FUZZY MODELING USING BAT ALGORITHM OPTIMIZATION METHOD FOR CLASSIFICATION

NOOR AMIDAH BINTI AHMAD SULTAN

Bachelor of Computer Science

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science in Computer Systems and Networking.

_____

(Supervisor's Signature)

Full Name   : Dr. Mohd Arfian Bin Ismail

Position   : Lecturer

Date   : 4th January 2019

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.

_____

(Student's Signature)

Full Name   : Noor Amidah Binti Ahmad Sultan

ID Number   : CA15092

Date   : 2nd January 2019

FUZZY MODELING USING BAT ALGORITHM OPTIMIZATION METHOD FOR CLASSIFICATION

NOOR AMIDAH BINTI AHMAD SULTAN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science (Computer Systems and Networking) with Honors

Faculty of Computer Systems and Software Engineering

UNIVERSITI MALAYSIA PAHANG

DECEMBER 2018

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Mohd Arfian bin Ismail for his germinal ideas, continuous encouragement, invaluable guidance and support in making this research possible. I am sincerely thankful for the time he spent proofreading and correcting my mistakes. I am extremely grateful to my beloved parent and family for keep on supporting and give advices and motivation for me in all time. To all my friends that keeps company me during all the time. Most of all, I would like to thanks to the Universiti Malaysia Pahang and Faculty of Computer System & Software Engineering for giving me the opportunity to write an honours thesis.

ABSTRAK

Pemodelan fuzzy adalah proses penjanaan parameter iaitu peraturan fuzzy dan keanggotaan. Peraturan fuzzy adalah bentuk keadaan kabur dan fungsi keahlian adalah fungsi umum penunjuk dalam set klasik. Untuk mewujudkan parameter, terdapat banyak masalah yang timbul dalam proses pemodelan fuzzy. Masalah yang timbul adalah data tidak lengkap dan saiz data adalah besar. Data yang tidak lengkap berlaku apabila beberapa data diproses hilang sehingga gagal mencatatnya. Masalah saiz data berlaku apabila data tidak dapat diproses kerana ia terlalu rumit. Untuk menyelesaikannya, kaedah Bat Algoritma dilaksanakan dalam kaedah pengoptimuman dalam pemodelan fuzzy untuk klasifikasi. Penggunaan Algoritma Bat ini untuk mencapai tujuan kajian ini. Tujuan penyelidikan ini adalah untuk mengaplikasikan dan menilai penggunann Bat Algoritma untuk klasifikasi. Untuk melakukan eksperimen dalam penyelidikan ini, terdapat beberapa dataset yang digunakan iaitu dataset WBCD, Haberman Survival dataset dan Pima India dataset. Enjin Sazanov yang merupakan enjin java fuzzy digunakan untuk memohon Bat Algoritma dalam eksperimen. Nilai parameter sudah ditetapkan untuk digunakan dalam setiap dataset dalam eksperimen. Keputusannya, ketepatan purata tertinggi dihasilkan iaitu 96.91% dengan menggunakan WBCD dataset. Ketepatan purata Bat Algoritma adalah membandingkan dengan kaedah lain. Setiap dataset menghasilkan peraturan fuzzy dan keanggotaan. Keseluruhan kajian ini, objektif kajian ini dapat dicapai.

ABSTRACT

Fuzzy modeling is a process of generating parameters which are fuzzy rule and membership function. Fuzzy rule is a form of a fuzzy condition and membership function is a generality of indicator function in classical sets. In order to create parameters, there are many problems arise in the process of fuzzy modeling. The problems are data incomplete and the size of the data is large. Data incomplete is happened when some of data to process is missing so it failed to record it. Problem of the size of data happen when data cannot be process because of it too complex. In order to solve it, Bat Algorithm method is implement in to optimization method in fuzzy modeling for classification. The study of Bat Algorithm is achieved the purpose of this research. The purpose of this research is to apply and evaluate the performance of Bat Algorithm for classification. In order conduct an experiment in this research, there are several dataset is use which are WBCD dataset, Haberman's Survival dataset and Pima Indian dataset. A Sazonov Engine which is a fuzzy java engine is use to apply Bat Algorithm in the experiment. The value of parameter is already set to use when applying every dataset in an experiment. As a result, the highest average accuracy generated which is 96.91% by using WBCD dataset. The average accuracy of Bat Algorithm is comparing with other methods. Every dataset is producing the best fuzzy rule and membership function. Overall of this research, the objectives of this research is achieved.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

ANN   Neural Network

BPNN   Back Propagational Neural Network

SVM   Support Vector Machine

LIST OF ABBREVIATIONS

ANN   Neural Network

BPNN   Back Propagational Neural Network

SVM   Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1   Background of Study

Nowadays, there are most issues arise in fuzzy systems are interpretability and robustness (Pogio, Smale, 2003). These two issues have an effected on traditional data-driven approaches. Interpretability is deals with the relation occurring between two heterogenous entities which are fuzzy system and human. Meanwhile, robustness is generalization capability and numerical stability of the identified model. These interpretability and robustness are main advantages to fuzzy systems (Guglielmann, 2008).

Classification is a process of grouping some of an elements. Hence, some of problem in classification arise which are inaccurate data and incomplete data. Due to these problem, fuzzy system is the best method to use as a solution (Gadaras and Mikhailov, 2009). The fuzzy system has two parameters which are fuzzy rules and fuzzy membership function. Fuzzy rules is a form of a conditional statement which is *IF x is A, THEN y is B* where *x* and *y* are linguistic variables (Dubois, Prade, 1996) and membership function is a generality of indicator function in classical sets. Fuzzy parameters is needed in order to perform its operation. Before fuzzy parameters are use, the identification of fuzzy parameters is needed and it known as fuzzy modeling.

Process to generates parameter is called fuzzy modeling. Problems that arise in fuzzy modeling are large data and incomplete of data. By applying optimization method in fuzzy modeling, these issues can be solve. In this research, fuzzy modeling using Bat Algorithm for classifcation was presented. The Bat Algorithm is one of the method that

is used for optimization in fuzzy modeling. The Bat algorithm is a nature-based algorithm based on prey-tracking behaviour that proposed by Yang in 2010.

1.2 Problem Statement

In this research, most of the problem arise in fuzzy modeling are interpretability and robustness. Table 1.1 shows summary of problem which are have been carried out.

Table 1.1   Summary of Problem Statement

No	Problem	Description	Effect
1.	Limitation of data	Data limit will make the classification process of data is complicated	This problem will effect on creating a fuzzy modeling
2.	Identification of many fuzzy parameters (Takagi & Sugeno, 1985)	Different size of data will use by the system	Fuzzy rules and fuzzy membership function become complex when involve large data and complex system

1.3 Goal and Objective of Research

The goal of this research is to optimize the fuzzy modeling by using Bat Algorithm method for classification problem. In order to achieve the goal, several of objectives are required to be made. Below are list of objectives in this study:

i.   To study Bat algorithm for fuzzy modeling.
ii.   To apply Bat algorithm method for fuzzy modeling.

iii.   To evaluate the performance of the Bat Algorithm method using classification problem.

1.4   Scope of Research

The scope this research can be listed as follows :

i.   The scope of research focus on classification problem.

ii.   The optimization method that will be used is Bat algorithm.

iii.   Using datasets from WBCD dataset, Haberman's Survival dataset and Pima Indian dataset.

1.5   Significance of Research

The significance of this research are :

i.   Process of altering and adjusting fuzzy model can perform automatically.

ii.   Many engineering problems can be solve by its efficiency and robust.

iii.   Increase the global search capability of Bat Algorithm.

1.6   Research Organization

This research consists of five chapters which are introduction, literature review, methodology, implementation and conclusion.

i.   Chapter 1: Introduction. This chapter about an overview of the research which is consist of introduction, problem statements, goals and objectives, scopes, significance and research of organization.

ii.   Chapter 2: Literature review. This chapter about discussion of a few of related research papers. All the relevant books or journal was being discussed.

iii.   Chapter 3: Methodology. This chapter about the method that was be used in conducting this research and discuss the flow process of the research.

REFERENCES

Aalaei, S., Shahraki, H., Rowhanimanesh, A., & Eslami, S. (n.d.). Feature selection using genetic algorithm for breast cancer diagnosis : experiment on three different datasets, (6).

Agarwal, P., & Mehta, S. (2014). Nature-Inspired Algorithms: State-of-Art, Problems and Prospects. *International Journal of Computer Applications*, *100*(14), 14–21. https://doi.org/10.5120/17593-8331

Babuška, R., & Verbruggen, H. B. (1996). An overview of fuzzy modeling for control. *Control Engineering Practice*, *4*(11), 1593–1606. https://doi.org/10.1016/0967-0661(96)00175-X

Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1882–1889. https://doi.org/10.1021/ci0341161

Chanas, S., & Kuchta, D. (1996). A concept of the optimal solution of the transportation problem with fuzzy cost coefficients. *Fuzzy Sets and Systems*, *82*(3), 299–305. https://doi.org/10.1016/0165-0114(95)00278-2

Cuevas, E., & Reyna-orta, A. (2014). A Cuckoo Search Algorithm for Multimodal Optimization, *2014*.

Fallahnezhad, M., Hassan, M., & Zaferanlouei, S. (n.d.). Author â€™ s personal copy A Hybrid Higher Order Neural Classifier for handling classification problems. https://doi.org/10.1016/j.eswa.2010.06.077

Gigras, Y., Gupta, K., & Choudhary, K. (2015). A Comparison between Bat Algorithm and Cuckoo Search for Path Planning, 4459–4466. https://doi.org/10.15680/ijircce.2015.0305067

Hayashi, Y., & Nakano, S. (2016). Informatics in Medicine Unlocked Use of a Recursive-Rule eXtraction algorithm with J48graft to achieve highly accurate and concise rule extraction from a large breast cancer dataset. *Informatics in Medicine Unlocked*, *1*(2015), 9–16. https://doi.org/10.1016/j.imu.2015.12.002

Luan, C., & Dong, G. (n.d.). Experimental Identification of Hard Data Sets for Classification and Feature Selection Methods with Insights on Method Selection.

Pan, F., Li, X., Long, T., Hu, X., Ren, T., & Du, J. (2013). Novel magnetic fi eld computation model in pattern classi fi cation, (May 2016). https://doi.org/10.1109/JSEE.2013.00100

Rahman, J., Suri, H. S., & Abedin, M. (2018). Accurate Diabetes Risk Stratification Using

Machine Learning : Role of Missing Value and Outliers, 1–17.

Rodrigues, L. (2016). Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection, (October 2015).

Saad, A., Dong, Z., & Karimi, M. (2017). A Comparative Study on Recently-Introduced Nature-Based Global Optimization Methods in Complex Mechanical System Design. *Algorithms*, *10*(4), 120. https://doi.org/10.3390/a10040120

Sabin, D., & Freuder, E. C. (1999). Optimization methods for constraint resource problems. *AAAI Workshop on Configuration*, 83–89.

Testing, N., Sazonov, E., & Halabe, U. B. (2003). Fuzzy Logic Expert System For Automated Damage Detection From Changes In Strain Energy Mode Shapes FUZZY LOGIC EXPERT SYSTEM FOR AUTOMATED DAMAGE DETECTION FROM CHANGES IN STRAIN, (November 2012). https://doi.org/10.1080/10589750290026556

Toan, T. Q., Sorokin, A. A., Thi, V., & Trang, H. (2018). Fuzzy system for controlling queue size of packets in telecommunication nodes, (1).

Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Informatics in Medicine Unlocked Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, *10*(December 2017), 100–107. https://doi.org/10.1016/j.imu.2017.12.006

Ya, C., & Uriarte-arcia, A. V. (2014). One-Hot Vector Hybrid Associative Classifier for Medical Data Classification, *9*(4), 1–13. https://doi.org/10.1371/journal.pone.0095715

Yang, X. S., & He, X. (2013). Bat algorithm: literature review and applications. *International Journal of Bio-Inspired Computation*, *5*(3), 141. https://doi.org/10.1504/IJBIC.2013.055093

Ye, N., Roosta-Khorasani, F., & Cui, T. (2017). Optimization Methods for Inverse Problems, 1–13. Retrieved from http://arxiv.org/abs/1712.00154

Zhao, J. (2016). The bat algorithm and its parameters The bat algorithm and its parameters, (June). https://doi.org/10.1201/b18592-237