Empirical Evaluation of Word Representation Methods in the Context of Candidate-Job Recommender Systems

Gazmira Brahushi
Durham College
Oshawa, ON, Canada
gazmira.brahushi@dcmail.ca

Uzair Ahmad
Durham College
Oshawa, ON, Canada
uzair.ahmad@durhamcollege.ca

2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI) | 979-8-3503-2088-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCMI56532.2022.10068466

***Abstract*—In this paper, we have evaluated our hybrid two-way recommendation system with expert-ranked resumes and job descriptions. The aim of the paper is to compare the lists produced by the recommendation system with human-ranked lists for candidate and job descriptions. Firstly, we set up four scenarios such as the matching of resume to resumes, job to jobs, resume to jobs, and job to resumes, and prepared a human ranking based on the content similarity on a total of 400 documents. Based on this annotated corpus we tested our system to calculate the cosine-similarity-based ranking for each scenario using the Global Vectors for Word Embeddings and Term Frequency-Inverse Document Frequency representations. Finally, we compared the similarities of human ranked lists and system-ranked lists by using the Rank Biased Overlap (RBO) similarity score. In both methods, GloVe and TF-IDF, the median RBO between human-ranked lists and system ranked are greater than 0.5. The highest median score is achieved on TF-IDF with a slight difference compared to GloVe apart from the ranking of resume-to-resume scenario where the variation between the two methods is considerable. This is due to the similarity between human-ranked lists and program-generated lists.**

Keywords—GloVe, TF-IDF, RBO

I. Introduction

In 2022, using job search websites is one of the most popular ways for job seekers to find work prospects and for companies to find the right candidates. There are several online job boards available, as well as apps, publishers, google search, platforms for social media networking organizations, and firm websites. One important and widely used today is LinkedIn, a professional network that helps to get close recruiters and candidates and helps both to create their network connections with other professionals. Based on the latest statistics of June 2022, 20 million Canadians use LinkedIn. [1]. Another job board that proudly stands in over 60 countries of the world is Indeed. It accounts for 16.6 percent of the market share among all other job boards. [2]. Talent.com is another opportunity for candidates and companies to find the ideal match. It is currently in 78 countries with a strong presence in the Canadian market. [3].

II. Related Work

The authors introduced a job recommendation system in [4] that employed numerous machine learning algorithms to learn about the applicants' job selection process. They achieved the best result on Bi-LSTM with the attention model because of their tests. Furthermore, they use non-machine learning approaches to handle job and applicant cold-start issues. They supplemented the model Bi-LSTMs with occupations performed by similar applicants and similar positions held by the candidate.

A lot of work has been conducted on the LinkedIn recruitment process. In [5] they have proposed an online recommendation system consisting of two services: retrieval service and scoring /ranking service. The retrieval service processes a user query creates a candidate set of members along with their features and computes the vector embedding for the query in parallel. The second pass scoring involves the development of similarity features based on the query and member embeddings, and the top-ranked results are returned.

In [6] researchers have provided an offline and online architecture to implement a hybrid model approach for talent searching on LinkedIn. They present an entity-personalized talent search model that combines GLMix and GBDT models to deliver personalized talent suggestions based on nonlinear tree interaction characteristics obtained by the GBDT.

The researchers at [7] have introduced a hybrid recommender system for job-seeking and recruiting websites. The data extracted from the website is modeled using a multi-relational graph and their solution has two main features; it uses content-based similarity relation and interaction-based reactions.

III. Methodology

We used in this paper 400 records of jobs and resumes and then based on their content similarity, for each job and resume we predicted ranked lists by considering the top 10, top 20, top 50, and top 100 predictions. In order to generate these predictions, we applied 2 methods; the first one is by using Global Vectors for Word Embeddings (GloVe) [8] and the second is by using Term Frequency-Inverse Document Frequency (TF-IDF) representation.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on July 12,2025 at 04:33:19 UTC from IEEE Xplore. Restrictions apply.

IV. Hybrid System Architecture

A. Hybrid System Components

Fig. 1 presents the components of our hybrid recommendation system. The data is fed into the Data Preparation module where it is first parsed into a readable format. Parsed data is then cleaned using various Natural Language Processing (NLP) techniques. The data is then annotated, and the Spacy module is trained for the extraction of entities. Using the trained model, the corpus data is reduced to relevant entities for better performance.

Associate rule mining is used to find patterns in the data and extract relevant rules. These rules help us to augment user queries.

The query engine uses various methods such as TF-IDF, Word2vec, and Doc2Vec to gain recommendations for job search as well as candidate resume search from the augmented query.

The TF-IDF is frequently used in the disciplines of text mining and information retrieval to determine the connection between each word in a collection of documents. They are specifically used to extract core words (keywords) from texts, determine search ranking, compute degrees of similarity across documents, and more.[9] The TF in TF-IDF refers to the use of particular words inside texts. In documents, words having a high TF value play a significant role. The DF reveals how frequently a certain term appears in the group of documents. It counts the number of times the term appears across several papers rather than just one. High DF value words do not have importance because they are used often across all publications. As a result, the importance of words in all papers is determined using the IDF, which is the inverse of the DF. This refers to how prevalent or uncommon a term is throughout the full collection of documents. Based on this approach if the term in question is widely used and appears in several papers the IDF would be close to zero, otherwise, it would go close to 1.[9].

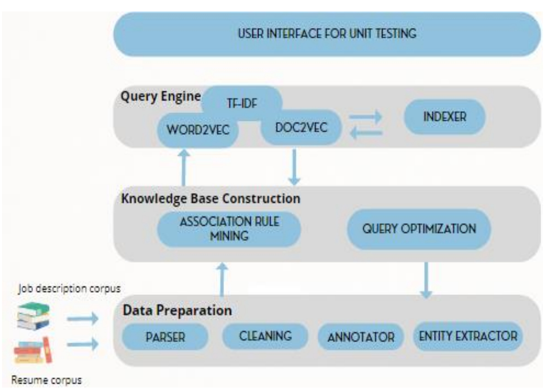

Fig. 1. System architecture.

Word2Vec is a method for learning word associations in a task involving natural language processing. In order to detect synonyms and antonyms of words or to suggest a word to finish an unfinished phrase, the word2vec algorithms apply a neural network model. Word2vec utilizes a collection of what are known as vectors of integers to represent any unique word. Using Word2Vec, there are two main ways to represent words.

- Continuous Bag of Words (CBOW) Approach - Using the context of the words around them, this method predicts the words that can fit in the center of an unfinished phrase, helping to complete it in part.
- The Skip-Gram technique - used to forecast surrounding words or context words given a current word in the same phrase. Each word in the big corpus is entered into the Skip-gram model, and the hidden or embedding layer then predicts the context words using the embedding weights.[10].

Doc2vec aims to generate a numeric representation of the document regardless of the length of the document. However, unlike words, documents do not have logical structures as words do, thus a different approach must be devised. Mikilov and Le proposed adding a second vector (Paragraph ID) as can be seen in Fig. 2.

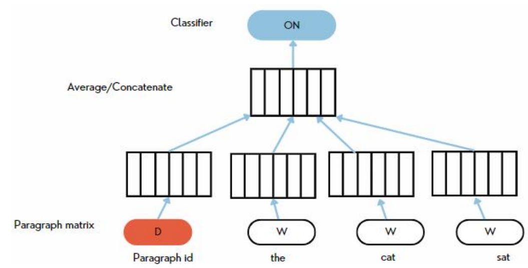

Fig. 2. PV-DM Model [7].

It is only a minor addition to the actual Continues Bag of Words (CBOW) model. But in order to anticipate the following word, they introduced a second feature vector that is document-unique rather than only utilizing words. As a result, the document vector D is trained together with the word vectors W, and at the conclusion of training, it contains a numerical representation of the document. The model used is named Distributed Memory Version of Paragraph Vector (PV-DM). It functions as a recollection that recalls what is absent from the current context. The document vector aims to represent the notion of a document, whereas the word vectors express the concept of a word.[10]

The indexer is used to fire queries for either job positions or Resume searches. It takes as an argument the query document and applies cosine similarity. It is employed to measure how similar the two vectors are, regardless of scale. It is specifically used in mathematics to calculate the cosine of the perspective between non-zero vectors projected in a multidimensional space. When the angle is smaller, the cosine similarity is better, and vice versa. [11]

Lastly, a user interface designed using Flask is used to help Job Seekers and Employers search for what they desire.

B. GloVE

Global Vectors for Word Embeddings (GloVe) is an unsupervised learning method for producing vector representations of words. It maps the words into a space where the semantic similarity between the words is determined by the

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on July 12,2025 at 04:33:19 UTC from IEEE Xplore. Restrictions apply.

distance between the words. It was introduced in 2014 and is a Stanford open-source project.

In our paper, we have used the pre-trained GloVe model. It includes English word vectors that have been trained using the combined Wikipedia 2014 and Gigaword 5th Edition corpora.[5]

V. Data Preparation

In this paper, we have considered data from jobs and resumes description. Names of occupations and resumes are shown in table 1. The columns of the table represent respectively the candidate profiles and job types used in our experiments.

TABLE I. Resumes and Job Occupation Types Used for Our Experiments

Candidate Profile	Job Types
Accounting Clerk	Account Manager
Administrative assistant	Affiliate Marketing
Assistant basketball coach	Assistant Manager
Assistant Manager	Business Analyst
Assistant Teacher	Construction Manager
Biomedical engineer	Data Analyst
Construction manager	Electrician
Customer service representative	Programmer
IT support	Support Specialist
Data Analyst	Data Analyst
Programmer	Programmer
Grade teacher	Teacher

Fig.3 represents all the preprocessing phases. Firstly, we manually filtered out 10 jobs and resumes by selecting more relevant resumes and job descriptions with similar meanings based on their content. In the following step, we applied a manual ranking for a specific resume, and how similar that resume was to other resumes by defining lists with 10 elements, where the 1st element was more meaningful related to a specific resume and the last element was less related to it. The same procedure was applied also for similarities of jobs. In addition, we have ranked manually, for a specific resume the job's content related to it, and for a particular job, resumes that were more meaningful related to that. We took into consideration 100 instances for each scenario. As we can see from the schema the result of this data preparation process is the application of glove embedding or TF-IDF matrix on corpus data of resume or job.

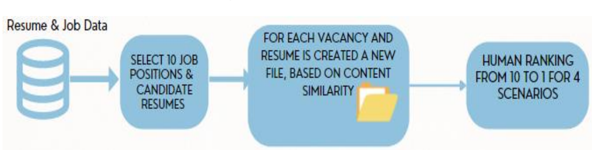

Fig. 3. Manual ranking process.

VI. Evaluation Methodology

In this study, we will apply rank-biased overlap (RBO). This measurement is based on a straightforward user model where the user compares the overlap of the two ranks at gradually increasing deeper levels. Some of its most important features are it gives greater importance to the differences at the top of the list rather than the differences at the bottom; it is consistent regardless of the depth made accessible by the list provider; and it does all of the abovementioned while putting the fewest assumptions possible on the data, none of which should go against the nature of indefinite rankings.[12] Fig. 4, explains schematically how the RBO similarity measure works. This function takes as a parameter 2 ranked lists which might be definite or indefinite and a coefficient of p. The latter controls the degree of the top weightiness of the RBO score. In the first scenario if the p<0.9 then the first elements of the list have the highest weight, otherwise when parameter p=0.9 then the first 10 elements ranked have 85% of the evaluation weight. If we want to consider giving weight to more than 10 elements, then we would increase the value of p as we could see in the 3rd and 4th scenarios.

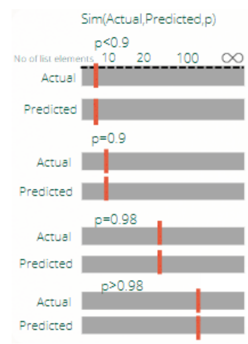

Fig. 4. RBO similarity score [9].

VII. Testing Procedure

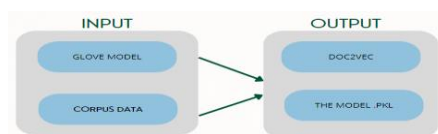

Fig. 5. Step 1, global vectors for word embeddings (GloVe).

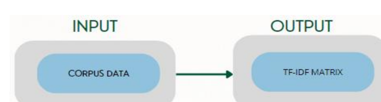

Fig. 6. Step 1, term frequency-inverse document frequency (TF-IDF method).

We have tested our hybrid recommendation system by using GloVe embeddings and the TF-IDF method. In Fig. 5 about the GloVe embeddings, we took as input the glove pretrained model and the corpus data jobs or resumes, and the output of this first step was the Doc2Vec representation and the binary representation of the model. If we compare the same step for the TF-IDF matrix we have used only the corpus data to generate that matrix, Fig. 6.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on July 12,2025 at 04:33:19 UTC from IEEE Xplore. Restrictions apply.

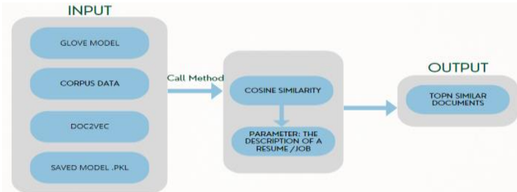

Fig. 7. Applying cosine similarity, GloVe embedding.

Fig. 7 shows the process of the application of cosine similarity, by using GloVe. We used as input the GloVe pre-trained model, the corpus data the Doc2Vec, and the generated model. pkl. Next, we applied cosine similarity by giving as a parameter the description of a job or a resume and by producing in this way top N similar documents where n might be 10, 20,50,100.

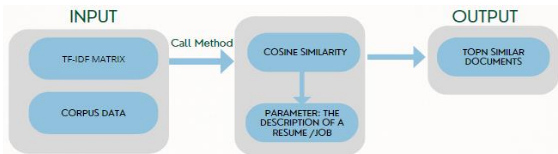

Fig. 8. Applying cosine similarity, the TF-IDF matrix.

The TF-IDF representation differs from the previous method because it uses fewer inputs in this phase and the other part is the same as can be inferred from Fig. 8.

The last part of our testing process is the RBO application on our data, Fig. 9. At this moment we have used our manual ranked lists created during the preprocessing phase and the lists generated by using our system, to find how similar are they both.

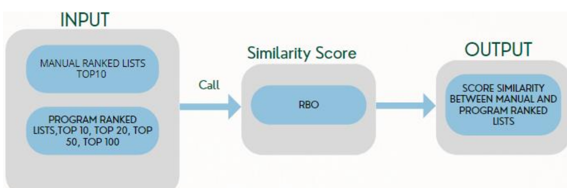

Fig. 9. RBO scoring.

VIII. Results

A. Global Vectors for Word Embeddings (GloVe), Score Visualization

After calculating the Rank Biased Overlap (RBO) score between the manual ranking list and the prediction ranking list by selecting the top 10, top 20, top 50, and top 100 of the resume or job predictions we have created box plots to understand better the distribution of each score.

In Fig. 10, our first scenario contains the ranking of similarities between each resume. Some inferences from the above graph:

- The distribution of RBO scoring between manual lists and prediction list top 10, top 20, and top 100 is almost the same, with data positively skewed.
- The distribution of RBO scoring between manual lists and prediction list top 50 is quite symmetric, which means that the median score value separates almost a similar portion of the scores.
- All the graphs contain outliers on the upper side of the box.

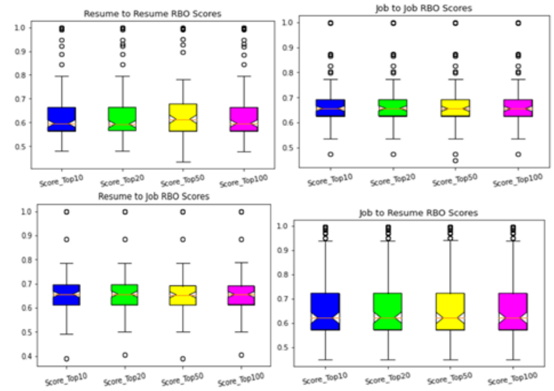

Fig. 10. GloVe representation: Box plots of the RBO scores.

Job-to-job scenarios and resume-to-job have almost the same distribution. Some observations from the above graph:

- The distribution of RBO scoring between manual lists and prediction lists top 10, top 20, top 50, and top 100 is quite symmetric.
- Box plots contain some outlier values on the lower side of the plot and more on the upper side.

Our last scenario contains the ranking of similarities of each job with respect to resumes. Some observations from the above graph:

- The distribution of RBO scoring between manual lists and prediction list top 10, top 20, top 50, and top 100 is almost the same, with data positively skewed.
- Box plots contain some outlier values on the upper side of the plot.

B. Term Frequency-Inverse Document Frequency (TF-IDF), Score Visualization

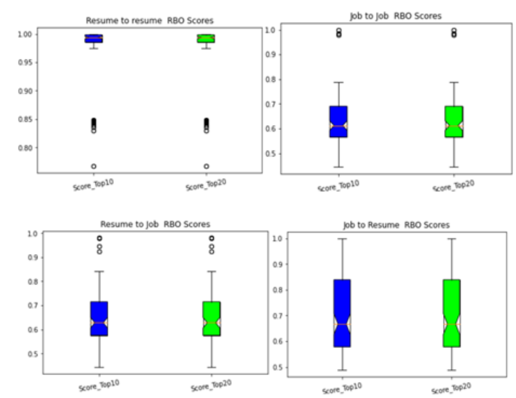

Fig. 11. TF-IDF presentation; Box plots of the RBO scores.

In Fig. 11, we have presented the top 10 and top 20 because the values of the RBO for the top 50 and top 100 are the same

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on July 12,2025 at 04:33:19 UTC from IEEE Xplore. Restrictions apply.

as the top 10. If we revise resume to resume for the previous method and TF-IDF method, we can note how much these graphs change. Some inferences from resume-to-resume scores:

- The distribution of RBO scoring between manual lists and prediction list top 10 and top20 is the same.
- It is evident that the majority of values are nearly one. That means human-ranked resumes and program-ranked resumes are almost similar.
- Since the pattern values are higher, all the values below 0.85 are considered outliers.

Based on TF-IDF representation our second scenario contains the ranking of similarities between each job and the third one contains the ranking of similarities of each resume with respect to jobs. Some observations from the above graphs:

- The distribution of RBO scoring between manual lists and prediction list top 10, and top 20 is the same, there are no changes in the values with data positively skewed
- Box plots contain some outlier values on the upper side.

Lastly, based on TF-IDF representation contains the ranking of similarities of each job with respect to resumes. Some observations from the above graph:

- The distribution of RBO scoring between manual lists and prediction list top 10, and top 20 is the same with data positively skewed.
- There are no outliers.
- Most of the values are below 0.85.

C. *Results*

TABLE II. TOP 10 MEDIAN SCORE

	Resume to Resume	Job to Job	Resume to Job	Job to Resume
Glove	0.5951	0.6555	0.6558	0.6208
TF-IDF	0.9937	0.6125	0.6279	0.6668

Table 2. shows the comparison of the median RBO score between Global Vectors for Word Embeddings (GloVe) and Term Frequency-Inverse Document Frequency (TF-IDF) representation. This table contains the median RBO score as a result of the comparison of manual ranked lists with the top 10 predicted lists for each scenario. Apart from the resume-to-resume scenario for the TF-IDF representation which has a significant difference from the GloVe embedding method, the other median RBO scores have slight changes between the two methods for each scenario. In the case of resume-to-resume, from our experiment, we have noticed that the similarity between the manual ranked list and the predicted list is higher, which is the reason why the median RBO score is the highest for the TF-IDF representation.

IX. CONCLUSION

In this paper, we took into consideration the description of resumes and vacancies. We selected 10 relevant jobs and candidate profiles to create manual human-ranked lists. Then we applied Rank Biased Overlap (RBO) to compare the lists generated by the hybrid recommendation system with the top 10 human-ranked lists for candidate and job descriptions. In our experiments we used two methods for word representation: Term Frequency-Inverse Document Frequency (TF-IDF) and Global Vectors for Word Embeddings (GloVe). The evaluation conducted showed there was a considerable similarity (minimum median value 0.59), between human-ranked lists and system-ranked lists. Furthermore, the median RBO between the two methods was almost the same, apart from a resume-to-resume scenario where the median RBO for TF-IDF is the highest value because of the high similarity between top 10 system-ranked lists and top 10 human-ranked lists.

REFERENCES

[1] Published by S. Dixon, & 11, J. (2022, July 11). Canada LinkedIn users by age group 2022. Statista. Retrieved September 13, 2022, from https://www.statista.com/statistics/1134991/linkedin-user-share-in-canada-age/#:~:text=As%20of%20June%202022%20in,of%20professional%20networking%20service%20LinkedIn

[2] Team, C. C. (2022, May 9). ZipRecruiter vs indeed. CareerCloud. Retrieved September 13, 2022, from https://www.careercloud.com/news/ziprecruiter-vs-indeed

[3] Talent.com. Betterteam. (n.d.). Retrieved September 13, 2022, from https://www.betterteam.com/ca/talent-com#:~:text=Talent.com%20is%20available%20in,free%20trial%2C%20no%20card%20required

[4] Nigam, A., Roy, A., Singh, H., & Waila, H. (2019). Job recommendation through progression of job selection. 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS). https://doi.org/10.1109/ccis48116.2019.9073723

[5] Ramanath, R., Inan, H., Polatkan, G., Hu, B., Guo, Q., Ozcaglar, C., Wu, X., Kenthapadi, K., & Geyik, S. C. (2018). Towards deep and representation learning for talent search at linkedin. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. https://doi.org/10.1145/3269206.3272030

[6] Ozcaglar, C., Geyik, S., Schmitz, B., Sharma, P., Shelkovnykov, A., Ma, Y., & Buchanan, E. (2019). Entity personalized talent search models with tree interaction features. The World Wide Web Conference. https://doi.org/10.1145/3308558.3313672

[7] Lu, Y., El Helou, S., & Gillet, D. (2013). A recommender system for job seeking and recruiting website. Proceedings of the 22nd International Conference on World Wide Web. https://doi.org/10.1145/2487788.2488092

[8] Pennington, J. (n.d.). Glove: Global vectors for word representation. Retrieved September 13, 2022, from https://nlp.stanford.edu/projects/glove/

[9] Qaiser, S., & Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to documents. International Journal of Computer Applications, 181(1), 25–29. https://doi.org/10.5120/ijca2018917395

[10] Le, Q.;& Mikolov, T.. (2014). Distributed Representations of Sentences and Documents. Proceedings of the 31st International Conference on Machine Learning, in Proceedings of Machine Learning Research 32(2):1188-1196 Available from https://proceedings.mlr.press/v32/le14.html.

[11] Job recommendation system using hybrid filtering - itm-conferences.org. (n.d.). Retrieved September 1, 2022, from https://www.itm-conferences.org/articles/itmconf/pdf/2022/04/itmconf_icacc2022_02002.pdf

[12] Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. ACM Transactions on Information Systems, 28(4), 1–38. https://doi.org/10.1145/1852102.1852106

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on July 12,2025 at 04:33:19 UTC from IEEE Xplore. Restrictions apply.