Eng Int Syst (2021) 5: 331–337

Engineering
Intelligent Systems

Network Awareness Data Fusion Algorithm Based on Fuzzy Time Series

Minchuan Wang*

School of Information Engineering, Zhengzhou Institute of Technology, Zhengzhou 450000, China

There are two problems with the traditional fusion algorithm: low detection accuracy and redundancy of outliers, which decrease fusion accuracy. A fusion algorithm based on fuzzy time series was proposed to distinguish speech register from network perceptual data. Based on the third-order historical data, fuzzy relations were established to classify them according to their differences. In this paper, a new prediction algorithm is proposed for the sensing of network data, and the actual data conversion of sensor nodes is determined according to certain principles. Experiments show that the fuzzy time series model has better prediction accuracy and improves the fusion accuracy.

Keywords: Fuzzy Time Series, Network-Aware Data, Fusion Algorithm, Fuzzy Relationship, Discourse Domain Division

1. INTRODUCTION

Wireless sensor networks contain thousands of sensing nodes that are deployed in a perceptual environment by random distribution. In order to enhance the robustness and accuracy of all the information, the sensor nodes must be overlaid with each other during their placement. However, this means that there is some redundancy in the data collected by the sensor nodes. Because sensor nodes have limited energy, storage space and computing power, the transmission of redundant data will consume excessive energy and shorten the lifetime of the whole network (Novelli et al., 2016; Wang and Liang, 2015). In order to avoid this problem, the sensor network needs to use the data fusion technology in the network to collect and process the data obtained from multiple sensor nodes in order to obtain more accurate and complete information.

Data fusion has several advantages: it can save energy, improve the efficiency of data collection, enhance data accuracy, and obtain comprehensive information. The advent of data fusion has taken the research focus of wireless sensor networks from an address-centric approach to a data-centric approach (Semmens et al., 2016). The data collected by the sensor nodes form a sequence in chronological order. Each sequence contains the historical data that produced the sequence system. How to find out the statistical characteristics of the corresponding systems and the laws of their development based on these sequences is well worth studying and exploring (Hollinger et al., 2015). The current algorithm ignores the trend factors in the time series, leading to a drastic reduction in the prediction accuracy of the network-aware data fusion (Nie and Wu, 2017). This study proposes a fuzzy time series prediction model for network-based data fusion, and has achieved satisfactory results (Farah et al., 2019; Norhidayu et al., 2019; Zoraiz et al., 2019).

*Corresponding Author e-mail: zzddwmc@126.com

2. NETWORK AWARENESS DATA FUSION ALGORITHM BASED ON FUZZY TIME SERIES

2.1 Discourse Domain Division and Network-Aware Data Fuzzification

From the existing research algorithms, it can be concluded that the classification of a discourse domain plays an important role in the prediction of a fuzzy time series model, and the result directly affects the accuracy of the prediction of network perception data fusion (Gaxiola et al., 2015). When classifying discourses according to their domains, a relationship analysis is conducted to improve the accuracy of the fuzzy times series prediction (Chen, 2015). The algorithm is:

Step 1: Calculate the absolute length and relative length between consecutive web-aware historical data. For the processing of historical data, the absolute length and the relative length of the data between two adjacent points is obtained using the following calculations, the calculation is:

The absolute length between two consecutive network-aware historical data is calculated with:

$$abs\_dif(t) = |x(t) - x(t-1)| \tag{1}$$

The relative length between network-aware historical data is calculated with:

$$rel\_dif(t) = |x(t) - x(t-1)/x(t-1)| \tag{2}$$

where $x(t)$ and $x(t-1)$ represent any two points of network-aware historical data, and the value of t is $t = 2, \cdots, n$.

Step 2: Calculate the absolute length and relative length of the historical data. On the basis of step 1, the mean value of the absolute length and the average value of the relative length of the entire network-aware historical data are obtained. The expressions are as follows.

The average absolute length is:

$$average\_dif = \sum_{t=2}^{n} (abs\_dif(t)) / (n-1) \tag{3}$$

The average relative length is:

$$ratio\_dif = \sum_{t=2}^{n} (rel\_dif(t)) / (n-1) \tag{4}$$

where n represents the number of network-aware historical data points.

Step 3: Determine the number of clusters. The whole historical data points are processed, and the minimum and maximum points in the historical data points are counted. $Min(x(t))$ and $Max(x(t))$ represent the minimum and maximum values in historical data; the specific number of clusters is given by the expression:

$$NUM = \frac{1}{ratio\_dif} \times \frac{Max(x(t))}{Min(x(t))} \tag{5}$$

Step 4: Improve fuzzification. The second subdivision of the existing fuzzy relationship is processed.

First, classify and organize all historical network-aware data, and count different historical network-aware data in each small discourse domain of $u_i (i = 1, 2, \cdots, NUM)$, and count the historical network-aware data number of $n_i (i = 1, 2, \cdots, NUM)$ included in each discourse domain. Then, the different discourse domains $n_i (i = 1, 2, \cdots, NUM)$ in the sensor are divided according to the number of historical network perceptions $n_i (i = 1, 2, \cdots, NUM)$ contained in the sensor. For example, if the points of historical network-aware data contained in the discourse domain u_I is n_I, then the u_I is divided twice into n_I second-level discourse domain intervals of equal length, which can be expressed as:

$$\begin{aligned}
u_I^1 &= [Min(x(t)) - spec/10 + l \times 1/n_1, Min(x(t)) \\
&\quad - spec/10 + l \times 2/n_1] \\
u_I^2 &= [Min(x(t)) - spec/10 + l \times 2/n_2, Min(x(t)) \\
&\quad - spec/10 + 1 \times 3/n_2] \\
&\vdots \\
u_I^{n_I} &= [Min(x(t)) - spec/10 + l \times (n_I - 1)/n_I, Min(x(t)) \\
&\quad - spec/10 + l]
\end{aligned} \tag{6}$$

According to the above division rules, the entire discourse domain can be divided into:

$$\begin{aligned}
u_I^1 &= [Min(x(t)) - spec/10 + l \times 1/n_I, Min(x(t)) \\
&\quad - spec/10 + l \times 2/n_I] \\
u_I^2 &= [Min(x(t)) - spec/10 + l \times 2/n_I, Min(x(t)) \\
&\quad - spec/10 + l \times 3/n_I] \\
&\vdots \\
u_I^{n_I} &= [Min(x(t)) - spec/10 + l \times (n_I - 1)/n_1, \\
&\quad Min(x(t)) - spec/10 + l] \\
u_2^1 &= [Min(x(t)) - spec/10 + l + l \times 1)/n_2, Min(x(t)) \\
&\quad - spec/10 + l + l \times 2/n_I] \\
&\vdots \\
u_{NUM}^{n_{NUM}} &= [Min(x(t)) - spec/10 + (NUM - 1) \times l \\
&\quad + (n_{NUM} - 1)/n_{NUM}, Min(x(t)) - spec/10 \\
&\quad + NUM \times l]
\end{aligned} \tag{7}$$

According to the second division, all the historical network perception data are fuzzified and the corresponding historical network perception data are fuzzified (Cheng et al., 2016), such as historical network perception data. At the same time, we can count the points of historical network-aware data belonging to the interval u_I, the historical network-aware data $x(t) \in u_I^m$ is recorded. At this time, the historical network-aware data in the sensor is marked as A_{i-1+B}, where $B = m/n_I$ represents the historical network-aware data type, and m represents the degree of membership.

2.2 The Establishment of Network-Aware Data Fuzzy Relationship

Historical data will have a great impact on the forecast results, but this does not mean that all the historical data should be

Table 1 Distribution of Logical Relationships.

Type	Relationship	Fuzzy relationship	Name
1	$I < j < k$	Ai, A_j, $A_k \rightarrow A_l$	"Up-Up" trend fuzzy logic group
2	$i < j = k$	Ai, A_j, $A_k \rightarrow A_l$	"Up-Equal" trend fuzzy logic group
3	$i = j < k$	Ai, A_j, $A_k \rightarrow A_l$	"Equal-Up" trend fuzzy logic group
4	$i = j = k$	Ai, A_j, $A_k \rightarrow A_l$	"Equal- Equal" trend fuzzy logic group
5	$i > j > k$	Ai, A_j, $A_k \rightarrow A_l$	"Down-Down" trend fuzzy logic group
6	$i > j = k$	Ai, A_j, $A_k \rightarrow A_l$	"Down-Equal" trend fuzzy logic group
7	$i = j > k$	Ai, A_j, $A_k \rightarrow A_l$	"Equal-Down" trend fuzzy logic group
8	$i < j > k$	Ai, A_j, $A_k \rightarrow A_l$	"Up-Down" trend fuzzy logic group
9	$i > j < k$	Ai, A_j, $A_k \rightarrow A_l$	"Down-Up" trend fuzzy logic group

used as the research object to establish the corresponding forecasting model. Because the impact of historical data on the predictions is very progressive, the closer the historical data is from the predicted time point, the greater is the impact on the predicted value, and the stronger is its internal relationship. In historical data with long-term forecasting, its influence on the forecasting value will be small, and can be ignored in many cases. Therefore, the amount of historical data required to establish the relevant fuzzy relationship will also be the decisive factor that affects the final forecast result. This paper considers the third-order historical data, establishes the related fuzzy relationships, and classifies the different fuzzy relationships so as to establish different adaptive models for different fuzzy relationships (Lu et al., 2015).

Step 5: Determine the fuzzy relationship group.

Any third-order fuzzy relationship can be expressed as $A_i, A_j, A_k \rightarrow A_l$. Assuming we take the fuzzy relationship:

$$\begin{gathered} A_{\chi_1}, A_{\chi_2}, A_{\chi_3} \rightarrow A_{\chi_4} \\ A_{\chi_2}, A_{\chi_3}, A_{\chi_4} \rightarrow A_{\chi_5} \\ \vdots \\ A_{\chi_i}, A_{\chi_{i+1}}, A_{\chi_{i+2}} \rightarrow A_{\chi_{i+3}} \\ \vdots \\ A_{\chi_n}, A_{\chi_{n+1}}, A_{\chi_{n+2}} \rightarrow A_{\chi_{n+3}} \end{gathered} \tag{8}$$

Nine different types of relationships can be derived from the fuzzy order relationship above, as shown in Table 1.

By studying the different fuzzy logic groups obtained by dividing the hierarchical fuzzy relationship, a corresponding adaptive relationship is established. It can effectively deal with different situations, whereas the single prediction model is not comprehensive and effective for prediction. This classification of the forecast model proposed in this paper can be adapted to different fuzzy logic groups.

2.3 Data Fusion Algorithm Based on Discourse Domain Division and Network-Aware Data Fuzzification

Traditional algorithms ignore the trend factor in time series. In this section, algorithms for the training phase and prediction phase of sensor network perceptual data are proposed. The trend values are extracted from the sensor time series to improve the prediction accuracy (Wang et al., 2016; Bas et al., 2015).

(A) Training phase

After the fuzzy relationship has been extracted, the following algorithm can be used to obtain the network-aware data fusion forecast trend:

(1) During the maximum initialization, the sequence $\omega = 5$, and the trend mark value $f = 1$;

(2) Use the first-order and second-order fuzzy relationships to predict the next data value of network-aware data and compare the accuracy. The expression is:

$$PA = |For - Act| \tag{9}$$

where *For* represents the predicted value of network-aware data fusion; Act represents the actual value of network-aware data fusion. The timing number for selecting a person with a small *PA* value is set as the trend mark value f;

(3) Predict the next fused data value of network-aware data by using the f order and the $f + 1$ order, and if $PA_f \leq PA_{f+1}$, the trend flag value $f = f - 1$, otherwise $f = f + 1$. If $f = 0$, return to step (2); If $PA_\omega \geq PA_{\omega+1}$, and ω is the next network-aware data fusion predicted mark value, the trend mark value $f = \omega - 1$;

(4) Use the trend mark value obtained in step (3) and predict the next observed value according to the fuzzy relationship of network perceived data calculated in Section 2.2;

(5) Repeat steps (3) and (4) until the network-aware data is processed.

The above algorithm can extract trend values during the training phase. Obtaining and de-blurring the predicted values in the process should follow this principle: If a rule set can point to more than one category, the predicted value is the average of these categories. For example:

Assuming there is network aware data rule cluster $A_L \cdot A_{L-1}, \cdots, A_1 \rightarrow A'_1, A'_2, \cdots, A'_p$ with L–order fuzzy relationship, then the network-aware data fusion predicted value is expressed as:

$$For = \frac{\sum_{t=1}^{p} \cdot \omega}{PA} \tag{10}$$

Table 2 Comparison of Prediction of Enrolment at Alabama University With Two Different Algorithms.

Year	Admission numbers	Proposed algorithm	Traditional algorithm	Year	Admission numbers	Proposed algorithm	Traditional algorithm
1991	13044		13418	2002	15432		
1992	13562	14189	13434	2003	15496	16365	15531
1993	13887	14425	14332	2004	15154	15648	15513
1994	14694	14568	15152	2005	15171	15649	15211
1995	15478	15587	15524	2006	15983	16523	15628
1996	15321	15647	15576	2007	16849	16234	16471
1997	15703	15664	15834	2008	18153	17069	17458
1998	15861	16101	16348	2009	18196	18326	18612
1999	16812	16189	16588	2010	18960	19002	19234
2000	16917	17088	16231	2011	19234	19003	19452
2001	16379	17106	15512	2012	18896	19002	19030
RMSE		532.08		RMSE		512.37	324.68

Table 3 Comparison of TAIEX and RMISE Prediction From 2010 to 2015 Obtained by Different Algorithms.

Year	Algorithm		RMISE
	Proposed algorithm	Traditional algorithm	
2010	93.28	102.98	108.69
2012	105.69	100.87	97.68
2013	104.32	110.35	94.10
2014	63.28	89.35	86.88
2015	48.98	54.68	79.08

Table 4 The RMSE and RMISE Predictions of TAEEX for 2006–2015 With the Proposed Method.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
RMSE	59.18	145.12	106.56	68.14	43.18	92.28	52.62	44.02	65.13	79.28
RMISE										77.34

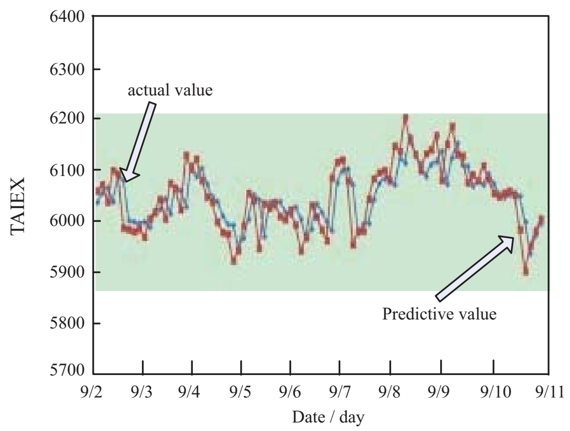

Figure 1 Real-Time and Predictive Curve of Network-Aware Data Fusion in September, 2006.

(B) Forecast phase

In this phase, the network perceived data obtained in the training phase is used to predict the perceived data of the network based on the fuzzy relationship and the trend value of the network-aware data.

First of all, according to the trend value, the fuzzy relationship of the corresponding network sensing data is dynamically selected to predict the sensor data at the next moment. If the trend value $f_t = L$ and $f_t \geq f_{t-1}$, the fuzzy relationship between L order and $L + 1$ order is selected to predict the next moment perception information, otherwise the L and the $L - 11$ order percentile relationship fuzzy relationship forecast will be used to achieve the fusion of network perception data. If $L = 1$, choose the first and second order fuzzy relationship.

Since both fuzzy orders are used in the forecasting process, the following rules are proposed to choose the final forecasting value:

If $f_t \geq f_{t-1}$, and $Act \geq Act_{t-1}$, then $For = \max(For_{f_t}, For_{f_{t-1}})$;

If $f_t \leq f_{t-1}$, and $Act \leq Act_{t-1}$, then $For = \min(For_{f_t}, For_{f_{t-1}})$; Otherwise $For = (For_{f_t},\ For_{f_{t-1}})/2$.

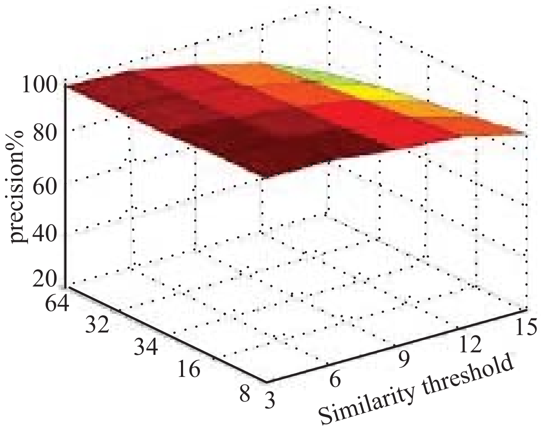

Figure 2 Mean Value of Network-Aware Data Fusion Accuracy With Different Data Sizes and Similarity Thresholds.

3. THE APPLICATION OF THE ALGORITHM AND EXPERIMENTAL RESEARCH

The proposed algorithm is compared with the existing traditional algorithms in predicting the number of students enrolled at Alabama University. In order to determine the validity of the proposed algorithm, the forecast results for TAIEX of the proposed algorithm are compared with results obtained by current algorithms.

Table 2 compares the number of students enrolled at Alabama University from 1991 to 2012, and compares the root mean square error prediction results of the proposed method with those of the traditional algorithm. It can be seen that the proposed algorithm is obviously superior to the traditional algorithm. The proposed algorithm has the smallest RMSE value, indicating that the prediction accuracy of the proposed algorithm is higher than that of the traditional algorithm.

For the TAIEX forecast, two experiments were conducted using the proposed algorithm.

The first experiment predicts the TAIEX in September 2014. The annual historical network perception data is divided into two parts. The historical network perception data from January to September are the training data, and the appropriate distance parameter ρ can be obtained through experiments to determine the best clustering number. The historical network perception data from October to November is the test data.

Table 3 compares the RMSE and RMISE values obtained by the proposed algorithm with those of the traditional algorithm for the TAIEX fusion results from 2010 to 2015. It can be seen that the proposed algorithm outperforms the existing traditional algorithms and the proposed algorithm has the smallest RMISE value. This shows that the accuracy of the proposed method in predicting network data fusion is higher than that achieved by the traditional algorithm.

The results of the second experiment are shown in Table 4, which predicts the RMSE of the predicted TAIEX fusion results over the decade from 2006 to 2015 and lists the RMISE. It can be seen that RMISE for this decade is small and comparable to the previous RMISE values obtained by both experiments, demonstrating the consistency of the proposed method.

Figure 1 shows the actual value of fusion results from September 2 to September 30 of TAIEX in 2006; the predicted value of the fusion result was obtained using the proposed algorithm. It proved that the prediction results obtained by the proposed algorithm are ideal, as well as demonstrating its validity.

The accuracy of detecting isolated points during network-aware data fusion is verified by using precision, recall, and F-measurement. The accuracy of the calculation is based on the correctness of the results of the sample; the accuracy and the recall rate indicate the accuracy of the process, and are widely used to evaluate the correctness of the classification algorithm. In experiments, the accuracy indicates the success rate of the network-aware data fusion algorithm to identify truly isolated points. The recall shows the percentage of isolated points identified by the fault-tolerant data fusion algorithm in the actual isolated points. The F-measurement is the harmonic mean of the first two.

Experimental tests were conducted on different sizes of network-aware data and similarity thresholds. By changing the size of network perception data, the probability of detecting isolated points and the efficiency of node communication are changed. Increasing the size of network-aware data reduces the probability of detecting isolated points, but increases the communication efficiency of the correspondent node (LSH code added). Similarly, an increase in the similarity threshold (without units) also reduces the probability of detecting isolated points, which seriously affects the results of network-aware data fusion.

The test results are shown in Figures 2, 3 and 4. It can be seen from the figures that the proposed algorithm can detect isolated points well with high accuracy, recall rate and F-measure mean with different numbers of network-aware data and similarity thresholds. When the size of the data is 64 bits and the similarity threshold is 15, the lowest observed mean of precision, the mean of recall rate, and the mean of F-measures can reach 0.73, 0.71 and 0.70, respectively.

Based on the above experiments, we tested the accuracy of network-aware data fusion using different similarity thresholds. In the experimental environment shown in Figure 1, the upper limit of error data sent by sensor nodes is 22%. Test results are reported in Figure 5.

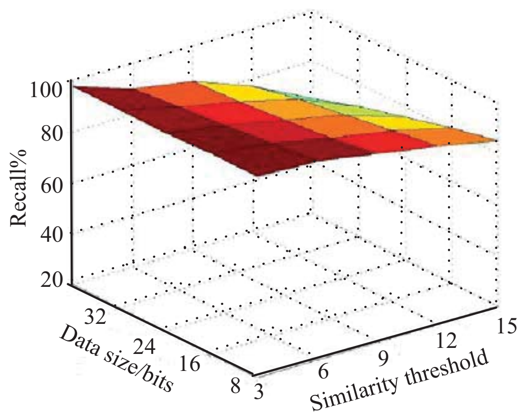

Figure 3 Mean Value of Network Perceived Data Fusion Recall Rate With Different Data Sizes and Similarity Thresholds.

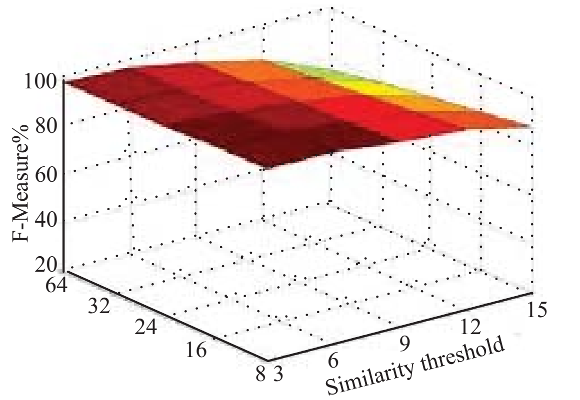

Figure 4 Mean Value of Network-Aware Data Fusion With Different Data Sizes and Similarity Threshold.

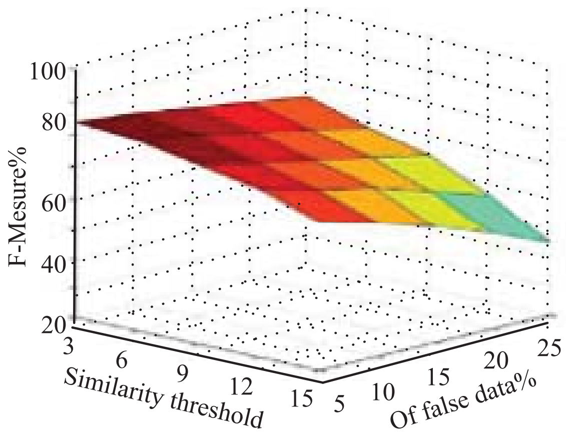

Figure 5 Mean Value of Network-Aware Data Fusion Accuracy With Different Amounts of False Data.

The accuracy of network-aware data fusion is affected by the accuracy of isolated point detection during data fusion. If all isolated points are detected during network-aware data fusion, the network-aware data convergence point will not receive any false information. That will produce fusion results with 100% accuracy. As can be seen from Figure 5, the greater the percentage of false data, the greater the negative impact on data fusion. This is because, when incorrect data in the network increases, some erroneous data will have enough minimal support to be identified as an isolated point. In addition, since the fuzzy time series prediction model used in this paper can accurately predict the fusion results of network-aware data at the next moment, the stability is better. In summary, the proposed algorithm can achieve accurate prediction, reduce the amount of data transmission errors of network-aware data, and increase the accuracy of fusion.

4. CONCLUSION

In this paper, we take into account the trend factors in the time series collected by sensor nodes, and improve the traditional time series model to make it possible to extract the trend values in the time series during the training phase. With the aid of this trend, we can improve the accuracy of detecting the network's perceived isolated points during data fusion, and eliminate redundancy. The experimental results showed that the fuzzy time series model can be used to predict the WSN-aware data with high accuracy, and highly-accurate fusion results can be obtained.

ACKNOWLEDGEMENTS

Henan Science and Technology Department Science and Technology Research Project: Wireless Sensor Network Intelligent Irrigation Gateway Design and Data Fusion Research (162102210046); Library Based PLC robot system (152102210003).

REFERENCES

1. Bas, E., Egrioglu, E., Aladag, C.H., et al. Fuzzy-time-series network used to forecast linear and nonlinear time series. Appl Intell, 2015, 43(2): 343–355.
2. Chen, M.Y., Chen, B.T. A hybrid fuzzy time series model based on granular computing for stock price forecasting. Inform Sciences, 2015, 294(2): 227–241.
3. Cheng, S.H., Chen, S.M., Jian, W.S. Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures. Inform Sciences, 2016, 27(C): 272–287.
4. Farah R.R., Hairulnizam M., Shahreen K., Mohd S.A. (2019). An Image-Based Captcha System Using Click. Acta Electronica Malaysia, 3(1): 23–25.
5. Gaxiola, F., Melin, P., Valdez, F., et al. Generalized type-2 fuzzy weight adjustment for backpropagation neural networks in time series prediction. Inform Sciences, 2015, 325(C): 159–174.
6. Hollinger, G.A., Yerramalli, S., Singh, S., et al. Distributed data fusion for multirobot search. IEEE T Robot, 2015, 31(1): 55–66.
7. Lu, W., Chen, X., Pedrycz, W., et al. Using interval information granules to improve forecasting in fuzzy time series. Int J Approx Reason, 2015, 57(1): 1–18.
8. Nie, Y., Wu, J.J. An online time series prediction method and its application. Journal of Beijing University of Technology, 2017, 3(3): 386–393.
9. Norhidayu M.Y., Nurhidayah N., Noor Faraliza S., Salasiah M.S., Nor Farmiza T. (2019). Recognizing the Personal Competencies of Future Information Professionals. Acta Informatica Malaysia, 3(1): 21–23.
10. Novelli, A., Tarantino, E., Fratino, U., et al. A data fusion algorithm based on the Kalman filter to estimate leaf area index evolution in durum wheat by using field measurements and Moderate Resolution Imaging Spectroradiometer surface reflectance data. Remote Sens Lett, 2016, 7(5): 476–484.
11. Semmens, K.A., Anderson, M.C., Kustas, W.P., et al. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. Remote Sens Environ, 2016, 185(185): 155–170.
12. Wang, S., Liang, X. A parameter estimation framework for Multiscale Kalman Smoother algorithm in precipitation data fusion. Water Resour Res, 2015, 50(11): 8675–8693.
13. Wang, W., Wang, B., Sun, J.L. Based on two layers of cross mapping large data fusion algorithm robustness. Computer Simulation, 2016, 33(10): 236–239.
14. Zoraiz N., Ahthasham S., Summaya A., Haroon K. (2019). Review Paper on Bitcoin Technology. Information Management and Computer Science, 2(1): 01–03.