

Deep Learning application for spectral emissivity targets recognition

Maria Luján Iglesias¹, Pâmela Suélen Käfer¹, Sílvia Beatriz Alves Rolim¹, Gabriel Prates Hallal¹ and Jean Marcel de Almeida Espinoza². ¹Federal University of Rio Grande do Sul (UFRGS) - State Research Center for Remote Sensing and Meteorology. Av Bento Gonçalves, 9500, Campus do Vale, 91501-970, Porto Alegre, RS, Brazil. ²Federal Institute of Rio Grande do sul (IFRS) – Campus Rio Grande. [Contact: lujaniglesias@gmail.com]

Copyright 2019, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 16th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 19-22 August 2019.

Contents of this paper were reviewed by the Technical Committee of the 16th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

In this work, we use a simple deep learning model to classify a set of emissivity spectrum from samples into six different classes. We test the network for only one hidden layer, changing the hidden units. The results showed an accuracy of 1.0 after 200 iteration steps (epochs) for 100 hidden units. The novelty of the work resides in the fact that there is no precedent for it to be applied for the classification of emissivity values in the range of thermal infrared, being this range of the electromagnetic spectrum, so important for geological mapping and mineral exploration.

Introduction

Vibrational spectroscopy provides a means for studying geologic materials both in the laboratory and in remote sensing applications (Ruff *et al.* 1997). The study of minerals and rocks using this technique is based on the principle that molecules vibrate as they interact with propagating electromagnetic (EM) energy (Lane and Christiansen 1998). The fundamental frequencies of internal molecular vibration for most geologic materials occur in the mid-infrared range of the EM spectrum, and for this reason it is a very useful, non-invasive technique for characterization of materials for planetary exploration. In despite of this, it is not easy for the human eye to identify a material from its emissivity spectrum.

In this sense, machine-learning techniques are becoming increasingly important. In particular, deep learning has proven to be both a major breakthrough and an extremely powerful tool in many fields (Zhu *et al.* 2017)

Remote-sensing scientists have exploited the power of deep learning to tackle different challenges and instigated a new wave of promising research. Deep learning involves a class of models which try to hierarchically learn deep features of input data. The network is first layer-wise initialized via unsupervised training and then tuned in a supervised manner. In this scheme, high-level features can be learned from low-level ones, whereas the proper features can be formulated for pattern classification in the end (CHEN *et al.*, 2014). The most commonly-used neural network model for classification in remote sensing is the multi-layer perceptron trained by the back-propagation algorithm (Rumelhart *et al.* 1986).

Given a dataset of shape (n,m) , where n is the number of data samples and m is the length of each data sample. The basic goal of a neural network is to learn a function

from a set of fixed-size inputs $\mathbf{X} = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, $x_i \in \mathbf{R}^m$, to a fixed-size output $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$, $y_i \in \mathbf{R}^m$. The network consists of the input layer (Figure 1), one or more intermediate or hidden layers consisting of a set of units or nodes, and an output layer.

In a feed-forward neural network, the layers are ordered and each unit of a layer connects to some subset of the units of previous layers. A layer is dense if each of its units connects to all of the units in the previous layer. A network is deep if there are several hidden layers. Each connection between units in adjacent layers has a weight and each unit has a bias.

Method

During the training, after computing the weighted sum of the inputs, an *activation function* is calculated. These two steps can be formulated as

$$a^l = f(w^l a^{l-1} + b^l) \quad (1)$$

where the w^l is the weight matrix for each layer, initialized with random values, and b^l is the bias vector for each layer, initialized with zeros. The activation function of a node defines the output of that node. This output is then used as input for the next node and so on.

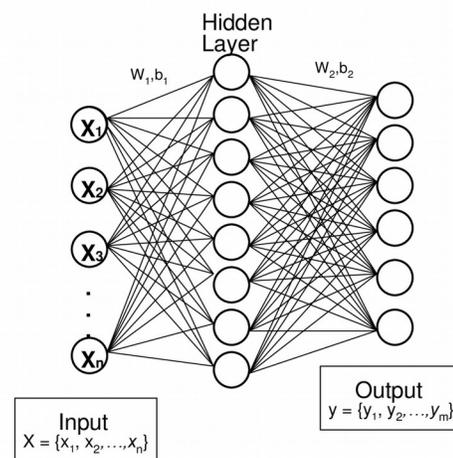


Figure 1. Diagram of the proposed neural network for the spectrum recognition (The hidden layer units are reduced for simplicity)

The nonlinear functions $f(\cdot)$ are generally chosen to be sigmoidal functions such as the logistic sigmoid or the 'tanh' function (Bishop, 2006).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The objective of training a classifier is to find the weights and biases which minimize the cost function $C(w,b)$. This cost is calculated with the *cross-entropy* error as:

$$c(w,b) = \frac{-1}{m} \sum_{i=1}^m \sum_{k=1}^n [y_{ik} \log(a_{ik}) + (1 - y_{ik}) \log(1 - a_{ik})]$$

where n denotes the input vector size and m denotes the number of examples in the dataset. The inner summation is over the input dimension, whereas the outer is over a whole dataset. Each of the y_{ik} are the real-valued output. Following, backpropagation is performed. In order to update the weights and bias for improve the results.

The goal of backpropagation is to compute the partial derivatives of the cost function with respect to any weight and bias in the network.

Is about understanding how changing the weights and biases in a network changes the cost function.

Because $c(w,b)$ is a continuous function of w and b , its smallest value will occur at a point in weight space such that the gradient error function vanishes, so that

$$\nabla c(w,b) = 0 \quad (3)$$

as otherwise we could make a small step in the direction of $-\nabla c(w,b)$ and thereby further reduce the error. Our goal is to find w and b such that $c(w,b)$ takes its smallest value. To solve the equation (3) we resort to iterative numerical procedures. The simplest approach to using gradient information is to perform a small step in the direction of the negative gradient, so that

$$w^l = w^l - \eta \frac{\partial c(w^l, b^l)}{\partial w^l} \quad (4)$$

$$b^l = b^l - \eta \frac{\partial c(w^l, b^l)}{\partial b^l} \quad (5)$$

where $\eta > 0$ is known as the *learning rate*. After each such update, the gradient is re-evaluated for the new weight vector and the process repeated. At each step the weights and biases are moved in the direction of the greatest rate of decrease of the error function, and so this approach is known as *gradient descent optimization algorithm* to adjust the weight of neurons by calculating the gradient of the loss function (Nielsen, 2015).

The backpropagation algorithm is summarized below. Implementation details can be found in most neural network books (e.g. Bishop, 1995).

1. Initialize network weights
2. present first input vector, from training data, to the network.
3. Propagate the input vector through the network to obtain an output.
4. calculate an error signal by comparing actual output to the desired (target) output.
5. propagate error signal back through the network.

6. adjust weights to minimize overall error.

7. repeat steps 2-7 with next input vector, until overall error is satisfactorily small.

The backpropagation algorithm contains two adjustable parameters, the learning rate and a momentum term, which can assist the training process in avoiding this (Gardner, 1998). The learning rate determines the step size taken during the iterative gradient descent learning process. If this is too large then the network error will change erratically due to large weight changes, with the possibility of jumping over the global minima. Conversely, if the learning rate is too small then training will take a long time. The momentum term is used to assist the gradient descent process if it becomes stuck in a local minimum. By adding a proportion of the previous weight change to the current weight change (which will be very small in a local minimum) it is possible that the weights can escape the local minimum.

To test our Neural Network Classifier, we use six different targets (Figure 1). The emissivity curves S01, S02, S03 and S04 correspond to samples of sand from four locations into the dunes of Bujurú Beach, in the South Coast of Rio Grande do Sul state, Brazil. They differ from each other because of their Titanium concentration. The green curve S05 correspond to sand from the dunes of Cidreira Beach, in the North Coast of Rio Grande do Sul state, Brazil. And the red curve S06 correspond to the emissivity of Puyehue volcano ashes (erupted in 2011, south of Chile).

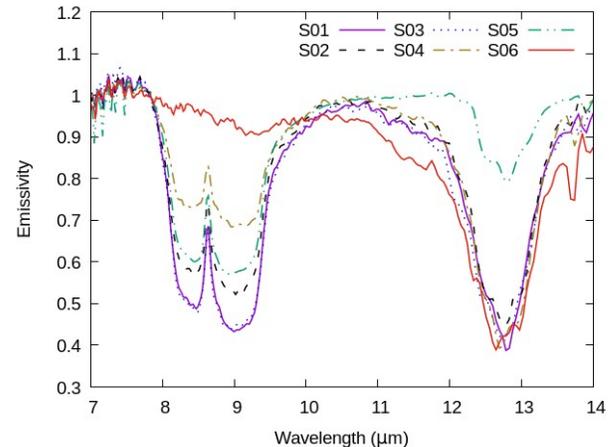


Figure 2. Emissivity spectral features for the six different samples tested in the neural network.

The samples were labeled as [0,1,2,3,4,5,6] respectively. The emissivity curves were measured under controlled temperature and humidity conditions, using a portable field spectrometer μ FTIR Model 102. The equipment allows the acquisition of infrared spectra of natural surfaces between 2 and 24 μm with spectral resolution of 2, 4 and 16 cm^{-1} , using a Michelson interferometer with two KBr prisms. The database created with the spectral emissivity of the samples consisted of 132 samples, 80 per cent for training the model and 20 per cent for testing it.

Although the curves within each sample group are very similar, there are small variations due to the temperature of the sample at the time of measurement.

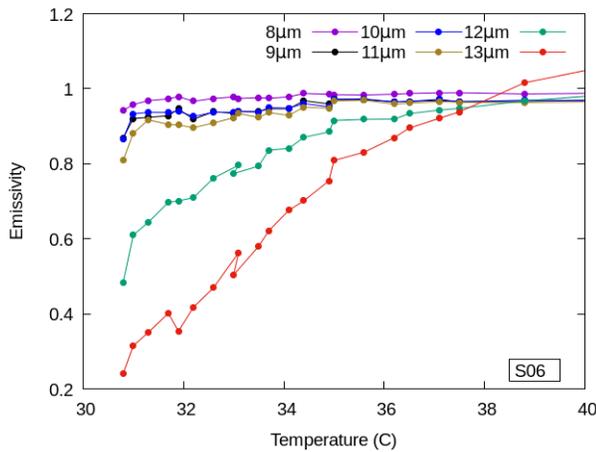


Figure 3. Emissivity in function of Temperature of the sample S06 for different wavelength.

This can be seen on Figure 3, where depending on the wavelength chosen, the emissivity remains almost constant independent of the temperature (from 8 to 11 µm) and for 12 and 13 µm the values increases in a more pronounced way.

This highlights the need to have a good algorithm that classifies the different targets into the class to which they belong, in spite of existing variations within the same class. Since, the human eye is unable to perform.

Results

We use 105 training samples for train the model using first 2 hidden units in the hidden layer, 10 and 100. After the training step we proceed to test the model in the 27 surplus samples. The results showed in Figure 4 met our expectations.

As expected, by placing more hidden units the algorithm *learns faster* and as a consequence the accuracy is significantly higher for 100 hidden layers than for 2 in a few hundred of iteration steps (*epochs*).

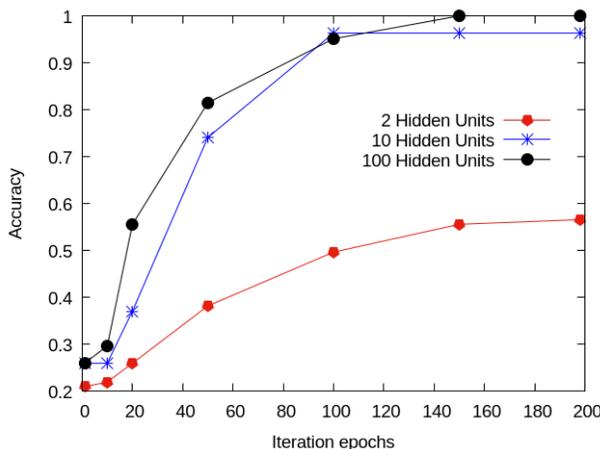


Figure 4. Accuracy in the predicted values for the test sample and for three values of hidden units in the hidden layer.

This result shows the importance of the application of new techniques, such as neural networks, in all branches of science and in particular in remote sensory, where we meet daily with the challenges of finding reliable algorithms when classifying. especially when we deal with emissivity spectra.

Conclusions

In this work we proposed an emissivity spectral classification using a neural network approach. It has been shown to be a useful tool for prediction and classification. It was shown that this technique is very powerful giving us a perfect accuracy in a few iteration steps.

This work is a preliminary result of a major project, We are increasing our data bank, adding a greater variety of samples, with different types of soils and minerals.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES), finance code 001.

References

BISHOP, C. M. (1995) Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

BISHOP, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.

CHEN, Y., LIN, Z., ZHAO, X., WANG, G., and GU, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094-2107, 2014.

GARDNER, M. & DORLING, S. (1998). Artificial neural networks (the multilayer perceptron)---A review of applications in the atmospheric sciences. *Atmospheric Environment*, 32, 2627--2636.

NIELSEN, M. A.; "Neural Networks and Deep Learning", Determination Press, 2015,

RUFF, S. W., P. R. Christensen, P. W. Barbera, and D. L. Anderson (1997), Quantitative thermal emission spectroscopy of minerals: A laboratory technique for measurement and calibration, *J. Geophys. Res.*, 102(B7), 14899–14913.

LANE, M. D., CHRISTENSEN, P.R., Thermal Infrared Emission Spectroscopy of Salt Minerals Predicted for Mars, *Icarus*, Volume 135, Issue 2, 1998, Pages 528-536, ISSN 0019-1035,

PAL, S.K., Mitra, S. "Multilayer perceptron, fuzzy sets, and classification," in *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 683-697, Sept. 1992.

ZHU, X. X. *et al.*, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8-36, Dec. 2017.