Extended Abstract

Probabilistic inference for solving structured MDPs and POMDPs

Marc Toussaint
School of Informatics, University of Edinburgh,
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK

Inference on structured domains has made considerable advances in recent years: Variational inference has been proposed as a basic methods to decompose the inference process (as on factored HMMs, Ghahramani & Jordan 1995), message-passing algorithms (such as loopy belief propagation, expectation propagation, or exact inference via the Junction Tree Algorithm, Minka 2001) can efficiently handle a very broad range of structured models. Extensions of particle filters allow for more versatile belief representations in continuous domains (Klaas et al. 2006), and interesting techniques for efficient inference in relational models are currently developed (Chavira et al. 2006).

To scale up to more realistic scenarios, planning (or model-based Reinforcement Learning) in stochastic environments equally needs to cope with *structured* descriptions of the environment, e.g., factored, hierarchical, and mixed (continuous/discrete) state representations. Approaches here include work on Factored Markov Decision Processes (Boutilier et al. 1995; Koller & Parr 1999; Guestrin et al. 2003; Kveton & Hauskrecht 2005), abstractions (Hauskrecht et al. 1998), and relational models of the environment (Zettlemoyer et al. 2005).

From a complexity theoretic point of view, the equivalence between inference and planning in well-known (see, e.g., Littman et al. 2001). However, we ask how such approaches to structured probabilistic inference can *exactly* be transferred to the problem of planning, in other words, whether one can translate the problem of planning directly to a problem of inference. The aim is to connect both fields more strongly and eventually to apply efficient methods of probabilistic inference directly in the realm of planning.

In my presentation I will report on our work on establishing equivalence between likelihood maximization and maximization of the expected future discounted return and applying the new technique of inference planning on different domains.

Inference planning in MDPs. In (Toussaint & Storkey 2006) we developed the general framework for inference planning in MDPs. One interesting aspect of this approach is to consider the total time of a Dynamic Bayesian Network as a random variable (formally this amounts to considering a mixture of finite-time DBNs) without compromising the efficiency of inference. In this way we can cope with arbitrary and discounted rewards and efficiently compute posterior distributions over actions, states and the total time conditioned on reward. An EM-algorithm is used for computing optimal policies.

An EM-algorithm for learning FSCs in POMDPs. In (Toussaint & Harmeling 2007) we propose using inference planning on POMDPs. Here, an EM-algorithm is used to learn the parameters of a Dynamic Bayesian Network model of the agent (which could be a finite state controller or any other structured control policy). The approach compares favorably to gradient-based approaches and allows to consider more interesting (structured) internal architectures of the agent than plain finite state controllers. For instance, for a specific architecture we can learn suitable internal nodes that represent basic reactive behaviors like aisle following in a maze.

Inference planning in robotic domains. In aiming to apply the new techniques on real world problems we considered the problem of trajectory planning for a redundant robotic system under collision constraints (Toussaint & Goerick 2007). We formulated a factor graph representation of the problem and solved the inference problem using (loopy) belief propagation. The approach proved efficient and is promising to extend to multiple (factored) constraints and state variables. Current research tries to extend this technique to planning sequences when multiple parallel controllers (motor primitives) are used to generate complex movements.

References

Boutilier, C., R. Dearden, & M. Goldszmidt (1995). Exploiting structure in policy construction. In *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI 1995)*, pp. 1104–1111.

Chavira, M., A. Darwiche, & M. Jaeger (2006). Compiling relational bayesian networks for exact inference. *International Journal of Approximate Reasoning* **42**, 4–20.

Ghahramani, Z. & M. I. Jordan (1995). Factorial hidden Markov models. In *Advances in Neural Information Processing Systems, NIPS*, Volume 8, pp. 472–478. MIT Press.

Guestrin, C., D. Koller, R. Parr, & S. Venkataraman (2003). Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research (JAIR)* **19**, 399–468.

Hauskrecht, M., N. Meuleau, L. P. Kaelbling, T. Dean, & C. Boutilier (1998). Hierarchical solution of Markov decision processes using macro-actions. In *Proc. of Uncertainty in Artificial Intelligence (UAI 1998)*, pp. 220–229.

Klaas, M., M. Briers, N. de Freitas, A. Doucet, S. Maskell, & D. Lang (2006). Fast particle smoothing: If I had a million particles. In *23nd Int. Conf. on Machine Learning (ICML 2006)*.

Koller, D. & R. Parr (1999). Computing factored value functions for policies in structured MDPs. In *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI 1999)*, pp. 1332–1339.

Kveton, B. & M. Hauskrecht (2005). An MCMC approach to solving hybrid factored MDPs. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*.

Littman, M. L., S. M. Majercik, & T. Pitassi (2001). Stochastic boolean satisfiability. *Journal of Automated Reasoning 27*(3), 251–296.

Minka, T. (2001). A family of algorithms for approximate bayesian inference. PhD thesis, MIT.

Toussaint, M. & C. Goerick (2007). Probabilistic inference for structured planning in robotics. Submitted to *Int. Conf. on Robotics and Automation (ICRA 2007)*.

Toussaint, M. & S. Harmeling (2007). Learning state representations with probabilistic inference for solving POMDPs. Submitted.

Toussaint, M. & A. Storkey (2006). Probabilistic inference for solving discrete and continuous state Markov Decision Processes. In *23nd Int. Conf. on Machine Learning (ICML 2006)*.

Zettlemoyer, L., H. Pasula, & L. P. Kaelbling (2005). Learning planning rules in noisy stochastic worlds. In *Proc. of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*.