International Journal of Engineering

Journal Homepage: www.ije.ir

RESEARCH NOTE

A Novel Type-2 Adaptive Neuro Fuzzy Inference System Classifier for Modelling Uncertainty in Prediction of Air Pollution Disaster

A. Safari[a], R. Hosseini*[a], M. Mazinani[b]

[a] *Department of Computer Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Electronic Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran*

PAPER INFO

Paper history:
Received 29 July 2017
Received in revised form 27 August 2017
Accepted 08 September 2017

Keywords:
Fuzzy Logic
Type-2 Fuzzy Set
Adaptive Neuro Fuzzy Inference System
Air Pollution Disaster

ABSTRACT

Type-2 fuzzy set theory is one of the most powerful tools for dealing with the uncertainty and imperfection in dynamic and complex environments. The applications of type-2 fuzzy sets and soft computing methods are rapidly emerging in the ecological fields such as air pollution and weather prediction. The air pollution problem is a major public health problem in many cities of the world. Prediction of natural phenomena always suffers from uncertainty in the environment and incompleteness of data. However, various studies have been reported for prediction of the air quality index but all of them suffer from uncertainty and imprecision associated to the incompleteness of knowledge and imprecise input measures. This article takes advantages of learning of adaptive neural networks alongside in new environment. Furthermore, it presents an Adaptive Neuro-Type-2 Fuzzy Inference System (ANT2FIS) to address the uncertainty and imprecision in air quality prediction. The data set of this study was collected from Tehran municipality official website for the last five years (2012-2017). The results reveal that the ANT2FIS prediction method is more reliable and is capable of handling uncertainty compared to the other counterpart methods. The performance results on real data set show the superiority of the ANT2FIS model in the prediction process with an average accuracy of 94% (AUC 99%) compared to other related works. These results are promising for early prediction of the natural disasters and prevention of its side effects.

***doi**: 10.5829/ije.2017.30.11b.16*

1. INTRODUCTION

Air pollution problem is a global problem that may harm humans or other living creatures in the environment. Prediction of air pollution is a universal important issue because that the prediction of critical events should guide decision making the polluted weather is affecting public health and ecosystem. Air pollution once thought of as entirely a localized issue, and now is recognized as a complex and force major problem which is also subjected to regional and global influences. The motivation of this research is to assist accurate and reliable prediction of AQI to improve public health and to manage one of dangerous humanity crisis, i.e., air pollution disaster. The most effective substance in air pollution is the particular matter (PM). Unlike most other pollutants (Co, So_2, No_2), PM cannot be characterized by time and space variations of the mass consolidation of a single compound [1]. Thus, the different rate of air pollution in the ambient air makes the modeling or prediction of air quality a more uncertain process. PM can have classified into two principal categories in size and morphology, PM 2.5 (diameter smaller than two micrometers) and PM 10. The most difficult challenge in air quality prediction is the behavior of PM on invariant weather and seasons. Table 1 shows five main effective features on AQI.

*Corresponding Author's Email: *rahil.hosseini@qodsiau.ac.ir* (R. Hosseini)

2. TYPE-2 FUZZY SET THEORY

Type-2 fuzzy sets, generalized form of type-1 fuzzy sets, are capable of handling uncertainty in complex environments [2].

Please cite this article as: A. Safari, R. Hosseini, M. Mazinani, A Novel Type-2 Adaptive Neuro Fuzzy Inference System Classifier for Modelling Uncertainty in Prediction of Air Pollution Disaster, International Journal of Engineering (IJE), TRANSACTIONS B: Applications Vol. 30, No. 11, (November 2017) 1746-1751

TABLE 1. Feature description of the of AQI

Features	Description
Sulfur Dioxide (SO_2)	This is one of the causes for concern over the environmental impact of the use of fuels as power sources. Breathing in SO_2 is attendant with increased lung indications and disease
Nitrogen Dioxide (NO_2)	Inhalation of such particles may cause or worsen respiratory diseases such as emphysema, bronchitis. It may also aggravate existing heart disease
Carbon Monoxide (CO)	The health threat from CO at low levels is important for those who hurt from cardiovascular diseases like angina pectoris
Particulate Matter (PM)	EPA grouped particle matter pollution into two classes: particles smaller than 10 µm and 2 .5 µm (PM10 and PM25)

The MF of a type-1 fuzzy set is a crisp number, while in a type-2 fuzzy set (T2FS), the MF is a subset of a fuzzy set and is itself fuzzy [3]. The mf of a type-2 fuzzy set which is defined by the function, is called the secondary MF of a type-2 fuzzy set [3] where a fuzzy set can be well-defined, as the area of a secondary mf is called the primary MF. As we said before, FOU is the two-dimensional support of $\tilde{A}$, follow as [4]:

$$FOU\left(\tilde{A}\right) = \{(x.u) \in X \times U \,|\, \tilde{\mu}A(x.u) > 0\} \quad (1)$$

In the Figure 1, the FOU primary membership functions are shown.

The universal architecture of a T2FLS is shown in Figure 2. The important components in the architecture of a T2FLS: fuzzifier, rules, inference engine and output producer, which contains a type reducer and a defuzzifier. The inference engine combines rules and gives a mapping from the input type-2 fuzzy sets to the output type-2 fuzzy sets.

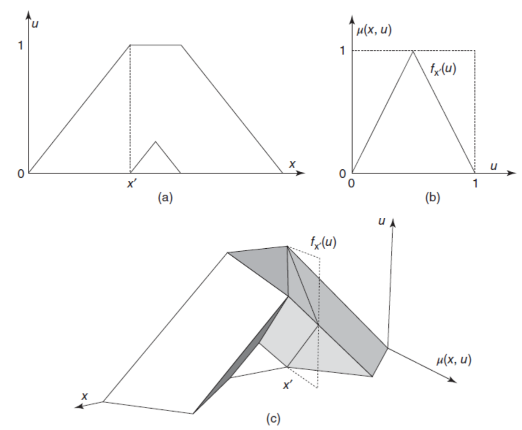

Figure 1. Footprint of Uncertainty Scheme (a) FOU with primary membership at x', (b) two possible secondary membership functions associated with x', and, (c) the resulting 3D type-2 fuzzy set [4]

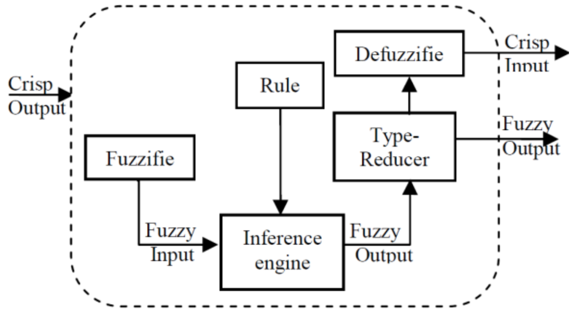

Figure 2. General Structure of Type-2 Fuzzy System [5]

Multiple antecedents are associated by the t-norm and the membership of input and output sets are combined using the sup-star arrangement. Multiple rules are combined using the t-conorm operation [5].

2. 1. The Concept of Type-2 Anfis Classifier

Being static and non-flexibility are two major weak points of a fuzzy inference system. The purpose of using the artificial neural network (ANN) beside of the fuzzy inference system is that neural networks are adaptive, online and strongly capable to learn from previous experiences. The hybridization of the type-2 fuzzy system and neural network make a powerful tool for handling uncertainty and dynamic conditions in an AQI prediction model [5].

3. REVIEW OF RELATED INTELLIGENT MODELS FOR AIR QUALITY PREDICTION

In this part, some articles are reviewed which used computational intelligence methods for predicting, analyzing, measuring and modeling the air quality index problem.

A fuzzy based air quality index (FAQI) method for air quality valuation has proposed by Sowlat et al. [6]. This study represents and validates such a fuzzy based algorithm based on Mamadani inference engine with one antecedent in rules. In this article, ten parameters divided into two groups were selected to cover all possible significant pollutants in the study area according to their strengths along with their impacts on human health. This article shows that an AQI based only on measures air pollutants may not be typical of the air quality. This brings about the need for a more comprehensive index, involving all the major pollutants and a boundary result in consequent part for uncertainty in boundary decision making.

An adaptive neuro-fuzzy based modeling for prediction of AQI has proposed by Yildirim and Bayramoglu [7]. The proposed method which analyzes the city of Zonguldak is a coastal town in Turkey.

Their ANFIS model has created by the subtractive clustering technique. First-order Sugeno model has chosen as inference system and neural networks are trained by the hybrid method. For defining membership functions, they applied Gaussian type membership functions. They found that the model is indicating pleasant prediction limits between 75–90% and 69–80% for SO_2 and TSP. They claimed that other researchers cab uses a better training dataset to forecast the AQI ranks with more accuracy.

In Ref. [8] they proposed three neural network methods for nonlinear modeling and one technique for linear for prediction of urban air quality in Lucknow city (India). In non-linear methods, they used MLPN, RBFN, and GRNN which are three of the most generally used neural networks constructions in literature for regression problems. They found that the main advantage of a GRNN is the extremely rapid training which does not need an iterative training technique.

PM10 forecasting based on various soft computing techniques has been proposed by Ababneh [9]. The main goal of their research was to improve accurate and effective air pollution predicting model for PM10 pollutants. As a base discipline, they used the artificial neural network for forecasting PM10 index. For training of system, they worked with general back-propagation learning algorithm. Also, they used hyperbolic tangent and linear activation functions for hidden and output layers. The comparison of the reviewed method and some other related works [10-12] bring it on Table 2 respectfully.

4. PROPOSED TYPE-2 ANFIS PREDICTIVE CLASSIFIER

4. 1. The Structure of Type-2 ANFIS One of the remarkable advantages of ANFIS systems is their capability in creating the proper rules and membership functions automatically, based on the input data used for training [13].

TABLE 2. Comparison of related intelligent models for AQI prediction

Methods	Advantages	Inputs	Outputs
New Type-1 Fuzzy [10]	Reliable and interpretable.	10	2
A new method based on ANFIS [11]	Determine the optimum number of fuzzy MFs.	18	8
ANFIS [12]	Acceptable performance	7	2
PLSR, MPR and ANN [13]	Capture the non-linearity in the data	5	2
Soft Computing [14]	Tried basic methods	8	2
k-means clustering, ANN [15]	Accurate in complex time series analysis	2	5
Fuzzy Genetic [16]	Evolutionary MFs creation	1	5

Proposed Type-2 ANFIS can predict better results and make a rational output for the unseen input data. The structure of the proposed Type-2 ANFIS method has the following steps:

1. The first layer is input variables.
2. The second layer involves of fuzzifies that drive inputs to the type-2 fuzzy terms used in the knowledge base (Rules).
3. The third layers include input MFs.
4. The fourth layer includes nodes demonstrating the rules of the knowledge base (Normalization).
5. The fifth layer contains output terms MFs of type-1 fuzzy.
6. The sixth layer calculates the fuzzy output for the output variables in whole system and aggregation.
7. The seventh layer takes in the defuzzification.

In the other side for defining the Antecedent MFs, the primary MFs for every antecedent are FSs described by Gaussian with uncertain means which is showed in Equation (2) [14].

$$\mu_k^i(X_k) = \exp\left[-\frac{1}{2}\left[\frac{X_k - m_k^i}{\sigma_k^i}\right]^2\right] \quad (2)$$

where, $m_k^i \in \lfloor m_{k1}^i . m_{k2}^i \rfloor$ is the uncertain mean, with k =1,2 (the number of antecedents) and i = (1,2, 3…) (number of rules), and σ_k^i is the standard deviation. The means of the antecedent fuzzy sets are homogeneously distributed over the whole input space.

4. 2. The Rules of Classifier Fuzzy inference is the process of formulating the mapping from given input(s) to output(s) using fuzzy logic [15]. The rules are heuristically extracted after a long duration search in main global space. One of the most effective steps in this section is the tuning of rules. The genetic algorithm (GA) was applied to make tuned rules as well as possible. The tuning process in this work has five main steps: handling inconsistent rules, management of duplicate rules, redundancy management, handling reduction and conjunction operators and linguistic hedges. After some generation, 19 robust rules are applied for the proposed classifier.

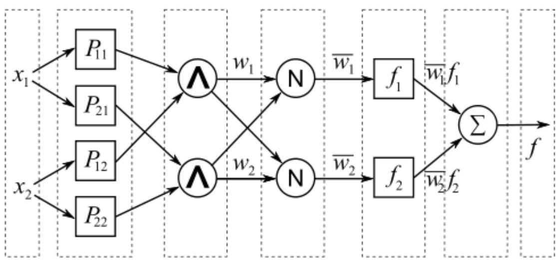

Figure 3. Type-2 ANFIS Structure

To make the process robust and independent from the weight choice for the fitness function, the type-2 fuzzy set theory goal was to dynamically adapt toward the search direction in the space of solutions. The selection structure was based on tournament technique; but to avoid from high selection pressure, elitism strategy was applied. In order to tackle the overfitting problem, a post-pruning technique was used. The structure of some rules used in rule-base as follows:

If PM-2.5 is " Low" and NO_2 is "low" then AQI is "Good."

If CO is "Medium" and PM-2.5 is "Medium" then AQI is "Unhealthy."

If PM-2.5 is "High" and SO_2 is “Medium” then AQI is "Unhealthy for Sensitive Groups."

If PM-10 is "Extremely High" and NO2 is “High” then AQI is "Dangerous."

Figure 4 shows the Gaussian input MFs of the proposed method.

The aim of the combination of neural networks and the type-2 fuzzy set was to make a more transparent, interpretable and accurate predictive classification system. The type-2 ANFIS can predict better and make a rational output for the unseen samples. The whole data set was separated into three parts; training set, validation set, test set. For this purpose, a 10-Fold cross validation was used to validate the results and make the classifier more reliable. The idea is to randomly divide the data into 10 (K) equal-sized parts. It leaves out part k, proper the model to the other K-1 parts (joint), and then gets the prediction for the left-out k-th part. This is done in turn for each part k = 1, 2, K and then the results are combined. The block diagram of the classification procedure is shown in the following diagram in Figure 5.

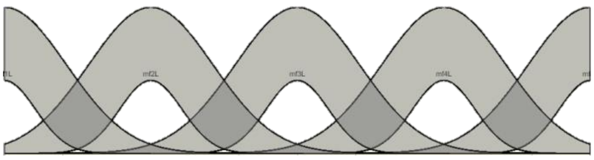

Figure 4. Gaussian Type-2 Membership Function for Input Variables

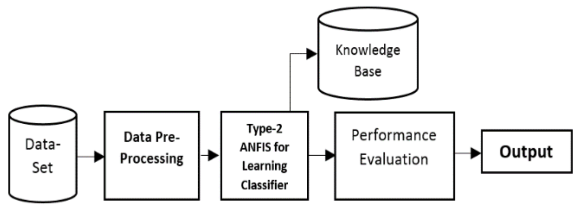

Figure 5. The procedure of Proposed Type-2 ANFIS classifier

5. EXPERIMENTAL RESULTS (CASE STUDY-TEHRAN PROVINCE)

To evaluate the performance of T2ANFIS method, a case study was directed in Tehran, the capital of Iran. Pollutant concentrations, including average of CO, SO_2, PM10, PM2.5, and NO_2 are daily measured in an average of different sampling stations for the last five years (2012-2017). The study results in high concentrations of air pollutants, especially in the case of pollutants mainly with automated-origin. The Tehran municipality data set in the last five years has been frequently used for the evaluation of classifiers. The data set was monitored from 13 different stations in a whole time-series duration based on daily records. The data set used in this research contains 10213 samples. The outputs of the classifier were defined as Good (14%), Moderate (21%), Unhealthy for Sensitive Group (32%), Unhealthy (21%) and Dangerous (12%).

5. 1. Air Quality Index (AQI)

In this study, five major air pollutants (CO, SO_2, PM10, PM2.5, and NO_2) were considered. Their concentrations are classified into five different categories according to frequency breakpoints. Table 3 shows the five main classes of AQI for input variables.

5. 2. Performance Evaluations

In order to evaluate the performance of the proposed model for AQI index prediction, a ROC curve analysis was conducted. The AQI problem is a multi-class classification, because of high dimension and massive complexity of surface under the curve (SUC) is used binaries technique (One Vs. One) to report the average of five class as a ROC curve [16]. In order to estimate classification accuracy through ROC curve analysis method as a widely accepted standard for performance evaluation of the classification models. To avoid from over-fitting and over-learning in the proposed method, the cross-validation strategy is used to handle the probable issue. Figure 6 shows the precision of the proposed classifier for AQI measurement in a daily time-series.

TABLE 3. Classification levels of Air Quality of Input Parameters and Variables [17]

Inputs	Low	Mid	High	Very High
NO_2	0-0.105	0.10-0.21	0.21-0.31	0.31-0.42
SO_2	0-0.065	0.06-0.13	0.13-0.19	0.19-0.26
CO	0-5.50	5.51-11	11-16.50	16-22
PM_{10}	0-60	61-120	121-220	221-320
$PM_{2.5}$	0-15.4	15.5-40.4	40.5-65.4	65-150

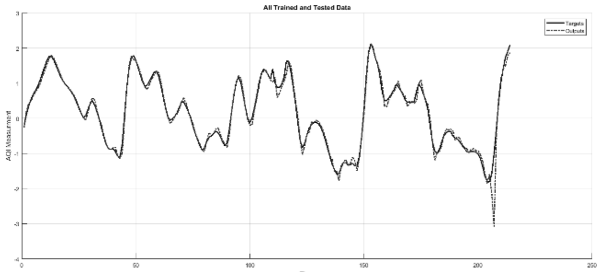

Figure 6. Precision measurement of output classifier in compare with system targets

Figure 7 shows the ROC curve of the proposed method in compare of our earlier work (Type-1 ANFIS) while using the same dataset. The type-2 ANFIS is obviously had better performance compare with the standard type-1 ANFIS classifier, where μ (i) AND μ (J) are the means ROC curve accuracy of the 10-fold cross-validation as follow as Equation (3) [18]:

$$\mu_i = \frac{1}{10}\sum_{k=1}^{10} AUCj \quad (3)$$

In Table 4, a comparison of the proposed method and earlier works for AQI prediction with different number of input variables are presented.

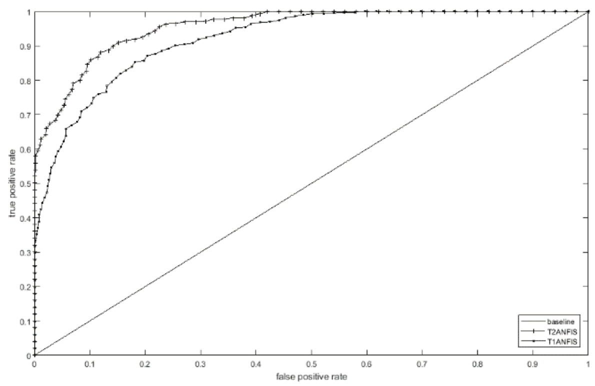

Figure 7. ROC Curve for proposed method and earlier authors work

TABLE 4. Comparison of proposed methods of AQI Prediction (NP*: Not Reported, CI**: Confidence Interval)

Method	Inputs	Evaluation	Accuracy %	CI**
Type-1 Fuzzy [10]	10	FSE	89	NP*
ANFIS [14]	7	RMSE	90	NP
PLSR, MPR [12]	6	MSE	89	NP
ANN Methods [13]	6	RMSE	88	NP
Type-1 Fuzzy (This Work)	5	ROC Curve	87	[74, 88]
Type-1 ANFIS (This Work)	5	ROC Curve	90	[90.1, 94]
Type-2 ANFIS (This Work)	5	ROC Curve	94	[90.3, 94]

The type-1 fuzzy classifier, Type-1 ANFIS and the type-2 ANFIS classifier are our works, respectfully compared with other related works.

6. CONCLUSION

This study presented an adaptive Neuro- Type-2 Fuzzy inference model to estimate the AQI factors over an urban area such as Tehran. The ANT2FIS approach for air quality index is capable of predicting other uncertain and imprecise conditions such as natural phenomena environment. To evaluate the methodology, it is applied to the Tehran standard AQI data set for the last five years. The average accuracy of type-2 ANFIS classifier, using five input variables and after applying a 10-fold cross-validation, was 94% with a 95% confidence interval (CI) of [90.33 94], which is comparative the average outcome of the earlier methods. Furthermore, the ANT2FIS has shown a good trade-off between accuracy, precision and interpretability in a complex dimensional phenomenon. The proposed method has the competence to accomplish more uncertainties in all part of prediction system and is appropriate for pattern classification problems when have noisy training and imprecise data set or lack of expert's knowledge. Our future work is to design an automatic adaptive and self-adaptive rule extraction model from the input data set for AQI prediction problem to reduce the number of rules and its complexity based on type-2 evolutionary algorithms.

7. REFERENCES

1. Sher, E., "Handbook of air pollution from internal combustion engines: Pollutant formation and control, Academic Press, (1998).
2. Hosseini, R., Ellis, T., Mazinani, M. and Dehmeshki, J., "A genetic fuzzy approach for rule extraction for rule-based classification with application to medical diagnosis", in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Citeseer., (2011), 05-09.
3. Sadeghian, A., Mendel, J. and Tahayori, H., "Advances in type-2 fuzzy sets and systems: Theory and applications", Springer, Vol. 301, (2013), ISBN: 978-1-4614-6666-6.
4. Hosseini, R., Qanadli, S.D., Barman, S., Mazinani, M., Ellis, T. and Dehmeshki, J., "An automatic approach for learning and tuning gaussian interval type-2 fuzzy membership functions applied to lung cad classification system", ***IEEE Transactions on Fuzzy Systems***, Vol. 20, No. 2, (2012), 224-234.
5. Mendel, J., Hagras, H., Tan, W.-W., Melek, W.W. and Ying, H., "Introduction to type-2 fuzzy logic control: Theory and applications, John Wiley & Sons, (2014).
6. Sowlat, M.H., Gharibi, H., Yunesian, M., Mahmoudi, M.T. and Lotfi, S., "A novel, fuzzy-based air quality index (FAQI) for air quality assessment", ***Atmospheric Environment***, Vol. 45, No. 12, (2011), 2050-2059.

7. Yildirim, Y. and Bayramoglu, M., "Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of zonguldak", ***Chemosphere***, Vol. 63, No. 9, (2006), 1575-1582.
8. Singh, K.P., Gupta, S., Kumar, A. and Shukla, S.P., "Linear and nonlinear modeling approaches for urban air quality prediction", ***Science of the Total Environment***, Vol. 426, (2012), 244-255.
9. Ababneh, M.F., Ala'a, O. and Btoush, M.H., "Pm10 forecasting using soft computing techniques", ***Research Journal of Applied Sciences, Engineering and Technology***, Vol. 7, No. 16, (2014), 3253-3265.
10. Shahraiyni, H.T., Sodoudi, S., Kerschbaumer, A. and Cubasch, U., "A new structure identification scheme for anfis and its application for the simulation of virtual air pollution monitoring stations in urban areas", ***Engineering Applications of Artificial Intelligence***, Vol. 41, (2015), 175-182.
11. Elangasinghe, M., Singhal, N., Dirks, K., Salmond, J. and Samarasinghe, S., "Complex time series analysis of pm 10 and pm 2.5 for a coastal site using artificial neural network modelling and k-means clustering", ***Atmospheric Environment***, Vol. 94, (2014), 106-116.
12. Shad, R., Mesgari, M.S. and Shad, A., "Predicting air pollution using fuzzy genetic linear membership kriging in gis", ***Computers, Environment and Urban Systems***, Vol. 33, No. 6, (2009), 472-481.
13. Moghadam-Fard, H. and Samadi, F., "Active suspension system control using adaptive neuro fuzzy (anfis) controller", ***International Journal of Engineering-Transactions C: Aspects***, Vol. 28, No. 3, (2014), 396-401.
14. Mendez, G. and De Los Angeles Hernandez, M., "Interval type-2 anfis", ***Innovations in Hybrid Intelligent Systems***, (2007), 64-71.
15. Khezri, R., Hosseini, R. and Mazinani, M., "A fuzzy rule-based expert system for the prognosis of the risk of development of the breast cancer", ***International Journal of Engineering Transactions A: Basics***, Vol. 27, No. 10, (2014), 1557-1564.
16. Bernard, S., Chatelain, C., Adam, S. and Sabourin, R., "The multiclass roc front method for cost-sensitive classification", ***Pattern Recognition***, Vol. 52, (2016), 46-60.
17. "Air quality index", in U.S. Environmental Protection Agency (EPA), Office of Air Quality Planning and Standards Outreach and Information Division Research Triangle Park, Outreach and Information Division Research Triangle Park, NC., (2014 of Conference).
18. Ara, A.L., Tolabi, H.B. and Hosseini, R., "Dynamic modeling and controller design of distribution static compensator in a microgrid based on combination of fuzzy set and galaxy-based search algorithm", ***International Journal of Engineering-Transactions A: Basics***, Vol. 29, No. 10, (2016), 1392-1400.

A Novel Type-2 Adaptive Neuro Fuzzy Inference System Classifier for Modelling Uncertainty in Prediction of Air Pollution Disaster

RESEARCH NOTE

A. Safari[a], R. Hosseini[a], M. Mazinani[b]

[a] *Department of Computer Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Electronic Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran*

PAPER INFO

Paper history:
Received 29 July 2017
Received in revised form 27 August 2017
Accepted 08 September 2017

Keywords:
Fuzzy Logic
Type-2 Fuzzy Set
Adaptive Neuro Fuzzy Inference System
Air Pollution Disaster

چکیده

تئوری مجموعه فازی نوع ۲ یکی از قوی ترین ابزارها برای تقابل با عدم قطعیت در یک محیط پویا و پیچیده است. در سال های اخیر کاربردهای مجموعه های فازی نوع ۲ و روش های محاسبات نرم به سرعت در زمینه های زیست محیطی مانند آلودگی هوا و پیش بینی آب و هوا مورد استفاده قرار گرفته است. مشکل آلودگی هوا یک مشکل عمده بهداشت عمومی و زیست محیطی در بسیاری از شهرهای جهان است. پیش بینی پدیده های طبیعی همیشه از عدم اطمینان در محیط و ناقص بودن داده ها رنج می برند. در این زمینه مطالعات زیادی برای پیش بینی شاخص های کیفیت هوا گزارش شده است، اما همه آنها از عدم اطمینان و ابهام در ارتباط با ناقص بودن دانش و اندازه گیری های نامناسب رنج برده اند. این مطالعه مزایای یادگیری و قدرت سازگاری شبکه های عصبی تطبیقی فازی را در محیط های جدید نشان میدهد. علاوه بر این، این پژوهش ارائه یک سیستم استنتاج فازی تطبیقی عصبی نوع ۲ (ANT2FIS) برای رفع عدم اطمینان و عدم ابهام در پیش بینی کیفیت هوا را ارئه داده است. مجموعه داده های این مطالعه از پایگاه رسمی شهرداری تهران طی پنج سال گذشته (۲۰۱۲–۲۰۱۷) جمع آوری شده است. نتایج نشان می دهد که پیش بینی بر اساس روش ANT2FIS قابل اعتماد تر است و قابلیت بیشتری در مدیریت عدم قطعیت در مقایسه با سایر روش های مرتبط را داراست. نتایج عملکرد مدل ارائه شده در این پژوهش بر روی داده های واقعی نشان دهنده قدرت سیستم ANT2FIS در فرآیند پیش بینی شاخص وضعیت هوا با صحت عملکرد میانگین ۹۴٪ (AUC ۹۹٪) در مقایسه با سایر کارهای مرتبط است.

doi: *10.5829/ije.2017.30.11b.16*