Argument Invention from First Principles

Anonymous ACL submission

Abstract

Competitive debaters often find themselves facing a challenging task – how to debate a topic they know very little about, with only minutes to prepare, and without access to books or the Internet? What they often do is rely on "first principles", commonplace arguments which are relevant to many topics, and which they have refined in past debates. In this work we aim to explicitly define a taxonomy of such principled recurring arguments, and, given a controversial topic, to automatically identify which of these arguments are relevant to the topic. As far as we know, this is the first time that this approach to argument invention is formalized and made explicit in the context of NLP. The main goal of this work is to show that it is possible to define such a taxonomy. While the taxonomy suggested here should be thought of as a "first attempt" it is nonetheless coherent, covers well the relevant topics and coincides with what professional debaters actually argue in their speeches, and facilitates automatic argument invention for new topics.

1 Introduction

In his treatise *De Inventione* Cicero defines the five canons of classical rhetoric as: inventio (invention), dispositio (arrangement), elocutio (style), memoria (memory), and pronuntiatio (delivery). The first of these, *Inventio*, is defined as a systematic search for arguments (Glenn et al., 2008), with applicability to wide variety of situations often seen as a desired property (Lauer, 2004). This problem has been referred to in the context of NLP as the task of *Argument Invention* (Walton and Gordon, 2012, 2017), but did not receive abundant attention.

One natural way people go through the process of *inventio* is to look for arguments in relevant texts, or, if they are familiar with the topic, rely on their knowledge and *memoria* for doing

so. This is reminiscent of the way Argument Mining algorithms operate (see e.g. Lippi and Torroni, 2015, 2016). However, we often find ourselves in situations where that is not possible. For example, when arguing politics over lunch, we might find ourselves backed into a corner, facing a topic with which we are not very familiar, but somehow nonetheless need to justify or oppose. This often happens because we were initially arguing some *principle*, and now we need to apply it to an unfamiliar example.

Professional debaters often face this problem. Presented with an unfamiliar topic they need to quickly come up with relevant arguments. The main technique for doing so is called arguing from "first principles" – relying on a "bank" of principled arguments, which are relevant to a wide variety of topics[1].

A common example is the Black market argument: banning a product or a service may lead to the creation of a black market, which, in turn makes products or services obtained therein less safe, leads to exploitation, attracts criminal elements, and so on. Hence, even if we agree that something should not be encouraged, it is advisable to have it legal and regulated.

This kind of argument can be made, *mutatis mutandis*, when debating quite different topics, such as legalizing organ trade, or banning pornography. However, it is not always relevant when debating whether to legalize or ban something; e.g., when debating whether to legalize polygamy or ban breastfeeding in public, the black market argument seems less appropriate.

Here, we aim to create a knowledge base of such principled arguments, which, when given a topic for debate or a critical essay, would readily yield the relevant ones. We do this in a framework of

---

[1]This is reminiscent of, yet different from, the Aristotelian use of this term, which refers to self-evident propositions.

certain types of *motions* (section 3). Specifically, we define several commonplace themes which are likely to be a point of contention – that is, where arguments of opposing stance can be made around this theme[2]. We show that for most motions there exist relevant arguments within the suggested knowledge base, and that they can be identified automatically with reasonable precision and recall. Moreover, we show that professional human debaters often allude to such arguments when they debate.

2 Related work

Previous computational work on argument invention was mainly done within the field of argument mining, where – as the name implies – the focus is on identifying arguments within a given text. Most works (e.g. Stab and Gurevych, 2014; Palau and Moens, 2009; Eger et al., 2017) assume that a relevant text is provided, while some include the task of extracting such text from a large open-domain corpus (e.g. Levy et al., 2014; Rinott et al., 2015; Shnarch et al., 2018; Al-Khatib et al., 2016; Levy et al., 2018). The work here complements such techniques by providing a dataset of arguments whose manual construction facilitates automatic retrieval for topics of interest and ensures quality, validity, style and so on.

A somewhat similar approach is suggested in work of Walton and Gordon (2017, 2018), where arguments are constructed from a database (Reed and Rowe, 2004) of smaller argumentative building blocks. However, these building blocks are topic-specific and can not readily provide arguments for topics not in the database.

The attempt to categorize arguments by looking for commonalities dates back to ancient times, such as Aristotle's list of 28 topoi (Aristotle and Kennedy, 1991). Modern works, such as Perelman (1971); Walton et al. (2008); Walton (2013), expanded on these ideas, similarly focusing on how an argument's conclusion is inferred from its premises. Unlike these efforts, the taxonomy suggested here is of recurring principled semantic themes. That is, arguments which in this work would be categorized as belonging to a specific argument theme could be of various topoi and follow different argumentation schemes.

In modern competitive debating the notion of commonalities between topics is prevalent due to the advantages they serve in overcoming knowledge barriers and in speeding up argument generation[3]. Armed with limited facts on the topic, the task of locating recurring patterns in order to argue the motion abstractly is composed of understanding what are the fundamental 'clashes' in the debate (cf. Sonnreich (2012), "debating from first principles"), similar to the taxonomy herein.

Our approach bears similarities to work in social sciences that attempts to describe different types of information *framing*, usually in the context of the news media (e.g. Semetko and Valkenburg, 2000). Recurring themes, like *Fairness and equality* or *Crime and punishment*, can be identified in the way the news media frames a certain policy issue or event (Card et al., 2015). de Vreese (2005) differentiates between specific and generic frames, characterizing the latter as those that can be applied to a wide range of events and contexts. Similarly, our work aims to categorize commonplace themes and identify their relevance (at a considerably larger scale), in the context of framing a topic that is subject to debate.

Our work also has some commonalities with psychological research on ideology (e.g. Altemeyer, 1981; Sidanius and Pratto, 2001; Jost et al., 2003). For example, Everett (2013) lists a 12-item scale to assess conservative ideology - of these, some map to our taxonomy (e.g. Welfare), while others are too specific. Moreover, conservatism in itself gives rise to one class of recurring arguments in our work.

3 Definitions

In the context of parliamentary debate, a *motion* is a proposal that is to be deliberated by two sides (government and opposition). Here we formally define a motion as a pair *(action, topic)*, where *topic* is a Wikipedia title (or a redirect to one), and *action* is a term coming from a closed set of allowed actions (Appendix A), and describes the government's proposal w.r.t. *topic*. For example, the motion *(ban, smoking)* should be interpreted as the government suggesting to ban smoking, and can be explicitly phrased as "we should ban smoking"[4]. Note that not all combinations of *action* and *topic* make for a good motion – the implied proposal should be worth deliberating; one for which

---

[2]In the context of debates, these are called "clashes".

[3]See for example: http://www.wudc2018.mx/training-materials.html

[4]In the context of competitive debate, the phrasing would usually be "This House would ban smoking".

reasonable arguments can be made by either side.

One often discerns between *policy motions* where the government proposes a concrete policy, and *analysis motions* in which the government declares its opinion on the *topic*. For the sake of simplicity and brevity (and with some abuse of notation) we will ignore this distinction. In particular, one of the allowed "actions" is *brings more harm than good*, which would usually be considered as indicating an analysis motion.

We define a *Class of Principled Arguments*, or *CoPA*, as a set of arguments revolving around a principled recurring theme (we define the name of the CoPA as this theme), alongside a set of motions to which this theme is relevant. Formally, a CoPA is a pair $c = (A, M)$, where A is a set of arguments, and M is a set of motions, s.t. every $a \in A$ is an argument that can plausibly be made when deliberating any motion $m \in M$. For every $a \in A$ we say that a is an argument in c, and similarly that every $m \in M$ is a motion in c, and that m and c *match*.

In this work we focus on modelling debate *clashes*, and hence are interested in A's which contain arguments of opposing stances towards the class's theme. For the sake of simplicity we consider only A's of size 2, and only simple arguments, which are essentially just claims or premises (appendix B). Note that this is indeed a simplification, and that for many CoPAs several distinct arguments can be naturally included in A (cf. Section 7).

The pair of claims may directly contradict each other, denoting a disagreement about facts. But by and large we tried to select pairs of claims that people tend to agree with[5], but would assign different valuation depending on their point of view (cf. Kock, 2009). For example, one of the CoPAs we define is *Clean energy*, with A = {"Humanity must embrace clean energy in order to fight climate change", "Ecological concerns add further strain on the economy"}, and M including motions such as *(subsidize, renewable energy)* and *(fight, global warming)*. Most people would agree that on the one hand climate change is a problem, and on the other that moving towards clean energy will be expensive. When debating motions where clean energy is a relevant theme the two sides are likely to agree that both claims have some merit, yet disagree on which supersedes the other. The valuation might very well depend on their subjective viewpoints, but also on the specific motion.

4 Data

4.1 Initial data

The definition above is a functional one, oriented towards facilitating labeling motion-class matches, so some care is required in CoPA construction. A pair of claims with one saying that the policy will not work, and the other that it will, defines a CoPA that essentially covers all policy motions. Conversely, very particular claims will yield a class for which relevant motions are hard to come by.

The set discussed in this work was defined based on the following guidelines:

1. One is able to define two concise claims of opposing stance towards the CoPA's theme.

2. One can think of at least three motions (not necessarily from the initial set) which would belong to the CoPA, and are not overly similar to one another.

Our motivation for requirement 1 was to model "clashes", the recurring themes in debates which are points of contention, since our main use-case is, given a motion, to suggest argumentative text with a clear stance towards the motion. Other use cases may relax this requirement, according to their goals. Requirement 2 ensures that the CoPA indeed captures a recurring theme, rather than a specific one.

Two annotators were presented with these guidelines and an initial list of 100 motions to make the task more concrete (see supplementary materials). They authored a list of about 60 CoPAs which was manually curated by two of the authors to a more concise, final list of 37 CoPAs, to avoid redundancy and ease the following labeling task.

Appendix B lists these 37 CoPAs and the claims therein. They are quite varied – some revolve around public policy (e.g. *Environment, Public health*), others on basic rights and freedoms (e.g. *Right to privacy, Freedom of religion*), some on the effect of a policy (e.g. *Black market, Greater good*), while yet others are very general (*Fixable, Conservatism*, and *Framework*)[6].

---

[5]This is somewhat similar to what Perelman (1971) call "Values" and also close to what Aristotle calls *Endoxa* (Aristotle and Kennedy, 1991).

[6]When analyzing our data, it will sometimes be interesting to omit these three classes, to ascertain they do not significantly skew the results.

Next, the same two annotators annotated all 100 motions for membership in the suggested CoPAs. In total, 92 motions were matched to at least one CoPA, and on average each motion was matched to 2.03 CoPAs. In our dataset, the greatest number of CoPAs a motion is a member of is 5, with two motions achieving this number – (legalize, prostitution) and (ban, infant circumcision). The full annotation is provided in the supplementary materials.

In order to validate this annotation, a sample of motion-CoPA pairs was annotated via crowdsourcing. For each motion, argument pairs from 2 randomly chosen non-matching CoPAs and (up to) 2 randomly chosen matching ones were annotated by 5 labelers. Average inter-annotator Cohen's kappa agreement was 0.63. Then, taking the majority vote for each pair as the crowdsourced label, we computed agreement between it and our initial labeling, yielding a kappa score of 0.78. These indicate a rather high agreement, especially in the context of computational argumentation (Passonneau and Carpenter, 2014; Habernal and Gurevych, 2016).

4.2 Expanded data

The initial construction of CoPAs was done with the aim of identifying themes which are recurrent in general, not just in the initial 100 motions. To verify that they generalize to other motions, we collected 589 additional motions, and annotated them for CoPA membership (the same annotators who did the initial annotation). On this new dataset, we found that 503 motions were matched to at least one CoPA (85%), and on average each motion was matched to 1.94 CoPAs. Hence, while our modeling may be biased by the initial set of motions, it seems to generalize well to new ones.

As with the initial set of motions, we used crowd-sourced annotations of a similarly-sampled portion of the dataset to verify the full annotation, attaining an average inter-annotator kappa score of 0.60, and a kappa score of 0.76 when comparing the majority vote to the full annotation.

The full dataset can be found in the supplementary material.

4.3 CoPA claims in recorded speeches

It is natural to ask whether the claims authored for each CoPA are an artificial construct for facilitating motion assignment, or are actual claims, likely to be made by people deliberating these motions.

To this end we considered the speeches recorded by Mirkin et al. (2018). Each such speech is given in the context of a motion, all of which are included in our dataset. For each motion we extracted the CoPAs to which it belongs according to our annotation, yielding 184 speeches with at least one matching CoPA. 7 Figure-Eight[7] annotators were presented with speeches in both audio and written form, alongside the claims from the matching CoPAs. They were asked whether each claim was (i) explicitly made by the speaker, was (ii) implicit in the speech or was (iii) not mentioned at all. A total of 800 (speech, claim) pairs were annotated, with one half of them being claims of a stance opposing that of the speaker.

In order to analyze agreement between annotators, we considered (i) and (ii) as a positive label and the latter as negative. The average inter-annotator Cohen Kappa score was 0.54. Moreover, since we showed both CoPA claims to the annotators we checked whether claims whose stance opposed that of the speaker were ever marked positive. With only 5% of the annotations being so, we concluded that the annotation was of reasonable quality (cf. section 6.2).

5 Matching Methods

Having a sizable dataset of (motion, CoPA) pairs, we examined several classifiers over it. That is, given a motion and a CoPA, the classifier aims to determine whether they match. Since the CoPAs are quite varied, we examined various classifiers, some focused on a motion's action, some on its topic, and some on a combination of both.

Specifically, we examined the following classifiers: **By action (BA-k):** Some actions are strongly indicative for (some of) the CoPAs a motion belongs to. To utilize this, this classifier trains by computing, for each allowed action a, and each CoPA c, the probability $p(c, a)$ that a motion with action a will belong to c. Prediction for a new motion $m = (a, t)$ is done by assigning each CoPA c the score $p(c, a)$. In addition, if the number of (training-set) motions in c with action a is less than some parameter k, this method makes no prediction.

By topic, nearest neighbors (KNN): Given a left-out motion, $m = (a, t)$ the algorithm goes over the motions $m_i = (a_i, t_i)$ in the training set, looking for those such that t_i is most similar to t (using

---

[7]https://www.figure-eight.com/

4

the similarity measure of Ein Dor et al., 2018). It keeps only those whose similarity is above a threshold of 0.5. If there are less than 3 such motions, no prediction is made. Otherwise it takes the (at most) top 5 motions. For each CoPA c, the assigned score is the fraction of these motions which belongs to c. This is then used to predict membership.

By topic, word2vec features (W2V): Each motion $m = (a, t)$ is represented as the word2vec (Mikolov et al., 2013) embedding vector of t (if t is a multi-word expression the vectors are summed and normalized). This vector is then used as a feature vector for a logistic regression classifier. That is, each CoPA is assigned the classification score of the classifier so trained. As a safeguard mechanism, we also determine an actions blacklist, B_c, for each CoPA c. An action a is in B_c if in the training set no motion with action a is in c. During prediction, if the left-out motion's action is in B_c, it will not be predicted as belonging to c.

By topic, Naive Bayes (NB) and Recurrent Neural Network (RNN): Following the work of Rabinovich et al. (2018) each motion $m = (a, t)$ is associated with a set of retrieved sentences containing the term t. For a given CoPA c, all the sentences associated with motions in c are considered as positive examples, and those of motions not in c as negative examples. For *NB*, A Naive Bayes classifier is then trained over the unigrams of these sentences, and uses its score for prediction. In addition, the same blacklist safeguard mechanism as for *W2V* above is used.

Similarly, these sentences were used to train an RNN to differentiate between positively-labeled sentences and negatively-labeled ones. See Rabinovich et al. (2018) for more details on these methods.

By topic and action (LR): We defined 17 features based on similarities between a motion and a CoPA, and on co-occurrence counts, similar to the one used in *BA-k*. A logistic regression classifier was trained and scored on the resulting feature vectors over pairs of (motion, CoPA). See appendix C for details.

Ensemble: For completeness, all 6 methods above were aggregated by simply assigning each CoPA the highest score it attained among all of them. We note that this is a very naive approach; while all methods produce scores in $[0, 1]$, it is not clear that they are comparable. In practice one would probably use an aggregation method that differentiates between the different classification methods, and between different CoPAs.

All classifiers (except one[8]) were evaluated in a leave-one-motion-out framework, over all motions and over relevant CoPAs. That is, each classifier was trained and tested 689 times – in each iteration it was trained over 688 motions and the relevant CoPAs, and then predicted whether the left-out motion matched these CoPAs. More precisely, each CoPA is assigned a score. We vary the score threshold, and determine membership by whether the assigned score exceeds the threshold.

6 Results

6.1 Complete dataset

All in all, our dataset describes the motion-CoPA relations of 689 motions and 37 CoPAs. Figure 1 shows a histogram of the CoPA sizes in this dataset. The two biggest CoPAs (*Fixable* and *Conservatism*) include nearly one third of the motions (207 motions, and 211 motions respectively), while at the other end, the class *Self determination* contains only 3 motions. Most CoPAs (32 out of the 37) are of modest size, containing less than 10% of the motions.
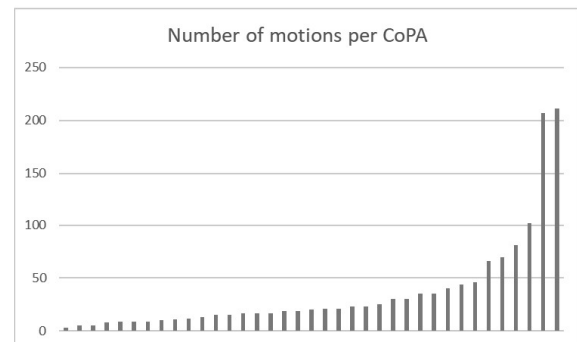

Figure 1: Distribution of the number of motions per CoPA.

Importantly, the CoPAs capture different facets of a motion, rather than induce a partition of the motions set. On average, a CoPA has a non empty intersection with 11.95 other CoPAs, with the average intersection size being 21% of the CoPA size. Figure 2 shows the inter-connectivity

---

[8]For technical reasons we trained and evaluated the RNN method using 3-fold cross-validation.

graph among CoPAs. The CoPA *Self determination* (which has only 3 motions in it) is an isolated vertex in this graph, but other than that the graph is connected. This is especially noteworthy, considering that many of the CoPAs are rather small. Figure 5 in Appendix D shows a heat-map of overlap sizes.
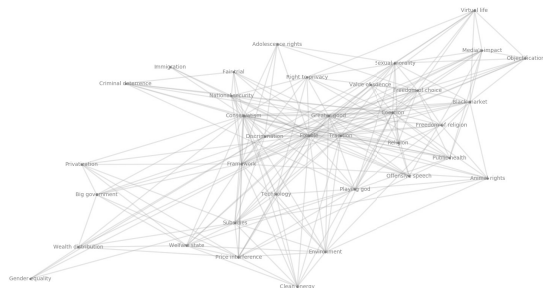

Figure 2: Graph of CoPAs, where edges indicate non-empty intersection and distance between vertices is indicative of intersection size. Not shown is "Self determination", an isolated vertex.

In the complete dataset, 87% of the motions belong to at least one CoPA, and on average each motion belongs to 1.95 CoPAs. That is, while this is certainly only a first step toward modeling principled recurring arguments, the suggested CoPAs are indeed a concise set that covers distinct argumentative themes and offers a good coverage w.r.t. the world of motions defined here.

6.2 CoPA claims in recorded speeches

Of the 184 annotated speeches of Mirkin et al. (2018), 87% had at least one CoPA-claim annotated as positive[9], and in total, 66% of the 400 (speech, claim) pairs (where the stance of the claim and the speaker were aligned) were marked as positive. However, in the vast majority of cases the claim was marked as implicit in the speech – according to the annotation only 10% of the speeches contain a CoPA-claim explicitly, and only 5% of the pairs are labeled as an explicit mention.

One reason for this may be the three "general" CoPAs, since their claims are so general that they would usually be at least implicit in a speech. When removing these CoPAs from the analysis 62% of the speeches have at least one positive claim, and 39% of the pairs are positive. Hence, even without these classes, most speeches implicitly mention at least one claim from the dataset.

---

[9] A pair is considered positive if a majority of annotators chose option (i) or (ii); cf. section 4.3.

This is probably due to the rather generic phrasing of the claims, which in the first place were constructed to be applicable "as-is" in multiple contexts. In other words, this annotation not only confirms that the CoPA claims convey arguments actually alluded to by humans, but that they do so at a rather high level, and so capture arguments that are not only plausible for a motion but also probable.

Conversely, we also examined, for each CoPA, in what fraction of the speeches to which it is matched is its claim annotated as positive. Of the 37 CoPAs, 29 match motions in Mirkin et al. (2018). For all but one (*Sexual morality*), in at least 25% of the relevant speeches, the CoPA's claim (of the correct stance) was labeled positive. For 24 CoPAs at least 50% were so labeled.

6.3 Motion-CoPA matching

As noted in section 5, we evaluated the proposed matching methods in a leave-one-out framework. For the action-based method, *BA-k*, we set $k = 5$ and consider only CoPAs which contains at least 5 motions with the same action. For the topic-based methods we considered only CoPAs which were manually marked as topic-based (see appendix B) and contain at least 10 motions. The *LR* method was naturally evaluated on all CoPAs.

Figure 3 describes the precision-recall trade-off for each of the 7 methods from section 5, which is computed over all (motion, CoPA) pairs: precision is the fraction of matching pairs whose score is above the threshold from among all pairs with such a score; recall is the ratio between the number of matching pairs with such a score, and the total number of matching pairs.

Note that for methods which look at only a subset of the CoPAs recall is bound to be low, since recall calculation takes into consideration all CoPAs, not just this subset.

With the task of *Argument Invention* in mind, a use-case of interest is, given a motion, to provide (at least) one CoPA from which argumentative content can be extracted. Accordingly, Figure 4 evaluates the precision for the highest scoring CoPA of each motion – for a given threshold, the figure depicts the fraction of motions whose highest scoring CoPA is both a match and above the threshold, as a function of the fraction of motions for which at least one CoPA passes the threshold. As can be seen, for a threshold that yields
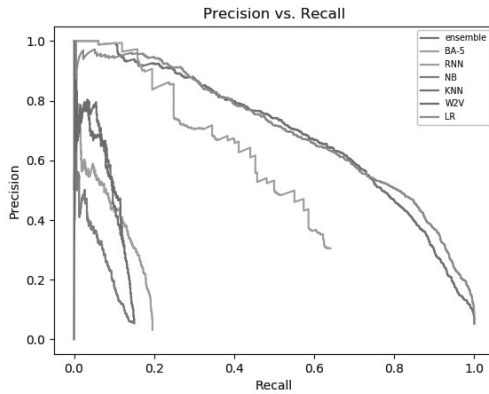
6

Figure 3: Precision-Recall curve for the various motion-CoPA matching methods.

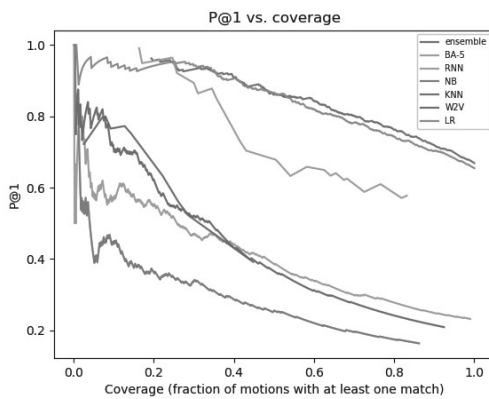CoPA prediction for half the motions, the ensemble method has 86% precision for its top prediction.

Figure 4: P@1 vs. coverage of the various motion-CoPA matching methods.

Finally, recall that the three "general" CoPAs (*Conservatism, Fixable, Framework*) might dominate the predictions analyzed above. Omitting them from the analysis does reduce precision somewhat, but nonetheless, the top prediction of the ensemble method for a threshold yielding a prediction for half the motions attains a precision of 75% (Figure 6 in appendix D; For this analysis the "general" classes were not included in the recall computation).

A naive baseline would always (and only) predict the CoPA with the largest number of motions as a match. When considering all CoPAs, this attains a precision of 30% (for *Conservatism*), and when omitting the three general CoPAs, a precision of 12% (for *Coercion*).

7 Discussion

The most basic argument model is probably Aristotle's *categorical syllogism*, which consists of a major premise, a minor premise and a conclusion (Aristotle and Kennedy, 1991); with the minor premise being a categorical proposition connecting between the major premise and the conclusion. The canonical example is:

> All men are mortal.
> Socrates is a man.
> Therefore, Socrates is mortal.

It is interesting to consider the CoPAs and the claims they contain in this context. When aiming to argue for a motion, and identify a CoPA to which it belongs, one can create a syllogistic-like argument as follows. The major premise would be the CoPA claim, the minor premise would explain why the motion is a member of the CoPA, and the conclusion would be that the motion should stand[10].

For example, when deliberating the motion (further exploit, solar energy), and identifying that it belongs to the CoPA *Clean energy*, the resulting (heuristic) argument could be:

> Humanity must embrace clean energy in order to fight climate change.
> Solar energy is a form of clean energy.
> Therefore, humanity must further exploit solar energy.

A major challenge in constructing the CoPAs herein was finding an explicit phrasing for the arguments, one that would be suitable without further context. One example for this is the *backlash* argument – an argument stating that implementing the policy will be counter-productive, since it will create a backlash reaction. While this is a common argument, arguing why a backlash reaction will occur and how it will be counterproductive may well depend on context. Moreover, it is difficult to phrase an appropriate claim of the opposite stance without further context.

However, the CoPAs can actually provide us with an appropriate context needed for phrasing such arguments. Thus, one could phrase more

_____

[10] Since the major premise here is not a categorical proposition, the argument will not be true in the propositional, *modus ponens* sense. But if the claim is indeed an *endoxa*, the argument should be one that most people consider plausible.

7

specific backlash arguments for the CoPA *Subsidies* or *Coercion* (with different phrasings), and use them when the CoPA is matched to a motion. In other words, we can expand the set of CoPA arguments to include more than just 2 claims; it could include further instances of principled arguments, each perhaps tailored to the specific CoPA. Recall that our motivation is *Argument Invention* – when a CoPA is matched, the underlying system can present all arguments the CoPA contains.

Indeed, with the aim of assisting critical writing in mind, one need not limit CoPAs to claims or even to coherent arguments. CoPAs could very well be rhetorical *loci* for relevant anecdotes, proverbs, memes, quotes from famous people and so on. They could also include text written in different styles to accommodate different types of presentations (*pronuntiatio*). In response to a topic, a system making use of the CoPAs knowledge base could present all these texts to the users, or filter them according to their preferences.

An interesting research direction in this respect is to include in a CoPA, for each claim it contains, a rebuttal argument that counters it. This can enable an argumentative dialog system (Rach et al., 2018), along the lines alluded to in Mirkin et al. (2018) – one can envision a system that performs *listening comprehension*, identifies the relevant CoPAs, checks whether any of the claims in the CoPA were mentioned in the audio, and, for those that do, responds with the rebuttal arguments matching this claim. This is similar to scripted dialog systems, with the important difference that the texts are not written for a specific scenario. they are principled arguments which can be used in many different contexts, allowing for an open-domain dialog system. We intend to describe such a system in future work.

Furthermore, a CoPA can include complex argumentative structures such as those in AraucariaDB (Reed and Rowe, 2004), from which multi-layered arguments can be constructed, e.g. using the Carneades Argumentation System of Walton and Gordon (2012). That is, instead of having such data per topic, as is currently the case in AraucariaDB, having such data for commonplace principled arguments facilitates their use over a wide range of topics. Note that for this the stance of the argument w.r.t the CoPA and the motion is important. For the sake of simplicity and brevity we have ignored this issue in this manuscript, but the

relevant stance labeling is available in the supplementary material.

In the field of computational argumentation, *de novo* argument synthesis has received relatively little attention. One naive attempt is that of Bilu and Slonim (2016), where claims are generated by pasting together a topic and short predicate. The framework suggested here may provide a richer and more stable basis for argumentative text generation. That is, a CoPA may include structured data which describes its principal theme. Then, when presented with a motion in this CoPA, the system would automatically generate, *de novo*, argumentative text based on this structured data *and the topic*. For example, this could be an NLG neural net trained on a large corpus of claims extracted using argument mining for motions in the CoPA.

Finally, let us reappraise the basic intuition of the corpus-wide argument-mining approach to argument invention – that an effective argumentation is one that draws on the widest possible array of proofs and arguments. Rhetoricians have characterized the art of convincing as starting from general and basic views, facts and opinions accepted by everyone (Perelman, 1971; Kock, 2009). In other words, an efficient argument starts not from the most original and unseen premises, but from what the public takes as consensual, and only then progresses to what is controversial. Therefore, the need for principled arguments is not only a question of time and practicality, but also stems from the essential nature of rhetoric: it is the necessity to call on the general views and opinions shared by everyone and to show that they ratify the desired conclusion.

8 Conclusion

We presented a novel framework – 689 controversial motions with a variety of topics and actions – in which the *Argument Invention* task can be formalized and assessed. We formalized the notion of commonplace principled arguments, and suggested a concrete and diverse taxonomy for them. While this taxonomy can certainly be expanded and refined, it nonetheless has the basic desired properties: most motions in our framework belong to it, annotators tend to agree on CoPA-motion matching, this matching can be done automatically with reasonable success, and human debaters tend to allude to the ascribed arguments when debating these motions.

References

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. 2016. Cross-domain mining of argumentative text through distant supervision. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: human language technologies*, pages 1395–1404.

Bob Altemeyer. 1981. *Right-wing authoritarianism*. University of Manitoba press.

Aristotle and George Alexander Kennedy. 1991. *On rhetoric: A theory of civic discourse*. Oxford University Press New York.

Yonatan Bilu and Noam Slonim. 2016. Claim synthesis via predicate recycling. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 525–530.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, pages 438–444.

Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 11–22.

Liat Ein Dor, Alon Halfon, Yoav Kantor, Ran Levy, Yosi Mass, Ruty Rinott, Eyal Shnarch, and Noam Slonim. 2018. Semantic relatedness of wikipedia concepts-benchmark data and a working solution. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Jim AC Everett. 2013. The 12 item social and economic conservatism scale (secs). *PloS one*, 8(12):e82131.

Cheryl Glenn, Melissa A Goldthwaite, and Robert Connors. 2008. *The St. Martin's guide to teaching writing*. Bedford/St. Martin's.

Ivan Habernal and Iryna Gurevych. 2016. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional lstm. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1589–1599.

John T Jost, Sally Blount, Jeffrey Pfeffer, and György Hunyady. 2003. Fair market ideology: Its cognitive-motivational underpinnings. *Research in organizational behavior*, 25:53–91.

Christian Kock. 2009. Choice is not true or false: The domain of rhetorical argumentation. *Argumentation*, 23(1):61–80.

Janice M Lauer. 2004. *Invention in rhetoric and composition*. Parlor Press LLC.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500.

Ran Levy, Ben Bogin, Shai Gretz, Ranit Aharonov, and Noam Slonim. 2018. Towards an argumentative content search engine using weak supervision. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2066–2081.

Marco Lippi and Paolo Torroni. 2015. Argument mining: A machine learning perspective. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 163–176. Springer.

Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2):10.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Shachar Mirkin, Guy Moshkowich, Matan Orbach, Lili Kotlerman, Yoav Kantor, Tamar Lavee, Michal Jacovi, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. 2018. Listening comprehension over argumentative content. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 719–724.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107. ACM.

Rebecca J Passonneau and Bob Carpenter. 2014. The benefits of a model of annotation. *Transactions of the Association for Computational Linguistics*, 2:311–326.

Chaim Perelman. 1971. The new rhetoric. In *Pragmatics of natural languages*, pages 145–149. Springer.

Ella Rabinovich, Benjamin Sznajder, Artem Spector, Ilya Shnayderman, Ranit Aharonov, David Konopnicki, and Noam Slonim. 2018. Learning concept abstractness using weak supervision. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4854–4859. Association for Computational Linguistics.

Niklas Rach, Saskia Langhammer, Wolfgang Minker, and Stefan Ultes. 2018. Utilizing argument mining techniques for argumentative dialogue systems. In *Proceedings of the 9th International Workshop On Spoken Dialogue Systems (IWSDS)*.

Chris Reed and Glenn Rowe. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979.

Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence-an automatic method for context dependent evidence detection. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 440–450.

Holli A Semetko and Patti M Valkenburg. 2000. Framing european politics: A content analysis of press and television news. *Journal of communication*, 50(2):93–109.

Eyal Shnarch, Carlos Alzate, Lena Dankin, Martin Gleize, Yufang Hou, Leshem Choshen, Ranit Aharonov, and Noam Slonim. 2018. Will it blend? blending weak and strong labeled data in a neural network for argumentation mining. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 599–605.

Jim Sidanius and Felicia Pratto. 2001. *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Tim Sonnreich. 2012. Monash association of debaters guide to debating (tips, tactics and first principles).

Christian Stab and Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1501–1510.

Claes H. de Vreese. 2005. News framing: Theory and typology. *Information Design Journal*, 13:51–62.

Douglas Walton. 2013. *Argumentation schemes for presumptive reasoning*. Routledge.

Douglas Walton and Thomas F Gordon. 2012. The carneades model of argument invention. *Pragmatics & Cognition*, 20(1):1–31.

Douglas Walton and Thomas F Gordon. 2017. Argument invention with the carneades argumentation system. *SCRIPTed*, 14:168.

Douglas Walton and Thomas F Gordon. 2018. How computational tools can help rhetoric and informal logic with argument invention. *Argumentation*, pages 1–27.

Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation schemes*. Cambridge University Press.