MM6761: Take-home Assignment 1

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

February 24, 2024

1 Chen et al. (2017)

The set of model parameters in this paper is: $\{L, \hat{B}^S, Q, B^S, B\}$, and each element in this set itself is a collection of various parameters.

- Based on the notation, please specify the focal model in the paper (i.e., MCH), the UCH model, as well as the FM model. Please also briefly discuss their differences.

  Answer:

  MCH (Section 2.3 on page 142):

  The ratings vector $\mathbf{Y}_i$ is generated based on β_i, which is generated based on the segment-level partworths β_i^S.

$$\mathbf{Y}_i = \mathbf{X}_i\beta_i + \epsilon_i \tag{1}$$
$$\beta_i = \beta_i^S + \eta_i. \tag{2}$$

  UCH:

  The ratings vector $\mathbf{Y}_i$ is generated based on β_i, which is generated based on the population-level partworths $\bar{\beta}$. Different from MCH, in UCH, all the individual-level parthworths are drawn from a common, population-level partworths vector.

  Hence, we keep Equation (1) the same, and twist Equation (2) to specify the UCH model.

$$\beta_i = \bar{\beta} + \eta_i. \tag{3}$$

  FM:

  The ratings vector $\mathbf{Y}_i$ is generated based on β_i. Among the population, there are S segments. Individual i belongs to a particular segment and adopts the segment-level partworths directly. Different from MCH, in FM, the individual-level parthworths are exactly the same as the segment-level partworths.

  Hence, we keep Equation (1) the same, and twist Equation (2) to specify the FM model.

$$\beta_i = \beta_i^S. \tag{4}$$

- In Section 2.3 (page 142), right above Assumption 1 (A1), the authors said that "a closer examination reveals that learning $\{B^S, B\}$ is sufficient, as ..." Please explain why.

Answer:

First, β_i, β_i^S, and $\hat{\beta}_i^S$ are all $n \times 1$ column vectors, where n is the number of elements in $\mathbf{X}_i$. Thus, $\hat{B}^S \triangleq \{\hat{\beta}_l^S\}_{l=1}^L$ is a $n \times L$ matrix, and both $B^S \triangleq \{\beta_i^S\}_i^I$ and $B \triangleq \{\beta_i\}_i^I$ are $n \times I$ matrices.

The number of unique columns in B^S is L. Also, if we start from $i = 1$ to $i = I$, and retain only the unique columns and arrange them by the appearance order, then we obtain $\hat{B}^S$. Finally, with L and $\hat{B}^S$, we can easily compute Q.

2 Text Embeddings

Please describe the similarities (one to two aspects) and main differences (one to two aspects as well) between the two major approaches to construct text embeddings: Dhillon and Aral (2021), and Ansari, Li, and Zhang (2018).

Answer:

Main differences:

- Model: DA builds a matrix factorization (unsupervised) model and incorporates it into a deep neural network structure, while ALZ adopts the supervised latent Dirichelet allocation (LDA) approach.
- Heterogeneity: DA models dynamic user heterogeneity (via u_i^t, see Section 4.2 on page 1063), while ALZ models static user heterogeneity (via $\boldsymbol{\gamma}_i$, see Figure 4 and Section 4.6 on page 993).
- Estimation: DA uses standard optimization methods to train its model, while ALZ uses stochastic variational Bayesian approach (i.e., a combination of optimization and Bayesian estimation).
- Many more points.

Similarities:

- Task: Both are about text analysis using advanced machine learning methods.
- Findings: Both uncover consumer content preferences (relatively static in ALZ, but over time as well in DA).
- Maturity: Both papers are solid in terms of both methodology and substantive findings.
- Many more points if you will.