International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Comparative Analysis of Word Embedding Techniques for Proposition Extraction in Concept Map Mining

[1]Bramhankar Smita Gulabrao, [2]Dr. Pooja Sharma

[1,2] Department of Computer Science, Kalinga University, Naya Raipur, Chhattisgarh, India

Abstract

A vital part of knowledge representation, concept map mining allows for the extraction of structured data from unstructured text. Among the most difficult tasks in this field is proposition extraction, which entails finding relevant connections between ideas. By converting words into high-dimensional vector representations that contain syntactic and semantic links, word embedding approaches have greatly enhanced natural language processing (NLP) applications. Various fields, including AI, computer science, biology, and engineering, are represented in the collection, which also includes instructional materials, Wikipedia pages, and scholarly books. In order to ensure that the dataset was consistent and of high quality, it was preprocessed before training began. We looked examined Word2Vec, GloVe, FastText, and BERT, four word embedding models, to see how well they could extract relevant propositions. Accuracy, recall, and F1-score were used to evaluate performance. First place went to BERT with the greatest F1-score, followed by FastText, GloVe, and Word2Ve, according to the findings. Training BERT took 600 seconds, but GloVe only needed 90 seconds, indicating that BERT needed much more computing resources. Finding the right method for proposition extraction in knowledge representation tasks may be challenging; this work sheds light on the trade-offs between accuracy and computing efficiency in embedding models.

Keywords: Proposition Extraction, Word Embedding, Dataset, Knowledge, Map Mining

I.INTRODUCTION

Through nodes and linkages, concept maps provide effective tools for knowledge representation that help one to see connections between many ideas. By allowing methodical and ordered information retrieval, these maps are useful in many fields like knowledge management, artificial intelligence, and education. Still, the automated creation of concept maps from textual data is difficult and calls for efficient approaches to extract relevant propositions defining the links between ideas. In this sense, Natural Language Processing (NLP) and Machine Learning (ML) approaches have become very important; word embedding methods are especially important for extracting propositions for idea map mining.

By transforming words into continuous vector representations that capture syntactic and semantic links, word embeddings have transformed natural language processing. To better understand word similarities and interconnections, word embeddings use large-scale corpus-based learning, in contrast to standard natural language processing methods like statistical and rule-based approaches, which frequently fail to retain contextual meaning. Words' contextual and relational meaning could not be completely captured by the frequency-based models used in the older approaches, such as Term Frequency-Inverse Document Frequency (TF-IDF) and n-grams. The inflexibility of rule-based systems un dealing with varied language patterns was a result of their reliance on predetermined syntactic structures. Traditional natural language processing methods struggled due to these restrictions to glean useful propositions for the purpose of creating concept maps. Word2Vec, GloVe, and FastText are examples of word embedding models that take contextual use into account to provide a more dynamic representation of words and improve concept link detection. The preparation of textual input is the first step in the multi-stage process of concept map development utilizing word embeddings. Due to its inherent noise and irrelevant components, raw text data is not always useful for concept extraction. We apply preprocessing techniques including stemming, lemmatization, tokenization, and stop-word deletion to improve the quality of the words that are extracted. While tokenization breaks down text into individual words or phrases, stop-word deletion gets rid of frequent terms like "is," "the," and "and" that don't add anything to the content. Words can be consistently represented by reducing them to their root forms by

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

stemming and lemmatization. Word embedding techniques are employed to transform words into numerical vector representations once the text is preprocessed. Word meaning and connections within the corpus are both captured by these vectors.
You may use pre-trained models or train your own to build word embeddings. Learning word vectors from a given dataset during training from scratch guarantees domain-specific relevance but demands a lot of computational power and text data. Another option is to use a pre-trained model. Google Word2Vec, Stanford GloVe, and Facebook FastText are all examples of such models that have already learnt to represent words from large text corpora. You may fine-tune these pre-trained embeddings for domain-specific applications, but they are great at capturing generic language patterns. Key concepts and their relationships are identified using a variety of computer approaches after word vectors have been collected.

Role of Word Embedding in Proposition Extraction

By converting words into dense vector representations that reflect their semantic and contextual links, word embedding is crucial in proposition extraction. In concept map mining, propositions are ideas (nodes) and their connections (edges), hence it is crucial to precisely find significant correlations between words. Semantic variants, polysemy, and contextual comprehension all challenge conventional keyword-based and rule-based approaches. By storing words in a multi-dimensional space, word embedding methods solve these problems by enabling similar words to be closer to one another depending on their meaning rather than just lexical similarity.

By use of static word representations produced by embedding models such as Word2Vec, GloVe, and FastText, one may find possible idea pairings depending on word co-occurrence and contextual similarity. These methods help to capture synonymy, analogy, and word connections, thereby improving the correctness of the acquired assertions. They could, however, not be able to differentiate word meanings in many circumstances. By producing dynamic, context-aware embeddings that guarantee word meanings are interpreted according on the surrounding text, transformer-based models such BERT and RoBERTa bypass this restriction. In complicated idea maps, where words could have many meanings based on their use, this is very helpful.

Using word embeddings helps proposition extraction to become more semantically accurate, scalable, and automated, hence lowering the need for intensive human annotations. Embedding-based approaches also help to extract domain-specific propositions, therefore enhancing performance in specialized sectors as scientific research, education, and healthcare. Word embedding is ultimately a useful technique for improving concept map generating and knowledge representation systems as it improves the capacity to harvest structured information from unstructured material.

Popular Word Embedding Techniques for Proposition Extraction

Extraction of meaningful propositions in concept map mining depends on precisely capturing semantic links between words. By expressing words as dense vectors in a high-dimensional space where related words are put closer to one another, word embedding methods assist to accomplish this. There have been created many word embedding techniques, each with special advantages and uses. Word2Vec, GloVe, FastText, and Transformer-based models like BERT and RoBERTa are the most often used approaches for proposition extraction in idea maps.

Word2Vec (CBOW and Skip-gram)

Originally presented by Mikolov et al. (2013), Word2Vec is among the first and most powerful word embedding models. It learns word representations from vast text corpora using a neural network-based method. Word2Vec consists of two primary architectures:

- **Continuous Bag of Words (CBOW):** This approach uses the words of surrounding context to forecast a target word. It generates a probability distribution over possible target words from many words as input. Computationally efficient, CBOW performs well when trained on big datasets.
- **Skip-gram:** Skip-gram forecasts surrounding words given a single target word unlike CBOW. It is appropriate for proposal extraction in concept maps where important connections may not always be in close proximity as it is more successful in catching unusual words and long-distance relationships between words.

Effective capturing of semantic and syntactic links between words by Word2Vec helps to find relevant concept pairings in proposition extraction. It does not, however, take word order or subword information, which might be a drawback in challenging proposal mining applications.

GloVe (Global Vectors for Word Representation)

Developed by Pennington et al. (2014), GloVe is another often used word embedding method with an eye on global word co-occurrence data from a corpus. Whereas Word2Vec depends on local context windows, GloVe generates word vectors by factorizing a word co-occurrence matrix. The main point is that their statistical distribution across a huge text corpus embeds word meanings.

GloVe's capacity to record local and worldwide word connections is one of its advantages as it helps in proposal extraction. GloVe can predict similarities between words even in cases where they do not occur often by using co-occurrence probability. In concept map mining, where many ideas may have indirect but significant connections, this quality makes it very useful. GloVe does not take subword information, however, and struggles with out-of-vocabulary (OOV) terms.

FastText (Subword-based Embeddings)

Facebook AI Research's FastText expands Word2Vec by adding subword data. FastText views each word as a set of character n-grams rather than as atomic units. For instance, the word "embedding" may be dissected into subword elements like "emb,," "eddi,," and "ding." These subword representations enable the model to extend more widely to morphologically rich languages and unusual terms.

FastText provides a major benefit for proposal extraction in managing variances in word forms and domain-specific terminology. In specialist domains where concept names could have many variants, this is very helpful. FastText may also deduce embeddings for unseen words, therefore addressing the OOV issue typical in Word2Vec and GloVe. FastText could still find it difficult to convey complex contextual meanings, so transformer-based models become more useful even with these benefits.

Transformer-based Embeddings (BERT, RoBERTa)

Particularly BERT (Bidirectional Encoder Representations from Transformers) and its improved counterpart RoBERTa (Robustly improved BERT Pretraining Approach), transformer-based models mark a major development in word embeddings. Transformers provide dynamic embeddings depending on word context unlike other methods producing stationary word vectors.

- **BERT:** Designed by Google AI, BERT is a deep bidirectional model learning word representations concurrently from left and right contexts. BERT is therefore rather good in deciphering word meanings in many circumstances. BERT can separate polysemous terms from extract more precise connections between ideas in proposition extraction in concept maps.
- **RoBERTa:** Using bigger datasets and eliminating the Next Sentence Prediction (NSP) target, RoBERTa enhances upon BERT's training efficiency. Better contextual knowledge and more accuracy in NLP tasks, including proposition extraction, follow from this.

Transformer-based embeddings are perfect for complicated concept linkages in concept maps as they shine in managing contextual variations. For fine-tuning, which might be difficult in certain applications, they do, however, need substantial computer resources and big labeled datasets.

II. REVIEW OF LITERATURE

Nugumanova, A. et al., (2021) One way to visualize information is via a concept map, which takes an input word or domain and represents it conceptually. You may save time reading and analyzing by creating a concept map that shows the interconnected ideas in a book or topic and how they relate to one another. But making concept maps is a tedious and time-consuming procedure. The topic of automatically generating concept maps from NLP texts is now the subject of much study. There is a lot of real-world value to the issue, but theoretically, the solutions are mostly language-dependent. Kazakh and other low-resource languages have a significant challenge when trying to use such approaches since they rely on high-quality annotated linguistic resources. We show our experimental work on automatically creating idea maps from English, Kazakh, and Russian literature and examine the challenges connected to language-dependent techniques. Using the example of a popular language-dependent approach called ReVerb, which was first created for English; we investigate the problems we have seen with the Kazakh and Russian languages.

Chandrasekaran, Dhivya et al., (2021) one area of Natural Language Processing that still needs further investigation is semantic textual similarity. There has been a lot of study in this area, and new transformer-based models have been able to get almost flawless results on benchmark datasets like the STS Dataset and the SICK dataset. In this research, we examine the datasets' sentences and determine how different word embeddings react to complicated sentences. Fifteen human annotators contributed semantic similarity scores to 50 phrase pairings that we used to construct a complicated sentence dataset in this work. An examination of readability is carried out to bring attention to the fact that the proposed dataset and the current benchmark datasets include phrases that are more difficult. In addition, we compare the suggested dataset to the current benchmark datasets and evaluate the performance of several language models and word embeddings. Findings reveal that embedding model performance is significantly affected by phrase complexity, with Pearson's and Spearman's correlation dropping by 10-20% as a consequence.

Pinandito, Aryo et al., (2021) Digital concept maps not only helped students learn, but they also made learning more interesting, interactive, and enjoyable. The core activity of the learning framework known as Kit-Build concept map is the decomposition of concept maps. A teacher's concept map is crucial in Kit-Build because students learn by assembling digital concept maps from a collection of teacher's map components. Kit-Build teacher concept maps should include the following elements: the learning context and method, the students' current level of comprehension, the instructor's goal and objective, and a set of focus questions. Because the broad idea maps produced by the automated generation technique are deemed inappropriate for the reasons stated, the semi-automatic approach is favored. With the addition of a new support function, the Kit-Build idea mapping tool can now produce concept maps semi-automatically using the idea Map Mining technique. This research presents the architecture of the idea map creation process's extraction and summarization step, which recommends the extracted concepts and proposition, triples to educators. This research set out to address that gap by investigating the efficacy of the support system's recommendations in a real-world setting for teaching English as a foreign language reading comprehension using Kit-Build. Specifically, the results indicated that the recommended Kit-Build idea map writing assistance tool is best suited for more thorough editing of concept maps.

Martinez-Rodriguez, Jose et al., (2020) The Semantic Web defines best practices for the representation of data pertaining to actual things (entities) and the relationships between them (properties). People and apps may benefit from this when it comes to sharing and using information. Unfortunately, much of the data is embedded in words written in normal language, which lack the linguistic description and structural preparation necessary for computers to handle the data directly. So, the trick is to figure out what data can be represented and then extract it. Therefore, this article lays forth a plan to use Semantic Web standards to extract data from phrases and their representations. In order to construct RDF triples (Subject-Predicate-Object structures), our approach makes use of Information Extraction tasks and a hybrid semantic similarity measure to get entities and relations from a Knowledge Base. These are then linked to persons and properties. Our technique outperforms a pattern-based method from the literature in terms of accuracy, and the trials show that it is feasible.

Hayashi, Toshitaka et al., (2019) there are a number of Natural Language Processing (NLP) jobs that Word2vec can simplify. We assume it has the potential to divide the word2vec vector space into positive and negative dimensions. Therefore, Sentiment Analysis activities may be accomplished with word2vec. Our prior work suggested a technique for sentence-level sentiment analysis that makes use of word embeddings (WEMB). The sentence vector is computed using Word's vectors from WEMB. The polarity and sentence vector are used to train the classification model. Once trained, the model can guess the unlabeled sentence's polarity. Unfortunately, the method's treatment of all words as equal weight while creating a phrase vector rendered it inadequate. To address this issue, we provide a solution in this article. When determining the sentence vector, we give each word a weight based on how significant it is. When compared to an approach that does not take word significance into account, the suggested method yields better results in terms of accuracy. But compared to the cutting edge, there is still a huge gap. We showcase upcoming projects and talk about the next upgrade.

Lorenzetti, Carlos et al., (2016) In order to collect and transmit human knowledge, electronic concept maps that are connected to other concept maps and multimedia resources may provide extensive knowledge models. This article discusses and assesses several approaches that might aid experts in expanding current knowledge models by generating suggestions for new, contextually relevant subjects using data retrieved from the web. There are two main areas of investigation into the process of topic generation for knowledge model extension: first, how to extract topic discriminators and descriptors from concept maps; and second, how to apply these extractors and descriptors to find novel and relevant candidate topics on the web. First, the essay establishes the theoretical groundwork for a "topic suggested" to facilitate information search inside the framework of a knowledge model that is being built, in order to answer these concerns. Presenting and evaluating algorithms used in EXTENDER, a subject recommendation tool, it is built on this architecture. An extensively used framework for facilitating knowledge modeling using concept maps, EXTENDER has been designed and tested inside CmapTools. Nevertheless, the algorithms' adaptability allows them to be used for a wide range of knowledge modeling systems and for search on the web in particular.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Bojanowski, Piotr et al., (2016) for several NLP applications, continuous word representations learned on large unlabeled corpora prove to be invaluable. Assigning a unique vector to each word allows many popular models to learn such representations to disregard word morphology. This is a problem, particularly for languages that are very rich in morphology and have a big vocabulary full of uncommon terms. Each word is shown as a bag of character n-grams in the skip-gram model, which is the basis of our innovative technique in this work. Each letter n-gram has its own vector representation, and words are just the total of all these representations. Because of how fast our technique is, we can rapidly train models on massive corpora. We test the acquired word representations on word analogies and similarity tasks across five languages.

Chen, Nian-Shing et al., (2008) the idea map and its many uses, particularly in online education, have been highlighted in recent studies. As an example, creators of adaptive learning materials often find themselves referring to domain-specific idea maps. In addition, idea maps may display both superficial and deep understanding of a topic. Learners may experience less information overload and learning disorientation if domain knowledge is graphically represented, according to research in the literature. Building idea maps, however, is a labor-intensive and expensive process that has traditionally depended on domain specialists. Because of the dynamic nature of new fields like e-Learning, creating concept maps for them is an even greater challenge. Building e-Learning domain idea maps from scholarly publications is the goal of this work. We use text-mining methods to automatically build e-Learning domain idea maps from a small set of relevant journal articles and conference papers. The created concept maps may serve as a valuable resource for those new to the area of e-Learning, educators creating adaptive learning materials, and students grasping the big picture of e-Learning domain knowledge.

III.RESEARCH METHODOLOGY

Dataset

The dataset used for this research includes an anthology of scholarly works, course materials, and Wikipedia articles spanning fields as diverse as engineering, computer science, biology, and artificial intelligence. Because of the dense conceptual frameworks in these texts, we believe they are perfect for testing proposition extraction techniques in concept map mining.

The dataset was preprocessed to improve data quality and guarantee consistency before word embedding methods were used. Following the cleaning of the dataset, several word embedding algorithms for proposition extraction were trained and evaluated using it.

Word Embedding Techniques

By using numerical vector representations, word embeddings are able to capture the semantic meaning and connections of words. Word2Vec, GloVe (Global Vectors for Word Representation), FastText, and BERT (Bidirectional Encoder Representations from Transformers) are the four word embedding approaches that were examined in this research for the purpose of proposition extraction.

The preprocessed dataset was used to train each embedding model. Then, their performance in proposition extraction was assessed using F1-score, recall, and precision.

IV.RESULTS AND DISCUSSION

Table 1: Performance Metrics for Proposition Extraction

Model	Precision	Recall	F1-Score
Word2Vec	72.3%	68.5%	70.3%
GloVe	74.1%	70.2%	72.1%
FastText	78.5%	74.6%	76.5%
BERT	86.7%	82.4%	84.5%

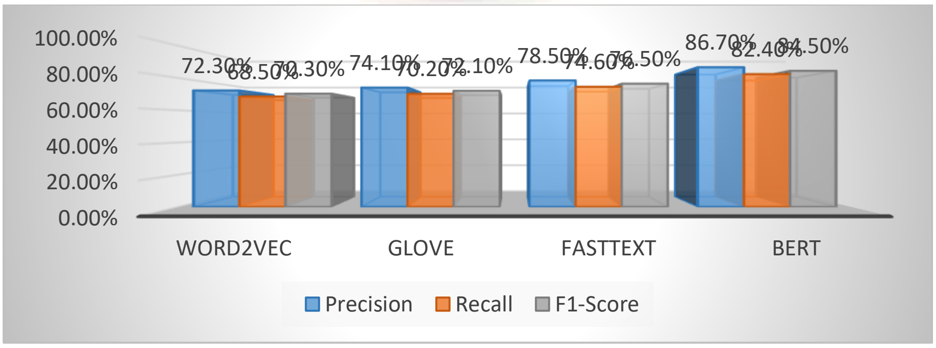

Figure 1: Performance Metrics for Proposition Extraction

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Table 1 shows that out of the four models, BERT had the best performance metrics, showing that it could accurately extract meaningful propositions with a balanced trade-off between recall and precision (86.7% precision, 82.44% recall, and 84.5%). Following closely behind, FastText achieved an F1-score of 76.5%. Its proficiency in handling uncommon and misspelled words is bolstered by its subword-level representation. While Word2Vec achieved the lowest F1-score at 70.3%, GloVe did somewhat better with a score of 72.1%. So, while Word2Vec does a good job of capturing semantic links, it can't compete with more sophisticated models when it comes to contextual depth.

Table 2: Execution Time (in Seconds) for Different Models

Model	Training Time	Inference Time
Word2Vec	120 sec	0.25 sec
GloVe	90 sec	0.22 sec
FastText	135 sec	0.30 sec
BERT	600 sec	1.10 sec

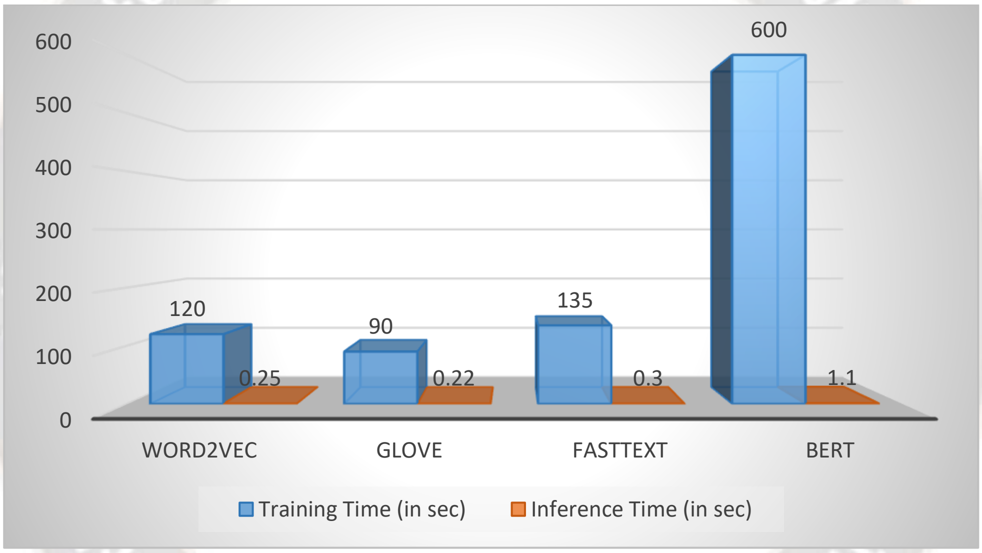

Figure 2: Execution Time (in Seconds) for Different Models

With a training time of 600 seconds and an inference time of 1.10 seconds, BERT was the most resource-intensive model, according to the execution time metrics in Table 2. However, it achieved the greatest accuracy in proposition extraction. Contrarily, GloVe's inference time was the lowest at 0.22 seconds and training time was the shortest at 90 seconds, suggesting that it was efficient in both training and real-time processing. Word2Vec was a reasonably efficient model that follows with a 120-second training time and a 0.25-second inference time. FastText was somewhat slower than Word2Vec and GloVe, taking 135 seconds for training and 0.30 seconds for inference, but it provided balanced performance in proposition extraction.

Table 3: Memory Usage for Different Models

Model	Memory Usage (MB)
Word2Vec	300
GloVe	250
FastText	400
BERT	1200

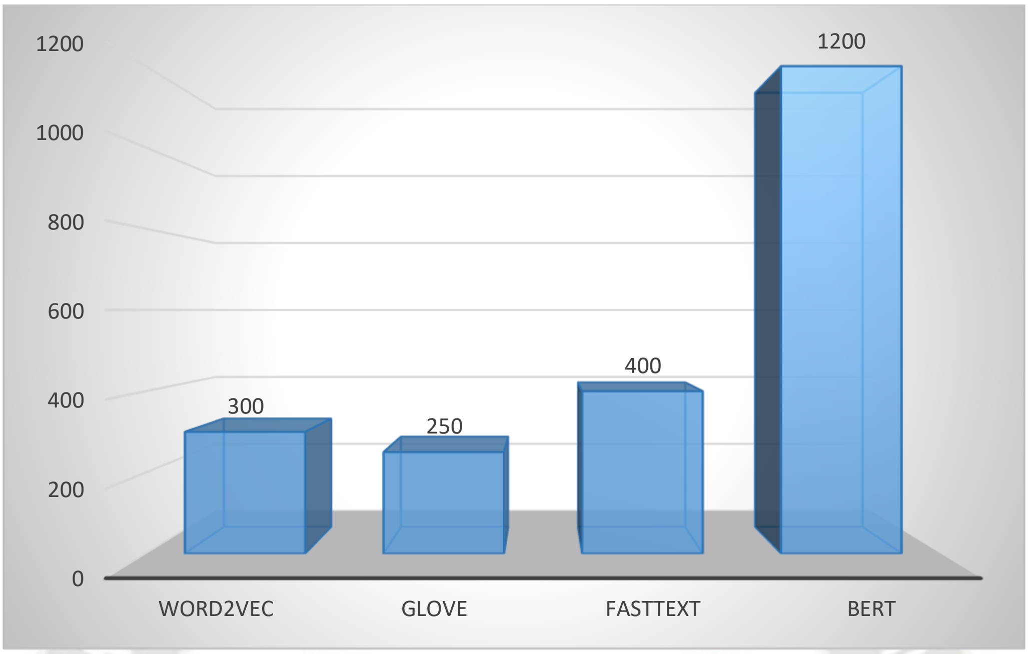

Figure 3: Memory Usage for Different Models

Table 3 emphasizes the compromise between model complexity and memory economy. BERT is the most resource-intensive model as its deep contextual learning causes the largest memory consumption—1200 MB. Conversely, GloVe (250 MB) and Word2Vec (300 MB) are lightweight and provide effective word connections with less computational need. FastText (400 MB) improves unusual word handling by adding subword representations, therefore balancing memory consumption with speed.

V.CONCLUSION

It is clear from the findings that BERT can successfully capture contextual connections, since it beats classic embedding models such as Word2Vec, GloVe, and FastText in terms of recall, accuracy, and F1-score. Nevertheless, BERT's implementation should be carefully assessed in light of available resources due to its substantial computational cost. By providing comparable accuracy with minimal computational needs, FastText provides a balanced option. For applications that need quicker processing speeds, GloVe and Word2Vec are still good choices, even if they aren't great at capturing complicated meanings. As a whole, this study demonstrates how to improve proposition extraction in NLP and knowledge representation by choosing an embedding model that strikes a good balance between accuracy and efficiency.

REFERENCES: -

[1] A. Nugumanova, A. Tlebaldinova, Ye. Baiburin, and Ye. Ponkina, "Natural Language Processing Methods For Concept Map Mining: The Case For English, Kazakh And Russian Texts," *Journal of Mathematics, Mechanics and Computer Science*, vol. 112, no. 4, pp. 102-108, 2021.

[2] A. Pinandito, D. Prasetya, Y. Hayashi, and T. Hirashima, "Semi-Automatic Concept Map Generation Approach of Web-Based Kit-Build Concept Map Authoring Tool," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 8, pp. 50-70, 2021.

[3] D. Chandrasekaran and V. Mago, "Comparative Analysis of Word Embeddings in Assessing Semantic Similarity of Complex Sentences," *IEEE Access*, vol. PP, no. 99, pp. 1-1, 2021.

[4] Z. Shao *et al.*, "Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 2, pp. 539-551, 2020.

[5] J. Martinez-Rodriguez, I. Lopez-Arevalo, and A. Rios-Alvarado, "Mining information from sentences through Semantic Web data and Information Extraction tasks," *Journal of Information Science*, vol. 48, no. 2, pp. 1-3, 2020.

[6] S. Alomari and S. Abdullah, "Improving an AI-Based Algorithm to Automatically Generate Concept Maps," *Computer and Information Science*, vol. 12, no. 4, pp. 72–83, 2019.

[7] R. Nacheva *et al.*, "Concept Map Mining Approach Based on the Mental Models Retrieval," *TEM Journal*, vol. 8, no. 4, pp. 1484-1489, 2019.

[8] R. Nacheva, L. Todoranova, S. Sulova, and B. Penchev, "Concept Map Mining Approach Based on the Mental Models Retrieval," *TEM Journal*, vol. 8, no. 4, pp. 1484-1491, 2019.

[9] T. Hayashi and H. Fujita, "Word Embeddings-based Sentence-Level Sentiment Analysis considering Word Importance," *Acta Polytechnica Hungarica*, vol. 16, no. 7, pp. 7-24, 2019.

[10] L. Yao, Y. Zhang, Q. Chen, H. Qian, B. Wei, and Z. Hu, "Mining coherent topics in documents using word embeddings and large-scale text data," *Engineering Applications of Artificial Intelligence*, vol. 64, no. 2, pp. 432-439, 2017.

[11] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, no. 1, pp. 4-9, 2016.

[12] C. Lorenzetti, A. Maguitman, D. Leake, F. Menczer, and T. Reichherzer, "Mining for Topics to Suggest Knowledge Model Extensions," *ACM Transactions on Knowledge Discovery from Data*, vol. 11, no. 2, pp. 1-30, 2016.

[13] S. Lee, Y. Park, and W. C. Yoon, "Burst analysis for automatic concept map creation with a single document," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8817-8829, 2015.

[14] T. Atapattu, K. Falkner, and N. Falkner, "Evaluation of concept importance in concept maps mined from lecture notes," in *Proceedings of the 6th International Conference on Computer Supported Education*, 2014, pp. 75-84.

[15] W. C. Yoon, S. Lee, and S. Lee, "Burst analysis of text document for automatic concept map creation," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer, Cham, 2014, pp. 407-416.

[16] J. Reis, A. Gaia, and R. Viégas Jr, "Concept Maps Construction Based on Exhaustive Rules and Vector Space Intersection," *IJCSNS International Journal of Computer Science and Network Security*, vol. 14, no. 7, pp. 26-31, 2014.

[17] I.Qasim, J.-W. Jeong, J.-U. Heu, and D.-H. Lee, "Concept Map Construction From Text Documents Using Affinity Propagation," *Journal of Information Science*, vol. 39, no. 6, pp. 719-736, 2013.

[18] K. Zubrinic, D. Kalpic, and M. Milicevic, "Constructing concept maps for adaptive learning systems based on data mining techniques," *Expert Systems with Applications*, vol. 40, no. 7, pp. 2746-2755, 2013.

[19] K. Zubrinic, D. Kalpic, and M. Milicevic, "The automatic creation of concept maps from documents written using morphologically rich languages," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12709-12718, 2012.

[20] N.-S. Chen, Kinshuk, C.-W. Wei, and H.-J. Chen, "Mining e-Learning domain concept map from academic articles," *Computers & Education*, vol. 50, no. 3, pp. 1009-1021, 2008.