Transformer-based Multi-Task Learning for NWDAF

Hyeonjae Jeong and Sangheon Pack
School of Electrical Engineering, Korea University, Seoul, Korea.
Email: qeqe@korea.ac.kr, shpack@korea.ac.kr

***Abstract*—With the transition toward 6G, the importance of enhancing NWDAF with advanced AI/ML capabilities is rapidly growing. Recent studies have explored the use of multi-task learning (MTL) to analyze networks by leveraging task inter-dependencies. However, conventional MTL approaches based on hard or soft parameter sharing fail to sufficiently capture the complex relationships among tasks. To address this limitation, we propose a transformer-based MTL model that explicitly models inter-task dependencies. By incorporating transformer, our method enables more fine-grained task relation modeling.**

***Index Terms*—Transformer, NWDAF, Multi-task Learning**

I. Introduction

The advent of 6G networks demands a paradigm shift in network intelligence, where the network data analytics function (NWDAF) must evolve into an AI-native component capable of supporting dynamic, data-driven decision making. Traditional approaches have primarily relied on single-task learning or conventional multi-task learning (MTL) frameworks with hard and soft parameter sharing [1], [2]. While these methods have demonstrated benefits in reducing training cost and sharing information across tasks, they remain limited in their ability to capture the complicated inter-dependencies that naturally exist among diverse analytic tasks in mobile core networks. Recent research highlights the necessity of exploiting such task relationships to improve both predictive accuracy, especially in scenarios where analytic outcomes are interlinked. To overcome the shortcomings of existing MTL strategies, this work introduces a transformer-based MTL model that explicitly models task-to-task dependencies, enabling more fine-grained representation learning and enhancing NWDAF's capability to provide accurate and efficient network analytics in 6G environments.

II. Transformer-based Multi-task Learning

The proposed model adopts a transformer-based architecture to enhance MTL for NWDAF in 6G networks, addressing the limitations of conventional hard and soft parameter sharing. The framework is organized into four main stages: input embedding, sequence integration, shared encoding, and task-specific decoding.

As shown in Figure 1, in the embedding stage, each task-specific input sequence is first projected into a fixed-dimensional representation through a dedicated linear embedding layer. To encode the source of each sequence, a task type embedding is added, and positional encoding is applied

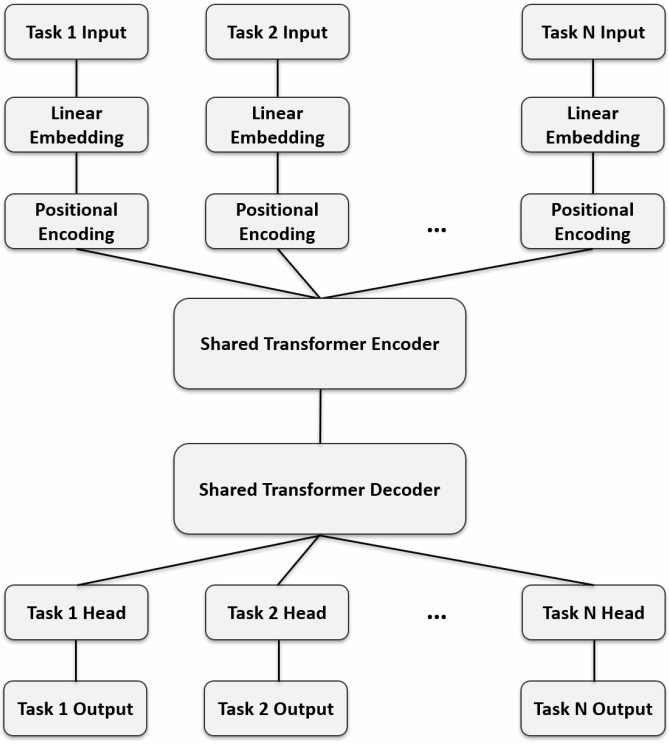

Fig. 1. Architecture of transformer-based multi-task Learning.

to preserve the temporal order of tokens. These representations are then concatenated into unified sequence that aggregates heterogeneous task information into a common feature space.

This integrated sequence is processed by the shared transformer encoder, which employs multi-head self-attention to learn deep task-aware representations. The encoder captures both global dependencies across tasks and local task-specific correlations, enabling more effective modeling of inter-task relationships compared to traditional parameter sharing strategies.The model introduces a set of learnable task query vectors, one per task, which serve as inputs to the transformer decoder. Through a combination of self-attention and cross-attention, the decoder refines these queries by selectively attending to the shared encoder memory. The resulting task-specific representations are passed to lightweight output heads, each implemented as a feed-forward projection network tailored to the objective of its corresponding analytic task, such as abnormal behavior detection, dispersion prediction, NF load

TABLE I
MODEL PERFORMANCE (MAE | RMSE)

Model	Abnormal Behavior	Dispersion	NF load	Service Volume
Hard Sharing MTL	1.47 | 2.1	1.07 | 1.25	1.92 | 2.33	1.29 | 1.9
Soft Sharing MTL	1.27 | 1.82	1.05 | 1.23	1.65 | 2.03	1.02 | 1.63
Transformer-based MTL (Proposed)	**0.51 | 0.52**	**0.89 | 0.93**	**1.81 | 1.9**	**0.86 | 1.17**

estimation, or service traffic volume forecasting.

By combining shared representation learning in the encoder with task-specialized decoding, the model explicitly captures task inter-dependencies while preserving flexibility for heterogeneous prediction requirements. Compared with hard sharing, which enforces a single representation across all tasks, and soft sharing, which loosely regularizes independent models, this encoder–decoder design enables more adaptive and fine-grained knowledge transfer, ultimately improving both predictive accuracy and resource efficiency in NWDAF's analytic functions.

III. EVALUATION RESULTS

We use a large-scale mobile traffic dataset collected in milan, Italy, to evaluate realistic NWDAF scenarios. We utilize data from Nov 1–10 for training, Nov 11–12 for validation, and Nov 13–14 for testing, after normalizing input features. Four representative prediction tasks are defined: abnormal behavior detection, service traffic volume prediction, dispersion analysis, and NF load prediction. All models are implemented in PyTorch 2.0 with comparable size and trained on an NVIDIA RTX 3060 GPU. Model performance is assessed using mean absolute error (MAE) and root mean squared error (RMSE), capturing both average and large prediction errors. We set hard parameter sharing MTL and soft parameter sharing MTL as the baselines.

Table I summarizes the predictive performance of different MTL architectures across four analytic tasks in NWDAF. The results clearly demonstrate the superiority of the proposed transformer-based MTL model compared to baseline models.

For abnormal behavior prediction, the proposed model achieves a substantial improvement, reducing the MAE and RMSE to 0.51 and 0.52, compared to 1.47/2.1 and 1.27/1.82 for hard and soft sharing, respectively. This highlights the ability of the transformer architecture to capture fine-grained temporal and spatial dependencies that are critical for detecting anomalous traffic behaviors. Similarly, in dispersion analysis, the proposed model yields the lowest errors 0.89/0.93, outperforming both hard 1.07/1.25 and soft sharing 1.05/1.23. The performance gains in these two tasks confirm that explicitly modeling inter-task dependencies provides tangible benefits when tasks are strongly correlated.

In the case of NF load estimation, the improvements are less pronounced, with the proposed model achieving 1.81/1.9 compared to 1.92/2.33 and 1.65/2.03. While the transformer still provides competitive accuracy, the relatively modest gains suggest that NF load prediction may rely on features less directly shared with other tasks, making it harder to benefit from cross-task information. Finally, for service volume prediction, the transformer-based MTL again achieves the best results , significantly outperforming both baselines.

Overall, these results confirm that the proposed transformer-based MTL framework provides consistently superior performance across tasks, with particularly strong gains in abnormal behavior and service volume prediction. By explicitly modeling task-to-task dependencies through attention mechanisms, the model achieves more effective knowledge transfer than conventional hard or soft sharing, thereby enhancing both predictive accuracy and robustness of NWDAF analytics in 6G environments.

IV. CONCLUSION

In this paper, we proposed a transformer-based MTL for NWDAF in the context of 6G networks. Unlike conventional MTL approaches based on hard or soft parameter sharing, our model leverages the encoder–decoder structure of the transformer to explicitly capture inter-task dependencies through attention mechanisms. Future work will explore resource-efficient model designs to enable NWDAF to deliver advanced intelligence with minimal resource consumption. By incorporating lightweight attention mechanisms and adaptive inference strategies, we aim to extend the benefits of transformer-based MTL while ensuring that analytics functions can be deployed efficiently even under constrained computational and energy budgets in next-generation mobile core networks.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-II221015, Development of Candidate Element Technology for Intelligent 6G Mobile Core Network) and National Research Foundation (NRF) of Korea Grant funded by the Korean Government (MSIT) (No. RS-2024-00341965).

REFERENCES

[1] A. P. Iyer *et al.*, "Mitigating the Latency-Accuracy Trade-off in Mobile Data Analytics Systems," in *Proc. ACM Mobile Computing and Networking (MOBICOM)*, New Delhi, India, October 2018.

[2] H. Lian *et al.*, "Towards Effective Personalized Service QoS Prediction from the Perspective of Multi-Task Learning," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 2587–2597, September 2023.