

SENTIMENT ANALYSIS: RETHINKING OUR APPROACH

by Dogan Can

A NOTE ON TERMINOLOGY

‘The beginning of wisdom is the definition of terms.’

Socrates

Opinion Mining (Information Retrieval)

=

Sentiment Analysis (NLP)

CHALLENGES

- Coming up with the right set of keywords non-trivial for humans [Pang et al., 2002]
- Sentiment expression usually subtle, difficult to identify without context
- The fine line between facts and opinions: factual statements may convey opinion, e.g. “long battery life”

CHALLENGES

- Non-trivial to recognize opinion holders [Kim & Hovy, 2005]
- Sentiment is context/topic sensitive, e.g., “Go read the book.” (movies vs. books)
- Discourse structure crucial, order trumps frequency, e.g.,

This film should be *brilliant*. It sounds like a *great* plot, the actors are *first grade*. However, it can't hold up.

SENTIMENT POLARITY CLASSIFICATION

- Types:
 - Binary (Thumbs up/down)
 - Multi-class (Positive/Negative/Mixed/Neutral)
 - (Ordinal) Regression (movie ratings)
 - Ranking (reviews)
 - Agreement (pair of reviews)

SENTIMENT POLARITY CLASSIFICATION

- Extract information on why reviewers liked or disliked something, e.g., “pros and cons” [Kim and Hovy, 2006]
- Rating is different from (opinion) *strength*, e.g., you can have a *strong opinion* about something being “mediocre”.
- Neutral is not lack of opinion.
- Neutral may be perceived as negative. [Cabral, 2006]

SUBJECTIVITY DETECTION

- Distinguishing subjective vs objective is harder than polarity classification. [Michalcea et al., 2007]
- Clues of subjectivity: low frequency words, collocations, adjectives, etc. [Wiebe, 2004]
- TREC Blog track and other recent work

JOINT TOPIC-SENTIMENT ANALYSIS

- Usual assumption: one subject/topic per document
- “Topic filtering and subjectivity filtering are complimentary.”
[Riloff, 2005]
- Especially relevant in the case of
 - Comparative statements
 - Statements mentioning various aspects of the same item

VIEWPOINTS & PERSPECTIVES

- Most work on political texts focus on general attitudes that are not necessarily targeted at a particular issue/subject.
- Opinions on a collection of bundled attitudes and beliefs, e.g. , **liberal** vs. **conservative** vs. **libertarian**
- Extract perspective-related info that may not correspond to sentiment polarity

OTHER NON-FACTUAL INFO

- Recognition of affect: anger, disgust, fear, happiness, sadness, surprise.
- Humor/sarcasm recognition and generation [Mihalcea, 2006]
- Genre detection, e.g. editorial vs news
- Source style, e.g. author identification
- Deceptive language

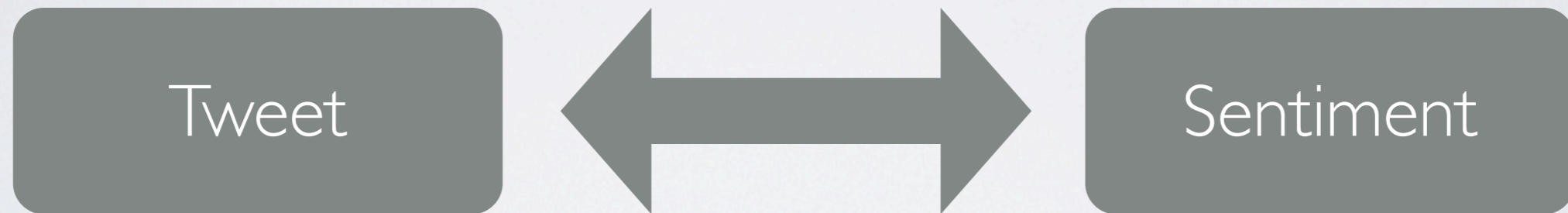
FEATURES

- Presence vs. frequency [Pang et al., 2002]
- Singletons are high-precision subjectivity indicators [Wiebe, 2004]
- Rare terms (OOVs) or novel versions correlate with subjectivity [Yang, 2006]
- Position info [Pang, 2002; Kim & Hovy, 2006]
- N-grams useful? [Pang, 2002 vs. Dave 2003]

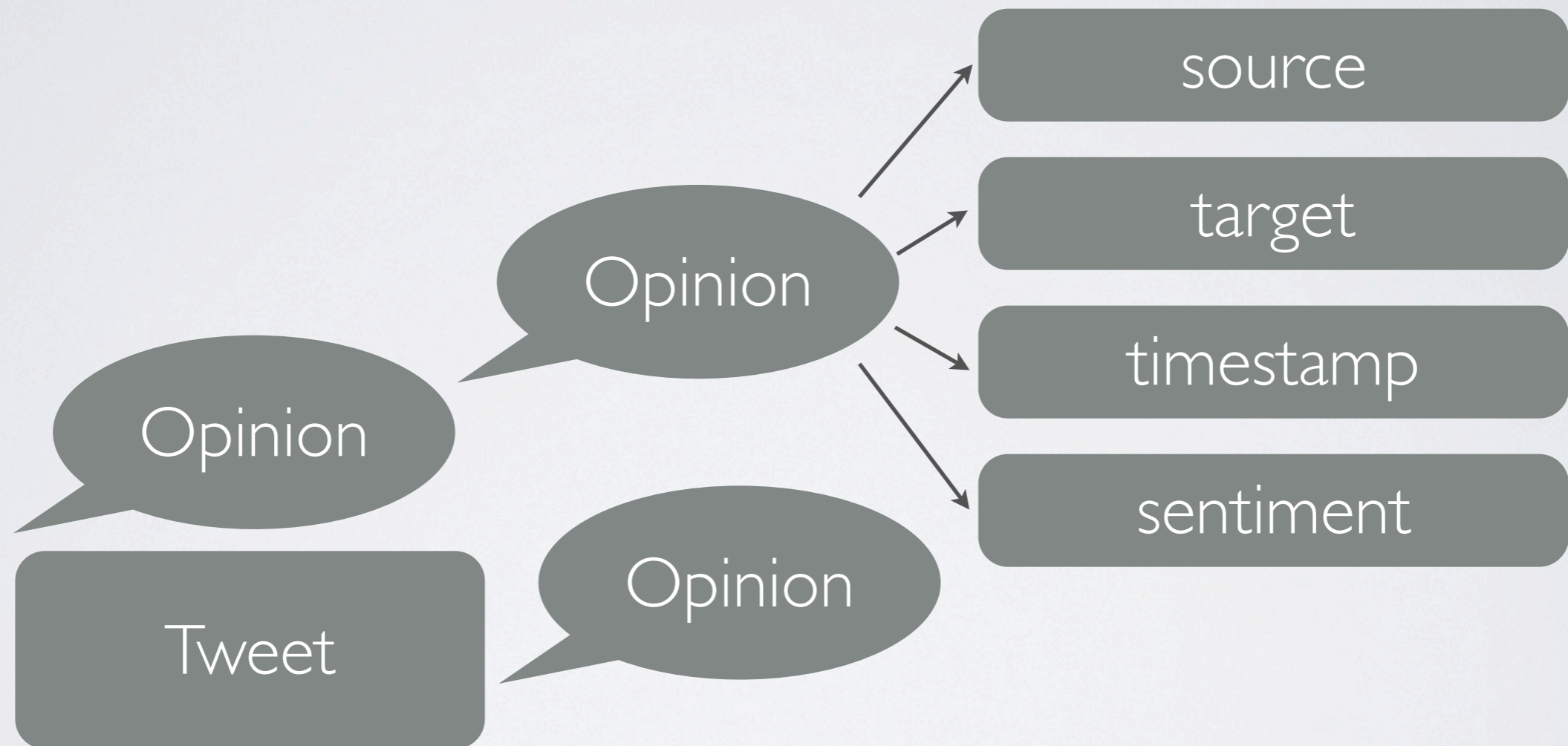
FEATURES

- Contrastive distance between terms
- More complicated methods:
 - Subsumption hierarchy to identify complex lexical features [Riloff, 2006]
 - Markov Blanket Classifier to extract a dependency structure [Airoldi, 2006]
 - Part-of-Speech tags (e.g. to identify adjectives), word sense disambiguation, syntax (e.g. dependency parsing) [various works]
- Negation

OUR CURRENT APPROACH



STRUCTURED APPROACH



SOURCE: OPINION OF WHOM?

- A complex object: collection of **attributes**
- Usually the Tweet author but can also be a group which includes the author or somebody mentioned in the tweet
- Can be thought as a collection of attributes, e.g. a **republican**, a **pundit**, a **celebrity**
- Has connections with other opinion holders, e.g. retweet network, follower/following network

TARGET: OPINION ABOUT WHAT?

- A complex object: collection of **attributes**
- A political candidate or an aspect of a candidate, e.g. Romney, Romney's foreign policy, Obama's health policy
- Sentiment is a like a profile 
e.g. health policy **positive**, foreign policy **negative**, ...

EXAMPLE

 **Nancy Giles** @nancygilesnyc 5 May

The diversity of Obama supporters is beautiful to me; the lack-of diversity in the **#Romney** crowd is stunning.
[#WhoDoesTheGOPRepresent?](#)

Expand

@nancygilesnyc

obama.supporter_diversity

May 5

positive

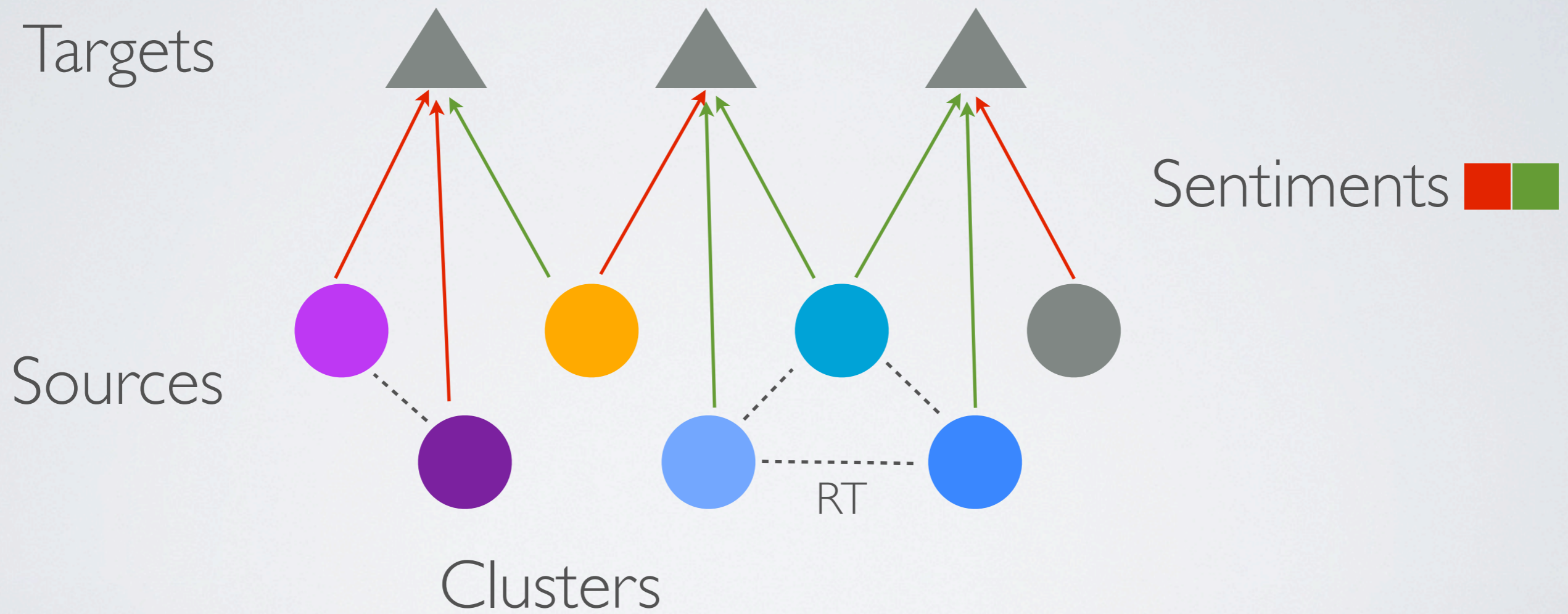
@nancygilesnyc

romney.supporter_diversity

May 5

negative

OPINION NETWORK



OPINION EXTRACTION

- Sub tasks:
 - Low level opinion annotation schemes
[Wiebe, 1999, 2005]
 - Identifying opinion expressions, their holders and targets
[Bethard et al., 2004; Breck, 2007; Choi
2005,2006;Kim&Hovy 2005;Wiebe 2006]

OPEN QUESTIONS

- What is a good annotation scheme to run on Amazon Turk?
- How to implement this scheme? client-side JS + server-side python + mysql ?
- What are the targets/attributes we are interested in?
- How to learn them from data in an unsupervised or semi-supervised fashion?
- How to identify opinion expressions, holders, targets?

REFERENCES

- Bo Pang and Lillian Lee (2008) “Opinion Mining and Sentiment Analysis”, *Foundations and Trends in Information Retrieval* 2(1-2), pp. 1–135, 2008. [See the references at the end of this review article]