Computer Science and Information Systems 17(1):93–115 https://doi.org/10.2298/CSIS180601015G

Climate Change Opinions in Online Debate Sites

Adrian Groza[1], Pinar Ozturk[2], Radu Razvan Slavescu[1], and Anca Marginean[1]

[1] Technical University of Cluj-Napoca
400128 Cluj-Napoca, Romania
{Adrian.Groza,Radu.Razvan.Slavescu,Anca.Marginean}@cs.utcluj.ro
[2] Norwegian University of Science and Technology
Trondheim, Norway
pinar@ntnu.no

Abstract. Debate sites in social media provide a unified platform for citizens to discuss controversial questions and to put forward their ideas and arguments on the issues of common interest. Opinions of citizens may provide useful knowledge to stakeholders but manual analysis of arguments in debate sites is tedious, while computational support to this end has been rather scarce. We focus here on developing a technical instrumentation for making sense of a set of online arguments and aggregating them into usable results for policy making and climate science communication. Our objectives are: (i) to aggregate arguments posted for a certain debate topic, (ii) to consolidate opinions posted under several but related topics either in the same or different debate site, and (iii) to identify possible linguistic characteristics of the argumentative texts. For the first objective, we propose a voting method based on subjective logic [13]. For the second objective, we assess the semantic similarity between two debate topics based on textual entailment [28]. For the third objective, we employ various existing methods for lexical analysis such as frequency analysis or readability indexes. Although we focused here on the climate change, the method can be applied to any domain.

Keywords: online debate analysis, aggregation of individual opinions, web text analysis, decision support for policy making.

1. Introduction

Policy makers, managers and social scientists are interested in opinions of stakeholders on issues of environmental, societal and political consequences. Although social media has proven to be a precious data source for studying how people use public arena for communicating their ideas and opinions [7,34], debate sites have not been in research focus to the same extent as other online platforms. The objective of this work is to investigate what kind of information can be extracted from individual opinions posted on debate sites. We focus here on the climate change problem because it is a matter that interests many people who may have different opinions and arguments. However, the method is general and can be used in other areas as well.

Debate sites are structured according to topics, (e.g. ”global warming”). Anybody may post a question (e.g., ”Is global warming affecting the planet?”) or a hypothesis (e.g., “Global warming is affecting the planet”), and anybody can post his or her opinion related to this question/hypothesis. In the rest of this paper we use ”hypothesis” regardless of the initiating post being in affirmative or interrogative format. The responses are votes (e.g., yes/no, pro/against or agree/disagree), which are optionally accompanied by an argumentative text. From the debate analysis perspective, the debate sites therefore possess a distinguished advantage: people's opinions about a debate topic are intrinsically labeled as pro or against, which enables automated extraction of labeled arguments.

It is not unusual that the same or similar hypotheses are discussed in more than one thread in the same debate site, even synchronously, because the debate sites do not offer a service for detecting such redundancy. For example, we noticed that *"Climate change affects the earth"* and *" Global warming affects our planet"* were debated at almost the same time within the same debate community. Moreover, the same hypothesis can be posted on the distinct debate community, where it may attract more (or less) negative (or positive) arguments. Thus, there is need for computational methods to handle these situations in order to have a clearer picture on what is debated related to a topic of interest. Hence, we propose here a computational method and a tool to facilitate a high level view on what is debated online.

There are several challenges in making sense of online debates. First, redundancy occurs when a person posts an existing hypothesis again, with a different wording. Second, the number of responses vary significantly across topics/hypotheses which makes it difficult to compare the degree of support for two hypotheses, one with tiny and the other with massive discussions. Third, different hypotheses may be considered the same for a specific purpose, for example, of a policymaker and hence the responses to them may need to be merged. Fourth, there are several debate sites, which we call "communities", independent from each other but discussing similar or the same topics. Gathering a consolidated opinion across these communities will provide a better insight into public opinions. However, it is not trivial to assess the semantic similarity between hypotheses and hence to extract collective opinions of people from distinct debate sites.

We used debate sites to extract an annotated corpus of climate change arguments in natural language. The motivation is that existing corpora for climate change are based either on media [4], or tweets [14,23]. Both sources do introduce specific disadvantages for natural language processing. First, arguments conveyed in media are too large and sparse within an article or news. Second, arguments in tweets do not follow a specific set of grammar rules. We consider that arguments from debate sites are more adequate for natural language processing (NLP), as arguments are smaller than media documents and they are grammatically more correct than tweets. Moreover, the existing corpora contain arguments labeled as pro or against either manually by external human annotators or automatically (e.g. based on machine learning). Differently, the arguments in our corpus are labeled by the conveyor of the argument himself/herself. That is, the confidence in the labels is higher. Hence, such a corpus can be useful for researchers in natural language arguments or argument mining.

Our objectives are (1) to aggregate arguments posted for a certain hypothesis, (2) to consolidate opinions posted under several but related hypotheses either in the same or different debate site, and (3) to identify possible linguistic characteristics of the argumentative texts. Note that we reserve the term *aggregation* for a summary of opinions under a specific hypothesis posted in one thread, while *consolidation* is used whenever two separated threads about a topic can semantically be merged.

For the first objective, we proposed a vote-based method based on Subjective Logic [13]. For the second objective, we assess the semantic similarity between two hypotheses based on textual entailment [28]. For the third objective, we employ various existing lexical analysis instrumentations such as frequency analysis or readability indexes. Although we focused here on the climate change, the method can be applied to any domain.

A social scientist using our ARGSENSE tool can obtain answers related to the following research questions:

Q_1**:** Are the arguers within a community apriori prone to accept or to reject a hypothesis?
Q_2**:** Which hypotheses are most (dis)believed or (un)popular in a community?
Q_3**:** Do the pro arguments have a different lexicon than the counter ones?
Q_4**:** Does an interrogation have more pros or more cons arguments than an affirmation?
Q_5**:** Are the pro arguments more readable than the con arguments?
Q_6**:** Is the length of hypothesis correlated with the number of arguments it receives?
Q_7**:** Does the formulation of the hypothesis itself (e.g., interrogative or affirmative) influence the degree of interest in the debate?

In the rest of the paper, section 2 browses related work on analysing climate change arguments. Section 3 introduces the climate change argument corpus that we harvested from debate sites, and the architecture of ARGSENSE for supporting the analysis of the online debates. Section 4 presents our vote-based method for argument aggregation based on subjective logic. Section 5 presents a method based on textual entailment for consolidating opinions of related debate topics. Section 6 applies opinion aggregation and opinion consolidation in the climate change domain. Section 7 applies lexical analysis for supporting social scientists on the climate change corpus. Finally, we review the findings through a concluding section.

2. Related work

Related work is restricted to our running scenario: climate change. We approach related work from three perspectives. First, we introduce different approaches for analysing online arguments on global warming. Second, we browse the existing corpora on climate change. Third, we present related tools that support understanding of climate change.

2.1. Opinion aggregation

We are aware of the limitations and risks of aggregating data from online sources. For example, people from online communities disagree far more on climate change than climate experts do. The scientific community has reached a consensus that the rise of average temperature is mostly caused by human activity. It has been argued by Boussalis et al. [4] that the lack of awareness or understanding of the scientific evidence is due to a ”coordinated and well-funded counter-movement of climate skeptics“. Relevant to the topic of climate change denial is the investigation in [32]. Here, Washington et al. have analysed the argumentative patterns in climate change literature and have identified five types of climate change denial argument: i) conspiracy theory, ii) fake expert, iii) impossible expectations, iv) misrepresentation or logical fallacy v) cherry-picking. From the argumentation technologies viewpoint, these results [32] open the way towards argumentation schemes [31] for climate change.

In the same line of aggregating expert opinions instead of non-expert arguments is the work of HaDoung et al. that compared several procedures to aggregate expert opinion based on an Transferable Belief Model. The approach in [10] has been tested with 16 expert real-world datasets on climate sensitivity. The experts are firstly clustered into fields of taught. In each group, beliefs are aggregated using a cautious conjunction operator. Across groups, a non-interactive disjunction is used. In our case, the arguers are partitioned into different debate sites. Within a community, the argument properties (belief, disbelief, ignorance) are consolidated based on the similarity, contradictory and entailment relations.

Our method for representing individual votes is inspired from the subjective logic of [32]. Lioma et al. have used subjective logic for interactive information retrieval [18]. Subjective logic has been used to model representation of information needs as uncertain beliefs. In [18], the same information need can have various textual representation. Similarly, our debate topics or hypotheses have different supporting arguments.

For consolidation of opinions posted under different (but semantically related) hypothesis, we used natural language processing techniques such as textual entailment. Textual entailment is used here to compute similarity, contradiction and entailment between hypotheses. The latter one includes a supervised machine learning component, that benefits from our crawled labeled arguments. Moreover, in line with [2] our method for detecting similarity, contradiction and entailment relies also on external knowledge resources like Wordnet [22] and VerbOcean [8].

2.2. Climate change debate corpora

Existing corpora for climate science are based on media documents [4], or tweets [14,23]. Boussalis et al. [4] has collected 16.000 documents for compiling a corpus of contrarian literature on climate change. 19 organisations known to argue for climate skepticism were included. The corpus was used to analyse the skeptical discourse on global warming over the period 1998-2013. Kirilenko et al. [14] have collected 1.8 million tweets on climate change between 2012 and 2013 in five languages with 41% of the daily discussion of climate change on Twitter originates from the USA and 13% from the UK. Tweets have also been collected for the stance detection dataset [23]. The corpus contains 4870 tweets in five domains: ”atheism”, ”climate change is a real concern”, ”feminist movement”, ”Hillary Clinton” and ”legalisation of abortion”. Each tweet was annotated by at least eight respondents in the CrowdFlower (http://www.crowdflower.com) platform.

From the natural language perspective, our climate change corpus has a technical advantage over the tweets-based corpora or document-based corpora. The advantage comes from the size and structure of the arguments. Tweets are charactered by flexible grammar structure, and lots of links or hash-tags. Hence, natural language processing is based more or less on lexical analysis within a statistical framework. Differently, media documents or journal articles are large and hence it is difficult to automatically filter the relevant information. Our corpus contains small arguments supporting and attacking a debate topic. The size of each pair of arguments makes it possible to effectively apply technical instrumentations such as textual entailment.

Regarding the size of each text, the corpus of Kwon et al. [16] is similar. Kwon at al. have collected 119 public comments about Environmental Protection Agency’s proposed emission standard rule on hazardous pollutants. The comments have been classified in three classes: support, oppose or propose a new idea. Each comment has been annotated by at least two coders. The inter-annotator agreement based on Cohen's Kappa coefficient [15] has been only 0.62. This low value signals the uncertainties rising when the comments are manually labeled by human annotators. Our larger corpus (11.653 arguments compared to 119 comments) does not have this labeling uncertainty, as each label was available from the debate sites. Hence, an advantage of our corpus is that the classification of an argument as positive or negative is given by the conveyor of the argument and not by an external annotator.

Chalaguine and Schulz [6] have focused on how convincing arguments are in online debates. The corpus contains arguments collected from 32 debates on 16 topics from *createdebate.com* and *procon.org*. The collected arguments have been used to generate 16k pairs of arguments. Each pair has been classified by human annotators as more convincing versus less convincing. Note that the arguments are not restricted to only one domain (i.e., climate change). The more topics in the corpus, the more difficult for machine learning to learn the language model. Note also that assessing the strength of an arguments in 16K pairs is a highly subjective task, given the bias due to personal beliefs, preferences and background of the annotators. Moreover, the randomly generated pairs contain arguments from different topics, hence more difficult to assess their strength. Chalaguine and Schultz have used the corpus by extracting 70 features for each pair (POS, statistics on texts, etc.) and fed these features to a neural network.

Hence, our climate change corpus provides the following advantages compared to existing corpora used in the argumentation mining community. First, the size of each argument is relatively small (compared to large scientific documents) and has a proper grammar (compared to tweets). Second, the corpus is restricted to a single domain (i.e. climate change), hence it facilitates building a better language model, in case machine learning is used. Third, classification of arguments into pros and cons is guaranteed to reflect the conveyor's real intention, and does not include inherent errors of manual annotation [16,23].

2.3. Tools for climate change understanding

The ARGSENSE tool aims to enhance understanding on climate change topic as it appears in public arena. Henceforth, this section browses related tools used to support stakeholders to understand climate science.

Launched in November 2014, the Web-based tool VisAdap [1] aims to increase understanding of anticipated risks from climate change. These climate change risks are addressed from the perspective of the target group of home-owners. The challenge is that home-owners have socio-cognitive barriers to adapt to the new risks caused by climate change. Risk impact of climate change is anticipated for a given region over the coming 40-60 years. With 16,000 users within 8 months [1], VisAdap has proved a popular tool to support individual decision making when buying or building a house. Currently, VisAdap is designed based on norms and values of home-owners collected from direct interviews. Instead of interviews, ARGSENSE relies on arguments conveyed in the public arena. In the same line of understanding people opinions on a narrow topic, ARGSENSE can perform topic-based analysis. The topic can be selected based on the target stakeholders, including the home-owners. For home-owners, ARGSENSE can report on the arguments related to debate topics such as flooding or storm damage. ARGSENSE is able to signal topics which are not accepted by a community. This result guidelines a policymaker regarding which information to communicate on platforms like VisAdap.

Launched in 2016, the AGCLIMATE hub (http://AgClimate4U.org) aims to transform heterogeneous climate change datasets into usable information in agriculture. Here, the stakeholders are crop farmers and agricultural advisors [3]. To understand their needs, methods from social sciences have been used: surveys, focus groups, interviews, and network analysis. Findings from these methods have been used to design various decision support tools for graphical visualisation of climate data, crop fields, or irrigation investment. ARGSENSE follows the same ”useful to usable” paradigm: the argumentation dataset is analysed from different perspectives to present usable findings to a policymaker. Data in AGCLIMATE is obtained by querying the Web services of various providers (i.e. National Oceanic Atmospheric Administration, Midwestern Regional Climate Center, USDA National Agricultural Statistics Service) of climate change structured data. Differently, data in ARGSENSE is obtained through crawling and it is in textual form. Hence, natural language processing techniques (such as textual entailment) were used to aggregate arguments. These aggregation contributes to increase usability for policymakers and social change agents.

The two works above [1,3] have focused on understanding the views on climate change of *people*. Differently, Mayer et al. [21] have recently focused on understanding mental models on climate change of the *scientists*. Mayer et al. [21] have argued that understanding the decisions for model design are important for the informed use of tools built on that models. The research method has been to qualitatively analyse semi-structured interviews with eleven interdisciplinary experts (i.e., climate scientists, economists, decision analysts) who lead projects in the field of climate risk management. For managing climate change risks, Mayer et al. have been interested in analysing decision's *justification* and *explanations*. In the same line of focusing on expert knowledge and not public arguments, Huang et al. [12] have developed an expert system for integrating climate change impact in the petroleum industry with the aim to support formulation of the relevant adaptation policies. Similarly, Qin et al. [27] have proposed an expert system to assess the impact of climate change on socio-economic and environmental factors. These relevant factors are further used by the expert system to formulate adaptation policies. In this paper, we were interested to analyse, people's *arguments*. These may lead to an interesting interplay between arguments, justifications and explanations [17,30].

The VA-TURF tool [20] aims to assess the vulnerability of coastal fisheries ecosystems. VA-TURF includes socio-economical aspects and enhances planning of coastal communities to climate change impacts. The research method has been to assess vulnerability directly by fishers, barangay leaders, residents, and local executive staff. These stakeholders can discuss within the VA-TURF system on risks associated to a particular climate change scenario. ARGSENSE can also provide insights on a particular climate change topic. Yet the target communities are different in VA-TURF and ARGSENSE : VA-TURF encourages local community-level actions. Differently, ARGSENSE analyses arguments conveyed by a debate community that is globally distributed.

The ArgueApply has been launched in 2017 as a mobile application for generic debates [26]. The argumentation model of ArgueApply is based on four different relations: support, strong support, attacks, and strong attacks. The voting method counts strong support twice compared to the support relations. This refinement is necessarily in our view to overcome cases in which most of the semantics of abstract argumentation output the empty set. Differently, we have only the support and attack relations. Instead, our approach based on subjective logic can be used to distinguish between the following cases: i) let a debate topic supported by 3 arguments and rejected by 2; ii) consider also a debate topic supported by 30 arguments and attacked by 20. The ignorance level in our argumentation model allows to distinguish between such arguments. The topic with 30/20 arguments is considered more accepted in a debate community. The capacity of modeling ignorance level is why we used subjective logic for our argumentation model. Moreover, our ARGSENSE tool complements its vote-based method with text analysis capabilities (textual entailment, lexical analysis).

3. Methods and tool

This section describes 1) the climate change argument corpus that we crawled for our experiments, and 2) the architecture of the ARGSENSE tool that we developed to facilitate the analysis of online debates.

First, we created a corpus (denoted cc) for the *Climate Change* domain from the three debate sites we selected: ForAndAgainst (henceforth faa), Debate.org (deb) and Debatepedia (dbp). All debate hypotheses discussing climate change were filtered based on the Wikipedia glossary of climate change. First, the crawled opinions are automatically structured in tuples $\langle h, t, l \rangle$, where h represents the debate hypothesis, t the argument in natural language (optional, hence may be empty), and l is the label of the vote pro (i.e., yes or agree) or cons (no or disagree) (see Table 1). Note that the label (pro or con) is provided by the conveyor of the argument. This label of the argument is automatically crawled from the webpage. This nice feature of the debate sites makes them an ideal source for extracting arguments which are already classified (i.e. labeled).

Table 1. Sample of tuples $\langle h, t, l \rangle$ in the climate change corpus.

Hypothesis (h)	Argument (t)	Label (l)
Climate change is man-made.	Human carbon emissions have accelerated global warming ...	pro
Climate change is man-made.	The climate has changed through history due to natural cycles.	cons
Should government adopt emissions trading to combat global warming?	Emissions trading encourages investments in technologies.	pro

There are 1,793 hypotheses in the corpus, and total 11,653 separate responses, i.e., arguments for the whole hypotheses repertoire. The cc corpus was obtained by crawling three debate communities: faa with 142 debates containing 877 arguments, deb with 742 topics containing 6,026 arguments, dbp with 909 debates on climate change attracting 4,750 arguments. With the resulted total of 11,653 arguments, the climate change corpus is, to our knowledge, the largest corpus of labeled arguments on climate change [3].

[3] The ARGSENSE tool and the climate change corpus are available at http://users.utcluj.ro/∼agroza/projects/argclime.

Second, we built the ARGSENSE tool to support the analysis of people's opinions expressed in debate sites. The system is helpful for social scientists and policymakers in getting an insight into people's attitudes toward the controversial issues of worldwide interest. ARGSENSE has two architectural components relying on a *vote-based method* and *text-based methods* respectively (see Fig. 1).

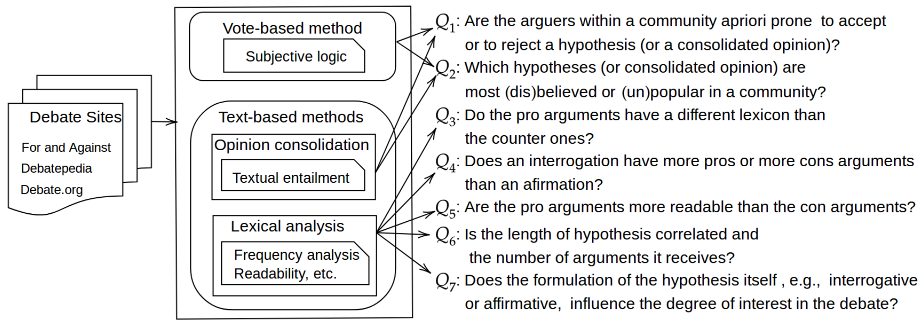


Fig. 1. ARGSENSE investigation domain. A voting method based on subjective logic is proposed to rank the debate topics based on belief, disbelief and popularity in a community of arguers. An opinion consolidation method is proposed to aggregate arguments from related debate topics. This supports a more accurate view on the same questions Q_1 and Q_2. Lexical analysis uses off-shelf frequency analysis tools to support social scientists and science communicators with questions Q_3 to Q_7.

The vote-based method takes tuples $\langle h, t, l \rangle$ by crawling the debate sites and aggregates votes l (of type pro or cons) for the same debate topic h. The aggregation is based on subjective logic. Subjective logic allows also to quantify belief and disbelief in h, but also the degree of ignorance in a community with respect to a debate topic h. The vote-based method helps a social scientist with answer to questions Q_1 and Q_2.

Text-based methods have two components: opinion consolidation and lexical analysis.

By *opinion consolidation* we mean the operation of aggregating arguments of semantic similar hypotheses. The semantic similarity relation is computed using textual entailment. Textual entailment identifies hypotheses representing the same debate topic, but posted using different words (e.g. *Climate change is manmade* and *Global warming is caused by humans*). Such related hypotheses can also be posted in different debate communities or posted in the same community but at different time points. To better support the social scientist, we need to consider all the arguments posted for or against all the hypotheses representing the same debate topic. We call this process opinion consolidation. Opinion consolidation is based on textual entailment and it represents the main conceptual proposal of this research.

Note that the vote-based method can be applied either on a single topic or on the consolidated topic that includes arguments from all related debates. In Fig. 1, this is illustrated by the fact that questions Q_1 and Q_2 can be applied both on a single hypothesis or on the consolidated debate topic. Throughout the paper, we use the term "aggregation" for a summary of opinions (vote-based) under a specific hypothesis, while ”consolidation” is used whenever two separately posted hypotheses can semantically be merged.

Lexical analysis identifies linguistic features of argumentative texts. One can investigate if a community of people uses specific linguistic patterns, and whether these patterns depend upon the topic or whether the discourse is supporting or countering. The results are presented visually through graphs, rankings, and the identified lexical patterns. The methods used for lexical analysis are not new - we use readability indexes, sequential pattern mining, statistical analysis. Instead, these features help a social scientist or policymaker for answering questions Q_3 to Q_7.

4. Aggregation of arguments for an individual hypothesis

Now we describe the method for translating the individuals' arguments posted under one thread of particular hypothesis in one debate site into an *aggregated opinion*.

To represent aggregated opinions we use subjective logic [13,9], which originally was developed for belief representation in knowledge bases. In subjective logic, an opinion ω on a given state of a system x is represented in terms of four quantities: $\omega_x = (b_x, d_x, u_x, a_x)$, where b_x represents an individual's degree of belief that the particular state x is true, d_x stands for disbelief and shows the belief that a state is false, and u_x is the uncertainty about the state. The parameter a_x is a measure of the *prior* probability of the truth value of x. In our case, the state x represents the hypothesis h for which people have provided arguments.

Differently from [13], we prefer the term *ignorance* instead of *uncertainty*, as it fits better to our task of assessing the degree in which a community is interested in a specific topic. Differently from [13], we also introduce the notion of community, to count only the arguments conveyed within a community or arguers.

The aggregated opinion of a community α about a hypothesis *h* is defined by:

Definition 1. *The opinion ω_h^α regarding the perceived truth value of hypothesis h by community α is a quadruple $\omega_h^\alpha = \langle b_h, d_h, i_h, a_h^\alpha \rangle$, where b_h represents the degree of belief (amount of evidence supporting h), d_h represents the disbelief (amount of evidence attacking h) and i_h represents the degree of ignorance about h with*

$$b_h + d_h + i_h = 1, \qquad \{b_h, d_h, i_h\} \in [0,1]^3 \tag{1}$$

The parameter a_h^α is a measure of the prior probability of the truth value of h in the community α. Hence, a_h^α is a feature of the community α. With no apriori information about α, we consider that a hypothesis has equal chances to be accepted or rejected.

In our framework, evidence for h are the arguments supporting or attacking h. For community α, let $\mathcal{A}_h^+$ be the set of arguments supporting h, and $\mathcal{A}_h^-$ the set of arguments attacking h. Let $e_h = |\mathcal{A}_h^+|$ be the number of arguments supporting h, and $n_h = |\mathcal{A}_h^-|$ the number of arguments attacking h. The parameters b_h, d_h and i_h are computed with:

$$b_h = \frac{e_h}{e_h + n_h + 1/a_h^\alpha} \tag{2}$$

$$d_h = \frac{n_h}{e_h + n_h + 1/a_h^\alpha} \tag{3}$$

$$i_h = \frac{1/a_h^\alpha}{e_h + n_h + 1/a_h^\alpha} \tag{4}$$

Example 1 illustrates the opinion ω_h^α for the h=*"Climate change is man-made"*.

Example 1. Assume h="Climate change is man-made" receives $A_h^+ = \{t_1, t_2, t_3, t_4, t_5\}$ and $A_h^- = \{t_6, t_7, t_8\}$. With no apriori information about community α ($a^0 = 0.5$), we have $b_h = 5/(5+3+2) = 5/10$, $d_h = 3/(5+3+2) = 3/10$, $u_h = 2/(5+3+2) = 2/10$. That is the opinion $\omega_h^\alpha = \langle 0.5, 0.33, 0.22, 0.5 \rangle$.

For particular values for b_h, d_h or i_h, special types of opinions can be defined: i) *vacuous opinion*: $i_h = 1$ (maximum ignorance, when no argument is available for h); ii) *dogmatic opinion*: $i_h = 0$ (no ignorance; theoretically, this happens if the number of arguments is infinite); iii) *neutral opinion*: $b_h = d_h$; iv) *equidistant opinion*: $b_h = d_h = i_h$; v) *pure opinion*: $b_h = 0$ or $d_h = 0$; vi) *negative opinion*: $b_h < d_h$ (when $d_h = 1$ we have an *absolute negative opinion*);vii) *positive opinion*: $b_h > d_h$.

The fourth parameter a^α is global to the community α where h is debated. With no apriori information regarding the acceptance of h by a community of agents, a^α defaults to 0.5. More accurate representation of a^α is obtained on the basis of the distribution of positive and negative opinions. Let $\mathcal{P}^\alpha$ be the set of hypotheses in a debate community α having more positive opinions than negative ones, given by $\mathcal{P}^\alpha = \{h \in \mathcal{H}^\alpha | e_h > n_h\}$. Let $\mathcal{N}^\alpha$ be the set of hypotheses in the community α having more negative opinions, given by $\mathcal{N}^\alpha = \{h \in \mathcal{H}^\alpha | n_h > e_h\}$. With this interpretation we have:

$$a^\alpha = \frac{|\mathcal{P}^\alpha|}{|\mathcal{P}^\alpha| + |\mathcal{N}^\alpha|}, \qquad \forall h \in \mathcal{H}^\alpha \tag{5}$$

The remaining $\mathcal{E}^\alpha = \mathcal{H}^\alpha \setminus \mathcal{P}^\alpha \setminus \mathcal{N}^\alpha$ is the set of neutral hypotheses in α.

A topic h is not necessarily independent from all other topics in the same community. There can be topics claiming the contrary of h or topics claiming the same idea of h but with different linguistic expressions. Therefore, we are interested next in exploiting these inter-relations between hypotheses in α, to obtain a clearer and consolidated opinion.

5. Consolidation of opinions from related hypotheses

If two hypotheses are semantically close to each other, we may want to consolidate the opinions expressed for them, because it may give more information about people's attitude towards the underlying debate topic. Such hypotheses may be posted in one debate site or different ones. The question is then how to judge semantic closeness between two hypotheses. Our computational method uses three relations for semantic closeness: similarity, contradiction and entailment.

Example 2 (Similar hypotheses). Consider g="Climate change is manmade" and h="Global warming is human made". Since g is similar to h, their supporting and attacking arguments can be aggregated.

Let $h, g \in \mathcal{H}^{\alpha}$, $e_h = |\mathcal{A}_h^+|$, $n_h = |\mathcal{A}_h^-|$, $e_g = |\mathcal{A}_g^+|$, $n_g = |\mathcal{A}_g^-|$.

Definition 2 (Consolidating opinions for similar hypotheses). *If h is similar to g ($h \sim g$) then the number of positive and negative arguments for computing the consolidating opinion $\hat{\omega}_h^{\alpha}$ are:*

$$\hat{e}_h = \hat{e}_g = e_h + e_g \tag{6}$$
$$\hat{n}_h = \hat{e}_g = n_h + n_g \tag{7}$$

Example 3 (Contradictory hypotheses). Let g="Climate change is a natural cycle". As h claims the opposite of g, the supporting arguments for h are the attacking arguments for g, while the supporting arguments for g attack h.

Definition 3 (Consolidating opinions for contradictory hypotheses). *If h contradicts g ($h \sim \neg g$) then the number of positive and negative opinions for computing the consolidated opinion $\hat{\omega}_h^{\alpha}$ are:*

$$\hat{e}_h = \hat{n}_g = e_h + n_g \tag{8}$$
$$\hat{n}_h = \hat{e}_g = n_h + e_g \tag{9}$$

Example 4 (Entailed hypotheses). Let k="Climate-induced changes are likely to cause effects involving many species of plants and animals" and l="Animals can be affected by climate changes". As k entails l, supporting arguments for k also support the particular claim l. But the supporting arguments for l do not necessarily support the more general hypothesis k. Instead, the attacking arguments of l also attack k. Arguments attacking k do not necessarily attack l.

Definition 4 (Consolidating opinions for entailing hypotheses). *If h entails g ($h \xrightarrow{ent} g$) then the number of positive and negative arguments for computing the consolidating opinion $\hat{\omega}_h^{\alpha}$ are:*

$$\hat{e_h} = e_h \tag{10}$$
$$\hat{e}_g = e_h + e_g \tag{11}$$
$$\hat{n}_h = n_h + n_g \tag{12}$$
$$\hat{n}_g = n_g \tag{13}$$

Three properties hold for our consolidation method:

1. less ignorance: based on the consolidated values $\hat{e_h}$ for supporting arguments and $\hat{n}_h$ for attacking arguments.
2. belief consistency: if h entails another hypothesis g, then b_h is expected to be smaller than b_g. That is: $(h \xrightarrow{ent} g) \Rightarrow (\hat{b}_h \leq \hat{b}_g)$.
3. sub-additivity of belief: if $\hat{b}_h + \hat{b}_{\neg h} < 1$.

The technical difficulty is to automatically identify these three relations: similarity, contradiction and entailment. For this task, we used the Excitement Open Platform for Textual Entailment (EOP) [19,24]. From EOP, the Biutee algorithm [28] was preferred due to its ability to interleave knowledge from lexical resources (e.g. WordNet, VerbOcean, Wikipedia) with the language model obtained with supervised learning. Biutee converts the text into the hypothesis via a sequence of transformations. The sequence of transformations is run over the syntactic representation of the text. On the parse tree, different entailment transformations can be applied, like lexical rules (e.g. CO2 $\rightarrow$ gas) or paraphrasing rules (e.g. A affects Y $\leftrightarrow$ Y is affected by X). As these relations are usually insufficient, they are complemented with transformations from a language model. The language model is learned based on a corpus of labeled pairs of text and hypothesis. The logistic-regression is the default algorithm used by Biutee. Given all possible transformations, Biutee applies the Stern et al. search algorithm [29] to find a proof that transforms the text into the hypothesis. The availability of this proof is another reason of using Biutee in our approach.

Algorithm 1 formalises our entailment-based method for computing consolidated opinions. The method starts by training the textual entailment machinery with the available tuples $\langle h, t, l \rangle$ of labeled arguments. Here we exploited the advantage that the arguments are already labeled as pro or cons by their own creators. Based on the labeled pairs, we used the max entropy classification algorithm to generate a language model for climate change arguments. The resulted model contains linguistic patterns in the climate change corpus for entailment and contradiction between each hypothesis h and its supporting and attacking arguments t.

Our trick was to use this learned model to compute now the entailment relations l between pairs of hypotheses $\langle h_1, h_2, l \rangle$ instead of a pair of hypothesis and one of its arguments $\langle h, t, l \rangle$. Hence, we fed Biutee (line 11) with: i) two hypotheses h and g, ii) the model for the climate change corpus, and iii) lexical knowledge bases like WordNet or VerbOcean (see Algorithm 1). Biutee will interleave domain-specific knowledge (encapsulated in the $model$) and domain-independent knowledge (i.e. WordNet, VerbOcean) to search for contradictory or entailment relations between h and g. If a contradictory relation is found, then the parameters $\hat{e}_h$ and $\hat{n}_h$ are computed based on Equations (8) and (9). If an entailment relation is found between h and g, we check if the relation is symmetric (i.e. g entails h too). In this case (line 14), we consider the two hypotheses are semantically similar and equations (6) and (7) are applied. Otherwise, we apply equations (10), (11), (12) and (13). Note that the same hypothesis can be in various relations with other hypotheses at the same time: contradiction (line 9), entailment (lines 18-19), or similarity (lines 14-16).

6. Opinion aggregation and consolidation on the climate chance

This section applies opinion aggregation and opinion consolidation on the climate change corpus. We start by ranking the topics in the climate change corpus based on degree of belief, disbelief, or ignorance. Then we identify similar topics and we consider all their arguments in order to make a more clear picture on the ongoing debates.

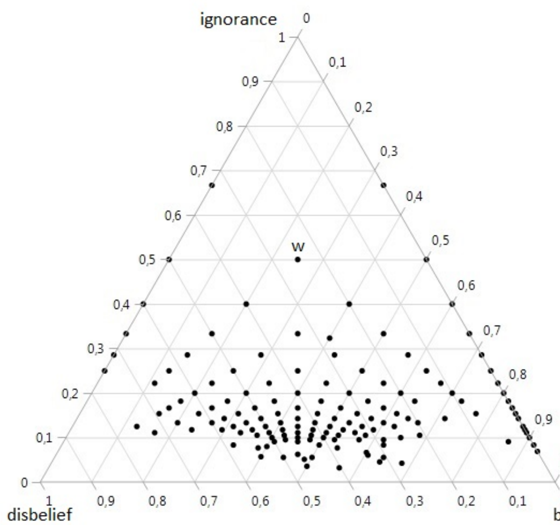
Algorithm 1: Consolidating opinions with textual entailment.

Input: α, corpus of hypotheses and labeled arguments $\langle h, t, l \rangle$
Input: kb, lexical knowledge bases (e.g. WordNet, VerbOcean)
Output: $\hat{\omega}_h^\alpha$, consolidated opinion for each h in α
for $\langle h, t, l \rangle \in \alpha$ **do**
    $model \leftarrow trainBiutee(h, t, l)$
end for
for $h \in \mathcal{H}^\alpha$ **do**
    $e_h \leftarrow |\mathcal{A}_h^+|, n_h \leftarrow |\mathcal{A}_h^-|$
    for $g \in \mathcal{H}^\alpha \setminus \{h\}$ **do**
        $e_g \leftarrow |\mathcal{A}_g^+|, n_g \leftarrow |\mathcal{A}_g^-|$
        $rel \leftarrow BiuteeEntail(h, g, model, kb)$
        if $rel \equiv \neg$ **then**
            $\hat{e}_h = \hat{n}_g \leftarrow e_h + n_g$
            $\hat{n}_h = \hat{e}_g \leftarrow n_h + e_g$
        end if
        if $rel \equiv \xrightarrow{ent}$ **then**
            if $BiuteeEntail(g, h, model, kb) \equiv \xrightarrow{ent}$ **then**
                $\hat{e}_h = \hat{e}_g \leftarrow e_h + e_g$
                $\hat{n}_h = \hat{e}_g \leftarrow n_h + n_g$
            else
                $\hat{e}_g \leftarrow e_h + e_g$
                $\hat{n}_h \leftarrow n_h + n_g$
            end if
        end if
        $\hat{\omega}_h^\alpha \leftarrow computeConsolidatedOpinion(\hat{e}_h, \hat{n}_h, a^\alpha)$
    end for
    return $\hat{\omega}_h^\alpha$
end for

6.1. Opinion aggregation

The voting based method based on subjective logic applied on the climate change corpus provides insights regarding Q_1: *Are the arguers within a community apriori prone to accept or reject a hypothesis?* In the climate change corpus a hypothesis has on average 4.5 supporting arguments and 3.62 attacking arguments. With $|\mathcal{P}^{cc}| = 943$ positive hypotheses and $|\mathcal{N}^{cc}| = 453$ we have $a^{cc} = 0.67$. On average, the degree of belief is a little larger than disbelief. Hence, members of the communities from which the arguments were collected seem to be prone to accept a given hypothesis.

All hypotheses in climate change corpus are depicted with barycentric coordinates in Fig. 2. Closer to the top are the hypotheses with high ignorance. Neutral opinions are on the median from the top. On the right part are positive opinions, and on the left part are the negative ones. By pressing on each of the opinion point, ARGSENSE provides details on that hypothesis or set of hypotheses.



No. of hypotheses	cc	= 1,793
Positive hypotheses	$\mathcal{P}^{cc}$	= 943
Negative hypotheses	$\mathcal{N}^{cc}$	= 453
Neutral hypotheses	$\mathcal{E}^{cc}$	= 397
Positive arguments	$\mathcal{A}^{+}_{cc}$	= 6,662
Counter arguments	$\mathcal{A}^{-}_{cc}$	= 4,991
Acceptance prone	a^{cc}	= 0.67
Ignorance interval	i	$\in$ [0.04..0.66]
Belief interval	b	$\in$ [0.14..0.93]
Disbelief interval	d	$\in$ [0..0.72]

Fig. 2. Depicting 1,793 hypotheses in the climate change corpus with barycentric coordinates. Each point depicts the set of hypotheses with the same coordinates. For instance, the opinion point $w = \langle 0.25, 0.25, 0.5, a^0 \rangle$ corresponds to 145 neutral hypotheses.

Figure 2 also shows that no vacuous opinions exist in our climate change corpus. The highest degree of ignorance is 0.66, given by hypotheses with only one argument. Still, there are 194 opinions with this high degree of ignorance, representing 11% from the total of 1793 hypotheses. Among them, 88% are pure positive and 12% are purely negative. With 35 positive arguments and 25 counter arguments, the hypothesis with the smallest degree of ignorance is *"Global warming is a natural cycle"*. Note that the second hypothesis with the smallest ignorance is *"Mankind is the main cause of global warming"*, which claims the opposite of h_0. There are 57 purely negative opinions, with the highest disbelief assigned to *"Is there a climate change conspiracy behind global warming and* *global cooling theories?"* (5 negative arguments and 0 positive). Instead, there are 396 purely positive opinions. The purely positive opinion with the highest degree of belief is *"The Kyoto protocol would harm the American economy"*. With 22 supporting arguments and 0 against, it has a belief of 0.91. The percentage of pure opinions (25%) is quite high. No equidistant opinion exists in the corpus. There are 22% neutral opinions. Most of them are supported by one argument and attacked by one argument. There are 53% positive opinions and 25% negative opinions.

Table 2. Answering to Q_2: Which hypotheses are most believed/disbelieved or popular/unpopular in a community?

Most believed hypothesis	e_h	n_h	b_h	d_h	i_h
The Kyoto protocol would harm the American economy.	22	0	0.94	0	0.06
Colonizing the Moon is critical for human survival.	18	0	0.92	0	0.08
Solar shading is a just response to irreversible global warming.	18	0	0.92	0	0.08
Most disbelieved hypothesis	e_h	n_h	b_h	d_h	i_h
People can relax. Global warming is a sham.	3	13	0.17	0.74	0.09
Is cap-and-trade better at reducing emissions?	3	13	0.17	0.74	0.09
Are oil sands bad for climate change?	3	10	0.21	0.69	0.1
Is injecting sulphur dioxide into the atmosphere a good idea?	3	8	0.24	0.64	0.12
Most popular hypothesis	e_h	n_h	b_h	d_h	i_h
Global warming is a natural cycle.	35	25	0.57	0.41	0.02
Mankind is the main cause of global warming.	28	26	0.5	0.47	0.03
Should we actually have a purge?	19	18	0.49	0.47	0.04
Manmade global climate change is real and a threat.	16	18	0.45	0.51	0.04
Most unpopular hypothesis	e_h	n_h	b_h	d_h	i_h
Global warming causes earthquakes.	1	0	0.4	0	0.6
The sun causes global warming.	1	0	0.4	0	0.6
All natural disasters are related to global warming.	1	0	0.4	0	0.6

By ranking the topics based on belief, disbelief, and popularity, ARGSENSE supports Q_2 as depicted in Table 2. Note that the most believed hypotheses are all pure opinions, given by no counter arguments for them ($d_h = 0$). Differently, none of the four most disbelieved topics is pure, given by the existence of pro arguments for them ($b_h > 0.17$). With 35 pro and 25 con arguments, the most popular topic has an ignorance of 0.02. Note that the first two most popular hypotheses belong to the same topic - real cause of global warming - but they claim opposite statements. From the ignorance value perspective, this result is consistent with the interpretation that the cause of global warming has been the most interesting topic for the arguers. From the psychological perspective, the result might indicate that the way in which the topic of the debate is formulated does influence the output of the debate: in both cases people seem to rather support the claim in the topic. The "natural cycle" hypothesis h is supported with a belief of $b_h = 0.57$, while the "human cause" hypothesis g is also supported with $b_g = 0.5$, even if g claims the opposite thing as h. One might expect that believing in h means a disbelief in g or a belief in g would be consistent with a disbelief in h. Given the nature of each debate, various factors may contribute to the above belief inconsistency at the community level.

From the opposite perspective, bottom part of Table 2 presents the hypotheses that people seem not to be interested in. There are 409 debates with only one positive argument and no attacking argument. Comparing the results of the most popular with the most ignored topics indicates that popular hypotheses are more general. Hidden variables, like time of issuing the debate, might be a cause of this lack of interest[4].

6.2. Consolidating opinions across hypotheses

The language model of climate change corpus was obtained by training the Biutee with on the cc corpus with 6,662 entailment pairs and 4,991 non-entailment. The entailment pairs correspond to pairs of hypotheses with supporting arguments, while non-entailment correspond to pairs of hypotheses with attacking arguments. We used the max entropy classification algorithm to generate the language model. WordNet [22] and VerbOcean [8] were used as external knowledge resources. From Wordnet, the following relations were considered during the search process for a useful transformation: synonym, derivationally related, hypernym, instance hypernym, member holonym, part holonym, substance meronym, entailment. Only the first sense was used for a depth limit of 2 in the Wordnet taxonomy.

We illustrate one entailment proof computed with the Biutee method. The proof lists the transformations applied on the parse tree of the hypothesis h_1 (see Table 3). Knowledge from Wordnet and VerbOcean was used, but also the learned model from the training examples. The resulted sentence *Climate change can affect animals* is compared with the hypothesis h_2, and the entailment label is applied. Note that the confidence for this decision is 0.61. On the entire corpus we have a confidence of 0.65. This value is similar to cases in which texts are manually labeled by two human annotators. For instance, Kwon et al. have reported an inter-annotator agreement of 0.62 [16].

Having entailment/non-etailment relations computed for the debate topics, we can now apply our consolidation method to aggregate arguments of similar topics, as exemplified in the next subsection. To illustrate the consolidation method in case of entailment consider the pair of hypotheses *h="Mankind is the main cause of global warming"* and *g="Global warming is real"*. h non-explicitly assumes that global warming is real and questions only its cause. Note that the assumption of h is the claim in g. Therefore we consider h entails g. In our corpus, we found that h is supported by 28 arguments and attacked by 26, while g is supported by 4 arguments and attacked by 4. That is $b_h = 0.5$ and $b_g = 0.46$. Because $h \xrightarrow{ent} g$ and $b_h > b_g$, the consistency property of belief does not hold for h and g. Instead, after applying the consolidation method in case of entailment, the consolidated belief becomes consistent and also the ignorance decreases. Based on equations (10) and (11), $\hat{e}_g = e_h + e_g = 28 + 4 = 32$ and $\hat{n}_h = n_h + n_g = 26 + 4 = 30$, while $\hat{e}_h = e_h = 28$ and $\hat{n}_g = n_g = 4$. The consolidated opinion for h is $\hat{\omega}_h^{cc} = \langle 0.47, 0.5, 0.03, 0.67\rangle$ and for g is $\hat{\omega}_g^{cc} = \langle 0.85, 0.11, 0.04, 0.67\rangle$. As the consolidated belief $\hat{b}_h < \hat{b}_g$, the belief consistency property holds between the entailing hypothesis h and g.

---

[4] For instance, the Marrakesh Climate Change Conference - November 2016 has not managed to trigger many debates, as all the debates site were invaded by debates related to the USA elections.

Table 3. Proof of entailment between two debate topics h_1 and h_2.

h_1: Climate-induced changes are likely to cause a series of cascading effects involving many species of plants and animals.

h_2 : Animals can be affected by climate changes.

Proof: Entailment, score = 0.6172

Substitute: ”climate-induced”(JJ) to: ”climate”(NN) (Multi-Word, remove words)

Substitute: ”involve”(VBG) to: ”affect”(VERB) with confidence:0.5 (Verb ocean)

Substitute: ”animal”(NNS) to: ”animals”(NNS)

Syntactic extraction rule: ”Relative Clause - Extract reduced relative clause to independent sentence”

The part-of-sentence (bag of words) is: ”affect effects ”

Insert <(5) ”by”, IN, prep> under <(12) ”affect”, VERB>(”by” costs -5.4143)

Insert <(2) ”can”, MD, aux> under <(12) ”affect”, VERB>(”can” costs -7.5089)

Existing-Word-Insert <(7) ”change”, NNS, pobj> under <(5) ”by”, IN> (”change” costs -8.2133)

Existing-Word-Insert <(6) ”climate”, NN, nn> under <(7) ”change”, NNS> (”climate” costs -10.7516)

Syntactic substitution rule: ”Coordination - Delete a verbal nominal or adjectival conjunct”

The part-of-sentence (bag of words) is: ”and Animals plants ”

Syntactic substitution rule: ”Possessive - Substitute an ”of-construction” with a nn modifier”

The part-of-sentence (bag of words) is: ”Animals of species”

Move node <(18) ”animals”> to <(12) ”affect”> with relation ’nsubjpass’(costs -3.00, change context)

Climate changes are likely to cause a series of cascading effects involving many species of plants and animals.

Climate changes are likely to cause a series of cascading effects *affecting* many species of plants and animals.

Climate changes are likely to cause a series of cascading effects affecting many species of plants and *animal*.

Climate change *can by affecting* many species of plants and animals.

Climate change *can affecting* many species of animals.

Climate change *can affect* animals.

To illustrate the sub-additive property of consolidated belief, consider the contradictory hypothesis h=*“Mankind is the main cause of global warming.“* and k=*”Global warming is a natural cycle”*. Semantically, h is opposite of k. In the climate change corpus, the non-additive property does not hold for h and k ($b_h = 0.5$, $b_k = 0.57$). Instead, after applying the accrual of arguments in case of the contradictory relation, the belief becomes consistent and also the ignorance decreases. Based on equations (8) and (9), $\hat{e}_h = \hat{n}_k = e_h + n_k = 28 + 25 = 53$ and $\hat{n}_h = \hat{e}_g = n_h + e_k = 26 + 35 = 61$. the consolidated opinion for h is $\hat{\omega}_h^{cc} = \langle 0.46, 0.53, 0.01, 0.67 \rangle$ and for g is $\hat{\omega}_g^{cc} = \langle 0.53, 0.47, 0.01, 0.67 \rangle$. As for the consolidated belief $\hat{b}_h + \hat{b}_g = 0.46 + 0.53 = 0.99 < 1$, then the belief consistency property holds.

Opinion consolidation was used here as a general method for enriching the set of arguments for a given hypothesis, thus diminishing its ignorance.

7. Argumentative-text characteristics

Argumentative text characteristics are used by social scientists, policymakers or science communicators to better understand the communities of arguers and to design effective ways to communicate science or policies to target audience. ARGSENSE is able to analyse differences between linguistic patterns used in pro and counter arguments, to assess the correlation between the popularity of a debate with how the debate topic was posted, or to compute the readability of pro and counter arguments. We exemplify the lexical analysis of ARGSENSE by answering questions Q_3 to Q_7 on the climate change corpus.

Q_3: *Do the pro arguments have a different lexicon than the counter arguments?* Different lexicon might be an indicator to the social scientist that one party of the debate is sensible to different aspects as the other party.

To detect possible differences, we searched for the most frequent words in pro and cons arguments. For instance, if we denote by f_{20}^+ and f_{20}^- the sets of the 20 most frequent words in the set of pros and cons, we obtained $f_{20}^+ \backslash f_{20}^- = \{emissions,\ greenhouse,\ water\}$ and $f_{20}^- \setminus f_{20}^+ = \{ice,\ increase,\ opponent\}$. These results suggest that proponents of climate change are concerned with emissions and greenhouse, while the opponents rise arguments related to ice. Interestingly, the ice related counter-argument is a common misconception related to climate change [11]. ARGSENSE was able to signal that this misconception is also spread over the debate sites.

Q_4: *When does a debate get more pros than cons, when formulated as a statement or as a question?* We are interested whether posting a hypothesis in affirmative or interrogative form could modify its chances to accumulate more arguments on one side or another. In the climate change corpus, 382 affirmative hypotheses received more pro arguments and 83 of them got more counter arguments. For interrogative topics, 561 got more pros and 370 more counter arguments. Fisher's exact test indicates a very strong statistical correlation ($p < 0.0001$) between the type of hypothesis and its chances to get more positive than negative arguments. The odds ratio value for the given example is 3.04, showing that the chances to have a winner are more than three times higher when the sentence is in affirmative than in interrogative.

Q_5: *Which is the readability of the arguments conveyed in a debate?* This provides an insight on the writing and reading comprehension skills of a community of arguers. An expert in science communication uses such readability indexes to adapt its arguments to the target audience. The science communicator should balance between simplifying the text and retaining technical details.

ARGSENSE is able to compare pro and counter arguments based on six readability indexes (Table 4). Coleman Liau and Automated Readability indexes rely on counting characters, words and sentences. The other indexes consider number of syllables and complex words. For more about readability formulas the reader is referred to [33]. No matter the readability index, the values for the positive and negative arguments are extremely similar. That is, no side uses more complex words than the other. The science communicator has to design ways to convey scientific results with the same readability indexes as the target audience or community or arguers [5].

Table 4. Readability indexes for pro ($\mathcal{A}_{cc}^+$) and against ($\mathcal{A}_{cc}^-$) arguments.

Readability index	Flesch Kincaid Reading Ease	Flesch Kincaid Grade Level	Gunning Fog Score	SMOG	Coleman Liau	Automated Readability
$\mathcal{A}_{cc}^+$	58.40	8.73	11.21	8.73	11.80	8.01
$\mathcal{A}_{cc}^-$	59.77	8.58	11.17	8.62	11.45	7.77

Q_6: *Is there a correlation between the length of a hypothesis and the number of its arguments?* We investigated whether heuristics like "the shorter the hypotheses, the more arguments" can be used by a debater to decide how to formulate the debate topic. The average number of words in $\mathcal{H}^{dbp}$ is 8.67. The correlation between the length of the hypothesis and the ignorance on it is -0,01. Similarly, the average number of words in $\mathcal{H}^{deb}$ is 9.67. The correlation between the length of the hypothesis and the ignorance on it is 0.12. Based on these two low values, we can conclude that for both communities deb and dbp, the length of the hypothesis does not influence the number of arguments.

Q_7: *Does a query trigger more interest than a statement?* We analysed if a debate topic posted as *Will the planet adapt to global warming?* will attract more arguments than the version *The planet will adapt to global warming*. We evaluated the ignorance level for each hypothesis in the affirmative and interrogative form. Fig. 3 gives the cumulated percentages of hypotheses in affirmative and interrogative format for a given ignorance threshold. For example, if we specify an ignorance threshold 0.1, there are 12.75% of interrogative hypotheses, but only 7.40% of affirmative hypotheses. The percentage of interrogative hypotheses is always higher than its affirmative counterpart, which makes us believe interrogative hypotheses have higher chances to get more intense discussion.

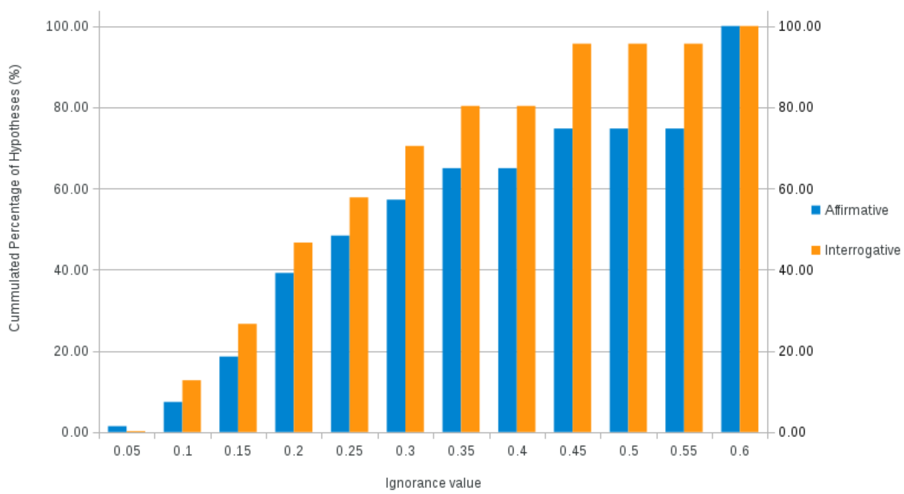


Fig. 3. Interrogative hypotheses attract more arguments than claiming hypotheses.

To summarise, the findings of lexical analyses related to questions Q_3 to Q_7 are: 1) proponents of climate change are more interested in: *emissions*, *greenhouse* and *water*, while the opponents convey more counter-arguments regarding *ice*; 2) affirmative hypotheses have three times higher chances to win compared to interrogative hypotheses ($p < 0.0001$); 3) pros and cons have the same readability values; 4) the length of the hypothesis does not influence the number of arguments for it; 5) interrogative hypotheses have higher chances to attract more arguments than in affirmative form.

8. Conclusions

We analysed arguments conveyed in public arena related to climate change. The four contributions of this research are: 1) the argumentative climate change corpus, 2) the

ARGSENSE social software for understanding public opinion on climate change, 3) the computational methods for aggregating and consolidating arguments, and 4) the lexical analysis of climate change arguments.

First, the climate change corpus is, to our knowledge, the largest corpus of labeled arguments on climate change.

Second, ARGSENSE makes sense of a set of arguments on climate change and aggregates them into usable results for policy making and climate science communication. ARGSENSE employs a voting method based on Subjective Logic to rank the debate topics according to their belief, disbelief or popularity within a community of arguers. Thus, ARGSENSE is in line with the recent trend to support scientific discovery [25] and to enhance the climate science cyber-infrastructure from *useful to usable* decision support tools [3].

Third, social sciences need to extend their instruments to be able to measure worldview on debate issues. Our method enhances the capabilities of social science to measure public support, disagreement or ignorance. It employs textual entailment to find similarity, contradiction or entailment between natural language arguments.

Fourth, we investigated in which way those interested in promoting public engagement need to pay attention towards linguistic aspects of communicating climate science. Our lexical analysis performed on arguments conveyed by people found that: 1) conveyors of pro arguments are interested in: *emissions*, *greenhouse* and *water*, while conveyors of con arguments in *ice*; 2) sequences in positive arguments do not overlap with sequences in negative arguments; 3) affirmative hypothesis have three times higher chances to win compared to interrogative hypotheses ($p < 0.0001$); 4) both pros and cons have the same readability; 5) no influence has been found between the length of the hypothesis and the number of arguments for it; 6) interrogative hypotheses have higher chances to attract more arguments than in affirmative form. Such lexical findings can be used by stakeholders to figure out how to communicate their point of view more effectively to the public.

One limitation of our approach regards the difficulty accurately assess the semantic similarity between topics. Even if the confidence in computing this semantic similarity is in the same range with the confidence of human annotators, it remains quite low: 0.65. This value is an average computed on the entire corpus. To tackle this limitation, one could explore the following two main directions. One direction is to apply the opinion consolidation method only to pairs of hypotheses for which the entailment/nonentailment is computed with high confidence. Another direction is to fine tune the parameters of the Biutee method related to: i) learning algorithm, ii) search process, or iii) external knowledge bases. We envisage the following possibilities. First, there are parameters of the learning algorithm used to build the language model. We run experiments only the max entropy classification algorithm. Second, there are parameters of the search step used to build the proof for entailment or nonentailment. We used all relations from the Wordnet and a depth limit of 2 in the Wordnet taxonomy. Third, one can add domain specific knowledge bases. That is, instead of relying only on general lexical resources (Wordnet, VerbOcean, Wikipedia) one can convert domain ontologies (for climate change in our case) to a rule-based format required by the Biutee method.

Another research line to be pursued is detecting repetitive arguments, either in verbatim copies or in semantically equivalent rephrasing. Here, we considered only the number of arguments and semantic similarity between topics. To overcome this, multiple dimensions can be considered, like argument provenance or time of issue. Such direction can be integrated into the larger context of research on fake arguments, collusion of argument proponents, or on how arguments propagate in public arena or in specific communities.

Acknowledgments. This work was funded by the EEA and the Norwegian Financial Mechanisms 2009-2014 under the contract 73264/02.09.2016 "Increasing understanding on climate change through public discourse analyse and stakeholders modeling" between Norwegian University of Science and Technology, Trondheim, Norway and Technical University of Cluj-Napoca, Romania.

References

1. Facilitating climate change adaptation through communication: Insights from the development of a visualization tool. Energy Research & Social Science 10, 57 – 61 (2015)
2. Batanović, V., Bojić, D.: Using part-of-speech tags as deep-syntax indicators in determining short-text semantic similarity. Computer Science and Information Systems 12(1), 1–31 (2015)
3. Biehl, L.L., Zhao, L., Song, C.X., Panza, C.G.: Cyberinfrastructure for the collaborative development of u2u decision support tools. Climate Risk Management 15, 90–108 (2017)
4. Boussalis, C., Coan, T.G.: Text-mining the signals of climate change doubt. Global Environmental Change 36, 89–100 (2016)
5. de Bruin, W.B., Morgan, M.G.: Reflections on an interdisciplinary collaboration to inform public understanding of climate change, mitigation, and impacts. Proceedings of the National Academy of Sciences p. 201803726 (2019)
6. Chalaguine, L.A., Schulz, C.: Assessing convincingness of arguments in online debates with limited number of features. In: Proceedings of the SR Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics. pp. 75–83 (2017)
7. Chaniotakis, E., Antoniou, C., Pereira, F.: Mapping social media for transportation studies. IEEE Intelligent Systems 31(6), 64–70 (2016)
8. Chklovski, T., Pantel, P.: Verbocean: Mining the web for fine-grained semantic verb relations. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (2004)
9. Groza, A., Ozturk, P., Slavescu, R.R., Marginean, A., Prasath, R.: Analysing climate change arguments using subjective logic. In: 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP). pp. 37–44. IEEE (2018)
10. Ha-Duong, M.: Hierarchical fusion of expert opinions in the transferable belief model, application to climate sensitivity. International Journal of Approximate Reasoning 49(3), 555 – 574 (2008)
11. Hamilton, L.C.: Did the arctic ice recover? demographics of true and false climate facts. Weather, Climate, and Society 4(4), 236–249 (2012)
12. Huang, Y., Huang, G., Hu, Z., Maqsood, I., Chakma, A.: Development of an expert system for tackling the public's perception to climate-change impacts on petroleum industry. Expert Systems with Applications 29(4), 817 – 829 (2005), http://www.sciencedirect.com/science/article/pii/S0957417405001077
13. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(03), 279–311 (2001)
14. Kirilenko, A.P., Stepchenkova, S.O.: Public microblogging on climate change: One year of twitter worldwide. Global Environmental Change 26, 171–182 (2014)
15. Kraemer, H.C.: Kappa coefficient. Wiley StatsRef: Statistics Reference Online (1982)
16. Kwon, N., Zhou, L., Hovy, E., Shulman, S.W.: Identifying and classifying subjective claims. In: Proceedings of the 8th Int. Conf. on Digital government research: bridging disciplines & domains. pp. 76–81. Digital Government Society of North America (2007)

17. Letia, I.A., Groza, A.: Arguing with Justifications between Collaborating Agents, pp. 102–116. Springer Berlin Heidelberg, Berlin, Heidelberg (2012), http://dx.doi.org/10.1007/978-3-642-33152-7_7
18. Lioma, C., Larsen, B., Schütze, H., Ingwersen, P.: A subjective logic formalisation of the principle of polyrepresentation for information needs. In: Proceedings of the third symposium on Information interaction in context. pp. 125–134. ACM (2010)
19. Magnini, B., Zanoli, R., Dagan, I., Eichler, K., Neumann, G., Noh, T.G., Pado, S., Stern, A., Levy, O.: The excitement open platform for textual inferences. In: Association for Computational Linguistics (System Demonstrations). pp. 43–48 (2014)
20. Mamauag, S.S., Alino, P.M., Martinez, R.J.S., Muallil, R.N., Doctor, M.V.A., Dizon, E.C., Geronimo, R.C., Panga, F.M., Cabral, R.B.: A framework for vulnerability assessment of coastal fisheries ecosystems to climate change - tool for understanding resilience of fisheries (VA-TURF). Fisheries Research 147, 381 – 393 (2013), //www.sciencedirect.com/science/article/pii/S0165783613001719
21. Mayer, L.A., Loa, K., Cwik, B., Tuana, N., Keller, K., Gonnerman, C., Parker, A.M., Lempert, R.J.: Understanding scientists' computational modeling decisions about climate risk management strategies using values-informed mental models. Global Environmental Change 42, 107 – 116 (2017), http://www.sciencedirect.com/science/article/pii/S0959378016306197
22. Miller, G.: WordNet: An electronic lexical database. MIT press (1998)
23. Mohammad, S.M., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: A dataset for detecting stance in tweets. In: Proceedings of 10th edition of the the Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia (2016)
24. Padó, S., Noh, T.G., Stern, A., Wang, R., Zanoli, R.: Design and realization of a modular architecture for textual entailment. Natural Language Engineering 21(02), 167–200 (2015)
25. Pankratius, V., Li, J., Gowanlock, M., Blair, D.M., Rude, C., Herring, T., Lind, F., Erickson, P.J., Lonsdale, C.: Computer-aided discovery: Toward scientific insight generation with machine support. IEEE Intelligent Systems 31(4), 3–10 (July 2016)
26. Pührer, J.: Argueapply: A mobile app for argumentation. In: International Conference on Logic Programming and Nonmonotonic Reasoning. pp. 250–262. Springer (2017)
27. Qin, X.S., Huang, G.H., Chakma, A., Nie, X., Lin, Q.: A MCDM-based expert system for climate-change impact assessment and adaptation planning–a case study for the Georgia Basin, Canada. Expert Systems with Applications 34(3), 2164–2179 (2008)
28. Stern, A., Dagan, I.: The BIUTEE research platform for transformation-based textual entailment recognition. LiLT (Linguistic Issues in Language Technology) 9 (2014)
29. Stern, A., Stern, R., Dagan, I., Felner, A.: Efficient search for transformation-based inference. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. pp. 283–291. Association for Computational Linguistics (2012)
30. Svrcek, M., Kompan, M., Bielikova, M.: Towards understandable personalized recommendations: Hybrid explanations. Computer Science and Information Systems 16(1), 179–203 (2019)
31. Walton, D., Reed, C., Macagno, F.: Argumentation schemes. Cambridge Univ. Press (2008)
32. Washington, H.: Climate change denial: Heads in the sand. Routledge (2013)
33. Weiss, K.D., Vargas, C.R., Ho, O.A., Chuang, D.J., Weiss, J., Lee, B.T.: Readability analysis of online resources related to lung cancer. Journal of Surgical Research 206(1), 90–97 (2016)
34. Zhang, H., Alim, M.A., Li, X., Thai, M.T., Nguyen, H.T.: Misinformation in online social networks: Detect them all with a limited budget. ACM Trans. Inf. Syst. 34(3), 18:1–18:24 (Apr 2016), http://doi.acm.org/10.1145/2885494

Adrian Groza is assoc prof. at the Dept. of Computer Science at Technical University of Cluj-Napoca, Romania. His current research regards knowledge representation, argumentative agents, explainable AI.

Pinar Ozturk is assoc. prof. at the Dept. of Computer and Information Science at Norwegian University of Science and Technology (NTNU). Before joining NTNU, Ozturk was a senior research scientist at SINTEF. Ozturk's research area is artificial intelligence methods for decision making.

Radu Razvan Slavescu got his PhD from the Technical University of Cluj-Napoca, Romania with a thesis focused on trust modeling and its application in multiagent systems. Now, Radu is with the Intelligent Systems group at the Computer Science Department of the same University. His domains of interest include Artificial Intelligence, Multiagent Systems and trust modeling, and Natural Language Processing, in which he has authored 40+ papers.

Anca Marginean got her PhD fromthe Techical University of Cluj-Napoca, Romania in 2013. Her interests regard natural language processing and Semantic Web.

Received: October 11, 2018; Accepted: September 17, 2019.