Why deterministic logic is hard to learn but Statistical Relational Learning works

Marc Toussaint – October 7, 2010

Consider a standard regression setup: we have data $\{(x, y)\}$. See figure 1 – a story I am sure you already know... (a) assumes that there is no uncertainty in the world/data and tries to model the data. This assumption ties you down to model every observation literally; e.g., in a Bayesian setup with a GP without observation noise. Eventually, this prohibits generalization, simplification, abstraction, regularization, finding a compact description, etc. On the contrast, (b) assumes uncertainty in the observations. This uncertainty (e.g., in a GP prior) is the key that allows you to fit a generalizing, smooth, regularized, abstracting,... function.

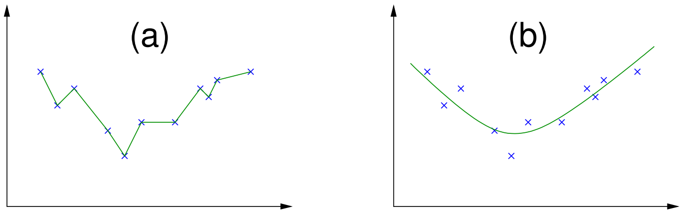

Figure 1: (a) Modelling (e.g. with a GP) when assuming there is *no observation noise*, (b) Modelling with observation noise. Lesson: we need uncertainty to learn compact models!

What has that to do with logic? Classical logic was deterministic. I claim that the core problem with classical logic is that *learning* doesn't work.

Statistical Relational Learning (SRL) is sometimes called a "marriage between logic and probabilities". That is, we can do proper statistical learning: trading between complexity and accuracy via regularizations; imposing regularizations based on minimum description length; learning simplifying, abstracting, compact models. All of this is *only* possible because we assume that the observations and models are probabilistic. We need this uncertainty to simplify, to compact our representation of previous experience!

Consider the example of learning a transition model $P(s'|a, s)$ from experiences (in the context of model-based Reinforcement Learning). We have list of (s, a, s') experiences. If we do not assume uncertainty and try to fit deterministic logic rules this will fail just as figure (a)! Only when the assumption of uncertainty "unleashes" the model and opens the door for simplification, abstraction and compactification: E.g., the agent may think "maybe this experience was only a statistically insignificant incident and I can build it into my model as an uncertain probabilistic event rather than introducing a special rule – then I can simplify my rule set much more." Uncertainty is the key for such thinking.

Therefore, SRL is not *merely* a marriage between logic and statistical learning – it is crucial to get logic working!

Dagstuhl Seminar Proceedings 10302
Learning paradigms in dynamic environments
http://drops.dagstuhl.de/opus/volltexte/2010/2801

As a final remark, in our work we do not care about 'logic' in the computational sense. That is, we do not use Prolog or logic inference or theorem proving as a computational paradigm. Instead, from our point of view, the logic descriptions are just complex feature descriptors. So we should call it "statistical learning with complex feature descriptors formulated in the language of logic formulas".

References

Tobias Lang and Marc Toussaint. Relevance grounding for planning in relational domains. In *Proc. of the European Conf. on Machine Learning (ECML 2009)*, volume 5781/2009 of *Lecture Notes in Computer Science*, pages 736–751. Springer, 2009a. ISBN 978-3-642-04179-2.

Tobias Lang and Marc Toussaint. Approximate inference for planning in stochastic relational worlds. In *Proc. of the 26rd Int. Conf. on Machine Learning (ICML 2009)*, pages 585–592. ACM, 2009b. ISBN 978-1-60558-516-1.

Tobias Lang and Marc Toussaint. Planning with learned stochastic relational rules. *Journal of Artificial Intelligence Research*, 39:1–49, 2010.

Tobias Lang, Marc Toussaint, and Kristian Kersting. Exploration in relational worlds. In *Proc. of the European Conf. on Machine Learning (ECML 2010)*, 2010.