Biraj Silwal

Kathmandu, Nepal · silwalbiraj@gmail.com · jarib047.github.io · (+977) 9808725497

Education

Tribhuwan University - Institute of Engineering, Pulchowk Campus — Lalitpur, Nepal
Master of Science in Computer Engineering, Data Science and Analytics — Feb 2022 – Jun 2024
Thesis: Interpreting Pre-trained Word Embeddings using Parts of Speech Metarepresentations
GPA: 84.13/100 ⇒ Equivalent WES iGPA: 4.00/4.00

Tribhuwan University - Kantipur Engineering College — Lalitpur, Nepal
Bachelor's Degree in Computer Engineering — Jan 2017 – Jan 2022
GPA: 65.7/100 ⇒ Equivalent WES iGPA: 3.53/4.00

Work Experience

Nepal Telecom — Kathmandu, Nepal
Computer Engineer — Feb 2024 – *current*

- Developed and managed a platform to generate year-on-year employee performance reviews.
- Managed and operated GIS tools for extending optical network availability.

Cedar Gate Technologies — Lalitpur, Nepal
Software Engineer, Data Processing — Apr 2022 – Oct 2022

- Developed an authentication system for the Data Processing department and automated the generation of daily data deliverables.
- Managed the Data Processing platform and ensured operational continuity by monitoring Production Engineers' workflows.

Publications

- **B. Silwal**, "Interpreting Pre-trained Word Embeddings using Parts of Speech Metarepresentations, *Graduate thesis*, https://jarib047.github.io/files/thesis.pdf.
- **B. Silwal**, "Syntactic Representations Enable Interpretable Hierarchical Word Vectors, *arXiv preprint*, https://arxiv.org/abs/2411.08384.
- **B. Silwal**, 'Fine-Tuning Small Embeddings for Elevated Performance, *arXiv preprint*, https://arxiv.org/abs/2411.18099.

Research

Editing Embeddings for Fine-Tuning LLMs — Sep 2024 - *current*

- Reconfigured various Large Language Models to isolate the generated word embeddings and post-process them as required.
- Applied various post-hoc methods based on biasing and projection, following the works of *Representation Editing* and *Representation Finetuning* to fine-tune embeddings for downstream tasks.
- Analyzed the impact of these methods on various LLM layers using intervention mechanism to establish mechanistic interpretability.

Syntactic Representations & Hierarchical Word Vectors — July 2023 - Feb 2024

- Extracted syntactic regularities from pre-trained word vectors via means of post-processing to create a novel notion of Syntactic Representation, which is interpretable in terms of parts of speech.

- Created Hierarchical word vectors, which work as a representation of the hierarchical aspect of the human learning process, by associating Syntactic Representations with their respective pre-trained word vectors.
- Evaluated and reported the performance of various forms of the Hierarchical vectors against the pre-trained vectors, in an array of benchmark evaluation tests.

Fine-tuning pre-trained BERT models using social media data Dec 2022 - Mar 2023

- Evaluated various BERT models pre-trained on Nepali language to establish a baseline model and an oracle.
- Extracted, stored and pre-processed social media data i.e posts and tweets in Nepali language.
- Fine-tuned the identified baseline model using the pre-processed data and finally compared the results with that of the baseline model and the oracle.

Predicting the outcomes and scorelines of football matches Dec 2019 - Feb 2021

- Identified, implemented and reported the use of multiple classification and regression algorithms to predict the outcomes and scorelines of various football matches.
- Submitted as a partial requirement for the completion of the Bachelors' degree.

Awards

Four year merit scholarship

- Awarded for exceptional performance in the 2016 Institute of Engineering entrance examinations; ranked **297**th out of nearly **15000** applicants.
- Covered all admissions, tuition and exam fees for the duration of the Bachelors' degree.

Skills

Technical Skills

- **Proficient:** Python, NumPy, Scikit-Learn, PyTorch, Tensorflow, Git, LaTeX, SQL, Keras, Pandas
- **Familiar:** Linux, C, C++, R, Apache Spark, Matplotlib, Docker

Relevant Courses

- Artificial Intelligence, Machine Learning & Computational Intelligence, Probability and Statistics, Modern Natural Language Processing, Fundamentals of Data Science and Analysis, Data Mining, Big Data Technologies, Big Data Analytics, Information Visualization, Optimization Theory and Techniques