Keous: a Tool for Mitigating Media Bias using Targeted Sentiment Analysis

Milo Knell

Abstract

Americans are more polarized than ever, and reside in increasingly separated ideological echo chambers. The news media is a polarizing force that contributes to this problem by presenting different analysis and coverage of events. This paper presents Keous, a tool to mitigate media bias by presenting articles on the same topic with different perspectives. Keous enables users to see different sides of every story to mitigate the bias of the news media, reduce the effects of ideological echo chambers, and facilitate diversity of thought. It finds articles on the same topic by fine-tuning BERT using triplet loss between headline and body texts to create semantically dense headline embeddings, and then performs OPTICS clustering on those embeddings. It uses another BERT model to perform state of the art targeted sentiment analysis on Sentihood (general only) and the newly created US Political News dataset. Within each topic cluster, the most dis-similar articles in opinion are selected using a novel dis-similarity function.

1 Introduction

Americans are growing increasingly polarized, according to a 2014 survey by the Pew Research Center [1]. As the median conservative and liberal grow apart, animosity and distrust grow. The same survey found that 27% of Democrats and 36% of Republicans view the other party as a threat to the well being of the nation. This polarization can have serious consequences such as in elections [2].

The news media's coverage of events can be a driver of polarization through the way it describes and analyzes events. This paper presents Keous, a tool to mitigate media bias by finding articles on the same topic with different perspectives. The goal of this system is to show users multiple viewpoints to reduce political echo chambers and polarization and promote diversity of thought.

In this paper, an article's meaningful content is defined as the entities within (people or groups). Keous aims to analyze the way an article refers to these entities to categorize the article's perspective. Consider the sentences "Donald Trump is awesome, but Bernie sucks!" and "Warren for life! Never Trump". They have targeted sentiments of {Donald Trump: 1, Bernie: -1} and {Warren: 1, Trump: -1}, which are used to define perspective in a matrix of entity sentiments. On both the benchmark dataset SentiHood [14] and the newly created US Political News dataset, Keous achieves state of the art results in targeted sentiment analysis. All code used can be found in the publicly available repository on GitHub [2].

The processing pipeline of Keous is as follows: scrape news websites to collect articles, generate headline embeddings and create clusters, predict sentiment, link entities, and create pairs.

2 Related work

2.1 News media bias mitigation

Only two similar works exist. AllSides, an online news aggregator, uses preexisting left-right source leanings to display an article from a left and a right

[1] https://www.people-press.org/2014/06/12/political-polarization-in-the-american-public

[2] https://github.com/Keous/keous

source on the same topic. This does not allow for articles to have different stances than the outlet that made it, only compares on the left-right spectrum, and has human bias.

NewsCube from Park et al. 2009 [13] defines the content using aspects, which are found through term frequency weighting. Articles on the same aspect with different viewpoints (also calculated with term weighting) are then presented.

2.2 Targeted Sentiment analysis

Sentiment analysis was first pioneered by Pang et al. in 2002 [4], which branched into ABSA (aspect based sentiment analysis) to find a fine-grained sentiment associated with a specific aspect, like price or safety. TABSA (targeted aspect based sentiment analysis) was then proposed by Saeidi et al. in 2016 [14] with SentiHood, which takes the aspects from ABSA and applies them to specific targets. The current state of the art on TABSA was achieved by Sun et al. in 2019 with BERT-pair [15] on Sentihood which will be compared to Keous.

Sentiment analysis on Twitter data was started by Go et al. in 2009 [7], and in 2011 Jiang et al. [10] noted that 40% of errors in Twitter sentiment analysis came from not considering targets. They created a rules-based method for targeted sentiment analysis. This led to top targeted sentiment analysis models like TDParse [17] and the target dependant and target connection LSTMs [16].

Sentiment analysis in the news domain has little prior work, but in 2004 Grefenstette et al. [8] performed targeted sentiment analysis using lexical methods. They do not provide a benchmark for comparison.

3 Creation of US Political News dataset

To accurately predict targeted entity sentiment for categorizing perspective, a training dataset is needed in the domain of news media. 100 articles over the last 12 years were selected with 20 from the New York Times, 30 from CNN, and 50 from Fox News.

3.1 Procedure

Named entities were tagged using spaCy [9], and for each named entity workers on Amazon Mechanical Turk (MTurk) were asked to identify the way the article tries to make that entity appear in one of five categories: *Very positive*, *Positive Neutral/NA*, *Negative*, and *Very negative*. Annotators were instructed to use *Neutral/NA* when the sentiment towards the entity was neutral or simply did not exist.

Articles were each annotated by one reviewer. Articles whose sentiments were all marked the same or left blank were thrown out. After removing articles, 92 were left: 46 from Fox News, 29 from CNN, and 17 from the New York Times.

3.2 Dataset

The 92 articles have 6871 entities tagged with the following breakdown of sentiment: *Very negative: 526, Negative: 1506, Neutral/NA: 2755, Positive: 1506, Very positive: 966*. The strong skew towards Neutral/NA is mostly due to the tagger marking entities that are not relevant to the sentiment of the sentence. For example, "Joe Biden's America is safer than ever" would have the targeted sentiments {Joe Biden: 2, America: 0}.

4 Methods

In this paper, an article's meaningful content to defined by the entities within (people or groups), and the following sections describe how Keous uses this definition to create pairs on the same topic with opposing opinion. Keous makes use of pre-trained BERT (Bidirectional Encoder Representation from Transformers) [6] models. BERT is pretrained on a large corpus with a masked language model where it predicts hidden words, and a next sentence prediction model where it predicts if one sentence follows after the other. It makes use of a bidirectional transformer stack and mutliheaded self attention to create context dependant word embeddings. Keous uses the BERT Base model, with 12 layers and a hidden layer size of 768.

4.1 Topic Clustering

Popular topic modeling options like LDA [3] are not fine-grained enough for the task of grouping news threads. Simple word co-occurrence is too simple to encode complex semantic relationships used in news headlines. By fine tuning a BERT model using triplet loss, the semantic density of the headline embedding is maximized. Triplet loss is defined by the equation below, where a is the anchor input, p is positive input, and n is negative input.

$$L(a, p, n) = max\{d(a_i, p_i) - d(a_i, n_i) + margin, 0\}$$

This loss function ensures that the positive output will be closer to the anchor than the negative. The anchor is set as the body text of a news article, the positive to be the corresponding headline, and the negative to be a headline from a different article. This maximizes the difference between headline body pairs, providing headline embeddings that encode the most relevant information for clustering. The network was trained on 20,000 articles over 2 epochs with a learning rate of $3e-5$ and a mean pooling strategy. The OPTICS algorithm [1] with $min\_samples = 4$ was chosen to create the topics.

4.2 Targeted Sentiment Analysis

To predict sentiment, the BERT body-level embeddings trained in 4.1 were used as the basis for targeted sentiment analysis with a 5 output linear layer with softmax activation and a dropout layer. Training data from the US Political News dataset was split into paragraphs, and for each entity in a paragraph it was copied and the entity replaced with *[TARGET]*. For example, *Alice and Bob went for a walk* would be split into two samples: *[TARGET] and Bob went for a walk* and *Alice and [TARGET] went for a walk*. The dataset is shuffled and split into a train and test set. The model is trained over 6 epochs with a learning rate of $2e-5$ with a linear warmup over 10% of the training steps, and a dropout probability of 0.5.

4.3 Entity Linking

The same entities must be linked together for accurate comparison between articles. Future work could use neural models like [11]. Keous uses a simple method knowledge base method using NeuralCoref (based on [5]): for each coreference cluster all stop words are removed and the remaining words are counted: {Harris: 11, Kamala Harris: 3, Sen. Kamala Harris: 1, HARRIS: 1}. The mention cluster is vectorized by the weighted average of word2vec [12] vectors $V \in \mathbb{R}^{n \times d}$ and mention vectors $M \in \mathbb{R}^n$ (entity counts) where d is the dimensions of the embedding and n is the number of groups in the cluster. This computation is written most efficiently as $M \cdot V$.

To load a document into the knowledge base, each mention cluster is vectorized as described above, then cosine similarity is computed with all other mention clusters in the knowledge base. The highest similarity cluster in the knowledge base is chosen, and the mention counts are combined ({Obama: 11, Barrack Obama: 3} + {Obama: 7, Barrack Obama: 2, Former President Obama: 1} would become {Obama: 18, Barrack Obama: 5, Former President Obama: 1}). If no existing mention cluster is found similar enough (defined by the parameter min_cos_link, set to 0.85) then the current cluster is added to the knowledge base for the future.

Once the knowledge base is constructed, entity matrices can be created. Two matrices initially filled with zeros are created: $sentiment, mention \in \mathbb{R}^{a \times n}$ where a is the number of articles being processed and n is the number of clusters. The sentiment matrix holds the total sentiment, and the mention matrix holds the total mentions. For each entity in an article, the most similar mention cluster is found in the knowledge base. For an entity in article i whose knowledge base mention is j with a sentiment s, the following operations are performed: $sentiment[i][j] \mathrel{+}= s$ and $mention[i][j] \mathrel{+}= 1$.

4.4 Pair Selection

The stance of an article can be represented by the entities $\{e_0, e_{,1}, ..., e_n\}$, given that for each entity there is a sentiment $S_i \in (-1, 1)$ and a weight (occurrences in the article) $W_i \in \mathbb{N}$. This paper presents a new opinion dis-similarity function between two articles with indices a and b.

$$D(S_a, W_a, S_b, W_b) = \sum_{i=0} \hat{W_{ai}} \hat{W_{bi}} |S_{ai} - S_{bi}|$$

Where $S_a, W_a, S_b, W_b \in \mathbb{R}^n$ are selected as slices from the sentiment and mention matrices created in section 4.3.

This can be thought of as absolute difference in opinion ($\sum |S_{ai} - S_{bi}|$) weighted by cosine similarity ($\sum \hat{W_{ai}} \hat{W_{bi}}$). The weighting is to make sure that a difference in opinion is meaningful in the article. Dissimilarity is computed pairwise amongst all articles in each topic cluster, and the highest pair in each with a minimum entity cosine similarity of *min_cos_topic* (set to 0.5) are selected.

5 Experiments

5.1 Targeted Sentiment analysis

Keous is compared to the current state of the art model on Sentihood, BERT-pair [15] which generates auxiliary sentences and evaluates them using BERT [6]. Keous will also be compared to TDParse [17], a top Twitter targeted sentiment analysis model that uses a heavily feature engineered linear SVM. Spacy's neural parser [9] is used for dependancy parsing. All models are compared on accuracy and macro averaged f1 score. The neural methods (Keous and BERT-pair) are averaged over 5 runs.

Sentihood, created by Saeidi et al. [14] contains aspect-level sentiments for targeted entities. For comparison only the general sentiment for each target entity is used, and all sentences with the sentiment *None*1 are left out, making a binary classification between *Positive* and *Negative*.

Model	**Accuracy**	**F1**
Keous	96.1	94.0
BERT-pair-QA-M	92.7	80.1
TDParse	83.0	74.0

Table 1: Performance on benchmark dataset Sentihood [14]

On the US Political News dataset, Keous is compared on a 5-way sentiment scale between *Very positive*, *Positive Neutral/NA*, *Negative*, and *Very negative*. TDParse cannot be compared on this non-binary classification task.

Model	**Accuracy**	**F1**
Keous	46.2	31.2
BERT-pair-QA-M	39.5	30.4
TDParse	-	-

Table 2: Performance on Us Political News dataset

6 Conclusion

This paper presents a method to mitigate media bias by showing articles on the same topic with different perspectives, making use of novel headline-body triplet loss to create semantically dense headline embeddings, state of the art targeted sentiment analysis on in-domain and benchmark data, a simple entity linker, and a novel dis-similarity function. The pairs generated by Keous will be posted at `https://www.keous.ai` weekly, where they can be read by the general public.

References

[1] Mihael Ankerst, Markus Breunig, Hans-Peter Kriegel, and Joerg Sander. Optics: Ordering points to identify the clustering structure. volume 28, pages 49–60, 06 1999.

[2] Dan Bernhardt, Stefan Krasa, and Mattias Polborn. Political Polarization and the Electoral Effects of Media Bias. (1798), 2006.

[3] David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 05 2003.

[4] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *CoRR*, cs.CL/0205070, 2002.

[5] Kevin Clark and Christopher D. Manning. Deep reinforcement learning for mention-ranking coreference models, 2016.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[7] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *Processing*, 150, 01 2009.

[8] Gregory Grefenstette, Yan Qu, James G. Shanahan, and David A. Evans. Coupling niche browsers and affect analysis for an opinion mining application. In *Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval*, RIAO '04, page 186–194, Paris, France, 2004.

[9] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

[10] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 151–160, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[11] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. End-to-end neural entity linking, 2018.

[12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[13] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. Newscube: Delivering multiple aspects of news to mitigate media bias. pages 443–452, 04 2009.

[14] Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1546–1556, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

[15] Chi Sun, Luyao Huang, and Xipeng Qiu. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence, 2019.

[16] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

[17] Bo Wang, Maria Liakata, Arkaitz Zubiaga, and Rob Procter. TDParse: Multi-target-specific sentiment recognition on twitter. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 483–493, Valencia, Spain, April 2017. Association for Computational Linguistics.