

# Assignment 2 Report

Michele Luca Contalbo, Ehtsham Akhter, Vida Zahedi, Yiran Zeng

December 18, 2021

## Abstract

This paper is a report which describes the second assignment given at the Natural Language Processing course at the University of Bologna. The assignment is centered on a small portion of the **fact checking** problem, which aims to establish whether a given statement is supported or refuted by a set of evidences. We were asked, given the **FEVER** dataset, to create some neural network models able to face this problem and compare their results. The tested models differ on how the sentence embeddings are extracted and combined together. Based on the metrics used (accuracy, f1-score, recall and precision), among all the architectures tested, the RNNs models provide better results.

## Keywords

fact checking, FEVER, RNN, sentence embedding

- **Verifiable**: whether the claim is verifiable or not;
- **Claim**: fact to be verified;
- **Evidence**: data structure composed by IDs that can be associated to the claim;
- **Label**: whether the Evidence Supports, Refutes or has NotEnoughInfo on the claim.

For this assignment, we are not interested in non verifiable claims, hence some preprocessing is done in order to filter out data which is not useful for our problem. In particular, we will not consider the **Verifiable** feature and focus only on **Evidence** supporting or refuting its corresponding **Claim**. Also, the **Evidence** feature has been modified in order to contain text and not IDs.

## 1 Introduction

In this section we will provide information about the assignment, the dataset structure and its preprocessing.

In this assignment we need to check if some claims are supported or refuted by set of evidences. The dataset used is **FEVER**, which as been devised with the idea to aid creating software able to tackle false information coming from unreliable sources.

### 1.1 Dataset Structure and Preprocessing

The dataset is about facts taken from Wikipedia documents and it consists of 185,445 claims manually verified and classified as **Supported**, **Refuted** or **NotEnoughInfo**. The dataset is composed by the following features:

- **ID**: id associated to the fact to verify;

### 1.2 Text Preprocessing

Standard text preprocessing strategies have been applied with the aim to reduce noise and focus on more semantically meaningful words. Each word in all sentences, after having created the vocabulary, have all been converted to their respective vocabulary IDs. The sentences have been padded to 122, which is the maximum sentence length between **Claim** and **Evidence** in training, validation and testing set.

## 2 Architectures and Evaluation Methods

### 2.1 General Structure

Each model differs from each other on how sentence embedding are extracted and combined together. Each architecture has the following structure:

Architecture General Structure	
Layer name	Input $\rightarrow$ Output size
Embedding	122 $\rightarrow$ (122,50)
Sentence Embedding	—
Merging	—
Dense	50 or 100 (concatenation) $\rightarrow$ 2

The Embedding Layer is fed with 2 inputs of size 122. The embedding matrix, used in the first layer, has been obtained from **GloVe**<sup>1</sup> with **embedding size** of 50. In the last one, **softmax** as been used as activation. In each layer, **masking** has been applied to avoid considering padded values. The models have been trained on 5 **epochs** and **batch size** of 64, with **Adam** as optimizer and **cross entropy** as loss function.

## 2.2 Sentence Embedding

The models implemented to obtain sentence embeddings are the following:

- RNN layer with last output as sentence embedding;
- RNN layer with mean of all outputs as sentence embedding;
- MLP layer with reshaping;
- Mean of token embeddings (**bag of vectors**)

The following table summarizes each of their structure:

Sentence Embedding Structure		
Model	Layers	Input $\rightarrow$ Output size
RNN with last output	RNN	(122,50) $\rightarrow$ 50
RNN with output mean	RNN	(122,50) $\rightarrow$ (122,50)
	Average Sentence <sup>2</sup>	(122,50) $\rightarrow$ 50
MLP	Reshape	(122,50) $\rightarrow$ 6100
	Dense	6100 $\rightarrow$ 256
	Dense	256 $\rightarrow$ 128
	Dense	128 $\rightarrow$ 50
Bag of vectors	Average Sentence	(122,50) $\rightarrow$ 50

<sup>2</sup><https://nlp.stanford.edu/projects/glove/>

<sup>2</sup>Lambda Layer

## 2.3 Merging

After having embedded the **Claim** and **Evidence** sentences, we must decide how to merge the results. The following strategies have been tested:

- **Concatenation**: concatenation of claim and evidence sentence embeddings (size doubled);
- **Sum**: sum of claim and evidence sentence embeddings;
- **Average**: average of claim and evidence sentence embeddings.

## 3 Results

The results section is probably next easiest to write after the Methods section, since it essentially boils down to presenting your data. If anything, the production of good, high quality figures is the most important and potentially time-consuming part of this. However, make sure to not analyze any of your results here! All of that belongs in the discussion.

Including figures into **L<sup>A</sup>T<sub>E</sub>X** can seem intimidating at first, but Overleaf makes it easy: simply click the 'Project' button above, select 'Files', and upload away from your computer. Then, insert the file name into the appropriate section of the code below. Figure 1 shows the output of such code. A pretty good guide to formatting figures can be found at [https://en.wikibooks.org/wiki/LaTeX/Floats,\\_Figures\\_and\\_Captions#Figures](https://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions#Figures).

```
\begin{figure}
\centering
\includegraphics[width=0.4\textwidth]{test.png}
\caption{Hello!}
\end{figure}
```

## 4 Discussion

And here is the 'meat' of the paper, so to speak. This is where you interpret your results, pointing out interesting trends within your data and how they relate to your initial hypothesis. This is also the place to justify your methodology, if you're so inclined (i.e. Why did you specifically use a certain statistical test over another? Why this tool over that tool?). Lastly, you're going to want to discuss potential sources of error. Make sure to make explicit reference to figures/tables when discussing your data; it can be helpful to walk the reader through your own personal interpretation of each figure in order. Although we recommend looking at past winning papers over at the STEM Fellowship Journal's website

anyways, referring to those papers might prove most helpful when it comes to writing your discussion.

## **Conclusions**

What are the long-term implications of your findings? Wrap up your discussion succinctly while pointing out the significance of your work as well as it what it means for the fields you examined as much as possible. Lastly, suggest ideas for future studies that could build on your work, and justify why they might be useful. Otherwise, you're all done!

## **Acknowledgements**

Anyone to thank/credit for helping your team along the way? This is the place to do it!