WIS @ the NTCIR-12 Temporalia-2 Tasks

Yue Zhao and Claudia Hauff

Web Information Systems Group, TU Delft

{Y.Zhao-1, C.Hauff}@tudelft.nl

Subtasks:
Temporal Intent Disambiguation (TID).

Keywords:
Temporal Intent, Query Intent Disambiguation, Time-series data, Wikipedia

Abstract:
Our approach focuses on the question of whether temporal signals, extracted from publicly available, external data sources (in this case the **Wikipedia page view** stream), as features in a machine learning setup are beneficial for this task.

Intuitions:

Time-series data of queries on Google Trend is a good indicator to show how users' interests of queries change over time. However, the disadvantage is following:

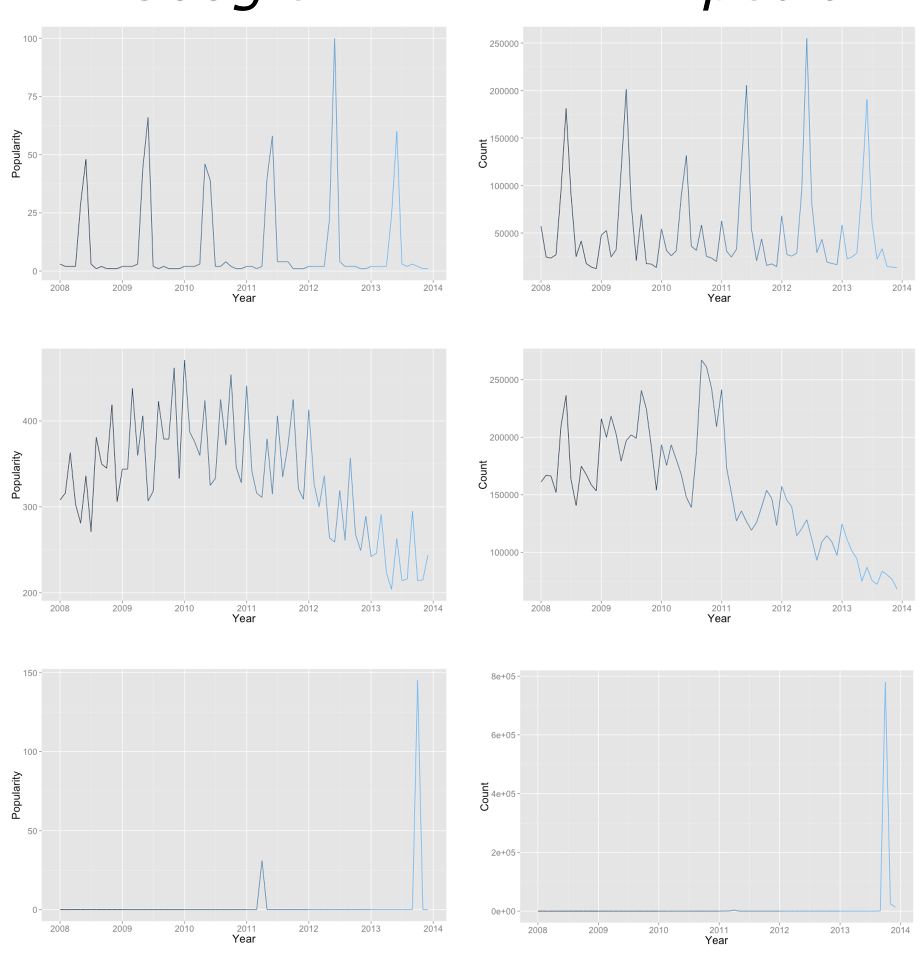
- No absolute frequencies are available
- It is unknown what data pre-processing & cleaning steps occurred
- The aggregations occur at a month-by-month level

Google *Wikipedia*

French Open

Free Dictionary

History of Government Shutdowns



Methods:

Features are extracted from query content, temporal expressions and time-series data of Wikipedia page views of best-match concepts

	Query Content Features
F1	Lemmas
F2	Named Entities
F3	Verb Tense: Uppermost Verb Tense (*UVB_tense*) and Verb Tense with Lemma (*tense_lemma*)
	Temporal Expression Features
F4	ref_{past}: number of TEs referring to past times with respect to the query issue time
	ref_{future}: number of TEs referring to future times with respect to the query issue time
	$same_{Y}$: number of TEs referring to the same year as the query issue time
	$same_{YM}$: number of TEs referring to the same year & month as the query issue time
	$same_{YMD}$: number of TEs referring to the same year & month & day as the query issue time
F5	$lemY_{past}$: number of numerical lemmas referring to past years with respect to the query issue time
	$lemY_{future}$: number of numerical lemmas referring to future years with respect to the query issue time
	$lemY_{same}$: number of numerical lemmas referring to same years with respect to the query issue time
	Time-Series Features
F6	*Sparsity*: indicates whether time-series data exists or not, and whether time-series data is sparse or not
F7	*Seasonality*: represented by the cosine similarity between the time-series data itself and its seasonal component generated through the Holt-Winter decomposition
F8	*Autocorrelation*: measures the periodicity of the time-series data by comparing the past 12 months of data to the same time period a year earlier
F9	$\{ref_{view\_D}, ref_{view\_MD}\}$: difference between the query issue month (month/day combination) and the month (month/day combination) the concept had the most pageviews in our Wikipedia pageview traces
F10	The *MEAN*, standard deviation (*STD*) and *MEDIAN* of the concept's time-series data are also computed

Mapping queries to Wiki concepts: Only the best-match concept is leveraged as we consider it to be the best representative of the entire query.

Probabilistic classification: Query with 4 temporal intents (P=x1, R=x2, F=x3, A=x4) is transformed into 100 sample with single intent setting: $10 \times x_i$ samples for intent i

Runs & Results:

1. The 3 runs submitted by WIS group:
   - **WIS-TID-E-1**: 227 query-content features, PCA with 50 components, Ridge regressor.
   - **WIS-TID-E-2**: query-content features + time-series features, PCA with 50 components, Ridge regressor.
   - **WIS-TID-E-3**: 227 query-content features, PCA with 100 components, SVM with RBF kernels.

2. Results overview of our submitted runs according to the official evaluation metrics.

			Per-Class Absolute Error			
Runs	**Cos Sim**	**MAE**	*Past*	*Recency*	*Future*	*Atemporal*
WIS-TID-E-1	0.792	0.215	0.211	0.154	0.204	0.291
WIS-TID-E-2	0.773	0.219	0.205	0.159	0.206	0.306
WIS-TID-E-3	0.791	0.197	0.151	0.146	0.204	0.288

Results Analysis:

1. What is the effects of features in 3 runs?

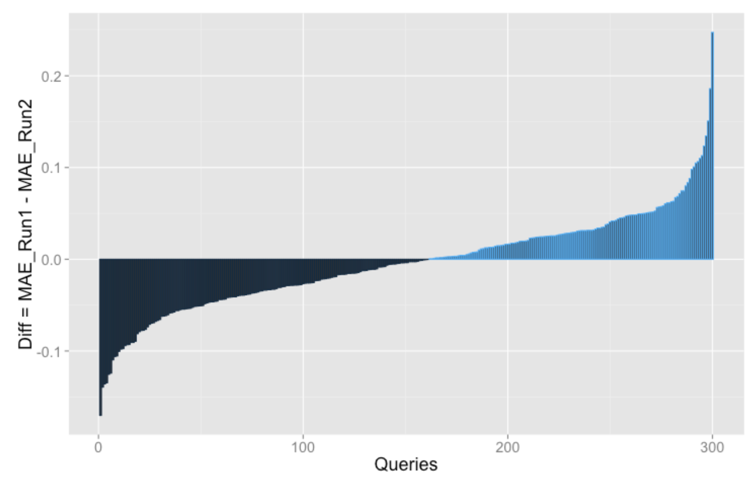
Ablation study of our submitted runs according to the official evaluation metrics.

MAE	**WIS-TID-E-1**	**WIS-TID-E-2**	**WIS-TID-E-3**
Baseline	0.215	0.219	0.197
- *Lemma&NN*	+0.0128	+0.0034	+0.0275
- *TE*	+0.0075	+0.0053	+0.0119
- *Verb*	+0.0052	-0.0007	-0.0103
- *Wiki&Type*	–	-0.0036	–
- *Sparsity*	–	-0.0043	–
- *Season*	–	-0.0011	–
- *AutoCor*	–	+0.0003	–
- *Ref*	–	-0.0002	–
- *Stats*	–	+0.0003	–

Cos Sim	**WIS-TID-E-1**	**WIS-TID-E-2**	**WIS-TID-E-3**
Baseline	0.792	0.773	0.791
- *Lemma&NN*	-0.0437	-0.0085	-0.0614
- *TE*	-0.0208	-0.0186	-0.0314
- *Verb*	-0.0118	+0.0016	+0.0252
- *Wiki&Type*	–	+0.0148	–
- *Sparsity*	–	+0.0147	–
- *Season*	–	+0.0048	–
- *AutoCor*	–	-0.0003	–
- *Ref*	–	+0.0006	–
- *Stats*	–	-0.0010	–

- *MAE* and *Cos Sim* show similar trends in each feature.
- Time-series features reduce the impact of content features
- Features of verb tense play different roles in different models.

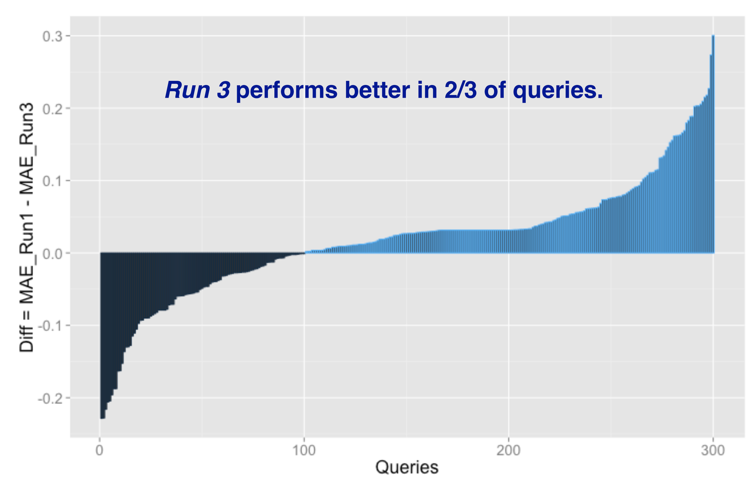
2. What is the impact of temporal features generated from Wikipedia page views?



Queries that WIS-TID-E-2 outperforms WIS-TID-E-3

#069 estimate on the Debt Crisis in Greece
#165 sovereign debt crisis
#021 PS4 Release Date
#131 The first walk into outer space was taken by a Soviet
#012 summer days
#085 timetable bus suzhou
#113 NBA Finals
#104 snake dishes have become popular in recent years
#013 The Following Recap
#092 baseball playoffs

3. What is the impact of predictor choice (regressor v.s. classifier) ?



++++ WIS-TID-E-1 outperforms WIS-TID-E-3 ++++

#106 Barack Obama is the 44th US president
#100 science fiction is a popular
#117 is it easy to find a job in hong kong
#175 How college is different from high school
#009 When Does Time Change

Atemporal **intents are estimated as** ***Future*** **intent**

++++ WIS-TID-E-3 outperforms WIS-TID-E-1 ++++

#069 estimate on the Debt Crisis in Greece
#043 The original building was built in 1710
#219 When Did WW2 Start
#026 when does fall start
#165 sovereign debt crisis

Verb tense **plays an important role in the prediction.**

TUDelft Delft University of Technology

SIKS