# Comparative Study on Word2Vec ML models on Resume Screening Dataset

##### **Vidit Khemka\*, Anchal Gupta\*, Bharat Bhalla\***

\* Department of Computer Science and Engineering, Delhi Technical Campus

***Abstract*-** We introduce a comparative study on different Word2Vec Machine Learning Models (Elmo, Universal sentencing encoder, GLOVE 84B) on predicting the best match of resume with pre-defined constraints on resume as part of resume screening process. Finding the best ML Model that suits our specific application is the focus of this study.

***Index Terms***- About four key words or phrases in alphabetical order, separated by commas. Keywords are used to retrieve documents in an information system such as an online journal or a search engine. (Mention 4-5 keywords)

1. Introduction

Resume screening process involves high level of human intelligence to filter resume according to a given requirements. In this article we have address the issue of resume screening process by using machine learning models (Elmo, Universal sentencing encoder, GLOVE84B). Conclusively we want to compare which machine learning model performs the best under given set of constraints.

We are converting the given set of constraints to a vector representation and also converting resume to a vector. This vector conversion is done by one of the given Machine Learning (Word2Vec) Models:

1. Elmo
2. Universal sentencing encoder
3. GLOVE84B

After this we are applying kNN Algorithm to the resume vector with respect to given constraints vectors. Finding the highest number of matches to given constraints amounts to indexing that specific resume of high match.

Machine Learning (Word2Vec) Models which best represents our Text to Vector will help choose the best resume

1. Study of Word2Vec Models
   * + 1. Elmo:

Elmo is a deep contextualized word representation that models both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). These word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre-trained on a large text corpus. They can be easily added to existing models and significantly improve the state of the art across a broad range of challenging NLP problems, including question answering, textual entailment and sentiment analysis.

**Salient features**

ELMo representations are:

* *Contextual*: The representation for each word depends on the entire context in which it is used.
* *Deep*: The word representations combine all layers of a deep pre-trained neural network.
* *Character based*: ELMo representations are purely character based, allowing the network to use morphological clues to form robust representations for out-of-vocabulary tokens unseen in training.
  + - 1. Universal sentencing encoder:

The best sentence encoders available right now are the two Universal Sentence Encoder models by Google. One of them is based on a Transformer architecture and the other one is based on Deep Averaging Network (DAN). They are pre-trained on a large corpus and can be used in a variety of tasks (sentimental analysis, classification and so on). Both models take a word, sentence or a paragraph as input and output a 512-dimensional vector.

The transformer model is designed for higher accuracy, but the encoding requires more memory and computational time. The DAN model on the other hand is designed for speed and efficiency, and some accuracy is sacrificed.

* + - 1. GLOVE 84B:

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space

https://nlp.stanford.edu/projects/glove/

1. WRITE DOWN YOUR STUDIES AND FINDINGS

Now it is the time to articulate the research work with ideas gathered in above steps by adopting any of below suitable approaches:

## A. Bits and Pieces together

In this approach combine all your researched information in form of a journal or research paper. In this researcher can take the reference of already accomplished work as a starting building block of its paper.

Jump Start

This approach works the best in guidance of fellow researchers. In this the authors continuously receives or asks inputs from their fellows. It enriches the information pool of your paper with expert comments or up gradations. And the researcher feels confident about their work and takes a jump to start the paper writing.

## B. Use of Simulation software

There are numbers of software available which can mimic the process involved in your research work and can produce the possible result. One of such type of software is Matlab. You can readily find Mfiles related to your research work on internet or in some cases these can require few modifications. Once these Mfiles are uploaded in software, you can get the simulated results of your paper and it easies the process of paper writing.

As by adopting the above practices all major constructs of a research paper can be written and together compiled to form a complete research ready for Peer review.

1. GET PEER REVIEWED

Here comes the most crucial step for your research publication. Ensure the drafted journal is critically reviewed by your peers or any subject matter experts. Always try to get maximum review comments even if you are well confident about your paper.

**For peer review send you research paper in IJSRP format to** [**editor@ijsrp.org**](mailto:editor@ijsrp.org)**.**

1. IMPROVEMENT AS PER REVIEWER COMMENTS

Analyze and understand all the provided review comments thoroughly. Now make the required amendments in your paper. If you are not confident about any review comment, then don't forget to get clarity about that comment. And in some cases there could be chances where your paper receives number of critical remarks. In that cases don't get disheartened and try to improvise the maximum.

**After submission IJSRP will send you reviewer comment within 10-15 days of submission and you can send us the updated paper within a week for publishing.**

This completes the entire process required for widespread of research work on open front. Generally all International Journals are governed by an Intellectual body and they select the most suitable paper for publishing after a thorough analysis of submitted paper. Selected paper get published (online and printed) in their periodicals and get indexed by number of sources.

**After the successful review and payment, IJSRP will publish your paper for the current edition. You can find the payment details at:** [**http://ijsrp.org/online-publication-charge.html**](http://ijsrp.org/online-publication-charge.html)**.**

1. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

Acknowledgment

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

References

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems* (Book style)*.* Belmont, CA: Wadsworth, 1993, pp. 123–135.
3. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
4. B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
5. E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” *IEEE Trans. Antennas Propagat.*, to be published.
6. J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” *IEEE J. Quantum Electron.*, submitted for publication.

Authors

**First Author** – Author name, qualifications, associated institute (if any) and email address.

**Second Author** – Author name, qualifications, associated institute (if any) and email address.

**Third Author** – Author name, qualifications, associated institute (if any) and email address.

**Correspondence Author** – Author name, email address, alternate email address (if any), contact number.