

News Images in MediaEval 2020

Benjamin Kille¹, Andreas Lommatzsch¹, Özlem Özgöbek²

¹Berlin Institute of Technology, Germany

²Norwegian University of Science and Technology, Norway

benjamin.kille@dai-labor.de, andreas.lommatzsch@dai-labor.de, ozlem.ozgobek@ntnu.no

ABSTRACT

Images play an important role in online news reading behavior. They attract users' attention and they can decide whether users pay attention to some content over other content. News Images in MediaEval 2020 aims to gain more insight into the interplay of news images and news consumption. Within this task, participants get access to a large set of articles and accompanying images. The task consists of two separate subtasks. Participants can choose to participate in both or one of them. In the first subtask, participants have to predict which images and articles have been paired by the publisher. In the second subtask, participants have to estimate the likelihood that users will click recommendations consisting of pairs of articles and images. This paper describes the task setting in detail and draws connections to existing research. The overview illustrates the metrics and evaluation procedures that are used.

1 INTRODUCTION

Online news articles are multimodal: the textual content of an article is often accompanied by an image. The image illustrates the text's content and attracts readers' attention. Research in both multimedia and recommender systems domains generally assumes a simple relationship between images and text occurring together. For instance, image captioning [4] often assumes that the caption quite literally describes the image's scenery. However, other research shows that when images accompany news articles, the relationship becomes more complicated [7]. The MediaEval 2020 News Images Task investigates the real-world relationship of news text and images in more depth, in order to understand its implications for journalism and news recommender systems.

The task branches into two subtasks, both of which participants can address using text-based or image-based features. The first subtask focuses on predicting which images and articles have been paired by the publisher, whereas the second subtask focuses on estimating the likelihood that users will click recommendations consisting of pairs of articles and images. Given these two subtasks, the ultimate objective of this task is to gain additional insight about i) the relationship of news text and the images accompanying them. ii) the connection between the image and title shown by a recommender system to users, and iii) the tendency of users to click on the recommended article. In particular, the main focus of this task is research that transcends conventional work in the area of image concept detection and that includes aspects of images that go beyond their literally depicted content (such as quality, style, and framing).

2 BACKGROUND AND RELATED WORK

The *Multimedia Evaluation Benchmark* (MediaEval) investigates the intersection of multimedia and recommender systems for the third time in 2020. In 2018, the *NewsREEL Multimedia*¹ task provided data from multiple publishers concerning the interaction of users with content. In 2019, a subtask of the *Multimedia RecSys*² featured similar data.

Publishers employ news recommender systems to personalize their services [5]. This emergence of 'fake news' has fueled the interest in news recommender systems [9]. Research has picked up on the demand and established venues to discuss the relation of news recommendation and misinformation [8]. Besides the recommended content, researchers devote more and more attention to the presentation. The research distributes across different areas. Research on image analysis produces tools and models to extract better features from image data. Recommender system research strives to better understand personalization and user behavior. The subfield dedicated to news recommendation deals with the particularities of news. For instance, users exhibit a session-based interest as opposed to long-term interests with regard to music, literature, or television. The subfield of multimedia recommendation delves deeper into how content can contribute to generating recommendations for users.

3 TASK DESCRIPTION

The task seeks to explore the relation between images and articles. We define two subtasks, either or both of which participants can choose to take part in.

3.1 Task 1: Image-Text Re-Matching

In practice, publishers employ staff to search for images to accompany news articles. In many cases, the employees have access to imagery from the event. Sometimes, they select images from a database (e.g. stock images). As a result, readers encounter pairs of articles and images. This subtask has removed the link between images and articles. Thus, participants separately get a list of articles and images. Participants must develop suited models to reconstruct the link between articles and images. These models can help us to understand what makes an image fitting to an article.

3.2 Task 2: News Click Prediction

Publishers continuously monitor users' interactions with their online services. Webservers record clicks to provide the basis for optimization. The servers' logs reveal that some articles attract more views than others. We hypothesize that images play a role in users' complex decision making. The evaluation data has the

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

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

click statistics removed. Participants must develop suited models to estimate the likelihood of clicks. These models can reveal what makes an image appealing to users.

Both subtasks investigate news consumption behavior. We will assess submission both in terms of quantitative performance—i.e. measured by the evaluation metrics—as well as qualitative insight into the interplay between images and news consumption.

4 DATASET

Server logs, covering a three-month period, constitute the building block for this task’s data set. The logs have been obtained from a big German publisher. They comprise information related to articles, images, and interactions with users. The data set represents articles with a reference, the link to the article, the title, and a text snippet of at most 256 characters. The data set presents images as the pair of a reference and the link to the image. As the publisher maintains the copyright for the images, participants need to download them individually. Interactions between users and content occur in three ways: reading, being recommended articles, and clicking these recommendations. Reading and clicking on recommendations are interactions triggered from the user. Generating recommendations is triggered by the system.

The data set comes in three batches. The first and second batch constitute the data designated to train the models. These batches include a mapping between articles and images as well as the interaction statistics. The third batch splits into separate files for the articles and images. Besides, the batch omits the interaction statistics. In addition to the images, the data includes tags derived from the images using the ImageNet model [2]. Participants can use the tags a textual representation of the images.

Table 1: Data Set Statistics. The number of cases refers to both articles and images. Cases with articles using the same image have been removed. The estimated download time has been measured at the Technische Universität Berlin with a standard laptop.

Feature	Batch 1	Batch 2	Batch 3
Time Span	January 2019	February 2019	March 2019
Purpose	Training	Training	Evaluation
No. Cases	4688	4676	4114
Download Time	45 min	45 min	40 min

Table 1 summarizes the data set. All batches contain between 4000 and 5000 pairs of articles and images. The cases have been assigned to the batches based on the chronology of the log files. Participants ought to be able to obtain the images in less than three hours with a standard internet connection.

5 EVALUATION

The third batch of the data set lacks both the link between articles and images and the interaction data. The two subtasks challenge participants to reestablish them. Participants can submit up to five runs for each of the two subtasks.

5.1 Task 1: Image Task Re-Matching

The evaluation set contains 4114 images and articles each. A valid submission pairs exactly one image with exactly one article. Specifically, the participants have to submit a file with two columns. The first column must contain the image references (i.e., `iid`). The second column must contain the article references (i.e., `aid`). To compare submissions, the evaluation protocol computes the proportion of correctly matched pairs. For instance, if participants accurately matched 1000 pairs, the score will be 1000 divided by 4114 or $\approx 24.3\%$.

5.2 Task 2: News Click Prediction

The training data reveals how often the system has recommended each article and how often these recommendations have resulted in clicks. This information remains hidden for the evaluation data. Participants must estimate the chance of an image being clicked. Hence, a valid submission presents two columns. The first column contains the image reference (i.e., `iid`). The second column features a numerical value corresponding to the likelihood of a click for that particular image. The evaluation protocol follows a three-step procedure. First, the protocol eliminates all images that had not been displayed to users at least 100 times. This step is necessary for robust calculation of the evaluation scores. As a result, the evaluation set retains 2329 images. Without this step, an image which has been displayed twice and clicked once would obtain a high score without meaningfully reflecting the performance of the approach. Second, the protocol sorts all images according to their estimated likelihoods. Third, the protocol compares the obtained ranking to the actual ranking to compute the precision. Precision quantifies the proportion of relevant items ranked at the top of the list. In this task, we have identified the 85 images with the highest likelihood of being clicked. Hence, we compute precision as the proportion of those images ranked in the top 85 images in the submission.

5.3 Run Description

Participants report results in dedicated working notes. The results ought to highlight their reasoning, qualitative insights, and critical reflections about what can be deduced from the quantitative results. Participants can submit up to five runs for each subtask.

6 CONCLUSION

Understanding the complicated relation of content and presentation remains a tough challenge. Various external factors impede drawing conclusion from data samples. This task strives to shed light on a subject that has become increasingly relevant, which is related to images and their strong influence on the perception and the authenticity of news. The presence of ‘fake news’ threatens social cohesion. Insights into the effect of content presentation yield the potential to safeguard against the erosion of trust into media. Knowing what features to consider when detecting fake news can help publishers to prevent their spread.

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] 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.
- [3] 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>
- [4] 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>
- [5] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems—Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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>