Information Sciences 415–416 (2017) 53–69

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Multi-target support vector regression via correlation regressor chains

Gabriella Melki [a], Alberto Cano [a,*], Vojislav Kecman [a], Sebastián Ventura [b,c]

[a] Department of Computer Science, Virginia Commonwealth University, USA
[b] Department of Computer Science and Numerical Analysis, University of Cordoba, Spain
[c] Department of Computer Science, King Abdulaziz University, Saudi Arabia Kingdom

A R T I C L E I N F O

Article history:
Received 10 August 2016
Revised 18 January 2017
Accepted 18 June 2017
Available online 19 June 2017

Keywords:
Multi-target regression
Multi-output regression
Regressor chains
Support vector regressor

A B S T R A C T

Multi-target regression is a challenging task that consists of creating predictive models for problems with multiple continuous target outputs. Despite the increasing attention on multi-label classification, there are fewer studies concerning multi-target (MT) regression. The current leading MT models are based on ensembles of regressor chains, where random, differently ordered chains of the target variables are created and used to build separate regression models, using the previous target predictions in the chain. The challenges of building MT models stem from trying to capture and exploit possible correlations among the target variables during training. This paper presents three multi-target support vector regression models. The first involves building independent, single-target Support Vector Regression (SVR) models for each output variable. The second builds an ensemble of random chains using the first method as a base model. The third calculates the targets' correlations and forms a maximum correlation chain, which is used to build a single chained support vector regression model, improving the models' prediction performance while reducing the computational complexity. The experimental study evaluates and compares the performance of the three approaches with seven other state-of-the-art multi-target regressors on 24 multi-target datasets. The experimental results are then analyzed using non-parametric statistical tests. The results show that the maximum correlation SVR approach improves the performance of using ensembles of random chains.

1. Introduction

In supervised learning, *single-target* (ST) models are trained to predict the value of a single, categorical or numeric, target attribute of a given example. In some cases, more than one target, or output, can be associated with a single sample input. These situations are handled by a generalization of ST learning, which involves predicting these multiple outputs concurrently, and is known as multi-target (MT) learning [1,5]. Specifically, MT learning includes *multi-target regression* (MTR), which addresses the prediction of continuous targets, *multi-label classification* [48] which focuses on binary targets, and *multi-dimensional classification* which describes the prediction of discrete targets [5,38].

* Corresponding author.
E-mail address: acano@vcu.edu (A. Cano).

http://dx.doi.org/10.1016/j.ins.2017.06.017
0020-0255/

Multi-target prediction has the capacity to generate models representing a wide variety of real-world applications, ranging from natural language processing [25] to bioinformatics [34]. Other application areas include ecology [1], gene function prediction [27], predicting the quality of vegetation [22,28], stock price index forecasting [46], and operations research [5,23].

Constructing models for these types of real-world problems presents many challenges, such as missing data, (due to targets not being observed or recorded), and noisy data (due to instrument, experimental or human error), and the curse of high dimensionality. Along with these challenges, the most difficult task is identifying relationships between the input data and its corresponding output value. In the context of multi-target modeling, multiple outputs must now be trained against, which inherently adds computational complexity. The targets may or may not be correlated, and the corresponding model must accommodate for both scenarios. However, a characteristic of the MT datasets used in these applications and elsewhere, is that they are generated by a single system, most likely indicating that the nature of the outputs captured has some structure [23]. Even though modeling the multi-variate nature and complex relationships between the target variables is challenging [5], they are more accurately represented by an MT model.

Several base-line methods have been proposed for solving multi-target tasks such as Multi-Objective Random Forests [28], Boosted Neural Networks [22], Ensembles of Trees [30], and many others. Support Vector Machines are a popular set of linear and non-linear supervised machine learning algorithms with a strong theoretical basis on Vapnik-Chervonenkis theory [13]. It has previously been shown that they outperform most algorithms in terms of performance, scalability, and the ability to efficiently deal with outliers [16,26]. Input space dimensionality does not have an adverse effect on the model training time, and furthermore, the final model produced is sparse, allowing for quick predictions.

There are two main approaches for using such base-line methods in the context of MT learning. The first being *problem transformation* methods, or *local* methods, in which the multi-target problem is transformed into multiple single-target problems, each solved separately using classical methods, as described above. The second being *algorithm adaptation* methods, or *global*, or *big-bang* methods, that adapt existing single-target methods to predict all the target variables simultaneously [5,27]. Using *problem transformation* algorithms for a domain of t target variables, t predictive models must be constructed, each predicting a single-target variable [27]. Prediction for an unseen sample would be obtained by running each of the t single-target models and concatenating their results. Conversely, when using *algorithm adaptation* algorithms for the same domain of t target variables, only one model would need to be constructed which would output all t predictions.

Literature shows that *algorithm adaptation* methods perform better than *problem transformation* methods [27,39]. The most valuable advantage of using multi-target techniques is that, not only are the relationships between the sample variables and the targets exploited, but the relationships between the targets amongst themselves are as well [3,9]. Single-target techniques, on the other hand, eliminate any possibility of learning from the possible relationships between the target variables because a single, independent model is trained for each target separately [4]. Another advantage of MT techniques is model interpretability [1,46]. A single multi-target model is highly more interpretable than a series of single-target models because it not only exploits the relationship between the data and targets, but also the targets amongst themselves. Not only is a single MT model more interpretable, but it could also be considerably more computationally efficient to train, rather than training multiple single-target models individually [2].

This paper proposes three novel approaches to solving multi-target regression problems. The objective of this research topic is to investigate the performance changes when building a regression model using two distinct *algorithm adaptation* chaining methods versus building independent single-target models for each target variable using a novel framework. The main contributions presented in this paper include:

- Evaluating the performance of a Support Vector Regressor (SVR) as a multi-target to single-target *problem transformation* method to determine whether it outperforms current state-of-the-art ST algorithms. We analyze its performance as a base-line model for MT chaining methods due to the fact that ST methods do not account for any correlation among the target variables.
- Building an MT ensemble of randomly chained SVR models (SVRRC), an *algorithm adaptation* approach, inspired by the state-of-the-art chaining classification method, Ensemble of Random Chains Corrected (ERCC) [46], to investigate the effects and advantages of exploiting correlations among target variables during model training, in the context of regression problems. The main issues to be investigated with this approach are the *randomness* of the created chains because they might not capture of correlations between the targets, as well as the time taken to build all the regressors in the ensemble.
- Proposing an MT *algorithm adaptation* model of SVRs that builds a unique chain, capturing the maximum correlation among target outputs, named SVR Correlation Chains (SVRCC). The advantages of using this *maximum correlation chain* approach include exploiting the correlations among the targets which leads to an improvement in model prediction performance, and a reduction in computational complexity because a single SVR-chain model is trained, rather than building an ensemble of 10 base regressors.

The experimental study evaluates and compares the performance of the three approaches, together with 7 other state-of-the-art multi-target regressors, on a set of 24 datasets with varied input size, dimensionality, and output targets. The results of the experiments are analyzed using non-parametric statistical analysis, namely the Bonferroni-Dunn, Holm, and Wilcoxon tests [19]. These post-hoc tests involve multiple comparisons among the algorithms, where they show significant differences in model performances across all datasets. The statistical analysis of the experiments presented in this paper shows the increase in performance of the support vector regressors, specifically, the maximum correlation chain, SVRCC.