Breaking Domain Barriers: Mixture of Experts for Cross-Domain Fake News Detection

Angelica Liguori[1*], Francesco Sergio Pisani[1], Carmela Comito[1], Massimo Guarascio[1], Giuseppe Manco[1]

[1]Institute for High Performance Computing and Networking, Italian National Research Council, Via P. Bucci 8/9C, Rende, 87036, Italy.

*Corresponding author(s). E-mail(s): angelica.liguori@icar.cnr.it;
Contributing authors: francescosergio.pisani@icar.cnr.it;
carmela.comito@icar.cnr.it; massimo.guarascio@icar.cnr.it;
giuseppe.manco@icar.cnr.it;

Short Summary of the contribution
In this work, we introduce MERMAID, a hierarchical ensemble framework designed to improve fake news detection across multiple domains. MERMAID is based on a mixture of experts that leverages the knowledge from different specialized large language models to classify examples from unknown domains. MERMAID embeds a model merging mechanism that reduces the number of different experts in a preliminary stage before feeding the MoE ensemble.

What is the main claim of the paper? Why is this an important contribution to the machine learning literature?
The paper introduces MERMAID, a novel hierarchical ensemble framework for cross-domain fake news detection that eliminates the need for retraining on new datasets. This work advances domain generalization by enabling fake news detection in unseen domains without labeled target data. Unlike existing methods, MERMAID optimizes expert selection by reducing the number of models while maintaining high performance.

What is the evidence you provide to support your claim? Be precise.
We evaluate our framework through three complementary analyses across five real-world datasets. First, we conduct a comparative study against four state-of-the-art methods. Next, we perform an ablation study to isolate the impact of individual

components and quantify their contributions. Finally, we introduce a few-shot learning analysis to assess adaptability: when minimal labeled data from unseen domains becomes available (e.g., 0.02%), MERMAID achieves up to a $\sim 30\%$ improvement over domain-adaptation approaches. In addition to this quantitative results, we conducted a qualitative analysis of expert selection dynamics within the MoE framework by examining how the gating mechanism distributes importance across different expert models on various datasets.

What papers by other authors make the most closely related contributions, and how is your paper related to them?

Our work falls within the field of multi- and cross-domain fake news detection. The main related works are as follows:

- Li Y, Lee K, Kordzadeh N, et al (2021) Multi-Source Domain Adaptation with Weak Supervision for Early Fake News Detection. In: IEEE International Conference on Big Data, pp 668–676.
- Wang Y, Ma F, Jin Z, et al (2018) EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, p 849–85.
- Shu K, Mosallanezhad A, Liu H (2022) Cross-Domain Fake News Detection on Social Media: A Context-Aware Adversarial Approach. Frontiers in Fake Media Generation and Detection pp 215–23.

Our solution differs significantly from these works as it handles domain transfer and adaptation. Cross-domain models focus on transferring generalizable knowledge from a source domain, assuming that the patterns learned from the source domain can be adapted to improve performance on the target domain. Our approach does not rely on knowledge transfer from a predefined source domain. Instead, it learns specialized models for different domains and combines them through a MoE framework to dynamically determine the best-performing models for predicting new data. Moreover, Cross-domain models typically require further training or fine-tuning on the target domain before achieving optimal performance. This necessitates additional labeled samples from the target domain, which might not always be available. Our approach does not require additional training or fine-tuning when encountering a new domain. Instead, it selects the most reliable domain-specific models from the ensemble, making it fully cross-domain without explicit adaptation.

Have you published parts of your paper before, for instance in a conference? If so, give details of your previous paper(s) and a precise statement detailing how your paper provides a significant contribution beyond the previous paper(s).

A very preliminary version of this work was published in the following paper, which first introduced the idea of developing a cross-domain hierarchical ensemble model leveraging Large Language Models to distinguish between real and fake content:

- Comito C., Guarascio M., Liguori A., Manco G., Pisani F.S.: *Beyond the Horizon*: Using Mixture of Experts for Domain Agnostic Fake News Detection. Discovery Science 2024. DOI: https://doi.org/10.1007/978-3-031-78980-9_25

In the following, the main points of novelty and difference with the conference paper are summarized:

- Unlike most existing methods, which rely on domain adaptation requiring prior knowledge of the target domain, MERMAID adopts a strategy to handle the domain generalization problem. The ensemble strategy combined with the usage of LLMs ensures that fake news detection remains effective even in unseen domains, distinguishing it as one of the few methods capable of performing cross-domain detection without needing retraining on new datasets.
- As introduced above, our framework integrates multiple specialized language model-based fake news detectors through a Mixture-of-Experts (MoE) mechanism, dynamically selecting the most relevant models for each input. Unlike our previous work, MERMAID incorporates a model merging mechanism that reduces the numbers/combines different experts in a preliminary stage before feeding the MoE ensemble. This additional step enhances the efficiency and coherence of expert selection, ensuring more effective processing of diverse inputs.
- We evaluate our framework through three complementary analyses across five real-world datasets. First, we conduct a comparative study against four state-of-the-art methods. Next, we perform an ablation study to isolate the impact of individual components and quantify their contributions. Finally, we introduce a few-shot learning analysis to assess adaptability: when minimal labeled data from unseen domains becomes available (e.g., 0.02%), MERMAID achieves up to a $\sim 30\%$ improvement over domain-adaptation approaches.
- We conduct a qualitative analysis of expert selection dynamics within the MoE framework by examining how the gating mechanism distributes importance across different expert models on various datasets, revealing dataset-dependent patterns in expert activation.
- We have extended the related works section and conducted a more comprehensive literature review.