AN EFFECTIVE FUZZY C-MEAN AND TYPE-2 FUZZY LOGIC FOR WEATHER FORECASTING

[1]AHMAD SHAHI, [2]RODZIAH BINTI ATAN, [3]MD. NASIR SULAIMAN

[1]Master Student, Department of Computer Science and IT, UPM University, Malaysia, 43400.

[2]Dr., Department of Computer Science and IT, UPM University, Malaysia, 43400.

[3]Assoc.Prof., Department of Computer Science and IT, UPM University, Malaysia, 43400.
E-mail: shahi@ieee.org, rodziah@fsktm.upm.edu.my, nasir@fsktm.upm.edu.my

ABSTRACT

Meteorological forecasting is applicable for versatile applications. Accurate weather prediction saves lives, money and time in both local and global area. Forecasting accuracy is still not accurate because of the uncertain (fuzzy) data of nature, due to several reasons including: incomplete data, hand writing error, device error, precision of measurements and discreet description of connective phenomena Inherent part reflecting our understanding of things. On the other hand in global area with large amount of data to process whole the data is time consuming, thus, to improve the quality of data and execution time, we need to manage the uncertainty of data and extract desired data. Therefore the uncertainty management and process the data demand intelligent methods with knowledge based approaches. This paper reviews challenges in this field and compares advantages and drawbacks of the existing methods that essentially are only applicable for local area. Finally we proposed a hybrid technique for new research based on fuzzy c-mean clustering technique and type-2 fuzzy logic that is useable in both local and global area. Finally we show our experiments and prove that hybrid technique performs better than existing weather prediction methods in low error rate.

Keywords: *Type-2 Fuzzy Logic, ANN, Fuzzy C-Mean Clustering, Nero-Fuzzy, Markov Fourier, CBR.*

1. INTRODUCTION

Meteorological forecasting is an important issue in research. Typically, the forecasting is performed at “global level”, by gathering data in a large geographical region [3].Weather prediction can have many different forms, depending on required applications. For example in airport it is far more important to know about the climate visibility a few hours ahead rather than temperature. Also in buildings, weather predictions can play an important role in saving the energy. This can be done by controlling the heat and cool factors of buildings [5]. On the other hand, data uncertainty plays a special role in these environment and such data are not perfect from many reasons: incomplete data, precision of measurements discreet description of connective phenomena, inherent part reflecting our understanding of things, keyboard error, hand writing error, device error [6].

However Meteorological data are uncertain (fuzzy) in nature and Information on weather is vaguely defined [10]. Uncertain data is consisting of noises and outliers that decrease quality of data. Noise is a random error or a parasite that comes from sensor network, error hand writing, and keyword error writing and so on. Noise is invalid value but outlier is valid it means, An outlier is an observation of the data that deviates from other observations so much that it arouses suspicions that it was generated by a different mechanism from the most part of data [23].however by removing noise and detecting outliers we can get high quality of data. On the other hand, weather prediction system is developed based on the data set, therefore to get accurate model based on the dataset we should improve the quality of dataset to get true result. Consequently in this paper we reviewed several methods which have been focusing in this area (weather prediction) so far.

These methods are not considering noise and outlier parameters and only applicable to local area. Moreover, global area prediction by these methods are not possible, because large time consuming process. Section 2 overviews the current techniques

in weather prediction. The rest of paper is organized as follows:
Section 3 analyzes the characteristics several of weather prediction methods. We described in Table 1 formal for quick review. Section 4 describes our approach based on Fuzzy C-Mean clustering and type-2 fuzzy logic system called hybrid method. Finally in section 5 we present results and discussion. Section 6 concludes the paper.

2. LITERATURE REVIEW

2.1 Numerical Weather Prediction (NWP)

History of numerical weather prediction, since the groundbreaking work of V. Bjerknes (1904) and LF. Richardson (1922), the challenge of weather forecasting has been related to an initial value conditions problem of mathematical physics (based on the non-linear equations governing fluid flow) and has been approached using numerical means. The success of the first numerical prediction by Charney, Fjortoft and von Neumann (1950) launched the unprecedented trend of recent decades in scientific research on earth sciences. It is also considered that Dr. Robert's over all contributions to the efficiency of NWP model in term of computer time resource is an order of magnitude, hence arguably equivalent to one to two generations of High Performance Computing (HPC) system. Numerical weather prediction models currently rely increasingly on High Performance Computing (HPC) systems that will soon make it possible to expand and significantly improve meteorological computer applications to an ever greater extent. The impact of extreme weather phenomena on society is growing more and more costly, causing infrastructure damage, injury and the loss of life. One aim of meteorological research is to limit the impact of extreme summer and winter weather through improved detection and through faster and more accurate prediction techniques [9].

Canadian meteorological scientists that use very high-resolution simulation of severe weather precipitation systems to prepare future NWP operational applications need to seek high performance computer resource outside the country. Internationally, the increasing demand for accurate weather and environmental prediction has led to significant attention being given to investments in numerical weather prediction and HPC system. The roadmap in Weather and Environmental Numerical Prediction will demand to train researchers with high skill in computationally intensive problems and with sustained access to significant high performance computing in Canada [9].

2.1.1. Strengths and issues of Numerical Weather Prediction

The advantages of Numerical Weather Prediction (NWP) to achieve very high-resolution simulation of severe weather precipitation systems to prepare future NWP operational applications and increase high performance computer resource and get high quality and accurate results[9].

Drawbacks of NWP:

- As mentioned above, numerical weather prediction needs high performance computing and memory space to get more accuracy [9].
- Numerical method has been investigated as the common method to forecast weather which involves a complex of mathematical computing [2].
- Traditional weather forecast by simulating the atmosphere using systems of mathematical equations which is known as Numerical Weather Prediction (NWP) is not good enough to create an optimal controller for the heating and cooling of buildings. That is why a predictor adjusted to the needs for such an engineering use, is to be created [5].

2.2. Markov-Fourier Gray Model

Traditional prediction approaches are either based only on global prediction or only on local prediction. From simulation results, it can be seen that local prediction schemes usually have better performance in predicting time series than global prediction schemes do. In this paper Ref.[8] has proposed an approach that can incorporate both types of prediction to increase prediction accuracy. The proposed Markov–Fourier gray model (MFGM) prediction approach uses a gray model to roughly predict the next datum from a set of the most recent data and a Fourier series to fit the residual errors produced by the gray model. Finally, Markov state transition matrices are employed to recode the global information generated also by the residual errors. By combining a local predicted value obtained by a Fourier series and a global estimated error obtained by the Markov forecasting method, the approach can predict the future weather more accurately [8].

2.2.1. Strengths and issues of Markov-Fourier Gray Model

The simulation results in Ref.[8] showed that this approach could predict the future more accurately than the other methods did. Besides, if the Markov matrices are only used to record short-term information, even though the prediction accuracy can be improved, it still cannot outperform the global Markov matrices. In other words, the global Markov matrices indeed can record global information to enhance the prediction capability. Finally, it should be noted that if Markov–Fourier gray model is used to predict white noise, the Markov correction mechanism might actually degrade the prediction performance [8].

2.3. Artificial Neural Networks (ANNs)

In Malaysia, numerical method is the commonly used method to forecast weather which involves a complex mathematical computing. The models used in forecasting are supplied by other counties such as Europe and Japan. The goal of Ref. [2] method is to forecast weather using another technology known as artificial neural network. This system is capable to learn the pattern of rainfall in order to produce a precise forecasting result. The supervised learning technique is used in the learning process [2].

James W. Taylor and Roberto Buizza [7] has investigated the use of weather ensemble predictions in the application of ANNs to load forecasting for lead times from one to ten days ahead. A weather ensemble prediction consists of multiple scenarios for a weather variable. In addition uses these scenarios to produce multiple scenarios for load. Besides considered, using of these new weather forecasts in the nonlinear modeling environment of ANNs.
The results show that the average of the load scenarios is a more accurate load forecast than the produced using traditional weather forecasts [7].

2.3.1. Strengths and issues of ANN

-In contrast with numerical weather prediction, artificial neural network decrease complexity of mathematical and computing and increase accuracy.
-The back propagation algorithm has been used in artificial neural network has a number of inherited Limitations:

1) Need heavy computational requirements.
2) Non existence of ANN design methodologies for deciding the value of learning rate and momentum.

Many researches and experiments done in weather prediction with neural networks hint that a forecast can be done with at least as good accuracy as the models that are used today. But the reliability issue of a trained network is still in doubt [2].

2.4. ANN and RPROP with dynamic Tunneling

A short-term temperature forecasting method based on artificial neural networks has been presented in Ref. [1]. An improved learning algorithm of neural network, RPROP, combined with a new efficient computational technique, dynamic tunneling technique is used to train neural network, for short, GDT. These two techniques are repeated alternatively processed to avoid local minima and result into a global optimization. The proposed networks are trained with actual data of the past 24 months (1999- 2000) and are tested with data of 6 months (2001.1~2001.3,2001.7~2001.9), which come from several meteorological stations around or in Chongqing, China [1].

In AI systems, especially artificial neural networks have been used to predict average temperature in some weather forecasting products, in which standard back-propagation (BP) is applied generally. Although BP training has proved to be efficient in a variety of applications, its convergence tends to be very slow and often yields sub-optimal local solutions [25][1].

2.4.1 Strengths and issues

ANN consisted of a gradient descent technique called RPROP with dynamic tunneling technique. Giving faster convergence and global optimization [3], when compare with standard Back Propagation (BP), the method (RPROP) leads to faster convergence obviously and obtain a closer result to fact, but reliability issue of a trained network is still in doubt [1].

2.5. Nero-Fuzzy Logic System

Nero-fuzzy Logic system has been using for the control of heating and cooling of buildings. A weather predictor with and without limited knowledge is to be developed. In Ref. [5] has been emphasized on the summer data and thus on the control of cooling of buildings. The problem became to some extent pattern recognition and regression task, hoping that these patterns revealed in the past information will keep repeating again in the future. Weather data is available and artificial neural networks and fuzzy logic systems are selected to be used amongst the several methods for creating models able to predict. The weather can portrayed as a vague concept a combination of the effectiveness of ANNs and the vagueness of fuzzy systems might prove more robust [5].

2.5.1 Strengths and issues on Nero-Fuzzy

The results of this model built with the combination of fuzzy system and ANN showed little improvement. One can say that the ANN with the specific architecture (which is the most widely used in regression problems) cannot improve anymore. However results did show a slight improvement. However we don't want to concentrate on this method because it is useful for local data [5]. By using Nero, a great disadvantage of ANNs is:

- Referred as "black boxes" because it is usually not possible to extract explicit knowledge from them [11] [12].
- The exact analysis of ANNs is rather difficult because of the complexity of the network mapping and the nondeterministic nature of many successfully completed training producer. hence the size of a hidden layer is usually determined experimentally and by empirical guidelines[12].

2.6 Fuzzy logic and Case-Base Reasoning

A fuzzy logic based methodology for knowledge acquisition is developed and used for retrieval of temporal cases in a case-based reasoning (CBR) system. The methodology is used to acquire knowledge about what salient features of continuous-vector, unique temporal cases indicate significant similarity between cases. Such knowledge is encoded in a similarity-measuring function and thereby used to retrieve k nearest neighbors (k-nn) from a large database. CBR is recommended to developers who are challenged to reduce the knowledge acquisition task, avoid repeating mistakes made in the past, reason in domains that have not been fully understood or modeled, learn over time, reason with incomplete or imprecise data and concepts, provide a means of explanation, and reflect human reasoning and these are some of the challenges faced by developers of weather forecasting systems [4].

A fuzzy k-nn based prediction system, called WIND-1, is proposed, implemented, and tested. Its unique Component is an expertly-tuned fuzzy k-nn algorithm with a temporal dimension. This algorithm is tested with the problem of producing 6-hour predictions of cloud ceiling and visibility at an airport, given a database of over 300,000 consecutive hourly airport weather observations (36 years of record). Its prediction accuracy is measured with standard meteorological statistics and compared to a benchmark prediction technique, persistence climatology (PC). In realistic simulations, WIND-1 is significantly more accurate. It takes about one minute for WIND-1 to produce a forecast [4].

2.6.1 Strengths and issues

The advantages of combination of fuzzy logic and case-based reasoning reduce the knowledge acquisition task and avoid repeating mistake made in past and also use common words to directly acquire domain knowledge about feature salience. It is also has drawbacks of having not fully understood or modeled, learn over time reason with incomplete and or imprecise data and concepts [4].

2.7. Fuzzy logic and Clustering Analysis

A new related coefficient comparison method is introduced to categorize the variables that have influence on electric load into some clustering groups, and their membership functions are set as input variables in the form of natural language using fuzzy set theory. The history data are clustered into output variables through the adaptive neuron-fuzzy inference system ANFIS so as to ensure the minimal errors of their member functions. The rules that link the input variables with output variables are set on the basis of practice data and expert knowledge. This method (clustering) is an effective tool to deal with some special variables such as monthly variables, weekly variables and hourly variables and efficiently closed to the actual load forecasting in practice [13].

2.7.1. Strengths and issues

Adaptive Nero-Fuzzy Inference System (ANFIS) is used to ensure the minimal errors of their membership function and design membership functions [13]. Therefore, by utilizing ANFIS in fuzzy logic it has drawbacks of neural network as explained in section 2.5.1.

2.8 .Type-2 Fuzzy logic system

The local meteorological forecasting an local level and historical data has been done by type-2 fuzzy Logic system.. A classical Fuzzy logic system (FLS), also denoted as Type-1 FLS, rules play a central role in the FLS framework. Rules can be provided by experts or can be extracted from numerical data [3].type-1 fuzzy logic has been done in deferent applications [14, 15, 16, 17, 18, 19] especially meteorological application [10, 20, 21] [3].

The disadvantage of Type-1 FLS does not succeed in handling the intrinsic uncertainty that the data hold, thus provoking a local drift of the short-term predicted values whereas a notable error in long-term prediction and is sensitive to noise and outlier as well.

The concept of embedded Type-1 FLS in Type-2 FLS [4] deals with the capability of Type-2 FLS of considering and handling the dispersion of the values of the measurements around a mean, easily predictable data. Using Type-2 FLS is also very important because this class of systems has shown small sensitivity to measurements uncertainty and noise. The whole system has been designed and simulated on historical meteorological data previously acquired. This system allows to achieve more accurate forecasting results even many hours in advance. The study of the evolution of this data permits to foresee the meteorological situation in a certain place. Because of the large amount of data to be processed, this procedure is a time consuming and needs very powerful computers. It is still sensitive to noise and less than type-1 fuzzy logic, however [3].

2.8.1 Strengths and issues of type-2 fuzzy logic System

In contrast with type-1 fuzzy logic, type-2 has small sensitivity to noise and suitable to take into account non-stationary noise both in training and operating phase, however, type-2 fuzzy logic has still sensitivity to measurements uncertainty and noise and the big drawbacks of type-2 fuzzy logic system is processing a large amount of data is time-consuming and needs very powerful computer [3].

3. COMPARATIVE ANALYSIS

From above discussion, the existing weather prediction techniques are only support to small scale of data and applied only to static environment with use of historical data. In other words, these methods are not applicable to real time data in dynamic environments.

In Table 1, we presented quick view of these methods with their contribution in order to improve accuracy or performance for meteorological forecasting.

4. PROPOSED HYBRID METHOD

As mentioned in previous section, these methods are vulnerable to noise, computation complexity or large time overhead.

By analyzing results of the Table 1, type-2 fuzzy logic system [3] is able to manage the uncertainty in weather forecasting. But this system is vulnerable to noise and outliers. Moreover to process a large amount of data needs high performance computing and large time consuming.

We proposed a hybrid method using “Fuzzy c-mean clustering and Type-2 fuzzy logic system” to overcome above difficulties.

Among artificial intelligent techniques, type-2 fuzzy logic system is suitable technique that can manage the uncertainty in weather forecasting [24].Type-2 FLS has been done by Ref. [3].Fuzzy C-Mean clustering that is one of data mining technique is utilized with Type-2 FLS to detect outliers and cluster desired data to improve accuracy and efficiency. But in this paper we have investigated the accuracy of hybrid method and shown in section 5.

Tabel1: Comparative analysis of methods in Meteorological forecasting

Researcher(s)	Method(s)	Contribution	Data Collection	Advantage(s)	Disadvantage (s)	Result(s)
Gilbert Brunet	NWP	R&D Activities and increasing memory space	Vancouver Island and Oklahoma	Increasing HPC performance, quality and accuracy	-complexity of mathematical computing. -no mentioned the percentage of accuracy	Accuracy and performance
Abdul Manan, Chia Su C`huan	ANNs	Supervised learning technique	Area of Senai,Johor,Malaysia. Data input: time of day, pressure, dry bulb temperature, wet bulb temperature, dew point, wind speed, wind direction, cloud cover, whether it rained in past hour	Decrease Complexity of mathematical computing and increase accuracy	Small sensitive to noise/precision	Accuracy: (78%)
Zheng Qin,Haoliang wang	ANNs	A gradient descent technique called RPROP with dynamic tunneling technique.	Trained with actual Data of past 24 months(1999-2000)and are tested with data of 6 months(2001.1~2001.3,2 001.7~2001.9) data set are from (Chongqing, China)	Giving faster convergence and global optimization	reliability issue of a trained network is still in doubt	Closer result to the fact(satisfied result)
Alexandros Mountis and Geoff Levermore	Fuzzy logic	ANNs	The historical data consis ted of hourly collection of Manchester International Airport's weather measurement for the years 1982-1995 examines only on summer data	The accuracy of fuzzy system depend on how well its parameters are defined and parameters is adjusted by NN	-ANNs are referred as "Black box", because it is usually not possible to extract explicit knowledge from them. -Complexity networks and nondeterministic nature of many successfully completed training procedures.	-Provide satisfactory result. -Mean absolute error(MAE): MAE(nero-fuzzy):1.7096 MAE(ANN): 1.8517 MAE(Fuzzy):3.0306.
Bjarne K.hansen and Denis Rordan	Fuzzy set theory	Case base reasoning (k-nn method)	-It is tested with problem of producing 6-hour predictions of cloud ceiling and visibility at an airport -database of over 300,000 consecutive hourly airport weather observation(36 years of records)	-reduce the knowledge acquisition task. -avoid repeating mistake made in past. -Use common words to directly acquire domain knowledge about feature salience.	Have not fully understood or modeled ,learn over time, reason with incomplete and or imprecise data and concepts.	Airport weather predictions More accurate which will make air travel safer and make airlines more profitable.
WEI DING,FU-GUI DONG, SHANG –DONG YANG	Fuzzy Logic	-Clustering technique -ANFIS	-----------------	-Effective tool to deal with some special variable -using ANFIS to ensure the minimal errors of their membership function and design MFs	-Drawbacks of NN -sensitive to the noise	To get more efficiency and closed to actual load forecasting in practice.
A.Mencattini,M.Salmeri	Fuzzy Logic System	Type-2 Fuzzy (type-reduction)	-About temperature and humidity that are collected from by the Neuronica Lab at Politecnico of Torino(Italy) -The measurement have been performed every 900 s,that is quarter of Hour.	Suitable to take into account non-stationary noise (small sensitive to noise) both in training and operating phases.	-Still has sensitivity to measurements uncertainty and noise. -processing a large amount of data is time-consuming and needs very powerful computer.	Achieve more accurate forecasting results even many hours in advance. Humidity MAE: 0.3380 Temperature MAE:0.0822
Shun-Feng Su, *Member*, *IEEE*, Chan-Ben Lin, and Yen-Tseng Hsu	Markov–Fourier gray Model (MFGM)	-----------------	------------------	- MFGM can provide the best performance among existing prediction schemes. -global information encoded in the Markov matrices indeed can provide useful information for predictions.	-Markov matrices are only used to record short-term information, even though the prediction accuracy can be improved, it still cannot outperform the global Markov matrices.	Examplel: Smooth function MAE:0.003 Example2 : Stock Market MAE:15.7 Example3: Time series MAE:0.0158

4.1. Methodology

The methodology consists of four stages of processes to increase the accuracy of uncertain data and improve efficiency:

1) In first stage, getting dataset form devices (sensor, satellite, etc) that are unclean data.
2) In next stage the preprocessing on to the data to clean the data (i.e. missing value, inconsistent data and noise).
3) Third step, due to outliers in dataset and vast amount of data, we will utilize c-mean clustering algorithm to detect anomaly data (outlier) to increase precision of prediction and cluster data based on their similarities to enhance performance.
4) Finally the use of type-2 fuzzy logic system on cleaned dataset and define its methods (i.e. Gaussian membership function) to get correct decision for data and get best result for prediction.

The numbered processes have been charted in Fig.1 and will be further explained in the next sections for a more detailed description.

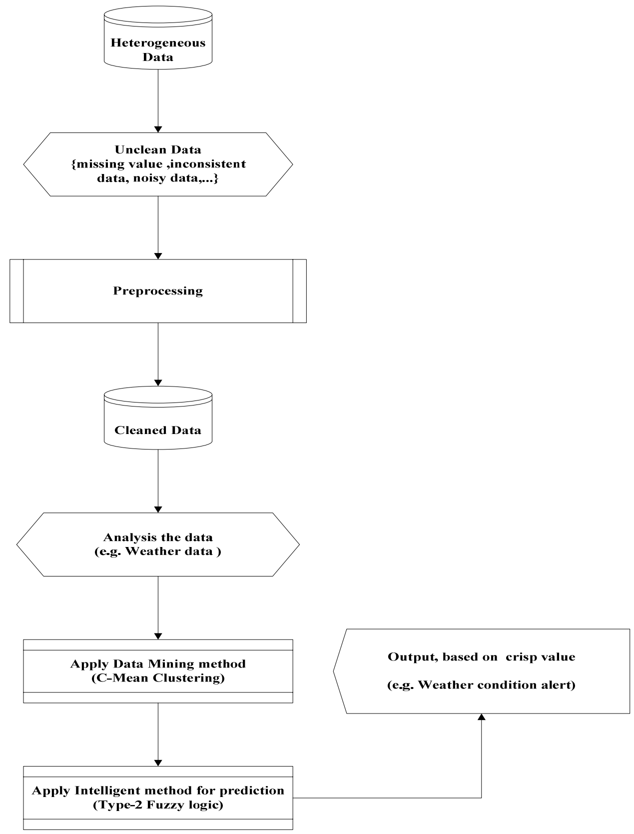

Fig.1. Flowchart of Methodology

4.2. Preprocessing

Today's real-world databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size (often several gigabytes or more) and their likely origin from multiple, heterogeneous sources [8]. Low-quality data will lead to low-quality mining results. Therefore we should utilize preprocessing in order to help improve the quality of the data and, the efficiency and ease of the mining process. There are a number of data preprocessing techniques that has been shown in Fig. 2:

Data cleaning: can be applied to remove noise and correct inconsistencies in the data.

Data integration: merges data from multiple sources into a coherent data store, such as a data warehouse.

Data transformation: such as normalization, may be applied. For example, normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements [22].

Data reduction: Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume.

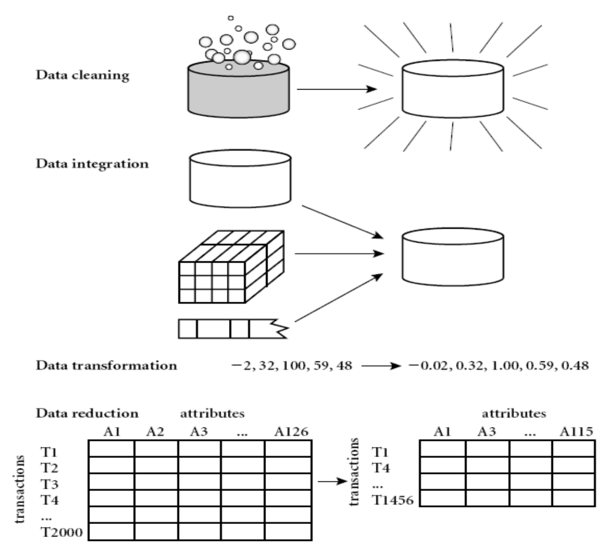

Fig. 2.Flowchart Forms of data preprocessing [22].

4.3. Clustering technique

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the

same cluster and are dissimilar to the objects in other clusters.
There are techniques of clustering to group the data that one of techniques that we are going to use is based on distance-base and its Flow chart has been shown in Figure 3.
Fuzzy c-mean clustering (FCM),from the huge number of clustering methods (see e.g. [4]) We focus in our methodology on fuzzy clustering (see [5]), since the degree of Membership of an object to the classes found provides a strong tool for the identification of changing class structures. We are working with fuzzy c-means (FCM) in order to build an Initial classifier and to update our classifier in each cycle but the presented methodology can be extended to any other technique which determines such degrees of membership (e.g. probabilistic clustering). Details on fuzzy c-means and a general presentation of fuzzy clustering can be found, Ref. [6].
The outputs of FCM are defined as below:

c	Number of classes
n	Number r of objects
p	Number of features describing each object
x_j	Feature vector of object j, j=1, …, n
v_i	Center of class i, i=1,…, c
μ_{ij}	Degree of membership of object j to class i, i=1,…, c, j=1,…, n
M^t	(c×n)-matrix in cycle t with μ_{ij} at position (i; j), i=1,…, c, j=1,…,n

4.4. Type-2Ffuzzy logic system

Type-2 fuzzy logic system handle a different kind of uncertainty than that handled by a non-singleton type-1 FLS [7]. A Non-Singleton system deals with the uncertainty in the input, whereas a type-2 system deals with the uncertainty in our knowledge about the system. Hence, type-2 systems should be robust to rule uncertainties.
Fig 4 shows the structure of a type-2 FLS and we briefly explain the structure:
Fuzzification: The fuzzifier maps the crisp input into a fuzzy set. This fuzzy set will, in general, be a type-2 set in the case of a type-2 FILS [26].
Inference Engine: The inference engine combines rules and gives a mapping from input type-1 fuzzy sets to output type-1 fuzzy sets. Multiple antecedents in rules are connected by the t-norm (corresponding to intersection of sets). The membership grades in the input sets are combined with those in the output sets using the sup-star composition. Multiple rules may be combined using the t-conorm (corresponding to union of sets) operation or during defuzzification [26].
Rules: Fuzzy systems are based on the preliminary information given to the system as fuzzy rules. These rules, which are written as linguistic commands, are usually not so precise. In fact, they are written to enable decision to be made in cases where there is imprecise or no preliminary information about the system under considerations. The following rules represent instances of generic fuzzy rules:

-IF “Salary is high” then “Tax is High”
-IF “Speed is Low” then “Accident Probability is Low”
-F “Left Obstacle is Near” and “Front Obstacle is Near” then “Turn Right Quickly” and “Reduce Speed”

The above rules may have single or multiple antecedents and/or consequences [26].
Type-Reduction and Defuzzification : In a type-1 FLS, the defuzzifier produces a crisp output From the fuzzy set that is the output of the inference engine, output is obtained from a type-1 set. In the type-2 case, an operation analogous to type-1 defuzzification gives a type-l set from a type-2 set (which is the output of the inference engine). We call this process type-reduction rather than defuzzification. The type reduced set can further be defuzzified to obtain a crisp output [26].

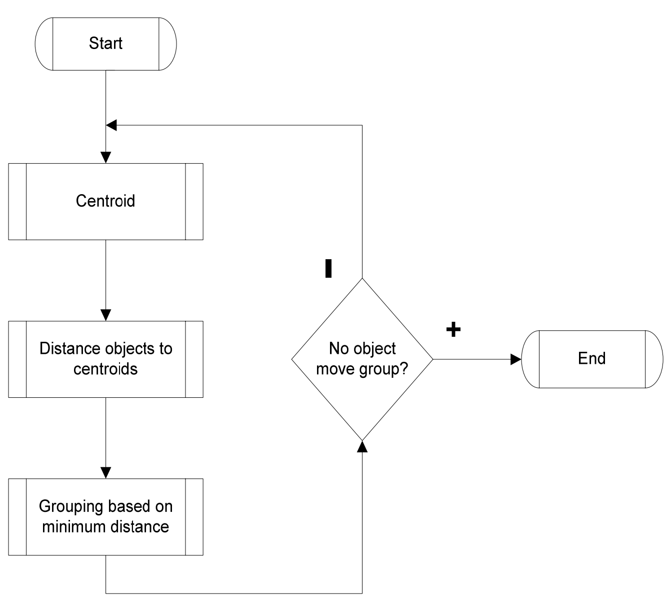

Fig. 3. Flowchart of distance-base method [22].

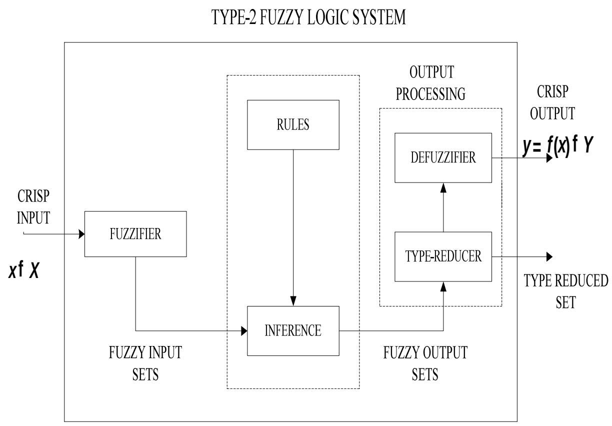

Fig. 4. Structure of type-2 fuzzy logic [3].

5. RESULTS AND DISCUSSION

The dataset for preprocessing is taken from Neuronica Lab at Politecnico of Torino (Italy) containing temperature data and humidity [3]. The preprocessing data is preprocessed by WEKA software to remove missing value and inconsistent data, and then is used as an input for FCM to remove noises , detect outliers and extract desired data after mining is utilized for type-2 Fuzzy logic method to make decision. Figure 5 and 6 show the results in two dimensions, value of sample (Temperature) and time.

In Fig.5 shows released desired line graph are actual data and released obtained line graph are resulted from type-2 FLS technique. The Fig 5 shows that resulted data are not close to real data; this difference comes due to uncertainty in dataset and low quality by existing, noise, outliers, and so on. Therefore to achieve high quality of data we used data mining technique to improve the accuracy and quality of data. We used fuzzy c-mean clustering (FCM) method to remove noisy data and detect outliers and then extract desired clustered data. The Fig 6 shows the result of hybrid method (FCM and Type-2 FLS).

To evaluate the results we used standard measurement called Mean Absolute Error (MAE) and Mean Square Error (MSE) which formulated as below:

If O_t is the actual observation for time period t *and* P_t is the forecast for the same period, then the error is defined as:

Error: $$e_t = O_t - P_t$$

Since there are observations and predictions for *n* time periods, then there will be *n* error terms, and the following standard statistical measures can be defined:

Mean Absolute Error:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Mean Square Error:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e^2{}_t$$

The MAE is defined by first making each error positive by taking its absolute value and then averaging the results. A similar idea is behind the definition of MSE. Here the errors are made positive by squaring each one, and then the squared errors are averaged. The MAE has the advantage of being more interpretable and easier to explain to non-specialists. The MSE has the advantage of being easier to handle mathematically. Each of these statistics deals with measures of accuracy whose size depends on the scale of the data [5].

In such circumstances, Table2 shows results of error value for type-2 FLS and hybrid method (type-2 FLS with FCM). For MAE measurement is 0.9947 and 0.1697 using type-2 FLS and hybrid method (type-2 FLS and FCM) respectively. These values in MSE measurement are 6.8124 and 0.0630. The result shows the hybrid method achieve minimum error rate in both MAE and MSE.

These effective results are due to the mechanism of FCM. FCM improves the quality of data by removing noises and detecting outliers. Thus FCM is a significant technique that can improve the accuracy of weather situation.

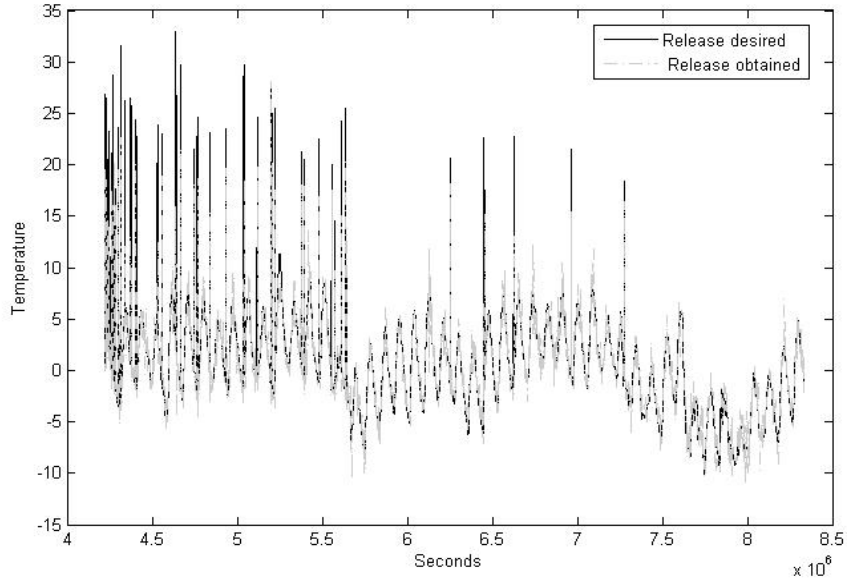

Fig. 5. Outcome of Type-2 Fuzzy logic system

Table 2. Measuring accuracy by utilizing Type-2 FLS and Both Type-2 FLS and C-mean clustering

Error Measurements	Type-2 FLS	C-Mean clustering and Type-2 FLS
MAE	0.9947	0.1697
MSE	6.8124	0.0630

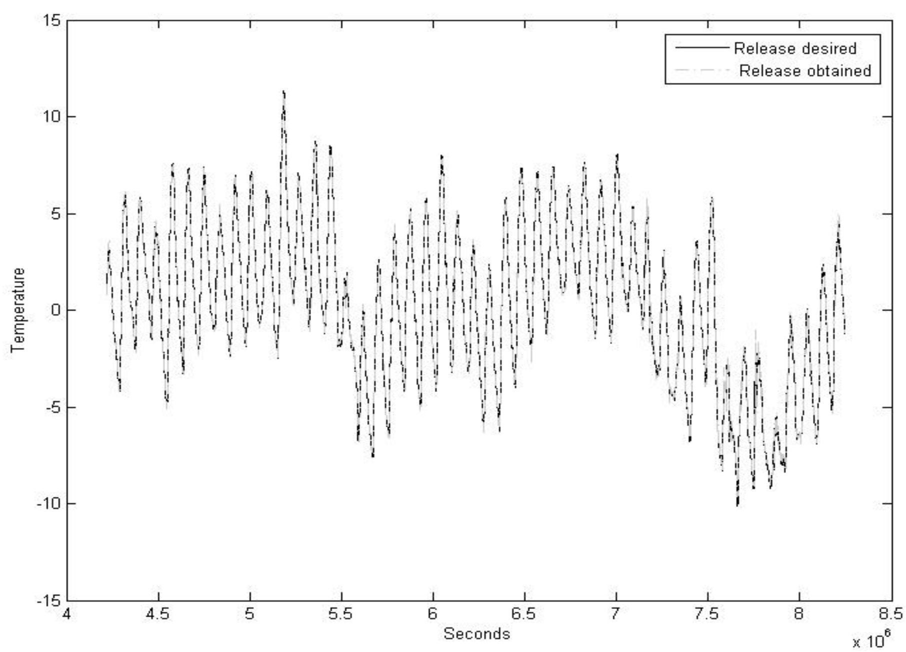

Fig. 6. Outcome of combination of Fuzzy C-Mean clustering and Type-2 FLS.

6. CONCLUSION

The use of heterogeneous data available causes increased uncertainty. Vast amount of data from dynamic environment will be very uncertain and the prediction of future behavior of such situation is very hard and if it is possible, the result will not be accurate and lead to get false prediction, on the other hand, a large amount of data should be processed, this procedure is time consuming and needs very powerful computers. In such circumstances, we proposed an effective method based on Fuzzy c-mean clustering and type-2 fuzzy logic. This proposed method is very effective on uncertainty data which lead to true prediction. In this paper type-2 FLS have been using in local area and it is not applicable for large amount of data because of its time-consuming procedure. The results proved hybrid method is applicable for both local and global area.

REFERENCES

[1]. Zheng Qin, Jinmin Yang, Haoliang Wang, Bin Wang, Jianjun Zou, Neural Network based on Dynamic tunneling technique for Weather Forecast,prdc,11th Pacific Rim International Symposium on Dependable Computing (PRDC'05), 2005, pp. 393- 397.

[2]. Abdul Manan Ahmad, Chia Su Chuan,Fatimah Mohamad, Weather Prediction Using Artificial Neural Network, The International Conference on Circuits/Systems Computers,2002.

[3]. A. Mencattini1, M. Salmeri1, S. Bertazzoni1, R.Lojacono1, E. Pasero2, W. Moniaci, *Local* Meteorological Forecasting by Type-2 Fuzzy Systems Time Series Prediction, IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Giardini Naxos, Italy, 20-22 July 2005.

[4].Bjarne K. Hansen and Denis Riordan,Weather prediction using Case-Base reasoning and Fuzzy set theory", Master of Computer Science Thesis,Technical University of Nova Scotia, Halifax, Nova Scotia, Canada,2005.

[5].A. Mountis and G. Levermore, Weather Predictor for Feedforward Control Working on The Summer Data , Univ. of Manchester, UK, IEEE,2005,pp. 334-340.

[6].Dana Klimešová, *GIS AND UNCERTAINTY MANAGEMENT*, ISPRS Commission VII Mid- Term Symposium "Remote Sensing: From Pixels to Processes", Enschede, the Netherlands, 8-11 May 2006.

[7]. James W. Taylor and Roberto Buizza,Neural Network Load Forecasting With Weather Ensemble Predictions, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 17, NO. 3, AUGUST 2002.

[8].Shun-Feng Su, *Member, IEEE*, Chan-Ben Lin, and Yen-Tseng Hsu,*A High Precision Global Prediction Approach Based on Local Prediction Approaches*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETIC— PART C:APPLICATIONS AND REVIEWS, VOL.32, NO. 4, NOVEMBER 2002.

[9].Gilbert Brunet,*The First Hundred Years of Numerical Weather Prediction*, Recherche en Prevision Numerique Environment Canada, Canada, 19th International Symposium on High Performance Computing Systems and Applications (HPCS'05) 1550-5243/05 $20.00 © 2005 IEEE.

[10].Hansen, B.SIGMAR: A fuzzy expert system For critiquing marine forecasts, AI Applications, Vol. 11, No. 1, 1997,pp.59-68.

[11]. Detlef Nauck, Frank Klawonn and Rudolf Kruse,Foundations of neuro-fuzzy systems, New York; Chichester: Wiley, 1997.

[12].Chin-Teng Lin and C.S. George Lee, Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems, Upper Saddle River; NJ: Prentice-Hall,1996.

[13].WEI DING, FU-GUI DONG, SHANG-DONG YANG , LOAD FORECASTING BASED ON CLUSTERING ANALYSIS USING FUZZY LOGIC, the Fourth International Conference onMachine Learning and Cybernetics, Guangzhou, 18-21 August 2005.

[14].Dokas, I. LOMA: A Web Based Fuzzy Expert System for Accident Prevention in Landfills,ADVISES Research Training Network, Young Researchers Meeting, York University, 25th May 2005.

[15].Kochukuttan,H.andA.Chandrasekaran. Development of a Fuzzy Expert System for Power Quality Applications, Proceedings of the 29th Southeastern Symposium on System Theory (SSST '97),1997.

[16].Wong, F., K. Leung and W. Lam. An Expert Computer System (ABVAB) on Gyneco logical Consultation, Journal of the Hong Kong Medical Association, Vol.41(1),1989.

[17]. Ngai, E. and F. Wat,Design and development of a fuzzy expert system for hotel selection,Omega: The International Journal of Management Science, Vol.31(4), 2003, pp. 275-286.

[18]. Mahabir, C., F. Hicks, and A. Robinson. Application of fuzzy logic to forecast seasonal runoff,Journal of Hydrologic Processes, Vol. 17,2003, pp. 3749–3762.

[19].Sheremetov, L., I. Batyrshin, J. Martinez, H. Rodriguez, and D. Filatov. Fuzzy Expert System for Solving Lost Circulation Problem, Proceedings of the Fifth International Conference on Hybrid Intelligent Systems (HIS'05),2005.

[20].Mahabir, C., F. Hicks, and A. Robinson. FORECASTING ICE JAM RISK AT FORT MCMURRAY, AB_ USING FUZZY LOGIC, Proceedings of the 16th IAHR International Symposium on Ice , Dunedin, New Zealand,, 2002.

[21].Hadjimichae,M.,P. Kuciauskas., P. Tag., R. Bankert. and J. Peak. A Meteorological Fuzzy Expert System Incorporating Subjective User Input, Knowledg And Information Systems,Vol. 4 (3), 2002, pp. 350-369.

[22].Jiawei Han and Micheline Kamber. Data Mining:Concepts and Techniques, Second Edition.Jim Gray,Microsoft Research. TheMorgan Kaufman Series in DataManagement Systems. University of Illinois at Urbana-Champaign: Diane Cerra, 2006 by Elsevier.

[23]. G. Williams, R. Baxter, H. He, S. Hawkins and L. Gu, "A Comparative Study for RNN for Outlier Detection in Data Mining". In Proceedings of the 2nd IEEE International Conference on Data Mining, page 709, Maebashi City, Japan, December 2002.

[24]. Jia Zeng and Zhi-Giang Liu,"Type-2 fuzzy sets for Pattern Recognition: The State-of-the-Art",Journal Of Uncertain Systems, Vol.1, No.3, pp.163-177,30 January 2007.

[25].P.R.Chowdhury,Y.P.Singh,and .A.Chansarkar, "Use of Dynamic Tunneling withBack propagation in Training Feedforward Neural Networks", *IEEE Region 10 International Conference on Global Connectivity in Energy, Computer,Communication and Control*, Vol.1, Dec. 1998 pp.29 – 32

[26].Javid Taheri, and Albert Y. Zomaya. Chapter 7 ." FUZZY LOGIC. The University of Sydney. 221-52.