Robust Photo Retrieval Using World Semantics

Hugo Liu, Henry Lieberman
Software Agents Group
MIT Media Laboratory

LREC 2002 Workshop:
Semantics for IR
2002.06.02
Las Palmas, Gran Canaria
Agenda

• The problem of photo retrieval
  • Traditional methods for improving robustness
  • Using world semantics to improve robustness
  • OMCS: A corpus of commonsense
  • Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
• Mechanisms for managing noise and ambiguity
• Conclusions and future work
The Problem of Photo Retrieval

Problem:

- There exists a database of photos, each annotated with keywords describing the subject of that photo.

- When retrieving these photos by keywords, exact keyword matching is too brittle, and so relevant photos might be missed by a query due to differences in vocabulary (e.g. “car” / “automobile”) and missed semantic connections (e.g. “bride” / “wedding”).

- The goal is to improve the robustness of retrieving photos.
Agenda

• The problem of photo retrieval

• **Traditional methods for improving robustness**
  • Using world semantics to improve robustness
  • OMCS: A corpus of commonsense
  • Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
  • Mechanisms for managing noise and ambiguity
  • Conclusions and future work
Can thesauri and dictionaries help improve robustness?

- Photos with keyword annotations resemble documents
- Thus we can apply the same IR techniques of query expansion by thesauri and dictionaries
  - e.g. naive synonym expansion:
    - “car” ➔ (“car”, “vehicle”, “automobile”)
  - e.g using WordNet lexical-based nymic relations
    - “cat” ➔ (“cat”, “feline”, “animal”, “paw”)
- Can also use word co-occurrence statistics, etc.
Can thesauri and dictionaries help improve robustness?

- **Answer:** No, not exactly
- **Synonyms, co-occurrence lists, etc**
  - good for bridging simple vocabulary differences like “car” and “automobile”
- **Lexical-based taxonomic/nymic relations**
  - More powerful than thesauri b/c hierarchical
  - Has potential to bridge some semantically connected concepts
Thesauri and dictionaries
(Limitations)

- thesauri (flatly structured associations)
  - Only good for associations of depth 1
    - Relevance drops off drastically after 1 level

- dictionaries (graph structured lexical associations)
  - However, falls short because there are very few nymic relations, and they are lexically motivated relations (e.g. hypernym, hyponym, meronym, etc), rather than motivated by the relations of concepts in the everyday world.
  - Linkages between lexical items are few and far apart because there are very few nymic relations.. Therefore not adequate for creating a robust semantic net.
  - Also, lexical items are generally too formal and theoretical to be practical (dictionary bias)
    - e.g. WordNet emphasizes cats as a kind of feline, people emphasize cats as pets!
Agenda

- The problem of photo retrieval
- Traditional methods for improving robustness
- Using world semantics to improve robustness
  - OMCS: A corpus of commonsense
  - Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
- Mechanisms for managing noise and ambiguity
- Conclusions and future work
How can world semantics help improve robustness?

- **Observation**
  - *photos are snapshots of the everyday world* (e.g. golden gate bridge) and *social situations* (e.g. party, recital, vacation, wedding)
  - knowing the structure and *semantics* of the everyday world can help us to guess what other annotations might be appropriate for that same photo.
  - e.g.:
    - Photo has annotation “bride”
    - World semantics guesses “wedding”, “groom”, “wedding veil” etc. and these become expansions to the original annotation “bride”
    - This photo’s retrieval is now more robust because typing in “wedding” will bring up this photo of the “bride”

- **We hope to exploit world semantics (== “commonsense”) to anticipate and improve robust photo retrieval.**
World semantics and commonsense

• The source of the world semantic knowledge for our mechanism is Open Mind Common Sense (OMCS).

• Our mechanism compiles the OMCS corpus into a world semantic resource

• This resource is modeled as a spreading activation network, where nodes represent commonsense concepts (e.g. “bride”) and edges represent weighted semantic proximity between concepts.
Agenda

• The problem of photo retrieval
• Traditional methods for improving robustness
• Using world semantics to improve robustness
• **OMCS: A corpus of commonsense**
  • Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
• Mechanisms for managing noise and ambiguity
• Conclusions and future work
Source of Commonsense

• Open Mind Common Sense (Singh, 2002)
  – Publicly acquired common knowledge about the world and social relations
  – Has about 420,000 commonsense facts
  – Represented as fill-in-the-blank English sentences
  – Ontology of about 30 commonsense relations

• Sample Sentences:
  – Something you may find in a restaurant is a waiter.
  – Something that might come after a wedding is a wedding reception.
Comparison of OMCS

• Comparison:
  – CYC (Lenat, 2002)
  – 3 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
Open Mind Redeeming Qualities

• Good granularity for our problem domain
  – i.e.: events, social composition, etc.

• Many simple, binary relations

• Our approach is tolerant to noise
  – Each commonsense annotation expansion doesn’t count much
  – It combines a lot of “partial evidence”

• In the context of our larger system, automated photo suggestion (retrieval) is “soft” (AI)
  – Unlike “hard” uses like question answering, and general problem solving, photo suggestion isn’t mission-critical
Agenda

- The problem of photo retrieval
- Traditional methods for improving robustness
- Using world semantics to improve robustness
- OMCS: A corpus of commonsense
- **Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval**
- Mechanisms for managing noise and ambiguity
- Conclusions and future work
The goal is to expand annotations using commonsense

- 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
Building the graph

• Sample mapping rule from OMCS to pred-arg structure:
  - somewhere THING1 can be is PLACE1
    somewherecanbe
    THING1, PLACE1
    0.5, 0.1

• Differentiate Forward/Backward weight
  - Small → big (e.g. bride → wedding) preferred

• 20 mapping rules applied to get 50,000 pred/arg structures (mostly binary)
Mapping Rules

- Mapping Rules target this subset of OMCS:
- 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.
Cleaning Args in Pred/Args

- Apply a sieve grammar to normalize concepts
  - **Noun Phrase**:
    - (PREP) (DET|POSS-PRON) NOUN
    - (PREP) (DET|POSS-PRON) NOUN NOUN
    - (PREP) NOUN POSS-MARKER (ADJ) NOUN
    - (PREP) (DET|POSS-PRON) NOUN NOUN NOUN
    - (PREP) (DET|POSS-PRON) (ADJ) NOUN PREP NOUN
  - **Activity Phrase**:
    - (PREP) (ADV) VERB (ADV)
    - (PREP) (ADV) VERB (ADV) (DET|POSS-PRON) (ADJ) NOUN
    - (PREP) (ADV) VERB (ADV) (DET|POSS-PRON) (ADJ) NOUN NOUN

- Examples: “to cook some dinner” → “cook dinner”, “in the playground” → “playground”
Building concept node graph

- 400,000+ sentences in OMCS corpus
- 50,000 predicate argument structures extracted
- 20 predicates in mapping rules
- 30,000 concept nodes
- 160,000 edges
- average branching factor of 5

Diagram:

- Nodes: bride, groom, wedding
- Edges: FoundTogether, PartOf
- Probabilities: 0.1, 0.2, 0.5
Spreading activation algorithm

- Origin node is annotation, activates with energy=1
- Next, nodes 1 hop away are activated, then 2 hops, etc
- Nodes will activate if energy meets threshold (e.g. 0.2)
  - Because edges are a measure of supposed semantic connectedness, lower energy $\Rightarrow$ less semantic relevance
- Given edge (A,B), activation score of B:

$$AS(B) = AS(A) \times \text{weight}(edge(A,B))$$
Agenda

• The problem of photo retrieval
• Traditional methods for improving robustness
• Using world semantics to improve robustness
• OMCS: A corpus of commonsense
• Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
• **Mechanisms for managing noise and ambiguity**
• Conclusions and future work
Mechanisms for managing noise and ambiguity

- **Sieve Grammar**
  - Unconstrained sentence blobs are normalized
- **Concept Mapper**
  - Morphological stripping
  - Lemmatisation
- **Reweighting the graph to keep semantic “focus”**
  - Clustering/reinforcement
  - inverse popularity heuristics
- **“Fail-soft” applications.**
Reweighting edges

- 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)

\[ \text{newWeight} = \text{oldWeight} \times \text{discount} \]

\[ \text{discount} = \frac{1}{\log(\alpha \times \text{branchingFactor} + \beta)} \]
More Example Expansions

>>> expand('london')
('england', '0.9618')
('ontario', '0.6108')
('europe', '0.4799')
('california', '0.3622')
('united kingdom', '0.2644')
('forest', '0.2644')
('earth', '0.1244')

>>> expand("symphony")
('concert', '0.5')
('piece of music', '0.4')
('theatre', '0.2469')
('screw', '0.2244')
('concert hall', '0.2244')
('xylophone', '0.1')
('harp', '0.1')
('viola', '0.1')
('cello', '0.1')
('wind instrument', '0.1')
('bassoon', '0.1')
('bass fiddle', '0.1')
Agenda

- The problem of photo retrieval
- Traditional methods for improving robustness
- Using world semantics to improve robustness
- OMCS: A corpus of commonsense
- Building a world semantic resource from a commonsense corpus and using this resource to improve the robustness of photo retrieval
- Mechanisms for managing noise and ambiguity
- **Conclusions and future work**
Conclusions

• A generic commonsense corpus, OMCS, was used to make a user interface more adaptable and flexible.

• OMCS was noisy and ambiguous and efforts were made to clean up the corpus
  – Mapping rules
  – Concept mapper
  – Clustering and popularity heuristics

• How do u?
Future Work

- Make an evaluation
- Incorporate word-sense tagged OMCS 2
- Migrate from pattern matching parser to broad-coverage parser