

Identifying forest change processes using MODIS EVI timeseries data and random forest classification. An implementation towards an operational forest monitoring

Clausnitzer, Siegfried, Christian Hüttich, Jonas Eberle, Christiane Schmallius
Uni Jena

Das vorrangige Ziel meiner Master-Abschlussarbeit war die Klassifizierung von Waldveränderungsprozesse anhand von MODIS-EVI-Zeitserien (Terra, 13 Jahre, 23 Zeitschritte pro Jahr). Als zweites Ziel galt es, das Verfahren als operationalen Dienst in einem Webportal anzubieten. Drei Veränderungsprozesse standen dabei im Fokus: Holzeinschlag, Feuer und Windwurf.

Zur Entwicklung des Verfahrens standen 50 Zeitserien je Klasse zur Verfügung. Die Vorauswahl erfolgte hierbei mittels BFAST, einem R-Paket zur Detektion von Umbrüchen (d. h. Veränderungsprozessen) in vorwiegend phenologischen Zeitserien. Die Charakterisierung jeder Zeitserie erfolgte zum einen mittels der Differenz simpler statistischer Maße, wie dem Mittelwert oder der Standardabweichung. Zum anderen wurde die Software TIMESAT eingesetzt, um konkrete phenologische Charakteristika, wie Länge oder Amplitude je Vegetationsperiode, zu errechnen. In allen Fällen wurde die Differenz zwischen nach und vor dem Umbruch gebildet (Charakteristik_nach_Umbruch - Charakteristik_vor_Umbruch). TIMESAT bietet immens viele Konfigurationsmöglichkeiten der einzelnen Parameter, woraus ein neues Problem wuchs: Welche Parameter sollte man verwenden? Welche Parameter sind wichtig für die Klassifikation, welche nicht? Die Befürchtung lag nahe, das der manuelle Ausschluss von Parametern mit Sicherheit auch einen Verlust an Information darstellen würde.

Eine große Datenlage führt unweigerlich zu der Thematik "Big Data", welche sich vornehmlich mit der Extraktion von Wissen aus großen Datenmengen beschäftigt. Ein bekanntes Verfahren in diesem Feld ist Random Forest, eigentlich ein Klassifikationsverfahren, was ganz nebenbei Informationen aus der Datenlage generiert. Hierbei wird eine Vielzahl von Entscheidungsbäumen, sprich ein Wald bzw. Forest, dazu benutzt, um eine Klasse zu wählen. Jeder einzelner Baum wird mit einem zufällig (Random) gewählten Test-Zeitserien trainiert, mit einem ebenso zufällig gewählten Teilmenge der zur Verfügung stehenden Parameter. Für jeden einzelnen Parameter kann dann festgestellt werden, wie groß dessen Einfluss auf die korrekte Klassifikation des Test-Datensatzes (Zeitreihe) ist, genannt „Importance“.

Im Falle meiner Arbeit habe ich 15 000 Bäume generiert, um sicherzustellen, dass jeder Parameter und jeder Testdatensatz zum Training mehrmals verwendet wurde. Das sichert die Plausibilität des Verfahrens. Der so generierte „Random Forest“ kann dann zur Klassifikation verwendet werden. Der Wichtigkeit der Variablen („Importance“) kann dann dazu verwendet werden, um Variablen mit wenig Einfluss für das Klassifikationsergebnis zu erkennen und aus dem Klassifikationsverfahren zu entfernen, indem ein neuer

Random Forest generiert wird. In meinen Fall wurde der Klassifikator einmal verfeinert, indem eine Untergrenze für die „Importance“ gesetzt wurde. Der so verfeinerte Random Forest diente dann zur Klassifikation.

Das Verfahren erzielt momentan eine Genauigkeit („Overall Accuracy“) von 92 %. Des weiteren kann aufgezeigt werden, welche Charakteristika den größten Einfluss haben. Dabei zeigen die einfachen statistischen Maße einen sehr großen Einfluss. Der Einfluss vieler mittels TIMESAT generierten Maße ist dahingegen recht gering. Über dies hinaus, lässt das Klassifikationsergebnis Aussagen zur den zugrunde liegenden Prozessen hinter Waldveränderung zu.

Das Verfahren zeigt sehr großes Potential, mit den immensen Datenmengen in der Fernerkundung umzugehen. Dennoch zeigt die Anwendung auf andere Gebiet außerhalb der Testflächen erhebliche Mängel auf. Der Einbezug von mehr Datenquellen (auch außerhalb der Fernerkundung) versprechen eine erhebliche Verbesserung in der Robustheit. Des weiteren ist auch die Einbeziehung von weiteren Datensätzen unabdingbar, vor allem um die räumliche und zeitliche Variabilität von Waldveränderungsprozessen zu erfassen.