A Selection Approach: Discriminative Model for Nominal Attributes-Based Distance Measures

Authors : Fang Gong

Abstract : Distance measures are an indispensable part of many instance-based learning (IBL) and machine learning (ML) algorithms. The value difference metrics (VDM) and inverted specific-class distance measure (ISCDM) are among the top-performing distance measures that address nominal attributes. VDM performs well in some domains owing to its simplicity and poorly in others that exist missing value and non-class attribute noise. ISCDM, however, typically works better than VDM on such domains. To maximize their advantages and avoid disadvantages, in this paper, a selection approach: a discriminative model for nominal attributes-based distance measures is proposed. More concretely, VDM and ISCDM are built independently on a training dataset at the training stage, and the most credible one is recorded for each training instance. At the test stage, its nearest neighbor for each test instance is primarily found by any of VDM and ISCDM and then chooses the most reliable model of its nearest neighbor to predict its class label. It is simply denoted as a discriminative distance measure (DDM). Experiments are conducted on the 34 University of California at Irvine (UCI) machine learning repository datasets, and it shows DDM retains the interpretability and simplicity of VDM and ISCDM but significantly outperforms the original VDM and ISCDM and other state-of-the-art competitors in terms of accuracy.
Keywords : distance measure, discriminative model, nominal attributes, nearest neighbor

Conference Title : ICDM 2020 : International Conference on Data Mining
Conference Location : Dublin, Ireland
Conference Dates : July 30-31, 2020