Expert Stance Graphs for Computational Argumentation

Orith Toledo-Ronen   **Roy Bar-Haim**   **Noam Slonim**

IBM Research - Haifa

{oritht,roybar,noams}@il.ibm.com

Abstract

We describe the construction of an *Expert Stance Graph*, a novel, large-scale knowledge resource that encodes the stance of more than 100,000 *experts* towards a variety of controversial *topics*. We suggest that this graph may be valuable for various fundamental tasks in computational argumentation. Experts and topics in our graph are Wikipedia entries. Both automatic and semi-automatic methods for building the graph are explored, and manual assessment validates the high accuracy of the resulting graph.

1 Introduction

Background knowledge plays an important role in many NLP tasks. However, *computational argumentation* is one area where little work has been done on developing specialized knowledge resources.

In this work we introduce a novel knowledge resource that may support various tasks related to argumentation mining and debating technologies. This large-scale resource, termed *Expert Stance Graph*, is built from *Wikipedia*, and provides background knowledge on the stance of experts towards debatable topics.

As a motivating example, consider the following *stance classification* setting, where the polarity of the following expert opinion on *Atheism* (*Pro* or *Con*) should be determined:

> *Dawkins sums up his argument and states, "The temptation (to attribute the appearance of a design to actual design itself) is a false one, because the designer hypothesis immediately raises the larger problem of who designed the designer. The whole problem we started out with was the problem of*

explaining statistical improbability. It is obviously no solution to postulate something even more improbable." (Dawkins, 2006, p. 158)

Inferring the stance directly from the above text is a difficult and complex task. However, this complexity may be circumvented by utilizing background knowledge about (Richard) Dawkins, who is a well-known atheist. Dawkins' page in Wikipedia[1] includes various types of evidence for his stance towards atheism:

1. *Categories:* Dawkins belongs to the following Wikipedia categories: *Antitheists, Atheism activists*, *Atheist feminists* and *Critics of religions*[2].

2. *Article text:* The article text contains statements such as *"Dawkins is a noted atheist"* and *"Dawkins is an outspoken atheist"*.

3. *Infobox:* Dawkins has a *known-for* relation with *"criticism of religion"*.

2 Expert Stance Graphs

The *Expert Stance Graph (ESG)* is a directed bipartite graph comprising two types of nodes: (a) *concept* nodes, which represent debatable topics such as *Atheism, Abortion, Gun control* and *Same-sex marriage*, and (b) *expert* nodes, representing persons whose stance towards one or more of the concepts can be inferred from Wikipedia. Stance is represented as labeled directed edges from an expert to a concept, e.g. *Richard Dawkins* $\xrightarrow{Pro}$ *Atheism*. Each concept and each expert have their own article in Wikipedia. We use the term *Experts* inclusively to refer to academics,

---

[1] https://en.wikipedia.org/wiki/Richard_Dawkins

[2] Inferring *Pro* stance for *Atheism* from *Critics of religions* depends on knowing the contrast relation between *Atheism* and *Religion*.

writers, religious figures, politicians, activists, and so on.

3 Applications

Expert opinions are highly valuable for making persuasive arguments, and expert evidence (premise) is a commonly used type of argumentation scheme (Walton et al., 2008). Rinott et al. (2015) describe a method for automatic evidence detection in Wikipedia articles. Three common evidence types are explored: *Study*, *Expert*, and *Anecdotal*. The proposed method uses type-specific features for detecting evidence. For instance, in the case of expert evidence, a lexicon of words describing persons and organizations with relevant expertise is used.

The process of incorporating expert opinions on a given topic into an argument involves several steps. First, we need to retrieve from our corpus articles that contain expert opinions related to the given topic. Second, the exact boundaries of these opinions should be identified. Finally, the stance of the expert opinion towards the topic (Pro or Con) should be determined, to ensure it matches the stance of the argument we are making. Each of these steps is a challenging task by itself.

The expert stance graph may facilitate each of the above subtasks. If an expert E is known to be a supporter or an opponent of some topic T, then the Wikipedia page of E is likely to contain relevant opinions on T. Furthermore, a mention of E can be a useful feature for identifying relevant expert opinions for T in a given article.

Finally, perhaps the most important use of the graph for expert evidence is stance classification. Previous work on stance classification has shown that it can be much improved by utilizing external information beyond the text itself. For example, posts by the same author on the same topic are expected to have the same stance (Thomas et al., 2006; Hasan and Ng, 2013). Similarly, as shown in the previous example, external knowledge on expert stance towards a topic can improve stance classification of expert opinions.

4 Building the Graph

We consider two complementary settings for building the graph: (a) *Offline*, in which the set of concepts is predefined, and minimal human supervision is allowed, and (b) *Online*, where our goal is to find ad-hoc *Pro* and *Con* experts for an unseen concept, in a fully-automatic fashion.

For both settings, our approach is based on Wikipedia categories and lists, which have several advantages: (a) they provide an easy access to large collections of experts, (b) their stance classification is relatively easy, and (c) their hierarchical structure can be exploited.

4.1 Concepts

Offline construction of the graph starts with deriving the set of concepts. We started with Wikipedia's list of controversial issues[3], which contains about 1,000 Wikipedia entries, grouped into several top-level categories. We manually selected a subset of 12 categories , and filtered out the remaining 3 categories.[4]

One of the authors selected from the remaining list concepts that represent a two-sided debate (*Meaning of life*, for instance, is a controversial topic but does not represent a two-sided debate). Persons and locations were filtered out as well. This list was expanded manually by identifying relevant concepts in Wikipedia article titles that contain the words "Debate" or "Controversy". Finally, two annotators assessed the resulting list according to the above guidelines. Concepts that were rejected by both annotators were removed. The final list contained 201 concepts.

4.2 Candidate Expert Categories

Next, we search relevant Wikipedia categories and lists for each concept. The process starts with creating search terms. The concept itself is a search term, as well as any lexical derivation of the concept that represents a person (e.g. *Atheism→Atheist*), which we term *person derivations*. Person derivations are found using WordNet (Miller, 1995; Fellbaum, 1998): we look for lexical derivations of the concept that have *"person"* as a direct or inherited hypernym.

We then find all Wikipedia categories and lists[5] that contain the search terms. For example, given the search terms *atheism* and *atheist*, some of the categories found are *Atheism activists, American*

---

atheists, List of atheist authors, Converts to Christianity from atheism or agnosticism and *Critics of atheism*. The set of categories is further expanded with subcategories of the categories found in the previous step. This step adds more relevant categories that do not contain the search terms, such as *Antitheists* for *Atheism*. To avoid topic drifting, we only add one level of subcategories.

Next, the persons associated with each category[6] are identified by considering outgoing links from the category page which are of type "Person", based on DBPedia's `rdf:type` property for the page (Lehmann et al., 2014). Categories with fewer than five persons are discarded. We also removed three concepts, for which the number of categories was too large: *Christianity*, *Catholicism*, and *Religion*. The resulting set included 4,603 categories containing 121,995 persons. Categories were found for 132 of the 198 concepts.

4.3 Category Stance Annotation

Finally, category names are manually annotated for stance. The annotation process has two stages: first, determine whether the category explicitly defines membership in a group of persons. For instance, *Swedish women's rights activists* and *Feminist bloggers* meet this criterion, but *Feminism and history* does not. We apply this test since we observed that it is much easier to predict with confidence the stance of persons in these categories.

Categories that do not pass this filter are marked as *Irrelevant*. Otherwise, the annotators proceed to the second stage, where they are asked to determine the stance of the persons in the given category towards the given concept, based on the category name. Possible labels are:

1. *Pro:* supporting the concept.

2. *Con:* opposing the concept.

3. *None:* The stance towards the concept cannot be determined based on the category name.

For instance, for the concept *Communism* we will have *British communists* and *Canadian Trotskyists* classified as *Pro*, *Moldovan anticommunists* classified as *Con*, and *Western writers about Soviet Russia* classified as *None*. Annotators may also consider direct parent categories for determining stance. In the previous example, knowing that *Canadian communists* is a parent category

---

[6]For convenience, we will refer in the following to categories and lists collectively as "categories".

Polarity	Concepts	Categories	Experts
Pro	105	3,221	93,570
Con	40	272	10,666

Table 1: Statistics on the Expert Stance Graph

of *Canadian Trotskyists* may help classifying the latter as *Pro* for *Communism*.

The categories were labeled by a team of six annotators, with each category labeled by two annotators. The overall agreement was 0.92, and the average inter-annotator Cohen's kappa coefficient was 0.79, which corresponds to *substantial agreement* (Landis and Koch, 1997). Cases of disagreement were labeled by a third annotator and were assigned the majority label. Category annotation was completed rather quickly - about 260 categories were annotated per hour. The total number of annotation hours invested in this task was 37.

The resulting ESG is composed of all experts in the categories labeled as *Pro* and *Con*. A total of 104,236 experts were found for 114 out of the 132 concepts, and for 31 concepts, both *Pro* and *Con* experts were found. The number of concepts, categories and experts for each stance is given in Table 1. As shown in the table, the vast majority of categories and experts found are *Pro*. Overall, our method efficiently constructs a very large ESG, while only requiring a small amount of human annotation time.[7]

4.4 Category Stance Classification

The offline list of concepts we started with is unlikely to be complete. Therefore, we would like to be able to find on-the-fly *Pro* and *Con* experts also for new, unseen concepts. This requires the development of a stance classifier for categories. We randomly split the 198 concepts into two equal-size subsets and used one subset for development and the other for testing. As a result, the 132 concepts for which categories were found are split into a development set, containing 69 concepts and their associated 2,069 categories, and a test set, containing 63 concepts and 2,534 categories. The development set was used for developing a simple rule-based classifier.

The logic of the rule-based classifier is sum-

---

[7]The IBM Debating Technologies group in IBM Research has already released several data resources, found here: `https://www.research.ibm.com/haifa/dept/vst/mlta_data.shtml`. We aim to release the resource presented in this paper as well, as soon as we obtain the required licenses.

```
Input: category CAT ; concept C ; person derivation PD for
   the concept
Output: stance classification of CAT into PRO/CON/NONE
   if CAT =~ critics of C then
      return CON
   else if CAT =~ anti|former|... PD then
      return CON
   else if CAT =~ PD dissident|... then
      return CON
   else if CAT =~ PD then
      return PRO
   else if CAT =~ anti|former|... C PERSON then
      return CON
   else if CAT =~ C PERSON then
      return PRO
   else
      return NONE
```

Algorithm 1: Category stance classification

	Predicted	Correct	Labeled	P	R
	Categories				
Pro	1,298	1,182	1,738	91.1	68.0
Con	144	140	186	97.2	75.3
	Experts				
Pro	28,693	25,754	41,701	89.8	61.8
Con	4,113	4,016	6,912	97.6	58.1

Table 2: Category stance classification results

	Predicted	Correct	Labeled	P	R
	Manual Annotation				
Pro	200	181	189	90.5	95.8
Con	200	173	178	86.5	97.2
	Classifier				
Pro	76	60	189	78.9	31.7
Con	87	81	178	93.1	45.5

Table 3: Expert stance assessment

marized in Algorithm 1. "=~" denotes pattern matching, and PERSON is any hyponym of the word "person" in WordNet, e.g. *activist, provider*, and *writer*. "..." denotes omission of some lexical alternatives.

The algorithm is first applied to the category itself, and if it fails to make a *Pro* or *Con* prediction (i.e returns *None*), it is applied to its direct parent categories, and the classification is made based on the majority of their *Pro* and *Con* predictions.

Table 2 shows the performance of the classifier on the test set, with respect to both categories and experts. Expert-level evaluation is done by labeling all the experts in each category with the category label. The following measures are reported for *Pro* and *Con* classes: number of predictions, number of correct predictions, number of labeled instances for this class, precision (P) and recall (R). Overall, the classifier achieves high precision for Pro and Con, both at the category and at the expert level, while covering most of the labeled instances. Yet, the coverage of the classifier is incomplete. As an example of its limitations, consider the categories *American pro-choice activists* and *American pro-life activists*, which are *Pro* and *Con* abortion, respectively. Their stance cannot be determined from the category itself according to our rules, because they do not contain the concept *Abortion*, and both were added as subcategories of *Abortion in the United States*, a category that does not have a clear stance (and indeed has both *Pro* and *Con* subcategories).

5   Expert-Level Assessment

So far we assumed that experts' stances can be predicted precisely from their category names. In this section we put this assumption to the test. We sampled 200 Pro experts and 200 Con experts from the test set. The polarity of the experts was derived from the manual labeling of their category. For each sampled instance, we first randomly selected one of the concepts in the test set, and then randomly picked an expert with the requested polarity. If the concept did not have any experts with that polarity, the above procedure was repeated until such an expert was found.

We then asked three human annotators to determine the stance of the experts towards their associated concept (*Pro/Con/None*), based on any information found on their Wikipedia page, and considered the majority label. As with the previous task, the annotators achieved *substantial agreement* (average kappa of 0.65). We evaluated the expert stance inferred from the category labeling by both the manual annotation and the rule-based classifier against these 400 labeled experts. The results are summarized in Table 3.

For the manual annotation, we see that the category name indeed predicts the expert's stance with high precision. In most of the misclassifications cases, the annotators could not determine the stance from the expert's web page. This discrepancy is partially due to the fact that the expert's page shows categories containing the expert, but does not display lists and parent categories containing the expert, which are available for category-based stance annotation. The precision of the classifier on this sample is also quite good (better for *Con*), but while we are able to identify a substantial part of the experts, recall still leaves much room for improvement.

6 Conclusion and Future Work

We introduced *Expert Stance Graphs*, a novel, large scale knowledge resource that has many potential use cases in computational argumentation. We presented an offline method for constructing the graph with minimal supervision, as well as a fully-automated method for finding experts for unseen concepts. Both methods show promising results.

In future work we plan to improve coverage by considering additional sources of information, such as the text of the expert's page in Wikipedia. We will also apply the graph in different tasks related to the detection and stance classification of expert evidence.

We also plan to enrich the graph with additional types of knowledge, which may be utilized to predict missing stance edges. Semantic relations between concepts, such as contrast (e.g. *Atheism* vs. *Religion*), may support such inferences, as experts are expected to have opposite stances towards contrasting concepts. Another possible extension of the graph is *influence* links between experts, which may indicate similar stances for these experts. Influence information is available from Wikipedia infoboxes.

Finally, we would like to apply collaborative filtering techniques to predict missing expert-concept stance relations. This is based on the intuition that experts who tend to have same (or opposite) stances on a set of topics, are likely to follow a similar pattern on topics for which we only have partial stance information.

References

Richard Dawkins. 2006. *The God Delusion*. Bantam Press.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. Language, Speech and Communication. MIT Press.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of IJCNLP*.

J. R. Landis and G. G. Koch. 1997. The measurements of observer agreement for categorical data. *Biometrics*, 33:159–174.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2014. DBpedia - a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*.

G. A. Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM*, pages 39–41.

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 EMNLP*.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of EMNLP*.

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