



Listes de contenus disponibles sur: [Scholar](#)

Comparative Analysis of Word2Vec and GloVe with LSTM for Sentiment Analysis: Accuracy and Loss Evaluation on Twitter Data

Journal homepage: ijssass.com/index.php/ijssass

Comparative Analysis of Word2Vec and GloVe with LSTM for Sentiment Analysis: Accuracy and Loss Evaluation on Twitter Data [☆]

Joshua Sopuru ^a, Adah Alubo ^b, Princess Chinemerem Iloh ^c, Oluwaseun Augustine Lottu ^d

- A. Assistant Professor Dr. Girne American University, Cyprus
 B. University of Hull UK, Data Science and artificial Intelligence
 C. Teesside university UK
 D. Stanmore College IT and Engineering Lecturer, UK

Received 23 August 2023; Accepted 10 December 2023
 Available online 5 January 2024

ARTICLE INFO

Keywords:

Sentimental analysis
 Word2Vec
 Artificial intelligence
 GloVe
 LSTM
 Twitter dataset

ABSTRACT

Artificial Intelligence (AI) is witnessing an increase in textual data from diverse sources such as social media, online reviews, and blogs. This textual data, rich in sentiments and emotions, has become a valuable asset for understanding public opinion and societal trends. Conventional sentiment analysis methods, relying on lexicon-based approaches and machine learning models, faced challenges in handling linguistic subtleties and contextual nuances. The advent of deep learning, particularly Long Short-Term Memory (LSTM) architecture, has revolutionized sentiment analysis by enabling automated pattern extraction from raw textual data. This article investigates the efficacy of Word2Vec and GloVe models in combination with LSTM for sentiment analysis using a Twitter dataset.

I. Introduction

In the context of the rapidly expanding field of Artificial Intelligence (AI), the surge in digital content generated via social media, online reviews, and blogs has ushered in an era of unprecedented textual data accumulation, capturing a diverse range of sentiments and emotions (Annett & Kondrak, 2008). This evolving textual corpus offers a unique avenue to fathom public opinion, customer feedback, and societal trends. Within Natural Language Processing (NLP), sentiment analysis is a pivotal subfield, seeking to computationally decode the Sentiment or emotional tenor embedded within the textual content. Its relevance spans business, marketing, politics, and the social sciences, magnifying its significance in today's data-driven landscape (Sopuru & Akkaya, 2019).

Conventional sentiment analysis methodologies heavily leaned on lexicon-based approaches and machine learning models fortified with manually crafted features (Araque et al., 2017). However, these methodologies frequently grappled with the intricacies of linguistic subtleties, contextual nuances, and the ever-shifting dynamics of language evolution. The emergence of deep learning has engendered a paradigm shift in sentiment analysis (Minaee et al., 2021). This new approach enables the automated extraction of intricate patterns and representations directly from raw textual data, surmounting several constraints imposed by conventional methods.

This article embarks on a targeted investigation, aiming to juxtapose the efficacy of Word2Vec and GloVe models in conjunction with the Long Short-Term Memory (LSTM) deep learning algorithms. This exploration is conducted utilizing a dataset drawn from the realm of Twitter. The ensuing analysis involves a comprehensive evaluation grounded in diverse performance metrics. Ultimately, the insights from this comparative analysis will serve as a bedrock for offering informed recommendations, specifically concerning

applying these algorithms within sentiment analysis.

Background and Motivation

Sentiment analysis has a rich history dating back to the early days of NLP research. Initial approaches employed rule-based methods that matched words or phrases in a text against predefined sentiment lexicons (Singh & Paul, 2021). While these approaches were simplistic and lacked context awareness, they laid the foundation for subsequent advancements. Machine learning methods, mainly supervised classification algorithms, gained traction as larger labeled datasets became available. These approaches utilized engineered features such as bag-of-words representations, n-grams, and hand-crafted linguistic features. However, the effectiveness of these methods was often limited by their reliance on shallow, human-crafted features that needed to be more robust to capture the intricate relationships present in natural language.

Deep learning, a subset of machine learning that involves training complex neural networks (Chassagnon et al., 2020), emerged as a transformative paradigm for various NLP tasks. In the context of sentiment analysis, deep learning models, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and more recently, transformer-based architectures like BERT and GPT, have shown remarkable performance gains (Topal et al., 2021). These models can automatically learn hierarchical representations of text, enabling them to capture intricate dependencies and contextual information, thus overcoming the limitations of hand-crafted features.

Scope and Aims

The primary objective of this study is to evaluate and contrast the performance of Word2Vec and GloVe methodologies in constructing word embeddings for sentimental analysis. Employing LSTM neural networks, we empirically assess the effectiveness of these methods using a Twitter dataset. Our specific aims encompass

measuring the proficiency of each method in capturing word semantics within the dataset, identifying contexts where their strengths lie, and delineating potential constraints in their applicability. This investigation aims to understand the comparative capabilities and limitations of Word2Vec and GloVe techniques in a time-bound manner.

II. Literature Review: Deep Learning Approaches for Sentiment Analysis - Word2Vec and GloVe

The emergence of deep learning techniques has revolutionized the field of sentiment analysis, enabling the automatic extraction of nuanced patterns and representations from raw textual data. In this literature review, we focus on two widely adopted deep learning approaches, Word2Vec and GloVe, which have significantly contributed to enhancing the accuracy and contextual understanding of sentiment analysis tasks.

Word2Vec: Distributed Word Representations

Word2Vec, introduced by Mikolov et al. (2013), is a milestone in word embeddings. It is based on the idea that words with similar meanings often appear in similar contexts. It aims to map words into continuous vector spaces so that their spatial relationships reflect their semantic meanings. This approach fundamentally changed the representation of words from sparse, high-dimensional one-hot encodings to dense, continuous-valued vectors.

The two primary architectures of Word2Vec are the Continuous Bag of Words (CBOW) and the Skip-gram model. CBOW predicts a target word from its surrounding context words, while Skip-gram predicts surrounding context words from a target word. These models effectively capture semantic relationships between words, enabling them to encode semantic similarities and analogies. This has proven beneficial for sentiment analysis as it helps models identify contextually similar words and phrases associated with different sentiments.

Researchers have integrated Word2Vec embeddings into sentiment analysis pipelines, achieving substantial performance improvements. Acosta et al. (2017) utilized pre-trained Word2Vec embeddings as input to a convolutional neural network (CNN) for sentence classification tasks. The model demonstrated strong performance on various sentiment analysis datasets, showcasing the effectiveness of Word2Vec embeddings in capturing contextual information critical for sentiment interpretation.

Word2Vec and LSTM

Numerous researchers have employed classification models based on Word2Vec and LSTM in their studies. For instance, Xiao, Wang, and Zuo (2018) developed an innovative patent text classification model that integrates Word2Vec and Long Short-Term Memory (LSTM) to categorize patent texts within the security domain efficiently. Their model capitalizes on patent text attributes, employing a multifaceted strategy. In their work, Initially, common terms like "the invention," "involvement," and "utility model" were added to the stop word list during text preprocessing, aiming to optimize storage and processing efficiency. Moreover, integrating a pre-trained Word2Vec model played a pivotal role in addressing the dimensionality challenges often encountered in conventional methodologies.

The pivotal phase of this approach encompassed the training of the LSTM classification model, facilitating the extraction of significant textual features for patent text categorization. Their dataset of 50,000 patent documents was divided into training and test sets to gauge model efficacy, maintaining a 4:1 ratio. Performance evaluation metrics, encompassing accuracy and the Receiver Operating Characteristic (ROC) curve, were employed to assess classification outcomes. Impressively, the proposed method yielded a classification accuracy rate of 93.48%.

Furthermore, they conducted a comprehensive

comparative analysis to benchmark the LSTM classification model against alternative techniques. These alternatives encompassed the K Nearest Neighbor (KNN) classification model, the Convolutional Neural Network (CNN) classification model, and hybrid models that amalgamate CNN and Word2Vec. Their results underscored the superiority of the proposed approach in effectively categorizing patent texts within the security domain.

GloVe: Global Vectors for Word Representation

Global Vectors for Word Representation (GloVe), introduced by Pennington et al. (2014), is another influential approach for learning word embeddings. GloVe leverages both global and local word co-occurrence statistics to construct word embeddings. It captures the relationships between words by optimizing a global word-word co-occurrence matrix to generate embeddings that reflect semantic similarities.

GloVe embeddings have shown remarkable results in sentiment analysis due to their ability to capture subtle semantic relationships. Various studies have employed GloVe embeddings in different architectures. Sachin et al. (2020) combined GloVe embeddings with a recursive neural network (RNN) for sentiment classification, achieving state-of-the-art performance on multiple benchmark datasets. The combination of GloVe embeddings' ability to capture global semantic information with the sequence modeling capacity of RNNs proved highly effective.

GloVe and LSTM

Similar to the works of Xiao, Wang, and Zuo (2018), Vijayvergia and Kumar (2021) introduced an innovative approach to emotion detection that leverages selective integration of strengths from shallow models. This methodology integrates the GloVe word model and LSTM architecture, demonstrating its effectiveness in improving real-time application performance. Their proposed model achieved 86.16% accuracy in 00.98

milliseconds per input. Therefore, the experiments show that the proposed model outperforms state-of-the-art models. Moreover, the computational cost shows that the proposal may be used for real-time applications.

Comparative Analysis

Based on previous research, Word2Vec and GloVe have their strengths for understanding sentiments in text. Word2Vec is a good fit for understanding short pieces of text like tweets and reviews because it focuses on how words are used together. On the other hand, GloVe looks at more significant patterns in how words are used, which makes it better for longer and more complicated texts.

This study seeks to determine which of these models works better for analyzing sentiment in tweets when combined with LSTM. To do that, we are asking two questions:

Research Questions

- A. Which model, when combined with LSTM has better accuracy?
- B. Which model, when combined with LSTM, has more loss?

III. Methodology

Dataset

The dataset used in this study was sourced from a Twitter chat to assess sentiments available on Kaggle. The dataset comprises 75,682 records, each with three independent variables: ID, Company, and Message. The primary dependent variable of interest is the Sentiment associated with each message.

Data Preprocessing

Before conducting the sentiment analysis, the dataset underwent preprocessing to ensure its suitability for machine learning. The preprocessing steps included tasks such as text normalization, tokenization, and removal of special characters and stop words. This process aimed to clean and standardize the textual data, enhancing the quality of subsequent analyses.

Feature Description

ID: A unique identifier assigned to each record, enabling traceability and reference.

Company: The entity or organization associated with the message, providing context to the message content.

Message: The textual content of the message extracted from the Twitter chat. This is the primary input for sentiment analysis.

Dependent Variable

The Sentiment variable serves as the dependent variable for this study. It represents the Sentiment expressed within each message and is categorized into classes such as positive, negative, or neutral. The sentiment labels provide the ground truth for training and evaluating the sentiment analysis models.

Word Embeddings:

Two prominent word embedding techniques, Word2Vec and GloVe, will be utilized to convert the tokenized words into meaningful vectors.

Word2Vec:

The Word2Vec algorithm creates word embeddings by learning the context of words in a corpus. The Skip-gram or Continuous Bag of Words (CBOW) architectures will be employed to train the Word2Vec model. The formula for calculating the probability of a context word given a target word in Skip-gram is:

$$P = (\omega_{context} | \omega_{target}) = \frac{1}{1 + e^{-v_{context} \cdot v_{target}}}$$

GloVe:

The GloVe algorithm constructs word vectors by considering global word co-occurrence statistics. The model aims to minimize the difference between the dot product of word vectors and the logarithm of the co-occurrence probability. The loss function for GloVe is:

$$J = \sum_{i=1}^v \sum_{j=1}^v f(X_{ij})(v_i^T - \log X_{ij})^2$$

LSTM Model:

The Long Short-Term Memory (LSTM) architecture will be employed to incorporate sequence information and

capture long-term dependencies within the reviews.

IV. Experimental Setup:

The dataset will undergo division into training and testing subsets, with 98.69% of the records (74,682 out of 75,682) allocated for training, and the remaining 1.31% (1,000 out of 75,682) reserved for testing purposes. In pursuit of constructing distinct sentiment analysis models, both embedding models, namely Word2Vec and GloVe, will be united with the LSTM architecture. This integration will result in the development of two separate sentiment analysis models.

Evaluation Metrics:

Two crucial metrics will be employed to assess the performance of the models:

Accuracy: The proportion of correctly predicted sentiment labels over the total predictions. The formula for accuracy is:

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Total\ Number\ of\ Prediction} \times 100$$

Loss: The difference between the predicted sentiment values and the actual sentiment labels. This metric indicates how well the model's predictions align with the true data. The loss function for a binary classification task like sentiment analysis can be the binary cross-entropy loss:

$$Loss = \frac{1}{N} \sum_{i=1}^N Y_i \cdot (\log(Y_i) + (1 - Y_i) \cdot \log(1 - Y_i))$$

Where:

N = the number of samples.

Y_i = true sentiment label for the i th sample.

y_i = predicted sentiment value for the i th sample.

Analysis and Comparison:

The obtained accuracy and loss values from both Word2Vec-LSTM and GloVe-LSTM models will be compared to determine which combination performs

better for sentiment analysis on the Twitter dataset.

V. Results

A. Word2Vec on LSTM

The results of employing Word2Vec in conjunction with the LSTM architecture for sentiment analysis are presented in Table 1. This table showcases the loss and accuracy values achieved across the 10 epochs of training.

Table 1: Loss and Accuracy Results for Word2Vec on LSTM

Epoch 1/10	0.0456	1.0000
Epoch 2/10	1.8226e-04	1.0000
poch 3/10	4.0122e-05	1.0000
Epoch 4/10	.2042e-05	1.0000
Epoch 5/10	4.0084e-06	1.0000
Epoch 6/10	1.3974e-0	1.0000
Epoch 7/10	5.0107e-07	1.0000
Epoch 8/10	1.8454e-07	1.0000
Epoch 9/10	7.0455e-08	1.0000
Epoch 10/10	2.8330e-08	1.0000

Test Loss: 1.782346714662708e-08

Test Accuracy: 1.0

Across the epochs, it is evident that the loss values progressively decrease, converging to minuscule magnitudes. This signifies the model's consistent refinement in its predictive capacity as training advances. The accuracy values consistently remain at their peak of 1.0000, reflecting the model's ability to make accurate predictions.

The outcomes depicted in Table 1 underscore the proficiency of the Word2Vec-LSTM combination in

effectively learning and predicting sentiment patterns from the input data. The convergence of both low loss and high accuracy values substantiates the model's aptitude in capturing intricate relationships within the text data.

B. GloVe and LSTM

The results obtained by integrating the GloVe word model with the LSTM architecture for sentiment analysis are presented in Table 2. This table provides an overview of the loss and accuracy values achieved throughout the 10 epochs of training.

Table 2: Loss and accuracy Result for GloVe on LSTM

Epoch 1/10	0.0032	1.0000
Epoch 2/10	9.2801e-06	1.0000

poch 3/10	2.5660e-06	1.0000
Epoch 4/10	8.9359e-07	1.0000
Epoch 5/10	3.3573e-07	1.0000
Epoch 6/10	1.3049e-07	1.0000
Epoch 7/10	5.1895e-08	1.0000
Epoch 8/10	2.1154e-08	1.0000
Epoch 9/10	8.9301e-09	1.0000
Epoch 10/10	2.6776e-09	1.0000

Test Loss: 2.677595478672856e-09

Test Accuracy: 1.0

The progression of loss values across the epochs in Table 2 follows a pattern of consistent reduction, reaching remarkably diminutive values. This pattern signifies the model's ability to optimize its predictions in relation to the true sentiment labels as training unfolds. Simultaneously, the accuracy values maintain a constant peak of 1.0000, indicating the model's precision in making accurate predictions.

The findings presented in Table 2 underscore the competence of the GloVe-LSTM fusion in effectively capturing and comprehending sentiment nuances embedded within the input text data. The coalescence of minimal loss and perfect accuracy values further substantiates the model's proficiency in extracting meaningful insights from the text and utilizing them for accurate sentiment classification.

VI. Discussion of Results

Focusing on the performance of the Word2Vec-LSTM and GloVe-LSTM models in terms of accuracy and loss as seen in the results, we move to answer our research questions.

1. Which model when combined with LSTM has better accuracy?

Both the Word2Vec-LSTM and GloVe-LSTM models consistently exhibit perfect accuracy (1.0000) across all

epochs, as indicated by Tables 1 and 2. This implies that both models are remarkably adept at correctly classifying Sentiment within the provided data. The similarity in accuracy performance suggests that both embedding techniques, Word2Vec and GloVe, are effective in conjunction with LSTM for sentiment analysis on the given dataset.

2. Which model when combined with LSTM has more loss?

Comparing the loss values from Tables 1 and 2, it's apparent that the GloVe-LSTM model consistently yields lower loss values throughout the training process. The loss values for GloVe-LSTM are notably smaller than those for Word2Vec-LSTM. This indicates that the GloVe-LSTM model better aligns its predictions with the actual sentiment labels, resulting in fewer discrepancies between the predicted and actual values. Therefore, the GloVe-LSTM model incurs "more loss" in the sense that its predictions are closer to the ground truth data, indicating superior convergence.

In summary, both the Word2Vec-LSTM and GloVe-LSTM models demonstrate exceptional accuracy, suggesting their proficiency in capturing sentiment patterns. However, the GloVe-LSTM model outperforms the Word2Vec-LSTM model in terms of lower loss values, showcasing its ability to minimize discrepancies

between predictions and actual sentiment labels. This observation indicates that the GloVe-LSTM model is more precise in its predictive capacity, making it a stronger candidate for sentiment analysis on the given dataset.

VII. Conclusion

This study went into the realm of sentiment analysis by examining the performance of two powerful word embedding techniques, Word2Vec and GloVe, in conjunction with the Long Short-Term Memory (LSTM) architecture. The research aimed to address two pivotal questions: which model, when combined with LSTM, attains better accuracy, and which model demonstrates more effective loss management. Through meticulous experimentation and analysis, the insights drawn from this study illuminate the efficacy of these models and their practical implications in sentiment analysis.

The dataset, sourced from a Twitter chat available on Kaggle, comprises 75,682 records, each characterized by three independent variables: ID, Company, and Message. The primary focal point centers on the dependent variable, Sentiment, associated with each message. The objective was to harness these variables to ascertain sentiment patterns effectively.

The experimental results, showcased in Tables 1 and 2, have offered intriguing revelations. Both the Word2Vec-LSTM and GloVe-LSTM models demonstrated exceptional prowess by achieving perfect accuracy (1.0000) throughout the training process. This affirms that both embedding techniques, Word2Vec and GloVe, are adept at accurately classifying Sentiment within the dataset.

Regarding loss management, the GloVe-LSTM model emerged as the frontrunner. With consistently lower loss values compared to Word2Vec-LSTM, the GloVe-LSTM model exhibited greater precision in aligning its predictions with actual sentiment labels. This highlights its ability to capture and interpret sentiment nuances

more effectively.

In conclusion, this study does not only sheds light on the capabilities of Word2Vec and GloVe embedding techniques but also highlights the crucial role of the LSTM architecture in sentiment analysis. The exceptional accuracy achieved by both models underscores their proficiency in deciphering Sentiment, while the lower loss values achieved by the GloVe-LSTM model emphasize its superior predictive precision. The findings presented here contribute to a deeper understanding of the interplay between embedding techniques and LSTM in sentiment analysis, offering valuable insights for applications in sentiment assessment, market research, and social media analysis.

Study Limitations

While this research sheds light on the performance of Word2Vec and GloVe in combination with the LSTM architecture for sentiment analysis, there are certain limitations that warrant acknowledgment. Notably, the study hinges on the utilization of a single dataset sourced from a Twitter chat on Kaggle. This constraint may potentially limit the generalizability of the findings to broader sentiment contexts. Moreover, the analysis focuses solely on the application of LSTM, omitting exploration of other deep neural network architectures that could provide nuanced insights into model performance.

Directions for Future Studies

To comprehensively unravel the intricacies of the models under evaluation, future studies could extend the research in several directions:

- **Diverse Datasets:** Incorporating multiple datasets from various sources and domains could provide a more comprehensive understanding of the models' adaptability and performance across different contexts.
- **Comparative Architectures:** Examining a broader range of deep neural network architectures, such as

Convolutional Neural Networks (CNNs) or Transformer-based models, could elucidate the strengths and weaknesses of each architecture when coupled with Word2Vec and GloVe embeddings.

- **Hyperparameter Tuning:** Exploring various hyperparameter configurations for both embedding techniques and LSTM could uncover optimal settings for achieving even better performance.
- **Fine-grained Sentiment Analysis:** Investigating the models' performance in distinguishing fine-grained sentiments (e.g., positive, negative, neutral) within the dataset could offer insights into their ability to discern subtleties in sentiment expression.
- **Domain-specific Analysis:** Conducting sentiment analysis in specialized domains, such as financial markets, healthcare, or politics, could unveil how well the models generalize to unique contexts.

In conclusion, while this study makes valuable contributions to the understanding of Word2Vec and GloVe in tandem with LSTM, it represents just a stepping stone in a broader research landscape. Future endeavors that encompass diverse datasets, a range of neural network architectures, and domain-specific analyses have the potential to enrich our understanding of sentiment analysis models and their real-world applications.

References

Acosta, J., Lamaute, N., Luo, M., Finkelstein, E., & Andreea, C. (2017). Sentiment analysis of twitter messages using word2vec. Proceedings of student-faculty research day, CSIS, Pace University, 7, 1-7.

Annett, M., & Kondrak, G. (2008). A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Advances in Artificial Intelligence: 21st Conference of the Canadian Society for Computational Studies of Intelligence*, Canadian AI 2008 Windsor, Canada, May 28-30, 2008 Proceedings 21 (pp. 25-35). Springer Berlin Heidelberg.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77, 236-246.

Chassagnon, G., Vakalopolou, M., Paragios, N., & Revel, M. P. (2020). Deep learning: definition and perspectives for thoracic imaging. *European radiology*, 30, 2021-2030.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 54(3), 1-40.

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).

Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S., & Nagrath, P. (2020). Sentiment analysis using gated recurrent neural networks. *SN Computer Science*, 1, 1-13.

Sopuru, J. C., & Akkaya, M. (2019). Guide for Modelling a Network Flow-Based Detection System for Malware Categorization: A Review of Related Literature. *Applying Methods of Scientific Inquiry Into Intelligence, Security, and Counterterrorism*, 150-178.

Singh, P. K., & Paul, S. (2021). Deep learning approach for negation handling in sentiment analysis. *IEEE Access*, 9, 102579-102592.

Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring transformers in natural language generation: Gpt, bert, and xlnet. arXiv preprint arXiv:2102.08036.

Vijayvergia, A., & Kumar, K. (2021). Selective shallow models strength integration for emotion detection using GloVe and LSTM. *Multimedia Tools and Applications*, 80(18), 28349-28363.

Xiao, L., Wang, G., & Zuo, Y. (2018, December). Research on patent text classification based on word2vec and LSTM. In *2018 11th International Symposium on Computational Intelligence and Design (ISCID) (Vol. 1, pp. 71-74)*. IEEE.

☆ Comparative Analysis of Word2Vec and GloVe with LSTM for Sentiment Analysis: Accuracy and Loss Evaluation on Twitter Data