MPII at the NTCIR-14 WWW-2 Task

Andrew Yates

Max Planck Institute for Informatics

MAX-PLANCK-GESELLSCHAFT

Motivation

Opportunity to evaluate NIR model (participating in pool)

- Previously evaluated on TREC Web Track 09-14 (WSDM '18, EMNLP '17)
- With long queries (TREC description)
- Re-ranking results from unsupervised model

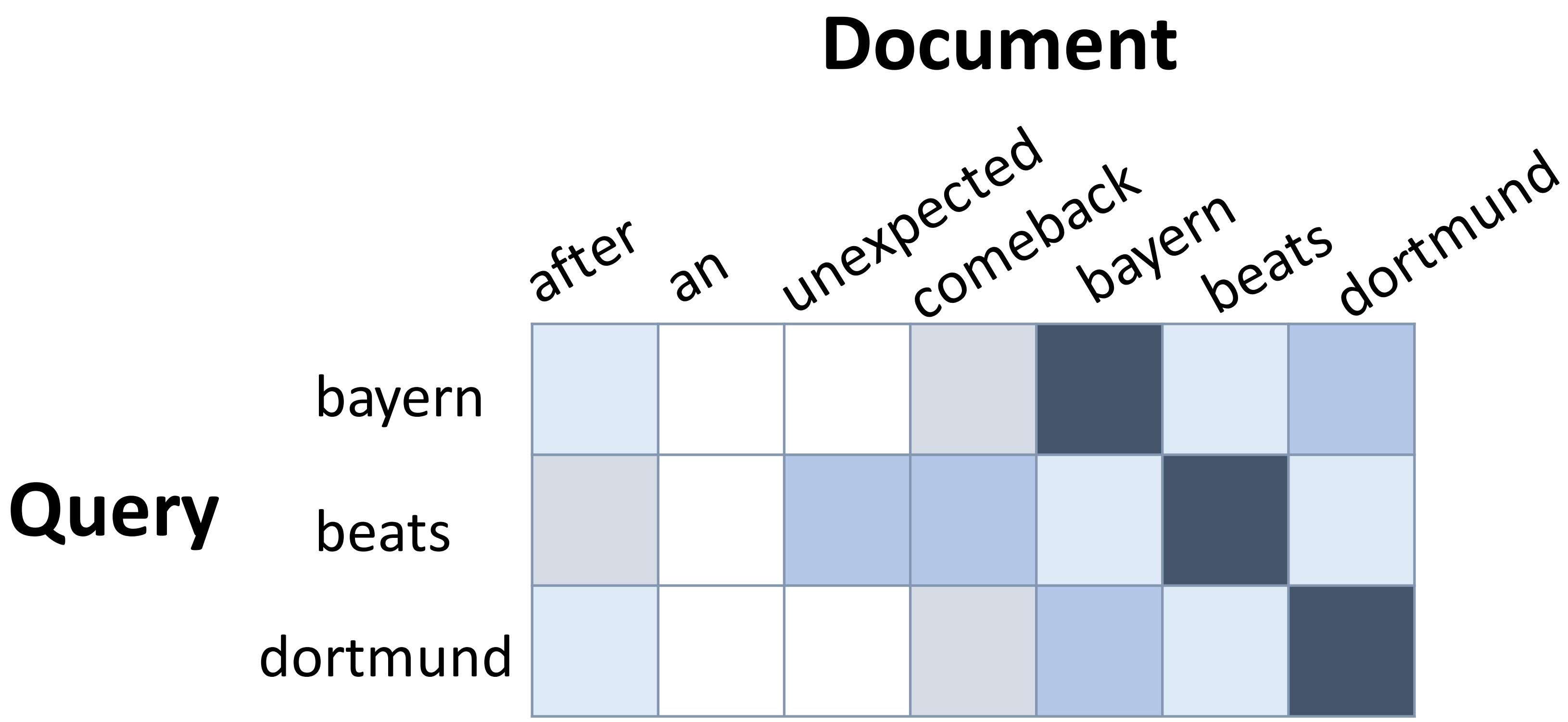
Significant improvement with a strong signal from WSDM '18?

How does it compare to BM25 with short queries (& pool)?

Outline

- Model summary (PACRR & Co-PACRR)
- Parameters varied
- Experimental setup
- Results

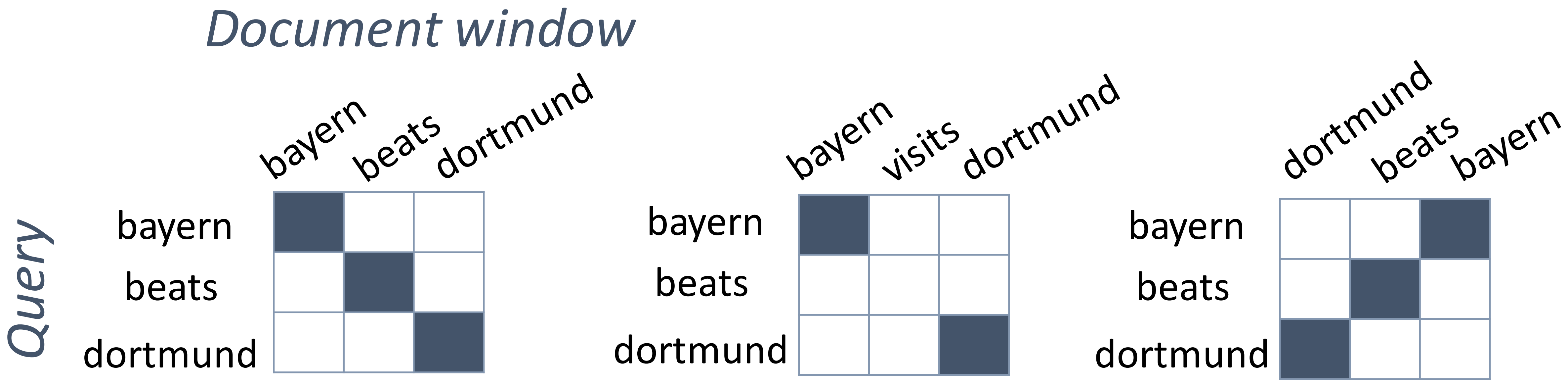
Input Representation



Query-document similarity matrix

- word2vec similarity
- One matrix for each document

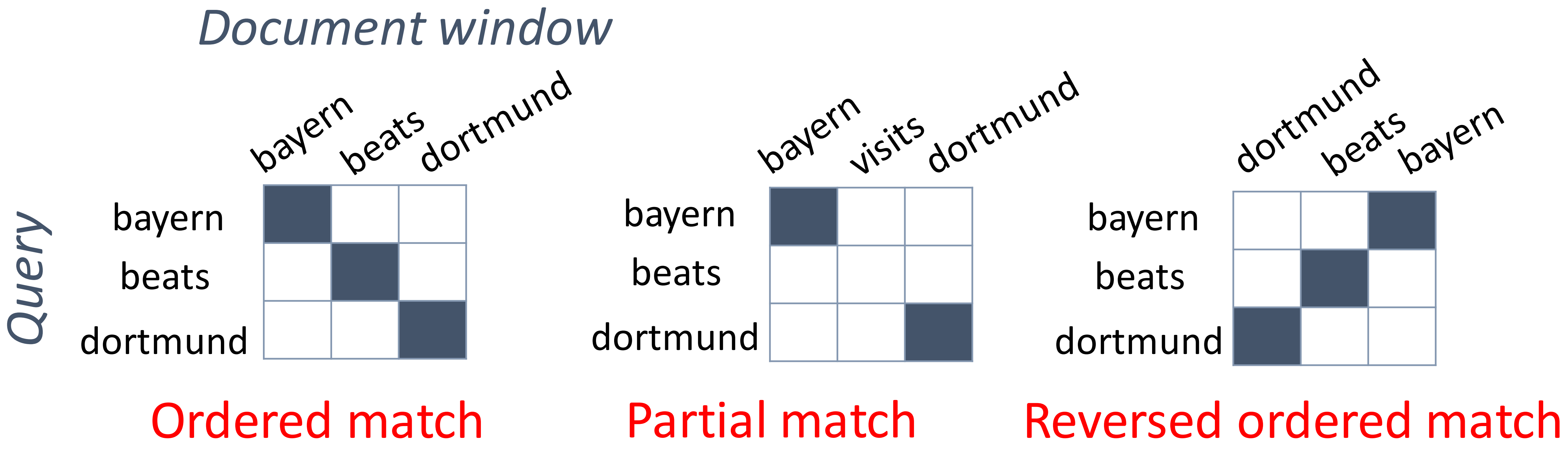
Using Positional Information



Match patterns (Convolutional kernels)

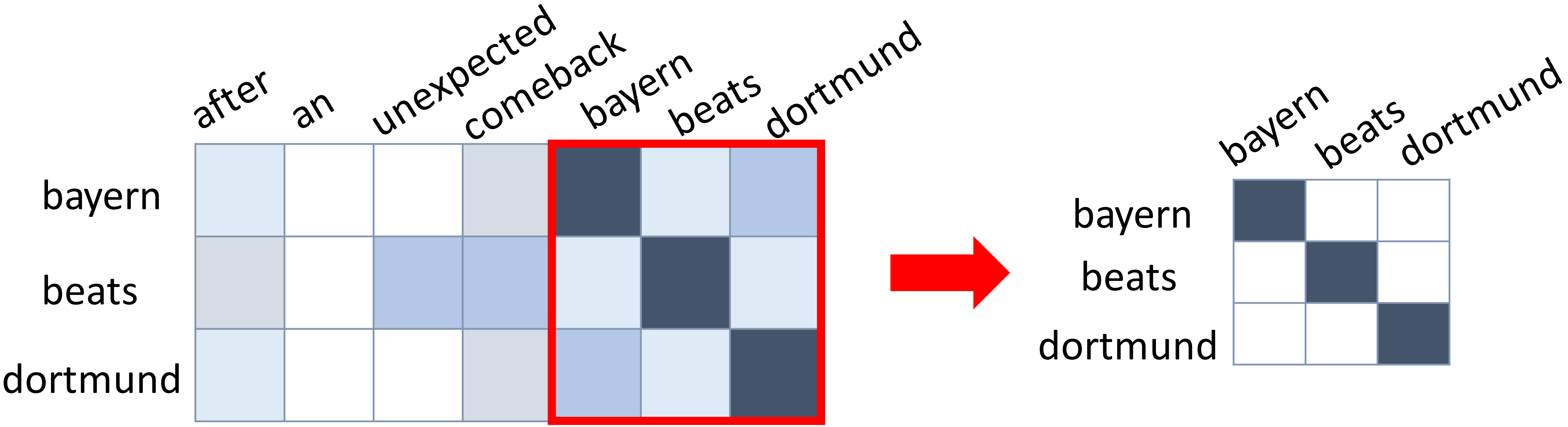
PACRR: A Position-Aware Neural IR Model for Relevance Matching.
K Hui, A Yates, K Berberich, G de Melo. In: EMNLP '17.

Using Positional Information



PACRR: A Position-Aware Neural IR Model for Relevance Matching.
K Hui, A Yates, K Berberich, G de Melo. In: EMNLP '17.

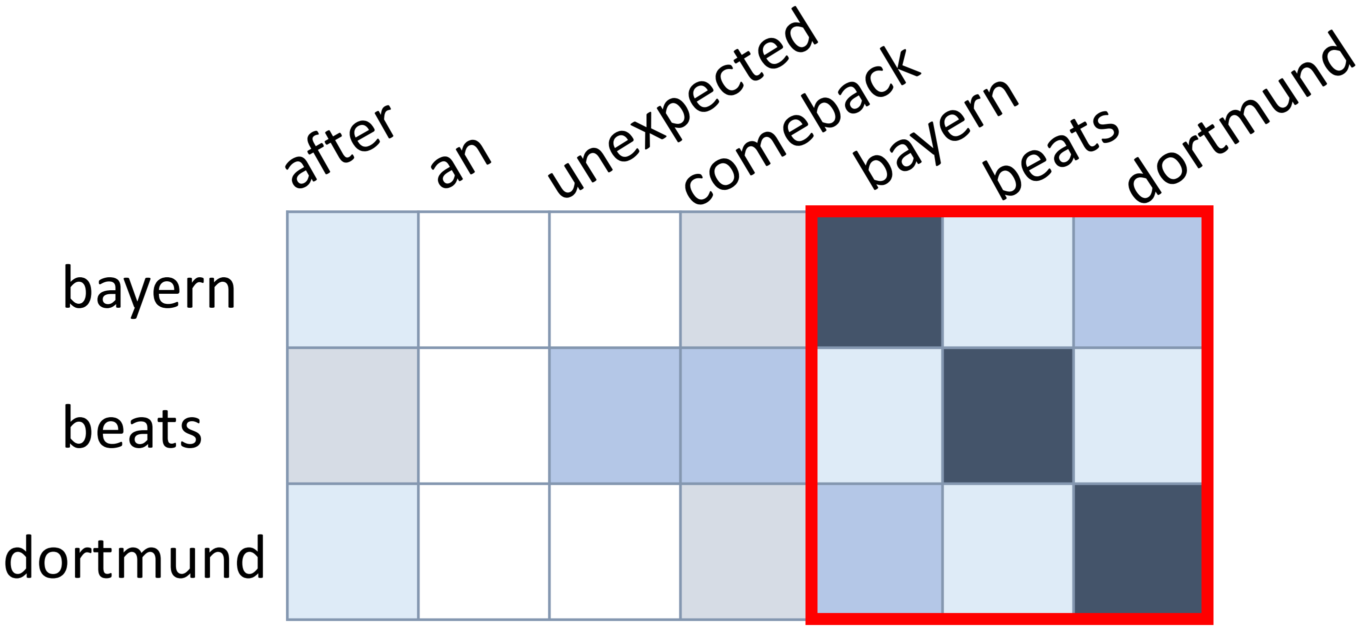
Using Positional Information

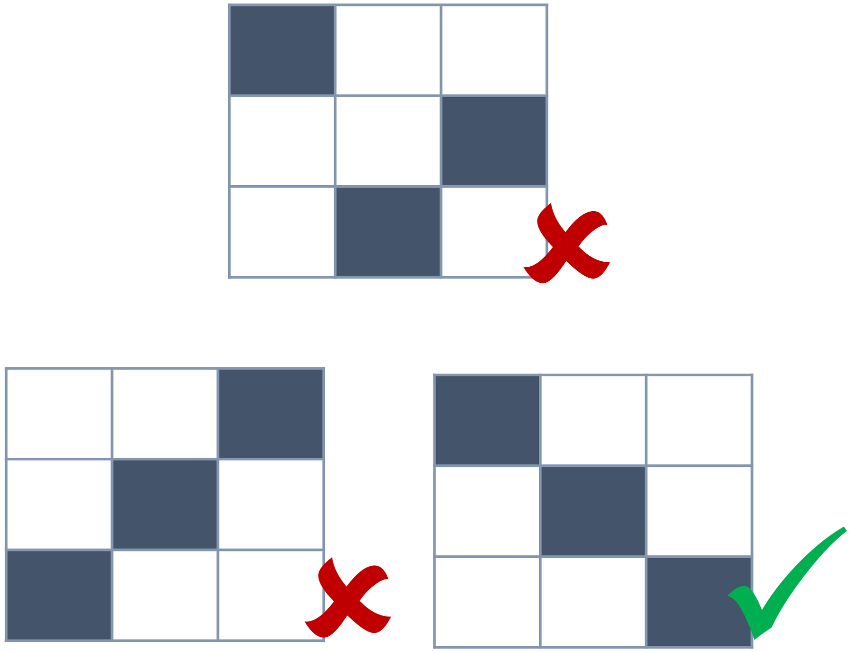


Matches are local: consider *NxN* regions of the matrix

PACRR: A Position-Aware Neural IR Model for Relevance Matching.
K Hui, A Yates, K Berberich, G de Melo. In: EMNLP '17.

Using Positional Information



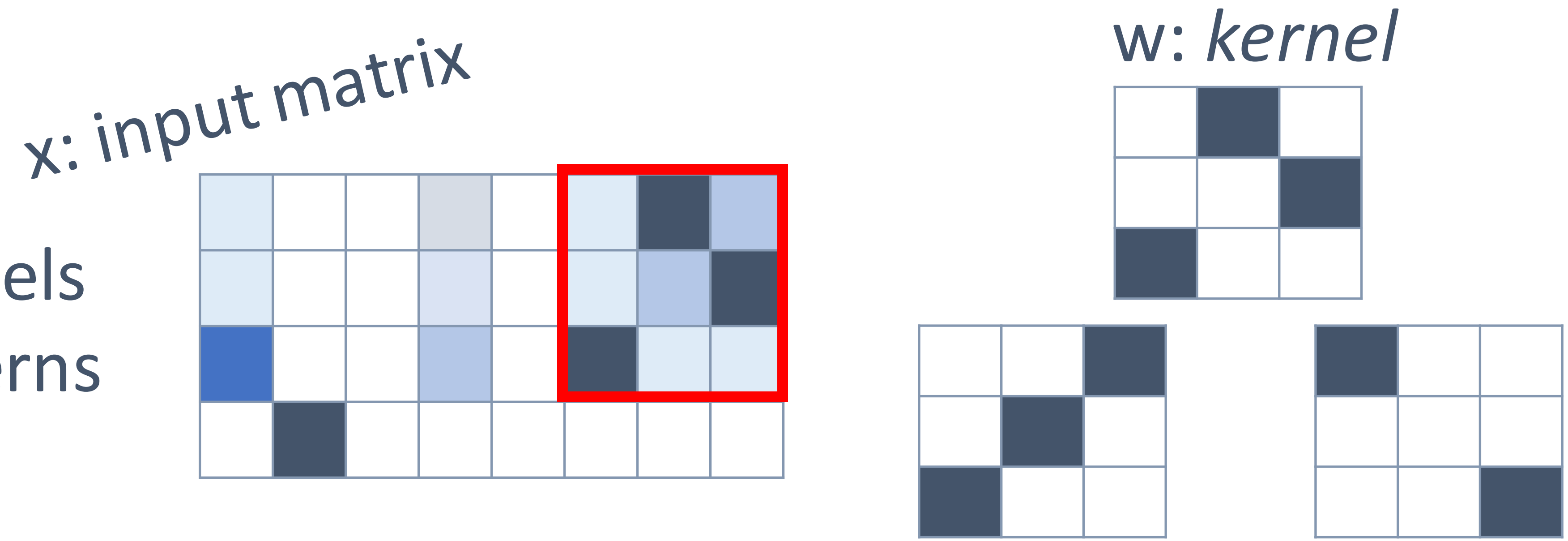


Patterns are exclusive: each region is best matched by a single pattern

PACRR: A Position-Aware Neural IR Model for Relevance Matching.
K Hui, A Yates, K Berberich, G de Melo. In: EMNLP '17.

PACRR: Position-Aware Convolutional Recurrent Relevance Matching

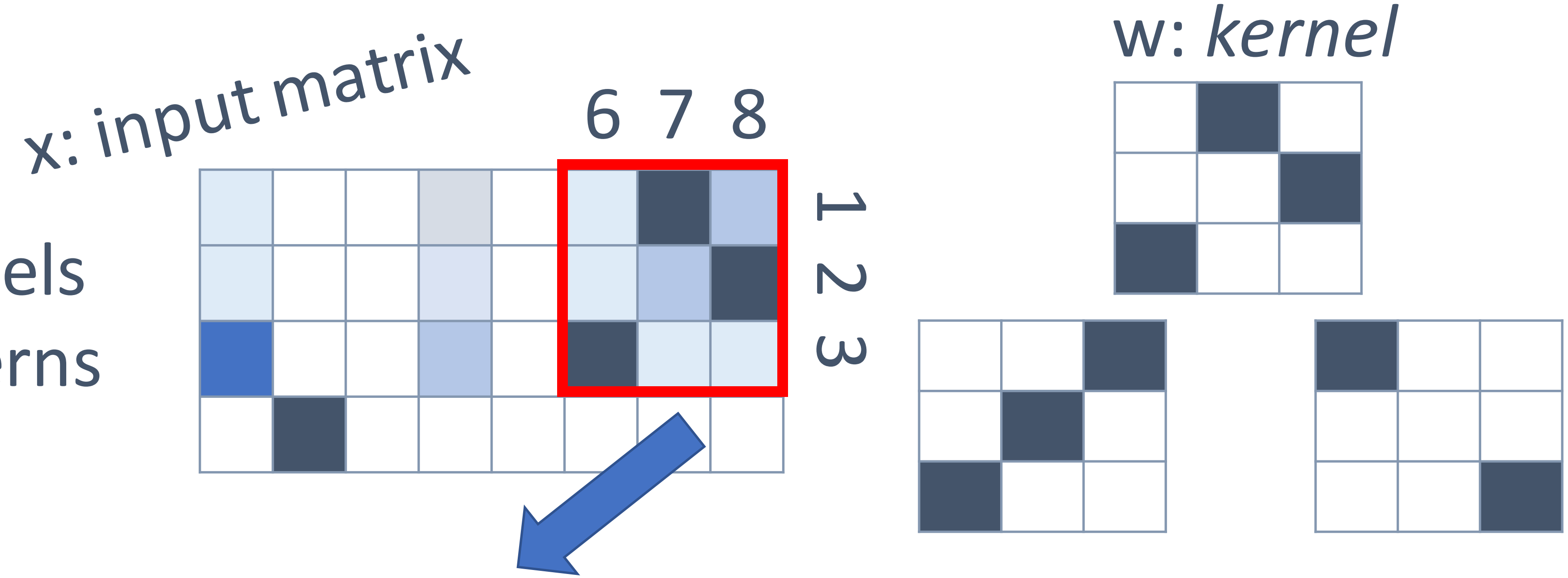
(1) CNN kernels capture patterns



PACRR: A Position-Aware Neural IR Model for Relevance Matching.
K Hui, A Yates, K Berberich, G de Melo. In: EMNLP '17.

PACRR: Position-Aware Convolutional Recurrent Relevance Matching

(1) CNN kernels capture patterns

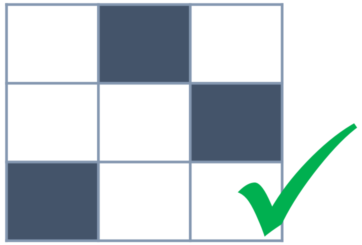


Signal for this region:

$$w_{1,1}\, x_{1,6} + w_{1,2}\, x_{1,7} + w_{1,3}\, x_{1,8} + \ldots + w_{2,1}\, x_{2,6} + \ldots\, w_{3,3}\, x_{3,8}$$

PACRR: Position-Aware Convolutional Recurrent Relevance Matching

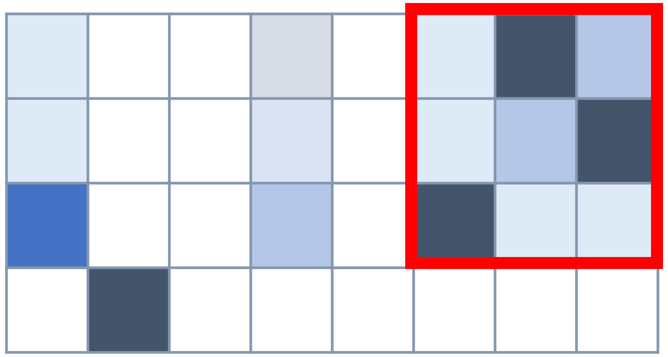
Best-matching pattern

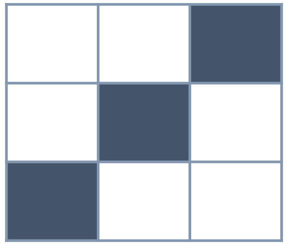


Signal: 1.0

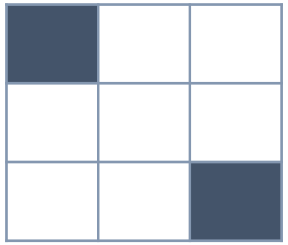
(1) CNN kernels capture patterns

(2) Max pool kernels





Signal: 0.3



Signal: 0

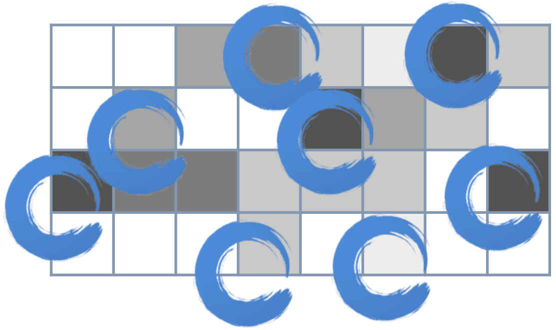
PACRR: Position-Aware Convolutional Recurrent Relevance Matching

(1) CNN kernels capture patterns

(2) Max pool kernels

(3) K-max pool query signals from doc regions

K=2



PACRR: Position-Aware Convolutional Recurrent Relevance Matching

(1) CNN kernels capture patterns

(2) Max pool kernels

(3) K-max pool query signals from doc regions

For each query term, we now have:

- K-max match signals for unigrams
- K-max match signals for bigrams
- …
- K-max match signals for n-grams

PACRR: Position-Aware Convolutional Recurrent Relevance Matching

max planck institut informatik

(1) CNN kernels capture patterns

(2) Max pool kernels

(3) K-max pool query signals from doc regions

(4) Combination function (FC layers) produce a score for each query term

(5) Document score is the summation

[Steps 4 & 5 differ from original papers]

PACRR: Position-A
Recurrent Relevar

Related to MatchPyramid, but
e.g., different pooling strategies

A Study of MatchPyramid Models on Ad-hoc Retrieval. L. Pang, Y. Lan, J. Guo, J. Xu, Z. Cheng. Neu-IR '16 SIGIR Workshop.

(1) CNN kernels capture patterns

(2) Max poo kernels

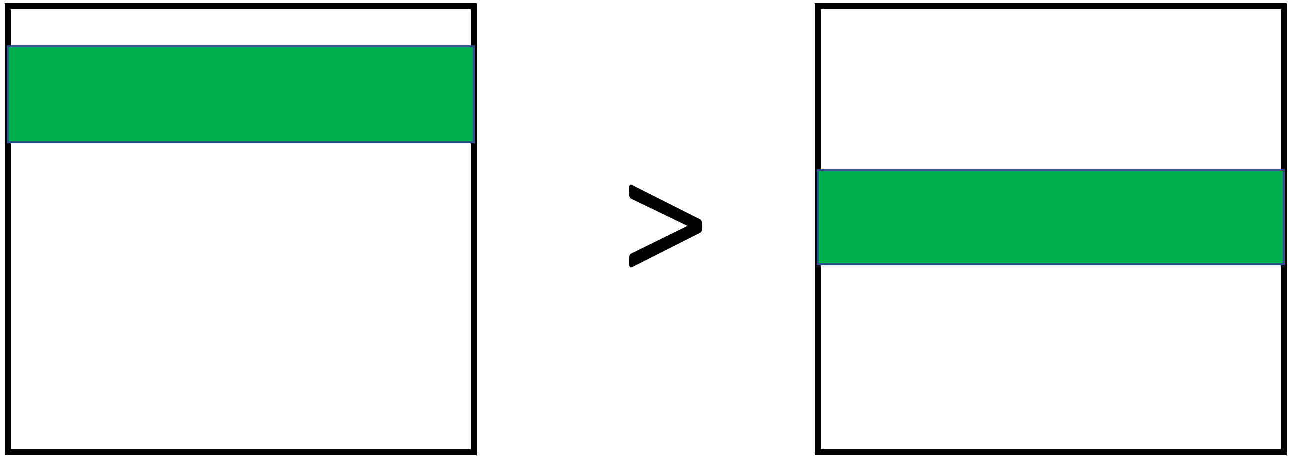
(3) K-max pool query signals from doc regions

(4) Combination function (FC layers) produce a score for each query term

(5) Document score is the summation
[Steps 4 & 5 differ from original papers]

Variant: Cascade Pooling

- Inspired by cascade model
  An experimental comparison of click position-bias models. Craswell et al. WSDM '08.
- Prefer document with earlier relevant information
- One of several improvements in Co-PACRR (WSDM '18)



Document A *Document B*

Variant: Cascade Pooling

For each query term, PACRR retains top *k* match signals

- Cascade Pooling: repeat for different document cutoffs
- Top *k* signals from the first 50% of the document
- Top *k* signals from the entire document

Query term FC receives match signals from different cutoffs

Co-PACRR: A Context-Aware Neural IR Model for Ad-hoc Retrieval.
K Hui, A Yates, K Berberich, G de Melo. In: WSDM '18.

Parameters Varied

1. Cascade pooling used? (3 with, 2 without)
2. Size of *k*-max pooling (top 5 vs. 15)
3. Size of fully connected layers that score query term (2x8 or 1)

Experimental Setup

- Train on TREC WT09-13 judgments
- WT14 and WWW-1 used for validation
- Using best weights on WWW-1 (after sanity checking on WT14), re-rank BM25 run provided by organizers

Results & Conclusion

- No significant improvement between any pair of runs
- No significant improvement over BM25
- Given past results, minD >= 0.1 seems large

Results & Conclusion

- No significant improvement between any pair of runs
- No significant improvement over BM25
- Given past results, minD >= 0.1 seems large

Recent work building on PACRR (and other NIR models):

CEDR: Contextualized Embeddings for Document Ranking.

S. MacAvaney, A. Yates, A. Cohan, N. Goharian. SIGIR '19.

Thanks!