Unveiling Hidden Intentions

Gerardo Ocampo Diaz and **Vincent Ng**
Human Language Technology Research Institute
University of Texas at Dallas
Richardson, TX 75083–0688
{godiaz,vince}@hlt.utdallas.edu

Abstract

Recent years have seen significant advances in machine perception, which have enabled AI systems to become grounded in the world. While AI systems can now "read" and "see", they still cannot read between the lines and see through the lens, unlike humans. We propose the novel task of *hidden message and intention identification*: given some perceptual input (i.e., a text, an image), the goal is to produce a short description of the message the input transmits and the hidden intention of its author, if any. Not only will a solution to this task enable machine perception technologies to reach the next level of complexity, but it will be an important step towards addressing a task that has recently received a lot of public attention, political manipulation in social media.

1 Introduction

In an increasingly online world, information reaches far and wide. Social media, forums, and messaging platforms have enabled citizens to speak their mind and facilitated the rapid proliferation of a wide range of content: from personal anecdotes, videos, and memes to political views and full-blown opinion and news articles from other users and third parties. The ubiquity and reach of online platforms empowers individuals to share information and unite behind common goals. At the same time, however, it exposes them to complexly orchestrated campaigns of manipulation that aim to influence public opinion and achieve political goals. Such objectives might include inducing fear and distrust, smearing or discrediting, raising support for political entities, creating and reinforcing division and discord in societies, and generating action or inaction among the population.

In recent years, known instances of such political manipulation have increased at an alarming rate. Noteworthy examples include ISIS' use of social media for propaganda and recruitment (Farwell 2014), state-sponsored interference in the 2016 U.S. presidential election (DiResta et al. 2018) (which included, among other things, spreading false information to prevent people from voting and disseminating content attacking the presidential candidates), public opinion manipulation through Twitter bots in Venezuela (Forelle et al. 2015) and crowdsourcing in Vietnam (Pham 2013), and, more recently, the dissemination of pro-government and anti-activist memes and posts on Twitter and Facebook during the 2019 Hong Kong protests (Gleicher 2019; Safety 2019). Research and journalistic evidence suggests that the popularity of engaging in such campaigns is only increasing worldwide (Bradshaw and Howard 2018), and that political manipulation in social media is carried out not only by foreign but also domestic parties (Barrett 2019; WNYC Studios 2019).

Several strategies prevail in these political manipulation campaigns. A common strategy is *framing*, which refers to the way the media presents information to the audience with the goal of influencing the choices they make about how to process that information. As an example, consider the recent protests in Hong Kong. A journalist who focused on reporting the police's overuse of weapons in curbing the protests and how many protesters were injured as a result will likely trigger a negative sentiment towards the police. In contrast, a journalist who focused on reporting how violent the protesters were (e.g., throwing bricks at the police, setting fire in the street, preventing tourists from entering the airport's security area) will likely trigger a negative sentiment towards the protesters. More sophisticated campaigns may involve a combination of this and other tactics.

How can artificial intelligence (AI) be used to counter political manipulation? We propose to use AI to identify the hidden messages behind a perceptual input, be it an image, a video, a speech, or a text document, along with the intention or goal of the author behind it. Specifically, we define the novel task of *hidden message and intention identification* as follows. Given a perceptual input, the goal is to automatically generate a short description (e.g., a phrase, a clause, or a sentence) of the message(s) conveyed by the input, as well as a separate description of the intention(s) that exists behind each message. Note that the surface message and intention of an author are different from his/her hidden message and intentions. Consider again the example on the Hong Kong protests. Regardless of how the protests were framed, the (typical) surface intention of the journalists is to *report on the ongoing protest events in Hong Kong*, with no particular message being relayed by journalists other than the facts

reported in their articles. However, if a report focuses on the violent acts carried out by the protesters, then the report could implicitly convey messages such as *The protesters are violent*, *The protesters are criminals*, or *The protesters are dangerous*, which would inevitably sway the public's opinion on the protests. From this, one could say that the author of said report could have the intention of *Discrediting the protesters* or *Supporting the government*.

AI could therefore alleviate political manipulation by allowing the public to understand the hidden messages and intentions of an author and using this additional information to determine how trustworthy a given source of information is. Note that our proposed task is different from intent recognition in dialogue systems, where the goal is to recognize the surface intent underlying an utterance (Larson et al. 2019) (e.g., is it a "request" or a "confirmation"?).

While we use political manipulation as an example to motivate the task of hidden intention and message identification, being able to unveil an author's hidden intentions and identify his/her messages has broader implications in AI. While recent advances in computer vision and speech and language processing have enabled AI systems to "see" the objects in an image and "read" the words in a document, these systems still cannot see through the lens and read between the lines. In other words, they are still far from being able to understand the message(s) the author intends to convey in an image or a document, unlike humans. Being able to understand these hidden intentions and messages like humans requires that a machine achieve a deeper level of understanding of a perceptual input, enabling machine perception to get one step closer to human perception. Oftentimes, world knowledge (e.g., that protesters and the government are antagonists in the 2019 Hong Kong protests, and that support for police implies support for government) is needed to achieve a deep understanding of a perceptual input, but acquiring and incorporating world knowledge into AI systems remains a key research challenge.

While our task involves unveiling the hidden intention(s) of an author of a text or an image, as well as identifying the hidden message(s) it conveys, no one other than the author can say for sure what his/her real hidden intention(s) and message(s) are. Hence, in reality, our task involves inferring the hidden message(s) and intention(s) of an author *as perceived by the human audience*. Note that this realistic version of the task aligns perfectly with our goal of making machine perception one step closer to human perception.

The rest of the paper is organized as follows. Sections 2 and 3 discuss the challenges involved in identifying an author's intention(s) and message(s) from two major kinds of perceptual input: memes/images and text. Section 4 outlines the first step towards automatic identification of hidden intentions and messages. Finally, we present concluding remarks in Section 5.

2 Memes and Images

Memes — user-created combinations of pictures and images overlaid with text — frequently flood Facebook, Twitter, Instagram, and Reddit feeds, and are also frequently shared directly through instant messaging on social circles. These relatively simple pieces of media are easy to make, tend to be catchy, and have a lower cognitive cost when compared to traditional pieces of media like text and videos, and are therefore easily viralized online. Although memes were first deployed mainly for comedic purposes, nowadays users use them to express opinions on a variety of topics, including politics, events, religion, personal values, and legislation. While a lot of recent research has focused on detecting malicious content in text (Hanselowski et al. 2018; Kiesel et al. 2019), there is suggestive evidence that images and memes are a key tool for the transmission of ideological and political content (Barrett 2019).

Adding to the well-known ad campaign carried out by the Internet Research Agency (IRA) targeting American politics in 2016, a recent report commissioned by the U.S. Senate Intelligence Committee (DiResta et al. 2018) points out that during the 2016 interference campaign on the U.S. elections, the IRA enjoyed more user engagement on Instagram (an image-centric platform) than on any other social platforms. Researchers also found that, between 2015 and 2018, there were 187 million user engagements with IRA material on Instagram. Recently, Facebook and Twitter have released statements detailing the detection of pro-government memes in their respective platforms in the wake of the 2019 Hong Kong protests (Safety 2019; Gleicher 2019).

Challenges Identifying a meme's messages and intentions is not a trivial task. The role of text and images is not the same across memes: text can be used to show quotes, label the entities involved in a meme, explicitly or implicitly state the message of a meme, or put images in context. Understanding memes (and the messages behind them) requires world knowledge such as: a) identifying specific entities present in an image, b) relating the visual structure of a meme to established types of memes, and c) identifying symbols, signs or figures and understanding how they shape the message of a meme. Figure 1 shows three memes, whose corresponding hidden messages and intentions can be found in the caption. For example, the leftmost meme never explicitly names specific entities, yet the images can be inferred to refer to the Hong Kong protesters. The second meme shows a Facebook ad with images of religious figures such as Satan and Jesus. Here, text is used to label a dialogue between the figures of Satan and Jesus, and an association is made between Clinton and Satan based on the dialogue. However, understanding that this is bad requires the knowledge that Satan and Jesus are figures typically associated with strongly negative and positive valences respectively. Furthermore, knowing that Jesus and Satan are rival figures, it should not be difficult to see that the author tries to associate Trump with Jesus. The rightmost meme is quite different from the other two memes because the imagery of kermit the frog has nothing to do with the topic of the meme — kermit the frog drinking a beverage indicates that this meme is indirectly criticizing something or someone. Here, text content is used to explicitly state the subject matter that "democrats are going to plow on until they lose all", but the text also states that "but that is none of my business". While at face value the text might indicate that the meme is merely pointing a fact out, the imagery of kermit the frog helps clar-

Figure 1: Example memes. The first meme conveys hidden messages such as *The protesters are dangerous/violent/evil*, with the hidden intention of *Discrediting protesters/Legitimizing the government's anti-protest actions*. The second meme conveys hidden messages through associations: Satan is associated with Hillary Clinton, while Jesus is implicitly associated with Trump, reflecting a negative stance on Hillary and a positive stance on Trump; the corresponding intentions would be *Convince people to vote for Donald Trump/Slander Hillary Clinton/Support Donald Trump*. The third meme conveys the hidden message that *Democrats are making bad decisions*, with the hidden intention of *Criticizing Democrats*.

ify that the meme is indeed criticizing Democrats.

Related work While computer vision researchers have long worked on tasks such as object recognition and image captioning, systems mostly focus on describing content at face value rather than reasoning over the meaning of the content. Popular datasets such as COCO (Lin et al. 2014), ImageNet (Deng et al. 2009), and Conceptual Captions (Sharma et al. 2018) are mainly concerned with recognizing objects in photographs and providing a description of the content in pictures (e.g., "a dog running on a beach"). Such systems are definitely useful for understanding memes. For example, it might be useful to know that the second meme in Figure 1 depicts "Jesus and Satan arm-wrestling", as this helps identify key figures and a context of conflict. However, as we explained above, they are not direct solutions to the problem. Past research on automatic meme captioning (Wang and Wen 2015) is closer to image captioning in this sense.

3 Text

Hidden messages and intentions on text content exist on a spectrum: on one hand, rants and opinion pieces may use inflammatory language and explicitly support or attack specific ideas, values or entities; on the other, content creators may use subtle language cues, cherry-pick facts, or frame information in ways that might transmit messages to the reader implicitly. For example, if an author wishes to manipulate readers into being more accepting of a particular political party's campaign to establish stricter immigration laws, one could try to impress upon readers the idea that *immigrants are dangerous* (hidden message), or simply frame concepts related to immigration law or said political party's decisions in a positive manner, either by implying positive associations (for example, associating the party with well-respected figures), or by implying that *the party knows what it is doing* (hidden message). By doing this, the author might achieve his/her intention of *Eliciting support for the party's campaign*. As another example, consider the following excerpt from an article on the seizure of a fentanyl shipment by Mexican authorities in 2019 (100percentfedup.com 2019):

> Today, the Mexican Navy stopped a massive shipment of fentanyl coming from Shanghai, China that was headed to the notorious Sinaloa Cartel. Ryan Saavedra of the Daily Wire did some math and discovered the Mexican Navy apprehended enough fentanyl to kill over 7 billion people.
>
> President Trump has been very vocal about his commitment to stop the opioid crisis in America.
>
> DEA Special Agent Clyde E. Shelley has identified fentanyl as the number one threat causing the opioid epidemic in America.
>
> The Mexican drug cartels have been sneaking illegal drugs like fentanyl, across our southern border for decades, while Democrat lawmakers fight to keep our borders open.

The article starts by introducing an objective event: the seizure of a fentanyl shipment. It then provides estimates *.. enough fentanyl to kill over 7 billion people* to develop a sense of urgency in readers. It then jumps to state that President Trump has been vocal about his commitment to stop the opioid crisis in America, an important issue for U.S. society which has garnered significant attention recently. Note that this jump implies a negative stance on China with the opioid crisis in the U.S.. After a (true) quote meant to raise a sense of urgency again, the article points out that Mexican cartels have been sneaking illegal drugs through the southern U.S. border while "Democrat lawmakers fight to keep our borders open", this last part implies an association between keeping

borders open and drug trafficking as well as with the opioid crisis, and is phrased in a way that seems to imply that *Democrat lawmakers are ignoring the opioid crisis* at best and that *Democrat lawmakers are confabulating to keep the opioid crisis going* at worst.

Although this appears to be an informative news article, it effectively transmits several hidden messages. Specifically, it intends to depict President Trump as a positive figure, who is committed to solving a problem for U.S. citizens, while depicting Democrats in a negative light, possibly holding China accountable for the opioid crisis in the U.S., and implying that keeping borders open is making the problem worse. Further, in light of recent events this article could be seen as an attempt to covertly legitimize President Trump's trade war with China and immigration policy.

Challenges For systems to be able to extract the aforementioned hidden messages, they need to a) understand the messages implied by rhetoric (e.g., how jumping from fentanyl and China to the opioid crisis establishes a connection between them), b) acquire world knowledge (such that the opioid crisis is an important problem for Americans), and c) establish connections between the entities, events and actions in the text and the policies, entities, or events not mentioned (such as drawing a connection between immigration policy and keeping borders open). In addition, systems should be able to understand implicit expressions of sentiment (like the negative sentiment implied on Democrats). It would perhaps also be useful to have systems perform *fact verification* (Thorne et al. 2018), since false claims can be indicative of manipulative behavior. However, as this example shows, content creators need not make false claims to spin content in a way that aligns with their political interests.

Related work Research relevant to this task includes fine-grained sentiment analysis (Pang and Lee 2008; Liu 2015), where the goal is to determine the sentiment (positive/negative/neutral) expressed towards a particular target/entity. Although fine-grained sentiment analysis is a well-researched task, the majority of research on the area is focused on *explicit* sentiment rather than *implicit* sentiment, which may more likely appear in well-crafted manipulative text content (compare, for example, "this phone is amazing!" with "this phone fits in my pocket easily"). Another related task is stance detection (Agrawal et al. 2003; Thomas, Pang, and Lee 2006; Balahur, Kozareva, and Montoyo 2009; Murakami and Raymond 2010; Somasundaran and Wiebe 2010; Wang and Rosé 2010; Anand et al. 2011; Biran and Rambow 2011; Hasan and Ng 2013; Mohammad et al. 2016; Ruder et al. 2018; Sun et al. 2018). Stance detection aims to infer the author's stance (*for* or *against*) towards a particular topic based on what s/he wrote and can therefore be viewed as a special case of hidden intention identification. A somewhat relevant task is argument persuasiveness scoring (Persing and Ng 2015; Habernal and Gurevych 2016a; 2016b; Wei, Liu, and Li 2016; Stab and Gurevych 2017; Eger, Daxenberger, and Gurevych 2017; Persing and Ng 2017), where the goal is to determine how persuasive an argument is. Typically, the persuasiveness of an argument can affect the ease with which the reader can infer the author's hidden intention(s) and message(s).

4 The First Step: Corpus Construction

Given the challenges in hidden intention and message identification, how can we address this task? Since we want a machine to learn how humans perceive hidden intentions and messages, we propose to employ machine learning (ML), particularly deep learning, given their successful applications to vision and natural language processing (NLP) tasks.

The question, then, is: how can we create an annotated corpus of training instances, each of which is a perceptual input paired with the corresponding (short) description of its author's hidden intention(s) and message(s)? We can rely on a handful of human experts. However, this will unlikely yield a sufficiently large corpus. While a large corpus is generally needed for successful application of deep learning, we believe that a large training corpus is essential for a task as complex as hidden intention and message identification, as a large corpus will likely encode a lot of the world knowledge needed for the task (see Sections 2 and 3). One way to automatically obtain a large amount of training data is *distant supervision* (Mintz et al. 2009). For instance, stances taken by Fox News can be assumed to be pro-Republican. While we believe that distant supervision can help to a certain extent, it probably cannot cover different kinds of intentions and messages, particularly the more complex ones. We believe the most promising approach to create a large annotated training set in a reasonably short period of time is *crowdsourcing* (Brabham 2013), where we hire workers on crowdsourcing platforms such as Amazon Mechanical Turk to obtain multiple annotations for each perceptual input.

Given a text or an image, the three annotation tasks are to (1) determine whether the text or image conveys hidden messages, and if yes, (2) what these messages are and (3) what the intention(s) of the author are. A natural question is: will different workers always agree on all or even one of these tasks? The answer is no. If the answer were yes, then the political manipulation problem would not exist: people are manipulated in part because not all of them understand these hidden intentions and messages.

Moreover, the same intentions and messages can be lexically realized in different ways. For instance, in the Hong Kong protest example, "pro-government" and "pro-police" can be viewed as similar, if not equivalent, intentions. Finally, it is possible that workers produce different descriptions because the given perceptual input may be associated with more than one intention or message. Hence, we believe that one should not be overly concerned that workers produce different descriptions for the same training instance.

A relevant question, then, is: what if the workers generate contradictory descriptions for the same training instance? We believe that this is unlikely. While workers can biased by their beliefs and cultural context, this would most likely translate into some workers being more sensitive towards certain content or ideas than others (e.g., a strongly religious worker might be more sensitive to content which criticizes religion than an agnostic worker). We would not expect workers to interpret pro-Democrat content as pro-Republican, unless they are confused between Democratic

or Republican agendas and values or have no such knowledge at all. Noisy labels produced as a result of this kind of situation can be easily singled out as long as the majority of the labels are consistent with each other.

The next question is: can all workers fail to identify hidden messages and intentions even if they exist? While the examples discussed earlier provide suggestive evidence that it is possible to identify hidden intentions and messages, nothing guarantees that they can always be identified, especially if they are more subtle than those in the political domain. While this remains a research question, we may reduce the likelihood that this happens by hiring workers with varied backgrounds (e.g., they are in different parts of the world with different education levels).

Another relevant question is: why formulate this task as a *generation* task, where the goal is to generate a short natural language description of the message(s) and the intent(s), rather than as a *classification* task, where the goal is to identify the message(s) and intent(s) from a predefined set? To formulate this as a classification task, one will have to define a taxonomy of intentions and messages. Not only may the number of intents and messages be large, but the they may be domain-dependent and hence have to be designed for each domain of interest. Even if we manage to design this taxonomy, inter-annotator agreement among the crowdsourced workers could be low if the taxonomy is sufficiently complex, as it is typically unrealistic to expect these workers to understand the meaning associated with each intent in a complex taxonomy. Given the successful application of recurrent neural networks (Jain and Medsker 1999) to natural language generation tasks, we believe that it is a good idea to cast it as a generation task.

5 Concluding Remarks

We introduced the novel task of hidden intention and message identification and discussed its possible application to *memes* and *text content*, along with the challenges inherent to these types of media, such as the interpretation of visual structure, entity recognition, the interpretation of symbols and figures for memes and implicit sentiment detection, the interpretation of prose, and the automatic acquisition of world knowledge and entity associations for text. Further, although not in the scope of this paper, this task could be expanded to include other types of media such as *videos*; specifically political shows and recorded public statements, where emotion recognition from images (You et al. 2015; 2016), speech cues, facial expressions and bodily gestures (Jiang, Xu, and Xue 2014; Pereira et al. 2016) could be useful in identifying hidden messages and intentions. We believe that the task has broader implications in AI, as the ability to understand hidden messages and intentions like humans requires that a machine achieve a deeper level of understanding of a perceptual input, enabling machine perception to get one step closer to human perception.

Work on this task could lead to the development of useful tools for researchers, governments, and citizens that could be applied to problems such as: (1) detection of bias in news articles and media outlets; (2) identification of sponsored propaganda and manipulation campaigns; (3) discovery and study of political echo chambers and radicalization campaigns; and (4) detection of fear-inducing content meant to destabilize the economy. On the other hand, it is also easy to see how this technology could be weaponized to hamper free-speech, with journalists and sites being automatically flagged based on the messages expressed by their content and those who consume said content being profiled.

Given the complexity the task and the problems inherent to building a dataset for it, we recommend that a shared task on this problem be organized with the goal of identifying interested researchers and jumpstarting work on it.

Acknowledgments

We thank the three anonymous reviewers for their comments on an earlier draft of the paper. This work was supported in part by NSF Grants IIS-1528037 and CCF-1848608.

References

100percentfedup.com. 2019. Breaking: Massive fentanyl shipment from China seized in mexicoenough fentanyl to kill estimated 7 billion. https://100percentfedup.com/breaking-massive-bust-on-mexican-border-enough-fentanyl-to-kill-7-billion-peopleshipment-reportedly-came-from-china/.

Agrawal, R.; Rajagopalan, S.; Srikant, R.; and Xu, Y. 2003. Mining newsgroups using networks arising from social behavior. In *WWW*, 529–535.

Anand, P.; Walker, M.; Abbott, R.; Fox Tree, J. E.; Bowmani, R.; and Minor, M. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 1–9.

Balahur, A.; Kozareva, Z.; and Montoyo, A. 2009. Determining the polarity and source of opinions expressed in political debates. In *CICLing*, 468–480.

Barrett, P. M. 2019. Disinformation and the 2020 election: How the social media industry should prepare. Technical report, NYU Stern Center for Business and Human Rights.

Biran, O., and Rambow, O. 2011. Identifying justifications in written dialogs. In *ICSC*, 162–168.

Brabham, D. C. 2013. *Crowdsourcing*. The MIT Press.

Bradshaw, S., and Howard, P. N. 2018. Challenging truth and trust: A global inventory of organized social media manipulation. Technical report, Working Paper 2018.1. Oxford, UK: The Computational Propaganda Project.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Li, F.-F. 2009. ImageNet: A large-scale hierarchical image database. In *CVPR*, 248–255.

DiResta, R.; Shaffer, K.; Ruppel, B.; Sullivan, D.; Matney, R.; Fox, R.; Albright, J.; and Johnson, B. 2018. The tactics & tropes of the Internet Research Agency. https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf.

Eger, S.; Daxenberger, J.; and Gurevych, I. 2017. Neural end-to-end learning for computational argumentation mining. In *ACL*, 11–22.

Farwell, J. P. 2014. The media strategy of ISIS. *Survival* 56(6):49–55.

Forelle, M.; Howard, P.; Monroy-Hernández, A.; and Savage, S. 2015. Political bots and the manipulation of public opinion in Venezuela. *arXiv preprint arXiv:1507.07109*.

Gleicher, N. 2019. Removing coordinated inauthentic behavior from China. https://newsroom.fb.com/news/2019/08/removing-cib-china/.

Habernal, I., and Gurevych, I. 2016a. What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation. In *EMNLP*, 1214–1223.

Habernal, I., and Gurevych, I. 2016b. Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM. In *ACL*, 1589–1599.

Hanselowski, A.; Avinesh, P.; Schiller, B.; Caspelherr, F.; Chaudhuri, D.; Meyer, C. M.; and Gurevych, I. 2018. A retrospective analysis of the fake news challenge stance-detection task. In *COLING*, 1859–1874.

Hasan, K. S., and Ng, V. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *IJCNLP*, 1348–1356.

Jain, L. C., and Medsker, L. R. 1999. *Recurrent Neural Networks: Design and Applications*. Boca Raton, FL, USA: CRC Press, Inc.

Jiang, Y.-G.; Xu, B.; and Xue, X. 2014. Predicting emotions in user-generated videos. In *AAAI*, 73–79.

Kiesel, J.; Mestre, M.; Shukla, R.; Vincent, E.; Adineh, P.; Corney, D.; Stein, B.; and Potthast, M. 2019. SemEval-2019 Task 4: Hyperpartisan news detection. In *SemEval*.

Larson, S.; Mahendran, A.; Peper, J. J.; Clarke, C.; Lee, A.; Hill, P.; Kummerfeld, J. K.; Leach, K.; Laurenzano, M. A.; Tang, L.; and Mars, J. 2019. An evaluation dataset for intent classification and out-of-scope prediction. In *EMNLP-IJCNLP*, 1311–1316.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common objects in context. In *ECCV*, 740–755.

Liu, B. 2015. *Opinion Mining, Sentiment Analysis, and Opinion Spam Detection*. Cambridge University Press.

Mintz, M.; Bills, S.; Snow, R.; and Jurafsky, D. 2009. Distant supervision for relation extraction without labeled data. In *ACL-IJCNLP*, 1003–1011.

Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; and Cherry, C. 2016. SemEval-2016 Task 6: Detecting stance in tweets. In *SemEval*, 31–41.

Murakami, A., and Raymond, R. 2010. Support or oppose? Classifying positions in online debates from reply activities and opinion expressions. In *COLING: Posters*, 869–875.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1-2):1–135.

Pereira, M. H. R.; Pádua, F. L. C.; Pereira, A. C. M.; Benevenuto, F.; and Dalip, D. H. 2016. Fusing audio, textual, and visual features for sentiment analysis of news videos. In *ICWSM*, 659–662.

Persing, I., and Ng, V. 2015. Modeling argument strength in student essays. In *ACL-IJCNLP*, 543–552.

Persing, I., and Ng, V. 2017. Why can't you convince me? Modeling weaknesses in unpersuasive arguments. In *IJCAI*, 4082–4088.

Pham, N. 2013. Vietnam admits deploying bloggers to support government. https://www.bbc.com/news/world-asia-20982985.

Ruder, S.; Glover, J.; Mehrabani, A.; and Ghaffari, P. 2018. 360° stance detection. In *NAACL HLT: Demonstrations*, 31–35.

Safety, T. 2019. Information operations directed at Hong Kong. https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html.

Sharma, P.; Ding, N.; Goodman, S.; and Soricut, R. 2018. Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning. In *ACL*, 2556–2565.

Somasundaran, S., and Wiebe, J. 2010. Recognizing stances in ideological on-line debates. In *NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 116–124.

Stab, C., and Gurevych, I. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics* 43(3):619–659.

Sun, Q.; Wang, Z.; Zhu, Q.; and Zhou, G. 2018. Stance detection with hierarchical attention network. In *COLING*, 2399–2409.

Thomas, M.; Pang, B.; and Lee, L. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *EMNLP*, 327–335.

Thorne, J.; Vlachos, A.; Cocarascu, O.; Christodoulopoulos, C.; and Mittal, A. 2018. The fact extraction and VERification (FEVER) shared task. In *First Workshop on Fact Extraction and VERification*, 1–9.

Wang, Y.-C., and Rosé, C. P. 2010. Making conversational structure explicit: Identification of initiation-response pairs within online discussions. In *NAACL HLT*, 673–676.

Wang, W. Y., and Wen, M. 2015. I can has cheezburger? A nonparanormal approach to combining textual and visual information for predicting and generating popular meme descriptions. In *NAACL HLT*, 355–365.

Wei, Z.; Liu, Y.; and Li, Y. 2016. Is this post persuasive? Ranking argumentative comments in online forum. In *ACL: Short Papers*, 195–200.

WNYC Studios. 2019. A progressive activist defends his deceptive tactics. https://www.wnycstudios.org/story/progressive-activist-defends-his-deceptive-tactics.

You, Q.; Luo, J.; Jin, H.; and Yang, J. 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *AAAI*, 381–388.

You, Q.; Luo, J.; Jin, H.; and Yang, J. 2016. Building a large scale dataset for image emotion recognition: The fine print and the benchmark. In *AAAI*, 308–314.