Time Expression Resolution for Social Media Data

Jeniya Tabassum, Alan Ritter and Wei Xu
Computer Science and Engineering
Ohio State University
{bintejafar.1,ritter.1492,xu.1265}@osu.edu

Abstract

We describe TweeTIME, a minimally supervised time resolver for social the media domain that learns from large quantities of unlabeled data and requires no hand-engineered rules or hand-annotated training corpora. Our system achieves 0.68 F1 score on the end-to-end task of resolving date expressions, outperforming a broad range of state-of-the-art systems.

1 Introduction

Temporal expressions are words or phrases that refer to dates, times or durations. Resolving time expressions is an important task in information extraction (IE) that enables downstream applications such as calendars or timelines of events (Derczynski and Gaizauskas, 2013; Do et al., 2012; Ritter et al., 2012; Ling and Weld, 2010), knowledge base population (Ji et al., 2011), information retrieval (Alonso et al., 2007), automatically scheduling meetings from email and more. Previous work in this area has applied rule-based systems (Mani and Wilson, 2000; Bethard, 2013; Chambers, 2013) or supervised machine learning on small collections of hand-annotated news documents (Angeli et al., 2012; Lee et al., 2014).

Social media especially contains time-sensitive information and requires accurate temporal analysis, for example, for detecting real-time cybersecurity events (Ritter et al., 2015; Chang et al., 2016), disease outbreaks (Kanhabua et al., 2012) and extracting personal information (Schwartz et al., 2015). However, most work on social media simply uses generic temporal resolvers and therefore suffers from suboptimal performance. Recent work on temporal resolution focuses primarily on news articles and clinical texts (UzZaman et al., 2013; Bethard and Savova, 2016).

In this paper, we present a new minimally supervised approach to temporal resolution that requires no in-domain annotation or hand-crafted rules, instead learning from large quantities of unlabeled text in conjunction with a database of known events. Our approach is capable of learning robust time expression models adapted to the informal style of text found on social media. Our best model achieves a 0.68 F1 score when resolving date mentions in Twitter. This is a 17% increase over SUTime (Chang and Manning, 2012), outperforming other state-of-the-art time expression resolvers HeidelTime (Strötgen and Gertz, 2013), TempEX (Mani and Wilson, 2000) and UWTime (Lee et al., 2014) as well. Our approach also produces a confidence score that allows us to trade recall for precision. To the best of our knowledge, TweeTIME is the first time resolver designed specifically for social media data. This is also the first time that distant supervision is successfully applied for end-to-end temporal recognition and normalization.

2 System Overview

Our system consists of two major components – a **Temporal Recognizer** and a **Temporal Normalizer**. The subsystems are shown in Figure 1:

2.1 Temporal Recognizer

The goal of the recognizer is to predict the temporal tag of each word, given a sentence (or a tweet) $\mathbf{w} = w_1, \ldots, w_n$. We propose a multiple-instance learning model and a missing data model that are capable of learning word-level taggers given only sentence-level labels.

Our recognizer module in is built using a database of known events as *distant supervision*. We assume tweets published around the time of a known event that mention a central entity are also likely to contain time expressions referring

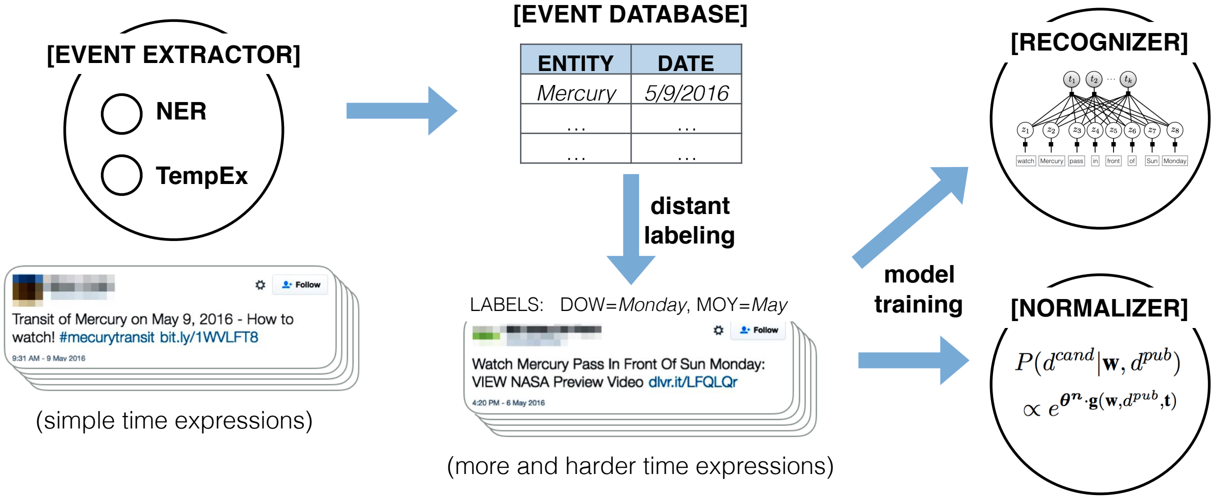

Figure 1: TweeTIME system diagram of model training.

to the event's date. For each event, we gather all tweets that contain the central entity and are posted within 7 days of the date when the event happend. We then label each tweet based on the event date in addition to the tweet's creation date.

Unlike supervised learning, where labeled instances are provided to the learner, in multiple instance learning scenarios (Dietterich et al., 1997), the learner is only provided with bags of instances labeled as either positive (where at least one instance is positive) or all negative. This is a close match to our problem setting, in which sentences are labeled with tags that should be assigned to one or more words. While learning, we never directly observe the words' tags ($\mathbf{z} = z_1, \ldots, z_n$). Instead, they are latent and we only observe the date of an event mentioned in the text, from which we derive sentence-level binary variables $\mathbf{t} = t_1, \ldots, t_k$ corresponding to temporal tags for the sentence. Following previous work on multiple-instance learning (Hoffmann et al., 2011; Xu et al., 2014), we model the connection between sentence-level labels and word-level tags using a set of deterministic-OR factors ϕ^{sent}.

The overall conditional probability of our model is defined as:

$$\begin{aligned} &P(\mathbf{t}, \mathbf{z}|\mathbf{w}; \boldsymbol{\theta}^r) \\ &= \frac{1}{Z} \prod_{i=1}^{k} \phi^{sent}(t_i, \mathbf{z}) \times \prod_{j=1}^{n} \phi^{word}(z_j, w_j) \\ &= \frac{1}{Z} \prod_{i=1}^{k} \phi^{sent}(t_i, \mathbf{z}) \times \prod_{j=1}^{n} e^{\boldsymbol{\theta}^r \cdot \mathbf{f}(z_j, w_j)} \end{aligned} \tag{1}$$

where $\mathbf{f}(z_j, w_j)$ is a feature vector and

$$\phi^{sent}(t_i, \mathbf{z}) = \begin{cases} 1 & \text{if } t_i = true \wedge \exists j : z_j = i \\ 1 & \text{if } t_i = false \wedge \forall j : z_j \neq i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

While the multiple-instance learning assumption works well much of the time, it can easily be violated – there are many tweets that mention entities involved in an event but that never explicitly mention its date. To handle this we adopt the missing data model of Ritter et. al. (2013). This model splits the sentence-level variables, t into two parts : m which represents whether a temporal tag is mentioned by at least one word of the tweet, and t' which represents whether a temporal tag can be derived from the event date. A set of pairwise potentials $\psi(m_j, t'_j)$ are introduced that encourage (but don't strictly require) agreement between m_j and t'_j, that is:

$$\psi(m_j, t'_j) = \begin{cases} \alpha_p, & \text{if } t'_j \neq m_j \\ \alpha_r, & \text{if } t'_j = m_j \end{cases} \tag{3}$$

Here, α_p (Penalty), and α_r (Reward) are parameters for the MiDaT model. α_p is the penalty for extracting a temporal tag that is not related to the event-date and α_r is the reward for extracting a tag that matches the date.

2.2 Temporal Normalizer

The Temporal Normalizer is built using a log-linear model which takes the tags $\mathbf{t}$ produced by the Temporal Recognizer as input and outputs one or more dates mentioned in a tweet. We formulate date normalization as a binary classification problem: given a tweet $\mathbf{w}$ published on date d^{pub}, we consider 22 candidate target dates ($\mathbf{w}$, d_l^{cand}) such that $d_l^{cand} = d^{pub} + l$, where $l = -10, \ldots, -1, 0, +1, \ldots, +10$, limiting the possible date references that are considered within 10 days before or after the tweet creation date, in addition to $d_l^{cand} = null$ (the tweet does not mention a date). The normalizer is similarly trained using the event database as distant supervision. The probability that a tweet mentions a candidate date is estimated using a log-linear model:

$$P(d^{cand}|\mathbf{w}, d^{pub}) \propto e^{\boldsymbol{\theta^n} \cdot \mathbf{g}(\mathbf{w}, d^{pub}, \mathbf{t})} \tag{4}$$

where $\boldsymbol{\theta^n}$ and $\mathbf{g}$ are the parameter and feature vector respectively in the Temporal Normalizer.

3 Results

To evaluate the final performance of our system and compare against existing state-of-the art time resolvers, we randomly sampled 250 tweets from 2014-2016 and manually annotated them with normalized dates; note that this is a separate date range from our weakly-labeled training data which is taken from 2011-2012. The final performance of our system, compared against a range of state-of-the-art time resolvers is presented in Table 1.

System	Prec.	Recall	F-value
TweeTIME	**0.58**	**0.70**	**0.63**
SUTime	0.54	0.64	0.58
TempEx	0.56	0.58	0.57
HeidelTime	0.43	0.52	0.47
UWTime	0.39	0.50	0.44

Table 1: Performance comparison of TweeTIME against state-of-the-art temporal taggers.

We see that TweeTIME significantly outperforms SUTime, TempEx, HeidelTime and UWTime on this challenging task of time expression resolution.

4 Conclusion

We presented a time resolver for social media domain that can learn from large amounts of unlabeled text using distant supervision. Our method extracts word-level temporal tags from tweets and combine them with a variety of other features in a novel date-resolver that predicts normalized dates referenced in a Tweet. Our proposed time resolver outperforms SUTime, TempEx, HeidelTime and UWTime on this challenging dataset for time normalization on the challenging social media domain.

References

Omar Alonso, Michael Gertz, and Ricardo Baeza-Yates. 2007. On the value of temporal information in information retrieval. In *ACM SIGIR Forum*. ACM, volume 41, pages 35–41.

Gabor Angeli, Christopher D Manning, and Daniel Jurafsky. 2012. Parsing time: Learning to interpret time expressions. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*.

Steven Bethard. 2013. A synchronous context free grammar for time normalization. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Steven Bethard and Guergana Savova. 2016. SemEval-2016 Task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*.

Nathanael Chambers. 2013. NavyTime: Event and time ordering from raw text. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)*.

Angel X Chang and Christopher D Manning. 2012. SUTime: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*.

Ching-Yun Chang, Zhiyang Teng, and Yue Zhang. 2016. Expectation-regulated neural model for event mention extraction. *Proccedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Technologies (NAACL)* .

Leon Derczynski and Robert J Gaizauskas. 2013. Temporal signals help label temporal relations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence* 89(1).

Quang Xuan Do, Wei Lu, and Dan Roth. 2012. Joint inference for event timeline construction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP)*.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL)*.

Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2011. Overview of the tac 2011 knowledge base population track. In *Proceedings of the Fourth Text Analysis Conference (TAC)*.

Nattiya Kanhabua, Sara Romano, Avaré Stewart, and Wolfgang Nejdl. 2012. Supporting temporal analytics for health-related events in microblogs. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*.

Kenton Lee, Yoav Artzi, Jesse Dodge, and Luke Zettlemoyer. 2014. Context-dependent semantic parsing for time expressions. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*.

Xiao Ling and Daniel S Weld. 2010. Temporal information extraction. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*.

Inderjeet Mani and George Wilson. 2000. Robust temporal processing of news. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL)*.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*.

Alan Ritter, Evan Wright, William Casey, and Tom Mitchell. 2015. Weakly supervised extraction of computer security events from Twitter. In *Proceedings of the 24th International Conference on World Wide Web (WWW)*.

Alan Ritter, Luke Zettlemoyer, Mausam, and Oren Etzioni. 2013. Modeling missing data in distant supervision for information extraction. *Transactions of the Association for Computational Linguistics (TACL)* 1:367–378.

H Andrew Schwartz, Greg Park, Maarten Sap, Evan Weingarten, Johannes Eichstaedt, Margaret Kern, Jonah Berger, Martin Seligman, and Lyle Ungar. 2015. Extracting human temporal orientation in Facebook language. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*.

Jannik Strötgen and Michael Gertz. 2013. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation* 47(2):269–298.

Naushad UzZaman, Hector Llorens, James Allen, Leon Derczynski, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 Task 1: TEMPEVAL-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)*.

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, and Yangfeng Ji. 2014. Extracting lexically divergent paraphrases from Twitter. *Transactions of the Association for Computational Linguistics (TACL)* 2(1).