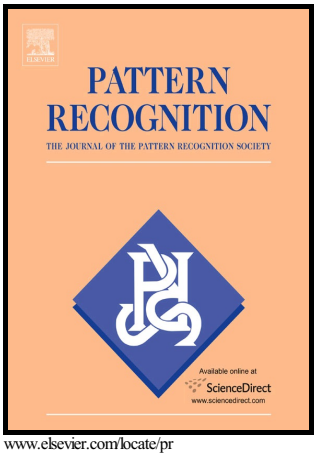
Author's Accepted Manuscript

A New Distance Measure for Non-Identical Data with Application to Image Classification

Muthukaruppan Swaminathan, Pankaj Kumar Yadav, Obdulio Piloto, Tobias Sjöblom, Ian Cheong



www.elsevier.com/locate/pr

PII: S0031-3203(16)30334-X
DOI: http://dx.doi.org/10.1016/j.patcog.2016.10.018
Reference: PR5926

To appear in: *Pattern Recognition*

Received date: 24 May 2016
Revised date: 12 October 2016
Accepted date: 15 October 2016

Cite this article as: Muthukaruppan Swaminathan, Pankaj Kumar Yadav, Obdulio Piloto, Tobias Sjöblom and Ian Cheong, A New Distance Measure for Non-Identical Data with Application to Image Classification, *Pattern Recognition* http://dx.doi.org/10.1016/j.patcog.2016.10.018

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A New Distance Measure for Non-Identical Data with Application to Image Classification

Muthukaruppan Swaminathan[a,b], Pankaj Kumar Yadav[a], Obdulio Piloto[c], Tobias Sjöblom[d], Ian Cheong[a,b,*]

[a] *Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore*
[b] *Department of Biological Sciences, National University of Singapore, Singapore*
[c] *Entopsis LLC, 601 West 20th St, Hialeah, FL 33136, USA*
[d] *Rudbeck Laboratory SE-751 85 Uppsala, Uppsala University, Sweden*

Abstract

Distance measures are part and parcel of many computer vision algorithms. The underlying assumption in all existing distance measures is that feature elements are independent and identically distributed. However, in real-world settings, data generally originate from heterogeneous sources even if they do possess a common data-generating mechanism. Since these sources are not identically distributed by necessity, the assumption of identical distribution is inappropriate. Here, we use statistical analysis to show that feature elements of local image descriptors are indeed non-identically distributed. To test the effect of omitting the unified distribution assumption, we created a new distance measure called the Poisson-Binomial Radius (PBR). PBR is a bin-to-bin distance which accounts for the dispersion of bin-to-bin information. PBR's performance was evaluated on twelve benchmark data sets covering six different classification and recognition applications: texture, material, leaf, scene, ear biometrics and category-level image classification. Results from these experiments demonstrate that PBR outperforms state-of-the-art distance measures for most of the data sets and achieves comparable performance on the rest, suggesting that accounting for different distributions in distance measures can improve performance in classification and recognition tasks.

Keywords: Poisson-Binomial distribution, semi-metric distance, non-identical data, distance measure, image classification, image recognition

*Corresponding author
Email addresses: muthu@tll.org.sg (Muthukaruppan Swaminathan), obdulio.piloto@entopsis.com (Obdulio Piloto), tobias.sjoblom@igp.uu.se (Tobias Sjöblom), ian@tll.org.sg (Ian Cheong)

Accepted manuscript

Download English Version:

https://daneshyari.com/en/article/4969817

Download Persian Version:

https://daneshyari.com/article/4969817

Daneshyari.com