Using Embeddings for Crowd Based Requirements Engineering

Context

▸ Development teams of popular mobile apps can receive thousands of user reviews daily.

▸ These developers communicate with each other using completely different communication channels, such as issue trackers.

Problem

▸ The nature of the content in issues differs starkly from what most users write about in their reviews.

▸ Issues may lack the steps to reproduce a bug or insights for a new feature request, which might be present in user reviews.

Previous Work

▸ Haering *et al.* proposed a method to automatically match user reviews and issues containing bug reports using text embeddings.

▸ Pilone *et al.* evolved the approach to work with feature requests and multiple languages.

Approach
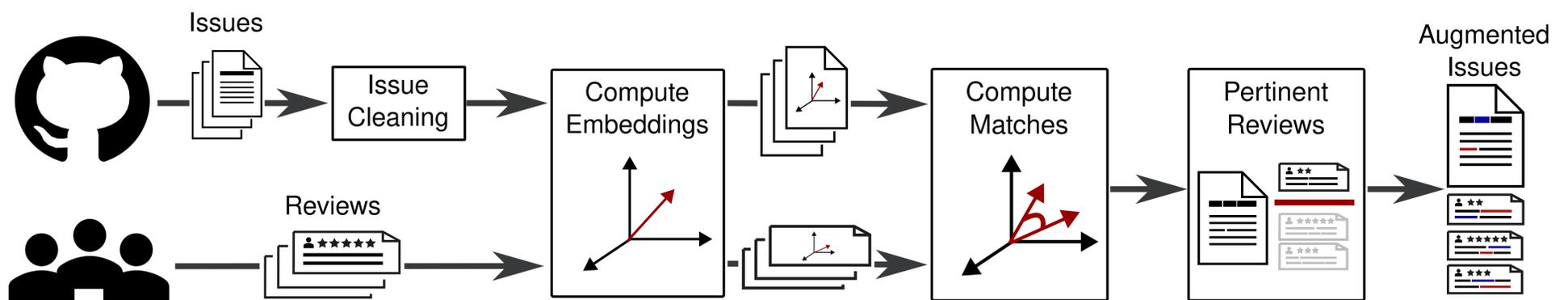
Step-by-Step

1. Collect all of the project issues and reviews.
2. Clean and filter the collected data.
3. Compute the *text embedding* of each review and issue (concatenating the title to the body), using the best-ranked model on the STS task of the MTEB benchmark, *jina-embeddings-v3* [3, 5].
4. Knowing that the distance between two embeddings correlates to the semantic difference between the two fragments, find the 5 most similar reviews to each issue or vice-versa [1, 2].

Results

Matches found by Embedding Method
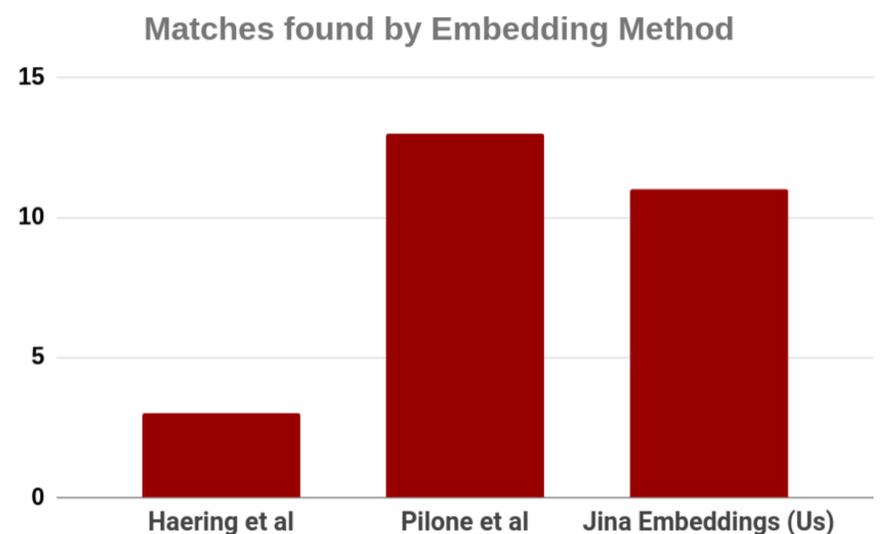
Preliminary Evaluation

▸ We reproduced the evaluation of Pilone *et al.* [4] using the BikeSP project, with over 600 issues and 69 user reviews.

▸ For the evaluation, we tested if the approach correctly identified the issue related to 29 user reviews.

▸ Besides collecting new issues from the project, we used a new multi-lingual model (*jina-embeddings-v3* [5]) that suppressed the need for translating the issues and reviews.

Discussion

▸ The results using *jina-embeddings-v3* [5] were slightly less favorable than that of Pilone *et al.* [4].

▸ The inclusion of new unrelated issues might have negatively impacted the results.

▸ Additionally, the multi-lingual model might be less accurate with texts in Portuguese than the previous model was with artifacts translated to English.

References

▸ Haering, Marlo et al. (2021). "Automatically Matching Bug Reports With Related App Reviews". In: *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pp. 970–981. DOI: 10.1109/ICSE43902.2021.00092.

▸ Jiao, Qilu and Shunyao Zhang (2021). "A Brief Survey of Word Embedding and Its Recent Development". In: *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Vol. 5, pp. 1697–1701. DOI: 10.1109/IAEAC50856.2021.9390956.

▸ Muennighoff, Niklas et al. (2023). *MTEB: Massive Text Embedding Benchmark*. arXiv:2210.07316. arXiv: 2210.07316 [cs.CL]. URL: https://arxiv.org/abs/2210.07316.

▸ Pilone, Arthur et al. (2024). "Multilingual Crowd-Based Requirements Engineering Using Large Language Models". In: *Proceedings of the Brazilian Symposium on Software Engineering (SBES)*. SBC, pp. 679–685. DOI: 10.5753/sbes.2024.3646.

▸ Sturua, Saba et al. (2024). *jina-embeddings-v3: Multilingual Embeddings With Task LoRA*. arXiv:2409.10173. arXiv: 2409.10173 [cs.CL]. URL: https://arxiv.org/abs/2409.10173.

Arthur Pilone Maia da Silva, Edllyn de Jesus Silva

arthurpilone@usp.br, edllyn@usp.br

Department of Computer Science — University of São Paulo

IME