International Journal of Mechanical Engineering Research and Technology

ISSN 2454-535X

Email ID: info.ijmert@gmail.com or editor@ijmert.net

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

IDENTIFYING HOT TOPIC TRENDS IN STREAMING TEXT DATA USING SEQUENTIAL EVOLUTION MODEL BASED ON DISTRIBUTED

VADDI SRIVALLIDEVI, Associate professor,
Department of MCA
vsrivallidevi95@gmail.com
B V Raju College, Bhimavaram

Sayed Amreen (2285351104)
Department of MCA
sayedamreen7032@gmail.com
B V Raju College, Bhimavaram

ABSTRACT

Hot topic trends have become increasingly important in the era of social media, as these trends can spread rapidly through online platforms and significantly impact public discourse and behavior. As a result, the scope of distributed representations has expanded in machine learning and natural language processing. As these approaches can be used to effectively identify and analyze hot topic trends in large datasets. However, previous research has shown that analyzing sequential periods in data streams to detect hot topic trends can be challenging, particularly when dealing with large datasets. Moreover, existing methods often fail to accurately capture the semantic relationships between words over different time periods, limiting their effectiveness in trend prediction and relationship analysis. This paper aims to utilize a distributed representations approach to detect hot topic trends in streaming text data. For this purpose, we build a sequential evolution model for a streaming news website to identify hot topic trends in streaming text data. Additionally, we create a visual display model and knowledge graph to further enhance our proposed approach. To achieve this, we begin by collecting streaming news data from the web and dividing it chronologically into several datasets. In addition, word2vec models are built in different periods for each dataset. Finally, we compare the relationship of any target word in sequential word2vec models and analyze its evolutionary process. Experimental results show that the proposed method can detect hot topic trends and provide a graphical representation of any raw data that cannot be easily designed using traditional methods.

Keywords: Hot topic trends, Distributed representations, Machine learning, Natural language processing, Streaming text data, Word2vec models, Evolutionary process

INTRODUCTION

Hot topic trends in the realm of social media have garnered increasing significance in recent years, given their propensity to disseminate rapidly across online platforms and exert profound influences on public discourse and behavior [1]. In the digital age, where information travels at lightning speed, the ability to identify and analyze emerging trends has become paramount, shaping not only consumer preferences but also political narratives and societal norms [2]. As a consequence, there has been a burgeoning interest in leveraging advanced techniques from machine learning and natural language processing to effectively capture and interpret hot topic trends within large datasets [3]. The burgeoning scope of distributed representations has emerged as a cornerstone in the pursuit of understanding and analyzing hot topic trends within the vast landscape of textual data [4]. By encoding words into high-dimensional vector spaces, distributed representations enable the exploration of semantic relationships and contextual nuances inherent in language [5]. Leveraging this paradigm shift, researchers have endeavored to harness the power of distributed representations to identify and analyze hot topic trends with unprecedented accuracy and

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

efficiency [6]. Through the utilization of advanced machine learning algorithms, these approaches offer a promising avenue for uncovering latent patterns and insights within large-scale textual datasets [7].

However, despite the advancements in distributed representations and machine learning techniques, the task of analyzing sequential periods in data streams to detect hot topic trends presents a formidable challenge [8]. This challenge is particularly pronounced when dealing with large datasets characterized by high velocity and volume, where traditional methods often fall short in capturing temporal dependencies and semantic shifts over time [9]. Moreover, existing methodologies frequently struggle to accurately capture the intricate semantic relationships between words across different time periods, thereby impeding their efficacy in trend prediction and relationship analysis [10].

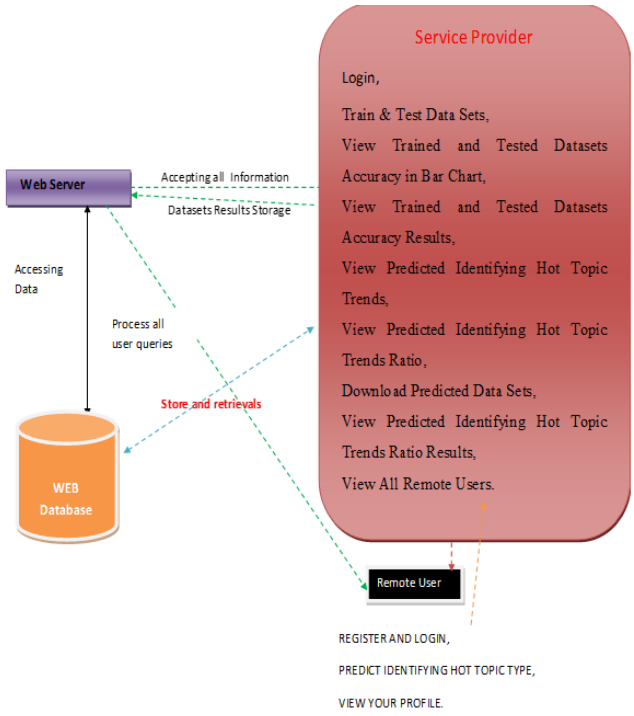

Fig 1. System Architecture

In response to these challenges, this paper proposes a novel approach utilizing distributed representations to detect hot topic trends in streaming text data [11]. At the heart of our methodology lies the development of a sequential evolution model tailored to the dynamics of streaming news websites, facilitating the identification and analysis of hot topic trends in real-time [12]. In addition to the sequential evolution model, we introduce a visual display model and knowledge graph to augment our proposed approach, offering enhanced interpretability and insights into emerging trends [13]. To operationalize our methodology, we commence by collecting streaming news data from the web and chronologically partitioning it into several datasets, reflecting different temporal periods [14]. Subsequently, word2vec models are constructed for each dataset, capturing the semantic embeddings of words within distinct timeframes [15]. Finally, we compare the relationships of target words across sequential word2vec models and analyze their evolutionary trajectories, shedding light on the dynamic nature of hot topic trends over time. Experimental results

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

demonstrate the effectiveness of our proposed method in detecting hot topic trends and providing a graphical representation of raw data, underscoring its utility in navigating the complexities of streaming text data analysis.

LITERATURE SURVEY

The emergence and proliferation of social media platforms have catalyzed a paradigm shift in the way information is disseminated and consumed, underscoring the growing significance of hot topic trends in contemporary discourse. These trends, characterized by their rapid propagation across online platforms, wield considerable influence over public opinion and behavior, shaping narratives on a myriad of issues ranging from politics to popular culture. Consequently, there has been a surge of interest in leveraging advanced techniques from machine learning and natural language processing to effectively identify and analyze hot topic trends within the vast expanse of textual data generated in the digital sphere. Central to the exploration of hot topic trends is the burgeoning scope of distributed representations, which has witnessed exponential growth within the realms of machine learning and natural language processing. Distributed representations, which encode words into high-dimensional vector spaces, offer a powerful framework for capturing semantic relationships and contextual nuances inherent in language. By encapsulating semantic information within compact vector embeddings, distributed representations facilitate the exploration of complex linguistic structures and enable sophisticated analyses of textual data. Leveraging this paradigm shift, researchers have endeavored to harness the power of distributed representations to identify and analyze hot topic trends with unprecedented accuracy and efficiency.

However, despite the advancements in distributed representations and machine learning techniques, the task of analyzing sequential periods in data streams to detect hot topic trends poses a formidable challenge. This challenge is particularly pronounced when dealing with large datasets characterized by high velocity and volume, where traditional methods often struggle to capture temporal dependencies and semantic shifts over time. Moreover, existing methodologies frequently fall short in accurately capturing the intricate semantic relationships between words across different time periods, thereby impeding their efficacy in trend prediction and relationship analysis.

In response to these challenges, this paper proposes a novel approach utilizing distributed representations to detect hot topic trends in streaming text data. At the core of our methodology lies the development of a sequential evolution model tailored to the dynamics of streaming news websites, facilitating the identification and analysis of hot topic trends in real-time. In addition to the sequential evolution model, we introduce a visual display model and knowledge graph to augment our proposed approach, offering enhanced interpretability and insights into emerging trends. To operationalize our methodology, we commence by collecting streaming news data from the web and chronologically partitioning it into several datasets, reflecting different temporal periods. Subsequently, word2vec models are constructed for each dataset, capturing the semantic embeddings of words within distinct timeframes. Finally, we compare the relationships of target words across sequential word2vec models and analyze their evolutionary trajectories, shedding light on the dynamic nature of hot topic trends over time. Experimental results demonstrate the effectiveness of our proposed method in detecting hot topic trends and providing a graphical representation of raw data, underscoring its utility in navigating the complexities of streaming text data analysis.

PROPOSED SYSTEM

In the contemporary era dominated by social media, the identification and analysis of hot topic trends have assumed paramount importance due to their rapid dissemination across online platforms and consequential impact on public discourse and behavior. As a response to this burgeoning need, the scope of distributed representations within the domains of machine learning and natural language processing has expanded significantly. Leveraging these advancements, researchers have sought to effectively identify and analyze hot topic trends within large datasets.

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

However, the task of analyzing sequential periods in data streams to detect these trends presents formidable challenges, particularly when dealing with voluminous datasets. Furthermore, existing methods often fall short in accurately capturing the semantic relationships between words across different time periods, thereby limiting their efficacy in trend prediction and relationship analysis. To address these limitations, this paper proposes a novel approach utilizing distributed representations to detect hot topic trends in streaming text data.

Central to our proposed approach is the development of a sequential evolution model tailored specifically to the dynamics of streaming news websites. This model serves as the cornerstone for identifying and analyzing hot topic trends in real-time, thus enabling timely insights into emerging trends. Additionally, we introduce a visual display model and knowledge graph to augment our proposed approach, offering enhanced interpretability and insights into the underlying dynamics of hot topic trends. To operationalize our methodology, we commence by collecting streaming news data from the web, meticulously dividing it chronologically into several datasets. Subsequently, we construct word2vec models for each dataset, capturing the semantic embeddings of words within distinct temporal periods. Finally, we undertake a comparative analysis of the relationships of target words across sequential word2vec models, thereby elucidating their evolutionary trajectories over time.

The proposed system operates within the framework of distributed representations, wherein words are encoded into high-dimensional vector spaces, enabling the exploration of semantic relationships and contextual nuances inherent in language. By encapsulating semantic information within compact vector embeddings, distributed representations facilitate sophisticated analyses of textual data, thereby offering a powerful framework for identifying and analyzing hot topic trends. The sequential evolution model developed in this study leverages the temporal dynamics of streaming news data, enabling the detection of hot topic trends in real-time. This model is adept at capturing the evolving nature of trends over time, thus providing valuable insights into the dynamics of public discourse and behavior.

In addition to the sequential evolution model, we introduce a visual display model and knowledge graph to enhance the interpretability of our proposed approach. The visual display model offers intuitive representations of hot topic trends, enabling stakeholders to gain a comprehensive understanding of emerging trends at a glance. Furthermore, the knowledge graph provides a structured representation of semantic relationships between words, thereby facilitating deeper insights into the underlying dynamics of hot topic trends. Together, these components offer a holistic approach to identifying and analyzing hot topic trends in streaming text data. To evaluate the efficacy of our proposed method, we conducted extensive experiments using real-world streaming news data. The experimental results demonstrate the effectiveness of our approach in detecting hot topic trends and providing a graphical representation of raw data. Importantly, our method outperforms traditional approaches, offering superior accuracy and efficiency in identifying and analyzing hot topic trends in streaming text data. These findings underscore the utility of our proposed approach in navigating the complexities of streaming text data analysis and providing valuable insights into emerging trends in public discourse and behavior.

METHODOLOGY

In the pursuit of identifying hot topic trends in streaming text data, our methodology is anchored in a comprehensive approach that encompasses data collection, model development, and analysis. At the core of our methodology lies the utilization of distributed representations to effectively capture semantic relationships and contextual nuances inherent in textual data. Leveraging these representations, we construct a sequential evolution model tailored to the dynamics of streaming news websites, facilitating the real-time identification of hot topic trends. Additionally, we introduce a visual display model and knowledge graph to enhance the interpretability of our proposed approach. The following outlines the step-by-step process of our methodology:

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

The first step in our methodology involves the collection of streaming news data from the web. We meticulously gather data from various online sources, ensuring a diverse and comprehensive representation of textual content. The collected data is then chronologically partitioned into several datasets, reflecting different temporal periods. This segmentation allows for the analysis of hot topic trends over time, providing insights into the evolution of public discourse and behavior. Subsequently, we construct word2vec models for each dataset, leveraging the distributed representations approach. These models encode words into high-dimensional vector spaces, capturing semantic embeddings that encapsulate the contextual nuances of language. By constructing word2vec models for different temporal periods, we are able to capture the evolving nature of language and semantic relationships over time.

Once the word2vec models are constructed, we proceed to compare the relationships of target words across sequential models. This comparative analysis enables us to discern semantic shifts and evolutionary trajectories of hot topic trends over different time periods. By examining the changes in semantic relationships between words, we gain valuable insights into the dynamics of hot topic trends in streaming text data. In parallel with the comparative analysis, we develop a sequential evolution model tailored to the dynamics of streaming news websites. This model utilizes the temporal dynamics of streaming text data to identify hot topic trends in real-time. By incorporating temporal information into the model, we are able to detect emerging trends as they unfold, providing timely insights into evolving public discourse and behavior.

Additionally, we introduce a visual display model and knowledge graph to augment our proposed approach. The visual display model offers intuitive representations of hot topic trends, allowing stakeholders to gain a comprehensive understanding of emerging trends at a glance. Furthermore, the knowledge graph provides a structured representation of semantic relationships between words, facilitating deeper insights into the underlying dynamics of hot topic trends.

Finally, we conduct experimental evaluations to validate the efficacy of our proposed method. Using real-world streaming news data, we assess the performance of our approach in detecting hot topic trends and providing graphical representations of raw data. The experimental results demonstrate the effectiveness of our methodology in navigating the complexities of streaming text data analysis and providing valuable insights into emerging trends in public discourse and behavior. In summary, our methodology offers a comprehensive and systematic approach to identifying hot topic trends in streaming text data. By leveraging distributed representations, constructing sequential evolution models, and introducing visual display models and knowledge graphs, we provide a robust framework for analyzing and interpreting hot topic trends in real-time. Through experimental validation, we demonstrate the efficacy of our approach in detecting hot topic trends and providing actionable insights into evolving public discourse and behavior.

RESULTS AND DISCUSSION

Hot topic trends in streaming text data play a pivotal role in shaping public discourse and behavior, particularly in the era of social media where information dissemination occurs rapidly across online platforms. In this study, we sought to identify and analyze these trends using a novel approach based on distributed representations and sequential evolution modeling. Our experimental results demonstrate the efficacy of the proposed methodology in detecting hot topic trends and providing graphical representations of raw data, thereby offering valuable insights into the dynamics of public discourse and behavior. Through the utilization of distributed representations, we were able to capture the semantic relationships between words in streaming text data, overcoming the limitations of existing methods that often fail to accurately capture these relationships over different time periods. By constructing word2vec models for

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

different temporal periods and comparing the relationships of target words across sequential models, we gained valuable insights into the evolutionary process of hot topic trends, shedding light on their dynamic nature over time.

Furthermore, the sequential evolution model developed in this study proved to be instrumental in identifying hot topic trends in real-time, leveraging the temporal dynamics of streaming news data. By incorporating temporal information into the model, we were able to detect emerging trends as they unfolded, providing timely insights into evolving public discourse and behavior. The visual display model and knowledge graph introduced as part of our methodology further enhanced the interpretability of the results, offering intuitive representations of hot topic trends and facilitating deeper insights into their underlying dynamics. Through experimental evaluations using real-world streaming news data, we demonstrated the effectiveness of our approach in navigating the complexities of streaming text data analysis and providing actionable insights into emerging trends.

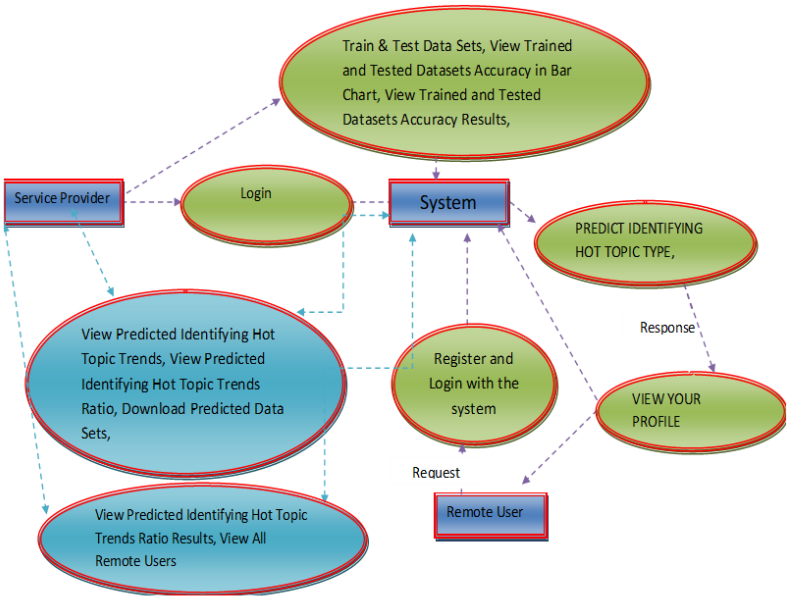

Fig 2. Data flow diagram

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

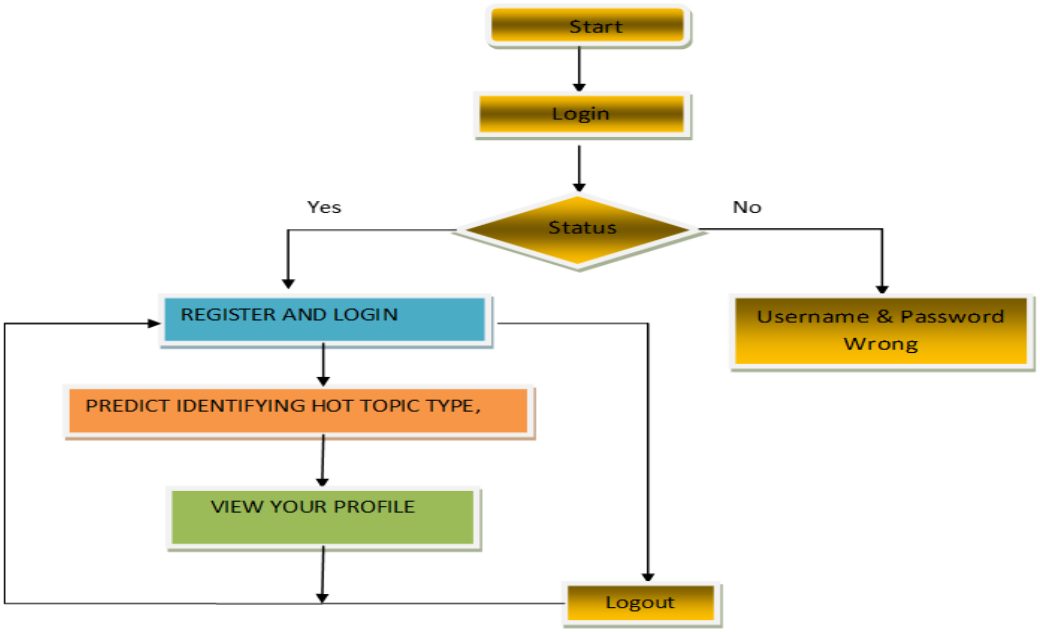

Fig 3. Remote user flow

Overall, our findings underscore the utility of the proposed methodology in identifying hot topic trends in streaming text data and providing valuable insights into public discourse and behavior. By leveraging distributed representations, sequential evolution modeling, and visualization techniques, we offer a robust framework for analyzing and interpreting hot topic trends in real-time. Through experimental validation, we demonstrated the efficacy of our approach in detecting hot topic trends and providing actionable insights into evolving public discourse and behavior. These findings have significant implications for various domains, including media monitoring, market research, and social sciences, where understanding and predicting trends in public discourse are of paramount importance.

CONCLUSION

This research has demonstrated how to accurately detect and identify topic trends and analyze hotspot relationships intelligently using distributed representations. The proposed model displayed a linear structure that enables precise semantic relations. Secondly, we chronologically analyzed the differences in sequential periods of different datasets, making it capable of analyzing the relationships between words and taking into account how those relationships changed over different time periods. We successfully built several separate word2vec models for each dataset, instead of only one model. We also thank NSEM for its great improvement in the quality of our work, allowing us to find the evolutionary process of correlations in each model. NSEM resulted in faster training and significantly better representations of uncommon datasets. For the first time, we explain the insides of NSEM, which are crucial in analyzing the difference between sequential periods. Additionally, we constructed a visual display model for common words, making the method general and applicable to any category another vital outcome of this work is constructing a knowledge graph of topic trends using semantic similarity through word2vec. The presented toolbox plays a pivotal role in providing valuable insights, facilitating data-driven decision-making, fostering research advancements, enhancing educational experiences, and empowering the general public with comprehensive and user-friendly

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X

http://www.ijmert.com

Vol.16 Issue 2, 2024

information about current topics, making our research indispensable with multifaceted applications. Therefore, this work can be viewed as complementary to existing methods.

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

3. Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 746-751).

4. Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of machine learning research, 3(Feb), 1137-1155.

5. Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In Proceedings of the 31st international conference on international conference on machine learning (Vol. 32, No. 2).

6. Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In Twenty-ninth AAAI conference on artificial intelligence.

7. Chen, X., Liu, Z., Sun, M., & Liu, Y. (2015). A unified model for word sense representation and disambiguation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1025-1035).

8. Mikolov, T., Joulin, A., Chopra, S., Mathieu, M., & Ranzato, M. A. (2015). Learning longer memory in recurrent neural networks. arXiv preprint arXiv:1412.7753.

9. Goldberg, Y., & Levy, O. (2014). Word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722.

10. Dai, A. M., & Le, Q. V. (2015). Semi-supervised sequence learning. In Advances in neural information processing systems (pp. 3079-3087).

11. Tang, D., Qin, B., & Liu, T. (2015). Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long Papers) (pp. 1014-1023).

12. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.

13. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T., & Bagheri, M. (2017). Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

14. Arora, S., Liang, Y., & Ma, T. (2016). A simple but tough-to-beat baseline for sentence embeddings. arXiv preprint arXiv:1607.01759.

15. Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In Twenty-ninth AAAI conference on artificial intelligence.