Semi-Automated Argumentative Analysis of Online Product Reviews

Adam WYNER [b,1], Jodi SCHNEIDER [a], Katie ATKINSON [b] and Trevor BENCH-CAPON [b]

[a] *Digital Enterprise Research Institute, National University of Ireland*
[b] *Department of Computer Science, University of Liverpool*

Abstract. Argumentation is key to understanding and evaluating many texts. The arguments in the texts must be identified; using current tools, this requires substantial work from human analysts. With a rule-based tool for semi-automatic text analysis support, we facilitate argument identification. The tool highlights potential argumentative sections of a text according to terms indicative of arguments (e.g. 'suppose' or 'therefore') and domain terminology (e.g. camera names and properties). The information can be used by an analyst to instantiate argumentation schemes and build arguments for and against a proposal. The resulting argumentation framework can then be passed to argument evaluation tools.

Keywords. text analysis, social media, argumentation schemes

Introduction

On the Internet, people produce large corpora of unstructured textual information. The complexity and volume of information require that we provide tools to extract, analyse, and structure it into a form suitable for further processing. The intended use of the information guides, in part, what is extracted and how it is structured. For example, companies, political parties, and governments may want to understand information in a corpus in order to identify dispositions on products, candidates, and policy proposals. To understand the information better, text analytic tools can be applied to the corpora to extract *sentiments* (positive or negative statements), identify *terminological trends* (e.g. product mentions), and *metadata* (locations, time, networks of users, etc) [10].

Though the results of such analysis may identify some of the structure within the textual source, they do not help to understand *discourse relations between statements in the corpus* that indicate a conclusion, a premise, or a contrast with some other statement. Identifying and extracting information according to these discourse relations would enable the reconstruction of *arguments* from the text. Broadly, there remains a large gap between natural language text and a formal representation that is amenable to automated processing for reasoning or argumentation. This paper presents a novel step towards filling the gap.

[1] Corresponding Author E-mail: adam@wyner.info

We explore a corpus of comments in an Internet forum about purchasing a camera. Within this corpus, there are various dialogical activities to examine, e.g. *persuasion*, *negotiation*, *deliberation*, and others [16]. Moreover, there is a wide range of explicit and implicit discourse relations [17]. In both areas, research is ongoing, and it is not yet clear how to systematically identify either the relevant dialogue type or discourse relations. To focus on argumentation identification in the corpus, we consider explicit indicators of argument, where participants present their views on the merits and demerits of the camera under consideration. Such an analysis brings to the forefront the *dialectical, qualitative* aspect of the overall discussion in contrast to the *quantitative* aspects found in sentiment, terminology, and metadata analyses.

In quantitative analyses, a minority view may have lesser impact, while in a dialectical, qualitative analysis, the minority view may have very high impact. For example, where only one out of a hundred reviews expresses dislike for the product, then this might be dismissed, while it may be that this one review raises a critical point that significantly attacks the other reviews and becomes highly relevant to making a decision. In other words, the influence of an argument is not measured by the number of statements in a certain classification, but in relation to its attacks on other arguments and its defence against attack. Identifying the main points of a review and their justifications along with the relationships (e.g. attack) between reviews would give analysts insight into the significance of individual comments and the network of comments taken together.

The paper presents an innovative development of argumentation theory, integrating research about discourse indicators, domain ontologies, argumentation schemes, and argumentation frameworks. We introduce a new argumentation scheme relevant to our target domain, an analysis of the domain, and selected argumentation components. We develop a novel application that instantiates these elements in a rule-based, semi-automated text analytic support tool that deterministically outputs passages based on explicit lists and rules. From the output, the analyst can examine, select, and manually structure the passages into an argumentation scheme [15]. Such instantiated schemes can then be abstracted, represented in an argumentation framework [3], and automatically evaluated [4]. We believe that the tool facilitates argument analysis by identifying relevant passages. We also clarify the aspects that remain to be developed further.

In Section 1, we present the use case and materials for our tool. Section 2 lays out the three levels of textual analysis - domain terminology, discourse indicators, and argumentation scheme. In Section 3, we review the General Architecture for Text Engineering (GATE) framework and the particular tool components built in the framework. This is followed by sample output in Section 4 and some discussion of system evaluation. Sample instantiated argumentation schemes derived from the tool output are provided in Section 5; we discuss how such arguments would be used in argument evaluation. We discuss the tool, related work, and future directions in Section 6.

1. Use Case and Materials

As a use case, we take reviews about buying the (arbitrarily chosen) Canon PowerShot SX220 HS Digital Camera from the Amazon e-commerce website.[2], where a very typi-

[2] Accessed 23.3.2012. `https://www.amazon.co.uk/product-reviews/B004M8S152/re`

cal question is: *Which camera should I buy?*. This choice has several advantages. First, decision support is needed to help people make concrete, personalised decisions. Second, manufacturers and retailers have a commerical interest in understanding consumer thinking and behaviour. Third, reviews are a popular genre of web commentary, with inherent disagreement, a clear overall topic/subject matter, and large amounts of readily available online data. Fourth, our results can be generalised to other similar forums.

There are 84 reviews in our corpus. Of these, we analysed all 46 five-star (highly favourable) reviews and all 5 three-star (moderately critical) reviews. In addition, threaded reviews (comments on reviews, and so on) are enabled. Our tool helps draw out the argumentative relationships among the reviews.

In product reviews, contributors comment about what to buy, justify their comments, and disagree with other comments about the merits, importance, or use of particular features. In these situations, understanding the justifications and disagreements about facts, consequences, values, and point of view is key to making sense of the overall discussion. For subjective aspects, the impact of the statement may depend on the extent to which consumers share values and viewpoints. Such qualitative aspects of the comments are not captured by quantitative measures of the discussion; the most popular comment may not advance the analysis or may only sway individuals susceptible to popular opinion.

2. Components of Analysis

We have identified five "tiers" of analysis: a consumer argumentation scheme, a set of discourse indicators, sentiment terminology, a user model, and a domain model. Broadly speaking, each user (given his model of himself and other users) justifies whether or not to purchase a camera, where the justification is expressed in terms of the elements of the tiers of analysis. We outline the tiers and some of their relationships; however, we also present an analytical approach, leaving the models to be fully specified.

Consumer Argumentation Scheme (CAS) Reviews are a social space with purposeful discussions that include argumentative content, where the purpose is to decide what camera to buy. Derived from the Value-based Practical Reasoning argumentation scheme [1], we propose an argumentation scheme for camera-buying that can be used to represent the arguments concerning a course of action relative to preferences and values. Even where people agree on the facts of a situation, they may make different rational choices of products based on their preferences and values.

For our purposes, we can take certain aspects as fixed, namely the attributes of the user (user model) and the camera (domain model), so they are beyond the scope of the argument. In the context of product purchase recommendations, the main point is that the owner should benefit from a camera that has particular properties, where the properties promote a value. For example, a camera with a fast shutter speed can be construed to promote the value of ease, because buying and hence owning such a camera would allow a person to take pictures rapidly; if one wants to satisfy that value, then one should buy a camera with the relevant property. We focus on such arguments, presented as a schema:

Consumer Argumentation Scheme (CAS):
Premise: Camera *X* has property *P*.
Premise: Property *P* promotes value *V* for agent *A*.
Conclusion: Agent *A* should *Action1* camera *X*.

We note that arguments about properties are objective while arguments about values are inherently subjective. What counts as promoting a value depends on the particular agent; further, any particular value may or may not be important to that agent. For our purposes, the premise about values encapsulates this information.

Additional propositions (given their justification) may imply or contradict the premises of the CAS. There may be rich *cascades of arguments*, but we leave specification of additional schemes to future development. There are various ways to critique the argument, considering: whether an item has some specific properties (relatively uncommon as this is objective); whether a property promotes some value; whether one value is more important than another, and others.

Discourse Indicators Discourse indicators are overt linguistic expressions of discourse relations within or between statements [17], binding statements into larger scale textual units such as an argument. The analysis of discourse indicators and relations is complex: there is a range of indicators, a spectrum of associated discourse relations, multiple senses for indicators (e.g. *but* has several meanings), and implicit discourse relations. However, in this study, we keep to a closed class of explicit indicators. The explicit indicators are used to signal potentially relevant passages; it remains for the analyst to resolve ambiguities in context. Out of the range of indicators, we use the following.

> *Indicators of premise*: after, as, because, for, since, when, assuming, ...
>
> *Indicators of conclusion*: therefore, in conclusion, consequently, ...
>
> *Indicators of contrast*: but, except, not, never, no, ...

Sentiment Terminology In addition, we make use of *sentiment* terminology that signals lexical semantic contrast: *The flash worked poorly* is the semantic negation of *The flash worked flawlessly*, where *poorly* is a negative sentiment and *flawlessly* is a positive sentiment. An extensive list of terms is classified according to a sentiment scale from highly negative to highly positive [8].

User Models The properties of users contribute to the substance of the user's review and the user's reaction to other reviews. The properties can be given in several subclasses such as:

> *User's parameters*: Age, gender, education, previous camera experience, ...
>
> *User's context of use*: party, indoors, sport, travel, desired output format, ...
>
> *User's constraints*: cost, portability, size, richness or flexibility of features, ...
>
> *User's quality expectations*: colour quality, information density, reliability, ...

The properties can be used to describe subclasses of users. For instance, a professional photographer is likely to have a selection of cameras tuned to particular purposes, while a casual snapper may use a single camera for all purposes. Value judgements about properties typically relate to a class, e.g. action photographers value quick shutter time, which is not as relevant to portrait photographers. We leave for future development an ontological analysis.

Camera Domain We have terminology from the camera domain that specifies the objects and properties that are relevant to the users. Analysing the corpus, consumer report magazines (e.g. Which?), and a camera ontology[3], we identified some of the prominent terminology such as properties with binary values (such as *has a flash*), properties with ranges (such as the number of megapixels, scope of the zoom, or lens size), and multi-slotted properties (e.g. the warranty).

3. Components of the Tool

To recognise the textual elements of Section 2, we use the GATE framework [2], which we briefly describe along with our approach to analysis, the representation of textual elements using GATE, components discussed in Section 2, and outline the *processing pipeline*. In Section 4, we provide the results of sample queries.

GATE is a framework for language engineering applications, which supports efficient and robust text processing [2]; it is an open source desktop application written in Java that provides a user interface for professional linguists and text engineers to bring together a wide variety of natural language processing tools and apply them to a set of documents. The tools are formed into a pipeline of natural language processors. Our approach to GATE tool development follows [18], which is: bottom-up, rule-based, unweighted, modular, iterative, incremental, among others. Once a GATE pipeline has been applied to a corpus, we can view the annotations of a text either *in situ* or extracted using GATE's ANNIC (ANNotations In Context) corpus indexing and querying tool.

For our purposes, we emphasise the role of *gazetteers* and *JAPE rules*, which form the *bottom level* of the analysis: a gazetteer is a list of words that are associated with a central concept; JAPE rules are transductions that take annotations and regular expressions as input and produce annotations as output. We have gazetteers for the terminology for discourse indicators, sentiment terminology, user models, and camera domain.

The gazetteer associates textual passages in the corpus that match terms on the lists with an annotation, e.g. a token term *aperture* is annotated with *cameraproperty*. The gazetteer thus annotates related terms (e.g. *aperture* and *shutter speed*) with the same annotation. The annotations introduced by gazetteers are used by JAPE rules, creating annotations that are visible as highlighted text and are easily searchable in ANNIC. Querying for an annotation, we retrieve all the terms with the annotation. The annotations can also be used to create higher level annotations, e.g. all base level annotations for negative terminology can also receive a global annotation as *Negative*.

4. Output and Queries

Once the corpus is annotated, we can view the annotations *in situ*. In Figure 1, we have a paragraph from a five star review, where the annotations indicate CameraProperty and PremiseIndicator (differentiated by colour in the orginal). We see justification for the statement *I adore this camera*. The justification has two parts that are conjoined by *and*: *I get a higher precentage of good shots out of it than from any other camera I have*; *the video is ridiculously good*.

[3] `http://www.co-ode.org/ontologies/photography/`

I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system

Figure 1. Graphical display of text with annotations

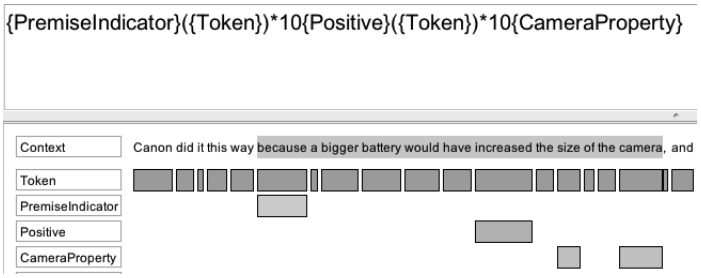

Figure 2. {PremiseIndicator}({Token})*10{Positive}({Token})*10{CameraProperty}

Alternatively, we can use the ANNIC tool to index and query a database of annotated text. Searching in the corpus for single annotations returns all those strings that are annotated with the search annotation along with their context and source document. Complex queries can also be formed. A query and a sample result appear in Figure 2, where the query finds all sequences of annotated text where the first string is annotated with *PremiseIndicator*, followed by zero to ten other *Tokens*, followed by a string with a *Positive* sentiment, another zero to ten *Tokens*, and finally ending with a *CameraProperty*. The search returned 42 candidate structures. The extract identifies a justification for an aspect of the camera design, which could be used to counter a negative comment about the battery size. With the query language, we can search for any of the annotations in the corpus in any order and refine them. In this way, the analyst can explore the annotated corpus for relevant passages. In the next section, we discuss how these outputs can be used to reconstruct arguments.

5. Analysis of Arguments and their Evaluation

The objective of the tool is to find specific patterns of terminology in the text that can be used to instantiate the CAS argumentation scheme both for and against purchase of the Canon PowerShot SX220 HS Digital Camera. We iteratively search the corpus for properties, instantiate the argumentation scheme, and identify attacks. Once we have instantiated arguments in attack relations, we may evaluate the argumentation framework. Our focus in this paper is the identification of arguments and attacks from the source material rather than evaluation. It is important to emphasise that we provide an analyst's *support tool*, so some degree of judgement is required.

From the results of queries on the corpus, we have identified the following premises bearing on *image quality*, where we paraphrase the source and infer the values from context. Agents are also left implicit, assuming that a single agent does not make contradictory statements. The premises instantiate the CAS in a positive form, where *A1* is an argument for buying the camera, and in a negative form, where *A2* is an argument for not buying the camera.

A1. P1: The pictures are perfectly exposed.
P2: The pictures are well-focused.
P3: No camera shake.
P4: Good video quality.
V1: These properties promote image quality.
C1: Therefore, you (the reader) should by the Canon SX220.
A2. P5: The colour is poor when using the flash.
P6: The images are not crisp when using the flash.
P7: The flash causes a shadow.
V2: These properties demote image quality.
C2: Therefore, you (the reader) should not buy the Canon SX220.

For *A1* to be an "good" argument for buying the camera, the argument *A2* must be defeated. We defeat *A2* as we find propositions (e.g. neg-P5) that are not attacked and that are negations of premises (e.g. P5) of *A2*:

neg-P5: For good colour, use the colour setting, not the flash.
neg-P6: No need to use flash even in low light.
neg-P7: There is a corrective video about the flash shadow.

Some of the incompatibilities between statements must be inferred for they do not have an obvious form of logical contradictions. Rather, they appear as implications from *alternative ways of addressing* the problem. For instance, *P5: The colour is poor when using the flash* is countered with *neg-P5*, which is a suggestion not to use the flash and to use a colour setting. This highlights the need to develop *richer tools to identify negation*. Given the arguments and propositions, the rebuttal of *A1* is defeated, and the reader may take *A1* as an undefeated argument to buy the Canon SX220.

6. Related Work and Future Directions

We have provided an implementation of a rule-based text analytic tool to support analysts in identifying relevant textual passages that can be reconstructed into argumentation schemes and attacks. As such, it is a semi-automated system. We do not, then, evaluate it according to *recall* and *precision* [7], which are appropriate for a domain with a well-developed *gold standard corpus*. Indeed, the tool could be helpful in developing such a gold standard corpus. The tool is a significant advance over current, graphically-based argument extraction tools that rely on the analysts' unstructured, implicit, non-operationalised knowledge of discourse indicators and content [14,11,6]. While arguments have been analysed from web-based texts, no tool is provided [5,13]. Two logic programming approaches have been proposed. In [9], statements are classified according to rhetorical roles using full sentence parsing and semantic translation. In [12], highly structured argumentative texts are processed using a stand-alone tool that is not integrated with other NLP tools. These logic programs do not use argumentation schemes, user or domain models, or support annotation search.

The tool can be further developed by adding argumentation schemes, disambiguating discourse indicators [17], and developing the user models. Other aspects of the analysis, value and domain terminology, require further structure. Asides from extending the corpus, more elaborate query patterns could be executed to derive more specific results. The open plan of the analysis and the flexibility of the tool provide a platform for future, detailed solutions to problems such as negation and ambiguity.

Acknowledgements

The second author's work was supported by Science Foundation Ireland under Grant No. SFI/09/CE/I1380 (Líon2) and a Short-term Scientific Mission grant from COST Action IC0801 on Agreement Technologies. The first author gratefully acknowledges support by the FP7-ICT-2009-4 Programme, IMPACT Project, Grant Agreement Number 247228. The views expressed are those of the authors.

References

[1] K. Atkinson and T. J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[2] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, pages 168–175, 2002.

[3] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[4] U. Egly, S. A. Gaggl, and S. Woltran. ASPARTIX: Implementing argumentation frameworks using answer-set programming. In M. G. de la Banda and E. Pontelli, editors, *ICLP*, volume 5366 of *Lecture Notes in Computer Science*, pages 734–738. Springer, 2008.

[5] S. Heras, K. Atkinson, V. J. Botti, F. Grasso, V. Julián, and P. McBurney. How argumentation can enhance dialogues in social networks. In P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari, editors, *COMMA*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 267–274. IOS Press, 2010.

[6] A. D. Liddo and S. B. Shum. Cohere: A prototype for contested collective intelligence. In *ACM Computer Supported Cooperative Work (CSCW 2010) - Workshop: Collective Intelligence In Organizations - Toward a Research Agenda*, Savannah, Georgia, USA, February 2010.

[7] R. Mitkof, editor. *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 2003.

[8] F. Å. Nielsen. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CoRR*, abs/1103.2903, 2011.

[9] V. Pallotta and R. Delmonte. Automatic argumentative analysis for interaction mining. *Argument and Computation*, 2(2-3):77–106, 2011.

[10] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, Jan. 2008.

[11] G. Rowe and C. Reed. Argument diagramming: The Araucaria Project. In A. Okada, S. B. Shum, and T. Sherborne, editors, *Knowledge Cartography: Software Tools and Mapping Techniques*, pages 163–181. Springer, 2008.

[12] P. Saint-Dizier. Processing natural language arguments with the <TextCoop>platform. *Argument & Computation*, 3(1):49–82, 2012.

[13] F. Toni and P. Torroni. Bottom-up argumentation. In *Proceedings of the First International Workshop on the Theory and Applications of Formal Argumentation (TAFA11)*, volume 11, pages 291–303, Barcelona, Spain, 2011.

[14] T. van Gelder. The rationale for Rationale. *Law, Probability and Risk*, 6(1-4):23–42, 2007.

[15] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Erlbaum, Mahwah, N.J., 1996.

[16] D. Walton and E. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY series in logic and language. State University of New York Press, 1995.

[17] B. Webber, M. Egg, and V. Kordoni. Discourse structure and language technology. *Natural Language Engineering*, December 2011. Online first.

[18] A. Wyner and W. Peters. On rule extraction from regulations. In K. Atkinson, editor, *Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference*, pages 113–122. IOS Press, 2011.