SCV 2008: December 1-3, Taipei, Taiwan (Abstract)

NONPARAMETRIC COMPARISON OF ROC CURVES TESTING EQUIVALENCE AND CLUSTERING

Jaromír Antoch, Luboš Prchal, Pascal Sarda

Keywords: ROC curves, binary classification, collocations, clustering.

Receiver operating characteristic (ROC) curves are a popular and widely used tool that can help to summarize the overall performance of diagnostic methods and/or classifiers assigning individuals $g \in \mathcal{G} = \mathcal{G}_0 \cup \mathcal{G}_1, \mathcal{G}_0 \cap \mathcal{G}_1 = \emptyset$, into one of the groups $\mathcal{G}_0$ or $\mathcal{G}_1$. Figure 1 provides examples of ROC curves, that visualizes the probability of correctly classified $\mathcal{G}_1$ individuals against the probability of misclassified $\mathcal{G}_0$ individuals for all critical values of the diagnostic method.

In practice, ROC curves are often used to compare several diagnostic methods (classifiers). Typically it is assumed that the method with its ROC curve closest to the point $[0, 1]$ is the best one for the particular problem. However, we point out that the optimality criteria can be more complex and may depend on actual situation. Figure 1 displays three plots, each with a pair of ROC curves corresponding to different association measures. It illustrates three typical situations one can meet in practice.

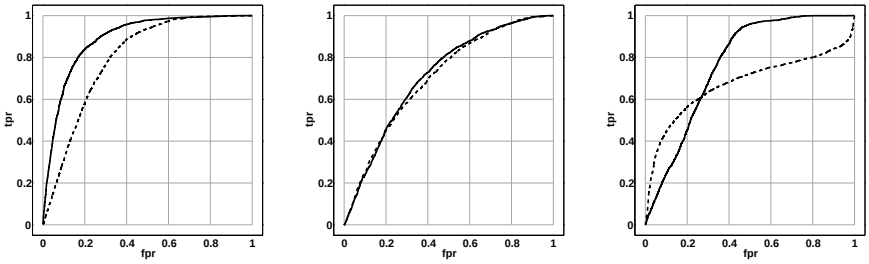

Figure 1. Examples of ROC curves for several linguistic measures.

First, everyone would agree that the solid curve in Figure 1a outperforms the dashed one. Figure 1b seems to be the opposite case, because both association measures provide, at least optically, equivalent ROC curves. Finally, concerning the overall performance of the presented ROC curves, the situation in Figure 1c is not at all clear. In all three cases, nevertheless, a natural question arises: *Are these ROC curves significantly different?*

There exist several methods for testing the difference between two ROC curves, some of them being listed in references. During the lecture the problem of testing equivalence of two ROC curves will be addressed and illustrated on a real data set stemming from the field of computational linguistics. A transformation of corresponding ROC curves, that motivates a test statistic based on a distance of two empirical quantile processes will be suggested, its asymptotic distribution shown and a simulation scheme enabling to find critical values proposed. The procedure is applied to the ROC curves measuring quality of the automatic collocation extraction. It is shown that obtained p-values can be used as a distance between the curves and enable ROC curves clustering.

References

Bamber D. (1975). *J. Math. Psych.* **12**, 387 – 415.
Beam C. A. and Wieand H. S. (1991). *Biometrics* **47**, 907 – 919.
Chen C.-H. (2002). *Statist. Sinica* **12**, 7 – 29.
DeLong E. R., DeLong D. M. and Clarke-Pearson D. L. (1988). *Biometrika* **44**, 837 – 846.
Greenhouse S. W. and Mantel N. (1950). *Biometrics* **6**, 399 – 412.
Venkatraman E. S. and Begg C. B. (1996). *Biometrika* **83**, 835 – 848.
Wieand S., Gail M. H., James B. R. and James K. L. (1989). *Biometrika* **76**, 585 – 592.
Zhou X. H., McClish D. K. and Obuchowski N. A. (2002). J. Wiley, New York.

Acknowledgement: Work has been supported by grants GAČR 201/06/0186 and MSM 0021620839. Authors thank Chun-houh Chen and Institute of Statistical Science Academia Sinica for hospitality, support and thorough care.

Addresses: Dept. of Probability and Statistics, Charles University in Prague, Sokolovská 83, CZ – 186 75 Praha 8, Czech Republic; {jaromir.antoch,lubos.prchal}@mff.cuni.cz
Institut de Mathématiques de Toulouse, UMR 5219, Université Paul Sabatier, 118 route de Narbonne, F – 310 62 Toulouse cedex, France; sarda@cict.fr