The Personal Touch: Rethinking News Categories through Author Profiles

Vara Prasad Gudi
vgudi@umass.edu
University of Massachusetts
Amherst, Massachusetts, USA

ABSTRACT

In the contemporary landscape, large language models (LLMs) such as ChatGPT have become indispensable for a multitude of tasks, ranging from answering queries to generating and summarizing content. However, a notable drawback surfaces when users receive uniform responses to identical queries, overlooking individual preferences and diverse backgrounds. This paper aims to overcome this limitation by delving into the realm of personalized large language models. Specifically, our focus lies in advancing the field of Personalized News Categorization, wherein the goal is to assess the capability of a language model to classify news articles written by a user (journalist). By infusing personalization into news categorizations, the research aims to address diversity in information categorization.

ACM Reference Format:
Vara Prasad Gudi. 2023. The Personal Touch: Rethinking News Categories through Author Profiles. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the realm of information overload, the accurate categorization of news articles is paramount for efficient information retrieval. This project addresses this challenge by introducing a unique model that tailors the categorization process to the writing style and interests of individual authors. The primary objective is to develop a model capable of classifying news articles into 15 distinct categories, such as politics and entertainment. Unlike traditional models, this approach embraces personalization, recognizing that an author's distinct style and interests play a crucial role in shaping the content they produce.

The validation and test dataset in a User-based Separation setting comprises articles with corresponding category labels, forming the foundation for a basic categorization model. However, to infuse personalization, the project leverages additional information from the authors' past works, treating them as individual profiles. This involves generating a query, ϕq, from the input article text, which is then used by a retrieval model, R, to identify the top k most relevant profile articles. To construct a personalized prompt, ϕp, the input article is concatenated with the retrieved profile articles. This composite prompt captures the topical context of the current article, as well as the nuanced writing style and specific interests of the author. The resulting prompt is then fed into the categorization model, enabling it to make personalized category predictions.

The effectiveness of personalization is assessed and compared to a non-personalized model using a comprehensive test set. The project encompasses key components, including query generation, retrieval, prompt construction, and training the categorization model. Iterative refinement processes are employed to continuously enhance the accuracy of personalized predictions. The output manifests as a set of JSON-formatted predictions, detailing the model's categorization of articles within the test set. This innovative approach not only improves accuracy but also lays the foundation for a more nuanced understanding of authors' contributions within the dynamic landscape of news categorization.

2 PERSONALIZED NEWS CATEGORIZATION

To make the language model more personalized, a common approach is to include information from the user's profile in the input. However, user profiles are often too long for large language models and can be costly to process. Instead, we suggest a solution where, for each input, we retrieve specific personalized details from the user profile and add them to the language model prompt.

2.1 Retrieval-augmented Method

The image shows a diagram of the retrieval-aligned method for personalizing large language models (LLMs). The method works by first retrieving personalized user profile items for a given input. These items can be anything from the user's name and interests to their past interactions with the LLM. Once the user profile items have been retrieved, they are used to construct a prompt for the LLM. This prompt is then used to generate the desired output. The retrieval-aligned method has several advantages over other methods for personalizing LLMs. First, it is able to personalize the LLM with a limited amount of user-specific information. Second, it is able to generalize to a wide range of tasks. Third, it is able to improve the performance of the LLM on all data and tasks. Here is a more detailed explanation of the different components of the diagram:

- User profile: This represents the user's profile, which can contain information such as the user's name, interests, and past interactions with the LLM.
- Retrieval function: This function retrieves personalized user profile items for a given input.
- Prompt construction function: This function constructs a prompt for the LLM using the retrieved user profile items.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

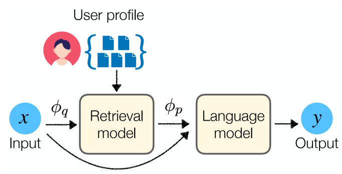

Figure 1: An overview of the retrieval-augmented method for personalizing LLMs. ϕq and ϕp represent query and prompt construction functions.

- Language model: This is the LLM that is being personalized.
- Input: This is the input that is given to the system.
- Output: This is the output that is generated by the system.

The retrieval-aligned method is a powerful technique for personalizing LLMs. It is able to personalize the LLM with a limited amount of user-specific information, generalize to a wide range of tasks, and improve the performance of the LLM on all data and tasks.

2.2 Problem Formulation

The goal is to build a model that can categorize news articles into one of 15 categories (politics, entertainment, etc) based on the writing style and interests of the author. The training data consists of articles and their category labels. This is used to train a basic categorization model. The validation and test data also include past articles written by each author as their "profile". To personalize, a query is generated from the input article text using ϕq. This query retrieves the top k most relevant profile articles using a retrieval model R. A prompt ϕp is constructed by concatenating the input article with the k retrieved profile articles. This personalized prompt better captures the author's interests and style and is fed to the categorization model. The model makes a personalized category prediction based on the prompt. Its accuracy is evaluated on the test set. Personalization is expected to improve over a non-personalized model as the author profile provides useful context. The main components are query generation, retrieval, prompt construction, and training the underlying categorization model. Iterative refinement can improve results. The final output is the model's predicted categories for the test set in JSON format.

3 METHODOLOGY

For the tasks we're working on, each user profile has a lot of information about the user. However, because language models like ours have limits on how much information they can use at once, it's only practical to use a smaller part of this information as input. Also, not all the information in a user's profile may be important for the specific task the user is trying to do. To address this, we suggest coming up with solutions that involve retrieving the most relevant information to enhance the process.

3.1 Data Preparation

In the initial data preparation phase, the raw data of User-based Separation setting is extracted and stored in a comprehensive list. To streamline the prototype model, a subset of the data is selected, and the JSON format is converted into a list structure. Subsequently, distinct lists are created to segregate the articles, queries, profiles, and corresponding categories for efficient organization. Text preprocessing techniques are then applied to refine the data, encompassing tasks such as lowercase conversion, tokenization, normalization, stopping, and Porter stemming. These steps collectively contribute to the transformation of the raw text into relevant tokens, optimizing the data for further analysis. Special attention is given to profiles, which undergo pre-processing to construct a user model. With both the input data and user profile model prepared, the information is seamlessly passed to the Retrieval model, specifically utilizing the BM25 algorithm. This comprehensive approach ensures a well-structured and refined dataset, laying the groundwork for effective retrieval model performance.

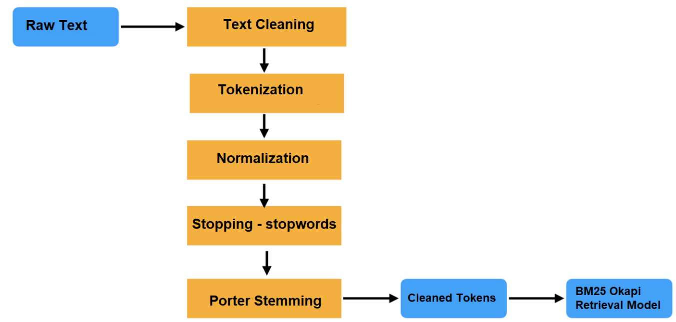

Figure 2: Text Pre-processing and Data Preperation

3.2 Retrieval Model (R)

After completing the Data Preparation stage, the next crucial step is the implementation of the Retrieval Model (R) using the Rank-BM25 algorithm, specifically the BM25Okapi variant. During the Initialization phase, an instance of the BM25 class is created, requiring a list of lists of strings representing document tokens as input. This class reads and indexes the corpus text. Subsequently, the process involves the Ranking of documents, wherein queries are presented, and the algorithm determines the relevance of documents based on their scores. To ensure a fair comparison, the queries are tokenized, and similar preprocessing steps applied to the documents are carried out. Rather than obtaining document scores, an alternative approach is to retrieve the top documents using the bm25 gettopN function. By executing these steps, the retrieval model yields the top documents for each article query. These identified documents, along with the corresponding articles, are then employed to construct a comprehensive prompt for further analysis or downstream tasks, ensuring a refined and efficient retrieval process.

3.3 Prompt Construction

To construct prompts for our Large Language Model (LM), we initiate the process with user articles serving as the primary input. The Per Profile Entry Prompt (PPEP) is formulated by incorporating the article's category, denoted as "Pi[category]," where "Pi[text]" represents the content of the article. The PPEP function iterates over each entry in the retrieved user profile, generating individualized prompts. Subsequently, we employ a prompts template, utilizing functions like "concat" to concatenate strings strategically,

ID	Article	Top Document	Category	Prompt
110	Calls for presidential impeachment have cast a...	The Electoral College will cast its vote for P...	politics	The category for the article: '['The Electoral...
111	With less than 10 weeks to go before the midte...	The Electoral College will cast its vote for P...	politics	The category for the article: '['The Electoral...
112	The U.S. has the most powerful military in the...	President Putin's ultimate ambitions are not k...	politics	The category for the article: '["President Put...

Table 1: Prompt Engineering: Prompt to the LLM

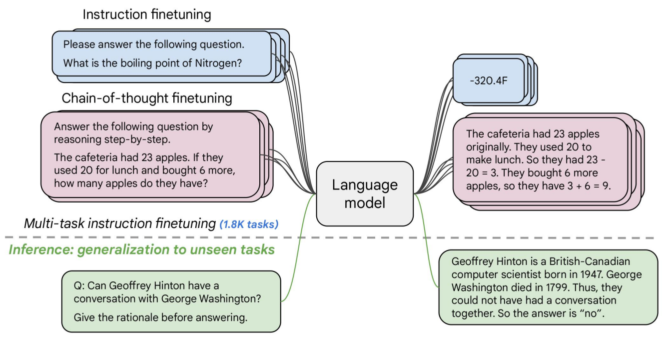

Figure 3: FLAN-T5-large

augmenting the LM input with the user profile details. The Aggregated Input Prompt (AIP) is then crafted by combining the PPEP with the relevant query and the top retrieved documents achieved through BM25. This unified prompt incorporates user-specific information and context from the retrieved documents to enhance the LM's performance. Simultaneously, we meticulously compile a dataframe featuring columns such as article, top relevant document, its corresponding category, and the generated prompt.

This comprehensive approach ensures a well-structured and contextually enriched input for the Large Language Model, optimizing its understanding and performance across diverse tasks.

3.4 Language Model (LLM)

The implementation of Flan-T5-large, a variant of the T5 language model designed for few-shot learning, is outlined. The "large" designation in Flan-T5-large denotes an increased model size, featuring a greater number of parameters for potentially more nuanced text understanding and generation. The project focuses on personalized news categorization, utilizing a sophisticated approach with Flan-T5-large. The curent step commences with meticulous prompt construction and fine-tuning to ensure prompts align precisely with the LLM requirements. These prompts, when paired with relevant articles, constitute a comprehensive inputs fed into the Flan-T5-large model. Leveraging its advanced few-shot learning capabilities, the model efficiently processes this information, categorizing each article based on contextual nuances embedded in the prompts. The resulting categorizations, systematically stored in a list, are anticipated to exhibit nuanced and accurate outcomes, reflecting the model's advanced understanding and processing prowess. This systematic approach streamlines the categorization process, underscoring a high degree of personalization and relevance in the output, showcasing the innovative use of AI in enhancing news categorization.

3.5 Evaluation

During the evaluation phase, a standardized process was implemented to gauge the performance of a Language Model (LLM) on a News Categorization task. This assessment primarily focused on comparing the model's predicted outputs for news articles with corresponding validation gold labels, which served as the ground truth for the task and were provided in JSON format. Concurrently, the LLM's predictions for the same set of articles were stored in a distinct JSON file. The execution of the evaluation script, encapsulated within LaMPEvaluation, involved the utilization of task-specific evaluation logic. This class internally defined custom metrics tailored to the task's nature, particularly designed for text classification scenarios. The evaluation process aimed to measure the model's proficiency in accurately categorizing news articles based on the provided ground truth. It encompassed functions for post-processing text classification and text generation predictions, creating custom evaluation metrics such as F1 and accuracy for classification tasks. Overall, the evaluation comprised the parsing of command-line arguments, loading of gold labels, generation of predictions, post-processing of results, and computation of evaluation metrics for news categorization in a personalized LLM.

4 RESULTS AND ANALYSIS

Based on the results from Table 2, let's analyze the performance of the different models:

The base model (FlanT5-base-finetuned) outperforms both of your models in terms of accuracy and F1 score. It seems to be a more effective model for the news categorization task based on the provided metrics. My Model 2 (FlanT5-xxl-zeroshot + BM25) performs better than Your Model 1 (FlanT5-large-fewshot + BM25) in terms of both accuracy and F1 score.

Model Comparison:
Accuracy: The base model has the highest accuracy, indicating better overall classification performance. However, all models seem to be relatively close in terms of accuracy.
F1 Score: F1 score considers both precision and recall. Your Model 2 has the highest F1 score, indicating better balance between precision and recall. This might be important if there is an uneven distribution of classes.

Model Fine-Tuning and Few-shot Learning:
My Model 1 (FlanT5-large-fewshot + BM25) involves few-shot learning, which might not have performed as well as the base model in this specific task. Consider experimenting with different few-shot learning techniques or hyperparameter tuning to improve performance.

Model Size Impact:
My Model 2 (FlanT5-xxl-zeroshot + BM25) is a larger model (xxl) compared to the base model. It shows improvement in F1 score, suggesting that a larger model might capture more complex patterns in the data.

Understanding:
Enhancing the accuracy and F1 score can be achieved through

Type	Model Name	Accuracy	F1 Score
Base Model	FlanT5-base-finetuned+ BM25	0.734	0.6132
My Model	FlanT5-large-fewshot+ BM25	0.6512	0.4708
My Model	FlanT5-xxl-zeroshot+ BM25	0.699	0.6360

Table 2: Performance of Multiple Models

the fine-tuning of a model's hyperparameters. Fine-tuning allows customization of the model to a specific task or domain, leading to superior accuracy and efficiency compared to utilizing a pre-trained model alone.

4.1 Comparison with LAMP Base Model

In comparing My Model 1 with the Base Model in this model, we observed that despite using a more advanced language model (LLM) in My Model 1 (FlanT5-large-fewshot+) without fine tuning, the accuracy and F1 score were slightly lower than the Base Model (FlanT5-base-finetuned+ BM25). This suggests that, in this context, fine-tuning the Base Model with additional data might have played a more crucial role in achieving better performance than using a larger language model alone. While My Model 1 with FlanT5-large showed promise, it indicates that the balance between a sophisticated LLM and effective fine-tuning strategies is crucial for optimal results in personalized news categorization.

5 RESEARCH WORK

The research problem at the core of this study revolves around refining and optimizing two critical components in the news categorization process: query generation and prompt engineering. Specifically, the investigation aims to discern the impact of these enhancements on the relevance of retrieved profile documents and the subsequent categorization outcomes.

5.1 Improved Query Generation and Relevance of Retrieved Profile Documents

The initial aspect of the research problem focuses on the efficacy of improved query generation in the context of news categorization. The pre-processing techniques applied to the article inputs, including Lowercase, Tokenization, Normalization, Stopping, and Stemming, set the stage for a comprehensive exploration. Leveraging advanced natural language processing techniques such as SpaCy and NLTK, the user profiles are also prepared for input. The subsequent application of the BM25 Retrieval model facilitates the retrieval of top documents for each query, thereby forming the basis for profile document categorization. The research seeks to delve into how these refined queries, coupled with meticulous pre-processing, and delves into the impact of appropriately generating queries on the improvement of relevance, precision and other metrics in the retrieved profile documents.

5.2 Writing Style Influence on Categorization Outcome through Prompt Engineering

The second dimension of the research problem involves the influence of writing style on the categorization outcome, achieved through prompt engineering. The retrieved top documents are strategically amalgamated with the actual input articles to create a Per Profile Entry Prompt (PPEP) and a Prompt Function - Aggregated Input Prompt (AIP). This novel approach integrates user-specific characteristics into the prompt design. The subsequent application of these tailored prompts to a Language Model (LLM) for news categorization introduces a user-centric dimension to the categorization process. It seeks to understand how tailoring prompts to an author's writing style can enhance the precision and relevance of sorting news articles

5.3 RELATED WORK - FUTURE SCOPE

Research on personalization in information access, driven by initiatives like the Netflix Challenge, spans academic and industrial realms, with a focus on recommender systems, search applications, and dialogue agents in NLP. Dialogue agent research involves constructing realistic conversational data, using crowd-workers, Reddit, and Weibo, and annotating movie dialogue datasets with narrative character personas. Personalization extends to diverse domains, including leveraging reviews and recipes for personalized generation tasks, exploring parameter-efficient models for personalized translation, and creating a personalized headline generation dataset from Microsoft News. The research also encompasses fundamental language modeling problems, utilizing user data from platforms like Reddit, Facebook, and Twitter. Applications of personalized language models to tasks such as stance classification and demographic inference are explored. Additionally, the connection between personalization and accommodating human label variation in NLP tasks is highlighted, as seen in studies on modeling annotators in classification tasks reliant on social norms. The collective research underscores the significance of personalization in tailoring information access and language understanding systems to individual users across various NLP domains.

6 CONCLUSION

In conclusion, this research project contributes to the advancement of language model benchmarking with a focus on personalized news articles classification. The introduced LaMP benchmark undergoes thorough evaluation, employing a range of language models and retrieval techniques to establish baseline performance. Through extensive experimentation, we delve into the selection of user profile entries for generating personalized prompts. The final section of this report highlights the performance of leading large-scale language models on the LaMP benchmark, providing a comprehensive comparison with the base paper results. The ensuing analysis and insights derived from this study contribute valuable perspectives to the evolving landscape of language model research and personalized content classification.

REFERENCES

[1] Salemi, A., Mysore, S., Bendersky, M., Zamani, H. (2023). LaMP: When Large Language Models Meet Personalization. University of Massachusetts Amherst.

[2] Ao, X., Wang, X., Luo, L., Qiao, Y., He, Q., Xie, X. (2021). PENS: A dataset and generic framework for personalized news headline

generation. In Proceedings of ACL-IJCNLP 2021 (pp. 82–92).

[3] Song, H., Wang, Y., Zhang, K., Zhang, W.-N., Liu, T. (2021). BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data. arXiv:2106.06169 [cs.CL].

[4] Dudy, S., Bedrick, S., Webber, B. (2021). Refocusing on Relevance: Personalization in NLG. In Proceedings of EMNLP 2021 (pp. 5190–5202).

[5] Croft, B. W., Cronen-Townsend, S., Lavrenko, V. (2001). Relevance feedback and personalization: A language modeling perspective. In DELOS Workshop.

[6] Das, A. S., Datar, M., Garg, A., Rajaram, S. (2007). Google news personalization: Scalable online collaborative filtering. In Proceedings of WWW '07 (pp. 271–280). ACM