Exploring Natural Language Inference: A Comparative Analysis of SVM and BiLSTM Methods

MANCHESTER 1824
The University of Manchester

Group 47: Yida Zhuge & Haojing Tong

The University of Manchester: Department of Computer Science

ABSTRACT

In this study, we analyze Logistic regression and Bidirectional Long Short-Term Memory (BiLSTM) networks for Natural Language Inference (NLI), aiming to improve logical relationship prediction between premise and hypothesis. Utilizing a 26K premise-hypothesis pair dataset, our Logistic regression approach reached a validation accuracy of 66% and a BiLSTM approach achieved a higher validation accuracy of 68%.

INTRODUCTION

Natural Language Inference is a fundamental task in natural language processing involves determining the logical relationship between pairs of text fragments, typically referred to as the premise and the hypothesis. The goal of NLI is to infer whether the meaning of the hypothesis can be logically derived or inferred from the meaning of the premise.

The motivation behind NLI lies in its relevance to various downstream NLP applications such as question answering, information retrieval, dialogue systems, and sentiment analysis.

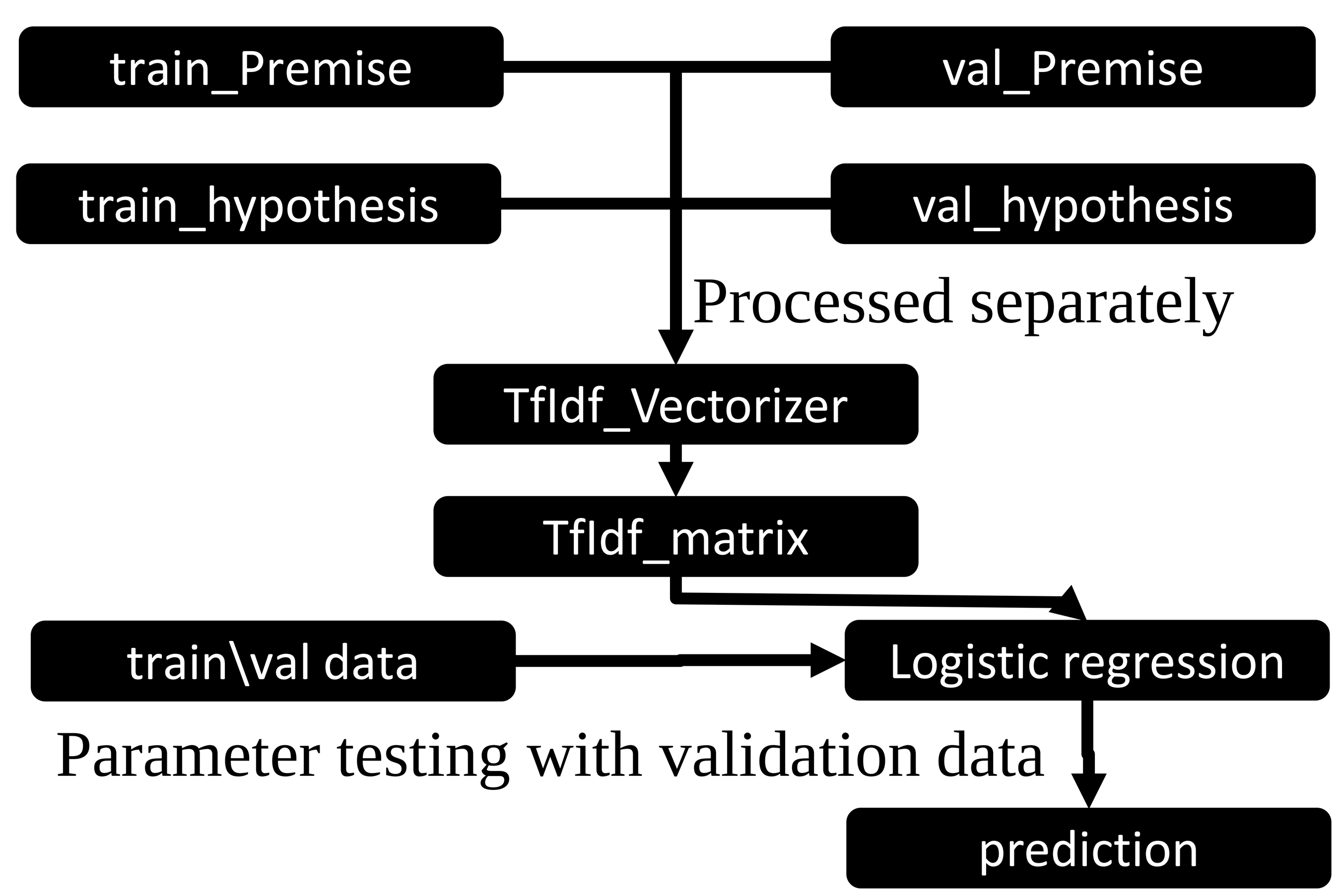
Moreover, NLI serves as a benchmark task for evaluating the capabilities of NLP models in understanding and reasoning about natural language. Progress in NLI research contributes to advancements in machine comprehension, logic-based reasoning, and semantic understanding, ultimately driving the development of more sophisticated NLP systems capable of handling complex language understanding tasks.

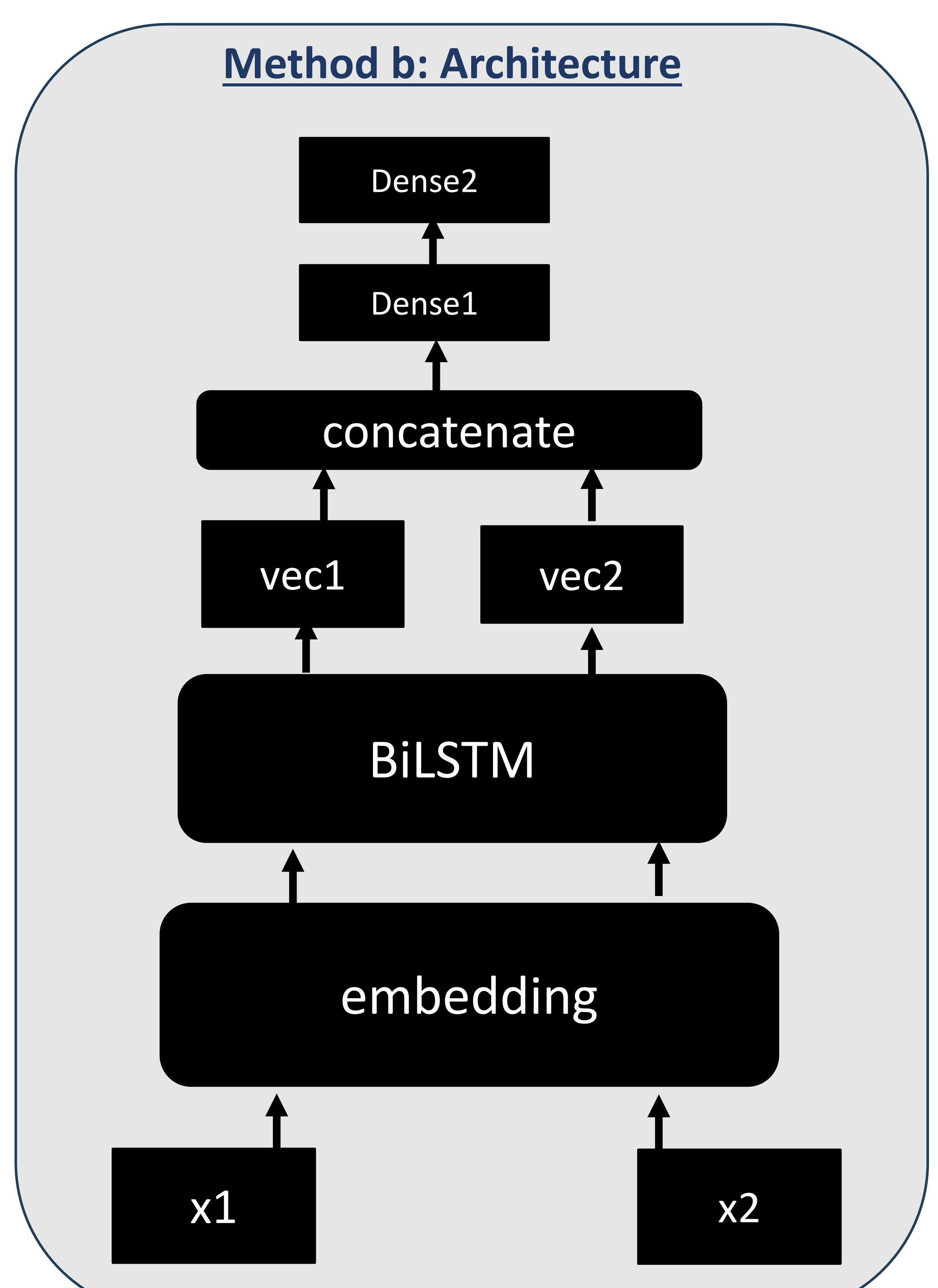
Materials & Methods

Premise	hypothesis	label
Year is 1800	It is 1800 year	1
…	…	…
It's a circle	It's a square	0

Train data(26k pairs with label)

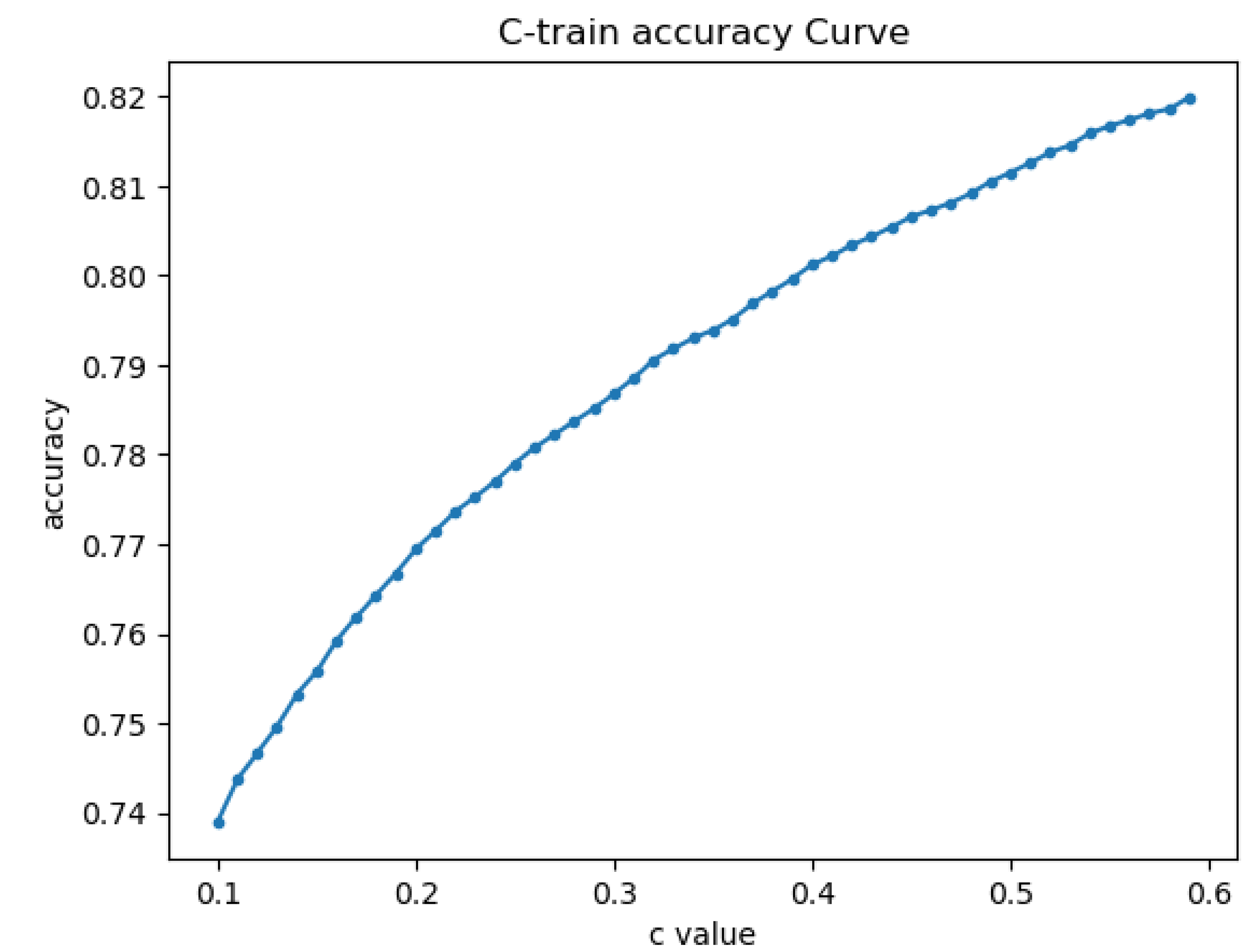
Method a: Logistic regression

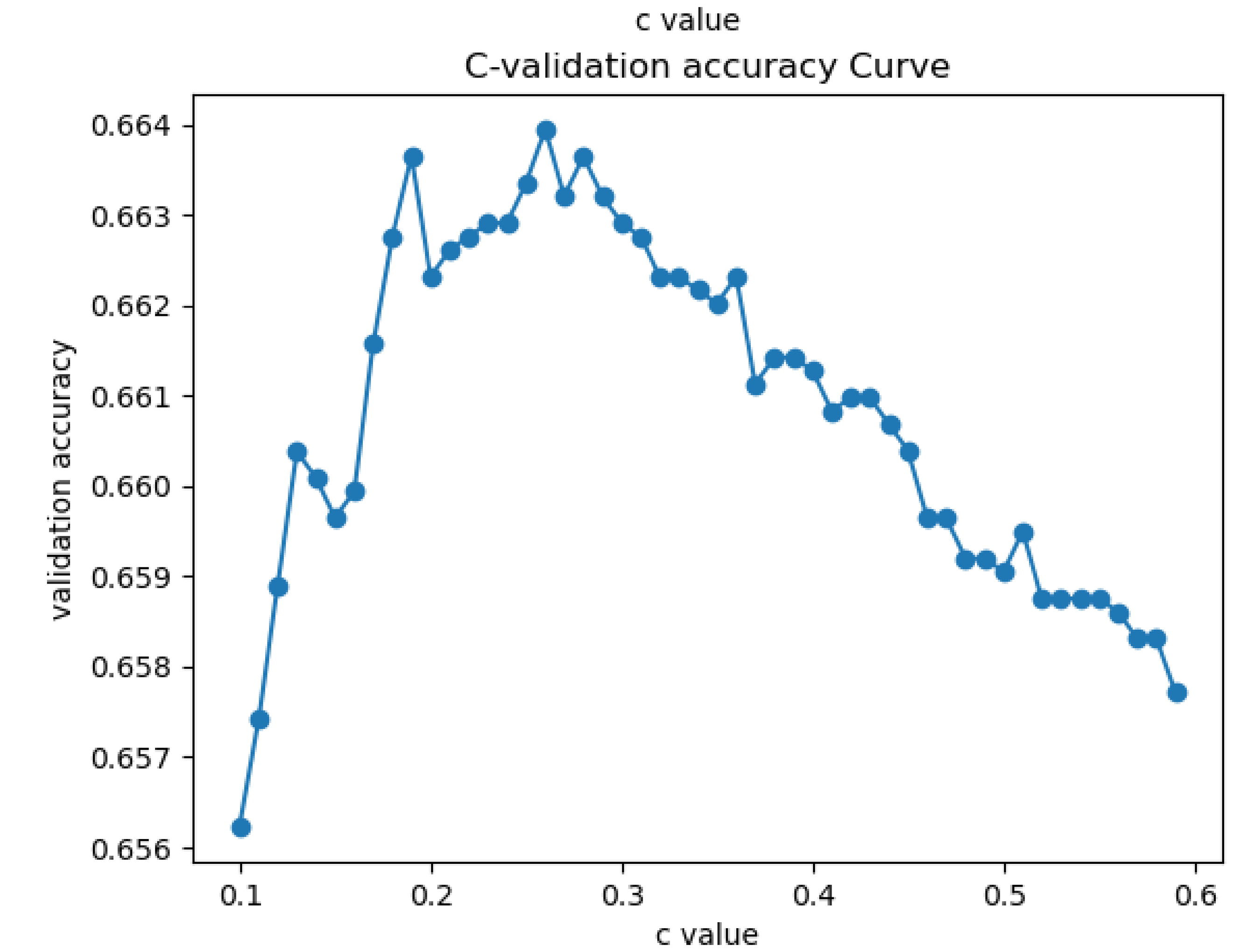




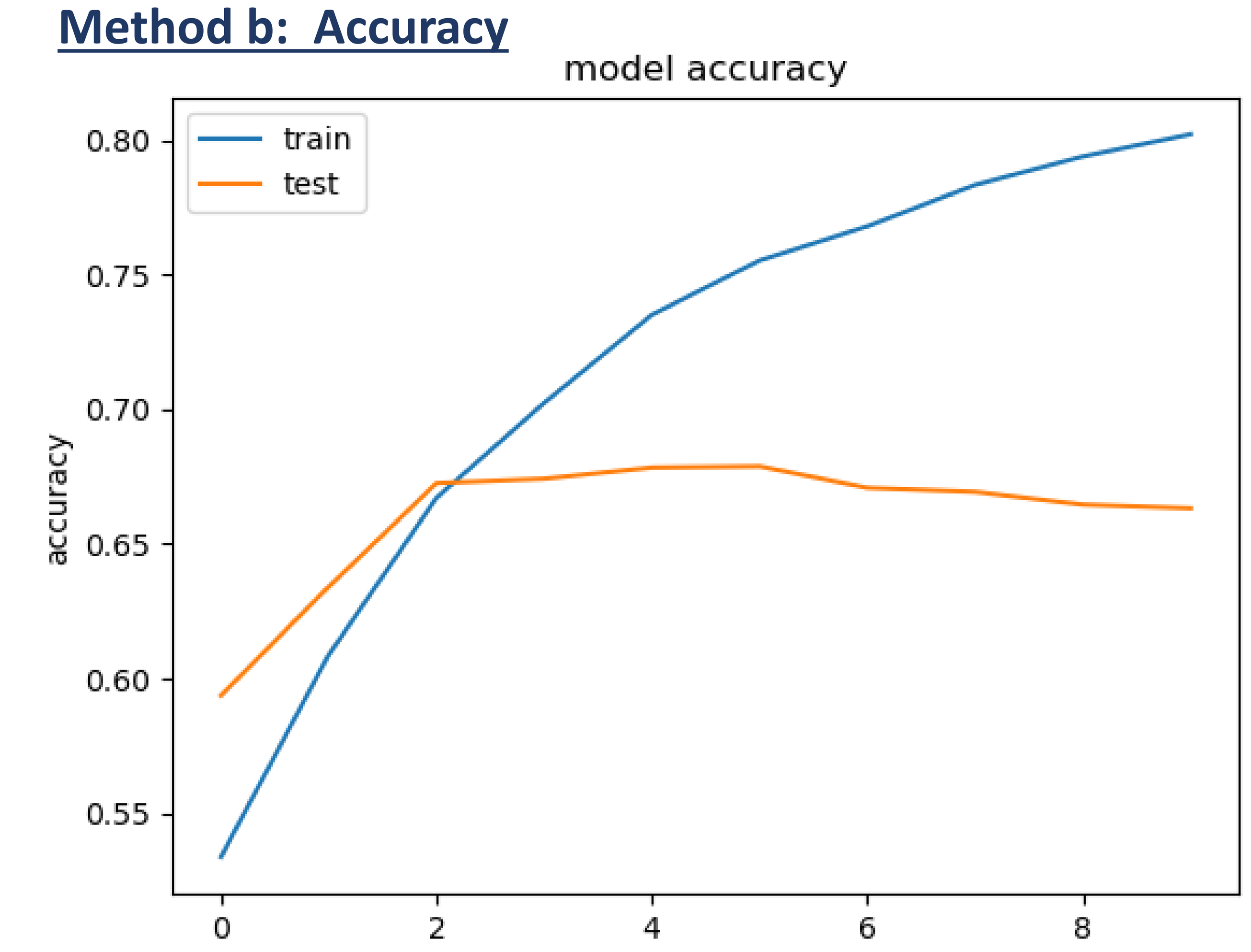
RESULTS

Method a: Logistic regression Evaluating Mitrix





Method b: Accuracy



CONCLUSIONS

This project aimed to leverage a Logistic regression and BiLSTM architecture to enhance the accuracy of Natural Language Inference. Our model was trained using a 26K dataset.

Before Logistic regression, TfIdf feature matrix is created for each word. Premise and hypothesis are trained separately and merged then to get a better validation result. The validation accuracy reached 66% when regularization parameter = 0.26 with a 76% train accuracy.

BiLSTM achieved a final training accuracy of 80% and a validation accuracy of 68.5% at epoch 4.

The model's robust performance suggests its applicability in real-world Natural Language Inference tasks. Future work could explore the BERT method mentioned in the instruction, which has 78% accuracy as baseline model.

Comparing the time cost of training model, Logistic regression cost 5 seconds per modelling while BiLSTM cost 6 minutes in the experiment. BiLSTM have better performance and higher time costs.

REFERENCES

[1] Aarne Talman, Anssi Yli-Jyr ̈a, and J ̈org Tiedemann. "Natural language inference with hierarchical bilstm max pooling architecture". In: arXiv preprint arXiv:1808.08762 (2018).

[2]Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb). 2014 Feb 15;24(1):12-8. doi: 10.11613/BM.2014.003. PMID: 24627710; PMCID: PMC3936971.