ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 5, May 2014

Efficiency And Time Complexity Improvement Of Hybrid Fuzzy-Neuro Model Using Rule Pruning

Author: Mosam Patel[1]

Affiliation: GTU,Gujarat[1]

ABSTRACT

Hybrid models have attracted many researchers since last decade. Many hybrid models are proposed using soft computing skills. Hybrid fuzzy-neuro model is basically used for applications or projects that require high accuracy along with good comprehensibility. Neural Networks are known for their high accuracy, whereas fuzzy logic is known for its interpretable results. Fuzzy-Neuro model incorporate the goodness of both the systems, but is suffering from the curse of generating large and complex rule base. None of the existing rule pruning methods work upon an all length rule base. So a new rule pruning procedure for all length rules based on confidence and support is proposed here. A simplified rule base is obtained by the proposed system while getting the accuracy at par with existing system. The experiment carried out using the 3 well-known datasets for Credit risk analysis prove that the proposed rule pruning method for HFNN helps to improve upon its time complexity and efficiency.

General terms

Datamining, Fuzzy Data Classification.

Keywords

Fuzzy Logic, Neural Network, Confidence, Support, Fuzzy Classification, Hybrid Fuzzy-Neuro Model.

1. INTRODUCTION

Since last decade, the greedy researchers have been making use of the combination of the Fuzzy logic systems and the Neural Networks, so as to get the benefits of both of the systems. Fuzzy-Neuro models thus incorporate the goodness of both the systems. But the main problem with such systems is its less interpretability/comprehensibility. So to make such systems more interpretable rule simplification techniques are used. Rule Simplification methods decrease the complexity of rules thus making them easily interpretable. Since last decade, the Credit Risk Management has grabbed many researchers' attention, to work on it and to improve the efficiency and accuracy of the methodology used for credit risk calculations. Credit Risk analysis helps in evaluating the credit worthiness of the customer.

The fuzzy logic systems are used for applications that have high level of human interaction like taking decisions etc, and also used for applications that require interpretable outputs. The plus points of Fuzzy systems are that, it is capable to represent the uncertainties of the human knowledge using linguistic variables. Because of the natural rules representation, the results are easily interpretable. The Neural Networks are basically used for learning the examples. The plus points of neural networks are it has high learning capacity, generalization capacity and its robustness in relation to disturbances. Two possible samples of HFNN Models are shown in figure 1.1. As the hybrid system contains goodness of both of these systems, the resulting hybrid Fuzzy-Neuro system is interpretable, self learned and comprehensible.

The three conditions for a KB to be interpretable are 1. Use interpretable fuzzy partitions, 2. Use a small number of rules, and 3. Use compact rules for large systems. Rule pruning tends to simplify the rule base. The rules generated can be simplified in number of ways: removing redundant rule, shortening the rule and merging the rules. This way the rules can be simplified thus increasing the interpretability of system. I am proposing a solution to credit risk analysis's methodology that achieves interpretability, but not at the cost of accuracy. The efficiency is intended to improve by minimizing the number of rules generated. I propose a new solution for credit risk analysis that is more interpretable and at the same time, take care that accuracy is not suffered. The interpretability planned to be achieved is in terms of reducing the number of rules generated by simplifying the rule base.

2. LITERATURE SURVEY

The first study of the neuro-fuzzy systems began in the 90's decade, with Jang, Lin and Lee in 1991, Berenji in 1992 and Nauck from 1993, etc. The majority of the earlier applications were of process control. Gradually, the applications of these combo spread for all the areas of the knowledge like, data

ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 5, May 2014

analysis, data classification, imperfections detection and support to decision-making, etc.

In papers [1] [8] [7] [6] the authors have presented different hybrid soft computing methods to do CRA. These methods are successful in achieving the accurate results, but where [6] produced just accurate results, [8] [7] produced accurate but less comprehensible results (due to unprocessed input variables). In papers [2] [3] [4] [5] [9] various rule pruning methods for fuzzy rules is presented. These methods however generating all length rules but is not feasible to apply for reducing an already present all length rule base. From various literatures surveyed it's clear that HFNN method in [1] processed input variables to find their relevance in deciding class label and produced accurate results but has following limitations: a) The method used to learn most relevant attributes from NN is vulnerable to dataset with more number of attributes. b) The Rule set generated by this method is large and redundant. However due to shortcoming of existing rule pruning methods a new rule pruning method working for already present all length rule base is required.

3. PROPOSED SYSTEM

I am proposing a combined HFNN system with rule pruning so as to make the rule base simple, thus increasing the interpretability. The rules are treated as the association rules and from the formula proposed in [5] the support and confidence of each rule is found. The most relevant rules are then learned from the existing rule base using the proposed rule pruning method based on support and confidence. The rules that are redundant and having lower support and confidence are pruned off.

3.1 Comparison of HFNN with proposed HFNN with rule pruning

Table 1 Comparison of existing and proposed system

HFNN	**Proposed HFNN with Rule Pruning**
1. Fuzzified data is encoded <u>Example:</u> - If after fuzzification the data falls into membership function 3 out of total 5 membership functions for that attribute, then, the encoded form of that data is 0 0 1 0 0.	1. As per [10], the encoding step is required to specify the detail of fuzzy members in NN layers and it is required to do so if rules are to be induced from the weights of neurons of NN.
2. The number of input neurons in NN is equal to the total number of the fuzzy members of all attributes of dataset. <u>Example</u>: - For 20 input attributes, each having 4 fuzzy members the total number of input neurons in NN is 80.	The number of input neurons in NN is equal to the total number of attributes of dataset. <u>Example</u>: - For 20 input attributes, the total number of input neurons in NN is 20.
3. The dataset is divided into 3 parts: Training set, validating set and testing set.	3. The dataset is divided into 2 parts: Training set and testing set.
4. The most relevant attributes are selected by trying out different combinations of the input attributes and keeping those attributes that give highest accuracy for NN with validating set. <u>Example</u>: - If there are n attributes then total (2^n) -1 different combinations are generated. So for 16 attributes 65,535 combinations are to be tried. For 20 attributes 10,48,575 combinations are to be tried.	4. The most relevant attributes are evaluated from the training set by finding the worthiness of a subset of attributes by considering the individual predictive ability of each attribute along with the degree of redundancy between them. The subsets of attributes that are highly correlated with the class while having lower inter-correlation are preferred.
5. The rules that are used 0 or 1 times are pruned, and rest rule set is used for fuzzy classification.	5. The redundant rules are pruned using the proposed rule pruning procedure and resultant rule set is used for fuzzy classification.

3.2 Proposed system's flow

The highlighted portion in the graph shows the steps in which modifications to the existing system as mentioned in table 1 is done.

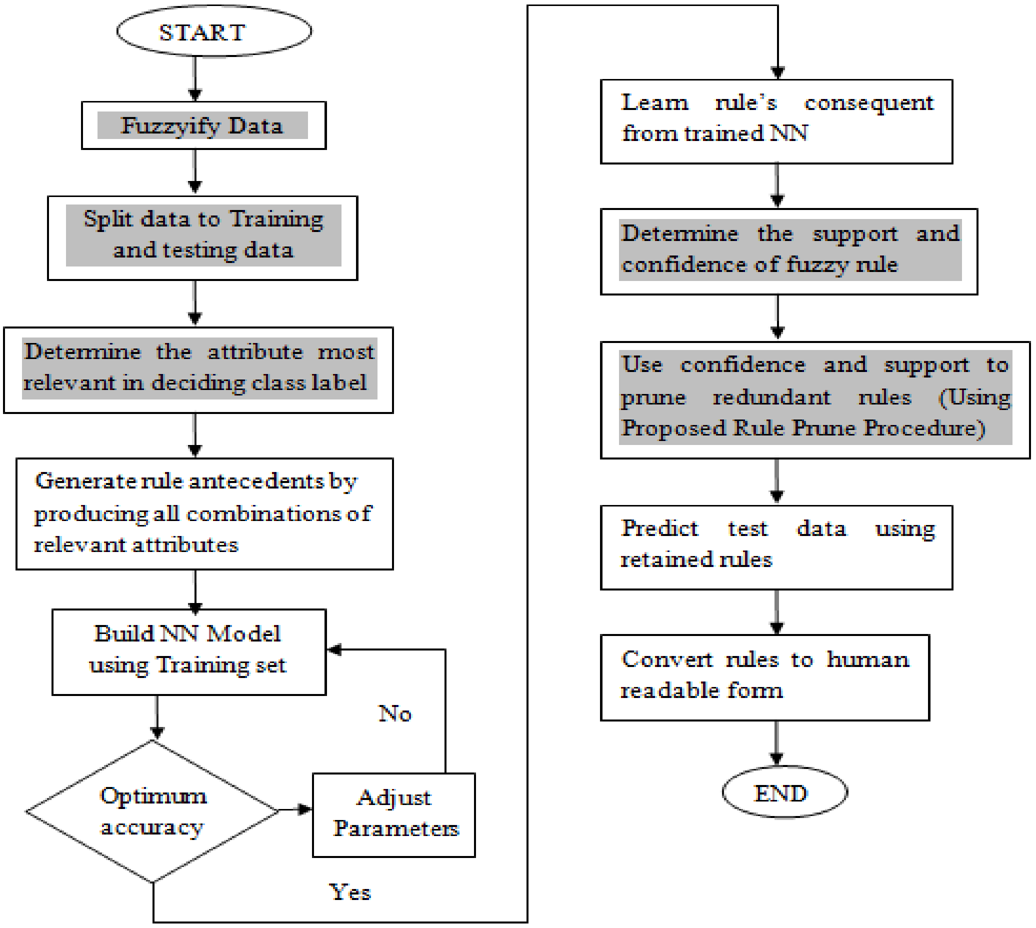

Figure 1 Proposed HFNN system's flow

3.3 Experimental Parameters

- The parameters considered for credit risk analysis is various borrower characteristics – such as character (reputation), capital (leverage), capacity (volatility of earnings), residence type, as well as credit history, etc.
- For Neural Networks the parameters like Hidden layer size, no. of epochs, learning rate, momentum and the output labels are considered.
- For fuzzy system the triangular fuzzy membership function is taken for fuzzifying the data. Fuzzy rules are used for classification of testing data.
- For Fuzzy classification accuracy, number of fuzzy rules, number of antecedents per rule, Support and confidence of fuzzy rule and time required to classify are considered.

4. EXPERIMENTAL RESULTS

The experiment was carried out on three datasets[11]: German bank dataset, Bank Marketing dataset and Adult income census dataset. The adult dataset contained missing values for categorical attributes that were replaced by maximum occurring category.

The table 2 shows the time complexity analysis of HFNN and proposed HFNN with rule pruning for all three datasets.

Table 2 Time complexity analysis

	HFNN	**HFNN with Rule pruning**
German Bank Dataset (20 input attributes and 1000 instances)	1. Attribute selection from NN based on accuracy. Time Taken : 3 ½ days	1. Attribute selection based on context sensitivity for class. Time taken : 10 msec
	2. No Rule Pruning	2. Rule pruning

		Time Taken: 0.9 sec
	3. Classification Time Taken: 0.63 sec	3. Classification Time Taken: 0.4 sec
Bank Marketing dataset (16 input attributes and 45211 instances)	1. Attribute selection from NN based on accuracy obtained. Time Taken : 3 days	1. Attribute selection based on context sensitivity for class. Time taken : 9 msec
	2. No Rule Pruning	2. Rule pruning Time Taken: 8.87 sec
	3. Classification Time Taken: 16.2 sec	3. Classification Time Taken: 1.45 sec
Adult income census dataset (14 input attributes and 48842 instances)	1. Attribute selection from NN based on accuracy obtained. Time Taken : 10 hrs	1. Attribute selection based on context sensitivity for class. Time Taken : 10 msec
	2. No Rule Pruning	2. Rule pruning Time Taken: 12.8 sec
	3. Classification Time Taken: 56.12 sec	3. Classification Time Taken: 1.5 sec

The Table 3 shows efficiency analysis of rules from HFNN and HFNN with rule pruning.

Table 3 Efficiency Analysis

	German Bank Dataset		**Bank Marketing dataset**		**Adult income dataset**	
	HFNN	HFNN with Rule pruning	HFNN	HFNN with Rule pruning	HFNN	HFNN with Rule pruning
No. of rules	749	14	2243	22	12499	26
No. of antecedants / rule	3.23	1.64	3.94	2.18	4.75	3.19

The table 4.3 shows the accuracy results of the two methods (with 70-30% training-testing set split.)

	German Bank Dataset	**Bank Marketing dataset**	**Adult income dataset**
HFNN	100	97.05	95.44
HFNN with Rule pruning	99	97.05	95.38

5. CONCLUSION

The proposed rule pruning method worked well in simplifying the large rule base and thus increasing the interpretability of fuzzy rules. The various credits risk analyzers and credit score models had shown a kind of trade-off between accuracy and the interpretability which was observed in the proposed method in negligible amount. The proposed HFNN with Rule pruning outperforms HFNN in terms of improvement in time complexity and efficiency.

6. ACKNOWLEDGMENTS

I would like to thank my source of inspiration Mr. Ghanshyam Pandey for supporting me to carry out a research on the underlying topic.

7. REFERENCES

[1] Elmer P. Dadios, James Solis, "*Fuzzy-Neuro Model for Intelligent Credit Risk Management*", Intelligent Information Management, 2012, 4, 251-260 doi:10.4236/iim.2012.425036 Published Online October 2012 (http://www.SciRP.org/journal/iim)

[2] Jose. M. Alonso, Luis Magdalena, Serge Guillaume, *"HILK: A New Methodology for Designing Highly Interpretable Linguistic Knowledge Bases Using the Fuzzy Logic Formalism",* International journal of intelligent systems, vol. 23, 761–794 (2008) Wiley Periodicals, Inc.

[3] Min-You Chen∗, D.A. Linkens,"*Rule-base self-generation and simplification for data-driven fuzzy models* ", Fuzzy Sets and Systems 142 (2004) 243–265. Elseiver ScienceDirect.

[4] K. Shehzad, "*Simple Hybrid and Incremental Postpruning Techniques for Rule Induction*", IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 2, February 2013

[5] Hisao Ishibuchi, Member, IEEE, and Takashi Yamamoto, Student Member, IEEE, "*Rule Weight Specification in Fuzzy Rule-Based Classification Systems*", IEEE Transactions On Fuzzy Systems, Vol. 13, No. 4, August 2005.

[6] Lean Yu, Shouyang Wang, Kin Keung Lai, "*Credit risk assessment with a multistage neural network ensemble learning approach*", Expert Systems with Applications 34 (2008) 1434–1444. Elseiver ScienceDirect.

[7] Chih-Fong Tsai, Ming-Lun Chen, "*Credit rating by hybrid machine learning techniques*", Applied Soft Computing 10 (2010) 374–380. Elseiver ScienceDirect.

[8] Nan-Chen Hsieh, "*Hybrid mining approach in the design of credit scoring models*", Expert Systems with Applications 28 (2005) 655–665. Elseiver ScienceDirect.

[9] D. P. Pancho, J. M. Alonso, O. Cordón, A. Quirin, and L. Magdalena, "*FINGRAMS: visual representations of fuzzy rule-based inference for expert analysis of comprehensibility*", *IEEE* Transactions on Fuzzy Systems, doi:10.1109/TFUZZ.2013.2245130, 2013.

[10]http://www.scholarpedia.org/article/Fuzzy_neural_network

[11] http://archive.ics.uci.edu/ml/datasets.html

