Unsupervised Detection of Argumentative Units though Topic Modeling Techniques

Alfio Ferrara and **Stefano Montanelli**
Dipartimento di Informatica, Università degli Studi di Milano
Via Comelico 39, 20135 - Milano, Italy
{alfio.ferrara,stefano.montanelli}@unimi.it

Georgios Petasis
Institute of Informatics and Telecommunications,
National Centre for Scientific Research (N.C.S.R.) "Demokritos"
P.O. BOX 60228, Aghia Paraskevi, GR-153 10, Athens, Greece
petasis@iit.demokritos.gr

Abstract

In this paper we present a new *unsupervised* approach, "Attraction to Topics" – $\mathcal{A}2\mathcal{T}$, for the detection of argumentative units, a sub-task of argument mining. Motivated by the importance of topic identification in manual annotation, we examine whether topic modeling can be used for performing unsupervised detection of argumentative sentences, and to what extend topic modeling can be used to classify sentences as claims and premises. Preliminary evaluation results suggest that topic information can be successfully used for the detection of argumentative sentences, at least for corpora used in the evaluation. Our approach has been evaluated on two English corpora, the first of which contains 90 persuasive essays, while the second is a collection of 340 documents from user generated content.

1 Introduction

Argument mining involves the automatic discovery of *argument components* (i.e. claims, premises) and the *argumentative relations* (i.e. supports, attacks) among these components in texts. Primarily aiming to extract arguments from texts in order to provide structured data for computational models of argument and reasoning engines (Lippi and Torroni, 2015a), argument mining has additionally the potential to support applications in various research fields, such as opinion mining (Goudas et al., 2015), stance detection (Hasan and Ng, 2014), policy modelling (Florou et al., 2013; Goudas et al., 2014), legal information systems (Palau and Moens, 2009), etc.

Argument mining is usually addressed as a pipeline of several sub-tasks. Typically the first sub-task is the separation between argumentative and non-argumentative text units, which can be performed at various granularity levels, from clauses to several sentences, usually depending on corpora characteristics. Detection of argumentative units (AU)[1], as discussed in Section 2, is typically modeled as a fully-supervised classification task, either a binary one, where units are separated in argumentative and non-argumentative ones with argumentative ones to be subsequently classified in claims and premises as a second step, or as a multi-class one, where identification of argumentative units and classification into claims and premises are performed as a single step. According to a recent survey (Lippi and Torroni, 2015a), the performance of proposed approaches depends on highly engineered and sophisticated, manually constructed, features.

However, fully-supervised approaches rely on manually annotated datasets, the construction of which is a laborious, costly, and error-prone process, requiring significant effort from human experts. At the same time, reliance on sophisticated features may hinder the generalisation of an approach to new corpora types and domains (Lippi and Torroni, 2015a). The removal of manual supervision through exploitation of *unsupervised approaches* is a possible solution to both of the aforementioned problems.

1.1 Motivations of our work

Topics seem to be related to the task of argument mining, at least for some types of corpora, as topic

[1] Also known as "Argumentative Discourse Units – ADUs" (Peldszus and Stede, 2013).

identification frequently appears as a step in the process of manual annotation of arguments in texts (Stab and Gurevych, 2014a). However, despite its apparent importance in manual annotation, only a small number of studies have examined the inclusion of topic information in sub-tasks of argument mining. Habernal and Gurevych (2015) have included sentiment and topic information as features for classifying sentences as claims, premises, backing and non-argumentative units. A less direct exploitation of topic information has been presented in (Nguyen and Litman, 2015), where topics have been used to extract lexicons of argument and domain words, which can provide evidence regarding the existence of argument components.

In this paper we propose "Attraction to Topics" – $\mathcal{A}2\mathcal{T}$, an unsupervised approach based on topic modeling techniques for detecting argumentative discourse units at sentence-level granularity (a sub-task known as "argumentative sentence detection"). The goals of $\mathcal{A}2\mathcal{T}$ are twofold. On the one side, $\mathcal{A}2\mathcal{T}$ enforces identification of sentences that contain argument components, by also distinguishing them from the non-argumentative sentences that do not contain argument components. On the other side, $\mathcal{A}2\mathcal{T}$ classifies the discovered argumentative sentences according to their role, as *major claims*, *claims*, and *premises*.

The rest of the paper is organized as follows: Section 2 presents an overview of approaches related to argument mining focusing on the detection of argumentative units, while Section 3 presents our approach on applying topic modeling for identifying sentences that contain argument components. Section 4 presents our experimental setting and evaluation results, with Section 5 concluding this paper and proposing some directions for further research.

2 Related work

Almost all argument mining frameworks proposed so far employ a pipeline of stages, each of which is addressing a sub-task of the argument mining problem (Lippi and Torroni, 2015a). The segmentation of text into argumentative units is typically the first sub-task encountered in such an argument mining pipeline, aiming to segment texts into argumentative and non-argumentative text units (i.e. segments that do contain or do not contain argument components, such as claims or premises). The granularity of argument components is text-dependant. For example, in Wikipedia articles studied in (Rinott et al., 2015), argument components spanned from less than a sentence to more than a paragraph, although 90% of the cases was up to 3 sentences, with 95% of components being comprised of whole sentences.

Several approaches address the identification of argumentative units at the sentence level, a sub-task known as "argumentative sentence detection", which typically models the task as a binary classification problem. Employing machine learning and a set of features representing sentences, the goal is to discard sentences that are not part (or do not contain a component) of an argument. As reported also by Lippi and Torroni (2015a), the vast majority of existing approaches employ "classic, off-the-self" classifiers, while most of the effort is devoted to highly engineered features. A plethora of learning algorithms have been applied on the task, including Naive Bayes (Moens et al., 2007; Park and Cardie, 2014), Support Vector Machines (SVM) (Mochales and Moens, 2011; Rooney et al., 2012; Park and Cardie, 2014; Stab and Gurevych, 2014b; Lippi and Torroni, 2015b), Maximum Entropy (Mochales and Moens, 2011), Logistic Regression (Goudas et al., 2014, 2015; Levy et al., 2014), Decision Trees and Random Forests (Goudas et al., 2014, 2015; Stab and Gurevych, 2014b).

However, approaches addressing this task in a semi-supervised or unsupervised manner are still scarce. In (Petasis and Karkaletsis, 2016) an unsupervised approach is presented, which addresses the sub-task of identifying the main claim in a document by exploiting evidence from an extractive summarization algorithm, TextRank (Mihalcea and Tarau, 2004). In an attempt to study the overlap between graph-based approaches and approaches targeting extractive summarization with argument mining, evaluation results suggest a positive effect on the sub-task, achieving an accuracy of 50% on the corpus compiled by Hasan and Ng (2014) from online debate forums and on a corpus of persuasive essays (Stab and Gurevych, 2014a). Regarding semi-supervised approaches, Habernal and Gurevych (2015) propose new unsupervised features that exploit clustering of unlabeled argumentative data from debate portals based on word embeddings, outperforming several baselines. This work employs also topic modeling as one of its features, by including as features the

distributions of sentences from LDA (Blei et al., 2003).

Topic modeling has been mainly exploited for identification of argumentative relations and for extraction of argument and domain lexicons. In Lawrence et al. (2014), LDA is used to decide whether a proposition can be attached to its previous proposition in order to identify non directional relations among propositions detected through classifiers based on words and part-of-speech tags. LDA has been also used to mine lexicons of argument (words that are topic independent) and domain words (Nguyen and Litman, 2015), by post-processing document topics generated by LDA. These lexicons have been used as features for supervised approaches for argument mining (Nguyen and Litman, 2016a,b). However, to the best of our knowledge, no prior approach has applied topic modeling to argumentative sentence detection in an unsupervised setting, which is the featuring aspect of the proposed $\mathcal{A}2\mathcal{T}$ approach presented in the following.

3 Topic modeling for argument mining

Given a document corpus, topic modeling techniques can be employed to discover the most representative topics throughout the corpus, and to provide an assignment of documents to topics, meaning that the higher is the assignment value of a document to a certain topic, the higher is the probability that the document is "focused" on that topic.

The idea of $\mathcal{A}2\mathcal{T}$ is that an argumentative unit is a sentence highly focused on a specific topic, namely a sentence with high assignment value to a certain topic and low assignment value to the other topics. To this end, $\mathcal{A}2\mathcal{T}$ introduces the notion of *attraction* with the aim at recognizing the sentences highly focused on specific topics, that represent the recognized argumentative units. In the following, the $\mathcal{A}2\mathcal{T}$ approach and related techniques are described in detail.

3.1 $\mathcal{A}2\mathcal{T}$ approach

The schema of the $\mathcal{A}2\mathcal{T}$ approach is shown in Figure 1. Consider a corpus of texts $\mathcal{C} = \{c_1, \ldots, c_n\}$, where a text $c_i \in \mathcal{C}$ is a sequence of sentences, like for example an essay, a web page/post, or a scientific paper. The ultimate goal of the $\mathcal{A}2\mathcal{T}$ approach is to derive a set of argumentative units $\mathcal{U} = \{\langle s_1, c, l\rangle, \ldots, \langle s_h, c, l\rangle\}$, where

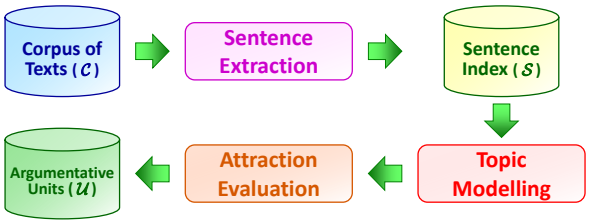

Figure 1: Schema of the $\mathcal{A}2\mathcal{T}$ approach

s_i is a sentence containing an argumentative unit, c is the text containing s, and l is the argumentative role expressed by the unit (e.g., major claim, claim, premise). The $\mathcal{A}2\mathcal{T}$ approach is articulated in the following activities:

Sentence extraction. $\mathcal{A}2\mathcal{T}$ approach is characterized by the use of topic modeling at sentence-level granularity. For this reason, a pre-processing step of the corpus $\mathcal{C}$ is enforced based on conventional techniques for sentence tokenization, words tokenization, normalization, and indexing (Manning et al., 2008). The result is a sentence set $\mathcal{S} = \{\langle \overrightarrow{s_1}, c, pos_1\rangle, \ldots, \langle \overrightarrow{s_m}, c, pos_m\rangle\}$, where $\overrightarrow{s_i}$ is the vector representation of the sentence s_i and c, pos are text and position in the text where the sentence appears, respectively. The sentence set is stored in a sentence index for efficient access of $\mathcal{S}$ elements.

Topic modeling. The set of extracted sentences $\mathcal{S}$ is used as the document corpus on which topic modeling is applied. The result of this activity is twofold. First, topic modeling returns a set of topics $\mathcal{T} = \{t_0, \ldots, t_k\}$ representing the latent variables that are most representative for the sentences $\mathcal{S}$. Second, topic modeling returns a distribution of sentences over topics $\theta = \{\theta_{s_1}, \ldots, \theta_{s_m}\}$. In particular, $\theta_{s_i} = [p(t_0|s_i), \ldots, p(t_k|s_i)]$ is the probability distribution of the sentence s_i over the set of topics $\mathcal{T}$, where $p(t_j|s_i)$ represents the probability of the topic t_j given the sentence s_i (i.e., the so-called assignment value of s_i to t_j).

Attraction evaluation. The notion of attraction is introduced to measure the *degree of focus* that characterizes sentences with respect to the emerged topics. To this end, the distribution of sentences over topics θ is exploited with the aim at determining the best topic assignment for each sentence of $\mathcal{S}$. The result is an attraction set $\mathcal{A} = \{\langle s_1, a_1\rangle, \ldots, \langle s_m, a_m\rangle\}$ where s_i is a sentence of $\mathcal{S}$ and a_i is its corresponding attraction

value.

Sentence labeling. By exploiting the attraction set $\mathcal{A}$, labeling has the goal to determine the sentences of $\mathcal{S}$ that are more focused on a specific topic, according to the hypothesis that those sentences are the argumentative units. In a basic scenario, labeling consists in distinguishing between sentences that are argumentative units ($l = au$) and sentences that are not argumentative units ($l = \overline{au}$). In a more articulated scenario, labeling consists in assigning a role to sentences that are recognized as argumentative units. For instance, it is possible to distinguish argumentative-unit sentences that are claims ($l = cl$), major claims ($l = mc$), or premises ($l = pr$). A sentence s recognized as argumentative unit is inserted in the final set $\mathcal{U}$ with the assigned label and it is returned as a result of $\mathcal{A}2\mathcal{T}$.

3.2 $\mathcal{A}2\mathcal{T}$ techniques

In $\mathcal{A}2\mathcal{T}$, the *sentence extraction* step is enforced by relying on standard techniques for representing documents in terms of feature vectors and bag of words (using *tf-idf* as weighting scheme) (Castano et al., 2017). Probabilistic topic modeling is exploited to enforce the subsequent *topic modeling* step. Probabilistic topic models are a suite of algorithms whose aim is to discover the hidden thematic structure in large archives of documents, namely *sentences* in $\mathcal{A}2\mathcal{T}$. The idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al., 2003). Probabilistic topic modeling algorithms infer the distribution θ of documents over topics and the distribution ϕ of words over topics, by sampling from the bag of words of each document. In our approach, we choose to exploit the Hierarchical Dirichlet Process (HDP). With respect to other algorithms (such as LDA), HDP has the advantage to provide the optimal number of topics instead of requiring to set such a number as input (Teh et al., 2006).

Attraction evaluation. The notion of *attraction* is introduced in $\mathcal{A}2\mathcal{T}$ to capture the intuition that argumentative units are related to the distribution of sentences over topics. Consider a set of sentences $\mathcal{S}$ and the distribution θ of sentences over the set of topics $\mathcal{T}$. The more the distribution θ_{s_i} of a sentence s_i over the topics is unequal, the more s_i is *focused* on a topic, thus suggesting s_i as a possible argumentative unit. A further feature that attraction aims to capture is that argumentative units often appear either at the beginning or at the end of texts. The attraction a_i of a sentence s_i is calculated as follows:

$$a_i = K\varphi_{s_i} + (1 - K)\frac{\rho_{s_i}}{\sum_{s_j \in c} \rho_{s_j}},$$

$\varphi_{s_i} = \max(\theta_{s_i})$ is a measure of how much s_i is focused on a topic and $\rho_{s_i} = \alpha f(pos_i)^2 + \beta f(pos_i) + \gamma$ is a parabolic function over the position of the sentence in c. In particular, given $L(c)$ as the number of sentences in c, $f(pos_i) = \left|\frac{L(c)}{2} - pos_i\right|$ such that $f(pos_i)$ is higher when s_i appears either at the beginning or at the end of c. The parameters α, β, γ determine the shape of ρ_{s_i}. $K \in [0, 1]$ is a constant value used to balance the role of focus and position in calculating the attraction. The attraction a_i can be interpreted as the probability of a sentence s_i to contain an argumentative unit. According to this interpretation, given s_i, also the contiguous sentences s_{i-1} and s_{i+1} have a chance to be argumentative units. As a result, given the calculated attraction set $\mathcal{A}$, we update the attraction values a_i through an interpolation mechanism based on the Savitzky-Golay smoothing filter (SGF) (Savitzky and Golay, 1964), so that $\mathcal{A} := SGF(\mathcal{A})$.

In Figure 2, an example of attraction evaluation is provided by showing the values of φ, ρ, attraction, and interpolated attraction for all the sentences within one considered student essays included in the corpus from (Stab and Gurevych, 2014a) (see Section 4).

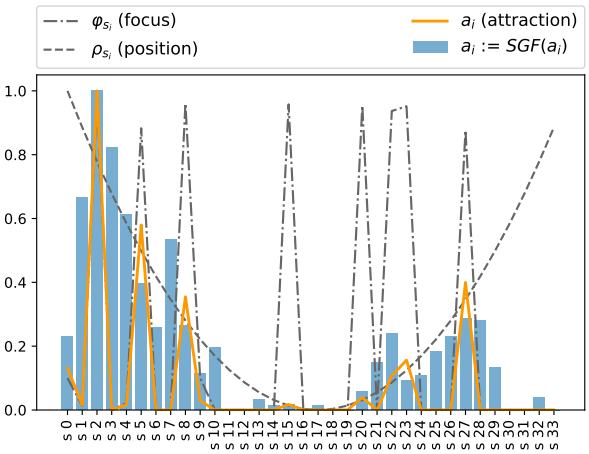

Figure 2: Attraction evaluation for the sentences of a considered text

Sentence labeling. Sentence labeling has the goal to turn attraction values into labeled categories. Consider a set of possible labels $\mathcal{L} = \{l_1, \ldots, l_g\}$, each one denoting a possible argumentative role that can be assigned to a sentence. Given a set of attraction values $\mathcal{A}$, a threshold-based mechanism is enforced to assign labels to sentences according to the following scheme:

$$\begin{array}{ccc} a_i < \tau_1 & : & s_i \leftarrow l_1 \\ \tau_1 \leq a_i < \tau_2 & : & s_i \leftarrow l_2 \\ \ldots & \ldots & \ldots \\ a_i \geq \tau_{g-1} & : & s_i \leftarrow l_g \end{array}$$

where $\tau_1 < \tau_2 < ... < \tau_{g-1}$ ($\tau_1, \ldots \tau_{g-1} \in (0, 1]$) are prefixed threshold values. The result of sentence labeling is a partition of $\mathcal{S}$ into g categories with associated labels.

In the experiments, we discuss two different strategies for sentence labeling. The first one is a *two-class labeling* strategy where the possible labels for a sentence are argumentative unit (au) and non-argumentative unit ($\overline{au}$. The second strategy is a *multi-class labeling* in which the possible labels of a sentence are non-argumentative unit $\overline{au}$, premise (pr), claim (cl), and major claim (mc).

4 Experimental results

For evaluation of the proposed $\mathcal{A}2\mathcal{T}$ approach, we have used two English corpora. The first corpus (C1 in the following) is a collection of 90 student persuasive essays (Stab and Gurevych, 2014a) which has been manually annotated with major claims (one per essay), claims and premises at the clause level. In addition, the corpus contains manual annotations of argumentative relations, where the claims and premises are linked, while claims are linked to the major claim either with a support or an attack relation. Inter-annotation agreement has been measured to unitized alpha (Krippendorff, 2004) $\alpha_U = 0.724$. These 90 essays consist of a total of $1,675$ sentences (from which 19.3% contain no argument components), with an average length of 18.61 ± 7 sentences per essay, while the 5.4% of sentences contain a major claim, 26.4% contain a claim, and 61.1% contain a premise.

The second corpus (C2 in the following) has been compiled and manually annotated as described in (Habernal and Gurevych, 2017). This corpus focuses on user generated content, including user comments, forum posts, blogs, and newspaper articles, covering several thematic domains from educational controversies, such as home-schooling, private vs. public schools, or single-sex education. Containing in total 340 documents, the corpus has been manually annotated with an argument scheme based on extended Toulmin's model, involving claims, premises, and backing, rebuttal, refutation argument units. The corpus contains documents of various sizes, with a mean size of 11.44 ± 11.70 sentences per document, while the inter-annotator agreement was measured as $\alpha_U = 0.48$. The corpus consists of 3,899 sentences, from which 2,214 sentences (57%) contain no argument components.

Both corpora have been preprocessed with NLTK (Loper and Bird, 2002) in order to identify tokens and sentences. Then, each sentence was annotated as argumentative or non-argumentative, depending on whether it contained an argument unit (i.e. a text fragment annotated as major claim, claim, or premise). In addition, each argumentative sentence was further annotated with one of *major claim*, *claim*, and *premise*, based on the type of the contained argumentative unit. For the second corpus, which utilizes a richer argument scheme, we have considered backing, rebuttal and refutation units as premises. This second corpus does not contain units annotated as major claims. The following three tasks have been executed:

- Task 1: Argumentative sentence identification – given a sentence, classify whether or not it contains an argument component.
- Task 2: Major claim identification – given a argumentative sentence, classify whether or not it contains a major claim.
- Task 3: Argumentative sentence classification – given a sentence, classify the sentence as *major claim*, *claim*, *premise*, or *non-argumentative*.

Baseline. As a baseline for comparison against our approach, we created a probabilistic classifier of sentences which evaluates the probability $p(l = au|s_i)$ as follows. Given the text c containing $L(c)$ sentences s_i, let be $\zeta_c \sim Dir(\boldsymbol{\alpha})$ the probability distribution of the sentences in c, such that $\zeta_c^{s_i} \sim p(l = au|s_i)$. The $L(c)$ parameters $\boldsymbol{\alpha}$ used to generate ζ_c are defined such that $\alpha_i = \left|\frac{L(c)}{2} - pos_i\right|$. The rationale of this procedure is to bias the random assignment of a sentence to the au label in favor of sentences appearing either in the beginning or in the end of a text. This bias attempts to model empirical evi-

dence that in several types of documents, the density of argumentative units in various sections of documents depends on the structure of documents. The beginning and end of a document are expected to contain argumentative units in structured documents like news, scientific publications, or argumentative essays (Stab and Gurevych, 2017), where major claims and supporting premises are frequently found in the beginning of documents, with documents frequently ending with repeating the major claims and supporting evidence.

4.1 Task 1: Argumentative sentence identification

The goal of Task 1 is to associate each sentence of the corpora to a label in $\mathcal{L} = \{au, \overline{au}\}$ by following a two-class labeling strategy (see Section 3). As a first experiment, we performed sentence labeling with different threshold ranging from 0 to 1 with step 0.05. In Figure 3, we report the precision, recall, and F1-measure for $\mathcal{A}2\mathcal{T}$ and for the baseline. In addition, we report also the results of applying sentence labeling based on φ and ρ (the components of attraction) separately. The parameter K for attraction calculation has been set to 0.5. Since $\mathcal{A}2\mathcal{T}$ is an unsupervised method, there is no easy way to define the threshold parameter τ, which has been empirically defined to $\tau = 0.3$. The different behavior of $\mathcal{A}2\mathcal{T}$ with respect to the baseline is shown in the confusion matrices reported in Figures 4 and 5.

From Figure 3, we can see that $\mathcal{A}2\mathcal{T}$ is significantly better than the baseline, especially for the C1 corpus. A characteristic of this corpus is that argumentative units are frequently located in the introduction or the conclusion of an essay, which is also reflected by the baseline that achieved an F1-measure of 0.35 for a threshold of $\tau = 0.05$ (with the baseline being particularly precise, suggesting that argumentative units are very frequently at the beginning and end of essays). Both components of attraction (φ and ρ) perform well, with the topic component φ being slightly better than position information ρ, both in precision and recall. The results are similar for corpus C2, with $\mathcal{A}2\mathcal{T}$ surpassing the baseline, although $\mathcal{A}2\mathcal{T}$ advantage in precision is smaller. As shown in the confusion matrix of Figure 5, the main source of error is the large number of false positives for the au class, proposing more argumentative units than what have been manually identified in corpus C2. This can be attributed to the sparseness of argumentative units in the C2 corpus, with almost 60% of the sentences being non-argumentative.

4.2 Task 2: Major claim identification

As a second experiment, we exploited probabilities associated with sentences to perform a ranked evaluation. In particular, we calculated two measures, namely P that is the area the under the precision-recall curve and R that is the area under the receiver operating characteristic (ROC) curve. In this experiments, we used different criteria for defining the true labels: in PCM, an annotated sentence in the corpus is considered a true argumentative unit if it is either a premise, a claim, or a major claim; in CM only claims and major claims are taken as valid au; in M only major claims are taken into account. Results are reported in Table 1.

Table 1: Area under the precision-recall (P) and the ROC (R) curves

	C1			C2	
P	PCM	CM	M	PCM	CM
$\mathcal{A}2\mathcal{T}$	0.79	0.31	0.08	0.26	0.19
φ	0.84	0.29	0.06	0.19	0.1
ρ	0.68	0.29	0.09	0.24	0.19
Baseline	0.68	0.31	0.11	0.16	0.06
R	PCM	CM	M	PCM	CM
$\mathcal{A}2\mathcal{T}$	0.4	0.52	0.62	0.7	0.76
φ	0.52	0.51	0.53	0.58	0.57
ρ	0.16	0.52	0.77	0.69	0.77
Baseline	0.16	0.53	0.79	0.31	0.18

4.3 Task 3: Argumentative sentence classification

The goal of Task 3 is to associate each sentence of the corpora to a label in $\mathcal{L} = \{\overline{au}, pr, cl, mc\}$ by following a multi-class labeling strategy (see Section 3). In particular, we adopted the thresholds $[0.1, 0.3, 0.5]$. This task is challenging since it is required to distinguish the different role played in argumentation by sentences that are often very similar from the terminological point of view. The confusion matrix for corpus C1 is shown in Figure 6, while Figure 7 shows the confusion matrix for corpus C2. Both $\mathcal{A}2\mathcal{T}$ and the baseline achieve low results, but the accuracy of $\mathcal{A}2\mathcal{T}$ is 0.3 against the 0.1 of the baseline. From Figure 6 we see that $\mathcal{A}2\mathcal{T}$ achieved good results for premises, and quite good results for claims, although distinguishing between claims and premises is challenging for the $\mathcal{A}2\mathcal{T}$ approach. In particular, the role

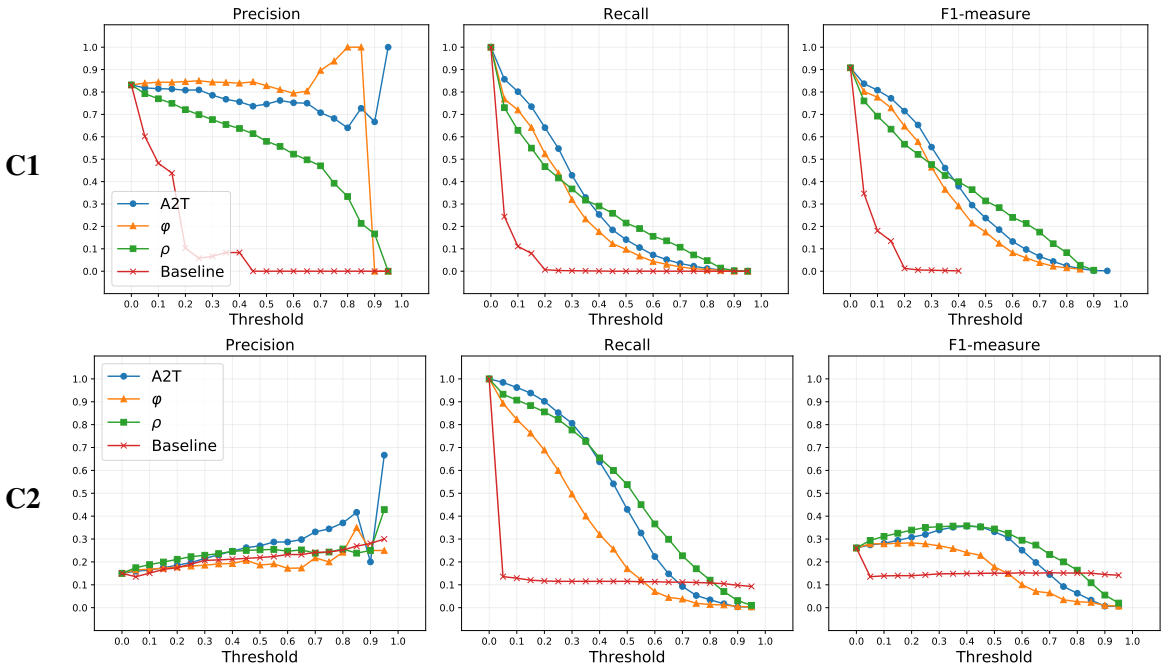

Figure 3: Precision, Recall and F1-measure with different thresholds

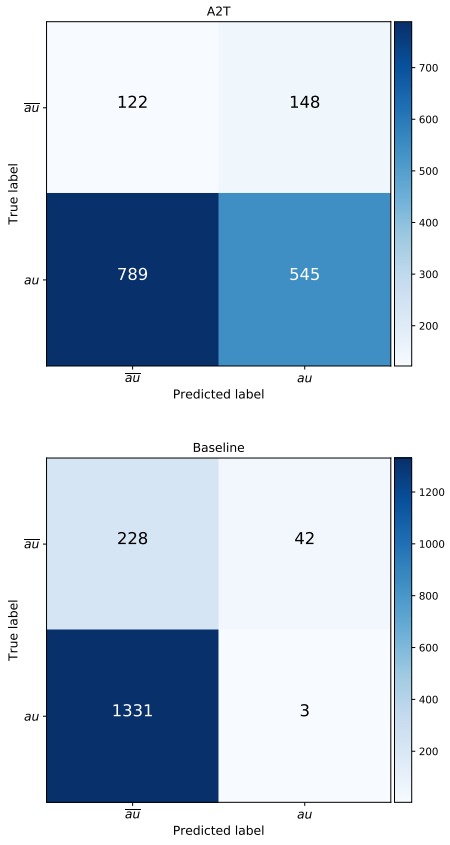

Figure 4: Two-class confusion matrices for corpus C1 (Threshold $\tau = 0.3$)

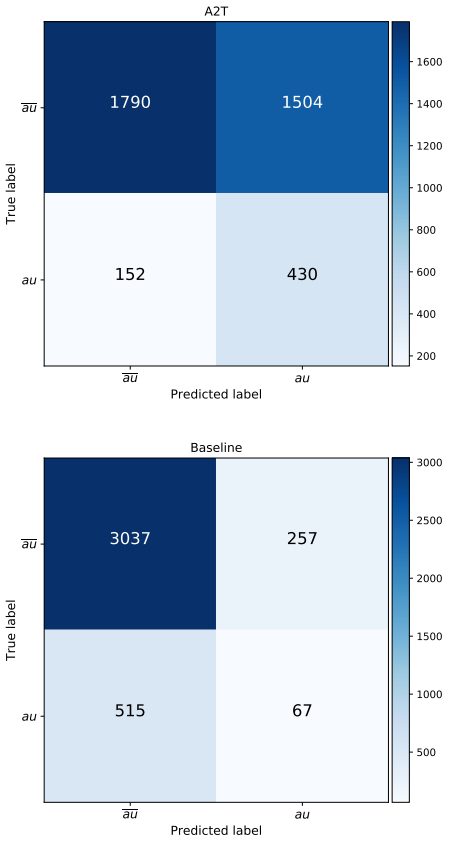

Figure 5: Two-class confusion matrices for corpus C2 (Threshold $\tau = 0.3$)

of sentences may change in different texts so that claims in one context are premises in another. This kind of contextual shift is only partially addressed by $\mathcal{A}2\mathcal{T}$, because the only contextual information we take into account is topic distribution. To the end of improving the understanding of the context,

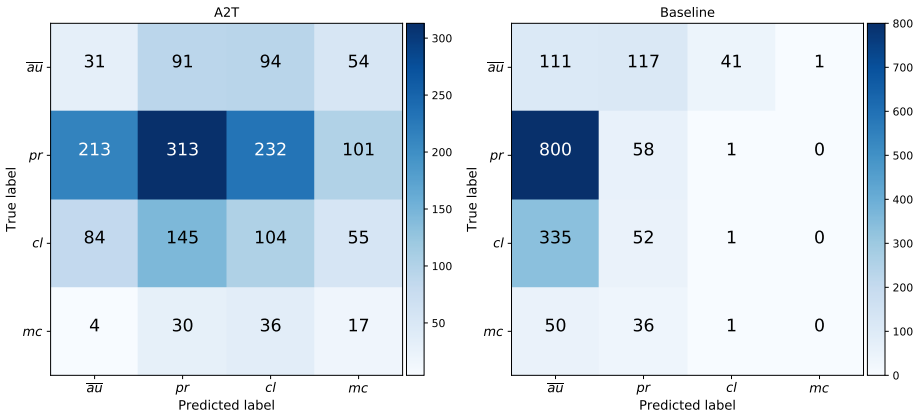

Figure 6: Multi-class confusion matrices for corpus C1

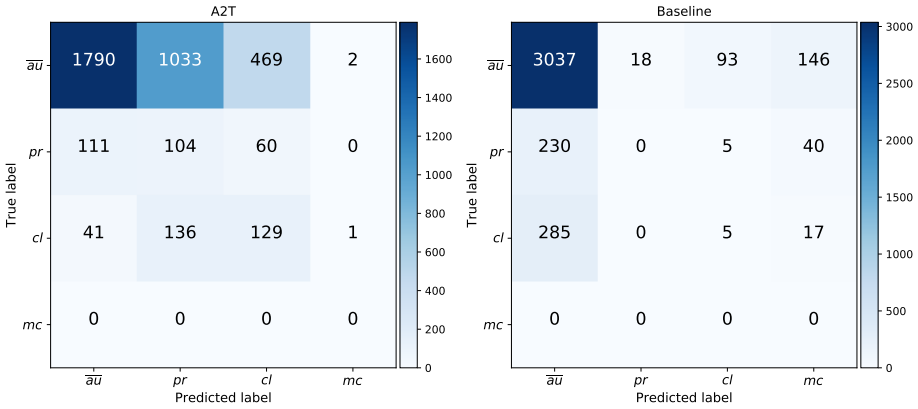

Figure 7: Multi-class confusion matrices for corpus C2

it may be useful to work also on semantic relations holding among sentences. This is actually one of the future tasks in our research work.

Another specific challenge emerges when we consider the corpus C2. Indeed, C2 contains a limited number of argumentative sentences with respect to the corpus size. In this case, since we analyze all the sentences according to their bag of words, we tend to overestimate the number of argumentative units, collecting a relatively high number of false positives.

4.4 Lessons learned from error analysis

A first evidence emerging from the analysis of confusion matrices for both corpora C1 and C2 is that the role of sentences is strictly dependent on the type of documents. C1 contains structured essays of various topics, while C2 provides conversational texts extracted from blogs and chats. In the first case, the number of argumentative units is higher than in the second one. In particular, for C2 we overestimated the probability of sentences to be an argumentative unit. This is mainly due to the fact that those sentences contain words that are semantically related to the main topic of the conversation although they are not playing a role in the argumentation. An example is the following sentence, taken from a document associated with the topic "school": "*why do some parents not think their kids can attain?*". The sentence is clearly part of a conversation and it has been annotated as a non argumentative unit because it is a question. However, since it contains words that are relevant for the topic (i.e., parents, kids, attain), $\mathcal{A}2\mathcal{T}$ associates the sentence with a good level of attraction, labeling it as a premise. In order to address this kind of false positives, we aim in our future work to study the dependency relations among sentences in text (such as question-answers) to the goal of achieving a better insight of the sentences role.

A second lesson learned from error analysis concerns the distinction between claims and premises. This confusion is evident especially when dealing with corpus C1. An example is given by the following two sentences, taken from

an essay about the role of sports in favor of peace.

- (s1) *for example, when Irak was hardly struck by the second gulf war, its citizens tried to catch any incoming news about the football-world cup through their portable receivers.*

- (s2) *thus, world sports events strongly participate in eventually pulling back people towards friendship and peace*

The sentence (s1) has been annotated as a premise, while (s2) as a claim. In our classification, they are both claims. The reason is that they both contain topic-related words and their position in text is similar. The main distinction is the presence of the expression "for example" in the first sentence which qualifies it as a premise. To this end, in our future work we aim at adding some special words (such as "for example", "therefore") in the background knowledge of the classifier, in order to improve the capability of discriminating premises and claims.

5 Concluding remarks

In this paper, we present the "Attraction to Topics" – $\mathcal{A}2\mathcal{T}$ unsupervised approach for detecting argumentative discourse units, at sentence-level granularity. Motivated by the observation that topic information is frequently employed as a sub-task in the process of manual annotation of arguments, we propose an approach that exploits topic modeling techniques in order to identify argumentative units. Since manual supervision is not required, $\mathcal{A}2\mathcal{T}$ has the potential to be applicable on documents of various genres and domains. Preliminary evaluation results on two different corpora are promising. First, $\mathcal{A}2\mathcal{T}$ performs significantly better than the baseline on argumentative sentence detection on both corpora. Second, $\mathcal{A}2\mathcal{T}$ exhibits good results for classifying argumentative sentences as major claims, claims, premises, and non-argumentative units, at least for the first corpus, which has a low rate of non-argumentative sentences (20%).

Regarding directions for further research, there are several axes that can be explored. Evaluation on a larger set of annotation corpora will provide enhanced insights about the performance of the proposed approach on different document types. Our preliminary results showed that despite good recall on multiple corpora, achieving also good precision can be a challenging task in documents where argumentative units are sparse, and false positives can be an issue. In this context, we would like to also exploit other types of relations, and extend our method with other kinds of similarities over sentences.

References

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022. http://dl.acm.org/citation.cfm?id=944919.944937.

Silvana Castano, Alfio Ferrara, and Stefano Montanelli. 2017. Exploratory analysis of textual data streams. *Future Generation Computer Systems* 68:391–406.

Eirini Florou, Stasinos Konstantopoulos, Antonis Koukourikos, and Pythagoras Karampiperis. 2013. Argument extraction for supporting public policy formulation. In Piroska Lendvai and Kalliopi Zervanou, editors, *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, LaTeCH@ACL 2013, August 8, 2013, Sofia, Bulgaria*. The Association for Computer Linguistics, pages 49–54. http://aclweb.org/anthology/W/W13/W13-2707.pdf.

Theodosis Goudas, Christos Louizos, Georgios Petasis, and Vangelis Karkaletsis. 2014. Argument extraction from news, blogs, and social media. In Aristidis Likas, Konstantinos Blekas, and Dimitris Kalles, editors, *Artificial Intelligence: Methods and Applications: 8th Hellenic Conference on AI, SETN 2014, Ioannina, Greece, May 15-17, 2014. Proceedings*, Springer International Publishing, Cham, pages 287–299. https://doi.org/10.1007/978-3-319-07064-3_23.

Theodosis Goudas, Christos Louizos, Georgios Petasis, and Vangelis Karkaletsis. 2015. Argument extraction from news, blogs, and the social web. *International Journal on Artificial Intelligence Tools* 24(05):1540024. https://doi.org/10.1142/S0218213015400242.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 2127–2137. http://aclweb.org/anthology/D15-1255.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics* 43(1):125–179. https://doi.org/10.1162/COLI_a_00276.

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 751–762. http://www.aclweb.org/anthology/D14-1083.

Klaus Krippendorff. 2004. Measuring the reliability of qualitative text analysis data. *Quality and Quantity* 38(6):787–800. https://doi.org/10.1007/s11135-004-8107-7.

John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft. 2014. Mining arguments from 19th century philosophical texts using topic based modelling. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, Baltimore, Maryland, pages 79–87. http://www.aclweb.org/anthology/W/W14/W14-2111.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In Jan Hajic and Junichi Tsujii, editors, *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*. ACL, pages 1489–1500. http://aclweb.org/anthology/C/C14/C14-1141.pdf.

Marco Lippi and Paolo Torroni. 2015a. Argument mining: A machine learning perspective. In Elizabeth Black, Sanjay Modgil, and Nir Oren, editors, *Theory and Applications of Formal Argumentation: Third International Workshop, TAFA 2015, Buenos Aires, Argentina, July 25-26, 2015, Revised Selected Papers*. Springer International Publishing, Cham, pages 163–176. https://doi.org/10.1007/978-3-319-28460-6_10.

Marco Lippi and Paolo Torroni. 2015b. Context-independent claim detection for argument mining. In *Proceedings of the 24th International Conference on Artificial Intelligence*. AAAI Press, IJCAI'15, pages 185–191. http://dl.acm.org/citation.cfm?id=2832249.2832275.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*. Association for Computational Linguistics, Stroudsburg, PA, USA, ETMTNLP '02, pages 63–70. https://doi.org/10.3115/1118108.1118117.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*. Association for Computational Linguistics, Barcelona, Spain, pages 404–411. http://www.aclweb.org/anthology/W/W04/W04-3252.pdf.

Raquel Mochales and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law* 19(1):1–22. https://doi.org/10.1007/s10506-010-9104-x.

Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*. ACM, New York, NY, USA, ICAIL '07, pages 225–230. https://doi.org/10.1145/1276318.1276362.

Huy Nguyen and Diane J. Litman. 2015. Extracting argument and domain words for identifying argument components in texts. In *Proceedings of the 2nd Workshop on Argumentation Mining, ArgMining@HLT-NAACL 2015, June 4, 2015, Denver, Colorado, USA*. The Association for Computational Linguistics, pages 22–28. http://aclweb.org/anthology/W/W15/W15-0503.pdf.

Huy Nguyen and Diane J. Litman. 2016a. Context-aware argumentative relation mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics. http://aclweb.org/anthology/P/P16/P16-1107.pdf.

Huy Nguyen and Diane J. Litman. 2016b. Improving argument mining in student essays by learning and exploiting argument indicators versus essay topics. In Zdravko Markov and Ingrid Russell, editors, *Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2016, Key Largo, Florida, May 16-18, 2016.*. AAAI Press, pages 485–490. http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS16/paper/view/12791.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*. ACM, New York, NY, USA, ICAIL '09, pages 98–107. https://doi.org/10.1145/1568234.1568246.

Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, Baltimore, Maryland, pages 29–38. http://www.aclweb.org/anthology/W/W14/W14-2105.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to argumentation mining in texts: A survey. *Int.*

J. Cogn. Inform. Nat. Intell. 7(1):1–31. https://doi.org/10.4018/jcini.2013010101.

Georgios Petasis and Vangelis Karkaletsis. 2016. Identifying argument components through textrank. In *Proceedings of the 3rd Workshop on Argument Mining (ArgMining2016)*. Association for Computational Linguistics, Berlin, Germany, pages 56–66. http://aclweb.org/anthology/W/W16/W16-2811.pdf.

Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence - an automatic method for context dependent evidence detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 440–450. http://aclweb.org/anthology/D15-1050.

Niall Rooney, Hui Wang, and Fiona Browne. 2012. Applying kernel methods to argumentation mining. In G. Michael Youngblood and Philip M. McCarthy, editors, *Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference, Marco Island, Florida. May 23-25, 2012*. AAAI Press. http://www.aaai.org/ocs/index.php/-FLAIRS/FLAIRS12/paper/view/4366.

Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36(8):1627–1639.

Christian Stab and Iryna Gurevych. 2014a. Annotating argument components and relations in persuasive essays. In Junichi Tsujii and Jan Hajic, editors, *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pages 1501–1510. http://www.aclweb.org/anthology/C14-1142.

Christian Stab and Iryna Gurevych. 2014b. Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 46–56. http://www.aclweb.org/anthology/D14-1006.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics* 0(ja):1–62. https://doi.org/10.1162/COLI_a_00295.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. 2006. Hierarchical dirichlet processes. *Journal of the American Statistical Association* 101(476):1566–1581. https://doi.org/10.1198/016214506000000302.