Text Mining of China's New Vocational Education Law to Enhance Power Grid Technical Skills Training

Guanghong Li[1,a*], Jing Wang[2,b], Chen Liang[3,c]

{liguanghong@outlook.com[a], 15262455@qq.com[b],liangchen319@yeah.net[c] }

Electric Grid Operations Training Department of State Grid Technology College, Jinan 250002,China[1,2]
Substation Operation and Maintenance Center of Taian Electric Supply Company,Taian 271000,China [3]

Abstract. In response to China's 2022 Vocational Education Law revision, this study employed text mining for thematic analysis better grasp the new principles and fresh perspectives. Preprocessing involved Python libraries for data cleaning and feature extraction. The LDA model revealed three key topics: industry-education integration, educational quality improvement, and teaching enhancement. Corresponding measures include the establishment of a dual training system, skills competition framework, and a layered teacher competency model. These initiatives align vocational education with industry needs and enhance workforce skills, ensuring a smoother transition into the modern era

Keywords: text mining; feature word selection; LDA model; industry-education integration; educational quality improvementt; teaching enhancement

1 Introduction

In 2022, China introduced the first revision to the Vocational Education Law over the past 26 years. China has established the world's largest vocational education system which has played a crucial role in supporting the country's economic and social development. The central challenge now lies in how vocational education can adapt to the requirements of the new era's economy and industrial development. This entails establishing a modern vocational education system capable of nurturing highly skilled technical workers to meet the demands of enterprises and promote high-quality employment for laborers.

The integration of industry and education is a key focus in the ongoing reforms of vocational education in the new era [1]. Concurrently, collaboration between educational institutions and businesses has become a vital pathway for talent development in vocational education. With businesses taking a prominent role in education, enhancing the cooperation between schools and enterprises for talent cultivation is a common challenge for vocational schools and companies [2]. Vocational education faces significant issues, such as a mismatch with industry talent needs and a disconnect from specific corporate job requirements. This results in a gap in school-enterprise collaboration, where schools show enthusiasm while businesses are less engaged.

To better grasp the novel ideas, new principles, and fresh perspectives introduced by the recent Vocational Education Law, a text mining approach has been employed for analysis. This method aims to identify key themes and subsequently initiate a series of reforms centered around these themes. The ultimate goal is to enhance the provision of technical skill training to various

NMDME 2023, October 13-15, Xi'an, People's Republic of China

DOI 10.4108/eai.13-10-2023.2341290

employees of the State Grid, including various technical skill training sessions, competency assessments, and competitive evaluations within the field of power distribution automation.

2 Text Preprocessing

Text preprocessing is an essential task in the fields of natural language processing and text analysis. It involves a series of cleaning and processing operations performed on raw text data before further analysis can take place, preparing the data for more effective analysis and modeling[3]. In this study, text cleaning, stopword removal processes are conducted using the Python jieba package, meanwhile tokenization relies on Python spaCy package.

2.1 Text Cleaning

Removing special characters, punctuation marks and other noisy data from the text.

2.2 Tokenization

Tokenization is splitting the text into words or tokens to divide the text into meaningful units. The jieba package [4] offers three distinct modes of segmentation: maximum accuracy, precise segmentation, and fast segmentation. To ensure precision in the analysis, this research opts for the maximum accuracy mode.

2.3 Stopword Removal

Eliminating common stopwords which typically carry little information value in text analysis tasks [5], such as "the," "is," "in," etc.

2.4 Stemming and Lemmatization

Converting words into their base forms to reduce lexical diversity and enhance consistency in the model.

3 Key Topic Revealing

The selection of text feature words is the process of identifying the words within a document that best represent the document's main content. Subsequently, clustering analysis is conducted on the feature words to determine the optimal number of topics. The use of perplexity serves as the evaluation criterion to ensure objectivity and accuracy in the results.

3.1 Text Feature Word Selection

TF-IDF method is commonly employed for selecting document feature words. TF denotes term frequency, signifying the frequency of a word's occurrence within the current document. On the other hand, IDF represents inverse document frequency, serving as a measure of how frequently a word appears across all documents. The specific computation formula is as follows:

$$TF - IDF(w) = TF(w) \times IDF(w) \tag{1}$$

$$TF(w) = \frac{N_w}{N_{total}} \tag{2}$$

N_w is the total number of word w in the document D_j, N_{total} is the total word count in the document D_j.

$$\mathrm{ID}F(w) = \log_2(\frac{M}{ND_w + 1}) \tag{3}$$

M is the total number of documents in the document set D, ND_w is the total number of documents involve word w.

3.2 Determine optimal number of topics

A lower perplexity indicates a better model differentiation, meaning that the topic selection is more accurate.

$$perplexity(D \mid M) = \exp\left|\frac{\sum_{i=1}^{M} \log(p(d_i))}{\sum_{i=1}^{M} N_i}\right| \tag{4}$$

M is the total number of documents in the document set D , N_i is the total number of words in the text D_i , $p(d_i)$ represents the probability of the LDA model generating text d_i.

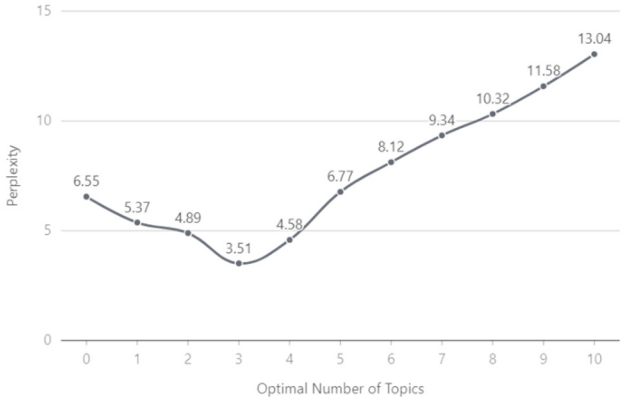

Fig. 1. Optimal number of topics.

The perplexity results are calculated as below in the **Table 1**:

Table 1. Perplexity Results

Optimal Number of Topics	Perplexity
0	6.55
1	5.37
2	4.89
3	3.51
4	4.58

5	6.77
6	8.12
7	9.34
8	10.32
9	11.58
10	13.04

As the results from **Table 1** and **Fig 1**, when the perplexity is at its lowest, the optimal number of topics is determined to be 3. Subsequently, we will proceed with the LDA model analysis.

3.3 LDA Model Generation

For modeling in this study, we employed the Gensim library in Python. Gensim is primarily designed for handling unstructured text and utilizes unsupervised learning to express topic vectors within the latent layer of text. It is a versatile and widely-used natural language processing tool that finds extensive application in the modeling process of LDA models [6].

Following the extraction of three topics as outlined in step B, the computed results are presented in the **Table 2** as below:

Table 2. LDA Model Results

Topic 1	Topic 2	Topic 3
Enterprise	Educational Quality Evaluation	Major
Industry-Education Integration	On-the-Job Internship	Curriculum
School-Enterprise Cooperation	Credit Courses	Talent Development Plan
Engineering Education Integration	Integrated Training	Online ourses
Integration of Moral and Technical	Education Training Programs	Internship Training

Topic 1 centers about nurturing individuals, with terms such as "enterprise," "industry-education integration," and "school-enterprise cooperation" pointing to the emphasized role of enterprises in education, as clarified by the New Vocational Education Law. It underscores the need to strengthen collaboration between educational institutions and businesses, enhancing teacher quality through approaches like "engineering education integration" and the integration of moral and technical education.

Topic 2 places emphasis on the educational model, highlighting the activation of work positions and school courses through integrated training to optimize the evaluation of educational quality.

Topic 3 revolves around the nurturing platform, focusing on the improvement of teacher professional competence and subject expertise, ultimately enhancing their teaching capabilities.

4 Corresponding Practices and Measures

After establishing our development strategy and themes, we will implement corresponding measures to facilitate training and support the State Grid Corporation's endeavors in technical skill development.

4.1 Constructing a Dual Training System for Schools and Enterprises

Aiming at topic1, collaborative training between educational institutions and businesses is proposed, which plays a pivotal role in enhancing the capabilities of "dual-teacher" educators. In terms of platform development, our team established a specialized training and evaluation center for distribution automation, which is prominently tailored to the industry. Additionally, we have engaged in a joint initiative with Jinan University, fostering a mutually beneficial flow of teachers and industry collaboration.

To strengthen the development of our system, we have implemented a dedicated plan to enhance the qualifications of trainers, aligning closely with the advancements in distribution automation technology. Furthermore, we rigorously enforced comprehensive training and practical experiences for all members within the professional group. Our training system encompasses online and offline components, including the national training program, provincial training initiatives, school-based training and industry internships, ensuring a tailored approach to training that meets specific needs.

4.2 Establishing a Skills Competition System

To support topic 2, the concept of "Work-Course-Competition-Certification" is introduced, which involves aligning work positions (Work), curriculum systems (Course), professional skills competitions (Competition), and vocational skill certifications (Certification). Our team has harnessed this concept to design specialized courses based on job skill standards, fostering the development of highly skilled individuals. Additionally, through professional skills competitions, we provide a platform for participants to showcase their high-end skills, thereby leading educational reform and promoting the convergence of "Work-Course-Competition-Certification" to cultivate highly skilled individuals.

1) "Work" involves the formulation of curriculum standards and clear teaching directions. These curriculum standards must align with job requirements, correspond to professional standards, and reflect work processes. By using real company work tasks as a foundation and incorporating enterprise-certified content, along with actual case studies, we aim to cultivate students' analytical and problem-solving abilities.

Our team has upgraded and revamped both offline training facilities and online course content to stay current with new knowledge, technologies, processes, and methods in the industry.

2) "Course" serves as the core and foundation of educational reform. To drive a "classroom revolution" and accommodate a diverse student body, we have enhanced the learner-centered professional and curriculum teaching evaluation system.

We have developed modular training and teaching content tailored to different levels of training recipients, ranging from vocational college students, newly hired employees of the State Grid Corporation of China (SGCC), various short-term training course attendees, to market-oriented and internationally-trained participants. We base these modules on the actual needs of learners and students, develop training and teaching resources, and conduct training and teaching activities.

3) "Competition" represents the high-end demonstration and benchmarking of course teaching. We uphold the principle of "Promoting Teaching through Competition, Promoting

Research through Competition, Promoting Construction through Competition, and Promoting Improvement through Competition." We have established a four-tier teacher teaching ability competition system, encompassing departments, schools, provinces, and the nation, to elevate course teaching standards. We also adhere to "Promoting Training through Competition" by integrating competition projects into the "dual-teacher" teacher training program. The competition content is transformed into teacher training content, and the standards migrate to teacher evaluation standards. This approach promotes a transformation in teachers' ideologies, concepts, and thinking patterns, enhancing their practical skills and information technology teaching abilities.

Our group won the first prize in the 2022 National Vocational College Skills Competition in teaching abilities.

4) "Certification" represents the assessment of course learning and industry validation. By developing and integrating various vocational skill assessment certificates, qualifications, and level certificates, we incorporate comprehensive abilities required for vocational activities and personal career development into certificates, enhancing students' employability and entrepreneurship skills.

To accomplish this, our team has introduced vocational qualifications and the "1+X" vocational certification system into the curriculum [7]. This enriches the course content, improves teaching methods, enhances the enjoyment of learning, and elevates teaching quality.

4.3 Constructing a Layered Capability Model

Teaching entails the transmission of knowledge, provision of guidance, and resolution of inquiries. Teacher development represents an evolving journey that encompasses the continuous accumulation, cultivation, and enhancement of educators' knowledge, skills, and personal attributes throughout their careers.

Our team is committed to nurturing "dual-qualified" instructors, ensuring that skills development and certification align with educators' career stages and the corresponding competency requirements. We introduce innovation by implementing a blended teaching model that seamlessly combines online and offline components. Additionally, we advocate for the creation of high-quality, structured teaching innovation teams, with a focus on enhancing teachers' competencies in ethics, pedagogy, comprehensive education, and self-directed growth.

To achieve these objectives, we have developed a competency requirement chart for "dual-qualified" educators [8], formulated and refined tiered standards, and conducted internal training initiatives. Our vigilant monitoring of cutting-edge applications in power distribution automation, coupled with the strict enforcement of comprehensive rotational training within our professional cohorts, is crucial. Furthermore, we actively participate in corporate internships.

5 Conclusions

The identification of key themes and the formulation of corresponding measures represent pivotal aspects of this study's endeavor. By employing a text mining approach, we discerned three principal topics: industry-education integration, the enhancement of educational quality, and teacher development. In response, we have devised a comprehensive set of measures. These

measures encompass the creation of a dual training system, the establishment of a skills competition framework, and the development of a layered teacher competency model. These strategic initiatives are meticulously designed to align vocational education with contemporary industry demands, facilitating the cultivation of a highly skilled workforce primed for success in the modern era.

References

[1] Zeng Tianshan, Li Jiehao, " The New Vocational Education Law Ensures the High-Quality Development of Vocational Education" China Vocational Education, vol. 2022, no. 16, pp. 16-22, 2022.

[2] Y. Hu and K. Li, "Top-level Optimization and Value Orientation for the High-quality Development of Vocational Education - Based on the Examination of the New Vocational Education Law," Education and Vocation, vol. 2022, no. 17, pp. 35-39, 2022.

[3] S. M. H. Dadgar, M. S. Araghi and M. M. Farahani, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," 2016 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, India, 2016, pp. 112-116

[4] S. Niklaus, N. C. D. Fonseca, S. Handschuh, "A Survey on the Application of SpaCy in Medical and Clinical Text Analysis", 2021 IEEE Seventh International Conference on Healthcare Informatics (ICHI), 2021, pp. 1-5.

[5] Y. Qin, R. Chen, M. Chen, S. Yu and X. Wu, "DFCN: Deep Fully Convolutional Network for Speech Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1333-1345, 2020

[6] Y. Bai, L. Guo, Q. Yu, B. Gao, "An Efficient Randomized Algorithm for Topic Modeling in Big Data", IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 9, pp. 2451-2464, 1 Sept. 2016.

[7] Li, R., Luo, J. and Li, H., 2021. Research on the Construction of 1+X Certificate System under the Background of New Engineering. Journal of Higher Education in Heilongjiang, 39(07), pp.126-129.

[8] Y. Cheng and W. Wu, "Dual-Qualified Teachers: Challenges and Opportunities," in IEEE Transactions on Education, vol. 63, no. 4, pp. 317-326, Nov. 2020.