Extraction of high-level features and labels in multi-label classification problems

Marilyn Bello[1,2], Gonzalo Nápoles[2,3], Ricardo Sánchez[1], Koen Vanhoof[2], and Rafael Bello[1]

[1] Computer Science Department, Central University of Las Villas, Cuba
[2] Faculty of Business Economics, Hasselt University, Belgium
[3] Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands
mbgarcia@uclv.cu

1 Introduction

Deep learning [4] is a promising avenue of research into the automated extraction of complex data representations at high levels of abstraction. Such algorithms develop layered, hierarchical learning architectures of data representations, where higher-level features are defined in terms of lower-level features. Pooling layers [4, 6] provide an approach to downsampling feature maps by summarizing the presence of features in patches of the feature map. They focus on data with a well-defined structure where the term feature neighborhood makes sense.

However, while it is interesting to recognize faces, or classify objects in images and videos, the truth is that there are other domains in which the data do not have a topological organization [7, 8]. For example, when using numerical descriptors to encode a protein, it might happen that two distant positions in the sequence are close to each other in the tri-dimensional space. That behavior could not be captured with local pooling methods. In those cases, using standard pooling operators might have little sense, even when the problem at hand could benefit significantly from a deep learning solution. Besides, although these operators are able to deal with both single-label and multi-label classification (MLC) problems [3], they are specifically aimed at reducing feature space. However, in the case of multi-label data, we can benefit significantly from implementing similar operations on the label space.

Hence, we propose a deep neural architecture to extract high-level features and labels in MLC problems [1, 2]. This approach, unlike the classic use of pooling, does not pool pixels but problem features or labels. The following sections provide a brief description of our proposal.

2 Bidirectional Deep Neural Network

This architecture, proposed in [1], is composed of several stacked association-based pooling layers, which are built starting from the features and the labels at the same time. The first pooling layer is composed of neurons denoting the problem features and labels, whereas in deeper pooling layers the neurons denote high-level features and labels extracted during the construction process. Each pooling layer uses a function that detects pairs of highly associated neurons (i.e. they fulfill a certain association threshold) while performing an aggregation operation to derive the pooled neurons. We use Pearson's correlation to estimate the association degree between two neurons. We compute the correlation matrix among features and labels, and derive the degree of association of the pooled neurons from the degree of association between each pair of neurons in the previous layer. The pooling process is repeated over aggregated features and labels until a maximum number of pooling layers is reached.

Once the high-level features and labels are extracted from the dataset, they are connected together with one or several hidden processing layers. These hidden layers are equipped with either ReLU, sigmoid or hyperbolic tangent transfer functions, therefore conferring the neural system with prediction capabilities. Finally, a decoding process [5] is performed, which connects the high-level labels to the original ones by means of one or more hidden processing layers.

3 Computing the degree of association among neurons from granulation entropy

In [2], we present a new method that replaces the correlation measure (i.e. that quantifies the association between two neurons) with another one that computes the entropy in the information granules that are generated from two features or labels. Unlike the pooling approach proposed in [1], this proposal does not require that either the features or labels have a certain degree of correlation with each other. The rationale behind the proposal suggests that two features (or labels) can be associated if the granulations generated from them have equal entropy [9]. Therefore, the proposal consists in obtaining a universe granulation, where each feature (or label) defines an indiscernibility relation, and the information granules are the set of indiscernible objects with respect to the feature (or label) under consideration. In this way, it is verified if the coverings (or partitions) generated by two features (or labels) induce similar entropy values.

4 Concluding Remarks

The numerical simulations on several MLC datasets show a significant reduction in the number of problem features and labels (i.e. a reduction of up to 96% and 87%, respectively), without affecting network's discriminatory capability. Having a smaller neural system implies that the training time is smaller when compared with a model that uses the full set of features and labels. Despite of the relatively good results reported by the model in [1], the function used to quantify the association between problem variables does not seem to be suitable for datasets having poor correlation among their features or labels. In order to mitigate this, a variant based on granulation entropy in [2] is proposed.

References

1. Bello, M., Nápoles, G., Sánchez, R., Bello, R., Vanhoof, K.: Deep neural network to extract high-level features and labels in multi-label classification problems. Neurocomputing **413**, 259–270 (2020)
2. Bello, M., Nápoles, G., Sánchez, R., Vanhoof, K., Bello, R.: Feature and label association based on granulation entropy for deep neural networks. In: International Joint Conference on Rough Sets. pp. 225–235. Springer (2020)
3. Charte, F., del Jesus, M.J., Rivera, A.J.: Multilabel classification: problem analysis, metrics and techniques. Springer (2016)
4. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT press Cambridge (2016)
5. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science **313**(5786), 504–507 (2006)
6. Lee, C.Y., Gallagher, P.W., Tu, Z.: Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In: Artificial intelligence and statistics. pp. 464–472 (2016)
7. Wan, S., Duan, Y., Zou, Q.: Hpslpred: an ensemble multi-label classifier for human protein subcellular location prediction with imbalanced source. Proteomics **17**(17-18), 1700262 (2017)
8. Xia, Y., Chen, K., Yang, Y.: Multi-label classification with weighted classifier selection and stacked ensemble. Information Sciences (2020)
9. Yao, Y.: Probabilistic approaches to rough sets. Expert systems **20**(5), 287–297 (2003)