Exploring Temporal Information in Search Queries

Xin Kang† and Yunong Wu‡ and Fuji Ren†
†Faculty of Engineering, Tokushima University
‡Department of R&D, Business Big Data Co., Ltd.
kang-xin@is.tokushima-u.ac.jp, wuyunong@brandbigdata.com, renfu@tokushima-u.ac.jp

Abstract
We carefully develop a series of temporal features based on the general knowledge underlying Wikipedia, and construct a deep neural network with a softmax layer for disambiguating people’s temporal intents in web search queries. We analyze the importance of different temporal features, and discuss the impact of neural network structures to the TID results.

Introduction
Web search queries are short (Fig. 1). Table 1 shows some easy and difficult samples for TID. The temporal information about searching events (e.g. iphone 6 and memorial day) which only exists in the human knowledge, must be explored for computers to understand people’s temporal intuitions.

The KGO team carries forward their previous work (as TUTA1) in the NTCIR-11 TQIC subtask, by focusing on the development of temporal features and the construction of a deep neural network with a probabilistic interpretation, to solve the TID problem.

Temporal Feature Extraction

Explicit Temporal Features The Uppermost Verb Tense feature indicates the tense of a query, which could be directly obtained by picking tense information (part-of-speech tag) from the main (uppermost) verb in a query [2].

The Time Gap feature represents the difference between an explicit time expression like year 1974 and the query submission time, with year, month, day, season, and period normalized for time differentiation.

Implicit Temporal Features The Temporal Named Entity feature extracts temporal information of a named entity in query, by exploring the temporal knowledge from Wikipedia.


Step 2. Parse summary and extract the Time Gap features.

Step 3. Resolve the correlation between Time Gap and Named Entity $e$:

$$s(x, e) = \cos \left( \frac{1}{\sqrt{\sum_j v_j + \sum_k v'_k}} \sum_j v_j(x) \sum_k v'_k(e) \right), \tag{1}$$

where $C(t)$ and $C(e)$ are the contexts for $t$ and $e$, respectively, and $v(w)$ is a 1000-dimensional semantic vector generated by a word2vec model [1]. Specifically, we extract the closest predicate to $t$ as its context $C(t)$, and extract words in query except $e$ as its context $C(e)$. C(September 9, 2014) $=$ [unrelated], C(September 9, 2014) $=$ [released], C(Iphone 6) $=$ [release, date, for], and get the correlations.


Step 5. Extract the time feature: memorial day.

Step 6. Extract the country information: default United States.

Step 7. Extract the year information: default query submission year.

Step 8. Query holiday date from HDB.

Step 9. Query submission year.

Step 10. Date differentiation: DIFF Future.

Other Features The People and Time feature clusters people and events based on their semantic similarities in TextRazor Entity Extraction (to decrease the feature space and avoid over-fitting).

What Time did Super Bowl Start? How did Amy Winehouse Die? DECEASED_PERSON, MEASURED_PERSON, PERSON

The Lemma feature normalizes words in a query. Part-of-speech tags in verb lemmas are kept for their tense information.

Neural Network Construction

We construct a deep neural network model to generate the probabilistic temporal predictions

$$\pi_t = \text{softmax}[\mathbf{W}_t]$$

$$H(p, \pi_t) = -\sum_{t \in \mathcal{T}} \pi_t \log \pi_t,$$ \tag{3}

with cross entropy as its cost function

$$\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \text{loss}(p, \pi_t).$$ \tag{4}

and the cosine similarity

$$\text{sim}(p, \pi_t) = \frac{1}{\sqrt{\sum_{i \in \mathcal{T}} p_i^2 \times \sum_{i \in \mathcal{T}} \pi_i^2}}, \tag{5}$$

Then we incorporate the TQIC data (400 samples) into the TID Dry Run set (93 samples) for training deep neural network, assuming that in TQIC there is no possibility mass for temporal labels except the tagged one, and test on the TID Formal Run set (300 samples). Parameters are selected through 5-fold cross-validation, based on the averaged per-class absolute loss

$$\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \text{loss}(p, \pi_t),$$

and the cosine similarity

$$\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \text{sim}(p, \pi_t).$$

Conclusions

• Exploring the temporal information from Wikipedia and resolving the temporal correlation with vectorized context-similarity.

• Deriving the abstract temporal features and generating the probabilistic temporal predictions with a deep neural network.

• Run-2 with fewer layers achieved the best loss score and Run-3 with more layers rendered the highest similarity score.

• Examining the association between temporal features and temporal labels suggested some directions for improvement.

References
[1] Ildio Ewiski word2vec model 1000 dimensions. 2015.