Learning Time to Event with Context-sensitive Neural Networks

1 Introduction

User-generated content in social media platforms provide important and timely indicators on the spontaneous and often genuine views of users about real world events. In addition to discussing past events, users often discuss major current or forth coming events. Such temporality of events gives rise to two major NLP tasks:

- Distinguishing whether a given event happened in the past, is currently ongoing, or will be happening in the future, and
- Identifying the amount of time to forthcoming events. The former task is crucial for many temporal tasks including timeline generation , temporal relation extraction, and news summarization.

The above problems have been recently introduced owing to the first attempt by [2] who have developed a corpus containing *past*, *on-going* and *future* labels for news events. Authors tackled the problem by developing SVMs and convolutional neural network approaches. Other researches have developed techniques to precisely predict the *time of occurrence* for events [3]. Their approaches heavily depend on annotated temporal links that capture the temporal relation between pairs of event mentions and time expressions in data. In addition, [5] have developed a new annotation scheme to anchor events to time; TimeBank articles [4] have been partially annotated according to this scheme leading to the TimeBank-EventTime corpus. Recently, [6] proposed an effective CNN model which is enhanced by position embeddings and achieved the state-of-the-art results on TimeBank-EventTime corpus.

In contrast to news, textual content in social media introduce new NLP challenges. Besides being short and context-less, there are large amounts of convoluted textual content with informal and abbreviated words which challenges NLP techniques for text understanding. More importantly, a large fraction of event-relevant content do not carry any explicit temporal information, making some previous approaches not suitable as they require explicit temporal information in data. In addition, the time distribution of event-relevant content are often heavily skewed because the majority of content are posted in hours before and after events. Such skewed distributions introduces new challenges to learning approaches especially for short-lived events. The above challenges inspire our work on time to event prediction using user-generated content in social media platforms.

1.1 Problem definition

In this project we want to solve **Time to Event Prediction** in Twitter: given a tweet about a forthcoming event, the task is to predict the amount of time to the day on which the event will occur. We model this problem as a regression task. we plan to design a neural network models with attention mechanism that attend to temporal expression in the context. Existing models assume each input text contains explicit temporal information about the target event. Although such temporal information are informative (as they provide important linguistic clues that can help the learner), they may not be available in most tweets content. Our goal is to address this gap in existing models and provide a better understanding about temporality and its linguistic aspects for time to event detection.

2 Dataset

We plan to collect tweets about football matches of the England Premier League (EPL) following the same approach as [3]. These types of matches have the advantage that users tweet about them with distinctive hashtags by convention[1] which leads to thousands of tweets posted per match and also we plan to use tweets about previous nlp venues including NAACL, ACL, EMNLP, EACL and etc.

3 Evaluation Method

We evaluate performance of the model in term of Mean Absolute Error (MAE) which measures the difference (in hours) between the actual time to event and the estimated one. We want to compare our finding with several relevant researches in this domain such as TACL2018 [5] and state of the art embedding representation like BERT [1] and also we plan to define a few statistic baselines

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Ruihong Huang, Ignacio Cases, Dan Jurafsky, Cleo Condoravdi, and Ellen Riloff. Distinguishing past, on-going, and future events: The eventstatus corpus. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 44–54, 2016.

[3] Ali Hürriyetoglu, Nelleke Oostdijk, and Antal van den Bosch. Estimating time to event from tweets using temporal expressions. *EACL 2014*, pages 8–16, 2014.

[4] James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. The timebank corpus. In *Corpus linguistics*, volume 2003, page 40. Lancaster, UK., 2003.

[5] Nils Reimers, Nazanin Dehghani, and Iryna Gurevych. Temporal anchoring of events for the timebank corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2195–2204, 2016.

[6] Nils Reimers, Nazanin Dehghani, and Iryna Gurevych. Event time extraction with a decision tree of neural classifiers. *Transactions of the Association of Computational Linguistics*, 6:77–89, 2018.

[1]For example, fans of ArsenalARSand ChelseaCHEteams uses #ARSCHEhashtag to tweet about the match between them when Arsenal is host.