

Employing Recent Advances in Machine Learning for Opinion Summarization

Claire Cardie

Department of Computer Science
Cornell University





**Homeland
Security**



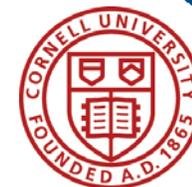
CERATOPS

Center for Extraction and Summarization
of Events and Opinions in Text

Janyce Wiebe, U. Pittsburgh

Claire Cardie, Cornell U.

Ellen Riloff, U. Utah



Cornell University

Where Our Work Fits In

- Consumer of advances in machine learning
 - Natural language learning
- Data = text from multiple genres and domains
- Transform documents and entire text collections into more useful (structured) representations
 - Databases
 - Graph-based summaries

Subjective Language

- **Subjective** sentences express *private states*, i.e. internal mental or emotional states
 - speculations, beliefs, emotions, evaluations, goals, opinions, judgments, ...

- (1) Jill said, "I *hate* Bill."
- (2) John *thought* he won the race.
- (3) Jane *hoped* for good weather.

Opinion Extraction and Summarization

- Extract non-factual information from text
 - Basic, low-level relations (database)
- Summarize in the form of graphs
- Hopefully provide insights that would not otherwise be easily accessible

WARNING:

NYTimes Oct06: "creepy and Orwellian"

Plan for the Talk

- Opinion summaries

 – Examples

- Constructing the summaries
- Open Problems

Fine-grained Opinions

Australian press has launched a bitter attack on Italy after seeing their beloved Socceroos eliminated on a controversial late penalty. Italian coach Lippi has also been blasted for his comments after the game.

In the opposite camp Lippi is preparing his side for the upcoming game with Ukraine. He hailed 10-man Italy's determination to beat Australia and said the penalty was rightly given.

Fine-grained Opinion Extraction

“The Australian Press launched a bitter attack on Italy”

- Five components
 - **Opinion trigger**
 - **Polarity**
 - positive
 - negative
 - neutral
 - **Strength/intensity**
 - low..extreme
 - **Source** (opinion holder)
 - **Target** (topic)

Opinion Frame

Source: “The Australian Press”

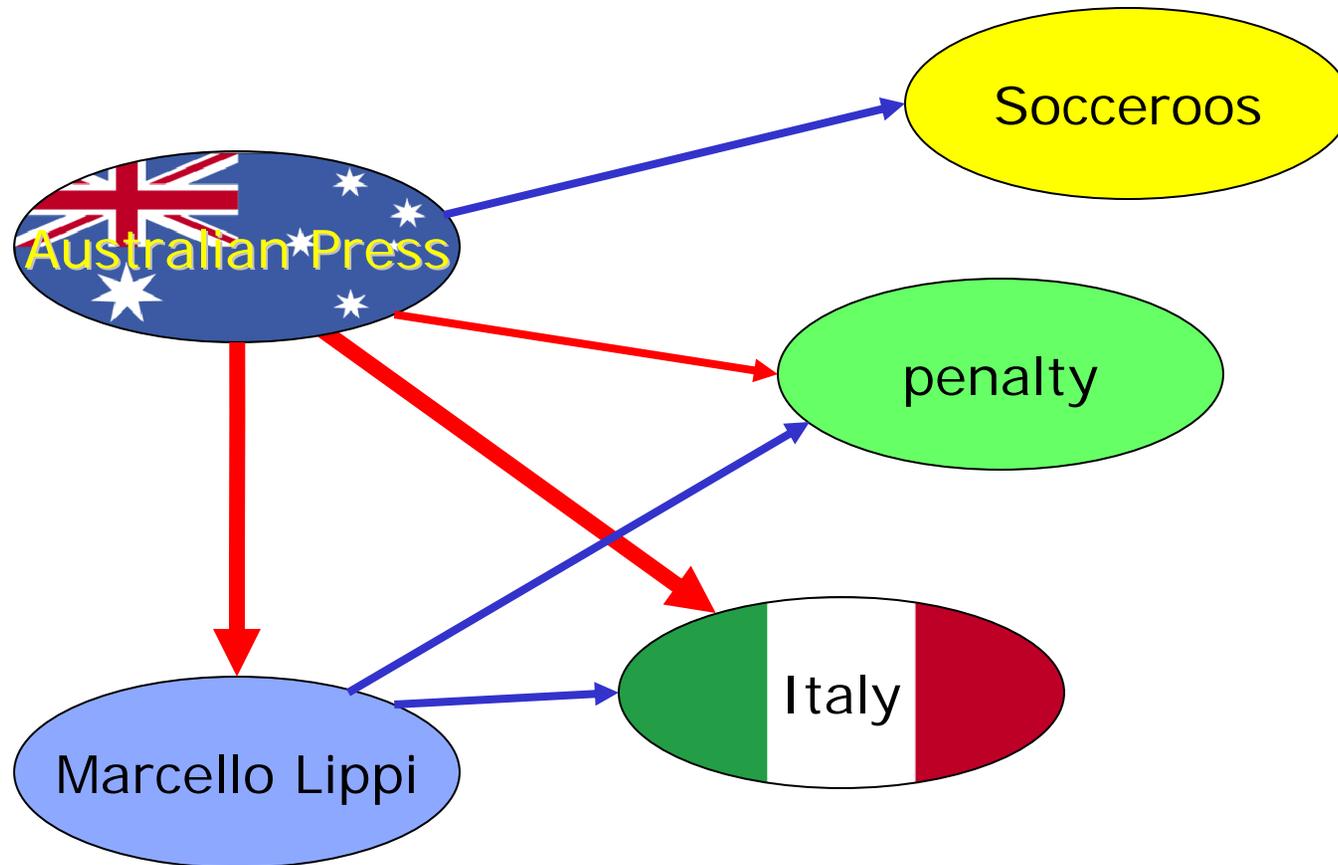
Polarity: negative sentiment

Intensity: high

Target: “Italy”

Trigger: “launched a bitter attack”

Opinion Summary



MPQA Summary Viewer

File Edit Mode Layout

The interface displays a graph with three red 'SOURCE' nodes at the top and three white 'TARGET' nodes at the bottom. A central yellow node is labeled 'the Zimbabwean head...'. Arrows point from the sources to the targets: a red arrow from 'Jack Straw' to the central yellow node, and blue arrows from 'the African bloc' and 'people of Zimbabwe' to the central yellow node. A black arrow points from the rightmost target to the central yellow node.

Node Type: Source
 Fraction: 2/2
 Percentage: 100.0%
 ID: straw
 Canonical Ref: Jack Straw

List of Changes:
 Layout reset to Source Target View.
 Layout reset to Nested Source Target View.
 Source Graph Cell ID : w - partners removed.

Source Node List

- mugabeetc (Mugabe, his government and his party, Zanu-PF)
- africaobservers (Observer missions from South Africa, Nigeria)
- mugabe (the Zimbabwean head of state)
- afobs (African observers)
- tsvang (Mugabe's main challenger, Morgan Tsvangirai of the opposition)
- namibians (The Namibians)
- zimesn (the Zimbabwe Election Support Network)
- zimpeople (people of Zimbabwe)
- africa (the African bloc)
- west (Western governments)
- sadc (Southern African Development Community (SADC) Partners)
- straw (Jack Straw)

government has rejected all criticism, describing the election as "exemplary, free, fair and peaceful."

Western governments disagree.

Western governments are lining up to condemn the Mugabe government, while the African bloc appears to have opted for solidarity in support of Mugabe.

Jack Straw, the foreign secretary of the former colonial power, **Britain**, which is locked in a verbal war with Mugabe, said the Zimbabwean head of state was bent on achieving "one goal, power at all costs. It is no surprise that this outcome has now been achieved."

start | Exceed | mpqa | ctc-u... | Inbox... | 12-m... | NetBe... | MPQA... | suzyq... | 1:01 PM

Example

The Annual Human Rights Report of the US State Department has been strongly criticized and condemned by many countries. Though the report has been made public for 10 days, its contents, which are inaccurate and lacking good will, continue to be commented on by the world media.

Many countries in Asia, Europe, Africa, and Latin America have rejected the content of the US Human Rights Report, calling it a brazen distortion of the situation, a wrongful and illegitimate move, and an interference in the internal affairs of other countries.

Recently, the Information Office of the Chinese People's Congress released a report on human rights in the United States in 2001, criticizing violations of human rights there. The report quoting data from the Christian Science Monitor, points out that the murder rate in the United States is 5.5 per 100,000 people. In the United States, torture and pressure to confess crime is common. Many people have been sentenced to death for crime they did not commit as a result of an unjust legal system. ...

Example

The Annual Human Rights Report of the US State Department has been **strongly criticized and condemned** by many countries. Though the report has been made public for 10 days, its contents, which are **inaccurate** and **lacking good will**, continue to be commented on by the world media.

Many countries in Asia, Europe, Africa, and Latin America have **rejected the content of the US Human Rights Report**, calling it **a brazen distortion** of the situation, **a wrongful and illegitimate move**, and an **interference** in the internal affairs of other countries.

Recently, the **Information Office of the Chinese People's Congress** released a **report on human rights in the United States** in 2001, **criticizing violations of human rights there**. The report quoting data from the Christian Science Monitor, points out that the murder rate in the United States is 5.5 per 100,000 people. In the **United States**, torture and pressure to confess crime is **common**. Many people have been sentenced to death for crime they did not commit as a result of an **unjust legal system**. ...

Too Many Opinion Frames

<writer>: onlyfactive <writer>: expr-subj (medium) <many-countries>:
neg-attitude (high) → <report>

<writer>: neg-attitude (medium) → <report> <writer>: neg-attitude (medium)
<writer>: neg-attitude (medium)

<writer>: onlyfactive <many-countries>: neg-attitude (medium)
→ <report> <many-countries>: extreme

<many-countries>: neg-attitude (high, high, medium)

<writer>: onlyfactive

<china-report>: neg-attitude (medium) → <US>

<writer>: onlyfactive <china-report>: onlyfactive

(medium) → <US> <writer>: expr-subj (low) <writer>: neg-attitude (low)

→ <US> <writer>: expr-subj (low) <writer>: neg-attitude (medium)

<writer>: onlyfactive

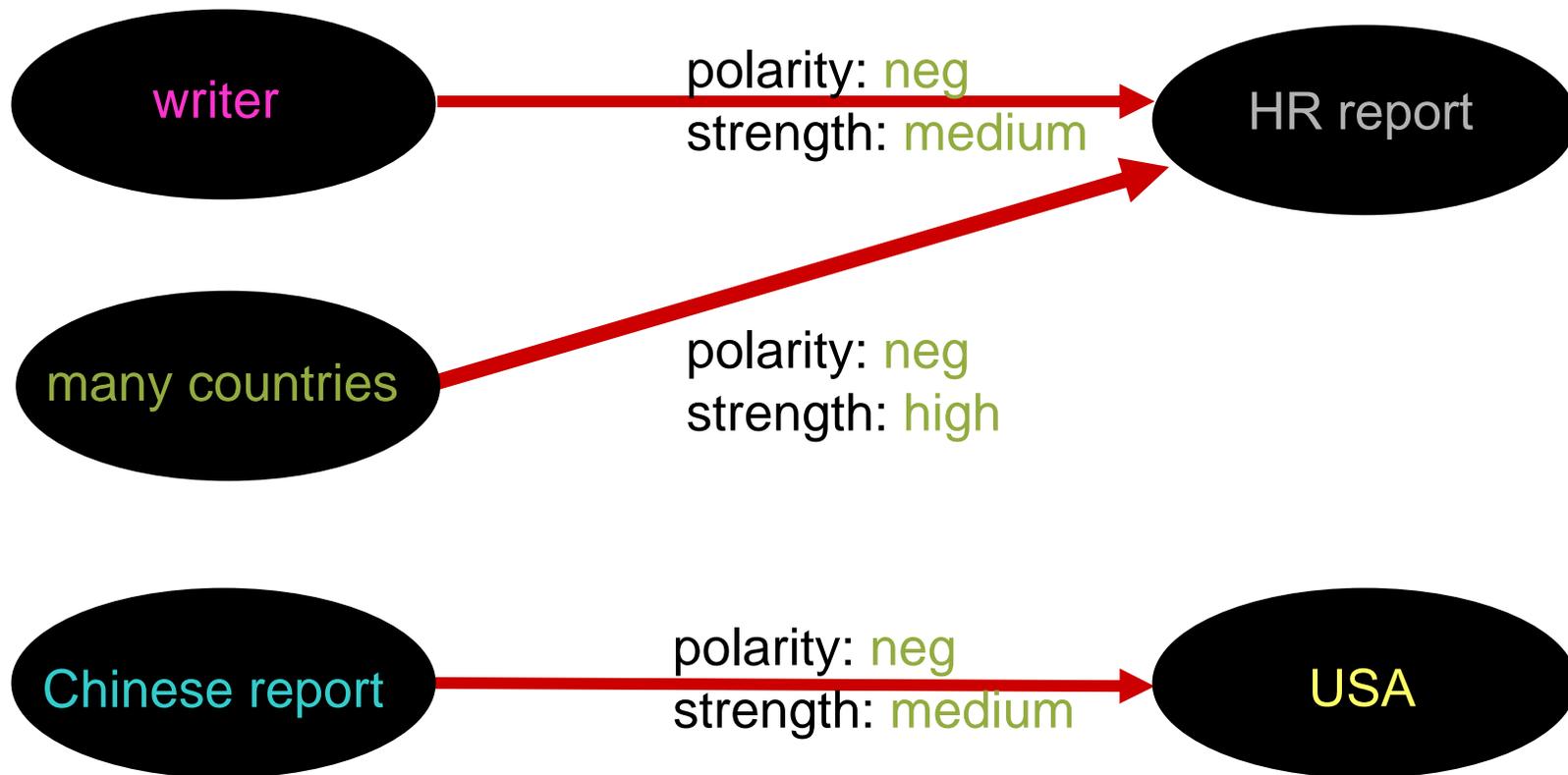
<writer>: onlyfactive

<writer>: neg-attitude (low) → <US>

<writer>: expr-subj (low)

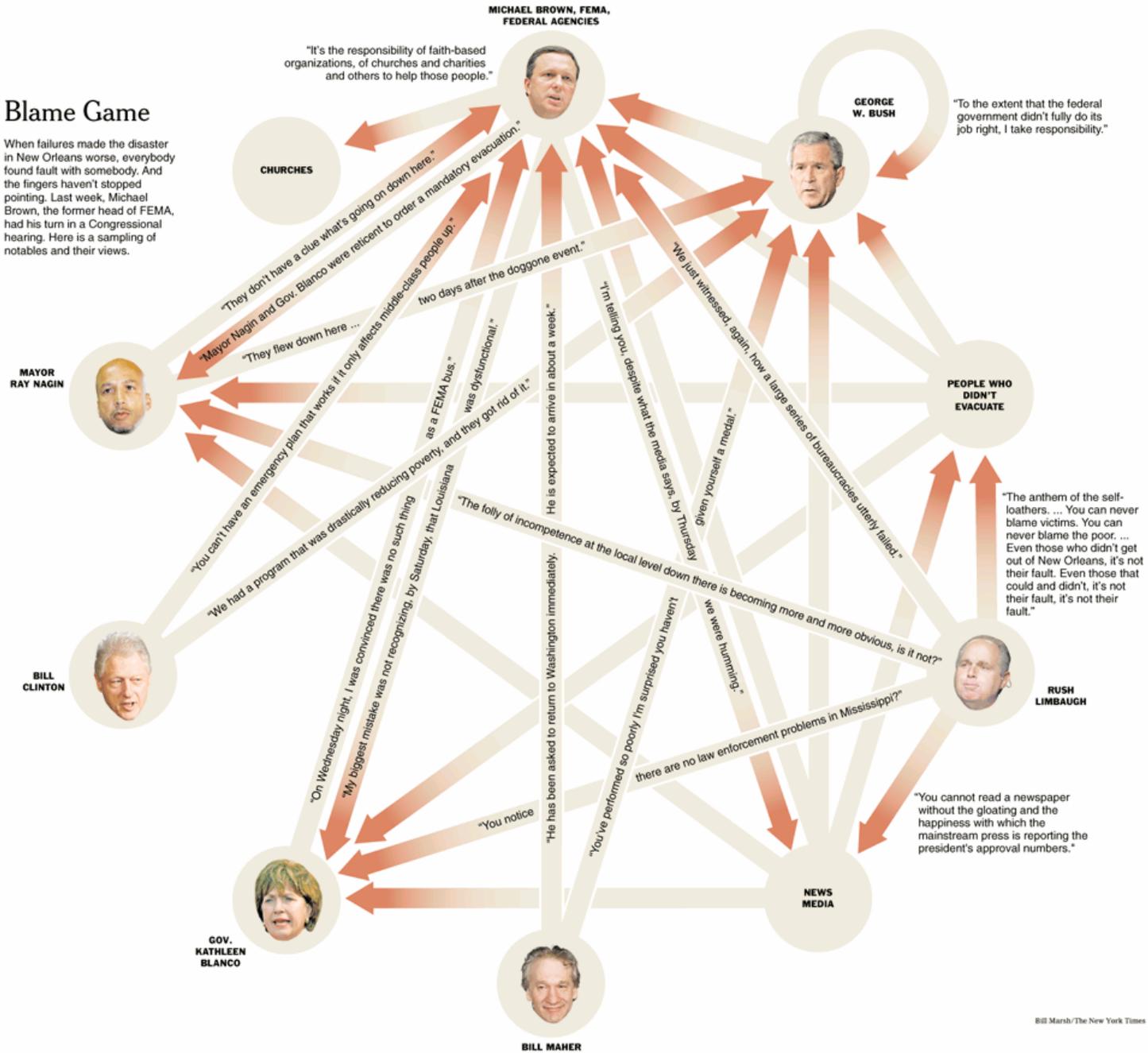


Opinion Summaries



Blame Game

When failures made the disaster in New Orleans worse, everybody found fault with somebody. And the fingers haven't stopped pointing. Last week, Michael Brown, the former head of FEMA, had his turn in a Congressional hearing. Here is a sampling of notables and their views.



Bill Marsh/The New York Times

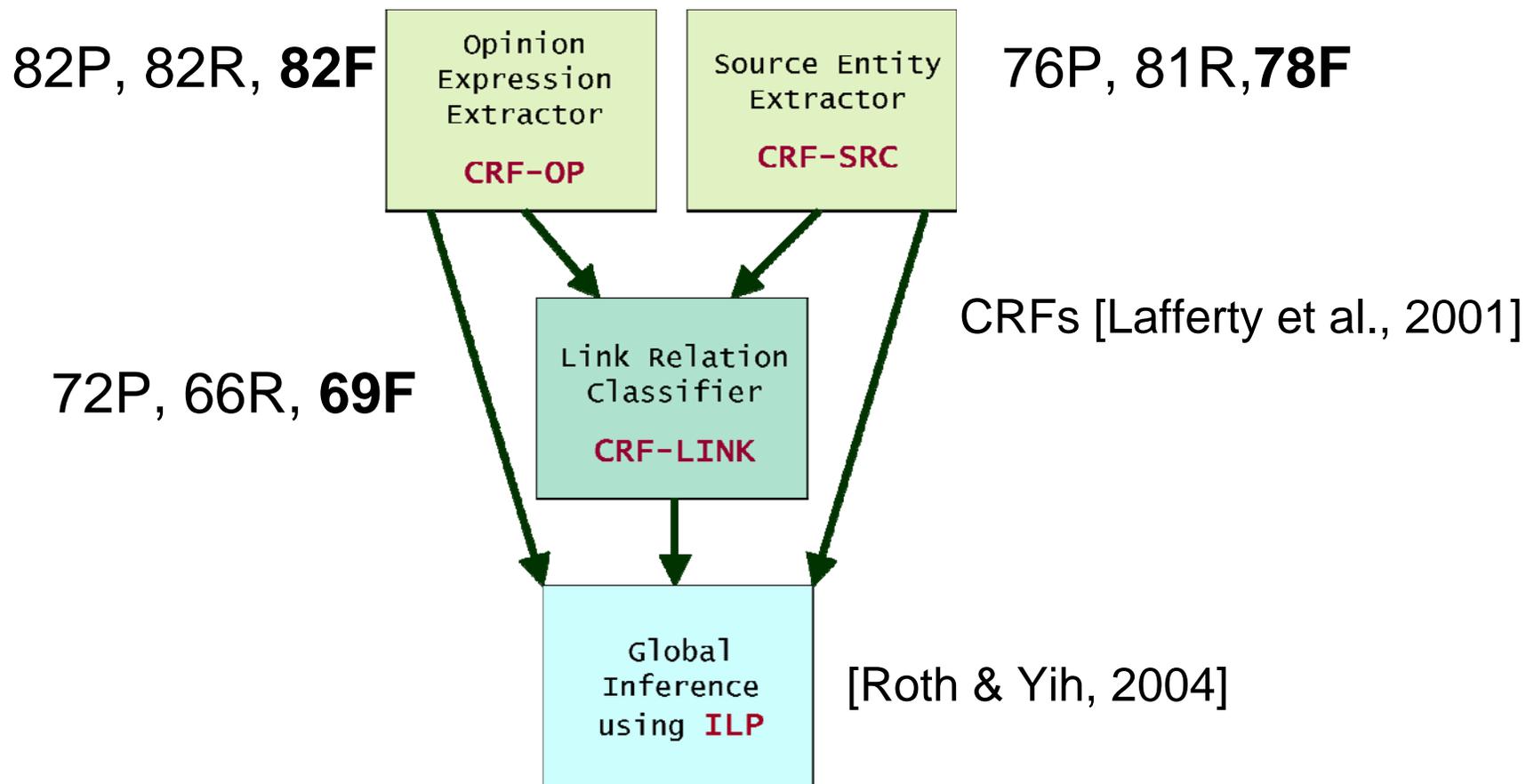
Bill Marsh/The New York Times

Constructing Summaries

- ➔ Generate opinion frames
- Source
 - Opinion trigger  expresses
 - Polarity
 - Strength
 - ~~Topic/target~~
- Group related opinions together
- By Source
 - ~~By Topic~~
- Aggregate multiple (conflicting) opinions from the same source on ~~the same topic~~
- User chooses strategy

Opinion Frame Extraction via CRFs and ILP

- Joint extraction of entities and relations



Constructing Summaries

■ Generate opinion frames

.78F
.82F

- Source
- Opinion trigger ← expresses .69F
 - Polarity
 - Strength
- ~~Topic/target~~



Group related opinions together

- By Source
 - ~~By Topic~~
- Aggregate multiple (conflicting) opinions from the same source on ~~the same topic~~
 - User chooses strategy

Partially Supervised Clustering for Source Coreference Resolution

- Labels for non-source NPs are unavailable

Australian press has launched a bitter attack on **Italy** after seeing **their** beloved Socceroos eliminated on a controversial late penalty.

Italian coach Lippi has also been **blasted** for his comments after the game.

In the opposite camp **Lippi** is preparing his side for the upcoming game with Ukraine. **He** hailed 10-man **Italy's** determination to beat Australia and said the **penalty** was rightly given.

[following Li & Roth, 2005; Finley & Joachims, 2005; McCallum & Wellner, 2003]

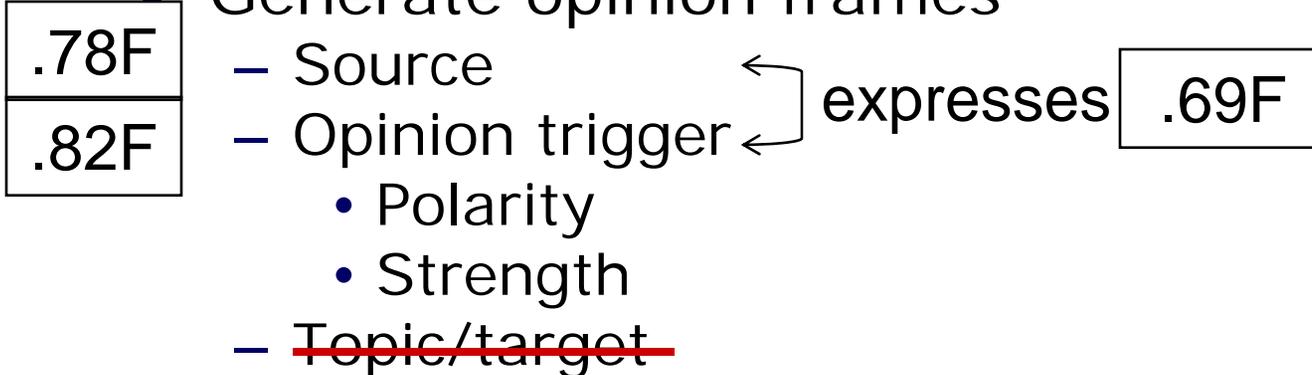


Partially Supervised Clustering

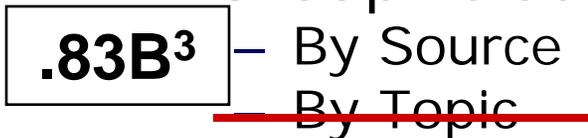
- Extend rule-learning algorithm to learn pairwise classification function in the context of single-link clustering.
 - Exploit complex structure of coreference resolution
- During rule construction, consider the effect of the rule on the overall clustering of items
 - Compute transitive closure **including** the unlabelled pairs
 - Calculate performance **ignoring** the unlabelled pairs

Constructing Summaries

■ Generate opinion frames



■ Group related opinions together



➔ Aggregate multiple (conflicting) opinions from the same source on ~~the same topic~~

– User chooses strategy

.40-.50F

Problems

- Combining dozens of linguistic classifiers/
sequence taggers
 - Focus on increasing recall levels
- Re-training required when domain or
genre changes
 - Semi-supervised learning? Active learning?
- How can we best incorporate user
feedback in the final system
 - During analysis/interpretation?
 - Fixing errors in final output?