

Data Mining in und mit Zeitreihen (SAS und die BILD-Zeitung, ein Beispiel aus der Vertriebsanalyse im Verlagswesen)

Stefan Jelinek, Armin Penneckendorf, Michael Wigbers, Walter Baum
Axel Springer Verlag AG, Hamburg – Logistik / Vertrieb-Marktanalyse

Basierend auf mehrjährigen Tageszeitreihen des Bild-Zeitungsverkaufes von bis zu 120 000 Einzelhändlern wird durch die Kopplung der SAS-Module BASE, ETS und STAT, in Verbindung mit einer webbasierten Ausgabe, ein vollautomatisches Zeitreihen-Mining-Modell vorgestellt, das wöchentlich die komplexen Marktstrukturen des Bild-Zeitungsverkaufes beschreibt. Der Schwerpunkt des anwendungsorientierten Vortrages liegt dabei neben der Darstellung der performanten Kopplung der einzelnen SAS-Module auf der Diskussion der Möglichkeiten und Gefahren, die sich aus der Verbindung von Zeitreiheninformationen mit räumlich strukturellen Informationen über die Verkaufsstellen und Märkte ergeben.

Die Verkäufe der Bild-Zeitung werden tagesbasiert erfasst und stehen in einem laufenden Prozess mit bis zu drei Wochen Verzögerung zu Analyse Zwecken zur Verfügung. Weiterhin sind umfangreiche Datensätze zur Beschreibung der Einzelhändler und der auf statistischen Bezirken basierenden Raumeinheiten (ca. 32 000), in denen die Händler geokodiert sind, vorhanden.

Aus den Verkaufszeitreihen werden über die Prozeduren EXPAND und ARIMA Trend und Saisonkoeffizienten ermittelt und den Strukturvariablen der Händler gegenübergestellt. Zur weiteren Beschreibung werden spezielle Merkmale der jeweiligen Zeitreihe (z.B. Verkaufsklassen, Streuung, Ausverkäufe, Samstagsindex etc.) extrahiert. Ausgewählte Zeitreihenmerkmale werden dabei mit Strukturvariablen (z.B. Geschäftsarten) zu logischen Komplextypen (Funktionstypen) zusammengeführt. Ergebnis ist dann eine Matrix aus drei Trendgrößen (langfristiger-, mittelfristiger- und Vorjahrestrend), denen ca. 25 Merkmalsausprägungen der Einzelhändler gegenüberstehen (Klassifikationsvariablen). Die Trendparameter werden pro Wochentag ermittelt, somit stehen jede Woche ca. 600 000 Fälle zur Verfügung.

Diese Matrix wird durch einen entropiebasierten Entscheidungsbaum jeweils für die drei Trendparameter segmentiert und die Gruppen über die Prozedur ANOVA statistisch bewertet. Die Darstellung erfolgt über eine webbasierte Oberfläche im Vertriebs-Informations-Portal des Verlages.

In einem Ausblick wird gezeigt, wie über eine automatische Mustererkennung versucht wird, die komplexe und teilweise unübersichtliche Struktur des Entscheidungsbaumes zu Kernaussagen zusammenzufassen. Weiterhin wird diskutiert, welche Probleme sich bei dem Versuch ergeben, Punktinformationen (Einzelhändler) mit räumlichen Informationen (soziodemographische Struktur) zu koppeln und wie angestrebt wird, dieses Dilemma zu umgehen.