Journal of Mathematical Psychology 70 (2016) 21–34

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Similarity, kernels, and the fundamental constraints on cognition

Reza Shahbazi [a], Rajeev Raizada [b,1], Shimon Edelman [c,*,1]

[a] Weill Cornell Medicine, Cornell University, United States
[b] Department of Brain and Cognitive Sciences, University of Rochester, United States
[c] Department of Psychology, Cornell University, United States

HIGHLIGHTS

- Cognitive necessities.
- Kernels.
- Similarity measurement.
- Tuning curves.

ARTICLE INFO

Article history:
Received 8 March 2015
Received in revised form
1 November 2015
Available online 8 January 2016

Keywords:
Kernel
RKHS
Similarity
Dimensionality
Complexity
Linear separability
Nonlinear

ABSTRACT

Kernel-based methods, and in particular the so-called kernel trick, which is used in statistical learning theory as a means of avoiding expensive high-dimensional computations, have broad and constructive implications for the cognitive and brain sciences. An equivalent and complementary view of kernels as a measure of similarity highlights their effectiveness in low-dimensional and low-complexity learning and generalization — tasks that are indispensable in cognitive information processing. In this survey, we seek (i) to highlight some parallels between kernels in machine learning on the one hand and similarity in psychology and neuroscience on the other hand, (ii) to sketch out new research directions arising from these parallels, and (iii) to clarify some aspects of the way kernels are presented and discussed in the literature that may have affected their perceived relevance to cognition. In particular, we aim to resolve the tension between the view of kernels as a method of raising the dimensionality, and the various requirements of reducing dimensionality for cognitive purposes. We identify four fundamental constraints that apply to any cognitive system that is charged with learning from the statistics of its world, and argue that kernel-like neural computation is particularly suited to serving such learning and decision making needs, while simultaneously satisfying these constraints.

1. Motivation and plan

The concept of similarity is widely used in psychology. Historically, in a philosophical tradition dating at least back to Aristotle, it has served as a highly intuitive, unifying slogan for a variety of phenomena related to categorization. Here's how Hume put it in the *Enquiry* (1748):

> ALL our reasonings concerning matter of fact are founded on a species of Analogy, which leads us to expect from any cause the same events, which we have observed to result from similar causes. Where the causes are entirely similar, the analogy is perfect, and the inference, drawn from it, is regarded as certain and conclusive. [...] Where the objects have not so exact a similarity, the analogy is less perfect, and the inference is less conclusive; though still it has some force, in proportion to the degree of similarity and resemblance.

In the past century, psychologists have turned similarity into a powerful theoretical tool, most importantly by honing the ways in which similarity can be grounded in multidimensional topological or metric representation spaces (see Osgood, 1949 for an early example) or in situations where a set-theoretic approach may seem preferable (Tversky, 1977).

Sometimes criticized as too loose to be really explanatory (e.g., Goodman, 1972), the concept of similarity has eventually been given a mathematical formulation, including a derivation from first principles of the fundamental relationship between similarity and

* Corresponding author.
E-mail address: edelman@cornell.edu (S. Edelman).

1 This work was supported in part by US NSF grant # 1228261 to R. Raizada and S. Edelman.

http://dx.doi.org/10.1016/j.jmp.2015.11.004
0022-2496/

generalization, and its empirical validation (Shepard, 1987). The mathematical developments, in particular, have solidified similarity's status as a theoretical-explanatory construct in cognitive science (Ashby & Perrin, 1988; Edelman, 1998; Goldstone, 1994; Medin, Goldstone, & Gentner, 1993; Tenenbaum & Griffiths, 2001; for a recent review, see Edelman & Shahbazi, 2012).

In the present paper, we explore the parallels between the psychological construct of similarity and its recent mathematical treatment in the neighboring discipline of machine learning, where a family of classification and regression methods has emerged that is based on the concept of a kernel (Schölkopf & Smola, 2002). Insofar as kernels (described formally in a later section) involve the estimation of distances between points or functions (Jäkel, Schölkopf, & Wichmann, 2008, 2009), they are related to similarity. At the same time, there seems to be a deep rift between the two.

On the one hand, similarity-based learning and generalization has long been thought to require low-dimensional representations, so as to avoid the so-called "curse of dimensionality" (Bellman, 1961; Edelman & Intrator, 1997, 2002), as well as to promote the economy of information storage and transmission (Jolliffe, 1986; Roweis & Saul, 2000). Moreover, as no two measurements of the state of the environment are likely to be identical, some abstraction is necessary before learning becomes possible, which calls for information-preserving dimensionality reduction (Edelman, 1998, 1999). On the other hand, the best-known kernel methods, based on the Support Vector Machine idea (Cortes & Vapnik, 1995; Vapnik, 1999), involve a massive increase in the dimensionality of the representation prior to solving the task at hand.

We attempt to span this rift by seeking a common denominator for some key ideas – and, importantly, their mathematical treatment – behind similarity and kernels. In service of this goal, we first identify, in Section 2, four fundamental constraints on cognition, having to do with (i) measurement, (ii) learnability, (iii) categorization, and (iv) generalization. In Section 3, we then show that while on an abstract-functional or task level these constraints appeal to the concept of similarity, on an algorithmic computational level they call for the use of kernels. Section 4 revisits some standard notions from the similarity literature in light of this observation. In Section 5, we illustrate the proposed synthesis by pairing the methods that it encompasses with a range of cognitive tasks and suggest some ways in which these methods can be used to further our understanding of computation in the brain. Finally, Section 6 offers a summary and some concluding remarks.

2. Fundamental constraints on cognition

2.1. A fundamental constraint on measurement

Perception in any biological or artificial system begins with some measurements performed over the raw signal (Edelman, 2008, ch.5). In mammalian vision, for instance, the very first measurement stage corresponds to the retinal photoreceptors transducing the image formed by the eye's optics into an array of neural activities. The resulting signal is extensively processed by the retinal circuitry before being sent on to the rest of the brain through the optic nerve.

Effectively, a processing unit at any stage in the sensory pathway and beyond "sees" the world through some measurement function $\phi(\cdot)$. Importantly, the measurement process is, at least in the initial stages of development, *uncalibrated*, in the sense that the precise form of the measurement function is not known – that is, not explicitly available – throughout the system. For example, the actual, detailed weight, timing profiles, and noise properties of the receptive field of a sensory neuron are implicitly "known" to the neuron itself (insofar as these parameters determine its response to various types of stimuli), but not to any other units in the system. Indeed, for the usual developmental reasons, those parameters vary from one neuron to the next in ways that are underspecified by the genetic code shared by all neurons in an organism.

Even if the system learns to cope with this predicament (as suggested by some recent findings; Pagan, Urban, Wohl, & Rust, 2013), such learning can only be fully effective if driven by calibrated stimuli, which are by definition not available in natural settings. Moreover, a system that relies on learning, be it as part of its development or as part of its subsequent functioning, it must either (i) simultaneously learn the structure of the data and its own parameters, or (ii) learn the former while being insensitive to the latter.

These considerations imply the following fundamental challenge:

Measurement Any system that involves perceptual measurement is confronted with unknowns that it must learn to tolerate or factor out of the computations that support the various tasks at hand, such as learning and categorization (see Tables 4 and 5).

To the best of our knowledge, this is the first statement of the measurement constraint in the literature. On a somewhat related note, Resnikoff (1989) observed that the general measurement uncertainty principle, as formulated by Gabor (1946), is important for understanding perception. For a recent review of uncertainty in perceptual measurement and the role of receptive field learning under this uncertainty, see (Jurica, Gepshtein, Tyukin, & van Leeuwen, 2013).

2.2. Three fundamental constraints on learning

In learning tasks, the need to generalize from labeled to unlabeled data (in supervised scenarios) or from familiar to novel data (in unsupervised scenarios) imposes certain general constraints on the computational solutions (Geman, Bienenstock, & Doursat, 1992). Although here we focus on categorization, where the goal is to learn class labels for data points, these constraints apply also to regression, where the goal is to learn a functional relationship between independent and dependent variables (Bishop, 2006).

According to the standard formulation in computational learning science, the problem of learning reduces, on the most abstract level of analysis, to probability density estimation (Chater, Tenenbaum, & Yuille, 2006). Indeed, the knowledge of the joint probability distribution over the variables of interest allows the learner to compute, for a query point, the value of the dependent variable, given the observed values (measurements) of the independent variables.[3] This basic insight serves as a background for the present discussion.

In this section, we briefly discuss the constraints that apply to (i) the computation of *similarity* among stimuli, (ii) to the *dimensionality* of representation spaces, and (iii) to the *complexity* of the decision surfaces.

2.2.1. Similarity

Estimating the *similarity* among stimuli is arguably the most important use to which sensory data could be put. As mentioned in the introduction, similarity constitutes the only principled basis for generalization Shepard (1987). Therefore, any non-trivial learning from experience (Edelman, 1998; Edelman & Shahbazi, 2012; Hume, 1748; Shepard, 1987) faces the following challenge:

[3] In this sense, the joint probability distribution over the representation space is the most that can be known about a problem. To know more – for instance, to know the directions of causal links between variables – observation alone does not usually suffice (Pearl, 2009; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). This topic is beyond the scope of the present survey.