, December 1999.
[12] J. Li, G. Dong, K. Ramamohanarao, L. Wong. Machine Learning, **54**(2),99(2004).
[13] W. Li, J. Han, J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01), pages 369{376, San Jose, CA, November 2001.
[14] B. Liu, W. Hsu, Y. Ma. Integrating classification and association rule mining. In Proceedings of the Fourth ACM SIGKDD International Conference on knowledge Discovery and Data Mining, pages 80-86, New York, NY, August 1998.
[15] K.Wang, S. Zhou, Y. He. Growing decision tree on support less association rules. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'00), pages 265{269, Boston, MA, August 2000.
[16] Y. Wang, A. K. C. Wong. IEEE Trans. On Knowledge and Data Engineering, **15**(3), 764 (2003).
[17] X. Yin, J. Han. CPAR: Classification based on predictive association rules. In Proceedings 2003 SIAM International Conference on Data Mining (SDM'03), pages 331-335, San Francisco, CA, May 2003.
[18] B. Liu, W. Hsu, Y. Ma. Integrating classification and association rule mining. In Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 80{86, New York, NY, August 1998.
[19] A. K. C. Wong, Y. Wang. IEEE Transactions on Systems, Man, and Cybernetics, **33**(1), 114 (2003).
[20] E. Alpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, 2nd edition 2010
[21] A. R. Webb,. Statistical Pattern Recognition. John Wiley & Sons, 2nd edition. 2002
[22] N. J. Nilsson, The mathematical foundations of learning machines. Morgan Kamann Publishers Inc., San Francisco, CA, USA. 1990.
[23] J. P. Marques de Sá,. Pattern Recognition, Concepts, Methods, and Applications. Springer-Verlag. 2002
[24] G. Palm, F. Schwenker, F. T. Sommer, A. Strey, Neural Associative Memories. Biological Cybernetics, **36**, 36 1993.
[25] B. S P. E., Prasad, Y. Sagar, , P. S., Rama, International Journal of Advanced Computer Sciences and Applications, **1**(6), 124 (2011).
[26] M. Aldape-Pérez, C. Yañez-Marquez, O. Camacho-Nieto, & J. Argüelles-Cruz, Computer Methods and Programs in Biomedicine, **106**(3), 287 (2011)
[27] A. Frank, A. Asuncion, 2010. {UCI} Machine Learning Repository. Available at: http://archive.ics.uci.edu/ml
[28] F. Paulin, A. Santhakumaran, International Journal on Computer Science and Engineering (IJCSE), **3**(1), 327 (2011).
[29] R. Santiago-Montero, C. Yañez-Marquez, J. L. Diaz-de-Leon, Research on computing Science: Pattern Recognition, Advances and Perspectives, 2002, pp. 449-460.

*Corresponding author: david.gutierrez@itleon.edu.mx