EXTRACTION OF HIGH-LEVEL FEATURES AND LABELS IN MULTI-LABEL CLASSIFICATION PROBLEMS

MARILYN BELLO, GONZALO NÁPOLES, RICARDO SÁNCHEZ, KOEN VANHOOF, RAFAEL BELLO

MULTI-LABEL CLASSIFICATION

Suppose $X = R^d$ denotes the d-dimensional instance space, and $L = \{l_1, l_2, \ldots, l_k\}$ denotes the label space with k being the possible class labels. For each multi-label example (x_i, L_i), $x_i \in X$ is a d-dimensional feature vector $(x_{i1}, x_{i2}, \ldots, x_{id})$ and $L_i \subseteq L$ is the set of labels associated with x_i. For any unseen instance $x \in X$, the multi-label classifier $h(\cdot)$ predicts $h(x) \subseteq L$ as the set of proper labels for x.

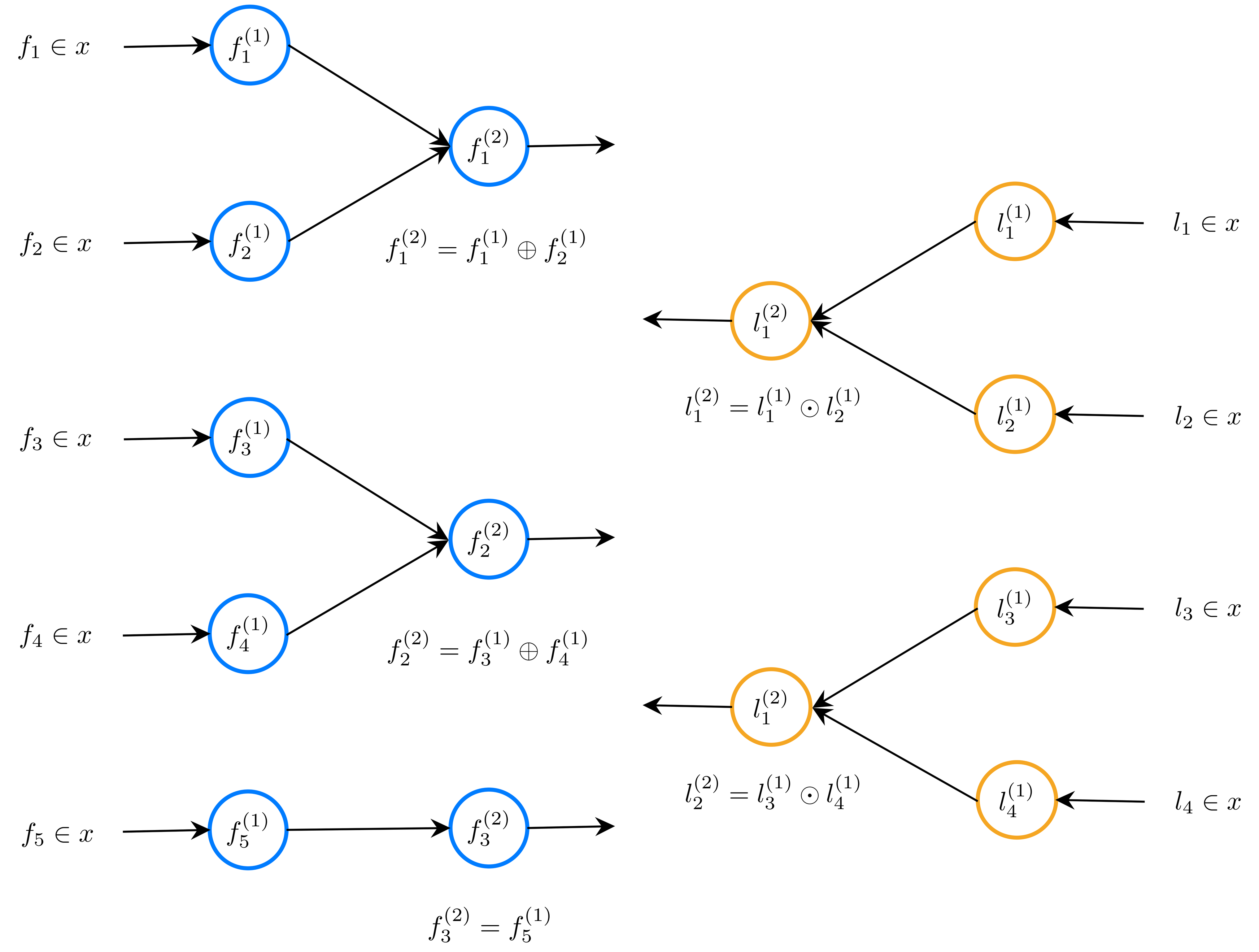
INTRODUCTION

- Deep learning is a promising avenue of research into the automated extraction of complex data representations at high abstraction levels.
- Several authors have proposed multi-label classification solutions inspired by deep learning techniques, in which the data have a topological organization.
- There are other domains in which the data do not have a topological organization, and using standard pooling operators might have little sense.

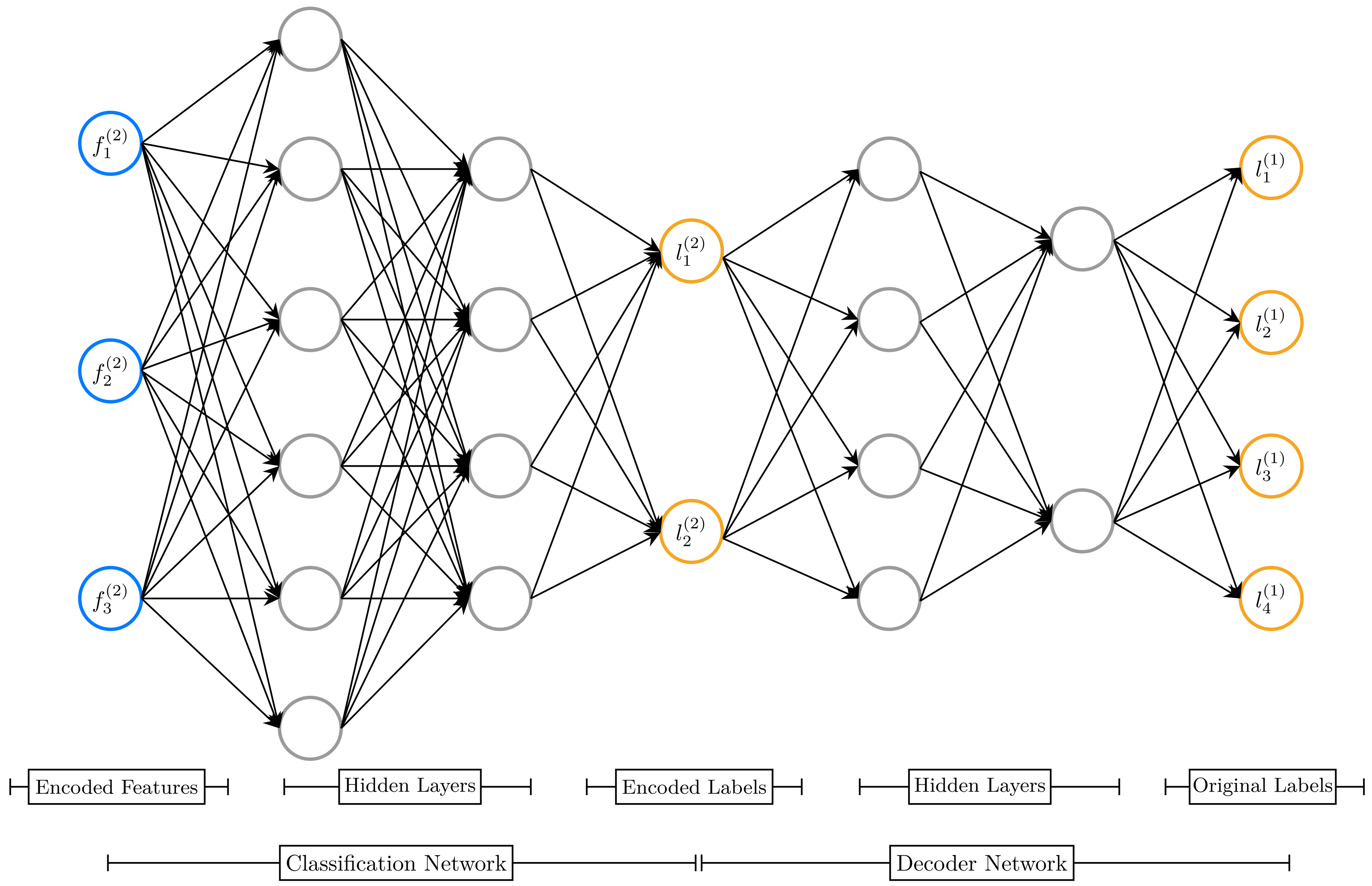
PROPOSAL BASICS

- It is composed of several stacked association-based pooling layers.
- Unlike the classic use of pooling, does not pool pixels but problem features or labels.
- The first pooling layer comprises neurons denoting the problem features and labels, whereas, in deeper pooling layers, the neurons represent high-level features and labels extracted.
- Each pooling layer uses a function that detects pairs of highly associated neurons while performing an aggregation operation.
- High-level features and labels extracted are connected with several hidden processing layers.
- A decoding process is performed, connecting the high-level labels to the original ones through one or more hidden layers.

BIDIRECTIONAL ASSOCIATION-BASED POOLING



BIDIRECTIONAL DEEP NEURAL ARCHITECTURE



DEGREE OF ASSOCIATION

A ***first variant*** uses Pearsons correlation to estimate the association degree between two neurons. As a ***second variant***, the entropy is computed in the information granules generated from two features or labels. The rationale behind this variant suggests that two features (or labels) can be associated if the granulations generated from them have equal entropy.

DISCUSSION

- The numerical simulations on several MLC datasets show a significant reduction in the number of problem features and labels, i.e. reducing up to 96% and 87%, respectively.
- Pearsons correlation does not seem suitable for datasets having a low correlation among their features or labels.
- The second variant reported higher reduction values in datasets having low correlation values among their features and labels.
- Our proposal does not aim to increase the prediction rates but reduce the number of deep feed-forward neural network parameters without harming their discriminatory power.
- Our results cry for the implementation of a convolutional operator to also increase networks discriminatory power.

SCAN FOR FULL PAPERS

CONTACT INFORMATION

mbgarcia@uclv.cu
G.R.Napoles@tilburguniversity.edu