哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY 深圳研究生院 Shenzhen Graduate School 智能计算研究中心 (Intelligent Computing Research Center, ICRC)

HITSZ-ICRC at NTCIR-11 Temporalia Task

Yongshuai Hou, Cong Tan, Jun Xu,Youcheng Pan, Qingcai Chen and Xiaolong Wang
Key Laboratory of Network Oriented Intelligent Computation
Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China
houyongshuai@hitsz.edu.cn

Introduction

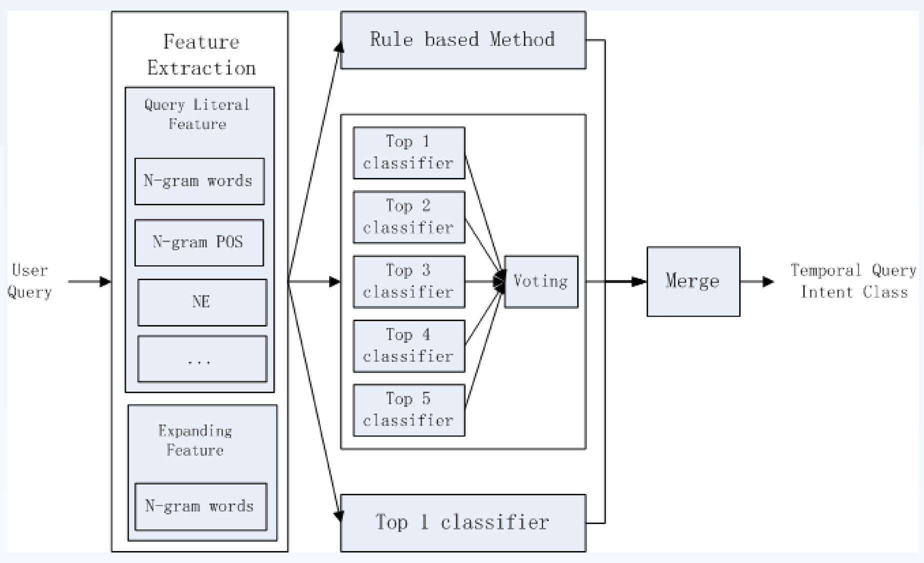
This paper presents methods HITSZ-ICRC group used in the NTCIR-11 Temporalia challenge at NTCIR-11.

- Rule based method and multi-classifier voting method were used for TQIC subtask
- Relevant score weight sum and learning to rank method were used for TIR subtask

Temporal Query Intent Classification

Temporal query intent classification method:

- Rule based
  Use Rule set to classify queries that contain time word or time-sensitive word, like "Movies *2012*", "*long term* weather *forecast*".
- Machine learning
  Train classifier for temporal query using machine learning algorithms , including logistic regression, SMO, HNB, etc.
- Multi-Classifier voting
  Vote results from the top N best classifiers trained
- Result merging
  Merge results gotten from rule based method and multi-classifier voting method



Temporal query intent classifying process

Feature List used:
N-gram term of query
POS n-gram
Named entity
Normalized date
Date distance
Time-sensitive word
N-gram term of SE result

Temporal Information Retrieval

Temporal Information Retrieval method:

- Candidate relevant document retrieving
  Index document set and retrieve search topic using BM25 language model
- Temporal relevant
  Judge temporal relevant between temporal search subtopic and relevant document base the date distance of search date and time expression in the document

$$dis_i = Dq - DX_i$$

$$C_i = \begin{cases} future, & if\, dis_i < 0 \\ past, & if\, dis_i > B_p \\ recency, & if\, 0 \le dis_i \le B_r \end{cases}$$

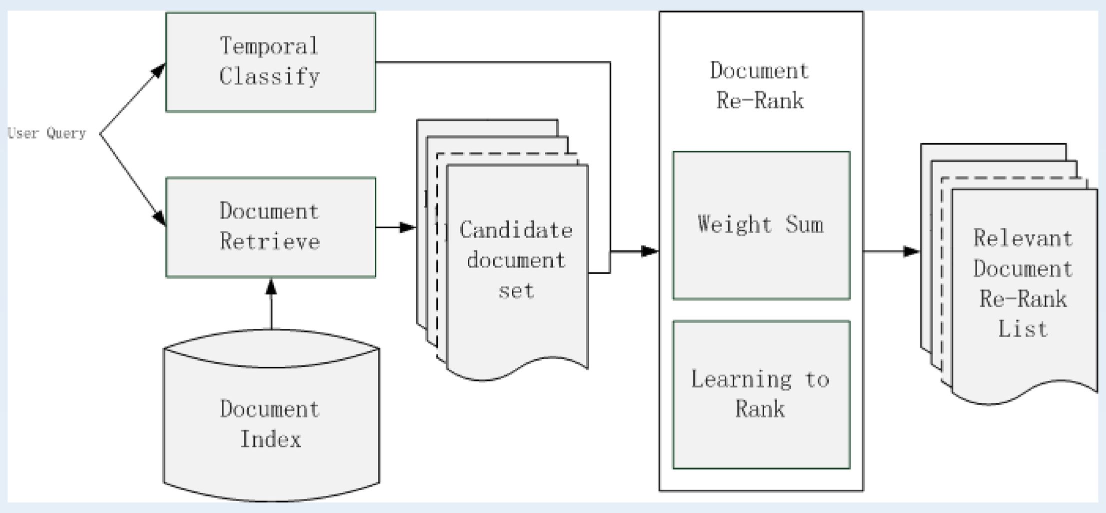
Where *Dq* is search date of the topic, DX_i is normalized time expression in document, B_p is the classification boundary for *past* class time expression, B_r is the classification boundary for *recency* class time expression. B_p=B_r=300 (days) here.

- Temporal search subtopic recognition
  Use the method in subtask TQIC to recognize temporal search subtopic class
- Relevant document re-ranking
  Re-rank relevant document list base the temporal relevant and content relevant between search subtopic and candidate document
  Two methods were tried:
  - Relevant score weight sum

$$R = \alpha R_c + (1-\alpha) R_t$$

Where *R* is the document final relevant score to the search subtopic, R_c is the content relevant score, R_t is the temporal relevant score, α is the weight coefficient and α≥0, α≤1.

- Learning to rank
  Use date distance as rank feature for learning to rank method
  Feature List for learning to rank:
  ➢similarity between search topic and document title
  ➢similarity between search topic and document content
  ➢similarity between search subtopic and document title
  ➢similarity between search subtopic and document content
  ➢BM25 relevant score between search topic and document
  ➢BM25 relevant score between search subtopic and document
  ➢temporal relevant score



Temporal information retrieval process

Table 1. Coefficient value for relevant score weight sum method

Subtopic Class	Coefficient
past	0.85
recency	0.73
future	0.76
atemporal	1

Evaluation

Data set

- "LivingKnowledge news and blogs annotated sub-collection" document corpus, contains 3.8M documents from blogs and news sources.
- 300 formal run queries for TQIC testing
- 50 formal run search topics each with 4 subtopics for TIR testing

Evaluation results of subtask TQIC

Table 2. Results evaluation of TQIC formal runs

runID	Correct Number	Precision
PrW	207	**69 %**
PrWsQW	203	67.67%
qRPrHNB	201	67%

Table 3. Precision of each class in TQIC formal runs

runID	atemporal	future	past	recency
PrW	**70.67%**	64%	78.67%	**62.67%**
PrWsQW	69.33%	66.67%	77.33%	57.33%
qRPrHNB	57.33%	**68%**	**81.33%**	61.33%

- Run *PrW*: merged results of PRISM rule set and multi-classifier voting method. The features used are expanding features from SE.
- Run *PrWsQW*: merged results of PRISM rule set and multi-classifier voting method. The features used include query literal features and expanding features from SE.
- Run *qRPrHNB*: merged results of manual rule set, PRISM rule set and HNB classifier.

Evaluation results of subtask TIR

Table 4. Results evaluation of TIR subtask runs

runID	nDCG @20	AP@20	P@20	nERR @20
BW	0.4544	0.4587	0.5895	0.6056
BWCC	0.4554	0.4599	0.5902	0.6064
LTRNC2	**0.4768**	**0.483**	**0.6018**	**0.6313**

Table 5. nDCG@20 of each class in TIR formal runs

runID	atemporal	future	past	recency
BW	0.4669	0.4607	0.4005	0.4897
BWCC	0.4678	0.4593	0.403	0.4915
LTRNC2	**0.5092**	**0.4804**	**0.4227**	**0.495**

- Run *BW*: used the relevant score weight sum method to re-rank, and used the original class information of search subtopic.
- Run *BWCC*: used the relevant score weight sum method to re-rank, use the classifying result of classifier in subtask TQIC as subtopic class, and did not use the original class information of search subtopic.
- Run *LTRNC2*: used learning to rank method (LambdaMART algorithm here) to re-rank, and did not use the original class information of search subtopic.

Conclusion

Merging results of rule based method and multi-classifier voting is effective for TQIC subtask.

Both relevant score weight sum method and learning to rank method are effective for TIR subtask, and learning to rank method is more effective here.