**引言**

在现代软件开发过程中，日志解析扮演着至关重要的角色，其核心任务是从大量非结构化的日志数据中精准地提取出结构化的信息，以便于后续的分析和处理。然而，传统的日志解析器在面对日志格式的多样性和复杂性时，往往无法高效地完成解析任务，这不仅增加了开发和维护的成本，还严重影响了下游日志分析任务的性能和准确性。鉴于此，开发出更加高效、准确且能够适应多种日志格式的日志解析方法，对于提升软件系统的整体质量和运维效率具有极为重要的意义，这不仅能帮助开发团队更好地理解和优化系统运行状态，还能为软件的持续改进和创新提供有力支持。

**问题分析**

**需求收集**

通过团队内的交流以及分析公开的日志数据集和相关文献，了解到相关人员对日志解析准确性和效率的需求。同时，从系统层面考虑，需要解析器能够快速适应不同系统产生的日志格式变化。

**需求分析**

关键需求包括高解析精度、快速适应新日志格式、低资源消耗等。约束条件包括有限的训练数据、模型推理速度等。

**现有方案评估**

传统日志解析方法如基于正则表达式、频繁模式挖掘等，虽能处理部分日志，但在面对复杂多样的日志格式时，准确性和适应性不足。基于机器学习的方法虽有一定改进，但对大量标注数据依赖性强，且难以处理未见过的日志格式。

**新解决方案**

在日志解析领域，利用大型语言模型（LLM）的 In-Context Learning（ICL，上下文学习）能力，可以通过在提示（prompt）中巧妙地嵌入少量精心挑选的标注样本及其对应的标签，有效地引导 LLM 深入学习日志模板的语义特征。这种方法不仅能够让 LLM 更加精准地理解日志模板的结构和语义，还能显著降低模型训练过程中的复杂性和资源消耗。通过这种方式，LLM 可以在无需大量标注数据的情况下，快速适应新的日志格式和任务，从而在日志解析任务中表现出更高的效率和准确性。这种方法的实施，为日志解析领域带来了新的突破，为处理复杂多样的日志数据提供了更加高效和灵活的解决方案。

**结论和展望**

通过引入**ICL**，**LLM**可以更好地利用预训练模型的语义理解能力，通过优化样本选择和提示构建，提高日志解析的准确性和效率。这些改进措施不仅在技术层面提升了 LLM 在日志解析任务中的性能表现，使其能够更快速、更准确地处理各种复杂的日志数据，而且在实际应用中具有重要意义。它们有效减少了对大规模标注数据的依赖，降低了数据准备和模型训练的成本和时间，使得 LLM 更加适合应用于实际场景中的日志解析任务，能够更好地满足软件开发和运维过程中对高效、准确日志解析的需求。

1.Introduction

In modern software development, log parsing plays a crucial role. Its core task is to accurately extract structured information from a large volume of unstructured log data for subsequent analysis and processing. However, traditional log parsers often fail to efficiently complete the parsing task when faced with the diversity and complexity of log formats. This not only increases the cost of development and maintenance but also severely impacts the performance and accuracy of downstream log analysis tasks. Therefore, developing more efficient, accurate, and adaptable log parsing methods for various log formats is of great significance for improving the overall quality and operational efficiency of software systems. This can not only help development teams better understand and optimize system performance but also provide strong support for the continuous improvement and innovation of software.

### Problem Analysis

#### Requirement Collection

Through internal team discussions and analysis of public log datasets and relevant literature, we identified the need for log parsing accuracy and efficiency. From a system perspective, the parser needs to quickly adapt to changes in log formats generated by different systems.

#### Requirement Analysis

Key requirements include high parsing accuracy, rapid adaptation to new log formats, and low resource consumption. Constraints include limited training data and model inference speed.

#### Existing Solution Assessment

Traditional log parsing methods, such as regular expressions and frequent pattern mining, can handle some logs but lack accuracy and adaptability when dealing with complex and diverse log formats. Machine learning-based methods have shown some improvements but rely heavily on large amounts of labeled data and struggle with unseen log formats.

### New Solution

In the field of log parsing, leveraging the In-Context Learning (ICL) capability of Large Language Models (LLMs) allows for the effective guidance of LLMs to deeply learn the semantic features of log templates by embedding a small number of carefully selected labeled samples and their corresponding labels into prompts. This method not only enables LLMs to more accurately understand the structure and semantics of log templates but also significantly reduces the complexity and resource consumption of model training. In this way, LLMs can quickly adapt to new log formats and tasks without the need for large amounts of labeled data, thus demonstrating higher efficiency and accuracy in log parsing tasks. The implementation of this method brings new breakthroughs to the field of log parsing, providing a more efficient and flexible solution for handling complex and diverse log data.

### Conclusion and Outlook

By introducing ICL, LLMs can better utilize the semantic understanding capabilities of pre-trained models. Optimizing sample selection and prompt construction enhances the accuracy and efficiency of log parsing. These improvements not only boost the performance of LLMs in log parsing tasks, enabling them to handle complex log data more quickly and accurately, but also hold significant practical importance. They effectively reduce the dependence on large-scale labeled data, lowering the costs and time associated with data preparation and model training. This makes LLMs more suitable for log parsing tasks in real-world scenarios, better meeting the needs for efficient and accurate log parsing in software development and operations.