Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews

Pavithra Rajendran
University of Liverpool

Danushka Bollegala
University of Liverpool

Simon Parsons
King's College London

Abstract

Enthymemes, that are arguments with missing premises, are common in natural language text. They pose a challenge for the field of argument mining, which aims to extract arguments from such text. If we can detect whether a premise is missing in an argument, then we can either fill the missing premise from similar/related arguments, or discard such enthymemes altogether and focus on complete arguments. In this paper, we draw a connection between explicit vs. implicit opinion classification in reviews, and detecting arguments from enthymemes. For this purpose, we train a binary classifier to detect explicit vs. implicit opinions using a manually labelled dataset. Experimental results show that the proposed method can discriminate explicit opinions from implicit ones, thereby providing encouraging first step towards enthymeme detection in natural language texts.

1 Introduction

Argumentation has become an area of increasing study in artificial intelligence (Rahwan and Simari, 2009). Drawing on work from philosophy, which attempts to provide a realistic account of human reasoning (Toulmin, 1958; van Eemeren et al., 1996; Walton and Krabbe, 1995), researchers in artificial intelligence have developed computational models of this form of reasoning. A relatively new sub-field of argumentation is *argument mining* (Peldszus and Stede, 2013), which deals with the identification of arguments in text, with an eye to extracting these arguments for later processing, possibly using the tools developed in other areas of argumentation.

Examining arguments that are found in natural language texts quickly leads to the recognition that many such arguments are incomplete (Lippi and Torroni, 2015a). That is if you consider an argument to be a set of *premises* and a *conclusion* that follows from those premises, one or more of these elements can be missing in natural language texts. A premise is a statement that indicates support or reason for a conclusion. In the case where a premise is missing, such incomplete arguments are known as *enthymemes* (Walton, 2008). One classic example is given below.

Major premise *All humans are mortal (unstated).*

Minor premise *Socrates is human (stated).*

Conclusion *Therefore, Socrates is mortal (stated).*

According to Walton, enthymemes can be completed with the help of common knowledge, echoing the idea from Aristotle that the missing premises in enthymemes can be left implicit in most settings if they represent familiar facts that will be known to those who encounter the enthymemes. Structured models from computational argumentation, which contain structures that mimic the syllogisms and argument schemes of philosophical argumentation will struggle to cope with enthymemes unless we can somehow provide the unstated information.

Several authors have already grappled with the problem of handling enthymemes and have represented shared common knowledge as a solution to reconstruct these enthymemes (Walton, 2008; Black and Hunter, 2012; Amgoud and Prade, 2012; Hosseini et al., 2014).

In this paper, we argue that there exists a close relationship between detecting whether a particular statement conveys an explicit or an implicit opinion, and whether there is a premise that supports the conclusion (resulting in a argument) or not (resulting in an enthymeme). For example, consider the following two statements S_1 and S_2:

$S_1 =$ *I am extremely disappointed with the room.*

$S_2 =$ *The room is small.*

Both S_1 and S_2 express a negative sentiment towards the room aspect of this hotel. In S_1, the stance of the reviewer (whether the reviewer is in favour or against the hotel) is explicitly stated by the phrase *extremely disappointed*. Consequently, we refer to S_1 as an *explicitly opinionated* statement about the room. However, to interpret S_2 as a negative opinion we must possess the knowledge that being small is often considered as negative with respect to hotel rooms, whereas being small could be positive with respect to some other entity such as a mobile phone. The stance of the reviewer is only implicitly conveyed in S_2. Consequently, we refer to S_2 as an *implicitly opinionated* statement about the room. Given the conclusion that this reviewer did not like this room (possibly explicitly indicated by a low rating given to the hotel), the explicitly opinionated statement S_1 would provide a premise forming an argument, whereas the implicitly opinionated statement S_2 would only form an enthymeme. Thus:

Argument

Major premise *I am extremely disappointed with the room.*

Conclusion *The reviewer is not in favour of the hotel.*

whereas:

Enthymeme

Major premise *A small room is considered bad (unstated).*

Minor premise *The room is small.*

Conclusion *The reviewer is not in favour of the hotel.*

Our proposal for enthymeme detection via opinion classification is illustrated in Figure 1, and consists of the following two steps. This assumes a separate process to extract the ("predefined") conclusion, for example from the rating that the hotel is given.

Step-1 Opinion structure extraction

a. Extract statements that express *opinions* with the help of local sentiment (positive or negative) and discard the rest of the statements.

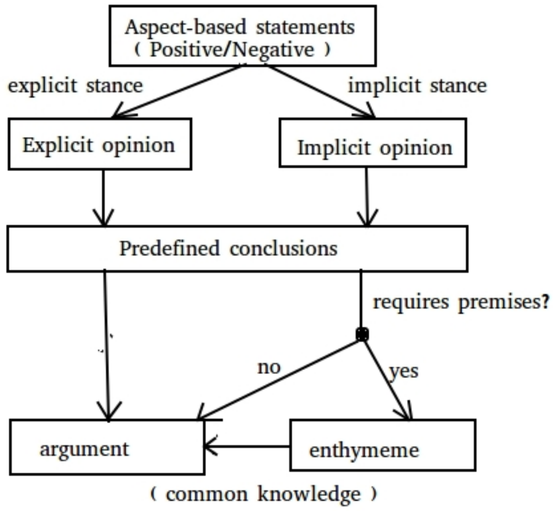

Figure 1: The relationship between explicit/implicit opinions and arguments/enthymemes.

b. Perform an aspect-level analysis to obtain the aspects present in each statement and those statements that include an aspect are considered and the rest of the statements are discarded.

c. Classify the *stance* of statements as being explicit or implicit.

Step-2 Premise extraction

a. Explicit opinions paired with the predefined conclusions can give us complete arguments.

b. Implicit opinions paired with the predefined conclusions can either become arguments or enthymemes. Enthymemes require additional premises to complete the argument.

c. Common knowledge can then be used to complete the argument.

This process uses both opinion mining and stance detection to extract arguments, but it still leaves us with enthymemes. Under some circumstances, it may be possible to combine explicit and implicit premises to complete enthymemes.

To see how this works, let us revisit our previous example. The explicitly opinionated statement *"I am extremely disappointed with the room"* can be used to complete an argument that has premise *"the rooms are small and dirty"*, which was extracted from the review, and a conclusion that *"The hotel is not favored"* which comes from the fact that the review has a low rating.

Argument

Major premise *I am extremely disappointed with the room.*

Minor premise *the rooms are small and dirty*

Conclusion *The reviewer is not in favour of the hotel.*

While developing this approach is our long-term goal, this paper has a much more limited focus. In particular we consider Step 1(c), and study the classification of opinions into those with explicit stance and those with implicit stance.

We focus on user reviews such as product reviews on Amazon.com, or hotel reviews on TripAdvisor.com. Such data has been extensively researched for sentiment classification tasks (Hu and Liu, 2004; Lazaridou et al., 2013). We build on this work, in particular, aspect-based approaches. In these approaches, sentiment classification is based around the detection of terms that denote *aspects* of the item being reviewed — the battery in the case of reviews of portable electronics, the room and the pool in the case of hotel reviews — and whether the sentiment expressed about these aspects is positive or negative.

Our contributions in this paper are as follows:

- As described above, we propose a two-step framework that identifies opinion structures in aspect-based statements which help in detecting enthymemes and reconstructing them.
- We manually annotated a dataset classifying opinionated statements to indicate whether the author's stance is explicitly or implicitly indicated.
- We use a supervised approach using the SVM classifier to automatically identify the opinion structures as *explicit* and *implicit* opinions using the n-grams, part of speech (POS) tags, SentiWordNet scores and noun-adjective patterns as features.

2 Related work

Argument mining is a relatively new area in the field of computational argumentation. It seeks to automatically identify arguments from natural language texts, often online texts, with the aim of helping to summarise or otherwise help in processing such texts. It is a task which, like many natural language processing tasks, varys greatly from domain to domain. A major part of the challenge lies in defining what we mean by an argument in unstructured texts found online. It is very difficult to extract properly formed arguments in online discussions and the absence of proper annotated corpora for automatic identification of these arguments is problematic.

According to Lippi and Torroni (2015a) who have made a survey of the various works carried out in argument mining so far with an emphasis on the different machine learning approaches used, the two main approaches in argument mining relate to the extraction of abstract arguments (Cabrio and Villata, 2012; Yaglikci and Torroni, 2014) and structured arguments.

Much recent work in extracting structured arguments has concentrated on extracting arguments pertaining to a specific domain such as online debates (Boltužić and Šnajder, 2014), user comments on blogs and forums (Ghosh et al., 2014; Park and Cardie, 2014), Twitter datasets (Llewellyn et al., 2014) and online product reviews (Wyner et al., 2012; Garcia-Villalba and Saint-Dizier, 2012). Each of these work target on identifying the kind of arguments that can be detected from a specific domain.

Ghosh et al. (2014) analyse target-callout pairs among user comments, which are further annotated as stance/rationale callouts. Boltuzic and Snaider (2014) identify argument structures that they propose can help in stance classification. Our focus is not to identify the stance but to use the stance and the context of the relevant opinion to help in detecting and reconstructing enthymemes present in a specific domain of online reviews.

Lippi and Torroni (2015b) address the domain-dependency of previous work by identifying claims that are domain-independent by focussing on rhetoric structures and not on the contextual information present in the claim.

Habernal et al. (2014) consider the context-independent problem using two different argument schemes and argues that the best scheme to use varies depending upon the data and problem to be solved. In this paper, we address a domain-dependent problem of identifying premises with the help of stance classification. We think that claim identification will not solve this problem, as online reviews are rich in descriptive texts that are mostly premises leading to a conclusion as to whether a product/service is good or bad.

There are a few papers that have concentrated on identifying enthymemes. Feng and

Hirst (2011) classify argumentation schemes using explicit premises and conclusion on the Araucaria dataset, which they propose to use to reconstruct enthymemes. Similar to (2011), Walton (2010) investigated how argumentation schemes can help in addressing enthymemes present in health product advertisements. Amgoud et al. (2015) propose a formal language approach to construct arguments from natural language texts that are mostly enthymemes. Their work is related to mined arguments from texts that can be represented using a logical language and our work could be useful for evaluating (Amgoud et al., 2015) on a real dataset. Unlike the above, our approach classifies stances which can identify enthymemes and implicit premises that are present in online reviews.

Research in opinion mining has started to understand the argumentative nature of opinionated texts (Wachsmuth et al., 2014a; Vincent and Winterstein, 2014). This growing interest to summarise what people write in online reviews and not just to identify the opinions is much of the motivation for our paper.

3 Method

3.1 Manual Annotation of Stance in Opinions

We started with the ArguAna corpus of hotel reviews from TripAdvisor.com (Wachsmuth et al., 2014b) and manually separated those statements that contained an aspect and those that did not. This process could potentially be carried out automatically using opinion mining tools, but since this information was available in the corpus, we decided to use it directly. We found that many of the individual statements in the corpus directly refer to certain aspects of the hotel or directly to the hotel itself. These were the statements we used for our study. The rest were discarded.[1]

Each statement in the corpus was previously annotated as positive, negative or objective (Wachsmuth et al., 2014b). Statements with a positive or negative sentiment were more opinion-oriented and hence we discarded the statements that were annotated as objective. A total of 180 reviews then gave us 784 opinions. Before we annotated the statements, we needed to define the possible (predefined) conclusions for the hotel reviews, and these were:

Conclusion 1 The reviewer is in favor of an aspect of the hotel or the hotel itself.

Conclusion 2 The reviewer is against an aspect of the hotel or the hotel itself.

We then annotated each of the 784 opinions with one of these conclusions. This was done to make the annotation procedure easier, since each opinion related to the conclusion forms either a complete argument or an enthymeme. During the annotation process, each opinion was annotated as either explicit or implicit based on the stance definitions given above. The annotation was carried out by a single person and the ambiguity in the annotation process was reduced by setting out what kind of statements constitute explicit opinions and how these differ from implicit opinions. These are as follows:

General expressive cues Statements that explicitly express the reviewer's views about the hotel or aspects of the hotel. Example indicators are *disappointed, recommend, great.*

Specific expressive cues Statements that point to conclusions being drawn but where the reasoning is specific to a particular domain and varies from domain to domain. Examples are *"small size batteries"* and *"rooms are small"*. Both represent different contextual notions, where the former suggests a positive conclusion about the battery and the latter suggests a negative conclusion about the room. Such premises need additional support.

Event-based cues Statements that describe a situation or an incident faced by the reviewer and needs further explanation to understand what the reviewer is trying to imply.

Each statement in the first category (general expressive) is annotated as an explicit opinion and those that match either of the last two categories (specific expressive, event-based) were annotated as non-explicit opinions. The non-explicit opinions were further annotated as having a neutral or implicit stance. We found that there were statements that were both in favor of and against the hotel and we annotated such ambiguous statements as being neutral.

[1]The remaining statements could potentially be used, but it would require much deeper analysis in order to construct arguments that are relevant to the hotels. The criteria for our current work is to collect simpler argument structures that can be reconstructed easily, and so we postpone the study of the rest of the data from the reviews for future work.

Explicit stance	Implicit stance
i would not choose this hotel again.	the combination of two side jets and one fixed head led to one finding the entire this bathroom flooded upon exiting the shower.
great location close to public transport and chinatown.	the pool was ludicrously small for such a large property, the sun loungers started to free up in the late afternoon.
best service ever	the rooms are pretentious and boring.

Table 1: A few examples of statements from the hotel data that are annotated with explicit and implicit stances.

From the manually annotated data, 130 statements were explicit, 90 were neutral and the rest were implicit. In our experiments, we focussed on extracting the explicit opinions and implicit opinions and thus ignored the neutral ones. Table 1 shows examples of statements annotated as explicit and implicit.

As shown in Figure 1, explicit opinions with their appropriate conclusions can form complete arguments. This is not the case for implicit opinions. Implicit opinions with their appropriate conclusions may form complete arguments or they may require additional premises to entail the conclusion. In this latter case, the implicit opinion and conclusion form an enthymeme. As discussed above, we may be able to use related explicit opinions to complete enthymemes. When we look to do this, we find that the explicit opinions in our dataset fall into two categories:

General These explicit opinions are about an aspect category, which in general, can be related to several sub-level aspects within the category.

Specific These explicit opinions are about a specific aspect and hence can only be related to that particular aspect.

To illustrate the difference between the two kinds of explicit claim, let us consider three examples given below.

- A1 : "The front desk staffs completely ignored our complaints and did nothing to make our stay better". *(implicit)*
- A2 : "The front desk staff are worst". *(specific explicit)*
- A3 : "I am disappointed with the overall customer service!" *(general explicit)*

In this case, both the specific opinion A2: "The front desk staff are worst", and the general opinion A3: "I am disappointed with the overall customer service" will work to complete the argument because the aspect *front desk staff* of the specific explicit opinion A2 matches that of the implicit statement A1. However, if the implicit statement was about another aspect (say the room cleaning service), then A2 woud not match the aspect, whereas the more general statement A3 would.

Having sketched our overall approach to argument extraction and enthymeme completion, we turn to the main contribution of the paper — an exploration of stance classification on hotel review data, to demonstrate that Step 1(c) of our process is possible.

3.2 Learning a Stance Classifier

Since we wish to distinguish between explicit and implicit stances, we can consider the task as a binary classification problem. In this section we describe the features that we considered as input to a range of classifiers that we used on the problem. Section 4 describes the choice of classifiers that we used.

The following are a set of features that we used.

Baseline As a baseline comparison, statements containing words from a list of selected cues such as excellent, great, worst etc. are predicted as explicit and those that do not contain words present in the cue list are predicted as implicit. The criteria followed here is that the statement should contain atleast one cue word to be predicted as explicit. The ten most important cue words were considered.

N-grams (Uni, Bi) Unigrams (each word) and bigrams (successive pair of words).

Part of Speech (POS) The NLTK[2] tagger helps in tagging each word with its respective part of speech tag and we use the most common tags (noun, verb and adjective) present in the explicit opinions as features.

[2]Natural Language Toolkit, `www.nltk.org`

Classifier	Explicit	Implicit
Logistic Regression	0.44	0.86
MultinomialNB	0.27	0.85
Linear SVM	**0.75**	**0.90**

Table 2: F1-scores of 5-fold cross validation results performed with different classifiers. The bold figures are the highest in each column.

Part of Speech (POS Bi) As for POS, but we consider the adjacent pairs of part of speech tags as a feature.

SentiWordNet score (senti) We used the SentiWordNet (Baccianella et al., 2010) lexical resource to assign scores for each word based on three sentiments i.e positive, negative and objective respectively. The positive, negative and objective scores sum up to 1. We use the individual lemmatized words in a statement as an input and obtain the scores for each of them. For each lemmatized word, we obtain the difference between their positive and negative score. We add up the computed scores for all the words present in a statement and average it which gives the overall statement score as a feature.

Noun-Adjective patterns Both the statements in general expression cues and specific expressions cues contain combinations of noun and adjective pairs. For every noun present in the text, each combination of adjective was considered as a noun-adjective pair feature.

In addition to these features, each token is paired with its *term frequency*, defined as:

$$\frac{\textit{number of occurrences of a token}}{\textit{total number of tokens}} \quad (1)$$

Thus rather than a statement containing several instances of a common term (like “the”), it will contain a single instance, plus the term frequency.

4 Experiments

Having covered the features we considered, this section describes the experimental setup and the results we obtained. We used the scikit-learn toolkit library to conduct three experiments.

4.1 Classifier

The first experiment was to train different classifiers — Logistic Regression, Multinomial Naive Bayes and Linear SVM — using the basic unigrams and bigrams as features and determine the best classifier for our task. Table 2 gives the 5 cross-fold validation F1-score results with the linear SVM classifier performing best. We used the scikit-learn GridSearchcv function to perform an evaluative search on our data to get the best regularization parameter value for the linear SVM classifier. This was C=10.

4.2 Training data

Having picked the classifier, the second experiment was to find the best mix of data to train on. This is an important step to take when we have data that is as unbalanced, in terms of the number elements of each type of data we are classifying, as the data we have here. The manually annotated statements were divided into two sets — training set and test set. We collected 30 explicit and 150 implicit opinions as the test set. These were not used in training. We gathered the remaining 100 explicit opinions and created a training set using these statements and a variable-sized set of implicit opinions. For each such training set, we ran a 5 fold cross-validation and also tested it against the test set that we had created. We use the linear SVM classifier to train and test the data with the basic features (unigrams and bigrams respectively). The mean F1-scores for the cross-validation on different train sets and the F1-scores on the test set for both explicit and implicit opinions are shown in Figure 2. The plot also contains the false positive rate for the test set with respect to different training sets.

4.3 Features

Given the results of the second experiment, we can identify the best size of training set, in terms of the number of explicit and implicit opinions. Considering Figure 2, we see that a training set containing 100 explicit and 250 implicit opinions will be sufficient. With this mix, the false positive rate is close to minimum, and the performance on the test set is close to maximum. We then carried out a third experiment to find the best set of features to identify the stances. To do this we ran a 5 fold cross-validation on the training set using the all the features described in Section 3.2 — in other words we expanded the feature set from just unigrams and bigrams — using both individual features and sets of features. We also performed the same experiment using these different features on

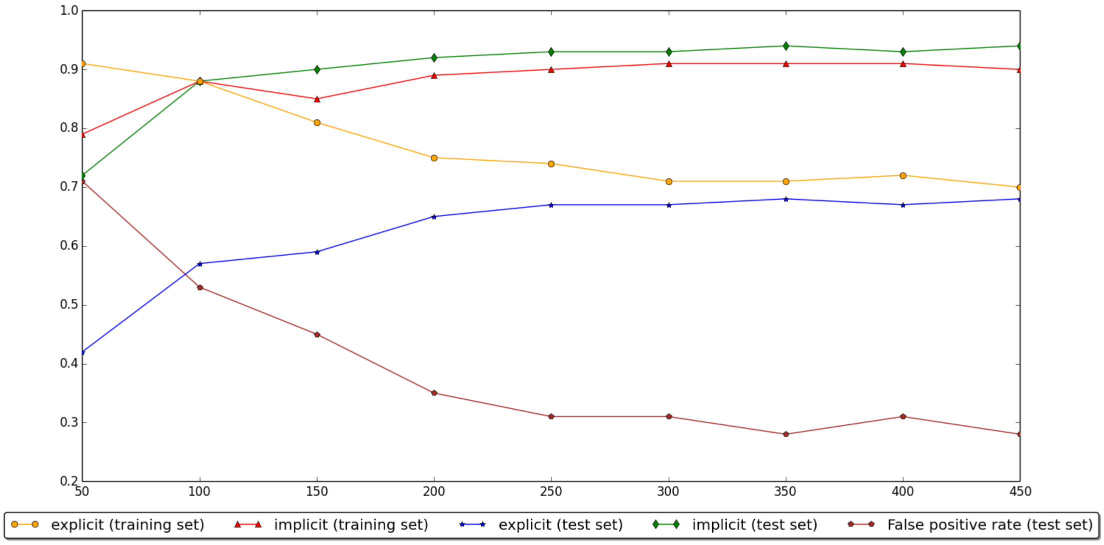

Figure 2: The results of the experiment to identify the best mix of explicit and implicit stances for training. The training set contained 100 explicit stances and as many implicit stances as indicated on the x-axis. The graph shows the cross-validation F1 scores for the training sets, and the corresponding F1 scores obtained on the test set. False positive rates for the test set with respect to each training set are also plotted.

the test set.

4.4 Results

Table 3 contains the results for the third experiment. The best performance results are highlighted — the highest values in each of the first four columns (classification accuracy) are in bold, as is the lowest value in the final column (false positive rate). We see that the basic features, unigrams and bigrams, give good results for both the cross-validation of the training set and for the test set. We also see that while the sentiment of each statement was useful in determining whether a statement is an opinion (and thus the statement is included in our data), sentiment does not help in distinguishing the explicit stance from the implicit stance which is why there is no improvement with the SentiWordNet scores as features. This is because both positive and negative statements can be either implicit or explicit. In contrast, the special features that include the noun-adjective patterns along with unigrams and bigrams gave the best performance for the test set, and also produced the lowest false positive rate.

4.5 Top 10 features

The linear SVM classifier gives the best performance results and thus we use the weights of the classifier for identifying the most important features in the data. The classifier is based on the following decision function:

$$y(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} + b \tag{2}$$

where w is the weight vector and b is the bias value. Support vectors represent those weight weight vectors that are non-zero, and we can use these to obtain the most important features. Table 4 gives the most important 10 features identified in this way for both explicit and implicit opinions.

5 Conclusion

In this paper, we focus on a specific domain of online reviews and propose an approach that can help in enthymemes detection and reconstruction. Online reviews contain aspect-based statements that can be considered as stances representing for/against views of the reviewer about the aspects present in the product or service and the product/service itself. The proposed approach is a two-step approach that detects the type of stances based on the contextual features, which can then be converted into explicit premises, and these premises with missing information represents enthymemes. We also propose a solution using the available data to represent common knowledge that can fill in the missing information to

Features	*Training set*		*Test set*		
	F1 Score		*F1 Score*		*False positive rate*
	Explicit	Implicit	Explicit	Implicit	
Baseline	0.73	0.88	0.67	0.92	0.41
Uni	0.74	**0.90**	0.65	0.92	0.40
Uni +Bi	0.75	**0.90**	0.70	**0.94**	0.30
Uni + Bi + POS	0.74	0.90	0.68	0.93	0.33
Uni + Bi + POS + POS Bi	0.72	0.89	**0.71**	**0.94**	**0.26**
Uni + Bi + POS + POS Bi + Senti	0.66	0.89	0.68	0.94	0.3
Uni + Bi + POS + Senti	0.73	**0.90**	0.70	**0.94**	0.31
Uni + Bi + Noun-Adj patterns	**0.77**	**0.90**	**0.72**	**0.94**	**0.26**

Table 3: The results of the experiment to identify the best feature set. The table gives the F1 scores for training set and test set using different sets of features. False positive rate on the test set is also listed. All results were obtained using the Linear SVM classifier except the baseline classifier. The bold numbers are the highest classification rates in each column, or the lowest false positive rate for the column, as appropriate.

Explicit	Weight	Implicit	Weight
excellent	4.18	Adj + Noun	-1.26
location	3.43	the hotel	-1.25
great	2.55	nice star	-1.23
experience	2.02	fairly	-1.08
recommend	1.91	hotel the	-1.03
was excellent	1.84	helpful + Noun	-0.96
hotel	1.61	location but	-0.95
service	1.48	hotel with	-0.94
extremely	1.45	advice stay	-0.94
was great	1.43	hotel stars	-0.94

Table 4: List of the 10 most important features present in explicit and implicit stances with their weights

complete the arguments. The first-step requires automatic detection of the stance types — explicit and implicit, which we have covered in this paper. We use a supervised approach to classify the stances using a linear SVM classifier, the best performance results on the test set with a macro-averaged F1-scores of 0.72 and 0.94 for explicit and implicit stances respectively. These identified implicit stances are then explicit premises of either complete arguments or enthymemes. (If they are premises of complete arguments, there are other, additional premises.) The identified explicit stances can then represent common knowledge information for the implicit premises, thus becoming explicit premises to fill in the gap present in the respective enthymemes.

6 Future work

The next steps in this work take us closer to the automatic reconstruction of enthymemes. The first of these steps is to look to refine our identification of explicit premises (and thus complete arguments, circumventing the need for enthymeme reconstruction). The idea here is that we believe that since we are currently looking only at the sentence level, we may be misclassifying some sentences as expressing implicit opinions when they include both implicit and explicit opinions. To refine the classification, we need to examine sub-sentential clauses of the sentences in the reviews to identify if any of them express explicit opinions. If no explicit opinions are expressed in any of the sub-sentential clauses, then the whole sentence can be correctly classified as a implicit opinion, and along with the predefined conclusion will become an enthymeme. The second of the steps towards enthymeme reconstruction is to look to use related explicit opinions to complete enthymemes, as discussed in Section 3.1. Here the distinction between general and specific opinions becomes important, since explicit general opinions might be combined with any implicit opinion about an aspect in the same aspect category, while explicit specific opinions can only be combined with implicit opinions that relate to the same aspect. Effective combination of explicit general opinions with related implicit opinions requires a detailed model which expresses what "related" means for the relevant domain. We expect the development of this model to be as time-consuming as all work formalising a domain. Another issue in enthymeme reconstruction is evaluating the output of the process. Identifying whether a given enthymeme has been successfully turned into a complete argument is a highly subjective task, which will likely require careful human evaluation. Performing this at a suitable scale will be challenging.

References

Leila Amgoud and Henri Prade. 2012. Can AI models capture natural language argumentation? In *IJCINI'12*, pages 19–32.

Leila Amgoud, Philippe Besnard, and Anthony Hunter. 2015. Representing and reasoning about arguments mined from texts and dialogues. In *ECSQARU'15*, pages 60–71.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC'10*, pages 2200–2204.

Elizabeth Black and Anthony Hunter. 2012. A relevance-theoretic framework for constructing and deconstructing enthymemes. *J. Log. Comput.*, 22(1):55–78.

Filip Boltužić and Jan Šnajder. 2014. Back up your stance: Recognizing arguments in online discussions. In *ACL'14*, pages 49–58.

Elena Cabrio and Serena Villata. 2012. Combining textual entailment and argumentation theory for supporting online debates interactions. In *ACL'12*, pages 208–212.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *ACL'11*, pages 987–996.

Maria Paz Garcia-Villalba and Patrick Saint-Dizier. 2012. A framework to extract arguments in opinion texts. In *IJCINI'12*, volume 6, pages 62–87.

Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui. 2014. Analyzing argumentative discourse units in online interactions. In *ACL'14*, pages 39–48.

Ivan Habernal, Judith Eckle-Kohler, and Iryna Gurevych. 2014. Argumentation mining on the web from information seeking perspective. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, pages 26–39.

Seyed Ali Hosseini, Sanjay Modgil, and Odinaldo Rodrigues. 2014. Enthymeme construction in dialogues using shared knowledge. In *COMMA'14*, pages 325–332.

Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI'04*, pages 755–760.

Angeliki Lazaridou, Ivan Titov, and Caroline Sporleder. 2013. A Bayesian model for joint unsupervised induction of sentiment, aspect and discourse representations. In *ACL'13*, pages 1630–1639.

Marco Lippi and Paolo Torroni. 2015a. Argument mining: A machine learning perspective. In *TAFA'15*, pages 163–176.

Marco Lippi and Paolo Torroni. 2015b. Context-independent claim detection for argument mining. In *IJCAI'15*, pages 185–191.

Clare Llewellyn, Claire Grover, Jon Oberlander, and Ewan Klein. 2014. Re-using an argument corpus to aid in the curation of social media collections. In *LREC'14*, pages 462–468.

Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In *ACL'14*, pages 29–38.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to argumentation mining in texts: A survey. In *IJCINI'13*, volume 7, pages 1–31.

Iyad Rahwan and Guillermo R. Simari, editors. 2009. *Argumentation in Artificial Intelligence*. Springer Verlag, Berlin, Germany.

Stephen Toulmin. 1958. *The Uses of Argument*. Cambridge University Press, Cambridge, England.

Frans H. van Eemeren, Rob Grootendorst, Francisca S. Henkemans, J. Anthony Blair, Ralph H. Johnson, Erik C. W. Krabbe, Christian Plantin, Douglas N. Walton, Charles A. Willard, John Woods, and David Zarefsky. 1996. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Lawrence Erlbaum Associates, Mahwah, NJ.

Marc Vincent and Grégoire Winterstein. 2014. Argumentative insights from an opinion classification task on a French corpus. In *New Frontiers in Artificial Intelligence*, pages 125–140.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, and Gregor Engels. 2014a. Modeling review argumentation for robust sentiment analysis. In *ICCL'14*, pages 553–564.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, Gregor Engels, and Tsvetomira Palakarska. 2014b. A review corpus for argumentation analysis. In *ICCLITP'14*, pages 115–127.

Douglas N. Walton and Erik C. W. Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, USA.

Douglas N. Walton. 2008. The three bases for the enthymeme: A dialogical theory. *J. Applied Logic*, 6(3):361–379.

Douglas Walton. 2010. The structure of argumentation in health product messages. *Argument & Computation*, 1(3):179–198.

Adam Wyner, Jodi Schneider, Katie Atkinson, and Trevor J. M. Bench-Capon. 2012. Semi-automated argumentative analysis of online product reviews. In *COMMA'12*, pages 43–50.

Nefise Yaglikci and Paolo Torroni. 2014. Microdebates app for Android: A tool for participating in argumentative online debates using a handheld device. In *ICTAI'14*, pages 792–799.