How not to suck at TIMEX annotation

L. Derczynski

University of Sheffield, Regent Court, Sheffield S1 4DP, UK

Abstract. Automatic annotation of temporal expressions has been attempted independently many times in the past decade. Results are starting to reach good levels but there always remains a class of problematic temporal expressions. In this paper, we present an analysis of mistakes and suggest remedies in each case. We then attempt an overall characterisation of errors and suggest future directions for improving temporal annotation.

1 Introduction

This paper concerns the identification and interpretation of expressions in language that represent times. For example, ...

#TODO – Distinguish dates, times, durations, sets; outline timex3 standard

In the next section, we review automatic approaches in timex recognition, up to the state of the art. We then describe an array of recognition and normalisation errors in sections 3 and 4 respectively. We then analyse the errors at a high level in Section 5 and finally conclude in Section 6.

2 Previous work

TempEval is a communal evaluation exercise [10] that assesses automatic temporal annotation. It includes set tasks, a fixed data set, and many systems. One of the tasks is temporal expression annotation. We will use this exercise as a basis for describing the state of the art.

#TODO – Outline methods used by current systems (integrated rule based: HeidelTime [8], semantic roles: TipSEM [4], sophisticated parsing: TRIPS [9], separate recognition and normalisation rules: USFD2 [2], NER and gazetteer: ANNIE [1], GUTime-based: TERNIP [6])

For annotation corpora, we informally tested timex recognition and normalisation systems on TimeBank [7], a gold-standard corpus, and the TAC KBP 2011 Source collection [1], a collection of 1.8 million documents from diverse sources. All examples are drawn from these resources; none are synthetic. We have intentionally not identified which systems produced which problems.

[1] Available from the Linguistic Data Consortium, reference LDC2010E12.

3 Recognition errors

3.1 Fall (the season)

To me, they all <TIMEX3 tid="t11" type="DATE" value="2009-FA">***fall*** </TIMEX3> *into the "drinks I'm supposed to love but actually normal coffee is better" category.*

Description Here, the word *fall* has been marked as a deictic temporal expression, and interpreted as a season using DCT to resolve the year. This is incorrect; in fact it describes a verbal event *fall into*. This is a sense disambiguation problem. To resolve this correctly, we could use either the following word, which is syntactically dominated by *fall*, or part of speech tagging to note that it is not a noun. Named entity recognition may help too.

Remedy: PoS tagging is a lightweight process, compared to NER and syntactic parsing. Any modern tagger (e.g. NLTK's built-in) can disambiguate the sense of "fall":

```
('To', 'TO'), ('me', 'PRP'), (',', ','), ('they', 'PRP'), ('all',
'VBP'), ('fall', 'VBP'), ('into', 'IN'), ...
```

3.2 Ranges of times

pinball's intended to give a real quick bite of entertainment, somewhere ***around 3 to 5 minutes****.*

Description: A duration which has a range of lengths is described, but nothing is annotated. In this case, the presence of a plural time unit (e.g. *minutes*, *months* prompts us to interpret a preceding modifier (if there is one) as part of a duration temporal expression. Modifiers could be quantitative (numeric, a numeric range) or qualitative (e.g. *some*, *a few*) and perhaps used to express a bound rather than an interval (e.g. *less than four months*). Word sense is important here; compare *a ten-minute run* with *four minute chefs*. Plurality and hyphenation ought to provide disambiguation clues.

Remedy: If we find a plural temporal unit noun, annotate it as a temporal expression, including adjacent previous modifier.

3.3 Years in words

*It's "****the two thousands two thousand tens****" etc etc, 2010 sounds like what my team is down by at halftime.*

Description: A year might be represented in words; *nineteen sixty nine*, for example. There may be hyphens between each word, indicating the connection between them and the bounds of the expression. These will occur in the same contexts as numeric dates, and so a perfect NER system should pick them up.

Remedy: One can convert textual dates into numbers and use immediate context (and perhaps a sanity check) to see if it is a timex. For example, a preceding word of "number" – *number two thousand and twelve* – indicates a non-temporal expression, whereas a preceding temporal signal [3] is an indicator of temporality (e.g. *since twenty ten*, *in the sixties*).

3.4 Sports terms

It's "the two thousands two thousand tens" etc etc, 2010 sounds like what my team is down by at ***halftime****.*

Description: A special sporting term is used to describe a time. It is preceded by a preposition that reinforces its use as a temporal expression.

Remedy: Develop a gazetteer containing these terms, and annotate them as temporal expressions. Surrounding context (e.g. temporal signals) can be used to disambiguate terms with more than one sense. Consider also preceding words that may mark the expression as a set, like *every*.

3.5 Decade names

The Roaring Twenties*. ...* ***Seventies*** *music.* ***Eighties*** *dances. This is a problem.*

Description: Duration temporal expressions, such as "the eighties", are sometimes only detected if preceded by a determiner "the". This constraint needs to be relaxed, while still distinguishing temporal from non-temporal usages of the word (e.g. those related to temperature).

Remedy: In the context of sentences consisting only of an NP, annotate these phrases as timexes. Where there is greater context, attempt word sense disambiguation using contextual clues.

3.6 Missed anchored duration

And then during ***the next 30 seconds****,*

Description: Here we have another quantifier/time-unit construction, which is not annotated, though has a pattern that is quite direct and unambiguously temporal.

Remedy: When we find a temporal signal, followed by a determiner, modifier, and time unit, annotate everything from the determiner to the unit as a temporal expression.

3.7 Missed unanchored duration

Kirk is the only woman to swim the 100-meter breaststroke in ***under 58 seconds*** *(57.77).*

Description: In this instance a modifier, a quantity and a temporal unit form a duration.

Remedy: Given a set list of modifiers, numbers detectable via a strict set of patterns, and a list of temporal units, we can detect durations accurately.

3.8 Implicit units

*Kirk is the only woman to swim the 100-meter breaststroke in under 58 seconds (**57.77**).*

Description: Sometimes we will encounter a number that lends precision to a previously specified duration; the context is usually sports timings. A later sentence may include "Megan Jendrick who posted a time of 58.87". This number needs to be annotated as a duration, using the units given earlier.

Remedy: If we find a number represented in digits and: (1) A duration timex including units of seconds, minutes or hours is in the past sentence; (2) the immediate context doesn't indicate against a temporal expression, then annotate the digits using the units in the previous expression.

3.9 Ages

*with her husband and <TIMEX3 tid="t38" type="DURATION" value="P3Y">**3-year** </TIMEX3>-old daughter*

Description: Here, a quantifier and time unit that are part of an adjectival construction describing and age are instead marked as a temporal expression.

Remedy: If the word immediately after the timex is *old*, do not annotate it as a timex.

4 Normalisation errors

4.1 Misalignment

The counterculture of <TIMEX3 tid="t20" type="DATE" value="206X"> the sixties </TIMEX3>.

Description: DCT in this case was 2009-12-10; the sixties is referring to the 1960s, not the 2060s. We have no immediate past-tense verbal clues.

Remedy: Adopt a sliding window approach to timex normalisation where a context (such as decade or month) is not present. Set this so that the majority of the window is in the past by default, allowing for tense-based adjustment, much like that used for weekday anchoring in [5].

5 Analysis

5.1 Commonly ignored information

Use the part of speech, it is easy.

5.2 Annotation guideline interpretation

Discuss how/where automated systems deviate from TIMEX annotation guidelines

6 Conclusion

The author would like to thank Dan Thomas of Adestra Ltd. for his valuable comments, and to acknowledge the UK Engineering and Physical Science Research Council's support in the form of a doctoral studentship.

References

1. Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.
2. L. Derczynski and R. Gaizauskas. Usfd2: Annotating temporal expresions and tlinks for tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 337–340. Association for Computational Linguistics, 2010.
3. L. Derczynski and R. Gaizauskas. Annotation and uses of temporal signals. *in preparation*, 2011.
4. H. Llorens, E. Saquete, and B. Navarro. Tipsem (english and spanish): Evaluating crfs and semantic roles in tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 284–291. Association for Computational Linguistics, 2010.
5. P. Mazur and R. Dale. What's the date?: high accuracy interpretation of weekday names. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 553–560. Association for Computational Linguistics, 2008.
6. C. Northwood. Ternip: Temporal expression recognition and normalisation in python. 2010.
7. 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, page 40, 2003.
8. J. Str
"otgen and M. Gertz. Heideltime: High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 321–324. Association for Computational Linguistics, 2010.
9. N. UzZaman and J.F. Allen. Trips and trios system for tempeval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 276–283. Association for Computational Linguistics, 2010.
10. M. Verhagen, R. Sauri, T. Caselli, and J. Pustejovsky. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62. Association for Computational Linguistics, 2010.