Knowledge-Driven Argument Mining: what we learn from corpus analysis

Patrick Saint Dizier

HAL Id: hal-01436201

https://hal.science/hal-01436201

Submitted on 16 Jan 2017

Knowledge-Driven Argument Mining:
what we learn from corpus analysis

Patrick SAINT-DIZIER [a,1],

[a] *IRIT-CNRS, 118 route de Narbonne, 31062 Toulouse Cedex France*

Abstract. Given a controversial issue, argument mining from texts in natural language is extremely challenging: besides linguistic aspects, domain knowledge is often required together with appropriate forms of inferences to identify arguments. Via the the analysis of various corpora, this contribution explores the types of knowledge that are required to develop an efficient argument mining system.

Keywords. Argument mining, Natural language processing, Knowledge representation

1. Aims and Challenges of Argument Mining

One of the main goals of argument mining is, given a controversial issue, to identify, in a set of texts, the arguments for or against that issue. Arguments are difficult to identify, in particular when they are not adjacent to the controversial issue, possibly not in the same text, because their linguistic, conceptual or referential links to that issue are rarely direct and explicit. Arguments are often evaluative natural language statements which become arguments because of the specific relations they have with another evaluative statement.

Except in specific contexts, most statements do not have any specific linguistic mark that would allow to directly identify as arguments. It is difficult to identify whether a statement is a support or an attack of a controversial issue, and what it precisely attacks or supports. Argument mining has to deal with two major problems: (1) **reference**: the argument is thematically related to the issue, and (2) **relatedness**: what facets of the issue are involved, how and how much.

Argument mining is an emerging research area which introduces new challenges both in natural language processing (NLP) and in argumentation. It requires the combination of linguistic analysis and language processing with artificial intelligence technology. Argument mining research applies to written texts, e.g. (Mochales Palau et ali.., 2009), (Kirschner et ali., 2015), for example for opinion analysis, e.g. (Villalba et al., 2012), mediation analysis (Janier et al. 2015) or transcribed dialogue analysis, e.g. (Budzynska et ali., 2014), (Swanson et ali., 2015). The analysis of the NLP techniques relevant for argument mining from annotated structures is analyzed in e.g. (Peldszus et al. 2016). Annotated corpora are now available, e.g. the AIFDB dialog corpora or (Walker et al., 2012). These corpora are very useful to understand how argumentation is realized in

---

texts, e.g. to identify argumentative discourse units (ADUs), linguistic cues (Nguyen et al., 2015), and argumentation strategies, in a concrete way, possibly in association with abstract argumentation schemes, as shown in e.g. (Feng et al., 2011). Finally, reasoning aspects related to argumentation analysis are developed in e.g. (Fiedler et al., 2007) and (Winterstein, 2012) from a formal semantics perspective with a study of concessive and contrastive connectors. In opinion analysis, the benefits of argument mining are not only to identify the customers satisfaction level, but also to characterize why customers are happy or unhappy. Abstracting over arguments allows to construct summaries and to define customer preferences or value systems (e.g. low fares are preferred to localization or quality of welcome for some categories of hotel customers).

This paper focuses on the corpus construction and analysis and identifies and categorizes the needs in terms of knowledge to perform efficient argument mining.

2. Corpus Construction and Analysis

To explore and characterize the forms of knowledge that are required to develop argument mining in texts, we constructed and annotated four corpora based on four independent controversial issues. These corpora are relatively small, they are designed to explore the problem, and to elaborate the main features of a more extended empirical analysis and model, but not to design a comprehensive knowledge-driven argument mining system.

For this first experiment, we considered the four following issues, which involve very different types of arguments, forms of knowledge (concrete or relatively abstract) and language realizations. These issues are: (1) Ebola vaccination is necessary, (2) Women's conditions have improved in India, (3) The development of nuclear plants is necessary, and (4) Organic agriculture is the future.

For each of these issues, the corpus constructed and the different arguments found (eliminating duplicates or closely related arguments) are summarized in Table 1. The text fragments which are investigated are extracts from various sources where these issues are discussed, in particular: newspaper articles and blogs from associations. These are documents accessible to a large public, with no professional consideration. Language is French (glosses are given is paper) or English. For example: issue 1: The Lancet (UK journal), a French journal that develops results from the Howard Hughes Medical Institute, a web site: Ebola vaccines, therapies, and diagnostics Questions and Answers ; issue 2: http://saarthakindia.org/womens_situation_India.html, http://www.importantindia.com/20816/ women-in-india-role-and-status-of-women-in-india/; issue 3: Pour/Contre le nucleaire (les centrales) web site in French, etc.

Issue	Corpus size (text extracts)	nb. of annotated different arguments	overlap rate: average nb of similar arguments
(1)	16 texts, 8300 words	50	4.7
(2)	10 texts, 4800 words	27	4.5
(3)	7 texts, 5800 words	31	3.3
(4)	23 texts, 6200 words	22	3.8
Total	56 texts, 25100 words	130	4.07

Table 1. Corpus typology

This corpus shows that the argument diversity per issue is not very large. A preliminary annotation task includes the annotation of all the arguments that have been found. In a second stage, arguments judged by the annotator to be similar or redundant are bound and count for a single utterance. A high overlap rate has been observed: while there are original arguments, authors tend to borrow quite a lot of material from each other, for example in (4) an average repetition rate of 3.8 has been observed. This overlap rate clearly depends on the annotator (personal analysis and knowledge of the issue domain) and is somewhat subjective. In spite of this subjectivity, this rate gives an interesting rough redundancy level. A more detailed analysis of those repetitions would be of much interest from a rhetorical and sociological perspective.

An argument and its context (the discourse structures that modify it) are tagged between XML <argument> tags with attributes. The argument kernel is tagged <main arg>. Attributes informally characterize the knowledge that is required. Tags encode typical features associated with arguments (text span, polarity, strength) and specific features proper to our investigation; these are:

- the **text span involved**, which ranges from a few words to a short paragraph. Arguments are numbered for referencing needs,

- the **discourse relations** associated with the argument, these are annotated using the tags defined in our TextCoop platform (Saint-Dizier, 2012).

- the **polarity of the argument** w.r.t. the issue which has one of the following values: support, concession (argumentative concession is a weak support), contrast (a weak attack), and attack. The concession and contrast categories have been introduced to account for cases where the support or the attack are weak. These however need to be defined more precisely (formally and in annotation guidelines). Concessions and contrasts are both discourse structures and criteria to evaluate argument polarity.

- the **conceptual relation with the issue**: why it is an attack or a support,

- the **knowledge involved**, when appropriate, to identify the argument: list of the main concepts used. These come preferably from a predefined domain ontology, or from the annotator intuitions, if none is available. This list may be quite informal, it nevertheless contributes to identify the nature of the knowledge involved to identify arguments,

- the **strength of the argument**.

A tagged argument for issue (1) is the following:

<argument nb= 11, polarity= concession ,
relationToIssue= limited proofs of efficiency and safety of vaccination,
conceptsInvolved= efficiency measure, safety measures, test and evaluation methods,
strength= moderate>
<concession> Even if the vaccine seems 100% efficient and without any side effects on the tested population, < /concession>
<main arg> it is necessary to wait for more conclusive data before making large vaccination campaigns < /main arg>
<elaboration> The national authority of Guinea has approved the continuation of the tests on targeted populations.</elaboration> < /argument>.

From our manual analysis, the following argument polarities are observed: attacks: 53 occurrences, supports: 33, argumentative concessions: 21, argumentative contrasts: 19 and undetermined: 4. The corpus shows a tendency to argue against an issue: attacks and contrasts = 55%, supports and concessions = 41%. The need of knowledge to iden-

tify the relation between an issue and an argument is summarized in Table 2.

Issue	need of knowledge nb of cases (rate)	total number of concepts involved (estimate)
(1)	44 (88%)	54
(2)	21 (77%)	24
(3)	18 (58%)	19
(4)	17 (77%)	27
Total	100 (77%)	124

Table 2. Evidence for Knowledge

These figures show that for about 77% of the arguments, some form of knowledge is involved to establish an argumentative relation between a statement and a controversial issue. An important result is that the number of concepts involved is not very large: 124 concepts for 100 arguments over 4 domains. Even if the notion of concept remains vague, these results are nevertheless interesting to develop large argument mining systems. These considerations are exemplified below.

3. Analysis of the types of knowledge involved in argument identification

Our analysis below tends to suggest that the Generative Lexicon (Pustejovsky 1986) (GL), with some improvements, is an adequate representation framework to deal with knowledge based argument mining. The main other lexical semantics approaches such as FrameNet or VerbNet mainly concentrate on verb's predicate argument structure and adjuncts: they characterize the roles that these elements (NPs, PPs a,d S) play in the verb and proposition meaning. According to our observations, this are not central features for knowledge-based argument mining. The GL has a relatively simple argument structure feature, which could be enhanfced with FrameNet data, however, w.r.t. our analysis, the features of much interest are the purposes and goals of an object or action, its origin, and its uses. These are those which are evaluated in arguments. These, to the best of our knowledge, are specific features of the GL Qualia structure, in particular the telic and the agentive roles. A main limitation is that the GL has very little resources available.

The GL is an attempt to structure lexical semantics knowledge in conjunction with domain knowledge via a decompositional view of lexical meaning. The Qualia structure of an entity is composed of four fields called roles:
- the constitutive role describes the various parts of the entity and its physical properties, it may include subfields such as material, parts and components,
- the formal role describes what distinguishes the entity from other objects, i.e. the entity in its environment, in particular the entities which are more generic. It may also be structured into several subroles such as shape, dimensions, position, etc.
- the telic role describes the entity functions, uses, roles and purposes,
- the agentive role describes the origin of the entity, how it was created or produced.
Let us now develop the analysis of each corpus.

Main concepts in arguments related to issues (1) and (4)

For issues (1) and (4), arguments mainly attack or support salient features of the main concepts of the issue and closely related ones by means of various forms of evalu-

ations. Samples of arguments found for issue (1) are:

Supports: *vaccine protection is very good; Ebola is a dangerous disease; high contamination risks; vaccine has limited side-effects, no medical alternative to vaccine, etc.*

Attacks: *limited number of cases and deaths compared to other diseases; limited risks of contamination, ignorance of contamination forms; competent staff and P4 lab difficult to develop; vaccine toxicity and high side-effects,*

Concessions or Contrasts: *some side-effects; high production and development costs; vaccine not yet available; ethical and freedom problems.*

For issue (1), the concept *vaccine* is the root of the system. The facets of this concept and the closely related ones used in arguments can be organized as follows:
(1) the parts of a vaccine: the *adjuvant* and the active principle; (2) its super types: a vaccine is a kind of medicine; and (3) the most central aspects of the concept vaccine w.r.t. argument mining, namely its purposes, goals and consequences, and how it is created, tested and sold. For example, the concepts of *side-effect* and *toxicity* are consequences of using a medicine; the concept of *contamination* is related to one of the purposes of a vaccine, namely to avoid *disease dissemination*; and *production costs* are related to the creation and development of any product, including medicines and vaccines, etc.

Without knowing that a vaccine protects humans from getting a disease, it is not possible, e.g. to say that *prevents high contamination risks* is a support for issue (1). Similarly, without the knowledge that the active principle of a vaccine is diluted into an adjuvant that is also injected, it is not possible to say that *the adjuvant is toxic* is an attack, this statement could be e.g. purely neutral or irrelevant to the issue.

The terms used in this short analysis: purpose, properties, creation and development, etc. are foundational aspects of the structure of a concept, relatively well defined in the GL. Using this knowledge allows to identify arguments that attack or support an issue and how. Positive or negative terms found in statements are not sufficient to determine arguments, as shown and illustrated in (Saint-Dizier 2016).

Main concepts in arguments related to issue (2)

The arguments related to issue (2) mainly involve comparisons with men's living conditions or refer to general principles of human welfare. Some arguments are justified by means of figures while others remain vague, possibly not up-to-date or relative to specific situations: women living and social conditions may indeed evolve over time, and may differ depending on locations. Samples of arguments found for issue (2) are:

Supports: *increased percentage of literacy among women; women are allowed to enter into new professional fields; at the upper primary level, the enrollment increased from 0.5 million girls to 22.7 million girls.*

Attacks: *practices of female infanticide, poor health conditions and lack of education still persisting; home is women's real domain; they are suffering the violence afflicted on them by their own family members; malnutrition is still endemic.*

No concessions or contrasts have been observed: argument polarities are very clearcut, with a very positive or negative tonality, proper to highly controversial and overheated debates.

The concepts used in arguments related to issue (2) concentrate on facets of humans in the society. The identification of these concepts is crucial to characterize argumentative relations between issue (2) and their polarity. For example, improving literacy means higher education, better jobs and therefore more independence and social recognition, which are typical of living condition improvements.

The concepts used in arguments for issue (2) can be classified into two categories:
- those related to the services provided by the society to individuals: education, safety, health, nutrition, human rights, etc.
- those related to the roles or functions humans can play in the society: job and economy development, family development, cultural and social involvement, etc.
The first category is close to the formal role of the GL Qualia, while the latter is close to the telic role. These categories act as types which are characterized by more precise predicates in Qualia roles. For example the family category may include:

```
have(X, family), develop(X, equality in family), educate(X,
children), feed(X, family), choose(X, husband/wife).
```

Arguments pros or cons develop evaluations of these concepts for women. In addition, lexical inferences must be developed to establish a complete link between words found in arguments and the main concepts illustrated above. For example, *malnutrition* is provoked by a 'lack of food' and entails 'health problems'.

Main concepts in arguments related to issue (3)

In the case of issue (3) (nuclear plants are necessary), supports or attacks mainly involve comparisons between various sources of energy. Samples of arguments are:

Supports: *energy independence; creates high technology jobs; risks are over-estimated; wastes are well-managed and controlled by AIEA; preserves the other natural resources,*

Attacks: *there are alternative solutions with less pollution; alternatives create more jobs than nuclear; there are risks of military uses : more dangerous than claimed; nuclear plants have high maintenance costs, etc.*

Concessions or Contrasts: *nuclear plants use dangerous products, but we know how to manage them; difficult to manage nuclear plants, but we have competent persons,*

In terms of language realizations, arguments are essentially constructed on the basis of comparatives and facts related to consequences, purposes or uses of nuclear plants. Utterances are quite developed and refer to knowledge about energy, pollution, health, or various forms of dangers, etc.

In terms of knowledge, nuclear plants have relatively complex agentive (construction, maintenance) and telic (roles, purposes) roles which must be decomposed into facets. For this latter facet, subtypes can be defined which develop its main purposes and secondary purposes and consequences, such as the production of nuclear wastes. The same structures must be developed for other sources of energy (coal, sea, wind, etc.) so that comparisons can be evaluated. Most arguments lack precise comparative data, e.g. wastes are all said to be a nuisance, but the degree is not given. These arguments require knowledge to be evaluated against precise data, otherwise they rather play on the pathos. These are called underspecified arguments. Finally, some arguments are fallacious, for example, the 'energetic independence' advocated as a support for nuclear plants is not acceptable since uranium is bought abroad: the use of knowledge contributes to detect them.

4. Summary: from concepts to knowledge representation for argument mining

The introduction of knowledge in argument mining is a crucial feature which allows to improve the identification of:
(1) the potential relatedness of an evaluative statement with a controversial issue,

(2) which aspect(s) of the issue it deals with, characterized by a specific facet in a Qualia role, identified by one or more predicates,

(3) the argument polarity: attack, support, concession, contrast,

(4) neutral, irrelevant, underspecified (when knowledge is not comprehensive enough to decide on the polarity of an argument) and fallacious arguments.

The fully annotated corpus (French and English) will be shortly made available.

In terms of language realizations, the following three main types of argument realizations are observed:

(1) use of **evaluative expressions** (52% of cases), in attribute-value form, where the attribute is proper to the concepts of the controversial issue or of closely derived concepts: *Vaccine development is very expensive, adjuvant is toxic.* Resources which are required are a set of Qualias representing the concepts, the lexical terms which are associated with each element of the Qualia and their quasi-synonyms,

(2) use of **comparatives** applied to related concepts (e.g. electricty produced by nuclear energy compared to coal or wind, 21% of the cases): Qualia of related concepts are needed with lexical realizations, e.g. *number of sick people much smaller than for Malaria, nuclear wastes are more dangerous than coal wastes*.

(3) **evaluation of uses, consequences or purposes** of the main concept of the issue (27% of the cases) requires Qualia structures and associated lexical entries and, when relevant, lexical inference and presuppositions associated with those lexical entries. Examples are: *vaccine prevents bio-terrorism, women are allowed to enter into new profesional fields.*

Knowledge-based argument mining requires relatively complex sources of knowledge, however, an important result is that for a given issue, only a few conecpts (and their Qualias) are needed. We have breifly shown elements of GL Qualia structures seem to be an appropriate knowledge representation framework, with an adequate conceptual typing (e.g. various forms of telicity) which can be combined with lexical data and lexical inference. In addition, this approach allows to:

- reason about arguments on the basis of a domain knowledge representation,

- develop conceptually relevant forms of argument synthesis or clustering, based on Qualia roles.

An open issue remains the acquisition of the Qualia structure data. Given that relatively few Qualias are involved for a controversial issue, some features can be developed manually. An important investigation direction is the acquisition of additional features, probably based on bootstrapping techniques.

5. Conclusion

In this short paper, we present a preliminary analysis of the different forms of knowledge which are frequently required to relate arguments with a given controversial issue. Our short corpus analysis shows that the type of information proposed by the Qualia structure of the GL is a useful knowledge source: object parts, action or object properties, uses, functions, creation and development are often the main topics for arguments. Obviously, this analysis must be expanded in various directions before any development of a knowledge-driven argument mining system. A more detailed empirical analysis is necessary, with a larger corpus, several annotators, and annotation guidelines before conluding that e.g. the Qualia structure is the main resource that is needed.

References

[1] Budzynska, K., Janier,M., Reed, C., Saint-Dizier, P., Stede, M., Yakorska, O., A model for processing illocutionary structures and argumentation in debates. In proc. LREC, 2014.

[2] Feng, V. W., Hirst, G., Classifying arguments by scheme. In Proceedings of the 49th ACL: Human Language Technologies, Portland, USA, 2011.

[3] Fiedler, A., Horacek, H., Argumentation within deductive reasoning. International Journal of Intelligent Systems, 22(1):49-70, 2007.

[4] Janier,M., Reed, C., Towards a Theory of Close Analysis for Dispute Mediation Discourse, Journal of Argumentation, 2015.

[5] Kirschner, C., Eckle-Kohler, J., Gurevych, I., Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications. In: Proceedings of the 2nd Workshop on Argumentation Mining, Denver, 2015.

[6] Mochales Palau, R., Moens, M.F., Argumentation mining: the detection, classification and structure of arguments in text. Twelfth international ICAIL'09, Barcelona, 2009.

[7] H., Nguyen, D., Litman, D., Extracting Argument and Domain Words for Identifying Argument Components in Texts. In: Proc of the 2nd Workshop on Argumentation Mining, Denver, 2015.

[8] Peldszus, A., Stede, M., From argument diagrams to argumentation mining in texts: a survey. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 2016.

[9] Pustejovsky, J., The Generative Lexicon, MIT Press, 1995.

[10] Saint-Dizier, P., Processing natural language arguments with the TextCoop platform, journal of Argumentation and Computation, vol 3(1), 2012.

[11] Swanson, R., Ecker, B., Walker, M., Argument Mining: Extracting Arguments from Online Dialogue, in proc. SIGDIAL 2015, USA.

[12] Villalba, M.G., Saint-Dizier, P., Some Facets of Argument Mining for Opinion Analysis, COMMA, Vienna, IOS Publishing, 2012.

[13] Walker, M., Anand, P., Fox Tree, J.E., Abbott,R., King, J. A Corpus for Research on Deliberation and Debate. Proc. of the Language Resources and Evaluation Conference (LREC), Istanbul, 2012.

[14] Winterstein, G., What but-sentences argue for: An argumentative analysis of 'but', in Lingua 122, 2012.