

Text Classification Model Through Maximizing Posterior Evidence Over Informative Sub-space

**Zhiwei SUN, Jun BAI, Zhuofan CHEN, Chen LI,
Wenge RONG, Zhang XIONG**

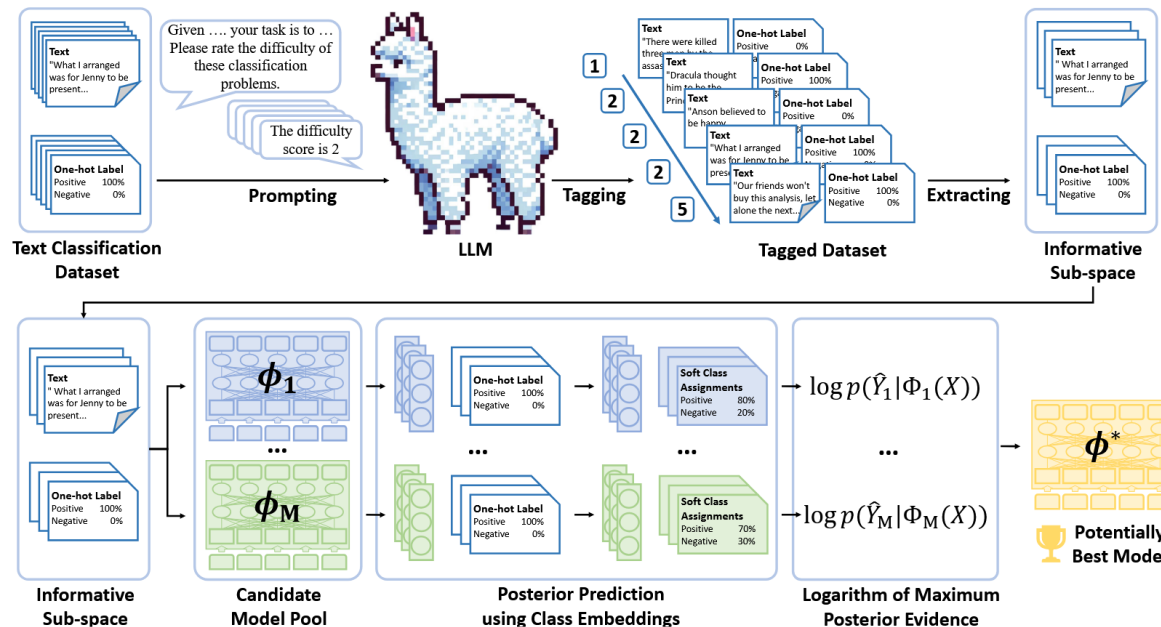
Frontiers of Computer Science, DOI: [10.1007/s11704-025-41380-7](https://doi.org/10.1007/s11704-025-41380-7)

Problems & Ideas

- Task of model selection
 - Model selection aims at selecting the model most suitable to the fine-tuning on the given dataset from candidate models, in which the brute force fine-tuning for all candidates should be avoided.
- Problems of model selection approaches for text classification:
 - The uncertainty of sample in categorization has been ignored.
 - Using all samples to estimate the transferability, where the easy samples and too difficult samples are noises that hinder the elicitation of true transferability of model.
- Ideas: Leveraging soft class assignments to capture the fine-grained uncertainty in categorization, and forming informative sub-space to prevent from the influence of noisy samples.

Main Contributions

- Contributions:
 - A novel “posterior prediction using class embeddings” method to quickly assess the affinity of each text to class embeddings, thereby calculating soft class assignments;
 - A novel metric (Logarithm of Maximum Posterior Evidence) is computed on samples with appropriate difficulty levels which constitute high-information subspaces within the dataset.



Samples of suitable difficulty are selected using an LLM, soft class assignments are computed, and model transferability is measured by the log evidence of these assignments.