Learning Latent Factor Models of Human Travel: Appendix

Michael Guerzhoy[1,3] **Aaron Hertzmann**[2,3]
[1]LEAR, INRIA Grenoble [2]Adobe Systems, Inc [3]University of Toronto

Abstract

In this document, we provide details on fitting the models described in [1]. We show how to compute the partial derivatives of the Negative Log-Likelihood (NLL) of the data under a single-cluster model in order to use them for conjugate gradient optimization. We then show how to optimize the NLL under the multiple-cluster model using a generalized EM algorithm.

1 The Single-Cluster Model

Recall that in our single-cluster model, the probability of transitioning from i to j in time interval τ is given by:

$$P_{ij\tau} = \frac{\exp\left(\rho(d_{ij}, \tau) + \alpha_j + \mathbf{u}_i^T \mathbf{v}_j\right)}{\sum_\ell \exp\left(\rho(d_{i\ell}, \tau) + \alpha_\ell + \mathbf{u}_i^T \mathbf{v}_\ell\right)} \tag{1}$$

The Negative Log-Likelihood (NLL) of the data under the single-cluster model is then:

$$\begin{aligned} NLL &= -\log \prod_{ij\tau} (P_{ij\tau})^{N_{ij\tau}} = -\sum_{ij\tau} N_{ij\tau} \log P_{ij\tau} && (2) \\ &= -\sum_{ij\tau} N_{ij\tau} \left[\rho(d_{ij}, \tau) + \alpha_j + \mathbf{u}_i^T \mathbf{v}_j - \ln \sum_\ell \exp(\rho(d_{i\ell}, \tau + \alpha_\ell + \mathbf{u}_i^T \mathbf{v}_\ell)) \right] && (3) \\ &= -\sum_{\tau d} N_{\tau d} \rho(d, \tau) - \sum_j N_j \alpha_j - \sum_{ij\tau} N_{ij\tau} \mathbf{u}_i^T \mathbf{v}_j + \sum_{ij\tau} N_{ij\tau} \ln \sum_\ell \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell + \mathbf{u}_i^T \mathbf{v}_\ell) && (4) \end{aligned}$$

Here, $N_{ij\tau}$ is the number of transitions from i to j in time interval τ (recall that the map as well as the time and the distances are discretized) and d_{ij} is the (discretized) distance between i and j. We abuse notation slightly by indicating different marginal histograms of N by different subscripts: $N_{\tau d}$ is the number of transitions over distance d in time interval τ, and N_j is the number of transitions that end at destination j.

Taking derivatives with respect to the model parameters yields:

$$\frac{\partial NLL}{\partial \alpha_j} = -N_j + \sum_{i\tau} N_{i\tau} \frac{\exp(\rho(d_{ij}, \tau) + \alpha_j)}{\sum_\ell \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)} \tag{5}$$

$$= -N_j + \sum_{i\tau} N_{i\tau} P_{ij\tau} \tag{6}$$

$$\frac{\partial NLL}{\partial \rho_{\tau d}} = -N_{\tau d} + \sum_{ij} N_{ij\tau} \frac{\sum_{\ell : d_{i\ell} = d} \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)}{\sum_\ell \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)} \tag{7}$$

$$= -N_{\tau d} + \sum_i N_{i\tau} P_{i\tau d} \tag{8}$$

$$\frac{\partial NLL}{\partial u_i} = -\sum_{j\tau} N_{ij\tau} v_j + \sum_{j\tau} N_{ij\tau} \sum_\ell v_\ell P_{i\ell\tau} \tag{9}$$

Note that the derivatives with respect to the model parameters ρ and α (Eqs. 6 and 8) can be naturally interepreted as the difference between the expected counts of transitions and the observed counts of transitions:

$$-N_j + \sum_{i\tau} N_{i\tau} P_{ij\tau} = -N_j + E[N_j] \tag{10}$$

$$-N_{\tau d} + \sum_i N_{i\tau} P_{i\tau d} = -N_{\tau d} + E[N_{\tau d}] \tag{11}$$

2 The Multiple-Cluster Model

Under the Cluster model, the probability of the trajectory $L_1^s, L_2^s, ..., L_K^s(s)$ of individual s is

$$P_s^{(c)} = P(L_1^s) \prod_{k=2}^{K(s)} P(L_k^s | L_{k-1}^s, \tau_{k-1}, c) \tag{12}$$

$$= P(L_1^s) \prod_{k=2}^{K(s)} P^{(c)}_{L_{k-1}^s L_k^s \tau_{k-1}} \tag{13}$$

where $P^{(c)}$ corresponds to P in Eq. 1, as parametrized by cluster c. The NLL for a single individual s is then

$$NLL(s) = \sum_c \pi_c P_s^{(c)} \tag{14}$$

where π_c is the prior probability of cluster c. The NLL for the entire dataset is:

$$NLL = \sum_s NLL(s) = \sum_s \sum_c \pi_c P_s^{(c)} \tag{15}$$

We optimize the NLL using a generalized EM algorithm [2], using exact E-steps and gradient descent M-steps. Each individual s has a soft-assignment variable $\gamma_{s,c}$ corresponding to the probability that the individual lies in cluster c. It is updated during the E-step:

$$\gamma_{s,c} \leftarrow \frac{P_s^{(c)}}{\sum_x P_s^{(x)}} \tag{16}$$

The generalized M-step consists of three updates. First, the cluster probabilities are updated:

$$\pi_c \leftarrow \frac{\sum_s \gamma_{s,c}}{S} \tag{17}$$

where S is the total number of individuals. Second, the parameters ρ of the individual clusters are set by minimizing the expected NLL of the individual clusters with respect to ρ:

$$\begin{aligned}
\mathcal{Q}(\rho_{1:C}, \alpha, \mathbf{u}, \mathbf{v}) &\doteq \sum_{s,c} \gamma_{s,c} \log P_s^{(c)} && (18) \\
&= \sum_{s,c} \gamma_{s,c} \sum_{k=2}^{K(s)} \log P(L_k^s | L_{k-1}^s, c) && (19) \\
&= \sum_{c} \sum_{s} \gamma_{s,c} \sum_{ij\tau} N_{ij\tau}^s \log P_{ij\tau}^{(c)} && (20)
\end{aligned}$$

where $\doteq$ indicates the omission of additive constants. Note that the individual clusters are separate in this summation, so that the optimization can be broken into C separate optimizations, one for each cluster, where the contribution of user s to cluster c is weighted by $\gamma_{s,c}$. Performing the conjugate gradient optimization in parallel on multiple cores takes no more time than optimizing the non-clustered model.

Third, to optimize the shared parameters, we observe that the derivative of $\mathcal{Q}(\theta_{1:C}, \alpha, \mathbf{u}, \mathbf{v})$ is a linear combination of derivatives that we can compute since they are analogous to the derivatives of the NLL of the single-cluster model. For example,

$$\frac{\partial \mathcal{Q}(\rho_{1:C}, \alpha, \mathbf{u}, \mathbf{v})}{\partial \mathbf{u}_i} = \sum_{c} \sum_{s} \gamma_{s,c} \frac{\partial}{\partial \mathbf{u}_i} \sum_{ij\tau} N_{ij\tau}^s \log P_{ij\tau}^{(c)} \qquad (21)$$

References

[1] M. Guerzhoy and A. Hertzmann. Learning latent factor models of human travel. In *NIPS Workshop on Social Network and Social Media Analysis: Methods, Models and Applications*, 2012.

[2] R. M. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer Academic Publishers, 1998.