Sentiment Analysis

Advanced Topics in Language Processing

Karo Moilanen

05-FEB-2008

karo.moilanen@comlab.ox.ac.uk

Sentiment Analysis

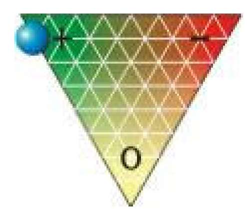
- Detection and analysis of expressions of
  - opinions, emotions, evaluations, beliefs, and speculations
  - cognitive *private* states that are not open to objective outside observation or verification
  - author/speaker sentiment, not sentiment evoked in the reader/hearer
  - difficult (impossible?) to formalise
  - no right answer, just points of view
- This lecture focuses on text, not speech

Sentiment Analysis

- "Clear" cases
  - *She is the **mother of all** mothers. She possessed character traits that other women **should strive** to meet. **Holiness**, **piety**, and **humbleness** are **valued very ...***
  - ***HOT!haha**.**Well** a **nice** Yamaha drum but **not properly handled** wif **care**,i **had to** ... it sounds like.....**noise**?**really noisy** songs,vocals **totally crap.haha**!so i ...*
- Less clear cases
  - *The patient scale consists of nine items: six subscales related to major lung **cancer symptoms** (**appetite**, **cough**, **dyspnea**, **fatigue**, **hemoptysis**, and **pain**), ...*
  - *By midday, the FTSE Eurofirst 300 **rebounded** 1.9 per cent 1,354.11, while Germany's Xetra Dax **rose** 2.1 per cent to 6,991.72 and in Paris the CAC 40 **added** 2.5 per cent to 4,988.80.*

Key Dimensions

- Subjectivity
  - subjective (*genius*), objective/factual (*hypertext markup language*)
- Polarity
  - positive (*love*), neutral (*fish*), mixed (*love-hate*), balanced
- Strength
  - weak (*nice*), strong (*extremely nice*)



- Affect
  - key universal categories (`anger`, `disgust`, `fear`, `joy`, `sadness`, `surprise`, ...)

Image from: Esuli,A. (2008). SentiWordNet. http://sentiwordnet.isti.cnr.it

Applications

- Classification
  - clustering, filtering
- Rating
  - ranking and labelling
- Summarisation
- Comparative analysis
  - pros vs. cons, support vs. opposition
- Role analysis
  - opinion holder, target
- Question Answering
- Lexicographic work

Applications

- Extraction
  - Information Extraction, Information Retrieval, Text Mining
- Affect analysis
  - Question/Answering and Dialogue
  - affective computing and HCI
- Tracking and visualisation
  - time series, regularities and irregularities, plotting, data summarisation, GUI design
- Financial analysis
- Media, marketing, financial, legal, and governmental organisations

Terminological Variance

- *Fully* standardised terminology still to emerge
  - subjectivity, sentiment, evaluation, private state, affect, emotion, opinion, attitude, appraisal, point of view, colouring, stance, perspective, tone, intent, modulation, …
  - analysis, classification, mining, tagging, extraction, …
  - positive, negative, thumbs up, thumbs down, good, bad, neutral, non-neutral, sentiment-bearing, recommended, not recommended, …
  - polarity, valence, (semantic) orientation, direction, slant, …
  - reversal, switcher, flipper, changing word, increase/decrease words…
  - strength, potency, force, ...
  - source, experiencer, holder, target, …

Levels of Analysis

- Document-level: *[(movie review)], [(news article)], …*
- Paragraphs: *[(conclusion)], …*
- Snippets: *[(search engine results)], [(n chars/words)], …*
- Sentences: *[The govt handled the crisis well], …*
- Clauses: *[It's not perfect] [but it is still useful], …*
- Phrases: *Life's nothing but [a bowl of rotten cherries], …*
- Words: *The [govt] [handled] the [crisis] [well], …*
- Morphemes: *[de-][conflict], [hope][-less], …*
- Senses: *[depressed] (button) vs. [depressed] (person), …*

Levels of Analysis

- Entities/objects: *[digital camera X], [person Y], [company W]…*
- Features/subparts: *[battery], [price], [keyboard layout], …*
- Roles: *according to [person Z], [person Q] claimed that, …*

Atomic Sentiment Carriers

- Single-word atomic carriers
  - *[victory]*$^{(+)}$, *[wrath]*$^{(-)}$, *[fondle]*$^{(+)}$, *[mutilate]*$^{(-)}$, *[brilliant]*$^{(+)}$, *[corny]*$^{(-)}$, *[admirably]*$^{(+)}$, *[badly]*$^{(-)}$,…
  - adjectives, adverbs, nouns, and verbs
- Multi-word atomic carriers
  - *[do away with smth]*$^{(-)}$, *[get it together]*$^{(+)}$, *[friendly fire]*$^{(-)}$, *[police state]*$^{(-)}$, *[de rigueur]*$^{(+)}$, *[head and shoulders above smth]*$^{(+)}$, *[on cloud nine]*$^{(+)}$, *[off the beaten track]*$^{(+)}$, *[live in an ivory tower]*$^{(-)}$, *[an old hand at smth]*$^{(+)}$, *[step on smb's toes]*$^{(-)}$, ...
- The expression inventory is vast
- Noticeably many infrequent instances

Case Study 1: Movie Reviews

- SA seen as an extension to standard text classification tasks
  - classify documents based on subject matter (SPORTS/ECONOMICS/…)
  - classify documents based on overall polarity
  - standard supervised text classification tools and techniques
- Learners
  - Naïve Bayes, MaxEnt, SVM
- Training
  - 700 (+) and 700 (-) movie reviews from the web
  - 3-fold cross validation

[Pang et al. (2002)]

Case Study 1: Movie Reviews

- Features
  - unigrams, bigrams, adjectives, POS tags, position in document
  - no stemming or stoplists
  - rudimentary support for negation
    - append a `NOT` tag to each word between a negator (e.g. *not*, *isn't*) and the first punct token on the right
  - represented as binary (presence) or frequency values

[Pang et al. (2002)]

Case Study 1: Movie Reviews

- Performance
  - random-choice baseline: 50
  - human-selected unigram baseline: 58 ~ 64
  - 82.9 accuracy (SVM, unigrams, binary)
- A further sentential subjectivity filter
  - classifier to filter out objective sentences
  - polarity classification proper done only subjective extracts
  - 86.4 accuracy
  - maintain the same level of performance with only 60% of the words in the text

[Pang et al. (2002)] [Pang & Lee (2004)]

Case Study 1: Movie Reviews

- Challenges
  - cf. accuracies of 90+ (with more classes) in standard text classification tasks
  - bag-of-features classifiers ignore discourse-temporal developments in text (e.g. the *“thwarted expectation”* rhetorical device in reviews)
  - reviews discuss multiple issues
    - events and actors in the movie
    - the style and art of the movie (the movie as a whole)
    - elaborative and contrasting devices
  - in web data, reviewers’ ratings and labels can be arbitrary
  - out-of-domain features perform poorly

[Pang et al. (2002)]

Case Study 2: PMI-IR

- Assumption
  - the polarity (Semantic Orientation (SO)) of a word tends to correspond to the polarities of its neighbours
- The polarity of a term is calculated using
  - a Pointwise Mutual Information (PMI) score of the term against key paradigmatic polarity terms
    - Pos = *{good, nice, excellent, positive, fortunate, correct, superior}*
    - Neg = *{bad, nasty, poor, negative, unfortunate, wrong, inferior}*
  - a “live” search engine (or a static corpus)
  - phrasal templates
    - `[JJ][NN]` *“romantic ambience”*, `[RB][JJ]` *“very cool”*

[Turney (2002)] [Turney & Littman (2003)]

Case Study 2: PMI-IR

- Context window
  - the original method used AltaVista's `NEAR` (±10 words) search operator (now deprecated)
  - using the `AND` search operator yields inferior results

$$PMI(t, t_i) = \log \frac{\#(\text{“}t \ \texttt{NEAR} \ t_i\text{”})}{\#(\text{“}t\text{”})\#(\text{“}t_i\text{”})}$$

$$SO(t) = \Sigma_{t_i \in Pos} PMI(t, t_i) - \Sigma_{t_i \in Neg} PMI(t, t_i)$$

- Also indicates strength of sentiment

[Turney (2002)] [Turney & Littman (2003)]

Formula from: Esuli,A. (2006). Opinion Mining.

http://medialab.di.unipi.it/web/Language+Intelligence/OpinionMining06-06.pdf

Case Study 2: PMI-IR

- Accuracy of term polarities
  - 3596 unambiguous words (adjectives, adverbs, nouns, and verbs) labelled manually as *positive* (1614) and *negative* (1982) from General Inquirer
  - 82.8 accuracy
  - 95+ when "mild" words are excluded
  - 97.1 for top (=highest confidence) 25% of words (899 words)
  - corpus size is crucial (61.3-68.7 accuracy using a 10-million word corpus (cf. hundred billion words))

[Turney & Littman (2003)]

Case Study 2: PMI-IR

- Document-level sentiment
  - the average of the SO scores of all adjectives and adverbs
- Performance
  - 74 accuracy on 410 reviews
  - 66 (movie reviews)
  - 84 (car reviews)

Extracted Phrase	Part-of-Speech Tags	Semantic Orientation
online experience	JJ NN	2.253
low fees	JJ NNS	0.333
local branch	JJ NN	0.421
small part	JJ NN	0.053
online service	JJ NN	2.780
printable version	JJ NN	-0.705
direct deposit	JJ NN	1.288
well other	RB JJ	0.237
inconveniently located	RB VBN	-1.541
other bank	JJ NN	-0.850
true service	JJ NN	-0.732
Average Semantic Orientation		0.322

[Turney (2002)]

Image from: Turney, P. (2002).

Case Study 3: Contextual Polarity

- Calculating the contextual polarities of sentiment expressions using
  - supervised machine learning (BoosTexter AdaBoost.HM, 5000 rounds boosting)
  - subjective expressions from the MPQA corpus annotated with contextual polarity tags (`+,-,n,both`) as training and test data
  - sentiment lexicon of 8000+ (manually and automatically derived) entries tagged with the above polarity tags and reliability indicators (`strong, weak`)
  - dependency parser
  - [neutral/polar classifier] → [polar classifier] architecture

[Wilson et al. (2005)]

Case Study 3: Contextual Polarity

- Task
  - classify the contextual polarity of subjective expressions which contain subjectivity clues from the lexicon
  - give each clue instance its own label
  - NB. expression boundaries are not determined
- Baseline
  - accuracy of always using the prior polarity of a clue instance as its contextual polarity: only 48

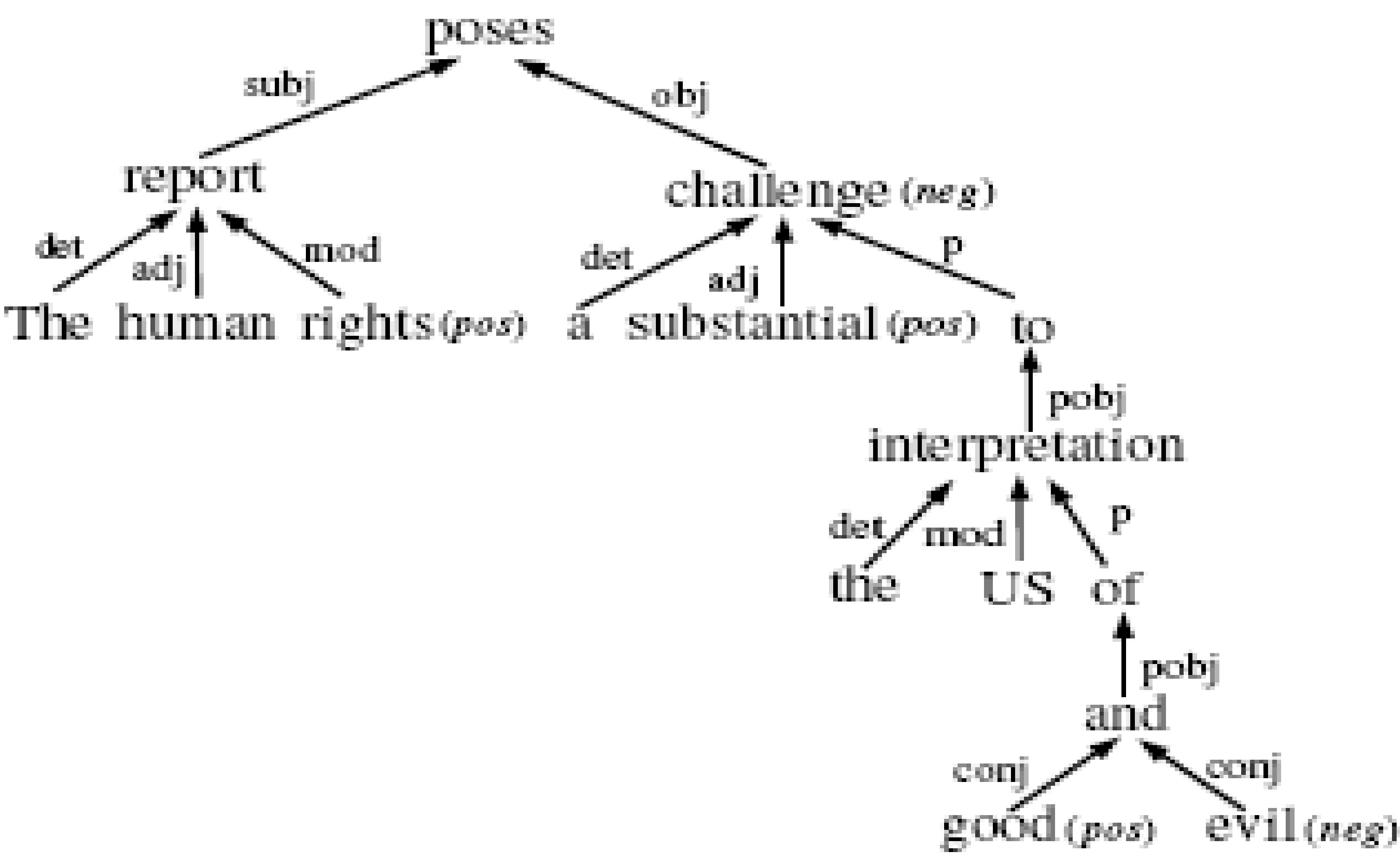
[Wilson et al. (2005)]

Case Study 3: Contextual Polarity

- Neutral/polar classifier
  - word features
    - token (*loves);* POS (`V`); trigram context (*she loves me*); prior polarity (+); reliability (`strong`)
  - modification features
    - preceded by (`ADJ/ADV/INTENSIFIER`); is intensifier (*terribly*); modifies or is modified by a (`strong/weak`) clue
  - structural features
    - is in subject (*criminals exist)*; is in copula (*is wonderful)*; is passive (*were destroyed)*

[Wilson et al. (2005)] [Polanyi & Zaenen (2006)]

Case Study 3: Contextual Polarity



[Wilson et al. (2005)]

Image from: Wilson, T. et al. (2005).

Case Study 3: Contextual Polarity

- Neutral/polar classifier
  - sentence features
    - # of `strong` and `weak` clues in previous, current, next sentences; POS tags counts in current sentence
  - document features
    - one of 15 document topics (`ECONOMICS`, `KYOTO PROTOCOL`,...)

[Wilson et al. (2005)]

Case Study 3: Contextual Polarity

- Polar classifier
  - word features
    - token (*loves*)*;* prior lexeme polarity (+)
  - negation features
    - is negated (*didn't love)*; is negated subject (*not a single volunteer came forward)*
  - modification features
    - modifies (*substantial <u>challenge</u>)* or is modified (*<u>substantial</u> challenge)* by a (`+,-,n,both`) token, or (`not mod`)

[Wilson et al. (2005)] [Polanyi & Zaenen (2006)]

Case Study 3: Contextual Polarity

- Polar classifier
  - conjunction features
    - is in a conjunction (*sweet and <u>sour</u>*) (`+,-,n,both,not mod`)
  - polarity shifter features
    - is general polarity shifter (*hardly successful)*; is positive polarity shifter (*alleviate smth);* is negative polarity shifter (*lack of smth)*

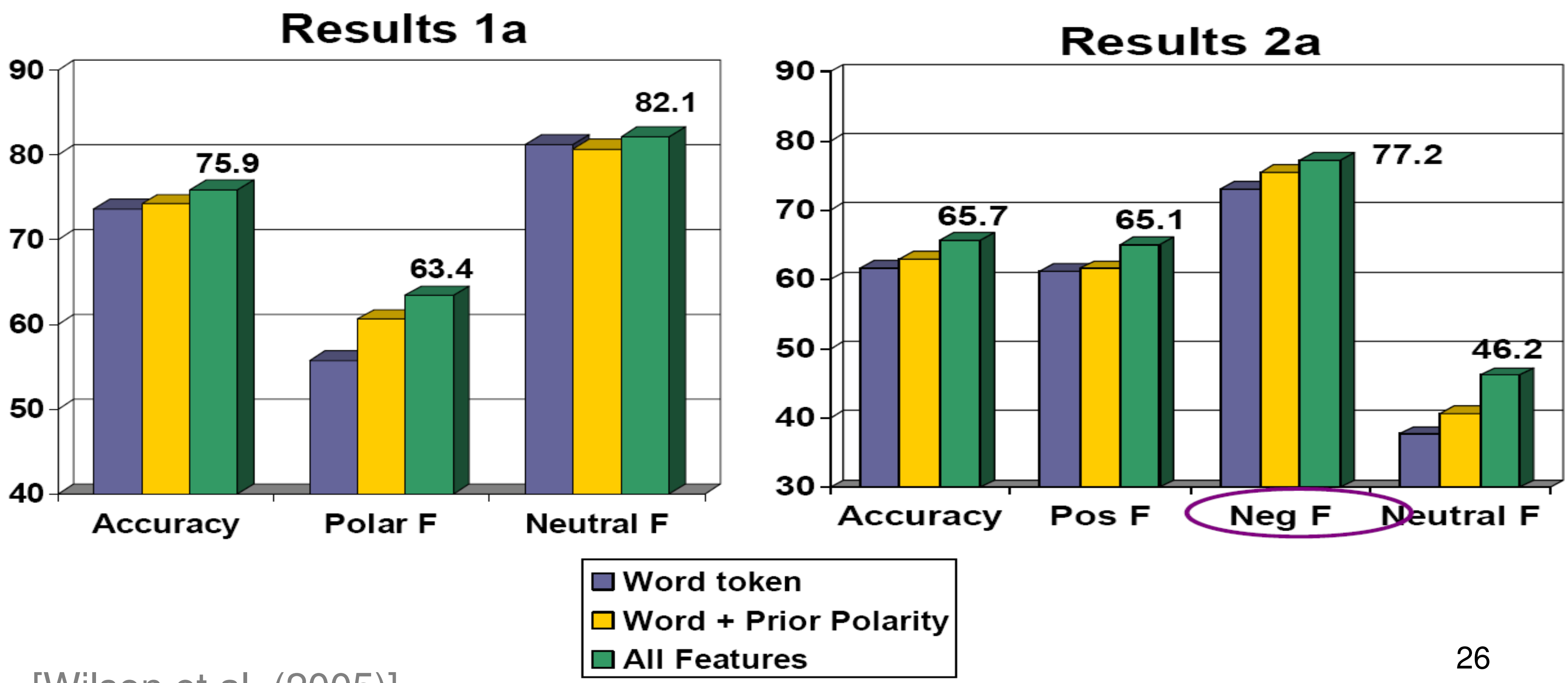
[Wilson et al. (2005)] [Polanyi & Zaenen (2006)]

Case Study 3: Contextual Polarity

- Performance

neutral/polar classifier

polar classifier



[Wilson et al. (2005)]

Images from: Wiebe, J, (2007). http://www.cs.pitt.edu/~wiebe/pubs/papers/EUROLAN07/eurolan07wiebe.ppt

Case Study 4: Product Features

- Descend to (sub)-sentential levels
  - expressing individual parts, features, and components of X
  - laptop *{ [+price], [-keyboard], [+sound card], ... }*
- Tasks
  - #1: identify and extract relevant features
    - term extraction and co-reference resolution
  - #2: determine the sentiment of features
    - <u>sentiment analysis proper</u>
  - #3: generate feature summary
    - clustering and ranking

Case Study 4: Product Features

- Common review formats
  - pros, cons, and detailed review sections (Epinions.com)
  - pros and cons sections (C|net.com)
  - free format (Amazon.com)
- Challenges
  - potentially noisy domain (misspellings, fragmentary language)
  - product features described using many expression types, not just NPs
  - relevant vs. irrelevant sections
  - unmentioned (implied) features

Case Study 4: Product Features

- *Creative Soundblaster Live 5.1: This is a* ***good*** *all around {**board**} with* ***good*** *{**software**} but the* ***lack*** *of {**support**} and* ***problems*** *related with this board if something goes* ***wrong*** *is* ***frustrating*** *and to be* ***avoided****.*
- *Ok, I am on my computer and put a cd in turn the music on, But the {**sound**} is just not what I expect, I am very* ***disapointed****. ...*
- *The SoundBlaster Live 5.1 from Creative is a {**well-priced**} (oem version) sound card with* ***excellent*** *{**features**} and {**sound**}.* ***Marred*** *by {**incompatibilty**} with many motherboards and needs to be {**set up**} by a real expert if you have a VIA chipset on your board.*
- *I have been using these {**speakers**} for some time and {**they**}* ***rock*** *...*
- *Although some have said about* ***random clicks****,* ***snaps*** *and* ***crackles*** *which apparently is a common issue with some people, I have not come across* ***such problems*** *...*

Excerpts from: http://www.reviewcentre.com/fi94-brand-Creative.html

Case Study 4: Product Features

- Term extraction (Hu & Liu 2004)
  - find explicit frequent noun/NP features (≤ 3 words)
  - POS tagging, simple NP and VP chunking
  - association mining (CBA, Apriori) to find frequent itemsets with 1% minimum support
  - discard compact features
    - *I had searched for a* ***digital camera*** *for 3 months* [compact]
    - *The !**camera** does not have a !**digital** zoom* [not compact]
  - discard redundant single-word features
    - *!manual* is a subset of *manual mode* or *manual setting*
    - minimal p-support ≥ 3

[Hu & Liu (2004)]

Case Study 4: Product Features

- Term extraction (Yi & Niblack 2005)
  - pre-specified subject terms (*camera*) with *part-of* (*lens*) and *attribute-of* (*resolution*) terms
  - definite sentence-initial base NPs (`the JJ* N+`)
  - keep *bnp*s with highest likelihood ratio (*lr*) scores (*-2logλ*) across topical (*D+*) and non-topical (*D-*) documents

	D_+	D_-
bnp	C_{11}	C_{12}
$\overline{bnp}$	C_{21}	C_{22}

$$\begin{aligned} r_1 &= \frac{C_{11}}{C_{11}+C_{12}} \\ r_2 &= \frac{C_{21}}{C_{21}+C_{22}} \\ r &= \frac{C_{11}+C_{21}}{C_{11}+C_{12}+C_{21}+C_2} \end{aligned}$$

$$\begin{aligned} lr = {} & (C_{11}+C_{21}) \cdot log(r) \\ & +(C_{12}+C_{22}) \cdot log(1-r) \\ & -C_{11} log(r_1) - C_{12} log(1-r_1) \\ & -C_{21} log(r_2) - C_{22} log(1-r_2) \end{aligned}$$

$$-2log\lambda = \begin{cases} -2 * lr & \text{if } r_2 < r_1 \\ 0 & \text{if } r_2 \geq r_1 \end{cases}$$

[Yi & Niblack (2005)]

Case Study 4: Product Features

- Sentiment analysis (Hu & Liu 2004)
  - list of opinion adjectives from the review corpus
  - the effective opinion of a frequent feature is the nearest adjective modifying the feature N/NP
    - *The [strap] is **horrible** and gets in the way of parts of the camera you need access to*
  - the noun/NP nearest to the opinion word invokes an infrequent feature
    - *Equally, the [bass]? can seem a little **lightweight** on thunderous [tracks]? but does retain focus.*

[Hu & Liu (2004)]

Case Study 4: Product Features

- Sentiment analysis (Yi & Niblack 2005)
  - sentiment lexicon (`beautiful JJ +`)
  - sentiment extraction database
    - `<predicate><sent_category: +/-|[~]source><target>`
  - the sentiment of the (SubjP/ObjP/ComplP/PrepP) `source` phrase is transferred towards a (SubjP/ObjP/PrepP) `target` phrase
    - `<impress><+><PrepP(by)>`: *I was impressed by <u>the build quality</u>*
    - `<offer><ObjP><SubjP>`: *<u>Dabs.com</u> offers high quality products*
  - negation inside phrases and predicates

[Yi & Niblack (2005)]

Case Study 4: Product Features

- Performance
  - Hu & Liu (2004)
    - tested on 500 manually-annotated product reviews of five products
    - 69.3 (r), 64.2 (p) on opinion sentence extraction
    - 84.2 on sentence polarity assignment
  - Yi et al. (2005)
    - tested on manually-annotated camera and music reviews
    - 56 (r), 87 (p)

[Hu & Liu (2004)][Yi & Niblack (2005)]

Case Study 5: Compositionality

- Goal
  - to calculate in a systematic way the polarities of syntactic constituents as a function of the polarities of their subconstituents
- Assumption
  - if the meaning of a sentence is a function of the meanings of its parts…
  - then the global polarity of a sentence is a function of the polarities of its parts
- Need
  - POS tagging, chunking, and ‘deep’ parsing

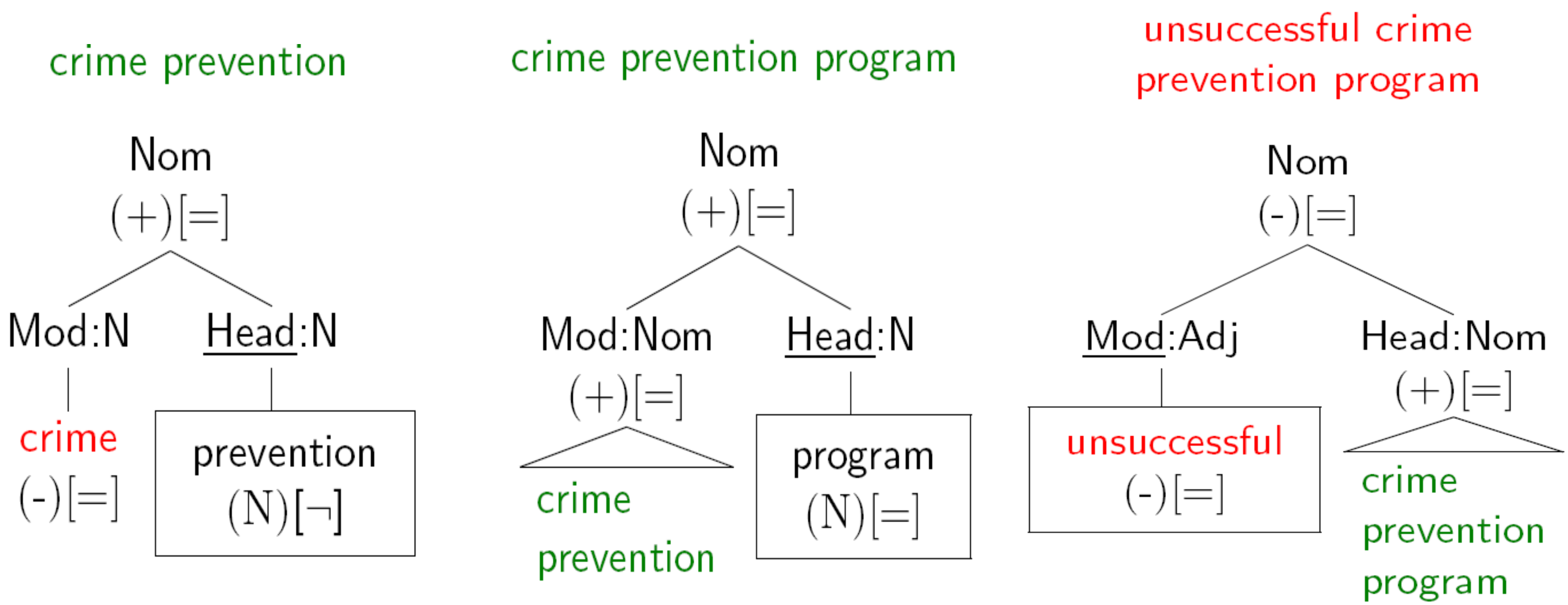
[Moilanen & Pulman (2007)]

Case Study 5: Compositionality

- Combine two elements
  - two morphemes, words, phrases, clauses, …
- Element ranking
  - each syntactic situation (i.e. node in the parse tree) determines which one dominates
- Polarity ranking
  - `{(+),(−),(M)} > (n)`
- Lexical tags
  - equal (`[=]`), reverse (`[¬]`)

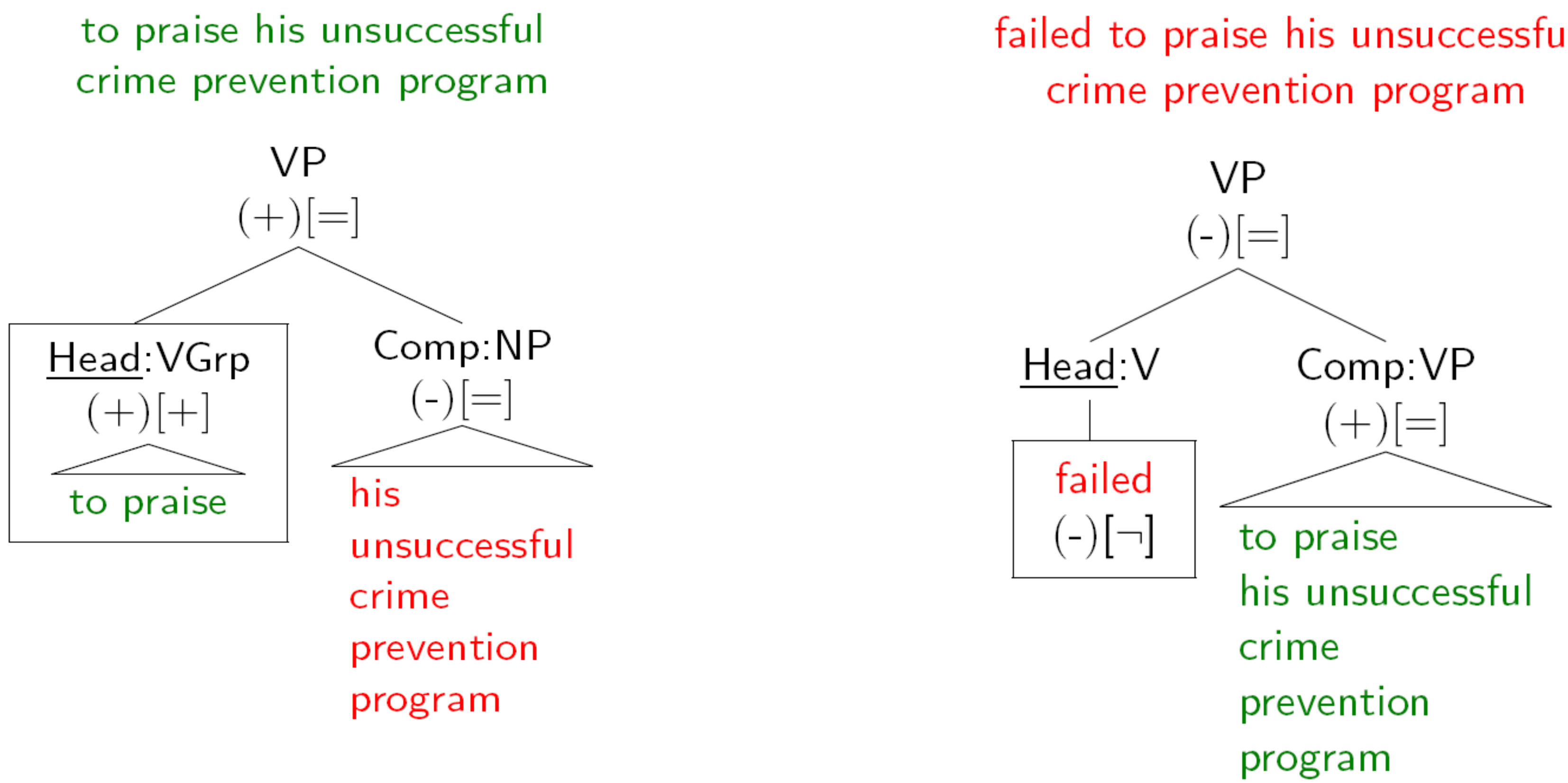
[Moilanen & Pulman (2007)]

Case Study 5: Compositionality



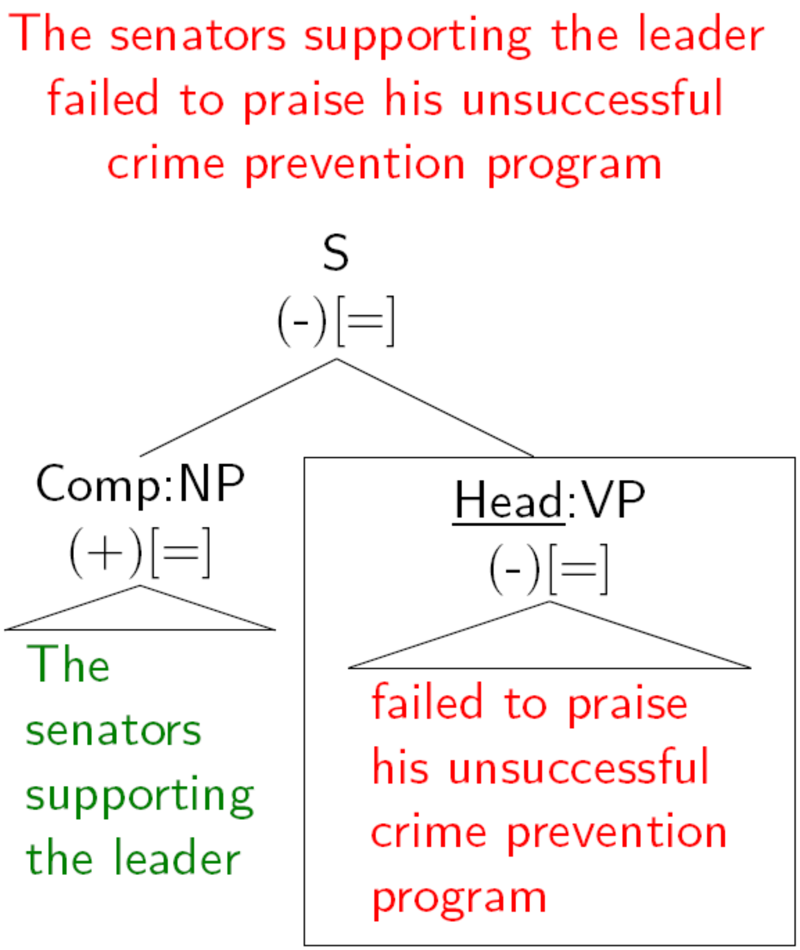
[Moilanen & Pulman (2007)]

Case Study 5: Compositionality



[Moilanen & Pulman (2007)]

Case Study 5: Compositionality



- Performance
  - short headlines
    - 63 - 76.3
    - 81.7 - 90 (strong cases)
  - NPs
    - 72.5 - 85.5
    - 79.2 - 89.1 (strong cases)
- Challenges
  - tagging and parsing errors
  - lexicon coverage
  - paralogical devices

[Moilanen & Pulman (2007)]

Evaluation Issues

- Sentiment is an inherently fuzzy phenomenon
- Typical gold standards
  - a very small number of annotators (2)
  - relatively small ad hoc hand-labelled datasets or large unverified hand-/automatically-labelled datasets
  - not all datasets are publicly available
- Common source of confusion in annotations
  - point of view, neutral polarity, category boundaries, category labels and scales, text region widths, …

Evaluation Issues

- Reported inter-annotator agreement scores
  - vary depending on the task, genre, level of analysis, and labels used
- What constitutes an acceptable baseline/upper bound for the algorithm?
- Few attempts to measure human performance
- Do not treat gold standards as holy
- Extremely high results may indicate data overfitting

Corpora and Datasets

- Hand-labelled datasets
  - not many available
  - preferred option
  - limited size can be a problem
- Automatically-compiled datasets
  - abundant raw data available
  - noisy labels, rankings, and language
- Tag diversity
  - mapping between different annotation/analytical frameworks can introduce further complications

Corpora and Datasets

- MPQA:
  - phrases and sentences annotated with subjectivity, polarities, and roles
  - 535 documents, 11114 sentences
  - http://www.cs.pitt.edu/wiebe/mpqa
- SemEval-07 Task 14
  - -100 to +100 polarity scale, six affect categories
  - 1200 news headlines
  - http://www.cse.unt.edu/~rada/affectivetext/
- Movie Review Data:
  - 1000 (+) and 1000 (-) reviews
  - 5331 (+) and 5331 (-) sentences / snippets
  - 5000 subjective and 5000 objective sentences
  - documents with ratings
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/
- FBS product features, 5 and 9 products
  - http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
- NPS: FBS product features with added base NP boundaries
  - (Contact me)
- The web…

Lexica

- General Inquirer
  - a manually-compiled word list, multi-dimensional affect tags
  - http://www.wjh.harvard.edu/%7Einquirer/
- SentiWordnet
  - WordNet 2.0 synsets annotated automatically with (+), (-), and (n) scores
  - http://sentiwordnet.isti.cnr.it/
- WordNetAffect
  - a subset of WordNet synsets labelled (semi-)automatically with multi-dimensional affect tags
  - http://wndomains.itc.it/download.html
- Sentiment and subjectivity clues by Wiebe et al.
  - http://www.cs.pitt.edu/~wiebe/pubs/index.html
- Affective Norms for English Words (ANEW)
  - a manually-compiled word list, multi-dimensional tags
  - http://csea.phhp.ufl.edu/media/anewmessage.html

References

- Hu, M. & Liu, B. (2004). Mining and Summarizing Customer Reviews. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004). Seattle. http://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf
- Moilanen, K. & Pulman, S. (2007). Sentiment Composition. In Proceedings of Recent Advances in Natural Language Processing (RANLP 2007). Borovets, Bulgaria, pp. 378–382. http://users.ox.ac.uk/~wolf2244/sentCompRANLP07Final.pdf
- Pang, B. & Lee, L. (2004). A Sentimental Education: Sentiment Analysis using Subjectivity Summarization Based on Minimum Cuts. In Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics. Barcelona. pp. 271–278. http://www.cs.cornell.edu/home/llee/papers/cutsent.home.html
- Pang, B. et al. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of EMNLP-02, Conference on Empirical Methods in Natural Language Processing. Philadelphia. pp. 79–86. http://www.cs.cornell.edu/home/llee/papers/sentiment.home.html
- Polanyi, L. & Zaenen, A. (2006). Contextual Valence Shifters. In J. Shanahan, Y. Qu, and J. Wiebe (Eds.). Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series, Vol. 20. Springer. pp. 1–10.
- Turney, P. (2002). Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics. pp. 417–424. http://iit-iti.nrc-cnrc.gc.ca/publications/nrc-44946_e.html

References

- Turney, P. & Littman, M. (2003). Measuring Praise and Criticism: Inference of Semantic Orientation from Association. In ACM Transactions on Information Systems, 21(4), October. pp. 315–346. http://iit-iti.nrc-cnrc.gc.ca/publications/nrc-46516_e.html
- Yi, J. & Niblack, W. (2005). Sentiment mining in WebFountain. In the Proceedings of the 21st International Conference on Data Engineering (ICDE 2005). pp. 1073–1083. http://ieeexplore.ieee.org/iel5/9680/30564/01410217.pdf

- Online bibliographies
  - http://patty.isti.cnr.it/~esuli/research/sentiment/biblio.html
  - http://www.cisr.iri.titech.ac.jp/~inui/SA/wg/index.html
  - http://www.cs.pitt.edu/~wiebe/pubs/papers/EUROLAN07/eurolan07bib.html