Entity Recognition Based on Knowledge Graph in Air Defense Domain

Zhigang Guo¹, Liuyang Yu²*, Gang Chen³, Xingxin Zhang⁴, Han Wei⁵, Yongwang Tang⁶

University of Information Engineering, Zhengzhou 450000, China

*Email: keep_er@163.com
²bcg20140801@stu.xjtu.edu.cn

Abstract. Air defense safety refers to the state in which the civil air transportation system is free from the threat or damage of illegal interference or disruptive behavior. It forms two important parts of aviation safety together with flight safety. It is an important part of civil aviation safety. This paper proposes an entity recognition algorithm based on knowledge graph, which aims to solve the problem of target recognition in the air defense field by combining pattern recognition and inconsistency verification methods based on knowledge graph. The effectiveness of the method is proved by verifying the air defense data in the simulation scenario. This method provides new research ideas for threat discovery in our country's air defense field.

1. Introduction

Air defense safety refers to the state in which the civil air transportation system is free from the threat or damage of illegal interference or disruptive behavior. It forms the two important parts of aviation safety together with flight safety. It is an important part of civil aviation safety. At present, most of the current air traffic control data organization methods are based on relational databases, which have the problem of "information overload and lack of knowledge". Although the amount of available data is large, the utilization rate of data is low due to the lack of a reasonable organizational structure, and the value of data is not fully utilized. Therefore, how to make full use of the existing information technology in the environment of big data to continuously improve the air defense security capability is a practical problem that needs to be solved urgently. As a structured semantic network, knowledge graph can accurately and vividly express the complex knowledge of the objective world in a graphical way[1]. It is a better way to organize and manage data. At present, it has been widely used in the fields of intelligent search, intelligence analysis, intelligent question answering, personalized recommendation, etc. By introducing knowledge graph technology into the air defense security field, the problems of poor aggregation, visualization and low utilization of traditional data organization methods can be effectively improved.

In this paper, by using the knowledge graph as a method of data organization and management, it can give full play to its ability to effectively integrate scattered and isolated information and reason about new knowledge between data. At the same time, in view of the target recognition problem that needs to be paid attention to in the air defense field, this paper proposes an entity recognition algorithm based on knowledge graph, combined with knowledge graph-based pattern recognition and...
inconsistent verification methods, aiming to maximize the advantages of knowledge graph. This paper proves the effectiveness of the method by analyzing the air defense data in the simulation scenario.

2. Related work

In recent years, with the continuous development of knowledge graph, pattern recognition and inconsistency verification technology based on knowledge graph has gradually matured. Although a lot of papers are focused on specific fields or applications, their methods and frameworks have certain commonalities. Such as the pattern recognition method based on knowledge graph proposed in literature[2] and the data inconsistency detection method based on rule map proposed in literature[3]. Literature[4] proposed a graph-based anomaly detection method. In the context of its proposed application, it also proved that graph-based anomaly detection is effective for the discovery of special nodes. Literature[5] presents a model which incorporates biomedical knowledge graph, graph embedding and deep learning methods for literature-based discovery. It suggests that incorporating knowledge graph and deep learning methods is an effective way for capturing the underlying complex associations between entities hidden in the literature. Many scholars have tried to convert the target recognition problem to the target recognition problem based on the knowledge graph. For example, Literature[6] proposed a target recognition method based on the knowledge graph, and simulated it for typical application scenarios.

In this paper, we first constructed a domain knowledge graph based on the actual situation of target recognition in the air defense field. A brief description of the graph is shown in Table 1. And based on the characteristics of the knowledge graph, we formulated a series of reasoning rules and finally a target recognition algorithm based on the knowledge graph was obtained, which realized the recognition and detection of flying targets.


| Table. 1 Brief description of knowledge graph |
|---------------------------------------------|
| **Flights** | **E-737** | Rate of Climb, Number of Engines, Maximum Speed, etc. |
| **Weapons** | AIM-120 Air-to-Air Missile | Guidance Method, Range, Power Unit, Country, etc. |
| **Electronic Systems** | RF-5000 Shortwave Radio | Communication Frequency Band, Working Mode, Modulation Mode, Equipment Type, etc. |
| **Communication Systems** | AN/ASN-141 Inertial Navigation System | Working Frequency Band, Interval Pulse, Equipment Type, etc. |
| **Countries** | United States | Name, Country, etc. |
| **Missions** | Daily Flight | Name, Threat Level, etc. |
| **Places** | Air Base | Name, Country, etc. |

3. Methodology

There are generally two assumptions about the knowledge graph: the closed world assumption: any relationship not represented by the existing triples is regarded as an error. And the open world assumption: the non-existent triples are explained is unknown. In this paper, we first simulate and generate multidimensional heterogeneous information, such as time and space information (including latitude, longitude, time, altitude, etc.), radiation source operating parameter information (including radar signals and their parameters, communication signals and their parameters, etc.), text information of the target flight plan, etc. Since the data is generated by simulation, the range of data generation can be restricted, so we can assume that the knowledge graph in this paper is closed.

In order to describe the statistical learning model of the knowledge graph, this article defines the following symbols:

(1) \( E = \{e_1, \ldots, e_N\} \): The collection of all entities and attributes in the knowledge graph;

(2) \( R = \{r_1, \ldots, r_N\} \): The collection of all relationships in the knowledge graph;
(3) \( x = (e_i, r_j, e_k) \): A certain triple extracted from the simulation data that may exist in the knowledge graph;
(4) \( y = 0 \) or \( 1 \): Binary random variable, the value is 1 when \( x \) exists, otherwise it is 0;
(5) \( X = \{x_1, x_2, \ldots\} \): The set of triples that may exist in the knowledge graph extracted from simulation data.

Our goal is to infer the name of the entity to be completed based on the set of all triples \( X \) to be completed, which belongs to the link speculation problem. We use the probability method to solve it.

**Step 1:** First extract the to-be-completed triples \( x_1, x_2, \ldots \) from the simulated simulation data that may exist in the knowledge graph, and assign these triples a probability of being in the knowledge graph based on expert knowledge, namely \( P(y=1) \).

For example, during a simulation process, the system receives the following data:

- **Data (1)**, Text message: The CDG2408 aircraft taking off from Japan will arrive on time;
- **Data (2)**, Radar signal (parameter): frequency = 210MHz, pulse interval pulse=0.5s, ...;
- **Data (3)**, Communication data (parameters): frequency = 20.8MHz, modulation = ASK, ...
- **Data (4)**, Track time and space data: including time, longitude, latitude, altitude, heading angle, etc.

The following triples can be extracted from the received data, where \( M \) is the node to be completed and the target to be identified and speculated:

1. \( "M1—Execution—Civil Aviation Mission" \) is a triple to be completed, namely \( x_1 \);
2. \( "APG-77—frequency_range—210MHz", "APG-77—pulse—0.5s", "APG-77—load—M2". \) And \( "APG-77 radar—load—M2" \) is a triple to be completed, namely \( x_2 \);
3. \( "HF-9500—frequency_range—20.8MHz", "HF-9500—modulation—ASK", "HF-9500—load—M3". \) And \( "HF-9500—load—M3" \) is a triple to be completed, namely \( x_3 \).

It can be seen from this example that not all the data can be directly mapped to a triple, such as data (4). Therefore, we divide the received data into two categories: for data that can be mapped to triples, perform step two, and for data that cannot be mapped to triples, perform step 3.

**Step 2:** After obtaining the above triples, we use the likelihood function \( L(M) \) to transform the prediction problem into an optimization problem. Specifically, first, we query the knowledge graph according to the elements in \( E \) and \( R \) involved in \( X \), and obtain the elements to be completed, which are \( M_1, M_2 \) and \( M_3 \). Then we obtain the confidence degree (score) of \( M_1=M_2=M_3=M \) according to the predefined score function \( G(M) \). In order to find \( M \) and its confidence, we use the likelihood function:

\[
f(x_1, \ldots, x_n \mid M) = f(x_1 \mid M) \times \ldots \times f(x_n \mid M) \tag{1}
\]

\[
L(M \mid x_1, \ldots, x_n) = f(x_1, \ldots, x_n \mid M) = \prod_{i=1}^{n} f(x_i \mid M) \tag{2}
\]

Then:

\[
M = \text{arg max } L(M \mid x_1, \ldots, x_n) \tag{3}
\]

In this example, \( x=3 \) in formulas (1) (2) (3). So far, we have inferred that \( M \) is a aircraft of A1 based on the received part of the multidimensional heterogeneous simulation data, and the confidence degree of the event \( S: "M \) is an A1 type aircraft" is \( p_1 \).

**Step 3:** For data that cannot be mapped to to-be-completed triples, we perform graph-based data inconsistency verification operations. Inconsistent verification includes four main steps:

1. Constructing the rule map;
2. Selection rules: Start from the top-level rules of the rule map to select the rules applicable to the current situation.
3. Conclusion inference: Inference is made through the rules selected in (2) and a certain conclusion is reached. Generally, we assume that the \( m \) selected rules are \( r_1, r_2, \ldots, r_m \). And the obtained \( m \) estimates are \( v_1, v_2, \ldots, v_m \). Then we use D-S evidence theory to infer \( (v_1, v_2, \ldots, v_m) \), and draw conclusion \( Y \).
4. Inconsistent verification: Judge whether the conclusion \( Y \) in (3) is consistent with the conclusion \( S \) in step 2. The main task of this step is to select an appropriate metric for the closeness
between the new inference result and the inference result of step 2, which is represented by the threshold \( T \) in this paper. If the proximity between them meets the given error range, we consider the new inference result to be consistent data, namely, \( P(S=Y) = 1 \).

For example, in the above example, we first infer that the flight target is \( M_4 \) based on expert knowledge and the confidence degree is \( p_2 \). At this time, \( Y \) is "\( M_4 \) is an A2 aircraft".

Then:
(1) If \( M_4 = M \) and \( p_2 \) is greater than the threshold \( T_1 \), then the event "\( M = M_4 \)" can be considered true, and the event "\( M \) is an A1 aircraft" is true;
(2) If \( M_4 \neq M \) and \( p_2 \) is less than the threshold \( T_2 \), event \( Y \) can be ignored, and the event "\( M \) is an A1 aircraft" is considered true;
(3) If \( M_4 \neq M \) and \( p_2 \) is greater than the threshold \( T_2 \), the event records "\( M_4 \) is an A2 aircraft" and "\( M \) is an A1 aircraft" are retained.

So far, we have used the pattern recognition and inconsistency verification technology of the knowledge graph and realize the identification and judgment of the target.

4. Results
In order to test the effect of the algorithm, we modify the code of the data simulation module and the recognition algorithm module during the test, and directly generate the target result in the simulation module. After the target is confirmed, we directly compare the difference between the identification result and the real result and get the test results. The test results are shown in Table 2. The data in the table are average values. The indicators are as follows:

- **Target Magnitude \( T \):** the magnitude of the total number of targets to be judged generated by the system for each test;
- **Data Magnitude \( D \):** The data level of all information generated by the system for each test;
- **Accuracy Rate \( P \):** the number of targets successfully identified by the system / the number of targets finally determined by all systems;
- **Recall Rate \( R \):** the number of targets successfully identified by the system / the number of targets to be identified by all input systems;
- \( F_1 = \frac{2 \times P \times R}{P + R} \).

| \( T \) | \( D \) | \( P \) | \( R \) | \( F_1 \) |
|---|---|---|---|---|
| \( 10^1 \) | \( 10^4 \) | 0.967 | 0.870 | 0.916 |
| \( 10^2 \) | \( 10^5 \) | 0.944 | 0.850 | 0.894 |
| \( 10^3 \) | \( 10^6 \) | 0.943 | 0.840 | 0.889 |
| \( 10^4 \) | \( 10^7 \) | 0.941 | 0.823 | 0.878 |

5. Conclusion
This paper proposes an entity recognition algorithm based on the knowledge graph by using map-based pattern recognition and inconsistent verification algorithms for entity recognition of the knowledge graph. And the feasibility of the algorithm is proved by verifying the simulation data in the simulation scene. This paper makes good use of the application value of the knowledge graph in the field of air defense, and also provides new research ideas for threat discovery in the field of air defense in our country.

References
[1] Zhao Z, Han S, So I. Architecture of knowledge graph construction techniques. International Journal of Pure and Applied Mathematics. 2018; 118(19):1869-1883.
[2] Jia B, Dong C, Chen Z, et al. Pattern Discovery and Anomaly Detection via Knowledge Graph[C].
[3] Ma J, Lu J, Zhang G, et al. A Rule-Map based Technique for Information Inconsistency Verification[C]. conference on decision and control, 2007: 296-301.
[4] W. Eberle, L. Holder, and B. Massengill. 2012. Graph-based anomaly detection applied to homeland security cargo screening. In Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference[C]. 382--387.

[5] Sang S. A Knowledge Graph based Bidirectional Recurrent Neural Network Method for Literature-based Discovery[C]// 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018.

[6] Knowledge Graph Application for Battlefield Airborne/Maritime Target Intelligent Recognition[J]. 2019, 010(003):1-5,24.