先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	ソフトウェア工学 (松本　健一　(教授))		
学籍番号	2411023	提出日	令和 7年 7月 22日
学生氏名	稲葉　真幸		
論文題目	Understanding the Characteristics of LLMs in Detecting Textually Dissimilar Duplicate Bug Reports		

要旨

In large-scale software development, hundreds of bugs are reported to bug management systems every day. This can lead to multiple people reporting the same bug, resulting in duplicate bug reports that are costly to identify. Among them, textually dissimilar duplicates are particularly difficult to detect. Because bug reports are written in natural language, the same issue can be described in many different ways. Conventional natural language processing and machine-learning methods struggle to identify such duplicates unless they fully capture the semantic meaning of each report. This thesis examines the characteristics of automatic detection of textually dissimilar duplicate bug reports using Transformer-based large language models, especially Sentence Transformer-based models that specialize in textual similarity. Quantitative evaluation shows that Sentence Transformer-based models achieve higher detection performance for both textually similar and dissimilar duplicate bug reports. In addition, a qualitative error analysis of misclassified cases and a case study of the best-performing model identify overarching factors that govern the success and failure of LLM-based methods, offering general insights for improving duplicate detection systems.