Convolutional Neural Networks for Attribute Classification

Juan Camilo Martinez
Universidad de los Andes
Departamento de Ingeniera Industrial
jc.martinez10@uniandes.edu.co

Abstract

In this work, a Convolutional Neural Network (CNN) is used to infer features and solve a multiclass, multilabel classification problem. Alternative methods for the solution of this problem are presented, and an evaluation is performed using the CelebA dataset.

Keywords: CNN, multiclass, multilabel, classification

1. Introduction

Convolutional Neural Networks (CNNs) are capable of mapping image data to an output variable while. They excel at processing data with a spatial relationship. CNNs can take image files directly as inputs without the need to perform feature extraction.

In this work, a CNN is used to perform classification on a subset of the CelebA database. CelebA is compossed of 202.599 images annotated with one or more of 40 attributes. As the attributes are not mutually exclusive (an image can have several of them or only one) this is a type of problem is known as multilabel classification.

2. Methodology

2.1. Multilabel Classification

Compared to binary classification (belongs to a class or not) and multiclass classification (belongs to exactly one of several classes) multilabel classification can be seen as an evolution of the previous two problems. As such, solutions for them can be tweaked to produce the desired results.

For instance, the problem can be thought of as a combination of binary classification problems. Multiple classifiers are trained, each one designed to separate one and only one class from the rest. Exemplar SVMs are a possible variation.

The problem can also be transformed into a multiclass one by defining every possible combination of labels as classes themselves.

Figure 1. Sample pictures of the CelebA database

2.2. CNN

The convolutional layer is what makes a CNN a CNN, so the architecture must include at the very least one. However, as the final output of the classifier is a vector with the classes the input belongs to, at least one fully connected layer must be present.

For this problem, a network based on a simplified AlexNet is implemented, composed of 3 convolutional layers, each connected to a ReLu activation layer and a max pooling layer. Two fully connected layers then reduce dimensionality until the class vector is the output. the implemented loss function is called BCELoss.

2.3. Ablation Studies

The removal of layers leads to a significant performance boost at the expense of complexity.

3. Results

4. Conclusions

The PHOW strategy for image classification was implemented for a small subset of the Imagenet database. This database includes classes that are very similar to each other (multiple classes of dogs and lizards, for example) and substantially different from each other. .

References