How to train PLSA?

How would you train the model?

Probabilistic Latent Semantic Analysis:

$$p(w|d) = \sum_{t \in T} p(w|t)\, p(t|d) = \sum_{t \in T} \phi_{wt} \theta_{td}$$

Parameters of the model:

- ϕ_{wt} – *probability of word* ***w*** *in topic* ***t***
- θ_{td} – *probability of topic* ***t*** *in document* ***d***

How would you train the model?

Log-likelihood optimization:

$$\log \prod_{d \in D} p(d) \prod_{w \in d} p(w|d)^{n_{dw}} \to \max_{\Phi, \Theta}$$

$$\Updownarrow$$

$$\sum_{d \in D} \sum_{w \in d} n_{dw} \log \sum_{t \in T} \phi_{wt} \theta_{td} \to \max_{\Phi, \Theta}$$

Given non-negativity and normalization constraints:

$$\phi_{wt} \geq 0 \qquad \theta_{td} \geq 0$$

$$\sum_{w \in W} \phi_{wt} = 1 \qquad \sum_{t \in T} \theta_{td} = 1$$

How would you train the model?

Log-likelihood optimization:

$$\log \prod_{d \in D} p(d) \prod_{w \in d} p(w|d)^{n_{dw}} \to \max_{\Phi, \Theta}$$

$$\Updownarrow$$

$$\sum_{d \in D} \sum_{w \in d} n_{dw} \log \sum_{t \in T} \phi_{wt} \theta_{td} \to \max_{\Phi, \Theta}$$

Given non-negativity and normalization constraints:

$$\phi_{wt} \geq 0 \qquad \theta_{td} \geq 0$$

$$\sum_{w \in W} \phi_{wt} = 1 \qquad \sum_{t \in T} \theta_{td} = 1$$

We have just plain texts

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

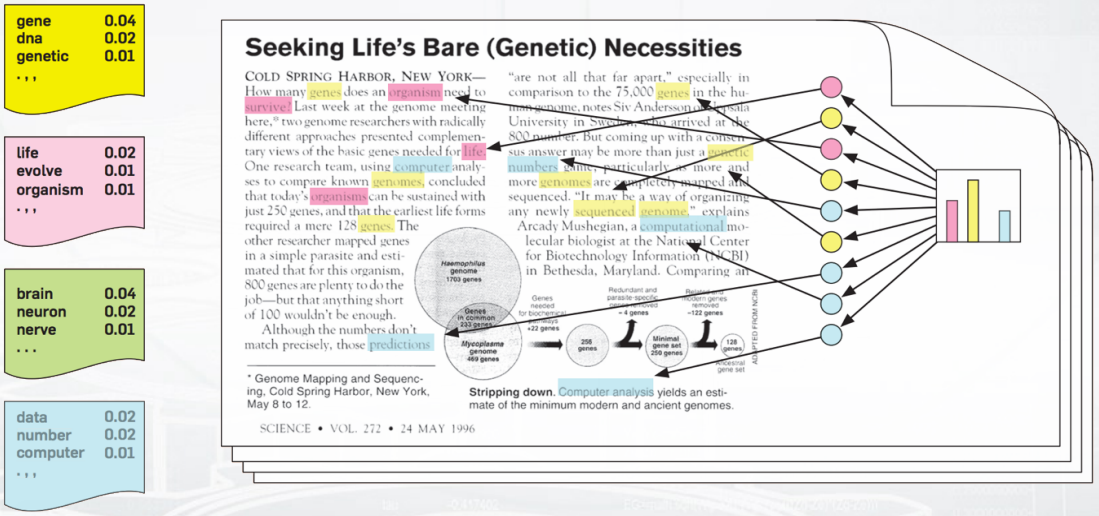
We have just plain texts

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

Topics

Documents

Topic proportions and assignments



If we knew topic assignments…

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

If we knew topic assignments…

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

We would just count:

$$p(w = sky|t) = \frac{n_{wt}}{\sum_w n_{wt}} = \frac{1}{4}$$

If we knew topic assignments…

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

We would just count:

$$p(w = sky|t) = \frac{n_{wt}}{\sum_w n_{wt}} = \frac{1}{4}$$

$$p(t = t|d) = \frac{n_{td}}{\sum_t n_{td}} = \frac{4}{54}$$

But we have just plain texts

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

But we have just plain texts

Pooh rubbed his nose again, and said that he hadn't thought of that. And then he brightened again, and said that, if it were raining already, the Heffalump would be looking at the sky wondering if it would clear up, and so he wouldn't see the Very Deep Pit until he was half-way down…

Idea! Let's estimate the topic assignment probabilities!

$$p(t|d,w) = \frac{p(w,t|d)}{p(w|d)} = \frac{p(w|t)p(t|d)}{p(w|d)}$$

Bayes rule *Product rule*

Put everything together: EM-algorithm

E-step:

$$p(t|d,w) = \frac{p(w|t)p(t|d)}{p(w|d)} = \frac{\phi_{wt}\theta_{td}}{\sum_{s\in T}\phi_{ws}\theta_{sd}}$$

M-step:

$$\phi_{wt} = \frac{n_{wt}}{\sum_w n_{wt}} \qquad n_{wt} = \sum_d n_{dw}\, p(t|d,w)$$

$$\theta_{td} = \frac{n_{td}}{\sum_t n_{td}} \qquad n_{td} = \sum_w n_{dw}\, p(t|d,w)$$