Statistics-based Approaches to Lexical Semantics

Martin Thorsen Ranang
Department of Computer and Information Science (IDI)

Trial Lecture, February 5th 2010
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Lexical Semantics

— “The study of how and what the words of a language denote.” (Pustejovsky, 1998)
— lexical semantic relations like: synonymy, antonymy (“close vs. distant”), hypo-/hypernymy (“car vs. vehicle”)
— polysemy (lexical ambiguity)
— selectional restrictions: “Joe ate <...> in a hurry.”
— Typical resources:
  • Dictionaries, Machine Readable Dictionaries (MRDs) (Wilks et al., 1996)
  • Ontologies and Semantic Networks
The Distributional Hypothesis

— “You shall know a word by the company it keeps.” Firth (1957).

— “There is a positive relationship between the degree of synonymy (semantic similarity) existing between a pair of words and the degree to which their contexts are similar.” (Rubenstein and Goodenough, 1965)

— “The meaning of entities, and the meaning of grammatical relations among them, is related to the restriction of combinations of these entities relative to other entities.” (Harris, 1968)
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Example Areas

— Word Sense Disambiguation (WSD)
— Natural Language Understanding (NLU) and Text Interpretation (TI)
— Machine Translation (MT)
— Information Retrieval (IR)

What parts of Natural Language Processing (NLP) are not affected by Lexical Semantics?
Outline

Introduction
  What is Lexical Semantics?
  Natural Language Processing (NLP) Applications
  My PhD Research

Statistics-based Approaches to Lexical Semantics
  Word Sense Disambiguation (WSD)
  Vector Space Model (VSM)
  Dimensionality Reduction
  Ontology Merging and Alignment

Summary
My PhD Research

— Developed a method for automatically mapping words from languages other than English to concepts in the Princeton WordNet by Miller et al. (1990); Fellbaum (1998)
WordNet Example

- dwelling_n_1, home_n_2, domicile_n_2, abode_n_2, habitation_n_2, dwelling_house_n_1
- building_n_1, edifice_n_1
- house_n_1
- mansion_n_2, mansion_house_n_1, manse_n_1, hall_n_11, residence_n_4
- country_house_n_1
- palace_n_1, castle_n_1
- manor_n_1, manor_house_n_1
- chateau_n_1
- window_n_1
- windowpane_n_1, window_n_6
Why Statistics-based?

— Frequencies of actual language usage
— Adapts to changes of the above
— Well suited to provide generalizations and to summarize features of huge text corpora.

(Manning and Schütze, 1999)
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Word Sense Disambiguation (WSD)

- **Bass**
  - *Morone saxatilis*
  - Tones of low frequency
  - Marchione bass guitar
Usage Context

— “He fished for **bass** using scented attractants.”
— “Joe played the **bass** fluently, while George played the piano.”
— “When the neighbors play their music I can’t hear the tune but can hear the **bass** tones.”
Word Sense Disambiguation (WSD)

— Two main approaches:

Integrated approach: postponed until semantic analysis; elimination of ill-formed semantic representations

Stand-alone approach: independent of, and prior to compositional semantic analysis; more often statistics-based
Supervised learning

Training: sense-tagged corpus; naïve Bayesian classifiers; feature vectors; “sliding window”
Feature vectors represent local context, and may include words and POS.

Application: Use the trained classifier on unseen ambiguous words, given a local-context feature vector
Statistics-based Stand-alone Approaches II

Bootstrapping
small number of training instances used as seeds; classifier trained through supervised learning

Unsupervised disambiguation
sense-discrimination, not sense tagging; groups of similar words, based on their local-context

Dictionary-based approach
Count overlap between sliding window and dictionary definition of candidate senses.
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Vector Space Model (Salton, 1971)

Term Frequency:

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

Inverse Document Frequency:

$$\text{idf}_i = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

Vector elements:

$$\text{w}_{i,j} = \text{tf}_{i,j} \cdot \text{idf}_i$$

Weight vector for doc $d$:

$$\text{v}_d = [w_{1,d}, w_{2,d}, \ldots, w_{N,d}]^T$$

Importance of term $i$ to doc $j$

Common words are less descriptive
Vector Space Model

- Enables comparison with other documents, based on content.
- Does it really describe a document’s meaning?
- Restrictions?
Semantic Augmentation of the Vector Space Model

Several attempts to improve document retrieval efficiency by incorporating lexical semantic information:

— Moldovan and Mihalcea (2000)
— Buscaldi et al. (2005)

No, or small, improvements to IR; some improvement for document classification.
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Latent Semantic Analysis (LSA) / Indexing (LSI)

— *Discrete* entities are mapped onto a *continuous* vector space;
— the mapping is determined by *global correlation patterns*; and
— *Dimensionality reduction is an integral part of the process*

(Landauer and Dumais, 1997; Ando, 2000; Bellegarda, 2007)
Dimensionality Reduction

— Singular Value Decomposition

Quantitative evaluation of different semantic word space models: Van de Cruys (2010)
Outline

Introduction
What is Lexical Semantics?
Natural Language Processing (NLP) Applications
My PhD Research

Statistics-based Approaches to Lexical Semantics
Word Sense Disambiguation (WSD)
Vector Space Model (VSM)
Dimensionality Reduction
Ontology Merging and Alignment

Summary
Lacher and Groh (2001) used signature \textit{tfidf} vectors for computing similarity between two ontology nodes.
Summary

— Lexical semantics
— How this relates to my PhD research
— Examples of statistics-based approaches to Lexical Semantics, including:
  • different Word Sense Disambiguation techniques
  • semantic augmentation of the vector space model
  • how LSA/dimensionality reduction of vector spaces handles synonymy
  • how statistics-based similarity measures are used to align and merge ontologies
References I


References II


References IV


References V


