CaptureBias: Supporting Media Scholars with Ambiguity-Aware Bias Representation for News Videos

Markus de Jong[2], Panagiotis Mavridis[1], Lora Aroyo[3], Alessandro Bozzon[1], Jesse de Vos[5], Johan Oomen[5], Antoaneta Dimitrova[3], and Alec Badenoch[4]

[1] Vrije Universiteit Amsterdam, User-Centric Data Science Group
{lora.aroyo,m.a.dejong}@vu.nl
[2] TU Delft, Web Information Systems
{p.mavridis,a.bozzon}@tudelft.nl
[3] Leiden University
a.l.dimitrova@fgga.leidenuniv.nl
[4] Utrecht University
A.W.Badenoch@uu.nl
[5] Beel en Geluid
{joomen,jdvos}@beeldengeluid.nl

Abstract. In this project we explore the presence of ambiguity in textual and visual media and its influence on accurately understanding and capturing bias in news. We study this topic in the context of supporting media scholars and social scientists in their media analysis. Our focus lies on racial and gender bias as well as framing and the comparison of their manifestation across modalities, cultures and languages. In this paper we lay out a human in the loop approach to investigate the role of ambiguity in detection and interpretation of bias.

Keywords: Bias detection · bias in news video files · ambiguity-aware bias representation · disagreement · machine learning · crowdsourcing · human in the loop

1 Introduction

The interpretation of textual and visual media is typically a subjective process where personal views and biases are becoming interlaced with and indistinguishable from the actual media content. For example, ethnic groups can be misrepresented by numbers in crime reports [10] and international news agencies can adjust the contents of their reports to tap into certain biases that they believe are present in the intended public [9]. So, the different points of view typically get expressed as a disagreement among different authors and consumers of the media content. The disagreement can be seen as a signal to identify the presence of ambiguity and has an effect on the detection of bias in visual and textual media, as well as on the understanding the meaning of the media message.

Studies of visual and textual media bias can be quite labor-intensive when performed manually [21], *e.g.* through labeling manually hundreds of hours of video [9]. With the exponential growth of visual (news) content, many machine learning and human computation approaches are emerging for the automation of the labeling, analysis and processing of video and textual material. In this work, we aim at further extending the state of the art for large-scale processing of textual and visual media to support media professionals, humanities and social science scholars in their process of analyzing news media (with respect to studying framing, gender and racial bias in news). The central point here is the study of content and semantic ambiguity when it comes to determining the topic, the events and the sentiment of the media material. Further, we aim to understand what causes this ambiguity, what are different types of ambiguity and how they influence the understanding and the capturing of bias in visual and textual media across different languages.

The concrete objectives of this research are to support typical *digital humanities* analysis tasks, *e.g.*

- *distant reading* of large collections of visual and textual news for understanding patterns and contexts framing, racial and gender bias in news over time and across different cultures and languages
- *close reading* of specific instances of visual media for understanding aspects, properties and causes of framing, racial and gender bias in news over time and across different cultures and languages.

Therefore, we investigate the role of ambiguity of the media content, as well as the ambiguity of the topic(s), context(s) and specific event(s) and entities depicted in the news media for the detection of framing, racial and gender bias. Our research is guided by the following hypotheses:

- There are different causes for disagreement in interpretation of visual media that will lead to different types of ambiguity;
- Ambiguity found in visual media can be related to subjectivity;
- Different types of ambiguity and subjectivity can be used to detect different types of biases, such as framing, racial bias and gender bias.

2 Related work

Here we present the related work on disagreement and ambiguity that occurs after annotation tasks. As mentioned, disagreement is a signal for ambiguity or subjectivity. Then ambiguity itself can also be a sign of subjectivity. Then these signals appear in the different manifestations of bias through misrepresentation of *entities* with the method of *framing* [9] or with different *sentiments* attached to these entities. Some of the entities that contain gender and race can also often be misrepresented [18, 10]. In the following we present the work that is related to the the detection of the above signals and bias manifestations.

Methods that study or leverage the disagreement in order to identify the quality of annotations done by a crowd exist. For instance, in computational linguistics [4] use Generalizability theory as a means to capture the reliability of an annotation and identify the reasons behind the level of confidence and reliability we can have over an annotation. In [17] they also use crowdsourcing for annotations and identify different subgroups of disagreement between crowd-workers for annotations and compare them with expert annotations. Also [8] propose a different measure for agreement that solves a number of problems that arise when other agreement measures are used for interval values. Instead they propose to reason about the type of agreement or disagreement by looking into the distribution of answers within an interval of values when suitable for the problem. On the other hand, [25] identify also disagreement and divergence into groups of coders and evaluate two tree based ranking metrics to compare disagreements.

Crowdtruth is a platform [16] that applies disagreement analytics to generate ground truth data with the use of crowdsourcing. It has been used to identify and name entities as well as determine annotation ambiguity [15], to detect language ambiguity in medical relations in texts [11] and to determine intrinsic ambiguity of events in video event detection [14]. Another automated method that uses the crowd predicts the ambiguity of images to assist in an crowdbased foreground object segmentation task [13].

Now, we take a look at the types of bias we are interested in: framing, racial bias and gender bias. We give a short definition of these, followed by related research methods for those biases.

A *frame* of a message can be described as 'highlighting some bits of information about an item that is the subject of communication, thereby elevating them in salience' [12], and the act of *framing* can be described as 'selecting and highlighting some features of reality while omitting others' [12]. For research purposes, it is therefore important to find the amount of attention that is given to a certain element (*e.g.* highlighting or downplaying) and what is omitted.

Gender and racial bias in media is most often investigated via certain *misrepresentations* and *presentations* of groups. An example of misrepresentation is when the number of group X shown on screen is not representative of the number of group X that are part of that society. An example of difference in presentation is when group X is presented or described in an different manner, *e.g.* shown in different sentiment than group Y or described with different adjectives, or when the focus lies on different properties of the groups. Therefore, the goals for investigating gender and racial bias here are (1) quantitative comparison with population statistics for misrepresentation, and (2) the rather more complex qualitative comparison between groups of the representation.

Framing can be investigated through manual thematic analysis [21]. However, automated methods also exist such as using keyword clustering to identify stakeholders standing on different sides [19]. Word-based quantitative text analysis and computer assisted methods have also been used, *e.g.* to identify interest group frames in the framing of environmental policy in the EU [5]. In the case of framing in video, we mentioned the investigation into framing in TV-news in countries that lie in overlapping spheres of influence of Russia and the EU [9], namely Belarus, Moldavia and Ukraine. In that study, 607 video news emissions were manually labeled on subject (EU, Russia), tone (positive, negative, neutral, none), theme (*e.g.* culture, history, security, values) and topic (*e.g.* external events or developments, human interest stories, visit from a state official). The relative number of reports on either EU or Russia was also compared. The results included statistics that showed different news channels aimed at particular local preferences (*e.g.* a shared religion, a shared history), but that (apart from the Russian channels) the news was in general most often balanced and neutral in tone and did not differ in tone towards either the EU or Russia.

As mentioned, research can discover racial bias expressed by discrepancies between actual on-screen role representation of ethnic groups and data from official statistics [10]. Example results from this 2017 investigation performed in Los Angeles showed that blacks were correctly reported as perpetrators, victims and police officers, and, while Latinos were accurately reported as perpetrators, they were underreported as victims and police officers. Whites were significantly overrepresented in all three categories. A similar quantitative comparison can be carried out to investigate gender bias, *e.g.* to investigate balanced reporting in sports [18]. This research also included qualitative research in which raters were asked to label announcer's language usage in relation to the athlete's gender (*e.g.* appearance, marital status) and imagery (*e.g.* active vs non-active pose, sports vs non-sports context). The researchers reported no significant quantitative gender bias, although there were still some differences found on other criteria. In other work, gender bias in Dutch newspapers expressed by stereotypical representation of male vs. female leadership in politicians was investigated with a dictionary approach [1].

To investigate framing and other biases, it is important to determine differences in message sentiment. Some automated text sentiment tools have been developed [20, 7] which are based on natural language processing (NLP). Voice tone is another possible source of sentiment analysis [26]. A relatively new modality in sentiment analysis is video, in which facial recognition techniques used to analyses actor's facial expression ('facial affect') [24]. Some work has also been done on creating an ensemble of all these sentiment analysis methods [22].

The methods put forward to analyze framing, gender and racial bias, however, do not make use of ambiguity in the crowd, even when such subjectivity may give us valuable information that could lead us to better detect bias and create better labels on subjective aspects as sentiment. Therefore, we propose an ambiguity-aware method that builds on CrowdTruth methodology [16] that will make use of ambiguity in the crowd to better detect bias.

3 The Approach: Disagreement-based Ambiguity for Bias Detection

We perform a number of knowledge acquisition experiments with media scholars and social scientists to determine aspects of bias in different modalities, cultures and languages. Next to this we also study ambiguity expressions, causes and types through crowdsourcing experiments for annotation of sentiment, topics, and opinions in news videos and articles. Main focus here is to understand (1) how disagreement is manifested as a signal for ambiguity, and (2) how ambiguity is related to subjectivity, and ultimately how these two lead to more accurate representation of bias in video and textual news. For this we apply, adapt and extend the CrowdTruth approach [3, 2, 16], which has been used to study disagreement-based ambiguity in various domains. We employ a hybrid human-machine system, where basic processing of both video and text material is performed to be used as a seed for the human computation tasks. Considering the large amount of video and text articles involved we envision an active learning cycle, where machine learning components continuously learn from humans-in-the-loop.

3.1 Dataset

Next, we describe the two types of data that we use and compare in our datasets: (1) textual and (2) video data.

Textual dataset Our textual dataset consists of news articles written in English from online sources such as: *e.g. BBC, The Guardian, CNN, Fox News, The New York Times, The Moscow Times, Sputnik, Breitbart News.* To identify target news events to study in videos, we use Wikipedia pages focusing on historical and political events [6]. Wikipedia provides crowd-sourced and editor-vetted articles from different contributors. We aim to extract event names and related event entities, e.g. people, organizations, locations and times and compare their representation in terms of opinions, perspectives and sentiment ground truth to compare the entities and facts presented within between different news sources.

Video dataset We perform experiments with a video dataset of short English language newsreels (*i.e.* a few minutes long with a spoken dialogue), accompanied by their metadata, *e.g.* short video description, title, tags, (auto-generated) subtitles and user comments. The videos in this dataset are collected from the following online news channels: *e.g. CNN, BBC, Al Jazeera, Sputnik, RT (formerly Russia Today), France24*. We also take advantage of the keyword annotated datasets on videos provided by YouTube in the YouTube8m dataset[7].

[6] Wikipedia: www.wikipedia.com
[7] YouTube-8M Dataset: https://research.google.com/youtube8m/

3.2 Data Preprocessing

We enrich the *subtitles*, *transcripts*, *in-video text* and *video metadata* with the set of events and related entities extracted from relevant Wikipedia pages and news articles.

Ambiguity signals in dataset We want to capture the different ambiguities from the dataset itself. For instance, using ControCurator[8] we process the comments from Wikipedia pages and YouTube videos from users in order to capture possible controversies. Also, for Wikipedia we can use a method similar to [23] in order to find controversial news articles from Wikipedia or Contropedia[9].

News event detection and data gathering After finding possible bias candidates with the use of the above tools from Wikipedia pages, we extract events using NLP processing. When Wikipedia articles are not present (for instance in the case of very recent news) we use different news article sources for the event and also make use of an initial video input from one source directly. We also use controversial video comments from these events, and, supported by Wordnet [10], we create seed words to assist a crowd to annotate an event. When the events are identified, we can collect video data from the different video channels of our initial dataset.

3.3 Disagreement for Bias Cues Extraction

In order to identify the framing, gender and racial bias introduced in news videos, we compare the information gathered from the video with Wikipedia and newspaper texts, as well as other videos (*e.g.* from other channels). When we are able to determine which main entities are related to an event, we can detect misrepresentions (of *e.g.* facts, actors) that might indicate framing. If a particular gender or race is misrepresented or represented in a certain way, we can infer gender and racial bias. As said, we base our bias cues on disagreement in both automatically extracted information and the crowd.

To be specific, order to be able to annotate videos for their events, we want to extract particular cues with both machine learning and human computation. Ideally, we want to identify with machine learning what needs to be annotated in the videos and transcripts by humans in order to find out *e.g.* what is being said, who is reporting, who is talking, how long are they talking, are they present at the scene of the news event?

To make use of all data modalities in our news videos, we investigate combining existing API's for textual, voice- and face-based sentiment analysis [22] in relation to the entities. Also, to be able to attach the entities to particular sentiments [6], we can compare different API's and state of the art methods and use their "disagreement" as a way to give a confidence to the combined output

[8] ControCurator: Crowds and Machines for Modeling and Discovering Controversy-http://controcurator.org/

[9] Contropedia: Analysis and visualization of controversies within Wikipedia articleshttp://contropedia.net/

[10] Wordnet: wordnet.princeton.edu/

and apply human computation to validate the sentiment analysis output from the machine learning methods. CrowdTruth[11] can be used to reason about the disagreement of the various subjects. Given that the crowd can also disagree for a particular subject, we investigate the reasons why the crowd could interpret a given message differently with regards to, for instance, their *demographics*.

4 Discussion

One of the limitations of our proposal is the lack of reliable data to capture 'opinion' neutral definition of recent events. As we use Wikipedia pages to extract both ground truth events to seed the search of these in media, as well as the intensity of edits and changes to these pages as an indication of possible controversy / bias or variety of opinions.

Acknowledgements

This research is supported by the Capture Bias project [12], part of the VWData Research Programme funded by the Startimpuls programme of the Dutch National Research Agenda, route "Value Creation through Responsible Access to and use of Big Data" (NWO 400.17.605/4174).

References

1. Aaldering, L., Van Der Pas, D.J.: Political leadership in the media: Gender bias in leader stereotypes during campaign and routine times. British Journal of Political Science p. 121 (2018). https://doi.org/10.1017/S0007123417000795
2. Aroyo, L., Welty, C.: The three sides of crowdtruth. Journal of Human Computation **1**, 31–34 (2014)
3. Aroyo, L., Welty, C.: Truth Is a Lie: CrowdTruth and the Seven Myths of Human Annotation. AI Magazine **36**(1), 15–24 (2015)
4. Bayerl, P.S., Paul, K.I.: Identifying sources of disagreement: Generalizability theory in manual annotation studies. Comput. Linguist. **33**(1), 3–8 (Mar 2007). https://doi.org/10.1162/coli.2007.33.1.3, http://dx.doi.org/10.1162/coli.2007.33.1.3
5. Boräng, F., Eising, R., Klüver, H., Mahoney, C., Naurin, D., Rasch, D., Rozbicka, P.: Identifying frames: A comparison of research methods. Interest Groups & Advocacy **3**(2), 188–201 (2014)
6. Calais Guerra, P.H., Veloso, A., Meira, Jr., W., Almeida, V.: From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 150–158. KDD '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/2020408.2020438, http://doi.acm.org/10.1145/2020408.2020438

[11] CrowdTruth: The Framework for Crowdsourcing Ground Truth Data http://crowdtruth.org/

[12] https://capturebias.eu/

7. Chaumartin, F.R.: Upar7: A knowledge-based system for headline sentiment tagging. In: Proceedings of the 4th International Workshop on Semantic Evaluations. pp. 422–425. Association for Computational Linguistics (2007)
8. Checco, A., Roitero, K., Maddalena, E., Mizzaro, S., Demartini, G.: Let's agree to disagree: Fixing agreement measures for crowdsourcing (October 2017), http://eprints.whiterose.ac.uk/122865/, © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org).
9. Dimitrova, A., Frear, M., Mazepus, H., Toshkov, D., Boroda, M., Chulitskaya, T., Grytsenko, O., Munteanu, I., Parvan, T., Ramasheuskaya, I.: The elements of russias soft power: Channels, tools, and actors promoting russian influence in the eastern partnership countries (2017)
10. Dixon, T.L.: Good guys are still always in white? positive change and continued misrepresentation of race and crime on local television news. Communication Research **44**(6), 775–792 (2017)
11. Dumitrache, A., Aroyo, L., Welty, C.: Crowdsourcing ground truth for medical relation extraction. arXiv preprint arXiv:1701.02185 (2017)
12. Entman, R.M.: Framing: Toward clarification of a fractured paradigm. Journal of communication **43**(4), 51–58 (1993)
13. Gurari, D., He, K., Xiong, B., Zhang, J., Sameki, M., Jain, S.D., Sclaroff, S., Betke, M., Grauman, K.: Predicting foreground object ambiguity and efficiently crowdsourcing the segmentation (s). International Journal of Computer Vision **126**(7), 714–730 (2018)
14. IEPSMA, R., GEVERS, T., INEL, O., AROYO, L.: Crowdsourcing for video event detection. In: Collective Intelligence (2017)
15. Inel, O., Aroyo, L.: Harnessing diversity in crowds and machines for better ner performance. In: European Semantic Web Conference. pp. 289–304. Springer (2017)
16. Inel, O., Khamkham, K., Cristea, T., Dumitrache, A., Rutjes, A., van der Ploeg, J., Romaszko, L., Aroyo, L., Sips, R.J.: Crowdtruth: Machine-human computation framework for sing disagreement in gathering annotated data. In: The Semantic Web–ISWC 2014, pp. 486–504. Springer (2014)
17. Kairam, S., Heer, J.: Parting crowds: Characterizing divergent interpretations in crowdsourced annotation tasks. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. pp. 1637–1648. CSCW '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2818048.2820016, http://doi.acm.org/10.1145/2818048.2820016
18. Kinnick, K.N.: Gender bias in newspaper profiles of 1996 olympic athletes: A content analysis of five major dailies. Women's Studies in Communication **21**(2), 212–237 (1998)
19. Miller, M.M.: Frame mapping and analysis of news coverage of contentious issues. Social science computer review **15**(4), 367–378 (1997)
20. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics. p. 271. Association for Computational Linguistics (2004)
21. Philo, G., Briant, E., Donald, P.: Bad news for refugees. Pluto Press (2018)
22. Poria, S., Peng, H., Hussain, A., Howard, N., Cambria, E.: Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. Neurocomputing **261**, 217–230 (2017)

23. Rad, H.S., Barbosa, D.: Identifying controversial articles in wikipedia: A comparative study. In: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration. pp. 7:1–7:10. WikiSym '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2462932.2462942, http://doi.acm.org/10.1145/2462932.2462942
24. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: A survey of registration, representation, and recognition. vol. 37, pp. 1113–1133. IEEE (2015)
25. Zade, H., Drouhard, M., Chinh, B., Gan, L., Aragon, C.: Conceptualizing disagreement in qualitative coding. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 159:1–159:11. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3173733, http://doi.acm.org/10.1145/3173574.3173733
26. Zhou, S., Jia, J., Wang, Q., Dong, Y., Yin, Y., Lei, K.: Inferring emotion from conversational voice data: A semi-supervised multi-path generative neural network approach (2018)