

Combining Argumentation and Aspect-Based Opinion Mining: The SMACk System

Mauro Dragoni, Célia da Costa Pereira, Andrea G. B. Tettamanzi, Serena Villata

▶ **To cite this version:**

Mauro Dragoni, Célia da Costa Pereira, Andrea G. B. Tettamanzi, Serena Villata. Combining Argumentation and Aspect-Based Opinion Mining: The SMACk System. AI Communications, 2018, 31 (1), pp.75 - 95. 10.3233/AIC-180752 . hal-01721538

HAL Id: hal-01721538
https://inria.hal.science/hal-01721538

Submitted on 9 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining Argumentation and Aspect-Based Opinion Mining: The SMACk System *

Mauro Dragoni[1], Celia da Costa Pereira[2], Andrea G. B. Tettamanzi[3], Serena Villata[3]
[1]*Fondazione Bruno Kessler, Trento, Italy*
[2]*Université Côte d'Azur, I3S, CNRS, France*
[3]*Université Côte d'Azur, CNRS, Inria, I3S, France*
E-mail:
dragoni@fbk.eu
celia.pereira|andrea.tettamanzi@unice.fr
villata@i3s.unice.fr

The extraction of the relevant and debated opinions from online social media and commercial websites is an emerging task in the opinion mining research field. Its growing relevance is mainly due to the impact of exploiting such techniques in different application domains from social science analysis to personal advertising. In this paper, we present SMACk, our opinion summary system built on top of an argumentation framework with the aim to exchange, communicate and resolve possibly conflicting viewpoints. SMACk allows the user to extract debated opinions from a set of documents containing user-generated content from online commercial websites, and to automatically identify the mostly debated positive aspects of the issue of the debate, as well as the mostly debated negative ones. The key advantage of such a framework is the combination of different methods, i.e., formal argumentation theory and natural language processing, to support users in making more informed decisions, e.g., in the context of online purchases.

Keywords: Argumentation Theory, Opinion Mining, Aspect Extraction

1. Introduction

Argumentation theory is a reasoning model based on the construction and evaluation of information pieces called *arguments*. Arguments are supposed to support, contradict, and explain statements, and they are used to support decision making [52]. Argumentation theory involves different ways for analyzing arguments and their relationships. A famous framework is the one called *abstract argumentation* proposed by Dung [19], which views each argument as an abstract entity and in which arguments are related to each other by means of attack relations. What distinguishes argumentation-based discussions from other approaches is that proposals can be supported by the arguments that justify, or oppose, them. This permits a greater flexibility than in other decision-making and communication schemes since, for instance, it makes it possible to persuade the other actors involved in the discussion to change their view of (or opinion about) a claim by identifying information or knowledge that is not being considered, or by introducing a new relevant factor in the middle of a negotiation or to resolve an impasse. Indeed, in the process of making a decision (or in a debate), we have to consider the opinion of all the people involved in the process. Such opinions are subjective statements that represent people's sentiments, emotions, perceptions, and mood about a particular object or subject.

Opinion mining [34] aims at analyzing texts about a given subject or topic, written in a natural language, to classify them with respect to the opinion they are supposed to convey or more specifically, with respect to the *polarity* (negative, positive or neutral) or emotions of the authors of such texts. Such natural language documents can be analyzed at three different levels: *(i)* at a sentence level — in this case we are interested in extracting the opinion or sentiment behind one single sentence; *(ii)* at a document level — in this case we are interested in the overall opinion or sentiment behind the entire document, and *(iii)* at an aspect level — in this case we are interested in a more fine-grained opinion or sentiment associated to the most relevant aspects or features in the document.

Concerning the particular case of sentiment analysis [34], the analysis at document and sentence levels allows us to extract a global feeling behind the document or the sentence. However, as pointed out by Chinsha and Joseph [10] for example, the authors of reviews often convey more than one single feeling in a single review by describing their sentiment about different as-

*This paper is an extended version of the demo paper published at IJCAI-2016 [16].

pects of the same subject. Let us consider the following text from a restaurant review: "Good food but the restaurant's location is too far from the city center"). In this example, we can distinguish two aspects, i.e., "food" and "location", with a positive sentiment with respect to the food, and a negative sentiment with respect to the location of the restaurant. If we contextualize this sentence in a debate, the final overall sentiment of the author about the restaurant will depend on how she/he will be convinced by others about this subject, i.e., it will depend on the strength of the arguments for and against the two considered aspects. However, a single system combining all these components is currently lacking.

More precisely, a framework aiming at combining aspect-based opinion mining methods to deal with the extraction of the aspects and their polarities in natural language documents together with formal argumentation theory to deal with the efficient computation of the arguments' strengths is needed. The advantage of this combined framework is twofold. On the one hand, we want to build an argumentation graph by extracting, from opinion content, triples structured as $\langle Aspect, Opinion, Attack \rangle$ or $\langle Aspect, Opinion, Support \rangle$, where *Aspect* is a property of an object (for instance, the "display" of a monitor), and *Opinion* is a value connected with such a property in the interval $[0,1]$, where 0 represents the most negative polarity ("bad" or "very bad", depending on the granularity of the scale of judgment), 1 is the most positive polarity ("good" or "very good"), and values inbetween represent graded or mixed opinions. On the other hand, we want to exploit implicit information users provide supporting their viewpoint by adopting quick tagging facilities (for example, the "like" and "dislike" buttons of some Web interfaces) to integrate the generated argumentation framework with information that is not explicitly mentioned in the text, but that we are able to infer through the analysis of user actions. For instance, if the triple $\langle$*"display"*$, 1, Support\rangle$ is extracted from a product review, and a user "likes" such review, it means that all the nodes that will be inserted in the argumentation graph will have a further *Support* information even if it has not been explicitly mentioned in textual form.

In this paper, we present SMACk 1.0, our argumentation-based opinion mining framework which is focused on the analysis of online user-generated content. More precisely, such a framework is based on the use of abstract argumentation theory [19], and it supports the detection and extraction of relevant opinions from a set of textual documents.

The paper is organized as follows: Section 2 discusses the related work and compares it to the proposed approach; in Section 3 some basics about abstract argumentation theory and fuzzy set theory are provided together with the description of the linguistic resources we used; Section 4 introduces the combined SMACk framework; Section 5 presents the fuzzy labeling algorithm for abstract argumentation introduced in [14] and its empirical evaluation, and in Section 6 we detail the approach implemented for extracting aspects from text. Then, Section 7 presents the user interface we implemented for supporting researchers and experts in data analysis. Finally, the framework is evaluated in Section 8. Conclusions end the paper.

2. Related Work

The proposed contribution embraces two research fields, argumentation theory and aspect-based opinion mining. To the best of our knowledge, it is the first time that these disciplines are combined for improving the effectiveness of a real-world sentiment analysis system. In this section, we give the context of each research field by highlighting the most relevant work which have inspired this contribution.

Aspect-Based Opinion Mining. Mining people's opinions may be included in the class of problems whose solution requires the processing of textual information. However, while techniques like information retrieval, for example, aim at processing facts or objective statements in order to extract useful information, opinion mining aims at extracting the views, sentiments, emotions, etc., from people's judgements.

Generally, given a text, the goal of opinion mining techniques is to infer the overall polarity of such a text by summarizing user's opinion [5]. Kumar and Ravi [30] survey published literature from 2002 to 2015 in which different approaches to opinion mining and sentiment analysis are described. In particular, we would like to stress the relevance with respect to what we are proposing here of the ones concerning the aspect-based sentence segmentation model and the ones concerning the aspect-based review summarization model.

Recently, the need of performing a more fine-grained analysis of the text emerged. Indeed, when we express our views about a restaurant, a movie, or the like, the reviews are not always global but express different sentiments about the different aspects

we deem interesting enough to influence our global sentiment. This task, known as "aspect-based sentiment analysis" [54], aims at detecting each single entity feature mentioned in a document and to infer the polarity of the text with respect to it. The importance of this type of analysis task has grown significantly in the last years. Initiatives dedicated to the comparison of real-world systems have also emerged, such as SemEval [46]. The systems usually presented at SemEval implement supervised approaches trained with sample datasets provided to participants.

Different kinds of supervised and unsupervised learning approaches have been proposed for extracting opinion targets and the opinion or sentiment conveyed by texts: most are based on conditional random fields (CRF) [26,11,66,40]; others are based on hidden Markov models (HMM) [28], sequential rule mining [33], dependency tree kernels [63], graph propagation [18], and clustering [56].

Supervised systems suffer from the necessity of exploiting manually labeled training data for building their models. The consequence is the limitation of the system validation in a production environment where such systems have to be applied to several domains. A time-saving way for overcoming this issue is the adoption of unsupervised approaches. Unsupervised approaches are mainly based on topic modeling [39,57,31,4,62,41,42] and syntactic rules designed using dependency relations [69,61,63,65,50]. There are also word frequency-based methods [25,55,47,68], word alignment methods [36], and label propagation methods [67].

Contributions described above are focused on the analysis of single documents and do not support opinion aggregation. If we extend the application of opinion mining techniques to a repository of documents, open challenges are (i) the detection of the most relevant aspects mentioned by users, (ii) how such aspects are considered by them, and (iii) the implementation of a scalable approach for managing a huge amount of documents.

The extraction of relevant information from a text is an activity usually performed by text summarization systems [43]. Text summarization aims to extract key information from a text, which is then presented to users as summaries or, in some cases, as lists of relevant keywords. The extraction task can be done on a single document or on a set thereof [35]. The main advantage of applying text summarization techniques is the possibility to relieve users from reading entire documents containing irrelevant details with respect to the topic of the document and the final goal of users.

While text summarization is the most suitable solution when parts of the documents are not particularly informative from the user perspective, they are not a solution when the goal is to go beyond the extraction of interesting information from texts.

In the scenario proposed in this paper, the adoption of text summarization approaches would not be the most suitable solution for the following reasons. First of all, the considered documents already contain only relevant information. Indeed, generally, when a user writes a review, she does not spend a lot of time and space to contextualize the review, but she provides only the most important aspects she wants to share with the community. Therefore, it is not requested to summarize documents content, but to aggregate the different opinions they express. Second, the desired output of the analysis of user-generated content is not only the extraction of the most relevant aspects, but the detection of the most debated ones about a particular topic. The approach has to be able to extract all the aspects of the given topic and to rank them from several perspectives (user agreement, user disagreement, polarities, etc.).

Two real-world implementations of unsupervised systems are presented by Popescu and Etzioni [48], and Bjorkelund *et al.* [3]. The former presents OPINE, a review-mining system able to perform aspect-based sentiment analysis and opinion ranking in an unsupervised way using relaxation labeling. The latter exploits aspects extracted from hotel reviews to support user decisions and to suggest effective visualizations of customers' reviews. This use of visualization is a remarkable common point with our work; however, we go steps beyond (i) by integrating argumentation theory in the aspect extraction task, (ii) by using external knowledge for clustering semantic similar aspects, (iii) by building a knowledge graph representing the connections between users and aspects extracted from a whole reviews' repository, and (iv) by integrating a labeling algorithm of the knowledge graph in order to detect most relevant information for supporting users' decision.

Argumentation and Opinion Mining. There exist only few approaches proposed in the literature coupling argumentation theory and opinion mining. Argumentation is applied to opinion mining by Grosse and colleagues [22]. In this article, the authors extract a particular version of arguments they called "opinions" based on incrementally generated queries. Given a query, they model an opinion supporting it as a set of aggregated tweets along with a prevailing sentiment, which

can be attacked by alternative counter-opinions. As a final result, they obtain what they call a *conflict tree*, rooted in the first original query, in a way that resembles dialectical trees in argumentation. Their final goal is to detect conflicting elements in an opinion tree to avoid potentially inconsistent information.

Argumentation mining [44,32] certainly shares some analogies with opinion mining and sentiment analysis, as highlighted also by Habernal et al. [23], even if the goals of the two differ: the goal of opinion mining is to understand what people think about something, while the aim of argumentation mining is to understand why, which implies looking for causes and reasons rather than just for opinions and sentiments. In the context of argument mining techniques, opinionated claim analysis has been addressed by Rosenthal and McKeown [53]. The authors released two datasets of 2,000 sentences each, with the purpose of extracting so-called opinionated claims. These consist of 285 LiveJournal blogposts and 51 Wikipedia discussion forums, respectively. They address a problem that is closer to sentiment analysis, and their aim is to detect assertions containing some belief, of the truth of which a user is attempting to convince an audience. Argumentative opinion analysis has been addressed by Saint-Dizier and colleagues [60], with the TextCoop platform, which basically constructs arguments from opinions and supportive elements such as illustrations and evaluative expressions, by using a set of handcrafted rules that explicitly describe rhetorical structures.

Both the goals and the adopted methodologies of these contributions are different from the one we present in this paper. Actually, the main difference is that we do not exploit opinion mining for argument mining tasks, e.g., argument or relation extraction, but we use abstract argumentation theory together with the fuzzy labeling algorithm, to organize the information extracted and classified through opinion mining methods, and support users in decision making.

3. Preliminaries

In this section, we provide some insights about abstract argumentation theory and fuzzy sets.

3.1. Abstract argumentation theory

We provide the basics of Dung-like abstract argumentation theory [19].

Definition: (Abstract argumentation framework) An abstract argumentation framework is a pair $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A}$ is a set of elements called *arguments* and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation called *attack*. We say that an argument A_i attacks an argument A_j if and only if $(A_i, A_j) \in \rightarrow$.

Dung [19] presents several acceptability semantics which produce zero, one, or several sets of accepted arguments. These semantics are grounded on two main concepts, called conflict-freeness and defence.

Definition: (Conflict-free, Defence) Let $C \subseteq \mathcal{A}$. A set C is *conflict-free* if and only if there exist no $A_i, A_j \in C$ such that $A_i \rightarrow A_j$. A set C *defends* an argument A_i if and only if for each argument $A_j \in A$ if A_j attacks A_i then there exists $A_k \in C$ such that A_k attacks A_j.

Definition: (Acceptability semantics) Let C be a conflict-free set of arguments, and let $\mathcal{D} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ be a function such that $\mathcal{D}(C) = \{A \mid C \text{ defends } A\}$.

- C is *admissible* if and only if $C \subseteq \mathcal{D}(C)$.
- C is a *complete extension* if and only if $C = \mathcal{D}(C)$.
- C is a *grounded extension* if and only if it is the smallest (w.r.t. set inclusion) complete extension.
- C is a *preferred extension* if and only if it is a maximal (w.r.t. set inclusion) complete extension.
- C is a *stable extension* if and only if it is a preferred extension that attacks all arguments in $\mathcal{A} \setminus C$.

The concepts of Dung's semantics are originally stated in terms of sets of arguments. We can also express these concepts using argument *labeling* [27,59, 7]. In a reinstatement labeling [7], an argument is labeled "in" if all its attackers are labeled "out" and it is labeled "out" if it has at least an attacker which is labeled "in".

Definition: (AF-labeling [7]) Let $\langle \mathcal{A}, \rightarrow \rangle$ be an abstract argumentation framework. An AF-labeling is a total function $lab : \mathcal{A} \rightarrow \{in, out, undec\}$. We define $in(lab) = \{A_i \in \mathcal{A} \mid lab(A_i) = in\}$, $out(lab) = \{A_i \in \mathcal{A} \mid lab(A_i) = out\}$, $undec(lab) = \{A_i \in \mathcal{A} \mid lab(A_i) = undec\}$.

Definition: (Reinstatement labeling [7]) Let lab be an AF-labeling. We say that lab is a reinstatement labeling iff it satisfies the following:

- $\forall A_i \in \mathcal{A} : (lab(A_i) = out \equiv \exists A_j \in \mathcal{A} : (A_j \rightarrow A_i \wedge lab(A_j) = in))$ and
- $\forall A_i \in \mathcal{A} : (lab(A_i) = in \equiv \forall A_j \in \mathcal{A} : (A_j \rightarrow A_i \supset lab(A_j) = out))$ and
- $\forall A_i \in \mathcal{A} : (lab(A_i) = undec \equiv \exists A_j \in \mathcal{A} : (A_j \rightarrow A_i \wedge \neg(lab(A_j) = out)) \wedge \nexists A_k \in \mathcal{A} : (A_k \rightarrow A_i \wedge lab(A_k) = in)$.

A reinstatement labeling is called *illegal* if the above conditions do not hold.

Definition: (Complete, grounded, preferred and stable labeling) *lab* is a complete labeling if there are no arguments illegally *in*, illegally *out*, or illegally *undec*. We say that *lab* is a

- *grounded*, iff $in(lab)$ is minimal (w.r.t. set inclusion);
- *preferred*, iff $in(lab)$ is maximal (w.r.t. set inclusion);
- *stable*, iff $undec(lab) = \emptyset$

3.2. Fuzzy Sets

Fuzzy sets [64] are a generalization of classical (crisp) sets obtained by replacing the characteristic function of a set A, χ_A, which takes up values in $\{0,1\}$ ($\chi_A(x) = 1$ iff $x \in A$, $\chi_A(x) = 0$ otherwise) with a *membership function* μ_A, which can take up any value in $[0,1]$. The value $\mu_A(x)$ or, more simply, $A(x)$ is the membership degree of element x in A, i.e., the degree to which x belongs in A.

A fuzzy set is completely defined by its membership function. Therefore, it is useful to define a few terms describing various features of this function. Given a fuzzy set A, its *core* is the (conventional) set of all elements x such that $A(x) = 1$; its *support*, $\text{supp}(A)$, is the set of all x such that $A(x) > 0$. A fuzzy set is *normal* if its core is nonempty. The set of all elements x of A such that $A(x) \geq \alpha$, for a given $\alpha \in (0,1]$, is called the α-cut of A, denoted A_α.

The usual set-theoretic operations of union, intersection, and complement can be defined as a generalization of their counterparts on classical sets by introducing two families of operators, called triangular norms and triangular co-norms. In practice, it is usual to employ the min norm for intersection and the max co-norm for union. Given two fuzzy sets A and B, and an element x,

$$(A \cup B)(x) = \max\{A(x), B(x)\}; \quad (1)$$

$$(A \cap B)(x) = \min\{A(x), B(x)\}; \quad (2)$$

$$\bar{A}(x) = 1 - A(x). \quad (3)$$

Finally, given two fuzzy sets A and B, $A \subseteq B$ if and only if, for every element x, $A(x) \leq B(x)$.

3.3. Sentiment Lexicons

Sentiment Lexicons are used for associating each term with a polarity value. Terms having such an association are called "opinion words" and they are used for estimating the polarity of a given sentence. Associating a polarity value to a specific word is a task that has been addressed by different perspectives. The results have been the availability of different resources that can be easily integrated within real-world systems. In SMACk, we decided to aggregate polarity values coming from three resources freely available: SenticNet [6], the General Inquirer vocabulary [1] [45], and the MPQA dictionary[2] [15].

SenticNet is a publicly available resource for opinion mining that exploits both artificial intelligence and semantic Web techniques to infer the polarities associated with common-sense concepts and to represent them in a semantic-aware format. In particular, SenticNet uses dimensionality reduction to calculate the affective valence of a set of Open Mind[3] concepts and it represents them in a machine accessible and processable format. The development of SenticNet was inspired by SentiWordNet [1], a lexical resource in which each WordNet synset is associated to three numerical scores describing how objective, positive, and negative the terms contained in each synset are. The differences between SenticNet and SentiWordNet are basically three: (i) in SentiWordNet, each synset is associated to a three-valued representation (the objectivity of the synset, its positiveness, and its negativeness), while in SenticNet there is only one value belonging to the $[-1,1]$ interval for representing the polarity of the concept; (ii) SenticNet provides the sentiment model of more complex common-sense concepts, while SentiWordNet is focused on assigning polarities to WordNet synsets: for instance, in SenticNet, complex concepts like "make good impression", "look attractive", "show appreciation" , "being fired", "leave behind", or "lose control" are used for defining positive or negative situations; and (iii) completely neutral concepts are not reported. SenticNet contains 30,000 polarity concepts and it may be connected with any kind of opinion mining application. For example, after the de-construction of the text into concepts through a semantic parser, SenticNet can be used to associate polarity values to these and, hence, to infer the overall

[1] http://www.wjh.harvard.edu/ inquirer/spreadsheet_guide.htm
[2] http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/
[3] http://realitycommons.media.mit.edu/

polarity of a clause, sentence, paragraph, or document by averaging such values.

The *General Inquirer* is an English-language dictionary containing almost 12,000 elements associated with their polarities in different contexts. Such dictionary is the result of the integration between the "Harvard" and the "Lasswell" general-purpose dictionaries as well as a dictionary of categories defined by the dictionary creators. When necessary, for ambiguous words, specific polarity for each sense is specified. For every word, a set of tags is provided in the dictionary. Only a subset of them are relevant to the opinion mining topic and, thus, exploited in this work:

- Valence categories: the two well-known "positive" and "negative" classifications.
- Semantic dimensions: these tags reflect semantic differential findings regarding basic language universals. These dimensions are: "hostile", "strong", "power", "weak", "submit", "active", and "passive". A word may be tagged with more than one dimension, if appropriate.
- Words of pleasure: these tags are usually also classified positive or negative, with virtue indicating strength and vice indicating weakness. They provide more focus than the categories in the previous two bullets. Such categories are "pleasure", "pain", "feel", "arousal", "emotion", "virtue", "vice".
- Words reflecting presence or lack of emotional expressiveness: these tags indicate the presence of overstatement and understatement; trivially, such tags are "overstated" and "understated".

Other categories indicating ascriptive social tags rather than references to places have been considered out of the scope of the opinion mining topic and have not been considered in the implementation of the approach.

Finally, *MPQA* is a sentiment lexicon built for Multi-Perspective Question Answering purposes. The lexicon contains around 8,222 terms annotated with their polarity ("positive", "negative", and "neutral") and with their intensity level ("strong" and "weak") and a set of 10,000 sentences manually annotated through the proposed annotation scheme. Indeed, besides the classic association $\langle word, polarity \rangle$, the MPQA lexicon implements a detailed annotation scheme that identifies key components and properties of opinions, emotions, sentiments, speculations, evaluations, and private states [51]. This annotation scheme covers a broad and useful subset of the range of linguistic expressions and phenomena employed in naturally occurring text to express opinion and emotion. The proposed annotation scheme is relatively fine-grained, annotating text at the word- and phrase-level rather than at the level of the document or sentence. For every expression of a private state in each sentence, a private state frame is defined. A private state frame includes the source of the private state (i.e., that whose private state is being expressed), the target (i.e., what the private state is about), and various properties involving intensity, significance, and type of attitude. An important property of sources in the annotation scheme is that they are nested, reflecting the fact that private states and speech events are often embedded in one another. The representation scheme also includes frames representing material that is attributed to a source, but is presented objectively, without evaluation, speculation, or other type of private state by that source.

The lists of terms contained in the resources presented above do not overlap completely. The strategy implemented within the SMACk platform considers words with a non-zero polarity value in at least one of the integrated resources. For example, the word "third" is not present neither in MPQA nor in SenticNet and has a polarity of 0 according to the General Inquirer. Consequently, it is not a valid opinion word. On the other hand, the word "huge" has a positive value of 0.069 in SenticNet, a negative value of -1 in MPQA and a value of 0 in the General Inquirer, therefore, it is evaluated as opinion word even if lexicons express contrasting values. In case of contrasting values, we did not investigate particular strategies for aggregating them. SenticNet already implements a continuous representation of polarity values. MPQA uses a discrete scale $[-1, 0, 1]$ that has been extended to $[-1, -0.5, 0, 0.5, 1]$ by halving -1 and 1 when the "weak" intensity level is present. For the General Inquirer the same strategy adopted for the MPQA lexicon has been adopted by exploiting the semantic dimension of the dictionary for halving the -1 and 1 values. Finally, the three values are aggregated by using the arithmetic average.

3.4. WordNet

WordNet[4] [21] is a large lexical database of English nouns, verbs, adjectives, and adverbs grouped into sets of cognitive synonyms called *synsets*, where each synset expresses a distinct concept. In particu-

[4] https://wordnet.princeton.edu/

lar, each synset represents a list of synonyms, intended as words that denote the same concept and that are interchangeable in many contexts. WordNet contains around 117,000 synsets linked to each other by a small set of "conceptual relations", i.e., synonymy, hypernymy, hyponymy, etc.. Additionally, a synset contains a brief definition ("gloss") and, in most cases, one or more short sentences illustrating the use of the synset members. Words having several distinct meanings are represented in as many distinct synsets. Even if WordNet superficially resembles a thesaurus, there are some important distinctions with respect to it. Firstly, WordNet does not define links between words, but between specific senses of words; this way, words that are found in close proximity to one another in the network are semantically disambiguated. Secondly, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than the similarity of their meanings. In the implemented system, Wordnet's compound names list has been used to detect word sequences that represent a single noun.

3.5. Stanford Core NLP

The preliminary textual analysis, consisting in converting the raw input text in an annotated and structured representation, is performed through the Stanford Core Natural Language Processing Library. Stanford CoreNLP is an integrated framework providing a wide range of natural language analysis tools. Each functionality is provided by a specific module. Below, we show the four modules of the CoreNLP library adopted within the SMACk system.

The *Pos Tagger* (Part Of Speech Tagger) is a software module aiming to assign a part of speech tag (such as noun, verb, adjective, etc.) to every word of a given sentence [29]. The *Coref Annotator* (Coreference resolution Annotator) generates co-reference Chain Annotations representing groups of words referring to the same entity [12]. Chains are used to resolve pronoun references. The *Parse Annotator* (Parser Annotator) [24] provides full syntactic analysis generating a tree grammar dependencies structure. Finally, the *Depparse Annotator* (Dependency Parser Annotator) [9] provides a representation of grammatical relations between words in a sentence producing graphs.

4. SMACk Framework

The SMACk framework is composed by three main elements:

1. The *Argument* module (presented in Section 5) is in charge of detecting and extracting relevant sentences from raw natural language text. As "relevant sentences", we mean sentences containing mentions of entity aspects or opinion words. This module also has to extract the timeline with which aspects are mentioned. This information is very important for building the argumentation graph.
2. The *Sentiment* module (presented in Section 6) is in charge of inferring the polarity associated with each aspect extracted by the Argument module. Such an inference task is performed by taking into account the timeline detected by the Argumentation module. This way, it is possible to continuously update the polarity of each aspect within the discourse.
3. The *Visualization* module (presented in Section 7) is in charge of showing in real-time (i) how the extracted aspects are connected to each other, (ii) how much such aspects are interesting from the users' community perspective, and (iii) how the polarity of each aspect changes during time.

In the remainder of the paper, we describe how these modules have been integrated within the platform, exploiting a real-world use case to clarify the goal and the results of SMACk. In particular, we consider a set of product reviews belonging to one of the categories used within the Amazon website. We analyze a set of $50,000$ reviews extracted from the Dranziera dataset [17].[5] Then, we provide a couple of images showing the user interface and how information is presented to the user.

Figure 1 provides a graphical representation of the workflow executed for the analysis of each document set. As document set, we mean a time-ordered set of opinions expressed by users about the same domain, where, for each domain, many aspects can be mentioned and not all aspects have to be mentioned in all documents. The adoption of the time-ordered constraint is mandatory for detecting the attacks and the supports with respect to a specific aspect. Such information is used for building the argumentation graph.

[5] http://goo.gl/7jK4Rp

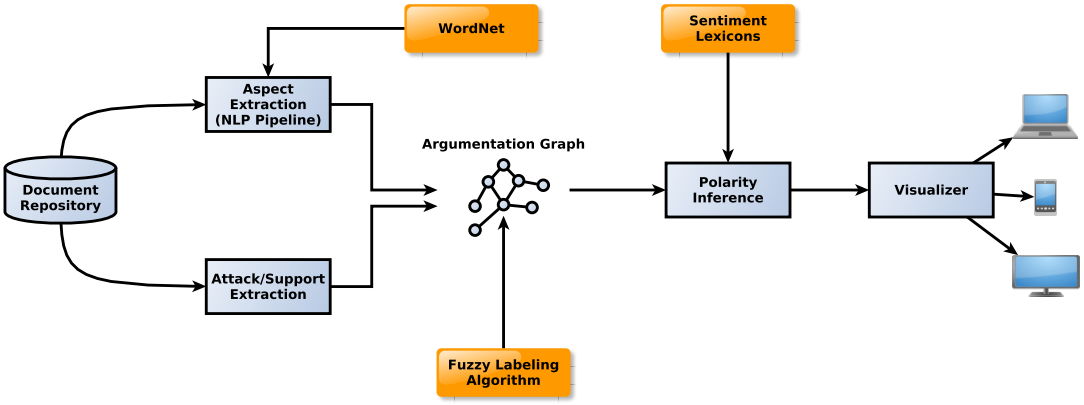


Fig. 1. The information workflow implemented within the SMACk system.

Construction of the argumentation graph. The first task is the construction of the argumentation graph given a set of user-generated contents about the same kind of product (for instance, "laptops"). For each text, the set of aspects and the qualities associated with them are extracted and stored under the form of structured information into knowledge nodes. Each node contains the following information:

- the average polarity expressed by users when they mention the specific aspect;
- the number of mentions of the aspect in the whole document set;
- the number of attacks to the opinion expressed with respect to a specific aspect; and,
- the number of supports to the opinion expressed with respect to a specific aspect.

All such information is exploited also by the Visualization Module (described in Section 7) for properly showing the graph summarization to users.

The result of this task is an argumentation graph where we have a node for each aspect, an edge connecting two or more aspects if they are mentioned in the same context, and a set of attack/support edges built based on the discussions extracted from the document set.

Extracting most interesting aspects. After the construction of the argumentation graph, the second task is the application of the labeling algorithm to compute the most interesting aspects emerging from the argumentation graph we constructed. In particular, the integration of the fuzzy labeling algorithm helps in the identification of the most interesting aspects from the user perspective.

This kind of analysis enables the extraction of relevant information that can be used for different purposes. For example:

- detecting which are the weaker aspects of a product (or a service) in order to drive future actions on possible product (or service) improvements;
- the exploitation of weak and strong aspects may inspire the development of personalized advertisement tools where information about the user viewpoints may help in making advertisement campaigns more effective;
- supporting users in making more informative decisions without the need to dig in the huge amount of (possibly technical) reviews about a product.

Besides the case study provided in this paper, the SMACk framework can be applied to several contexts with different levels of complexity. The main examples are in the social science domain, where a huge amount of texts needs to be analyzed for detecting the mood of people with respect to different debated topics, or the analysis of online user-generated content about products or services.

5. Fuzzy Labeling for Abstract Argumentation

In this section, we recall the main features of the fuzzy labeling algorithm for abstract argumentation [14]. For a complete description of the algorithm and its convergence theorem as well as the comparison with the related approaches, we refer the reader to [14]. Moreover, we report about the implementation we developed to test the performance of the algorithm [13].

In order to account for the fact that arguments may originate from sources that are trusted only to a certain

degree, the (crisp) abstract argumentation structure described in Section 3 may be extended by allowing gradual membership of arguments in the set of arguments $\mathcal{A}$. We have that $\mathcal{A}$ is a fuzzy set of trustworthy arguments, and $\mathcal{A}(A)$, the membership degree of argument A in $\mathcal{A}$, is given by the trust degree of the most reliable (i.e., trusted) source that offers argument A,[6]

$$\mathcal{A}(A) = \max_{s \in \mathrm{src}(A)} \tau_s, \tag{4}$$

where $\mathrm{src}(A)$ is the set of sources proposing argument A and τ_s is the degree to which source $s \in \mathrm{src}(A)$ is trusted.

Definition: (Fuzzy AF-labeling) Let $\langle \mathcal{A}, \rightarrow \rangle$ be an abstract argumentation framework. A fuzzy AF-labeling is a total function $\alpha : \mathcal{A} \rightarrow [0,1]$.

Such an α may also be regarded as (the membership function of) the fuzzy set of acceptable arguments: $\alpha(A) = 0$ means the argument is outright unacceptable, $\alpha(A) = 1$ means the argument is fully acceptable, and all cases inbetween are provided for.

Definition: (Fuzzy Reinstatement Labeling) Let α be a fuzzy AF-labeling. We say that α is a fuzzy reinstatement labeling iff, for all arguments A,

$$\alpha(A) = \min\{\mathcal{A}(A), 1 - \max_{B:B \rightarrow A} \alpha(B)\}. \tag{5}$$

The above definition combines two intuitive postulates of fuzzy labeling: (1) the acceptability of an argument should not be greater than the degree to which the arguments attacking it are unacceptable and (2) an argument cannot be more acceptable than the degree to which its sources are trusted: $\alpha(A) \leq \mathcal{A}(A)$.

We can verify that the fuzzy reinstatement labeling is a generalization of the crisp reinstatement labeling of Definition 3.1, whose *in* and *out* labels are particular cases corresponding, respectively, to $\alpha(A) = 1$ and $\alpha(A) = 0$. The intermediate cases, $0 < \alpha(A) < 1$ correspond to a continuum of degrees of "undecidedness", of which 0.5 is but the most undecided representative.

In order to compute, given a fuzzy argumentation framework, its fuzzy reinstatement labeling, we cast this problem as a problem of finding a solution to a system of n non-linear equations, where $n = \|\mathrm{supp}(\mathcal{A})\|$. Since iterative methods are usually the only choice for solving systems of non-linear equations, da Costa Pereira et al. [14] resort to this technique, by proposing an iterative algorithm.

We denote by $\alpha_0 = \mathcal{A}$ the initial labeling, and by α_t the labeling obtained after the t^{th} iteration of the labeling algorithm.

Definition: Let α_t be a fuzzy labeling. An iteration in α_t is carried out by computing a new labeling α_{t+1} for all arguments A as follows:

$$\alpha_{t+1}(A) = \min \left\{ \mathcal{A}(A), \frac{\alpha_t(A) + 1 - \max_{B:B \rightarrow A} \alpha_t(B)}{2} \right\}. \tag{6}$$

This defines a sequence $\{\alpha_t\}_{t=0,1,\ldots}$ of labelings which always converges to a limit fuzzy labeling [14]. Moreover, the convergence speed is linear: in practice, a small number of iterations is enough to compute the limit up to the desired precision. The fuzzy labeling of a fuzzy argumentation framework is thus the limit of $\{\alpha_t\}_{t=0,1,\ldots}$.

Definition: Let $\langle \mathcal{A}, \rightarrow \rangle$ be a fuzzy argumentation framework. A fuzzy reinstatement labeling for such argumentation framework is, for all arguments A,

$$\alpha(A) = \lim_{t \rightarrow \infty} \alpha_t(A). \tag{7}$$

An extensive evaluation of the performance and scalability of the fuzzy labeling algorithm with respect to a benchmark of abstract argumentation frameworks have been carried out in [13], to which we refer the interested reader. We first selected three existing datasets for abstract argumentation tasks used in the literature, namely the datasets created by Bistarelli *et al.* [2], by Cerutti *et al.* [8], and by Vallati *et al.* [58]. Moreover, we generated our own dataset of abstract argumentation frameworks by randomly combining some well known *graph patterns* in argumentation theory into 20,000 bigger argumentation frameworks. Secondly, we studied the behaviour of the algorithm with respect to the frameworks in the benchmark, to check whether their performance are satisfiable even considering huge and complex networks as those represented in the datasets, e.g., presenting an increasing number of strongly connected components. The experimental results reported in [13] clearly indicate that the fuzzy labeling algorithm scales up nicely in all circumstances, and is thus a viable argumentation reasoning tool.

Let us now consider a real world example, based on the data exploited by the SMACk framework. We

[6] Here, we suppose that the agent is optimistic. To represent a pessimistic behaviour, we should use the min operator, for example.

have the following 9 reviews published on a commercial website:

- *User1*: The display on this computer is the best I have seen in a very long time.
- *User5*: I hate the display screen and I have done everything I could do the change it.
- *User7*: The display is beyond horrible.
- *User8*: The display is awesome.
- *User10*: The colors of this display are very brilliant. Overall it is an amazing monitor.
- *User11*: This is the last time that I buy a display produced by them.
- *User14*: I do not agree with most of the reviews. This display is very nice by considering also its cheap price.
- *User15*: I like this display very much. I suggest it if you want to upgrade your hardware.
- *User18*: This display is the best choice I've done.

From these reviews, we construct the argumentation graph visualized in Figure 2 where the nodes represent the arguments (i.e., the aspects) and the edges represent attacks between arguments. Each node states whether a certain user expressed a positive or negative opinion about the product concerning a certain aspect, i.e., the *display* aspect of the computer. We assume that the reviews shown above have been published in the same temporal order. The reader may observe that the review provided by *User8* is in contrast also with the one provided by *User5*, but the attack is not reported in Figure 2. We must clarify that, in our approach, the argumentation graph is built by assuming that when a *UserX* provides a review, he/she attacks only the last review containing an opposite opinion with respect to the one provided by *UserX*. The rational behind this assumption is that our dataset does not contain details about the "topic-reply" relationship between messages. Thus, we decided to adopt this strategy to obviate this lack of knowledge. Part of our future work will focus on building datasets including also this information too.

We initialize each node in the argumentation graph to 1.0, meaning that at the beginning of the computation each user is considered to be trustworthy (as we do not have any prior knowledge about the trustworthiness degree of the users), and so holds for the information she conveys. This starting assumption is then confirmed or revised, depending on the attacks against the proposed arguments (i.e., aspects). Figure 3 shows the transition steps required to get to the final labeling, where U_1 represents the argument "User1, display, POSITIVE", and the same holds for the other arguments. In this specific real-world example, convergence is reached in 14 steps.

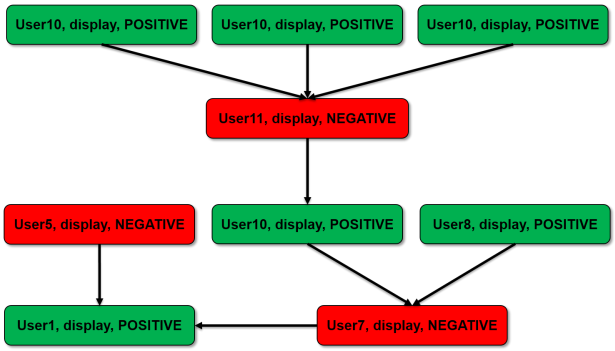


Fig. 2. Example of argumentation graph showing the aspects extracted from four reviews about the topic *display*.

6. Unsupervised Aspect-Extraction Algorithm

Here, we present the approach implemented for extracting aspects from text. The overall aspect extraction approach relies on the natural language processing (NLP) pipeline shown in the middle layer of Figure 4.

As it is shown on the bottom layer, the implemented pipeline exploits the three linguistic resources introduced in Section 3: a stopwords list,[7] sentiment lexicons, and a linguistic knowledge base. These resources are used by the a state of the art natural language processing library, namely Stanford Core NLP Library [38], shown in the top layer of Figure 4.

The pipeline is composed by five phases described as follows:

- Compound Noun Extraction. The first step consists in detecting the presence of compound names. This step is supported by the use of the POS-Tagger module provided by the Stanford Core NLP library and by WordNet (both introduced in Section 3). When two consecutive words are tagged as nouns by the POS-Tagger, their composition is searched within the WordNet dictionary. If the compound expression is found, it is tagged as compound name and used as a unique token, otherwise not.

[7] The used stopwords list is available at http://www.lextek.com/manuals/onix/stopwords1.html

t	$\alpha_t(U_1)$	$\alpha_t(U_5)$	$\alpha_t(U_7)$	$\alpha_t(U_8)$	$\alpha_t(U_{10})$	$\alpha_t(U_{11})$	$\alpha_t(U_{14})$	$\alpha_t(U_{15})$	$\alpha_t(U_{18})$
0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1	0.5	↓	0.5	↓	0.5	0.5	↓	↓	↓
2	0.25		0.25		0.5	0.25			
3	0.125		0.125		0.625	0.125			
4	0.0625		0.0625		0.75	0.0625			
5	0.0312		0.0312		0.843	0.0312			
6	0.0156		0.0156		0.906	0.0156			
7	0.0078		0.0078		0.945	0.0078			
8	0.0039		0.0039		0.968	0.0039			
9	0.0019		0.0019		0.982	0.0019			
10	0.0009		0.0009		0.990	0.0009			
11	0.0004		0.0004		0.994	0.0004			
12	0.0002		0.0002		0.997	0.0002			
13	0.0001		0.0001		0.998	0.0001			
14	0		0		0.999	0			

Fig. 3. Transition steps of the fuzzy labeling algorithm on the argumentation framework visualized in Figure 2.

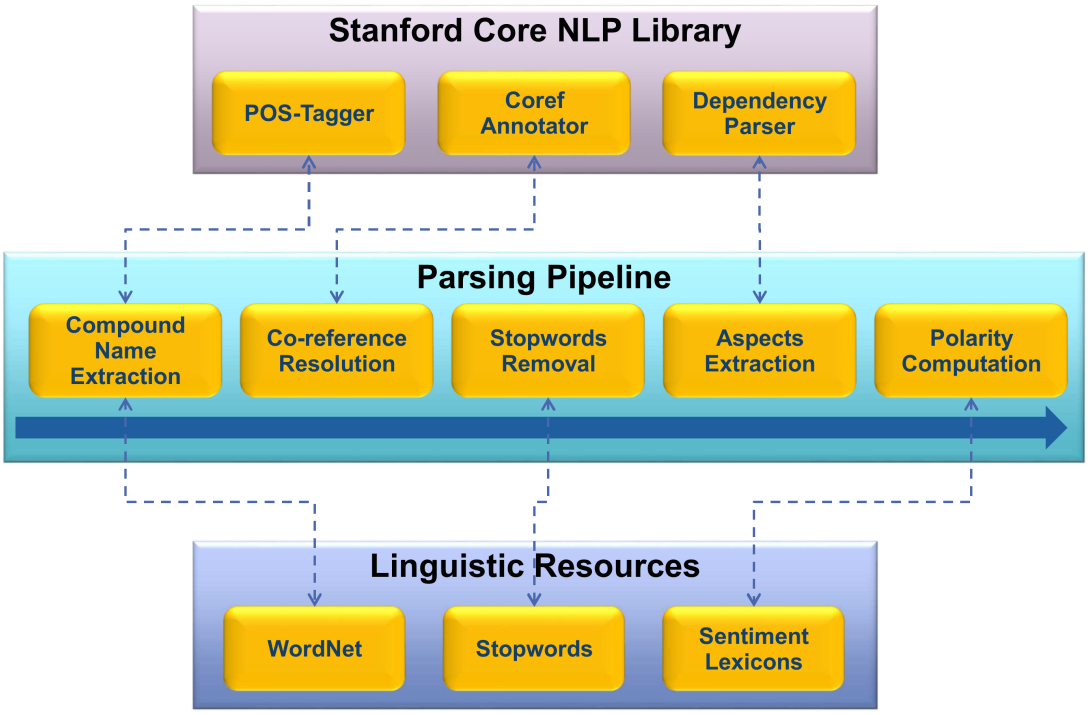


Fig. 4. The natural language processing pipeline implemented within the SMACk platform aiming to extract aspects and compute their polarity from the analyzed textual resources.

– Co-reference Resolution. This step consists in associating pronouns with the related noun (or compound noun). This is necessary for detecting all associations between opinion words and aspects. This operation is completely supported by the Coref Annotator. Refinements of the adopted algorithm are out of scope of this paper and they are part of future work.
– Stopwords Removal. Once compound names have been detected and pronouns have been replaced

with the right terms, the pipeline removes all stopwords from the text by exploiting the list mentioned above.
- Aspect Extraction. This step is the most important one and it consists on detecting the correct aspects contained in the text and the associated opinion words. Details about this step are provided below where we introduce the extraction algorithm adopted for analyzing provided text.
- Polarity Computation. Finally, this step is responsible of computing the polarity associated with each aspect extracted during the previous step. The overall polarity of an aspect A is computed by aggregating (as explained below) the single polarities of the opinion words associated with A. Single polarity values are extracted from the sentiment lexicons described in Section 3.

The aspect-extraction approach implemented within the SMACk platform is based on the Grammar-Dependencies-based algorithm described in [20]. Here, we briefly report how the algorithm works. For further details about the adopted strategy, we refer to [20].

The aspect-extraction algorithm analyzes the structure of the grammar dependency graph generated by the CoreNLP library for extracting the connections between aspects and opinions. Each dependency extracted by the CoreNLP library can be expressed by a triple: $\langle Relation_Type, Governor, Dependant \rangle$.[8] The most interesting peculiarity in using grammar dependencies graphs is the possibility of detecting sentiment relationships that are different from the "classic" *noun-adjective* combination.

Given a dependency node n, the algorithm checks if one of the following rules subsists:

Rule 1: If the relation type is an adjectival modifier ("amod"), a connection between an aspect and an opinion word is created if and only if the governor is an aspect and the dependant has a polarity value in at least one of the sentiment lexicons.

Rule 2: If the relation type is a nominal subject ("nsubj"), a connection between an aspect and an opinion word is created if and only if the governor has a polarity value in at least one of the sentiment lexicons and the dependant is an aspect.

Rule 3: If the relation type is a direct object ("dobj"), a connection between an aspect and an opinion word is created if and only if the governor has a polarity value in at least one of the sentiment lexicons and the dependant is an aspect.

By applying in ordered sequence the three rules mentioned above on the sentence "I enjoyed the screen resolution, it's amazing for such a cheap laptop", we obtain the following list of "aspect-opinion" associations:

$$\text{laptop} \leftarrow \{\text{cheap}\}$$
$$\text{screenresolution (it)} \leftarrow \{\text{amazing, enjoyed}\}$$

Afterwards, concerning the polarity computation task, given the associations above our system is able to infer, for both aspects "laptop" and "screen resolution", a positive polarity.

The polarity computation phase considers polarities of opinion words represented by trapezoidal fuzzy numbers in the $[-1,1]$ interval, reflecting the uncertainty resulting from the combination of different and possibly conflicting sentiment lexicons. The fuzzy polarities of the different opinion words occurring in a text for a given aspect are then aggregated, following the approach described in [18], by a fuzzy averaging operator.

7. User Interface

The SMACk platform is equipped with a simple user interface showing the result of documents analysis. Through such an interface, users are able to monitor the evolution of the argumentation graph and of the polarities associated with each extracted aspect. In the demo version presented in [16], we included three document sets related to three entities belonging to as many domains ("laptops", "restaurant", and "hotel"). The buttons for activating the demo working on each document set are placed in the right part of the interface as it is possible to observe in Figure 5.

This interface has been designed for supporting two kinds of users: "Researchers" and "Experts". The Researcher profile can observe how the argumentation graph is created during the documents analysis process. Figure 5 shows the construction of the argumentation graph during the analysis of the documents stream. Once the algorithm detects a new argument, it creates a colored node (green or red) into the graph containing a label composed by (i) the identifier of the opinion holder, (ii) the aspect mentioned in the review, and (iii) the polarity associated with the mentioned aspect. The polarity value affects the background color

[8] The meaning of each element of the triple together with all the possible relation type, can be found in the official Stanford Document available at http://nlp.stanford.edu/software/dependencies_manual.pdf

of the node: green for positive polarity, and red for the negative one. On the contrary, if the argument already exists, the algorithm evaluates if the content of the current review supports or attacks the existing argument. Based on this, the graph is updated accordingly.

The second profile supported by the tool is the "Expert" one. Experts can use an interface focusing on aspects monitoring in order to have a summary of the information associated with them. Figure 6 shows what happen while the system is running in this modality. In the right side of the graph area, two nodes are created: an "Agreement" node, colored in green, meaning the existence of users agreeing with opinions expressed for the connected aspects; and a "Disagreement" node, colored in red, meaning the existence of disagreements between users about opinions expressed for the connected aspects. The agreement or disagreement with an opinion are not directly connected with the polarity expressed for a particular aspect. For example, in the temporarily situation shown in Figure 6, four aspects have been detected so far: "food", "sushi", "place", and "space". For "food" and "space" there are both agreements and disagreements about expressed opinions; while, for the "sushi" and "place" there is an overall agreement about what has been expressed by users.

At the end of the textual analysis, the full graph is re-organized and each node is changed for showing (i) the level of interest of each aspect (i.e. how much users discussed about a specific product feature) and (ii) the overall polarity expressed about it. Figure 7 shows an example of such a result. For each node, we can distinguish two graphical elements supporting the analysis of the results. The first one is the border thickness of each node. The thicker the border, the more the aspect has been evaluated as interesting by the system. The second one is the background color of each node. We used a 11-gradient color scale for showing the overall polarity about each aspect. Such a color scale goes from green (positive polarity) to red (negative polarity).

Through this interface, users are able to analyze opinion streams and to have a real-time view on each aspect connected to a particular topic.

8. Evaluation

In this Section, we present the evaluation of the SMACk system. Such a system is evaluated under different perspectives aiming to show the efficiency and effectiveness of the modules implemented within the system. In [13], we already discussed about the efficiency of the fuzzy labeling algorithm supporting the discovery of the most interesting "nodes" in the generated graph. Here, we focus our evaluation on two perspectives:

- *Aspect extraction*. One of the tasks in charge to the Argumentation Module is the extraction of aspects from text. Such a task is important for defining, later in the analysis process, which aspects are the most significant ones. This evaluation task focused on measuring the effectiveness of the aspect-extraction approach.
- *Polarity detection*. The computation of the aspect's polarity enables the detection of which product features are strong or weak. The Sentiment Module is in charge of inferring the polarity of each aspect given the context in which such an aspect is included. Here, we measured the capability of the SMACk system of inferring the correct polarity.

Besides the effectiveness of the technological components, we provide also a discussion about the usability of the user interface and the direction we intend to follow for the evolution of the platform presented in this article.

8.1. Evaluation on Aspect Extraction

Table 1 reports the results obtained by our approach on the aspect extraction benchmark used in SemEval 2015 Task 12. The task consists in extracting from raw sentences the list of aspects towards which an opinion is expressed in the given text. For example, given the sentence: "This laptop is extremely portable and easily connects to WiFi at the library and elsewhere.", the term "laptop" has to be considered as aspect because opinions expressed through the terms "portable" and "easily connects" are associated with it. On the contrary, the term "library" does not have to be extracted because no opinions are associated with this term. Precision and recall are computed by using the well-known machine learning formulas [49] where "true positives" are the extracted aspects that are contained in the gold standard, "false positives" are the extracted aspects that are not contained in the gold standard, and "false negatives" are the aspects contained in the gold standard that are not detected by the algorithm. The algorithm has been tested on the "Restaurant" and "Laptop" datasets. The overall performance is in line

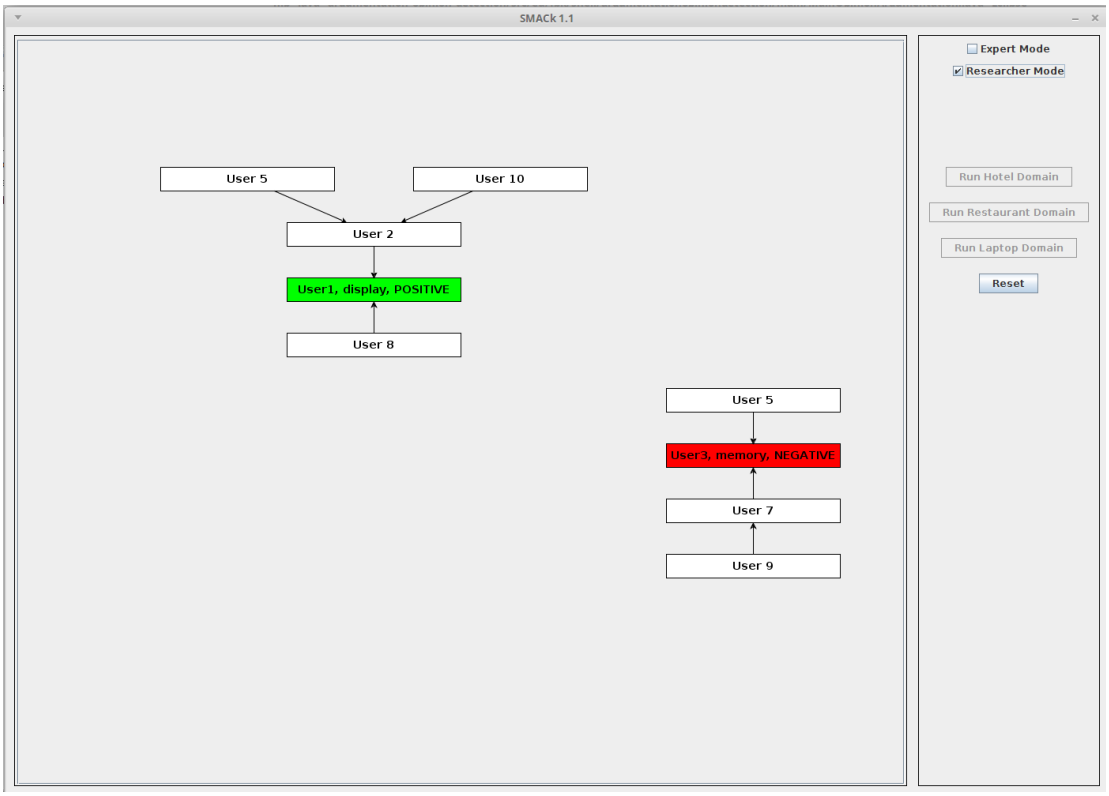


Fig. 5. The construction of the argumentation graph during the documents analysis process.

with the best systems participating in the evaluation campaign and, on the "Laptop" dataset, our aspect extraction approach recorded the best precision and F-measure. It is also important to highlight that all the systems we compared to, apply supervised approaches for extracting aspects, while our approach implements an unsupervised technique. This way, it is possible to implement the system in any environment without the requirement of training a new model.

Concerning the "Restaurant" domain, the gap between our approach and the best ones is given by the conservative strategy implemented for extracting aspects. One of the most common issue in unsupervised aspect-based approach is the extraction of false positive aspects [37]. The major consequence of such issue is the poor effectiveness of modules exploiting the outcome of the aspect extraction component. Unfortunately, the adoption of a conservative strategy leads to lower recall values. However, the latter is a preferable solution by considering the massive use of the aspects in the other components of the SMACk platform.

8.2. Evaluation on Polarity Computation

Table 2 reports the results of the polarity computation technique. The approach has been evaluated on the two datasets mentioned above. Here, we measured the accuracy of the polarity detection algorithm: given the set of opinion words associated with an aspect, such a polarity is computed by aggregating the fuzzy polarities of each opinion words. Results demonstrated the effectiveness of the polarity detection techniques implemented within the SMACk system by obtaining the best performance on the "Laptop" dataset, and the third best one on the "Restaurant" dataset. After a detailed analysis of the results, we noticed that the reason for which our approach performs better on the "Laptop" dataset is due to the simple language used for describing product features. Indeed, in the "Restaurant" dataset opinions are expressed in a more articulated way and sometimes the approach fails to detect the right polarity. Improvements in this direction will be part of the future work.

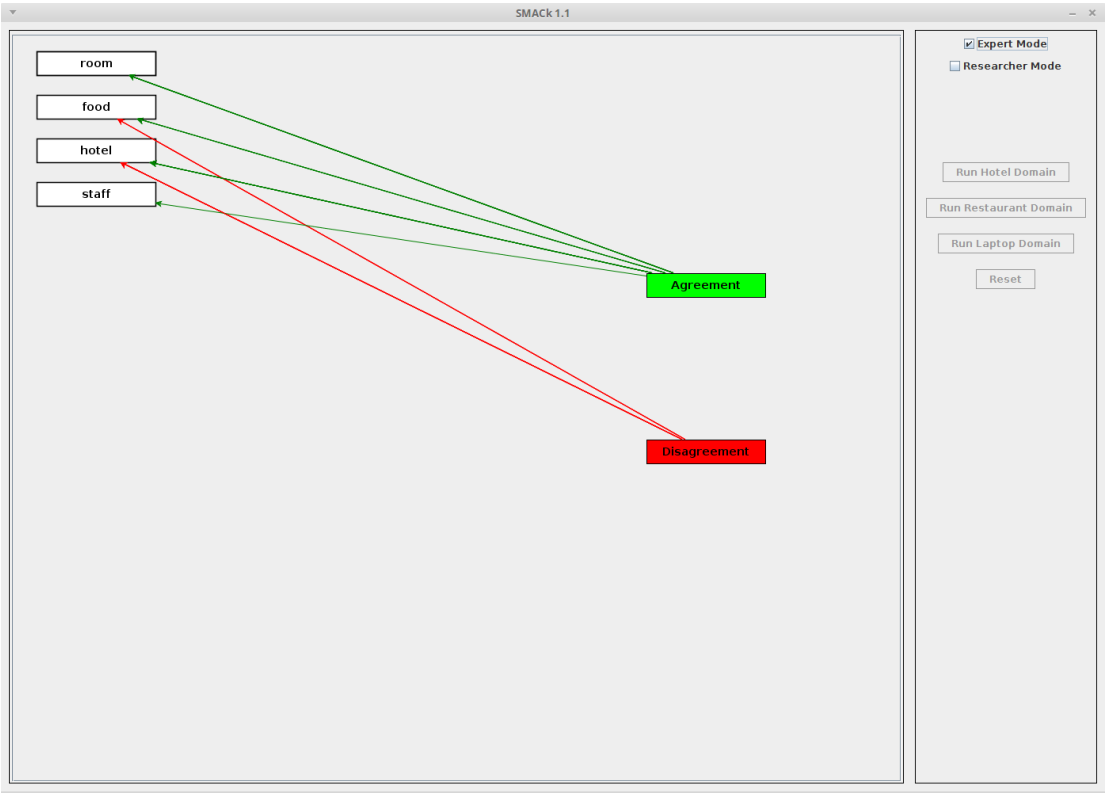


Fig. 6. The user interface during the analysis of a raw natural language text.

Table 1

Results obtained on the aspect extraction task, for the "Restaurant" and "Laptop" datasets, on the SemEval 2015 benchmark. For each dataset, we reported Precision, Recall, and F-Measure. Acronyms refer to the systems participated in the SemEval 2015 competition.

	Restaurant			**Laptop**		
System Acronym	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
IHS-RD-Belarus	**0.7095**	0.3845	0.4987	0.5548	0.4483	0.4959
LT3 pred	0.5154	0.5600	0.5367	-	-	-
NLANGP	0.6385	**0.6154**	**0.6268**	0.6425	0.4208	0.5086
sentiue	0.6332	0.4722	0.5410	0.5773	0.4409	0.5000
SIEL	0.6440	0.5135	0.5714	-	-	-
TJUdeM	0.4782	0.5806	0.5244	0.4489	**0.4820**	0.4649
UFRGS	0.6555	0.4322	0.5209	0.5066	0.4040	0.4495
UMDuluth	0.5697	0.5741	0.5719	-	-	-
V3	0.4244	0.4129	0.4185	0.2710	0.2310	0.2494
SMACk	0.6895	0.5368	0.6036	**0.6702**	0.4157	**0.5131**

8.3. Lessons Learned

Early in this section, we demonstrated the suitability of the components integrated within the SMACk platform. In Section 5, we showed how the system is efficient in resolving the labeling task for detecting the most important aspects contained in a set of user-generated documents. Furthermore, in the previous subsections, we validated the capability of the opinion mining module to extract the correct aspects

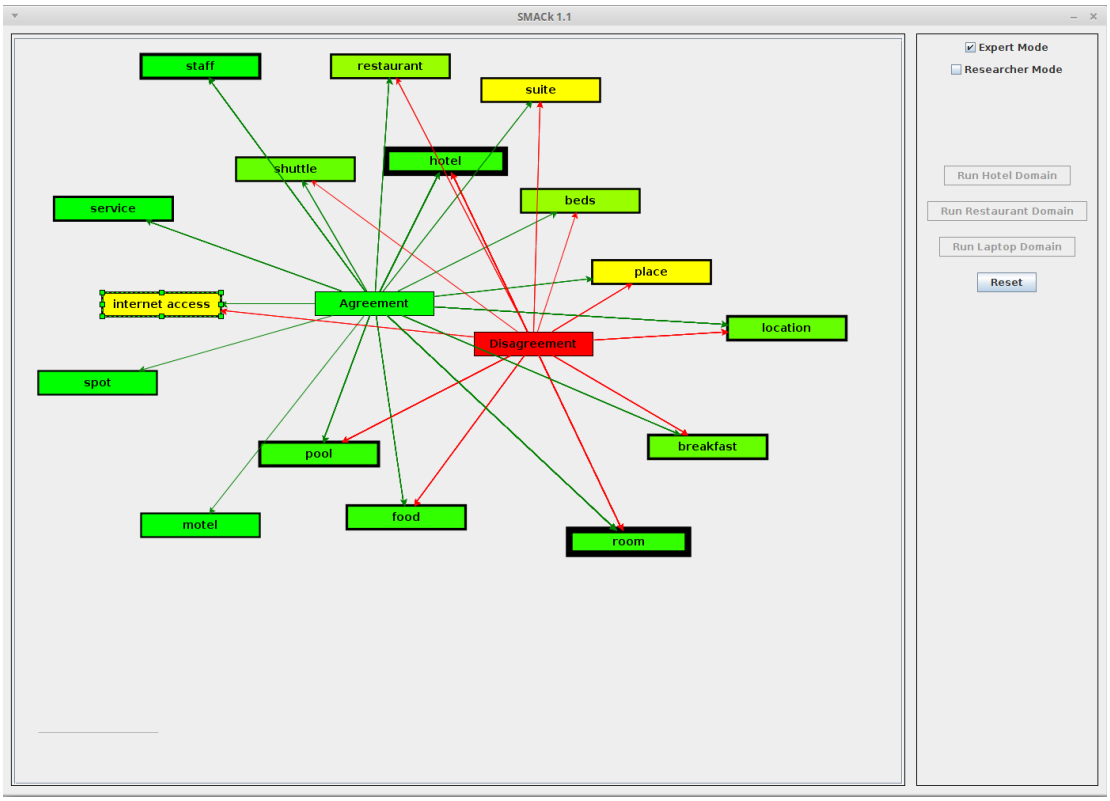


Fig. 7. The results of the analysis of a bunch of documents talking about a specific hotel.

Table 2

Results obtained concerning the computation of polarities associated with single aspects on the SemEval 2015 benchmark. For each dataset, we reported the accuracy obtained in computing polarities ("positive", or "negative"). Acronyms refer to the systems participated in the SemEval 2015 competition.

System Acronym	Acc. "Restaurant"	Acc. "Laptop"
ECNU	0.7810	0.7829
EliXa	0.7005	0.7291
lsislif	0.7550	0.7787
LT3	0.7502	0.7376
sentiue	**0.7869**	0.7934
SIEL	0.7124	-
SINAI	0.6071	0.6585
TJUdeM	0.6887	0.7323
UFRGS	0.7171	0.6733
UMDuluth	0.7112	-
V3	0.6946	0.6838
wnlp	0.7136	0.7207
SMACk	0.7794	**0.8589**

from a document and to infer the right polarity associated to each aspect.

Besides such validation tasks, the development of the SMACk platform allowed us to learn some lessons that will lead the improvement of the whole system. In particular, what we learned from this experience can be recognized in two main aspects: (i) efficient management of data streams, and (ii) understandability of the

user interface.

SMACk Efficiency The scenario used in this first prototype focused on using document sets having a limited number of items. By switching from a test environment to a more complex one, we noticed that the time needed for generating the argumentation graph increases significantly. This issue was related to the necessity of detecting, for each aspect that was already extracted, the presence of a support or attack edge. While a possible solution might be the parallelization of this task, some tricks have to be applied. Indeed, the constraint of analyzing documents by keeping the timing order in which they have been generated, requires to perform some checks based on the number of documents that we want to analyze at a certain time. Thus, by having, for example, a window of n documents that we want to parallelize, a possible strategy is to verify if there are conflicts between the aspects extracted from such documents. This way, we would be able to safely update the argumentation graph without losing potential edges.

User Interface Improvement The second lesson we learned from this work is related to which improvements should be carried out to the user interface for making the platform more appealing from the user's perspective. Users interviewed for judging the tool provided feedback that can be summarized in the following two issues.

- Contextual information into the argumentation graph: in this prototype we did not take into account the possibility of having different kinds of users: Basic and Advanced. While basic users can be satisfied from a simple graphical information supporting the detection of the most interesting aspects, advanced users wanted to see detailed information associated with each aspect, i.e. the polarity value, a summary of supports and attacks to each aspect, etc. This functionality will be included in the next version of the platform.
- Animate the evolution of single aspect: the second issue raised by the users was related to the impossibility of observing how each aspect "evolves" during the analysis of the data stream. In particular, a desiderata is the possibility of focusing on a single aspect and to observe how such aspect is attacked and/or supported through time. This feature has been considered as a valuable support for associating peaks of attacks or supports based on contextual events that cannot be tracked through the SMACk system.

The two points brought to light from users' feedbacks will be used as a starting point for improving the infrastructure of the SMACk platform. Thus it will be possible to employ such a platform in a larger scale context with the aim of increasing its technological readiness level.

9. Conclusions

In this paper, we have presented SMACk, a combined framework for argumentation-based opinion extraction from natural language documents. The goal of SMACk is to support users dealing with huge amounts of data, typically in the context of online purchases, to detect the most important aspects for them of the product to be bought, and consequently, to easily detect what are the main arguments raised in favor or against this product with respect to the specific aspect of interest. In order to address this challenging task, we combined two components, namely aspect-based opinion mining, and abstract argumentation theory. The overall framework is presented, and the evaluation results show the feasibility of the task fostering further research in this direction.

Several open challenges have to be considered for future work. First, we plan to improve the user interface of the system in order to enhance usability. The idea is to develop SMACk as a web application such that it could be plugged in different websites, e.g., for online purchase like Amazon. Second, the current version of SMACk works in an unsupervised way. On the one hand, this is a strong point of the proposed approach with respect to the state of the art; on the other hand, due to the availability of effective approaches about domain detection, we plan to exploit such approaches combined with word-embedding techniques for building linguistic models that can be used for polishing the set of extracted aspects. Finally, the version of SMACk discussed in this paper works only on English texts. The next step will focus on the integration of multi-language support.

References

[1] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*. European Language Resources Association, 2010.

[2] S. Bistarelli, F. Rossi, and F. Santini. Benchmarking hard problems in random abstract afs: The stable semantics. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 153–160. IOS Press, 2014.

[3] E. Bjørkelund, T. H. Burnett, and K. Nørvåg. A study of opinion mining and visualization of hotel reviews. In E. Pardede, editor, *The 14th International Conference on Information Integration and Web-based Applications & Services, IIWAS '12, Bali, Indonesia - December 3-5, 2012*, pages 229–238. ACM, 2012.

[4] S. Brody and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 804–812. The Association for Computational Linguistics, 2010.

[5] E. Cambria. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107, 2016.

[6] E. Cambria, R. Speer, C. Havasi, and A. Hussain. Senticnet: A publicly available semantic resource for opinion mining. In *AAAI Fall Symposium: Commonsense Knowledge*, 2010.

[7] M. Caminada. On the issue of reinstatement in argumentation. In *Procs. of JELIA*, volume 4160 of *LNCS*, pages 111–123. Springer, 2006.

[8] F. Cerutti, M. Giacomin, M. Vallati, and M. Zanella. An SCC recursive meta-algorithm for computing preferred labellings in abstract argumentation. In C. Baral, G. D. Giacomo, and T. Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.

[9] D. Chen and C. D. Manning. A fast and accurate dependency parser using neural networks. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[10] T. C. Chinsha and S. Joseph. A syntactic approach for aspect based opinion mining. In *Proceedings of the 9th IEEE International Conference on Semantic Computing, ICSC 2015, Anaheim, CA, USA, February 7-9, 2015*, pages 24–31, 2015.

[11] Y. Choi and C. Cardie. Hierarchical sequential learning for extracting opinions and their attributes. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden, Short Papers*, pages 269–274. The Association for Computer Linguistics, 2010.

[12] K. Clark and C. D. Manning. Entity-centric coreference resolution with model stacking. In *Association for Computational Linguistics (ACL)*, 2015.

[13] C. da Costa Pereira, M. Dragoni, A. G. B. Tettamanzi, and S. Villata. Fuzzy labeling for abstract argumentation: An empirical evaluation. In S. Schockaert and P. Senellart, editors, *Scalable Uncertainty Management - 10th International Conference, SUM 2016, Nice, France, September 21-23, 2016, Proceedings*, volume 9858 of *Lecture Notes in Computer Science*, pages 126–139. Springer, 2016.

[14] C. da Costa Pereira, A. Tettamanzi, and S. Villata. Changing one's mind: Erase or rewind? In T. Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 164–171. IJCAI/AAAI, 2011.

[15] L. Deng and J. Wiebe. MPQA 3.0: An entity/event-level sentiment corpus. In R. Mihalcea, J. Y. Chai, and A. Sarkar, editors, *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pages 1323–1328. The Association for Computational Linguistics, 2015.

[16] M. Dragoni, C. da Costa Pereira, A. G. B. Tettamanzi, and S. Villata. Smack: An argumentation framework for opinion mining. In S. Kambhampati, editor, *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 4242–4243. IJCAI/AAAI Press, 2016.

[17] M. Dragoni, A. Tettamanzi, and C. da Costa Pereira. Dranziera: An evaluation protocol for multi-domain opinion mining. In N. C. C. Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may 2016. European Language Resources Association (ELRA).

[18] M. Dragoni, A. G. Tettamanzi, and C. da Costa Pereira. Propagating and aggregating fuzzy polarities for concept-level sentiment analysis. *Cognitive Computation*, 7(2):186–197, 2015.

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

[20] M. Federici and M. Dragoni. Towards unsupervised approaches for aspects extraction. In M. Dragoni, D. R. Recupero, K. Denecke, Y. Deng, and T. Declerck, editors, *Joint Proceedings of the 2th Workshop on Emotions, Modality, Sentiment Analysis and the Semantic Web and the 1st International Workshop on Extraction and Processing of Rich Semantics from Medical Texts co-located with ESWC 2016, Heraklion, Greece, May 29, 2016.*, volume 1613 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[21] C. Fellbaum. *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, MA, 1998.

[22] K. Grosse, M. P. González, C. I. Chesñevar, and A. G. Maguitman. Integrating argumentation and sentiment analysis for mining opinions from twitter. *AI Commun.*, 28(3):387–401, 2015.

[23] I. Habernal, J. Eckle-Kohler, and I. Gurevych. Argumentation mining on the web from information seeking perspective. In E. Cabrio, S. Villata, and A. Z. Wyner, editors, *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing, Forlì-Cesena, Italy, July 21-25, 2014.*, volume 1341 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

[24] E. W. Hinrichs and D. Roth, editors. *Accurate Unlexicalized Parsing*, 2003.

[25] M. Hu and B. Liu. Mining and summarizing customer reviews. In W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, editors, *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pages 168–177. ACM, 2004.

[26] N. Jakob and I. Gurevych. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP 2010, 9-11 October 2010, MIT Stata Center, Massachusetts, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1035–1045. ACL, 2010.

[27] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *J. Log. Comput.*, 9(2):215–261, 1999.

[28] W. Jin, H. H. Ho, and R. K. Srihari. Opinionminer: a novel machine learning system for web opinion mining and extraction. In *KDD*, pages 1195–1204, 2009.

[29] C. M. Kristina Toutanova, Dan Klein and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *In Proceedings of HLT-NAACL 2003*, pages 252–259, 2003.

[30] P. R. Kumar and V. Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl.-Based Syst.*, 89:14–46, 2015.

[31] F. Li, M. Huang, and X. Zhu. Sentiment analysis with global topics and local dependency. In M. Fox and D. Poole, editors, *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. AAAI Press, 2010.

[32] M. Lippi and P. Torroni. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Techn.*, 16(2):10, 2016.

[33] B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In A. Ellis and T. Hagino, editors, *Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*, pages 342–351. ACM, 2005.

[34] B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In C. C. Aggarwal and C. X. Zhai, editors, *Mining Text Data*, pages 415–463. Springer, 2012.

[35] H. Liu, H. Yu, and Z. Deng. Multi-document summarization based on two-level sparse representation model. In *AAAI*, pages 196–202, 2015.

[36] K. Liu, H. L. Xu, Y. Liu, and J. Zhao. Opinion target extraction using partially-supervised word alignment model. In F. Rossi, editor, *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 2134–2140. IJCAI/AAAI, 2013.

[37] Q. Liu, Z. Gao, B. Liu, and Y. Zhang. Automated rule selection for aspect extraction in opinion mining. In Q. Yang and M. Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 1291–1297. AAAI Press, 2015.

[38] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations*, pages 55–60. The Association for Computer Linguistics, 2014.

[39] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, and P. J. Shenoy, editors, *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 171–180. ACM, 2007.

[40] M. Mitchell, J. Aguilar, T. Wilson, and B. V. Durme. Open domain targeted sentiment. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1643–1654. ACL, 2013.

[41] S. Moghaddam and M. Ester. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In W. Ma, J. Nie, R. A. Baeza-Yates, T. Chua, and W. B. Croft, editors, *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*, pages 665–674. ACM, 2011.

[42] A. Mukherjee and B. Liu. Aspect extraction through semi-supervised modeling. In *The 50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, July 8-14, 2012, Jeju Island, Korea - Volume 1: Long Papers*, pages 339–348. The Association for Computer Linguistics, 2012.

[43] A. Nenkova and K. McKeown. A survey of text summarization techniques. In C. C. Aggarwal and C. Zhai, editors, *Mining Text Data*, pages 43–76. Springer, 2012.

[44] A. Peldszus and M. Stede. From argument diagrams to argumentation mining in texts: A survey. *IJCINI*, 7(1):1–31, 2013.

[45] S. P.J, D. Dunphy, and S. Marshall. *The General Inquirer: A Computer Approach to Content Analysis*. Oxford, England: M.I.T. Press, 1966.

[46] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado, June 2015. Association for Computational Linguistics.

[47] A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 6-8 October 2005, Vancouver, British Columbia, Canada*. The Association for Computational Linguistics, 2005.

[48] A.-M. Popescu and O. Etzioni. *Extracting Product Features and Opinions from Reviews*, pages 9–28. Springer London, London, 2007.

[49] D. M. W. Powers. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[50] G. Qiu, B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, 2011.

[51] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, and D. Crystal. *A comprehensive grammar of the English language*, volume 397. Cambridge Univ Press, 1985.

[52] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[53] S. Rosenthal and K. McKeown. Detecting opinionated claims in online discussions. In *Sixth IEEE International Conference on Semantic Computing, ICSC 2012, Palermo, Italy, September 19-21, 2012*, pages 30–37. IEEE Computer Society, 2012.

[54] K. Schouten and F. Frasincar. Survey on aspect-level sentiment analysis. *IEEE Trans. Knowl. Data Eng.*, 28(3):813–830, 2016.

[55] G. Somprasertsri and P. Lalitrojwong. Extracting product features and opinions from product reviews using dependency analysis. In M. Li, Q. Liang, L. Wang, and Y. Song, editors, *Seventh International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2010, 10-12 August 2010, Yantai, Shandong, China*, pages 2358–2362. IEEE, 2010.

[56] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su. Hidden sentiment association in chinese web opinion mining. In *WWW*, pages 959–968, 2008.

[57] I. Titov and R. T. McDonald. A joint model of text and aspect ratings for sentiment summarization. In K. McKeown, J. D. Moore, S. Teufel, J. Allan, and S. Furui, editors, *ACL 2008, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, USA*, pages 308–316. The Association for Computer Linguistics, 2008.

[58] M. Vallati, F. Cerutti, and M. Giacomin. Argumentation frameworks features: an initial study. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 1117–1118. IOS Press, 2014.

[59] B. Verheij. Artificial argument assistants for defeasible argumentation. *Artif. Intell.*, 150(1-2):291–324, 2003.

[60] M. P. G. Villalba and P. Saint-Dizier. A framework to extract arguments in opinion texts. *IJCINI*, 6(3):62–87, 2012.

[61] B. Wang and H. Wang. Bootstrapping both product features and opinion words from chinese customer reviews with cross-inducing. In *Third International Joint Conference on Natural Language Processing, IJCNLP 2008, Hyderabad, India, January 7-12, 2008*, pages 289–295. The Association for Computer Linguistics, 2008.

[62] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In B. Rao, B. Krishnapuram, A. Tomkins, and Q. Yang, editors, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 783–792. ACM, 2010.

[63] Y. Wu, Q. Zhang, X. Huang, and L. Wu. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1533–1541. ACL, 2009.

[64] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[65] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain. Extracting and ranking product features in opinion documents. In C. Huang and D. Jurafsky, editors, *COLING 2010, 23rd International Conference on Computational Linguistics, Posters Volume, 23-27 August 2010, Beijing, China*, pages 1462–1470. Chinese Information Processing Society of China, 2010.

[66] M. Zhang, Y. Zhang, and D. Vo. Neural networks for open domain targeted sentiment. In L. Màrquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 612–621. The Association for Computational Linguistics, 2015.

[67] X. Zhou, X. Wan, and J. Xiao. Collective opinion target extraction in chinese microblogs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1840–1850. ACL, 2013.

[68] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu. Multi-aspect opinion polling from textual reviews. In D. W. Cheung, I. Song, W. W. Chu, X. Hu, and J. J. Lin, editors, *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009*, pages 1799–1802. ACM, 2009.

[69] L. Zhuang, F. Jing, and X. Zhu. Movie review mining and summarization. In P. S. Yu, V. J. Tsotras, E. A. Fox, and B. Liu, editors, *Proceedings of the 2006 ACM CIKM International Conference on Information and Knowledge Management, Arlington, Virginia, USA, November 6-11, 2006*, pages 43–50. ACM, 2006.