Karlsruhe Institute of Technology

Coherence and Associativity as a Basis of Cognitive Architecture

by Caslav Bozic

KIT SCIENTIFIC WORKING PAPERS 30

Impressum

Karlsruher Institut für Technologie (KIT)
www.kit.edu

2015

ISSN: 2194-1629

Coherence and Associativity as a Basis of Cognitive Architecture

Caslav Bozic
bozic@kit.edu
Karlsruhe Institute of Technology (KIT), Germany
caslav@drbozic.de
Dr. Bozic Infoservices, Germany

May 13, 2015

Abstract

In this working paper we explore a possibility of defining implementable cognitive architecture based on coherence and associativity. Drawing loosely on insights from cerebral cloning (Calvin), confabulation (Hecht-Nielsen), and coherence (Thagard), we propose a basis for a novel cognitive architecture.

1 Introduction

The paper is organized as follows: in Chapter 2 we describe existing work in this area which inspired the ideas presented here, or are in some way related to it. Chapter 3 introduces some of the most important terms and concepts for understanding the model, while Chapter 4 describes the model and the process itself. Chapter 5 concludes this paper.

2 Previous Work

Unified theories of cognition were defined in (Newell, 1990) as

> a single set of mechanisms for all of cognitive behavior.

Today, more than 25 years later, there is still no single set of mechanisms to describe cognition, at least not in the form that most of the scholars would agree upon or that could be used in a high-school-level textbook. We aim (as many authors before) at offering a solution to this puzzle.

2.1 Confabulation

In (Hecht-Nielsen, 2007), author offers a theory and describes possible underlying mechanisms for cognitive functions. He draws a parallel between muscle contractions and a *confabulation* operation within a *cognitive module*, a functional and anatomical segment of a brain. Fundamental component of the theory are *symbols* and *knowledge links*, but the author argues that Bayesian explanation of cognition is flawed, and what is actually happening in the brain is the maximization of *cogency* $p(\alpha\beta\gamma\delta|\lambda)$ (otherwise called likelihood). By postulating an expression for the knowledge link weights as $log_c(p(\alpha|\lambda)/p_0) + B$ (α represents *source symbol*, λ *target symbol*, and c, p_0, B are positive constants) with some additional evolutionally achieved mathematical conditions, the author draws equivalence between maximizing cogency and confabulation. In the same book there is a description of a computer system created to mimic the proposed mechanism, and extensive examples of its results in completing sentences are given. Results of experiments on different type of data are not disclosed.

2.2 Thought Evolution

Another book (Calvin, 1997), as well as (Calvin, 1996), feature the view that all cognitive processes are based on *thought evolution* - mechanism of competition between different thoughts. Spatiotemporal patterns of neuron firing are replicated within neighboring hexagonal segments of cortex, and by synchronization and amplification, or elimination of these patterns, the conscious thoughts emerge.

2.3 Coherence

The term *coherence* used in (Thagard, 2000) has somewhat different meaning compared to the usage of the term in this paper. The author is offering a computational approach to coherence and providing a potentially powerful theory of cognitive mechanisms. Coherence optimization, in this context, represents constraint satisfaction maximization, which can be achieved by a number of algorithms. To the selection of algorithms, the author says:

> The most psychologically appealing models of coherence optimization are provided by connectionist algorithms. These use neuronlike units to represent elements, and excitatory and inhibitory links to represent positive and negative constraints. Settling a connectionist network by spreading activation results in the activation (acceptance) of some units and the deactivation (rejection) of others.

The degree of constraint satisfaction represents a measure of coherence, e.g. when using goodness-of-fit measure, defined as $\sum_j \omega_{ij} a_i(t) a_j(t)$, the assigned activation values a represent acceptance or rejection of an element. This approach was used in implementing a computational model of explanatory coherence called ECHO, which has been applied to many cases from the history of science and legal reasoning.

2.4 Summary

The mechanisms described in this paper draw inspiration from aforementioned models. The main goal of the paper is to offer an implementable mechanism and describe an underlying cognition theory.

3 Terms

The basic terms in this text are:

- **Concept**: basic building block that facilitates processing. The only payload of a concept are other concepts that are related to that particular concept. Physiologically, it might be a temporal-spatial pattern of neuron firing.

- **Coherence sensation**: is closely related to the *aha! moment* and the sensation of truth/correctness. It is simply a pleasant physiological effect triggered by some concept being satisfied with an additional timing condition/constraint. Concept is satisfied when the same concept/pattern appears in a different location of the cortex.

- **Coherence potential**: is function of the triggered coherence sensation intensity, depending on the concept life time before satisfying. It is represented in Figure 1. It facilitates continuous shift in explanation complexity needed to trigger coherence sensation - when a relation between 2 concepts becomes well established in the mind, it appears with a shorter delay, thus allowing only low coherence sensation intensity. The solution becomes *trivial*. After a certain period of time the concept loses its coherence potential, making a quest for satisfying it *boring*.

- **Moment concept**: For each instance of relatively strong coherence sensation, a new concept is created (so called moment concept), and related to all concepts active in that moment (in the same area of cortex)

4 Process

When the concept becomes **active**, the goal is to **satisfy** it (by reaching it via another path in concept graph). Until that happens, the concepts related to the active concept are in turn activated (according to a priority). When the concept is satisfied, and if the coherence potential is above certain threshold, a new moment concept will be created, the concepts will be felt as true, confirmed (they *are*). If the coherence potential is under the threshold, this process **will not be recorded at all** and thus inaccessible to the mind.

In the most immediate instance of cognition, we are used to generating possible outcomes (activating related concepts) - jumping to conclusion - and

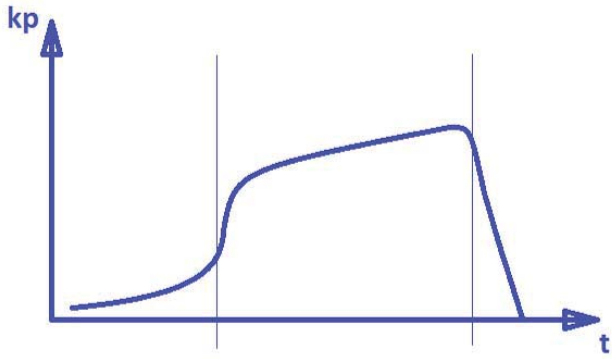

Figure 1: Coherence potential in time until concept satisfying

when one of them has been satisfied, and we feel the coherence sensation, we then rationalize backwards and create an appropriate *chain of logic*.

We are striving to feel the coherence sensation, which is a basic reward in learning and understanding. This is why e.g. a kid gets excited whenever it recognizes letter A in a text, and it explains why this excitement disappears as soon as it can be recognized without much effort - thus making space for more complex system explanations.

5 Conclusion

In this paper we offered to lay down the basic principles of an implementable cognitive architecture based on coherence and associativity. Further work is needed to arrive at fully implementable model, but we are confident it can be achieved by refining the basic principles presented here.

References

A. Newell, *Unified Theories of Cognition*. Harward University Press, 1990.

R. Hecht-Nielsen, *Confabulation Theory (The Mechanism of Thought)*. Springer, 2007.

W. H. Calvin, *How Brains Think (Evolving Intelligence, Then and Now)*. Weidenfeld & Nicolson, London, 1997.

W. H. Calvin, *The Cerebral Code (Thinking a Thought in the Mosaics of the Mind)*. The MIT Press, 1996.

P. Thagard, *Coherence in Thought and Action*. The MIT Press, 2000.

KIT Scientific Working Papers
ISSN 2194-1629