Predicting the Interest in News based On Image Annotations

Alexandru Ciobanu
Technische Universität Berlin
alexandru.ciobanu@campus.
tu-berlin.de

Andreas Lommatzsch
DAI-Labor, TU Berlin
andreas@dai-lab.de

Benjamin Kille
DAI-Labor, TU Berlin
benjamin.kille@dai-labor.de

ABSTRACT

In recent years, the World Wide Web has changed from text-focused web pages to multi-media sources featuring photos, videos, and audio. The worldwide growth of broadband connections has facilitated this trend and supports the spread of user-generated content. Navigating and finding interesting content has become a difficult challenge. In this paper, we present approaches which use visual features to predict how interesting a news article will be. This task is part of the NewsREEL Multimedia challenge. The challenge provides a large-scale data set of news items, images, and interactions. We implement a recommender system which can distinguish interesting articles from irrelevant ones based on image features. We evaluate the system's throughput and predictions. We explain our insights and outline ideas to apply the gained knowledge in additional domains.

KEYWORDS

Multimedia, News, Recommender Systems, Image Analysis

1 INTRODUCTION

The number of documents and news articles published on the World Wide Web has increased dramatically. Users struggle to find relevant items. Recommender systems support users by reducing information overload. They analyze users' behavior toward items and derive patterns to determine the most relevant items. Collaborative filtering and content-based filtering have become the most widely used algorithms for recommender systems.

Multimedia content—e.g. photos, videos, and audio—permeate our everyday lives. More and more services emerge that enable us to share photos and videos. Still, research on recommender systems has yet to leverage multimedia content. This work contributes to this effort by focusing on the use of image data for recommending news. In particular, we use methods which automatically determine fitting descriptors for images. The task asks us to estimate how interesting freshly published news articles will become. The evaluation setting equates interestingness with popularity due to the lack of user profiles. Hence, we focus on non-personalized recommender systems. We hypothesize that images play a decisive role as they capture users' attention. Thus, we use image annotations to implement an estimator.

The remainder of the paper is structured as follows: In Sec. 2 we recapitulate the scenario. Sec. 3 discusses related work. We present the approach in Sections 4. Subsequently, Sec. 5 illustrates the evaluation results. Finally, Sec. 6 details our findings and gives an outlook to future research.

2 PROBLEM DESCRIPTION

Several domains demand to estimate items' relevancy based on images. In this work, we address the task defined by NewsREEL Multimedia [7]. We determine the most relevant news items based on the multimedia dataset provided by the task organizers. The data include news articles, images displayed next to them, and interactions with readers. We report the evaluation metrics precision at ten (Prec@10) as well as precision at the top ten percent (Prec@10%). We consider each news portal ("domain") independently. More details can be found in [7].

3 RELATED WORK

Recommender systems support users in finding the most interesting information. Traditionally, recommender system analyze user profiles and provide recommendations based on the similarity in the user behavior ("Collaborative Filtering" [4, 5]). In the worldwide web, users can anonymously access most websites as they relinquish login procedures. As a result, systems lack access to comprehensive user profiles. They rely on session-based approaches or content-based filtering instead.

Item-based recommender algorithms correlate item features and user feedback which is taken to indicate the interest in the items. Item features can be defined based on the item content. Typically, text-mining approaches or semantic algorithms—describing the item based on ontologies—are used to obtain the item features [6, 8].

Reduced computational costs have facilitated deep learning approaches for recognizing patterns in images. Deep Learning frameworks, such as Tensorflow [1] or Keras [2] trained on large image collections try to automatically identify concepts in images and to label images with meaningful terms. The quality of the image annotations depends on the concrete scenario and the size of the training dataset.

The use of automatically computed image features for news recommender systems is still a topic for future research. Several case studies [3] suggest that there is a potential for developing useful recommender systems based on visual image features. This motivates us to implement new recommendation algorithms with image features. The subsequent sections explain our approach and the implementation.

4 APPROACH

We consider only the images displayed next to the news items. We ignore additional meta-data or textual features such as text snippets or headlines. In the first step, we annotate the images. We use Google Vision—Google's Image Annotation Service—to ensure reliable labels. We annotate all images provided by the dataset in NewsREEL Multimedia. Google Vision outputs a list of labels and their probabilities. We use the five most likely labels.

Having inspected the image annotations, we recognized the need to process the labels further. Many labels exhibited a too fine-grained level of details. Consequently, we have trimmed the labels to the first word. For instance, *"football equipment and supplies"* has become *"football"*.

Our approach assumes that the labels represent the key information to estimate how exciting news items are. We use the impression information in the training dataset—available for weeks one to three, and six to eight—to train an estimator. In other words, we calculate the number of impression for each label. Some labels appear in more articles than others. Still, readers' preferences remain uncertain. Thus, we normalize the labels' weights obtaining the average impression per label. As a result, we get three figures: the total number of impressions, the average number of impressions, and the number of articles linked to the label. We carry out the calculations for each news portal separately. This accounts for variations in topics amid publishers. Furthermore, the publishers vary concerning the number of impressions which could bias our features.

We estimate an item's popularity based on the five labels assigned to its accompanying image. Subsequently, we sort the items according to their scores and submit the top items to the task organizers.

5 EVALUATION

Our model uses 94 000 news articles and 704 000 labels. The automatic annotation failed for about 1.4% of all articles. In some cases, Google Vision failed to provide labels. In other cases, labels exhibited a low probability. We successfully process 96% of all items contained in the test set. For the remaining 4% either no label was found or the label did not exist in the training set. We explored a variety of hyper-parameters to optimize our estimates. For instance, we varied the number of labels and the weeks used to train the estimator. We have validated different settings. Eventually, we observed the best performance for the configuration with five labels and the entire training data. We have submitted these estimates to the task organizers. We obtained results regarding precision at ten, precision at the top ten percent, as well as average precision at the top ten percent.

Table 1 lists the results for publishers 13554, 17614, and 39234. The results show that the prediction quality highly depends on the news portal. Our approach performs very successfully for publisher 13554. Our method achieves 70% Precision@10, and 58% Precision@10%. Analyzing the model in detail, we find that photos of German car brands and items comparing different cars are popular on this domain; articles without car exterior photos (e.g. portraits, buildings and cockpit designs) get only a small number of impressions.

For the domain 17614 our approach outperforms the baseline [7], but reaches a lower precision for portal 13554. Analyzing the annotations most important for classifying the items on website 17614, we find that images annotated with police and transportation are popular in this domain.

For the domain 39234 the approach performs similar to the baseline. The big variance in the observed prediction performance seems

Table 1: Evaluation results for our approach grouped by news portals and weeks. Columns refer to precision at ten (P@10), precision at the top ten percent (P@10%), and average precision at the top ten percent (AP@10%).

	domain 13554			domain 39234			domain 17614		
Week	P@10	P@10%	AP@10%	P@10	P@10%	AP@10%	P@10	P@10%	AP@10%
04	0.70	0.63	0.58	0.10	0.14	0.11	0.10	0.25	0.21
10	0.70	0.56	0.51	0.00	0.14	0.10	0.00	0.28	0.19
11	0.70	0.58	0.51	0.00	0.15	0.10	0.00	0.27	0.19
12	0.70	0.54	0.52	0.10	0.17	0.11	0.00	0.27	0.19
avg.	0.70	0.58	0.53	0.05	0.15	0.11	0.03	0.27	0.20

to indicate that the computed annotations are only suitable for prediction the popularity of items in certain domains. Moreover, the importance of images for the popularity of images may differ on the considered news portals.

6 CONCLUSION

In this paper, we have presented several approaches for estimating the interests in news items based on visual features. Results show that our approach outperforms the baseline. Still, textual features seem to contain more information than visual features. We have observed varying levels of performance depending on the publisher. For some publishers—e.g. 13554—visual features perform far above the baseline. For other publishers, on the other hand, differences remain small.

Our approach determines fitting descriptors for images. Thereby, we optimize the recommendations indirectly. We suppose that readers engage with concepts related to labels. Alternatively, we could hypothesize that readers react more strongly to the image rather than the concept. If this thesis holds, we may be better off designing low-level image features.

We plan to extend this line of research. Currently, our system considers labels separately. We will develop a model for label categories which will allow us to improve the preprocessing. Besides, we will further analyze the labels for each domain. We expect manual inspect to provide valuable clues on how to improve performance.

REFERENCES

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, pages 265–283, Berkeley, CA, USA, 2016. USENIX Association.
[2] F. Chollet et al. Keras. https://keras.io, 2015.
[3] F. Corsini and M. Larson. CLEF NewsREEL 2016: Image based Recommendation. In *Working Notes of the 7th International Conference of the CLEF Initiative, Evora, Portugal*. CEUR Workshop Proceedings, 2016.
[4] J. Herlocker, J. Konstan, A. Bochers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99)*, 1999.
[5] Y. Koren and R. Bell. *Advances in Collaborative Filtering*, pages 145–186. Springer US, Boston, MA, 2011.
[6] A. Lommatzsch. *Semantic Movie Recommendations*, chapter 5, pages 133–154. Advances in Computer Vision and Pattern Recognition. Springer International Publishing, Smart Information Systems edition, 2015.
[7] A. Lommatzsch, B. Kille, M. Larson, F. Hopfgartner, and L. Ramming. NewsREEL Multimedia at MediaEval 2018: News Recommendation with Image and Text Content. In *Procs. of MediaEval 2018*.

[8] P. Lops, M. Gemmis, and G. Semeraro. Content-based recommender systems: State
     of the art and trends. In *Recommender Systems Handbook*, chapter 3, pages 73–105.
     Springer, 2011.