Journal of Computers Vol. 32 No. 6, 2021, pp. 168-175
doi:10.53106/199115992021123206014

Data Analysis Method of Talent Cultivation Based on Relational Graph

Aiyan Wu*, Yongmei Zhang

School of Information Science and Technology, North China University of Technology, Jinyuanzhuang 5, Beijing, China
wuaiyan@ncut.edu.cn

Received 1 July 2021; Revised 1 August 2021; Accepted 2 August 2021

Abstract. A large amount of data are produced in the process of postgraduate training. Based on relational graph technology, a complete technical scheme is proposed inclouding big data integration, relationship extraction and visualization. It can serve the training of high-level talents in Colleges and universities. Firstly, it analyzes the characteristics of postgraduate training big data, and puts forward the construction of multi-layer concept entity to complete the concept model of relation graph. And a classification algorithm based frequent item mining is proposed to analyze the employment data. Then, a communication architecture among indifferent relational databases and graph database is designed to solve the problem of access conflict. The research method completes the distributed integration of training big data, excavates and displays the training information and its rules to guide the education practice. Take the school of information science and technology of North China University of technology as an example to show the application effect. This method is suitable for expressing the relationship between complex structured big data in a certain professional field, especially the data sets stored in different relational databases. It can not only clearly show the relationship between different granularity entities, but also greatly improve the efficiency of relationship extraction.

Keywords: postgraduate, education quality, relation graph, big data

1 Introduction

Advanced information technology promotes the reform and innovation of education, and the current education informatization has entered the era of big data. A large amount of data are produced in the process of high-level talent training, such as teaching, management, practice and other training. These data contain important value [1-3]. So, it is necessary to transform them into visual information of assistant education management. In education management, the demand for data collaboration and depth analysis is increasingly strong. The mining and display of the relationship between data can more accurately judge the academic crisis, and guide the development of students with clear goals [4-5]. It's not easy to do that. Because it studies not only the curriculum achievement, but also a lot of individual development information [6]. The goal of this paper is to integrate and display main training information of individual, extract and display the relationship between various academic items and various positions. Finally it can provide services for talent training.

This research faces a huge data system, including course scores, dormitory members, laboratory team members, competitions, practical activities, research direction, occupation, etc. They are structured data stored in different types of relational databases, which come from different departments such as teaching, scientific research, management and so on. There are many data types and inconsistent storage formats. Therefore, if data is analyzed in the original relational database, the access efficiency must be very low. Therefore, the relationship graph is used in this paper, which can not only improve the operational

* Corresponding Author

efficiency, but also realize the visualization of the relationship between various entities. First, the relationship between entities is extracted from different relational databases and stored in the form of graph data. Then, the training data is analyzed by searching the graph data. Finally, the analysis results are displayed by relational graph.

According to the characteristics of personnel training data, this paper puts forward a construction method of relationship graph based on multi tier architecture and graph data analysis method for high-level personnel training. The main work of this paper is as follows:

(1) A multi-layer concept entity architecture is proposed to clarify the entities with different granularity and the relationships between them;

(2) The Apriori algorithm is improved, and a frequent item mining algorithm based on classification is proposed to realize the extraction of the relationship between training and employment;

(3) a communication architecture among indifferent relational databases and graph database is designed to solve the problem of access conflict, and realize the big data relationship graph of postgraduate cultivation.

2 Theoretical Basis

A large number of data sets are analyzed by specific algorithms, and then the relationships contained in different things are found. These relationships can connect the things into a data network, which is the relational graph [7-8]. Its concept comes from the social graph. Compared with knowledge graph, a relational graph can pay more attention to the relationship between entities rather than their knowledge structure [9-10].

A relationship graph is directional and can be expressed by RDF framework. Its model is shown in Fig. 1. *s* and *e* represent entities, and r_i represents relationships between entities.

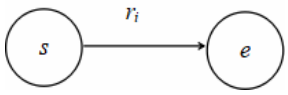

Fig. 1. RDF meta relation model

RDF framework is a set of triples, each of which is the relationship between a pair of entities in the relational graph. The relationship is not unique and repeatable. That is, the relationship in one triple can also be another. It can be expressed by the following formula.

$$RDF = \{R_1, R_2, \cdots, R_n\} \tag{1}$$

$$R_i = (s, r_i, e) \qquad R_i \in RDF \tag{2}$$

Relational graph technology can not only show information entities, but also show the association between entities [11]. So, it can make up for the defect that many visualization technologies are difficult to show information association. The development of big data technology has realized the extraction of information association from different relational databases. Therefore, relational graph technology has become one of the key technologies of big data visualization.

Domain specific graph focuses on architecture and professionalism between its concepts. In the process of graph construction, rule-based entity extraction method is the first to be used. This method is suitable for structured data, and entity extraction is carried out by string matching. In addition, statistical method is used to calculate the frequency of keywords, and then entities are extracted. It is suitable for unstructured data. Talent training data is structured data. According to the characteristics of domain data, entity extraction method based on keyword matching can be constructed. But, the data relationship of talent training is complex, and data granularity is not unique.

3 Relationship Graph Construction Method Based on Multi-layer Concept Entity

3.1 The Definition of Multi-layer Concept Entity

In order to explore the integration, analysis and visualization of talent cultivation data, we should first analyze the daily management needs, and then define the entities of the relationship graph, the attributes of the entities and the relationship between entities, and construct the conceptual model of the relationship graph. In the research of personnel training data, we find that students' learning is a big concept category, and it is difficult to clearly show the details of academic progress with the same granularity concept entity. Therefore, the construction method based on multi-layer concept entity is proposed.

A concept entity can be composed of many parts, so we can define concept entity, sub concept entity and their attributes, and the relationships between them. The definition of multi-layer concept entity can be expressed by formula 3 and formula 4. s and e are entities, e_i is the sub entity of e, and r is the association between s and e, which is also the association between entity s and sub entity e_i.

$$e = e_1 \cup e_2 \cup \cdots \cup e_n \tag{3}$$

$$(s,r,e) \Rightarrow (s,r,e_i) \quad \begin{cases} e_i \in e \\ i = 1 \cdots n \end{cases} \tag{4}$$

3.2 Conceptual Model of Relational Graph of Talent Cultivation Based on Multi-layer Entity

The data entity of talent cultivation is defined with the multi-layer concept entity method, which can more clearly express the talent training process. In practical research, two layers of concept entities and their attribute constraints are defined, as shown in Table 1.

Table 1. Definition of multi level concept entity

entity	subentity	attributes	constraints
graduate student	none	student number	character
		name	character
		age	integer
		gender	character
		student cadre	character
		grade	character
		major	character
		research direction	character
		main technology	character
tutor	none	name	character
		major	
study	credit	credits of degree courses	integer
		credits of elective courses	
		GPA	
		practice credits	
	paper	amount of issued papers	integer
		status of dissertation	character
	competition	name	character
		year	date
		level	character
		awards	character
	honor	year	date
		level	character
		name	character
job	none	name	character
		type	

Corresponding to the definition of two-level concept entities, the relationship between them can be divided into three categories: basic relationship, development relationship and attribute related relationship, as shown in Table 2.

Table 2. Relationship between entities

relationship type	relationship name	related entities	relationship description
basic relationship	academic guidance	postgraduates and tutors	graduate students guided by tutors in their studies.
	research team	postgraduates	graduate students with the same tutor
	dormitory members	postgraduates	graduate students living in the same dormitory
developing relationship	get	postgraduates and their studies	learning progress of Postgraduates
	employment	postgraduates and their jobs	jobs chosen by graduate students after graduation
attribute association	*N*-correlation	study and job	the degree of correlation between different jobs and postgraduate training

Relationship graph is a directed graph, and the relationship between entities is not unique. The relationship in one triplet can also be in another. In fact, a relationship is an attribute that represents the relationship between entities. The conceptual structure of talent cultivation data is shown in Fig. 2.

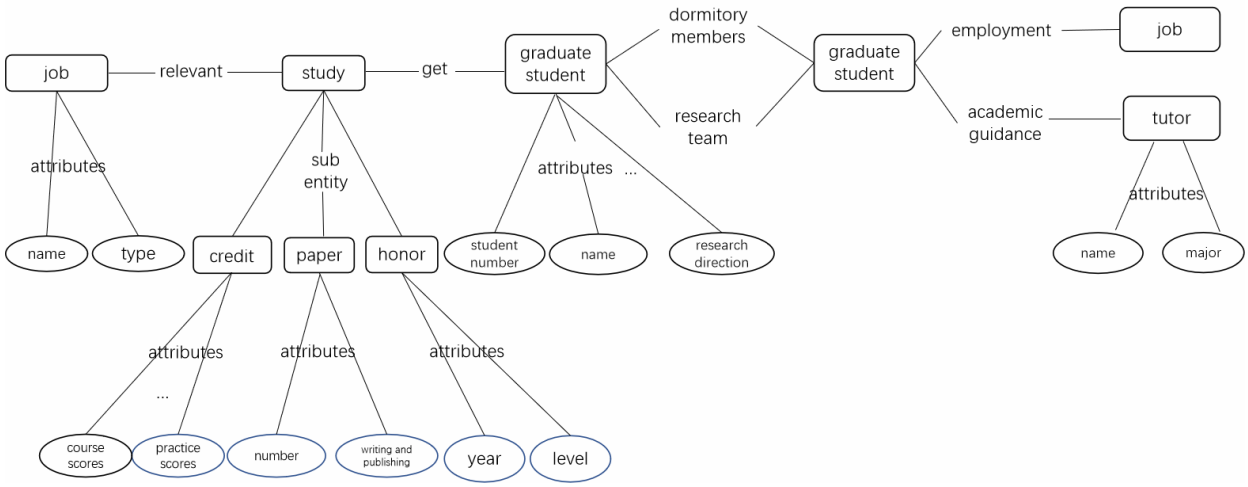

Fig. 2. Two layer conceptual structure of relation graph on talent training

3.3 Extraction and Storage of Relationships

The instantiation of the conceptual model is its data model. In practical research, graduate information are mostly structured data, but they are stored in different types of databases. In order to coordinate the communication between different databases and meet the application requirements, graph database is adopted. Entities and their relationships are extracted from relational database, represented by triples, and stored in graph data. All relationships are extracted first by year and then by profession. When the data is updated, only the new data is extracted and stored in the graph database. For example, ("Josn", "tutor", "Wang") means "Josn's tutor is Mr. Wang", as follows:

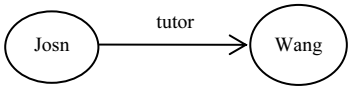

Fig. 3. Instance element

The algorithm is described as follows:

```
Node=year; /*extracted first by year*/
Traversal(node) /*access database by year*/
{
  Visit(node.major) /*then access data of the same grade by major*/
  While(not null)
  {
    If(record.value==keyword)
    Assign values to triples /* represented by triples*/
    }
}
```

3.4 Data Correlation Analysis

Different jobs have different technical requirements. In the process of postgraduate training, it is necessary to guide the development goals of postgraduates to match the job needs. High level personnel training can better serve the local economic development. Therefore, this paper analyzes the correlation between graduate development and job demand, and shows it in a relational graph, so as to facilitate the understanding and use of education managers. In practice, there are few attributes in association mining. The result of mining is not to obtain the maximum frequent itemsets, but to obtain the degree of correlation between different jobs and postgraduate training. The job property set is represented by A, and the set of graduate development events is represented by B.

$$A = \{\text{civil servants} \text{、} \text{instutions} \text{、} \cdots \text{、} \text{state} - \text{owed enterprises} \text{、} \text{private enterprises}\} \tag{5}$$

$$B = \{\text{GPA} \text{、} \text{competition} \text{、} \cdots \text{、} \text{research direction} \text{、} \text{main technology}\} \tag{6}$$

According to the research needs, Apriori algorithm is improved, and a frequent item mining algorithm based on classification is proposed. The main idea is to visit graduate employment information based on the year and major, and store it in the attached table according to the type of job. The element correlation is calculated by traversing the attached table. The algorithm is represented by the following formula.

$$\mu = \{\mu_1, \mu_2, \cdots, \mu_7\} \tag{7}$$

$$\begin{cases} \exists A_j AND(B_i ORB_i) \rightarrow \mu_i = \dfrac{\mu_i \times n + 1}{n + 1} \\ A_j \in A, B_i \in B \end{cases} \tag{8}$$

Here, μ is the correlation set, μ_i is the correlation between B_i and A_j, and n is the number of traversed records.

$$\delta_i = \begin{cases} \text{highly relevant} & 75\% \le \mu_i \\ \text{relevant} & 30\% < \mu_i < 75\% \\ \text{irrelvant} & \mu_i \le 30\% \end{cases} \tag{9}$$

Finally, the results are expressed as triples (A_j, δ_i, B_i) and imported into the graph database. For example: (“IT private enterprise”, “highly relevant“, “Java”) means IT private enterprises are more likely to choose graduate students who have mastered Java technology.

4 The Realization and Application of Relational Graph

4.1 Its Realization

An independent relation extraction module is constructed to extract relations from different databases of different departments. For example, the graduate students’ academic achievements are extracted from the database of the educational administration department, and the job information is extracted from the

employment department. All relationships are stored firstly in the graph database, and then extracted from the graph database for graph display.

Graph display mainly includes two parts. One is the graduate relationship map, which extracts the relationship between postgraduates and related entities from the graph database with student ID, and shows them directly in the form of map. Another is to show the relationship between various jobs and academic achievements. According to the job category, the related academic achievements are extracted from the graph database and their correlation degree is calculated with formulas 8 and 9, which is displayed in the form of relationship graph. The architecture of the program implementation is shown in Fig. 4.

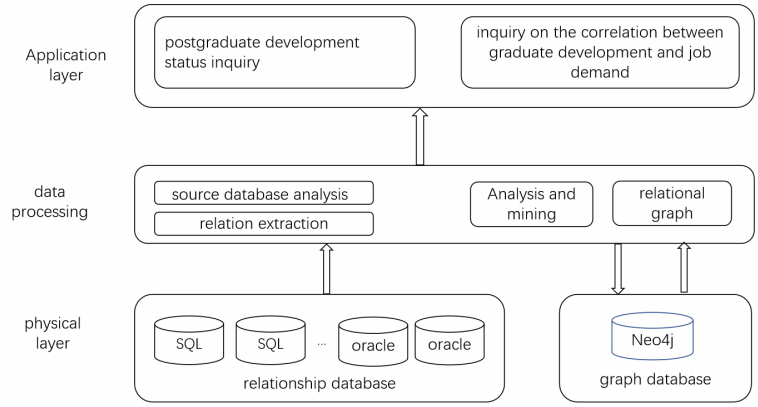

Fig. 4. Software architecture of relation graph on talent training

4.2 Application Cases

Taking the postgraduates in School of Information Science and Technology, North China University of Technology as an example is following. Teaching, scientific research, ideological education and employment are managed by different departments. The above scheme is used to build a collaborative management platform for postgraduate training. The platform includes course, practice, thesis opening, mid-term assessment, thesis review and defense, degree evaluation and other key links. And it can realize the management and monitoring of big data in the whole training process. The individual development of a postgraduate is shown in Fig. 5. At the same time, it can analyze the relationship between students' development and employment, and give feedback to guide students' development. This can help to build a classified guidance for graduate development with students' employment intention and employers' recruitment needs.

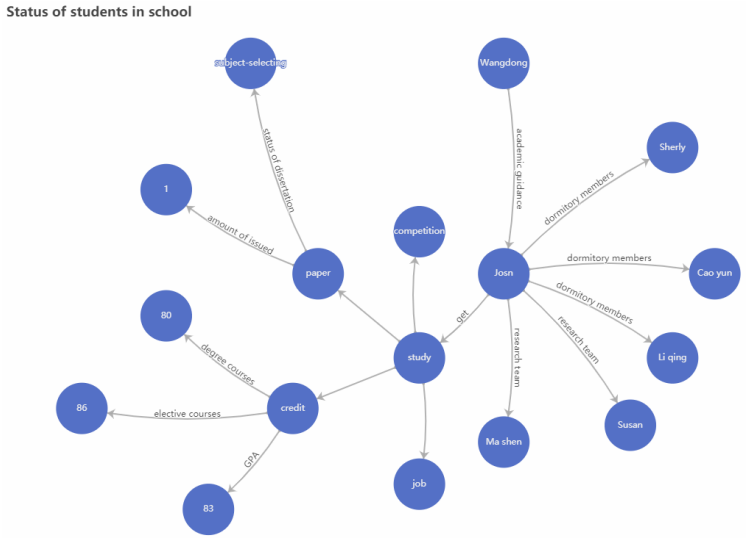

Fig. 5. Map of postgraduate individual development

From the map of graduate students' individual development, we can fully understand an individual's situation, especially the students with academic difficulties. For students with academic difficulties, it can help to analyze the problems, so that counselors can give timely guidance, and effectively help students out of academic difficulties. Since 2014, there are 16 students with academic early warning in the college. Students are effectively guided to improve their academic status in time with the platform. Using the statistical method in references 12 and 13, the median represents the average state of years of study. In this college, the median length of master's degree is shown in Fig. 6.

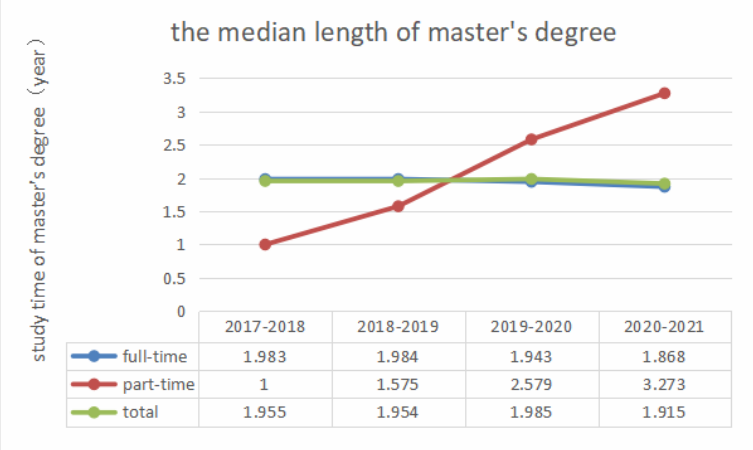

Fig. 6. Median length of master's degree

In addition, counselors can guide the development of graduate students with students' employment intention and employers' recruitment needs, which will promote the students to plan their own development path according to their future job requirements, and then clarify the learning objectives. This measure not only improves the training quality of graduate students and promotes their successful graduation, but also improves their employment satisfaction and employers' satisfaction to them. The graduate employment rate of the college has been nearly 100% for five consecutive years. And the average signing rate of employment agreement is more than 90%, which is higher than the local average.

5 Conclusion

The informatization process of higher education has completed the informatization stage of management platform. How to find the rules to guide the practical work from the complicated educational information? This is an important research issue in the new era of educational informatization. In this paper, a new construction method of relationship graph is proposed for multi granularity cultivation information. Practice has proved that this method can integrate distributed training information, mine and display training rules. It is very convenient for the information coordination in the process of graduate students' cultivation, and can improve the management level and the quality of training.

Acknowledgements

The paper is supported by the project (NO. KM202110009002).

References

[1] J. Wang, Y. Yang, Y. Zheng, From Big Data to Computational Pedagogy: Conceptual, Motivation and Outlet, China Academic Journal Electronic Publishing House (1)(2020) 85-91.

[2] Y. Feng, J. Zhong, X. Li, S. Zhou, Research and Practice in Collaborative Innovation Training of High-level Technology Personnel for Big Data Industry, China Academic Journal Electronic Publishing House (06)(2017)35-44.

[3] F. Cai, W. Zhou, X. Wang, Review of Postgraduate Education Research in China in 40 Years of Reform and Opening-up, Heilongjiang Researches on Higher Education 38(02)(2020) 108-113.

[4] X.-Y. Zhang, J.-H. Wang, D. Feng, Forty Years' High-Level Talent Research: Hot Topics and Prospects, Heilongjiang Researches on Higher Education 37(12)(2019) 39-43.

[5] S. Zhong, How Artificial Intelligence Supports Educational Revolution, China Academic Journal Electronic Publishing House (03)(2020)17-24.

[6] Z.-Q. Zou, Current Situation, Hot spots and Frontiers of Postgraduate Education Research in China, Heilongjiang Researches on Higher Education 36(09)(2018) 86-90.

[7] J. Yu, H.-Q. Huang, Y. Xi, Z.-X. Zhu, Interpersonal Relationship Reasoning Based on Knowledge Graph in Graph Database, Information Science 37(10)(2019) 8-12.

[8] X. Yang, Y. Wang, Y. Li, Y. Wang, PBasic Framework of the Opening and Sharing System of Educational Government Data, China Academic Journal Electronic Publishing House (09)(2020) 65-73.

[9] P. Yu, Y. Li, Research on Higher Education Data Ecological Governance Architecture Under the Perspective of Smart Campus, China Academic Journal Electronic Publishing House (05)(2020) 88-100.

[10] K. Chen, Y.-L. Tan, Knowledge Map of Ideological and Political Courses in China Based on CiteSpace, Heilongjiang Researches on Higher Education 38(02)(2020) 128-132.

[11] S.-J. Lin, P.-Q. Shen, Chinese University Logistics Research in New Era: Cooperation, Evolution, Hotspots and Prospects, Heilongjiang Researches on Higher Education 37(08)(2019) 89-93.

[12] NSF, Survey of Earned Doctorates, <https://www.nsf.gov/statistics/ srvydoctorates.pdf>, 2019.

[13] Z. Li, G. Lu, Y. Gao, T. Xiang, Big Data-based Analysis of Monitoring and Warning on Study Duration Doctoral Students Spent for PhD Degree, Journal of Graduate Education10(05)(2020) 15-22.