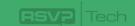
UWNLP at the NTCIR-12 Short Text Conversation Task

Anqi Cui (caq@uwaterloo.ca), Guangyu Feng (gfeng@uwaterloo.ca), Borui Ye (b7ye@uwaterloo.ca), Kun Xiong (xiongkun04@gmail.com), Xingyi Liu (liuxingyi99@gmail.com), Ming Li (mli@uwaterloo.ca)



University of Waterloo, Canada



his is a joint work with RSVP Technologies Inc., a cutting-edge Al and NLP R&D start-up (www.rsvptech.ca).

BACKGROUND

"Away from the city, so you can find your inner peace."









METHOD



1. Retrieve relevant posts

- Combine comments to their corresponding posts as documents
- Index all the post texts, with Apache Solr.
- Retrieve documents based on the keywords of the query.
- Relevant but may not appropriate as a conversation response.



3. Rank the document candidates

- (1) Length: Output longer comments in all retrieved candidates.
- (2) Max comment sim.: Output all comments $\{c\}$ of the best post p, where $score(p)=sim(q,p)+\max\{sim(q,c)\}$
- (3) Combined sim.: Output the best comments where score(c) = sim(q,p) + sim(q,c)



2. Score the posts & comments

Model

- Linear combination.
- Random forest*

Features

- Character-based: Length of the longest common substring, overlapping of any character.
- Word-based: Cosine similarity, overlap similarity, word order similarity, inverse document frequency scores*, latent semantic analysis similarity, Word2Vec similarity**.
- * Trained with 3,809 sentence pairs from search query logs, labeled manually
- ** Trained with 200 million question-answer pairs from Baidu Zhidao and Sogou Wenwen.

RESULTS

Ranking Models and Comment Selection Methods of the Submitted Runs

Run ID	Ranking model	Comment selection	
uwnlp-C-R1	Linear combination Random forest	Combined sim.	
uwnlp-C-R2		Max comment sim.	
uwnlp-C-R3		Length	
uwnlp-C-R4		Combined sim.	
uwnlp-C-R5		Max comment sim.	

Evaluated Results of the Submitted Runs

	Run ID	Mean nDCG@1	Mean P+	Mean nERR@10	
	uwnlp-C-R1	0.2767	0.4284	0.4095	
	uwnlp-C-R2	0.2767	0.3977	0.3740	
	uwnlp-C-R3	0.1733	0.2564	0.2255	
	uwnlp-C-R4	0.1033	0.2085	0.1867	
	uwnlp-C-R5	0.1067	0.1862	0.1732	

Number of Test Topics Containing Good Comments

Run ID	#L2@1	#L1+@1	#L2@10	#L1+@10
uwnlp-C-R1	17	25	64	93
uwnlp-C-R2	17	25	49	80
uwnlp-C-R3	8	12	32	59
uwnlp-C-R4	8	7	35	61
uwnlp-C-R5	8	8	30	53
Voting	19	42	47	79

Voting Select 10 comment candidates from each of the 5 runs, add up their rankings, and use the sum as the final ranking measure.

CONCLUSIONS

- The linear combination model works better than the random forest model, mainly because of the different distribution between the training data and test data.
- The combined similarity method works better, showing that comments also provide useful semantic information with the query.
- Longer comments carry irrelevant information topically.
- The merge-and-vote strategy finds out the top candidate better, but not for the top-ten candidates.