Supplementary: Learning to Plan Paths in Human Environments from Large Scale Preference Feedback

Ashesh Jain, Debarghya Das, Ashutosh Saxena

Department of Computer Science, Cornell University

1 Generative Model: Learning the Parameters

Given user preference data from PlanIt (Section 5), we learn the parameters of Eq. (1). Since our goal was to make the data collection easier for users, the labels we get are either bad, neutral or good for a particular segment of the video. The challenge is that we do not know which activity a_i is being affected by a given waypoint t_i during feedback. A waypoint could even be influencing multiple activities. For example, in Fig. 1 (left) a waypoint of a robot passing between the human and TV could affect multiple watching activities.

We therefore define a latent random variable $z_a^i \in \{0, 1\}$ for waypoint t_i, such that $p(z_a^i|E)$ (or η_a) is the (prior) probability of user data arising from activity a. Incorporating this parameter in Eq. (1) gives the following cost function:

$$\mathbf{\Psi}(\{t_1, .., t_k\}|E) = \prod_{i=1}^{k} \sum_{j=1}^{|E|} p(z_{a_j}^i|E)\ \Psi_{a_j}(t_i|E) \tag{1}$$

where $|E|$ is the number of activities in environment E.[1] Figure 1 (right) shows the generative process for user preference data. Since the number of activities vary across environments (with maximum of K activities), for each environment we choose $|E|$ prior probability parameters η_a, such that $a \in E$.

Training data: We obtain user preferences over n environments $E_1, .., E_n$. For each environment E we consider m trajectory segments $\mathcal{T}_{E,1}, .., \mathcal{T}_{E,m}$ labeled as bad by users. For each segment $\mathcal{T}$ we sample k waypoints $\{t_{\mathcal{T},1}, .., t_{\mathcal{T},k}\}$. We use $\Theta \in \mathbb{R}^{30}$ to denote the model parameters and solve the following maximum likelihood problem:

$$\begin{aligned}\Theta^* &= \arg\max_{\Theta} \prod_{i=1}^{n} \prod_{j=1}^{m} \mathbf{\Psi}(\mathcal{T}_{E_i,j}|E_i;\Theta) \\ &= \arg\max_{\Theta} \prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{l=1}^{k} \sum_{h=1}^{|E_i|} p(z_{a_h}^l|E_i;\Theta)\ \Psi_{a_h}(t_{\mathcal{T}_{E_i,j},l}|E_i;\Theta)\end{aligned} \tag{2}$$

Eq. (2) does not have a closed form solution. We there use Expectation-Maximization (EM) approach to learn the parameters. In the E-step, we calculate the posterior activity assignment $p(z_{a_h}^l|t_{\mathcal{T}_{E_i,j},l}, E_i)$ for all the waypoints and we update the parameters in the M-step.

[1] We extract the information about the environment and activities by querying OpenRAVE. In practice and in the robotic experiments, human activity information can be obtained using the software package by Koppula et al. [1].

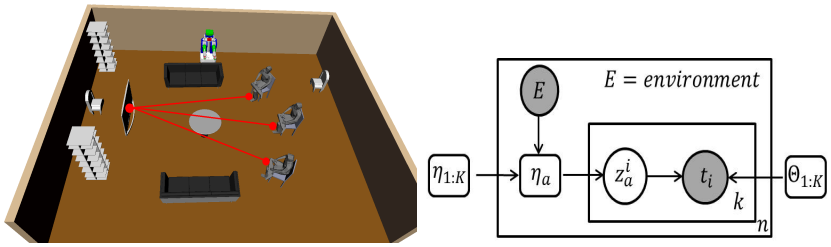

Fig. 1: **(Left)** An environment with three instances of watching activity. **(Right)** Generative process for modeling the user preference data. K is the number of activities.

E-step: In this step keeping the model parameters fixed we find the posterior probability of a waypoint t affecting an activity a_i.

$$p(z_{a_i}|t, E; \Theta) = \frac{p(z_{a_i}|E; \Theta)\Psi_{a_i}(t|E; \Theta)}{\sum_{h=1}^{|E|} p(z_{a_h}|E; \Theta)\Psi_{a_h}(t|E; \Theta)} \tag{3}$$

We calculate this posterior for every waypoint in our data set.

M-step: Using the posterior from E-step we update the model parameters in this step. Our affordance representation consists of three distributions, namely: Gaussian, von-Mises and Beta. The parameters of Gaussian, and mean (μ) of von-Mises are updated in a closed form. Following Sra [2] we perform first order approximation to update the variance (κ) of von-Mises. The parameters of beta distribution (α and β) are approximated using first and second order moments of the data.

Estimating von-Mises distribution parameters: von-Mises is parameterized by a scalar mean μ and variance κ. Mean for an activitiy a has closed form update expression:

$$\mu_a = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l|t_{\mathcal{T}_{E_i,j},l}, E_i)\mathbf{x}_{t_{\mathcal{T}_{E_i,j},l}}}{\|\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l|t_{\mathcal{T}_{E_i,j},l}, E_i)\mathbf{x}_{t_{\mathcal{T}_{E_i,j},l}}\|} \tag{4}$$

However, updating κ is not straightforward. We follow the first order approximation by Sra [2] and update κ as follows:

$$\kappa_a = \frac{\bar{R}(2-\bar{R}^2)}{1-\bar{R}^2} \tag{5}$$

$$\text{where, } \bar{R} = \frac{\|\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l|t_{\mathcal{T}_{E_i,j},l}, E_i)\mathbf{x}_{t_{\mathcal{T}_{E_i,j},l}}\|}{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l|t_{\mathcal{T}_{E_i,j},l}, E_i)} \tag{6}$$

Estimating Beta distribution parameters: Beta distribution is parameterized by two scalars α and β. We use method of moments to estimate these parameters. For an activ-

ity a, we first estimate first and second order moments i.e. sample mean and variance:

$$m_a = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)\bar{d}_{t_{\mathcal{T}_{E_i,j},l}}}{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)} \tag{7}$$

$$v_a = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)(\bar{d}_{t_{\mathcal{T}_{E_i,j},l}} - m_a)^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)} \tag{8}$$

We now estimate α and β using the first and second order moments of data:

$$\alpha_a = m_a \left(\frac{m_a(1-m_a)}{v_a} - 1 \right) \tag{9}$$

$$\beta_a = (1-m_a) \left(\frac{m_a(1-m_a)}{v_a} - 1 \right) \tag{10}$$

Estimating Gaussian distribution parameters: It is parameterized by a scalar mean g and variance σ. For an activity a we estimate parameters of Gaussian distribution in closed form.

$$g_a = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i) d_{t_{\mathcal{T}_{E_i,j},l}}}{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)} \tag{11}$$

$$\sigma_a = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)(d_{t_{\mathcal{T}_{E_i,j},l}} - g_a)^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{k} p(z_a^l | t_{\mathcal{T}_{E_i,j},l}, E_i)} \tag{12}$$

In above equations $d_{t_{\mathcal{T}_{E_i,j},l}}$ is the distance of waypoint $t_{\mathcal{T}_{E_i,j},l}$ from object/human.

Bibliography

[1] H. Koppula, R. Gupta, and A. Saxena. Learning human activities and object affordances from rgb-d videos. *IJRR*, 32(8), 2013.

[2] S. Sra. A short note on parameter approximation for von mises-fisher distributions: and a fast implementation of i s (x). *Computational Statistics*, 27(1), 2012.