Summarizing Multi-Party Argumentative Conversations in Reader Comment on News

Emma Barker and **Robert Gaizauskas**
Department of Computer Science
University of Sheffield
{e.barker, r.gaizauskas}@sheffield.ac.uk

Abstract

Existing approaches to summarizing multi-party argumentative conversations in reader comment are extractive and fail to capture the argumentative nature of these conversations. Work on argument mining proposes schemes for identifying argument elements and relations in text but has not yet addressed how summaries might be generated from a global analysis of a conversation based on these schemes. In this paper we: (1) propose an issue-centred scheme for analysing and graphically representing argument in reader comment discussion in on-line news, and (2) show how summaries capturing the argumentative nature of reader comment can be generated from our graphical representation.

1 Introduction

A very common feature of on-line news is a reader comment facility that lets readers comment on news articles and on previous readers' comments. What emerges are *multi-party conversations*, typically argumentative, in which, for example, readers question, reject, extend, offer evidence for, explore the consequences of points made or reported in the original article or in earlier commenters' comments. See, e.g. *The Guardian* on-line.

One problem with such conversations is that they can rapidly grow to hundreds or thousands of comments. Few readers have the patience to wade through this much content, a task made all the more difficult by lack of explicit topical structure. A potential solution is to develop methods to summarize comment automatically, allowing readers to gain an overview of the conversation.

Various researchers have already proposed methods to automatically generate summaries of reader comment (Khabiri et al., 2011; Ma et al., 2012; Llewellyn et al., 2014). These authors adopt broadly similar approaches: first reader comments are topically clustered, then comments within clusters are ranked and finally one or more top-ranked comments are selected from each cluster, yielding an extractive summary. A major drawback of such summaries is that they fail to capture the essential argument-oriented nature of these multi-way conversations, since single comments taken from clusters do not reflect the argumentative structure of the conversation. I.e. such summaries do not identify the *issues* about which commenters are arguing, the alternative *viewpoints* commenters take on the issues or key *evidence* supporting one viewpoint or another, which a truly informative summary must do.

By contrast, researchers working on argument mining from social media, including reader comment and on-line debate, have articulated various schemes defining argument elements and relations in argumentative discourse (e.g. Ghosh et al. (2014), Habernal et al. (2014), Swanson et al. (2015)). If such elements and relations could be automatically extracted then they could potentially serve as a basis for generating a summary that better reflects the argumentative content of reader comment. Indeed, several of these authors have cited summarization as a motivating application for their work. However, to the best of our knowledge none have proposed how, given a formal analysis of an conversation in response to a news article, one might produce a summary of that conversation. This is a non-trivial issue.

In this paper we make two main contributions. First (Section 2) we present a light-weight analytical framework consisting of various argument elements and relations, specifically developed to

Proceedings of the 3rd Workshop on Argument Mining, pages 12–20,
Berlin, Germany, August 7-12, 2016.

capture argument in reader comments and news and we show, via an example, how an analysis using this framework may be graphically represented. Secondly (Section 3), we make proposals for how summaries that capture the argument-oriented character of reader comment conversations could be derived from the graphical representation of the argument structure of a set of comments and the article, as presented in Section 2.

2 A Framework for Characterising Argument in Comment on News

2.1 Issues, Viewpoints and Assertions

From an idealised perspective, commenters address *issues*, hold *viewpoints* on issues and make *assertions*, which serve many purposes including directly expressing a viewpoint and providing evidence for an assertion (or viewpoint). Of course reader comments may also have other functions, e.g. expressing emotions or making jokes, but here we are primarily interested in their argumentative content. We expand on these terms as follows[1]:

Assertions A comment typically comprises one or more assertions – propositions that the commenter puts forward and believes to be true. Each assertion has a particular role in the local discourse. We find relations between assertions made within a comment, between assertions made in different comments and between assertions in comments and assertions in the article. Key relations between assertions include: *rationale* (one provides evidence to support another); *agree/disagree* (one agrees or disagrees with another).

Viewpoints Disagreement or contention between comments is a pervasive feature of reader comment and news. When an assertion made by one comment is contradicted by or contends i) an assertion expressed in another comment, or ii) an assertion reported in or entailed by something reported in the news article, each opposed assertion expresses a *viewpoint*. It follows that whether or not an assertion expresses a viewpoint is an emergent property of the discourse and only relative to the local discourse; it is not an inherent feature of the assertion itself.

[1]We would like to thank one our reviewers for pointing out close similarities between the framework we describe here and the IBIS framework of Kunz and Rittel (1970). In particular they share the ideas that issues are questions, are key primitive elements in a theory of argumentative discourse and emerge dynamically and recursively in argument.

Issues Implicitly related to notion of viewpoint is that of *issue*. We can think of an issue as a question or problem to which there are two or more contending answers. The space of possible answers is the set of related but opposed viewpoints expressed in the comment set. I.e. an issue is that which a viewpoint is a viewpoint *on*.

Issues may be expressed in various ways, e.g. (1) via a "whether or not"-type expression, e.g. *whether or not to lower the drinking age*; (2) via a yes-no question, e.g. *Should Britain leave the EU?*; (3) via a "which X?"-type expression when there are more than two alternatives, e.g. *Which was the best film of 2015?*. However, issues are rarely explicitly articulated in reader comments or in the initial news article. Rather, as the dialogue evolves, a set of assertions made by commenters may indicate a space of alternative, opposed viewpoints, and an issue can then be recognised.

Sub-issues frequently emerge within the discussion of an issue, i.e. issues have a recursive nature. When evidence proposed as support for a viewpoint on an issue is contended, the two contending comments, which may in turn attract further comments, become viewpoints on a new issue, subordinate to the first. Sub-sub-issues may arise below sub-issues and so on.

2.2 A Graphical Representation

In the previous sub-section we defined the key concepts in our approach to analysing argument in comment on news. To demonstrate how they can be used to analyse a particular news article plus comment set we propose a graphical representation of the argument structure, with indices that anchor the representation in textual elements. A graphical approach is well-suited to the task of identifying structural relations between elements in a scheme, particularly when some of the elements are abstractions not themselves directly represented in the text (as is widely recognised in the argumentation community (Reed et al., 2007; Conklin and Begeman, 1988)).

We introduce our graphical representation via an example. Figure 1 shows an extract from a *Guardian* news article about the controversy surrounding a town council's decision to reduce the frequency of bin collection, and 11 (of 248) comments posted in response to the article. Figure 2 shows a partial depiction of the issues, viewpoints and rationales and argumentative structure in this

[S0] **Rubbish?** [S1] **Bury council votes to collect wheelie bins just once every three weeks**
[S2] Locals fear the new move will lead to an increase in fly-tipping and attract foxes and vermin, but the council insists it will make the borough more environmentally friendly. [S3] Is it just a desperate cost cutting measure? ...
[S4] A council in Greater Manchester is to be the first in England to start collecting wheelie bins only once every three weeks, scrapping the current fortnightly collection. [S5] The controversial decision was unanimously passed by councillors in Bury on Wednesday night, despite fears fly tipping would increase. [S6] One councillor who voted for the motion accused her opponents of "scaremongering" after they warned rubbish would pile up and attract vermin. [S7] Another argued the money saved could be spent on more social workers.
[S8] It affects the grey bins used for general household waste which can't be recycled ... [S9] The Labour-run council claims the move is part of a strategy to turn Bury into a "zero waste borough", boost recycling and save money on landfill fees ... [S10] Many residents feel it is simply a desperate cost saving measure, after the town hall was told to make more than £32m of cuts over the next two years ...

Id	Poster	Reply	Comment
1	A		I can't see how it won't attract rats and other vermin. I know some difficult decisions have to be made with cuts to funding, but this seems like a very poorly thought out idea.
2	B	2 → 1	Plenty of people use compost bins and have no trouble with rats or foxes.
3	C	3 → 2	If they are well-designed and well-managed- which is very easily accomplished. If 75% of this borough composted their waste at home then they could have their bins collected every six-weeks. It's amazing what doesn't need to be put into landfill.
4	D	4 →1	It won't attract vermin if the rubbish is all in the bins. Is Bury going to provide larger bins for families or provide bins for kitchen and garden waste to cut down the amount that goes to landfill? Many people won't fill the bins in 3 weeks - even when there was 5 of us here, we would have just about managed.
5	E	5 → 1	Expect Bury to be knee deep in rubbish by Christmas it's a lame brained Labour idea and before long it'll be once a month collections. I'm not sure what the rubbish collectors will be doing if there are any. We are moving back to the Middle Ages, expect plague and pestilence.
6	F		Are they completely crazy? What do they want a new Plague?
7	G	7 →6	Interesting how you suggest that someone else is completely crazy, and then talk about a new plague.
8	H	8 →7	Do you think this is a good idea? We struggle with fortnightly collection. This is tantamount to a dereliction of duty. What are taxpayers paying for? I doubt anyone knew of this before casting their vote.
9	I	9→8	I think it is an excellent idea. We have fortnightly collection, and the bin is usually half full or less[family of 5].. Since 38 of the 51 council seats are held by Labour, it seems that people did vote for this. Does any party offer weekly collections?
10	G	10 →8	I don't think it's a good idea. But..it won't cause a plague epidemic.
11	G	11 →9	I live by myself, so my bin is going to be smaller ..but I probably have more bin-space-per-person. And I recycle everything I can possibly recycle, and make sure nothing slips through the net. Yet I almost fill my bin with food waste and the odd bit of unrecyclable packaging in a fortnight.. How are you keeping your bin so empty?

Figure 1: Part of a news article (top) and comments responding to it (bottom). Comments are taken from two threads in sequence but some intermediate comments have been omitted. Full article and comments at: `http://gu.com/p/4v2pb/sbl`.

example as a directed graph.

Nodes in the graph represent issues, viewpoints or assertions. Issues are distinguished by italics, e.g. *Is reducing bin collection to once every 3 weeks a good idea?* Nodes inside dashed boxes are implicit parts of the argument, i.e. are not directly expressed in the comments or article but are implied by them and allow other nodes which are explicit to be integrated into the overall argument structure. Nodes are labelled with abstract glosses of content explicitly or implicitly mentioned in the article, with repeated content represented only once. For nodes that are expressed or signalled in the news article or comments, a list of article sentence [sn] and comment [cn] indices is given, grounding the argument graph in the text.

Relations between nodes are indicated by directed edges in the graph. Orange edges indicate that the assertion at the tail of arrow is a viewpoint on the issue at the head, e.g. [less frequent collection] "will attract vermin" or "will not attract vermin". Blue edges indicate the assertion at the tail provides a rationale for the assertion at the head.

To create such a graph manually is a laborious process of iterative refinement: (1) Look for textual units expressing contention in the article. When spotted, formulate initial glosses of opposed viewpoints. (2) Read the comments in turn, by thread, looking for textual units expressing contention between comments or comments and the article. When spotted, formulate/refine glosses for opposed viewpoints (*will attract vermin*/*won't attract vermin*) and propose a potential issue. (3) Group similar and related comment together – requires re-reading earlier content to assess similarity and may result in refining earlier glosses of viewpoints and issues. (4) As rationale relations are recognised, add these in beneath viewpoint or

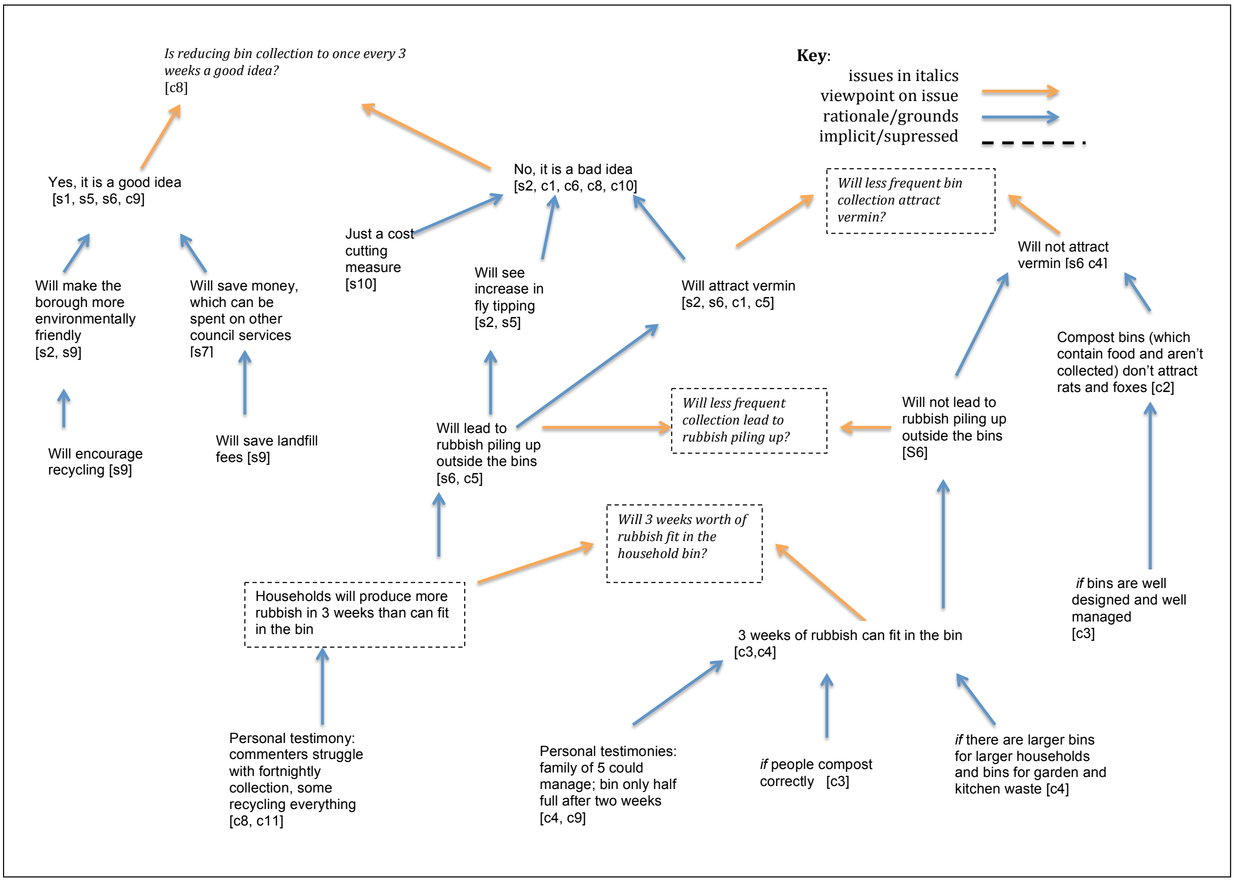

Figure 2: Argument graph for the article and comment subset shown in Figure 1

rationale nodes already in the graph. Implicit assertions or viewpoints may need to be added to capture the structure of the argument (e.g. *households will produce more rubbish in 3 weeks than can fit in the bin*). This should only be done when necessary to integrate parts of the argument that are explicit (i.e. not all implicit assertions need to be made explicit).

3 Generating Summaries from Argument Graphs

Argument graphs show the issues raised, viewpoints taken and rationales given across a (possibly very large) set of comments. Such a graph reveals the argumentative structure of the comment set in relation to the article, something that should be reflected in any argument-oriented summary. But clearly it contains far too much information to appear in a summary. How, then, can we use the graph structure and the information it contains to produce a summary?

The argument graph itself could be used as a *visual* presentation mechanism for the argumentative content of the comments and article. The user could, e.g., be allowed to expand or collapse nodes on demand, starting from the top-most issues(s) and viewpoints. While this is a promising line of work, here we focus on how to generate a *textual* summary from an argument graph. In the next subsection we discuss features of the graph one could take into account to generate a summary. Then we present two heuristic algorithms for generating summaries that exploit these features and illustrate the output of one of them by using it (manually) to produce a summary from an extended version of the argument graph presented in Figure 2. Finally, we compare that summary with a human summary of the same article and comments produced in a less prescriptive way.

3.1 Features for Summarization

Given an argument graph as presented in Section 2.2, a summarization algorithm could exploit both quantitative features of nodes in the graph as well as structural relations between nodes.

Features of *issue* nodes to use include:

Issue Index Count Number of comments or article sentences explicitly mentioning the issue.

Issue Id	Textual Gloss of Issue	Issue Index Count	Max Issue Depth	Sub Node Count	Sub Issue Count	Total Index Count
1	Is reducing bin collection to once every 3 weeks a good idea?	1	7	31	7	58
2	Will less frequent bin collection attract vermin?	0	6	21	5	38
3	Will less frequent collection lead to rubbish piling up?	0	4	15	3	24
4	Will 3 weeks of rubbish fit in the bin?	0	4	13	2	21
5	Will reducing collection encourage recycling?	0	3	10	2	13
6	Can people recycle/compost more rubbish than they do?	0	2	7	1	11
7	Are vermin attracted by the type of rubbish in the black/grey bin?	0	1	2	0	5
8	Do people in flats have composting facilities?	0	1	2	0	4

Table 1: Issues in Bin Collection Article and First 30 Comments

Id	Issue	Textual Gloss of Viewpoint on Issue	VPt Index Count	Total Index Count	Evid. Count	Total Evid. Count	Max VPt Depth	Evid. Nodes Contd
1	1	No it is a bad idea	6	21	3	7	4	4
2	1	Yes it is a good idea	4	14	3	7	4	2
3	2	Will attract vermin	4	12	2	4	3	3
4	2	Will not attract vermin	3	19	3	10	5	4
5	3	Will lead to rubbish piling up	2	5	1	2	2	1
6	3	Will not lead to rubbish piling up	1	12	1	6	4	2
7	4	3 weeks of rubbish can fit in the bin	2	11	3	5	3	1
8	4	3 weeks of rubbish cannot fit in the bin	0	3	1	1	1	0
9	5	Reducing collection will encourage recycling	1	6	2	3	2	1
10	5	Reducing collection will not encourage recycling	0	5	1	3	2	2
11	6	People don't recycle/compost everything they can	3	4	1	2	1	0
12	6	People recycle/compost everything they can already	1	5	2	2	1	1
13	7	Black/grey bins contain the type of rubbish that attracts vermin	3	3	0	0	0	0
14	7	Black/grey bins contain packaging that does not attract vermin	2	2	0	0	0	0
15	8	People in flats don't have composting facilities	3	3	0	0	0	0
16	8	People in flats have composting facilities	1	1	0	0	0	0

Table 2: Viewpoints (VPts) in Bin Collection Article and First 30 Comments

Maximum Issue Depth Count of all levels below an issue.

Sub-Node Count Count of all viewpoint and rationale nodes below the issue node. including points of contention.

Sub-Issue Count Count of all issues below a given issue.

Total Index Count Count of all indices on the issue itself and all sub-nodes.

Maximum Issue Depth, together with the Sub-node Count give a broad indication of the structural character of the argument. A shallow depth score of say 2, with a high number of nodes suggest that there are lots of different reasons given for supporting or not supporting something, but no complex case given to explain or justify the support for these positions. Sub-issue Count is indicative of the degree of contention on an issue and Total Index Count indicates the volume of explicit comment relating to an issue, i.e. is an indication of the number of participants in the conversation saying things related to the issue. Table 1, which is meant to be indicative only, shows these counts for a version of the argument graph of Figure 2 extended to include 30 comments (two full threads).

Features of *viewpoint* nodes to use include:

Viewpoint Index Count Count of the indices of comments or article sentences that directly support the viewpoint.

Viewpoint Total Index Count Count of all indices on the viewpoint, both direct and indirect; i.e. indices on the viewpoint and indices on all rationale nodes supporting the high level viewpoint (excludes indices on nodes contending any of the lower level rationales).

Evidence Count Count of the number of rationale nodes directly below a viewpoint.

Total Evidence Count Count of all nodes playing a role in supporting the viewpoint.

Maximum Viewpoint Depth Count of the levels of rationale given below the viewpoint.

Evidence Nodes Contended Count of the number of direct contentions to rationales supporting a viewpoint.

Viewpoint Index Count shows the strength of direct support for a viewpoint. Viewpoint Total Index Count provides an indication of both direct support for the viewpoint and support for the supporting arguments. Together, Evidence Count, Total Evidence Count and Maximum Viewpoint Depth indicate the structural complexity and detail of the supporting case. Evidence Nodes Contended indicates the degree to which rationales for a viewpoint are contended by counter arguments. Table 2 shows these counts for the extended version of the argument graph shown in Figure 2.

Table 2 defines measures relating to the position and popularity of issue and viewpoint nodes in an argument graph. The same measures can be calculated for *evidence* nodes.

3.2 Algorithms for Summarization

The counts specified in the previous section together with the structure of the argument graph can be used in many different ways to generate summaries. Here we mention just two as an indication of the space of possibilities. [2]

Simple Issue-oriented Summarizer One simple baseline is to list issues discussed, up to the summary length limit, ordered by whichever quantitative measure for issues is felt to best indicate significance. Choosing an ordering measure like Total Index Count places value on the number of commenters discussing the issue; choosing Sub-Node Count favours more elaborated arguments and Sub-Issue Count favours issues that give rise to more contention. In the example shown in Table 1 the various measures all correlate closely so the choice of which measure to use is arbitrary; however this need not always be the case.

[2]One important technical observation should be made In the example above the argument graph is a connected graph in which there is a unique issue node (the *root issue*): (1) to which all other nodes are connected either via viewpoint or rationale relations or via issues arising from contention of a rationale node otherwise related to the root issue, and (2) none of whose viewpoint nodes are rationales for other nodes in the argument. In Figure 2, e.g., while no issue node has a parent, the issue *Is reducing bin collection to once every 3 weeks a good idea?* is unique in that none of its viewpoint nodes is a rationale for any other node in the graph. In the general case, comment sets can give rise to multiple, unrelated root issues. We do not discuss such cases here, i.e. we assume the graph to be summarised is a connected graph with a single root issue. However, the algorithms discussed below could easily be extended to accommodate the more general case, e.g. by distributing the total summary length between each root-dominated sub-graph, possibly allowing "larger" sub-graphs, as determined by one of the measures above, a greater proportion of overall summary length.

Algorithm 1 Single Issue Summarizer

Require: An argument graph G; issue I in G; comparison functions $f_S(.,.)$, $f_E(.,.)$ for ordering viewpoint and rationale nodes respectively; a measure $Threshold_E$ on evidence nodes

1: $Summary \leftarrow []$
2: $Summary += I$
3: $S \leftarrow$ list of the viewpoints on I ordered by f_S
4: **for each** viewpoint s in S **do**
5:     $Summary += s$
6:     $R_s \leftarrow$ list of rationales for s ordered by f_E
7:     **for each** rationale node r_s in R_s **do**
8:         **if** $f_E(Threshold_E, r_s)$ **then**
9:             $Summary += r_s$
10:         **end if**
11:     **end for**
12: **end for**

Simple Argument-Oriented Summarizer The simple issue-oriented approach ignores information about viewpoints and rationales and about which sub-issues relate to which specific dominating issues. A more interesting approach to summarization should take this into account. One way to do this is shown in Algorithm 1, which outlines the logic for selecting the content for inclusion in an argument-oriented summary of one issue. Given an argument graph and an issue, the algorithm starts by including the issue in the summary, then for each viewpoint on the issue adds that viewpoint in an order defined over some feature of viewpoints (e.g. Total Index Count). As each viewpoint is added, evidence nodes for the viewpoint are added, ordered by some node feature (e.g. Evidence Nodes Contended), provided their count on this feature exceeds a threshold.

Algorithm 1 only summarizes a single issue. It could be used to generate a high level summary of an argument graph by calling it with the root node. Or, it could be extended to cover more of the graph in various ways. For example, after line 9, when the decision to add an evidence node r_s to the summary has been made, r_s could be checked to see if it is a viewpoint on a issue, i.e. if it has been contended. If yes, it could be added to a list $SubI$ of sub-issues to report in the summary. After line 12, $SubI$ could be sorted by some measure of importance and, possibly, thresholded or truncated to shorten it. The algorithm

Summary 1 (97 words) Many commenters were unhappy with the less frequent collections; some were struggling already with the fortnightly collection and were concerned with vermin or overflowing bins. A few commenters, however, countered that black bins were for non-food waste that would not attract vermin. Other commenters thought fewer collections were manageable if people recycled their food waste, garden waste, and any other recycleable materials. Few commenters, however, pointed out the lack of composting facilities for those living in some areas or flats. The council should provide more education and services in these areas to encourage more people to recycle.	**Summary 2** (112 words) The central issue discussed was whether reducing bin collection to once every 3 weeks is a good idea. Some argued it was a bad idea because it would lead to vermin being attracted and to an increase in flying tipping. Others argued it was a good idea as it would save money that could be spent on other council services and would make the borough more environmentally. Whether the proposal would lead to vermin being attracted was debated. Some argued they would because the proposal would lead to rubbish piling up in the streets. Others argued it would not as the proposal would not lead to rubbish piling up in the streets.

Figure 3: A human authored summary (Summary 1) and a potential automatic summmary (Summary 2) of the first 30 comments on the Bury Bin Collection Article

could then be called recursively on each of the issues in $SubI$ or $SubI$ could be returned and added to an agenda maintained by a higher level controlling algorithm, which calls Algorithm 1 iteratively on each of the issues in its agenda. Of course the summary must not exceed its length limit.

The algorithm only selects the content for inclusion in a summary and ignores details of how it is to be realised. A more or less mechanical surface realisation process could be used to generate a summary like that shown in Figure 3, Summary 2. In this summary for the extended argument graph underlying Tables 1 and 2, we assume the root issue has been summarised (sentences 1-3) and that one further issue (2) has also been chosen for inclusion using the sort of extension to the algorithm described in the last paragraph (sentences 4-6).

3.3 Comparison with Human Summaries

Figure 3 shows a human-authored summary of the first 30 comments on the bin collection article, created as part of a corpus of gold standard reader comment summaries (SENSEI Project, 2016). Annotators created the gold standard summaries using a novel 3 stage method: (1) each comment in the source set is annotated with a *label* (i.e. a mini-summary of the main points in the comment); (2) related labels are sorted into groups that the annotator believes will be helpful for writing an overview summary and a group label is produced to indicate common content in the group; (3) based on the analysis and annotations created in stages (1) and (2), an overview summary is written, which should identify the main points raised in the discussion, different views, areas of consensus, the proportion of comments addressing a topic or sharing a view, and strong feelings shown.

The human summary sentences shown above correspond very closely to elements in the graphical representation of the same 30 comments and while the summary addresses only a subset of the graph nodes, it does not introduce any additional content. This is a promising, if weak, form of validation as it suggests that summaries read off our argument graph using the algorithm of the last section are very similar to those produced by humans, given only modest direction. Further comparison of our gold standard human summaries with graphical representations of the same source texts might provide additional insights into how to refine algorithms for summary content selection.

4 Related Work

In recent years various authors have begun work on argument mining in on-line discussion forums (e.g., Cabrio and Vilatta (2012); Boltužić and Šnajder (2014); Swanson et al. (2015)) and reader comment on news (e.g., Sobhani et al. (2015); Carstens and Toni (2015); Sardianos et al. (2015)). While sharing some features, such as allowing multiple participants to exchange views, make claims and supply supporting arguments, these two sources of argumentative discourse also exhibit notable differences. For instance, in online discussion forums such as *debatepedia.org* or *convinceme.net*, debates are topically organized or tagged with key words, e.g. *climate change*, and a debate is typically framed by a starting motion or question and an example of a supporting or counter statement (similar to our notion of issue and viewpoint). In reader comment this structured information is missing and the debate is framed solely by a document (the article), with issues, as we define them, rarely explicitly signalled in the article or comments. Thus, the task of structuring the debate by discovering the issues, which our framework addresses, is a challenge of particular importance for reader comment.

Many authors propose models of argumentation and associated annotation schemes, e.g. Ghosh et al. (2014), Swanson et al. (2015), Carstens and Toni (2015). These models/schemes specify a set of argumentative elements and relations between them and, as noted by Peldszus and Stede (2013), approaches to argument mining typically address the subtasks of identifying, classifying and relating argumentative discourse units (ADUs) according to the types of ADU and argumentative relation specified in whatever model/scheme has been adopted. Our framework too relies upon defining and operationalising the identification of similar argument elements and relations (viewpoints and rationales in our case). However, with the exception of Kunz and Rittel (1970) we are not aware of any argumentation model that puts the notion of issue in the form of a question at the centre of the model and organises argument elements and relations around it.

Aside from differences in the text type addressed (reader comment rather than on-line debate) and the prominence given to notion of issue in our anaytical framework, our principal difference to other work in argument mining is the task we focus on: summarization. Some authors have cited summarization as a motivating end-user task, e.g. Swanson et al. (2015) and Misra et al. (2015). However, both these works aim at summarising an argument on a single topic like "gun control" across multiple dialogues and do not address the summarization of single, multi-party argumentative conversations that may address multiple issues, such as those found in reader comments. To the best of our knowledge no one has addressed the form that an end-user overview summary of reader comment might take or how it might be generated from the abstract representation of an argument, as we do in this paper.

5 Discussion and Future Work

In this paper we have defined notions of *issue*, *viewpoint* and *assertion* as part of a framework for analysing argumentative conversations such as those that appear in response to news articles in on-line news. We introduced a graphical representation for representing these argument elements and relations between them, such as *viewpoint-on* holding between viewpoints and issues and *rationale-for* holding between assertions and viewpoints/other assertions. We also discussed how an argument graph can be used to generate summaries of argumentative conversations, proposing features that could be extracted from an argument graph to assist in selecting content to be summarised and sketching two basic summarization algorithms suggestive of the space of possible algorithms that could be developed.

We are fully aware that our analytical framework, graphical representation and proposals for summarization algorithms are theoretical preliminaries and, while grounded in extensive observation and analysis of data, need to be implemented and empirically evaluated to be validated. This forms the core of our current and future work. Specifically we need to further develop, implement and evaluate (1) methods for reliably extracting an argument graph from news articles and comments (2) summarization algorithms of the sort outlined above. Building argument graphs is the greater of these challenges and is perhaps best approached by factoring it into sub-tasks, such as *candidate assertion detection*, *argumentative relation detection* and *issue identification*. Candidate assertion detection involves segmenting the text into clauses that could play a role in the argument. Argumentative relation detection involves identifying various relations between candidate assertions, such as identity, disagreement or contradiction and evidence or support. Issue identification involves detecting a disagreement or contradiction relation between assertions and establishing sufficient supporting argumentation for the opposed assertions and/or repetition across multiple participants in the conversation to deem them an issue. Building components to carry out these sub-tasks is likely to require the creation of annotated resources for training and testing. Existing supervised learning techniques can then be brought to bear. As well as implementing our proposals, further work should be carried out to refine and validate our analytical framework, e.g., by getting multiple analysts to generate argument graphs for a corpus of comment sets.

While these challenges are substantial we believe the proposals made in this paper provide a realistic framework to progress work on summarization of multi-party argumentative conversations.

Acknowledgements The authors would like to thank the European Commission for supporting this work, carried out as part of the FP7 SENSEI project, grant reference: FP7-ICT-610916.

References

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

Elena Cabrio and Serena Villata. 2012. Combining textual entailment and argumentation theory for supporting online debates interactions. In *The 50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 208–212.

Lucas Carstens and Francesca Toni. 2015. Towards relation based argumentation mining. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 29–34, Denver, CO, June. Association for Computational Linguistics.

Jeff Conklin and Michael L. Begeman. 1988. gIBIS: A hypertext tool for exploratory policy discussion. *ACM Trans. Inf. Syst.*, 6(4):303–331.

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.

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.

Elham Khabiri, James Caverlee, and Chiao-Fang Hsu. 2011. Summarizing user-contributed comments. In *Proceedings of The Fifth International AAAI Conference on Weblogs and Social Media (ICWSM-11)*, pages 534–537.

Werner Kunz and Horst W. J. Rittel. 1970. Issues as elements of information systems. Working Paper No. 131, Institute of Urban and Regional Development, Univ. of California, Berkeley, Calif.

Clare Llewellyn, Claire Grover, and Jon Oberlander. 2014. Summarizing newspaper comments. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014*, pages 599–602.

Zongyang Ma, Aixin Sun, Quan Yuan, and Gao Cong. 2012. Topic-driven reader comments summarization. In *Proceedings of the 21st ACM international Conference on Information and Knowledge Management*, CIKM '12, pages 265–274.

Amita Misra, Pranav Anand, Jean E. Fox Tree, and Marilyn Walker. 2015. Using summarization to discover argument facets in online idealogical dialog. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 430–440.

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.

SENSEI Project. 2016. The SENSEI corpus of human summaries of reader comment conversations in online news. Available from: nlp.shef.ac.uk/sensei/.

Chris Reed, Douglas Walton, and Fabrizio Macagno. 2007. Argument diagramming in logic, law and artificial intelligence. *Knowledge Engineering Review*, 22(1):87–109.

Christos Sardianos, Ioannis Manousos Katakis, Georgios Petasis, and Vangelis Karkaletsis. 2015. Argument extraction from news. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 56–66.

Parinaz Sobhani, Diana Inkpen, and Stan Matwin. 2015. From argumentation mining to stance classification. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 67–77.

Reid Swanson, Brian Ecker, and Marilyn Walker. 2015. Argument mining: Extracting arguments from online dialogue. In *Proceedings of the SIGDIAL 2015 Conference*, pages 217–226.