

Lexicon Induction

Melanie Bolla and Olga Whelan
Ling 575

Lexicon Induction

(and the problem it addresses)

Automatic extraction of semantic dictionaries from textual corpora

Some applications:

- collection of words belonging to the same semantic category (semantic lexicons)
- induction of translation pairs based on distributional properties

Lexicon induction compensates for the lack of existing annotated data on sentiment.

Papers

1. Vasileios Hatzivassiloglou and Kathleen McKeown (1997). *Predicting the Semantic Orientation of Adjectives*.
2. Ellen Riloff and Janyce Wiebe (2003). *Learning Extraction Patterns for Subjective Expressions*.
3. Peter D. Turney and Michael L. Littman (2003). *Measuring Praise and Criticism: Inference of Semantic Orientation from Association*.

Focus of papers

Lexicon Induction techniques for Sentiment Analysis

- polarity: (1), (3)
 - positive or negative (or neutral)
- subjectivity: (2)
 - subjective or objective

Predicting the Semantic Orientation of Adjectives (Hatzivassiloglou, McKeown)

- Important study on adjective polarity; influenced other, more recent works.

[Predicting the semantic orientation of adjectives](#)

V Hatzivassiloglou, KR McKeown - ... of the 35th Annual Meeting of the ..., 1997 - dl.acm.org

Abstract We identify and validate from a large corpus constraints from conjunctions on the positive or negative semantic orientation of the conjoined adjectives. A log-linear regression model uses these constraints to predict whether conjoined adjectives are of same or ...

[Cited by 1197](#) [Related articles](#) [All 28 versions](#) [Cite](#) [Save](#)

- Google Scholar citation count: 1197

Predicting the Semantic Orientation of Adjectives (Hatzivassiloglou, McKeown)

1. explored constraints on semantic orientation of conjoined adjectives
2. used a model to predict whether two adjectives share the same polarity
 - log-linear regression
 - morphology rules
3. assigned the adjectives to one of two groups of opposite orientation
 - iterative optimization - clustering algorithm
4. established the polarity of the group (positive or negative)
 - comparing average frequencies of the adjectives in each group

Hypothesis

- Conjunctions provide indirect information on orientation because they impose constraints on the semantic orientation of their arguments
- For most connectives (except *but*) the adjectives have the same orientation

The tax proposal was

*{ simple and well-received
simplistic but well-received
simplistic and well-received }

by the public.

- Synonyms have same orientation; antonyms have the opposite

Application: refining extraction of semantic similarities (antonyms, synonyms)

1. Data: adjectives and conjunctions

- POS-annotated WSJ corpus (21 million words)
 - selected adjectives appearing more than 20 times
 - labelled for polarity (1,336: +657 -679)
 - 500 labels validated by independent annotation (96.97%)
 - Two-level finite-state grammar collected 15,431 conjoined adjective pairs
 - morphological transformations ➡ 9,296 pairs
 - **Classification of conjunctions** - validates the hypothesis
 - parser classifies conjunctions
 - three-way cross-classification
- 77.84% -
same orientation!

2. Same or different polarity?

- baseline: all the conjunctions have the same orientation (except *but*)
- morphological analyzer - word formations often have the opposite polarity (*adequate - inadequate, thoughtful - thoughtless*)
- log-linear regression - uses info from different conjunction categories

Prediction method	Morphology used?	Accuracy on reported same-orientation links	Accuracy on reported different-orientation links	Overall accuracy
Always predict same orientation	No	77.84%	—	77.84%
	Yes	78.18%	97.06%	78.86%
<i>But</i> rule	No	81.81%	69.16%	80.82%
	Yes	82.20%	78.16%	81.75%
Log-linear model	No	81.53%	73.70%	80.97%
	Yes	82.00%	82.44%	82.05%

3. Finding groups with same polarity

- each pair of adjectives has a dissimilarity value [0, 1]
 - same orientation \longrightarrow low dissimilarity
 - different orientation \longrightarrow high dissimilarity
- these links form a graph; nodes are divided into two subsets based on orientation using non-hierarchical clustering algorithm

- create random partition; find P
- to minimize $\Phi(P)$ adjectives are iteratively

$$\Phi(P) = \sum_{i=1}^2 \left(\frac{1}{|C_i|} \sum_{\substack{x, y \in C_i \\ x \neq y}} d(x, y) \right)$$

moved from one cluster to another until $\Phi(P)$ can't be improved

4. Label Clusters for Polarity

- computing average frequency of words in each cluster
- group with higher average frequency is labelled as positive

WHY?

Vasileios Hatzivassiloglou and Kathleen McKeown (1993). *Towards The Automatic Identification Of Adjectival Scales: Clustering Adjectives According To Meaning*

- semantically unmarked adjectives are more frequent in oppositions (81%)
- unmarked members are almost always positive

Evaluation: sparse test set

Demonstrated how the performance depends on the corpus size and graph density:

A_{α} - subset of A including adj x iff there are at least α links L between x and other elements of A

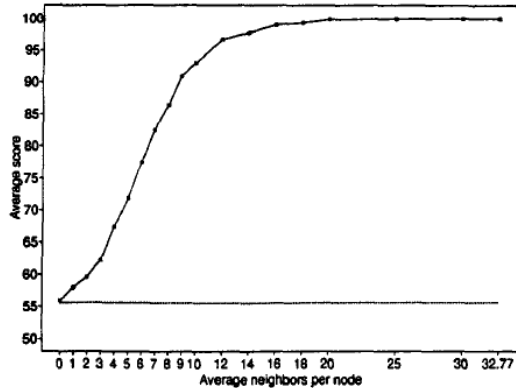
Accuracy grows with the number of links per adjective

α	Number of adjectives in test set ($ A_{\alpha} $)	Number of links in test set ($ L_{\alpha} $)	Average number of links for each adjective	Accuracy	Ratio of average group frequencies
2	730	2,568	7.04	78.08%	1.8699
3	516	2,159	8.37	82.56%	1.9235
4	369	1,742	9.44	87.26%	1.3486
5	236	1,238	10.49	92.37%	1.4040

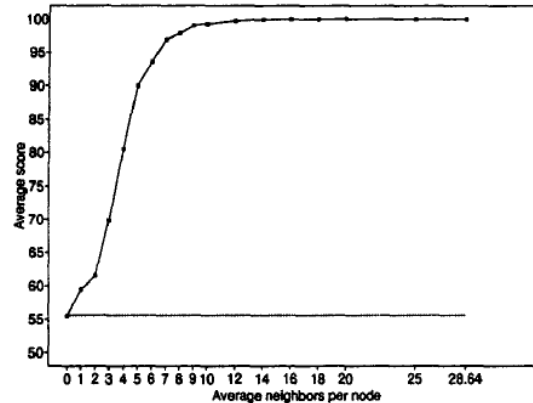
Evaluation - simulation experiments

Performance for a given level of precision P of identifying links and an average number of links k per adjective:

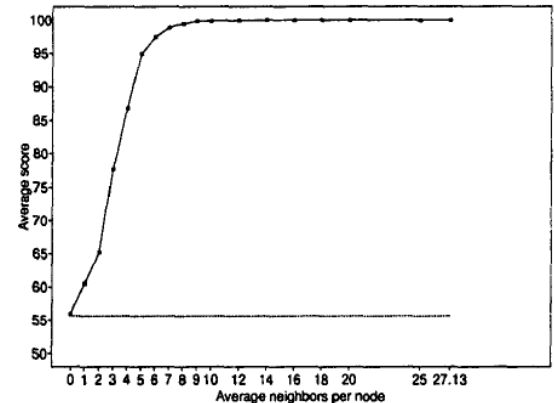
Even for low P and k , the ability to classify the adjectives correctly is very high
for $P=0.8$ and $k=12$ performance reaches 99%



(a) $P = 0.75$



(c) $P = 0.85$



(d) $P = 0.9$

Goals and achievements

- automatically establish semantic orientation of adjectives using indirect linguistic features extracted from corpus
 - orientation of conjoined adjectives using conjunction information
 - polarity of a group of adjectives with the same orientation based on their semantic relationships
- conjunctions place linguistic constraints on the adjectives they connect
- prove that relations between conjunctions and adjectives can be described in binary terms of *and* (interconnection) and *but* (contradiction)
- high level of precision can be achieved using a fairly small number of links between graph nodes

Why is it important?

- explores use of morphology in finding semantic orientation
- can compensate for impracticality of semantic information on polarity (i.e. definitions), which is unwieldy, rarely provided and often incomplete
- contribute to automatic identification of synonyms and antonyms, including contextually
- can be extended to other parts of speech and a broader set of conjunctions, as well as to, inversely, interpret the conjunctions themselves

What we learned

- positive adjectives have higher frequency
- corpus can be represented as graph
- a very basic baseline approach that assigns same-orientation link to all conjoined pairs with an exception for *but* works pretty well - 81.75% overall

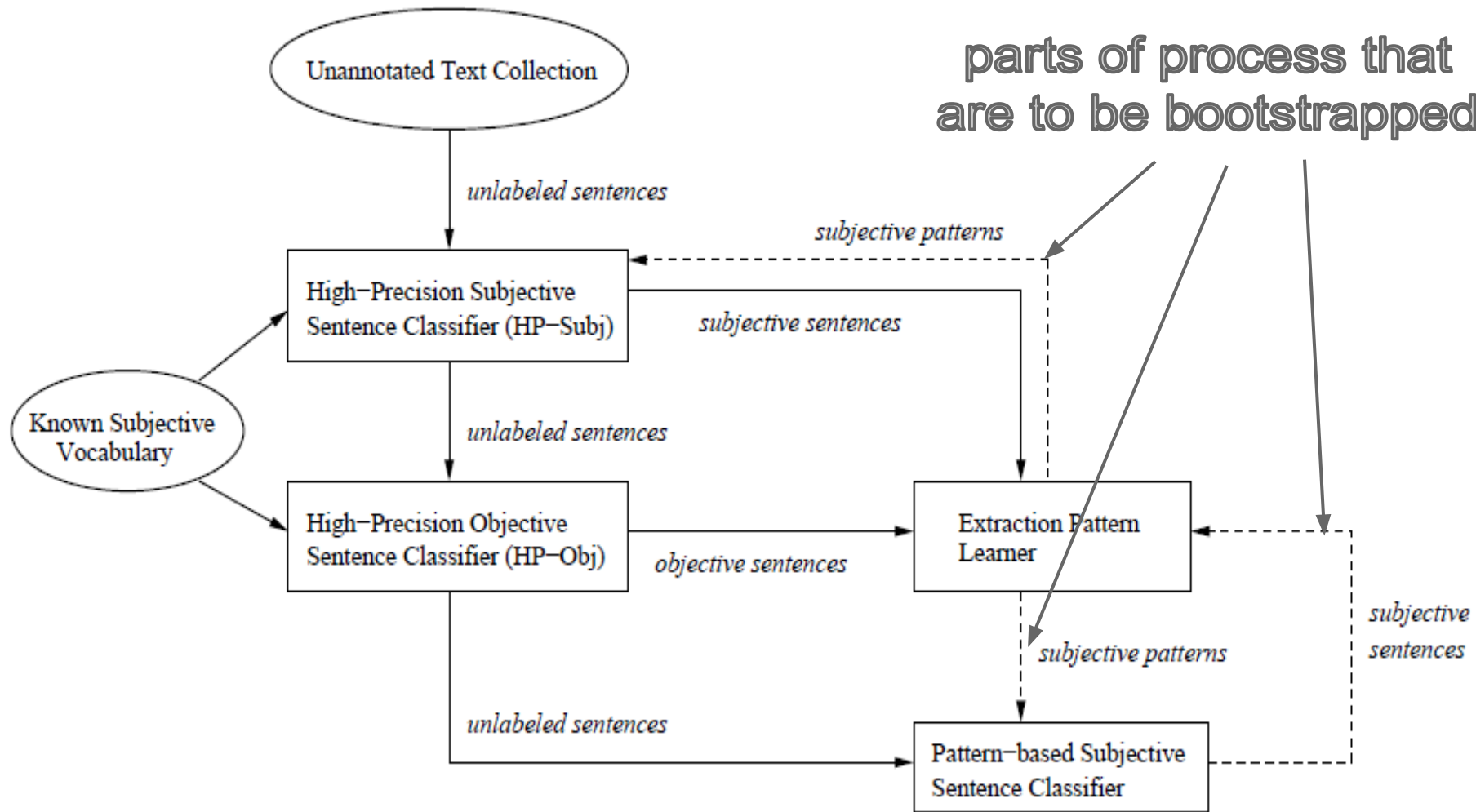
Critique

- Orientation labels
 - How were they assigned?
 - If automatically, what was the method?
 - If manually, did the authors perform it?
- Morphological analyzer
 - How elaborate was it?
 - Was there a list of affixes they considered to claim that adjectives related in form almost always have different semantic orientation?

Learning Extraction Patterns for Subjective Expressions (Riloff, Wiebe)

Bootstrapping process

1. high precision classifiers label unannotated data for training
 - a. subjective classifier (HP-Subj)
 - b. objective classifier (HP-Obj)
2. extraction pattern learner (similar to AutoSlogTS, (Riloff, 1996))
 - a. learn new subjective patterns from data output of (1)
3. identification of more subjective sentences due to learned patterns of (2)



1. HP-classifiers

Data for extraction patterns comes from FBIS foreign news documents

1. Subjectivity clues

- are lists of lexical items (words, N-grams)
- come from reliable manually developed or derived sources
- can be *strongly* and *weakly subjective*

2. HP-Subj

- 2+ strongly subjective clues; 91.5% precision, 31.9% recall

3. HP-Obj

- 1 or fewer weakly subjective clues; 82.6% precision, 16.4% recall

2. Learning subjective patterns

1. Syntactic templates applied to corpus - extraction patterns generated for every template that appears in corpus
2. Gather statistics on frequency of occurrence in subjective vs. objective sentences
3. Ranking the patterns using conditional probability measure + thresholds to ensure subjectivity

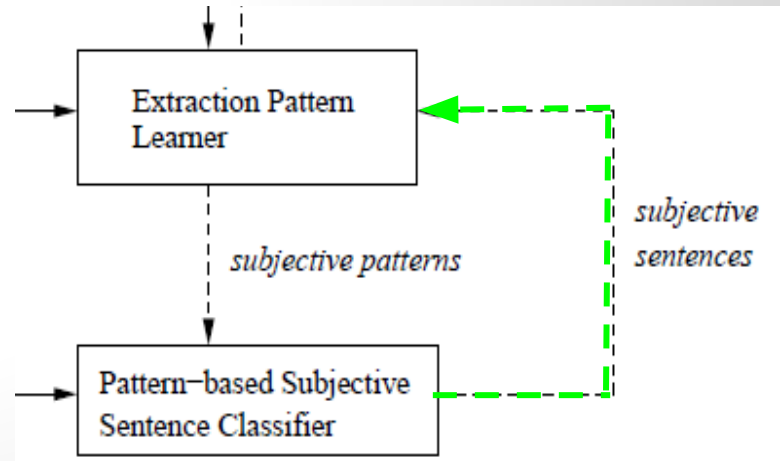
PATTERN	FREQ	%SUBJ
<subj> was asked	11	100%
<subj> asked	128	63%
<subj> is talk	5	100%
talk of <np>	10	90%
<subj> will talk	28	71%
<subj> put an end	10	90%
<subj> put	187	67%
<subj> is going to be	11	82%
<subj> is going	182	67%
was expected from <np>	5	100%
<subj> was expected	45	42%
<subj> is fact	38	100%
fact is <dobj>	12	100%

3. Finding new subjective sentences

New subjective sentences are fed back

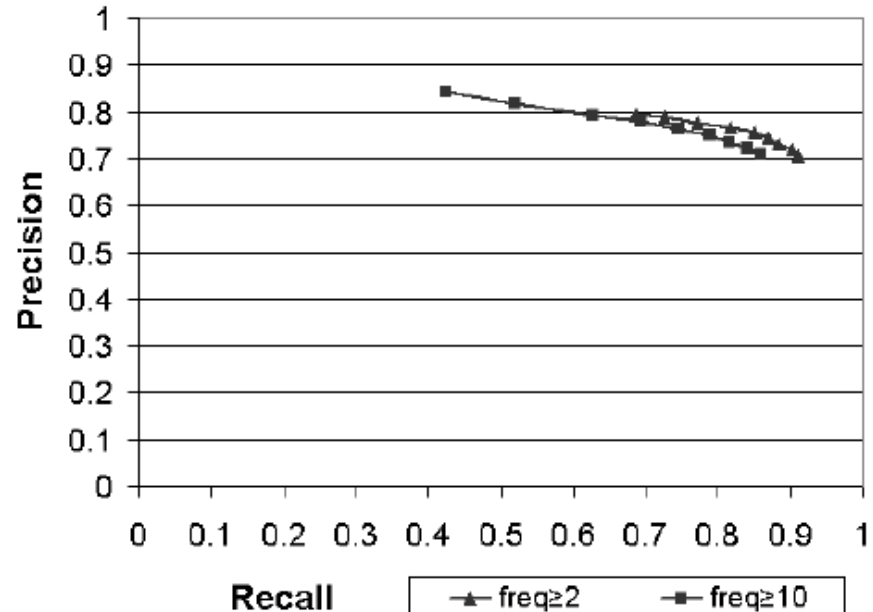
to the extraction pattern learner;

bootstrapping cycle is complete!



Evaluation - learning

- 210 sentences manually annotated for low/medium/high/extreme strength of private state - 90% agreement
- clear subjective, objective cases + borderline harder to discern
- precision measured for different frequency thresholds
- 71% < precision < 85%
extraction patterns are effective



Evaluation - bootstrapping

- Pattern-Based Subjective Classifier: 9,500 new subjective sentences (cf. with 17,000 of initially found by HP-classifiers)
- extraction pattern learner: 4,248 new patterns (less with stricter threshold)

new patterns allow to label more sentences as subjective without great loss of precision

HP-Subj		HP-Subj w/Patterns	
<i>Recall</i>	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>
32.9	91.3	40.1	90.2

Goals and achievements

- Goal: to bootstrap the process of learning subjective expressions and extracting them from unannotated data
 - HP classifiers automatically identify subjective/objective sentences in unlabelled text
 - output of HP classifiers can be used to train an algorithm learning subjective extraction patterns
 - new patterns can be used to grow the training set
- extraction pattern techniques allows the learning of linguistically rich data
- a corpus-based subjectivity extraction method may be more effective, since some subjective expressions are not perceived as such by humans

Why it is important?

- There is not enough subjectivity labelled data to use in machine learning, so, even a small percentage of sentences labelled by a HP classifier is a huge improvement.
- The approach allows classifying sentences for subjectivity, not entire texts.
- It helps to expand the set of reliable subjectivity extraction patterns.

What we learned

- An objectivity classifier can identify objective sentences based on the absence of subjective markers.
- Human input in learning extraction patterns algorithm can be substituted with a conditional probability measure ranking.
- Surprisingly, expressions involving noun 'fact' are correlated with subjectivity.

Critique

- How many bootstrapping cycles did they run?
- What were the results with strict thresholds on later cycles?

Conclusions

- Lexicon induction is a process that creates large amounts of data from small amounts of reliably (most likely manually) annotated data.
- It can be used across domains, as the data being created is based on to patterns extracted from the data.
- Some applications include polarity and subjectivity classification.