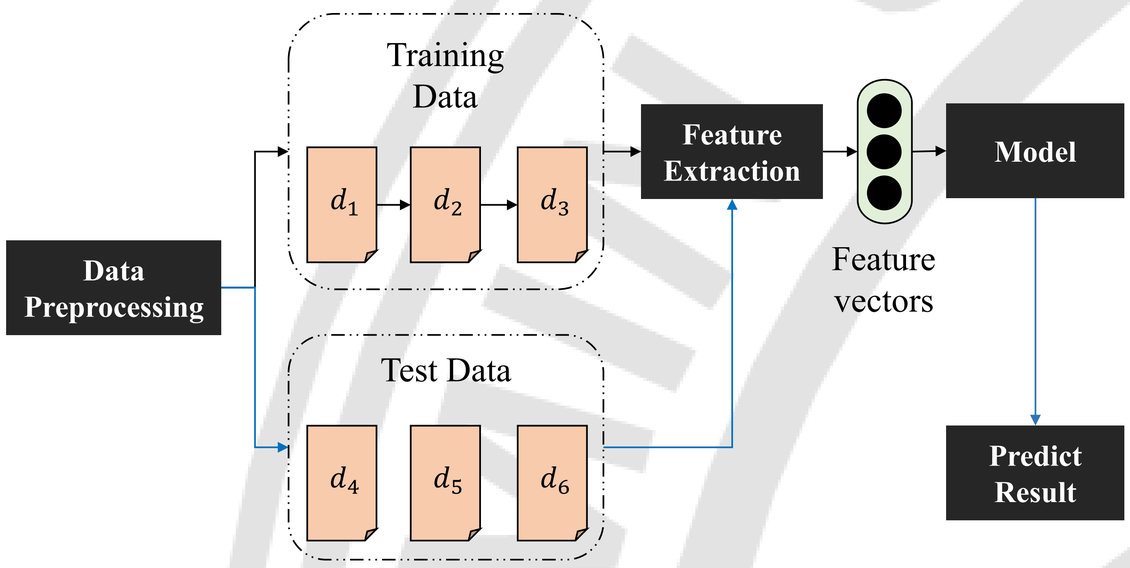
RUCIR at NTCIR-14 WWW-2 Task

Xue Yang, Shuqi Lu, Shijun Wang, Han Zhang and **Zhicheng Dou**
Renmin University of China, Beijing, P.R. China
{ruc_yangx, lusq, wangshijun, zhanghanjl, dou}@ruc.edu.cn

Data-Flow Overview



Step 1: Data Preprocessing

- Document Collection
  - Clueweb09, Clueweb12, SogouT-16, Sogou-QCL
- Relevance Judgement
  - TREC09 ~ TREC14, Sogou-QCL

Step 2: Feature Extraction

- Traditional Features
  - For different fields: (anchor), title, URL, body, whole

Name	Description
BM25	BM25 with default parameters
TF-IDF	TF-IDF model
LMIR	Language model with Dirichlet smoothing
TF	Sum of term frequency
IDF	Sum of inverse document frequency
DL	Document length
PM	Perfect match
CM	Complete match

- Embedding Features
- Deep Neural Features
  - ARC-I
  - ARC-II
  - DRMM
  - aNMM
  - MV-LSTM
  - DUET

Step 3: Training & Evaluation

- Model
  - Ranklib: LambdaMART
- Metric Evaluation
  - NDCG@K, Q@K, nERR@K

Experiment Results (Chinese)

Run	Query	Features	nDCG@10	Q@10	nERR@10
RUCIR-1	Description	Traditional Embedding	0.4515	0.4228	0.5792
RUCIR-2	Content	Traditional Embedding	**0.4866**	**0.4571**	**0.6044**
RUCIR-3	Description	Traditional	0.4503	0.4223	0.5630
RUCIR-4	Description	Traditional Deep Neural	0.4458	0.4226	0.5619
RUCIR-5	Description	Deep Neural	0.2745	0.2404	0.3832

Experiment Results (English)

Run	Query	Features	nDCG@10	Q@10	nERR@10
RUCIR-1	Description	Traditional	0.3137	0.2973	0.4469
RUCIR-2	Content	Traditional	**0.3489**	**0.3352**	**0.4917**
RUCIR-3	Description	Traditional Embedding	0.3137	0.2973	0.4469
RUCIR-4	Description	Traditional Deep Neural	0.3293	0.3094	0.4602
RUCIR-5	Description	Deep Neural	0.2876	0.2659	0.4188

* Embedding features are cosine similarity between query representation and document representation, which are obtained by averaging the word vectors in the text. Word vectors are trained by word2vec.

Conclusion

- [CN] Traditional Features + Embedding Features >> Other Runs; [EN] Traditional Features >> Other Runs (2 >> 1,3,4,5)
- Query Content > Query Description (CO > DE)
- Deep Neural Features << Other Runs (5 << 1,2,3,4)