ISSN: 2688-8653

A Method to Solve Sentence in Natural Language Data Mining Based on Knowledge Graph

Guoxing Lv, Chuansheng Wu

Liaoning University of Science and Technology, Liaoning, China.

Abstract

The implementation example of this project discloses a method and device for solving sentences of natural language data mining based on a knowledge map. The method includes: building a knowledge map of data mining process;Perform natural language processing on the received data mining statements described by natural language to extract the problem ontology and the relationship between the problem ontology;Standardize the problem ontology and the relationship between the problem ontology to obtain a standardized subgraph to be matched;The sub graph to be matched is matched with the similar sub graph of the knowledge map of the data mining process constructed to obtain the problem ontology mapping relationship between the solution sub graph, the sub graph to be matched and the solution sub graph;Get the abstract process of data mining solution associated with it according to the solution subgraph;Map the content corresponding to the standardized problem ontology node in the sub graph to be matched to the input parameters of the abstract process of data mining solution, and execute the solution process to get the output results.This project has greatly reduced the threshold of using big data mining technology.

Keywords

Knowledge Map, Natural Language, Data Mining, Statement.

1. Preface

In technology research and development, data mining is a process of discovering useful knowledge from data.It is widely used in decision-making analysis of banking, finance, insurance, retail, logistics, e-commerce, scientific research, biology, medicine, agriculture and other industries.In the information age, the scale of data continues to increase, and the demand for massive data mining will become more and more common. It has become an urgent problem to accurately and efficiently select useful knowledge from among them. At the same time, the heterogeneity of data also increases the difficulty of massive data processing.

At present, the data mining methods in the market mainly include: 1. Traditional data mining platform: the business personnel who understand the data and business situation propose the data mining requirements, and the technical personnel, after understanding the requirements, program to realize the data mining and output the mining results.Technicians may not be practical because they do not understand the business.In addition, every time the business personnel have a new idea, it needs to be handed over to the technical personnel for implementation, and it is impossible to quickly test and obtain results.2. Visual BI tool: it focuses on visualization, and needs to learn the syntax or operation method of BI tool programming, and requires users to understand the expertise in big data mining.However, it cannot solve complex problems and cannot analyze data with many dimensions.

ISSN: 2688-8653

2. Project content

The purpose of this project is to overcome the shortcomings of the existing technology, provide the method and device for solving the sentences of natural language data mining based on the knowledge map, realize the shielding of profound expertise in the field of big data mining for service users, simplify the modeling of complex service processes in the field, and greatly reduce the threshold for the use of big data mining technology,It has promoted the application of big mining technology in decision-making processing of various industries.

In order to achieve the above objectives, the project adopts the following technical solutions:

First, the statement solving method of natural language data mining based on knowledge map includes:

Constructing the knowledge map of data mining process;Perform natural language processing on the received data mining statements described by natural language to extract the problem ontology and the relationship between the problem ontology;Standardize the problem ontology and the relationship between the problem ontology to obtain a standardized subgraph to be matched;The sub graph to be matched is matched with the similar sub graph of the constructed data mining process knowledge map to obtain the problem ontology mapping relationship between the solution sub graph, the sub graph to be matched and the solution sub graph;Get the abstract process of data mining solution associated with it according to the solution subgraph;Map the content corresponding to the standardized problem ontology node in the sub graph to be matched to the input parameters of the abstract process of data mining solution, and execute the solution process to get the output results.

Build a knowledge map of data mining process, including: abstract the data mining method into a single solution function or multiple solution functions to reverse the solution process.

Standardize the problem ontology and the relationship between the problem ontology to obtain the standardized sub graph to be matched, including: mapping the problem ontology to the representation matrix through the language model;Calculate the similarity between the representation matrix and the problem ontology representation matrix in the data mining process knowledge map to obtain the similarity result;Set the problem ontology higher than the set threshold in the similarity result as the standardized problem ontology;Standardized subgraph to be matched is constructed with standardized problem ontology as node and problem ontology relationship as edge.

Map the content corresponding to the standardized problem ontology node in the sub graph to be matched as the input parameters of the abstract process of data mining solution, and execute the solution process to get the output results, including: judging whether the problem ontology in the sub graph is not fully matched;If all the matches are successful, the content corresponding to the standardized problem ontology node in the sub graph to be matched will be mapped to the input parameters of the data mining solution abstract process, and the solution process will be executed to obtain the output results.

On the second hand, the natural language data mining statement solving device based on knowledge map includes building unit, extraction unit, standardization processing unit, matching unit, acquisition unit and execution unit;The construction unit is used for constructing the knowledge map of data mining process;The extraction unit is used for performing natural language processing on the data mining statements described by the received natural language to extract the problem ontology and the relationship between the problem ontology;The standardization processing unit is used for standardizing the problem ontology and the relationship between the problem ontology to obtain a standardized sub graph to be matched;The sub graph to be matched is matched with the similar sub graph of the constructed data mining process knowledge map to obtain the problem ontology mapping

ISSN: 2688-8653

relationship between the solution sub graph, the sub graph to be matched and the solution sub graph;The acquisition unit is used to acquire the associated data mining solution abstract process according to the solution subgraph;The execution unit is used to map the content corresponding to the standardized problem ontology node in the sub graph to be matched into the input parameters of the data mining solution abstract process, and execute the solution process to obtain the output results.

The construction unit includes an abstract processing module;The abstract processing module is used to abstract the data mining method into a single solution function or a solution flow of multiple solution functions.

The standardization processing unit comprises a mapping module, a calculation module, a setting module and a construction module;The mapping module is used to map the problem ontology into a representation matrix through a language model;The calculation module is used to calculate the similarity between the representation matrix and the problem ontology representation matrix in the data mining process knowledge map to obtain the similarity result;The setting module is used to set the problem ontology higher than the set threshold in the similarity result as the standardized problem ontology;The construction module is used to construct a standardized subgraph to be matched with the standardized problem ontology as the node and the problem ontology relationship as the edge.

The further technical scheme is that: it also includes a judgment unit;The judgment unit is used to judge whether the problem ontology in the sub graph is not matched successfully;If all the matches are successful, the content corresponding to the standardized problem ontology node in the sub graph to be matched will be mapped to the input parameters of the data mining solution abstract process, and the solution process will be executed to obtain the output results.

In the third aspect, a computer device includes a memory, a processor and a computer program stored on the memory and capable of running on the processor. When the processor executes the computer program, it implements the above steps of solving the natural language data mining statement based on the knowledge map.

The fourth aspect is a computer-readable storage medium. The storage medium stores a computer program, and the computer program includes program instructions. When the program instructions are executed by the processor, the processor will execute the above steps of the natural language data mining statement solving method based on the knowledge map.

3. Implementation steps

A method for solving sentences of natural language data mining based on knowledge atlas includes the following steps: S10-S60.

S10. Build a knowledge map of data mining process.

In an embodiment, step S10 specifically includes the following steps: S11.

S11. The solution process of abstracting the data mining method into a single solution function or multiple solution functions is reversed.

The data mining process knowledge map is constructed through data mining methods. The commonly used data mining methods include correlation analysis, regression analysis, decision tree, etc., which can be abstracted into the solution process twist of a single solution function or multiple solution functions. The solution twist of multiple functions can use the process engine to configure the twist relationship.Therefore, a data mining method can form an abstract process of data mining solution.

Each solution process has corresponding input parameters and output results, and we take the input parameters and output results as the problem ontology;Solving abstract processes is actually a directed graph from solving parameters to outputting results, that is, problem

ISSN: 2688-8653

ontology relationship.The relationship between the problem book and the problem ontology can form a directed acyclic graph, which is named as the solution subgraph.

In order to facilitate subsequent standardization, each problem ontology is mapped into a multi-dimensional representation matrix through language model calculation in advance and stored in the knowledge map.

S20. Perform natural language processing on the received data mining statements described by natural language to extract the problem ontology and the problem ontology relationship.

Problem ontology extraction belongs to the task of named entity recognition, and problem ontology relationship recognition belongs to the problem of relationship classification between entities.The extraction of problem ontology and problem ontology relationship is processed by the natural language processing text information extraction model. The natural language processing text information extraction model can be any entity recognition and relationship extraction model, such as BERT like language models, CasRel and other joint entity recognition and relationship extraction models.By labeling the problem ontology and problem ontology relationship in the data mining statements described by natural language, training the model, and obtaining the natural language processing text information extraction model that can identify the problem ontology and problem ontology relationship.For the newly input data mining statements described by natural language, multiple problem ontologies and corresponding problem ontology relationships are identified through the text information extraction model of natural language processing.

In this embodiment, the data mining sentence described by the input natural language is: "The house is located near the Shenzhen Software Park, with a building area of 75 square meters, two rooms, one living room and one bathroom. It was built in 2005 and renovated in 2019. The house adopts Nordic style soft decoration, good lighting and convenient life. If you want to sell it, how much can the house price be set?".

After natural language processing text information extraction model, the problem ontology is extracted as follows:

Housing location information: Shenzhen Software Park;House size: 75;Number of bedrooms: two;Number of toilets: one;House construction time: 2005;House renovation time: 2019;House appearance: Nordic style;Output result: house price.

After the natural language processing text information extraction model, the problem ontology relationship is extracted as follows:

House location information+house size+number of bedrooms+number of toilets+house construction time+house renovation time+house appearance ->output results.

S30. Standardize the problem ontology and the relationship between the problem ontology to obtain a standardized subgraph to be matched.

Since the problem ontology is extracted from the data mining statements described by natural language, it needs to be mapped to the problem ontology in the data mining process knowledge map before sub map matching can be done.

In an embodiment, step S30 specifically includes the following steps: S301-S304.

S301. The problem ontology is mapped to the representation matrix through the language model.

S302. Calculate the similarity between the representation matrix and the problem ontology representation matrix in the data mining process knowledge map to obtain the similarity result.

S303. Set the problem ontology higher than the set threshold in the similarity result as the standardized problem ontology.

S304. Standardized subgraph to be matched is constructed with standardized problem ontology as node and problem ontology relationship as edge.

ISSN: 2688-8653

For step S301-S304, as shown in Figure 6, Figure 6 is a schematic diagram of the standardized sub map to be matched.Use language models in general or professional fields for semantic association. Various language models such as BERT can be used. After language model calculation, the extracted problem ontology can be mapped into a multi-dimensional representation matrix, and the similarity calculation can be done with the multi-dimensional representation matrix of the problem ontology in the data mining process knowledge map. Various similarity calculation methods can be used.The problem ontology with the highest similarity and higher than the set threshold is the standardized problem ontology.Finally, with the standardized problem ontology as the node, the extracted problem ontology relationship as the edge, and the standardized subgraph to be matched is obtained.

In this embodiment, as shown in Figure 7, Figure 7 shows the standardized sub map to be matched after standardization processing.

S40. Match the sub graph to be matched with the constructed data mining process knowledge map to obtain the problem ontology mapping relationship between the solution sub graph, the sub graph to be matched and the solution sub graph.

The obtained sub graph to be matched is matched with the similar sub graph of the data mining process knowledge map to obtain the solution sub graph with the highest matching similarity and higher than the set threshold, and the problem ontology mapping relationship between the sub graph to be matched and the solution sub graph.Similar subgraph matching can be realized by such methods as backtracking search based on depth search, or multi way join based on breadth first.When matching subgraphs, the data mining solution abstract process nodes in the data mining process knowledge map are ignored.

In this embodiment, the UllmannAlgorithm algorithm based on depth search and backtracking is used to obtain the matching subgraph as shown in Figure 8.

S50. Obtain its associated data mining solution abstract process according to the solution subgraph.

After the matched solution sub graph is obtained, its associated abstract process of data mining solution can be obtained.

S60. Map the content corresponding to the standardized problem ontology node in the sub graph to be matched as the input parameters of the data mining solution abstract process, and execute the solution process to obtain the output results.

The problem ontology mapping relationship between the sub graph to be matched and the sub graph to be solved, and the sub graph to be solved consists of the input parameters and output results of the data mining solution abstract process, so that the content corresponding to the problem ontology node in the sub graph to be matched can be mapped to the input parameters of the data mining solution abstract process, and the solution process can be executed according to the process twist relationship between the solution function and the multi function,The output results are obtained.

In one embodiment, before executing the solution process, it is also necessary to determine whether the problem ontology is completely matched in the process of matching and solving the subgraph.Therefore, the following steps are also included before step S60: S55.

S55. Judge whether the problem ontology in the sub graph is not matched successfully;If all matches are successful, the content corresponding to the standardized problem ontology node in the sub graph to be matched will be mapped to the input parameters of the data mining solution abstract process, and the solution process will be executed to get the output results.

If the problem ontology does not match all, the solution process cannot be executed, that is, the compilation fails.If the problem ontology matches all, the subsequent solution process can be executed.

ISSN: 2688-8653

4. Conclusion

The beneficial effects of this project compared with the existing technology are as follows: this project carries out natural language processing on the data mining statements described by natural language to extract the problem ontology and the problem ontology relationship, and then standardizes the problem ontology and the problem ontology relationship to obtain standardized sub maps to be matched;The sub graph to be matched is matched with the similar sub graph of the constructed data mining process knowledge map to obtain the problem ontology mapping relationship between the solution sub graph, the sub graph to be matched and the solution sub graph;Get the abstract process of data mining solution associated with it according to the solution subgraph;Map the content corresponding to the standardized problem ontology node in the sub graph to be matched to the input parameters of the abstract process of data mining solution, and execute the solution process to get the output results.It has realized shielding the profound professional knowledge in the big data mining field for service users, simplifying the modeling of complex domain service processes, greatly reducing the threshold for the use of big data mining technology, and promoting the application of big data mining technology in decision-making processing in various industries.

Reference

[1] A Fast and Effective Web Document Clustering Method [J]Zhang Rong. Computer Application Research, 2004 (04).

[2] Web Mining and Its Application in Competitive Intelligence System [J]Chen Pingli, Information Science, 2003 (09).

[3] Web based data mining [J]Li Haibin. Journal of Guilin Institute of Technology, 2003 (02).

[4] Research on Web Log Mining Technology [J]Jinwei. Computer CD Software and Application, 2012 (14).

[5] Research on the Application of Web Mining in Website Optimization [J]Yang Qinglian;Wang Jing. China Management Informatization, 2008 (15).

[6] An Improved Web Log Session Identification Method [J]Fang Yuankang;Hu Xuegang;Xia Qishou, Computer Technology and Development, 2008 (11).

[7] A Web mining recommendation method based on collaborative filtering [J]Chen Xi;Chen Xin. Journal of Beijing University of Information Technology (Natural Science Edition), 2013 (06).

[8] Research on the Application of Web Mining Technology in Social Network Analysis [J]Gao Hua, Science and Technology Information, 2013 (09).

[9] Optimized session identification method in Web log preprocessing [J]Fang Yuankang;Hu Xuegang;Xia Qishou, Computer Engineering, 2009 (07).

[10] Research progress and prospect of Web mining based on soft computing [J]Yi Gaoxiang;Hu Heping, Computer Engineering and Design, 2006 (10).