Argument extraction for supporting public policy formulation

Eirini Florou
Dept of Linguistics, Faculty of Philosophy
University of Athens, Greece
`eirini.florou@gmail.coml`

Stasinos Konstantopoulos
Institute of Informatics and
Telecommunications, NCSR 'Demokritos'
`konstant@iit.demokritos.gr`

Antonis Kukurikos   **Pythagoras Karampiperis**
Institute of Informatics and Telecommunications, NCSR 'Demokritos', Athens, Greece
`{kukurik,pythk}@iit.demokritos.gr`

Abstract

In this paper we describe an application of language technology to *policy formulation*, where it can support policy makers assess the acceptance of a yet-unpublished policy before the policy enters public consultation. One of the key concepts is that instead of relying on thematic similarity, we extract *arguments* expressed in support or opposition of positions that are general statements that are, themselves, consistent with the policy or not. The focus of this paper in this overall pipeline, is identifying arguments in text: we present and empirically evaluate the hypothesis that verbal *tense* and *mood* are good indicators of arguments that have not been explored in the relevant literature.

1 Introduction

The large-scale acquisition, thematic classification, and sentiment analysis of Web content has been extensively applied to brand monitoring, digital reputation management, product development, and a variety of similar applications. More recently, it has also seen application in public policy validation, where the 'brand' to be monitored is a publicized and widely commented government policy.

All these methods typically rely on the *semantic similarity* between a given text or set of terms and Web content; often using domain-specific ontological and terminological resources in order to measure this similarity. This approach, however, requires that all parties involved discourse on the same topic; that is to say, that we are seeking the collective opinion of the Web on a topic that has been publicized enough to attract the attention of the Web.

In this paper we present a slightly different approach, where we are looking for *arguments* expressed in support or opposition of opinions with little semantic similarity to our topic of interest. As a rough example, consider how drafting environmental policy can benefit from access to statistics about how people felt about industrial growth at the expense of environmental concerns when other policy in completely different domains was on public consultation: many of the arguments about the relative merits of industrial growth and environmental concerns can retain their structure and be thematically transferred to the new domain, helping draft a policy that best addresses people's concerns.

Of paramount importance for implementing such an approach is the linguistic tools for identifying arguments. In this paper, we first motivate the inclusion of argument extraction inside the larger policy formulation and validation cycle and present the position of an argument extraction tool inside a computational system for supporting this cycle (Section 2). We then proceed to a present the argument extraction literature (Section 3) and our hypothesis that verbal morpho-syntactic features are good discriminators of arguments (Section 4). We close the paper by presenting and discussing empirical results (Section 5) and concluding (Section 6).

2 Policy formulation and validation

Our work is carried out in the context of a project that develops computational tools for the early phases of policy making, before policy drafts have been made available for public consultation.[1]

At that stage, the policy's impact on public opinion cannot be estimated by similarity-based searching for relevant Web content, since the policy text has not been announced yet — or even fully authored for that matter. One of the core

---

[1] Full details about the project have been suppressed to preserve anonymity, but will be included in the camera-ready.

ideas of the project is that in order to assist the policy formulation process, a tool needs to estimate the acceptance of a yet unpublished document based on Web content that is not thematically similar, but is rather supporting or opposing a more general position or maxim that also supports or opposes the policy under formulation.

To make this more concrete, consider a new policy for increasing the penetration of wind power production, setting specific conditions and priorities. The project is developing an authoring environment where specific policy statements are linked to more general statements, such as:

(1) Greenhouse gas emissions should not be a concern at all.

(2) It is desired to reduce greenhouse gas emissions, but this should be balanced against other concerns.

(3) It is desired to reduce greenhouse gas emissions at all costs.

We have, thus, created a formulation of 'relevant content' that includes Examples 1 and 2 below. These are taken from different domains, are commenting policies and laws that are already formulated and made public, and can be used to infer the level of support for the new wind power policy although no textual similarity exists.

(4) In case hard packaging is made compulsory by law, producers will be forced to consume more energy, leading to more greenhouse gas emissions.

(5) Tidal power production does not emit greenhouse gases, but other environmental problems are associated with its widespread deployment.

Leaving aside the ontological conceptualization that achieves this matching, which is reported elsewhere, we will now discuss the language processing pipeline that retrieves and classifies relevant Web content.

Content is acquired via *focused crawling*, using search engine APIs to retrieve public Web pages and social network APIs to retrieve content from social content sharing platforms. Content is searched and filtered (in case of feed-like APIs) based on fairly permissive semantic similarity measures, emphasising a high retrieval rate at the expense of precision. As a second step, clean text is extracted from the raw Web content using the Boilerpipe library (Kohlschütter et al., 2010) in order to remove HTML tags, active components (e.g. JavaScript snippets), and content that is irrelevant to the main content (menus, ad sections, links to other web pages), and also to replace HTML entities with their textual equivalent, e.g., replacing '&' with the character '&'.

The resulting text is tokenized and sentence-splitted and each sentence classified as relevant or not using standard information retrieval methods to assess the semantic similarity of each sentence to the general policy statements. This is based on both general-purpose resources[2] and the domain ontology for the particular policy. Consecutive sentences that are classified as positive are joined into a *segment*.

The main objective of the work described here is the classification of these segments as being representative of a stance that would also support or oppose the policy being formulated, given the premise of the general statements (1)–(3). Our approach is to apply the following criteria:

- That they are semantically similar to the general statements associated with the policy.

- That they are arguments, rather than statements of fact or other types of prose.

- That their polarity towards the general statements is expressed.

In order to be able to assess segments, we thus need a linguistic pipeline that can calculate semantic similarity, identify arguments, and extract their structure (premises/consequences) and polarity (in support or opposition).

The focus of the work described here is identifying arguments, although we also outline how the features we are proposing can also be used in order to classify chunks of text as premises or consequences.

3 Related Work

The first approaches of argument extraction were concentrated on building wide-coverage argument

---

[2] WordNets are publicly available for both English and Greek, that is the language of the experiments reported here. Simpler semantic taxonomies can also be used; the accuracy of the semantic similarity measured here does not have a major bearing on the argument extraction experiments that are the main contribution of this paper.

structure lexicons, originally manually Fitzpatrick and Sager (1980, 1981) and later from electronic versions of conventional dictionaries, since such dictionaries contain morpho-syntactic features Briscoe et al. (1987). More recently, the focus shifted to automatically extracting these lexical resources from corpora Brent (1993) and to hybrid approaches using dictionaries and corpora.

Works using syntactic features to extract topics and holders of opinions are numerous (Bethard et al., 2005). Semantic role analysis has also proven useful: Kim and Hovy (2006) used a FrameNet-based semantic role labeler to determine holder and topic of opinions. Similarly, Choi and Cardie (2006) successfully used a PropBank-based semantic role labeler for opinion holder extraction.

Somasundaran et al. (2008; 2010) argued that semantic role techniques are useful but not completely sufficient for holder and topic identification, and that other linguistic phenomena must be studied as well. In particular, they studied discourse structure and found specific cue phrases that are strong features for use in argument extraction. Discourse markers that are strongly associated with pragmatic functions can be used to predict the class of content, therefore useful features include the presence of a known marker such as 'actually', 'because', 'but'.

Tseronis (2011) describes three main approaches to describing argument markers: Geneva School, Argument within Language Theory and the Pragma-dialectical Approach. According to the Geneva School, there are three main types of markers/connective, organisation markers, illocutionary function markers (the relations between acts) and interactive function markers. Argument within Language Theory is a study of individual words and phrases. The words identified are argument connectors: these describe an argumentative function of a text span and change the potential of it either realising or de-realising the span. The Pragma-dialectical Approach looks at the context beyond words and expressions that directly refer to the argument. It attempts to identify words and expressions that refer to any moves in the argument process. Similarly to Marcu and Echihabi (2002), the approach is to create a model of an ideal argument and annotate relevant units.

4   Approach and Experimental Setup

As seen above, shallow techniques are typically based on connectives and other discourse markers in order to define shallow argument patterns. What has not been investigated is whether shallow morpho-syntactic features, such as the tense and mood of the verbal constructs in a passage, can also indicate argumentative discourse.

Our hypothesis is that *future* and *conditional* tenses and moods often indicate conjectures and hypotheses which are commonly used in argumentation techniques such as *illustration*, *justification*, *rebuttal* where the effects of a position counter to the speaker's argument are analysed. Naturally, such features cannot be the sole basis of argument identification, so we need to experiment regarding their interaction with discourse markers.

To make this more concrete, consider the examples in Section 2: although both are perfectly valid arguments that can help us infer the acceptance or rejection of a policy, in the first one future tense is used to speculate about the effects of a policy; in the second example there is no explicit marker that the effects of large-scale tidal power production are also a conjecture.

Another difficulty is that conditional and future verbal groups are constructed using auxiliary verbs and (in some languages) other auxiliary pointers. Consider, for example, the following PoS-tagged and chunked Greek translation of Example 4:

(6) *[oi   paragogoi]$_{np}$*
     [the-NomPl   producers-NounNomPl]
     [tha   ipochreothoun]$_{vp}$
     [Pointer-Fut force-Perf-3PP]
     [na   katanalosoun]$_{vp}$
     [Pointer-Subj consume-Inf]
     'producers will be forced to consume'

In order to be able to correctly assign *simple future*, information from the future pointer 'tha' needs to be combined with the perfective feature of finite verb form. Conditionals, future perfect, past perfect, and similar tenses or moods like subjunctive also involve the tense of the auxiliary verb, besides the future pointer and the main verb.

We have carried out our experiments in Greek language texts, for which we have developed a JAPE grammar[3] that extract the tense and mood of

---

[3]JAPE is finite state transducer over GATE annota-

Table 1: Categories of morpho-syntactic features extracted from text segments.

Label	Description	Features
DM	Absolute number of occurrences of discourse markers from a given category	5 numerical features
Rel	Relative frequency of each of the 6 tenses and each of the 6 moods	12 numerical features
RCm	Relative frequency of each tense/mood combination (only for those that actually appear).	9 numerical features
Bin	Appearance of each of the 6 tenses and each of the 6 moods	12 binary features
Dom	Most frequent tense, mood, and tense/mood combination	3 string features
	TOTAL	41 features

each verb chunk. The grammar uses patterns that combine the features of pointers and auxiliary and main verbs, without enforcing any restrictions on what words (e.g., adverbs) might be interjected in the chunk. That is to say, the chunker is responsible for identifying verb groups and our grammar is restricted to propagating and combining the right features from each of the chunk's constituents to the chunk's own feature structure.

PoS-tagging and chunking annotations have been previously assigned by the ILSP suite of Greek NLP tools (Papageorgiou et al., 2000; Prokopidis et al., 2011), as provided by the relevant ILSP Web Services[4] to get PoS tagged and chunked texts in the GATE XML format.

At a second layer of processing, we create one data instance for each segment (as defined in Section 2 above) and for each such segment we extract features relating to verbal tense/mood and to the appearance of discourse markers. The former are different ways to aggregate the various tenses and moods found in the whole segment, by measuring relative frequencies, recording the appearance of a tense or mood even once, and naming the predominant (most frequent) tense and mood; tense and mood are seen both individually and as tense/mood combinations.

Furthermore, we have defined five absolute frequency features which record the matching against the several patterns and keywords provided for the following five categories of arguments:

- justification, matching patterns such as 'because', 'the reason being', 'due to', etc.

- explanation, matching patterns such as 'in other words', 'for instance', quotesfor this reason(s), etc.

- deduction, 'as a consequence', 'in accordance with the above', 'proving that', etc.

- rebuttal, 'despite', 'however', etc.

- conditionals, 'supposing that', 'in case that', etc.

All features extracted by this process are given on Table 1.

5 Results and Discussion

We have used the method described in Section 2 in order to obtain 677 text segments, with a size ranging between 10 and 100 words, with an average of 60 words. Of these, 332 were manually annotated to *not* be arguments; the remaining 345 positive examples were obtained by oversampling the 69 segments in our corpus that we have manually annotated to be arguments.[5]

We have then applied the feature extraction described in Section 4 in order to set up a classification task for J48, the Weka[6] implementation of the C4.5 decision tree learning algorithm (Quinlan, 1992). We have applied a moderate confidence factor of 0.25, noting that experimenting with the confidence factor did not yield any substantially different results.

In order to better understand the feature space, we have run a series of experiments, with quantitative results summarized in Table 2. The first

---

tions. Please see http://gate.ac.uk/sale/tao/splitch8.html The JAPE grammar we have developed will be made available on-line; location cannot be yet disclosed in order to preserve anonymity.

[4]Currently at http://ilp.ilsp.gr

[5]The data and relevant scripts for carrying out these experiments are available at http://users.iit.demokritos.gr/~konstant/dload/arguments.tgz

[6]Please see http://www.cs.waikato.ac.nz/ml/weka

Table 2: Precision and recall for retrieving arguments using different feature mixtures. Please cf. Table 1 for an explanation of the feature labels. The results shown are the 10-fold cross-validation mean.

Morpho-syntactic features used	With Discourse Markers			Without Discourse Markers		
	Prec.	Rec.	$F_{\beta=1}$	Prec.	Rec.	$F_{\beta=1}$
All	75.8%	71.9%	73.8%	75.5%	70.4%	72.9%
no Dom	79.8%	73.3%	**76.4%**	74.0%	71.9%	72.9%
no Rel	74.5%	72.8%	73.8%	73.1%	69.3%	71.1%
no RCm	76.3%	71.0%	73.6%	76.8%	70.1%	73.3%
no Bin	70.0%	70.4%	70.2%	66.7%	69.6%	68.1%
Rel	73.4%	75.9%	74.6%	70.3%	72.2%	71.2%
Dom	57.1%	98.8%	72.4%	54.9%	94.2%	69.4%
RCm	69.3%	66.7%	67.9%	71.9%	62.9%	67.1%
Bin	71.7%	49.9%	58.8%	70.1%	44.9%	54.8%
None	67.9%	20.9%	31.9%		—	

observation is that both morpho-syntactic features and discourse markers are needed, because if either category is omitted results deteriorate. However, not *all* morpho-syntactic features are needed: note how omitting the *Dom*, *Rel*, or *RCm* categories yields identical or improved results. On the other hand, the binary presence feature category *Bin* is significant (cf. 5th row). We cannot, however, claim that only the *Bin* category is sufficient, and, in fact, if one category has to be chosen that would have to be that of relative frequency features (cf. rows 6-9).

6 Conclusion

We describe here an application of language technology to *policy formulation*, and, in particular, to using Web content to assess the acceptance of a yet-unpublished policy *before* public consultation. The core of the idea is that classifying Web content as similar to the policy or not does apply, because the policy document has not been made public yet; but that we should rather extract arguments from Web content and assess whether these argue in favour or against general concepts that are (or are not) consistent with the policy being formulated.

As a first step to this end, our paper focuses on the identification of arguments in Greek language content using shallow features. Based on our observation that verb tense appears to be a significant feature that is not exploited by the relevant literature, we have carried out an empirical evaluation of this hypothesis. We have, in particular, demonstrated that the relative frequency of

each verb tense/mood and the binary appearance of each verb tense/mood inside a text segment are as discriminative of argumentative text as the (typically used) discourse markers; and that classification is improved by combining discourse marker features with our verbal tense/mood features. For doing this, we developed a regular grammar that combines the PoS tags of the members of a verb chunk in order to assign tense and mood to the chunk. In this manner, our approach depends on PoS taggin and chunking only.

In subsequent steps of our investigation, we are planning to refine our approach to extracting argument structure: it would be interesting to test if argument premises tend to correlate with certain tenses or moods, distinguishing them from from conclusions. Further experiments can also examine if the simultaneous appearance of concrete tenses at the same sentence is an indicator of an argument. Finally, we plan to examine the predicates of an argument, and especially if the head word of each sentence (be it verb or deverbal noun) and its seat at the boundaries of the sentence may contribute to extract an argument or not, especially for impersonal, modal, and auxiliary verbs.

Acknowledgements

References

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2005. Extracting opinion propositions and opinion holders using syntactic and lexical cues. *In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, Computing Attitude and Affect in Text: Theory and Applications. Springer.*

Michael Brent. 1993. From grammar to lexicon: unsupervised learning of lexical syntax. In *Computational Linguistics - Special issue on using large corpora: II, Volume 19 Issue 2*, pages 243–262.

Ted Briscoe, Claire Grover, Bran Boguraev, and John. Carroll. 1987. The derivation of a grammatically-indexed lexicon from the longman dictionary of contemporary english. In *Proceeding of ACL '87.*

Yejin Choi and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of EMNLP 2006.*

Eileen Fitzpatrick and Naomi Sager. 1980. *The Lexical subclasses of the LSP English Grammar.* Linguistic String Project, New York University.

Eileen Fitzpatrick and Naomi Sager. 1981. The lexical subclasses of the lsp english grammar. In *N. Sager (ed), Natural Language Information Processing, Addison- Wesley, Reading, Ma.*

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressedin online news media text. In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text.*

Christian Kohlschütter, Peter Fankhauser, and Wolfgang Nejdl. 2010. Boilerplate detection using shallow text features. In *Proc. 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010)* New York, USA.

Daniel Marcu and Abdessamad Echihabi. 2002. An unsupervised approach to recognizing discourse relations. In *Proc. ACL '02.*

Haris Papageorgiou, Prokopis Prokopidis, Voula Giouli, and Stelios Piperidis. 2000. A unified POS tagging architecture and its application to Greek. In *Proceedings of the 2nd Language Resources and Evaluation Conference (LREC 2000), Athens, Greece*, pages 1455–1462.

Prokopis Prokopidis, Byron Georgantopoulos, and Haris Papageorgiou. 2011. A suite of NLP tools for Greek. In *Proceedings of the 10th International Conference of Greek Linguistics (ICGL 2011), Komotini, Greece.*

J. Ross Quinlan. 1992. *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA, USA.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proc. NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text (CAAGET 2010)*

Swapna Somasundaran, Janyce Wiebe, and Joseph Ruppenhofer. 2008. Discourse level opinion interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, COLING '08, Stroudsburg, PA, USA. Association for Computational Linguistics.*

Assimakis Tseronis. 2011. From connectives to argumentative markers: A quest for markers of argumentative moves and of related aspects of argumentative discourse. *Argumentation*, 25(4):427–444.