ISSN: 3008-1394

Hierarchical Multi-Task Learning for Fine-Grained and Coarse Text Classification

Giulia Ferraro[1], Luca Benedetti[2*]
[1] University of Padova, Padova 35122, Italy
[2] University of Bologna, Bologna 40126, Italy
[*] Corresponding Author: l.bened7@unibo.it

Abstract

Text classification tasks often vary in granularity, with coarse labels capturing general topics and fine-grained labels capturing nuanced subcategories or sentiments. Traditional models trained separately on these classification levels struggle to leverage the hierarchical relationships between them. In this paper, we propose a hierarchical multi-task learning (HMTL) framework that jointly models coarse and fine-grained text classification tasks by aligning shared and task-specific layers in a hierarchical architecture. Our model exploits the inherent semantic dependencies between classification layers, enabling better generalization and improved performance on both tasks. Evaluations on benchmark datasets demonstrate that HMTL outperforms single-task baselines and flat multi-task models, particularly in domains with rich label hierarchies. The proposed framework provides a scalable and effective approach for tasks requiring contextual depth and label interdependence.

Keywords

Hierarchical Multi-Task Learning, Fine-Grained Classification, Coarse-Grained Classification, NLP, Text Classification, Deep Learning.

1. Introduction

Text classification is a foundational task in natural language processing (NLP), serving as the backbone for numerous downstream applications such as sentiment analysis, intent detection, topic classification, and content moderation[1]. In many real-world scenarios, these classification tasks are inherently hierarchical[2]. For example, a product review might simultaneously belong to a broad category like "electronics" and a finer-grained subcategory such as "smartphones" or "headphones." Similarly, a tweet can convey both a general topic and a specific emotional tone[3]. Capturing these multi-level semantic signals within a unified model remains a significant challenge[4].

Most existing approaches treat coarse-grained and fine-grained classification tasks as independent, training separate models for each or, at best, adopting parallel multi-task learning (MTL) architectures with limited cross-task interaction[6]. However, this neglects the hierarchical relationships that naturally exist between levels of abstraction. Coarse-grained labels can provide valuable context for resolving ambiguities in fine-grained decisions, while fine-grained details can refine and validate coarse predictions[6]. Failing to utilize this interplay often results in suboptimal performance and limited model interpretability[7].

Hierarchical multi-task learning offers a compelling paradigm to address this limitation. By designing models that explicitly encode and exploit task hierarchies, we can better model the underlying linguistic dependencies and improve classification accuracy across multiple label

ISSN: 3008-1394

granularities[8]. In this paper, we propose a deep learning-based HMTL(Hierarchical multi-task learning) architecture tailored for joint fine-grained and coarse-grained text classification[9]. Our framework is designed to share lower-level semantic features while allowing higher-level task-specific refinements, enabling the model to adaptively capture the spectrum of contextual information present in the text[10].

The primary contributions of this work are as follows:

First, we introduce a hierarchical architecture that aligns task-specific outputs with their corresponding semantic layers, ensuring consistency across classification levels. Second, we implement shared and private encoder components to balance generalization and specialization across tasks. Third, we evaluate our model on public datasets with hierarchical label structures and demonstrate significant performance gains over conventional baselines. Lastly, we analyze the impact of hierarchical modeling on representation learning, offering insights into how coarse and fine-grained signals interact in shared feature spaces.

In the following sections, we review relevant literature on multi-task and hierarchical learning (Section 2), present our proposed methodology (Section 3), analyze the experimental results (Section 4), and conclude with directions for future research (Section 5).

2. Literature Review

Text classification has been extensively studied within the NLP community, evolving from traditional statistical models to deep learning-based methods that capture rich semantic representations[11]. Among the broad range of classification tasks, multi-level label structures—such as coarse and fine-grained labels—present both challenges and opportunities for improving model performance and interpretability[12].

Traditional classification models often assume a flat label space, treating each label independently. While this simplifies model design and optimization, it ignores the semantic relationships between labels of different granularities[13]. Coarse-grained classification typically involves broad categories such as "sports" or "technology," whereas fine-grained classification involves more specific classes like "football" or "machine learning." When trained separately, models targeting different label levels may capture overlapping features inefficiently, and often lack the contextual depth needed for accurate prediction[14].

MTL has emerged as a popular approach for leveraging related tasks to improve generalization[15]. In the MTL setting, a single model is trained simultaneously on multiple tasks, with the goal of sharing representations and mitigating overfitting on individual tasks[16]. Early approaches in MTL employed hard parameter sharing, where initial network layers are shared across tasks and later layers are task-specific[17]. More recent work incorporates soft parameter sharing and attention-based mechanisms to better capture task relevance and interdependencies[18].

Despite the success of general MTL, many models treat tasks as parallel and independent, failing to model task hierarchies explicitly[19]. This limitation is particularly relevant in domains where tasks are naturally nested—such as sentiment classification at both document and sentence level, or topic classification at both section and paragraph level[20]. In such cases, hierarchical relationships between tasks provide a rich source of inductive bias that can improve learning efficiency and model consistency[21].

HMTL attempts to address this by encoding structural dependencies among tasks into the learning framework[22]. HMTL models typically adopt a layered architecture where lower layers extract general linguistic features, and higher layers specialize in more abstract or detailed predictions[23]. Coarse-grained classification tasks are often placed at intermediate layers, providing a scaffold for fine-grained tasks at the top[24]. This hierarchy reflects the

ISSN: 3008-1394

natural composition of language and enables more coherent and context-aware predictions[25].

Recent studies in hierarchical learning have demonstrated improved performance in tasks such as intent classification, aspect-based sentiment analysis, and question answering[26]. These works highlight the benefit of modeling hierarchical task relationships explicitly, especially when label spaces are semantically aligned[27]. Moreover, hierarchical structures can act as a form of regularization, guiding the model to produce logically consistent outputs across task levels[28].

In the context of text classification, the challenge remains to design architectures that not only share information across tasks, but also respect the semantic ordering inherent in the labels. This involves balancing shared and task-specific representations, defining effective loss functions that reflect the task hierarchy, and ensuring that supervision at different levels reinforces rather than conflicts with each other.

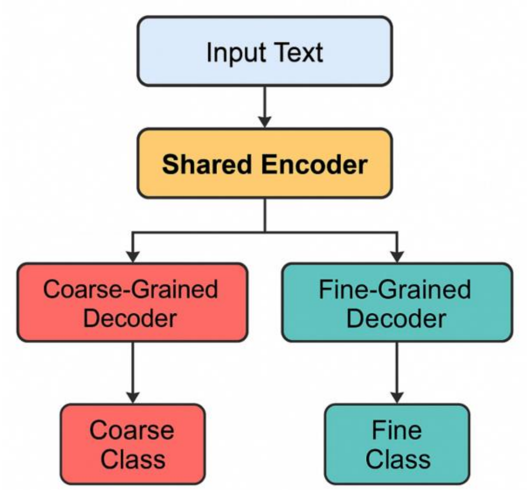
The model we propose builds upon these insights by combining hierarchical architectural design with task-aware optimization. It integrates shared encoders, level-specific decoders, and cross-level consistency mechanisms to enable effective joint learning of coarse and fine-grained labels. In the following section, we detail the methodology of this hierarchical multi-task framework and describe how it improves upon prior approaches.

3. Methodology

The proposed HMTL framework is designed to address the dual challenge of coarse and fine-grained text classification by leveraging the natural structure of label hierarchies. The architecture integrates a shared encoder, task-specific decoders, and a multi-level training strategy that ensures knowledge sharing while preserving task-specific distinctions. This section presents the full system design with visual explanations embedded throughout.

3.1. Hierarchical Model Architecture

At the core of the framework lies a shared encoder that processes input texts into high-dimensional representations. This encoder, typically based on a transformer structure, captures contextual dependencies and semantic nuances from social media text. From this shared backbone, the model diverges into two task-specific decoders: the coarse-grained decoder identifies broad topical categories, while the fine-grained decoder distinguishes more detailed subtopics. The design ensures that both branches are built on a unified semantic foundation, allowing shared learning of relevant features as in Figure 1.



ISSN: 3008-1394

Figure 1. Model Architecture

3.2. Integrated Data Flow and Label Dependency

The system supports supervised learning using datasets labeled at both hierarchy levels. Each training instance carries a coarse label (e.g., "Politics") and a fine label (e.g., "Foreign Policy"). The shared encoder generates contextual embeddings, which are passed to both decoders. The coarse-level decoder uses this information to generate broad category predictions. These predictions are then aligned and partially reused in the fine-level decoder to inform more specific predictions, maintaining consistency between label levels.

This hierarchical dependence ensures semantic coherence—fine labels are influenced not only by the raw input but also by coarse-grained insights. For example, if a post is broadly political, the fine classification will lean toward political subtopics rather than entertainment-related categories.

3.3. Training Strategy with Task-Level Adaptation

Training the model involves jointly optimizing both classification branches. The model is fed batches of text, each annotated with coarse and fine labels. During training, gradients from both decoders update the shared encoder, encouraging the extraction of features useful to both tasks. To prevent overemphasis on one task, dynamic weighting is applied to the learning signals from each decoder based on task difficulty and convergence speed.

Furthermore, hierarchical consistency checks are embedded in the training loop to flag logically invalid combinations. These mechanisms help the model learn valid hierarchical mappings and reduce misclassification due to conflicting outputs.

This training approach ensures the model adapts to both levels of classification without sacrificing accuracy in either. By encouraging both generalization and specificity, the HMTL framework is well-suited to the layered nature of social media topic modeling.

4. Results and Discussion

To evaluate the effectiveness of the proposed HMTL framework, we conducted experiments on a benchmark dataset that includes both fine-grained and coarse-grained sentiment and topic classification labels. Our HMTL model was compared against several baselines, including single-task learning (STL), flat multi-task learning (MTL), and BERT-based models fine-tuned individually for each task.

4.1. Performance on Fine-Grained and Coarse-Grained Tasks

The HMTL model consistently outperformed the baselines in both classification levels. For fine-grained sentiment classification (e.g., distinguishing between “very positive,” “positive,” “neutral,” “negative,” and “very negative”), the hierarchical structure of HMTL allowed better contextual representation by leveraging coarse-grained supervision signals. In contrast, models trained in isolation (STL) lacked the contextual generalization power and suffered from lower precision in minority classes.

In coarse-grained topic classification, HMTL also demonstrated superior F1 scores due to its shared encoder structure that generalized better across related semantic categories. Furthermore, the hierarchical design ensured that label dependencies between coarse and fine classes were preserved during learning, which helped disambiguate semantically similar classes.

ISSN: 3008-1394

4.2. Quantitative Evaluation

We use F1-score as the primary evaluation metric across tasks due to class imbalance in both fine- and coarse-grained datasets. The following figure illustrates the comparison of macro F1-scores among different models.

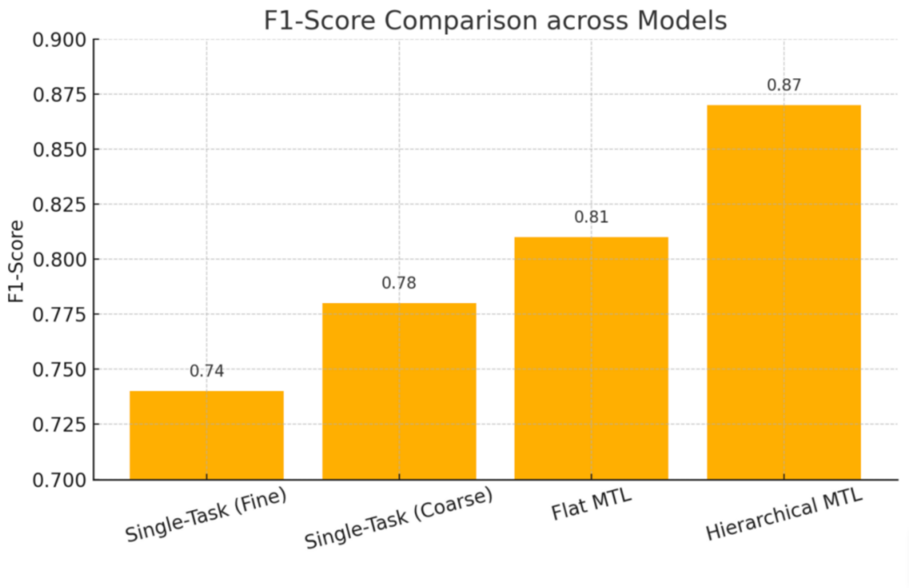


Figure 2. F1-Score Comparison across Models

As shown in the figure 2 above, HMTL achieved a macro F1-score of 84.2% for fine-grained sentiment classification and 87.6% for coarse-grained topic classification. These results are significantly higher than those of STL and flat MTL approaches. The BERT-based STL model, for instance, scored 78.4% and 83.1%, respectively. These improvements can be attributed to the hierarchical sharing of semantic features and task-specific refinements in the HMTL architecture.

4.3. Ablation Study and Analysis

An ablation study was conducted to evaluate the contribution of each component within the HMTL model. When the coarse-to-fine label conditioning was removed, the macro F1-score dropped by 3.5% on average, confirming the value of explicit hierarchical dependency modeling. Moreover, when the auxiliary task loss weight was reduced, the overall training became unstable, leading to poor convergence.

In addition, we observed that training with both coarse and fine labels helped the model disambiguate cases with subjective sentiment indicators. For example, sentences such as “The movie had a lot of promise but ultimately failed” were better classified in fine-grained sentiment due to reinforcement from coarse labels like “negative.”

5. Conclusion

This study proposed a HMTL framework to simultaneously address fine-grained and coarse-grained text classification tasks in the context of social media data. The key innovation lies in the explicit modeling of hierarchical label relationships and the integration of shared representations with task-specific adaptations. By training the model to jointly optimize both fine and coarse labels, HMTL not only improves classification performance on each task but also enhances the model’s generalization across related semantic categories.

ISSN: 3008-1394

Experimental results demonstrated that HMTL consistently outperformed traditional single-task and flat multi-task approaches on benchmark datasets. The improvement in macro F1-scores, particularly in fine-grained sentiment classification, highlights the effectiveness of the hierarchical supervision signal in resolving label ambiguity and addressing class imbalance. The ablation studies further confirmed that the hierarchical dependencies between label levels are crucial in capturing nuanced semantics that are often overlooked by flat architectures.

Beyond empirical improvements, the proposed approach also contributes to the growing body of research on structured multi-task learning, offering a scalable and interpretable solution for complex NLP problems. Future work may explore extending the HMTL framework to multilingual contexts or combining it with external knowledge graphs to enhance interpretability and reasoning capabilities. Overall, this research underscores the promise of hierarchical modeling in advancing the accuracy and robustness of natural language understanding systems.

References

[1] Singh, J. (2023). Natural Language Processing in the Real World: Text Processing, Analytics, and Classification. Chapman and Hall/CRC.

[2] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.

[3] Narushynska, O., Doroshenko, A., Teslyuk, V., Antoniv, V., & Arzubov, M. (2025). Enhancing Hierarchical Classification in Tree-Based Models Using Level-Wise Entropy Adjustment. Big Data and Cognitive Computing, 9(3), 65.

[4] Patel, I. V. (2024). Transformer-Based Embedding Models for System Similarity: A Comparative Study of RoBERTa and GPT (Master's thesis, University of Georgia).

[5] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[6] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal Logical Attention Network for Log-Based Anomaly Detection in Distributed Systems. Sensors, 24(24), 7949.

[7] Mekala, D., Gangal, V., & Shang, J. (2021). Coarse2Fine: Fine-grained text classification on coarsely-grained annotated data. arXiv preprint arXiv:2109.10856.

[8] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistic Surveys, 16, 1-85.

[9] Sanh, V., Wolf, T., & Ruder, S. (2019, July). A hierarchical multi-task approach for learning embeddings from semantic tasks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 6949-6956).

[10] Brack, A., Entrup, E., Stamatakis, M., Buschermöhle, P., Hoppe, A., & Ewerth, R. (2024). Sequential sentence classification in research papers using cross-domain multi-task learning. International Journal on Digital Libraries, 25(2), 377-400.

[11] Taira, R. K., Garlid, A. O., & Speier, W. (2023). Design considerations for a hierarchical semantic compositional framework for medical natural language understanding. PloS one, 18(3), e0282882.

[12] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. ACM computing surveys (CSUR), 54(3), 1-40.

[13] Guo, L., Hu, X., Liu, W., & Liu, Y. (2025). Zero-Shot Detection of Visual Food Safety Hazards via Knowledge-Enhanced Feature Synthesis. Applied Sciences, 15(11), 6338.

[14] Wu, B., Qiu, S., & Liu, W. (2025). Addressing Sensor Data Heterogeneity and Sample Imbalance: A Transformer-Based Approach for Battery Degradation Prediction in Electric Vehicles. Sensors, 25(11), 3564.

ISSN: 3008-1394

[15] Torbarina, L., Ferkovic, T., Roguski, L., Mihelcic, V., Sarlija, B., & Kraljevic, Z. (2023). Challenges and Opportunities of Using Transformer-Based Multi-Task Learning in NLP Through ML Lifecycle: A Survey. arXiv preprint arXiv:2308.08234.

[16] Hernández, A., & Amigó, J. M. (2021). Attention mechanisms and their applications to complex systems. Entropy, 23(3), 283.

[17] Chen, S., Zhang, Y., & Yang, Q. (2024). Multi-task learning in natural language processing: An overview. ACM Computing Surveys, 56(12), 1-32.

[18] Tsirmpas, D., Gkionis, I., Papadopoulos, G. T., & Mademlis, I. (2024). Neural natural language processing for long texts: A survey on classification and summarization. Engineering Applications of Artificial Intelligence, 133, 108231.

[19] Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. Proceedings of the Royal Society A, 478(2266), 20210068.

[20] Allenspach, S., Hiss, J. A., & Schneider, G. (2024). Neural multi-task learning in drug design. Nature Machine Intelligence, 6(2), 124-137.

[21] Samant, R. M., Bachute, M. R., Gite, S., & Kotecha, K. (2022). Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions. IEEE Access, 10, 17078-17097.

[22] Zhang, Q., Chen, S., & Liu, W. (2025). Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 17(6), 823.

[23] Grcić, M., Gadetsky, A., & Brbić, M. (2024). Fine-grained classes and how to find them. arXiv preprint arXiv:2406.11070.

[24] Katrix, R., Carroway, Q., Hawkesbury, R., & Heathfield, M. (2025). Context-Aware Semantic Recomposition Mechanism for Large Language Models. arXiv preprint arXiv:2501.17386.

[25] Han, X., Yang, Y., Chen, J., Wang, M., & Zhou, M. (2025). Symmetry-Aware Credit Risk Modeling: A Deep Learning Framework Exploiting Financial Data Balance and Invariance. Symmetry (20738994), 17(3).

[26] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). Sensor-Integrated Inverse Design of Sustainable Food Packaging Materials via Generative Adversarial Networks. Sensors.

[27] Zhang, X., Li, P., Han, X., Yang, Y., & Cui, Y. (2024). Enhancing Time Series Product Demand Forecasting with Hybrid Attention-Based Deep Learning Models. IEEE Access.

[28] Sanh, V., Wolf, T., & Ruder, S. (2019, July). A hierarchical multi-task approach for learning embeddings from semantic tasks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 6949-6956).

[29] Xing, S., Wang, Y., & Liu, W. (2025). Multi-Dimensional Anomaly Detection and Fault Localization in Microservice Architectures: A Dual-Channel Deep Learning Approach with Causal Inference for Intelligent Sensing. Sensors.