Impact of Style in Personalization of Large Language Models

Aashnna Soni, Balpreet Kaur, Prakriti Shetty

TABLE OF CONTENTS

01
RESEARCH PROBLEM

Defining our research problem

Personalisation in LLMs

- ❖ Personalization in LLMs aims to connect their universal capabilities with the rising demand for individualized interactions.
- ❖ Linguistic style plays a crucial role in language, and recent advances have focused on developing style representations using authorship verification tasks

Key PRIOR ART

- ❖ A.Salemi et al. (2023)
  - ➢ LaMP: a new framework for training and evaluating personalized LLMs
- ❖ Wegmann et al. (2022)
  - ➢ Sentence embedding model designed explicitly to encapsulate linguistic style
- ❖ Neelakanteswara et al. (2024)
  - ➢ Computes the average of Wegmann embeddings to encapsulate the overall stylistic tendencies of the author

RESEARCH QUESTION

Can we have an approach where we **combine content-dependent** and **content-independent** style representations to generate a personalised news headline for a given query document, such that it captures overall stylistic tendencies?

02

APPROACH

Logical and Mathematical basis
of our proposed solution

DENSE RETRIEVAL

1. Original query input to an LLM generates k query variants.
2. Dense retrieval and FAISS indexing to retrieve the top m semantically similar documents

STYLE EMBEDDINGS

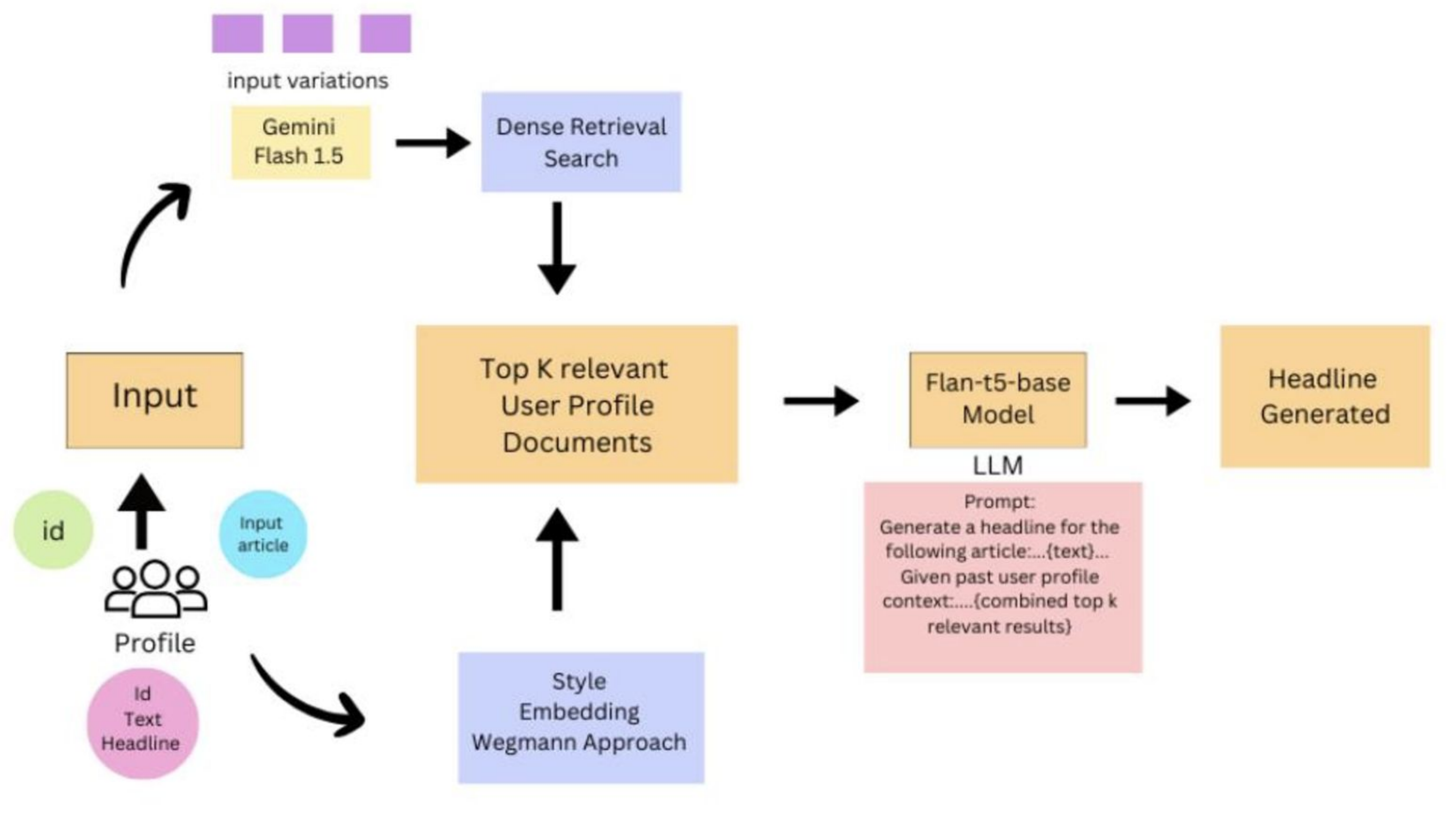
1. Employ Wegmann et al. (2022)'s model to extract style embeddings from the user profile documents
2. Compute the average of these embeddings to encapsulate the overall stylistic tendencies.
3. Arrange the inputs in descending order based on their cosine similarity to the average embedding.
4. Retrieve the top k results

UNION

HEADLINE GENERATION

1. Append the nuanced context of top k retrieved documents to the original query.
2. This expanded query is then sent as an input to the LLM, which then
3. generates our final output yi , the title of the query passage.

ARCHITECTURE DIAGRAM



03

RESULTS

Results and analysis

Results:

Evaluation Metrics: ROUGE-1 (unigram recall) and ROUGE-L (longest common subsequence) used to assess performance.

Best Performance: Hybrid approach outperforms all models in both ROUGE-1 (0.0985) and ROUGE-L (0.0894).

BM25 Model: Strong performance, second best for ROUGE-1 (0.0989) and ROUGE-L (0.0892).

Dense Search & Style Embedding: Lower scores, indicating reduced effectiveness in headline generation.

Metric	BM25	Style Embedding	Dense Search	Hybrid
Rouge-1	0.0989	0.0826	0.0787	0.0985
Rouge-L	0.0892	0.0787	0.0736	0.0894

04

CONCLUSIONS

Final Takeaways

CONCLUSIONS

- Hybrid model combines strengths of content-dependent and content-independent embeddings.
- Hybrid approach outperforms all models in both ROUGE-1 (0.0985) and ROUGE-L (0.0894).
- Traditional IR methods like BM25, when optimized, still perform well in text generation tasks.

FUTURE WORK

- Refining techniques: sentence embedding to extract relevant portions of the document.
- Alternative LMs: for better task learning

THANK YOU