

Multi-target Classification: Methodology and Practical Case Studies

Mark Last 

Department of Information Systems Engineering, Ben-Gurion University of the Negev,
84105 Beer-Sheva, Israel
mlast@bgu.ac.il

Abstract. Most classification algorithms are aimed at predicting the value or values of a single target (class) attribute. However, some real-world classification tasks involve several targets that need to be predicted simultaneously. The Multi-objective Info-Fuzzy Network (M-IFN) algorithm builds an ordered (oblivious) decision-tree model for a multi-target classification task. After summarizing the principles and the properties of the M-IFN algorithm, this paper reviews three case studies of applying M-IFN to practical problems in industry and science.

Keywords: Multi-target classification · Multi-objective info-fuzzy networks · Information theory · Decision trees

1 Introduction

As indicated in [1, 2], the assumption that a learning task has only one objective is very restrictive. Data instances in many real-world datasets may be simultaneously assigned multiple class labels related to multiple tasks, which may be strongly related, completely unrelated, or just weakly related to each other. Examples include a student's grades in several courses, multiple diagnoses of a given patient, multiple outputs of a software system, multi-trait prediction from genomic data [3], etc. More examples of concurrent learning tasks are discussed in [4].

In [5], we have presented a unified framework for single-objective and multi-objective classification, called an *extended classification task*, which includes the following components:

- $R = (A_1, \dots, A_n)$ - a non-empty set of n candidate input features ($n \geq 1$), where A_i is an attribute i . The values of these attributes (features) can be used to predict the values of class dimensions (see below).
- $O = (C_1, \dots, C_m)$ - a non-empty set of m class dimensions ($m \geq 1$). This is a set of tasks (targets) to predict. The *extended classification task* is to build an accurate model (or models) for predicting the values of all class dimensions, based on the corresponding *dependency subset* (or subsets) $I \subseteq R$ of selected input features. A special case of this task is *multi-label classification*, which allows multiple labels of the same dimension to be assigned to a given instance.

Section 2 of this paper outlines the methodology for inducing a multi-target model called Multi-objective Info-Fuzzy Network (M-IFN) and discusses its main characteristics. Section 3 reviews several case studies of applying M-IFN to practical problems in diverse branches of industry and science. Finally, in Sect. 4, we briefly discuss open challenges in multi-target classification.

2 Multi-objective Info-Fuzzy Networks

As indicated in [5], a multi-objective info-fuzzy network (M-IFN) is a multi-target extension of a single-objective info-fuzzy network (IFN). Similar to IFN, M-IFN has a single root node and an “oblivious read-once decision graph” structure, where all nodes of a given layer are labeled by the same feature and each feature is tested at most once along any path. It also has a target layer with a target node for each class label of every target. Every internal M-IFN node is shared among all targets, which makes it an extreme case of a Shared Binary Decision Diagram [6]. This implies that each terminal (leaf) node is connected to at least one target node associated with a value of every target.

Unlike CART [7], C4.5 [8], and EODG [9], the M-IFN construction algorithm has only the growing (top-down) phase, which iteratively chooses predictive features maximizing the decrease in the total conditional entropy of all targets. The top-down construction is pre-pruned by the Likelihood-Ratio Test. The details of M-IFN construction procedure are presented in [5].

In [10], we show the M-IFN algorithm to have the following important properties:

- The average conditional entropy of m targets in an n -input m -dimensional M-IFN model is not greater than the average conditional entropy over m single-target models S_i ($i = 1, \dots, m$) based on the same n input features. This inequality is strengthened if the multi-target M-IFN model is built upon more features than the single-target models. Consequently, we may expect that the *average accuracy* of a multi-target M-IFN model in predicting the values of m targets will not be lower, or even will be higher, than the average accuracy of m single-target models using the same set of predictive features.
- If all class dimensions (targets) are either mutually independent or totally dependent on each other, the input features selected by the M-IFN algorithm will minimize the joint conditional entropy of all targets, i.e. will provide the most accurate classification model for all target classes. The case of mutual independence extends the scope of multitask (transfer) learning [1], where all “extra” tasks (targets) are assumed to be related to the main classification target.

3 Case Studies

Our first case study [11] refers to prediction of grape and wine quality in a multi-year dataset provided by Yarden - Golan Heights Winery in Israel. For each grape field in every season, the Winery keeps record of 27 quality parameters (target variables) along with 135 candidate input features. Thus predicting grape and wine quality is clearly a multi-target

classification task. We have used M-IFN to identify the most significant predictive factors of grape and wine quality parameters. We have also shown that on average, single-target IFN models are significantly more accurate on this data than C4.5 decision-tree models whereas the M-IFN models are even more accurate than the single-target IFN models. This result agrees with the previously mentioned observation that the average accuracy of a single multi-target M-IFN model is not expected to be worse than the average accuracy of multiple single-target models using the same set of predictive features.

The second case study [12], partially supported by General Motors, deals with predicting the probability and the timing of vehicle failures based on an integrated database of sensor measurements and warranty claims. We have applied the IFN and M-IFN induction algorithms to a dataset of 46,418 records representing periodical battery sensory readings for 21,814 distinct vehicles of a high-end model. The prediction models have been evaluated by the area under ROC (Receiver Operating Characteristics) curves, also known as the Area under Curve, or AUC. Though the IFN and the M-IFN ROC curves for the target attribute *Battery Failure* are nearly identical, the multi-target approach has shown a clear advantage in terms of model comprehensibility as it reduced the total number of prediction rules by 33 %.

The third, more recent case study [13] is aimed at predicting the number and the maximum magnitude of seismic events in the next year based on the seismic events recorded in the same region during the previous years. The predictive features include six seismic indicators commonly used in earthquake prediction literature as well as 20 new features based on the moving annual averages of the number of earthquakes. We have evaluated eight classification algorithms on a catalog of 9,042 earthquake events, which took place between 01/01/1983 and 31/12/2010 in 33 seismic regions of Israel and its neighboring countries. The M-IFN algorithm has clearly shown the best result in terms of the Area under Curve (AUC) criterion, explained by its unique capability to take into account the relationship between two target variables: the total number of earthquakes and the maximum earthquake magnitude during the same year.

4 Conclusions

In this paper, we have presented the M-IFN (Multi-objective Info-Fuzzy Network) algorithm for inducing multi-target classification models. The algorithm's effectiveness and broad applicability have been demonstrated via case studies in three diverse fields: winemaking, predictive maintenance, and seismology. The multi-target classification domain is facing a number of exciting challenges such as semi-supervised learning from a subset of targets, handling delayed target values, and adapting deep learning algorithms for the multi-target classification task.

References

1. Caruana, R.: Multitask learning. *Mach. Learn.* **28**, 41–75 (1997)
2. Suzuki, E., Gotoh, M., Choki, Y.: Bloomy decision tree for multi-objective classification. In: PKDD 2001 (2001)

3. He, D., Kuhn, D., Parida, L.: Novel applications of multitask learning and multiple output regression to multiple genetic trait prediction. *Bioinformatics* **32**(12), i37–i43 (2016)
4. Caruana, R.: Multitask learning: a knowledge-based source of inductive bias. In: Proceedings of the 10th International Conference on Machine Learning, ML 1993, University of Massachusetts, Amherst (1993)
5. Last, M.: Multi-objective classification with info-fuzzy networks. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 239–249. Springer, Heidelberg (2004)
6. Bryant, R.E.: Graph-based algorithms for boolean function manipulation. *IEEE Trans. Comput.* **C-35-8**, 677–691 (1986)
7. Breiman, L., Friedman, J., Olshen, R., Stone, P.: *Classification and Regression Trees*. Wadsworth, Belmont (1984)
8. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco (1993)
9. Kohavi, R., Li, C.-H.: Oblivious decision trees, graphs, and top-down pruning. In: Proceedings of International Joint Conference on Artificial Intelligence (IJCAI) (1995)
10. Last, M., Friedman, M.: Black-box testing with info-fuzzy networks. In: *Artificial Intelligence Methods in Software Testing*, pp. 21–50. World Scientific, Singapore (2004)
11. Last, M., Elnekave, S., Naor, A., Shonfeld, V.: Predicting wine quality from agricultural data with single-objective and multi-objective data mining algorithms. In: *Recent Advances on Mining of Enterprise Data: Algorithms and Applications*, pp. 323–365. World Scientific, Singapore (2007)
12. Last, M., Sinaiski, A., Subramania, H.S.: Condition-based maintenance with multi-target classification models. *New Gener. Comput.* **29**(3), 245–260 (2011). Special Issue on Hybrid and Ensemble Methods in Machine Learning
13. Last, M., Rabinowitz, N., Leonard, G.: Predicting the maximum earthquake magnitude from seismic data in Israel and its neighboring countries. *PLoS ONE* **11**(1), e0146101 (2016)