Semantic Classification for Big Data Analysis

Amina Taouli, Djamel Amar Bensaber, Nabil Keskes, and Khayra Bencherif

LabRI Laboratoty, Ecole Supérieure Informatique, Sidi Bel Abbes, 22000, Algeria
{ a.taouli, d.amarbensaber, n.keskes, k.bencherif }@esi-sba.dz

Abstract. Big Data can be defined as a dataset that contains a huge volume of information which can be analyzed to discover unknown and useful patterns. In fact, Deep Learning is a technique that is used mainly to facilitate the Big Data Analysis by extracting complex abstractions of high level. Nevertheless, data heterogeneity represents a key challenge in the context of Big Data Analysis. Usually, data providers use different techniques to represent the same real-world object. Moreover, they have a lack of methods for the automation of the classification of the input images. Therefore, we have to add the semantic aspect in the classification for enhancing the Big Data Analysis process. In order to meet this requirement, we put forward an approach that allows semantizing the classification using Convolutional Neural Network and Semantic Memory. We apply the pooling operation to reduce the size of the input data. After a filter succession, the convolution maps are concatenated into a 1-D vector and then we use Semantic Memory to classify the input data.

1 Introduction

Nowadays, the total amount of data consumed and offered on the Web has increased. For this purpose, the researchers expected that in 2020 there will be more than 16 zettabytes of useful data Turner et al. (2014). In connection with this explosion, the term Big Data has emerged to refer to any collection of massive, high-speed and heterogeneous datasets. In this context, we face the main challenges of acquiring, integrating, analyzing and visualizing large amounts of data Curry (2014). This amount of data is captured in various forms containing unstructured data that account for 90% of all data.

In order to make the raw data acquired usable in the recommendation, decision and prediction systems, it is recommended to use Big Data Analysis (BDA). In fact, BDA is the process of adding structures to the data to find facts, relationships, patterns, and extracting hidden information.

Currently, the Deep Learning (DL) has a huge success in several areas such as speech recognition, image processing and natural language Mohammadi et al. (2017). In addition to the data growth, the DL plays an important role in providing predictive solutions for BDA that consist of having a better result, understanding and detecting relationships between data and predicting future instances.