Categorizing stackoverflow Questions Using A Tag Hierarchy

Philip Roozendaal (p.m.roozendaal@student.tudelft.nl)
Supervisors: Dr. Maliheh Izadi and Prof. Dr. Arie van Deursen

TU Delft

Background

- "Stack Overflow is a question and answer site for professional and enthusiast programmers"
- Stack Overflow currently contains over 22 million questions
- Thousands of new questions asked daily
- Tags used as an important tool to help organize and navigate

Research question

How accurate and useful is a tag predictor based on a tag hierarchy?

- **RQ1**: How accurate is the tag predictor for SO questions?
- **RQ2**: How should the tag hierarchy be organized?
- **RQ3**: How useful is the tag predictor for SO questions?

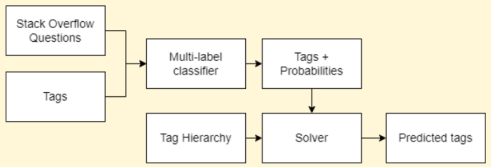

Figure 1

Building the tag-hierarchy

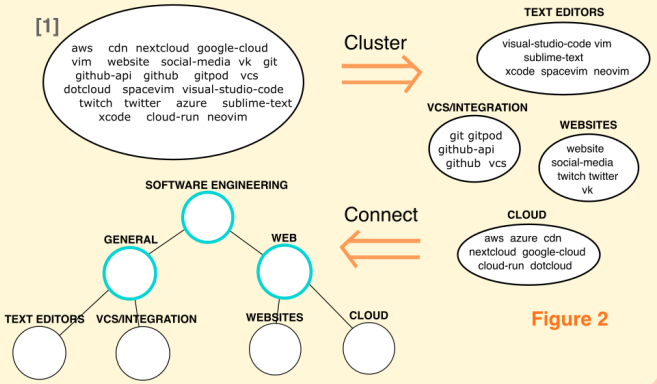

Figure 2

Applying the hierarchy

- 1. Predict probability for each possible tag using the baseline classifier.
- 2. Build a clique from the results, where each node represents a tag with probability, and each edge the closeness between two nodes in the hierarchy.
- 3. Find the N nodes that with the sum of the tag probabilities and edges obtain the highest possible score within the clique.

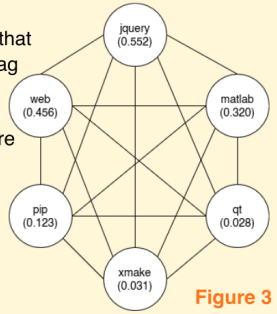

Figure 3

Setting up the Baseline

- DistilBERT used as the Baseline model [2]
  - State-of-the-art transformer
  - Fast learning, accurate results
- Questions split for 80% training and 20% testing
- Results
  - LRAP of 52%
  - F1 of 63%

Data Collection

- Stack Exchange Data Explorer (question data, tags)
- Stack Exchange REST API (related tags)
- Question data filtered by:
  - Posted after 2016 (focus on recent trends)
  - Contains at least one tag in hierarchy
- The first 20'000 questions are collected

Evaluation

F1 and LRAP scores both improved by 2% over the baseline!

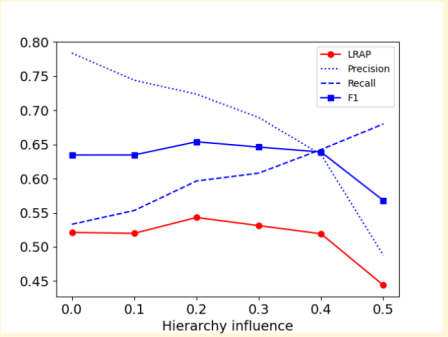

Figure 4

Survey Results

Rank Improved	Rank Deteriorated	A useful tag was added	A useful tag was removed
7	1	6	1

Figure 5

Conclusion

How accurate and useful is a tag predictor based on a tag hierarchy?

- **RQ1:** How accurate is the tag predictor for SO questions?
  It is at least 2% more accurate!
- **RQ2:** How should the tag hierarchy be organized?
  As a balanced tree with clusters that are not too small or too big
- **RQ3:** How useful is the tag predictor for SO questions?
  The survey has indicated that it adds a net benefit to usefulness over the baseline

References

- [1] Maliheh Izadi, Abbas Heydarnoori, Georgios Gousios. Topic recommendation for software repositories using multi-label classification algorithms, 2021.
- [2] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108, 2019
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018