

# Distributed Representation of Compositional Structure

Simon D. Levy

Washington and Lee University, USA

## INTRODUCTION

AI 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 asks whether they are *local* or *distributed* (van Gelder 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, traditional representations 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.

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). Such representations have a number of properties that make them attractive for knowledge representation (McClelland, Rumelhart, & Hinton 1986): they are robust to noise, degrade gracefully, and support graded comparison through distance metrics. These properties enable fast associative memory and efficient comparison of entire structures without unpacking the structures into their component parts.

This article provides an overview of distributed representations, setting the approach in its historical context. The two essential operations necessary for building distributed representation of structures—*binding* and *bundling*—are described. We present example applications of each model, and conclude by discussing the current state of the art.

## BACKGROUND

The invention of the backpropagation algorithm (Rumelhart, Hinton, & Williams 1986) led to a flurry of research in which neurally inspired models were applied to tasks for which the use of traditional AI data structures and algorithms were commonly assumed to be the only viable approach. A compelling feature of these new models was that they could “discover” the representations best suited to the modelling domain, unlike the manmade representations used in traditional AI. These discovered or learned representations were typically vectors of numbers in a fixed interval like  $[0, 1]$ , representing the values of the hidden variables. A statistical technique like principal component analysis could be applied to such representations, revealing interesting regularities in the training data (Elman 1990).

Issues concerning the nature of the representations learned by backpropagation led to criticisms of this work. The most serious of these held that neural networks could not arrive at or exploit systematic, compositional representations of the sort used in traditional cognitive science and AI (Fodor & Pylyshyn 1988). A minimum requirement noted by critics was that a model that could represent e.g. the idea *John loves Mary* should also be able to represent *Mary loves John* (systematicity) and to represent *John*, *Mary*, and *loves* individually in the same way in both (compositionality). Critics claimed that neural networks are in principle unable to meet this requirement.

Systematicity and compositionality can be thought of as the outcome of two essential operations: *binding* and *bundling*. Binding associates fillers (*John*, *Mary*) with roles (lover, beloved). Bundling combines role/filler bindings to produce larger structures. Crucially, representations produced by binding and bundling must support an operation to recover the fillers of roles: it must be possible to ask “Who did what to whom?” questions and get the right answer. Starting

around 1990, several researchers began to focus their attention on building models that could perform these operations reliably.

**VARIETIES OF DISTRIBUTED REPRESENTATION**

This article describes the various approaches found in the recent neural network literature to implementing the binding and bundling operations. Although several different models have been developed, they fall into one of two broad categories, based on the way that roles are represented and how binding and bundling are performed.

**Recursive Auto-Associative Memory**

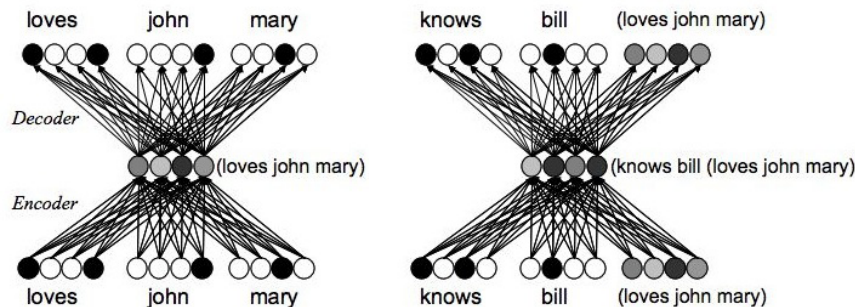
In Recursive Auto-Associative Memory, or RAAM (Pollack 1990), fillers are represented as relatively small vectors ( $N=10-50$  elements) of zeros and ones. Roles are represented as  $N \times N$  matrices of real values, and role/filler binding as the vector/matrix product. Bundling is performed by element-wise addition of the resulting vectors. There are typically two or three role matrices, representing general role categories like agent and patient, plus another  $N \times N$  matrix must to represent the predicate (*loves, sees, knows, etc.*). Because all vectors are the same size  $N$ , vectors containing bindings can be used as fillers, supporting structures of potentially unlimited complexity (*Bill knows Fred said John loves Mary.*) The goal is to learn a set of matrix values (weights) to encode a set of such structures.

In order to recover the fillers, a corresponding set of matrices must be trained to *decode* the vectors produced by the encoder matrices. Together, the encoder and decoder matrices form an autoassociator network (Ackley, Hinton, & Sejnowski 1985) that can be trained with backpropagation. The only additional constraint needed for backprop is that the vector/matrix products be passed through a limiting function, like the sigmoidal “squashing” function  $f(x) = 1 / (1 + e^{-x})$ , whose output falls in the interval (0,1). Figure 1 shows an example of autoassociative learning for a simple hypothetical structure, using three roles, with  $N = 4$ . The same network is shown at different stages of training (sub-tree and full tree) during a single backprop epoch. Note that the network devises its own compositional representations on its intermediate (“hidden”) layer, based on arbitrary binary vectors chosen by the experimenter. Unlike these binary vectors (black and white units), the intermediate representations can have values between zero and one (greyscale).

Once the RAAM network has learned a set of structures, the decoder sub-network should be able to recursively unpack each learned representation into its constituent elements. As shown in Figure 2, decoding is a recursive process that terminates when the decoder’s output is similar enough to a binary string and continues otherwise. In the original RAAM formulation, “similar enough” was determined by thresholds: if a unit’s value was above 0.8, it was considered to be on, and if it was below 0.2 it was considered to be off.

RAAM answered the challenge of showing how neural networks could represent compositional structures in a systematic way. The representations discovered by RAAM could be compared directly via distance metrics,

Figure 1. Learning the structure (*knows bill (loves john mary)*) with RAAM



4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/distributed-representation-compositional-structure/10295](http://www.igi-global.com/chapter/distributed-representation-compositional-structure/10295)

## Related Content

---

### Factors Influencing Patient Adoption of the IoT for E-Health Management Systems (e-HMS) Using the UTAUT Model: A High Order SEM-ANN Approach

Manish Dadhich, Kamal Kant Hiran, Shalendra Singh Rao and Renu Sharma (2022). *International Journal of Ambient Computing and Intelligence* (pp. 1-18).

[www.irma-international.org/article/factors-influencing-patient-adoption-of-the-iot-for-e-health-management-systems-e-hms-using-the-utaut-model/300798](http://www.irma-international.org/article/factors-influencing-patient-adoption-of-the-iot-for-e-health-management-systems-e-hms-using-the-utaut-model/300798)

### Epistemic Stewardship and Pedagogies of Reasoning in AI-Mediated Statistics Classrooms: Integrating Causal Identification Fairness Metrics

Ashraf Alam (2026). *Statistics Pedagogy in the Age of AI: Challenges, Strategies, and Global Case Studies* (pp. 143-178).

[www.irma-international.org/chapter/epistemic-stewardship-and-pedagogies-of-reasoning-in-ai-mediated-statistics-classrooms/411762](http://www.irma-international.org/chapter/epistemic-stewardship-and-pedagogies-of-reasoning-in-ai-mediated-statistics-classrooms/411762)

### Combining Supervised Learning Techniques to Key-Phrase Extraction for Biomedical Full-Text

Yanliang Qi, Min Song, Suk-Chung Yoon and Lori deVersterre (2011). *International Journal of Intelligent Information Technologies* (pp. 33-44).

[www.irma-international.org/article/combining-supervised-learning-techniques-key/50484](http://www.irma-international.org/article/combining-supervised-learning-techniques-key/50484)

### Android-Based Skin Cancer Recognition System Using Convolutional Neural Network

Sercan Demirci, Durmu Özkan and Ibrahim Halil Toprak (2021). *Diagnostic Applications of Health Intelligence and Surveillance Systems* (pp. 59-85).

[www.irma-international.org/chapter/android-based-skin-cancer-recognition-system-using-convolutional-neural-network/269029](http://www.irma-international.org/chapter/android-based-skin-cancer-recognition-system-using-convolutional-neural-network/269029)

### Analysing Twitter Data for Phishing Tweets Identification

Falah Hassan Ali Al-Akashi (2021). *International Journal of Intelligent Information Technologies* (pp. 1-11).

[www.irma-international.org/article/analysing-twitter-data-for-phishing-tweets-identification/277074](http://www.irma-international.org/article/analysing-twitter-data-for-phishing-tweets-identification/277074)