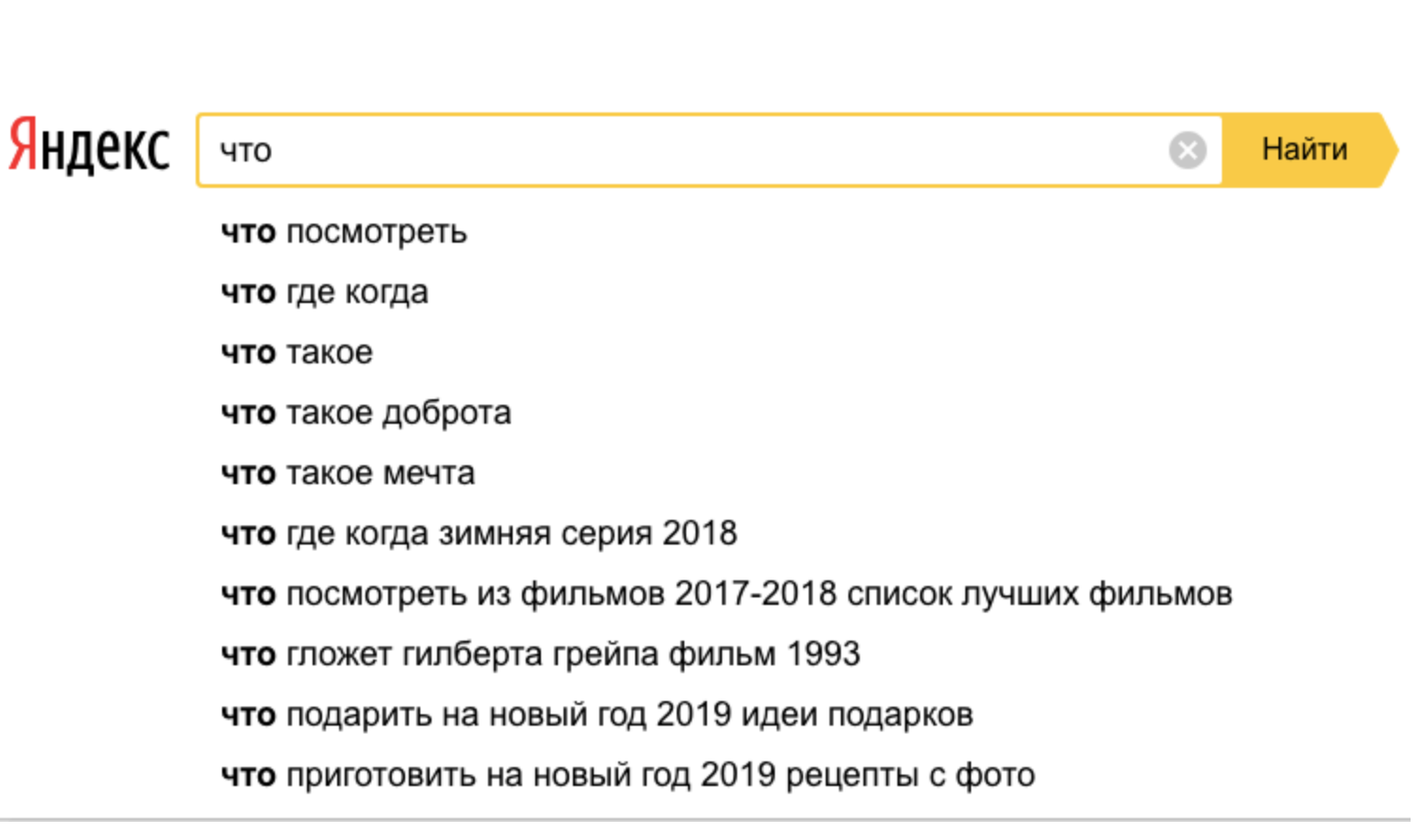
Языковые модели в задачах предиктивного пользовательского ввода

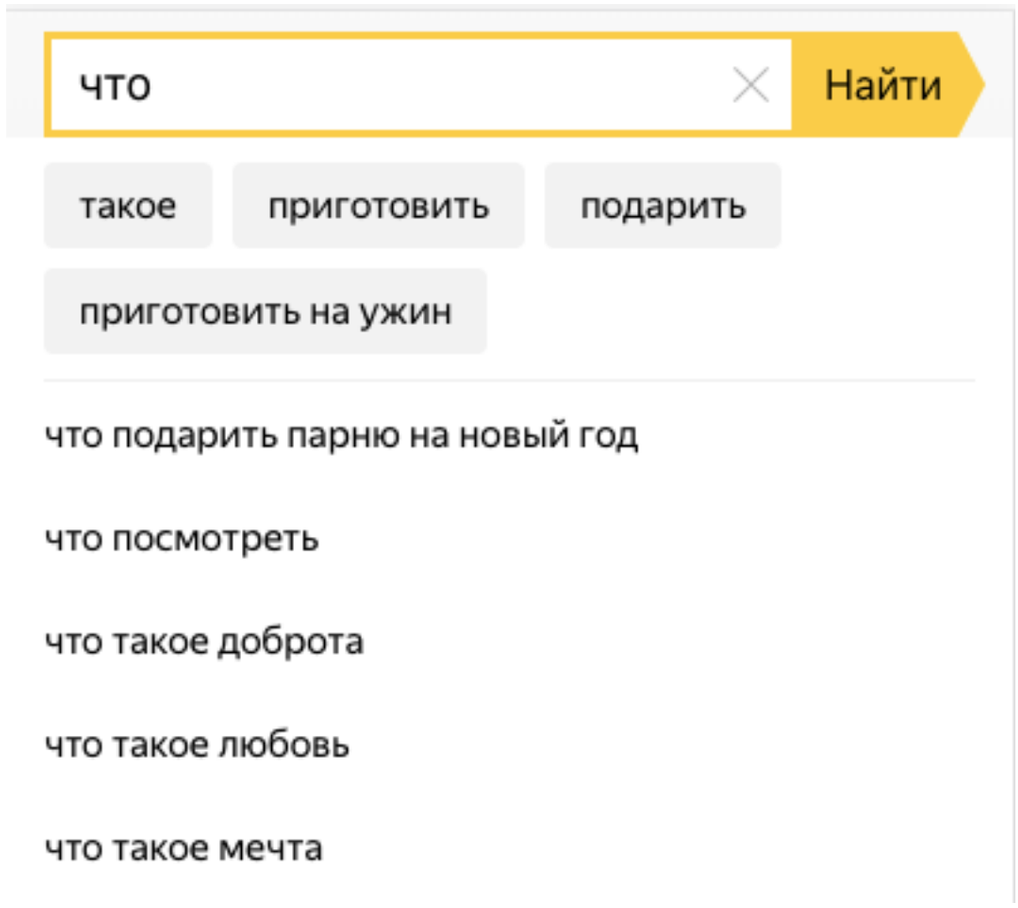
Студент: Турумтаев Галим
Научный руководитель: Шаграев Алексей

2018

Постановка Задачи:

Предсказание поискового запроса пользователя
с помощью языковых моделей





Проблемы:

- Скорость ответа
- Русский язык
- Другой синтаксис
- Сложные метрики
- Свежесть

Метрики

Онлайн:

- Время набора запроса / кол-во символов
- Клики на подсказки

Оффлайн:

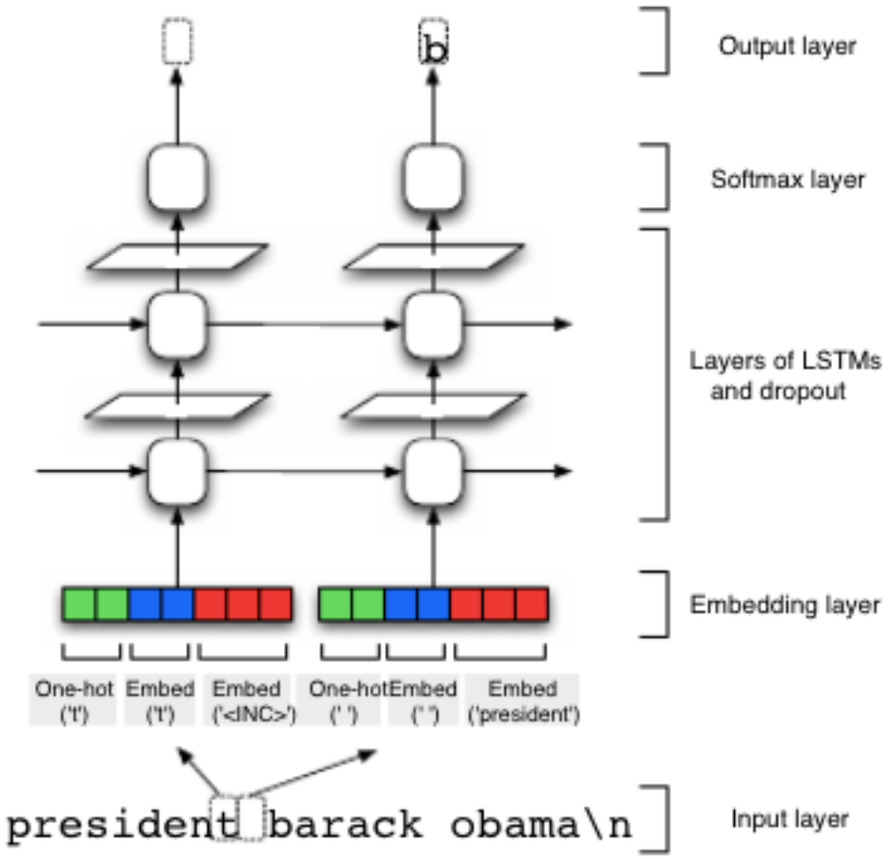
- MRR@k
- Recall@k

Что другие люди уже делали в этой области?

A Neural Language Model for Query Auto-Completion

Dae Hoon Park*
Huawei Research America
Santa Clara, CA, USA
daehpark@gmail.com

Rikio Chiba
Yahoo Inc.
Sunnyvale, CA, USA
rchiba@yahoo-inc.com



$$L = \sum_{q \in Q} \sum_{j=1}^{|q|-1} \log p(t_{j+1} | t_1, \ldots, t_j) \tag{4}$$

Table 2: MRR and PMRR evaluation results. The highest score for each metric is in bold face.

Model	MRR Seen	MRR Unseen	MRR All	PMRR Seen	PMRR Unseen	PMRR All
MPC [1]	0.428	0.000	0.171	0.566	0.000	0.225
Char. n-gram (n=7)	0.363	0.236	0.287	0.550	0.376	0.445
Mitra10K+MPC+λMART [12]	0.427	0.179	0.278	0.586	0.297	0.412
Mitra100K+MPC+λMART [12]	0.428	0.212	0.298	0.588	0.368	0.455
Proposed models						
NQLM(S)	0.381	0.287	0.325	0.557	0.460	0.499
NQLM(S)+WE	0.406	0.286	0.334	0.582	0.445	0.500
NQLM(L)+WE	0.419	0.303	0.349	0.589	0.465	0.514
NQLM(S)+MPC	0.433	0.287	0.346	0.580	0.460	0.508
NQLM(S)+WE+MPC	**0.434**	0.286	0.345	0.580	0.445	0.499
NQLM(L)+WE+MPC	**0.434**	0.303	**0.355**	0.580	0.465	0.511
NQLM(S)+MPC+λMART	0.428	0.288	0.344	**0.594**	0.465	0.516
NQLM(S)+WE+MPC+λMART	0.428	0.288	0.344	0.590	0.454	0.508
NQLM(L)+WE+MPC+λMART	0.428	**0.305**	0.354	0.593	**0.475**	**0.522**

SIGIR '17, August 07-11, 2017, Shinjuku, Tokyo, Japan

Personalized Language Model for Query Auto-Completion

Aaron Jaech and Mari Ostendorf
University of Washington
{ajaech, ostendor}uw.edu

Size	Model	Seen	Unseen	All
	MPC	.292	.000	.203
	Unadapted	.292	.256	.267
(S)	ConcatCell	.296	.263	.273
	FactorCell	.300	.264	.275
	Unadapted	.324	.286	.297
(B)	ConcatCell	.330	.298	.308
	FactorCell	.335	.298	.309

Table 2: MRR reported for seen and unseen prefixes for small (S) and big (B) models.

$$h_t = \sigma([w_t, h_{t-1}]\mathbf{W} + b + \mathbf{V}u) \quad (1)$$

$$h_t = \sigma([w_t, h_{t-1}]\mathbf{W}' + b) \quad (2)$$

$$\mathbf{A} = (u_i \times_1 \mathbf{Z}_L)(\mathbf{Z}_R \times_3 u_i) \quad (3)$$

Personalized neural language models for real-world query auto completion

Nicolas Fiorini
National Center for Biotechnology Information
National Library of Medicine, NIH
Bethesda, MD, USA
nicolas.fiorini@nih.gov

Zhiyong Lu
National Center for Biotechnology Information
National Library of Medicine, NIH
Bethesda, MD, USA
zhiyong.lu@nih.gov

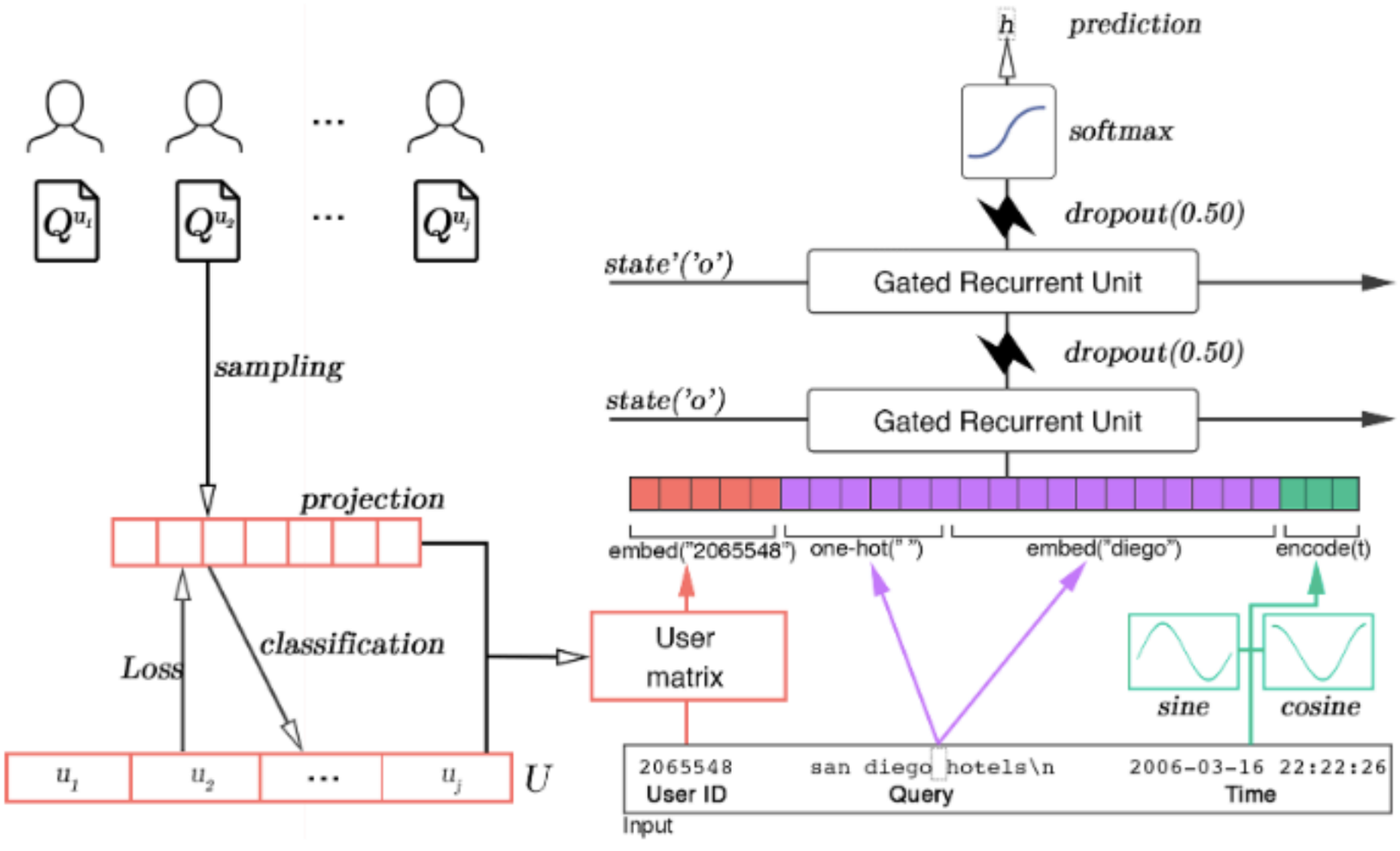


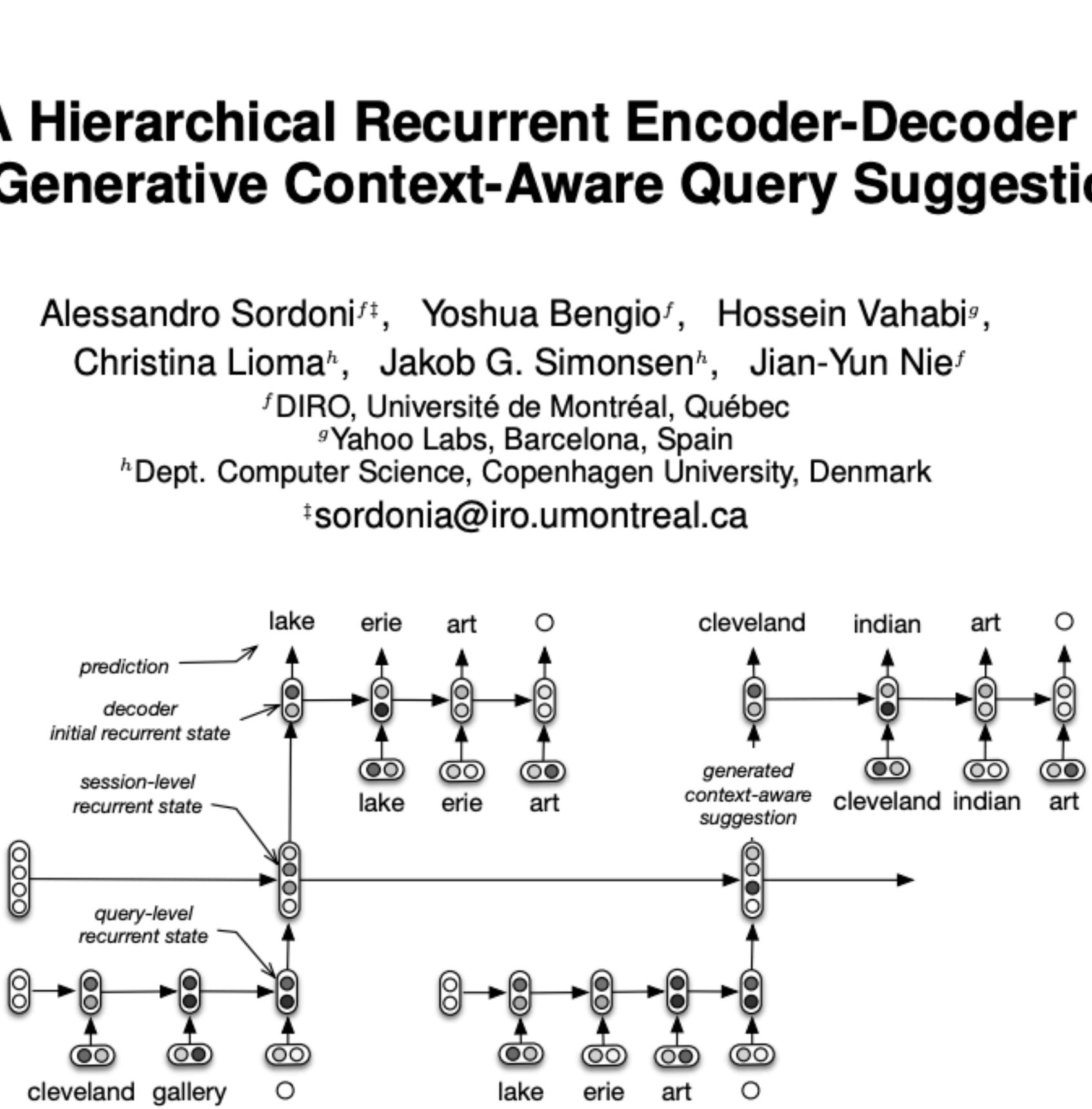
Figure 1: Architecture of our proposed model.

Table 1: MRR results for all tested models on the AOL and biomedical datasets with their average prediction time in seconds.

Model	AOL dataset				Biomedical dataset			
	MRR			Time	MRR			Time
	Seen	Unseen	All		Seen	Unseen	All	
MPC (Bar-Yossef and Kraus, 2011)	**0.461**	0.000	0.184	**0.24**	0.165	0.000	0.046	**0.29**
NQLM(L)+WE+MPC+λMART (Park and Chiba, 2017)	0.430	0.306	0.356	1.33	0.159	0.152	0.154	2.35
Our models in this paper								
NQAC	0.406	0.319	0.354	0.94	0.155	0.139	0.143	1.73
NQAC$_U$	0.417	0.325	0.361	0.98	**0.191**	0.161	0.169	1.77
NQAC$_{UT}$	0.424	0.326	0.365	0.95	0.101	**0.195**	0.157	1.81
NQAC$_{UT}$+D	0.427	0.326	0.366	1.32	0.186	0.185	0.185	2.04
NQAC$_{UT}$+MPC	**0.461**	0.326	0.380	0.68	0.165	**0.195**	**0.187**	1.20
NQAC$_{UT}$+MPC+λMART	0.459	**0.330**	**0.382**	1.09	0.154	0.179	0.172	2.01

A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion

Alessandro Sordoni[f‡], Yoshua Bengio[f], Hossein Vahabi[g],
Christina Lioma[h], Jakob G. Simonsen[h], Jian-Yun Nie[f]
[f]DIRO, Université de Montréal, Québec
[g]Yahoo Labs, Barcelona, Spain
[h]Dept. Computer Science, Copenhagen University, Denmark
[‡]sordonia@iro.umontreal.ca



Attention-based Hierarchical Neural Query Suggestion

Wanyu Chen
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
wanyuchen@nudt.edu.cn

Fei Cai*
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
caifei@nudt.edu.cn

Honghui Chen
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
caifei@nudt.edu.cn

Maarten de Rijke
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

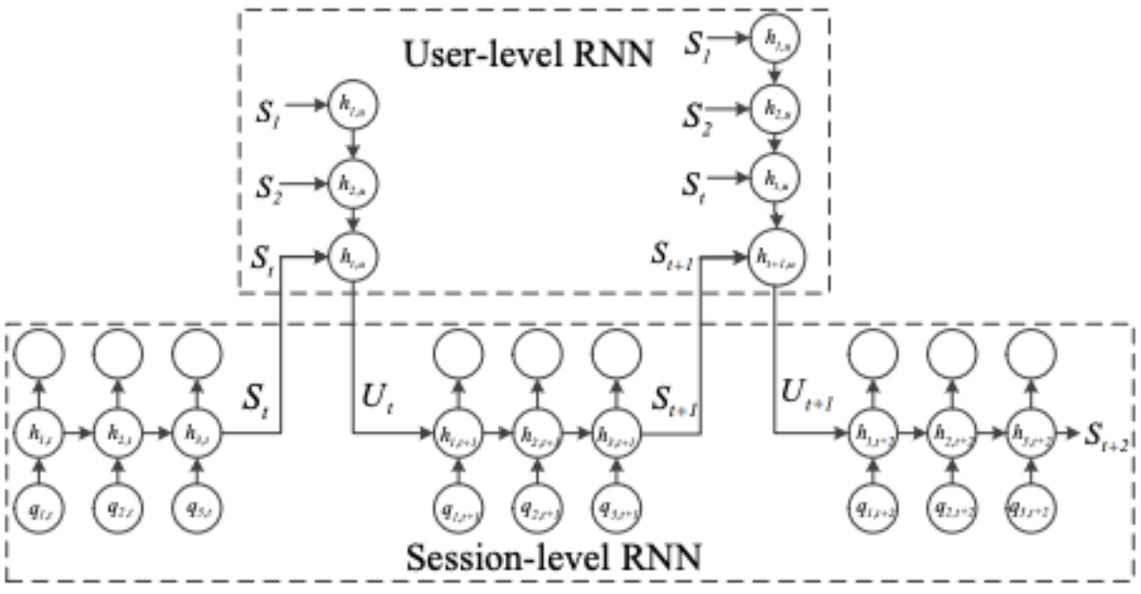


Figure 2: Structure of the HNQS model.

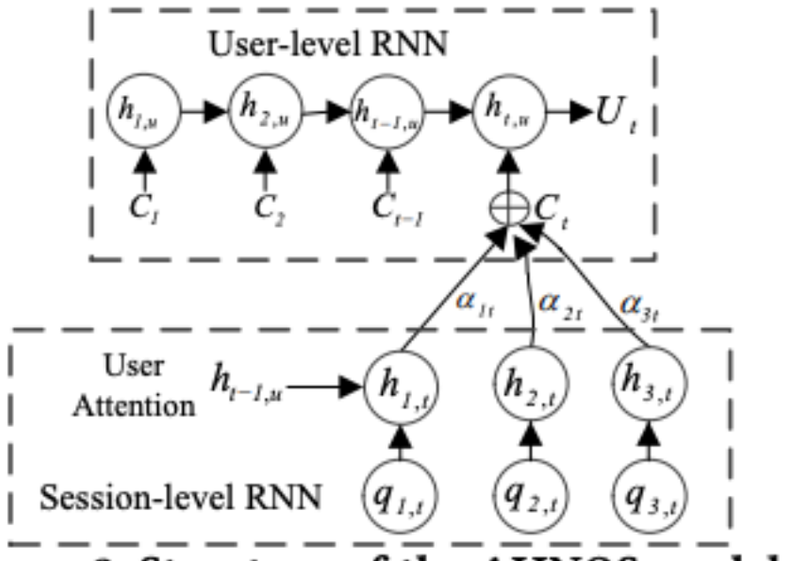


Figure 3: Structure of the AHNQS model.

План действий:

- Изучить литературу
- Взять / подготовить данные
- Попробовать разные ~~задачи~~ архитектуры
- Сравнить эти модели между собой
- Улучшать скорость / качество