An Intelligent Crisis-Mapping Framework for Flood Prediction

Siti Azirah Asmai, Zaheera Zainal Abidin, Halizah Basiron, Sabrina Ahmad

Abstract— This paper proposes a new framework for crisis-mapping with flood prediction model based on the crowdsourcing data. Crisis-mapping is still at infancy stage development and offers opportunities for exploration. In fact, the application of the crisis-mapping gives fast information delivery and continuous updates for crisis and emergency evacuation using sensors. However, current crisis-mapping is more to the information dissemination of flood-related information and lack of flood prediction capability. Therefore, this paper applied artificial neural network for flood prediction model in the proposed framework. Sensor data from the crowdsourcing platform can be used to predict the flood-related measures to support continuous flood monitoring. In addition, the proposed framework makes used of the unstructured data from the Twitters to support the flood warnings dissemination to locate flood area with no sensor installation. Based on the results of the experiment, the fitted model from the optimization process gives 90.9% of accuracy performance. The significance of this study is that we provide a new alternative in flood warnings dissemination that can be used to predict and visualized the flood occurrence. This prediction is significant to agencies and authorities to identify the flood risk before its occurrence and crisis-maps can be used as an analytics tool for future city planning.

Index Terms: Crisis-mapping, flood prediction, crowdsourcing, artificial neural network, visualization.

I. INTRODUCTION

Crowdsourcing data has become an active area used among internet user which offers a wealth of valuable information in many fields such as business, education, political, healthcare, disaster management, emergency response and more. With the Internet of Thing (IoT) technology, the masses of crowdsourcing data are available on the social media platforms such as Facebook, Twitter, YouTube and blog provide fast information and responses [1]. However, there is limited use of crowdsourcing data for crisis and disaster events. This is because the technology is too dependent on good data acquisition [2,3], best performance and required the proper techniques to make it intelligently to be used in the disaster situation [4]

Furthermore, crowdsourcing data can be in a form of structured and unstructured, which demand various techniques and approaches[5]. The structured data can be obtained in many platforms of applications. Some of the applications are integrated with the IoT devices i.e. sensor, which updates in daily basis.

On the other hand, unstructured data from the social media can be used to understand the situation and people in the affected area when they are communicating with each other [6]. However, with no proper textual data analysis i.e. text message may not capable to convey conceivable detail to guide the actions taken. In addition, the text also needs to be filtered and determine its trustworthiness so that it will be reliable for further analysis related to the crisis.

Therefore, in order to ensure a proper solution in crowdsourcing data, crisis-mapping is used as a mechanism to aid emergency response planning by performing and displaying data analytics during a crisis or disaster. Crisis or disaster map is referred as information gathering, visualizing and analysis of data during a crisis, including political, social and environmental ones with different development intensity. Crisis-mapping allows a large group of people to control, public, volunteers and crisis responders either from the site of the crisis or at a distance, to share information regarding on the crisis matter [7].

The first event that utilizes the crisis-mapping was the Haiti earthquake in 2010. Later, many crises and emergency cases have been reported using crisis-mapping. On the other hand, the natural disaster that highly reported in Malaysia is flood, since the location of Malaysia is influential by the monsoon. In 2015, the biggest flood occurred was in Kelantan, followed by other countries. Many people were homeless and killed. Thus, the application of crisis-mapping has been tremendously used throughout Malaysia and aided the emergency response team.

Based on the preliminary study, we found that the current flood visualization solutions are typically based on sensor-based information, which consumes a lot of time and resources. Even though the information extracted from the sensors, but the reliability data still remain a challenge. Flood-related data can be obtained in the respective duration but it often performs poor performance due to the sensor faults conditions [3].

Therefore, this paper proposes the framework of intelligent crisis mapping to visualize the flood event. At the early stage of this proposed crisis-mapping, this

crisis –mapping engine will analyze the flood-related data from the existing corporate database and a flood prediction model is constructed using a neural network technique. The values from the prediction model will be visualized on the map through a mobile application.

II. RELATED STUDY

Flood warning dissemination is one of the important efforts that is being taken seriously by many flood risk management authorities in order to continuous improving their flood warning and forecasting systems through a rigorous engineering and research methodology [4].

There are many studies that have used sensors for flood data collection [8–10]. For monitoring flood event, the river level sensor can be used to monitor flood from a base station. The sensors can show the level of river water and supposedly give the certain alerts for the users to prepare for evacuation and other actions. In another example, the volume of rainfall at the potential flood location area also can be used to indicate the flood occurrence and perform flood warnings [11].

Regardless of the sensor systems state-of-the-art, readings that lead to flood indicator may not be accessible or contains noise due to faulty and failure sensor. Thus, this problem leads to data no longer supported to the fast response and emergency evacuation. Therefore, in order to reduce the dependencies on the reliability of the sensors located in the crisis places, prediction data are needed to fulfill the reliability of the crisis flood information.

In addition, the places that are not installed with sensors are out of coverage in flood monitoring Therefore, the people who are affected by the flood will not be to receive the alert or any crucial information. In fact, people who are affected can be mobile users who are potential to be a contributor to real-time reports on the flood situation using social media such as Twitter, Facebook and YouTube. Furthermore, it is crucial to have a platform for handling massive information and voluminous from the crowd.

The crowdsourcing can contain all information in the form of unstructured textual data [12]. Supposedly, each word from the crowd should be analyzed intensively because it contains the important information that leads to how the flood prevention and actions can be taken and executed. However, this hand-label identification analysis may lead to a time consuming and resume huge resources. However, the crowd data can be arranged in pre-determined structure feedback, but it might mislead the structure of all information [13]. In another hand, if the information is simply captured without formal structure cluster, the informative crisis may difficult to be extracted and influence the effective flood evacuation.

Therefore, an integration mechanism should be formulated to identify the important information that related to flood either from the sensor data or from the huge of open–textual domain data to supply the crisis information for better flood prevention action. There are several studies proposed the similar applications. A study from Maidment [14] proposes a framework for flood prediction using

geospatial information for local flood emergency response and aim to obtain a real-time flood inundation map and flood emergency response. This crisis mapping will be helpful for disaster planning in addition to an actual emergency response to the floods. Machine learning techniques such as Neural Network can be used for flood prediction.

However, there is lack of studies use of the NN for flood prediction applications. Some of the predicted results from the NN models are unsatisfactory due to accuracy performance between the actual and predicted readings. In fact, the predicted reading from the NN prediction can be used when the monitored data unavailable because of the failure or faulty sensory.

III. THE PROPOSED FRAMEWORK

The proposed framework is Intelligent Crisis-Mapping for flood visualization. In fact, the proposed framework can be also integrated with existing crisis monitoring applications and devices. In this framework, there are four main phases of the framework: Data preparation, Feature Extraction, Prediction Mapping and Application as depicted in Fig. 1. The detail of the designed framework is addressed in the following subsections. In fact, the proposed framework can be also integrated with existing crisis monitoring applications and devices.
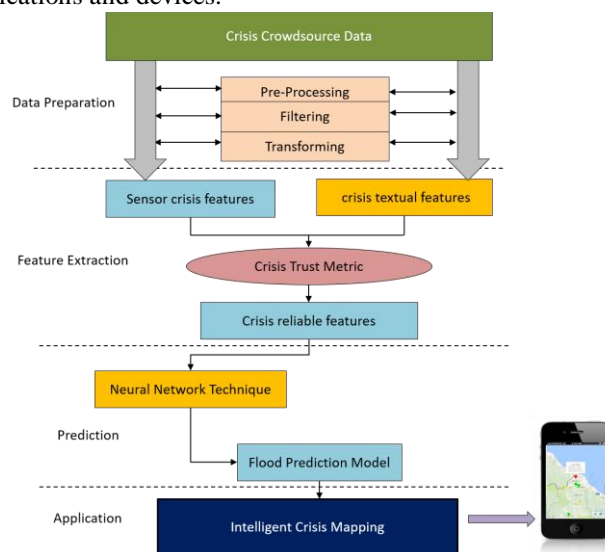

Fig. 1. Framework of Intelligent Crisis Mapping for Flood Visualization

a. Data Preparation

The data from crowdsourcing platforms such as public websites, Twitter and other social media platforms will be extracted. The method of the extracting the data is currently using web scrapping. There are two types of data used; sensory data and textual data. Both types of data will undergo a set of preparation process; pre-processing, filtering and transformation. It must be noted that not all the processes will be used simultaneously, it depends on the characteristic or state of the readiness of the

data. For sensory data, data that extracted from public crowdsourcing platform contains a missing value at a specific date, which needs to treat the missing values and filter the extreme values. For unstructured text data, the data is tokenized into single words and phrases, followed by removing all stop word. This textual data will be transformed in which a collection of text information that is related to crisis domain.

b. Feature Extraction

In the second phase, the process of extracting features for flood-related data task has been performed. These features extraction is divided into two parts. The first part, the structured data from sensory data is arranged with the targeted flood measure. The second part, the unstructured data that extracted from the social media platform for example Twitter is identified and differentiate based on the word list that contribute to tweet that referring to the flood in natural disaster domain. Since the word flood might be used by the Twitter user to describe another meaning, therefore, a list of stops word is used to filter the collected tweets. The stops word and related word list will be updated from time to time along with the exploration of new words or terms in flood domain.

Each tweet that has been tokenized into a single word at the data preparation phase as shown as in Fig. 2 is used to generate flood-related index is through a method called Term Frequency-Inverse Document Frequency (TF-IDF).

Fig. 2. Tokenization of Flood Tweets

TF-IDF computes the frequency of term or word that we assume to be related with flood domain, in each tweet or string of data. This numerical statistical method also is like a weighting factor that will increase proportionally to the frequency of the target word or term appear in the tweet. The higher the total frequency of all target words found within a tweet, a higher score will be given to the tweet. In other words, the flood-related index of the tweet increases as the final score from the TF-IDF method increase.

c. Crisis Trust Metric

Based on sensory and text features combination, the crisis parameters are identified for generating crisis trust metric. This crisis trust metric will be used to ensure that only reliable crisis features will be selected for flood prediction.

The crisis trust metric is responsible to filter the tweet with the weighted TF-IDF to ensure the reliability of the flood information is according to river level that obtained from the sensory data. If the flood data is in a normal condition, but some tweets show differently so it gives a "Low Index" of the crisis trust. Fig. 3 depicts the example of series of river level settings such as "Normal", "Alert", "Warning" and "Danger" from flood management authority's website that can be useful in calculating the "Crisis Trust Metric". These measurements in Fig. 3 will only be used once the text extracted from social media are successfully classified. The outcome from the crisis trust metric is the set of the crisis reliable features from sensors and also text classification data.

Fig.3. River Level Setting

d. Flood Prediction Model Generation

The flood prediction model is developed based on the crisis reliable features data by using Artificial Neural Network (ANN) technique. The ANN is used in the proposed framework to predict the targeted flood-related measure. By using the concept of learning through experience, ANN gathers the information and then detects the patterns and relationship in the data. An ANN architecture constitutes a computational model that contains hundreds of artificial neurons and connected with coefficients known as weights. This computational model is used to generate the desired output based on learning the information process from connecting neurons in the network. Due to its learning ability on the data, ANN can be applied in many different applications of industry, medicine, games, science and technology[15].

Typically, an ANN is designed for classification [16,17] and now ANN can also be applied in flood prediction. The implementation of ANN in flood prediction requires the input data to be arranged in sequence and segmented for constructing prediction outputs. To apply a neural network, several key parameters need to be considered, namely the number of hidden layers, the number of nodes and activation functions. Deciding the number of input and hidden nodes have always been an issue, which is having a smaller number of hidden nodes tended to have less adequate performance. Meanwhile, having too many nodes may increase the risk of over-fitting the data and impede generalization [15].

On the other hand, current analytics software such as SAS® Visual Data Mining

and Machine Learning, has the ability to determine the best architecture of the neural network model in order to overcome the problems in determining the best architecture for ANN model. By using local search optimization approach, the model parameters (or often called as hyper parameters) can be optimized during training for fitting the best model as explained in [18]. Once the optimized neural network model is completely obtained, the testing datasets are fed into the model and compared with the actual measures for accuracy performance.

e. Application

The proposed framework produces an Intelligent Crisis Mapping app. This mapping will use the new crowdsources data to generate the predicted flood measures through the developed model. With the integration of Google maps and properties of the location, the predicted flood based on the results of neural network on sensors data and textual analysis of tweets can be visualized in the map. Thus, this Intelligent Crisis Mapping can be developed either through mobile applications or the web applications for flood information dissemination.

IV. FIRST RESULTS

It this paper, we currently focus on developing a flood prediction model in the proposed crisis- mapping as the early stage of our study. We initially used sensory data collected from the authority of flood management. The experiment for the flood prediction model is based on the flood event in Kelantan state of Peninsular Malaysia. Sungai Kelantan base station is chosen as a case study for generating the flood prediction model as the basin of Sungai Kelantan are highly populated. The sensors are located at the observed stations of Sungai Kelantan and inputs from specifically two stations, which are Lebir and Sokok stations, are used to predict river level at Sungai Kelantan as shown in Fig. 4. The input parameters are based on the river levels and rainfall data.

Fig. 4. Map of the Sungai Kelantan

Table 1. Summary of the rain measures of Kelantan stations

Types of Dataset	Parameters	Stations		
		Lebir	*Sokor*	*Sungai Kelantan*
Training January 2013 – November 2014	Mean rainfall	6.19	7.53	8.97
	Minimum Rainfall	0.00	0.00	0.00
	Maximum Rainfall	140.50	159.00	213.70
	Mean River Level	23.81	15.58	8.83
	Minimum River Level	22.69	14.80	7.54
	Maximum River Level	35.91	22.02	17.64
Test 1 November 2015 – February 2016	Mean rainfall	7.50	8.92	11.63
	Minimum rainfall recorded	0.00	0.00	0.00
	Maximum Rainfall Recorded	79.10	139.10	177.60
	Mean River Level	23.94	15.50	10.39
	Minimum River Level	23.02	14.96	9.50
	Maximum River Level	31.47	20.20	14.97

Test 2 June 2016 – February 2017	Mean rainfall	7.76	4.33	18.07
	Minimum rainfall recorded	0.00	0.00	0.00
	Maximum Rainfall Recorded	216.70	72.50	450.70
	Mean River Level	24.27	15.61	10.52
	Minimum River Level	23.12	14.72	8.65
	Maximum River Level	34.71	22.07	18.20
Test 3 June 2017 – November 2017	Mean rainfall	9.90	9.96	7.49
	Minimum rainfall recorded	0.00	0.00	0.00
	Maximum Rainfall Recorded	125.80	154.20	74.70
	Mean River Level	23.97	15.53	9.46
	Minimum River Level	23.35	15.04	8.30
	Maximum River Level	32.34	21.78	15.61

The sensors readings that were taken from January 2013 until November 2014 are used for training the model. Three data sets from November 2015 until February 2016, June 2016 until February 2017 and June 2017 until November 2017 are used for testing the model and evaluate the accuracy performance. Table 1 shows the sensors readings from three stations which are Lebir, Sokok and Sungai Kelantan.

The optimization techniques are used with the training dataset to obtain the best possible architecture of the neural network model. As the result, multi-layered perceptron (MLP) is chosen by using Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) technique. The optimized neural network model is run with the tested data to observe the model accuracy performance.

The model accuracy performance is described in terms of the error of prediction or the residual between the actual and predicted values. In the literature, there are many performance assessment methods to measure the accuracy and each one has advantages and limitations. The most common performance accuracy measurements that can be used to evaluate the residual in prediction are the MAE and RMSE [19,20]. In this paper, the accuracy measurements namely mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) are used to check the accuracy performance of the optimized ANN. The MAE measures the average magnitude of the error in prediction values equally without considering its direction. Therefore, MAE is suitable to be used to describe uniformly

distribution errors [20]. The RMSE measures the average of the square of all the errors. Since the errors are squared before averaged, RMSE is an excellent tool to amplify and sensitively treat the large errors [21]. The MAE, MSE and RMSE can be calculated as shown in Table 2.

Table 2. Performance Metric

Performance Metric	Calculation
MAE	$\frac{1}{n}\sum_{t=1}^{n}\left\|y_t - \hat{y}_t\right\|$
MSE	$\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2$
RMSE	$\frac{1}{n}\sum_{t=1}^{n}\sqrt{\left(y_t - \hat{y}_t\right)^2}$

where n is the total number of the time series data, y_t and $\hat{y}_t$ are the actual and predicted value of the time series in time t respectively.

Table 3 outlines the performances result based on both training and testing data. From Table 3, the error of prediction is small and produced produce excellent accuracy with test datasets. Based on the RMSE results, it shows the accuracy of the prediction is 90.9% on average. Overall, the model can be used for flood applications.

Table 3. Model Performance Results

Dataset	Performance Measurements		
	Mean Absolute Error (MAE)	Mean Square Error(MSE)	Root Mean squared error (RMSE)
Training	0.0384	0.0026	0.0514
Test 1	0.0517	0.0051	0.0715
Test 2	0.0856	0.0102	0.1010
Test 3	0.0799	0.0100	0.1000

In this study, the flood application is developed on mobile-based to implement flood prediction and visualization. The application uses new flood-related data that are extracted from the local flood management authority website. The data extraction is done by requesting from website using the URL and reading the HTML page contents. With the targeted tag, the required new flood related-data are extracted into the database and used to generate the predicted river level.

Fig. 5. Flood Application

By creating an app, to map with markers, we produce a map that shows the location of the sensors or the target location with the threshold value for tweets collected in flood domain. We use three distinct colors for the marker to differentiate the flood prediction which is green, yellow and red, which represents low chance, medium chance and a high chance of flood respectively as shown in Fig. 5.

V. CONCLUSION

In this paper, a crisis mapping framework with flood prediction model is presented. This flood prediction model is our early stage of the crisis mapping development. For the development of flood prediction model, the training and the testing data from the sensors readings that installed in three different stations are observed. The fitted prediction model gives promising result nearly 90% accuracy with testing datasets. From this model, the flood-related data is predicted with the new crowdsource data. The predicted data is mapped and visualized on the app. The app is useful for fast response and emergency flood evacuation operations. In addition, this application can help authorities to identify the flood risk before its occurrence and crisis maps can use as an analytics tool for future city planning. As the future work, the developed app will combine the text data that extract from social media to visualize the flood in the map.

VI. ACKNOWLEDGMENT

REFERENCES

1. R.Q. Wang, H. Mao, Y. Wang, C. Rae, W. Shaw, Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data, Comput. Geosci. 111, 2018, 139–147.
2. L. Cheng, J. Niu, C. Luo, L. Shu, L. Kong, Z. Zhao, Y. Gu, Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks, Comput. Networks. 134, 2018, 66–77.
3. A.P. Das, S.M. Thampi, Fault-resilient localization for underwater sensor networks, Ad Hoc Networks. 55, 2017 132–142.
4. H. Kreibich, M. Müller, K. Schröter, A.H. Thieken, New insights into flood warning reception and emergency response by affected parties, Nat. Hazards Earth Syst. Sci. 17, 2017, 2075–2092.
5. E. Niforatos, P. Campos, A. Vourvopoulos, A. Doria, M. Langheinrich, Atmos, Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Adjun. Publ. - UbiComp '14 Adjun. 5, 2014, 135–138.
6. R.Q. Wang, H. Mao, Y. Wang, C. Rae, W. Shaw, Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data, Comput. Geosci. 111, 2018, 139–147.
7. I. Norheim-Hagtun, P. Meier, Crowdsourcing for crisis mapping in Haiti, Innov. Technol. Governance, Glob. 5, 2010, 81–89.
8. K. Auynirundronkool, N. Chen, C. Peng, C. Yang, J. Gong, C. Silapathong, Flood detection and mapping of the Thailand Central plain using RADARSAT and MODIS under a sensor web environment, Int. J. Appl. Earth Obs. Geoinf. 14 , 2012, 245–255.
9. T. Ali, L.T. Jung, I. Faye, Diagonal and Vertical Routing Protocol for Underwater Wireless Sensor Network, Procedia - Soc. Behav. Sci. 129 , 2014, 372–379.
10. L. Cheng, J. Niu, C. Luo, L. Shu, L. Kong, Z. Zhao, Y. Gu, Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks, Comput. Networks. 134 , 2018, 66–77. doi:10.1016/j.comnet.2018.01.012.
11. C.W. DAWSON, R. WILBY, An artificial neural network approach to rainfall-runoff modelling, Hydrol. Sci. J. 43, 1998, 47–66.
12. A. Moumtzidou, S. Andreadis, I. Gialampoukidis, A. Karakostas, S. Vrochidis, I. Kompatsiaris, Flood Relevance Estimation from Visual and Textual Content in Social Media Streams, Companion Proc. Web Conf. 2018. , 2018, 1621–1627.
13. S. Salleh, A. Asmai, H. Basiron, S. Ahmad, A Malay Named Entity Recognition Using Conditional Random Fields, in: 5th Int. Conf. Inf. Commun. Technol. (ICoIC7), 2017, 2017: pp. 44–49.
14. D.R. Maidment, Conceptual Framework for the National Flood Interoperability Experiment, J. Am. Water Resour. Assoc. 53 (2017) 245–257.
15. S.A. Asmai, B. Hussin, M.M. Yusof, A.S. Shibghatullah, Time Series Prediction Techniques for Estimating Remaining Useful Lifetime of Cutting Tool Failure,

Retrieval Number: B10580882S819/2019©BEIESP
DOI:10.35940/ijrte.B1058.0882S819

1309

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Int. Rev. Comput. Softw. 9 (2014) 1783–1790.

16. B.M. Zahran, Classification of Brain Tumor Using Neural Network, Int. Rev. Comput. Softw. 9, 2014, 673–678.

17. M.W. Ashour, F. Khalid, L.N. Abdullah, A.A. Halin, Artificial Neural Network-Based Texture Classification Using Reduced Multidirectional Gabor Features, Int. Rev. Comput. Softw. 9, 2014, 1007–1016.

18. P. Koch, B. Wujek, O. Golovidov, S. Gardner, Automated Hyperparameter Tuning for Effective Machine Learning, (2017) 1–23.

19. S. Ismail, A. Shabri, R. Samsudin, A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting, Expert Syst. Appl. 38 ,2011, 10574–10578.

20. T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature, Geosci. Model Dev. 7, 2014, 1247–1250.

21. S. Saigal, D. Mehrotra, Performance comparison of time series data using predictive data mining techniques, Adv. Inf. Min. 4, 2012, 57–66.

AUTHOR PROFILE

Dr. Siti Azirah Asmai is a senior lecturer at Faculty of Information and Communication at Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. She obtained her BSc in Computer Science from Universiti Teknologi Malaysia and MSc. in Information and Communication Technology for Engineer from Coventry University. She received her Ph.D. in Information and Communication Technology from UTeM in 2014. Her current research work focuses predictive analysis, Data analytics and Visualization, Time Series and Intelligent Computing

Dr. Zaheera Zainal Abidin, PhD in I.T. and Quantitative Sciences, MSc. in Quantitative Sciences, MSc. in Computer Networking (Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor) and Bachelor of Information Technology from University of Canberra, Australia. Senior Lecturer at Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (2009 – present), Universiti Kuala Lumpur (2005-2009) and ExxonMobil Kuala Lumpur Regional Center as a Project Analyst (2000-2001). Also, certified of Internet-of-Things specialist (CIoTS) and CISCO Academy Instructor. Research interest include Internet-of-Things, computer networking, biometrics, steganography and image processing.

Dr. Halizah Basiron is a senior lecturer at Faculty of Information and Communication at Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. She obtained her BSc in Computer Science from Universiti Teknologi Malaysia and MSc. in Interactive Computing System Design from University of Loughborough. She received her Ph.D. in computer Science from University of Otago in 2012. She does research in Artificial Intelligence especially in Natural Language Processing. She is now doing some research on Sentiment Analysis, Subjective Identification and Corpus Development on (multilingual) written text.

Dr. Sabrina Ahmad was born in Muar, Johor, Malaysia on August, 15 1978. She obtained her first degree in Information Technology (Hons.) from Universiti Utara Malaysia in 2000. In 2002, she obtained Master of Science (Computer Science – Real-time Software Engineering) from Universiti Teknologi Malaysia. Her doctoral degree (PhD in Computer Science) was obtained from The University of Western Australia in 2012. She is specialized in Software Engineering particularly in Requirements Engineering. Dr. Sabrina is currently a Coordinator for Centre of Advanced Computing Technology and a Senior Lecturer at Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka.