Causal Relation Extraction

Eduardo Blanco, Nuria Castell, Dan Moldovan
HLT Research Institute, TALP Research Centre, Lymba Corporation

LREC 2008, Marrakech

Introduction

- The **automatic detection and extraction of Semantic Relations** is a crucial step to improve the performance of several NLP applications (QA, IE, …)

- Example:
  - *Why do babies cry?*
  - *Hunger is the most common cause of crying in a young baby.*

- This work is focused on **Causal Relations**

Causal Relations

- **Relation between two events**: cause and effect
  - **cause** is the **producer** of the effect
  - **effect** is the **result** of the cause

- CAUSATION and other Semantic Relations
  - INFLUENCE(e1, e2) if e1 **affects the manner or intensity** of e2, but **not the occurrence**.
    - *Targeting skin cancer relatives improves screening*
  - CAUSATION(e1, e2) => TMP_BEFORE(e1, e2)

Causal Relations

- **Three** subtypes:
  - CONDITION if the **cause** is **hypothetical**
    - *If he were handsome, he would be married*
  - CONSEQUENCE if the **effect** is **indirect** or unintended
    - *His resignation caused regret among all classes*
  - REASON if it is a causation of **decision, belief, feeling or acting**
    - *I went because I thought it would be interesting*

Causal Relations

- **Marked or unmarked**
  - [marked] *I bought it* **because** *I read a good review*
  - [unmarked] *Be careful. It's unstable*

- **Ambiguity**
  - *because* **always** signals a causation
  - *since* **sometimes** signals a causation

- **Explicit or implicit**
  - [explicit] *She was thrown out of the hotel after she had run naked through its halls*
  - [implicit] *John killed Bob*

The Method

- Based on the use of **syntactic patterns** that may encode causation. We redefine the problem as a **binary classification**: *causation* or *¬causation.*
- **Manual classification** of 1270 sentences from TREC5 corpus, 170 causations found
- **Manual clustering** of the causations into syntactic patterns:

no.	Pattern	Productivity	Example
1	[VP **rel** C], [**rel** C, VP]	63.75%	*We didn't go* ***because*** *it was raining*
2	[NP **VP** NP]	13.75%	*The speech* ***sparked*** *a controversy*
3	[VP **rel** NP], [**rel** NP, VP]	8.12%	*More than a million Americans die* ***of*** *heart attack every year*
4	other	14.38%	*The lighting caused the workers to fall*

The Method

Syntactic patterns

- Since pattern 1 comprises more than half of the causations found, we focused this pattern
- The four most common relators encoding causation are *after*, *as*, *because* and *since*
- Example:
  - *He, too, [was subjected]$_{VP}$ to anonymous calls [after]$_{rel}$ [he [scheduled]$_{VPc}$ the election]$_{C}$*
- An instance not always encodes a causation:
  - *The executions took place a few hours **after** they announced their conviction*
  - *It has a fixed time, **as** collectors well known*
  - *It was the first time any of us had laughed **since** the morning began*

The Method

- We found 1068 instances in the SemCor 2.1 copus, 517 of which encoded a causation (i.e. the baseline is 0.516)
- Statistics depending on the relator:

Relator	Occurences encoding causation	Causations signaled
after	15.35 %	6.85 %
as	11.21 %	7.34 %
because	98.43 %	73.39 %
since	49.61 %	12.52 %

The Method

- relator = {after, as, because, since}

- relatorLeftModification = {POS tag}
- relatorRightModification = {POS tag}

- semanticClassVCause = {WordNet 2.1 sense number}
- verbCauseIsPotentiallyCausal = {yes, no}
  - A verb is potentially causal if its gloss or any of its subsumers' glosses contains the words *change* or *cause to*

- semanticClassVEffect = {WordNet 2.1 sense number}
- verbEffectIsPotentiallyCausal = {yes, no}

The Method

- For both VP, verb tense = {present, past, modal, perfective, progressive, passive}

- lexicalClue = {yes, no}
  - yes if there is a ‘,’, ‘and’ or another relator between the relator and VP_C
    - *He went as a tourist **and** ended up living there*
    - *City planners do not always use this boundary as effectively **as** they might*

The Method

Feature Selection

- **relator** = {after, as, because, since}
- **relatorLeftModification** = {POS tag}
- **relatorRightModification** = {POS tag}

- **semanticClassVCause** = {WordNet 2.1 sense number}
- **verbCauseIsPotentiallyCausal** = {yes, no}

- semanticClassVEffect = {WordNet 2.1 sense number}
- verbEffectIsPotentiallyCausal = {yes, no}

- For both VP, **verb tense** = {present, **past**, modal, **perfective**, progressive, passive}

- **lexicalClue** = {yes, no}

The Method

- As a Machine Learning algorithm, we used Bagging with C4.5 decision trees

- Results:

Class	Precision	Recall	F-Measure
causation	0.955	0.842	**0.895**
¬causation	0.869	0.964	**0.914**

Error Analysis

- Most of the causation are signaled by *because* and *since* (85.91%)
- The model learned is only able to classify the instances encoded by *because* and *since*
  - The results are good even though we discard all the causations signaled by *after* and *as*

- We can find examples belonging to different classes and with exactly the same values except for the semantic ones:
  - [causation]: *They [arrested]$_{VP}$ him after [he [assaulted]$_{VP}$ them]$_{C}$*
  - [¬*causation*]: *He [left]$_{VP}$ after [she [had left]$_{VPC}$]$_{C}$*

Error Analysis

- Paraphrasing doesn't seem to be a solution:
  - He left *after* she had left
  - He left *because* she had left

- Results obtained with the examples signaled by since:

Class	**Precision**	**Recall**	**F-Measure**
causation	0.957	0.846	**0.898**
¬causation	0.878	0.966	**0.920**

Conclusions and Further Work

- System for the detection of marked and explicit causations between a VP and a subordinate clause
- Simple and high performance

- Combine CAUSATION and other semantic relations:
  - CAUSATION(e1,e2), SUBSUMED_BY(e3,e1)=>CAUSATION(e3,e2)
  - CAUSATION(e1,e2), ENTAIL(e2,e3)=>CAUSATION(e1,e3)
- Causal **chains** and **intricate** Causal Relations
  - It is lined primarily by industrial developments and concrete-block walls because the constant traffic and emissions do not make it an attractive neighborhood

Questions?