Adapting recommendations to short-term shopping goals

Lukas Lerche, M.Sc.
TU Dortmund
LS XIII Informatik

Joint work with:
Dietmar Jannach
Matthäus Schmedding

Cooperation partner:
Tobias Lang
(Zalando)

A small example

- Past purchases:


- Present views:


- What to recommend?

Short-term shopping goals

- Assumption:
  - User has a reason to visit a web shop 
  - Strong indicator for the type of product he wants
  - But: His past behaviour might be important, too!

- RS has to adapt to the intended goal fast
  - Without forgetting the users general taste 

Research vs. reality

- Common in RS research:
  - Rating matrix (User/Item-matrix)
  - No contextual information

- Real-world scenario web shops:
  - Few ratings (if any)
  - Log data always available
  - Context known (item data, user demographics, …)

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

*anonymized/simplified

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

User-item assignment

*anonymized/simplified

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

Product information

*anonymized/simplified

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

Session information

*anonymized/simplified

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

4 different user actions

*anonymized/simplified

Log data structure*

Item	User	Brand	Color	Category	Session	Action
I081	U041	B072	Green	Shoes	S022	VIEW
I027	U023	B099	Blue	Trousers	S029	SALE
I012	U014	B076	Grey	Jackets	S037	WISH
I044	U064	B025	Red	Caps	S042	CART
I091	U088	B077	Yellow	T-shirts	S055	SALE
I032	U088	B077	Red	T-shirts	S055	SALE
I087	U088	B077	Black	T-shirts	S066	SALE
I011	U088	B066	Black	Shoes	S077	VIEW
I054	U088	B077	Green	Shoes	S077	VIEW

Actions of the example user

*anonymized/simplified

Research focus

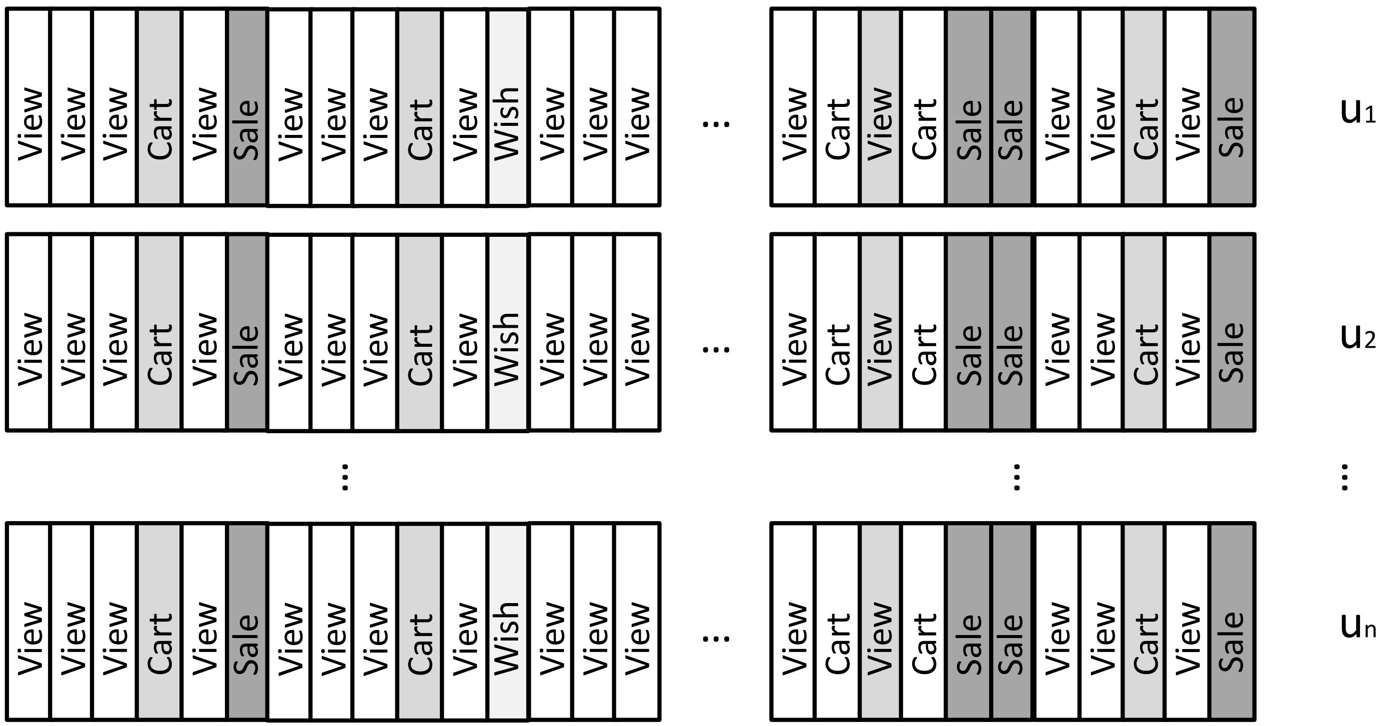
1. Creating an offline evaluation protocol with short-term shopping goals in mind

2. Adapting and comparing different recommendation techniques on short-term shopping goals

The protocol

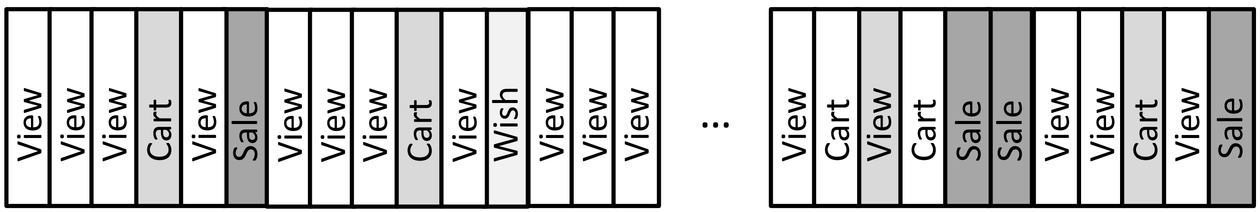
- Session-based evaluation protocol for offline log data
- Exploit available information about the sequence of the visitor actions
- Take the short-term interests into account by revealing recent actions to the recommender

Short-term adapting recommendation

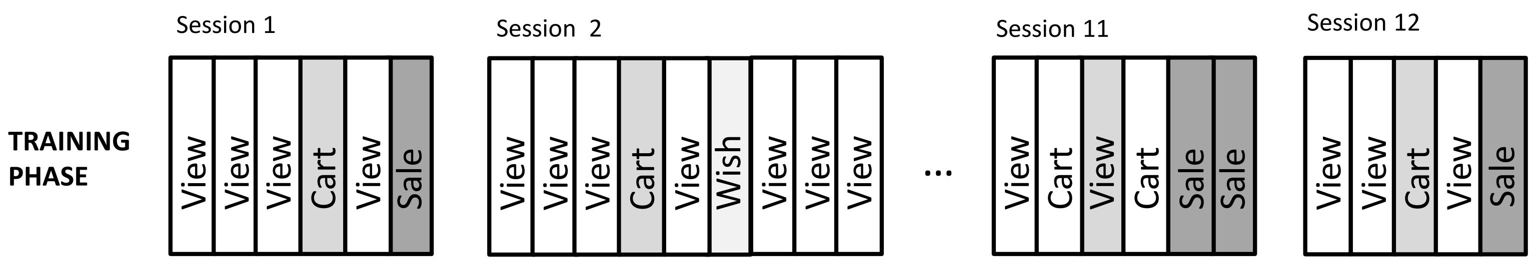


1) Get actions per user

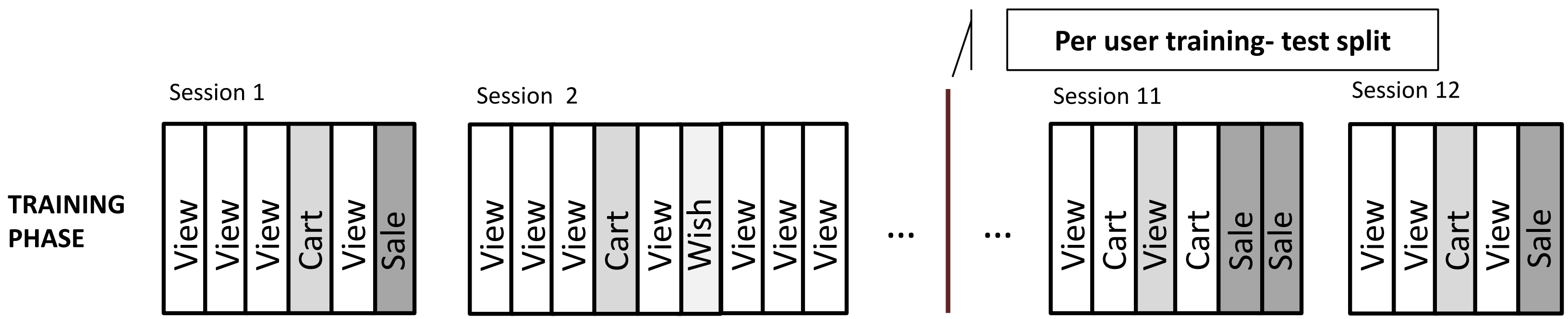
TRAINING PHASE



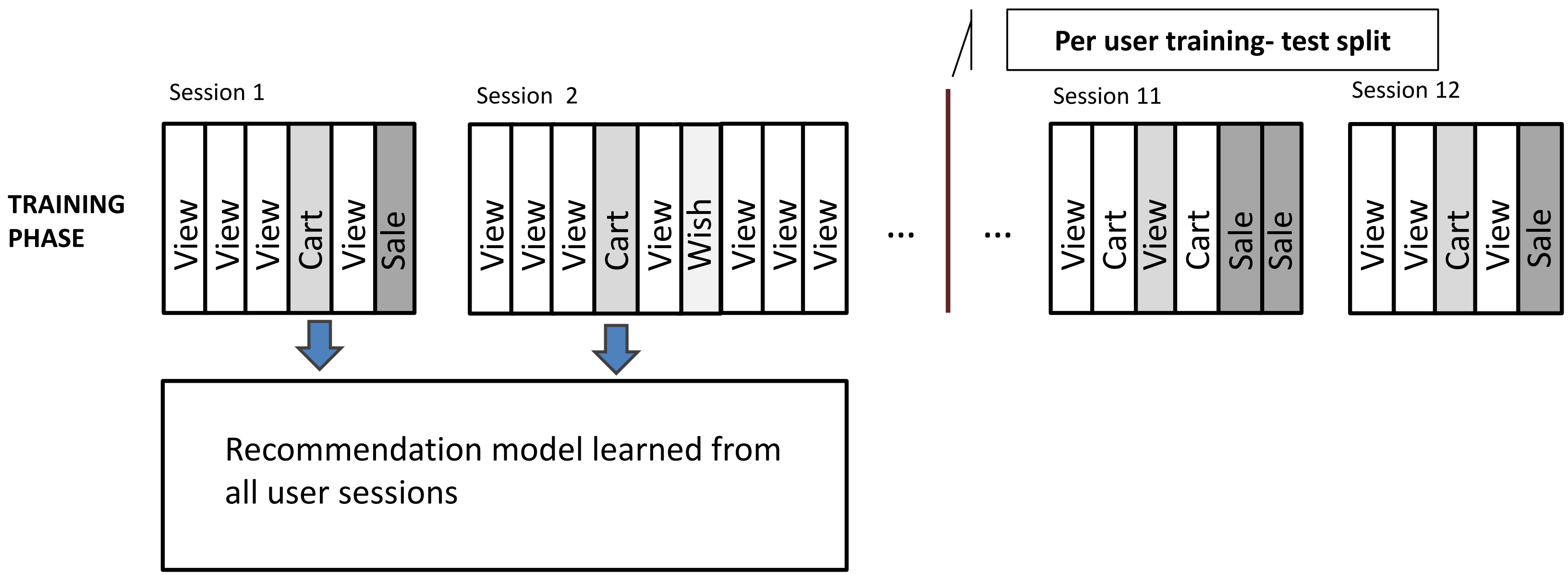
2) Group by session



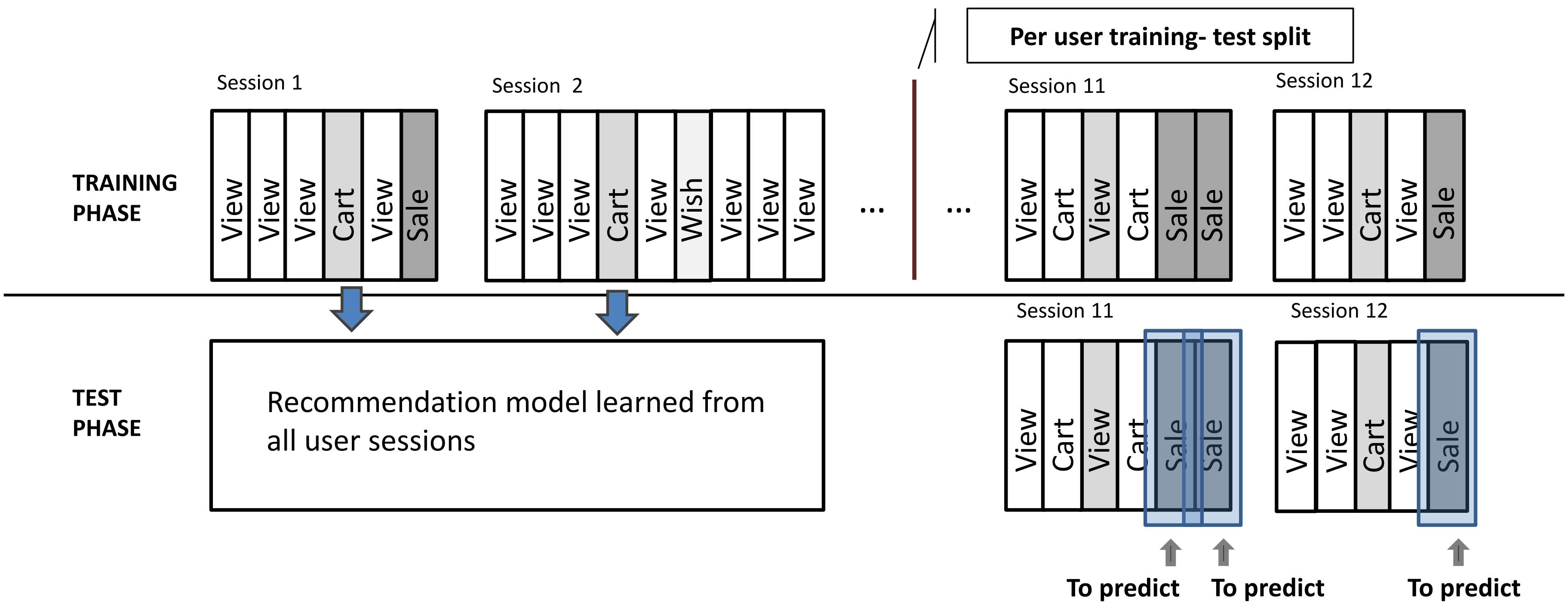
3) Split in training and test



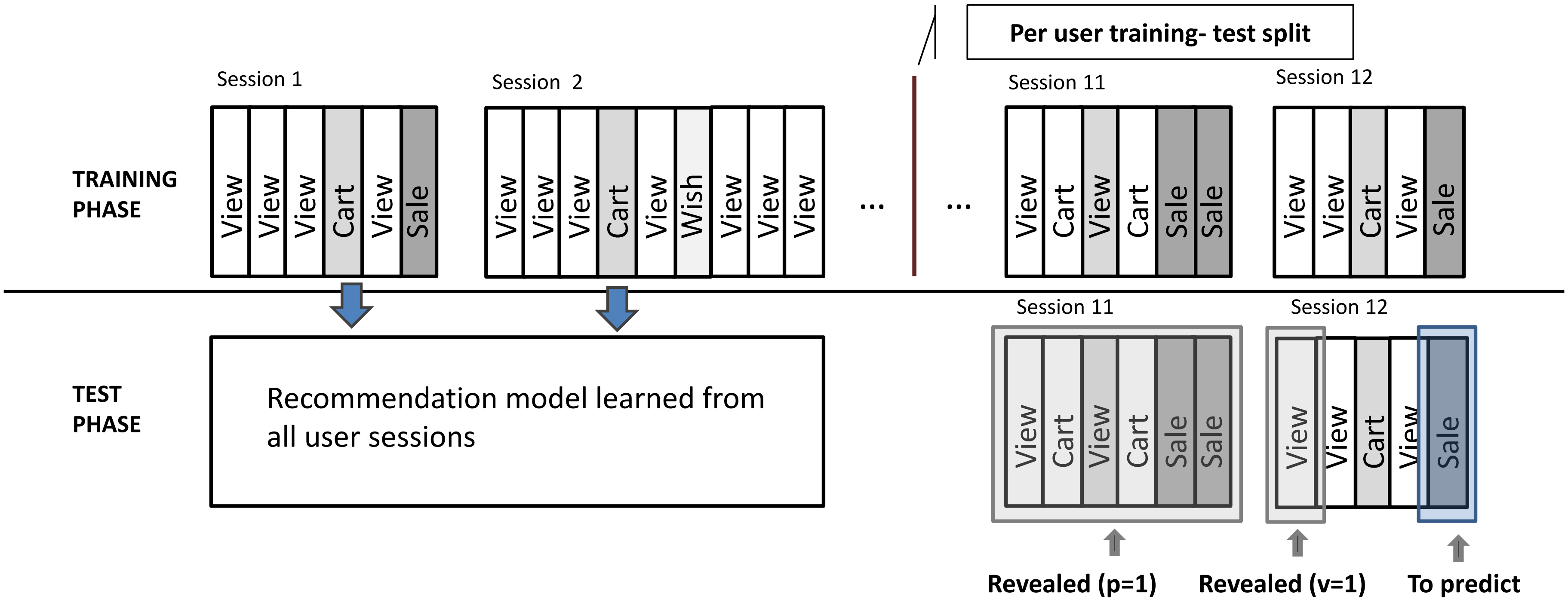
4) Train baseline recommender



5) Identify purchases to predict



6) Reveal short-term information



Experiments

- Benchmark of existing (standard) recommendation algorithms against hybrid algorithms that take the short-term interest into account.

- Result: Rather easy contextualization already leads to good performance improvements

Used data set

- Anonymized log data provided by Zalando
  - About 20 million logged actions
  - Naturally, many only-viewing users

	Sparse	Medium	Dense
Users	121,018	38,447	1,869
Items	40,526	19,061	2,208
Purchases	680,787	344,684	43,079
Views	9,807,282	3,929,813	117,734
Min. purchases/user	3	5	10
Min. purchases/item	3	5	10

Algorithms

Non-contextualized baseline strategies

- BPR

Learning-to-rank algorithm for implicit feedback with matrix factorization model

S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme;
BPR: Bayesian Personalized Ranking from Implicit Feedback; 2012

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN

Traditional collaborative filtering

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN
- PopRank

Unpersonalized, most viewed and purchased

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN
- PopRank
- Random

Lowest baseline, over all items

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN
- PopRank
- Random

Contextualization strategies

- CoOccur

Users who viewed A, also viewed B.

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN
- PopRank
- Random

Contextualization strategies

- CoOccur
- CoOccur-Filtering

As above, but ordered by baseline

Algorithms

Non-contextualized baseline strategies

- BPR
- Item-kNN
- PopRank
- Random

Contextualization strategies

- CoOccur
- CoOccur-Filtering
- FeatureMatching

Short-term content-based user profile with preferred brand and category

Algorithms

Non-contextualized baseline strategies

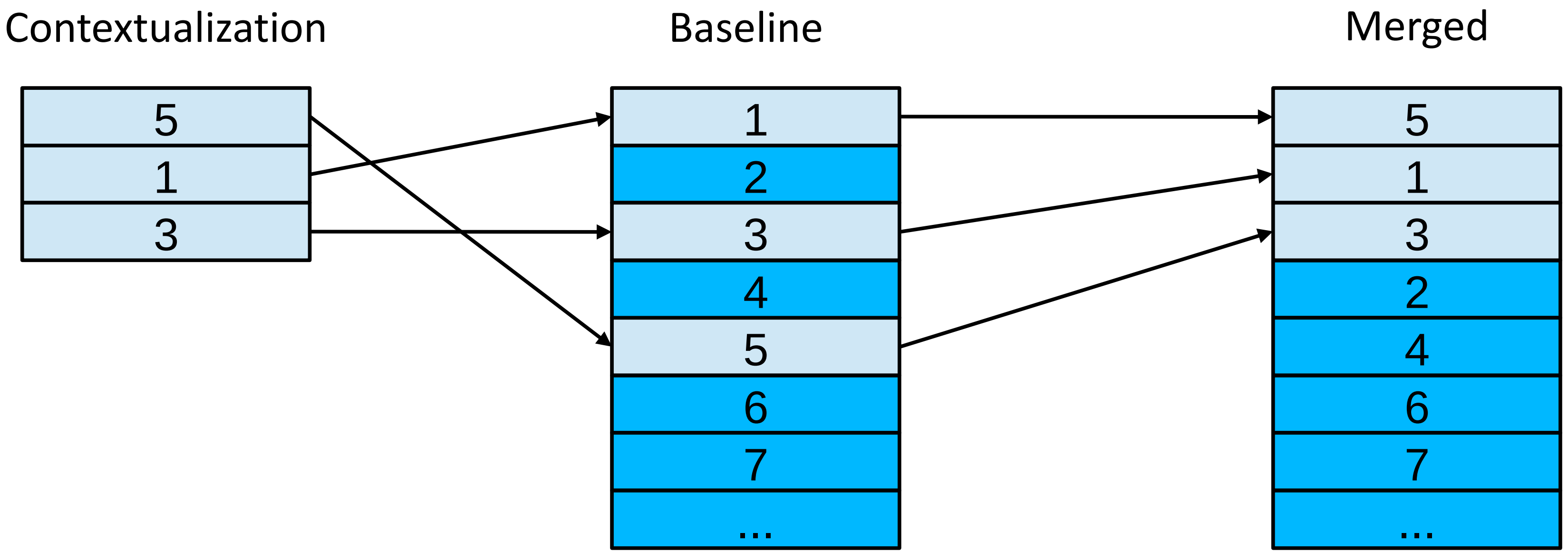
- BPR
- Item-kNN
- PopRank
- Random

Contextualization strategies

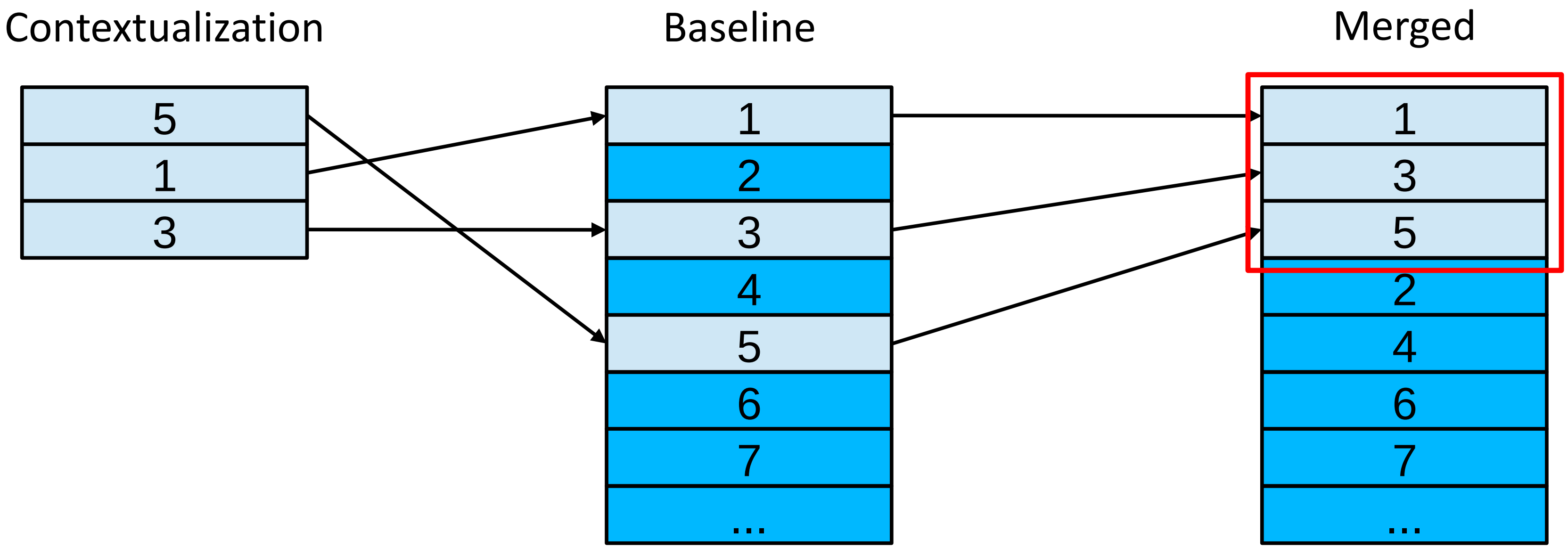
- CoOccur
- CoOccur-Filtering
- FeatureMatching
- RecentlyViewed

„Shopping history“

Hybridization



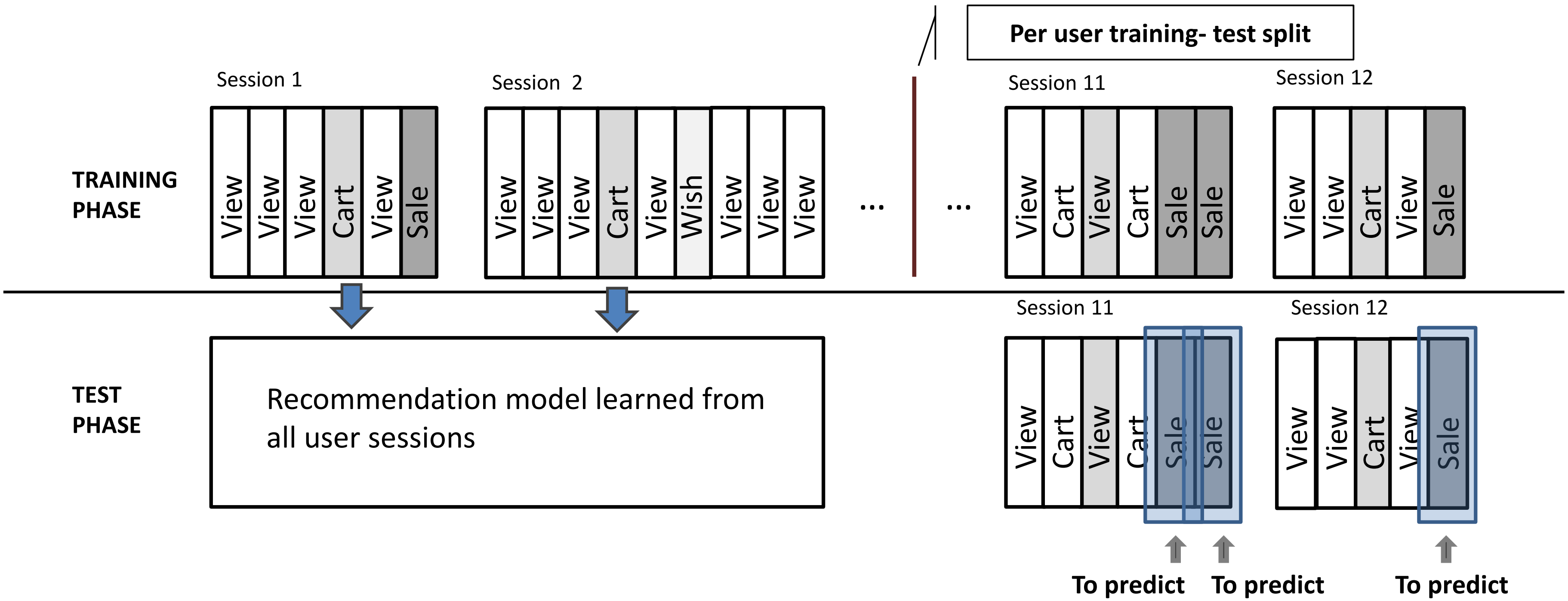
Hybridization



CoOccur-Filtering keeps baseline order

Evaluation metric

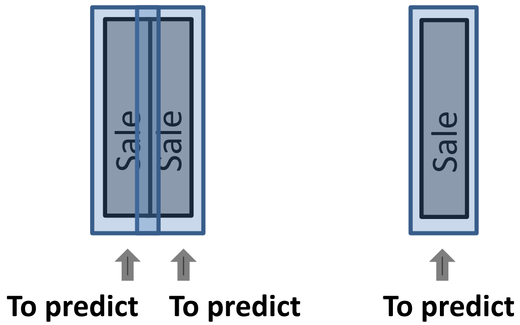
- Variant of the recall:



P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

Evaluation metric

- Variant of the recall:



P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

Evaluation metric

- Variant of the recall:

P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

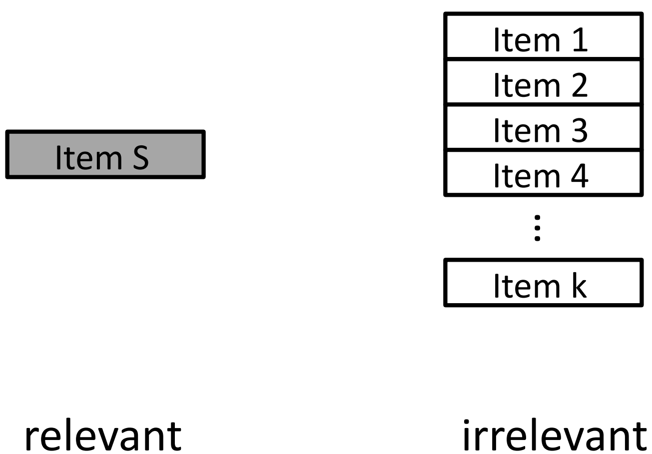
Evaluation metric

- Variant of the recall:

P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

Evaluation metric

- Variant of the recall:

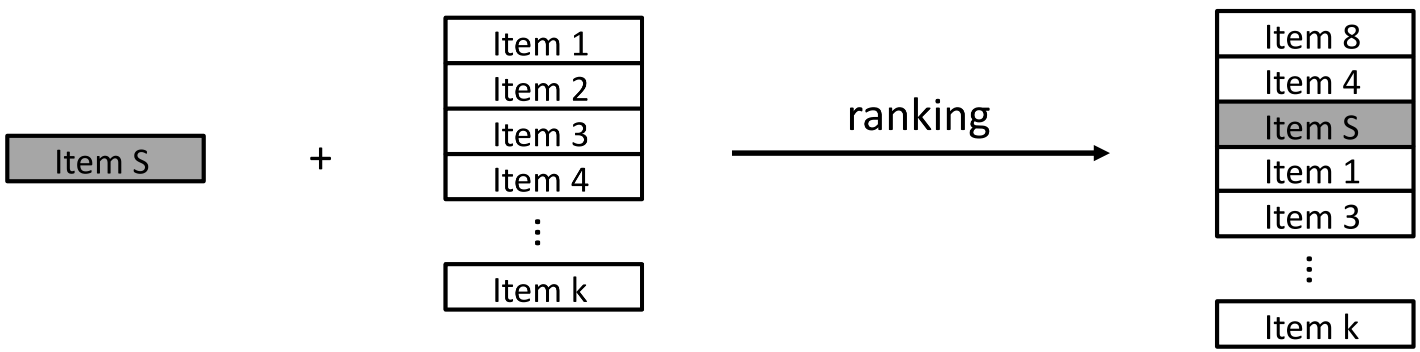


P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

k = 100

Evaluation metric

- Variant of the recall:

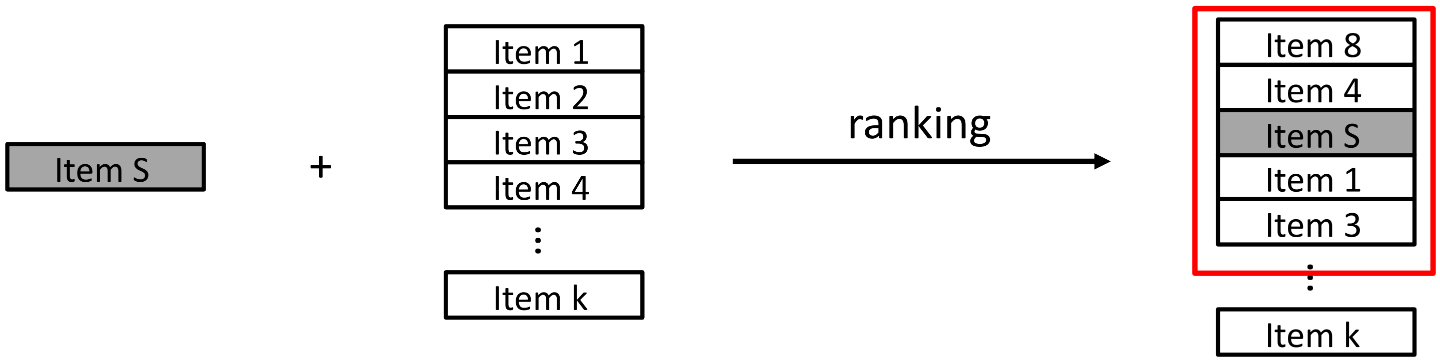


P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

k = 100

Evaluation metric

- Variant of the recall:



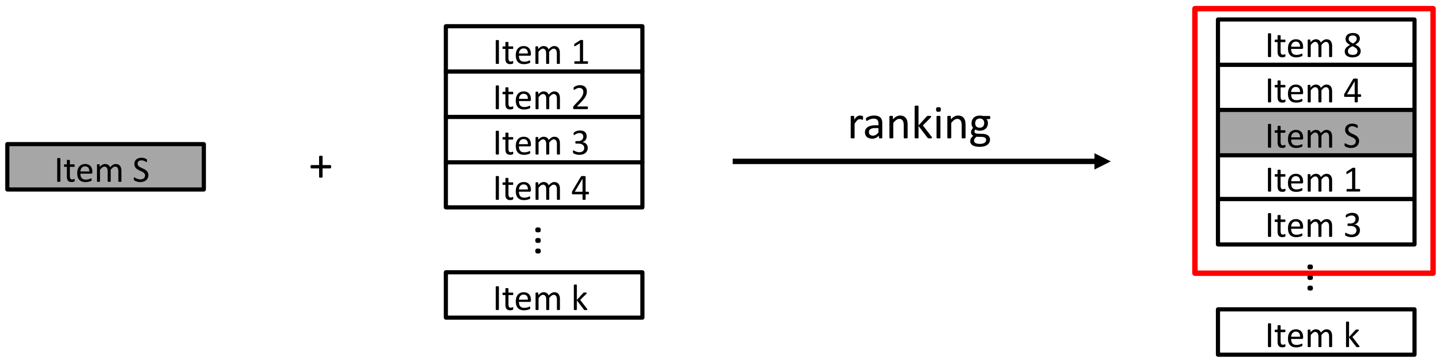
S in Top-N? → Hit!

P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

n=10 k = 100

Evaluation metric

- Variant of the recall:



- Variant of the precision: $\frac{1}{1+k} \cdot recall$

P. Cremonesi, Y. Koren, R. Turrin; Performance of Recommender Algorithms on Top-N Recommendation Tasks; 2010

n=10 k = 100

Results

Recall-variant
Dense subsample

	v=0, p=2	v=2, p=2	v=5, p=2	v=10, p=2	v=5, p=0
BPR			0.40		
PopRank			0.21		
ItemKNN			0.19		
Random			0.09		
CoOccur + ...					
BPR	0.38	0.47	0.49	0.52	0.48
PopRank	0.37	0.45	0.48	0.51	0.41
Random	0.31	0.44	0.48	0.50	0.40
CoOccur-Filter + ...					
BPR	0.39	0.47	0.48	0.50	0.48
PopRank	0.29	0.36	0.38	0.39	0.38
Random	0.23	0.34	0.35	0.37	0.36
FeatureMatching + ...					
BPR	0.41	0.65	0.71	0.76	0.71
PopRank	0.38	0.62	0.68	0.72	0.64
Random	0.31	0.60	0.66	0.73	0.61
RecentlyViewed + ...					
BPR	0.40	0.55	0.64	0.72	0.63
PopRank	0.36	0.53	0.62	0.71	0.52
Random	0.27	0.47	0.57	0.67	0.47
RecentlyViewed + FeatureMatching + ...					
BPR	**0.41**	**0.66**	**0.73**	**0.79**	**0.71**
PopRank	0.40	0.65	0.72	0.79	0.64
Random	0.32	0.63	0.70	0.78	0.62

Results

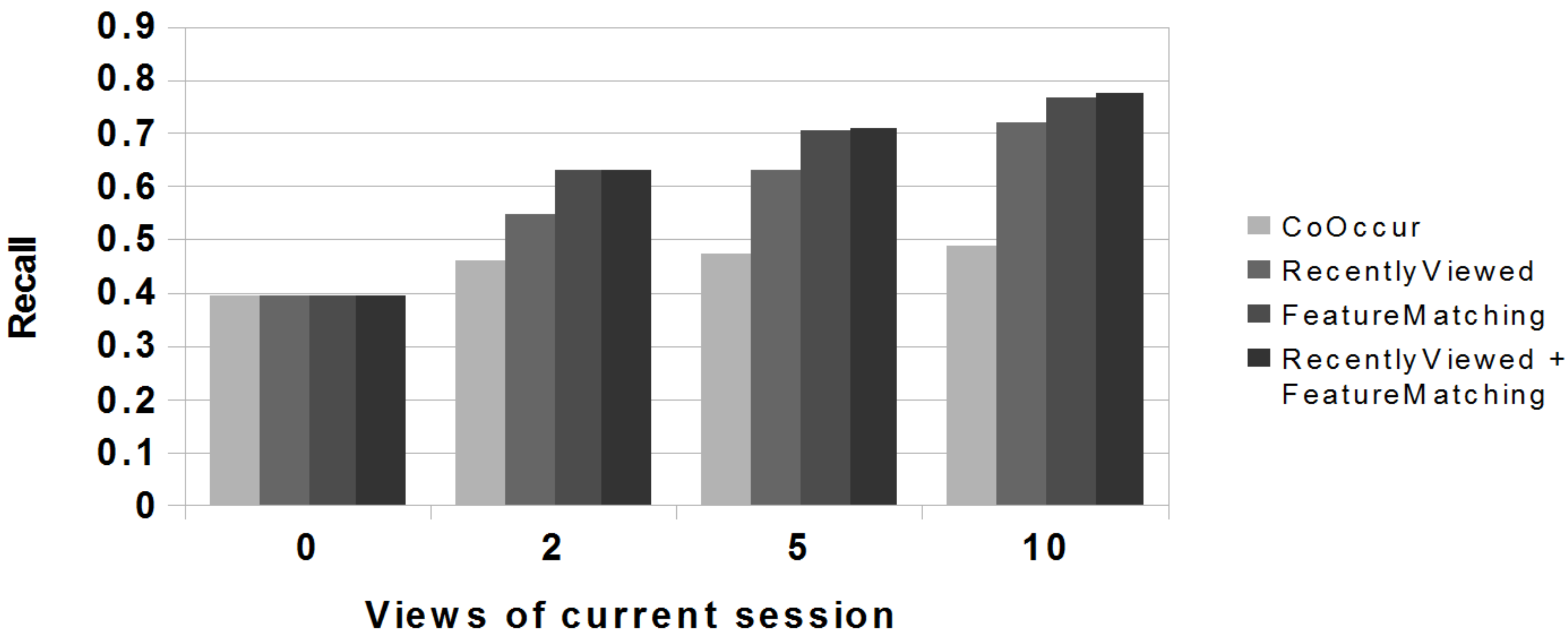
Recall-variant
Diffenent subsamples

	v=0, p=2	v=2, p=2	v=5, p=2	v=10, p=2	v=5, p=0
Dense: RecentlyViewed + FeatureMatching + ...					
BPR	**0.29**	**0.53**	**0.61**	**0.69**	**0.58**
PopRank	0.28	0.51	0.61	0.68	0.52
Random	0.25	0.50	0.60	0.69	0.51
Medium: RecentlyViewed + FeatureMatching + ...					
BPR	**0.53**	**0.69**	**0.73**	**0.78**	**0.72**
PopRank	0.51	0.68	0.72	0.77	0.65
Random	0.36	0.59	0.65	0.72	0.55
Sparse: RecentlyViewed + FeatureMatching + ...					
BPR	**0.59**	**0.75**	**0.79**	**0.83**	**0.77**
PopRank	0.58	0.73	0.78	0.82	0.72
Random	0.36	0.63	0.69	0.76	0.60

Also measured: MRR

Other dataset: Tmall (Alibaba Group, CN)

Influence of short-term actions



BPR, p = 2

Discussion / open problems

- Open challenges: How should specific visitor action be interpreted? Recommending items seen in the past?

- Limitations through offline evaluation
  - To which extent is prediction/recall a good metric?
  - What about obviousness?
  - User study/online experiment

- Domain dependant

Summary

- Taking short-term interests into account can improve the recommendation
- More realistic protocol for the aspect of short-term interest
- Even simple approaches perform well
- Domain specifics (brand-loyalty, categories) can be important