User onboarding

Valerio Perrone

March 15, 2017

- We will run LDA on ICML2017 papers once they are available and find K topics, which we label by hand.

- The frontend needs to present users with these topics at onboarding and save the selected topics for each user i into a binary vector $c_i = (c_{i1}, \ldots, c_{iK})$.

- For all i we read c_i from the database (how/where is it saved? TODO: coordinate with the backend) and update our prior over $\theta_i \sim \text{Dirichlet}(\alpha + c_{i1}, \ldots, \alpha + c_{iK})$ (setting for example $\alpha = 1$, but should be tuned) so that for each k

$$\hat{\theta}_{ik} = \frac{\alpha + c_{ik}}{K\alpha + c_{i1} + \cdots + c_{iK}}.$$

- Give $\hat{\theta}$ to CTR as an input. Note: we are assuming we do not infer θ.