Proceedings of the Sixteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2018)

Bayesian Model Selection in Statistical Construction of Justification

Hiroyuki Kido
Institute of Logic and Cognition
Sun Yat-sen University
No. 135, Xingang Xi Road, Guangzhou, 510275, P. R. China

Abstract

Argumentation mining involves identification of an attack relation between natural language sentences. Bayesian inference characterizing argument-based reasoning addresses this issue by calculating the posterior distribution over attack relations given acceptability statuses of arguments. This paper discusses the use of Bayesian model selection where graph-theoretic properties impose restrictions on the graphic structure of attack relations.

Introduction

Statistical construction of an explanation or justification is an important research issue of argumentation mining. This paper deals with a model selection problem associated with this issue. Suppose that Alice and Bob are planning to go out together and discussing where to go.

Alice Let's go to an opera show this weekend. (a)

Bob I feel like going to a soccer match. (b)

Alice I'm too sick to be outdoors for the soccer match. (c)

Alice I'm tired of watching losing games anymore. (d)

Given the above arguments, people would be able to guess that they form the argumentative structure shown on the center in Figure 1. However, this task is very difficult for machines even with state-of-the-art statistical natural language processing.

By contrast, this paper addresses the task by solving the inverse problem of argument-based reasoning. The inverse problem means to estimate the posterior distribution over attack relations given agents' beliefs regarding acceptability status of arguments. Now suppose that each agent accepts its own claims at the end of the argument. $\boldsymbol{acc}^A = \{a, c, d\}$ stands for Alice's belief, $\boldsymbol{acc}^B = \{b\}$ stands for Bob's one, and $\boldsymbol{acc} = (\boldsymbol{acc}^A, \boldsymbol{acc}^B)$ stands for their beliefs. What we want now is the posterior probability of an attack relation att given observed acceptability status $\boldsymbol{acc}$. Bayes' theorem allows the calculation as follows.

$$p(att|\boldsymbol{acc}) \quad \propto \quad p(\boldsymbol{acc}|att)p(att) \qquad (1)$$

However, we here need to manage a difficult problem causing poor explainability, predictability and time complexity.

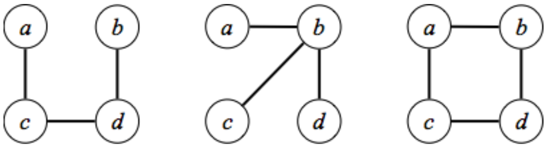

Figure 1: Path graph (left), tree graph (center) and connected bipartite graph (right) where each edge represents a symmetric attack relation.

In terms of the explainability and predictability, if all attack relations are taken into account then reflexive and transitive attack relations, for instance, become the subject of calculation. It would be difficult to expect, in general, that such unrealistic attack relations successfully explain observed acceptability status, and also predict unobserved acceptability status. Moreover, in terms of the time complexity, if all attack relations are taken into account then the calculation of expression (1) is analytically intractable in general. This is because the number of attack relations in consideration grows exponentially depending on the increase of the number of arguments. In fact, there are $2^{n\times n}$ attack relations given n arguments, and thus $65,536$ attack relations exist even in the case of $n = 4$. Sampling approaches for approximate posterior inference cannot fundamentally solve the problem in this situation because it is time-consuming to converge.

Method

We thus give the Bayesian network shown in Figure 2, and propose to use the Bayesian model selection technique, so-called empirical Bayes. Empirical Bayes methods assume models (i.e., deterministic parameters) defining a set (i.e., a hypothesis space) of attack relations defined in each model. They choose the best single model that maximizes the marginal likelihood. That is, we use the point estimate with type-II maximum likelihood in accordance with the following expression.

$$\begin{aligned}\hat{m} &= \arg\max_m p(\boldsymbol{acc}|m) \\ &= \arg\max_m \sum_{att} p(\boldsymbol{acc}|att)p(att|m) \qquad (2)\end{aligned}$$

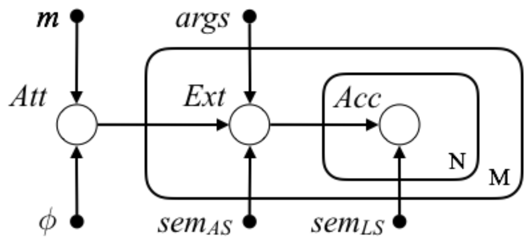

Figure 2: The Bayesian network for an argument model. m, ϕ, $args$, sem_{AS}, sem_{LS} are deterministic parameters for a model, an attack-relation prior, a set of arguments, acceptability semantics, and logical semantics, respectively. Att, Ext and Acc are random variables for attack relations, extensions and acceptability statuses, respectively.

The models we introduce here are graph-theoretic properties that impose restrictions on the graph structure of attack relations. They include from any directed graph, i.e., the most flexible model, to path graphs, i.e., relatively simple model. Now we suppose three undirected models: a path-graph model (i.e., connected line graphs) denoted by m_1, a tree-graph model (i.e., connected graphs without cycles) denoted by m_2, and a connected-bipartite-graph model (i.e., connected graphs without odd cycles), denoted by m_3. Figure 1 shows example graphs of each model.

The model $\hat{m}$ selected in expression (2) is intuitively a simple one that successfully explains acceptability status $\boldsymbol{acc}$. This intuition is formalized as a trade-off between data fitness corresponding to $p(\boldsymbol{acc}|att_i)$ and model complexity corresponding to $p(att_i|m)$.

For simplicity, we now assume that model complexity is defined with the number of possible attack relations. Given 4 arguments, there are 12 path graphs, 16 tree graphs and 19 connected bipartite graphs, and thus m_1 is the simplest model and m_3 is the most complex model. Although the simplest model m_1 is good in terms of model complexity, it does not have an attack relation that explains $\boldsymbol{acc}$. That is, for example, it makes no sense to think that the left attack relation in Figure 1 causes Alice and Bob's beliefs $\boldsymbol{acc}$. Next, the most flexible model m_3 is good in terms of data fitness because it includes the center (and correct) attack relation in Figure 1. However, it is not good in terms of model complexity because it includes other redundant attack relations including the right attack relation in Figure 1. In contrast to the models, m_2 is the best model relatively resolving the trade-off between model complexity and data fitness.

Correctness

The dataset we partially used is "Claim Stance Dataset (Bar-Haim et al. 2017)" provided by IBM Debater. 72 claims with topic "Violent Games" were collected from this dataset. We collected acceptability status of each of those claims from 100 anonymous individuals per each claim via an online survey. We averaged the statuses and then used the first 6 claims with negative status and 6 claims with positive status. We defined observations $\boldsymbol{acc}$ by those statuses, and used them in the argument model.

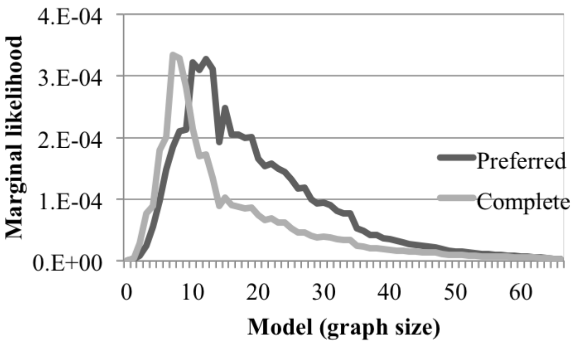

Figure 3: The marginal likelihood of the twelve claims given the graph-size models. We averaged five runs where each run sampled distinct attack relations from each model with a hundred trials.

A model is defined with a graph size, i.e., the number of edges of a graph. Given twelve claims, we have sixty-six, i.e., the combination ${}_{12}C_2$, models. Figure 3 shows the marginal likelihood of $\boldsymbol{acc}$ with respect to preferred and complete semantics. The x-axis shows the number M of graph edges, i.e., models. Here, we have assumed attack relations sampled from each model with a hundred trials. This allows us to avoid getting further involved with a complexity problem. In spite of the fact, it is observed that the empirical Bayes brings out the sharp peak at $M = 7$ for complete semantics and $M = 12$ for preferred semantics. This implies that the claims favor those models, and that the models based on graph sizes are reasonable choices to explain the statuses of the claims.

Conclusions

Focusing on graph-theoretic properties of argumentation frameworks, this paper introduced empirical Bayes in statistical construction of justification. We empirically discussed the effect of the use of the model selection technique.

Acknowledgments.

The author thanks Prof. Hu Liu for his kind support. This study was supported by NSSFC Grant 14ZDB015 and JSPS KAKENHI Grant Number 15KT0041.

References

Bar-Haim, R.; Bhattacharya, I.; Dinuzzo, F.; Saha, A.; and Slonim, N. 2017. Stance classification of context-dependent claims. In *Proc. of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 1*, 251–261.