Eva Lucero Pérez Salcedo A01568830
7th Semester
Artificial Emotions:Going beyond Artificial Intelligence
Instituto Tecnologico de Monterrey campus Guadalajara
Course provided by: Mahdi Zaarei and Alejandro de León Languré
25th September 2023

Word Embeddings for the Optimization of Feedback Comments

ABSTRACT
This paper examines how word embeddings have a specific role in methods where artificial intelligence can make the most of it while improving NLP algorithms in learning words and their patterns in training. Considering their importance, a subject to cover and improve is related to effective feedback in certain services or products, here is where word embeddings can facilitate information, improvement and constructive feedback to professional and even personal growth. It's purpose to approach the quality of feedback comments through the use of word embeddings can benefit from analyze and optimize feedback comments automatically where words and phrases can be more contextually relevant and understood in language processing. It is also addressed the difficulties and limitations including the memory capacity in some computational training. Our results show when properly implemented, word embeddings can contribute to more depth and quality from textual feedback transformed into numerical vectors having more effective decisions and recommendation to future compilation of information from the users. To summarize, this paper goes through the potential of word embeddings in feedback analysis making a more robust understanding of the text and contribute to better services, products and development in artificial intelligence.

INTRODUCTION
Word embeddings are a simple way to represent words and documents, its also seen as a word vector, is a numeric input representing a word in a lower-dimensional space, this way words with similar meanings have a similar representation inside a sentence, being approximated to understand the meaning of the word in particular. They are mostly used as input in machine learning models as well as their training to represent their numeric patterns for their learning. It's commonly used pre-trained models for this such as SpaCy and fastText. This information received in the converted form, is used by Natural Language Processing (NLP) algorithms to perform more easily this information, learning their patterns

and processing their textual meaning. With this summary of their structure and meaning, this next sections will provide a more complete understanding of word embedings, its techniques and applications in fields like processing feedback comments from users in a platform.

Word embedding

It is defined as a numeric vector input that lets words with similar meanings to have the same representation, they can help to understand or approximate to the meaning with words in a lower dimensional space. Its more easily built and faster with models that have graph embeddings like WordNet, for this exact reason pre-trained word embedding models are the best way to take advantage of this approach. It all started with the issue in deep learning and machine learning, it’s not always possible to take the information as a whole word letter by letter, for the computers it’s easier to take this information with numerical inputs, that’s why word embeddings are the solution in this case scenario where a lower-dimensional space captures inter-word semantics. “Each word is represented by a real-valued vector with tens or hundreds of dimensions.” [2]

Word2Vec

This method was developed in 2013 by Google, invented for training word embeddings and is based on a distributional hypothesis where it uses skip-grams or a CBOW. Here a vector is assigned to every word, it can begin with a random vector or one-hot vector. They have three layers, input, output and projection. It considers linguistic contexts by their order and history to have a pattern in the future. [1]

1. *Continuous Bag of Words (CBOW)*
   In CBOW, the model predicts the current word based on its context.
   The hidden layer is computed as the average of the input context word vectors, as well as the output. Where this formula takes part.

   $h = 1/C \sum_{i=1}^{C} v_{w_i}$

   $P(w_O | w_{I_1}, w_{I_2}, \ldots, w_{I_C}) = e^{u_{w_O} h} / \sum_{w=1}^{W} e^{u_w h}$

   - v_{wi} are the input vectors for the context words.
   - u_{wO} are the output vectors for the vocabulary words.
   - h is the hidden layer.
   - C is the number of context words.

- W is the number of words.

GloVe:Global Vector for word representation

This is another method for word embeddings, its done by iteration of the words, it was developed by Stanford by Pennington, et al. It has great performance in world analogy, it reduces the computational cost of training the model for the reason that it has simpler error function that has better findings in different word embeddings. [3]

> *"So, unlike Word2Vec, which creates word embeddings using local context, GloVe focuses on global context to create word embeddings which gives it an edge over Word2Vec. In GloVe, the semantic relationship between the words is obtained using a co-occurrence matrix." -Turing Knowledge Base.*

The function of GloVe is defined in the next formula:

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T w_j + b_i + b_j - log(X_{ij}))^2$$

- V is the size of the vocabulary.
- X_{ij} is the number of times word i occurs in the context of word j.
- w_i and w_j are the word vectors for words i and j.
- b_i and b_j are terms for i and j.
- $f(X_{ij})$ is a weighted function that helps with rare occurrences.

Implementation on Feedback comments

With the previous information, it's seen that word embeddings can improve the analysis and interpretation of feedback by understanding more deeply the context of the words used from the training on neural networks. They make sure that similar comments are recognized, even if the words used are different, leading to a better analysis of user input. This technology has a benefit in generating accurate and automatic replies, and for helping recommendation systems provide relevant suggestions. It can take a variety of words and their synonyms, ensuring a robust analysis of feedback. Word embeddings are also crucial in figuring out the feelings and opinions expressed in the comments, and in extracting the specific aspects of a product or service being discussed. Also, when combined with other language processing models, they enhance the performance of various text analysis tasks. In summary, using word embeddings in feedback comments makes a better customer support, helps in product development, and improves communication between users and systems, all of this while maintaining a highly informational approach. [4],[5]

Results

In this research, we discovered that using word embeddings to analyze feedback comments makes the system smarter and faster in understanding what people say. More specifically, models like Word2Vec and GloVe made it easier for the system to get the context, leading to a better and more detailed analysis. We found a increase in accuracy when sorting feedback into different feelings or topics, and the system responded quicker. The answers given were more accurate and fit the context better, and the system could understand different words that meant the same thing, making the feedback analysis more complete. However, on the drawbacks, we also noticed that these word embeddings take up a lot of computer memory and take longer to train. Looking forward, these results show that there's a lot of potential to make feedback analysis even better, by finding ways to train models more efficiently and integrate more advanced language understanding methods, while still being careful about computer resources. In the end, our work confirms that word embeddings are really important for making feedback analysis better, leading to benefit users and a more effective system. [6]

CONCLUSION

In conclusion, this essay explores how word embeddings, a type of artificial intelligence and language understanding tool, play an important role in making sense of feedback comments. By converting words into numbers, word embeddings help machines to better grasp and work with human language, ensuring that similar comments are accurately identified, no matter the choice of words. Techniques such as Word2Vec and GloVe have shown great effectiveness in understanding the context and subtle meanings in language, resulting in a better analysis of user feedback.

The advantages of using word embeddings are evident, including improved accuracy in determining the mood or topic of feedback, faster response times, and more fitting recommendations. However, it's also important to note the challenges, such as the need for substantial memory and long training periods. Despite these issues, the potential of word embeddings to transform feedback analysis, leading to improved customer support, product development, and communication between users and systems, is uncomparable. Looking forward, efforts should be made towards making model training more efficient and innovating more advanced methods of understanding language, all while being aware of the computer resources used. This leads to improved support for customers and better development of products and services. Overall, the work we've done shows that word embeddings are really important for making sense of user feedback and improving artificial intelligence tools.

REFERENCES

[1] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

[2] "Guide on Word Embeddings in NLP," Turing Knowledge Base, Sep. 28, 2023. [Online]. Available: https://www.turing.com/kb/guide-on-word-embeddings-in-nlp

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

[4] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[5] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

[6] Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing. Draft of the 3rd edition.

[7] F. Almeida and G. Xexeo. “Word Embeddings: A Survey.” arXiv.org e-Print archive. Accessed: Sep. 28, 2023. [Online]. Available: https://arxiv.org/pdf/1901.09069.pdf

[8] A. Bakarov. “A Survey of Word Embeddings Evaluation Methods.” arXiv.org e-Print archive. Accessed: Sep. 28, 2023. [Online]. Available: https://arxiv.org/pdf/1801.09536.pdf

[9] T. Schnabel, I. Labutov, and D. Mimno. “Evaluation methods for unsupervised word embeddings.” Cornell University. Accessed: Sep. 28, 2023. [Online]. Available: https://aclanthology.org/D15-1036.pdf