THE 14TH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING

wsdm 2021

Tutorial
Morning (March, 8)
9:30-12:30, GMT+2

Beyond Probability Ranking Principle: Modeling the Dependencies among Documents

Liang Pang[1], Qingyao Ai[2], Jun Xu[3]

1. CAS Key Lab of Network Data Science and Technology
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
2. The University of Utah, USA
3. Renmin University of China, Beijing, China

1. Introduction

1.1 The Ranking Problem

1.2 Organization of the tutorial

The Ranking Problem

Information Quota Increasing

Screen Size Decreasing

KB

m^2

No Ranking

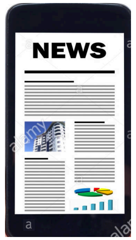
MB

Ranking with Index

dm^2

GB

Ranking with Query



cm^2

Ranking with Personalization

PB

Ranking with …

mm^2

The Ranking Application

Schedule

Half Day Tutorial 9:30 – 12:30	Time	Speaker
1. Introduction	9:30 – 10:10	Qingyao Ai
2. Ranking with Probability Ranking Principle (PRP)		
3. Limitations of PRP Principle		
5min Break		
4. Ranking with Sequential Dependency	10:15 – 11:05	Jun Xu
4.1 Heuristic Sequential Ranking Models		
4.2 Learning Sequential Ranking Models		
4.3 Challenges		
5min Break		
5. Ranking with Global Dependency	11:10 – 12:20	Liang Pang
5.1 List Inputted Global Ranking Models		
5.2 Set Inputted Global Ranking Models		
6. Conclusion		
7. Q&A and Open Discussions	12:20– 12:30	Liang Pang Qingyao Ai Jun Xu

2. Ranking with Probability Ranking Principle (PRP)

2.1 Feature based Ranking Methods

2.2 Neural IR Models

Similarity based Models

- Boolean Models
  - Relational algebra with algebraic expressions
  - Either fetch the document (1) or doesn’t fetch the document (0), there is no methodology to rank them

- Vector Space Models
  - Address the problem of the documents being partially matched
  - Term Frequency - Inverse Document Frequency (tf-idf)
  - Calculating cosine value between query weight vector and document weight vector

Probabilistic Models

- Introduced by Maron and Kuhns in 1960 and further developed by Roberston and other researchers.
- **Relevance** is expressed in terms of probability.

- BM25 Model
  - A bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document, regardless of their proximity within the document
- Language Model for IR
  - View each document as a language sample and estimate the probabilities of producing words
  - A query is treated as a generation process and retrieved documents are ranked based on P(Q|D)

Probability Ranking Principle (PRP)

Probability Ranking Principle

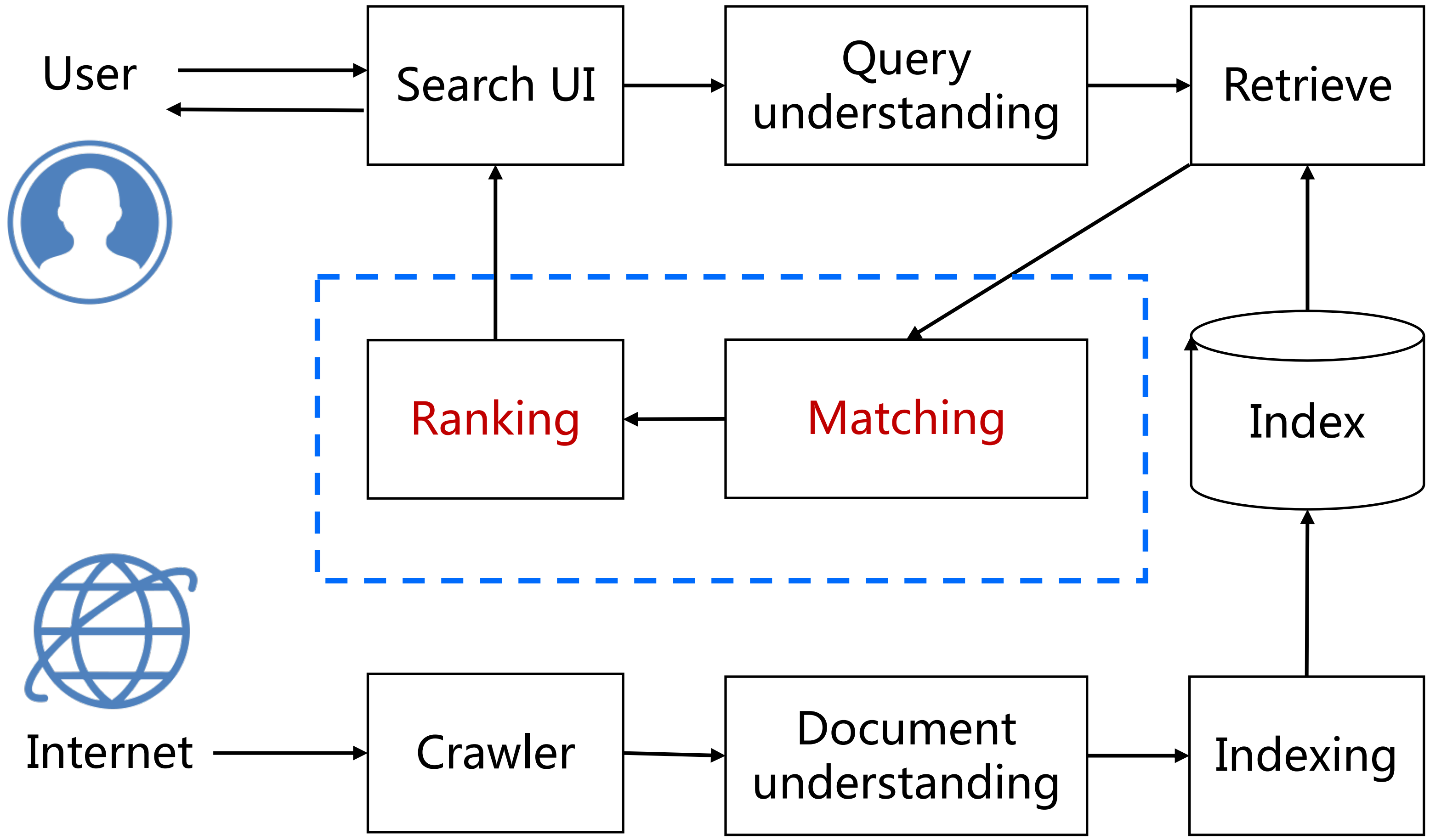
The principle that, for optimal retrieval, documents should be ranked in order of the probability of relevance or usefulness has been brought into question by Cooper.

--- Stephen E. Robertson

PRP assumes that each document has **a unique probability** to satisfy a particular information need. Therefore, the ranking scores of documents are **assigned separately** and are **independent to each other**.

Robertson, S.E. The Probability Ranking Principle in IR. *J. Documentation* (1977), 294304. Reprinted in *Readings in Information Retrieval*, K. Sparck Jones and P. Willett (eds), 1997, 281286.

Document Ranking under PRP



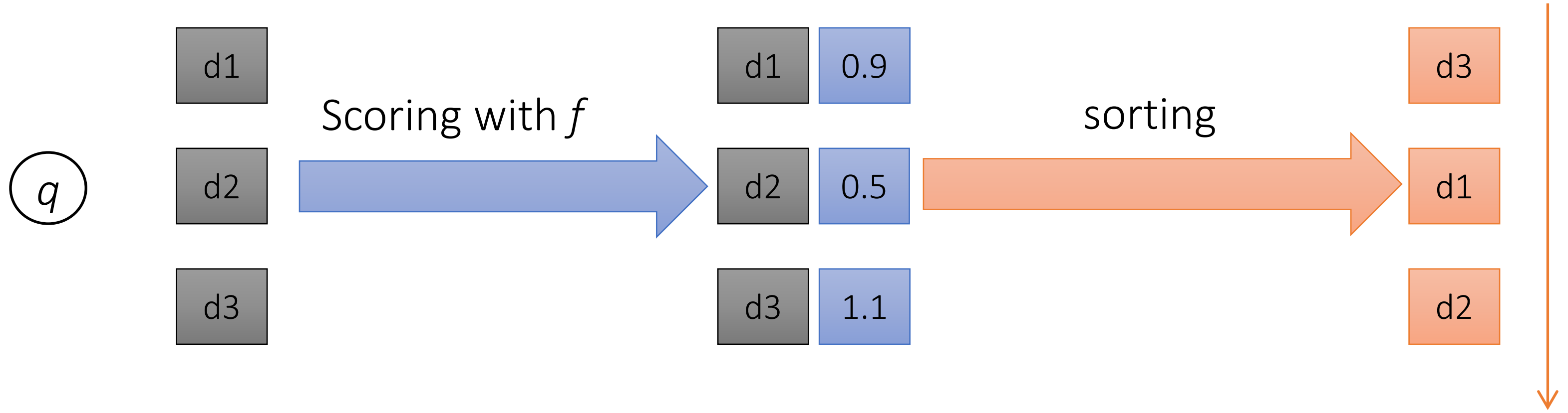
- Ranking criteria
  - Relevance
- Ranking models
  - Heuristics
    - BM25
    - LMIR
  - Learning to rank

2.1 Feature based Ranking Methods

- Heuristic Ranking Functions
- Learning to Rank Approaches

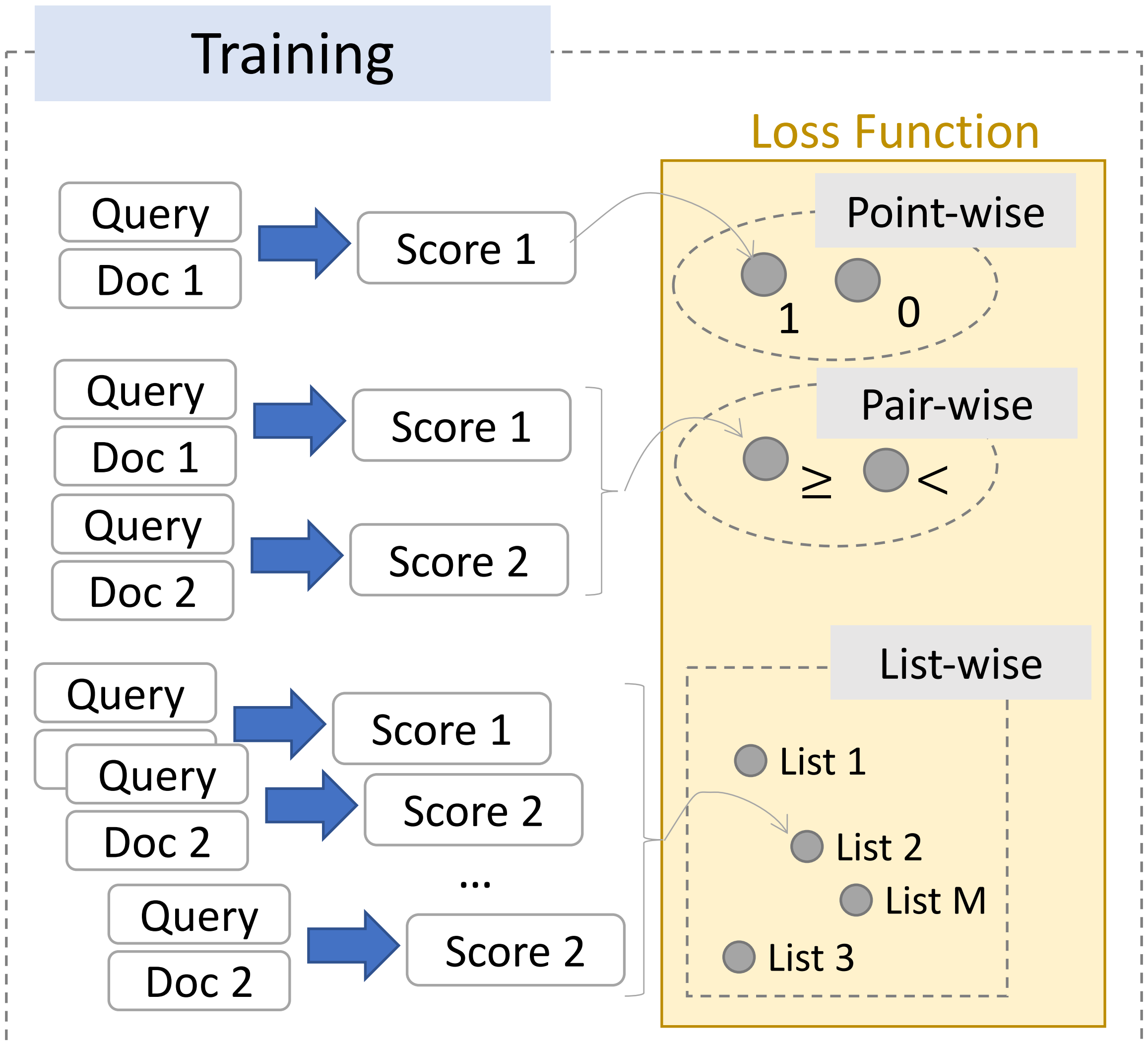
Ranking Model = Scoring + Sorting

Under Probability Ranking Principle …

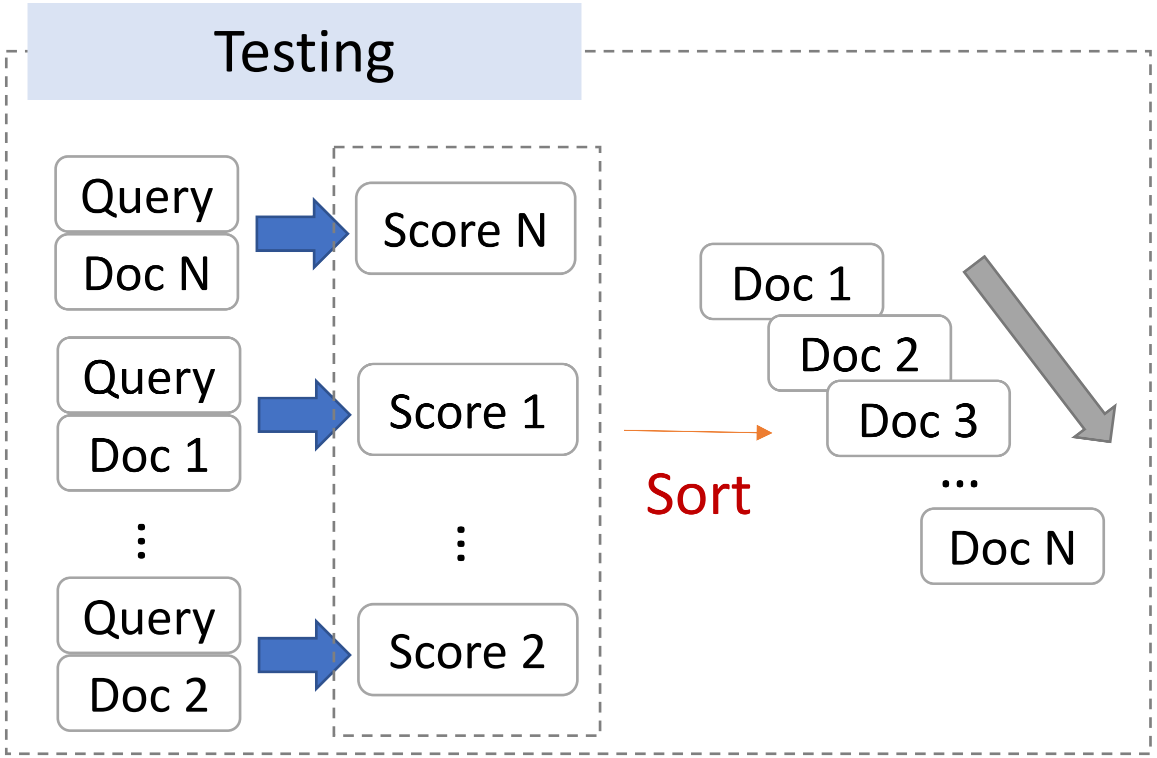


- Utility of a doc is independent of other docs
- Ranking as scoring + sorting
  - Each documents can be scored independently
  - Scores are independent of the rank

Standard Learning to Rank



Loss Function shapes the gaps between scores.



Uni-variate Scoring Function

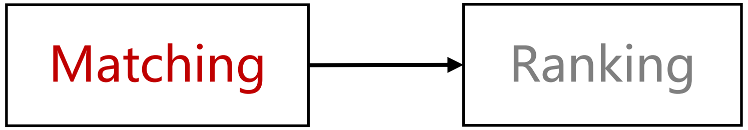
Model “ ➡ ” maps query-doc pair to a score.

$$f\left(\begin{matrix}\text{Query}\\ \text{Doc 1}\end{matrix}\right) \rightarrow \text{Score 1} \qquad f: \mathbb{D} \mapsto \mathbb{R}$$

2.2 Neural IR Models

- Methods of Representation Learning
- Methods of Matching Function Learning
- Methods of Combining

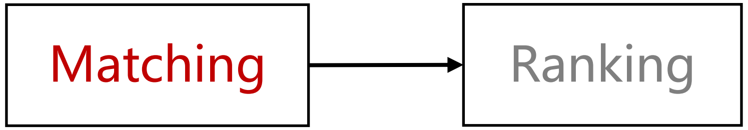
Biggest Challenge: Semantic Gap



- Same intent can be represented by different queries (representations)
- Search is still mainly based on term level matching
- Query document mismatch occurs, when searcher and author use different representations

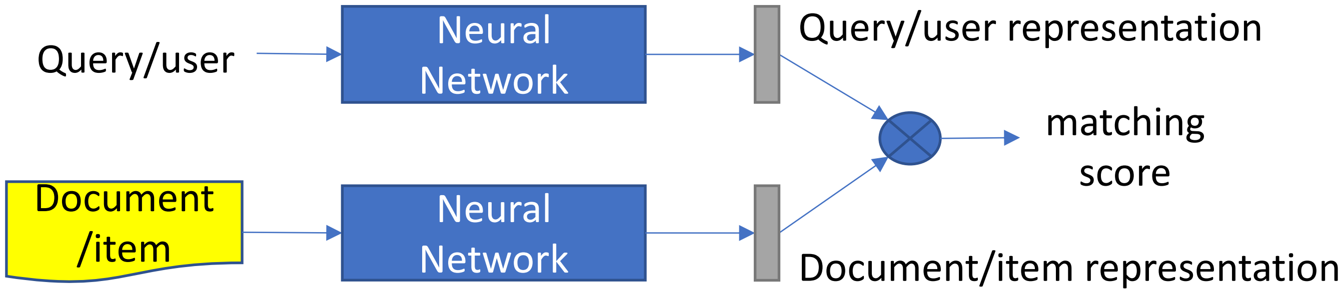
Deep Learning Paradigms for Matching

[Pang et al., AAAI ‘16; Wan et al., AAAI ‘16; Wan et al., IJCAI ’16]



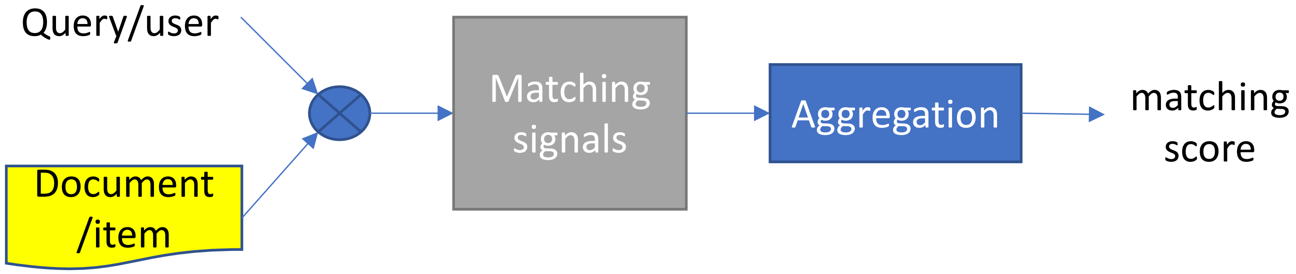
$$Match(q, d) = F(\phi(q), \phi(d))$$

- Methods of representation learning



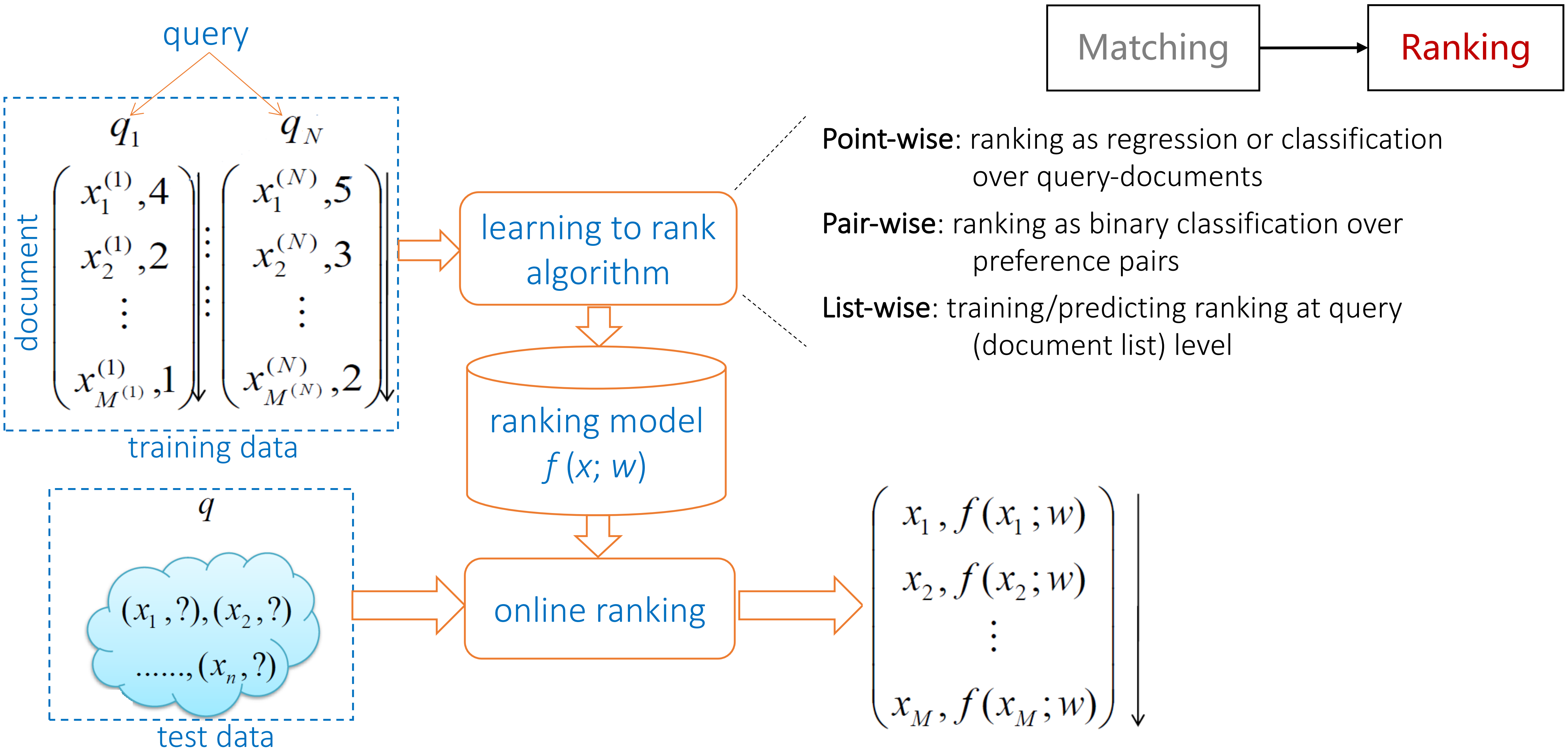
- Step 1: calculate representation $\phi(q)$
- Step 2: conduct matching $F\big(\phi(q), \phi(d)\big)$

- Methods of matching function learning



- Step 1: construct basic low-level matching signals
- Step 2: aggregate matching patterns

Learning to Rank Algorithm



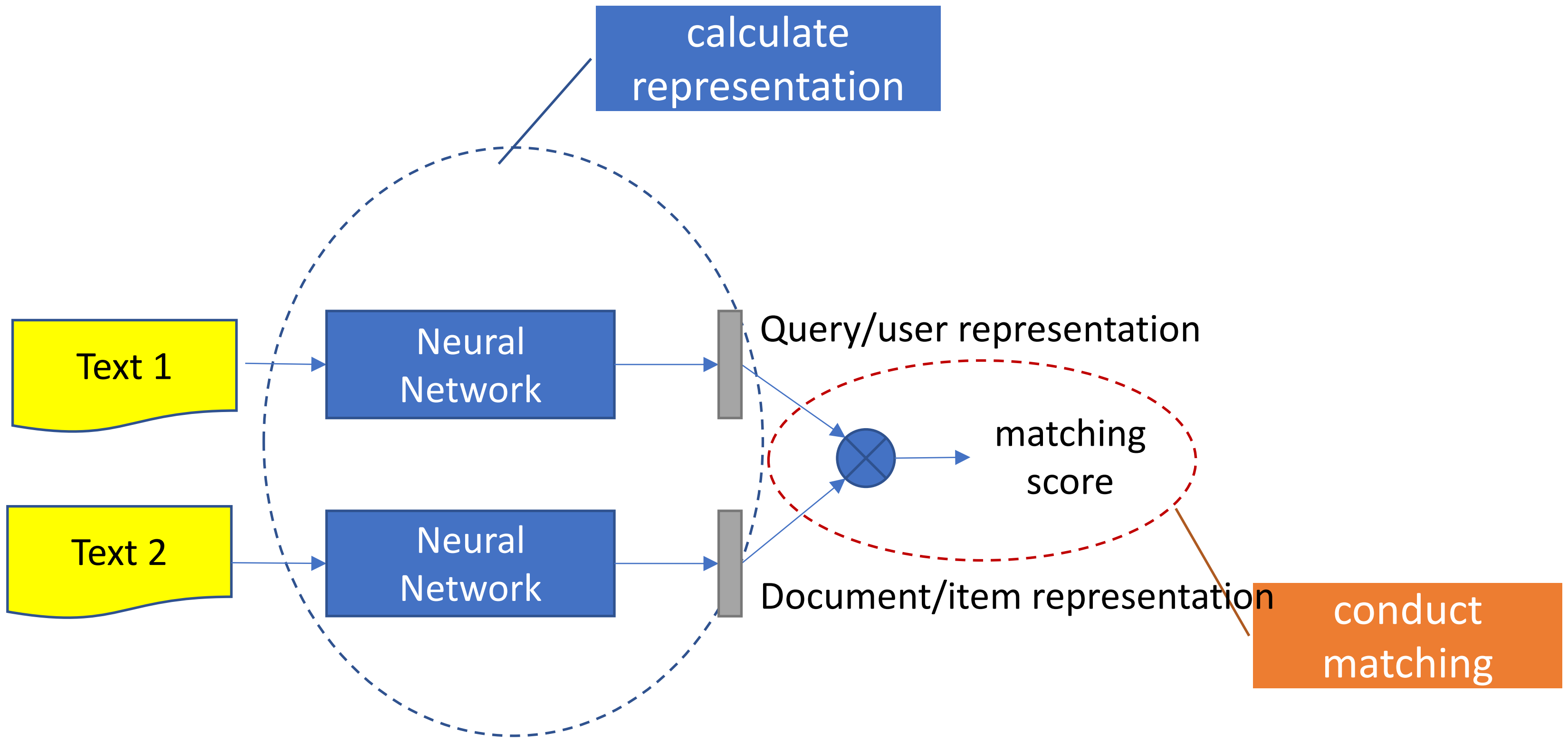
Point-wise: ranking as regression or classification over query-documents

Pair-wise: ranking as binary classification over preference pairs

List-wise: training/predicting ranking at query (document list) level

Methods of Representation Learning

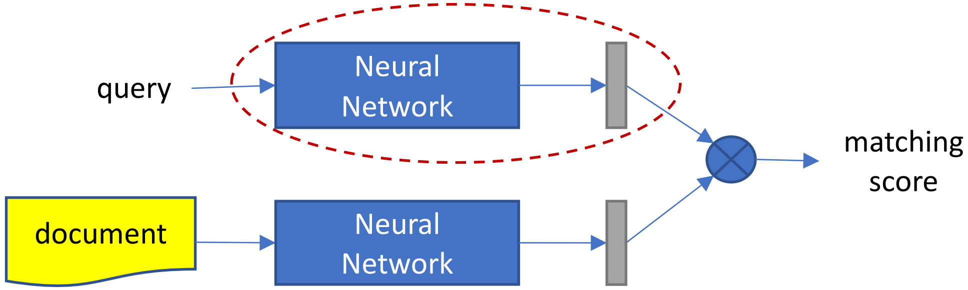
- Step 1: calculate representation $\phi(x)$
- Step 2: conduct matching $F\big(\phi(x), \phi(y)\big)$



Typical Methods of Representation Learning for Matching

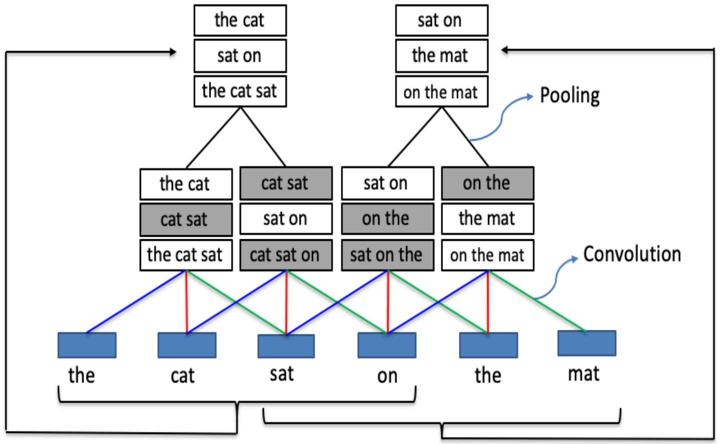
- Based on DNN
  - **DSSM**: Learning Deep Structured Semantic Models for Web Search using Click-through Data (Huang et al., CIKM ’13)
- Based on CNN
  - **CDSSM**: A latent semantic model with convolutional-pooling structure for information retrieval (Shen et al. CIKM ’14)
  - **ARC I**: Convolutional Neural Network Architectures for Matching Natural Language Sentences (Hu et al., NIPS ’14)
  - **CNTN**: Convolutional Neural Tensor Network Architecture for Community-Based Question Answering (Qiu and Huang, IJCAI ’15)
- Based on RNN
  - **LSTM-RNN**: Deep Sentence Embedding Using the Long Short Term Memory Network: Analysis and Application to Information Retrieval (Palangi et al., TASLP ’16)

Representation Learning

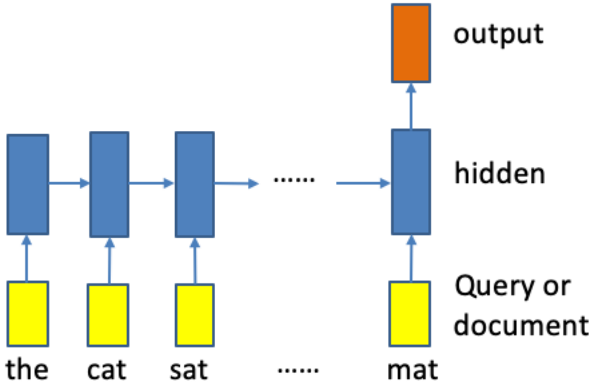


Modeling Order Information

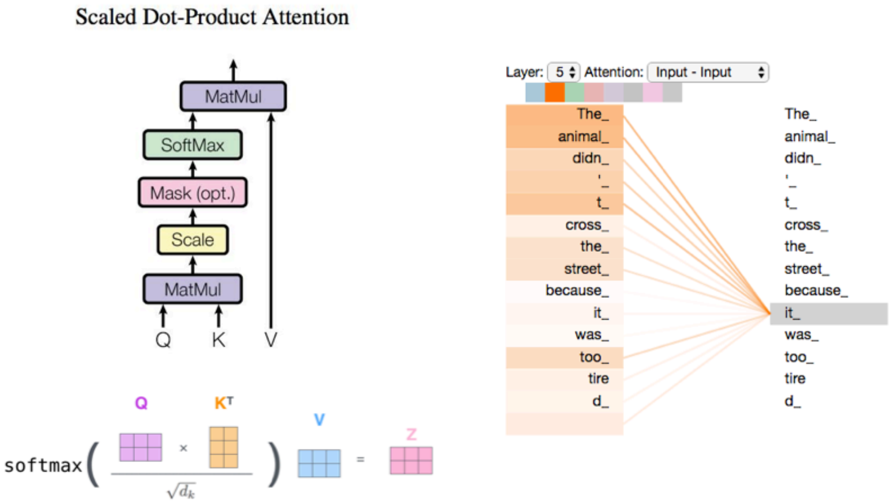
CNN



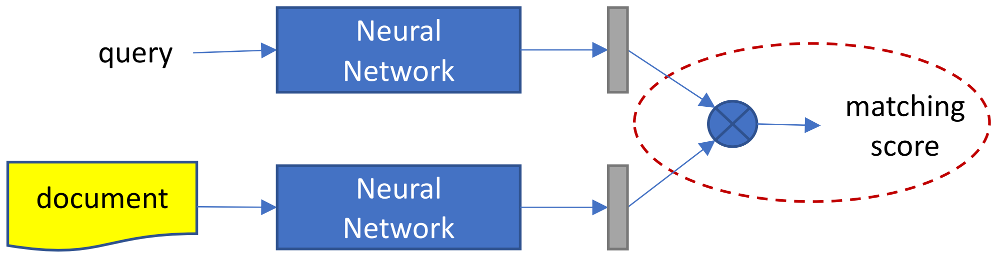
RNN



Self-Attention



Matching Function



- Given representations of query and document : q and d
- Similarity between these two representations:

  - Cosine Similarity (DSSM, CDSSM, RNN-LSTM)

$$s = \frac{q^T \cdot d}{|q| \cdot |d|}$$

  - Dot Product

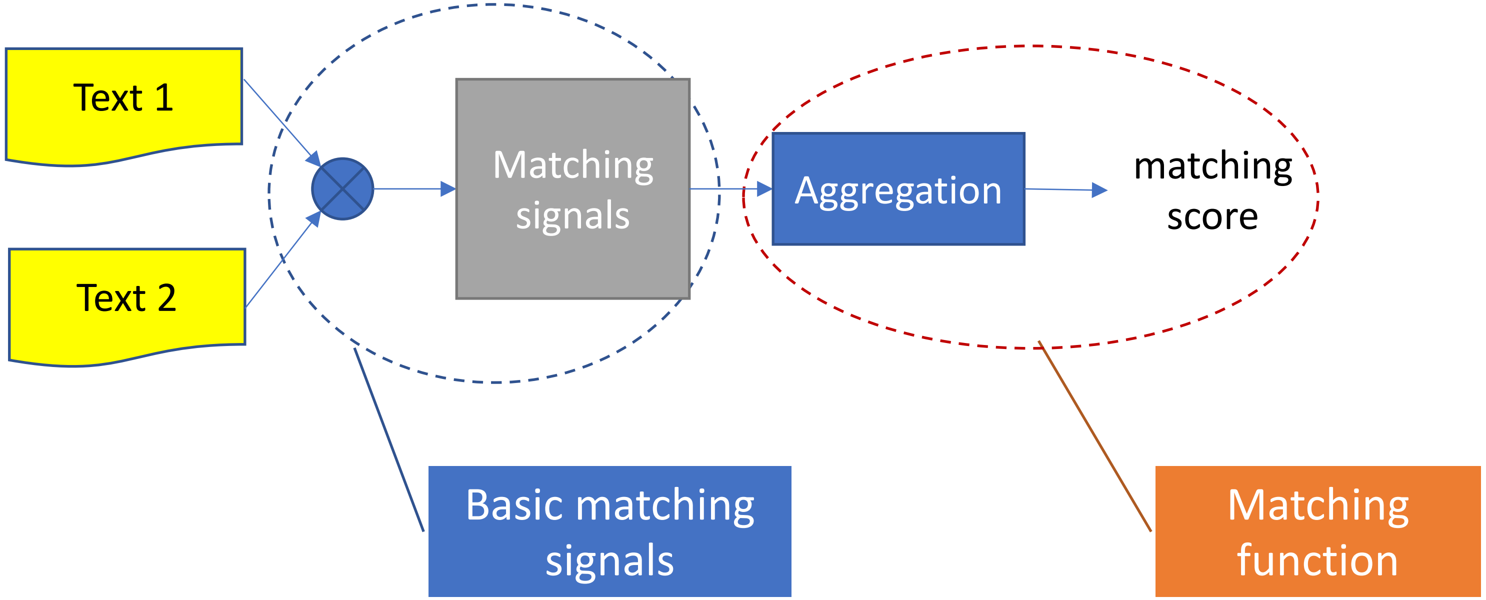
$$s = q^T \cdot d$$

  - Multi-Layer Perception (ARC-I)

$$s = W_2 \cdot \sigma \left(W_1 \cdot \begin{bmatrix} q \\ d \end{bmatrix} + b_1 \right) + b_2$$

Matching Function Learning

- Step 1: construct basic low-level matching signals
- Step 2: aggregate matching patterns



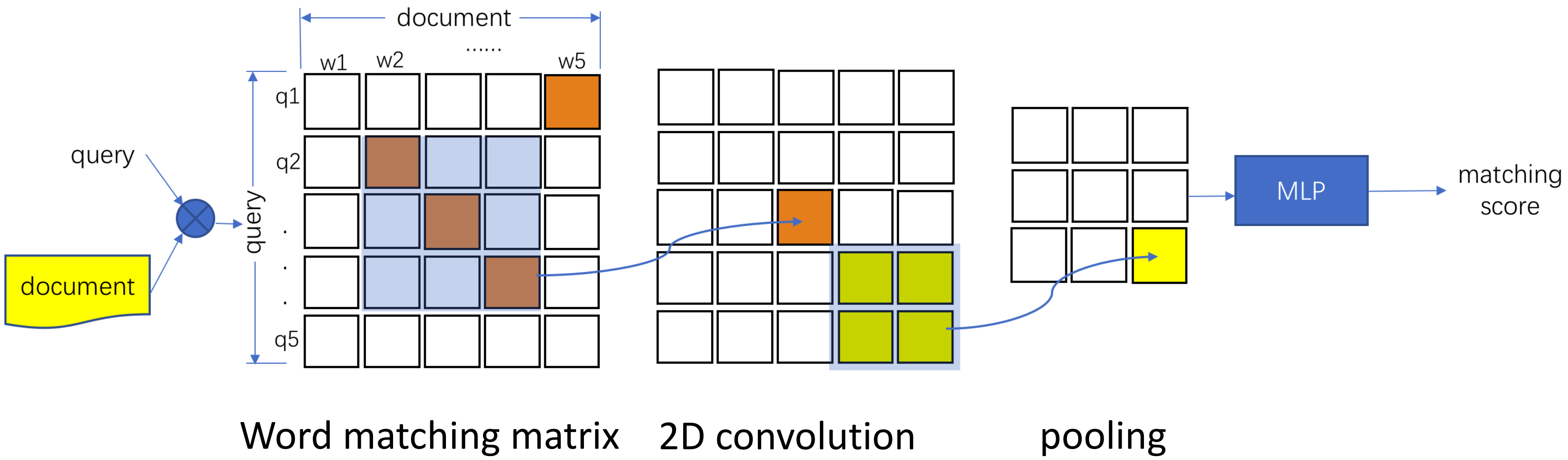
Typical Matching Function Learning Methods

- For short text (e.g., sentence) similarity matching
  - ARC II (Hu et al., NIPS ’14)
  - MatchPyramid (Pang et al., AAAI ’16)
  - Match-SRNN (Wan et al., IJCAI ’16)

- For query-document relevance matching
  - DRMM (Guo et al., CIKM ’16) and aNMM (Yang et al., CIKM ‘16)
  - K-NRM (Xiong et al., SIGIR ‘17) and Conv-KNRM (Dai et al., WSDM ‘18)
  - DeepRank (Pang et al., CIKM ‘17) HiNT (Fan et al., SIGIR ‘18) and PACRR (Hui et al., EMNLP ’17)

MatchPyramid (Pang et al., AAAI '16)

- Inspired by image recognition
- Basic matching signals: word-level matching matrix
- Matching function: 2D convolution + MDP



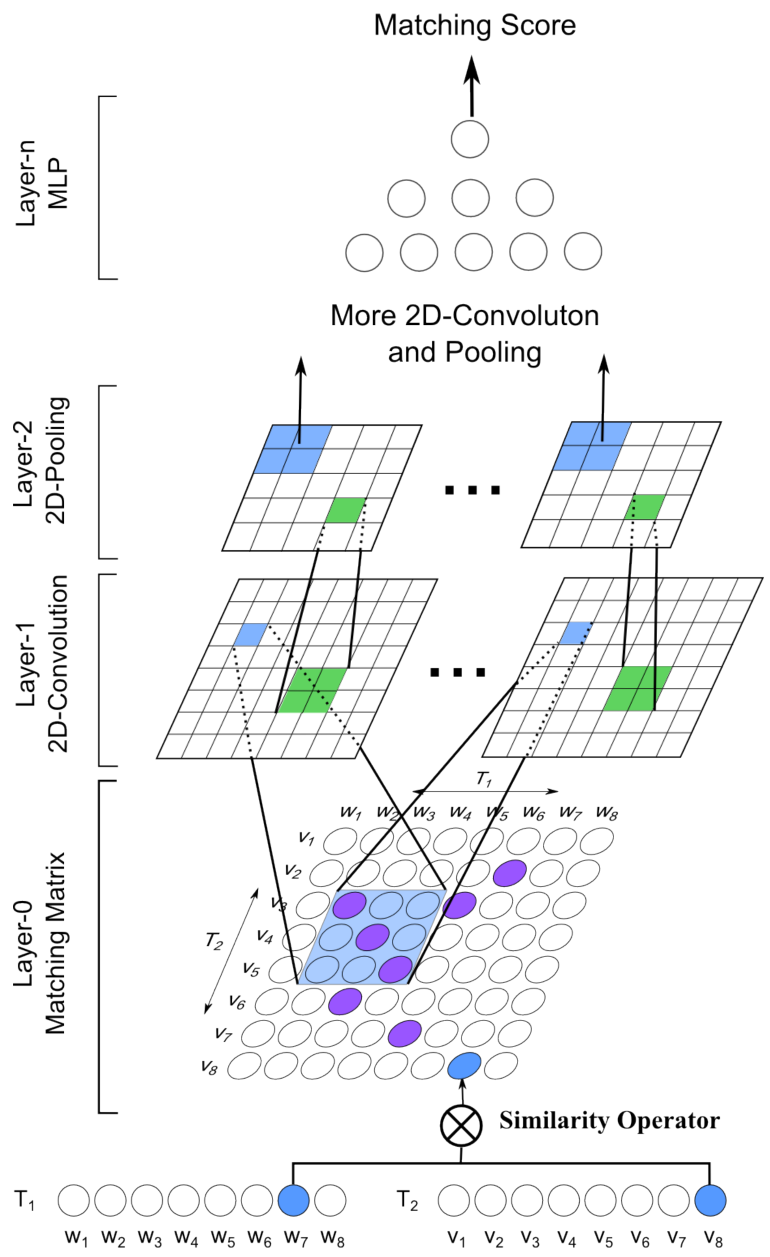
MatchPyramid

Hierarchical Convolution

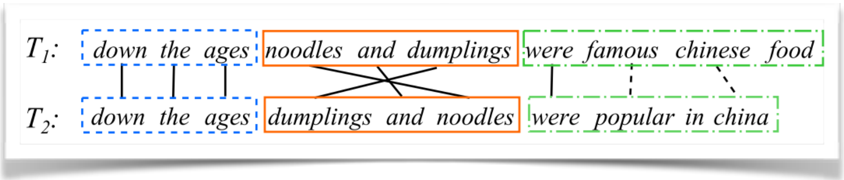
A Way to Capture Rich Matching Patterns

Matching Matrix

Bridging the Gap between Text Matching and Image Recognition



Similarity ≠ Relevance



- Similarity matching
  - Whether two sentences are semantically similar
  - Homogeneous texts with comparable lengths
  - Matches at all positions of both sentences
  - Symmetric matching function
  - Representative task: Paraphrase Identification

- Relevance matching
- Whether a document is relevant to a query
- Heterogeneous texts (keywords query, document) and very different in lengths
- Matches in different parts of documents
- Asymmetric matching function
- Representative task: ad-hoc retrieval

Relevance Matching

- Global Distribution of Matching Signals
  - DRMM (Guo et al., CIKM ’16) and aNMM (Yang et al., CIKM ‘16)
  - K-NRM (Xiong et al., SIGIR ‘17) and Conv-KNRM (Dai et al., WSDM ‘18)

- Local Context of Matching Positions
  - DeepRank (Pang et al., CIKM ‘17), HiNT (Fan et al., SIGIR ‘17) and PACRR (Hui et al., EMNLP ’17)

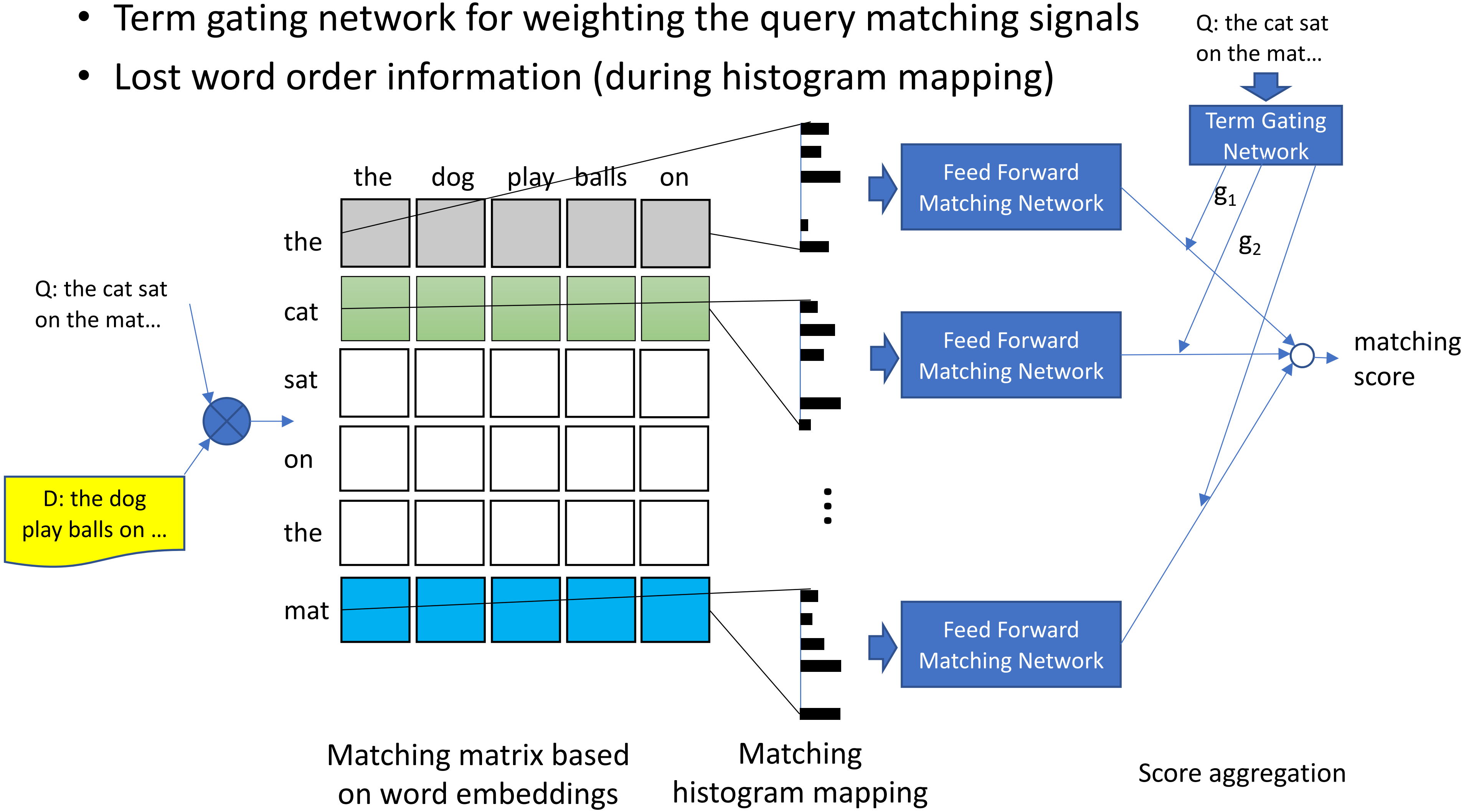
Global Distribution of Matching Signals

- Step 1: calculate matching signals for each query term
- Step 2: statistic each query term’s matching signal distributions
- Step 3: aggregate the distributions

- Pros
  - Matching between short query text and long document text
  - Robust: matching signals from irrelevant document words
- Cons: lost term order information

Deep Relevance Matching Model (DRMM) (Guo et al., CIKM '16)

- Matching histogram mapping for summarizing each query matching signals
- Term gating network for weighting the query matching signals
- Lost word order information (during histogram mapping)



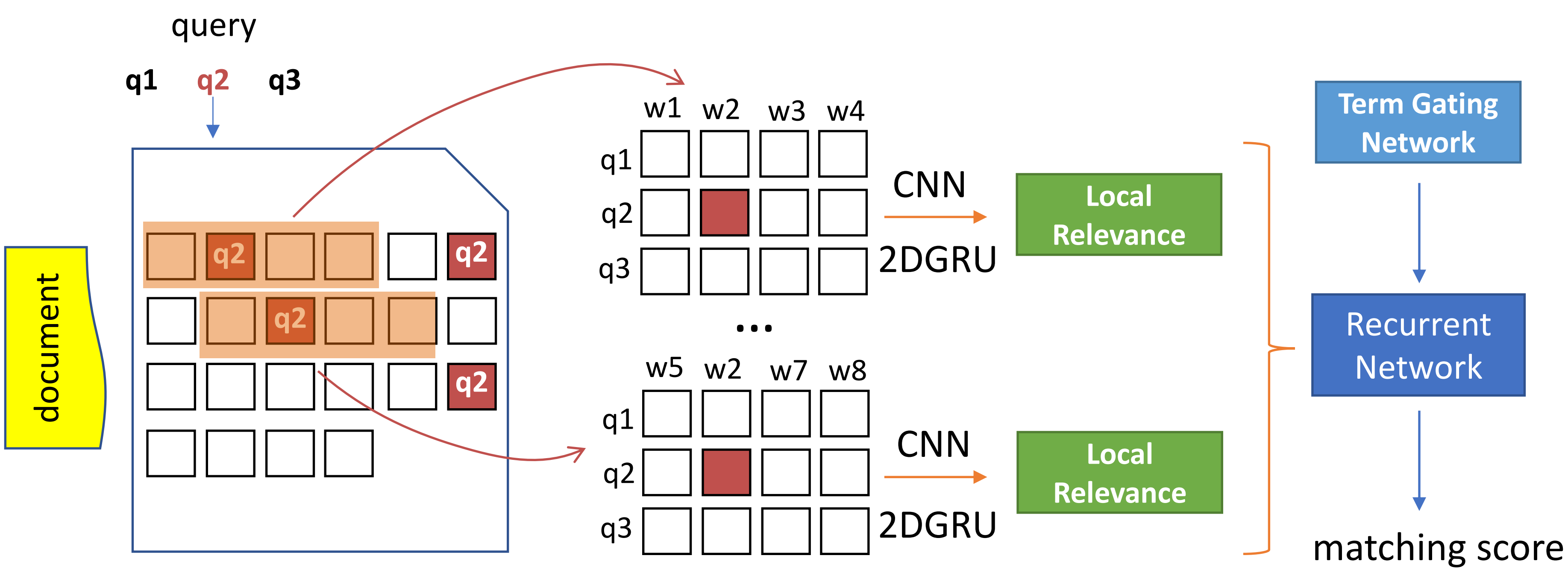
Local Context of Matching Positions

- Step 1: find matching positions for each query term
- Step 2: calculate matching signals within the local context
- Step 3: aggregate the local signals

- Advantages:
  - Matching between short query text and long document text
  - Robust: filtered out irrelevant context
  - Keep order information within the context

DeepRank (Pang et al., CIKM '17)

- Calculate relevance by mimicking the human relevance judgement process



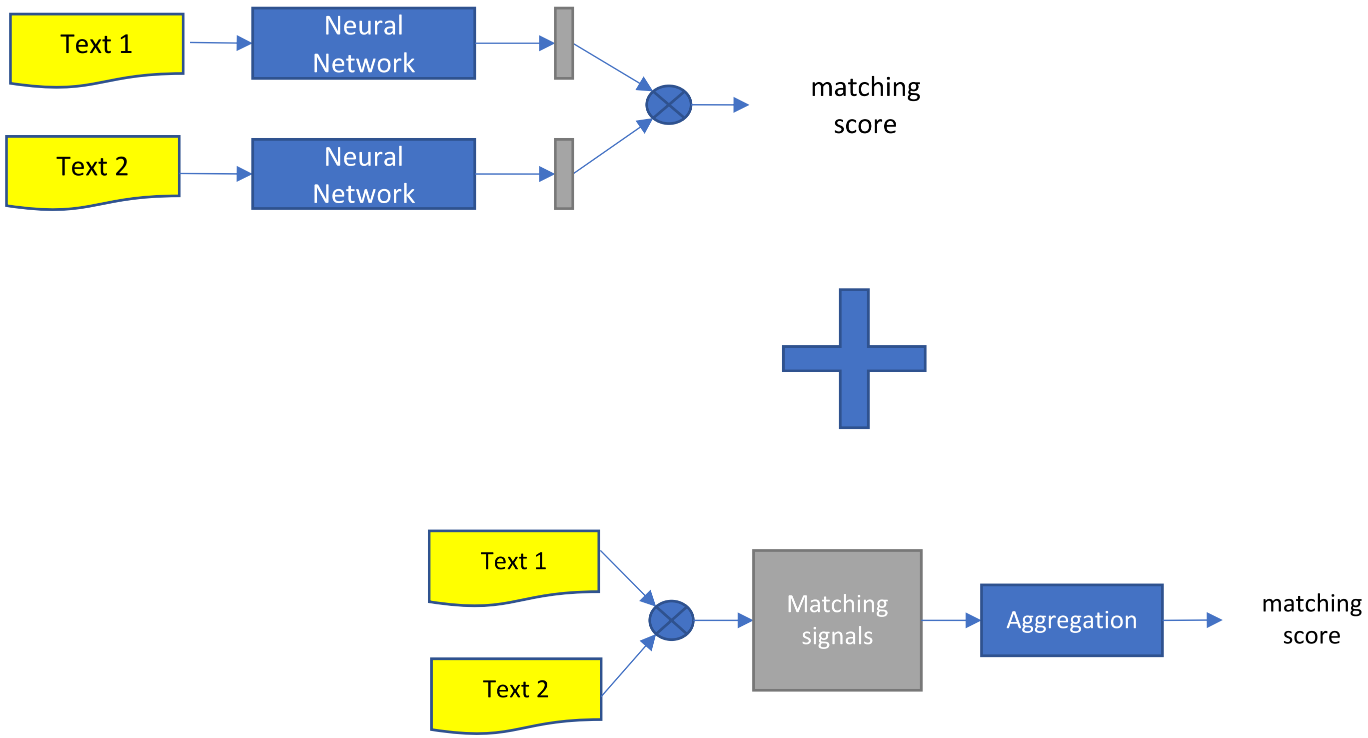
1. Detecting Relevance locations: focusing on locations of query terms when scanning the whole document

2. Determining local relevance: relevance between query and each location context, using MatchPyramid / MatchSRNN etc.

3. Matching signals aggregation:

$$F(\mathbf{q}, \mathbf{d}) = \sum_{w \in \mathbf{q}} (E_w \mathbb{I})^T \cdot \mathcal{T}(w)$$

Methods of Combination



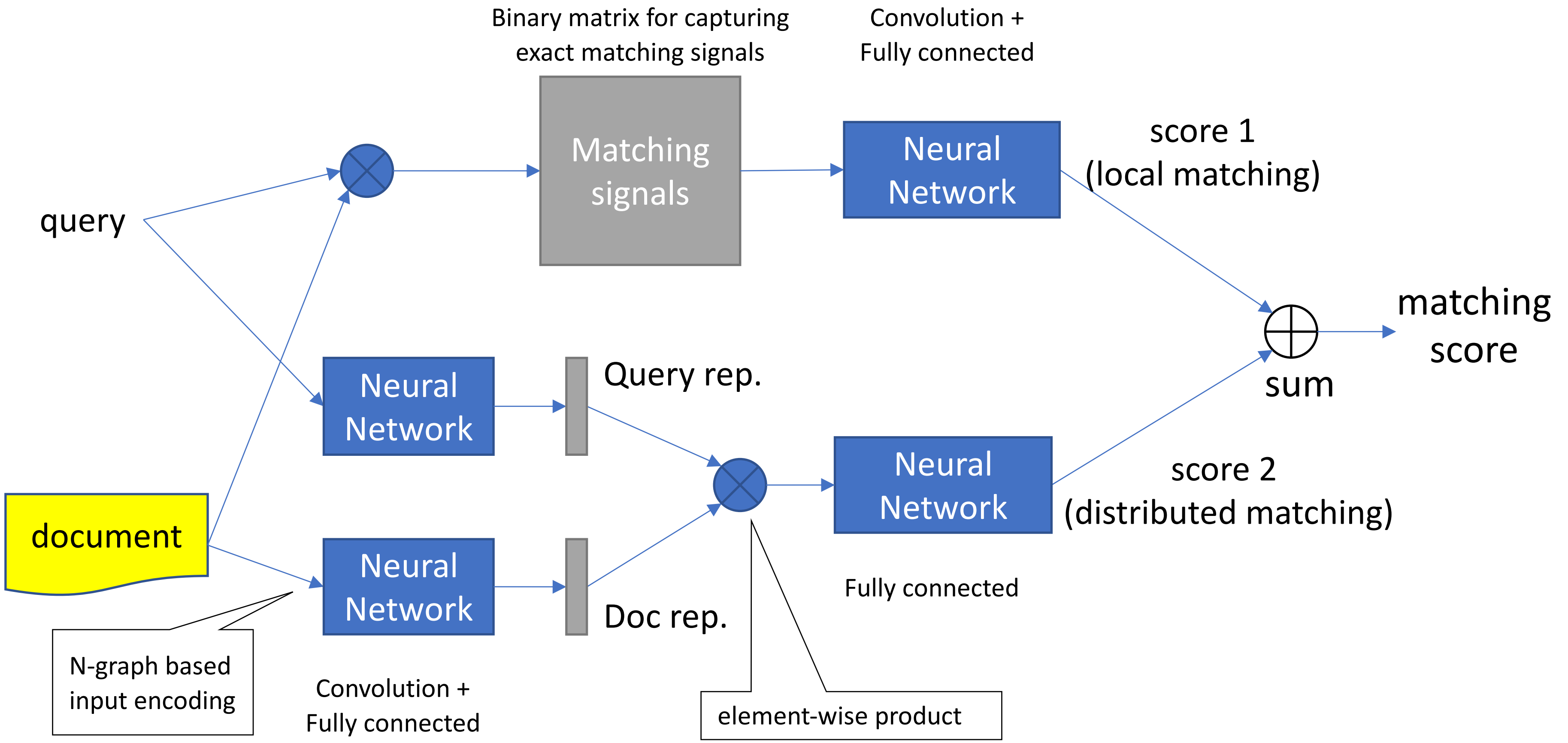
Typical Combination Methods

- Combined Strategy
  - DUET (Mitra et al., WWW ' 17)

- Coupled Strategy
  - BiMPM (Wang et al., IJCAI ' 17)
  - MwAN (Tan et al., IJCAI ' 18)
  - RE2 (Yang et al., ACL' 19)
  - BERT (Devlin et al., ACL' 19)

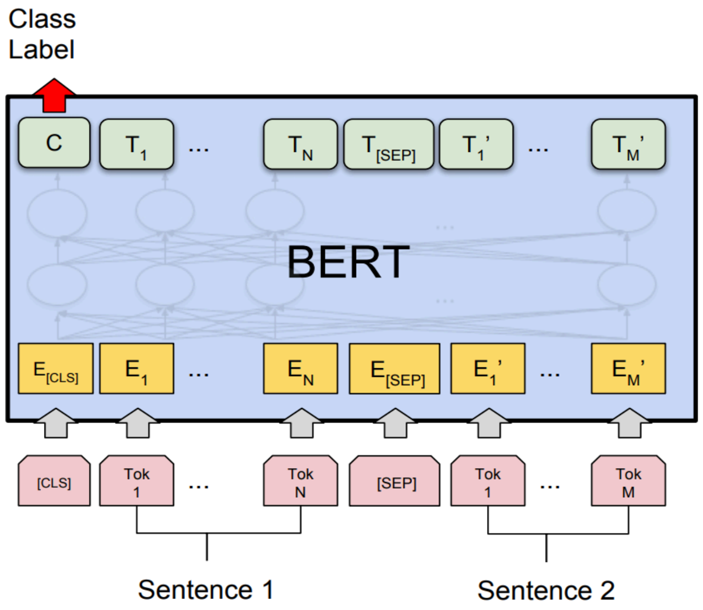
DUET (Mitra et al., WWW '17)

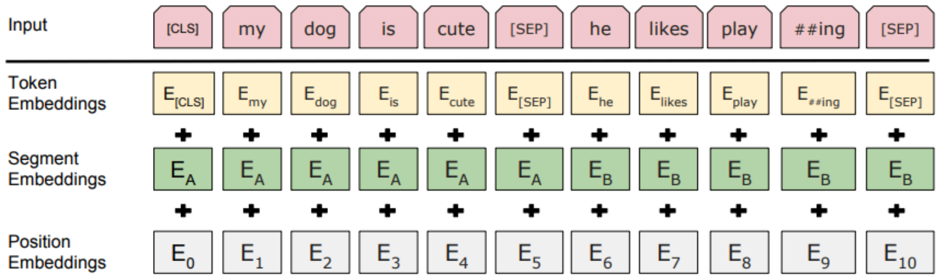
- Hypothesis: matching with distributed representations complements matching with local representations
  - Local matching: matching function learning
  - Distributed matching: representation learning



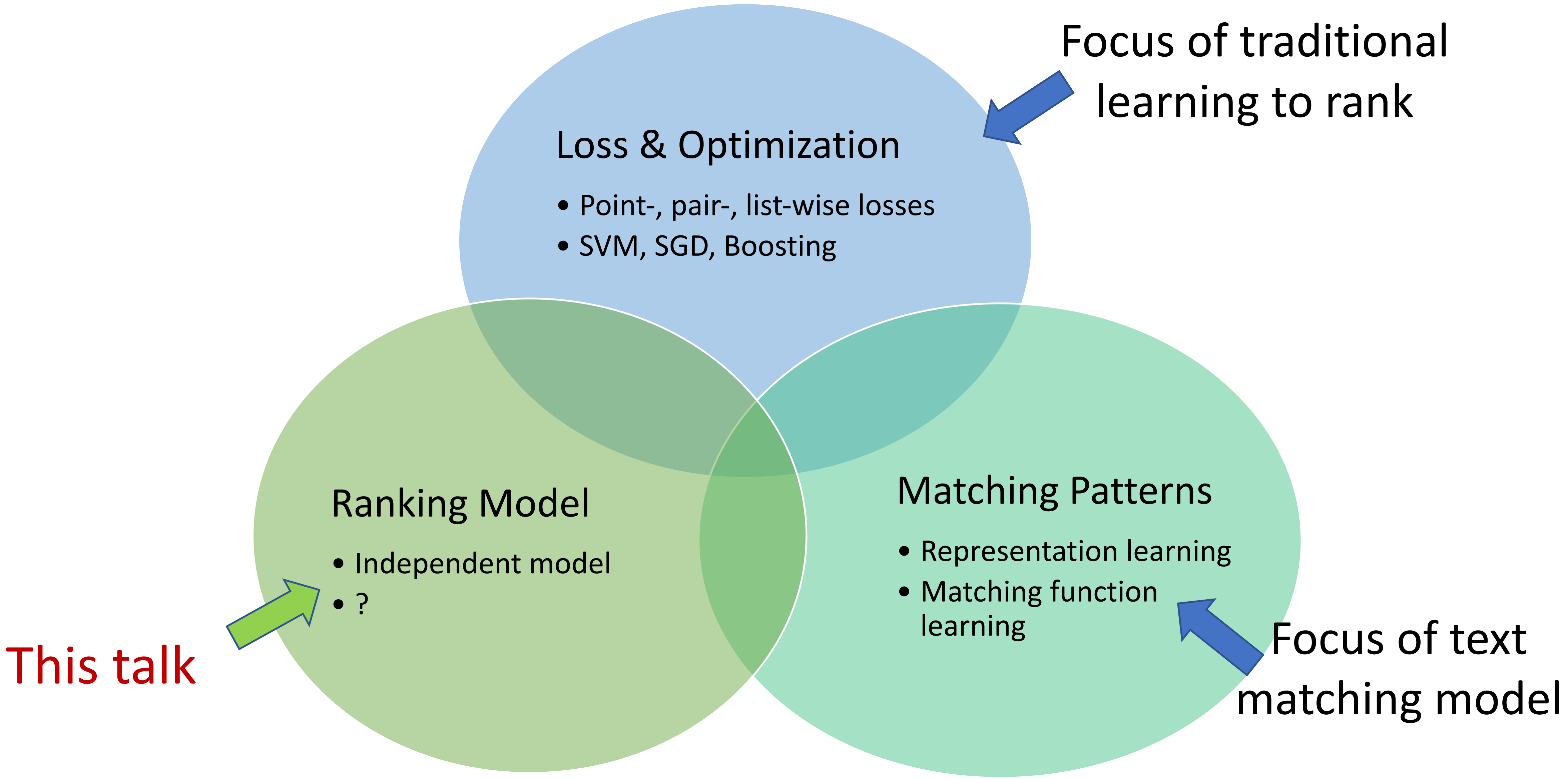
BERT (Devlin et al., ACL' 19)

- Problem: pretrained language information can improve the semantic matching problem
- Solution: Pre-training of Deep Bidirectional Transformers
  - Step 1: Pre-training BERT
  - Step 2: Finetuning on specific task





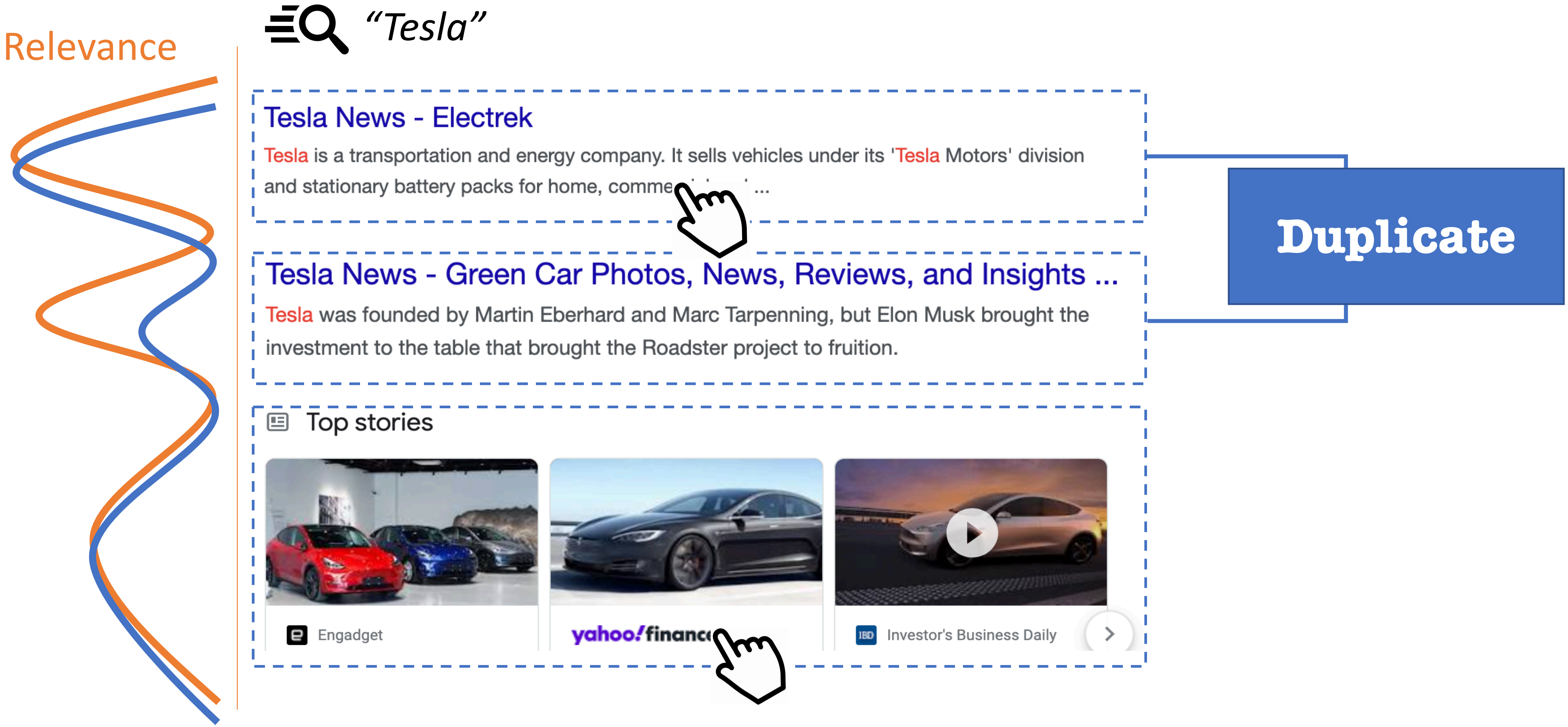
Building Ranking Models



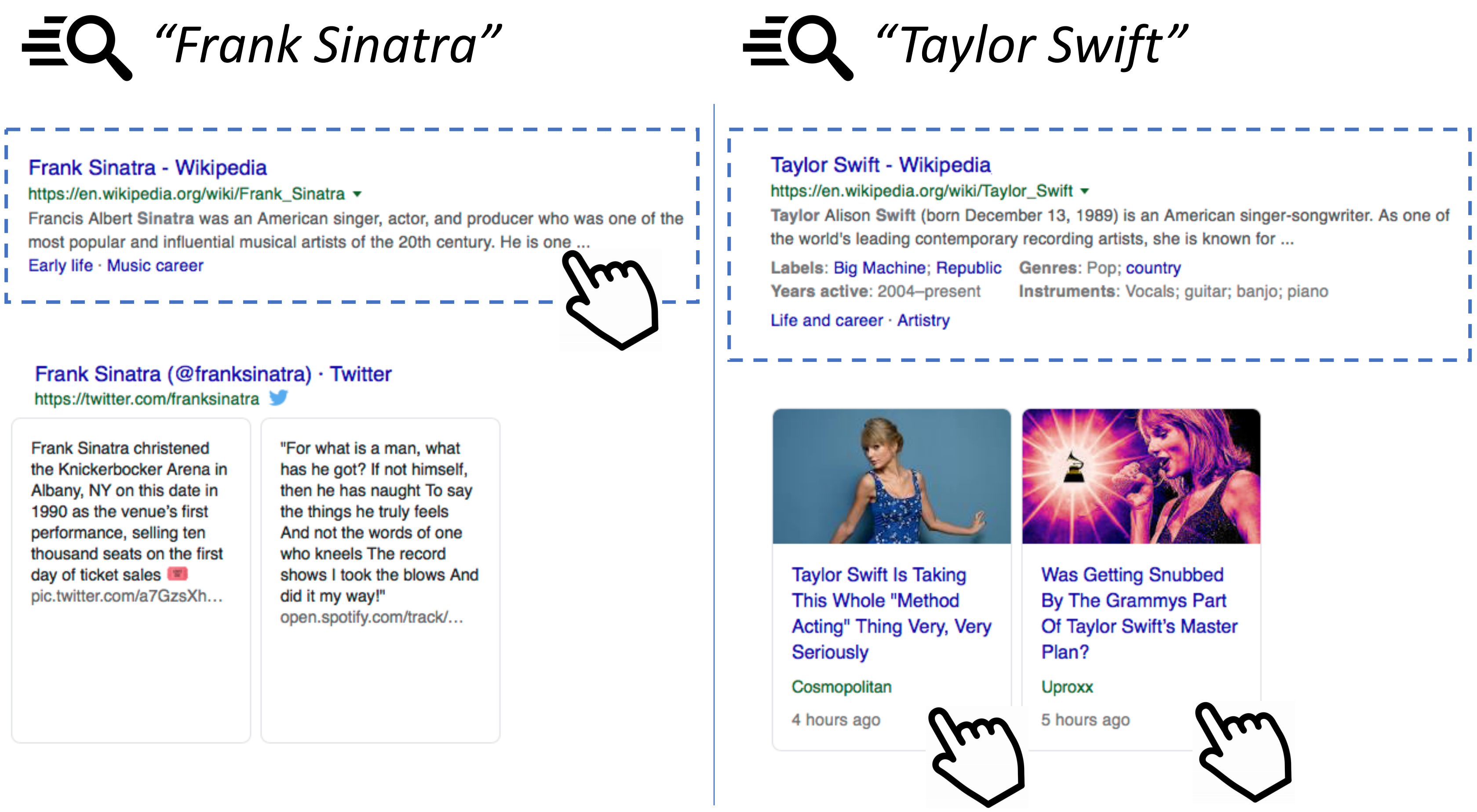
3. Limitations of PRP Principle

Is PRP Optimal?

Problem-1: Duplication of Documents

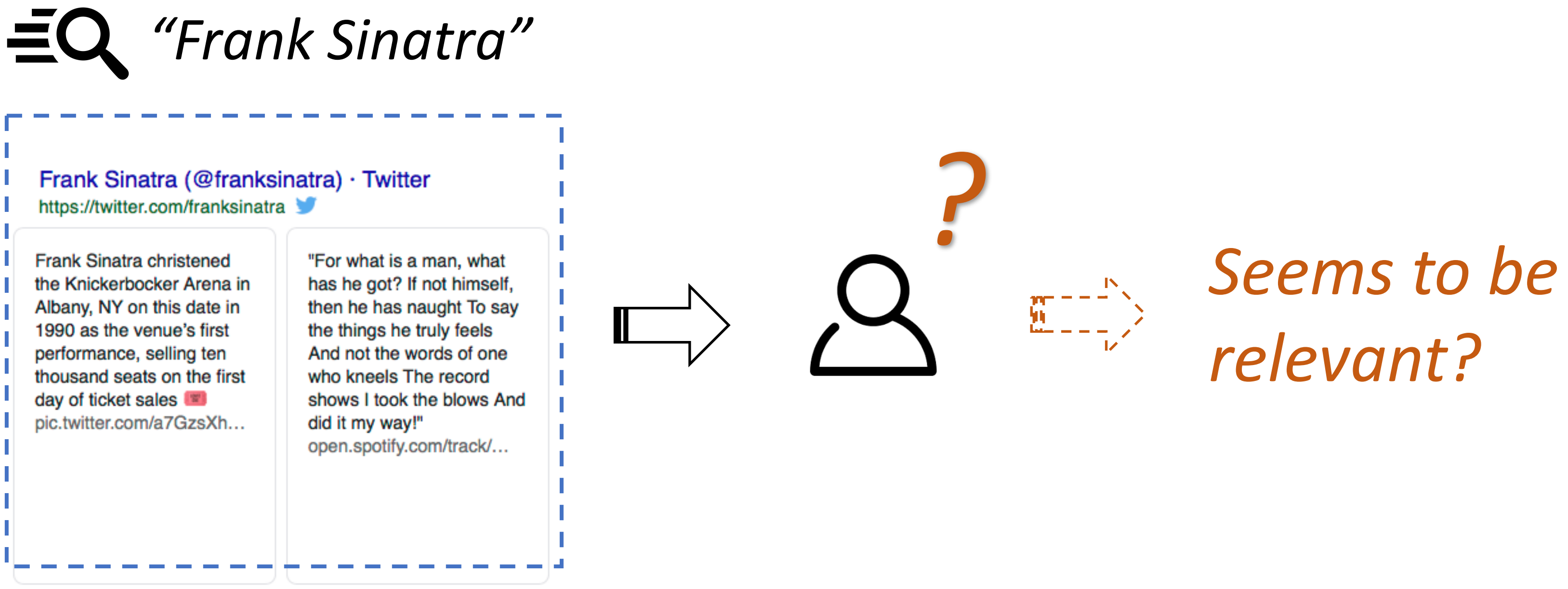


Problem-2: Relevance Depends on the Context



Qingyao Ai, Keping Bi, Jiafeng Guo and W. Bruce Croft. Learning a Deep Listwise Context Model for Ranking Refinement. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor Michigan, U.S.A. (SIGIR 2018).

Problem-3: Comparison Nature of Relevance



Problem-3: Comparison Nature of Relevance

"Frank Sinatra"

Frank Sinatra - Wikipedia
https://en.wikipedia.org/wiki/Frank_Sinatra
Francis Albert **Sinatra** was an American singer, actor, and producer who was one of the most popular and influential musical artists of the 20th century. He is one ...
Early life · Music career

Frank Sinatra (@franksinatra) · Twitter
https://twitter.com/franksinatra

Frank Sinatra christened the Knickerbocker Arena in Albany, NY on this date in 1990 as the venue's first performance, selling ten thousand seats on the first day of ticket sales
pic.twitter.com/a7GzsXh...

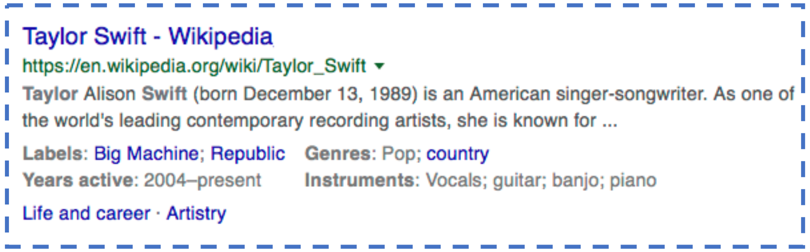
"For what is a man, what has he got? If not himself, then he has naught To say the things he truly feels And not the words of one who kneels The record shows I took the blows And did it my way!"
open.spotify.com/track/...

The first one is better!

Most results are fairly old

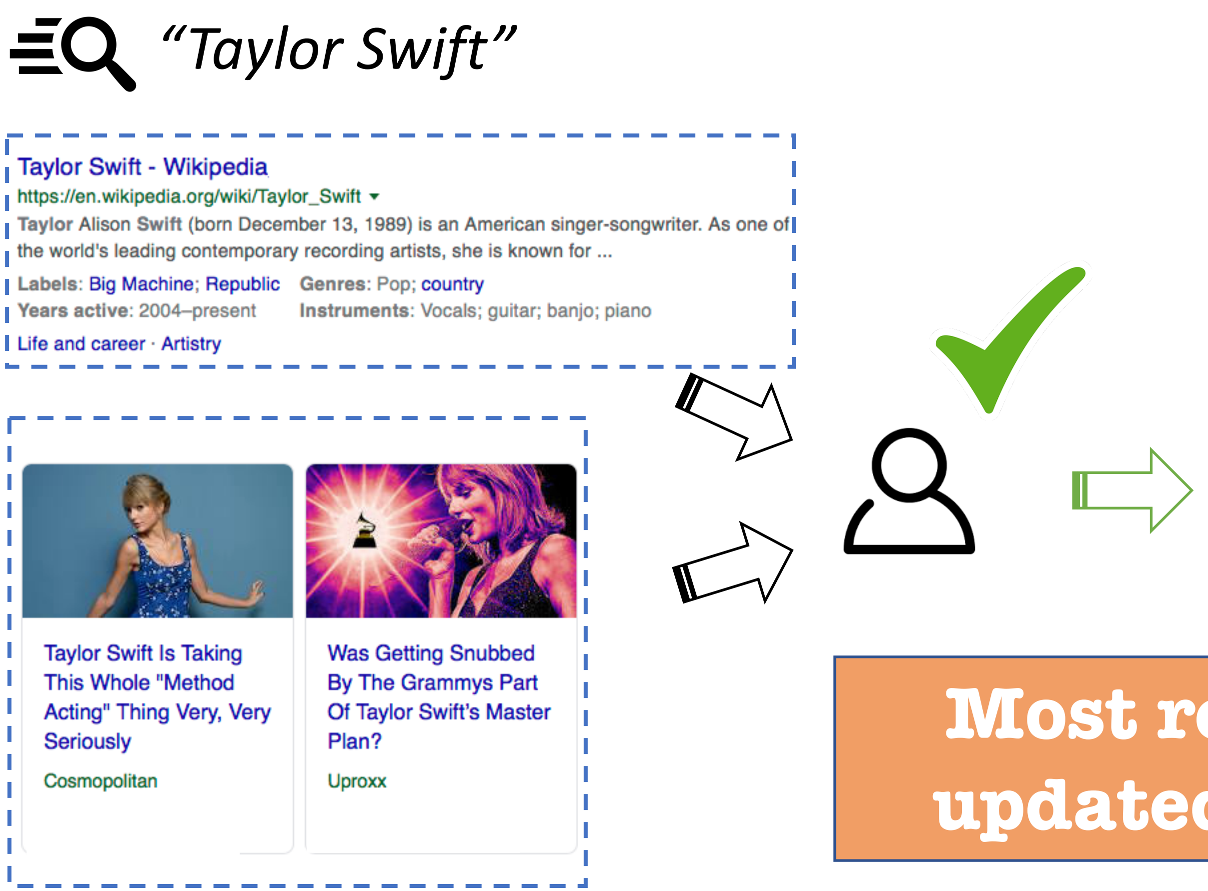
Problem-3: Comparison Nature of Relevance

“Taylor Swift”


Taylor Swift - Wikipedia
https://en.wikipedia.org/wiki/Taylor_Swift ▾
Taylor Alison **Swift** (born December 13, 1989) is an American singer-songwriter. As one of the world's leading contemporary recording artists, she is known for ...
Labels: Big Machine; Republic **Genres**: Pop; country
Years active: 2004–present **Instruments**: Vocals; guitar; banjo; piano
Life and career · Artistry

Seems to be relevant?

Problem-3: Comparison Nature of Relevance



The second one is better!

Most results are updated recently

PRP is Suboptimal

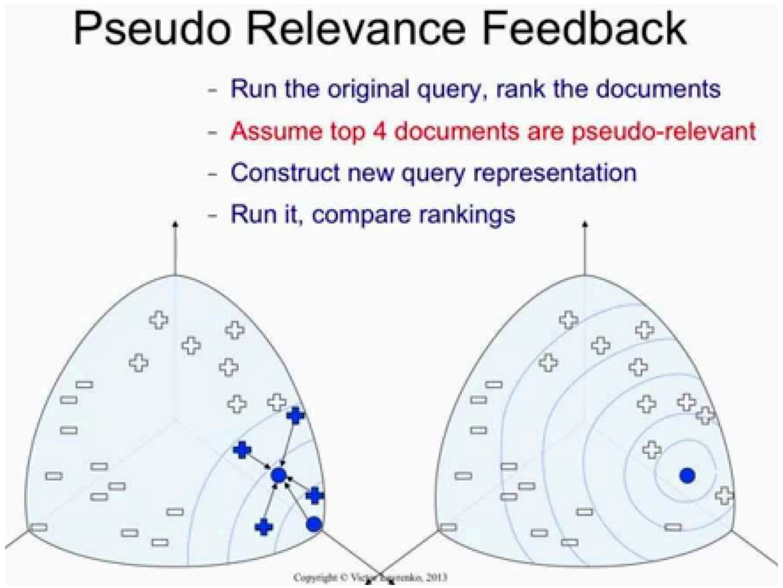
Beyond Probability Ranking Principle (PRP)

Each doc independently of the rest?

If not, PRP is not valid.

1. Cross-doc interactions & Local context information
   - pseudo relevance feedback
   - diversity ranking
   - Interactive information retrieval
   - query-dependent learning-to-rank

2. PRP works doc-by-doc but ranking is evaluated request-by-request
   - user's interactions with information retrieval systems show strong comparison patterns



Interactive Information Retrieval

Reference

- Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, Xueqi Cheng, Text Matching as Image Recognition, the 13th AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA. (AAAI 2016)
- Shengxian Wan, Yanyan Lan, Jiafeng Guo, Jun Xu, Liang Pang, Xueqi Cheng, A Deep Architecture for Semantic Matching with Multiple Positional Sentence Representations, the 13th AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA. (AAAI 2016)
- Shengxian Wan, Yanyan Lan, Jun Xu, Jiafeng Guo, Liang Pang, Xueqi Cheng, Match-SRNN: Modeling the Recursive Matching Structure with Spatial RNN, the 25th International Joint Conference on Artificial Intelligence, New York City, USA. (IJCAI 2016)
- Jiafeng Guo, Yixing Fan, Liang Pang, Liu Yang, Qingyao Ai, Hamed Zamani, Chen Wu, W. Bruce Croft and Xueqi Cheng. A Deep Look into Neural Ranking Models for Information Retrieval. Information Processing & Management (IPM).
- Robertson, S.E. The Probability Ranking Principle in IR. J. Documentation (1977), 294304. Reprinted in Readings in Information Retrieval, K. Sparck Jones and P. Willett (eds), 1997, 281286.
- Huang P-S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data//Proceedings of the 22nd ACM international conference on Conference on information and knowledge management. Amazon, India, 2013: 2333-2338
- Shen Y, He X, Gao J, et al. A latent semantic model with convolutional-pooling structure for information retrieval//Proceedings of the 23rd ACM international conference on Conference on information and knowledge management. New York, USA, 2014: 101-110.
- Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences//Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 2014: 2042-2050.
- Qiu X, Huang X. Convolutional neural tensor network architecture for community-based question answering//Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, 2015: 1305-1311.
- Palangi H, Deng L, Shen Y, et al. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(4): 694-707.

Reference

- Jiafeng Guo, Yixing Fan, Qiqing Yao, W. Bruce Croft, A Deep Relevance Matching Model for Ad-hoc Retrieval, the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016), Indianapolis, USA.
- Liu Yang, Qingyao Ai, Jiafeng Guo, W. Bruce Croft, aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model, the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016), Indianapolis, USA.
- Yixing Fan, Jiafeng Guo, Yanyan Lan, Jun Xu, Chengxiang Zhai, and Xueqi Cheng. Modeling Diverse Relevance Patterns in Ad-hoc Retrieval. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor Michigan, U.S.A. (SIGIR 2018).
- 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 International ACM SIGIR conference on research and development in information retrieval, pp. 55-64. 2017.
- Dai, Zhuyun, Chenyan Xiong, Jamie Callan, and Zhiyuan Liu. "Convolutional neural networks for soft-matching n-grams in ad-hoc search." In Proceedings of the eleventh ACM international conference on web search and data mining, pp. 126-134. 2018.
- Kai Hui, Andrew Yates, Klaus Berberich, and Gerard de Melo. 2017. A PositionAware Deep Model for Relevance Matching in Information Retrieval. In EMNLP ’17.
- Mitra, Bhaskar, Fernando Diaz, and Nick Craswell. "Learning to match using local and distributed representations of text for web search." In Proceedings of the 26th International Conference on World Wide Web, pp. 1291-1299. 2017.
- Wang, Zhiguo, Wael Hamza, and Radu Florian. "Bilateral multi-perspective matching for natural language sentences." In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 4144-4150. 2017.
- Yang, Runqi, Jianhai Zhang, Xing Gao, Feng Ji, and Haiqing Chen. "Simple and Effective Text Matching with Richer Alignment Features." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4699-4709. 2019.
- Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019, June. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

Thanks Q & A

Name: Liang Pang | Email: pangliang@ict.ac.cn