Social network and click-through prediction with factorization machines

Seung-hwan baek

2013.5.13

FM

$$\hat{y}^*(\mathbf{x}) := w_0 + \sum_{j=1}^{p} w_j \, x_j$$
$$+ \sum_{l=1}^{|\mathcal{B}|} \sum_{l'>l}^{|\mathcal{B}|} \sum_{j \in B_l} \sum_{j' \in B_{l'}} x_j \, x_{j'} \sum_{f=1}^{k} v_{j,f} \, v_{j',f}$$

Block 개념 사용

- acceleration을 하려고 노력(non zero element를 줄이자)
- 같은 block안의 interaction은 없는거라고 봄.

- 하나의 카테고리에 해당하는 모든 variable을 한 block으로
- User를 나타내는 attribute들을 하나의 block으로 등.

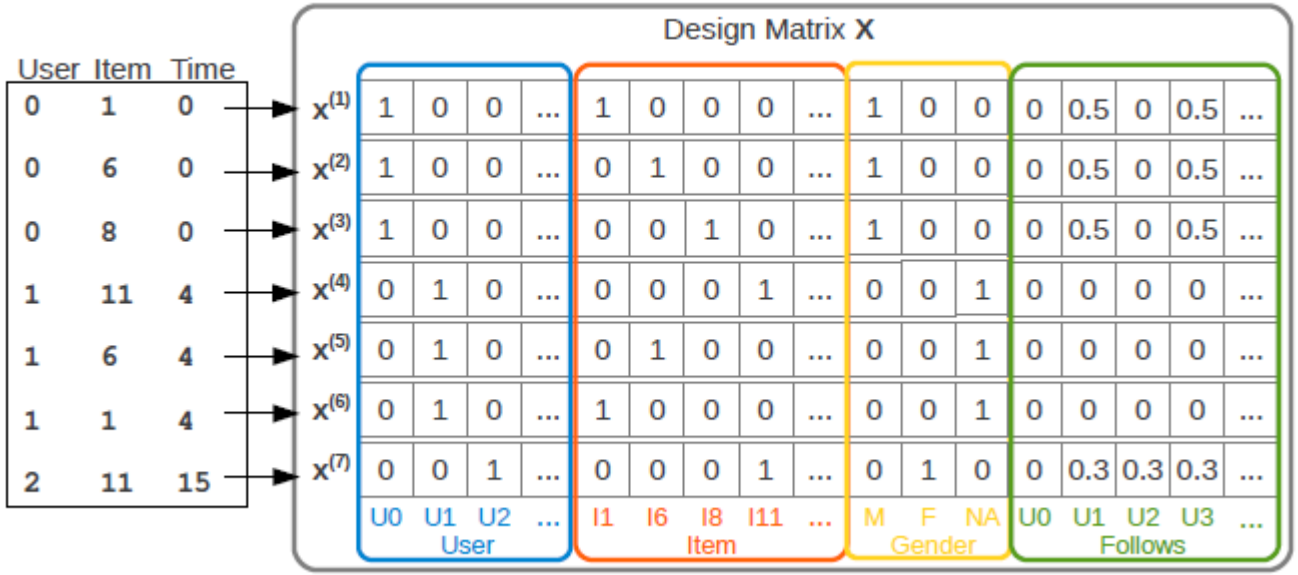
- One block per categorical variable

Categorical variable
categorical variable is a variable that can take on one of a limited, and usually fixed, number of possible values

Matrix

- Categorical variable: Binary로 저장
- time은 사용하지 않음, 그러나 sequential feature는 뽑아내서 사용.
- User action도 사용하지 않음

Column ➔ feature에 해당. 여기엔 총 4개의 category가 있고 매우 많은 predictor variable이 있음.



Markov Chain Monte Carlo inference(MCMC)

- Regularization parameter를 model에 넣어서 자동으로 결정되게 만듦.

- Standard deviation: 0.05
- # of factors: k

Method	k	# Samples	MAP3 (public)	MAP3 (private)
FM with user interactions	32	128	0.42405	0.41111
		256	0.42514	0.41192
FM without user interactions	22	128	0.42663	0.41491
		256	0.42802	0.41577
		384	0.42833	0.41582
Ensemble	n/a	n/a	0.42909	**0.41622**

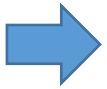
Predicate

1. Main
   - User ID
   - Item ID
2. User Attributes & Social Network
   - Age, gender, # of tweets of the user
   - Tags, keyword
   - Set of all users that the user follows
3. Sequential information

- Item info, user action table: not used.

Sequential information

- Log scale truncated duration of
  - this session
  - Previous session
  - Next session
  - Next next session
  - Next next next session
- Session index in descending order
- # of
  - Sessions in the next 60 secs.
  - Sessions in the previous 60 secs.
- Visit index

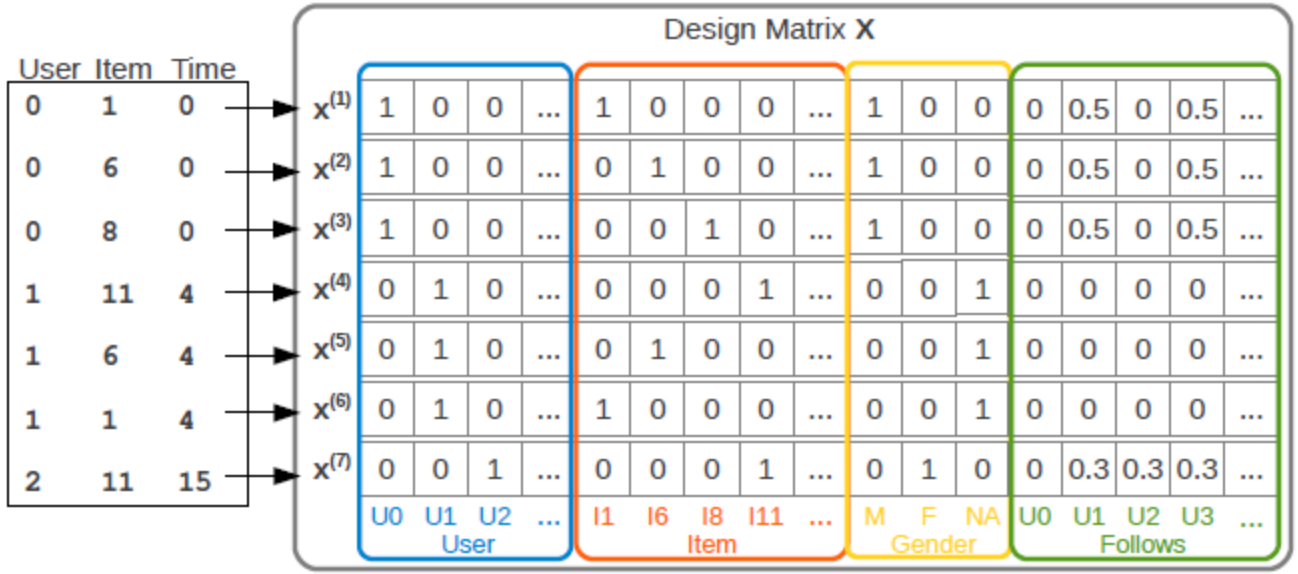


Categorical variable in binary form

Ensemble

- FM with user interaction
  - Block안에서 interaction 무시됨.
- FM without user interaction
  - 추가적으로 user variable 간의 interaction이 무시됨.

    Ex) Interaction between User id and age is removed.



Summary

- FM을 사용하는 이유: Large categorical domain의 variable
- MCMC를 사용
- Predicate variable을 어떤걸 사용하느냐
  - Sequential info도 사용
  - 사용하지 않는 정보도 있음: Item attribute, user action 등.
- Block의 개념을 이용해 acceleration시킴
  - 어떤 variable들을 하나의 Block으로 넣을 것인가
- Ensemble