A Gamified Approach To Automatically Detect Biased Wording And Train Critical Reading

1st Smilla Hinterreiter
Institute of Industrial Design
University of Applied Sciences Magdeburg-Stendal
Magdeburg, Germany
smilla1.hinterreiter@stud.h2.de

Abstract—**Biased media has an effect on the public perception of occurring events. By altering word choice, outlets can alter beliefs and views. A gold standard data set is needed to train sufficient classifiers that detect biased wording. This work aims to develop a game that trains players to read news critically while collecting their annotations. The vision is to tackle the complex problem of media bias detection with a very scalable, high quality, and gold standard data set to overcome the drawbacks of current models in the area.**

Index Terms—**Serious Game, GWAP, HMI, Media Bias, Text Analysis**

I. Introduction

Balanced and unbiased information is crucial to develop an understanding of current events [1]. Media bias in the form of slanted news coverage or internal news article bias can alter beliefs and impact the public perception of events [2]. This bias leads to the polarization of public opinion, making "Filter Bubbles" or "Echo chambers" more likely to form [3]. News outlets can be biased in multiple ways [4]. Bias by word choice is defined by how bias can manifest itself in article sentences by altering word choices that refer to the same concept [5]. Furthermore, bias is sensitive to the individual background of consumers [4], [6].

Existing data sets in the domain exhibit major drawbacks, especially regarding their size and annotator information [4]. No data set has yet offered information reliable enough to use the advantages of modern machine learning methods to present a meaningful media bias measure to any user or reader [7]. Especially, the cost of gold standard data in a domain where agreement is difficult to achieve turns the data set creation into a challenge [8]. Sentences and words from English news articles have to be labeled by hand to create the data set. Crowdsourcing through tools like Amazon Mechanical Turk[1] cannot achieve the quality level of expert annotations [4]. Contributors need to be trained to understand biased wording and practice annotating to reach the standard of experts.

II. Problems and Intentions

Creating a dataset on biased wording is a challenge since crowdsourcing doesn't achieve the needed quality, while paying experts is expensive and slow. With the goal of educating about media bias and training players in recognizing it, this work intends to find ways to use game design for text annotation collection. A mobile game with the annotation as a core mechanic will be developed and evaluated to find out how a game can collect reliable player-generated data about biased wording. Using the data, a classifier will be build[2]. The three primary goals are firstly, creating a scalable approach for collecting text annotation, secondly, providing players with fun, learning value, and a deeper understanding of the media system and thirdly, use the data gathered within the system to develop a classifier able to identify media bias automatically.

III. Related Work

Crowdsourcing and Citizen Science have been successfully used to make discoveries or collect data with the help of volunteers. Prominent examples are FoldIt and Galaxy Zoo. In FoldIt, over 57,000 players folded different protein structures, outperforming AI solutions and solving the structure of an HIV enzyme within three weeks [9]. In GalaxyZoo more than 200,000 volunteers made over 100 million galaxy classifications [10]. The ESP Game by von Ahn [11] was the first Game With A Purpose (GWAP) to use humans for tasks a computer cannot perform by transforming image labeling into a game mechanic. This work inspired the development of more GWAPs like Wormingo [12] and WordClicker [13].

IV. Concept

The player's goal is to manage a media outlet and gain as much in-game money and followers as they please. The core mechanic is to choose which topics and sentences to publish. Depending on the sentence, the outlet becomes more polarizing or slanted, attracting different followers. Since bias is affected by the players' background, an assessment of their slant has to be done early in the game based on the questionnaire for media bias assessment [8]. A survey will be conducted to sort them in one out of five media outlet houses. The game has to appeal to players from the start; the learning curve has to be shallow, it has to be easy to pick up and put down, and quickly accessible through smartphones or web browsers.

---

[1]https://www.mturk.com/, accessed on 2021-08-08.

[2]This extended abstract puts more focus on the game than the classifier.

A. Annotations

Before publishing, sentences need to be analyzed through the annotation game. The goal is to train critical reading, emphasize the fun parts of annotating and make the repetitive parts less dull through feedback mechanisms and motivators. The players will learn how to recognize biased wording through an interactive tutorial, which consists of multiple short lessons followed by a practice phase. Players begin with swiping single words that are biased or unbiased and build their way up to complex sentences. For each correct annotation (based on the gold standard data set), players receive rewards. If the sentence is based on the gold standard data set, players receive instant feedback. In higher levels, more sentences where the answer is yet unknown appear. Instead of getting the full reward, players receive a small part while the sentence moves to the 'Published' page. When enough players annotated the same sentence, players navigate to 'Published' to receive feedback on their annotations and collect the full reward.

B. User Group

A diverse audience is needed to get a balanced data set. The medium casual game was chosen to reach a broad audience. The early adopters will be young academics and students. The jury audience consists of senior academics with an interest in language. The game will be designed focusing on these groups while keeping it simple and accessible to include a broad audience in later adoption phases.

V. EVALUATION

First, the evaluation focuses on the quality of the annotations. The game data set will be compared to the gold standard data set BABE [7]. This will provide information about how effective the learning effect through the tutorial is. The second evaluation will focus on the player's experience, deploying F2P metrics [13]. This allows drawing a conclusion on whether the game is enjoyable for players, hence whether they will play and annotate more.

REFERENCES

[1] K. Mossberger, C. J. Tolbert, and R. S. McNeal, *Digital Citizenship: The Internet, Society, and Participation*. The MIT Press, Oct. 2007. DOI: 10.7551/mitpress/7428.001.0001.

[2] M. Recasens and D. Jurafsky, "Linguistic Models for Analyzing and Detecting Biased Language," in *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, vol. 1, 2013.

[3] C. R. Sunstein, "The Law of Group Polarization," *Journal of Political Philosophy*, vol. 10, no. 2, pp. 175–195, 2002, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9760.00148. DOI: 10.1111/1467-9760.00148.

[4] T. Spinde, L. Rudnitckaia, J. Mitrović, F. Hamborg, M. Granitzer, B. Gipp, and K. Donnay, "Automated identification of bias inducing words in news articles using linguistic and context-oriented features," *Information Processing & Management*, vol. 58, no. 3, p. 102 505, May 2021. DOI: 10.1016/j.ipm.2021.102505.

[5] R. Puglisi and J. M. Snyder, "Chapter 15 - Empirical Studies of Media Bias," in *Handbook of Media Economics*, ser. Handbook of Media Economics, S. P. Anderson, J. Waldfogel, and D. Strömberg, Eds., vol. 1, North-Holland, Jan. 2015, pp. 647–667. DOI: 10.1016/B978-0-444-63685-0.00015-2.

[6] T. Spinde, L. Rudnitckaia, F. Hamborg, and B. Gipp, "Identification of Biased Terms in News Articles by Comparison of Outlet-Specific Word Embeddings," in *Diversity, Divergence, Dialogue*, K. Toeppe, H. Yan, and S. K. W. Chu, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2021, pp. 215–224. DOI: 10.1007/978-3-030-71305-8_17.

[7] T. Spinde, M. Plank, J.-D. Krieger, T. Ruas, B. Gipp, and A. Aizawa, "Neural media bias detection using distant supervision with babe - bias annotations by experts," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, Dominican Republic, Nov. 1, 2021, published.

[8] T. Spinde, C. Kreuter, and W. Gaissmaier, "Do You Think It's Biased? How To Ask For The Perception Of Media Bias," 2021. DOI: 10.5281/ZENODO.4651186.

[9] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and F. Players, "Predicting protein structures with a multiplayer online game," galaxy, *Nature*, vol. 466, no. 7307, pp. 756–760, Aug. 2010. DOI: 10.1038/nature09304.

[10] M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg, "Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers," *Astronomy Education Review*, vol. 9, no. 1, 2010, Publisher: Portico. DOI: 10.3847/aer2009036.

[11] L. von Ahn, "Games with a Purpose," *Computer*, vol. 39, no. 6, pp. 92–94, Jun. 2006. DOI: 10.1109/MC.2006.196.

[12] D. Kicikoglu, R. Bartle, J. Chamberlain, and M. Poesio, "Wormingo: A 'true gamification' approach to anaphoric annotation," in *Proceedings of the 14th International Conference on the Foundations of Digital Games*, San Luis Obispo California USA: ACM, Aug. 2019, pp. 1–7. DOI: 10.1145/3337722.3341868.

[13] C. Madge, R. Bartle, J. Chamberlain, U. Kruschwitz, and M. Poesio, "Incremental Game Mechanics Applied to Text Annotation," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Barcelona Spain: ACM, Oct. 2019, pp. 545–558. DOI: 10.1145/3311350.3347184.