# Combining Argument Mining Techniques

**John Lawrence**
School of Computing
University of Dundee
UK
j.lawrence@dundee.ac.uk

**Chris Reed**
School of Computing
University of Dundee
UK
c.a.reed@dundee.ac.uk

## Abstract

In this paper, we look at three different methods of extracting the argumentative structure from a piece of natural language text. These methods cover linguistic features, changes in the topic being discussed and a supervised machine learning approach to identify the components of argumentation schemes, patterns of human reasoning which have been detailed extensively in philosophy and psychology. For each of these approaches we achieve results comparable to those previously reported, whilst at the same time achieving a more detailed argument structure. Finally, we use the results from these individual techniques to apply them in combination, further improving the argument structure identification.

## 1 Introduction

The continuing growth in the volume of data which we produce has driven efforts to unlock the wealth of information this data contains. Automatic techniques such as Opinion Mining and Sentiment Analysis (Liu, 2010) allow us to determine the views expressed in a piece of textual data, for example, whether a product review is positive or negative. Existing techniques struggle, however, to identify more complex structural relationships between concepts.

Argument Mining[1] is the automatic identification of the argumentative structure contained within a piece of natural language text. By automatically identifying this structure and its associated premises

and conclusions, we are able to tell not just *what* views are being expressed, but also *why* those particular views are held.

The desire to achieve this deeper understanding of the views which people express has led to the recent rapid growth in the Argument Mining field (2014 saw the first ACL workshop on the topic in Baltimore[2] and meetings dedicated to the topic in both Warsaw[3] and Dundee[4]). A range of techniques have been applied to this problem, including supervised machine learning (starting with (Moens et al., 2007)) and topic modelling ((Lawrence et al., 2014)) as well as purely linguistic methods (such as (Villalba and Saint-Dizier, 2012)); however, little work has currently been carried out to bring these techniques together.

In this paper, we look at three individual argument mining approaches. Firstly, we look at using the presence of discourse indicators, linguistic expressions of the relationship between statements, to determine relationships between the propositions in a piece of text. We then move on to look at a topic based approach. Investigating how changes in the topic being discussed relate to the argumentative structure being expressed. Finally, we implement a supervised machine learning approach based on argumentation schemes (Walton et al., 2008), enabling us to not only identify premises and conclusions, but to determine how exactly these argument components are working together.

Based on the results from the individual imple-

---

[1]Sometimes also referred to as Argumentation Mining

mentations, we combine these approaches, taking into account the strengths and weaknesses of each to improve the accuracy of the resulting argument structure.

## 2 Dataset

One of the challenges faced by current approaches to argument mining is the lack of large quantities of appropriately annotated arguments to serve as training and test data. Several recent efforts have been made to improve this situation by the creation of corpora across a range of different domains; however, to apply each of the techniques previously mentioned in combination means that we are limited to analysed data containing complete argumentation scheme specifications and provided along with the original text.

Although there are a number of argument analysis tools (such as Araucaria (Reed and Rowe, 2004), Carneades (Gordon et al., 2007), Rationale (van Gelder, 2007) and OVA (Bex et al., 2013)) which allow the analyst to identify the argumentation scheme related to a particular argumentative structure, the vast majority of analyses which are produced using these tools do not include this information. For example, less than 10% of the OVA analyses contained in AIFdb (Lawrence et al., 2012) include any scheme structure.

AIFdb still offers the largest annotated dataset available, containing the complete Araucaria corpus (Reed et al., 2008) used by previous argumentation scheme studies and supplemented by analyses from a range of other sources. Limiting the data to analyses containing complete scheme specifications and for which the original text corresponds directly to the analysis (with no re-construction or enthymematic content (Hitchcock, 1985) added) leaves us with 78 complete analyses (comprised of 404 propositions and 4,137 words), including 47 examples of the argument from expert opinion scheme and 31 examples of argument from positive consequences (these schemes are discussed in Section 5.)

## 3 Discourse Indicators

The first approach which we present is that of using discourse indicators to determine the argumentative connections between adjacent propositions in

| Relation Type | Words |
|---|---|
| Support | because, therefore, after, for, since, when, assuming, so, accordingly, thus, hence, then, consequently |
| Conflict | however, but, though, except, not, never, no, whereas, nonetheless, yet, despite |

Table 1: Discourse indicators used to determine propositional connections

a piece of text. Discourse indicators are explicitly stated linguistic expressions of the relationship between statements (Webber et al., 2011), and, when present, can provide a clear indication of its argumentative structure. For example, if we take the sentence "Britain should disarm because it would set a good example for other countries", then this can be split into two separate propositions "Britain should disarm" and "it (disarming) would set a good example for other countries". The presence of the word "because" between these two propositions clearly tells us that the second is a reason for the first.

Discourse indicators have been previously used as a component of argument mining techniques for example in (Stab and Gurevych, 2014), indicators are used as a feature in multiclass classification of argument components, with each clause classified as a major claim, claim, premise or non-argumentative. Similar indicators are used in (Wyner et al., 2012), along with domain terminology (e.g. camera names and properties) to highlight potential argumentative sections of online product reviews. By looking at discourse indicators in isolation, however, we aim to determine their ability to be used on their own as an argument mining method.

There are many different ways in which indicators can appear, and a wide range of relations which they can suggest (Knott, 1996). We limit our search here to specific terms appearing between two sequential propositions in the original text. These terms are split into two groups, indicating support and attack relations between the propositions. A list of these terms can be seen in Table 1.

| | p | r | f1 |
|---|---|---|---|
| Discourse Indicators | 0.89 | 0.04 | 0.07 |

Table 2: Comparison of the connections between propositions determined by discourse indicators and manual analysis

By performing a simple search for these terms across the text of each item in our corpus, we were able to determine suggested connections between propositions and compare these to the manual analyses. The results of this comparison can be seen in Table 2. In this case we look at the connections between the component propositions in the manually analysed argument structure (385 connections in total), and consider a connection to have been correctly identified if a discourse indicator tells us that two propositions are connected, and that the relation between them (support or attack) is the same as that in the manual analysis.

The results clearly show that, when discourse indicators are present in the text, they give a strong indication of the connection between propositions (precision of 0.89); however, the low frequency with which they can be found means that they fail to help identify the vast majority of connections (recall of 0.04). Additionally, the approach we use here considers only those discourse indicators found between pairs of consecutive propositions and, as such, is unable to identify connected propositions which are further apart in the text. Because of this, discourse indicators may provide a useful component in an argument mining approach, but, unless supplemented by other methods, are inadequate for identifying even a small percentage of the argumentative structure.

## 4 Topical Similarity

The next approach which we consider looks at how the changes of topic in a piece text relate to the argumentative structure contained within it. This method is similar to that presented in (Lawrence et al., 2014), where it is assumed firstly that the argument structure to be determined can be represented as a tree, and secondly, that this tree is generated depth first. That is, the conclusion is given first and then a line of reasoning is followed supporting this conclusion. Once that line of reasoning is exhausted, the argument moves back up the tree to support one of the previously made points. If the current point is not related to any of those made previously, then it is assumed to be unconnected.

Based on these assumptions we can determine the structure by looking at how similar the topic of each proposition is to its predecessor. If they are similar, then we assume that they are connected and the line of reasoning is being followed. If they are not sufficiently similar, then we first consider whether we are moving back up the tree, and compare the current proposition to all of those made previously and connect it to the most topically similar previous point. Finally, if the current point is not related to any of those made previously, then it is assumed to be unconnected to the existing structure.

Lawrence et al. perform these comparisons using a Latent Dirichlet Allocation (LDA) topic model, in our case however, the argument structures we are working with are from much shorter pieces of text and as such generating LDA topic models from them is not feasible. Instead we look at the semantic similarity of propositions. We use WordNet[5] to determine the similarity between the synsets of each word in the first proposition and each word in the second. This relatedness score is inversely proportional to the number of nodes along the shortest path between the synsets. The shortest possible path occurs when the two synsets are the same, in which case the length is 1, and thus, the maximum relatedness value is 1. We then look at the maximum of these values in order to pair a word in the first proposition to one in the second, and finally average the values for each word to give a relatedness score for the proposition pair between 0 and 1. Similarly to in (Lawrence et al., 2014), the threshold required for two propositions to be considered similar can be adjusted, altering the output structure, with a lower threshold giving more direct connections and a higher threshold greater branching and more unconnected components.

The results of performing this process using a threshold of 0.2, are shown in Table 3, and an example of the output structure can be seen in Figure 1.
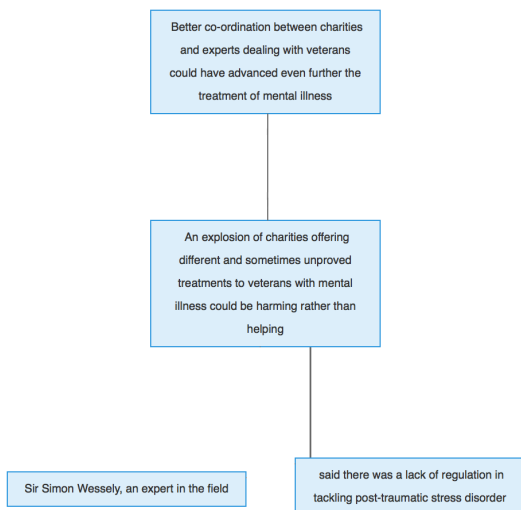
---

[5] http://wordnet.princeton.edu/

| | p | r | f1 |
|---|---|---|---|
| Non-directed | 0.82 | 0.56 | 0.67 |
| Current to previous | 0.63 | 0.43 | 0.51 |
| Previous to current | 0.19 | 0.13 | 0.15 |

Table 3: Topic Similarity Edge Predictions

For the results in Table 3, we consider a connection to have been correctly identified if there is any connection between the propositions in the manual analysis, regardless of direction or type. The standard output we obtain does not give any indication of the directionality of the connection between propositions, and these results are given in the first row of the table. The other two rows show the results obtained by assuming that these connections are always in one direction or another i.e. that the connection always goes from the current proposition to its predecessor or vice-versa.

Figure 1: Topic Structure



The results for non-directed connections are encouraging, as with the discourse indicators, precision (0.82) is higher than recall (0.56) suggesting that although this method may fail to find all connections, those that it does find can generally be viewed as highly likely. We can also see that the assumption of directionality from the current proposition to a previous proposition gives much better results than the other way around, suggesting that generally

when a point is made it is made to support (or attack) something previously stated.

## 5   Argumentation Scheme Structure

Finally, we consider using a supervised machine learning approach to classify argument components and determine the connections between them. One of the first attempts to use this kind of classification is presented in (Moens et al., 2007), where a text is first split into sentences and then features of each sentence are used to classify them as "Argument" or "Non-Argument". This approach was built upon in (Palau and Moens, 2009), where each argument sentence is additionally classified as either a premise or conclusion. Our approach instead uses argumentation schemes (Walton et al., 2008), common patterns of human reasoning, enabling us to not only identify premise and conclusion relationships, but to gain a deeper understanding of how these argument components are working together.

The concept of automatically identifying argumentation schemes was first discussed in (Walton, 2011) and (Feng and Hirst, 2011). Walton proposes a six-stage approach to identifying arguments and their schemes. The approach suggests first identifying the arguments within the text and then fitting these to a list of specific known schemes. A similar methodology was implemented by Feng & Hirst, who produced classifiers to assign pre-determined argument structures as one in a list of the most common argumentation schemes.

The main challenge faced by this approach is the need to have already identified, not just that an argument is taking place, but its premises, conclusion and exact structure before a scheme can be assigned. By instead looking at the features of each component part of a scheme, we are able to overcome this requirement and identify parts of schemes in completely unanalysed text. Once these scheme components have been identified, we are able to group them together into specific scheme instances and thus obtain a complete understanding of the arguments being made.

Several attempts have been made to identify and classify the most commonly used schematic structures (Hastings, 1963; Perelman and Olbrechts-Tyteca, 1969; Kienpointner, 1992; Pollock, 1995;

Walton, 1996; Grennan, 1997; Katzav and Reed, 2004; Walton et al., 2008), though the most commonly used scheme set in analysis is that given by Walton. Here we look at two of Walton's schemes, Expert Opinion and Positive Consequences. Each scheme takes the form of a number of premises which work together to support a conclusion (the structure of the two schemes used can be seen in Table 4.)

---

**Expert Opinion**
*Premise:* Source E is an expert in subject domain S containing proposition A [FieldExpertise]
*Premise:* E asserts that proposition A is true (false) [KnowledgeAssertion]
*Conclusion:* A is true (false) [KnowledgePosition]

---

**Positive Consequences**
*Premise:* If A is brought about, then good consequences will (may plausibly) occur [PositiveConsequences]
*Conclusion:* Therefore, A should be brought about [EncouragedAction]

---

Table 4: Argumentation schemes

The features of these common patterns of argument provide us with a way in which to both identify that an argument is being made and determine its structure. By identifying the individual components of a scheme, we are able to identify the presence of a particular scheme from only a list of the propositions contained within the text. In order to accomplish this, one-against-others classification is used to identify propositions of each type from a set of completely unstructured propositions. Being able to successfully perform this task for even one of the proposition types from each scheme allows us to discover areas of the text where the corresponding scheme is being used.

This classification was performed with a Naïve Bayes classifier implemented using the *scikit-learn*[6] Python module for machine learning, with the features described in Table 5. Part Of Speech (POS)

tagging was performed using the Python NLTK[7] POS-tagger and the frequencies of each tag added as individual features. The similarity feature was added to extend the information given by unigrams to include an indication of whether a proposition contains words similar to a pre-defined set of keywords. The keywords used for each type are shown in Table 6, and are based on the scheme definitions from Table 4 by manually identifying the key terms in each scheme component. Similarity scores were calculated using WordNet[8] to determine the maximum similarity between the synsets of the keywords and each word in the proposition. The maximum score for the words in the proposition was then added as a feature value, indicating the semantic relatedness of the proposition to the keyword.

| Feature | Description |
|---|---|
| Unigrams | Each word in the proposition |
| Bigrams | Each pair of successive words |
| Length | The number of words in the proposition |
| AvgWLength | The average length of words in the proposition |
| POS | The parts of speech contained in the proposition |
| Punctuation | The presence of certain punctuation characters, for example " " indicating a quote |
| Similarity | The maximum similarity of a word in the proposition to pre-defined words corresponding to each proposition type |

Table 5: Features used for scheme component classification

Table 7 shows the precision, recall and F-score obtained for each proposition type. The results show that even for a scheme where the classification of one proposition type is less successful, the results for the other types are better. If we consider being able to correctly identify at least one proposition type, then our results give F-scores of 0.93 and

---

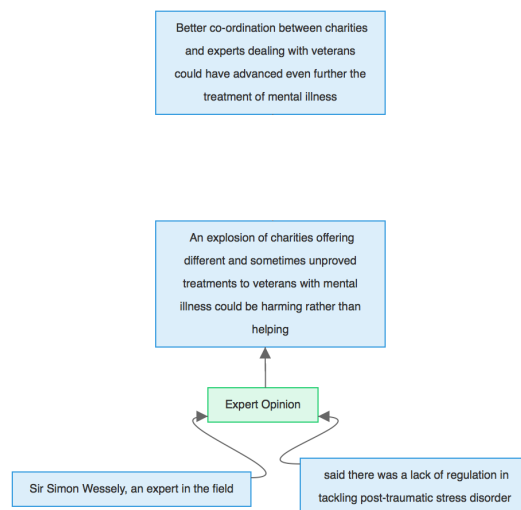| Type | Keywords |
|---|---|
| **Expert Opinion** | |
| FieldExpertise | expert, experienced, skilled |
| KnowledgeAssertion | said |
| KnowledgePosition | be (to be) |
| **Positive Consequences** | |
| PositiveConsequences | occur, happen |
| EncouragedAction | should, must |

Table 6: Keywords used for each proposition type

0.75 for locating an occurrence of each scheme type considered. This compares favourably with (Feng and Hirst, 2011), where the occurrence of a particular argumentation scheme was identified with accuracies of between 62.9% and 90.8%. Furthermore, Feng & Hirst's results only considered spans of text that were already known to contain a scheme of some type and required a prior understanding of the argumentative structure contained within the text, whereas the approach presented here does not have either of these requirements.

| | p | r | f1 |
|---|---|---|---|
| **Expert Opinion** | | | |
| FieldExpertise | 0.95 | 0.90 | 0.93 |
| KnowledgeAssertion | 0.74 | 1.00 | 0.83 |
| KnowledgePosition | 0.93 | 0.55 | 0.62 |
| **Positive Consequences** | | | |
| PositiveConsequences | 0.75 | 0.67 | 0.71 |
| EncouragedAction | 0.86 | 0.67 | 0.75 |

Table 7: Classifying scheme components

By looking further at each set of three propositions contained within the text, we can locate areas where all of the component parts of a scheme occur. When these are found, we can assume that a particular scheme is being used in the text and assign each of its component parts to their respective role. This gives us an automatically identified structure as shown in Figure 2, where we can see that the component parts of the scheme are completely identified, but the remaining proposition is left unconnected.

Figure 2: Scheme Structure



## 6 Combined Techniques

Having looked at three separate methods for automatically determining argument structure, we now consider how these approaches can be combined to give more accurate results than those previously achieved.

In order to investigate this, we tested a fixed subset of our corpus containing eight analyses, containing 36 pairs of connected propositions which we aim to identify. The remainder is used as training data for the supervised learning approach used to identify scheme instances. The use of such a fixed dataset allows us to compare and combine the computational methods used for discourse indicators and topical similarity with the supervised learning method used for scheme identification. The results of applying each approach separately are given in the first part of Table 8. In each case, the precision, recall and f1-score is given for how well each method manages to identify the connections between propositions in the set of analyses.

We can see from the results, that again the precision for discourse indicators is high, but that the recall is low. This suggests that where indicators are found, they are the most reliable method of determining a connection.

The precision for using schematic structures is also high (0.82), though again the recall is lower. In this case, this is due to the fact that although

this method can determine well the links between components in an argumentation scheme instance it gives no indication as to how the other propositions are connected.

Finally, topic similarity gives the poorest results, suggesting that this method be used to supplement the others, but that it is not capable of giving a good indication of the structure on its own.

Based on these results, we combine the methods as follows: firstly, if discourse indicators are present then they are assumed to be a correct indication of a connection; next, we identify scheme instances and connect the component parts in accordance with the scheme structure; and finally we look at the topic similarity and use this to connect any propositions that have previously been left out to the already identified structure. This combination of approaches is used to take advantage of the strengths of each. As previously discussed, discourse indicators are rare, but provide a very good indication of connectedness when they do occur, and as such applying this method first gives us a base of propositions that are almost certainly correctly connected. Scheme identification offers the next best precision, and so is applied next. Finally, although topical similarity does not perform as well as scheme identification and does not give an indication of direction or type of connection, it allows us to connect those propositions which are not part of a scheme instance.

Carrying out this combined approach gives us the results shown in the last row of Table 8. Again, the results are based on correctly identified connections when compared to the manual analysis. We can see that by combining the methods, accuracy is substantially improved over any one individual method.
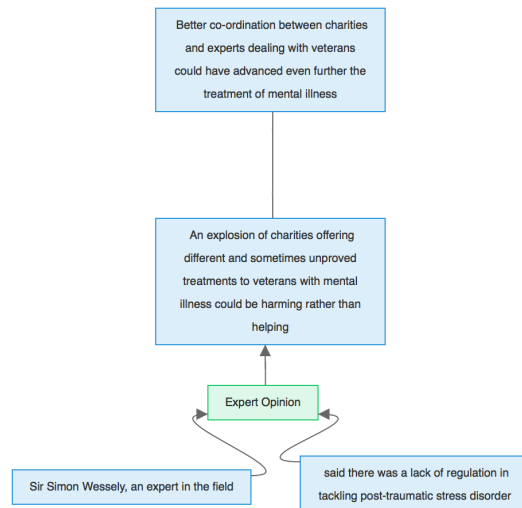
An example of the resulting structure obtained using this combined approach can be seen in Figure 3. If we compare this to a manual analysis of the same text (Figure 4), we can see that the structures are almost identical, differing only in the fact that the nature of the relationship between the premises "An explosion of charities offering different and sometimes unproved treatments to veterans with mental illness could be harming rather than helping" and "Better co-ordination between charities and experts dealing with veterans could have advanced even further the treatment of mental illness" is still unknown. We could make the further assumption, as detailed

|  | p | r | f1 |
|---|---|---|---|
| Discourse Indicators | 1.00 | 0.08 | 0.15 |
| Topic Similarity | 0.70 | 0.54 | 0.61 |
| Schematic Structure | 0.82 | 0.69 | 0.75 |
| **Combined Methods** | **0.91** | **0.77** | **0.83** |

Table 8: Identifying Argument Structure

in section 3 that the second proposition supports or attacks the first as it appears later in the text, and in so doing obtain a picture almost identical to that produced by manual analysis.
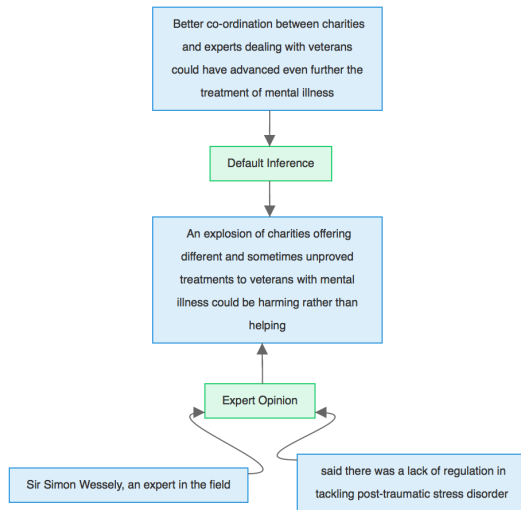


Figure 3: Combined

## 6.1 Proposition Boundary Learning

Until now, we have considered determining the argumentative structure from a piece of text which has already been split into its component propositions; however, in order to be able to extract structure from natural language, we must also be able to perform this segmentation automatically.

Text segmentation can be considered as the identification of a form of Elementary Discourse Units (EDUs), non-overlapping spans of text corresponding to the minimal units of discourse. (Peldszus and Stede, 2013) refers to these argument segments as 'Argumentative Discourse Units' (ADUs), and defines an ADU as a 'minimal unit of analysis', pointing out that an ADU may not always be as small as

Figure 4: Manual Analysis



an EDU, for example, 'when two EDUs are joined by some coherence relation that is irrelevant for argumentation, the resulting complex might be the better ADU'.

We now look at how well our combined approach performs on text which is segmented using Propositional Boundary Learning. This technique, introduced in (Lawrence et al., 2014), uses two naïve Bayes classifiers, one to determine the first word of a proposition and one to determine the last. The classifiers are trained using a set of manually annotated training data. The text given is first split into words and a list of features calculated for each word. The features used are given below:

**word** The word itself.

**length** Length of the word.

**before** The word before.

**after** The word after. Punctuation is treated as a separate word so, for example, the last word in a sentence may have an after feature of '.'.

**pos** Part of speech as identified by the Python Natural Language Toolkit POS tagger[9].

Once the classifiers have been trained, these same features are then determined for each word in the

---

test data and each word classified as either 'start' or 'end'. Once the classification has taken place, the individual starts and ends are matched to determine propositions, using their calculated probabilities to resolve situations where a start is not followed by an end (i.e. where the length of the proposition text to be segmented is ambiguous). Using this method, Lawrence et al. report a 32% increase in accuracy over simply segmenting the text into sentences, when compared to argumentative spans identified by a manual analysis process.
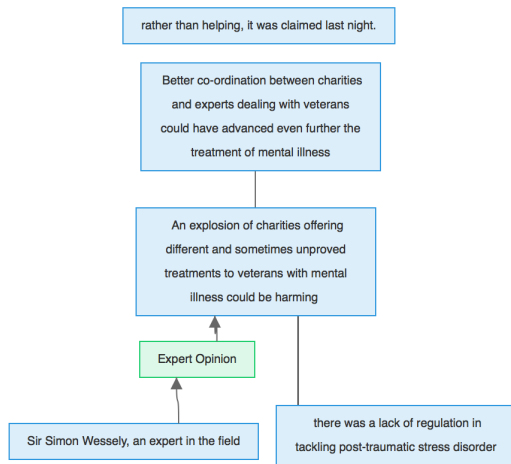
Performing this process on the text from the example in Figure 4 we obtain a list of five propositions:

1. An explosion of charities offering different and sometimes unproved treatments to veterans with mental illness could be harming

2. rather than helping, it was claimed last night.

3. Sir Simon Wessely, an expert in the field

4. there was a lack of regulation in tackling post-traumatic stress disorder

5. Better co-ordination between charities and experts dealing with veterans could have advanced even further the treatment of mental illness

Using these propositions as input to our scheme component classification identifies proposition 1 as an Expert Opinion KnowledgePosition, and proposition 3 as FieldExpertise, though fails to identify any of the propositions as a KnowledgeAssertion. Additionally, applying topical similarity to these propositions results in suggested connections from 1 to 4 and from 1 to 5.

The output from this process can be seen in Figure 5. Although this structure is not identical to that obtained using manually identified propositions, the similarity is strong and suggests that with improvement in the automatic segmentation of text into argument components, these techniques could be used to give a very good approximation of manual argument analysis.

Figure 5: Automatically Identified Propositions



## 7 Conclusion

We have implemented three separate argument mining techniques and for each achieved results comparable to those previously reported for similar methods.

In (Feng and Hirst, 2011), the occurrence of a particular argumentation scheme was identified with accuracies of between 62.9% and 90.8% for one-against-others classification. However, these results only considered spans of text that were already known to contain a scheme of some type and required a prior understanding of the argumentative structure contained within the text. By considering the features of the individual types of premise and conclusion that comprise a scheme, we achieved similar performance (F-scores between 0.75 and 0.93) for identifying at least one component part of a scheme.

We have shown that, although there are strengths and weaknesses to each of these techniques, by using them in combination we can achieve results that are remarkably close to a manual analysis of the same text. The accuracy we achieve for determining connections between propositions (f-score of 0.83) compares favourably with other results from the argument mining field. For example, in (Palau and Moens, 2009) sentences were classified as either premise (F-score, 0.68) or conclusion (F-score, 0.74), but in the case of our combined results, not only are we able to determine the premises and conclusion of an argument, but its schematic structure and the precise roles that each of the premises play in supporting the conclusion.

Finally, we have shown that by using Propositional Boundary Learning as an initial step in this process, we are able to take a piece of natural language text and automatically produce an argument analysis that still remains close to that determined by a manual analyst.

As the field of argument mining continues its dramatic growth, there are an increasing number of strategies being explored for contributing to the task. In building a simple algorithm for combining these techniques, we have demonstrated that it is quite possible to yield significant increases in performance over any single approach. This is in contrast to some other areas of text mining and machine learning in general, where combining different techniques is either not possible or else yields only marginal improvements. It seems likely that this strong complementarity in techniques for argument mining reflects a deep diversity not just in the techniques but in the underlying insights and strategies for identifying argument, which in turn reflects the breadth of philosophical, linguistic and psychological research in argumentation theory. We might hope as a consequence that as that research is increasingly tapped by algorithms for extracting various aspects of argument, so the combinations of algorithms become more sophisticated with ever better argument mining performance on unconstrained texts.

## Acknowledgments

# References

Floris Bex, John Lawrence, Mark Snaith, and Chris Reed. 2013. Implementing the argument web. *Communications of the ACM*, 56(10):66–73, Oct.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 987–996. Association for Computational Linguistics (ACL).

Thomas F Gordon, Henry Prakken, and Douglas Walton. 2007. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10):875–896.

Wayne Grennan. 1997. *Informal Logic: Issues and Techniques*. McGill-Queen's Press-MQUP.

Arthur C Hastings. 1963. *A Reformulation of the Modes of Reasoning in Argumentation*. Ph.D. thesis, Northwestern University.

David Hitchcock. 1985. Enthymematic arguments. *Informal Logic*, 7(2):289–98.

Joel Katzav and Chris Reed. 2004. On argumentation schemes and the natural classification of arguments. *Argumentation*, 18(2):239–259.

Manfred Kienpointner. 1992. *Alltagslogik: struktur und funktion von argumentationsmustern*. Frommann-Holzboog.

Alistair Knott. 1996. *A data-driven methodology for motivating a set of coherence relations*. Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh.

John Lawrence, Floris Bex, Chris Reed, and Mark Snaith. 2012. AIFdb: Infrastructure for the argument web. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 515–516.

John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft. 2014. Mining arguments from 19th century philosophical texts using topic based modelling. In *Proceedings of the First Workshop on Argumentation Mining*, pages 79–87, Baltimore, Maryland, June. Association for Computational Linguistics (ACL).

Bing Liu. 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666.

Marie-Francine Moens, Eric Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 225–230. ACM.

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. ACM.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.

Chaim Perelman and Lucie Olbrechts-Tyteca. 1969. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press.

John L Pollock. 1995. *Cognitive carpentry: A blueprint for how to build a person*. MIT Press.

Chris Reed and Glenn Rowe. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(4):961–980.

Chris Reed, Raquel Mochales Palau, Glenn Rowe, and Marie-Francine Moens. 2008. Language resources for studying argument. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC-2008)*, pages 91–100, Marrakech.

Christian Stab and Iryna Gurevych. 2014. Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 46–56, Doha, Qatar, October. Association for Computational Linguistics (ACL).

Tim van Gelder. 2007. The rationale for rationale. *Law, probability and risk*, 6(1-4):23–42.

M.G. Villalba and P. Saint-Dizier. 2012. Some facets of argument mining for opinion analysis. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 23–34.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Douglas Walton. 1996. *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Douglas Walton. 2011. Argument mining by applying argumentation schemes. *Studies in Logic*, 4(1):38–64.

Bonnie Webber, Markus Egg, and Valia Kordoni. 2011. Discourse structure and language technology. *Natural Language Engineering*, 18(4):437–490.

Adam Wyner, Jodi Schneider, Katie Atkinson, and Trevor Bench-Capon. 2012. Semi-automated argumentative analysis of online product reviews. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 43–50.