Journal of Theoretical Biology 275 (2011) 29–41

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Symbol manipulation and rule learning in spiking neuronal networks

Chrisantha Fernando [a,b,c,*]

[a] Department of Informatics, University of Sussex, Falmer, Brighton, BN1 9RH, London, UK
[b] MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK
[c] Collegium Budapest, Institute for Advanced Study, Szentháromság u. 2, H-1014 Budapest, Hungary

ARTICLE INFO

Article history:
Received 19 August 2010
Received in revised form
9 January 2011
Accepted 10 January 2011
Available online 13 January 2011

Keywords:
Spiking neural network
Neuronal replicator hypothesis
Learning classifier sysems
Spike-time-dependent plasticity

ABSTRACT

It has been claimed that the productivity, systematicity and compositionality of human language and thought necessitate the existence of a physical symbol system (PSS) in the brain. Recent discoveries about temporal coding suggest a novel type of neuronal implementation of a physical symbol system. Furthermore, learning classifier systems provide a plausible algorithmic basis by which symbol re-write rules could be trained to undertake behaviors exhibiting systematicity and compositionality, using a kind of natural selection of re-write rules in the brain, We show how the core operation of a learning classifier system, namely, the replication with variation of symbol re-write rules, can be implemented using spike-time dependent plasticity based supervised learning. As a whole, the aim of this paper is to integrate an algorithmic and an implementation level description of a neuronal symbol system capable of sustaining systematic and compositional behaviors. Previously proposed neuronal implementations of symbolic representations are compared with this new proposal.

1. Introduction

In a highly influential and controversial paper, Fodor and Pylyshyn provided an argument for the existence of a physical symbol system (PSS) in the brain (Fodor and Pylyshyn, 1988). A simple way to understand what is meant by a PSS is to map the cognitive concepts onto concepts from chemistry. This analogy should not be taken too far, and is best used as an intuition pump to broaden the way we think about symbol processing. We do not wish to claim that a neuronal physical symbol system is isomorphic to a chemical one, and so at each stage we discuss the differences as well as the similarities.

Chemistry deals with molecules that are composed of atoms. Structural relations between atoms define a molecule. There is a combinatorial syntax, i.e. a set of chemical structural constraints such as valance, charge, etc., that determine how atoms can legally join together to make molecules.

Furthermore, the structure of a molecule has information about its *chemical function or reactivity*, and this is systematically related to the function of its parts, e.g. the structure of the benzene ring *means* that it will react in a certain way in a given environment, and the fact that it has a methyl group *means* that this reactivity will be changed in a systematic way in that environment. This is a kind of internal semantics. Internal semantics deals with how a symbol contains information about the reactivity within the symbol system itself. It is also called compositionality, and we will discuss it shortly, but before this, we should highlight that there is another kind of semantics, which we call external semantics. External semantics deals with how a symbol contains information about the outside world (i.e. its semantic interpretability) and it is easiest to consider this for biochemical systems; here molecules can clearly be seen to also have semantic content, i.e. parts of the molecule may confer information about the environment external to the symbol system. For example, the conformation of hemoglobin can confer information about the oxygen saturation, or the concentration of a cell signaling molecule can confer information about glucose concentration.

Fodor and Pylyshyn proposed that there is a PSS implemented in neuronal structures that has similar properties to a molecular symbol system. To understand why, it is useful to compare chemical experiments with human language and thought. The properties of atoms and molecules described above give chemistry a special set of macroscopic characteristics. For example, chemistry is productive. The capacity for chemical reactivity is unlimited. Indefinitely many molecules can be produced allowing indefinitely many reactions. This is possible with only a finite set of distinct atomic types. Therefore, an unbounded set of chemical structures must be non-atomic. In the same way, an indefinite number of propositions can be entertained, or sentences spoken. This is known as the productivity of thought and language,

* Correspondence address: Department of Informatics, University of Sussex, Falmer, Brighton, BN1 9RH, London, UK.
E-mail address: ctf20@sussex.ac.uk

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.01.009

therefore neuronal symbols must have the same capacity for being combined in unlimited ways.

Secondly, chemistry is systematic, that is the capacity for atoms to be combined in certain ways to produce some molecules is intrinsically connected to their ability to produce others. Consider how a chemist might learn chemistry. There are rules of thumb that help a chemist to guess how atoms will form a molecule, and how that molecule will react based on its structure. A chemist does not learn just a list of valid molecules or reactions. In the same way, there is systematicity in language, e.g. the ability to produce or understand a sentence is intrinsically connected with the ability to produce and understand other sentences. This is because there is systematicity in the way that the physical symbols responsible for language can form symbol structures and there is systematicity in the way that this structure determines the reactivity of these structures. Because of this structure, languages need not be learned by learning a phrasebook. Languages have syntax. No English speaker can say A loves B, but not be able to say B loves A.

Thirdly, we have already discussed how the same atom makes approximately the same contribution to each molecule in which it occurs. This means that there is systematicity in reactivity (semantics) as well as in structure (syntax). This is known as compositionality. In the same way, lexical items in sentences have approximately the same contribution to each expression in which they occur. This is a property of internal meanings (semantics), i.e. what a structure means in terms of function.

The phenomena of language and thought imply a neuronal physical symbol system, i.e. a system with all the properties described above, in the same way as the phenomena of chemistry imply the existence of atoms and molecules. However, there are *extra* properties required of the PSS in cognition compared to the PSS of chemistry. The most important is the fact that cognition includes the capacity to *learn* an appropriate PSS, not just to implement it. The fact that children can learn and manipulate explicit rules (Clark, 1991; Karmiloff-Smith, 1996) implies the existence of a neuronal physical symbol system capable of forming structured representations (analogous to molecules) and learning rules for operating on these representations (analogous to reactions of molecules) (Marcus, 2001). A theory of a neuronal physical symbol system must also explain the capacity to infer grammar during the process of language acquisition (Steels and Szathmáry, 2008). Finally, symbol grounding (i.e. semantic interpretability) is also needed.

We believe that strong evidence for a neuronal PSS comes from the field of grammar learning. The following is an example of a behavior that is proposed to require neuronal symbols. Gary Marcus has shown that 7 month old infants can distinguish between sound patterns of the form ABA vs. ABB, where A and B can consist of different sounds, e.g. "foo", "baa" etc. Crucially, these children can generalize this discrimination capacity to new sounds that they have never heard before, as long as they are of the form ABA or ABB. Marcus claims that performance in this task requires that the child must extract "abstract algebra-like rules that represent relationships between placeholders (variables), such as "the first item X is the same as the third item Y", or more generally that "item I is the same as item J"" (Marcus et al., 1999). Several attempts have been made to explain the performance of these children without a PSS (e.g. using connectionist models) (Seidenberg and Elman, 1999) but Marcus (2001, p. 70) has criticized these as smuggling in symbolic rules in one way or another by design. For Marcus it seems that the system *itself* must discover the general rule. In summary, the problem with a large set of connectionist learning devices is that a regularity learned in one component of the solution representation is not applied/generalized effectively to another part (Marcus, 2001). Marcus calls this the problem of *training independence* (Marcus, 2001). Marcus considers this as one of the fundamental requirements for a learning system to be described as symbolic or rule based.

It is important to realize that it would be nonsense to claim that the brain is nothing but a physical symbol system. Indeed, we believe that a PSS is needed to explain only some relatively advanced aspects of cognition, e.g. some aspects of language and abstract thought. There is a huge amount of non-symbolic functionality possessed by neuronal processes. Non-symbolic learning mechanisms are probably utilized in the search for symbolic rules during ontogeny. An excellent example of such non-symbolic (connectionist) functionality is how the visual system can learn shift-invariance from a few training examples. The test for the trained system is to be able to determine whether two novel objects presented at different times or places are the same or different, e.g. for face recognition (Wiskott and Malsburg, 1995). Note that the same/different distinction is important in being able to solve Marcus' ABA vs. ABB task. Konen and von der Malsburg (1993) have shown how shift-invariant pattern recognition can be achieved by rapid reversible synaptic plasticity (dynamic link matching). Exactly this process can be applied to automatically learning to distinguish same and different in a symbol system if the symbol is represented on a grid in the same way as a visual image. Thus, non-symbolic mechanisms can be involved in the discovery of symbol systems. The framework presented in this paper emphasizes this interaction. Also, we certainly do not claim that physical tokens are as simple as the 3×3 spatiotemporal structures we discuss. Instead, we propose that such tokens will be grounded in non-arbitrary ways to sensory and motor systems. Due to space constraints, the symbol grounding problem (semantic interpretability) is not addressed in this paper, although it is dealt with thoroughly elsewhere in a manner which does not remove the need for a physical symbol system (Harnad, 1990). Some authors have claimed that the symbol grounding problem has actually been solved in robotics (Steels, 2007). However, in these cases, a physical symbol system is still required.

To summerise, the following is a definition of a physical symbol system of the type proposed to be required to explain the kinds of rule learning exhibited in Marcus's task above, adapted from Harnad (1990). A physical symbol system contains a set of arbitrary **atoms (or physical tokens)** that are manipulated on the basis of "**explicit rules**" that are likewise physical tokens or strings (or more complex structures consisting) of such physical tokens. The explicit rules of chemistry for example allow the calculation of reactions from the structure of atoms and molecules. The rule-governed symbol-token manipulation is based purely on the shape of the symbol tokens (not their "meaning"), i.e., it is **purely syntactic**, and consists of "rulefully combining" and recombining symbol tokens. There are primitive atomic symbol tokens and **composite symbol-token strings (molecules)**. The entire system and all its parts – the atomic tokens, the composite tokens, the syntactic manipulations both actual and possible and the rules – are all "**semantically interpretable**:" The syntax can be systematically assigned a meaning, e.g. as standing for objects, as describing states of affairs (Harnad, 1990). Semantic interpretability in molecular systems occurs in evolved biochemical systems, e.g. cell signaling molecules, transcription factors, etc., all convey meaning to a gene regulatory system about the state of the environment. Analogously, a neuronal symbol system contains information about the environment.

We will demonstrate how: (1) arbitrary physical tokens (atoms), (2) arranged into molecules or symbol structures,