Information Sciences 462 (2018) 242–261

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Visual-based analysis of classification measures and their properties for class imbalanced problems

Dariusz Brzezinski*, Jerzy Stefanowski, Robert Susmaga, Izabela Szczęch

Institute of Computing Science, Poznan University of Technology, ul. Piotrowo 2, 60–965 Poznan, Poland

ARTICLE INFO

Article history:
Received 21 July 2017
Revised 28 February 2018
Accepted 8 June 2018
Available online 15 June 2018

Keywords:
Classification performance measures
Visualisation
Barycentric system
Class imbalance

ABSTRACT

With a plethora of available classification performance measures, choosing the right metric for the right task requires careful thought. To make this decision in an informed manner, one should study and compare general properties of candidate measures. However, analysing measures with respect to complete ranges of their domain values is a difficult and challenging task. In this study, we attempt to support such analyses with a specialized visualisation technique, which operates in a barycentric coordinate system using a 3D tetrahedron. Additionally, we adapt this technique to the context of imbalanced data and put forward a set of measure properties, which should be taken into account when examining a classification performance measure. As a result, we compare 22 popular measures and show important differences in their behaviour. Moreover, for parametric measures such as the F_β and $IBA_\alpha(G\text{-}mean)$, we analytically derive parameter thresholds that pinpoint the changes in measure properties. Finally, we provide an online visualisation tool that can aid the analysis of measure variability throughout their entire domains.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Classification is one of the most important machine learning tasks, commonly applied to many real-world problems. One of the crucial ingredients of this supervised learning task is the selection of a performance measure that allows the user to discern good classifiers from bad ones. An appropriate measure should support choosing the best classifier among several candidates and help tune its parameters. As a result, the selected performance measure is responsible for the optimization of the learning process [10].

Although researchers often focus on overall predictive accuracy, which promotes recognizing the highest number of instances of all target classes, the choice of the evaluation measure is not always a unique and simple decision. This is due to the fact that many practical classification problems require more sophisticated approaches to dealing with errors referring to particular subsets of instances. This realization has paved the way to proposals and analyses of many other performance measures.

Unfortunately, comparing several alternative classification measures and selecting the most appropriate one is not an easy task, even for experts constructing learning systems. In practice, crucial aspects of measures, such as atypical values they can take, their monotonicity, or their symmetry, are rarely taken into account during measure selection. As the understanding of

* Corresponding author.
E-mail addresses: dariusz.brzezinski@cs.put.poznan.pl (D. Brzezinski), jerzy.stefanowski@cs.put.poznan.pl (J. Stefanowski), robert.susmaga@cs.put.poznan.pl (R. Susmaga), izabela.szczech@cs.put.poznan.pl (I. Szczęch).

https://doi.org/10.1016/j.ins.2018.06.020
0020-0255/© 2018 Elsevier Inc. All rights reserved.

measure properties is crucial for improving classification models and their learning process, we postulate the need for more research on analysing measures and developing methods that support researchers in this task.

The analysis of measure properties may be done by examining a measure's definition. However, theoretical investigations are often very laborious and time consuming, especially when multi-dimensional aspects provided by the confusion matrices need to be taken into account. Owing to these difficulties, such an analysis could be alternatively carried out with visual techniques, in order to aid researches in finding and interpreting measure properties. Such visual-based insight is of utmost importance especially for classification tasks with complex example distributions, such as class imbalanced data.

In imbalanced data one of the target classes, called the minority class, contains much less examples than the remaining (majority) classes. Imbalanced data constitutes a great difficulty for standard learning algorithms, as classifiers tend to be biased towards the majority class and misclassify minority examples even though their correct recognition is usually more important [23,29]. The prevalence of class imbalance in many practical tasks has led to the development of various methods for improving classifiers learning from skewed data [4,30,43]. In this context, much work has also been done in the field of classification measures. Since typical performance measures, such as classification accuracy, are not appropriate for imbalanced data [13,35], several more relevant metrics have been considered. The most popular ones include *precision, recall* (*sensitivity*), *specificity*, and their aggregates, e.g. *G-mean* or F_1*-score*. These and other measures for imbalanced data are typically defined on the basis of confusion matrices summarizing the predictions of a binary classifier. Looking into related studies, one can notice that the number of such measures is relatively high and that each represents different aspects of classification performance, often leading to quite different interpretations [23]. This shows that there is no single measure that is the best choice in all situations. However, which measure is used in a given problem seems to be, to a large extent, dictated simply by the measure's popularity rather than a thorough discussion of its properties. That is why supporting visual analysis of performance measures is particularly important for class imbalanced problems and there is a need for new analysis methods.

In this paper, we put forward a new visualisation technique for analysing entire domains of classification performance measures. The proposed visualisation depicts all possible configurations of predictions in a confusion matrix, regardless of the used classifier. The method adapts an approach originally created for rule interestingness measures to the context of classification [38]. Contrary to existing performance measure visualisations, such as ROC space [14], the proposed method presents measures in a space which is defined directly on elements of the confusion matrix, is easily interpretable in 3D, and remains defined for all elements of the domain.

Acknowledging the need for systematic discussion on properties of performance measures, we also put forward ten properties which should be taken into account in the context of imbalanced data. The proposed properties characterize the behaviour of measures, reveal unexpected or atypical values, and can help researchers select measures suitable for a given learning problem.

Consequently, we consider 22 classification measures, chosen from the literature for their popularity and diversity, and analyse them with respect to the proposed properties. The analysis is performed using our visualisation technique that allows to examine thoroughly the measures in complete ranges of their values. As a result, important differences in the measures' behaviour are highlighted, constituting a new theoretical contribution to research on class imbalance data and providing practical guidelines for selecting measures for particular classification problems. Finally, it is demonstrated that visualisations can also lead to analytical derivations of measure properties.

The main contributions of the paper are as follows:

- In Section 3, we present a technique for visualizing classification performance measures using the barycentric coordinate system and discuss its characteristics. Additionally, we present an online tool that implements the proposed technique and allows the analysis of both predefined as well as user-defined measures.
- In Section 4, we put forward ten properties, providing knowledge on the behaviour of classifier performance measures for imbalanced problems. The introduced properties involve maxima, minima, elements of symmetry, monotonicity, and undefined measure values.
- In Section 5, using the proposed visualisation technique we analyse and compare 22 classification measures with respect to the proposed properties. Moreover, we present a critical discussion on the applicability of measures with particular properties for imbalanced problems.
- In Section 6, we perform a set of case studies on how the proposed properties can be used to compare selected measures. More precisely, we study the differences between F_1*-score, G-mean, Mathews Correlation Coefficient*, and *Optimized Precision*, the effect of internal parametrisation on the F_β measure, and external parametrisation for IBA_α(*G-mean*). Apart from visual inspection, we analytically derive threshold parameter values for selected measures.
- In Section 7, we discuss the most important issues in analysing classification performance measures and draw lines of potential further investigations.

2. Related works

2.1. Classifier performance measures

Classifiers can be assessed in many aspects, such as their predictive ability, training time, memory usage, model complexity, interpretability, or other criteria [26]. In this paper, we consider predictive performance only and focus on measures

Download English Version:

https://daneshyari.com/en/article/6856241

Download Persian Version:

https://daneshyari.com/article/6856241

Daneshyari.com