UMassAmherst
College of Information & Computer Sciences
Center for Intelligent Information Retrieval

A Study of Neural Matching Models for Cross-lingual IR

Puxuan Yu and James Allan
CIIR, UMass Amherst

Email: pxyu@cs.umass.edu
Slides: cs.umass.edu/~pxyu/public/pdf/inter-clir-sigir20.pdf
Paper: https://dl.acm.org/doi/abs/10.1145/3397271.3401322

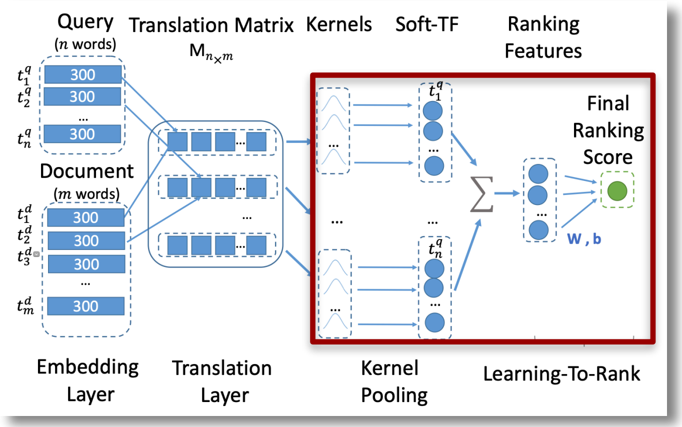
Motivation

- Neural matching models have been popular for mono-lingual IR
  - Mono-lingual word embeddings + matching architectures
- Cross-lingual word embeddings (CLWEs) are developing rapidly
  - Yet for CLIR, they are only used for query translation

- Why not use CLWEs + matching architectures for CLIR?
- Leads to end-to-end CLWEs learning for retrieval

Background

- Neural matching models:
  - ARC-I, ARC-II, MatchPyramid, DRMM, PACRR, Duet, KNRM, ……
  - Query-document interactions w/ different architectures on top
  - Some support end-to-end training



An illustration of the KNRM model [Xiong et al., SIGIR 2017].

Background

- CLWEs for CLIR [Litschko et al., SIGIR 2018]:
  - Top-1 query translation* (TbT-QT):
    - Translate each query term to its nearest document language term
    - Retrieve using query-likelihood model
  - Higher-level query document interactions (BWE-AGG):
    - Score = cos_sim(q, d). q/d vector is aggregated from CLWEs of its constituent terms (simply average or weighted by IDF)
    - Combining representations in late stages

We want to answer…

- Is query translation still a good idea?
- Adapting matching models from mono-lingual to cross-lingual:
  - Do they work well off-the-shelf?
  - Cross-lingual IR has (almost) no exact term matching signals
    - Is it important?
  - Word-pair similarity distributions are very different
    - Does it affect parameter selection?

Experiments setup

- Baselines: heuristics using CLWEs
- Matching models: MP, DRMM, KNRM
- Interactions: cosine, Gaussian, exact, hybrid
  - Simulated exact: 1 for cos > η, 0 otherwise
  - Hybrid: concat ranking features from exact and cosine
- Query translation-based matching (*-TbT-QT):
  - Top-1 query translation with CLWEs, then matching using cosine

Experiments setup

- Data:
  - CLEF 2000-2003
  - English query -> {Dutch, Italian, Finnish, Spanish} documents
  - Documents are truncated to first 500 tokens
- Metric: MAP on all queries (5-fold cross validation)
- CLWEs: pre-aligned fastText embeddings
- Training:
  - Pairwise hinge loss

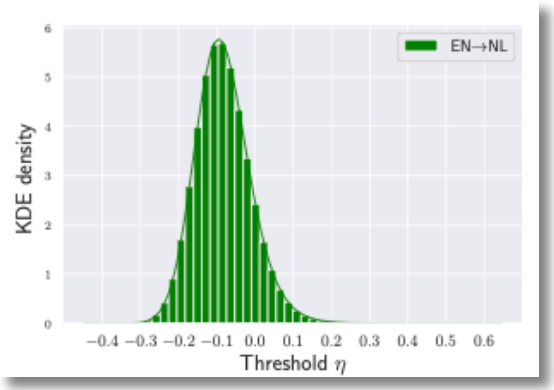
Results

- Neural matching models work off-the-shelf for CLIR
  - Start with lower-level query document interactions
  - DRMM works best. Attention on query terms could be a factor. Cosine as interaction function is robust.
- Query translation is unnecessary
  - {*-TbT-QT} is consistently worse than {*-Cosine}
- "Exact" matching gives strong results
  - Relevance is heavily dependent on top matching signals

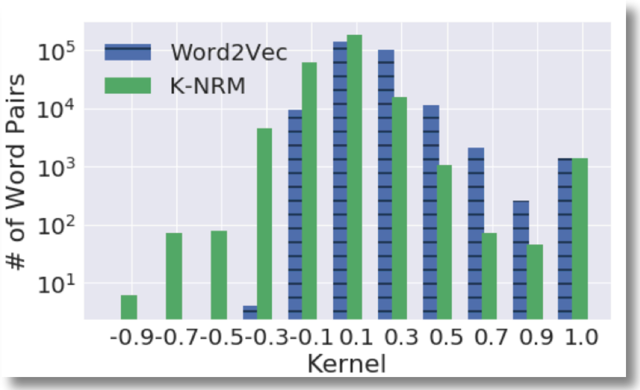
Lang. Pair	EN→ NL	EN→ IT	EN→ FI	EN→ ES
BWE-Agg-Add	.237	.173	.170	.297
BWE-Agg-IDF	.246	.178	.180	.298
TbT-QT-BM25	.240	.231	.122	.341
TbT-QT-QL	.297	.268	.126	.387
MP-Cosine	.348	.331	.254	.423
MP-Gaussian	.322	.319	.203	.405
MP-Exact	.327	.295	.202	.415
MP-Hybrid	.343	.326	.243	.427
MP-TbT-QT	.327	.300	.195	.409
DRMM-Cosine	**.374**	**.352**	**.304**	**.462**
DRMM-TbT-QT	.345	.324	.193	.450
KNRM-Cosine	.368	.313	.286	.423
KNRM-TbT-QT	.329	.288	.200	.405

Discussions

- Word-pair similarity distribution of CLWEs
  - Smaller mean, smaller variance, no skewness and low density at top similarity



EN query & NL document terms in CLWE space.



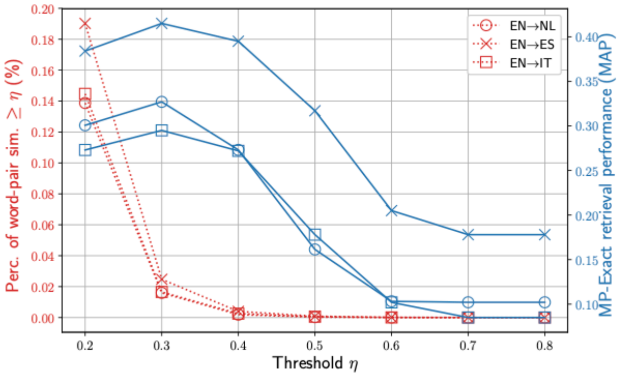
EN query & document terms in mono-lingual embedding space.
[Xiong et al., SIGIR 2017]

Discussions

- Top similarities are “flat”. Top translation term is not necessarily the best
  - Probably why TbT-QT-QL fails
- Exact matching peaks at certain similarity threshold across languages
  - Beneficial to considerate more than top-1 translation

EN	phone	telephones	Telephone	landline	rotary-dial
	0.818	0.761	0.720	0.694	0.669
ES	telefónicos	teléfono	telefónica	telefónia	telefóno
	0.535	0.522	0.522	0.520	0.520

Top-5 closest word to “telephone” in an English embedding space and CLWEs space (Spanish).



MP-Exact performance v.s. threshold.

Conclusions

- CLWEs + neural matching models work well for CLIR
  - DRMM works best; MatchPyramid is not bad compared with KNRM
- Directly model term-level query document interactions in the CLWEs space without query translation!
- Mostly top translations are helpful, but limiting to top-1 can be harmful
- CLWEs and mono-lingual word embeddings are distributed very differently
- Next step: end-to-end CLWEs learning for retrieval

UMassAmherst

College of Information & Computer Sciences

Center for Intelligent Information Retrieval

Puxuan Yu

cs.umass.edu/~pxyu

pxyu@cs.umass.edu