Identifying Time Intervals for Knowledge Graph Facts

Dhruv Gupta
Max Planck Institute for Informatics, Germany
Saarbrücken Graduate School of CS, Germany
dhgupta@mpi-inf.mpg.de

Klaus Berberich
Max Planck Institute for Informatics, Germany
htw saar, Germany
kberberi@mpi-inf.mpg.de

ABSTRACT

Knowledge graphs capture very little temporal information associated with facts. In this work, we address the problem of identifying time intervals of knowledge graph facts from large document collections annotated with temporal expressions. Prior approaches in this direction have leveraged limited metadata associated with documents in large collections (e.g., publication dates) or have limited techniques to model the uncertainty and dynamics of temporal expressions. Our approach to identify time intervals for time-sensitive facts in knowledge graphs leverages a time model that incorporates uncertainty and models them at different levels of granularity (i.e., day, month, and year). Evaluation on a temporal fact benchmark using two large news archives amounting to more than eleven million documents show the quality of our results.

1 INTRODUCTION

Journalists often face the challenging task of substantiating an investigative story with temporal facts in order to convey the validity of certain arguments. The tool of choice for many such scenarios are commercial search engines. The journalist often collects many documents using keyword queries, manually accesses and inspects documents for temporal facts and then compiles a spreadsheet of the observed dates. However, this manual process becomes impossible if the temporal evidence is spread over thousands of document and if multiple facts need to be verified.

In this work, we leverage temporal expressions in document contents to identify time intervals for temporal facts in knowledge graphs. A temporal fact is a sentence that connects an entity (person, organization, or location) via a temporal predicate to a temporal expression. Computational analysis of temporal expressions is challenging as they can be present at different granularities from precise timestamps to coarse level mentions of decades, e.g., the 60s. Furthermore, temporal expressions can be implicit and relative to dates mentioned elsewhere in the text, e.g., last sunday.

In short, our approach overcomes the above challenges as follows. First, we extract temporal facts using information extraction templates that rely on word sequences and temporal expressions. Second, we model the temporal expressions identified for the facts in a time model that understands uncertainty and incorporates temporal granularity. Third and finally, we aggregate the temporal evidence found and rank the time intervals that are of interest to the knowledge graph fact.

2 APPROACH

We next define concrete definitions and steps of our approach.

Temporal Facts. A temporal fact spotted in a document collection consists of natural language representation of a knowledge graph fact with a co-occurring temporal expression. Formally,

$$(\text{f})\textsc{act} \equiv \langle (\textsc{s})\textsc{ubject}, (\textsc{p})\textsc{redicate}, (\textsc{o})\textsc{bject} \rangle [\textsc{time}]. \qquad (1)$$

A natural language representation of a knowledge graph fact $\langle s, p, o \rangle$ is obtained by looking at all possible surface forms of all three arguments of a fact from paraphrase and surface form dictionaries. Thus, to detect a knowledge graph fact we have to detect the equivalent natural language representation:

$$\textsc{f} \equiv \langle [s_1 \ldots s_m], [p_1 \ldots p_n], [o_1 \ldots o_l] \rangle. \qquad (2)$$

Query Processing. For each surface form or paraphrase of the predicate in the natural language representation of the fact (Equation 2) we utilize an inverted index over phrases that detects the positional span of each mention of the named entity or predicate. By intersecting the document identifiers of the mentions we obtain a set of candidate documents that contain the surface forms of the arguments of the knowledge graph fact. We restrict ourselves to those documents in which the positional spans of the surface forms and predicate paraphrases occur within a sentence of the document. We further distill this candidate set of sentences by considering only those sentences that contain a temporal expression that co-occurs with the other arguments of the fact. Thus, with this final candidate set of sentences $\mathcal{S}$ we have identified all sentences in which the fact is materialized in natural language and an evidence in the form of a temporal expression is present.

Time Model. As mentioned earlier, temporal expressions are challenging to analyze as they are present at different granularities and are inherently uncertain in nature. For instance, the temporal expression the 60s is highly uncertain as to which time interval it refers to: is it referring to $[1962, 1965]$ or $[1963, 1967]$? However, if we can model every possible time interval referred by such temporal expressions then we can rely on more precise evidence in form of fine granular temporal expressions (e.g., 1963 to 1965) that can help us ascertain the time interval the fact may refer to. In order to model this uncertainty we rely on the uncertainty-aware time model proposed by Berberich et al. [1]. Each temporal expression T is modeled by a four-tuple that contains bounds to: when the time interval could have begun $[b_\ell, b_u]$ and when the time interval could have ended $[e_\ell, e_u]$. Thus, the temporal expression T is represented by the following four-tuple:

$$T = \langle b_\ell, b_u, e_\ell, e_u \rangle. \qquad (3)$$

Temporal expressions found co-occurring with the natural language representation of knowledge graph facts are then compared in this time model to aggregate and identify the time interval of relevance for the fact.

Identifying Time Intervals. Our hypothesis is that the time scope of a knowledge graph fact F can be found by aggregating all the temporal expressions T, found to co-occur with its mention (Equation 2) in text, in the uncertainty-aware time model (Equation 3). Following our earlier work [3], we define the probability of a time interval being relevant to a given fact F as follows:

$$P([b, e] \mid F) = \frac{1}{|\mathcal{S}|} \cdot \sum_{s \in \mathcal{S}} P([b, e] \mid s_{time}), \qquad (4)$$

where, $s \in \mathcal{S}$ denotes a sentence containing the fact F and s_{time} denotes the temporal expressions associated with the sentence s. The likelihood of a time interval being relevant to the fact F given the sentence s is aggregated as (following [1]):

$$P([b, e] \mid s_{time}) = \frac{1}{\mid s_{time} \mid} \sum_{T \in s_{time}} \frac{\mathbb{1}([b, e] \in T)}{\mid T \mid}. \qquad (5)$$

3 EVALUATION

We next describe the evaluation setup for the experiments and a discussion of the obtained results.

Document Collections and Annotations. We utilized two news document collections that comprise of more than eleven million documents. Sentence splitting and temporal expressions for all the documents in these collections were obtained by using the SUTime component of the Core NLP toolkit [5]. Each time annotation is normalized to a time interval in the following format: $[(b)egin = yyyy\text{-}MM\text{-}dd, (e)nd = yyyy\text{-}MM\text{-}dd]$. Furthermore, we translate these annotations into the uncertainty-aware time model as follows: $[b, e] \equiv \langle b, e, b, e \rangle$. Annotation and collection statistics are given in Table 1.

Knowledge Graph Facts. We utilize the temporal fact benchmark by Gerber et al. [2] for experimental evaluation. Since our approach is completely unsupervised in nature we considered all the positive examples in train and test splits in the benchmark. Furthermore, for reasons of efficiency and recall we disregard the paraphrases of the predicates and only consider spotting the surface forms of the subject and object within a sentence of a document. For the baseline we considered the time interval correct when the begin and end year of the temporal expression matched to that of the ground truth begin and end years. For our proposed method, we considered the time interval determined correct if the overlap between generated time interval and the ground truth time interval, which is at the year granularity, was over 75% with respect to the generated time interval.

Systems and Baselines. As a baseline, we consider a method that simply ranks the time annotations, when coarsened to year granularity, by its frequency. For each temporal fact in the benchmark, if the number of sentences matched exceeds a threshold we sample a small subset of sentences to execute our method and the baseline. Note that each sentence can contain more than one temporal expression. We set the threshold of the maximum number of evidences considered as 25.

Table 1: Document collection statistics.

COLLECTION	$n_{documents}$	$n_{sentences}$	$n_{temporal\ expressions}$
NEW YORK TIMES	1,855,623	54,024,146	15,411,681
GIGAWORD	9,870,655	181,386,746	72,247,124

Table 2: Results obtained for different collections.

SYSTEM	NEW YORK TIMES			
	P@1	**P@5**	**P@10**	**MRR**
BASELINE	0.21	0.12	0.12	0.25
PROPOSED METHOD	**0.26**	0.12	0.12	**0.30**

SYSTEM	GIGAWORD			
	P@1	**P@5**	**P@10**	**MRR**
BASELINE	0.24	0.13	0.12	0.29
PROPOSED METHOD	**0.34**	**0.16**	**0.15**	**0.39**

Metrics. We evaluate to observe at what rank we can identify the correct time interval for a given fact. This is measured by the mean reciprocal rank (MRR). Furthermore, we evaluate if we can identify the correct ground truth time interval at rank 1, 5, and 10. That is, we measure precision at rank 1 (P@1), rank 5 (P@5), and rank 10 (P@10). We report averages over these metrics for the temporal facts in the benchmark.

Results. The results for the two document collections are given in Table 2. For the New York Times document collection we were able to evaluate 526 temporal facts in the benchmark, while for the Gigaword collection we were able to evaluate 740 temporal facts out of 1500 facts in the benchmark. The remaining facts could not be evaluated by our method since we could not spot sentences containing the entity surface forms for them in the document collections. From Table 2 we see that our method identifies the correct time interval for the facts at around rank 3. Considering precision at one, our method outperforms the baseline for both collections. The results thus show that our method identifies the correct time intervals for facts by leveraging uncertainty and modeling complex temporal expressions at different granularity.

4 RELATED WORK

Temporal information associated with knowledge graph facts is of high importance. A notable attempt in this direction was by Talukdar et al. [6]. Their method to identify the temporal scope of a knowledge graph fact counted frequency of publication dates associated with the documents containing its mention. For validating the temporal scopes of facts Gerber et al. [2] rely only on year granularity temporal expression tokens. Thus, they disregard the rich temporal information that is conveyed by implicit and relative temporal expressions. More recently, Kuzey et al. [4] aimed at leveraging the temporal information associated with facts in the YAGO knowledge graph and temporal expressions associated with the mention of the fact in document collection for tagging "temponyms". However, all these approaches utilize a rather naïve notion of time i.e., simple frequency of timestamps. Without modeling the inherent uncertainty and dynamics associated with the temporal expressions, these prior approaches can not identify concrete time intervals associated with knowledge graph facts.

REFERENCES

[1] K. Berberich et al.. A Language Modeling Approach for Temporal Information Needs. In *ECIR 2010*.
[2] D. Gerber et al.. DeFacto-Temporal and Multilingual Deep Fact Validation. *Web Semantics* 35, P2 (Dec. 2015).
[3] D. Gupta and K. Berberich.. Identifying Time Intervals of Interest to Queries. In *CIKM 2014*.
[4] E. Kuzey et al.. As Time Goes By: Comprehensive Tagging of Textual Phrases with Temporal Scopes. In *WWW 2016*.
[5] C. D. Manning et al.. The Stanford CoreNLP Natural Language Processing Toolkit. In *ACL 2014*.
[6] P. P. Talukdar et al.. Coupled temporal scoping of relational facts. In *WSDM 2012*.