max planck institut
informatik

PACRR: A Position-Aware Neural IR Model for Relevance Matching

Kai Hui[1], Andrew Yates[1], Klaus Berberich[1], Gerard de Melo[2]
[1]Max Planck Institute for Informatics
{khui, kberberi, ayates}@mpi-inf.mpg.de
[2] Rutgers University, New Brunswick
gdm@demelo.org

MAX-PLANCK-GESELLSCHAFT

Conference on Empirical Methods in Natural Language Processing 2017

Motivation

- Decades of research in ad-hoc retrieval provides useful measures to boost the performance

- Unigram matching signals have been successfully incorporated in neural IR models [2,4]

- How to incorporate positional matching information remains unclear

Matching Information to Incorporate

QUERY

computer science course Denmark

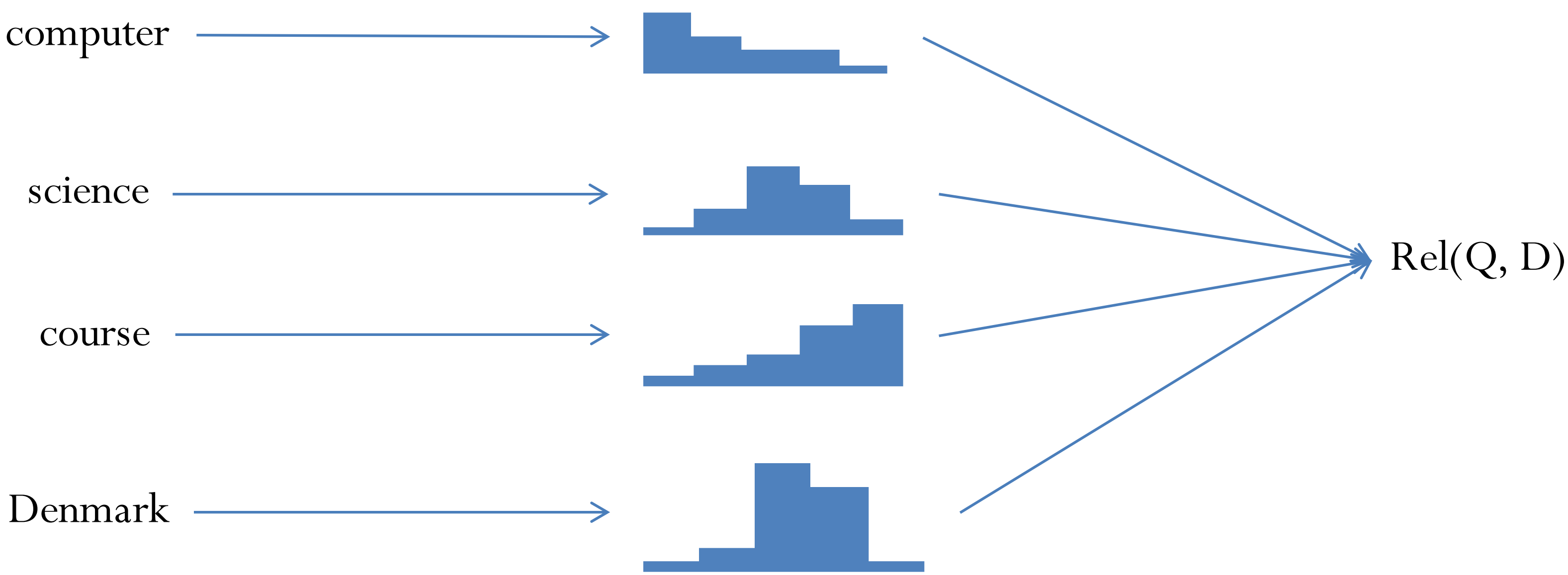
DOCUMENT

1. Institutes in **Denmark** provide graduate-level **courses** in **computer science.**
2. PCHandle is an online portal for purchasing personal **computers** in **Denmark.**

- Unigram matching: matching individual terms independently
- Term dependency: “computer science”
- Query proximity: the proximity between different matchings

Model Unigram Matching by Counting

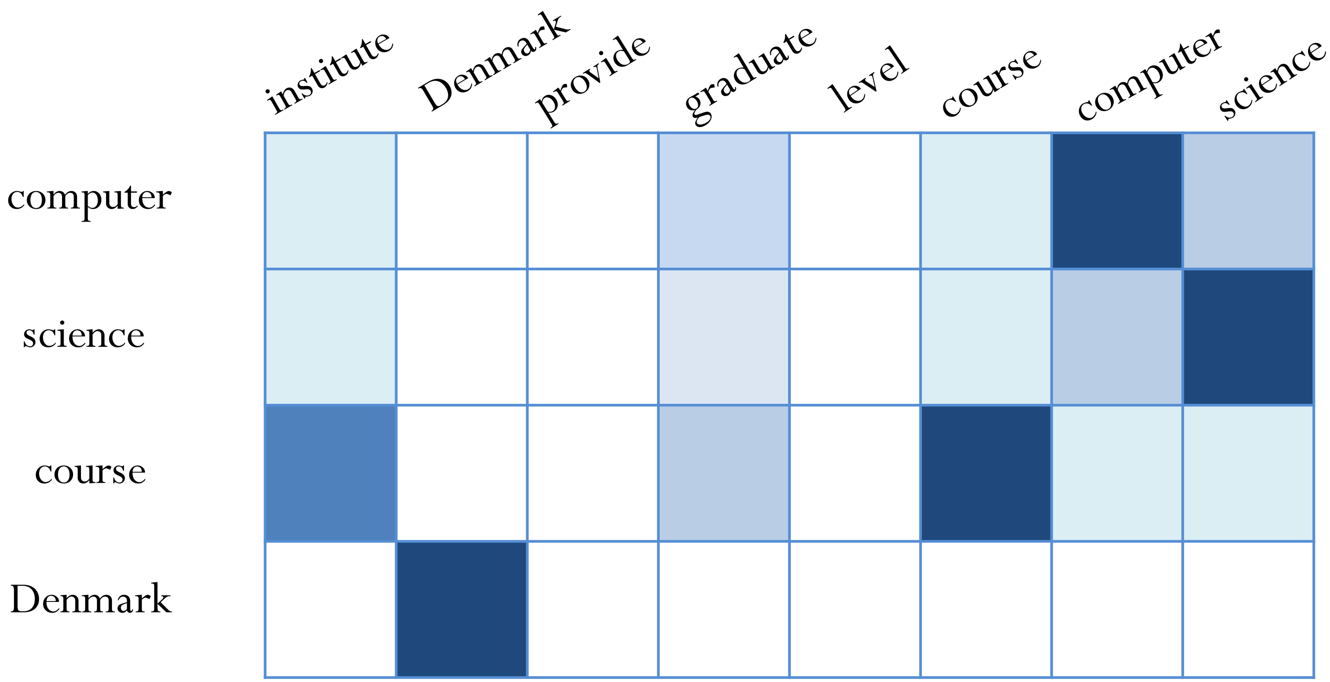
- Given a query Q and a document D
- Compute the semantic similarity between each term pair, where one term is from Q and another is from D (via word2vec)
- Group such similarity into bins and model the relevance between Q and D with a histogram [2]



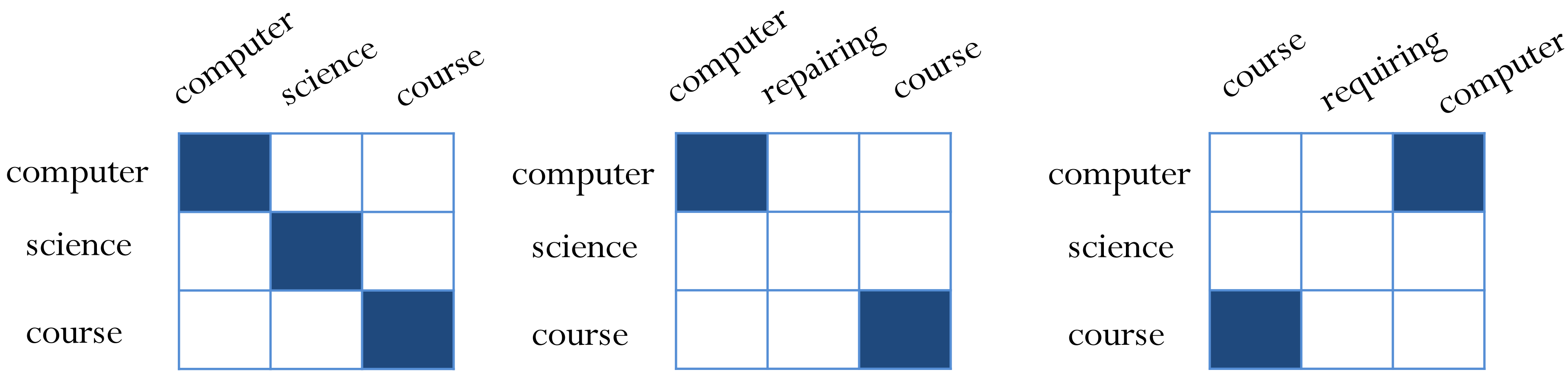
bag-of-word assumption
(independence among terms)

Beyond Unigram Matching: Model Positional Information

1) Retain the similarity into the similarity matrix, keeping both similarity and their relative positions [1,3,5]

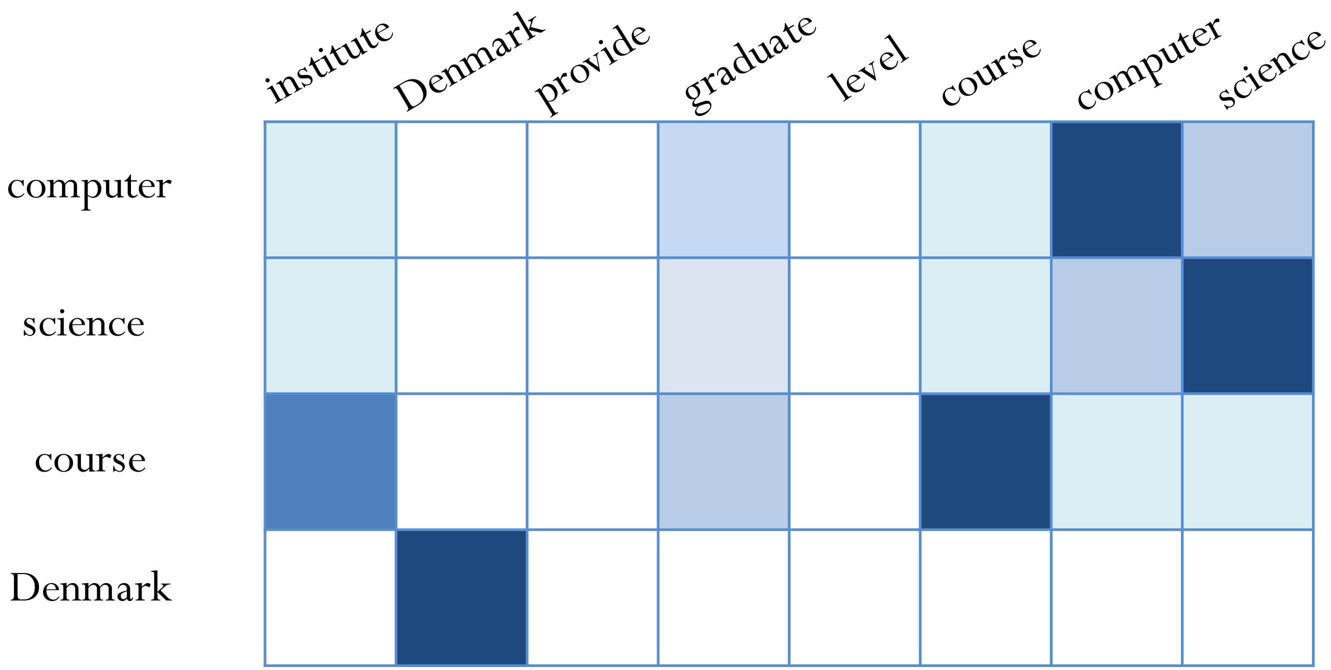


Beyond Unigram Matching: Model Positional Information



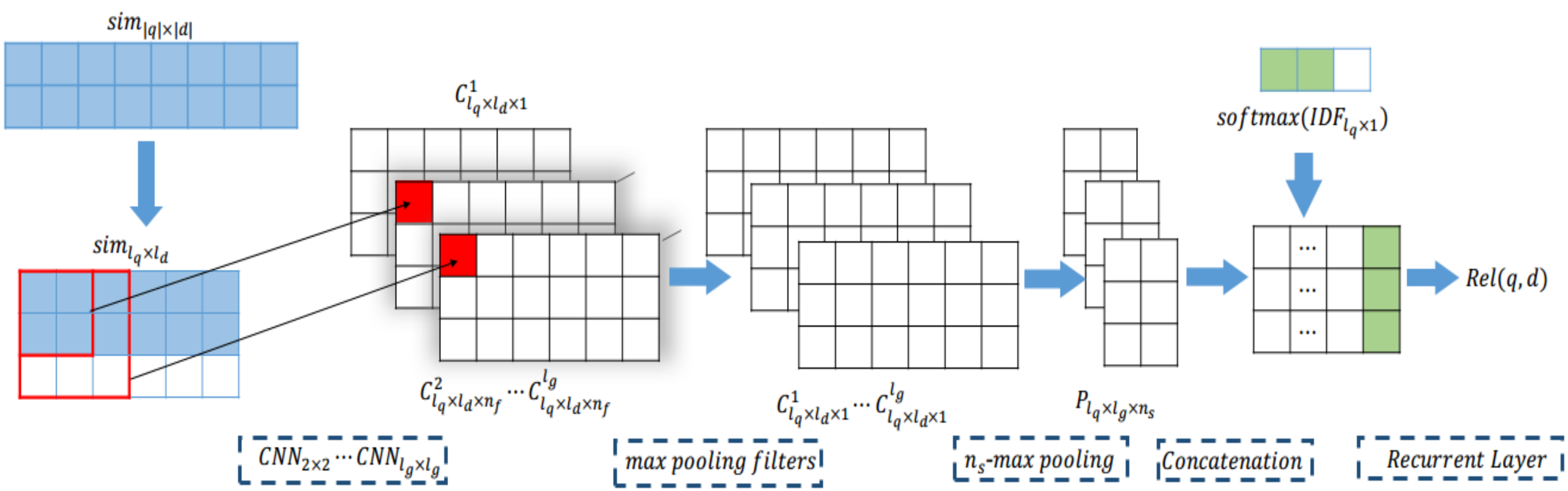
2) Matching could be modeled based on different local patterns in the similarity matrix
3) Individual text windows only include one salient matching pattern

Beyond Unigram Matching: Model Positional Information



4) Only retain the salient matching signals for individual query terms

PACRR: Position-Aware Convolutional Recurrent Relevance Matching



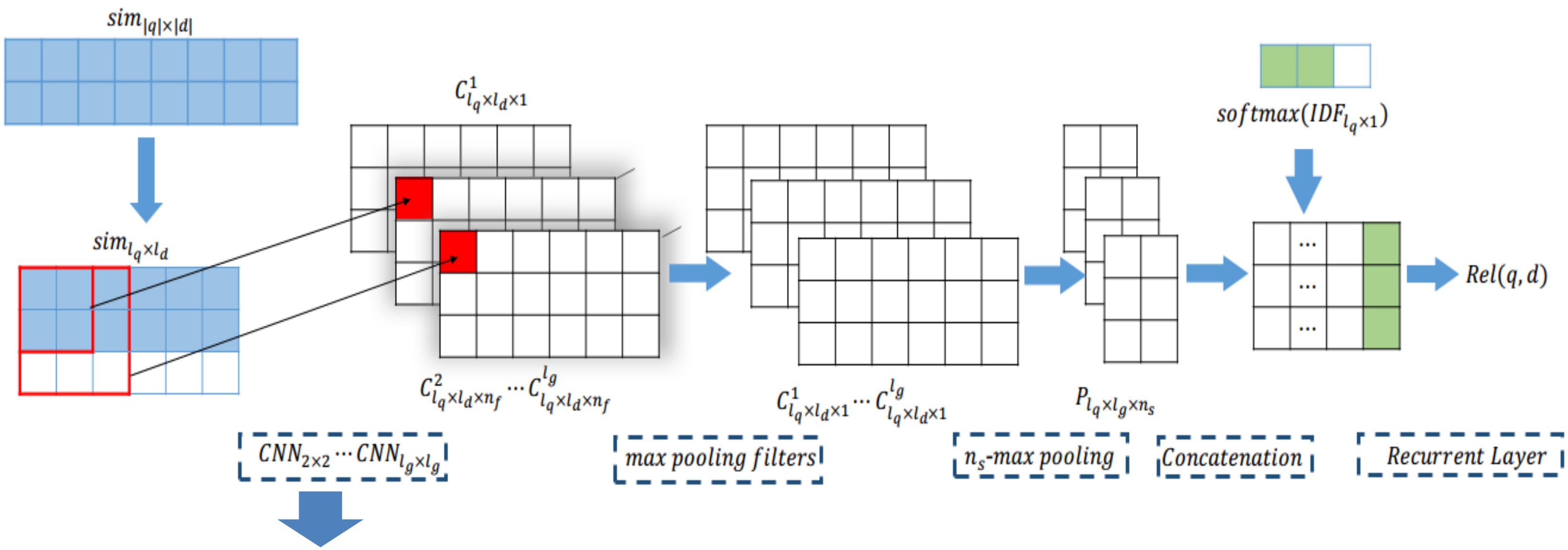
(1) CNN layers with different sizes: 2X2, 3X3, 4X4, etc..

(2) Max-pooling among filters

(3) K-max pooling: retain the k most salient signals for each query term

(4) LSTM layer for combination

PACRR: Position-Aware Convolutional Recurrent Relevance Matching



- CNN kernels (dozens of filters) in different sizes, corresponding to text windows with different length

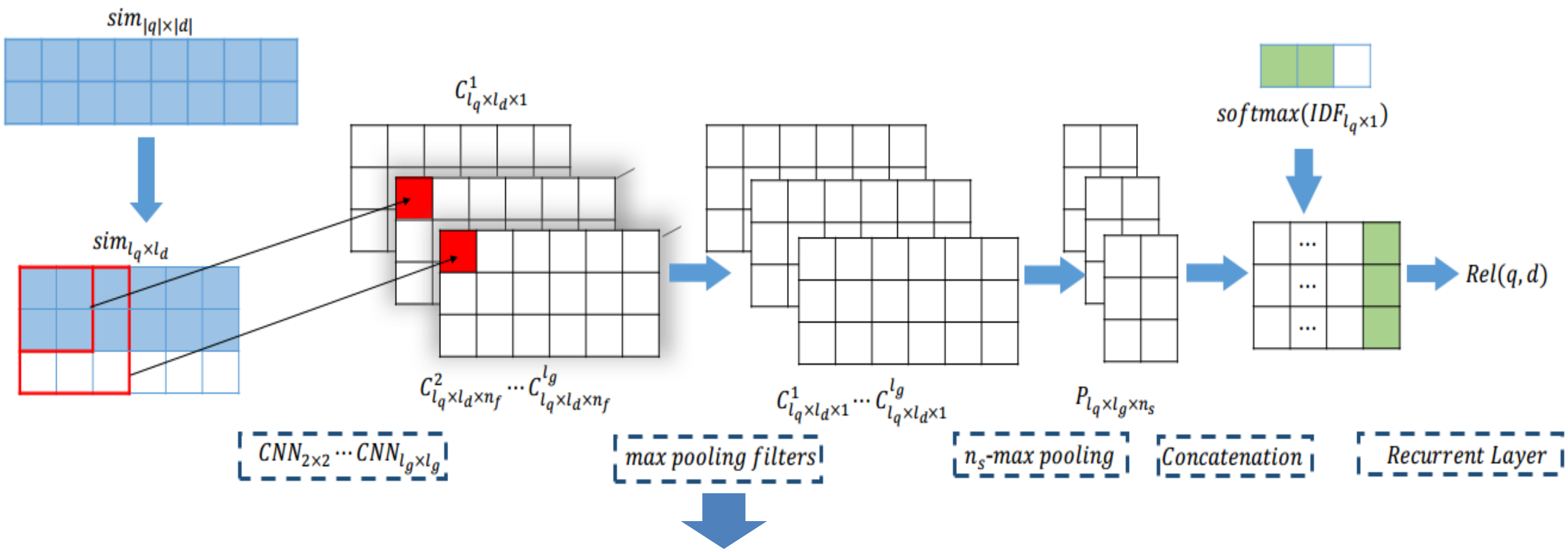
1	0		1	0		0	0
0	1		0	0		1	0

computer science, science course, etc..

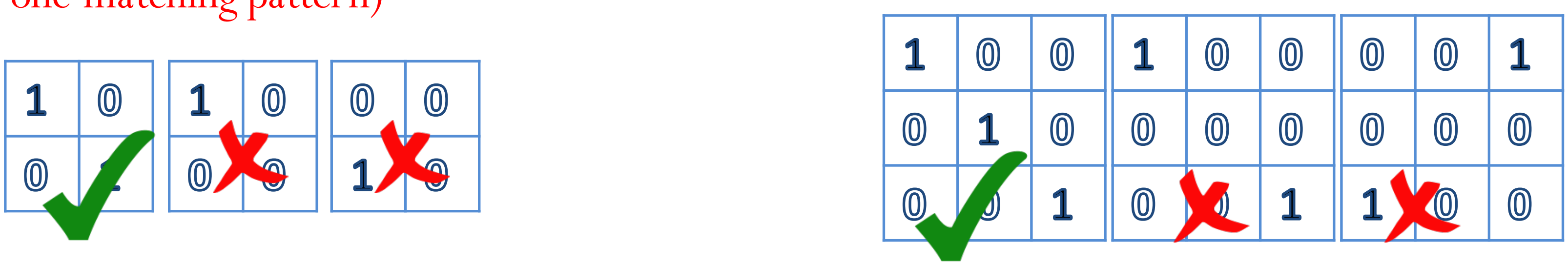
1	0	0		1	0	0		0	0	1
0	1	0		0	0	0		0	0	0
0	0	1		0	0	1		1	0	0

computer science course, science course Denmark, etc..

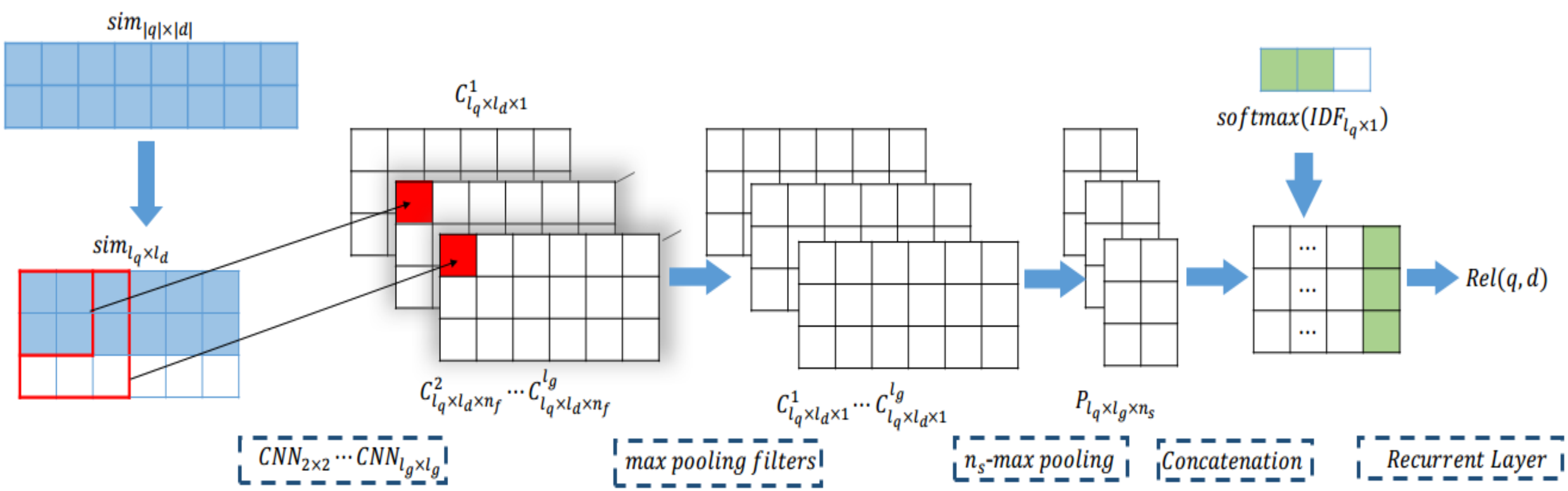
PACRR: Position-Aware Convolutional Recurrent Relevance Matching



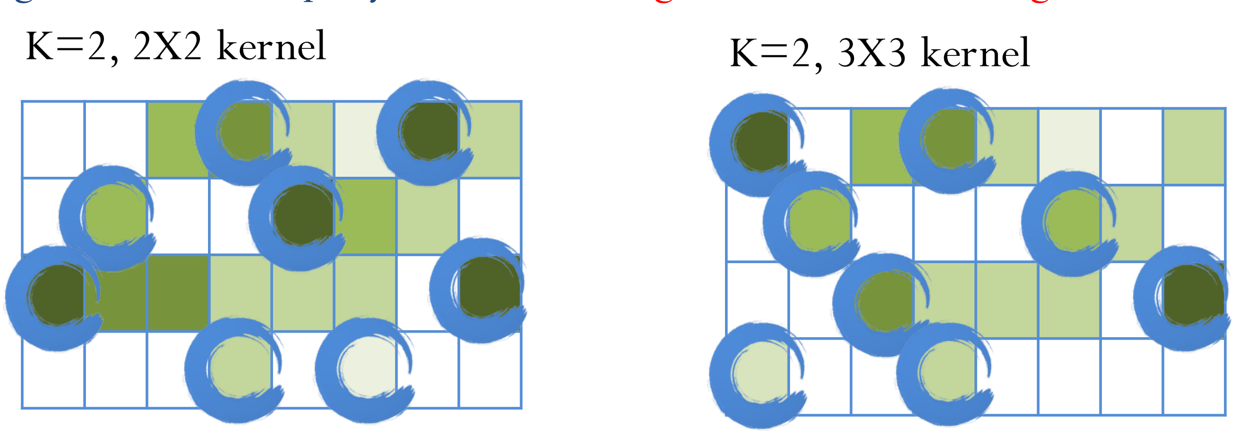
- Max pooling different filters for individual kernels (individual text windows at most include one matching pattern)



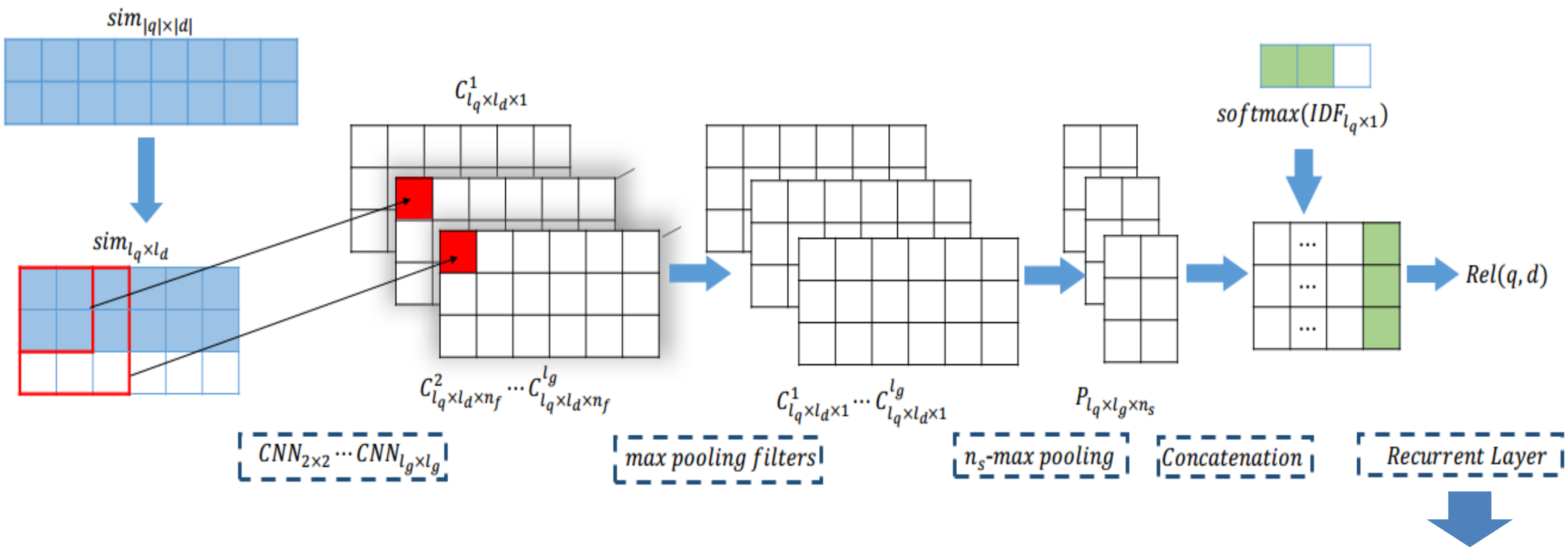
PACRR: Position-Aware Convolutional Recurrent Relevance Matching



- K-max pooling for individual query terms, retaining the k most salient signals for individual query terms



PACRR: Position-Aware Convolutional Recurrent Relevance Matching



- A LSTM layer combines signals on different query terms

Evaluation

- Based on TREC Web Track ad-hoc task 2009-2014, including 300 queries, 100k judgments and approx. 50 runs in each year

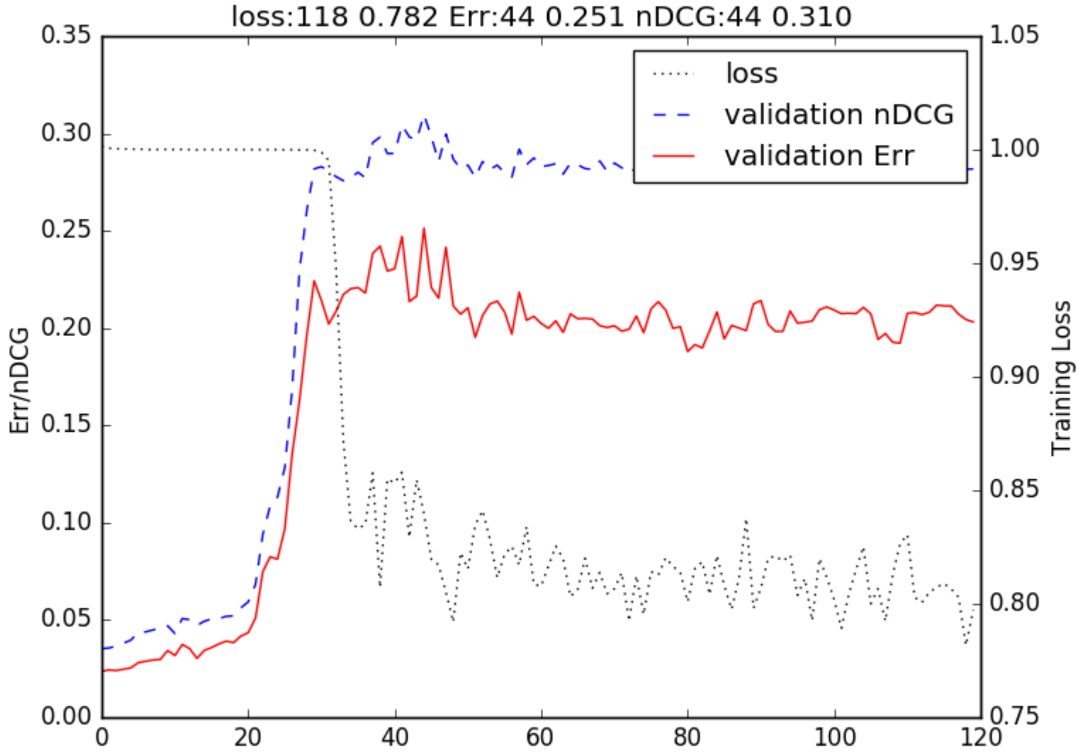
- Measures: ERR@20
  - A real value measure summarizing the quality of a ranking
  - The higher the better

- Baseline models: MatchPyramid [1], DRMM [2], local model in DUET [3], and K-NRM [4]

Training and Validation

- Employ five years (250 queries) for training and validation
- Randomly reserve 50 queries from the 250 queries for validation, and the model selection is per ERR@20
- Test on the remaining year (50 queries)

Training and Validation



The training loss, ERR@20 and nDCG@20 per iteration on validation data. The x-axis denotes the iterations. The y-axis indicates the ERR@20/nDCG@20 (left) and the loss (right).

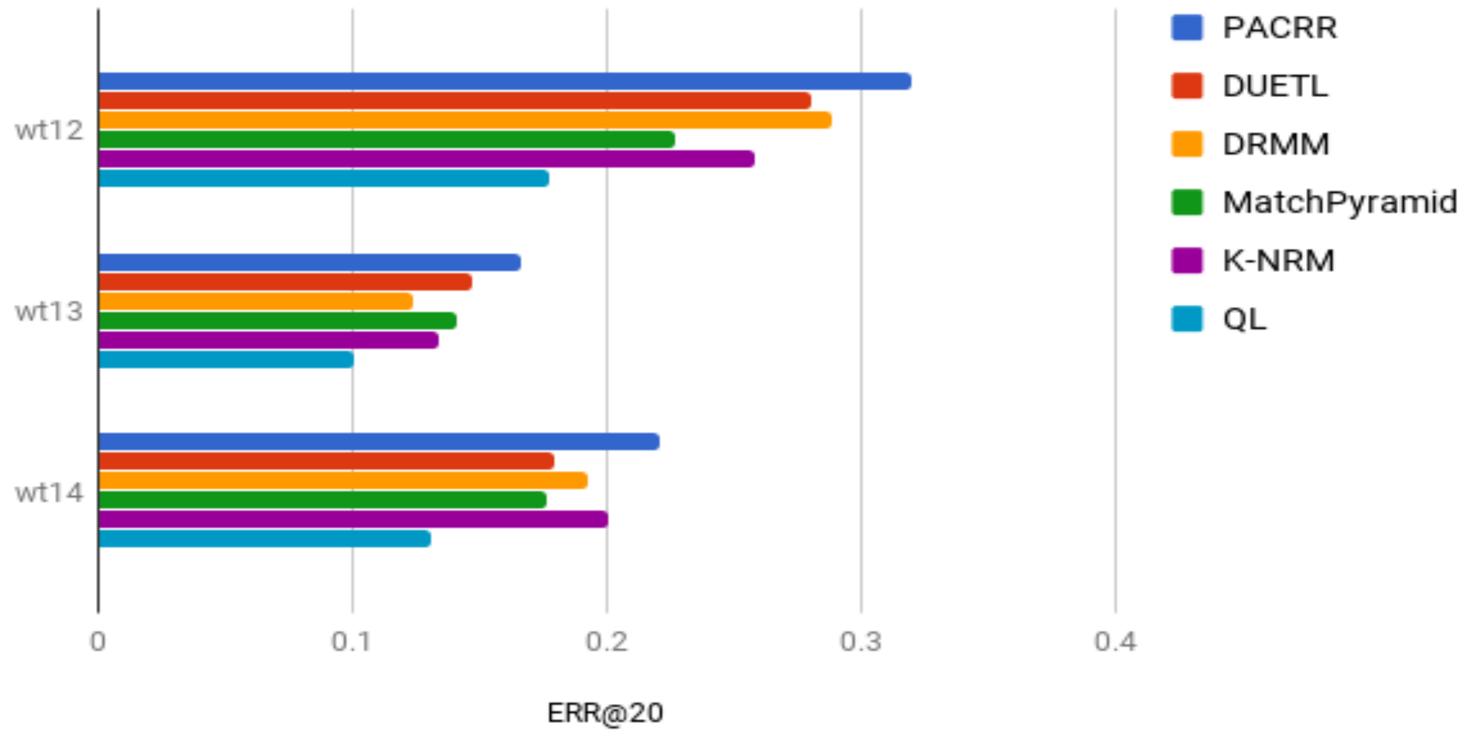
Result: **RerankSimple**

----How good a neural IR model can achieve by reranking QL baseline?

- ❑ The Neural IR model is employed as a re-ranker, making improvements by re-ranking top-k (e.g., top-30) search results from initial ranker
- ❑ Initial ranker can access the whole collection of documents
- ❑ Re-rank search results from a simple ranker, namely, query-likelihood model (QL)

Result: **RerankSimple**

----How good a neural IR model can achieve by reranking QL baseline?



- All neural IR models can improve based on QL search results .
- PACRR can achieve top-3 by solely re-ranking the search results from query-likelihood model.

Result: **PairAccuracy**

----How many doc pairs a neural IR model can rank correctly?

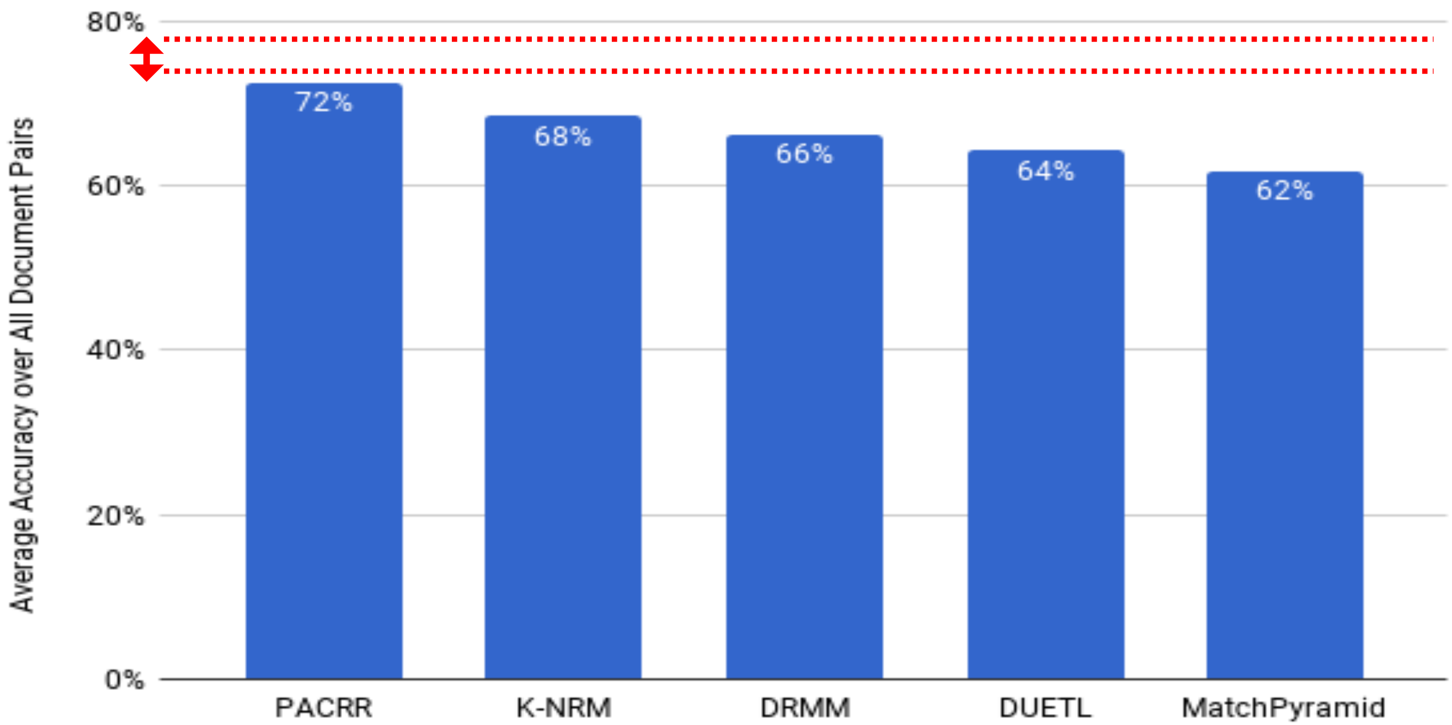
- Evaluate on pairwise ranking benchmark. Given (q, d_1, d_2),

  d_1 is more relevant or d_2 is more relevant?

- Cover all document pairs that are being predicted
- Calculate the accuracy: the ratio of the concordant pairs

Result: **PairAccuracy**

----How many doc pairs a neural IR model can rank correctly?



- The average accuracy for PACRR among different label pairs is 72%
- As reference, human accessors agree with each other by **74-77%** according to literature

Reference

[1] Pang, Liang, Lan, Yanyan, Guo, Jiafeng, Xu, Jun, and Cheng, Xueqi . "A Study of MatchPyramid Models on Ad-hoc Retrieval." In: Proceedings of the Neu-IR 2016 SIGIR Workshop on Neural Information Retrieval. Neu-IR '16

[2] Guo, Jiafeng, Fan, Yixing, Ai, Qingyao, and Croft, W. Bruce (2016). "A deep relevance matching model for ad-hoc retrieval." In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. CIKM '16

[3] Mitra, Bhaskar, Diaz, Fernando, and Craswell, Nick . "Learning to Match using Local and Distributed Representations of Text for Web Search." In: Proceedings of the 26th International Conference on World Wide Web. WWW '16

[4] Xiong, Chenyan, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. "End-to-end neural ad-hoc ranking with kernel pooling." In: Proceedings of the 40th annual international ACM SIGIR conference on Research and development in information retrieval. SIGIR '17

[5] Hui, Kai, Yates, Andrew, Berberich, Klaus, and Melo, Gerard de. "Position-Aware Representations for Relevance Matching in Neural Information Retrieval." In: Proceedings of the 26th International Conference on World Wide Web Companion. WWW '17

max planck institut
informatik

Thank You!

code: https://github.com/khui/repacrr

contact: khui@mpi-inf.mpg.de

MAX-PLANCK-GESELLSCHAFT