Temporal Relation Identification and Classification in Clinical Notes

ACM-BCB 2013

Jennifer D'Souza and Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas

Task Definition

•Given two entities (i.e. events or time expressions) in a text document classify them into one of a set of predefined temporal relations.

•Example BEFORE_OVERLAP

She had a normal pancreas at that time , however , hyperdense kidneys .

OVERLAP

Goal

•Advance the state-of-the-art in temporal relation classification in clinical notes by attempting fine-grained 12-class classification as opposed to broader 3-class classification (2012 i2b2 Challenge)

Dataset

i2b2 Clinical Temporal Relations Challenge Corpus (i2b2 Corpus) [Sun et al., 2013]

•310 de-identified discharge summaries annotated with 12 temporal relations.

- 190 training documents
- 120 test documents

•12 types of event-event, event-time temporal relations

Relation Type (%)	Inverse Relation Type (%)
Simultaneous (32.5%)	Overlap (40.2%)
Before (11.1%)	After (4.1%)
Before_Overlap (3.6%)	Overlap_After (11.6%)
During (2.7%)	During_Inv (4.5%)
Begins (5.5%)	Begun_By (1.4%)
Ends (2.3%)	Ended_By (3.1%)

Learning-based Baseline System

- **67 features taken from state-of-the-art systems**
  Lexical (17) , Grammatical (33), Entity attributes (8), Semantic (4), Distance (2), and Section creation time related (3)
- ***SVM**^{multiclass}* (Tsochantaridis et al., 2004)
- **Specialized Classifiers:** Four classifiers rather than just one shown to be better [Tang et al., 2012].
  1.Same-sentence event-event classifier
  2.Same-sentence event-time classifier
  3.Inter-sentence event-event classifier
  4.Inter-sentence coreferent classifier

Our Approach

Knowledge-Rich: Large-scale expansion of linguistic features including features based on *predicate-argument* and *discourse* relations.

Hybrid: System architecture that *combines rules and machine-learning*.

Knowledge-Rich Aspect

Five types of *novel features*

Type1: Pairwise Lexical Features

- Some of our baseline lexical features are computed based on either entity1 or entity2 but not both. This group includes pairwise versions of those lexical features to better capture the relationship of the two entities under consideration.
- One pairwise feature involves pairing up the type and modality of entity1 with the type and modality of entity2.
  - Type attribute encodes the type of the medical event, and modality attribute encodes whether the event happened in reality or not.
  - Example: Patient was given supplemental oxygen for shortness of breath.

    Event of type TREATMENT and modality FACTUAL

    Event of type PROBLEM and modality FACTUAL
  - From this example we create the feature: TREATMENT-FACTUAL-PROBLEM-FACTUAL
  - This feature encodes how factual medical events relate in time; it provides evidence that the temporal relation is OVERLAP_AFTER.
- Other pairwise lexical features are:
  - Entity head word pairs; Prepositional lexeme pairs; Preposition trace feature; and Verb POS trace feature.

Type2: Dependency Relation Features

- Motivation
  - Example: It is aggravated by activity.

    agent

    Event of type PROBLEM

    Event of type OCCURRENCE
  - Question: If we know that an OCCURRENCE is an agent to a PROBLEM, how will the two events relate in time?
  - Answer: The OCCURRENCE event and the PROBLEM event are SIMULTANEOUS.
  - Dependency relations in general enable such inferences and so we form features from them.
- 50 binary features in this group
  - For each dependency relation type among 25 others produced by the Stanford parser,
    - is the relation from entity1 to entity2?
    - or, is the relation from entity2 to entity1?

Type3: Webster and WordNet Features

- Motivation
  - Example: Her amylase was mildly elevated but has been down since then.
    - Knowledge from Webster: mildly elevated and down are antonyms.
    - Grammatically, mildly elevated and down are in contrast because of the coordinating conjunction *but.*
  - Question: If we know *two events mean opposite things*, and also that *they contrast each other grammatically*, how confidently can we associate them as being at different times temporally?
  - Answer: Statistically speaking, it is very likely that event mildly elevated is BEFORE event down, or event down is AFTER event mildly elevated.
  - We extract the following linguistic relations from Webster and WordNet to enable temporal class inferences as above.
    - Webster linguistic relations: Synonym, Antonym, Related-Word and Near-Antonym
    - WordNet linguistic relations: Hypernym, Hyponym, Troponym, and Similar
- 8 total binary features
  - For each type of linguistic relation,
    - is (event1, event2) ∈ t?
    - is (event2, event1) ∈ t?

Type4: Predicate-Argument Features

- Motivation
  - Example: Discussion should occur with the family about weaning him from medications to make him more comfortable.

    Verb ← Purpose Argument
  - Question: To accomplish a purpose, when should the action be taken?
  - Answer: The action weaning him must be taken BEFORE the purpose more comfortable is accomplished.
  - We use the following types of predicate-argument relations extracted automatically using the tool SENNA:
    - directional, manner, temporal, and cause.
- 8 total binary features
  - For each type of predicate-argument relation,
    - does event1 appear in event2's argument?
    - does event2 appear in event1's argument?

Type5: Discourse Relation Features

- Motivation
  - Example: *{_Argument1* At operation, there was no gross adenopathy, and it was felt that the tumor was completely excised.*}* *{_Argument2* The patient thereafter had a benign convalescence.*}*

    Event of type TREATMENT

    Event of type OCCURRENCE
    - Explicit Discourse Relation: Asynchronous

Type5: Discourse Relation Features (contd.)

- Question: If we know that a TREATMENT event is in a text segment (argument1) that is logically connected by an asynchronous relation to another text segment (argument2) containing an OCCURRENCE event, what is the temporal relation between the events?
- Answer: TREATMENT operation is BEFORE OCCURRENCE benign convalescence
- Some of the other discourse relations automatically extracted by Lin et. al.'s PDTB-style parser are:
  - Cause, Conjunction, Synchrony, Contrast, …
- 48 total binary features from 12 explicit discourse relations
  - For each type of discourse relation,
    - is event1 argument1, and event2 argument2?
    - is event2 argument1, and event1 argument2?

Hybrid Aspect

Manual Rules Development

- Rules are manually developed based on development data not used for evaluation.
- Rules are ordered in decreasing order of accuracy measured on development data.
- A new instance is classified using the 1st applicable rule in the ruleset.

Combining Rules and Machine Learning

- **2 Methods**:
  - **Method 1:** We employ all of the rules as *additional* features for training the temporal relation classifier.
  - **Method 2:** Given a test instance, we first apply to it the ruleset composed only of rules that are at least 75% accurate. If none of the rules is applicable, we classify it using the classifier employed in method 1.

RESULTS

	Features	All Rules	All Rules with accuracy >= 0.75	Features + Rules as Features	Rules + Features + Rules as Features
Feature Type	micro F	micro F	micro F	micro F	micro F
Baseline	55.3	--	--	--	--
+ Pairwise	55.5	37.6	14.5	57.2	57.8
+ Dependencies	55.5	40.0	16.2	57.4	58.1
+ WordNet	55.6	40.0	16.2	57.2	57.9
+ Webster	55.8	40.0	16.2	57.3	58.0
+ PropBank	55.8	45.4	21.3	57.6	59.7
+ Discourse	56.2	47.3	24.0	57.9	**61.1**

- Using all knowledge sources, the hybrid "Rules + Features + Rules as Features" architecture provides a 15% improvement over the baseline.