# Minimally Supervised Event Causality Identification

## EMNLP 2011

Presenter: Alessandra Cervone

February 24, 2015

# Outline

# Causality

# Causality

# Pointwise Mutual Information (PMI)

$$PMI(x, y) = log \frac{P(x,y)}{P(x)P(y)}$$

- A distributional similarity method that measures the degree of association between two elements $x$ and $y$ in context.

# Pointwise Mutual Information (PMI)

$$PMI(x, y) = log \frac{P(x,y)}{P(x)P(y)}$$

- A distributional similarity method that measures the degree of association between two elements $x$ and $y$ in context.
- If $PMI = 0$, $x$ and $y$ are not associated, if $PMI > 0$ $x$ and $y$ are associated.

# Integer Linear Programming (ILP)

- Linear programming is an optimization technique which allows to find the best value of a linear function (for ex. the minimum or the maximum value) subject to some linear *constraints*.

# Integer Linear Programming (ILP)

- Linear programming is an optimization technique which allows to find the best value of a linear function (for ex. the minimum or the maximum value) subject to some linear *constraints*.
- Example of linear programming problem: find the maximum value of $z = x + 2y$, under the constraints $x \geq 0$, $y \geq 0$ and $3x - y \geq 2$.

# Integer Linear Programming (ILP)

- Linear programming is an optimization technique which allows to find the best value of a linear function (for ex. the minimum or the maximum value) subject to some linear *constraints*.
- Example of linear programming problem: find the maximum value of $z = x + 2y$, under the constraints $x \geq 0$, $y \geq 0$ and $3x - y \geq 2$.
- Integer Linear programming is a type of Linear programming where at least some of the variables have to be integers.

# Automatic extraction of causality relation

## Example

The police arrested the man *because* he killed someone.

The police arrested the man. He killed someone.

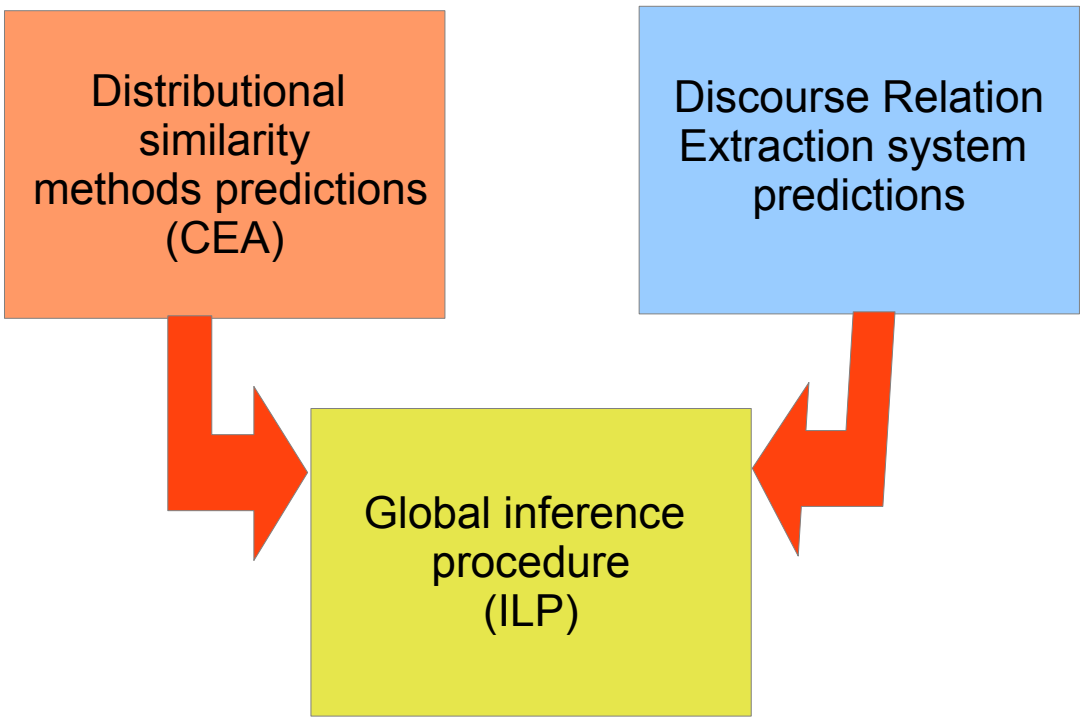
# Automatic extraction of causality relation

## Example

The police arrested the man *because* he killed someone.
The police arrested the man. He killed someone.

- Distributional similarity methods are generally used to extract causal relations.
- Claim:Discourse connectives (e.g. *because*) can be used to provide additional evidence to decide if two events (e.g. 'the police arrested the man' and 'he killed someone') are in a causal relation.

# System architecture



# Events definition

*Event*: an action or occurrence that happens with associated arguments.

$$e = p(a_1, a_2, ..., a_n)$$

- The predicate $p$ is defined as the word that triggers the presence of $e$ in the text, while $a_1, a_2, ..., a_n$ are its associated arguments.
- In the event 'The police arrested the man', $p$ = arrested, $a_1$ = the police, $a_1$ = the man.

# CEA

## Cause-Effect Association

$$CEA(e_i, e_j) = s_{pp}(e_i, e_j) + s_{pa}(e_i, e_j) + s_{aa}(e_i, e_j)$$

- The causality relation between two events $e_i$ and $e_j$ is the sum of:
- $s_{pp}$ = the association between event *predicates*.
- $s_{pa}$ = the association between the *predicate of an event* and the *arguments of the other event*.
- $s_{aa}$ = the association between event *arguments*

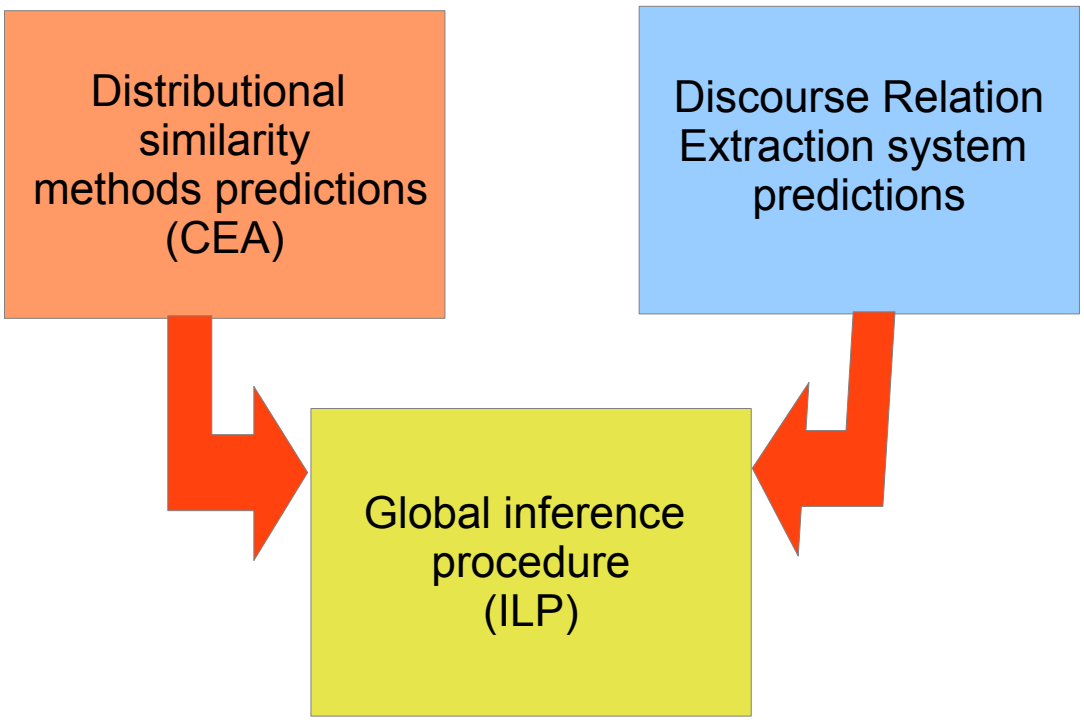
# CEA

## Cause-Effect Association

$$s_{pp}(e_i, e_j) = PMI(p^i, p^j) \times max(u^i, u^j) \times IDF(p^i, p^j) \times Dist(p^i, p^j)$$

- $PMI(p^i, p^j) \rightarrow$ assumption: event $e$ is a possible cause of event $e'$, if $e'$ happens more frequently with $e$, than by itself.
- $IDF(p^i, p^j) \rightarrow$ assumption: predicates appearing in many documents are probably not relevant.
- $max(u^i, u^j) \rightarrow$ assumption: predicates which appear most frequently with each other should be awarded.
- $Dist(p^i, p^j) \rightarrow$ assumption: event pairs that appear closer together have a higher weight.

# System architecture



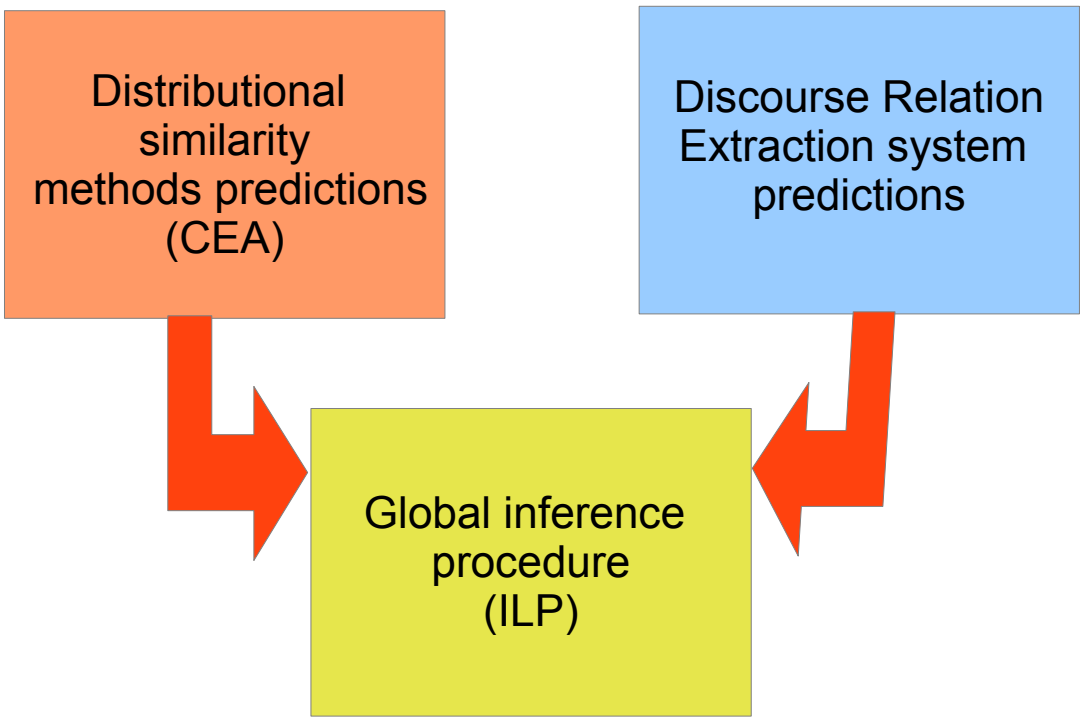
# Discourse Relations extraction

## Example

(The police arrested the man) *because* (he killed someone).

- Different discourse connectives evoke different discourse relations. In our example the relation is *Cause*, but there are many others (ex. *Concession*, *Contrast*, *Restatement*).
- The system used here considers only explicit connectives in text and automatically predicts their discourse relation with their relative text spans ('The police arrested the man' and 'he killed someone') using various features.

# System architecture



# Joint inference with ILP

- The ILP framework is used to define constraints between event pair causality extraction and discourse relation identification, that is *to force the predictions of the two systems to cohere with each other*.
- Prior to do so, however, the predictions of the CEA need to be binarized (in order to be included in the inferential process) as *causal* or $\neg$*causal*.

# Joint inference with ILP

Objective function: Find the max of

$$|L_{DR}| \sum_{c \epsilon C} \sum_{dr \epsilon L_{DR}} s_c(dr) \cdot x_{\langle c,dr \rangle}$$
$$+|L_{ER}| \sum_{ep \epsilon EP} \sum_{er \epsilon L_{ER}} s_{ep}(er) \cdot y_{\langle ep,er \rangle}$$

$L_{DR}$ = set of discourse relations,
$L_{ER}$ = set of event relation label ($causal$, $\neg causal$),
$s_c(dr)$= probability that connective $c$ is predicted to be of discourse rel $dr$,
$x_{\langle c,dr \rangle}$= binary indicator variable (=1 if $c$ labeled with $dr$),
$s_{ep}(er)$= CEA prediction score that event pair $ep$ takes on a given label er,
$y_{\langle ep,er \rangle}$= binary variable which equals 1 if $ep$ labeled as $er$.

# Joint inference with ILP

Example of constraint:

$$x_{\langle c,"\,Cause"\rangle} \leq \sum_{ep \epsilon L_{EP_c}} y_{\langle ep,"\,causal"\rangle}$$

$EP_c$ = set of event pairs that cross the two spans associated to $c$. This constraint means that if the discourse label *Cause* is assigned to $c$, then at least one of $ep_i, ..., ep_j$ must be labeled as *causal*.

# Evaluation

- According to the PARSEVAL scores on a human annotated test corpus, compared to other baseline systems (ex. simple PMI) the CEA approach alone obtained significantly better results.
- Moreover, the integration of the discourse relations information in CEA further increased the performance.

# Summary

- This article proposes a new approach to causal relation extraction.
- In particular, it shows how it is possible to integrate the predictions of two different systems using ILP.
- The integration of different sources of information for extracting causal relations seems to improve the performance of the system..
- Outlook
  - Implicit discourse connectives.
  - Apply ILP to other events relations.