Dirichlet Process Mixture Language Model

Justin T. Chiu

July 11, 2019

Abstract

We extend the work in Mielke and Eisner (2018) with a Dirichlet Process mixture language model.

1 Problem Statement

Open vocabulary LM with open class latent lexemes. Maybe explicitly model some other stuff with local RVs?

Let's see what we can discover...

2 Model

Generative story. Let α be a hyperparameter and $G_0 \sim N(0, I_d)$.

1. Draw $V_i \mid \alpha, x_{<t} \sim \text{Beta}(1, \alpha), i = \{1, 2, \ldots\}$, the parameters of the stick breaking process.
2. Draw $\eta_i^* \mid G_0 \sim G_0, i = \{1, 2, \ldots\}$, the embeddings for the latent lexemes.
3. For each index $t \in 1, \ldots, T$
   (a) Draw lexeme $Z_t \mid \{v_1, v_2, \ldots\} \sim \text{Cat}(\pi(\mathbf{v}))$
   (b) Draw word $X_t \mid z_t \sim p(x_t \mid \eta_{z_t}^*)$

asdf

2.1 Parameterization

3 Training and Inference

4 Anticipated Failure Cases

1.

References

Sebastian J. Mielke and Jason Eisner. 2018. Spell once, summon anywhere: A two-level open-vocabulary language model. *CoRR*, abs/1804.08205.