Multi-label classification via Error-Correcting Output Codes

Miguel Ángel Bautista[1,2], Oriol Pujol[1,2], and Sergio Escalera[1,2]

Computer Vision Center[1], University of Barcelona- Applied Math and Analysis Dept. [2]
E-mail: mbautista@ub.edu, opujol@ub.edu, sescalera@ub.edu

In general, the goal of multi-class classification problems is to find a function $f : X \rightarrow Y$, where X is the feature space and Y the label set (note $|Y| > 2$). Then every point $x \in X$ is assigned to a label $y \in Y$. In this kind of problems the set of labels Y is mutually exclusive. In this sense, any point x can only be assigned to one label y. Nevertheless, in the multi-label setting each point $x \in X$ can be assigned to l labels in Y, where $1 \leq l \leq (|Y| - 1)$. In this sense, the goal is to find a function f so that it minimizes the miss-classification cost. In the multi-label case, the cost if often defined as the hamming loss of the set of predicted labels $y^{'}$ with the true labels y of x, $HL_x(y, y') = \sum_{i=1}^{|Y|} I(y_i^{'}, y_i)$.

There are plenty of works in the field of multi-label classification [1, 2]. In [1] the authors propose and ensemble method to treat multi-label problems via decoupling the original problem in several multi-classification problems and combining the predictions. In addition, the authors in [1] model each class and then provide a multi-label output by learning a threshold on membership probability to each label in $|Y|$. Recently, in [2] the authors propose a method that is able to capture the relationships between labels by modelling the correlation among them. The main contribution of this works is to capture the relationships between the labels in Y in order to improve the final prediction.

Nevertheless, up to our knowledge, none of the previous works have tried to exploit the underlying hierarchical structure of labels, and thus, minimizing the risk of miss-classification. We propose to use the Error-Correcting Output Codes (ECOC) [3] framework to treat multi-label classification problems. In this sense, since problem-dependent coding strategies have shown good performance on literature [4], we propose to embed the hierarchical label structure inherent to any multi-label problem into the ECOC coding matrix, and thus, minimizing the final hamming loss on the label set of each point $x \in X$. An example of this procedure is shown in Figure 1. In this figure, we can see how the hierarchical structure of a multi-label toy problem is embeeded into the ECOC coding matrix, which provides a multi-label prediction at the decoding step.

Our preliminary results suggest the viability of such approach, obtaining performances similar to state-of-the-art multi-label approaches. In addition, since the number of classifiers used is small, the overall computational cost is also reduced.

References

[1] Grigorios Tsoumakas and Ioannis Katakis, “Multi-label classification: An overview” *Int J Data Warehousing and Mining*, 2007, 1-13, 2007.

[2] Yi Zhang and Jeff G. Schneider, ”Multi-Label Output Codes using Canonical Correlation Analysis” *JMLR - Proceedings Track*, 15, 873-882, 2011.

[3] Bautista, M.Á. and Escalera, S. and Baró, X. and Radeva, P. and Vitriá, J. and Pujol, O., ”Minimal Design of Error-Correcting Output Codes” *Pattern Recognition Letters*, 33, 693-702, 2012.

[4] Miguel Ángel Bautista and Oriol Pujol and Xavier Baró and Sergio Escalera, ”Introducing the Separability Matrix for Error Correcting Output Codes Coding” *Multiple Classifier Systems*, 1, 227-236, 2011.

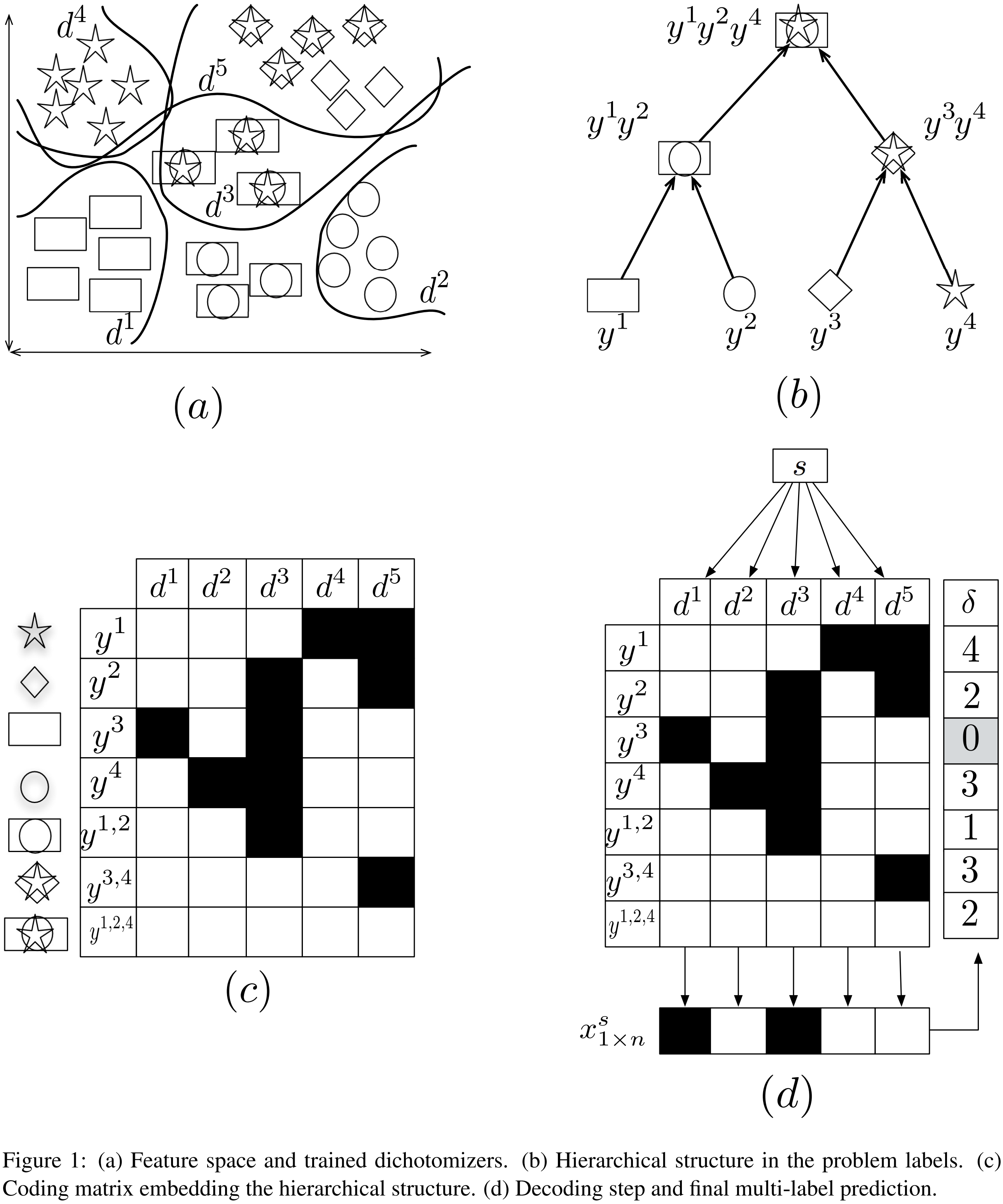

Figure 1: (a) Feature space and trained dichotomizers. (b) Hierarchical structure in the problem labels. (c) Coding matrix embedding the hierarchical structure. (d) Decoding step and final multi-label prediction.