

SLWWW at the NTCIR-13 WWW Task

Peng Xiao , Lingtao Li , Yimeng Fan and Tetsuya Sakai
Waseda University

Objective

For given query, we try to utilize two query methods to improve the search effectiveness.

Methods

- Centroid method

The centroid method represents q by:

$$S_{Cent}(t; q) \stackrel{\text{def}}{=} \exp(\cos(\vec{t}, \vec{q}_{Cent})) .$$

$$\vec{q}_{Cent} \stackrel{\text{def}}{=} \sum_{q_i \in q} \vec{q}_i .$$

\vec{t} : the L2-normalised Word2Vec vector representing term t .

q_i : query q 's i -th query term.

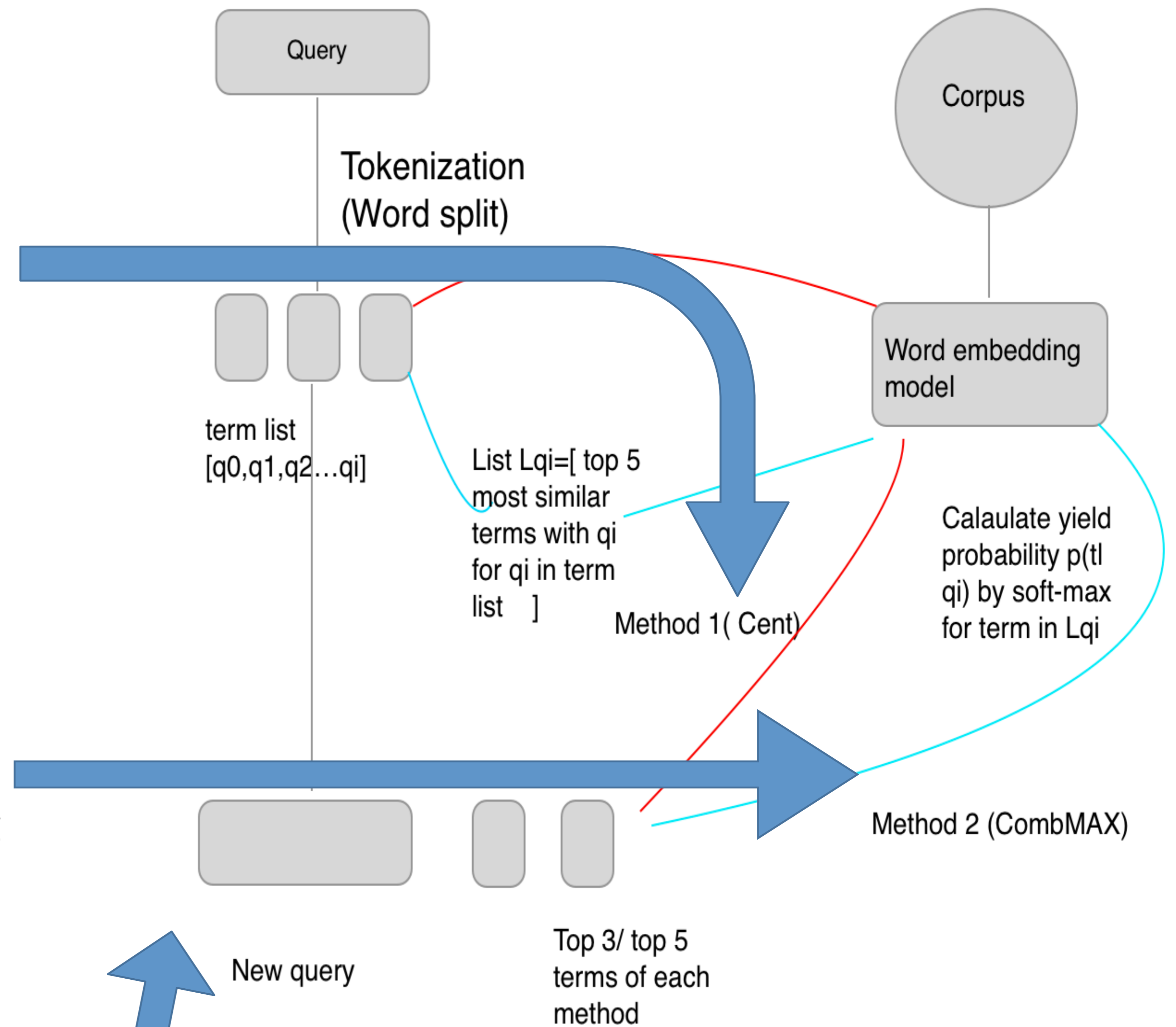
- CombMAX method

Select 5 most similar terms for each q_i according to $\cos(\vec{q}_i, t)$.

Similarities are softmax-normalised:

$$S_{CombMAX}(t; q) \stackrel{\text{def}}{=} \max_{q_i \in q} p(t|q_i) .$$

$$p(t|q_i) \stackrel{\text{def}}{=} \frac{\exp(\cos(\vec{q}_i, \vec{t}))}{\sum_{t' \in L_{q_i}} \exp(\cos(\vec{q}_i, \vec{t}'))} .$$



Query expansion

Select top 3, 5 terms according to the term selection scores, and confuse it with maximum likelihood estimate.

Data

SogouT16

Num of documents	Num of test topics
81,264	100

Results

Submitted runs

Table 1: Run descriptions

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	0.5341 *3,*4

Official results and baseline results. Pity..

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

Remove the topic 33 cause it lacks judged non relevant documents.

Conclusion

Applied the query expansion methods based on word embedding. But the result of the experiment is not ideal.

Further work

Try other models to utilize the user click information.