Towards a Structured Analysis of Approximate Problem Solving: a Case Study in Classification

Perry Groot Annette ten Teije Frank van Harmelen

Division of Mathematics and Computer Science,
Faculty of Sciences, Vrije Universiteit,
Amsterdam, The Netherlands

The full version of this paper appeared in: Principles of Knowledge Representation and Reasoning: Pr. of the 9th International Conference (KR-2004), p. 399–406, Whistler, BC, Canada.

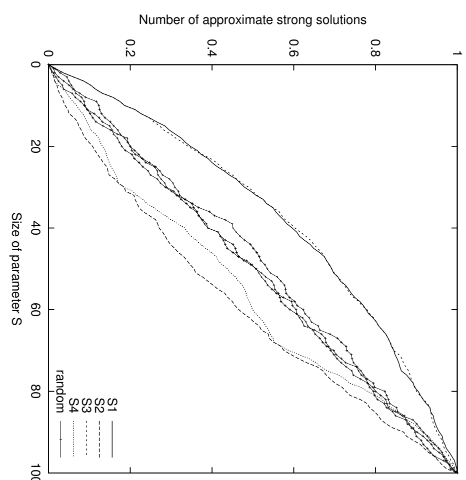
Introduction: The use of approximation as a method for dealing with complex problems is a fundamental research issue in Knowledge Representation. The literature on approximate and anytime reasoning of the last decade has studied specific *algorithms*: their anytime behaviour, their performance profiles, compositionality of such algorithms, monitoring and control of such algorithms, etc. However, a declarative characterisation of such algorithms is often lacking. On the other hand, *declarative formalisms* have been provided for approximate reasoning, but this work only tackles general logical deduction, and needs to be made more concrete for specific forms of AI problem solving. In this article we try to bridge the gap between these algorithm studies on the one hand and the analytical/declarative characterisations on the other hand. We present a case study in which we show how to do a structured analysis of an approximate entailment method for approximate problem solving.

Approximation method: The approximate entailment method focussed on is the method of Cadoli and Schaerf.[1] This method is general and can be applied to any problem that can be formalized in (propositional) logic and uses logical entailment for inferencing. This method has a number of desirable properties for an approximation method, but also has a number of limitations. In particular, it is not immediately obvious what the effect is of applying the method on a specific problem domain such as diagnosis or classification. Furthermore, the method uses a parameter S resulting in a whole spectrum of approximations that range from zero to optimal precision. Practical usefulness of the method therefore depends on the choice for S, making this choice a crucial part of the method. Currently, the method has not been evaluated beyond diagnosis and belief revision by means of a quantitative and qualitative analysis.

[1] Marco Schaerf and Marco Cadoli. Tractable reasoning via approximation. *Artificial Intelligence*, 74:249–310, 1995.

Domain: In classification the goal is to identify an object as belonging to a certain class. The object is described in terms of a (possibly incomplete) set of observations, i.e., attributes with some value. The classification criteria we consider are weak, strong, and explanative classification. In weak classification a class is a solution when it is consistent with the domain theory and the observations. In strong classification a class *c* is a solution when the domain theory together with *c* explains all observations. That is, we want candidate solutions to actually possess the properties that have been observed. In explanative classification a class is a solution if the class is explained by the observations. That is, a class is a candidate solution if all its properties are observed. Note that in strong classification a candidate solution may have *more* properties while in explanative classification a candidate solution may have *less* properties than the ones actually observed.

Structured analysis: This article shows how to do a structured analysis of the use of an approximate entailment method for approximate problem solving. The approach consists of two steps. **Theoretical analysis:** In this step the properties of the approximation method applied on a specific problem domain are analyzed by using the rules of the logic and the properties of the approximation method. This step tries to limit the choices for the parameter *S* resulting in a smaller search space of useful settings (and changes) for the crucial approximation parameter. **Empirical analysis:** In this step heuristics for the parameter *S* are determined and experiments are set up that measure the quality of these heuristics. The experiments include choosing problem instances, a quality measure for the heuristics, and a way to compare the measured qualities of the heuristics.



Results: The main results of the *theoretical analysis* are formulas obtained that describe the effect of replacing the entailment operator by the approximate entailment operator for the three classification forms. It was proven that, approximate weak classification behaves identical to weak classification that allows certain observations to be inconsistent. Approximate strong classification behaves identical to strong classification restricted to a subset of all classes. Approximate explanative classification behaves identical to classical explanative classification to which a set of classes is added.

Furthermore, the results of the theoretical analysis resulted in a number of restrictions for reasonable choices for the parameter *S* of the approximation method. These restrictions for the parameter *S* are used in the *empirical analysis* for developing concrete guidelines for choosing the parameter *S*. In the article, several heuristic orders are constructed for strong and explanative classification and validated using an empirical analysis. A typical outcome of such an empirical analysis is shown in the figure and is explained in more detail in the full paper.