Computationalizing a Toulmin Model for Argumentation Generation

Paul Reisert[†] *Naoya Inoue*[†] *Kentaro Inui*[†] *Toshihiko Yanase*[‡] *Kohsuke Yanai*[‡]
Tohoku University[†] *Central Research Laboratory, Hitachi, Ltd.*[‡]
{preisert,naoya-i,inui}@tohoku.ac.jp {toshihiko.yanase.gm, kohsuke.yanai.cs}@hitachi.com

1 Introduction

Responding to recent advances in the social networking technology such as weblogs and microblogs, a number of opinions of people across the world are widely spread on the Web. The automatic visualization of the debate structure of these opinions is a challenging problem, but will be extremely useful for many users. For example, it will be useful if we can automatically identify that *house should ban alcohol*, the opinion of blogger A, can be supported by *alcohol causes liver disease*, the opinion of blogger B_1, and can be refuted by *alcohol promotes good health*, the opinion of blogger B_2.

In order to accomplish this goal, we turn to argumentation mining, a recently popular field of study. Argumentation is the theory of reaching a conclusion for a claim based on some type of premises. Various types of argumentation mining methods have been proposed for areas such as legal documents [5] and online argument stance [3]. In this work, we focus on one particular argumentation model, the Toulmin model [13], due to its ideal structure for policy debates [1], as it can be applicable for daily arguments opposed to one specific domain. For example, claim, data, and warrant, all components of the Toulmin model, can be thought of as follows: a claim is something an individual believes, data is support, or evidence, to the claim, and a warrant is the link between the claim and data. Substituting a debate motion such as *We should ban alcohol* for a claim, one can utilize the Toulmin model for constructing various arguments to support a given motion while not only discovering support, but also the implicit warrant which can strengthen the effectiveness of the data.

To the best of our knowledge, no previous work has experimented with automatically constructing Toulmin instantiations through computational modeling for a given keyword. Therefore, in this work, we create a computational model for automatically constructing Toulmin-based argumentation structure from the Web for given keywords. The most challenging part of automatic construction of Toulmin instantiation is to recognize the semantic relation between statements. In the context of discourse relation recognition and QAs, significant amount of researches have been done [15, 11, 10, 7, 12]; however, there still remains many issues. Our idea is to represent the statements as excitation relations (i.e., PROMOTE(X, Y) or SUPPRESS(X, Y)) [6] as a start for Toulmin instantiation.

2 Related Work

To the best of our knowledge, no prior work has developed a computation model for automatically constructing Toulmin instantiations. However, various components of the Toulmin model have individually been researched and are discussed below.

The most similar work to ours is the automatic detection of implicit premises for Walton [16]'s argumentation schemes [5]. Similarly, we aim to make the implicit links in the Toulmin model explicit through computational modeling in order to assist with generating constructive debate speeches. In future work, we plan to adopt different argumentation theories.

Inspired by Hashimoto et al. [6]'s excitatory and inhibitory templates, in this work, we similarly compose a manual list of promote and suppress predicates.

Given a motion-like topic, previous work has found relevant claims to support the topic [7]. Other work has utilized a list of controversial topics in order to find relevant claim and evidence segments utilizing discourse markers [12]. Previous Why-QA work [15, 11, 10] has dealt with finding answers for questions such as *Why should alcohol be banned?*. In this case, a passage such as *Alcohol causes heart disease* can be retrieved; however, the passage is not necessarily concerned with *Why is heart disease negative?* which can act as a link between the question and answer. In this work, in addition to a claim and it data, or evidence, we explore finding the link, or warrant, between the claim and data in order to strengthen the relationship between both, one of the aspects of the Toulmin model.

In terms of determining stance, previous work has utilized attack or support claims in user comments as a method for determining stance [3]. In this work, we rely on our PROMOTE(X,Y) and SUPPRESS(X,Y) relations, coupled with positive and negative sentiment values, as a means to signify stance. Simultaneously, not only does this assist with stance, but it is an important feature for argument construction in our first round of constructing automatic Toulmin instantiations.

3 Toulmin Model

Toulmin was the first to believe that most arguments could simply be modeled using the following six components: claim, data, warrant, backing, qualifier, and rebuttal [13]. This model is referred to as the Toulmin model and is shown in Figure 1, along with an instantiation. In this work, we focus on constructing an argument consisting of a claim, data, warrant, as these three components make up the bare minimum of the Toulmin model. The claim consists of the argument an individual wishes for others to believe. Data consists of evidence to support the claim. However, in the event the data is considered unrelated to the claim by another individual, such as a member of a negative team in a policy debate, the warrant, although typically implicit, can explicitly be mentioned to state the relevance of the data with the claim.

In addition to the basic components, one individual may require more information to support the warrant. This component is referred to as backing, and we attempt to identify backing in this work. A qualifier consists of a component, such as a sentence or word, in which affects the degree of the claim. Finally, a rebuttal consists of a counter-claim to a claim. We leave the detection of qualifier and rebuttal for future work.

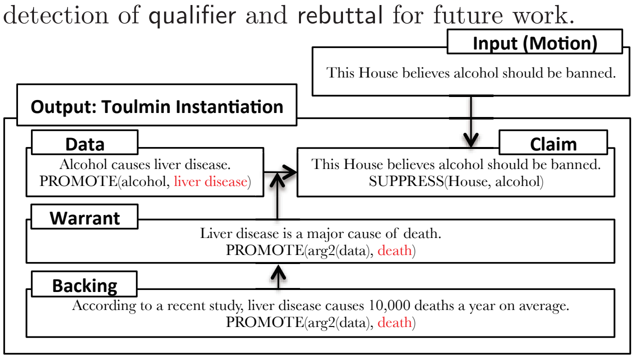

Figure 1: An Instantiation of the Toulmin Model. Red color represents negative sentiment.

4 Methodology

As shown in Figure 1, our task consists of the following: given a topic motion in the form PROMOTE(House,Y) or SUPPRESS(House, Y), where Y is a topic keyword, we instantiate a Toulmin model by first recognizing the topic motion as a Toulmin model claim, and through computational modeling, we generate the remaining Toulmin model arguments.

For instantiating a Toulmin model through computational modeling given a motion, or claim in the Toulmin model, we need to recognize the semantic relation between sentences in a corpus. For example, to find data of the claim, we need find a set of sentences that can serve as the evidence of the claim. However, as described in Section 2, there are still a lot of challenging problems in this research area.

Therefore, our idea is to focus on the sentences that can be represented by an excitation relation, namely PROMOTE(X, Y) or SUPPRESS(X, Y), which is inspired by [6]. Focusing on such sentences, we can recast the problem of semantic relation recognition between sentences as a simple pattern matching problem. For example, suppose we are given the claim SUPPRESS(government, riot). Then, we can find the supporting evidence of this claim by searching for sentences that match PROMOTE(riot, Z), where the sentiment polarity of Z is negative. Another example is shown in Figure 1.

4.1 Overview

We develop a two-staged framework for the automatic construction of Toulmin instantiations. First, we extract a set of claims represented by two-place predicates (e.g., *cause(alcohol, cancer)*) from a text corpus and generalize them into an excitation relation, namely either PROMOTE(X, Y) or SUPPRESS(X, Y). We then store the generalized relations into a database, which we call a *knowledge base*.

Second, given the motion claim that is also represented by a two-place predicate (e.g., *ban(house, alcohol)*) by the user, we find relevant claims from the knowledge base to identify data, warrant, and backing for the input motion claim. In the rest of this section, we elaborate on two processes one by one.

4.2 Knowledge Base Construction

For constructing a knowledge base of PROMOTE(X,Y) and SUPPRESS(X,Y) relations, we rely on a manually created list of verbs representing PROMOTE/SUPPRESS relations and parsed dependency output. Similar to Open Information Extraction systems [18, 4, 9, etc.], we extract a set of triples (A_1, R, A_2), where R is a verb matching a PROMOTE/SUPPRESS-denoting verb, A_1 is a noun phrase (NP) serving as the surface subject of R, A_2 is an NP serving as the surface object of R.

In our experiment, we utilized a collection of web pages extracted from ClueWeb12 as a source corpus of knowledge base construction. ClueWeb12[1] consists of roughly 733 million Web documents ranging from blogs to news articles. From 14K documents containing 2,422,108,179 sentences retrieved from ClueWeb12, we extract 79,574,542 relations from using a manually composed list of 40 PROMOTE (e.g. *increase, cause, raise*) and 76 SUPPRESS (e.g. *harm, kill, prevent*) predicates. We parse each document using Stanford CoreNLP [8] in order to utilize both dependency and named entity features. The details of our knowledge base can be found in Section 5.

At this time, we limit our extraction on a simple noun subject/direct objects opposed to passive sentences (e.g. *cancer is caused by smoking*).

4.3 Finding Toulmin Arguments

4.3.1 Data

Given the motion in the form of a triplet $I = (A_1, R, A_2)$, we first extract a set D of candidate triplets of data for the input motion I from the constructed knowledge base. As described in Section 3, data is defined as a statement that supports the input motion. We find a set of data triplets based on the following hypotheses:

- if the input motion is PROMOTE(X, Y), the supporting data can be in the following two forms: (i) PROMOTE(Y, Z), where the sentiment polarity of Z (henceforth, *sp*(Z)) is positive, or (ii) SUPPRESS(Y, Z), where *sp*(Z) is negative.
- if the input motion is SUPPRESS(X, Y), the supporting data can be either (i) PROMOTE(Y, Z), where

[1] http://www.lemurproject.org/clueweb12.php/

sp(Z) is negative, or (ii) SUPPRESS(Y, Z), where *sp*(Z) is positive.

For example, given the input motion *ban(house, alcohol)*, where *ban* is a SUPPRESSrelation, we extract (i) all PROMOTE relations in which its A_1 is *alcohol* and $sp(A_2)$ is negative (e.g., *cause(alcohol, liver disease)*), and (ii) SUPPRESS relations in which its A_1 is *alcohol* and $sp(A_2)$ is positive (e.g., *decrease(alcohol, life expectancy)*). We elaborate more on our sentiment calculation in Section 5.1.

4.3.2 Warrant and Backing

For each $d \in D$, we extract a set W_d of candidate warrants using the similar hypotheses in the data extraction step. As described in Section 3, warrant serves as the supporting evidence of d being a reason for the input motion I. For example, we need to find a statement that explains why *alcohol promotes lung cancer* supports *house should ban alcohol* (in this case, the statement such as *lung cancer causes death* can be a warrant).

To capture warrant of data d, we apply the following hypotheses if the input motion I is PROMOTE(X, Y):

- if d is PROMOTE(Y, Z), where *sp*(Z) is positive, the warrant of d can be either: (i) PROMOTE(Z, W), where $sp(W)$ is positive, or (ii) SUPPRESS(Z, W), where $sp(W)$ is negative.
- if d is SUPPRESS(Y, Z), where *sp*(Z) is negative, the warrant of d can be either: (i) PROMOTE(Z, W), where $sp(W)$ is negative, or (ii) SUPPRESS(Z, W), where $sp(W)$ is positive.

Similarly, if the input motion I is SUPPRESS(X, Y), the following rules are applied:

- if d is PROMOTE(Y, Z), where *sp*(Z) is negative, the warrant of d can be either: (i) PROMOTE(Z, W), where $sp(W)$ is negative, or (ii) SUPPRESS(Z, W), where $sp(W)$ is positive.
- if d is SUPPRESS(Y, Z), where *sp*(Z) is positive, the warrant of d can be either: (i) PROMOTE(Z, W), where $sp(W)$ is positive, or (ii) SUPPRESS(Z, W), where $sp(W)$ is negative.

For example, for the input motion *ban(house, alcohol)* and data *cause(alcohol, liver disease)*, we would have as a result *cause(liver disease, death)*, *suppress(liver disease, metabolism)*.

For extracting a set B_d of backing candidates for d, we used the W_d as B_d because backing is different from warrants in terms of informativeness. That is, backing also serves as the evidence of d as being a reason for the input motion I, but backing should be based on statistical data, such as a white paper.

4.3.3 Toulmin Instantiation

So far, we have a set D of candidate data, and for each $d \in D$, we have a set W_d, B_d of candidate warrants and backings. Extracting the source sentences of each set, we construct a set M of Toulmin instantiations, where $M = \{(miSent(sents(d)), liSent(sents(w)), miSent(sents(b)) \mid d \in D, w \in W_d, b \in B_d\}$, $sents(t)$ is a set of source sentences of t, *miSent(X)* is a function to return the most informative sentence among X, and *liSent(X)* is a function to return the least informative sentence among X. In our experiment, we simply define the informativeness of sentence based on the number of named entities in the sentence and the sentence length.

5 Experiment and Observations

Given four topic motions *ban(House, alcohol‖boxing‖gambling‖homework)*, we construct multiple Toulmin instantiations and present our results and findings for the method proposed in Section 4. In this experiment, we collect the top 10 most frequent relations for data and the top 5 most frequent relations for warrant and backing for evaluation.

5.1 Sentiment Polarity Calculation

For calculating the sentiment of each argument's head noun, we use SentiWordNet [2], Takamura et al. [14]'s sentiment corpus, and the Subjectivity Lexicon [17]. For each corpus, we assign a value of 1.0 if the sentiment is positive, -1.0 if negative, or otherwise neutral. We base positive and negative as a value greater than 0 and less than 0, respectively. In the case of SentiWordNet, we focus only on the top-ranked synset polarity value for each noun. Afterwards, we combine the values per noun and calculate sentiment using the following:

$$sp(w)=\left\{\begin{array}{l} pos \text{ if } num\_pos\_votes(w) \geq 2 \\ neg \text{ if } num\_neg\_votes(w) \leq -2 \\ neutral \text{ otherwise} \end{array}\right\}, \text{ where } w \text{ is}$$

the head noun of the direct object in each PROMOTE and SUPPRESS relation.

5.2 Results

The results of our knowledge base construction are shown in Table 1. *Positive*, *Negative*, and *Neutral* refer to the number of relations in which a relation's A_2 sentiment is positive, negative, and neutral, respectively.

Table 1: PROMOTE (PR) and SUPPRESS (SP) relations from our data set.

Type	Positive	Negative	Neutral	Total
PR	5,515,661	2,141,886	61,746,810	69,404,357
SP	452,673	467,182	9,250,330	10,170,185
Total	5,968,334	2,609,068	70,997,140	79,574,542

We manually evaluate our output and base our precision measure for both data, warrant, and backing on whether a each data is consistent with a claim, whether warrant is a valid link between claim and data, and whether backing is appropriate for the resulting data. Overall, we acheive a reasonable precision for data; however, in the case of warrant and backing, our system's performance was quite poor (as shown in Table 2).

We show a list of fully correct Toulmin instantiations in Table 3. We note that the resulting Toulmin instantiations for *alcohol* and *dehydration* resulted in consistent results. In other cases, as shown in Table 4, we found that many examples were not consistent due

Table 2: Precision of output

ban(A_1, A_2)	Data	Warrant	Backing
A_2=alcohol	0.90 (9/10)	0.24 (9/38)	0.14 (4/28)
A_2=boxing	0.57 (4/7)	0.16 (5/32)	0.31 (5/16)
A_2=gambling	0.56 (5/9)	0.15 (6/40)	0.13 (4/32)
A_2=homework	0.80 (4/5)	0.27 (8/30)	0.20 (5/25)
Total	0.71	0.20	0.18

Table 3: Sample of correct output for the topic *alcohol*

Argument	Sentence
Data	"Alcohol causes dehydration of your body cells", says John Brick, P.h.D...
Warrant	Dehydration can increase risk of falls.
Backing	Dehydration at any age can also increase the risk of ocular dryness, says Rachel Bishop, MD, chief of the consult services section at the National Eye Institute.
Warrant	Dehydration decreases your ability to do physical work.
Backing	Dehydration impairs the ability to regulate body temperature, reduces mental and physical work performance, and increases susceptibility to heat injuries/illnesses.
Warrant	Dehydration causes fatigue.
Backing	"Dehydration causes fatigue, and fatigue causes poor posture", says Falsone.

Table 4: Sample of incorrect output for various topics

Ex	Topic	Argument	Sentence
1	Boxing	Data	Boxing and full-contact martial arts pose an extremely high risk of serious and even blinding eye injuries.
		Warrant	High risks and threshold bring high returns.
2	Homework	Data	If you are worried that your homework may be affecting your health or safety, or that of your family, contact your union or an advice agency.
		Warrant	Health, balance and symmetry enhance the quality of any dog.
		Backing	(US National Center for Health) Using safe cleaners can improve the quality of both your working environment and the world at large.
3	Homework	Data	Homework inherently causes problems.
	Alcohol	Data	But for both men and women, alcohol can also bring on problems with the heart and arteries, as reported by MSNBC .
		Warrant	These problems affect people's activities of daily living, cause falls and injuries, and lead to depression and social isolation.

to relations alone lacking a sentence's contextual information. In the case of Example 1, we extracted the relation PROMOTE(boxing, high risk); therefore, all warrant arguments were related around *high risks* either promoting or suppressing something negative or positive, respectively. This produced undesirable results, as the phrase *eye injuries* was ignored but relevant to the term *high risks*. Furthermore, from Example 2, we see that although the data states that *homework affects health*, from the warrant, we see that *health enhances quality of a dog*. In addition to relations alone, we need world knowledge to provide information that *homework* is typically done by a *human*, for instance. Finally, from Example 3, we discover the phrase *these problems* in a data argument; however, we need to employ coreference resolution into our future work in order to determine if *these problems* are directly related to *heart and arteries problems* in order to acheive higher precision.

6 Conclusion and Future Work

In this work, we initiated a preliminary study composed of developing a computational model for the instantiation of Toulmin models given a topic motion keyword. We evaluated our system output and found that although data had nearly 70% precision, our precision for warrant and backing suffered due to inconsistencies with both coreference information and contextual information. In future work, we will expand upon our PROMOTE and SUPPRESS keyword list, and we will experiment with state-of-the-art relation extraction technologies, as our current implementation is based on simple extraction rules. In addition, we will expand upon the contextual information presented in a topic motion, such as location, in order to construct Toulmin models for more elaborative topics such as *ban alcohol in Japan*. In addition to coreference resolution, we would also like to extract relations without an exact match restriction using various technologies (e.g. word vectors). Finally, we will experiment with identifying qualifier and rebuttal into our computational model in future work.

Acknowledgments

This work was supported by the MEXT Research Student Scholarship and by the JSPS KAKENHI Grant No. 23240018.

References

[1] I. D. E. Association and R. Trapp. *The Debatabase Book: A Must-have Guide for Successful Debate*. International Debate Education Association, 2009.

[2] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *in Proc. of LREC*, 2010.

[3] F. Boltužić and J. Šnajder. Back up your stance: Recognizing arguments in online discussions. In *Proc. of the First Workshop on Argumentation Mining*, pages 49–58, 2014.

[4] A. Fader, S. Soderland, and O. Etzioni. Identifying relations for open information extraction. In *Proc. of EMNLP*, EMNLP '11, pages 1535–1545, 2011.

[5] V. W. Feng and G. Hirst. Classifying arguments by scheme. In *Proc. of ACL: HLT - Volume 1*, HLT '11, pages 987–996, 2011.

[6] C. Hashimoto, K. Torisawa, S. De Saeger, J.-H. Oh, and J. Kazama. Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In *Proc. of Joint Conference on EMNLP and CoNLL*, EMNLP-CoNLL '12, pages 619–630, 2012.

[7] R. Levy, Y. Bilu, D. Hershcovich, E. Aharoni, and N. Slonim. *Proc. of COLING 2014: Technical Papers*, chapter Context Dependent Claim Detection, pages 1489–1500. 2014.

[8] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proc. of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014.

[9] Mausam, M. Schmitz, R. Bart, S. Soderland, and O. Etzioni. Open language learning for information extraction. In *Proc. of the 2012 Joint Conference on EMNLP and CoNLL*, EMNLP-CoNLL '12, pages 523–534, 2012.

[10] J. Mrozinski, E. Whittaker, and S. Furui. Collecting a why-question corpus for development and evaluation of an automatic QA-system. In *Proc. of ACL: HLT*, pages 443–451, 2008.

[11] J.-H. Oh, K. Torisawa, C. Hashimoto, M. Sano, S. De Saeger, and K. Ohtake. Why-question answering using intra- and inter-sentential causal relations. In *Proc. of ACL: long papers*, pages 1733–1743, 2013.

[12] P. Reisert, J. Mizuno, M. Kanno, N. Okazaki, and K. Inui. A corpus study for identifying evidence on microblogs. In *Proc. of LAW VIII*, pages 70–74, 2014.

[13] S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

[14] H. Takamura, T. Inui, and M. Okumura. Extracting semantic orientations of words using spin model. In *Proc. of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 133–140, 2005.

[15] S. Verberne. Developing an approach for why-question answering. In *Proc. of EACL: Student Research Workshop*, EACL '06, pages 39–46, 2006.

[16] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.

[17] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of HLT and EMNLP*, HLT '05, pages 347–354. Association for Computational Linguistics, 2005.

[18] A. Yates, M. Cafarella, M. Banko, O. Etzioni, M. Broadhead, and S. Soderland. Textrunner: Open information extraction on the web. In *Proc. of HLT: NAACL: Demonstrations*, NAACL-Demonstrations '07, pages 25–26, 2007.