
JOHN PAPARRIZOS, University of Chicago, USA
CHUNWEI LIU, MIT, USA

The work on the reproducibility of this project is praiseworthy. All required dependencies and build steps are carefully noted in the provided git repository. A series of scripts allows to automatically rerun the experiments, reproduce the results, and recreate some of the plots in the paper. The reproduced results are similar to the values reported in the paper, and, importantly, all relationships between the compared methods are maintained.

1 INTRODUCTION
This is a reproducibility report for the paper [1]. To summarize, the central results and claims of the paper are supported by the submitted experiments. The key figures have been reproduced accurately enough. The reproducibility scripts are easy to use and well-documented.

2 SUBMISSION
The reproducibility submission consists of detailed instructions on project dependencies and how to rerun the experiments with Python scripts acting as a command-line entry point for the reviewer. Several Python scripts are provided for downloading the datasets, running experiments, and recreating results. Paper figures can be generated automatically by figure scripts with detailed log files.

- Github repository with code and scripts at: https://github.com/dlllb/coles-paper
- Data sources are given as bash scripts

3 HARDWARE AND SOFTWARE ENVIRONMENT
Table 1 describes the resources used in the original paper used and our reproducibility effort.

<table>
<thead>
<tr>
<th></th>
<th>Paper</th>
<th>Repro Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
<td>Tesla P-100 GPU</td>
<td>Tesla V-100 GPU</td>
</tr>
</tbody>
</table>

4 REPRODUCIBILITY EVALUATION
4.1 Process
The experiments are reproduced on the five datasets (Age group prediction competition, Churn prediction competition, Assessment prediction competition, Retail purchase history age group prediction, and Scoring competition) attached in this submission. The scripts run the proposed method and other baselines sequentially on those real-world scenarios. The plot scripts parse the produced results and generate figures shown in the paper. It was possible to follow the reproducibility instructions without the authors’ help.
4.2 Results

The following figures and tables have been reproduced: Figure 1, Figure 2, Figure 3, and Figure 4. The obtained numbers and the visual plots appear to be close enough to the paper’s reported values. The deviation is attributed to the differences in hardware. Most importantly, the relationships between different baselines’ performances match the ones reported and discussed in the paper.

5 SUMMARY

The major figures have been reproduced on the reproducibility platform. The ideas, claims, and findings supported by these figures are therefore reproduced as well.

REFERENCES

Fig. 2. Embedding dimensionality vs. quality on Age group (corresponding to original paper Fig 3a)

Fig. 3. Embedding dimensionality vs. quality on Churn (corresponding to original paper Fig 3b)
Fig. 4. Model quality for different dataset sizes (corresponding to original paper Fig 4)