Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-7, Issue-3, March 2025
DOI: 10.53469/jrse.2025.07(03).06

Personalized Information Generation: A Review

Tony Fredrick

Somany Institute of Technology and management, M. D. U Rohtak
fredrick0708@yahoo.com

Abstract: ***There are countless options for choosing news headlines, and finding the right balance between conveying an important message and capturing the reader's attention is the key to successful publishing. However, it is unfair to present the same information about the same topic to all readers; because no matter what the preferences and interests of different readers are, there may be confusion as to why a particular topic is presented to them and a good match may not be found between interests and requested topic. In this article, we present a new approach that addresses these problems by combining user profiling to provide personalized headlines and automated and human review methods to determine what you use for a particular headline. Our system uses a powerful key function to assign unique keywords to users based on their reading history, which is then used to transform generation titles. Through an in - depth analysis, we demonstrate the effectiveness of the proposed framework in delivering personalized headlines that suit the needs of the target audience. Our platform has the ability to improve the performance of data requests and facilitate the creation of personalized content.***

Keywords: news headlines, attention, publishing, Information

1. Introduction

Personalized Information systems such as Google News and Yahoo News help users discover topics relevant to their interests (Karimi et al., 2018). However, the system often displays the same title to all users, making it difficult for them to understand the relationship between their interests and the recommended content, which can reduce the effectiveness of the recommendation system. To solve this problem, we propose a new framework for generating unique, engaging headlines () that clearly show the relationship between the user's reading history and a particular article. Our platform has the ability to improve the performance of personalized information requests as well as recommendations for short videos, articles, recipes and more. Creating unique headlines is a difficult task due to the limitations of clarity and the need to capture readers' attention. A given headline should (a) clearly convey the main message of the article and (b) provide a clear line in the user's reading history, using only 10 words on average. There are two important issues in this process. First, a headline that encourages users to click but provides little information and doesn't convey an important story becomes clickbait rather than a useful headline. Second, it is difficult to find data sources, large amounts of data containing many specific topics and user profiles. Such a data set would be useful for the development of unique titles but is not currently available. The key to better branding is to develop a comprehensive approach that allows us to (a) understand users' interests based on their reading history, (b) create personalized headlines, and (c) evaluate the performance of these headlines based on user experience. . . Previous research on topic creation has focused primarily on creating headlines that summarize information or first sentences, but did not consider the potential benefits of personalization. In this work, we propose a network that includes user profiles2 and a comprehensive set of automated and human evaluation methods that users can choose to create personalized headlines that connect with different audiences.

Our approach focuses on learning useful features that combine the user's reading history with a collection of signed sentences. This type of user profile is efficient and flexible as the signature phrase can be easily updated as the user's interests change (Bansal et al., 2015). These signed sentences are derived from news stories based on the user's reading history through cross - learning, without the need for published data. For example, if the phrase Upper East Side appears frequently in a user's reading history, it may become a signature phrase for the user. These signature sentences do not need to appear verbatim in the user's reading history and can generate significant interest. For example, if the user's reading history includes the phrases Avengers and Hulk, this can show love for Marvel movies and be a signature for Marvel Studios. a sentence expressing this advantage. We create a database dataset trains the model to generate personalized headlines for a news article. Using signature phrases, our model is able to create a connection between the recommended article and the user’s interests, resulting in personalized headlines that are both engaging and anchored to the article to avoid clickbait. Evaluating personalized news headlines presents unique challenges. It would be ideal to have human evaluators judge the effectiveness of system headlines. Indeed, we have conducted a human evaluation in this study. However, this process is time - consuming and costly, making it impractical during the system development phase. Thus, we propose a comprehensive synthesis of automated and human evaluation methods to assess headline relevance and user preference. By using the signature phrases, we synthesize user profiles of various types.

We hypothesize that headlines tailored to these user profiles will be more popular among some users than general, impersonal headlines based on a popularity index. We also test various automated metrics to assess headline quality in terms of information delivery, relevance to the article at hand, and content

In this article we make the following contributions:

- present a comprehensive framework for delivering personalized news articles that convey the content of the article. The main message of the article attracts the reader's attention while also matching his interests. Our platform uses deep learning algorithms to extract signature phrases from users' reading history and use them to customize
- titles;

- We design automatic and human evaluation methods to evaluate the performance of titles based on their accuracy and preference.

We also compare the proposed level with the main generation head, show the results in the database, and determine the best direction for future research through in - depth analysis of the system results. Approaches Our aim is to create a user - friendly title that gives the gist of the information provided for a particular user. To achieve this, we developed three steps:

1) Signature language. Using the keyword generation module, we define the signature sequence Zd = {z1, z2. . . } contains various elements of d;
2) User signature selection expression. From the set of candidate signatures, we select the subset Zud ⊆ Zd corresponding to the interest of user u as the user's signature expression (section 3.2);
3) Signature – refers to the title. Based on the information d and the selected sentence of user signature Zu d, we generate the title content of article d according to the interests of user u.

Signature Phrases Identification

We ppproach this task as a script query where the template takes a data entry or header as input and outputs all signature phrases in the entry, separated by semicolons. We use the BART model estimated from the KPTimes3 dataset. KPTimes is a massive dataset containing 279, 000 pieces of information associated with signed sentences. Unlike other feature news sites that focus on scientific research articles, KPTimes focuses on extracting feature sentences from news stories, making it a perfect fit for our mission. The model is trained by minimizing the cross - entropy loss between predicted signature patterns and human signature assignments.

User Signature Selection

In this step, we track all sentences signed by candidates in Zd based on their engagement level and user u's reading history, and select the top candidates as the user's signature. Suppose that the user's history Hu can be defined as a list of titles of articles that the user has previously read, i. e. Hu = {t1, t2,. . . . We first transform each signed term zi ∈ Zd into a vector zi using the signature expression. To calculate the usability score of each candidate signature phrase, we consider two different user history matching strategies:

(1) Encryption code. We concatenate all titles in user Hu's reading history with additional semicolons to separate the titles. We then place the concatenated titles into the full title of the hu using historical encoding.

The inflection points of the signature expression zi ∈ Zd for u operators S (zi, Hu) are obtained by the dot product of the two parts:

$$S\ (zi, Hu) = z\text{T}i\ hu \qquad (1)$$

(2) Code for contacts. Each head tj ∈ Hu is encoded as a complete vector of tj using head coding. The user engagement score is then defined as the minimum dot product correlation between signed expressions zi and each head in the reading history: S (zi, Hu) = max

In practice, we train the user signature phrase selection model using an in - batch contrastive learning approach. We consider a batch of synthesized users {u1, u2, ・ ・ ・, uNB} where NB is the batch size, and each user ui has exactly one user signature phrase zi. The reading history Hi for user ui is then constructed by randomly sampling news articles whose candidate signature phrases contain zi, i. e., Hi = {d | zi ∈ Zd}. In this way, (zi, Hi) is considered as a positive pair, and (zi, Hj) (i≠ j) is considered as a negative pair.

Signature - Oriented Headline Generation

We design a user - specific title function as the production function. Given a data point d and a user u and user signature Zu
d ⊆ Zd, our goal is

t = [w1, w2,. . .] for the value d, where w i is the ith symbol in t.

In particular, the input of the generator is the combination of the user's signature Zu d with the information data d, and the output is the signature - based topic t. During training, Zu d is recognized from t, which is the real head of d. In the decision process, Zu d is defined from d itself and is chosen by the users of the selected signature, since it does not exist before the title t is created. Here we use BART as a generator for the production header.

Processing

In this section, we describe the process of corpus processing, including generating user synthesis and generating topic - based speech signatures. Our data is drawn from two existing databases Corpus Newsroom Gigaword Synthesized User Dataset Train

For each corpus, the synthesized user dataset is used to train the sentence selection module and evaluate the entire system, while the topic dataset is used to train the topic. build steps (no tests are set as the build steps are evaluated system - wide using the user test dataset). corpora: Wordroom (Grusky) and Gigaword (Rush et al.; Graff et al.,). The Newsroom corpus contains 995, 041 articles from both articles, in training, 108, 837 in validation, and 108, 862 in testing. The Gigaword corpus contains 7, 704, 419 examples in training, 394, 390 in validation, and 381, 045 in testing. For each corpus, we create two datasets: the user dataset and the main data generation. The first dataset is used to train the use of signatures for speech selection (section 3.2) and evaluate the entire system, while the second dataset is used to train signature - focused keywords (section 3.3).

Synthesized User Creation

As since real user data is not available, we create synthetic users that simulate real reading history. The synthesis process consists of the following steps:

(1) Identification of signed sentences in the entire dataset to generate a candidate test;

(2) Map each signed sentence to the data pipeline containing that sentence; . . provision of diverse learning (Article 4). But when evaluating the model, each user synthesizes 1 to 5

interesting sentences to simulate the real situation. It's important to note that it's easy to create custom topics for users with a simple reading history (for example, users whose reading history includes only one or two topics). To examine the impact of the number of interested users on the generated topic, we replicate 2000 engaged users with 1 ~ 5 interesting figures. In general, a unique title is only useful if the source material matches the user's interests. To determine effectiveness, for each user we choose to select one of the user's signature phrases and then select a news story that contains the selected phrase as the entry of the search query. This ensures that the information the title should create is relevant to the user. Details are explained in Chapter 5. One generation. To create a characteristic sentence for a topic, we use the characteristic sentence model to extract the characteristic sentence from the original topic. These generated sentences, along with relevant information data, are then entered into up to sample headlines to create original headlines. In our study, we reduced the number of all news stories to a maximum of 512 and only retained signed sentences that appeared in more than 10 news stories.

2. Experiments

We evaluate our proposed system from different perspectives, including empirical evaluation for specific generations (section 5.2), empirical evaluation (section 5.3), and elimination studies (section 5.4).5.1 Baseline methods

We compare the performance of our system with the following baseline methods:

PENSEBNR and PENS - NRMS (Ao et al., 2021) are model head - based LSTM methods. Both were trained on the PENS dataset but used different reading histories; The vanilla system is a BART - wide model that efficiently manages data generation without using signed sentences; Vanilla Man refers to the news writer's original title; The SP header uses the signature phrase that appears in the first human - written thread accompanying thread; SP - random selects a characteristic phrase from the dataset to guide the generation of the title. SP - vol.

Selectors Evaluation.
To evaluate signature selection performance, we track all sentences signed by candidates in the synthesis user's conversation and provide the following parameters: (1) Hit[at]K is the percentage of appearances at the highest K level where the correct signature sentence is found.; . We use a synthetic user evaluation dataset to evaluate both creation and signature selection.

Factors Affecting Headline Generation.
From our research, we found that the following factors affect the quality of the given topics: (1) The number of topics the user is interested in. As seen in Table 55, the results of evaluating the data points of connected users according to their categories. The fact that the number of interest points is different shows that as the number of interest points increases, the user's compatibility score decreases, while other scores remain the same. This shows that it is easy to create specific headlines so that users can read relevant information in a few sentences. However, even if the number of suggestions increases to 30, our proposed method still achieves a better climate change score than the vanilla system, while the performance and confidence scores are also comparable. (2) User signature number. Analysis of submitted headers showed that submitted headers could have contained event errors if the signature - referenced header generator had consistently taken the signature user as input. This is because the generator is forced to include unsigned sentences in the common header, as shown in the first example in Table 3. As a result, we only use one signature phrase to guide the generation header.

3. Conclusion

We exploring the creation of specific topics related to different user interests. We propose a basic approach to topic creation and data collection to support our classroom's education without the need for human - sourced data. We are also exploring evaluation methods that enable automatic evaluation of titles submitted from multiple sources. Our research shows the effectiveness of our proposed method.

Ethical Considerations

It is important to use the unique data generation technology concept efficiently and responsibly. Although the technique aims to improve user experience and experience, it can also be used to create headlines that may be of interest to an individual reader, which can lead to biased information. In this document, we have taken the necessary steps to protect your personal data. Our technique is based on the user's reading history, represented as a set of recently viewed content. Due to privacy concerns, no demographic information such as age, gender, or location is used or collected. We encourage the public to continue researching the impacts and consequences of this technology.

References

[1] Goyal, Tanya, Li, Junyi Jessy, and Durrett, Greg. News summarization and evaluation in the era of gpt - 3.
[2] Firoozeh, Nazanin, Nazarenko, Adeline, Alizon, Fabrice, and Daille, B´eatrice. Keyword extraction.
[3] Das, Debashis, Sahoo, Laxman, and Datta, Sujoy. A survey on recommendation system.
[4] Moghe, Nikita, Arora, Siddhartha, Banerjee, Suman, and Khapra, Mitesh M. Towards exploiting background knowledge for building conversation systems.
[5] Cai, Pengshan, " Generative Models for Personalized Information".
[6] Krapivin, Mikalai, Autaeu, Aliaksandr, and Marchese, Maurizio. Large dataset for key phrases extraction.
[7] Ao, Xiang, Wang, Xiting, Luo, Ling, Qiao, Ying, He, Qing, and Xie, Xing. PENS: A dataset and generic framework for personalized news headline generation.
[8] Beliga, Slobodan, Mestrovic, Ana, and Sanda. An overview of graph - based keyword extraction methods and approaches.
[9] Bernstein, Abraham, Vreese, Claes De, Helberger, Natali, Schulz, Wolfgang, and Zweig, Katharina A. Diversity, fairness, and data - driven personalization in (news) recommender system.

[10] Flek, Lucie. Returning the N to NLP: Towards contextually personalized classification models.
[11] Karimi, Mozhgan, Jannach, Dietmar, and Jugovac, Michael. News recommender systems – survey and roads ahead.
[12] Lin, Chin - Yew. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out.*
[13] Speretta, Mirco, and Gauch, Susan. Personalized search based on user search histories.
[14] Tanya Goyal, Junyi Jessy Li, Greg Durrett. News summarization and evaluation in the era of gpt - 3.
[15] Wu, Fangzhao, Qiao, Ying, Chen, Jiun - Hung, Wu, Chuhan, Qi, Tao, Lian, Jianxun, Liu, Danyang, Xie, Xing, Gao, Jianfeng, Wu, Winnie, and Zhou, Ming. MIND: A large - scale dataset for news recommendation.