



Forschungsbericht 2015



pixelio.de: Petra Engeljehringer

Risikomodellierung in der Kraftfahrthaftpflichtversicherung

Versicherungsverträge bieten einen Schutz gegen potentielle Risiken, jedoch ist unsicher, ob und in welcher Höhe Schäden tatsächlich eintreten. In der Risikomodellierung ist es deswegen von hoher Relevanz, vorhandene Daten zu nutzen und aus ihnen zu lernen. Dafür stehen statistische Verfahren zur Verfügung, die zur Prädiktion von Schadenaufwänden verwendet werden können.

Projektleitung

Prof. Dr. Viktor Sandor
Fakultät für angewandte
Natur- und Geistes-
wissenschaften

Projektmitarbeiter

- Daniela Sebald, B. Sc.
- Dipl.-Math. Monika Sussmann
- Prof. Dr. Ulrich Wellisch
Fakultät für angewandte
Natur- und Geisteswis-
sensschaften

Auftraggeber

Die Bayerische

Projektlaufzeit

Dez. 2014 – Juni 2015

In Zusammenarbeit mit dem Versicherungsunternehmen „Die Bayerische“ wird untersucht, ob die Risikomodellierung im Bereich der Krafthaftpflichtversicherung mit der „open source“ Statistiksoftware R umsetzbar ist. Hierzu werden Schadendaten aus den letzten Jahren analysiert, die anonymisierte Informationen zum Kunden, zum Fahrzeug und Versicherungsvertrag sowie zu entstandenen

Schäden enthalten, z.B. Alter des Versicherungsnehmers, Typklasse und Fahrleistung. Ziel ist es dabei, die wichtigsten Merkmale zu selektieren, die die aufgetretenen Schäden, d.h. Schadendurchschnitt und Schadenfrequenz, bestmöglich abbilden.

Zum Projektbeginn wurden die Schadendaten aus den letzten Jahren an die Projektgruppe übermittelt. Um richtige und schlüssige Prognosen

Überprüfung
der Daten

Deskripte
Analysen

Modellierung

Überprüfung
der Ergebnisse

zu erzielen, ist eine korrekte Datenbasis notwendig. Hierfür wird überprüft, ob die einzelnen Merkmale ausreichend gefüllt sind und ob unrealistische Werte enthalten sind. Falsche Werte werden entfernt, um die nachfolgende Modellierung nicht zu verfälschen.

Anschließend werden erste Zusammenhängestrukturen und Bestandsverteilungen mittels deskriptiver und explorativer Analysen aufgezeigt. Folgende Fragestellungen sind dabei von Interesse. Welche Bestandsverteilung liegt vor? Wie steht das Merkmal mit den Zielgrößen Schadenfrequenz und Schadendurchschnitt in Zusammenhang? Anhand von Visualisierungen und Kennzahlen werden Gruppierungen vorgenommen. Ausprägungen, die sich nicht unterscheiden, werden zu Gruppen zusammengefasst und nicht relevante Merkmale aus der weiteren Betrachtung ausgeschlossen.

In Abbildung 1 ist die Bestandsverteilung für ein Merkmal anhand eines Säulendiagramms dargestellt. Zusätzlich enthält die Grafik die gemittelte Schadenfrequenz je Ausprägung (blaue Linie).

Es weisen etwa 12 Prozent des Bestandes die Merkmalsausprägung 16 bei einer Frequenz von 5,8 Schäden je 100 Fahrzeugen auf.

Nach der Datenaufbereitung wird mit der Risikomodellierung begonnen. Als Zielgrößen werden der Schadendurchschnitt und die Schadenfrequenz betrachtet. Mittels generalisierter linearer Modelle und semi-automatisierten Variablenselektionsverfahren werden auf einem Teildatensatz (Lerndatensatz) die Variablen bzw. die Variablenkombination bestimmt, die mit den Responsevariablen in einem signifikanten Zusammenhang stehen. Die Güte dieser Modelle wird anhand unabhängiger Validierungs- und Testdaten beurteilt. Zur Darstellung der Anpassungsgüte und der Trenneigenschaft der Modelle eignen sich sogenannte Liftplots. In Abbildung 2 ist dieser Plot für die Zielgröße Schadenfrequenz dargestellt. Es entsteht durch das Modell eine Trennung von 13 Prozent zu ca. 3 Prozent. Die blaue Linie ist die gemittelte beobachtete Schadenfrequenz und die rote Linie stellt die dazugehörige Anpassung durch das Modell dar. Es ist somit eine gute Anpassung der Daten erzielt worden. Mittels Gütekriterien und Tests können ebenso quantitative Überprüfungen erfolgen.

Die selektierten Variablen für die Schadenfrequenz und den Schadendurchschnitt werden anschließend in ein multiplikatives Modell zur Vorhersage des Schadenbedarfes überführt, das sogenannte Risikomodell.

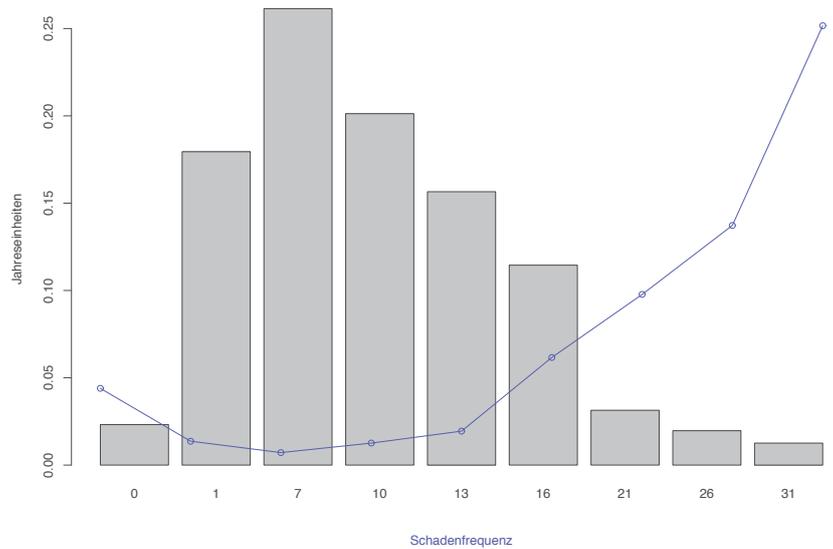


Abb. 1: Bestandsverteilung und entsprechende Schadenfrequenz

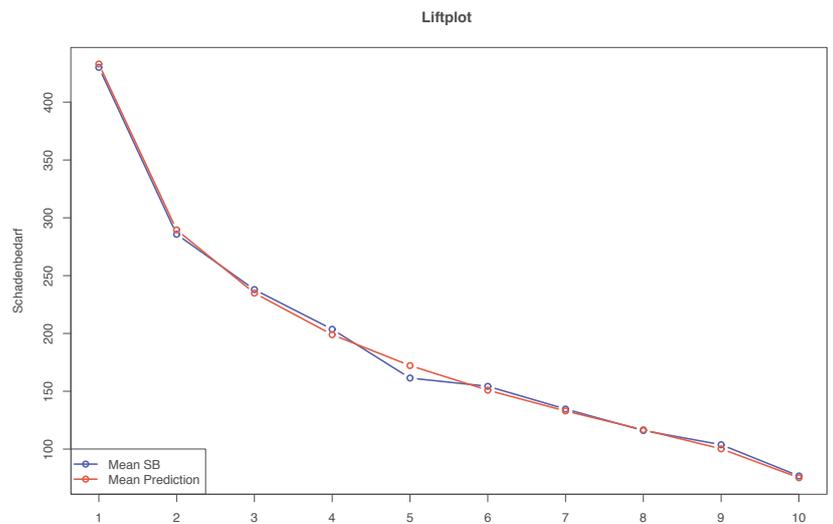


Abb. 2: Liftplot für Modellanpassung Schadenfrequenz

Dieses wird in Versicherungsunternehmen zur Abbildung der Risikokomponente von Verträgen verwendet. Unter Berücksichtigung von Kosten und Marketingkonzepten entsteht hieraus das Tarifmodell, womit schließlich die Versicherungsprämie eines künftigen Versicherungsvertrages bestimmt wird.

Besonderheit dieses Forschungsprojektes ist die Umsetzung der Risikomodellierung in der frei verfügbaren Statistiksoftware R. Häufig stehen Versicherungsunternehmen teure Tools zur Risikomodellierung zur Verfügung. Es wird jedoch gezeigt, dass identische Ergebnisse auch mit R erzielt werden können. Das Statistikprogramm beinhaltet eine

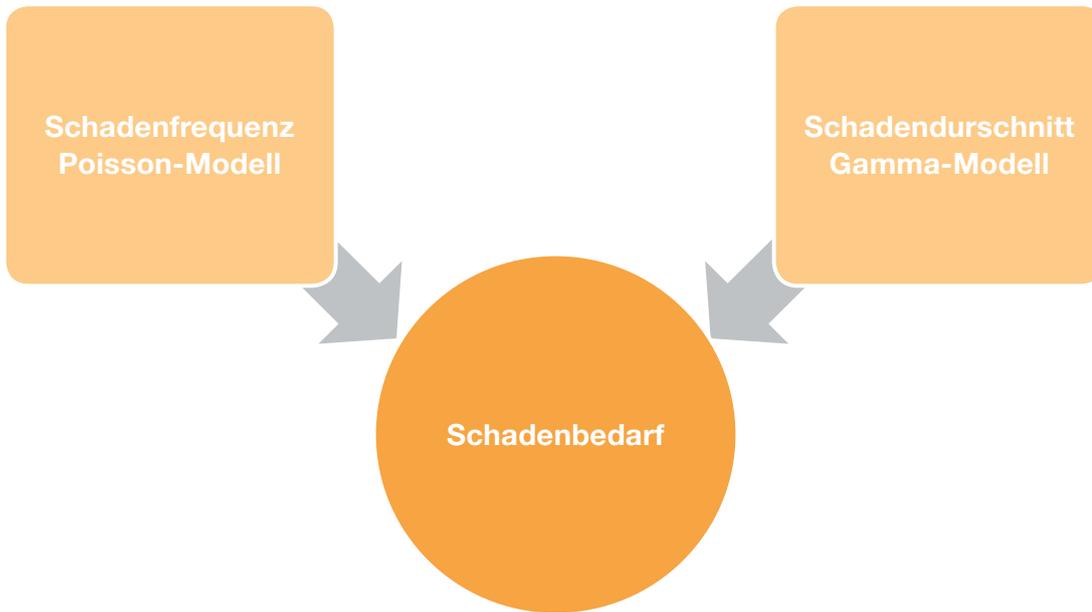


Abb. 3: Statistische Grundlagen

Vielzahl von Funktionen und Anwendungen, die einfach eingebunden werden können. Zusätzlich besteht auch die Möglichkeit eigene Anwendungen und Funktionen zu programmieren.

Statistische Grundlagen

Für die Modellierung der Schadenfrequenz und des Schadendurchschnittes werden Verteilungsannahmen getroffen, die in den generalisierten linearen Modellen berücksichtigt werden. Für die Schadenanzahl wird eine Poissonverteilung unter Berücksichtigung der Versicherungsdauer angenommen. Für die Schadendurchschnitte wird eine Gammaverteilung zu Grunde gelegt. Die Ergebnisse aus beiden Modellen werden anschließend multiplikativ zusammengeführt, um den Schadenbedarf abzubilden.

$$\begin{aligned}
 &\text{Schadenbedarf} \\
 &= \text{Schadenfrequenz} * \text{Schadendurchschnitt} \\
 &= \frac{\sum \text{Schadenanzahl}}{\sum \text{Jahreseinheiten}} * \frac{\sum \text{Aufwand}}{\sum \text{Schadenanzahl}}
 \end{aligned}$$

Nach Bestimmung der Einflussparameter für die einzelnen Merkmale können über einen additiven Zusammenhang die Responsevariablen geschätzt werden.

Zusammenfassung und Ausblick

Das Forschungsprojekt zeigt, dass die Modellierung von Schäden in der Versicherungsbranche gut mit der „open source“ Software R umsetzbar

ist. Es können komplexe Strukturen abgebildet werden und deskriptive Analysen erfolgen. Der Umgang mit großen Datenmengen ist teilweise problematisch und zeitintensiv. Die Variablenselektionsalgorithmen und iterativen Optimierungsalgorithmen sind sehr rechenintensiv. Hierfür sind ausführlichere Untersuchungen im Bereich sogenannter Big-Data Anwendungen notwendig. Zusätzlich ist die Untersuchung der Anwendbarkeit von weiterführenden statistischen Modellen von Interesse.

**Viktor Sandor, Daniela Sebald,
Monika Sussmann, Ulrich Wellisch**

Literatur

- [1] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [2] DAV, Arbeitsgruppe Tarifierungsmethodik. (2015). Aktuarielle Methoden der Tarifgestaltung in der Schaden-/Unfallversicherung. VWW Karlsruhe.