Information Extraction

- **Today**
  - Learning approaches
    - Weakly supervised methods
    - Fully automatic methods for IE
    - Named entity identification
    - Sequence-tagging methods
      - MEMM's
      - Opinion extraction
    - ILP for relation extraction

Hidden Markov Models

- A set of $N$ states
- A transition probability matrix $A$, each $a_{ij}$ representing the probability of moving from state $i$ to state $j$, s.t. $\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$
- A sequence of $T$ observations, each one drawn from a vocabulary $V = V_1, V_2, ..., V_N$
- A sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation $o_t$ being generated from a state $i$

HMMs for entity detection

- American Airlines
- A unit of AMR Corp.
- Immediately matched
- The move
- A spokesman
- Wagner said

HMM for weather Prediction

- $B_1$:
  - $P(HOT)$
  - $P(COLD)$
- $B_2$:
  - $P(HOT)$
  - $P(COLD)$

Could be Victim, Target, Person-IN, Person-OUT, etc.
HMM equations

Decoding/inference in HMMs

Classification approach???

End-to-end process
Feature extraction

- We’d like to be able to include lots of features as in classification-based approaches (e.g. SVMs, dtrees)

Not possible with HMMs

Maximum entropy Markov model (MEMM)

MEMM equations

MEMM for p-o-s tagging

- Condition on many features of the input
  - Capitalization
  - Morphology
  - Earlier words
  - Earlier tags
Decoding/inference in MEMMs

Information Extraction

- **Today**
  - Learning approaches
    - Weakly supervised methods
    - Fully automatic methods for IE from structured text
    - Sequence-tagging methods
      - MEMM's
      - Opinion extraction
    - ILP for relation extraction

Relation extraction

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)
Identifying Sources of Opinions

- Via CRF’s (extension of MEMM’s)

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

<The Washington Post> criticized <Obama>’s view on the oil crisis.
Results: Opinion Holders

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

• Better than a (very good!) pattern-learning IE approach (Riloff)
• Better than (very good!) semantic role labeling algorithms (Roth)
• But 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>

Joint extraction of entities and relations

k-best →

all source-opinion pairs

Opinion Expression Extractor
CRF-OP

Source Entity Extractor
CRF-SRC

Link Relation Classifier
CRF-LINK

Global inference using ILP

[Roth & Yih, 2004]
Constraints

- Binary integer variables $O_i, S_j, L_{ij}$
  - Weights for $O_i, S_j, L_{ij}$ are based on probabilities from individual classifiers

- Constraints
  $\forall i, \quad O_i = \sum_j L_{ij}$
  - Link coherency (only one link from each opinion)
  
  $\forall j, \quad S_j + A_j = \sum_i L_{ij}$
  - Link coherency (up to two links from each source)
  
  $\forall j, \quad A_j - S_j \leq 0$
  - Link coherency (preferably one link from each source)
  
  $\forall i,j, i < j, \quad X_i + X_j = 1, \quad X \in \{S, O\}$
  - Entity coherency (for all pairs of entities with overlapping spans)

- Objective function
  $f = \sum_i (w_i O_i) + \sum_j (w_j S_j) + \sum_j (w_j S_j) + \sum_{i,j} (w_{ij} L_{ij}) + \sum_{i,j} (w_{ij} L_{ij})$