Representing Stimulus Similarity

Daniel J. Navarro
Department of Psychology
University of Adelaide

December, 2002

Contents

Abstract

The practice of specifying stimulus representations using measures of similarity holds some status in cognitive modelling. Formal theories of cognitive processes such as generalisation, categorisation, identification, and recognition often employ these representations, and therefore rely on theories of stimulus similarity. With this reliance in mind, it is important to limit stimulus representations to those justified by observations of human thought and behaviour, rather than engaging in the questionable practice of specifying stimulus representations on the basis of introspection.

Over the last 50 years, psychologists have developed a range of frameworks for similarity modelling, along with a large number of numerical techniques for extracting mental representations from empirical data. This thesis is concerned with the psychological theories used to account for similarity judgements, as well as the mathematical and statistical issues that surround the numerical problem of finding appropriate representations. Specifically, the thesis discusses, evaluates, and further develops three widely-adopted approaches to similarity modelling: spatial, featural, and tree representation.

The spatial approach locates each stimulus in a multidimensional co-ordinate space, and assumes that the similarity between two stimuli is a function of how close they are to one another. Tree models represent stimuli as the terminal nodes in an acyclic graph, like a genealogical or taxonomic tree. The similarity between two stimuli is considered to be inversely related to the length of the unique path that connects them. Featural representations describe stimuli in terms of a set of characteristics that they either possess or do not possess. Featural similarity is assumed to be increased by shared features and decreased by distinguishing features.

This thesis develops three major themes. The first, discussed in detail in Chapter 3 but reiterated throughout, regards the important question of how to evaluate a representation. Any representation can be considered to be a model purporting to explain the set

of observed similarities, and should be evaluated as such. It is argued that the representation must not only provide a good fit to the data, but do so in the simplest possible manner, and should satisfy such qualitative constraints as interpretability and psychological plausibility. This thesis uses a Geometric Complexity framework to provide an appropriate trade-off between data-fit and model complexity. In doing so, expressions for the statistically principled Geometric Complexity Criterion are derived for several classes of similarity models. Furthermore, an extended investigation of the analytic properties of these measures for featural and tree models is presented, in order to provide an understanding of what makes one representation more complex than another. A briefer discussion of these issues with regard to spatial representations is also provided.

The second main aspect of this thesis is the discussion of featural representation in Chapter 4. Four theories of featural similarity are considered: the common features model, the distinctive features model, Tversky's seminal Contrast Model, and a new theory called the Modified Contrast Model. The Modified Contrast Model differs from Tversky's by assuming that each individual feature is declared to be a commonality or a distinction, rather than a weighted sum of both. These four theories are evaluated with regard to their psychological assumptions, their analytic properties, and their performance when applied to several empirical data sets. In addition to applying these models to pre-existing data, three new data sets are collected in this evaluation. These investigations suggest that the Modified Contrast Model may combine the common and distinctive elements required by a featural theory in a more plausible manner than Tversky's model.

The third theme to this thesis, discussed in Chapter 5, is the development of an approach to spatial representation that allows a single point to represent multiple stimuli. Based on the prototype theory of categorical structure, this approach enables a set of prototypes to be directly inferred from similarity data. The effectiveness of three "prototype scaling" algorithms are evaluated, and the best algorithm is then applied to several data sets in order to illustrate the potential of the approach.

Overall, it is argued that the different representational frameworks are each best suited to different domains, and that it is therefore important to consider their assumptions, and seek to fit models that are appropriate to the context. As discussed in Chapter 6, future work in this regard might treat similarity judgements as decision processes. Finally, no matter which similarity theory is adopted, the measure of a model should take account of its data-fit, its complexity, and its interpretability.