Analogical Integration of Semantic Roles with Vector Symbolic Architectures

Simon D. Levy (levys@wlu.edu)
Department of Computer Science, 407 Parmly Hall
Washington and Lee University
Lexington, VA 24450 USA

Abstract

Semantic roles describe "who did what to whom" and as such are central to analogy processing and other cognitive processes. For analogy processing, roles are usually specific to each predicate: for *loves* there is a LOVER and a BELOVED, for *eats* an EATER and an EATEN, etc. Language modeling, on the other hand, requires more general roles like AGENT and PATIENT in order to relate form to meaning in a parsimonious way. This paper presents a new model of semantic roles that addresses this dichotomy. The model uses a distributed representation scheme called Vector Symbolic Architectures (VSA) for representing roles and their fillers. Starting with specific roles, the model learns to generalize roles through exposure to language data, through a process that is itself analogical. The learning mechanism is simple and efficient, and its scaling properties are well-understood. The model is able to learn and exploit new representations without losing the information from existing ones. The contribution of the model to the study of analogy is thus twofold: it shows how representations needed for analogy processing can be accommodated within a more general theory of semantic roles, and suggests how important analogy may be to language learning. We present experimental data illustrating these principles, and conclude by discussing some implications for the relation between analogical processing and language.

KEYWORDS: *Analogy, learning, abstraction, distributed representations, semantic roles, language acquisition, grammatical constructions, Sapir-Whorf Hypothesis*

The Centrality of Semantic Roles

In modeling language and thought, perhaps no other concept has been as crucial as that of semantic roles. From practically-motivated work in artificial intelligence (Schank, 1972) to philosophical discussions of metaphor (Lakoff & Johnson, 1980) to theory-driven approaches in linguistics (Fillmore, 1968), the importance of representing "who did what to whom" has never been in dispute.

The manner in which such roles are specified depends largely on the phenomena of interest, as well as the implementation architecture. Analogy models built on predicate calculus (Falkenhainer, Forbus, & Gentner, 1990) rely mainly on positional notation; *e.g.*, loves(mary, john), sees(bill, fred). Some connectionist models of analogy, such as (Hummel & Holyoak, 2005) and (Eliasmith & Thagard, 2001), make use of predicate-specific roles. In such models these propositions would be represented as loves(LOVER=mary, BELOVED=john), sees(SEER=bill, SEEN=fred), with each capital-letter role implemented by a different unit or set of units, and the equal sign representing some sort of role/filler binding operation (discussed in more detail below). In general, one gets the sense that cognitive models of analogy and related processes view roles as specific to predicates and relations. This specificity is, to a large extent, what makes the problem of analogy interesting.

From a linguistic perspective, however, the story is quite different. In addition to the theoretical interest in characterizing the nature and relationships of thematic roles across the world's languages (Fillmore, 1968), there is a practical advantage to constraining the number of roles posited for a given language: without abstract roles like AGENT, PATIENT, INSTRUMENT, *et al.*, it is difficult to express generalizations relating form to meaning. For example, any rule-based system for generating or understanding English declaratives will need to express the generalization that the default order is AGENT-PREDICATE-PATIENT. With roles specific to each predicate, a different rule has to be stated for each new predicate. It is, however, clear that language users can exploit roles to build new constructions, even without any specific information about the meaning of a verb. Told that *Mary glorped John*, English speakers can construct related sentences like *John was glorped by Mary*, *Who did Mary glorp?*, etc.

Role Integration as Analogy

The present work takes its inspiration from the framework of Cognitive Grammar (Langacker, 1987), with its emphasis on general cognitive mechanisms and learning from examples. In such a framework, one sort of role conceptualization is seen as developmentally prior to, but crucially related to, the other. Because children can reason about the world well before they acquire adult-level language competence, many cognitive grammarians (notably (Tomasello, 1992)) favor a view in which general roles emerge from predicate-specific roles through exposure to language. Although this is the approach taken in the experiment described below, it is important to emphasize that the main issue is relating and integrating the two sorts of roles using a single mechanism. The model makes no commitment to a particular direction of change (specific $\rightarrow$ general or general $\rightarrow$ specific), and can accommodate either.

In English and many other languages, information about thematic roles is conveyed primarily through word order (*dog bites man* vs. *man bites dog*). As detailed in (Tomasello, 1992), adults' consistent use of word order provides the language learner with an opportunity to generalize roles across different predicates, once the the learner has acquired the concepts associated with particular words, plus the general lexical categories *noun* and *verb*. The fundamental insight on which our model is based is that *this process is itself analogical*. In a way reminiscent of visual analogy processing (Forbus & Tomai, 2005), repeated examples of correspondences between representational components (*noun/filler*, *verb/predicate*) supports integration of different structures into a single conceptual representation.

For example, a child learning English will be provided with numerous consistent associations between simple declarative sequences of the form

$$noun_i \ldots verb_j \ldots noun_k \tag{1}$$

and conceptual structures of the form

$$predicate_j(role_{j,l} = entity_i,\ role_{j,m} = entity_k) \tag{2}$$

The child eventually induces the generalized form

$$predicate_j(role_l = entity_i,\ role_m = entity_k) \tag{3}$$

where $role_l$ is AGENT and $role_m$ PATIENT. A similar story obtains for "free word order" languages like Russian and Japanese, with noun endings taking the place of word order in conveying thematic roles.

Issues in Representation

Cognitive models are often categorized in terms of the *connectionist* vs. *symbolic* distinction. In addition to being descriptively unhelpful, these terms are also typically conflated with a host of issues that may have nothing to do with the commitments entailed by a particular model.

A more useful distinction among cognitive representations is whether they are *local* or *distributed* (vanGelder, 1999). Traditional symbol systems (grammar, predicate calculus) use local representations: a given symbol has no internal content and is located at a particular address in memory. Although well-understood and successful in a number of domains, such representations are biologically unrealistic and suffer from brittleness. The number of possible items to be represented is fixed at some arbitrary hard limit, and a single corrupt memory location or broken pointer can wreck an entire structure. Similar difficulties arise in feature-based representations, and in connectionist models based on "grandmother cell" representations, like temporal synchrony networks (Shastri & Ajjanagadde, 1993).

In a distributed representation, on the other hand, *each entity is represented by a pattern of activity distributed over many computing elements, and each computing element is involved in representing many different entities* (Hinton, 1984). In addition to being more obviously similar to the way that the brain seems to represent concepts, such representations have a number of appealing properties for cognitive modeling (McClelland, Rumelhart, & Hinton, 1986): they are robust to noise, provide realistically "soft" limits on the number of items that can be represented at a given time, and support distance metrics. These properties enable fast associative memory and efficient comparison of entire structures without breaking down the structures into their component parts – precisely the sorts of operations that are useful in modeling analogy.

Distributed Representation of Role/Filler Bindings

A number of schemes have been developed for constructing distributed representations of role/filler bindings. The methods that have enjoyed the most success recently can be grouped under the heading *Vector Symbolic Architectures*, or VSAs (Gayler, 2003). VSA's rely on large vectors of random numbers to represent both roles and fillers, and use a fast, lossy compression function to bind a role and a filler representation into a vector of the same size as the role and the filler. Given a vector representing a role/filler binding, the filler (role) can be retrieved by applying the inverse of the compression function to this vector and the role (filler) vector. Because the compression function is lossy, retrieval produces a "noisy" version of the original vector. The noisy vector can be "cleaned up" by an associative memory, the simplest (but least neurally plausible) implementation of which is a table that stores all the original vectors and returns the one closest to the noisy version. If more fine-grained representations are needed, features can be implemented as random vectors that are summed element-wise to produce a vector representing the combination of the features. This summing operation can also be used to "bundle" different bindings into a coherent whole. For example, the proposition expressed by *Mary loves John* can be represented as

$$\mathsf{JOHN} + \mathsf{MARY} + \mathsf{LOVES} + \mathsf{LOVER} \circledast \mathsf{MARY} + \mathsf{BELOVED} \circledast \mathsf{JOHN} \tag{4}$$

where each capitalized word stands for the vector representation of the corresponding role or filler, $+$ is vector addition, and $\circledast$ is the binding operation. Bundling the representations of JOHN, MARY, and LOVES together with the role/filler bindings supports surface-level similarity of the sort observed in psychological experiments on analogy (Gentner & Toupin, 1986). By using bundles as fillers, recursive structures of arbitrary complexity can be built, subject to the "soft limit" phenomenon mentioned above.

The most popular variety of VSA is the Holographic Reduced Representation, or HRR (Plate, 2003), in which binding is implemented as circular convolution and unbinding as circular correlation (both of which can be performed using the Fast Fourier Transform). HRR has been used successfully in modeling analogy, with results similar to those found in experiments with human subjects (Eliasmith & Thagard, 2001).

Learning Thematic Roles with Holographic Representations

Because the concept of binding is general enough to encode sequencing information (absolute position of a word or its position relative to another), HRR can also be used to model word order in a psychologically realistic way (Jones & Mewhort, 2007). Based on this insight, we conducted two simple experiments to simulate learning of thematic roles from word order. These experiments tested the ability of an HRR-based model to generalize thematic roles within a fixed set of predicates and arguments, while maintaining the information from predicate-specific roles useful for processing analogies.

The training set consisted of HRR's for 50 two-place predicates like LOVES and SEES, and 10 fillers like JOHN and MARY. A simplifying constraint was that no predicate was allowed to have the same filler in both roles. The arrangement resulted in 4500 ($= 50 \times 10 \times 9$) unique propositions. Predicates, fillers, and roles were represented as 3000-dimensional vectors of Gaussian-distributed random numbers with mean of zero and a standard deviation of $\sqrt{1/3000}$, following the suggestion in (Plate, 2003). Propositions were constructed as in example (4) above. To simulate exposure to word-order consistencies as in (1) and (2) above, we used a random vector v_1 to represent the order *noun ... verb* and another vector v_2 to represent *verb ... noun*. The initial propositional HRR structures were rebuilt by replacing the predicate-specific first role with itself plus v_1, and the second role with itself plus v_2, and re-binding these summed roles to their fillers.

Experiment 1: Generalizing Thematic Roles

To model generalization of thematic roles, we took the vector average of all the integrated agent roles, thereby obliterating predicate-specific information. This average agent role was used as the cue to retrieve a filler from the HRR of each proposition. We did the same for the patient roles. Successful retrieval implies the ability to use a given proposition in various grammatical constructions, including ones to which the model has not been exposed. Retrieval was determined to be successful when the vector cosine of the (noisy) retrieved filler was highest with the original filler used to build the HRR for the proposition (as opposed to any of the other 10 possible fillers). Because we required both fillers to be retrieved successfully, chance performance on this task is 1%. In 10 trials with different sets of random vectors, we achieved a mean success of 99.8%, with a standard deviation of 0.5%. This result shows that the fillers of thematic roles were successfully recovered despite these roles being integrated with the analogical roles.

Experiment 2: Retrieving Analogical Roles

Analogy processing requires any representation of propositional content to support querying the concrete role played by a given entity. A representation of *Mary loves John*, for example, must report Mary's role as LOVER, and not merely something general like thematic role AGENT. To test the retrieval of such roles, we used the fillers in each proposition as retrieval cues. As in the experiment above, the results were highly successful: both of the analogical (specific) roles were retrieved correctly 97.5% of the time, with a standard deviation of 5% over the 10 trials (chance $<< 1\%$). As in the first experiment, the integration of thematic and analogical components into a single representation did not adversely affect the ability to retrieve crucial information.

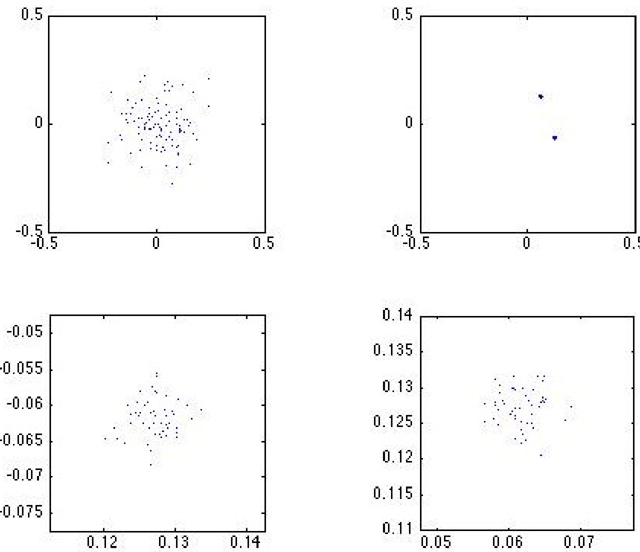

Figure 1: Principal Components Analysis of role vectors. Top left: before roles. Top right: after addition of theta roles. Bottom: detail from top right.

Discussion

The two experiments reported above show that our model supports both analogical and linguistic processes in a single distributed representation. Some critics of distributed representations have expressed concern about the opacity of their behavior; e.g., the potential difficulty of assigning credit to the appropriate component of a system in which information is distributed among all representational elements (Minsky & Papert, 1988). Dealing with this issue will likely require a more sophisticated notion of "component" than the one offered in the traditional cognitive science literature. Figure 1 shows a principal components analysis of role vectors from the experiments, before and after linguistic role integration. (The axes correspond to the two dimensions with the highest eigenvalues.) As the top right image in the figure suggests, there is indeed a strong internal structure to the integrated VSA representations, which reveals how they support the behavior found in the experiments. At a gross level, there are two broad role categories, corresponding to the linguistic roles AGENT and PATIENT. A closer look at this structure (bottom images) reveals a level of detail beyond the high-level agent/patient distinction visible in the middle image: at a more fine-grained level, there are still as many analogical role categories as there were originally, consistent with the results of the second experiment.

Conclusions and Future Work

Anyone who has struggled through Latin can recall the following experience: one first learns a few simple associations between a handful of noun cases (nominative, accusative, *et al.*) and thematic roles, but soon confronts the bewildering variety of uses made of the cases: dative of agent, ablative absolute, duration-of-time accusative, etc. This situation highlights exactly the issue addressed in the work described above: there is a complex, nuanced relation between the small set of abstract thematic roles in a given language and the multiplicity of roles needed for real-world cognitive tasks like analogy processing.

As compared with traditional discrete (localist) symbol systems like grammars, vector symbolic architecture offers a more supple, psychologically plausible basis for representing this kind of subtlety. Further, this representation can seamlessly integrate linguistic information into existing, prelinguistic representations, opening the door to exploring longstanding hypotheses on the possibility that language can influence thought (Whorf, 1939/2001).

Missing in our story is an explicit mechanism for transducing role representations to/from word sequences. Modeling these mechanisms in a neurally realistic manner is an active research topic (Dominey, Hoen, & Inui, 2006). The work presented here represents a first step toward integrating such "surface-level" information with a sufficiently rich semantic representation in a neurally plausible way.

As one reviewer of an earlier draft of this paper pointed out, we have not provided any examples of how our VSA-based model compares with other analogy models. A serious comparison like this is beyond the scope of this paper; however, we have noted that VSA models have been applied successfully to analogy problems (Eliasmith & Thagard, 2001). Although performance compared favorably with non-distributed or partially-distributed model, some researchers have cast doubt on the ability of distributed representation models to scale up to real-world problems like visual analogies in sketches (Forbus & Tomai, 2005). This criticism provides an interesting, concrete challenge for the VSA approach. The results reported in (Eliasmith & Thagard, 2001) were based on unrealistically small (512 element) vectors, presumably due to hardware limitations at the time the research was conducted. Mathematical analysis of convolution memories in the appendices to (Plate, 2003) show representational capacity that scales linearly with vector size. This sort of scalability suggests an experiment applying realistically-sized VSAs to the visual analogies task, to see whether Moore's Law has taken care of this issue.

Acknowledgments

The author thanks Ross Gayler, Harry Howard, Pentti Kanerva, and Tony Plate for inspiration and useful suggestions, and two anonymous reviewers for helpful criticism.

References

Dominey, P., Hoen, M., & Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, *18*, 2088-2107.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: a distributed model of analogical mapping. *Cognitive Science*, *25*(2), 245-286.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1990). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

Fillmore, C. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (p. 1-90). Holt, Rinehart.

Forbus, K. D., & Tomai, J. M. U. E. (2005). Analogical learning of visual/conceptual relationships in sketches. In *Aaai 2005* (p. 202-208).

Gayler, R. (2003). Vector symbolic architectures answer jackendoffs challenges for cognitive neuroscience. In P. Slezak (Ed.), *Iccs/ascs international conference on cognitive science* (p. 133-138). Sydney, Australia: University of New South Wales.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277-300.

Hinton, G. (1984). *Distributed representations* (Tech. Rep. No. CMU-CS-84-157). Computer Science Department, Carnegie Mellon University.

Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the lisa project. *Current Directions in Pychological Science*, *14*(3), 153-157.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford University Press.

McClelland, J., Rumelhart, D., & Hinton, G. (1986). The appeal of parallel distributed processing. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). MIT Press.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons*. MIT Press.

Plate, T. (2003). *Holographic reduced representation: Distributed representation for cognitive science*. CSLI Publications.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, *3*(4), pages 532–631.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning. *Behavioral and Brain Sciences*, *16*(3), 417-494.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

vanGelder, T. (1999). Distributed versus local representation. In *The mit encyclopedia of cognitive sciences* (p. 236-238). Cambridge, Massachusetts: MIT Press.

Whorf, B. L. (1939/2001). The relation of habitual thought and behavior to language. In J. A. Duranti (Ed.), *Linguistic anthropology: A reader.* Oxford, UK: Blackwell.