Extending Community QA Moderation Using Multi-View Learning and T5 Fine-Tuning

Vraj Shah
Vs921

Ayush Dodia
and179

Nithik Pandya
nhp46

Abstract

Community Question-Answering (CQA) platforms such as Stack Overflow face significant challenges in maintaining high content quality due to the overwhelming number of user-generated questions. Existing moderation approaches primarily rely on crowdsourced user feedback, which is time-consuming, subjective, and inconsistent. To address this, the study by Annamoradnejad et al. (2022) proposes a multi-view learning-based classifier that categorizes questions into High-Quality (HQ), Low-Quality Editable (LQ_EDIT), and Low-Quality Closeable (LQ_CLOSE). While their approach achieves 95.6% accuracy, it lacks explainability in domain-specific classification and does not effectively handle low-quality edits. This project builds on their work by improving model explainability and enhancing low-quality questions using a fine-tuned T5 model. Our approach introduces an iterative refinement loop that systematically improves LQ_EDIT questions until they meet high-quality standards. Additionally, we analyze and refine the original moderation pipeline to improve its robustness and adaptability. The expected outcomes include an enhanced multi-view model with improved explainability, a fine-tuned T5-based question rephrasing system, and a fully automated iterative moderation pipeline.

1 Introduction

Community Q&A websites are dense knowledge-sharing communities in which users post and answer technology questions. Keeping the quality and relevance of questions intact is critical to maintaining platform integrity, yet manual moderation won't work since too much material is uploaded every day. Stack Overflow and other websites are based on user-moderation mechanisms, whereby questions are labeled as inappropriate, modified, or closed through feedback and voting by the community. However, the process takes time and is subjective, leading to irregular moderation activity.

To address these problems, Annamoradnejad et al. (2022) proposed a multi-view classification model that predicts the quality of the question as high-quality, requiring edits or closure. Their approach does not rely on user interaction but instead utilizes textual, contextual, and statistical features extracted from the question. While their model exhibits high accuracy, it is opaque in the explanations of its classifications in various technical domains (e.g., AI, Data Structures, Java). Additionally, low-quality edits are not regularly refined, and thus, automated moderation is less effective. This project closes these gaps by implementing explainability mechanisms, question rephrasing based on T5, and an iterative refinement pipeline to increase moderated questions' overall quality and usability.

2 Dataset and Data Sources

Our project is currently working on the Stack Overflow dataset introduced in Annamoradnejad et al. (2022), consisting of 60,000 labeled questions on a range of programming topics. The question has title, body, tags, and a classification label (HQ, LQ_EDIT, or LQ_CLOSE). We will be working with real-world Stack Overflow questions with programming-related tags labeled (e.g., "Python," "Machine Learning," "Data Structures"). These tags are used to analyze patterns of classification among different technical domains.

While our current experiments make use of only Stack Overflow data, future implementation of this project will be supplemented with additional datasets from other well-known CQA platforms, such as Quora, Yahoo! Answers, and GitHub Discussions. Merging these diverse sources will help improve the generalizability of the model and the domain diversity of different Q&A environments.

3 Methodology and Implementation Plan

Our implementation is structured into four major steps.

3.1 Research and Updating the Moderation Pipeline

We will analyze the existing moderation pipeline introduced in Annamoradnejad et al. (2022) to identify bottlenecks, inefficiencies, and areas for optimization. Based on our findings, we will propose and implement improvements to further enhance the moderation pipeline.

3.2 Improving Explainability

We will do a feature importance analysis on the multi-view classifier to ascertain which characteristics—such as readability, length, and clarity—contribute most to predictions in order to improve model transparency. By categorizing error rates and accuracy variations for questions tagged with AI, Data Structures, Java, and other programming fields, we will also be able to see trends in question categorization across different technical domains. A better grasp of the model's behavior across various query kinds will be possible thanks to these revelations.

3.3 Handling Low-Quality Edits Using T5

Low-quality questions that can be fixed through edits (LQ_EDIT) require systematic improvement before being published. To address this, we introduce a T5-based text transformation model that rephrases low-quality questions into clearer, more structured formats. Instead of directly using the T5 model in moderation, we will first generate a new dataset where all LQ_EDIT questions are rephrased using the T5 model. This newly generated dataset will then be used to fine-tune the T5 model, ensuring that it learns from real-world low-quality questions and their improved versions.

3.4 Iterative Question Refinement Pipeline

To further enhance low-quality questions, we implement an iterative refinement loop that repeatedly improves a question until it reaches high-quality status. The process follows these steps:

1. Classify a new user-submitted question into HQ, LQ_EDIT, or LQ_CLOSE.
2. If classified as HQ, the question is approved without modification.
3. If classified as LQ_CLOSE, the question is directly removed as it is either spam, off-topic, or incomprehensible.
4. If classified as LQ_EDIT, it is passed through the fine-tuned T5-based rephrasing model to generate a higher-quality version.
5. Re-classify the transformed question using the multi-view classifier.
6. Repeat the process until the question is categorized as HQ.
7. Maintain a loop counter to track the number of iterations needed for improvement.

This automated refinement mechanism ensures that low-quality questions are progressively enhanced before they are finalized for public view.

3.5 Evaluation

We will be evaluating the performance of our multi-view classification model using standard performance metrics like accuracy, precision, recall, and F1-score. We will also be checking the confusion matrix and ROC curves to get more information about how the classifier performs on different technical domains. At the same time, the performance of the fine-tuned T5 model will be measured in terms of metrics such as BERT Score to quantify the quality of the rephrased questions produced and observe the number of iterations of the loop required in the iterative refinement pipeline to generate high-quality output.

4 Expected Outcomes

The expected outcomes of this project include an improved multi-view classifier with enhanced explainability in domain-specific question categorization, a fine-tuned T5 model capable of transforming low-quality questions into well-structured, high-quality posts, and a fully automated moderation pipeline that iteratively refines low-quality edits; performance evaluation will compare the original model against our modifications in terms of accuracy, explainability, and question clarity.

5 References

Annamoradnejad, I., Habibi, J., & Fazli, M. (2022). Multi-view approach to suggest moderation actions in community question answering sites. *Information Sciences, 600*, 144-154. https://doi.org/10.1016/j.ins.2022.03.085