

Preliminary Studies of Predictive Analytics Algorithm for Anticipating Mobile Network Performance Behaviour

Wahidah Hashim¹, Muhammad Idham Bin Abdul Halim¹, Ahmad Fadzil Ismail²,
Kok-Lim Alvin Yau³ and Mohammad Kamrul Hasan²

¹*College of Science Computer & Information Technology,
Universiti Tenaga Nasional, Kajang, Selangor, Malaysia*

²*Department of Electrical and Computer Engineering, International
Islamic University Malaysia, Gombak, Selangor, Malaysia*

³*Department of Computing and Information Systems
Sunway University, Selangor, Malaysia
wahidah@uniten.edu.my, iddy93@gmail.com*

Abstract

In this article, we propose the use of predictive analytics to monitor connection speed in mobile networks. Predictive analytics reduces switching delay which occurs when a network device disassociates and associates itself from one network to another. Using predictive analytics, historical data are recorded and a model is created to assist in formulating a suitable algorithm to predict the best network speed while reducing process complexity. We highlight the significance of the predictive analytics concept using multiple network devices. Real time data is collected by numerous speed measurements on a daily basis. The data is collected by recording the downlink and uplink speeds of network at a regular interval at a specific location. The predictive analytics technique is applied on the collected network dataset using WEKA tools to visualize the pattern of the network speed. The predicted network speed is calculated based on the visualizer classifier error known as J.48.

Keywords: *Predictive analytics, cognitive network selection, multiple network, switching algorithm, WEKA*

1. Introduction

One of the challenges of a network device to select the next reliable operating network is the time it takes to evaluate the connection speed of each candidate networks before a decision is made. Switching to the next reliable operating network is necessary due to the limitation of the device and network protocol that do not allow a device to stay connected with the current operating network, while at the same time scanning for other subscribed networks may offer better communication quality. Quality of the coverage provided by the telecommunication provider or even the device limitation to allow only manual selection are among examples of these limitations. Since simultaneous or parallel evaluation of the networks is generally not feasible, a device must disassociate, scan, evaluate and associate sequentially whenever the performance of the current operating network falls below an acceptable threshold, triggering a new connectivity. Figure 1 describes a common system architecture involving multiple networks [1-2].



Figure 1. Multiple Broadband Communication Devices Embedded with Network Selection Algorithms Reproduced from [1]

Depending on the system architecture, selecting the next reliable operating network can suffer up to a delay factor of 0.16, whilst successive selection can suffer up to a delay factor of 0.54 [1]. Such high delay factor makes successive selection unsuitable for real-time data communication, such as video call, although it can be acceptable in non-real time applications.

The successive selection model started off by connecting and scanning the download speed of each accessible connection and stores this information into a text file. The download speed data retrieved are ranked based on the highest towards the lowest speed which is stored in another text file for ease of data retrieval purpose. From the ranked data, the connection with the highest download speed is selected as a new connection. It then stays with the connection within a period of time which in this study specified as 30 minutes based on the latency of running the "speedtest" script embedded in the model. After 30 minutes, the model is initiated to repeat the overall scanning of the networks. Due to this, the scanning process will occupy the connectivity time. This is the main drawbacks of the model which results into the second selection model which is known as comparative selection.

Having discussed the switching network delay issue in the current model, a technique that can reduce such sequential process is by embedding predictive analytics into the decision making module. Predicted values, which are based on anticipation on historical and trend of the data for each network connection speed helps the decision making module to select the next operating network quicker. Data, in this context, refers to the recorded connection speed monitored within a certain period of time at a specific location under investigation. Greater amount of historical data improves the accuracy of the predictions [3-4].

Predictive analytics have been used frequently and broadly in various fields. Large corporations tend to use predictive analytics in order to gain insights on what their customers are interested in [5-6]. In the medical area, neural networks are applied to predict coronary artery disease amongst the patients [7]. This can be done by analyzing a dataset collected from patients and classifying common traits among those with the disease. In this paper, we scrutinize predictive analytics as a simple technique to model an algorithm that provides forecast on the best network connectivity access based on the

speed of past connections. This algorithm aims to provide a network device with continuous connection to the network while offering the highest network speed among multiple connections in a specific area [3].

However, using predictive analytics to represent data accurately can be an arduous task especially for those who are not familiar with its practice [8]. It becomes apparently difficult when predicting the expected values from a given dataset. This is because the connected network speed has an irregular pattern and has a large variance throughout the day. Before the analyzing and processing steps begin, it is recommended to check for reasonableness of the data set [9]. Checking data reasonableness is important in order to determine whether a value conforms to specified criteria or not. The autonomous network selection algorithm plays an important role in the presence of multiple networks. The real advantage of this algorithm is that the user does not need to manually configure the best connectivity, hence reducing the time delay when switching between networks.

The rest of the paper is organized as follows. Section 2 describes, in brief, the time delay description when switching between networks. Section 3 covers our proposed predictive analytics. Section 4 discusses our experimental set-up for measuring the network behaviour. Experimental results and discussion are discussed in Section 5. Finally, concluding remarks are highlighted in Section 6.

2. Time Delay when Switching between Multiple Networks

The academic community in the recent years has introduced various ways of reducing the time delay incurred when a network device switches between networks. A low-latency MAC protocol for wireless sensor networks (L-MAC) was proposed in [10] and a dynamic duty cycle control protocol was also introduced [11]. This alleged protocol aims to reduce end-to-end delay and serves as a means for energy conservation. With reference to a study in [1], the comparison of switching delays incurred for different selection models is depicted in Figure 2. It is clearly seen that the successive selection model experiences highest delay. This means that the time it takes to switch from one network to another can disrupt the aim of providing seamless connectivity. Note that, in the successive selection mechanism, there are several threshold updates on the instantaneous connection speed, and the update interval is 30 minutes.

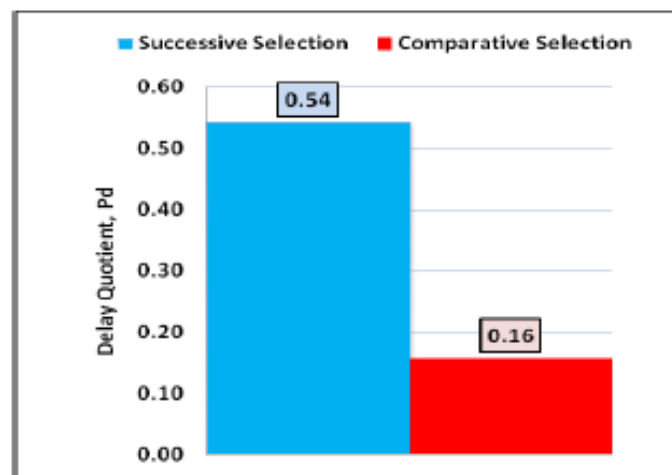


Figure 2. Delay Quotient from Comparative and Successive Models [1]

3. Predictive Analytics

For a single network connection, such as a device that holds only a single sim card, predictive analytics does not limit the capability of forecasting the network speed and identifying the pattern of the network behavior. In this scenario, predictive analytics technique can be developed on various platforms. Amongst the notable ones are R [12], WEKA [13], JMP [14], SAS [15] and even Microsoft Excel. Some of them are already bundled with default algorithms that can be used out of the box. Moreover, they are also equipped with robust visualization tools to represent the data in a more meaningful way. In addition to that, such software platform can incessantly show the final results and the expected output without too much complexity.

3.1. WEKA

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [13]. In this preliminary study, collected data is processed using Weka in a meaningful manner with supported knowledge and learning capability from historical data trend.

4. Experimental Setup

Figure 3 and Table 1 show the experimental setup and system parameters used while collecting data of connection speed.



Figure 3. Emulating the Empirical Experiment for Real Data Collection

Table 1. Parameters for Equipment Testing Network Speed

Parameters	Description
Spectrum band	HSPA – 2.1 GHz,
Operating System (OS)	Ubuntu 12.04 LTS
Location	Indoor Building
Environment	University (Banting)
Daily data monitoring period	8pm - 12pm (9 hours)
Network tested	3G
Specified download size	2 Mbit

We measure the network speed by using a specific software running on a Windows powered laptop. The connection speed is recorded every 30 minutes. For the purpose of data collection, the network speed was recorded from 8.00 pm to midnight.

5. Analysis of Results

The results from the initial speed test and the expected prediction are plotted on a graph and are analysed. Based on the results from Figures 4 and 5, it can be observed that the network speed at the campus area is changing quite rapidly. This can be surprising as there is a higher demand for fast network speed in the campus areas, yet as we can see from Figure 4, the speed never rises higher than 5 Mbps. Now, there can be various factors that influence this type of outcome. First of all, the convoluted and high volume of congestion in the data network has supplied the network speed at full capacity. Secondly, the carrier company might have throttled the data. Now, this is not inherently bad as we might have used up or nearly all of our allocated capacity. With the throttling of data, we can still connect to the internet but at a very low speed. The speed is drastically reduced, sometimes to as low as 32 kbps and sometimes the speed is at par with a 2G network.

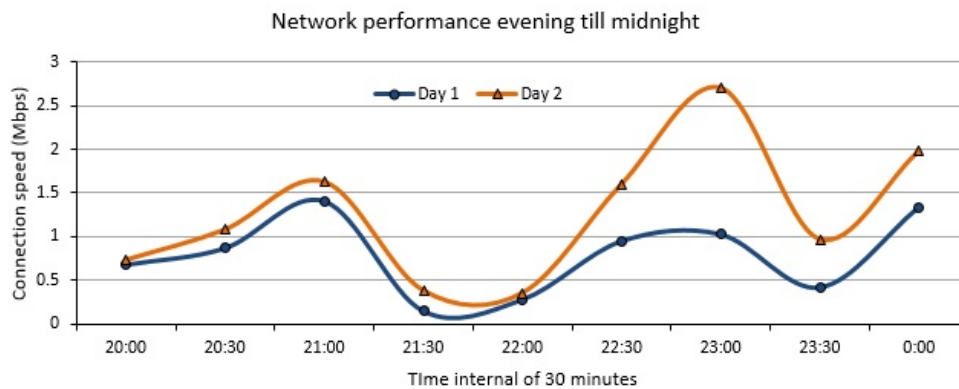


Figure 4. Network Performance in Murni Apartment, UNITEN from Evening to Midnight of for Two Separate Days

If we look at the results from Figure 4, we notice that the network speed is on the low side before 22:00pm for both days. The network speed starts to rise up to 71% after 10 pm.

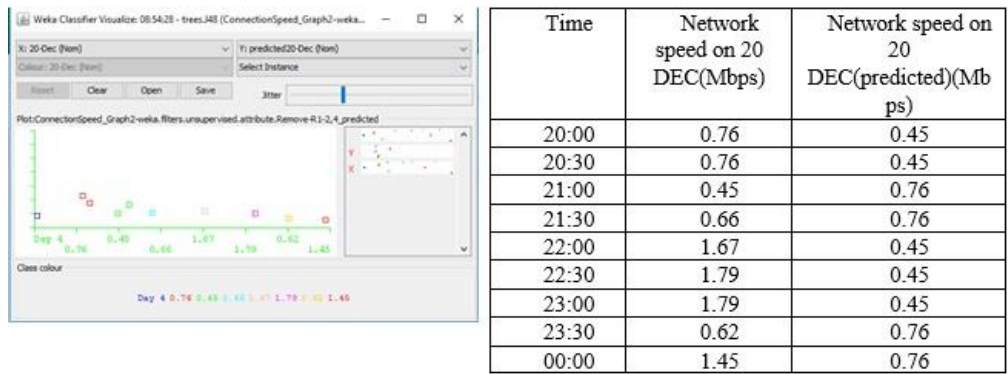


Figure 5. Predicted Network Speed for UNITEN Campus Area

From Figure 5, we run the predictive analysis using the data we already gathered for the two days in the UNITEN campus area. We use WEKA as our data mining tool. We loaded the dataset, albeit the input data and we classified the instances using the renowned J.48 classifier. J.48 is an algorithm used to generate a decision tree. It can be used as classification, so for this reason it is often referred to as a statistical classifier. The predicted network speed was plotted above. We also tweaked the jitter knob, so that we may get apparent results and avoid overlapping of the values.

Table in Figure 5 shows the value of the network speed retrieved from the dataset and the predicted speed that is the product from the visualizer classifier error predictive analysis tools in WEKA using the classifier J48. This is achieved by WEKA learning and processing the dataset to provide a reasonable prediction. It is observed that the network speed is quite far off most of the time after the predictive analysis algorithm has been applied. This implies that the accuracy of predictive analysis is still very low and needs to be worked enhance further. One of the main reasons for this low accuracy is due to the small collected real data. Data should be minimum three folds in order to ensure the predictive method has better accuracy.

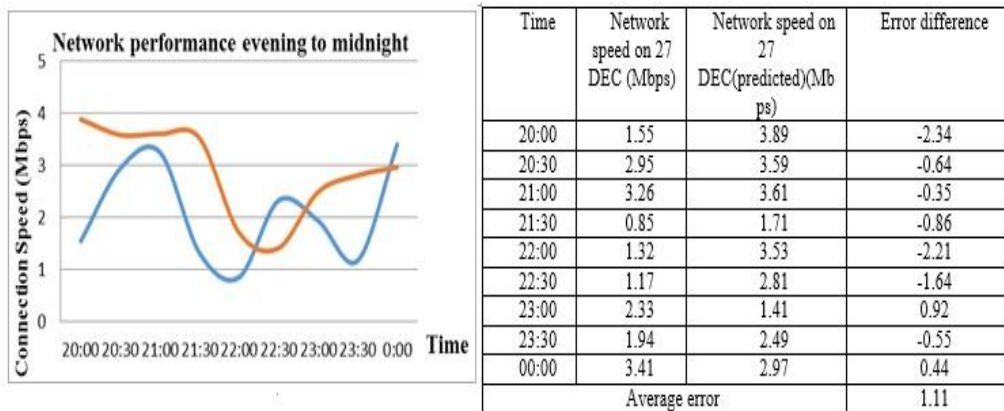


Figure 6. Network Performance in Banting Residential Area from Morning to Evening for Two Separate Days

On the second empirical data collection, we have studied the subscribed network performance for Banting residential area. The same duration of time interval was selected. The data gathered can be shown in Figure 6. The data then processed using Weka to get the predicted values. With this predictive values as shown in Figure 6 table, it is observed that the predicted values still very far from the actual value. The big difference can increase the error of prediction and lower the accuracy. In the table, the average error difference was calculated and the figure is 1.11. What this means is, for every predicted

value that the system provided we need to have a margin of accuracy about ± 1.11 . As an example, is the predicted value resulted into 3.59, the real value could be within the range of 2.48 or 4.70. In this case, the actually value was monitored as 2.95 which is closer to 2.48 lower range value.

The value of network speed was derived from the visualize classifier error. It is apparent that the predicted speed is higher than the real network speed on 27th Dec. This remarks an issue with the accuracy of the predictive analysis algorithm. This may be due to the lack of data observed. As more data is recorded, the accuracy of the algorithm increases.

6. Conclusion

The adoption of predictive analytics algorithm in resolving the issues of switching delays in multiple network connectivity device is described as the preliminary studies of this research work. By designing the predictive algorithm based on the recorded data, it is observed that the algorithm's prediction accuracy is an issue by giving quite a significant error of ± 1.11 value. It was found that the small sample data could be one of the major reason on the low accuracy percentage. For future work, we opt for a longer period of data collection to enhance the accuracy of the predictive algorithm and also to observe the time delay when switching between networks using the predictive algorithm approach.

Acknowledgements

We would like to state our outmost gratitude to the Ministry of Education Malaysia for sponsoring our project based on Fundamental Research Grant Scheme (FRGS) under grant number FRGS/1/2015/ICT02/UNITEN/02/1.

References

- [1] W. Hashim, A. F. Ismail, S. Dzulkifly, and N. A. A. Ghafar, "Cognitive Selection Mechanism for Indoor Propagation", *International Journal of Computer and Comm. Engineering*, vol. 2, no. 4, (2013).
- [2] W. Hashim, A. F. Ismail, S. Dzulkifly, and N. A. A. Ghafar, "Cognitive Selection Mechanism Performance in IEEE 802.11 WLAN", *International Journal of Computer and Comm. Engineering*, vol. 2, no. 4, (2013).
- [3] W. Hashim, A. F. Ismail, N. M. Anas, H. A. M. Ramli, and K. Abdullah, "Cognitive Power Adjustment for OFDM-Based Wireless Communication System: An Empirical Study", *Lecture Notes on Software Engineering*, vol. 3, no. 1, (2015).
- [4] "Oracle® Cloud: Working with Predictive Planning for Oracle Planning and Budgeting Cloud Service", (2015).
- [5] Y. Lu, "Integrating predictive analysis & social media", *IEEE Conference on Visual Analytics Science & Technology (VAST)*, Paris, France, (2014).
- [6] D. Pareit, B. Lannoo, I. Moerman, and P. Demeester, "The history of WiMAX: A complete survey of the evolution in certification and standardization for IEEE 802.16 and WiMAX", *IEEE Communication. Survey Tutorials*, vol. 14, no. 4, (2012).
- [7] A. Alfayly, I.-H. Mkwawa, L. Sun, and E. Ifeachor, "QoE-based performance evaluation of scheduling algorithms over LTE", *IEEE Globecom Workshops*, (2012).
- [8] E. W. Steyerberg, F. E. Harrell, G. J. J. M. Borsboom, M. J. C. Eijkemans, Y. Vergouwe, and J. D. F. Habbema, "Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis", *Journal Clin Epidemiol*, vol. 8, no. 75, (2001).
- [9] C. McCue, "Data Mining and Predictive Analysis", *Butterworth-Heinemann*, (2007).
- [10] Y. Yue and J. Liu, "A Low-Latency MAC Protocol for Wireless Sensor Networks", *Chinese Journal of Electron Devices*, vol. 3, no. 31, (2008).
- [11] X. Wang and X. Wang, "Dynamic Duty Cycle Control for End- to-End Delay Guarantees in Wireless Sensor Networks", *IEEE International Workshop on Quality of Service (IWQoS)*, Beijing, China, (2010).
- [12] J. Monogan, "An Introduction to R", *User Man*, (2009).
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software", *An update, SIGKDD Explor.*, vol. 1, no. 11, (2009).
- [14] B. Jones and J. Sall, "JMP statistical discovery software", *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 3, no. 3, (2011).

[15] R. N. Rodriguez, "SAS, Wiley Inter-discipline", Revision, Computer, Statistics, vol. 1, no. 3, (2011).

Authors



Wahidah Hashim, 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 publications and filed several patents on her research findings. She is a member of the IEEE and IACSIT.



Muhammad Idham Abdul Halim, 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 directly via email at iddy93@gmail.com.