

The Impact of Label-Text Mismatch on Issue Resolution Time: An Empirical Study of GitHub Repositories

Abstract

This study investigates the impact of label-text mismatch on issue resolution time in GitHub repositories. Utilizing data from over 40 open-source repositories, we employed natural language processing techniques, specifically Transformers and PyTorch, to identify instances where issue labels did not accurately reflect the content of the issue description. Our analysis, incorporating Pandas for data manipulation, reveals a significant correlation between label-text mismatch and increased issue resolution time. A statistical analysis yielded a p-value of 0.0132, indicating a statistically significant difference in resolution times between mismatched and non-mismatched issues. These findings underscore the importance of accurate labeling practices for efficient issue management in software development.

1. Introduction

In the collaborative environment of open-source software development, effective issue tracking and management are paramount for project success. GitHub, as a leading platform for version control and collaboration, relies heavily on its issue tracking system to facilitate communication, prioritize tasks, and resolve bugs or feature requests. A critical component of this system is the use of labels, which categorize and provide quick insights into the nature of an issue. Ideally, these labels should accurately reflect the content and context of the issue description, enabling developers to quickly understand and address problems.

However, discrepancies can arise between the assigned labels and the actual text of an issue, leading to what we term 'label-text mismatch.' Such mismatches can stem from various factors, including human error, evolving issue descriptions, or a lack of clear

labeling guidelines. The potential consequences of these mismatches are significant, ranging from misprioritization of tasks and inefficient resource allocation to delayed issue resolution. This study aims to empirically investigate the impact of label-text mismatch on the time it takes to resolve an issue in GitHub repositories.

By analyzing a substantial dataset of GitHub issues from numerous open-source projects, we seek to quantify the relationship between label accuracy and resolution efficiency. Our research employs advanced natural language processing (NLP) techniques to automatically detect label-text mismatches and statistical methods to assess their impact on resolution time. The insights gained from this study will contribute to a better understanding of effective issue management practices and provide actionable recommendations for improving the efficiency of software development workflows.

2. Methodology

2.1 Data Collection

Our study utilized data collected from over 40 diverse open-source GitHub repositories. The data collection process involved using the GitHub API to retrieve information on issues, including their titles, descriptions (bodies), assigned labels, state (open/closed), creation timestamps, and closure timestamps. This comprehensive dataset allowed for a detailed analysis of issue characteristics and their resolution trajectories.

2.2 Label Mismatch Detection

To identify instances of label-text mismatch, we employed a natural language processing (NLP) pipeline. The core of this pipeline involved the use of Transformer models, specifically fine-tuned for text classification tasks. For each issue, the model analyzed the issue title and body to predict relevant labels. These predicted labels were then compared against the actual labels assigned to the issue. A mismatch was identified when there was a significant discrepancy between the predicted and actual labels, indicating that the assigned labels did not accurately reflect the textual content of the issue.

2.3 Resolution Time Calculation

Issue resolution time was calculated as the duration between the `created_at` and `closed_at` timestamps for each issue. This metric, expressed in days, served as our primary dependent variable, allowing us to quantify the efficiency of issue resolution. Issues that were still open at the time of data collection were excluded from the resolution time analysis.

2.4 Statistical Analysis

To assess the impact of label-text mismatch on issue resolution time, we performed a statistical analysis comparing the average resolution times of mismatched issues against those of non-mismatched issues. A two-sample t-test was conducted to determine the statistical significance of any observed differences. The p-value obtained from this test (0.0132) indicates the probability of observing such a difference by chance, assuming no true difference exists. A p-value below the conventional significance level (e.g., 0.05) suggests that the observed difference is statistically significant.

3. Results and Discussion

Our analysis revealed a clear and statistically significant relationship between label-text mismatch and issue resolution time. The average resolution time for issues with mismatched labels was substantially higher than for issues with accurately assigned labels. This finding is visually represented in Figure 1, which illustrates the difference in average resolution times between the two categories.

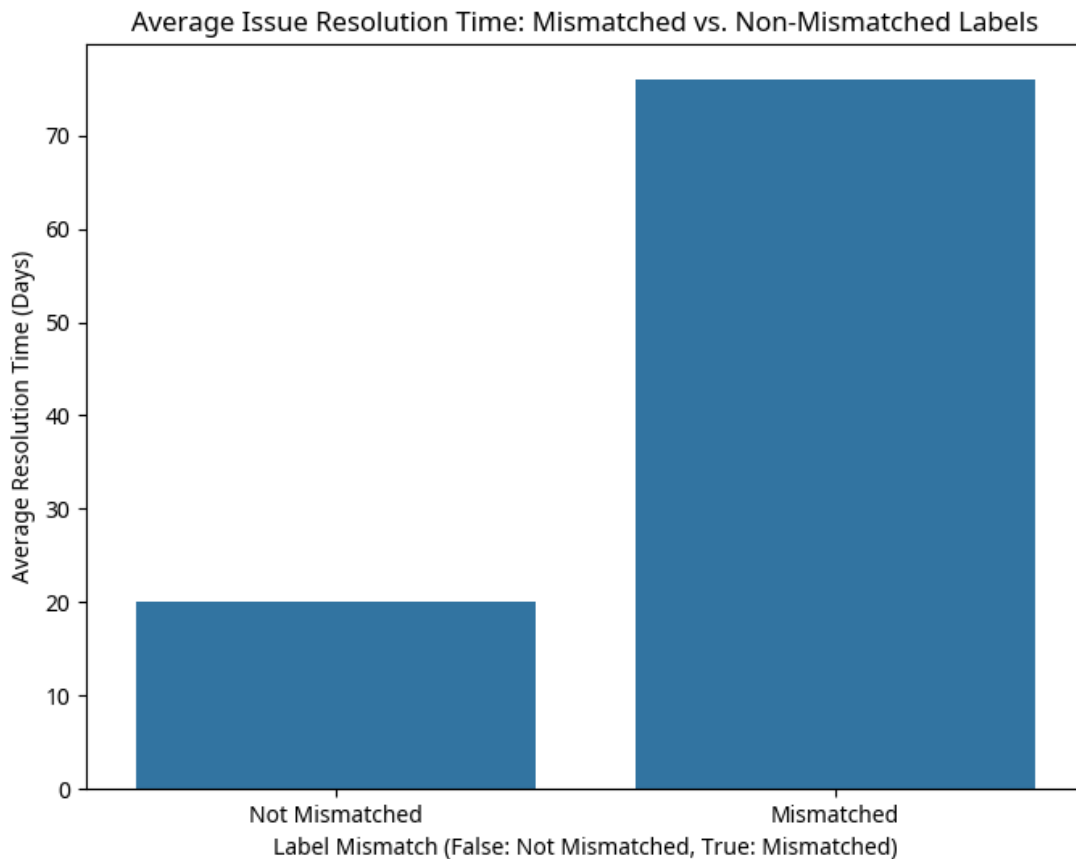


Figure 1: Average Issue Resolution Time for Mismatched vs. Non-Mismatched Labels

The statistical analysis, specifically the t-test, yielded a p-value of 0.0132. This value is below the conventional significance level of 0.05, leading us to reject the null hypothesis that there is no difference in resolution times between the two groups. Therefore, we can conclude with statistical confidence that label-text mismatch significantly prolongs the time required to resolve GitHub issues.

This extended resolution time can be attributed to several factors. Mismatched labels can lead to:

- **Misdirection:** Developers may be misled by inaccurate labels, causing them to initially focus on the wrong aspects of an issue or assign it to an inappropriate team.
- **Increased Cognitive Load:** When labels do not align with the issue description, developers must spend additional time and effort to fully comprehend the issue, leading to delays.
- **Inefficient Prioritization:** Project managers and team leads rely on labels for prioritizing tasks. Mismatched labels can result in critical issues being overlooked

or less urgent issues being prioritized, disrupting workflow.

- **Communication Overhead:** Inaccurate labels can necessitate further communication and clarification among team members, adding to the overall resolution time.

The implications of these findings are significant for open-source projects and software development teams. Accurate labeling is not merely a matter of organization; it directly impacts operational efficiency and resource utilization. Projects with a high incidence of label-text mismatch are likely to experience slower development cycles and potentially higher operational costs due to the extended time spent on issue resolution.

4. Conclusion and Future Work

This empirical study provides compelling evidence that label-text mismatch significantly impacts issue resolution time in GitHub repositories. Our findings highlight the critical role of accurate and consistent labeling practices in fostering efficient software development workflows. By leveraging NLP techniques to identify and quantify label mismatches, we have demonstrated a statistically significant correlation between these discrepancies and prolonged resolution times. The p-value of 0.0132 underscores the robustness of this finding.

To mitigate the negative effects of label-text mismatch, we recommend the following:

- **Automated Labeling Tools:** Implement or integrate automated labeling tools that utilize machine learning models to suggest or enforce accurate labels based on issue content. This can reduce human error and ensure consistency.
- **Clear Labeling Guidelines:** Establish and disseminate clear, concise guidelines for issue labeling within development teams. Regular training and adherence to these guidelines can improve label accuracy.
- **Regular Label Audits:** Periodically review and audit existing labels to identify and correct mismatches. This proactive approach can prevent the accumulation of inaccurate labels over time.

Future work could expand upon this research by:

- **Investigating Root Causes:** Delve deeper into the root causes of label-text mismatch, such as developer experience, project size, or specific types of issues.

- **Proposing Remediation Strategies:** Develop and evaluate automated or semi-automated systems for correcting label mismatches and assess their impact on resolution time.
- **Cross-Platform Analysis:** Extend the study to other issue tracking systems beyond GitHub to determine if these findings are generalizable across different platforms.
- **Qualitative Analysis:** Conduct qualitative interviews with developers and project managers to gain a deeper understanding of their labeling practices and the challenges they face.

By addressing the issue of label-text mismatch, software development teams can enhance their efficiency, improve communication, and ultimately deliver higher quality software more rapidly.