Logic Mining in Football Matches Using k-Satisfiability Based Reverse Analysis Method

Kho Liew Ching[1], Mohd Shareduwan Mohd Kasihmuddin[2*], Mohd. Asyraf Mansor[3], Aslina Baharum[4] and Saratha Sathasivam[5]

[1,2,5]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia*
[3]*School of Distance Education, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia*
[4]*Faculty of Computing and Informatics, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

This finding presents a logic mining technique to model the results (Win Draw / Lose) of the football matches played in English Premier League. In this research, a method namely *k* Satisfiability based reverse analysis method (*k*SATRA) is brought forward to obtain the logical relationship among the clubs in this league. The logical rule obtained from the football matches is used to categorize the results of future matches. *k*SATRA makes use of the advantages of Hopfield Neural Network and k Satisfiability representation. The dataset used in this study includes the data of 6 clubs in the league, which are composed of all league matches from 2014 to 2018. The effectiveness of *k*SATRA in obtaining logical rule in football matches is tested based on root mean square error (RMSE), mean absolute error (MAE) and CPU time. Results acquired from the computer simulation shows the robustness of *k*SATRA in exhibiting the performance of the clubs.

Keywords: 2 Satisfiability, Logic Mining, Hopfield Neural Network, 2 Satisfiability Reverse Analysis Method

I. INTRODUCTION

Football is a popular sport where two teams consists of eleven players each compete in. Football analytics is unable to garner much attention in academic literature because the match statistics that are available to public are very limited (Haaren & Van den, 2015). Nevertheless, there are still a few models that were proposed to predict the results of a football match. Tsakonas et al. (2002) proposed soft computing methods to predict the result of a football match in regard to neural networks, fuzzy rules and genetic programming approach. On the other hand, Baio and Blangiardo (2010) proposed a Bayesian hierarchical model to predict the final score of a football match.

Artificial neural network (ANN) learns from the biological nervous system of human beings, for example how information is processed by the brain (Rojas,1996).

Hopfield Neural Network (HNN) is one of the well-known networks implemented to solve various optimization problems (Hopfield & Tank, 1985). HNN shows outstanding learning behaviour. For example, productive learning and retrieval operation. Traditional HNN is susceptible to a few deficiencies (Gee et al., 1993) so logic programming is embedded to HNN as a single intelligent unit (Abdullah,1992). Logic mining in HNN was proposed by Sathasivam (2016) by applying Reverse Analysis method. This method can obtain the logical rule among neurons. 2 Satisfiability (2SAT) was discovered to enhance the representation of general SAT itself (Maknickas,2015). This makes 2SAT a suitable approach to represent logical rules in neural network. By considering only 2 literals per clause, the logical complexity in learning the relationship between the variables in real life problem decreases. By hybridizing Reverse Analysis and 2SAT, a new method, 2 Satisfiability Reversed Analysis method (2SATRA) will be

*Corresponding author: shareduwan@usm.my

utilized to obtain the logical rule of football matches.

II. 2SAT REPRESENTATION

2 Satisfiability (2SAT) is a logical rule which comprises of only 2 literals per clause. 2SAT is usually expressed as Boolean formulas called Conjunctive Normal Form (CNF) or Krom formulas. 2SAT consists of three components (Kasihmuddin,2017):

1. A set of x variables, $v_1, v_2, ..., v_x$

2. A set of literals. A literal can be any variable or a negation of any variable.

3. A set of y definite clauses, $C_1, C_2, C_3, ..., C_y$ linked by logical AND ($\wedge$). Each clause comprises of strictly 2 literals joined by just logical OR ($\vee$).

Each of the variable can only take bipolar value of 1 or -1 which represents true or false respectively. Explicit definition of the 2SAT formula P_{2SAT} is given by

$$P_{2SAT} = \bigwedge_{i=1}^{y} C_i \qquad (1)$$

where C_i is a list of clause with 2 variables each,

$$C_i = \bigvee_{i=1}^{y} (m_i, n_i) \qquad (2)$$

The primary aim of 2SAT representation is to discover the consistent interpretation that makes formula becomes satisfied (Kasihmuddin et al., 2017).

III. HOPFIELD NEURAL NETWORK (HNN)

HNN is one of the most used neural network models. It is a simple neural network model that has feedback connections. HNN systematically stores patterns as a content addressable memory (CAM) (Muezzinoglu et al., 2003). HNN is a network of N interconnected neurons where the output and input of each neuron is connected. The connection weight from neuron i to j is denoted by w_{ij}. In HNN, $w_{ij} = w_{ji}$, (symmetric networks), and $w_{ii} = w_{jj} = 0$ (no self-feedback connections). Let S_i be the state or output of the i th unit, q_i is the pre-defined threshold of unit i. For bipolar networks, S_i is either +1 or -1. General updating rule in HNN is given by:

$$S_i = \begin{cases} 1 & if \sum_j w_{ij} S_j > \theta_i \\ -1 & Otherwise \end{cases} \qquad (3)$$

The local field of the network is as follows:

$$h_i(t) = \sum_j w_{ij}^{(2)} S_j + w_i^{(1)} \qquad (4)$$

The updating rule will be:

$$S_i(t) = sgn[h_i(t)] \qquad (5)$$

The final state of neurons will be examined by using Lyapunov or energy function:

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij}^{(2)} S_i S_j - \sum_i w_i^{(1)} S_i \qquad (6)$$

The final energy of HNN is always decreasing with the dynamics. Energy value acquired from equation (6) will determine whether it is a local or global minimum energy. 2SAT in HNN is abbreviated as HNN-2SAT model.

IV. 2 SATISFIABILITY BASED REVERSE ANALYSIS METHOD (2SATRA)

Logic mining will execute efficiently if the most favourable HNN-2SAT model is used. The neurons (attributes) are represented in bipolar form {-1,1}. By acquiring the synaptic weight between 2 neurons, 2SATRA might be able to reveal the level of their connectedness. Therefore, Wan Abdullah's method (Abdullah,1992) is utilized in the learning phase of 2SATRA to find out the accurate synaptic weight between the two neurons. By considering both neurons C and D where $S_C \in \{-1,1\}$ and $S_D \in \{-1,1\}$, the possible 2SAT clause with its corresponding synaptic weight are summarized in Table 1.

Table 1. Possible 2SAT logic with its corresponding synaptic weight

Synaptic Weight	$P_1 = C \vee D$	$P_2 = \neg C \vee D$	$P_3 = C \vee \neg D$	$P_4 = \neg C \vee \neg D$
W_C	0.25	-0.25	0.25	-0.25
W_D	0.25	0.25	0.25	-0.25
W_{CD}	-0.25	0.25	-0.25	-0.25

As an example, given that neuron C and D shows 1 and -1 respectively, P_3 will be selected as the clause representation of the data set. In accordance with the nature of the neuron, 2SATRA will convert all the data sets into 2SAT logic. Algorithm below shows the implementation of 2SATRA.

Algorithm of Implementation of 2SATRA:

Step 1: Given binary learning and testing data set with outcome P_{learn} and P_{test}, convert all binary data set to bipolar representation where 0 denotes as -1 and 1 remains 1.

Step 2: Initialize synaptic weight of the neurons and assign all neurons with bipolar data obtained from Step 1.

Step 3: Segregate the collection of two neurons per clauses $C_1, C_2, C_3, ..., C_n$ that lead to $P_{learn} = 1$.

Step 4: Obtain P_{test} by comparing the frequency of the 2SAT clauses in the overall learning data set.

Step 5: Check clauses satisfaction of P_{best}.

Step 6: Derive the synaptic weight of P_{best} by using WA method in Table 1 and randomize the state of the neurons.

Step 7: Apply Sathasivam relaxation method (Sathasivam, 2010) to the network.

Step 8: Find the final state of neuron by computing the corresponding local field by using equation (4).

Step 9: Induce all possible 2SAT logic $P_1^B, P_2^B, P_3^B, ..., P_n^B$ from the neuron states.

Step 10: Examine all induced logic P_i^B by comparing the outcome of P_i^B with P_{best}.

Step 11: Calculate the root mean square error (RMSE), mean absolute error (MAE) and CPU time.

In learning data set, {Win/Draw, Lose} will be converted into bipolar representation {1,-1} respectively. Each football club will be represented in terms of neuron in 2SATRA. The respective football club and neuron is summarized in Table 2.

Table 2. Respective Football Club and Neuron

Neuron	Football Club
A	Arsenal (*Ar*)
B	Chelsea (*Ch*)
C	Liverpool (*Li*)
D	Man. City (*Mc*)
E	Man. United (*Mu*)
F	Tottenham Hotspurs (*Sp*)

In this research, England Premier League's data for the 2014-2018 seasons were used. The simulation of 2SATRA for the data set was executed using DEV C++ on Windows 8.1. The data set is divided into 60% for learning phase usage and the rest of 40% for testing phase. During retrieval phase, the HNN-2SAT model will undergo 100 trials with 100 combinations so that the alteration of the neuron states will increase (Sathasivam & Fen, 2013). The threshold CPU time for program execution is 24 hours.

V. RESULTS AND DISCUSSIONS

A total of 3 performance evaluation namely root mean square error (RMSE), mean absolute error (MAE) and CPU time were analysed to determine the accuracy and steadiness of HNN-2SAT in doing 2SATRA. NC is defined as the total number of clause and 1 clause has 2 neurons (attributes).
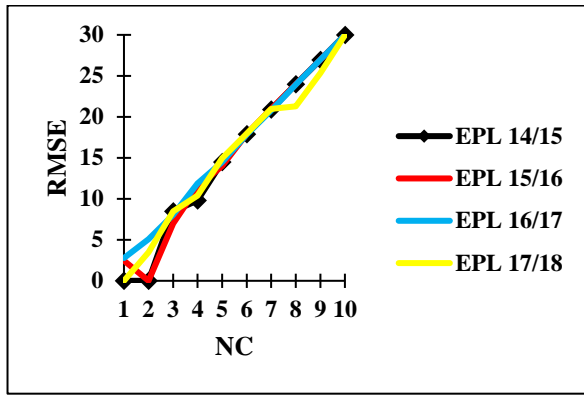
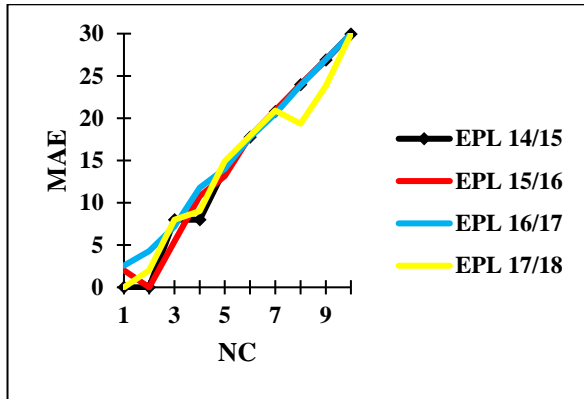Figure 2. RMSE for 4 EPL Seasons
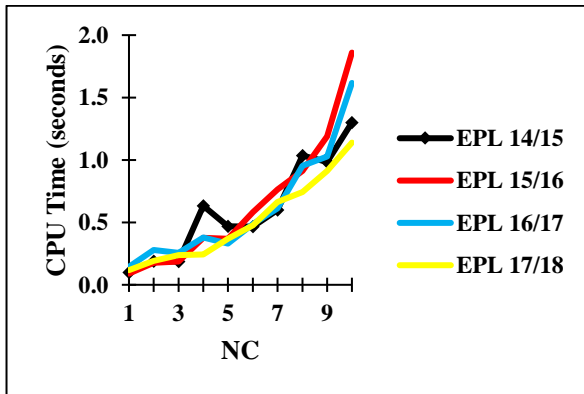

Figure 3. MAE for 4 EPL Seasons

Figure 4. CPU Time for 4 EPL Seasons

Figure 2 and Figure 3 showed the accuracy of HNN-2SAT in doing 2SATRA. Based on Figure 2 and Figure 3, it can be observed that at NC=1, the HNN-2SAT model had the most excellent results in terms of RMSE and MAE. The reason behind this was when the number of clause got larger, learning phase of 2SATRA got more complicated as HNN-2SAT had to discover the consistent interpretation for P_{best}. Conjointly, the learning error for 2SATRA increased as the number of neurons increased. 2SATRA achieved maximum value of RMSE and MAE when NC=10. In this case, the learning phase of 2SATRA reached a trial and error state and lead to RMSE and MAE accumulation. This phenomenon had a good agreement with the study by Sathasivam [13]. Figure 4 showed the steadiness of 2SATRA in logic mining. 2SATRA was capable of inducing optimal P_i^B in short CPU time. At NC=2, 2SATRA successfully induced optimal P_i^B in only 0.2 seconds. Sathasivam relaxation method [13] is able to lessen neuron oscillation that would prolong the CPU time and hence achieve sub-optimal P_i^B. During learning phase, P_i^B induced by 2SATRA managed to accomplish an accuracy of 63% (EPL 14/15) and 72% (EPL 15/16, EPL 16/17, EPL 17/18). This is because of the character of neuron in HNN, rather than oscillating, the neurons have always converged to minimum energy. The best induced logic, P_{best} and inconsistent interpretation, $P_{inconsistent}$ for each EPL seasons are summarized in Table 3.

Table 3. Best Induced Logic and Inconsistent Interpretation

EPL Season	Best induced logic, P_{best}	Inconsistent interpretation, $P_{inconsistent}$
14/15	$(Ar \vee Ch) \wedge (Li \vee Mc) \wedge (Mu \vee Sp)$	$(\neg Ar \wedge \neg Ch) \vee (\neg Li \wedge \neg Mc) \vee (\neg Mu \wedge \neg Sp)$
15/16	$(Ar \vee Ch) \wedge (Li \vee Mc) \wedge (Mu \vee Sp)$	$(\neg Ar \wedge \neg Ch) \vee (\neg Li \wedge \neg Mc) \vee (\neg Mu \wedge \neg Sp)$
16/17	$(Ar \vee Ch) \wedge (Li \vee Mc) \wedge (\neg Mu \vee Sp)$	$(\neg Ar \wedge \neg Ch) \vee (\neg Li \wedge \neg Mc) \vee (Mu \wedge \neg Sp)$
17/18	$(Ar \vee Ch) \wedge (Li \vee Mc) \wedge (\neg Mu \vee Sp)$	$(\neg Ar \wedge \neg Ch) \vee (\neg Li \wedge \neg Mc) \vee (Mu \wedge \neg Sp)$

According to Table 3, the relationship among the football clubs is shown. During EPL Season 14/15, in any given match week, if Arsenal, Liverpool and Manchester United lost their matches, the rest of the clubs such as Chelsea, Manchester City and Tottenham Hotspurs will have more player options during that week. With that advantage, club

such as Chelsea has the privilege to send their second-best team. The implication of the logical rule gives more training time to Chelsea's first team and focus on more important matches. This will reduce the number of injuries faced by the club. The results have shown that 2SATRA has decent potential to obtain logical rule that classifies the results of win/draw or lose for a football match.

VI.　SUMMARY

In this research, 2SATRA is shown to be a decent prediction system to model the results of football matches. The effectiveness of 2SATRA in doing logic mining is examined by using 4 EPL data sets. The results acquired shows that 2SATRA has decent potential to obtain optimal logic from learned data set. Future research could be done by integrating some metaheuristic algorithm to accelerate the process of learning phase of 2SATRA.

VII.　ACKNOWLEDGEMENT

VII. REFERENCES

Abdullah, W. A. T. W. (1992). Logic programming on a neural network. International journal of intelligent systems, 7(6), 513-519.

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37(2), 253-264.

Gee, A. H., Aiyer, S. V., & Prager, R. W. (1993). An analytical framework for optimizing neural networks. Neural Networks, 6(1), 79-97.

Haaren, V. J., & Van den Broeck, G. (2015). Relational learning for football-related predictions. In Latest Advances in Inductive Logic Programming, 237-244.

Hopfield, J. J., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. Biological cybernetics, 52(3), 141-152.

Kasihmuddin, M. S. M. (2017). Satisfiability logic programming incorporating metaheuristics in Hopfield neural networks. PhD Thesis. University of Science, Malaysia.

Kasihmuddin, M. S. M., Mansor, M., & Sathasivam, S. (2017). Robust Artificial Bee Colony in the Hopfield Network for 2-Satisfiability Problem. Pertanika Journal of Science & Technology, 25(2), 453-468.

Maknickas, A. A. (2015). How to solve kSAT in polynomial time. International Journal of Operational Research, 23(3), 257-267.

Muezzinoglu, M. K., Guzelis, C., & Zurada, J. M. (2003). A new design method for the complex-valued multistate Hopfield associative memory. IEEE Transactions on Neural Networks, 14(4), 891-899.

Rojas, R. (1996). Neural Networks: A Systematic Introduction. Berlin: Springer.

Sathasivam, S. (2006). Logic Mining in Neural Networks. PhD Thesis. University of Malaya, Malaysia.

Sathasivam, S. (2010). Upgrading logic programming in Hopfield network. Sains Malaysiana, 39(1), 115-118.

Sathasivam, S., & Fen, N. P. (2013). Developing agent-based modeling for doing logic programming in hopfield network. Applied Mathematical Sciences, 7(1), 23-35.

Tsakonas, A., Dounias, G., Shtovba, S., &Vivdyuk, V. (2002). Soft computing-based result prediction of football games. In The First International Conference on Inductive Modelling. Lviv, Ukraine.