Feature Representation and Organization Method for Public Opinion Big Data Based on Association Analysis

Pengju Wang¹, Huifeng Xue¹ and Feng Zhang²

¹School of Automation, Northwestern Polytechnical University, Shaanxi, China
²School of Information Engineering, Yulin University, Shaanxi, China

*Corresponding author e-mail: wangpengju@nwpu.edu.cn, 416848742@qq.com

Abstract. In order to improve the data mining level of public opinion big data, by analysing the relationship among semantic web, ontology and RDF (Resource Description Framework), the structure and organization method of public opinion big data association data are separated. Based on the analysis of network public opinion big data organization technology, and then the RDF structure definition of network public opinion knowledge map is given, and the construction method of RDF association analysis map of network public opinion knowledge is realized. Experimental research shows that the proposed method can provide more accurate correlation analysis of public opinion big data, which can effectively realize the analysis and management of public opinion hot events, and provide decision support for the management of network public opinion.

Keywords: Public Opinion; Big Data Fusion; RDF; Data Mining; Association Analysis

1. Introduction

Currently, BBS forum, blog, microblogging, social networking sites and mobile Internet social media representatives showed rapid changes [1]. The popularity of the Internet, the proliferation of Internet users, and the widespread use of Internet technology have subverted the previous pattern of public opinion dissemination, resulting in the weakening or absence of the traditional "gatekeeper" mechanism, which has a profound impact on society. With the help of network technology, the reality that everyone can speak, the Internet is becoming a popular tool for public opinion [2]. Internet monitoring is faced with a variety of information content, explosive growth of information volume, rapid spread of diffusion, coupled with its virtually, concealment, divergence, permeability and randomness, fragmentation of language and other characteristics, and easy to cause The gathering of groups and the exaggeration of the events, therefore, emergencies in public opinion will pose a threat to social public safety and have a huge impact on society [3].

Public opinion big data feature representation and organization method is complex system engineering, involving many theories and methods of system engineering [4]. Through the description and definition of structured, semi-structured and unstructured public opinion data, analyze the organizational structure of related data of public opinion big data, build a multi-source heterogeneous public opinion big data RDF metadata model, and then study the logical structure of related data [5, 6].
Realize the data specification description of the multi-source heterogeneous public opinion big data from the microscopic level, use the Meta model to reveal, share and link the fusion method of the public opinion big data from the macro level, and solve the correlation analysis resource description of the multi-source heterogeneous public opinion big data Problems with semantic links [7, 8].

This paper starts with the relevance of public opinion big data, analyzes the internal relationship between "granular computing" and "ontology", and studies the global semantic consistency and semantic relevance description framework based on multi granularity pattern discovery and fusion, then the decision-oriented granular computing and ontology-related multi-granular decision-making model are designed, and then the multi-granular semantic fusion method of big data suitable for different decision-making applications is proposed, and the multi-granular response mechanism of human-machine collaborative decision-making is clarified.

2. Analysis of the Relevant Situation of Public Opinion Management Decision under the Angle of Data Granulation and Knowledge Discovery

2.1. Definition of Public Opinion Management Decision Model
In the process of data granulation, ontology semantics, granular and granular computing are related to the general process of decision system, in the process of analysis, the general decision system can be composed of \( D, F, V, f \), expressed by \( S \), denoted as \( S = (D, F, V, f) \), in data granulation and knowledge discovery, it can be transformed into an ontology integration to facilitate the unified expression of semantic aspects. Where, \( D, F, V \) is a finite set of multiple types of data in the decision-making process, \( f \) is the related reasoning function, which can be applied to the axiomatic function of ontology. In the process of data granulation, the elements in \( D \) are all kinds of multi-source heterogeneous big data, which need to be granulated; \( F \) is the attribute set of big data and \( F = M \cup N \), \( M \cap N = \emptyset \), the elements in \( M \) are called conditional attributes, the elements in \( N \) are called decision attributes, these defined attributes can be used to use the concepts in the ontology, \( M \) and \( N \) can be mapped into relationships in ontology technology; \( V \) is called the value domain of attribute, and its elements are called attribute values, which can be used to represent instances in ontology modeling. By using the granular computing method to decompose the decision-making system, the decision-making in the decision-making system is transformed into the subsystem, which is convenient to realize the knowledge discovery modeling of big data based on the granular mechanism [9, 10].

In the decision rules, for the decision system \( S = (D, F, V, f) \), since \( F = M \cup N \) and \( M \cap N = \emptyset \), the \( M \) and \( N \) elements are conditional attribute and decision attribute set respectively, so the decision system is also recorded as \( S = (D, M \cup N, V, f) \), to further decompose \( S = (D, M \cup N, V, f) \) according to decision rules, set \( M = \{m_1, m_2, ..., m_m\}, N = \{n_1, n_2, ..., n_n\} \), to take two sets of attribute values \( V(v_1, v_2, ..., v_v) \) and \( V(q_2, q_2, ..., q_n) \), the decision rules of \( S = (D, M \cup N, V, f) \) can be defined as:

\[
\alpha \rightarrow \beta = (n_1, v_1) \land (n_2, v_2) \land ... \land (n_m, v_m) \rightarrow (m_1, q_1), (m_2, q_2), ..., (m_n, q_n)
\]  

(1)

2.2. Definition of Public Opinion Big Data Granulation and Knowledge Discovery Based on Granulation Mechanism
All For the decision system \( S = (D, F, V, f) \) defined above, we design \( \alpha \rightarrow \beta \) is the decision rule of \( S \). when the decision system driven by big data is very large, \( S \) needs to be divided into several subsystems, the decision with \( \alpha \rightarrow \beta \) is transformed into a subsystem, which can realize the decomposition and transformation of multi granularity decision system and cross granularity reasoning. By further decomposing the conditions and decision attributes in decision system
\[ S = (D, M \cup N, V, f) \]
set \( u_1, u_2, \ldots, u_n \in M \cup N \), there can be conditional attribute or decision attribute in attribute \( u_1, u_2, \ldots, u_n \). For \( n \) attribute values \( q_1, q_2, \ldots, q_n \in V \), set
\[
\phi = (u_1, q_1) \wedge (u_2, q_2) \wedge \ldots, \wedge (u_m, q_m)
\]
(2)

Then \( \phi \) is the corresponding particle on \( S \) when the attribute \( u_1, u_2, \ldots, u_n \) is fixed, the particle \(|\phi|\) changes with the change of attribute value \( q_1, q_2, \ldots, q_n \), so the set of particles can be obtained:
\[
L = \{ |\phi|, |\phi| \neq \emptyset, \phi = (u_1, q_1) \wedge (u_2, q_2) \wedge \ldots, \wedge (u_m, q_m), q_1, q_2, \ldots, q_n \in V \}
\]
(3)

We call \( L \) as a set of particles related to attribute \( u_1, u_2, \ldots, u_n \). Multiple decision-making subsystems can form multiple sets of particles in different spaces. For any decision rule
\[ S = (D, M \cup N, V, f) \]
of decision system \( \alpha \rightarrow \beta \), its corresponding decision can be transformed into a partition subsystem \( S^1 \), equivalently.

3. Experimental Analysis
For such structured micro data resources as real-time monitoring data, social and economic development data, public opinion subject data and basic information data stored in the relational database, it is proposed to transform them into RDF data model. Association data construction mainly completes the semantic description of data and RDF format transformation by building mapping files. For structured data extraction, there are many open-source software, such as Drupal, d2rq, triple and other tools, where, D2RQ function more comprehensive tool, you can achieve the data in a relational database into RDF format of resource data in order to facilitate analysis of data relationships with other RDF data, therefore, this paper aims to use D2RQ tools for the extraction of structured data. D2rq, proposed by Bizer, is an illustrative language, which can establish a mapping relationship between rdf:ns / OWL ontology and RDF tables, thus allowing applications supported by RDF to use non RDF relational data as virtual RDF data. The implementation structure is shown in Figure 2.

![Figure 1. D2rq system architecture](image)

The experimental environment of this experiment is win10 bit operating system, the software tools used are: Eclipse supporting Java program development; MySQL relational database management system for data storage and management; Jena2 is an open source tool supporting semantic web and associated data application development; d2rq version is d2rq-0.8.1. We use d2rq to convert the contents of the focus task body object stored in MySQL into RDF triplet files, some transformation contents are as follows:

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<rdft:RDF
xmlns:rdft="http://www.w3.org/1999/02/22-rdft-syntax-ns#"
xmlns:cd="http://www.yanan.cn/cd#"/>
```
The first line of this RDF document is an XML declaration. This XML declaration is followed by
the root element of the RDF document: < RDF: RDF >. Xmlns: RDF namespace, which specifies that
the element with the prefix RDF comes from the namespace "http://www.w3.org/1999/02/22-rdf-
syntax-ns". Xmlns: CD namespace, which specifies that the element with the prefix CD comes from
the namespace "www.yanan. cn/cd". The < RDF: description > element contains a description of the
resource identified by the RDF: about attribute. The elements <cd:artist>, <cd:country>,
<cd:company>, and so on are attributes of this resource. Table 2 is the data after triple data analysis,
The RDF triplet model after parsing is visualized as shown in Figure 2.

Table 1. Triple data analysis results

| Sn | Subject                  | Predicate          | Object           |
|----|--------------------------|--------------------|------------------|
| 1  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#artist | "WangPen"        |
| 2  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#country | "China"          |
| 3  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#company | "Yanan City"     |
| 4  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#date    | "2020.4.10"     |
| 5  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#arttitile | "Opition view"  |
| 6  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#homepag | "http://weibo.sina.cn" |
| 7  | http://www.yanan.cn/cd/weibo | http://www.yanan.cn/cd#borthyear | "1975"          |
Figure 2. Visualization diagram of triplet model after RDF analysis

Association data analysis uses RDF structured data, establishes semantic links between data, navigates between different URIs with RDF links, and provides users with association access between different data. The key technology of data association is to effectively identify all kinds of data, image, video and other resources in the big data to be fused by RDF, and then access and transfer the identified resources by protocol. All data sources are identified by URI identity, and a data object represents an RDF resource to be described.

4. Conclusion
As a new semantic network technology of data exchange and representation, association data can transform unstructured and heterogeneous data on the network into structured data that can be processed and understood by machines, which provides a good foundation for the organization of public opinion data.

The associated data analysis method inputs the multi-source and heterogeneous big data, and the output content is the RDF Meta model data information stored in a centralized way that conforms to the unified semantic description standard. The processing process includes the structured data object category, the multi-source and heterogeneous data type classification, the associated data description and definition, the structured data RDF model definition, the semi-structured data RDF model definition, and the non-knot The definition of the unified data model of structured data and the definition of the RDF meta model of unstructured data are processed, and the output is the RDF meta model information conforming to the unified description of multi-source heterogeneous data.

Acknowledgments
This work is partially supported by the Natural Science and Technology Project Plan in Yulin of China (2019-78-2, 2019- 2019-78-1, 2019-76-2), Natural Science Basic Research Plan in Shaanxi Province of China (2017NY-134, 2016KJXX-62), Funding Project for Department of Yulin University (16GK24,TZRC1801), and Thanks for the help.

References
[1] Kim Y, Kang N and Jung J et al, A review on the management of water resources information based on big data and cloud computing , Journal of Wetlands Research, 18(1) 100-112,2016
[2] F Zhang, H-F Xue, D-S Xu, Big data cleaning algorithms in cloud computing, International Journal of Online Engineering, 9( 3) 77-81, 2013
[3] Li Zhao-Xing, A multi objective optimization algorithm for recommender system based on
PSO, Computer Modelling and New Technologies, 18(7) 231-235, 2014
[4] Donkal, Gita, Verma, Gyanendra K, A multimodal fusion based framework to reinforce IDS for securing Big Data environment using Spark, Journal of Information Security and Applications, 43(1) 1-11, 2018
[5] Zhang Feng, Xue Hui-feng, Zhang Jing-Cheng, Multi-source Big data Dynamic Compressive Sensing and Optimization Method for Water Resources Based on IoT, Sustainable Computing: Informatics and Systems, 20: 210-219, 2018
[6] S. Singh and Y. Liu, A Cloud Service Architecture for Analyzing Big Monitoring Data, Tsinghua Science & Technology, 21(1) 50-70, 2016
[7] J. H. Huang and X. Yao, A Multi-Agent Evolutionary Algorithm for Software Module Clustering Problems, Soft Computing, 21(13) 3415-3428, 2017
[8] Kaleen R M, Muhammad F and Ibrahim B G et al, Remote access capability embedded in linked data using bi-directional transformation: Issues and simulation, Sustainable Cities and Society, (38) 662-674, 2018
[9] Isabel Torre-Bastida, Ana Del Ser and Javier Lana, Big Data for transportation and mobility: recent advances, trends and challenges, IET Intelligent Transport Systems, 12(8) 742-755, 2018
[10] Mihir Pant, Don Towsley and Dirk Englund et al, Percolation thresholds for photonic quantum computing, Nature Communications, 6 (3) 6-10, 2019