

## Abstract

Die Motivation für diese Dissertation ist dualer Natur: Einerseits ist die Dissertation methodologisch orientiert und entwickelt neue statistische Ansätze und Algorithmen für maschinelles Lernen. Gleichzeitig ist sie praktisch orientiert und fokussiert sich auf den konkreten Anwendungsfall von Produktretouren im Onlinehandel.

Die "data explosion", verursacht durch die Tatsache, dass die Kosten für das Speichern und Prozessieren großer Datenmengen signifikant gesunken sind (Bhimani and Willcocks, 2014), und die neuen Technologien, die daraus resultieren, stellen die größte Diskontinuität für die betriebliche Praxis und betriebswirtschaftliche Forschung seit Entwicklung des Internets dar (Agarwal and Dhar, 2014). Insbesondere die Business Intelligence (BI) wurde als wichtiges Forschungsthema für Praktiker und Akademiker im Bereich der Wirtschaftsinformatik (WI) identifiziert (Chen et al., 2012). Maschinelles Lernen wurde erfolgreich auf eine Reihe von BI-Problemen angewandt, wie zum Beispiel Absatzprognose (Choi et al., 2014; Sun et al., 2008), Prognose von Windstromerzeugung (Wan et al., 2014), Prognose des Krankheitsverlaufs von Patienten eines Krankenhauses (Liu et al., 2015), Identifikation von Betrug (Abbasi et al., 2012) oder Recommender-Systeme (Sahoo et al., 2012). Allerdings gibt es nur wenig Forschung, die sich mit Fragestellungen um maschinelles Lernen mit spezifischen Bezug zu BI befasst: Obwohl existierende Algorithmen teilweise modifiziert werden, um sie auf ein bestimmtes Problem anzupassen (Abbasi et al., 2010; Sahoo et al., 2012), beschränkt sich die WI-Forschung im Allgemeinen darauf, existierende Algorithmen, die für andere Fragestellungen als BI entwickelt wurden, auf BI-Fragestellungen anzuwenden (Abbasi et al., 2010; Sahoo et al., 2012). Das erste wichtige Ziel dieser Dissertation besteht darin, einen Beitrag dazu zu leisten, diese Lücke zu schließen.

Diese Dissertation fokussiert sich auf das wichtige BI-Problem von Produktretouren im Onlinehandel für eine Illustration und praktische Anwendung der vorgeschlagenen Konzepte. Viele Onlinehändler sind nicht profitabel (Rigby, 2014) und Produktretouren sind eine wichtige Ursache für dieses Problem (Grewal et al., 2004). Neben Kostenaspekten sind Produktretouren aus ökologischer Sicht problematisch. In der Logistikforschung ist es weitestgehend Konsens, dass die "letzte Meile" der Zulieferkette, nämlich dann wenn das Produkt an die Haustür des Kunden geliefert wird, am CO<sub>2</sub>-intensivsten ist (Browne et al., 2008; Halldórsson et al., 2010; Song et al., 2009). Werden Produkte retourniert, wird dieser energieintensive Schritt wiederholt, wodurch sich die Nachhaltigkeit und Umweltfreundlichkeit des Geschäftsmodells von Onlinehändlern relativ zum klassischen Vertrieb reduziert. Allerdings können Onlinehändler Produktretouren nicht einfach verbieten, da sie einen wichtigen Teil ihres Geschäftsmodells darstellen: So hat die Möglichkeit, Produkte zu retournieren positive Auswirkungen auf Kundenzufriedenheit (Cassill, 1998), Kaufverhalten (Wood, 2001), künftiges Kaufverhalten (Petersen and Kumar, 2009) und emotionale Reaktionen der Kunden (Suwelack et al., 2011). Ein vielversprechender Ansatz besteht darin, sich auf impulsives und kompulsives (LaRose, 2001) sowie betrügerisches Kaufverhalten zu fokussieren (Speights and Hilinski, 2005; Wachter et al., 2012). In gegenwärtigen akademischen Literatur zu dem Thema gibt es keine solchen Strategien. Die meisten Strategien unterscheiden nicht zwischen gewollten und ungewollten Retouren (Walsh et al., 2014). Das zweite Ziel dieser Dissertation besteht daher darin, die Basis für eine Strategie von Prognose und Intervention zu entwickeln, mit welcher Konsumverhalten mit hoher Retourenwahrscheinlichkeit im Vorfeld erkannt und rechtzeitig interveniert werden kann.

In dieser Dissertation werden mehrere Prognosemodelle entwickelt, auf Basis welcher demonstriert wird, dass die Strategie, unter der Annahme moderat effektiver Interventionsstrategien, erhebliche Kosteneinsparungen mit sich bringt.

## Abstract

The motivation for this dissertation is a dual one: On the hand it is methodological, as it introduces new statistical and machine learning approaches, on the other hand it is practical, as these approaches are applied to provide solutions for and study a real-world business problem, namely that of product returns in online retail.

The “data explosion” enabled by the fact that the costs of storing and processing large amounts of data have decreased significantly (Bhimani and Willcocks, 2014) and the new technologies resulting from this trend constitute the biggest disruption in business practise and business research since the rise of the internet (Agarwal and Dhar, 2014). In particular, Business Intelligence (BI) has been identified as an important research topic for both practitioners and academics in the field of Information Systems (IS) (Chen et al., 2012). Machine learning algorithms have been successfully applied to a large variety of BI problems, including sales forecasting (Choi et al., 2014; Sun et al., 2008), forecasting wind power output (Wan et al., 2014), analysis of patient outcome (Liu et al., 2015), fraud detection (Abbasi et al., 2012) or recommender systems (Sahoo et al., 2012). However, very little research is concerned with machine learning issues that are unique to BI: Even though existing machine learning algorithms are occasionally modified for a specific BI problem (Abbasi et al., 2010; Sahoo et al., 2012), IS research in BI as well as BI practice is generally limited to applying existing machine learning approaches and statistical concepts that were originally developed for other domains to specific BI problems (Wu et al., 2008; Chen et al., 2012). One of the two motivations for this dissertation is to close this gap.

This dissertation focuses on the important BI problems of product returns in online retail for an illustration and a practical application of the proposed concepts. Many online retailers fail to be profitable (Rigby, 2014) and product returns have been recognized as a major cause for this problem (Grewal et al., 2004). In addition to being a cost factor for online retailers, product returns are problematic from an environmental point of view: In the logistics literature, it is widely recognized that the “last mile” of the delivery chain, when the product is delivered from the store to the customer’s doorstep, is most CO<sub>2</sub>-intensive (Browne et al., 2008; Halldórsson et al., 2010; Song et al., 2009). Product returns repeat this energy-intensive step, thus decreasing the environmental friendliness of online retail as a business model relative to more traditional forms of retail. However, online retailers cannot simply prohibit product returns, because they are an essential part of their business model: It has been demonstrated that enabling customers to return unwanted products has a positive impact on customer satisfaction (Cassill, 1998), purchase rates (Wood, 2001), future buying behaviour (Petersen and Kumar, 2009) and customers’ emotional responses (Suwelack et al., 2011). A promising approach is to focus on impulsive or even compulsive shopping behaviour (LaRose, 2001) and fraudulent returns (Speights and Hilinski, 2005; Wachter et al., 2012). To date, there are no such strategies in the academic literature on the topic. In fact, most strategies are one-size-fits-all approaches which do not differentiate between wanted and unwanted returns (Walsh et al., 2014). Another motivation for this dissertation is to present the basis for a strategy for handling product returns that addresses the identified shortcoming in the extant literature, namely a strategy of prediction and prevention which identifies consumption patterns associated with a high probability of a product return and intervenes before the transaction even takes place.

This dissertation develops several prediction models that form the basis for such a strategy and show that it is feasible, given moderately effective interventions.