

Survey on the Family of the Recursive-Rule Extraction Algorithm

Yoichi Hayashi^{1,*}, Tomohiro Takagi¹, Hiroyuki Mori², Hiroaki Kikuchi³, Takamichi Saito¹, Hideaki Iiduka¹ and Sayaka Akioka²

¹Department of Computer Science Meiji University Tama-ku, Kawasaki 214-8571, Japan

²Department of Network Design Meiji University Nakano-ku, Tokyo 164-8525, Japan

³Department of Frontier Media Science Meiji University Nakano-ku, Tokyo 164-8525, Japan

Abstract: In this paper, we first review the theoretical and historical backgrounds on rule extraction from neural network ensembles. Because the structures of previous neural network ensembles were quite complicated, research on an efficient rule extraction algorithm from neural network ensembles has been sparse, even though a practical need exists for rule extraction in Big Data datasets. We describe the Recursive-Rule extraction (Re-RX) algorithm, which is an important step toward handling large datasets. Then we survey the family of the Recursive-Rule extraction algorithm, *i.e.* the Multiple-MLP Ensemble Re-RX algorithm, and present concrete applications in financial and medical domains that require extremely high accuracy for classification rules. Finally, we mention two promising ideas to considerably enhance the accuracy of the Multiple-MLP Ensemble Re-RX algorithm. We also discuss developments in the near future that will make the Multiple-MLP Ensemble Re-RX algorithm much more accurate, concise, and comprehensible rule extraction from mixed datasets.

Keywords: Ensemble Concepts, Rule Extraction, Re-RX Algorithm, Multiple-MLP Ensemble, Neural Network Rule Extraction, Neural Network Ensembles, Data Mining, Ensemble Learning.

1. INTRODUCTION

In this paper, we first survey the origin of neural network (NN) rule extraction, the incorporation of fuzziness in neural network rule extraction, the theoretical foundation of neural network rule extraction, previous rule extraction from neural network ensembles and the difficulties of previous neural network ensembles.

We also describe the Recursive-Rule extraction (Re-RX) algorithm and survey the family of this algorithm, *i.e.* the Multiple-MLP Ensemble Re-RX algorithm. Concrete applications in financial and medical domains show its usefulness. Furthermore, two promising ideas to considerably enhance the accuracy of the Multiple-MLP Ensemble Re-RX algorithm are presented. Finally, this paper discusses development of the Multiple-MLP Ensemble Re-RX algorithm for much more accurate, concise, and comprehensible rule extraction from mixed datasets. These developments will be offered in the near future.

1.1. Origin of Neural Network Rule Extraction

We first consider the layered connectionist model by Gallant [1] and Saito and Nakano [2] for rule extraction

in the medical domain. The inputs and outputs consist of crisp variables in all cases. Generally, the symptoms are represented by input nodes, and the diseases and possible treatments by intermediate and/or output nodes. The multilayer network described by Saito and Nakano was applied to the detection of *headache*. Headache patients respond to a questionnaire regarding the perceived symptoms and their responses constitute the input to the network.

In 1988, Gallant [1] created a model for handling *sacrophagal* problems. The model uses a linear discriminate network (with no hidden nodes) trained by the *simple pocket algorithm*.

Gallant's model incorporates inferencing and forward chaining, confidence estimation, backward chaining, and explanation of conclusions by IF-THEN rules. To generate a rule, the attributes with greater inference strength (magnitude of connection weights) are selected and a conjunction of the more significant premises is formed to justify the output concept.

1.2. Incorporating Fuzziness in Neural Network Rule Extraction

As an illustration of the characteristics of layered fuzzy neural networks for inferencing and rule generation, the models by Hayashi [3, 4] and Hudson *et al.* [5] are described first. A *distributed single-layer perceptron-based* model trained with the *pocket*

*Address correspondence to this author at the Department of Computer Science Meiji University Tama-ku, Kawasaki 214-8571, Japan;
Tel: +81-44-934-7475; Fax: +81-44-931-5161;
E-mail: hayashiyg@gmail.com

algorithm was used by Hayashi [3, 4] for diagnosing *hepatobiliary* disorders. All contradictory training data are excluded, because these cannot be handled by the model. The input layer consists of fuzzy and crisp cell groups, while the output is modeled only by fuzzy cell groups. The crisp cell groups are represented by m cells taking on two values, $\{(+1, +1, \dots, +1), (-1, -1, \dots, -1)\}$. Fuzzy cell groups, however, use binary m -dimensional vectors, each taking values of $\{+1, -1\}$. Linguistic terms of relative importance, such as very important and moderately important, are allowed in each proposition. Linguistic truth values, such as completely true, true, possibly true, unknown, possibly false, false, and completely false, are also assigned by the domain experts and depend on the output values. By using different linguistic truth values, a pattern belonging to more than one class can be modeled. Extraction of fuzzy IF-THEN production rules is possible by using a top-down traversal involving analysis of the node activation, bias, and the associated link weights.

1.3. Theoretical Foundation of Neural Network Rule Extraction

A fuzzy system adaptively infers and modifies its fuzzy association from representative numerical samples. Neural networks, in contrast, can *blindly* generate and refine fuzzy rules from training data. Fuzzy sets, considered advantageous in the field of logic, can easily handle higher-order processing. Higher flexibility is a characteristic feature of neural networks produced by learning and, hence, NNs are better suited for data-driven processing [6].

In 1993, Buckley, Hayashi, and Czogala [7] mathematically proved the equivalence of neural nets and fuzzy expert systems. In other words, they proved that we can describe the contents of trained neural networks by a set of linguistic IF-THEN rules. Moreover, this paper firmly established the theoretical foundation of neural network rule extraction.

Hayashi and Buckley [8] proved that 1) any rule-based fuzzy system can be approximated by a neural net, and 2) any neural net (e.g., feed forward net, multilayered net) can be approximated by a rule-based fuzzy system. This kind of equivalence between the fuzzy-rule-based system and neural networks has also been previously studied [7, 8, 9, 10].

1.4. Rule Extraction from Neural Network Ensemble

In the beginning of the 1990s, Hansen and Salamon [11] showed that the generalization ability of learning

systems based on artificial neural networks can be significantly improved through ensembles of artificial neural networks. That is, multiple artificial neural networks are trained and their predictions are combined via voting. Since combining works remarkably well, it became a very popular topic in both neural network and machine learning communities [12].

In general, a neural network ensemble is constructed in two steps: training a number of component neural networks, then combining the component predictions [13].

The rationale for considering a combination of methods is similar to that of ensemble NNs [14]. However, the ensemble is more robust and mitigates the effect of one method that gives bad results and ruins the performance [15].

Although many authors have generated comprehensible models from individual networks, much less work has been done in the explanation of neural network ensembles [16].

Bologna proposed the *Interpretable Multi-Layer Perceptron* (IMLP) and the *Discretized IMLP* (DIMLP) models with rules generated from neural network ensembles [17, 18]. The DIMLP is a special neural network model that generates symbolic rules to explain the knowledge embedded within the connections and the activation neurons. Bologna described how to translate symbolic rules into the DIMLP and how to extract rules from one or several combined neural networks. Rules are generated from a DIMLP network by the induction of a special decision tree and taking into account virtual hyper plane frontiers.

With the Rule Extraction From Neural network Ensemble (REFNE) algorithm proposed by Zhou *et al.* [12], attributes are discretized during rule extraction, whereas Bologna's rule extraction algorithm performs the discretization during learning through the use of staircase activation functions. Furthermore, rules generated by REFNE are limited to three antecedents, but DIMLP does not impose any constraints. Another important difference is that we extract unordered rules from DIMLP ensembles, whereas ordered rules are generated by REFNE. Bologna's rule extraction algorithm has no parameters; hence, it could be easier for a non-specialist in rule extraction to use the DIMLP ensemble rather than rule extraction techniques that require several parameters to be set [16].

The REFNE approach proposed by Zhou *et al.* is designed to extract symbolic rules from trained neural network ensembles that perform classification tasks. REFNE utilizes trained ensembles to generate a number of instances and then extracts rules from those instances. REFNE can gracefully break the ties in prediction made by individual neural networks [12].

Zhou *et al.* [13] analyzed the relationship between an ensemble and its component neural networks from the context of both regression and classification. Their work revealed that it may be better to ensemble *many* instead of *all* available neural networks at hand. This result is interesting because most approaches ensemble *all* available neural networks for prediction.

Adeodato *et al.* [19] showed that an ensemble of MLPs yields better results than does the single-MLP solution. For this purpose, the performance of the ensemble was compared to the average of the performances of each single MLP.

Bologna [16] pointed out that, although many authors have generated comprehensive models from individual networks, much less work has been done to explain ensembles of neural networks.

We carefully surveyed the previous work on rule extraction from neural network ensembles since 1988. The structure of previous neural network ensembles was quite complicated, although the learning capability was extremely high. Thus, the complicated structure meant that the rule extraction algorithm for a neural network ensemble was a difficult task. However, practical neural network ensembles are needed to realize the extremely high performance needs of various rule extraction problems in real life.

1.5. Difficulties of Previous Neural Network Ensembles

Because neural network ensembles are multiple individual neural networks, they present their own problems: their complexity is greater, rule extraction is more difficult, and they use more computing resources than should be necessary. In the neural network ensembles studied to date, research on methods such as weighted voting [20] and averaging [21] for integrating the output has been conducted. Algorithms that use bagging or boosting in the C4.5 algorithm have been presented, but these do not directly address the problems, because splitting a dataset into parts and applying the C4.5 algorithm to them generates rules that determine the class, and so the total output is classes that are broadly based on the rules. This in turn

means that the rules are numerous and redundant. Consequently, we believe that methods such as bagging or boosting cannot be assumed to adequately extract rules from a neural network ensemble.

Nevertheless, neural networks are known to be an effective method for real-world classification problems involving nonlinear data. Contrary to the standard explanation that neural networks operate as a “black box,” many studies have been conducted on the methods for rule extraction [22]. These studies can be seen as an outgrowth of the extraordinary advances that have been made in information technology (IT) and the ability of IT to easily handle Big Data. Extracting rules from neural networks is not simply a matter of breaking open the black box. From the perspective of data mining, rule extraction increases the opportunities to use neural network technology as a powerful data mining technology.

2. RECURSIVE RULE EXTRACTION ALGORITHM (RE-RX ALGORITHM)

The Re-RX algorithm [23] is designed to generate classification rules from datasets that have both discrete and continuous attributes. The algorithm is recursive in nature and generates hierarchical rules. The rule conditions for discrete attributes are disjointed from those for continuous attributes. The continuous attributes only appear in the conditions of the rules lowest in the hierarchy.

Here, we consider only two-group classification problems, although the algorithm that we are proposing can be easily extended to handle multiple groups. The outline of the algorithm is as follows.

Algorithm Re-RX(S, D, C)

Input: A set of data samples S having discrete attributes D and continuous attributes C.

Output: A set of classification rules.

1. Train and prune [24] an NN by using the dataset S and all of its D and C attributes.
2. Let D' and C' be the sets of discrete and continuous attributes, respectively, still present in the network, and let S' be the set of data samples correctly classified by the pruned network.
3. If $D' = \emptyset$, then generate a hyper plane to split the samples in S' according to the values of the

continuous attributes C' , and then stop. Otherwise, use only the discrete attributes D' to generate the set of classification rules R for dataset S' .

4. For each rule, R_i is generated:

If $\text{support}(R_i) > \delta_1$ and $\text{error}(R_i) > \delta_2$, then

- Let S_i be the set of data samples that satisfy the condition of rule R_i and let D_i be the set of discrete attributes that do not appear in rule condition R_i .
- If $D_i = \emptyset$, then generate a hyper plane to split the samples in S_i according to the values of their continuous attributes C_i , and then stop.
- Otherwise, call Re-RX (S_i, D_i, C_i).

The support of a rule is the percentage of samples that are covered by that rule. The support and the corresponding error rate of each rule are checked in step 4. If the error exceeds the threshold δ_2 and the support meets the maximum threshold δ_1 , then the subspace of this rule is further subdivided either by calling Re-RX recursively when discrete attributes are still not present in the rule conditions or by generating a separating hyper plane involving only the continuous attributes of the data. By handling discrete and continuous attributes separately, the Re-RX algorithm generates a set of classification rules that are more comprehensible than the rules having both types of attributes in their conditions.

3. MULTIPLE-MLP ENSEMBLE RE-RX ALGORITHM

In 2012, Hara and Hayashi [25, 26] proposed two kinds of Two-MLP ensembles by using the Recursive-Rule extraction (Re-RX) algorithm for data with mixed attributes: one kind of ensemble was for two-class classification and the other was for multiple-class classification. The accuracy of the proposed algorithm was far better than that of most previous algorithms.

In 2013, Hayashi *et al.* [27, 28, 29] presented the Three-MLP Ensemble Re-RX algorithm with an extremely high performance result for the German Credit dataset, which is a two-class mixed dataset.

The Re-RX algorithm [23] is an effective rule extraction algorithm for datasets that comprise both discrete and continuous attributes, and so it is a core part of the Three-MLP Ensemble Re-RX algorithm. The

Re-RX algorithm utilizes C4.5 [30] and back-propagation (BP) to train the MLPs recursively many times [23]. Thus, we believe that the Re-RX algorithm establishes a coexistence and a co-prosperity in both the neural network rule extraction community and the machine learning community.

Decision trees (also known as Classification Trees or hierarchical classifiers) started to play an important role in machine learning after the publication of Quinlan's Iterative Dichotomiser 3, known as ID3 [31]. Subsequently, Quinlan presented Classifier 4.5, known as C4.5, which is an advanced version of ID3. Since then, C4.5 has been considered to be the standard model in supervised classification.

Decision trees are models based on a recursive partitioning method that divides the dataset by using a single variable at each level. This variable is selected by a given criterion. Ideally, these models are sets of cases in which all the cases in a set belong to the same class [32].

The knowledge representation of decision trees has a simple structure. The structure can be interpreted as a compact rule set in which each node of the tree is labeled with an attribute variable that produces a different branch for each variable value (*i.e.*, a partition of the dataset). Leaf nodes are labeled with a class label.

The process for inferring a decision tree is mainly determined by the following: (i) the criteria used to select the attribute that should be placed in a node and branched; (ii) the criteria for stopping the tree branching process; (iii) the method for assigning a class label or a probability distribution to the leaf nodes; and (iv) the posterior pruning process used to simplify the tree structure [32].

Many rule extraction algorithms have been designed to generate classification rules from neural networks that have been trained to distinguish data samples from different classes. These algorithms frequently assume that the input data attributes are discrete in order to make the rule extraction process more manageable.

We consider the Multiple-MLP Ensemble to be a "virtual" ensemble system. The algorithm cascades BP to train the multiple neural-network ensemble. Thus, strictly speaking, the multiple NNs do not need to be trained simultaneously. In addition, this simple and new concept of rule extraction from the NN ensemble can

avoid previous complicated NN ensemble structures and rule extraction algorithms. The extracted rules maintain the high learning capabilities of NNs while expressing highly comprehensible rules.

We conducted a comprehensive performance comparison between the Three-MLP Ensemble Re-RX algorithm [33, 34] (shortened to “Three-MLP Ensemble”) and the Re-RX algorithm [23] and the variant[35] for eight kinds of real-life credit-risk two-class mixed datasets. In the Multiple-MLP Ensemble, each NN is an MLP. We also compared the performance of the Three-MLP Ensemble algorithm with various performances reported by recent classifier algorithms [32, 36, 37, 38, 39] on the same two-class mixed datasets used in the present experiments.

The Three-MLP Ensemble is shown as an example of the Multiple-MLP Re-RX algorithm [33, 34]. A schematic diagram of the Three-MLP Ensemble Re-RX algorithm is given as Figure. 1.

Algorithm Three-MLP Ensemble Re-RX (LD', LDf, LDff')

Inputs: Learning datasets LD', LDf, LDff'.

Outputs: Primary rule set, secondary rule set, and tertiary rule set.

1. Randomly extract a dataset of an arbitrary proportion from learning dataset LD, and name the set of extracted data as LD'.
2. Train and prune [24] the first NN by using LD'.
3. Apply the Re-RX algorithm to the output of step 2, and output the primary rule set.
4. Based on these primary rules, create a dataset LDf from a dataset (LD) that is not correctly classified by the rules.
5. Randomly extract the dataset of an arbitrary proportion from learning dataset LDf, and name the set of extracted data as LDf'.
6. Train and prune [24] the two ensemble NNs by using LDf'.
7. Apply the Re-RX algorithm to the output of step 6, and output the secondary rule set.
8. Integrate the primary rule set and the secondary rule set.
9. Based on the rules integrated in step 8, create a dataset LDff from a dataset (LD) that is not correctly classified by the rules.
10. Randomly extract data samples of an arbitrary proportion from learning dataset LDff, and name the set of extracted data as LDff'.
11. Train and prune [24] the ensemble of three NNs by using LDff'.
12. Apply the Re-RX algorithm to the output of step 11, and output the tertiary rule set.
13. Integrate the primary rule set, the secondary rule set, and the tertiary rule set.

4. STRATEGIC APPROACH FOR MULTIPLE-MLP ENSEMBLE RE-RX ALGORITHM

We first carefully reconsider the mechanism of the Re-RX algorithm [23] in the framework of the Multiple-MLP Ensemble Re-RX algorithm [33, 34]. The original target of the Re-RX algorithm [23] is the perfect or strict separation between discrete attribute rules and continuous attribute rules to achieve comprehensible rule extraction for mixed datasets. We believe that eliminating the continuous attributes C' to generate the set of classification rules [23] will reach the limit of accuracy. Furthermore, multiple use of the Re-RX algorithm is beyond the expectation of the authors in [23]. In our sufficient experience, we have learned that not only parameters such as δ_1 and δ_2 [23] but also the pruning rate in the Re-RX algorithm is quite sensitive to the accuracy of the Re-RX algorithm.

Paying attention to this fact and eliminating the continuous attributes C' , we proposed a strategic approach [40] for the Multiple-MLP Ensemble Re-RX algorithm. The idea of the strategic approach consists of two processes, non-pruning for trained ensemble neural networks without continuous attributes and relaxed rule generation with continuous attributes, to extract extremely accurate, comprehensible, and concise rules for mixed (*i.e.*, discrete and continuous attributes) datasets. We conducted experiments to find rules for four kinds of mixed datasets and compared the accuracy, comprehensibility, and conciseness against the Multiple-MLP Ensemble Re-RX algorithm. For Card1, Card2, Card3, and German datasets, the newly proposed strategic Multiple-MLP Ensemble Re-RX algorithm outperformed the previous Multiple-MLP Ensemble Re-RX algorithm.

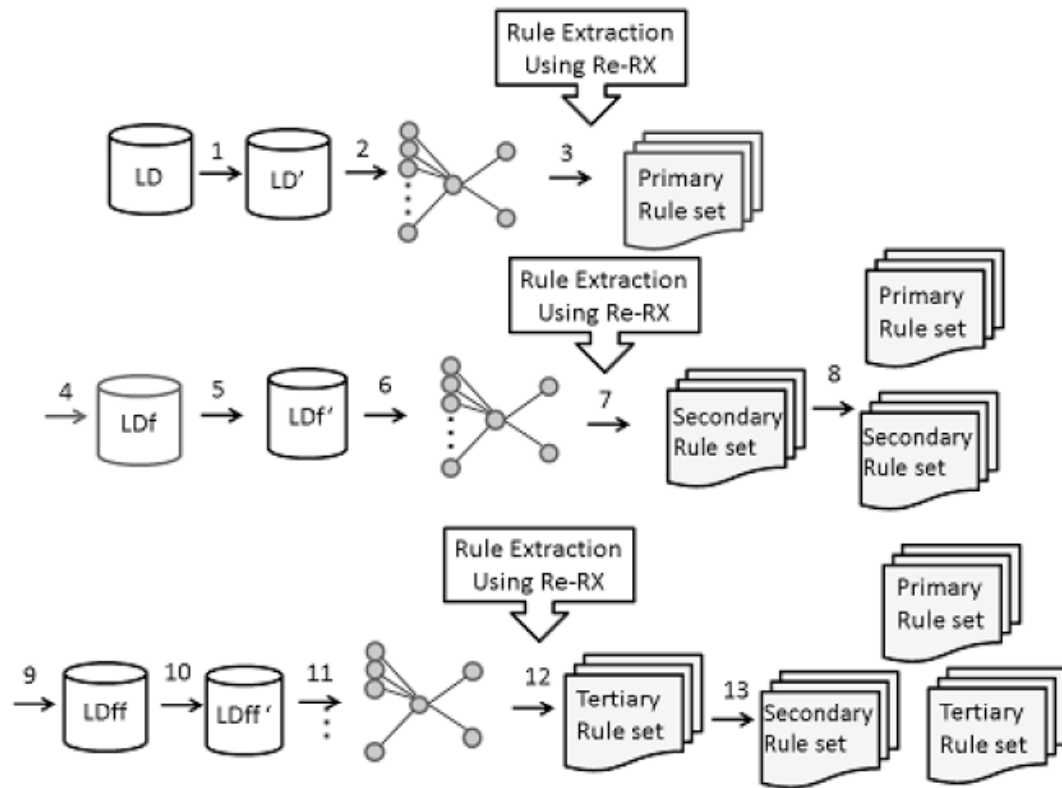


Figure 1: Schematic diagram of Three-MLP Ensemble by the Re-RX algorithm.

5. PERFORMANCE IMPROVEMENTS OF MULTIPLE MLP ENSEMBLE RE-RX ALGORITHM BY REPLACEMENT OF DECISION TREE UNIT

The concepts on conventional pruning used in J4.8 and grafting used in J48graft [41] are contrasting and complementary to each other. We believe that the performance of the Re-RX algorithm is much affected by the decision tree unit. After some consideration, we applied the grafting properties of J48graft in the Multiple MLP Ensemble Re-RX algorithm to improve the accuracies.

We proposed the replacement of the decision tree unit by J48graft in the Multiple-MLP Ensemble Re-RX algorithm ("Multiple-MLP Ensemble") [42] and conducted experiments to find rules for three kinds of mixed (*i.e.*, discrete and continuous attributes) datasets and compared the accuracy, comprehensibility and conciseness against the Multiple-MLP Ensemble Re-RX algorithm by a conventional decision tree unit, *i.e.*, J4.8. For the CARD1 and CARD3 datasets, the Multiple-MLP Ensemble by the J48graft much increased the accuracies of the Multiple-MLP Ensemble by J4.8.

6. APPLICATIONS OF MULTIPLE-MLP ENSEMBLE RE-RX ALGORITHM IN FINANCIAL AND MEDICAL DOMAINS

6.1. Application of Multiple-MLP Ensemble in Credit-Risk Evaluation [43]

Credit-risk evaluation is a challenging and important task in the domain of financial analysis. Many classification methods have been suggested in the literature to tackle this problem. We presented the results of an analysis of eight real-life credit-risk two-class mixed datasets (*i.e.*, discrete and continuous attributes) by using the Three-MLP Ensemble Re-RX algorithm (shortened to "Three-MLP Ensemble"). Clarifying the neural network decisions by explanatory rules that capture the learned knowledge embedded in the networks can help the credit-risk manager to explain why a particular applicant is classified as either bad or good. To compare the Three-MLP Ensemble performance, we executed comprehensive rule extraction experiments on real-life two-class mixed datasets commonly used for benchmarking studies in credit-risk evaluation. The extremely high accuracy of the Three-MLP Ensemble increased the accuracies by the Re-RX algorithm and the variant. In this paper, we

also compared the accuracy with that of the classifiers recently proposed. It is concluded that neural network rule extraction by the Three-MLP Ensemble is a powerful management tool that allows us to build advanced, comprehensible, and accurate decision-support systems for credit-risk evaluation.

6.2. Application of Multiple-MLP Ensemble in Medicine [44]

We compared the accuracies of rules extracted from multi-class mixed medical datasets by using the Three-MLP Ensemble Re-RX algorithm and the accuracies when using other rule extraction algorithms and classifiers. The Multiple MLP Ensemble Re-RX algorithm is able to handle mixed attribute datasets (*i.e.*, discrete and continuous) and uses a significant feature of the Re-RX algorithm proposed by Setiono. We used two medical datasets for rule extraction experiments: the Thyroid dataset and the Dermatology dataset. These two datasets consist of discrete attributes and continuous attributes. We explored the effectiveness of the Multiple MLP Ensemble Re-RX algorithm to extract extremely accurate rules for medical datasets. The extremely high accuracies in the Thyroid dataset and the Dermatology dataset obtained by the Multiple-MLP Ensemble Algorithm were superior to the accuracies obtained by previous rule extraction algorithms and classification methods. Thus, this paper opened the “black box” of trained neural network ensembles for medical datasets.

7. CONCLUSION

This paper first reviewed theoretical and historical backgrounds on rule extraction from neural network ensembles. Next, we described the Recursive-Rule extraction (Re-RX) algorithm. This paper surveyed the family of the Recursive-Rule extraction algorithm, *i.e.* the Multiple-MLP Ensemble Re-RX algorithm, and presented concrete applications in financial and medical domains that require extremely high accuracy for classification rules. Furthermore, this paper mentioned two promising ideas to considerably enhance the accuracy of the Multiple-MLP Ensemble Re-RX algorithm.

The original Re-RX algorithm proposed in [23] involved some human handling in the recursive process of the algorithm, and so the Re-RX algorithm has insufficient rule extraction speed for Big Data consisting of discrete and continuous attributes, *i.e.* huge mixed datasets.

The Re-RX algorithm requires many times of BP learning and pruning for MLP and generation of decision tree units such as C4.5. Thus, our important targets to improve the Multiple-MLP Ensemble Re-RX algorithm for practical use for Big Data analytics are high-speed, extremely accurate, concise, and comprehensible rule extraction.

We have been engaged in ongoing research for potential developments of the Multiple-MLP Ensemble Re-RX algorithm with dynamic and successive pruning rate changes [45] for more accurate, concise, and comprehensible rule extraction from large mixed datasets.

We already fully implemented the Multiple-MLP Ensemble Re-RX algorithm with dynamic and successive pruning rate changes [45] on a conventional laptop PC. In this case, the PC takes approximately 30 seconds to extract extremely accurate rules from commonly used big mixed datasets such as Bene2 with 7190 records and Thyroid with 7200 records. The extraction speed will be much faster than expected based on the expectations of the Multiple-MLP Ensemble Re-RX algorithm. On the extraction speed, our Multiple-MLP Ensemble Re-RX algorithm with dynamic and successive pruning rate changes will run 10 times faster, *i.e.*, less than a few seconds on a high-end commercial server.

It must be noted that a clear line can be drawn between high accuracies of the rule extraction using our Multiple-MLP Ensemble Re-RX algorithm with dynamic and successive pruning rate changes and the so-called recommendation systems and event prediction systems on some websites, which show ordinary accuracies.

ACKNOWLEDGEMENTS

We first would like to express our deepest appreciation to Prof. K. Fukumiya, Director General, Organization for the Strategic Coordination of Research and Intellectual Properties in Meiji University for the full support of our project: Developments of Extremely High Accuracy Rule Extraction Algorithm from Big Data and Wide Areas Applications. We also should express our deep appreciation to our dean, Prof. T. Arakawa, a head of the Strategic Coordination, Prof. K. Tsuchiya for their continuous encouragements.

REFERENCES

- [1] Gallant SI, “Connectionist expert systems,” *Commun. ACM*, 1988; 31: 152-169.

- [2] Saito K, Nakano R, "Medical diagnosis expert systems based on PDP model," in Proc. IEEE Int. Conf. Neural Netw, San Diego, CA1988; 1:255-1.262.
- [3] Hayashi Y, "Neural expert system using teaching fuzzy input and its application to medical diagnosis," Inform. Sci.: Applicat 1994; 1: 47-58.
- [4] Hayashi Y, "A neural expert system with automated extraction of fuzzy if-then rules and its application to medical diagnosis," in Advances in Neural Information Processing Systems, R. P. Lippmann, J. E. Moody, and D. S. Touretzky Eds. Los Altos, CA: Morgan Kaufmann 1991; 578-584.
- [5] Hudson DL, Cohen ME, Anderson MF. "Use of neural network techniques in a medical expert system," Int. J. Intell. Syst 1991; 6: 213-223.
- [6] Takagi H, "Fusion technology of fuzzy theory and neural network---Survey and future directions," in Proc. Int. Conf. Fuzzy Logic and Neural Networks, Iizuka, Japan, 1990; 13-26,
- [7] Buckley JJ, Hayashi Y, Czogala E. "On the equivalence of neural nets and fuzzy expert systems," Fuzzy Sets Syst 1993; 53(2): 129-134.
- [8] Hayashi Y, Buckley JJ. "Approximation between fuzzy expert systems and neural networks," Int. J. Approx. Res 1994; 10: 63-73
- [9] Buckley JJ, Hayashi Y. "Numerical relationship between neural networks, continuous function and fuzzy systems," Fuzzy Sets Syst 1993; 60(1): 1-8.
- [10] Buckley JJ, Hayashi Y. "Hybrid neural nets can be fuzzy controllers and fuzzy expert systems," Fuzzy Sets and Syst 1993; 60: 135-142.
- [11] Hansen LK, Salamon P. "Neural network ensembles," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.12, No.993-1001, 1990.
- [12] Zhou ZH, "Extracting symbolic rules from trained neural network ensembles," AI Communications, 2003; 16: 3-15,
- [13] Zhou ZH, "Ensemble neural networks: Many could be better than all," Artificial Intelligence 2002; 137: 239-263.
- [14] Yao X, Islam M, "Evolving artificial neural network ensembles," IEEE Computational Intelligence Magazine, 2008; 3(1): 31-42.
- [15] Khosravi A, *et al.*, "Comprehensive review of neural network-based prediction intervals and new advances," IEEE Trans. Neural Netw 2011; 22(9): 1341-1356.
- [16] Bologna G, "Is it worth generating rules from neural network ensembles?," J. of Applied Logic 2004; 2: 325-348.
- [17] Bologna G, "A study on rule extraction from several combined neural networks," Int. J. Neural Syst 2001; 11(3): 247-255.
- [18] Bologna G, "A model for single and multiple knowledge based networks, Artificial Intelligence in Medicine 2003; 28: 141-163.
- [19] Adeodato PJL, *et al.*, "MLP ensembles improve long term prediction accuracy over single networks," Int. J. Forecasting, 2011; 27: 661-671.
- [20] Alpaydin E, "Multiple Neural Networks and Weighted Voting," IEEE Trans. on Pattern Recognition, 1992; 2: 29-32.
- [21] Robert B, *et al.*, "Boosting of the margin: A new explanation for the effectiveness of voting methods," The Annals of Statistics, 1998; 26(5): 1651-1686.
- [22] Duch W, Setiono R, Zurada JM. "Computational intelligence methods for rule-based data understanding," Proceedings of the IEEE 2004; 92(5): 771-805.
- [23] Setiono R, Baesens B, Mues C. "Recursive neural network rule extraction for data with mixed attributes," IEEE Trans. Neural Netw 2008; 19: 299-307.
- [24] Setiono R, "A penalty-function approach for pruning feedforward neural networks," Neural Comp. 1997; 9(1): 185-204.
- [25] Hara A, Hayashi Y. "Ensemble neural network rule extraction using Re-RX algorithm," in Proc. of WCCI (IJCNN) 2012, June 10-15, Brisbane, Australia 2012; 604-609.
- [26] Hara A, Hayashi Y. "A new neural data analysis approach using ensemble neural network rule extraction" In Proc. 22nd Int. Conf. Artificial Neural Netw. (Lecture Notes in Computer Science, Vol.7552), Lausanne, Switzerland, September 2012; 515-522.
- [27] Hayashi Y, *et al.*, A New Approach to Three Ensemble Neural Network Rule Extraction Using Recursive-Rule eXtraction Algorithm, Inter. Joint Conf. Neural Networks (IJCNN 2013), Dallas, 2013; 835-841.
- [28] Hayashi Y, "Neural data analysis: Ensemble neural network rule extraction approach and its theoretical and historical backgrounds, Keynote Speech, In 12th Inter. Conf. Artificial Intelligence (ICAISC 2013) Zakopane, Poland, June 9-13, (Springer, LNAI Vol. 7894, Eds. L. Rutkowski *et al.*) 2013; 1-19.
- [29] Hayashi Y. "Neural network rule extraction by a new ensemble concept and its theoretical and historical background: A review, Inter. J. of Computational Intelligence and Applications 2013; 12(4): 1340006-1-1340006-22.
- [30] Quinlan JR, Programs for machine learning, Morgan Kaufmann Series in Machine Learning, San Mateo, CA, 1993.
- [31] Quinlan JR, "Induction of decision trees", Machine Learning, 1986; 1: 81-106,.
- [32] Abellan J, Mantas C. "Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring", Expert Systems with Applications 2014; 41: 3825-3830,.
- [33] Hayashi Y. *et al.*, "Comparative Study of Performances on the Family of the Recursive Rule Extraction Algorithm", Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), Hamburg, Germany, Sept. 15-19, 2014, S. Wermter *et al.* (Eds.): ICANN 2014, LNCS 8681 2014; 491-498.
- [34] Hayashi Y. *et al.*, "Characteristics and Potential Developments of Multiple-MLP Ensemble Re-RX Algorithm", Proceedings of the 21st International Conference on Neural Information Processing (ICONIP 2014), November 3-6, Kuching, Malaysia, L. C.Kiong *et al.* (Eds.), Lecture Note in Computer Science 2014; 8834: 619-627,.
- [35] Setiono R, *et al.*, "Rule extraction from minimal neural networks for credit card screening," Inter. J. of Neural Systems 2011; 21(4): 265-276.
- [36] Khashel M. "A bi-level neural -based fuzzy classification approach for credit scoring problems," Complexity 2013; 18(6): 46-57,
- [37] Brown I. Mues C. "An experimental comparison of classification algorithms for imbalanced credit scoring datasets," Expert Systems with Applications, 2012; 39: 3446-3453
- [38] Luukka P. "Nonlinear fuzzy robust PCA algorithms and similarity classifier," Expert Systems with Applications, 2010; 37: 8296-8302.
- [39] Tsai CF. "Combining cluster analysis with classifier ensembles to predict financial distress", Information Fusion, 2014; 4: 6-58.
- [40] Yoichi Hayashi and Shota Fujisawa, Strategic Approach for Multiple-MLP Ensemble Re-RX Algorithm, Proc. of International Joint Conference on Neural Network (IJCNN 2015), Killarney, Ireland, July 12-17, 2015 (submitted for presentation) and presented at Meiji-Czestochowa Joint Workshop on Big Data Engineering, Czestochowa, Poland 2014.
- [41] Webb GI. Decision Tree Grafting From the All-Tests-But-One Partition, Proceedings of the 16th International Joint conference on Artificial Intelligence (IJCAI), 1999; 2: 702-707.

- [42] Yoichi Hayashi, Yuki Tanaka, Performance Improvements of Multiple-MLP Ensemble Re-RX Algorithm by Replacement of Decision Tree Unit, Proc. of International Joint Conference on Neural Network, (IJCNN 2015), Killarney, Ireland, July 12-17, 2015 (submitted for presentation) and presented at Meiji-Czestochowa Joint Workshop on Big Data Engineering, Czestochowa, Poland 2014.
- [43] Yoichi Hayashi, Using Three-MLP Ensemble Re-RX algorithm and Recent Classifiers for Credit-Risk Evaluation, Proc. of International Joint Conference on Neural Network, (IJCNN 2015), Killarney, Ireland, July 12-17, 2015 (submitted for presentation) and presented at Meiji-Czestochowa Joint Workshop on Big Data Engineering, Czestochowa, Poland, October 7-8, 2014, European Journal Operational Research (submitted for publication)
- [44] Yoichi Hayashi and Tomohiro Hiraku, Application of Multiple-MLP Ensemble Re-RX Algorithm to Multi-Class Mixed Datasets in Medicine, Proc. of International Joint Conference on Neural Network, (IJCNN 2015), Killarney, Ireland, July 12-17, 2015 (submitted for presentation) and presented at Meiji-Czestochowa Joint Workshop on Big Data Engineering, Czestochowa, Poland, October 7-8, 2014, Artificial Intelligence in Medicine (submitted for publication).
- [45] Yoichi Hayashi, Satoshi Nakano and Ryutaro Ono, Multiple-MLP Ensemble Re-RX Algorithm with Dynamic and Successive Pruning Rate Changes, the 22nd International Conference on Neural Information Processing (ICONIP 2015), Istanbul, Turkey, November 9-12, 2015 (submitted for presentation)

Received on 8-12-2014

Accepted on 18-12-2014

Published on 12-1-2015

<http://dx.doi.org/10.15379/2410-2938.2014.01.01.04>

© 2014 Hayashi *et al.*; Licensee Cosmos Scholars Publishing House.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.