Cost-sensitive Classification: Status and Beyond

Hsuan-Tien Lin
Department of Computer Science and Information Engineering
National Taiwan University, Taipei 106, Taiwan

I. Introduction

Consider asking a doctor to check a patient and predict her/his health status as {`H1N1`-infected, `cold`-infected, `healthy`} [1]. In the following table, we can see the different costs that the society needs to pay in the nine different scenarios.

actual status \ diagnosis	H1N1	cold	healthy
H1N1	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. However, if an H1N1-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 H1N1 symptom to prevent the "1000000" level mis-prediction.

If we were to build an automatic system—a "computer doctor"—to make the diagnosis, how can the system use the cost information appropriately? Many real-world applications that share similar needs can be found in medical decision making, target marketing, and object recognition. Those applications belong to *cost-sensitive classification*. In fact, cost-sensitive classification can be used to express any finite-choice and bounded-loss machine learning problems [2]. Thus, it has been attracting much research attention in the past decade [3], [4], [5], [6], [7], [8], [2], [9], [10], [11], [12], [1], [13].

II. Status

Next, we discuss the state-of-the-art in cost-sensitive classification. We can further separate the problem to two cases: the binary case (when there are only two possible classes) and the multiclass one.

A. Binary Cost-sensitive Classification

The binary cost-sensitive classification problem considers only two kinds of costs: mis-predicting the first class as the second; mis-predicting the second class as the first. Thus, the problem is relatively simpler. In particular, it has been studied in detail from the theoretical perspective, and has reached a satisfactory performance from the empirical perspective. [4] was the first to lay down the theoretical foundation of binary cost-sensitive classification. More specifically, [4] showed that every such problem can be reduced to a cost-less classification one by the technique of re-weighting the importance of each example. [5] not only theoretically extended the approach in [4] to a more general setting, but also empirically applied the approach to make common binary classification algorithms cost-sensitive. Thus, in terms of both the theory and the practice, the case can be considered solved.

B. Multiclass Cost-sensitive Classification

Multiclass cost-sensitive classification can be much more difficult than the binary one. As pointed out in [8], the sound theoretical foundation in [4] cannot be directly extended to the multiclass case. Thus, it is challenging to design good algorithms from the theoretical perspective, and early approaches that deal with multiclass cost-sensitive classification are more-or-less heuristic.

For instance, [3] proposed the MetaCost algorithm, which was the first practical multiclass cost-sensitive classification algorithm. MetaCost can make any cost-less classification algorithm cost-sensitive. In particular, it takes the cost-less algorithm as both a pre-processor and a learner. The pre-processing step would re-label the training examples to fit the needs of cost-sensitive classification, and the learning step would then extract information from the re-labeled examples. Nevertheless, the re-labeling step is based on an ideal (possibly non-realistic) probabilistic assumption, which makes it hard to rigorously analyze the performance of the whole algorithm. The shortcoming restricts further development of practical tools based on MetaCost. Many other early approaches [14] suffer from similar shortcomings.

To design algorithms with stronger theoretical guarantee, many modern cost-sensitive classification approaches are reduction-based, just like the work of [5] for binary cost-sensitive classification. That is, they try to solve the cost-sensitive classification problem by transforming it to other known types of machine learning problems. Such an approach can bring in some immediate benefits. First, well-tuned existing learning algorithms can be readily transformed into good cost-sensitive classification ones, which saves immense efforts in design and implementation. Second, new learnability results for cost-sensitive classification can be easily derived from known ones for other problems, which saves tremendous efforts in theoretical analysis.

The work of [6] is a precursor of the reduction kind. They reduce multiclass cost-sensitive classification to multiclass cost-less classification. Nevertheless, the proposed boosting-based algorithm only applies to a limited range of cost-less classification algorithms, and thus its success is quite limited.

[8] analyzed the re-weighting approach from another angle. They showed that the weights for re-weighting can be systematically obtained by solving some linear equations, but the equations may not be solvable for some (many) cost-sensitive classification problems. Their work demonstrated the theoretical difficulty of reducing multiclass cost-sensitive classification to cost-less classification. [10] gave yet another theoretical analysis along the same direction of [8], and showed that reduction to cost-less classification cannot be done without a re-labeling step, which inevitably introduces noise that deteriorate the learning process. The work justifies the practical difficulty when directly reducing multiclass cost-sensitive classification to cost-less classification.

Another plausible route is to reduce multiclass cost-sensitive classification to binary cost-sensitive classification. The work of [9] casted the problem to a tournament design game, which resulted in a tree-structured decomposition of the problem. [2] and [11], on the other hand, formed the reduction with pairwise comparisons. [7] took a different way of reduction, and introduced randomness in forming the transformed binary classification problems. The three reductions are similarly promising in theory, but except for the ones in [2], [11], the empirical advantages and disadvantages are yet to be validated.

Yet another plausible route is to predict multiclass cost-sensitive classification to regression [12], [1], [13]. In the work of [1], there is a serious empirical comparison of regression-based reduction to classification-based ones based on some limited cost-sensitive classification settings. The empirical results demonstrate that regression-based reduction is quite promising.

III. Beyond

After a decade of studying cost-sensitive classification [3], what are the central questions now in this sub-field? Below I list some of my personal thoughts.

A. Where Does Cost Come From?

Cost-sensitive classification starts from the assumption that the cost is given based on the application needs. Although such an assumption offers a flexibility to the users of cost-sensitive classification algorithms, the truth is that the users may not be comfortable when assigning the costs. For instance, it can be a hard assertion to say “This cost component be 9 times larger than the other component rather than 9.1 times larger.”

One possibility is to view the costs as knobs one can turn in an equalizer. While such a tuning view is more-or-less standard in binary classification (see, e.g. LIBSVM [15] with the default -w option), the corresponding view is yet to be established for multiclass cost-sensitive classification. Also, for specific applications, there is a need to studying guidelines that assist users in turning the knobs.

The other possibility is to link the costs to the true loss function of interest in specific applications. The linking is not always an easy task but we started to get some limited success in the Yahoo! Learning to Rank Challenge 2010 [16].

Yet another possibility is to lessen the burden of the users by considering cost intervals instead of cost values. This is an interesting ongoing research direction [17].

B. What Are the Benchmarks?

Because of the difficulty in assigning costs, one should note that there are very few true publicly-available cost-sensitive classification data sets. KDDCup 1999 offers one such data set, but its cost matrix is not far from the cost-less one and is artificially given rather than coming from any true meanings. Most of the existing work on cost-sensitive classification (e.g. see [1]) focuses on a traditional benchmark [3] that uses publicly-available data sets while manually assigning the cost based on the size of each class along with some randomness.

The lack of publicly-available data set makes it hard to compare cost-sensitive classification algorithms based on the true needs. In fact, as can be seen from the previous section, there are not many comparative works for cost-sensitive classification yet. Seeking for true cost-sensitive classification data sets can be important for continuing research in cost-sensitive classification.

C. General Cost-sensitive or Not?

As mentioned, cost-sensitive classification can be used to express any finite-choice and bounded-loss machine learning problems [2]. Nevertheless, when encountering one such problem, should we use a general existing cost-sensitive classification algorithm or design specific algorithms to directly tackle the problem?

[10] showed that for the ordinal ranking problem, in which the classes are ordered ranks and the cost vector is V-shaped around the desired rank, general cost-sensitive classification algorithms achieve reasonable performance. Nevertheless, better specific algorithms *can* be designed by a deeper study of the problem.

Does this mean that cost-sensitive classification is not important? My humble opinion is no. Cost-sensitive classification offers an immediate entry point for simple and quick solutions (if mature algorithmic tools exist). Such an entry point can be important for tackling new machine learning problems that arise daily in the application domains. Then, after accumulating some understanding of the new problem from the general cost-sensitive perspective, one would be ready to further improve the performance from a more specific perspective.

IV. Summary

In this position paper, I listed some state-of-the-art tools on cost-sensitive classification, and discussed some continuing challenges. It is quite exciting to work on this general machine learning problem. I would hope to encourage the community to try some of those tools if there are application needs that can be met by cost-sensitive classification algorithms.

References

[1] H. Tu and H.-T. Lin, "One-sided support vector regression for multiclass cost-sensitive classification," in *Proceedings of ICML '10*, 2010.

[2] A. Beygelzimer, V. Daniand, T. Hayes, J. Langford, and B. Zadrozny, "Error limiting reductions between classification tasks," in *Machine Learning: Proceedings of the 22rd International Conference*, L. D. Raedt and S. Wrobel, Eds. ACM, 2005, pp. 49–56.

[3] P. Domingos, "MetaCost: A general method for making classifiers cost-sensitive," in *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM SIGKDD. ACM, 1999, pp. 155–164.

[4] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, B. Nebel, Ed. Morgan Kaufmann, 2001.

[5] B. Zadrozny, J. Langford, and N. Abe, "Cost sensitive learning by cost-proportionate example weighting," in *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003)*. IEEE Computer Society, 2003.

[6] N. Abe, B. Zadrozny, and J. Langford, "An iterative method for multi-class cost-sensitive learning," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, Eds. ACM, 2004, pp. 3–11.

[7] J. Langford and A. Beygelzimer, "Sensitive error correcting output codes," in *Learning Theory: 18th Annual Conference on Learning Theory*, ser. Lecture Notes in Artificial Intelligence, P. Auer and R. Meir, Eds., vol. 3559. Springer-Verlag, 2005, pp. 158–172.

[8] Z.-H. Zhou and X.-Y. Liu, "On multi-class cost-sensitive learning," in *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI-06*, AAAI. AAAI Press, 2006.

[9] A. Beygelzimer, J. Langford, and P. Ravikumar, "Multiclass classification with filter trees," 2007, downloaded from http://hunch.net/~jl.

[10] H.-T. Lin, "From ordinal ranking to binary classification," Ph.D. dissertation, California Institute of Technology, 2008.

[11] ——, "A simple cost-sensitive classification algorithm using one-versus-one comparisons," National Taiwan University, Tech. Rep., 2010.

[12] H. Tu, "Regression approaches for multi-class cost-sensitive classification," Master's thesis, National Taiwan University, 2009.

[13] H. Tu and H.-T. Lin, "Regression approaches for multi-class cost-sensitive classification," National Taiwan University, Tech. Rep., 2010.

[14] D. D. Margineantu, "Methods for cost-sensitive learning," Ph.D. dissertation, Oregon State University, 2001.

[15] C.-C. Chang and C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, National Taiwan University, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] M.-F. Tsai, S.-T. Chen, Y.-N. Chen, C.-S. Ferng, C.-H. Wang, T.-Y. Wen, and H.-T. Lin, "An ensemble ranking solution to the yahoo! learning to rank challenge," National Taiwan University, Tech. Rep., 2010.

[17] X.-Y. Liu and Z.-H. Zhou, "Learning with cost intervals," in *Proceedings of KDD 2010*, 2010.