Linked to Hired

Ali Sarosh Bangash
Computer Science and Engineering
University of South Florida
Tampa, USA
alibangash@usf.edu

Krish Veera
Computer Science and Engineering
University of South Florida
Tampa, USA
krishv@usf.edu

Tanay Jambli
Computer Science and Engineering
University of South Florida
Tampa, USA
tanayj@usf.edu

Guillermo Medina Nieves
Computer Science and Engineering
University of South Florida
Tampa, USA
gmedinanieves@usf.edu

Sumit Jadhav
Computer Science and Engineering
University of South Florida
Tampa, USA
sumitjadhav@usf.edu

Aravinda Venkatesan
Computer Science and Engineering
University of South Florida
Tampa, USA
aravindav@usf.edu

***Abstract*—The study investigates using language models to extract technical skills from LinkedIn profiles' 'About' sections. Two hypotheses are proposed: H_0 suggests no improvement between random word selection and base model keyword extraction, while H_a predicts substantial enhancement with base model keyword extraction. Findings reveal significant variations in profile completeness among occupations, with sales/marketing professionals exhibiting higher completeness. Gender disparities are noted, with men providing more personal information. Methodologically, the study employs BART-Large model for untrained and finetuned keyword extraction. Results demonstrate the superiority of the BART model over random selection and further enhancement with finetuning. The study contributes to understanding LinkedIn's use in recruitment processes and suggests future research directions.**

I. Motivation

Our motivation for this project is the current state of the Software Engineering Job Market which presents lots of challenges for both seasoned professionals and especially for newcomers. Among these challenges are constant rejections, layoffs, on top of a phenomenon of skill abundance. Despite the increasing demand for software engineers, the competition for coveted positions often results in a high rate of rejections, leaving many qualified candidates disillusioned and discouraged. The same can be said about layoffs, whether due to economic downturns, company restructuring, or technological shifts, adds an element of uncertainty and insecurity to the career trajectories of even experienced software engineers. There is also an abundance of required technical skills, which in turn causes employers to seek candidates with a diverse set of proficiencies. These skill expectations can overwhelm job seekers, leading to feelings of inadequacy or uncertainty about how to effectively navigate the job market. As a result, we wanted software engineers to have a way to search for the most relevant technical skills in the software engineering job market so they can focus their efforts on a specific set of skills and making navigating the job market easier. This is where we come in with a system, we developed capable of accurately extracting technical skills mentioned in LinkedIn postings. LinkedIn has become one of the most fundamental platforms for professional networking, boasting millions of users and an extensive database of job postings so by extracting the most prevalent technical skills of the job postings we can understand and get a good grasp of the market demand.

As for our research hypothesis, by utilizing language models inferences and fine-tuning techniques on a language model, we aim to develop a system capable of accurately extracting technical skills mentioned in LinkedIn postings' 'About' sections. We believe this approach will yield :

$H_\mathbf{0}$: No improvement in identifying relevant skills between random word selection vs. base model keyword extraction

$H_\mathbf{a}$: Substantial improvement in identifying relevant skills between random word selection vs. base model keyword extraction

II. Literature Review

A. LinkedIn and recruitment: how profiles differ across occupations

The paper aims to investigate the focal elements of LinkedIn profiles for hiring professionals and analyze profile variations across different occupations. It comprises two phases: qualitative interviews with NYC hiring professionals to identify key profile variables, followed by a quantitative analysis of profiles from HR, sales/marketing, and I/O psychology based on the identified variables. Findings reveal significant occupation-based differences in profile completion, with sales/marketing professionals exhibiting higher completeness. Gender disparities were noted, with men providing more personal information. The study underscores LinkedIn's incompleteness relative to traditional resumes. Theoretical implications include insights into LinkedIn's role in recruitment, while practical implications emphasize holistic profile assessment for recruiters and the importance of completeness for job seekers. Limitations include small sample sizes and future research recommendations for broader scope exploration.

B. Jobseeker-industry matching system using automated keyword selection and visualization approach

According to research presented by the Indonesian Journal of Electrical Engineering and Computer Science, "Jobseeker-industry matching system using automated keyword selection and visualization approach", the researchers set out to create some sort of visualization of extracted skills and present it to the graduates, so that it may help them in finding the right job that they need. As for our research, we are concerned with just how well a specific model performs, as is, and then determining how well a fine-tuned version of the model performs and compare it to the base. We are not concerned in developing visual to help graduates.

C. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension

The paper introduces BART, a denoising autoencoder designed for pretraining sequence-to-sequence models. BART's training process involves corrupting text with a customizable noising function and then training a model to reconstruct the original text. It adopts a standard Transformer-based architecture, which effectively generalizes previous pretraining schemes like BERT and GPT. The study evaluates various noising approaches and identifies optimal performance when employing techniques such as random sentence shuffling and a novel in-filling scheme. BART demonstrates strong efficacy in text generation tasks and performs comparably well in comprehension tasks. It achieves state-of-the-art results in abstractive dialogue, question answering, summarization tasks, and machine translation, surpassing previous benchmarks with notable improvements. A series of ablation experiments are conducted within the BART framework to analyze factors influencing end-task performance. Overall, the paper presents a comprehensive evaluation of BART's capabilities, showcasing its versatility and effectiveness across various natural language processing tasks.

III. Data Selection and Preparation

We obtained the dataset for our study from a publicly available Kaggle repository dedicated to software engineering job postings. This repository offered a comprehensive collection of job postings sourced from various platforms, providing a diverse range of data for analysis.

To ensure the dataset's suitability for our study, we conducted a manual curation process to select a subset of entries. From the original dataset comprising approximately 10,000 entries, we randomly sampled 600 job postings for analysis. Each selected entry's 'about' section, containing essential job role and requirement information, was meticulously copied and compiled. This manual extraction process ensured data accuracy and relevance to our research objectives. Subsequently, the curated dataset was utilized to conduct two key tasks: random baseline word selection and keyword selection using BART. These tasks were pivotal in evaluating the effectiveness of language model inferences and fine-tuning techniques for extracting technical skills mentioned in the job postings. Through these analyses, we aimed to gain insights into the performance of BART in identifying relevant skills within the software engineering domain. The curated subset was chosen to strike a balance between data volume and analytical feasibility, facilitating robust analysis within the constraints of available resources.

IV. Methods

The methods to disprove the null hypothesis that there would be no difference between random word selection, base model keyword extraction, and finetuning model keyword extraction were as follows.

The Large Language Model (LLM) used for this case was Facebook's open-source model BART-Large. LLMs are advanced artificial intelligence systems designed to understand and generate human-like text by learning from vast amounts of textual data. The BART-Large model is a sequence-to-sequence model, using a combination of encoder and decoder components. Despite being relatively small, trained on 406M data points with a max token limit of 1024, its compact size and low hardware cost make it a realistic and sufficient model to use. It was utilized in two ways: untrained keyword extraction with TF-IDF vectorization and trained keyword extraction with finetuning.

The dataset used for both untrained and trained keyword extraction was from the previous section, which detailed the creation and structure of the dataset.

For random word selection, a Python code split individual texts into lists of words, and numbers were chosen randomly and inserted into the lists using the Python Random module to ensure unbiased and unpredictable word selection.

Base model keyword extraction, termed "inferencing," involved preprocessing text data, extracting keywords using TF-IDF vectorization, and managing datasets for natural language processing tasks. Relevant libraries for this implementation included "datasets," "transformers," "pandas," "nltk," and "sklearn," facilitating handling and preprocessing datasets for machine learning tasks. Functions were implemented using these libraries: Load Data for standardizing input data format and Perform Keyword Extraction for computing TF-IDF scores, tokenizing text, limiting output to top features, and fitting the model to each text. The main execution flow involved dataset preparation, repeatedly loading data from different CSV seed files, converting them into pandas DataFrames and then into Dataset objects. Each dataset was split into training and testing sets with a 33% test size using train_test_split to ensure uniformity and reproducibility. Keyword extraction involved extracting keywords using the Perform Keyword Extraction function on each split, particularly the test set, storing extracted keywords in lists, and printing them out to visualize the most significant terms per document according to TF-IDF.

The finetuning of the BART-Large model involved providing the dataset to make the model understand specified keywords to look for and in what context to find them. The

Python code, utilizing the transformers library, fine-tuned a sequence-to-sequence model on a keyword extraction task. Relevant libraries included ”datasets,” ”accelerate,” ”transformers,” ”pandas,” ”PyTorch,” ”NumPy,” and ”sk-learn.” The data loading and preparation involved loading a dataset from a CSV file into a pandas data frame, splitting it into training, validation, and test sets, and converting data frames into Hugging Face Dataset objects grouped into a DatasetDict. Model setup involved setting up a BART-Large model and tokenizer. Text tokenization and processing ensured inputs did not exceed the model's maximum input size. Training involved specifying training settings with TrainingArguments and training the model using a Trainer object. Testing involved tokenizing and truncating the testing set to ensure processing by the model, running the testing data, and printing output alongside actual keywords for visual comparison. Finally, a hit rate function calculated hit rate scores, mean, standard deviation, and variance using NumPy functions.

V. Results

After conducting the random baseline test, the inference test, and the subsequent fine-tuning test, we now have the following results from our work. According to our random baseline test, the mean hit rate and standard deviation for each seed is displayed in Figure 1.

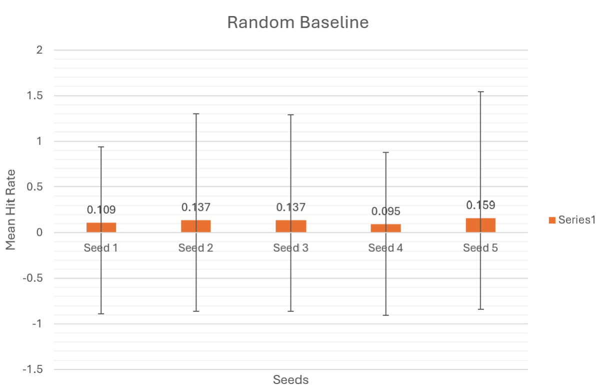

Fig. 1. Random Baseline Word Selection Results

Subsequently, the mean hit rate and standard deviation for the BART inference test is shown in Figure 2.

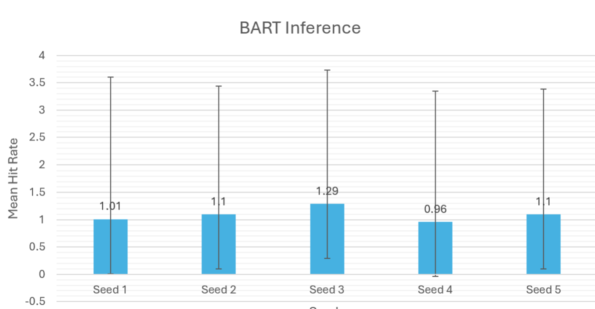

Fig. 2. BART Inference Results

If we combine the inference and baseline graph into one visual, we obtain Figure 3.

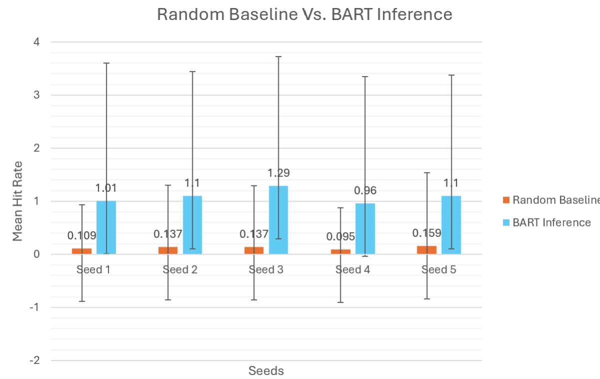

Fig. 3. Combined Results Visual

From the results, we can see the inference model outperforming the random baseline model by a definite margin. Across all five seeds, the mean hit rate for the BART model was performing well above the random baseline test. If we examine the the BART Fine-Tune model, we obtain Figure 4.

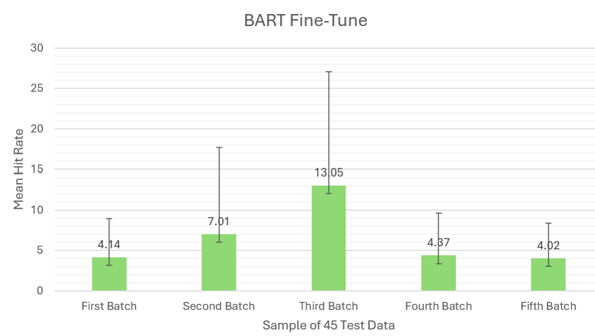

Fig. 4. BART Fine-Tune Visual

We've separated the BART Fine-Tune result from the preceding models because we used only one train-test split for our fine-tune model. We were more concerned as to how maximally optimized could we create a fine tune model that can achieve the best hit rate possible. For this, we had BART-large be trained on a train-test split of 555 to 45. We then split the results into different batch (five batches, with each batch being of size nine test data points). As we can see, the fine-tuned model overwhelmingly outperformed the prior models, albeit the different train and test split.

VI. Discussion

Our analysis revealed that there was a statistically significant difference between the two models where the base model keyword extraction shows significant improvement over the random word selection, providing evidence to reject the null hypothesis in favor of the alternative hypothesis. These results were expected as random word selection will always be worse just by its very nature. Likewise, it is to be expected that the base model keyword extraction will perform better, while this model hasn't been trained yet to find skills, its function is to find keywords in the text, and it is very likely that it will consider skills as keywords based on the emphasis most 'about' sections have on them. The Bart Finetuning model shows substantial advancements over the base model's keyword extraction, surpassing even the initial gap observed between random word selection and base model performance. This still falls under our expectations since in the Bart Finetuning we make the model understand that the specified keywords we look for are the skills and in what context it would find them in. While our study has showed promising insights into skill extraction methodologies, it is essential to acknowledge and address the limitations inherent in our approach. One of our biggest limitations is the constrained size of our dataset, a factor that inevitably impacts the robustness and generalizability of our results. Our current dataset comprises a modest selection of 600 'about' sections sourced from the original Kaggle dataset which contained close to 10,000 job postings. While this sample size has enabled us to observe notable enhancements in skill extraction performance, particularly with the implementation of Bart Finetuning, it is evident that our findings could be further strengthened with access to a larger and more diverse dataset. Therefore, if it were to be done with an even bigger dataset our models could better capture the intricacies of skill representation and context, leading to more robust and reliable skill extraction results. Additionally, another thing that can be improved is Finetuning the BART model on unique seeds which would enhance the accuracy and efficacy of our skill extraction framework. By leveraging unique seeds—distinctive identifiers or markers associated with specific skills or skill categories—we can tailor the fine-tuning process to focus on refining the model's understanding and recognition of these key elements. This targeted approach allows the model to hone its ability to detect and contextualize skills within 'about' sections with greater precision and granularity. Furthermore, fine-tuning on unique seeds enables the model to learn subtle nuances and variations in how skills are articulated and represented across different contexts, industries, and professions. As a result, the fine-tuned BART model becomes more adept at discerning relevant keywords and phrases indicative of sought-after skills, leading to improved accuracy and reliability in skill extraction tasks.

REFERENCES

[1] J. Zide, B. Elman, and C. Shahani-Denning, "LinkedIn and recruitment: How profiles differ across occupations," Employee Relations, https://www.emerald.com/insight/content/doi/10.1108/ER-07-2013-0086/full/html (accessed Apr. 29, 2024).

[2] N. Kamaruddin and A. W. A. Rahman, Jobseeker-Industry Matching System using automated keyword selection and visualization approach, https://www.researchgate.net/publication/332140232_Jobseeker-industry_matching_system_using_automated_keyword_selection_and_visualization_approach (accessed Apr. 29, 2024).

[3] M. Lewis et al., "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," arXiv.org, https://arxiv.org/abs/1910.13461 (accessed Apr. 29, 2024).

VII. CONTRIBUTIONS

For contribution, **Ali-Sarosh Bangash** was partially responsible for obtaining the BART model inference results and the generating the five different seeds for the models used. **Krish Veera** was responsible for doing the Excel manipulation and the random baseline model test. **Guillermo Medina Nieves** was also primarily responsible for obtaining the BART model inference results. **Tanay Jambli** was responsible for fine tuning the BART-Large model. **Aravinda Venkatasan** and **Sumit Jadhav** were both responsible for the literature review.