Recommendation System HFT Model [1]

Ximing

Contents

- Latent-Factor Recommendation Systems
- Latent Dirichlet Allocation
- HFT Model

Latent-Factor Recommendation Systems

$$rec(u, i) = \alpha + \beta_u + \beta_i + \gamma_u \cdot \gamma_i$$

- α : offset parameter
- β_u : user bias
- β_i : item bias
- γ_i : K-dimensional item factors
  - the extent to which items exhibit certain properties
- γ_u : K-dimensional user factors
  - the extent to which users are attracted to those properties

Latent-Factor Recommendation Systems

$$\min_{\Theta} \frac{1}{|\Gamma|} \sum_{r_{u,i} \in \Gamma} \left(rec(u,i) - r_{u,i} \right)^2 + \lambda \Omega(\Theta)$$

- Γ : training corpus of ratings
- $\Theta = \{\alpha,\ \beta_u,\ \beta_i,\ \gamma_u,\ \gamma_i\}$
- $\Omega(\Theta)$: regularizer
  - penalize complex models, e.g. $\|\gamma\|_2^2$
  - when items and users have few ratings, this would push γ_u and γ_i towards zero, leaving only the offset and bias to $rec(u,i)$
- gradient descent methods to solve this problem

Latent Dirichlet Allocation

$$p(\Gamma|\theta,\phi,z) = \prod_{d\in\Gamma}\prod_{j=1}^{N_d}\theta_{d,z_{d,j}}\phi_{z_{d,j},w_{d,j}}$$

- d : document
- $\theta_{d,k}$: the probability words in document d discuss topic k
  - θ_d follows the Dirichlet distribution
- ϕ_k : word distribution for topic k
  - probability that a particular word is used for that topic
- $z_{d,j}$: topic assignment for jth word in document d
  - integer between 1 and K

Latent Dirichlet Allocation

$$p(\Gamma|\theta,\phi,z) = \prod_{d\in\Gamma}\prod_{j=1}^{N_d}\theta_{d,z_{d,j}}\phi_{z_{d,j},w_{d,j}}$$

- $p(\Gamma|\theta,\phi,z)$
  - likelihood of corpus Γ
- $\theta_{d,z_{d,j}}$
  - likelihood of seeing particular topics
- $\phi_{z_{d,j},w_{d,j}}$
  - likelihood of seeing particular word for this topic

- $p(\Gamma|\theta,\phi,z)$ is maximized
  - $\Phi = \{\theta,\phi\}$ and z are traditionally updated via sampling

HFT Model

- Hidden Factors as Topics (HFT)
  - Latent-Factor Recommendation System
  - Latent Dirichlet Allocation
  - Combine these two ideas

- document d_i : all reviews of a particular item i
  - when users review items, they tend to discuss properties of the item more than they discuss their own personal preferences

HFT Model

- Transformation between $\theta_{i,k}$ and $\gamma_{i,k}$
  - $\gamma_{i,k}$ is the extent to which item i exhibit property k
    - $\gamma_i \in R^K$
  - $\theta_{i,k}$ is the probability reviews of item i discuss topic k
    - $\theta_i \in \Delta^K \quad (\text{i.e.,} \quad \theta_{i,k} \geq 0, \ \sum_k \theta_{i,k} = 1)$
  - If an item exhibits a certain property (high $\gamma_{i,k}$), this will correspond to a particular topic being discussed (high $\theta_{i,k}$)

$$\theta_{i,k} = \frac{\exp(\kappa \gamma_{i,k})}{\sum_{k'} \exp(\kappa \gamma_{i,k'})}$$

- κ is fit during learning, it controls the peakiness of the tranformation
  - κ large : users only discuss the most important topic
  - κ small : users discuss all topics evenly

HFT Model

$$\min f(\Gamma|\Theta, \Phi, \kappa, z) = \frac{1}{|\Gamma|} \sum_{r_{u,i} \in \Gamma} \left(rec(u,i) - r_{u,i}\right)^2 - \mu l(\Gamma|\theta, \phi, z)$$

- $\Theta = \{\alpha,\ \beta_u,\ \beta_i,\ \gamma_u,\ \gamma_i\}$: rating parameters
- $\Phi = \{\theta, \phi\}$: topic parameters
- κ : transformation parameter between θ and γ
- z : set of topic assignments for each word in corpus Γ

- $\frac{1}{|\Gamma|} \sum_{r_{u,i} \in \Gamma} \left(rec(u,i) - r_{u,i}\right)^2$ is the rating error
- $l(\Gamma|\theta, \phi, z)$ is the corpus likelihood
- μ is to trade-off the importance of these two effects

HFT Model

- Solving $\underset{\Theta,\Phi,\kappa,z}{\operatorname{argmin}} f(\Gamma|\Theta,\Phi,\kappa,z)$

- Stochastic optimization procedure
  - update $\Theta^{(t)},\Phi^{(t)},\kappa^{(t)} = \underset{\Theta,\Phi,\kappa}{\operatorname{argmin}} f\left(\Gamma\middle|\Theta,\Phi,\kappa,z^{(t-1)}\right)$
    - Gradient descent method L-BFGS
    - For $\Theta = \{\alpha,\ \beta_u,\ \beta_i,\ \gamma_u,\ \gamma_i\}$ and $\Phi = \{\theta,\phi\}$, θ is determined by γ_i
    - ϕ is determined by introducing additional variable ψ to make sure $\phi_k \in \Delta^N$

$$\phi_{k,w} = \frac{\exp(\psi_{k,w})}{\sum_{w'} \exp(\psi_{k,w'})}$$

  - sample $z_{i,j}^{(t)}$ with probability $p(z_{i,j}^{(t)} = k)\ \propto \theta_{i,k}^{(t)}\phi_{k,w_{i,j}}^{(t)}$
    - Assign each word to a topic randomly, with probability proportional to the likelihood of that topic occurring with that word
    - $\theta_{i,k}$ is the probability of the topic k being used for this item i
    - $\phi_{k,w_{i,j}}$ is the probability of the particular word $w_{i,j}$ being used for the topic k
  - the two steps are repeated until convergence

Reference

- [1] McAuley, J., and Leskovec, J. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems* (2013), ACM, pp. 165-172.