Reviews rebuttal
E-commerce products similarity
Project Report for NLP Course, Winter 2023

S. Rećko
WUT
01151399@pw.edu.pl

M. Sperkowski
WUT
01151430@pw.edu.pl

P. Tomaszewski
WUT
01151442@pw.edu.pl

K. Ułasik
WUT
01151444@pw.edu.pl

supervisor: A. Wróblewska
WUT
anna.wroblewska1@pw.edu.pl

1 Team 10 Rebuttal

First of all it is worth mentioning that the reviewing teams had only an early version of the report available, as our team misunderstood the deadline changes and delivered the report a little bit after the first deadline. The teams and supervisor were both informed of that issue and the teams were provided with updating reports as soon as significant changes were made.

- *lack of the report in github as of 8th December*
- As stated before, the report has been added few days later.
- *lack of interpretation/explanation of some of the plots and concepts present in the presentation (for example slide 20 - what is score?)*
- Explanations of main project elements have been expanded in the report.
- *no discussion of the results in the report*
- The early version of report didn't include the discussion as it was still in the works.
- *no reference to the "crucial requirement for the project" that is the time limit in the summary*
- The time measurements analysis for different models have been added to the report.
- *some phrases such as "Works okay, needs refinement and further experiments." should be revised*
- Those were work in progress replacements, all text is revised to scientific standards.
- *The polish dataset is introduced in the 3.1 section of the report but is not used anywhere later and there is even no EDA for it - maybe it will be better to remove it from the report.*
- The early experiments for the Polish dataset, and reasons for excluding it further have been described in the current report.
- *Figure 1 does not have code that can be used to reproduce it or a reference to its source.*
- That item at the time was taken from the WDC website, however it wasn't sufficiently explained. It is corrected in the current report.
- *The code seems reproducible as everything is divided into Classes that can be initialized and used in python notebooks. However, we could not run the specified notebook (demonstration.ipynb) that the authors point to since it depended on a data directory that wasn't part of the provided content. We were able to run the first part of the notebook which was "Demonstration of Raw Embedding Comparison".*
-
- *Another comment I wanted to add was that if it is specified an exact python version to run, then the list of packages to be install should be exhaustive, since you can assume that the person running your program might use a fresh environment (just an opinion)*
- For longevity of the project this sounds like a good idea. However for now the requirements is a sufficient specification. None of the used methods are deprecated and in open-source projects depreciation takes a significant amount of time, so it shouldn't become a problem in a long while.

2 Team 15 Rebuttal

- *A little short description of the Polish dataset. No experiments included with this dataset,*

even the unsuccessful ones (although some justification for this is stated).

- This has been updated in the current version of the report.

- *Limited literature, but the included positions are essential for the project. Unneeded distinction of sections Related works and Related Methodology.*

- The literature has been extensively reworked in the current version of the report.

- *Descriptions of the plots just summarize what they show and lack interpretations of the data and reasoning for their inclusion.*

- This issue has been expanded for some of the plots where we felt this was related.

- *Missing performance discussion, which was mentioned to be critical (only stated that DistilBERT should be faster).*

- Performance analysis has been added in the current version of the report.

- *Missing distinction of similarity score in the results into 3 classes (identical, closely related and entirely different).*

-

- *Our version of unfinished report lacks more extensive experiments description, conclusions and abstract.*

- These elements have been added in the current version of the report.

- *Try to include the Polish dataset, as the results may be quite interesting, even if unsuccessful.* The work done on the Polish dataset and reasons for dropping it have been added to the report.

- *Non-scientific citation of the main dataset - just a link in text.*

- Fixed.

- *Missing TF-IDF citation.*

-

- *There are no ethical concerns about the work.*

-

- *The use of only the gold standard with 1,100 pairs of offers may be not sufficient to train the network.*

- We do agree that this limit might be the reason for unsatisfactory results with fine tuning, however the computability available to us only allowed for this analysis. This issue has been described in the report.

- *The python packages were provided in unusual format (cell in ipynb) without versions (after a while the latest version may break your code – destroy reproducibility - The training process didn't set any random seeds so I highly doubt it would be possible to reproduce the results.*

-

- *Docstring used for functions/methods/classes doesn't resemble standard styles (e.g. docblockr, pep257, google, sphinx, numpy)*

-

- *Use of formatting tool would help you to keep the code up to PEP standards*

-

- *Division of code into separate files is good. However, sometimes it may be a little too much (e.g., trainer.py, training_func.py could have been in one file).*

-

- *From an OOP perspective the LanguageModelBase being a base for Trainer class is unintuitive. The method train() on LLM model that doesn't actually train model should have different name.*

-