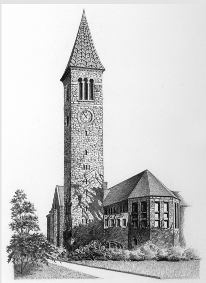
Finding and Extracting Opinions in On-line Text



Plan for the Talk

- Subjectivity and sentiment in language
- Opinion extraction
  - definition and examples
- Algorithms and evaluation
- Demo

Subjective Language

- **Subjective** sentences express *private states*, i.e. internal mental or emotional states
  - speculations, beliefs, emotions, evaluations, goals, opinions, judgments, …
    - Jill said, "I *hate* Bill."
    - John *thought* about whom to vote for.
    - Claire *hoped* her lecture would go well.

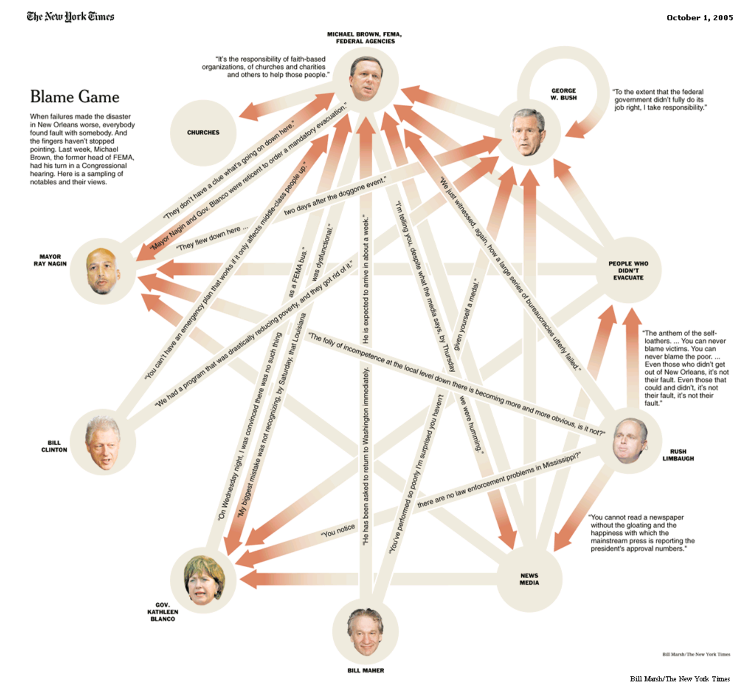
Subjectivity vs. Sentiment

- **Sentiment** expressions are a type of subjective expression
  - expressions of *positive* and *negative* emotions, judgments, evaluations, …
    - Jill said, "I *hate* Bill." -
    - John *thought* about whom to vote for.
    - Claire *hoped* her lecture would go well. +

In this talk, opinion = any subjective language

Why Study Opinions?

- Web Queries of a Subjective Nature
  - How have **business** views towards **global climate change** varied over the past decade?
  - What is the reaction in **Asia** to the to the **Bush policy towards the Kyoto Protocol**?
  - How have **consumers and businesses** responded to Gore's "**An Inconvenient Truth**"?
  - **Who** were the first people to propose **bailout options for banks** in the current economic crisis?
  - What does **Sarah Palin** think about **<X>**?



Research Trend

Factual and Event-based Text

Pang & Lee
[ACL 2002, ACL 2004, …]

Subjective Text

sentiment analysis tome
[Pang & Lee, 2008]

Plan for the Talk

- Subjectivity and sentiment in language
- Opinion extraction
  - definition and examples
- Algorithms and evaluation
- Demo

Fine-grained Opinion Extraction

…The Australian press has launched a bitter attack on Italy after seeing their beloved Socceroos eliminated on a controversial late penalty. Italian coach Lippi has 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 their winning penalty was rightly given.

…

Fine-grained Opinion Extraction

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 Opinions

"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)

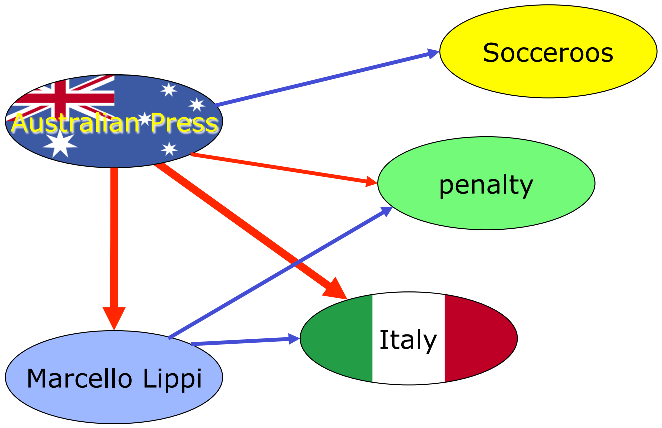
Opinion Frame
Polarity: negative
Intensity: high
Source: "The Australian Press"
Target: "Italy"

Example – 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.

Example – Opinion Summary



Cornell University

What makes this hard?

- MPQA corpus
  - 2812 opinion expressions (medium or higher intensity)
  - 4282 *content word* tokens
  - 49% are unique
- For words in these expressions that appear > 1 time
  - 38% appear in *both* subjective and objective contexts
    - *achieved* (2 subjective, 4 objective);
    - *against* (15 subjective, 40 objective);
    - *considering* (3 subjective, 7 objective);
    - *difficult* (7 subjective, 8 objective);
    - *fact* (14 subjective, 7 objective);

Cornell University [Wiebe, Wilson, Cardie 2004]

What makes this hard?

- MPQA corpus (dev set)
  - 41.5% of sentences are objective
  - 44.0% of sentences contain mixtures of opinions and objective speech events
    - Half of these contain 3 or more
  - Differing polarities for opinion expressions from the same sentence

	positive	negative	neutral	both
positive	36	32	44	2
negative	66	32	66	7
neutral	55	69	135	4
both	2	2	7	0

Thanks, Ainur!

Cornell University

Not so easy for people either…

Cornell University

Plan for the Talk

- Subjectivity and sentiment in language
- Opinion extraction
  - definition and examples
- Algorithms and evaluation
- Demo

Opinion Extraction and Aggregation

- Generate opinion frames
  - Source/opinion holder
  - Opinion expression
    - Polarity
  - ~~Topic/target~~
- Group related opinions together
  - By Source
  - ~~By Topic~~

Apply machine learning / statistical learning algorithms…

Cornell University

Supervised Inductive Learning

examples
[features + class]

↓

ML Algorithm

↓

(novel) examples [features] → statistical model (program) → class

Cornell University

Supervised Inductive Learning

examples **of (Source) NPs in context**
[features + class]

↓

ML Algorithm

↓

(novel) **NPs in context** [features] → statistical model (program) → **Source? Not Source?**

Cornell University

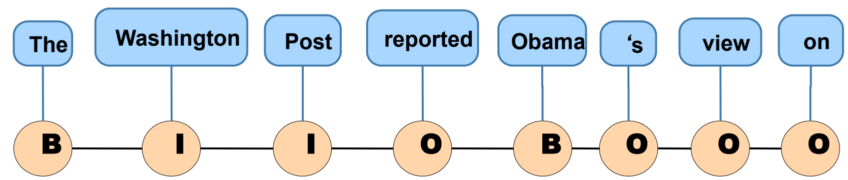
Source Annotations

<Australian press> has <u>launched a bitter attack</u> on **Italy** after seeing **<their>** <u>beloved</u> **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>** <u>hailed</u> 10-man **Italy**'s determination to beat Australia and said the **penalty** was <u>rightly given</u>.

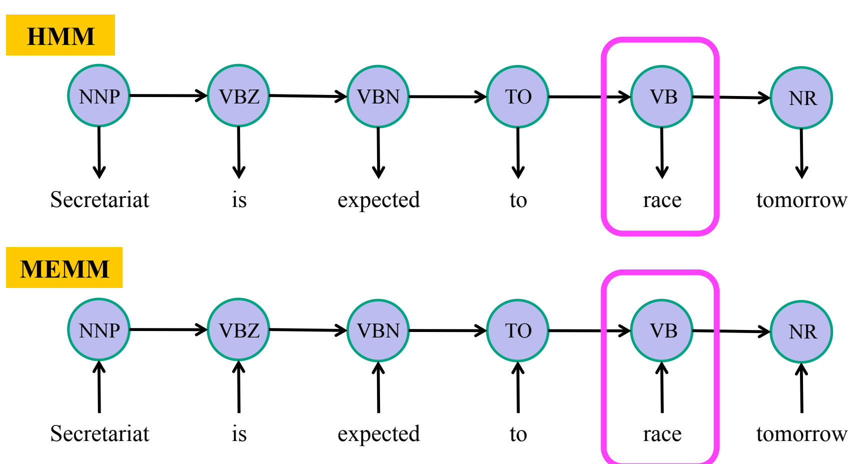
Identifying Sources of Opinions
…as an *Information Extraction* task

- Sequence tagging

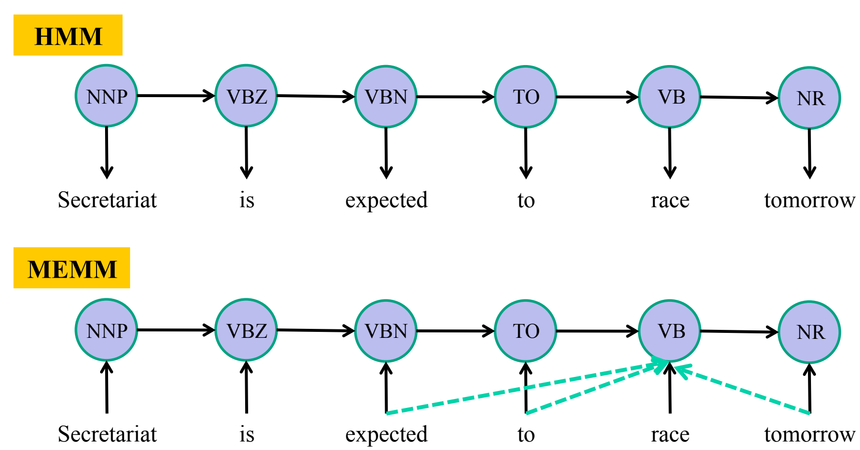


<The Washington Post> reported **<Obama>**' s view on the oil crisis.

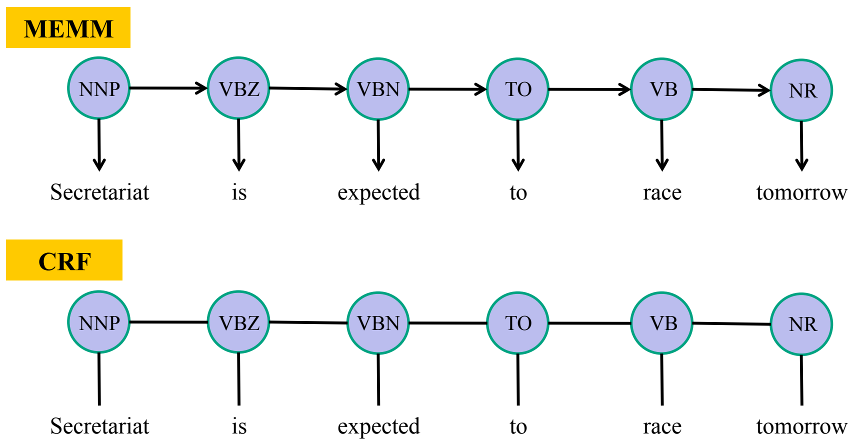
HMM v.s. MEMM



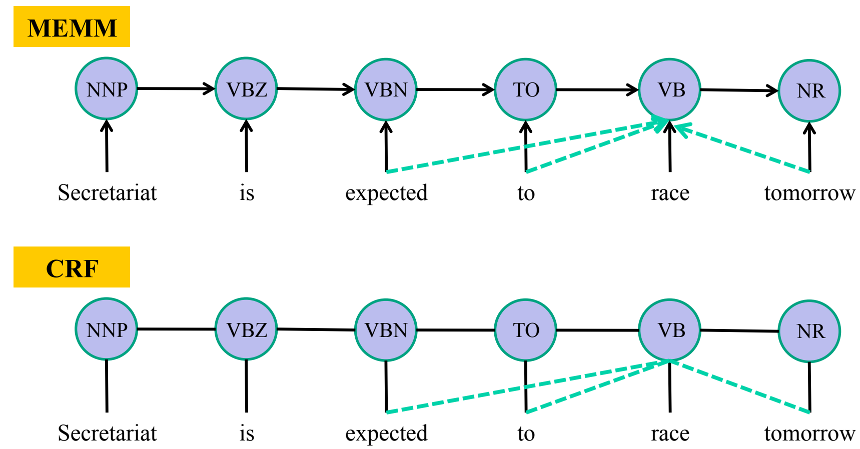
HMM v.s. MEMM



MEMM v.s. CRF



MEMM v.s. CRF



Conditional Random Fields

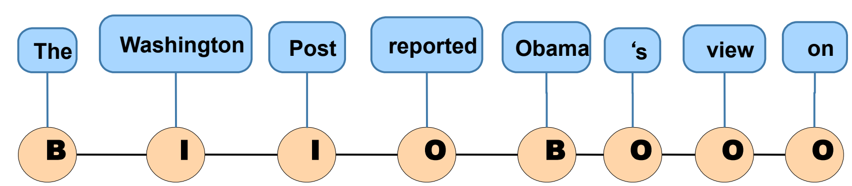
- Discriminative training
- Incorporation of arbitrary non-independent features (past + future)
  - semantic class, suffixes, constituent type, etc.

- Perform better than related classification and generative models (e.g. HMMs)
  - Part-of-speech tagging [Lafferty et al., 2001]
  - Noun phrase chunking [Sha and Pereira, 2003]
  - Human protein name tagging [Bunescu et al. 2004]

[Lafferty et al., 2001]

Identifying Sources of Opinions …as an *Information Extraction* task

- Sequence tagging



<The Washington Post> reported **<Obama>**' s view on the oil crisis.

Features for Source Extraction

- **Syntactically…**
  - *mostly* noun phrases
- **Semantically…**
  - entities that can bear opinions
- **Functionally…**
  - linked to opinion expressions

Features for Source Extraction

- Words [-4,+4]
- Capitalization
- Part-of-speech tags [-2,+2]
- Opinion phrase lexicon
  - Derived from training data
  - Wiebe et al.'s [2002] 500+ word lexicon
- Shallow semantic class information
  - Sundance partial parser and named entity tagger
  - WordNet hypernym
- Constituent type
- Grammatical role
  - Collins' parser
- Task-specific combinations
  - E.g., Parent contains opinion word

Evaluation

- MPQA data set (www.cs.pitt.edu/mpqa)
  - ~550 documents
  - Manually annotated w.r.t. fine-grained opinion information
  - Provides gold standard
- Automatically derive training/test examples
- 10-fold cross-validation
- Evaluation measures
  - Precision
  - Recall
  - F-measure

Results: Opinion Holders

- >82% precision (accuracy)
- ~60% recall (coverage)
- 69.4 F-measure

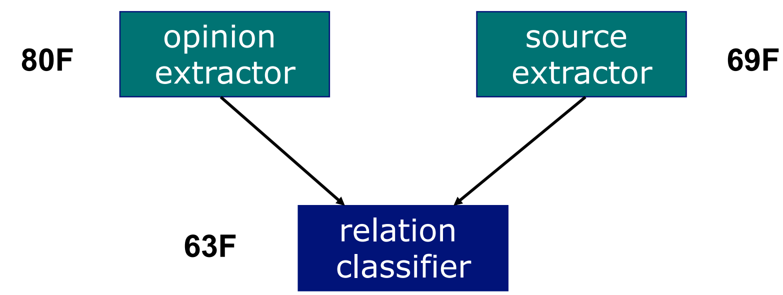
So there's a lot of room for improvement…

Errors

- False positives
  - Perhaps this is why **Fidel Castro has not spoken out** against what might go on in Guantanamo.
- False negatives
  - And for this reason, too, they have a moral duty to **speak out**, as **Swedish Foreign Minister Anna Lindh, among others**, did yesterday.
  - In particular, **Iran and Iraq are at loggerheads** with each other to this day.

Extracting and Linking to Opinions

- To be useful, we need to link sources to their opinions
  - <source> expresses <opinion>



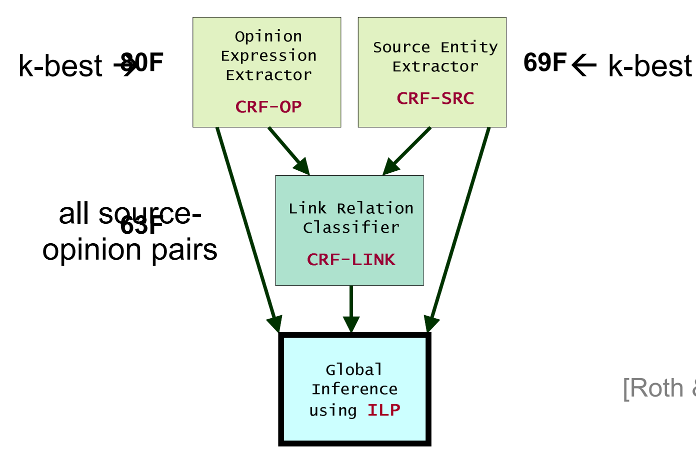
Research Trend: Structured Learning

- Beyond simple classification tasks
- Dependent/output variable has an internal structure
- Multiple dependent/output variables with dependencies or constraints among them

E.g. syntactic parse tree, source-expresses-opinion relation

Opinion Frame Extraction via CRFs and ILP

- Joint extraction of entities and relations



[Roth & Yih, 2004]

[Choi et al., EMNLP 2006]

Constraints

- Binary integer variables *O_i, S_j, L_i,j*
  - Weights for *O_i, S_j, L_i,j* are based on probabilities from individual classifiers
- Constraints

$\forall i,\ O_i = \sum_j L_{i,j}$: link coherency(only one link from each opinion)

$\forall j,\ S_j + A_j = \sum_i L_{i,j}$: link coherency(upto two links from each source)

$\forall j,\ A_j - S_j \le 0$: link coherency(preferably one link from each source)

$\forall i,j, i<j,\ X_i + X_j = 1,\ X \in \{S, O\}$

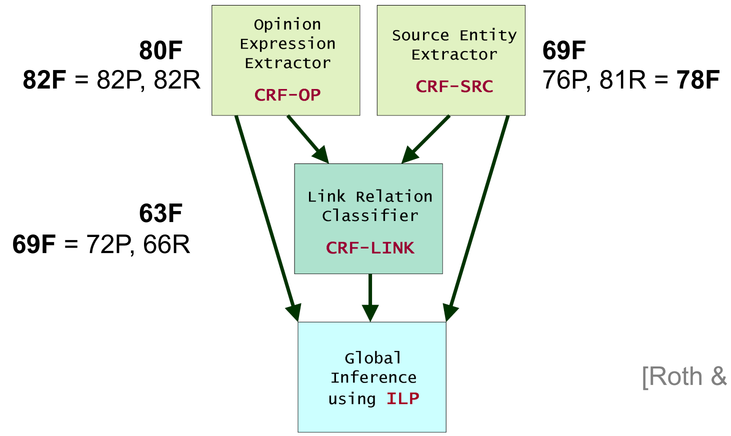
: entity coherency(for all pairs of entities with overlapping spans)

- Objective function

$$f = \sum_i (w_{O_i} O_i) + \sum_i (\overline{w}_{O_i} \overline{O}_i) + \sum_j (w_{S_j} S_j) + \sum_j (\overline{w}_{S_j} \overline{S}_j) + \sum_{i,j} (w_{L_{i,j}} L_{i,j}) + \sum_{i,j} (\overline{w}_{L_{i,j}} \overline{L}_{i,j})$$

Opinion Frame Extraction via CRFs and ILP

- Joint extraction of entities and relations



[Roth & Yih, 2004]

Update on Results

- Extracting Opinion Expressions with semi-Markov Conditional Random Fields (Yang & Cardie, EMNLP 2012)
  - Handles indirect (as well as direct) opinion expressions
    - The International Committee of the Red Cross, [as usual]$_{[ESE]}$, [has refused to make any statements]$_{[DSE]}$.
    - The Chief Minister [said]$_{[DSE]}$ that [the demon they have reared will eat up their own vitals]$_{[ESE]}$.
- Joint Inference for Fine-Grained Opinion Analysis (Yang & Cardie, submitted)
  - Handles source + target entities
  - Handles is-from + is-about relations
  - Handles cases with implicit sources

Sources and Coreference

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.

Somewhat different techniques…different talk ☺.

Appinions Demo

- Opinion frames
  - Opinion expressions
  - Opinion holder
  - Polarity
  - Topic
- Aggregation
  - w.r.t. opinion holder or topic
  - over time
- Use to derive social networks
  - Determine key influencers for particular topics