Cognitive Component Analysis

Ling Feng
Intelligent Signal Processing,
Informatics and Mathematical Modeling,
Technical University of Denmark, Denmark
lf@imm.dtu.dk

Abstract

Cognitive Science has attracted a new level of prosperity 'consciousness' of engineers during recent decades. One of the reasons is that people begin to wonder whether statistically optimal representations by a variety of machine learning methods are aligned with human cognitive activities. The evolution of human perception system and cognition system is a long-time adaptation process and an on-going interaction between natural environments and natural selection. During the course of charting family trees of various species, evolutionary biologists intended to distinguish between "primitive" and "derived" features. Primitive features group alliances agents; and derived features are prone to enlarge the difference among individuals within a specific agent. Wagensberg linked the difference of individuals to the importance of *independence* for successful 'life forms': *A living individual is part of the world with some identity that tends to become independent of the uncertainty of the rest of the world.* Wagensberg also points out that by creating alliances agents can give up independence for the benefit of a group, which in turns may increase independence for the group as an entity.

Our *independence* hypothesis has been inspired by intriguing facts from using independent component analysis (ICA) algorithm. As a consequence of evolution, human perception system can model complex multi-agent scenery. Humans' ability of using a broad spectrum of cues for analyzing perceptual input and identification of individual agents, has been studied and furthermore stimulated in computers. The resulting theoretically optimal representations achieved by using a variety of ICA closely resembles representations found in human perceptual systems on visual contrast detection, on visual features involved in color and stereo processing, and on representations of sound features.

As a consequence, COgnitive Component Analysis (COCA) was first brought to bear in 2005: the process of unsupervised grouping of data such that the resulting group structure is well-aligned with that resulting from human cognitive activity. We investigated the independent cognitive component hypothesis, which asks the question: *Do humans also use these information theoretically optimal 'ICA' methods in more generic and abstract data analysis.* COCA has been applied to broad topics to review low-level cognitive components. These evidences confirmed that ICA is relevant for representing semantic structure, in text and social networks and musical features; more strikingly for representing information embedded in speech signals, such as phoneme, gender, speaker identity, and even height.

Human learns strategies from one perceptual domain to another, and apply them in more or less distinct categories, such as events or objects grouping. In machine learning, the label structures found by unsupervised learning using some real world data sets, are consistent with the labels derived from human cognition. 'Ray'-structures in latent semantic analysis-like plots, are the key COCA phenomena. Evidences have been found that 'ray'-structures discovered by COCA, which are understood as the human cognition labels, coincide with labels of the samples in the relevant feature space. The fact that structures by COCA are aligned with labels structures, highlights the possibility of using unlabeled data in supervised learning methods.