Adopting Predictive Analytics Model for Mitigating Network Evaluation Delay in Best Connection Selection Mechanism

Muhammad Idham Abdul Halim[1], Wahidah Hashim[1], Ahmad Fadzil Ismail[2], Kok-Lim Alvin Yau[3], Mohammad Kamrul Hasan[2], Rajina M. A. Raj Mohamed[1]

*[1] College of Computer Science & Information Technology,
Universiti Tenaga Nasional, Selangor, Malaysia*
[2]Department of Electrical and Computer Engineering,

International Islamic University Malaysia, Selangor, Malaysia

[3]Department of Computing and Information Systems

Sunway University, Selangor, Malaysia
iddy93@gmail.com, wahidah@uniten.edu.my, af_ismail@iium.edu.my,

Abstract

Predicting for best network connection with highest accuracy of prediction deems to be a desire for anticipating most reliable network communication. Besides the ordinary function of any predictive mechanism to understand certain behaviour and trends, we have adopted such mechanism to mitigate network evaluation delays that occur every time switching to other network. In this paper, we propose a predictive analytics mechanism based on existing network selection models. The mechanism uses connection speed as main parameter to determine the network behaviour before a selection being made. We conducted a few predictive models that are able to predict the network speed based on the input data that we have recorded daily at an interval of 15 minutes. The root mean squared error (RMSE) and squared error (SE) values were calculated for each predictive model used. It was found that, the W-M5Rules and W-M5P have the lowest error difference amongst other tested predictive models. For our sample size dataset, it is appropriate to say that the W-M5Rules and W-M5P show the best accuracy that resembles the actual collected data.

Keywords: Predictive analytics, W-M5Rules, network speed, best connectivity access, mobile data, RapidMiner

1. Introduction

The abundance of wireless infrastructure has made it possible for the presence of multiple networks surrounding wireless user device that wishes to associate itself with any of these networks. However, due to protocol limitation of certain multiple backhaul communication devices, the association and at the same time evaluating other potential networks fail to be carried out in parallel. More intelligent system may have introduced three advanced network association steps. Firstly, the device measures the received signal strength indicator (RSSI) of the networks. Secondly, it ranks the networks based on RSSI. Thirdly, it associates itself with the network that provides the highest RSSI. The RSSI, which fluctuates in nature, changes with various factors, particularly the distance between the device with the network base station (BS). Figure 1 shows laboratory setup in our study on how network connection speed for two subscribed networks can be monitored. Two devices are configured to measure the network speed also known as throughput, which can be either the download or upload speed of the wireless MiFi. Further reading on previous works by the authors can be obtained from [1].

Figure 1. Laboratory Setup of Network Speed Data Collection for Two Subscribed Networks

This paper enhances the existing network selection mechanism in which a wireless user device must disconnect from the current operating network, which provides poor network performance, and reconnect to another available network, for best possible throughput value. This process can be tedious and cumbersome. The main challenge is that, the user device does not have any insights on the available networks (e.g., whether the available network has better or poorer connectivity). While conventional network selection mechanisms have been proposed in [2] and [3], an evaluative model with elements of cognitive selection mechanism was proposed in [1], and it has shown to be a better alternative, taking into account the connection speed values in contrast to the usual RSSI. In [4], a successive selection model was proposed as an improvement in [1]. While the successive and evaluative selection models have shown to improve network performance, both models have a lengthy network evaluation delay compared to the random selection method. This is because the successive and evaluative selection models scan for the speed of each network, and then rank them accordingly. This process is performed whenever a user device disconnects and reconnects to networks. This means that the evaluative delay increases with the number of available networks for the user device to choose.

Consequently, it contradicts with the advantage brought by automatic network selection mechanism, which aims to reduce the time incurred by the user device to select for best available networks. This paper embeds predictive analytics into a predictive model to reduce network evaluative delay. In general, the predictive model forecasts the download speed and eliminates the need to do speed test for each of the network during rescanning process. Various predictive models are presented, and the performance criteria applied are RMSE and SE to find the most suitable predictive model for a specific dataset under studies.

The rest of the paper is organized as follows. Section 2 describes, in brief, related predictive models that are suitable for network selection. Section 3 covers parameters used to measure the accuracy of the predictive models. Section 4 discusses our experimental set-up for adopting predictive analytics techniques. Experimental results and discussion are eloborated in Section 5. Finally, concluding remarks are highlighted in Section 6.

2. Predictive Analytics Models

An extensive amount of research has been conducted in predictive analytics. As a result, a variety of classifiers have been introduced. This section explains the basics of predictive algorithms adopted in this study. There are summarized in Table 1.

Table 1. Summary of Predictive Model under Study

Model	Description	Technique	Strength
M5 [5],[6]	2 sub categories: W-M5Rule & W-M5P	• Classify data into catergories for its decision tree based on certain rules for W-M5Rule • W-M5P generates a decision list for regression problems using separate-and-conquer techniques	Eliminates any possible data without any categories and rules [7], [8]
Conjunctive rule	Implement a single conjunctive rule learner that are capable of predicting nominal and numeric class labels [9]	• A rule usually are made up of antecedents "AND"ed • The value intended for prediction for the classification or estimate relationship calculation (regression).	Calculate the information gain to make the data more meaningful and accurate [10]
Gradient boosted	An ensemble of either regression or classification tree models	• Boosting is a flexible nonlinear regression procedure that helps improving the accuracy of trees • Calculation speed takes longer and human interpretability decreases. The gradient boosting method generalizes tree boosting to reduce these issues [12]	Consistently improving its estimations to garner appropriate predictive results
k-Nearest Neighbor (k-NN)	Based on learning by examples.	• The given test example was compared with training examples that are homogeneous to it • Each example is represented by a point in an n-dimensional space. • Amongst the simplest	When given an unfamiliar example, a k-nearest neighbor algorithm searches the pattern space for the k training examples that are closest to the unfamiliar example [11].
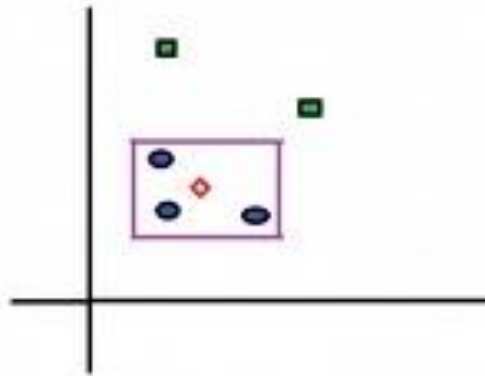
Figure 2 . k-NN Algorithm Theoretical Representation [13]

Figure. 2 depicts a distirbution of blue circles (BC) and green squares (GS) and a single red diamond (RD). For this example, RD, the unknown instance has to be either a member of BC or GS. To classify the class of the RD, the k-NN algorithm will associate RD with the closest instances circle which is apparently three BC. The three closest points to RD is all BC. Thus, it can be deduced that RD belongs to the class BC [13].

3. Measurement Parameters for the Selected Predictive Models

We investigate a number of predictive models on our collected dataset. These models were chosen for its ability to predict numerical labels and identify polynominal attributes, as the main objective of this predictive model is to anticipate network speed, which is a numeric value. The predictive models adopted in this study are listed as W-M5Rules, W-ConjuctiveRule, W-M5P, Gradient Boosted Trees, k-NN, W-IBk, W-Linear Regression

The performance criteria that we analyzed were both SE and RMSE. For each of the predictive models both of these performance criteria were calculated and compared to, respectively. The description for each of these parameters are given below

- **Square Error (SE)**

$$= \frac{|x_1 - y_1| + |x_2 - y_2| + \cdots + |x_i - y_i|}{i}$$

Where x_i is the actual network speed and y_i is the expected network speed and i is the number of network speed data collected.

- **Root Mean Squared Error (RMSE)**

$$= \sqrt{\frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_i - y_i)^2}{i}}$$

Where x is the actual network speed and y is the expected network speed and i is the number of collected network speed data.

RMSE is used very commonly to evaluate the accuracy of predictive models. This is due to the fact that it has some useful mathematical properties whereby the value is calculated using the difference in the actual network speed and the predicted network

speed. Then the difference was squared to remove any negative value from the operation. Simply, the RMSE is just the square root of the mean square error.

4. Experimental Setup

Parameters	Description
Spectrum band	Wi-Fi – 2.4 GHz
Modem Capacity	4G capable - up to 150 Mbps
Network Capacity	3G,4G – up to 75 Mbps
Operating System (OS)	Ubuntu 16.04.1 LTS
Location	Indoor Building
Campus	UNITEN, Putrajaya
Time interval for rescanning	15 minutes
Network tested	Wi-Fi
Average download file size	1 MB

Host	Telco A	Telco B
Network Capacity	Up to 75 Mbps	Up to 75 Mbps
Network Availability in Vicinity	3G & 4G	3G & 4G
Protocol	802.11n	802.11n
Network band	2.4 GHz	2.4 GHz
Network Channel	6	6
Capability of Modem	LTE-FDD: DL/UL 100/50Mbps; LTE-TDD:DL/UL 68/17Mbps	LTE-FDD: DL/UL 150Mbps/50Mbps; LTE-TDD: DL/UL 110Mbps/11Mbps

Figure 3 . System and Experimental Setup Parameters

A user connected to multiple networks was set up. The experiment design and specifications, which are shown in Figure 3 and Figure 1 consists of common peripherals that are easily accessible and available, including a Windows powered laptop, a broadband dongle, a valid data plan and a specific software for measuring the speed of the network. In general, we switched on the broadband dongle and wait for it to receive data, and then we measured the network speed using specified software running on a Windows powered laptop. The connection speed was recorded every 15 minutes. The interval of 15 minutes is considered as this is the interval for data transfer from smart meter to data collector.

ExampleSet (1717 examples, 1 special attribute, 5 regular attributes)

Row No.	Mbps Speed	Date	Type of Link	Test Size	MB's Speed	Host
1	0.910	Sep 21, 2016...	UP	768 kB	0.110	uniten
2	0.370	Sep 21, 2016...	DN	1.3 MB	0.050	uniten
3	1.040	Sep 21, 2016...	UP	936 kB	0.130	uniten
4	0.680	Sep 21, 2016...	DN	1.3 MB	0.080	uniten
5	0.370	Sep 21, 2016...	UP	256 kB	0.050	uniten
6	0.500	Sep 21, 2016...	DN	1.3 MB	0.060	uniten
7	0.910	Sep 21, 2016...	UP	608 kB	0.110	uniten

**Figure 4. Dataset of Network Speed Collected**

In Figure 4, the dataset that we have collected is partly presented. The Mbps speed is the numerical label that we aim to predict. The full log of the dataset consists of 1717 rows of data. The predictive model was implemented using open access tools known as the Rapidminer platform.

5. Analysis of Results

Table 2. Performance Criteria of Each Predictive Model

Model	Squared Error (SE)	Root Mean Squared Error (RMSE)
W-M5Rules	0.001	0.024
W-ConjunctiveRule	50.612	7.062
W-M5P	0.001	0.024
W-IBk	5.309	2.271
Gradient Boosted Trees	1.96	1.366
k-NN	136.219	11.615
W-Linear Regression	0.001	0.026

Table 2 shows the performance criteria of each predictive model. The squared error and root mean squared error for each model are presented. It is evident that the model with the lowest SE and RMSE are the M5-based W-M5Rules and W-M5P approaches as shown in Figure 5. Coming close is the W-Linear Regression model which only differs in RMSE. Another viable model for this dataset is the Gradient Boosted Trees. The other models do not seem suitable for this dataset. In this experiment, the M5 algorithm has shown to be superior compared to other algorithms, and further investigation is needed to investigate using other datasets.
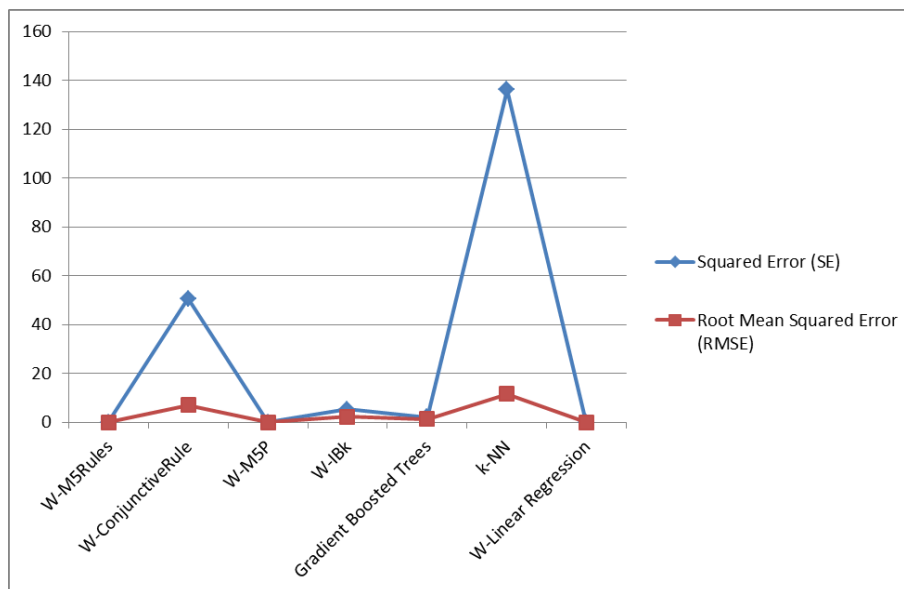
Figure 5. Comparing Performance Criteria of Various Predictive Models

6. Conclusion

In this paper, we investigate different predictive models for predicting the value of network speed and compare their results. The W-M5Rules and W-M5P algorithms have shown to achieve the lowest RMSE and SE amongst the other predictive models. Using our dataset, W-M5Rules and W-M5P have achieved the best result that is closest to the original value. For our future works we will include investigations using different datasets to test on its accuracy performance. With this promising results, using W-M5Rules and

W-M5P techniques are found to be the most suitable predictive models technique for anticipating network speed.

Acknowledgements

References

[1]   W. Hashim, A. F. Ismail, S. Dzulkifly and N. A. A. Ghafar, Cognitive Selection Mechanism for Indoor Propagation, vol. 2, no. 4, **(2013)**.
[2]   Y. Fukuda and Y. Oie, "Decentralized access point selection architecture for wireless LANs", IEICE Trans. Commun., E90–B, 9, **(2007)**.
[3]   F. Xu, X. Zhu, C. C. Tan, Q. Li, G. Yan and J. Wu, SmartAssoc: Decentralized access point selection algorithm to improve throughput, IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 12, **(2013)**.
[4]   W. Hashim, A. F. Ismail, N. A. Abd Ghafar and S. Dzulkifly, "Cognitive Selection Mechanism Performance in IEEE 802.11 WLAN", International Journal of Computer and Communication Engineering, vol. 2, no. 4, **(2013)**.
[5]   R. J. Quinlan, "Learning with Continuous Classes", In: 5th Australian Joint Conference on Artificial Intelligence, **(1992)**.
[6]   Y. Wang and I. H. Witten, "Induction of model trees for predicting continuous classes", In: Poster papers of the 9th European Conference on Machine Learning, **(1997)**.
[7]   W. A Jensen, "Decision Trees for Business Intelligence and Data Mining: UsingSAS® Enterprise MinerTM", Technometrics, vol. 50, no. 3, **(2008)**.
[8]   J. Urnkranz, "Separate-and-Conquer Rule Learning", Artif. Intell. Rev., vol. 13, **(1999)**.
[9]   W. N. H. W. Mohamed, M. N. M. Salleh and A. H. Omar, "A comparative study of Reduced Erro Pruning method in decision tree algorithms", IEEE International Conference on Control System, Computing and Engineering, ICCSCE, **(2012)**; Penang, Malaysia.
[10] Decision Tree, Saedsayad.com. Web. 21 Nov., **(2016)**.
[11] D. Aha and D. Kibler, "Instance-based learning algorithms", Machine Learning, vol. 6, no. 1, **(1991)**.
[12] T. Srivastava, "Introduction To K-Nearest Neighbors : Simplified", Analytics Vidhya. Web. 21 Nov., **(2016)**.
[13] R. M. GmbH, "Gradient Boosted Trees - Rapidminer Documentation", Docs.rapidminer.com. Web. 21 Nov., **(2016)**.

Authors

Muhammad Idham Abdul Halim, He is a Graduate Research Assistant at the Universiti Tenaga Nasional (UNITEN) in Selangor, Malaysia. He received his Bachelor's degree from UNITEN in Electrical & Electronics Engineering. He is currently researching in the intelligent algorithms to depict the best network speed for reliable connectivity. His area of interest in research is intelligent network selection algorithm, predictive analytics, IoT and cloud computing. He can be contacted at iddy93@gmail.com

Wahidah Hashim, She received her bachelor degree in Information Technology, Business Management and Language from University of York, UK in 1999. She then pursued her MSc in Multimedia Technology at University of Bath, UK in 2001. She completed her PhD studies from King's College London, UK in 2008 in the field of Telecommunication Engineering. She is currently at Universiti Tenaga Nasional, system and networking department, College of Computer Science and Information Technology as a principle lecturer since 2014. Prior to this, she was a staff researcher

at MIMOS Berhad, a Malaysian National R&D in ICT sector. Apart from her main task in doing research in cognitive radio, WLAN, OFDM, MIMO systems, IoT, Big Data and wireless system, she is actively involved in the development of Malaysia technical specification, standard and guidelines of wireless devices. Dr. Wahidah has published several publication and filed several patents on her research findings. She is a member of the IEEE and IACSIT.