Cumulative distribution networks: Graphical models for cumulative distribution functions

Jim C. Huang* and Brendan J. Frey
Probabilistic and Statistical Inference Group, University of Toronto
10 King's College Road, Toronto, ON, M5S 3G4, Canada
{jim,frey}@psi.toronto.edu
*: Presenting author

March 31, 2009

Probabilistic graphical models are widely used for representing dependencies between random variables in a wide variety of problems such as error-correction coding, information retrieval and computational systems biology. The underlying probability functions described by graphical models typically take on the form of the joint probability density function (PDF) of many continuous random variables, or the joint probability mass function (PMF) of many discrete random variables. For many real-world problems, graphical models for PDFs may present several computational challenges, such requiring one to introduce latent variables into a model in order to explain complex dependencies between observable variables and then having to marginalize out these latent variables. Such models can often be non-identifiable, as there are a potentially infinite number of possible latent variable models associated with any given model defined over observable variables. Often, the difficulty of performing inference and learning under such models increases significantly in such models so that approximate inference techniques may be required.

Another possible limitation is that the joint PDF/PMF itself might not be appropriate for some applications. In applications such as learning to rank, the *cumulative distribution function* (CDF) is a probabilistic representation which arises frequently. As a probability measure over inequality events of the type $X \leq x$, the CDF lends itself to problems such as learning to rank where the goal is to predict events of the above type. Examples of this type of problem include web search and document retrieval, predicting movie ratings or predicting multiplayer game outcomes with a team structure. In contrast to the canonical machine learning problems of classification or regression, in learning to rank we are required to learn some mapping from inputs to structured outputs in which multiple outputs are dependent for a given observation. Here we may wish to model not only stochastic ordering relationships between variables, but also statistical dependence relationships between variables.

To address the above issues, we present the cumulative distribution network (CDN) as a novel graphical model which describes the joint CDF of a set of variables (Huang and Frey 2008; Huang and Frey 2009a). In contrast to most previous graphical models, marginalization in CDNs involves tractable operations such as computing limits and computing the derivatives of local functions in the graph. In contrast to undirected graphs, the global normalization constraint in a CDN can be enforced locally for each function in the graph. We will show that the rules for assessing conditional independence in CDNs are unlike those for directed, undirected and factor graphs. We will show that the conditional independence criteria for CDNs defined over continuous variables in fact corresponds to those of bi-directed models for continuous variables (Richardson 2003). As a result, CDNs are a continuous parameterization for bi-directed models which allow us to represent complex dependencies between observable variables in a model with appropriate latent variables marginalized out. We will provide both sufficient and necessary conditions for a CDN to have a dual representation as a factor graph with latent variables, so that the joint probability model described by the CDN exhibits the same conditional independence relationships as that of a factor graph with additional latent variables introduced. This will then allow us to construct multivariate extreme value distributions for which both a factor graph and CDN representation exists.

We will discuss the problem of performing inference under CDNs, which we will show corresponds to computing derivatives of the joint CDF. We will describe a message-passing algorithm for inference in CDNs called the *derivative-sum-product algorithm* and demonstrate its use on a problem of structured ranking in multiplayer gaming. We will then present a general framework for structured ranking learning in which we are given many observations of partial orderings over many objects to be ranked and we wish to learn to rank these objects whilst accounting for a high degree of dependency between outputs (Huang and Frey 2009a). Results will be presented for applications of structured ranking learning to the problems of document retrieval (Burges *et al.* 2005; Liu *et al.* 2007; Xia *et al.* 2008) and regulatory sequence discovery in computational biology (Berger *et al.* 2006; Chen *et al.* 2007; Huang and Frey 2009b). For this class of applications, CDNs provide a graphical structure which includes previous probabilistic models for rank data, such as the Plackett-Luce, Bradley-Terry models (Marden 1995) and the methods of (Burges *et al.* 2005; Chen *et al.* 2007; Xia *et al.* 2008).

References

[Berger *et al.* 2006] Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep III, P.W. and Bulyk, M.L. 2006. Compact, universal DNA microarrays to comprehensively determine transcription factor binding specificities. *Nat. Biotechnol.* **24**: 1429-1435.

[Burges *et al.* 2005] Burges, C.J.C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N. and Hullender, G. (2005) Learning to rank using gradient descent. *In Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, 89-96.

[Chen *et al.* 2007] Chen, X., Hughes, T.R. and Morris, Q.D. 2007. RankMotif++: a motif-search algorithm that accounts for relative ranks of K-mers in binding transcription factors. *Bioinformatics* **23**: i72-i79.

[Huang and Frey 2008] Huang, J.C. and Frey, B.J. 2008. Cumulative distribution networks and the derivative-sum-product algorithm. *In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI).*

[Huang and Frey 2009a] Huang, J.C. and Frey, B.J. 2009. Structured ranking learning using cumulative distribution networks. *Advances in Neural Information Processing Systems (NIPS)* **21**.

[Huang and Frey 2009b] Huang, J.C. and Frey, B.J. 2009. STORMSeq: Ranking sequences by integrating experimental datasets with diverse computational predictions. *Submitted.*

[Liu *et al.* 2007] Liu, T-Y., Xu, J., Qin, T., Xiong, W. and Li, H. (2007) LETOR: Benchmark dataset for research on learning to rank for information retrieval. *LR4IR 2007, in conjunction with SIGIR 2007.*

[Marden 1995] Marden, J. I. (1995) Analyzing and modeling rank data. *CRC Press*, 1995.

[Richardson 2003] Richardson, T. S. (2003) Markov properties for acyclic directed mixed graphs. *Scand. J. Statist.* **30**, 145-157.

[Xia *et al.* 2008] Xia, F., Liu, T.Y., Wang, J., Zhang,W. and Li, H. (2008) Listwise approach to learning to rank - theory and algorithm. *In Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML).*

Topic: Graphical Models
Preference: Oral