SNS Text Mining and Search

Xiaohua Liu and Furu Wei

Natural Language Computing Group

Microsoft Research Asia

Agenda

- QuickView: A Research Platform of SNS Text Mining and Search
- SNS Text Mining
  - Semantic Role Labeling
  - Sentiment Analysis
- Future Work

Agenda

- QuickView: A Research Platform of SNS Text Mining and Search
- SNS Text Mining
  - Semantic Role Labeling
  - Sentiment Analysis
- Future Work

Microsoft QuickView beta

A Research Platform of SNS Text Mining and Search

System Overview

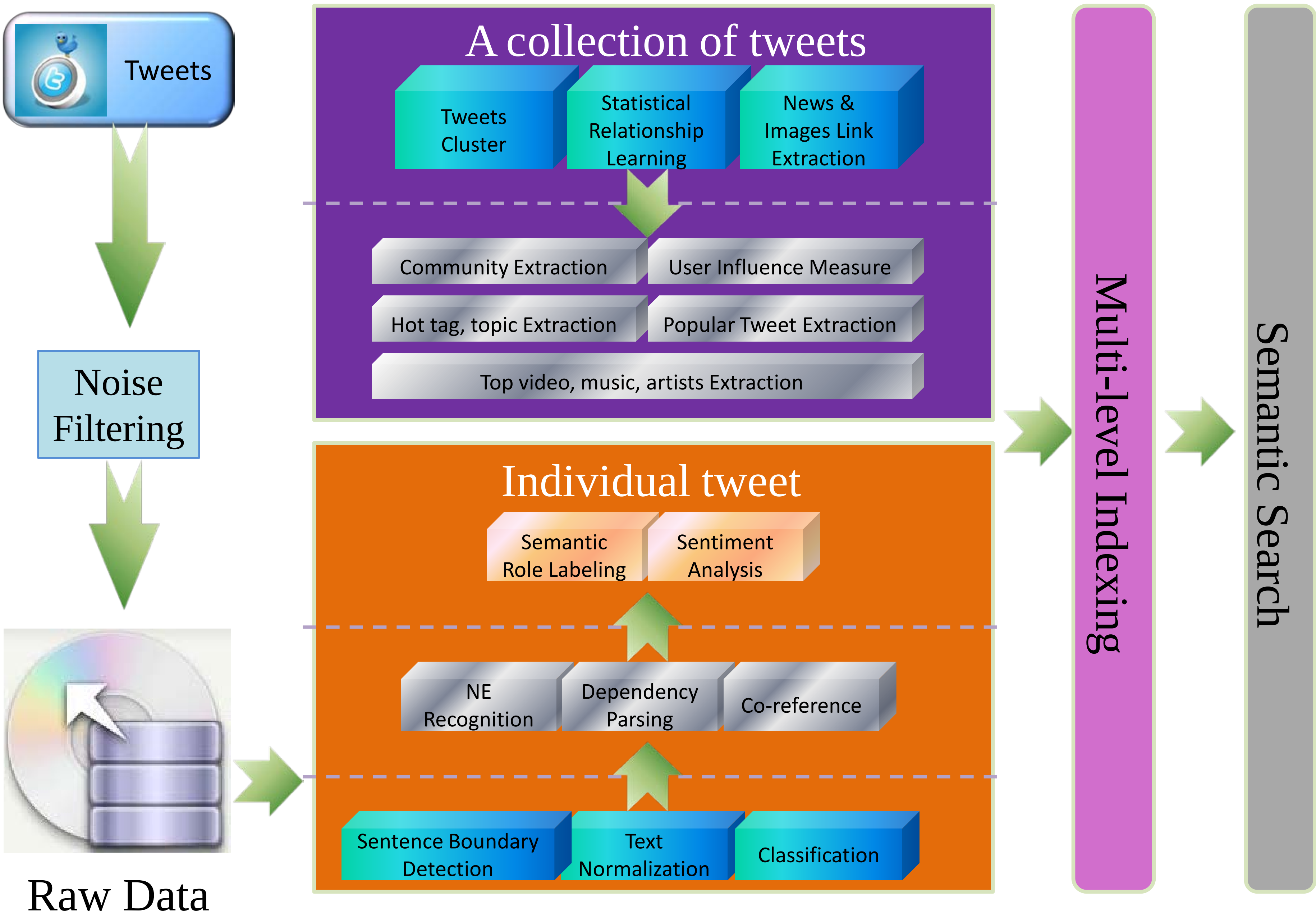
- Input
  - Tweets
  - Facebook updates
  - Chinese Weibo, e.g., Sina microblog
  - …
- Process
  - NLP pipeline for social content: POS, shallow parsing, SRL, etc.
  - Single instance level mining: entities, events, opinions, etc.
  - Collectively mining: hot topics, hot entities, hot opinions, etc.
- Output
  - A multi-level index of social contents
  - Recent statistical information: e.g., hot topics

Data processed per day

- Tweets
  - 1,500M English tweets
- Facebook
  - 15M updates
  - 100M likes
  - 8M fan pages
- SINA weibo
  - 1.5M micro-blogs

SNS Research Platform

- Twitter, FaceBook, Chinese micro-blogs
- 3000+ News sources
- Multiple languages
- Text mining
- Personalized search and recommendation


Tweets
Noise Filtering
Raw Data
A collection of tweets
Tweets Cluster
Statistical Relationship Learning
News & Images Link Extraction
Community Extraction
User Influence Measure
Hot tag, topic Extraction
Popular Tweet Extraction
Top video, music, artists Extraction
Individual tweet
Semantic Role Labeling
Sentiment Analysis
NE Recognition
Dependency Parsing
Co-reference
Sentence Boundary Detection
Text Normalization
Classification
Multi-level Indexing
Semantic Search

The QuickView Demo

Our Current Research

- Sentiment analysis
  - Target-dependent Twitter Sentiment Classification (ACL 2011)
  - User-level sentiment analysis incorporating social networks (KDD 2011)
  - A Graph-based Hashtag Sentiment Classification Approach. (CIKM 2011)
- NER
  - Recognizing Named Entities in Tweets (ACL 2011)
  - Mining Entity Translations from Comparable Corpora: A Holistic Graph Mapping Approach. (CIKM 2011)
- SRL
  - Collective Semantic Role Labeling (IJCAI 2011)
  - Improving Semantic Role Labeling using Semi-supervised Learning (AAAI 2011)
  - Semantic Role Labeling for News Tweets (Coling 2010)
- SNS Search
  - An Empirical Study on Learning to Rank of Tweets. (Coling 2010)
  - QuickView: Semantic Search For Tweets (SIGIR 2011 demo)

Agenda

- QuickView: A Research Platform of SNS Text Mining and Search
- SNS Text Mining
  - Semantic Role Labeling
  - Sentiment Analysis
- Future Work

Semantic Role Labeling for Tweets

Outline

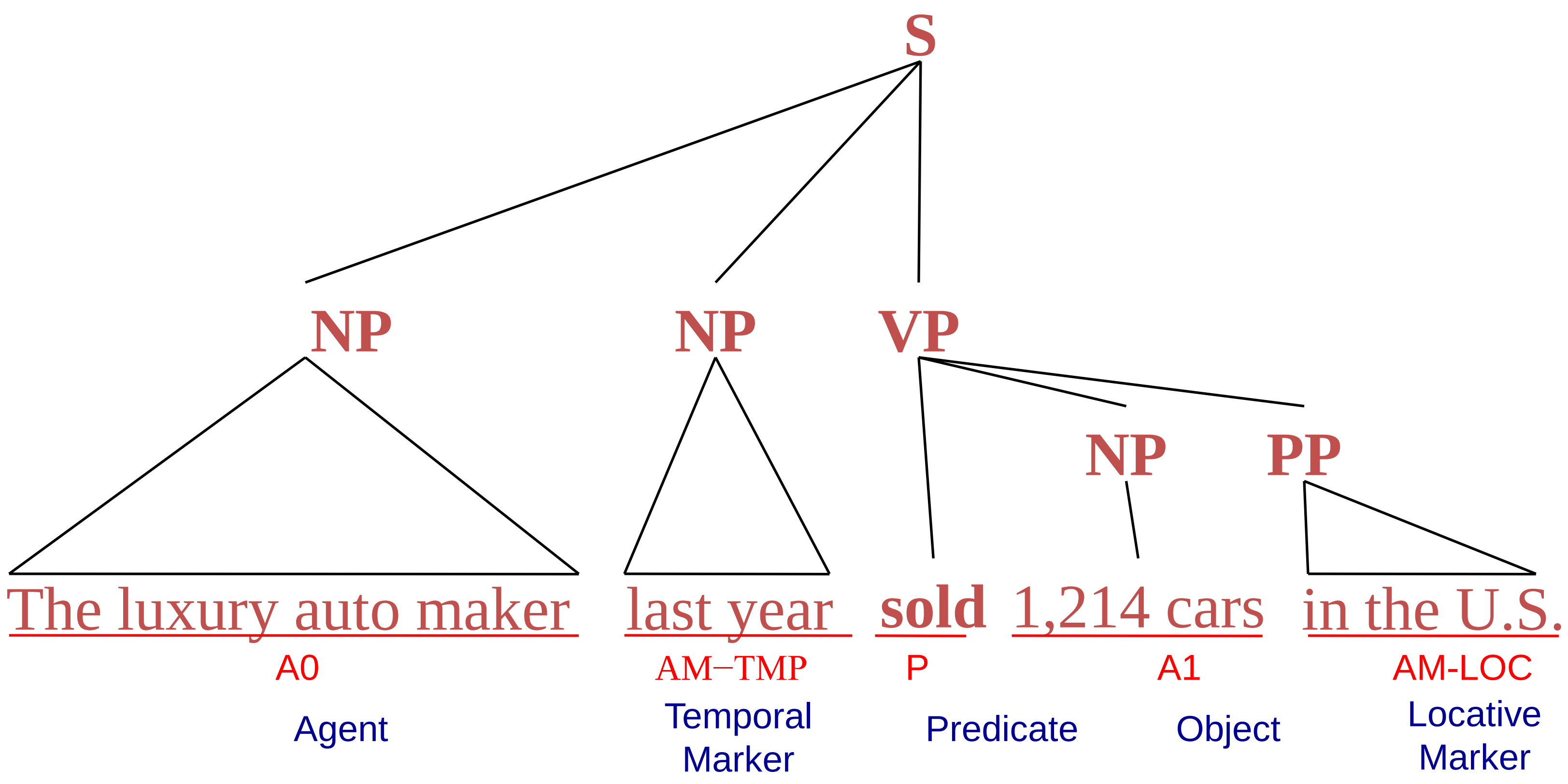
- Introduction
  - SRL task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

Outline

- Introduction
  - SRL task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

Semantic Role Labeling

- Detect basic event structures such as *who* did *what* to *whom*, *when* and *where*



Predicate

- Verbal predicate (PropBank)
  - *Chile [**earthquake**] $_{A0}$ shorten the [**day**] $_{A1}$*
- Other types of predicate (NomBank)
  - *[**Her**]$_{A0}$ gift of [**a book**]$_{A1}$ [**to John**]$_{A2}$*

Predicate Arguments

- Core arguments
  - A0, A1: agent and patient
- 13 adjunctive arguments
  - Temporal, manner, location, etc.
- Phrase level vs. word level argument
  - Word level: *Chile [**earthquake**]* $_{A0}$ *shorten the [**day**]* $_{A1}$
  - Phrase level: *[**Chile earthquake**]* $_{A0}$ *shorten [**the day**]* $_{A1}$

Evaluation of SRL

- Evaluation metrics
  - Precision
    - How many output labels are correct
  - Recall
    - How many labels in the gold standard dataset are correctly labeled
  - F1 score
    - The harmonic mean of Precision and Recall or
    - $F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$

Evaluation of SRL

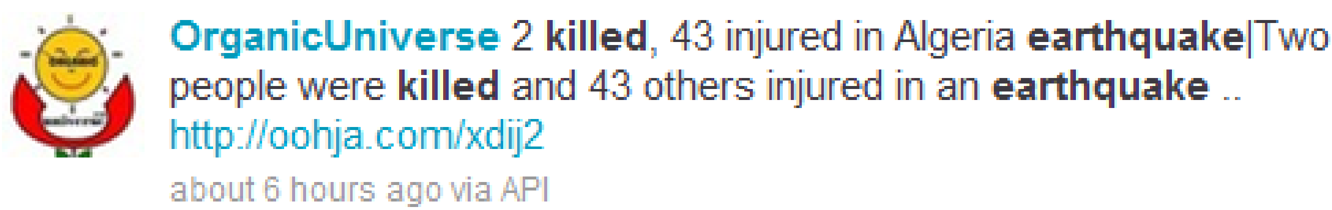
- Test datasets (from PropBank)
  - WSJ (Wall Street Journal) : mainly news
  - Brown: more balanced corpus, including news, reports and others
- The state-of-the-art results
  - CoNLL-2005 :81.52% F1 on WSJ
  - CoNLL-2008: 87.69% F1 on WSJ, 69.06% F1 on Brown
  - CoNLL-2009:80.47 F1 on WSJ
- Best systems are pipelined or based on MLN

Outline

- Introduction
  - SRL Task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

SRL Helps Twitter Search

- Twitter search is now keyword search, unable to answer questions, like how many people were killed in Algeria earthquake?



SRL Helps Twitter Search (2)

- SRL extracts who acted what
  - *oh yea and Chile [**earthquake**] $_{A0}$ the earth off it's axis according to NASA and **shorten** the [**day**] $_{A1}$ by a wee second :-(* → *[**earthquake**] $_{A0}$ **shorten** the [**day**] $_{A1}$*
  - Beyond keyword search, e.g., *what **shorten** the **day***?

SRL Helps Twitter Search

- SRL abstracts away syntax variances
  - *Chile Earthquake Shortened Earth Day*
  - *The Chile earthquake shortened the length of an Earth day*
  - *…*
  - → *[**earthquake**]* $_{A0}$ ***shorten*** *[**day**]* $_{A1}$

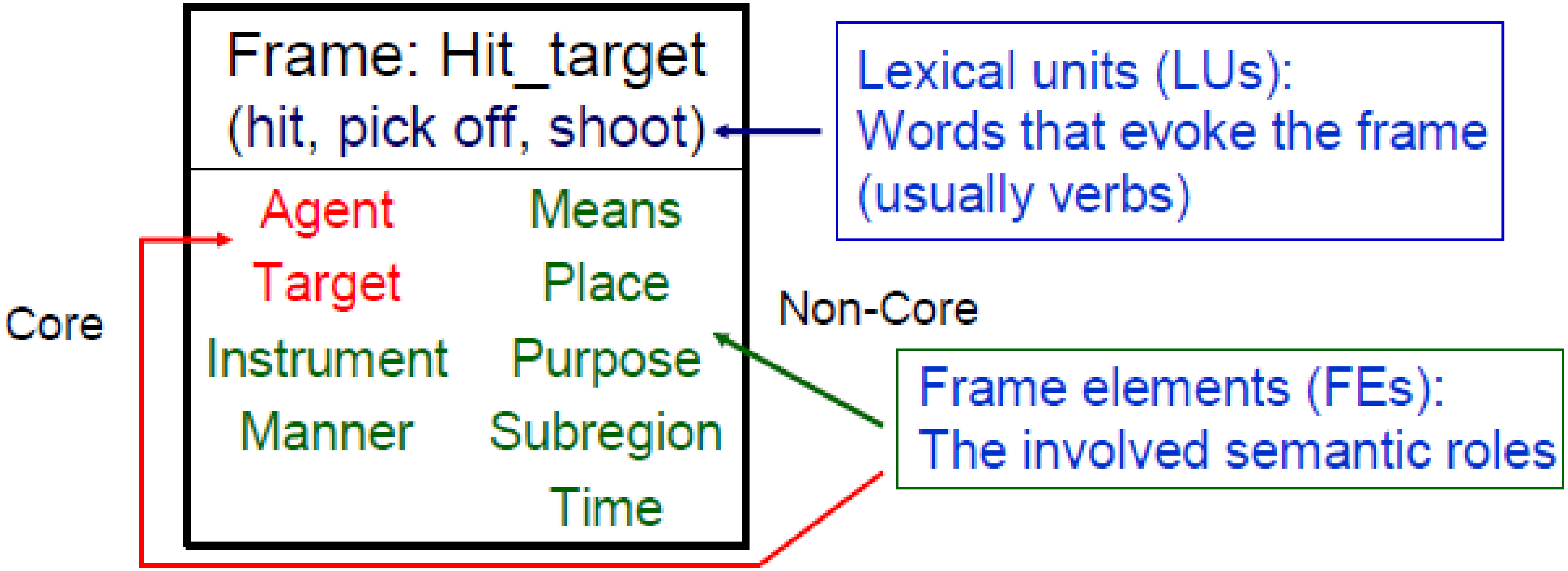
Outline

- Introduction
  - SRL Task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

FrameNet (Fillmore et al., 2004)

- Computational frame lexicon + corpus of examples annotated with semantic roles (mostly BNC)
  - ~800 semantic frames
  - >9,000 lexical units
  - ~150,000 annotated sentences

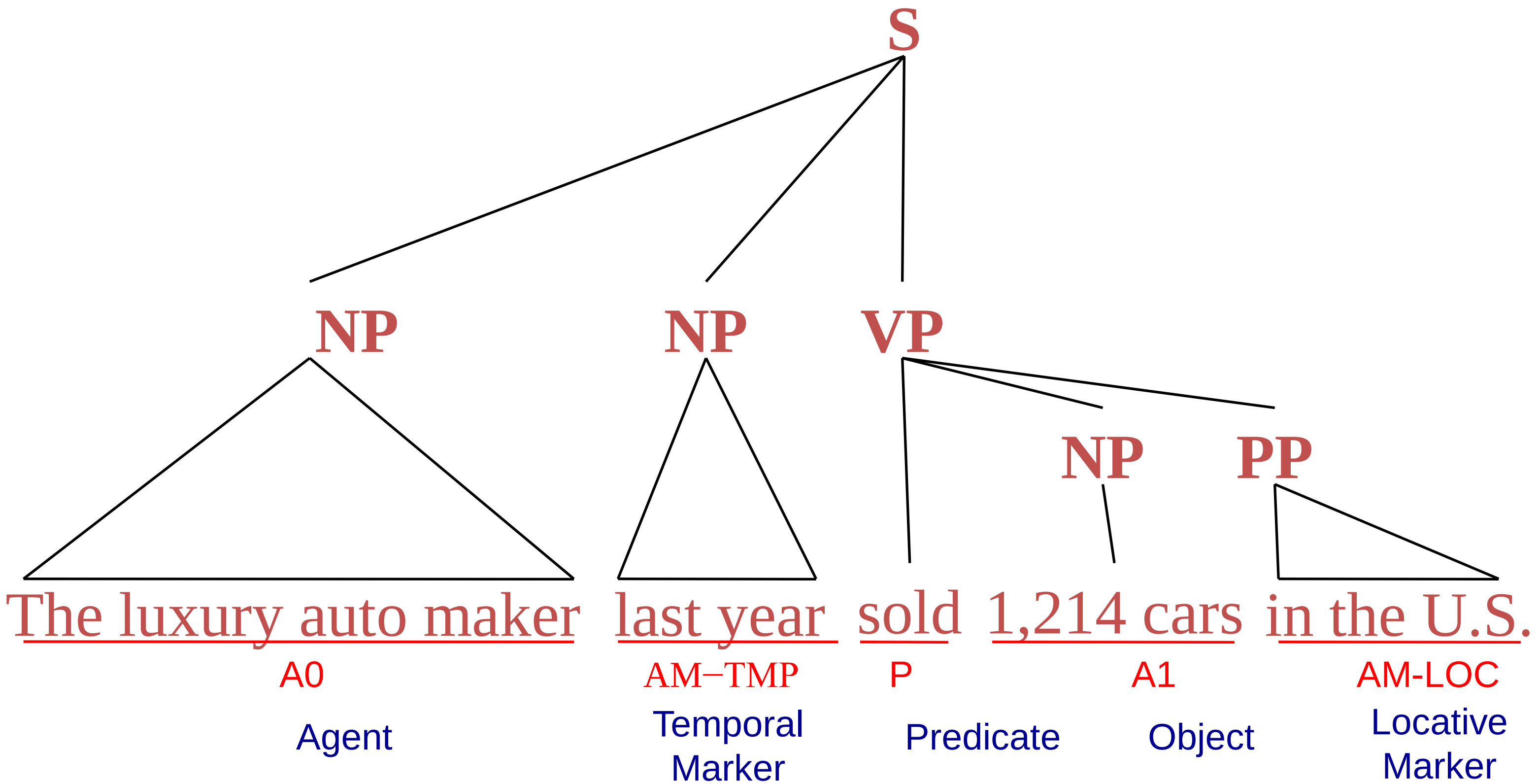
A Frame Example



[$_{\text{Agent}}$ *Kristina*] ***hit*** [$_{\text{Target}}$ *Scott*] [$_{\text{Instrument}}$ *with a baseball*] [$_{\text{Time}}$ *yesterday*].

PropBank (Palmer et al., 2005)

- The primary resource for research in SRL
- Annotation of all verbal predicates in Penn Treebank


S
NP
NP
VP
NP
PP
The luxury auto maker
last year
sold
1,214 cars
in the U.S.
A0
AM-TMP
P
A1
AM-LOC
Agent
Temporal Marker
Predicate
Object
Locative Marker

An Example: Argument Structure Depends on Verb and Its Meaning

sell.01: commerce: seller

A0="seller" (*agent*); A1="thing sold" (*theme*); A2="buyer" (*recipient*); A3="price paid"; A4="benefactive"

[Al Brownstein]$_{A0}$ sold [it]$_{A1}$ [for \$60 a bottle]$_{A3}$

sell.02: give up

A0="entity selling out"

[John]$_{A0}$ sold out

sell.03: sell until none is/are left

A0="seller"; A1="thing sold"; ...

[The new Harry Potter]$_{A1}$ sold out [within 20 minutes]$_{AM-TMP}$

NomBank (Meyers et al., 2004)

- Annotation of the nominal predicates in Penn TreeBank
  - [IBM] $_{A0}$'s appointment of [John] $_{A1}$
  - The appointment of [John] $_{A1}$ by [IBM] $_{A0}$
  - [John] $_{A1}$ is the current [IBM] $_{A0}$ appointee

Outline

- Introduction
  - SRL Task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

Typical systems

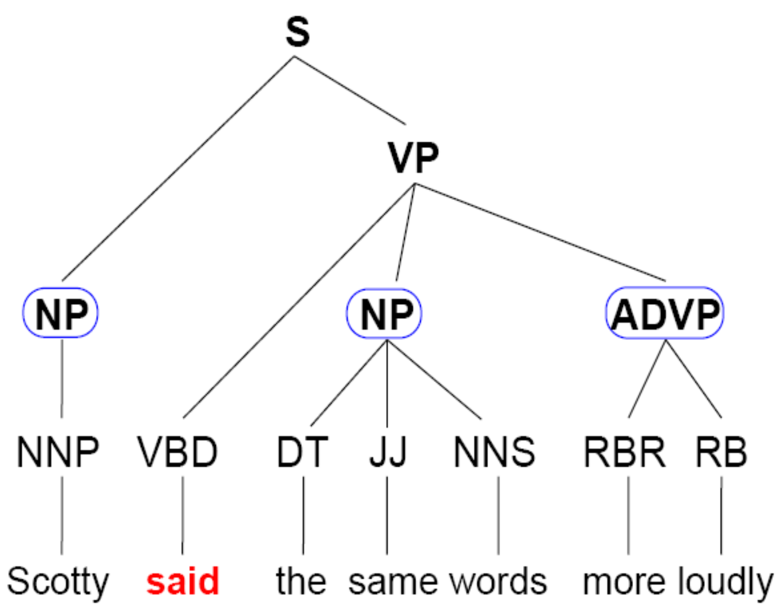
- Pipelined system
- System based on sequential labeling
- System using Markov Logic Networks
- Collective SRL (jointly conduct SRL on multi sentences)

Typical systems

- Pipelined system
- System based on sequential labeling
- System using Markov Logic Networks
- Collective SRL (jointly conduct SRL on multi sentences)

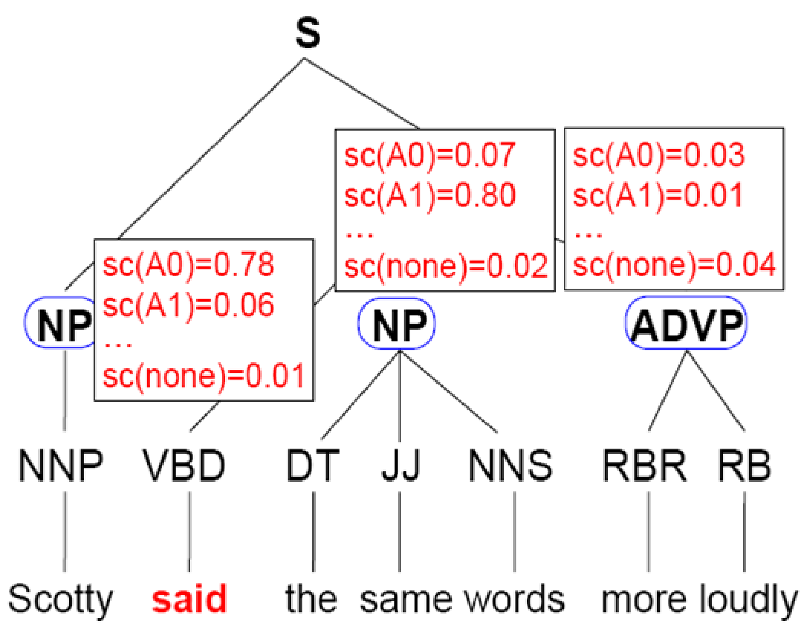
Pipelined SRL

- Argument candidates generation



Pipelined SRL

- Argument candidates generation
- Argument classification



Pipelined SRL

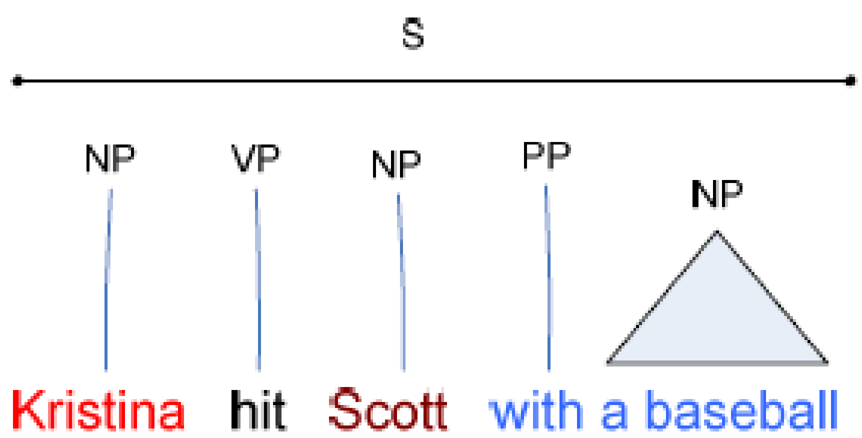
- Argument candidates generation
- Argument classification
- Global inference
  - Find the best solution from all possible solutions
    - E.g., Re-ranking of N best solutions(Haghighi et al., 2005; Toutanova et al., 2008)

Typical Systems

- Pipelined system
- System based on sequential labeling
- System using Markov Logic Networks
- Collective SRL (jointly conduct SRL on multi sentences)

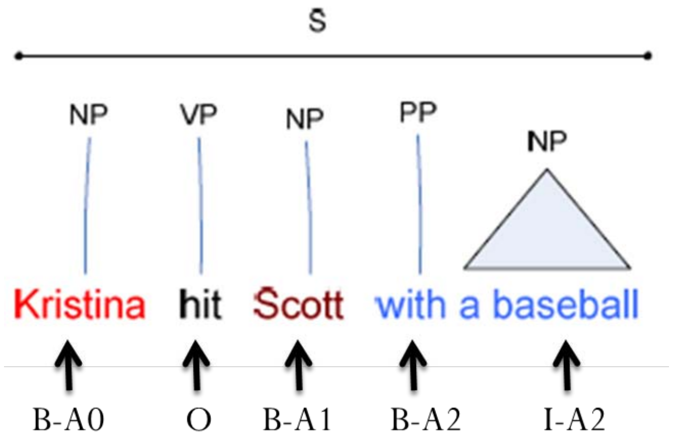
System Based on Sequential Labeling (Marques et al., 2005)

- Break into base chunks
  - Chunker: Yamcha (Kudo & Matsumoto, 2001)



System Based on Sequential Labeling

- Break into base chunks
- Labeling each chunk
  - B/I marks the beginning/ continuation of an argument span; and O non-arguments



Tool: CRF++ http://crfpp.sourceforge.net/

Typical Systems

- Pipelined system
- System based on sequential labeling
- System using Markov Logic Networks
- Collective SRL (jointly conduct SRL on multi sentences)

System using Markov Logic Networks (Sebastian Riedel and Ivan Meza-Ruiz,2008)

- Define formulae

$$lemma(p, +l_1) \wedge lemma(a, +l_2) \Rightarrow hasRole(p, a)$$

$$(role(p, a, r_1) \wedge r_1 \neq r_2 \Rightarrow \neg role(p, a, r_2))$$

System using Markov Logic Networks

- Define formulae
- Learning formula weights
  - To allocate high probability to correctly identified predicate argument structures

I swim $\rightarrow$ {lemma(1,I) , lemma(2, swim), isPredicate(2)} > {lemma(1,I) , lemma(2, swim), isPredicate(1)}

System using Markov Logic Networks

- Define formulae
- Learning formula weights
- Inference
  - Jointly determine predicate argument structures that best fit the formulae

Toolkit: thebeast http://code.google.com/p/thebeast/

Typical systems

- Pipelined system
- System based on sequential labeling
- System using Markov Logic Networks
- Collective SRL (jointly conduct SRL on multi sentences)

Task Definition of Collective SRL

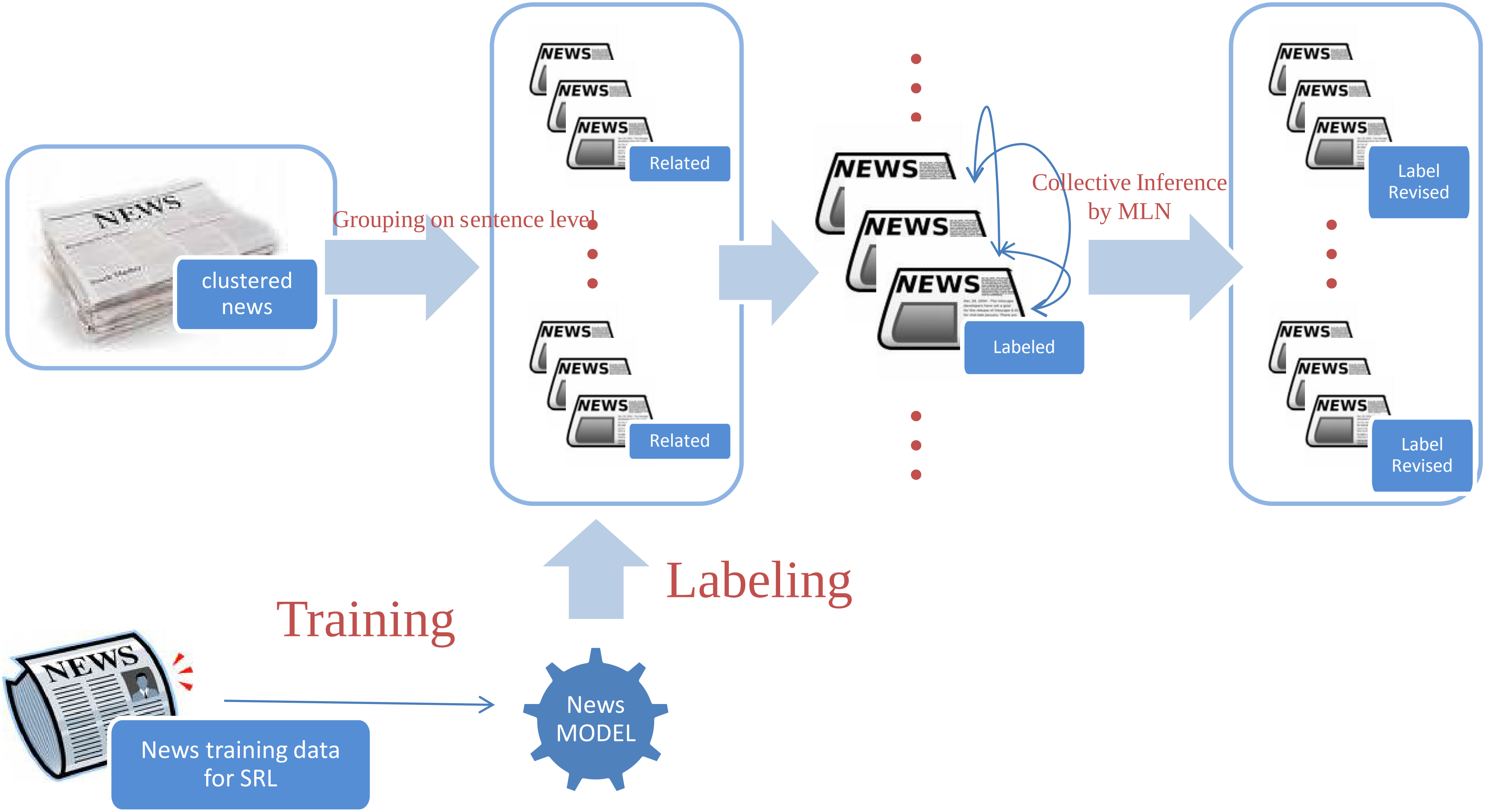
- Input: a set of sentences from news articles
  - *1. Hurricane **Ida**, the first Atlantic hurricane to **target** the U.S. this **year**, **plodded** yesterday **toward** the Gulf Coast…*
  - *2. Hurricane **Ida trudged toward** the Gulf Coast…*
  - *…*
- Output: predicate-argument-role structures
  - 1. (plodded, Ida, A0), (plodded, toward, AM-DIR), (target, Ida, A0), (target, U.S., A1), (target, year, AM-TMP)
  - 2. (trudged, Ida, A0), (trudged, toward, AM-DIR)

- Role sets (following PropBank)

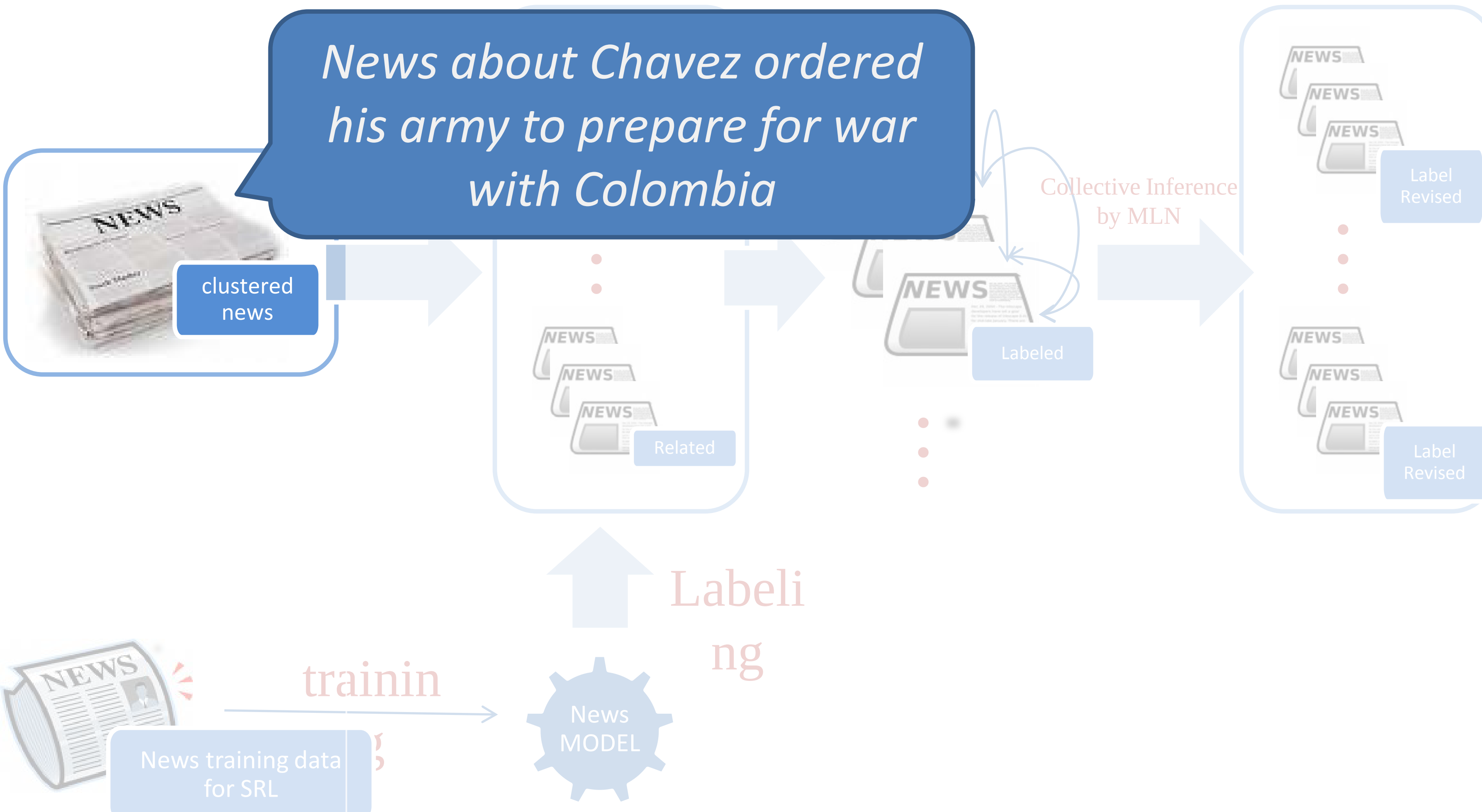
Collective SRL

- Motivated by the fact SRL on one sentence can help that on other differently phrased sentences with similar meaning
  - *A suicide* ***bomber*** *blew himself up Sunday in market in Pakistan's* ***northwest*** *crowded with shoppers ahead of a Muslim holiday,* ***killing*** *12 people, including a mayor who ….*
  - *Police in northwestern Pakistan say that a suicide* ***bomber*** *has* ***killed*** *at least 13 people and wounded dozens of others.*

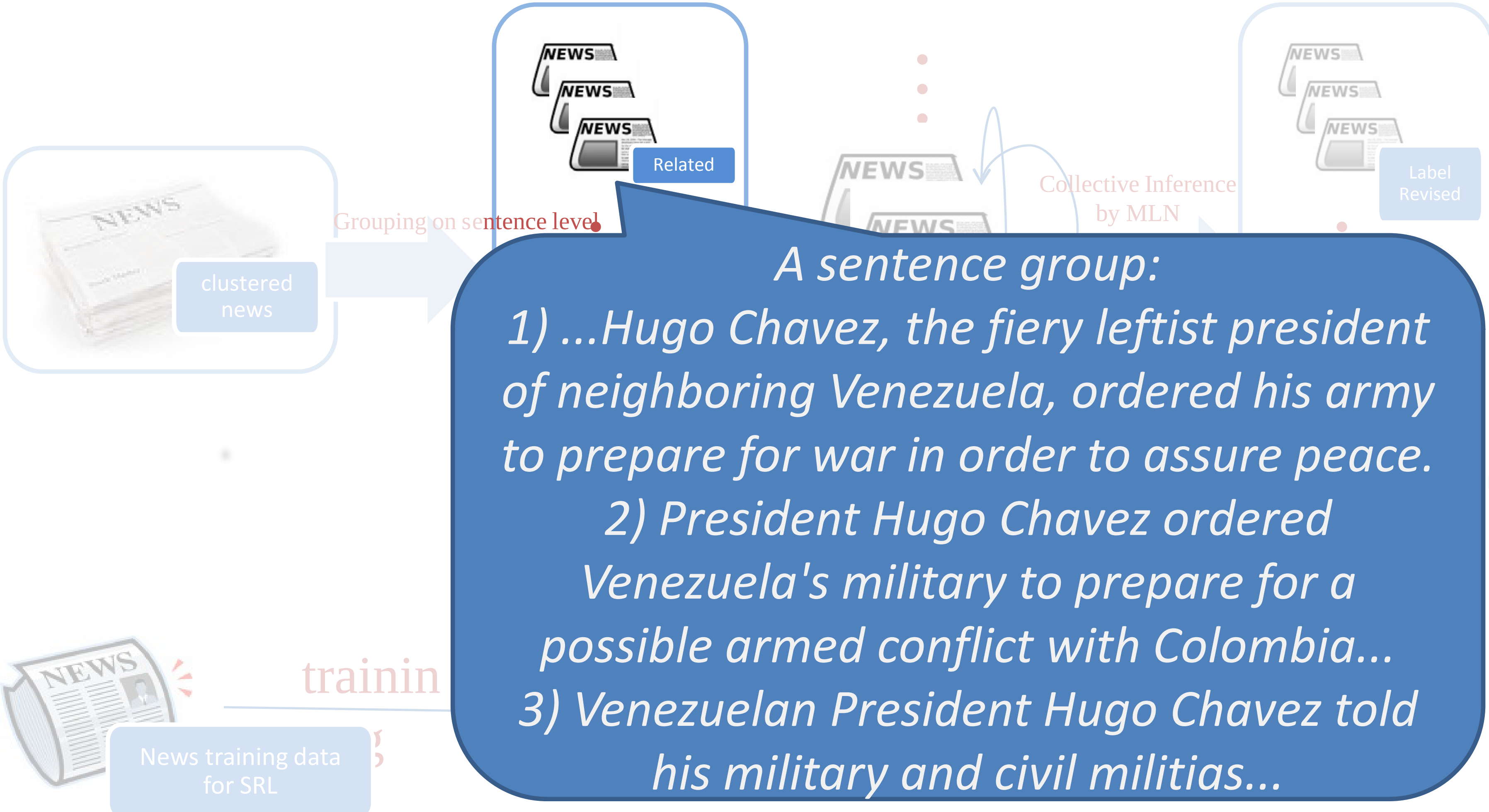
Implementation of Collective SRL
(Xiaohua Liu et al., 2010)



Implementation of Collective SRL



Implementation of Collective SRL



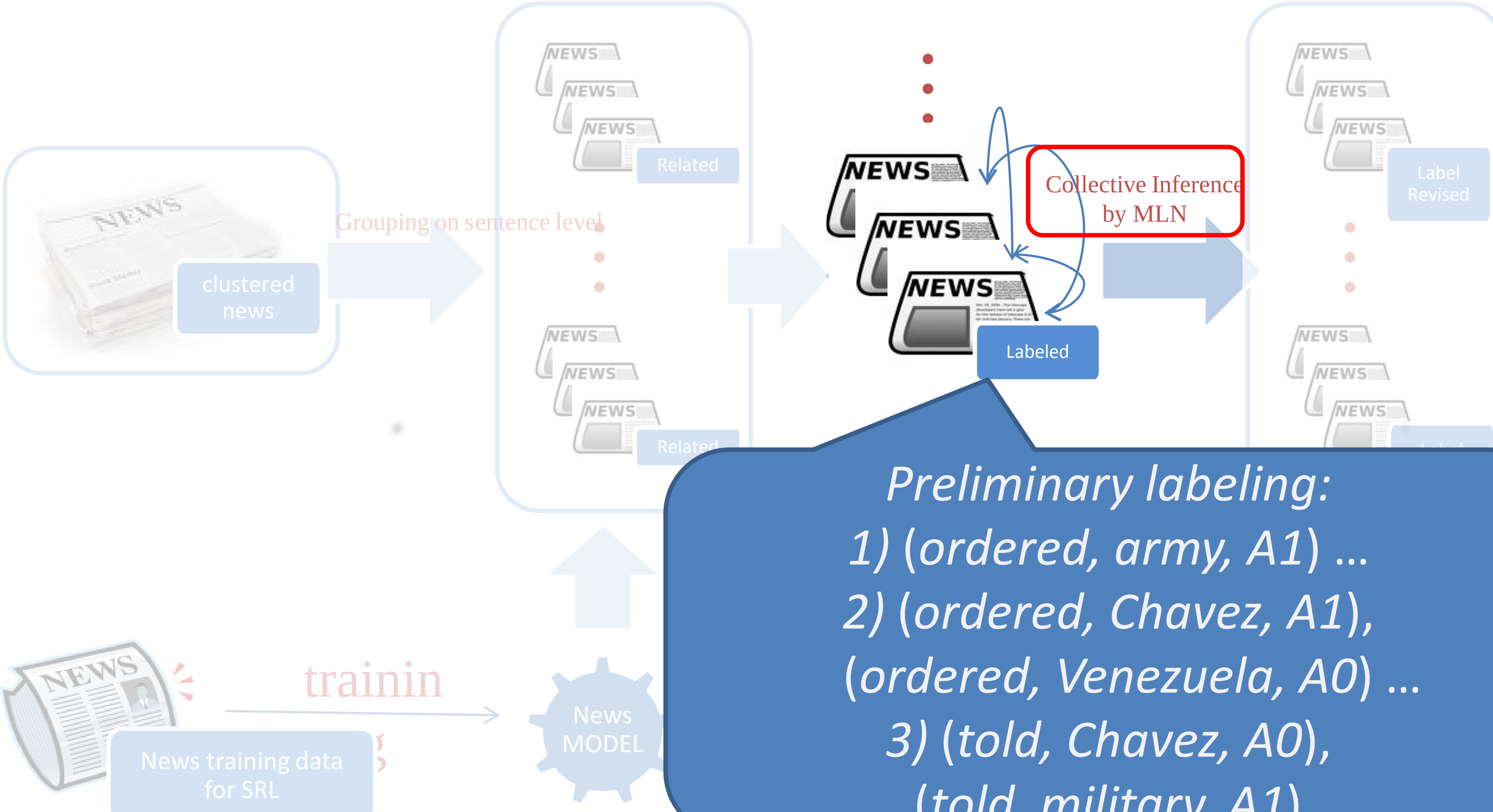
A sentence group:

1) …Hugo Chavez, the fiery leftist president of neighboring Venezuela, ordered his army to prepare for war in order to assure peace.

2) President Hugo Chavez ordered Venezuela's military to prepare for a possible armed conflict with Colombia…

3) Venezuelan President Hugo Chavez told his military and civil militias…

Implementation of Collective SRL



Preliminary labeling:
1) (ordered, army, A1) …
2) (ordered, Chavez, A1),
(ordered, Venezuela, A0) …
3) (told, Chavez, A0),
(told, military, A1) …

Implementation of collective SRL

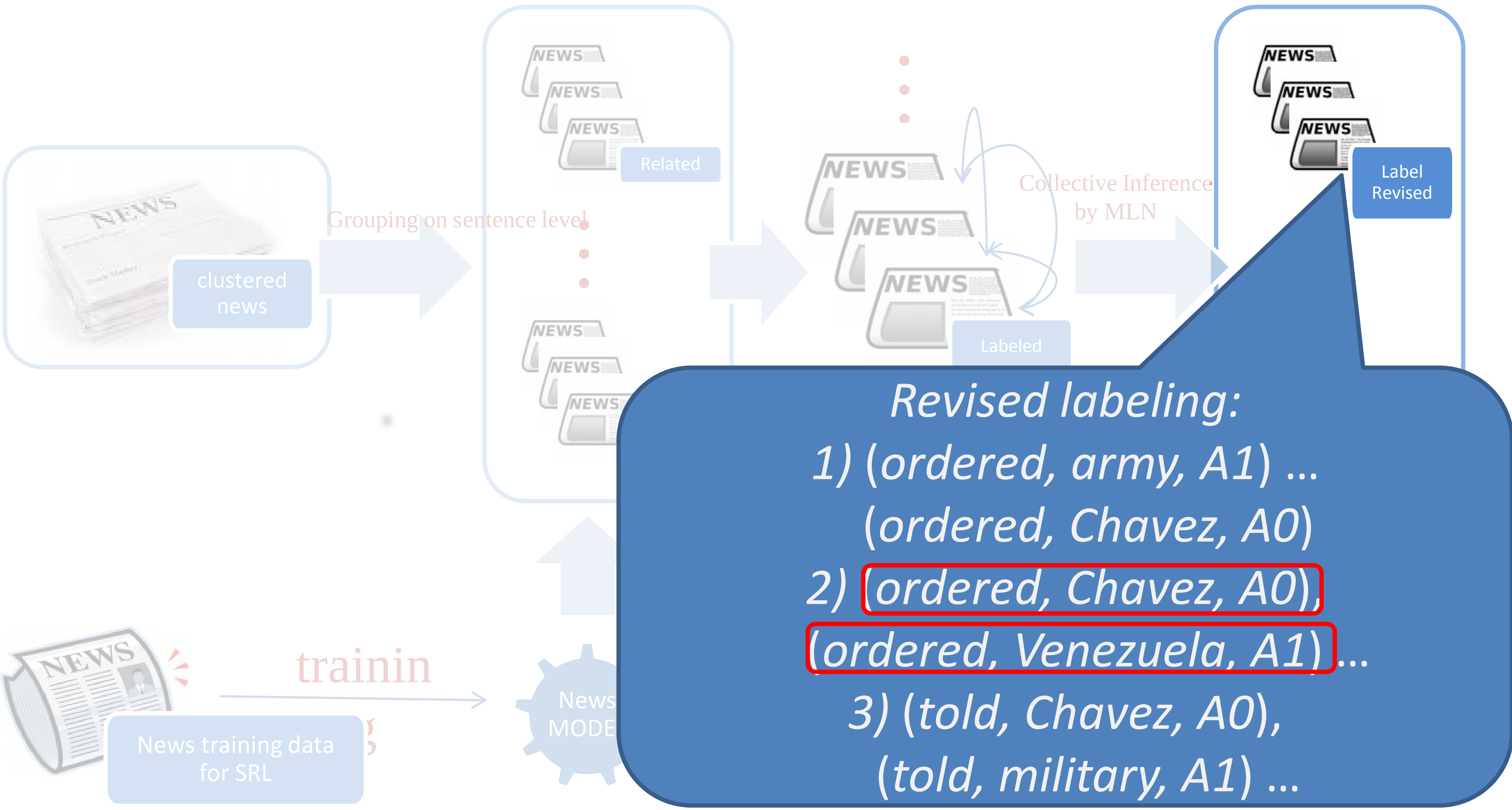
Collective inference with MLN:

introduce two formulas (the second is for collective inference)

$$role(s, p, a, +r) \Rightarrow final\_role(s, p, a, +r) \qquad (1)$$

$$s1 \neq s2 \wedge lemma(s1,p1,p\_lemma) \wedge lemma(s2,p2,p\_lemma) \wedge lemma(s1,a1,a\_lemma) \wedge lemma(s2,a2,a\_lemma) \wedge role(s2,p2,a2,+r) \Rightarrow final\_role(s1,p1,a1,+r) \qquad (2)$$

Implementation of Collective SRL

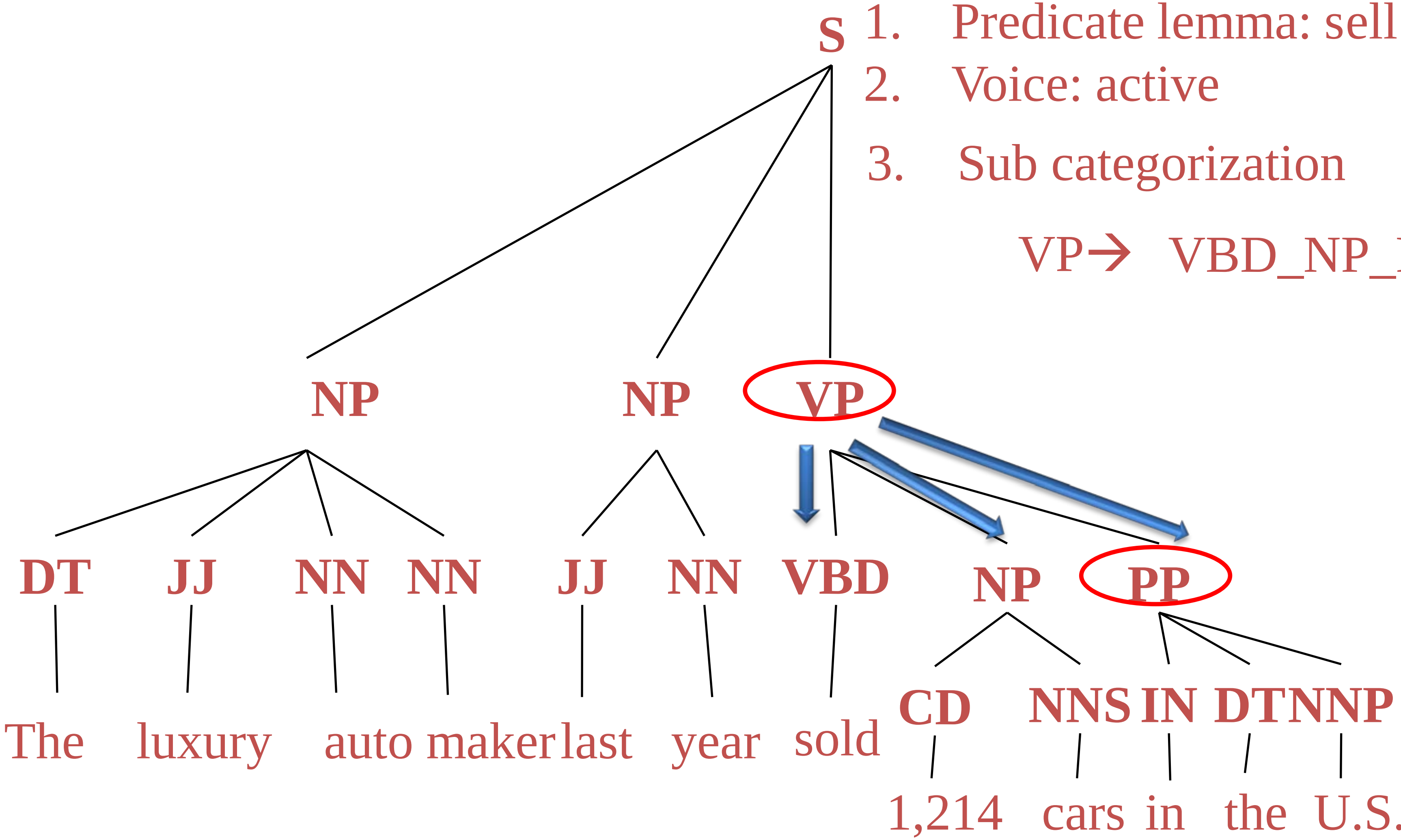


Experimental Results of Collective SRL

- Data
  - 1000 sentences from news clusters, grouped into 200 clusters
- Results (10-fold cross validation)

Systems	Precision	Recall	F-Score
Baseline	69.87%	59.26%	64.13%
Our method	67.01%	68.33%	67.66%

Feature Engineering in SRL



1. Predicate lemma: sell
2. Voice: active
3. Sub categorization

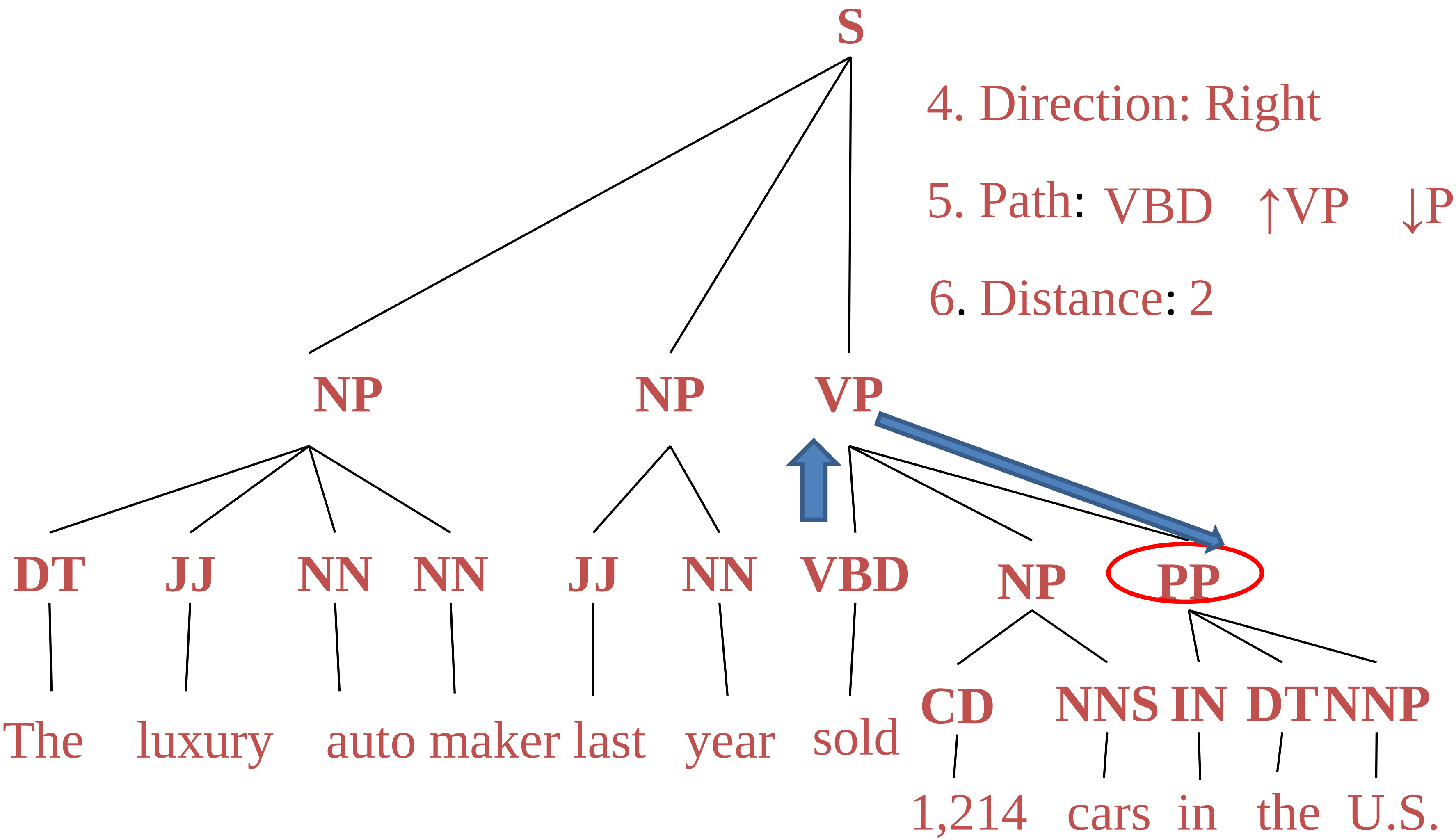
VP→ VBD_NP_PP

Feature Engineering in SRL

4. Direction: Right

5. Path: VBD ↑VP ↓PP

6. Distance: 2



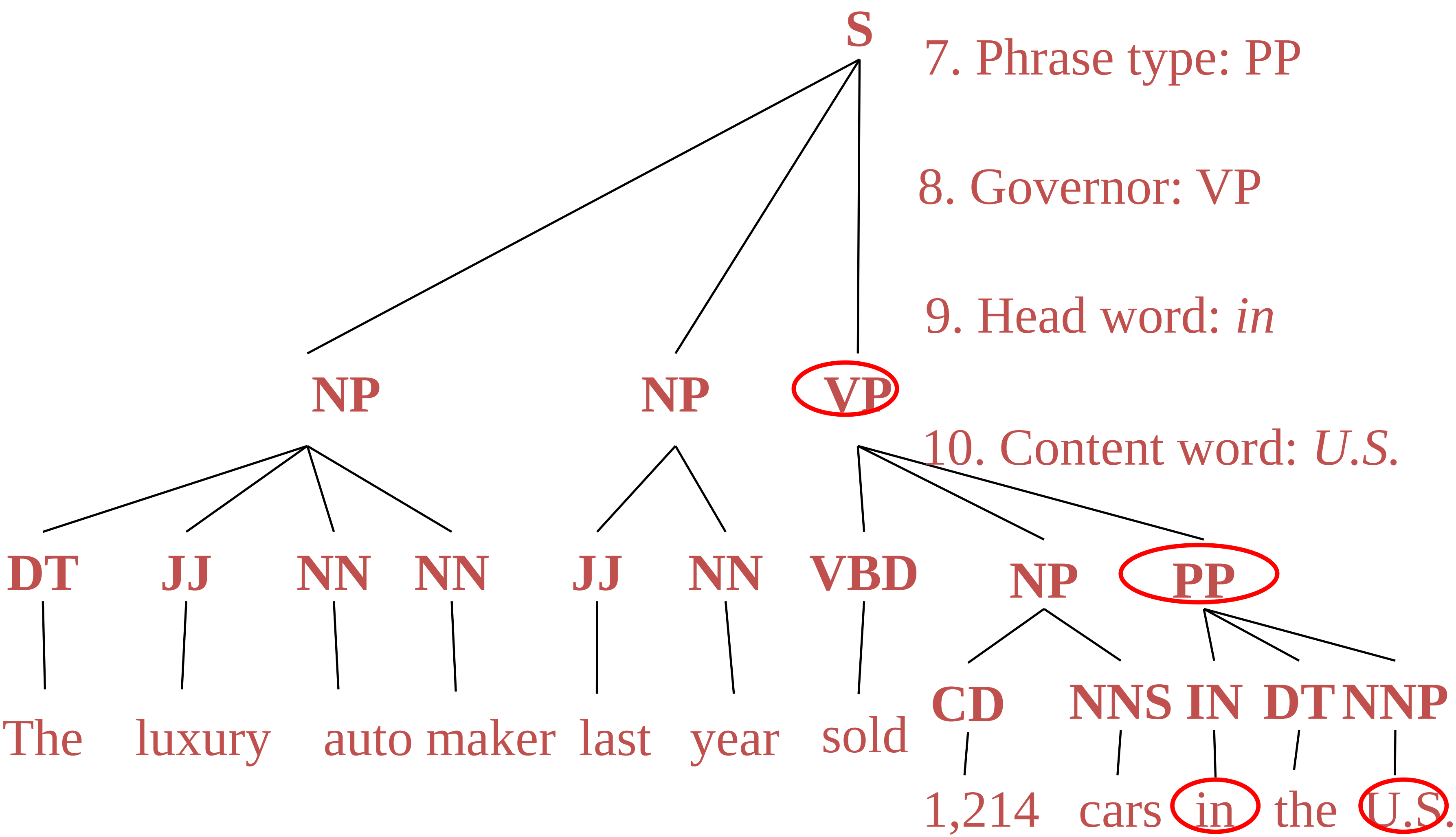
Feature Engineering in SRL

7. Phrase type: PP

8. Governor: VP

9. Head word: *in*

10. Content word: *U.S.*



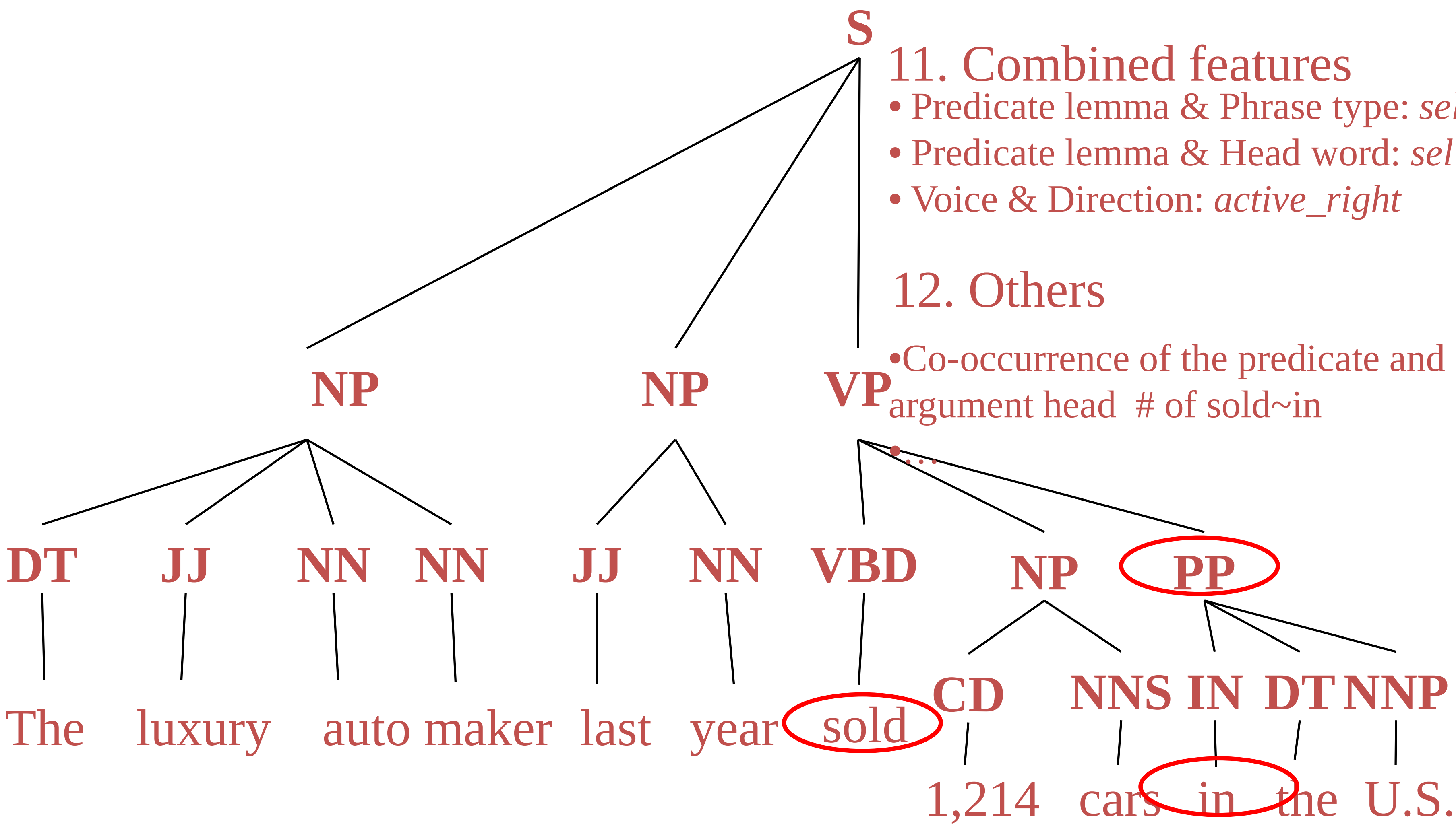
Feature Engineering in SRL

11. Combined features

- Predicate lemma & Phrase type: *sell*_PP
- Predicate lemma & Head word: *sell_in*
- Voice & Direction: *active_right*

12. Others

- Co-occurrence of the predicate and argument head # of sold~in
- …



Outline

- Introduction
  - SRL Task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

Task Definition of Tweet Level SRL

- Input: a tweet
  - *oh yea and Chile **earthquake** the earth off it's axis according to NASA and **shorten** the **day** by a wee second :-(*
- Output: predicate-argument structures
  - (shorten, earthquake, A0), (shorten, day, A1)

Research Challenges

- SRL system for news does not work: F1 90.0% → 43.3%
  - Reason: tweets are greatly different from news in written styles
    - Formal vs. informal; and Human edited vs. freely written; long vs. short;
  - Question: how to leverage existing SRL resources?

Research Challenges

- Building a SRL for tweets requires a huge number of training data
  - Manually labeling is prohibitively affordable
  - Question: can we train a system without much human labeling?

Outline

- Introduction
  - SRL Task definition
  - Application to twitter search
- General approaches to SRL
  - Resources
  - Typical systems
- SRL on tweets
  - Challenges
  - Method

SRL for Tweets

- SRL for news tweets
  - Focus on news tweets, tweets that report news
    - Relatively formal and less noisy
    - Easy to leverage related news
  - Using the redundancy between news and news tweets (Xiaohua Liu et al., 2010)
- SRL for general tweets
  - Collective SRL using clustering (Xiaohua Liu et al., 2011)
  - Enhancing SRL forTweets using Self-training (Xiaohua Liu et al., 2011)

SRL for News Tweets

Key Observations(1)

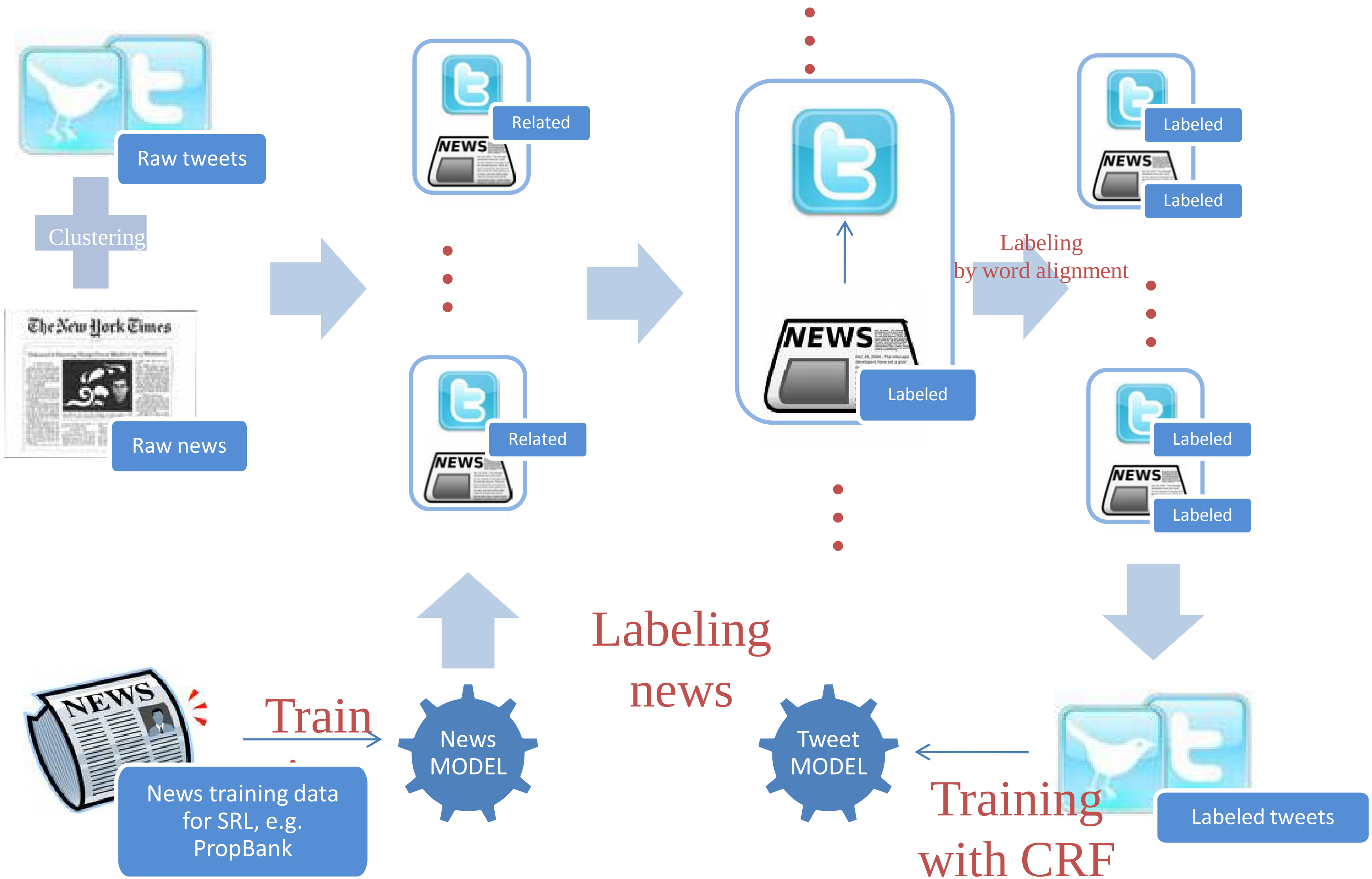
- There are strong content connection between news and tweets
  - Tweets directly excerpted from news articles or Links in tweets point to news articles

  - Official news that follow hot tweets
    - E.g., For *Chile earthquake* on Match 2nd, 2010, 261 news and 722 news tweets published on the same day that described this event

Key Observations(2)

- News and tweets that describe similar content often have similar predicate argument structures
  - *Chile Earthquake Shortened Earth Day*
  - *Chile Earthquake Shortened Day*
  - *oh yea and Chile earthquake the earth off it's axis according to NASA and shorten the day by a wee second :-(*

Label News Tweets with News



Label News Tweets with News

Label News Tweets with News

Tweet:

oh yea and Chile earthquake the earth off it's axis according to NASA and shorten the day by a wee second :-(

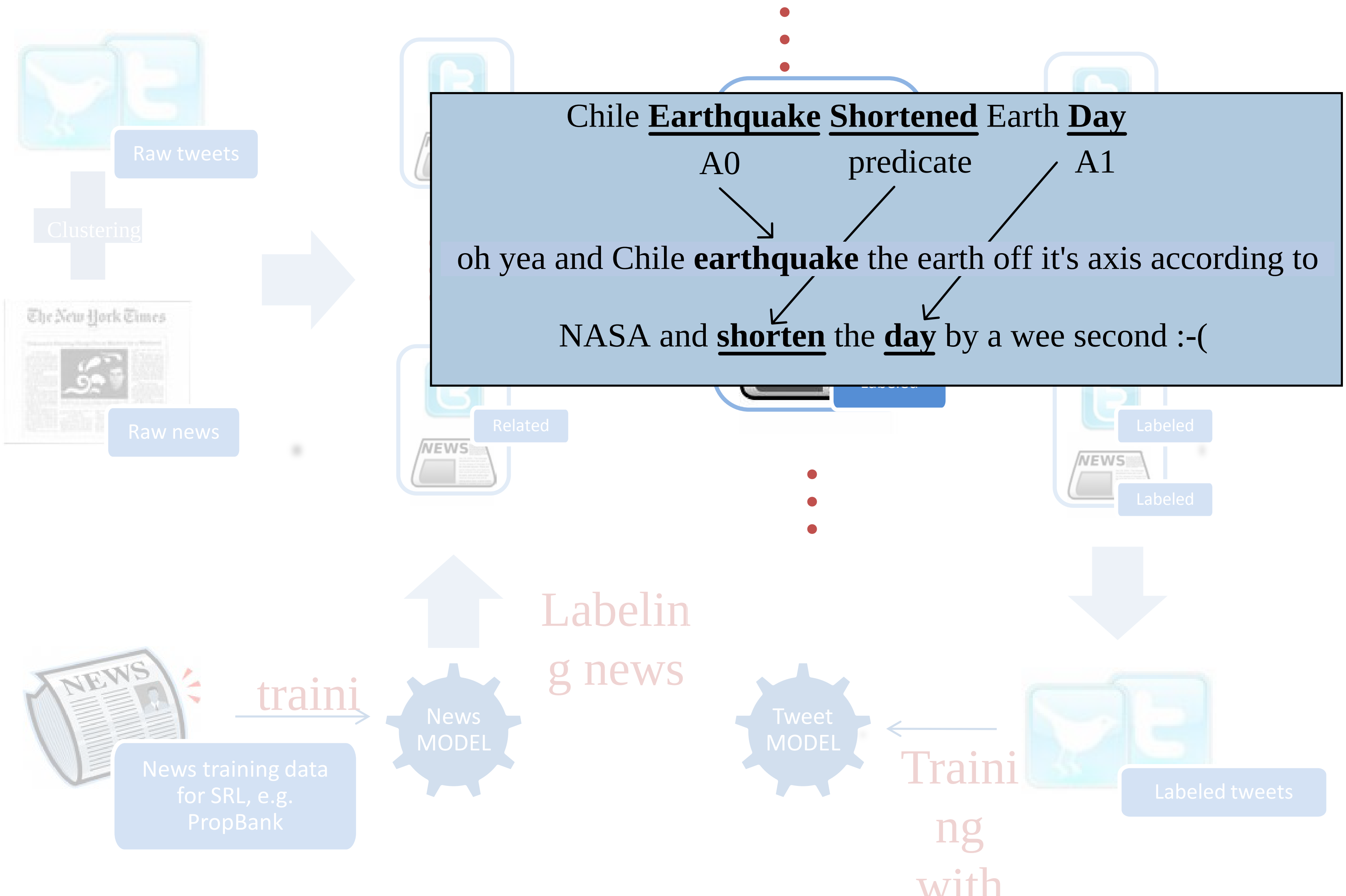
News:

Chile Earthquake Shortened Earth Day

Chile Earthquake Shortened Day

…

Label News Tweets with News



Label News Tweets with News

Conflict resolution

Conflicts are cases that violate any of the two structure constraints(Meza-Ruiz and Riedel, 2009)

1. one (predicate, argument) pair has only one role label in one sentence;

E.g., *(shorten, earthquake, A0) vs. (shorten, earthquake, A1)*

2. one predicate can have each of the proper arguments (A0~A5) once at most in one sentence.

E.g., *(shorten, earthquake, A0) vs. (shorten, axis, A0),*

News training data for SRL, e.g. PropBank

MODEL

MODEL

Traini ng with

Labeled tweets

Experiment Setting

- Evaluation metric: precision, recall and F1
- Baseline: SRL system trained on news (Meza-Ruiz and Riedel, 2009)
- Data preparation
  - Training dataset: 10,000 mechanically labeled tweets
  - Testing dataset: 1,110 human labeled tweets

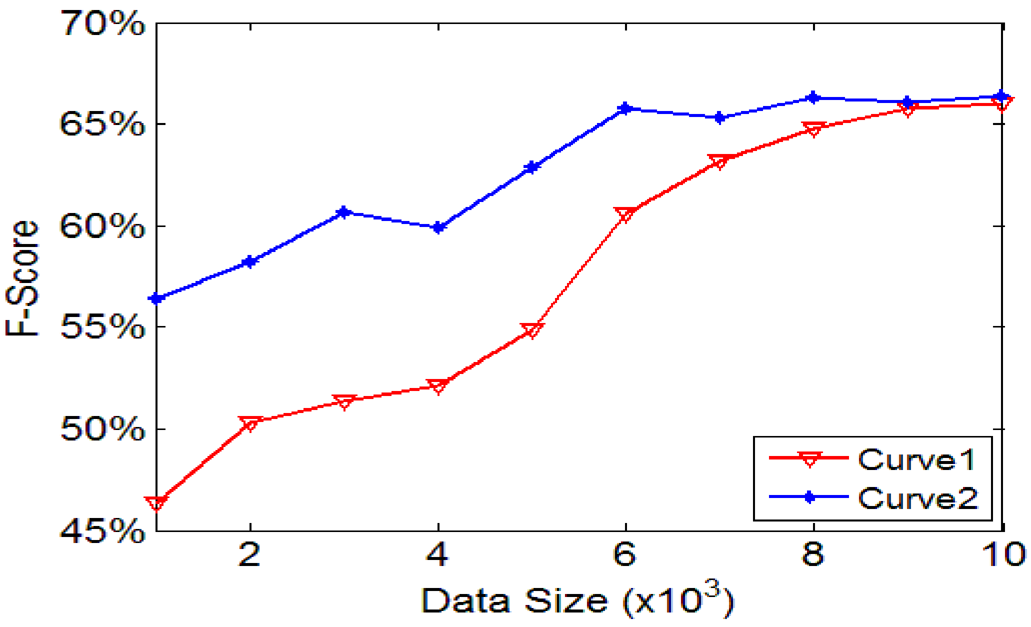
Experimental Results (1)

- Basic results
  - SRL-TS: our system; SRL-BS: baseline

	Precision	Recall	F1
SRL-BS	36.0 %	54.5%	43.3%
SRL-TS	78.0%	57.1%	66.0%

Experimental Results (2)

- Influence of training data size
  - Curve1: no test data is used for training
  - Curve2: half of the test data is used as training data



Collective SRL for Tweets with Clustering

Observation

- Great redundancy in tweets, i.e., the same predicate-argument structures occur in multiple tweets.
- Example:
  - oh yea and Chile earthquake the earth off it's axis according to NASA and shorten the day by a second :-
  - Chile Earthquake Shortened Earth Day
  - item Chile Earthquake Shortened Day

Collective SRL with Clustering

- Two round labeling to leverage the redundancy in tweets
  - #1 round: labeling on single tweet level using a conventional CRF based labeler
  - #2 round: labeling on cluster level using a CRF based labeler enhanced by cluster-level features

- Cluster-level features
  - The top 3 most frequent roles played by the word in the cluster

Algorithm 1 Collective SRL for tweets with Clustering.

Require: Tweet stream i;sequential labelers l_1, l_2;output stream o.

Initialize clusters cl: $cl = \varnothing$.
while Pop a tweet t from i and $t \neq null$ **do**
Put t to a cluster c: $(c, cl) = cluster(cl, t)$.
if $|c| > N$ **then**
Initialize cache of labeled results cs:$cs = \varnothing$.
for $\forall t^{'} \in c$ **do**
Label $t^{'}$ with l_1:$(t^{'}, \{(p, s, cf)\}) = label(l_1, t^{'})$.
for $\forall cf \in \{(p, s, cf)\} > \alpha$ **do**
Cache labeled results:$cs = cs \cup \{(t^{'}, p, s, cf)\}$.
end for
end for
for $\forall t^{'} \in c$ **do**
Label $t^{'}$ with l_2:$(t^{'}, \{(p, s, cf)\}) = label(cs, l_2, t^{'})$.
Output labeled results $(t^{'}, \{(p, s, cf)\})$ to o.
end for
Remove c from cl:$cl = cl - \{c\}$.
end if
end while
for $\forall c \in cl, \forall t^{'} \in c$ **do**
Label $t^{'}$ with l_1:$(t^{'}, \{(p, s, cf)\}) = label(l_1, t^{'})$.
Output labeled results $(t^{'}, \{(p, s, cf)\})$ to o.
end for
return o.

Implementation Details

- The CRF labelers
  - Training : CRF++
  - Decoder: Viterbi algorithm
  - Feature extraction: OpenNLP and the Stanford parser
- Clustering
  - Bottom-up online clustering based on merge
  - Features: bag-of-words
  - Similarity computing: cosine similarity, i.e.,

$$sim(\vec{t_1}, \vec{t_2}) = \frac{\vec{t_1} \cdot \vec{t_2}}{|\vec{t_1}||\vec{t_2}|}$$

Algorithm 2 Clustering a tweet.

Require: Clusters cl;tweet for clustering t.

Get the reference of the most near cluster c^* for t:$c^* = argmax_{c' \in cl} sim(t, c')$.
Get the similarity s between t and c^*: $s = sim(t, c^*)$.
if $s < \beta$ **then**
    Create a new cluster for t: $c^* = \{t\}$.
    Add c^* to cl:$cl = cl \cup \{c^*\}$.
    if $|cl| > M$ **then**
        Merge clusters:$cl = merge(cl)$.
    end if
end if
return c^* and cl.

Experiment Setting

- Dataset: 6,670 manually annotated dataset, randomly divided into three parts
  - 1,000 for development
  - 2,394 for training
  - The remaining for testing
- Baseline: a conventional CRF based labeler
  - BIO labeling schema: …<B-A0>earthquake<O> shorten<B-A1>day…
  - Features: lemma/POS tag of the current/previous/next token, the lemma of the predicate and its combination with the lemma/POS tag of the current token, the voice of the predicate (active/passive), the distance between the current token and the predicate, and the relative position of the current token to the predicate
- Evaluation metrics: Precision, Recall and F1

Experiment Results

- An absolute gain of 3.1% in terms of average F1 measure

System	Pre.(%)	Rec.(%)	F1(%)
SRL_{CL}	61.9	56.7	59.2
SRL_{BA}	62.7	50.8	56.1

- Main error sources:
  - Irregular words in tweets: "…thank youuuu sweedie pops…"
  - Unknown words in tweets: "Bacteria in the gut shown to lower obesity…"

Enhancing SRL for Tweets with Self-training

Motivation

- Use abundant unlabeled data to overcome the lack of annotated tweets
  - Repeatedly re-train the model using the tweets labeled by itself
  - Consider two factors while selecting: correctness of labeling and informativeness

Algorithm 1 Self-training based SRL for tweets.

Require: Tweet stream i;training tweets ts;output stream o.

Initialize two CRF based labelers l and $l^{'}$: $(l, l^{'}) = train(cl)$.
Initialize the number of new accumulated tweets for training n: $n = 0$.
while Pop a tweet t from i and $t \neq null$ **do**
    Label t with l:$(t, \{(p, s, cf)\}) = label(l, c, t)$.
    Label t with $l^{'}$:$(t, \{(p, s, cf)\}^{'}) = label(l^{'}, c, t)$.
    Output labeled results $(t, \{(p, s, cf)\})$ to o.
    if $select(t, \{(p, s, cf)\}, \{(p, s, cf)\}^{'})$ **then**
        Add t to training set ts:$ts = ts \cup \{t, \{(p, s, cf)\}\}$;$n = n + 1$.
    end if
    if $n > N$ **then**
        Retrain labelers:$(l, l^{'}) = train(cl)$;$n = 0$.
    end if
    if $|ts| > M$ **then**
        shrink the training set:$ts = shrink(ts)$.
    end if
end while

Features

- Lemma/POS tag of the current/previous/next token
- Lemma of the predicate and its combination with the lemma/POS tag of the current token
- The voice of the predicate (active/passive)
- The distance between the current token and the predicate
- The relative position of the current token to the predicate
- Dependencies parsing related features

Selection Criteria

- Select such tweets that the performance can be most improved if they are selected for training
  - Correctness : These tweets are correctly labeled
    - Train two independent models and consider a tweet is labeled correctly if the two models give the same results confidently
  - Informativeness: These tweets provide new information to the existing training set
    - If a labeled tweet is not much similar to any tweet in the training set

Algorithm 2 Selection of a training tweet.

Require: Training tweets ts; tweet t; labeled results by l $\{(p,s,cf)\}$; labeled results by l' $\{(p,s,cf)\}'$.

if $\{(p,s,cf) \neq \{(p,s,cf)\}'$ **then**
    return FALSE.
end if
if $\exists cf \in \{(p,s,cf)\} \cup \{(p,s,cf)\}' < \alpha$ **then**
    return FALSE.
end if
if $\exists t' \in ts\ sim(t,t') > \beta$ **then**
    return FALSE.
end if
return TRUE.

Experiment Setting

- Dataset: 7,171 manually annotated dataset, randomly divided into three parts
  - 583 as seeds
  - 5,421 for self-training development
  - The remaining for blind testing
- Baseline: a conventional CRF based labeler
- Evaluation metrics: Precision, Recall and F1

Experiment Results

- A gain of 3.4% F1

Table 2: Basic experimental results.

System	Pre.(%)	Rec.(%)	F1(%)
SRL_{SE}	59.2	45.9	51.7
SRL_{BA}	46.7	50.0	48.3

- Effects of correctness

System	Pre.(%)	Rec.(%)	F1(%)
SRL_{SE}	59.2	45.9	51.7
SRL_{SE-C}	48.4	36.5	41.6

- Effects of informativeness

System	#T	F1(%)
SRL_{SE}	2,557	51.7
SRL_{RD}	2,557	47.7
SRL_{SP}	4,000	42.5
SRL_{CF}	4,277	44.9

References

- Màrquez, Lluís. 2009. *Semantic Role Labeling Past, Present and Future*, Tutorial of ACL-IJCNLP 2009.
- Meza-Ruiz, Ivan and Sebastian Riedel. 2009. Jointly Identifying Predicates, Arguments and Senses using Markov Logic. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, pages: 155-163.
- Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Daniel Tse and Zhongyang Xiong. 2010. Collective Semantic Role Labeling on Open News Corpus by Leveraging Redundancy. COLING 2010
- Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Zhongyang Xiong and Changning Huang. 2010. Semantic Role Labeling for News Tweets. COLING 2010
- Xiaohua Liu, Kuan Li, Ming Zhou and Zhongyang Xiong. 2011. Collective Semantic Role Labeling for Tweets with Clustering. IJCAI 2011
- Xiaohua Liu, Kuan Li, Ming Zhou and Zhongyang Xiong. 2011. Enhancing Semantic Role Labeling for Tweets with Self-training. AAAI 2011

Agenda

- QuickView: A Research Platform of SNS Text Mining and Search
- SNS Text Mining
  - Semantic Role Labeling
  - Sentiment Analysis
- Future Work

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Sentiment Analysis (SA)

- Also known as **opinion mining**: to understand the attitude of a speaker or a writer with respect to some topic
  - The attitude may be their judgment or evaluation, their affective state or the intended emotional communication
  - Most popular classification of sentiment: positive or negative
- For example
  - *<u>The pictures</u> are very clear.*
  - *In his recent State of the Union address, **US President Bush** quite unexpectedly labeled <u>Iran, Iraq, and the DPRK</u> as an "axis of evil".*

Applications of SA

- Business intelligence system
- Purchase planning
- Public opinion management
- Web advertising

Sentiment Components

- **Holder**
  - who expresses the sentiment
- <u>Target</u>
  - what the sentiment is expressed to
- Polarity
  - the nature of the sentiment (e.g., positive/negative)
- *In his recent State of the Union address, **US President Bush** quite unexpectedly labeled <u>Iran, Iraq, and the DPRK</u> as an "axis of evil".*

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Holder Detection

- Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns
  - (Choi et al., HLT/EMNLP-05)

International officers believe that the EU will prevail.
International officers said US officials want the EU to prevail.

- View *source identification as an* information extraction task and tackle the problem using sequence tagging and pattern matching techniques simultaneously
  - *Linear-chain CRF model* to identify opinion sources
  - Patterns incorporated as features

CRF for Holder Detection

- Given a sentence X, to seek for a label sequence Y that maximizes

$$P(y|x) = \frac{1}{Z_x} \exp\left(\sum_{i,k} \lambda_k \, f_k(y_{i-1}, y_i, x) + \sum_{i,k} \lambda'_k \, f'_k(y_i, x) \right)$$

  - Y_i belongs to {'S', 'T', '-'}
  - λ_k and λ'_k are parameters, f_k and f'_k are feature functions
  - Z_x is the normalization factor

International	officers	believe	that	the	EU	will	prevail
S	T	-	-	-	-	-	-

Basic Features

- **Capitalization features:** all-capital, initial-capital
- **Part-of-speech features ([-2,+2]):** noun, verb, adverb, wh-word, determiner, punctuation, etc
- **Opinion lexicon features: [-1,+1]** whether or not the word is in the *opinion lexicon*
- **Dependency tree features**
  - the grammatical role of its chunk
  - the grammatical role of xi-1's chunk
  - whether the parent chunk includes an opinion word
  - whether xi's chunk is in an argument position with respect to the parent chunk
  - whether xi represents a constituent boundary
- **Semantic class features:** the semantic class of each word: authority, government, human, media, organization or company, proper name, and other

Extraction Pattern Learning

- Looking at the context surrounding each answer and proposes a lexico-syntactic pattern
  - *[They]$_h$ complained about the deficiencies of the benefits given to them.*
  - *<subj> complained*

- Compute the probability that the pattern will extract an opinion source

$$P(\text{source} \mid \text{pattern}_i) = \frac{\text{correct sources}}{\text{correct sources} + \text{incorrect sources}}$$

Extraction Pattern Features

- Four IE pattern-based features for each token xi
  - SourcePatt-Freq, SourcePatt-Prob,
  - SourceExtr-Freq, SourceExtr-Prob

  - Where
    - SourcePatt indicates whether a word activates any source extraction pattern. E.g., "complained" activates the pattern "<subj> complained"
    - SourceExtr indicates whether a word is extracted by any source pattern. E.g., "They" would be extracted by the *"<subj> complained"*

Experimental Results

- MPQA data
  - In total, 535 documents where targets are annotated by human
  - 135 as development set and feature engineering, and the remaining 400 for evaluation, performing 10-fold cross validation
- 3 measures: overlap match (OL), head match (HM), and exact match (EM)

		Recall	Prec	F1
Extraction Patterns	OL	48.5	81.3	60.8
	HM	46.9	78.5	58.7
	EM	41.9	70.2	52.5
CRF: basic features	OL	56.1	81.0	66.3
	HM	55.1	79.2	65.0
	EM	50.0	72.4	59.2
CRF: basic + IE pattern features	OL	59.1	82.4	68.9
	HM	58.1	80.5	67.5
	EM	52.5	73.3	61.2

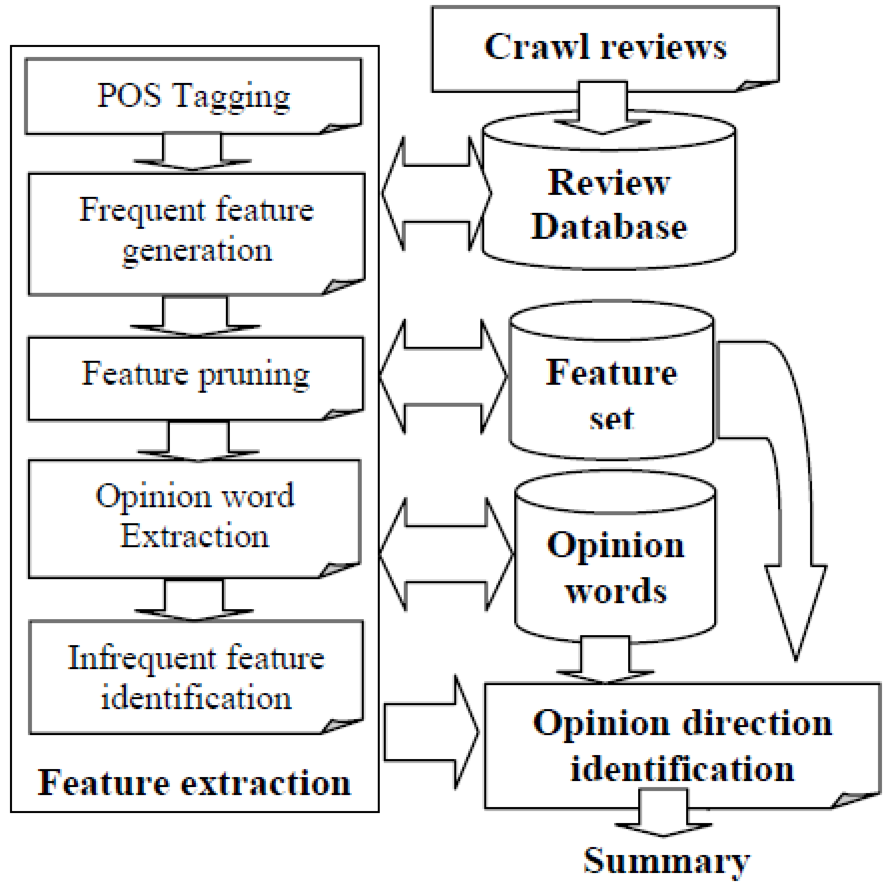
Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Target Detection

- Mining Opinion Features in Customer Reviews
  - (Minqing Hu and Bing Liu, AAAI 2004)
  - *Explicit feature*
    - *The pictures are very clear.*
  - *Implicit feature*
    - *While light, it will not easily fit in pockets. (size)*
- Task definition
  - Given a product name and all the reviews of the product, to find the features of the product that appear explicitly as nouns or noun phrases in the reviews

Approach Overview



Frequent Features Detection

- Association rule mining
  - Find frequent features with three words or fewer
  - Appears in more than 1% of the review sentences (minimum support)

- Feature Pruning
  - Compactness: compact in at least 2 *sentences*
  - *p-support (pure support):* a p-support lower than the minimum p-support (3)

Infrequent Feature Detection

- People use the same adjective words to describe different subjects
  - *"Red eye is very easy to correct."*
  - *"The camera comes with an excellent easy to install software"*
  - *"The pictures are absolutely amazing"*
  - *"The software that comes with it is amazing"*

Infrequent Feature Detection

- Opinion word identification
  - For each sentence in the review database, if it contains any frequent feature, extract the nearby *adjective as* opinion word
- Infrequent feature detection
  - For each sentence in the review database, if it contains no frequent feature but one or more opinion words, find the *nearest noun/noun phrase of the opinion word* as an infrequent feature

Experimental Results

- Data: customer reviews of five electronics products from Amazon.com and C|net.com

Product name	No. of manual Features	Frequent features (association mining)		Compactness pruning		P-support pruning		Infrequent feature identification	
		Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Digital camera1	79	0.671	0.552	0.658	0.634	0.658	0.825	0.822	0.747
Digital camera2	96	0.594	0.594	0.594	0.679	0.594	0.781	0.792	0.710
Cellular phone	67	0.731	0.563	0.716	0.676	0.716	0.828	0.761	0.718
Mp3 player	57	0.652	0.573	0.652	0.683	0.652	0.754	0.818	0.692
DVD player	49	0.754	0.531	0.754	0.634	0.754	0.765	0.797	0.743
Average	**69**	**0.68**	**0.56**	**0.67**	**0.66**	**0.67**	**0.79**	**0.80**	**0.72**

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Lexicon Based Polarity Classification

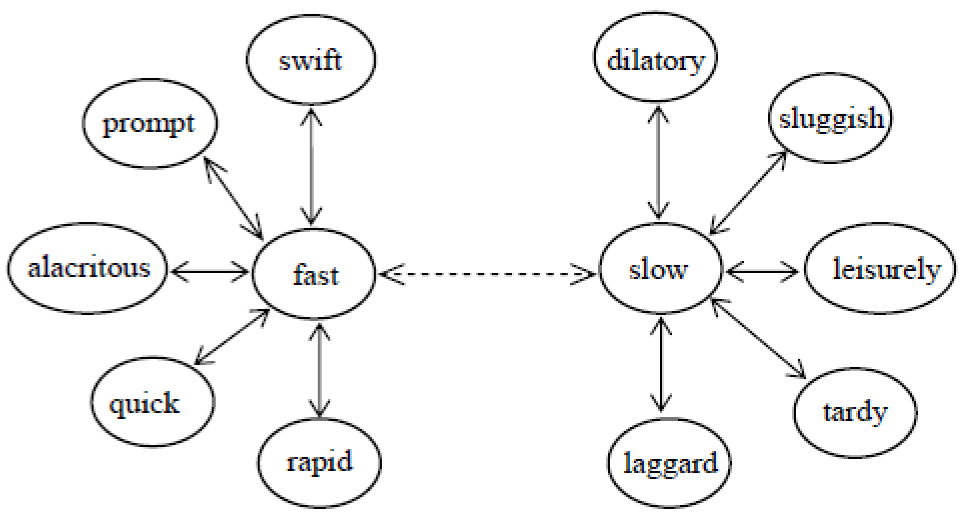
- Mining and Summarizing Customer Reviews
  - (Hu and Liu, KDD-2004)

- Basic idea
  - Use the dominant orientation of the opinion words in the sentence to determine the orientation of the sentence.
  - That is, if positive/negative opinion prevails, the opinion sentence is regarded as a positive/negative one.

Lexicon Building

(Hu and Liu, KDD-2004)

- Utilize the adjective synonym set and antonym set in WordNet to predict the semantic orientations of adjectives
  - Adjectives share the same orientation as their synonyms and opposite orientations as their antonyms.
- Start with several seeds, iteratively expand to cover most opinion words



Hatzivassiloglou and McKeown (1997)

- Predicting the Semantic Orientation of Adjectives
  - (Hatzivassiloglou and McKeown, ACL-97)

- Assumption: adjectives connected by “and”/”but” tend to have same/opposite polarities

The tax proposal was

1. simple and well-received
2. simplistic but well-received
3. *simplistic and well-received

by the public.

ML-based Approaches for Polarity Classification

- Thumbs up? Sentiment Classification using Machine Learning Techniques
  - (Pang et al., 2002)

- Basic idea
  - Treat sentiment classification simply as a special case of topic-based categorization
    - With the two “topics” being positive sentiment and negative sentiment
    - Use three standard algorithms: Naive Bayes classification, maximum entropy classification, and support vector machines

Approach Details

- Document representation
  - *Each* document *d is represented by a feature vector* $\vec{d} := (n_1(d), n_2(d), \ldots, n_m(d))$
  - $n_i(d)$ *could indicate presence, term frequency*
- Classification algorithms
  - Naive Bayes, Maximum Entropy, SVM

Data

- Movie reviews
  - From Internet Movie Database (IMDb)
    - http://www.cs.cornell.edu/people/pabo/movie-review-data/
    - http://reviews.imdb.com/Reviews/
  - 700 positive / 700negative

- Experiment setting for ML classifiers
  - 3-fold cross validation
  - Treating punctuation as separate lexical items
  - No stemming or stoplists were used

Experimental Results

- Baseline: use a few words written by human to classify

	Proposed word lists	Accuracy
Human 1	positive: *dazzling, brilliant, phenomenal, excellent, fantastic* negative: *suck, terrible, awful, unwatchable, hideous*	58%
Human 2	positive: *gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting* negative: *bad, cliched, sucks, boring, stupid, slow*	64%

- ML-based methods

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	”	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

Other Related Approaches

- Topic sentiment mixture
  - Mei et al., 2007
- Semi-supervised approach
  - Li et al., 2010
- Domain Adaptation
  - Blizter et al., 2007

Summary

1. Sentiment analysis refers to a set of subtasks
   - Holder, target, polarity

2. Sentiment analysis is a challenging task and more difficult than traditional topic-based classification
   - Understanding of the semantics is often needed
     - How could anyone sit through this movie?

   - Same word/phrase may have different polarities in different domains
     - An unpredictable movie (positive)
     - An unpredictable politician (negative)

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Twitter Sentiment Analysis

- Aiming to find positive and negative tweets about a given topic
  - Focusing on polarity classification
- Target-dependent sentiment classification
  - Given a target, classifying a tweet as positive, negative or neutral (no sentiment) towards the target
  - Input: a tweet "*Windows 7 is much better than Vista!*" and a target "*Windows 7*"
  - Output: positive

Advantages of Twitter SA

- Large amount
- Wide coverage of domain
- Fresh
- From grass roots

Special Challenges

- Short and ambiguous

- Informal and unedited texts
  - *"<u>another part of me</u> by Micheal Jackson is soo nicee! Loooveeeeee itttttttttt!"*

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Existing Twitter SA Systems

- Lexicon-based method
  - Twittratr
- Rule-based
  - Tweetfeel
- Machine learning based
  - Twitter sentiment
- Unknown
  - Twendz
  - Tweetsentiments

Twitrratr

- 

Microsoft | twitrratr - Windows Internet Explorer
http://twitrratr.com/search/Microsoft
Bing
Favorites
Microsoft | twitrratr
Page
Safety
Tools
twitrratr
TRACKING OPINIONS ON TWITTER
SEARCH
SEARCHED TERM
Microsoft
POSITIVE TWEETS
619
NEUTRAL TWEETS
9470
NEGATIVE TWEETS
317
TOTAL TWEETS
10406
5.95% POSITIVE
91.01% NEUTRAL
3.05% NEGATIVE
at a columbia engineering job fair. ny state dot is here. the city of portland. nyc dot. real networks. microsoft. great free stuff (view)
i love how unhelpful microsoft help articles are (view)
new microsoft blog: the windows 7 blog for developers http://is.gd/4isg i think this blog could also interesting for admins. (view)
microsoft gemini is what you always wanted excel to be. its an excellent step forward to bi for the masses. (view)
microsoft gemini is what you always wanted excel to be. its an excellent step forward to bi for the masses. (view)
@zik actually, i know people @ msnbc.com and already knew that... still, i don't think microsoft is happy with nbc making them look dumb. (view)
@zik actually, i know people @ msnbc.com and already knew that... still, i don't think microsoft is happy with nbc making them look dumb. (view)
Reading Harvard Business columnist Stew Friedman that says Microsoft could learn from Neil Young. Hhmm.. OK. (view)
Just ordered Microsoft Office 2007 Professional for $66 and change after shipping. Much better than the $390 that Amazon wants. (view)
@bryansimpson I have been hearing that a lot over the past month. Maybe them and Microsoft need to get together. (view)
Microsoft Ad Business Strong, But Display Ads Threatened: Among Microsoft's diverse revenue streams, its d.. http://bit.ly/2owGq0 (view)
Got myself a Microsoft Natural 4000 keyboard amd wireless optical mouse. Need a USB hub methinks (view)
[PCWorld] Microsoft Taps Telefonica to Deliver Live Messenger VOIP http://tinyurl.com/5g57pj (view)
Found this: meta name="GENERATOR" content="Microsoft FrontPage 4.0". I think I just threw up in my
why microsoft develop ie ? it sucks grrrrrr (view)
why microsoft develop ie ? it sucks grrrrrr (view)
why microsoft develop ie ? it sucks grrrrrr (view)
why microsoft develop ie ? it sucks grrrrrr (view)
why microsoft develop ie ? it sucks grrrrrr (view)
f*ck you internet explorer!! microsoft, you should be utterly ashamed of yourselves. (view)
f*ck you internet explorer!! microsoft, you should be utterly ashamed of yourselves. (view)
f*ck you internet explorer!! microsoft, you should be utterly ashamed of yourselves. (view)
f*ck you internet explorer!! microsoft, you should be utterly ashamed of yourselves. (view)
f*ck you internet explorer!!
Internet | Protected Mode: On
100%

Twitrratr

- http://twitrratr.com

- Feature
  - 3 classes (positive, negative, neutral)
  - Highlight the sentiment expressions

- Method
  - Lexicon-based
    - Words, phrases, emoticons (☺, :D, :-(…)
  - Manually-made lexicon
    - Still contains errors (e.g., *fail* in the positive list)
  - Simple string (not word) match (“*unhelpful*”)

Tweetfeel

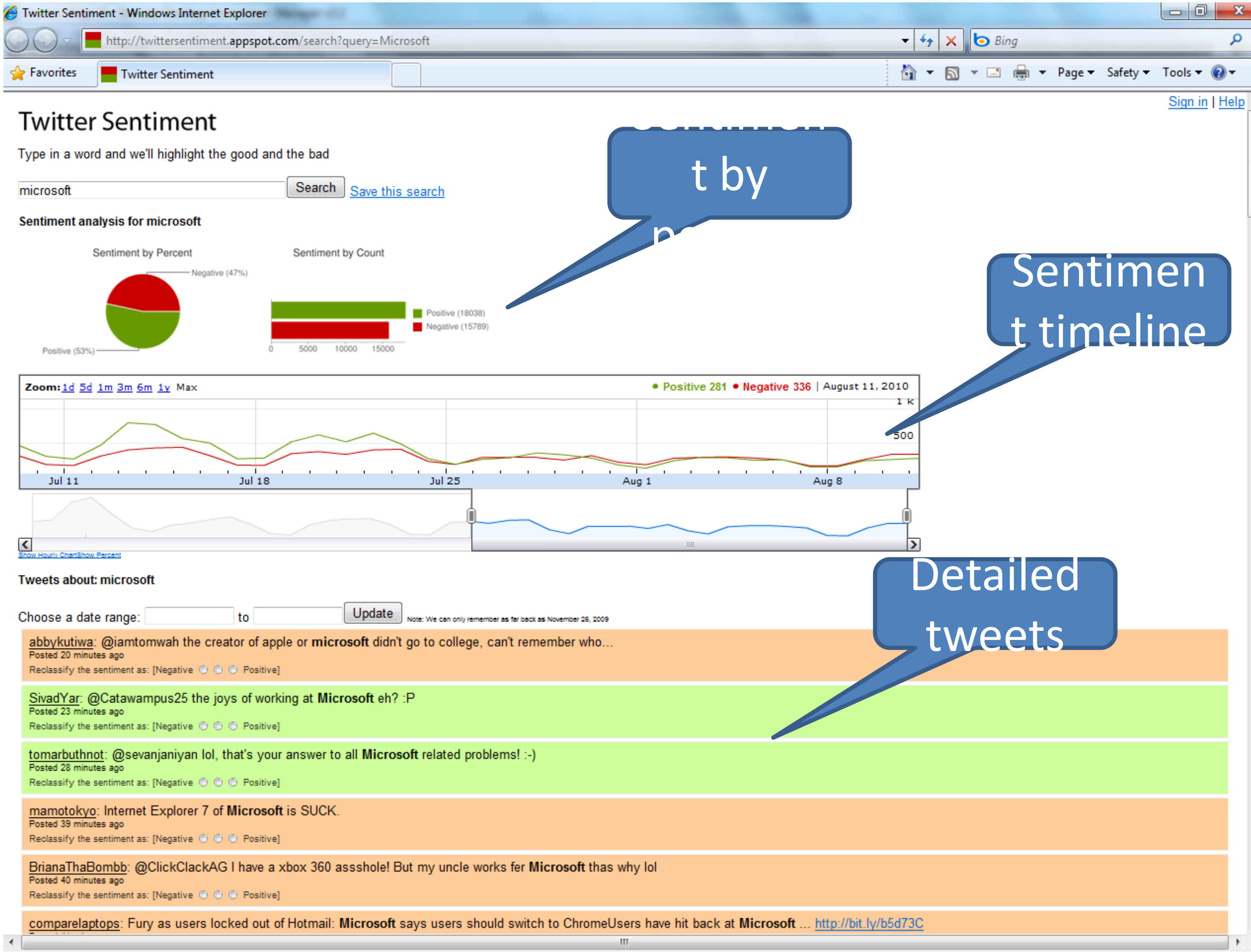
- 

What do tweeple think about microsoft? Check out Tweetfeel twitter sentiment here - Windows Internet Explorer
http://www.tweetfeel.com/#microsoft
Tweetfeel Biz | FAQ | Contact Us | Biz Login
tweetfeel
A New Level of Tweetfeel
Biz
microsoft
Search
Try some Twitter trends: Bondan Prakoso Scott Pilgrim Ramadan Inception Meteor Shower iOS Poulsen
60
89
= 60%
LOL RT @gmilh: RT @BillGayts I love microsoft prodott. I li prend and me li ficc in my ass.
@leighpenny1 Is it OK that I love microsoft Excel and would marry it if I could figure out have to have sex with it?
I love microsoft Office 2010....Yeah i kno im a nerd at heart!
RearType? Expliquenme el prototipo de Table que desarrolla microsoft #fail TOTAL
RT @BillGayts I love microsoft prodott. I li prend and me li ficc in my ass.
@dancerlindsey nope PS3..microsoft sucks
Read our FAQ Subscribe to our API Legal Stuff 100% Guarantee Share
Follow us
Email us
Brought to you by conversition
Powered by twitter
Done, but with errors on page.
Internet | Protected Mode: On
100%

Tweetfeel

- http://www.tweetfeel.com
- Feature
  - 2 classes: positive and negative
- Method
  - Probably rule based
    - Positive patterns
      - pos_verb [Query], [Query] pos_verb, [Query] is pos_adj
    - Negative patterns
      - neg_verb [Query], [Query] neg_verb, [Query] is neg_adj
  - High precision, low recall

Twitter Sentiment

- 

Twitter Sentiment

- http://twittersentiment.appspot.com
  - Created by some graduate students at Stanford University
- Features
  - 2 classes: positive and negative
  - Timeline: how the number of pos/neg sentiments change over time
  - Allows users to correct wrongly classified tweets
- Method
  - Machine learning-based (maximum entropy classifier)
  - Unsupervised training data construction by making use of emoticons (☺ for positive, ☹ for negative)

Summary

- Twitter SA has its own characteristics
  - Short, informal text
  - Pictograms (<3) and emoticons (☺, ☹, :D,…)
- However, not intensively studied yet
  - Traditional SA methods are employed
  - No paper published in top conferences yet
  - Lacking of large amount of publicly available annotated data for system evaluation and comparison
- Potential directions
  - Tweet normalization
  - Context aware sentiment analysis

References

- J. Blitzer, R. McDonald, and F. Pereira. Domain adaption with structural correspondence learning. In EMNLP, 2006.
- Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. 2005. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. Proceedings of HLT/EMNLP-05.
- Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant Supervision. http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf
- Vasileios Hatzivassiloglou and Kathleen McKeown. 1997. Predicting the semantic orientation of adjectives. In Proc. of the 35th ACL/8th EACL, pages 174–181.
- Hu, M., and Liu, B. 2004. Mining Opinion Features in Customer Reviews. To appear in AAAI'04, 2004.
- Hu, M., and Liu, B. 2004a. Mining and summarizing customer reviews. In Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177. ACM Press New York, NY, USA.
- S Li, CR Huang, G Zhou, SYM Lee. 2010. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 414–423, Uppsala, Sweden, 11-16 July 2010.
- Q.Mei, X. Ling,M.Wondra, H. Su, and C.X. Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the 16th International Conference onWorldWideWeb, pages 171–180.
- Bo Pang, Lillian Lee and Shivakumar Vaithyanathan: Thumbs up? Sentiment Classification using Machine. Learning Techniques. In Proc. Conf. on EMNLP 2002.

Outline

- Sentiment analysis
  - Introduction
    - Definition, application, components
  - Approaches for SA subtasks
    - Holder detection
    - Target detection
    - Polarity classification
- Twitter Sentiment Analysis
  - Goals and challenges
  - Existing systems
  - Target-dependent twitter sentiment analysis

Target-dependent Twitter Sentiment Analysis

- Twitter: an important source for mining people's sentiments
  - Many existing systems for Twitter sentiment analysis
    - Tweetfeel, Twendz, and Twitter Sentiment
  - Typical scenario: the user inputs a sentiment target as a query, and searches for tweets containing positive or negative sentiments towards the target

- Target-dependent Sentiment Classification of Tweets
  - Given a query, classify the sentiments of the tweets as positive, negative or neutral according to whether they contain positive, negative or neutral sentiments about that query
  - Here the query serves as the target of the sentiments

Related Work

- Machine learning based (target-indenpdent) twitter sentiment classification
  - Barbosa and Feng, 2010: Two-step approach to classify the sentiments of tweets using SVM classifiers
  - Davidiv et al., 2010 : Classify tweets into multiple sentiment types using hashtags and smileys as labels
  - Go et al., 2009: SVM classifier + collect training data using emoticons

Motivating Examples

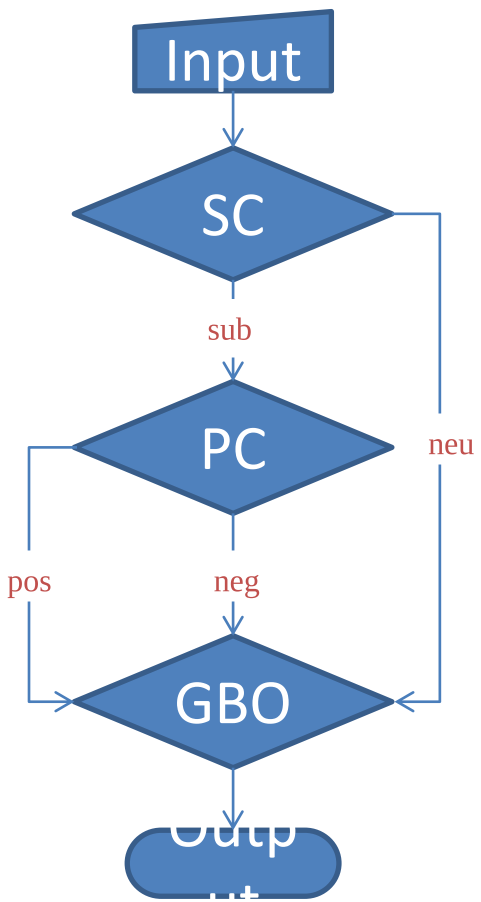
- Most current systems do not consider the target when classifying the sentiment
  - *"Bringing iPhone and **iPad** apps into cars? http://www.speakwithme.com/ will be out soon and alpha is **awesome** in my car."*
    - (positive by *Twitter Sentiment*)
  - *"Here's a **great** article about Monte Veronese cheese. It's in Italian so just put the url into **Google** translate and enjoy http://ow.ly/3oQ77"*
    - (positive by *Twitter Sentiment*)
  - *"No debate needed, heat **can't beat** **lakers** or celtics"*
    - *(negative by Twitter Sentiment)*

Motivations

- Consider the relation between the target and the sentiment word
  - *Windows 7 is much **better** than Vista!*
    - "Windows 7" is connected with "better" by a copula while "Vista" is connected by a preposition "than"
  - *People everywhere **love** Windows & vista. Bill Gates*
    - "love" is not connected to "Bill Gates"
- Consider the context of the tweet
  - *"First game: Lakers!"*
  - *Too short even for human to decide the polarity*

Overview of Our Approach

- Task definition
  - Input
    - a collection of tweets containing the target (or query)
  - Output
    - labels assigned to each of the tweets
- Three steps
  - Subjectivity classification (SC)
  - Polarity classification (PC)
  - Graph-based optimization (GBO)



Preprocessing

- Tweet normalization
  - A simple rule-based model
  - “gooood” to “good”, “luve” to “love”
- POS tagging
  - OpenNLP POS tagger
- Word stemming
  - A word stem mapping table (about 20,000 entries)
- Syntactic parsing
  - A Maximum Spanning Tree dependency parser (McDonald et al., 2005)

Subjectivity and Polarity Classification

- Binary SVM classifiers with linear kernel
  - Target-independent features
    - Content features
      - Words, punctuations, emoticons, and hashtags
    - Sentiment lexicon features
      - The number of positive or negative words in the tweet according to a sentiment lexicon (General Inquirer)
  - Target-dependent features

Target-dependent Features

- *Rules for generating target-dependent features*
  - *Subject/object of* a transitive verb wi
    - wi _arg2, e.g., “I love iPhone”, => “love_arg2”
    - wi_arg1, e.g., “Obama reaffirms ..” => reaffirm_arg1
  - *Subject of* a intransitive verb
    - Wi_it_arg1
  - *Head of* an adjective or noun
    - Wi_arg1
  - Connected by a copula with an adjective or noun
    - Wi_cp_arg1
  - w_i is an adjective or intransitive verb appearing alone as a sentence and *the target* appears in the previous sentence
    - Wi_arg
    - E.g., “*John did that. Great!*” => great_arg
  - w_i is an adverb, and the verb it modifies has the target as its subject
    - Arg1_v_wi
    - *E.g., “iPhone works better with the CellBand” => arg1_v_well*
- Handle negations by adding “*neg-*”
  - “*iPhone does not work better with the CellBand*” => *neg-work_it_arg1, arg1_v_neg-well*

Target Expansion

- Many sentiments are not expressed exactly towards the target
  - *“I am passionate about Microsoft technologies especially Silverlight.”*
  - *Microsoft vs. Microsoft technologies*
- Extended target identification
  - All noun phrases including the target
  - Mentions co-referring to the target
  - The top K nouns and noun phrases which have strong association with the target
  - Head nouns of all extended targets, which have strong association with the target

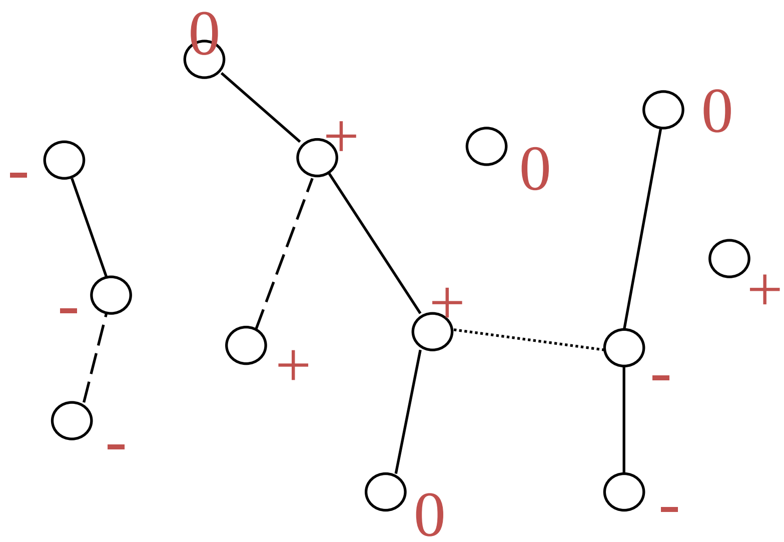
Examples for Target Expansion

- I am passionate about Microsoft technologies especially Silverlight.
- Oh, Jon Stewart. How I love you so …

- Top K nouns and noun phrases with “Lady Gaga”: ladygaga, dressing, songs …
- Microsoft Technologies

Graph-based Sentiment Optimization

- Relation types among the input tweets
  - Retweeting
  - Being published by the same person
  - Replying



Graph-based Sentiment Optimization

$p(c \mid \tau) \longrightarrow$ Relaxation Labeling (Angelova and Weikum, 2006) $\longrightarrow p(c \mid \tau, G)$

The output score by SC and PC to calculate p(c| τ)

Iteratively optimize p(c| τ,G)

$$p(c \mid \tau, G) = p(c \mid \tau) \sum_{N(d)} p(c \mid N(d))\, p(N(d))$$

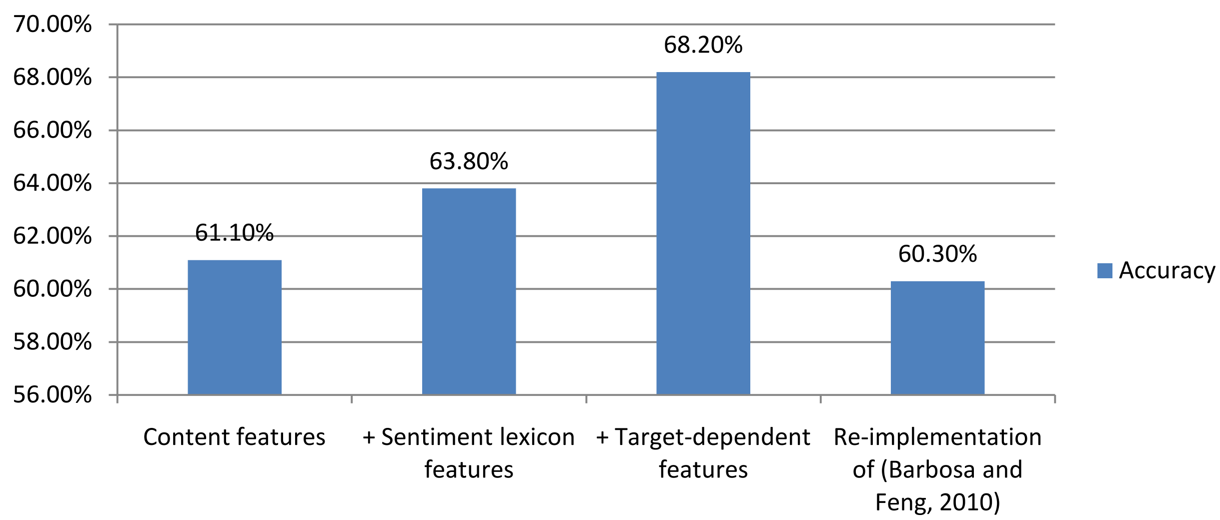
- *c* is the sentiment label of a tweet which belongs to {positive, negative, neutral}
- *G* is the tweet graph
- *N(d)* is a specific assignment of sentiment labels to all immediate neighbors of the tweet
- τ is the content of the tweet

Experimental Setting

- Raw data
  - 5 queries: *Obama, Google, iPad, Lakers, Lady Gaga*
  - 400 English tweets downloaded for each
- Annotation
  - 2 human annotators
  - 3 labels: positive, negative or neutral
  - 459 positive, 268 negative and 1,212 neutral tweets
- Inter-annotator study
  - For 86% of tweets, two annotators give identical labels
  - For 13%, neutral-subjective disagreement
  - For 1%, positive-negative disagreement

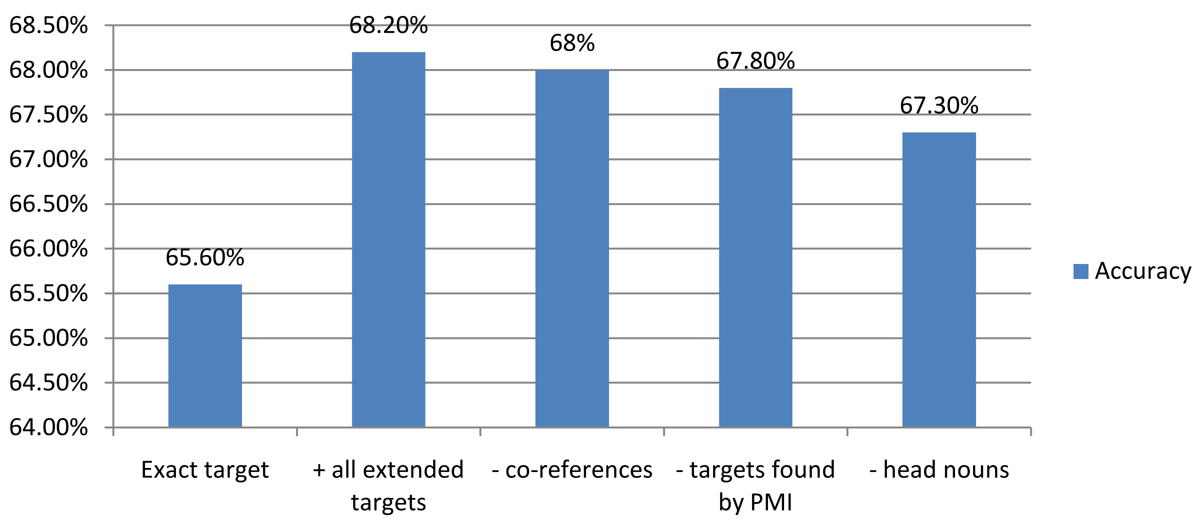
Subjectivity Classification Evaluation

- Data
  - 727 subjective (positive + negative) tweets and 1212 neutral tweets
  - 5 fold cross validation



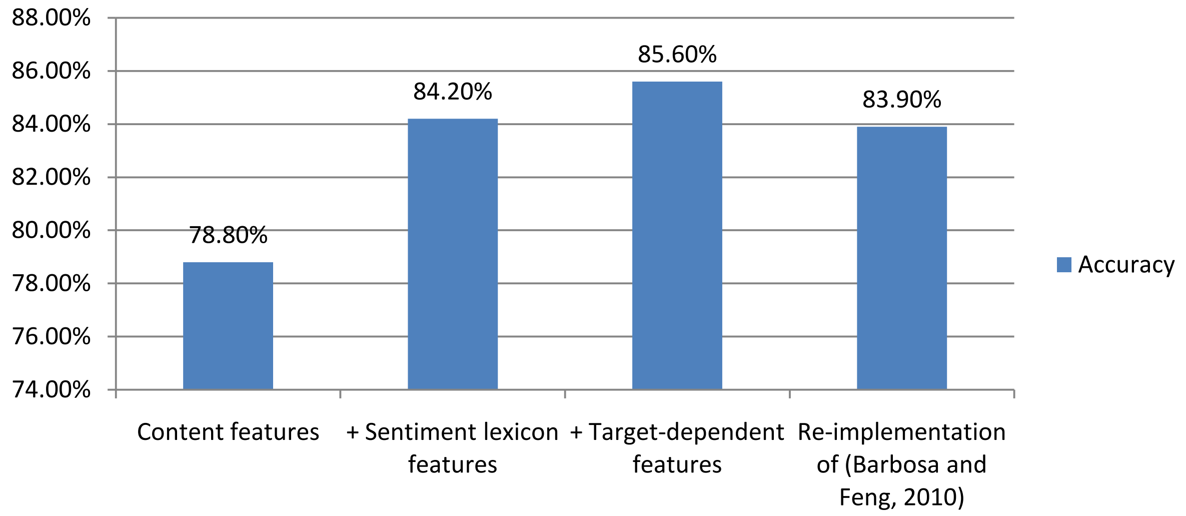
Evaluation of Extended Targets for Subjectivity Classification

- Data
  - 727 subjective (positive + negative) tweets and 1212 neutral tweets
  - 5 fold cross validation



Polarity Classification Evaluation

- Data
  - 268 negative and 459 positive tweets
  - 5 fold cross validation



Evaluation of Graph-based Optimization

- Data
  - 459 positive, 268 negative and 1,212 neutral tweets

System	Accuracy(%)	F1-score (%)		
		pos	neu	neg
Target-dependent sentiment classifier	66.0	57.5	70.1	66.1
+Graph-based optimization	68.3	63.5	71.0	68.5

Summary

- Target-dependent Twitter sentiment classification
  - Target-dependent features can improve the performance, especially for subjectivity classification
  - Incorporating related tweets can further improve the performance
- Future work
  - More types of extended targets
  - Exploring relations between Twitter accounts for classifying the sentiments of the tweets

Agenda

- QuickView: A Research Platform of SNS Text Mining and Search
- SNS Text Mining
  - Semantic Role Labeling
  - Sentiment Analysis
- Future Work

Research Task

- Fundamental NLP tasks
  - Tokenization, POS tagging, parsing, etc.
  - Text normalization
- Named entity extraction
  - Beyond PLO: movie, TV show, music, etc.
  - Normalization
- Hashtag understanding
  - Summarization
  - Relation recognition

Model and Approach

- Graph and graphical models
- Semi-supervised learning
- Online learning and data stream mining

Thanks