

# Bayesian Argumentation-Scheme Networks: A Probabilistic Model of Argument Validity Facilitated by Argumentation Schemes

Takahiro Kondo<sup>1</sup>, Koki Washio<sup>1</sup>, Katsuhiko Hayashi<sup>2</sup> and Yusuke Miyao<sup>1</sup>

<sup>1</sup>The University of Tokyo

<sup>2</sup>Gunma University

{takahiro\_kondo, kwashio, yusuke}@is.s.u-tokyo.ac.jp

khayashi0201@gmail.com

## Abstract

We propose a methodology for representing the reasoning structure of arguments using Bayesian networks and predicate logic facilitated by argumentation schemes. We express the meaning of text segments using predicate logic and map the boolean values of predicate logic expressions to nodes in a Bayesian network. The reasoning structure among text segments is described with a directed acyclic graph. While our formalism is highly expressive and capable of describing the informal logic of human arguments, it is too open-ended to actually build a network for an argument. It is not at all obvious which segment of argumentative text should be considered as a node in a Bayesian network, and how to decide the dependencies among nodes. To alleviate the difficulty, we provide abstract network fragments, called idioms, which represent typical argument justification patterns derived from argumentation schemes. The network construction process is decomposed into idiom selection, idiom instantiation, and idiom combination. We define 17 idioms in total by referring to argumentation schemes as well as analyzing actual arguments and fitting idioms to them. We also create a dataset consisting of pairs of an argumentative text and a corresponding Bayesian network. Our dataset contains about 2,400 pairs, which is large in the research area of argumentation schemes.

## 1 Introduction

We propose to formalize the reasoning structure of an argument towards quantitative validity assessment. To assess the validity, we need to understand what kind of reasoning is applied. The following example is a discussion on whether the death penalty deters crime.

**P:** We should introduce the death penalty to deter crime.

**Q:** Criminologists do not believe that it deters crime.

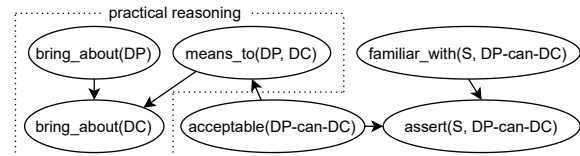


Figure 1: A Bayesian network on whether the death penalty deters crime. **DP**: introducing the death penalty, **DC**: deterring crime, **S**: criminologists.

P justifies the introduction of the death penalty as a means to deter crime. Q’s objection is justified based on expert opinion. By understanding how the argument is justified, we can identify more specific issues that need to be validated. For example, we can ask whether the goal of deterring crime is worth achieving, or whether the criminologist’s area of expertise is really relevant to the assertion.

We use Bayesian networks and predicate logic to represent the reasoning structure of an argument as shown in Fig. 1. We express the meaning of text segments using predicate logic. For example, ‘the death penalty is a means to deter crime’ is expressed as *means\_to(the death penalty, deterring crime)*. Then, we represent the dependencies between text segments using a Bayesian network. For example, whether or not a crime can be deterred depends on whether or not the death penalty has a deterrent effect on crime and whether or not the death penalty is actually introduced. This formalism is expressive and flexible to represent a variety of human justifications as it just models conditional dependencies. However, due to its flexibility, it is too open-ended to build a network for an argument. Why do we use the predicate *means\_to* for a node in the example network? How can we decide the dependencies among nodes?

To facilitate a network construction, we prepare abstract network fragments, called *idioms*, which represent typical argument justification patterns derived from argumentation schemes (Walton et al.,

2008). For example, one of the argumentation schemes, *practical reasoning*, is defined as follows.

**Major Premise:** I have a goal  $G$ .

**Minor Premise:** Carrying out this action  $A$  is a means to realize  $G$ .

**Conclusion:** Therefore, I ought (practically speaking) to carry out this action  $A$ .

We define an idiom for the justification of means to an end referring to practical reasoning. The dotted line in Fig. 1 uses this idiom. By preparing idioms, the network construction process is decomposed into **idiom selection**, **idiom instantiation**, and **idiom combination**. We define 17 idioms in total by referring to argumentation schemes as well as analyzing actual arguments and fitting idioms to them.

We also create a dataset consisting of pairs of an argumentative text and a corresponding Bayesian network. The problem is that annotating argumentation schemes in argumentative texts is too expensive. In addition to identifying justification patterns, we need to represent the reasoning structure with Bayesian networks, which is almost impossible. We solve the problem by adopting a procedure to generate a text from a built network inspired by OVERNIGHT, the data building method in semantic parsing (Wang et al., 2015). As a result, we can make about 2,400 argument-network pairs, which is large in the research area of argumentation schemes. Our dataset is publicly available at <https://github.com/mynlp/basn>.

In summary, our contribution is twofold.

- We propose a methodology for representing the reasoning structure of arguments using Bayesian networks and predicate logic facilitated by argumentation schemes. (Sec. 3)
- We create a dataset consisting of pairs of an argumentative text and a corresponding Bayesian network while reducing an annotation effort. (Sec. 4)

## 2 Related Work

### 2.1 Argument Validity Assessment

Wachsmuth et al. (2017) divided argumentation quality into three main dimensions, *Cogency*, *Effectiveness*, and *Reasonableness*, through comprehensive survey. Our method mainly focuses on cogency which relates to the acceptability of premises

and the relevance of a premise to a conclusion. Reasonableness is about acceptability in the dialectical context. Our method can serve in the dialectical context because we can model the validity of an argument and its counterargument together in the same Bayesian network. On the other hand, we do not focus on effectiveness which relates to an emotional appeal and the style of an argument.

The previous models for quantitatively assessing the quality of an argument, no matter which dimension of quality is being addressed, predict a score with either feature-based machine learning (Persing et al., 2010; Persing and Ng, 2013, 2015; Ghosh et al., 2016; Wachsmuth et al., 2016; Wachsmuth and Stein, 2017; Ke et al., 2019; Wachsmuth and Werner, 2020) or neural networks (Ke et al., 2018; Lauscher et al., 2020). In contrast, our method uses a probabilistic model to calculate the validity.

In argumentation theory, the probabilistic model for argument assessment is not mature. Baroni et al. (2018) provide a comprehensive survey of formal argumentation. In chapter 2 of the book, Prakken states ‘systematic studies of the combination of argumentation with probability were sparse.’ Our method is based on Bayesian networks. The use of Bayesian networks in argumentation has been studied through practice rather than theory, as in the case of the legal field discussed next.

### 2.2 Bayesian Networks with Argumentation Schemes

Argumentation schemes describe typical justification patterns in an argument. Walton et al. (2008) lists 60 representative argumentation schemes from various literature, and many studies refer to their list. In the list, argumentation schemes are defined by premises and a conclusion as shown in Sec. 1 and express the claim that a conclusion is plausible if all premises are true.

In the legal domain, argumentation schemes help to construct a Bayesian network though network construction is manual. In the analysis of court cases, it has been attempted to calculate the validity of hypotheses based on evidence or testimony using Bayesian networks (Fenton et al., 2013; Vlek et al., 2014). Since the arbitrariness of the network construction procedure is an issue, Timmer (2017) transforms an argumentation scheme that can be used in court cases into a fragment of a Bayesian network which provides a typical network structure. While the research shows that Bayesian networks

and argumentation schemes are compatible, argumentation schemes are used for assisting manual network construction by humans. It does not aim to automatically construct a network from the argumentative text. We combine Bayesian networks with predicate logic towards automatic network construction using natural language processing. In addition, we also prepare fragments for the argumentation schemes that are not covered in the legal domain, such as practical reasoning.

### 2.3 Datasets

A few datasets annotate argument structures in argument texts for the purpose of argument evaluation. [Stab and Gurevych \(2014\)](#) annotate argument components (major claim, claim, and premises) and relations (support and attack) in persuasive essays. [Ghosh et al. \(2016\)](#) apply the same annotation as [Stab and Gurevych \(2014\)](#) to TOEFL essays associated with holistic scores (high/medium/low). Our dataset consists of pairs of a short argumentative text and its corresponding Bayesian network for modeling the validity of arguments.

The major challenge in dataset creation is the difficulty of annotating argumentation schemes. Among existing datasets, the AraucariaDB corpus ([Reed et al., 2008](#)) is the largest as far as we know and includes approximately 660 manually annotated arguments ([Feng and Hirst, 2011](#)). [Lawrence et al. \(2019\)](#) state that annotated corpora of argumentation schemes are scarce, small, and unrepresentative. They provide an annotation tool to address the issue. We adopt a different approach inspired by [Wang et al. \(2015\)](#), in which they create training data for a semantic parser by generating canonical utterances from logical forms and paraphrasing them using crowdsourcing to reduce annotation effort. We employ a procedure to generate an argument text from a pre-built Bayesian network.

## 3 Bayesian Argumentation-Scheme Networks

Our goal is to represent the reasoning structure of arguments with Bayesian networks towards an argument validity assessment. Fig. 2 shows the overall process of network construction. First, we prepare abstract Bayesian network fragments which represent typical argument justification patterns derived from argumentation schemes. We call the abstract fragments *network idioms*. Then, through the analysis of an argument, appropriate idioms are

selected and embodied by assigning logical variables. We call the embodied fragments *network instances*. Finally, by combining the network instances, we can construct a Bayesian network for the entire argument.

In the following, we first describe a formalism of an argument network. Second, we explain how network idioms are defined and introduce three representative idioms. Third, we provide network combination rules to construct an entire argument network. Finally, we explain the advantages of our methodology.

### 3.1 Argument Reasoning Structure with Bayesian Networks and Predicate Logic

We formalize the reasoning structure of arguments with Bayesian networks and typed predicate logic. We represent a network node using a boolean random variable of a predicate logical expression. Logical variables (i.e. predicate arguments) are typed. A type is formally equivalent to using a specific (pre-defined) predicate to impose constraints on a logical variable. Edges represent conditional dependencies among boolean random variables.

### 3.2 Network Idioms

Idioms are abstract Bayesian network fragments that represent typical argument justification patterns. Argumentation schemes help us to identify justification patterns. We can determine the network structure of an idiom according to the characteristic of justification. For example, practical reasoning deals with the causal relationship between means and ends. We connect an edge from means to ends along the direction of causality. In addition, argumentation schemes can also be used as a reference to decide predicates of idioms. For example, one of the premises in practical reasoning, ‘Carrying out this action  $A$  is a means to realize  $G$ ’, can be expressed as  $means\_to(G, A)$  in predicate logic. Thus, we regard  $means\_to$  as a candidate of predicates.

In defining idioms, we refer to [Walton et al. \(2008\)](#)’s list of argumentation schemes as well as analyze actual arguments and fit idioms to them. In this work, we analyze arguments in ProCon.org<sup>1</sup>, which presents sourced pros and cons of debatable issues. As a result, we define a total of 17 idioms using a finite set of predicates and variable types. Currently, we define idioms for the justification

<sup>1</sup><https://www.procon.org/>

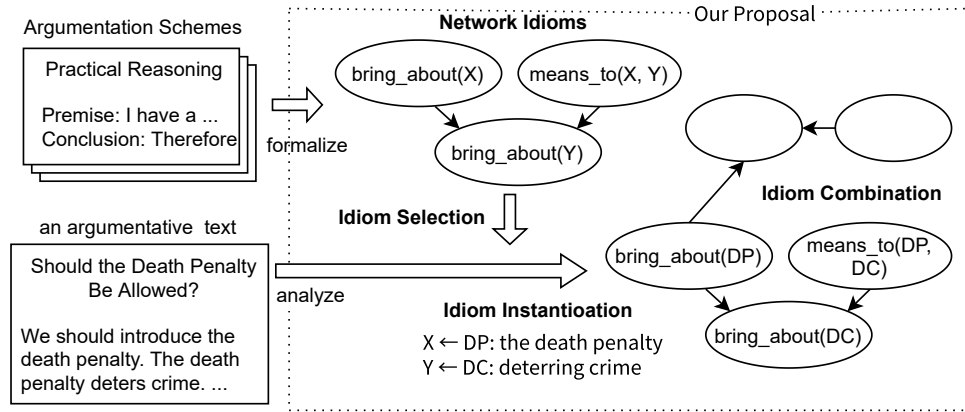


Figure 2: Overall Process of Our Methodology

patterns that appear in the actual argument we analyze. Our idiom covers 25 out of 60 argumentation schemes in the [Walton et al. \(2008\)](#)'s list.

Regarding the design of idioms, unlike annotation schemes of argument mining datasets ([Stab and Gurevych, 2017, 2014](#); [Peldszus and Stede, 2015](#)), the direction of an edge is not necessarily from a premise to a conclusion. The reason is that an idiom is designed to express the reasoning structure behind the justification of an argument. Although the justifications described by argumentation schemes often involve logical leaps, we can sometimes find a deductive relationship among premises and a conclusion. For example, practical reasoning concludes 'I ought to carry out this action  $A$ ' from the two premises, 'I have a goal  $G$ ' and 'carrying out this action  $A$  is a means to realize  $G$ .' That justification is fallacious from a formal logic perspective. Instead, if the two propositions, 'I carry out this action  $A$ ' and 'carrying out this action  $A$  is a means to realize  $G$ ', are true, then ' $G$  is realized' is deductively true. By expressing such a relationship among premises and a conclusion in argumentation schemes, we model what reasoning is behind them.

In this section, we introduce three representative idioms. All idioms and the definition of variable types are shown in [Appendix A](#), [Appendix B](#) respectively. In the following explanations, the types of logical variables are omitted. In the figures shown below, a node symbol  $p$  corresponds to a premise and a symbol  $c$  to a conclusion.

### Idiom 1: Means for Goal

This idiom expresses the justification of means to an end. [Fig. 3](#) shows the network structure. A conclusion node is  $c$ :  $bring\_about(X)$ .

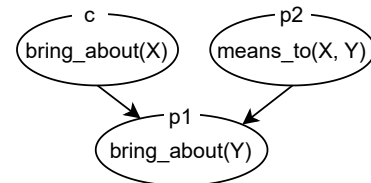


Figure 3: Means for Goal Idiom

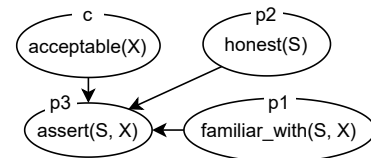


Figure 4: From Source Knowledge Idiom

Two premise nodes are  $p_1$ :  $bring\_about(Y)$  and  $p_2$ :  $means\_to(X, Y)$ . The edges are from  $c$  to  $p_1$  and from  $p_2$  to  $p_1$ . It expresses the relationship that  $X$  causes  $Y$ .

The description in the argumentation-scheme-like format is as follows.

**Premise1:** Bringing about  $Y$  is desirable.

**Premise2:** Bringing about  $X$  is a means to realize  $Y$ .

**Conclusion:** We should bring about  $X$ .

In the list of [Walton et al. \(2008\)](#), *practical reasoning* corresponds to this idiom. Moreover, *argument from need for help* and *argument from distress* also correspond by interpreting them as a justification of an action aimed at solving the problem.

### Idiom 2: From Source Knowledge

This idiom is about justification based on the knowledge of a source. [Fig. 4](#) shows the network structure. This idiom has four essential nodes. A con-

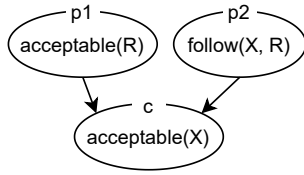


Figure 5: Rule or Principle Idiom

clusion node is  $c: acceptable(X)$ . Three premise nodes are  $p_1: familiar\_with(S, X)$ ,  $p_2: honest(S)$ , and  $p_3: assert(S, X)$ . The edges are from  $c$  to  $p_3$ , from  $p_1$  to  $p_3$ , and from  $p_2$  to  $p_3$ . It expresses the relationship between a hypothesis and evidence. A source’s assertion is the evidence that a hypothesis is true. This idiom has optional premise nodes  $o_1: position\_to\_know(S, X)$  and  $o_2: expert\_in(S, X)$ . Both nodes back up the source’s knowledge by linking them to  $p_1$ .

The description in the argumentation-scheme-like format is as follows.

- Premise1:**  $S$  is familiar with  $X$ .
- Premise3:**  $S$  is honest.
- Premise3:**  $S$  asserts  $X$ .
- Conclusion:**  $X$  is acceptable.

In the list of Walton et al. (2008), *argument from position to know*, *argument from expert opinion*, and *argument from witness testimony* correspond to this idiom.

### Idiom 3: Rule or Principle

This idiom is about justification based on whether a claim is in accordance with rules or norms. Fig. 5 shows the network structure. A conclusion node is  $c: acceptable(X)$ . Two premise nodes are  $p_1: acceptable(R)$  and  $p_2: follow(X, R)$ . The edges are from  $p_1$  to  $c$  and from  $p_2$  to  $c$ . It expresses the relationship between a general rule and a specific case. If the case is subject to the rule, the acceptability of the case follows the acceptability of the rule.

The description in the argumentation-scheme-like format is as follows.

- Premise1:**  $R$  is acceptable.
- Premise2:**  $X$  follows  $R$ .
- Conclusion:**  $X$  is acceptable.

There is a negation variant for this idiom.

- Premise1:**  $R$  must be obeyed.
- Premise2:**  $X$  does not follow  $R$ .
- Conclusion:**  $X$  is not acceptable.

In the list of Walton et al. (2008), no argumentation scheme corresponds to this idiom. One of the argumentation schemes, *argument from rule* assumes that a rule describes specific actions to follow. However, in the actual arguments we analyze, the main point at issue is whether individual cases fall under the rules like the Constitution.

### 3.3 Combination Rules

A Bayesian network for an entire argument often contains multiple network instances because multiple justifications may be combined, or a counterargument may be listed together. To construct an entire network, we need to combine network instances. We define the following four rules and *default inference* to merge or link nodes.

#### Rule 1: merge the nodes with the same logical expressions

Merge nodes if the logical expression of nodes is equivalent.

#### Rule 2: link a node to its negational form

Link a node to the node with the negational form. For example, link  $acceptable(X)$  to  $\neg acceptable(X)$ . The edge can be either direction. A boolean value is inverted through the edge.

#### Rule 3: link a propositional expression to its predicate expression

The same content can be expressed in two ways, predicates or propositions. For example,  $means\_to(X, Y)$  has the same meaning as  $acceptable(P)$  if  $P$  is a proposition that  $X$  is a means to  $Y$ . In such case, we link a propositional expression like  $acceptable(P)$  to a predicate expression like  $means\_to(X, Y)$ . A boolean value remains unchanged through the edge.

#### Rule 4: link nodes whose variables are synonymous

Link nodes if the predicates of two nodes are the same and logical variables are of synonymous meaning. For example, link  $bring\_about(X)$  to  $bring\_about(X')$  if the meaning of  $X$  is synonymous to that of  $X'$ . A boolean value remains unchanged through the edge.

#### Default Inference

Besides the formal rules, we may need to link nodes according to the reasoning that does not fit into any of the justification patterns of prepared network

idioms. When Visser et al. (2018) annotated argumentation schemes, they also define default inference as not fitting any of the 60 argumentation schemes.

### 3.4 Advantages of Our Method

The advantages of our method are threefold.

#### **Incorporating Argumentation Schemes into NLP Models for Argument Evaluation**

Argumentation schemes are useful for validating arguments as they help to teach critical thinking skills in pedagogy (Reed and Walton, 2001). However, few NLP models apply argumentation schemes to argument validity assessment. One reason for this could be that the mathematical formalism of argumentation schemes had not been established. We hope that our proposal will encourage more research on incorporating argumentation schemes into NLP models.

#### **Extending Existing Techniques of Formal Semantic Representation to Discourse Analysis**

NLP techniques for transforming the meaning of a text into a formal representation are mainly based on sentences as the unit of analysis. Empirical research for applying discourse-level formal representation to real-world examples has not progressed. Although our proposal is limited to the analysis of arguments, we extend existing techniques to discourse analysis.

#### **Flexibility to Describe Human’s Informal Logic and Applicability to Actual Arguments**

Human reasoning to justify arguments differs from formal logic. In general, a conclusion is not a logical consequence of premises. Humans make a variety of justifications expanding on deductive, inductive, and hypothetical reasoning. It is difficult to explain them in a unified manner. Our formalism is flexible enough to describe human reasoning because it just expresses the relationship between premises and conclusions with conditional probabilities. Though due to its flexibility, there is a large degree of freedom to express argument structure, the network can be patterned according to justification patterns provided by argumentation schemes.

## 4 Dataset Creation

We create a dataset consisting of pairs of relatively short English argument texts and their corresponding Bayesian networks. Our dataset not only helps

us to train a machine learning model but also empirically shows that our approach can model the reasoning structure of actual arguments.

As described in Sec. 2.3, annotating argumentation schemes is very costly. In addition, our proposed method requires annotating a network structure. Therefore, it is difficult to prepare a large amount of high-quality data by the procedure of giving an argument text and annotating a network. In this work, we adopt the procedure of generating an argument text from a pre-built network.

### 4.1 Dataset Creation Procedure

Dataset creation is done in the following steps.

#### **Create a Bayesian network from a wide range of discussions on a particular topic.**

ProCon.org is a website that collects a wide range of opinions for and against controversial issues, mainly in the United States. We select six topics (the death penalty, gun control, minimum wage, legal abortion, school uniforms, and nuclear power) that are also selected in the existing argument mining dataset (Stab et al., 2018). We divide each topic into several subtopics. For each subtopic, we summarize the main points of discussion on the site and create a Bayesian network by applying network idioms. For example, for the subtopic on the death penalty, ‘Does the death penalty deter crime?’, we create a network like Fig. 1. We also create a text corresponding to the network we have created. This procedure is carried out by the author.

#### **Decompose the built network into network fragments.**

We decompose the built Bayesian network, which represents the entire argument of each subtopic, into network fragments so that one fragment contains only one network instance as much as possible. For example, from the network of Fig. 1, only the part surrounded by the dotted line is cut out.

#### **Convert a network fragment into premises and a conclusion.**

The predicate logical expression of each node in the network instance is converted into natural language and divided into premises and a conclusion as the description of argumentation schemes. For example, the predicate expression *means\_to*(the death penalty, deterring crime) is converted to ‘The death penalty is a means to deter crime’ and classified as a premise. This pro-

Table 1: The number of premises-conclusion patterns. Patterns are counted by topic and by idiom.

Idiom \ Topic	death-pena	gun-ctrl	min-wage	abortion	sc-uniform	nuclear	all
Means for Goal	4	7	2	5	0	2	20
Goal from Means	0	0	0	0	0	5	5
From Consequence	14	11	19	10	19	5	78
Source Knowledge	4	0	0	0	2	0	6
Source Authority	0	1	0	3	2	0	6
Rule or Principle	8	4	3	4	0	0	19
Others	5	3	0	0	1	4	13
<b>Total</b>	35	26	24	22	24	16	147

cedure is almost done by a simple templating generation like ‘argument1 predicate argument2.’ The author adjusts it so that it is grammatically natural. As an example, the following premises and conclusion are generated.

**Premise1:** The death penalty is a means to deter crime.

**Premise2:** Deterring crime is desirable.

**Conclusion:** The death penalty is necessary.

### Generate an argument text from premises and a conclusion by crowdsourcing.

We present premises and a conclusion to crowdworkers and ask them to convert given contents into a natural language text while preserving the original meaning. In addition, the following two instructions are given to the crowdworkers.

You can omit phrases considered to be common sense like ‘Saving lives is desirable.’

Do not use the following phrases in your writing.

- follow, above, desirable, acceptable, necessary, true, questionable
- be means to/for, lead to, some may argue, either A or B
- X is evidence/example that Y, X is enough to achieve the goal
- respect to, obey, assert, credible, familiar, popular

The latter prohibits using idiom-derived keyphrases in the presented premises or conclusions.

Table 2: Subtopics of the death penalty in our dataset.

The death penalty
- Is the Death Penalty Immoral?
- Is the Death Penalty Unconstitutional?
- Has DNA Testing Led to Significant Improvements in the Criminal Justice System?
- Does the Death Penalty Deter Crime?
- Should the Death Penalty Be Used for Retribution?
- Does a Person’s Race Affect the Likelihood of Receiving the Death Penalty?
- Should the Death Penalty Be Abolished or Paused Because of the Alleged - - Possibility of an Innocent Person Being Executed?
- Is Life in Prison without Parole a Better Option Than the Death Penalty?

## 4.2 Results

We created networks for a total of 53 subtopics of 6 topics and extracted 147 network fragments from them. The subtopics of the death penalty are shown in Table 2. We use Amazon Mechanical Turk and ask 20 crowd workers to generate an argument text per premises-conclusion pair made from the network fragments. Workers were paid 0.15\$ per response. The number of subtopics per topic and the number of premises-conclusion patterns presented to the workers are shown in Table 1.

As a result of the crowdsourcing, 2,370 valid responses were obtained. 1,359 responses that did not follow the instructions were excluded as invalid responses. The vocabulary size and average sentence length (word counting) of the responses are shown in Table 3. The following metrics are measured in order to verify the diversity and bias of the created data since the procedure of present-

Table 3: Summary of crowdsourcing responses.

Topic	death-pena	gun-ctrl	min-wage	abortion	sc-uniform	nuclear	all
#responses	535	417	446	289	411	272	2,370
vocab size	1,052	776	549	692	735	609	2,493
sentence len	23.7	23.0	18.2	19.8	19.2	23.8	21.3

Table 4: Metrics of the diversity of our dataset.

Metric	death-pena	gun-ctrl	min-wage	abortion	sc-uniform	nuclear	all
<b>SIM-INOUT</b>	0.69	0.66	0.69	0.62	0.65	0.68	0.67
<b>SIM-RES</b>	0.57	0.54	0.59	0.48	0.54	0.55	0.55
<b>BECAUSE</b>	6.4	5.8	10.3	1.7	5.1	7.4	6.3

ing premises-conclusion pairs to the crowdworkers would result in a large bias. The results are shown in Table 4.

#### **SIM-INOUT: Cosine Similarity between Inputs and Responses in Crowdsourcing**

We measure the cosine similarity of the Bag of Words (BoW) representation between a premises-conclusion pair presented to crowdworkers and their response.

#### **SIM-RES: Cosine Similarity among Responses in Crowdsourcing**

We measure the BoW cosine similarity among the responses to the same premises-conclusion pair and average the values.

#### **BECAUSE: Percentage of ‘Conclusion because Premises’ Response Pattern**

A cloud worker may simply answer by combining premises with a conclusion using ‘because.’ We investigate the ratio of such responses. We divide responses into the first half and the second half separated by ‘because.’ Then, we calculate the BoW cosine similarity between the first half and a conclusion, and between the second half and premises. We compute the percentage of responses where both similarity values are more than 0.7.

### **4.3 Characteristics of Our Dataset**

The data size is large in the research area of argumentation schemes. We create 2,370 pairs of argumentative text and a corresponding Bayesian network in total. The number of arguments is larger than the largest existing dataset (Reed et al., 2008) annotated with argumentation schemes though the vocabulary size is smaller.

As for the diversity of the dataset, the bias is suppressed. Table 4 shows the values of metrics. The following example is the case where a SIM-INOUT (the BOW cosine similarity between a premises-conclusion pair and a response) is 0.67.

**Premises:** Gun ownership causes an accidental gun death. Accidental gun death is not desirable.

**Conclusion:** Gun ownership is not necessary.

**Worker’s Response:** Lots of accidental gun death results from gun ownership which is completely avoidable.

Since this is an average example in terms of the metric, we can infer that the responses are not biased toward uniformity. We also come to the same conclusion based on our observations of all the responses. Furthermore, the BECAUSE metric is 6.3, which indicates few workers used the response pattern of ‘Conclusion because Premises.’

### **4.4 A Baseline Model of Idiom Selection**

We set up a task to select idioms for argumentative texts to investigate how well the existing NLP model would work. We use idiom names listed in Table 1 as labels. Infrequent idioms are labeled as ‘others.’ We separate a dataset according to topics. The data for the death penalty is used as a test set and the rest as a training set.

We apply BERT for sequence classification (Devlin et al., 2019; Wolf et al., 2020) as a baseline. The pre-trained model is BERT-Base Uncased and is not finetuned. Since learning without no sampling leads to always predicting the most frequent idiom in training data, we employ



Table 5: The results of a baseline idiom classifier for a test set. Some idioms do not appear in the test set.

Idiom	Prec	Recall	F1
Means for Goal	0.283	0.374	0.287
Goal from Means	-	-	-
From Consequence	0.636	0.169	0.230
Source Knowledge	0	0	NA
Source Authority	-	-	-
Rule or Principle	0.321	0.735	0.430
Others	0.137	0.061	0.077

both oversampling and undersampling<sup>2</sup>. Parameters are optimized using AdamW (Loshchilov and Hutter, 2019). We perform 5fold cross-validation on the training set and determine hyperparameters. The learning rate is  $10^{-5}$  chosen from  $\{10^{-4}, 10^{-5}, 10^{-6}\}$  and the weight decay is  $10^{-3}$  chosen from  $\{10^{-2}, 10^{-3}, 10^{-4}\}$ . The max epoch is 400 and the batch size is 64. We repeat the experiment ten times. The averages of precision, recall, and f1 for each idiom are shown in Table 5. The average of overall accuracies for the test set is 27.5%. It indicates just selecting a suitable idiom is not an easy task.

## 5 Conclusion and Future Work

We propose a methodology for representing the reasoning structure of arguments using Bayesian networks and predicate logic. We define 17 network idioms derived from argumentation schemes to facilitate the network construction by giving typical network patterns. We also create a large dataset consisting of 2,370 pairs of argument text and a corresponding Bayesian network. In future work, we train a model for automatic network construction.

## Acknowledgements

We thank Naoya Inoue for providing information about literature on argumentation schemes and web resources for argumentative texts.

## References

Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre. 2018. *Handbook of formal argumentation*. College Publications.

<sup>2</sup><https://github.com/ufoym/imbalanced-dataset-sampler>

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Vanessa Wei Feng and Graeme Hirst. 2011. [Classifying arguments by scheme](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 987–996, Portland, Oregon, USA. Association for Computational Linguistics.

Norman Fenton, Martin Neil, and David A. Lagnado. 2013. [A general structure for legal arguments about evidence using bayesian networks](#). *Cognitive Science*, 37(1):61–102.

Debanjan Ghosh, Aquila Khanam, Yubo Han, and Smaranda Muresan. 2016. [Coarse-grained argumentation features for scoring persuasive essays](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 549–554, Berlin, Germany. Association for Computational Linguistics.

Zixuan Ke, W. Carlile, Nishant Gurrupadi, and Vincent Ng. 2018. [Learning to give feedback: Modeling attributes affecting argument persuasiveness in student essays](#). In *IJCAI*.

Zixuan Ke, Hrishikesh Inamdar, Hui Lin, and Vincent Ng. 2019. [Give me more feedback II: Annotating thesis strength and related attributes in student essays](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3994–4004, Florence, Italy. Association for Computational Linguistics.

Anne Lauscher, Lily Ng, Courtney Napoles, and Joel Tetreault. 2020. [Rhetoric, logic, and dialectic: Advancing theory-based argument quality assessment in natural language processing](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4563–4574, Barcelona, Spain (Online). International Committee on Computational Linguistics.

John Lawrence, Jacky Visser, and Chris Reed. 2019. [An online annotation assistant for argument schemes](#). In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 100–107, Florence, Italy. Association for Computational Linguistics.

I. Loshchilov and F. Hutter. 2019. [Decoupled weight decay regularization](#). In *ICLR*.

Andreas Peldszus and Manfred Stede. 2015. [An annotated corpus of argumentative microtexts](#). In *Proceedings of the First European Conference on Argumentation: Argumentation and Reasoned Action*, Lisbon, Portugal.

- Isaac Persing, Alan Davis, and Vincent Ng. 2010. [Modeling organization in student essays](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 229–239, Cambridge, MA. Association for Computational Linguistics.
- Isaac Persing and Vincent Ng. 2013. [Modeling thesis clarity in student essays](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 260–269, Sofia, Bulgaria. Association for Computational Linguistics.
- Isaac Persing and Vincent Ng. 2015. [Modeling argument strength in student essays](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 543–552, Beijing, China. Association for Computational Linguistics.
- C. Reed and D. Walton. 2001. Applications of argumentation schemes. In *Proceedings of the 4th Conference of the Ontario Society for the Study of Argument (OSSA2001)*.
- Chris Reed, Raquel Mochales Palau, Glenn Rowe, and Marie-Francine Moens. 2008. [Language resources for studying argument](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- 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, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Christian Stab and Iryna Gurevych. 2017. [Parsing argumentation structures in persuasive essays](#). *Computational Linguistics*, 43(3):619–659.
- Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018. [Cross-topic argument mining from heterogeneous sources](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics.
- Sjoerd T. Timmer. 2017. *Designing and Understanding Forensic Bayesian Networks using Argumentation*. Ph.D. thesis, Utrecht University.
- J. Visser, John Lawrence, Jean H. M. Wagemans, and C. Reed. 2018. Revisiting computational models of argument schemes: Classification, annotation, comparison. In *COMMA*.
- Charlotte Vlek, Henry Prakken, Silja Renooij, and Bart Verheij. 2014. [Building bayesian networks for legal evidence with narratives: a case study evaluation](#). *Artificial Intelligence and Law*, 22.
- Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2016. [Using argument mining to assess the argumentation quality of essays](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1680–1691, Osaka, Japan. The COLING 2016 Organizing Committee.
- Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. [Computational argumentation quality assessment in natural language](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187, Valencia, Spain. Association for Computational Linguistics.
- Henning Wachsmuth and Benno Stein. 2017. A universal model for discourse-level argumentation analysis. *ACM Transactions on Internet Technology (TOIT)*, 17:1 – 24.
- Henning Wachsmuth and Till Werner. 2020. [Intrinsic quality assessment of arguments](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6739–6745, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.
- Yushi Wang, Jonathan Berant, and Percy Liang. 2015. [Building a semantic parser overnight](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1332–1342, Beijing, China. Association for Computational Linguistics.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

## A All Idioms

### A.1 Cause to Effect

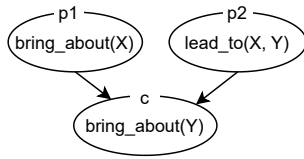


Figure 6: Cause to Effect Idiom

Variable type:  $X$  is Action or Case.  $Y$  is Case.  
 Referenced argumentation schemes: argument from cause to effect.

### A.2 From Consequence

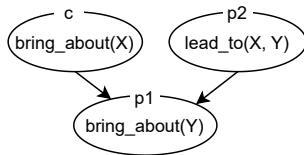


Figure 7: From Consequence Idiom

Variable type:  $X$  is Action or Case.  $Y$  is Case.  
 Referenced argumentation schemes: argument from consequence, pragmatic argument from alternatives, argument from threat, argument from fear appeal, argument from danger appeal.

### A.3 Goal from Means

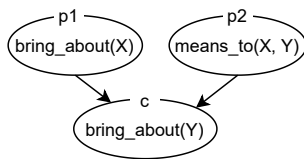


Figure 8: Goal from Means Idiom

Variable type:  $X$  is Action or Case.  $Y$  is Case.

### A.4 Means for Goal

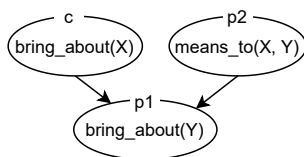


Figure 9: Means for Goal Idiom

Variable type:  $X$  is Action or Case.  $Y$  is Case.

Referenced argumentation schemes: practical reasoning, argument from need for help, argument from distress.

### A.5 From Source Knowledge

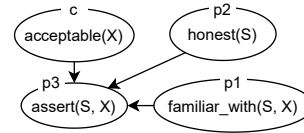


Figure 10: From Source Knowledge Idiom

Variable type:  $X$  is Proposition.  $S$  is Person(s) or Thing(s).

Referenced argumentation schemes: argument from position to know, argument from expert opinion, argument from witness testimony, argument from popular practice.

### A.6 From Source Authority

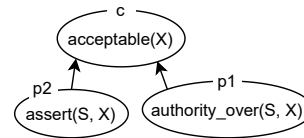


Figure 11: From Source Authority Idiom

Variable type:  $X$  is Proposition.  $S$  is Person(s) or Thing(s).

Referenced argumentation schemes: argument from popular opinion, argument from rules.

### A.7 Rule or Principle

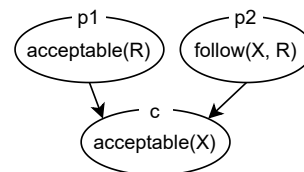


Figure 12: Rule or Principle Idiom

Variable type:  $R$  is Proposition.  $X$  is Action or Case.

## A.8 Interpretation

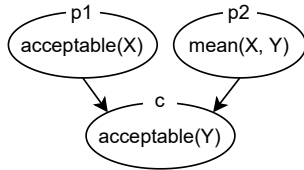


Figure 13: Interpretation Idiom

Variable type: Both  $X$  and  $Y$  are Proposition.

## A.9 From Evidence

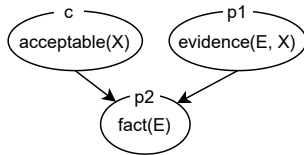


Figure 14: From Evidence Idiom

Variable type:  $X$  is Proposition.  $E$  is Case.  
Referenced argumentation schemes: argument from sign, argument from evidence to a hypothesis.

## A.10 From Example

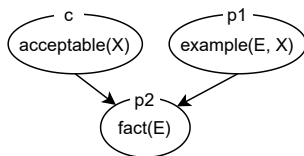


Figure 15: From Example Idiom

Variable type:  $X$  is Proposition.  $E$  is Case.  
Referenced argumentation schemes: argument from example.

## A.11 Constraint Idioms

The network structure of an idiom in this group is either Fig. 16 or Fig. 17. The predicate *CONSTRAINT* is defined for each idiom.

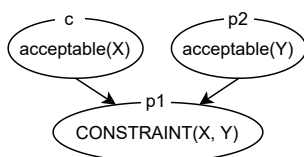


Figure 16: Constraint Idiom1

Variable type: Both  $X$  and  $Y$  are Action or Case.

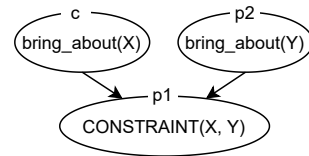


Figure 17: Constraint Idiom2

Variable type: Both  $X$  and  $Y$  are Action or Case.

### A.11.1 Either Constraint

Use *either* as *CONSTRAINT*.

Referenced argumentation schemes: argument from alternatives.

### A.11.2 Contradict Constraint

Use *contradict* as *CONSTRAINT*.

Referenced argumentation schemes: argument from oppositions.

### A.11.3 Consistent Constraint

Use *consistent* as *CONSTRAINT*.

Referenced argumentation schemes: pragmatic inconsistency, argument from inconsistent commitment.

### A.11.4 Comparative Constraint

Use *better\_than* as *CONSTRAINT*.

## A.12 Verbal Classification

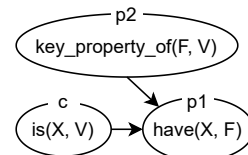


Figure 18: Verbal Classification Idiom

Variable type:  $X$  is Thins(s).  $V$  is Concept.  $F$  is Property.

Referenced argumentation schemes: argument from verbal classification, argument from definition to verbal classification.

## A.13 Verbal Evaluation

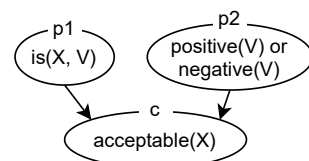


Figure 19: Verbal Evaluation Idiom

Variable type:  $X$  is Thins(s).  $V$  is Concept.

## A.14 Correlation to Cause

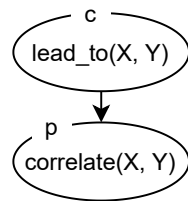


Figure 20: Correlation to Cause Idiom

Variable type:  $X$  is Action or Case.  $Y$  is Case.  
 Referenced argumentation schemes: argument from correlation to cause.

## B Variable Types

Table 6: Variable Types

Type	Description and Examples
Person(s)	A person or persons. e.g. Criminologists.
Thing(s)	A thing or things. e.g. The death penalty.
Action	An action. e.g. Introduce the death penalty.
Case	A situation or state of affairs. e.g. The death penalty has been introduced.
Proposition	A rule, an assertion, a fact, etc. e.g. The death penalty deters crime.
Concept	A concept or a term. e.g. Retribution.
Property	A characteristic of something. e.g. Punishment in proportion to serious crime.