Drawing Parallels between Multi-Label Classification and Multi-Target Regression

Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, and Ioannis Vlahavas

Department of Informatics, Aristotle University of Thessaloniki,
54124 Thessaloniki, Greece
{greg,espyromi,vlahavas}@csd.auth.gr

Abstract. Learning from multi-label data has recently received increased attention by researchers working on machine learning and data mining for two main reasons. The first one is the ubiquitous presence of multi-label data in application domains ranging from multimedia information retrieval to tag recommendation, query categorization, gene function prediction, medical diagnosis, drug discovery and marketing. The other reason is a number of challenging research problems involved in multi-label learning, such as dealing with label rarity, scaling to large number of labels and exploiting label relationships (e.g. hierarchies), with the most prominent one being the explicit modelling of label dependencies.

Multi-label learning is closely related to multivariate regression, also known as multi-output or multi-target regression, which aims at predicting multiple real-valued target variables instead of binary ones. Despite that multi-target regression is a less popular task, it still arises in several interesting domains, such as predicting the wind noise of vehicle components, stock price prediction, ecological modelling and more recently energy-related forecasting, such as wind and solar energy production forecasting and load/price forecasting.

Multi-label learning is often treated as a special case of multi-target regression in statistics. However, we could more precisely state that both are instances of the more general learning task of predicting multiple targets, which could be real-valued, binary, ordinal, categorical or even of mixed type. The baseline approach of learning a separate model for each target applies to both learning tasks. Most importantly, they share the same core challenge of modelling dependencies among the different targets.

In this talk we will emphasize this tight relationship between multi-label classification and multi-target regression by discussing their similarities, while also highlighting their differences. We will further discuss techniques that can inherently handle both tasks. Motivated from their tight connection, we will discuss pathways for transferring recent advances in the more popular multi-label learning task to multi-target regression and will give examples of our recent work in this direction[1][2]. Finally, we will present our recent extension of the Mulan library[3] with methods for multi-target regression.

[1] http://arxiv.org/abs/1211.6581
[2] http://arxiv.org/abs/1404.5065
[3] http://mulan.sourceforge.net