Conceptual Spaces for Artificial Intelligence: Formalization, Domain Grounding, and Concept Formation

Lucas Bechberger[*]
Institute of Cognitive Science, Osnabrück University
lucas.bechberger@uni-osnabrueck.de

In artificial intelligence, one can distinguish two layers of knowledge representation: On the one hand, there is the symbolic layer, where abstract knowledge is represented in a structured, logic-based format. On the other hand, there is the subsymbolic layer, where perceptual knowledge is stored in a numeric way, e.g., in the form of weights within a neural network. Ultimately, both approaches will have to be combined in order to arrive at a truly integrated system. It is however still unclear how exactly to accomplish this.

The highly influential framework of conceptual spaces [2] proposes to solve this problem by using an intermediate conceptual layer based on geometric representations: One can identify abstract symbols from the symbolic layer with regions in a high-dimensional space whose dimensions are based on subsymbolic perceptual processing.

Although the framework offers an elegant geometric way of representing conceptual knowledge, it does not provide concrete mechanisms that describe how the overall structure of the conceptual space and the concepts in it can be learned. Concept formation [3], i.e., the process of incrementally creating a meaningful hierarchical categorization of unlabeled data points, can potentially fill this gap.

My research goal is to devise a concept formation process for the conceptual spaces framework that discovers meaningful concepts in an unsupervised way based on perceptual input. A successful implementation of such a system requires three ingredients:

[*] ORCID: 0000-0002-1962-1777

1. A mathematical formalization of the conceptual spaces framework which is both thorough and easily implementable.

2. A principled way of grounding the dimensions of a conceptual space in subsymbolic perceptual processing.

3. A concept formation algorithm that groups points in the conceptual space into meaningful regions.

In my talk, I will first introduce my mathematical formalization of the conceptual spaces framework along with its implementation. This formalization aims to represent correlations between domains (such as "red apples are sweet and green apples are sour") in a geometric way. I will illustrate that convex sets prevent such a geometric representation of correlations, whereas star-shaped sets do not. The proposed formalization includes not only a parametrically describable class of star-shaped sets, but also a large variety of operations on these sets.

Next, I will present my proposal for grounding the dimensions of a conceptual space in subsymbolic perceptual processing. I propose to use neural networks like InfoGAN [1] or β-VAE [4] for learning the dimensions of domains with an unclear internal structure (e.g., shape). These neural networks have been shown to extract meaningful features from unlabeled data sets – meaningful features which can potentially serve as dimensions of a conceptual space. I will illustrate this idea by some preliminary results obtained so far.

Finally, I will sketch the envisioned concept formation process: I will argue that an incremental clustering algorithm is needed and that my proposed formalization is capable of supporting it. I will further argue that language games [5] can provide valuable additional constraints for this clustering process by enforcing a conceptualization that is grounded not only in perception, but also in communication.

References

[1] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, 2016.

[2] Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought.* MIT press, 2000.

[3] John H. Gennari, Pat Langley, and Doug Fisher. Models of Incremental Concept Formation. *Artificial Intelligence*, 40(1-3):11–61, September 1989.

[4] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In *5th International Conference on Learning Representations*, 2017.

[5] Luc Steels. *The Talking Heads Experiment: Origins of Words and Meanings.* Language Science Press, 2015.