Stance-based Argument Mining – Modeling Implicit Argumentation Using Stance

Michael Wojatzki
Language Technology Lab
University of Duisburg-Essen
Duisburg, Germany
michael.wojatzki@uni-due.de

Torsten Zesch
Language Technology Lab
University of Duisburg-Essen
Duisburg, Germany
torsten.zesch@uni-due.de

Abstract

A major remaining challenge in argument mining is implicitness. We propose to model implicit argumentation using explicit stances and the overall stance of a debate. Our evaluation on a social media corpus shows that our model (i) can be reliably annotated even on noisy data and (ii) has the potential to improve the performance of automated argument mining.

1 Introduction

Argument mining aims at an automated analysis of persuasive communication. One yet unsolved problem is that –especially in informal settings – argumentation is often done implicitly. For instance, in a debate on atheism, one may observe an utterance such as *Bible: infidels are going to hell* or even shorter *#JesusOrHell*. In the context of a debate about atheism, both utterances implicitly express the argument that the author is against atheism, because the bible says that this will result in a stay in hell after death. However, both claims are never explicitly mentioned.

Typically, models of argument mining assume that an argument consists of at least an explicit *claim* and a number of optional supporting structures such as *premises* (Palau and Moens, 2009; Peldszus and Stede, 2013). Figure 1a shows an example of the simplest manifestation of these *claim-premise* schemes. However, in implicit argumentation the claim usually needs to be inferred, as it is not explicitly expressed (see figure 1b). We argue, that in the absence of explicit information, the claim always corresponds to the overall stance in the debate in which the utterance is made. *Stance* can be defined as being in favor or against a defined target such as a controversial topic, e.g. *being in favor of atheism* or *being against it* (Mohammad et al., 2016). Thus, one may always

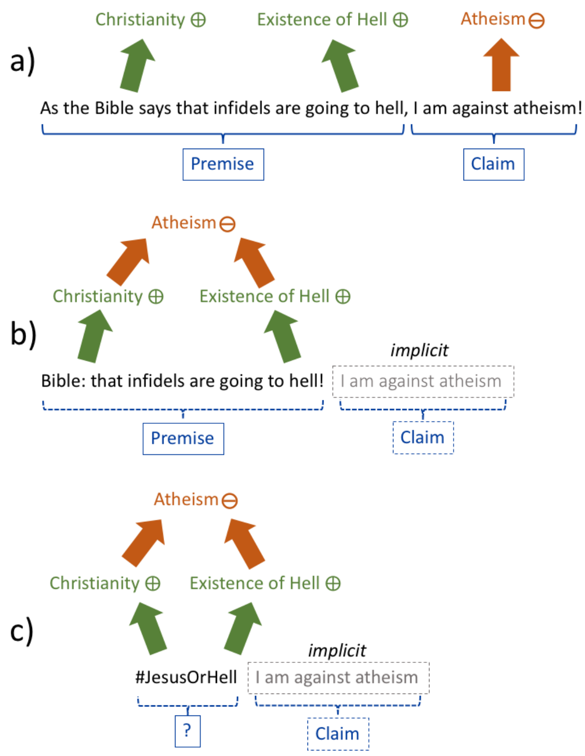

Figure 1: Stance-based vs. *claim-premise* model

transform a stance into a claim of the form *I am* {*in favor*|*against*} [TARGET]. As further illustrated in figure 1c, the claim-premise scheme is also not well suited for fragments like *#JesusOrHell*, while the fragment clearly invokes stances on christianity and the existence of hell.

In this paper, we show how explicitly expressed stances and the (possibly implicit) debate stance can be used as a proxy for argumentation. Compared to the traditional models of argument mining, our model has the advantage that stances are more easily derived and frequent than rich rhetorical structures. As we enforce explicit stances to be backed by direct textual evidence, we ensure a high reliability of the model. We argue that our model is especially useful for argument mining on social media texts, as the informal mode of communication leads to a high proportion of implicit arguments.[1] We annotate a corpus of noisy Twitter messages and show that our model can be reliably annotated and that it has the potential of improving the automated classification of stances as well as of traditional models of argument mining.

2 Related Work

Our model aims at capturing stances as a proxy for implicit arguments. Thus it cannot be directly compared with more complex models that assume typed relations between their components such as the *claim-premise* scheme. Here, we only discuss approaches linking stance and argumentation, and discuss the related work with respect to applicability to different text genres and inter-annotator agreement.

Boltužić and Šnajder (2014) use a set of predefined phrases such as *It is discriminatory to refuse gay couples the right to marry* and align them to stance labeled debate utterances. They report an agreement (Fleiss' Kappa) of 0.46 and 0.51 for the two debates in their corpus. Sobhani et al. (2015) also align predefined phrases with stance labeled comments but only indirectly relate them to the texts by mapping them to extracted statistical topic models. They state that this reduces the annotation effort, but the agreement remains rather low at 0.56 (Cohen's) kappa for tagging the arguments. Conrad et al. (2012) manually model two hierarchies of argumentative phrases with positive and negative stance as root nodes. Each hierarchy consists of more general phrases (such as *bill is politically motivated*) which are refined by phrases in the lower level of the hierarchy. After extensive training of the annotators, they reach a (Cohen's) kappa of 0.68. Hasan and Ng (2014) use argumentative phrases which have been previously extracted from the text. On four different domains they reach a (Cohen's) kappa of 0.78-0.82 on utterance level and of 0.61-0.67 on sentence level.

We thus conclude that enforcing an explicit grounding of annotation decisions in an utterance can be more reliably annotated than annotations that are mainly based on the interpretations of the annotators. Thus, in our model we only annotate stances if they have some explicit anchor in the text. For example, we would annotate a negative stance towards same-sex marriage (abbreviated notation: *Same-Sex Marriage* $\ominus$) only for a sentence like *gay marriage is a sin* where the stance is explicitly expressed, but not for a sentence like *as a true conservative, I trust in every word of the Bible* where *Same-Sex Marriage* $\ominus$ can only be inferred implicitly.

Misra et al. (2015) and Swanson et al. (2015) apply text summarization techniques to extract central propositions and then group them by a similarity measure which incorporates stance. For instance, if two statements relate to the same target, but express different polarities they are considered to be *roughly equivalent*. Consequently, stance is modelled only indirectly but may be inferred from the grouping of statements by the similarity measure. In addition, their approach relies on text summarization which does not make sense for very short texts such as the shown examples.

Another group of approaches deals with detecting agreement or disagreement between consecutive utterances (Ghosh et al., 2014; Clos et al., 2016), which could be interpreted as a stance towards the target that is mentioned in the first utterance. These models require a set of utterances organized in a conversation which limits the applicability. As we have seen from the examples in figure 1, even a single fragment can contain an argument. Our model should thus be applicable to single utterances, and not rely on a minimum text length.

It should be noted that all above mentioned studies have been carried out on data with relatively elaborate discussions, e.g. from dedicated web-based debating portals. We argue that applying those models to social media data like Tweets would result in considerably lower agreement, as the data contains a much higher proportion of implicit and less-elaborated arguments.

3 Modeling Arguments Using Stances

In order to solve the major challenge of implicit arguments that cannot be modeled well with existing approaches, we introduce a new model based on a *debate stance* that will in most cases be implicit, but can be inferred from one or more *explicit stances* that rely on textual evidence from the utterance. We thereby assume that an utterance is always made in the context of a certain debate.

Figure 2 gives on overview of the model which we metaphorically describe as an iceberg. In the

[1] For instance, the annotation of a comparatively elaborate social media corpus by Habernal et al. (2014) shows that almost half of the claims are implicit.

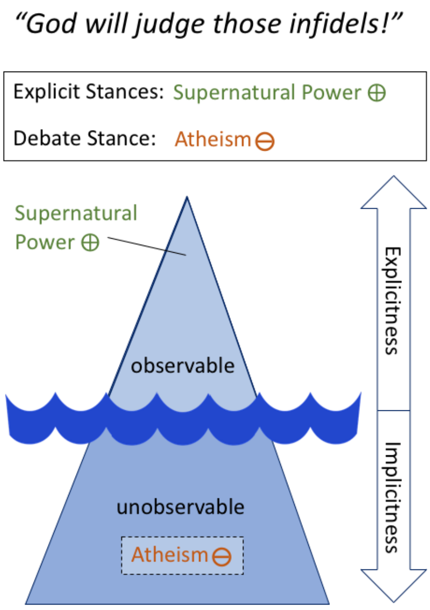

Figure 2: Our model and the iceberg metaphor for capturing implicit argumentation by using the the components (i) explicit stances and (ii) a debate stance.

context of a debate about atheism, an utterance like *God will judge those infidels!* is like the visible (explicit) part of the iceberg. It expresses a stance in favor of a supernatural power (*Supernatural Power* ⊕), while the actual stance on the debate target of atheism (*Atheism* ⊖) is not visible but must be inferred. Note that the debate stance might also be explicitly expressed (see figure 1a), but in implicit argumentation it has to be derived from the explicit stances.

In principle, each utterance evokes a large set of implicit stances (in a similar way as the iceberg contains a lot of invisible ice below the waterline). For instance, one may infer that a person uttering *Bible: infidels are going to hell!* is probably in favor of praying and might have a negative stance towards issues such as abortion, same-sex marriage, etc. However, we argue that being in favor of Christianity already implicitly covers these stances under a common sense interpretation. Depending on the present informational need these targets may be more or less relevant.

For modeling stance, we can build on plenty of research (Anand et al., 2011; Somasundaran and Wiebe, 2009; Sridhar et al., 2014; Hasan and Ng, 2013) and even a shared task on automatic stance detection (Mohammad et al., 2016). These works commonly define stance as being in favor of or against a given target. Consequently, stance is a tuple consisting of a target and a stance expression such as *Atheism* ⊕ or *Atheism* ⊖.

Debates can be categorized in two sided debates in which authors can take a pro or contra stance and more open debates which may contain several other targets (e.g. *What evidence do we have for global warming?*). However, we argue that each of the targets in an open debate – e.g. a certain piece of evidence for global warming – can be considered as a two sided debate. I.e. an authors may agree or disagree on an elevated sea level as evidence of global warming. Moreover, if one acknowledges that the participants in a two sided debate also discuss certain sub-topics, the separation between two sided debates and open debates vanishes.

Debate Stance As described above, we refer to the (frequently implicit) stance towards the target of the whole debate as *debate stance*. For instance, if in the context of an atheism debate someone describes their personal faith, we may assume that they want to communicate the fact that they are against atheism. Note that exactly the same utterance might not communicate a stance against atheism in the context of another debate such as on the importance of charity.

Explicit Stances While the overall debate stance may be implicit, there has to be some explicit information in the utterance that enables this inference. Otherwise the goal of the persuasive utterance (i.e. convincing someone or at least expressing her standpoint) cannot be achieved. As a stance can always be transformed into a claim which can be considered as the minimum constituent of an argument (Habernal et al., 2014; Palau and Moens, 2009), we argue that the minimal information that has to be provided in a persuasive utterance is a stance towards some target.

Given a stance, humans can infer the argument using a common sense interpretation. If one states *God will judge those infidels* (*Believe in God* ⊕) in an atheism debate, one can infer stances such as *being a infidel is a sin* ⊕, *God punishes infidels* ⊕ and the debate stance *Atheism* ⊖. If an author wants to deviate from this interpretation, they need to communicate this explicitly, e.g. by adding *but the constitution grants religious freedom* (*religious freedom* ⊕).

From lexical priming studies it is known that the perception of words can activate knowledge about associated concepts or real-world events (Jones and Estes, 2012; Hare et al., 2009). Since there is also strong evidence for priming effects of stimuli other than words (Tulving and Schacter, 1990), we conclude that priming should be applicable to stances as well and therefore forms the underlying mechanism of implicit argumentation.

Thereby, our model of implicit argumentation aligns with the *Relevance Theory* proposed by Sperber and Wilson (1986) and the *Cooperative Principle* by Grice (1970) as we also assume that utterances provide hints on the intended meaning to the recipient. Particularly, our model shares the assumption of *Relevance Theory* that the precision of statements is such that a receiver can decode the meaning only by incorporating the context.

Selection of Stance Targets As indicated by our iceberg metaphor (see figure 2), just a small proportion of the argument is observable but the larger part is hidden from sight. The granularity of the stance targets has thereby to be considered with respect to the present informational needs. If one wants to get a more general view on the examples in figure 1, one could fall back to the target *belief in a supernatural power* which is also less explicitly covered. Depending on what degree of explicitness is chosen, an utterance can thereby express more than one explicit stance. Analogously, the unobservable parts of the argument vary in the degree of their implicitness. The degree of implicitness is seen here as the strength to which other stances are primed by the explicit part. For instance, if one claims the existence of hell, one affirms the existence of heaven with a small degree of implicitness but a stance about reincarnation is taken only very implicitly.

What level of granularity should be chosen is an open research question. As demonstrated by Conrad et al. (2012), a too fine grained distinction has the consequence of a sparse distribution which makes it difficult to derive relations between components of their model or to enable automated classification. Thus, selecting the most explicit targets does not appear to be the appropriate level to gain comprehensive insights on how taking a stance in a debate is manifested by explicit stances. However, if a target is too implicit, it might be invoked by authors in favor of the debate target as well as against the target.

4 Corpus Annotation

In order to show that our approach is indeed viable, we conduct an annotation study on social media data from the SemEval 2016 task 6 on stance detection. This enables us (i) to assess how reliably our model can be annotated, (ii) to examine what insights we can get by inspecting usage patterns of explicit stances, and (iii) to estimate how well our model can be assigned automatically. We now describe in more detail the utilized data, the annotation process, and how we derived the targets in a granularity that we found to be appropriate.

4.1 Data

As our argumentation scheme is centered around stance, we rely on data used by the first shared task on automated stance detection (Mohammad et al., 2016) which enables us to consider the present work in this context. A relevant property of the data, as stated by the task organizers, is that it contains a high proportion of tweets that do not explicitly mention the target and therefore can be considered as implicit utterances.

We focus here on Subtask A with tweets about five targets which are annotated for being in favor/against a target or if neither such inference is likely. We limit our study to 733 tweets on *Atheism* (513 from the training set and 220 from the test set), as we found the topic to require less knowledge about specific political events.

4.2 Derivation of Targets

In our model, choosing the right number and granularity of targets is crucial. On the one hand, they have to be expressive enough to capture differences in nuanced argumentation. On the other hand, they should not be too fine grained as this would result in very sparse distributions that cannot be handled by automated methods. Therefore, we utilize a semi-automated, bottom-up approach that focusses on concepts that are mostly explicitly expressed by named entities and nouns. We consider the 50 most frequent concepts. It should be noted that in this corpus of Twitter messages on Atheism, the *atheism* appears exactly once and the *atheist* only 6 times. This indicates that implicit argumentation is prevalent in social media.

As we want to ensure that the targets used enable us to differentiate the authors' positions sufficiently, we also consider the degree of association between nouns and named entities to the stances *Atheism* $\oplus$

and *Atheism* ⊖. In detail, we compute the collocation coefficient *Dice* (Smadja et al., 1996) for each word, and selected the 25 words which are most strongly associated with *Atheism* ⊖ and *Atheism* ⊕.

We found the resulting concepts to be too numerous and too fine-grained to be used in our model. We thus, manually group concepts into more coarse-grained targets. For instance, concepts such as *Bible* and *Jesus* are grouped into the target *Christianity*. A potential criticism of our approach is that at this stage of our work, we can not evaluate whether the set is best possible choice. We plan to shed light on this aspect in future research. The final set of derived, explicit targets is shown in table 1.

4.3 Annotation Process

Using the selected data, we let three annotators (two undergraduate and one graduate student of cognitive science) identify stances towards the derived targets and the debate target. In order to familiarize the annotators with our model, we previously trained them on a small data set that is comparable in its social media character but concerns a different target.

Since the data partly contains utterances which cannot be understood without further context, we give annotators the option to mark them accordingly. Irony is another phenomenon, which influences the interpretability. Therefore, we asked the annotators to annotate the tweets for irony as well.

Since it is still possible that our annotators interpret the tweets differently than in the original annotation, we re-annotated the debate stance using the original questionnaire described in Mohammad et al. (2016). While annotating explicit stances, the annotators had the instruction to only annotate stances towards targets if they have textual evidence for it.

5 Evaluation

In this section, we evaluate the annotated data. For this purpose, we first analyze the reliability of the annotation on different levels of granularity using Fleiss' Kappa (κ). For the analysis, we exclude tweets that are annotated for irony and understandability issues. However, we found that the annotators rarely agree on these phenomena as we get a κ of only 0.06 for understandability and a κ of 0.23 for irony. Therefore, we only exclude 18 tweets in which at least two annotators share the same judgment, which results in 715 tweets for the final corpus.

5.1 Inter Annotator Agreement

Since the explicit targets are annotated on the basis of textual evidence, we expect a high level of agreement. The notation of explicit targets should also result in a strong agreement of the annotation of the debate stance because it enforces a deep analysis of the communicative goal of an utterance. As shown in figure 3, we obtain a Fleiss' κ of 0.72 for the annotation of the debate stance. Unfortunately, we cannot compare our agreement to the originally SemEval data, as the organizers do not report a chance corrected agreement measure for their final decision. Also not directly comparable is the agreement of Sobhani et al. (2015) as they report weighted κ. We argue that their weighted κ of 0.62 is in a range similar to ours.

In figure 3, we also show the agreement for the explicit targets. Since explicit stances have a similar, deriving function like the argumentative phrases proposed by Conrad et al. (2012) and Hasan and Ng (2014), we compare our agreements to theirs which does not exceed a Cohen's κ of 0.68. Two targets (*Christianity* and *Islam*) yield especially high agreement above 0.8, because they are associated with clear signal words such as *Jesus* and *Quran* and other markers such as the numerical reference to biblical passages. Other targets such as *Secularism* and *Freethinking* are rather abstract. They hardly involve special signal words but still gain high agreements of a κ above 0.7, which shows that our annotators did not just learn to recognize certain keywords, but can also reliably annotate more abstract targets. This is further supported by the fact that the agreement for the annotation of *no explicit target* is also in this range. The targets *USA*, *Religious Freedom*, *Same-Sex Marriage*, and *Life After Death* yield only a moderate agreement between 0.4 and 0.6. An error analysis for the target *Same-Sex Marriage* shows that there is disagreement if the tweet contains a stance towards gay rights in general but not to gay marriage. We therefrom see two possibilities here to improve the agreement: On the one hand, we could choose more comprehensive targets such as *gay rights* to cover the combined positions. On the other hand, we could train the annotators to more consistently account for such differences. A rather low κ of 0.31 is obtained for the target *No Evidence*.

Explicit Target	Description	Examples for Textual Evidence
Christianity	belief in the religion Christianity	Jesus, Christ, Bible, Mary Mother of God, Catholic, Gospel
Freethinking	idea that truth should be formed on the basis of logic, reason, and empiricism	#freethinking, #DogmasNeverHelp
Islam	belief in the religion Islam	Quran, Ummah, Hadith, Mohammed, Allah
No Evidence	idea that there is no evidence for religion	there is no evidence for God
Life After Death	believe in an existence after death	paradise, heaven, hell, Dschanna
Supernatural Power	belief in a supernatural being or an abstract supernatural power	God, Lord, Jesus, holy spirit, Allah, Ganesha, destiny, predestination
USA	United States of America	our country, our state, America, US
Conservatism	the conservative movement in the USA	republicans, #tcot, tea party
Same-Sex Marriage	the right of a same-sex couples to marry	gay marriage, same-sex marriage
Religious Freedom	everyone should have the freedom to have and practice any religion	#religiousfreedom, right to choose your religion
Secularism	religion and nation should be strictly separated	separation of church and state, #secularism

Table 1: Explicit targets which are semi-automatically derived for the debate target *Atheism*

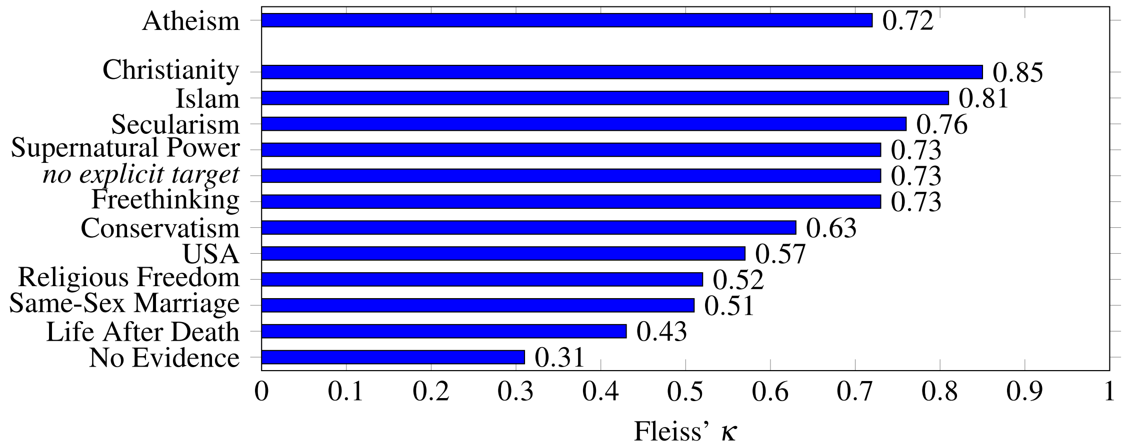

Figure 3: Inter-annotator agreement of the debate stance *Atheism* and explicit stances

Regarding this target, we observe that annotators sometimes deviated from our guidelines and incorporated different degrees of inferred knowledge as they used *Bill Nye* or *Richard Dawkins*[2] as anchors for their decisions, although the utterance contains no explicit stance in favor of *No Evidence*.

Finally, we obtain a κ of 0.63 for the joint decision on both the debate and the explicit targets. Note that this agreement is not directly comparable with the approaches from related work, as they only implicitly model the debate stance, do not report agreements of a joint decision or rely on stances that are determined by the structure of the data. The obtained inter-annotator agreement shows that our model can be annotated reliably and that the recognized difficulties may be compensated by a better training of the annotators and a better selection of targets.

[2]famous supporters of the position that there is no evidence for religion

5.2 Stance Pattern Analysis

In order to inspect usage patterns of explicit stance taking, we must agree on one annotation for each tweet. Since we do not assume that there are differences in the quality of the three annotators, we rely on a majority vote to compile a final annotation.

Figure 4 visualizes the frequency of the explicitly taken stances for Atheism $\oplus$ and Atheism $\ominus$. It shows that there are significant differences in the argumentation patterns between the two camps. As expected, if advocates of atheism are against a target such as *Christianity*, the opponents are mostly in favor of it or do not mention it. This pattern is also observable for the reverse case such as for *Freethinking*. Note that utterances addressing the target *Same-Sex Marriage* are exclusively annotated for expressing no stance towards *Atheism*. Further exceptions are the targets *USA* and *Religious Freedom* that are positively mentioned by both camps. However, a deeper analysis shows that

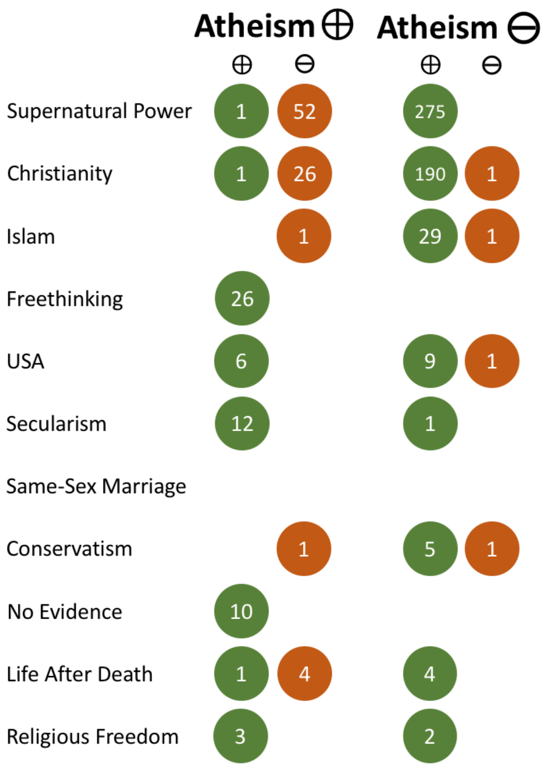

Figure 4: Frequency of explicit stances grouped according to debate stance

these targets always occur together with other targets which seem to be more relevant for the debate stance.

In order to analyze stance patterns in more details, we show which other stances are used together with the target *Supernatural Power* (the most frequent target in both camps) in figure 5. We observe that authors that are against Atheism use *Christianity* ⊕ together with *Supernatural Power* ⊕ in 50% of all cases. In contrast, authors that are in favor of Atheism only combine *Supernatural Power* ⊖ with *Christianity* ⊖ in 13% of all cases. The figure also shows that the other explicit stances only play a subordinate role in the combination with those targets.

From these analyses we can conclude that stable patterns of argumentation using explicit stances other than the debate stance exist. This is a strong indication for the validity of our assumption that the debate stance can be inferred from explicitly expressed stances.

5.3 Automatically Assigning Stances

We now want to examine how well the two main components of our model – the explicit stances and the debate stance – can be automatically assigned.

Target (# instances)	Majority Class Baseline	Our Approach
Supernatural Power (335)	.53	.78
Christianity (223)	.69	.79
Islam (43)	.94	.95

Table 2: Explicit stance classification (only showing targets occurring in at least 5% of all instances)

Feature Set	F_1
majority class baseline	.49
n-gram	.66
explicit stance$_{predicted}$	.65
explicit stance$_{oracle}$	.88

Table 3: Debate stance classification

We re-implement a state-of-the-art classifier (Mohammad et al., 2016) using the DKPro TC framework[3] (Daxenberger et al., 2014) and leave the development of sophisticated classification models to future research. For preprocessing, we rely on the DKPro Core framework[4] (Eckart de Castilho and Gurevych, 2014) and apply a twitter-specific tokenizer (Gimpel et al., 2011). In all experiments, we use ten-fold cross-validation and report micro averaged F_1.

Explicit Stances As the results from the stance detection task in SemEval-2016 (Mohammad et al., 2016) indicate, a support vector machine with a linear kernel equipped with simple word and character n-gram features is the state of the art in automated stance prediction. Table 2 shows the results of our reimplementation of state-of-the-art classifier (using weka's SMO) and the majority class baseline for comparison. The results indicate that the two most frequent targets can be classified with success, if one relates them to the majority class baseline. We observe that each target has its own linguistic markers such as the use of Arabic terms if one is in favor of Islam. Therefore, we argue that these peculiarities can be targeted even better by specialized features.

The analysis in table 2 excludes targets that have a insufficient coverage (less than 5% of all instances) to train a meaningful model. A possibility to deal with this sparsity may be to incorporate unlabelled data such as demonstrated for traditional models by Habernal and Gurevych (2015).

[3] version 0.8.0
[4] version 1.7.0

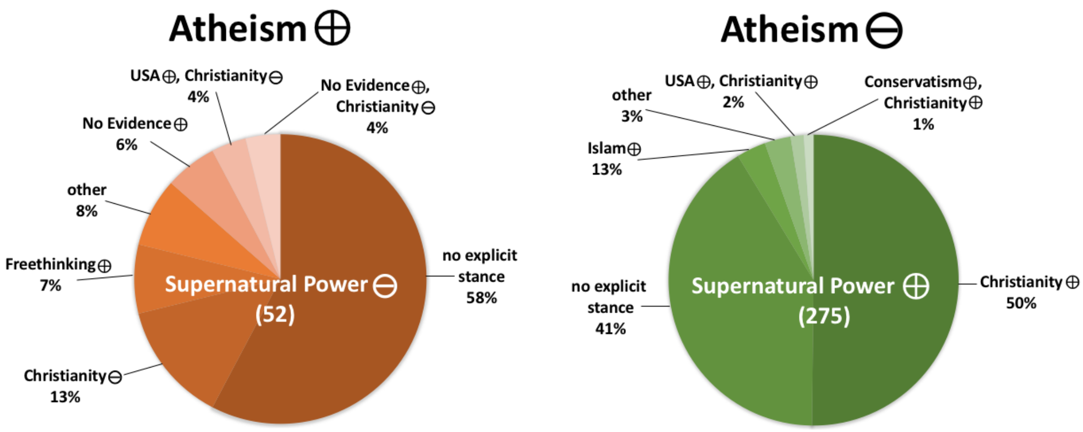

Figure 5: Most frequently used, explicit stances and the percentage shares to which they cooccur with other explicit stances

Debate Stance Table 3 shows the results obtained for automatically assigning the debate stance. Besides the majority class baseline ($F_1 = .49$), we use the same setup as for the explicit stances to train an n-gram based classifier and obtain an F_1 of .66. In order to evaluate the usefulness of explicit stances for inferring the debate stance, we use the predictions from the previous experiment as features for a decision tree classifier (J48). This stacked classifier performs on par (.65) with the n-gram based classifier. It seems that the quality of predicting explicit stances is not yet good enough to improve over the state-of-the-art without incorporating general n-gram features.

In order to estimate the potential of explicit stance features for classifying the debate stance, we add an oracle condition to our experiments in which we assume that the classification of explicit stances is done correctly. This classifier using only the manually annotated explicit stances yields an F_1 score of .88 showing that large improvements over the state of the art are possible if explicit stances can be more reliably classified. We believe that this is indeed possible as explicit stances are always grounded in the text itself, while the debate stance might only be indirectly inferred.

6 Conclusions and Future Work

We have identified implicitness as a major remaining problem in argument mining. Implicit arguments are only poorly supported by textual evidence and need to be inferred. We propose to model implicit argumentation by explicit stances and that cover more implicit stances and – most importantly – the overall stance that is taken in a debate. As we thereby enforce that the explicit stances are assigned with respect to textual evidence, we can ensure that our model is grounded on the actual utterances and less on their interpretation. As we argue that stances can always be interpreted as claims, our approach is interpretable in the form of a *claim-premise* scheme and therefore takes a step in bridging the gap between argument mining and stance detection. We provide evidence that this model can be reliably annotated, even on such a challenging domain as social media. In addition, we demonstrate that the model has the potential to boost performance in the automated detection of debate stance and traditional argument mining. We make the annotated data publicly available[5].

As this is a first attempt on modeling implicit arguments using stances, we see several lines of future research. First, we want to examine how the degree of granularity of explicit targets effects the quality of the model. Furthermore, we want to enhance our approach with an automated derivation of these targets. Finally, we want to improve the automatic assignment of explicit stances to unleash the full potential of explicit stances for argument mining.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media". We also want to thank our annotators Dominik Lawatsch and Niklas Meyer.

[5]http://www.ltl.uni-due.de/stance-based-am

References

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 1–9, Stroudsburg, USA.

Filip Boltužić and Jan Šnajder. 2014. Back up your stance: Recognizing arguments in online discussions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 49–58, Baltimore, USA.

Jérémie Clos, Nirmalie Wiratunga, Stewart Massie, and Guillaume Cabanac. 2016. Shallow techniques for argument mining. In *European Conference on Argumentation (to appear)*, Lisbon, Portugal.

Alexander Conrad, Janyce Wiebe, et al. 2012. Recognizing arguing subjectivity and argument tags. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 80–88, Stroudsburg, USA.

Johannes Daxenberger, Oliver Ferschke, Iryna Gurevych, Torsten Zesch, et al. 2014. Dkpro tc: A java-based framework for supervised learning experiments on textual data. In *ACL (System Demonstrations)*, pages 61–66, Baltimore, USA.

Richard Eckart de Castilho and Iryna Gurevych. 2014. A broad-coverage collection of portable nlp components for building shareable analysis pipelines. In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT*, pages 1–11, Dublin, Ireland.

Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui. 2014. Analyzing argumentative discourse units in online interactions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 39–48, Baltimore, USA.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47, Portland, USA.

Herbert P. Grice. 1970. *Logic and conversation*, volume 3. Academic Press.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the EMNLP*, pages 2127–2137, Lisbon, Portugal. Association for Computational Linguistics.

Ivan Habernal, Judith Eckle-Kohler, and Iryna Gurevych. 2014. Argumentation mining on the web from information seeking perspective. In *ArgNLP*.

Mary Hare, Michael Jones, Caroline Thomson, Sarah Kelly, and Ken McRae. 2009. Activating event knowledge. *Cognition*, 111(2):151–167.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the IJCNLP*, pages 1348–1356, Nagoya, Japan.

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the EMNLP*, pages 751–762, Doha, Qatar.

Lara L. Jones and Zachary Estes. 2012. Lexical priming: Associative, semantic, and thematic influences on word recognition. *Visual word recognition*, 2:44–72.

Amita Misra, Pranav Anand, JEF Tree, and MA Walker. 2015. Using summarization to discover argument facets in online idealogical dialog. In *Proceedings of the NAACL HLT*, pages 430–440, Denver, USA.

Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation (to appear)*, San Diego, USA.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107, New York, USA.

Andreas Peldszus and Manfred Stede. 2013. Ranking the annotators: An agreement study on argumentation structure. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 196–204, Sofia, Bulgaria.

Frank Smadja, Kathleen R. McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational linguistics*, 22(1):1–38.

Parinaz Sobhani, Diana Inkpen, and Stan Matwin. 2015. From argumentation mining to stance classification. In *Proceedings of the NAACL HLT 2015*, pages 67–77, Denver, USA.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 226–234, Singapore.

Dan Sperber and Deirdre Wilson. 1986. Relevance: communication and cognition. *Language in Society*, 17(04):604–609.

Dhanya Sridhar, Lise Getoor, and Marilyn Walker. 2014. Collective stance classification of posts in online debate forums. In *Proceedings of the joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pages 109–117, Baltimore, USA.

Reid Swanson, Brian Ecker, and Marilyn Walker. 2015. Argument mining: Extracting arguments from online dialogue. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 217–226, Prague, Czech Republic.

Endel Tulving and Daniel L. Schacter. 1990. Priming and human memory systems. *Science*, 247(4940):301–306.