From NLI to Classification: Entailment Learning for Low-Resource Text Classification

Rumana Ferdous Munne[1] Noriki Nishida[1] Shanshan Liu[1]
Narumi Tokunaga[1] Yuki Yamagata[2,3] Kouji Kozaki[4] Yuji Matsumoto[1]
[1]RIKEN Center for Advanced Intelligence Project (AIP)
[2]RIKEN R-IH,[3]RIKEN BRC
[1]Osaka Electro-Communication University
{rumanaferdous.munne, noriki.nishida, shanshan.liu, narumi.tokunaga
yuki.yamagata, yuji.matsumoto}@riken.jp
kozaki@osakac.ac.jp

Abstract

In real-world scenarios, text classification often faces the challenge of limited labeled data, especially for rare or emerging classes. Traditional methods struggle in these situations, requiring new approaches that can generalize to unseen or sparsely annotated classes. This challenge is particularly common in the biomedical field, where data is expensive to annotate, and new diseases and treatments frequently emerge. This paper proposes an entailment-based framework for zero and few-shot text classification by reframing the task as a natural language inference (NLI) problem. Leveraging pre-trained language models, the approach infers labels for unseen classes without additional fine-tuning. For few-shot scenarios, minimal task-specific fine-tuning significantly enhances performance. Our findings highlight the potential of entailment-based learning as a versatile and effective paradigm for text classification in low-resource environments.

1 Introduction

Text classification is a fundamental task in natural language processing (NLP), with applications in areas like sentiment analysis and document categorization. However, in real-world settings, a major challenge is the scarcity of labeled data, especially for rare or emerging classes. This is particularly problematic in domains like biomedicine, where data is expensive to annotate and often sparse for specific conditions. Traditional supervised methods struggle to generalize to unseen classes, making zero-shot and few-shot learning approaches increasingly important for addressing these challenges.

In the biomedical field, accurate text classification is essential for tasks such as annotating electronic health records (EHRs), extracting information from scientific literature, and validating biomedical ontologies. The complexity of medical terminology and the constant emergence of new diseases further complicate the creation of comprehensive labeled datasets. This makes biomedical text classification an ideal application for zero-shot and few-shot learning, where minimal or no labeled data is required to classify new, unseen categories.

In this paper, we propose an entailment-based learning framework that leverages pre-trained language models for zero- and few-shot text classification. By reframing text classification as a natural language inference (NLI) problem, our approach evaluates the relationship between input text and class-specific hypothesis templates. This allows the model to determine whether a given text entails a specific class label without requiring task-specific fine-tuning. By using large-scale pre-trained models, we avoid the need for extensive labeled data while still achieving effective classification for previously unseen classes.

To enhance performance in few-shot settings, we further explore the integration of task-specific fine-tuning on a minimal amount of labeled data. This enables our model to adapt to specific domains and improve classification accuracy in scenarios where only a small set of labeled examples is available. Through extensive experiments across multiple benchmark datasets, we demonstrate that our method

achieves competitive results in zero-shot settings and consistently outperforms existing baselines when fine-tuned with limited annotations.

Our findings underscore the potential of entailment-based learning as a flexible and powerful approach to text classification in low-resource environments, particularly in the biomedical domain. By combining the strengths of pre-trained models with the power of textual entailment reasoning, our framework provides a scalable solution to the challenges posed by sparse labeled data, enabling accurate classification across a broad range of medical and scientific texts.

Experimental results shows that our method achieves competitive results in zero-shot settings and consistently outperforms existing baselines when fine-tuned with limited annotations.

2 Related Works

Natural Language Inference (NLI) has become pivotal for text classification tasks, especially in zero-shot and few-shot settings. Early datasets like SNLI [1] and MNLI [2] laid the foundation for NLI models, which treat text classification as an entailment problem—predicting whether a text input entails a target class hypothesis. Yin et al. (2019) proposed zero-shot text classification as a textual entailment problem [3], while Gera (2022) [4] introduced a self-training approach for this task. Other studies, such as Koutsomitropoulos (2021) [6] , applied zero-shot learning to validate biomedical ontology annotations, and Pamies (2023) enhanced zero-shot classification using weak supervision and entailment [8]. Additionally, GPT models, including ChatGPT, have been applied to zero-shot tasks like clinical NER [9]. These developments highlight NLI's potential as a versatile tool for tackling classification challenges across various domains.

3 Method

By leveraging pre-trained NLI models, classification tasks can be reframed as entailment problems, where the objective is to determine whether a given hypothesis (e.g., a process label) is entailed by an input text (the premise). This approach provides a systematic and flexible way to address classification challenges involving limited labeled data and complex semantic relationships. In this work, we introduce an Entailment-Based Text Classification model designed to overcome the challenges of classifying biomedical text. In this framework, input text passages are treated as premises, predefined class labels are reformulated as hypotheses, and the model predicts which labels are logically entailed by the text. Figure 1 illustrates the functionality of our model. By leveraging the strengths of NLI, our approach effectively bridges the gap between textual data and semantic classification tasks in the biomedical domain.

3.1 Converting Labels into Hypotheses

The first step in our approach is to convert the target class labels into hypotheses suitable for an NLI-based classification task. This involves transforming class label names into a format compatible with textual entailment. We experimented with two hypothesis templates: one using the exact class label as it is, and another rephrased into a descriptive format, such as "This text is about <class label>." These templates enable the model to establish the relationship between the input text passage (the premise) and each potential class label (the hypothesis). Our experiments found that the direct label template yielded the best results.

3.2 Converting Classification Data into Entailment Data

To adapt classification tasks to the NLI framework, we reformulated the dataset into an entailment-compatible format. For each data split (train, dev, and test), each input text passage (the premise) was paired with a positive hypothesis corresponding to its true label, while the remaining labels were paired as negative hypotheses. In the zero-shot setup, unseen labels were excluded from training and evaluated only during testing to ensure a true zero-shot scenario. In the few-shot setup, a limited number of labeled examples were introduced for fine-tuning, enabling the model to better learn the mapping between text and hypotheses while still maintaining generalization to unseen labels.

3.3 Entailment Model Learning

We utilized widely recognized state-of-the-art pretrained models BART-Large-MNLI [7] for text classification based on natural language inference (NLI). Our proposed approach explores two primary setups: Zero-Shot Learning and Few-Shot Learning.

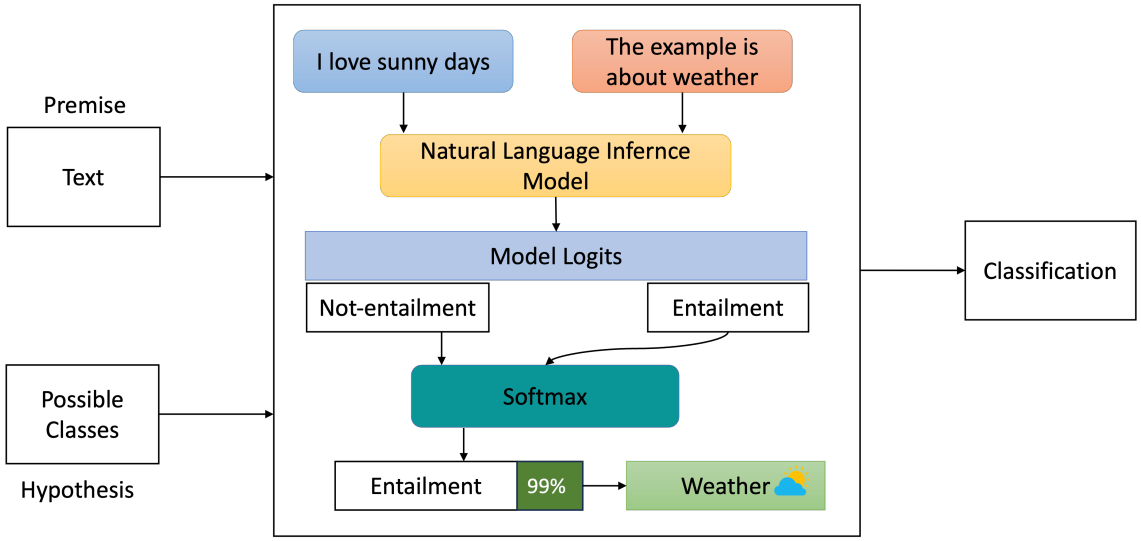

Figure 1 Entailment Based Classification

3.3.1 Zero-Shot Learning

In the zero-shot setup, we directly applied the pretrained entailment models to the test sets without any task-specific fine-tuning. This invloves no training phase and the model solely leverages its pretrained knowledge to classify entirely unseen labels based solely on the input text and hypotheses.

3.3.2 Few-Shot Learning

In the few-shot setup, we fine-tuned the pretrained NLI models using our small-scale dataset. This fine-tuning process allowed the model to learn domain-specific patterns from a limited set of labeled examples, improving its ability to generalize to both seen and unseen labels. By leveraging this small-scale dataset, we significantly enhanced the model's performance on classification tasks. Since the original dataset was not designed for a Natural Language Inference (NLI) task, we created a custom NLI dataset to adapt it for this purpose. We converted the positive examples into a pairwise format, where each example consists of a text input passage and a corresponding hypothesis, labeled with "entailment." In cases where the dataset contained multiple related classes, we generated several positive pairwise examples, each corresponding to a different class. Additionally, for each positive example, we generated a random negative example by pairing the premise with an unrelated hypothesis, labeled as "not-entailment." This balanced approach provided the model with an equal distribution of entailed and non-entailed pairs, ensuring effective training.

Class Name	#training	#test	Total
Neoplasms	2530	633	3163
Digestive system diseases	1195	299	1494
Nervous system diseases	1540	385	1925
Cardiovascular diseases	2441	610	3051
General pathological conditions	3844	961	4805
Total	**11550**	**2888**	**14438**

Table 1 Class distributions within the Medical Abstracts dataset.

4 Dataset

We utilized the Medical Abstracts dataset for our experiments. The raw Medical Abstracts dataset was sourced from Kaggle. This dataset originally comprises 28,880 medical abstracts categorized into five distinct classes of patient conditions, though only about half of the data is annotated. The original annotations consisted of numerical labels only. To create a usable medical text classification dataset, we processed the corpus by selecting only the labeled medical abstracts, assigning descriptive labels to the corresponding classes, and dividing the data into training and test sets.

Table 1 provides a summary of the processed Medical Abstracts dataset. Additionally, Table 2 outlines the inferred label keywords for each class. The processed corpus is made publicly available under the Creative Commons CC BY-SA 3.0 license at https://github.com/sebischair/Medical-Abstracts-TC-Corpus.

Class Name	Label Keywords
Neoplasms	neoplasms
Digestive system diseases	intestine, system, diseases
Nervous system diseases	nervous, system, diseases
Cardiovascular diseases	cardiovascular, diseases
General pathological conditions	general, pathological, conditions

Table 2 Class names and their corresponding label keywords.

Settings	Model	F1-score
Similarity Based	Lbl2TransformerVec*	56.46
Zero-shot	Zero-shot Entailment*	57.28
	Proposed zero shot model	57.19
Few-shot	Proposed few shot model	**67.34**

Table 3 Comparison of F1-scores across different models and settings. *Results from Schopf et al. [5]

5 Experiment

5.1 Implementation

We evaluate our zero-shot model using BART-Large-MNLI for its expertise in textual entailment task. For the few-shot model, we fine-tune the pretrained BART-Large-MNLI model using our limited NLI training data, which is constructed from the training dataset through the process described in 3.2. All training sessions utilized the Adam optimizer with a learning rate of 2×10^{-5} and a weight decay of 0.01. The fine-tuning process was executed using PyTorch and the Hugging Face Transformers library.

5.2 Result and Discussion

The primary objective of this task is to classify medical abstracts into five condition classes using a processed dataset. Our model demonstrates competitive performance compared to existing state-of-the-art methods. Specifically, we compare our approach with the Lbl2TransformerVec and Zero-shot Entailment models, as reported by Schopf et al. (2022)[5] in Table 5.2. The Lbl2TransformerVec model utilizes label embeddings combined with transformer-based architectures for similarity-based classification, while the Zero-shot Entailment model applies a zero-shot learning approach using a pre-trained DeBarta architecture.While this Entailment model is conceptually similar, it features distinct task formulations, as well as design and implementation adaptations.

In the zero-shot setting, our proposed model achieves an F1-score of 57.18, which is highly comparable to the Zero-shot Entailment model's score of 57.88. This result demonstrates that our model effectively handles classification tasks without requiring task-specific fine-tuning. However, the true strength of our approach lies in the few-shot setting, where our model achieves a significant improvement with an F1-score of 67.34. This represents a nearly 10% increase in performance compared to the zero-shot setting, underscoring the benefits of fine-tuning on a small, task-specific NLI (Natural Language Inference) dataset derived from the training split of the Medical Abstracts dataset.

The fine-tuning process enables our model to address the challenges of classifying instances with previously unseen or partially seen labels more effectively. By allowing the model to generalize in these scenarios, we address a critical need in medical text classification, where labeled data is often sparse, and emerging conditions frequently lead to new, unseen categories.

Overall, this result underscores the value of integrating fine-tuning on task-specific datasets with a robust pre-trained model, paving the way for more effective handling of challenging classification scenarios in biomedical contexts.

6 Conclusion

This paper presented an entailment-based learning framework for zero- and few-shot text classification, with a particular focus on addressing challenges in the biomedical domain. By leveraging pre-trained language models and reformulating classification as a natural language inference (NLI) problem, our approach effectively handles unseen classes without requiring extensive labeled data.

The integration of fine-tuning with minimal labeled data further demonstrated significant improvements in classification accuracy, highlighting the adaptability of the framework to low-resource settings. These results underscore the potential of NLI-based entailment learning as a powerful and scalable solution for text classification, particularly in domains like biomedicine, where annotated data is scarce and new categories continuously emerge.

References

[1] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pages 632–642, 2015.

[2] Adina Williams, Nikita Nangia, and Samuel Bowman. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**,pages 1112–1122, 2018.

[3] Wenpeng Yin, Jamaal Hay, and Dan Roth. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pages 3914–3923, 2019.

[4] Ariel Gera, Alon Halfon, Eyal Shnarch, Yotam Perlitz, Liat Ein-Dor, and Noam Slonim. Zero-shot text classification with self-training. In **Conference on Empirical Methods in Natural Language Processing**, 2022.

[5] Tim Schopf, Daniel Braun, and Florian Matthes. Evaluating unsupervised text classification: Zero-shot and similarity-based approaches. In **Proceedings of the 2022 6th International Conference on Natural Language Processing and Information Retrieval**, pages 6–15, 2022.

[6] Dimitrios Koutsomitropoulos. Validating ontology-based annotations of biomedical resources using zero-shot learning. In **The 12th International Conference on Computational Systems-Biology and Bioinformatics**, pages 37–43, 2021.

[7] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pages 7871–7880, July 2020.

[8] Marc Pàmies, Joan Llop, Francesco Multari, Nicolau Duran-Silva, César Parra-Rojas, Aitor González-Agirre, Francesco Alessandro Massucci, and Marta Villegas. A weakly supervised textual entailment approach to zero-shot text classification. In **Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pages 286–296, 2023.

[9] Yan Hu, Qingyu Chen, Jingcheng Du, Xueqing Peng, Vipina Kuttichi Keloth, Xu Zuo, Yujia Zhou, Zehan Li, Xiaoqian Jiang, Zhiyong Lu, and others. Improving large language models for clinical named entity recognition via prompt engineering. In **Journal of the American Medical Informatics Association**, page ocaad259, 2024.