Computer Physics Communications 185 (2014) 1172–1176

Contents lists available at ScienceDirect

Computer Physics Communications

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

Visualization of the significance of Receiver Operating Characteristics based on confidence ellipses☆

Nicholas V. Sarlis *, Stavros-Richard G. Christopoulos

Solid State Section and Solid Earth Physics Institute, Physics Department, University of Athens, Panepistimiopolis, Zografos 157 84, Athens, Greece

A R T I C L E   I N F O

Article history:
Received 20 June 2013
Received in revised form
19 November 2013
Accepted 4 December 2013
Available online 27 December 2013

Keywords:
Receiver Operating Characteristics (ROC)
Complex systems
Systems obeying power laws
Significance level
p-value

A B S T R A C T

The Receiver Operating Characteristics (ROC) is used for the evaluation of prediction methods in various disciplines like meteorology, geophysics, complex system physics, medicine etc. The estimation of the significance of a binary prediction method, however, remains a cumbersome task and is usually done by repeating the calculations by Monte Carlo. The FORTRAN code provided here simplifies this problem by evaluating the significance of binary predictions for a family of ellipses which are based on confidence ellipses and cover the whole ROC space.

Program summary

Program title: VISROC.f

Catalogue identifier: AERY_v1_0

Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AERY_v1_0.html

Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland

Licensing provisions: Standard CPC licence, http://cpc.cs.qub.ac.uk/licence/licence.html

No. of lines in distributed program, including test data, etc.: 11511

No. of bytes in distributed program, including test data, etc.: 72906

Distribution format: tar.gz

Programming language: FORTRAN.

Computer: Any computer supporting a GNU FORTRAN compiler.

Operating system: Linux, MacOS, Windows.

RAM: 1Mbyte

Classification: 4.13, 9, 14.

Nature of problem:
The Receiver Operating Characteristics (ROC) is used for the evaluation of prediction methods in various disciplines like meteorology, geophysics, complex system physics, medicine etc. The estimation of the significance of a binary prediction method, however, remains a cumbersome task and is usually done by repeating the calculations by Monte Carlo. The FORTRAN code provided here simplifies this problem by evaluating the significance of binary predictions for a family of ellipses which are based on confidence ellipses and cover the whole ROC space.

Solution method:
Using the statistics of random binary predictions for a given value of the predictor threshold ε_t, one can construct the corresponding confidence ellipses. The envelope of these corresponding confidence ellipses is estimated when ε_t varies from 0 to 1. This way a new family of ellipses is obtained, named *k*-ellipses, which covers the whole ROC plane and leads to a well defined Area Under the Curve (AUC). For the latter quantity, Mason and Graham [1] have shown that it follows the Mann–Whitney U-statistics [2] which can be applied [3] for the estimation of the statistical significance of each *k*-ellipse. As the transformation

☆ This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding author. Tel.: +30 210 7276736.
E-mail address: nsarlis@phys.uoa.gr (N.V. Sarlis).

http://dx.doi.org/10.1016/j.cpc.2013.12.009

is invertible, any point on the ROC plane corresponds to a unique value of k, thus to a unique p-value to obtain this point by chance. The present FORTRAN code provides this p-value field on the ROC plane as well as the k-ellipses corresponding to the (p =)10%, 5% and 1% significance levels using as input the number of the positive (P) and negative (Q) cases to be predicted.

Unusual features:
In some machines, the compiler directive -O2 or -O3 should be used to avoid NaN's in some points of the p-field along the diagonal.

Running time:
Depending on the application, e.g., 4s for an Intel(R) Core(TM)2 CPU E7600 at 3.06 GHz with 2 GB RAM for the examples presented here

References:

[1] S.J. Mason, N.E. Graham, Quart. J. Roy. Meteor. Soc. 128 (2002) 2145.

[2] H.B. Mann, D.R. Whitney, Ann. Math. Statist. 18 (1947) 50.

[3] L.C. Dinneen, B.C. Blakesley, J. Roy. Stat. Soc. Ser. C Appl. Stat. 22 (1973) 269.

1. Introduction

Receiver Operating Characteristics (ROC) graphs is a technique currently used [1–7] for estimating the predictability of various complex systems and has already found useful applications in various fields like medicine, e.g. see [8], meteorology [9,10], etc. As suggested by Fawcett [11] ROC graphs depict the trade off between hit rates and false alarm rates with a conceptually simple way which also applies to the case of skewed class distributions which is usually the case in the physics of complex systems.

We limit ourselves in the case of binary predictions. In this case, there are two classes of events **p** or **n**; for example in the case of extreme events in complex systems an event is classified as **p** if its magnitude exceeds a given threshold M_t otherwise it is considered **n**. Before the occurrence of each event a predictor ϵ is assigned based on some prediction algorithm and a hypothesized class **P** or **N** for the forthcoming event is decided on the basis of whether ϵ *exceeds* or not, respectively, the predictor threshold value ϵ_t. If the hypothesized class is **P** and the event is **p** we have a successful prediction called True Positive (TP). Similarly, if the hypothesized class is **N** and the event is **n** we again have a successful prediction called True Negative (TN). If, however, the hypothesized class is **P** and the event is **n** we have a False Positive (FP) unsuccessful prediction. Finally, if the hypothesized class is **N** and the event is **p** we have a False Negative (FN) unsuccessful prediction. Assuming that in total we examine P events of the **p** class and Q events of the **n** class, one defines [11] the True Positive rate (TPr) – or hit rate (H) – as the ratio of the totality of TP's over P,

$$H \equiv \frac{|TP|}{P} = \frac{|TP|}{|TP| + |FN|}, \tag{1}$$

and the False Positive rate (FPr) – or false alarm rate (F) – as the ratio of the totality of FP's over Q,

$$F \equiv \frac{|FP|}{Q} = \frac{|FP|}{|FP| + |TN|}. \tag{2}$$

The ROC curve is obtained as we vary ϵ_t and plot H as a function of F, e.g. see Fig. 1.

In the example of Fig. 1, we depict the ROC curves for three different values of M_t which result [6] when considering as predictor the number ϵ_k of successive extrema of the aftershock magnitude time series m_{k+1} in the case of the 1999 Hector Mine earthquake. As we vary the threshold ϵ_t an ROC curve is obtained, e.g. see the results depicted by the red squares which correspond to a target magnitude $M_t = 5.0$. In this particular example, the hypothesized class is either **P** or **N** depending on whether ϵ_k is *smaller* or larger, respectively, than ϵ_t. We observe that the diagonal (black solid line) corresponds to random predictions and would have been obtained as an ROC curve for a random predictor if both P and Q had tended to infinity. If we repeat the experiment using the definition of the hypothesized classes of the previous paragraph, i.e., interchanging **P** with **N**, we obtain the ROC curves that lie in the lower triangle which are also depicted with thick lines without symbols in Fig. 1 which are symmetric images of the originals with respect to the center $(F_c, H_c) = (1/2, 1/2)$. Thus, the statistical significance of an ROC curve depends on its deviation from the $H = F$ diagonal and by negating the condition used for the construction of the hypothesized classes we can 'reflect' an ROC curve with respect to the center of the ROC space. Moreover, in the present example, as we change M_t we drastically affect the number P of **p** events, see Table 1, because earthquake magnitudes follow the Gutenberg–Richter law [12]. The corresponding ROC curves may become more distant from the diagonal, but this may be misleading as their statistical significance varies when P (and Q) change. It is the aim of the present paper to present a plausible visualization method for the statistical significance of an ROC curve together with a FORTRAN code that generates the corresponding significance intervals. In Section 2, we will present the proposed k-ellipses that cover the ROC space and obtain the related statistics. Section 3 discusses the implementation of the method in FORTRAN and Section 4 summarizes the results.

2. The *k*-ellipses family

In order to estimate the statistical significance of an ROC curve, we need to compare it with similar results obtained when using a random predictor. Without loss of generality, we assume that the predictor threshold ϵ_t varies in the range [0, 1]. Then, as a random predictor we can consider a uniformly distributed random number u_i in the same interval. Under these assumptions, the conditional probabilities $P(\mathbf{P}|\mathbf{p})$ and $P(\mathbf{P}|\mathbf{n})$ to obtain the hypothesized class **P** under the assumption that the event is either **p** or **n** are both equal to ϵ_t. Thus, the number l of TP's as well as the number m of FP's follow the binomial distribution with attempt probability ϵ_t for P and Q attempts, respectively. The mean value of the hit rate H and the false alarm rate F result in

$$\langle H \rangle = \frac{\langle l \rangle}{P} = \epsilon_t, \tag{3}$$

$$\langle F \rangle = \frac{\langle m \rangle}{Q} = \epsilon_t, \tag{4}$$