CSED524-Homework 4

Inhyuk Na

2018/11/25

1 Introduction

In the lecture, we've learned about Latent Semantic Analysis(LSA) and Probabilistic Latent Semantic Analysis(PLSA). In this report, we introduce LSA and PLSA briefly, and explain what are the strengths and weakness of PLSA, compared to LSA. All of the contents are based on [2].

2 LSA and PLSA

The key idea of LSA is to map high-dimensional count vectors to a lower dimensional representation in a so-called latent semantic space. The goal of LSA is to find a data mapping which reveals semantical relations between the entries of interest. The mapping is restricted to be linear and is based on a Singular Value Decomposition of the co-occurrence table.

PLSA is a probabilistic variant of LSA. By [2], PLSA has a sound statistical foundation and defines a proper generative model of the data. PLSA is based on the aspect model. The aspect model is a latent variable model for co-occurrence data which associates an unobserved class variable $z \in \mathcal{Z} = \{z_1, ..., z_K\}$ with each observation. The joint probability model can be written either

$$P(d, w) = P(d) \sum_{z \in \mathcal{Z}} P(w|z)P(z|d) \text{ (asymmetric parameterization), or}$$

$$P(d, w) = \sum_{z \in \mathcal{Z}} P(z)P(d|z)P(w|z) \text{ (symmetric parameterization).}$$

Thus, we can do the maximum likelihood estimation via EM algorithm.

3 Strenghs and weakness of PLSA compared to LSA

By Hofmann[2], PLSA has many strengths compared to LSA. First, the mixture approximation $\boldsymbol{P}$ of the co-occurrence table is a well-defined probability distribution and factors have a clear probabilistic meaning, where we define $\boldsymbol{P} = \hat{\boldsymbol{U}}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{V}}^t$, $\hat{\boldsymbol{U}} = (P(d_i|z_k))_{i,k}$, $\hat{\boldsymbol{V}} = (P(w_j|z_k))_{j,k}$, and $\hat{\boldsymbol{\Sigma}} = (P(z_k))_k$. However, LSA does not define a properly normalized probability distribution and the approximation $\tilde{\boldsymbol{N}}$ may even contain negative entries. Thus, the directions in the PLSA space are interpretable as multinomial word distributions, but there is no obvious interpretation of the directions in the LSA latent space. In addition, PLSA can take advantage for model selection and complexity control, by well-established statistical theory. However, LSA demand on ad hoc heuristics.

However, PLSA has weakness compared to LSA, which is the computational complexity. Thus, EM algorithm of PLSA is only guaranteed to find a local maximum of the likelihood function, but SVD of

LSA is guaranteed to find exact solution. However, Hofmann's experiments[2] show that EM algorithm is not significantly slower than SVD, and PLSA experimentally outperforms than LSA.

Hofmann shows the strengths of PLSA by two experiments. One is related to perplexity minimization and the other is related to information retrieval. By perplexity evaluation, PLSA outperforms than LSA. In information retrieval, PLSA also outperforms than LSA and PLSA also has robust properties. Thus, Hofmann says that PLSA also allow us to systematically combine different models according to Bayesian model combination scheme.

Without comparing with LSA, David Blei[1] points out some weakness of PLSA. First, PLSA is not a well-defined generative model of documents, and there is no natural way to use it to assign probability to a previously unseen document. Second, the number of parameters grows linearly with the number of training documents. PLSA tries to overcome this by tempering heuristic, but it has been shown that overfitting can occur. David Blei says that his Latent Dirichlet allocation overcomes these weaknesses.

References

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.