Short-term Traffic Flow Forecast on Basis of PCA- Interval Type-2 Fuzzy System

Zhengyang Qu¹, Jun Li¹
¹Automation & Electrical Engineering School, Lanzhou Jiaotong University, China
²lijun691201@mail.lzjtu.cn

Abstract. According to the time series of urban traffic flow, a prediction method on basis of type-2 fuzzy logic is proposed under the theoretical framework of fuzzy logic. The prediction model of the interval type-2 fuzzy system is established, and the BP algorithm is used to adjust the coefficients of the antecedent and consequent of fuzzy rules. In this paper, the algorithm is verified by the measured data of road network, and compared with other fuzzy methods, BP algorithm and support vector machine (SVM). According to the experimental data that type-2 fuzzy logic system has higher prediction accuracy.

1. Introduction
The traffic flow system is a huge system with strong uncertainty and randomness. For this nonlinear system, the general parameter model is difficult to predict traffic flow, and the prediction accuracy is lower. At present, Kalman filter[1], SVM and BP neural network [2] have been successfully applied to traffic flow prediction. Through these examples, it is proved that the traffic flow system can be modeled by these forecasting methods, thus overcoming some nonlinear problems in the forecasting process[3]. However, for complex traffic flow systems, when the predicted model contains uncertain information, such as noise, holidays, etc., the prediction accuracy and robustness may be reduced.

As a convenient and fast time-series modeling tool[4], The type-2 fuzzy logic system has been successfully adopted in a lot of fields, including Mackey-Glass chaotic time-series forecast, short-term wind speed prediction, chemical process identification, etc.and has a good development prospect and potential. Interval type-2 fuzzy set simplifies type-2 fuzzy set, and its purpose is to effectively reduce the tedious computation of type-2 fuzzy system. When using a fuzzy system to predict nonlinear data, the input dimension will be higher, based on the existing successful models, this article proposes a way combing principal component analysis and type-2 fuzzy system, which is not only applied to traffic flow prediction, but also compared with other forecasting methods such as SVM. The results show that this method has higher accuracy.

2. The principal component analysis
The essence of primary component discussion is a statistical approach of dimension decrease. Utilizing orthogonal change, the original random vectors with correlated components are transformed into new random vectors with uncorrelated components[5], thus eliminating multicollinearity and reducing noise interference in data. The dimension of the measured traffic flow data is reduced, which effectively reduces the calculation difficulty. The algorithm steps are as follows:

Step1: The data sets input for a given dimension:

\[
\hat{x} = \{\hat{x}_1, K, \hat{x}_n\} \in R^{p \times N}
\] (1)

Published under licence by IOP Publishing Ltd

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Journal of Physics: Conference Series 2171 (2022) 012051 doi:10.1088/1742-6596/2171/1/012051
Step2: Calculate the covariance matrix and its eigenvalues and eigenvectors. Eig(•)-Eigenvalue decomposition function:

\[
[U, A] = \text{eig}(NC)
\]  
(2)

\[
C = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)(x_i - m)^T
\]  
(3)

Step3: Projecting the primary data into a space:

\[
x_i = U_i^T \bar{x}_i, \quad i = 1, K, N
\]  
(4)

The following is an extraction method of the preprocessed data set:

\[
X = \{x_i, K, x_N\} \in \mathbb{R}^{p \times N}
\]  
(5)

3. Type-2 fuzzy system

The computational complicatedness of type-2 fuzzy data is very complex, especially for the type-reduction[6], so to simplify the calculation, let the secondary grade be 1. In this way, we can get an interval type-2 fuzzy system: means an interval type-2 fuzzy set. When \(\forall x \in X, \mu_2(x, u, \rho)=1\) there are:

\[
\bar{y} = \int_{x \in X} \int_{u \in U} 1/(x, u) \, du \, dx \quad J \in [0,1]
\]  
(6)

Fuzzy rules are used to deal with fuzzy input generated by the membership function. Let be the \(i\)-th ingredient of input, \(F_i\) be the \(i\)-th antecedent component of the \(l\)-th rule, which is a fuzzy set corresponding to \(x_i\), and \(y_i\) represent the output function of the \(l\)-th rule. In this paper, TSK’s consequent[7] is used, and the fuzzy rules are as follows:

\[
x_i \in F_i, x_j \in F_j, \ldots, x_p \in F_p, \text{then}, y_i = f_i(x_1, x_2, \ldots, x_p)
\]  
(7)

In which \(F_i\) is the antecedent of interval type-2 FLS responding to \(x_i\). It is illustrated by the uncertain mean Gaussian primary membership role.

\[
\mu_{F_i}(x_k) = \exp \left\{ -\frac{1}{2} \left( \frac{x_k - m_{F_i}}{\sigma_{F_i}} \right)^2 \right\}, \quad m_{F_i} \in [m_{F_i}^-, m_{F_i}^+]
\]  
(8)

where, \(k = 1, \ldots, p, \quad i = 1, \ldots, M\).

Construct the points in the set as Gaussian fuzzy numbers:

\[
\mu_{x_i}(x_k) = \exp \left\{ -\frac{1}{2} \left( \frac{x_k - x_i^*}{\sigma_{x_i}} \right)^2 \right\}
\]  
(9)

\(\sigma_{x_i}\) is the standard deviation of the input set, \(x_i^*\) is a center of fuzzy sets

Fuzzy inference engine: fuzzy reasoning is the key to a fuzzy system. Inference engine calculates the membership function of FLS output according to input and the antecedent of the rule

\[
\mu_{y_l}(y) = \prod_{i=1}^{N} \mu_{B_i}(y)
\]  
(10)
B is a fuzzy logical system output set;

Type-reducer: This is a unique feature of a type-2 fuzzy system[8], and it must be decreased to a type-1 fuzzy set before getting accurate data at last. In this paper, the centroid reduction method is adopted.

Defuzzifiers: through the centroid method, the output result of the precise device is:

$$y_c(x) = \frac{\sum_{i=1}^{N}y_i \mu_B(y_i)}{\sum_{i=1}^{N} \mu_B(y_i)}$$

(11)

4. Optimization parameters of BP algorithm

An input-output data pair \((x^{(j)}, y^{(j)}); j=1, \ldots, N\) is given, for designing a fuzzy logic model. Reduce the error by updating the design parameters[9]:

$$e^{(j)} = \frac{1}{2}[f(x^{(j)}) - y^{(j)}]^2$$

(12)

The backpropagation algorithm as described below:

Step1: Initialize all parameters and set the training counter

Step2: Set up counter \(t=1\) for the collecting data sets.

Step3: Computed \(y_{j1}, y_{j2}\):

$$y_{j1} = \frac{\sum_{i=1}^{M} f_i^j y_i^j}{\sum_{i=1}^{M} f_i^j} = y_{j1} \left( f^j, f^{L+1}, f^{M}, y_{j1}, y_{j2} \right)$$

(13)

$$y_{j2} = \frac{\sum_{i=1}^{M} f_i^j y_i^j}{\sum_{i=1}^{M} f_i^j} = y_{j2} \left( f^j, f^{R+1}, f^{M}, y_{j1}^j, y_{j2}^j \right)$$

(14)

Step4: Calculated defuzzified output

$$y_r(x^{(j)}) = \frac{y_{j1}(x^{(j)}) + y_{j2}(x^{(j)})}{2}$$

(15)

Step5 Set \(t=t+1\), when \(t=N+1\), jump to Step 6; or turn to Step 3;

5. Example of traffic flow prediction

Based on the application research of type 2 fuzzy sets, this paper constructs a novel short-term traffic flow prediction framework [10], and makes a prediction model for the traffic flow of an intercontinental highway abroad [11-12]. Firstly, based on the C-means clustering method, this paper improves the Gaussian interval type 2 fuzzy set, and fuzzifies the traffic data of each period (freedom, congestion, congestion). Through clustering, single-point data are transformed into cluster intervals and embedded type-1 fuzzy sets through these data. On this basis, the interval type-2 fuzzy sets are generated by the Join operation. Finally, the prediction results are generated by type-reduction and defuzzification. In this paper, 2660 groups of data are selected for testing, and the first 2500 groups are training data sets. The remaining 100 groups are test data sets, and Figure 1 is the basic prediction model.
Different methods were used in the tests, the antecedent has 5 fuzzy sets[15], and set principal component $p = 4$, set the study rate to 0.3. After 15 cycles of testing and training, the RMSE was computed. Table 1 shows several fuzzy methods for different parameters.

| FLS      | The mean and standard deviation of antecedent | Consequent               |
|----------|----------------------------------------------|--------------------------|
| A1-C0 TSK FLS | $m_k^l, \sigma_k^{r} = \sigma_s$ | $c_j^l \in [0, 0.3]$   |
| A1-C1 TSK FLS | $m_k^l, \sigma_k^{r} = \sigma_s$ | $c_j^l \in [0, 0.3], s_j^l \in [0, 0.03]$ |
| A2-C0 TSK FLS | $[m_k^l - 0.2 \sigma_s, m_k^l + 0.2 \sigma_s], \sigma_k^{r} = \sigma_s$ | $c_j^l \in [0, 0.3]$   |
| A2-C1 TSK FLS | $[m_k^l - 0.2 \sigma_s, m_k^l + 0.2 \sigma_s], \sigma_k^{r} = \sigma_s$ | $c_j^l \in [0, 0.3], s_j^l \in [0, 0.03]$ |

The following figure displays the contrast result between the forecast value and the actual value of the type-2 fuzzy system after 15 cycles of training. The blue line in the chart represents the actual output of traffic flow, the red line indicates the predicted output of the model, and Figure 3 displays the test mistake. According to the outcomes, the Type-2 TSK prediction approach shows a good role.

**Fig. 1** Structure of Traffic flow Prediction model

**Fig. 2** Traffic Flow prediction on basis of type-2 TSK FLS
Table 2 shows the prediction errors based on BP neural network, SVM method, and four different FLSs methods on the test set. It can be seen that the prediction precision of the FLSs approach is higher than that of the SVM and BP approaches. Among them, A2-C1 TSK FLS prediction method has the highest accuracy, which shows that the fuzzy logic system has very strong application potential in dealing with nonlinear data such as traffic flow.

| Method                 | RMSE  |
|------------------------|-------|
| A1-C0 TSK FLS          | 0.0619|
| A1-C1 TSK FLS          | 0.0648|
| A2-C0 TSK FLS          | 0.057 |
| A2-C1 TSK FLS          | 0.0557|
| BP                     | 0.1119|
| SVM                    | 0.058 |

6. Conclusion
The forecasting method on basis of PCA-interval type-2 FLS can deal with nonlinear data with strong randomnesses, such as traffic flow, based on uncertain rules, which can effectively avoid the problem of fuzzy rule explosion and improve the forecasting accuracy. Aiming at strong nonlinearity, a prediction model based on interval type-2 TSK method is proposed. The difference with the type-1 fuzzy system, BP neural network, and SVM method, it shows that the interval type-2 method has higher accuracy and better prediction performance.

References
[1] Yang G F, Xu R, Qing M, Zheng K L 2017 Short-term traffic prediction based on ARMA and Kalman filter Journal of Zhengzhou University 38(02) pp 36-40
[2] Xiong W Q, Yan X B, Jiang S X 2015 Short-term Traffic Flow Prediction base on BP Neural Network and Fuzzy Inference System Intelligent Computer and Applications 5(02) pp 43-46, 51
[3] Shiliang S, Changshui Z, Guoqiang Y A 2006 Bayesian network approach to traffic flow forecasting IEEE Transactionson Intelligent Transportation Systems 7 (1) pp 124-132
[4] Zhao L 2017 Short-Term Traffic Flow Prediction Based on Interval Type-2 Fuzzy Neural Networks IEEE Transactions on Fuzzy Systems 99 pp 231-246
[5] Wang X H 2019 short-term wind power prediction based on PCA-interval type-2 FLS Acta Energiae Solaris Sinica 40(03) pp 608-619
[6] Wu D, Mendel J 2019 Recommendations on Designing Practical Interval Type-2 Fuzzy Systems Engineering Applications of Artificial Intelligence 85 pp 182-193
[7] Mendel J M, John R I, and Liu F 2006 Interval type-2 fuzzy logic systems made simple IEEE Trans. 14(6) pp 808–821
[8] Yang J, Zhu J W, Liu B, Feng C, Zhang H L 2019 Short-term passenger flow prediction of urban rail transit based on composite model Transportation System Engineering and Information 19(03) pp 119-125
[9] Zhao L 2017 Short-Term Traffic Flow Prediction Based on Interval Type-2 Fuzzy Neural Networks IEEE Transactions on Fuzzy Systems 99 pp 231-246
[10] Li R M, Jiang C Y, Zhu F H, Chen X L Traffic flow data forecasting based on interval type-2 fuzzy sets theory Journal of Automatica Sinica IEEE/CAA
[11] Li R M, Liang Q H 2019 Reliability evaluation of artificial traffic system based on Interval-type-2 fuzzy set 45(10) pp 1915-1922
[12] Mendel J M and Wu H 2007 Type-2 fuzzistics for non-symmetric interval type-2 fuzzy sets: Forward problems IEEE Trans. Fuzzy Syst. 15(5) pp 916-930
[13] Pan Y P, Huang D P, Sun Z H 2011 Overview of type-2 fuzzy logic control Control Theory & Applications 28(001) pp 13-23
[14] Jeong Y S, Byon Y J 2013 Supervised weighting-online learning algorithm for short-term traffic flow prediction IEEE Control Systems Magazine 16(3) pp 34-45
[15] He Y Y 2020 Research on Short-term Traffic Flow prediction model in Intelligent transportation system Modern Computer 16 pp 8-11