

Question about the paper : Evolutionnary Hierarchical Dirichlet Processes for Multiple Correlated Time Varying Corpora

1 Question 1:

When computing the table T_{jk}^t , do you consider that :

1. For $t=0$, π_{jk}^{t-1} does not exist and then :

$$T_{jk}^t / \beta_k^t, \pi_{jk}^{t-1}, N_{jk}^t \sim CRP(\alpha_0^t \beta_k^t, N_{jk}^t)?$$

and not

$$T_{jk}^t / \beta_k^t, \pi_{jk}^{t-1}, N_{jk}^t \sim CRP(\alpha_0^t v_j^t \pi_{jk}^{t-1} + \alpha_0^t (1 - v_j^t) \beta_k^t, N_{jk}^t)$$

2. For $t=T$, $T_{jk}^{t \Rightarrow t+1}$ is assumed to be equal to 0 ?

For recap, the procedure you wrote is the following one :

$$(T_{jk}^{t \Rightarrow t+1}, T_{jk}^{0 \Rightarrow t+1}) \sim \text{Multinomiale}(T_{jk}^{t+1}, [p, 1 - p]), (22)$$

with

$$p = \frac{v_j^{t+1} \pi_{jk}^t}{(1 - v_j^{t+1}) \beta_k^{t+1} + v_j^{t+1} \pi_{jk}^t}$$

$$T_{jk}^t / \beta_k^t, \pi_{jk}^{t-1}, N_{jk}^t \sim CRP(\alpha_0^t v_j^t \pi_{jk}^{t-1} + \alpha_0^t (1 - v_j^t) \beta_k^t, N_{jk}^t)$$

2 Question 2:

As explained in (20), (21) and (22), to compute Z_{ji}^t , after computing the posterior distribution of Z_{ji}^t / x_{ji}^t , we find that:

$$P(z_{ji}^t = k / x_{ji}^t) \sim P(z_{ji}^t = k / \pi_j^t) \cdot P(x_{ji}^t / z_{ji}^t = k \dots)$$

we know that:

$$P(z_{ji}^t = k / \pi_j^t) = \pi_{jk}^t$$

moreover,

$$P(x_{ji}^t / z_{ji}^t = k \dots) = \frac{\Gamma(n+1) \Gamma(\sum_{a \in A, w} X_{aw} + \alpha_w) \prod_{w=1}^W \Gamma(\alpha_w + x_{jiw}^t + \sum_{a \in A} X_{aw})}{\Gamma(\sum_{a \in A, w} X_{aw} + \alpha_w + x_{jiw}^t) \prod_{w=1}^W [\Gamma(x_{jiw}^t + 1) \Gamma(\alpha_w + \sum_{a \in A} X_{aw})]}$$

with $A = ((i, j, t), Z_{ji}^t = k)$.

The last computation appears because $P(x_{ji}^t/z_{ji}^t = k...)$ is a posterior distribution of multinomial and dirichlet distributions.
I used this technique to optimize the weights and this is not the one you used in your HDP code in Java on Github.

3 Question 3:

When the results of your implementation are given in Table (3), you say that $\phi_k, 1 = 0.1, 0.30....$ But actually, if I am not wrong, $\phi_k \sim Dir(\mu)$ where μ can be computing aposteriori. So my question is :

Are the ϕ_k 's given in Table(3) samples or means of the law of ϕ_k ?

4 Question 4:

How long did your algorithm take to give results on real data set?

When my algorithm runs on real data set, it converges to a single topic. Did you have this problem ?

Thank you very much for your time, and feel free to give me any advice about my research.

Sincerely
Florian

A useful Java Code

<https://github.com/yqsong/dirichlet-processes/tree/master/DirichletProcesses/src/model/clustering>

Link to your article

<https://dl.acm.org/citation.cfm?id=1835940>

Link to my implementation

https://github.com/BastinFlorian/EvoHDP/blob/master/Evo_hdp.ipynb.