Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 192–199

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS
October 30 – November 1, 2017, Chicago, Illinois, USA

The Uncertainty Area Metric: a Method for Comparing Learning Machines on What They Don't Know

Walker H. Land,Jr.[a] and J. David Schaffer[b]

[a] Retired Emeritus Research Professor, Department of Bioengineering, Binghamton University, Binghamton, NY 13902 USA
[b] Institute for Multigenerational Studies, Binghamton University, Binghamton, NY 13902 USA

Abstract

Schaffer and Land [14] described a method whereby a machine intelligence (MI) process can "know what it doesn't know." In this paper, the concept is illustrated by three examples: the GRNN oracle ensemble method that combines multiple SVM classifiers for detecting Alzheimer's type dementia using features automatically extracted from a speech sample, an Evolutionary Programming and Adaptive Boosting hybrid and a Generalized Regression Neural Network hybrid for classifying breast cancer. The authors assert it is (1) applicable quite directly to a great many other learning classifier systems, and (2) provides an intuitive approach to comparing the performance of different classifiers on a given task using the size of the "area of uncertainty" as a measure of performance metric. This paper provides support for these assertions by describing the steps needed to apply it to a previously published study of breast cancer benign / malignancy prediction, and then illustrates how this "area of uncertainty" may be computed, which is a work in progress, using the GRNN oracle results and a resultant Bayesian network from the Alzheimer's speech research study.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: mixture of expert systems; GRNN oracle; EP-AB hybrid; breast cancer; Bayesian networks; machine intelligence; area of uncertainty

1. Introduction and Background

In the field of Artificial Intelligence (AI), there has long been interest in providing some capability for an AI system to explain how it reached its conclusions. In the days of rule-base expert systems, this capability could be

Corresponding author. Walker H. Land
E-mail address: wland@binghamton.edu

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.067

provided by meta-rules [4]. More recently, however, more complex systems such as deep learning networks with thousands of internal parameters trained on sometimes millions of training cases, have proven much more difficult for providing any schemes for self-explanation [5]. Here, we propose a much simpler method that, while not providing any kind of explanation, is at least capable of providing some level of risk assessment by providing a "flag" along with some of its predictions indicating that it is aware that these predictions may be unreliable. We are confident that this method is very widely applicable – basically to any learning classifier system that has been trained on a set of cases where it knows that some of these cases it has been unable to correctly classify, in addition to not being hampered by explaining why the AI process arrived at a specific decision.

In the background paper, Schaffer and Land [14] presented the development of a general approach for a machine intelligence (MI) algorithm for a system to "know what it doesn't know"; that is, when a predictor/classifier has the ability to know when it's predictions or classifications are unreliable. The methodology presented there is asserted to be applicable to just about any learning predictor/detector/diagnostic MI system which has been trained on a set of known cases. This overview approach is described and illustrated by applying a GA-SVM-GRNN-oracle hybrid to the detection of those sample cases that have Alzheimer's disease from normal controls by using features extracted from a sample of speech and the mini-mental state exam (MMSE).

The method is basically to define a region of uncertainty in feature space around each trouble-maker (TM) case: a training case that the trained classifier system still gets wrong. Two problems that the proposed approach overcomes are (1) the training data sets tend to be somewhat sparse in a high-dimensional feature space, and (2) a precise definition is needed for this region of uncertainty. The first problem they solve by applying the t-SNE algorithm of van der Maatens and Hinton [13] to get a projection of the training cases in a low-dimensional (2D) space where the case distributions may be better distributed. The second problem is addressed by linearly interpolating the known prediction errors (< 0.5 for correctly classified cases, and >0.5 for TM cases) to locate estimated cut-points where the prediction error is equal to 0.5. They then provide an algorithm that chooses a set of these cut-points from the closest neighbours of the TM case using a user-defined smoothness parameter. They then propose that one may compare two or more alternative classifier approaches to a given data set by examining the different proportions of the 2D reduced feature space that are the regions of uncertainty for the alternative classifiers.

This approach differs from the traditional approach of comparing classifiers in terms of classification accuracy. It may be applied simply to a training set without doing cross-validation, but of course, it may also be applied in this way, but retains the need to specify a decision threshold. Another traditional method for comparing classifiers that does not require a decision threshold is to compute the area under the ROC curve (AUC). This method also differs from a simple accuracy comparison in that it combines knowledge of the number of errors (TMs) with a heuristic for quantifying the extent of uncertainty in a projected feature space.

In the next two sections of this paper we provide illustrations of how this process may be used. We show how a learning classifier method consisting of evolutionary programming (EP) optimized multilayer perceptron (MLP) classifiers applied using the Adaboost approach (EP-AB hybrid) applied to a task of predicting the malignancy of a breast lesion using clinical features and BI-RADS features from mammograms, induces the same type of TMs as the GRNN oracle approach. We then outline how a Bayesian network (BN) induced for the same Alzheimer's speech detection task can give rise to an area of uncertainty calculation that is directly comparable to that from the GRNN oracle on the same task. This research is an ongoing process, but we believe readers will find the concept appealing.

Nomenclature

AB	Adaptive Boosting, the Adaboost algorithm
AD	Alzheimer's Disease Group
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
BI-RADS	Breast Imaging-Reporting and Data System
BN	Bayesian Network
CAS	Complex Adaptive System
EP	Evolutionary Programming
FN	False Negative
FP	False Positive