MM6761: Take-home Assignment 3

Dai Yao (dai@{yaod.ai, mle.bi})

April 6, 2024

1 Puranam, Narayan and Kadiyali (MKSC 2017), and Liu and Toubia (MKSC 2018)

For Puranam et al. (2017), how is the model estimated? Can we estimate the model without estimating Equation (2) first, and then use the outcome to estimate Equation (2)?

Answer:

The parameters in the model can be split into three major categories: (1) those related to the classical topic models, including the hyperparameters α_d and β_d, the document level topic proportions θ_d, the vector of word level assignments of topics z_d, and the topic level parameter ϕ_k; (2) λ_k which capture the effects of observed characteristics of the document on α_d (Equation 1 on page 729), and (3) ν_{kd} which are the coefficients from the DiD specification (Equation 2 on page 729).

Since λ_k affect z (Equation 5 on page 730, Equation 7 on page 731), the parameters in (1) and (2) are hence estimated together, using a combination of Gibbs sampler (to estimate everything except λ_k) and Stochastic Expectation Maximization (to estimate λ_k). After obtaining an estimate of θ_{kd}, ν_{kd} is estimated by a simple linear regression. Hence, we can estimate the model without ν_{kd} first, and then we can use the outcomes in each iteration of the MCMC to estimate ν_{kd}.

For Liu and Toubia (2018), how is the model estimated? Can we estimate the model without estimating Equation (3) first, and then use the outcome to estimate Equation (3)?

Answer:

The parameters in the model can be split into two major categories: (1) those related to the classical topic model for the queries and the web pages, including ϕ_k, θ_p and z_p, and θ_q and z_q, and (2) R which capture the semantic relationships between θ_p and θ_q.

They use a Gibbs sampler to estimate ϕ_k, θ_p, z_p, and z_q first since the posterior distributions of them are conjugate. They then use a Metropolis-Hastings algorithm to estimate θ_q because the posterior distribution (Equation A.5 on page 947) is no longer conjugate. They finally use maximum likelihood estimator to estimate R since it only depends on θ_p and θ_q.

We cannot estimate the model without R first, because R affects both θ_p (Equation A.4 on page 947) and θ_q (Equation A.5 on page 947) in their estimation process. To draw some analogy, R are similar to λ in Puranam et al (2017) in terms of the linkage with the main model.

For Liu and Toubia (2018), what's your opinion about Equation (3)? Why do you think it is innovative? Do you see any potential problems? If so, what are the problems?

Answer:

Please see the detailed discussions of main problem of the model in the commentary paper for this paper, which is included in the reading list (Zhang, Yao, Zhang, 2021).

2 Chakraborty, Kim, and Sudhir (2022)

Comparing Chakraborty et al. (2022) to the previous two papers, what are the main differences in terms of model development (2 to 3 main points)? By model development, you may focus on how machine learning based algorithms are used in the model development process, and how these algorithms are integrated with traditional marketing models into a unified framework.

Answer:

- Both Puranam et al (2017) and Liu and Toubia (2018) extend the classical topic model (i.e., LDA) and incorporate new features. Chakraborty et al (2022) do not develop new machine learning algorithms. Instead, they use a deep learning hybrid convolutional-LSTM neural network to convert review texts to numeric attribute sentiment scores. The main body of the paper is a structural model about reviewers' rating behaviors which also account for missing attributes in reviewers' ratings.

- In Chakraborty et al (2022), the deep learning model and the structural model are clearly separated from each other, while in the previous two studies, they only have one model and the various components in each model are closely related to each other.

- In Chakraborty et al (2022), the Bayes rule is used to impute missing attribute score at the segment level. No missing data is accommodated in the previous two studies.

What are the variables in X^q_{jk} and X^w_{ijk}, respectively? How A_{ijk} is computed if it is missing?

Answer:

The coefficients for X^q_{jk} are α_k (Equation 2 on page 609), hence, the covariates include those reported in Table 7 (page 615). The coefficients for X^w_{ijk} are δ^g_k, hence, the covariates include those reported in Table 8 (page 617).

When A_{ijk} is missing, they use the Bayes rule to calculate the probability for A_{ijk} to realize each rating score s (Equation 5 on page 610). The idea is:

- Given X^q_{jk} and C^g_k, we can infer the probability for A_{ijk} to realize a specific score s (Equation 3 on page 609);

- Given X^w_{ijk} (in which A_{ijk} is a covariate (see Sentiment 1, 2, 4, 5 covariates in Table 8 on page 617), we can infer the probability for A_{ijk} to be missing.

- Given the two probabilities, we can apply the Bayes rule to impute the missing value (Equation 5 on page 610).