Mangalath, P., Quesada, J., & Kintsch, W. (2004). Analogy-making as Predication Using Relational Information and LSA Vectors. In K. D. Forbus , D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. Chicago: Lawrence Erlbaum Associates.

Analogy-making as Predication Using Relational Information and LSA Vectors

Praful Mangalath, José Quesada, Walter Kintsch ([praful , quesadaj, wkintsch]@psych.colorado.edu)
Institute of Cognitive Science, University of Colorado, Boulder
Boulder, CO 80309-0344 USA

Current models of analogy comprehension use hand - coded representations. As Hummel and Holyoak (2003) put it, "the problem pf hand-coded representations is among the most serious problems facing computational modeling as a scientific enterprise: All models are sensitive to their representations, so the choice of representation is among the most powerful wild cards at the modeler's disposal" (p. 247). French (2002) reviews different computational models of analogy-making, and points out one of the most fundamental problems of the field: case representations are authored (hand-coded) to make the model work. Between the challenges and future directions he presents "the systematic exploration of experimenter-independent representation-building and learning mechanisms" (p. 204).

In this poster, we propose LSA as a method to generate the much-wanted non-hand-coded representations. However, LSA has severe limitations to represent structure. Turney and Littman (2003) pointed out that the similarity of *semantic relations* between words is not directly reducible to the semantic similarity of individual words. This is also the leitmotiv of some analogy models like Gentner's (1983; 1989).Thus, LSA alone would fail to explain analogy, where relations (structure) between words are fundamental. We use a predication (Kintsch, 2001) to represent structure comparisons in the LSA semantic space. Predication is able to select the features (neighbors) of one component of the analogy (the source) that are relevant to the other (the target).

Table 1(a): a sample SAT question.

	Ostrich : bird
(a)	Lion : Cat
(b)	Goose : Flock
(c)	Ewe : Sheep
(d)	Cub : Bear
(e)	Primate: Monkey

Table 1(b): predication using analogy domains

	Number	Percent	T&L (2003)
Correct	157	41.20%	47.10%
Incorrect	150	40.10%	51.60%
Skipped	67	17.20%	1.30%
Total	374	100%	100.00%
Precision	157/307	0.51%	47.70%
Recall	157/374	0.42%	47.10%
F		0.46%	47.10%

We calculated the predication vectors for all the targets and alternatives of 374 items from the Scholastic aptitude test (SAT). This dataset of analogies was collected by Turney and Littman (2003). An example of a SAT item can be seen in Table 1(a). To calculate the correct alternative, we computed the cosine between the target vector and each alternative, and selected the alternative with the highest cosine. However, this method had poor results: using LSA this way leaves out most of the relational information. For example, relations such as is-a, part-of, causal-agent-of, etc. are all substituted by a very basic semantic distance measure when we compute the cosine between the target and the alternatives. To include this relational information in the comparison, we constructed a set of ten possible relations between the components in the 374 SAT analogies (table 2). Then we computed the cosine between the list of words that define the analogy domain and each analogy predication vector in the dataset. That is, for each analogy we created a vector of ten features, where each feature indicates how similar the analogy is to each of the analogy domains. For example, Ostrich::bird would load primarily in the taxonomy and Hyponymy domain components, but also in endonymy, synonymy, and degree. Then, we correlated these loading vectors for the target and each alternative, and selected the alternative that best correlated with the target to solve the SAT question.

Table 2: Ten analogy domains and their characteristic words

Hyponymy	X is a type of Y (for example - Maple:Tree)	[Subordinate of, superordinate to, rank, class, category, family, genus, variety, type of, kind of, hyponym]
Degree	X means Y at a certain degree (Pour:Drip)	[level, stage, point, magnitude, extent, greater, lesser, intensity, severity, extreme, degree]
Meronymy	The parts of X include the Ys (Body:Arm)	[part, whole, component, made up of, portion, contains, constituent, segment, piece of, composite, meronym]
Taxonomy	X is an item in the category Y	(Milk:Beverage)[classification, containing, structure, relationship, hierarchy, system, framework, taxonym]
Synonymy	is the same as Y (Work:Labor)	[equivalent, equal, likeness, match, interchangeable, alike, same as, similar, close to, like, synonym]
Antonymy	is the opposite of Y (Find:Hide)	[opposite, unlike, different, antithesis, opposed, contradiction, contrast, reverse, anti, not the same as, antonym]
Characteristic	X is a characteristic of Y (Dishonesty:Liar)	[indicative, representative of, typical of, feature, attribute, trait, property, mannerism, facet, quality, characteristic]
Plurality	X is many Ys (Throng:People)	[mass, bulk, several, many, lots of, numerous, crowd, group, more, number, plural]
Endonymy	X entails Y (Coop:Poultry)	[entails, require, evoke, involve, suggest, imply, presuppose, mean]
Use	X is used to Y (Scissors:Cut)	[do with, manipulate, operate, function, purpose, role, action, utilize, employ, use]

The results are displayed in Table 1(b). The performance of our model is very close to the state of the art in automatic analogy making when considering correct answers (42% vs. 47%, Turney & Littman, 2003), and precision, recall and F measures. Furthermore, our model is psychologically plausible.