Information Measure Toolbox for Classifier Evaluation on Open Source Software Scilab

Bao-Gang Hu[1,2]
[1] NLPR/LIAMA, Institute of Automation
[2] Beijing Graduate School
Chinese Academy of Sciences
P.O. Box 2728, Beijing, 100080 China
hubg@nlpr.ia.ac.cn

Abstract

Classifier evaluation has been a wide and long time concern in the areas of pattern recognition [1] and data processing [2]. In recent years, it becomes a hot topic in machine learning, data mining, and web search [3-5]. A selection of proper evaluation measures is still an open problem in machine learning. This is particularly true for classifications. If considering binary classifications, one can find out over thirty measures available for uses [5]. However, most of the existing measures are generally unable to handle the generic classifications including rejection class. In engineering practice, rejection, or abstention, on samples is a common scheme which generally improves the classification accuracy [6-7]. However, the conventional performance measures may fail to provide an objective ranking on two classifiers if they hold different accuracy and rejection rates.

This work is an extension of our early studies [8-10] on developing the information measures for classifier evaluations, particularly on the programming implementation. In our previous studies, we demonstrated that one of the most advantages of information measures over the conventional performance measures is the ability of handling classifications with rejection operations. This work, falling within OpenPR Project, aims to build up a novel toolbox of information measures for classifier evaluations on open-source software Scilab. We select Scilab platform due to its powerfulness and flexibility in applications from both research and educational purposes.

To simplify the study, we presume that the basic data available for classifier evaluations is a confusion matrix. For a generic classification problem, an augmented confusion matrix, **C**, is considered below in which one column for a rejected class is added on a conventional confusion matrix:

$$\boldsymbol{C}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1\mathrm{m}} & c_{1(m+1)} \\ c_{21} & c_{22} & \cdots & c_{2\mathrm{m}} & c_{2(m+1)} \\ & & \cdots & & \\ c_{m1} & c_{m2} & \cdots & c_{mm} & c_{m(m+1)} \end{bmatrix}, \tag{1}$$

where c_{ij} represents the sample number of the ith class that is classified as the jth class. The total class number is m. The row data corresponds to the exact classes, and the column data corresponds to the prediction classes. The last column represents a rejected class. In [10], we have examined mostly-commonly used information measures, such as mutual information [11], KL divergence [11], and cross entropy [1]. For the reason of normalization and symmetric property requirements, twenty four normalized information measures, denoted by NI_k (k=1,2, ..., 24), are formed for a systematic comparison. For realizing an objective evaluation, we apply Shannon entropy for the definition of each measure.

In the implementation of programing, the specific attention is made on the singularity aspect. When calculating entropy, one may encounter two types of singularities, namely

a) removable singularity (say, 0*log0 = 0), and
b) absolute singularity (say, a*log0 = -∞).

Removable singularity is a specific feature for some function at singularity point the function is undefined, but its limit exists at this point. In the program on Scilab, we proposed the specific procedures to deal with two types of singularities. In the program, by adding a very small bias, eps=2.2e-64, to the confusion matrix **C**, one can obtain the reasonably good value at the removable singularity point. The

program will output “S” for the NI when the absolute singularity is detected in the calculation. This output will indicate that the current NI is unable to delivery an applicable measure value from the given confusion matrix.

Numerical examples are given by two-class and three-class problems respectively for cases including rejection operations. Based on the program, users are able to obtain the empirical findings about cons and pros of each measure from the numerical examples. They can also test their own examples easily, or enrich the toolbox by adding novel information measures in a similar way. To the author's knowledge, the program developed in this work seems to be its first one in the public on the subject of classifier evaluations. We hope the information measure toolbox with its program in this work will promote more systematic studies on the comparisons among both information-based measures and performance-based measures.

Keywords: Abstaining classifiers, entropy, mutual information, divergence, evaluation, open source, toolbox

References:

[1] Duda, R.O. Hart, P.E., Stork, D.: Pattern Classification (2nd eds.). John Wiley, NY (2001)

[2] Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendon Press, London (1995)

[3] Japkowicz, N., “Why Question Machine Learning Evaluation Methods”, AAAI Workshop on Evaluation Methods for Machine Learning, (2006)

[4] Huang, J. and Ling, C.X., "Constructing New and Better Evaluation Measures for Machine Learning". In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), 859-864, (2007).

[5] Lavesson, N., Davidsson, P., “Analysis of Multi Criteria Methods for Algorithm and Classifier. Evaluation”. In: Proc. 24th Annual Workshop of the Swedish, (2007)

[6] Tortorella, F., “An Optimal Reject Rule for Binary Classifiers”, In: Joint IAPR International Workshops SSPR 2000 and SPR 2000, Lecture Notes in Computer Science, No. 1876, 611-620, Alicante, Spain, 2000, Springer-Verlag.

[7] Pietraszek, T., “Optimizing abstaining classifiers using ROC analysis”, in: Proceedings of the 22nd International Conference on Machine Learning, 2005.

[8] Wang, Y. and Hu, B.-G., “A study on integrated evaluating kernel classification performance using statistical methods”, Chinese Journal of Computers, 31, 64-74, 2008 (in Chinese)

[9] Hu, B.-G., Wang, Y., “Evaluation Criteria Based on Mutual Information for Classifications Including Rejected Class. Acta Automatica Sinica. 34, 1396--1403 (2008)

[10] Hu, B.-G., He, R., and Yuan, X.-T., “Information-Theoretic Measures for Objective Evaluation of Classifiers”, submitted to a journal, (2009)

[11] Cover, T.M. and Thomas J.A., Elements of Information Theory. John Wiley, NY (2006)