Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

SBERT-based Deep Learning model for mapping of PEOs and POs with Justification Rubrics

Affia Thabassum[1], M. Mohammed Thaha[2*], A. Abudhahir[3]

[1]Department of Computer Applications,
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, Tamil Nadu, India
[2]Department of Computer science and Engineering
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, Tamil Nadu, India
[3]Department of Electronics and Instrumentation
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, Tamil Nadu, India
[2]mohammedthaha@crescent.education

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Introduction: Aligning Program Educational Objectives (PEOs) with Program Outcomes (POs) is a key step in developing a meaningful engineering curriculum. It ensures that what students learn is in line with both academic goals and industry needs. However, doing this manually can be time-consuming and biased.

Objectives: To support curriculum designers in developing more cogent and industry-relevant engineering education programs by developing an automated, objective, and efficient system that uses natural language processing to assess and align Program Educational Objectives (PEOs) with Program Outcomes (POs).

Methods: This study explores a more efficient and objective method using Natural Language Processing (NLP), specifically the Sentence-BERT (SBERT) model, to compare the meanings of PEOs and POs. We used real data from the Civil Engineering Department at B.S. Abdur Rahman Crescent Institute of Science and Technology. Since the dataset was limited we applied text augmentation techniques like synonym replacement, random insertion or deletion, and shuffling to create a more robust dataset. Cosine similarity was used to measure how closely each PEO aligns with the POs, and the results were categorized into High, Medium, Low, or No Similarity based on expert-defined thresholds.

Results: The results show that this approach is effective in identifying meaningful connections between PEOs and POs. It offers a helpful tool for curriculum designers and academic reviewers who want a clearer, more consistent way to evaluate and improve educational programs.

Conclusions: This study provides a way to connect Program Educational Objectives (PEOs) to Program Outcomes (POs) using the SBERT model. Using text augmentation approaches and fine-tuning SBERT, we successfully categorized the similarity scores into four groups: High, Medium, Low, and No Similarity. Implementing a rubric-based evaluation adds a new level of understanding to the model's judgments, enabling more informed and logical instructional planning. Future studies can concentrate on improving text augmentation methods and investigating alternative transformer-based models to BERT in order to further improve the mapping process.

Keywords: Program Educational Objectives (PEOs) , Program Outcomes (POs), Engineering curriculum, Natural Language Processing (NLP), Sentence-BERT (SBERT)

INTRODUCTION

Program Educational Objectives (PEOs) and Program Outcomes (POs) are important parts of engineering education. PEOs describe what students should be able to do a few years after they graduate, while POs describe the knowledge and skills they should have by the time they finish their degree. It is important to connect PEOs and POs

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

clearly so that the curriculum stays relevant, meets educational goals, and satisfies accreditation standards. Normally, teachers and curriculum experts do this mapping manually, which takes a lot of time and can sometimes lead to mistakes or personal bias—especially when there are many PEOs and POs. To solve this, our study uses a smart language model called Sentence-BERT (SBERT), which can understand the meaning of sentences and find how similar they are. We used PEO and PO data from the Civil Engineering Department at B.S. Abdur Rahman Crescent Institute of Science & Technology.

Using SBERT, we calculated the semantic similarity scores between each PEO-PO pair and these scores are classified into four categories: High, Medium, Low, and No Similarity. To increase our model's overall robustness and capacity, we used various text augmentation techniques such as synonym replacement, random insertion, random deletion, and text shuffling. Thus, these methods allowed us to expand the dataset and allowed the model to learn from a broader range of sentences and vocabularies. The augmented dataset was used to fine-tuning and training the SBERT model, while the original PEO-PO pairs were used for testing.

One special feature of our work is a rubric system, which explains each similarity level and also gives examples of alternative PEOs for every category. This helps teachers and curriculum designers better understand and improve the alignment. In the end, our method makes the PEO-PO mapping process faster, clearer, and more reliable. The findings state that it can significantly reduce the time and effort involved in curriculum alignment while supporting more consistent and data-driven educational planning.

Reimers and Gurevych (2019) introduced Sentence-BERT (SBERT). This initial sentence transformer model is a refined version of BERT achieved by utilising a Siamese network (Schroff et al., 2015). SBERT helps quickly create and compare sentence meanings in just a few seconds. It reduced the amount of computing work required for large text collections and it performed better than supervised models like as InferSent (Conneau et al., 2017) and Universal Sentence Encoder (Cer et al., 2023) in terms of Semantic Textual Similarity (STS). SBERT helped them improve how sentences were represented using deep learning models and inspired many new methods. However, using BERT embeddings directly for sentence representation had some problems, To solve this, Li et al. (2020) introduced a method called BERT-flow, which changes BERT embeddings to fix uneven distribution (anisotropy). In a same way, Su et al. (2021, 2023) proposed BERT-whitening, which uses a technique of machine learning to make the sentence representations more balanced and reduce their size.

Sentence embedding, also known as sentence representation learning, is a rapidly growing area of research. Most techniques in this field fall in two main categories: supervised and unsupervised methods. Supervised methods—such as SBERT (Reimers & Gurevych, 2019), InferSent (Conneau et al., 2017), and the Universal Sentence Encoder (Cer et al., 2023) require labeled data for training, while unsupervised methods do not rely on annotated datasets. In recent years, several unsupervised sentence embedding methods based on contrastive learning have been introduced, including IS-BERT (Zhang et al., 2020), DeCLUTR (Giorgi et al., 2021), CT (Carlsson et al., 2021), SimCSE (Gao et al., 2021), and DiffCSE (Chuang et al., 2022). These approaches concentrate on generating positive and negative pairs in an unsupervised way to improve sentence representations. However, most existing methods for sentence embedding are computationally expensive.

BERT (Devlin et al., 2018) can be used as a cross-encoder for tasks involving sentence or phrase pair scoring. In this setup, the two input texts are separated by a special [SEP] token, and multi-head attention is applied across all tokens together. Although this method achieves strong performance on many sentence pair tasks, it has a key limitation: it does not produce independent sentence embeddings.To address this, SBERT (Reimers & Gurevych, 2019) was introduced. It modifies BERT to encode each sentence separately, followed by applying mean pooling to the output. This results in fixed-size sentence embeddings, which can be easily compared using similarity measures like cosine similarity.

METHODS

We obtained a dataset of PEOs and POs from the Civil Engineering Department at B.S. Abdur Rahman Crescent Institute of Science & Technology. Table 1 lists the programme Edcuational Objectives and table 2 provides the programme outcomes.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1: Program Educational Objectives (PEOs)

PROGRAMME EDUCATIONAL OBJECTIVES	
PEO1	Exhibit expertise in Planning, Design, Execution and Maintenance of Civil Engineering works with environmental care.
PEO2	Design and construct Civil Engineering Infrastructure with emphasis on Durability and Sustainability.
PEO3	Develop and execute Civil Engineering projects with social relevance aiming for rural and urban development.
PEO4	Pursue Research in complex Civil Engineering problems involving multidisciplinary aspects and provide sustainable solutions.
PEO5	Exercise leadership with an ethical approach, perform in teamwork with good communication skills, and excel in cost and time management.

Table 2: Program Outcomes (Pos)

PROGRAMME OUTCOMES	
PO1	Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to solve complex engineering problems.
PO2	Identify, formulate, research literature, and analyze complex engineering problems using first principles of mathematics and sciences.
PO3	Design solutions for complex engineering problems and system components or processes with appropriate considerations for health, safety, and environment.
PO4	Use research-based knowledge and methods including experiments, data analysis, and synthesis to provide valid conclusions.
PO5	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools to complex engineering activities.
PO6	Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to professional engineering practice.
PO7	Understand the impact of professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8	Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.
PO9	Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11	Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12	Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Text Augmentation: A variety of text augmentation techniques are used to generate PEO-PO pairs to train the model as the original dataset was less.

Text augmentation techniques like Random Insertion, random deletion, text shuffling, and synonym replacement were used to generate the training dataset. More than 1,00,000 augmented PEO-PO pairs were generated for training.

Text Augmentation Techniques: Random insertion: means adding new words, particularly synonyms of existing words, randomly in a sentence. This technique adds variety to the sentence structure and adds length which makes the model more robust to different sentence formations.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Random deletion: means taking out words from a sentence at random. This makes the sentence shorter and helps the model focus on the most important words, making it better at handling incomplete or missing information.

Text shuffling: means changing the order of words or phrases in a sentence. This helps the model understand that even if the structure changes, the meaning can still stay the same.

Synonym replacement: means swapping words in a sentence with other words that have the same meaning. This shows the model different ways to say the same thing, helping it learn to understand a wider variety of sentences.

Model Training and fine-tuning*:* To map Program Educational Objectives (PEOs) to Program Outcomes (POs), we used the Sentence-BERT (SBERT) designed for sentence-level embeddings. The augmented PEO-PO pairs are used as training data. We used a pre-trained model that generated high-quality sentence embeddings and fine-tuned it based on our parameters, such as batch size, loss function, epochs, and step size.

- Batch Size: The batch size specifies how many pairs of PEO and PO are processed before updating the model's parameters
- Cosine Similarity Loss Function: This function computes the cosine of the angle between the embedding vectors of a PEO-PO pair, producing a similarity score ranging from 0 to 1
- Epoch: An epoch is one full pass through the training data. Multiple epochs help the model better learn the relationship between PEOs and POs by updating its parameters with each pass.
- Warmup Step: To prevent the model from making large, destabilising parameter updates at the start of training, we included a warmup step.

Threshold Determination and Testing: Initially, 15 experts categorised the similarity scores for the original data - Program Educational Objectives (PEOs) and Program Outcomes (POs) as High, Medium, Low, and No Similarity. Following this categorisation, the qualitative assessments were converted into quantitative scores ranging from 0 to 1, with 0 indicating no similarity and 1 indicating high similarity. Converting qualitative assessments to quantitative scores involves assigning numerical values to the different categories of qualitative data This conversion facilitated the setting of appropriate thresholds for categorising similarity scores in a more standardised manner.

Table 3: Expert labelled data for PEO1-PO1 pair

PEO-PO Pair	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Expert 12	Expert 13	Expert 14	Expert 15
PEO1-PO1	High	Medium	High	High	High	Medium	High	High	Medium	High	High	High	High	Medium	High

The pair number in table 3 indicates the specific PEO1-PO1 combination being evaluated. The qualitative labels are given by each expert for the PEO-PO pair. Labels are "High", "Medium", "Low", and "No" similarity. The average of the numerical scores assigned to each label and calculated by summing up the numerical values corresponding to each expert's label for a specific pair and dividing by the number of experts. These qualitative labels are then converted to quantitative scores and thresholds were set for High, Medium, Low and No similarity.

Once the model is trained on the augmented pairs of PEO and PO, the original dataset was given to test the model. The model produces the mapping based on the threshold set and gives us High, Medium, Low or No Similarity mapped pairs as categorized outputs.

To validate and interpret the model's output, we created rubrics where the model offers alternative PEO statements for various similarity levels. We employed text augmentation techniques on the original data again for rubric generation. Text augmentation techniques like random addition, deletion, and synonym replacement to generate these variations. After that, each PEO-PO pair was mapped using cosine similarity, and the least, median, and highest similarity pairs were identified for further justification.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

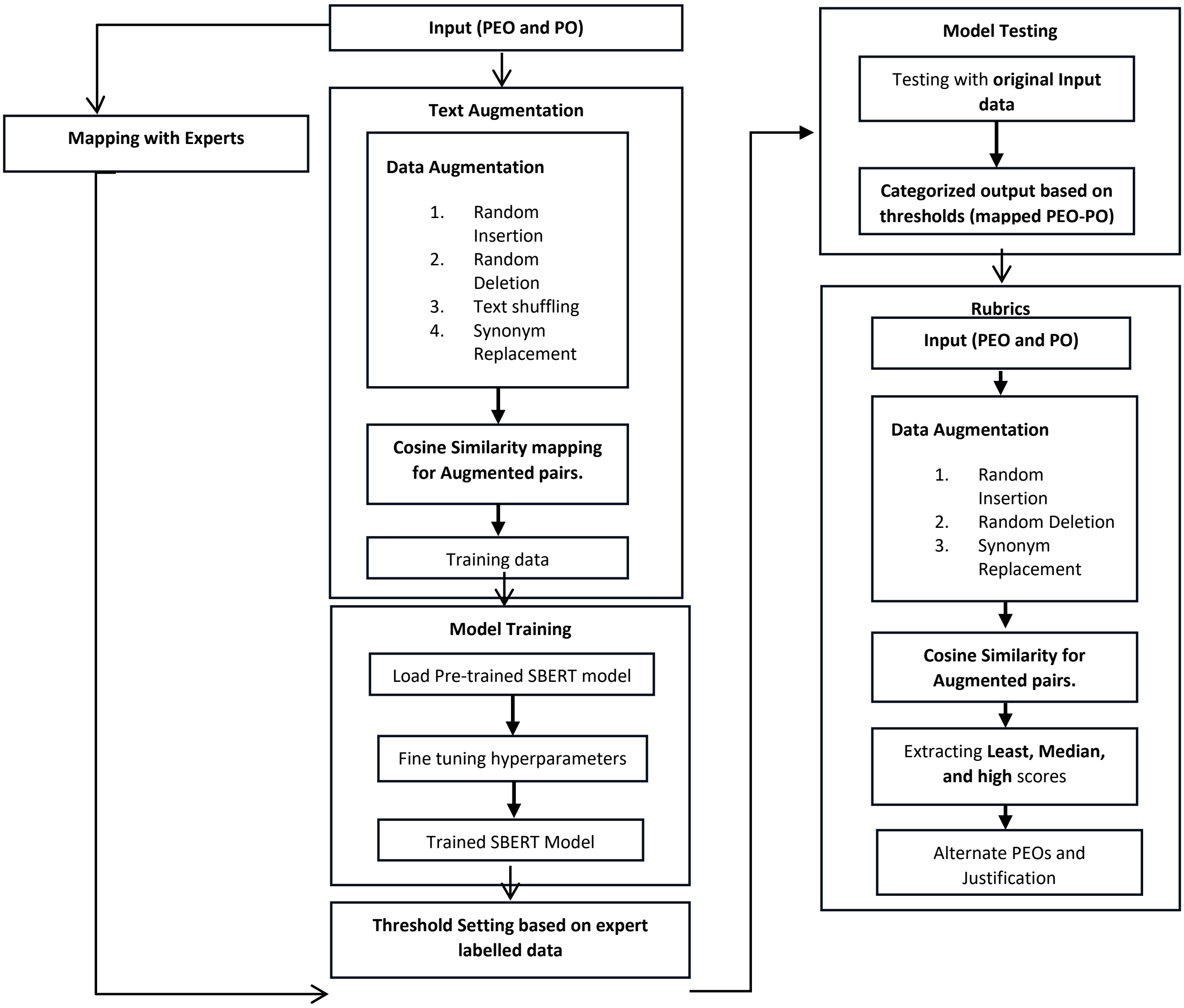

Figure 1 Mapping process of PEO's and PO's

RESULTS

The data was acquired from the Civil Engineering Department at B.S. Abdur Rahman Crescent Institute of Science & Technology with 5 PEOs and 12 POs, as given in Tables 1 and 2. Since the dataset was minimal and using it to train the model would not be efficient, text augmentation techniques such as Random insertion, Random Deletion, Text Shuffling, and Synonym Replacement were implemented, and 250 PEOs and 600 POs with a total of 1,50,000 pairs of PEO-PO were generated for training.

To effectively map Program Educational Objectives (PEOs) to Program Outcomes (POs), we used the Sentence-BERT (SBERT) designed for sentence-level embeddings. Specifically, we used the "paraphrase-MiniLM-L6-v2" pre-trained SBERT model, which is optimized for generating high-quality sentence embeddings and is well-suited for tasks requiring semantic similarity measurements. The model was pre-trained on large-scale text corpora to learn general language representations.

This pre-training provides the model with a robust understanding of language, which is important when fine-tuning it for specific tasks like PEO-PO mapping. These augmentations ensure that the model could handle different ways of expressing the same underlying educational objectives and outcomes. A batch size of 16 is used to balance between computational efficiency and the model's ability to learn from various examples within one batch. The main goal of fine-tuning was to help the model learn how to tell the difference between high, medium, low, and no similarity in PEO-PO pairs. We used a cosine similarity loss function for this. It calculates how close two sentences are by measuring the angle between their vector forms and it gives a score between 0 and 1. The model is trained to

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

give higher scores to pairs that are closely related and lower scores to those that are not closely related. The fine-tuning process was carried out over six epochs. To prevent the model from making large, destabilising parameter updates at the beginning of training, we included a warmup step. During this initial phase, the learning ratebasically, the step size used to update model parameters was gradually increased. This warmup step allows the model to converge more easily to an optimal solution, lowering the risk of overshooting and improving overall training stability. Table 4 shows the optimization parameters

Table 4: Fine-tuning

	Parameters	Values
1	Batch Size	16
2	Loss Function	Cosine Similarity Loss Function
3	Training Duration (Epoch)	6 epochs
4	Step Size	10

The training was carried out on Google Colab using an A100 GPU and 16 GB of RAM, which provided the necessary computational resources to fine-tune the SBERT model efficiently. It took about 6 hours to complete.

Converting Qualitative Labels to Quantitative Scores

To enable a more systematic analysis, the qualitative labels provided by the experts on table 3 were converted into numerical values. We assign the following numerical values to each qualitative label:

- High Similarity was assigned a score of 1.0.
- Medium Similarity was assigned a score of 0.6.
- Low Similarity was assigned a score of 0.3.
- No Similarity was assigned a score of 0.0.

This conversion is crucial because it allows us to perform mathematical and statistical analysis on the expert-labelled data. For each PEO-PO pair, the qualitative labels from all 15 experts were converted to their corresponding numerical values in Table 5. The average score for each pair was then calculated. This average score represents the consensus among the experts on the similarity between that PEO and PO.

Table 5: Quantitative Scores for expert labelled data

PEO-PO Pair	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Expert 12	Expert 13	Expert 14	Expert 15	Average Score
PEO1-PO1	1.0	0.6	1.0	1.0	1.0	0.6	1.0	1.0	0.6	1.0	1.0	1.0	1,0	0.6	1.0	0.92

The average similarity score is 0.92, which indicates a high similarity according to our thresholds. This was repeated for all pairs and to set a threshold value for mapping. Calculation for Pair PEO1-PO1:

- Expert 1 labelled "High" (1.0), Expert 2 labelled "Medium" (0.6), Expert 3 labelled "High" (1.0), and so on.
- Sum of scores: (1.0 + 0.6 + 1.0 + ... + 1.0) = 13.8
- Average Score: 13.8 / 15 = 0.92

This process was repeated for every PEO-PO pair.

Setting of thresholds: The process of setting thresholds involved determining cut-off points that would categorize the average similarity scores into High, Medium, Low, and No Similarity categories as shown in table 6.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 6: Setting of thresholds

Similarity	Threshold
High	0.55 to 1.0
Medium	0.3 to 0.55
Low	0.1 to 0.3
No Similarity	Less than 0.1

Analysing the Distribution of Scores

High Similarity (0.55 to 1.0):

- The upper threshold was set at 1.0 (perfect similarity), which corresponds to the highest possible agreement among experts.
- The lower threshold for High Similarity was set at 0.55. This value was chosen because it is above the median score and indicates a strong consensus among experts that the PEO and PO are closely related.
- Scores in this range indicate that most experts rated the PEO-PO pair as high. The cut-off at 0.55 was chosen based on expert threshold and ensures that the pairs with a strong consensus are labelled as High Similarity.

Medium Similarity (0.3 to 0.55):

- The upper threshold for Medium Similarity was set at 0.55. This value was chosen to mark the boundary where a strong relationship ends, and a moderate relationship starts.
- The lower threshold for Medium Similarity was set at 0.3. This value is below the median and specifies a moderate consensus among experts that the PEO and PO have some degree of relatedness but not a strong one.
- Scores in this range indicate that there is some agreement among experts that the PEO and PO are related but not strongly. The cut-off at 0.3 ensures moderately similar pairs are not grouped with the low-similarity pairs.

Low Similarity (0.1 to 0.3):

- The upper threshold for Low Similarity was set at **0.3**. This value was chosen to differentiate between pairs with weak relationships.
- The lower threshold for Low Similarity was set at **0.1**. This value indicates a weak relationship with minimal agreement among experts.
- Scores in this range indicate weak similarity, where only a few experts might see a connection between the PEO and PO. The lower boundary at 0.1 ensures that these weak connections are still recognized but are not mistaken for more meaningful relationships.

No Similarity (0.0 to 0.1):

- The upper threshold for No Similarity was set at **0.1**. This value was chosen to indicate a lack of meaningful relationship between the PEO and PO, where experts generally agreed that the two are not related.
- The lower threshold was naturally set at **0.0** because it represents the absence of similarity.
- Scores in this range indicate that the PEO and PO are generally not considered related by the experts. This range captures the pairs with little to no alignment, as the majority perceives.

Testing the Model: The trained model was tested on the original dataset of PEOs and POs obtained from the Civil Engineering Department at B.S. Abdur Rahman Crescent Institute of Science & Technology, given in Table 1. The

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

model calculated similarity scores were categorized into High, Medium, Low, and No Similarity based on the thresholds determined from the expert-labelled data given in Table 6. The mapping done by the SBERT model for the given thresholds is categorised in Table 7.

Table 7: Categorization Results of Similarity Scores

PEO-PO Pair	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
PEO1	High	Medium	High	Low	High	High	High	Medium	Low	Medium	High	Medium
PEO2	Medium	Low	High	Low	Medium	Medium	High	Medium	Low	Medium	Medium	Low
PEO3	Medium	Low	Medium	Low	Medium	Medium	High	Medium	Low	Medium	Medium	Medium
PEO4	High	High	High	Medium	High	Medium	High	Medium	Low	Medium	High	Medium
PEO5	Low	Low	Low	Low	Low	Medium	Medium	High	High	Medium	High	Low

Rubric-Based Evaluation and Justification: To further validate and understand the SBERT model's decisions, we developed a rubric that provides alternative PEO statements for different similarity levels. When a PEO-PO pair was mapped as "High" by the model, the rubric suggests what the PEO statement will be if it were categorized as "Medium," "Low," or "No Similarity." The rubric helps explain why the model gave a certain similarity score and helps us explore how closely or loosely PEOs and POs can match. To support this, we again used data augmentation techniques on the PEOs and POs. This allowed us to create different versions of the same PEO or PO, helping the model understand a wider range of sentence forms and meanings. The augmentation techniques used are Random Insertion, Random Deletion and Synonym Replacement. For each PEO, 10 unique augmented versions are generated using these techniques, and similarly, **10** unique versions are created for each PO. The augmented PEOs are then paired with the augmented Pos. The cosine similarity is calculated for each PEO-PO pair, and the result is categorized into one of four categories. The dataset then Extracted Least, Median, and Highest Similarity:

- The least similarity pair is the one with the lowest cosine similarity score.
- The median similarity pair is the one with the middle cosine similarity score (found by dividing the group into two halves).
- The highest similarity pair is the one with the highest cosine similarity score.

By following this approach, the dataset provides a comprehensive view of how PEOs align with POs across different similarity levels, which is valuable for educational assessments and program design. In Table 6, the model has mapped PEO1-PO1 as “high.” The rubrics provide PEO statements as to why the model did not map other similarities and what PEO statement can be given for the model to map medium, low, or no similarity. The justifications are provided for the alternate PEO statements. An example of PEO1-PO1 is mentioned in Table 8.

Table 8: Rubric-Based Evaluation Example for PEO1 and PO1

Similarity Level	PEO Statement	Justification
High	Exhibit expertise in Planning, Design, Execution and Maintenance of Civil Engineering works with environmental care.	This statement directly aligns with the knowledge and application focus in PO1, hence categorized as High.
Medium	Exhibit direct expertise in civil Planning, Design, Execution, and Maintenance of Civil Engineering works with environmental care	This statement is less specific and focuses on proficiency rather than expertise, thus categorized as Medium.
Low	Exhibit expertise in Planning, Design, Execution, and of Engineering works with care.	This statement is very general and lacks the depth required for a strong alignment with PO1, thus categorized as Low.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The rubric was used to validate the SBERT model's categorisations for all PEO-PO pairs. This process included creating alternative PEO statements for each similarity level and assessing their compatibility with the corresponding POs. No similarity was categorized for a few pairs as it did not find similarity between them. The rubric-based approach provided a clear rationale for the model's categorisation, ensuring that the mappings were both computationally valid and pedagogically sound.

A qualitative analysis was carried out to assess the model's ability to capture semantic relationships. The analysis included a manual review of selected PEO-PO pairs and their respective similarity scores.

Table 9: Qualitative Analysis of Selected PEO-PO Pairs

PEO	PO	Calculated Similarity by model	Model calculated Similarity	Expert Similarity
PEO1	PO1	0.56	High	High
PEO2	PO1	0.48	Medium	Medium
PEO3	PO1	0.46	Medium	Medium
PEO4	PO1	0.64	High	High
PEO5	PO1	0.20	Low	Low

The qualitative analysis, the calculated similarity scores for each PEO-PO pair by the SBERT model were compared with expert-labeled similarity categories, which were categorised according to the established thresholds for High, Medium, Low, and No Similarity. For example, the pair PEO1-PO1 had a calculated similarity of 0.56, which the model classified as "High." Since this score is higher than the high similarity threshold (0.55 to 1.0), most experts probably also classified this pair as "High," representating strong agreement between the model and expert assessments. Similarly, PEO2-PO1 and PEO3-PO1 had calculated similarity scores of 0.48 and 0.46, respectively, which the model classified as "Medium," which the experts also considered medium similarity (0.3 to 0.55). The PEO4-PO1 pair was similarly classified as "High," matching the expected expert judgement for this high similarity score, with a similarity of 0.64. Finally, the model classified the PEO5-PO1 pair as "Low" according to the expert threshold for low similarity, which is between 0.1 and 0.3, based on their computed similarity of 0.20. The alignment between the expert labels and the model's categorisations validates the SBERT model's applicability for this task by showing how well it captures and reflects the semantic relationships between POs and PEOs.

The findings show that the SBERT model is a useful tool for educational assessment since it can map PEOs to POs in an efficient manner. The model seems to effectively capture the semantic connections between educational goals with high accuracy and minimal error. The model's ability to generalise was improved using text augmentation techniques. However, depending too much on text augmentation techniques could result in noise, and more investigation is required to improve these approaches.

Our approach has limitations despite promising results. While improving robustness, text augmentation techniques may introduce noise which compromises the similarity score accuracy. Not all aspects of semantic relationships may be captured by using cosine similarity as the only evaluation metric. Further research endeavours will centre on optimising text augmentation techniques and exploring alternative transformer-based models, like RoBERTa and T5, to augment the data and the mapping procedure even more. Furthermore, integrating additional assessment metrics and carrying out more thorough qualitative analyses will yield a more thorough evaluation of the model's efficacy.

CONCLUSION

This study offers a method to apply the SBERT model to map Program Educational Objectives (PEOs) to Program Outcomes (POs). We accurately classified the similarity scores into four categories High, Medium, Low, and No Similarity using text augmentation techniques and fine-tuning SBERT. A new dimension to understanding the model's decisions is attained through the implementation of a rubric-based evaluation, allowing for better-informed and more rational educational planning. To further enhance the mapping process, upcoming research can focus on refining text augmentation techniques and exploring different transformer-based models other than BERT.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

REFERENCES

[1] Shaikh, S., Daudpotta, S. M., & Imran, A. S. (2021). Bloom's learning outcomes' automatic classification using lstm and pretrained word embeddings. IEEE Access, 9, 117887-117909.
[2] Li, Y., Rakovic, M., Poh, B. X., Gaševic, D., & Chen, G. (2022). Automatic classification of learning objectives based on bloom's taxonomy. In Educational Data Mining 2022. International Educational Data Mining Society.
[3] Syeed, M. M., Shihavuddin, A. S. M., Uddin, M. F., Hasan, M., & Khan, R. H. (2022). Outcome based education (OBE): Defining the process and practice for engineering education. IEEE Access, 10, 119170-119192.
[4] Anderson, J. (2020). The TATE model: a curriculum design framework for language teaching. ELT Journal, 74(2), 175-184.
[5] F. Schroff, D. Kalenichenko, and J. Philbin. "FaceNet: A unified embedding for face recognition and clustering". In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07, pp. 815–823, 2015. doi: 10.1109/CVPR.2015.7298682.
[6] N. Reimers and I. Gurevych. "Sentence-BERT: Sentence embeddings using siamese BERT-networks". In Proceedings of the Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, pp. 3982–3992, 2019. doi: 10.18653/v1/d19-1410.
[7] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes. "Supervised learning of universal sentence representations from natural language inference data". In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 670–680, 2017. doi: 10.18653/v1/d17-1070.
[8] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. John, N. Constant, M. Guajardo-Cespedes, Yuan, C. Tar, Y. Sung, B. Strope, and R. Kurzwell. "Universal Sentence Encoder" 2018, Accessed: May 22, 2023. [Online]. Available: http://arxiv.org/abs/1803.11175.
[9] B. Li, H. Zhou, J. He, M. Wang, Y. Yang, and L. Li. "On the sentence embeddings from pre-trained language models". In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 9119–9130, 2020. doi: 10.18653/v1/2020.emnlp-main.733.
[10] J. Su, J. Cao, W. Liu, and Y. Ou. "Whitening Sentence Representations for Better Semantics and Faster Retrieval" 2021, Accessed: Jul. 15, 2023. [Online]. Available: http://arxiv.org/abs/2103.15316.
[11] Y. Zhang, R. He, Z. Liu, K. H. Lim, and L. Bing. "An unsupervised sentence embedding method by mutual information maximization". In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1601–1610, 2020. doi: 10.18653/v1/2020.emnlp-main.124.
[12] J. Giorgi, O. Nitski, B. Wang, and G. Bader. "DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations". In 59th Annual Meet of Association for Computational Linguistics and Proceedings of 11th International Joint Conference on Natural Language Processing, pp. 879–895, 2021. doi: 10.18653/V1/2021.ACL-LONG.72.
[13] F. Carlsson, E. Gogoulou, E. Ylipää, A. C. Gyllensten, and M. Sahlgren. "SEMANTIC RE-TUNING WITH CONTRASTIVE TENSION". In Proceedings of 9th International Conference on Learning Representations, 2021.
[14] T. Gao, X. Yao, and D. Chen. "SimCSE: Simple Contrastive Learning of Sentence Embeddings". In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 6894–6910, 2021. doi: 10.18653/v1/2021.emnlp-main.552.
[15] Y. S. Chuang, R. Dangovski, H. Luo, Y. Zhang, S. Chang, M. Soljačić, S. Li, W. Yih, Y. Kim, and J. Glass. "DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings". In Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4207–4218, 2022. doi: 10.18653/v1/2022.naacl-main.311.
[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.