Abstract

In recent research the classifications of imbalanced data sets have received considerable attention. It is natural that due to the class imbalance the classifier tends to favour majority class. In this paper we investigate the performance of different methods for handling data imbalance in the microcalcification classification which is a classical example for data imbalance problem. Micro calcifications are very tiny deposits of calcium that appear as small bright spots in the mammogram. Classification of microcalcification clusters from mammograms plays an important role in computer-aided diagnosis for early detection of breast cancer. In this paper, we review in brief the state of the art techniques in the framework of imbalanced data sets and investigate the performance of different methods for microcalcification classification.

References

On the Classification of Imbalanced Datasets

- Chao Chen, Andy Liaw, and Leo Breiman, "Using random forest to learn imbalanced data," www.stat.berkeley.edu/tech-reports/666.pdf
- Son Lam Phung, Abdesselam Bouzerdoum, Giang Hoang Nguyen, "Learning pattern classification tasks with imbalanced data sets," http://ro.uow.edu.au
- Piyasak Jeatrakul, Kok Wai Wong, and Chun Che Fung, "Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm," http://researchrepository.murdoch.edu.au
- Yanmin Sun, Mohamed S Kamel, and Yang Wang, "Boosting for learning multiple classes with imbalanced class distribution," Proceedings of the Sixth International Conference on Data Mining (ICDM&apos;06), 0-7695-2701-9/06 © 2006 IEEE.
- Chao Chen and Mei-Ling Shyu, "Clustering-based binary-class classification for imbalanced data sets," rvc.eng.miami.edu/paper/2011/IRI11_clustering.pdf
On the Classification of Imbalanced Datasets

- Jie Gu, "Random Forest Based Imbalanced Data Cleaning and Classification", http://lamda.nju.edu.cn/conf/pakdd07/dmc07/reports/P251.pdf

- Giang Hoang Nguyen Abdessalem, "Learning pattern classification tasks with imbalanced data sets",
- Nitesh V. Chawla, Nathalie Japkowicz, Aleksander Kołcz, "Editorial: Special issue on learning from imbalanced data sets", Sigkdd Explorations. Volume 6, Issue 1 - Page 1-6, 2004
- Haibo He, and Edwardo A. Garcia, "Learning from Imbalanced Data", IEEETransactions on knowledge and data engineering, Vol. 21, No. 9, September 2009
- Xing-Ming Zhao, Xin Li, Luonan Chen, and Kazuyuki Aihara1, "Protein classification with imbalanced data", 2008; 70:1125–1132. VVC 2007

On the Classification of Imbalanced Datasets

- Rolando R. Hernández-Cisneros and Hugo Terashima, "Evolutionary Neural Networks Applied To The Classification Of Microcalcification Clusters In Digital Mammograms", 2006 IEEE Congress on Evolutionary Computation Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006

Index Terms

Computer Science  Artificial Intelligence

Keywords

Classification  Microcalcification  Imbalanced Data Sets  Mammography