Computational Modelling of Word Sense Sentiment

Fangzhong Su

Supervisors: Katja Markert & Eric Atwell

Natural Language Processing Group
School of Computing

October 29, 2007

Outline

Introduction

- Traditional Data Mining
  Text Mining: topic/fact (sports, finance, politics, science,etc..)
- Opinion Mining
  Opinion/Attitude: thumb-up (positive) or thumb-down (negative)
- Examples
  Movie/product reviews
  Social event/policy survey

Motivation

Potential Application:

- Automatic summarization of reviews
- Opinion Extraction
- Detection of hostile messages
- ...

Task

- Aim
  Predict the semantic orientation of word senses

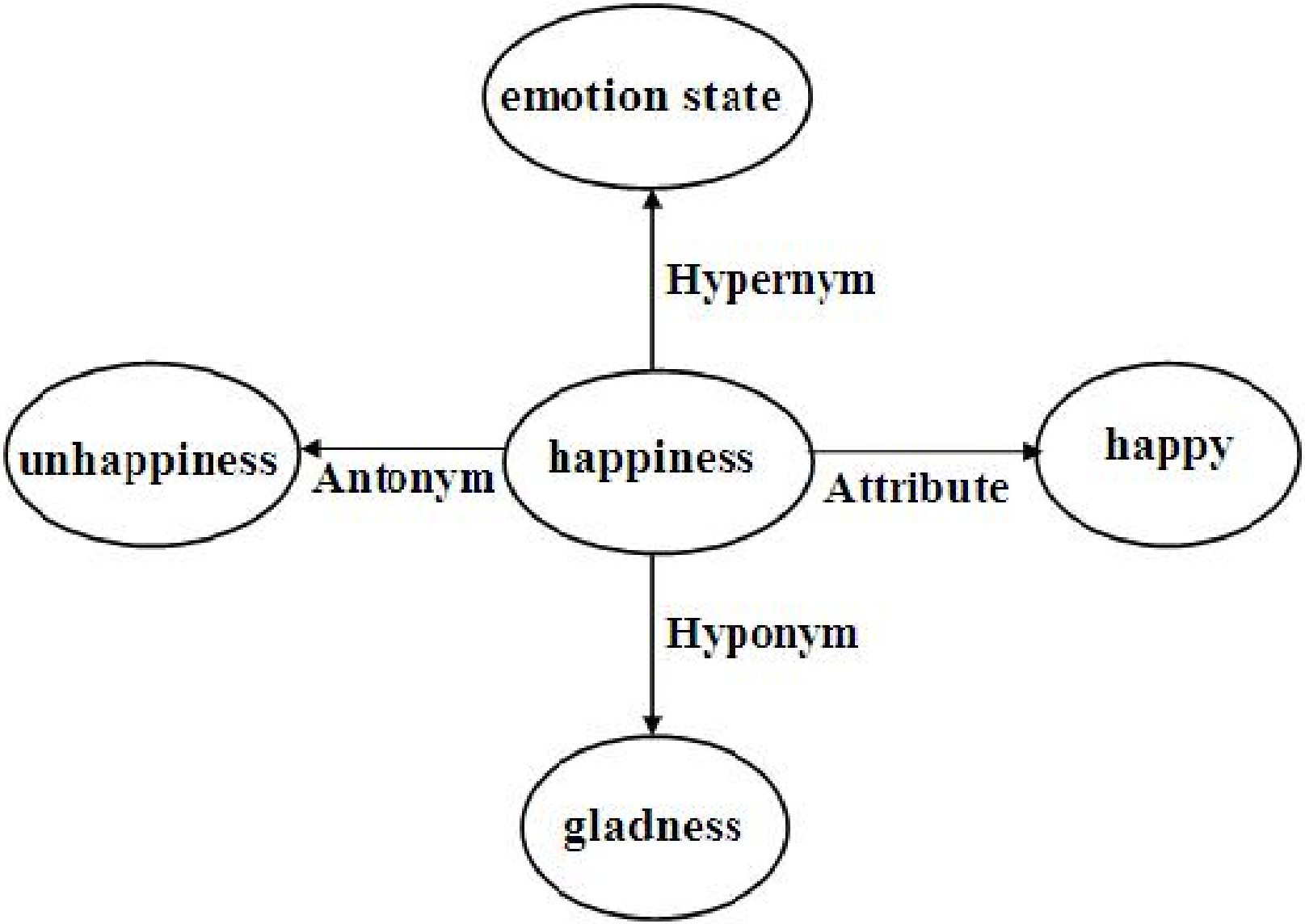
- Task Description
- Input
  A collection of word sense: $ws_1, ws_2, ws_3 \ldots ws_n$
- Output
  The semantic orientation of the corresponding word senses: $SO(ws_1), SO(ws_2), SO(ws_3) \ldots SO(ws_n)$

WordNet

- A large lexical database of English[1, 2]
- Synsets: Synonym Set
  - definition of the words
  - each synset corresponds to a distinct concept
- Relations

  Synonym, antonym, similar-to, attribute, hypernym, hyponym...

Example


emotion state
Hypernym
unhappiness
Antonym
happiness
Attribute
happy
Hyponym
gladness

Semantic Orientation

- Subjectivity
  Subjective(S), Objective(O), Both(B)

  Opinion, emotion, evaluation, speculation etc...[3]
- Connotation

  Polarity: positive(S:P/O:P), negative (S:N/O:N) and no connotation (O:O),varied(S:V)
- Current Work

  Start from subjectivity categorization

  S (S:P, S:N and S:V)

  O (O:P, O:N and O:O)

Examples

- winner, victor-(the contestant who wins the contest)

Examples

- winner, victor-(the contestant who wins the contest) O:P
- death, decease – (the event of dying or departure from life; "her death came as a terrible shock")

Examples

- winner, victor-(the contestant who wins the contest)
  O:P
- death, decease – (the event of dying or departure from life; "her death came as a terrible shock")
  O:N
- above-(appearing earlier in the same text; "flaws in the above interpretation")

Examples

- winner, victor-(the contestant who wins the contest)
  O:P
- death, decease – (the event of dying or departure from life; "her death came as a terrible shock")
  O:N
- above-(appearing earlier in the same text; "flaws in the above interpretation")
  O:O

Examples

- happy-(enjoying or showing or marked by joy or pleasure or good fortune: "a happy smile"; "spent many happy days on the beach"; "a happy marriage")

Examples

- happy-(enjoying or showing or marked by joy or pleasure or good fortune: "a happy smile"; "spent many happy days on the beach"; "a happy marriage") S:P
- nasty, awful – (offensive or even (of persons) malicious; "in a nasty mood"; "a nasty accident"; "a nasty shock"; "a nasty smell"

Examples

- happy-(enjoying or showing or marked by joy or pleasure or good fortune: "a happy smile"; "spent many happy days on the beach"; "a happy marriage")
  S:P
- nasty, awful – (offensive or even (of persons) malicious; "in a nasty mood"; "a nasty accident"; "a nasty shock"; "a nasty smell"
  S:N
- alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse – (a person who drinks alcohol to excess habitually)

Examples

- happy-(enjoying or showing or marked by joy or pleasure or good fortune: "a happy smile"; "spent many happy days on the beach"; "a happy marriage")
  S:P
- nasty, awful – (offensive or even (of persons) malicious; "in a nasty mood"; "a nasty accident"; "a nasty shock"; "a nasty smell"
  S:N
- alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse – (a person who drinks alcohol to excess habitually)
  Both

Terms

- Semantic Orientation of words

Terms

- Semantic Orientation of words

  Example:

word	polarity
pig	objective
positive	positive
negative	negative

Terms

- Semantic Orientation of words

  Example:

word	polarity
pig	objective
positive	positive
negative	negative

"One word only corresponds to one polarity"

Terms

- Semantic orientation of word senses
  Example:[1, 2]
  (1)hog, pig, grunter, squealer, Sus scrofa – (domestic swine)(objective)
  (2)slob, sloven, pig, slovenly person – (a coarse obnoxious person)(negative)
  (3)positive – (characterized by or displaying affirmation or acceptance or certainty etc.; "a positive attitude")(positive)
  (4)positive – (greater than zero; "positive numbers")(objective)

Terms

- Semantic orientation of word senses
  Example:[1, 2]
  (1)hog, pig, grunter, squealer, Sus scrofa – (domestic swine)(objective)
  (2)slob, sloven, pig, slovenly person – (a coarse obnoxious person)(negative)
  (3)positive – (characterized by or displaying affirmation or acceptance or certainty etc.; "a positive attitude")(positive)
  (4)positive – (greater than zero; "positive numbers")(objective)

  "One word may have several different word senses, which may correspond to different semantic orientations"

Proposed Algorithms

- Basic idea

  Semi-supervised learning(SSL)[4, 5, 6] with constraints[7, 8]
- Intuition
  - A small number of labeled data
  - A large number of unlabeled data
- Constraints

Constraints

- WordNet synset relations[1, 9]

  Synonym, antonym, direct hyponym...etc

- Morphological information

  Prefix: happy-unhappy, convenient-inconvenient...etc

  Suffix: happy-happiness...etc

Algorithm Framework

1. Train a classifier on labeled data only
2. Apply the classifier to unlabeled data, assign probabilistic values
3. Incorporate the constraints information to reweigh the probabilistic value
4. Update the parameters of the classifier
5. Go back to step 2 until convergence

Alternative Proposal

Monosemous words (=word senses)

- Characteristics
  (1)infrequent word
  (2)each word corresponds to only one sense
  (3)single polarity
- Strategy
  (1)compute co-occurrence frequency with subjective paradigm words [10]
  (2)Threshold setting for polarity filtering

Completed Work

Annotation Study

- Micro-WNOp corpus[11]
  3 groups, 1105 WordNet synsets
- Agreement statistic
  agreement $\geq$ 83%
- Kappa statistic
  kappa$\geq$ 76%

Completed Work

Annotation Study

- Micro-WNOp corpus[11]
  3 groups, 1105 WordNet synsets
- Agreement statistic
  agreement $\geq$ 83%
- Kappa statistic
  kappa$\geq$ 76%
- Conclusion—"a well-defined task"

Completed Work

Baseline Measurement

- Majority category: 62.5%
- Rule-based Method: 67.7%
- Classification[12, 13] without feature reduction: 64.5%
- Classification with feature reduction
  - Word Frequency threshold: 63.4%
  - SVD(Singular Value Decomposition): 69.1%

Future Work

- Implementation
  - Semi-supervised learning with constraints
  - Semantic orientation detection of monosemous synsets
- Exploration of potential evaluation methodology
- Embedding sentiment information of word senses in other opinion mining tasks.

Reference I

[1] Christiane Fellbaum.

Wordnet:an electronic lexical database.

The MIT Press, Cambridge, MA., 1999.

[2] Wordnet.

http://wordnet.princeton.edu/.

[3] Janyce Wiebe and Rada Mihalcea.

Word sense and subjectivity.

ACL, 2006.

[4] Xiaojin Zhu.

Semi-supervised learning literature survey.

Technical report, University of Wiscosin-Madison, Department of Computer Science, 2005.

Reference II

[5] Xiaojin Zhu.

Semi-supervised learning.

ICML 2007 tutorial,Corvallis, OR, USA, http://pages.cs.wisc.edu/ jerryzhu/icml07tutorial.html, 2007.

[6] Kamal Nigam.

Using unlabeled data to improve text classification.

Doctorial Dissertation, Computer Science Department, Carnegie Mellon University, 2001.

[7] Yoshimasa Tsuruoka and Junichi Tsujii.

Training a naive bayes classifier via the em algorithm with a class distribution constraint.

Proceeding of the Seventh Conference on Natural Language Learning(CoNLL), pages 127–134, 2003.

Reference III

[8] Mingwei Chang, Lev Ratinov, and Dan Roth.

Guiding semi-supervision with constraint-driven learning.

ACL'07, Prague, Czech Republic, 2007.

[9] Andrea Esuli and Fabrizio Sebastiani.

Sentiwordnet: A publicly available lexical resource for opinion mining.

Proceedings of LREC 2006 - 5th Conference on Language Resources and Evaluation, Genova, Italia, 2006.

[10] Peter Turney and Michael L.Littman.

Measuring praise and criticism: Inference of semantic orientation from association.

ACM Transaction on Information Systems, 21(4):315–346, 2003.

[11] Micro-wnop.

http://www.unipv.it/wnop/, 2007.

Reference IV

[12] Ian H. Witten and Eibe Frank.

Data mining: Practical machine learning tools and techniques.

2nd Edition, Morgan Kaufmann,San Francisco, 2005.

[13] Weka.

http://www.cs.waikato.ac.nz/ml/weka/, 2005.

Any questions?