

Verbesserung von Prognosen für Restaurants mittels Boosting-Methoden

Das Ziel der Arbeit ist, verschiedene Modelle zur Verbesserung von Prognosen für Restaurants zu beurteilen und eine Empfehlung abzuleiten.

Als Ausgangspunkt gilt ein Prognosetool der Prognolite GmbH, welche Dienstleistungen im Bereich der Informationstechnologie, insbesondere im Datenmanagement und der Visualisierung und dem Verkauf von damit zusammenhängenden Waren, erbringt. Dazu gehört ein Prognosetool für die Anzahl der verkauften Menus pro Viertelstunde für die Eidgenössische Technische Hochschule (ETH) Zürich. Diese Anzahl wurde bereits mit dem Random Forest prognostiziert. Nun möchte die Prognolite GmbH herausfinden, ob ihr bestehendes Prognosetool mit dem neusten und populärsten Boostingverfahren weiter verbessert werden kann.

Zur Analyse und Beurteilung der allfälligen Verbesserung der bestehenden Prognose werden in dieser Bachelorarbeit zwei Boosting-Methoden angewendet. Zum einen das Generalized Boosted Regression Model, welches das R Package «gbm» verwendet. Der Algorithmus ist fest implementiert und es können dadurch nur wenige Parameter angepasst, respektive getunt werden. Die Verlustfunktion (die quadratische Abweichung) kann nur aus einer bestimmten Anzahl von Funktionen gewählt werden. Zum anderen das Extreme Gradient Boosting, welches das R Package «xgboost» benutzt. In diesem Modell ist der Algorithmus im Package bereits implementiert, lässt aber viel Spielraum, ihn anzupassen. Beispielsweise kann die Verlustfunktion individuell definiert werden. Da die Prognosen schlussendlich mit dem Prozentsatz des absoluten Fehlers der Prognose (MAPE) verglichen werden, wird die MAPE-Funktion respektive eine quadratische Ersatzfunktion als Verlustfunktion definiert.

Die Qualität der verschiedenen Modelle wurde mit der Residuen-Analyse untersucht. Summarisch lässt sich festhalten, dass die Residuen mehrheitlich unauffällig sind. Einige Ansatzpunkte zur weiteren Verbesserung des aktuellen Prognosetools der Prognolite GmbH konnten jedoch identifiziert werden:

- Das Extreme Gradient Boosting-Modell ist im Vergleich zum aktuellen Random Forest-Modell leicht besser, dies gilt insbesondere in bestimmten Auswertungsbereichen (90 %- und 95 %-Quantil). Das Generalized Boosted Regression Model weist hingegen diese Vorteile nicht auf.
- Die Analyse- und Auswertungsansätze könnten bei allen drei Modellen optimiert werden. Beispielsweise mit zusätzlichen erklärenden Variablen oder beim Ansatz, die Version des Prognosetools mit den Prognosen pro Tag zu wählen.

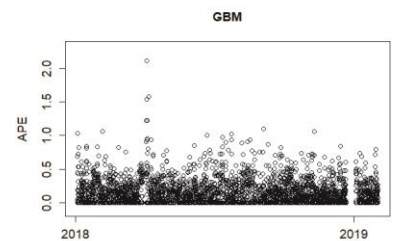


Diplomandin

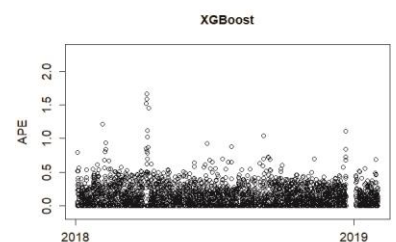
Tamira Franceschetti

Dozent

Thoralf Mildenerger



Absoluter prozentualer Fehler der
Prognose des Generalized Boosted
Regression Model



Absoluter prozentualer Fehler der
Prognose des Extreme Gradient
Boosting