A SEMANTIC UNSUPERVISED LEARNING APPROACH TO WORD SENSE DISAMBIGUATION

Dissertation Presentation

April 4, 2018

Dian I. Martin

Presentation Overview

- Background
- LSA-WSD Approach
- Word Importance in a Sentence
- Automatic Word Sense Induction
- Automatic Word Sense Disambiguation
- Future Research

THE PROBLEM

WORD SENSE DISAMBIGUATION (WSD): WHICH SENSE OF A WORD IS BEING USED IN A GIVEN CONTEXT?

Mowing the lawn was a hard task for the little boy.

The boxer threw a hard left to the chin of his opponent.

WSD

Multiple Meanings = Different Word Senses

All Word Senses = Word Definition

Two WSD Tasks

Sense Discovery

Determine all the senses for a target word, *word A.*

Sense Identification

Determine which sense a target word, *word A*, is being used in a particular context.

WSD Approaches

A Priori Knowledge

- Dictionary-based or Knowledge-based methods
- Supervised methods
- Minimally supervised methods

No A Priori Knowledge

- Unsupervised methods

WSD Applications

To name a few …

- Any NLP application
- Information retrieval
- Text mining
- Information Extraction
- Lexicography
- Educational applications
- Analysis of the learning system

LSA-WSD APPROACH

An unsupervised algorithm for automated WSD

Latent Semantic Analysis

Unsupervised Learning Algorithm

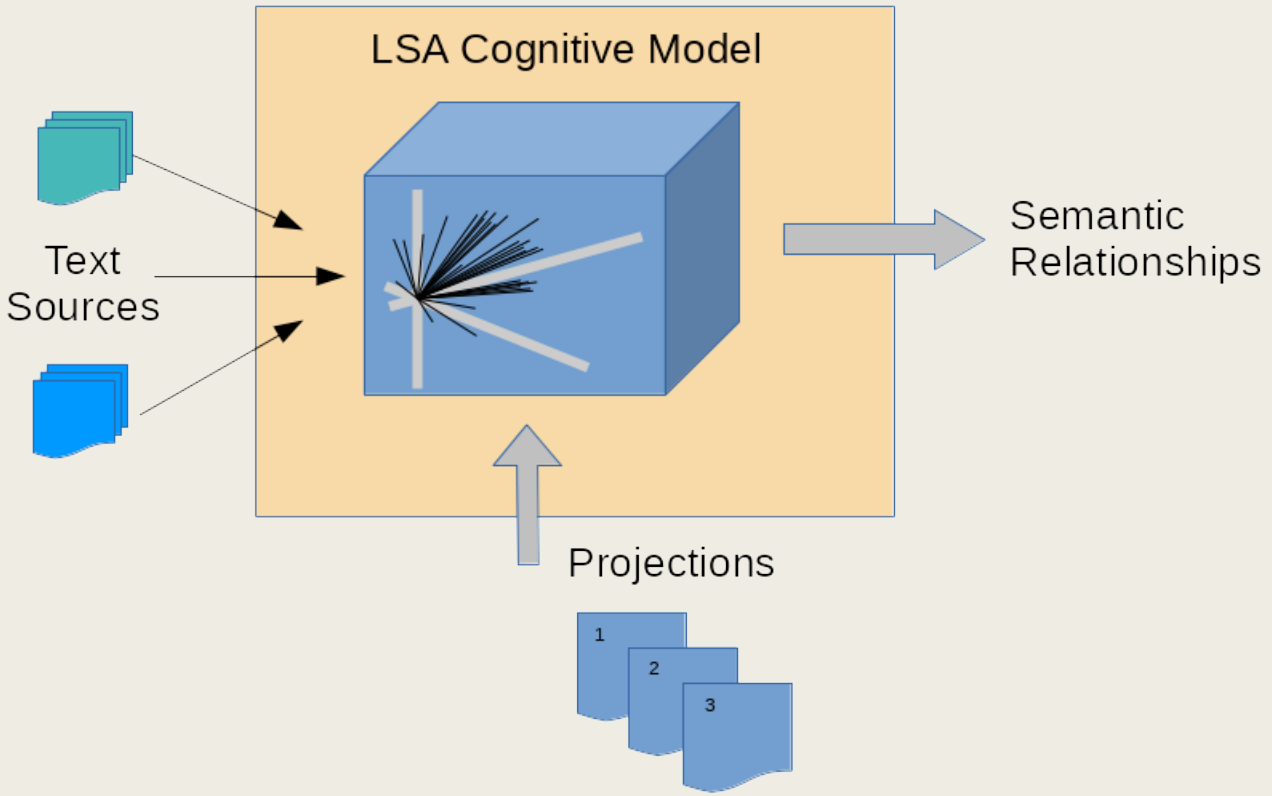
- Represents a cognitive model
- Mimics human learning
- Many applications where LSA-based learning system (LS) has simulated human knowledge
  - *Essay grading*
  - *Interactive auto-tutors*
  - *Synonym tests*
  - *Text comprehension*
  - *Summarization feedback*

Compositionality Constraint

The meaning of a document is the sum of the meaning of the terms that it contains.

The meaning of a term is defined by all the contexts in which it does and does not appear.

LSA-Based Learning System



Latent Semantic Analysis (LSA)

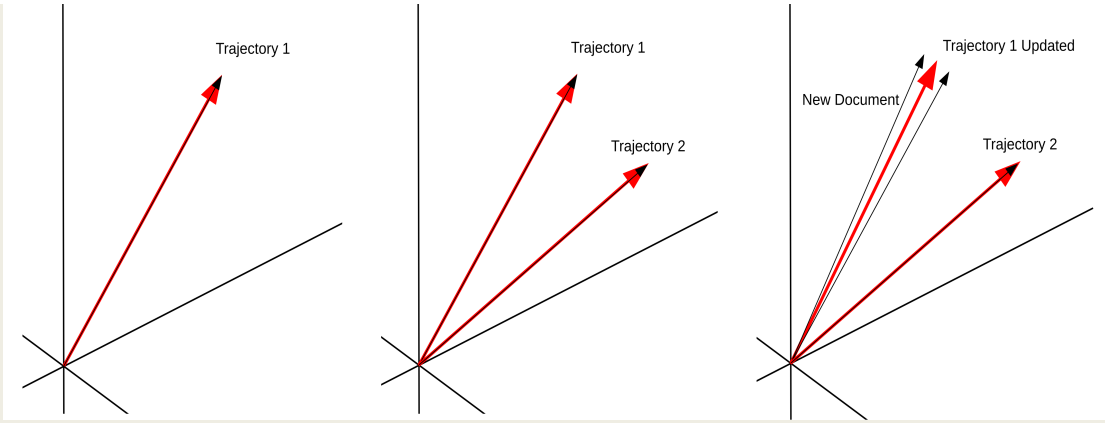
- Text => Term x Document (TD) matrix
- TD matrix => Weighted TD matrix
- Weighted TD matrix => Singular Value Decomposition (SVD)
- SVD => Term vectors and Document vectors
- Term vectors => Projections
- Vector comparisons => Semantic Similarity

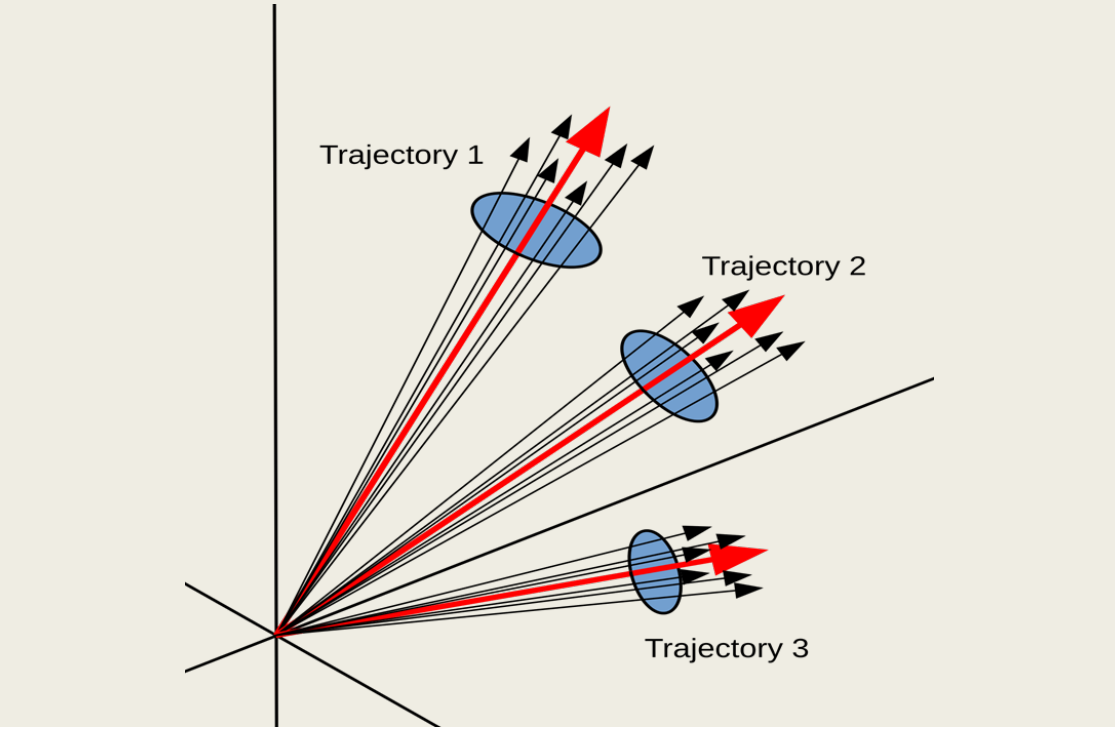
LSA-WSD Approach: Sense Discovery

Semantic Mean Clustering (SMC)

Sentence clustering (sentclusters)

Synonym clustering (synclusters)





LSA-WSD Approach: Sense Identification

For given target word and particular context:

- Map sentence or context into LSA semantic space
- Determine closest cluster
- Closest cluster identifies the sense

Document Collections

Document Set	# Documents	# Sentences	# Unique Words
Grade Level A 150K	162777	1955690	141252
Grade Level B 150K	162845	1958077	141774
Grade Level A 200K	209365	2503308	162295
Grade Level B 200K	209423	2503697	162308
Grade Level Unique A 200K	196261	2309345	164940
Grade Level Unique B 200K	196262	2306918	164975
Grade Level A 250K	259847	3099118	182492
Grade Level B 250K	260059	3097901	182311
News A 200K	200000	2782399	254236
News B 200K	200000	2781141	255640

WORD IMPORTANCE IN A SENTENCE

Finding adequate contexts to use in sentence clustering for deriving senses for a target word.

Word Importance
3 Questions

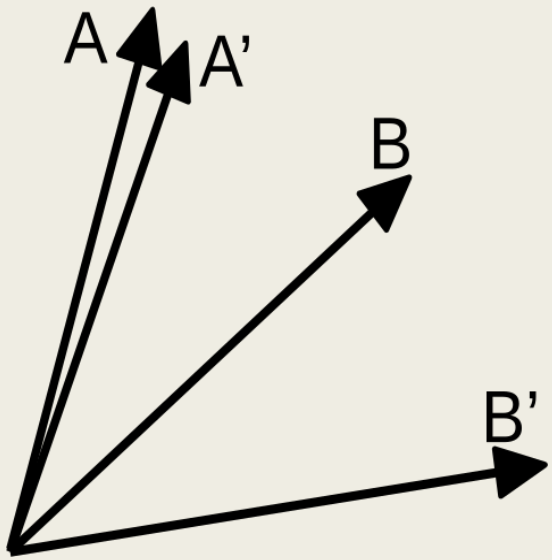
- Does sentence length have an impact on the importance of a word in a sentence?
- Are there specific words that never contribute or always contribute to the meaning of a sentence?
- How often do sentences have important words, ones that contribute notably to the meaning of the sentence?

Cosine Impact Value (CIV)

Determine impact of a word on the meaning of a sentence:

- Project the sentences with and without target word into the LSA semantic space
- Compute cosine similarity between them (CIV)

CIV has inverse relationship with impact of a word on the meaning of a sentence

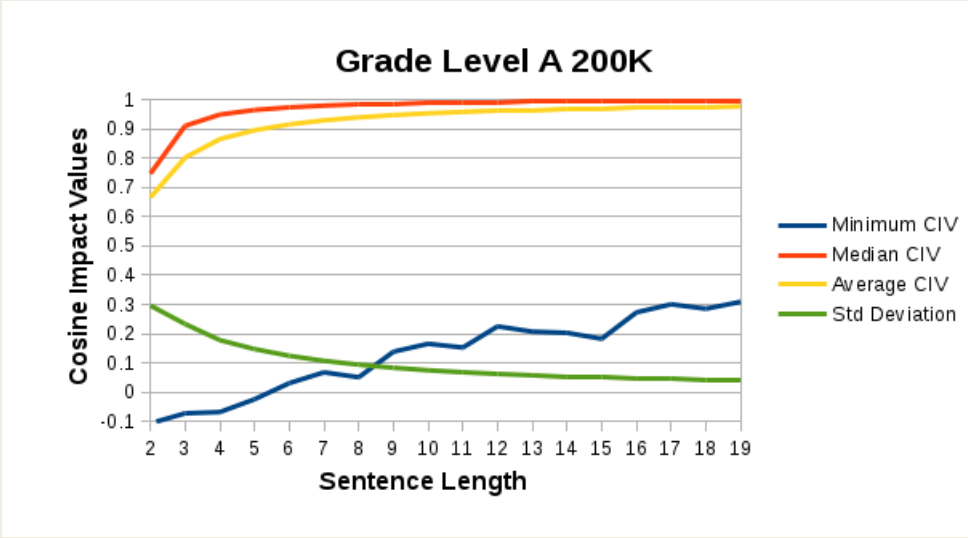


Cosine Impact Values Calculated

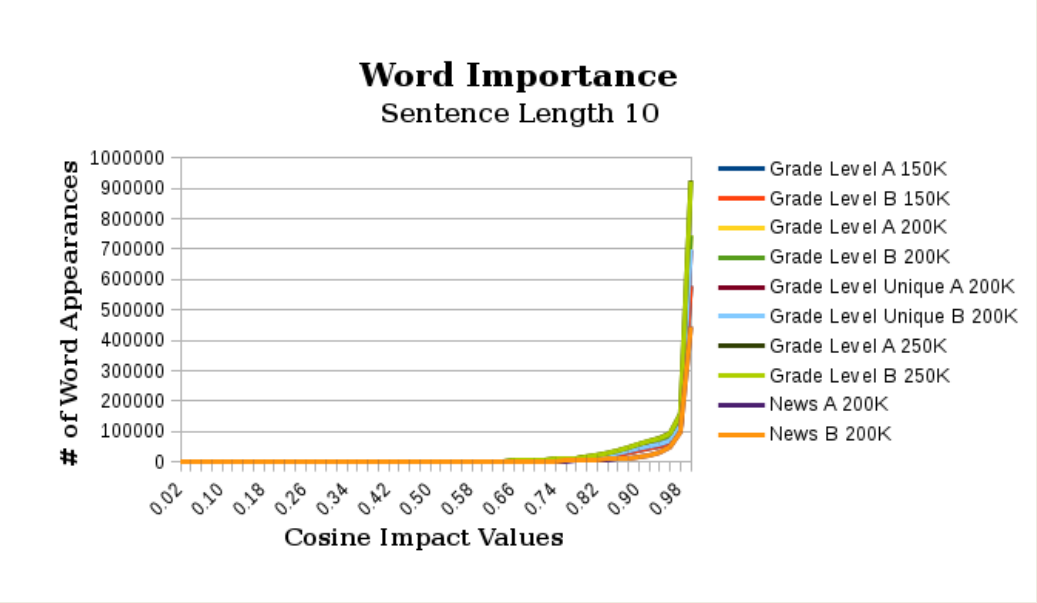
To identify a general indicator of word importance, consider:

- Sentences of lengths two or greater
- Sentences of lengths 2 to 19 for the grade level document set
- Sentences of lengths 10 to 32 for the news document set
- Each word in each of these sentences
- Each of the **234,568,429** CIVs

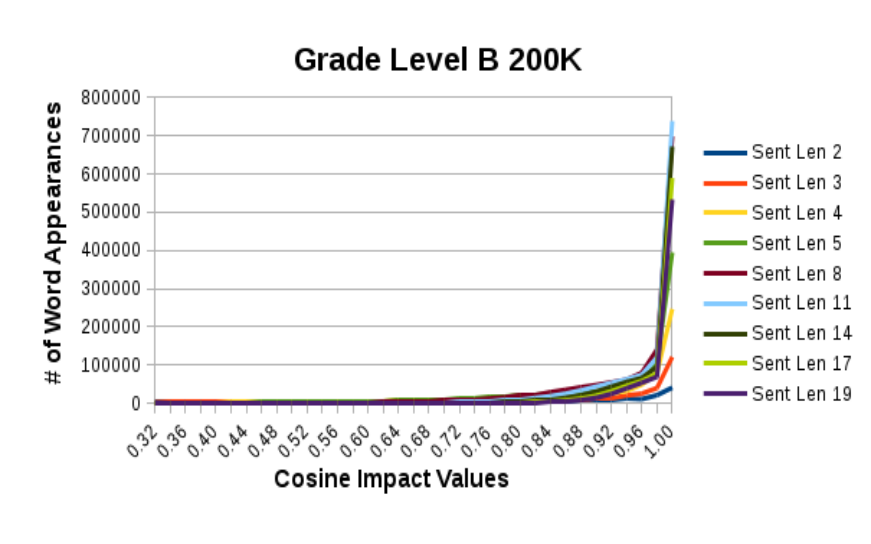
Effect of Sentence Length on Word Importance

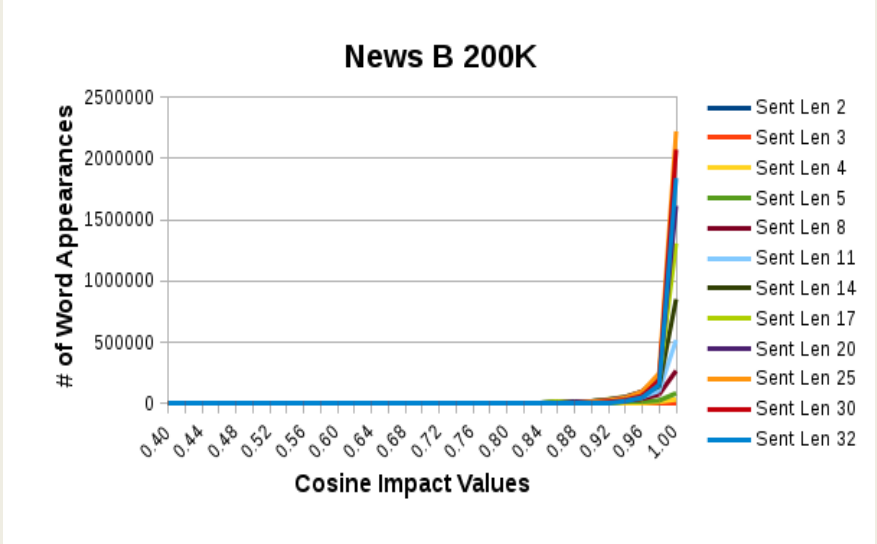


Distribution of CIVs for Sentence Length Ten

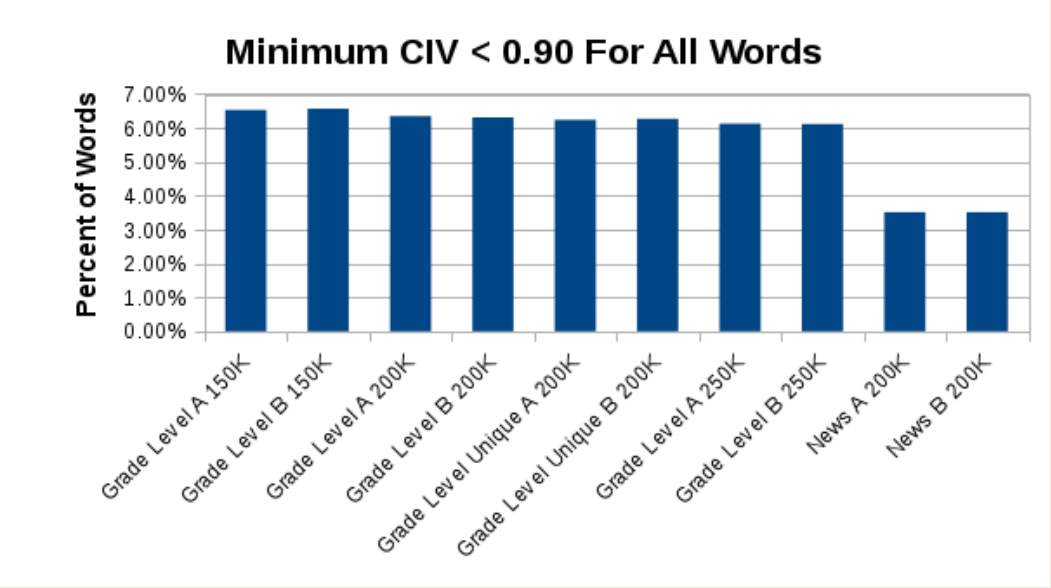


Distribution of CIVs for Different Sentence Lengths for a Document Collection

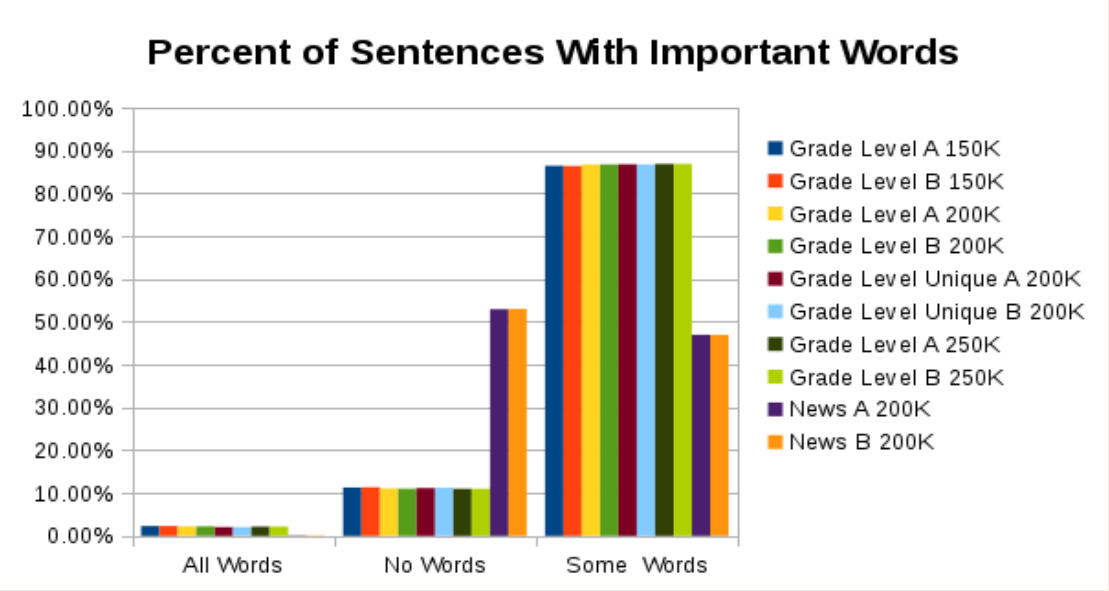




Word Characteristics for Word Importance in a Sentence



Appearance of Important Words in Sentences



Word Importance Observations

- CIV of 0.90 determines individual importance for a word on the meaning of a sentence
- Few words in a corpus, less than 7%, are important to one or more sentences in which they appear
- Words that are always important to the meaning of the sentences in which they are appear are nouns
- Majority of sentences do contain at least one important word
- Sentences of length four or less generally contain all important words
- As sentence length increases, individual word importance decreases
- Corpus size and content did not have an effect on word importance measures

WORD SENSE INDUCTION

Step 1 in LSA-WSD approach:
The automatic discovery of the possible word senses for a given word.

Creating the Learning System (LS)

- Precursor to Word Sense Induction (WSI)
- WSI dependent on the knowledge contained in LS
- Just as humans determination of senses is different so will senses of WSI systems
- LSA-based LS beneficial for deriving senses indicative a particular learner or domain
- Used two document collections of 200K documents from each source in WSI experiments

Clustering Expectations

- Items would be evenly distributed across individual clusters
- Outliers an anomaly – obscure sense or noise?
- Singleton clusters not desirable
- All items in one cluster – one sense discovered or multi-sense?

Target Words

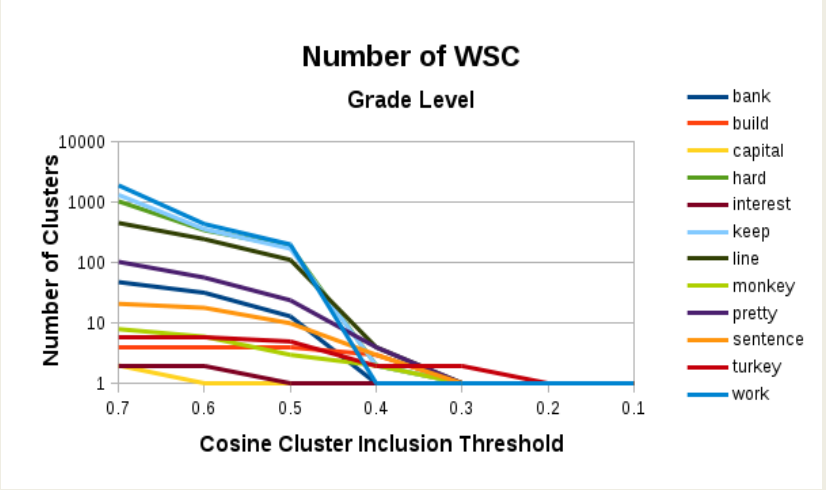
bank	interest	pretty
batch	keep	raise
build	line	sentence
capital	masterpiece	serve
enjoy	monkey	turkey
hard	palm	work

Sense Discovery with Sentclusters

WSI Experiments using sentclustering (cluster sentences with SMC) for a target word:

1. *All sentences vs. important word set*
2. *Determining appropriate clusters*
3. *Larger grade level LS*
4. *Different source for LS and sentences*
5. *Augmented sentence vector*
6. *Sentence with target word removed*

Problem: Multi-sense cluster



Senses Induced using Sentclusters for the Target Word bank

WSC #	# in Cluster	Example sentences
1	1	Bits of broken shell lie on the sunny bank.
2	2	The bank was held up. The bank held Arncaster's mortgage.
3	1	She retrieved the shopping bags and hurried to the bottle bank.
4	1	They walked from bank to bank.
5	74	The Brickster was a bank robber. In the bank, Mark goes up to a teller. In my bank, one quarter goes CLANK. "My piggy bank," Slither said. There's one hiding in the bushes on the bank. She does a perfect cannonball from the mossy bank. Sunny squinted, searching her memory bank.

Sense Discovery with Synclusters

- Examine meaning of target word by examining words close to it within the LSA-based learning system
- Embedded in the term vector is all the senses of the term
- Separate senses by clustering synonyms based on cosine similarity
- Top *k* terms closest to target word are clustered by SMC
- Closest word to centroid of word sense clusters (WSC) is the identifier for the cluster

Top 100 Closest Words to the Target Word bank

Terms 1-12	13-24	25-36	37-48	49-60	61-72	73-84	85-96	97-100
bank	current	boatmen	deposit	monongahela	riffles	wading	sandbars	uminpeachable
banks	raft	canoe	loan	paddle	snags	riverside	portage	potomac
downstream	tributary	steamboat	willows	reeds	money	narmada	bills	marshy
riverbank	barge	footbridge	nashua	cash	shallows	rhadamnanthus	swift	spanned
upstream	steamboats	flood	riverbed	ferryman	creek	cocytus	sawmills	
river	muddy	ferrymen	barges	boatman	conononka	radarscope	padding	
rapids	eddies	dammed	paddled	riverbanks	savings	insecttortured	mississippi	
downriver	bluffs	bottomlands	tributaries	dams	flowing	shallow	damming	
dam	levee	sandbar	thames	rafts	bottomland	waterfall	meander	
upriver	gorge	flatboats	midstream	headwaters	creeks	waded	murky	
bridge	flatboat	robb	canal	silt	watercourse	overhanging	platte	
flowed	bend	stream	countercurrents	poling	poled	crossing	riverboat	

WSCs Discovered Using Synclusters for the Target Word bank

Grade Level LS

WSC	# in WSC	WSC Descriptor	Next closest words	Cosine between *bank* and WSC centroid
WSC 1	93	downstream	river, rapids, upstream, riverbank	0.78
WSC 2	6	money	bills, cash, savings, loan	0.51

News LS

WSC	# in WSC	WSC Descriptor	Next closest words	Cosine between *bank* and WSC centroid
WSC 1	88	banks	banking, deposits, bankers, lending	0.78
WSC 2	9	rates	interest, reserve, mortgage, discount	0.36
WSC 3	1	finance		0.21
WSC 4	1	manages		0.21

WSCs Discovered Using Synclusters

- Target word: sentence

Grade Level LS => 1 WSC => spelling

News LS => 1 WSC => prison

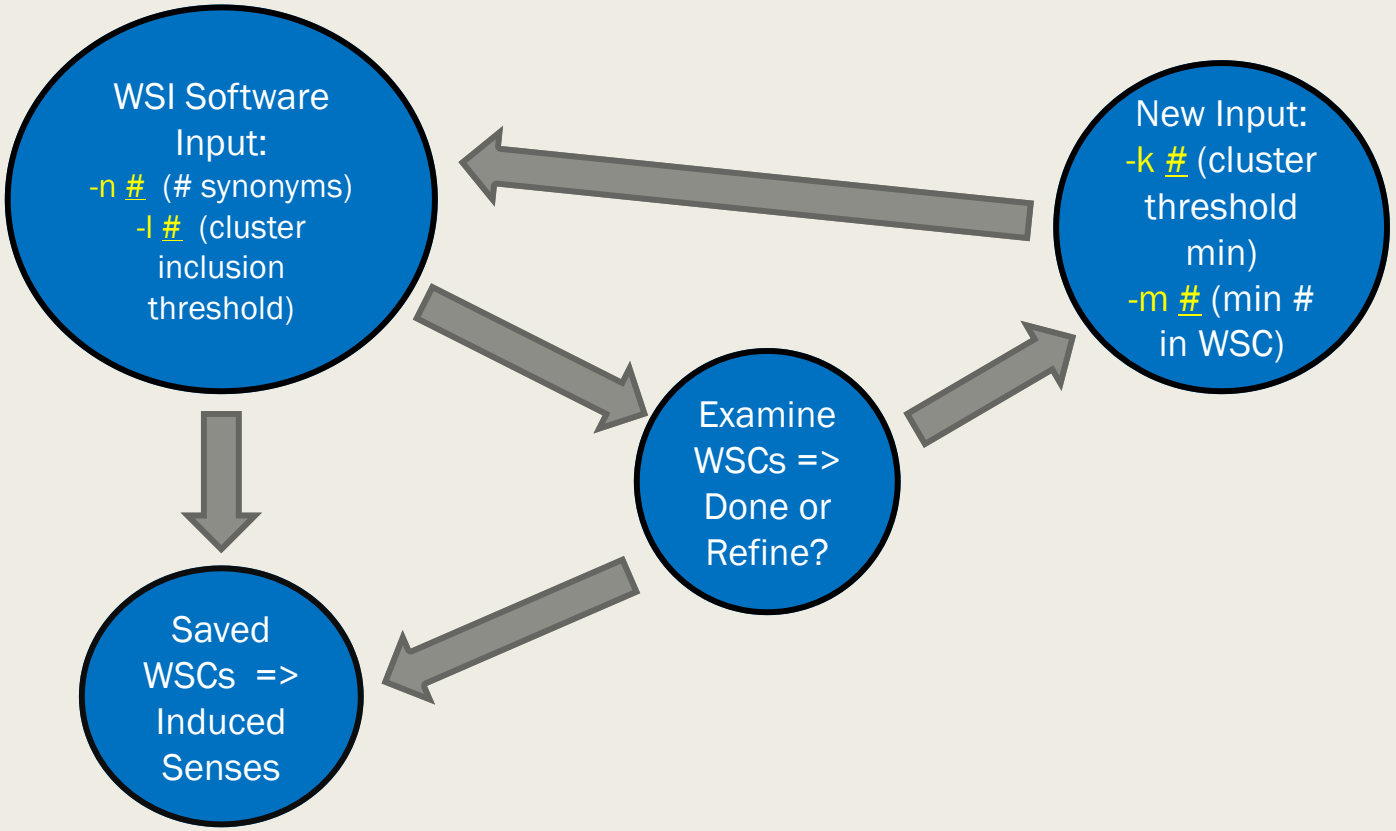
- Target word: raise

Grade Level Learning System			News Learning System		
WSC Label	# in WSC	Cosine to Centroid	WSC Label	# in WSC	Cosine to Centroid
money	71	0.57	increases	11	0.58
raised	2	0.55	funds	4	0.50
crops	6	0.50	tax	26	0.48
support	6	0.37	interest	4	0.38

Synclusters

- Produced reasonable results
- Candidate WSCs should have cosine similarity between centroid and target word > 0.35
- Can be applied to any word in document collection
- Allows for user refinement of candidate WSCs
- Two learning systems not equal in their representation of knowledge

User Input to Synclustering



WORD SENSE IDENTIFICATION

Step 2 in LSA-WSD approach:
Determine in which sense a target word is used in a particular context.

Two Methods to Determine Correct Sense for a Sentence

Synonym Replacement (SR) Method

SentA = Original Sentence

For each WSC derived from synclustering:

1. *SentB = Original Sentence with target word replaced with WSC identifier*
2. *Project SentA and SentB into the LSA semantic space*
3. *Compute cosine between projections*

Determine highest cosine similarity and corresponding cluster is the identified word sense

Context Comparison (CC) Method

SentA = Original Sentence

Remove target word from *SentA*

Project *SentA* into the LSA semantic space

For each WSC derived from synclustering:

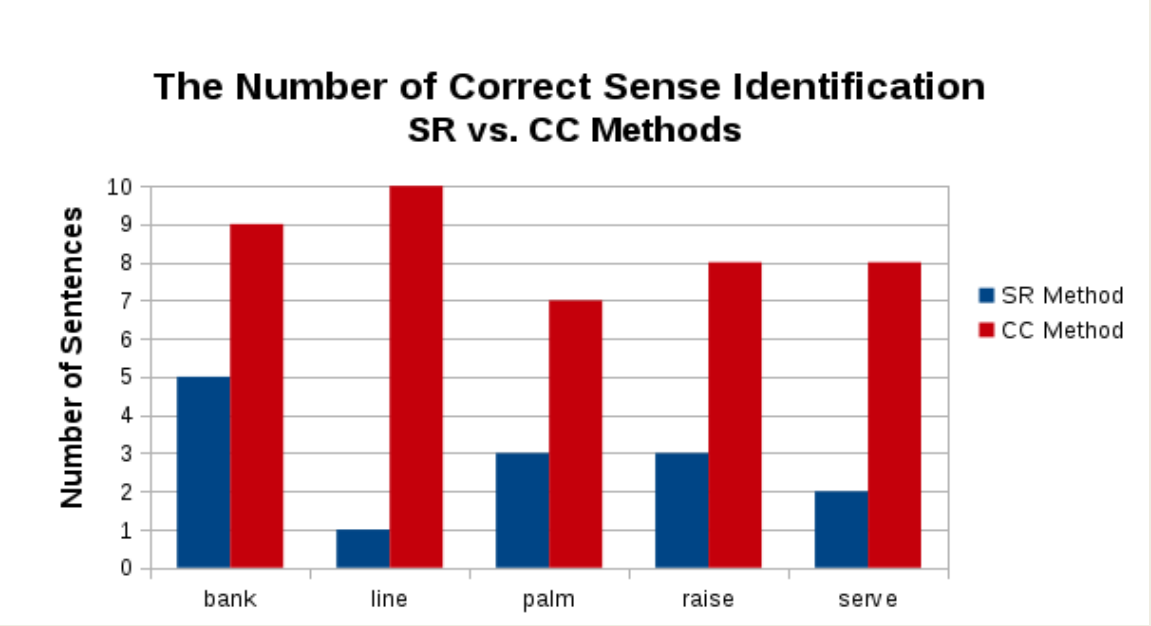
Compute cosine between projection and centroid vector

Determine highest cosine similarity and corresponding cluster is the identified word sense

Test Sentences for WSD Task for the Word *line*

Annotated WSC Label	Sentences Using *line* in this Sense
zone (line marked on a field or court)	Jackie stepped to the line and dropped in both foul shots. Jim plowed forward to stop the quarterback from reaching the goal line.
assonance (line of poetry)	The pattern of stressed and unstressed syllables discernible in a line of poetry has been analyzed in order to determine whether the line follows an iambic or a dactylic or an anapestic metrical arrangement. Each stanza has eight lines.
bait (line on a fishing rod)	He reeled in the line and bent the pole. He cast out his line.
horizontal (mathematical term for a line or lines in particular directions)	The curved line represents the variation of voltage in the signal. Draw a horizontal line above the vertical line.
ahead (line marking the starting point or finishing point in a race)	Matthew dashed across the finish line. I crossed the finish line, jogged to a stop, and kneeled on the cinders, breathing deeply.
Different sense	The workers would build them on a moving assembly line.
Ambiguous sense	Hold the line a minute, Diane.

Number of Times Word Senses Correctly Identified in Context



Outcomes of Automatic Disambiguation

- CC method performed the best at identifying the sense of a target word in a given context
- CC method identified the correct sense 84% of the time
- Cosine similarity measure produced by the CC method appears to provide information about the confidence of the sense identification

CONCLUSION AND FUTURE RESEARCH

The LSA-WSD approach to automated WSI and WSD is a SUCCESS!

What next?

Word Importance in a Sentence

Conclusions

- Importance of a word to the meaning of a sentence can be computed by the CIV
- Words with CIV less than 0.90 are primary contributors to meaning of sentence
- Most words had a small impact on meaning of a sentence
- Majority of the sentences in the corpora contained at least one important word
- All words contribute to the meaning of sentence (compositionality constraint)
- Corpus size and content did not have an observable effect on word importance

Future research

- First of its kind using the LSA-based LS
- Can be extended to apply sentence importance within a document, or sub-part of text
- Apply word importance to other applications

  For example: Educational settings

Automated Word Sense Induction

Conclusions

- Sentclustering was able to discover multiple senses for a target word, but there was an existence of a multi-sense cluster
- Synclustering produced reasonable results for sense discovery
  - *Candidate WSCs should have a cosine similarity between centroid and target word > 0.35*
  - *Grade level LS produced more broad-based results*
  - *Senses can be induced for any word*
  - *Can be a semi-supervised system*

Future Research

- Refine sentclustering:
  - *Use human annotated sentences*
  - *Secondary processing of multi-sense clusters*
- More experimentation using synclustering:
  - *Induce senses for more words*
  - *Define the number of synonyms to use and the optimal cluster inclusion threshold*

Automated Word Sense Disambiguation

Conclusions

- Two methods were considered
- The CC method produced more accurate results, identifying the correct sense for a target word within context sentences 84% of the time
- Cosine similarity measure gives a degree of confidence in the CC method

Future Research

- Other methods can be tried
- More sentences need to be tested for further validation and generalization of results

Broader Implications of This Research

Evaluation of the LS

LSA-WSD system can

- Indicate how well a LS knows the senses of a word
- Help to define the body of knowledge and use of language captured in the LS
- Guide the creation of a LS for use in different applications

Automation of the Tasks

LSA-WSD system can

- Produce a viable unsupervised, automated system for both WSI and WSD tasks
- Be adapted to different languages or updated when language changes
- Be used for different applications

THANK YOU!

Questions?

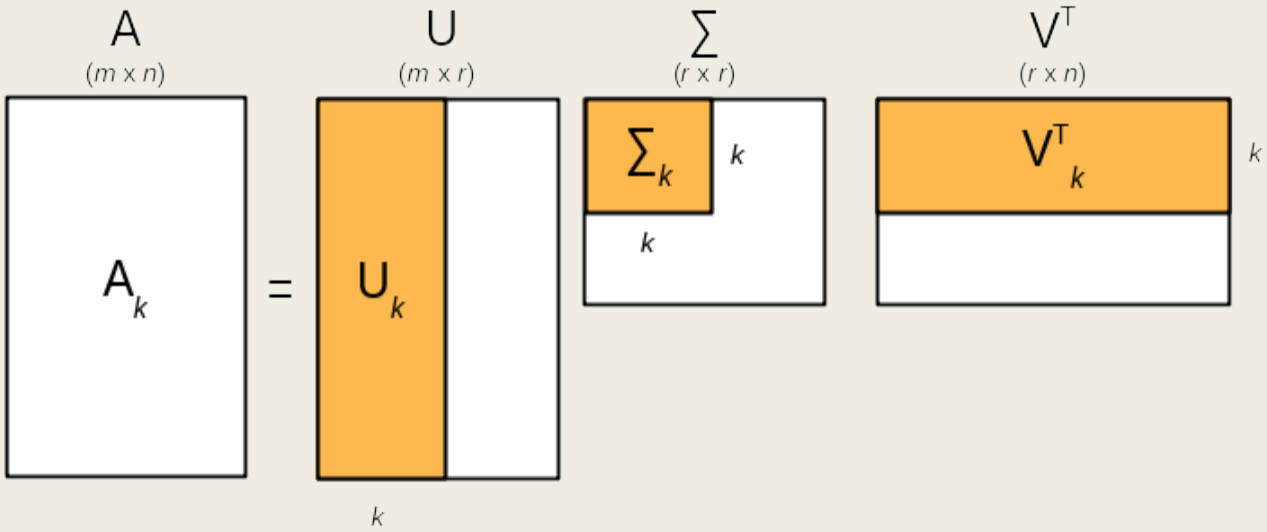
WSD History

- 1940s: Distinct task in machine translation
- 1950s: Turing's revelation
- 1960s: Bar-Hillel's assessment, progress stalled
- 1970s: Automated WSD in artificial intelligence (AI) approaches
- 1980s: Turning point for WSD: word experts, Lesk algorithm, polarized words
- 1990s: Development of WordNet, Senseval workshop, statistical revolution
- 2000s; Development of many different approaches

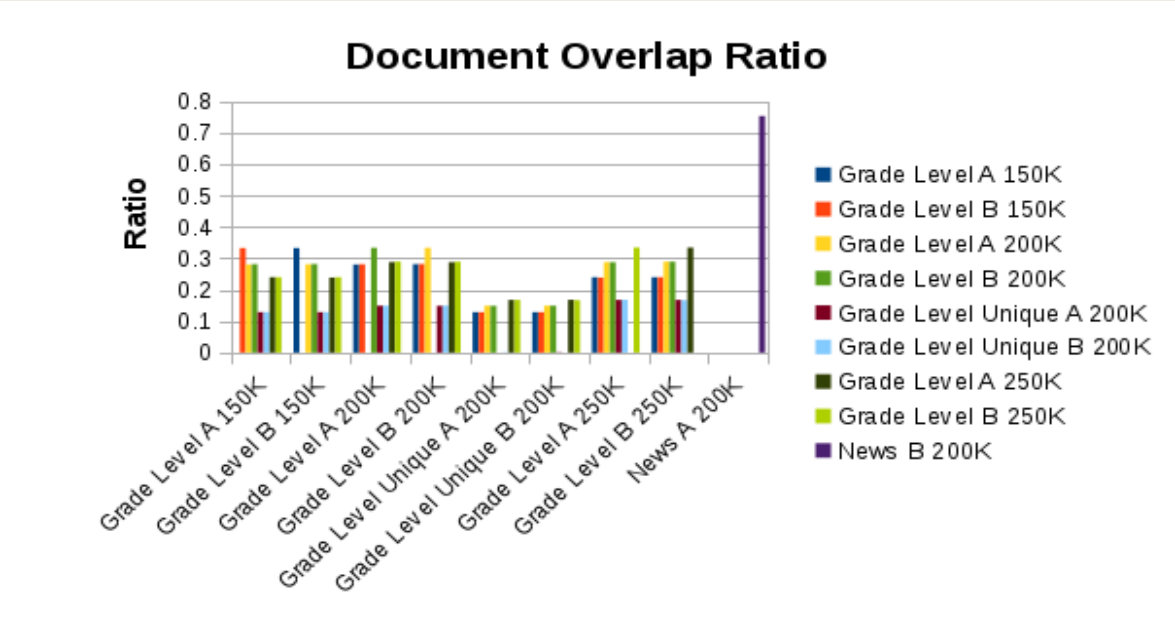
Local and Global Weighting

- Local functions: Normalize term frequency (*tf*) within each context
- Global functions: Normalize the global term frequency (*gf*) across all contexts
- Generally *log-entropy* applied within LSA-based learning system
- Log ($\log(tf + 1)$) => approximates the simple growth of standard learning
- Entropy ($(1 + \sum \frac{p_{ij}\ log_2 p_{ij}}{log_2\ (n)}$, where $p_{ij} = \frac{tf_{ij}}{gf_i}$, for term i and document j)) => estimates the degree to which observing a term indicates the context in which it appears

Singular Value Decomposition



Document Overlap Ratio



Sentence Length Comparisons Between Corpora