Semantic Understanding and Commonsense Reasoning in an Adaptive Photo Agent

Xinyu Hugo Liu
Thesis Advisor: Henry Lieberman
Software Agents Group
MIT Media Laboratory
hugo@media.mit.edu

EECS MasterWorks Colloquium talk
29 Apr 2002
Cambridge, CA
The Big Picture

 Traditional Computer Programs
- Users adapt to software interface
- Makes mistakes obvious to humans
- Static

 More Intelligent and Adaptive
1. Software adapts to user. Understand English
2. Has some common sense
3. Learns and improves over time
   Not a complete set of desirable properties.

This thesis investigates these 3 properties in the photo storytelling domain.
Highlights

- **ARIA Photo Agent**
  Facilitates storytelling with photos (i.e. email, web page)
  Annotates photos based on use
  Dynamically retrieves photos to facilitate storytelling

- **Deeper Semantic Understanding → Better annotation**
  A world-aware text parser for better understanding and information extraction

- **Common Sense → More Robust Photo Retrieval**
  People can understand that the annotations “bride” and “wedding” are connected…and so should ARIA

- **Learning Personal common sense**
The User Interface

Email Headers

Text Pane

Photo Pane (ordered by relevance)

Double-click to edit annotations
Hi Jane!

Last weekend, I went to Ken and Mary’s wedding in San Francisco. Here’s a picture of the two of them at the gazebo.

Regards,
John
A Baseline Approach

- Extract all keywords from text surrounding photo

- Jane
- Last
- Weekend
- Went
- Ken
- Mary’s
- Wedding
- San
- Francisco
- Picture
- Gazebo

Doesn’t make sense!
Rather brittle!
Our approach

- Use concepts, not keywords
- Identify semantically important concepts

- Who: Ken, Mary
- What: wedding
- When: June 12, 2001, 1pm
- Where: gazebo, San Francisco, CA

How did ARIA learn this annotation?
“Hi Jane!

Last weekend, I went to Ken and Mary’s wedding in San Francisco.

Here’s a picture of the two of them at the gazebo.
Last weekend, I went to Ken and Mary’s wedding in San Francisco. Here’s a picture of the two of them at the gazebo.
Mechanism for understanding

Tokenize and Normalize
Anaphora Resolution / Context Attacher
Temporal UA runs

POS Tagging
TBL Tagger (Brill, 1994) on Penn Treebank tagset

Syntactic Parsing
Link Grammar (Sleator & Temperley, 1993)
outputs a constituent parse

Semantic Processing
transforms a syntactic parse tree → semantic parse
ontology of tags for event structure (Jackendoff, 1983, 1990)
Not as deep as logical formulas i.e. CYC-NL (Burns & Davis), UNITRAN (Dorr, 1992), but also not as brittle

Frame Extraction

ARIA_DATE_INTERVAL{2001-6-10,2001-6-12},
[PERSON $user] [ACTION go] [TRAJECTORY to [PERSON Ken] and [PERSON Mary] ’s [EVENT wedding] [LOCATION in [PLACE San Francisco]] [SBR .] [PLACE Here] [STATE is] [THING a picture [PROPERTY of [PERSON the two of them] [LOCATION at [PLACE the gazebo]]]]] SBR
Critique

The pragmatist says:

Why go through all this effort to parse text? Why not just keyword spot PLACES, EVENTS, NAMES?

Response

That would make a great baseline

Getting the constituent structure is important for:

- Learning bindings between concepts, e.g. “Here is a picture of George, who was the minister at our wedding and also a great family friend.” (Learning)
- Exploit structure to induce semantic importance
  - “Here is my wife posing with foobar in front of the Eiffel Tower
Part II: Retrieval

- As user types, the photos in the shoebox dynamically reorder (based on relevance)
- A baseline approach: “exact keyword matching”
  Problem is.. not robust morphological variations (e.g. sisters =/= sister)
  semantic connections missed (e.g. wedding =/= bride)
Our Approach

*Missed Semantic Connections*

A good candidate for commonsense reasoning

1. **Observation:** In consumer photography, events and situations tend to be somewhat stereotyped and predictable:
   - brides are usually found at weddings
   - a birthday cake may be found at a birthday party
   - picnics may be located in parks

   (note the granularity of knowledge)

2. So, given the annotation “bride”, commonsense tells us to also add “wedding”
Each annotation undergoes common sense expansion, which produces other semantically related annotations. For example:

```python
>>> expand("bride")
(('love', '0.632'), ('wedding', '0.5011'))
(('groom', '0.19'), ('marry', '0.1732'))
(('church', '0.1602'), ('marriage', '0.1602'))
(('flower girl', '0.131') ('happy', '0.131'))
(('flower', '0.131') ('lake', '0.131'))
(('cake decoration', '0.131') ('grass', '0.131'))
(('priest', '0.131') ('tender moment', '0.131'))
(('veil', '0.131') ('wife', '0.131'))
(('wedding dress', '0.131') ('sky', '0.131'))
(('hair', '0.1286') ('wedding bouquet', '0.1286'))
(('snow covered mountain', '0.1286'))
```

Not perfect or complete, be can still be useful!
Source of Commonsense Knowledge

- **Open Mind Common Sense (Singh, 2002)**
  - Publicly acquired common knowledge about the world
  - Has about 400,000 commonsense facts
  - Represented as fill-in-the-blank English sentences

- **Sample Sentences:**
  - Something you may find in a restaurant is a waiter.
  - Something that might come after a wedding is a wedding reception.
Comparision

 выбранный

 Contrast:
 CYC (Lenat, 1998)
 1 million hand crafted assertions
 represented more formally, i.e. as logic
 not publicly available

 Caveats of OMCS
  More ambiguity than CYC
   Word senses not disambiguated
  Coverage of topics is spotty
  Culturally specific (CYC is too)
   i.e.: common knowledge for middle-class USA
    • Wedding scenario is a perfect example
Applying commonsense to ARIA

- OMCS: Good granularity for our problem domain
  i.e.: events, social composition, etc.

- Full commonsense inference is hard.
  Combinatorial explosion
  Deep understanding/representation needed

- A more robust approach is needed
  Treat reasoning as an associative mechanism
Expanding annotations Using Spreading Activation

- Use OMCS to build a semantic resource
- Choose spreading activation network
 Nodes are concepts, edges are semantic relation
  e.g.: bride and wedding are connected through foundAt edge
SAN applications to IR: Salton & Buckley (1988)
Enables shallow commonsense reasoning (only applying transitivity inference pattern)
On-the-fly photo suggestion demands efficiency
A world-semantic resource

Pattern-Matching Parser compiles OMCS to Predicate-Argument structures

\[ \text{e.g.: MayBeFoundAt(bride, wedding)} \]

Modes of Reasoning Desired

- **Classification**: A dog is a pet
- **Spatial**: San Francisco is part of California
- **Scene**: Things often found together are: restaurant, food, waiters, tables, seats
- **Purpose**: A vacation is for relaxation; Pets are for companionship
- **Causality**: After the wedding ceremony comes the wedding reception.
- **Emotion**: A pet makes you feel happy; Rollercoasters make you feel excited and scared.
Building concept node graph

- 20 pattern rules => 50,000 pred-arg structs
- 30,000 concept nodes
- 80,000 pairs of directed edges
- Average branching factor is 5
Spreading activation algorithm

- Distance from origin $\leftrightarrow$ semantic distance
- Connected nodes activate if energy meets some relevance threshold (e.g. 0.2)
- Given edge $(A,B)$, activation score of $B$:

$$AS(B) = AS(A) \times \text{weight}(edge(A,B))$$
Ambiguity and Noise Management

- Two opportunities to reweight edges to improve semantic relevance
  - Reinforcement
  - Inverse Popularity

- First, Reinforcement
  - Simple Idea
  - Nodes connected thru multiple paths are more semantically related
  - Applies to k-clusters (mutually reinforced)
Inverse Popularity

- Punish a node for having too many children
- Often symptomatic of word sense collision (since OMCS is not sense disambiguated)
- Sense collisions lead to noise
- Discount by equation (above-right)

$$newWeight = oldWeight \cdot discount$$

$$discount = \frac{1}{\log(\alpha \cdot branchingFactor + \beta)}$$
System Demo
Summary of Contributions

- Developed a robust world-semantic parser
  - Deeper than keyword spotting.. allows structural induction, learning relationships...
  - YET, not full blown logical representation.. more robust and doesn’t require enormous lexical resources like argument structure..

- Integrated Commonsense Reasoning
  - Large scale commonsense knowledge
  - Semantic associations help reduce missed connections
  - Many modes of reasoning performed (Social, Geospatial, Causal, Emotional)

- Learning and user modeling integrated with ordinary commonsense
Pointers, etc.

Papers:


Access to papers: google for “hugo liu”