

News Images in MediaEval 2021

Benjamin Kille¹, Andreas Lommatzsch², Özlem Özgöbek¹, Mehdi Elahi³, Duc-Tien Dang-Nguyen^{3,4}

¹Norwegian University of Science and Technology, Norway

²Berlin Institute of Technology, Germany

³University of Bergen, Norway

⁴Kristiania University College, Norway

benjamin.u.kille@ntnu.no, andreas.lommatzsch@tu-berlin.de, ozlem.ozgobek@ntnu.no,

mehdi.elahi@uib.no, ductien.dangnguyen@uib.no

ABSTRACT

Most news outlets offer a multi-modal user experience. Besides texts, readers encounter images, audio, video, and interactive elements. News Images strives to understand better how images affect news consumption. Participants gain access to a large scale data set of news articles and images. The task consists of two subtasks. Participants can engage in both or one of them. In the first subtask, participants must predict which images publishers paired given news articles with. In the second subtask, participants must estimate the chance that users will pay attention to pairs of articles and images. This paper describes the settings in detail and draws connections to existing research.

1 INTRODUCTION

The news landscape features a multimodal mix of content. Frequently, images accompany the text to draw attention. Research concerning multimedia and recommender systems usually assumes a simple relationship between images and text. For instance, research on image captioning [5] assumes the caption to quite literally describe the image's scenery. However, research on the connection between text and images of news articles indicates a more complicated relationship [8]. This is where the News Images task in MediaEval 2021 comes into play. The task investigates this relationship to understand its implications for journalism and news personalisation.

The task comprises two subtasks, both of which participants can tackle using text-based or image-based features. For the first task, the link between a set of articles and images has been removed. Participants must re-establish which images the publisher had assigned to articles. For the second task, participants must estimate the chance that users paid attention to articles. Thereby, we are trying to understand if images increase the users' attention to news articles. Ultimately, we seek to gain further insight about the relationship of text and images and the reactions of users to recommendations consisting of headlines and images. Particularly, the task aims to surpass conventional work in the area of image concept detection. In other words, we hope to capture aspects of images that exceed the literally depicted content such as quality, style, and framing.

2 BACKGROUND AND RELATED WORK

The *Multimedia Evaluation Benchmark* (MediaEval) examines the intersection of multimodality and recommendation for the fourth time in 2021. In 2018, the *NewsREEL Multimedia*¹ task offered data from several publishers. In 2019, a subtask of the *MultimediaRecsys*² features similar data. In 2020, the *NewsImages*³ task provided data covering three months of news. In 2021, we have extended the coverage of previously released data set by another month.

News outlets have introduced personalisation in the form of recommender systems [6]. News recommender systems help users finding relevant content. Still, most recommender systems rely on information extracted from ratings and texts; the consideration of image data could help to further improve recommendations. On the other hand, personalisation has introduced some issues. The emergence of 'fake news' has raised some red flags [13]. 'Fake news' often put data or images in an misleading context. Thus a fine-grained analysis of images and text helps to get a better understanding of this phenomenon.

Research has recognised the importance of topics related to news personalisation. As a result, more and more venues for discussion have been created [10]. Besides, recent years have seen an upward trend concerning research on multimodal recommender systems. For instance, Truong and Lauw [12] investigate how to leverage multimodal user feedback, Salah et al. [11] prepare a framework for multimodal recommender systems and Oramas et al. [9] examine the use of multimodal data for music recommendation. For a comprehensive review, we refer to [2].

News recommender system research strives to learn about the positive and negative effects of personalisation. The News Images task supports the research toward multimodality. We want to learn more about how images affect news readers' experience.

3 TASK DESCRIPTION

News Images explores the interplay of text and imagery for news consumption. The task defines subtasks. Participants can take part in either or both of the subtasks.

3.1 Task 1: Image-Text Re-Matching

Publishers manually equip news articles with images. The content curators will take images from the event if available. Otherwise, they can use images from designated databases such as stock images. Consequently, we encounter pairs of articles and images. For this

Copyright 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

MediaEval'21, December 13-15 2021, Online

¹<https://www.multimediaeval.org/mediaeval2018/newsreelmm/>

²<https://www.multimediaeval.org/mediaeval2019/mmrecsys/>

³<https://multimediaeval.github.io/editions/2020/tasks/newsimages/>

subtask, we have removed the links between images and articles. The subtask asks participants to re-establish the correct links.

3.2 Task 2: News Click Prediction

News outlets monitor users' engagement with their channels. Their webservers record interactions between readers and news articles. We expect images to play a role in readers' complicated decision making on whether or not to read an article. For this task, we have concealed the interaction statistics of the test set. Participants must predict the articles which attracted most user engagement.

Both subtasks focus on news consumption. We assess submissions both in terms of quantitative performance and qualitative value. More qualitative aspects concern the increased understanding of images' effect on news consumption.

4 DATASET

The task's underlying data is derived from four months worth of webserver log files of a German news publisher. The data contains information related to articles, images, and interactions with users. Each article and images has a reference number assigned. Articles' metadata includes the URL, title, and a text snippet of at most 256 characters. Participants have to download the images as we lack the necessary copyright to distribute them. Users have engaged with articles in three ways: accessing the article, seeing recommendations, and clicking on the recommendations. While the system delivers the recommendations, users explicitly choose to read articles and click on recommendations.

The data set comprises five batches. The first three batches constitute the training data for both subtasks. The training data contains the links between articles and images as well as the interaction statistics. The fourth batch contains the test data for subtask Image-Text Re-Matching. Therein, the link between articles and images has been removed. The fifth batch contains the evaluation data for subtask News Click Prediction. Therein, the interaction statistics have been removed.

Table 1 illustrates the data set. The data have been split chronologically to guarantee meaningful results. All batches contain between 1900 and 2700 articles and images. Downloading a batch can take around 45 minutes with a standard broadband internet connection.

Table 1: Data Set Statistics. The data set is split into five batches each corresponding to one month of webserver logs. The number of cases refers to both articles and images.

Feature	Period	Purpose	No. Cases
Batch 1	Jan 2019	Training	2539
Batch 2	Feb 2019	Training	2604
Batch 3	Mar 2019	Training	2387
Batch 4	Apr 2019	Evaluation (Re-matching)	1915
Batch 5	May 2019	Evaluation (Prediction)	2223

5 EVALUATION

Participants must re-match articles and images and predict how much attention articles attract. For each subtask, participants can submit up to five runs.

5.1 Task 1: Image Task Re-Matching

The link between articles and images has been removed for the evaluation set. Participants deliver a ranked list of at most 100 candidate images for each article. The evaluation set contains 1915 articles and images. Participants must provide a file with 101 columns separated by a tab character. The first column contains the article reference. The second column contains the most likely matched image. The third column contains the second best match and so on. For each article, we compute the precision at a set of cut-off points. For instance, we can check whether the first items contain the actually linked image. Averaging over all articles, we obtain the quantitative evaluation metric: precision at rank N for $N \in \{1, 5, 10, 20, 50, 100\}$.

5.2 Task 2: News Click Prediction

The training partitions reveal how frequently articles have been read by users. In the evaluation partition, the reading statistics remain hidden. Participants must estimate the statistics. We are less interested in precise estimates. Instead, we want to know whether the estimator can discern the most interesting articles. Consequently, participants provide a file with article references populating the first column. The second column shows the estimated reading statistics. The estimates have to be numeric such that we can define a ranking. The evaluation takes the ranking and computes the precision at N for $N \in \{1, 5, 10, 20, 50, 100\}$. The one hundred articles with most accesses constitute the ground truth.

5.3 Run Description

Participants inform about their ideas and discuss the evaluation results in working notes. The working notes highlight their reasoning, qualitative findings, and critical reflections about what can be deduced from the quantitative results. Participants may compare the results from different runs and analyze the finding with respect to result quality, computational complexity, and the used resources. The discussion of the results should take into account the specific properties of the dataset and explain how the finding can be used in related scenarios. Ultimately, the participants ought to describe what they have learned and how their insights can help to move the research forward.

6 CONCLUSION

How users interact with digital content on a larger scale remains hard to understand. Multimodal news presentation becomes more important. Publishers aim to retain their readership by informing and entertaining. Consequently, understanding users' preferences toward content can give them competitive advantages. News Images wants to shed light on the relation between images and text in the news domain. Better understanding this relationship can help to prevent harm to the news eco-system, for instance in the form of increased polarisation and diminished trust in information.

ACKNOWLEDGMENTS

We would like to thank plista for kindly providing the real world data. Further, we thank Martha Larson for her support.

REFERENCES

- [1] Francesco Corsini and Martha Larson. 2016. CLEF NewsREEL 2016: Image based Recommendation. In *Working Notes of the 7th International Conference of the CLEF Initiative, Evora, Portugal*. CEUR Workshop Proceedings.
- [2] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–38.
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 248–255.
- [4] Mouzhi Ge and Fabio Persia. 2017. A Survey of Multimedia Recommender Systems: Challenges and Opportunities. *International Journal of Semantic Computing* 11, 03 (2017), 411–428. <https://doi.org/10.1142/S1793351X17500039>
- [5] MD. Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga. 2019. A Comprehensive Survey of Deep Learning for Image Captioning. *ACM Comput. Surv.* 51, 6, Article 118 (Feb. 2019). <https://doi.org/10.1145/3295748>
- [6] Mozghan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems—Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.
- [7] Andreas Lommatzsch, Benjamin Kille, Frank Hopfgartner, Martha Larson, Torben Brodt, Jonas Seiler, and Özlem Özgöbek. 2017. CLEF 2017 NewsREEL Overview: A Stream-based Recommender Task for Evaluation and Education. In *8th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF 2017)*. Springer.
- [8] Nelleke Oostdijk, Hans van Halteren, Erkan Başar, and Martha Larson. 2020. The Connection between the Text and Images of News Articles: New Insights for Multimedia Analysis. In *Proceedings of The 12th Language Resources and Evaluation Conference*. 4343–4351.
- [9] Sergio Oramas, Oriol Nieto, Mohamed Sordo, and Xavier Serra. 2017. A deep multimodal approach for cold-start music recommendation. In *Proceedings of the 2nd workshop on deep learning for recommender systems*. 32–37.
- [10] Özlem Özgöbek, Benjamin Kille, Jon Atle Gulla, and Andreas Lommatzsch. 2019. The 7th international workshop on news recommendation and analytics (INRA 2019). In *Proceedings of the 13th ACM Conference on Recommender Systems*. 558–559.
- [11] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *J. Mach. Learn. Res.* 21 (2020), 95–1.
- [12] Quoc-Tuan Truong and Hady Lauw. 2019. Multimodal review generation for recommender systems. In *The World Wide Web Conference*. 1864–1874.
- [13] Xinyi Zhou and Reza Zafarani. 2020. A Survey of Fake News. *Comput. Surveys* 53, 5 (Sep 2020), 1–40. <https://doi.org/10.1145/3395046>