Report: Experiments on α updates

Arijus Pleska

This report assesses some experiments performed on reproducing α parameters used in a generative data process. Note that the report is structures in the following sections: 1) defining the experiment settings; 2) assessing the experiment results; 3) settings some questions to be discussed during next meeting.

The Experiment Settings

The intention of the carried experiments is to identify the optimal settings for the Metropolis–Hastings algorithm application. To start with, I have generated a synthetic corpus; the parameters used in the corpus generation will allow to assess the performance achieved in the experiments. The corpus generation parameters are set as follows:

- The number of topics: K = 2;
- The number of documents (time-slices): T = 20;
- The size of vocabulary: V = 10;
- The number of words per document t: $N_t \sim \text{Pois}(\lambda), \quad \lambda = 1000.$

Further, to consider the initial settings of α_k development over documents, α_0 is a sine curve and α_1 is a cosine curve; the corresponding softmax expressions of the curves, i.e $\mu = \text{softmax}(\alpha)$, are illustrated in Figure 1 below.

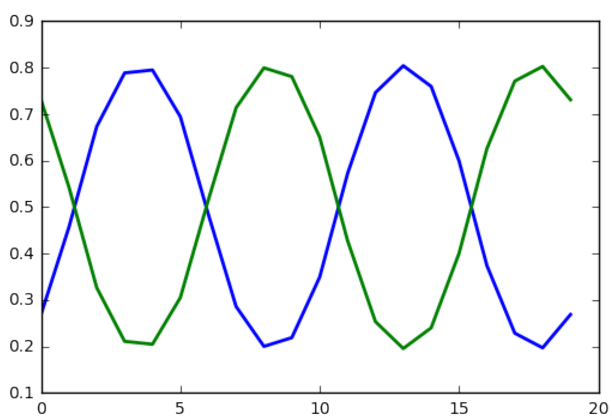


Figure 1: The values of μ used in the generative process.

Speaking of β, it was initially predefined and kept constant throughout the dynamic generative process; β is illustrated in Figure 2 below.

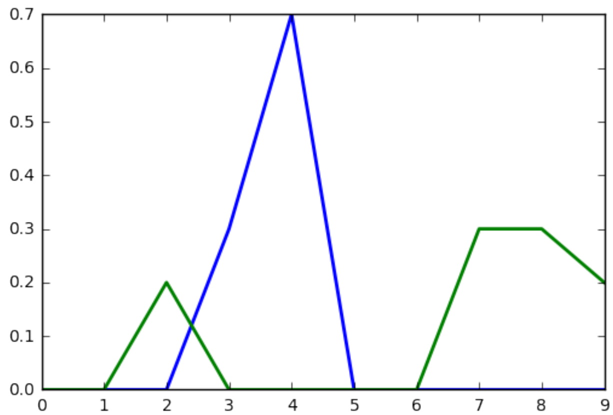


Figure 2: The values of β used in the generative process.

Note that the latter β values were applied to the autoregressive topic model for the α update experiments.

The experiment results

The first experiment is focused on discovering the choice of the variances. To be more specific, the alpha update is based on three different variances were used: the 'initial' variance $\sigma_0^2 I$ to induce α_t at $t = 0$, the 'basic' variance $\sigma^2 I$ to induce α_t at $t > 0$, and the 'proposed' variance $\delta^2 I$ to induce α'_t at $t = 0$; also, note that α'_t at $t > 0$ were induced using the 'basic' variance.

For both experiments, the number of autoregressive iterations was set to 500 and σ^2 was set to 0.1. The resulting plots of μ with different values of σ_0 and δ^2 are illustrated in Figures 3, 4, 5 below.

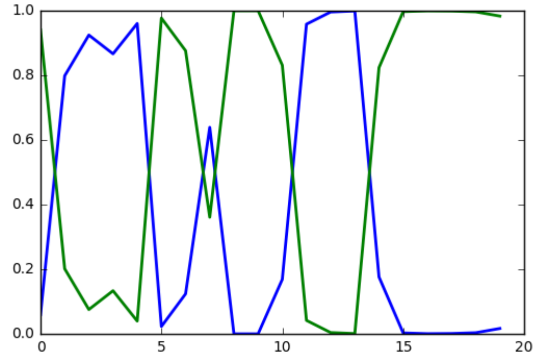
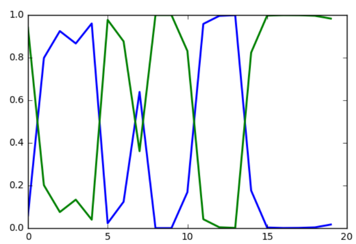
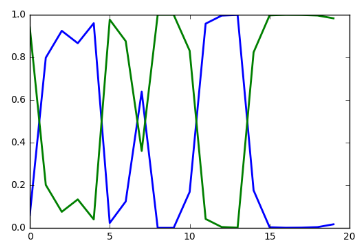


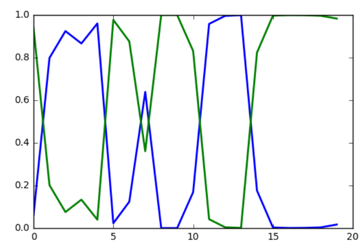
Figure 3: The values of β used in the generative process.



(a) A gull



(b) A gull2



(c) A tiger

Figure 4: Pictures of animals

The second experiment was carried to determine the impact of the α update in recovering the original topic fluctuations in the synthetic corpus. For this reason, the autoregressive part of the dynamic topic was disabled. The topic assignments to the documents of the last iteration are visualised in Figure 6 below.

Observations

For the last section of this report, I have set some questions to be addressed during next meeting; these are listed below:

- In some cases, the larger number of iterations reduced the performance in reproducing the initial α;

Current Stage

During the experiment, I have used the following settings:

- The synthetic data has been created by inducing the previously implemented dynamic topic modelling (DNT) generative process:
  - The number of documents: $|\mathrm{D}| \approx 6000$;
  - The size of the vocabulary: $|\mathrm{V}| \approx 2000$;
  - The number of words per document: $\mathrm{N}_d \approx 20, \quad \forall d \in \mathrm{D}$;
  - Instead of intensity values, it is assumed that the document dictionaries contain word counts. For example, $d_{111} = \{v_{20} : 15, v_{40} : 5\}$.
- The number of topics: $K = 10$;
- The number of time-slices: $T = 50$;
- The alpha at $t = 0$: $\alpha_0 \sim \mathcal{N}(\mu_0, \sigma_0^2 I), \quad \mu_0 = 0.1, \quad \sigma_0^2 = 0.2$;
- The alphas at $t > 0$: $\alpha_t \sim \mathcal{N}(\alpha_{t-1}, \sigma^2 I), \quad \sigma^2 = 0.1$;
- The candidate alphas: $\alpha_t' \sim \mathcal{N}(\alpha_t, \delta^2 I), \quad \delta^2 = 2$;
- The acceptance rate: $r_t = \min(1, p(\alpha_t')/p(\alpha_t))$;
- The probability of the state: $p(\alpha_t) = p(\alpha_t|\alpha_{t-1}) \cdot p(\alpha_{t+1}|\alpha_t) \cdot \pi(\alpha_t)$, where π is a mapping to the mean parameterisation;

The rationale of the implementation follows the following principle: α_t is set to α_t' on the successful 'toss' based on r_t. Also, the variances are tuned to obtain $r_t \approx 30\%$.

Issues

My uncertainties with the proposed solution are the following:

- The estimation of $p(\alpha_t)$:
  - The third term of the expression, $\pi(\alpha_t)$, represents the topic distribution in documents in time-slice t;
  - The current model treats the vocabulary term distributions over the topics, β, to have same values; therefore, this term was omitted – it cancels out upon the estimation of r_t;
  - The first (and second) term $p(\alpha_t|\alpha_{t-1})$ is drawn from $\mathcal{N}(\alpha_{t-1}, \sigma^2 I)$.
- Since α_t is a vector, the initial r_t is a vector as well.