Université de Montréal at the NTCIR-11 IMine Task

Arbi BOUCHOUCHA, Jian-Yun NIE and Xiaohua LIU

Dept. of Computer Science and Operations Research
RALI, Université de Montréal
Québec, Canada
{bouchoar, nie and liuxiao}@iro.umontreal.ca

Embedding Framework: Proposed Approach

Objective Function

$$\min \sum_{r \in R} \sum_{e_i, e_j \in E, i \neq j} \frac{1}{2} . w_r . \left(sim\left(\vec{e_i}, \vec{e_j}\right) - sim_r\left(e_i, e_j\right) \right)^2$$

subject to $\|\vec{e}\|_2^2 \leq 1, e^k \geq 0, k = 1,2,\ldots,N, \forall e \in E$

Maximal Marginal Relevance-based Expansion (MMRE) †

$$\vec{e^*} = argmax_{e \in E-ES} \left(\beta . sim(\vec{e}, \vec{q}) - (1-\beta) . max_{\vec{e'} \in ES}\, sim(\vec{e}, \vec{e'}) \right)$$

where $sim\left(\vec{e_i}, \vec{e_j}\right) = \vec{e_i} . \vec{e_j} = \sum_{k=1,\ldots,N} e_i^k . e_j^k$ for any two vectors $\vec{e_i}$ and $\vec{e_j}$

Resource-based Similarity Functions

- For Wikipedia, ConceptNet, query logs and documents feedback: same definitions as in Bouchoucha et al. †
- For query suggestions:

$$sim_{QS}\left(e_i, e_j\right) = \frac{2 . n\left(e_i, e_j\right)}{n\left(e_i\right) + n\left(e_j\right)}$$ where $n(e_i)$ (resp. $n(e_j)$) is the number of times term e_i (resp. e_j) appears in query suggestions, and $n(e_i, e_j)$ is the number of times when both expansion terms e_i and e_j appear in the suggestions of the same query.

Our Participation: Subtopic Mining & Document Ranking subtasks

Subtask 1: Subtopic Mining (English)

- Query Classification:

- 33 features derived from the resources

Eg. NumTerms, ClarityScore, WikiLength, ClickEntropy, AvgCommonNodes, etc.

- Query Disambiguation and Predicting Subtopic Importance:

- Disambiguate each ambiguous query to generate the first-level subtopics (query interpretations or sub-queries) using Wikipedia + query logs

Subtask 2: Document Ranking (English)

- Selective diversification (depending on the class of the query)
  - **Clear query**: No need for diversification (results of the baseline)
  - **Broad query**: Search results of the expanded query (expansion terms are those obtained by the embedding framework)
  - **Ambiguous query**: Greed selection of documents from different sets

$$d^* = \arg\max_{d \in \cup D_i - S} \left(\frac{rel(d) . score_q(i)}{rank(d)} \right)$$

where D_i is the set of document corresponding to the i[th] sub-query

$$score_q(i) = \frac{n_q(i)}{|E_q|}$$

is the score of the i[th] sub-query of *q*, where E_q is the set of terms added to reformulate *q*, and $n_q(i)$ is the number of terms from E_q that appear in the i[th] Wikipedia page of *q*.

Submitted Runs

- **1st run**: (UM13-S-E-1A ; UM13-D-E-1A) - five resources
- **2nd run**: (UM13-S-E-2A ; UM13-D-E-2A) - four resources (discard query suggestions)
- **3rd run**: (UM13-S-E-3A ; UM13-D-E-3A) - one single resources (query logs)

Experiments and Results

Used Resources

- Five different resources: English Wikipedia dumps of July 8th, 2013; MSN query logs of 2006; ConceptNet 5; Top 50 feedback documents; and query suggestions from Bing, Google and Yahoo!

Query Classification

- We used SVM-Light tool for non-linear SVM (with RBF kernel)
- 450 training queries publicly available in http://www.ccc.ipt.pt/~ricardo/datasets/GISQC_DS.html

Results

Table 1: Query Classifier Performance

Query class	Precision	Recall	F1
Ambiguous	75.00%	56.30%	64.30%
Broad	44.83%	76.47%	56.52%
Clear	88.89%	47.06%	61.54%

Table 2: Overall subtopic mining results

runs	Hscore	Fscore	Sscore	H-measure
1st run	0.2056	**0.1624**	**0.0059**	0.0047
2nd run	**0.2064**	**0.1624**	**0.0059**	**0.0049**
3rd run	0.1766	**0.1624**	0.0049	0.0037

Table 3: Overall document ranking results

runs	Coarse-grain	Fine-grain
1st run	**0.6254**	**0.5566**
2nd run	0.6001	0.5309
3rd run	0.4474	0.3770

Table 4: Coarse-grain results (first-level subtopic)

runs	AP	RBP	nDCG	ERR	I-rec	D#-nDCG
1st run	**0.5479**	**0.1655**	**0.5108**	0.4236	**0.7899**	**0.6511**
2nd run	0.4782	0.1489	0.4750	**0.4251**	0.7520	0.6137
3rd run	0.2520	0.1025	0.3162	0.2880	0.5692	0.4397

Table 5: Fine-grain results (second-level subtopic)

runs	AP	RBP	nDCG	ERR	I-rec	D#-nDCG
1st run	**0.5479**	**0.1480**	**0.4629**	**0.2628**	**0.6310**	**0.5469**
2nd run	0.4782	0.1340	0.4301	0.2602	0.5874	0.5089
3rd run	0.2520	0.0915	0.2901	0.1807	0.3798	0.3331

Discussions

- **Table 1**: Our classifier succeeds to correctly classify about 60% of the queries
  - Failure to distinguish between broad and clear queries
- **Table 2**: 1st and 2nd run lead to very comparable results

➔ Query suggestions seems to be not helpful to improve results! (need further investigation in the future to confirm…)

- **Table 2**: 1st and 2nd run vs. 3rd run

➔ Using multiple resources yields to improved results compared to using a single one

- **Table 3**, **4** and **5**: Best performance in document ranking is obtained by our 1st run

➔ Combining multiple resources yields to a better coverage of query aspects.

Conclusions

- We experimented a new approach for DQE which learns query aspects by selecting (good) expansion terms from different resources.
- Our best document ranking run in IMine task is ranked **No. 2 of all 15 runs** in terms of coarse-grain and fine-grain results.

† A. Bouchoucha, X. Liu, and J.-Y. Nie. "Integrating Multiple Resources for Diversified Query Expansion". In *Proc. of ECIR*, pp. 437-442, Amsterdam, Netherlands, 2014.