

Angewandte Statistik zur Analyse von Nutzerdaten in der Energiewirtschaft

Detlef E. Schröder

Abstract: Individualisierung ist ein wesentliches Schlagwort der Energiewirtschaft und das nicht nur in der Tarifentwicklung, sondern auch in der Prognose und Energieerzeugung. Dazu werden Nutzerdaten erhoben und für die Prognose verwendet. Bei den dabei anfallenden Datenmengen ist eine effiziente Verarbeitung und effektives statistisches Ergebnis entscheidend. Beides soll anhand von Beispielen und den Möglichkeiten der Oracle Datenbank verdeutlicht werden. Die freie Statistische Analyseumgebung R spielt in diesem Zusammenhang eine immer bedeutendere Rolle und soll hier auch zur Anwendung kommen.

Keywords: Prognose, Datenbank, Analysemodelle, Nutzerverhalten, R, Data Mining, Smart Meter, Big Data, Oracle

1 Individualisierung in der Energiewirtschaft

Die Energiewirtschaft ist von großen Veränderungen betroffen, die sowohl die Erzeugung, als auch den Verbrauch von Energie betreffen. Die Produktion ist individualisiert, in dem jeder Verbraucher mit einer entsprechenden Photovoltaik, Wind oder Kraft-Wärme-Kopplung Anlage auch zum Produzenten werden kann. Der Markteintritt ist in diesem Bereich sehr einfach geworden und die Hürden geschrumpft. Dem gegenüber ist die Produktion zur Versorgungssicherheit in seiner Komplexität gestiegen.

Auf Grund dieser Effekte wird es für die Energiewirtschaft immer entscheidender auch die Verbraucher Seite zu individualisieren. Dies kann über die Preis und Vertragsbildung geschehen. Dazu ist es aber notwendig aussagekräftige Modelle zur Verbrauchsprognose zu erstellen und entsprechend schnell zur Anwendung bringen zu können. Die Prognosen Qualität, Ex-ante, lässt sich wesentlich über die erklärenden Variablen steuern und die Funktionsform der Beziehungen [WL24]. Dank den Smart Meter Messwerten und der Big Data Technologie ist die Zahl der erklärenden Variablen enorm gestiegen. Diese mit den in relationaler Struktur vorhandenen Größen in Verbindung zur Anwendung zu bringen stellt allerdings eine Herausforderung dar. Oracle stellt mit der Datenbank, der zur Verfügung stehenden Statistik und der Big Data Ergänzung die Mittel für diese Prognosen zur Verfügung.

2 Statistik mittels SQL

Für Oracle bietet SQL eine verbindende Sprache zur Verwendung sämtlicher Daten. Diese können in der Form eine relationalen Struktur, dem Ergebnis einer R Funktion oder als Big Data Struktur vorliegen. Alle Formate lassen sich mittels SQL verarbeiten und analysieren. Somit stellt SQL die Basis für die Analyse der Nutzerdaten zur Verfügung.

2.1 Prognose-Erstellung

Die Prognose Erstellung folgt den Regeln und dem Ablauf des Data Mining Prozesses. Fragestellung, Daten identifizieren, Datenqualität sicherstellen, Algorithmus identifizieren und Anwenden. Das Anwenden erfolgt über einfaches SQL, das nach folgendem Muster aufgebaut sein kann:

```
SELECT kundenummer, PREDICTION (FOR verbrauch USING *)  
OVER() AS verbrauchsprognose FROM alledaten WHERE  
bedingungen
```

Im Muster einer Analytischen SQL Funktion wird mit PREDICTION die Data Mining Funktionalität der Datenbank genutzt, um die Vorhersage des verbrauch mit allen Einflussgrößen aus alledaten zu ermitteln. Dabei kann PREDICTION die Oracle SQL eigene Funktion sein, oder eine spezielle R Funktion. alledaten kann dabei relational, Big Data oder das Ergebnis einer anderen R Funktion sein, die Daten als Frame zur Verfügung stellt. Die bedingungen sind dann auf die Daten bezogen und können natürlich in diesem Beispiel vor Allem die Zeitpunkte für die zu berechnenden Verbrauchsmengen betreffen oder natürlich räumlich einschränken.

2.2 Prognose-Anwendung

Aufgrund der einheitlichen Sprache SQL kann jedes SQL fähige Werkzeug diese Prognosen im Dialog verwenden oder die Werte können relational gespeichert werden, pro Kunde ein individuelles Verbrauchsmodell, oder mehrere für unterschiedliche Situationen (tagsüber, abends, Fußballspiel, ...)

Wesentlich ist, das aufgrund der Möglichkeiten der Datenbank keine Einschränkung was Datenmenge oder Performance angeht, vorhanden sind. Somit kann auf ein aufwändiges Sampling verzichtet und die statistische Genauigkeit erhöht werden, da Sampling-Fehler ausgeschlossen sind. Einer Individualisierung steh also nichts im Wege

3 Literaturverzeichnis

[WL24] Wirtschaftslexikon24.com – Stichwort Prognosequalität. [Link](#)