Study on Automatic Defect Report Classification System with Self Attention Visualization

Rin Hirakawa\textsuperscript{a}, Keitaro Tominaga\textsuperscript{b}, Yoshihisa Nakatoh\textsuperscript{a}
\textsuperscript{a}Kyushu Institute of Technology, Kitakyushu, Fukuoka Japan
\textsuperscript{b}Panasonic System Design Co., Ltd., Yokohama, Kanagawa Japan
nakatoh@ecs.kyutech.ac.jp

Abstract— In recent years, software in devices such as smartphones and tablets has become increasingly multifunctional, and the use of OSS has become essential. In software development using large-scale OSS, it is important to report defects to appropriate personnel promptly. In this paper, we propose a method to classifying defect reports into appropriate categories using fine-tuned BERT and visualize self-attention information. In the evaluation, category classification was performed using defect reports of the actual OSS project. The F1 score was 0.87, which indicated that high-accuracy classification was possible. Also, the visualization results show that category-specific words can be extracted.

I. INTRODUCTION

In recent years, software functions of devices such as smartphone, tablet, and car display audio are becoming increasingly multifunctional in response to user requests. Because smart devices require software to be released in a short period of time, OSS (Open Source Software) is commonly used for a variety of functions. Benefits of using OSS include faster development times, new features through updates and open internal implementations. However, in software that uses large-scale OSS, when a bug occurs, it takes a long time to identify the OSS that is causing the bug from the symptoms, which causes a delay in the development cycle.

Systems development using OSS typically involves a tester uploading "Tickets" describing the defect to the defect management system and manually assigning it to the appropriate personnel. Although quick and accurate classification is required in order to solve defects quickly, the number of defects increases as the development scale of OSS increases, and the manual and time costs required for classification become a burden on the progress of the project.

In this research, we propose a system that automates the categorization of defects using fine-tuned BERT as a tool to shorten the development cycle of smart device software using OSS. The bug report description is divided into subwords and then entered into BERT so that it can be categorized into the OSS category, taking into account the rare word information specific to the project. The proposed method also considers visualizing the self-attention information of the model during prediction by a directed graph. Visualizing the words referenced by the model in automatic classification increases the chances of noticing misclassification and helps quickly understanding the contents of a defect report.

II. PROPOSED METHOD

A. Overview

Our proposed system is for defect tickets uploaded on a defect management system such as Redmine. We use only the bug description in category classification. The bug description is divided into tokens using the WordPiece tokenizer, and then classified into categories by the BERT model fine-tuned with the defect reports. Categorized defect tickets are assigned to each person in charge with category information and Self-attention visualization results (Fig. 1).

B. WordPiece Tokenizer

WordPiece\textsuperscript{[1]} is the technique that divides unknown words into subwords so that they can be interpreted as combinations of known vocabularies. The bug report contains many domain-specific vocabularies, but WordPiece can be used to extrapolate meaning from general vocabularies.

C. BERT

BERT\textsuperscript{[2]} is attracting attention as a general-purpose neural language model that can be adapted to various tasks. BERT (base model) consists of 12 layers, each containing 12 self-attention heads. Each head is known to pay attention to different words, and we attempt to visualize these self-attention.

![Fig. 1. Overview of proposed system.](image-url)
III. EXPERIMENT

In order to verify the effectiveness of the proposed method, we conduct category classification experiments using actual OSS project defect report data[3]. The data covers 44049 Apache defect reports collected from 1999 to 2017.

A. Preprocessing

First, the data containing the duplication and NA value is removed, and the defect description shorter than 20 characters is deleted. Then, the defect reports of the top five classes in the number of data are extracted, and data columns other than product category labels and descriptions are deleted.

B. fine-tuning BERT

Divide the data resulting from the pre-processing at the rate of 6:2:2 for training, validation, and testing. Using the training and validation data, fine-tune the BERT for the category classification under the conditions shown in Table 1.

C. Category Classifications and Visualization

The test data is classified by the fine-tuned BERT to obtain the confusion matrix and F1 score. Then Self-attention is visualized using randomly sampled data from the test data. A directed graph is created based on the obtained self-attention, and the page rank algorithm is applied under the condition of alpha = 0.9. Visualization displays a directed graph with edges and nodes weighted by the strength of attention and the value of Page Rank.

IV. RESULT

The confusion matrix obtained from the category classification experiment is shown in Fig. 2. As a result, the F1 score (weighted average) was 0.87, indicating that the overall defect report could be classified with high accuracy, but the confusion matrix suggested that misclassification was caused by an imbalance in the number of class data. An example of the visualization data is shown in Fig. 3. In this example, the product-specific vocabulary is highlighted, indicating that the surrounding vocabulary is paying strong attention.

V. CONCLUSION

In this study, we carried out classification experiments of defect reports and visualization of self-attention using fine-tuned BERT, and confirmed that classification with high accuracy is possible. As future tasks, we plan to develop a method to improve misclassification based on visualized self-attention and extract only information useful for developers.

ACKNOWLEDGMENT

We thank Keiichi Tokuyama and his section members for their helpful feedback on the paper. This work is supported by a grant from Panasonic System Design.

REFERENCE