Research on the Path of Enhancing Employment and Entrepreneurship Ability of Deaf College Students Based on Knowledge Graph

Pengyu Liu*
Special Education College Changchun University, Changchun 130022, China

Abstract—Enhancing employment capabilities and selecting suitable career paths are crucial for deaf university students. The advancement of knowledge graph technology has opened up technical possibilities for career decision-making among these students. This paper calculates user preferences and introduces an exponential decay function integrated with a time factor to accurately reflect the dynamic changes in user interest preferences over time. Leveraging knowledge graphs for personalized recommendations, the study proposes recommending necessary skills to enhance employment and entrepreneurial capabilities among students. Additionally, it employs knowledge graphs to suggest more suitable career paths for deaf university students. Finally, through empirical validation, the paper demonstrates the effectiveness of the proposed hybrid clustering and interest-based collaborative filtering recommendation algorithm.

Keywords—Knowledge graph; hearing impaired college students; employment and entrepreneurial ability; interest matching; feature extraction

I. INTRODUCTION

Enhancing employability is crucial for hearing-impaired college students, who often face numerous challenges in the job market, such as difficulties in accessing information, limited communication abilities, and social prejudice. Improving their employability not only increases their competitiveness in the workplace but also promotes their social integration and self-fulfillment. This enhancement encompasses several aspects, including the acquisition of professional skills, the development of workplace soft skills, and the strengthening of adaptability and autonomous learning abilities. Achieving these improvements requires the collective effort of schools, governments, and various societal sectors through systematic education, training, and support measures to help hearing-impaired students better meet the demands of the job market [1]-[2].

Currently, the employment situation for hearing-impaired college students remains challenging. Despite the increase in employment opportunities due to the widespread availability of education and heightened social awareness, the overall employment rate is still relatively low. Many employers harbor misconceptions and biases about the abilities of hearing-impaired students, leading to discrimination and unfair treatment during the job search process [3]. Additionally, hearing-impaired students face significant limitations in career choices, as many high-skill, high-income professions remain inaccessible to them. Therefore, improving their employment situation requires a multi-faceted approach involving policy measures, corporate responsibility, and social support to foster a more inclusive and diverse employment environment [4].

A knowledge graph is a knowledge representation method based on graph structures, which expresses entities and their relationships through nodes and edges. The concept of the knowledge graph was introduced by Google in 2012, aiming to provide a semantically rich knowledge base through ontology standardization and information integration. Knowledge in knowledge graphs is represented and stored in the form of triples (e.g., <entity, relationship, entity> or <entity, attribute, attribute value>), enabling the structured management and querying of complex knowledge. Widely used in search engines, intelligent recommendations, and natural language processing, knowledge graphs offer users more accurate and relevant information services.

Knowledge graph technology holds significant potential to enhance the employment and entrepreneurship capabilities of hearing-impaired college students [5]-[6]. First, by constructing a knowledge graph related to these students, one can comprehensively understand their educational background, skill sets, and career interests, thus providing them with the most suitable career recommendations. For instance, a knowledge graph can integrate job requirement information from various industries and match it with the skills and interests of hearing-impaired students, offering personalized career advice.

Second, knowledge graph technology can be used to design personalized learning paths and skill training programs. Based on the career goals of hearing-impaired students, knowledge graphs can recommend relevant learning resources and courses, helping them systematically improve their professional skills and workplace soft skills. Additionally, knowledge graphs can track learning progress, provide real-time feedback and improvement suggestions, ensuring the effectiveness and relevance of the training.

Moreover, knowledge graph technology can support the entrepreneurial activities of hearing-impaired students. By analyzing market demands and industry trends, knowledge graphs can offer entrepreneurial guidance and resource support, helping students identify market opportunities and formulate scientific business plans. Knowledge graphs can also connect entrepreneurs with investors, mentors, and partners.

*Corresponding Author.
This work is supported by 2024 Jilin Provincial Department of Education Employment and Entrepreneurship and Student Management Special Project (number: JJKH20240769JY).
creating an ecosystem that supports the entrepreneurship of hearing-impaired students.

In summary, leveraging knowledge graph technology to recommend suitable career paths for hearing-impaired students and enhance their employment and entrepreneurship capabilities can significantly improve their integration into society and create more value for the community. Through scientific technological methods and systematic support measures, we can effectively improve the employment situation of hearing-impaired students, fostering their professional development and personal growth.

II. LITERATURE REVIEW

In this chapter, we introduce the development of knowledge spectrum and employability and summarize the research gaps by highlighting them.

A. Knowledge Graph

Numerous scholars abroad have made significant contributions to the development of knowledge graph technology. Hildrun Kretschmer, a renowned German scientometrician, has achieved notable results in the study of three-dimensional models of scientific collaboration, significantly advancing the field of knowledge graphs [7]-[8]. E.C. Noyons and colleagues at Leiden University in the Netherlands developed a set of mathematical methods for bibliometric mapping, further enhancing the development of knowledge graph technology.

Xiao et al. [9] provided a comprehensive review of knowledge graphs in manufacturing process planning. Analyzed the key technologies of process knowledge graph, including process knowledge representation, process knowledge extraction, process knowledge graph construction, process knowledge graph refinement, process knowledge graph validation, and process generation. Wang et al. [10] suggest using knowledge graphs for code or API recommendations, vulnerability mining, and localization to improve development and design efficiency and accuracy. Fettach et al. [11] use knowledge graphs to represent these data is useful for determining job market demand and establishing better evaluation methods.

The development of knowledge graph construction techniques has been rapid and increasingly sophisticated.

B. Employment and Entrepreneurship Level of Hearing-Impaired College Students

In the "Internet Plus" era, technology offers hearing-impaired college students opportunities to overcome barriers, allowing them to leverage information technology and choose home-based work through new media platforms [12]-[14]. This approach not only improves their employment prospects but also fosters innovation in employment models, aiding their adaptation to the workplace and societal environments. "Internet Plus" has revitalized the employment landscape in China, creating new opportunities. The employment models for hearing-impaired college students in the context of digitalization can be categorized into three types: direct employment, outsourced employment, and self-employment. However, these students face numerous challenges in employment and entrepreneurship due to limitations in educational attainment, knowledge reserves, and access to higher education, compounded by insufficient policy support and financial incentives for entrepreneurship.

Analyzing the opportunities and challenges for flexible employment of hearing-impaired college students through the Internet reveals that, on one hand, the development of the Internet has improved their employment quality and provided more job opportunities. On the other hand, the core human capital of hearing-impaired students in the market remains relatively low, necessitating continuous enhancement of their human capital. Scholars have proposed the concept of the "digital divide," indicating that hearing-impaired students cannot enjoy equal rights in acquiring and using relevant skills compared to other groups, leading to new issues of information inequality, which pose significant challenges to their employment [15].

Currently, the Internet facilitates employment by expanding the social networks of hearing-impaired students, thereby mitigating the spatial and temporal limitations caused by physical disabilities. Artificial intelligence (AI) technology, based on computer science, can significantly enhance the labor skills of hearing-impaired students, improving the quality of employment [16]-[17].

C. Research Gaps

In summary, the enhancement of employment and entrepreneurship capabilities for hearing-impaired college students involves addressing several critical issues:

1) Analyzing the Required Employment and Entrepreneurship Skills for Hearing-Impaired Students: It is essential to identify and understand the specific skills and competencies that hearing-impaired students need to succeed in the job market and entrepreneurial ventures. This involves assessing their educational backgrounds, existing skill sets, and the unique challenges they face.

2) Utilizing Knowledge Graphs to Recommend Relevant Skills: Knowledge graphs can play a pivotal role in guiding hearing-impaired students towards acquiring the necessary skills. By mapping out the relationships between various skills, job requirements, and educational resources, knowledge graphs can provide personalized recommendations for skill development, thereby enhancing their employability and entrepreneurial capabilities.

3) Leveraging Knowledge Graphs to Suggest Suitable Career and Entrepreneurship Paths: In addition to skill recommendations, knowledge graphs can be employed to suggest the most suitable career and entrepreneurial paths for hearing-impaired students. By integrating data on industry trends, job market demands, and individual preferences, knowledge graphs can help students identify and pursue opportunities that align with their strengths and interests.

The aforementioned points are crucial in improving the employment and entrepreneurship prospects of hearing-impaired college students. Given these challenges, this paper will further explore in subsequent sections how innovative
technological approaches and methodologies can be utilized to help hearing-impaired students more effectively acquire the necessary skills and competencies. By addressing these issues, we aim to provide a comprehensive framework that supports their professional development and integration into the workforce.

III. PROPOSED NATURAL LANGUAGE PROCESSING MODEL

This chapter introduces our proposed methods for improving the employment and entrepreneurship level of hearing-impaired college students.

To analyze the employment and entrepreneurship capabilities of hearing-impaired students, we extracted data for the 2021 to 2023 cohorts from the academic administration system, including student registration records, academic transcripts, and course schedules. This data was then preprocessed using Python and office tools. Preprocessing was necessary for several reasons:

1) Filtering irrelevant information: The raw data contained numerous irrelevant entries that needed to be removed before importing into the Neo4j graph database.

2) Data quality issues: Many students had incomplete or poor-quality data due to multiple course failures, missed exams, or withdrawals. This data required cleaning and processing to be usable.

3) Removing redundancy: The integrated data had redundancies that needed to be eliminated to ensure consistency and accuracy without altering the original data.

We employed the Neo4j graph database for storing knowledge points. Unlike traditional relational databases, which store data in table fields, graph databases store data and the relationships between data on nodes and edges. In a graph database, these are known as "nodes" and "relationships." Each relationship consists of a start node, an end node, and an edge pointing from the start to the end node. All nodes in the database are interconnected by various relationships. Graph databases also support traditional database functionalities such as adding, deleting, modifying, and searching data.

We structured the student-related information knowledge graph into three main components: student information nodes, employment and entrepreneurship capability nodes, and course information nodes.

1) Student information nodes: The attributes of the "Student Information" node are defined as shown in Table I. The node is named 'S' with the label 'student'.

2) Employment and entrepreneurship information node: The "employment and entrepreneurship information" node contains seven attributes. Table II shows the attribute name and description of the employment and entrepreneurship information node. The node name is T, the label is ability, and the node contains seven attributes.

3) Course information node: This paper contains seven attributes of the "Course Information" node. Table III shows the attribute name and description of the "Course information" node. This node is created after the relationship between the course line and student information is stripped.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Attribute Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num</td>
<td>student number</td>
</tr>
<tr>
<td>Name</td>
<td>name</td>
</tr>
<tr>
<td>Sex</td>
<td>gender</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Attribute Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>prof</td>
<td>profession</td>
</tr>
<tr>
<td>Nature</td>
<td>Nature of company</td>
</tr>
<tr>
<td>city</td>
<td>city</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Attribute Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>kclb</td>
<td>Course category</td>
</tr>
<tr>
<td>kcbh</td>
<td>Course number</td>
</tr>
<tr>
<td>kcmc</td>
<td>Course title</td>
</tr>
<tr>
<td>xf</td>
<td>Credit hour</td>
</tr>
<tr>
<td>dkjs</td>
<td>Substitute teacher</td>
</tr>
<tr>
<td>ksxz</td>
<td>Nature of examination</td>
</tr>
<tr>
<td>ksf</td>
<td>Examination method</td>
</tr>
</tbody>
</table>

By organizing the data in this manner, we aim to build a comprehensive and interconnected knowledge graph that facilitates a deeper understanding of the capabilities and needs of hearing-impaired students. This graph can then be used to develop personalized recommendations for skill development and career paths, ultimately enhancing their employment and entrepreneurship opportunities. Further sections of this paper will delve into the specifics of constructing this knowledge graph and the methodologies employed to leverage it for improving the career prospects of hearing-impaired students.

Building on the construction of the student information knowledge graph, we utilized student data to perform hybrid clustering on student users. Following this, we calculated the similarity between each student cluster's characteristics and the attributes defined for various employment and entrepreneurship directions. To enhance the clustering effectiveness, we employed a Canopy+Bi-Kmeans hybrid clustering model. This combination offers several advantages: it strengthens the robustness of individual clustering against noise and accelerates the similarity computation process. The flowchart of the Canopy+Bi-Kmeans algorithm is shown in Fig. 1.
Expanding on the Canopy+Bi-Kmeans hybrid clustering model, the Canopy method serves as an initial, coarse-grained clustering step, which identifies the approximate clusters or "canopies" where points are grouped based on a loose distance threshold. This step reduces the search space for the subsequent, more precise clustering method, Bi-Kmeans. The Bi-Kmeans algorithm then refines these clusters by iteratively minimizing the within-cluster variance, resulting in more accurate and well-defined clusters.

The integration of these two methods leverages the strengths of each: Canopy’s efficiency in handling large datasets and reducing computational complexity, and Bi-Kmeans’ precision in fine-tuning the cluster boundaries. This hybrid approach not only improves the clustering quality but also significantly enhances computational efficiency, making it suitable for large-scale educational datasets.

By combining these methods, the Canopy+Bi-Kmeans hybrid model efficiently narrows down the data points into manageable clusters, which are then accurately refined. This approach is particularly beneficial in educational data mining, where large and diverse datasets are common.

Fig. 1 illustrates the flowchart of the Canopy+Bi-Kmeans algorithm, detailing the steps involved in the hybrid clustering process. Through this method, we aim to provide a robust framework for identifying and analyzing student clusters, thereby facilitating personalized recommendations for enhancing their employment and entrepreneurship skills.

Hearing-impaired college students can rate projects based on their personal interests. Tags play a crucial role in helping these students understand the content and attributes of the projects more deeply. By analyzing the number of tags, we can infer user preferences. However, this often leads to an overemphasis on current trendy tags, resulting in less accurate recommendations when users opt for less popular tags [18]. Consequently, this approach fails to fully capture and reflect users’ interests and preferences.

To address this issue, we employ the Term Frequency-Inverse Document Frequency (TF-IDF) method to calculate user preferences. TF-IDF is a statistical measure used to evaluate the importance of a keyword within its dataset (as shown in Eq. (1)) [19]. This method helps balance the weight of popular and less popular tags, providing a more accurate reflection of user preferences.

By implementing the TF-IDF approach, we aim to improve the recommendation system’s accuracy, ensuring that the preferences of hearing-impaired college students are adequately represented and that they receive more personalized and relevant project suggestions. This adjustment allows for a better alignment of user interests with the recommended content, enhancing the overall user experience.
The $P_{ua}$ value is directly proportional to the degree of preference. Here, $P_{ua}$ represents the preference value of user $u$ for the project tag $a$. $n$ denotes the total number of projects, and $s$ denotes the total number of project tags. $\sum_{i=1}^{n} r_{ui} \cdot f_{ia}$ indicates the number of times user $u$ has tagged with tag $a$; $\sum_{i=1}^{n} \sum_{s=1}^{s} r_{ui} \cdot f_{ia}$ represents the total number of times user $u$ has tagged all projects; $num_{m}$ represents the total number of users. $num_{ua}$ represents the number of users who have tagged with tag $a$, $\sum_{i=1}^{n} 1 \sum_{s=1}^{s} f_{ia}$ denotes the total number of tags, and $\sum_{i=1}^{n} f_{ia}$ denotes the total number of tag $a$.

Eq. (1) demonstrates that if a user’s selected tag is infrequently chosen and comprises a smaller proportion of the entire tag set, it more accurately reflects the user’s preferences, thus enhancing recommendation efficiency. For instance, when recommending employment directions in cloud computing and big data, we primarily analyze the courses closely related to this field and the student’s grades in these courses. If a student’s performance in courses related to cloud computing and big data is significantly higher than in other courses, the likelihood of recommending cloud computing and big data as an employment or entrepreneurship direction increases.

$$P_{ua} = \frac{\sum_{i=1}^{n} r_{ui} \cdot f_{ia}}{\sum_{i=1}^{n} \sum_{s=1}^{s} r_{ui} \cdot f_{ia}} \cdot \log \left( \frac{\sum_{i=1}^{n} \sum_{s=1}^{s} f_{ia}}{\sum_{i=1}^{n} f_{ia}} \right)$$

(1)

During the recommendation process for employment and entrepreneurship directions, if the student’s interest includes "Network and Information Security," the recommendation degree for this direction will be elevated.

Traditional recommendation algorithms often use static tag identifiers for user preferences, typically represented by 0 and 1. This approach implies that the recommendation impact of these tags remains constant at all times, which is not effective for recommendations requiring temporal sensitivity. If user interest preferences are not considered dynamic over time, recommendations may not align with actual user preferences [20]. In reality, user interests are often dynamic and change over time [21]. To address this issue, at least two surveys should be conducted before making employment and entrepreneurship recommendations to capture changes in student interests. Recent user behavior is more relevant to recommendations than earlier behavior, so higher weight is assigned to recent tags to ensure timeliness and improve recommendation efficiency.

We introduce Eq. (2), which utilizes an exponential decay function incorporating a time factor to accurately reflect changes in user interest preferences over time. This approach ensures that recommendations remain relevant and aligned with the user’s current interests.

$$T_{ui} = \exp \left( - \frac{t_{now} - t_{ui}}{T_{att}} \right)$$

(2)

Among them, $T_{ui} \in (0,1)$ represents the time weight of user $u$ for project $i$. $T_{att}$ denotes the time window parameter, which signifies the duration of user preference interest. $t_{now}$ is the time of the most recent survey collection on student interests, and $t_{ui}$ is the time of the previous survey. $T_{att}$ is the time decay parameter, representing the rate of interest preference decay. $\frac{t_{now} - t_{ui}}{T_{att}}$ is rounded up in the calculation, and $T_{s} \cdot \left| \frac{t_{now} - t_{ui}}{T_{s}} \right|$ indicates the time segment in which the user’s project evaluation occurred. If a user’s interest remains unchanged over a year, with months as the statistical unit, then $T_{s} = 12$. If recommendations are made within the same academic year after the user evaluates the project, i.e., $t_{now} - t_{ui} \leq 12$, the user’s interest begins to decay after 12 months, with a decay period of 12 months, and the decay coefficient remains the same within the decay cycle.

When using the TF-IDF method to calculate user interest preferences, we integrate a time-weighted decay function (Eq. (3)) to derive user interest preferences and update the values in the user tag matrix accordingly. Finally, the normalized Euclidean distance gives the Eq. (4).

$$P_{ui} = \frac{\sum_{i=1}^{n} r_{ui} \cdot f_{ia} \cdot T_{ui} \cdot \log \left( \frac{\sum_{i=1}^{n} \sum_{s=1}^{s} f_{ia}}{\sum_{i=1}^{n} f_{ia}} \right)}{\sum_{i=1}^{n} \sum_{s=1}^{s} r_{ui} \cdot f_{ia}}$$

(3)

$$sim_{1}(u, v) = \frac{1}{1 + \frac{\sum_{i=1}^{n} (u_{i} - v_{i})^2}{\sum_{i=1}^{n} (u_{i} \cdot v_{i})}}$$

(4)

Generally, when calculating similarity, personal attributes of users, such as gender, are not typically considered. Therefore, we have incorporated user attributes and integrated these fundamental user attributes into the similarity calculation.

The similarity of gender attributes is represented by Eq. (5).

$$sim_{2}(u, v) = \begin{cases} 0, & X_{u} \neq X_{v} \\ 1, & X_{u} = X_{v} \end{cases}$$

(5)

In Eq. (6), $u$ and $v$ represent different users, $X_{u}$ and $X_{v}$ denote the genders of users $u$ and $v$ respectively. By integrating user interest preferences and attributes, we derive a comprehensive similarity score, forming a novel similarity calculation model. Here, $\lambda \in [0,1]$ serves as a weighting coefficient. The value of $sim(u, v)$ is inversely proportional to the similarity between the two users.

$$sim(u, v) = \lambda \cdot sim_{1}(u, v) + (1 - \lambda) \cdot sim_{2}(u, v)$$

(6)

Subsequently, predicting user ratings for items and making recommendations are expressed as shown in Eq. (7). Here, $\bar{r}_{u}$ represents the average rating given by user $u$ for evaluated items, $\bar{r}_{v}$ denotes the average rating given by neighboring user $v$ for evaluated items, $N_{u}$ signifies the nearest neighbors of target user $u$, $v$ denotes users in the neighbor set who have rated the item $i$, $r_{ui}$ denotes the rating given by user $v$ for item $i$, and $sim(u, v)$ represents the similarity between users $u$ and $v$.

$$P_{ui} = \bar{r}_{u} + \frac{\sum_{v \in N_{u}} \left( sim(u, v) \cdot (r_{ui} - \bar{r}_{ui}) \right)}{\sum_{v \in N_{u}} sim(u, v)}$$

(7)

IV. EXPERIMENT AND VERIFICATION

In this section, we will verify the validity of the proposed method based on the experimental data set we collected.
A. Experimental Environment

The student employment and entrepreneurship direction recommendation system based on knowledge graph is implemented on B/S architecture. The specific system development environment is depicted in Table IV. We analyze collected data on course grades and corresponding behavioral data of hearing-impaired university students to construct a knowledge graph and generate student profiles. The system then provides recommendations for employment and entrepreneurship directions based on relevance, ranking the top three directions for recommendation. Additionally, the system recommends courses related to the user's interests in employment and entrepreneurship directions to enhance relevant skills.

Fig. 2 illustrates the verification process of the knowledge graph-based employment and entrepreneurship direction recommendation system. Initially, data integration from relevant business systems associated with student users and data obtained from the internet is stored in a data warehouse using ETL (Extract-Transform-Load) tools. Subsequently, data preprocessing and feature engineering are conducted to build persistent structures of personal user profiles for hearing-impaired university students. Finally, leveraging the knowledge graph, employment and entrepreneurship direction recommendations are applied based on student user profiles.

B. Evaluation Parameter

After several rounds of training, the scoring error is reduced and the optimal parameter recommendation model is obtained. The measure we use is the mean absolute error (MAE), shown by Eq. (8).

\[
MAE = \frac{\sum_{u \in Test} |P_{ui} - r_{ui}|}{\sum_{u \in Test} |r_{ui}|}
\]  

(8)

Where \(r_{ui}\) denotes the true rating of user u for item i, and \(P_{ui}\) represents the predicted rating of user u for item i, the Mean Absolute Error (MAE) is calculated as the average absolute difference between \(P_{ui}\) and \(r_{ui}\) across the test set. A lower MAE score indicates better model performance and is evaluated using the formula:

C. Test and Evaluation

As shown in Table V, we determined the values of \(T_s, T_{att}\), and \(\lambda\) through ablation experiments. The gray shading indicates the highest scores achieved. From the ablation experiments, it is evident that when \(T_s = 5\), \(T_{att} = 60\), and \(\lambda = 0.4\), we achieve optimal performance.

Furthermore, as shown in Table VI, after setting the aforementioned parameters, we conducted ablation experiments to compare the impact of the \(k\) value in the algorithm and benchmarked it against other state-of-the-art (SOTA) algorithms. A smaller MAE value indicates better recommendation performance. It is evident from the results that our proposed algorithm achieves the highest accuracy compared to the other three algorithms tested. For instance, when the number of nearest neighbors is set to 25, the MAE reaches its minimum value across all tested algorithms. Our algorithm improves the performance by 5.25% compared to the second-ranked algorithm. The experiments demonstrate that our proposed method achieves lower scores, confirming its effectiveness.

![Fig. 2. NLL-test loss.](image-url)
outstanding performance on the datasets we collected, showcasing excellent capabilities and results, its broader application requires more comprehensive validation. To ensure the accurate and sustained improvement of employment and entrepreneurial capabilities among deaf university students, we advocate for further empirical research on feasibility, effectiveness, security, and other aspects. This endeavor will help confirm the practical potential of our approach and guide its future dissemination and application in the educational domain.

ACKNOWLEDGMENT

Fund project:2024 Jilin Provincial Department of Education Employment and Entrepreneurship and Student Management Special Project (number:JJKH20240769JY).

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES


