Grounding Ontologies in the External World

Antonio CHELLA
University of Palermo and ICAR-CNR, Palermo
antonio.chella@unipa.it

Abstract. The paper discusses a case study of grounding an ontology in the external world by a cognitive architecture for robot vision developed at the RoboticsLab of the University of Palermo. The architecture aims at representing symbolic knowledge extracted from visual data related to static and dynamic scenarios. The central assumption is the principled integration of a robot vision system with a symbolic system underlying the knowledge representation of the scene. Such an integration is based on a conceptual level of representation intermediate between the sub-symbolic processing of visual data and the declarative style employed in the ontological representation.

Keywords. Conceptual Spaces, Symbol Grounding

1. Introduction

The symbol grounding problem, as stated by Stevan Harnad [1], roughly concerns how to interpret a formal symbol system in terms of the entities in the external world. For example, an instance of the problem is the interpretation of the symbol “Hammer#1” in a formal symbol system by the corresponding hammer in the real world. The problem is crucial especially for autonomous agents because an autonomous agent has to find the meaning of its symbols in the inner structures of the agent itself. Harnad discusses the capabilities of neural networks as candidate mechanisms able to solve the problem.

The paper claims that an intermediate representation of a geometric kind is a better candidate for the symbol grounding problem. It is well known that neural networks present several problems [2]. They are opaque, i.e., it is difficult to understand the behavior of a neural network simply by analyzing its weights and the activation levels of its units. Moreover, a neural network needs a massive training set of labeled examples. After a neural network is trained, it is quite difficult to add new examples without restarting the training phase from scratch. The compositionality of concepts in neural networks is another well-known problem of a complicated solution.

The theory of conceptual spaces provides instead a robust geometric framework for the grounding of ontologies of symbols in a cognitive agent that overcomes many of the limitations of neural network representations.

2. Conceptual Spaces

A conceptual space (CS) is a metric space in which entities are characterized by some quality dimensions [3]. Examples of such aspects could be color, pitch, volume, spatial coordinates, and so on. Some dimensions are closely related to the sensorial inputs of the system; others may be characterized in more abstract terms. The dimensions of a conceptual space represent qualities of the external environment independently from any linguistic formalism or description. In this sense, a conceptual space comes before any symbolic characterization of cognitive entities.

An important aspect of the theory of conceptual spaces is the definition of a metric function in CS. In brief, the distance between two points of a CS computed according to such a metric function corresponds to a measure of the similarity between the entities corresponding to the points.

Another pillar of CS theory is the role of convex sets of points in the conceptualization. According to psychological literature (see, e.g., [4]), the so-called natural categories represent the most informative level of categorization in taxonomies of real-world entities. They are the most differentiated from one another and constitute the preferred level for reference. Also, they are the first to be learned by children, and categorization at their level is usually faster. The theory of conceptual spaces assumes the so-called Criterion P, according to which natural categories correspond to convex sets in some suitable CS. As a consequence, *betweenness* is significant for natural categories, in that for every pair of points belonging to a convex set (and therefore sharing some features), all the points between them also belong to the set itself, and share in their turn the same features.

Conceptual spaces, as discussed in detail in [2], are more transparent than neural networks; they can be built even by a small set of examples; they are more suitable for incremental learning; the problem of compositionality may be taken into account more quickly and naturally. See [5] for up to date discussion on the relationships between conceptual spaces and structures in the brain.

3. A Cognitive Architecture

Based on these ideas, a cognitive architecture for robot vision has been developed at the RoboticsLab of the University of Palermo.

3.1. The three areas

The design is subdivided into three main areas: the *subconceptual*, the *conceptual* and the *linguistic* areas. The *subconceptual* area is related to the processing of data coming from the sensors. Here, information is not yet organized concerning conceptual structures and categories. Instead, in the *linguistic* area, representation and processing are based on a logic-oriented formalism based on description logic [6]. In this area, ontologies may be suitably represented.

The *conceptual* area is based on the theory of conceptual spaces previously outlined. It is an intermediate level of representation between the sub-conceptual and the linguistic areas. Here, data is organized in conceptual structures that are independent of symbolic description. The symbolic ontology of the linguistic area is then interpreted on aggregations of these structures. The conceptual space acts as a workspace in which low-level and high-level processes access and exchange information from bottom to top and from top to bottom.

The three areas of the architecture are parallel computational components working together on different commitments. There is no privileged direction in the flow of information among them: some computations are strictly bottom-up, with data flowing from the subconceptual up to the linguistic through the conceptual area; other calculations combine top-down with bottom-up processing.

3.2. The case of static scenes

In the case of grounding ontologies related to static scenes [7], we take into account a suitable conceptual space where each point corresponds to a geometric entity. Then, “natural” concepts such as boxes, cylinders, spheres, correspond to convex sets of points in the considered conceptual space. A symbol like “Box#1” thus corresponds to an item in the CS belonging to the convex set of boxes.

Composite objects cannot be described by single points in this CS. To represent these objects, we naturally assume that they correspond to sets of points in CS. For example, a chair can be easily described as the set of its constituents, i.e., its legs, its seat and so on. Analogously, a hammer may be considered as composed of two geometric entities: its handle and its head. So, a generic composite object is described as the set of points corresponding to its components.

The concept of hammer thus is described in CS as a set of pairs, each of them is made up of the two elements of a real hammer, i.e., its handle and its head. Let us suppose for simplicity that the hammer handle is typically a cylinder, while the hammerhead is usually a box. Then, the handle of the hammer will be grounded in the CS on the subset of the set of points corresponding to the concept of the cylinder, while the head of the hammer will be grounded on the suitable subset of points corresponding to the concept of the box.

Thus, the symbol "Hammer#1,” corresponding to a specific instance of a hammer, will correspond to a specific pair of points in the conceptual space: one point of the pair will belong to the proper subset of cylinders while the other point will belong to the subset of boxes. In turn, these points are linked to the corresponding entities in the external world thanks to the subconceptual area that processes the data coming from the sensors of the system.

3.3. The focus of attention

To identify in the CS the set of components of a composite object as the hammer that is described at the symbolic level, we define a *focus of attention* mechanism acting as a light spot that sequentially scans the conceptual space.

In the beginning, the focus of attention explores a zone in the conceptual space where a point is expected that matches one of the components of the composite object, for example, the point corresponding to the hammer handle. If this expectation is satisfied, then the focus of attention searches for a second component of the composite object (e.g., a second point corresponding to the hammerhead, with suitable shape and appropriate spatial arrangement). This process is iterated until all such expectations are satisfied, and therefore there is enough evidence to assert that a composite object as the hammer is present in the scene.

The focus of attention is controlled by two different modalities, namely the *linguistic* modality and the *associative* modality. According to the linguistic modality, the focus of attention is driven by the symbolic knowledge explicitly stored in the ontology in the linguistic area.

For example, let us suppose that the system stored in its ontology the description of the hammer as composed by a head and by a handle. When the system recognizes a point in CS as a possible part of a hammer (e.g., as its handle), it generates the hypothesis that a hammer is present in the scene, and therefore it searches the CS for the lacking parts (in this case, the hammer's head).

When different types of objects have similar parts (e.g., similar handles), various competing hypotheses are generated, the most plausible of which wins on the others, and is accepted by the system.

According to the associative modality, the focus of attention is driven by an associative mechanism based on learned expectations. Let us suppose that the system has experienced several scenes where a hammer is present along with a nail. As a consequence, the system determines to associate hammers and boxes; when a hammer is present in the scene, it expects to find a nail in the surroundings.

A natural way to implement the focus of attention as it has been described before is to employ an associative memory. A mechanism based on a suitable associative neural network that implements both the linguistic and the associative modalities of the focus of attention is discussed in [7].

The ideas previously summarized for the analysis of static scene are generalized to ground ontologies related to dynamic scenes analysis [8], robot actions [9], robot self-recognition [10], robot self-consciousness [11], and recently to model ontologies related with music perception [12].

4. Conclusions

Conceptual spaces offer a robust theoretical framework for the development of a conceptual semantics for symbolic ontologies that can account for the grounding of symbols in the data coming from robot vision. In this sense, conceptual spaces could give a relevant contribution to a better integration of robot vision and ontology-based AI techniques in the design of autonomous agents.

Acknowledgement

The author would like to thank Salvatore Gaglio, Marcello Frixione, Antonio Lieto, Peter Gärdenfors, Christian Balkenius, Roberto Pirrone, Haris Dindo, Irene Macaluso, Ignazio Infantino, Valeria Seidita, Geraint Wiggins, Nicola Guarino, Stefano Borgo, among others, for discussions about the topics of the paper.

References

[1] S. Harnad, The symbol grounding problem, *Physica D: Nonlinear Phenomena*, **42** (1990), 335-346.

[2] A. Lieto, A. Chella, M. Frixione, Conceptual spaces for cognitive architectures: a lingua franca for different levels of representation, *Biologically Inspired Cognitive Architectures*, **19** (2017), 1-9.

[3] P. Gärdenfors, *Conceptual spaces: The geometry of thought.* MIT Press, Cambridge, MA, 2000.

[4] E. Rosch, Cognitive representations of semantic categories, *Journal of Experimental Psychology: General*, **104** (1975), 192-233.

[5] C. Balkenius, P. Gärdenfors, Spaces in the Brain: From Neurons to Meanings, *Frontiers in Psychology*, **7** (2016), 1820.

[6] R.J. Brachman, D.L. McGuinness, P.F. Patel-Schneider, L. A. Resnick, A. Borgida, Living with CLASSIC: when and how to use a KL-ONE-like language, in: J.F. Sowa (ed.), *Principles of semantic networks: explorations in the representation of knowledge*, 401-456, Morgan Kaufmann, San Mateo, CA, 1991.

[7] A. Chella, M. Frixione, S. Gaglio, A cognitive architecture for artificial vision, *Artificial Intelligence*, **89** (1997), 73-111.

[8] A. Chella, M. Frixione, S. Gaglio, Understanding dynamic scenes. *Artificial Intelligence,* **123** (2000), 89-132.

[9] A. Chella, S. Gaglio, R. Pirrone, Conceptual representations of actions for autonomous robots, *Robotics and Autonomous Systems,* **34** (2001), 251-263.

[10] A. Chella, M. Frixione, S. Gaglio, Anchoring symbols to conceptual spaces: the case of dynamic scenarios, *Robotics and Autonomous Systems,* **43** (2003), 175-188.

[11] A. Chella, M. Frixione, S. Gaglio, A cognitive architecture for robot self-consciousness, *Artificial Intelligence in Medicine,* **44** (2008), 147-154.

[12] A. Chella, A cognitive architecture for music perception exploiting conceptual spaces, in F. Zenker, P. Gärdenfors (eds.), *Applications of conceptual spaces. The Case for Geometric Knowledge Representation,* 187–203, Springer, Heidelberg, 2015.