Construction and Application of a Major-Specific Knowledge Graph Based on Big Data in Education

https://doi.org/10.3991/ijet.v17i07.30405

Zhihao Wu[1], Fang Jia[2](✉)
[1] College of Foreign Languages, Cangzhou Normal University, Cangzhou, China
[2] English Section of Public Course Teaching Department, Hebei university of Chinese medicine, Shijiazhuang, China
jiafang@hebcm.edu.cn

Abstract—The main problems lying in the learning process of learners in different majors are that they study blindly and do not have a complex knowledge structure, which are seriously affecting their learning effects. But if a knowledge graph can be modeled for the major-specific curriculum system using the quantitative method from the perspective of knowledge network, it may be able to improve the existing teaching problems and optimize the teaching quality. The existing major-specific knowledge graphs were all constructed in an abstract form, ignoring the inherent prior learning relationship between teaching units and curriculum knowledge. To this end, taking English major as an example, this paper studied the construction and application of a major-specific knowledge graph based on the big data in education. Firstly, the English major-specific knowledge graph was modeled, the calculation process of node importance was shown, and a localized graph of the knowledge network of English major courses was given. Then, a multi-node feature selection framework for the English major-specific knowledge graph was constructed based on the context of nodes, and the importance of the top k nodes in the constructed knowledge graph was extracted using the multi-node feature extraction technology. After that, the experimental results verified the stability and connectivity of the nodes in the constructed knowledge graph.

Keywords—major-specific knowledge, knowledge graph, node importance, node features

1 Introduction

With the rapid development of science and technology, new knowledge and technologies are rapidly growing, leading to increasingly large and complex knowledge systems for all kinds of majors [1-8]. When studying the major-specific knowledge, however, learners currently have some problems. For example, they study blindly and do not have a complex knowledge structure, which are seriously affecting their learning effects [9-14]. The existing English major-specific knowledge system has been relatively mature and stable, but in the actual teaching process, there are also some problems in both teaching and learning [15-18]. If a knowledge graph can be modeled for the English major-specific curriculum system using the quantitative method from the perspective of knowledge network, some problems in the English major teaching plan and mode may be found out so that what areas to improve and optimize in English teaching will be known.

With the rapid development of science and technology and the continuous update of knowledge, college teachers and students are required to grasp the latest development trends in the fields related to the curriculum knowledge faster, more comprehensively and more accurately. To this end, many schools and colleges have digitized their educational resources. However, there is no unified knowledge representation structure in the traditional educational resource sharing, which prevents the learning resources from being shared and reused in a satisfactory way. Pei et al. [19] used the ontology technology to build a curriculum knowledge map, which can not only reflect the relationships between knowledge modules, but also realize the mining and representation of more knowledge relations and types such as tacit knowledge to a certain extent. Bielefeldt [20] pointed out that the body of knowledge has outlined the skills and competencies required for a licensed professional engineer and described the skills and competencies that should be acquired as part of an accredited bachelor's degree. Zhang and Li [21] reviewed the literatures on higher vocational classroom research in CNKI, and used research methods such as knowledge graph analysis to draw knowledge graphs of high-frequency keyword relationship maps and keyword co-occurrence maps. At last, the trend of higher vocational classroom research was proposed through the centrality analysis and multi-dimensional scale analysis Atlas of SPSS. Kobets et al. [22] used the cognitive modeling methods of weakly structured systems to construct cognitive maps of educational processes, demonstrated the possibility of using cognitive modeling and cognitive maps in college knowledge management systems and identified target risk factors, basic risk factors and risk factors affecting the quality of educational processes in higher education systems. In order to reveal the research characteristics and status quo of big data in education, Jiang et al. [23] used the software CiteSpace V to analyze 2052 papers on the national knowledge infrastructures in China. The research results show that "ideological and political education" is the most frequently used keyword, and "e-schoolbag", "education reform" and "MOOC" have the longest exposure.

Through the analysis of the existing research results, it can be seen that although the knowledge network has been applied to various scientific fields, few people have integrated it with the major-specific knowledge graphs, especially with the current English major teaching reform. Most of the existing research on major-specific knowledge graphs is based on knowledge relation, such as the English grammar knowledge graph with English grammar rules as the basic nodes. These knowledge graphs were constructed in an abstract form, ignoring the inherent prior learning relationship between teaching units and curriculum knowledge. To this end, this paper conducted research on the construction and application of a major-specific knowledge graph based on the big data in education. Section 2 of the paper first models the English major-specific knowledge graph, shows the process of how the node importance is calculated and gives a localized graph of the knowledge network of the English major courses; Section 3 builds a multi-node feature selection framework for the English major-specific knowledge graph based on the context of the nodes, and extracts the importance of the top k nodes in the constructed knowledge graph using the multi-node feature extraction technology. The experimental results prove the stability and connectivity of the nodes in the constructed knowledge graph.

2 Modeling of the English major-specific knowledge graph

2.1 Model construction

To build the major-specific knowledge graph based on the multi-source big data in education and study the node feature extraction method, the hierarchical framework of the system is given in Figure 1.

The knowledge relations in the English major-specific knowledge graph consisting of three layers can be represented by an adjacency matrix. Suppose that the English major-specific knowledge graph is represented by O_1, that the teaching unit knowledge graph is represented by O_2, and that the course knowledge graph is represented by O_3. Since each adjacency map has a corresponding adjacency matrix, it can be expressed as matrix O_1, O_2, O_3 or map O_1, O_2, O_3 below. O_1 is $O_1=(U,D)$, where $U=\{u_1,u_2,...,u_m\}$ is the set of knowledge points in the knowledge graph, and $D=\{d_{i,j}|i,j=1,2,...,m\}$ is the set of knowledge relations, where $d_{i,j}$ is the directed connection line between knowledge points u_i and u_j in the graph. The parameter representations are similar for the knowledge graphs O_2 and O_3.

In the modeling of the English major-specific knowledge graph O_1, only the direction is considered and the weight is ignored, and at the same time, the closeness of the relations between knowledge points is also ignored. Eq. (1) shows the expression of the matrix O_1:

$$O_1 = \begin{bmatrix} o_{1,1} & o_{1,2} & \cdots & o_{1,1422} \\ o_{2,1} & o_{2,2} & \cdots & o_{2,1422} \\ \cdots & \cdots & \cdots & \cdots \\ o_{1422,1} & o_{1422,2} & \cdots & o_{1422,1422} \end{bmatrix} \tag{1}$$

If the knowledge point u_i is the prior knowledge point of the knowledge point u_j, the value of the binary function $o_{i,j}$ is 1; if the knowledge point u_j is the prior learning knowledge point of the knowledge point u_i, the value of the binary function $o_{i,j}$ is 0.

Since the weights of knowledge graph O_2 and O_3 are obtained based on their connection relationships with the knowledge graph O_1, the matrices O_2 and O_3 cannot be used directly in the analysis of the English major-specific knowledge graph. The effects of the teaching unit and course knowledge scales on the weights of the constructed knowledge graphs cannot be ignored.

In order to eliminate the effect of the knowledge scale, it is necessary to normalize the data in the matrices O_2 and O_3. Eq. (21) shows the specific process:

$$H = \left[h_{i,j} \right] = \left[\frac{o_{i,j}}{\sqrt{o_{i,j} \cdot o_{j,j}}} \right] \tag{2}$$

Suppose that H is the normalized knowledge graph, with the same number of dimensions as that of the corresponding matrix of the original graph. In the knowledge graphs constructed based on teaching units and course knowledge points, the dependency between two teaching units or two courses u_i and u_j is represented by $h_{i,j}$.

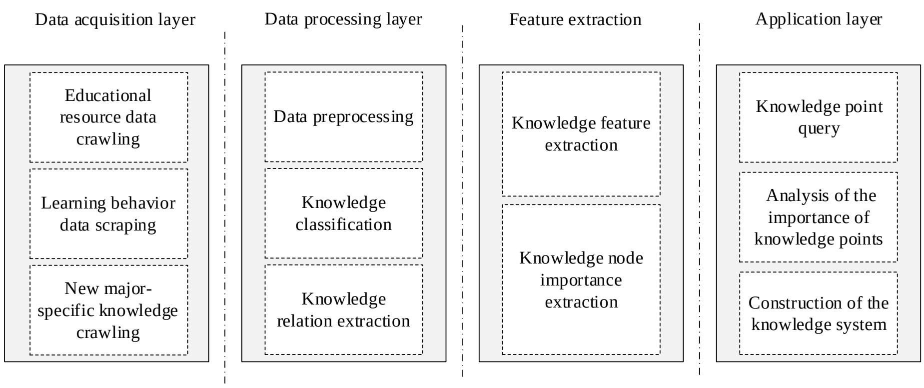

Fig. 1. Hierarchical architecture of the major-specific knowledge graph

2.2 Analysis of the knowledge graph

The node degree in the knowledge graph reflects the importance of the knowledge point it represents in the whole knowledge system of the English major. Knowledge points with large node degrees should be studied as key knowledge by students of English major. The English major-specific knowledge graph can be regarded as a directed network, so the actual significance of its out-degree and in-degree needs to be considered. Assuming that the number of nodes in the knowledge graph is represented by m, and the node relationship is represented by h_{ij}, Eq. (3) and (4) show the out-degree and in-degree calculation formulas of the knowledge graph:

$$l_i^{out} = \sum_{j=1}^{m} h_{ij} \tag{3}$$

$$l_i^{in} = \sum_{j=1}^{m} h_{ji} \tag{4}$$

The English major-specific knowledge graph is composed of several course knowledge graphs, and the degree of closeness between the course knowledge graphs can be characterized by the clustering coefficient of the English major-specific knowledge graph. Eq. (5) gives the calculation formula of the clustering coefficient $\Phi(i)$ of each node in the English major-specific knowledge graph:

$$\Phi(i)=\frac{\left|\left\{d_{j,l}:u_j,u_l\in K(i),d_{j,l}\in D_i\right\}\right|}{l_i\left(l_i-1\right)} \tag{5}$$

Based on the calculation results of the above formula, the clustering coefficient of the entire major-specific knowledge graph is further calculated as follows:

$$\bar{\Phi}=\frac{1}{m}\sum_{i=1}^{m}\Phi(i) \tag{6}$$

In order to measure the connectivity between a node and other nodes in the knowledge graph, the parameter "node betweenness" is introduced. Assuming that the number of the shortest paths between two teaching units or courses u_i and u_j is represented by χ_{ij}, and that the number of the shortest paths passing through node u_i among all shortest paths between u_i and u_j is represented by $\chi_{ij}(\tau)$, Eq. (7) shows the formula to calculate the node betweenness of any node as follows:

$$\begin{aligned}\Phi_Y(\tau)&=\sum_{i\neq j\neq\tau}\gamma_{ij}(\tau)\\ \gamma_{ij}(\tau)&=\frac{\chi_{ij}(\tau)}{\chi_{ij}}\end{aligned} \tag{7}$$

2.3 Calculation of importance

The most important purpose of applying various statistical characteristics of knowledge graph in the analysis of English major-specific knowledge graph is to find out the important nodes in the knowledge graph, as only by emphasizing the teaching and learning of the important teaching units or knowledge points in courses in English major study can we ensure the stability and connectivity of the learners' English major-specific knowledge structures.

In this paper, node proximity, neighborhood, criticality and importance degree are used to define the importance of nodes in the knowledge graph. Assuming that e_{ij} is the length of the shortest path with u_i as the starting point and u_j as the ending point, Eq. (8) shows how to calculate the proximity $DJ(i)$ of u_i:

$$DJ(i)=\frac{1}{\sum_{j=1}^{m}e_{ij}},\ j\neq i \tag{8}$$

Assuming that the degree of the node u_i is represented by $l=|\gamma_{li}|=\Sigma_{uj\in\gamma li}o_{ij}$, Eq. (9) gives the judgment formula of the neighborhood of u_i:

$$\gamma_{li} = \left\{ u_j \mid u_j \in U, o_{ij} = 1, j = 1, 2, ..., m \right\} \tag{9}$$

In the neighborhood γ_{li} of the node u_i with the degree l, if l is greater than 2, assuming that the number of the shortest paths between any pair of nodes passing through u_i is represented by $R(i)$, and that the number of the shortest paths that do not pass through u_i is represented by $Y(i)$, then the criticality of the node u_i can be calculated according to Eq. (10):

$$L(i) = \frac{R(i)}{R(i) + Y(i)} \tag{10}$$

If l is equal to 1, then the criticality $L(i)$ of u_i is equal to 0. The node importance can be used to measure the local and global importance of each node in the knowledge graph. Eq. (11) gives the calculation formula of the node importance $SQ(i)$:

$$SQ(i) = DJ(i) L(i) = \frac{R(i)}{\left[\sum_{j=1}^{m} e_{ij} \right] \left[R(i) + Y(i) \right]} \tag{11}$$

It can be seen from the above analysis that, to calculate $SQ(i)$, $L(i)$ needs to be calculated first, and that to calculate γ_{li}, $R(i)$ and $Y(i)$ need to be obtained first. Therefore, the connectivity of u_i in its neighborhood is mainly affected by the nodes in its neighborhood and the intersection of the neighborhoods of the two nodes in its neighborhood. The key domain of the node u_i can be calculated according to Eq. (12):

$$G_i = \left\{ u_r \mid u_r \in \left(\gamma_{l_r r} \bigcap_{\forall u_r, u_j \in \gamma_{li}} \gamma_{l_k k} \right) \bigcup \gamma_{li} \right\} \tag{12}$$

If the set of shortest paths between any two nodes u_r and u_j in the neighborhood γ_{li} of the node u_i is represented by $K(u_r,u_j)$, there is:

$$K(u_r, u_j) = \left\{ \left\{ u_r, u_j \right\} \text{or} \left\{ u_r, u_j, u_j \right\} \right\} \tag{13}$$

Assuming that there are q_{rj} shortest paths between u_r and u_j, then $R(i)$ can be calculated according to Eq. (14):

$$R(i) = \sum_{G_i} r(i), where\ r(i) = \begin{cases} 1/q_{rj}, u_i \in K(u_r, u_j) \\ 0, u_i \notin K(u_r, u_j) \end{cases} \tag{14}$$

$Y(i)$ is calculated as:

$$Y(i) = \sum_{G_i} y(i), where\ y(i) = \begin{cases} 0, u_i \in K(u_r, u_j) \\ 1, u_i \notin K(u_r, u_j) \end{cases} \tag{15}$$

After the calculation of $R(i)$ and $Y(i)$ is completed, the criticality $L(i)$ can be further calculated. The Dijkstra's algorithm based on the greedy idea is used to calculate the shortest path e_{ij} of all nodes, and then $DJ(i)$ can be obtained. Based on the obtained $L(i)$ and $DJ(i)$, $SQ(i)$ can be calculated.

Considering there are so many teaching units or course knowledge points contained in the English major-specific knowledge network, this paper selected some knowledge points of the course "English Linguistics" for demonstration. Figure 2 clearly shows the local topology of the core knowledge point network of the course. The nodes of different sizes in the graph indicate their different degree values.

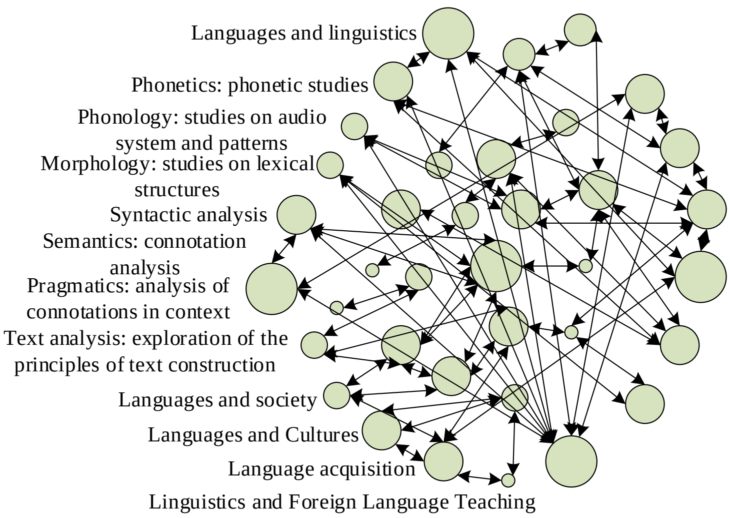

Fig. 2. Localized graph of the knowledge network of an English course

3 Extraction of knowledge graph node features

With the continuous expansion of the English major-specific knowledge graph, the attributes of the nodes representing teaching units or course knowledge points will continue to increase, which will lead to sparse or redundant associations between node attributes and thus make learners lost in the knowledge. Therefore, this paper proposed a multi-node feature selection framework for the English major-specific knowledge graph based on the context of the nodes, and extracted the importance of the top k nodes in the constructed knowledge graph using the multi-node feature extraction technology. Figure 3 presents the multi-node feature selection model for the major-specific knowledge graph.

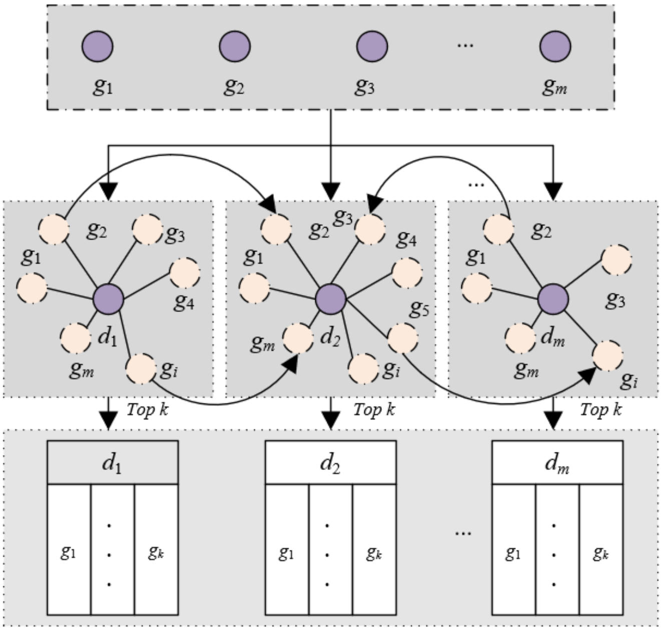

Fig. 3. Multi-node feature selection model for the major-specific knowledge graph

In this paper, the importance of node features is measured based on the IT-IDF algorithm. Eq. (16) shows the calculation formula of the INF(g) of feature g:

$$INF(g) = log\left(\frac{m}{\left|\left\{d \mid g \in GR(d)\right\}\right|}\right) \tag{16}$$

It can be seen from the above formula that the more nodes there are with a certain feature g, the less feature information g contains, and the smaller the INF(g) value is. Assuming that the triple composed of the nodes and their attribute features is represented by Ψ, the IT(g) feature is calculated in this paper based on the attribute name AN(g) and the attribute value AV(g):

$$\begin{aligned} IT(\varepsilon) = log\left|\Psi\right| \exists d, g : \tau \text{"} appears\ in \text{"} H\ and \\ \tau \equiv \left(d\ AN(g)\ AV(g)\right)\ and\ AV(g) = \varepsilon \}\right| \end{aligned} \tag{17}$$

From the above formula, it can be seen that the more times the triple Ψ composed of the attribute feature with the value ε and its corresponding nodes appears in the knowledge graph, the more important the feature information contained in the attribute feature corresponding to the value is, and the more popular the knowledge is. Eq. (18) shows the node feature importance evaluation formula that considers both the INF(g) feature and the IT(g) feature:

$$NCI(g) = INF(g) * IT(\mathrm{val}(g)) \tag{18}$$

In order to fully consider the redundancy and correlation of the core node feature set, the node feature similarity function is introduced in this paper. Suppose that the similarity calculation function based on feature edit distance is represented by SIMED( ), that the calculation function for the semantic similarity of two features by SIMED( ), and that the calculation function for the local graph structure similarity of two features by SIMST( ). Eq. (19) shows the formula for calculation of the similarity between the features g^{j}_{di} and g^{k}_{dl}:

$$SIM\left(g^{l}_{d_i}, g^{k}_{d_j}\right) = \frac{SIM_{ED}\left(g^{l}_{d_i}, g^{k}_{d_j}\right) + SIM_{SE}\left(g^{l}_{d_i}, g^{k}_{d_j}\right) + SIM_{ST}\left(g^{l}_{d_i}, g^{k}_{d_j}\right)}{3} \tag{19}$$

When i and j are equal, the above formula is used to calculate the feature similarity within a node in the knowledge graph, and when i and j are not equal, it is used to calculate the feature similarity between nodes. Suppose that the knowledge point strings that need similarity calculation are represented by r_i and r_j, where the knowledge point strings can be presented in a mixed form of Chinese and English. Suppose that the length of r_j is represented by $W(r_j)$, that $SIMED(r_i,r_j)\in[0,1]$, and that $Eg(r_i,r_j)\in[0,\max(W(r_i), W(r_j)]$. Eq. (20) gives the formula for calculation of feature similarity based on edit distance:

$$SIM_{ED}\left(r_i, r_j\right) = \frac{1}{3\sum_{g\in\{z,t,q\}}} 1 - \frac{E_g\left(r_i, r_j\right)}{max\left(W\left(r_i\right), W\left(r_j\right)\right)} \tag{20}$$

SIMSE( ) can be calculated by the traditional semantic sequence kernel function. Eq. (21) calculates the similarity between the local graph structures of the given two features f^{a}_{ei} and f^{b}_{ei} as follows:

$$SIM_{ST}\left(\mathrm{val}\left(g^{l}_{d_i}\right), \mathrm{val}\left(g^{l}_{d_j}\right)\right) = \frac{\overrightarrow{\mathrm{val}\left(g^{l}_{d_i}\right)} \bullet \overrightarrow{\mathrm{val}\left(g^{k}_{d_j}\right)}}{\left|\overrightarrow{\mathrm{val}\left(g^{l}_{d_i}\right)} \| \overrightarrow{\mathrm{val}\left(g^{k}_{d_j}\right)}\right|} \tag{21}$$

Based on the three constraints - high importance of internal features, low similarity between internal features, and high similarity between entities, the multi-node core features are extracted from the knowledge graph. It is required to extract a core feature set with a length of l for each node in the relevant node set $\{d_1,d_2,...,d_m\}$. If it is converted into a 0-1 knapsack problem, the objective function is set as follows:

$$\begin{aligned} &\max imize \sum_{i=1}^{|Res(d)|} \sum_{j=i}^{|Res(d)|} \sum_{x=1}^{|GR(d_i)|} \sum_{y=1}^{|GR(d_j)|} q_{g^{x}_{d_i}, g^{y}_{d_j}} * a_{i,x} * a_{j,y} \\ &s.t. \sum_{x=1}^{|GR(d_j)|} a_{i,x} \le min\left\{L, |GR(d_i)|\right\}, a_{i,x} \in \{0,1\} \end{aligned} \tag{22}$$

Assuming that the preset size of the set of core features that need to be extracted for each node is represented by Φ, and that the knapsack profit brought by the selected features g^{x}_{di} and g^{y}_{di} by $q_{x,y}$, the calculation formula is as follows:

$$q_{x,y} = \begin{cases} \eta_1 \cdot rank\left(g_{d_i}^{x}\right),\ if\ \ d_i = d_j, x = y \\ -\eta_2 \cdot sim\left(g_{d_i}^{x}, g_{d_j}^{y}\right),\ if\ \ d_i = d_j, x \neq y \\ \eta_3 \cdot sim\left(g_{d_i}^{x}, g_{d_j}^{y}\right), if\ \ d_i \neq d_j \end{cases} \tag{23}$$

The value range of the weight coefficients η_1, η_2, and η_3 in the above formula is [0, 1]. η_1 and η_2 are used to adjust the novelty of the features within the nodes in the knowledge graph, and η_3 is used to adjust the degree of mutual influence between nodes. The larger the η_3 is, the more correlated the extracted node features are, and conversely, the less correlated they are.

4 Experimental results and analysis

Table 1. Courses contained in the modeling of the English major-specific knowledge graph

No.	Course	No.	Course	No.	Course
1	Comprehensive English	11	International Finance	21	Interpretation Techniques & Practice
2	English & American Literature	12	English Reading	22	International Trade Practices
3	Advanced English	13	English Viewing, Listening and Speaking	23	Ideological and Moral Cultivation and Basic Law Education
4	Introduction to English Literature	14	English Grammar and Substantive Writing	24	College Chinese
5	Introduction to Linguistics	15	Introduction to English Graduation Thesis Writing	25	Physical Education
6	Special Course of Contemporary English and American Culture	16	English Speeches and Presentations	26	Introduction to Maoism and Socialist Theoretical System with Chinese Characteristic
7	English Lexicology	17	Advanced English Listening	27	English Prose
8	The Society and Culture of Major English-speaking Countries	18	English Stylistics	28	English & American Literature
9	Principles of Economics	19	English-Chinese Translation of Business Documentation	29	International Business Simulation Training
10	International Trade	20	Chinese-English Translation	30	

This paper analyzed the structural characteristics of the constructed English major-specific knowledge graph. According to the statistics of degrees, the English major-specific knowledge graph is closer to $DF(l)\sim\mu l^{-\eta}$. Figure 4 shows the distribution function of the out-degrees and in-degrees of nodes, which is similar to the power-law distribution. In the fitting function of the in-degree distribution, η = 0.7642 and μ = 0.371, and in the fitting function of the out-degree distribution, η= 0.6977 and μ = 0.314.

As can be seen from the figure, most of the teaching units or course knowledge points in the English major-specific knowledge graph have only a few connections in knowledge, which indicates that the constructed knowledge graph conforms to the scale-free network feature. In view of this feature, it is necessary to strengthen the learning of the teaching units or course knowledge points with high connections in knowledge in the Englis major teaching. If such nodes are missing, it will seriously affect the learning of subsequent teaching units or course knowledge points, and exert a great impact on the stability and connectivity of learners' English knowledge structures.

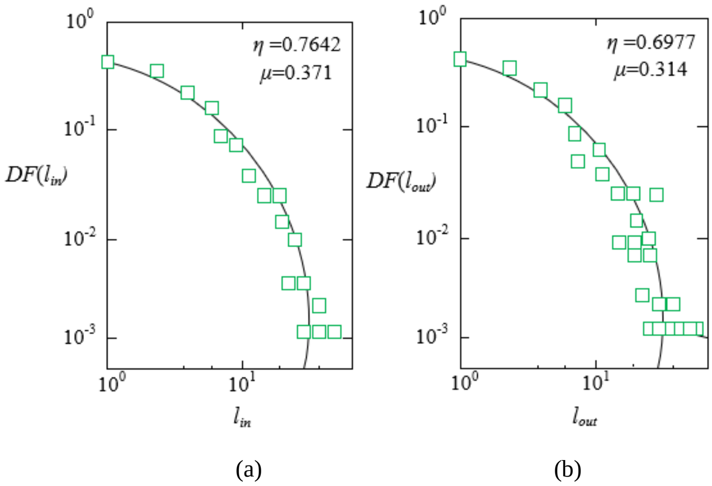

Fig. 4. Fitting functions of the in-degree and out-degree distributions of nodes

For the convenience of calculation, it is assumed that the length of each connection line of the English major-specific knowledge graph are equal. Next, the importance of different teaching units or courses is calculated, with the results shown in Table 2. Due to the large number of nodes, only part of the courses are listed here.

It can be seen from Table 2 that, the courses with high importance in English major are all the basic courses and specialized courses located in the center of the knowledge graph, which play more important roles in promoting the stability and connectivity of the English major-specific knowledge graph. While the basic courses such as Chinese-English translation are the basis of English learning, their values are not high in the importance evaluation, but that does not mean such courses are not important; instead, it is because such courses are mostly located at the endings of the knowledge graph. When the global importance of the major-specific knowledge graph is considered, the importance values of these courses will decrease.

The descriptive statistics of the structure attributes and centrality of the knowledge graphs for the 10 teaching units of the course "English Linguistics" are shown in Tables 3 and 4. The mean centrality of the sub-network of the unit "Pragmatics: Analysis of Connotations in Context" is 14.217, greater than those of the knowledge graphs of other teaching units, showing that the knowledge points in the knowledge graph of this teaching unit are highly connected with other knowledge points. When the knowledge points in the knowledge graph of this teaching unit change, it is very likely that the other knowledge points will also change. Therefore, the knowledge points in the teaching unit "Pragmatics: Analysis of Connotations in Context" play an important basic role in the whole course "English Linguistics", and mastering the knowledge points in this teaching unit will have a positive effect on the learning of the subsequent courses or teaching units and even help learners master the whole English major-specific knowledge system to some extent.

Table 2. Calculation results of node importance

Course	Node proximity	Course	Node criticality	Course	Node importance
Comprehensive English	0.048	Advanced English	1.001	Comprehensive English	0.045
English & American Literature	0.056	Introduction to English Literature	0.702	English & American Literature	0.041
Advanced English	0.051	Introduction to Linguistics	0.635	Advanced English	0.026
Introduction to English Literature	0.042	Special Course of Contemporary English and American Culture	0.528	Introduction to English Literature	0.014
Introduction to Linguistics	0.046	Principles of Economics	0.511	Introduction to Linguistics	0.014
Special Course of Contemporary English and American Culture	0.035	International Trade	0.462	Introduction to Linguistics	0.011
English Lexicology	0.033	International Finance	0.402	Special Course of Contemporary English and American Culture	0.013
The Society and Culture of Major English-speaking Countries	0.031	English Reading	0.425	Principles of Economics	0.011
English Speeches and Presentations	0.033	English Viewing, Listening and Speaking	0.403	International Trade	0.011
Advanced English Listening	0.031	International Business Simulation Training	0.415	International Finance	0.008

Table 3. Structural attribute values of the knowledge graph of each teaching unit

Teaching unit No.	1	2	3	4	5	6	7	8	9	10
Number of nodes	175	89	97	102	131	115	102	75	66	45
Number of connection lines	2526	853	914	1025	1154	1138	1268	652	564	385
Network density	0.074	0.0.98	0.094	0.091	0.075	0.093	0.115	0.138	0.146	0.185
Average path length	1.859	1.905	1.936	1.948	1.907	1.962	1.833	1.858	1.811	1.848
Mean clustering coefficient	0.915	0.859	0.862	0.874	0.885	0.894	0.958	0.882	0.869	0.847

Table 4. Descriptive statistics of the centrality of the knowledge graph for each teaching unit

Teaching unit No.	1	2	3	4	5
Mean centrality	12.859	8.495	9.153	9.257	11.263
Minimum out-degree/in-degree	2	2	2	3	3
Maximum out-degree/in-degree	174	85	96	112	105
Standard deviation of out-degree/in-degree	12.958	8.457	9.152	9.638	9.748
Teaching unit No.	6	7	8	9	10
Mean centrality	8.749	12.592	8.519	8.695	8.742
Minimum out-degree/in-degree	4	2	3	4	2
Maximum out-degree/in-degree	102	116	75	62	43
Standard deviation of out-degree/in-degree	9.624	10.847	12.625	8.953	8.624

5 Conclusions

This paper studied the construction and application of a major-specific knowledge graph based on the big data in education. Firstly, the English major-specific knowledge graph was modeled, the calculation process of node importance was shown, and a localized graph of the knowledge network of English major courses was given. Then, a multi-node feature selection framework for the English major-specific knowledge graph was constructed based on the context of nodes, and the importance of the top k nodes in the constructed knowledge graph was extracted using the multi-node feature extraction technology. The experiment section listed the courses contained in the modeling of the English major-specific knowledge graph, and analyzed the structural characteristics of the constructed English major-specific knowledge graph. It also showed the fitting functions of the in-degree and out-degree distributions of nodes and the calculation results of node importance, selected the courses with higher importance. According to the descriptive statistics of the structural attribute values and centrality of the knowledge graph of each teaching unit obtained, the teaching unit that plays an important basic role in the content of the course “English Linguistics” was identified.

6 References

[1] Wang, J.G., Qi, Y.M. (2020). Influence of teacher-student matching on knowledge innovation of graduate students. International Journal of Emerging Technologies in Learning, 15(20): 87-101. https://doi.org/10.3991/ijet.v15i20.17421

[2] Squires, A., Pyster, A., Sauser, B., Olwell, D., Enck, S., Gelosh, D., Anthony, J. (2010). Applying systems thinking via systemigrams™ for defining the body of knowledge and curriculum to advance systems engineering (BKcase) project. In INCOSE International Symposium, 20(1): 739-753. https://doi.org/10.1002/j.2334-5837.2010.tb01101.x

[3] Al-Abdullatif, A.M., Gameil, A.A. (2020). Exploring students' knowledge and practice of digital citizenship in higher education. International Journal of Emerging Technologies in Learning, 15(19): 122-142. https://doi.org/10.3991/ijet.v15i19.15611

[4] Bots, P.W., Thissen, W.A. (2000). Negotiating knowledge in systems engineering curriculum design: Shaping the present while struggling with the past. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 30(2): 197-203. https://doi.org/10.1109/5326.868441

[5] Yu, X., Stahr, M., Chen, H., Yan, R. (2021). Design and implementation of curriculum system based on knowledge graph. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 767-770. https://doi.org/10.1109/ICCECE51280.2021.9342370

[6] Fazlollahtabar, H., Muhammadzadeh, A. (2012). A knowledge-based user interface to optimize curriculum utility in an e-learning system. International Journal of Enterprise Information Systems (IJEIS), 8(3): 34-53. https://doi.org/10.4018/jeis.2012070103

[7] Squires, A.F., Olwell, D.H., Ferris, T.L., Hutchison, N.A., Pyster, A., Enck, S., Gelosh, D. (2011). Developing systems engineering graduate programs aligned to the body of knowledge and curriculum to advance systems engineering (BKcase (tm)) guidelines. https://doi.org/10.21236/ADA522654

[8] Lu, W., Zhang, J. (2011). Construction of curriculum ontology aiming at educational service support: an ontology approach for knowledge-intensive service systems. In 2011 International Joint Conference on Service Sciences, 315-318. https://doi.org/10.1109/IJCSS.2011.70

[9] Choo, D., Dushyanthen, S., Gray, K., Capurro, D., Merolli, M., Chapman, B. E., Lyons, K. (2022). Designing a professional development online short course to foster Learning Healthcare Systems. International Journal of Medical Informatics, 158: 104666. https://doi.org/10.1016/j.ijmedinf.2021.104666

[10] Hsiao, Y.C., Shiao, Y.T., Yang, Y. H. (2019). The role of self-regulated learning in the workplace professional English e-learning course. In Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning, 205-210. https://doi.org/10.1145/3306500.3306531

[11] Bozic, M., Certic, J.D., Vukelic, M., Cizmic, S. (2018). New instructional approach for fostering generic and professional competences: case study of the project and problem based learning engineering practice course. The International journal of engineering education, 34(5): 1581-1591.

[12] González Lara, G.S., Ferchow, E., Eguia, L., Loera, V., Velázquez, L. (2017). Art and technology: course designed on competency-based learning to form the professional profile of the actual engineer. 2017 Research in Engineering Education Symposium, Research in Engineering Education Symposium.

[13] Lieberoth, A., Hanghøj, T. (2017). Developing professional "game teacher" repertoires: Describing participants and measuring effects in a Danish college course on game based

learning. In Proceedings of the 11th European Conference on Game-Based Learning. ECGBL 2017, 377-386.
[14] Li, Y. (2021). Construction and practice of blended learning mode of English writing course based on information technology. In 2021 4th International Conference on Information Systems and Computer Aided Education, 1032-1037. https://doi.org/10.1145/3482632.3483077
[15] Fang, Y., Li, J. (2021). Application of the deep learning algorithm and similarity calculation model in optimization of personalized online teaching system of English course. Computational Intelligence and Neuroscience, 2021. https://doi.org/10.1155/2021/8249625
[16] Zhang, M. (2021). Diversification of the evaluation mode of English mixed gold course learning mode in colleges and universities based on mixed learning. In International Conference on Data and Information in Online, 431-437. https://doi.org/10.1007/978-3-030-77417-2_36
[17] Huang, M. (2021). A study of English course evaluation, satisfaction and perception toward virtual learning online from an empirical perspective. In 2021 International Conference on Internet, Education and Information Technology (IEIT), 350-353. https://doi.org/10.1109/IEIT53597.2021.00083
[18] Yu, J. (2020). The online English course activities design based on the motivation mechanism of adult online learning. In 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), 277-282. https://doi.org/10.1109/CSEI50228.2020.9142487
[19] Pei, P., Xuejing, D., Deqing, Z. (2020). Construction of curriculum knowledge map based on ontology. In Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence, 259-265. https://doi.org/10.1145/3438872.3439091
[20] Bielefeldt, A. (2009). Mapping an undergraduate curriculum onto the environmental engineering body of knowledge. In 2009 Annual Conference & Exposition, 14-860. https://doi.org/10.18260/1-2--4925
[21] Zhang, W., Li, L. (2020). The knowledge map analysis about the theme and trend of classroom study in higher vocational education in China. In Journal of Physics: Conference Series, 1453(1): 012144. https://doi.org/10.1088/1742-6596/1453/1/012144
[22] Kobets, N., Kovaliuk, T., Mozoliova, D. (2019). Cognitive modeling and cognitive map applying to the knowledge management in the higher education system. In International Conference on Information Systems Architecture and Technology, 63-73. https://doi.org/10.1007/978-3-030-30443-0_6
[23] Jiang, C., Wang, Q., Qing, W., Zhu, L., Cheng, S., Wang, H. (2019). The present of education big data research in China: Base on the bibliometric analysis and knowledge mapping. In Journal of Physics: Conference Series, 1187(5): 052039). IOP Publishing. https://doi.org/10.1088/1742-6596/1187/5/052039

7 Authors

Zhihao Wu, male, ethnic Hui, was born in Cangzhou, northern China's Hebei Province in 1986, holds a master's degree in Foreign Linguistics and Applied Linguistics, and now works as a lecturer of College English, Cross-Cultural Communication, Intensive English Reading, Extensive English Reading, and Listening at Cangzhou Normal University. The main research directions include second Language acquisition, cross-cultural communication, and English language teaching. He chaired

a provincial-level project, 5 municipal- and department-level projects, and one school-level educational reform project. He has published 10 papers (email: wzh1986@caztc.edu.cn).

Fang Jia, female, ethnic Han, was born in Shijiazhuang, northern China's Hebei Province in 1978, holds a master's degree in English, and now works as an associate professor and English teacher of public courses at Hebei University of Chinese Medicine. The main research directions include English teaching and cross-cultural communication. She chaired 2 provincial-level projects, 3 municipal- and department-level projects and 4 school-level projects, won 4 academic awards at school/ department level or above and 4 other academic awards, and co-edited 2 textbooks. She has published 11 papers in national core journals (being the first author of 7 papers) (email: jiafang@hebcm.edu.cn).

Article submitted 2022-02-07. Resubmitted 2022-03-12. Final acceptance 2022-03-14. Final version published as submitted by the authors.