NTCIR-12
with the 7th International Workshop on EVIA
Tokyo, 7-10 June, 2016

WHUIR at the NTCIR-12 Temporalia Temporal Intent Disambiguation Task

Sisi Gui
School of Information Management, Wuhan University, China
sgui0229@whu.edu.cn

The Task

The huge volume of web pages makes the **time** an important factor in **information retrieval**. Using temporal information, e.g. the temporal intent of user's queries, may improve IR systems' accuracy.

This task is to **estimate a distribution of four temporal intent classes** (Atemporal, Past, Recent, or Future) for a given query.

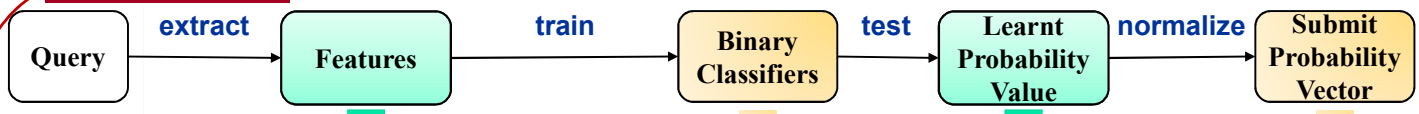
Observation

- Query logs are difficult to obtain
- Previous work shown results on queries features outperformed better than features from retrieved docs

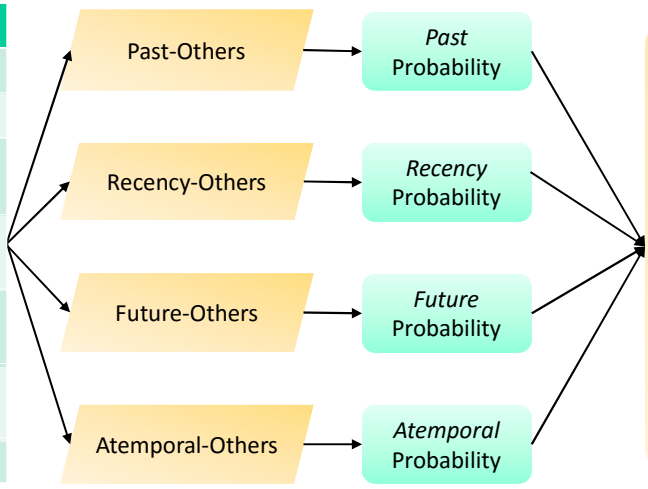
Main Idea

- Extract features from queries only
- Machine Learning + Multi-Class SVR

Research Design



Features	Notes
Name Entity	location, person, organization
Query Length	without stop-words removal
Year Number	words to express year; e.g. "2007","2013"
Core Verb Tense	verb reflecting the really tense of the whole sentence
Domain Keywords	the most frequent words shown in certain category
Time Gap	time interval between query focus time and submitted time
Temporal words	TempoWordNet



$\mathbf{V} = (p_1, p_2, p_3, p_4)$

$$\sum_{i=1}^{4} p_i = 1$$

p_i : query's normalized probability value for *Past*, *Recency*, *Future*, *Atemporal* respectively

Experiment

- **Dataset**: TID English dataset
  - Dry Run: 73 training queries + 20 test queries
  - Formal Run: 300 test queries
- **Off-the-shelf Tools**
  - Stanford NLP
  - TempoWordNet
  - LIBSVM
- **Results**

RUN ID	svm type	kernel type	Cosine Similarity	Per-class Absolute Loss
WHU1	epsilon-SVR	sigmoid	0.6196	0.2662
WHU2	nu-SVR	linear	0.5225	0.2921
WHU3	nu-SVR	polynomial	0.6933	0.2520

- **Best Run Discussion**
  - *Atemporal* is hard to estimate
  - estimate more probability values to some 0 probability value temporal class

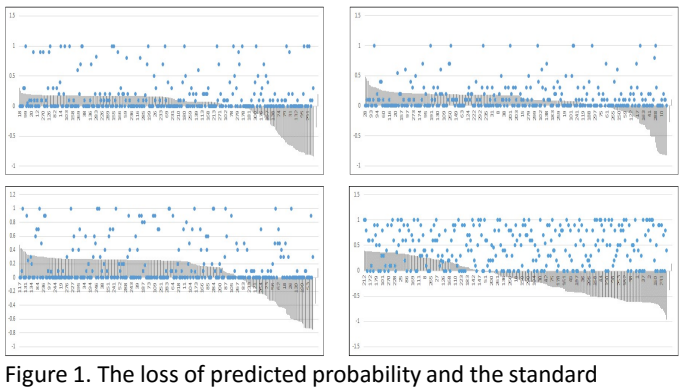


Figure 1. The loss of predicted probability and the standard probability for four categories