

Clustering von Fahrzeugenkern basierend auf Telematikdaten

Eine Versicherung hat einen Drive Recorder eingesetzt, um das Fahrverhalten von Junglenkern aufzuzeichnen. Die Installation des Drive Recorders geschah dabei auf freiwilliger Basis. Die Versicherung kann die Daten analysieren und Risikoprävention betreiben, um die Schadenfälle zu minimieren. Die Kunden konnten dabei je nach Fahrverhalten von verbilligten Prämien profitieren. In der vorliegenden Arbeit werden die Junglenker aufgrund ihrer Fahrzeugnutzung gruppiert, um deren Unterschiede zu analysieren und das Schadenrisiko der Gruppen abzuschätzen.

Dazu werden mit einem Geocustering-Algorithmus die Start- und Endpunkte aller Fahrten eines Kunden räumlich geclustert. Selten frequentierte Orte werden dabei als Ausreisser ausgewiesen.

Anschliessend wird die Anzahl der Fahrten von und in die verschiedenen Cluster gezählt. In Kombination mit der Laufzeit des Drive Recorders kann ausgewiesen werden, ob Verbindungen täglich, wöchentlich, monatlich oder noch seltener gefahren werden.

Für das Clustering der Kunden werden als zusätzliche Variablen die total gefahrene Distanz, die Startzeit der einzelnen Fahrten und die Unterscheidung zwischen Wochentag oder Wochenende für jeden Kunden verwendet. Diese Variablen werden anteilmässig pro Häufigkeit berechnet, sodass eine Differenzierung von beispielsweise täglichen Nachtfahrten sowie wöchentlichen Wochenendfahrten gemacht werden kann. Zum Clustern werden zwei verschiedene Algorithmen verwendet. Der CLARA-Algorithmus ist ein partitionierendes Verfahren und erstellt explizit getrennte Cluster. Der hierarchische Algorithmus produziert verschachtelte Cluster, welche als Dendrogramm dargestellt werden. Beim CLARA-Algorithmus resultieren 12 Clustergruppen, das hierarchische Clustering liefert bei 10 Gruppen optimale Resultate. Zum Schluss werden die Kundencluster mit den Schadendaten der jeweiligen Kunden verknüpft, um die Cluster basierend auf der durchschnittlichen jährlichen Schadensumme pro Kunde zu vergleichen.

Lenkende, welche viele Kilometer, insbesondere am Wochenende oder während der Nacht, fahren, haben eine tendenziell höhere durchschnittliche Schadensumme pro Jahr. Kunden, welche hingegen gesamthaft wenig fahren und ihr Fahrzeug weniger auf unbekanntem Strecken benutzen, verursachen eine tiefere durchschnittliche Schadensumme pro Jahr. Für detailliertere Aussagen ist der Beobachtungszeitraum der Drive Recorder Daten jedoch zu kurz, da nur wenige hundert Schadenfälle vorliegen.



Diplomierende
Robin Jezler
Raffael Pfister

Dozierende
Martin Frey
Thoralf Mildenerger



Abbildung 1 zeigt einen Ausschnitt der Start- und Endpunkte von Fahrten eines Mitarbeiters der Versicherung. Die Daten werden anhand der Geo-Koordinaten mit dem Geocustering-Algorithmus DBSCAN geclustert



Abbildung 2 zeigt die Punkte aus der Abbildung 1 nach dem Clustern. Mit den Farben werden dabei die unterschiedlichen Cluster visualisiert. Graue Punkte zeigen Ausreisser.