Message Passing and Expectation Propagation

Christoph Dehner
dehner@in.tum.de

29 May 2017

Probabilistic graphical models like Markov Random Fields or Bayesian Networks provide clear and illustrative ways to describe probabilistic processes. In such models, nodes represent random variables and edges their conditional dependencies. The joint distribution of all involved variables can be expressed as a product of factors f_i.

Typical tasks in graphical models are are to do Bayesian inference to extract marginal distributions of specific variables or to find maximum apriori estimates. The paper presents two approaches for that. Firstly, two belief propagation algorithms for exact inference are presented. In the second part expectation propagation as a method for approximate inference is introduced.

The sum-product algorithm performs exact inference in a graphical model. It bases on the naive approach of marginalizing out all but one variables of the joint distribution. The necessary computational effort for that is exponential in the number of variables. Expressing the joint distribution as a product of factors from the graphical model allows to exchange summation and multiplication. By this way, the complexity of marginalization can be reduced to linear cost in the number of involved variables. Analogously, the max-sum algorithm presents a efficient solution for finding a maximum apriori estimates.

Both algorithms operate on factor graphs, a structure which makes the factors of the joint distribution explicit and can be extracted from graphical models. The two belief propagation algorithms consist of local messages passed through the factor graph and are only exact for trees.

Therefore, the second part of the paper deals with approximated inference in general models. Next to methods like loopy belief propagation, which generalize the message passing approach to arbitrary graph, expectation propagation as a very popular method for approximate inference is covered.

Expectation propagation approximates a distribution $p(\theta)$ with a member of the exponential family $q(\theta)$ by minimizing the KL divergence $KL(p||q)$. This is similar to variational inference, but with a swapped order in the KL divergence. The approximation $q(\theta)$ is refined iteratively by matching the moments of one factor at a time of the joint distribution while all other factors are fixed. Compared to variation inference, the final algorithm is not guaranteed to converge any more. However, if it does, it often achieves greater accuracy.