Biologically Inspired Cognitive Architectures 19 (2017) 1–9

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

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

Research article

Conceptual Spaces for Cognitive Architectures: A *lingua franca* for different levels of representation

Antonio Lieto [b,a,e,*], Antonio Chella [c,a], Marcello Frixione [d]

[a] ICAR-CNR, Palermo, Italy
[b] University of Turin, Dip. di Informatica, Torino, Italy
[c] University of Palermo, DIID, Palermo, Italy
[d] University of Genoa, DAFIST, Genova, Italy
[e] National Research Nuclear University, MEPhI, Moscow, Russia

ARTICLE INFO

Article history:
Received 26 June 2016
Revised 9 October 2016
Accepted 9 October 2016

Keywords:
Knowledge representation
Cognitive Architectures
Cognitive modeling
Conceptual Spaces

ABSTRACT

During the last decades, many Cognitive Architectures (CAs) have been realized adopting different assumptions about the organization and the representation of their knowledge level. Some of them (e.g. SOAR (Laird, 2012)) adopt a classical symbolic approach, some (e.g. LEABRA O'Reilly and Munakata (2000)) are based on a purely connectionist model, while others (e.g. CLARION (Sun, 2006)) adopt a hybrid approach combining connectionist and symbolic representational levels. Additionally, some attempts (e.g. biSOAR) trying to extend the representational capacities of CAs by integrating diagrammatical representations and reasoning are also available (Kurup & Chandrasekaran, 2007). In this paper we propose a reflection on the role that Conceptual Spaces, a framework developed by Gärdenfors (2000) more than fifteen years ago, can play in the current development of the Knowledge Level in Cognitive Systems and Architectures. In particular, we claim that Conceptual Spaces offer a *lingua franca* that allows to unify and generalize many aspects of the symbolic, sub-symbolic and diagrammatic approaches (by overcoming some of their typical problems) and to integrate them on a common ground. In doing so we extend and detail some of the arguments explored by Gärdenfors (1997) for defending the need of a conceptual, intermediate, representation level between the symbolic and the sub-symbolic one. In particular we focus on the advantages offered by Conceptual Spaces (with respect to symbolic and sub-symbolic approaches) in dealing with the problem of compositionality of representations based on typicality traits. Additionally, we argue that Conceptual Spaces could offer a unifying framework for interpreting many kinds of diagrammatic and analogical representations. As a consequence, their adoption could also favor the integration of diagrammatical representation and reasoning in CAs.

Introduction

Within the field of cognitive modeling, it is nowadays widely assumed that different kinds of representation are needed in order of accounting for both biological and artificial cognitive systems. Examples are the broad class of neural network representations (including deep neural networks); the vast family of symbolic formalisms (including logic and Bayesian or probabilistic ones); analogical representations such as mental images, diagrammatic representations, mental models, and various kinds of hybrid systems combining in different ways the approaches mentioned above.

All these methods are successful in explaining and modeling certain classes of cognitive phenomena, but no one can account for all aspects of cognition. This problem also holds if we consider some recent successful artificial systems. For example, the *Watson* system is based on a probabilistic system able to reason on enormous amounts of data, but it mostly fails to account for trivial common-sense reasoning (see Davis & Marcus (2015), p. 94). Similarly, the *AlphaGo* system (Silver et al., 2016), based on massive training of deep neural networks, is impressively successful in the well-defined domain of the *Go* game. However, it is not able to transfer its approach in general or cross-domain settings. In general, this is a classical obstacle of neural networks: in order to solve a particular problem they need to be trained by a suitable and vast training set. Then, however, how to employ the learned strategies to solve similar problems is still an open issue.[1]

* Corresponding author at: University of Turin (Dept. of Computer Science) and ICAR-CNR, Palermo, Italy.
E-mail address: lieto@di.unito.it (A. Lieto).

[1] This issue is also explicitly reported by Hassabis in an interview published on Nature http://goo.gl/9fUy4Z.

http://dx.doi.org/10.1016/j.bica.2016.10.005
2212-683X/

Based on this evidence, our claim is that the Knowledge Level of cognitive artificial systems and architectures can take advantage of a variety of different representations. In this perspective, the problem arises of their integration in a theoretically and cognitively motivated way. While, in fact, existing hybrid systems and architectures (see e.g. Sun (2006)) are able to combine different kinds of representations (see for example the class of neuro-symbolic systems in D'Avila Garcez, Lamb, & Gabbay (2008)), nonetheless this kind of integration is usually *ad hoc* based (Chella, Frixione, & Gaglio, 1998) or, as we will show in the following sections, is only partially satisfying. Our hypothesis is that Conceptual Spaces can offer a *lingua franca* that allows to unify and generalize many aspects of the representational approaches mentioned above and to integrate them on common ground.

The paper is organized as follows: in the Section "Heterogeneity of representations in cognitive science: The case of concepts" we report how, in Cognitive Science research, the problem of conceptual representations intended as a heterogeneous phenomenon has gained attention and experimental support in the last decades. In the Section "Representational formalisms and approaches in AI", we consider this pluralistic representational stance in the area of Artificial Intelligence by focusing on some of the most widely known representational approaches adopted in literature. The Sec tion "Conceptual Spaces as a *lingua franca*" provides a synthetic description of Conceptual Spaces, the representational framework that we propose for the connection of the different representational levels used in different CAs. In the Section "On the advantages of Conceptual Spaces" we outline the advantages offered by the Conceptual Spaces representation used as a grounding layer for the classical AI approaches reviewed in the Section "Represent ational formalisms and approaches in AI". In doing so we extend and detail some of the arguments explored by Gärdenfors (1997) for defending the need of a conceptual, intermediate, representation level between the symbolic and the sub-symbolic one. Conclusions end the paper.

Heterogeneity of representations in cognitive science: the case of concepts

In this section, we present some empirical evidence from Cognitive Science that favor the hypothesis of the heterogeneity of representations in cognitive systems and architectures. In particular, we take into account two classes of evidence concerning conceptual representations: the description of non-classical concepts (Se ction "Representing non-classical concepts") and the application of the dual process distinction to conceptual knowledge (Section " Dual-process oriented conceptual representations").

Representing non-classical concepts

In Cognitive Science, different theories about how humans represent, organize and reason on their conceptual knowledge have been proposed. In the traditional view, known as the classical or Aristotelian theory, concepts are defined as sets of necessary and sufficient conditions. Such theory was dominant in philosophy and psychology from the antiquity until the mid-70s of the last century, when the empirical results of Rosch (1975) demonstrated the inadequacy of such a theory for ordinary common sense concepts. These results showed that familiar concepts often exhibit *typicality* effects. The results obtained by Rosch have had a crucial importance for the development of different theories of concepts trying to explain various representational and reasoning aspects concerning typicality. Usually, such theories are grouped into three broad classes: (i) the prototype theories, developed starting from the work of Rosch; (ii) exemplars theories; and (iii) theory-theories (see e.g. Murphy (2002) and Machery (2009) for a detailed review of such approaches). All of them are assumed to account for some aspects of the typicality effects in conceptualization (such as that one of common-sense categorization).

According to the prototype view, knowledge about categories is stored using prototypes, i.e., representations of the best instance of a category. For example, the concept CAT coincides with a representation of a typical cat. In the simpler versions of this method, prototypes are represented as (possibly weighted) lists of features.

According to the exemplar view, a category is represented as set of specific exemplars explicitly stored within memory: the mental representation of the concept CAT is thus the set of the representations of (some of) the cats encountered during lifetime.

Theory-theories approaches adopt some form of holistic point of view about concepts. According to versions of theory-theories, concepts are analogous to theoretical terms in a scientific theory. For example, the concept CAT is individuated by the role it plays in our mental theory of zoology. In other versions of the approach, concepts themselves are identified with micro-theories of some sort. For example, the concept CAT is a mentally represented micro theory about cats.

Despite such approaches have been historically considered as competitors, since they propose different models, and they have different predictions about how the humans organize and reason on conceptual information, various works (starting from Malt (1989)) showed that they are eventually not mutually exclusive. Rather, they seem to succeed in explaining different classes of cognitive phenomena. In particular, empirical data - i.e., behavioral measures as categorization probability and reaction times - suggest that subjects use different representations to categorize. Some people employ exemplars, a few rely on prototypes, and others appeal to both exemplars and prototypes. Some representations seem to be more suitable for certain tasks, or for certain categories. Also, this distinction seems to have also neural plausibility witnessed by many empirical results (the first in this line is due to Squire & Knowlton (1995)).

Such experimental results led to the development of the so-called *heterogeneous hypothesis* about the nature of conceptual representations, according to which concepts do not constitute a unitary phenomenon. In particular, different types of conceptual representations are assumed to exist. All such representations represent different bodies of knowledge associated with the same category. Each body of conceptual knowledge is thus manipulated by various processes involved in multiple tasks (e.g. recognition, learning, categorization).

Dual-process oriented conceptual representations

A further divide between different kinds of conceptual representations refers to the dual process hypothesis about reasoning and rationality. According to *dual process* theories (Evans & Frankish, 2009; Kahneman, 2011; Stanovich & West, 2000) two different types of cognitive processes and systems exist, which have been called respectively *System(s) 1* and *System(s) 2*.

System 1 processes are automatic. They are phylogenetically older and shared by humans and other animal species. They are innate and control instinctive behaviors, so they do not depend on training or particular individual abilities and, in general, they are cognitively undemanding. They are associative and operate in a parallel and fast way. Moreover, *System 1* processes are not consciously accessible to the subject.

System 2 processes are phylogenetically recent and are peculiar to the human species. They are conscious and cognitively penetrable (i.e. accessible to consciousness) and based on explicit rule following. As a consequence, if compared to *System 1*, *System 2*