

On The Role of Dataset Complexity in Case-Based Reasoning

Derek Bridge

Department of Computer Science,
University College, Cork
Ireland
d.bridge@cs.ucc.ie

There has been an amount of recent interest in characterizing the complexity of datasets, especially with regard to how this predicts and affects classifier performance. Many complexity measures, especially ones that focus on geometrical properties of the dataset, have been proposed. But this body of work has, as yet, had little influence on Case-Based Reasoning (CBR) research.

In this talk, I review some of the complexity measures and consider their suitability for use in CBR, for example whether they can handle symbolic-valued attributes, which are common in the kinds of domains in which CBR is used. I look at a handful of measures that have been proposed from within the CBR research community, and assess their usefulness. Finally, I consider the usefulness of the complexity measures both from the perspective of case-based classification, but also from the perspective of case base maintenance.