Workshop on Survey Methodology:

Big data in official statistics

Block 7: Remote sensing data

20 May 2019,

Brazilian Network Information Center (NIC.br),

São Paulo, Brazil

Jan van den Brakel

Statistics Netherlands and Maastricht University

Introduction

Remote sensing data:

- Satellite and aerial images:
  - Measuring change in land use, air quality, desertification, deforesting, urban sprawl, etc
  - Measuring poverty and well being constructs in combination with survey data
- Road sensor data:
  - Measuring traffic intensity
  - Potential indicator economic growth

Measuring urban extension with satellite images

- CBS project; work by Liana Curier
- Data
  - MODIS Terra satellite: NDVI (250m—500m) using the Normalized Difference Vegetation Index
  - Training set: CORINE Land cover: inventory of land cover in 44 classes
- Methods
  - Classify pixels with land cover classes in the Netherlands:
    * Agricultural areas (white)
    * Artificial surfaces (red)
    * Forest and semi natural areas (green)
    * Water bodies (blue)
    * Wetlands (purple)

  - Machine learning algorithms:
    * Random forest
    * k Nearest Neighbour
    * Support Vector Machines
    * Deep learning using convolutional neural networks

- First results
  - Best classifier: random forest
  - Average accuracy 87% for land use in 5 classes
  - Validation with CORINE land cover data base

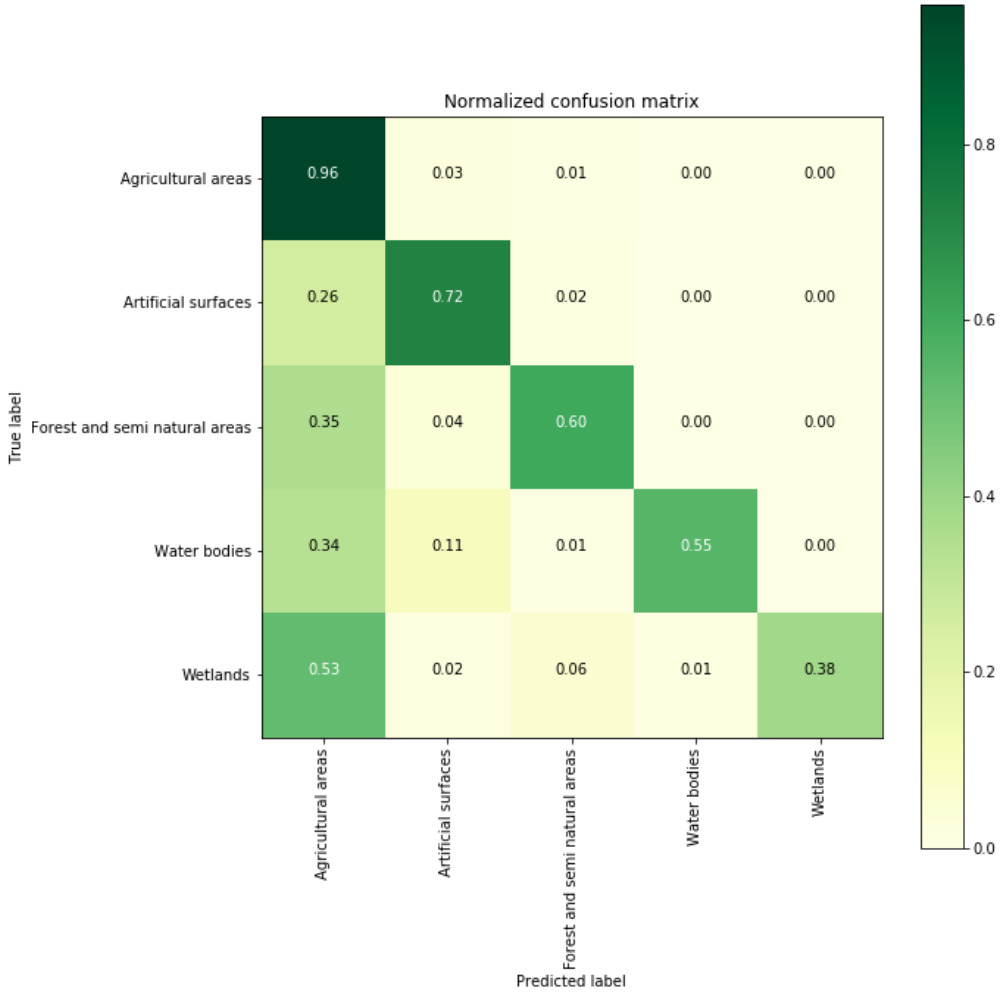

Figure 1: Confusion matrix of the Random Forest experiment on the proposed NDVI datasets.

Results RF:

- Presentation:
  - Colour of the inner circle indicates the prediction
  - Colour of the outer circle represent the true classification of CORINE land use

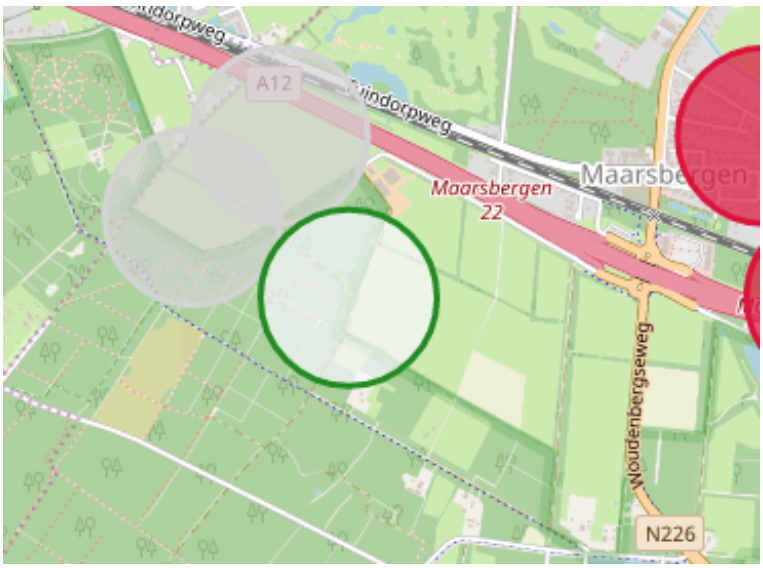

Figure 2: Misclassification Example

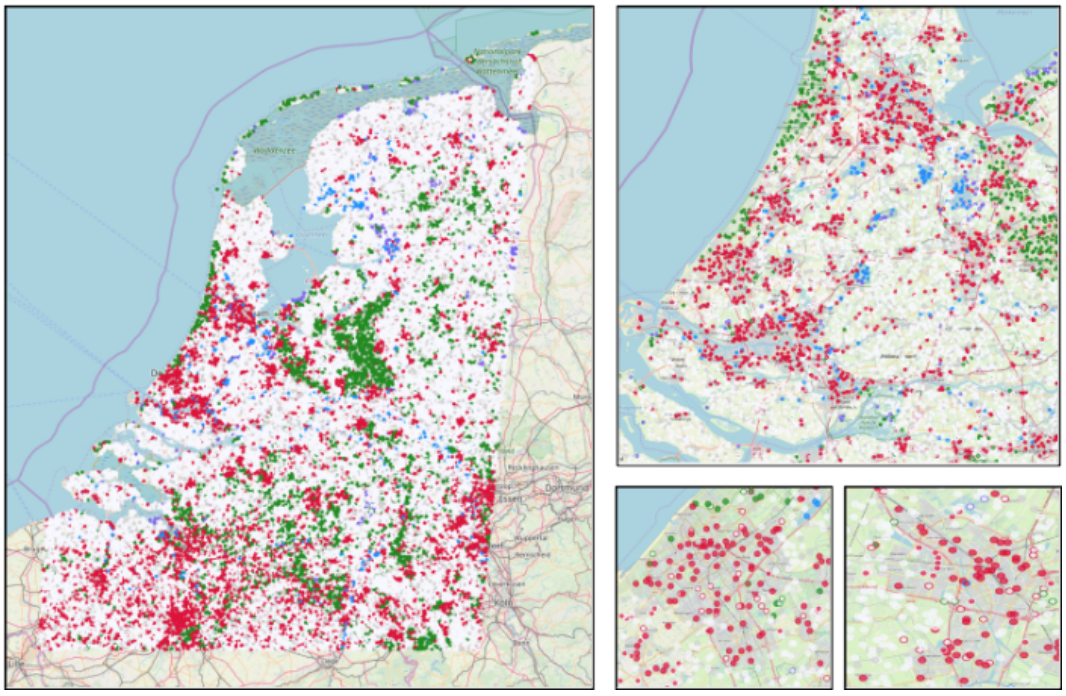

Figure 3: Mapping RF land cover classification. Each MODIS NDVI pixel is represented by a circle. Inner circle represent the prediction from the trained RF classifier, outer circle represent the land cover classification from 2012 CORINE land cover database. Agricultural areas, Artificial surfaces, Forest and semi natural areas Water bodies and Wetlands are marked in white, red, green, blue and purple respectively.

Measuring solar panels with aerial images

- CBS project; work by Tim de Jong
- Estimating the amount of photo-voltaic solar power to evaluate the energy transition
- Data:
  - Aerial pictures of the Netherlands with a resolution of 25cm $\times$ 25cm
  - Test set: pictures annotated manually
- Method:
  - Picture-based classifier: deep learning or convolutional neural network
  - A standard VGG16 convolutional neural network

Aerial images:

- Convolutional neural network
- For classification of pictures with and without solar panels

Figure 4: Example of a picture-based classification: a tile with a negative classification (left, without solar panels) and a tile with a positive classification (right, with solar panels).

Neural network:

- Input layer, hidden, layer output layer
- Nodes are connected via activation functions (sigmoide, relu, ...)
- Fully connected neural network
- Can be considered as multiple logistic regression functions
- Model can be made more complex by adding more hidden layers and / or more nodes

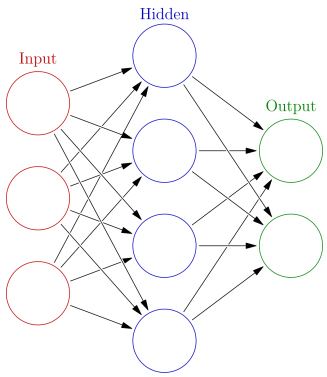

Figure 5: Fully connected neural net work.

Picture classification:

- Convolutional neural network
- Many hidden layers
- First set of layers: convolutional filters
  - Extract features from pictures
- Last set of hidden layers:
  - Fully connected
  - Classification based on the features extracted with the convolutional layers

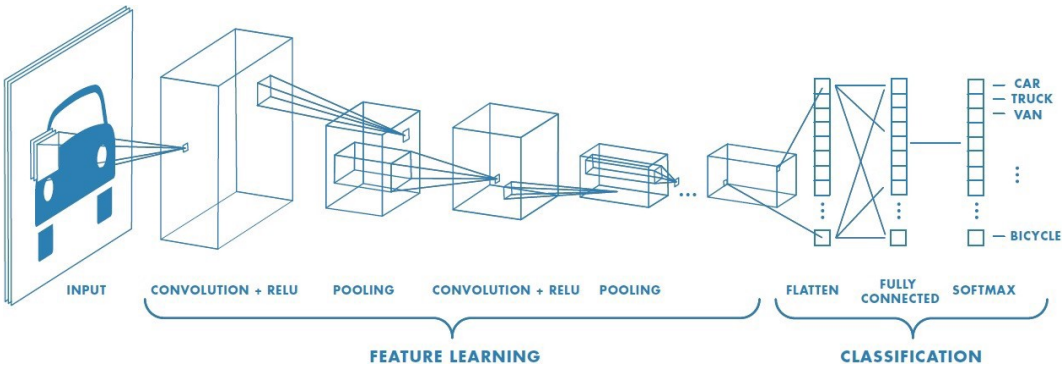

Figure 6: Convolutional neural network.