Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

SaEP12.3

A Combined Fuzzy Cognitive Map and Decision Trees Model for Medical Decision Making

Elpiniki Papageorgiou, Chrysostomos Stylios, *Member, IEEE* and Peter Groumpos *Member, IEEE*

***Abstract*—Fuzzy Cognitive Maps (FCMs) are an efficient modeling method providing flexibility on the simulated system's design. They consist of nodes-concepts and weighted edges that connect the nodes and represent the cause and effect relationships among them. The performance of FCMs is dependent on 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. This research proposes a successful attempt to combine fuzzy cognitive maps with decision tree generators. A combined Decision Tree-Fuzzy Cognitive Map (DT-FCM) model is proposed when different types of input data are available and the behavior of this model is studied. In this research work, we introduce a new hybrid modeling methodology for decision making tasks and we implement the proposed methodology at a medical problem.**

I. INTRODUCTION

Decision analysis is based on a number of quantitative methods that aid in choosing amongst alternatives [1]. Traditional decision analysis is used to indicate decisions favouring good outcomes even though there is an uncertainty surrounding the decision. Furthermore, the value of each possible outcome of a decision, whether measured in costs and benefits or utility, is usually variable.

As the number of treatment options and policy choices has exploded and the cost of medical research has skyrocketed, the "best" treatment for all clinical situations cannot be determined by conducting randomized controlled trials. Therefore, traditional decision analysis, in combination with sensitivity analysis, has become a standard methodology for using existing data and expert opinion to examine effectiveness and cost-effectiveness issues in health care [2-4].

Over the last years, several approaches have been examined, some attempts hybridise decision trees with other machine learning techniques, such as neural networks [5-7], Bayesian networks [8-9]. For instance, neural networks trained with back-propagation can represent a larger number of concepts than a decision tree. A decision tree though, with its greedy heuristics has the advantage of being able to quickly map out the general form of the concept space. A lot of the work in this area has of course been concerned with how to map the representations of decision trees to the necessary neural network representations [10-11].

Decision trees (DTs) belong to the state-of-the art techniques, which are used to make decisions from a set of instances. There are two types of nodes in a decision tree: decision nodes and leaves [12-14]. Leaves are the terminal nodes of the tree and they specify the ultimate decision of the tree. Decision nodes involve testing a particular attribute. Usually, the test at a decision node compares an attribute value with a constant. The decision tree is typically constructed by means of a ''divide-and-conquer'' approach. See Babic, et al. [15] for discussion of the methodology building DTs and Podgorelec et al. [16] for a bibliography and evaluation of DTs in the medical literature.

In medical decision making (diagnosing, classification, etc.) there are many situations where decision must be made effectively and reliably. Conceptual simple decision making models with the possibility of automatic learning are the most appropriate for performing such tasks. Decision trees are a reliable and effective decision making technique that provide high classification accuracy with a simple representation of gathered knowledge and they have been used in different areas of medical decision making [17,18]. Also, the soft computing technique of FCMs has been used for medical decision tasks in radiotherapy and for classification tasks in bladder tumour grading [19,20].

Where the DT is constructed, it is easy to convert the tree into a rule set by deriving a rule for each path in the tree that starts at the root and ends at the leaf node. The leaves of the DTs are used as concepts of the FCM model whereas their association rules through their confidence levels are used to initialize FCM weights. These rules in a form of IF-Then rules can be used to determine the weight settings and values of the FCM model. The enhanced FCM system, after the training through the unsupervised Nonlinear Hebbian Learning algorithm, is used to indicate the decision [21].

In this paper, the background of DTs and FCMs are presented and how they are combined to succeed decision making with the emphasis on existing and possible future applications in medicine. Also, the proposed DT-FCM model is implementing for the case of bladder tumours, the first results are encouraging.

II. DECISION TREES GENERAL CONCEPTS

Inductive inference is the process of moving from concrete examples to general models, where the goal is to learn how to classify objects by analyzing a set of instances

P. Papageorgiou is with Laboratory for Automation and Robotics, University of Patras, 26500 Patras, Greece (Corresponding author Tel +302610997293; Fax: +302610997309; Email: epapageo@ee.upatras.gr)

C. Stylios is with the Dept. of Communications, Informatics and Management, TEI of Epirus, Kostakioi, Artas, Greece (email: stylios@teiep.gr)

P. Groumpos is with Laboratory for Automation and Robotics, University of Patras, 26500 Patras (Email: groumpos@ee.upatras.gr)

(already solved cases) whose classes are known. Instances are typically represented as attribute-value vectors. Learning input consists of a set of such vectors, each belonging to a known class, and the output consists of a mapping from attribute values to classes. This mapping should accurately classify both the given instances and other unseen instances.

A decision tree is formalism for expressing such mappings and consists of tests or attributes nodes linked to two or more subtrees and leafs or decision nodes labelled with a class which means the decision. A test node computes some outcome based on the attribute values of an instance, where each possible outcome is associated with one of the subtrees. An instance is classified by starting at the root node of the tree. If this node is a test, the outcome for the instance is determined and the process continues using the appropriate subtree. When a leaf is eventually encountered, its label gives the predicted class of the instance. The finding of a solution with the help of decision trees starts by preparing a set of solved cases.

A decision tree can be built from a set of training objects with the "divide and conquer" principle. When all objects are of the same decision class (the value of the output attribute is the same) then a tree consists of a single node—a leaf with the appropriate decision. Otherwise an attribute is selected which value is of at least two different decision classes and a set of objects is divided according to the category of the selected attribute. The selected attribute builds an attribute (test) node in a growing decision tree, for each branch from that node the inducing procedure is repeated upon the remaining objects regarding the division until a leaf (a decision) is encountered [15,16].

The leaves of a decision tree are decisions and represent the value classes of the decision attribute – decision classes. Usually the members of a set of objects are classified as either positive or negative instances (for example ill and healthy patients), generally this approach has to be extended with multi-class decision making, enabling one to differentiate between various decision classes (to determine patients as low, medium and high for example). R. Quinlan influenced a large part of the research on Decision Trees construction [22].

The main advantages of DTs are their interpretability and the easy derivation of understandable and flexible (i.e., non-fixed) decision rules. In a decision tree, the "depth" of the tree only determines the maximum number of conditions that is used in decision rules. This is a maximum and non-fixed number. For this reason, the idea to integrate DTs with the following mentioned FCMs into a new decision tool was conceived.

III. Brief Description of Fuzzy Cognitive Maps

An FCM model stores the existence knowledge in the kind of concepts and in the type and value of the interconnections between concepts. Generally, concepts reflect attributes, characteristics, qualities and senses of the system. Each concept represents one of the key-factors of the modeled system and its value is represented by a number A_i. Interconnections among concepts of FCM signify the cause and effect relationship that a concept has on the others.

The structure of the FCM model can be viewed as a recurrent artificial neural network, where concepts are represented by neurons and causal relationships between concepts by weighted links connecting the neurons. These weighted interconnections represent the direction and degree with which a concept influences the value of the interconnected concepts [23]. A simple FCM with 5 concepts and 10 interconnections is illustrated in Figure 1.

Each interconnection between two concepts C_i and C_j, has a weight w_{ij}, belonging to the interval [-1,1]. Generally, the value of each node is calculated, computing the influence of other nodes to the specific node, by applying the following calculation rule:

$$A_i^{(k+1)} = f(A_i^{(k)} + \sum_{\substack{j \neq i \\ j=1}}^{N} w_{ij} \cdot A_j^{(k)}) \qquad (1)$$

where $A_i^{(k+1)}$ is the value of node C_i at time $k+1$, $A_j^{(k)}$ is the value of node C_j at time k, w_{ij} is the weight of the interconnection between node C_i and node C_j and f is the sigmoid threshold function.

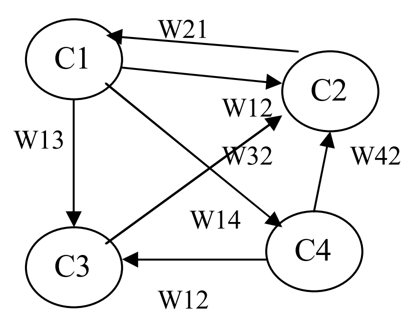

Fig.1 A simple Fuzzy Cognitive Map

Learning algorithms are means to increase the efficiency and robustness of FCMs, by selecting and modifying the FCM weights. Until now, two unsupervised learning algorithms, the Active Hebbian Learning and the Nonlinear Hebbian Learning have been proposed to train FCMs [21,24].

It may be interesting to explore the possibility of combining the advantages of decision tree induction in terms of understanding and simplicity with the advantages of FCMs in terms of modelling and simulation. In this paper, the results of such a study are reported and a new decision making methodology is proposed that combines FCMs with DTs, namely Decision Tree-Fuzzy Cognitive Map (DT-FCM) method.

IV. DESIGN THE DECISION TREE-FUZZY COGNITIVE MAP MODEL

As it has already been stated, the central idea of the proposed technique is to combine a decision tree (created by any decision tree algorithm, for example ID3) with the FCM. The derived FCM model is subsequently trained using an unsupervised learning algorithm to achieve improved decision accuracy. In this paper, the C4.5 has been chosen as a typical representative of the decision tree approach [22]. Similarly, the Nonlinear Hebbian Learning (NHL) algorithm is chosen as a representative of unsupervised FCM training.

The DT-FCM's function is briefly outlined in (Fig. 2). If there is a large number of input data then the quantitative data are used to induce a decision tree and qualitative data (through experts' knowledge) are used to construct the FCM model. The FCM's flexibility is enriched by the fuzzification of the strict decision tests (derived fuzzy IF-THEN rules to assign weights direction and values). Finally, the derived FCM model (new weight setting and structure) is trained by the unsupervised NHL algorithm to achieve a decision.

This methodology can be used for three different circumstances, depending on the type of the initial input data: (1) when the initial data are quantitative, the DT generators are used and an inductive learning algorithm produce the fuzzy rules which then are used to update the FCM model construction; (2) when experts' knowledge is available, the FCM model is constructed and through the unsupervised NHL algorithm is trained to calculate the target output concept responsible for the decision line; and (3) when both quantitative and qualitative data are available, the initial data are divided and each data type is used to construct the DTs and the FCMs separately. Then the fuzzy rules induced from the inductive learning restructure the FCM model enhancing it. At the enhanced FCM model the training algorithm is applied to help FCM model to reach a proper decision.

The new technique has three major advantages. First, the association rules derived from the decision trees 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 decision tree rules into an FCM also specifies the weight assignment through new cause-effect relationships among the FCM concepts. Third, as will be demonstrated through the experiments, this technique fares better than the best decision tree inductive learning technique and the FCM decision tool.

Actually, our research group works on improving the medical diagnosis process by different means: (1) introducing a methodology based on FCMs for decision making in complex medical systems where experts' knowledge is available; (2) constructing modular FCMs for characterizing tumor grading; and (3) certain histopathological features-attributes (for example, cell distribution, nuclei, mitosis, necrosis) have been converted into discrete values, although their conceptual vagueness could be quantified by the degree of membership of a numerical value in a fuzzy set. Thus their values would be a user defined finite set of linguistic values.

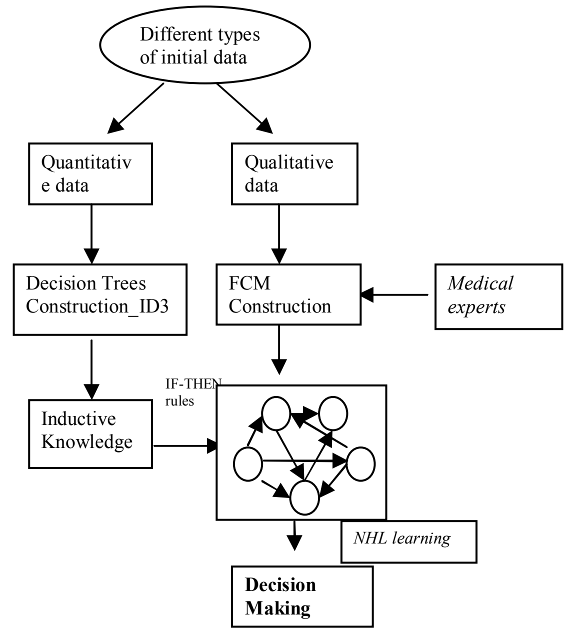

Fig. 2: The decision making system constructed by Decision Trees and Fuzzy Cognitive Maps

V. IMPLEMENTATION OF THE DT-FCM MODEL FOR BLADDER TUMOR GRADING

Ninety-two cases of urinary bladder cancer were collected from the archives of the department of pathology of University Hospital of Patras Greece. Histopathologists-experts had diagnosed 63 cases as low-grade and 29 as high-grade using conventional WHO grading system. Following grade diagnosis, each tissue section was evaluated retrospectively, using a list containing eight well documented in the bibliography histopathological criteria essential for tumour grading [20]. The FCM model for tumor grading had been developed and presented analytically in [20]. The FCM grading tool was able to give distinct different values for the majority of high-grade and low-grade cases using a simple Bayesian classifier for the output data. Except the experts' knowledge for determining FCM model, quantitative data for the eight main histopathological features [20,25] were also available and used for constructing DT. Then through the inductive learning procedure, a set of association rules were derived. Some of the best association rules, based on their confidence levels, are given in the Table1. The necessary If-Then rules were induced and introduced in the FCM model enhancing its initial structure.

After the development of the DT-FCM model and the determination of specifications for the implementation of the NHL algorithm, the hybrid system was used to examine cases and assigned sensitivity and specificity for grading bladder tumors. The same data set that used in previously proposed FCM-TG model, were also used to evaluate the performance of the DT-FCM methodology in categorizing tumors as low grade or high grade. The results for average sensitivity and average specificity for the ninety two bladder tumour cases were 80% and 90% respectively using the DT-FCM, whereas the resulting accuracies for low grade and high grade cases were 79% and 87.5% through the FCM grading tool [25].

TABLE I
EXAMPLE OF ASSOCIATION RULES DERIVED FROM DECISION TREES

Rules	Result/Decision Leaves
Cell-size=uniform, mitosis=absent rate	Grade Low
Cell-distribution=even, nucleoli=inconspicuous	Grade Low
Cell-distribution=clustered, cell-size=pleomorphic	Grade High
Nuclei=uniform, mitosis=absent rate	Grade Low
Cell-size=uniform, cell-number=numerous, nucleoli=inconspicuous	Grade Low

Our obtained results through the implementation of the proposed DT-FCM methodology are very promising and encourage us to continue our effort towards this direction.

VI. CONCLUSION

In this research work, the idea of combining decision tress with FCMs was explored in order to maintain the potential advantages of both techniques. The new integrated system has been introduced to assist medical decision making process. This work proposes a new framework of Fuzzy Cognitive Map utilizing Decision Trees that updates the traditional Fuzzy Cognitive Map and has better characteristics. This paper proposes the inclusion of decision tree generators in the structure of the FCM, and the new DT-FCM system gives better results. The performance of the new methodology can deal with different kind of input data eliminating numerical errors. In future work, we are going to compare the results of the DT-FCM system for assessing tumours' grading with other machine learning approaches.

ACKNOWLEDGMENT

The work of E.I. Papageorgiou was supported by a postdoctoral research grant from the Greek State Scholarships Foundation (I.K.Y.).

REFERENCES

[1] J.H.C. Sox, M.A. Blatt, M.C. Higgins, K.I. Marton, *Medical Decision Making*, Boston, Massachusetts: Butterworths, 1988.
[2] T.G. Dietterich, "Machine-learning research: Four current directions", The AI Magazine, 18(4), pp. 97–136, 1998. URL: http://citeseer.nj.nec.com/dietterich97machine.html
[3] B. Cremilleux, and C. Robert, "A theoretical framework for decision trees in uncertain domains: Application to medical data sets," *Lecture Notes in Artificial Intelligence,* Vol. 1211, pp. 145–156, Springer-Verlag, 1997.
[4] J.R.Quinlan, "Decision trees and decision making," *IEEE Trans System, Man and Cybernetics,* 20(2), pp. 339–346, 1990.
[5] F D'alche-Buc, D Zwierski, J.Nadal, "Trio learning: a new strategy for building hybrid neural trees," *Neural Syst.* Vol.5(4), pp.255–74, 1994.
[6] R. Krishnan, G. Sivakumar, & P. Bhattacharya, "Extracting decision trees from trained neural networks," *Pattern Recognition*, 1999, 32(12):1999-2009.
[7] M. Kubat, "Decision trees can initialize radial-basis function networks," *IEEE Trans Neural Networks*, 9(5):813-821, 1998.
[8] D. Janssens, G. Wets, T. Brijs, K. Vanhoof, T. Arentze, H. Timmermans, "Integrating Bayesian networks and decision trees in a sequential rule-based transportation model," *Europ J Operat Research*, in press (2005).
[9] D. Heckerman, D. Geiger, D.M Chickering, "Learning Bayesian networks: the combination of knowledge and statistical data," *Machine Learning*, 20 (1995) 197-243.
[10] W. Leow, R. Setiono, "On mapping decision trees and neural networks," *Knowledge Based Systems* 12, 1999, 95–99.
[11] G.P.J. Schmitz, C. Aldrich, & F.S. Gouws, "ANN-DT: An algorithm for extraction of decision trees from artificial neural networks," *IEEE Transactions on Neural Networks*, 10(6), pp.1392-1401, 1999).
[12] R. Setiono, H. Liu,. "A connectionist approach to generating oblique decision trees," *IEEE Transactions on Systems, Man, Cybernetics — Part B: Cybernetics*, Vol. 29(3), pp.440-444, 1999.
[13] D.McSherry, "Strategic induction of decision trees", *Knowledge-Based Systems*, 12(5-6), pp. 269-275, 1999.
[14] J.R. Quinlan, "Induction of decision trees," *Machine Learning*, 62(1):81-106, 1986.
[15] S.H. Babic, P. Kokol, V. Podgorelec, M. Zorman, M. Sprogar, M.M. Stiglic, "The Art of Building Decision Trees", *J Med Syst.*, 24(1), pp. 43-52, 2000.
[16] V. Podgorelec, P. Kokol, S B. tiglic, and I. Rozman, "Decision Trees: An Overview and Their Use in Medicine," *Journal of Medical Systems, Vol. 26, No. 5, October 2002*
[17] P. Kokol, *et al,* "Decision trees and automatic learning and their use in cardiology," *J. Med. Systems* 19(4): 1994.
[18] Z. Ping, C. Lihui, "A novel feature extraction method and hybrid tree classification for handwritten numeral recognition," Pattern Recognition Letters 23, pp. 45, January 2002.
[19] E., Papageorgiou, C. Stylios, P. Groumpos, "An Integrated Two-Level Hierarchical Decision Making System based on Fuzzy Cognitive Maps (FCMs)," *IEEE Trans Biomed Engin*, Vol. 50(12), pp.1326-1339, December 2003.
[20] E.I., Papageorgiou, P.,Spyridonos, P., Ravazoula, C.D. Stylios, P.P. Groumpos, G. Nikiforidis, "Advanced Soft Computing Diagnosis Method for Tumor Grading:, *Artif Intell Med*, Vol. 36 (2006) 59-70.
[21] E.I. Papageorgiou, P.P. Groumpos, "A weight adaptation method for fine-tuning Fuzzy Cognitive Map causal links," *Soft Computing Journal*, Vol. 9 pp. 846-857, 2005.
[22] J.R. Quinlan, *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann, 1993.
[23] B.Kosko, *Neural Networks and Fuzzy Systems*, Prentice-Hall, New Jersey, (1992).
[24] E.I. Papageorgiou, C.D. Stylios, P.P. Groumpos, "Active Hebbian Learning to Train Fuzzy Cognitive Maps," *Int. J. Approx. Reasoning*, Vol. 37 (2004) 219-249.
[25] E.I. Papageorgiou, P. Spyridonos, P. Ravazoula, C.D. Stylios, P.P. Groumpos, G Nikiforidis, "The Challenge of Using Soft Computing Techniques for Tumor Characterization", *LNCS* 3070, pp. 1031-1036, 2004.