Using Time-Series for Temporal Intent Disambiguation in NTCIR-12 Temporalia

Dan Li, Xiaoxia Liu, Yunxia Zhang, Degen Huang, Jingxiang Cao

Dalian University of Technology, China

linda_2013, liuxxivy, zhangyunxia@mail.dlut.edu.cn, huangdg, caojx@dlut.edu.cn

Abstract

We propose a method that integrating text features and time-series features for Temporal Intent Disambiguation task. The AvgCosine and AvgAbsLoss of our best run reach 0.732 and 0.210 on the 300 samples of NTCIR-12 Temporalia .

Introduction

• Problems

Traditional TQIC methods mainly use text features of queries to predict the underlying temporal intents. However, this is not sufficient due to:

1) The challenge that queries with explicit temporal expressions are few (only 1.5%) for TQIC.
2) The limited temporal information mined from languages that are lack of morphologic changes.

• Motivations

Human discovers temporal intents based not only on surface query texts, but also on **the timelines of target queries in their minds**. Hence a new source of temporal information is needed for TID.

Method

• Architecture

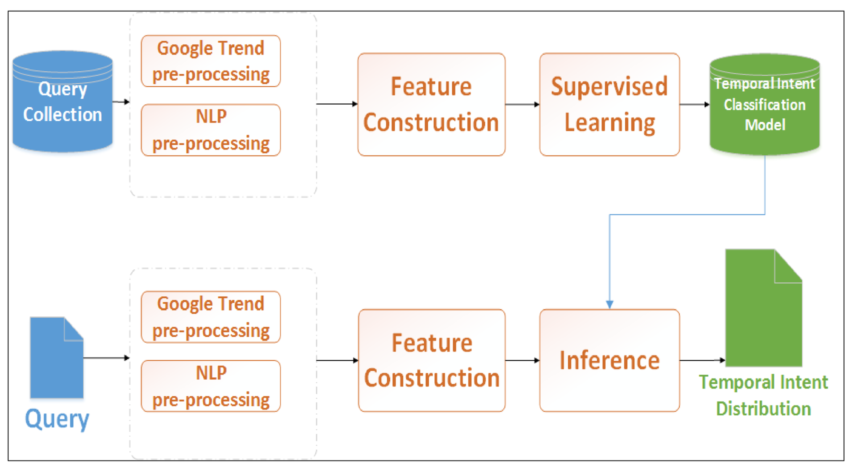

Figure 1. The architecture to model temporal intent category distribution

• Pre-processing

1. NLP pre-processing
   – Stanford CoreNLP Package for tokenization, lemmatizing, POS tagging, parsing
   – SUTime for temporal expression recognition
2. GT pre-processing
   – Time Domain Data (TDD) crawled from Google Trends
   – Frequency Domain Data (FDD) made by *periodogram* in TSA package in R

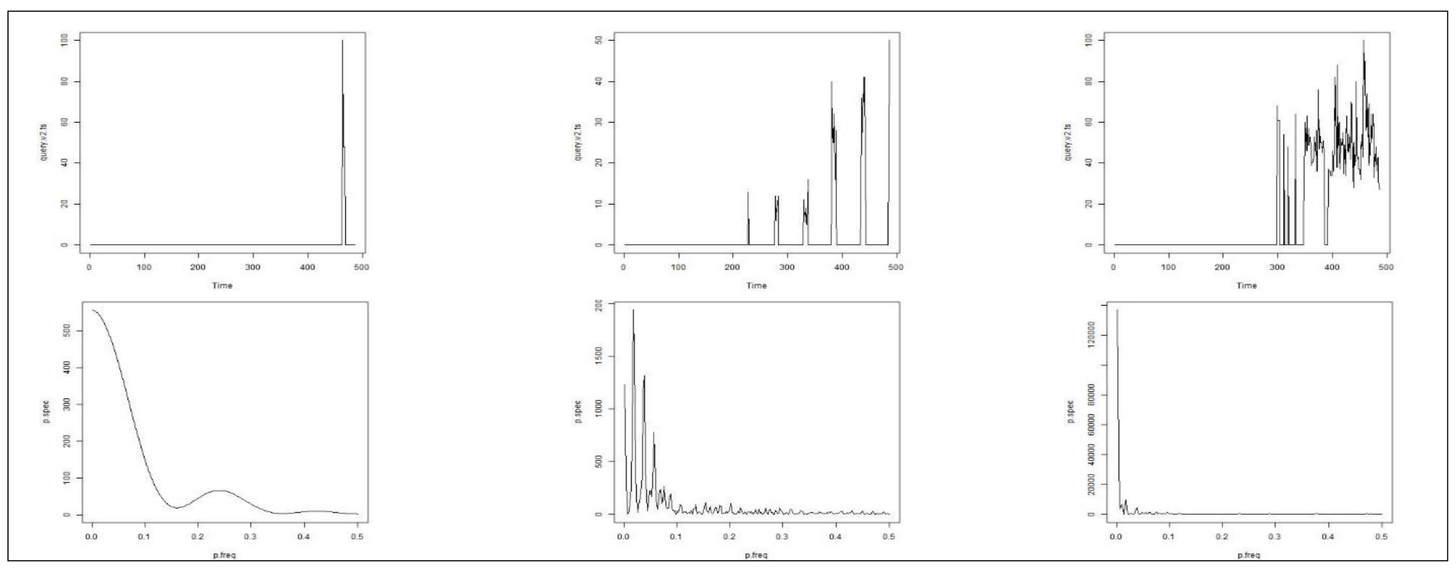

Figure 2. Examples of pairs of TDD (first line) and FDD (second line). The horizontal axis of TDD is time (in week), and the vertical axis is standardized search frequency from 0 to 100. The horizontal axis of FDD is frequency (in Hz), and the vertical axis is the spectral density at corresponding frequencies. The three query examples are "end of twinkies", "nba playoff's scores", and "when was America discovered".

• Feature description

Table 1. Five feature groups

Feature Group	Feature	Value	Description
Trigger word	*whether_past* *whether_recency* *whether_future* *whether_atemporal*	Boolean	Four groups of trigger word sets
Word POS	*head_word_POS* *verb_tense*	Adapted Stanford POS tag set	Head words obtained from Stanford dependency parsing result
Word temporal probability¶	*past probability* *recency probability* *future probability* *atemporal probability*	Real number: [0,1]	Bayesian Probability
Explicit time gap	*explicit_time_gap_past* *explicit_time_gap_recency* *explicit_time_gap_future*	Boolean	Subtraction of the standardized date from the issue date
Implicit time gap (Time-series)	*rule-based time gap*¶ *rule_based_past* *rule_based_recency* *rule_based_future*	Booolean	Subtraction of the estimated date from the issue date
	time-series statistics¶ *max, mean, variation, sr, mr, stridency*	Real number: (-∞,+∞)	Statistics discriminating the four categries

Word temporal probability¶

The temporal intent category distribution of a word w :

$$p(c_i|w) = \frac{p(w|c_i)p(c_i)}{p(w)} = \frac{p(w|c_i)p(c_i)}{\sum_{j=1}^{4} p(w|c_j)p(c_j)}, i = 1,2,3,4.$$

The temporal intent category distribution of a query q:

$$p(c_i|q) = \sum_{w \in q} p(c_i|w), \qquad i = 1,2,3,4.$$

Rule-based time gap¶

Input TDD and FDD of a query; **Output** time gap of a query

```
1. DETECT_PERIOD(TDD, FDD)
2. if periodical:
3.     time gap = COMPUTE_MAXIMUM_POINT(TDD, FDD)
4. else if occasional:
5.     time gap = COMPUTE_PEAK_POINT(TDD, FDD)
6. Else if atemporal:
7.     time gap = None
```

Time-series statistics¶

$$sr = \frac{f_{max} - \max(\{f_1, f_2, \cdots, f_T\} - \{f_{max-d}, \cdots, f_{max-1}, f_{max}, f_{max+1}, \cdots, f_{max+d}\})}{\sum_{t=1}^{T} f_t}$$

$$mr = \frac{f_{max}}{\sum_{t=1}^{T} f_t}$$

$$stridency = \frac{f_{max}}{\sum_{f_t > \delta f_{max}} f_t}$$

• Classification Model

Logistic Regression model in Scikit-learn package

Experiments

• Data

Training set: 93 samples from the dry run of NTCIR-12 Temporalia and 300 samples from the formal run of NTCIR-11 Temporalia

Test set: 300 samples from the formal run of NTCIR-12 Temporalia

• Runs

In RUN1, four groups of features are used including trigger word, word POS, explicit time gap, word temporal probability. In RUN2, we add feature rule-based time gap based on RUN1. In RUN3, we replace feature rule-based time gap with time-series statistics.

Table 2. Average cosine and average absolute loss for RUN1, RUN2, and RUN3

	AvgCosine	AvgAbsLoss
RUN1	0.728	0.208
RUN2	**0.732**	0.210
RUN3	0.727	0.212

- The feature time-series statistics and rule-based time gap do not help much here. It is a challenge but a chance to improve time-series features in the future.

Table 3. Confusion matrix for RUN1, RUN2, RUN3 and manual results compared with the standard result

		STANDARD			
		P	R	F	A
RUN1	P	34	2	10	25
	R	4	18	6	25
	F	1	4	39	14
	A	5	10	8	95
RUN2	P	33	3	11	32
	R	4	17	6	33
	F	1	4	39	13
	A	6	10	7	81
RUN3	P	34	2	11	32
	R	4	17	5	25
	F	1	4	40	14
	A	5	11	7	88
MANUAL	P	42	3	3	15
	R	0	23	6	8
	F	0	3	37	6
	A	2	5	17	130

- The high performance of manual result indicates that TID task is reasonable and is hopeful to be further improved. The major errors and challenges are related to *atemporal* category.

Conclusion

We participated in the TID subtask of the NTCIR-12 Temporalia task and submitted three runs. The temporal intent probability distribution of four categories of the 300 test queries are predicted through logistic regression in all the three runs. In RUN1, four groups of features are used including trigger word, word POS, explicit time gap, temporal probability of words. Implicit time gap is added in the form of rule-based time gap in RUN2 and in the form of time-series statistics in RUN3. RUN2 performs slightly better than the rest two runs with AvgCosine of 0.732 and AvgAbsLoss of 0.210.

The future improvement may lie in three directions. First, discriminate the atemporal queries with temporal ones in the early step and classify the rest three ones later. Second, extend the probability distribution of the word to its synonyms using methods like WordNet or word embedding. Finally, preprocess the time-series data of queries by removing long-term trends and seasonal changes.

Acknowledgements

This work was supported by the National Nature Science Foundation of China (No. 61173100 61272375) and National Social Science Foundation of China (No. 15BYY175).