Project 3: Title Generation with T5 and Mistral

Mohammad Abu Tareq Rony
abutareq@iastate.edu

Abstract

In this project, we investigate the automatic generation of titles for animal genetics papers of Animal QTLdb. We use the `QTL_text.json` data set to train and evaluate two different approaches. First, we fine-tune T5-small encoder decoder model that learns to generate titles directly from abstracts. Second, we apply an instruction-tuned Mistral 7B model in a 10-shot in-context learning setting without any parameter updates. On the held-out test set, the T5-small model achieves BLEU 11.36, ROUGE-2 0.2144, and ROUGE-L 0.3619. The Mistral 7B model reaches BLEU 8.40, ROUGE-2 0.1940, and ROUGE-L 0.3458, showing strong semantic alignment even without fine-tuning. Based on these results, we use the fine-tuned T5-small model to generate titles for the abstracts in `QTL_test_unlabeled.tsv`, providing a practical workflow for automated title generation in QTLdb-related curation tasks. The **Project Code** is available at: `NLP Project 3`

1 Introduction

Scientific abstracts often contain detailed descriptions of research objectives, methods, and findings. Converting this information into a short and meaningful title can be challenging, especially when large collections of papers must be curated. Automated title generation offers a way to reduce manual effort, and recent advances in language models make this task increasingly feasible.

In this project, we explore two different approaches for generating scientific titles from abstracts. The first approach fine-tunes a T5 model on the QTL dataset, allowing it to learn the typical structure and style of real paper titles. The second approach uses the Mistral 7B model in a 10-shot setting, where titles are generated through a simple prompt without any model training.

Our goal is to compare how these two strategies perform on the same dataset. T5 represents a supervised learning method, while Mistral reflects a prompt-based method that relies only on in-context examples.

2 Methodology

2.1 Dataset Description and Preprocessing

The dataset in `QTL_text.json` contains animal genetics papers, and we use only two fields for title generation:

- **Abstract**: A textual summary describing the purpose, methods, and findings of the study.
- **Title**: The reference title, used as the target output for training and evaluation.

After cleaning, the data is split into training and development sets for fair evaluation.

2.2 Models Description

We use a fine-tuned T5-small model (Raffel et al., 2020) and a 10-shot prompted Mistral 7B model for title generation. A summary of the key features of each model is given in Table 1. T5 maps the abstract to a title (Eq. 1), while Mistral uses a prompt-based input (Eq. 2).

$$f_{\text{T5}}(x) \rightarrow y_{\text{title}} \tag{1}$$

$$f_{\text{Mistral}}(\text{prompt} + \text{abstract}) \rightarrow y_{\text{title}} \tag{2}$$

2.3 Evaluation Metrics

We evaluate the generated titles using three standard text generation metrics: BLEU, ROUGE-2, and ROUGE-L. A simplified BLEU formula is;

$$\text{BLEU} = \text{precision of matching } n\text{-grams}. \tag{3}$$

ROUGE-2 is computed as;

$$\text{ROUGE-2} = \frac{\text{bigrams matched}}{\text{bigrams in reference}}. \tag{4}$$

ROUGE-L is defined as;

$$\text{ROUGE-L} = \frac{\text{length of LCS}}{\text{length of reference}}. \tag{5}$$

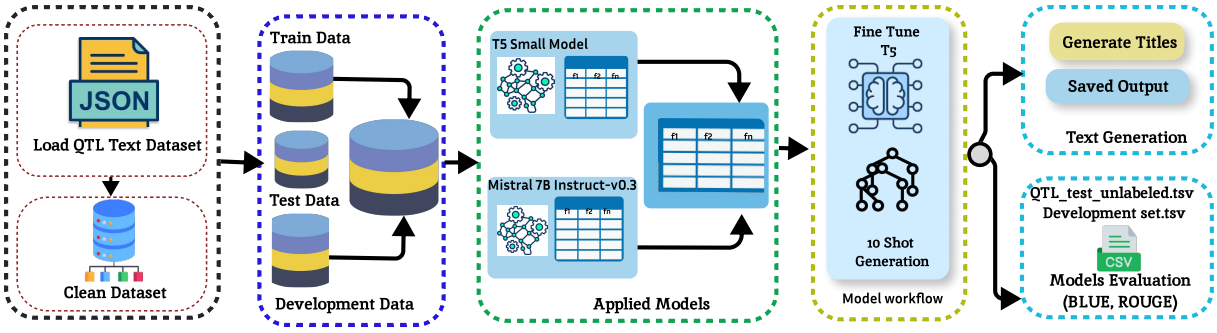

Figure 1: **Complete workflow for scientific title generation using T5-Small fine-tuning and Mistral-7B zero-shot prompting**

Table 1: Summary of the experimental setup for the T5-small and Mistral-7B models.

Feature	T5-small	Mistral 7B
Model type	Encoder–decoder (Seq2Seq)	Decoder-only (causal LM)
Learning setup	Supervised fine-tuning	10-shot in-context prompting

Table 2: Development & test scores for title generation.

Metric	T5-small	Mistral 7B
Development Metrics		
BLEU	12.63	9.62
ROUGE-2	0.1989	0.1837
ROUGE-L	0.3445	0.3353
Test Metrics		
BLEU	11.36	8.40
ROUGE-2	0.2144	0.1940
ROUGE-L	0.3619	0.3458

Table 3: Sample generated titles from the test set.

PMID	Generated Title
34902587	Epidemiology of a Porcine Circovirus Type 3 in a Boar Stud.
35451025	Simulations of a Three Breed Cross breeding Genome Using Imputed Data.

3 Results

We evaluate both models using BLEU, ROUGE-2, and ROUGE-L. The development set is used for metric computation, while the test set is used to compare final title generation quality. Table 2 presents the scores for both models. T5-small shows strong performance on both sets, while Mistral 7B produces competitive results, especially in ROUGE metrics, despite using only a 10-shot prompt. Example outputs from the test set are shown in Table 3.

4 Conclusion

This project compared two approaches for generating scientific titles from abstracts. The T5-small model learned the task through supervised fine-tuning, while the Mistral 7B model generated titles using a 10-shot prompt without additional training. Both methods produced clear and meaningful titles showing different strengths.

References

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Use of AI Tools

AI-assisted writing tools were used to improve grammar, clarity, and formatting.