Acta Agraria Kaposváriensis (2006) Vol 10 No 3, 89-107
Kaposvári Egyetem, Állattudományi Kar, Kaposvár
University of Kaposvár, Faculty of Animal Science, Kaposvár

Fuzzy Association Rule Mining for the Analysis of Historical Process Data

F.P. Pach[1], A. Gyenesei[2], S. Németh[1], P. Árva[1], J. Abonyi[1]
[1] University of Veszprém, Departement of Process Engeneering, Veszprém, Hungary
[2] Department of Knowledge and Data Analysis, Unilever Research Vlaardingen, The Netherlands

ABSTRACT

Process data collected during the operation of complex production processes can be used for system identification, process monitoring and optimization. This work presents a new algorithm that is able to extract useful knowledge from data. The extracted information is given in the form of association rules. Association rule mining finds interesting association or correlation relationships among a large set of data items. The large itemsets can be related to the frequent events of a process, and this is useful for detect unknown relationships among the process variables, reduct the models of the system, estimate the product quality and build a classifier. The proposed method based on the Apriori algorithm, but the main idea is incorporate fuzziness (fuzzy logic increases the interpretability of the model and tolerance against measurement noise and uncertainty). The general applicability and efficiently of the developed tool are showed by an application study, one general example for the feature (input) selection problem and the analysis of a polymerization process data. Moreover the proposed classifier is used for three general used classification problems.

(Keywords: fuzzy logic, classification, association rules, knowledge discovery, polymerization)

ÖSSZEFOGLALÁS

Fuzzy asszociációs szabálybányászat hisztorikus folyamat adatok elemzésére

Pach[1] F.P., Gyenesei[2] A., Németh[1] S., Árva[1] P., Abonyi[1] J.
[1]Veszprémi Egyetem, Folyamatmérnöki Tanszék, Veszprém
[2]Adatelemző Központ, Unilever Kutatóközpont Vlaardingen, Hollandia

A bonyolult gyártási folyamatok irányítása során keletkező folyamatadatok felhasználhatók rendszerazonosításra, folyamat monitorozásra és optimalizálásra. A cikk egy olyan új algoritmust mutat be, amelynek segítségével hasznos információkat nyerhetünk ki ezen folyamatadatokból. A bemutatott eljárás a feltárt információkat asszociációs szabály formájában jeleníti meg. Nagy adathalmazokban asszociációs szabálykereséssel érdekes összefüggéseket tárhatunk fel az egyes elemek között. A megtalált összefüggésekre, mint nagy elemhalmazokra hivatkozunk, amelyek gyakori események együttes előfordulásai lehetnek egy-egy folyamaton belül. Az általunk javasolt eljárás az APRIORI algoritmust használja alapul, azonban annak fuzzy módosítását alkalmazzuk (a fuzzy logika növeli a modell értelmezhetőségét, alkalmas zajos adatok és a bizonytalanság kezelésére), lehetővé téve a folyamatváltozók rejtett összefüggéseinek feltárását, segítségével megbecsülhető a termék minősége, továbbá osztályozási modelleket is generálhatunk. Az eljárás általános alkalmazhatóságát és *hatékonyságát egy tanulmánnyal szemléltetjük, ahol az első példa modell struktúra meghatározása egy polimerizációs reaktor esetében, majd az eljárás osztályozási teljesítményét vizsgáljuk három széles körben használt osztályozási problémán.*

(Kulcsszavak: fuzzy logika, osztályozás, asszociációs szabályok, tudásfeltárás, polimerizáció)

INTRODUCTION

One of the most popular research tasks in data mining is the *discovery of frequent itemsets and association rules*. The problem originates in market basket analysis which aims at understanding the behavior of retail customers, or in other words, finding associations among the items purchased together *Agrawal* (1994). A famous example of an association rule in such a database is “diapers => beer”, i.e. young fathers being sent off to the store to buy diapers, reward themselves for their trouble. Because of the practical usefulness of association rule discovery, this approach can be applied in various research areas.

This paper presents two applications of this data mining tool:

A. *Feature selection method* based on fuzzy association rule mining
B. *Associative classification method* based on fuzzy rule base

Feature Selection Methods

Real-world data analysis, data mining and modeling problems typically involve a large number of potential variables. The number of these variables should be minimized, especially when the model is nonlinear and contains many parameters. Therefore, effective methods for feature selection (also called structure selection) are very important for any modelling exercise. This paper proposes a new data-driven method for the structure selection of nonlinear models that can be represented by the following model: y=*f*(x), where *f*(.)is a nonlinear function and **x** represents the vector of the input variables of the model. For dynamical systems, the input-selection problem includes the choice of the model’s order (number of lagged inputs and outputs used as regressors) and the number of pure time delays.

A large number of structure-selection methods, like correlation or principal component analysis have been introduced for linear models. Several information-theoretical criteria have been proposed for the structure selection of linear dynamic input-output models. Examples of the classical criteria include the final prediction error and the Akaike information criterion *Akaike* (1974). Subsequently, Schwartz and Rissanen later developed the minimum description length criterion, which was proven to produce consistent estimates of the structure of linear dynamic models *Liang* (1993). With these tools, determining the structure of linear systems is a rather straightforward task. However, these methods usually fail to discover the significant inputs in real-world data, which are almost always characterized by nonlinear dependencies. Relatively little research has been carried out on the structure selection for nonlinear models. In the paper of *Aguirre* (1995), it is argued whether a certain type of term in a nonlinear model is spurious. In *Aguirre* (1996), this approach is used for the structure selection of polynomial models, and an alternative solution is introduced by initially conducting a forward search through the many possible candidate model terms before performing an exhaustive all-subset model selection on the resulting model. A backward search approach based on orthogonal parameter estimation is also applied *Korenberg* (1988) and *Mendes* (2001).

Rule Based Classification

Database mining problems involving classification can be viewed within a common framework of rule discovery *Agrawal* (1993). Effective development of data mining techniques to discover knowledge from training samples for classification problems is evitable. Moreover, it is necessary to develop effective methods for classification problems in industrial engineering. A classification data set is normally in the form of a relational table, which is described by a set of distinct attributes (discrete and continuous). Each data record (or example) is also labelled with a class label. The *classification* is to build a model - based on training data set - called *classifier* to predict future data objects for which the class label is unknown. *Rule-based classification systems* have been widely used in real world applications because of the easy interpretability of rules. The left side of the rule is the antecedent part, that determines the condition and the right side of the rule, the consequence part is one class label. Therefore rules are in form: $X \rightarrow c_i$.

The *traditional rule-based classifiers* prefer small rule sets to large rule sets, but small classifiers are sensitive to the missing values in unseen test data. Many techniques (decision trees *Quinlan* (1992), rule learning *Clark* (1989), Naïve-Bayes classification *Duda* (1973), statistical approaches *Lim* (2000)) and systems (~ rule induction algorithms as C4.5 *Quinlan* (1992), *Clark* (1989), and RIPPER *Cohen* (1995)) have been developed. These techniques have largely focused on finding compact, representative subsets of rules that can be used for prediction.

On the other hand, studies propose approach within data mining have concentrated on using exhaustive search to find all high quality association rules that satisfy a set of constraints (typically based on support and confidence). Recently, these two approaches have been integrated in that a number of different research groups have developed tools for classification based on association rule discovery (*associative classification*). Clearly, both the computational complexity and the number of rules produced grow exponentially for association rule mining. Minimum support holds the key for the success of the model. Although the complete set of rules may not be directly used for classification, effective and efficient classifiers have been built using the rules. The most of these methods work in two main phases:

1. discovering all association rules,
2. organizing the resulted association rules in a classification model.

In CBA *Liu* (1998) (Classification Based on Associations) a set of high confidence rules is selected from classification association rules to form a classifier (this method is also used in msCBA). The selection of rules is based on a total order defined on the rules.

The msCBA *Liu* (2000) uses multiple class minimal support in rule generation (i.e., each class is assigned a different minimal support), rather than using only a single minimal support as in CBA.

CMAR *Li* (2001) (Classification based on Multiple class-Association Rules) method extends an efficient frequent pattern mining method, FP-growth, constructs class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage.

CPAR *Yin* (2003) (Classification based on Predictive Association Rules) combines the advantages of both associative classification and traditional rule-based classification. Instead of generating a large number of candidate rules as in associative classification, CPAR adopts a greedy algorithm to generate rules directly from training data.

CAEP *Dong* (1999) (Classification by Aggregating Emerging Patterns) Emerging patterns (EPs) are item sets whose supports change significantly from one dataset to another; they were recently proposed to capture multi-attribute contrasts between data classes, or trends over time.

ADT *Wang* (2000) (Association based Decision Tree) combines the richness of association rules and the accuracy-driven pruning of decision tree induction. To give DT induction the full pruning power, all confident association rules are used without any support requirement.

LB *Meretakis* (1999) (Large Bayesian) Item sets provide local descriptions of the data. This work proposes to use item sets as basic means for classification purposes too. To enable this, the concept of class support sup of an item set is introduced, i.e., how many times an item set occurs when a specific class c is present.

CorClass *Zimmermann* (2004) (Correlated Association Rule Mining for Classification) first discovers all correlated association rules (adapting a technique by Morishita and Sese) and then applies the discovered rule sets to classify unseen data. The key advantage of CorClass, as compared to other techniques for associative classification, is that CorClass directly finds the associations rules for classification by employing a branch-and-bound algorithm.

The major strength of such systems is that they are able to use the most accurate rules for classification because their rule learners aim to find all rules. This explains their good performance in general.

The paper is organized as follows. *Section 1* shows the base of *fuzzy association rule mining* (counting the fuzzy support, mining frequent item set and generation of rules). *Section 2* illustrates how the fuzzy association rule mining algorithm can be used to determine the relationships of variables in a function, select the model structure of a linear and non-linear model, or select the most relevant features that apply to determine product quality in a production process. *Section 3* presents the associative classification method based on fuzzy rule base. In *Section 4* the rule pruning methods are detailed. General applicability and efficiently of the developed tool are showed by an application study in *Section 5*.

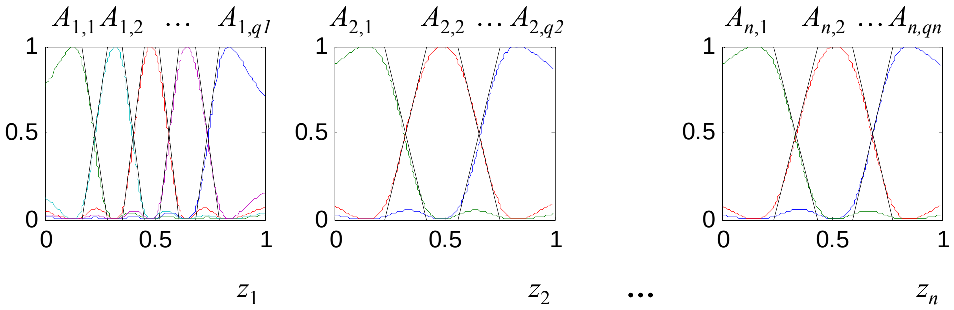
FUZZY ASSOCIATION RULE MINING

Generate a Fuzzy Dataset

The original data (*D* may include crisp values (continuous and discrete) for each attribute) available for the identification of the model is arranged into a matrix $\mathbf{D} = [\mathbf{X}\ Y]_{N\times n+1}$, where $\mathbf{X} = [x_{i,k}]_{N\times n}$ is the input matrix (the *i*th input vector denoted by $\mathbf{x}_i = [x_{i,1}, x_{i,2},\ldots, x_{i,n}]^T$) where $i = 1,\ldots,N$ and *N* represents the number of data samples and *n* represents the number of input variables. The output variable „matrix" $Y = [y_i]_{N\times 1}$ is a column vector, because the searched rules have only one item in the consequent part. This data must be transformed into a fuzzy dataset to allow fuzzy association rule mining. Therefore, the first step of the algorithm generates a new fuzzy dataset from the original dataset by user specified fuzzy sets. Where crisp sets are being applied instead of fuzzy ones, this step could be considered a discretization of the numeric (quantitative) attributes of the original dataset. The discretization of the data into disjoint subsets (partitioning) for each variable is referred as binning, since the partitions (intervals) defined on the quantitative features can be considered as bins. Instead of quantizing the input data into hard subsets, fuzzy Gustafson-Kessel (GK) clustering *Gustafson* (1979) is used to partition all the candidate regressors into fuzzy subsets. As a result, for each input z_j, the cluster centers v^j and the partition matrix $\mathbf{U}^j \in [0,1]^{nj\times N}$ are obtained, where elements of the partition matrix represents the membership of the $z_{j,k}$ data in the ith cluster ($k = 1,\ldots,N$ and $i = 1,\ldots,q_j$, where the number of clusters is q_j). The resulting clusters can be directly used to generate the fuzzy data, e.g. $A_{j,i}(x_{j,k}) = \mathbf{U}^j_{k,i}$. Beside this nonparametric definition of the membership values it is advantageous to design parameterized membership functions to represent the $A_{j,i}(x_{j,k})$ fuzzy sets. For this purpose, trapezoidal membership functions can be used, see in *Figure 1*. Each trapezoid is represented by four parameters related to the shoulders and the legs of the trapezoid: $a_{j,i}$, $b_{j,i}$, $c_{j,i}$ and $d_{j,i}$. At the current implementation, the position of the shoulders is determined based on a threshold value where $\mathbf{U}^j_{k,i} > 0.9$. The legs of the membership functions were defined to obtain a Ruspini type partition, as $\sum_{i=1}^{n_i} A_{j,i}(x_{j,k}) = 1, \forall\ j,k,$ i.e. $a_{j,i} = c_{j,i-1}$ and $d_{j,i} = b_{j,i+1}$

Figure 1

Trapezoidal membership functions



1. ábra: Trapéz tagsági függvények

Definition 1: the ith *input fuzzy data* (for attributes $z_1 z_{2,\ldots,} z_n$) is denoted by $\mathbf{t}_i$:

$$t_i=[u_1,u_2,\ldots,u_l,\ldots,u_n,y_i], \tag{1}$$

where $\mathbf{u}_l$ includes the fuzzy membership values (between 0 and 1) of the ith input data $\mathbf{x}_i$ for all trapezoids membership functions: $A_{l,1},\ldots, A_{l,ql}$ on the l^{th} attribute, and q_l denotes the number of trapezoids on the lth attribute.

$$U_l=[A_{l,1}(x_{i,l}),A_{l,2}(x_{i,l}),\ldots,A_{l,q_l}(x_{i,l})], \tag{2}$$

and fuzzy models can be identified from such data by generating a fuzzy rule base with rules in the form of:

$$\text{Rl: If } x_1 \text{ is } A_{1,l} \text{ and } \ldots \text{ and } x_n \text{ is } A_{n,l} \text{ then } y_l \text{ is } B_l, \tag{3}$$

where R_j denotes the l^{th} rule, l=1,…,M. The M denotes the number of the rules. $A_{l,l},\ldots,A_{n,l}$ are the antecendent fuzzy sets described by membership functions, and y_l is the output of lth rule, where B_j is value of a categorical variable or a class label in case of classification data set. In regard to our goal of generating fuzzy rule base, in this section we focus on the problem of mining fuzzy association rules. Such rules can be discovered in two steps: (1) *mining frequent itemsets*, and (2) *generating association rules* from the discovered set of frequent itemsets. For both steps we have to define the concept of fuzzy support, it is used as a criterion in deciding whether a fuzzy itemset (association rule) is frequent or not, therefore we first introduce the basic definitions and notations that are needed in frequent itemset and association rule mining.

Counting the Fuzzy Support

Let $D_F = \{t_1, t_2, \dots, t_N\}$ be a transformed fuzzy dataset of N tuples (data points) with a set of variables $\mathsf{Z} = \{Z_1, Z_2, \dots, Z_{n+1}\}$ (where the z_{n+1} is the output variable y) and let $C_{i,j}$ be an arbitrary fuzzy interval (fuzzy set) associated with attribute Z_i in Z. From this point, we use the notation $<Z_i:C_{i,j}>$ for an *attribute-fuzzy interval pair*, or simply *fuzzy item*. An example could be <Age:Young>. For *fuzzy itemsets*, we use expressions like <Z:C> to denote an ordered set $Z \subseteq \mathsf{Z}$ of attributes and a corresponding set C of some fuzzy intervals, one per attribute, i.e. $<Z:C> = [<Z_{i1}:C_{i1,j}> \cup <Z_{i2}:C_{i2,j}> \cup \dots \cup <Z_{iq}:C_{iq,j}>]$, $q \leq n+1$. In the literature, the fuzzy support value has been defined in different ways. Some researchers suggest the minimum operator as in fuzzy intersection, others prefer the product operator (as examples, see [22, 23]). They can be defined formally as follows.

Definition 2: value $t_k(z_i)$ for attribute z_i, then *fuzzy support* of <Z:C> with respect to D is defined as

$$FS(Z:C) = \frac{\sum_{k=1}^{N} \min_{\langle z_i:c_{i,j}\rangle \in \langle Z:C\rangle} t_k(z_i)}{N} \tag{4}$$

or

$$FS(Z:C) = \frac{\sum_{k=1}^{N} \prod_{\langle z_i:c_{i,j}\rangle \in \langle Z:C\rangle} t_k(z_i)}{N} \tag{5}$$

We treat memberships as probabilities and therefore prefer the product form. A fuzzy support reflects how the record of the identification dataset support the itemset.

Definition 3: an itemset <Z:C> is called *frequent itemset* if its fuzzy support value is higher than or equal to a user-defined minimum support threshold σ.

The following example illustrates the calculation of the fuzzy support value. Let <X:A> = [<Balance:medium> ∪ <Income:high>] be a fuzzy itemset, the dataset shown in Table 1. The fuzzy support of <X:A> is given by:

$$FS(X:A) = \frac{0.5 \cdot 0.4 + 0.8 \cdot 0.4 + 0.7 \cdot 0.7 + 0.9 \cdot 0.3 + 0.9 \cdot 0.6}{5} = 0.364 \tag{6}$$

Table 1

Example database containing membership values

⟨Balance:medium⟩	⟨Credit:high⟩	⟨Income:high⟩
0.5	0.6	0.4
0.8	0.9	0.4
0.7	0.8	0.7
0.9	0.8	0.3
0.9	0.7	0.6

1. táblázat: Tagsági értékeket tartalmazó példa adatbázis

Mining Frequent Itemsets

As mentioned above, the first subproblem of discovering fuzzy association rules is to find all frequent itemsets. The best-known and one of the most commonly applied frequent pattern mining algorithms, *Apriori*, was developed by *Agrawal* (1994). The name is based on the fact that the algorithm uses prior knowledge of frequent itemsets already determined. It is an iterative, breadth-first search algorithm, based on generating stepwise longer *candidate* itemsets, and clever pruning of non-frequent itemsets. Pruning takes advantage of the so-called *apriori* (or *upward closure*) *property* of frequent itemsets: all subsets of a frequent itemset must also be frequent. Each candidate generation step is followed by a counting step where the supports of candidates are checked and non-frequent ones deleted. Generation and counting alternate, until at some step all generated candidates turn out to be non-frequent. A high-level pseudocode of the algorithm is given in the following:

Algorithm *Mining Frequent Fuzzy Itemsets* (minimum support σ, dataset *D*)

```
$k = 1$
$(C_k;D_F) = Transform(D)$
$F_k = Count(C_k, D_F, \sigma)$
while $|C_k| \neq 0$ do
    $inc(k)$
    $C_k = Generate(F_{k-1})$
    $C_k = Prune(C_k)$
    $F_k = Count(C_k, D_F, \sigma)$
    $F = F \cup F_k$
end
```

The subroutines are outlined as follows:

- *Transform*(*D*): Generates a fuzzy database D_F from the original dataset *D* (denoted by Step 1 in the following sections). At the same time the complete set of candidate items C_1 is found.
- *Count*(C_k , D_F, σ): In this subroutine the fuzzy database is scanned and the fuzzy support of candidates in C_k is counted. If this support is not less than minimum support σ for a given itemset, we put it into the set of frequent itemsets F_k.
- *Generate*(F_{k-1}): Generates candidate itemsets C_k from frequent itemsets F_{k-1}, discovered in the previous iteration *k-1*. For example, if F_1={⟨Balance:high⟩,⟨Income:high⟩} then C2={⟨Balance:high⟩∪⟨Income:high⟩}.
- *Prune*(C_k): During the prune step, the itemset will be pruned if one of its subsets does not exist in the set of frequent itemsets *F* (denoted by Step 2 in the following sections).

Generate Fuzzy Association Rules

Since the rules are generated from the frequent itemsets, the generation of fuzzy association rules (denoted by Step 3 in the following sections) becomes relatively straightforward. More precisely, each frequent itemset <Z:C> is divided into a consequent <Y:B> and antecedent <X:A>, where $X \subset Z$, $Y = Z-X$, $A \subset C$ and $B = C-A$.

Definition 4: a *fuzzy association rule* can be represented in the form of

If *X* is *A* **then** *Y* is *B* (7)

or in more compact form of

$$\langle X:A\rangle \Rightarrow \langle Y:B\rangle \qquad (8)$$

Definition 5: *confidence* of a fuzzy association rule$\langle X{:}A\rangle \Rightarrow \langle Y{:}B\rangle$ is defined as

$$\mathrm{FC}(X : A \Rightarrow Y : B) = \frac{\mathrm{FS}(\langle X : A\rangle \Rightarrow \langle Y : B\rangle)}{\mathrm{FS}(X : A)} \tag{9}$$

which can be understood as the conditional probability of <Y:B>, namely P(<Y:B>|<X:A>).

Definition 6: an association rule is *strong rule* if its support and confidence exceeds a given minimum support (σ) and minimum confidence threshold (γ). Since the rules are generated from frequent itemsets, they satisfy the minimum support automatically.

Using our sample database (Table 1), the fuzzy confidence value of the rule "If Balance is medium and Income is high then Credit is high" is calculated as

$$\mathrm{FC}(X : A \Rightarrow Y : B) = \frac{0.278}{0.364} = 0.766 \tag{10}$$

Association rules mined using the above support-confidence framework are useful for many applications. However, a rule might be identified as interesting when, in fact, the occurrence <X:A> does not imply the occurrence of <Y:B>. The occurrence of an itemset <X:A> is independent of the itemset <Y:B> if FS(Z:C) = FS(X:A)×FS(Y:B), otherwise itemsets <X:A> and <Y:B> are dependent and correlated as events. The correlation between the occurrence of <X:A> and <Y:B> can be measured by computing the interestingness of a given rule:

$$\mathrm{Fcorr}\left(\langle X : A\rangle, \langle Y : B\rangle\right) = \frac{\mathrm{FS}(Z : C)}{\mathrm{FS}(X : A) \cdot \mathrm{FS}(Y : B)} \tag{11}$$

If the resulting value of (11) is less than 1, then the occurrence of <X:A> is negatively correlated with the occurrence of <Y:B>. If the resulting value is grater than 1, then <X:A> and <Y:B> are positively correlated, meaning the occurrence of one implies the other. If the resulting value is near to 1 then <X:A> and <Y:B> are independent and there is no correlation between them.

MODEL STRUCTURE SELECTION

This section illustrates how the previously presented fuzzy association mining algorithm can be used to select the most relevant features of a datadriven model. The proposed method - MOSSFARM (Model Structure Selection by Fuzzy Association Rule Mining) - consists of the following steps:

Step 1: Transform crisp dataset into fuzzy
Step 2: Mine frequent itemsets
Step 3: Generate fuzzy association rules

Step A/4: Aggregate the rules for the selection of the input var*iables*
Step A/5: Determine the output by the rule base

Since in the previous section all of the functions needed to mine general fuzzy association rules were considered (*Step 1-3*), this section will focus on the remaining steps (*Step A/4* and *Step A/5*) that are needed to solve the studied feature selection problem.

Selection of the Relevant Input Variables

In some cases, not only is the generation of interesting fuzzy (association) rules, such as **If** *X* is *A* **then** *Y* is *B*, important, but it is necessary to select the most important input variables (feature selection). For this purpose, it is useful to aggregate the support, the confidence, and the correlation of the individual rules. A given *X* set of the input variables represent a certain class of rules (and frequent itemsets). Hence, it is possible to aggregate the measures of these rules $X \in R$, where *R* represents the set of the interesting rules:

$$\mathrm{FS}_X = \sum_{X \in R} \mathrm{FS}\big(\langle X : A \rangle \cup \langle Y : B \rangle\big) \tag{12}$$

$$\mathrm{FC}_X = \sum_{X \in R} \mathrm{FC}\big(\langle X : A \rangle \Rightarrow \langle Y : B \rangle\big) \frac{\mathrm{FS}\big(\langle X : A \rangle \cup \langle Y : B \rangle\big)}{\sum_{X \in R} \mathrm{FS}\big(\langle X : A \rangle \cup \langle Y : B \rangle\big)} \tag{13}$$

$$\mathrm{Fcorr}_X = \sum_{X \in R} \mathrm{Fcorr}\big(\langle X : A \rangle, \langle Y : B \rangle\big) \frac{\mathrm{FS}\big(\langle X : A \rangle \cup \langle Y : B \rangle\big)}{\sum_{X \in R} \mathrm{FS}\big(\langle X : A \rangle \cup \langle Y : B \rangle\big)} \tag{14}$$

With the help of the models the different sets of the input variables can be ordered, and based on this information a decision can be made about the model-structure for the linear or non-linear models of a system.

RULE BASED CLASSIFICATION

Fuzzy Association Rule Based Classifier

After all fuzzy association rules are generated (see *Step 1-3* in previous section) a fuzzy association rule based classifier can be built from the resulted rule base and the trapezoidal membership functions (generated by the partition method, an example is depicted in *Figure 1*). The fuzzy rule base includes the rules in the form of Expression 3. In a classification dataset the values of the output variable are the class labels. Therefore consequent part of the *j*th rule includes only the class label, $B_j \in \{C_1, C_2, \dots, C_{qy}\}$ where q_y are the number of classes.

Definition 7: A rule is said to fire when the conditions upon which it depends occur. Since these conditions are defined by fuzzy sets which have degrees of membership, a rule will have a *firing strength, β_j*. The firing strength is determined by the mechanism which is used to implement the *and* in the above expression 12, in this paper the product of the degrees of membership will be used:

$$\beta_j(x_k) = \prod_{\langle z_i : c_{i,j} \rangle \in \langle Z:C \rangle} t_k(z_i) \ , \tag{15}$$

used in the rule R_j on the k^{th} input $\mathbf{x}_k$. The class label is determined by the aggregating individual contributions based on the rule consequents (*B*):

$$w_k = \sum_{B} \beta_r(x_k), \ r = 1, \dots, M, \tag{16}$$

where the w_k is a ($q_y \times 1$) sized column vector, that includes the weights of classes. The class with the maximal weight will be the output of classifier model.

$$\hat{y}_k = \arg\max(w_k) \tag{17}$$

The main steps of the classification mechanism are the following:

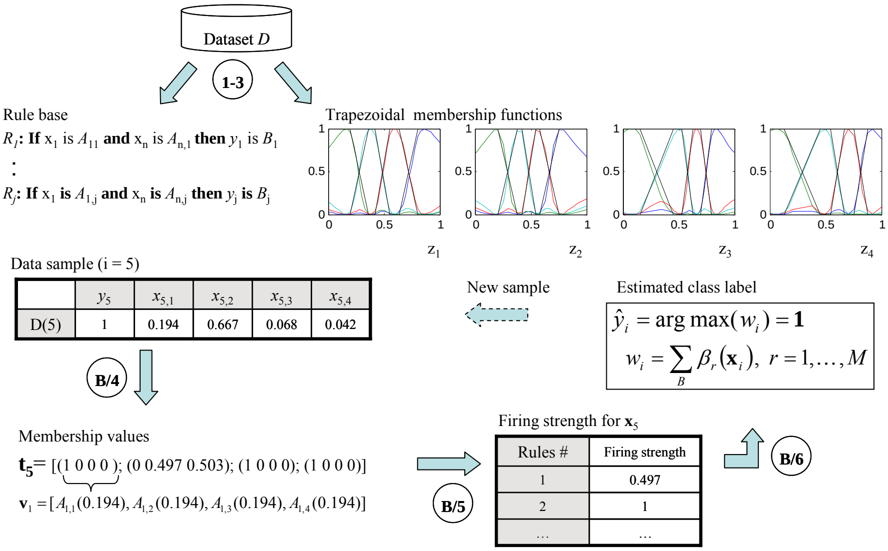
```
Step 1-3. generate the rule base and the trapezoidal membership functions

for $i = 1,\dots, N$
    Step B/4. calculate the $t_k$ membership values for data input $\mathbf{x}_k$ (Eq. 1)
    Step B/5. determine the firing strengths of all the rules (Eq. 15)
    Step B/6. aggregate individual contributions and determine $\hat{y}_k$ (Eq. 16-17)
end
```

In the previous sections *Step 1-3* are discussed, the *Step B/4 - B/6* are needed to solve a classification problem. The *Figure 2* shows an example for a classification method.

Figure 2

Classification Method


Dataset D
1-3
Rule base
R_1: If x_1 is A_{11} and x_n is $A_{n,1}$ then y_1 is B_1
:
R_j: If x_1 is $A_{1,j}$ and x_n is $A_{n,j}$ then y_j is B_j
Trapezoidal membership functions
1
0.5
0
0
0.5
1
z_1
z_2
z_3
z_4
Data sample (i = 5)
New sample
Estimated class label
$\hat{y}_i = \arg\max(w_i) = \mathbf{1}$
$w_i = \sum_B \beta_r(\mathbf{x}_i),\ r = 1,\dots,M$
B/4
Membership values
$\mathbf{t}_5$ = [(1 0 0 0); (0 0.497 0.503); (1 0 0 0); (1 0 0 0)]
$\mathbf{v}_1 = [A_{1,1}(0.194), A_{1,2}(0.194), A_{1,3}(0.194), A_{1,4}(0.194)]$
B/5
Firing strength for $\mathbf{x}_5$
B/6

	y_5	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$
D(5)	1	0.194	0.667	0.068	0.042

Rules #	Firing strength
1	0.497
2	1
…	…

2. ábra: Az osztályozás menete

RULE BASE PRUNING

Rule interesting measure and rule pruning

Association rule mining often results in a huge amount of rules. Attempts to reduce the size of the resulted rule base can be roughly divided to two categories *Goethals* (2005).

(1) In *subjective approach*, the user is offered some tools to specify which rules are potentially interesting and which are not, such as templates *Klemettinen* (1994) and constraints *Goethals* (2000) and *Ng* (1998).

(2) In *objective approach*, user-independent quality measures are applied on association rules. While interestingness is user-dependent to a large extent, objective measures are needed to reduce the redundancy inherent in a collection of rules. The objective approaches can be further categorized by whether they measure each rule independently of other rules (e.g., using support, confidence, or lift) or address rule redundancy in the presence of other rules (e.g., being a rule with the most general

condition and the most specific consequent among those having certain support and confidence values). Obviously only approaches of the latter type can potentially address redundancy between rules.

In *Jaroszewicz* (2003) contributions in both directions are contained: first a new interestingness measure is given generalizing three important known measures: chi-squared, entropy gain and Gini gain, and second, a method of pruning association rules using the Maximum Entropy Principle is presented. In the method CBA *Liu* (1998) pruning is done using the pessimistic error based method in C4.5. It prunes a rule *R* as follows: if rule *R*'s pessimistic error rate is higher than the pessimistic error rate of rule R^{-} (obtained by deleting one condition from the conditions of *R*), then rule *R* is pruned. For the computation steps of method, see *Quinlan* (1992).

In this paper a confidence measure based pruning method is proposed, because it is easy to use remove the complex rules, and the pruned rule base is efficient used for the classification.

Rule Base Pruning Based on Confidence

The number of all rules in the rule base is *M*, but the rules can be diverse length (a *rule length* is the number of antecedent fuzzy sets) and fuzzy confidence values belong to all rules. The advantage of the application of the fuzzy confidence measure is served by the monotonic property: if given a rule of length *i* with an FC value in the rule base, and a rule of length (*i*+1) contain added input variable, the FC value of the rule improves or it does not change. Based on this a rule base pruning algorithm has been developed that removes the unnecessarily complex rules. The rule pruning method can be formalized in the following:

The method starts with the comparing of the longest rules with the smaller. A large rule which contains the smaller rules are deleted from the rule base when the maximal FC value of the smaller rules is higher then the FC value of the large rule minus ε , the *correction factor* (initially is set to 2 percent). This rule pruning method gives smaller rules in the rule base. However the pruned rule base includes much less rules, the new classifier has about equal classification accuracy as with the use of unpruned rule base.

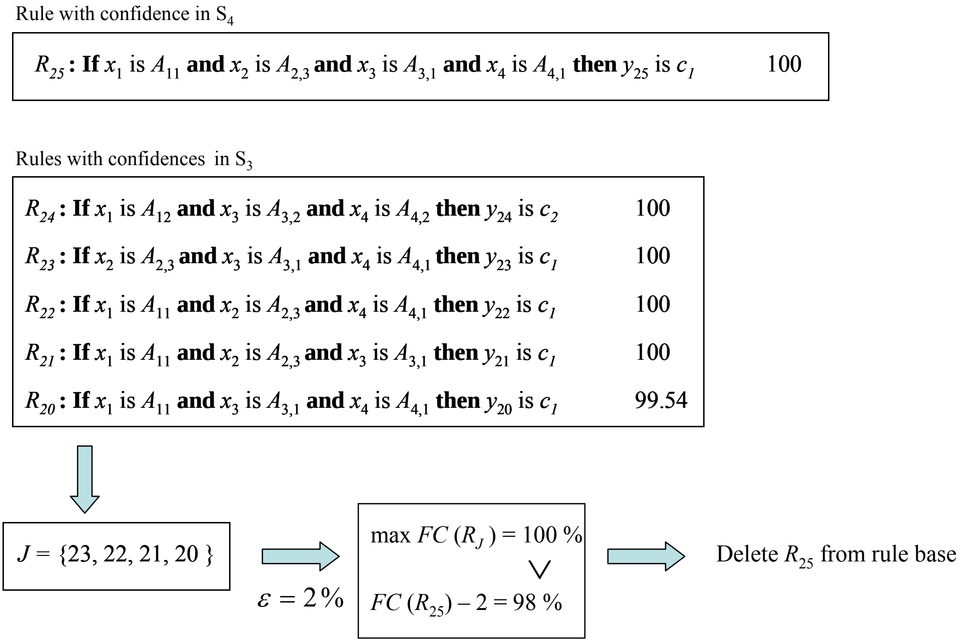
```
given the sets of several length rules: $S_{1,\dots,}S_L$
$L$=max length($R_l$), $l$ = 1,…,$M$
$J$ is an empty set

for $i=L,\dots,2$
    for all $R\in S_i$
        for all $R'\in R_{i-1}$
            if $size(R\cap R')$ = i
                $J = J\cup$ index of $R'$
            end
        end
        if max($FC(R_J)$) > $FC(R)$-ε
            delete $R$ from the rule base
        end
        clear $J$
    end
end
```

The *Figure 3* shows an example for this pruning method.

Figure 3

Example for Rule Pruning Method



Rule with confidence in S_4

R_{25}: **If** x_1 is A_{11} **and** x_2 is $A_{2,3}$ **and** x_3 is $A_{3,1}$ **and** x_4 is $A_{4,1}$ **then** y_{25} is c_1 100

Rules with confidences in S_3

R_{24}: **If** x_1 is A_{12} **and** x_3 is $A_{3,2}$ **and** x_4 is $A_{4,2}$ **then** y_{24} is c_2 100

R_{23}: **If** x_2 is $A_{2,3}$ **and** x_3 is $A_{3,1}$ **and** x_4 is $A_{4,1}$ **then** y_{23} is c_1 100

R_{22}: **If** x_1 is A_{11} **and** x_2 is $A_{2,3}$ **and** x_4 is $A_{4,1}$ **then** y_{22} is c_1 100

R_{21}: **If** x_1 is A_{11} **and** x_2 is $A_{2,3}$ **and** x_3 is $A_{3,1}$ **then** y_{21} is c_1 100

R_{20}: **If** x_1 is A_{11} **and** x_3 is $A_{3,1}$ **and** x_4 is $A_{4,1}$ **then** y_{20} is c_1 99.54

$J = \{23, 22, 21, 20\}$

$\varepsilon = 2\%$

max $FC(R_J) = 100\,\%$

$\vee$

$FC(R_{25}) - 2 = 98\,\%$

Delete R_{25} from rule base

3. ábra: Példa szabálytisztításra

The general applicability and efficiently of the developed tool are showed by an application study in the next Section. One general example for the feature (input) selection problem and the analysis of a polymerization process data. Moreover three general used classification problems show the applicability of the proposed classification method.

APPLICATION STUDY

Model structure selection problem

The first application example is originating from *Doyle* (1995), the aim is to generate the model order of a dynamical system based on the data generated by a simulation model of a continuous polymerization reactor. While the regression-tree induction method which applied to all of the 941 data points selected all of the eight variables x_1 through x_8, i.e. $y_{k+1} = f(y_k, y_{k-1}, \dots, y_{k-4}, u_k, u_{k-1}, \dots, u_{k-4})$, the MOSSFARM selected only 2-3 length model structures. The first five structures are depicted in Table 1. As can be seen, the model was able to select the correct model structure $y_{k+1} = f(y_k, u_k, u_{k-1})$ *Rhodes* (1998) where the number of the clusters are set to three for all input variables, and five for the output one (σ=1% and γ=75%).

Table 2

Selected model structure for a continuous polymerization reactor data

structure # (1)	selected variables (2)	FS (3)	FC (4)	Fcorr (4)
1	x_1, x_5, x_6	0.14	89	612
2	x_5, x_6, x_7	0.13	87	603
3	x_5, x_6	0.17	82	575
4	x_1, x_5, x_8	0.16	85	574
5	x_1, x_5	0.16	85	573

2. táblázat: Kiválasztott modell struktúrák a folytonos polimerizációs reaktor adatai alapján

Struktúra száma(1), Kiválasztott változók(2), Támogatottság(3), Bizonyosság(4), Korreláció(5)

Classification problems

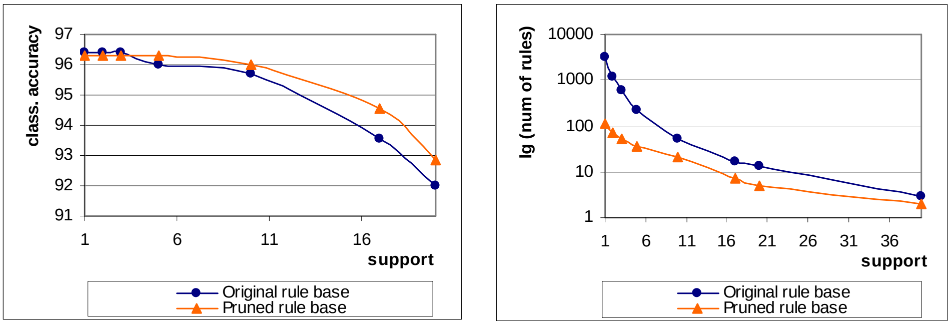
As it was presented the proposed method can be used not only for model structure or feature selection, it is also applicable for solving classification problems. The ***Wisconsine, Iris*** and ***Wine*** data sets are widely used classification problem for testing a new classification method.

The ***Wisconsin*** *Breast Cancer* data (WBCD) is available from the University of California, Irvine (UCI, *URL: http://www.ics.uci.edu/_mlearn/),* is a real classification problem.

The aim of the classification is to distinguish between benign and malignant cancers based on the available nine measurements: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nuclei, and mitosis. The attributes have integer value in the range [1;10]. The original database contains 699 instances however 16 of these are omitted because these are incomplete, which is common with other studies. The class distribution is 65.5% benign and 34.5% malignant, respectively.

Figure 4

Results of *WISCONSIN* problem



4. ábra: Eredmények a Wisconsin probléma esetében

Table 3 shows the summary results of the classsification where the confidence is 90 percent, and these results are depicted in *Figure 4*. While the value of the minimal support condition is high (~10-20%) the classification accuracy at the pruned rule base is higher than at the original rule base, but for low support the proposed method give better classification results at the original rule base. The disadvantage of the full rule base there are too many rules, e.g. accuracy is 96,42% but the number of rules is 594, if σ=3. The advantage of the pruned rule base is the „small" rule number with slightly less classification accuracy.

Table 3

Summary results of *WISCONSIN* problem

Support % (1)	Number of rules in original rule base (2)	Accuracy % (3)	Number of rules in pruned rule base (4)	Accuracy % (5)
20	13	91,99	5	92,85
17	17	93,56	7	94,56
10	52	95,70	21	95,99
5	219	95,99	37	96,28
3	594	96,42	54	96,28
2	1170	96,42	69	96,28
1	3185	96,42	109	96,28

3. táblázat: Eredmények összefoglalása a Wisconsin probléma esetében

Támogatottság(1), Szabályszám tisztítás előtt(2), Pontosság(3), Szabályszám tisztítás után(4), Pontosság(5)

The Wisconsin Breast Cancer data are widely used to test the effectiveness of classification and rule extraction algorithms. *Nauck* (1999) combined neuro-fuzzy techniques with interactive strategies for rule pruning to obtain a fuzzy classifier. An initial rule-base was made by applying two sets for each input, resulting in 2^9 = 512 rules which was reduced to 135 by deleting the non-firing rules. A heuristic data-driven learning method was applied instead of gradient descent learning, which is not applicable for triangular membership functions. Semantic properties were taken into account by constraining the search space. They final fuzzy classifier could be reduced to two rules with five to six features only, with a 95.06% classification accuracy. Rule-generating methods that combine GA and fuzzy logic were also applied to this problem *Pena-Reyes* (2000). In this method the number of rules to be generated needs to be determined a priori. This method constructs a fuzzy model that has four membership functions and one rule with an additional else part. *Setiono* (2000) has generated similar compact classifier by a two-step rule extraction from a feedforward neural network trained on preprocessed data.

As *Table 4* shows, our classifier generate more rules, but these rules are short, therefore MOSSFARM is a compact classifier with such high accuracy. It give 95.85% classification accuracy with the following parameters: the number of clusters (c) are three for all input variables, minimal support is 12 and minimal conditions is 90 percent. With the parameters, *c=3, σ=11%, γ=97%* the accuracy will be higher: 96.42%.

Table 4

Classification rates and model complexity for classifiers constructed for the *WISCONSIN* problem

	Method (1)	**Num of rules (2)**	**Num of conditions (3)**	**Accuracy (4)**
Setiono	NeuroRule 1e	1	4	97.36
Setiono	NeuroRule 1f	4	4	97.36
Setiono	NeuroRule 2a	3	11	98.10
Pena-Reyes & Sipper	Fuzzy-GA1	1	4	97.07
Pena-Reyes & Sipper	Fuzzy-GA2	3	16	97.36
Nauck and Kruse	NEFCLASS	2	10-12	95.06
This paper	MOSSFARM (c=4, σ=12, γ=90)	11	16	95.85
This paper	MOSSFARM (c=3, σ=11, γ=97)	8	14	96.42

4. táblázat: Osztályozási pontosság és model komplexitás a Wisconsin probléma esetében

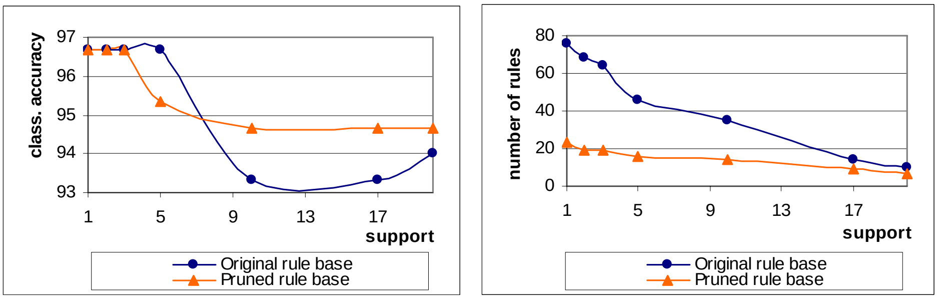
Támogatottság(1), Szabályszám(2), Pontosság(3)

The ***Iris*** dataset is available from UCI. It is perhaps the best known database to be found in the pattern recognition literature. One of the most frequently referred paper in the subject is is Fisher's. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are not linearly separable from each other. The results of our method are depicted in *Figure 5.*

While the value of support is low, the classification result is equal (Figure 6), but for high support values the classification results are higher at the pruned rule base where the number of rules are greatly smaller (with 66%). Therefore the method with rule pruning, gives high classification results by smaller rule base.

Figure 5

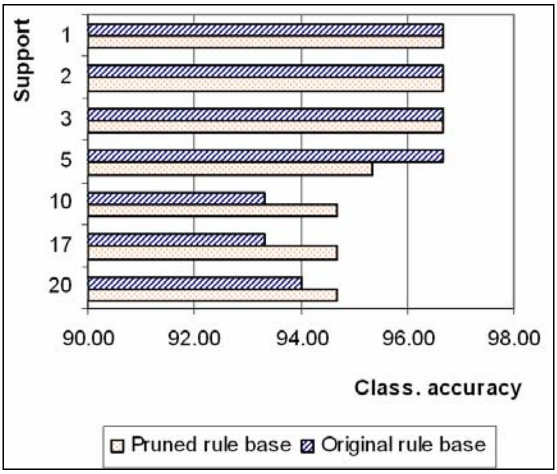
Results of *IRIS* class. problem



5. ábra: Eredmények az Iris probléma esetében

Figure 6

Comparation of results for the original and the pruned rule base at *IRIS* dataset



6. ábra: Eredmények az Iris problémánál az eredeti és a tisztított adatbázisok alkalmazásával

The ***Wine*** data contains chemical analysis of 178 wines grown in the same region in Italy but derived from three different cultivars. The data contains 59 instances from the first class, 71 from the second class, and 48 from the third class. The problem is to distinguish the three different types based on 13 continuous attributes derived from chemical analysis. As a pre-processing procedure, we normalized all attribute values into real numbers in the unit interval [0, 1]. Thus the wine data were transformed into a three-class pattern classification problem in the 13-dimensional unit cube $[0, 1]^{13}$. The results are showed in *Table 5*, where the number of clusters (c) and the minimal support condition are changed, but the confidence is always 90%.

Table 5

Classification rates and complexity at *WINE* dataset in function of clusters (partitions)

Support	**c = 3**		**c = 4**		**c = 5**		**c = 6**		**c = 7**		**c = 8**	
(1)	rules (2)	acc (3)	rules	acc	rules	acc	rules	acc	rules	acc	rules	acc
17	19	94.94	4	87.64	3	89.33	2	86.52	2	85.95	2	79.21
14	36	96.06	12	91.01	6	89.88	4	90.45	4	89.33	2	79.21
12	69	96.02	23	92.13	10	91.01	5	92.7	5	91.57	5	88.76
10	131	97.19	46	97.19	15	91.57	12	95.5	**12**	**96.63**	7	94.38
7	295	97.75	132	97.19	72	95.50	38	97.19	28	97.19	16	97.19

5. táblázat: Osztályozási pontosság és model komplexítás a partíciószám függvényében a Wine probléma esetében

Támogatottság(1), Szabályszám(2), Pontosság(3)

With smaller support value higher classification accuracy is resulted, but the classifier is too large (there are many rules). If the number of clusters (paritions) for all input variable is higher, the classifier will be compact, and comparatively accurate, see e.g. if *c=7, σ=10%, γ=90%*, the bold values in the *Table 5*.

CONCLUSIONS

In this paper we showed a new model-free, fuzzy association rule based method for the selection of the important variables of a data-driven model. The results show that the developed tool provides an efficient method for determining the order and structure of models, moreover it can be the base of classifier building. The method was able to select the correct model structure on the data generated by a simulation model of a continuous polymerization reactor. The proposed approach is also used for classification problems: Wisconsin, Iris and Wine are widely used to test the effectiveness of classification and rule extraction algorithms. Our new method give high classification accuracy and acceptable classifier complexity. The proposed methods have been implemented as a MATLAB program, and will be free available from *www.fmt.vein.hu/softcomp*.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of the Cooperative Research Centre (VIKKK, project II-1A) and the Hungarian Research Found (OTKA, No. T037600, No. T049534).

REFERENCES

Agrawal, R., Srikant R. (1994). Fast algorithm for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases. 487–499.

Agrawal, R., Imielinski T., Swami, A. (1993). Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering. 5. 6. 914–925.

Aguirre, L.A., Billings, S.A. (1995). Improved structure selection for nonlinear models based on term clustering. Int. J. Control, 62. 569–587.

Aguirre, L.A., Mendes, E.M.A.M. (1996). Global nonlinear polynomial models: Structure, term clusters and fixed points. Int. J. Bifurcation Chaos, 6. 279–294. p.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control, 19. 716–723.

Clark, P., Niblett T. (1989). The CN2 induction algorithm. Machine Learning, 3. 261–283.

Cohen, W. (1995). Fast effective rule induction. Proceedings of the 12th International Conference on Machine Learning, Tahoe City, CA, Morgan Kaufmann 115-123.

Dong, G., Zhang, X. Wong, L., Li, J. (1999). CAEP: classification by aggregating emerging patterns. Second International Conference on Discovery Science.

Doyle, F.J., Ogunnaike, B.A., Pearson, R.K. (1995). Nonlinear model-based control using second-order volterra models. Automatica, 31. 697–714.

Duda, R., Hart, P. (1973). Pattern Classification and Scene Analysis. JohnWiley & Sons : New York

Goethals, B., den Bussche, J.V. (2000). On supporting interactive association rule mining. Proceedings of the Second International Conference on Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science, Springer, 1874. 307-316.

Goethals, B., Muhonen, J., Toivonen, H. (2005). Mining Non-Derivable Association Rules SIAM International Data Mining Conference, Newport Beach, California

Gustafson, D.E., Kessel, W.C. (1979). Fuzzy clustering with fuzzy covariance matrix. In: Proceedings of the IEEE CDC, San Diego, 761–766.

Hong, T.P., Kuo, C.S., Chi, S.C. (1999). Mining association rules from quantitative data. Intelligent Data Analysis, 3. 5. 363–376.

Jaroszewicz, Sz. (2003). Information - theoretical and combinatorial methods in data mining. PhD Dissertation, University of Massachusetts, Boston

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, A.I. (1994). Finding interesting rules from large sets of discovered association rules. Third International Conference on Information and Knowledge Management (CIKM'94), Gaithersburg, MD, USA, ACM. 401-407.

Korenberg, M., Billings, S.A., Liu, Y., McIlroy, P. (1988). Orthogonal parameter estimation algorithm for nonlinear stochastic systems. Int. J. Control, 48. 193–210.

Kuok, C.M., Fu, A., Wong, M.H. (1998). Mining fuzzy association rules in databases. ACM SIGMOD Record, 27. 1. 41–46.

Li, W., Han, J., Pei, J. (2001). CMAR: Accurate and efficient classification based on multiple class-association rules. In: Proceedings of the 2001 IEEE International Conference on Data Mining (eds.: Cercone, N., Lin, T.Y., Wu, X.), San Jose, California, USA, IEEE Computer Society 369-376.

Liang, G., Wilkes, D., Cadzow, J. (1993). Arma model order estimation based on the eigenvalues of the covariance matrix. IEEE Trans. Signal Process, 41. 10. 3003-3009.

Lim, T.S., Loh, W.Y., Shih, Y.S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning, 40. 203-228.

Liu, B., Hsu, W., Ma, Y. (1998). Integrating classification and association rule mining. KDD'98, New York

Liu, B., Ma, Y., Wong C.K. (2000). Improving an Association Rule Based Classifier. Principles of Data Mining and Knowledge Discovery, 504-509.

Mendes, E.M.A.M., Billings, S.A. (2001). An alternative solution to the model structure selection problem. IEEE Trans. Syst. Man Cybernetics, Part A: Syst. Humans, 31. 6. 597-608.

Meretakis, D., Wuthrich, B. (1999). Extending Naive Bayes Classifiers Using Long Itemsets. Knowledge Discovery and Data Mining, 165-174.

Nauck, D., Kruse, R. (1999). Obtaining interpretable fuzzy classification rules from medical data. Artif. Intell. Med. 16. 149–169.

Ng, R.T., Lakshmanan, L.V.S., Han, J., Pang, A. (1998). Exploratory mining and pruning optimizations of constrained association rules. Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, SIGMOD Record, 27. 2. 13–24.

Pena-Reyes, C.A., Sipper, M. (2000). A fuzzy genetic approach to breast cancer diagnosis. Artif. Intell. Med. 17. 131–155.

Quinlan, J.R. (1992). C4.5: program for machine learning. Morgan Kaufmann : San Mateo, CA.

Rhodes, C., Morari, M. (1998). Determining the model order of nonlinear input/output systems. AIChE Journal, 44. 151–163.

Setiono, R. (2000). Generating concise and accurate classification rules for breast cancer diagnosis. Artif. Intell. Med. 18. 205–219.

Wang, K., Zhou, S., He, Y. (2000). Growing decision tree on support-less association rules. In: KDD'00, Boston, MA

Yin, X., Han, J. (2003). CPAR: Classification based on predictive association rules. In: Proceedings of 2003 SIAM International Conference on Data Mining (SDM'03)

Zimmermann, A., Raedt L.D. (2004) CorClass: Correlated Association Rule Mining for Classification. Discovery Science, 7th International Conference, Padova, Italy, 60-72.

Corresponding author (*Levelezési cím*):

János Abonyi
University of Pannónia, Department of Process Engineering
H-8201, Veszprém, P.O. Box 158
Pannon Egyetem, Folyamatmérnöki Tanszék
8201, Veszprém, Pf. 158.
Tel.: 36-88-624-447, Fax: 36-88-624-171
e-mail: abonyij@fmt.uni-pannon.hu