Open Research Online

The Open University's repository of research publications
and other research outputs

Summarising the points made in online political debates

Conference or Workshop Item

For guidance on citations see FAQs.

oro.open.ac.uk

Summarising the points made in online political debates

Charlie Egan
Computing Science
University of Aberdeen
`c.egan.12@aberdeen.ac.uk`

Advaith Siddharthan
Computing Science
University of Aberdeen
`advaith@abdn.ac.uk`

Adam Wyner
Computing Science
University of Aberdeen
`azwyner@abdn.ac.uk`

Abstract

Online communities host growing numbers of discussions amongst large groups of participants on all manner of topics. This user-generated content contains millions of statements of opinions and ideas. We propose an abstractive approach to summarize such argumentative discussions, making key content accessible through 'point' extraction, where a point is a verb and its syntactic arguments. Our approach uses both dependency parse information and verb case frames to identify and extract valid points, and generates an abstractive summary that discusses the key points being made in the debate. We performed a human evaluation of our approach using a corpus of online political debates and report significant improvements over a high-performing extractive summarizer.

1 Introduction

People increasingly engage in and contribute to online discussions and debates about topics in a range of areas, e.g. film, politics, consumer items, and science. Participants may make points and counterpoints, agreeing and disagreeing with others. These online argumentative discussions are an untapped resource of ideas. A high-level, summarised view of a discussion, grouping information and presenting points and counter-points, would be useful and interesting: retailers could analyse product reviews; consumers could zero in on what products to buy; and social scientists could gain insight on social treads. Yet, due to the size and complexity of the discussions and limitations of summarisers based on sentence extraction, much of the useful information in discussions is inaccessible.

In this paper, we propose a fully automatic and domain neutral unsupervised approach to abstractive summarisation which makes the key content of such discussions accessible. At the core of our approach is the notion of a 'point' - a short statement, derived from a verb and its syntactic arguments. Points (and counter-points) from across the corpus are analysed and clustered to derive a summary of the discussion. To evaluate our approach, we used a corpus of political debates (Walker et al., 2012), then compared summaries generated by our tool against a high-performing extractive summariser (Nenkova et al., 2006). We report that our summariser improves significantly on this extractive baseline.

2 Related Work

Text summarisation is a well established task in the field of NLP, with most systems based on sentence selection and scoring, with possibly some post-editing to shorten or fuse sentences (Nenkova and McKeown, 2011). The vast majority of systems have been developed for the news domain or on structured texts such as science (Teufel and Moens, 2002).

In related work on mailing list data, one approach clustered messages into subtopics and used centring to select sentences for an extractive summary (Newman and Blitzer, 2003). The concept of recurring and related subtopics has been highlighted (Zhou and Hovy, 2006) as being of greater importance to discussion summarisation than the summarisation of newswire data. In 'opinion summarisation', sentences have also been grouped based on the

feature discussed, to generate a summary of all the reviews for a product that minimised repetition (Hu and Liu, 2004). There has also been interest in the summarisation of subjective content in discussions (Hu and Liu, 2004; Lloret et al., 2009; Galley et al., 2004).

In addition to summarisation, our work is concerned with argumentation, which for our purposes relates to expressions for or against some textual content. Galley et al. (2004) used adjacency pairs to target utterances that had been classified as being an agreement or disagreement. Others have investigated arguments raised in online discussion (Boltuzic and Šnajder, 2015; Cabrio and Villata, 2012; Ghosh et al., 2014). A prominent example of argument extraction applies supervised learning methods to automatically identify claims and counter-claims from a Wikipedia corpus (Levy et al., 2014).

In this paper, we explore the intersection of text summarisation and argument. We implement a novel summarisation approach that extracts and organises information as points rather than sentences. It generates structured summaries of argumentative discussions based on relationships between points, such as counterpoints or co-occurring points.

3 Methods

Our summariser is based on three components. The first robustly identifies and extracts points from text, providing data for subsequent analysis. Given plain text from a discussion, we obtain (a) a pattern or signature that could be used to link points – regardless of their exact phrasing – and (b) a short readable extract that could be used to present the point to readers in a summary. A second component performs a number of refinements on the list of points such as removing meaningless points. The third component builds on these extracted points by connecting them in different ways (e.g., as point and counterpoint, or co-occurring points) to model the discussion. From this, it formulates a structured summary that we show to be useful to readers.

3.1 Point Extraction

We use the notion of a 'point' as the basis for our analysis – broadly speaking it is a verb

and its syntactic arguments. Points encapsulate both a human-readable 'extract' from the text as well as a pattern representing the core components that can be used to match and compare points. Extracts and patterns are stored as attributes in a key-value structure that represents a point.

Consider the sentence from a debate about abortion: *"I don't think so, an unborn* **child** *(however old)* **is** *not yet a* **human***."* Other sentences may also relate to this idea that a child is not human until it is born; e.g. *"So you say: children are not complete humans until birth?"* Both discuss the point represented by the grammatically indexed pattern: `child.subject be.verb human.object`. Note that we are at this stage not concerned with the stance towards the point being discussed; we will return to this later. To facilitate readability, the extracted points are associated with an 'extract' from the source sentence; in these instances, *"an unborn child is not yet a human"* and *"children are not complete humans until birth?"* Generation of points bears a passing resemblence to Text Simplification (Siddharthan, 2014), but is focussed on generating a single short sentence starting from a verb, rather than splitting sentences into shorter ones.

Points and extracts are derived from a dependency parse graph structure (De Marneffe et al., 2014). Consider:
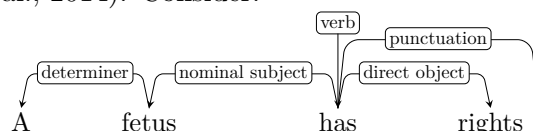
Here, the nominal subject and direct object relations form the pattern, and relations are followed recursively to generate the extract. To solve the general case, we must select dependency relations to include in the pattern and then decide which should be followed from the verb to include in the extract.

3.1.1 Using Verb Frames

We seek to include only those verb dependencies that are required by syntax or are optional but important to the core idea. While this often means using only subject and object relations, this is not always the case. Some dependencies, like adverbial modifiers or parataxis, which do not introduce information relevant to

```
<FRAME>
  <DESCRIPTION primary="NP V NP"
        secondary="Basic Transitive"/>
  <EXAMPLES>
    <EXAMPLE>
      Brutus murdered Julius Cesar.
    </EXAMPLE>
  </EXAMPLES>
  <SYNTAX>
    <NP value="Agent"><SYNRESTRS/></NP>
    <VERB/>
    <NP value="Patient"><SYNRESTRS/></NP>
  </SYNTAX>
  <SEMANTICS>...</SEMANTICS>
</FRAME>
```

Figure 1: VerbNet frame for 'murder'

the point's core idea are universally excluded. For the rest, we identify valid verb frames using FrameNet, available as part of VerbNet, an XML verb lexicon (Schuler, 2005; Fillmore et al., 2002). Represented in VerbNet's 274 'classes' are 4402 member verbs. For each of these verb classes, a wide range of attributes are listed. FrameNet frames are one such attribute, these describe the verb's syntactic arguments and the semantic role of each in that frame. An example frame for the verb 'murder' is shown in Figure 1. Here we see that the verb takes two Noun Phrase arguments, an Agent ('murderer') and Patient ('victim').

We use a frame's syntactic information to determine the dependencies to include in the pattern for a given verb. This use of frames for generation has parallels to methods used in abstractive summarisation for generating noun phrase references to named entities (Siddharthan and McKeown, 2005; Siddharthan et al., 2011). To create a 'verb index', we parsed the VerbNet catalogue into a key-value structure where each verb was the key and the list of allowed frames the value. Points were extracted by querying the dependency parse relative to information from the verb's frames.

With an index of verbs and their frames, all the information required to identify points in parses is available. However, as frames are not inherently queriable with respect to a dependency graph structure, queries for each type of frame were written. While frames in different categories encode additional semantic information, many frames share the same basic syntax. Common frames such as `NounPhrase Verb NounPhrase` cover a high

percentage of all frames in the index. We have manually translated such frames to equivalent dependency relations to implement a means of querying dependency parses for 17 of the more common patterns, which cover 96% of all frames in the index. To do this, we used the dependency parses for the example sentences listed in frames to identify the correct mappings. The remaining 4% of frames, as well as frames not covered by FrameNet, were matched using a 'Generic Frame' and a new query that could be run against any dependency graph to extract subjects, objects and open clausal complements.

3.1.2 Human Readable Extracts

Our approach to generating human readable extracts for a point can be summarised as follows: recursively follow dependencies from the verb to allowed dependents until the sub-tree is fully explored. Nodes in the graph that are related to the verb, or (recursively) any of its dependents, are returned as part of the extract for the point. However, to keep points succinct the following dependency relations are excluded: *adverbial clause modifiers, clausal subjects, clausal complement, generic dependencies and parataxis.* Generic dependencies occur when the parser is unable to determine the dependency type. These either arise from errors or long-distance dependencies and are rarely useful in extracts. The other blacklisted dependencies are clausal in nature and tend to connect points, rather than define them. The returned tokens in this recursive search, presented in the original order, provide us with a sub-sentence extract for the point pattern.

3.2 Point Curation

To better cluster extracted points into distinct ideas, we curated points. We merged subject pronouns such as 'I' or 'she' under a single 'person' subject as these were found to be used interchangeably and not reference particular people in the text; for example, points such as `she.nsubj have.verb abortion.dobj` and `I.nsubj have.verb abortion.dobj` were merged under a new pattern: `PERSON.nsubj have.verb abortion.dobj`. Homogenising points in this way means we can continue to rely on a cluster's size as a measure of importance in the summarisation task.

A number of points are also removed using a series of 'blacklists'. Based on points extracted from the Abortion debate (1151 posts, ~155000 words), which we used for development, these defined generic point patterns were judged to be either of little interest or problematic in other ways. For example, patterns such as `it.nsubj have.verb rights.dobj` contain referential pronouns that are hard to resolve. We excluded points with the following subjects: *it_PRP, that_DT, this_DT, which_WDT, what_WDT*. We also excluded a set of verbs with a `PERSON` subject; certain phrases such as "*I think*" or "*I object*" are very common, but relate to attribution or higher argumentation steps rather than point content. Other common cue phrases such as *"make the claim"* were also removed.

3.3 Summary Generation

Our goal is the abstractive summarisation of argumentative texts. Extracted points have 'patterns' that enable new comparisons not possible with sentence selection approaches to summarisation, for instance, the analysis of counter points. This section describes the process of generating an abstractive summary.

3.3.1 Extract Generation

Extract Filtering: In a cluster of points with the same pattern there are a range of extracts that could be selected. There is much variation in extract quality caused by poor parses, punctuation or extract generation. We implemented a set of rules that prevent a poor quality extract from being presented.

Predominantly, points were prevented from being presented based on the presence of certain substring patterns tested with regular expressions. Exclusion patterns included two or more consecutive words in block capitals, repeated words or a mid-word case change. Following on from these basic tests, there were more complex exclusion patterns based on the dependencies obtained from re-parsing the extract. Poor quality extracts often contained (on re-parsing) clausal or generic dependencies, or multiple instances of conjunction. Such extracts were excluded.

Extract Selection: Point extracts were organised in clusters sharing a common pattern of verb arguments. Such clusters contained all the point extracts for the same point pattern and thus all the available linguistic realisations to express the cluster's core idea. Even after filtering out some extracts, as described above, there was still much variation in the quality of the extracts. Take this example cluster of generated extracts about the Genesis creation narrative:

> "The world was created in six days."
> "The world was created in exactly 6 days."
> "Is there that the world could have been created in six days."
> "The world was created by God in seven days."
> "The world was created in 6 days."
> "But, was the world created in six days."
> "How the world was created in six days."

All of these passed the 'Extract Filtering' stage. Now an extract must be selected to represent the cluster. In this instance our approach selected the fifth point, "*The world was created in 6 days.*" Selecting the best extract was performed every time a cluster was selected for use in a summary. Selections were made using a length-weighted, bigram model of the extract words in the cluster, in order to select a succinct extract that was representative of the entire cluster.

Extract Presentation: Our points extraction approach works by selecting the relevant components in a string for a given point, using the dependency graph. While this has a key advantage in creating shorter content units, it also means that extracts are often poorly formatted for presentation when viewed in isolation (not capitalised, leading commas etc.). To overcome such issues we post-edit the selected extracts to ensure the following properties: first character is capitalised; ends in a period; commas are followed but not preceded by a space; contractions are applied where possible; and consecutive punctuation marks condensed or removed. Certain determiners, adverbs and conjunctions (because, that, therefore) are also removed from the start of extracts. With these adjustments, extracts can typically be presented as short sentences.

3.3.2 Content Selection:

A cluster's inclusion in a particular summary section is a function of the number of points

in the cluster. This is based on the idea that larger clusters are of greater importance (as the point is made more often). Frequency is a commonly used to order content in summarisation research for this reason; however in argumentative texts, it could result in the suppression of minority viewpoints. Identifying such views might be an interesting challenge, but is out of scope for this paper.

Our summaries are organised as sections to highlight various aspects of the debate (see below). To avoid larger clusters being repeatedly selected for each summary section, a list of used patterns and extracts is maintained. When an point is used in a summary section it is 'spent' and added to a list of used patterns and extracts. The point pattern, string, lemmas and subject-verb-object triple are added to this list. Any point that matches any element in this list of used identifiers cannot be used later in the summary.

3.3.3 Summary Sections:

A summary could be generated just by listing the most frequent points in the discussion. However, we were interested in generating more structured summaries that group points in different ways, i.e. counterpoints & co-occurring points.

Counter Points: This analysis was intended to highlight areas of disagreement in the discussion. Counterpoints were matched on one of two possible criteria, the presence of either negation or an antonym. Potential, antonym-derived counterpoints, for a given point, were generated using its pattern and a list of antonyms. Antonyms were sourced from WordNet (Miller, 1995). Taking `woman.nsubj have.verb right.dobj` as an example pattern, the following potential counterpoint patterns are generated:

- `man.nsubj have.verb right.dobj`
- `woman.nsubj lack.verb right.dobj`
- `woman.nsubj have.verb left.dobj`

Where there were many pattern words with antonym matches, multiple potential counter point patters were generated. Such hypothesised antonym patterns were rejected if the pattern did not occur in the debate. From the example above, only the first generated pattern: `man.nsubj have.verb right.dobj` appeared in the debate.

Negation terminology was not commonly part of the point pattern, for example, the `woman.nsubj have.verb right.dobj` cluster could include both "*A woman has the right*" and "*A woman **does not** have the right*" as extracts. To identify negated forms within clusters, we instead pattern matched for negated words in the point extracts. First the cluster was split into two groups, extracts with negation terminology and those without. The Cartesian product of these two groups gave all pairs of negated and non-negated extracts. For each pair a string difference was computed, which was used to identify a pair for use in summary. Point-counterpoint pairs were selected for the summary based on the average cluster size for the point and counterpoint patterns. In the summary section with the heading "people disagree on these points", only the extract for the point is displayed, not the counterpoint.

Co-occurring Points: As well as counterpoints we were also interested in presenting associated points, i.e. those frequently raised in conjunction with one another. To identify co-occurring points, each post in the discussion was first represented as a list of points it made. Taking all pairwise combinations of the points made in a post, for all posts, we generated a list of all co-occurring point pairs. The most common pairs of patterns were selected for use in the summary. Co-occurring pairs were rejected if they were too similar – patterns must differ by more than one component. For example, `woman.nsubj have.verb choice.dobj` could not be displayed as co-occurring with `woman.nsubj have.verb rights.dobj` but could be with `fetus.nsubj have.verb rights.dobj`.

Additional Summary Sections: First, points from the largest (previously) unused clusters were selected. Then we organised points by topic terms, defined here as commonly used nouns. The subjects and objects in all points were tallied to give a ranking of topic terms. Using these common topics, points containing them were selected and displayed in a dedicated section for that topic.

Layout Summary

People disagree on these points:
    People are automatically responsible.
    Guns were illegal.
    They are no more dangerous than any other firearm out there.
    Criminals have guns.
    Clearly having guns.
    Carry a loaded gun.
Commonly occurring points made in the discussion were:
    Keep arms in our homes.
    Kill many people.
    Own a gun there.
Users that talk about X often talk about Y.
    (X) I should not have a gun. (Y) Guns make you safer.
    (X) Concealed carry permit holder. (Y) Carry permit for 20 years.
    (X) Bear arms, in actuality. (Y) Bear arms for the purpose of self-defense.
Common points made in the discussion linking terms were:
    I grew up in a house hold around firearms.
    You need a machine gun to kill a deer.
    You take guns off the streets.
Points for commonly discussed topics:
    Gun
      On how to use guns.
      Get guns off the street.
      Them carry a gun.
    People
      People kill people.
      Guns don't kill people.
      The militia are the people.
    Government
      Protect ourselves from a tyrannical government.
      Fighting a tyrannical government.
      Limit democratically-elected governments.
Points about multiple topics:
    More people are accidentally killed by their own guns.
    Get the guns and kill the other people.
    Cars kill many more people per year than guns.
People ask questions like:
    Do you really need a machine gun to kill a little deer?
    Do guns make you safer?

Plain Summary

    People are automatically responsible.
    Guns were illegal.
    They are no more dangerous than any other firearm out there.
    Criminals have guns.
    Clearly having guns.
    Carry a loaded gun.
    I should not have a gun.
    Guns make you safer.
    Concealed carry permit holder.
    Carry permit for 20 years.
    Bear arms, in actuality.
    Bear arms for the purpose of self-defense.
    Keep arms in our homes.
    Kill many people.
    Own a gun there.
    I grew up in a house hold around firearms.
    You need a machine gun to kill a deer.
    You take guns off the streets.
    More people are accidentally killed by their own guns.
    Get the guns and kill the other people.
    Cars kill many more people per year than guns.
    On how to use guns.
    Get guns off the street.
    Them carry a gun.
    People kill people.
    Guns don't kill people.
    The militia are the people.
    Protect ourselves from a tyrannical government.
    Fighting a tyrannical government.
    Limit democratically-elected governments.
    Do you really need a machine gun to kill a little deer?
    Do guns make you safer?

Figure 2: Examples of Layout vs Plain Summaries

Most large clusters have a pattern with three components. Points with longer patterns are less common but often offer more developed extracts (e.g. "*The human life cycle begins at conception.*") Longer points were selected based on the number of components in the pattern. An alternative to selecting points with a longer pattern is to instead select points that mention more than one important topic word. Extracts were sorted on the number of topic words they include. Extracts were selected from the top 100 to complete this section using the extract selection process.

As a final idea for a summary section we included a list of questions that had been asked a number of times. Questions were much less commonly repeated and this section was therefore more an illustration than a summary.

4 Evaluation

Studies were carried out using five political debates from the Internet Argument Corpus (Walker et al., 2012): creation, gay rights, the existence of god, gun ownership and healthcare (the 6th, abortion rights, was used as a development set). This corpus was extracted from the online debate site 4forums (www.4forums.com/political) and is a large collection of unscripted argumentative dialogues based on 390,000 posts.

25 Study participants were recruited using Amazon Mechanical Turk who had the 'Masters' qualification. Each comparison required a participant to read a stock summary and exactly one of the other two (plain and layout) in a random order. Figure 2 provides examples of these, which are also defined below.
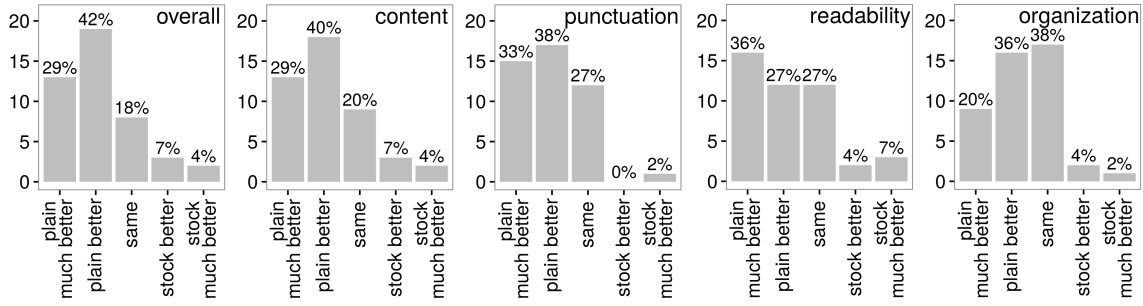
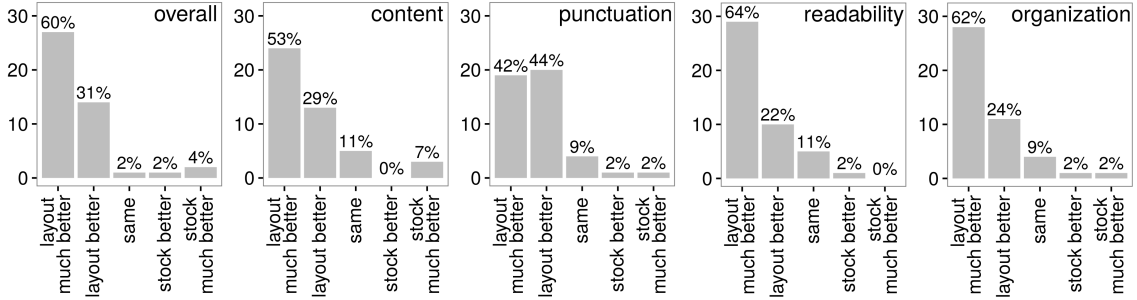Figure 3: Counts of participant responses when comparing of Plain & Stock.

Figure 4: Counts of participant responses when comparing of Layout & Stock.

- **Stock**: A summary generated using an implementation of the state of the art sentence extraction approach described by Nenkova et al. (2006)
- **Plain**: A collection of point extracts with the same unstructured style and length as the Stock summary.
- **Layout**: A summary adds explanatory text that introduces different sections of points.

The Stock summaries were controlled to be the same length as other summary in the comparison for fair comparison. Participants were asked to compare the two summaries on the following factors:

- **Content Interest / Informativeness**: The summary presents varied and interesting content
- **Readability:** The summary contents make sense; work without context; aren't repetitive; and are easy to read
- **Punctuation & Presentation**: The summary contents are correctly formatted as sentences, punctuation, capital letters and have sentence case
- **Organisation:** Related points occur near one another

Finally they were asked to give an overall rating and justify their response using free text. 9 independent ratings were obtained for each pair of summaries for each of the 5 debates using a balanced design.

4.1 Results

The study made comparisons between two pairs of summary types: Plain vs. Stock and Layout vs. Stock.

All of the five comparison factors presented in Figure 3 show a preference for our Plain summaries. These counts are aggregated from all Plain vs. Stock comparisons for all five political debates. Each histogram represents 45 responses for a question comparing the two summaries on that factor. The results were tested using Sign tests for each comparison factor, with 'better' and 'much better' aggregated and 'same' results excluded. The family significance level was set at $\alpha = 0.05$; with $m = 5$ hypotheses - using the Bonferroni correction (α/m); giving an individual significance threshold of $0.05/5 = 0.01$. 'overall', 'content', 'readability', 'punctuation' and 'organisation' were all found to show a significant difference $(p < 0.0001$ for each); i.e. even unstructured summaries with content at the point level was overwhelmingly preferred to state of the art sentence selection.

Similarly, when Layout was also compared against Stock on the same factors, we observed an even stronger preference for Layout summaries (see Figure 4). To test the increased

	Plain (A) vs. Stock (B)	Layout (A) vs. Stock (B)	*Row Total*
A much better	13	27	40
A better	19	14	33
A same	8	1	9
B better	3	1	4
B much better	2	2	4
Column Total	45	45	90

Table 1: Plain & Layout vs Stock responses contingency table

preference for Layout vs. Stock, compared to Plain, we used a One-sided, Fisher's Exact Test. Taking the 45 responses for 'overall' ratings from both comparisons, the test was performed on the contingency table (Table 1). The p-value was found to be significant ($p = 0.008$); i.e the structuring of points into sections with descriptions is preferred to the flat representation.

4.2 Discussion

The quantitative results above show a preference for both Plain and Layout point-based summaries compared to Stock. We had also solicited free textual feedback; these comments are summarised here.

Multiple comments made reference to Plain summaries having fewer questions, less surplus information and more content. Comments also described the content as being "proper English" and using "complete sentences". Comments also suggested some participants believed the summaries had been written by a human. References were also made to higher level properties of both summaries such as "logical flow", "relies on fallacy", "explains the reasoning" as well as factual correctness. In summary, participants acknowledged succinctness, variety and informativeness of the Plain summaries. This shows points can form good summaries, even without structuring into sections or explaining the links.

For comments left about preferences for Layout summaries, references to organisation doubled with respect to preferences for Plain. Readability and the idea of assimilating information were also common factors cited in justifications. Interestingly, only one comment made a direct reference to 'categories' (sections) of the summary. We had expected more references to summary sections. Fewer comments in this comparison referenced human

authors; sections perhaps hint at a more mechanised approach.

5 Conclusions and Future Work

We have implemented a method for extracting meaningful content units called points, then grouping points into discussion summaries. We evaluated our approach in a comparison against summaries generated by a statistical sentence extraction tool. The comparison results were very positive with both our summary types performing significantly better. This indicates that our approach is a viable foundation for discussion summarisation. Moreover, the summaries structure the points; for instance by whether points are countered, or whether they link different topic terms. We see this project as a step forward in the process of better understanding online discussion.

For future work, we think the approach's general methods can be applied to tasks beyond summarisation in political debate, product reviews, and other areas. It would be attractive to have a web application that would take some discussion corpus as input and generate a summary, with an interface that could support exploration and filtering of summaries based on the user's interest, for example, using a discussion-graph built from point noun component nodes connected by verb edges.

Currently the approach models discussions as a flat list of posts — without reply/response annotations. Using hierarchical discussion threads opens up interesting opportunities for Argument Mining using points extraction as a basis. A new summary section that listed points commonly made in response to other points in other posts would be a valuable addition. There is also potential for further work on summary presentation. Comments by participants in the evaluation also suggested that it would be useful to present the frequencies for points to highlight their importance, and to be able to click on points in an interactive manner to see them in the context of the posts.

References

[Boltuzic and Šnajder2015] Filip Boltuzic and Jan Šnajder. 2015. Identifying prominent arguments in online debates using semantic textual similarity. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 110–115.

[Cabrio and Villata2012] Elena Cabrio and Serena Villata. 2012. Combining textual entailment and argumentation theory for supporting online debates interactions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 208–212. Association for Computational Linguistics.

[De Marneffe et al.2014] Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D Manning. 2014. Universal stanford dependencies: A cross-linguistic typology. In *LREC*, volume 14, pages 4585–4592.

[Fillmore et al.2002] Charles J Fillmore, Collin F Baker, and Hiroaki Sato. 2002. The framenet database and software tools. In *LREC*.

[Galley et al.2004] Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 669. Association for Computational Linguistics.

[Ghosh et al.2014] 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.

[Hu and Liu2004] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

[Levy et al.2014] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 1489–1500.

[Lloret et al.2009] Elena Lloret, Alexandra Balahur, Manuel Palomar, and Andrés Montoyo. 2009. Towards building a competitive opinion summarization system: challenges and keys. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium*, pages 72–77. Association for Computational Linguistics.

[Miller1995] George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, November.

[Nenkova and McKeown2011] Ani Nenkova and Kathleen McKeown. 2011. Automatic summarization. *Foundations and Trends in Information Retrieval*, 5(2-3):103–233.

[Nenkova et al.2006] Ani Nenkova, Lucy Vanderwende, and Kathleen McKeown. 2006. A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 573–580. ACM.

[Newman and Blitzer2003] Paula S Newman and John C Blitzer. 2003. Summarizing archived discussions: a beginning. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 273–276. ACM.

[Schuler2005] Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

[Siddharthan and McKeown2005] Advaith Siddharthan and Kathleen McKeown. 2005. Improving multilingual summarization: using redundancy in the input to correct mt errors. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 33–40. Association for Computational Linguistics.

[Siddharthan et al.2011] Advaith Siddharthan, Ani Nenkova, and Kathleen McKeown. 2011. Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4):811–842.

[Siddharthan2014] Advaith Siddharthan. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.

[Teufel and Moens2002] Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4):409–445.

[Walker et al.2012] Marilyn A. Walker, Jean E. Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012*, pages 812–817.

[Zhou and Hovy2006] Liang Zhou and Eduard H
Hovy. 2006. On the summarization of dy-
namically introduced information: Online dis-
cussions and blogs. In *AAAI Spring Sympo-
sium: Computational Approaches to Analyzing
Weblogs*, page 237.