

Research with binary classification

Optimal Binning for a Variance Based Alternative of Mutual Information in Pattern Recognition

Atilla Farkas^{a,*}, György Kovács^b

^aUniversity of Debrecen, Faculty of Informatics, PO Box 201, Debrecen 4002, Hungary

^bAnalyst Mode Ltd., Apolló street 5, Borsnagyháza 4303, Hungary

Abstract

Mutual information (MI) is a widely used similarity measure in pattern recognition. MI uses entropy as a measure of uncertainty to quantify the structural similarity of two vectors. Replacing entropy with variance as a measure of uncertainty, an analogous class of similarity measures can be defined and estimated by regression techniques. Recently, the non-linear piecewise constant regression (PWCR) has been proposed to derive similarity measures of this scheme, leading to competitive alternatives of MI. Although PWCR is based on binning, the optimal binning technique for certain problems remained an open question. In this paper, we show mathematically that the optimal binning needs to be aligned with the expected relationship between the vectors being compared. In general, approximately optimal binnings can be found by combinatorial optimization, and in certain cases the optimal binning can be determined by k-means clustering. The theoretical findings are supported by numerical experiments that show a 2.5% increase in the AUC score in simulated pattern recognition scenarios and improved feature rankings in feature selection problems. The results suggest that the proposed binning techniques could improve the performance of PWCR-derived similarity measures in real-world applications.

Keywords: dissimilarity, template matching, matching by tone mapping, optimal binning, explained variance

MSC: 61A05, 61A10, 62M05, 62D15

1. Introduction

Mutual information (MI) and its variations [28, 20, 35, 29, 6] have been successfully applied in a variety of pattern recognition problems. Feature selection in supervised learning is usually based on MI [36], and as a similarity measure, MI is a popular choice in problems where patterns subject to non-linear distortions are to be recognized (for example, template matching [25], registration [26], and disparity map estimation [14]).

In its original formulation in statistics [8], MI describes the interdependence of two random variables ζ and ξ by quantifying the information gained about ζ (ξ) by observing ξ (ζ), and the information is measured in terms of entropy. When ζ and ξ are statistically independent, MI becomes zero [8]. In contrast, MI is the entropy of ξ when ζ is completely determined by ξ [8].

In pattern recognition, binned and normalized similarity measures are usually preferred to facilitate the interpretation of similarity values and their comparability between different schemes [5]. Consequently, normalized forms of MI have been proposed [30]. One of the most common formulations is the uncertainty coeff-

^{*}Corresponding author

Email address: atilla.farkas@uni-due.de (Atilla Farkas), gyorgy.kovacs@am-ltd.com (György Kovács)

Preprint submitted to Neurocomputing

March 25, 2021

Manual extraction of specifics

Experiment
(e.g. $p = 20$, $n = 300$)

Reported scores
(e.g. $acc = 0.938$,
 $sens = 0.758$,
 $spec = 0.953$)

Uncertainty
(e.g. $\epsilon = 0.001$)

Consistency testing

No inconsistency

The scores could be yielded from the experiment

The proposed consistency testing algorithms

Inconsistency

The scores could NOT be yielded from the experiment