Bayesian Allocation Model and Non-negative Matrix Factorization

Barış Can Esmer, Yusuf Hakan Kalaycı
Advisor : Ali Taylan Cemgil
Boğaziçi Üniversitesi

Abstract

We investigate a stochastic process called "Bayesian Allocation Model" where tokens are allocated randomly to a tensor and the probability for each index is specified by a graphical model. By integrating out some parameters of the model analytically, one obtains a Polya Urn Model, which enables one to calculate statistics about the model more efficiently by respecting the sparsity of the tensor.

By exploiting the relationship between this dynamic generative model and non-negative matrix/tensor factorization, we can look at NMF from another perspective.

Introduction

In the last decades, several different approaches for the analysis of relational data have been proposed. Topic models and matrix/tensor decompositions are some of the basic tools that have been widely adopted in various application domains such as

- Recommendation systems
- Knowledge base completion
- Algorithm representations of tensors

Tensors appear as a natural candidate for the representation of relational data when two or more entities are involved. Exact or approximate decomposition can be states as $X \approx WH$ which may provide a latent representation of data that can be used for

- Prediction
- Anomaly detection
- Denoising

In addition to matrix factorization, Non-negative matrix factorization has important properties making it popular reasearch topic:

- Clustering property
- Ease of interpretation

Bayesian Allocation Model

Bayesian Allocation model (BAM) is a generative dynamical model that is easy to describe.

$$\theta_k(:) \sim \mathcal{D}(\alpha_i(:))$$
$$\theta_{i|k}(:,k) \sim \mathcal{D}(\alpha_{i|k}(:,k))$$
$$\theta_{j|k}(:,k) \sim \mathcal{D}(\alpha_{j|k}(:,k))$$
$$\lambda \sim \mathcal{GA}(a,b)$$
$$S_{ikj} \sim \mathcal{PO}(\lambda\theta_{i|k}(i,k)\theta_{j|k}(j,k)\theta_k(k))$$

With this model NMF of X can be found by considering $p(\theta,\lambda|S_{ij}=X_{ij})$. Finding most likely θ, λ will yield us the factorization of $X \approx W \times H$ with defining:

- $W = \theta_{i|k}(i,k)$
- $H = \theta_{j|k}(j,k) \times \theta_k(k) \times \lambda$

Also, Jacobian Transformation will give us the result of KL-NMF:

$$W_{:,k} \sim \mathcal{D}(\alpha_{i,k})$$
$$H_{k,j} \sim \mathcal{GA}(\alpha_{j|k}, b)$$
$$X_{i,j} \sim \mathcal{PO}(\sum_k W_{ik}H_{kj})$$

Polya Urn Interpretation of BAM

Pólya Urn Models are a family of statistical models that exhibits self-reinforcing property, which can be expressed the *"the rich gets richer"* property.

$X_1, X_2, \dots$ is said to be *exchangeable* if for any permutation $X_{i_1}, X_{i_2}, \dots$ we have

$$\mathbb{P}(X_1, X_2, \dots) = \mathbb{P}(X_{i_1}, X_{i_2}, \dots)$$

If we consider the predictive distribution of BAM, we have

$$\mathbb{P}(s^\tau \mid s^{1:\tau-1}) = p(s^\tau \mid S^{\tau-1}) = p(S^\tau \mid S^{\tau-1})$$

The first equality follows from the fact that the random variables $\{s^t\}_{t=1}^{\tau}$ are exchangeable and so does the second one since the increments are by 1.

$$\text{BAM } (j \leftarrow k \rightarrow i)$$

$$\mathbb{P}(s^\tau_{ikj} = 1 \mid S^{\tau-1}) = \frac{\alpha_{+k+} + S^{\tau-1}_{+k+}}{\alpha_+ + S^{\tau-1}_+} \frac{\alpha_{+kj} + S^{\tau-1}_{+kj}}{\alpha_{+k+} + S^{\tau-1}_{+k+}} \frac{\alpha_{ik+} + S^{\tau-1}_{ik+}}{\alpha_{+k+} + S^{\tau-1}_{+k+}}$$

Marginal Likelihood & Monte Carlo Algo

Let $S_{i+j} = \sum_k S_{ikj}$ and call $\mathbb{P}(S_{i+j} = X)$ the marginal likelihood of X.

$$\mathbb{P}(S_{i+j} = X) = \mathbb{P}(S_{i+j} = X \mid S_+ = T)\mathbb{P}(S_+ = T)$$

Focusing on first term,

$$\mathbb{P}(S_{i+j} = X \mid S_+ = T) = \sum_{S^T} \mathbb{1}[S^T_{i+j} = X]\mathbb{P}(S^T)$$

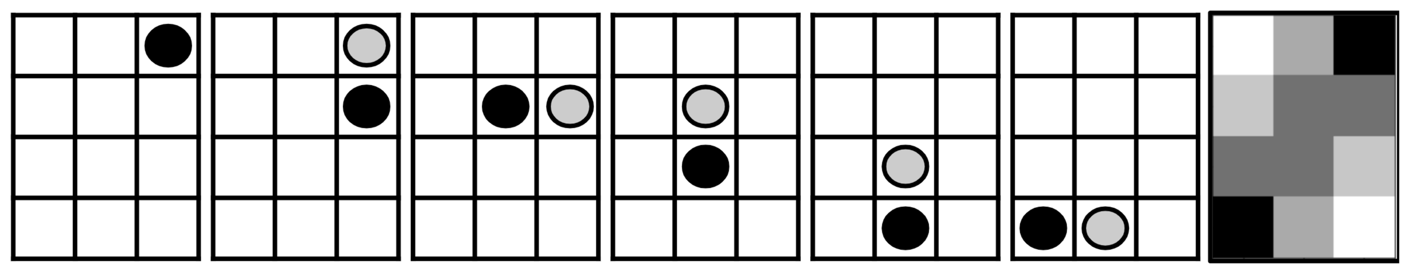
and its Monte Carlo estimate is equal to

$$\mathbb{P}(S_{i+j} = X \mid S_+ = T) = \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}[S^{T(m)}_{i+j} = X]$$

Sequential Importance Sampling

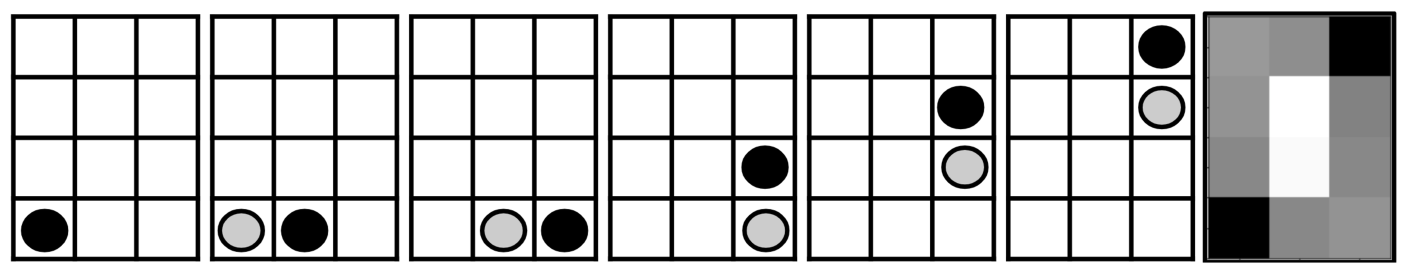
- Backward Proposal

$$q_{back}(s^{1:T}) = q_{back}(s^{1:T}_{i,j})q(s^{1:T}_k \mid s^{1:T}_{i,j})$$



- Forward Proposal

$$q_{forw}(s^{1:T}) = q_{forw}(s^{1:T}_{i,j})q(s^{1:T}_k \mid s^{1:T}_{i,j})$$



Experiments

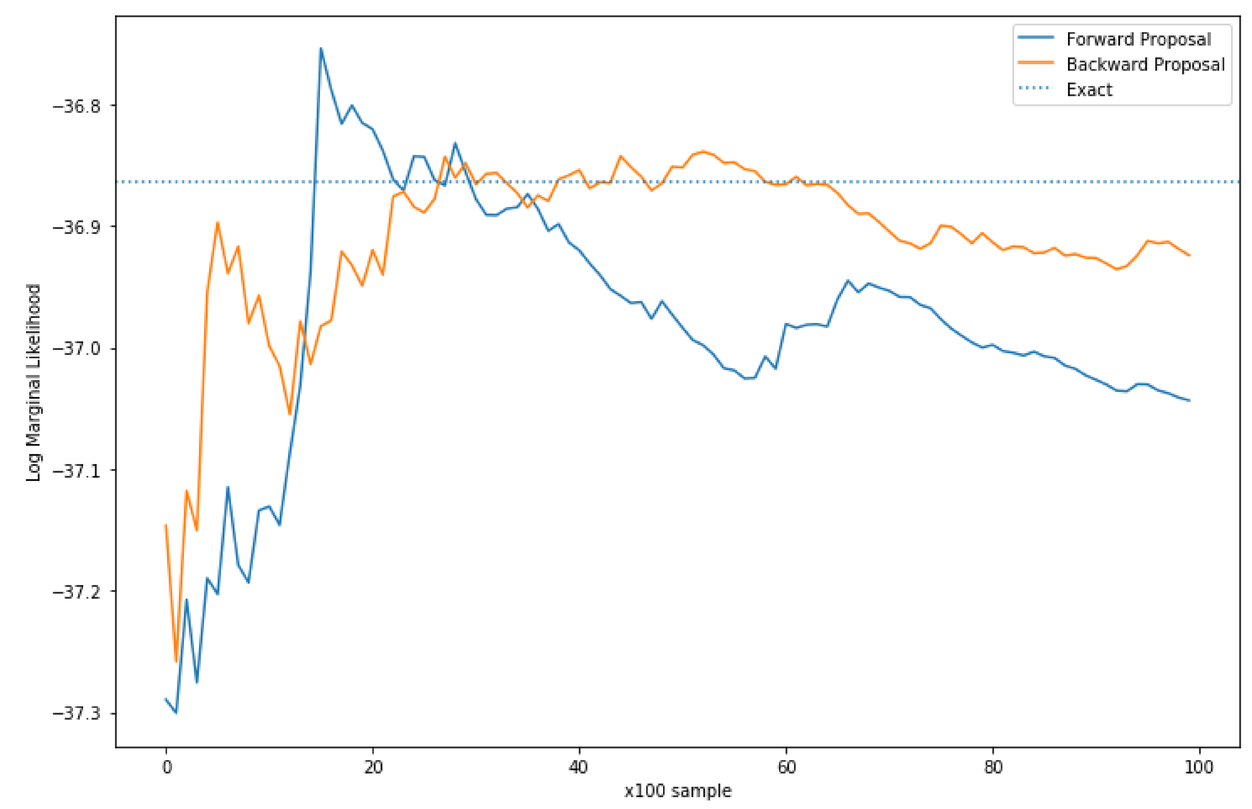


Fig.1 Forward proposal vs Backward proposal

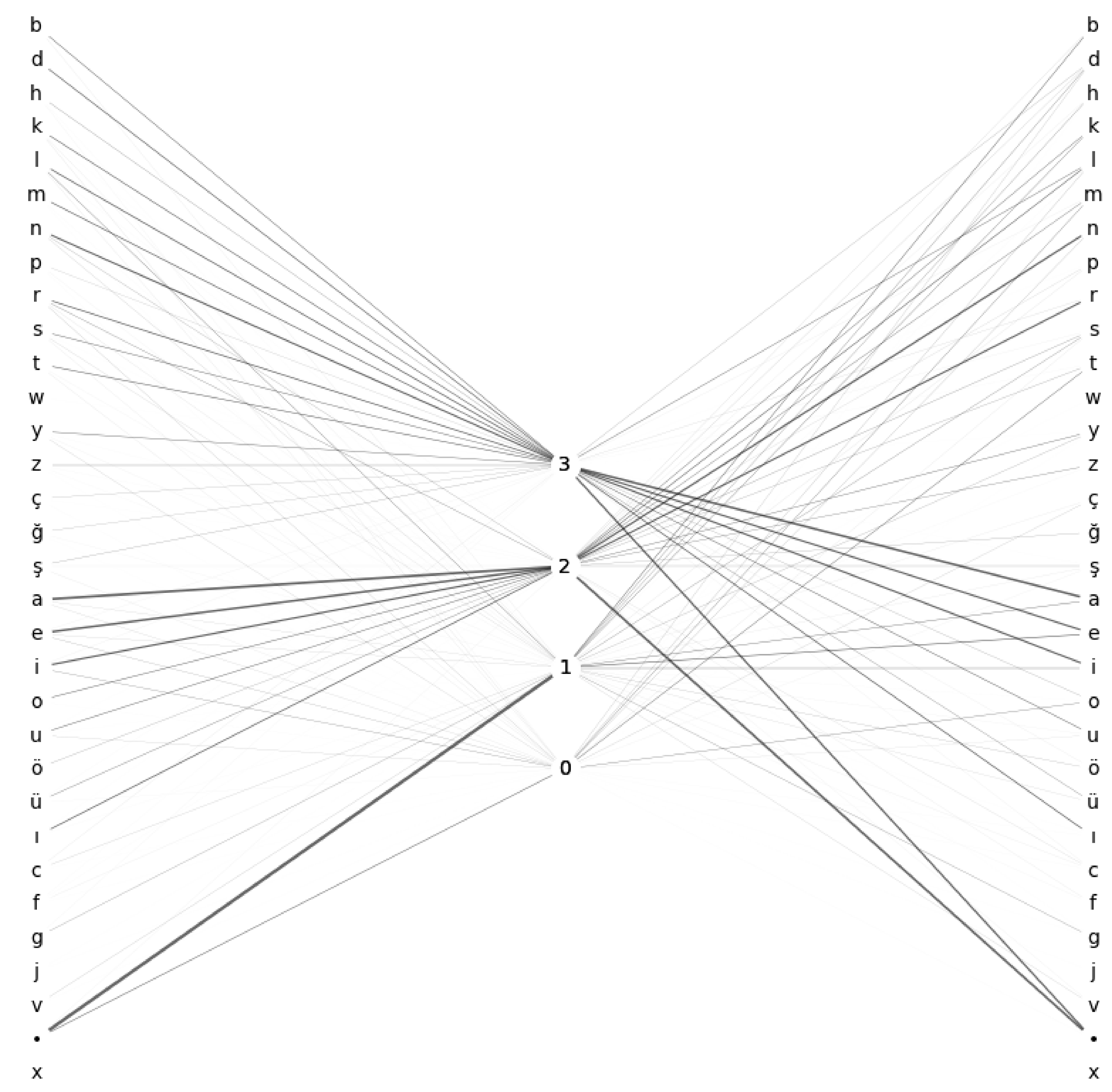


Fig.2 Decomposition of letter transition (bigram) data of Hamlet (Turkish).

Conclusion and Acknowledgement

In this project, we familiarized ourselves with topics such as Bayesian Allocation Model, Nonnegative Matrix Factorization and Polya urn models. Specifically, we investigated the relationship between KL-NMF and BAM.

We would like to thank **Burak Kurutmaz** and **Melih Barsbey** for their generous time and efforts that made this project possible.

Reference

1. Ali Taylan Cemgil, Burak Kurutmaz, Umut Şimşekli, Sinan Yıldırım, *Bayesian Allocation Model: Scoring and Model Selection for Nonnegative Tensor Decompositions by simulating Polya Urns Preprint*
2. Hosam Mahmoud, *Polya Urn Models*, Chapman and Hall/CRC, 1st Edition, 2008.