

Word Sense Disambiguation

Part I – Introduction

Alexander Fraser

CIS, LMU München

2015.12.01 WSD and MT

- Please read Navigli chapters 1 and 2 if you haven't already
- Next week: I might present an exercise on classification (using Wapiti)
 - How many people already did this in IE?
 - It is slightly different this year

Outline

- Introduction
 - Definitions
 - Ambiguity for Humans and Computers
 - Very Brief Historical Overview
 - Theoretical Connections
 - Practical Applications
- Methodology

Definitions

- *Word sense disambiguation* is the problem of selecting a sense for a word from a set of predefined possibilities.
 - Sense Inventory usually comes from a dictionary or thesaurus.
 - Knowledge intensive methods, supervised learning, and (sometimes) bootstrapping approaches
- *Word sense discrimination* is the problem of dividing the usages of a word into different meanings, without regard to any particular existing sense inventory.
 - Unsupervised techniques

Outline

- Definitions
- Ambiguity for Humans and Computers
- Very Brief Historical Overview
- Theoretical Connections
- Practical Applications

Computers versus Humans

- *Polysemy* – most words have many possible meanings.
- A computer program has no basis for knowing which one is appropriate, even if it is obvious to a human...
- Ambiguity is rarely a problem for humans in their day to day communication, except in extreme cases...

Ambiguity for Humans - Newspaper Headlines!

- DRUNK GETS NINE YEARS IN VIOLIN CASE
- FARMER BILL DIES IN HOUSE
- PROSTITUTES APPEAL TO POPE
- STOLEN PAINTING FOUND BY TREE
- RED TAPE HOLDS UP NEW BRIDGE
- DEER KILL 300,000
- RESIDENTS CAN DROP OFF TREES
- INCLUDE CHILDREN WHEN BAKING COOKIES
- MINERS REFUSE TO WORK AFTER DEATH

Ambiguity for a Computer

- The fisherman jumped off the **bank** and into the water.
- The **bank** down the street was robbed!
- Back in the day, we had an entire **bank** of computers devoted to this problem.
- The **bank** in that road is entirely too steep and is really dangerous.
- The plane took a **bank** to the left, and then headed off towards the mountains.

Outline

- Definitions
- Ambiguity for Humans and Computers
- **Very Brief Historical Overview**
- Theoretical Connections
- Practical Applications

Early Days of WSD

- Noted as problem for Machine Translation (Weaver, 1949)
 - A word can often only be translated if you know the specific sense intended (A bill in English could be a pico or a cuenta in Spanish)
- Bar-Hillel (1960) posed the following:
 - Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.
 - Is “pen” a writing instrument or an enclosure where children play?

Since then...

- 1970s - 1980s
 - Rule based systems
 - Rely on hand crafted knowledge sources
- 1990s
 - Corpus based approaches
 - Dependence on sense tagged text
 - (Ide and Veronis, 1998) overview history from early days to 1998.
- 2000s
 - Hybrid Systems
 - Minimizing or eliminating use of sense tagged text
 - Taking advantage of the Web

Outline

- Definitions
- Ambiguity for Humans and Computers
- Very Brief Historical Overview
- **Interdisciplinary Connections**
- Practical Applications

Interdisciplinary Connections

- Cognitive Science & Psychology
 - Quillian (1968), Collins and Loftus (1975) : spreading activation
 - Hirst (1987) developed marker passing model
- Linguistics
 - Fodor & Katz (1963) : selectional preferences
 - Resnik (1993) pursued statistically
- Philosophy of Language
 - Wittgenstein (1958): meaning as use
 - “For a *large* class of cases - though not for all - in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language.”

Outline

- Definitions
- Ambiguity for Humans and Computers
- Very Brief Historical Overview
- Theoretical Connections
- Practical Applications

Practical Applications

- Machine Translation
 - Translate “bill” from English to Spanish
 - Is it a “pico” or a “cuenta”?
 - Is it a bird jaw or an invoice?
- Information Retrieval
 - Find all Web Pages about “cricket”
 - The sport or the insect?
- Question Answering
 - What is George Miller’s position on gun control?
 - The psychologist or US congressman?
- Knowledge Acquisition
 - Add to KB: Herb Bergson is the mayor of Duluth.
 - Minnesota or Georgia?

References

- (Bar-Hillel, 1960) The Present Status of Automatic Translations of Languages. In Advances in Computers. Volume 1. Alt, F. (editor). Academic Press, New York, NY. pp 91-163.
- (Collins and Loftus, 1975) A Spreading Activation Theory of Semantic Memory. Psychological Review, (82) pp. 407-428.
- (Fodor and Katz, 1963) The structure of semantic theory. Language (39). pp 170-210.
- (Hirst, 1987) Semantic Interpretation and the Resolution of Ambiguity. Cambridge University Press.
- (Ide and Véronis, 1998) Word Sense Disambiguation: The State of the Art. Computational Linguistics (24) pp 1-40.
- (Quillian, 1968) Semantic Memory. In Semantic Information Processing. Minsky, M. (editor). The MIT Press, Cambridge, MA. pp. 227-270.
- (Resnik, 1993) Selection and Information: A Class-Based Approach to Lexical Relationships. Ph.D. Dissertation. University of Pennsylvania.
- (Weaver, 1949): Translation. In Machine Translation of Languages: fourteen essays. Locke, W.N. and Booth, A.D. (editors) The MIT Press, Cambridge, Mass. pp. 15-23.
- (Wittgenstein, 1958) Philosophical Investigations, 3rd edition. Translated by G.E.M. Anscombe. Macmillan Publishing Co., New York.

Outline

- Introduction
- Methodology
 - General considerations
 - All-words disambiguation
 - Targeted-words disambiguation
 - Word sense discrimination, sense discovery
 - Evaluation (granularity, scoring)

Overview of the Problem

- Many words have several meanings (homonymy / polysemy)

–Ex: “**chair**” – furniture or person

–Ex: “**child**” – young person or human offspring

- Determine which sense of a word is used in a specific sentence

- **Note:**

- often, the different senses of a word are closely related

- Ex: **title** – right of legal ownership
– document that is evidence of the legal ownership,

- sometimes, several senses can be “activated” in a single context (co-activation)

- Ex: “*This could bring competition to the trade*”
competition: – the act of competing
– the people who are competing

Word Senses

- The *meaning* of a word in a given context
- Word sense representations
 - With respect to a dictionary

chair = a seat for one person, with a support for the back; "he put his coat over the back of the chair and sat down"

chair = the position of professor; "he was awarded an endowed chair in economics"

- With respect to the translation in a second language

chair = chaise

chair = directeur

- With respect to the context where it occurs (discrimination)

"Sit on a *chair*" "Take a seat on this *chair*"

"The *chair* of the Math Department" "The *chair* of the meeting"

Approaches to Word Sense Disambiguation

- Knowledge-Based Disambiguation
 - use of external lexical resources such as dictionaries and thesauri
 - discourse properties
- Supervised Disambiguation
 - based on a labeled training set
 - the learning system has:
 - a training set of feature-encoded inputs AND
 - their appropriate sense label (category)
- Unsupervised Disambiguation
 - based on unlabeled corpora
 - The learning system has:
 - a training set of feature-encoded inputs BUT
 - NOT their appropriate sense label (category)

All Words Word Sense Disambiguation

- Attempt to disambiguate all open-class words in a text

“He put his suit over the back of the chair”

- Knowledge-based approaches
- Use information from dictionaries
 - Definitions / Examples for each meaning
 - Find similarity between definitions and current context
- Position in a semantic network
 - Find that “table” is closer to “chair/furniture” than to “chair/person”
- Use discourse properties
 - A word exhibits the same sense in a discourse / in a collocation

All Words Word Sense Disambiguation

- Minimally supervised approaches
 - Learn to disambiguate words using small annotated corpora
 - E.g. SemCor – corpus where all open class words are disambiguated
 - 200,000 running words
- Most frequent sense

Targeted Word Sense Disambiguation

- Disambiguate one target word

“Take a seat on this **chair**”

“The **chair** of the Math Department”

- WSD is viewed as a typical classification problem
 - use machine learning techniques to train a system
- Training:
 - Corpus of occurrences of the target word, each occurrence annotated with appropriate sense
 - Build feature vectors:
 - a vector of relevant linguistic features that represents the context (ex: a window of words around the target word)
- Disambiguation:
 - Disambiguate the target word in new unseen text

Targeted Word Sense Disambiguation

- Take a window of n word around the target word
- Encode information about the words around the target word
 - typical features include: words, root forms, POS tags, frequency, ...
 - An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
 - Surrounding context (local features)
 - [(guitar, NN1), (and, CJC), (player, NN1), (stand, VVB)]
 - Frequent co-occurring words (topical features)
 - [*fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band*]
 - [0,0,0,1,0,0,0,0,0,0,1,0]
 - Other features:
 - [followed by "player", contains "show" in the sentence, ...]
 - [yes, no, ...]

Unsupervised Disambiguation

- Disambiguate word senses:
 - without supporting tools such as dictionaries and thesauri
 - without a labeled training text
- Without such resources, word senses are not *labeled*
 - We cannot say “**chair/furniture**” or “**chair/person**”
- We can:
 - Cluster/group the contexts of an ambiguous word into a number of groups
 - *Discriminate* between these groups without actually labeling them

Unsupervised Disambiguation

- Hypothesis: same senses of words will have similar neighboring words
- Disambiguation algorithm
 - Identify context vectors corresponding to all occurrences of a particular word
 - Partition them into regions of high density
 - Assign a sense to each such region

“Sit on a **chair**”

“Take a seat on this **chair**”

“The **chair** of the Math Department”

“The **chair** of the meeting”

Evaluating Word Sense Disambiguation

- Metrics:
 - Precision = percentage of words that are tagged correctly, out of the words addressed by the system
 - Recall = percentage of words that are tagged correctly, out of all words in the test set

– Example

- | | |
|------------------------------------|------------------------------|
| • Test set of 100 words | Precision = $50 / 75 = 0.66$ |
| • System attempts 75 words | Recall = $50 / 100 = 0.50$ |
| • Words correctly disambiguated 50 | |

- Special tags are possible:
 - Unknown
 - Proper noun
 - Multiple senses
- Compare to a gold standard
 - SEMCOR corpus, SENSEVAL corpus, ...

Evaluating Word Sense Disambiguation

- Difficulty in evaluation:
 - Nature of the senses to distinguish has a huge impact on results
- Coarse versus fine-grained sense distinction

chair = a **seat** for one person, with a support for the back; "he put his coat over the back of the chair and sat down"

chair = the position of **professor**; "he was awarded an endowed chair in economics"

bank = a **financial institution** that accepts deposits and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the mortgage on my home"

bank = a **building** in which commercial banking is transacted; "the bank is on the corner of Nassau and Witherspoon"

- Sense maps
 - Cluster similar senses
 - Allow for both fine-grained and coarse-grained evaluation

Bounds on Performance

- Upper and Lower Bounds on Performance:
 - Measure of how well an algorithm performs relative to the difficulty of the task.
- Upper Bound:
 - Human performance
 - Around 97%-99% with few and clearly distinct senses
 - Inter-judge agreement:
 - With words with clear & distinct senses – 95% and up
 - With polysemous words with related senses – 65% – 70%
- Lower Bound (or baseline):
 - The assignment of a random sense / the most frequent sense
 - 90% is excellent for a word with 2 equiprobable senses
 - 90% is trivial for a word with 2 senses with probability ratios of 9 to 1

References

- (Gale, Church and Yarowsky 1992) Gale, W., Church, K., and Yarowsky, D. *Estimating upper and lower bounds on the performance of word-sense disambiguation programs* ACL 1992.
- (Miller et. al., 1994) Miller, G., Chodorow, M., Landes, S., Leacock, C., and Thomas, R. *Using a semantic concordance for sense identification.* ARPA Workshop 1994.
- (Miller, 1995) Miller, G. Wordnet: A lexical database. ACM, 38(11) 1995.
- (Senseval) Senseval evaluation exercises <http://www.senseval.org>

Slide sources

- Many slides today are from a Mihalcea and Pedersen tutorial in 2005, given at several locations including ACL and AAI

Outlook

- Navigli is useful background material for the literature review
Referat subjects
 - I do not expect the Referats to parallel Navigli though (too much material in Navigli!)
 - If you have any questions about this, please ask or send me an email
- Please read Navigli Sections 1 and 2 (first 15 pages) for next week
 - If you have time, also look at Section 3 briefly
 - (I will ask you to read Sections 3 and 5 for the following week, see the web page)
- Check the web for where we will meet next week!

- Thanks for your attention!