Analysis and Refinement of Temporal Relation Aggregation

Taylor Cassidy\textsuperscript{1,2}, Heng Ji\textsuperscript{3}

\textsuperscript{1}IBM T. J. Watson Research Center \textsuperscript{2}U.S. Army Research Lab
\textsuperscript{3}Computer Science Department, Rensselaer Polytechnic Institute
Outline

- Background: Adding time information to relations
- Aggregation and Classification Challenges
- Solutions
- Future work
Temporal Slot-filling: Gather and aggregate temporal information about the same relation across multiple documents.

School Attended: University of Houston

Mentions

Individuals

... But when?
What is “Temporal Information”?  

- **Focus:** Information about a relation between two named entities expressible in terms of its start and end time
- **Simplifying Assumption:** relations are “interval shaped”
  - One occurrence per relation
  - No Gaps
- **Task:**
  - **Input:** mention of a relation between two named entities in a document.
  - **Given:** corpus of millions of newswire, weblog, and discussion forum documents
  - **Output:** aggregated temporal information about the relation that can be inferred (collectively) from the entire corpus
Jones married Norton Simon in 1971, in a ceremony on a yacht in the English Channel after a courtship of three weeks. The four-tuple temporal representation is:

\[
\text{spouse}(\text{Jennifer Jones, Norton Simon}) = \langle 1971-01-01, 1971-12-31, 1971-01-01, \infty \rangle
\]

*(Amigó et al., 2011)*
Scoring Metric*

- **Temporal Quality**
  - Let $S = <t^{(1)}, t^{(2)}, t^{(3)}, t^{(4)}> $ be system output, $G = <g^{(1)}, g^{(2)}, g^{(3)}, g^{(4)}> $ be gold standard
  
  \[
  Q(S) = \frac{1}{4} \sum_{i} \frac{c}{c+|t^{(i)} - g^{(i)}|}
  \]

  - An error of $c$ time units produces a 0.5 score; scores produced with $c = 1$ year
  - Each element in tuple is scored independently

- **Overall Metric**
  \[
  P = \frac{\sum_{S^i \in C(S)} Q(S^i)}{M}
  \]
  \[
  R = \frac{\sum_{S^i \in C(S)} Q(S^i)}{N}
  \]

  - $M$: the number of system output tuples
  - $N$: the number of gold standard tuples
  - $C(S)$: the number of instances that have correct slot fills

*(Ji, Grishman, & Dang, 2011)*
Evaluation Examples

Infinite = 10000  -Infinite = 0
C_{cons} = C_{vag} = 5

\[ \begin{align*}
&Q = 0.25 \\
&Q = 0.61 \\
&Q = 0.33 \\
&Q = 0.5 \\
&Q = 1
\end{align*} \]
Approach Overview

(Ji et al., 2013 KIS)
Approach Overview

(Ji et al., 2013 KIS)

Distant Supervision

Web Docs

Free Base

SVM Classifiers

Information Retrieval
(query expansion etc.)

Pre-processing
(Entity Resolution, Parsing, etc.)

Relevant Sentences
(contain temporal information)

Temporal Classification
(NONE, BEGINNING, ENDING, WITHIN, THROUGHOUT)

Temporal Aggregation

Final Four-Tuple
\[ <t^{(1)}, t^{(2)}, t^{(3)}, t^{(4)}> \]
## Intermediate Relations

| Int. Rel     | four-tuple                      |
|--------------|---------------------------------|
| BEGINNING    | $< s, e, s, \infty >$          |
| ENDING       | $< -\infty, e, s, e >$          |
| BEG_AND_END  | $< s, e, s, e >$                |
| WITHIN       | $< -\infty, e, s, \infty >$    |
| THROUGHOUT   | $< -\infty, s, e, \infty >$    |
| BEFORE_START | $< e, \infty, e, \infty >$     |
| AFTER_END    | $< -\infty, s, -\infty, s >$  |
| BEFORE_START*| $< s, \infty, s, \infty >$     |
| AFTER_END*   | $< -\infty, e, -\infty, e >$  |
| NONE         | $< -\infty, \infty, -\infty, \infty >$ |

### Aggregation Operator

A spokeswoman for the Norton Simon Museum says Jones died Thursday ...

\[
< -\infty, 2009-12-11, -\infty, 2009-12-11>
\]

\[<\text{spouse}(Jones, Simon), 2009-12-11> \text{ classified as AFTER_END} \]

### Final Four-tuple

\[<1971-01-01, 1971-12-31, 1971-01-01, 2009-12-11>\]
How to Aggregate?

1. Gather instances from the corpus in which a relation \( r \) co-occurs with a time expression \( \gamma \).

2. Label each classification instance \( <r, \gamma> \) with an intermediate relation (e.g. WITHIN).

3. Aggregate all \( <r, \gamma>_i \) to yield one final four-tuple for \( r \).

A four-tuple is “consistent” iff.

1. \( t^{(1)} \leq t^{(2)} \)
2. \( t^{(3)} \leq t^{(4)} \)
3. \( t^{(1)} \leq t^{(4)} \)
How to Aggregate?

- Simple Aggregation:

  \[ T_{\text{final}} = \langle \max t^{(1)}, \min t^{(2)}, \max t^{(3)}, \min t^{(4)} \rangle \]

  - Take the latest \( t^{(1)} \) and \( t^{(3)} \) from all four-tuples, and the earliest \( t^{(2)} \) and \( t^{(4)} \) from all four-tuples.

- Problem: systems yield many incorrect four-tuples, resulting in a poor final four-tuple under max-constrain

- Previous work’s solution: order four-tuples by classifier confidence**

* (Ji, Grishman, Dang, 2011) **(Ji et al., 2013 KIS)
How to Aggregate four-tuples?

- Order four-tuples by confidence.
- Apply max-constrain to the top two.
- Accept the resulting four-tuple as the new #1 only if it is consistent
- Repeat

**Iterative Pair-Wise Max-constrain**

\[ T_0 \text{ and } T_1 \text{ are inconsistent – don’t aggregate!} \]

\[ T_0 \quad < -\infty, 1999-12-31, 1999-01-01, 1999-12-31 > \quad 99\% \]

\[ T_1 \quad < -\infty, 2009-07-31, 2009-07-01, \infty > \quad 71\% \]

\[ T_2 \quad < -\infty, 1998-12-31, 1998-01-01, \infty > \quad 65\% \]

\[ \text{classifier confidence: } \]

\[ < -\infty, 1998-12-31, 1999-01-01, 1999-12-31 > \]

*(Ji et al., 2013 KIS)*
How to Aggregate four-tuples?

- **Problem**: systems yield many incorrect four-tuples
- **Previous work’s solution**: order four-tuples by classifier confidence
- **Remaining Problems**:
  - Many incorrect four-tuples have high classifier confidence.
  - In particular: many relations are classified as holding at the document creation time
  - It only takes one “bad apple” to derail the process and corrupt the final four-tuple.
Suppose $T_1$ is incorrect, but was assigned high confidence...

|     | < $\infty$, 1999-12-31, 1999-01-01, 1999-12-31 > | 71% |
|-----|---------------------------------|-----|
| $T_0$ | < $\infty$, 2009-07-31, 2009-07-01, $\infty$ > | 99% |
| $T_1$ | < $\infty$, 1998-12-31, 1998-01-01, $\infty$ > | 65% |
"Bad Apple" effect

- **T₁ is incorrect**, but has high confidence

|   |   |   |   |
|---|---|---|---|
| T₁ | < -∞, 2009-07-31, 2009-07-01, ∞ > | 99% |
| T₀ | < -∞, 1999-12-31, 1999-01-01, 1999-12-31 > | 71% |
| T₂ | < -∞, 1998-12-31, 1998-01-01, ∞ > | 65% |

- Only four-tuples consistent with the incorrect T₁ can contribute to the final four-tuple
How to Aggregate?

- **Problem:** systems yield many incorrect four-tuples
- **Previous work’s solution:** order four-tuples by classifier confidence
- **Remaining Problems:**
  - Many incorrect four-tuples have high classifier confidence
  - In particular: many relations are classified as holding at the document creation time
  - It only takes one “bad apple” to derail the process and corrupt the final four-tuple.

- **Solutions:**
  - (1) Fix some common errors using rules
  - Use background knowledge to (2) eliminate or fix bad four-tuples and (3) add additional relevant four-tuples
(1) Post Relational States

- Verb Phrases use tense and aspect to help locate their predication time
  - tense is anaphoric (Higginbotham, 2006; Partee, 1984) and “binds” to a time
- NP’s lack these clues (in English).
  - We rely on context, world knowledge, reasoning, etc., to locate the properties NP’s denote.
- **An Exception**: modifiers like *former, then-, and ex-* introduce a *post-state* (Tonhauser, 2002)
- **Solution**: apply simple rules to correctly relate NP’s modified by *former, then-, and ex-* to the Document Creation Time
(2) Entity Existence Constraints

- **Common cause:** deceased person’s relations are mentioned in a newspaper article. Systems treat these as true at the time of publication.

- **Solution:** Obtain entity lifespan information from Wikipedia and convert **WITHIN** labels to **AFTER_END** or **BEFORE_START** for out-of-lifespan time-expressions.
This process eliminates many four-tuples that are incorrect, yet have a high classifier confidence!
(3) Inter-Relation Dependencies

Input relation: title(Chudinov, Prime Minister)

Extract inter-relation dependencies from Wikipedia

Hand code interval based reasoning rules for four-tuples

Retrieve documents relevant to dependent relations

Apply inference rules

"Prime Minister Daniar Usenov said at a press conference Monday that the government will uphold its current economic measure"

title(Chudinov, Prime Minister)
BEFORE
title(Usenov, Prime Minister)

AFTER_END* @ Monday
Classification Results

| Settings                  | P   | R   | F   |
|---------------------------|-----|-----|-----|
| Baseline System           | .337| .294| .314|
| Baseline + (1)            | .341| .298| .318|
| Baseline + (1) + (2)      | .353| .309| .329|
| Baseline + (1) + (2) + (3)| .360| .315| .336|

- Rule based approach to handling temporal nominal modifiers outperforms statistical methods
- Background knowledge about entity existence helps eliminate inaccurate temporal information
- Background knowledge about relation ordering facilitates addition of mutually beneficial temporal information
Remaining Analysis & Challenges

- Title relation attributed via Noun Phrase
  - e.g., (former) President Smith
  - Under what circumstances does relation hold at DCT? At implicit and explicit predication times related to surrounding verbs?

- Relation ordering information
  - Under what circumstances does temporal information about one relation inform that of another?
  - How to extract relation ordering information automatically from text?
“O’Donnell ... suggested **Wednesday** that the Obama administration - particularly **Vice President Joe Biden**, who **represented** Delaware in the Senate for **decades** - was behind them.”

- Is $title(Joe Biden, Vice President)$ true at
  - the predication time of $represented$?
  - ... at Document Creation Time?
- How do we know?
  - “Represent Delaware” not consistent with “Vice President” (of the US)
“Hidden former”

“Isakov is the latest of several top officials from the country’s caretaker government to resign so they can prepare for elections. Others include First Deputy Prime Minister Almazbek Atambayev, …”

\textit{title(Atambayev, First Deputy Prime Minister) relation not true at DCT}\nIt is clear that Atambayev already has resigned\n
In November 2000, \textbf{Chinese President Jiang Zemin} paid a state visit to Laos, the first visit to Laos by a Chinese president\n
\textit{title(Jiang Zemin, President) not true at DCT}\nThis sentence is from a list of historical visits by Chinese Presidents to Laos
To what time does "former" apply?

Secretary of State Hillary Rodham Clinton says former Philippines President Corazon Aquino “helped bring democracy back” to her country after years of authoritarian rule.”

Is title(Corazon Aquino, President) true at the predication time of helped? … at Document Creation Time?

Former US President Bill Clinton and US journalists Euna Lee and Laura Ling returned Wednesday from North Korea, one day after North Korea’s leader Kim Jong-II pardoned the two women

Is title(Bill Clinton, President) true at the time of returned? … at Document Creation Time?

Future work: rigorously define elements of context that govern temporal interpretation of NP predicates
In this work we assumed precedence relations were given.

- title(Minow Dean) and title(Kagan, Dean) are temporally linked.
Extracting Relation Ordering From Text (Future Work)

“Martha Minow, who succeeded Kagan as Harvard Law School dean, …”

title(Minow, Dean) and title(Kagan, Dean) are temporally linked

Future work: Extract Precedence (and other temporal) relations from raw text
Thank You