数据挖掘推动客户数据驱动车辆的开发进程

WEGENER Jan[1]，VAN PUTTEN Sebastiaan[2]，
NEUBECK Jens[3]，WAGNER Andreas[1,3]

（1. 斯图加特大学 汽车工程学院,斯图加特70049,德国；2. 奥迪股份公司,英戈尔施塔特85049,德国；
3. 斯图加特汽车工程与车辆发动机研究所,斯图加特 70049,德国）

摘要：对于汽车等面向消费者的行业,数据驱动的产品开发是一项关键的系统工程。数据驱动方法的前提是数据本身,由于现代车辆的联网能力不断提高,使得汽车制造商能够以内部总线信号的形式记录并储存客户数据。由于这些数据并不用于外部,只用于内部通信以保证车辆的安全性和功能性,这给这些数据的应用带来了一定的困难,因此汽车行业的主要问题在于如何利用数据挖掘技术从这些数据中提取客户需求及相关信息。为此,对上述数据应用问题进行了文献调研,并在此基础上开展数据挖掘的模拟研究,以确定现有数据挖掘过程在需求获取领域的适用性；进而提出一种扩展程序,使得数据挖掘技术能够应用于相关客户数据之中,从而加快车辆开发的进程。

关键词：车辆开发；数据驱动；数据挖掘；跨行业数据挖掘标准流程；数据驱动需求获取

中图分类号：U461　　　　　**文献标志码**：A

Data Mining as an Enabler for Customer Data Driven Vehicle Development

WEGENER Jan[1], *VAN PUTTEN Sebastiaan*[2], *NEUBECK Jens*[3], *WAGNER Andreas*[1,3]

（1. University of Stuttgart Institute of Automotive Engineering，70049 Stuttgart，Germany；2. AUDI AG，85049 Ingolstadt，Germany；3. Forschungsinstitut für Kraftfahrwesen und Fahrzeugmotoren Stuttgart，70049 Stuttgart，Germany）

Abstract：Data-driven product development is a key technology for systems engineering especially for consumer-oriented industries such as the automotive industry. The basic prerequisite for all data driven approaches is data itself. Due to the increasing networking capabilities of modern vehicles，automotive manufactures are able to record and store customer data in the form of internal vehicle bus signals. The challenge in using this data is that it is not designed for external use，but for internal communication to ensure the safety and functionality of the vehicle. Therefore，the main question is how to extract customer needs and consumer-relevant information within this data using the process of data mining（DM）. Consequently，in this paper，a literature review on the aforementioned use case is conducted. Based on the literature research，a DM simulation game is conducted to determine the suitability of existing DM processes in the area of requirements elicitation. Finally，a process extension is proposed that helps to systematically focus the DM process on customer-relevant information and thus accelerate the overall process.

Key words：vehicle development；data driven；data mining（DM）；cross-industry standard process for DM（CRISP-DM）；data driven requirements elicitation

---

1　Introduction

The development of modern systems faces several challenges across all industries. System Design often still emerges bottom up from individual pieces，instead of top down from an architecture[1]. In addition the complexity of those systems is increasing faster than it can be managed[2]. Especially the latter is highly relevant in the automotive industry where the complexity has significantly increased over the last decades[3]. A trend that will only continue with the advent of autonomous driving and the continuous efforts in drive train electrification[4]. Modern vehicles are basically turning into high performance computing platforms[2]. On top of that OEMs are accelerating the time to market for innovations and inventions[5].

All of these factors increase the planning efforts in all development phases[2]. OEMs are incentivized to create a development methodology, which avoids errors and reduces their consequences in case they occur[1]. Requirements elicitation as the first part of system development therefore becomes all the more important. Specifically, the identification and analysis of the demands and the fulfilment of the expectations of consumers.

First steps in data driven automotive systems engineering (ASE) haven been made by Bach et al[6] by proposing a complementing data driven approach. On the basis of real world test data from prototypes and series cars for quality assessment, they proposed methodologies for using the recorded data in each of the development phases of the standard automotive systems engineering process phases. In the case of requirements elicitation, they recommend to use generic statistical methods to answer simple w-questions like why, when, where etc. certain things happen. They suggest that deeper analyses of the dependencies between signals might uncover useful information but do not go any further on this[6]. The intention of this paper is to enhance the requirements elicitation phase of ASE by proposing a data mining (DM) methodology to uncover useful hidden information in recorded vehicle data.

First, the state of the art for ASE is briefly presented followed by a more detailed report of DM, presenting the common methodologies and differences between them. Then, the methodology used to derive the basic approach for automotive DM is explained, followed by an exemplary application of said approach. Lastly, the methodology is concluded and an outlook for the use of DM in ASE is given. Figure 1 shows the general procedure from data collection to data driven requirements.
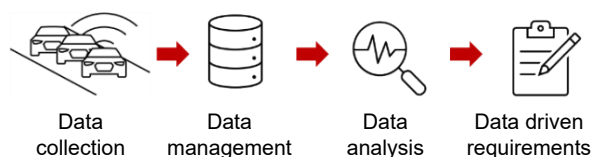

Data collection → Data management → Data analysis → Data driven requirements

Fig.1 General procedure from data collection to data driven requirements

2 State-of-the-art

The state-of-the-art in systems engineering is first briefly presented, followed by the state-of-the-art in DM in a bit more detail.

2.1 Automotive systems engineering

Systems engineering in general and automotive systems engineering especially are key areas of expertise for managing complex systems in the automotive world[1]. It is based on the knowledge of systems science and expanding this into the domain of technical development[7]. This is also reflected in the term itself consisting of the two words systems and engineering, combining two distinct perspectives into one, which provides a systematic way of development[7]. Mastering the systems engineering approach avoids development arbitrariness, provides verifiable, traceable goals and progress by transferring deep understanding about the system to solutions[1]. The International Council on Systems Engineering defines systems engineering as "a transdisciplinary and integrative approach to enable the successful realization, use, and retirement of engineered systems, using systems principles and concepts, and scientific, technological, and management methods"[8].

Application assistance for systems engineering are provided by generic process models such as the V-model according to VDI 2206[9], waterfall model, spiral model or VDI 2221[10]. The underlying structure of these processes and sub processes are based on a fundamental understanding of how to transform customer needs to a solution[11]. These generic processes are often adapted and improved by organizations. The most mature versions get implemented as internal development processes and form the basis for industry standards such as the software process improvement and capability determination (SPICE) model[12]. SPICE is published by the Institute of Electrical and Electronics Engineering as a standard for software development processes[11]. SPICE is part of the ISO/IEC 15504[13] standard that manages the analysis, assessment and improvement of processes[11]. Several domains have

adopted and adapted the model to their specific requirements，such as medical SPICE（Medi SPICE）space SPICE（SPICE 4 Space），Automotive SPICE（ASPICE），and many more[11]. The ASPICE model boosts the division of the system （the vehicle）into smaller logical more convenient to work with subsystems.

2.2　DM

The term DM is used to refer to just one-step of one of the earliest DM methodologies called Knowledge Discovery in Databases（KDD）[14-15]. Today it refers to the whole methodology and is used synonymously with the terms Knowledge Discovery （KD）and KDD[16-17]. The area of DM describes the development of methods and techniques that extract useful information from data[14,17-18]. The core problem is how low-level data that is too extensive for direct comprehension can be converted into other formats，which，depending on the application，are more compact（reports），more abstract（models and equations）or in general more useful[14]. Discovered Knowledge should be applicable to new data，novel and comprehensible at the latest after some postprocessing[14].

The tools and methods used in DM are framed inside of systematic processes. Motivated by the fact that blind application of said techniques can lead to the discovery of meaningless and invalid knowledge[14,16,19]. This kind application called data dredging resulted in a negative connotation of the term DM when it first arose[14]. Data mining models provide a set of processing steps that can be pursued when executing DM projects. By detailing the procedures of each individual step the DM models help to plan，work through and reduce costs of the overall project[16, 20-21]. Since the original KDD process a multitude of different DM models have been proposed[16]. Two extensive surveys one from Mariscal et al[17] and one from Plotnikova et al[21] list 14 different DM processes. Most of them related to either the initial KDD process or the Cross Industry Standard for DM（CRISP-DM）[17, 21-22]. In addition to the aforementioned surveys there have been several other researchers analysing and comparing the various

DM processes Kurgan and Musilek[16]，Marbán et al[23] and Martínez-Plumed et al[24]. All of these studies emphasize the similarities as well as the historical development of those DM processes. KDD being the initial process that inspired many alternations of its original ideas and CRISP-DM being the standard model borrowing ideas from its most important predecessors and simultaneously laying the groundwork for many following proposals[17, 21, 23-24].

The main distinction among the models lies in the recommended number and scope of their specific steps[16]，often splitting up tasks into two steps that are one in another model and vice versa[17]. Nevertheless，most of the models are subject to the same basic structure，sequentially completing tasks with similar content，often supported by the iterative nature of the processes involving iterations and loops between different steps[16]. The common steps or tasks of the processes are：domain understanding，data preparation，modeling and evaluation[16]. The specifics of these steps are explained in in the next paragraph by presenting the CRISP-DM process model.

Developed by an industry consortium CRISP-DM is designed to be domain independent[17]，which lead to a wide adoption by industry and research communities[23]. As a result CRISP-DM is often cited to be the de facto standard for DM[16-17]. The main steps of CRISP-DM as shown in Fig. 2 are described below.
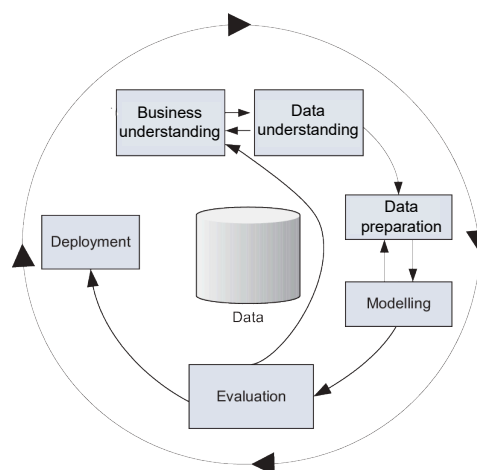

Fig.2　Process model Cross Industry Standard for DM (CRISP-DM)[25]

Business Understanding：This first phase focuses on understanding the project goals and requirements from a business perspective. Based on that knowledge the task gets translated into a DM problem definition including a preliminary plan for achieving the set goals[22].

Data Understanding：This phase begins with an initial data collection and continues with activities to familiarize oneself with the data，identify data quality problems，gain first insights into the data or uncover interesting subsets in order to form hypotheses for hidden information[22].

Data Preparation：The data preparation phase deals with all tasks to transform the initial raw data into the final dataset that is being fed to the modeling tools in the next phase. As indicated in Fig. 2 these and the next phase are likely to be performed multiple times and without any determined order[22]. Data preparation very often requires significant amount of manual data manipulation which is difficult to automate[16]. Because of this the data preparation phase is by far the most time consuming phase taking up 45%—60% of the overall time spent on each DM project[16].

Modeling：In the modeling phase numerous modeling techniques are selected and utilized depending on the DM problem. The model parameters are optimized through iterative application. The selected modeling technique often determines the specific form of the fed data. Therefore，as mentioned above stepping one step back to data preparation is often unavoidable[22].

Evaluation：This second to last phase deals with evaluating the developed model (possibly models) including thoroughly reviewing the executed steps along the way. The main task is to determine whether all goals set during the business understanding phase are sufficiently met. The evaluation phase ends with a decision on how to use the DM results[22].

Deployment：In the last phase, the evaluated model(s) are deployed to enable data driven decisions and support in the business process for the end customer. Depending on the end customers' requirements deployment might be as straightforward as generating a report or as complicated as implementing a repeatable DM process[22].

3 Method

In this section，the practical approach to find a suitable DM methodology is described first. Subsequently，the DM framework proposed in this paper is presented，resulting from the accumulated knowledge from the application of the former. The goal being to find the most suitable DM methodology for ASE. As mentioned in the state of the art DM section, there have been several models developed throughout recent years. These standardized processes like CRISP-DM，KDD，SEMMA etc. are designed to be industry agnostic and as such might not fulfill the requirements of each domain[26]. The sheer number of processes combined with their industry agnostic characteristics make the model selection a daunting task[16].

3.1 Simulation game for processing different theoretical requirement elicitation examples through DM

In order to find the most suitable DM methodology for ASE a simulation game on how to theoretically address three different DM problems is conducted，focusing on how to repeatedly and reliably discover hidden knowledge on how the customer uses the vehicle. Question 1 (Q1)：How does the customer use the manual shift mode of the automatic transmission? Question 2 (Q2)：How does the customer use the preinstalled navigation system? Question 3 (Q3)：What is the customer's maximum speed? Three different types of questions were specifically picked to yield a wide field of possible DM tasks. Each question intended to focus on different aspects of the customers interactions with the vehicle：Q1 to focus on interactions with the mechanical part of the vehicle in this case the different shift input methods (paddles / stick)，Q2 to focus on the interactions with the digital part of the vehicle and Q3 to focus on the whole vehicle including its surroundings like traffic or weather conditions. Each

of these questions was examined separately in a DM simulation game in which the individual steps were processed one after the other. Due to the abstract nature of a simulation game, the focus could inevitably not be on specific models, the associated parameter tuning, data cleaning or something similar. Instead, the focus is primarily on the domain understanding, data understanding and evaluation phases.

In order not to get tangled up by the specifics of each individual DM process, the simulation game was performed using only the CRISP-DM methodology to first get a baseline of how to generally apply DM projects in the context of ASE. Documenting the steps along the way enabled the analysis of the three applications and the search for similarities as well as differences to deduct more general recommendations on automotive bus mining. Originally, the purpose of the analysis was to see if any of the other techniques offered significant advantages over the CRISP-DM methodology. Partly during the application, but at least during the analysis, it became clear that most of the DM methods were applicable in principle, but only with the right framework supporting them. Three critical application weaknesses were discovered which, if circumvented, increase the chances of a successful DM project and thus the discovery of new insights into customer behavior.

The first and biggest weakness is right at the start of the process in the business understanding step. The simulation game shows that the instructions, due to their industry-agnostic design, are too broad and too unspecific for a simple, reproducible application. In particular, if the goal of the DM project is to discover new knowledge, the business understanding should be based on a system which ensures that no aspect of the customer behavior regarding the initial question is excluded. Without some sort of systematically approach, discovering hidden knowledge is like searching for the needle in the haystack.

The second issue discovered is that the data preprocessing of the vehicle bus signals is in some cases so complex that a separate DM process is

necessary. The main reason for this is how the data is sent over the bus. Since the data is designed for internal communication and not for retrospective analysis, theoretically simple data preparation task requires a great deal of effort. An example of this is the determination of the type of road the vehicle is on (e. g., motorway, country road or city road). Depending on the data protection regulations, it may be forbidden to save the GPS position of the vehicle, as well as the use of the navigation system data, as these also contain personal data. The only way around these regulations is to design models that can determine the type of road based on other signals that are not relevant to data protection. This can quickly escalate in complexity if not only simple street types are to be determined, but more detailed levels are desired, for example, to distinguish between freeway entrances and exits or country roads with and without serpentines.

The last issue results from the previous one. Since the data preprocessing step can in some cases lead to separate DM tasks, which in turn require their own data preprocessing steps, the time required for the entire DM project is greatly underestimated. This requires some sort representation since the time spent on the entire project as well as the individual steps is an important aspect of the overall acceptance as well as transparency of such DM projects.

3.2 Proposed DM framework

The framework orients itself on the four fundamental steps of all DM processes, domain understanding, data understanding, modeling, and evaluation (compare Section 2.2). The overarching idea behind the proposed framework is to categorize customer behavior by means of use cases that are as detailed as possible and then to use them in the form of metadata for further analysis. The categorized metadata serve as a simplified representation of all vehicle bus signals relevant to the initial usage question. The whole framework is depicted in Fig. 3.

The proposed DM framework starts with a usage question, similar to the Q1, Q2 and Q3 questions above. Based on this question, the first process step takes place：**Use Case Derivation**. As alluded to
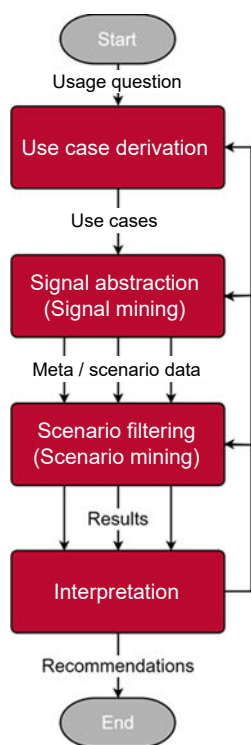
Fig.3 Process model of proposed DM framework

above, the task of this step is to map any user behaviour related to the usage question in the form of use cases. In order to be able to create as accurate a picture as possible of customer behaviour, the simple but effective method of w-questions is recommended. This technique provides a repeatable system for categorization while allowing a differentiated view of customer usage. Furthermore, each question offers the possibility for further detailing. For example, if the question "*Where?*" is considered, this can be answered with increasing level of detail by: Country, state, county, city, etc. If the entire fleet of customer vehicles of a manufacturer or even group is examined, region specific requirements might be derived as a result. The question of the type of road, which was already mentioned in the last section, also forms another layer of the "*Where?*" question and can again be considered in various degrees of detail. Other notable examples, are the questions "*Why?*" and "*How?*" These are usually more complex and therefore require more effort than others, but they can also tell us more about customer behaviour. An example is question Q3, where it can be more interesting to find out why customers did not drive

faster than their actual speed. At the same time, the question is much more difficult to answer due to the numerous possible influences such as: weather conditions, surrounding traffic or in case of electric vehicles the state of charge of the battery. Logically, depending on the questions to be investigated, different detailed w-questions and consequently different use cases arise. This enables a further benefit of the Domain Understanding method proposed by us, since all questions and use cases can be recorded in the form of catalogues and used for other analyses. This ensures a high degree of reproducibility.

Once all use cases have been captured, the next process phase **Signal Abstraction (Signal Mining)** begins. The goal of this phase is to map the previously derived use cases with suitable bus signals in order to record the occurrence of the use case in customer behaviour in the form of scenario/meta data. This is where the aforementioned varying complexity in mapping the use cases becomes apparent. Without privacy laws, for example, all of the "*Where?*" questions mentioned above could be answered simply by an accurate GPS position of the vehicle. If these laws apply, however, other solutions must be found. For example, rule-based or classification algorithms that can map the searched use case by combinations of other signals. As soon as such a separate DM task arises, established DM methods such as CRISP-DM can be used.

With the generated metadata, the main analysis phase **Scenario Filtering (Scenario Mining)** can start either exclusively on a metadata basis or in addition to raw data. The aim is to gain new insights and information about customer behaviour through mining, e. g. in the form of structure-discovery processes or intelligent combination of metadata, e. g. through conditional scenarios. Processes like CRISP-DM can also be used here. Finally, in the last phase, **Interpretation**, the collected partial results are brought together and evaluated in relation to the initial question. The desired overall process product is an objective, data driven requirements recommendation.

4　Experimental applications of proposed framework

In this section，part of the exemplary application of the proposed DM framework is presented based on question Q3：What is the customer's maximum speed？ As a preliminary disclaimer，the data used for the application was recorded from test vehicles and not from customer vehicles. Accordingly，this is only a presentation of the methodology and not the data or the knowledge gained as there are likely to be differences in the driving behavior of test drivers and normal customers. Nevertheless， the data are recordings from real road traffic on German roads，so except for the driver，there should be no difference to the data that could be collected from customers. Thus， there is nothing to prevent the methods developed from also being applicable to customer data.

This is primarily focused on the middle two process phases and in particular on a question already mentioned in the previous part：Why did the driver not speed up？ The identification and subsequent combination of the constraining influences allows a conditional statement about the speed chosen by drivers when they are not constrained by external influences such as surrounding traffic or speed limits. This can provide a new perspective on the same data，

increasing the quality of the information and providing new knowledge about the data.

In the first phase of the process，the w-question method was used to identify numerous influences that prevent drivers from driving faster， such as the type of road（especially in Germany with its mostly unrestricted autobahns）， speed limits，vehicles ahead，and weather conditions. The mapping of these identified influences onto simple categorical metadata in the second phase epitomizes the varying complexity mentioned in the introduction of the framework. Speed limits and temperature（as a simplified representation of weather conditions）could be mapped directly through simple bus signals. In order to be able to map the existence of a vehicle in front by means of a categorical variable，a more complex rule-based algorithm had to be used and，due to data protection regulations in Germany，even a machine learning classification algorithm for determining the road type. The classification model required extensive preprocessing to match the GPS positions recorded by test vehicles， for which other data protection regulations apply，to different road types with the help of OpenStreetMap[27] data. The model could then be trained and validated based on the matched road types. Fig. 4 shows a section of a recorded journey，with the unclassified route at the top and the road types classified with the model at the bottom.

Fig.4　Exemplary representation of road classification, unclassified trip sections above and classified sections below

Both the rule-based vehicle detection and the road type classification are examples of complex preprocessing and separate DM，respectively. This underlines why the associated process step is called signal mining. Furthermore，both methods illustrate

the reusability of the developed methods，as both can be used for other problems without further ado.

The derived metadata could be used in the next step of the process to make a conditional statement about the question listed at the beginning. Based on

the individual metadata, scenarios of varying detail were formed, and one potentially limiting environmental factor was filtered per scenario layer. Fig. 5 shows the distribution of the speed driven by a boxplot for each scenario. Scenario 1 shows the unfiltered speed. Scenario 2 shows a filter level in which only measured values on motorways and motorway-like road types are displayed. Scenario 3 additionally filters all readings with detected speed limits and Scenario 4 filters all readings where a vehicle in front of the measuring vehicle was detected. The last filter stage in Scenario 5 filters acceleration and deceleration phases of motorway entrances and exits, as the desired cruising speed has not yet been reached.
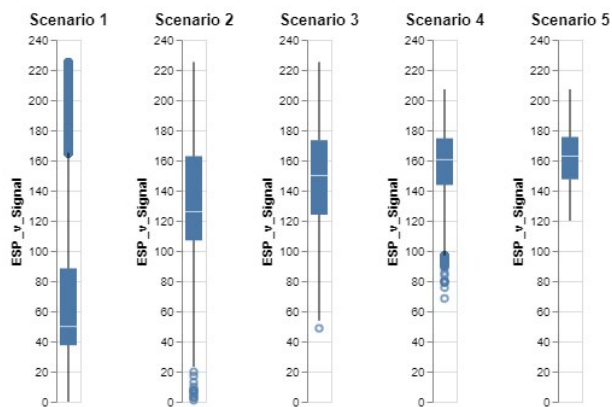

Fig.5　Representation of the speed distribution per scenario, each scenario filters one factor influencing the choice of speed

It is probably not possible to identify all influences on the choice of speed, let alone filter them. Other influences such as temperature were not relevant for this evaluation, as all rides were recorded under similar conditions. Nevertheless, the scenario illustrates a clear trend. The lower the number of factors on the choice of speed, the higher the drivers choose their own speed. Also remarkable is the decreasing dispersion of the recorded speeds per scenario, which can be easily recognized by the shrinking interquartile range of the individual boxplots.

With regard to the issue under consideration, it becomes clear how important the methods such as the scenario filtering shown here are for requirements elicitation in the automotive sector. Assume that for a new vehicle project, a requirement for the maximum speed of the vehicle is sought. Based on the unfiltered data (Scenario 1), it would be reasonable to set a requirement recommendation for the maximum speed to a value around the upper whisker of the boxplot (about 160 km/h). In this way, only about 1.5% of the recorded values would be unattainable for the new vehicle. However, if only the data where drivers were not restricted in their choice of speed were consulted (Scenario 5), the same approach would result in a recommendation for a maximum speed of over 200 km/h. At the same time, the data show that the first requirement recommendation would result in about 50% of all driving situations in which the driver has a free choice of speed being unattainable for a vehicle developed according to this requirement. This could lead to customers either being dissatisfied with this product or even ruling out a purchase from it. Alternatively, the data could be used to set a threshold that determines how much current customer behavior may be constrained by the requirements of the new product. For example, if the goal is to maintain 90% of the current possible customer behavior in terms of the maximum speed driven, the maximum speed requirement may be reduced to 185 km/h.

The example illustrates the added value of conditional information for requirements elicitation. The method presented here allows the user to compile his own brand-relevant scenarios via the collection of metadata and to design his products specifically according to a scenario-dependent product usage analysis, resulting in objective, data driven and customer-centric products.

5　Conclusions

Data driven customer integration in the product development process is one of the key technologies that needs to be mastered in order to remain competitive in a world of increasing product complexity and growing customer market power. One approach to master this challenge is customer

DM to support the requirements elicitation process. On the basis of an intensive DM simulation game，this paper worked out why the existing DM processes are not suitable for this purpose. The identified weaknesses of the existing processes as well as the collected insights of the business game were used to design a suitable DM process for customer DM. This proposed methodology was explained in detail and then the added value of this was presented through an application example. One of the key aspects of the proposed methodology is the categorization of customer behavior through metadata. The analysis of this metadata provides a reliable and repeatable system that facilitates the discovery of new knowledge in the data. For this reason，it can be stated that metadata categorization is a key technique for understanding the intricacies of the inner workings and interactions of modern vehicle bus systems，and thus for understanding and mapping customer behavior. In the long term，there is the possibility of categorizing every aspect of customer-vehicle interaction to enable true big data analytics.

A major downside of this approach，lies in the fact that this type of data analysis is by definition reactive. Only the customer behavior that is possible in the context of the current vehicle can be analyzed. Therefore，this approach must always be used in conjunction with other proactive customer analyses.

Research needs for the process model proposed here include the suitability of the developed methods in the real requirements derivation process. Subsequently，it would be particularly interesting to see to what extent currently existing requirements deviate from the purely data driven requirements. Furthermore，it should be investigated which methods are best suited to discover new patterns and thus new knowledge from large categorized customer databases through unsupervised learning methods.

References：

［1］ MAURER M，WINNER H. Automotive systems engineering［J］. Berlin：Springer，2013.

［2］ D'AMBROSIO J，SOREMEKUN G. Systems engineering challenges and MBSE opportunities for automotive system design［C］// 2017 IEEE International Conference on Systems，Man，and Cybernetics（SMC）. Banff：IEEE Xplore，2017：2075.

［3］ MASCHOTTA R，WICHMANN A，ZIMMERMANN A，et al. Integrated automotive requirements engineering with a SysML-based domain-specific language［C］// 2019 IEEE International Conference on Mechatronics（ICM）. Ilmenau，Germany：IEEE Xplore，2019：402. DOI：https://doi. org/10.1109/ICMECH.2019.8722951.

［4］ FANK P，BOJA D，ABTHOFF T. Big data driven vehicle development – Technology and potential［C］// 2020 Internationales Stuttgarter Symposium. Wiesbaden：Springer Fachmedien Wiesbaden，2020：315. DOI：https://doi. org/10.1007/978-3-658-30995-4_31.

［5］ FRAGNER A，KREIS A，HIRZ M. Virtual tools to support design and production engineering：Early detection of stone chips to optimize production processes［C］// 2020 IEEE 7th International Conference on Industrial Engineering and Applications（ICIEA）. Paris：IEEE，2020：399. DOI：https://doi.org/10.1109/ICIEA49774.2020.9102004.

［6］ BACH J，LANGNER J，OTTEN S，et al. Data-driven development，a complementing approach for automotive systems engineering［C］// 2017 IEEE International Systems Engineering Symposium（ISSE）. IEEE，2017：1. DOI：https://doi.org/10.1109/SysEng.2017.8088295.

［7］ BAJZEK M，FRITZ J，HICK H. Systems engineering principles［C］// HICK H，KÜPPER K，SORGER H. Systems Engineering for Automotive Powertrain Development. Cham：Springer International Publishing，149. DOI：https://doi.org/10.1007/978-3-319-99629-5_7.

［8］ SILLITTO H，MARTIN J，MC KINNEY D，et al. Systems Engineering and System Definitions［Z］. San Diego：INCOSE － International Council on Systems Engineering，2019.

［9］ VDI. Entwicklung cyber-physischer mechatronischer systeme（CPMS）：VDI/VDE 2206［S］. 2020［2021-10-11］. https://www. vdi. de/richtlinien/details/vdivde-2206-entwicklung-cyber-physischer-mechatronischer-systeme-cpms.

［10］ VDI. Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte：VDI 2221［S］. 1993［2021-10-11］. https://www. vdi. de/richtlinien/details/vdi-2221-methodik-zum-entwickeln-und-konstruieren-technischer-systeme-und-produkte.

［11］ BAJZEK M，FRITZ J，HICK H. Systems engineering processes［C］// HICK H，KÜPPER K，SORGER H. Systems Engineering for Automotive Powertrain Development. Cham：Springer International Publishing，2021. 235. DOI：https://doi.org/10.1007/978-3-319-99629-5_9.

［12］ ESCH J，RETTMANN A，MARZINEAK S. A systems engineering approach to electromagnetic compatibility［C］// LIEBL J. Der Antrieb von morgen 2021. Berlin：Springer Berlin Heidelberg，2021：167. DOI：https://doi.org/10.1007/978-3-662-63403-5_11.

［13］ International Organization for Standardization. Information technology — Process assessment：ISO/IEC 15504-5：2012 ［S］. ［2021-10-11］. https://www. iso. org/standard/60555. html.

［14］ FAYYAD U, PIATETSKY-SHAPIRO G, SMYTH P. From data mining to knowledge discovery in databases［J］. AI Magazine, 1996, 17：37.

［15］ FAYYAD U, PIATETSKY-SHAPIRO G, SMYTH P. The KDD process for extracting useful knowledge from volumes of data［J］. Commun ACM, 1996, 39(11)：27. DOI：https:// doi.org/10.1145/240455.240464.

［16］ KURGAN L A, MUSILEK P. A survey of knowledge discovery and data mining process models［J］. The Knowledge Engineering Review, 2006, 21(1)：1. DOI：https://doi.org/ 10.1017/S0269888906000737.

［17］ MARISCAL G, MARBÁN Ó, FERNÁNDEZ C. A survey of data mining and knowledge discovery process models and methodologies ［J］. The Knowledge Engineering Review, 2010, 25 (2) ： 137. DOI： https://doi. org/10.1017/ S0269888910000032.

［18］ HAN J W, KAMBER M, PEI J. Data Mining：Concepts and Techniques［M］. ［S. l.］：Elsevier professional, 2011.

［19］ ROTONDO A, QUILLIGAN F. Evolution Paths for Knowledge Discovery and Data Mining Process Models［J］. SN Computer Science, 2020, 1(2)：109. DOI：https://doi.org/ 10.1007/s42979-020-0117-6.

［20］ CIOS K J, KURGAN L A. Trends in Data Mining and Knowledge Discovery ［C］// PAL N R, JAIN L. Advanced Techniques in Knowledge Discovery and Data Mining. London：Advanced Information and Knowledge Processing, Springer London, 2005：1. DOI：https://doi.org/10.1007/1-84628-183-0_1.

［21］ PLOTNIKOVA V, DUMAS M, MILANI F. Adaptations of data mining methodologies：A systematic literature review［J］. PeerJ Computer Science, 2020, 6(2)：e267. DOI：https:// doi.org/10.7717/peerj-cs.267.

［22］ CHAPMAN P, CLINTON J, KERBER R, et al. CRISP-DM 1.0. Step-by-step data mining guide［A］. ［S. l.］：CRISP-DM Consortium, 2000.

［23］ MARBÁN Ó, MARISCAL G, SEGOVI J. 2009. A data mining & knowledge discovery process model ［C］// PONCE J, KARAHOC A. Data Mining and Knowledge Discovery in Real Life Applications. ［S. l.］ ： I-Tech Education and Publishing, 2009. DOI：https://doi.org/10.5772/6438.

［24］ MARTÍNEZ-PLUMED F, CONTRERAS-OCHANDO L, FERRI C, et al. CASP-DM context aware standard process for data mining ［DB/OL］. (2017-09-19) ［2021-05-28］. https:// doi.org/10.48550/arXiv.1709.09003.

［25］ PECHENIZKIY M, PUURONEN S, TSYMBAL A. Does relevance matter to data mining research?［C］// KACPRZYK J, LIN T Y, XIE Y, et al. Data Mining：Foundations and Practice：Part of the Studies in Computational Intelligence book series (SCI, volume 118). Berlin：Springer Berlin Heidelberg, 2008：251. DOI：https://doi. org/10.1007/978-3-540-78488-3_15.

［26］ PLOTNIKOVA V, DUMAS M, MILANI F. Adapting the CRISP-DM data mining process：A case study in the financial services domain［C］// CHERFI S, PERINI A, NURCAN S. Research Challenges in Information Science：Lecture Notes in Business Information Processing. Cham：Springer International Publishing, 2021. 55. DOI：https://doi.org/10.1007/978-3-030-75018-3_4.

［27］ OpenStreetMap. org. Map data ［DB/OL］. ［2021-05-20］. https://www.openstreetmap.org.