A Preliminary Functional Analysis of Memory in the Word Sense Disambiguation Task

Nate Derbinsky & John E. Laird
University of Michigan

Long-Term Memory Systems

Class of mechanism to…

- **Encode** experience
- **Store** internally
- Support **retrieval**

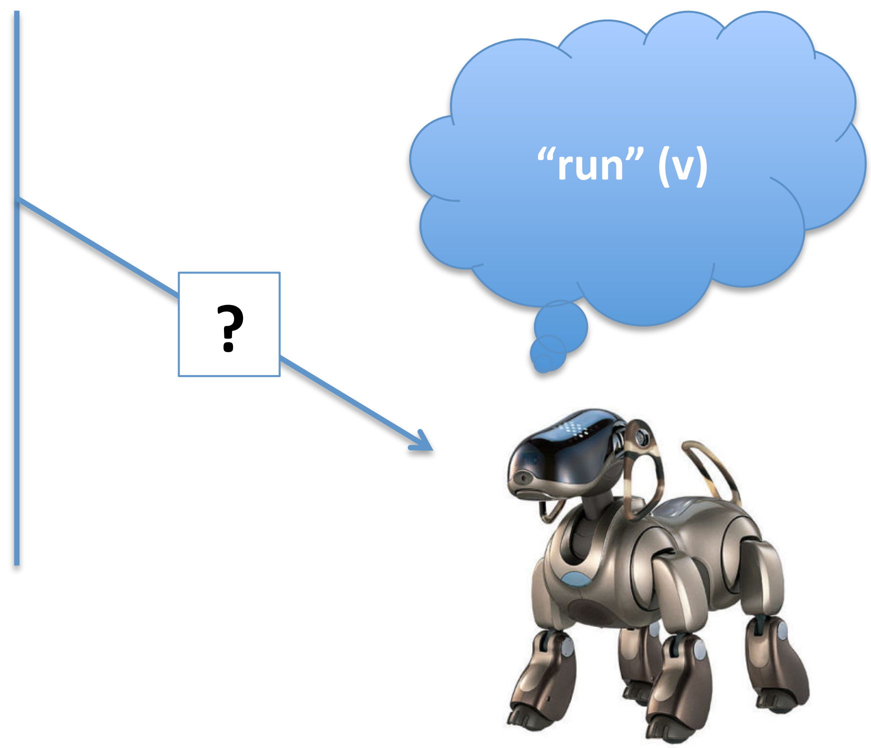
Our **goal**: develop and evaluate a suite of mechanisms that are <u>effective</u> and <u>efficient</u> across a variety of tasks.

Ambiguous Cues

Long-Term Memory

…

Agent



Problem: Supporting Ambiguous Cues

Given…

- large store of knowledge;
- and a cue that pertains to multiple previously encoded memories…

support retrievals that are **effective** and **efficient** across a variety of tasks.

Prior Work: Base-level Activation

Rational analysis posited that human memory optimizes over history of past memory access

- Anderson & Schooler, 1991

Current methods do not scale to large stores of knowledge

- Douglass, Ball, & Rodgers, 2009
- Derbinsky, Laird, & Smith, 2010

This Work

Task analysis. Word sense disambiguation and the SemCor data set

Effectiveness. Demonstrate the functional benefit of biasing retrievals towards past memory access

Efficiency. Preliminary evaluation of heuristics to support historically biased retrievals

Word Sense Disambiguation (WSD)

Task. Computationally identify the meaning of words in context.

Our focus is not language processing, therefore we simplify the problem formulation.

Our WSD Formulation

<u>Input</u>

- Sequence of sentences (sequence of words)
- Each word specified as lexical string and part-of-speech (noun, verb, adjective, adverb)

<u>Given</u>

- Machine Readable Dictionary (MRD): for each word…
  - Set of available senses: for each sense…
    - Definition
    - Tag frequency

WSD Example

Input

Sentence

He will be succeeded by Ivan Allen Jr., who became a candidate in the Sept. 13 primary after Mayor Hartsfield announced that he would not ***run*** *for reelection.*

Word

"run" (v)

MRD

a) (0) "become undone; 'the sweater unraveled'

b) (0) "come unraveled or undone as if by snagging; 'Her nylons were running'"

c) (0) "reduce or cause to be reduced from a solid to a liquid state, usually by heating; 'melt butter'; 'melt down gold'; The wax melted in the sun'"

d) (3) "cause to perform; 'run a subject'; 'run a process'"

…

h) (7) "run, stand, or compete for an office or a position; 'Who's running for treasurer this year?'"

…

r) (106) "move fast by using one's feet, with one foot off the ground at any given time; 'Don't run—you'll be out of breath'; 'The children ran to the shore'"

…

(41 total options)

WSD Evaluation Data Set

Corpus. SemCor

>185,000 annotated words

MRD. WordNet v3

>212,000 word senses

WSD Task Analysis

Effective Sense Size. A measure of the difficulty of disambiguating a word via a random selection strategy

Measured as…

$$\text{ESS} = \frac{\#\ \textbf{available}\ \text{senses}}{\#\ \textbf{appropriate}\ \text{senses}}$$

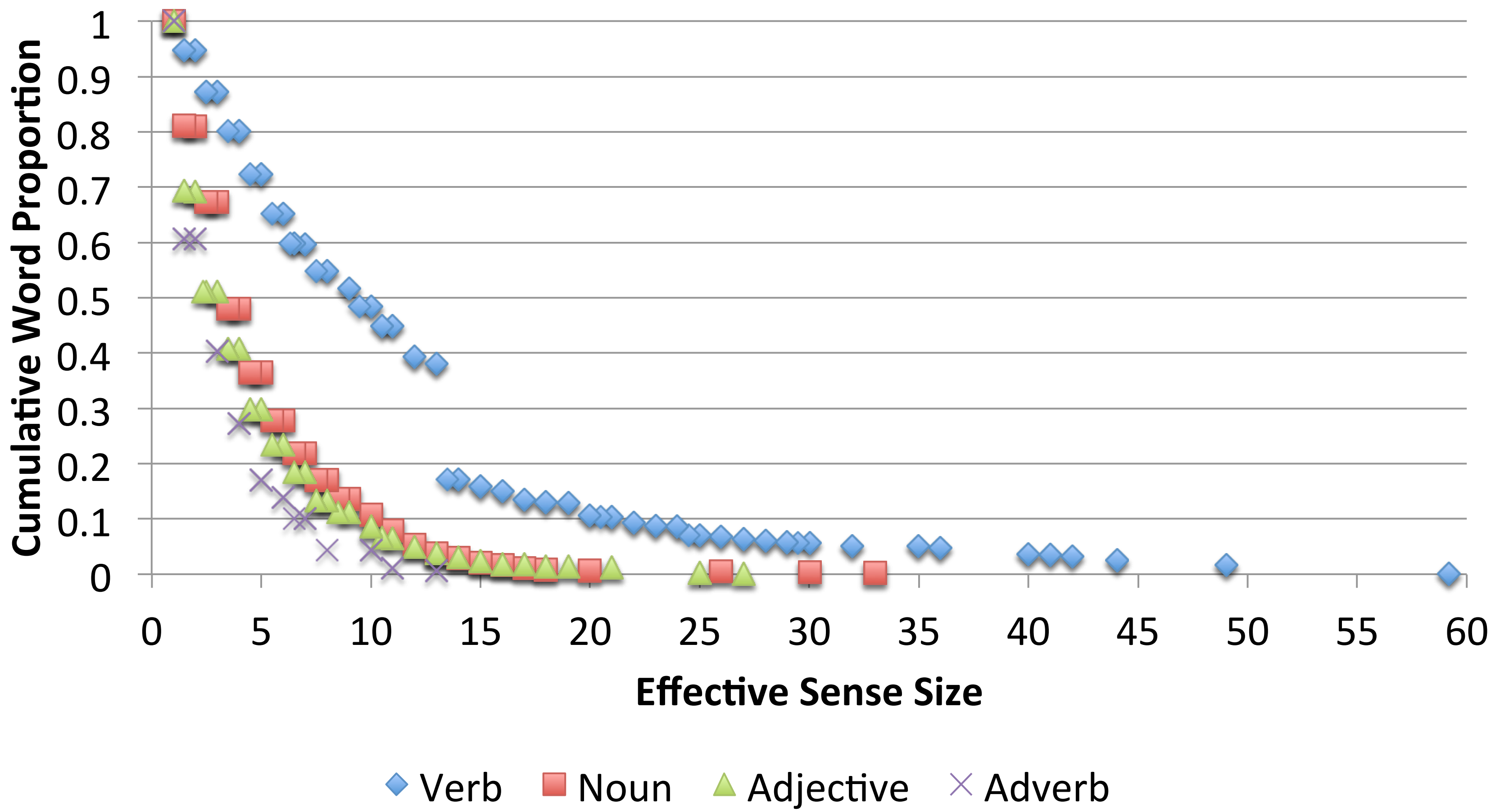
Data Set Analysis

	Adverb	Adjective	Noun	Verb
ESS=1	39%	30%	19%	5%
Median ESS	2.5	3	3.5	9.5

Overall…

- Average ESS: 2-3
- Expected accuracy: **38.73%**

SemCor Analysis



Assessing Effectiveness

1. Evaluation Methodology

2. Non-Memory Baselines

3. Memory Models

Evaluation Methodology

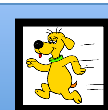
Task

...

7. "announce" (v)
8. "not" (r)
9. "run" (v)
10. "reelection" (n)

...

1. "run" (v)

2. 

3. 

Agent

Non-Memory Baselines

Algorithm	Accuracy
Random	38.73%
Lesk (1986)	63.40%
Simplified Lesk (Kilgarriff & Rosenzweig, 2000)	65.52%
Static Frequency (most common sense)	**76.39%**

Notes:

- These algorithms do not learn from task feedback
- Lesk depends upon definition quality
- Frequency depends upon corpus sense distribution representativeness

Memory Models

$>^{?}$ 76.39%

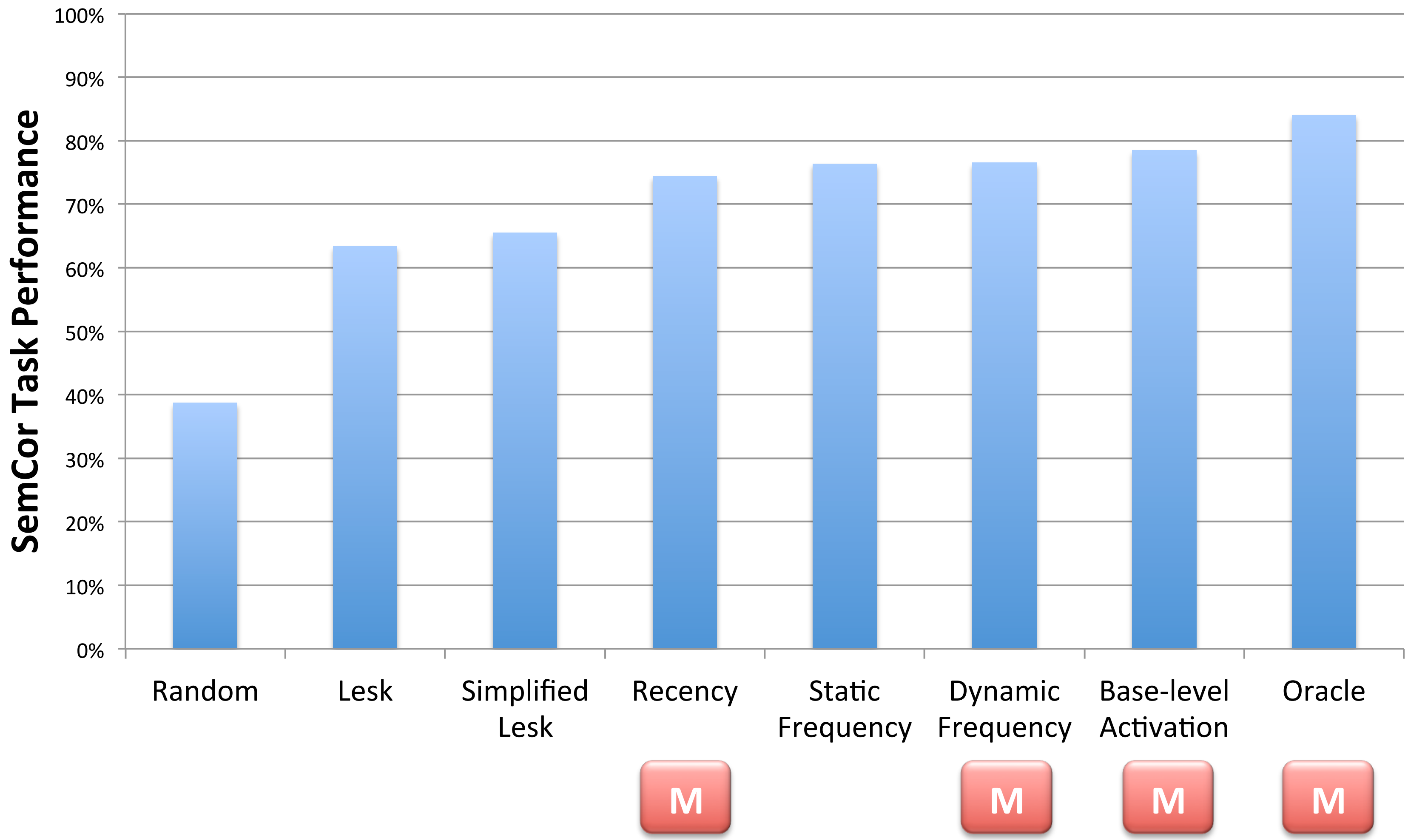
	Intuition	Run 1	Run 10
Recency	Select most recently encountered sense (surrogate for context)	72.34%	74.43%
Frequency	Select most frequently encountered sense (empirical sense commonality)	71.69%	**76.53%**
Base-level	Exponential decaying memory access history (rational analysis)	74.45%	**78.47%**
Oracle	Max(recency, frequency)	**79.51%**	**84.08%**

Base-level Activation

$$\ln(\sum_{j=1}^{n} t_j^{-d})$$

Term	Meaning
n	Number of memory accesses
t_j	Time since the j^{th} access
d	Free decay parameter

Effectiveness: Summary



Towards Efficiency

Focus. Base-level activation

- High WSD performance
- Commonly used in cognitive modeling community

Challenge. Exponential decay of all memories at each time step

Approach. Heuristically update memory activation

$$\ln\left(\sum_{j=1}^{n} t_j^{-d}\right)$$

Evaluation Metrics

Updates. Number of memories for which to re-compute activation

- Average
- Maximum (~ reactivity)

Validation. Percentage of word sense selections that are identical to naïve implementation

Efficiency Summary

	Intuition	Avg. Updates	Max. Updates	Validation
Naïve	For all candidate memories, re-compute	2.94	31	100%
Stable	Only re-compute correct memories	0	0	72.85%
NT	Only re-compute for high (N)umber of accesses or recent (T)ime of access	1.32	31	100%
NTM	NT + incremental (M)aintenance	1.74	6	99.87%

Summary

- Analyzed WSD task in context of comparing and evaluating memory models
- Demonstrated effectiveness of memory models that incorporate history of access
- Evaluated initial heuristics for efficient support

Future work. Additional data sets, tasks, memory models, and evaluation in real systems that incorporate additional knowledge.