

A MULTIDIMENSIONAL DATA
DESCRIPTOR TOOL BASED ON
FUZZY MIN MAX NEURAL NETWORK
ALGORITHM

FATIN NURJANNAH BINTI SALAUDDIN

Bachelor of Computer Science (Graphics and
Multimedia Technology)

UNIVERSITI MALAYSIA PAHANG



SUPERVISOR'S DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of ~~*Doctor of Philosophy/ Master of Engineering/ Master of Science~~ In *PSM*

A handwritten signature in black ink, appearing to read 'Mohammed Fatah'.

(Supervisor's Signature) Full

Name : *Mohammed Fatah*

Position :

Date : *10/1/19*



STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

A handwritten signature in black ink, appearing to be 'Fatin Nurjannah Binti Salauddin', written over a horizontal line.

(Student's Signature)

Full Name : FATIN NURJANNAH BINTI SALAUDDIN

ID Number : CD15052

Date : 09 January 2019

A MULTIDIMENSIONAL DATA DESCRIPTOR MODEL BASED ON FUZZY
MIN MAX NEURAL NETWORK ALGORITHM

FATIN NURJANNAH BINTI SALAUDDIN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
for the award of the degree of
Bachelor of Computer Science (Graphic and Multimedia Technology)

Faculty of Computer System & Software Engineering
UNIVERSITI MALAYSIA PAHANG

MAY 2018

ACKNOWLEDGEMENTS

All praises to Allah S.W.T because of the strength given and his bless as I able to complete this project. I would like to express my sincere appreciation to the individuals that have been involved in helping me to completed this project.

Also, my appreciation and thanks goes to my supervisor, Dr. Mohammed Falah Mohammed for all the guidance, assistance and all the comment that help me to complete this project according the schema and make me understand more about this project.

Last but not least, I would like to thanks my parents, En. Salauddin bin Mohd Hashim and Ainul Mardliah, and also to my siblings that always give me a moral support and motivation for me to finish this project.

ABSTRAK

Tujuan projek ini adalah untuk memperkenalkan pelbagai teknik dan kaedah yang berbeza yang pernah digunakan sebelum ini untuk menganalisis data. Objektif utama adalah untuk membina alat bagi data analisis untuk data yang terdiri dalam pelbagai dimensi. Dalam projek ini, Teknik yang akan digunakan adalah teknik Fuzzy Min Max. Dengan menggunakan method ini, visualisasi data akan memaparkan nilai minimum dan maksimum dalam nilai antara kosong ke satu. Dalam projek ini, pelbagai teknik akan dibandingkan dan menentukan teknik yang paling sesuai. Teknik - teknik yang dibandingkan adalah semua teknik yang ada dalam Neural Network seperti K-Nearest Neighbour Fuzzy Min Max, general Reflex Fuzzy Min Max, dimana untuk mendapatkan idea berkaitan method, algorithm, teknik dan konsep untuk projek dan kajian yang sedia ada. Selain itu, dalam projek ini, tiga sistem yang sedia ada akan dibandingkan dari segi kebaikan dan keburukan setiap sistem, seperti Tableau Public, Qlikview dan IBM InfoSphere Streams. Projek ini dibuat menggunakan Matlab Programming. Oleh itu, dalam projek ini, akan diterangkan lebih lanjut mengapa saya menggunakan Fuzzy Min Max berbanding dengan kaedah yang lain.

ABSTRACT

The purpose of this project is to introduce different techniques and methods that have been used before to analyze the data. The main objective is to build an data analytic tool for a multidimensional dataset. In this project, the technique that will be used is Fuzzy Min Max method. By using this method, the data visualization will displaying the minimum and maximum value in range of zero to one. In this project, it will be compare a few techniques and determine which techniques is the most suitable. The techniques is in the Neural Network which has a few popular techniques such as K-Nearest Neighbour Fuzzy Min Max, general Reflex Fuzzy Min Max to get some idea of methodologies, algorithm, techniques and concept of the whole existing project and research study. Besides that, in this paper, it will also compare three existing system such as Tableau Public, Qlikview and IBM InfoSphere Streams. They have been compare for their advantages and disadvantages. The implementation of Fuzzy Min Max Neural Network technique has been applied using Matlab Programming. Therefore, in this project, it will explain more about why I am using the Fuzzy Min Max method rather than other methods.

TABLE OF CONTENT

DECLARATION	
TITLE PAGE	
ACKNOWLEDGEMENTS	iiiiiii
ABSTRAK	iv
ABSTRACT	v
TABLE OF CONTENT	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xxi
CHAPTER 1 INTRODUCTION	12
1.1 Background	1-3
1.2 Problem Statement	3
1.3 Aim and Objectives	4
1.4 Scope	4
1.5 Significance of this project	4
1.6 Thesis Organization	4-5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Neural Network	6-7

2.2.1	Fuzzy Min Max Neural Network	7-8
2.2.2	K-Nearest Fuzzy Min Max	8-9
2.2.3	General Reflex Fuzz Min Max	9-10
2.3	Summary	11
2.4	Existing system of data analytical tools	11-12
2.4.1	Tableau public	12
2.4.2	Qlikview	12
2.4.3	IBM Infosphere Streams	12
2.5	Summary	13
2.6	Conclusion	14
	CHAPTER 3 METHODOLOGY	15
3.1	Introduction	15
3.2	Methodology	15-16
3.2.1	Analyse Background Study	16
3.2.2	Literature review	16-17
3.3	Fuzzy Min Max Neural Network	17-23
3.4	Context diagram and Use Case diagram	23
3.5	Flowchart and dialogue diagram	24
3.6	Proposed User interface	25

3.7	Hardware and Software	25
	3.7.1 Hardware requirement and specification	25
	3.7.2 Software requirement and specification	26-27
3.8	Gantt Chart	27
3.9	Conclusion	27
CHAPTER 4 RESULT AND DISCUSSION		
4.1	Introduction	28
4.2	Result Discussion	29
CHAPTER 5 CONCLUSION		
5.1	Introduction	30
5.2	Research Constraints	30-31
5.3	Future Work	31
REFERENCES		32-34
APPENDIX A SAMPLE APPENDIX 1		35

LIST OF TABLES

Table No	Table Title	Page Number
2.1	Advantages and Disadvantages	11
2.2	Advantages and Disadvantages	13
3.1	Hardware Requirements and Specifications	26
3.2	Software Requirements and Specifications	26

LIST OF FIGURES

Figure number	Figure title	Page number
Figure 2.1	General Reflex FMM	Page 10
Figure 3.1	Research Methodology	Page 14
Figure 3.2	FMM learning algorithm	Page 19
Figure 3.3	Example of expansion problem	Page 21
Figure 3.4	Hyperbox overlap test	Page 21
Figure 3.5	FMM pseudocode	Page 22
Figure 3.6	EFMM pseudocode	Page 23
Figure 3.7	Context Diagram for data analytic tool	Page 23
Figure 3.8	Use Case for data analytic tool	Page 24
Figure 3.9	Flowchart	Page 24
Figure 3.10	Dialogue Diagram	Page 25

LIST OF ABBREVIATIONS

Abbreviation	Title
ANN	ARTIFICIAL NEURAL NETWORK
FMM	FUZZY MIN MAX
EFMM	ENHANCED FUZZY MIN MAX
KNFMM	K-NEAREST FUZZY MIN MAX
NN	NEURAL NETWORK

CHAPTER 1

INTRODUCTION

1.1 Background

As the world are developing more and more, big companies are getting success and success, where the increasing amount of people means there are many data and information to be analyze. In fact, data analyses have been used in different areas of life like in finance, science and research, and business insight [1]. That analysis covering different kind of datasets including supervised, and unsupervised data sets [2]. Here and after, the focus will be on the supervised data set as one of the most important data in classification process.

In fact, analysing could help to diagnosis the limitation or the strength of the data and that could lead to improve the system performance. The data analysing are been use in many area such as medical, education, business and extra [3]. There are many data analytics tools that have been used in that area, for example, for the medical area, the most popular is the Hadoop (Apache platform) which is the open source distributed data processing. Initially, Hadoop was developed for the routine functions. Hadoop have been used widely for process extremely large amount of dataset and also can serve as two roles which are as analytical tools and data organizer. Even though Hadoop has many advantages, however person with Hadoop skill are rarely to be found due to the difficulty of installation and configuration processes [4].

Beside Hadoop, there are other tools that been used in others area such as Tableau and IBM. Tableau tool is useful for the data visualization process. It is very easy to be used even if the user doesn't have a strong background skill about it. Tableau using the simple drag and drop method which allow user to directly interact,

visualizing data, controls the analyses process, and generate the final report [5]. Although Tableau has many advantages, it is not completely stable due to its open source functionality. In case of IBM, it design to support IBM's global campaign for Business Analytics and Optimization (BAO). IBM has three popular platforms that supposed to support a real-time analytics processing (RTAP), such as, IBM's acquisition of Netezza which introduced in 2010, IBM InfoSphere BigInSights in 2011, and IBM InfoSphere Streams [6]. All these three IBM products can define the modern analytics, include built-in text analytics and provide the velocity for the big data on structured and unstructured data. But, not all tools can cooperate with the big data volumes at one time, then it will not support the specific analytic modelling if users need it, added to that, processing the data can be very slow [6]. Considering all these problems, there is a need to use a specific tool that could provide an easy and clear understanding for the nature of data in different domains. Hear, the focus will be on the classification domain. In order to keep focus, here and after, the focus will be on classification techniques.

In fact, analysing data could help to improve the classification performance for neural network, by avoiding some of the existing limitations and produce better classification accuracy [7]. Classification is one of the three major components for data mining. It is the technique that used to classify each item from a set of data into a predefined class. Classification are being used widely as it is a major technique in Data Mining, where it used in various fields such as historical credit rating, employment history, number and types of investment [8], and extra.

In fact, there are many techniques that have been designed to handle different pattern classification problems, such as, decision tree which is a simple technique comparing to other but it straightforward applied to solve different classification problems [9, 10]. Another techniques example are the rule-based classifier [11], support vector machines [12], naïve Bayes classifier [13], and the artificial neural networks which mainly been used in various field such as forecasting, data compression and extra for used in engineering purposes. [14].

Even though there were many attempts to enhance the classification performance by proposing different techniques, however, the door is still open to further enhance the prediction performance. One of the ways that could really help to improve classifiers performance could be achieved through data analysing process. This kind of process normally helps researchers to understand the nature of data, identify the limitations/drawbacks, then proposed a suitable solution in order to overcome the existing limitations. Based on that, many application and tools have been proposed for data analysis, for example, Artificial Neural Network (ANN) which has emerged as one of the popular methods in tackling pattern classification problems. ANNs are useful for handling noisy data collected from real environments. Even though the number of ANN model that have been developed from the first neural model are high, the focus of this thesis is on one of the latest proposed ANN for patterns classification, which is the Fuzzy Min-Max (FMM) neural network, in order to develop a useful pattern classification model. Based on that, we are going to propose a classification. Further discussion about the FMM neural network will be introduced in chapter 2.

1.2 Problem Statement

As highlighted in the introduction, there are many data analytic tool where a user can used freely according to their preference, however, all the data analytic have their own limitations or drawbacks that leads to the new tool. Therefore, in this research, it is focusing to develop a new data analytic tool that can improve the available tool for the visualization, which by using the Fuzzy Min Max technique. The tool will display the data in the range of 0 to 1. It will also display the overlap area between the classes of data.

REFERENCES

- [1] <https://www.digitalvidya.com/blog/data-analytics-applications/> (online)
- [2] Koturwar, P., Girase, S. and, & Debajyoti, M. (2014). A Survey of Classification Techniques in the Area of Big Data. *International Journal of Advance Foundation and Research in Compute*, 1(11), 1–7.
- [3] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [4] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
<https://doi.org/10.1186/2047-2501-2-3>
- [5] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [6] Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter, 19(4)*, 1–34.
- [7] Mohammed, M. F., & Chee Peng Lim. (2015). An Enhanced Fuzzy Min–Max Neural Network for Pattern Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(3), 417–429.
<https://doi.org/10.1109/TNNLS.2014.2315214>
- [8] <https://www.digitalvidya.com/blog/data-analytics-applications/> (online)
- [9] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [11] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.

[11] Data Mining – Rule Based Classification (online)

https://www.tutorialspoint.com/data_mining/dm_rbc.htm

[12] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts ,
Decision Trees , and.

[13] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts ,
Decision Trees , and.

[14] Kesavaraj, G., & Sukumaran, S. (2013). A Study On Classification Techniques in
Data Mining.

[15] Christos Stergiou and Dimitrios Siganos (2015). Neural Networks (Online).

https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction
to neural networks

[16] [25] Simpson, P. K. (1992). Fuzzy Min-Max Neural Networks-Part, 3(5).

[17] [23] [26] Mohammed, Mohammed Falah, and Chee Peng Lim. "An Enhanced
Fuzzy Min & Max Neural Network for Pattern Classification". 2013 IEEE Transactions
on Neural Networks and Learning Systems 26.3 (2015): 417-429. Web

[18] [24] Mohammed, Mohammed Falah, and Chee Peng Lim. "Improving the Fuzzy
Min-Max Neural Network with A K-Nearest Hyperbox Expansion Rule for Pattern
Classification". Applied Soft Computing 52 (2017): 135-145. Web.

[19] Nandedkar, A. V, & Biswas, P. K. (2017). A general reflex fuzzy min-max neural network, *14*(2007), 2017.

[20] <https://intellipaat.com/blog/what-is-tableau/> (online)

[21] <http://www.learnallbi.com/what-is-qlikview/> (online)

[22] https://www.ibm.com/support/knowledgecenter/en/SSCRJU_4.0.1/com.ibm.streams.welcome.doc/doc/ibminfospherestreams-introduction.html (online)