Temporal Relation Extraction: The Event Ordering Task

Hugo O. Sousa[1]

[1]INESC TEC, Portugal, R. Dr. Roberto Frias, Porto

Abstract

Although most Natural Language Processing tasks, such as Text Classification and Natural Language Translation, have experienced a major performance improvement due to recent advances in neural network architectures, Temporal Relation Extraction remains an open challenge. This leaves the door open for new research questions. In this paper, we provide a brief summary of the task and some of the recent efforts that have been made to solve it. In addition, some research opportunities yet to be explored are also discussed.

Keywords

Temporal Relation Extraction, Information Retrieval, Natural Language Processing

1. Temporal Relation Extraction

Temporal Relation Extraction (TRE) is a Natural Language Processing (NLP) task focused on classifying the temporal relationship between entities, typically events or temporal expressions, found in a text. A model that can accurately classify such relationships would be able to place events in a timeline, making it temporal-aware. This temporal knowledge could then be used in any time-sensitive NLP task, such as text summarization, natural language translation, question answering, or used more widely in knowledge bases. Despite many efforts in recent years, neural network architectures fail to make the leap in effectiveness already seen in other NLP tasks, thus making TRE an open challenge.

The roots of this task can be traced back to 2002, the year the TERQAS workshop took place. This workshop produced two important results: the Time Markup Language (TimeML) [1], the first annotation scheme that annotates temporal relationships; and `TimeBank` [2], the first corpus annotated with temporal relationships. Since then, many annotation schemes and datasets have been proposed. Some with the aim of making the annotation more complete as in `TimeBank-Dense` [3] and `TDDiscourse` [4], others to cope with the specificities of other languages, such as the `French TimeBank` [5], the `Portuguese TimeBank` [6] and the `Hindi TimeBank` [7]. MATRES [8] is a more comprehensive effort, with the authors annotating multiple time axis of the text. In addition, domain-specific datasets were also annotated, such as THEE [9] for event-based surveillance systems in public health and THYME [10] for health records. Another effort that was of great importance for the TRE task were the `SemEval` competitions, most notably the TempEval shared tasks held in three different years 2007 [11], 2010 [12], and 2013 [13].

Due to the low annotator agreement, the tendency over the years has been to simplify and refine the annotation scheme. For example, `TimeBank` was annotated with all 13 Allen interval relations [14], whereas in `TimeBank-Dense` the relation set consisted of only 6 interval relations. Also in MATRES [8], the authors argue that the inter-annotator agreement was much lower for the relation between the end-points of events, so they decided to focus the annotation only on the start-points.

The considerable number of datasets and annotation schemes makes it difficult to determine which model is the state of the art in TRE. To this regard, we have been working to create a Python package to facilitate comparison between different models. This will provide a common ground between them that the research community can build upon.

But it seems that despite many efforts made in recent years to train deep neural networks [15, 16, 17, 18], the state of the art models often rely on hand-craft rules [19, 20, 21, 17, 22] that are domain-specific and laborious to develop. Another approach to TRE is to train the model to identify the absolute time at which each event occurred in the narrative [19, 23]. After identifying the absolute time of each event, they can be placed in a timeline, where inferring their relations is trivial.

2. Research Questions

There are many research questions for this task, that are worth to be discussed. For example, classifying all 13 Allen interval relationships is typically difficult due to the fact that most relationships are underrepresented, leading to an unbalanced dataset. This problem can be solved by transforming the interval relations into point relations between the start and endpoints of each interval

DESIRES 2021 – 2nd International Conference on Design of Experimental Search Information REtrieval Systems, September 15–18, 2021, Padua, Italy

hugo.o.sousa@inesctec.pt (H. O. Sousa)
0000-0003-3226-9189 (H. O. Sousa)

[24]. Doing this will result in only three relations: EQUAL, BEFORE or AFTER. Making it easier to train the model.

When computing temporal closure [25] in a dataset annotated with Allen relations, it is common to derive relations that may have more than one relation. For example, if A, B and C are events and the annotation says that A OVERLAPS B and B OVERLAPS C than, the relation between A and C can be MEETS, OVERLAPS or BEFORE. This opens the door for another possibility yet to be explored that is to stage the problem as a Reinforcement Learning task [26]. In this framework, we can take full advantage of temporal closure by rewarding the model for any of the three relationships, whereas this would not be possible in conventional deep neural networks.

Another interesting approach that could be promising are Graph Neural Networks (GNN) [27]. Temporal relations have the natural structure of a graph, where the nodes are events or temporal expressions and the edges are the relations between them. GNN have demonstrated the ability to take advantage of this rich structure, making it a promising avenue for future research.

Acknowledgement

This work has been carried out as part of the project Text2Story, financed by the ERDF European Regional Development Fund through the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185).

References

[1] J. Pustejovsky, J. M. Castano, R. Ingria, R. Sauri, R. J. Gaizauskas, A. Setzer, G. Katz, D. R. Radev, Timeml: Robust specification of event and temporal expressions in text., New directions in question answering 3 (2003) 28–34.

[2] J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, et al., The timebank corpus, in: Corpus linguistics, volume 2003, Lancaster, UK., 2003, p. 40.

[3] T. Cassidy, B. McDowell, N. Chambers, S. Bethard, An annotation framework for dense event ordering, Technical Report, CARNEGIE-MELLON UNIV PITTSBURGH PA, 2014.

[4] A. Naik, L. Breitfeller, C. Rose, Tddiscourse: A dataset for discourse-level temporal ordering of events, in: Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue, 2019, pp. 239–249.

[5] A. Bittar, P. Amsili, P. Denis, L. Danlos, French timebank: an iso-timeml annotated reference corpus, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 130–134.

[6] F. Costa, A. Branco, Temporal information processing of a new language: Fast porting with minimal resources, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 671–677.

[7] P. Goel, S. Prabhu, A. Debnath, P. Modi, M. Shrivastava, Hindi timebank: An iso-timeml annotated reference corpus, in: 16th Joint ACL-ISO Workshop on Interoperable Semantic Annotation PROCEEDINGS, 2020, pp. 13–21.

[8] Q. Ning, H. Wu, D. Roth, A multi-axis annotation scheme for event temporal relations, arXiv preprint arXiv:1804.07828 (2018).

[9] J. Niu, V. Ng, G. Penn, E. E. Rees, Temporal histories of epidemic events (thee): a case study in temporal annotation for public health, in: Proceedings of The 12th Language Resources and Evaluation Conference, 2020, pp. 2223–2230.

[10] W. F. Styler IV, S. Bethard, S. Finan, M. Palmer, S. Pradhan, P. C. De Groen, B. Erickson, T. Miller, C. Lin, G. Savova, et al., Temporal annotation in the clinical domain, Transactions of the association for computational linguistics 2 (2014) 143–154.

[11] M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, J. Pustejovsky, Semeval-2007 task 15: Tempeval temporal relation identification, in: Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), 2007, pp. 75–80.

[12] M. Verhagen, R. Sauri, T. Caselli, J. Pustejovsky, Semeval-2010 task 13: Tempeval-2, in: Proceedings of the 5th international workshop on semantic evaluation, 2010, pp. 57–62.

[13] N. UzZaman, H. Llorens, L. Derczynski, J. Allen, M. Verhagen, J. Pustejovsky, Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations, in: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 2013, pp. 1–9.

[14] J. F. Allen, Maintaining knowledge about temporal intervals, Communications of the ACM 26 (1983) 832–843.

[15] M. D. Ma, J. Sun, M. Yang, K.-H. Huang, N. Wen, S. Singh, R. Han, N. Peng, Eventplus: A temporal event understanding pipeline, arXiv preprint arXiv:2101.04922 (2021).

[16] Q. Ning, S. Subramanian, D. Roth, An improved neural baseline for temporal relation extraction,

arXiv preprint arXiv:1909.00429 (2019).

[17] R. Han, I. Hsu, M. Yang, A. Galstyan, R. Weischedel, N. Peng, et al., Deep structured neural network for event temporal relation extraction, arXiv preprint arXiv:1909.10094 (2019).

[18] H. Wang, M. Chen, H. Zhang, D. Roth, Joint constrained learning for event-event relation extraction, arXiv preprint arXiv:2010.06727 (2020).

[19] Q. Do, W. Lu, D. Roth, 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, 2012, pp. 677–687.

[20] Q. Ning, Z. Feng, D. Roth, A structured learning approach to temporal relation extraction, arXiv preprint arXiv:1906.04943 (2019).

[21] Q. Ning, Z. Feng, H. Wu, D. Roth, Joint reasoning for temporal and causal relations, arXiv preprint arXiv:1906.04941 (2019).

[22] R. Han, Q. Ning, N. Peng, Joint event and temporal relation extraction with shared representations and structured prediction, arXiv preprint arXiv:1909.05360 (2019).

[23] A. Leeuwenberg, M.-F. Moens, Towards extracting absolute event timelines from english clinical reports, IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020) 2710–2719.

[24] C. Freksa, Temporal reasoning based on semi-intervals, Artificial intelligence 54 (1992) 199–227.

[25] M. Verhagen, Times between the lines, Brandeis University, Massachusetts (2004).

[26] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[27] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, M. Sun, Graph neural networks: A review of methods and applications, CoRR abs/1812.08434 (2018). URL: http://arxiv.org/abs/1812.08434. arXiv:1812.08434.