Temporal-Contextual Recommendation in Real-Time

Yifei Ma* Balakrishnan (Murali) Narayanaswamy*

Haibin Lin Hao Ding

yifeim@amazon.com

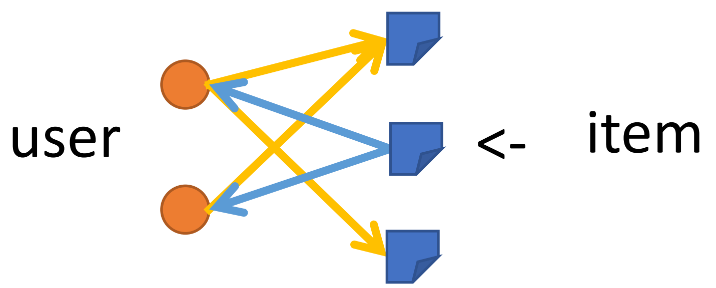
*Equal contribution authors

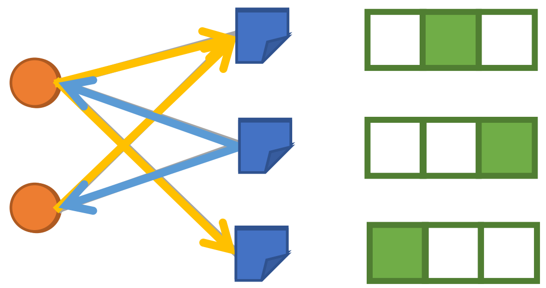
Objective

- Real-time recommendation without ML/DL experience
- One network architecture to fit most application scenarios

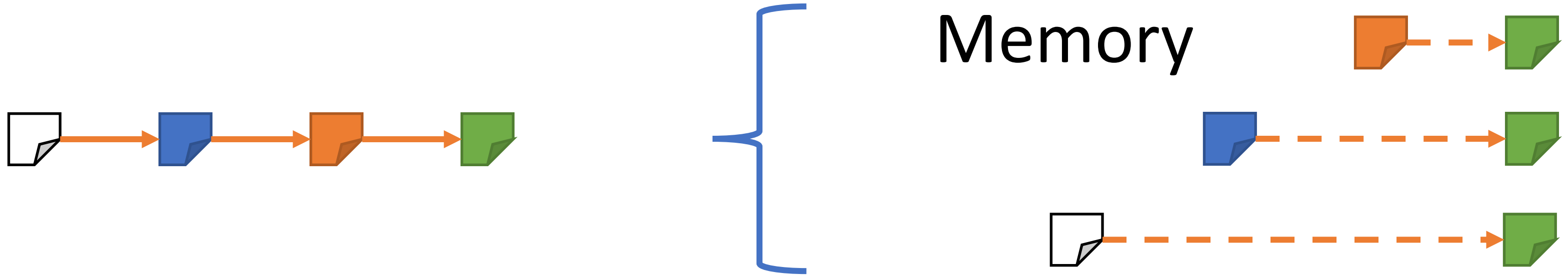
Background

- Most rec sys are built from user/item latent representations





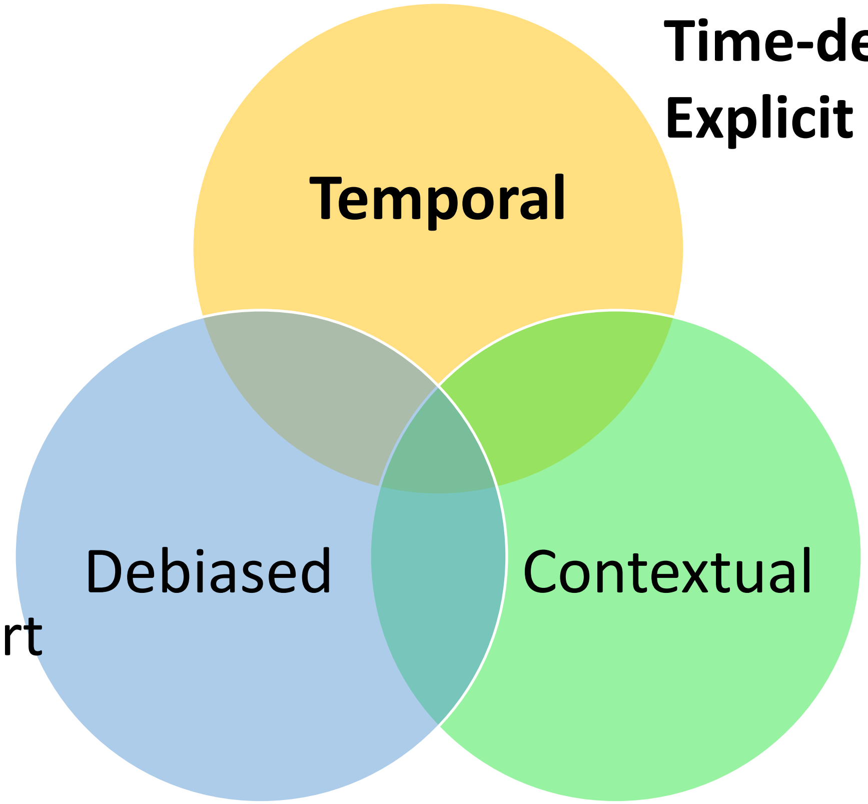
- Sequence models with ordered user histories



- However, there are still gaps in practice
  - Metadata
  - Temporal drifts

Our Contributions

Time-delta changes
Explicit contexts

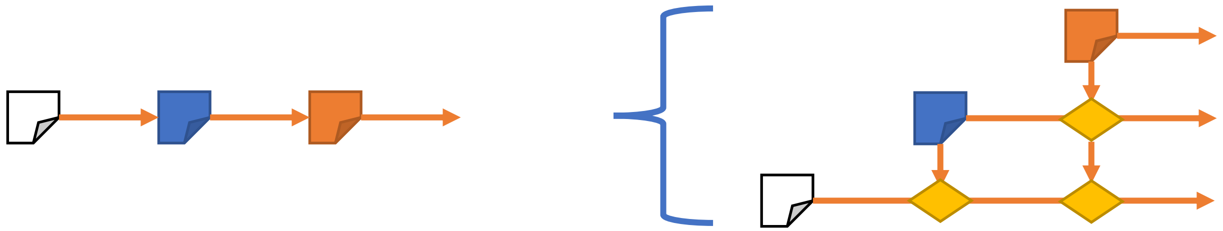


Continuous item cold-start
Negative sampling

Recommend by
item meta-data

Sequence Model and Context Changes

- Exponential Moving Average (EMA) – smooth changes in context



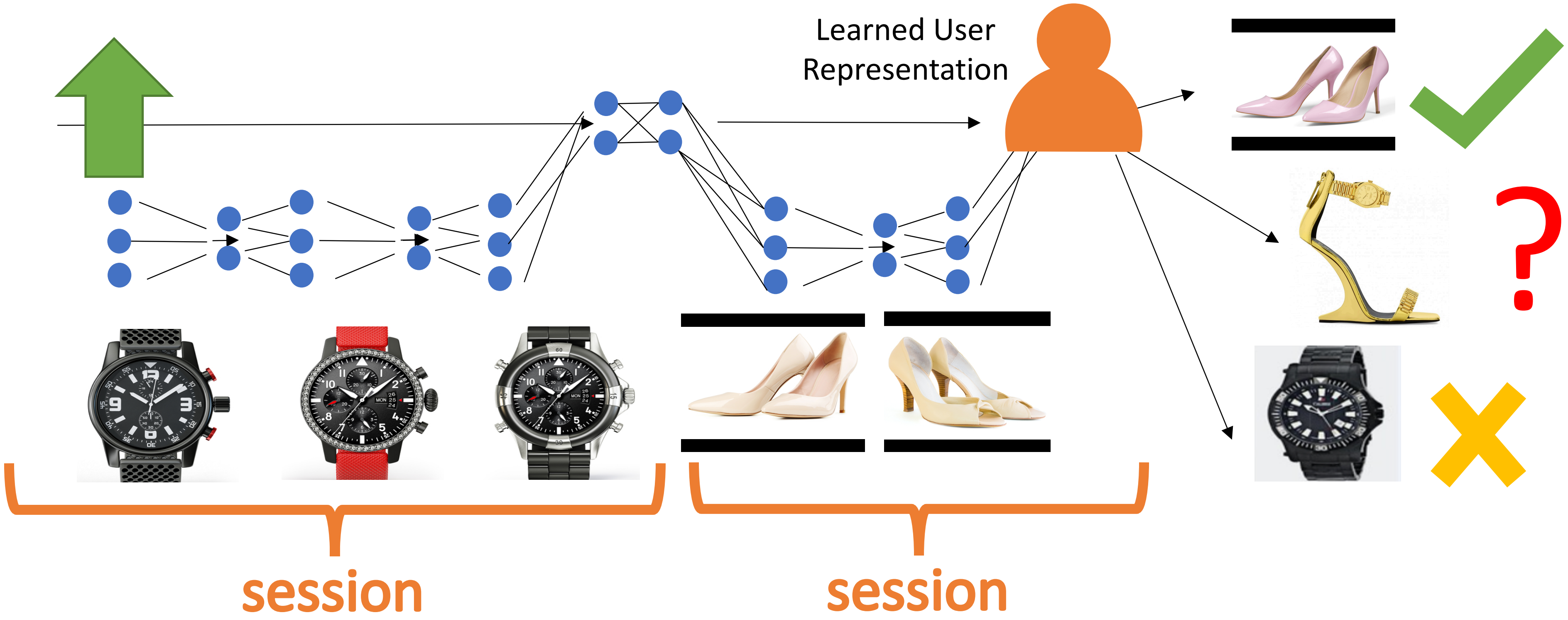
- Gating Recurrent Units (GRUs) – implicit context changes

Marry has a little lamb

Subject Verb Object

  - Back Propagation Through Time (BPTT) – active memory of ~15 items

Explicit Context Changes



👍Skip connection achieves long-term memory effects [Quadrana et al., 2017]

👎Slower in minibatch training due to irregular computation flows

Remap time-delta as RNN input?

Efficient implementation inspired from https://gluon-nlp.mxnet.io/

Recommendation Changes with Time-Delta

- Δt between last visit and next recommendation
- As Δt increases, items become more general

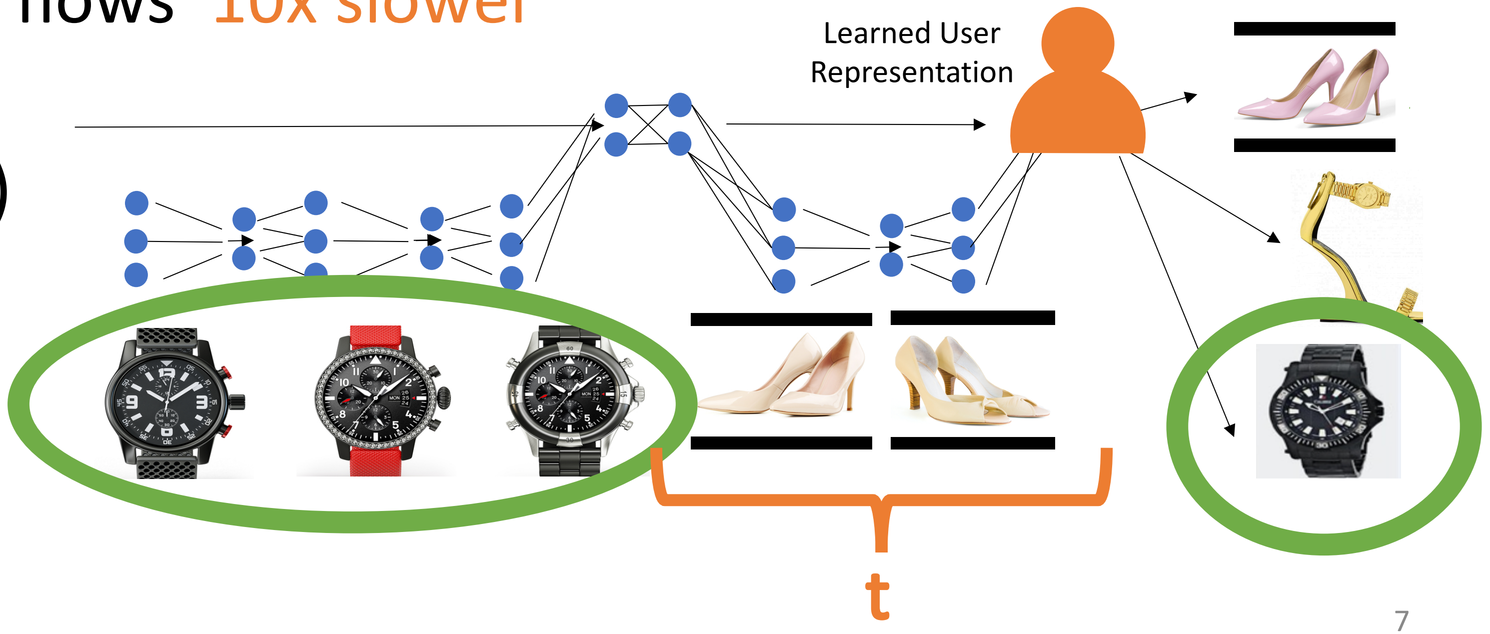
Δt	title	genres	popularity
0	Purple Rose of Cairo	Comedy|Drama	0.000236
60	Unbearable Lightness	Drama	0.000209
3600	Local Hero	Comedy	0.000195
86400	Big	Comedy|Drama	0.001130

F. Maxwell Harper and Joseph A. Konstan. 2015. The **MovieLens** Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages. http://dx.doi.org/10.1145/2827872

Synthetic Memory Benchmark

Reoccurring purchase after a noisy session; memory capacity = t

- RNN: not sensitive to explicit context changes
- HGRU [Quadrana et al.]: 50<t<60
- Irregular computation flows 10x slower
- HRNN [ours]: 10<t<20
(Meet most use-cases)



Speed and Performance Trade-Offs

- Subreddit-interactions prediction (18271 users, 27453 subreddits)
- HRNN: GRU + time-delta (ours)
- HGRU: GRU + time-delta + hierarchy
- HRNN (ours) meets most use cases

+0.16

+0.05 10x slower

Hit@5 on reddit recommendation

	GRU	HRNN	HGRU	Pop	BPR[5]
50 hidden	0.26	0.42	0.47	0.11	0.39

https://www.kaggle.com/colemaclean/subreddit-interactions

Speed and Performance Trade-Offs

- Subreddit-interactions prediction (18271 users, 27453 subreddits)
- HRNN: GRU + time-delta (ours) +0.16
- HGRU: GRU + time-delta + hierarchy +0.05 10x slower
- HRNN (ours) meets most use cases

Hit@5 on reddit recommendation

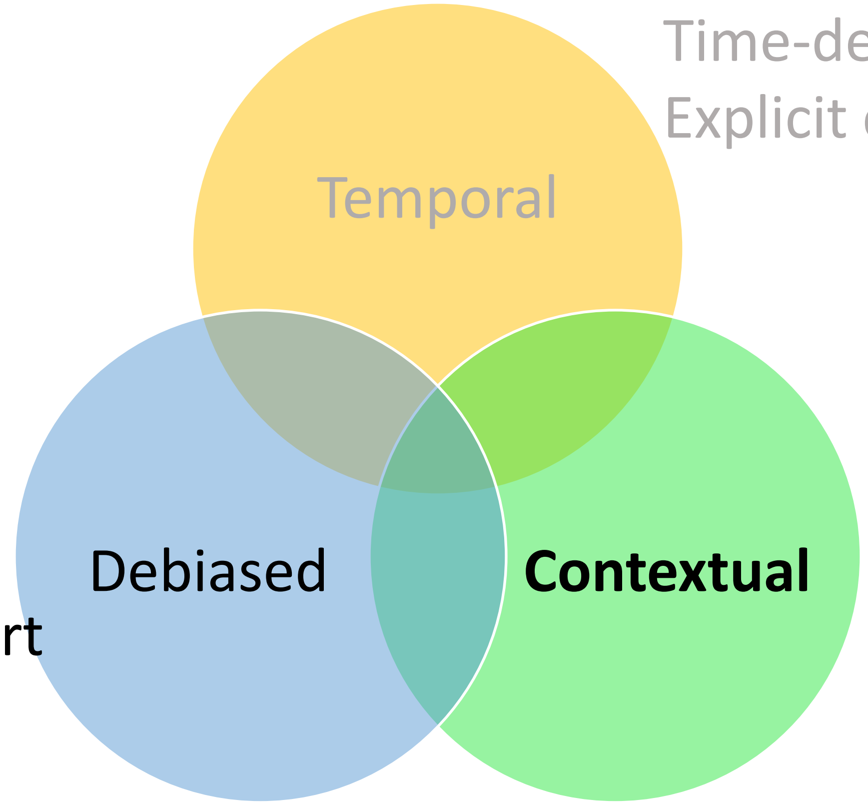
	GRU	HRNN	HGRU	Pop	BPR[5]
50 hidden	0.26	0.42	0.47	0.11	0.39
200 hidden	0.55	0.55	0.55		

+0.29

- Caveat:
  - Largest benefit comes from model tuning
  - Public datasets have limited complexity

Contributions

Time-delta changes
Explicit contexts



Continuous item cold-start
Negative sampling

Recommend by item meta-data

Recommend by Item Meta-Data

Factorization machines [Rendle 2011]: a game of point collection

Item	Rain	No rain	Wind	No wind	Scores
Umbrella (w_1)	1	0	0	1	1
Jacket (w_2)	1	0	1	0	0
Nothing (w_3)	0	1	1	1	2
Context	**0**	**1**	**0**	**1**	

Meta score = [context meta-data] * [item meta-data]

HRNN score = [user embedding] * [item embedding]

A hybrid model adapts to both types of data sources!

Recommend New Movies in a Hold-Out Set

- Remove a random half of unique movies from training
- Expect similar genres top rank top within the hold-out set

Genres in user history

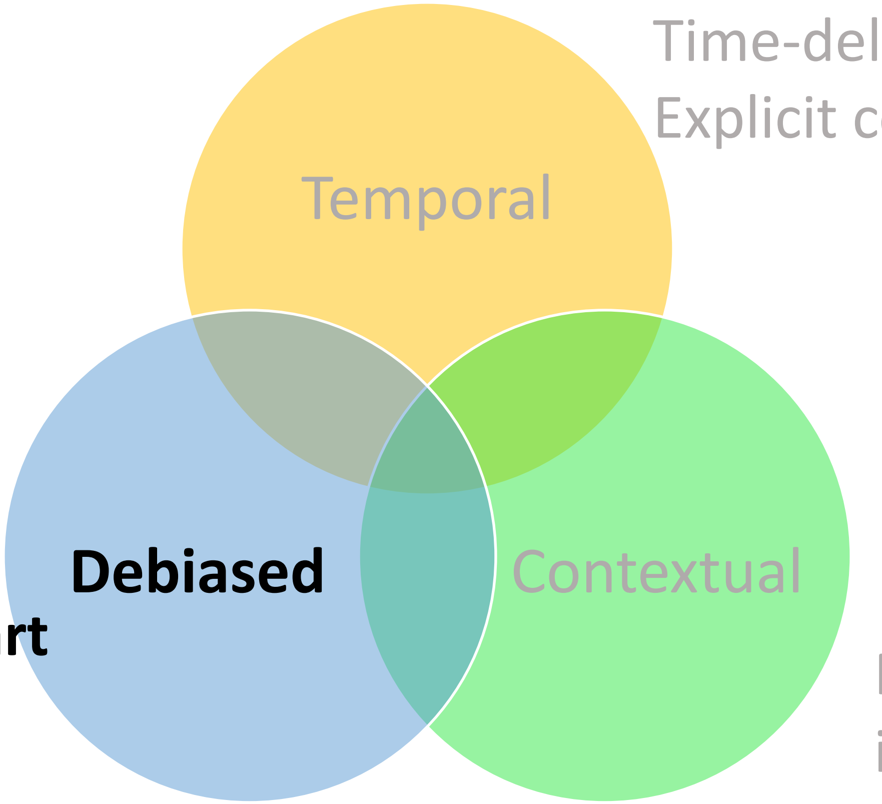
Braveheart (1995)	Action|Drama|War
Farinelli: il castrato (1994)	Drama|Musical
Edward Scissorhands (1990)	Drama|Fantasy|Romance
Back to the Future (1985)	Adventure|Comedy|Sci-Fi
Road Warrior, The (Mad Max 2) (1981)	Action|Adventure|Sci-Fi
Grumpy Old Men (1993)	Comedy
Halloween: The Curse of Michael Myers (Hallowe...	Horror|Thriller

Genres in top-ranked hold-out movies

Making Contact (a.k.a. Joey) (1985)	Fantasy|Horror|Sci-Fi
Decoys (2004)	Comedy|Horror|Sci-Fi
Critters 4 (1991)	Comedy|Horror|Sci-Fi
TerrorVision (1986)	Comedy|Horror|Sci-Fi
Tremors II: Aftershocks (1996)	Comedy|Horror|Sci-Fi
Evil Aliens (2005)	Comedy|Horror|Sci-Fi
Thing with Two Heads, The (1972)	Comedy|Horror|Sci-Fi

- Limitations?
  - When warm items present, underscore cold items due to low confidence

Contributions



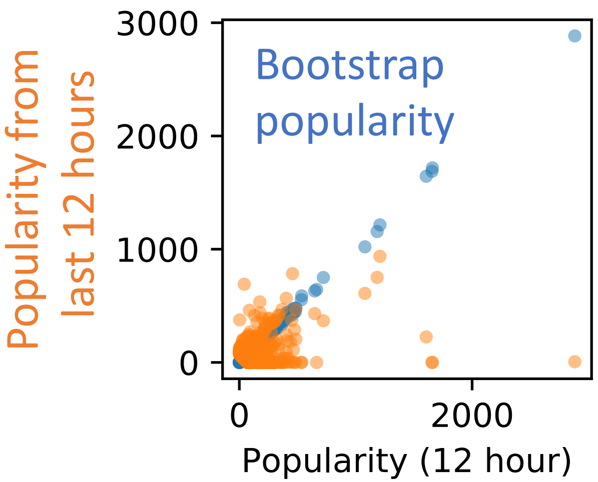
Time-delta changes
Explicit contexts

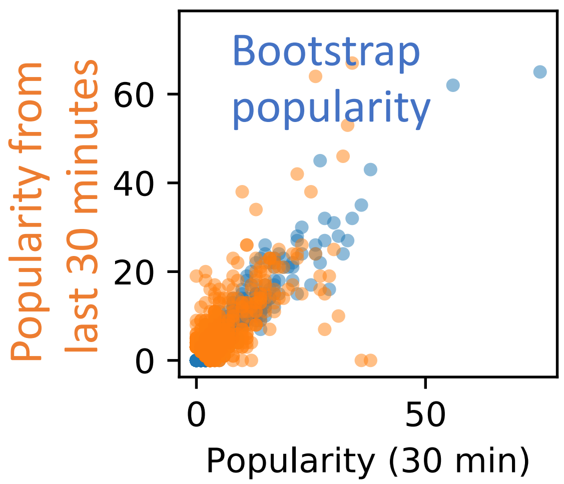
Continuous item cold-start
Negative sampling

Recommend by
item meta-data

Temporal Drifts in Item Popularity

- In news or media domains, new items easily drive >50% of traffic;
- Count global item popularity in Time Intervals (T)
  - T = 12 hours, 30 minutes, etc.
  - Test with previous T
  - Null = bootstrap from same T



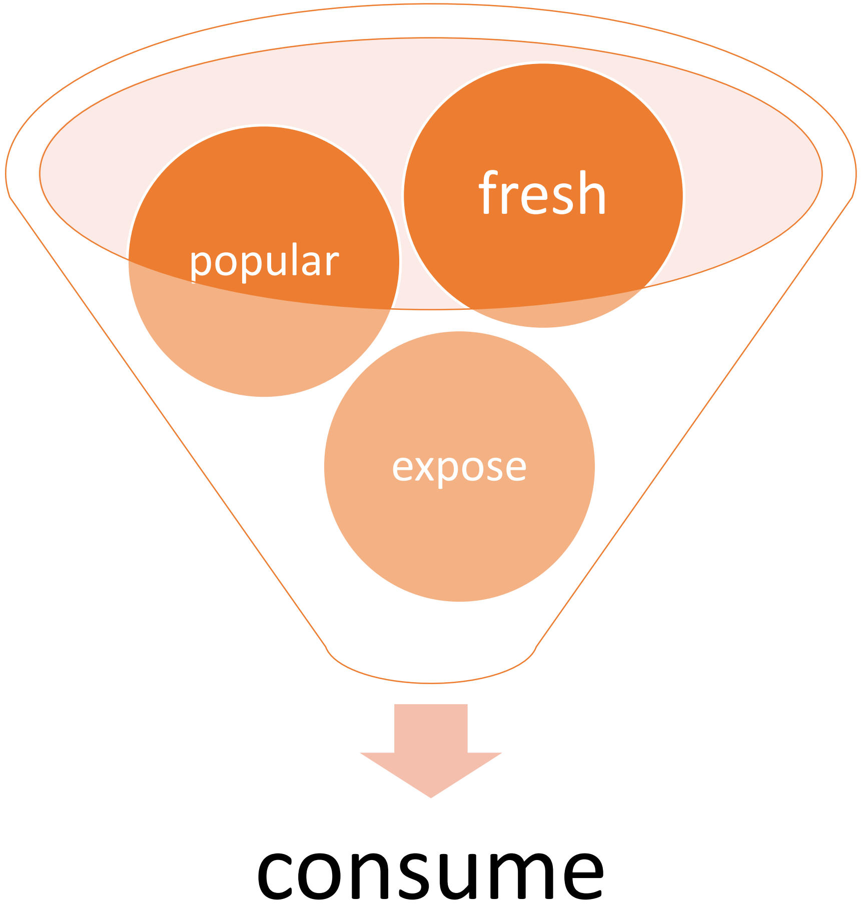


- How to use the stale information?

Stale information after T = 12 hours

No significant drifts in T = 30 minutes

Item Trend Debiasing



We learn to attribute between:

P(exposure | freshness, popularity)

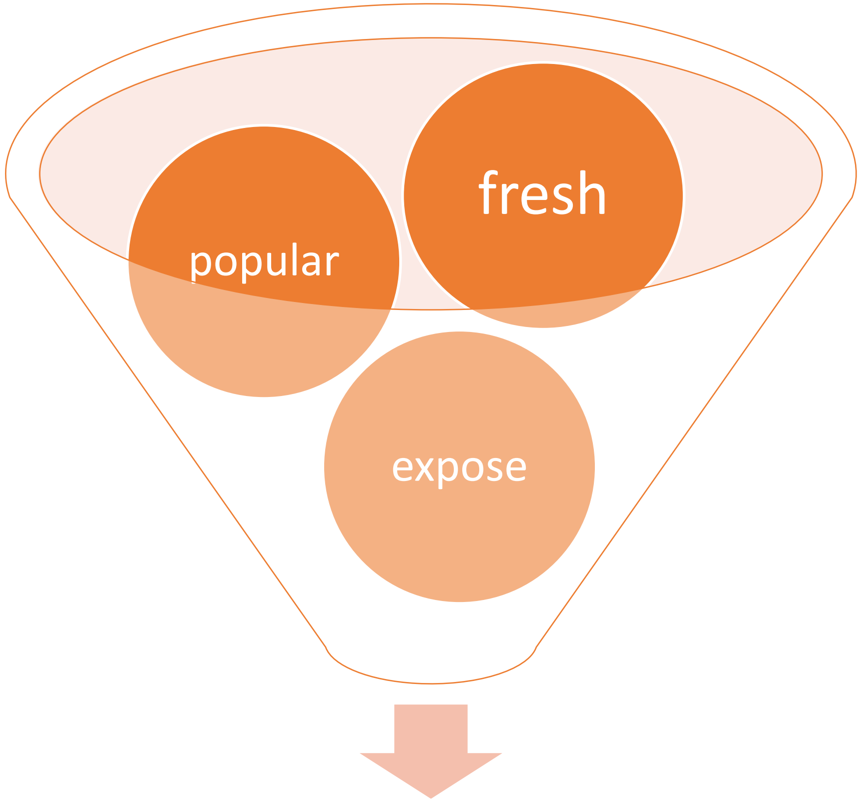
P(consumption | exposure, user preference)

Related to Inverse Propensity Scoring

More theoretical insights in [Ma et al., 2019]

Model inspired by [Wu et al., 2017]

Negative Sampling



consume

10x speed-up; 2x better performance in unit time

Vanilla RNN softmax on every item

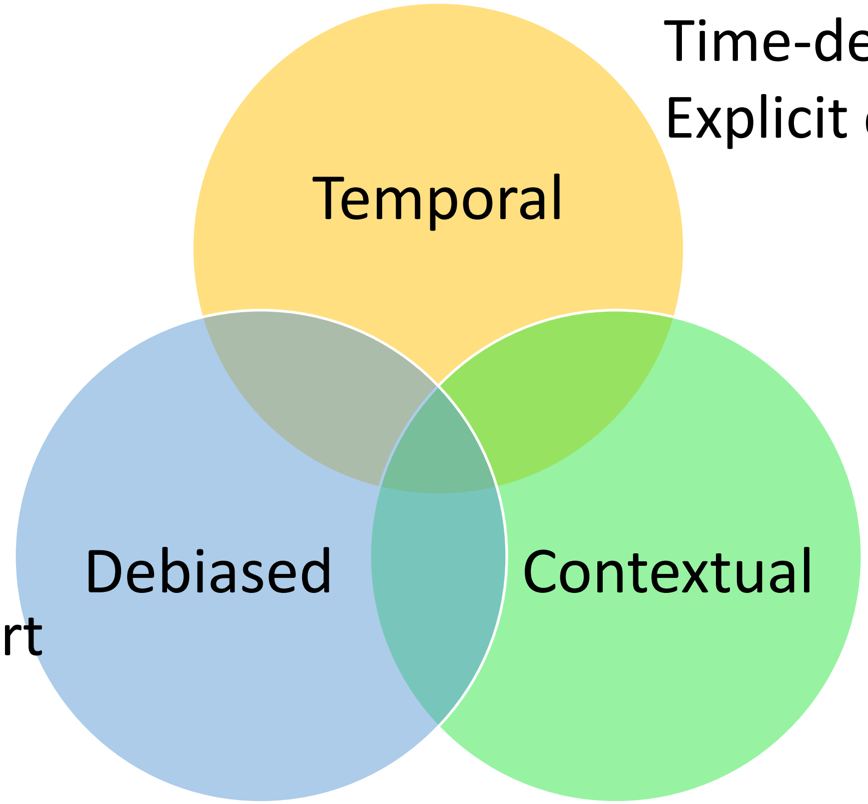
sample from popular items + debiasing

	Output	ml-1m	ml-10m	ml-20m	Taobao
Output size (m)	IS	62	255	362	1087
	Dense	1683	65 134	131 263	1 183 451
Throughput (#items/sec)	IS	23k	20k	**20k**	**7.8k**
	Dense	23k	23k	17k	631
PPL	IS	377	405	455	**17.6k**
	Dense	409	439	494	119k
NDCG	IS	0.128	0.123	0.12	**0.15**
	Dense	0.123	0.119	0.115	0.08

https://tianchi.aliyun.com/dataset/dataDetail?dataId=649

Bring Everything Together

Time-delta changes
Explicit contexts



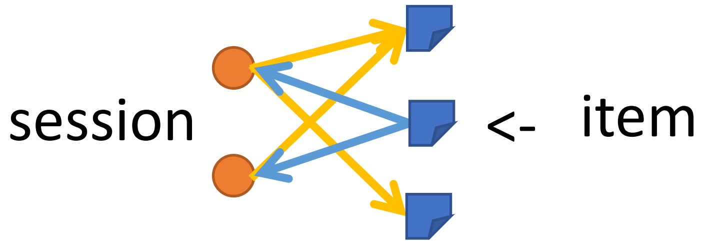
Continuous item cold-start
Negative sampling

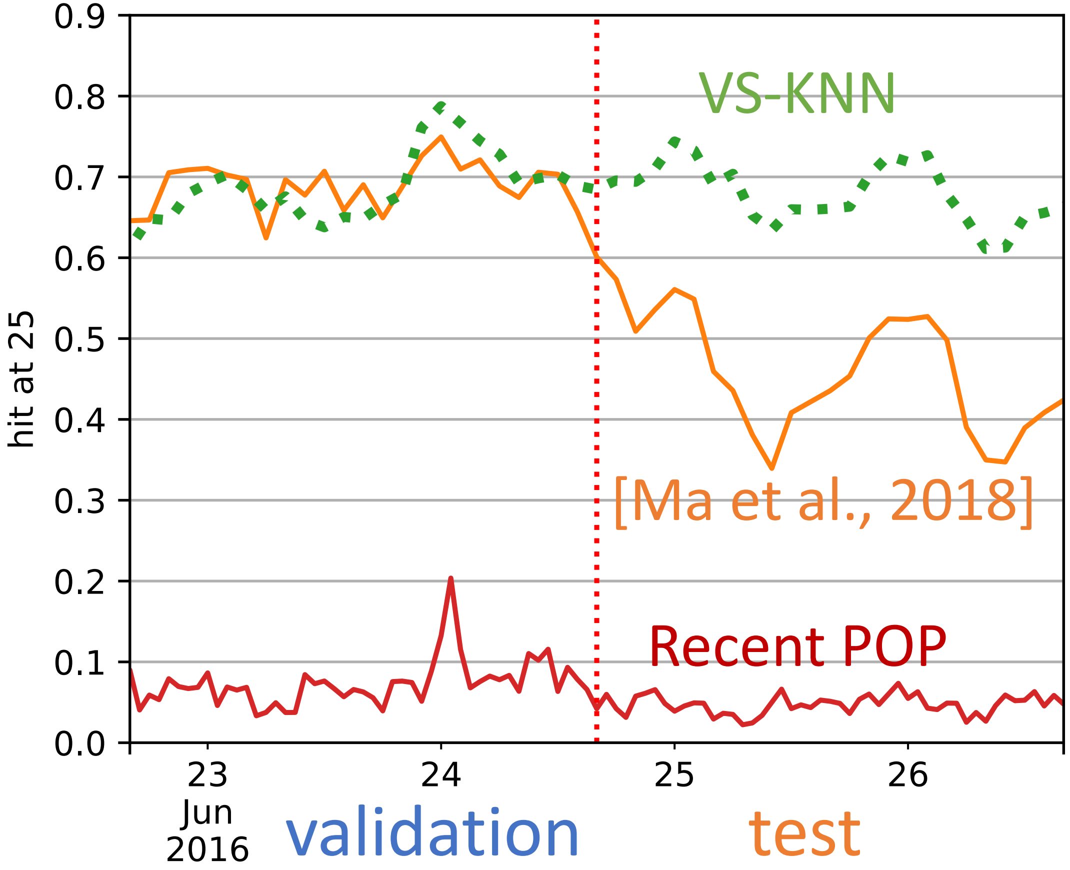
Recommend by
item meta-data

Case Study: Outbrain News Recommendation

14 days of news articles; Online VS-KNN > Offline GRU

[Ludewig & Jannach, 2018] https://github.com/mjugo/StreamingRec





Reproduce on larger set of items

- Sample 1% of users
- Filter user_activities >= 10
- Filter item_views >= 100

Retain 4M views from

- 377k users & 31k items

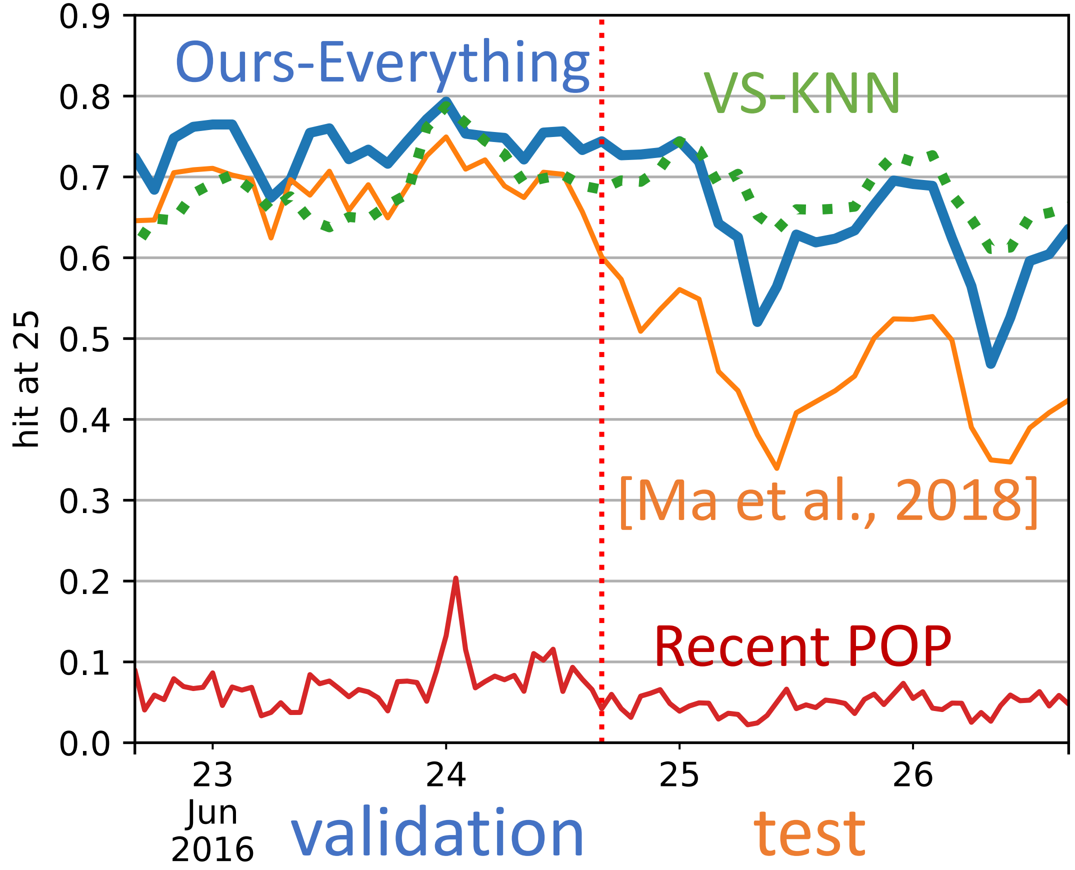
https://www.kaggle.com/c/outbrain-click-prediction

Case Study: Outbrain News Recommendation

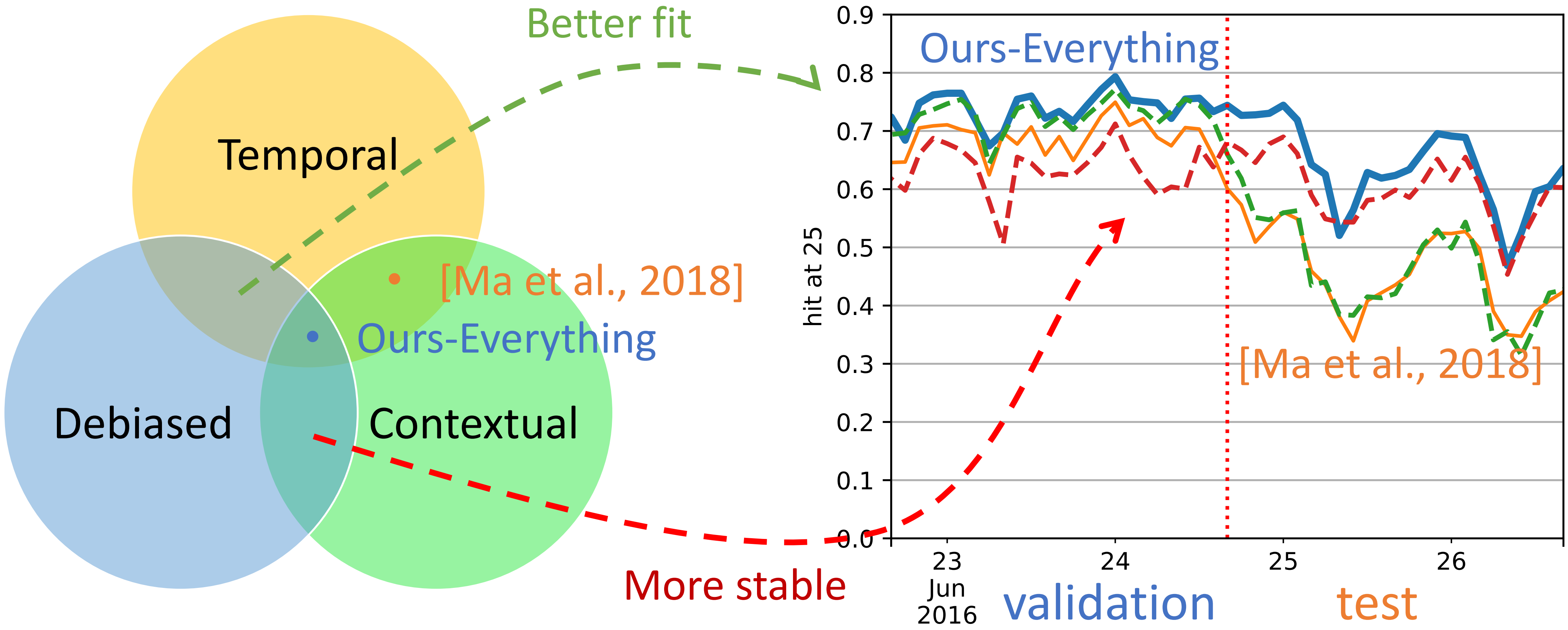
Item trend debiasing

- Item freshness since release
- Total views in last hour

Improves validation & test with long histories & futures!

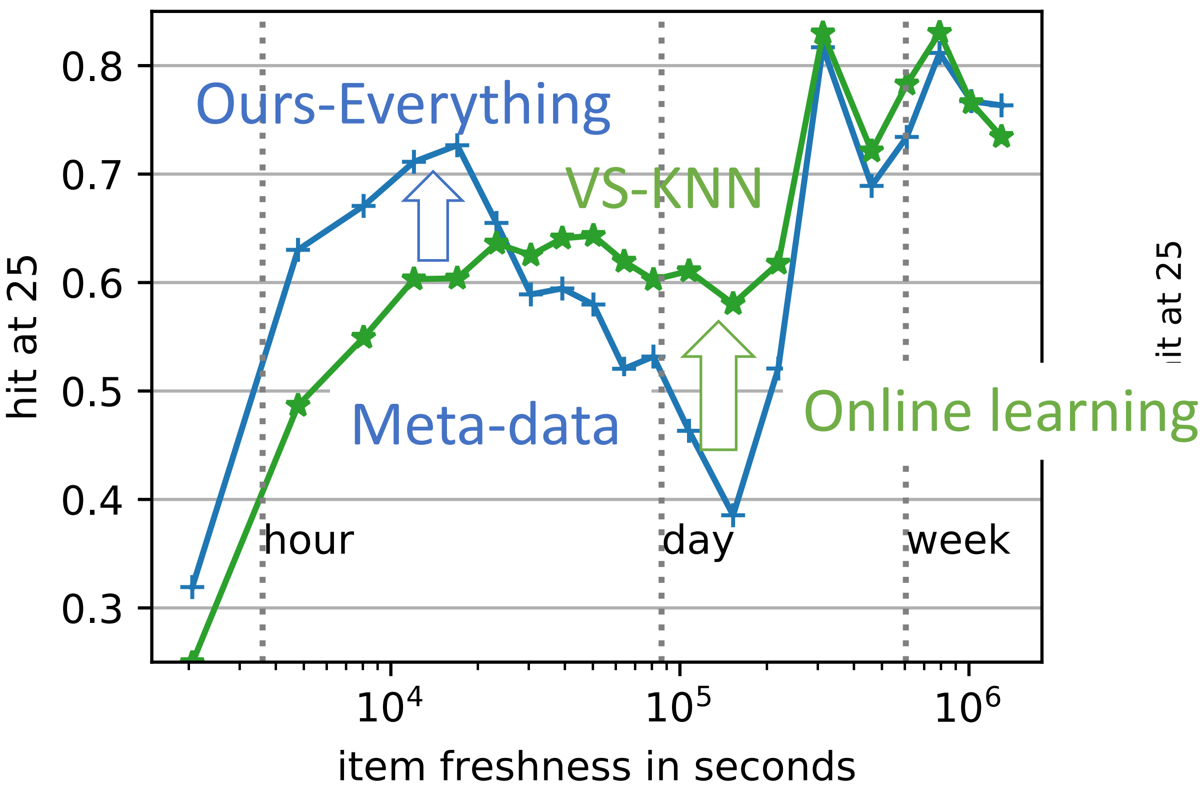


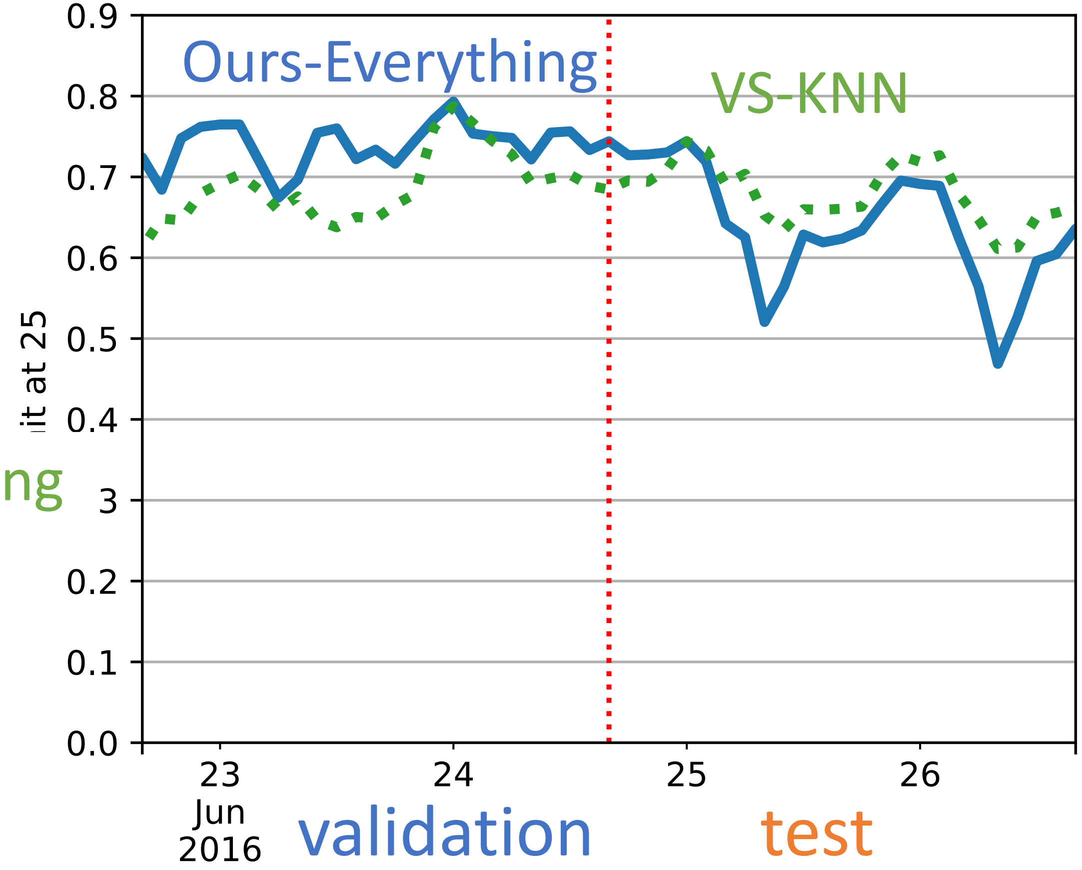
Case Study: Outbrain News Recommendation



Case Study: Outbrain News Recommendation

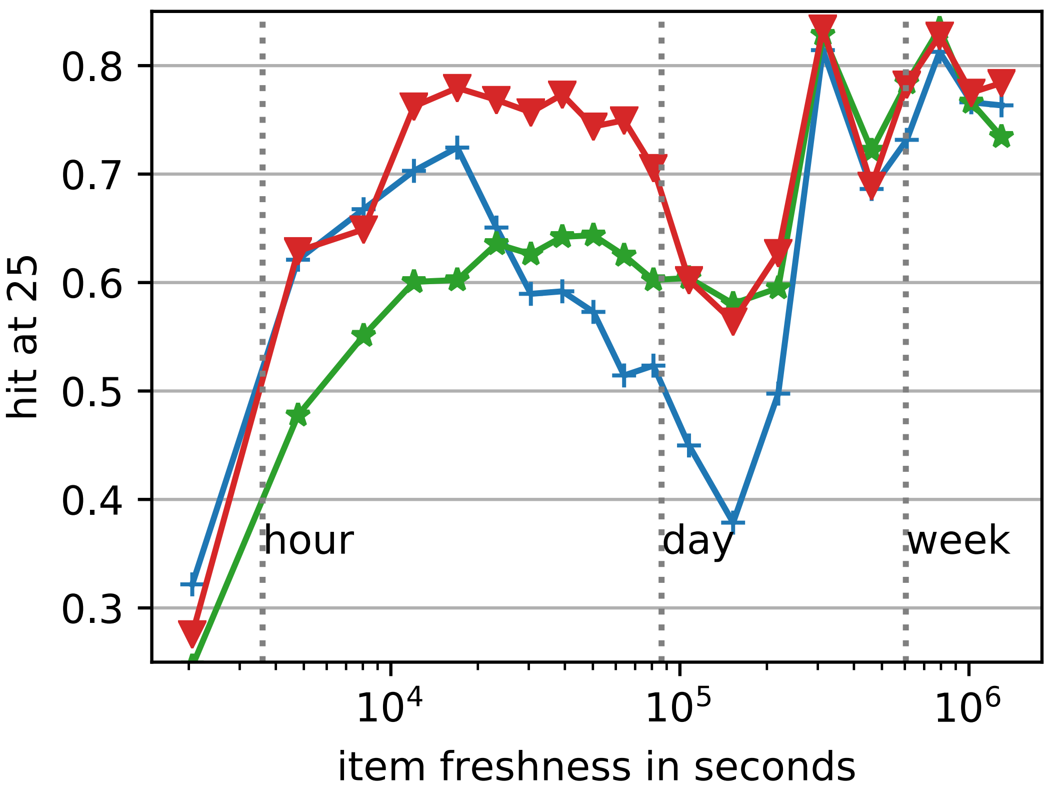
VS-KNN depends on good cold-start item recs in the first place

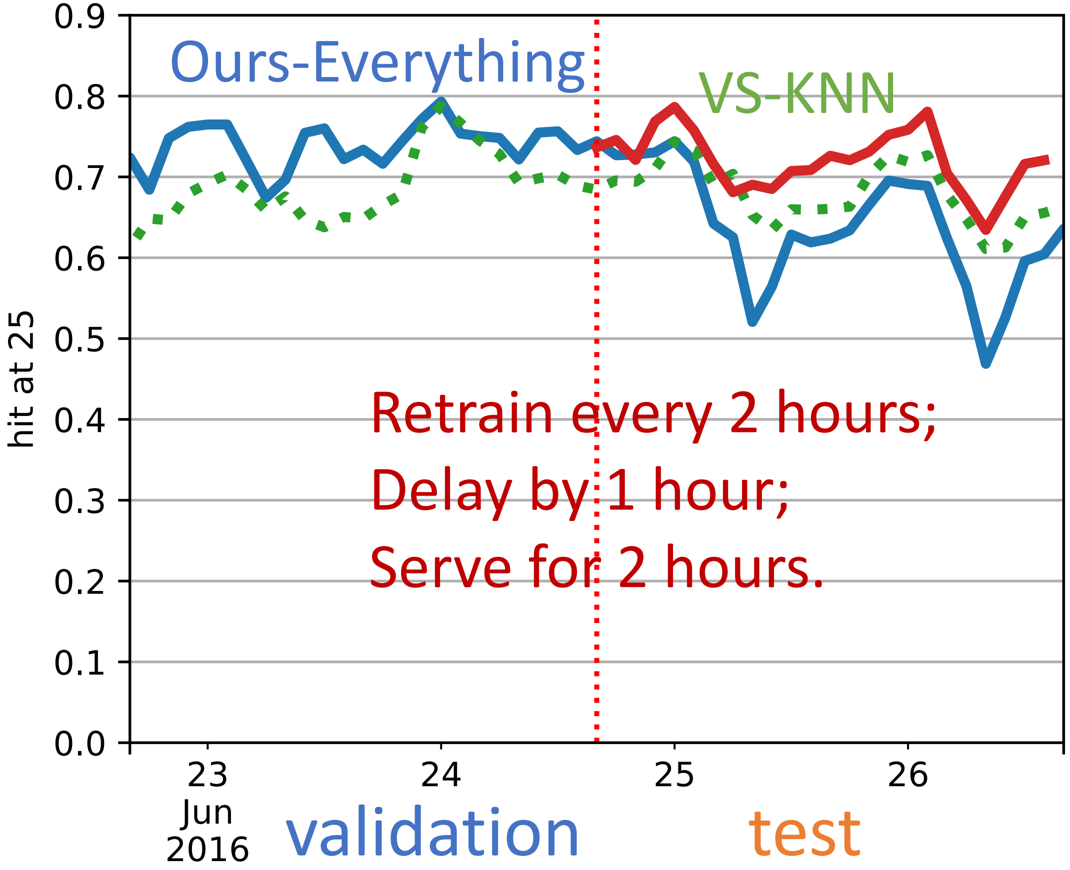




Case Study: Outbrain News Recommendation

Retrained Ours-Everything seems to bridge the gaps.





Conclusion

- Fill the practicality gap of session-aware RNN models
- Use temporal, contextual, side information
- Address time-varying confounding variables
- More theoretical details in paper
- Adopted by Amazon Personalize

https://github.com/aws-samples/amazon-personalize-samples/