# Argumentation Mining In Twitter: A Study Of Controversial Topics

**Aseel Addawood**
Illinois Informatics Institute
University of Illinois at Urbana-Champaign
**aaddaw2@illinois.edu**

## Abstract

In recent years, social media has revolutionized how people communicate and share information. Besides connecting with friends, one important function of social media is to share opinions with others. Microblogging sites such as Twitter have often provided an online forum for social activism. When users debate controversial topics on social media, they typically share different types of evidence to support their claims. Classifying these types of evidence can provide an estimate for how adequately the arguments have been supported. In this paper, we first introduce a manually built gold standard dataset of 3,000 tweets related to the recent FBI and Apple encryption debate. We develop a framework for automatically classifying six evidence types typically used on Twitter to discuss this debate. Our findings show that a Support Vector Machine (SVM) classifier trained with N-grams and additional features can capture different ways of representing evidence on Twitter, exhibiting significant improvements over the unigram baseline and achieving a F1 macro-average of 82.8%. Recommendations and insights into this task are also shared in this paper to support others working on similar tasks.

*This work is part of a published work in the 3rd Workshop on Argument Mining, ACL 2016. (Addawood & Bashir, 2016)*

## 1 Introduction

Social media has grown dramatically over the last decade. Researchers turn to social media via online posts as a source of information to explain many aspects of human experience (Gruzd & Goertzen, 2013). Due to the textual nature of online users' self-disclosures of their opinions and views, social media platforms present a unique opportunity for further analysis of shared content, including the means by which controversial topics are argued. On social media sites, especially Twitter, user text contains arguments with inappropriate or missing justifications—a rhetorical habit we do not usually encounter in professional writing. One way to handle such faulty arguments is to simply disregard them and focus on extracting arguments containing proper support (Cabrio & Villata, 2012; Villalba & Saint-Dizier, 2012). However, in some cases, what seems to be missing evidence is an unfamiliar or different type of evidence. Thus, recognizing the appropriate type of evidence can be useful in assessing the viability of users' supporting information and, in turn, the strength of their whole argument.

The motivation for this study is to facilitate online users' search for information concerning controversial topics. Social media users are often faced with information overloads about any given topic, and understanding positions and arguments in online debates can potentially help users formulate stronger opinions on controversial issues as well as foster better personal and group decision-making (Freeley & Steinberg, 2013). Analyzing argumentation from a computational linguistics point of view has led recently to a new field called argumentation mining, which examines the ways in which people disagree, debate, and form a consensus.

Argumentation mining focuses on identifying and extracting the argumentative structures of documents. In this study, we describe a novel and unique benchmark data set achieved through a simple argument model and elaborate on the associated annotation process. Unlike the classical

Table 1. Summary of evidence type classification results using one-vs.-all in %

| Feature Set | NEWS vs. All | | | BLOG vs. All | | | NO EVIDENCE vs. All | | | Macro Average F1 |
|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 | |
| Uni.(Baseline) | 76.8 | 74 | 73.9 | 67.3 | 64.4 | 63.5 | 78.5 | 68.7 | 65.6 | 67.6 |
| Basic Features | 84.2 | 81.3 | 81.3 | 85.2 | 83 | 82.9 | 80.1 | 75.5 | 74.4 | 79.5 |
| Psychometric Features | 62 | 61.7 | 57.9 | 64.6 | 63.7 | 63.5 | 59.2 | 58.9 | 58.6 | 60 |
| Linguistic Features | 65 | 65.3 | 64.2 | 69.1 | 69 | 69 | 63.1 | 62.6 | 62.4 | 65.2 |
| Twitter-Specific Features | 65.7 | 65.2 | 65 | 63.7 | 63.6 | 63.6 | 68.7 | 68.1 | 67.9 | 65.5 |
| All features | 84.4 | 84 | **84.1** | 86 | 85.2 | **85.2** | 79.3 | 79.3 | **79.3** | **82.8** |

Toulmin model (Toulmin, 2003), we search for a simple and robust argument structure comprised only of two components: a claim and associated supporting evidence.

Previous research has shown that a claim can be supported using different types of evidence (Rieke & Sillars, 1984). The annotation that is proposed in this paper is based on the type of evidence one uses to support a position in a given debate.

## 2 Method

This study uses Twitter as its main source of data. Crimson Hexagon (Etlinger, 2012) was used to collect tweets from January 1, 2016 to March 31, 2016. The tweets concerned the recent Apple/FBI encryption debate. The search criterion for this study looked for tweets that contained the word "encryption" anywhere in its text.

To perform argument extraction from a social media platform, we followed a two-step approach. The first step was to identify sentences containing an argument. The second step was to identify the evidence type found in those tweets classified as argumentative. Annotators were asked to annotate each tweet as either having or not having an argument based on the type of evidence used in the tweet. Two annotators were trained, and a three-iteration procedure was taken to ensure the validity of the annotation. A total of 3,000 tweets were annotated. These tweets were coded into one of seven evidence types: *News media account, Expert opinion, Blog post, Picture, Other, No evidence, Non-Argument.*

We proposed a set of features to characterize each type of evidence in our collection. We identify four types of features based on their scope. *Basic Features* refer to N-gram features, which rely on the word count (TF) for each given unigram or bigram that appears in the tweet. *Psychometric Features* refer to dictionary-based features. They are derived from linguistic inquiry and word count (LIWC) (Pennebaker, 1997; Pennebaker & Francis, 1996). *Linguistic Features* encompass four types of feature. The first is grammatical features. The second type is LIWC summary variables. The newest version of LIWC includes four new summary variables (analytical thinking, clout, authenticity, and emotional tone), which resemble "person-type" or personality measures. The third type is sentiment features. For the final type, subjectivity features, we used Wilson et al.'s (2005) subjectivity clue lexicon to identify the subjectivity type of tweets. *Twitter-Specific Features* refer to characteristics unique to the Twitter platform.

## 3 Results

Our first goal was to determine whether a tweet contained an argument. We used a binary classification task in which each tweet was classified as either argumentative or not. As a first step, we compared classifiers that have frequently been used in related work: Naïve Bayes, Support Vector Machines (SVM), and Decision Trees (J48). The best overall performance was achieved using SVM, which resulted in an 89.2% F1 score for all features as compared to the unigram model.

Our second goal was to perform evidence type classification. Results across the training techniques were comparable; the best results were again achieved using SVM, which resulted in a 78.6% F1 score by combining all features. In Table 1, we computed Precision, Recall, and F1 scores with respect to the top three evidence types, employing one-vs.-all classification problems for evaluation purposes. We chose these three evidence types since all other types were too small and could have led to biased sample data. The results show that the SVM classifier achieved a F1 macro-averaged score of 82.8%.

We performed feature analyses to investigate the most informative feature for each class. There are different features that work for each class. For example, Twitter-specific features such as title, word count, and the number of words per sentence are good indicators of the NEWS evidence type. One explanation for this is that people often include the title of a news article in the tweet with the URL, thereby engaging the Twitter-specific features more fully.

A combination of linguistic features and psychometric features best describe the NO EVIDENCE classification type. Furthermore, in contrast to blogs, users not using any evidence tend to express more positive emotions. This may imply that they are more confident about their opinions. There are, however, shared features used for both Blog and NO EVIDENCE types, such as the use of first-person singular and the colon. One explanation for this is that since blog posts are often written in a less formal, less evidence-based manner than news articles, they are comparable to tweets that lack sufficient argumentative support.

## 4 Conclusion

In this paper, we have presented a novel task for automatically classifying argumentation on social media for users discussing controversial topics, such as the recent FBI and Apple encryption debate. We classified six types of evidence people use in their tweets to support their arguments. This classification can help predict how arguments are supported. We have built a gold standard dataset of 3,000 tweets from the recent encryption debate. We find that Support Vector Machines (SVM) classifiers trained with N-grams and other features capture the different types of evidence used in social media and demonstrate significant improvement over the unigram baseline, achieving a macro-averaged F1 score of 82.8 %.

One consideration for future work is how to classify the stance of tweets by using machine learning techniques to understand a user's viewpoint and opinions about a debate. Another consideration for future work is to explore other evidence types that may not be presented in our data for so that they may be generalizable to other datasets.

## 5 Recommendations

While conducting this work, we encountered some challenges. This section highlights these challenges and how we handled them. One difficulty was the informal format of social media text. Social media text does not follow any guidelines or rules for the expression of opinions. Consequently, many messages contain improper syntax or spelling, which presents a significant for extracting meaning from them. However, social media text can provide great insight into public opinions, attitudes, and behaviors. Understanding public opinions and attitudes towards controversial topics may help scholars, law enforcement officials, and policymakers to develop better policies and guidelines.

Annotating tweets related to controversial topics such as the encryption debate requires annotators who not only understand the English language, including its informal cultures, but also understand the encryption debate as a whole. Another challenge of annotating the data was related to the language and structure of tweets, which tend to use informal and incoherent text. In addition, it is important to note that although our classification achieved a high score in the selected debate topic, these results may not be generalizable to other domains without further investigation.

Working with social media data such as Twitter always raises ethical concerns. These include the privacy and anonymity of online users, data ownership, and data security. Good ethical practice is essential for conducting social media research. However, the lack of relevant guidance and guidelines for conducting this type of research makes it reliant on the researchers' background and values, including compliance with rules and regulations established by the research community (i.e., the Institutional Review Board (IRB), the Health Insurance Portability and Accountability Act (HIPAA)), intellectual property law, copyright, and data/API term of use.

## References

Addawood, A. A., & Bashir, M. N. (2016). *"What is your evidence?" A study of controversial topics on social media*. In Proceedings of the 3rd Workshop on Argument Mining.

Cabrio, E., & Villata, S. (2012). *Combining textual entailment and argumentation theory for supporting online debates interactions.* Paper presented at the Proceedings of the 50th Annual

Meeting of the Association for Computational Linguistics: Short Papers-Volume 2.

Etlinger, S., & Amand, W. . (2012). Crimson Hexagon [Program documentation]. Retrieved from http://www.crimsonhexagon.com/wp-cotent/uploads/2012/02/CrimsonHexagon_Altimeter_Webinar_111611.pdf

Freeley, A. J., & Steinberg, D. L. (2013). *Argumentation and debate*: Cengage Learning.

Gruzd, A., & Goertzen, M. (2013). *Wired academia: Why social science scholars are using social media*. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences (HICSS).

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological science, 8*(3), 162-166.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion, 10*(6), 601-626.

Rieke, R. D., & Sillars, M. O. (1984). *Argumentation and the decision making process*: Addison-Wesley Longman.

Toulmin, S. E. (2003). *The uses of argument*: Cambridge university press.

Villalba, M. P. G., & Saint-Dizier, P. (2012). Some Facets of Argument Mining for Opinion Analysis. *COMMA, 245*, 23-34.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis*. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing.