

Statistik im und für das WWW

Christoph Freudenthaler

Universität Hildesheim

Die klassische (mathematische) Statistik beschäftigt sich intensiv mit der Effizienz von Schätzern und Schätzmethoden. Der Grund für diesen Fokus auf Effizienz liegt in den grundsätzlich kleinen Stichprobengrößen, die aus finanzieller (Kosten pro Studienteilnehmer) oder persönlicher (nur wenige Patienten mit seltener Krankheit) Ressourcenknappheit resultieren.

Im digitalen Zeitalter und insbesondere im World-Wide-Web (WWW) ist es jedoch kein Problem Unmengen an Benutzerdaten zu sammeln. Während der Benutzer eines Online-Shops seine Produkte per Click in seinen Warenkorb legt, wird im Hintergrund seine Aktivität mitprotokolliert und die damit verbundene Information de facto kostenlos ermittelt. So entstehen riesige Datenmengen, in denen Effizienz anders definiert ist: Anstelle von maximaler Schätzgenauigkeit und sparsamen Modellen werden Rechenzeit und benötigter Speicherplatz genauso wie automatisierte Modellwahl immer wichtiger.

Dieser Vortrag beschäftigt sich mit neuen Anwendungen im WWW, den damit verbundenen neuen Anforderungen an statistische Methoden und stellt erste erfolgreiche Ansätze vor.