Social Network and Click-through Prediction with Factorization Machines

Steffen Rendle

Universität Konstanz

Social Network Analysis, University of Konstanz

KDDCup Workshop, 12th August 2012, Beijing

Outline

Problem Setting

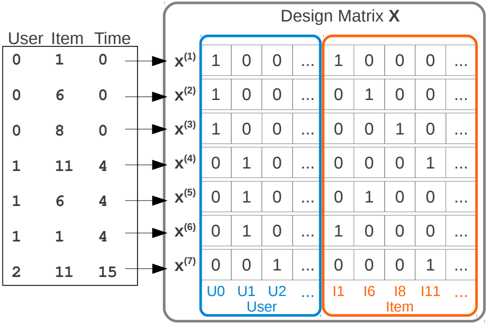
- Main variables in both tasks are IDs, i.e. categorical variables of very large domain.
  - Estimating variable interactions is challenging.
- Both tasks include many additional variables.
  - Flexible model, i.e. feature engineering is desirable.

Design Matrices

User	Item	Time
0	1	0
0	6	0
0	8	0
1	11	4
1	6	4
1	1	4
2	11	15

Problems can be encoded with a real valued design matrix: $X \in \mathbb{R}^{n \times p}$.

Design Matrices



Problems can be encoded with a real valued design matrix: $X \in \mathbb{R}^{n \times p}$.

Design Matrices

User	Item	Time
0	1	0
0	6	0
0	8	0
1	11	4
1	6	4
1	1	4
2	11	15

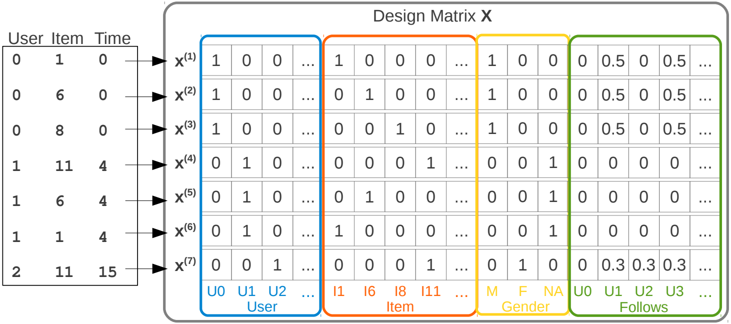
Design Matrix **X**

	User				Item					Gender			Follows				
	U0	U1	U2	...	I1	I6	I8	I11	...	M	F	NA	U0	U1	U2	U3	...
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	1	0	0	0	0.5	0	0.5	...
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	1	0	0	0	0.5	0	0.5	...
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	1	0	0	0	0.5	0	0.5	...
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	0	1	...	0	0	1	0	0	0	0	...
$\mathbf{x}^{(5)}$	0	1	0	...	0	1	0	0	...	0	0	1	0	0	0	0	...
$\mathbf{x}^{(6)}$	0	1	0	...	1	0	0	0	...	0	0	1	0	0	0	0	...
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	0	1	...	0	1	0	0	0.3	0.3	0.3	...

Problems can be encoded with a real valued design matrix: $X \in \mathbb{R}^{n \times p}$.

- many problems can be described.
- many models can be applied.

Design Matrices



Problems can be encoded with a real valued design matrix: $X \in \mathbb{R}^{n \times p}$.

- many problems can be described.
- many models can be applied.
  - Here only Factorization Machines are used.

Factorization Machines

- Let $\mathbf{x} \in \mathbb{R}^p$ be an input vector with p predictor variables.
- Model equation (degree 2):

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{p} w_i \, x_i + \sum_{i=1}^{p} \sum_{j>i}^{p} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \, x_i \, x_j$$

- Model parameters:

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^p, \quad \mathbf{V} \in \mathbb{R}^{p \times k}$$

- Properties:
  - Computing model equation is in $\mathcal{O}(N_z(\mathbf{x}) \, k)$
  - FMs are multilinear.

[Rendle, ICDM 2010]

Factorization Machines: Learning/ Inference

- FMs can be learned for regression and classification with
  - Stochastic Gradient Descent (SGD)
  - Alternating Least Squares (ALS) / Coordinate Descent (CD)
  - Markov Chain Monte Carlo (MCMC)

- Computational complexity $\mathcal{O}(N_z(X)\,k)$ (per iteration)

[Rendle, TIST 2012; Freudenthaler et al., NIPS-WS 2011]

Outline

Problem Setting

Track 1: Follower Prediction in a Social Network
- Feature Engineering
- Results

Track 2: Clickthrough Prediction

Summary

Track 1: Follower Prediction in a Social Network

Learning Task

- Binary classification problem (evaluated with ranking measure)

Track 1: Follower Prediction in a Social Network

Learning Task

- Binary classification problem (evaluated with ranking measure)

Predictor Variables

- Main variables
  - UserID
  - ItemID
- User Attributes & Social Network
  - age, gender, number tweets of the user
  - tags, keywords
  - set of all users that the user follows (from the social network)

Track 1: Follower Prediction in a Social Network

Learning Task

- Binary classification problem (evaluated with ranking measure)

Predictor Variables

- Main variables
  - UserID
  - ItemID
- User Attributes & Social Network
  - age, gender, number tweets of the user
  - tags, keywords
  - set of all users that the user follows (from the social network)
- Sequential information

Track 1: Follower Prediction in a Social Network

Learning Task

- Binary classification problem (evaluated with ranking measure)

Predictor Variables

- Main variables
  - UserID
  - ItemID
- User Attributes & Social Network
  - age, gender, number tweets of the user
  - tags, keywords
  - set of all users that the user follows (from the social network)
- Sequential information
- Other information has not been used (e.g. item info, user action table)

Sequential Information

User	Item	Time
0	1760350	0
0	1774722	0
0	786313	0
1	1775029	4
1	1902321	4
1	462104	4
2	1774509	15
2	1774717	15
2	1775024	15
...		
0	1774957	238
0	1775024	238
0	696419	238
...		
0	1928678	242
0	2105464	242
0	647356	242
...		
0	1774708	304
0	1775009	304
0	1928678	304
...		
0	372259	49922
0	458026	49922
0	563514	49922
...		

Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	
1	1902321	4	
1	462104	4	
2	1774509	15	
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	
0	1775024	238	
0	696419	238	
...			
0	1928678	242	
0	2105464	242	
0	647356	242	
...			
0	1774708	304	
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	
0	458026	49922	
0	563514	49922	
...			

Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	
0	1775024	238	
0	696419	238	
...			
0	1928678	242	
0	2105464	242	
0	647356	242	
...			
0	1774708	304	
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	
0	458026	49922	
0	563514	49922	
...			

Sequential Information

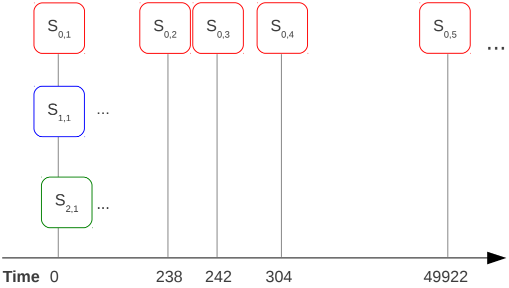
User	Item	Time	**Session**
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	$S_{2,1}$
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	$S_{0,2}$
0	1775024	238	
0	696419	238	
...			
0	1928678	242	$S_{0,3}$
0	2105464	242	
0	647356	242	
...			
0	1774708	304	$S_{0,4}$
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	$S_{0,5}$
0	458026	49922	
0	563514	49922	
...			

Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	$S_{2,1}$
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	$S_{0,2}$
0	1775024	238	
0	696419	238	
...			
0	1928678	242	$S_{0,3}$
0	2105464	242	
0	647356	242	
...			
0	1774708	304	$S_{0,4}$
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	$S_{0,5}$
0	458026	49922	
0	563514	49922	
...			

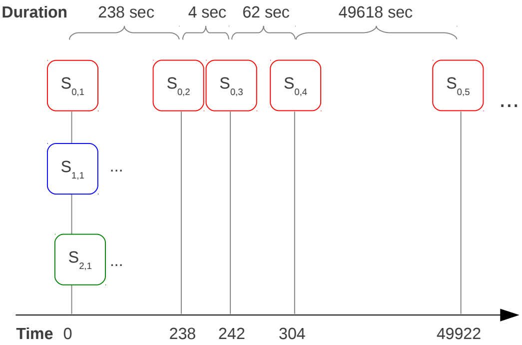
Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	$S_{2,1}$
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	$S_{0,2}$
0	1775024	238	
0	696419	238	
...			
0	1928678	242	$S_{0,3}$
0	2105464	242	
0	647356	242	
...			
0	1774708	304	$S_{0,4}$
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	$S_{0,5}$
0	458026	49922	
0	563514	49922	
...			



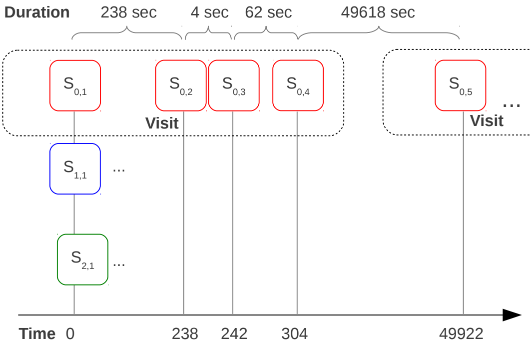
Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	$S_{2,1}$
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	$S_{0,2}$
0	1775024	238	
0	696419	238	
...			
0	1928678	242	$S_{0,3}$
0	2105464	242	
0	647356	242	
...			
0	1774708	304	$S_{0,4}$
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	$S_{0,5}$
0	458026	49922	
0	563514	49922	
...			



Sequential Information

User	Item	Time	Session
0	1760350	0	$S_{0,1}$
0	1774722	0	
0	786313	0	
1	1775029	4	$S_{1,1}$
1	1902321	4	
1	462104	4	
2	1774509	15	$S_{2,1}$
2	1774717	15	
2	1775024	15	
...			
0	1774957	238	$S_{0,2}$
0	1775024	238	
0	696419	238	
...			
0	1928678	242	$S_{0,3}$
0	2105464	242	
0	647356	242	
...			
0	1774708	304	$S_{0,4}$
0	1775009	304	
0	1928678	304	
...			
0	372259	49922	$S_{0,5}$
0	458026	49922	
0	563514	49922	
...			



Sequential Information

For each user-item-time triple, the following sequential information is generated:

- duration of this session
- duration of previous session
- duration of next session
- duration of next next session
- duration of next next next session
- session index
- session index in descending order
- number of sessions in the next 60 seconds
- number of sessions in the previous 60 seconds
- visit index

Results

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**

Outline

Problem Setting

Track 1: Follower Prediction in a Social Network

Track 2: Clickthrough Prediction
- Feature Engineering
- Results

Summary

Track 2: Clickthrough Prediction

Learning Task

- Weighted regression problem (evaluated with ranking measure)

Track 2: Clickthrough Prediction

Learning Task

- Weighted regression problem (evaluated with ranking measure)

Predictor Variables

- Main variables
  - AdID
  - UserID
  - QueryID
  - position in the impression list.
- Attributes
  - user gender
  - user age
  - query tokens
  - title tokens

Track 2: Clickthrough Prediction

Learning Task

- Weighted regression problem (evaluated with ranking measure)

Predictor Variables

- Main variables
  - AdID
  - UserID
  - QueryID
  - position in the impression list.
- Attributes
  - user gender
  - user age
  - query tokens
  - title tokens
- Other information has not been used (e.g. URL, DescriptionID, etc.)

ID Model

- Main variables
  - AdID
  - UserID
  - QueryID
  - position in the impression list.

Variable	Levels overall	Levels in test	Levels only in test
Ad ID	670,560	300,012	28,853
User ID	23,907,495	3,263,681	1,883,948
Query ID	26,243,385	3,801,978	2,121,309

ID Model

- Main variables
  - AdID
  - UserID
  - QueryID
  - position in the impression list.
- Most levels of UserID and QueryID in test do not appear in training set.

Variable	Levels overall	Levels in test	Levels only in test
Ad ID	670,560	300,012	28,853
User ID	23,907,495	3,263,681	1,883,948
Query ID	26,243,385	3,801,978	2,121,309

ID Model

- Main variables
  - AdID
  - UserID
  - QueryID
  - position in the impression list.
- Most levels of UserID and QueryID in test do not appear in training set.
- SGD with early stopping is chosen as learning algorithm for ID based models.

Variable	Levels overall	Levels in test	Levels only in test
Ad ID	670,560	300,012	28,853
User ID	23,907,495	3,263,681	1,883,948
Query ID	26,243,385	3,801,978	2,121,309

Attribute Model

- ID variables of user and query are replaced with attributes:
  - AdID
  - user gender
  - user age
  - query tokens
  - title tokens
  - position in the impression list

Attribute Model

- ID variables of user and query are replaced with attributes:
  - AdID
  - user gender
  - user age
  - query tokens
  - title tokens
  - position in the impression list
- MCMC is used for learning.

Mixed Model

- Both attributes and IDs are used:
  - AdID
  - UserID
  - QueryID
  - user gender
  - user age
  - query tokens
  - position in the impression list.

Mixed Model

- Both attributes and IDs are used:
  - AdID
  - UserID
  - QueryID
  - user gender
  - user age
  - query tokens
  - position in the impression list.
- SGD with early stopping is chosen as learning algorithm.

Results

Model	Inference	wAUC (public)	wAUC (private)
ID-based model ($k = 0$)	SGD	0.78050	0.78086
Attribute-based model ($k = 8$)	MCMC	0.77409	0.77555
Mixed model ($k = 8$)	SGD	0.79011	0.79321
Final ensemble	n/a	0.79857	**0.80178**

Ensemble

- Rank positions (not predicted clickthrough rates) are used.
- The MCMC attribute-based model and different variations of the SGD models are included.

Outline

Problem Setting

Track 1: Follower Prediction in a Social Network

Track 2: Clickthrough Prediction

Summary

Summary

- Main variables in both tasks have large categorical domains.
  - Interactions are hard to estimate.
  - Factorization Models are appealing.
- Many predictor variables are available.
  - Feature-engineering based techniques facilitate model definition.
- Factorization Machines combine flexibility of feature-engineering with advantages of factorization models.