Proceedings of HLT-NAACL 2003
Short Papers , pp. 55-57
Edmonton, May-June 2003

Inferring Temporal Ordering of Events in News

Inderjeet Mani
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
imani@mitre.org

Barry Schiffman
Columbia University
1214 Amsterdam Avenue
New York, NY 10027
bschiff@cs.columbia.edu

Jianping Zhang
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
jzhang@mitre.org

Abstract

This paper describes a domain-independent, machine-learning based approach to temporally anchoring and ordering events in news. The approach achieves 84.6% accuracy in temporally anchoring events and 75.4% accuracy in partially ordering them.

1 Introduction

Practical NLP applications such as text summarization and question-answering place increasing demands on the processing of temporal information. In multi-document summarization of news, it is important to know the relative order of events so as to correctly merge and present information. In question-answering, one would like to be able to ask when an event occurs, or what events occurred prior to a particular event. Such capabilities presuppose an ability to infer the temporal order of events in discourse.

A number of different knowledge sources appear to be involved in inferring event ordering (Lascarides and Asher 1993), including tense and aspect (1), temporal adverbials (2), and world knowledge (3).

(1) Max *entered* the room. He *had drunk/was drinking* the wine.

(2) A drunken man *died* in the central Phillipines *when* he *put* a firecracker under his armpit.

(3) U. N. Secretary- General Boutros Boutros-Ghali Sunday *opened* a meeting of ....Boutros-Ghali *arrived* in Nairobi from South Africa, …

As (Bell 1999) has pointed out, the temporal structure of news is dictated by perceived news value rather than chronology. Thus, the latest news is often presented first, instead of events being described in order of occurrence (the latter ordering is called the narrative convention).

This paper describes a domain-independent approach to temporally anchoring and ordering events in news. The approach is motivated by a pilot experiment with 8 subjects providing news event-ordering judgments which revealed that the narrative convention applied only 47% of the time in ordering the events in successive past-tense clauses. Our approach involves mixed-initiative corpus annotation, with automatic tagging to identify clause structure, tense, aspect, and temporal adverbials, as well as tagging of reference times and anchoring of events with respect to reference times. We report on machine learning results from event-time anchoring judgments.

2 Linguistic Processing

The time expression tagger TempEx (Mani and Wilson 2000) tags and assigns values to temporal expressions, both "absolute" expressions like "June 1, 2001" and relative expressions like "Monday". It was cited in (Mani and Wilson 2000) as achieving a .83 F-measure against hand-annotated data. Inter-annotator reliability across 5 annotators on 193 TDT2-documents was .79F for extent and .86F for time values, with TempEx scoring .76F (extent) and .82F (value) on these documents.

The clause tagger (CLAUSE-IT) identifies top-level clauses (C), top-level clauses with gapped subjects (GC), e.g., "<C>He returned the book</C> <GC>and went home</GC>", relative clauses (RC), and complement clauses (CO), which include all non-finite clauses.

Our pilot experiment also revealed that the proportion of clauses with explicit time expressions (TIMEX2) is approximately 25%, suggesting that anchoring the events to just the explicit times wouldn't be sufficient. The system accordingly computes a reference time (Reichenbach 1947) value (**tval**) for each clause, defined to be either the time value of an explicit temporal expression mentioned in the clause, or, when the explicit time expression is absent, an implicit time value inferred from context.

To generate this **tval** feature, the simple algorithm in Figure 1 was used. The system also anchors the event's time with respect to the tval (at, before, or after) when the tval is an explicit reference time. This feature is called **anchor-explicit.** All in all, the features shown in Table 1 were computed for each clause.

Set initial tval to document-creation-date.
For each clause:

1. If clause has explicit time, then set its tval to it.
2. If clause-type is relative clause, assume its tval is inaccessible to later discourse.
3. If clause verb is of type reporting verb, set tval to document-creation-date.
4. If clause is inside quotes inherit tval from embedding clause.
5. Otherwise, pick most recent tval.

Figure 1: Algorithm for Computing Reference Time (tval)

CTYPE: clause is a regular clause, complement clause, or relative clause
CINDEX: subclause index
PARA: paragraph number
SENT: sentence number
SCONJ: subordinating conjunction (e.g., while, since, before)
TPREP: preposition in a TIMEX2 PP
TIMEX2: string in the TIMEX2 tag
TMOD: temporal modifier not attached to a TIMEX2, (e.g., after [an altercation])
QUOTE: number of words in quotes
REPVERB-P: reporting verb in clause
STATIVE-P: stative verb in clause
ACCOMP-P: accomplishment verb
ASPECTSHIFT: shift in aspect from previous clause
G-ASPECT: grammatical aspect {progessive, perfect,nil}
TENSE: tense of clause {past, present, future, nil}
TENSESHIFT: shift in tense from previous clause
ANCHOR_EXPLICIT: {<, >, =, undef}
TVAL: reference time for clause, i.e., a time value

Table 1: Linguistic Features for each Clause[1,2]

[1] The statives and accomplishments were computed from UMaryland's LCS lexicon, based on (Dorr and Olsen 1997)

3 Learning Anchoring Rules

A human unconnected with our project corrected the tval, based on a set of annotation guidelines, on a sample of 2069 clauses extracted at random from the North American News Corpus. She also anchored the event's time with respect to the tval (AT, BEF, AFT, or undefined). This feature (not a machine feature) is called **anchors.**

The corrections showed that the algorithm in Figure 1 was right on tval for 1231 out of 2069, giving an accuracy of 59%. Tracking the sequence of corrected tvals revealed that the tval of the previous clause was kept 65.75% of the time, that it reverted to some other previous tval 22.99% of the time, and that it shifted to a new tval 11.26% of the times. Most of the errors in computed tvals had to do with the tval being assigned erroneously the document date rather than reverting to a non-immediately previous tval. Finally, the **anchor-explicit** relation is correct 83.8% of the time; however, just guessing "at" for the explicit anchor will get an accuracy of 90.2%.

	ANCHORS	**TVAL-MOVES**
MAJORITY	(AT) 76.9	(KEEP) 65.75
C5.0 Rules	80.2 (±1.8)	71.8 (±0.5)

Table 2: Accuracy of Anchoring Rules

We then used this training data to train a statistical classifier, C5.0 Rules (Quinlan 1997), to learn (1) **anchors** relation rules and (2) rules for tracking the **tval moves** (keep, revert, shift) across successive clauses. The accuracy of anchors rules as well as tval change rules are shown in Table 2. It can be seen that accuracy of machine learning here is significantly better than the majority class. The tval, tense, and tense shift play a useful role in anchoring, revealing that the tval is a useful abstraction. Here are some of the rules learnt (here t_e is the clause index, assumed to stand for the event time of the clause):

If no sconj and no tmod and no tprep and tval-class =day then anchors(AT, t_e, tval) 80.4% accurate (156 examples).

If tense is present and no sconj and tval-class=month then anchors(AT, t_e, tval) 77.8 (7).

If tense is present perfect and no sconj, then

See www.umiacs.umd.edu/ ~bonnie/ LCS_ Database_Documentation.html.

[2] Since the TIMEX2 and tval values form an open class, they were automatically grouped into classes based on the granularity of the time expression, namely, {time-of-day, day, week, month, year, or non-specific}.

anchors(BEF, t_e, tval) 83 (4).
If tense shift is present2past and no explicit time and no sconj, then anchors(AT, t_e, tval) 90 (30)

4 Partially Ordering Links

Based on the best machine-learned rules for the **anchors** relation, **anchors** tuples are generated for each document. The tvals in the document's anchor tuples are also partially ordered, yielding tuples consisting of ordered pairs of tvals. The two sets of tuples are then used to provide a partial ordering of events in the document, in the form of **links** tuples: *links*(R, e_i, e_j), where e_i and e_j are the events corresponding to clauses i and j, and R is in {at, bef, aft, or undefined}. One of the authors evaluated the partial ordering for accuracy, on seven documents[3]. The results of this evaluation are shown in Table 3. #Correct-anchor is the number of the anchors tuples correctly classified and #total is the total number of anchors tuples classified. Link Recall is the percentage of human generated links tuples (723 in all) that are correctly identified by machine learned rules. Link Precision is the percentage of the machine generated links tuples that are correct.

#Clauses	**#Words**	**#correct-anchor / #total-anchor**	**Link Recall**	**Link Precision**
40	525	15/18 (83.3%)	44/65 (67.7%)	53/63 (84.1%)
18	335	12/13 (92.3%)	59/59 (100%)	59/62 (95.2%)
27	509	17/22 (77.2%)	23/40 (57.5%)	23/58 (39.7%)
38	617	21/27 (77.8%)	94/172 (54.7%)	94/190 (49.5%)
22	296	11/12 (91.7%)	39/42 (92.9%)	39/49 (79.6%)
14	242	6/7 (85.7%)	6/6 (100%)	6/7 (85.7%)
35	447	28/31 (90.3%)	297/339 (87.6%)	289/335 (86.3%)
194	**2971**	**110/130 (84.6%)**	**562/723 (77.7%)**	**563/764 (73.7%)**

Table 3: Document-Level Accuracy of Learnt Rules

5 Conclusion

Overall, our approach achieves 84.6% accuracy in anchoring events and 75.4% F-measure in partially ordering them. These numbers compare favorably with the previous literature: (Filatova and Hovy 2001) obtained 82% accuracy on anchoring for a single type of event/topic on 172 clauses, while (Mani and Wilson 2000) obtained accuracy of 59.4% on anchoring over 663 verb contexts. Our approach is also distinct in its use of human experimentation, machine learning and the variety of linguistically motivated features (including temporal adverbials) that are brought to bear.

Future work will examine the role of aspectual features, learning from skewed distributions dominated by AT (an overwhelming majority of news events occur at the reference times), and the incorporation of unsupervised learning methods.

References

A. Bell. News Stories as Narratives. In A. Jaworski and N. Coupland, The Discourse Reader, Routledge, 1999, 236-251.

E. Filatova, and E. Hovy. Assigning Time-Stamps to Event-Clauses. Workshop on Temporal and Spatial Information Processing, ACL'2001, Toulouse, 88-95.

A. Lascarides and N. Asher. Temporal Relations, Discourse Structure, and Commonsense Entailment. 1993. Linguistics and Philosophy 16, 437-494.

I. Mani and G. Wilson. Robust Temporal Processing of News. ACL'2000, 69-76.

R. Quinlan. 1997. C5.0. www.rulequest.com.

H. Reichenbach. The tenses of verbs. In H. Reichenbach, Elements of Symbolic Logic. Macmillan, 1947, Section 51, 287-298.

[3] Note that the naïve algorithm for tval is only 59% correct. While improvements to the naïve algorithm are clearly possible based on the corrected tval, to adequately test the machine learnt rules we use the corrected tval.