



Learning Temporal Ambiguity in Web Search Queries

Conference on Information and Knowledge Management (CIKM 2017)
November 6-10 2017 | Singapore

	Behrooz Mansouri	M.Sadegh Zahedi	Maseud Rahgozar	UNIVERSITY OF WOLLONGONG IN DUBAI 	Farhad Oroumchian		Ricardo Campos
	University of Tehran, Iran				University of Wollongong in Dubai, UAE		Polytechnic Institute of Tomar, INESC TEC – LIAAD, Portugal
	b.mansouri@ut.ac.ir	s.zahedi@ut.ac.ir	rahgozar@ut.ac.ir		farhadoroumchian@uowdubai.ac.ae		ricardo.campos@ipt.pt

Introduction

Time has strong influence on web search. The temporal intent of the searcher adds an important dimension to the relevance judgments of web queries. However, lack of understanding their temporal requirements increases the ambiguity of the queries, turning retrieval effectiveness improvements into a complex task. In this paper, we propose an approach to classify web queries into four different categories considering their temporal ambiguity. For each query, we develop features from its search volumes and related queries using Google trends and its related top Wikipedia pages. Our experiment results show that these features can determine temporal ambiguity of a given query with high accuracy. We have demonstrated that a Multilayer Perceptron Networks can achieve better results in classifying temporal class of queries in comparison to other classifiers.

Approach

To detect the different types of seasonal queries, we expand the queries with two types of features: (i) time-series; (ii) content features; (iii) Wikipedia feature.



Time Series Features



Related Queries Features



Wikipedia Feature

Time series built on query frequency using  from 2004 to 2017.

Related queries using Google Trends from 2004 to 2017.

Top-20 related Wikipedia pages

- Autocorrelation
- Seasonality
- Kurtosis
- Randomness Test
- Sum of squared errors
- Modality
- Mean

- The ratio of related queries containing a year and the number of total related queries
- The number of unique years mentioned in related queries.

- The number of year expressions in the name of top related pages

Results

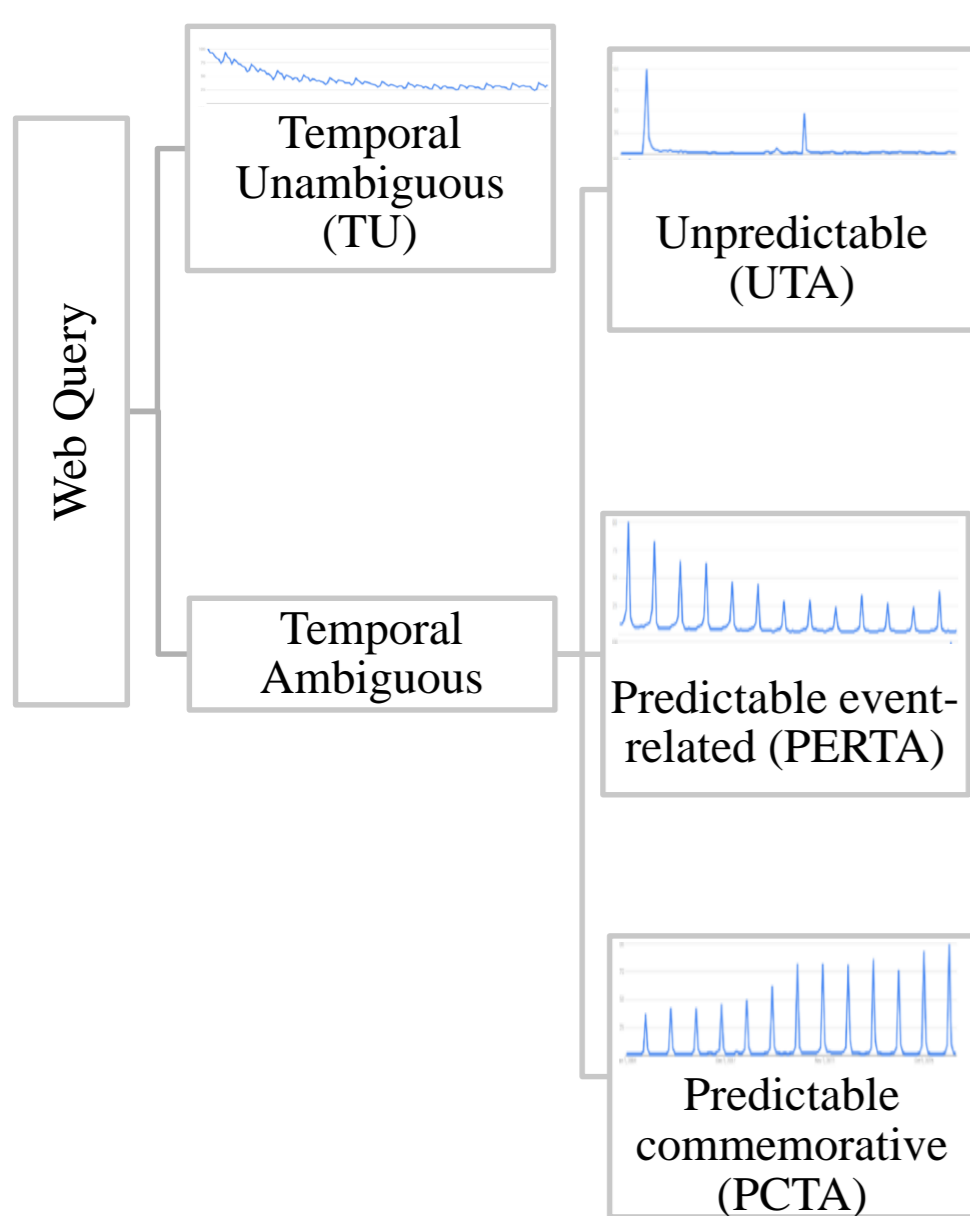
Model	Precision	Recall	F-measure
MLP	0.868	0.868	0.867
SVM	0.773	0.766	0.758
Random Forest	0.815	0.820	0.817
Adaboost	0.793	0.836	0.814
Naive Bayes	0.788	0.790	0.788

Performance of different classifiers

Classified \ Real	Classified			
	TU	UTA	PERTA	PCTA
TU	174	7	2	2
UTA	13	87	3	2
PERTA	4	3	75	18
PCTA	4	2	6	98

Confusion matrix for the MLP classifier

Temporal Classification Taxonomy



Category	Examples
TU	Computer Science, Secure passwords, Fashion
UTA	Messi Injury, Tsunami, Tom Hanks movie
PERTA	US presidential election, Golden globe awards, The US Open
PCTA	Santa Claus, Valentine's Day, September 11th, trick or treat

Dataset

500 queries were selected from the works of Gupta and Berberich; and Jones and Diaz, and labeled by 3 professional editors.

<http://dbrg.ut.ac.ir/TemporalAmbiguousQueryDataset/>

Query Class	#Queries	Query Class	#Queries
TU	185	PERTA	100
UTA	105	PCTA	110

Queries Collection Summary.

Literature cited

- Campos, R., Dias, G., and Jorge, A. (2011). What is the Temporal Value of Web Snippets. In WWW-TWAW'11, pp. 9-16.
- Campos, R., Dias, G., Jorge, A., and Jatowt, A. (2014). Survey of Temporal Information Retrieval and Related Applications. In CSUR, 47(2). Article No.: 15.
- Gupta D. and Berberich, K (2015). Temporal Query Classification at Different Granularities. In SPIRE'15, pp. 156-164.
- Jones R. and Diaz, F (2007). Temporal Profiles of Queries. In TOIS, 25(3).
- Metzler, D., Jones, R., Peng, F., and Zhang, R. (2009). Improving Search Relevance for Implicitly Temporal Queries. In SIGIR'09, pp. 700-701.
- Shokouhi M (2011). Detecting Seasonal Queries by Time-Series Analysis. In SIGIR'11, pp. 1171-1172.
- Zhao, Y., and Claudia H. (2016). Temporal Query Intent Disambiguation using Time-Series Data. In SIGIR'16, pp 1017-1020.
- Kulkarni, A., Teevan, J., Svore, K. M., and Dumais, S.T. (2011). Understanding Temporal Query Dynamics. In WSDM'11, pp. 167-176.

Conclusions

In this paper, we proposed an approach for identifying different types of temporal ambiguous queries. We extracted features from search frequency volume and related queries using Google trends data and expanded our queries with these features. As a further additional knowledge, we also used top related Wikipedia pages title in order to extract year expressions for each query. A Multilayer perceptron neural network was used for temporal classification of queries. We have demonstrated that a reasonably good accuracy could be achieved for most of the categories. In future work, we plan to improve our categorization techniques by employing more distinctive features within a web retrieval search engine and by using standard test collections for experimental procedures.

Further information

- www.ccc.ipt.pt/~ricardo/
- <http://dbrg.ut.ac.ir/>

Download this Poster

