

2014 – 02

Performance Measures in
Classification Problems with
Class-Imbalanced Data

Bo-Shiang Ke and Yuan-chin Ivan Chang

Dec. 09, 2014

Performance Measures in Classification Problems with Class-Imbalanced Data

Bo-Shiang Ke¹ and Yuan-chin Ivan Chang²

¹Institute of Statistics, National Chiao Tung University, 1001,
University Road, Hsinchu, 30010, Taiwan
e-mail: naivete0907@gmail.com

²Institute of Statistical Science, Academia Sinica, 128, Academia
Road Section 2, Taipei, 11529, Taiwan
e-mail: ycchang@stat.sinica.edu.tw

Summary

Supervised classification problem has become the major issue and enormous techniques are invented to deal with this problem in all kinds of different practical considerations. Some performance measures have accompanied with these techniques. However, the perplexities arise from the choice among these measures, for each measure may have its own purpose. This paper aims not only to provide the overview of traditional criteria but also to incorporate more useful evaluation measures in applications; besides, we also investigate the connections among them as well as the behavior while facing imbalanced data set.

Key words: AC_1 statistic; lift index; normalized mutual information; balanced accuracy; cumulated gain chart; cumulated lift chart.

1. Introduction

Supervised statistical learning always play an important role in our daily life. For example, we would like to know the relationship between salary and the education level, to distinguish whether patients suffer specific disease or not, and so forth. Roughly speaking, supervised statistical learning