A Thesis for the Degree of Ph.D. in Engineering

Cost-Efficient Text Classification without Manually Labeled Data

September 2025

Graduate School of Science and Technology,
Keio University

Yejian Zhang

別表６
（３）

Keio University

Thesis Abstract

No.

Registration Number	☑ "KOU" □ "OTSU" No. *Office use only	Name	ZHANG, Yejian

Thesis Title

Cost-Efficient Text Classification without Manually Labeled Data

Thesis Summary

Text classification helps manage large volumes of unstructured data. Its applications are manifold: analyzing software user feedback to guide developers in maintenance and evolution; performing sentiment analysis on customer reviews to help enterprises enhance products and services; detecting toxic comments to filter harmful content; and conducting topic categorization to organize and retrieve documents efficiently. However, most machine-learning-based approaches rely heavily on large amounts of manually labeled data for training. Moreover, when viewed through the lens of software engineering for machine learning, practical deployment faces additional hurdles, including the need for domain expertise and substantial computational resources. This dissertation investigates how to design cost-efficient and accessible ML-based text classification systems that require little or no manual labeling, while also aiming to lower the barriers related to computational resource consumption and the need for expertise.

Two complementary, cost-efficient text classification systems are presented in this dissertation. The first system trains a domain-adapted language model through self-supervised learning and combines unsupervised clustering with active learning to identify the most informative instances for labeling. This system enables accurate classification of software engineering-oriented game user reviews while requiring only a small number of labeled instances. The second system embeds a large language model as an oracle within an active learning loop, delivering cost-efficient cross-task text classification—including sentiment analysis, news topic categorization, and toxic comment detection—without any manually labeled data.

The proposed systems were evaluated through comprehensive experiments. The first achieved 88.8% accuracy on Steam game reviews with minimal supervision and demonstrated the reuse potential. The second system consistently delivered high accuracy across all three text classification tasks without manual labels. Moreover, compared to directly using the GPT model for text classification, the second system achieved over 93% of its classification performance while requiring only approximately 6% of the associated time and financial cost.

Overall, this dissertation proposes cost-efficient and accessible solutions for text classification without relying on manually labeled data. By integrating large language models, self-supervised learning, active learning, and unsupervised learning, the proposed systems demonstrate the feasibility of building high-performance classification systems under limited supervision and resource-constrained conditions. The experimental results validate not only their technical effectiveness and generalization capabilities but also their practical value in scenarios where manual annotation is costly or infeasible. These findings contribute toward advancing more sustainable and adaptable machine-learning-based text classification technologies across diverse application scenarios.