Latent Dependency Forest Models (Supplementary Material)

Shanbo Chu, Yong Jiang and Kewei Tu
School of Information Science and Technology
ShanghaiTech University, Shanghai, China
{chushb, jiangyong, tukw}@shanghaitech.edu.cn

An Example of Using LDFMs to Model CSI

The assignments of the variables can influence the distribution over the dependency structures. In this way, LDFMs can model CSI to some extent. Here is an example of using LDFM-S to model three binary variables X_1, X_2 and X_3.

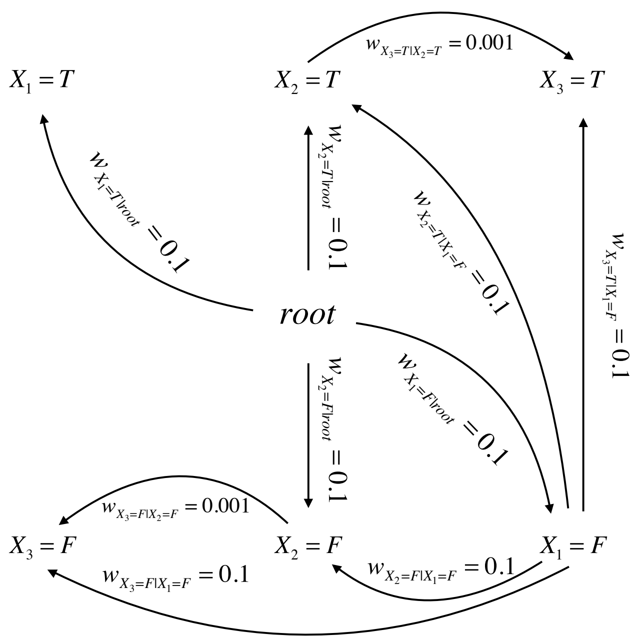

Figure 1: All possible pairwise dependencies between the three variables and a root node. Each dependency has a weight and only the dependencies with non-zero weights are shown. The weight $w_{s|x_i}$, which is the probability of generating a stop node given the assignment $X_i = x_i$ is not drawn for simplicity, but it can be computed using the normalization condition discussed in the *LDFM-S* subsection in the main text.

Figure 1 gives an example of using LDFM-S to model CSI. The conditional probabilities of the two variables X_2 and X_3 given X_1 can be computed using the formula in the *LDFM-S* subsection and they are shown in Table 1. It can be seen that when $X_1 = T$, X_2 and X_3 are strongly dependent; when $X_1 = F$, they are only weakly dependent.

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Table 1: The conditional probabilities of the two variables X_2 and X_3 given X_1

X_1	X_2	X_3	$P(X_2, X_3 \mid X_1)$
T	T	T	0.5
T	T	F	0
T	F	T	0
T	F	F	0.5
F	T	T	0.251
F	T	F	0.249
F	F	T	0.249
F	F	F	0.251

The Derivation Details

We show the details of deriving the probability of generating an assignment $\mathbf{x}$ discussed in the *LDFM* subsection in the main text.

$$\begin{aligned} p(\mathbf{x}) &= \sum_{\hat{T}} p(\hat{T}) p(\mathbf{x}|\hat{T}) \\ &= \sum_{\hat{T}} p(\hat{T}) \sum_{M} p(\mathbf{x}, M, \hat{T}) \\ &= \beta \sum_{\hat{T}} \sum_{M} \prod_{(i,j) \in E_{\hat{T}}} w_{x_j|x_i} \\ &= \beta n! \sum_{T \in T(G_{\mathbf{x}})} \prod_{(i,j) \in E_T} w_{x_j|x_i} \\ &= \beta n! Z_{\mathbf{x}} \propto Z_{\mathbf{x}} \end{aligned}$$

where $\hat{T}$ is the uniformly generated tree structure and M is a mapping from the n variables to the n nodes of the tree structure $\hat{T}$, β is the constant value of $p(\hat{T})$. $\beta n!$ is a constant w.r.t. $\mathbf{x}$. Here we have $n!$ because for each spanning tree T of $G_{\mathbf{x}}$, each permutation of the n variables is generated differently (i.e., corresponds to a different $\langle T, M \rangle$ pair).

Table 2: The maximum of CLL and CMLL normalized by the number of query variables. The bold numbers mark the best performance.

Dataset	Asia	Child	Alarm	Insurance	Sachs	Water	Win95pts	Hepar2	Hailfinder
5000 training samples; 40% Query, 30% Evidence									
BN	-0.274	-0.721	-0.436	-0.565	-0.675	-0.474	-0.229	-0.509	-1.223
DN	-0.268	-0.634	-0.317	-0.499	-0.610	-0.407	-0.185	-0.490	-1.089
SPN	-0.262	-0.63	**-0.277**	-0.476	-0.644	-0.415	**-0.118**	-0.489	**-0.941**
MT	-0.262	-0.707	-0.343	-0.557	-0.647	-0.435	-0.121	-0.507	-1.241
LDFM	**-0.258**	-0.609	-0.293	**-0.460**	**-0.605**	**-0.399**	-0.166	-0.481	-0.991
LDFM-S	-0.263	**-0.607**	**-0.291**	-0.462	-0.613	-0.462	-0.130	**-0.480**	-0.987
5000 training samples; 30% Query, 40% Evidence									
BN	-0.266	-0.711	-0.411	-0.589	-0.655	-0.437	-0.187	-0.497	-1.088
DN	-0.237	-0.610	-0.303	-0.528	-0.589	-0.391	-0.148	-0.479	-0.985
SPN	-0.229	-0.619	**-0.272**	-0.506	-0.620	-0.402	**-0.114**	-0.481	**-0.893**
MT	-0.226	-0.698	-0.348	-0.603	-0.620	-0.427	-0.116	-0.499	-1.188
LDFM	**-0.230**	-0.609	-0.288	**-0.481**	**-0.581**	**-0.383**	-0.129	-0.461	-0.908
LDFM-S	-0.235	**-0.588**	-0.286	-0.482	-0.586	-0.461	-0.124	**-0.459**	-0.904

The Evaluation Results of LDFM-S

We report the results of LDFM-S and LDFM trained on the 5000-sample datasets and evaluated by using Gibbs sampling on two different proportions of dividing the query and evidence variables in Table 2. It can be seen that LDFM-S has similar performance to LDFM on most datasets, but achieves significantly better results on the Win95pts dataset and significantly worse results on the Water dataset. Therefore, it may depend on the dataset as to whether modeling distributions over tree structures is useful.

More Evaluation Results

In the *Experiments* section in the main text, we report the evaluation results of two proportions of dividing the query and evidence variables (40% query, 30% evidence and 30% query, 20% evidence). In Table 3 we report the evaluation results of the other two proportions (30% query, 40% evidence and 20% query, 30% evidence).

Table 3: The maximum of CLL and CMLL normalized by the number of query variables. The bold numbers mark the best performance.

Dataset	Asia	Child	Alarm	Insurance	Sachs	Water	Win95pts	Hepar2	Hailfinder
5000 training samples; 30% Query, 40% Evidence									
BN	-0.266	-0.711	-0.411	-0.589	-0.655	-0.437	-0.187	-0.497	-1.088
DN	-0.237	-0.610	-0.303	-0.528	-0.589	-0.391	-0.148	-0.479	-0.985
SPN	-0.229	-0.619	**-0.272**	-0.506	-0.620	-0.402	**-0.114**	-0.481	**-0.893**
MT	-0.226	-0.698	-0.348	-0.603	-0.620	-0.427	-0.116	-0.499	-1.188
g-LDFM	-0.230	**-0.609**	-0.288	**-0.481**	**-0.581**	**-0.383**	-0.129	**-0.461**	-0.908
t-LDFM	**-0.210**	-0.630	-0.349	-0.556	-0.590	-0.389	-0.159	-0.464	-1.019
2000 training samples; 30% Query, 40% Evidence									
BN	-0.266	-0.764	-0.469	-0.599	-0.669	-0.466	-0.195	-0.506	-1.099
DN	-0.242	-0.626	-0.312	-0.550	-0.597	-0.404	-0.154	-0.497	-0.999
SPN	-0.232	-0.634	**-0.300**	-0.519	-0.634	-0.406	**-0.117**	-0.479	-0.911
MT	-0.228	-0.719	-0.371	-0.623	-0.638	-0.446	-0.132	-0.521	-1.250
g-LDFM	-0.229	**-0.587**	-0.303	**-0.487**	-0.581	**-0.378**	-0.126	**-0.462**	**-0.900**
t-LDFM	**-0.218**	-0.640	-0.392	-0.555	**-0.576**	-0.402	-0.169	-0.465	-1.013
500 training samples; 30% Query, 40% Evidence									
BN	-0.288	-0.776	-0.500	-0.690	-0.721	-0.480	-0.236	-0.534	-1.314
DN	-0.240	-0.654	-0.348	-0.706	-0.625	-0.418	-0.186	-0.498	-1.058
SPN	-0.243	-0.752	-0.425	-0.637	-0.741	-0.512	**-0.147**	-0.520	-1.140
MT	-0.234	-0.961	-0.569	-0.811	-0.710	-0.562	-0.183	-0.647	-2.226
g-LDFM	-0.238	**-0.609**	**-0.331**	**-0.509**	-0.596	**-0.390**	-0.152	-0.473	**-0.947**
t-LDFM	**-0.220**	-0.650	-0.359	-0.562	**-0.579**	-0.398	-0.167	**-0.473**	-1.029
5000 training samples; 20% Query, 30% Evidence									
BN	-0.217	-0.724	-0.432	-0.585	-0.698	-0.448	-0.217	-0.505	-1.164
DN	-0.198	-0.655	-0.316	-0.524	-0.626	-0.411	-0.174	-0.487	-1.068
SPN	-0.188	-0.671	**-0.297**	-0.514	-0.660	-0.428	**-0.133**	-0.491	-0.982
MT	-0.189	-0.733	-0.355	-0.586	-0.659	-0.445	-0.134	-0.506	-1.277
g-LDFM	-0.192	**-0.644**	-0.309	**-0.480**	**-0.617**	**-0.398**	-0.145	**-0.468**	**-0.972**
t-LDFM	**-0.166**	-0.678	-0.352	-0.544	-0.618	-0.404	-0.168	-0.471	-1.068
2000 training samples; 20% Query, 30% Evidence									
BN	-0.217	-0.759	-0.486	-0.600	-0.697	-0.464	-0.214	-0.510	-1.159
DN	-0.194	-0.659	-0.330	-0.547	-0.638	-0.418	-0.177	-0.492	-1.073
SPN	-0.189	-0.682	-0.324	-0.518	-0.667	-0.431	**-0.137**	-0.492	-0.993
MT	-0.189	-0.750	-0.376	-0.601	-0.669	-0.455	-0.148	-0.522	-1.316
g-LDFM	-0.191	**-0.637**	**-0.319**	**-0.483**	-0.620	**-0.402**	-0.143	**-0.471**	**-0.946**
t-LDFM	**-0.169**	-0.687	-0.394	-0.551	**-0.618**	-0.408	-0.183	-0.472	-1.062
500 training samples; 20% Query, 30% Evidence									
BN	-0.223	-0.778	-0.522	-0.664	-0.716	-0.479	-0.262	-0.535	-1.431
DN	-0.192	-0.675	-0.362	-0.551	-0.676	-0.428	-0.199	-0.499	-1.116
SPN	-0.192	-0.795	-0.444	-0.627	-0.765	-0.517	**-0.161**	-0.529	-1.219
MT	-0.191	-0.915	-0.522	-0.721	-0.721	-0.511	-0.184	-0.607	-2.120
g-LDFM	-0.198	**-0.648**	**-0.335**	**-0.497**	-0.642	**-0.406**	-0.168	-0.478	**-0.990**
t-LDFM	**-0.172**	-0.690	-0.368	-0.551	**-0.613**	-0.410	-0.175	**-0.476**	-1.075