Towards Argumentation-based Classification

Matthias Thimm
Universität Koblenz-Landau
Germany

Kristian Kersting
Technische Universität Darmstadt
Germany

1 Introduction

Classification is the problem of categorizing new observations by using a classifier learnt from already categorized examples. In general, the area of *machine learning* [Mitchell, 1997] has brought forth a series of different approaches to deal with this problem, from decision trees to support vector machines and others. Recently, approaches to *statistical relational learning* [De Raedt *et al.*, 2016] even take the perspective of knowledge representation and reasoning into account by developing models on more formal logical and statistical grounds. In this position paper, we envisage to significantly generalize this reasoning aspect of machine learning towards the use of *computational models of argumentation* [Baroni *et al.*, 2011], a popular approach to commonsense reasoning, for reasoning within machine learning. More concretely, we consider the following two-step classification approach. In the first step, rule learning algorithms are used to extract frequent patterns and rules from a given data set. The output of this step comprises a huge number of rules (given fairly low confidence and support parameters) and these cannot directly be used for the purpose of classification as they are usually inconsistent with one another. Therefore, in the second step, we interpret these rules as the input for approaches to structured argumentation, such as ASPIC$^+$ [Modgil and Prakken, 2014] or DeLP [Garcia and Simari, 2004]. Using the argumentative inference procedures of these approaches and given a new observation, the classification of the new observation is determined by constructing arguments on top of these rules for the different classes and determining their justification status.

The use of argumentation techniques allows to obtain classifiers, which are by design able to *explain* their decisions, and therefore addresses the recent need for *Explainable AI*: classifications are accompanied by a dialectical analysis showing why arguments for the conclusion are preferred to counterarguments. Argumentation techniques in machine learning also allows the easy integration of additional expert knowledge in form of arguments.

While there are some previous works considering the combination of machine learning and computational argumentation techniques—see e. g. [Možina *et al.*, 2008; Riveret and Governatori, 2016]—, the proposed two-step process offers a novel perspective on this combination, which is likely to bring new insights on the general relationship between machine learning and knowledge representation and reasoning. Preliminary experiments already suggest that our framework can yield performance comparable to state-of-the-art, while being explainable.

2 Proposed approach

We illustrate the goals of our envisioned approach using a classical example for a (multi-class) classification problem, the "Animals with Attributes" data set[1] (we only consider the base package with the class/attribute table). This dataset describes 50 animals, e. g. ox, mouse, dolphin, using 85 binary attributes such as "swims", "black", and "arctic". Using a first-order logic representation this data can be represented as a set of ground literals such as

$$swims(dolphin), \neg black(dolphin), \neg arctic(dolphin), \ldots$$

Now given the truth values of some attributes of a new animal, say a kangaroo, the classification tasks consists of predicting the values of the remaining attributes, e. g. given the fact that a kangaroo is orange and that it hops, does it live in the arctic? We address this task by first applying *association rule mining* such as the well-know Apriori algorithm [Agrawal and Srikant, 1994]. The output is a set of association rules such as "animals with flippers usually live in the ocean" which can be modeled as

$$flippers(X) \rightarrow ocean(X)$$

As the rules are mined based on frequent patterns and ignore logical coherency, they may be contradictory to each in other in certain cases. For example, another mined rule could be

$$big(X) \rightarrow \neg ocean(X)$$

saying that big animals usually do not live in the ocean. However, as a dolphin both has flippers and is big, the above two rules would therefore result in a contradiction and no meaningful classification could be given in this case. It is not surprising that rule mining algorithms are rarely used for classification purposes in this manner. We, however, add another second step to our classification approach by taking the output of the rule mining algorithm, i. e., a set of rules, as the input of an approach to structured argumentation such as ASPIC$^+$ [Modgil and Prakken, 2014] or DeLP [Garcia and Simari, 2004]. In these approaches, rules are not just applied in a

[1] `http://attributes.kyb.tuebingen.mpg.de`

direct fashion but arguments are build for all alternative conclusions and compared through e. g. a dialectical procedure in order to determine a consistent set of conclusions. Assume that another rule mined in the first step is

$$big(X), blue(X) \rightarrow ocean(X)$$

meaning that big and blue animals do indeed live in the ocean. Using *specificity* [Stolzenburg *et al.*, 2003] as a comparison criterion between conflicting arguments the conflict can be resolved because this final argument defeats the less specific second argument. We call this general approach *Argumentation-based Classification* (AbC). It is customizable by employing different rule mining algorithms in the first step and different approaches to structured argumentation in the second step. Moreover, besides using classical (qualitative) approaches to structured argumentation in the second step we can also make use of argumentative approaches incorporating quantitative uncertainty such as [Rienstra, 2012; Alsinet *et al.*, 2008]. By doing so, we can make use of additional quantitative information of the rules mined in the first step. For example, the confidence value of a rule can be interpreted as a conditional probability, i. e., the ratio of the probability of the conjunction of the head and body of the rule over only the body of the rule. This information can be used during the argumentation process in order to make more accurate predictions.

Making use of argumentation in classification allows the user to also inspect the reasoning process of why a certain prediction has been made, i. e., the resulting argumentative classification approaches are explainable by design. Formalisms such as DeLP conduct a dialectical analysis where all arguments contributing to the matter of deciding whether a certain statement is true. This analysis can be shown to the user in order to explain why a certain decision has been made. For example, above we would get the explanation "A dolphin lives in the ocean because it is blue, despite the fact that it is big". Users can then evaluate this reasoning and, if they are not satisfied with the explanation, pose a new argument for a different conclusion.

3 Preliminary results and conclusion

In order to assess the feasibility of our envisaged approach, we already implemented a first version of Argumentation-based Classification using the standard Apriori algorithm [Agrawal and Srikant, 1994] for rule mining and DeLP [Garcia and Simari, 2004] as the structured argumentation approach. We applied the rule miner to the "Animals with Attributes" data set with minimum confidence 0.9 and minimum support 0.8. We only mined rules with up to 3 elements in the body and 1 element in the conclusion. All rules with confidence value 1 were interpreted as strict rules, the remaining rules were interpreted as defeasible rules. This resulted in 254 strict and 621 defeasible rules. To these rules we added all but one randomly chosen attribute fact of some randomly chosen animal and asked DeLP whether the remaining attribute is warranted (note that DeLP has a three-valued answering behaviour: yes/no/undecided). We repeated this experiment 1000 times. While in about 70% of the times, DeLP could not classify the attribute (answer "undecided") it never misclassified any attribute and therefore classified 30% correctly, e. g., it never answered "no" when the correct answer was "yes". However, slightly changing the parameters of the experiment (such as minimum support and minimum confidence) would increase and decrease these values, while still not misclassifying any attribute. Note that using the mined rules directly as a classifier results in inconsistent classifications most of the time. We found these initial results encouraging and it is likely that a more careful setup, analysis, and evaluation will improve them significantly.

References

[Agrawal and Srikant, 1994] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings VLDB'94*, pages 487–499, 1994.

[Alsinet *et al.*, 2008] T. Alsinet, C. I. Chesñevar, L. Godo, and G. R. Simari. A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems*, 159(10):1208–1228, 2008.

[Baroni *et al.*, 2011] P. Baroni, M. Caminada, and M. Giacomin. An Introduction to Argumentation Semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[De Raedt *et al.*, 2016] L. De Raedt, K. Kersting, S. Natarajan, and D. Poole. *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2016.

[Garcia and Simari, 2004] A. Garcia and Guillermo R. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1–2):95–138, 2004.

[Mitchell, 1997] Tom Mitchell. *Machine Learning*. McGraw-Hill Education, 1997.

[Modgil and Prakken, 2014] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.

[Možina *et al.*, 2008] M. Možina, M. Guid, J. Krivec, A. Sadikov, and I. Bratko. Fighting knowledge acquisition bottleneck with argument based machine learning. In *Proceedings ECAI'08*, 2008.

[Odom and Natarajan, 2016] P. Odom and S. Natarajan. Actively interacting with experts: A probabilistic logic approach. In *Proceedings ECML PKDD 2016*, part II pages 527–542, 2016.

[Rienstra, 2012] T. Rienstra. Towards a probabilistic Dung-style argumentation system. In *Proceedings AT2012*, 2012.

[Riveret and Governatori, 2016] R. Riveret and G. Governatori. On learning attacks in probabilistic abstract argumentation. In *Proceedings AAMAS'16*, 2016.

[Stolzenburg *et al.*, 2003] F. Stolzenburg, A. Garcia, C. I. Chesnevar, and G. R. Simari. Computing generalized specificity. *Journal of Non-Classical Logics*, 13(1):87–113, 2003.