Cloud Implementation of a Neural Classifier for Remotely Sensed Hyperspectral Images

J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, F. del Frate

Hyperspectral Computing Laboratory (HyperComp). Department of Computer Technology and Communications. Escuela Politecnica de Caceres, University of Extremadura. Avenida de la Universidad sn, E-10003 Caceres, SPAIN. **juanmariohaut@unex.es**

Introduction

Remotely sensed hyperspectral imaging (RS-HSI) is able to collect hundreds of images captured at different wavelengths for the same area on the Earth's surface, characterizing in unique way the observed materials at a pixel level through their spectral signatures. In this context, new HSI missions have been already planned by national agencies that might be very interesting for regional applications, for example in the field of agriculture. In fact, the use of RS-HSI data in precision agriculture field has a great impact, due to its powerful ability to accurately discriminate the types of crops and their state through the behavior of solar spectrum that is absorbed and reflected by them [1], helping to detect plagues and diseases in crop areas and supporting decision-making [3].
However, the high dimensionality of these data volumes, coupled with the current rate of data collection carried by airborne and satellite-platform sensors (such as the AVIRIS or the EO-1 Hyperion, among others) make the development of highly computational-efficient techniques a significant research effort, due to the complexity of processing and storing such great amount of data. Moreover, most HSI processing algorithms exhibit high computational complexity. As a result, there is a need to develop implementations of such algorithms on high performance computing (HPC) architectures. In particular, cloud computing technologies are able to provide both distributed storage and processing of large HSI repositories which are typically located in different datacenters. This work presents a new cloud computing implementation (developed using Apache Spark) of a neural network classifier for RS-HSI data.

Distributed Framework Design

Related with the computing engine, as Fig. 1 shows, the proposed distributed environment is based on: 1) *OpenStack*, a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, providing Infrastructure as a Service (IaaS), which allows to abstract and manage physical machines that will give the support to virtual machines, and 2) *Apache Spark*, a widely used framework for large-scale data processing on cloud computing architectures, which implements a fault-tolerant abstraction for in-memory cluster computing, and provides fast and general data processing on large distributed platforms.

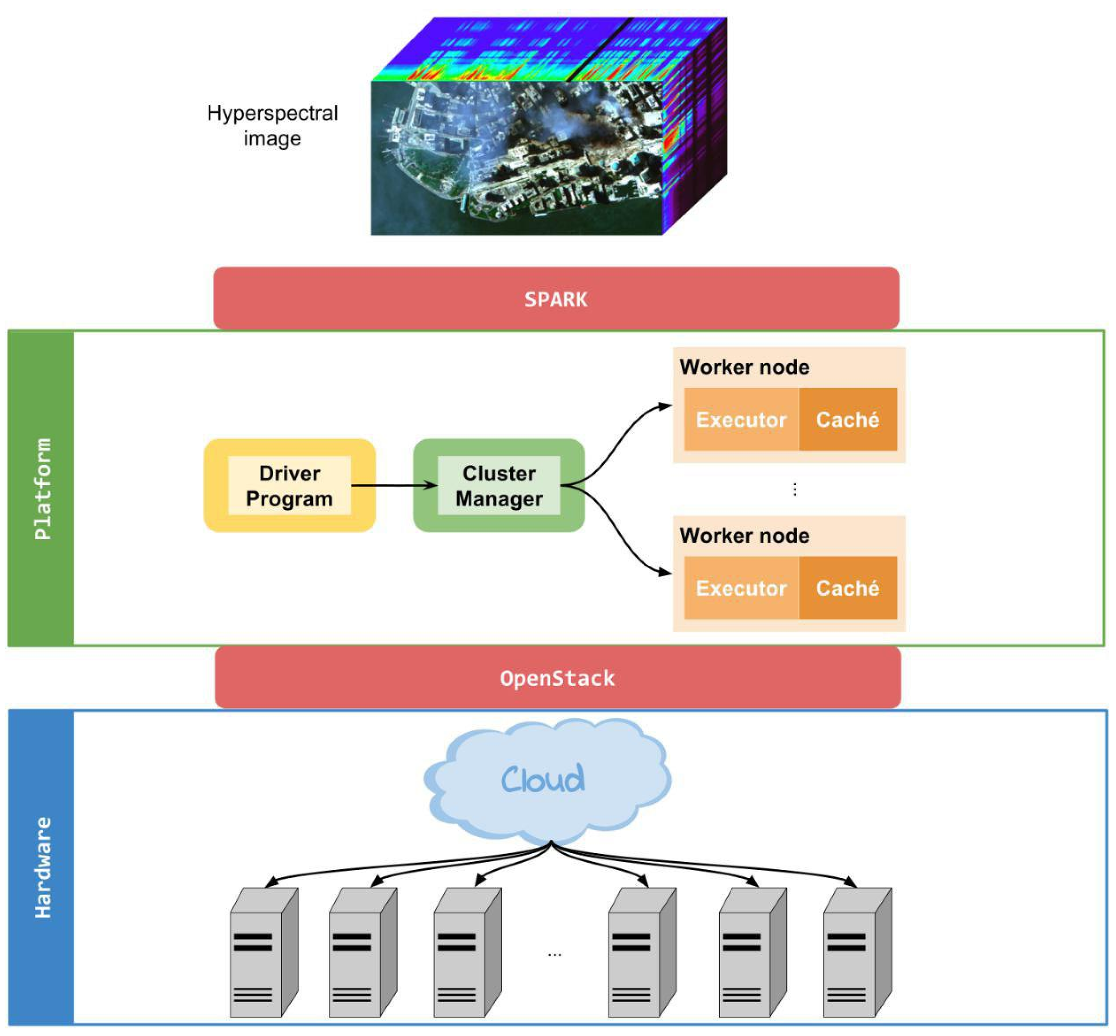

Figure: 1. Integrated Apache Spark and OpenStack framework

In addition, the distributed programming model follows the MapReduce [2] scheme provided by Apache Spark. In this sense, HSI data is organized as key/value pairs to which several actions and transformations are applied, creating a hierarchy of task and subtask that are executed in parallel inside each worker node.

Distributed Implementation

The following algorithm concisely describes the steps followed in the implementation of the distributed neural network for the classification of RS-HSI data:

Algorithm 1 MLP Algorithm

procedure MLP_CLASSIFIER($n_{iterations}$, $n_{tolerance}$)
  for $i < n_{iterations}$ & $t < n_{tolerance}$ **do**
    Master node sends parameters to slaves ▷ + slaves, - calculation + communications
    Slaves compute its gradient of error $\nabla E(\mathbf{X})_s$.
    Slaves send the gradient to master node.
    Master node obtains the final gradient $\nabla E(\mathbf{X}) = \sum_s \nabla E(\mathbf{X})_s$,
    Master node calculates the update of model's weights $\Lambda\mathbf{W}$
  end for
end procedure

The proposed approach has been tested using the **Big Indian Pines** dataset, which has a size of 2678×614 pixels. It was collected over an agricultural area with regular crop patches and irrecular zones of forest. It contains 220 spectral bands in the range from 400 to 2500 nm, with spectral resolution of 10 nm, moderate spatial resolution of 20 nm and 16 bits of radiometric resolution. The percentage of pixels with ground truth information is 20.33% (334245 out of 1644292 pixels) and the total number of classes is 58, which come in the form of a single label assignment per pixel. In the following we show the speedups obtained by our newly developed cloud implementation.

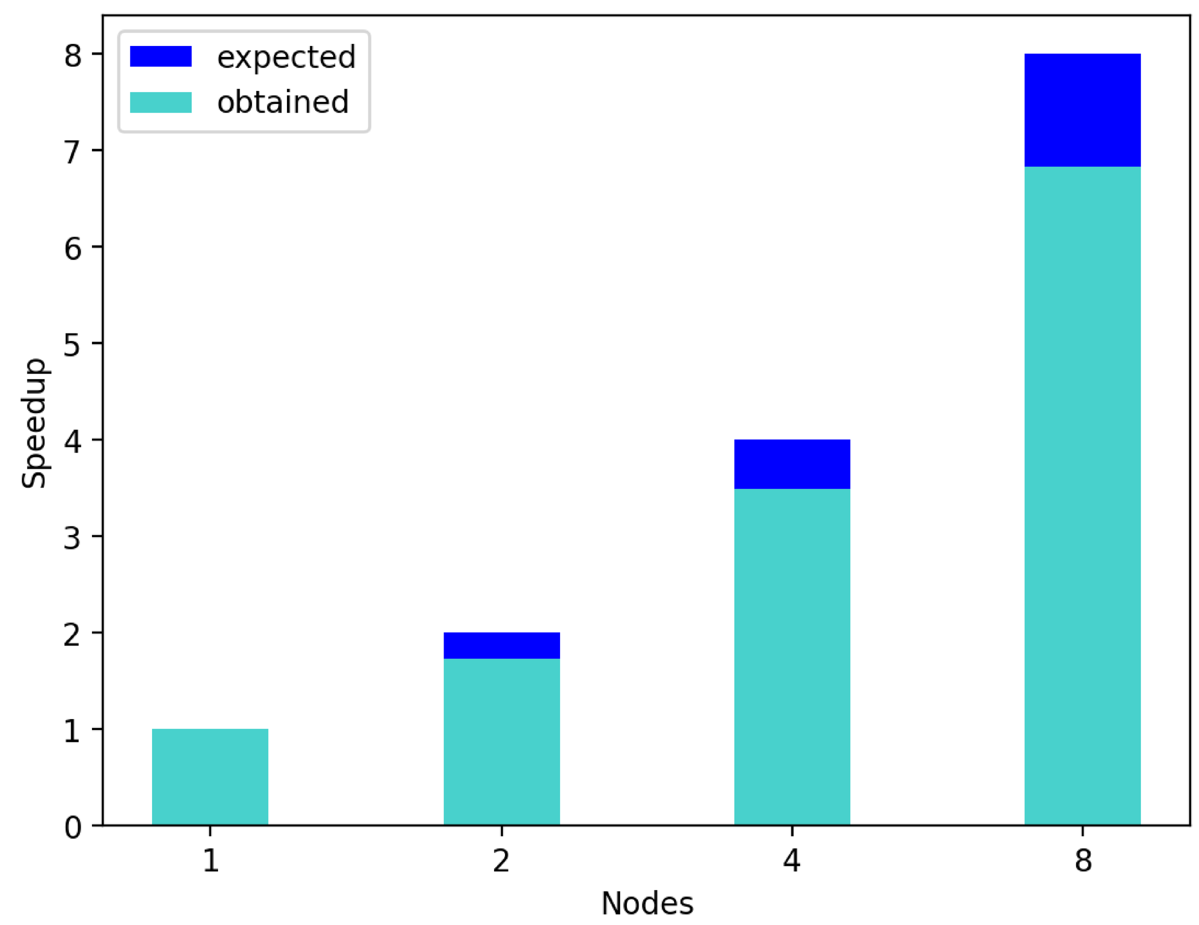

Figure: 2. Speed up obtained with 1, 2, 4 and 8 worker nodes over Big Indian Pines dataset

Results and Conclusions

In this work, a cloud computing implementation of a neural classifier has been developed for RS-HSI data classification on Apache Spark platforms. Experimental results, indicated in Fig. 2, show the effectiveness of the proposed distributed implementation with large HSI datasets in terms of computational performance and classification accuracy, obtaining **6.81x speedup** with 8 worker nodes versus 1 worker node. The proposed implementation allows for the exploration and analysis of large HSI data repositories, as the processing times can be significantly reduced, and the images can be processed in distributed fashion. This is important, since HSI repositories are characterized by their distributed nature. As future work, we will implement other algorithms for hyperspectral data classification using cloud computing platforms.

Acknowledgements

This work has been supported by Ministerio de Educación (Resolución de 26 de diciembre de 2014 y de 19 de noviembre de 2015, de la Secretaría de Estado de Educación, Formación Profesional y Universidades, por la que se convocan ayudas para la formación de profesorado universitario, de los subprogramas de Formación y de Movilidad incluidos en el Programa Estatal de Promoción del Talento y su Empleabilidad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016. This work has also been supported by Junta de Extremadura (decreto 297/2014, ayudas para la realización de actividades de investigación y desarrollo tecnológico, de divulgación y de transferencia de conocimiento por los Grupos de Investigación de Extremadura, Ref. GR15005). This work has been additionally supported by the Spanish Ministry of Economy under the projects ESP2016-79503-C2-2-P and TIN2015-63646-C5-5-R.

References

[1] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan. Hyperspectral vegetation indices and novel algorithms for predicting green lai of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sensing of environment*, 90(3):337–352, 2004.

[2] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei. Parallel and Distributed Dimensionality Reduction of Hyperspectral Data on Cloud Computing Architectures. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP(99):1–9, 2016.

[3] M. Zhang, Z. Qin, X. Liu, and S. L. Ustin. Detection of stress in tomatoes induced by late blight disease in california, usa, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 4(4):295–310, 2003.