

# Research on Expressway Travel Time Prediction Based on Exclusive Disjunctive Soft Set

Qingqing Wang<sup>1</sup>, Huamin Li<sup>1,a</sup> and Weixin Xiong<sup>2</sup>

<sup>1</sup> School of Automation, Chongqing University, Chongqing, 400044, China

<sup>2</sup> School of Mechanical Engineering, Chongqing University, Chongqing, 400044, China

**Abstract.** In order to study the prediction problem of expressway travel time, due to the ambiguity and uncertainty in the road traffic system, the travel time prediction model is established based on the exclusive disjunctive soft set theory. Through the parameter reduction theory of soft set, the main influence factors are extracted, and the mapping relationship between the influence factors and the travel time is obtained through the exclusive disjunctive soft set decision system. The travel time model is established based on the soft set theory, and the travel time is calculated through the mapping relationship. The experimental results show that, compared with the BPR function model, the travel time model based on the exclusive disjunctive soft set theory reduces the prediction error and effectively improves the calculation accuracy of the travel time.

## 1 Introduction

With the rapid development of intelligent transportation system, the traffic information that can be obtained and provided is more and more abundant, and the travel time prediction is an indispensable part of the advanced highway travel information system. Facing the increasingly prominent traffic problems, dynamic traffic control and traveller behavior decision-making require highway travel time prediction. It is important to improve the accuracy of the estimation when considering the estimated value, it is of great practical significance to quantitatively describe the uncertainty of the predicted value.

The researches on travel time mainly focus on the construction of prediction models and travel time function forms, which have achieved rich research results. More Rohan applied artificial neural networks to road traffic flow prediction, which reduced the difficulty of establishing accurate mathematical models, while ensuring high prediction accuracy [1]. Mori et al. introduced the literature on travel time model and thoroughly classified existing technologies, most of the attention was focused on travel time estimation and travel time prediction [2]. Allstr M et al. combined parametric and nonparametric traffic state prediction techniques through assimilation in aggregating Kalman filtering. For nonparametric prediction, the neural network method was adopted; the parameter prediction was performed by using the cell transmission model with velocity as the state [3]. Celikoglu et al. investigated the performance of the flow model in providing efficient travel time estimation for varying flow patterns of freeway traffic by adopting the two-phase fundamental diagram [4]. Zhong et al. proposed

an online travel time prediction model with emphasis on capturing the effects caused by anomalies. A Functional Principal Component Analysis (FPCA) framework was adopted to forecast link travel times based on historical data and real-time measurements [5]. Sehyun Tak et al. proposed the new algorithm called multi-level k-Nearest Neighbor (k-NN), which was designed for predicting travel time with higher computational efficiency and prediction accuracy, and provided the data fusion method to overcome the problems of data errors that combined the traffic data from ILDs and DSRC [6]. Lu et al. proposed a piecewise function based on traffic flow and vehicle type in order to estimate the travel time function of heterogeneous flow, and used the VISSIM model of actual traffic data calibration to generate traffic data for estimating parameters [7]. Chen et al. used the data source from highway charge system to establish the prediction model based on Bootstrap to improve the reliability of point prediction in travel time [8]. Wang et al. used the K nearest neighbour algorithm to predict the short-term travel time of the expressway based on the location and time information of the vehicle entering and leaving the toll station in the highway toll data [9]. The experimental results showed that the algorithm could effectively improve the prediction accuracy of the model, but the impact of weather conditions on traffic conditions was not considered, and the historical database wasn't classified according to different weather conditions.

However, when studying the travel time, the following problems are rarely considered: the judgment of the traffic participants on the traffic state is inherently vague, not precise. Under the premise of ensuring the effective provision of information, traffic participants choose

<sup>a</sup> Corresponding author: huaminli@cqu.edu.cn

different paths by judging road congestion conditions, etc., but their judgments on road congestion conditions are qualitative, not quantitative. That is to say, the ambiguity should be fully considered for the estimation and prediction of the travel time. Soft sets are powerful theories for dealing with ambiguity and uncertainty, its advantages are not only for quantitative analysis, but also for qualitative analysis [10-12]. However, the existing soft set is mainly the aspect of complete information, but the aspect of incomplete information system needs to be studied. From the point of view of data mining, this paper introduces the exclusive disjunctive soft set to describe the incomplete information system, and quantifies the data in the incomplete information system by redefining the characteristic functions in the exclusive disjunctive soft set.

The rest of the paper is organized as follows. Section 2 introduces the basic concepts. Section 3 describes the prediction model and algorithms used in the study. Section 4 uses the numerical examples to verify the validity of the model, and the model is compared with the prediction result of the BPR function model. Finally, conclusions are discussed in Section 5.

## 2 Preliminary

### 2.1. Soft set

Molodtsov proposed the new uncertainty mathematics tool of soft set theory [13]. For the first time, he systematically introduced some basic theories and applications of soft sets, and used soft set as a general mathematical tool for dealing with uncertain, fuzzy and inaccurate objects. Soft set is a new method for solving uncertain problems that can't be solved by traditional methods, the biggest difference between the soft set and the traditional data method is that it can set different forms of parameters according to its own preferences, which can solve the problem of ambiguity and uncertainty well.

**Definition 2.1**  $(F, E)$  is a soft set on the domain  $U$ , if and only if  $F$  is a mapping of  $E$  into the set of all subsets of the set  $U$ [13].

**Definition 2.2** Let  $P(U)$  be the fuzzy set of the domain  $U$ ,  $A_i \subset E$ , and the two-group  $(F_i, A_i)$  is called fuzzy soft set defined on the universe  $U$ , where  $F_i$  is a mapping[14]:

$$F_i: A_i \rightarrow P(U) \quad (1)$$

### 2.2 Exclusive disjunctive soft set

**Definition 2.3**  $(F, E)$  is a soft set on the domain  $U$ ,  $x$  is an element of the domain  $U$ ,  $e$  is an element of parameter set  $E$ .  $\xi_x(e)$  represents the  $e$ -element feature function of soft set  $(F, E)$ , defined as

$$\xi_x(e) = \begin{cases} 1, & x \in F(e) \\ 0, & x \notin F(e) \end{cases} \quad (2)$$

**Definition 2.4**  $(F, E)$  is a soft set on the domain  $U$ , where  $F$  is the mapping,  $F: B \rightarrow P(U)$ ,  $B$  is a non-empty parameter set,  $(F, B)$  is the exclusive disjunctive soft set, if and only if  $(F, B)$  meets

- (i)  $\bigcup_{e \in B} F(e) = U$ ;
- (ii) Supposing  $x$  is an element of the domain, if

$\sum_{e \in B} \xi_x(e) > 1$ ,  $\xi_x(e)$  can be interpreted as :

- a) For each  $x$  is unknown, but  $\sum_{e \in B} \xi_x(e) = 1$ ;
- b) And  $x$  is called an indeterminate element of soft set  $(F, B)$ [15].

Both the bijective soft set and the exclusive disjunctive soft set are developed on the basis of soft set, the bijective soft set is a special exclusive disjunctive soft set. The bijective soft set doesn't contain uncertain elements, and exclusive disjunctive soft set can be used to describe an incomplete information system.

### 2.3 Exclusive disjunctive soft set decision system

Soft set theory as a new mathematical method to solve the problem of uncertainty obtains fast development, the paper introduces the exclusive disjunctive soft set, and proposes to systemize the incomplete information into the exclusive disjunctive soft set decision system, so that the exclusive disjunctive soft set can process the decision analysis based on the incomplete information system. In the process of quantification, it can make up for the information that may be missing, and comprehensively reflect the overall information of the system.

**Definition 2.5** Supposing  $(F_i, E_i)$  ( $i = 1, 2, 3, \dots, n$ ) is  $n$  exclusive disjunctive soft sets defined on the domain  $U$ , where  $E_i \cap E_j = \emptyset$ , ( $i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, n, i \neq j$ ),  $(G, B)$  is the bijective soft set defined on the domain  $U$ , and  $B \cap E_i = \emptyset$  ( $i = 1, 2, 3, \dots, n$ ),  $(G, B)$  are the soft set of decisions. Supposing  $(F, E) = (\bigcup_{i=1}^n (F_i, E_i))$ , the triplet  $((F, E), (G, B), U)$  is called the exclusive disjunctive soft set decision system defined in the domain  $U$ . The dependence of the exclusive disjunctive soft set decision system is denoted by  $\kappa$ , which is defined as[15]:

$$\kappa = \gamma(\bigwedge_{i=1}^n (F_i, E_i), (G, B)) \quad (3)$$

## 3 Travel time model based on exclusive disjunctive soft set

### 3.1. Problem formulations

In the case of the same road network conditions of the transportation system, the intuitive factors affecting the travel time are mainly road capacity, traffic flow, vehicle speed and occupancy, including: time factors (general time, peak hours, etc.), weather conditions (sunny, cloudy, rainy, etc.), traffic events (road construction, traffic failure, traffic control, etc.), etc.. The travel time is often caused by one or more of these factors, because different factors have different degrees of influence on the travel time, and the same factor has different influences on the travel time on different occasions, the more important factors are extracted and the relatively minor factors are removed. Combined with actual traffic data and road traffic conditions, according to expert experience, data and natural language are transformed into a fuzzy vector map of the domain of travel time. The parameter reduction theory of soft set is used to eliminate the redundant information and preserve the mapping features that can't be removed [16].

According to the definition of soft set, firstly establishing a mapping relationship between the influence factors of road traffic (including qualitative and quantitative factors) and the travel time, so as to quantify the fuzzy factors [17]. A fuzzy soft set  $(F, X)$  on the domain  $S = (S_1, S_2, \dots, S_n)$ ,  $S_i$  represents the power set of travel time under different traffic conditions. For the parameter set  $X$ ,  $X = X_1 \cup X_2 \cup \dots \cup X_i$ ,  $X_i$  represents the  $i$ th factor that affects travel time.

### 3.2 Construction of the travel time prediction model based on exclusive disjunctive soft set

On the basis of the above, giving the following definitions: **Definition 3.1** Universe  $U = \{T_1, T_2, T_3\}$  indicates the formula mode of travel time under different traffic conditions. There is currently no uniform regulation on the division of road traffic states, but we think it is possible to satisfy the following possibility: when the number of vehicles is less than or equal to the maximum value to ensure that the vehicle is free to travel, the maximum value indicates the actual traffic capacity of the road section, the travel time is generally stable and small, and the travel time is denoted as  $T_1$ ; when the number of vehicles exceeds the threshold, the threshold represents the traffic jam of the road section, the travel time gradually changes and increases, and the travel time is denoted as  $T_2$ ; when there is severe congestion on the road, few vehicles can enter the network, and the travel time required is longer, the travel time is denoted as  $T_3$ .  $E$  indicates the parameter set,  $E = \{E_1, E_2, \dots, E_i\}$ ,  $E_i$  represents the  $i$ th factor of affecting travel time. Here, the soft set  $(F, E)$  describes the prediction of the travel time.

The specific steps to establish the travel time prediction model are as follows:

**Step 1** Considering the uncertain and fuzzy factors, according to the parameter reduction theory of soft set, and retaining the important influence factors, and eliminating the secondary factors.

**Step 2** Constructing the exclusive disjunctive soft set decision system  $(\cup_{i=1}^n (F_i, E_i), (G, B), U)$ ,  $(G, B) = (F_j, E_j)$ .

**Step 3** Establishing the membership function and the mapping relationship between the influence factors and travel time according to the decision rules, and then quantifying the fuzzy factors to calculate the dependence  $\kappa$ ,  $\kappa = \gamma(\bigwedge_{i=1}^n (F_i, E_i), (F_j, E_j))$  of the exclusive disjunctive soft set decision system  $(\cup_{i=1}^n (F_i, E_i), (F_j, E_j), U)$ .

**Step 4** Establishing the parameter form of the soft set between the travel time and the influence factors.

**Step 5** According to the mapping relationship, predicting the travel time.

According to the mapping relationship, the following expression of the travel time is obtained,  $P(U) = \left\{ T_a^0, T_a^0 \left( \frac{Q_a - C_a}{Z_a - C_a} + e^{\alpha\kappa\beta} \right), T_a^0 \left( \frac{Q_a - Z_a + \varepsilon}{\varepsilon} + e^{\alpha(\kappa + \delta)\beta} \right) \right\}$ , corresponding to  $T_1, T_2, T_3$  of the domain  $U$ , which respectively indicate the prediction model of the travel time. For any of the above traffic conditions, the corresponding mapping relationship can be found in the power set.

According to the congestion state of the road, it is

divided into three sections according to the probability of occurrence. Traffic states serves as a classification criteria. Because different influence factors can cause different traffic conditions, which decide how long the travel time will take. In general, when  $\kappa \in [0, 0.3)$ , the prediction model of the travel time is  $T_1$ , when  $\kappa \in [0.3, 0.7)$ , the prediction model of the travel time is  $T_2$ , when  $\kappa \in [0.7, 1.0]$ , the prediction model of the travel time is  $T_3$ .

Among them,  $T_a^0$  indicates the time of travel on the free flow on the road section  $a$ ,  $Q_a$  indicates the traffic flow on the road section  $a$ .  $C_a$  indicates the actual traffic capacity of the road section  $a$ , and the threshold value  $Z_a$  indicates the traffic jam of the road section  $a$ , these two parameters can be obtained by consulting relevant experts or by calculating the arithmetic mean value of the historical record of the section.  $\kappa$  denotes the dependence of the exclusive disjunctive soft set decision system. Usually,  $\alpha = 0.15, \beta = 4.0, \varepsilon > 0, \delta > 0$ , they are the constants.

## 4 Numerical example

### 4.1. Experimental data source

The data in this paper is derived from UK Highway Network, which can help find and use open government data developed by data.gov.uk, using a variety of sources to estimate traffic flow parameters such as travel time and speed, including automatic license plate recognition (ANPR) cameras, on-board global positioning systems (GPS), and induction circuits built into the road, including average traffic time, vehicle speed, road occupancy, traffic flow, etc. In order to facilitate data selection, the data is first preprocessed. This paper chooses to use the road segment (A1 northbound at a minor junction between A167 near Washington Birtley and A692 (120014301)) as the research object of this paper, and calculates the average travel time of the two stations. Collecting 10 consecutive days of data, the collection interval was 15 minutes, and a total of 500 sets of data were collected. Taking a certain section as the predicted section, the data collected in the first 9 days of the section is taken as the learning sample, and the data on the 10th day of the section is the test sample, and the newly established prediction model is used to predict the travel time.

According to the relevant information and expert experience, and parameter reduction theory of soft set [16-17]. This paper mainly considers objective factors (traffic flow, vehicle speed), weather conditions and traffic event as the main factors affecting travel time.  $E$  represents the factor parameter set that affects the travel time,  $E = E_1 \cup E_2 \cup E_3 \cup E_4$ ,  $E_1$  indicates the quality of the weather condition,  $E_1 = \{\text{sunny, cloudy, rainy}\}$ ;  $E_2$  indicates the magnitude of the traffic flow per unit lane, the parameter set of traffic flow are defined as large, general and small;  $E_3$  indicates the magnitude of the average speed, the parameter set of average speed are defined as fast, general and slow;  $E_4$  indicates the emergency of road traffic,  $E_4 = \{\text{road construction, traffic failure, traffic control, normal}\}$ , among them, the traffic control refers to the control measures for vehicles and personnel in some or all

traffic sections for certain safety reasons, such as road and bridge construction, etc.. The membership functions of each set are formulated as shown in Table 1, Table 2, Table 3, Table 4 [18].

**Table 1.** Membership function of weather condition.

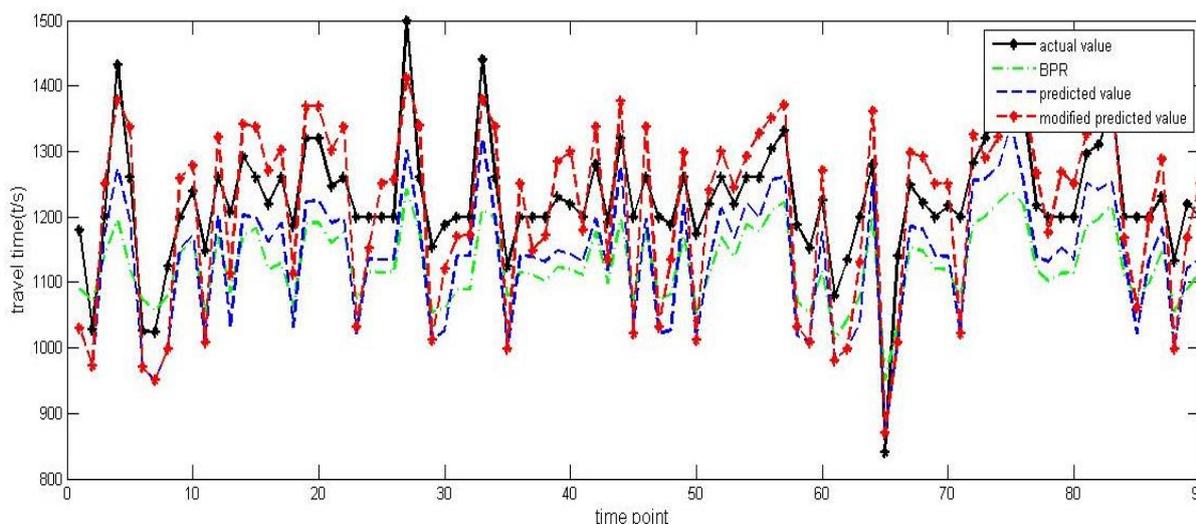
| Weather factor       | Rainy | Cloudy | Sunny |
|----------------------|-------|--------|-------|
| Degree of membership | 1.0   | 0.5    | 0     |

**Table 2.** Membership function of traffic flow.

| Language value | Traffic flow (pcu) |     |     |     |      |
|----------------|--------------------|-----|-----|-----|------|
|                | 240                | 480 | 720 | 960 | 1700 |
| Large          |                    | 0.1 | 0.7 | 0.9 | 1.0  |
| General        |                    | 0.6 | 1.0 | 0.6 |      |
| Small          | 1.0                | 0.8 | 0.2 |     |      |

**Table 3.** Membership function of average speed.

| Language value | Average speed (km/h) |     |     |
|----------------|----------------------|-----|-----|
|                | 60                   | 75  | 90  |
| Fast           |                      | 0.8 | 1.0 |
| General        | 0.5                  | 0.7 | 0.4 |
| Slow           | 0.7                  | 0.2 |     |



**Figure 1.** Comparison of results of three calculation models for travel time of road section.

**Table 4.** Membership function of emergency.

| Emergency            | Road construction | Traffic failure | Traffic control | Normal |
|----------------------|-------------------|-----------------|-----------------|--------|
| Degree of membership | 0.7               | 0.6             | 0.5             | 0      |

In summary, the dependence of the exclusive disjunctive soft set decision system was calculated. Through the mapping relationship [19] and the decision rule, the corresponding travel time prediction model was obtained from the interval range of the dependency degree. In order to verify the validity of the travel time prediction model established based on the exclusive disjunctive soft set theory, it was compared with the calculated value of the BPR function model. When only considering factors such as traffic flow and speed, comparing the prediction results of the new model and BPR function, and the result at this time was called the predicted value. When adding weather factors and emergencies of road traffic, comparing the prediction results of the prediction model, the result at this time was called the modified predicted value, and the results of the three calculation models of the road section travel time were compared as shown in Fig.1, in the figure, X-axis represented the number of samples tested and Y-axis represented the travel time, the unit was t/s.

#### 4.2. Analysis of results

Three sets of prediction experiment results were output, and the evaluation indicators were used for data analysis. The indicators for the prediction results are selected as follows:

$$MAPE = \frac{\sum_{i=1}^N | \frac{T_p - T_a}{T_a} |}{N} \quad (4)$$

$$RMSPE = \sqrt{\frac{\sum_{i=1}^N (\frac{T_p - T_a}{T_a})^2}{N}} \quad (5)$$

The two evaluation indicators are: MAPE (Mean Absolute Percent Error) is the average of the absolute error between the true value and the predicted value, which can reflect the actual situation of the prediction error; RMSPE (Root- Mean Squarer Phase Error) is the measure of the deviation between the predicted value and the observed value. Where  $T_p$  is the predicted value,  $T_a$  is the actual value, and  $N$  is the predicted number. Table 5 shows the results of the comparison of the indicators of the two optimization methods.

**Table 5.** Comparison of prediction error results.

|       | Modified predicted value | Predicted value | BPR    |
|-------|--------------------------|-----------------|--------|
| MAPE  | 0.0573                   | 0.0704          | 0.0811 |
| RMSPE | 0.0046                   | 0.0065          | 0.0075 |

From the comparison curve of travel time, the calculation results of the two models (BPR function model and travel time model based on exclusive disjunctive soft

set) can better estimate the actual travel time. However, it is easy to know through the error analysis table: Under the same conditions, predicting the travel time based on the exclusive disjunctive soft set theory, MAPE and RMSPE of the experiment reduce by 0.0107 and 0.001 respectively. Taking into account the weather and the sudden situation of the actual traffic, the modified predicted model predicts better results than the original model, MAPE and RMSPE of the experiment reduce by 0.0131 and 0.0019, it can be seen that the accuracy of the prediction is obviously improved.

## 5 Conclusions

In this paper, the influence of ambiguity and uncertainty on travel time is studied based on traffic flow data, and using the exclusive disjunctive soft set decision system and soft set theory to establish the corresponding prediction model and verify it with actual data. The experimental results show that, compared with the BPR function, the model established in this paper can improve the accuracy of prediction. Certainly, the results of this paper are only a relatively preliminary study, and there are still many aspects that need to be discussed in depth. For example, when dealing with fuzzy factors, there is inevitably some subjectivity, which leads to some deviation of the prediction results, this requires multiple observations, consultation with multiple experts, and repeated correction of relevant data to improve the prediction process and results. For the influx of massive data, whether the calculation speed can meet the requirements of real-time applications in real projects, and so on.

## Acknowledgements

The project is supported by the National Key R&D Program, China (Grant No. 2018YFB1600600).

## Reference

1. R. More, A. Mugal, S. Rajgure, et al. Road traffic prediction and congestion control using artificial neural networks. *International Conference on Computing, Analytics and Security*. IEEE. 52-57, (2016).
2. U. Mori, A. Mendiburu, M. Álvarez, et al. A review of travel time estimation and forecasting for Advanced Traveller Information Systems. *Transportmetrica A: Transport Science*. 11(2):119-157, (2015).
3. A. Allström, M.J. Ekstr, R.D. Gundleg, et al. Hybrid Approach for Short-Term Traffic State and Travel Time Prediction on Highways. *Transportation Research Record: Journal of the Transportation Research Board*. 2554:60-68, (2016).
4. H.B. Celikoglu. Flow-based freeway travel-time estimation: A comparative evaluation within dynamic path loading. *IEEE Transactions on Intelligent Transportation Systems*. 14(2):772-781, (2013).
5. R.X. Zhong, J.C. Luo, et al. An Online Travel Time Prediction Model Based on Functional Data Analysis. *Transportation Research Board 96th Annual Meeting*. (2017).
6. S. Tak, S. Kim, K. Jang, et al. Real-Time Travel Time Prediction Using Multi-Level k-Nearest Neighbor Algorithm and Data Fusion Method. *2014 International Conference on Computing in Civil and Building Engineering*. (2014).
7. Z.Y. Lu, Q. Meng, G. Gomes, et al. Estimating link travel time functions for heterogeneous traffic flows on freeways. *Journal of Advanced Transportation*. 50(8), 1683-1698, (2016).
8. J.N. Chen, X. Zhang, S.R. Zhang. Analysis and empirical study of highway travel time interval prediction based on Bootstrap-KNN. *Control and Decision*. 33(11):2080-2086, (2018).
9. X. Wang, X.H. Chen, X.M. Yang. Short-term travel time prediction of expressway based on K nearest neighbor algorithm. *China Journal of Highway and Transport*. 28(01), 102-111, (2015).
10. X. Zhi, W.J. Chen, L.L. Li. A method based on interval-valued fuzzy soft set for multi-attribute group decision-making problems under uncertain environment. *Knowledge and Information Systems*. 34(03), 653-669, (2013).
11. X. Zhi, K. Gong, Y. Zou. A combined forecasting approach based on fuzzy soft sets. *Journal of Computational and Applied Mathematics*. 228 (1):326-333, (2009).
12. Y.C. Shao. An application of fuzzy rough sets in predicting on urban traffic congestion. *14th International Symposium on Distributed Computing & Applications for Business, Engineering and Science*. IEEE. 208-211, (2016).
13. D. Molodtsov. Soft set theory—first results. *Computers and Mathematics with Applications*. 37(4-5), 19-31, (1999).
14. P.K. Maji, R. Biswas, A.R. Roy. Soft set theory. *Computers and Mathematics with Applications*. 45(4-5), 555-562, (2003).
15. X. Zhi, K. Gong, et al. Exclusive disjunctive soft sets. *Computers & Mathematics with Applications*. 59(6), 2128-2137, (2010).
16. S. Danjuma, M.A. Ismail, T. Herawan. An alternative approach to normal parameter reduction algorithm for soft set theory. *IEEE Access*. 5, 4732-4746, (2017).
17. K. Gong, P.P. Wang, X. Zhi. Bijective soft set decision system based parameters reduction under fuzzy environments. *Applied Mathematical Modelling*. 37(6), 4474-4485, (2013).
18. R. Li, C. Jiang, F. Zhu, et al. Traffic Flow Data Forecasting Based on Interval Type-2 Fuzzy Sets Theory. *IEEE/CAA Journal of Automatica Sinica*. 3(2), 141-148, (2016).
19. H.L. Yang, Z.L. Guo. Kernels and closures of soft set relations, and soft set relation mappings. *Computers and Mathematics with Applications*. 61(3), 651-662, (2011).