Solving Semantic Analogy Problems Using ConceptNet

Mika Braginsky and Will Whitney

December 13, 2012

Summary

We implemented a system for solving SAT analogy problems, using ConceptNet. The system attempts to find the relationship between the target pair of words (such as *mason*:*stone*) and each option pair (such as *teacher*:*chalk*, *carpenter*:*wood*, *soldier*:*gun*, *photograph*:*camera*, *book*:*word*), scores the similarity of each option's relationship to the target's relationship, and selects the option with the highest score. On a dataset of 374 questions, it achieves an accuracy rate of 29.7%, with an accuracy rate of 47% on the 99 questions for which it can obtain at least minimal information.

1   Problem Overview

Reasoning by analogy has long been recognized in the cognitive sciences as a key component of cognition, in that it forms the basis for processes such as decision making, object recognition, face perception, problem solving, and creativity. Modeling this type of reasoning and enabling computer systems to make use of it should therefore be an important step in enabling artificial intelligence systems to emulate vital aspects of cognition.

As a way of exploring the more general phenomenon of reasoning by analogy, we implemented a system for solving analogy questions from the SAT (a standardized test used for college admissions). SAT analogy questions are specifically designed to test students' skill in solving problems using analogy-driven logic, so they are a reasonable test base for evaluating the ability of a computer system to reason by analogy.

An SAT analogy question consists of a target pair of words and five option pairs of words. The task is to select the option pair that "best expresses a relationship similar to that expressed in the original pair", as stated in the test's directions. For example:

```
ostrich:bird
```
(a) `lion:cat`

(b) `goose:flock`

(c) `ewe:sheep`

(d) `cub:bear`

(e) `primate:monkey`

To solve the problem, one must determine the relationship between each word pair:

```
An OSTRICH is a species of BIRD
```
(a) `LION is a species of CAT`

(b) `FLOCK is a group of GOOSE`

(c) `EWE is a female SHEEP`

(d) `CUB is a young BEAR`

(e) `MONKEY is a sub-group of PRIMATE`

Solving the problem requires a lot of semantic knowledge about how concepts are related to each other. In this question, the solver needs to not only know what ostriches and birds and lion and cats are, but also how ostriches are related to birds, how lions are realted to cats, etc. This knowledge must be also capture a rather specific level of detail – if the solver only knew that an ostrich is a type of bird, it would conclude that the same relation holds not just between lion and cat, but also between ewe and sheep and between cub and bear.

In general, the solver must make use of extensive information about the relationships between concepts and make inferences about the similarities between relationships. Achieving competence on this task, then, is an indicator of considerable mastery of reasoning abilities and common sense knowledge, both for a human and a computer system.

2 Previous Work

A number of algorithms have been devised to approach this problem (Table 1 shows their accuracy on a common test of 374 questions). Broadly speaking, they fall into two categories: lexicon-based and corpus-based (some are hybrid, a mix of both). The lexicon-based algorithms tend to use information from WordNet, a database of English words that encodes many lexical and some semantic relationships. The corpus-based algorithms use a corpus of questions to train their algorithm for the problem.

While some of the previous algorithms achieve almost human-level performance, they don't faithfully capture the nature of a human approach to solving the problem. Humans doesn't need to be "trained" on hundreds of questions before being able to answer further ones, and humans utilize much more complex and extensive semantic knowledge than can be provided by WordNet.

Reference for algorithm	Type	Correct
Random guessing	*Random*	*20.0%*
Jiang and Conrath (1997)	Hybrid	27.3%
Lin (1998)	Hybrid	27.3%
Leacock and Chodrow (1998)	Lexicon-based	31.3%
Hirst and St.-Onge (1998)	Lexicon-based	32.1%
Resnik (1995)	Hybrid	33.2%
Turney (2001)	Corpus-based	35.0%
Mangalath et al. (2004)	Corpus-based	42.0%
Veale (2004)	Lexicon-based	43.0%
Bicici and Yuret (2006)	Corpus-based	44.0%
Herdadelen and Baroni (2009)	Corpus-based	44.1%
Turney and Littman (2005)	Corpus-based	47.1%
Turney (2012)	Corpus-based	51.1%
Bollegala et al. (2009)	Corpus-based	51.1%
Turney (2008)	Corpus-based	52.1%
Turney (2006a)	Corpus-based	53.5%
Turney (2006b)	Corpus-based	56.1%
Average US college applicant	*Human*	*57.0%*

Table 1: Results of various algorithms on a common 374 question data set.[1]

---

[1] http://aclweb.org/aclwiki/index.php?title=SAT_Analogy_Questions_%28State_of_the_art%29

3   Approach

Attempting to solve analogy problems in a more plausibly human-like way requires access to a great deal of semantic information. Our approach was to obtain this information using ConceptNet, a semantic network that attempts to encode "common sense" knowledge and other information that computers should know about the world[2].

ConceptNet is a directed graph where the nodes are concepts and the edges are relationships between concepts. Its knowledge is assembled from a variety of sources, such as Wikipedia, Wiktionary, WordNet, and direct human contributions. It contains basic semantic relationships and everyday knowledge, such as:

dog —IsA→ pet

learn —MotivatedByGoal→ know more information

as well as more specialized cultural and scientific knowledge:

saxophone —UsedFor→ jazz

plutinos —can cross→ Neptune's orbit

If ConceptNet were identical to the full space of concepts and relationships known to a human, using its knowledge would allow a system to solve analogy questions as well as human. Such a system would determine the relationship between each word pair by finding the path from one word to the other, and then select an answer by picking the option with the same path (relationship) as the target pair.

However, ConceptNet is far from perfect – in trying to find the relationship between two concepts, ConceptNet has both not enough data and noisy data. On one hand, it has not enough data in that it just doesn't know enough about some concepts, such that finding a path between them is impossible. At the same time, its relationships are often very broad, so for some concepts it finds many paths between them. For example, for the concepts "note" and "music", it finds that:

---

[2]ConceptNet is developed by the Digital Intuition group, MIT Media Lab.

```
music —have or involve→ note

note —RelatedTo→ music

music —HasProperty→ note

note —in→ music
```

In light of noisy and insufficient data, our approach was to:

- Find the list of paths for the target pair

- Find the list of paths for each option pair

- For each option, compare the option's list of paths to the target's list of paths. Score their similarity using a scoring function.

- Pick the option with the highest similarity score

The closer that ConceptNet's knowledge is to a human's knowledge, the more similar this algorithm is to a human's problem solving approach.

4 Implementation

TODO: explain our imeplementation: queries to ConceptNet, parallelization, finding paths, similarity metric

5 Results

TODO: show our results, discuss error types

6 Further Work

TODO: give options of ways this could be improved/extended