Nuance:

A Complex Network of Concepts

Adam Olenderski[1,3] **and René Doursat**[2,3]

[1]Robotics Research Group, [2]Brain Computation Laboratory
[3]Department of Computer Science and Engineering
University of Nevada, Reno

Nuance: A Complex Network of Concepts

1. Introduction
2. Prototype Model
3. Preliminary Results
4. Discussion

Nuance: A Complex Network of Concepts

1. Introduction
   - Motivation: From Symbols to Meaning
   - Denotation and Connotation
   - A Linguistic Model of Cognition
2. Prototype Model
3. Preliminary Results
4. Discussion

1. Introduction

➢ Motivation: From Symbols to Meaning

- ✓ unraveling how language correlates with the mind's conceptual organization is a core challenge of cognitive science and AI
  - natural language processing/translation
  - human-computer interaction
  - text information retrieval
  - conceptual Web search
- ✓ common-knowledge associations, e.g., between 'throw' and 'ball', are not found in dictionaries
- ✓ yet, they reveal our fundamental cognitive frames of reference (a.k.a. semantic fields, stereotypes, scenarios, etc.)

→ *how can we capture, model and use these frames?*

1. Introduction

➢ Denotation and Connotation

(1) It was a dark and stormy night

- ✓ what does this sentence *mean*? (denotation)
  - dark = little light
  - stormy = rain, lightning, thunder
  - night = no sun
- ✓ what other meaning does this sentence convey? (connotation)
  - fear, apprehension, suspense
  - violence, tumult
- ✓ where does this extra meaning come from?
  - cognitive frames of reference, themselves created by:
  - real-world, nonlinguistic experience (perception and action)
  - linguistic experience (written and oral)

1. Introduction

➢ A Linguistic Model of Cognition

- ✓ how can we simulate frames of reference?
- ✓ one way is cognitive linguistic, linking language with perception
  - nonlinguistic, iconic representations: visual scenes, etc.
- ✓ another way would be trying to infer frames of reference from purely **linguistic usage**
  - statistical co-occurrence of words in fully formed written text and spoken language: how often/strongly words are related
  - the whole written record of human experience is now almost entirely accessible via Internet

→ *opens the way to automated statistical parsing on a big scale*

Nuance: A Complex Network of Concepts

1. Introduction
2. Prototype Model
   - From Text Corpora to Word Clusters to Concepts
   - A Network-Database of Cluster-Concepts
   - Creating the Network: Nodes and Links
   - Querying the Network
3. Preliminary Results
4. Discussion

2. Prototype Model

➢ From Word Clusters to Concepts

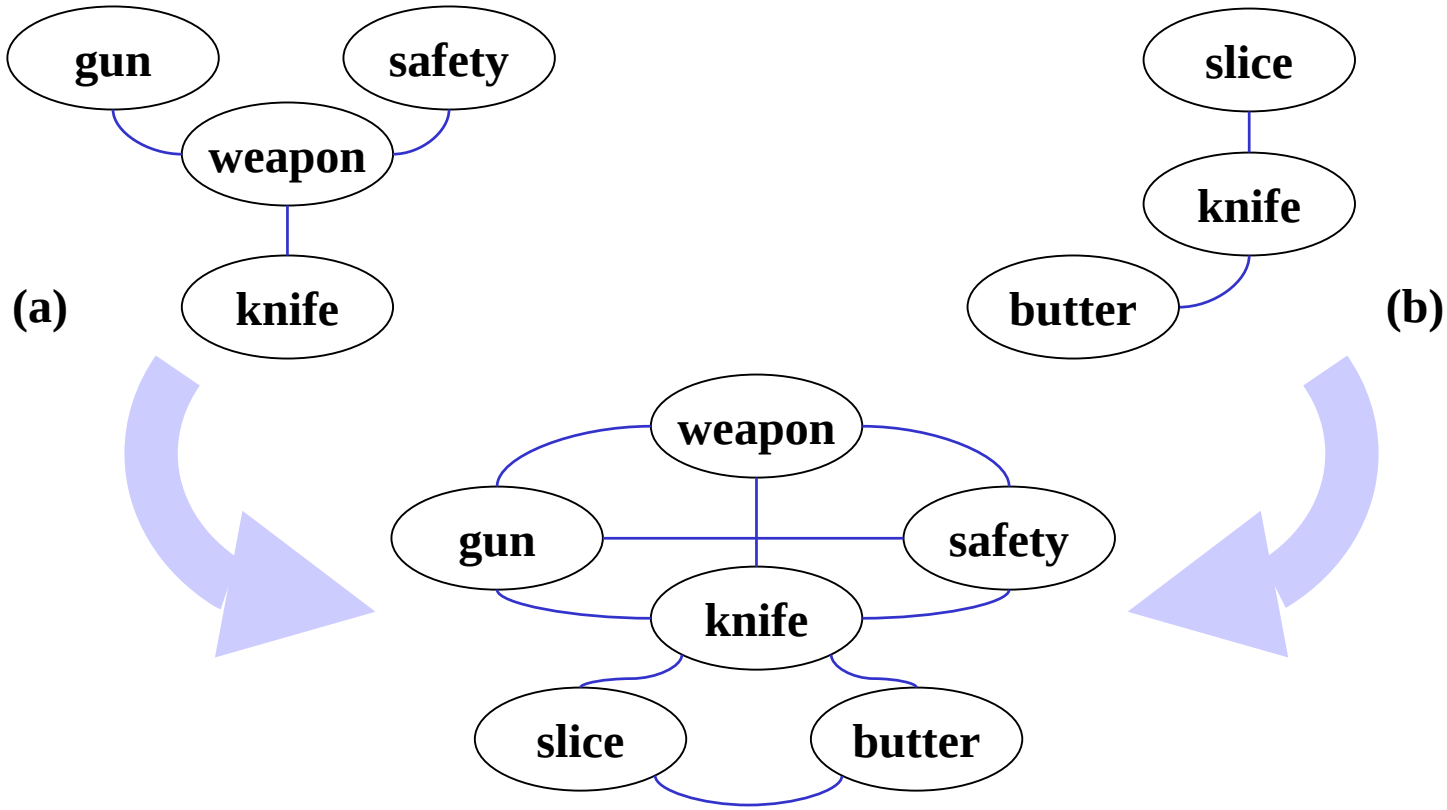
- ✓ the model is based on the premise that **concepts** are best captured by **word clusters** instead of single words
- ✓ a single word can belong to the intersection of several clusters representing different concepts
  - homonyms
    - game: chess, play, cards, tv, snacks, …
    - game: hunt, animal, wild, rifle, …
  - different usages
    - game: chess, play, cards, tv, snacks, …
    - game: joke, psychology, scheme, social, …
  - nuances
    - game: chess, play, cards, tv, snacks, …
    - game: competition, baseball, sports, tv, snacks, …

2. Prototype Model

➢ From Text Corpora to Word Clusters

(a) For your safety, no knife, gun, or other weapon is allowed

(b) A knife can slice through butter



2. Prototype Model

➢ A Network-Database of Cluster-Concepts

- ✓ we exploit the combinatorial power of networks to express semantic and cognitive concepts as supra-word entities
- ✓ we propose an algorithm to create a network-database of such word clusters by scanning and merging text corpora
- ✓ then, the network-database can be queried to retrieve cluster-concepts by selectively activating some of their nodes
- ✓ there is no predefined list of cluster-concepts: new word combinations might emerge from the connectivity of the network, depending on the input query

2. Prototype Model

➢ Creating the Network — Nodes

- ✓ when scanning the text, the system records the words encountered in the text and their location
  - if the word does not have a node, create a new node
  - add the location or “address” of the word to a list of addresses maintained by the word’s node
  - a word address is hierarchical, for example a quintuplet: *{document, section, paragraph, sentence, rank within sentence}*

2. Prototype Model

➢ Creating the Network — Nodes

Don't Call It Negotiating

You need not give up on a debt-free life. **Students** or their parents should phone or visit the aid office and request an appeal of their financial aid offer if they believe the initial offer does not meet their needs. . .

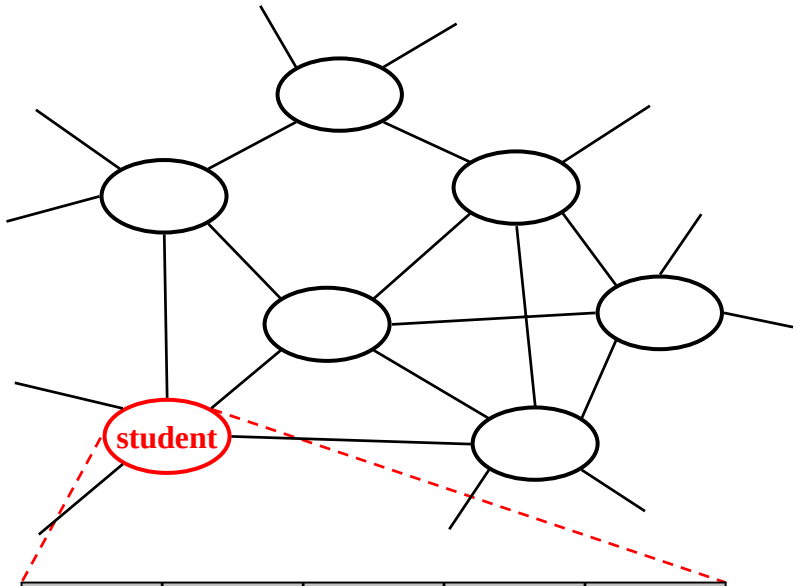
The Cost of College

By a **student**'s senior year of high school, most parents are struggling to save up any extra cash for college. . .

1. Test prep

Earning a certain AP test score may allow a **student** to get college credit and bypass some freshman classes. . .

Then there's the SAT. It's up in cost this year from $29.50 to $41.50, largely because it now includes a writing component. About half of **students** now take the SAT more than once. . .



doc	sect	parag	sent	rank
1	1	1	2	1
2	1	1	1	3
2	2	1	1	10
2	2	2	3	4

2. Prototype Model

➢ Creating the Network — Links

- ✓ the system then creates links between all nodes as follows:
  - given a pair of nodes, compare each address of the first word to each address of the second word
  - the distance between two word addresses is the compound effect of 5 factors, one factor for each level of the hierarchy
    - if the two addresses are in the same document, multiply by 1.01, otherwise by .8
    - if the two addresses are in the same section, multiply by 1.02, otherwise by .85
    - for same paragraph: × 1.05, otherwise .9
    - for same sentence: × 1.2, otherwise .95
    - if the two words are adjacent: × 1.5, otherwise .99

2. Prototype Model

➢ Creating the Network — Links

Don't Call It Negotiating

You need not give up on a debt-free life. Students or their parents should phone or visit the aid office and request an appeal of their financial aid offer if they believe the initial offer does not meet their needs. . .

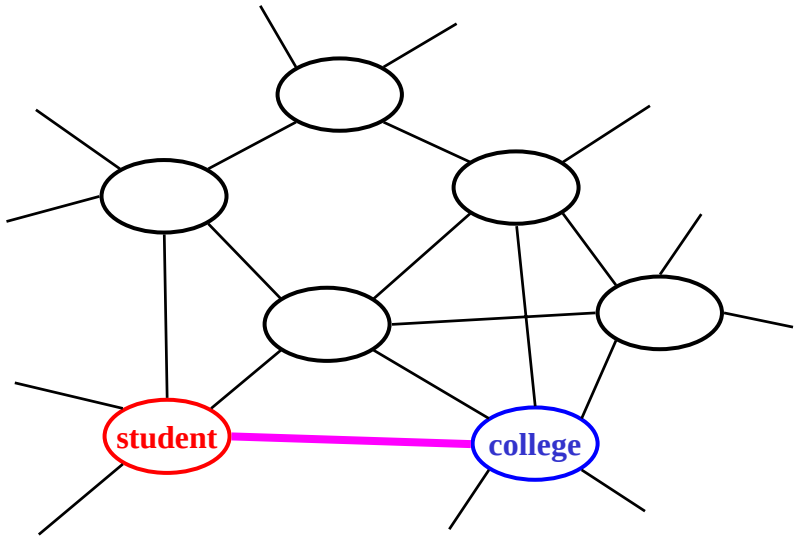
The Cost of College

By a student's senior year of high school, most parents are struggling to save up any extra cash for **college**. . .

1. Test prep

Earning a certain AP test score may allow a **student** to get **college** credit and bypass some freshman classes. . .

Then there's the SAT. It's up in cost this year from $29.50 to $41.50, largely because it now includes a writing component. About half of students now take the SAT more than once. . .



doc	sect	parag	sent	rank
2	2	1	1	10
× 1.01	× .85	[× .9	× .95	× .99] = .73
2	1	1	1	21
2	2	1	1	13
× 1.01	× 1.02	× 1.05	× 1.2	× .99 = 1.29

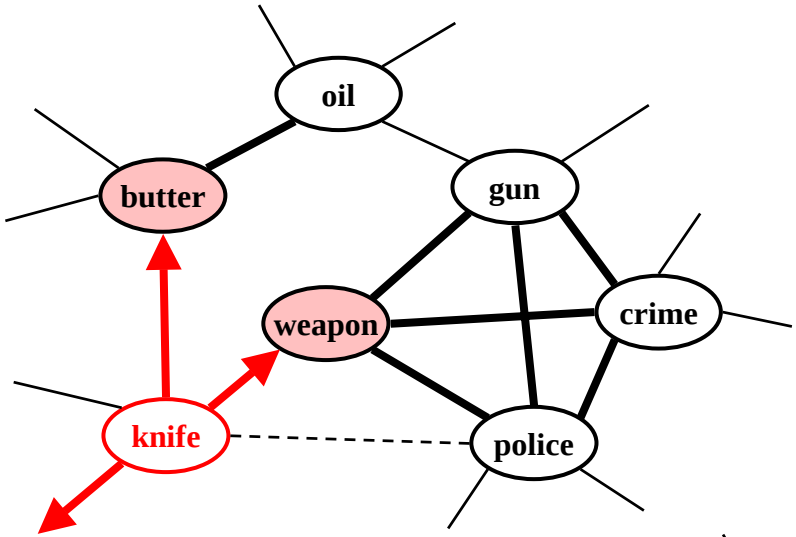
2. Prototype Model

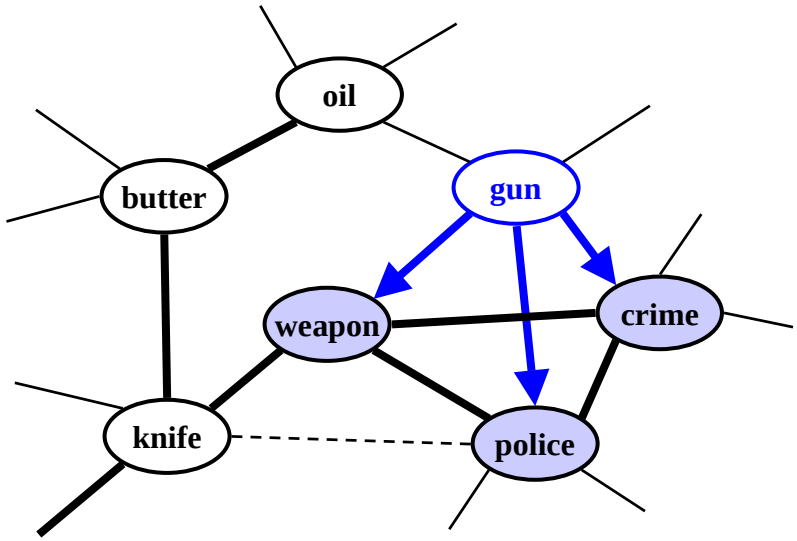
➢ Querying the Network

- ✓ the user may examine connected portions of the network by inputting queries
  - for now, a query is simply a list of words
- ✓ the system replies with output concepts
  - for now, a concept is also a list of words representing the cluster that was activated in the network by the query words
  - each query word activates the N words (e.g., 20) to which it is most strongly connected—its "preferred neighborhood"
  - the resulting concept words are at the intersection of all the preferred neighborhoods of the query words

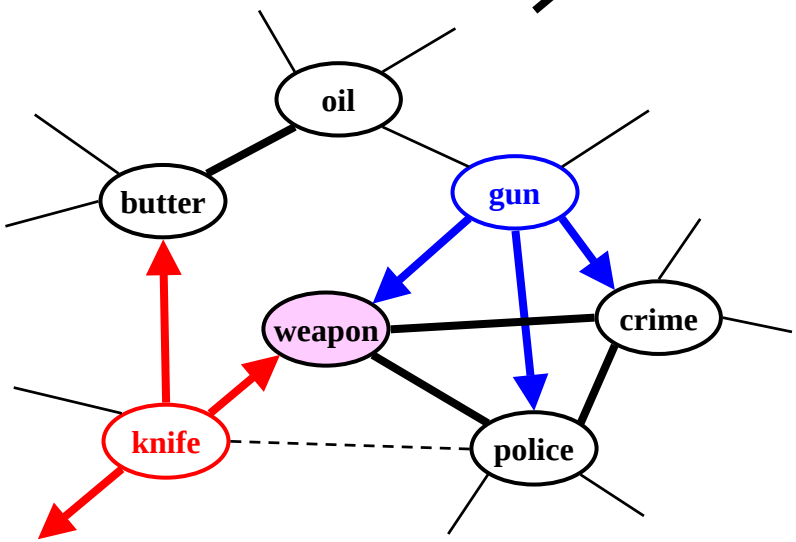
2. Prototype Model

- Querying the Network





knife, gun



knife, gun, weapon

Nuance: A Complex Network of Concepts

1. Introduction
2. Prototype Model
3. Preliminary Results
   - A Simple Command-Line Program (Demo)
4. Discussion

3. Preliminary Results

- A Simple Command-Line Program (Demo)
  - ✓ processing an input text
  - ✓ looking at the result file listing all the words and their links
  - ✓ querying the network with a few words
  - ✓ getting the output cluster-concepts

Nuance: A Complex Network of Concepts

1. Introduction
2. Prototype Model
3. Preliminary Results
4. Discussion

4. Discussion

➢ (Immediate) Future Directions

- ✓ integrating truly large text corpora, using a Web crawler
- ✓ exploration of model rules
  - network building: weight computation schemes
  - network querying: cluster activation schemes
- ✓ exploration of parameter space, self-tuning, optimization
- ✓ network structure analysis using complexity metrics: clustering coefficient, average path length, power law exponent, etc.

Nuance: A Complex Network of Concepts

1. Introduction
2. Prototype Model
3. Preliminary Results
4. Discussion