Compositionality in Neural Systems

Barbara Hammer

University of Osnabrück,
Department of Mathematics/Computer Science,
D-49069 Osnabrück, Germany,

e-mail: hammer@informatik.uni-osnabrueck.de,
phone:+49 541 969 2488,
fax: +49 541 969 2770

Introduction

In real life, people deal with composite structures: Written English language is built of 26 characters and a few additional symbols which form syllables, words, sentences, articles, roadmaps, and finally the 'Handbook of Brain Theory'. Spoken language consists of raw acoustic waves at a basic level; at a higher level, it can be decomposed into phonemes which form words, sentences, a poem or speech. Visual data can be decomposed into pixels with various colors and intensities; alternatively, the raw image data may be represented by features like edges or texture, which are grouped to complex contours, objects, and, finally, a complete scene like the image of our grandmother who sits knitting on a chair. Moreover, not only real life data are processed as composite objects, artificial data created by humans or virtual objects have a composite structure, either: Web sites consist of single pages with head and body, links, tabulars, figures, enumerations, *etc.* Computer programs decompose into procedures and functions, or objects and methods. Logical formulas and terms are recursive objects built of symbols for constants, variables, and functions, logical connectives, and quantors.

Speaking of neural systems, two questions arise:

(i) Artificial neural networks are developed in order to model important aspects of the human brain and to explain how biological neural networks process information; a common characteristic of the way in which data are created, processed, and stored by humans is compositionality. *Which neural dynamics allow composite structures, grouping and binding to emerge* in such a way that the single parts can be restored rapidly and, at the same time, the whole composite structure can be identified with a single object?

(ii) Artificial neural networks are a powerful and universal machine learning tool which is used in scientific and industrial applications for data contained in a finite dimensional vector space. Every day's data is composite. *How can we adapt standard neural techniques to composite structures* such that connectionistic methods can be used in these every day's areas of application and combined with other compositional machine learning tools?

Naturally, these two questions constitute extreme positions in a single problem spectrum: How is information processed in the brain? Since it is unlikely that models which are satisfactory with regard to both, efficiency and biological plausibility, will come about in the near future, the focus can lie on one of the two aspects: One can develop practically applicable and efficiently trainable approaches or one can design biologically plausible and universal systems. Most existing approaches lie somewhere in between, which is, of course, partially due to the fact that biological systems are indeed very effective, suggesting the universal principle of compositionality and dynamic binding, *e.g.* in the visual system (Bienenstock 1996) (see DYNAMIC LINK ARCHITECTURE).

Properties of Compositionality

What are general properties of composite objects like the sentences 'John loves Mary' or 'Mary loves that John loves Mary'? The structures are composed of *basic primitives*, here the words 'John', 'Mary', 'loves', and 'that'. The primitives are instances of a certain *type*, e.g. a noun or verb. Composite expressions arise if the primitives are combined in specified *relations.* The sentence 'John loves Mary' is an instance of a relation of the form 'subject predicate object', where 'subject' and 'object' can be instantiated with a specific noun, for example. Alternatively, these positions may be filled with composite objects such as 'John loves Mary' in the example 'Mary loves that John loves Mary'. The complexity of recursively generated structures is usually not limited a priori. As an example, we could build the sentences 'John loves that Mary loves that John loves Mary.' or 'Mary loves that John loves that Mary loves that John loves Mary.' and so on. Hence, as pointed out in van Gelder (1990), the combination of primitives of certain types within constituency relations yields composite structures of a priori *unlimited complexity and an infinite number of possible combinations.*

The *semantics* of composite objects is determined by the semantics of the simple primitives and their relation in the structure. The primitives alone do not determine the semantics. The sentences 'John loves Mary' and 'Mary loves John', for example, have identical constituents but different meanings. Moreover, the decomposition of complex objects into simpler parts and their interpretation may depend on the whole structure: the meaning of 'her' in the sentence 'Mary loves that John loves her' depends on the context of the part 'John loves her'. Conversely, the same situation may be described with different composite objects, such as 'Mary loves that John loves Mary' or 'Mary loves that John loves her'. *I.e.*, parts of a structure may be substituted by entirely different representations without affecting the semantics. Hence, the whole structure as well as the involved primitives and their relation should be available for referring the semantics.

Though an unlimited number of combinations of the primitives in constituency relations is possible in principle, commonly only *a small part* of all possible combinations is used in practice and makes sense. Either basic syntactic rules or semantic limitations restrict the variety of possible compositions: 'That John loves Mary loves Mary.' does not make sense, for example, where the composite structure 'That John loves Mary' is used as subject. Restrictions for possible combinations are often context dependent. For example, the sentence ' John loves himself' would be preferred to ' John loves John' unless the word 'John' refers to two different persons.

Humens are capable of *understanding and producing* composite objects. Moreover, they can deal with new complex structures although they have never seen the specific combination of primitives before. As an example, having read the above sentences concerning John and Mary and the sentence 'Mary loves Peter.', people would be capable of understanding the sentence 'John loves that Mary loves Peter.' – and they would possibly infer that this sentence is false. People can infer the meaning of partially new structures. They have knowledge about primitives, relations, and additional information like rules for the composition or experience with similar structures. Moreover, compositional structures play an important role if the capability of humans for analogical-based reasoning is investigated (see ANALOGY-BASED REASONING AND METAPHOR).

Appropriate artificial neural systems which are used for processing compositional data should take these properties into account. While they have to deal with an unrestricted amount of partially new and ambiguous data, they can use the data's sparsity and hierarchical structure for efficient processing.

Neural Systems and Compositionality

Popular neural methods perform pattern recognition for which classical statistics constitutes a well founded theory (see PATTERN RECOGNITION). Standard neural networks are adapted to real vectors contained in a finite dimensional Euclidian real-vector space. Concerning compositional structures, the question arises how to encode and process an arbitrary amount of information with this finite dimensional machinery.

Compared to their biological counterparts, artificial neural networks often neglect the fine temporal structure of spike trains (see OSCILLATORY AND BURSTING PROPERTIES OF NEURONS and ADAPTIVE SPIKE CODING): standard artificial networks process real values or binary values in a well defined topological order. Real values correlate to the mean spiking frequence of the biological neurons, hence the local temporal structure and respective correlation of spikes is not taken into account. Binary values might encode single spikes, although the topology allows processing at a discrete or fixed time scale only. Experiments provide evidence that the fine temporal structure may indeed carry important information for encoding complex scenes in biological networks (Engel et. al. 1992). However, the precise way in which information is encoded in the respective parts of the brain is not yet understood. Moreover, some effects can possibly already be explained at the abstract level of rate coding (see RATE CODING). Hence it is worth considering the whole spectrum of neural architectures capable of dealing with compositionality. Our taxonomy for characterizing the various approaches is based on the connection structure of the neural architectures and on their fundamental dynamical behavior.

A key property of appropriate systems is their capability of processing a priori unlimited information. *Static solutions* with feedforward networks have the advantage that efficient training algorithms are readily available (see BACKPROPAGATION – GENERAL PRINCIPLES AND ISSUES FOR BIOLOGY). Unfortunately, their capacity is either limited, or they need an a priori unlimited amount of neural resources.

Recurrent neural networks use the additional degree of freedom provided by a priori unlimited processing time in order to map the information appropriately. Recurrent networks may be either *partially recurrent* where the processing dynamics is determined by the respective data structure. The restriction to limited recursive data structures allows the immediate generalization of efficient standard learning tools (see RECURRENT NETWORKS: SUPERVISED LEARNING). Alternatively, the networks may be *fully recurrent*; then the dynamics is determined by the respective process. This allows more flexibility, but the processing time cannot be limited a priori and alternative learning algorithms are necessary (see TEMPORAL SEQUENCES: LEARNING AND GLOBAL ANALYSIS). In particular, continuous time fully recurrent neural networks can use the fine temporal structure for encoding complex information such as in networks of spiking neurons (see INTEGRATE-AND-FIRE NEURONS). These approaches are biologically plausible and flexible. Unfortunately no efficient learning algorithm comparable to standard backpropagation for feedforward networks has been established for networks of spiking neurons until today.

Static Approaches

The *single neuron doctrine* assumes that each simple or composite object is represented by the activity of a specific neuron for this object. Neurons for complex objects are hierarchically connected to neurons representing their simple parts and they only become active if all neurons for their parts become active. Such 'grandmother cells' have been found in the visual system of the macaque monkey, for example (see SPARSE CODING IN THE PRIMATE CORTEX). Some invariance, for example with respect to the specific lighting, may be involved in the hierarchical computation, such that the neuron which encodes our grandmother fires independent of the specific context. Naturally, this approach requires additional neural resources for each new primitive, relation, or composite object, hence suffers from combinatorial limitations. Moreover, it is not obvious how entirely new combinations of well known ingredients could be processed properly, *i.e.* processed in such a way that the possibility of analogies and transfer is taken into account. However, the hierarchical feedforward computation enables to identify each neuron with a specific meaning and to use standard neural techniques.

Alternatively, composite objects may be represented in a *distributed manner* via the activation of a group of neurons for the single features (see READING OUT POPULATION CODES). This reduces the number of necessary neurons, but it is not obvious how invariance and not instantiated or new features are to be integrated. The necessary amount of resources depends on the respective task and cannot be uniformly limited. Moreover, the composition of two objects to a composite object is to be distinguished from the simple superposition of both's activations; otherwise, the respective decomposition would become ambiguous.

Naturally, both encodings may be combined in *localized distributed representations* of objects which correspond to the firing of neurons at a certain area of the pool. This encoding is suggested by the presence of topology preserving maps in biological systems where similar stimuli cause neural activities in similar neural regions (see OCULAR DOMINANCE AND ORIENTATION COLUMNS).

Actually, either localized or distributed encoding of composite objects in a vector space with fixed dimension is up to now the standard encoding for practical applications (see LOCALIZED VERSUS DISTRIBUTED REPRESENTATIONS). Hierarchical extraction of relevant features and finally of the respective class is a promising technique successfully applied in various approaches like the recognition of visual objects or text processing (Mel and Fiser 2000; Riesenhuber and Poggio 1999). Naturally, given the limits with respect to the number of recognized objects, static approaches suffer from a priori limitations and can only explain parts of biological information processing.

Partially Recurrent Systems

Discrete time partially recurrent neural networks are widely used in time series prediction, speech recognition, or processing of sequences of real vectors in general (Kremer 2001) (see LANGUAGE PROCESSING and ADAPTIVE CONTROL: NEURAL NETWORKS APPLICATIONS). They deal with sequences of real vectors as opposed to simple real vectors and hence are capable of processing very simple composite objects with a priori unlimited informational content. Their dynamics directly mirror the recursive structure of the data: A standard feedforward network encodes in its internal activations the context of the computation, *i.e.* the first part of a sequence, and recursively maps one entry of the sequence after another to new contexts depending on its internal state. Since the dynamic is fixed according to the data structure, standard gradient descent techniques can be used for supervised learning of mappings with sequences as inputs or outputs (see RECURRENT NETWORKS: SUPERVISED LEARNING).

The a priori unlimited information in the data structure changes the learning theoretical properties compared to standard feedforward networks: On the one hand, the power of recurrent networks can be demonstrated by relating them to classical symbolic computing mechanisms like Turing machines (Hammer 2002) (see NEURAL AUTOMATA AND COMPUTATIONAL COMPLEXITY). On the other hand, valid generalization can no longer be guaranteed for training set sizes which are both independent of the underlying input distribution and independent of the specific output of the training algorithm (Hammer 2002) (see PAC LEARNING AND NEURAL NETWORKS and VAPNIK-CHERVONENKIS DIMENSION OF NEURAL NETWORKS).

One can think of the recursive dynamics of simple recurrent networks as a recursive coding of sequences to distributed representations in a finite dimensional vector space. This idea can immediately be generalized to more complex composite objects provided they have a recursive structure: Trees consist of a label and a fixed number of subtrees. Hence a network which shall encode trees instead of simple sequences can recursively map the root's label and the already computed codes for the subtrees to a code for the entire tree. Conversely, it can decode a distributed representation of a tree via computing the root's label and codes for the subtrees which can be recursively processed further (Frasconi, Gori, and Sperduti 1997; Hammer 2002). This mechanism can be found in various *structured connectionistic systems*. Note that terms and formulas have a natural representation as tree structures: The single symbols are contained in the respective nodes and the subformulas or subterms, respectively, correspond to subtrees. Hence networks capable of encoding or decoding tree structures constitute a universal mechanism which enables the use of neural techniques in symbolic domains.

Concrete implementations of this basic idea differ in the way of how the respective networks are trained. With the dynamics being determined by the data structure, gradient descent techniques can be used for supervised learning of general mappings with trees as input or output. Recursive networks are trained directly for the specific learning problem (Frasconi, Gori, and Sperduti 1997; Hammer 2002). The recursive autoassiociative memory trains only encoding and decoding such that their composition yields the identity on the data, leading to a universal encoding (Frasconi, Gori, and Sperduti 1997; Sperduti 1994). Holographic reduced representation uses a fixed transition function which is not trained at all (Plate 1995). These approaches have been successfully applied in different areas of application such as chemistry, theorem proving, or language processing (Frasconi, Gori, and Sperduti 1997; Hammer 2002). Learning is quite similar to the training of simple recurrent networks what concerns practical algorithms as well as theoretical properties like approximation and generalization capability (Hammer 2002).

However, this approach is limited to recursive compositions with a well defined recursive structure; cyclic structures cannot be processed in this way. Moreover, access to single parts of recursive structures may be time consuming and subject to noise depending on the level of recurrence. There exist fundamental mathematical limitations for the possibility of coding infinite tree structures in a finite dimensional vector space: Encoding has to be nested or fractal such that the Euclidian metric is no longer appropriate for the resulting distributed representations (Hammer 2002). Hence various neural methods which are based on the Euclidian metric cannot be used for further processing. Reliable decoding is a difficult task with a lower bounded complexity (Hammer 2002) (see VAPNIK-CHERVONENKIS DIMENSION OF NEURAL NETWORKS). Hence neural networks with the above dynamics are not appropriate for efficient decoding a large amount of distributed data.

Fully Recurrent Systems

Fully recurrent neural systems are networks where the neuron's activations evolve in time in a continuous or discrete manner. Commonly, the dynamics can be described by nonlinear difference equations in discrete time or differential equations in continuous time. The main difference to partially recurrent networks consists in the fact that the processing dynamics and consequently the required computation time are not directly determined by the data structures. Time and complexity for computation and representation of information are a priori unlimited. The systems may use the temporal structure of the neuron's activations for storing important information in specific spatio-temporal activation patterns such as synfire chains, *i.e.* successive spiking of specified neurons in precise time intervals (see SYNFIRE CHAINS). As proposed in Bienenstock (1996), for example, synchronous oscillation of different neurons or assemblies may indicate that they represent objects which are bound together (see SYNCHRONIZATION, BINDING, AND EXPECTANCY and ADAPTIVE SPIKE CODING). This type of binding could easily be further processed with coincidence detectors. Alternatively, information may be stored in simple localized or distributed patterns of the neuron's activities as in the static and partially recurrent approaches. Various systems obey a gradient dynamics as an example and converge to a fixed point, which contains the important information (see TEMPORAL SEQUENCES: LEARNING AND GLOBAL ANALYSIS).

Concrete implementations differ considerably in their complexity and intention: Several approaches merely demonstrate that important effects such as *oscillation, synchronization, and coincidence detection* can be found in experiments with biological neural activities and they can be simulated in an artificial, though biologically plausible environment (see CHAOS IN NEURAL SYSTEMS, COLLECTIVE BEHAVIOUR OF COUPLED OSCILALATORS, and CORTICAL POPULATION DYNAMICS AND PSYCHOPHYSICS). In particular in the context of networks of spiking neurons, methods which allow a mathematical investigation of complex systems have been developed in the last years (see INTEGRATE-AND-FIRE NETWORKS). First steps into possible training mechanisms use the principles of self-organization and Hebbian learning (see POST-HEBBIAN LEARNING RULES). Binding via the temporal structure and synchronous oscillation is a biologically plausible mechanism which is supported by experimental results (Engel et. al. 1992). It is not yet obvious how complex recursive and hierarchical structures can be represented in this way since methods like iterated period doubling or a superposition of the oscillation, for example, are restricted by the computation accuracy of neurons. Moreover, efficient learning algorithms for practical applications are not yet available.

Other approaches *develop practical tools for concrete tasks* which involve binding mechanisms such as the problem of feature grouping or edge detection in images (see PERCEPTUAL GROUPING AND FIGURE GROUND SEPARATION). Grouping may be encoded in synchronously oscillating neurons or in the localized activation of specific neurons in a limiting state which minimizes an energy function such as in the competitive layer model (Wersing, Steil, and Ritter 2001). Most approaches are based on appropriate excitation of similar neurons and an inhibition of cells with dissimilar activation. Again, only limited possibilities of automatic learning of the connections are available so far.

Finally, recurrent systems for *complex analogical reasoning and symbol processing* have been proposed (see STRUCTURED CONNECTIONIST MODELS). Popular approaches are LISA, SHRUTI, and INFERNET, which have in common that binding is realized via synchronous oscillation of neurons or pools of neurons (Hummel and Holyoak 1997; Shastri 2002; Sougné 1999). Structures are represented through localized or distributed cell activations. Rules correspond to specific neural connections which allow human-like analogical reasoning and are mostly hand encoded. They are capable of simulating various effects and limitations of human-like reasoning. However, like most fully recurrent approaches, the systems suffer from the lack of universal and efficient training algorithms.

Discussion

Compositionality as a common principle of information processing should find its counterpart in artificial neural systems. Solutions may either enhance static feedforward systems and represent the objects by static activation patterns, they may rely on recursive and adaptive encoding mechanisms in partially recurrent networks, or they may use complex dynamics and the fine temporal structure of the neuron's activation for reliable representation. Naturally, it is possible to transfer more practical tools and theoretical guarantees from classical network techniques to these systems if they are similar to classical systems. It should be pointed out that, until today, it is still disputed whether artificial neural networks are at all capable of adequately handling compositional data, and if so, which type of networks is the most suitable one. Remarkable results have already been obtained with simple recurrent networks, whereas some researchers argue that more complicated dynamics or dynamics similar to classical symbolic processing mechanisms are necessary for successful modeling within the context of compositionality (see CONNECTIONIST AND SYMBOLIC REPRESENTATIONS and (Elman 1998) and references therein for a discussion on this subject).

Important directions of further research within this context include the following: real neural codes in biological systems are to be explored. In particular, it has to be examined whether effects like synchronous activation indeed contain information that is necessary for the representation of relations, for example. Oscillations could constitute a byproduct of information processing which merely enables efficient adaptation in biological systems.

Restrictions of compositionality are to be explored and formalized. In practice, only a sparse subset of all possible combinations of primitives and relations occurs. Mathematical analysis of structure processing systems implies usually a worst case analysis and might indicate, for example, that static approaches are not appropriate for this field. However, neural mechanisms which are not capable of representing arbitrary composite objects in principle might be well suited for restricted, though important domains.

References

* Bienenstock, E. 1996. Composition. In: Brain Theory - Biological Basis and Computational Theory of Vision, A. Aertsen and V. Braitenberg (eds.), Elsevier, pp. 269-300.

Elman, J. 1998. Generalization, simple recurrent networks, and the emergence of structure. In: Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, M.A. Gernsbacher and S.J. Derry (Eds.), Mahwah, NJ: Lawrence Erlbaum Associates.

Engel, A. K., P. König, A. K. Kreiter, T. B. Chilien, and W. Singer, 1992. Temporal coding in the visual cortex: New vista on integration in the nervous system. Trends in Neurosciences, 15:218-225.

* P. Frasconi, M. Gori, and A. Sperduti. A general framework for adaptive processing of data sequences. IEEE Transactions on Neural Networks, 9(5):768-786, 1997.

* van Gelder, T. 1990. Compositionality: a connectionist variation on a classical theme. Cognitive Science, 14:355-384, 1990.

* Hammer, B. 2002. Recurrent networks for structured data – a unifying approach and its properties. To appear in Cognitve Systems Research.

Hummel, J. E. and K. J. Holyoak, 1997. Distributed representation of structure: A theory of analogical access and mapping. Psychological Review, 104(3):427-466.

* Kremer, S. C. 2001. Spatio-temporal connectionist networks: A taxonomy and review. Neural Computation, 13(2):249-306.

Mel, B. and J. Fiser, 2000. Minimizing binding errors using learned conjunctive features. Neural Computation, 12:247-278.

Plate, T. 1995. Holographic reduced representations. IEEE Transactions on Neural Networks, 6(3):623-641.

Riesenhuber, M. and T Poggio, 1999. Are cortical models really bound by the 'binding problem'? Neuron, 24:87-93.

Shastri, L. 2002. Advances in SHRUTI - A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. To appear in Applied Intelligence.

Sougné, J. P. 1999. INFERNET: A neurocomputational model of binding and inference. PhD Thesis, Université de Liège.

Sperduti, A. 1994. Labeling RAAM. Connection Science, 6(4):429-459.

Wersing, H., J. J. Steil, and H. Ritter, 2001. A competitive layer model for feature binding and sensory segmentation of features. Neural Computation, 13(2):357-387.