SLWWW at the NTCIR-13 WWW Task

Peng XIAO ,**Yimeng FAN**, Lingtao Li, Tetsuya Sakai

Waseda University

WASEDA University

Outlines

Objective

- Chinese Subtask of the We Want Web task
- Our goal : improve the search effectiveness
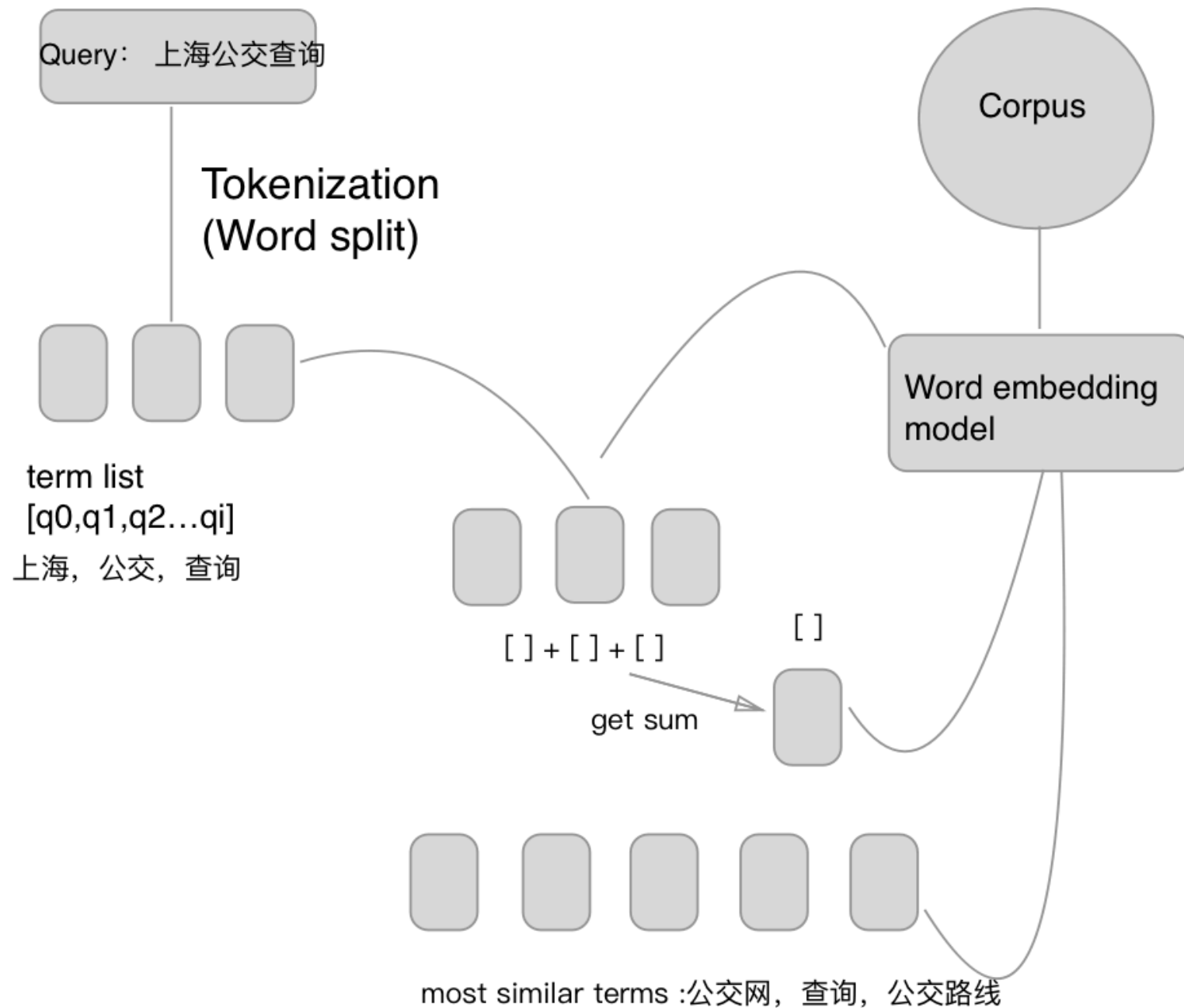
Outlines

Data

- Around 81,000 documents
- 100 topics
- Train: 92 topics with around 45,000 user logs; Test: 100 topics with around 88,000 user logs (never used)
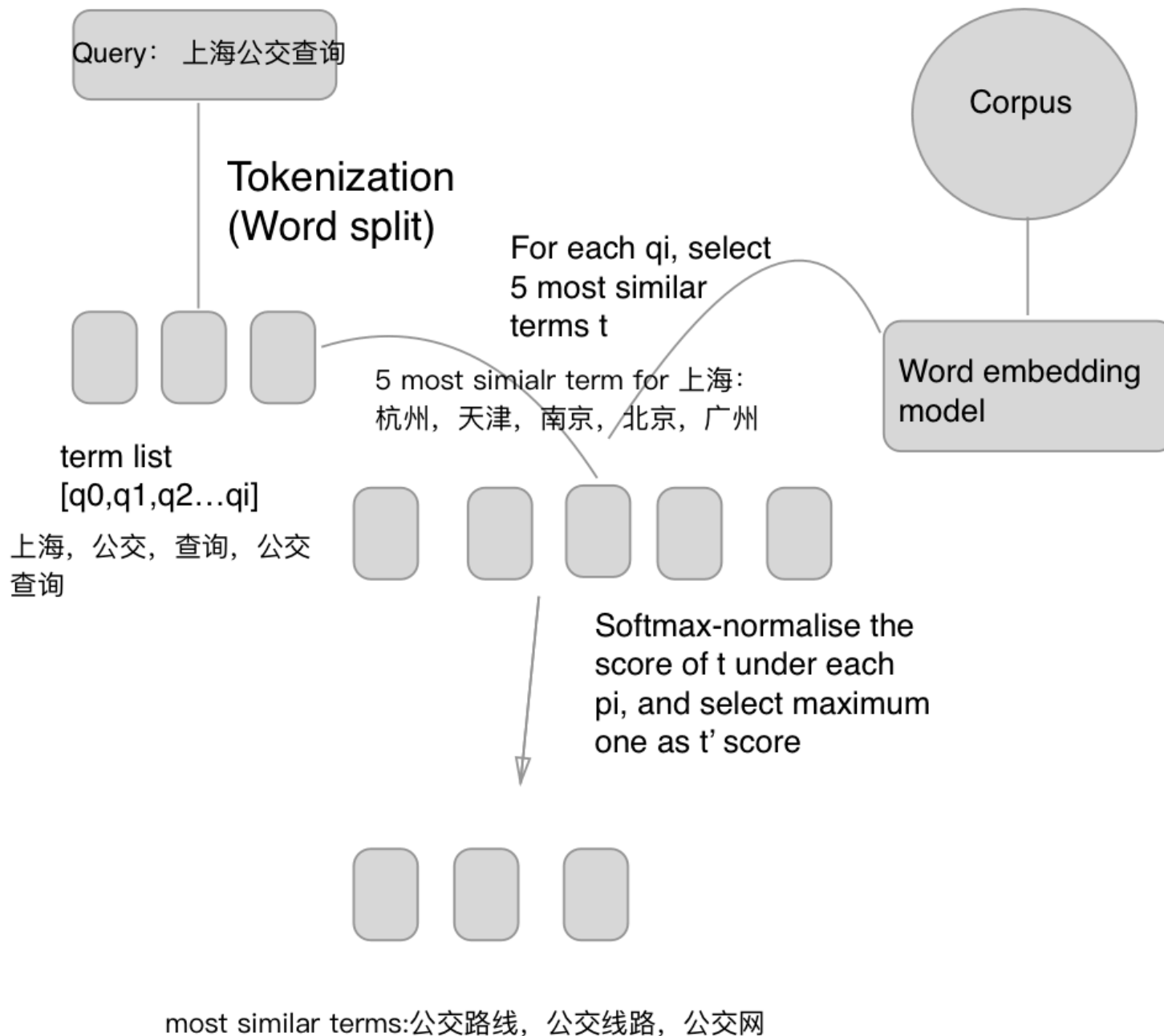- 200d word2vec model (full corpus of SogouT 16)

Outlines

Centroid

Query: 上海公交查询

Corpus

Tokenization
(Word split)

term list
[q0,q1,q2...qi]
上海，公交，查询

Word embedding
model

[ ] + [ ] + [ ]

[ ]

get sum

most similar terms :公交网，查询，公交路线

CombMAX

Query: 上海公交查询

Tokenization
(Word split)

Corpus

For each qi, select
5 most similar
terms t

Word embedding
model

term list
[q0,q1,q2...qi]

5 most simialr term for 上海:
杭州，天津，南京，北京，广州

上海，公交，查询，公交
查询

Softmax-normalise the
score of t under each
pi, and select maximum
one as t' score

most similar terms:公交路线，公交线路，公交网

Query expansion

Score of Cent/CombMAX

Score of maximum lilkelihood estimate

normalise and fuse

Select top3/top5 terms

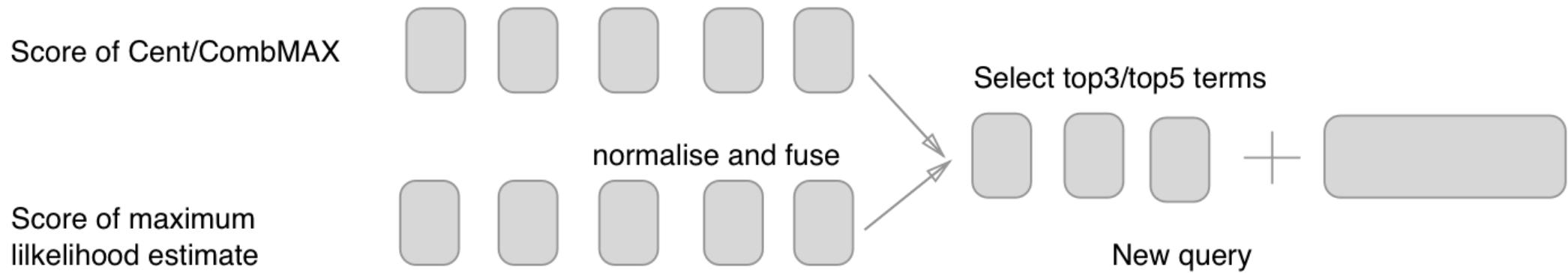New query

Outlines

1. Objective
2. Data
3. Query expansion based on word embedding
4. Official result and analysis
5. Conclusion

Official result

Run name	Term scoring method	#Expansion terms
SLWWW-C-NU-Base-1	Centroid	3
SLWWW-C-NU-Base-2	CombMAX	3
SLWWW-C-NU-Base-3	CombMAX	5
SLWWW-C-NU-Base-4	Centroid	5

run	Mean MSnDCG@10	Mean Q@10	Mean nERR@10
SLWWW-C-NU-Base-1	0.3206	0.3094	0.4753
SLWWW-C-NU-Base-2	0.3225	**0.3099**	0.4723
SLWWW-C-NU-Base-3	0.2909	0.2838	0.4327
SLWWW-C-NU-Base-4	0.2991	0.2949	0.4406
baseline	**0.3235**	0.2522	$\mathbf{0.5341}^{*3,*4}$

Condensed-list measures

r	relevance level
1	$L9$
2	$L0$
3	unjudged
4	$L0$
5	unjudged
6	unjudged
7	unjudged
8	$L0$
9	unjudged
10	$L0$

r	relevance level
1	$L9$
2	$L0$
3	$L0$
4	$L0$
5	$L0$
6	$\uparrow L0$
7	$\uparrow L0$
8	$\uparrow L0$
9	$\uparrow L3$
10	$\uparrow L0$

Office result after removing topic 33

run	Mean MSnDCG@10	Mean Q@10	Mean nERR@10
SLWWW-C-NU-Base-1	0.3177	0.3058	0.4715
SLWWW-C-NU-Base-2	0.3196	**0.3064**	0.4686
SLWWW-C-NU-Base-3	0.2878	0.2799	0.4285
SLWWW-C-NU-Base-4	0.2961	0.2912	0.4365
baseline	**0.3208**	0.2486	**0.5311**^{*3,*4}

underestimated

Condensed-list measure scores after removing Topic 0033

run	Mean MSnDCG@10	Mean Q@10	Mean nERR@10
SLWWW-C-NU-Base-1	0.3177	0.3058	0.4715
SLWWW-C-NU-Base-2	0.3196	0.3064	0.4686
SLWWW-C-NU-Base-3	0.2878	0.2799	0.4285
SLWWW-C-NU-Base-4	0.2961	0.2912	0.4365
baseline	**0.5259**^{*1*2*3,*4}	**0.5068**^{*1*2*3,*4}	**0.6999**^{*1*2*3,*4}

overestimated

Outlines

Conclusions

- Applied query expansion based on centroid and combMax methods.

- For the regular evaluation, Base-3 and Base-4 statistically significantly underforms the baseline in terms of Mean nERR@10 (underestimated)

- Based on condensed-list measures, all four submitted runs statistically significantly underform the baseline.(overestimated)

- The true effectiveness scores for the baseline run should lie somewhere between the regular evaluation and the condensed-list measures evaluation

Reference

- [1] S. Kuzi, A. Shtok, and O. Kurland. Query expansion using word embeddings. In Proceedings of ACM SIGIR 2016, pages 1929–1932, 2017.

- [2] C. Luo, T. Sakai, Y. Liu, Z. Dou, C. Xiong, and J. Xu. Overview of the NTCIR-13 we want web task. In Proceedings of NTCIR-13, 2017.

- [3] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. https://arxiv.org/abs/1301.3781, 2013.

- [4] T. Sakai. Alternatives to bpref. In Proceedings of ACM SIGIR 2007, pages 71–78, 2007.

- [5] T. Sakai. Metrics, statistics, tests. In PROMISE Winter School 2013: Bridging between Information Retrieval and Databases (LNCS 8173), pages 116–163, 2014.

Thank you for

- Organizers of NTCIR13