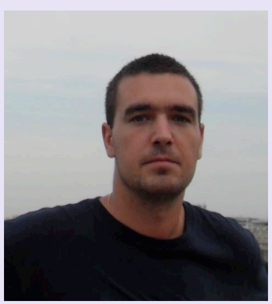
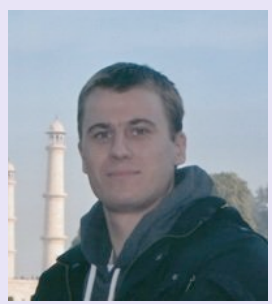
Hidden CRF with Distributed User Embeddings for Ad Targeting

Nemanja Djuric **Vladan Radosavljevic** **Mihajlo Grbovic** **Narayan Bhamidipati**

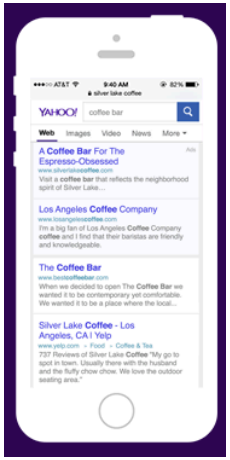
ICDM 2014
December 17th, 2014, in Shenzhen, China

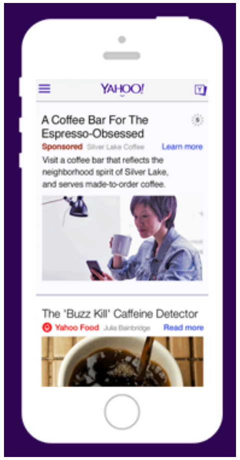
Overview of the talk

- What and why?
- Methodology
  - Distributed embeddings of user activities
  - Hidden Conditional Random Fields
- Results
- Conclusion

Introduction

- Millions of people visit Yahoo websites daily to search, read articles, check for email, …
- Targeted online ads: right time, right place, right person
  - Improves online experience
  - Brings benefits for both publishers and advertisers
- **Problem:** estimate users' propensity to click on an ad or purchase an item





Traditional approaches

- Problem is defined as a classification or a ranking task
- Learn a model based on:
  - Demographic information
  - Information aggregated over time based on activity history (e.g., pages visited, queries issued, ads viewed or clicked on, …)
  - Immediate search queries
- However, actual sequence of user actions is not used

Proposed approach

- We propose to infer user modes from a sequence of activities on a session level: *shopper* vs. *non-shopper*
  - Increases the value of each shown ad
  - Reduces the number of shown ads
- Example of a search sessions:

time

Label: shopper	Label: Non-shopper
halloween city stores	how to feed a baby
female star trek characters	how to keep a baby quiet
female star trek officers	how to update ur iphone 5
female star trek officers	how to update your iphone 5
star trek officers uniform	how to update ur iphone 5

Methodology: Data representation

- Many challenges ahead
  - Sparsity (e.g., large activity space)
  - Scalability (millions and millions of records)
- Embedding of user activities
  - Sessionize a sequence of user activities $\mathbf{s} = (x_1, \ldots x_T)$

session as a sentence

“halloween city stores” “female star trek characters” “female star trek officers” “star trek officers uniform”

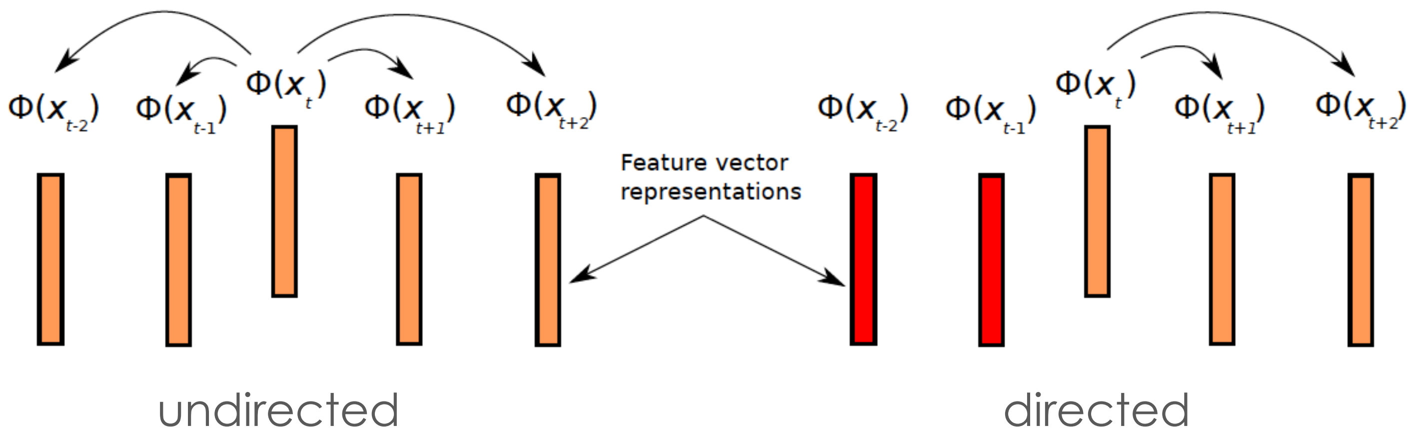
activity as a word

  - Low-dimensional activity representation using *directed skip-gram*: activity x_t represented as *D* dimensional vector $\Phi(x_t)$

“*halloween city stores*” →

1.2	0.1	3.7	-0.8	2.2

Methodology: Directed skip-gram



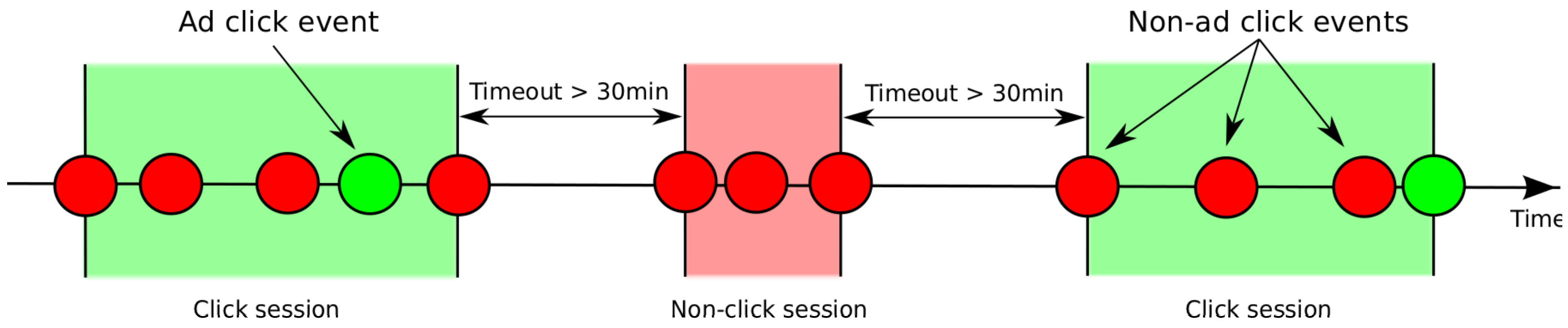
- The directed skip-gram: maximizes log-probabilities of future activities x_{t+i} given their preceding activity x_t

$$\text{maximize} \quad \frac{1}{T}\sum_{t=1}^{T}\sum_{1\leq i\leq l}\log \mathbb{P}(x_{t+i}|x_t)$$

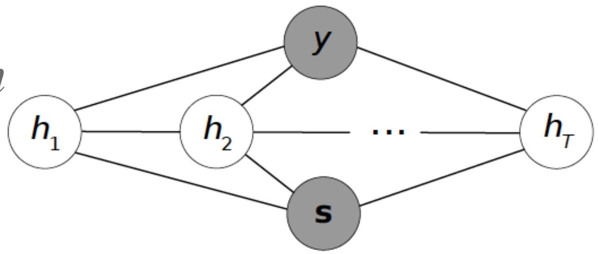
- The model focuses on predicting subsequent activities instead of both past and future contexts

Methodology: Hidden CRF

- We consider a task of inferring user modes: *shopper* vs. *non-shopper*



- Use HCRF to model user behavior
  - Each activity assigned latent variable h
  - Label: shopper vs. non-shopper



Methodology: Hidden CRF

- Hidden CRF models distribution

$$\mathbb{P}(y, \mathbf{h}|\mathbf{s}, \mathbf{w}) = \frac{1}{Z(\mathbf{h}, \mathbf{s}, \mathbf{w})} \exp\left(\Psi(y, \mathbf{h}, \mathbf{s}, \mathbf{w})\right)$$

$$\Psi(\mathbf{h}, y, \mathbf{s}, \mathbf{w}) = \sum_{t=1}^{T} \sum_{k=1}^{|\mathcal{H}|} \sum_{d=1}^{D} w_{kd}^{(1)} f_{kd}^{(1)}(x_t, h_t) + \sum_{t=1}^{T} \sum_{k=1}^{|\mathcal{H}|} \sum_{r=1}^{|\mathcal{Y}|} w_{kr}^{(2)} f_{kr}^{(2)}(y, h_t) + \sum_{t=2}^{T} \sum_{k_1, k_2=1}^{|\mathcal{H}|} \sum_{r=1}^{|\mathcal{Y}|} w_{k_1 k_2 r}^{(3)} f_{k_1 k_2 r}^{(3)}(y, h_{t-1}, h_t)$$

$f_{kd}^{(1)}(x_t, h_t) = \Phi_d(x_t) \cdot I(h_t = k)$ compatibility of activity representations and hidden states

$f_{kr}^{(2)}(y, h_t) = I((h_t = k) \wedge (y = r))$ compatibility between hidden state h_t and session-level label y

$f_{k_1 k_2 r}^{(3)}(y, h_{t-1}, h_t) = I((h_{t-1} = k_1) \wedge (h_t = k_2) \wedge (y = r))$ compatibility between neighboring hidden states and the label

- Training: $\textbf{maximize} \ \log \mathbb{P}(\mathbf{y}|\mathbf{s}, \mathbf{w}) = \sum_{i=1}^{N} \log \mathbb{P}(y_i|\mathbf{s}_i, \mathbf{w}) - \lambda \mathbf{w}^{\mathrm{T}} \mathbf{w}$
- Inference: $\hat{y} = \operatorname{argmax}_y \left(\sum_{\mathbf{h}} \mathbb{P}(y, \mathbf{h}|\mathbf{s}, \mathbf{w}) \right)$

Experiments: Setup

- Activity representation learned on one month of Yahoo log data (October 2013)
- Vector dimension set to 100, window length to 5
- Labeled data for HCRF
  - Balanced set of ~500,000 click/non-click or purchase/non-purchase sessions
  - Minimum session length 5
- Competing methods
  - logistic regression, SVM (using average of vectors in session)
  - HMM, CRF (each activity is labeled with session label)
- 5-fold crossvalidation

Experiments: Activity embeddings

- Learned distributed activity representations give very intuitive results
- Example: nearest neighbors with respect to cosine distance for search query “*baseball*”

Non-shopper

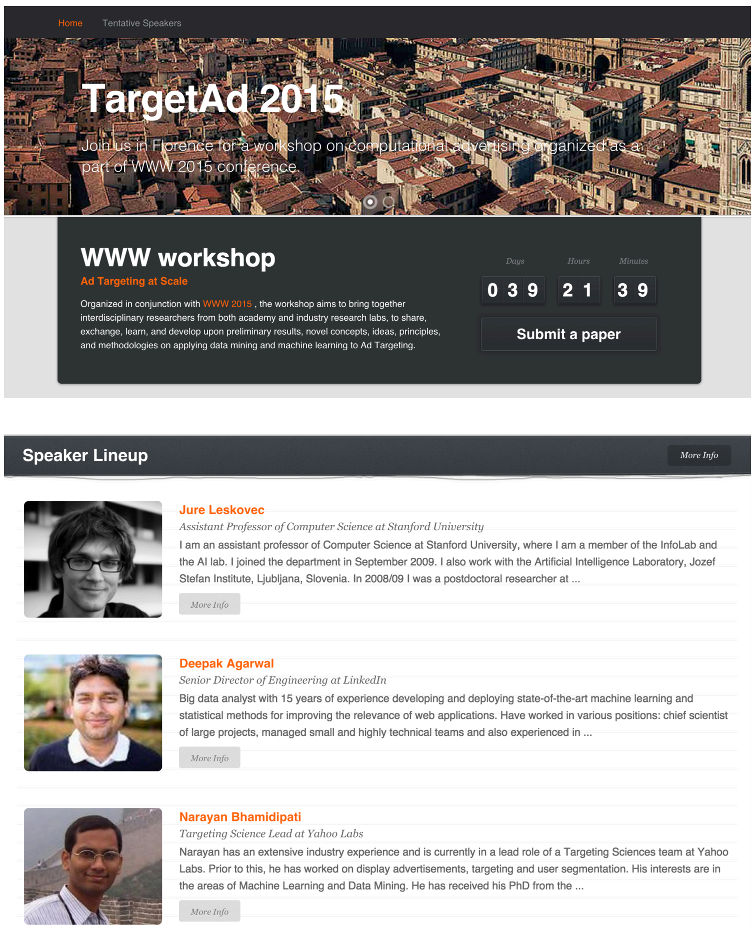
Shopper

Experiments: Results

- Relative improvement over logistic regression model, where either clicks or purchases are used as labels
- HCRF outperforms alternatives for all accuracy measures
- Directed skip-gram outperformed undirected approach

		SVM	**HMM**	**CRF**	**HCRF**$_{undir}$	**HCRF**$_{dir}$
Click pred.	Prec.	7.82	1.03	3.89	9.06	12.9
	Rec.	8.84	1.10	3.59	9.85	11.1
	F1	8.29	1.06	3.74	9.39	12.1
Purch. pred.	Prec.	1.03	0.17	0.86	2.07	2.58
	Rec.	2.04	0.41	0.20	2.65	3.47
	F1	1.57	0.30	0.50	2.38	3.06

www.targetad-workshop.net



The topics of this workshop include, but are not limited to, the following areas:

- Machine learning-based Ad Targeting
- Ad Targeting in Social Networks
- Ad Targeting and Big Data
- Ad Targeting on Mobile Devices
- Big Data Platforms
- Recommender systems
- Pay-per-install targeting strategies
- In-game and in-app advertising
- Deep Learning for Ad Targeting
- Machine Learning in Sponsored Search
- Contextual advertising
- Large-scale user modeling for Ad Targeting
- Data-driven methods for interest targeting
- Privacy-preserving targeting
- Behavioral targeting
- Personalization
- Post-conversion Feature Attribution
- Learning to rank for interest targeting
- Click-conversion modeling

The workshop is organized by the following industry researchers:

- Vanja Josifovski (Google)
- Olivier Chapelle (Criteo Labs)
- Mihajlo Grbovic (Yahoo Labs)
- Vladan Radosavljevic (Yahoo Labs)
- Nemanja Djuric (Yahoo Labs)

Program committee of the workshop includes the following area experts:

- Shuang Yang (Twitter)
- Minmin Chen (Criteo Labs)
- Junfeng Pan (Facebook Inc.)
- Kosta Ristovski (Hitachi Labs)
- Sujith Ravi (Google Research)
- Slobodan Vucetic (Temple University)
- Abhimanyu Das (Microsoft Research)
- Zornitsa Kozareva (Yahoo Labs)
- Suleyman Cetintas (Yahoo Labs)
- Fabrizio Silvestri (Yahoo Labs)
- Dmitry Pechyony (Microsoft)
- Onno Zoeter (Xerox Research)

Conclusions

- The proposed directed skip-gram architecture finds useful representations of user actions
- HCRF significantly outperformed the baseline approaches
- Ongoing/future work
  - Incorporating context, demographics, location, weather, and other useful signals
  - Multi-vector representations of a single activity

Thank you!

- Questions?