Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 15, no. 3, March 2019, pp. 1014-1022
DOI: 10.23940/ijpe.19.03.p31.10141022

Target Recognition and Behavior Prediction based on Bayesian Network

Chao Lin[a] and Yanan Liu[b,*]

[a]*The 44 unit of PLA 92941 Army, Huludao, 125000, China*
[b]*China Research Institute of Radiowave Propagation, Qingdao, 266000, China*

Abstract

The identification of target identity attributes and its behavioral prediction are important means for providing command and decision support in modern warfare. This paper analyzes the key steps in the process of target recognition and behavior prediction and proposes a target recognition and behavior prediction model based on the Bayesian network. In the simulation example, by integrating the sensor information and combining expert knowledge, the model can effectively and accurately conduct battlefield situational awareness. Combined with the background of big data, this paper introduces the distributed processing system Hadoop, and prospects its application in target recognition and behavior prediction.

Keywords: target identification; behavioral Prediction; Bayesian network; big data

(Submitted on November 5, 2018; Revised on December 6, 2018; Accepted on January 3, 2019)

1. Introduction

Modern war is a war of information. Both sides spare no efforts to acquire intelligence by using advanced reconnaissance techniques. Battlefield target identification and behavior prediction are important foundations for realizing the design of the command system, the disposition of weapons and equipment, the management of personnel allocation and the research of strategy and tactics. It can provide the commanders with quick and accurate decision basis and support [1].

The complex battlefield electromagnetic environment is poor, and the received information is often vague and uncertain. All above situations make the target recognition and behavior prediction an uncertain multi-source fusion reasoning process. The Bayesian Network (BN), based on probability theory and graph theory, is a model for uncertainty expression and probability inference [2]. Laurutzen improved it and laid a foundation for the practical use of BN [3]. In 1992, the Hugin software developed by Aalborg University was the first commercial software of Bayesian network, which greatly promoted the use of BN [4].

Bayesian networks are widely used in various fields such as fault detection, industrial control, medical diagnosis, and the military [5]. In the military field, Shi Zhifu used BN to fuse the multi source sensor information and achieve comprehensive recognition of target aircraft, which effectively utilizes information complementarity of a single sensor [6]. Ma Zhijun established the equipment damage model based on BN and verified the feasibility and effectiveness of Bayesian networks in equipment damage location [7]. Laskey studied the field of situation estimation and proposed a Bayesian network model for situation assessment [8]. Based on the Bayesian Network, Daniel.W.F proposed a model for operational damage assessment [9]. The model can be used to evaluate the target damage effect in real time synthetically, with all kinds of prediction, the collected information and expert knowledge considered.

2. Basic Theory about BN

In the process of simulating the uncertainty representation, reasoning and learning of causality, the Bayesian network provides a method to visualize the knowledge diagram. BN is a directed acyclic graph consisting of two parts. Part of the

* Corresponding author.
E-mail address: lc19851225@126.com

qualitative information, that is, the network structure, uses nodes and edges to represent the reality of things and their associations; The other part, that is, network parameters, represents the quantity of correlation degree between variables through the conditional probability table [10].

In the Bayesian network, nodes without direct connection are conditionally independent. In this case, the joint probability can be decomposed into the following Formula (1):

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \tag{1}$$

Conditional independence greatly reduces the complexity of the probabilistic model in the computation process. Assuming all the variables in the model 'Rain' have two different states, the joint probability will be as follow:

$$\begin{aligned} P(C,S,R,W) = P(C) \times P(S \mid C) \times \\ P(R \mid C,S) \times P(W \mid C,S,R) \end{aligned} \tag{2}$$

In order to calculate the upper formula, a total of 13 independent probabilities are required. However, conditional independence is applied to break it down.

$$\begin{aligned} P(C,S,R,W) = P(C) \times P(S \mid C) \times \\ P(R \mid C) \times P(W \mid S,R) \end{aligned} \tag{3}$$

Joint probability can be obtained by only nine independent probabilities. Compared with the original, conditional independence can greatly reduce the computational complexity of Bayesian networks.

Bayesian network inference refers to the calculation of posterior probabilities of interested variables after giving evidence variables according to Bayes formula:

$$P(Y \mid X) = \frac{P(X,Y)}{P(X)} \tag{4}$$

However, Bayesian network inference is an NP-hard problem [11], and therefore derives a series of reasoning algorithms. Exact inference is a very important method in common inference algorithms, such as the joint-tree algorithm [12], variable elimination algorithm [13] and so on.

The Bayesian network parameter learning is a process to learn the conditional probability distribution from a given sample data under a certain structure, that is to update the original prior distribution of the network variables. Commonly, the maximum likelihood estimation algorithm [14] and the expected maximum value algorithm [15] are effective tools.

The Bayesian network structure learning refers to the use of training sample sets to determine the appropriate Bayesian network topology combining prior knowledge as much as possible. Currently, structure learning algorithms are mainly divided into three categories: methods based on dependency analysis [16] methods based on scoring search [17].

3. Bayesian Network for Target Identification and Behavior Prediction

Combat Identification (CID) is the foundation for understanding the battlefield situation and seizing the opportunity of combat. CID and behavioral prediction are the prerequisites for the situation and threat assessment (STA). Its significance lies in combining the observed target events and actions with the real-time environment of the battlefield, and through certain information processing methods, to provide fast, effective, and accurate data support for the battlefield command.

3.1. BN Model

In 1995, Endsley proposed a concept of situational awareness, which refers to the perception of situational elements within a certain time and space environment, and the understanding of the information obtained, which in turn forms the state of the momentary state of these situational elements[18]. He put forward the situational awareness model as shown in Figure 1.

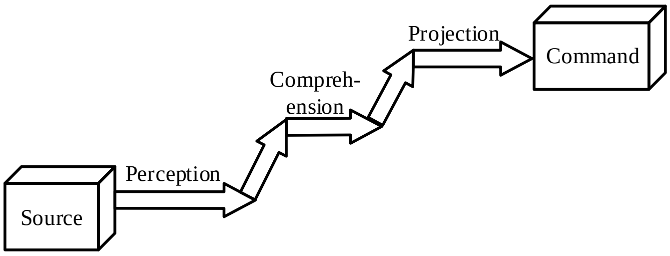

Figure 1. Endsley situational awareness model

Referring to Figure 1. model, the Bayesian network-based target recognition and behavior prediction model can be divided into three levels. First, preprocess the sensor information. Then, get the recognition result as well as the prediction using BN inference block. Finally, show them. The overall framework of the system is shown in Figure 2.

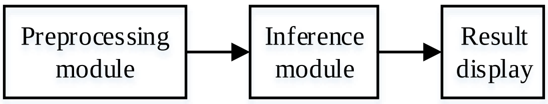

Figure 2. Battlefield situational awareness system model

Among them, the sensor data processing module is mainly to discretize the continuous value of variables such as the flight speed and altitude based on prior knowledge of statistical data. Here, taking the speed parameter as an example. The target speed fuzzy variable set is defined as: speed = {fast, medium, slow}, and the speed interval is given in Table 1.

Table 1. The true value of the target speed

Slow (m/s)	Medium (m/s)	Fast (m/s)
≤ 150	150 ~ 450	≥ 450

The triangle wave function is used to represent it with two centers 150m/s and 450m/s respectively, and the deviation is taken to divide the speed into regions. The membership function of the velocity can be expressed as follows:

$$\text{Slow Range: } y=\begin{cases}1, & x\leq 120\\ 2-\dfrac{1}{60}x, & 120\leq x\leq 180\end{cases}$$

$$\text{Normal Range: } y=\begin{cases}\dfrac{1}{60}x-2, & 120\leq x\leq 180\\ 1, & 180\leq x\leq 360\\ 3-\dfrac{1}{180}x, & 360\leq x\leq 540\end{cases}$$

$$\text{Fast Range: } y=\begin{cases}\dfrac{1}{180}x-2, & 360\leq x\leq 540\\ 1, & x\geq 540\end{cases}$$

Figure 3 is the discrete curve of the formula above. The overlapping part of the two discrete values is expressed in the form of soft evidence. Literature [19] introduces the use of soft evidence in Bayesian network inference.

The Bayesian network inference module is the core of the entire battlefield target recognition and behavior prediction system, which directly affects the production of command decision. Therefore, it is very important to build a precise network. This paper constructs a BN based on the JDL model (The Joint Directors of Laboratories information fusion model) proposed by the United States Department of Defense Laboratory joint leadership [20-21]. The JDL model is a 5-layer information fusion model, as shown in Figure 4.

In the JDL model, layer 0 processes signal information, which generally refers to the data acquired by the sensor, such as the speed and other parameters. Layer 1 is an object-level division. Specifically, it can identify the type, platform, or even the class of the target. Moreover, layer 2 usually tells you whether the target is friendly or not. The last level is to evaluate the threat, that is, to predict the next state of the target. The fourth layer is mainly an auxiliary processing of information.

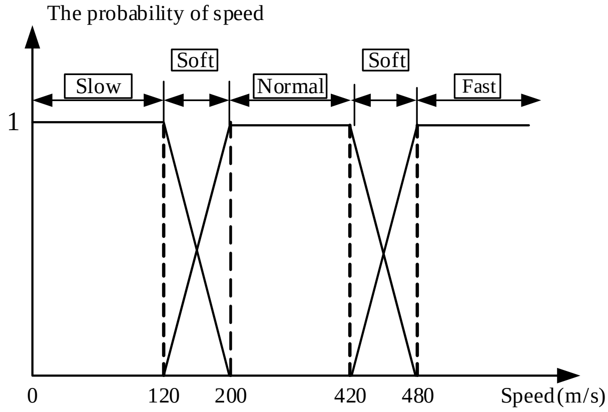

Figure 3. Velocity parameter preprocessing threshold quantization function

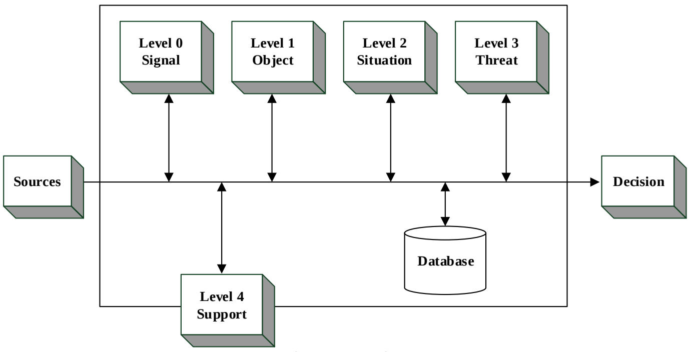

Figure 4. JDL information fusion model

Nowadays, typical intelligence resources are radar, ESM, enemy-used sensors and data links. The results of target recognition and behavior prediction may be different levels including target category, platform, type, country, and enemy-matter attributes [22]. In other words, building a battlefield target recognition and behavior prediction system is a multi-level decision problem. Combing the JDL model, expert knowledge, and historical data, this paper establishes the following BN model shown in Figure 5.

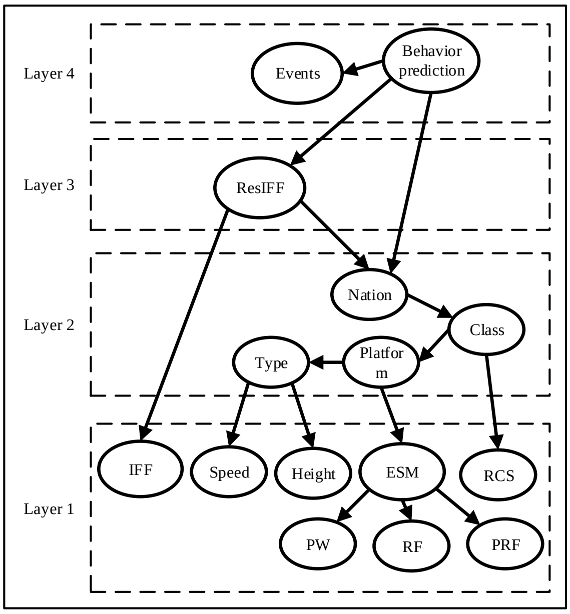

Figure 5. Target identity recognition and behavior prediction model

It should be noted that data preprocessing corresponds to Layer1 in the model and gives the discrete results for follow-up work. In Layer2, radar reconnaissance and communication reconnaissance are usually used to determine the motion attributes and radiation characteristics of targets. In the process of measuring the motion attributes, the type of aircraft is determined mainly based on the speed and altitude. The observation variables in the radiant feature reconnaissance process are often radar carrier frequency (RF), the pulse recurrence frequency (PRF) and the pulse width (PW). Once these parameters are obtained, the radar type can be matched. Generally speaking, aircraft platform has a one-to-one corresponding radar. In addition, the radar cross-sectional area (RCS) plays a very important role in determining the type of aircraft. In the third layer, the enemy and self-attributes are determined through the use of country information and the use of airborne identification of friend or foe (IFF) system. Finally, at the behavioral forecasting level, the results above work. Besides, temporary military exercises, hostile relations in wartime, and interactions in peacetime and other specific events are considered.

3.2. Bayesian Network Parameter Settings

Parameters in BN reflect the degree of dependence between various nodes and affect the confidence in the process of information transmission.

For the root node, its prior probability is usually determined by statistical data. If there is no prior information, it is generally distributed according to equal probability principle; For other nodes, the parameter means conditional probability. The determination of parameters requires the integration of various aspects of knowledge, including statistical knowledge and expert experience, etc. This paper first lists the state set of each node in BN, shown in Table 2.

Table 2. Air battle situation awareness node state set

Layer1	Speed	(Fast, Normal, Slow)	Altitude	(High,Medium, Low)
	PW	(Wide, Normal, Narrow)	PRF	(More,General, Less)
	RF	(HF, MF, LF)	IFF	(Yes, No)
	RCS	(Large, Medium, Small)		
	ESM	AN/APG-81, AN/APG-78, AN/APG-77, AN/APG-70, MMCR, AN/APG-68, AN/APG-63, N-019M, N011M, AN/APQ-181, APEAJIET, others		
Layer2	Type	(Fighter, bomber, helicopters, Transport)		
	Nation	(America, Russia, Japan, South Korea, India, Malaysia)		
	Platform	F-22, F-35, F-15, F-16, Mig, Su, B-2, Tu-160, AH-64, Ka-52, C130, IIyushinII-76		
	Class	F-35A, F-22A, F-15E, F-15J, F-15K, F-16C/D, Mig-29, Su-30, B2-A, Tu160, AH-64D, Ka-52, C130, IIyushinII-76		
Layer3	ResIFF	(Friends, supposed friends, general, skeptical, hostile, uncertain)		
Layer4	Temporary events	(Joint military, Wartime, Peacetime)		
	Behavior prediction	(Active, Harassing, Silent)		

Conditional probability values come from statistical knowledge, combined with expert experience. For example, for the determination of Type, count the total number of every type of aircraft, and then regard the number of occurrences of this type as the numerator and the total number as the denominator. The ratio is preliminarily identified as CPD. Then, it will be modified moderately according to the expert's judgment on the situation and other military information reports. After the initial probability determined, inference can be performed to make the information flow in the network so as to realize the recognition and prediction.

3.3. Simulation

The simulation experiment adopts Netica software to establish the model and calculate an inference result. Netica supports the joint-tree algorithm and the input of soft evidence and hard evidence. Assume that under the global pattern of peaceful development and local conflicts, an unknown aircraft intends to enter. After the target is locked, it is found to have a speed of 480m/s and a height of 11235m. Through the preprocessing module, soft evidence information can be obtained: soft-speed (fast: 0.7; normal speed: 0.3; low: 0.0), soft-height (high: 0.8; medium: 0.2, low: 0.0). The evidence is shown in Table 3.

In addition to detecting the above information, IFF system gives a negative response. Combing the current international situation, that is, the theme of peace development, target recognition and behavior prediction system shows the CID result and conducts behavioral prediction. The identity recognition curve is shown in Figure 6.

Table 3. Evidence table at various times

T	Speed	Height	RF	PW	PRF	RCS
0	(.7,.3,.0)	(.8,.2,.0)	--	--	--	--
1	(.8,.2,.0)	(.8,.2,.0)	--	--	--	--
2	(.8,.2,.0)	(.8,.2,.0)	--	--	--	--
3	(.9,.1,.0)	(.8,.2,.0)	(.2,.7,.1)	--	--	--
4	(.9,.1,.0)	(.9,.1,.0)	(.1,.8,.1)	(.1,.2,.7)	--	--
5	(.9,.1,.0)	(.9,.1,.0)	(.1,.9,.0)	(.1,.1,.8)	(.8,.2,.0)	--
6	(1.,.0,.0)	(.9,.1,.0)	(.1,.9,.0)	(.0,.1,.9)	(.9,.1,.0)	(.1,.2,.7)
7	(1.,.0,.0)	(1.,.0,.0)	(.0,1.,.0)	(.0,.1,.9)	(.9,.1,.0)	(.0,.2,.8)
8	(1.,.0,.0)	(1.,.0,.0)	(.0,1.,.0)	(.0,.0,1.)	(1.,.0,.0)	(.0,.1,.9)
9	(1.,.0,.0)	(1.,.0,.0)	(.0,1.,.0)	(.0,.0,1.)	(1.,.0,.0)	(.0,.0,1.)

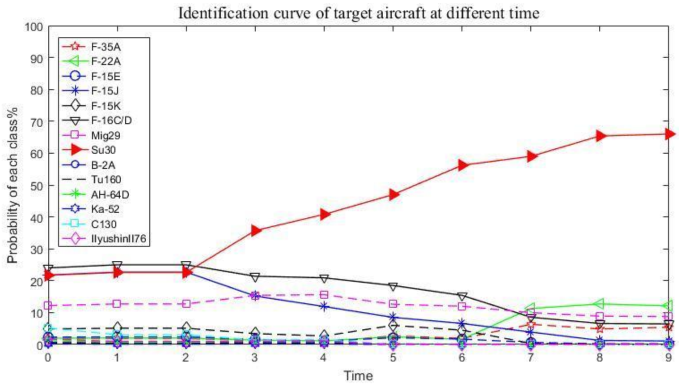

Figure 6. Target aircraft identification curve

Figure 7 shows the results of Netica's software for target identification and behavioral prediction.

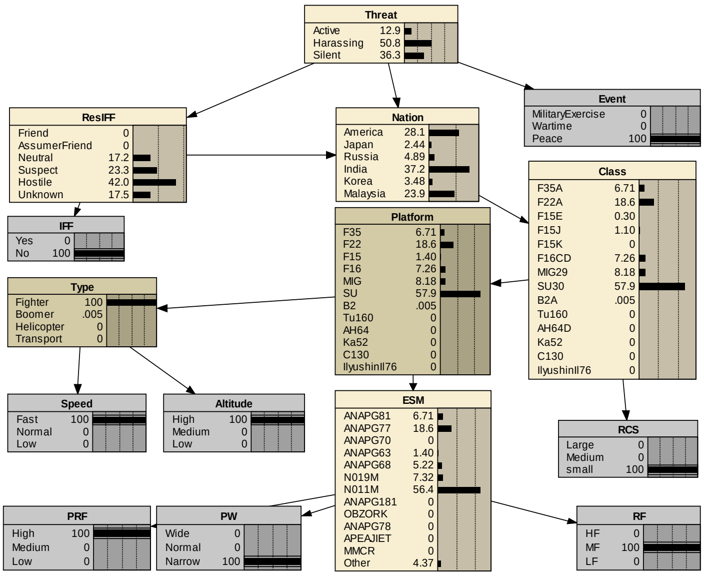

Figure 7. Target identification and behavior prediction results

From the results, it can be seen that the type of this target is most likely the Su-30. Investigation shows that this series fighter is mainly equipped in Russia, India, Malaysia. With the help of the IFF system, it was learned that the country attribute is most likely to be India. During the peacetime, this fighter is predicted to have the highest possibility of harassment, followed by silent retreat. The above simulation results are in line with the actual situation, indicating that the Bayesian network-based target recognition and behavior prediction model can effectively fuse information and make full use of complementary information to improve inference results. This system can provide data support for battlefield command and decision making with practical significance.

4. Bayesian Network Target Recognition in Big Data Background

With the continuous development of modern science and technology, the electronic reconnaissance techniques improved a lot, and reconnaissance methods have become more and more abundant. The reconnaissance data presented the following characteristics: (1) Diversified sources, including real-time data from communication devices, radar, and IFF, etc., as well as external system data such as artificial lines, sea/air situation information and images. (2) Large quantity, with increasing storage technology and the accumulation of time, the amount of data is increasing, usually counted in TB or even PB; (3) It has a complex format, the data format is not only including the structure one, but also including the semi-structured, even it will include the unstructured type [23-24].

Big data technology can enhance the function of command and control as well as ensure the performance of the information and decision [25]. In depth fusion of big data and battlefield recognition and prediction system, it is necessary to collect and process data from different data sources. How to store and deal with mass data, so as to make more accurate, rapid and effective system is a difficult problem.

The traditional data processing model adopts a centralized data computing pattern. However, in the face of an explosively increasing complex type structure data, this method cannot achieve efficient storage management, and there is a significant shortage of computational efficiency. Therefore, it is necessary to use distributed computing technology.

Hadoop is a relatively common distributed computing framework. It was created by Doug Cutting, the founder of the Apache Lucene project. Its design ideas originate from the Google File System [26] and Map-Reduce [27]. In [28], the various components of Hadoop are described detaily. The core of Hadoop are HDFS (Hadoop Distributed File System) and MapReduce programming model.

HDFS is a distributed file system that provides a highly fault-tolerant and high-throughput massive data storage solution with features such as self-healing, high scalability, high reliability, and low cost. HDFS adopts a Master/Slave architecture, as shown in Figure 8. Among them, NameNode is the master node, responsible for task scheduling and namespace management. The DataNode is a slave node and handles data as the user or master node ask.

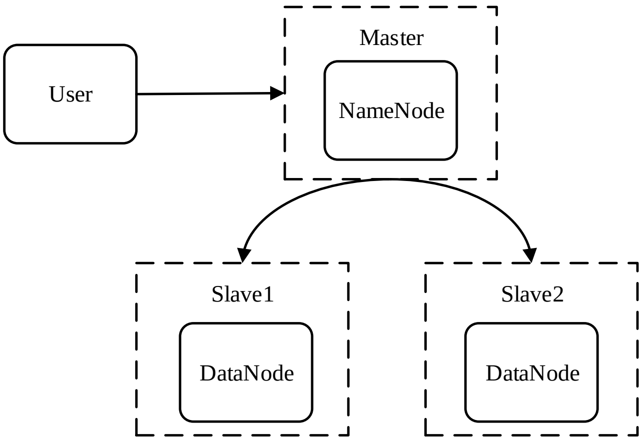

Figure 8. HDFS basic framework

MapReduce is a software framework. Spilt segmentation is performed on a task first. Then, the segmentation will be mapped to corresponding key values and sorted through shuffle. Finally, reduction operations are performed to achieve parallel data processing. The basic framework is shown in Figure 9.

Massive data make traditional stand-alone BNs hard to work, and problems such as poor storage capacity, slow calculation speed, and inflexible processing come. The combination of Hadoop and Bayesian networks can solve the above problems. In the process of using Hadoop to deal with BN problems, how to convert the operations in BN inference into the corresponding Map and Reduce functions becomes the key point.

In [29], a MapReduce model is used to implement the joint-tree algorithm. In the upward collection and downlink distribution, the computation of cut sets' marginal probability can be regarded as the Map process, and the update of cluster's potential energy is considered as Reduce. In [30], the author used the MapReduce model to implement the EM algorithm, where the BN is decomposed into several factors. Literature [31] proposed an analysis framework that converts the Bayesian network model and MapReduce model to each other, and pointed out that the MapReduce model has better processing ability for semi-structured and unstructured data. Document [32] applies the distributed computing framework MapReduce to the Bayesian network parameter learning of incomplete data, and uses different mappers to handle the observed different states of the nodes so as to get an average value. Then, use the reducer to accumulate each value to achieve a classical expectation maximization algorithm.

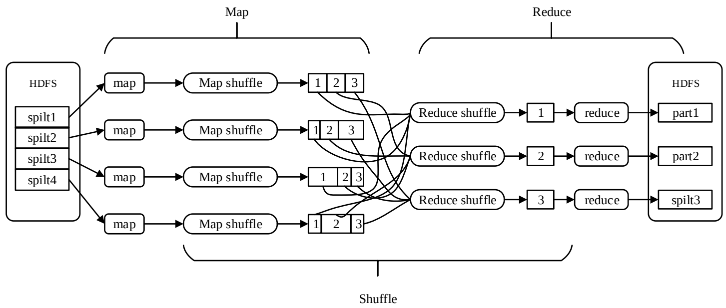

Figure 9. MapReduce processing flow

5. Conclusions

This paper studies the application of the Bayesian network model in target recognition and behavior prediction in battlefield environment. The JDL model is for reference to build the system. Simulation experiments were conducted with real air combat data. Simulation results show that the model effectively integrates various sensors and makes full use of real-time information and expert knowledge. It can provide command personnel with a quick and accurate decision support.

This paper also focuses on the application of Hadoop in the reconnaissance process with big data. We talk about the core parts of Hadoop and summarize the previous implementation of Bayesian network based on the Hadoop platform. How to effectively build a Hadoop-based Bayesian network model for better use in battlefield situational awareness will be the next step.

References

1. M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, Vol. 37, No. 1, pp. 32-64, 2016
2. J. Pearl, "Fusion, Propagation, and Structuring in Belief Networks," Elsevier Science Publishers Ltd., 1986
3. S. L. Lauritzen and D. J. Spiegelhalter, "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems," *Journal of the Royal Statistical Society*, Vol. 50, No. 2, pp. 157-224, 1988
4. Hugin Expert, "The Leading Decision Support Tool," (https://www.hugin.com)
5. J. Liu, "Bayesian Network Inference on Risks of Construction Schedule-Cost," in *Proceedings of International Conference of Information Science and Management Engineering*, pp. 15-18, IEEE Computer Society, 2010
6. Z. Shi, Y. Guo, and Y. Li, "Intelligent Decision-Making Modeling for Unmanned Aerial Vehicle based on Probability Graphical Models," in *Proceedings of IEEE Fifth International Conference on Advanced Computational Intelligence*, pp. 304-308, IEEE, 2012
7. Z. J. Ma, Q. Shi, and B. Li, "Battle Damage Assessment based on Bayesian Network [C]," in *Proceedings of Eighth Acis International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, IEEE*, pp. 388-391, 2007
8. G. Laskey, W. Navy, Y. Dc, et al., "Combat Identification with Bayesian Networks," *Combat Identification with Bayesian Networks*, 2002
9. D. W. Franzen, "A Bayesian Decision Model for Battle Damage Assessment," Master's Thesis, Helsinki University of Technology, 1999
10. J. C. Ramirez, G. Munoz, and L. Gutierrez, "Fault Diagnosis in an Industrial Process using Bayesian Networks: Application of the Junction Tree Algorithm," in *Proceedings of Electronics, Robotics and Automotive Mechanics Conference*, pp. 301-306, IEEE Computer Society, 2009
11. G. F. Cooper, "The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks (Research Note)," Elsevier Science Publishers Ltd., 1990
12. P. P. Shenoy and G. Shafer, "Axioms for Probability and Belief-Function Propagation," Readings in Uncertain Reasoning, Morgan Kaufmann Publishers Inc., pp. 169-198, 1990
13. N. L. Zhang and D. Poole, "A Simple Approach to Bayesian Network Computations," 1994
14. M. C. Golumbic, "Algorithmic Graph Theory and Perfect Graphs," Academic Press, pp. 207-208, 1980
15. S. L. Lauritzen, "The EM Algorithm for Graphical Association Models with Missing Data," *Computational Statistics & Data*

Analysis, Vol. 19, No. 2, pp. 191-201, 1995
16. J. Cheng, D. A. Bell, and W. Liu, "Learning Belief Networks from Data: An information Theory based Approach," *Artificial Intelligence*, Vol. 137, No. 1-2, pp. 325-331, 1997
17. G. F. Cooper and E. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, Vol. 9, No. 4, pp. 309-347, 1992
18. M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, Vol. 37, No. 1, pp. 32-64, 2016
19. A. B. Mrad, V. Delcroix, S. Piechowiak, et al., "Understanding Soft Evidence as Probabilistic Evidence: Illustration with Several Use Cases," in *Proceedings of International Conference on Modeling*, *Simulation and Applied Optimization*, pp. 1-6, IEEE, 2013
20. R. Pan, Y. Peng, and Z. Ding, "Belief Update in Bayesian Networks using Uncertain Evidence," in *Proceedings of IEEE International Conference on Tools with Artificial Intelligence*, pp. 441-444, IEEE, 2006
21. A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL Data Fusion Model," in *Proceedings of SPIE - The International Society for Optical Engineering*, Vol. 3719, pp. 430-441, 1999
22. T. M. Schuck, J. B. Hunter, and D. D. Wilson, "Multi-Hypothesis Structures and Taxonomies for Combat Identification Fusion," in *Proceedings of 2004 Aerospace Conference*, Vol. 3, pp. 2026, IEEE, 2004
23. X. Q. Cheng, X. L. Jin, Y. Z. Wang, et al., "A Literature Review on Big Data System and Analysis Technology," *Journal of Software*, Vol. 9, pp. 1889-1908, 2014
24. S. O. Fadiya, S. Saydam, and V. V. Zira, "Advancing Big Data for Humanitarian Needs," *Procedia Engineering*, Vol. 78, pp. 88-95, 2014
25. Fazal-e-Amin, A. S. Alghamdi, I. Ahmad, and T. Hussain, "Big Data for C4i Systems: Goals, Applications, Challenges and Tools," in *Proceedings of Fifth International Conference on the Innovative Computing Technology* (*INTECH 2015*), pp. 89-93, Galcia, 2015
26. S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google File System," *Acm Sigops Operating Systems Review*, Vol. 37, No. 5, pp. 29-43, 2003
27. J. Dean and S. Ghemawat, "Map Reduce: Simplified Data Processing on Large Clusters," ACM, 2008
28. T. White, "Hadoop - The Definitive Guide 4e," Hadoop: The Definitive Guide, O'Reilly Media, Inc., pp. 1-4, 2015
29. N. Ma, Y. Xia, and V. K. Prasanna, "Parallel Exact Inference on Multicore using MapReduce," in Proceedings of the 2012 IEEE 24^{th} International Symposium on Computer Architecture and High Performance Computing, pp. 187-194, 2012
30. Y. Zhao, J. Xu, and Y. Gao, "A Parallel Algorithm for Bayesian Network Parameter Learning based on Factor Graph," in *Proceedings of IEEE International Conference on Tools with Artificial Intelligence*, pp. 506-511, IEEE Computer Society, 2013
31. N. Jongsawat and W. Premchaiswadi, "Solving the NP-Hard Computational Problem in Bayesian Networks using Apache Hadoop MapReduce," in *Proceedings of International Conference on ICT and Knowledge Engineering*, pp. 1-5, IEEE, 2013
32. A. Basak, I. Brinster, X. Ma, et al., "Accelerating Bayesian Network Parameter Learning using Hadoop and MapReduce," in *Proceedings of International Workshop on Big Data*, *Streams and Heterogeneous Source Mining*: *Algorithms*, *Systems*, *Programming Models and Applications*, *ACM*, pp. 101-108, 2012