International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7556-7565

https://doi.org/10.70517/ijhsa463648

Research on Automatic Generation of Civics Teaching Content for Engineering Management Course Based on Natural Language Generation Technology

Qian Li[1,*]
[1] Department of Engineering Management, Chengdu Jincheng College, Chengdu, Sichuan, 611731, China
Corresponding authors: (e-mail: JC20240201@163.com).

Abstract Currently, the teaching of Civics in engineering management courses faces the problems of difficulty in selecting teaching materials and lagging behind in content updating, and teachers have invested a lot of time and energy in integrating professional knowledge with Civics elements. The rapid development of natural language processing technology provides technical support for solving these problems and realizes the efficient automated generation of Civics teaching content for engineering management courses by automatically filtering and matching relevant Civics elements through algorithms. In this study, we first constructed the database and knowledge graph of Civics in engineering management courses, and designed the automatic screening algorithm for Civics elements. Then the keywords are extracted using the TF-IDF algorithm, and the BERT and GPT-2 models are used to generate the Civics text content. Finally, the generation effect is evaluated by content quality score, keyword coverage and student score improvement rate. The results show that the percentage of content with 0.4 to 1.0 quality scores reaches 74.4%, and the students' scores are improved by 7.2% compared with the traditional textbook after teaching Civics content based on natural language processing technology. The Jaccard similarity coefficient test shows that the average value of corpus overlap rate under the same topic is 28.71%, and the overlap between different topics is 29.31%. The results of the student feedback experiment show that the difficulty coefficient of the test paper is 0.64 falling in the moderate category, and the total average score of the content difficulty is 4.175.This study proves that natural language generation technology has a good prospect of application in the automated generation of the content of the teaching of Civics and Politics in engineering management courses, and it can effectively improve the quality of teaching and students' learning effect.

Index Terms natural language generation, engineering management, curriculum civics, knowledge graph, BERT model, TF-IDF algorithm

I. Introduction

"Ideological and political education" refers to a comprehensive educational concept that combines various courses and ideological and political theory courses in the same direction in the form of building a full-staff, full-course, and full-curriculum education pattern to form a synergistic effect, and takes "cultivating people with virtue" as the fundamental task of education, which can not only alleviate the embarrassment of one-way education in ideological and political courses, but also highlight the mainstream value orientation, and realize the combination of knowledge transfer and value guidance [1]-[4]. Traditional engineering management education is based on the application concept of serving the society, and is committed to cultivating students' innovative consciousness and practical ability, which is characterized by four features: practicality, comprehensiveness, ethicality and globalization [5]-[7]. With the implementation of the “One Belt, One Road” initiative and the rise of “New Engineering”, engineering management talents need not only to continue to improve their professionalism, but also to steadily improve their political awareness [8], [9]. Therefore, the course ideology of engineering management teaching aims to seek the internal unity of engineering management education and course ideology reform [10]. However, the traditional process of engineering management course production requires teachers to do a lot of editing work, which is inefficient and difficult to flexibly respond to the needs of knowledge updating [11], [12]. The automated generation of teaching content for engineering management courses based on natural language generation (NLG) technology has become an important way to solve this problem [13], [14].

NLG technology is an important research direction in the field of artificial intelligence, aiming to use computer algorithms and models to generate texts that conform to human language habits and norms [15], [16]. With the development of deep learning habits and big data, NLG technology has made significant progress in the fields of machine translation, chatbots, and intelligent document generation [17]-[19]. In teaching content generation, the personalized supply of automated Civics teaching resources can be achieved through the use of NLG technology, and this automated content generation is conducive to improving teaching efficiency [20], [21]. However, teachers cannot completely rely on the content generation of NLG technology, and only by combining the traditional content generation method with automated generation and realizing the complementarity can it better serve the teaching of Civics and Politics in engineering management courses [22]-[24].

The construction of curriculum ideology and politics in colleges and universities is an important measure for the implementation of the fundamental task of establishing moral education, which is of great significance to the cultivation of students' correct values and family and national sentiments. As an important field of science and technology, the construction of the Civics and Politics of Engineering Management courses should not only teach professional knowledge, but also incorporate the ideological connotations of patriotism education, scientific and technological service to the country, and artisan spirit. At present, there are three dilemmas in the teaching of Civics in engineering management courses: firstly, the Civics elements in the existing teaching materials are insufficient, and the professional teachers face difficulties in selecting suitable Civics cases; secondly, the Civics content is not sufficiently integrated with the professional knowledge, and it lacks systematicity and relevance; thirdly, the Civics elements have a long updating cycle, and it is difficult to reflect the needs of the times in a timely manner. Teachers need to invest a lot of time in collecting, screening and integrating civic and political materials, which is a heavy workload and inefficient. These problems constrain the improvement of the quality of Civics teaching in engineering management courses, and affect the realization of the goal of cultivating moral character.

Natural language processing technology has been developing rapidly in recent years, and has achieved remarkable results in text generation, information extraction, knowledge map construction and other aspects. The application of this technology to the automated generation of the teaching content of the course's civic politics can effectively solve the current dilemma faced by the construction of the civic politics of the engineering management course. By establishing the correlation model between professional courses and Civics elements, and realizing the intelligent screening of Civics materials and automatic generation of content, it can not only greatly improve the efficiency of teachers' lesson preparation, but also ensure the quality and relevance of the generated content, and provide technical support for the construction of Civics in courses.

Based on natural language processing technology, this study explores the automated generation method of Civics teaching content for engineering management courses. The study is divided into three stages: first, constructing the database and knowledge map of engineering management course Civics to lay the data foundation for automated generation; second, designing the automatic screening algorithm of Civics elemental materials, realizing the accurate screening of Civics materials through text pre-processing, key knowledge point matching degree calculation and the application of the BERT model; and lastly, generating the Civics textual content based on the natural language processing technology, which includes keyword extraction, text content generation and classification. The generation effect is evaluated by indicators such as content quality score, keyword coverage rate and students' score improvement rate, which verifies the application value of natural language generation technology in the automated generation of Civics teaching content for engineering management courses, and provides new ideas and methods for the construction of Civics in higher education courses.

II. NLP-based Generation of Civics Teaching Content for Engineering Management Courses

II. A. Construction of Civic Database for Engineering Management Course

II. A. 1) Civics database for engineering management courses

According to the Guidelines for the Construction of Civics and Politics in Higher Education Courses, the organization of the content of course civics and politics should be centered on ideals and beliefs, with love for the party, patriotism, socialism, the people and the collective as the main line, and systematic education around the focus of family and national sentiment and cultural literacy.

This paper takes the engineering management course as the research object, combines the characteristics of teacher education, professional characteristics and the background of the times, and combs the framework of the Civics and Politics of the engineering management course and the knowledge points to help establish a database. The featured civic and political content includes the civic and political elements of science and technology, education, craftsmanship, professional competence and engineering ethics, and the civic and political framework of the engineering management course is constructed.

Civic and political examples are selected from sources that mainly consider famous quotes, celebrity stories or national events, with traceable sources, authentic data, and themes that are in line with the characteristics of college students. For each leaf node on the Civics Framework of Engineering Management Program, 45 related Civics instances are manually completed, i.e., the first batch of database data is obtained, including Civics instances and their corresponding types, with a total of 50 types of data.

II. A. 2) Engineering Management Program Knowledge Mapping

After constructing the database and importing the knowledge points in Building Construction Technology, it is then necessary to construct the knowledge points and Civics types, as well as the correlation relationship between each knowledge point.

For this purpose, knowledge mapping technology is introduced to correlate the knowledge of different subject areas and help teachers better combine the elements of Civics with the knowledge of each subject area.

Logically, knowledge mapping is divided into a data layer and a schema layer. The data layer expresses facts, such as entities, attributes and attribute values, in ternary groups, which can be stored using databases such as Neo4j graphs. The schema layer standardizes the facts in the data layer through ontology libraries, provides conceptual templates for knowledge storage, enhances the organization of the knowledge structure, and reduces data redundancy.

Constructing a knowledge graph can be accomplished through knowledge extraction and knowledge fusion of raw data. Knowledge extraction generally includes three parts: entity extraction, relationship extraction and attribute extraction, which can be accomplished using ternary groups (entity 1, relationship and entity 2).

II. A. 3) Primitive Civics Element Acquisition

Focused web crawler, also known as topic web crawler. The technology starts crawling from the initial link, and then based on the analysis algorithm to evaluate and retain links related to the topic, put into the queue of links waiting to be crawled, marking to prevent repeated crawling, and then through a certain search strategy to obtain the link again until it meets certain conditions In order to further improve the accuracy and efficiency of the crawling, clustering algorithms are introduced into the process of knowledge crawling. The main links include:

(1) For the links to be crawled, a score F1 is given based on the analysis algorithm.

(2) Clustering a certain number of crawled links and calculating the similarity F2 between the links to be crawled and those already crawled.

(3) Sort the links and crawl the links with higher scores.

II. B. Design of automatic filtering algorithm for Civics element materials

II. B. 1) Text pre-processing process

Text preprocessing is an important step in improving the accuracy of subsequent analysis. The preprocessing process in this study includes the following key steps:

(1) Noise removal: regular expressions are used to effectively filter out the non-essential elements in the text (e.g., HTM tags, special symbols, etc.) to ensure that the text is clean and lay the foundation for subsequent processing.

(2) Chinese Lexical Segmentation: The jieba tool is used for efficient lexical segmentation, adapting to the characteristics of Chinese without space separation, capable of dealing with complex vocabulary structures, such as idioms, slang, etc., and supporting lexical annotation and keyword extraction functions, which enhances the flexibility of analysis.

(3) Deactivation filtering: After word segmentation, the deactivation list is customized to remove common non-substantive words, such as “the”, “had”, etc. This not only streamlines the data volume, but also improves the focus and accuracy of text analysis. The list of deactivated words is integrated with the existing Chinese deactivation lexicon and customized according to the content of the study.

II. B. 2) Calculation of the matching degree of professional key knowledge points

In order to accurately detect whether key concepts of mobile application development are covered in the text, an exhaustive list of keywords mobile_app_dev_keywords was created, which contains terms such as various key technologies, tools, frameworks, and best practices for mobile application development.

In terms of specific implementation, for the target text, it is also preprocessed first, including the removal of irrelevant symbols, word splitting, and the removal of words in the customized deactivation word list, and then the preprocessed set of text words is intersected with the set of keywords for mobile application development, the number of keywords shared between the two is calculated, and the intersection of the set of text keywords and mobile_app_dev_keywords is calculated. keywords and the intersection of the text keyword collection with mobile_app_dev_keywords, and the result obtained reflects the degree of knowledge related to mobile application development involved in the text. By counting the size of the intersection, the relevance and coverage of the text content to the knowledge points in the field of mobile application development can be quantitatively assessed. This approach helps to filter out materials such as literature, tutorials, or case studies that are closely related to mobile application development from a large amount of text resources.

II. B. 3) Application of BERT model in text embedding vector computation

The BERT model [25] performs well in areas such as named entity recognition and sentiment analysis, and can effectively capture the deep semantic information of text. Mathematically, the embedded vector of BERT model can be expressed by the following equation:

$$Embedding(v) = BERT(tokenize(v)) \tag{1}$$

where v is the input text, $tokenize(v)$ is the text after disambiguation, and $BERT$ is the $BERT$ model function.

In this study, the BERT model is used to obtain the embedding vectors of the text, which can effectively capture the deep semantic information of the text, and the main steps are to segment and encode the preprocessed text, input the encoded text into the pre-trained BERT model, and obtain the output embedding vectors.

This study combines the advantages of the two methods, on the one hand, using TF-IDF and keyword matching to assess the similarity of the text on surface features, on the other hand, using the BERT model and cosine similarity to measure the fit of the text at the semantic level, so as to realize the all-around consideration and filtering of the relevance of the course content and the news material.

II. B. 4) Screening mechanism and scoring system design

Next, the system utilizes the pre-constructed Civic and Political Element Keyword Database to count the number of occurrences of keywords related to Civic and Political Education in the news materials, so as to derive the “Civic and Political Element Score”. By calculating the intersection size of keywords between the text of news materials and the Civic and Political Keywords Database, the system evaluates the value of the news in spreading the concept of Civic and Political Education.

In order to further improve the evaluation accuracy, the system combines the deep semantic similarity between the news material and the course content calculated by the BERT model - “BERT similarity” - to capture more complex semantic associations.

Ultimately, the comprehensive scoring system integrates the scores of each aspect through the following modified formula:

$$\begin{aligned} CombinedRelevance(C,N) = w_1 \times KeywordOverlap(C,N) \\ +w_2 \times CosineSimilarity(C,N) \end{aligned} \tag{2}$$

$$\begin{aligned} MatchScore(C,N) = w_3 \times CombinedRelevance(C,N) \\ +w_4 \times BERTSimilarity(C,N) \end{aligned} \tag{3}$$

where C denotes course content, N denotes news material, $KeywordOverlap(C,N)$ is the value of the keyword intersection score, and $CosineSimilarity(C,N)$ is the cosine similarity computed based on the $TF-IDF$ vector. And $BERTSimilarity(C,N)$ is the deep semantic similarity obtained by $BERT$ model. The weighting coefficients w_1, w_2, w_3 and w_4 can be adjusted according to the actual application scenarios and needs to balance the weight of each score item in the final matching score. Such a comprehensive calculation ensures a full-blooded and objective assessment of the content relevance of news materials and mobile application development courses as well as the value of Civic and Political education. In this paper, from the consideration of the importance of each factor, set $w_1 = 0.3, w_2 = 0.7, w_3 = 0.7, w_4 = 0.3$.

II. C. Civics content generation based on natural language processing technology

II. C. 1) Keyword extraction

Keyword extraction is divided into two methods: keyword assignment and keyword extraction. The former is selected based on the existing thesaurus, and the latter is selected through algorithmic analysis. In this paper, we focus on exploring the latter, and specifically practiced the algorithm based on TF-IDF [26], the graph-based sorting algorithm TextRank, and the keyword extraction algorithm based on the LDA topic model. The following is an example of the TF-IDF-based algorithm, which discusses how to extract the keywords of the examples of Civics elements.

The TF-IDF-based algorithm evaluates the category differentiation ability of words by analyzing the word frequency (TF) and inverse document frequency of the words, which is calculated as:

$$\text{Word frequency (TF)= the number of times a certain word appears in an article} \tag{4}$$

$$\text{Word frequency (TF)} = \frac{\text{The number of times a word appears in the article}}{\text{The total number of words in the article}} \tag{5}$$

$$\text{Inverse Document Frequency (IDF)} = \log\left(\frac{\text{Total number of document corpus}}{\text{Number of documents containing the word} + 1}\right) \tag{6}$$

If more articles contain a word, the smaller the IDF is; to avoid the denominator in equation (6) being zero, the number of documents containing the word is added to one.

The formula for TF-IDF is:

$$\text{TF-IDF= Word Frequency (TF)- Inverse Document Frequency (IDF)} \tag{7}$$

It can be seen that TF-IDF is directly proportional to the number of occurrences of a word in a text and inversely proportional to the number of occurrences of the word in all texts.The higher the TF and the lower the IDF, the better the category differentiation ability and the more suitable for keywords.

The keyword extraction algorithm based on TF-IDF has wide applicability and can be used for keyword extraction tasks in most texts.

II. C. 2) Content Generation of Civics Texts

In this paper, we adopt the summary generation technique and practically compare the more popular GPT-2 algorithm and Bert algorithm.

(1) Algorithm based on GPT-2: The pre-training model released by OpenAI organization in 2018 based on the GPT model, which is essentially an autoregressive model, consists of a multi-layer unidirectional Transformer's decoder part.The task of the GPT model is to predict what words (tokens) will appear next. For example, if a sentence is obtained in the training text, the GPT model will first give it a “BOS” (Begin of Sentence) token, then output a word vector, and then predict what the next token should be based on this word vector.

(2) Bert-based algorithm: a Transformer-based language model that uses the Encoder module to encode using context. The model pre-training utilizes a large amount of linguistic data to learn linguistic syntax as well as semantic information, which can encode the text and generate the corresponding representation vectors.

In the process of generating the content of the Civics text, we use the representation vectors encoded by Bert, and we have to weight the representation vectors of each word, select the key words, etc., calculate the importance of each word, and select the key information from them to generate the text.

II. C. 3) Civics Text Content Classifier

The text categorization task is to assign one or more labels to the processed material that are set in advance. We focus on Bert-based text categorization and EasyDL-based text categorization methods.

Bert-based text classification: firstly, the semantics of the text is encoded by a pre-trained language model, and then the encoded semantic vectors are used as features input to a classifier for text classification.

EasyDL-based text classification: based on the model of text classification-multilabel in this platform, the classification labels are customized to achieve automatic classification of text content, and each text can belong to multiple classification labels at the same time.

III. The effect of automated generation of the content of the Civics teaching in engineering management courses

III. A. Selection of assessment indicators

For the assessment of teaching effectiveness, the following assessment metrics can be selected. First, the content quality score is assessed using the open source tool Natural Language Data Package, which evaluates the generation quality score of a single sentence or multiple sentences (e.g., entire documents and paragraphs) through the functions that come with the tool. Second, keyword coverage is used using the TF-IDF method. Firstly, the frequency of occurrence of each word in the content is calculated, after that the inverse frequency of the word in the content set is calculated, and finally the two are multiplied to get the weight of the word, and the word with higher weight is the keyword, and the algorithm tends to filter out the common words and keep the important words. The keyword coverage can be derived from the occurrence of keywords in the content. Third, academic performance is improved. The effectiveness of the generated content for knowledge transfer is assessed by testing students' mastery of key concepts and knowledge. Compare the changes in the performance of students using the natural language processing technology-based Civics based on natural language processing technology-based Civics content with those using traditional textbooks in the exams to assess the teaching effectiveness of the generated content. Fourth, the rate of student score improvement. It is necessary to follow the same environment to compare the changes in the scores of students using Civics content based on natural language processing technology with those using traditional textbooks in the exam to assess the teaching effectiveness of the generated content. Comparison is made by comparing the scores of two different classes of the same grade in the same school using

Civics content based on natural language processing technology and traditional textbooks respectively. For learning achievement improvement, the formula for the student score improvement rate S can be expressed as:

$$S = \frac{G-T}{T} \times 100\% \tag{8}$$

where G denotes the average score of students using the content of the engineering management course; and T denotes the average score of students using the traditional textbook.

III. B. Assessment of Content Generation for Civics Teaching in Engineering Management Courses

III. B. 1) Content quality score

In order to assess the quality of the generated content, the distribution table of the quality score of the generated content using BERT is shown in Figure 1. As can be seen from the figure, a high percentage of content with high quality scores is extracted, focusing on the percentage of content with quality scores from 0.4 to 1.0 reaching 74.4%, to achieve efficient generation of content.

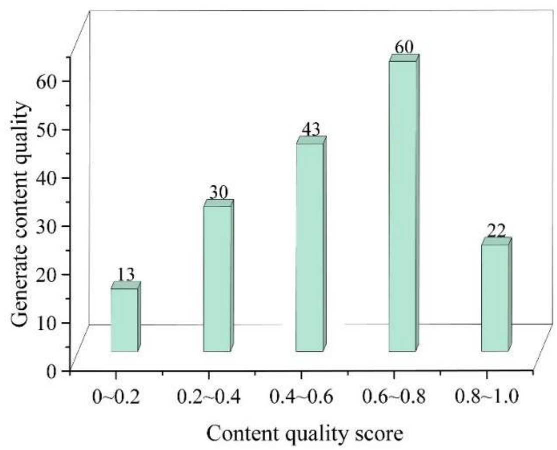

Figure 1: Generated content

III. B. 2) Changes in coverage

The coverage of the generated content is measured using keyword coverage, and the change in coverage over time is shown in Figure 2. In the figure, 1 to 8 represent the learning progress of Civics based on natural language processing technology, and a larger number represents more learning progress of Civics based on natural language processing technology. As the course unfolds, the keyword coverage of the generated content is getting bigger and bigger, and it can be concluded that after the steps of data collection and preprocessing, key information extraction, content generation content embellishment and so on the content of BERT technology can realize fast and efficient generation, which provides a new solution for the application of Civics content generation and automation based on natural language processing technology.

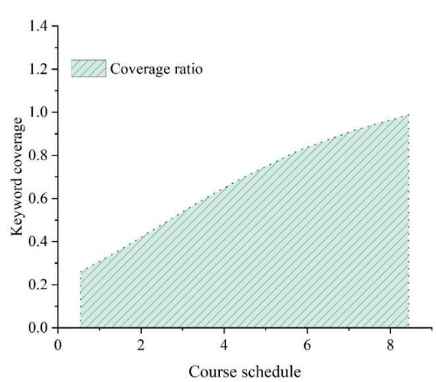

Figure 2: coverage area

III. B. 3) Rate of improvement in student scores

The average scores of the final exams of students in different classes of the same grade and the same major were derived by comparing the average scores of the grades taught using the content of the Civics course with the average scores of the grades taught using the traditional textbooks. The score improvement rate is shown in Table 1. From the table, it can be seen that the students' scores are improved by 7.2% after using the Civics content teaching based on natural language processing technology compared with the scores after using the traditional textbook teaching, and there is an improvement in the scores. This shows that the generation of Civics content based on natural language processing technology is effective and has a high degree of matching.

Table 1: Score improvement rate

Courses	The average content of the microcourses	The average score of traditional textbooks	Score improvement rate /%
Project management	85.3	81.2	5.04
Engineering contract management	86.2	83.7	2.99
Project cost management	94.3	85.6	10.16
Engineering economics	85.4	79.4	7.56
Engineering valuation	89.7	78.5	14.27
Managerial economics	90.5	87.4	3.55
Organizational behavior	95.4	88.9	7.31
Financial management	626.8	584.7	7.20

III. C. Validation and Discussion

III. C. 1) Data overlap test

Jaccard similarity coefficient is a metric used to compare the similarity of two sets, by partitioning the text into a collection of words, and then calculating the size of the intersection and concatenation to derive the similarity between the two, which is widely used in the field of data mining, information retrieval, natural language processing and other fields [27]. The same application of python to carry out the Jekyll and Hyde similarity calculation, and then visualize and analyze the percentage, it can present the overlap between the corpus more intuitively.

First, the overlap of the corpus under the same topic is compared. Taking the four pieces of corpus under the theme of engineering economy as the sampling object for the calculation as shown in Figure 3, the average value of the two-two overlap rate is 28.71%, which is slightly higher than 25%. The dialog corpus of the same topic allows a certain degree of overlap due to the content association.

Secondly, the overlap between the corpus of different topics is compared. Taking the eight corpus under Engineering Economy Theme 1 and Engineering Management Theme 2 as the object of study, they are organized into two documents respectively by theme, and the calculation results in a content overlap rate of 29.31%. The overlap between texts of different themes is higher than that under the same theme, which shows that ChatGPT does not make a clear distinction between the contents regarding the differences in themes.

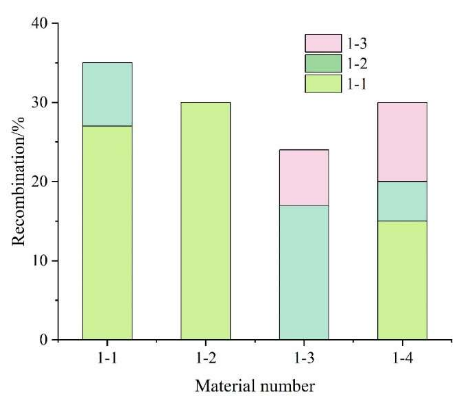

Figure 3: The recombination of the four dialogues in the project economic subject

Again, a two-by-two overlap comparison of the 60 corpus and output as an area plot as shown in Fig. 4 can more clearly observe the overlap trend among the generated texts. There are four groups with over 80% overlap, the highest group with 99.19% overlap, and three groups with 60% overlap, while most of the other corpus has a floating overlap of 20%, with the lowest overlap of 1.5%. The low differentiation between the dialogues under each theme and the high repetition rate make it unsuitable to be presented directly as teaching content.

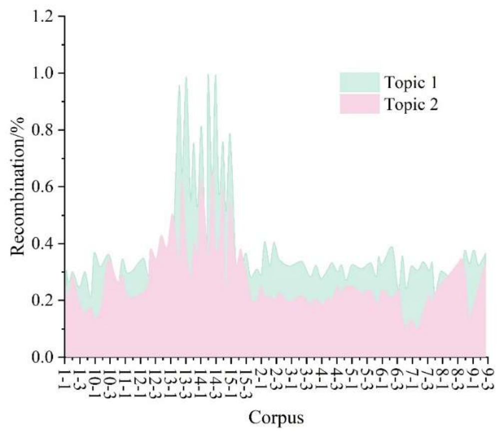

Figure 4: The two sides of the whole language are compared

III. C. 2) Student feedback

To verify the match between the above findings and the actual learning situation, an experiment based on students' feedback was set up. Two versions of engineering project management test papers were designed based on the results of formulaic measurements, the A version of the test paper included the corpus with the lowest (15-2 Project management) and highest (1-3Engineering construction) scores in the Flesch ease-of-reading analysis, and the B version included the corpus with bipolar scores (7-1Organizational behavior, 11-1Economic construction) in the McApline EFLAW and the Dale-Chall analyses of the corpus with the polar scores (7-1Organizational behavior, 11-1Economic management), mirroring and overlapping each other for validation. Students were asked to rate the difficulty of each conversation based on a Likert scale. The experimental subjects were second-year engineering management students in a higher vocational college, and the A and B versions of the test papers were randomly distributed in equal numbers, with a total of 150 test papers distributed and 131 recovered, with a recovery ratio of 87.33%.

The analysis was carried out using the descriptive statistics function in Excel, resulting in a combined mean score of 65.78 (out of 100) for the A and B versions of the test paper, with a difficulty coefficient of 0.64 falling in the moderate category, close to the norm score (60). The median is 63.5, the standard deviation is 21.0781, and the variance is 405.327, indicating a high degree of dispersion in the score data, and a wide variation in the difficulty of the topics in the same test paper, which is in line with the topic design. The kurtosis value of -0.7152 is relatively flat, and the skewness value is -0.1442, the data distribution is relatively symmetrical, which is consistent with the characteristics of normal distribution as shown in Figure 5.

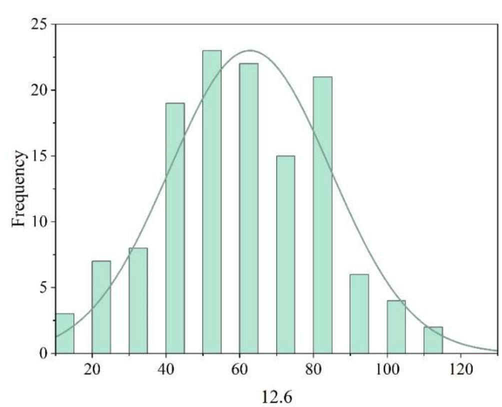

Figure 5: The student feedback experiment results in the normal distribution

According to the results of the students' difficulty scores on the four articles using the Likert scale (1 very easy-2 easy-3 moderate-4 difficult-5 very difficult), Figure 6 shows that the overall average score is 4.175, which is between "moderate" and "difficult".The mean values of the four corpora did not differ much, with a maximum of 3.67 and a minimum of 3.34. All the corpora conformed to a normal distribution except for the 11-1Economic management corpus, which had a peak of 1.37, with a slightly sharper peak in the data distribution, and a concentration of extremes, which indicated that students' judgments of the difficulty of this corpus were more polarized.

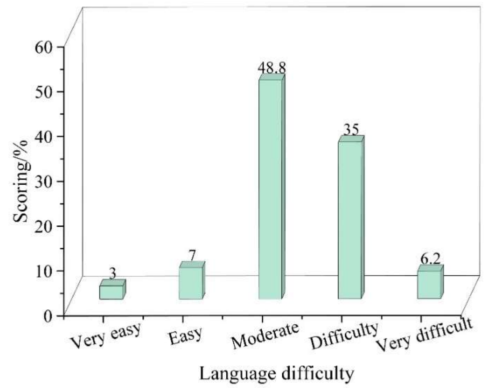

Figure 6: Students score results for difficulty in corpus

IV. Conclusion

The automated generation system of Civics teaching content for engineering management courses based on natural language generation technology shows good application results. The content quality assessment found that 74.4% of the content generated using BERT had quality scores in the range of 0.4 to 1.0, which proved that the system was capable of generating higher-quality Civics content. The analysis of keyword coverage shows that after the steps of data collection and preprocessing, key information extraction, content generation and embellishment, the content generated by the BERT technology achieves a wider knowledge coverage, providing a new solution for the automated application of course Civics content.

The comparison of students' score improvement rate shows that after using natural language processing technology-based Civics content teaching, students' scores are improved by 7.2% compared with using traditional teaching materials, with the highest improvement rate of 14.27% in the engineering valuation course. The Jaccard similarity coefficient test shows that the average value of corpus overlap rate under the same topic is 28.71%, and the overlap between corpus of different topics is 29.31%, and the overlap of part of the corpus is as high as 99.19%, which indicates that there is a certain degree of duplication of the generated content, which needs to be further optimized.

The results of the student feedback experiment show that the difficulty coefficient of the test paper is 0.64, which is in the moderate category. The total average score of the students' rating of the difficulty of the generated content is 4.175, which is between “moderate” and “difficult”, and there is not much difference in the difficulty ratings of different corpora. These data show that the Civics content based on natural language generation technology is in line with the cognitive level of engineering management students and can effectively support teaching activities. Taken together, natural language generation technology has significant advantages and broad prospects in the automated generation of Civics teaching content for engineering management courses.

References

[1] Yang, Y., Yu, W., & He, B. (2025). Teaching Design and Practice of Ideological and Political Education in Cartography Course. International Journal of Social Science and Education Research, 8(4), 210-217.

[2] Zheng, W., & Lu, Q. (2025). Resource Allocation and Optimal Path Selection of Civic and Political Education Curriculum Based on Multi-Objective Optimization. J. COMBIN. MATH. COMBIN. COMPUT, 127, 359-378.

[3] Heng, C., Yuhao, W., & Zixuan, Z. (2025). Research on Curriculum Ideological and Political Education in Architectural Programs under the Background of Urban Micro-renewal: Taking the Principles of Interior Design Course as an Example. Frontiers in Educational Research, 8(2).

[4] Wang, Y. (2019, August). Research on Teaching Reform of Public Basic Courses in Colleges and Universities under the Concept of “Course-based Ideological and Political Education”. In 2019 5th International Conference on Social Science and Higher Education (ICSSHE 2019) (pp. 73-76). Atlantis Press.

[5] Mettler, T., & Pinto, R. (2015). Serious games as a means for scientific knowledge transfer—A case from engineering management education. IEEE Transactions on Engineering Management, 62(2), 256-265.

[6] Pacher, C., Woschank, M., Zunk, B. M., & Gruber, E. (2024). Engineering education 5.0: a systematic literature review on competence-based education in the industrial engineering and management discipline. Production & Manufacturing Research, 12(1), 2337224.

[7] Lantada, A. D. (2020). Engineering education 5.0: Continuously evolving engineering education. International journal of engineering education, 36(6), 1814-1832.

[8] Ma, Y., & Liang, R. (2025). Research on Ideological and Political Teaching of Engineering Management Curriculum under the Background of New Era. Frontiers in Educational Research, 8(3).
[9] Xinhai, Z., Min, Y., Jie, B., Hongliang, Z., Ren, L., Xuemei, L., & Jun, L. (2022). The Construction of Engineering Ethics and Management Curriculum System Based on Curriculum Ideology and Politics. Journal of Education and Information Technology, 1(1), 13.
[10] Sun, G. (2023, July). Reform and Exploration of Ideological and Political Education in the Course of Engineering Project Management. In 2nd International Conference on Humanities, Wisdom Education and Service Management (HWESM 2023) (pp. 424-429). Atlantis Press.
[11] Lima, R. M., Mesquita, D., Rocha, C., & Rabelo, M. (2017). Defining the Industrial and Engineering Management Professional Profile: a longitudinal study based on job advertisements. Production, 27, e20162299.
[12] Ahern, A., Dominguez, C., McNally, C., O'Sullivan, J. J., & Pedrosa, D. (2019). A literature review of critical thinking in engineering education. Studies in Higher Education, 44(5), 816-828.
[13] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.
[14] Foster, M. E. (2019). Natural language generation for social robotics: opportunities and challenges. Philosophical Transactions of the Royal Society B, 374(1771), 20180027.
[15] Shin, J., & Nam, J. (2021). A survey of automatic code generation from natural language. Journal of Information Processing Systems, 17(3), 537-555.
[16] Li, H. (2018). Deep learning for natural language processing: advantages and challenges. National Science Review, 5(1), 24-26.
[17] Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2), 139-172.
[18] Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., ... & Hon, H. W. (2019). Unified language model pre-training for natural language understanding and generation. Advances in neural information processing systems, 32.
[19] Dale, R. (2020). Natural language generation: The commercial state of the art in 2020. Natural Language Engineering, 26(4), 481-487.
[20] Bouayad-Agha, N., Casamayor, G., & Wanner, L. (2014). Natural language generation in the context of the semantic web. Semantic Web, 5(6), 493-513.
[21] Li, C., & Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. International Journal of Artificial Intelligence in Education, 31, 186-214.
[22] Zhang, R., Guo, J., Chen, L., Fan, Y., & Cheng, X. (2021). A review on question generation from natural language text. ACM Transactions on Information Systems (TOIS), 40(1), 1-43.
[23] Saha, B. (2025). Generative AI for Text Generation: Advances and Applications in Natural Language Processing. Journal of Computer Allied Intelligence (JCAI, ISSN: 2584-2676), 3(1), 77-91.
[24] Chengwei, Y. A. O., Chunhui, C. H. E. N., & Mei, C. H. E. N. (2024). Artificial intelligence natural language generation experiment and instruction design for the liberal arts students. Experimental Technology and Management, 41(4), 177-184.
[25] Zizhao Zhang,Xinyue Yang,Liping Sun,Yu Sun & Jichuan Kang. (2025). Research on constructing and reasoning the collision knowledge graph of autonomous navigation ship based on enhanced BERT model. Expert Systems With Applications,278,127429-127429.
[26] Emre Delibaş. (2025). Efficient TF-IDF method for alignment-free DNA sequence similarity analysis. . Journal of molecular graphics & modelling,137,109011.
[27] Arnold Hien,Noureddine Aribi,Samir Loudni,Yahia Lebbah,Abdelkader Ouali & Albrecht Zimmermann. (2024). Mining diverse sets of patterns with constraint programming using the pairwise Jaccard similarity relaxation. Constraints,(prepublish),1-32.