­Literature Review

**Predicting Political Bias in Media**

*Andrew Quirk & Cody Hankins. “CS224U” final project at Stanford CS department. Unpublished?*

@inproceedings

→ In terms of overall goal and motivation, fairly similar to what had inspired me with regard to my thesis idea. “The end goal is to educate the user: informing them, from a non-biased model, about the content they are consuming and if it is skewed toward a more liberal or a more conservative point of view.”

→ Corpus: transcripts of Congress speeches from 1993-2012. Labeled with time and speaker (for political affiliation).

→ Added: scraped articles from popular news websites. Library: newspaper3k. Used bias chart to assign political affiliation based on source.

→ Avoided POS labeling of texts for computational reasons.

→ aggregate word embeddings: cosine similarity/Jaccard similarity: measures similarity between two vectors of words. Congress speeches from the same year had 97% vector similarity.

→ Since the previous metric wouldn’t tell the difference between parties talking about the same issues (since many words would be repeated), they added other models to capture the higher-order relationships. The goal was to train a neural network classifier.

→ Bag-of-words model: Continuous Bag of Words (CBOW), which uses surrounding word context, Skipgram embeddings which rely on pairs of words (bigrams?) (FastText library). CBOW was most effective.

→ Produced word associations for each party for each “topic”. For example, the word “immigrants” had Republican associations with ‘lowskilled’, ‘wages’, ‘dropouts’, and Democrat associations with ‘nationals’, ‘apprehended’, ‘syndicates’.

→ Found a decline in accuracy as time goes on, which makes sense since they trained on Congress earlier sessions and rhetoric shifts over time. Generalization to modern data would be a problem if research was continued.

→ An idea they had was to use talk show transcripts as data. (Anderson Cooper, Bill O’Reilly, etc.)

**Political Ideology Detection Using Recursive Neural Networks**

*Iyyer et al, 2014.*

@inproceedings{iyyer2014political,

title={Political Ideology Detection Using Recursive Neural Networks},

booktitle={Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics},

author={Iyyer, Mohit and Enns, Peter and Boyd-Graber, Jordan and Resnik, Philip},

year = {2014}}

→ Notes that existing models focus on wordlists/bag-of-words. Goal was to model the compositional aspect of language.

→ “Bias may be localized to a small portion of the document, undetectable by coarse-grained methods.” Therefore, they decide to detect ideological bias on the sentence level. *“Ideological bias*” = if the “author’s political position […] is evident from the text”.

→ RNNs have achieved previous success in sentence-level NLP tasks, and do not rely on “hand-made lexicons, dictionaries or rule sets” (thereby distinguishing themselves from previous approaches to ideological bias).

→ “We initialize our model with 300-dimensional *word2vec* toolkit vectors generated by a continuous skip-gram model trained on around 100 billion words from the Google News corpus.” Vectors have linear relationships (mathematically close vectors = semantically close words). To keep these relationships as phrases are formed by combining words, “we initialize our left and right composition matrices such that parent vector *p* is computed by taking the average of children *a* and *b*.”

→ Initially used Convote, the dataset on congressional transcripts from 2005. Extracted sentences from this corpus which had a higher likelihood of political bias (since most sentences were generic/didn’t have bias) using features from Yano et al. (2010).

→ Developed a new political ideology dataset annotated at the phrase level. This started with the Ideological Books Corpus (IBC) by Gross et al. (2013) – “books and magazine articles written between 2008-2012 by authors with well-known political leanings”, documents “manually labeled with coarse-grained ideologies (right, left and center) as well as fine-grained ideologies (e.g. religious-right, libertarian-right) by political science experts”. They did the same sentence selection process as on Convote to filter out those sentences which were likely to have political bias. Then they reduce the proportion of neutral sentences in the dataset, because they want to focus on right/left, and used an active learning tool called DUALIST to label the remaining sentences with ‘neutral’ or ‘biased’ and then conservative or liberal. Finally, they used Crowdflower to crowdsource human annotations on the sentence and the phrase level (using syntax parsing). They end up with 3,412 sentences including 13,640 annotated nodes.

→ A note: “As we get farther from the root of the tree, nodes are more likely to be neutral.” Interesting to find evidence of syntactic relations illuminating bias by the way phrases are put together.

→ Baselines: bag of words, phrase-level annotations, syntactic pseudo-word features (Greene & Resnik 2009) from dependency relations, logistic regression model trained on the average of the pretrained word embeddings for each sentence. Baselines ranged from 50% (completely random) to ~67% (logistic regression model). They are predicting right/left, no neutral.

→ Feature vectors start for each node in the tree of a sentence and then are percolated to the root of the tree to obtain final instance representations, which are passed to the RNN.

→ Three RNN models were trained: one with random initial parameters and sentence-level labels, one with word2vec initialization and sentence-level labels, and one with word2vec embeddings and annotated phrase labels in addition to sentence-level labels. This last one performed the best at about 70% accuracy.

→ Found that best RNN model “is able to accurately model the compositional effects of bias in sentences with complex syntactic structures”. For example, “free market ideology” changing once it reaches a higher level phrase as “made worse by”.

→ The model can make errors when polarity switches such as the above occur at higher levels in the tree – perhaps since some information is lost at ever propagation step.

**Linguistic Models for Analyzing and Detecting Biased Language**

*Recasens et al, 2013.*

@inproceedings{recasens2013linguistic,

title={Linguistic Models for Analyzing and Detecting Biased Language},

author={Recasens, Marta and Danescu-Niculescu-Mizil, Cristian and Jurafsky, Dan},

year = {2013}}

→ Analyzes “real instances of human edits designed to remove bias from Wikipedia articles”. ‘NPOV’ corpus. Their dataset was cut to only those sentences which had one NPOV (neutral point of view) edit involving one word.

→ Predictive task: given a biased sentence, what is the bias-inducing word? Difficult even for humans.

→ Revealed two types of bias: epistemological and framing.

→ *epistemological bias* includes examples of assumptions/presupposed propositions in the text/focuses on a phrase’s believability. Some linguistic features:

→ factive verbs “presuppose the truth of their complement clause”. ‘reveal’ vs ‘indicate’; ‘he realized’ vs ‘his stand was’

→ entailments “are directional relations that hold whenever the truth of one word or phrase follows from another” - like implications of certain words. “Murder” vs “kill”; “was coerced into accepting” vs “accepted”

→ assertive verbs “are those whose complement clauses assert a proposition”, like ‘stated’ vs ‘claimed, or ‘said’ versus ‘pointed out’

→ hedges are “used to reduce one’s commitment to the truth of a proposition” - like adding ‘may’ or ‘possibly’

→ note that epistemological bias is *bidirectional*: it can happen when doubt is cast upon a fact which is generally accepted as true, or when an uncertain fact is stated with certainty

→ *framing bias* includes examples of perspective-specific words or praising. Some linguistic features (tend to be fairly word-specific, as opposed to epistemological bias): subjective intensifiers, one-sided terms (‘liberated’ vs ‘captured’, ‘pro-life’ or ‘anti-abortion’)

→ Logistic regression model trained on feature vectors for each word produces, for each sentence, the three highest ranked words by probability to be biased.

→ Features: lots. Ones that describe the word under analysis (lemma, POS), or its surrounding context (POS of word to left, presence of hedge in context), context being a 5 gram window (two to left, two to right). Predefined lists of hedges, factives, assertives, implicatives, entailments, subjectivity lexicon, sentiment lexicon, bias lexicon.

→ Outperforms sampling of baselines; including top 3 choices, results approach 60% accuracy.

→ Human annotation produced only 30% accuracy. (Suggests something about the dataset?)

**Mapping the echo-chamber: detecting and characterizing partisan networks on Twitter**

*Nourbakhsh et al, 2017.*

@inproceedings{nourbakhsh2017mapping,

title={Mapping the echo-chamber: detecting and characterizing partisan networks on Twitter},

author={Nourbakhsh, Armineh and Liu, Xiaomo and Li, Quanzhi and Shah, Sameema},

year = {2017}}

→ Paper ended up being not too relevant to me – it focuses on identifying linked networks on social media of like-minded individuals (echo chambers).

→ Does aim to identify the level of bias/partisanship “indicated by the homogeneity of content”, as well as identifying “topics or phrases that the community is vulnerable to spreading misinformation about”, and is “ideology-agnostic”.

→ Kaggle fake news dataset – they extract the domains included and find tweets from users who cite that domain. From those tweets and accounts, they extract the domains cited by each user, and then the bias of those domains (just by similarity to the Kaggle domains and associated labels). From that they label the relative partisanship of an account (which measures “the account’s tendency to share from similar domains”).

→ They find communities by mapping accounts which share many domains. They (with only brief description) describe a clustering model to find different ‘echo chambers’.

**Shedding (a thousand points of) light on biased language**

*Yano et al, 2010.*

@inproceedings{yano2010shedding,

title={Shedding (a thousand points of) light on biased language},

author={Yano, Tae and Resnik, Philip and Smith, Noah A.},

booktitle={Proceedings of NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk},

pages={152--158},

year = {2010}}

→ “Empirical study of linguistic indicators of bias in the political domain, using text drawn from political blogs.”

→ “Bias” is defined separately from “subjectivity” in NLP – subjective language is used to communicate the user’s own opinions or thoughts as opposed to factual information. Bias, on the other hand, is a “tendency or preference towards a particular perspective, ideology or result”.

→ To collect sentences for the study, they used gathered lists of “sticky” (strongly associated) bigrams in left/right-leaning texts. There were about 500 bigrams in each. Also extracted 4 categories of emotional words: negative, positive, causation, anger; finally, they extracted “kill” verbs – verbs with negative meanings that carry intention in the transitive state (like kill).

→ Annotation results carried about half ‘unbiased’ labels, and quite a few ‘biased, but not sure’ labels.

→ Agreement among annotations was reasonable/consistent.

**Automatic Detection of Political Opinions in Tweets**

*Maynard et al, 2012.*

@inproceedings{maynard2012automatic,

title={Automatic Detection of Political Opinions in Tweets},

author={Maynard, Diana and Funk, Adam},

booktitle={Proceedings of the 8th International Conference on The Semantic Web},

pages={88--99},

year = {2012}}

→ Refers to tweets as “microposts” and discusses the challenges related to working with them as opposed to longer texts.

→ History of sentiment detection can be split into lexicon-based methods and machine-learning methods. The former neglects contextual information. Tweets “typically do not contain much contextual information and which assume much implicit knowledge”. Less grammatical, more ‘extralinguistic’ cues such as emoticons and hashtags. This means that parsers will struggle with tweets. In addition, irony/sarcasm are common in tweets and difficult for machines to handle.

→ In their system, first tuples are formed to carry out sentiment analysis on each tweet. (Person, Opinion, Political Party), so (Bob Smith, pro, Labour). Then opinions by the same person can be collected. It can also see if changes in opinion have happened over time.

→ Corpus: political tweets over pre-election period in UK in 2010. Includes metadata though they do not take this information in their features.

→ Linguistic pre-processing: tokenization, POS tagging, morpological analysis, sentence splitting, but not full parsing. They do do named entity recognition (ANNIE, a tool that is part of GATE).

→ Phases of opinion detection: detecting pos/neg/neu words (affect annotation), “identifying factual or opinionated versus questions or doubtful statements”, “identifying negatives”, and “detecting extra-linguistic clues such as smileys. Further detail:

→ For the affect annotations, they use a gazetteer/sentiment dictionary, but does take in context to make sure the intended meaning of the word matches the one in the dictionary.

→ They also keep a gazetteer of frequently occuring hashtags.

→ They make sure to differentiate questions/unsure statements from opinionated phrases. The former would get a neutral opinion, not a biased one.

→ They not only identify opinions of the tweet-authors, but also extract other people’s opinions mentioned in tweets, e.g. “Vote for Labour. Harry Potter would.” (Harry Potter, pro, Labour)

→ Results are high precision at the expense of recall. Out of 1000 tweets, the system identified 143 as opinionated.

→ Problem: lack of world knowledge base. Tweet likens politician to Voldemort; model can’t understand this.

→ “isolation of powerful indicator words, politicized and carrying connotation or bias”

**Measuring ideological proportions in political speeches**

*Sim et al, 2013.*

@inproceedings{sim2013measuring,

title={Measuring ideological proportions in political speeches},

author={Sim, Yanchuan and Acree, Brice and Gross, Justin H. and Smith, Noah A.},

booktitle={EMNLP},

year = {2013}}

→ Goal: measure candidate’s ideological positioning from speeches. “Do political candidates in fact stray ideologically at opportune moments?”

→ model to infer mixtures of ideological positions in documents

→ Lexicons of “cues” (terms associated with an ideology) are automatically extracted from a corpus of books and magazines, manually labeled (for coarse and fine ideologies, and occasionally topics) by a political science expert who is also one of the paper authors.

→ Uses SAGE (sparse additive generative models, Eisenstein et al 2011) to assign probability to a text as if it were a bag of terms, parameterizing the distribution using a generalized linear model. They use it to define the probability of each term, depending on the attributes (in this case, the ideology labels) of the text it is in. They extracted a total of 8,483 cues.

→ Generally: “The term selection method we have described can be understood as a form of feature selection that reasons globally about the data and tries to control for some effects that are not of interest (topic or document idiosyncrasies).”

→ A speech is therefore presented as “a sequence of cues interspersed with lags”, which are the irrelevant filler text between ideological cues. The idea is that nearby cues will be of the same ideologies.

→ Bayesian statistics & probability used to evaluate uncertainty

→ Model name: CLIP (cue-lag ideological proportions), HMM where states correspond to ideologies and emissions are a cue and the lag amount after it.

→ a domain-specific topology relating ideologies to one another is also used as input, in order to “encode the closeness of different ideologies and, by extension, the odds of transitioning between them within a speech”. This defines transitions for states in the HMM.

→ There is a “restart” probability for the lag, to capture the idea that a longer lag after a cue term increases entropy over the next ideology state (meaning it’s more likely that it could be a different ideology).

→ They want to measure the amount of time spent in each ideology over each part of the campaign. Evaluation for this is difficult and they take a series of hypothesis that they measure against the model and baseline.

→ Notes: using candidates’ names as cue terms was a systematic issue.

→ In general this approach seems a little different from others that I’ve been reading, and while their goal differs from mine, I might come back to this later for ideas.

**Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective**

*Ahmed and Xing, 2010.*

@inproceedings{ahmed2010staying,

title={Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective},

author={Ahmed, Amr and Xing, Eric P.}

booktitle={EMNLP},

year = {2010}}

→ Lexical variation in a document because of 3 things: writer ideological belief, topical content, and topic-ideology interaction.

→ Goal is to be able to factor a document collection into a representation of all three of these things.

→ Uses Topic Models (Blei et al 2003). Multi-view Latent Dirichlet Allocation (mview-LDA). Views the word content of each document as the result of the interaction between the document’s ideological and topical dimensions.

→ Each ‘topic’ in mview-LDA is represented by a word distribution. Each document is tagged with the ideology it represents, which is a ‘topic’ of its own. There are also ideology-specific topics (how each ideology addresses each topic).

→ Bayesian approach, sampling from a posterior distribution to estimate parameters.

→ small datasets, ex ~600 articles and a trimmed vocab size of 4100 words.

→ View Extraction

**What drives media slant? Evidence from US daily newspapers**

*Gentzkow and Shapiro, 2010.*

@inproceedings{gentzkow2010whatdrives,

title={What drives media slant? Evidence from US daily newspapers},

author={Gentzkow, Matthew and Shapiro, Jesse M.},

booktitle={Econometrica},

pages={35--71},

year = {2010}}

→ slant index to rate ideological leaning of newspapers, governed by frequency of use of partisan collocations of 2-3 tokens

→ phrases such as “death tax”, “war on terror”, “war in Iraq” - they use these to identify partisanship, not general bias

→ finds that for areas, profit-maximizing slant is close to local newspapers’ actual slant – much of the paper is devoted to this economic modeling of media slant

→ they give their method of phrase selection – possibly replicable?

**Opinion mining and sentiment analysis**

*Pang and Lee, 2008.*

@inproceedings{pang2008opinion,

title={Opinion mining and sentiment analysis},

author={Pang, Bo and Lee, Lillian},

booktitle={Foundations and trends in information retrieval},

year = {2008}}

→ overall view of approaches for opinion extraction – this is a long survey

→ 4.1.2 is the relevant part: subjectivity detection and opinion identification. It’s pretty short, but there is a list of relevant work. Compares the problem of subjectivity detection with “genre classification”.

→ for feature ideas: term PRESENCE: binary feature vectors wherein entries merely declare whether a term occurs or not (for review polarity classification)

→ position information – place of a token within a textual unit

→ research has shown a high correlation between the presence of adjectives and sentence subjectivity

→ also nouns, verbs, in particular subjective nouns (“concern”, “hope”)

→ modality identification (opinion, assertion or description) using dependency tree syntax features

→ very domain-specific models/have domain transfer problems

→ unsupervised lexicon induction to create sentiment databases

→ “One interesting characteristic of document-level sentiment analysis is the fact that a document can consist of sub-document units (paragraphs or sentences) with different, sometimes opposing labels, where the overall sentiment label for the document is a function of the set or sequence of labels at the sub-document level.” Use these relationships to help achieve a more accurate labeling overall.

→ for example, ‘evaluative’ vs. ‘non-evaluative’ segments in a text.

→ Discourse structure: can model global sentiment of a document as a trajectory of local sentiment. Position matters for sentiments in a document.

**Sentence subjectivity detection with weakly-supervised learning**

*Lin et al, 2011.*

@inproceedings{lin2011sentence,

title={Sentence subjectivity detection with weakly-supervised learning},

author={Lin, Chengua and He, Yulan and Everson, Richard},

booktitle={Proceedings of AFNLP 2011},

pages={1153--1161},

year = {2011}}

→ Sentence-level subjectivity detection using latent Dirichlet allocation (LDA).

→ Weakly-supervised generative model learning using only a small set of domain-independent lexical subjectivity clues as input.

→ another common solution: use a small number of seed words with known polarity to infer the polarity o fa large set of unidentified terms, using mutual information and context.

→ subjLDA is “essentially a four-layer Bayesian model”. Draws subjectivity label for the sentence, sentiment labels for each word in the sentence, word conditioned on sentiment label

→ tested on MPQA dataset

→ used subjClue and SentiWordNet lexica: only took strongly subjective clues, discarding weakly subjective ones. “The rationale behind the filtering is that while a strongly subjective clue is seldom used without a subjective meaning, weakly subjective clues are ambiguous, often having both subjective and objective uses.” Final lexicon subset was 477 positive and 917 negative words.

→ They also used SentiWordNet’s neutral words, but found that there was little improvement from their inclusion.

→ Baseline accuracy was 63%, subjLDA gets 71.2%.

→ Argue that “appropriate filtering of neutral words is necessary in order to avoid introducing bias into model learning”; their dominance in terms of frequency biases a model toward the objective class.

**Lexical and hierarchical topic regression**

*Nguyen et al, 2013.*

@inproceedings{nguyen2013lexical,

title={Lexical and hierarchical topic regression},

author={Nguyen, Viet-An and Boyd-Graber, Jordan and Resnik, Philip},

booktitle={NIPS},

pages={1106--1114},

year = {2013}}

→ Similar to the previous one: using (supervised hierarchical) latent Dirichlet allocation (LDA).

→ focuses on agenda setting, framing, and topic modeling: how different language users approach different topics. “Prediction of perspective”.

→ Not super relevant, since it works on hierarchical topic extraction.

**A joint topic and perspective model for ideological discourse**

*Lin et al, 2008.*

@inproceedings{lin2008jointtopic,

title={A joint topic and perspective model for ideological discourse},

author={Lin, Wei-Hao and Xing, Eric and Hauptmann, Alexander},

booktitle={Machine Learning and Knowledge Discovery in Databases},

pages={17--32},

year = {2008}}

→ Hypothesize at the core that “ideological perspectives [are] reflected in lexical variations”

→ Develop a statistical model for ideological discourse, in which lexical variations are “encoded in a word’s topical and ideological weights”. The model uncovers these weights from texts and predicts its ideological perspective.

→ Again – this is not the most relevant for me, since this wants to differentiate ideological perspectives from one another, whereas I am just looking for the presence of essentially any ideology. However, there are some things to take away.

→ “Word frequency in ideological discourse should be determined by how much a word is related to a text’s topic (i.e. *topical*) and how much authors holding a particular ideological perspective emphasize or de-emphasize the word (i.e., *ideological*). A model for ideological discourse should take both topical and ideological aspects into account.”

**Learning subjective language**

*Wiebe et al, 2004.*

@inproceedings{wiebe2004learning,

title={Learning subjective language},

author={Wiebe, Janyce and Wilson, Theresa and Bruce, Rebecca and Bell, Matthew and Martin, Melanie},

year = {2004}}

→ Perhaps useful if I can use it to extract my own vocabulary of subjective words specific to the political news domain?

→ Hapax legomena: the set of words that appear just once in the corpus. “feature with high frequency and significantly higher precision than baseline”

→ Collocations: they describe a method to automatically identify collocational clues, such as fixed n-grams (‘of the century’ – note that this includes noncontent words/stopwords). “The method is then used to identify an unusual form of collocation: one or more positions in the collocation may be filled by any word that is unique in the test data.”

→ and third type of subjectivity clue: adjective and verb features identified “using the results of a method for clustering words according to distributional similarity”; they suggest that words may be in similar distributions because

→ “Many good clues of subjectivity occur with low frequency”; in fact, “uniqueness in the corpus is an informative feature for subjectivity classification”. (Note: remove numbers.) Not only that, but the precision of unique and other low-frequency words increases with corpus size.

→ Precision: the precision of a set of types S is the number of instances of members of S in opinion(/biased) pieces over the total number of instances of members of S in the data. Baseline precision is the # of word instances in opinion pieces over total number of word instances.

→ Data are noisy: “opinion pieces contain objective sentences, and nonopinion pieces contain subjective sentences”.

→ For collocations: first extract all n-grams up to n=4 from the data. Constituents are word-step, part-of-speech pairs, ex. In-prep the-det can-noun. Select a subset of the n-grams based on precision: must be greater than baseline, and greater than the maximum precision of its constituents (otherwise just take the smaller n-gram).

→ Special type of collocation called Unique Generalized N-Gram: placeholders for unique word. Replace all hapax legomena with “UNIQUE” (remember POS tag).

→ Distributional Similarity: they give an algorithm starting with a set of seed words that gives clusters for different parts of speech/syntactic relationships. For each seed set, the precision of the set in the training data is calculated and if it is greater than a threshold T, retain the set.

→ An idea explored is density parameters: if the subjectivity elements are more likely to be subjective if they are surrounded by other subjective elements.

→ “Using the k-nearest-neighbor classification algorithm with leave-one-out cross-validation, a classification accuracy of 94% was achieved on a large test set, with a reduction in error of 28% from the baseline.”

**More than words: syntactic packaging and implicit sentiment**

*Greene and Resnik, 2009.*

@inproceedings{greene2009syntactic,

title={More than words: syntactic packaging and implicit sentiment},

author={Greene, Stephan and Resnik, Philip},

booktitle={NAACL},

year = {2009}}

→ Instead of focusing on purely lexical features, this looks at “syntactic packaging”

→ passive voice use (“Mistakes were made”)

→ They give linguistic justification for these kinds of syntactic framing, focusing on certain types of verbs that can carry implications, and then empirical confirmation of these theoretical concepts.

→ Problem: the important properties of the verbs used (volition, causation, telicity etc) are not directly observable. Hence OPUS: observable proxies for underlying semantics. Idea is to “use observable grammatical relations … as proxies for the underlying semantic properties that gave rise to their syntactic realization using those relations.”

→ Used a specially constructed corpus on the death penalty to test the predictive power of their extracted features.

**Which side are you on? Identifying perspectives at the document and sentence levels**

*Lin et al, 2006.*

@inproceedings{lin2006identifying,

title={Which side are you on? Identifying perspectives at the document and sentence levels},

author={Lin, Wei-Hao and Wilson, Theresa and Wiebe, Janyce and Hauptmann, Alexander},

booktitle={Proceedings of CoNLL 2006},

pages={109--116},

year = {2006}}

→ Identifying which perspective a document here, like other articles mentioned, is not exactly what I am doing but is tangentially relevant.

→ Bitterlemons corpus on Israeli/Palestinian conflict (this is not the only paper here in which this has been used).

→ Focus on word usage/lexical features, uses machine learning to uncover word patterns. Doesn’t seem to add much new to the discussion.

**Opinion Observer: analyzing and comparing opinions on the Web**

*Liu et al, 2005.*

@inproceedings{liu2005opinion,

title={Opinion Observer: analyzing and comparing opinions on the Web},

author={Liu, Bing and Hu, Minqing and Cheng, Junsheng},

booktitle={WWW 2005},

pages={342--351},

year = {2005}}

→ This is a product to analyze customer reviews and quickly pick out the most salient points/pros and cons. In their own words, related but quite different from sentiment classification. I still read it to see if they used any inspiring features.

→ In terms of features, mostly POS tags and n-grams, so nothing particularly surprising or exciting.

**Recognizing stances in ideological on-line debates**

*Somasundaran and Wiebe, 2010.*

@inproceedings{somasundaran2010recognizing,

title={Recognizing stances in ideological on-line debates},

author={Somasundaran, Swapna and Wiebe, Janyce},

booktitle={Proceedings of NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text},

pages={116--124},

year = {2010}}

→ Ideological stance recognition – focusing on *arguing*

→ “**Obviously** that hasn’t happened, and to be completely objective (as all scientists **should be**) **we must** lean on the side of **greatest evidence** which at the present time is for evolution. [side: against the existence of God]”

→ Used the MPQA corpus for arguing subjectivity and generated an arguing lexicon.

→ Modal verbs such as must, should, ought. (Would these show up in my news articles…? Check.)

→ When searching for n-gram matches in the arguing lexicon: “prevent the same text span from matching twice – once a trigram match is found, a substring bigram (or unigram) match with the same text span is avoided”.

→ Done on a sentence level. An opinion (ap for arguing-positive and an for arguing-negative) is attached to all the content words in the sentence (target) to construct opinion-target features.

→ “For every modal verb detected, three features are created by combining the modal verb with its subject and object. Note that all the different modals are replaced by ‘should’ when creating features. This helps to create more general features.”

→ Sentiment-based features (independent of arguing features). Sentiment lexicon (Wilson et al 2005) includes pos, neg and subj words that are neutral in polarity (like absolutely, believe, think). Polarity for each sentence is calculated (Vote and Flip from Choi and Cardie 2009 – counts hits and takes into account negation) and assigned to each content word.

→ Baseline unigram model pretty consistently outperforms baseline sentiment model (though I wonder if this would be different in my case, where I am not trying to differentiate stances but just detect a stance at all).

**Contrasting opposing views of news articles on contentious issues**

*Park et al, 2011.*

@inproceedings{park2011contrasting,

title={Contrasting opposing views of news articles on contentious issues},

author={Park, Souneil and Lee, Kyungsoon and Song, Junehwa},

booktitle={Proceedings of ACL 2011},

pages={340--349},

year = {2011}}

→ Disputant = person who takes a position and participates in contention around an issue. Opponent-based frame – focuses on the relation between disputants. Ex. ‘president office vs. opposition parties’.

→ Less relevant to my work but still an interesting paper/idea.

**Recognizing arguing subjectivity and argument tags**

*Conrad et al, 2012.*

@inproceedings{conrad2012recognizing,

title={Recognizing arguing subjectivity and argument tags},

author={Conrad, Alexander and Wiebe, Janyce and Hwa, Rebecca},

booktitle={Proceedings of ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics},

pages={80--88},

year = {2012}}

→ Mentioned: less precise words like “almost” - indicators of subjectivity?

→ Their task is twofold: identify expressions of arguing subjectivity, and label each one with an appropriate argument tag.

→ constructed their own labeled dataset

→ focus is on sentence-level argument detection rather than document-level stance classification

→ discourse relations: could (for example) the presence of concessions indicate neutrality/non-subjectivity in terms of political partisanship?

→ use a distributional model to capture less frequent words that are similar to other more frequent words

→ on the sentence level, subjectivity classification benefits from additional context (so preceding and following sentences)

→ features: unigrams, sentiment (2 binary features indicating pos or neg), 15 binary features for kinds of discourse relationships, and distributional model features – for top 5 most similar and top 10 for each noun and verb. This last set of features plays the largest role in improving performance.

**Cats rule and dogs drool!: Classifying stance in online debate**

*Anand et al, 2011.*

@inproceedings{anand2011classifying,

title={Cats rule and dogs drool!: Classifying stance in online debate},

author={Anand, Pranav and Walker, Marilyn and Abbott, Rob and Fox Tree, Jean E. and Bowmani, Robeson and Minor, Michael},

booktitle={Proceedings of ACL-HLT 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis},

pages={1--9}.

year = {2011}}

→ long term goal is to understand dialogic and discourse structure of these arguments

→ they are looking specifically to identify the agreement/disagreement relations – not really my task

→ Naive Bayes classifier with features: post info, unigrams, bigrams, cue words (initial uni/bi/trigrams), repeated punctuation (?? !! ?!), dependency tags, LIWC (Linguistics Inquiry Word Count tool) features such as words per sentence/pronouns/pos and neg emotion words, POS tags of head words (??) called generalized dependencies, opinion dependencies (dep/gen dep features for MPQA opinion words), context features

→ over all topics, it was difficult to beat the unigram baseline

→ both kinds of dependency features showed an overall lack of impact

**News reporting bias detection prototype**

*Leban et al, 2014.*

@inproceedings{leban2014news,

title={News reporting bias detection prototype},

author={Leban, Gregor and Košmerlj, Aljaž and Belyaeva, Evgenia and Fortuna, Blaž},

year = {2014}}

→ US news sources mostly source Associated Press

→ Tabloid outlets show longer headlines shorter articles, more colorful language using adjectives and adverbs

→ Broad dataset spanning the whole world

→ types of bias covered: article length differences, grammatical differences, readability differences, geographical bias, topic coverage bias, speed of reporting bias, newswire citation bias, similarity in coverage of events, content similarity

→ Article length difference seems to vary among publishers, which is the opposite of what I want

→ Overall this survey seems to focus on differences between specific publishers instead of more general differences between types of news sources or bias itself. They used features like adjectives, adverbs, pronouns etc and found measurable differences, but this serves to me as mostly a warning to make sure that I am not learning publisher-specific writing features.

**Online News Media Bias Analysis using an LDA-NLP Approach**

*Doumit and Minai, 2012.*

@inproceedings{doumit2012online,

title={Online News Media Bias Analysis using an LDA-NLP Approach},

author={Doumit, Sarjoun and Minai, Ali},

year = {2012}}

→ More Latent Dirichlet Allocations. This also has a focus on different agents/writers.

→ Includes a lot of semantics – taking the semantics of original factual news, looks at how bias gets embedded into the “very fabric” - the semantic description – of the news.

→ Not really an approach I think I can replicate – or results that I want to – but the results in extraction are very interesting.

**Do media companies drive bias? Using sentiment analysis to measure media bias in newspaper tweets**

*Thomsen, 2018.*

@inproceedings{thomsen2018media,

title={Do media companies drive bias? Using sentiment analysis to measure media bias in newspaper tweets},

author={Thomsen, Taylor},

year = {2018}}

→ Takes an already existing sentiment model and runs it on GOP/Dem-related tweets from different news providers – so focuses (a) on the publishers and (b) on differentiating parties, not to mention (c) not creating their own model doesn’t give me very much to learn from here.

**Exploration of classifying sentence bias in news articles with machine learning models**

*Bellows 2018*

@inproceedings{bellows2018classify,

title={Exploration of classifying sentence bias in news articles with machine learning models},

author={Bellows, Martha},

year = {2018}}

→ Overview of what is being done in the field currently.

→ Different embeddings considered: word2vec, GloVe, fastText

→ Models considered: SVM, NN, CNN, RNN, with results showing no front-runner

→ “Bias at the news outlet level usually entails what pieces the outlet decides to cover and how articles are featured on their websites and newspapers. At the article level, it is somewhat similar. What perspectives are used, the title, and which photographs are chosen can all contribute to the article’s bias. At the sentence level, the syntax and semantics plays an important role in contributing to bias. Finally, the word level clearly is based on what words are chosen.” their focus is sentence level bias – not including bias of selection/omission

→ Lots of ways to do word embeddings: count vectors, term frequency, co-occurrence matrix, continuous bag of words, term frequency-inverse document frequency, skip-gram, etc.

→ Author collected articles, separated into sentences. “Each date was then separated out by news source and 150 sentences were randomly subsampled without replacement.” → about 1050 sentences taken for each date. Combined, shuffled, saved. Manually inspected all sentences. 876 per date in the end.

→ Made gold standard sentences by taking 500 sentences, ending up with 374 after manual inspection/removing.

→ All batches submitted to Amazon’s Mechanical Turk, then filtered to those which received a unanimous label. Overall, 4754 sentences with a total of 13140 annotations for regular and 1870 with 15010 for gold. Used majority voting and culled more after that for a final sentence count of 2143.

→ Project had only two labels: biased and unbiased.

→ word2vec, GloVe, fastText. But first, preprocessing: lowercasing, removing non-alphanumeric characters, removing stopwords, tokenized.

→ Gives best resulting hyperparameters for each model (SVM, NN, CNN, RNN). RNN was actually LSTM

→ Baseline was TF-IDF + Naive Bayes.

→ None stand out above the others, although they all outperformed the baseline. TF-IDF/NB = 68%, all other models between 74 and 78%.

→ Future research: sentence embedding? Doc2vec, skip-thought vectors, weighted word vector technique (this is in her references).

**Large-scale Sentiment Analysis for News and Blogs**

*Godbole et al, 2007*.

@inproceedings{godbole2007largescale,

title={Large-scale Sentiment Analysis for News and Blogs},

author={Godbole, Marta and Srinivasaiah, Manjunath and Skiena, Steven},

booktitle={Proceedings of International Conference on Web and Social Media}

year = {2007}}

→ Goal: development of a system to return pos/neg analysis of news.

→ Defined separate lexicons for each of seven sentiment dimensions (general, health, crime, sports, business, politics, media).

→ “We developed an algorithm to expand small dimension sets of seed sentiment words into full lexicons.” Uses recursion and path-following to expand the lexicons. They found a high correlation of their algorithmically-generated lexicons to past manually-curated ones.

→ When scoring sentiment data, they account for modifiers and negations next to their lexicon words.

→ They aggregate sentiment data over all articles to evaluate ‘world\_polarity’ for an entity for the whole time period, and ‘entity\_polarity’ for an entity on that particular day. They also evaluate ‘world\_subjectivity’ and ‘entity\_subjectivity’ for all news. Then they provide lists of individuals or news organizations which are particularly positive or negative.

**Recognizing contextual polarity in phrase-level sentiment analysis**

*Wilson et al, 2005.*

@inproceedings{wilson2005recognizing,

title={Recognizing contextual polarity in phrase-level sentiment analysis},

author={Wilson, Theresa and Wiebe, Janyce and Hoffman, Paul},

year = {2005}}

→ Approach to phrase-level sentiment analysis: differ between neutral or polar, and then for polar between negative and positive.

→ Two-step process: starts with a dictionary of clues marked with prior polarity. Extracts phrases with instances of those clues from a corpus. The phrases are marked as neutral or polar. Then, in the second step, polarity is disambiguated (positive, negative, both, neutral).

→ MPQA Opinion Corpus

→ They found that there were many examples of the subjectivity clues appearing in contexts that did not share the prior polarity of the clues, indicating that context is crucial.

→ Generally a strict rule-based system in the second part of the process, finding modifiers and negators and adverbs in relevant locations in order to classify the polarity of the phrase.

→ Note to self: I used this database significantly in the NLP final project and if I decide to use it here I can look back/possibly reuse some code.

**Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews**

*Turney, 2002.*

@inproceedings{turney2002semantic,

title={Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews},

author={Recasens, Marta and Danescu-Niculescu-Mizil, Cristian and Jurafsky, Dan},

booktitle={Proceedings of ACL 2002},

pages={417--424},

year = {2002}}

→ Fairly simplistic lexicon-based pos/neg review classification.