Biologically Inspired Cognitive Architectures II:
Papers from the AAAI Fall Symposium (FS-09-01)

To Cognize Is to Categorize Revisited: Category Theory Is Where Mathematics Meets Biology

Jaime Gómez and Ricardo Sanz
Universidad Politécnica de Madrid,
José Gutiérrrez Abascal, 2. Madrid 28006 Spain
jd.gomez@upm.es

Abstract

This paper claims for a shift towards ”the formal sciences” in the cognitive sciences. In order to explain the phenomenon of cognition, including aspects such as learning and intelligence, it is necessary to explore the concepts and methodologies offered by the formal sciences. In particular, category theory is proposed as the most fitting tool for the building of an unified theory of cognition.

This paper proposes a radically new view based in category theory. A cognitive model is *informally* defined as a mapping between two different structures, while a structure is the set of components of a system and their relationships.

Put *formally* in categorical terms, a model is a functor between categories that reflects the structural invariance between them.

In the paper, the theory of categories is presented as the best possible framework to deal with complex system modeling -ie: biologically inspired systems that transcend and offer a much more powerful tool kit to deal with the phenomenon of cognition that other purely verbal tools like the psychological categories that Rosch or Harnad refer.

Introduction

“The aim of cognitive sciences always was –and still is today– the mechanisation of the mind and not the humanisation of the machine”. (Dupuy 2009)

The cyberneticians put the focus on modeling the mind, by doing this, they radically changed the status of the mental. The mind could no longer be understood with metaphors or allegories -e.g: Freud, Proust; but rather, it could be incorporated as a new domain into the scientific activity, which essentially consists in constructing models of the phenomena observed.

The acceptance of the mind as another domain within the scientific repertoire, relies upon the hypothesis that what makes any physical cognitive agent understand through means of models, can itself be modeled.

But there exists a danger here, as models become more simplified than the phenomena, it is nature which has to be explained by the model and not the other way around.

However, despite the daunting complexity of the brain, its dynamic organisation which is reflected by stable patterns, is preserved and at least, in principle, it is amenable to a formal representation.

Overcoming the structure-function approach

Cognition extracts from the environment features on the basis of which a system reacts. Based upon this definition, the terms environment and feature are key and must be explored in more detail. A theory is needed for the measurement of the neural activity, which is able to identify the ecologically relevant stimuli and to interpret the neural activity, as independent as possible from the experimenter’s suppositions and beliefs.

As Herbert Simon said, there is no principled way to see the brain and environment as separate systems, the brain and the environment form a nearly-decomposable system.

According to Duncan Luce (Luce 1995), as honest scientists, as far as we disentangle the structure of the physical objects we are studying, we must to begin to describe it in formal terms.

The formalisation of the structure of a system is not just an issue of honesty, but a requirement to be fulfilled in order to attain the *character* of a system -i.e: a pattern, conducting a behavior, defined in mathematical terms.

A mathematical toolkit

Admittedly, mathematical formalisms do not tell anything per se about the way physical systems behave. Rather, a mathematical theory of natural phenomena is, in the best case, a bunch of definitions, axioms and theorems that for some reason, happen to provide an appropriate framework able to describe, predict and explain better than any other conceptual framework, the way in which nature does behave.

We would like to define in the theory of categories, some key terms that are relevant in order to build such explanatory framework for cognition. (Pierce 1991)

Definition: category consists of the following :

i. a collection of objects A, B, C . . .

ii. a collection of arrows (or morphisms) f,g,h . . .

iii. operations assigning to each arrow f between two objects A, B, a domain and a codomain such that if $f : A \rightarrow B$ then domf=A and codf=B

iv. a composition rule, assigning to each pair of arrows, f, g $f : A \rightarrow B, g : B \rightarrow C$ a composite arrow $(f \circ g) : A \rightarrow C$. Note that in order to exist, composition of two arrows f and g needs that cod(f) = dom(g).
The composition is subject to the following axioms:

iv.a associativity: The operator $\circ$ is associative, given $f : A \rightarrow B$, $g : B \rightarrow C$, and $h : C \rightarrow D$, $(h \circ g) \circ f = h \circ (g \circ f)$ and using the composition rule it results: $((h \circ g) \circ f)(a) = h(g(f(a))) = (h \circ (g \circ f))(a)$

iv.b identity: every object A has an identity function $1_A : A \rightarrow A$ satisfying the identity law for every arrow $f : A \rightarrow B$, $1_B \circ f = f$ and $f \circ 1_A = f$

A category is *anything* that satisfies this definition, the objects can be sets, groups, monoids, vector spaces... or neurons in the hippocampus.

Definition: isomorphism is a morphism $f : A \rightarrow B$ such that it exists a morphism $g : B \rightarrow A$ and $g \circ f = 1_A$ and $f \circ g = 1_B$.

There is a number of mathematical domains that can be formulated as categories, for example, the category of groups, the category of rings, the category of vector space or the category of topological spaces. The non-algebraic categories like category of neurons, or the category of brain areas, once mathematically formalised, become operational and can be used to describe the structure of the physical phenomena.

There is also a category of categories, in which objects are categories and morphisms, the structure preserving maps between categories, called functors.

A functors F is a structure preserving map between categories.

Definition: Let C and D be categories, the functor $F : C \rightarrow D$ maps each object $C' \in C$ to $F(C') \in D$ and each arrow in C, $f : A \rightarrow B$, to the arrow in D $F(f) : F(A) \rightarrow F(B)$. The functor is subject to the following two conditions:

i. composition is preserved. For $f : A \rightarrow B$ and $g : B \rightarrow C$ $F(g \circ f) = F(g) \circ F(f)$

ii. identities are preserved, so for any object A in the category C, $F(1_A) = 1_{F_A}$

Case study: The reverse inference trouble in brain studies

Direct inference dictates which brain areas are active given a cognitive process. Inverse inference determines from the activation of a brain region, which particular cognitive process is engaged.

In (J.Gómez, R.Sanz, and C.Hernández 2008) it is stated that direct inference can be defective in terms of precision, while reverse inference can also be a logical fallacy. The problem with reverse inference is that the activation of one area can be incidental to a cognitive process. This is a methodological defect almost ubiquitous in the community of brain imaging studies.

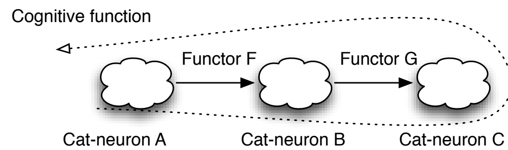

Figure 1: Cat-neurons are categories whose objects are neurons and their arrows synapses. The functors are structure-preserving maps between the categories. In this description, the reverse inference problem consists of determining the category of functors and natural transformations

Category theory brings a mathematical toolkit able to express with rigor, concepts such as the multi-realizability of one function by different structures, or the reducibility of one complex structure in some simpler pattern.

Structure is not just a 3-dimensional physical structure, this is a layman's view of structure. Using the resource of mathematics, we can go deeper into this concept. Indeed, symmetry and continuity are structures that can be realised in both physical and non-physical systems.

It goes without saying that one system has many possible structures, the structure however, is contingent upon the behavior of the system that is considered.

As any partition of a system could be its structure, it is necessary to make a distinction between relevant and irrelevant structures(for some observed behavior). The relevant structures are those that are preserved under transformation. Consequently, the structure must reflect the invariance through transformations, which is embodied in the functor. Figure 1

Conclusions

The time has come to set the agenda for a "hard cognitive science". To that end, the authors propose to translate into mathematical terms key concepts, that until now, have been used loosely. Category theory is presented as a suitable framework for complex system modeling, and as a sophisticated toolkit for mental theories.

Acknowledgements

We acknowledge the support of the European Commission through grant *ICEA: Integrating Cognition, Emotion and Autonomy*

References

Dupuy, J.-P. 2009. *The Mechanization of Mind: On the Origins of Cognitive Science*. MIT Press.

J.Gómez; R.Sanz; and C.Hernández. 2008. Cognitive ontologies: Mapping structure and function of the brain from a systemic view. *AAAI Fall symposium, BICA 2008*.

Luce, D. 1995. Four tensions concerning mathematical modeling in psychology. *Annual Reviews of Psychology, 46, 1-26*.

Pierce, B. 1991. *Basic Category Theory for Computer Scientists*. MIT Press.