Lejiang GUO[1,2], Fangxin CHEN[2] , Chao GAO[1], Wei XIONG[1]

Huazhong University of Science and Technology (1), Air Force Radar Academy (2)

Performance Measurement Model of Multi-Source Data Fusion Based on Network Situation Awareness

***Abstract**. In order to solve the problem of the energy optimization and assessment with situational awareness performance in the wireless sensor network, the function model is proposed based on the potential relationship between the trend of events and trend assumptions. Research from data integration, correlation analysis, events and prediction method, this paper analyzes the ability of the decision node to obtain and process information. It derivates the multi-sensor network situational awareness performance measurement formula, integrates network situation element, draws a trend calculation curve to achieve the network trend prediction. Experiments show that the proposed method can effectively predict the future trend of the network, the trend calculation results and time series analysis results is better than other network trend assessment algorithms.*

***Streszczenie.** Zaproponowano model sieci bezprzewodowej czujników bazujący na potencjalnych zależnościach między trendami i zdarzeniami. Artykuł analizuje możliwość decyzji węzłów do realizacji przesyłu informacji. Eksperymenty potwierdziły, że metoda pozwala na efektywne prognozowanie trendu. (**Model pomiaru możliwości wieloźródłowej fuzji danych bazującej na wiedzy o bezprzewodowej sieci czujników**)*

Keywords: Situation Awareness, Prediction, Modelling, Assessment Index, Performance Metrics.
Słowa kluczowe: prognozowanie, sieć bezprzewodowa czujników, fuzja danych

Introduction

With the extensive application of computer network and expanding the network size, the structure of network is more and more complex. The problem of the network management technology is gradually revealed. Using a single network technology to network management has been unable to meet the requirement of network, so the situational awareness came into being[1]. Data fusion as a key technology is a hot topic of the recent research, which aims to reduce the cognitive burden on the network, primarily through the integration in the primary data on the basis of information provided by the entity level, similar information aggregation, information extraction and deep-seated target gradually divided into a cluster of higher abstraction levels. Multi-source information fusion and target recognition is to support decision-making with the comprehensive utilization of computer, information processing, and artificial intelligence technology. In multi-source information fusion Multi-source information fusion and comprehensive recognition will make full use of resources, dominate by a variety of observational reasonable and complementary information, and redundant information in space and time, and produce the consistency of the observed environment explanation or description of fusion results based on some optimization Criteria[2-4]. The result is the separation of various data sources to obtain observational information, and export more effective information through the optimal combination. The network situation awareness is to evaluate system dynamic response in the entire network to predict the future development trends and provide a reliable reference. The future sensor networks need to collect and integrate a large amount of information quickly and effectively. The face requirements of a large quantity of dense, multi-source information processing have been given the higher level to provide accurate and reliable target track, the credibility of target attribute.

With the neural network to the network security situation value forecast, the neural network algorithms rely mainly on empirical risk minimization principle, easily leads to the generalization ability reducing and the model structure is difficult to determine. Network situational awareness during the real-time online prediction system, this shortcoming of the neural network has been further enlarged. Effective way to avoid the problems faced by the above algorithm in support vector regression algorithm to predict the absolute error is small, forecasting the correct trend rate can accurately predict the development trend of the network trend. Compared with RBF neural algorithm, the more conducive value is of real-time network trend prediction analysis.

The Problem Description and Model

Collection to set the network target of a moment, according to the status information for each of the target collection, the various objectives is divided into multiple disjoint cluster, described in mathematical language, that is too form a partition of the collection [5].

$\bigcup_{i=1}^{n}\psi_i = \Omega \quad \forall i \neq j, \psi_i \bigcap \psi_j = \phi$ Wherein each set corresponds to a cluster, the target set by the number of clusters. The characteristics of matrix representation $F = [f_1, f_2, \ldots\ldots, f_n]$,the physical characteristics of each component has each of the target. The objective situation i of the network at the moment network t ,the information vectors $f_i = [f_1(t), f_2(t), \ldots, f_m(t)]$ Elements in the feature vector to the target location, type, $Q(F_0, F_1) = Q(F_1 \mid F_0) Q(F_0)$ speed and other features. Make $Q(\bullet)$ to produce the quality of the information in the various processes, F_0 is characteristics of cyber, $Q(F_0) = 1$ .The information field has three sub-domains: the sensor sub-domains, the integration sub-domain and the communication network sub-domain, as shown in Figure 3, each sub-domain conversion can be used to describe the conditional probability[6].

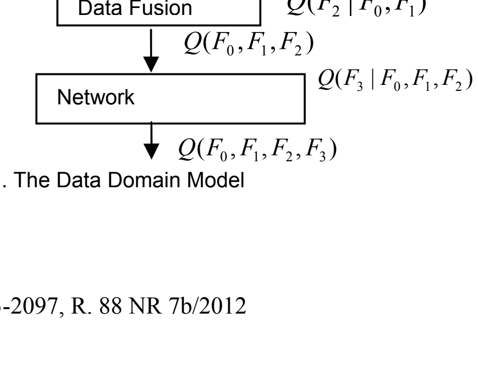

Fig.1. The Data Domain Model

1) Sensor sub-domain: $Q(F_0,F_1)=Q(F_1\mid F_0)Q(F_0)$

2) Integration of sub-domain:

$Q(F_0,F_1,F_2)=Q(F_2\mid F_0,F_1)Q(F_0,F_1)$

3) Communication sub-domain:

$Q(F_0,F_1,F_2,F_3)=Q(F_3\mid F_0,F_1,F_2)Q(F_0,F_1,F_2)$

Sensor Sub-domain Analysis Model

Completeness Objective network based time t in the number M , N is found target, from the conditional probability, $N=M\times P_M\times R(d)$, P_M is the probability of the target is not obscured, $R(d)$ is the performance function of detection range d . The completeness of the metric as follows: $Q_{com}(F_1\mid F_0)=NM=P_M R(d)$

Correctness If R_i is characteristic of i sensor, n is the total number report, the practical features is estimated to: $\mu = \frac{1}{n}\sum_{i=1}^{n} R_i$ .Based estimation $A=\mid E[\mu]-\mu\mid$, μ is true character value. A_j is to estimate the deviation of the characteristics j , the correctness measure: $Q_{\cot}(F_1\mid F_0)=e^{-\sum_{j=r}^{m}A_j}$

Timeliness Located from the sensor to detect the time between the reports of the receiver sensor fusion domain is t_s . The timeliness metric is: $Q_{cut}(F_1\mid F_0)=e^{-t}$

Integration of the Sub-domain Analysis Model

Completeness The i intelligence source in the completeness of the fusion center processing follows $c_i=a+\alpha(1-e^{-\sigma t})$, a expresses sensor itself which is integrate part of the target detection accounts percentage . $0<a<1; a+\alpha$ is indicating that the fusion centre can carry out sub-Class part of the percentage of the maximum detection points. $0<a+\alpha\le 1;\sigma$ is class speed. c_i is the intelligence source i completeness of processing in the fusion center, c_c is the facilities in the central fusion target detection report completeness, reliability theory can be complete integration of sub domains measure: $Q_{com}(F_2\mid F_0,F_1)=\left[1-\prod_{i=1}^{k}(1-c_i)\right]c_c$

Correctness $a_j=e^{-v_j}$, which V_j is the integration of information on the feature j variance, α_j indicats the relative importance of weights, β_j is weight the relative importance, i is the integration of information correctness $\sum_{j=1}^{m}a_j=1,\sum_{i=1}^{m}\beta_i=1$.

Integration sub-domain correct metric:

$Q_{\cot}(F_2\mid F_0,F_1)=\sum_{i=1}^{n}\beta_i W_i=\sum_{i=1}^{n}\sum_{j=1}^{m}\beta_i\alpha_j a_j$

Timeliness t_f is integration of information, resulting in the time required by the timeliness metric:

$Q_{cut}(F_2\mid F_0,F_1)=e^{-y}$

Communication Network Sub-domain Model

Completeness Set p_i to information received from the fusion facilities probability, k is the number of users. Assuming the independence, the probability of receiving the information without any constraints to complete network: $Q_{com}(F_3\mid F_0,F_1,F_2)=N=\prod_{i=1}^{k}p_i$.

Correctness Set all users get the correct probability $\Pr$ of trend information. $\Pr$ is the central integration of the joint probability. i is correct probability: trend information $\Pr_i=P_i(F_3,F_2)=P_i(F_3\mid F_2)P(F_2)$. $P_i(F_3\mid F_2)$ is correctly received from the central fusion facility i .It depends on the central integration of the reliability of the communication network between facilities and user. $P(F_2)$ is correct transmission probability. The correctness of communication network sub-domain measure:

$Q_{\cot}(F_3\mid F_0,F_1,F_2)=\Pr=\prod_{i=1}^{k}P(F_2)p_i$.

Timeliness Set t_n the average transmission delay networks, timeliness metric $Q_{cut}(F_3\mid F_0,F_1,F_2)=e^{-t_n}$.

Performance Metrics between Situation Awareness

Connections of the situation and trends represent dependencies between nodes. If the situation node S is independent among the phase system, it can use the layer tree structure to represent the relationship between the nodes of situation. $Bel(s)=\sum_{i=1}^{n}Bel(s_i)$ .The connection between the situation node tree networks can be seen as a single set of elements based on hierarchical assumptions. Assume that the evidence came from the event direct effect on the trend of node S , the ratio: $\lambda=P(e\mid S)/P(e\mid -S)$, λ reflects the extent of the evidence e to support the situation S . After evidence, the confidence of the trend nodes is updated as follows: $Bel'(s)=\alpha\lambda Bel(S)$ .The normalization factor: $\alpha=[\lambda Bel(S)+1-Bel(S)]^{-1}$.

Update the node degree of confidence; S spread the following message to its neighboring and the parent node: $m_1^+=Bel'(S), m_2^+=\alpha$, S descendants of node (such as Z), the confidence level of the parent node (X) and the brother node (Y) are updated as follows:

$$Bel'(S)=m^-Bel'(Z)$$

$$Bel'(X)=m_2^+[Bel(X)-Bel(S)]+m_1^+$$

$$Bel'(Y)=m_2^+Bel(Y)$$

Update the confidence of the node X , then its parent node spread the news: $(m_1^+)'=Bel'(X),(m_2^+)'=m_2^+$, The connection between the situation nodes can be seen as the level network, evidential reasoning algorithm can be applied to multi-level assumptions. The ability to multi-sensor network of cooperation effect is the key to improve cooperation after the situational awareness of decision-making node. Situational awareness refers to the various elements of the environment within a specific time and space perception, understanding and prediction of the elements of meaning elements state. It is the integration of the results of the current comprehension, the previous decision-makers the knowledge, experience and the actual situation. The results of the situational awareness of decision makers is not only related with the trend of the amount of information received, so the decision makers ability to understand. Therefore, the situational awareness of the decision node includes the ability to obtain information and information processing capabilities of the decision node. Suppose the network of cooperation in decision-making node situational awareness effectiveness of its information retrieval performance and information processing performance is proportional to the product that:

$U = U_{acq} \bullet U_{awa}$, U is the decision node in the cooperative network performance, U_{acq} is the performance of situational awareness and information of the decision node , U_{awa} is the decision nodes in the information processing performance, the three functions are continuous function of time t . U_{acq} and U_{awa} will be quantitative analysis. Before analysis, it is necessary to do the following four basic assumptions.

The Design of Prediction Model and Dynamic Comprehensive Evaluation Method

Situational awareness system operates in real-time online. Each processing data collection time interval calculates the value of network security posture to form a time series data sets. The prediction model first normalized data set recently for a fixed period of time data, on this basis construct the sample data, sample data is divided into N -dimensional vector. Predict the function of the first m -dimensional input vector after the n -dimensional output vector ($N = m + n$), a set of (m, n) constitutes a training data. Then initialize the model training parameters, through the formation of the training data to train the prediction model. Frequent data collection refers to the frequency exceeds a certain threshold of data in the data set. Suppose a collection of data for the number of data items N , and support given $s(0,1)$, all occurrences of more than a data item s_N is called frequent item. Mining frequent data collection has a wide range of applications in the areas of database, data mining and network. Mining data streams frequently is a challenging problem. Unlimited features of the data stream, the algorithm can only scan data again. By again scanning the data stream to calculate the precise frequency storage space $O(M)$, M is the collection of all the different values in the data stream may be infinite. Thus, accurate mining frequent item in the data stream is usually impossible. Therefore, it is focused on mining frequent data stream approximation algorithms[7-8]. Mining data streams frequently approximation algorithm is characterized by one pass, using only a small storage space, query results, although the real results of the algorithm is an approximation, but the algorithm can guarantee the error between the approximate query results and actual results does not exceed the user-specified interval range. The design idea of the algorithm is as follows:

(1)Assume that the degree of error u and support s , these two values are taken into account specified.

(2) S is given when query is issued, it can be $(u,1)$ to any value.

(3) The algorithm save the data stream $1/u$ samples, along with the continuous arrival of the data stream data dynamically to the principle of maintaining the sample collection.

(4) When the query support S is frequent, the algorithm uses the sample collection gives the approximate results and provide analysis of use to the next step.

The algorithm includes the statistical analysis and query processing o. When the data stream of the arrival of new data, statistical analysis algorithm is responsible for maintaining the sample set D ; When a user makes frequent item queries, query processing algorithm is responsible for output query results. Before prediction, firstly data cleaning, analysis, data transformation and the statute of pre-processing step to improve the accuracy of the forecasting process, effectiveness and scalability[9]. Prediction methods can compare and evaluate with accuracy, speed, robustness, type, scalability, and interpretability criteria. In order to further improve the prediction accuracy, it needs to assess the accuracy of the predictor. Assessment is based on the random division of the accuracy of the given data, commonly uses technology to keep the random sub-sample, cross-validation.

Dynamic Comprehensive Evaluation Method

For three different times of the comprehensive evaluation such as the history, status and future, it is necessary to establish the corresponding period of the evaluation index system. Without loss of generality, the evaluation of the different groups can be used in each period, followed by $x_1, x_2, , x_{mk}$ $(k = 1,2,3)$ and $\{r_{ij}\}(i = 1,2,,n;\ j = 1,2,,m_k;\ k = 1,2,3)$

(1) on the s_1 historical performance evaluation. On the time interval $[k_0, k_0 + T - 1]$ in the past, it takes a dynamic evaluation function as follows: $u_i(k) = \sum_{j=1}^{m} w_{ij}(k) r_{ij}(k)$ $k = k_0, k_0 + 1,, k_0 + T - 1$

Where $w_j(k)$ is the weighting function, $w_j(k) \geq 0, \sum w_j(k) = 1, u_i(k)$ expressed the comprehensive evaluation value of s_1 at the moment k , $r_{ij}(k)$ is the evaluation of the value x_j at the moment k , T is a known positive integer. Characterize the overall operation level of the system s_1 in the time interval $[k_0, k_0 + T - 1]$ is given.

$$y^{(1)} = \frac{1}{T-1} \sum_{k=k_0}^{k-\gamma-1} \sum_{j=1}^{m} w_j(k) r_{ij}(k) \quad i = 1,2,\cdots,n$$

(2)Evaluation on t s_i status, the whole time given that the status $k_0 + T$, in order to run n status of the system $s_1, s_2, \cdots, s_n$ to distinguish, by the Scatter method to determine the weight coefficient μ_i , in the moment

$k = k_0 + T$, the status quo is represented by the following formula: $y^{(2)}{}_i = \sum_{j=1}^{m} \mu_j r_{ij} (k_0 + T) \quad i = 1,2,\cdots,n$

(3)The evaluation of the development trend, it make N appropriate known constants, the next time interval $(k_0 + T, k_0 + T + N)$, if the evaluation of s_i the expected average x_i or r_{ij} , $j = 1,2,\cdots,m$ when the future development trend s_i can be used $y^{(3)}{}_i = \sum_{j=1}^{m} \rho_j r_{ij}$ $i = 1,2,\cdots,n$, ρ_{ij} is the corresponding weight coefficient.

(4) Dynamic Comprehensive Evaluation. The system s_i set History, status, future in one of the dynamic value of comprehensive evaluation by the following formula is given by: $y_i = \sum_{j=1}^{3} \lambda_j (y_i^{(j)} - y^{*(j)})^2 \quad i = 1,2,\cdots,n$.

Respectively, where $y^{*(j)}$ is the ideal value $y^{(j)}$, or $y^{*(j)} = \max_{1 \le i \le n} \{ y_i^{(j)} \}$, $j = 1,2,3$, $\lambda_1, \lambda_2, \lambda_3$ is given the right weight coefficient in advance, and $\sum_{j=1}^{3} \lambda_i = 1$, according to the value y_i of small to large to be sorted or classified, in order to achieve n system dynamic $s_1, s_2, \cdots, s_n$ evaluation purposes.

Experimental Results

OPNET (optimization Network Engineering Tool) is of network performance simulation analysis tool that can be used to design and simulation of network protocols, and the establishment of local and wide area network model. OPNET Modeller can integrate the network system, information network group to build a variety of data links, wireless communication networks and wireless sensor networks, subnets, and simulation experiments. In Fig.2 and Fig.3, it is the experimental results respectively under high load, the large-scale case, and normal circumstances.

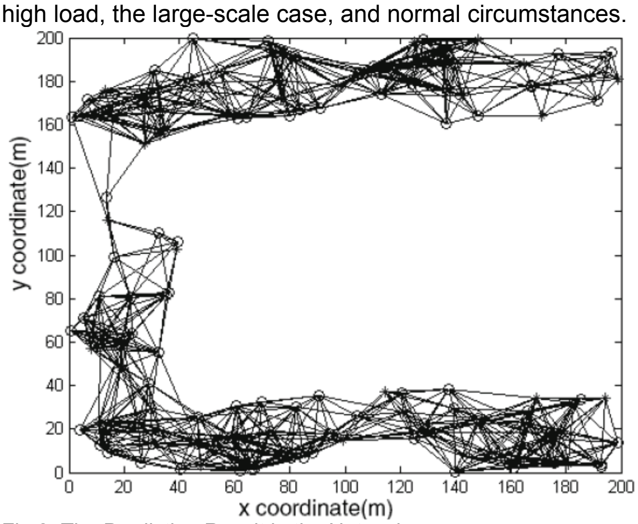

Fig.2. The Prediction Result in the Network

According to the simulation under normal circumstances, It can see the performance of the system simulation in five minutes before running low, mainly due to the link between the various network groups networking, network groups and network groups to establish the start of each unit; after a period of warm-up, the system gradually stabilized, the combined effectiveness situation information and tactical information curve light Micro-shocks, in the performance of trend information, due to the limited distribution of the system or network congestion, distribution capabilities to reduce. Decline in the high-load large-scale case, the system performance, it increases warm-up event, and situation information, Information and command and control information shocks. The comparison of both experimental cases, the information resource evaluation method has a certain sensitivity and rationality in order to better validate the method. The correctness of this paper build a simulation environment based on the integrated network operations system in the same network size. Through the normalized average performance, it compared to the results of the assessment of the information resource evaluation method and the classic method, such as the experimental results in Fig.3

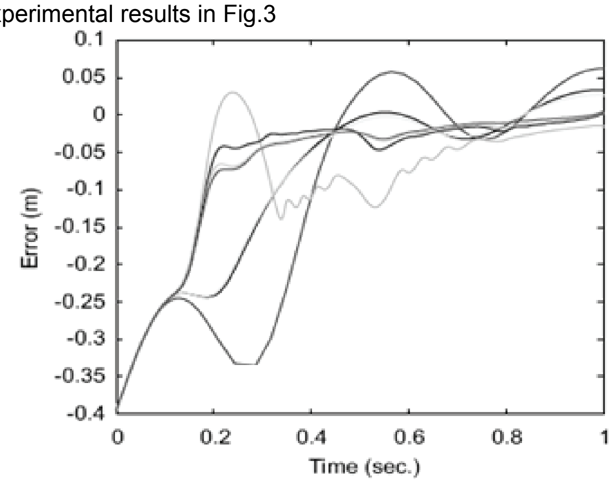

Fig.3.The Error comparison of algorithm

From the experiments, the equivalent of network size and the number of units, respectively, the average performance of the information resources is assessed; the different ways to get the results is normalized. From the experimental results, it calculates the proposed information resource assessment methodologies close to the results with the results calculated by the classical the methods, indicating that the method is effective. From the experiment, the proposed method is better in the general case in classic method. The reasons of this phenomenon is mainly to further enhance the synergy of collaboration capabilities, the information gain and the combat units in the system is the next step is to be solved. Experiments show the system construction method is reasonable, the performance of its information resources is considerable, but also illustrates its information resources, evaluates the correctness and validity of the method.

Conclusion

Network situational awareness system as a new field of study has very important significance to improve the emergency response capacity of the network. Many experts and research institutions have been significant research in this field and try situational awareness for the large-scale network. On the basis of the analysis of system functional requirements, multi-source sensor network and situational awareness system and it gives the detailed system-level conceptual model and all levels of functional entities design. Therefore, how to reduce the amount of computation, increase calculation speed is still current need to solve the problem. Massive real-time processing of network data, situation assessment and prediction of the accuracy, trend visualization, and other aspects also need to make a long-term and meticulous research. In addition, it is clear from the overall level, the level of components and component relationships to guide engineering practice and the key technology to further develop reduction algorithm of situation assessment index system.

REFERENCES

[1] Scholtz, J.C.,Antonishek, B.,Young, J.D. Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces.IEEE Transactions on Systems and Humans. 4 (2005),450-459.

[2] Roman R,Lopez J, Gritzalis, S.Situation awareness mechanisms for wireless sensor networks.Communications Magazine, 4 (2008), 102-107

[3] Ellerbroek J, Visser M,van Dam S.B.J,Mulder M,van Paassen M.M.Design of an Airborne Three-Dimensional Separation Assistance Display. IEEE Transactions on Systems and Humans, 5 (2011), 863-875

[4] Yongmian Zhang, Qiang Ji.Efficient Sensor Selection for Active Information Fusion. IEEE Transactions on Cybernetics, 3 (2010) ,719-728

[5] Holsopple J, Sudit M, Nusinov M, Liu D,Haitao Du,Shanchieh Yang. Enhancing situation awareness via automated situation assessment. Communications Magazine, 3 (2010) , 146-152

[6] Ellerbroek J,Visser M,van Dam S.B.J,Mulder M,van Paassen M.M. Design of an Airborne Three-Dimensional Separation Assistance Display.IEEE Transactions on Systems and Humans, 5 (2011), 863-875

[7] Pederson T, Janlert L.-E,Surie, D. A Situative Space Model for Mobile Mixed-Reality Computing.Pervasive Computing, 4 (2011), 73-83

[8] Anagnostopoulos C, Hadjiefthymiades S.Enhancing Situation-Aware Systems through Imprecise Reasoning.IEEE Transactions on Mobile Computing, 10 (2008), 1153-1168

[9] Valera M, Velastin S.A,Ellis A,Ferryman J.Communication Mechanisms and Middleware for Distributed Video Surveillance.IEEE Transactions on Circuits and Systems for Video Technology, 12 (2011), 1795-1809

***Authors**: dr Lejiang Guo, a Senior IEEE member, the department of control science and engineering, Huazhong University of Science and Technology, Wuhan, 430074, China. E-mail: radar_boss@163.com. Prof Fangxin Chen, the department of Early Warning Surveillance Intelligence, Air Force Radar Academy, Wuhan,430019, China. E-mail:pxial@msn.com. dr* Cao Gao, *the department of control science and engineering, Huazhong University of Science and Technology, Wuhan, 430074, China. E-mail: leezbaby@163.com. dr* Wei Xong,*the department of control science and engineering, Huazhong University of Science and Technology, Wuhan, 430074, China. E-mail: xiongwei@163.com.*