Variational Methods for Inference

Dexiong Chen   Chia-Man Hung   Baoyang Song

January 4, 2017

One of the most important problems in graphical models is the probabilistic inference. It consists of deducing and computing properties including marginal probabilities or conditional probabilities of an underlying distribution represented as a graphical model. For graphs with simple structure such as trees, the inference problem can be exactly solved by message-passing algorithms taking form of sum-product or max-product algorithm, which explore the conditional independence properties present in the graph and has only linear complexity in the number of nodes. These algorithms can also be extended to arbitrary graphs by using junction tree representation. However, the time complexity will also be increased to be exponential in the size of the maximal clique in the junction tree, which makes the exact computation intractable. Thus, a variety of approximation procedures have been developed and studied. One of the fundamental approaches is to design algorithms involving Monte Carlo methods, referred as Markov Chain Monte Carlo (MCMC). The idea is simply sampling a Markov Chain that converges to the distribution of interest. These approaches possess theoretical guarantee and simple implementation. Nevertheless, sampling methods can be very slow to converge and lack stopping criterion [1].

In this project, we will study another approach, namely variational methods for computing approximations of marginal probabilities. These methods are based on variational principle, which generally converts a complex problem to a simpler problem by including additional parameters following convex duality and conjugate. There exist several formulations for variational inference and we will focus on mean field, loopy sum-product or belief propagation and structured mean field. The objective of the project is to understand and analyze these methods and compare them with the MCMC methods such as Gibbs sampling, applied on an Ising model.

According to the principle of maximum entropy, many graphical models can be viewed as exponential families and we can assume that the distribution of interest p is in one exponential family $q_\theta(x) = \exp(\theta^T \phi(x) - A(\theta))$. The convexity of the log partition function A allows us to find the connection between A and its conjugate dual function A^*

$$A(\theta) = \sup_{\mu \in \mathcal{M}} (\theta^T \mu - A^*(\mu)). \tag{1}$$

where $\mathcal{M}$ is the marginal polytope. One may show that the supremum of this equation is attained by the moment matching condition $\mu = \mathbb{E}_\theta[\phi(x)]$, which is the goal of our inference problem.

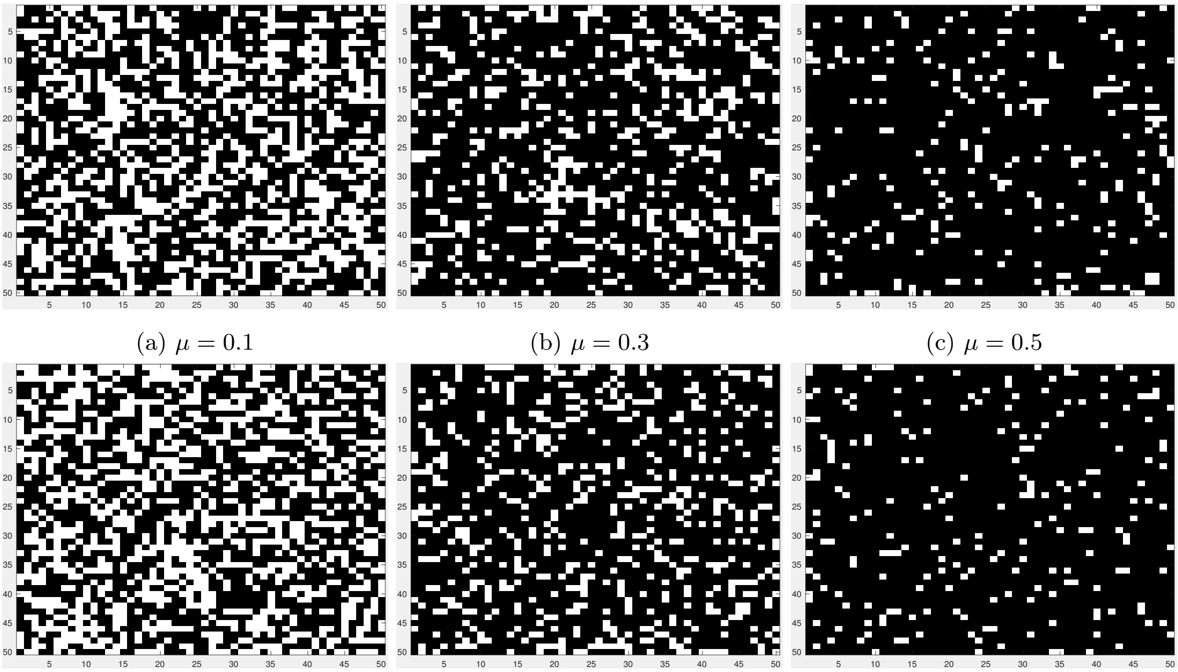
Loopy belief propagation. The message-passing algorithms listed above can be reinterpreted as a variational formulation via the Bethe variational problem.

Mean field. The main obstacles of the variational formulation 1 are the lack of the knowledge about the constraint set $\mathcal{M}$ and the conjugate function A^*. The key idea of mean field method is to limit the optimization in a tractable subset of distributions in such a way that both two terms are easy to characterize.

Structured mean field. In order to adapt mean field to more specific and structural graphs, we can incorporate in the above subset additional structure.

Simulation. We have implemented the Gibbs sampling algorithm and the naive mean field algorithm on a square-lattice Ising model. Figure 1 shows some instances with different increasing correlations.

Next steps. We will investigate and implement the loopy belief propagation and structured mean field algorithms.



(a) $\mu = 0.1$ (b) $\mu = 0.3$ (c) $\mu = 0.5$

(d) $\mu = 0.1$ (e) $\mu = 0.3$ (f) $\mu = 0.5$

Figure 1: Top: Gibbs sampling; Bottom: Mean-field.

References

[1] Martin J Wainwright and Michael I Jordan. "Graphical models, exponential families, and variational inference". In: *Foundations and Trends® in Machine Learning* 1.1-2 (2008), pp. 1–305.