

Non-Time-Series-based Random-Forest for Land Cover Classification

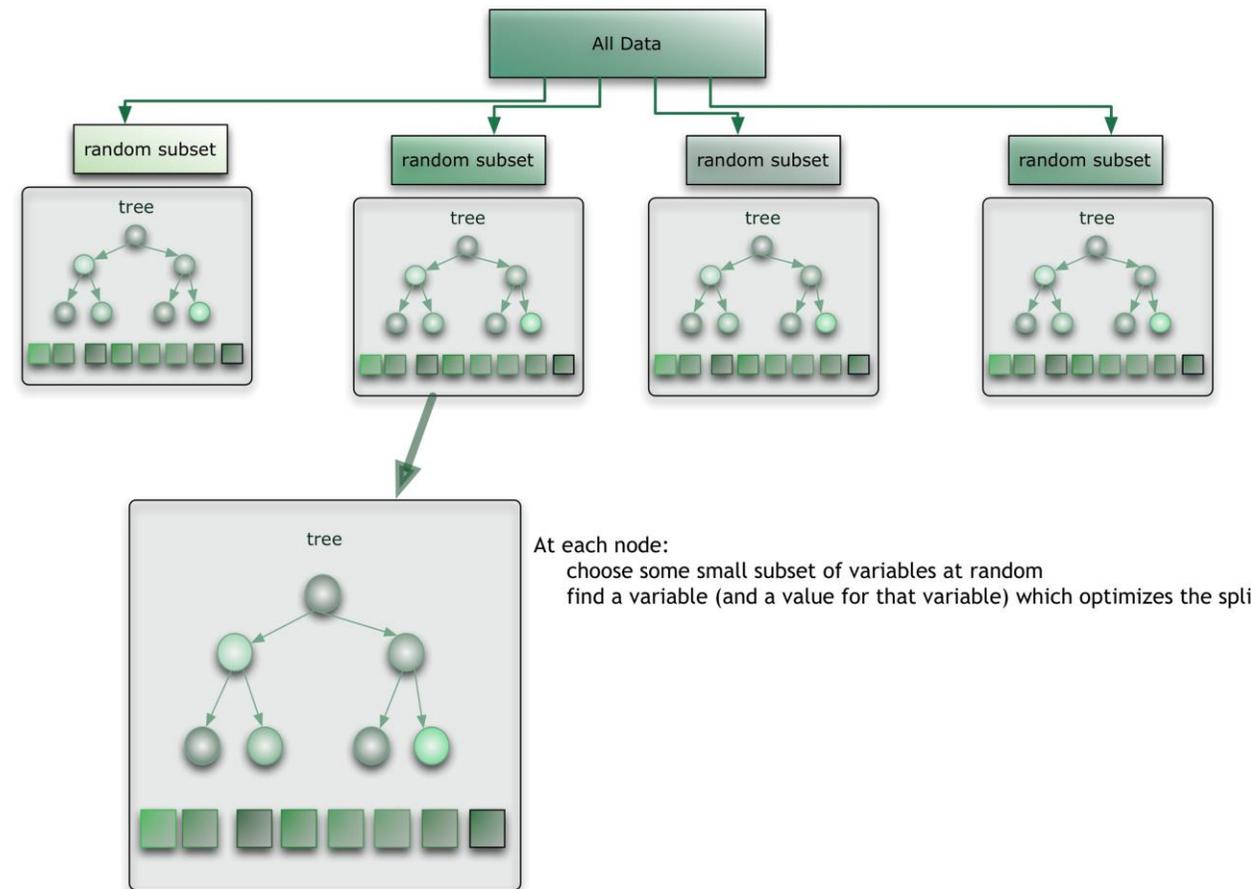
ALEXEJ SHABAS

Gliederung

1. Random Forest
 1. Vorteile/Nachteile
2. Was bedeutet genau „Non-Time-Series-based“
3. Mögliche Features
4. Quellen

Random Forest

- Forest: mehrere DTs
- Random:
 - Zufällige Teilmenge
 - Zufällige Features pro Knoten
 - bestes Feature aus der Teilmenge für Split
- Entscheidung über „Abstimmung“



Random Forest – Vorteile/Nachteile

■ Vorteile:

- Einfach zu verwenden
- wenig Overfitting
- Höhere Genauigkeit als bei DT
- Parallel trainierbar

■ Nachteile:

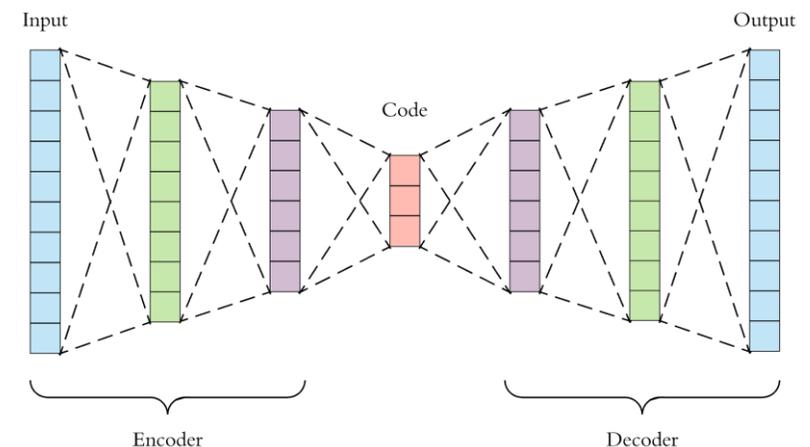
- Nicht leicht zu interpretieren (Black Box)
- Rechenintensiv
- Inputparameter (Anzahl der DTs und Größe der Featureteilmenge pro Knoten)

Was bedeutet genau „Non-Time-Series-based“

- Temporale Abhängigkeit wird nicht explizit modelliert
- Jeder Zeitpunkt einer Zeitreihe wird als ein eigenständiges Feature interpretiert
- → Feature-Engineering nötig

Mögliche Features

- Gegebene Zeitreihen + Ableitung neuer Zeitreihen (Vegetation Index)
 - Jeder Zeitpunkt als eigenständiges Feature (Binning für Feature Reduktion)
- Ableitung statistischer Werte aus den Zeitreihen (davor Noise Reduktion z.B. Moving Average)
 - Durchschnitt, Min, Max, Range, Standard Abweichung, ...
 - 100te Features (TSFRESH Framework)
 - Aufteilung der Zeitreihe in Saisons → statistische Werte pro Saison
- Autoencoder für Featureextraktion
 - Zeitreihen normalisieren
 - Fully Connected „Denoising Autoencoder“ trainieren z.B. $f(\text{Zeitreihe}) = \text{movingAverage}(\text{Zeitreihe})$
 - Mittlerer Layer: Features für Random Forest
- → Experimentieren:
 - Anzahl der Zeitreihen (Zeitreihen mit vielen Missingvalues ignorieren)
 - Anzahl der statistischen Features
 - Nur statistische Features vs nur Autoencoder-Features
 - Kombination der statistischen Features mit Autoencoder-Features
 - PCA auf Featurevektoren (Featurereduktion)



Mögliche Features

- Mögliche Vegetation Indizes:

- $NDVI = \frac{(NIR-RED)}{(NIR+RED)}$

- $EVI2 = 2.5 * \frac{(NIR-RED)}{NIR+(2.4*RED)+1}$ (Alternative zu EVI ohne BLUE-Channel)

- $MSAVI2 = \frac{(2*NIR+1 - \sqrt{(2*NIR+1)^2 - 8*(NIR-RED)})}{2}$

- $SAVI = \frac{(NIR-RED)}{(NIR+RED+0.5)} * 1.5$

- $OSAVI = \frac{(NIR-RED)}{(NIR+RED+0.16)}$

Quellen

- <http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics>
- Gómez, Cristina, Joanne C. White, and Michael A. Wulder. "Optical remotely sensed time series data for land cover classification: A review." ISPRS Journal of Photogrammetry and Remote Sensing 116 (2016): 55-72.
- Jiang, Zhangyan & Huete, Alfredo & Didan, K & Miura, Tomoaki. (2008). Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment. 112. 3833-3845. 10.1016/j.rse.2008.06.006.
- <https://medium.com/regen-network/remote-sensing-indices-389153e3d947>
- <https://github.com/blue-yonder/tsfresh>