Advances in Cost-sensitive Multiclass and Multilabel Classification

Hsuan-Tien Lin
htlin@csie.ntu.edu.tw
National Taiwan University
Taipei, Taiwan

ABSTRACT

Classification is an important problem for data mining and knowledge discovery and comes with a wide range of applications. Different applications usually evaluate the classification performance with different criteria. The variety of criteria calls for cost-sensitive classification algorithms, which take the specific criterion as input to the learning algorithm and adapt to different criteria more easily. While the cost-sensitive binary classification problem has been relatively well-studied, the cost-sensitive multiclass and multilabel classification problems are harder to solve because of the sophisticated nature of their evaluation criteria. The tutorial aims to review current techniques for solving cost-sensitive multiclass and multilabel classification problems, with the hope of helping more real-world applications enjoy the benefits of cost-sensitive classification.

CCS CONCEPTS

• **Computing methodologies** → **Cost-sensitive learning**.

KEYWORDS

cost-sensitive learning, multiclass classification, multilabel classification

ACM Reference Format:
Hsuan-Tien Lin. 2019. Advances in Cost-sensitive Multiclass and Multilabel Classification. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3292500.3332283

1 INTRODUCTION

Classification is an important machine learning problem for data mining and knowledge discovery. Traditionally, the regular classification problem aims at minimizing the error rate of mis-prediction. Nevertheless, the error rate may not be the best optimization criterion for some practical applications. For instance, consider asking a doctor to check a patient and predict her/his health status as {`H7N9`-infected, `cold`-infected, `healthy`} [15]. In the following table, we can see the different costs that the society needs to pay in the nine different scenarios.

diagnosis / actual status	`H7N9`	`cold`	`healthy`
`H7N9`	0	10000	1000000
`cold`	100	0	3000
`healthy`	100	30	0

The rows represent the actual patient status, and the columns represent the diagnosis made by the doctor. For instance, on any correct diagnosis, the society pays no (additional) cost. Somehow if an H7N9-infected patient is predicted as cold-infected or healthy, the whole society may suffer from a huge amount of cost. On the other hand, if a cold-infected patient is predicted as healthy, the society needs to pay some cost—but not as serious as the ones paid in the previous scenario. These different costs are important for a human doctor when making any diagnosis. For instance, the doctor would be very careful on the slightest H7N9 symptom to prevent the "1000000" level mis-prediction.

If we were to build an automatic system—a "computer doctor assistant"—to help with the diagnosis, how can the system use the cost information appropriately? Such a cost-sensitive classification problem can be very different from the regular classification one, and is needed in applications like targeted marketing, information retrieval, medical decision making, object recognition and intrusion detection.

Cost-sensitive binary classification problem has been studied since the 90s, resulting in sampling and re-weighting tools that continue to influence many real-world applications [6, 17]. In the past 20 years, researchers have advanced those tools to tackle more complicated problems, including multiclass and multilabel classification ones. The tutorial aims to review and summarize those advances to allow more real-world applications to enjoy the benefits of cost-sensitive classification. The advances range from the Bayesian approaches that consider costs during inference, to reduction-based approaches that transform the cost-sensitive classification task to other tasks, to deep learning approaches that plug the costs into the optimization and feature-extraction process. We discuss the relationship between the approaches as well as their practical usage. We will also introduce some success in data mining applications, such as improving the performance of a real-world bacteria classification system and tackling the class-imbalance problem of KDDCup 1999.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
KDD '19, August 4–8, 2019, Anchorage, AK, USA
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332283

2 TUTORIAL OUTLINES

If time allows, the tutorial plans to cover the following works on cost-sensitive multiclass and multilabel classification.

2.1 Cost-sensitive Multiclass Classification

- motivation of cost-sensitive multiclass classification
- Bayesian perspective: Bayes-optimal decision making and the MetaCost algorithm [5]
- reduction to binary classification: one-versus-one decomposition [1, 13] and tree decomposition [2]
- reduction to regression: SVM [15] and deep learning [3] algorithms

2.2 Cost-sensitive Multilabel Classification

- motivation of cost-sensitive multilabel classification
- Bayesian perspective: the PCC algorithm with beam search [4, 11] and its connection to MetaCost [5]
- reduction to multiclass classification: the PRAKEL algorithm [16] and its ancestor [14]
- reduction to binary classification: the CFT algorithm [12] and its connection to tree decomposition [2]
- deep learning: the LLSL algorithm [7]

2.3 Applications of Cost-sensitive Classification

- multiclass: bacteria classification for bioinformatics [8] with error-and-cost-sensitive classification [9]
- multiclass: intrusion classification for security in KDDCup 1999 [9] and the imbalanced learning problems [10]
- multilabel: tag classification for social network analysis [14]

3 PRESENTER INFORMATION

Professor Hsuan-Tien Lin received his Ph.D. in Computer Science from California Institute of Technology in 2008. He joined the Department of Computer Science and Information Engineering at National Taiwan University as an assistant professor in August 2008, and has been a professor since August 2017. He is currently also the Chief Data Science Consultant of Appier, a startup company that specializes in making AI easier in business domains.

From the university, Prof. Lin received the Distinguished Teaching Award in 2011 and three Outstanding Teaching Awards between 2016 and 2018. He co-authored a top-selling machine learning textbook “Learning from Data” and offered two popular Mandarin-teaching MOOCs based on the textbook. He co-led the teams that won six KDDCup world champions consecutively between 2010 and 2013.

ACKNOWLEDGMENTS

The author thanks current and past members of the computational learning laboratory at National Taiwan University for their many works on cost-sensitive classification. This tutorial presentation is partially supported by the Ministry of Science and Technology of Taiwan under MOST 107-2628-E-002-008-MY3 and National Taiwan University under NTU-CDP-108L7738.

REFERENCES

[1] Alina Beygelzimer, Varsha Dani, Tom Hayes, John Langford, and Bianca Zadrozny. 2005. Error-Limiting Reductions Between Classification Tasks. In *Proceedings of the International Conference on Machine Learning (ICML)*. 49–56.

[2] Alina Beygelzimer, John Langford, and Pradeep Ravikumar. 2009. Error-Correcting Tournaments. In *Proceedings of the Conference on Algorithmic Learning Theory (ALT)*. 247–262.

[3] Yu-An Chung, Hsuan-Tien Lin, and Shao-Wen Yang. 2016. Cost-aware Pre-training for Multiclass Cost-sensitive Deep Learning. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. 1411–1417.

[4] Krzysztof Dembczyński, Weiwei Cheng, and Eyke Hüllermeier. 2010. Bayes Optimal Multilabel Classification via Probabilistic Classifier Chains. In *Proceedings of the International Conference on Machine Learning (ICML)*. 279–286.

[5] Pedro Domingos. 1999. MetaCost: A General Method for Making Classifiers Cost-Sensitive. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 155–164.

[6] Charles Elkan. 2001. The Foundations of Cost-Sensitive Learning. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*. 973–978.

[7] Cheng-Yu Hsieh, Yi-An Lin, and Hsuan-Tien Lin. 2018. A Deep Model with Local Surrogate Loss for General Cost-sensitive Multi-label Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

[8] Te-Kang Jan, Hsuan-Tien Lin, Hsin-Pai Chen, Tsung-Chen Chern, Chung-Yueh Huang, Bing-Cheng Wen, Chia-Wen Chung, Yung-Jui Li, Ya-Ching Chuang, Li-Li Li, Yu-Jiun Chan, Juen-Kai Wang, Yuh-Lin Wang, Chi-Hung Lin, and Da-Wei Wang. 2011. Cost-sensitive Classification on Pathogen Species of Bacterial Meningitis by Surface Enhanced Raman Scattering. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 390–393.

[9] Te-Kang Jan, Da-Wei Wang, Chi-Hung Lin, and Hsuan-Tien Lin. 2012. A Simple Methodology of Soft Cost-sensitive Classification. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 141–149.

[10] Salman H. Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A. Sohel, and Roberto Togneri. 2018. Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data. *IEEE Transactions on Neural Networks and Learning Systems* 29, 8 (2018), 3573–3587.

[11] Abhishek Kumar, Shankar Vembu, Aditya Krishna Menon, and Charles Elkan. 2013. Beam search algorithms for multilabel learning. *Machine Learning* 92, 1 (2013), 65–89.

[12] Chun-Liang Li and Hsuan-Tien Lin. 2014. Condensed Filter Tree for Cost-Sensitive Multi-Label Classification. In *Proceedings of the International Conference on Machine Learning (ICML)*. 423–431.

[13] Hsuan-Tien Lin. 2014. Reduction from Cost-Sensitive Multiclass Classification to One-versus-One Binary Classification. In *Proceedings of the Asian Conference on Machine Learning (ACML) (JMLR Workshop and Conference Proceedings)*, Vol. 39. 371–386.

[14] Hung-Yi Lo, Shou-De Lin, and Hsin-Min Wang. 2014. Generalized k-Labelsets Ensemble for Multi-Label and Cost-Sensitive Classification. *IEEE Transactions on Knowledge and Data Engineering* 26, 7 (2014), 1679–1691.

[15] Han-Hsing Tu and Hsuan-Tien Lin. 2010. One-sided Support Vector Regression for Multiclass Cost-sensitive Classification. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*. 1095–1102.

[16] Yu-Ping Wu and Hsuan-Tien Lin. 2017. Progressive k-Labelsets for Cost-Sensitive Multi-Label Classification. *Machine Learning* 106, 5 (May 2017), 671–694.

[17] Bianca Zadrozny, John Langford, and Naoki Abe. 2003. Cost-Sensitive Learning by Cost-Proportionate Example Weighting. In *Proceedings of the International Conference on Data Mining (ICDM)*. 435–442.