UNIVERSITI PUTRA MALAYSIA

INTEGRATING TYPE-2 FUZZY LOGIC SYSTEM WITH FUZZY C-MEANS CLUSTERING FOR WEATHER PREDICTION

AHMAD SHAHI SOOZAEI

FSKTM 2011 30

INTEGRATING TYPE-2 FUZZY LOGIC SYSTEM WITH FUZZY C-MEANS CLUSTERING FOR WEATHER PREDICTION

By

AHMAD SHAHI SOOZAEI

Thesis submitted to the School of Graduate Studies, Universiti Putra Malaysia, in fulfilment of the Requirements for the Degree of Master of Science

February 2011

DEDICATION

To

My Beloved Mother and Brother,

And

The Soul of My Father

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

INTEGRATING TYPE-2 FUZZY LOGIC SYSTEM WITH FUZZY C-MEANS CLUSTERING FOR WEATHER PREDICTION

By

AHMAD SHAHI SOOZAEI

February 2011

Chairman: Rodziah binti Atan, PhD

Faculty: Computer Science and Information Technology

Today's world concerns more about the impact of weather in the development of society. An accurate weather prediction system can act in as a vital role for making crucial decision on the life and property issues. In this research, the temperature attributes of weather is considered. Accurate weather temperature prediction can be achieved with regards to the quality of data processed. Fundamentally, weather prediction is complex due to the heterogeneous and enormous data. There are also factors such as outliers, noise and overlapped data which cause an increase of uncertainty in the data. Therefore, the assurance of data quality is associated by isolating these uncertainties factors. The quality of data is foreseen to increase the accuracy of prediction. However, most researchers in this domain do not consider the importance of data quality in their researches. In the existing prediction methods, Type-2 fuzzy logic is the proper method to deal with the uncertainty. In fuzzy systems, the relation between

uncertainty of input data and fuzziness is expressed by membership functions. However, if the regions of the data of different classes are highly overlapping or contain noise and outliers, the value of membership function will be misleading. This effect is known as the membership un-robustness. Furthermore, the result or decision produced will not be accurate and lead to false prediction. Thus, overlapped data and uncertainty are two important issues which affect the quality of data.

In this thesis, a method is proposed to predict next temperature value with high accuracy. The proposed method is based on combination of statistic equation with Fuzzy C-Mean (FCM) clustering and Type-2 fuzzy logic system (Type-2 FLS) with gradient descent algorithm. The statistic equation with FCM can be applied to handle outliers and cluster desired data and gradient descent in Type-2 FLS is utilized to tune the membership function parameters. Another feature of the proposed method is improvement in the performance time (run time) by clustering the desired data.

The proposed method has been validated by experiments using Italy and New York weather temperature dataset. The findings show that the accuracy of this method for prediction next value increased as compared to base method. The accuracy percentage of proposed method on the Italy dataset was found to increase accuracy up to 89.6%. For New York dataset, the proposed method was found to increase accuracy up to 91% as compared to 67% by the base method. The performance time of the proposed method has improved 52% and 49% in comparison to base method for Italy and New York dataset respectively. The results prove that the proposed method is more efficient than base method in accuracy and performance time basis.

Abstrak tesis yang dikemukakan kep ada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

INTEGRASI SISTEM LOGIK KABUR TYPE-2 DENGAN PENGGUGUSAN KABUR C-MEANS BAGI RAMALAN CUACA

Oleh

AHMAD SHAHI SOOZAEI

Februari 2011

Pengerusi: Rodziah binti Atan, PhD

Fakulti: Sains Komputer dan Teknologi Maklumat

Dunia hari ini mengambil berat tentang kesan cuaca terhadap pembangunan masyarakat. Sistem ramalan cuaca yang tepat boleh memainkan peranan penting dalam pembikinan keputusan yang rumit berkaitan dengan isu kehidupan dan harta benda. Di dalam kajian ini, atribut suhu dalam cuaca diambil kira. Ramalan suhu cuaca yang tepat boleh dicapai bergantung kepada kualiti data yang diproses. Secara asas, ramalan cuaca adalah kompleks disebabkan oleh kepelbagaian serta saiz data yang sangat besar. Terdapat juga faktor seperti unsur luaran, gangguan dan data bertindan yang menyebabkan peningkatan ketidak-tentuan data. Oleh itu, memastikan kualiti data adalah berhubung rapat dengan pengasingan faktor tidak-tentu ini. Kualiti data dilihat mampu untuk meningkatkan ketepatan ramalan. Walaubagaimana pun, kebanyakan penyelidik dalam domain ini tidak mengambil kira kepentingan kualiti data dalam penyelidikan mereka. Dalam kaedah ramalan sedia ada, logik kabur Type-2 adalah kaedah yang sesuai bagi mengendalikan ketidak-tentuan. Dalam sistem kabur, perkaitan antara input data dan kekaburan dinyatakan oleh fungsi keahlian. Namun begitu, jika kawasan data bagi kelas

yang berbeza terlalu bertindan atau mengandungi gangguan dan unsur luaran, nilai bagi fungsi keahlian akan tersasar. Kesan ini dikenali sebagai keahlian tidak kukuh. Tambahan lagi, hasil atau keputusan yang dikeluarkan akan menjadi tidak tepat dan menjurus kepada ramalan palsu. Jadi, data bertindan dan ketidak-tentuan adalah dua isu penting yang memberi kesan terhadap kualiti data.

Dalam tesis ini, satu kaedah dicadangkan untuk meramal nilai suhu seterusnya berketepatan tinggi. Kaedah yang dicadangkan adalah kombinasi persamaan statistik dengan penggugusan *Fuzzy C-Mean* (FCM) serta sistem logik kabur *(Type-2 FLS)* dengan algoritma pengurangan berkala. Persamaan statistik dan FCM diaplikasi bagi mengendali unsur luaran/gangguan dan penggugusan data diingini dan pengurangan berkala Type-2 FLS bagi penalaan parameter fungsi keahlian. Ciri lain kaedah cadangan ini adalah penambah baikan masa pelaksanaan (masa larian) dengan menggugus data yang dikehendaki.

Kaedah cadangan telah disahkan melalui uji kaji menggunakan set data suhu cuaca Italy dan New York. Penemuan menunjukkan peningkatan pencapaian pada kaedah ini berbanding kaedah asas. Peratusan ketepatan bagi kaedah cadangan untuk set data Italy didapati meningkat sebanyak 89.6%. Bagi set data New York, kaedah cadangan telah meningkatkan ketepatan sehingga 91% dibandingkan dengan 67% bagi kaedah asas. Masa pelaksanaan bagi kaedah cadangan telah meningkat 52% dan 49% lebih pantas bagi set data Italy dan New York berbanding dengan kaedah asas. Hasil kajian membuktikan bahawa kaedah yang dicadangkan adalah lebih berkesan dibandingkan dengan kaedah asas bagi ketepatan dan masa pelaksanaan.

ACKNOWLEDGEMENTS

First and foremost, Alhamdulillah for giving me the strength, patience, courage and determination in completing this work. All grace and thanks belongs to Almighty Allah.

I would like to express my deep gratitude to my mother and brother for providing me the opportunity to continue my master's program and financial support. In addition, I am grateful to my supervisor Dr. Rodziah binti Atan for her kind assistance, critical advice, encouragement and suggestions during the study and preparation of this thesis. Moreover, I appreciate her encouragement to provide the opportunity to attend for conferences and submit to journals. I truly appreciate the time she devoted in advising me and showing me the proper directions to continue this research and for her openness, honesty and sincerity.

I would also like to express my gratitude to my co-supervisor Associate Professor Dr. Md. Nasir Sulaiman for his kind assistance and important advice, to whom I am grateful for his practical experience and knowledge that made an invaluable contribution to this thesis.

Finally, I would like to extend my gratitude to the Dean and members of FSKTM for their endless support and facility they provided for Master and PhD research students.

APPROVAL

I certify that an Examination Committee has met on date of viva to conduct the final examination of **Ahmad Shahi Soozaei** on his thesis entitled "**INTEGRATING TYPE-2 FUZZY LOGIC SYSTEM WITH FUZZY C-MEANS CLUSTERING FOR WEATHER PREDICTION** " in accordance with Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Examination Committee are as follows:

Mohamed Othman, PhD
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hamidah Ibrahim, PhD
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Norwati Mustapha, PhD
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(External Examiner)

Muhammad Suzuri Hitam, PhD
Faculty of Science and Technology
Universiti Malaysia Trengganu
(External Examiner)

_____

NORITAH OMAR, PhD
Assoc. Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirements for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Rodziah binti Atan, PhD
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Md. Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

HASANAH MOHD GHAZALI , PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously and is not concurrently submitted for any other degree at Universiti Putra Malaysia or at any institution.

AHMAD SHAHI SOOZAEI

Date: 2 February 2011

TABLE OF CONTENTS