

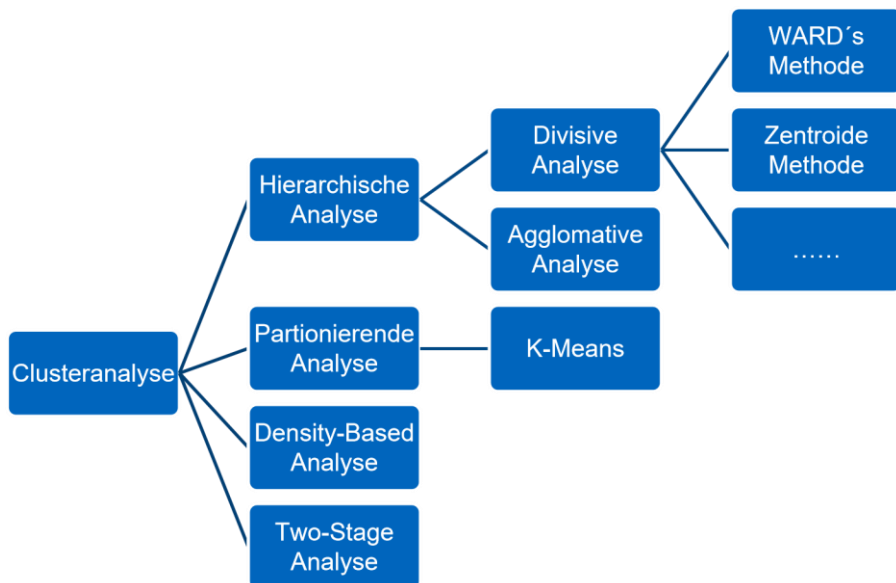
Analyse der Nachfragecluster des Verkehrsflusses auf einem Autobahnabschnitt mithilfe von maschinellem Lernen

Bachelor's Thesis von Huiling Jiang

Mentor(in/innen/en):

Dr.-Ing. Lisa Kessler

Barbara Metzger



Insgesamt analysiert die Arbeit Verkehrsdaten der deutschen Autobahn A9 unter Verwendung von Methoden des maschinellen Lernens, einschließlich der Verwendung von Clustering-Algorithmen und Python-Programmierung.

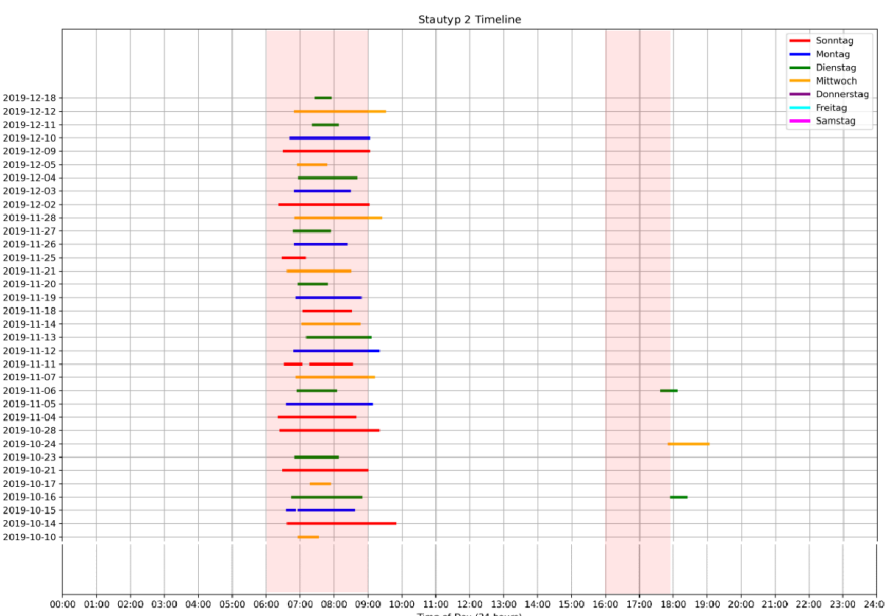
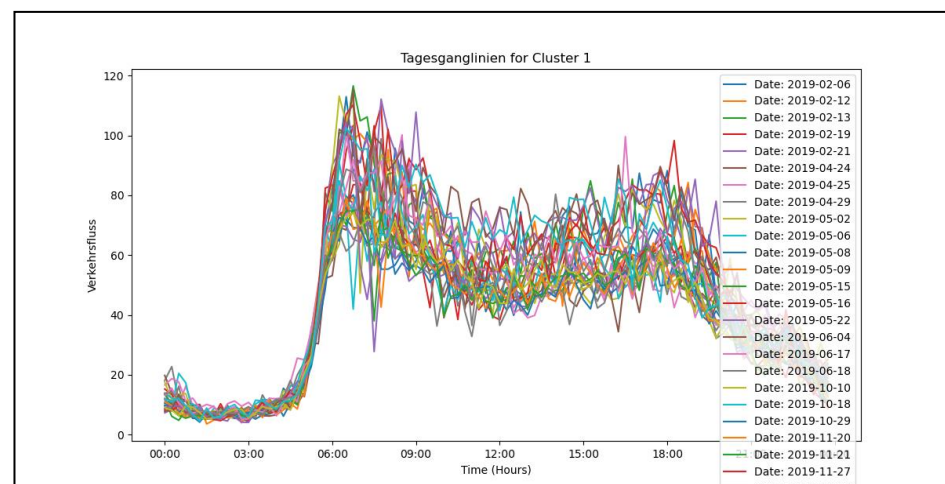
Die Clusteranalyse wird hauptsächlich in vier Kategorien unterteilt: hierarchische Analyse, partitionierende Analyse, dichtebasierte Analyse und zweistufige Analyse, die mehrere Methoden kombiniert. Hierarchisches Clustering kann auch unterteilt werden in Divisive Analyse und Agglomerative Analyse unterteilt werden. Die partitionierende Analyse ist vor allem für die k-Means Methode bekannt. Dies ist eine sehr weit verbreitete Clustermethode.

Bei der Analyse der Daten von Verkehrsfluss habe ich das hierarchische Clustering als Hauptclustermethode gewählt. Dies liegt daran, dass hierarchisches Clustering anwendbar für Zeitreihen Daten ist und kein Anfangswert erforderlich ist. Da es sich bei der Ganglinie um Zeitreihendaten handelt, benötige ich DTW als Ähnlichkeitsmetrik.

Für den ganzen Datensätzen ist eine Vorklassifizierung sinnvoll, wobei Arbeitstagen und Wochenenden im Voraus zu gruppieren.

Das Ergebnis von Gruppe des Wochenendes zeigt, dass in jedem Cluster erhebliche Abweichungen vorhanden sind und die Überlappung in diesem Bereich nicht sehr hoch ist. Das liegt wahrscheinlich daran, dass die Leute am Wochenende freier und vielfältiger in ihrer Gestaltung sind. Sie sind nicht an feste Arbeitszeiten oder Schulzeiten gebunden.

Im Gegensatz dazu ist es bei der Arbeitstagen Gruppe zu erkennen, dass die Überlappung der Linien innerhalb der Arbeitstags-Cluster sehr hoch ist und diese Zeitreihen sehr ähnliche Merkmale aufweisen. Die überwiegende Mehrheit der übrigen Cluster weist eine sehr ähnliche Tendenz auf. Wie in der Abbildung zu sehen ist, weisen beispielsweise in Cluster 1 an Arbeitstagen eine Tendenz mit einer sehr ausgeprägten Spitze kurz vor 6:00 Uhr und einer zweiten Spitze um 18:00 Uhr auf. Spezielle Cluster weist Feiertagen auf.



Im zweiten Teil wird Stautypen analysiert. Das Beispiel im Diagramm ist der Megastau. Bei den Megastau-Daten habe ich ein klares Muster gefunden. Fast der gesamte Megastau tritt während der morgendlichen Rushhour an Wochentagen auf, also zwischen 6 und 9 Uhr morgens. Daraus lässt sich ableiten, dass das Auftreten von Megastau sehr eng damit zusammenhängt, ob es sich um die morgendliche Rushhour an Werktagen handelt oder nicht. Das ist ein sehr bemerkenswertes Nachfragecluster. Der Grund dafür ist verständlich, denn während der morgendlichen Rushhour an Werktagen haben die Menschen ähnliche Pendelzeiten, und die meisten Menschen haben in dieser Zeit Pendelbedarf, was den Verkehrsfluss erhöht und viele Staus verursacht. Dieses Nachfragebündel ist sehr regelmäßig und repräsentativ, da die Pendelzeiten feststeht und kaum von anderen Faktoren wie dem Wetter beeinflusst wird.

Im anderen Saarten weist kein deutliche Eigenschaft auf.