A Conceptual Approach for Optimizing Distribution Logistics using Big Data

Research-in-Progress

Tobias Engel
Technische Universität München
Chair for Information Systems – I17
tobias.engel@in.tum.de

Oleksandr Sadovskyi
Technische Universität München
Chair for Information Systems – I17
o.sadovskyi@tum.de

Markus Böhm
Technische Universität München
Chair for Information Systems – I17
markus.boehm@in.tum.de

Robert Heininger
Technische Universität München
Chair for Information Systems – I17
robert.heininger@in.tum.de

Helmut Krcmar
Technische Universität München
Chair for Information Systems – I17
krcmar@in.tum.de

Abstract

Big data analytics creates new opportunities and potentials in the field of supply chain management (SCM). Specifically, linking inter-firm supply chain processes of two entities such as freight forwarder and final customer were identified as relevant areas for performance improvements. Automatic analysis of data from sources such as mobile equipment, sensor networks, and geospatial devices can significantly improve accuracy of SCM transportation processes; thus contributing to supply chain performance by minimizing delivery attempts, and ensures higher customer satisfaction, as deliveries are carried out when customers are able to receive them. Therefore, we develop a distribution concept targeted at realization of this potential. We chose the online shopping domain for our concept, as current research provides evidence that delivery of goods acquired in online stores often causes multiple delivery attempts, and many firms struggle with providing good service to the customers. Finally, we discuss limitations, implications and possibilities for future research.

Keywords

Big Data, Supply Chain Management, Distribution concept, Actor Network Theory, Online Shopping

Introduction

In the last three years big data has become an important topic in business and academia, and one of the most promising emerging technologies (Manyika et al. 2011; Markl et al. 2013; McAfee and Brynjolfsson 2012). Researchers state that in the upcoming years big data technologies will provide tools and resources that will shape the future of important industries like manufacturing, retail, advertising, telecommunications, transportation and healthcare (Hessman 2013; Kayyali et al. 2013; Manyika et al. 2011; Young and Pollard 2012). Due to currently low proliferation of big data-based business models, every new effective implementation in an organization can potentially offer significant competitive advantage in the form of innovation, cost leadership or operational effectiveness. Potential for effective big data integration exists in almost all sectors, spanning from small non-profit governmental and scientific institutions to large enterprises (Manyika et al. 2011).
Big data is also expected to improve supply chain collaboration and supply chain processes such as inventory management, transportation management and relationship management (Waller and Fawcett 2013). A recent literature review (Sadovskyi et al. 2014) shows that there exists a significant shortage of new theoretical and practical developments in this direction. Existing big data concepts in the field of Supply Chain Management (SCM) attempt to target relevant tasks such as faster tracking and classification of goods (Swaminathan 2012), collecting data for transportation logistics planning and scheduling (Gorodetsky et al. 2012), and data analytics for so-called “health checking” of suppliers and creditors or keeping track of partners’ compliance conditions using web mining (Young and Pollard 2012).

Potentially, big data can be utilized for the forecasting of delivery times and route optimization by taking into account various data such as traffic and weather information or driver characteristics. Other important fields of big data application is the analysis of sensor information for better inventory management and merchandising decisions (Waller and Fawcett 2013). Increasing utilization of GPS-capable equipment together with the further evolvement of sensor networks and the concept of “Internet of Things” offer strong capabilities for big data based automation in the distribution part of the SCM (Ashton 2009). Spatial data, generated by many personal mobile devices and logistics aiding equipment can serve as basis for optimizing SCM processes, resulting in improved customer satisfaction and a decrease of operational costs (Agouris et al. 2011; Becker et al. 2010; Laurila et al. 2012).

There also exists a need for a concept to optimize the routing processes (Shekhar et al. 2012; Valerio et al. 2009; Waller and Fawcett 2013). Such a concept has to allow organizations to link upstream and downstream data on a granular level for each entity such as a customer to improve external transportation processes. This concept can be used to improve supply chain performance, for example in delivery service companies, to ensure personal delivery hand-over at all times and avoid multiple delivery attempts caused by customers’ absence (Fang and Zhang 2005; Weltevreden 2008).

Therefore, this study is concerned with the development of a conceptual approach aimed at improving distribution processes. The approach includes the elicitation of requirements for the concept and the subsequent development of a framework for a big data supported approach to the problem. Furthermore, security aspects and limitations are considered in this work. We contribute to the theoretical base by proposing a novel approach to utilize big data in logistics. Findings described in this paper can be further used to implement a prototype, which will allow supply chain practitioners to increase their distribution efficiency, and improve their competitive position.

The remaining sections of the paper are organized as follows: the next section briefly describes our research approach and the findings of a survey to derive requirements for our concept. Section three provides an outline of the concept and its security aspects. In the subsequent sections, we further discuss the concept and its implications, assess the limitations as well as possibilities for future research, and close with a brief conclusion.

Design methodology and survey findings

This study is aimed to develop a big data based concept in logistics and follows the three-cycle design model proposed by Hevner (2007), see Figure 1. As the Design Cycle, this model specifies iterative development and refinement of the concept until all functional and non-functional requirements are fulfilled. Requirements for the Design Cycle as well as domain-related information are provided by the Relevance Cycle. In our research, we use customer survey as method to elicit requirements, provide rationales, and isolate application domains of the concept. The Rigor Cycle on the other hand provides a knowledge base for the development process and serves as theoretical and practical solution source for design subtasks. During the development, we use results of previous research, including (Pospiech and Felden 2012; Sadovskyi et al. 2014; Waller and Fawcett 2013).

Therefore, this study follows an approach that is comprised of: (1) reviewing the big data knowledge base, (2) developing a big data concept using the ANT theory aimed at realizing the opportunities, (3) capturing relevant customer needs and expectations in a particular field of logistics, and (4) discussing its implications.
Knowledge Base (Foundations)

Generally, the term “big data” refers to qualitatively and quantitatively new types of data, as well as to revolutionary new means of its collection, storage and processing. To characterize “big data” it is often spoken of the so-called three V’s (Chen et al. 2012; Davenport et al. 2012; McAfee and Brynjolfsson 2012): Volume, Velocity, and Variety. Volume refers to the proliferation of digital technology across value chains of organization. Velocity describes the speed of new data creation (creating the need for real-time data analysis). Variety reflects elements such as data from trucks or mobile phones. Thus, the actor network theory (ANT) is a suitable approach for our research (see below). To differentiate big data from business analytics, organizations blend innovative technology and skills to harness the three V’s and transform their data into useful business information. In the context of big data, this is called veracity (Hessman 2013).

The importance of big data is reflected by the constant development of new tools and techniques that make big data analytics more affordable and easier to implement (Manyika et al. 2011). Further, the number of reviews, case studies and theoretical publications devoted to the topic, as well as the amount of conferences held in the area of data mining and data analytics is growing each year (Sadovskyi et al. 2014). In consequence, the interest in big data increases and accelerates the adoption of big data-enabled business models in the industry (Manyika et al. 2011).

Sadovskyi et al. (2014) found that many theoretical and practical developments aim to improve internal processes such as operations due to the significant amount of business-related data that the companies usually accumulate throughout their business processes. However, to optimize external processes such as supply chain activities, technological and organisational changes have to be implemented across supply chain partners to enable mutual information sharing. While the exchange of information among partners is a known problem that hinders effective functioning of supply chain processes, new technologies and concepts support the information exchange process from various elements/objects such as trucks or mobile phones. For example, RFID and sensor networks allow faster tracking and classification of goods (Swaminathan 2012). Further, cloud computing enables the collection of data, and big data applications enable real-time optimisation for certain areas such as transportation planning and scheduling (Hessman 2013).

As distribution processes involve at least two supply chain partners (customer, distributor, and perhaps the supplier) and further equipment (elements/objectives), we chose the actor network theory (ANT) as theoretical foundation for our research. By choosing ANT, we can model various elements, even though these elements are dissimilar. The ANT was developed in order to analyze relationships between objects that are influenced by both human and non-human powers. This theory’s main tenet is that often systems and their constituents as objects have a complex socio-technical nature (Walsham 1997). Such objects need to be viewed as heterogeneous while clear object decomposition in social and technical parts is often not feasible. Moreover, interactions between the objects are also inseparably socio-technical.

For example, in our case seemingly technical task of delivery service is further complicated by customer behavior, who is inclined to change his location. Therefore, approaching this task as solely human-related or technically-related is unlikely to produce an effective solution. The actor network theory is a suitable instrument for modeling and analyzing last-mile delivery service as it is capable of taking into account
the complex technical and social relationships that appear between interacting entities such as customers, delivery agents, vehicles and mobile devices. ANT can also be used to inform the Design Science approach.

**Environment (Application Domain)**

In the field of distribution logistics, a customer-oriented approach is essential for any organizations success (Monczka et al. 2008). However, in reality, existing solutions do not provide organizations with all necessary tools. This is further complicated by the fact that customers become more dynamic and the percentage of online shopping activities steadily increases, resulting in a growing amount of packages that need to be distributed (Fang and Zhang 2005). Therefore, delivery service organizations constantly search for areas of improvement and develop adequate reaction to emerging trends, as this can be seen as one of the key success factors towards gaining competitive advantage.

In order to elicit the requirements for our concept, we have conducted a survey among the consumers of B2C delivery services. Particularly, we have targeted customers of online retail businesses who comprise one of the fastest growing shares of package delivery service users (Weltevreden 2008).

Survey participants were polled on their experiences with delivery services, particularly focusing on current issues and future transformations of the industry, viewed from a customer perspective. The goals of the survey included (1) determining the overall level of customer satisfaction with delivery services, (2) definition of the most frequent problems that the users encounter and gathering suggestions for possible solutions, and (3) identification of user preferences in order to locate general areas of improvement in the logistics industry.

The results of the survey include information obtained from 122 respondents, residing primarily in Europe. The age of respondents spans from 21 to 68 years. The survey was conducted online between October and November of 2013.

Participants could state their problems in a free text field. During the analysis, we clustered the stated problems into the following categories: late delivery, delivery when recipient is not available, lost packages, other (this category includes special cases such as missing notifications, damaged packages, destinations not found by delivery agent etc.). The analysis of the survey results shows that almost 50% of the respondents experience some sort of problems with delivery with at least 29% of the customers not getting their packages on time (see Figure 2).

![Figure 2. Frequent issue types, encountered by survey respondents.](image)

In order to look more deeply into how this issue can be addressed, we further investigated the personal preferences of that part of the audience which have indicated lower satisfaction scores in relation to the services (see Figure 3). This exposed that 38.7% of unsatisfied respondents name “Customizable delivery time” and 19.3% of respondents state “Customizable delivery destination” as their first priorities for the enhancement of delivery services. This speaks for the fact that current delivery service models lack flexibility. Traditional delivery models appear to no longer fit to the lifestyles of people, in particular online-shoppers, who are becoming more mobile and flexible (Weltevreden 2008). Further findings reveal that customers now tend to value punctuality of delivery more than speed and even more than cost.
Such a shift in the clients’ priorities is another indicium that customers wish to be put in control and be able to decide themselves where and when they want to receive their packages.

In addition, customers have expressed other interesting suggestions during the survey, which include delivery notifications over e-mail and SMS, real-time GPS-enabled package tracking, configurable delivery time slots before ordering, delivery outside of normal working hours and automatic delivery to a reserved address in case the customer is not at home.

**Big Data Distribution Concept**

In order to address the presented customer concerns, we have developed a framework that is comprised of three key components (see Figure 4). From right to the left: (1) personal mobile devices of the customers (smartphones), (2) a big data analytics computation center, residing in the data processing facilities of the delivery service provider, and (3) delivery vehicles, equipped with global positioning technology.

The information flow between the system components functions as follows: customers share their personal calendar as source of their prospective location information as well as global positioning data updates with the computation center through a dedicated mobile application, installed on their personal mobile device. On the other side, sensors inside of each delivery vehicle register all packages that are currently on board and send this information to the computational center. In addition, delivery vehicles constantly gather and transmit their location and road situation data. Under road situation information we understand real-time routing-relevant data such as traffic information, locations of road construction sites and detours, weather conditions and so forth.
On the basis of the road situation data and dynamic customer data, an analysis is conducted, which provides more accurate routing information and computation of a more efficient delivery schedule. Customer prospective positioning information is being utilized to compute delivery proposals in terms of a specific place and time of delivery, which are sent back to the customers’ personal mobile devices. Consequently, customers receive a possibility to choose one of the locations that are listed in their calendar with a suggested time for delivery. After selecting one of the options by the customer, the computation center recalculates the most optimal route, sends a routing update and the booked delivery time slot information to the corresponding delivery vehicle. In order to optimize routing and serve customers in the best way, such negotiation rounds need to be conducted with each customer.

After the negotiation with all customers is complete and the route is communicated, the delivery vehicle driver only receives subsequent routing updates on the basis of changing user locations and/or road situations. For the case that a customer changes its planned location in the calendar or data readings from the GPS sensor of its phone do not correspond to the planned location, this will be recognized by the mobile application. The user will then receive a query if they are experiencing changes in their calendar, and if the user confirms, an update will be sent to the computational center where the route will be recalculated and the delivery vehicle driver will be notified.

The technical side of the concept is as follows: Inside the vehicle, packages are registered based on their tracking number during loading and unloading using RFID technology. Their tracking numbers are collected and an inventory list of the truck is created. This list is sent to the computation center via a dedicated communication interface using 3G wireless Internet connection. The computation center links tracking numbers to the customer data and recognizes which users need to be contacted for further information. Subsequently, users are notified with a short text message and are urged to launch the preinstalled application on their mobile device to start the communication with the computation center, which then sends specific information requests back to the application.

After receiving the information request, the mobile application asks the user to grant temporary access to the calendar and GPS navigation unit of its mobile device. If the access is granted, data is captured by the mobile application and transferred over an encrypted channel to the computation center, where it is provisionally stored in a database. The secure communication channel is negotiated on top of a mobile broadband Internet connection (3G). It is important to mention that dynamic data, such as GPS positions of the delivery vehicles and road situation readings are constantly being collected and create large amounts of data, which need to be analyzed in real-time for route optimization purposes. This is the reason why an in-memory database needs to be used in this concept. As soon as necessary information and location data from the customers is gathered, real-time big data analysis methods are invoked, and an optimal route for the vehicle is computed. Following, the users are notified over the already established secure communication channel about the delivery time and place possibilities.

In order to achieve security and safety of user data, no personal information shall be transferred over the communication network. For example, the communication with delivery vehicles only includes package tracking number and communication with mobile device is conducted through a separate secured channel for each connected user. No user location data shall be accessible outside of the in-memory database at any time and shall be deleted immediately after the delivery is complete and the communication session with customer mobile device is closed. Only the road situation information is constantly kept in the database in order to facilitate delivery route computation for all vehicles and improve route proposals based on historic data. An important note in this regard is that the customer remains the single owner of his location data and according to the concept, only temporarily lends it to the computation center, solely for the purpose of routing analytics and planning.

The ANT allows to model interactions between objects have complex social, technological and political nature (Tatnall and Gilding 1999). In the course of this study we used this approach to build a holistic understanding of the issue and conceptualize communicating entities as well as interactions between them. As the result of such modeling, additional concerns were recognized and possible countermeasures considered. The theory can be used to link the solution knowledge base (rigor cycle) with the requirement pool (relevance cycle) consequently facilitating the chosen design science approach.
Discussion and Implications

In this work, we develop a concept aimed to improve distribution processes taking the needs of online shopping customers into account. Thus it contributes to the scarce literature on utilizing big data for supply chain management (Waller and Fawcett 2013). The described concept is meant to resolve the insufficient flexibility of conventional delivery methods using a big-data approach that incorporates the latest data-intensive technological developments such as sensor networks, global positioning technology and big data analytics.

We found that currently existing concepts which are meant to target flexibility issues, including packstations (Lampe and Stölzle 2012) and package boxes (Schnedlitz et al. 2013) do not address the customizable delivery destination issue and also require the installation of additional equipment near the customers' place of residence. Such additional setups can be relatively costly and labor consuming. There also exist limitations for residents of large apartment buildings and other densely inhabited areas, where package boxes cannot be installed for every tenant due to a lack of free space. Packstations have drawbacks in remote rural areas, where their installation is hard to justify economically. Both concepts are also potentially prone to robbery. The survey, conducted within this study shows that 49% of respondents are willing to pay a small fee for using dynamic delivery services, but only 29% are willing to invest a larger sum to install a package box.

In addition, concepts such as packet shops and locker points (Schäfer 2011; Weltevreden 2008) also do not address flexibility in terms of place of delivery and require additional effort from customers to pick up their packages. Another constraint is that customers need to adapt to the working hours of the packet shops. In general, existing ideas target the organizational side of the distribution process without attempting to utilize more sophisticated technology to improve and automate information flows within the processes.

Therefore, we developed a dynamic approach that delivery service organizations can use in order to improve existing processes or define new business models, subsequently achieving better customer satisfaction and lower transportation costs. The described concept addresses these issues by automatically collecting supply chain and supplementary data from the delivery services company's infrastructure using sensors, and customer location data. This information is combined and analyzed in order to provide better routing decisions and achieve higher flexibility in terms of time and place of delivery, effectively increasing customer satisfaction.

Furthermore, as mentioned by Sadovskyi et al. (2014), most organizations still struggle to introduce interorganizational information flows in the field of SCM. The concept presented in this paper contributes to theoretical and practical research in this direction by linking upstream and downstream supply chain information, and subsequently combining it with dynamic road conditions data to benefit operations and outbound flow. In consequence, this helps to optimize routing and serve customers who dynamically change their location. Moreover, the concept has potential benefits for delivery service providers by ensuring successful delivery at first attempt, and therefore offer reduction in transportation costs.

Limitations and future research

The main challenge that the proposed concept is confronted with, is the security of the customers’ personal information. Security requirements are important, as users are becoming more protective about their personal data. Thus a trade-off between the amount of personal data and the degree of flexibility needs to be made, when realizing this concept. Customer acceptance of the concept might be impaired if too much personal data needs to be made available to a service provider. In this regard, other, less information-intensive variations of the concept are also possible. For example, using information only from the customer calendar without dynamic verification through GPS data or even a minimal provision of specific calendar time entries upon request can be seen as alternatives. Another concern is that the realization possibility for certain smartphone manufacturers may also be a challenge, as not all mobile device operating systems allow clear application access delimitation across mobile device components.

Although the described concept is specific to B2C distribution functions, we believe that it may serve as a starting point for future projects aimed at the enhancement of B2B supply chain processes based on interorganizational information sharing. Furthermore, this concept illustrates how big data analytics may
be utilized in defining novel approaches to address existing customer needs and therefore create competitive advantage by redefining business processes. The paper also contributes to a theoretical exploration of information sharing patterns among multiple supply chain partners. Further research could use our insights to improve SCM processes using real-time analytics, making them more beneficial to a broader customer base. A prototype, realizing the described concept can be developed by extending the mutual application of design science and actor network theory in the course of the future research. Drawing from ANT, we will model the actors, elements and objects and their interrelations in more detail for the foundation of our prototype.

Conclusion

The rapidly growing amount of available data along with the emergence of new technological tools provide new opportunities for the enhancement and redefinition of supply chain processes in order to gain competitive advantage by reducing costs, improving operations or creating innovative solutions. This research develops a concept for a data-intensive application to improve distribution processes of firms and minimize delivery attempts, resulting in increased customer satisfaction by providing dynamic functions as desired by customers.

REFERENCES


