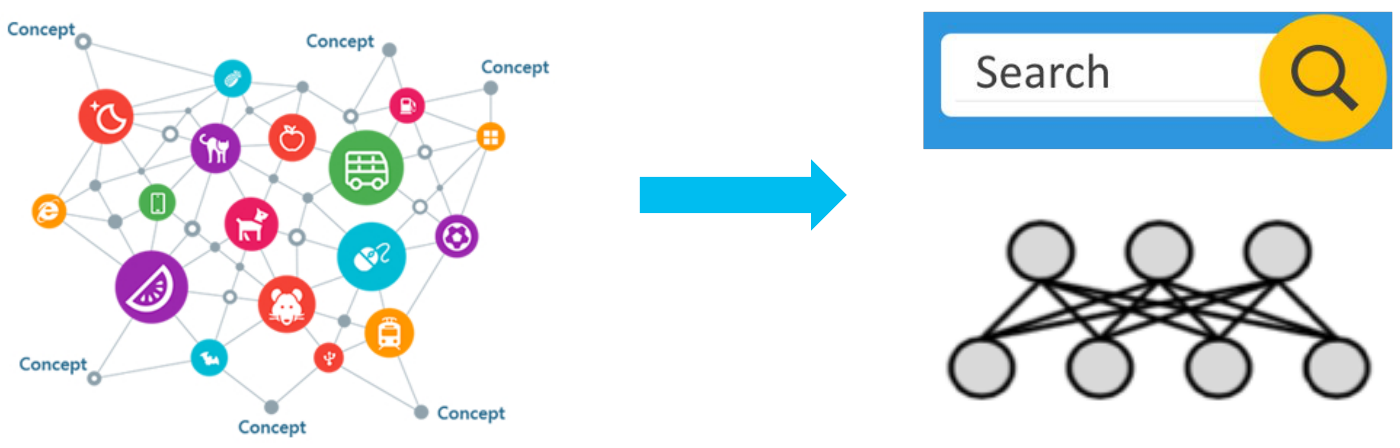
Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval

Zhenghao Liu[1], Chenyan Xiong[2], Maosong Sun[1], Zhiyuan Liu[1]

1.Tsinghua University 2.Carnegie Mellon University

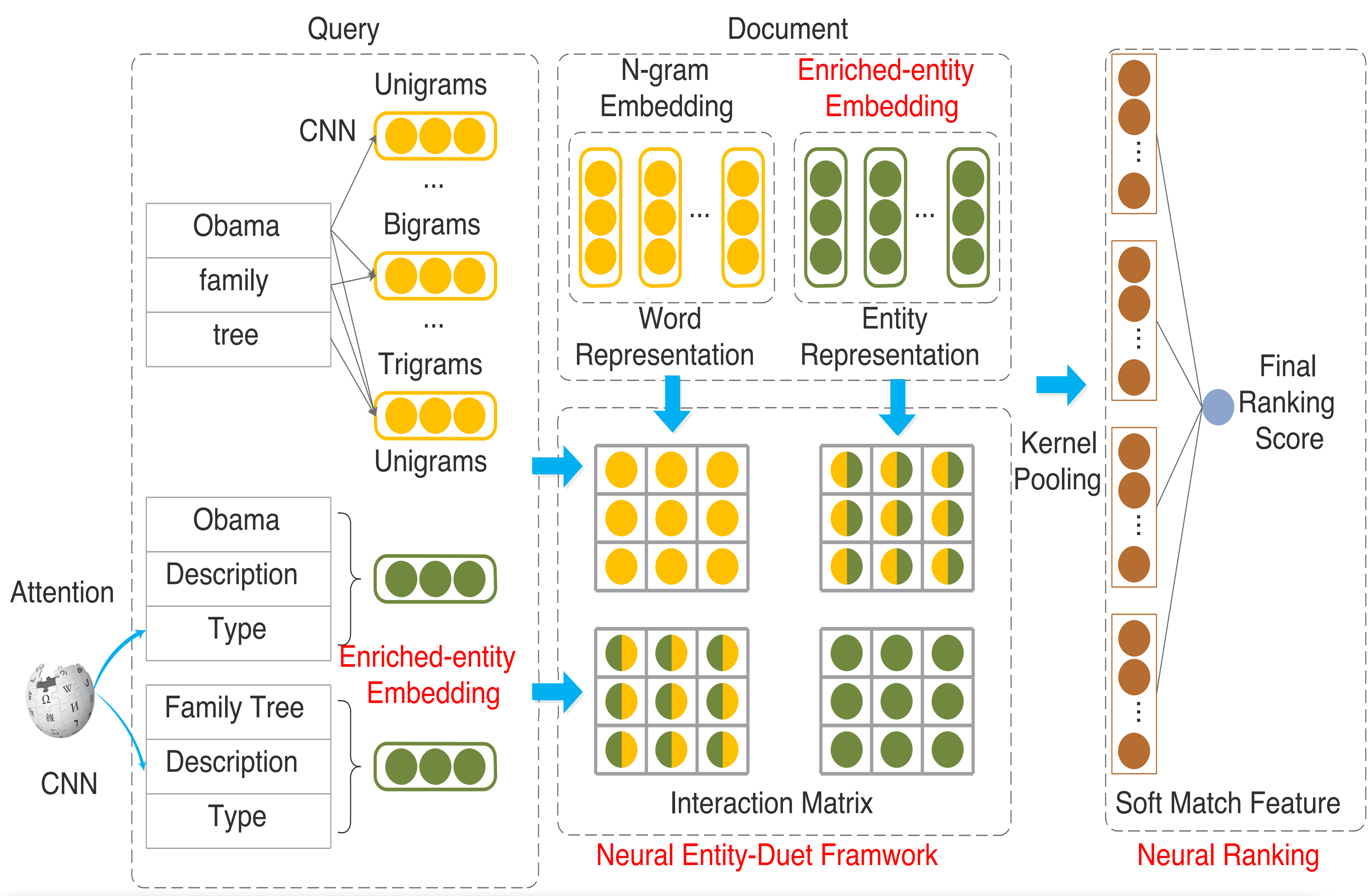
Motivation and Background

- **Queries and documents often match based on knowledge**
  - **Query:** "Meituxiuxiu web version"
  - **Document:** "Meituxiuxiu web version: An online picture processing tools"
  - **Meituxiuxiu web version:** Meituxiuxiu is the most popular Chinese image processing software, launched by the Meitu company
- **Our motivation is to study the effectiveness of knowledge graph semantics in state-of-the-art neural ranking models**



Entity-Duet Neural Ranking Model (EDRM)

- **Enriched-entity Embedding**
  - Integration of knowledge graph semantics
- **Neural Entity-Duet Framework**
  - Multi-level soft matches in the embedding space
- **Integration with Kernel based Neural Ranking (K-NRM)**
  - K-NRM and Conv-KNRM are state-of-the-arts, which calculate n-gram and entity cross matches with Gaussian Kernels
  - K-NRM -> **EDRM-KNRM**
  - Conv-KNRM->**EDRM-CKNRM**



Experimental Methodology

- **Dataset:**
  - Sogou query log
  - About 100K training queries and 1K testing queries
- **Knowledge Graph:**
  - CN-DBpedia, a Chinese knowledge graph
  - Entities in both queries and documents are linked with CMNS
- **End-to-end Training:**
  - Train on relevance labels estimated by a click model (DCTR), about 8500k training pairs
  - Test on two click model labels (DCTR->Testing-SAME and TACM->Testing-DIFF) and raw user clicks (Testing-RAW)

Experimental Results

Overall Performance

	Testing-SAME				Testing-DIFF				Testing-RAW	
Method	NDCG@1		NDCG@10		NDCG@1		NDCG@10		MRR	
BM25	0.142	-46%	0.287	-32%	0.163	-46%	0.325	-23%	0.228	-34%
RankSVM	0.146	-45%	0.309	-26%	0.170	-43%	0.352	-17%	0.224	-35%
Coor-Ascent	0.159	-40%	0.355	-15%	0.209	-30%	0.378	-11%	0.242	-30%
DRMM	0.137	-48%	0.313	-25%	0.213	-29%	0.359	-15%	0.234	-32%
CDSSM	0.144	-46%	0.333	-21%	0.183	-39%	0.353	-16%	0.231	-33%
MP	0.218	-17%	0.379	-10%	0.197	-34%	0.345	-18%	0.240	-30%
K-NRM	0.265	–	0.420	–	0.300	–	0.423	–	0.345	–
Conv-KNRM	0.336	27%	0.481	15%	0.338	13%	0.432	2%	0.358	4%
EDRM-KNRM	0.310	17%	0.455	8%	0.333	11%	0.434	3%	0.362	5%
EDRM-CKNRM	**0.340**	**28%**	**0.482**	**15%**	**0.371**	**24%**	**0.451**	**7%**	**0.389**	**13%**

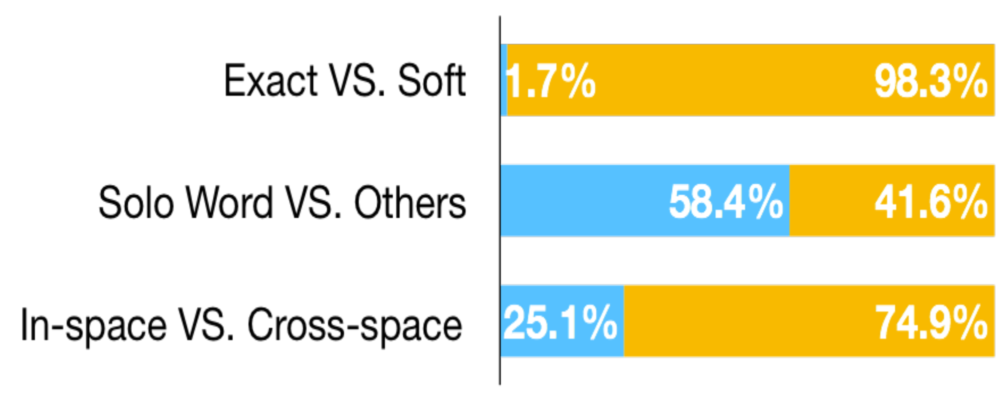
On Testing-SAME

- Significant improvement compared to K-NRM
- Little improvement compared to Conv-KNRM
- Conv-KNRM is able to learn phrases matches (entity) from data

On Testing-DIFF and Testing-RAW

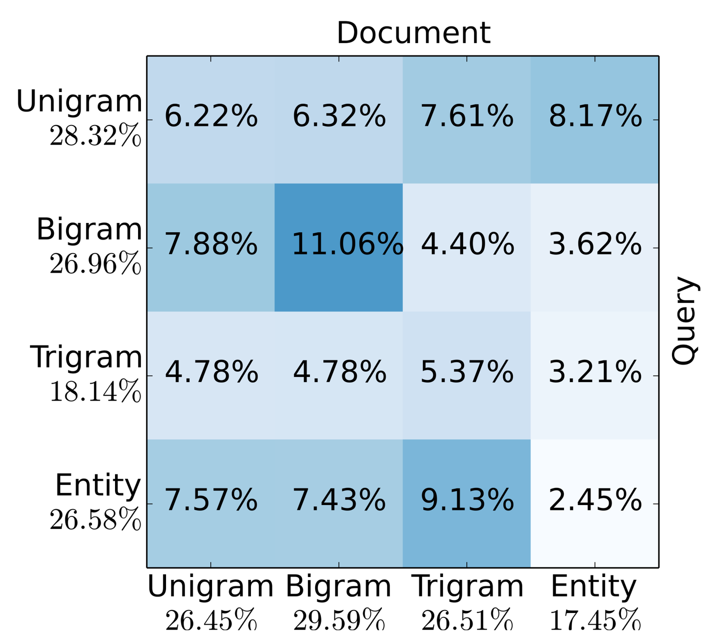
- Significant improvement compared to K-NRM and Conv-KNRM
- EDRM shows generalization ability

Ranking contribution for EDRM-CKNRM



Overall kernel weight

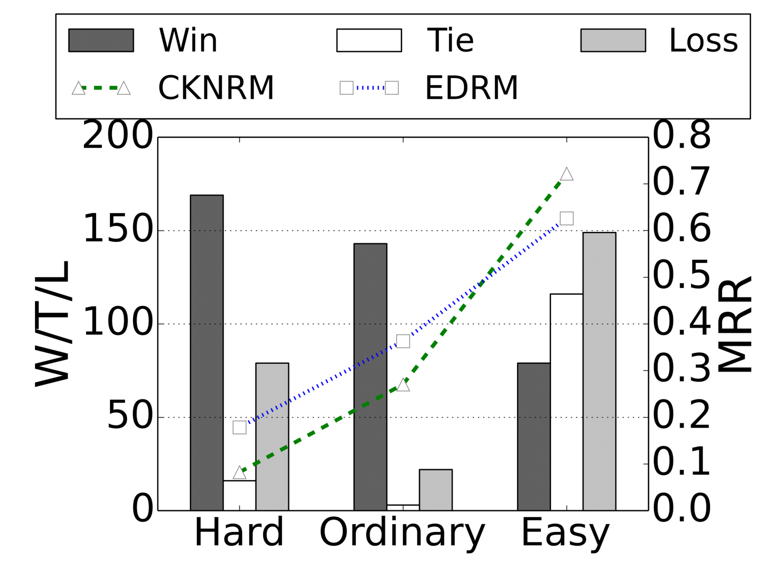
- Most of the weight goes to soft match
- Entity related matches play an important role
- Cross-space matches are more important



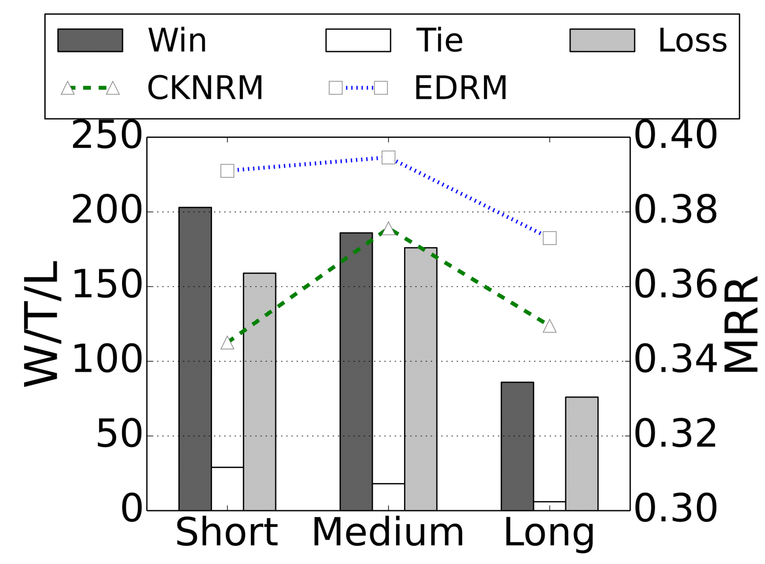
Individual kernel weight

- N-grams and entities are important components which share almost uniformly distributed weight

Performance on Different Scenarios



Query Difficulty Scenario



Query Length Scenario

- Greatest improvement on short and hard queries
- Knowledge are more crucial for the limited query text

Conclusion

- **Knowledge based Neural Ranking Model:**
  - Integrate knowledge graph semantics in state-of-the-art neural ranking models
  - Entity types and descriptions are external embeddings to match entities and n-grams
- **End-to-end Training with User Clicks:**
  - A data-driven combination of entity-oriented search and neural information retrieval
- **Effectiveness and Generalization ability:**
  - Show greater advantage on hard and short queries
  - Improve performances on more difficult testing scenarios

Paper

Codes

Email: liu-zh16@mails.tsinghua.edu.cn