A Large-Scale Knowledge Base for ACT-R

Dario Salvucci
Drexel University

Approach

- From Newell's "20 Questions" paper (1973)…

One Program for Many Tasks

The third alternative paradigm I have in mind is to stay with the diverse collection of small experimental tasks, as now, but to construct a single system to perform them all. This single system (this model of the human information processor) would have to take the instructions for each, as well as carry out the task. For it must truly be a single system in order to provide the integration that we seek.

Approach

- Unified models of cognition
  - Consider a single cognitive model with...
    - a single (initial) set of declarative chunks
    - a single (initial) set of production rules
    - with (initially) fixed parameter settings
    - on a fixed cognitive architecture
  - ... and try to account for behavior across a range of diverse domains
- Two components
  - Procedural knowledge
  - Declarative knowledge

Procedural Knowledge

- What procedural knowledge does such a unified model require?
  - basic procedural skills
    - e.g., clicking an interface icon, typing a key
  - instruction-following skills
    - listening to and encoding instructions
    - following them to generate actions
    - [building on work by Taatgen, Anderson, et al. in ACT-R; by Lewis, Huffman, Laird, et al. in Soar]
  - indexicals
    - references to items, visual or cognitive
    - e.g., experimenter's word "target" as an imaginal slot name
    - e.g., experimenter "finger" as location of specified information

Current Task Domains

- Paired Associates
- Tracking & Choice
- Equation Solving
- Menu Selection
- Dual-Choice (PTS)
- Dual-Choice (PRP)
- Driving & Dialing

Task	R	Err
Paired RT	>.99	0.11
Paired Correct	0.97	0.08
Tracking Error	0.97	0.10
Tracking RT	0.69	0.09
Equation Gazes	0.93	0.30
Equation GazeDur	>.99	0.16
Menu RT	0.90	0.57
Menu FirstFix	0.96	0.09
DualChoice1 RT	0.93	0.12
DualChoice2 RT	0.77	0.20
Driving-Dialing RT	0.99	0.10
Driving-Dialing LD	0.96	0.09
Driving-Dialing LV	>.99	0.07

Declarative Knowledge

- So this model can do 7 tasks...
  but it basically doesn't *know* anything
  - It knows math facts, and that's practically all
  - It doesn't know what New York is, or who Bob Dylan is, or what a platypus is.
- Unified models will need more knowledge.
  - Something akin to Cyc, Scone, etc.
  - But available to ACT-R models, and with psychologically plausible activations

Knowledge Source

- Derive knowledge from a huge existing source: Wikipedia
  - full text isn't very useful, but...
  - Wikipedia includes "infoboxes" with basic attribute-value pairs
  - DBpedia has cleaned up the data and made them available
  - Representation as triples:

```
Bruce Springsteen,   birthplace,   New_Jersey
```

Bruce Springsteen

Background information	
Also known as	The Boss, Bad Scooter
Born	September 23, 1949 (age 62) Long Branch, New Jersey, United States
Genres	Rock, folk rock, heartland rock, hard rock, roots rock
Occupations	Musician, Songwriter
Instruments	Vocals, guitar, harmonica, bass guitar, piano, percussion, banjo, drums, keyboards
Years active	1969–present
Labels	Columbia
Associated acts	The E Street Band, Steel Mill, Miami Horns, The Sessions Band
Website	www.brucespringsteen.net

Represer

- Instead o
  chunk, th
  triplet ch
  - basicall
  - or: *pre*
- Why?
  - knowin
    - e.g.,
  - attribut
  - sometir
  - attribut

Bruce Springsteen	background	solo singer
Bruce Springsteen	alias	The Boss
Bruce Springsteen	birth date	1949-09-23
Bruce Springsteen	birth place	Long Branch, New Jersey
Bruce Springsteen	birth place	New Jersey
Bruce Springsteen	birth place	United States
Bruce Springsteen	instrument	Singing
Bruce Springsteen	instrument	Guitar
Bruce Springsteen	instrument	Harmonica
Bruce Springsteen	instrument	Piano
Bruce Springsteen	genre	Rock music
Bruce Springsteen	genre	Heartland rock
Bruce Springsteen	genre	Folk rock
Bruce Springsteen	genre	Roots rock
Bruce Springsteen	genre	Americana (music)
Bruce Springsteen	occupation	Musician
Bruce Springsteen	occupation	Singer-songwriter
Bruce Springsteen	active years start year	1972
Bruce Springsteen	record label	Columbia Records
Bruce Springsteen	associated band	E Street Band
Bruce Springsteen	associated band	Steel Mill
Bruce Springsteen	associated band	Miami Horns
Bruce Springsteen	associated band	The Sessions Band
Bruce Springsteen	associated musical artist	E Street Band
Bruce Springsteen	associated musical artist	Steel Mill
Bruce Springsteen	associated musical artist	Miami Horns
Bruce Springsteen	associated musical artist	The Sessions Band
Bruce Springsteen	homepage	http://www.brucespringsteen.net/
Bruce Springsteen	instrument	Fender Telecaster
Bruce Springsteen	instrument	Fender Esquire
Bruce Springsteen	instrument	Takamine Guitars
Bruce Springsteen	instrument	Hohner

Chunks

- Most chunks come from "Infobox properties"

Aristotle	birth year	-0384
Aristotle	death year	-0322
Aristotle	notable idea	Golden mean (philosophy)
Aristotle	notable idea	Reason
Alabama	demonym	Alabamian
Alabama	capital	Montgomery, Alabama
Alabama	largest city	Birmingham, Alabama
Alabama	area total	1.35765E11
Abraham Lincoln	birth date	1809-02-12
Abraham Lincoln	birth place	Hardin County, Kentucky
Abraham Lincoln	death date	1865-04-15
Abraham Lincoln	resting place	Oak Ridge Cemetery
Abraham Lincoln	spouse	Mary Todd Lincoln
Algeria	anthem	Kassaman
Algeria	currency	Algerian dinar
Algeria	capital	Algiers
Algeria	official language	Arabic
Algeria	official language	French language
Ayn Rand	notable work	The Fountainhead
Ayn Rand	notable work	Atlas Shrugged
Amphibian	kingdom	Animal
Amphibian	phylum	Chordate

Chunks

- Types come from "Infobox types"

Autism	isa	disease
Aristotle	isa	person
Aristotle	isa	philosopher
Alabama	isa	place
Alabama	isa	populated place
Alabama	isa	administrative region
Abraham Lincoln	isa	person
Abraham Lincoln	isa	politician
Abraham Lincoln	isa	president
Academy Award	isa	award
Algeria	isa	populated place
Algeria	isa	country
Ayn Rand	isa	person
Ayn Rand	isa	writer
Amphibian	isa	species
Amphibian	isa	eukaryote

Chunks

- Mapping string → symbol from "Name redirects"

```
Madonna (entertainer)      name     "Madonna"
Madonna (entertainer)      name     "Lourdes Leon Ciccone"
Madonna (entertainer)      name     "Madonna Louise Veronica Ciccone"
Madonna (entertainer)      name     "Madonna Ciccone"

Muammar al-Gaddafi         name     "Muammar Qaddafi"
Muammar al-Gaddafi         name     "Mohammar Qaddafi"
Muammar al-Gaddafi         name     "Gadaffi"
Muammar al-Gaddafi         name     "Gadhafi"
Muammar al-Gaddafi         name     "Gaddafi"
Muammar al-Gaddafi         name     "Muammar Gaddafi"
Muammar al-Gaddafi         name     "Mu'ammar Al Qathafi"
Muammar al-Gaddafi         name     "Moammar Ghadafi"
…
```

Declarative Knowledge

- Final knowledge base
  - >20 million chunks describing ~2.3 million objects
- Implementation for retrieval
  - SQLite database, 2.8 GB
  - if information not found in main memory, check "extended" memory database
    - not a theoretical claim — purely for efficiency, and leaves well enough alone for main memory
  - Java ACT-R only right now (LISP ACT-R would just need a SQLite wrapper)
  - In general, retrieval takes <1 second

Declarative Knowledge

- Base-level activation
  - Assume that a person's exposure to each fact is proportional to the *N* mentions/links in Wikipedia
  - Base-level activation = log(2*N*)
  - Example: retrieve "isa musician"

```
Bob_Dylan
Elvis_Presley
David_Bowie
Madonna_(entertainer)
John_Lennon
```

  - Note: all chunks about a person/object have the same base-level activation (we need more data to do better)

Declarative Knowledge

- Spreading activation
  - computed as usual... except that associations boost associated symbols, not chunks
  - still working out the details

$$A_i = B_i + \sum_k \sum_j W_{kj} S_{ji} + \varepsilon$$

$$S_{ji} = S - \ln(fan_j)$$

New York	name	"New York"
New York	capital	Albany, New York
New York	largest city	New York City
New York City	name	"New York"
New York City	leader title	Mayor
New York City	leader name	Michael Bloomberg
New York City	population	8391881

Declarative Knowledge

- Demonstration procedural knowledge
  - currently, a very simple parse/retrieval system
  - when a phrase is heard, retrieve "name" chunk to map phrase → symbol (with context)
    - e.g., "New York" → New_York_City or New_York
  - parse simple questions...

What is the capital of New_York?
What is the population of New_York?
What actor is a star of Airplane?
What athlete is a star of Airplane?
What actor was born in Philadelphia?
Who was the director of Philadelphia?
What musician was born in New_Jersey?
Who is Mean_Joe_Greene?
What president is Princeton the alma_mater of?

Final Points

- Three good things
  - step toward unified models of cognition
  - knowledge re-use (see: Susan Chipman)
  - potential for HCI applications (understanding people)
- Two issues
  - deeper questions of representation
  - interaction with natural language mechanisms
- One confession
  - I'm not sure what to do with this.