Preface to INRA 2022

Özlem Özgöbek[1], Andreas Lommatzsch[2], Benjamin Kille[1], Peng Liu[1],
Edward C. Malthouse[3] and Jon Atle Gulla[1]

[1]*Norwegian University of Science and Technology, 7491 Trondheim, Norway*
[2]*Berlin Institute of Technology, 10623 Berlin, Germany*
[3]*Northwestern University, USA*

Introduction

The 10th International Workshop on News Recommendation and Analytics (INRA 2022)[1] was held in conjunction with the 45th International ACM Conference on Research and Development in Information Retrieval (SIGIR 2021), July 11–15, 2022 in Madrid, Spain in a hybrid setting. The main goal for the workshop was to facilitate the exchange of ideas and networking among researchers, media companies, and practitioners within the news recommendation and analytics domain.

News consumption has shifted online and increasingly toward social media. Intelligent systems have entered the picture to support users in finding relevant content. These news recommender systems alleviate the problem of information overload via personalizing information access. At the same time, societies face new problems. More and more people doubt the credibility of news—a phenomenon referred to as 'fake news.' News platforms can create filter bubbles, where users only encounter information and opinions that reinforce and conform to their own beliefs. Businesses have emerged for collecting, analyzing, and marketing data concerning readers' interactions with information sources. Some news outlets blend information and advertising in 'native advertising.' Social media's influence on readers' information consumption has reached levels that let governments call into question their status as platform providers.

The previous nine editions of the International Workshop on News Recommendation and Analytics (INRA) have served as a forum to discuss research, technological advancements, and societal impact of artificial intelligence in the news ecosystem. The importance of artificial intelligence and machine learning in the news domain keeps increasing. Data journalism

[1]https://research.idi.ntnu.no/NewsTech/INRA

has become more relevant. In the Covid-19 pandemic, many news outlets started to integrate interactive infographics on their websites. Simultaneously, we observe that multimedia becomes better accessible for AI technology through the advancements in deep learning. The pre-trained language models, e.g., BERT and GPT, which achieve remarkable performances on various NLP tasks, have shown their great potential to enhance news text modeling for recommendations. The increasing focus on the news domain let us to believe that a forum for in-depth discussion about new technologies and trends in news is more relevant than ever.

About the Workshop

The workshop focused on three main themes: news personalization, news analytics, and the psychological, societal, and ethical aspects of news personalization systems. The following topics of interest included in the call for papers for this workshop:

- **News Personalization**
  - Context-aware news recommender systems,
  - News recommendation in social media,
  - Multi-modal news recommendation,
  - User behavior analysis and user interest modeling in the news domain,
  - User modeling and user profiling,
  - Applications of data mining for personalized search and navigation,
  - Personalized news user interface and visualization,
  - Diversity and multi-perspectivity in news personalization and recommendation,

- **News Analytics**
  - News semantics and ontologies,
  - Adaptive and personalized news summarization, categorization, and opinion mining,
  - Social Graph and heterogeneous network analysis ,
  - User segmentation and community discovery,
  - Big data technologies for news streams,
  - News framing research,
  - Automated news generation ,
  - News political leaning and tone,
  - News trends and evolution,

- **Psychological, Societal, and Ethical Aspects of News Personalization Systems**
  - Privacy and security issues,
  - Clickbait, fake news, and misinformation detection,
  - Diversity and fairness of news search/recommendation,
  - Bias in online news,
  - Transparency and explainability, and
  - Emotion and cognition in news reception.

INRA 2022 received four submissions. Each paper was reviewed by at least two reviewers in a single-blind reviewing process. Thereof, three papers were accepted for presentations. Abstracts of these papers can be found in the Accepted Papers Section below. The workshop was organized in a hybrid setting where all sessions were available digitally.

In addition to the accepted papers, the workshop included a keynote speech, two invited talks and a panel discussion. The keynote speech was given by Dr. Yong Zheng from Illinois Institute of Technology, USA. Dr. Zheng's keynote speech, titled "Multi-Objective Optimization (MOO) and News Recommendations," included an introduction of multi-objective optimization, revisited different applications and discussed the existing and novel applications of MOO in news recommender systems.

Two invited talks were given by Dr. Andreas Lommatzsch, who is on the faculty of Elektrotechnik und Informatik at the Technische Universität Berlin, and Srivas Prasad who is the Question Understanding team leader at Bloomberg LP. Dr. Lommatzsch's talk titled "Understanding the role of images in news" discussed the of assigning images to a news article text. Images in news typically should catch the reader's attention and highlight an interesting aspect of the news. Since for most recent news events no photos are available, images in news are often taken from stock photo databases or old photos only indirectly related to the news image. Since the images included in a news article has a strong influence on the perception of news, it is important to understand the relation between news texts and images. Relevant use cases are for instance: Finding images for news texts not having an image yet, the personalization of news or the detection of fake news. In the talk Dr. Lommatzsch highlighted the challenges in detail, explained the strengths and weaknesses of existing methods, and introduced the MediaEval NewsImage challenge providing a platform for implementing and evaluating new methods.

Srivas Prasad shared his experiences from Bloomberg News in this talk titled "AI and Financial News at Bloomberg." He discussed the problem of detecting relevant information in social media and delivering reliable, valuable information in real time.

The panel discussion, titled "Future of News Ecosystems," stimulated interesting discussion about what the future technological developments may bring in news rooms and to the readers. The panel discussion was moderated by Özlem Özgöbek from Norwegian University of Science and Technology and had four participants both from academia and industry. The panel participants were Yong Zheng from Illinois Tech, Fangzhao Wu from Microsoft Research Asia, Rich Hanes from Bloomberg and Preslav Nakov from Hamad Bin Khalifa University. The workshop closed with an open discussion session.

Accepted Papers

INRA 2022 has accepted three papers after the single-blind reviewing process by at least two reviewers. Here, we list the papers, authors, and the abstract.

Relevancy and Diversity in News Recommendations

Shaina Raza, Chen Ding There are some unique challenges associated with news recommendation systems, such as changing user preferences over dynamically generated news articles.

Diversity is required for a news recommender system to expose users to a wide range of information. We propose a deep neural network with a two-tower architecture that learns news item representations through a news item tower and user representations via a query tower. In our approach, we present an enhanced vector for each query and news item in order to introduce information interaction between these entities. We also offer a negative sampling approach to tackle the selection bias of implicit user feedback. Further, we introduce a category loss function that aligns the uneven news category representations of items to introduce diversity in the news recommendations being produced. Experimental results on two benchmark datasets demonstrate that our proposed architecture performs better than the baselines and achieves both relevancy and diversity.

Understanding the Relation of User and News Representations in Content-Based Neural News Recommendation

Lucas Möller, Sebastian Padó A number of models for neural content-based news recommendation have been proposed. However, there is limited understanding of the relative importance of the three main components of such systems (news encoder, user encoder, and scoring function) and the trade-offs involved. In this paper, we assess the hypothesis that the most widely used means of matching user and candidate news representations is not expressive enough. We allow our system to model more complex relations between the two by assessing more expressive scoring functions. Across a wide range of baseline and established systems this results in consistent improvements of around 6 points in AUC. Our results also indicate a trade-off between the complexity of news encoder and scoring function: A fairly simple baseline model scores well above 68% AUC on the MIND dataset and comes within 2 points of the published state-of-the-art, while requiring a fraction of the computational costs.

A Benchmark for Text Classification in News Recommendations

Xinyi Li, Edward C. Malthouse Text classification is an important task in natural language processing. In the current era, people mainly obtain information from online news resources. It is then important to have an automatic and accurate news classifier to categorize every day's news stories such that readers can find articles of interested more easily. We use news story data from the McClatchy organization to establish benchmarks on how accurately stories can be classified by multiple existing deep learning classifiers. Among the models we evaluated, Bidirectional Encoder Representations from Transformers (BERT) provides the best accuracy, macro-averaging precision, micro-averaging precision, macro-averaging recall and micro-averaging recall. Different from many other news benchmark data set, McClatchy provides both headline and full-text for each news story. We compare the performance of every deep learning-based classifier using headlines versus full-texts—the top three predicted categories include the labeled value 95% of the time with full-texts training and 92% with headlines only. Furthermore, the defined topics in McClatchy are not mutually exclusive. Some predictions identified as inaccurate are in fact classified into reasonable topics. We further provide a visualization of stories from various defined topics. The predicted results and the visualization of news stories illustrate the untrustworthiness of labeled classes and the intrinsic

difficulty of categorizing news stories.

Organizing Committee

- Özlem Özgöbek, Norwegian University of Science and Technology, Norway
- Andreas Lommatzsch, Berlin Institute of Technology, Germany
- Benjamin Kille, Norwegian University of Science and Technology, Norway
- Peng Liu, Norwegian University of Science and Technology, Norway
- Edward C. Malthouse, Northwestern University, USA
- Jon Atle Gulla, Norwegian University of Science and Technology, Norway