A Probabilistic Model for Component-Based Shape Synthesis

Supplementary Material

Evangelos Kalogerakis Siddhartha Chaudhuri Daphne Koller Vladlen Koltun

Stanford University

Likelihood evaluation and parameter estimation. The first task in evaluating the score for a test structure G is to estimate the MAP parameters $\tilde{\Theta}_G$ by maximizing the product

$$\tilde{\Theta}_G = \arg\max_{\Theta} P(\mathbf{O} \mid G, \Theta) P(\Theta \mid G).$$

Here, the first term is the likelihood function and the second term is the parameter prior. Unfortunately, the product cannot be optimized in closed form, because the likelihood is a function of the unknown values of the hidden random variables:

$$P(\mathbf{O} \mid G, \Theta) = \prod_k \sum_{R_k, \mathbf{S}_k} P(O_k, R_k, \mathbf{S}_k \mid \Theta).$$

Therefore, we use the expectation-maximization (EM) algorithm to optimize the parameters iteratively. The algorithm starts with an initial assignment to the values of the shape style R and component style S_l for each category label l. The initial assignment is obtained by k-means clustering on the feature space $\{\mathbf{C}_l, \mathbf{D}_l\}$ under the Euclidean metric to obtain initial values of S_l for each training example. Then we perform k-medoids on the feature space $\{\mathbf{S}, \mathbf{N}\}$ under the Hamming metric to get initial values for the shape style R for each training example. In both cases, we repeat the clustering with random starting points until we find the assignments that minimize the sum of distances of the data points to their closest cluster centers.

The EM algorithm alternates between two steps: the M-step in which the parameters $\tilde{\Theta}_G$ are re-estimated based on the current assignments to the hidden random variables, and the E-step in which the algorithm performs inference to find probabilistic assignments to the hidden variables for each training example. In the M-step, the MAP estimates are computed using a Dirichlet prior distribution for the parameters of the CPDs of the discrete random variables and a normal-Wishart distribution for the parameters of the CPDs of the continuous random variables. The updates to the parameters are computed as follows:

Given the probabilities estimated in the previous E-step, we compute for each shape k:

$$M[R = r] = \sum_k P(R_k = r \mid O_k),$$
$$M[S_l = s] = \sum_k P(S_{l,k} = s \mid O_k),$$
$$M[S_l = s, R = r] = \sum_k P(S_{l,k} = s, R_k = r \mid O_k).$$

The parameters for the probability table for R are estimated as:

$$q_r = \frac{M[R = r] + \alpha}{K + \alpha|\mathcal{R}|},$$

where the hyperparameter α is set to 0.1. Then for each remaining discrete random variable T with a single parent U and every value u of U, the corresponding CPT parameters are estimated:

$$q_{t|u} = \frac{M[T = t, U = u] + \alpha}{M[U = u] + \alpha|\mathcal{T}|}.$$

For discrete random variables with multiple parents $\mathbf{U}$, we fit sigmoid CPDs using iterative reweighted least-squares [Bishop 2006]. Then we compute for every value $\mathbf{u}$ in the value space $\mathcal{U}$ of $\mathbf{U}$:

$$q_{t|\mathbf{u}} = \frac{P(T = t \mid \mathbf{U} = \mathbf{u}) + \alpha/K}{1 + \alpha|\mathcal{T}|/K},$$

where $P(T = t \mid \mathbf{U} = \mathbf{u})$ is the output of the estimated sigmoid functions.

For the case of a continuous random variable $\mathbf{C}_l$ with no continuous parents, the parameters of its conditional linear Gaussians are updated as follows [Gauvain and Lee 1994; Geiger and Heckerman 1994; Koller and Friedman 2009] (we omit the conditioning on discrete parents for notational clarity):

$$\boldsymbol{\phi}_l = \frac{\boldsymbol{\mu}_l + \sum_{k=1}^{K} P(S_{l,k} = s)\mathbf{C}_{l,k}}{1 + M[S_l = s]},$$

$$\Sigma_l \equiv \Sigma_{ll} = \frac{\Omega + \sum_{k=1}^{K} P(S_{l,k} = s)(\mathbf{C}_{l,k} - \boldsymbol{\phi}_l)(\mathbf{C}_{l,k} - \boldsymbol{\phi}_l)^{\mathrm{T}}}{1 + M[S_l = s]} + \frac{(\boldsymbol{\mu}_l - \boldsymbol{\phi}_l)(\boldsymbol{\mu}_l - \boldsymbol{\phi}_l)^{\mathrm{T}}}{1 + M[S_l = s]}.$$

where $\Omega = 10^{-4}\mathbf{I}$ is a regularization parameter. If the number of the above parameters is larger than the number of training instances, we only consider diagonal covariance matrices. The parameter $\boldsymbol{\mu}_l$ is set to be the mean of the features in the component category l. If $\mathbf{C}_l$ has continuous parents $\{\mathbf{C}_{l'}\}$, then the parameters are updated as follows:

$$\Sigma_{ll'} = \frac{\sum_{k=1}^{K} P(S_{l,k} = s)(\mathbf{C}_{l,k} - \boldsymbol{\phi}_l)(\mathbf{C}_{l',k} - \boldsymbol{\phi}_{l'})^{\mathrm{T}}}{M[S_l = s]},$$
$$\boldsymbol{\phi}_{l,0} = \boldsymbol{\phi}_l - \Sigma_{ll'}\Sigma_{l'l'}^{-1}\boldsymbol{\phi}_{l'},$$
$$\boldsymbol{\phi}_{l,l'} = \Sigma_{ll'}\Sigma_{l'l'}^{-1}$$
$$\Sigma_l = \Sigma_{ll} - \Sigma_{ll'}\Sigma_{l'l'}^{-1}\Sigma_{l'l}.$$

If the number of the parameters in $\boldsymbol{\phi}_{l,l'}$ is larger than the number of training instances, we use only the three first scale features in $\mathbf{C}_l$ for estimating $\boldsymbol{\phi}_{l,l'}$.

In the E-step, inference is performed to estimate probabilities of assignments $P(R_k \mid O_k)$ and $P(S_{l,k} \mid O_k)$ for each training data instance. Inference for the hidden random variables can be performed using variable elimination. Given observed data O_k for a source shape k, we compute $P(R, S_l \mid O_k)$ for a label $l \in \mathcal{L}$ using the following formula:

$$P(R, S_l \mid O_k) = \frac{P(R, S_l, O_k)}{P(O_k)},$$

where

$$P(R,S_l,O_k) = \sum_{\mathbf{S}\setminus\{S_l\}} P(R) \prod_{l'\in\mathcal{L}} P(S_{l'} \mid R) P\left(N_{l',k} \mid R, \pi(N_{l'})\right)$$
$$P\left(\mathbf{C}_{l',k} \mid S_{l'}, \pi(\mathbf{C}_{l'})\right)$$
$$\prod_{l^* \text{ adj } l'} P\left(\mathbf{D}_{l',l^*,k} \mid S_{l'}, \pi(\mathbf{D}_{l',l^*})\right)$$
$$= P(R)\; P(S_l \mid R)\; P(N_{l,k} \mid R, \pi(N_l))$$
$$P(\mathbf{C}_{l,k} \mid S_l, \pi(\mathbf{C}_l)) \prod_{l' \text{ adj } l} P(\mathbf{D}_{l,l',k} \mid S_l, \pi(\mathbf{D}_{l,l'}))$$
$$\prod_{l^*\in\mathcal{L}, l^*\neq l} \sum_{S_{l^*}} P(S_{l^*} \mid R) P(N_{l^*,k} \mid R, \pi(N_{l^*}))$$
$$P(\mathbf{C}_{l^*,k} \mid S_{l^*}, \pi(\mathbf{C}_{l^*}))$$
$$\prod_{l^{**} adj. l^*} P(\mathbf{D}_{l^*,l^{**},k} \mid S_{l^*}, \pi(\mathbf{D}_{l^*,l^{**}})).$$

and

$$P(O_k) \;=\; \sum_{R,S_l} P(R,S_l,O_k).$$

In the above formulas $\pi(.)$ denotes the parents of a variable and $\mathbf{D}_l = \{D_{l,l*}\}$ represents the discrete features for each label l - in our case, these discrete features store the number of adjacent components for each label l^* adjacent to label l.

The EM algorithm iterates until convergence: it is stopped when the parameters differ less than 0.001% at maximum compared to the previous iteration.

Likelihood of the fictitious dataset $\mathbf{O}^*$. The term $P(\mathbf{O}^* \mid G, \tilde{\Theta}_G)$ is the likelihood of the estimated MAP parameters assuming the training dataset was completed with probabilistic assignments to the hidden random variables. Given these probabilistic assignments found in the last step of the EM algorithm, the likelihood is computed as follows:

$$P(O^* \mid G, \tilde{\Theta}_G) = \prod_R P(R)^{M[R]}$$
$$\prod_{l\in\mathcal{L}} \Bigg[\prod_R \prod_{S_l} P(S_l \mid R)^{M[S_l,R]}$$
$$\prod_{R,\pi(N_l)} \prod_{N_l} P(N_l \mid R, \pi(N_l))^{M[N_l,R,\pi(N_l)]}$$
$$\prod_{l' \text{ adj } l} \prod_{S_l,\pi(\mathbf{D}_{l,l'})} \prod_{\mathbf{D}_{l,l'}} P(\mathbf{D}_{l,l'} \mid S_l, \pi(\mathbf{D}_{l,l'}))^{M[\mathbf{D}_{l,l'},S_l,\pi(\mathbf{D}_{l,l'})]}$$
$$\prod_{S_l,\pi(\mathbf{C}_l)} \prod_k P(\mathbf{C}_{l,k} \mid S_l, \pi(\mathbf{C}_{l,k}))^{P(S_{l,k}|O_k)} \Bigg],$$

where $M[\cdot]$ measures the number of times one or more discrete random variables take particular values. For the case of hidden random variables, this is replaced by the expected counts, e.g. $M[R=r] = \sum_k P(R_k = r \mid O_k)$.

Marginal likelihood of the fictitious dataset $\mathbf{O}^*$. The marginal likelihood $P(\mathbf{O}^* \mid G)$ is the likelihood times the parameter prior integrated over the parameter space, assuming the training dataset was completed with probabilistic assignments to the hidden random variables. For complete datasets, the marginal likelihood can be evaluated analytically in the case of Dirichlet priors for the discrete random variables and normal-Wishart priors for the conditional Gaussian random variables [Geiger and Heckerman 1994]:

$$P(\mathbf{O}^* \mid G) = \int_\Theta P(\mathbf{O}^* \mid \Theta, G) P(\Theta \mid G)\, \mathrm{d}\Theta$$
$$= \frac{\Gamma(|R|\alpha)}{\Gamma(|R|\alpha + K)} \prod_R \frac{\Gamma(\alpha + M[R])}{\Gamma(\alpha)}$$
$$\prod_{l\in\mathcal{L}} \Bigg[\prod_R \frac{\Gamma(|\mathcal{S}_l|\alpha)}{\Gamma(|\mathcal{S}_l|\alpha + M[R])} \prod_{S_l} \frac{\Gamma(\alpha + M[S_l,R])}{\Gamma(\alpha)}$$
$$\Bigg(\prod_{R,\pi(N_l)} \frac{\Gamma(|\mathcal{N}_l|\alpha)}{\Gamma(|\mathcal{N}_l|\alpha + M[R,\pi(N_l)])}$$
$$\prod_{N_l} \frac{\Gamma(\alpha + M[N_l,R,\pi(N_l)])}{\Gamma(\alpha)} \Bigg)$$
$$\Bigg(\prod_{l' \text{ adj } l} \prod_{S_l,\pi(\mathbf{D}_{l,l'})} \frac{\Gamma(|\mathcal{D}_{l,l'}|\alpha)}{\Gamma(|\mathcal{D}_{l,l'}|\alpha + M[S_l,\pi(N_l)])}$$
$$\prod_{\mathbf{D}_{l,l'}} \frac{\Gamma(\alpha + M[\mathbf{D}_{l,l'},S_l,\pi(N_l)])}{\Gamma(\alpha)} \Bigg)$$
$$\prod_{S_l,\pi(\mathbf{C}_l)} \frac{T_{\mathbf{C}_l}(\mathbf{C}_l, S_l, \pi(\mathbf{C}_l))}{T_{\mathbf{C}_l}(\pi(\mathbf{C}_l), S_l, \pi(\mathbf{C}_l))} \Bigg],$$

where $\Gamma(\cdot)$ is the gamma function, and

$$T_{\mathbf{C}}(\mathbf{u}, S_l, \mathbf{v}) = 2\pi^{(-pM[S_l,\pi(\mathbf{C})=\mathbf{v}]/2)} \left(\frac{1}{1 + M[S_l,\pi(\mathbf{C})=\mathbf{v}]} \right)^{\frac{p}{2}}$$
$$\cdot \frac{c(p,p)}{c(p, p + M[S_l,\pi(\mathbf{C})=\mathbf{v}])} |\Omega|^{\frac{p}{2}} |\Sigma_{\mathbf{u}}|^{-(p+M[S_l,\pi(\mathbf{C})=\mathbf{v}])/2}.$$

Here p is the number of dimensions of $\mathbf{u}$, and

$$c(p,t) = \left[2^{\frac{dt}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{t+1-i}{2} \right) \right]^{-1}.$$

References

BISHOP, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer-Verlag.

GAUVAIN, J., AND LEE, C. 1994. Maximum A Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. *IEEE Transactions on Speech and Audio Processing 2*, 291–298.

GEIGER, D., AND HECKERMAN, D. 1994. Learning Gaussian Networks. Tech. Rep. MSR-TR-94-10.

KOLLER, D., AND FRIEDMAN, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press.