NEURO-FUZZY, FUZZY DECISION TREE AND ASSOCIATION RULE BASED METHODS FOR FUZZY COGNITIVE MAP GRADING PROCESS

E.I. Papageorgiou*, P.P. Groumpos

*Department of Electrical and Computer Engineering, University of Patras, Rion 26500, Tel: +302610997293, epapageo@ece.upatras.gr, groumpos@ece.upatras.gr

Keywords: decision making, fuzzy cognitive map, rule extraction, neuro-fuzzy methods, fuzzy rules.

Abstract

This work focuses on the formalization of a Fuzzy Cognitive Map based decision support system using fuzzy If-Then rules (extracted from data) accompanied with the available experts’ knowledge. The proposed approach is applied to build a Fuzzy Cognitive Map (FCM) grading tool, an advanced FCM-based model used for prediction. The FCM is a modeling methodology based on exploiting knowledge and experience. It can handle uncertainty and can be constructed by experts’ knowledge and the proposed fuzzy rules. The performance of FCMs is known to be sensitive to the initial weight setting and architecture. This shortcoming can be alleviated and the FCM model can be enhanced if a fuzzy rule base (IF-THEN rules) is available. The paper reports a successful attempt to combine FCMs with neuron-fuzzy, fuzzy decision tree and association rule based methods. These methods extract the available knowledge from data in the form of fuzzy rules and insert them into the FCM grading tool used for decision making tasks. This rule base could be derived by association rules, neuron-fuzzy approaches and knowledge extraction methods in general. In this research work, our scope is to introduce a new framework for decision making tasks and the proposed methodology is implemented in a well-known benchmark medical problem with real clinical data.

1 Introduction

This paper reports a first attempt to handle different data types for decision support tasks in medicine with soft computing techniques. It presents a trial to construct an advanced framework of combining different knowledge extraction methods for advanced Fuzzy Cognitive Mapping Decision Support in medical applications.

A decision-making procedure is a complex process that has to take under consideration a variety of interrelated functions. In Decision Support Systems (DSS) we are not only interested on the accuracy and prediction of the results (as in classification and data mining techniques) but for the transparency and interpretability of the results from the medical practitioner who uses the DSS in his daily clinical practice.

The a priori knowledge about a problem to be solved is frequently given in a symbolic, rule-based form. Extraction of knowledge from data, combining it with available symbolic knowledge, and refining the resulting knowledge-based expert systems is a great challenge for computational intelligence. Reasoning with logical rules is more acceptable to human users than the recommendations given by black box systems [1], because such reasoning is comprehensible, provides explanations, and may be validated by human inspection. It also increases confidence in the system, and may help to discover important relationships and combination of features, if the expressive power of rules is sufficient for that.

In the area of analysis different data types, the interpretability and simplicity of fuzzy systems is the key advantage. Fuzzy systems are not better function approximators or classifiers than other approaches. If we want to keep the model simple, the prediction is usually less accurate [2]. This means fuzzy systems should be used for data analysis, if an interpretable model is needed that can also be used to some extent for prediction. Furthermore, interpretability should not mean that anybody can understand a fuzzy system. It means that users, who are at least to some degree experts in the domain where the data analysis takes place, can understand the model. Obviously we cannot expect a lay person to understand a fuzzy system in a medical domain. It is important that the medical expert who uses the model should understand it [3].

A large number of possible computational intelligence approaches, from extraction knowledge from databases, have been developed [4]. Frequently, machine learning systems can be used to develop the knowledge bases used by expert systems. Given a set of clinical cases that act as examples, a machine learning system can produce a systematic description of those clinical features that uniquely characterize the clinical conditions. This knowledge can be expressed in the form of simple rules, often used for decision making in medicine [5,6, 7].

Among the wide range of possible approaches, the neuro-fuzzy, the fuzzy decision trees and the association rule based
methods were selected to extract the knowledge and construct a compact and useful fuzzy rule base. In developments of the new rule based methods for prediction applications besides the retention and enhancement of achieved accuracies (in the classification problems), the one of the most important objects is to enlarge the interpretability of the rules. Taking this aspect into account a possible way is the adaptation of fuzzy logic.

In a recent study [8], two approaches were proposed based on the association rule and decision tree inductive learning method, for generation of fuzzy rule base. Rule-extraction methods should not be judged only on the basis of the accuracy of the rules, but also on their simplicity and comprehensibility. Rule bases are efficiently used in many areas but this paper concentrates only in their use for medical decision making.

In this paper a fuzzy decision tree, a neuron-fuzzy system and a fuzzy association rule based method (Section 2) are introduced for fuzzy rule base generation. Our main goal is to show how construct compact fuzzy rule bases which are used for FCM based grading process and more general how they can be applied to form a DSS based on FCM methodology.

Fuzzy Cognitive Maps (FCMs) constitute an attractive modeling technique for complex systems. FCMs were proposed by Kosko to represent the causal relationship between concepts and analyze inference patterns [9]. FCMs represent knowledge in a symbolic manner and relate states, processes, events, values and inputs in an analogous manner. The knowledge which has been accumulated for years on the operation and behavior of a system can be adequately explicit using FCMs.

Fuzzy Cognitive Map can be seen as a collection of the rules such that it not only concerns the relationships between the causes and effects, but also considers the relationships among the causes. Therefore, it provides a stronger reasoning ability than rule-based reasoning and it can be used to model complex relationships among different concepts.

The conventional expert systems require the construction of a knowledge base which is elicited from the experts’ experience. Compared to expert systems, FCMs are relatively quicker and easier to acquire knowledge and experience, especially exploiting the human approaches who do not usually think in equations but in words. Using the FCMs construction methods someone can have as many knowledge sources as wanted with diverse knowledge and different degrees of expertise. These knowledge sources can all be combined into one FCM [10]. The main advantage is that there is no restriction on the number of experts or on the number of concepts.

The flexibility of FCMs in system design, model and control, as well as their learning properties [11,12,13], make their choice attractive for a variety of modeling and support decision tasks. FCMs were used in many disciplines for easy comprehension of complex social systems and for decision-making [14-17]. Also, few frameworks based on fuzzy cognitive maps for the task of reasoning and learning have been proposed [18,19]. Furthermore, multiple FCM models have been proposed for medical decision support systems (see in [13,20-22]). In previous works, FCM based methods have been proposed for characterizing tumours’ malignancy (urinary bladder and brain tumors) and a FCM grading tool, namely FCM-GT, was proposed for each case problem to classify the degree of tumour malignancy. The FCM-GT was constructed from the available knowledge from histopathologists-experts and the cause-effect relationships among the concepts were trained using an unsupervised learning algorithm [20,21].

This work proposes a methodology to develop a framework for decision support in medical systems using FCMs constructed by efficient and important fuzzy rules. The paper shows how some efficient computational intelligence techniques (based on fuzzy decision trees [23], association rules [24,25] and neuro-fuzzy networks (NEFCLASS) [4,26]) are very useful rules to rule extraction and data understanding and to construct FCM for decision support. The selected rules from the generated fuzzy rule base contain only the most important and most confidential rules that could be used for FCM grading process.

2 Three Rule Extraction Methods

The huge amount of medical data and the different sources of medical information make the task of decision making difficult and complex. Data mining and knowledge processing systems are intelligent systems that used in medicine for the tasks of diagnosis, prognosis, treatment planning and decision support [2].

There are several methods proposed for logical rule generation combining different data types (machine learning, fuzzy decision trees, association rules, Bayesian networks, neural networks, pattern recognition). We have selected the more powerful of these algorithms for the knowledge extraction from data and development of the specific fuzzy rule base; the neuro-fuzzy networks, the fuzzy decision trees and the fuzzy association rules. They have been proved from the literature that give better rules and keep the level of interpretability and accuracy in the classification tasks [1,2].

The interpretability of a fuzzy system – especially if applied in data analysis – is one of its key advantages. To support the readability of a fuzzy model resulting from a training process, one should use approaches that keep the learning algorithms simple – and therefore understandable – and do not touch the semantics of the underlying fuzzy models. The FCM keeps the transparency and interpretability of the process.
2.1 Neuro-fuzzy based method

Neuro-fuzzy systems play a very prominent role and they are applied to a variety of data analysis problems like classification, function approximation or time series prediction. Fuzzy data analysis in general and neuro-fuzzy methods in particular makes it easy to strike a balance between accuracy and interpretability. This is an interesting feature for intelligent data analysis.

An important aspect of intelligent data analysis is to select an appropriate model with the application in mind. It may be necessary to sacrifice precision for interpretability, i.e., a suitable balance between model complexity and comprehensibility, between precision and simplicity must be found. Intelligent data analysis also requires the selection of appropriate algorithms for the process of creating a model. Neuro-fuzzy models combine the learning capability of neural networks with the representational power of fuzzy inference systems, thus producing systems that can acquire knowledge from data and represent it in form of fuzzy rules, [1,4], [2], [3]. Unfortunately, the interpretability of fuzzy knowledge acquired by a neuro-fuzzy system may be heavily compromised by the learning phase of the network.

In order to extract knowledge to be judged “interpretable”, a set of properties on knowledge bases must be formulated. A lot of work has been done in this sense, resulting in different proposals of formal properties [27,28,29]. The definition of such properties leads to the development of learning methods to induce interpretable fuzzy rules from data [1,4], [2], [3], but many of these methods use time-intensive techniques.

From decision support systems we expect that they can learn, adapt to the users preferences, filter information, act on the behalf of the user, simplify complex information, are simple to use, etc. [1]. Neuro-fuzzy methods can help in achieving some of these goals, especially if we apply neuro-fuzzy methods that focus on interpretability. Neuro-fuzzy methods use different approaches to create rule bases. The neuro-fuzzy based method (Figure 1, on the center) is consisted on the following steps:

Step 1: In the first step a partitioning method is needed to get discrete data elements on continuous attributes. The applied method is a fuzzy clustering algorithm to determine trapezoidal fuzzy membership functions for each attribute (two fuzzy sets were considered for NEFCLASS).

Step 2: The user sets the maximum number of rules to n. The NEFCLASS approach is analyzed and transformed by a rule learning algorithm into a fuzzy rule base.

Step 3: The fuzzy rule learning algorithm processes the training data and determines the best consequent for a rule by a performance measure. In addition, the algorithm tries to reduce the size of the rule base by selecting only a number of rules depending on their performance.

Step 4: A post pruning process that deletes the unnecessary long rules is implemented.

The fuzzy rule learning algorithm proposed by Nauck [4] is the more efficient one, providing fuzzy rule base that can be used to build advance and enhancement FCM DSS systems.

![Figure 1. Main steps of the neuron-fuzzy (center), fuzzy decision tree (left) and the association rule (right) based methods.](image)

2.2 Fuzzy decision tree based method

Till recently years, many fuzzy decision tree induction algorithms have been introduced [31]. Fuzzy decision trees represent the discovered rules far natural for human (for example thanks to the linguistic variables). The [32] takes a detailed introduction about the non fuzzy rules and the different kind of fuzzy rules.

In classification problems the continuous attributes in the input domain need partitioning. For example in Figure 1 the attribute cell size is partitioned into two overlapped partitions (two fuzzy sets) small and large. Many type of membership functions can be used (triangular, trapezoids, Gaussian, etc.) for partitions. While the papers in the literature discuss various methods, this paper focuses only the a priori partition based fuzzy decision tree induction algorithms. At the a priori based methods, the partition step is ahead the tree induction step. The a priori partition and fuzzy decision tree based extraction method is showed at follows [33].

This approach (Figure 1, on the left) consists on the following steps:

Step 1: A fuzzy clustering algorithm is used for input domain partition. The supervised method takes into account the class labels during the clustering. Therefore the resulted partitions, the fuzzy membership functions (fuzzy sets) represent not only the distribution of data, but the distribution of the classes too.

Step 2: During a pre-pruning method the resulted partitions could analyze and combine the unduly overlapped fuzzy sets.
The results of the pre-pruning step are input parameters (beside data) for the tree induction algorithm. The applied tree induction method is the FID (Fuzzy Induction on Decision Tree) algorithm by C. Z. Janikow [34].

Step 4: The fuzzy ID3 is used to extract rules which are then used for generating fuzzy rule base.

Step 5: While the FID algorithm could generate larger and complex decision tree as it is necessary, therefore a post pruning method is applied. The rule which yields the maximal fulfillment degree in the least number of cases is deleted. This method provides compact fuzzy rule base that can be used for building FCM-DSS.

2.3 Fuzzy association rule based method

The association rule mining algorithms are the most frequently used data mining tools in rule extraction besides the others methods. Many different methods have been developed [3], [21,17,20], [35-37] but two main steps are common in most of them. The mining starts with frequent item set searching (it is defined first in paper [38]) then association rules are generated from the large item sets. The selection of an appropriate algorithm depends on the structure (sparse, dense) and the size of the analyzed database. Additionally the application area influences also notable the suitable methods.

An associative method which serves compact fuzzy rule bases from data is applicable in this work to build accurate fuzzy logic-based decision making tools. The next subsection presents the used fuzzy association rule based method. The following steps describe the proposed approach (Figure 1, on the right):

Step 1: A partitioning method is needed to get discrete data elements on continuous attributes. The applied method is a fuzzy clustering algorithm to determine trapezoidal fuzzy membership functions for each attributes.

Step 2: While the membership functions as fuzzy sets are counted for fuzzy items, the frequent item sets are searched on easy way. The membership values determine the supports of the items. The searching of the larger item sets is based on the a priori-principle [5].

Step 3: The association rules with class label in the consequent part are generated from the frequent item sets.

Step 4: A correlation measure selects the classification rules determine most the results of prediction. Only the positive correlated, above the average rules are stored in the rule base. These rules are called important rules.

Step 5: The unnecessary complex, redundant and conflict rules are searched during a post-pruning method [8]. The selected rules are removed from the rule base therefore only the most important and most confidential rules could be use for FCM grading tools.

2.4 Fuzzy cognitive map representation

FCM is a knowledge-based technique that follows an approach similar to the human reasoning and human decision-making process. FCM consists of nodes (concepts) which illustrate the different aspects of the system’s behavior. These nodes (concepts) interact with each other showing the dynamics of the model. Human experts who supervise a system and know its behavior under different circumstances develop a FCM model of the systems in such a way that their accumulated experience and knowledge are integrated in the causal relationships between factors/characteristics/components of the FCM model [10]. Figure 2 illustrates a graphical representation of a FCM.

Figure 2: A simple Fuzzy Cognitive Map

Human knowledge and experience are reflected in the selection of concepts and weights for the interconnections between concepts of the FCM. Each node-concept represents one of the key-factors of the modeled system and it is characterized by a number $A_i$ which represents its value. Interconnections among concepts of FCM signify the cause and effect relationship one concept has on the others. The cause and effect interconnection between two concepts $C_j$ and $C_i$ is described with the weight $w_{ji}$ taking value in the range –1 to 1. The value $A_i$ of the concept $C_i$ expresses the degree of its corresponding physical value. At each simulation step, the value $A_i$ of a concept $C_i$ is calculated by computing the influence of other concepts $C_j$’s on the specific concept $C_i$ following the calculation rule:

$$A_i^{(k+1)} = f(A_i^{(k)} + \sum_{j=1}^{N} A_j^{(k)} \cdot w_{ji})$$

Where $A_i^{(k+1)}$ is the value of concept $C_i$ at simulation step $k+1$, $A_j^{(k)}$ is the value of concept $C_j$ at simulation step $k$, $w_{ji}$ is the weight of the interconnection from concept $C_j$ to concept $C_i$ and $f$ is a sigmoid threshold function [39], that have been selected since the values $A_i$ of the concepts, lie within $[0,1]$.

The development and design of the appropriate fuzzy cognitive map for the description of a system requires the contribution of human knowledge. The experts develop fuzzy cognitive maps using an interactive procedure of presenting their knowledge on the operation and behavior of the system.

The procedure for constructing fuzzy cognitive maps is as follows: experts define the main concepts that represent the model of the system; they describe the structure and the interconnections of the network using fuzzy conditional statements. The fuzzy if-then rule that experts use to describe the relationship among concepts assumes the following form, where $A$ and $B$ are linguistic variables:

$IF$ value of concept $C_i$ is $A$ $THEN$ value of concept $C_j$ is $B$.

The overall rule describes the causal relationship between the value of concept $C_i$ and concept $C_j$, thus the weight between concept $C_i$ and concept $C_j$ can be inferred from the rule. The calculation of the overall rule for each interconnection is
based on the fuzzy inference Tsukamoto model [39] and the overall output is taken as the weighted average of each rule’s output.

3 Framework design for medical decision making

There is a necessity to develop a framework extracting the best fuzzy rules from data mining techniques for FCM grading process based on advanced methods of data mining. As it has already been stated, the central idea of the proposed technique is to combine different data driven methods to extract the available knowledge from data and to generate fuzzy If-Then rules. The resulted fuzzy rule base is applied to build an FCM grading tool used for prediction, or it can be applied to form a FCM-DSS. The derived FCM-DSS model is subsequently trained using an unsupervised learning algorithm to achieve improved decision accuracy and interpretability.

FCM-GT can be created by rules derived from data using association rules, fuzzy decision trees or neuron-fuzzy methods to create rule bases. This model can be constructed combining knowledge from the available fuzzy rule base. A simple discriminant method is used for the characterization of output concepts. According to the desired values of output concepts, the FCM-GT reaches a decision about the grading process and the degree of tumour malignancy (for details see e.g. in [20]).

Figure 3: The proposed approach for medical decision making using FCM grading tool The proposed approach for medical decision making using FCM-GT is not supposed to function as an automated process to create a fuzzy classification system, but as a tool to support user in constructing a decision support system. The new technique has three major advantages. First, the association rules derived from the knowledge extraction, data mining and pattern recognition techniques have a simple and direct interpretation and introduced in the initial FCM model to update its operation and structure. For example, a produced rule can be: If the variable 1 (input variable) has feature A Then the variable 2 (output variable) has feature B.

Second, the procedure that introduces the fuzzy rules evidences into an FCM also specifies the weight assignment through new cause-effect relationships among the FCM grading tool concepts. Third, this technique fares better in respect to the transparency and interpretability of the results from the medical expert than each one of data mining and the inductive learning techniques.

4 Illustrative application

This section shows an empirical analysis of the classification efficiency of the proposed algorithms. The Wisconsin breast cancer dataset [40] is one of the favourite benchmark datasets for testing classifiers. Properties of cancer cells were collected for 699 cases, with 458 benign (65.5%) and 241 (34.5%) malignant cases of cancer. They include nine attributes: clump thickness, uniformity of cell size, uniformity of cell shape, epithelial cell size, marginal adhesion, bare nuclei, normal nucleoli, and mitosis, all taking integer values ranging from one to ten. The problem is to distinguish between malignant and benign cancer cells.

In this section, we report some results obtained from experiments run with the proposed approach of FCM-GT and compared with experiments run with each one of the three programs of NEFCLASS (available in [26]), fuzzy decision inductive learning algorithm (available in [34]) and fuzzy association rule algorithm proposed in [24].

At first a partition method is used to assign the fuzzy sets of each one variable. These fuzzy sets according to the proposed framework are inserted into the three advanced computational approaches (Figure 1) to generate fuzzy If-Then rules. The selected a priori partition method was the supervised Gath-Geva clustering algorithm [41,42]. The number of the initial number of the partitions for all attributes were equal with the number of classes, two. The resulted partitions are represented in Figure 4.

NEFCLASS neuro-fuzzy system [26] with two trapezoidal membership functions per input feature generated 12 rules, and the “best per class” rule learning gave 95.4% correct answers. An example of some rules derived from NEFCLASS, have found:

- If Uniformity of Cell Size=1 or 2 Then benign,
else malignant
- If Uniformity of Cell Size Bland Chromatin
  Then malignant, else benign

The FCM-GT for breast cancer cannot be created initially
from doctors-experts for the problem of WBC data because
there is not available the experts’ knowledge to construct the
neuro-rule algorithm. In this case problem, the
FCM grading tool is constructed from the most important and
most confidential fuzzy rules (derived from rule base) to
determine the cause and effect relationships among the
nodes/concepts. The experts only determined the nodes of the
FCM model that represent the nine histological attributes and
the output node which represents the type of tumor cancer
(benign or malignant). 16 from the 699 instances are omitted
because these are incomplete. Only the full completed cases
are used in this empirical study.

The classification accuracy is measured by 10-fold cross
validation using a simple discriminant method on the values
of output concept that represents the type of tumour
malignancy (see an example in [20]). The average accuracy is
95.85% with 15 rules (from the three approaches). The result
of this new approach is useful and comparable with the other
methods already used for the same data set of WBC [4].

Table I summarizes the results from different approaches
presented in the literature. All experiments were run with 10-
fold cross-validation. C4.5 was run with standard
configuration. In NEFCLASS, for each attribute, a fuzzy
partition with two fuzzy sets was created, which were evenly
distributed over the attribute’s domain. The rule pruning
algorithm was also implemented on fuzzy sets since we tried
to generate comprehensible decision making models, a trade-off
between precision and complexity should be found.

These results were compared with other approaches already
applied in the same data set of WBC, the well-known
decision tree learner C4.5 (Release 8): [43,44], the
NEFCLASS J version, which can generate a fuzzy rule based
classifier by coupling neural networks with fuzzy systems
[45], the neuron-rule algorithm [46], the fuzzy-GA approach
[47] and the supervised fuzzy clustering approach [48]. The
models generated by these programs are compared w.r.t.
precision, complexity, and interpretability of the outputs.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree C4.5 [43]</td>
<td>95.1%</td>
</tr>
<tr>
<td>C4.5 rules [44]</td>
<td>95.4%</td>
</tr>
<tr>
<td>NEFCLASS [26]</td>
<td>95.06%</td>
</tr>
<tr>
<td>NEFCLASS J version [45]</td>
<td>96.7%</td>
</tr>
<tr>
<td>Neuro-rule [46]</td>
<td>98.1%</td>
</tr>
<tr>
<td>Supervised fuzzy clustering [48]</td>
<td>95.57%</td>
</tr>
<tr>
<td>Fuzzy-GA approach [47]</td>
<td>97.36%</td>
</tr>
<tr>
<td>Fuzzy Decision Trees [34]</td>
<td>96.0%</td>
</tr>
<tr>
<td>Proposed approach for FCM-GT</td>
<td>95.85%</td>
</tr>
</tbody>
</table>

Table I: Accuracies for a number of computational
approaches

The FCM-GT does not solve classification problems better
than other approaches do, for example, fuzzy decision trees,
neuron-fuzzy learning methods, fuzzy-genetic approaches are
better classification processes, but its simplicity and linguistic
interpretability are important factors for the acceptance of a
solution.

4 Conclusion

Our main aim of the proposed framework was not to achieve
better accuracies or to present a better classifier, but to
introduce a new FCM-GT enhancement by fuzzy rule base
constructed by efficient extraction of knowledge methods.
The new advanced FCM-DSS is simple, less complex,
transparent and interpretable to be accepted for medical
applications. The distinguishing feature of such FCM-DSS is
its situations with large amount of data, not enough
knowledge from experts and difficulty to handle the available
knowledge from many different sources of information.

In our opinion, revival of interest in DSS research crucially
depends on new frameworks and architectures that would
extend the cognitive capacities of DSS to meet the real world. We hope that our work points in this direction. Currently the work is underway to produce detailed specifications or comprehensive guidance on how to use the generic approach in other application domains.

Acknowledgements

The work of E.I. Papageorgiou was supported by a postdoctoral research grant from the Greek State Scholarship Foundation “I.K.Y.”.

References


(IIZUKA’96), Iizuka, Fukuoda, Japan, pp. 618-621, (1996).


