Contextual RNNs for Recommendation

Elena Smirnova and Flavian Vasile

DLRS 2017

Agenda

1. Setup
2. Motivation
3. **Recurrent** Neural Networks
4. **Contextual Recurrent** Neural Networks
5. Experimental evaluation

Setup

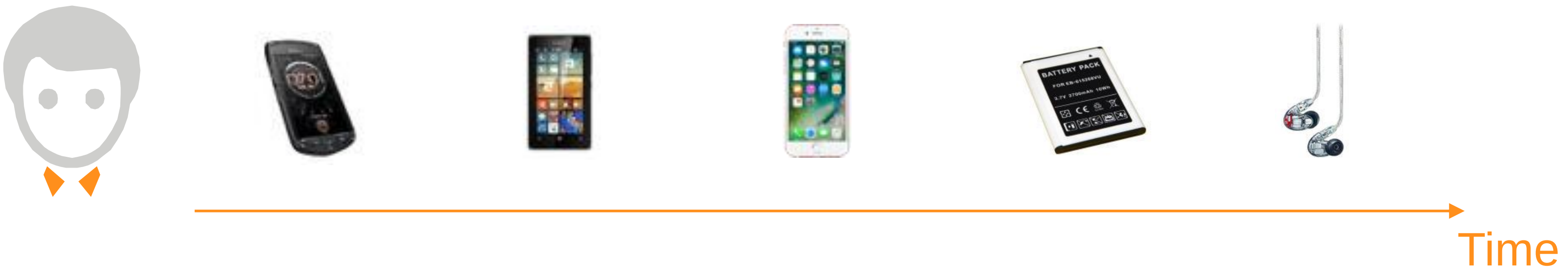
Sequences of user-item interactions

Contextual information

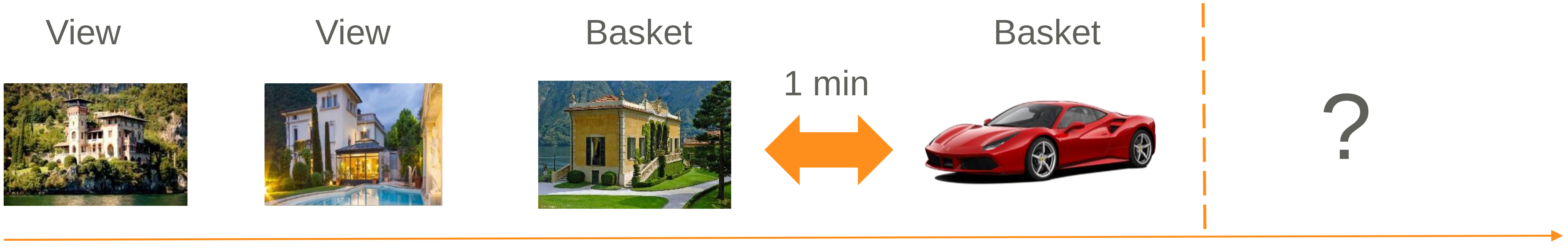
- Event type
- Timestamp

Examples: shopping, music listening, video watching

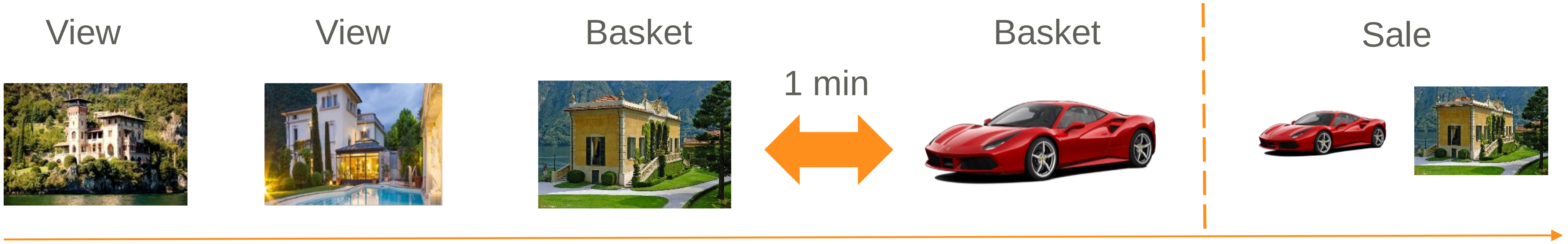
Next event prediction as recommendation task



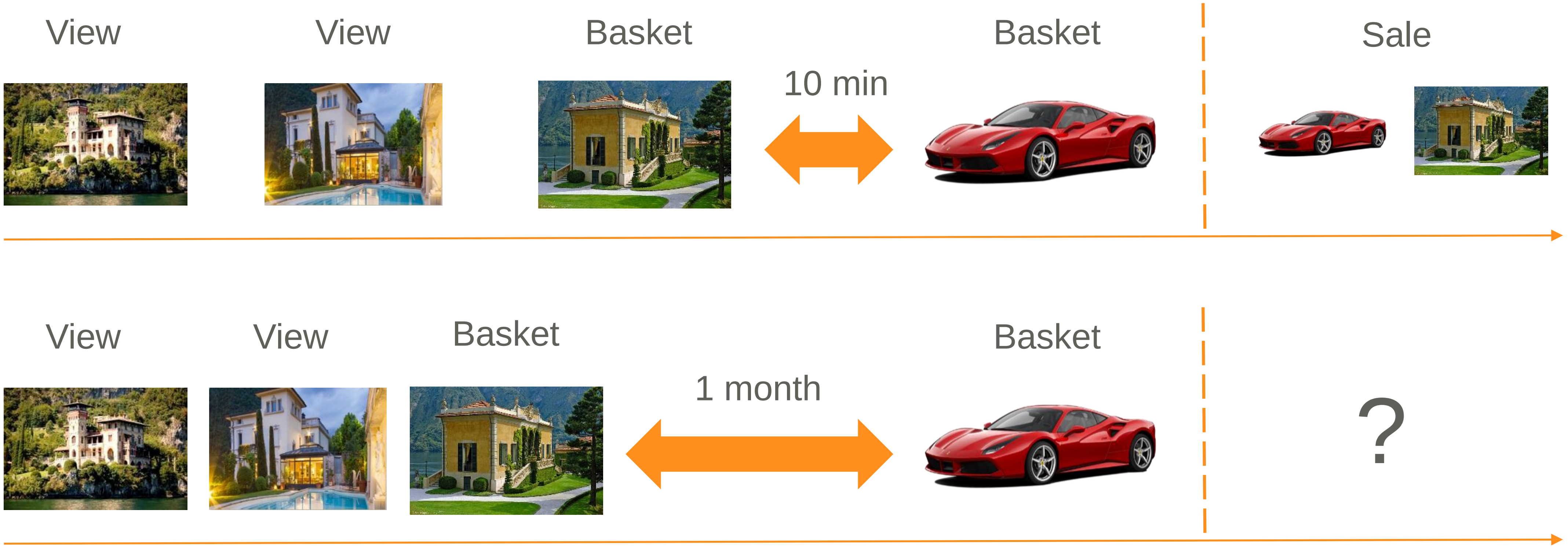
Motivation



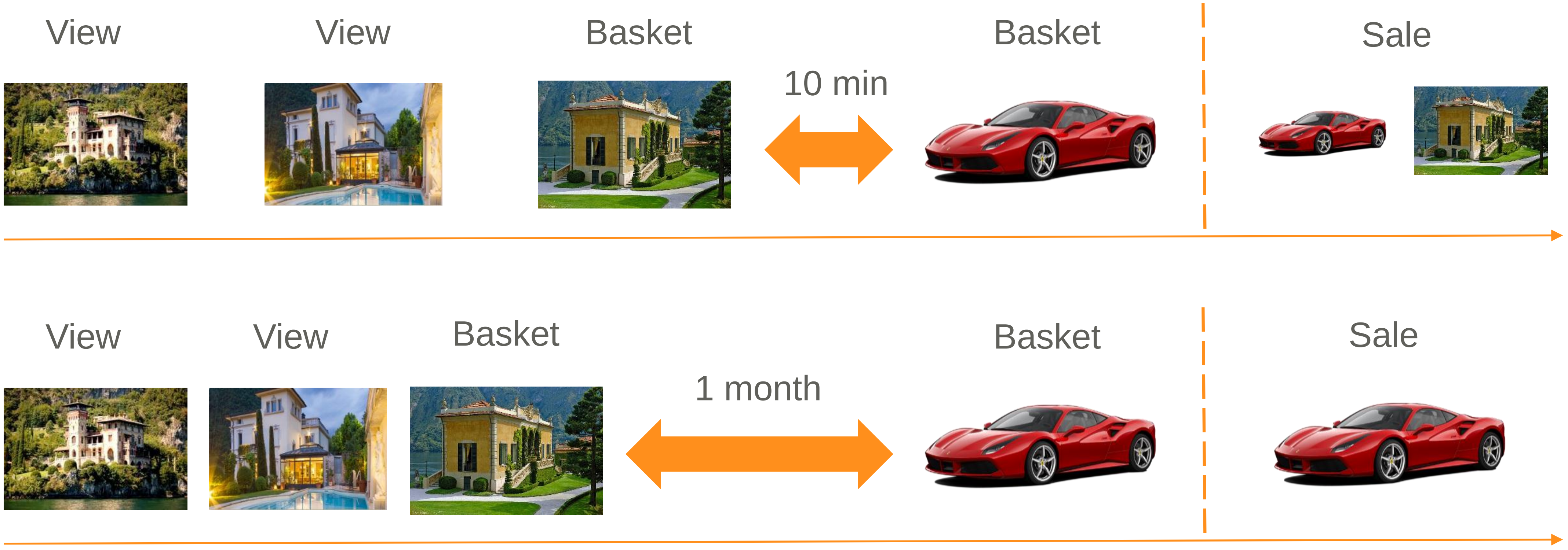
Motivation



Motivation


View
View
Basket
10 min
Basket
Sale
View
View
Basket
1 month
Basket
?

Motivation


View
View
Basket
10 min
Basket
Sale
View
View
Basket
1 month
Basket
Sale

Setup (cont'd)

Sequences of item-context pairs

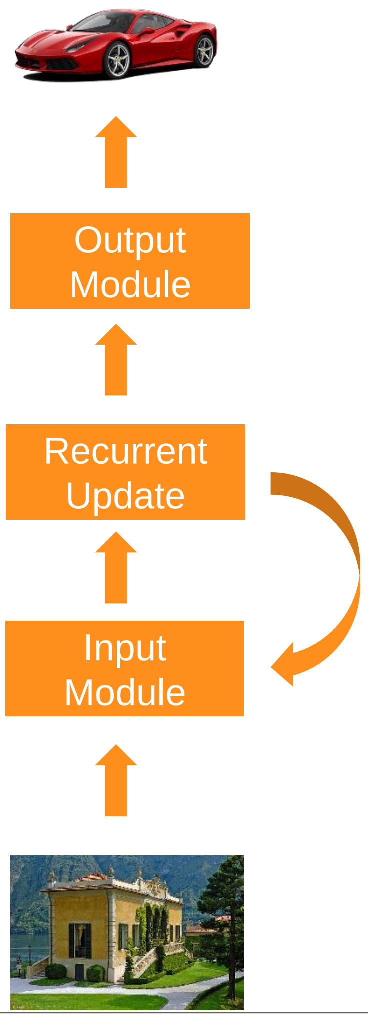
$$X = \{(x_t, c_t)\}, t = 1..T$$

To produce likely continuations of a sequence, we define a probability distribution over sequences

$$p(X) = p(x_1, .., x_T, c_1, .., c_T) = \prod_{t=1}^{T} p(x_t | c_t, x_{<t}, c_{<t})$$

We model $p(x_t | c_t, x_{<t}, c_{<t})$ using RNN

Recurrent Neural Network



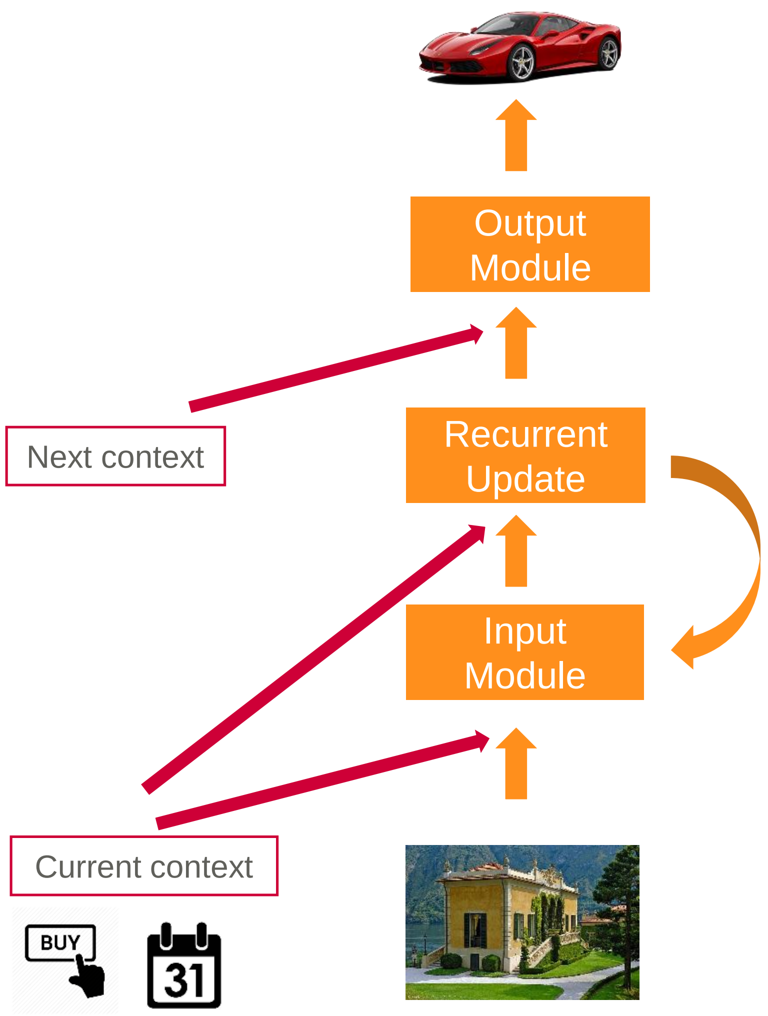
$$o_t = softmax(Vh_t)$$

$$h_t = f(x_t^{embed}, h_{t-1}) = \tanh(W_{hh}h_{t-1} + W_{xh}x_t^{embed} + b)$$

$$x_t^{embed} = x_t V$$

$$x_t = [0, \ldots, 1, \ldots, 0]$$

Conditioning RNN on Context



$$h_t^c = \theta^{out}(h_t, c_{t+1}), o_t = softmax(V h_t^c)$$

$$h_t = f^c(x_t^c, h_{t-1})$$

$$x_t^{embed} = x_t V, x_t^c = \theta^{in}(x_t^{embed}, c_t)$$

$$x_t = [0, .., 1, .., 0], c_t = \left[c_t^{evt}; c_t^{time}\right]$$

RNNs: Conditioning input and output

Concatenation $[x_t; c_t]$

no effect of context on item representation
additive term in state vector definition

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + W_{ch}c_t + b)$$

Multiplicative interaction $x_t \odot Cc_t$

tighter binding

changes item similarities based on context

Concatenation and multiplicative interaction $[x_t \odot Cc_t; c_t]$

[1] T. Mikolov and G. Zweig. 2012. Context dependent recurrent neural network language model.
[2] R. Kiros, R. Zemel, and R. Salakhutdinov. 2014. A Multiplicative Model for Learning Distributed Text-based Attribute Representations.

RNNs: Conditioning hidden dynamics

In the case of concatenation, context plays a limited role

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t + W_{ch} c_t + b)$$

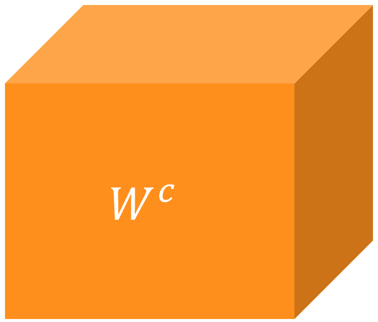
$$\frac{\partial h_t}{\partial h_{t-n}} = \prod_{k=t-n+1}^{t} W_{hh}^T diag(tanh'(W_{hh} h_{t-1} + W_{xh} x_t + W_{ch} c_t + b))$$

Ideally, a transition matrix per context

$$h_t = \tanh(W_{hh}^c h_{t-1} + W_{xh}^c x_t + W_{ch} c_t + b)$$

But impractical

context dimensions



RNNs: Conditioning hidden dynamics

Multiplicative term to gate the transition matrix on context

$$h_t = \tanh((W_{hh}h_{t-1} + W_{xh}x_t) \odot Uc_t + b)$$

Gradient contains context more explicitly

$$\frac{\partial h_t}{\partial h_{t-n}} = \prod_{k=t-n+1}^{t} W_{hh}^T diag(Uc_k) diag(tanh'(((W_{hh}h_{k-1} + W_{xh}x_k)) \odot Uc_k + b))$$

[3] Y. Wu, S. Zhang, Y. Zhang, Y. Bengio, and R. Salakhutdinov. 2016. On Multiplicative Integration with Recurrent Neural Networks.

Experimental Setup

Datasets

User browsing and purchasing activity on multiple e-commerce websites

YooChoose (RecSys Challenge’15)	8M sessions	37K products
Criteo Internal dataset	2.3M sessions	370K products

Contextual features

- Time gap
- Event type (view, basket, sale)
- Time (month, hour, day of week)

Recall@10

Model	YooChoose	Internal Dataset
Baselines		
CoVisit	0.374	0.329
BagOfItems	0.443	0.354
GRU RNN without context	**0.562**	**0.454**
Contextual RNNs		
Condition input, output and hidden transition	**+6.6%**	**+4.3%**

Zoom on uplift of Recall@10

As business, we are interested in sale events

As recommendation system, we are interested in recommending non-historical items -- items that user has not seen previously

	YooChoose	Internal dataset
Sale events	+12.5%	+6.1%
Non-historical products	+10.2%	+9.4%

Conclusion

- New ways of internalizing context into RNN
- Multiplication alone is more powerful than concatenation
- Conditioning on context in input / output and the hidden dynamics of the RNN brings the best uplift in performance
- Improved modeling of rare events and hard cases like non-historical items

Thank you!