Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 11
November 2023

Investigation of Deep Learning Algorithms to Reduce the Text Complexity

[1]Mohammad Jafarabad, [2]Behrouz Minaei Bidgoli
[1]Department of Computer and Information Technology Qom University
Qom, Iran
[2]Department of Computer engineering University of Science and Technology
Tehran, Iran

Abstract- Converting complex texts to simple texts and reducing the complexity of scientific articles are among the issues that have received more attention with the development of deep learning techniques. The simplification of texts greatly helps to expand the dissemination of knowledge. Deep learning semantic similarity algorithms are used to recognize words vector with high accuracy. In this research, by introducing machine learning algorithms such as word2vec, glove, fast text, Bert and Elmo, we will introduce the best algorithm with high efficiency in recognizing and solving the complexity of texts.

Keywords- text mining, Bert algorithm, Elmo algorithm, deep learning, simplification.

Introduction

One of the problems that have not been completely solved in the computer world today is the problem of converting complex texts into simple texts. The complexity of language makes it difficult to work simultaneously on all its parts (phonology, semantics, etc.). Therefore, in this study, we examine the use of deep learning algorithms to recognize and eliminate the complexity of words in scientific texts.

If we implement a plan to turn complex knowledge texts into simple, of course, a big step can be taken to facilitate access to knowledge needs [1]. Any project that can turn written knowledge sources in English into a simple, can be a world-wide knowledge-based project on text simplification.

In the second part of this research, we discuss the background of semantic similarity methods. In the third section, we will introduce and study the state of art methods of neural network algorithms. In the fourth section, we will introduce the appropriate algorithm for simplifying scientific texts and in the fifth section, conclusions will be expressed.

Background Research

Text mining is defined as a knowledge processing in which a user deals with a set of documents and a set of analysis tools. Similar to data mining, text mining seeks to obtain useful information from data sources based on identity and interesting pattern exploration. In the case of text mining, there is a set of documents, but there isn't interesting patterns can be found among the records of the official database.

In terms of complexity and difficulty of languages, English is not comparable to a language like Chinese [2]. Many of English words are rooted in ancient Greek and Latin, Therefore, it is necessary to think of a solution to solve the complexities of English language.

Language is one of the most important mental phenomena that have been prevalent among people in different parts of the world for many years. There are currently about 7,000 languages in the world. Of these, 2,400 are endangered and 840 of them are used in New Guinea alone. Linguists classify language in the areas of syntax, phonology (the place of chained and super-chained phonetic elements in the language system) and semantics (examining meanings in human languages).

One of the most important issues in recognizing complexity is familiarity with semantic similarity. When a complex word is found in the text, the word closest in meaning should be replaced. Therefore, it can be said that in a way, most text simplification activities end with recognizing semantic similarity. Of course, other parameters may be used to assess simplicity. In the following, we will review the most common methods of recognizing lexical similarity in the past. [3]

A. Explicit Semantic Analysis (ESA)

In order to process natural language, we need to acquire vast amounts of universal knowledge. Past work on semantic relevance was based on statistical techniques that did not use background or prior knowledge. Explicit semantic analysis was designed as a way to improve text classification and was based on calculating the semantic relationship using cosine similarity between vectors. This method provides meaning in the multidimensional space

of natural concepts derived from Wikipedia, the largest available encyclopedia.

In natural language processing and information retrieval, Explicit Semantic Analysis (ESA) is a display of text that uses body of documents as a knowledge base. The document set in this analysis is the Wikipedia text set. Explicit semantic analysis is a method that extracts the meaning of texts in the multidimensional space of concepts derived from the source. Machine learning techniques are used to present the explicit meaning of any text as a measured vector.[4]

Explicit semantic analysis is a technique for calculating "semantic relevance" between texts using distributive information. The ESA represents text as content vectors explicitly defined by humans. The amplitude of each dimension in the vector is equal to the weight of the text related to that dimension / explicit concept.

B. Latent Semantic Analysis

Latent Semantic Analysis (LSA) is one of the best and most well-known dimensional algorithms used in data retrieval. Its most attractive practical feature is the ability to interpret the dimensions of the vector space result as conceptual contents and, in fact, to analyze the conceptual relationship between the implicitly executed terms during matrix analysis. LSA often does not perform well in large heterogeneous assemblies.[5] Many successful techniques related to dimensionality in clustering and document retrieval have been proven.

C. Generalized Latent Semantic Analysis (GLSA)

Latent Conceptual Analysis (GLSA) is a framework for calculating the conceptual motivation of document terms and vectors. In contrast to LSA and other dimensional algorithms used for documents, this generalization focuses on the calculation of terminology vectors and document vectors computed as linear combinations. Thus, unlike LSA, GLSA is not based on qualitative word document vectors.

The GLSA method is able to combine any similarity in the term space with any dimensional reduction method. The traditional term-document matrix is used in the last step to prepare the weights in the linear composition of the term vector. One of the advantages of providing a glossary-based GLSA document is that there is no out-of-sample problem for new documents.[6] It is true that this problem exists for new terms, but new terms occur at much lower rates of documentation. In addition, these rare and new terms do not help much in classifying or retrieving documents. GLSA provides good flexibility in detecting signs and symbols of semantic connection between expressions.

D. Mutual information

Mutual information is a measure of the amount of information that a random variable has about other random variables, and the reduction in uncertainty of one random variable is due to the knowledge of another[7]. The amount of mutual information is defined according to the following relation:

$$X\ ;\ Y\ ) \ = \sum_{x\in X}\sum_{y\in y}\mathcal{P}(x\ ,y)\ \log\ \frac{\mathcal{P}(x\ ,\ y)}{\mathcal{P}(x)\mathcal{P}(y)}(\text{I}$$

E. Pointwise Mutual Information

Pointwise Mutual information is a measure of the interdependence between two random variables. This criterion has been proven to be a good dependency criterion in many natural language processing applications. Research on the generalization of Pointwise mutual information was presented in two main articles. The first examined the interaction of several random variables in the field of information theory, and the second presented an initial generalization of reciprocal information based on the concept of conditional disorder. [8]

F. Interaction information

Interaction information, also called co-information, is conditional on the concept of reciprocal information. Conditional reciprocal information is the reciprocal information of two random variables conditioned on the third variable[9].

$$\sum_{x\in X}\ \sum_{y\in Y}\ \sum_{z\in Z}\mathcal{P}(x,y,z)\log\ \frac{\mathcal{P}(X,Y|Z)}{\mathcal{P}(X|Z)\mathcal{P}(Y|Z)}$$

= I (X; Y|Z)

Which can also be written as follows:

$$\sum_{x\in X}\sum_{y\in Y}\sum_{z\in \mathbb{z}}\mathcal{P}(X,Y,Z)\log\frac{\mathcal{P}(Z)\mathcal{P}(X,Y,Z)}{\mathcal{P}(X,\ Z)\mathcal{P}(Y,Z)}$$

G. Normalized Information Distance (NID)

There are universal criteria for data types, and they are one of the main goals in clustering theory. The normal distance of information is a measure of the global distance for different types of objects. For the two strings X and Y, the NID criterion is defined as follows:

$$d(X,Y)=\frac{max\{\mathcal{K}(X|Y),\mathcal{K}(Y|X)\}}{max\{\mathcal{K}(X),\mathcal{K}(Y)\}}$$

Where K (x | y) is the length of the shortest program in which output X is obtained for input Y, and K (x) is the length of the shortest program in which output X is output for empty input. [10]

H. Vector Generation of an explicitly defined Multidimensional Semantic Space (VGEM)

VGEM is a combination of probabilistic and vector-based approaches that can be used to calculate the semantic similarity of words and even a group of words. This method can also be used for large datasets and there is no limit to the size of the word store used.

Similar to all measurements based on the VGEM vector, it defines the degree of similarity between two words, and this value is equal to the cosine of the angle between the two vectors. Words with a cosine close to zero are less similar, and for the closer the cosine is near to one. A value of 1 means that the words have the same meaning.

The main advantage of VGEM over semantic measurement of probabilistic relations is that it can also calculate the relationship between several words. A semantic measurement of probabilistic relations cannot find similarities between two paragraphs. VGEM, like other vector-based measures, can simply represent a paragraph or even an entire document as a vector, and then compare vector to vector in that semantic space. [11]

Disadvantages of VGEM include the need to pre-process the vocabulary. On the other hand, semantic measurement of relationship cannot analyze a very large set of texts. Also, this method is not suitable for analyzing texts that are constantly changing, even adding a single word may require reprocessing the whole text.

I. Normalized google Distance (NID)

Words and phrases acquire their meanings through their use in society . Each word has a relative meaning against other words and phrases. For computers, "database" is equivalent to the word "community", and "database search method" is equivalent to the word "use". Google similarity interval is a new theory of similarity between words and phrases based on information distance and Kolmogorov complexity. To consolidate and generalize this idea, the World Wide Web is used as a database, and Google as a search engine. This method also applies to other search engines and databases[12].

In defining the distance between the searched terms x and y according to Google Normal Distance (NGD) we have:

$$NGD(x,y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}}$$

Where f (X) represents the number of pages containing the occurrence of X and F (X, Y) represents the number of pages containing the occurrence of both x and y, and N represents the total number of web pages indexed by Google. We perform the calculations using the binary logarithm specified by "log".

J. SimRank

Similarity rank is a general measure of semantic similarity between objects that uses graph modeling theory to do so. This criterion can be generalized to any domain with object-to-object relationships. The similarity rank determines the structural similarity between objects based on their relationship to other objects. The main idea of the similarity rank states: "Two objects are similar if they are referenced by similar objects."

In other words, object “a” is considered similar to object “b” if they are referenced by objects “c” and “ d”, respectively, and “c” and “d” themselves are similar. This method can be used to determine the similarity between two texts, two web pages, two words and two paragraph. It is important to note that similarity rankings can be used for domains that are structured and have an acceptable level of object-distinguishable relationships[13].

Investigation of deep learning methods for semantic similarity

In recent years, deep learning has moved towards the introduction of words in vectors [14]. This scientific leap has made it possible to understand the real meaning of words with the help of deep neural networks with high accuracy in high dimensions. These algorithms started with word2vec and glove, and in recent years methods such as Elmo, Context2Vec, Bert, and Fast text have been added to previous methods. In this section, we will describe the proposed methods in-depth algorithms.

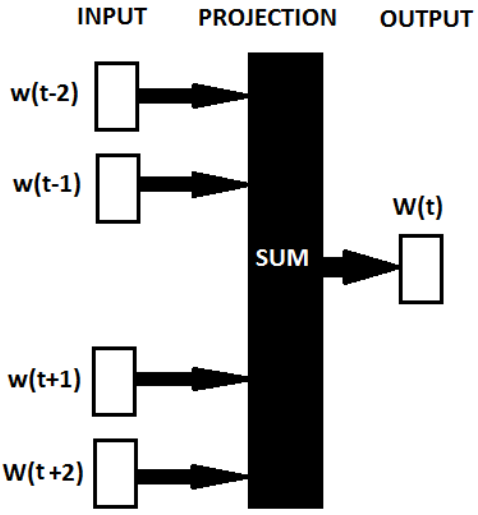
A. word2vec

The paper [15] presents two new architectures called Continuous Bag-of-Words and Continuous Skip-gram for calculating word vectors in a very large data set And the results are better performed based on the types of neural networks. Due to advances in similarity detection, the rate of increase in computational cost is very low [16]. For example, the proposed method has designed a high-quality word learning vector for 106 billion words in

less than a day. These vectors provide the state-of-the-art of the word in measuring the syntactic and semantic similarity of words.

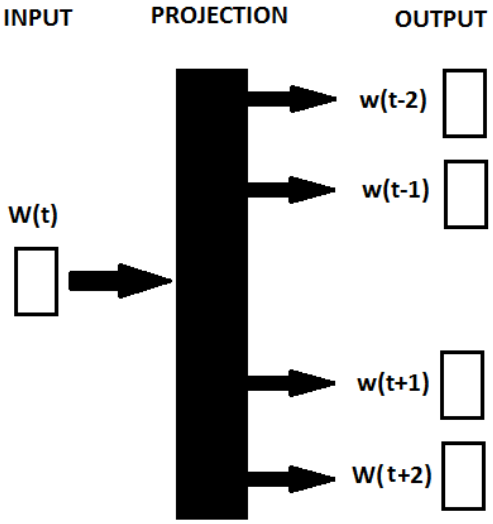
In the mentioned research, high quality is achieved in displaying words. Neural network training is done using the DesBelief method, although it is also possible to run several data training methods in parallel on the network. An example of a combination training method is the Adagrad method. In training, the priority is to choose the method with the least complexity. The language model can be taught in two steps. In the first step, the continuous word vector is taught using the simple model, and then the NNLM is taught the N gram mode on top of these word distribution displays.

The first proposed model, called the continues word bag, is similar to NNLM, in which the hidden layer is removed and the projection layer is shared for all words. This throws all the words into the same position. (Their vectors are averaged.) In this architecture, the current words are predicted according to the text. This architecture was proposed in 2013. In this architecture, called the word bag model, the order of the words has no effect on the passing time of the launch. (Memory has no effect.) This architecture works best with a logarithmic linear classification of four words at the input and one word at the output. Like NNLM, in CBOW the weight matrix between the input and projection layers are shared equally for all word positions. (Figure 1)



Continuous bag of words [15]

The second architecture proposed in this paper is the Continuous Skip-gram, which is similar to CBOW. In the previous model, the current word was predicted. But in this method, instead, we try to reach the meaning of the word based on the classification of words in similar sentences. That is, we consider each current word as an input for a logarithmic linear classification through the projection layer, and the prediction is based on a definite range before and after the word in the category. This study showed that increasing this definite range improves the quality of word vectors, but it also increases computational complexity. The words that are farther away from the current word have less to do with the word. (Figure 2)



Continuous Skip-gram [15]

Article [17] uses a combination of Word2Vec and SVM in sentiment analysis. The test results show that adding SVM can perform the task of classifying sentiment better than using any other model. However, the Word2vec model alone has the lowest accuracy (only 0.70).

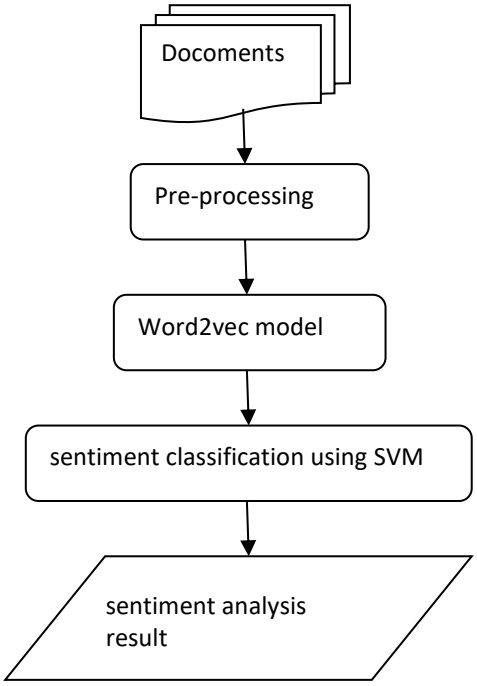
Figure 3 shows the Word2Vec combination model for the article support vector machine [17]. Pre-processing is done before starting the main process. Some of the steps performed in this step are tokenize and remove the excess characters. After performing the preprocessing step, the word vector is created using Word2Vec. Word vectors are used as classification features. Word2Vec can generally help improve classification performance, because in Wor2Vec similar words have similar vectors. Finally, in the last step, the reviews are classified into positive or negative categories.

B. Glove

In the semantic vector space model, each word is represented by a real vector, which has many applications in document classification, answering questions, and information retrieval. Angle between two word vectors is the main quality method in determining the distance between two words. In the GLOVE method,

the words are embedded based on a previous coding. For example, the word king is indicated in coding with man and queen with woman. This method is obtained by generalizing local background window methods and latent semantic analysis. In GLOVE, statistical information is prepared in a more strong way than latent semantic analysis, and a more appropriate content is used to extract the statistics. This method is based on regression models. Data training is also done in GLOVE. This method is 75% more effective than LSA in reasoning (From whole to part) (analogy) and is better than LSA in detecting similarity. [18] combination of Word2Vec and SVM [17]

In [19], in-depth learning is implemented by embedding modeling to analyze sentiment from online perspectives. Opinions about the types of products and restaurants have an important impact on decision making. This article uses Yelp Restaurant Reviews to predict the content of the reviews. This article uses GloVe, Word2vec, and CNN to implement the proposed emotion analysis system.



C. Fast text

In the past decades, the approach to sentiment analysis has been to use traditional classification. Deep learning models have recently achieved remarkable results when applied to computer diagnostics and speech recognition as well as natural language processing (NLP). Using word vectors is essential to represent words presented in a vocabulary in a dense format. These displays are usually called word embedding. Although not a new idea, word embedding became popular in 2013 after the release of pre-trained Word2Vec by Google. After that, other words such as fast text appeared. [20]

Article [21] performs linear analysis and optimization of text classification with Fast Text. This paper demonstrates that simple linear classifiers can compete with complex deep learning algorithms in text classification programs. Vocabulary and linear classification techniques are as accurate as or only slightly lower than fast text in deep learning algorithms, although this method is faster. Fast text can be turned into a simple equivalent classifier.

The exceptional performance of Fast Text is not surprising. This is the result of a very simple classification algorithm and its very professional execution in C++. The high accuracy of the very simple Fast Text algorithms clearly indicates that the problem of text classification is not yet sufficiently understood. Due to the great complexity of nonlinear classification models, their direct analysis is very difficult. A good understanding of simple linear classification algorithms such as Fast text is the key to building good nonlinear text classifiers. This was the main motivation for analyzing the Fast Text classification algorithm.

D. Elmo

Embedding Elmo is a difficult NLP architecture. In the two-way Elmo model, three 1024 dimensional vectors are generated in each step for one sentence. Article [22] suggests that they learn with a specific weight. However, this is a proposed weight scheme and is not practical for some tasks, and as shown, it does not necessarily perform optimally. This paper concludes that using hole third layer of the published language model often reduces performance. By performing the weight learning process, it is concluded that only the first and second layers are able to improve performance.

The article [23] analyzes the text summarization. This article uses the CNN dataset. The amount of data produced by humans is very large. people generates about 2 quintiles of unstructured data every day, and that number is expected to grow. With such a huge amount of information, the need for fast access and extensive understanding of textual data is becoming more and more. Automatic text summarization is a powerful tool that can be a useful solution to this problem.

Article [23] code has been implemented by crawling news articles from cnn.com and integrating important news points to form a polynomial summary. The summary consists of 300,000 samples. The second is the recently published Newsroom dataset, which contains 1.3 million article summaries.

Article [23] used the "Original (5.5B)" Elmo model. Bilateral LSTM with 256 neurons was also used. It should be noted that text summarization is one of the most complex tasks of text mining and to date has not been very accurate. At best, with other algorithms, this was done with 43% accuracy. But in the current article, this was done with an accuracy of 16 to 38 percent, which shows that Elmo has not been able to solve the problem of summarization at least 50 percent to date.

E: Bert

Burt algorithm is one of the algorithms that have paid more attention to the content. Panda2011, Penguin 2012, Hummingbird 2013, Mobile 2015, RankBrain 2015, Medic 2018 and Bert 2019 are some of the latest deep learning algorithms that Google has used for its search engine.

Article [24] is an experimental study of multitasking in Bret to extract biomedical text. Deep learning in natural medical and clinical language is useful for things like identifying and normalizing existence and extracting relationships. Extensive empirical studies have been conducted on evaluation criteria in biomedical understanding.

The article [25] examines the application of the Bret algorithm in text classification. Text classification is a classic issue. To classify text and other NLP tasks, we can avoid teaching a new model and move forward using pre-trained word embedding models (CoVe, GloVe, word2vec, Elmo, and Bert).

Introducing the best deep learning methods to simplify scientific texts

One of the reasons for the decrease in the readability of texts is their lexical complexity. Simplification can greatly help increase readers understanding and increase human reading and knowledge. Scientific texts often fall into the category of complex texts due to their complex vocabulary. [26]

A. Research experiments

Based on the article [26] complex texts can be created for reasons such as the presence of pronouns, words with low repetition frequency or long sentences. Article [26] examines the discussion of low-frequency words, which is one of the most important reasons for complexity. Then, by presenting an algorithm, it has introduced how to recognize words based on the complexity of frequency and character.

Accordingly, we selected a series of highly complex sentences. These two sentences have relatively the same meaning. These pear to pear corresponding sentences with similar concepts have been shown in table 1.

TABLE I. Ten pear to pear corresponding sentences with similar concepts

A1	The new strategy of the OPEC summit is to use all the capacities of Iran, Oman, Qatar, Kuwait and Saudi Arabia.
A2	The newly policy of Organization of the Petroleum Exporting Countries is to use the full potential of the Persian Gulf countries.
B1	World War II veteran rescued from Corona
B2	A US federal soldier defeats Quid 19 in Second World War
C1	The Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field Supports victims of armed conflict.
C2	The First Geneva Convention covers military and civilian casualties of war.
D1	The Gulag Archipelago is about the repressive Soviet regime
D2	An Experiment in Literary Investigation by Solzhenitsyn deals with the collapse of former socialist Russia
E1	Round robin tournament is a competition in which each contestant meets all other contestants in turn.
E2	All play all tournament is an elimination tournament, in which participants are eliminated after playing with all the players.

For each sentence, the corresponding vector Glove, Word2vec, and Elmo are created. Corresponding sentences have a relatively similar percentage of complexity. Finally, the similarity of the sentences is examined in a peer-to-peer manner. The more similar algorithm is the best one to identifying complex

concepts in the texts. Table 2 shows the similarities of the corresponding vectors in the different algorithms.

Table II. evaluate the similarity of sentences with the same meaning

Similarity distance / Pear to pear sentences	Word2vec	Glove	Elmo
A1,A2	0.71	0.76	0.80
B1,B2	0.67	0.65	0.64
C1,C2	0.71	0.79	0.81
D1,D2	0.61	0.77	0.76
E1,E2	0.51	0.56	0.74

The results of the experiment in the complexity discussion of 5 pairs of highly similar sentences, often with a same percentage of complex words, show that the Elmo algorithm in the experiment detects an average of 6% more similarity complex texts than Glove. Glove also detects an average of 3% better complexity than Word2vec.

B. Analysis of text complexity

The paper [27] provides a general framework for text simplification systems. An important point made in that article is that in the simplification process, although tasks such as deleting passive sentences or reducing sentence length can be helpful, but the most effective factor in creating a simple text is the use of more simple words.

Among the introduced algorithms for recognizing lexical similarity and lexical substitution, Elmo will be the most efficient and appropriate algorithm for simplification. The main reason is that due to the ignorance of most words in complex texts, some algorithms, such as glove or GLSA, which are LSA-based and analyze words based on previous experience, have difficulty recognizing new and complex words. word2vec also has the problem of identifying new words.

Fast text, although is fast, still offers a solution for speed, and this solution is not to solve the problem of new words. Among the introduced algorithms, Elmo is an algorithm that can correctly identify new words by analyzing characters and processing characters. Unlike other algorithms, this algorithm is in terms of characters, which makes it possible to represent new and unseen words in terms of structural form.

Articles [28], [29], [30] and [31] are examples of recent simplification research, most of them have used Elmo. Of course, a study of this series of articles shows that Bret can also be effective in recognizing unknown words. As mentioned in the previous section, Elmo also has good results in text summarizing.

The Elmo algorithm was developed by Allen in 2018 and its method is different from the old and traditional algorithms. This algorithm uses deep and two-way LSTM. Unlike other algorithms that build a dictionary for words and assign a vector to each word, Elmo analyzes the words in the context in which they are used.

Despite the rapid development of sentence embedding methods in recent years, it is challenging to provide a comprehensible assessment. The article [32] deals with the evaluation of sentence embedding methods in linguistic exploration issues with Elmo. Over the years, significant improvements have been made in the development of sentence analysis. In this paper, a conceptual evaluation of recent methods has been performed using a variety of linguistic exploration issues such as exploration tasks.

Article [33] shows that the more we move towards the hidden inner (upper) layers, the more complex semantic information will be represented. In the following, they will introduce the architecture of each layer from the lowest to the highest levels. This model represents and adapts the input sentence words to each other in three different ways: Glove, context Elmo, and convolutional neural network coding.

Conclusion

The expansion of cyberspace has increased the volume of waste data. Such an issue has caused people to turn to insignificant information instead of gaining knowledge due to the attractiveness of the news. Knowledge often has a complex conceptual structure in scientific and academic texts. For this purpose, it is necessary to carry out a process to control knowledge texts before publication and to create a requirement to publish knowledge in simple language. Creating software can be useful in this field.

Currently, machine learning algorithms are very popular in simplification studies. Language has a complex structure. The letters are combined in different order to form completely different words, and sometimes one word has several meanings. Combining words and

sentences in different order can have completely different moods and make a difference in meaning.

Using word embedding, semantic and syntactic information of words can be received from a large number of unlabeled data. Word embedding has been used in many works in natural language processing to produce more effective word representation. The issue addressed in this study was which of the Elmo, FastText, GloVe, Word2Vec, and Bert methods can be of great help in simplifying text research?

Studies have shown that Elmo can be more helpful to text simplification programmers in recognizing the equivalent of complex words due to their greater familiarity with unfamiliar words and working with characters. On the other hand, Bert has had relatively good results in recent research and has been able to become the algorithm used by Google in 2019. Therefore, these two algorithms can have better results in simplifying or summarizing complex texts, especially scientific texts.

References

[1] Volodymyrovych TY, Tetiana K, Yaroslavovych TB. Distance learning technologies in online and mixed learning in pre-professional education of medical lyceum students. J Adv Pharm Educ Res. 2021;11(4):127- 35.

[2] Mekaeil NST. Investigating the relationship between learning anxieties and learning strategies used by EFL teachers and students. J Adv Pharm Educ Res. 2022;12(4):77-81.

[3] Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. International Journal of Computer Applications, 68(13), 13-18.

[4] Gabrilovich, E., & Markovitch, S. (2009). Wikipedia-based semantic interpretation for natural language processing. Journal of Artificial Intelligence Research, 34, 443-498.

[5] Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1-2), 177-196.

[6] Matveeva, I., & Levow, G. A. (2006, June). Graph-based generalized Latent Semantic Analysis for document representation. In Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing (pp. 61-64).

[7] Rossi, F., Lendasse, A., François, D., Wertz, V., & Verleysen, M. (2006). Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. Chemometrics and intelligent laboratory systems, 80(2), 215-226.

[8] Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL, 31-40.

[9] Van de Cruys, T. (2011, June). Two multivariate generalizations of pointwise mutual information. In Proceedings of the Workshop on Distributional Semantics and Compositionality (pp. 16-20).

[10] Terwijn, S. A., Torenvliet, L., & Vitányi, P. M. (2011). Nonapproximability of the normalized information distance. Journal of Computer and System Sciences, 77(4), 738-742.

[11] Grintsvayg, A., Veksler, V. D., Lindsey, R., & Gray, W. D. (2007). Vector Generation of an Explicitly-defined Multidimensional Semantic Space. In Proceedings of ICCM Eight international conference on Cognitive Modeling, Oxford, UK (pp. 231-232).

[12] Cohen, A. R., & Vitányi, P. M. (2013). Normalized google distance of multisets with applications. arXiv preprint arXiv:1308.3177.

[13] Jeh, G., & Widom, J. (2002, July). SimRank: a measure of structural-context similarity. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 538-543).

[14] Adiga R, Biswas T, Shyam P. Applications of Deep Learning and Machine Learning in Computational Medicine. J Biochem Technol. 2023 Jan 1;14(1)1-6.

[15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[16] Tetiana, R., Inna, K., Iryna, N., Natalia, S., Liudmyla, K., Olexandr, B. and et al. Digital component of professional competence of masters of pharmacy in the framework of blended learning. Arch Pharm Pract 2021;12(1):98-102

[17] Fauzi, M. A. (2019). Word2Vec model for sentiment analysis of product reviews in Indonesian language. International Journal of Electrical and Computer Engineering, 9(1), 525.

[18] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[19] AlSurayyi, W. I., Alghamdi, N. S., & Abraham, A. Deep Learning with Word Embedding Modeling for a Sentiment Analysis of Online Reviews.

[20] Santos, I., Nedjah, N., & de Macedo Mourelle, L. (2017, November). Sentiment analysis using convolutional neural network with fastText embeddings. In 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-5). IEEE.

[21] Zolotov, V., & Kung, D. (2017). Analysis and optimization of fasttext linear text classifier. arXiv preprint arXiv:1702.05531.

[22] Reimers, N., & Gurevych, I. (2019). Alternative weighting schemes for elmo embeddings. arXiv preprint arXiv:1904.02954.

[23] Mastronardo, C., & Tamburini, F. (2019). Enhancing a Text Summarization System with ELMo. In CLiC-it.

[24] Peng, Y., Chen, Q., & Lu, Z. (2020). An Empirical Study of Multi-Task Learning on BERT for Biomedical Text Mining. arXiv preprint arXiv:2005.02799

[25] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. In China National Conference on Chinese Computational

[26] Mirzaei, B., Jafarabad, M., Karimi, H., Khastavaneh, M. J., Parveh, A., Dehno, S. B., ... & Safari, A. (2015, December). The effect of simplification based on words and their role in approving or rejecting articles. In 2015 IEEE International Conference on Progress in Informatics and Computing (PIC) (pp. 71-76). IEEE.

[27] Khandelwal, M., & Jafarabad, M. SOFTWARE FEASIBILITY STUDY TO TRANSFORM COMPLEX SCIENTIFIC WRITTEN KNOWLEDGE TO AClear, RATIONALE AND SIMPLE LANGUAGE.

[28] Maruyama, T., & Yamamoto, K. (2020). Extremely Low-Resource Text Simplification with Pre-trained Transformer Language Model. International Journal of Asian Language Processing, 30(01), 2050001.

[29] Scarton, C., Madhyastha, P., & Specia, L. (2020). Deciding when, how and for whom to simplify. In ECAI 2020 (Vol. 325, pp. 2172-2179). IOS Press.

[30] Gupta, H., & Patel, M. (2020, October). Study of Extractive Text Summarizer Using The Elmo Embedding. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 829-834). IEEE.

[31] Gooding, S., & Kochmar, E. (2019, November). Recursive Context-Aware Lexical Simplification. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 4855-4865).

[32] Perone, C. S., Silveira, R., & Paula, T. S. (2018). Evaluation of sentence embeddings in downstream and linguistic probing tasks. arXiv preprint arXiv:1806.06259.

[33] Sanh, V., Wolf, T., & Ruder, S. (2019, July). A hierarchical multi-task approach for learning embeddings from semantic tasks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 6949-6956).Rossi, F., Lendasse, A.