

DEEP NEURAL NETWORKS BASED SEMANTIC SEGMENTATION FOR OPTICAL TIME SERIES

Wei Yao, Mihai Datcu

Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany

ABSTRACT

Semantic segmentation or classification for satellite image time series (SITS) is a rarely touched topic, partly due to the difficulty in having the data, but more due to the unreachable task. In this research, we propose a dataset which consists of the Landsat image time series, with the purpose of performing multi-spectral semantic segmentation. As there is no ground truth information, we used unsupervised clustering to group time series into clusters, then Long short term memory (LSTM) unit based Recurrent neural networks (RNN) has been trained. We investigate the accuracy values for our test image patches, around 40% accuracy has been achieved for the sequence classification.

Index Terms— Satellite image time series, RNN, LSTM, Temporal pattern

1. INTRODUCTION

In remote sensing area, semantic segmentation always remains to be a challenging task, meanwhile, it's also a critical step for various applications afterwards. There are already a few researches focus on semantic segmentation from satellite imagery. However, a limited number of researches have been done regarding satellite image time series. Previously, researchers extract information rely on domain expertise.

Along with its overwhelming popularity among various research areas, deep learning techniques have shown surprising performances in image understanding, machine translation, time series prediction, etc. We remote sensing practitioners are also gradually following the trend, trying to utilize the powerful deep learning techniques to solve our specific problems. There are already some pioneers. Marc has proposed to get a fixed pixel-wise semantic classification label, later he used a sequence-to-sequence model which is able to obtain pixel-wise labels at each time [3]. Lichao has proposed a RNN to learn spectral-spatial-temporal features for

change detection in multispectral imagery [4]. More applications and analysis need to be endeavored.

Within the framework of the European Union project ECO-POTENTIAL, one of our goals is data mining for better understanding our surrounding environment. Hence in this research, we investigate the potential of deep learning methods for multi-spectral satellite image time series. Specifically, our dataset contains 300 Landsat satellite image time series with a size of 3030x1474 pixels which covers the Doñana national park in Spain.

Our challenge lies in that there is no ground truth information, and we are interested in finding the locations where show similar temporal patterns. Regarding temporal patterns, they can show repetitive pattern, sudden-change pattern, no-change pattern, etc. We prepared our datasets by using unsupervised clustering, then study and discover knowledge via experiments on deep neural networks. We choose the well-known Long Short Term Memory (LSTM) based Recurrent Neural Networks (RNN) for our experiments. In conclusion, the following points show our main contributions:

- We introduce a Landsat image time series dataset which consists of 299 scenes that covers Doñana national park for pixel-wise semantic segmentation.
- We tackle the problem of lacking ground truth information.
- We build multi-spectral models, based on deep learning LSTM-RNN learning scheme.

1.1. Processing Flow

In our approach, we have developed a tool which visualizes the image time series, choose different bands or characteristic indices (vegetation index, water index), and plot indices curves. In order to tackle the problem of lacking ground truth information and simplify the manual labeling, the dataset is cleaned and grouped into clusters as training dataset. Then it will be put into deep learning models to learn abstract and detailed representations. A

simple LSTM based Recurrent Neural Network (RNN) is applied to our time series; a LSTM RNN variant model which integrates LSTM based RNN and AutoEncoder has been tested.

The processing flow is composed of the following steps:

1. Preparation of training dataset (cleaning, clustering)
2. Data Observations
3. Application of LSTM model to the training dataset
4. Generation of the intermediate representations.

2. PREPARATION OF THE TRAINING DATASET

2.1. Dataset

Our dataset is Landsat products which cover the protected area of the Doñana National Park in southwestern Spain. It is an imagery time series which covers 299 scenes range from year 1999 to 2017. The Landsat SITS dataset is based on the WGS84 coordinate reference system. The images are located between the northwestern coordinates (-6.905383W, 37.221228N) and the southeastern coordinates (5.957231W, 36.76N).

The Landsat imagery consists of 6 bands. From band 1 to band 6, they are: blue, green, red, near infrared, shortwave infrared 1 and shortwave infrared 2. Figure 1 shows the paddy area at four time indices: 55, 110, 165 and 220.

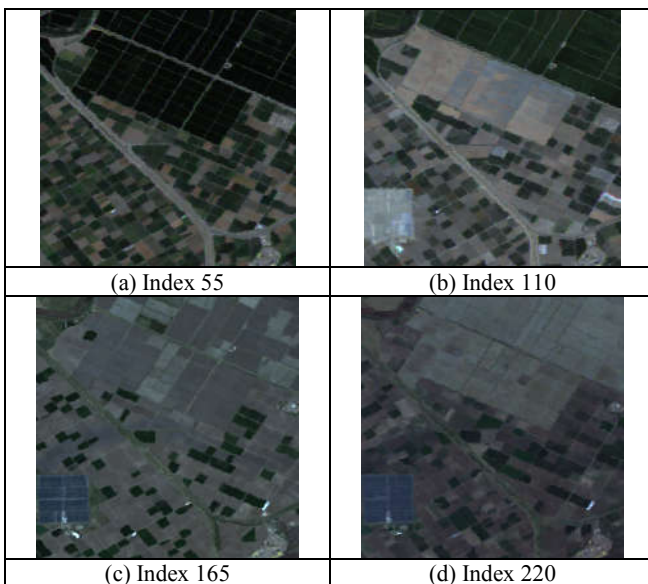


Figure 1. Image Time Series: Paddy Example.

2.2. Data Cleaning

By viewing the RGB bands, scenes with strong artefacts, scenes with clouds, scenes are saturated were rejected for further analysis. After removing them, there are 220 scenes left. By analysing NDVI curve, those scenes with abnormal values (e.g., very large number) are rejected.

2.3. k-Means Clustering

For each pixel location, the time series sequence is used as a data sample. All land regions and small parts of the ocean have been used as experimental data areas. After that, the data is subsampled: one data point from every 5 rows and every 5 columns. The subsampled data is taken as the input for k-Means clustering method, in order to group similar patterns together.

The blue band time series with a size of 1474 rows and 3030 columns have been grouped into 10 clusters. Figure 2 and Figure 3 show one example cluster and its covered mask which extracts the paddy area.

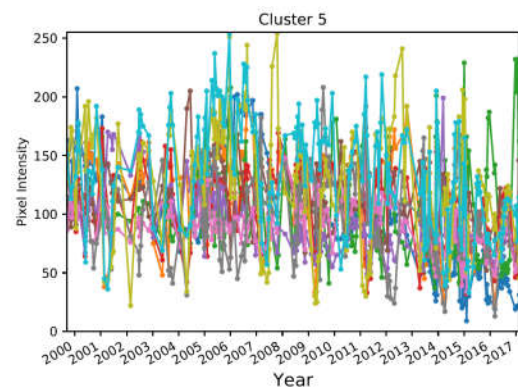


Figure 2. Cluster of NDVI Time Series.



Figure 3. Cluster Overlay: Paddy Areas.

3. METHODOLOGY

3.1. Landsat Product Characteristics

For Landsat imagery, five different kinds of indexes can be calculated. They describe the Landsat product characteristics [5]. Here we list the Normalized Difference Vegetation Index (NDVI), which is calculated as a ratio between the red (R) and near infrared (NIR) values in traditional fashion.

$$(NIR - R) / (NIR + R) \quad (1)$$

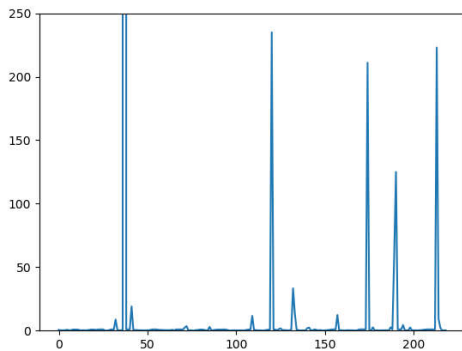


Figure 4. NDVI Curve of One Location.

Figure 4 shows the NDVI curve for one location within the paddy areas. Regarding temporal patterns, by data observations, we have noticed at least three temporal patterns: locations show repetitive pattern which may represent vegetation's seasonal ripe and harvest situation; locations show sudden-change pattern which may represent marshlands where are flooded in winter time; locations show no-change pattern which may represent roads or dunes where no changes when time passes by.

3.2. Recurrent Neural Networks (RNN)

Deep learning methods requires very little engineering by hand, they are very beneficial for us, aim at processing satellite imagery, by taking the advantages of applying them for the increases amount of available computation and data.

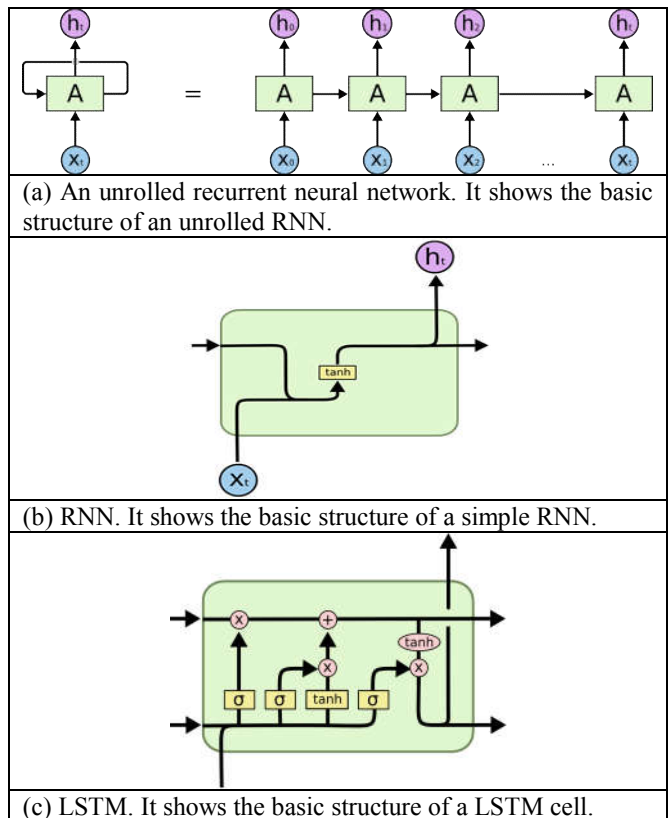
Especially for time series, Recurrent Neural Network (RNN) is a family of neural networks which process sequential inputs, such as speech and language. RNNs process one element from an input sequence at a time, a 'state vector' is kept in their hidden units which implicitly contains information of the history of all the past elements of the sequence [1].

3.3. Long Short Term Memory (LSTM)

With the RNNs framework, the Long Short Term Memory (LSTM) cell is proposed to overcome the long computing time problem caused by the normal recurrent backpropagation process in RNNs [2]. Hence, LSTM cell based RNNs use special hidden units which are able to keep memory for a long time. A LSTM cell consists of input gate, output gate and memory gate. Specifically, LSTM networks or GRU based RNNs are used for the encoder and decoder networks for machine translation [1].

3.4. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is a slightly variation on the LSTM. It combines the forget and input gates into a single "update gate", also merges the cell state and hidden state. GRU is simpler than standard LSTM model, and has been growing increasingly popular.



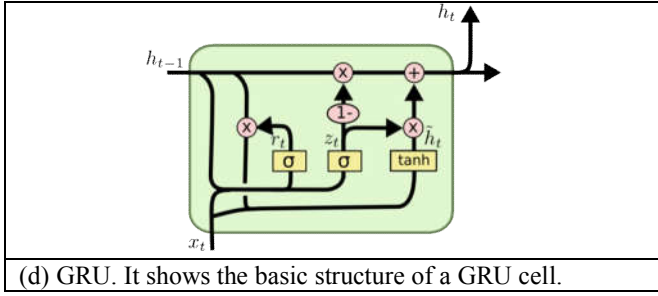


Table 1. LSTM, RNN Illustrations [6].

Table 1 shows illustrations which explain the concepts of RNN, LSTM cell and GRU cell.

4. EXPERIMENTS

4.1. Segmentation Experiment

For our preliminary experiments, we have taken 10 grouped clusters as our labels, we have performed a set of experiments with different parameter settings by using the LSTM based RNN:

Layers	3	3	3	3	3	5	7	10
Cells	16	32	64	96	128	32	32	32
Batch Size	64	64	64	96	128	64	64	64
Accuracy	0.12	0.35	0.40	0.38	0.38	0.36	0.34	0.37

Table 2. The First Segmentation Results.

Table 2 shows the first performance with different parameter settings. The best network performance with an accuracy of 40% has been achieved by networks with hyper-parameter settings {LSTM layers = 3, Cells = 64}.

4.2. Conclusion and Outlook

Facing the common problem of lacking of ground truth information, in this work, we start analyzing the data with the help of domain knowledge. By using unsupervised clustering for the vegetation index along time, the data are grouped into 10 clusters. The clustered data are used as training dataset for RNN deep neural networks. Around 40% accuracy is obtained. However, the result is very preliminary. More analysis will be performed in the future publications.

5. ACKNOWLEDGEMENT

Part of the work was conducted within the EU project ECO-POTENTIAL.

6. REFERENCES

- [1] Deep Learning, Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Nature, pp. 436-444, Vol. 521, 28 May 2015.
- [2] Long Short Term Memory, Hochreiter, S., Schmidhuber, J., Neural Computing, Vol. 9, pp. 1735-1780.
- [3] Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders, Marc Rußwurm, Marco Körner, ISPRS International Journal of Geoinformatics, 2018, pp. 1-19.
- [4] Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery, Lichao Mou, Lorenzo Bruzzone and Xiao Xiang Zhu, submitted to IEEE TGARS, March 6, 2018.
- [5] Product Guide: Landsat Surface Reflectance-Derived Spectral Indices, Department of the interior U.S. geological survey, version 3.6, December 2017.
- [6] Understanding LSTM Networks, <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>