Full-Cycle Data Analysis Platform Construction Scheme for the Truck Industry

Jinhuan Zang, Heling Mao* and Lu Zhang

{*Corresponding author: maoheling@catarc.ac.cn}

{ zangjinhuan@catarc.ac.cn, zhanglu2022@catarc.ac.cn}

China Auto Information Technology (Tianjin) Co., Ltd., Tianjin, China

Abstract. Due to the popularity of data acquisition tools, data will be generated throughout the life cycle of automobiles. Key information needs to be extracted to reflect market changes, competitive enterprise dynamics, technology trends from the data. In this study, data processing processes and tools in the industry are investigated, and key technologies such as data dimension reduction, data fusion and data governance are integrated by using the coupling relationship of various links to connect data resources with each other through ID identification. Finally, the visualization platform and analysis report can be used to condense data analysis results and expert wisdom, provide decision support for product planning and market strategy of automobile and parts enterprises, and digitally empower the value chain of automobile.

Keywords: Truck Industry Data, Big Data, Lake Warehouse Integration, Big Data Platform Construction

1. Introduction

With the achievements of digital technology in various industries across the country, it is the only way for various industries to demonstrate the development trend of the industry with digital tools and make corporate decisions more based on the current economic and technological environment. Each process of the truck industry will generate data, and the data of each link will be collected by portals, apps and other channels, and eventually form data resources, which are dispersed among different data holders such as truck enterprises, dealers, and maintenance factories. Therefore, using data management tools such as data dimensionality reduction and data batch processing to build a data lake covering the whole life cycle of trucks, and provide decision-making support for enterprises through data analysis, is the hot spot of current data processing practice.

At present, there are abundant data collection and analysis studies on the truck industry at home and abroad. Cui Xin's research[1] is based on the data generated in the process of consumer use, and creatively combines the static attributes and dynamic driving behavior characteristics of vehicles to establish a mathematical model, and finally gets a vehicle portrait, so as to help manufacturers understand the

ICSETPSD 2023, November 17-18, Coimbatore, India

DOI 10.4108/eai.17-11-2023.2342644

use scenario of the vehicle. Malikov A. A.[2] and Miguel Pablo Martinez de[3] take full advantage of customer feedback to propose measures to respond more quickly and efficiently to customer vehicle requirements. The research of the above scholars focuses on some data types of the truck industry, but the truck industry has many data types, and different types of data need to be analyzed after coupling relationship.

After the realization of data collection, it is necessary to provide data services for the demand side of data in an appropriate way. Huang Yajuan[4] and Bai Xuesen[5] realized data service based on visual interface by using visual large screen and remote monitoring cloud platform respectively.

The existing research focuses on the analysis of data use in the truck industry from the perspective of modeling analysis and visualization platform. However, it has not been able to open up the data, especially the current truck industry data sources throughout the whole life cycle of the truck, and the data generated by different links is difficult to retrieve each other. Based on the data coupling relationship, this study connects the data of the whole life cycle, and then builds a big data decision support platform.

2. Big Data Platform Construction Contents

2.1 Data Sources for the Truck Industry Connected Big Data Platform

2.1.1. Truck Industry Data

Truck industry data mainly refers to the data generated during the development of the truck industry. Collectively, truck industry data includes data generated during the life cycle of trucks, such as design, research and development, production, sales, use, replacement and end-of-life. By analyzing the data of truck life cycle, we can effectively integrate the Internet technology with truck Industry[6].

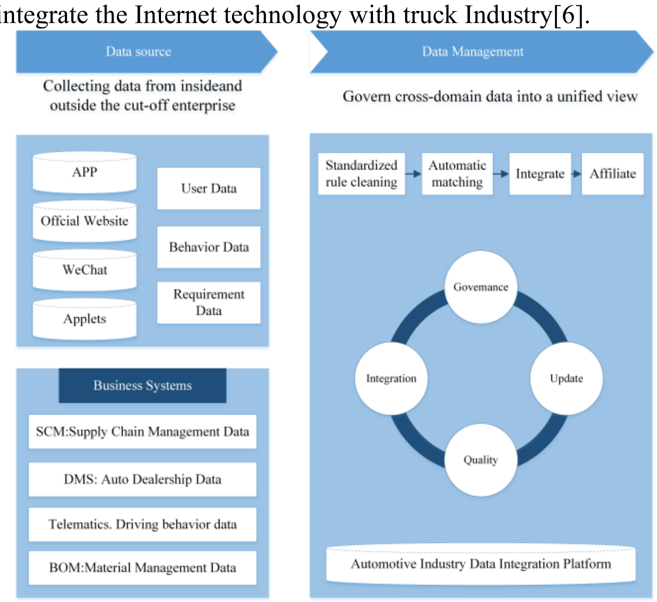

Fig. 1. Source and integration of industry data

The source of data for the truck industry is partially in the data hub within the truck enterprise (e.g. customer data, production data, design and R&D data, material data, etc.) A portion is stored outside the truck enterprise, such as dealership self-developed data platforms, third-party web pages and programs, etc. (e.g. car sales data, car replacement data). As Figure 1 illustrates the two sources of data (in-enterprise and out-of-enterprise), the data comes from a wide range of sources, but all are unified in the truck industry data integration platform for processing and realizing data value.

2.1.2. Consumer Preference Data

The proportion of consumers trading in and purchasing additional cars is gradually expanding, and the demand for cars among the post-90s is growing. This segment of consumers is more concerned about personalized experience, so "customer retention marketing" based on consumer preference data is particularly important.

Consumer preference data mainly refers to the various behavioral data generated during the process of car viewing, as well as car purchase and use. By establishing online and offline digital user touchpoints and conducting data analysis on consumer preference data, the truck industry can be better developed to meet consumers' consumption needs.

2.1.3. Truck Market Data

Truck's market data mainly refers to the data generated by the truck industry in the process of production and sales. Generally speaking, the truck market data includes truck market share, truck models, truck sales, truck parts supporting, truck recycling rate, truck ownership and other related contents. The truck market data includes a wide range and more contents, which directly affects the construction content and quality of the connected big data platform of truck industry.

2.1.4. Government Policy Data

Government policy data mainly comes from government policy support or industry restrictions on the truck industry. Government policy data usually includes data on policy subsidies, industry taxes, carbon emissions, vehicle energy composition, policies and regulations of the truck industry. For each government policy, it is necessary to determine its favorable truck models, favorable enterprises, and separately organize and summarize, combined with market data, to present the impact of the policy on the market.

Table 1. Data characteristics and data processing methods

Data type	Data characteristic	Processing method
Truck Industry Data	Reflecting the industry situation, dynamic data and static data mixed	Data-cleaning, data fusion, and finally data analysis
Consumer Preference Data	Related to the real-time operation of the user and has strong immediacy	Real-time data batch processing
Truck Market Data	Collected after a certain period(Every week or month)	Data-cleaning, data analysis
Government Policy Data	Released irregularly, the interval is generally long, but the impact is	In-depth analysis combined with expert opinions

	large	

The processing methods of the above four kinds of data are shown in Table 1.

2.2 Requirements for Building A Connected Big Data Platform for the Truck Industry

In terms of hardware, the truck industry connected big data platform has to establish a large number of data computers, distributed data storage. On the software side, a fast and efficient data processing layer, data storage layer, and data service layer should be established[7].

2.2.1. Data processing layer

The data processing layer takes the data collected and stored in the DW layer and processes it to meet the needs of use and distribution in the business.

The tools available for the data processing layer are: custom Shell language packages, JAVA packages, Spark distributed storage, SQL jobs, etc.

```
update tmp_qa set CXJB='Car_A' where CXJB='Car_Compact';
update tmp_qa set CXJB='Car_A0' where CXJB in ('Car_Small','CAR_Small');
update tmp_qa set CXJB='Car_A00' where CXJB='Car_Mini';
update tmp_qa set CXJB='Car_B' where CXJB='Car_Medium';
update tmp_qa set CXJB='Car_C' where CXJB='Car_Large';
update tmp_qa set CXJB='MPV_A' where CXJB='MPV_Compact';
update tmp_qa set CXJB='MPV_C' where CXJB='MPV_Large';
update tmp_qa set CXJB='MPV_B' where CXJB='MPV_Medium';
update tmp_qa set CXJB='MPV_A0' where CXJB='MPV_Small';
update tmp_qa set CXJB='SUV_A' where CXJB='SUV_Compact';
update tmp_qa set CXJB='SUV_C' where CXJB='SUV_Large';
update tmp_qa set CXJB='SUV_B' where CXJB='SUV_Medium';
update tmp_qa set CXJB='SUV_A0' where CXJB='SUV_Small';

```

Fig. 2. SQL Job - Data Cleaning

As in Figure 2, a model-level mapping relationship is constructed based on the consensus of the automotive industry (e.g., data mapping of Compact ↔ Car Compact ↔ Car_A) and cleaned with SQL job's update method in conjunction with the needs of business analysis.

```
* sqoop_nev_allcyc_x...
#!/bin/bash
CONNECTURL=jdbc:oracle:thin:@10[redacted]
#
ORACLENAME=autoserver
#
ORACLEPASSWORD=[redacted]
oralceTableName=TMP_XNY_2023
#
columns=xliang,nian,yue,province,city,qxian,qhua,jb,sfxgou,developedtier,region,provinceen,
,syquan,syxzhi,dbjb,dbjbxf,dblevel1,dblevel2,dblevel3,xkjb,sgmw_level,yfljb,clxh,jt,zzshang,
,qymc,ppai,cxing,cxing2,zzshang_en,ppai_en,cxing_en,xfschang_db,xfschang,cxxshi_db,cxxshi,
,cxjbie,cxjbie2,ppsx,cxi,sfdsc,qdxsxf,qdxs,cchang,ckuan,cgao,zju,zzl,zbzl,cyshu,zgcs,mfdj,mfqy,
,rlzl_en,rlzl,pliang,plml,nj,fdjwz,fdj_jqfs,pfbz,gl,zhyh,bsq_xh,bsq_cj,bsqlxing,bsqlxing2_final
,dwei,zdsjia,zgsjia,lts,xgltgg,ltlx,ql,hl,tk,bpb,lwcc,fhzs,sdjb,dczl,dczzzlnum,dczzzl,dcrl,
,maxdcrl,dczcqy,dczcdy,maxdczcdy,dcdtxh,dcdtdy,dcdtqy,djxh,djlx,djsl,djqy,djlqfs,djkzqqy,
,zckzqlx,zckzqqy,bmsqy,czcdjxh,czcdjqy,sfwjcd,sfykhzz,xslcgk,maxxslcgk,xslcds,maxxslcds
#
hdfsPath=/webdata/data/$oralceTableName
#
$SQOOP_HOME/bin/sqoop import -D mapreduce.map.memory.mb=4096 -D mapreduce.map.java.opts=-
-Xmx1024m --append --connect $CONNECTURL --username $ORACLENAME --password $ORACLEPASSWORD --
-target-dir $hdfsPath  --num-mappers 1 --table $oralceTableName --columns $columns --verbose --
-fields-terminated-by '\001'
```

Fig. 3. Custom Shell Language Packages - Data Pulling

Data validation must be performed at all times during each step of data processing, customizing the data for correctness validation, fill rate validation, and value range validation. After the checks pass the tables are pushed or pulled into the distributed storage structure using ETL tools. The "data pull" is the connection to the data source in the target server, For example, if the storage tool of the data source is Oracle, you can write the url configuration using jdbc to connect to Oracle and define the fields and storage location for pulling the data in the Shell language package to pull the data to the local storage medium such as HDFS, as in Figure 3.

"Data Push" enables connection from a data source to a target database and pushing it into the target database. Figure 4 implements the same table transfer from HDFS to the message queue using Hive Sql.

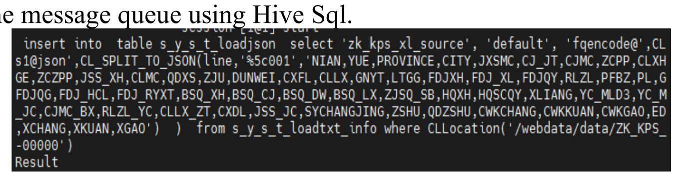

Fig. 4. SQL Job - Data Push

2.2.2. Data storage layer

After data collection is completed, we must focus on data storage construction to ensure that the collected data and analyzed data can be effectively stored and utilized. Use Kafka, MetaQ, Notify and other message queue tools, and MySQL, Redis, HDFS and other data storage tools to achieve different kinds of data classification storage and aggregation. At the same time in the storage layer need to build data computing model to meet the needs of the truck industry for data computing operation[8].

2.2.3. Static data and dynamic data integration

Big data in the truck industry is divided into static data and dynamic data according to the frequency of updating and collection. Static data is data that is processed every N days(N>1),Delayed time granularity is "days". For example, the production, sales and inventory data of truck industry, dealers may take inventory once every few weeks or even months; the update of government policy data is also in "days". Static data can be called and transferred by writing a data interface under the premise of the original data storage method[9].

Dynamic data refers to data that is updated in "seconds" or even "milliseconds", such as consumer web browsing data, user driving behavior data, etc[10]. As shown in Figure 5, the logic of how static and dynamic data are collected and eventually applied separately is shown.

Static and dynamic data integration based on message queue tools such as Kafka, MetaQ, Notify, etc., data storage tools such as MySQL, Redis, HDFS, etc., and data migration tools such as Hive Sql, Spark Sql, etc. The final data model is selected for calculation based on the data application requirements.

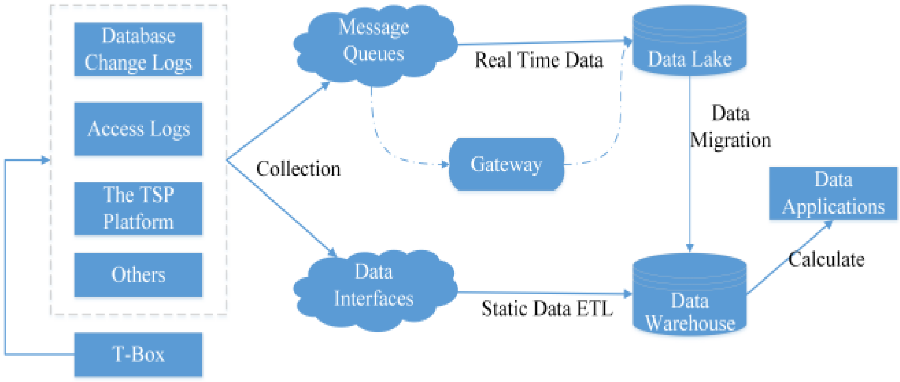

Fig. 5. Data integration logic

In order to connect data from various data sources, ID identification technology is needed. There are many ID identification options for customer big data, including name, ID, phone number, mobile device number, etc. However, all of them have their limitations. Using name as ID will inevitably lead to renaming and name change, so it is impossible to qualify who is the target of data generation. The same person may use different phone numbers in two operations, which is not conducive to the complete establishment of a large database of consumer behavior, and mobile device numbers face the same problem. Only the ID number can avoid the above problems, but it is difficult to be used by enterprises because it is more difficult to collect.

ID-Mapping technology can achieve a comprehensive and accurate positioning of customers: the above-mentioned types of ID identification is distilled into "One-ID", as shown in Figure 6.

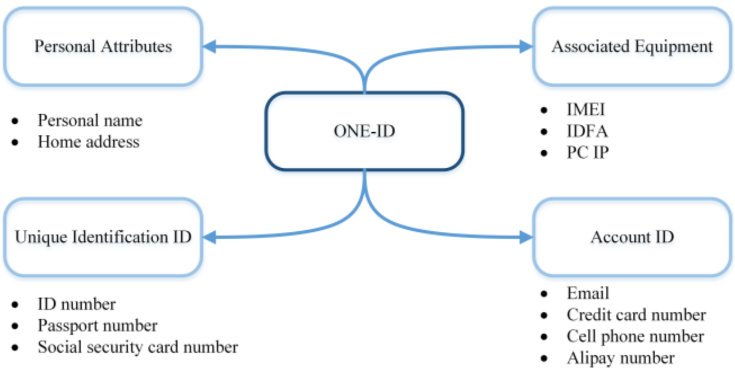

Fig. 6. One-ID identifies the user

Not only the consumer data can be precisely positioned with ID-Mapping technology, but also corporate users, dealers and other data subjects can be distinguished by building One-ID, which unifies the identification methods like a "puzzle" and ultimately realizes the efficient use of data..

After the establishment of One-ID, the vehicle in the sale, use, maintenance of the process of the vehicle identification ID and user (personal users, corporate users) associated (such as Figure 7), the entire life cycle of the vehicle for data management, build a complete user portrait[11],This will improve the production and operation of truck companies.

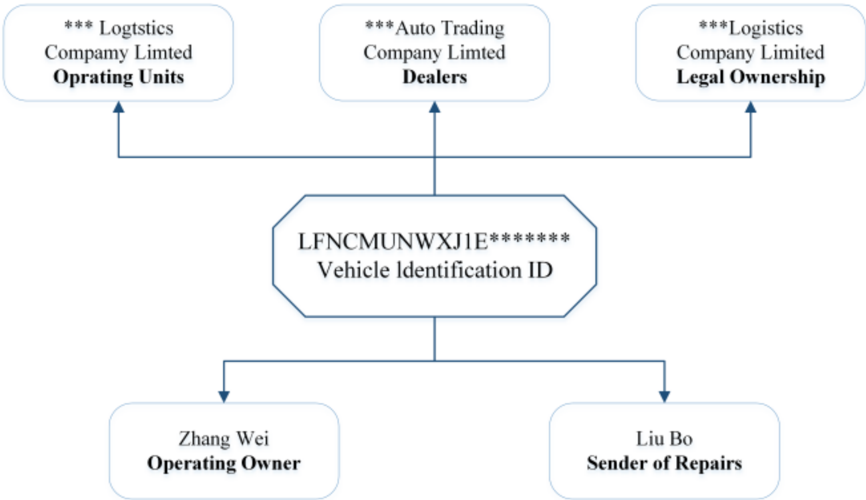

Fig. 7. User Big Data based on user One-ID

3. Data Analysis Platform Construction Solution

3.1 Visualization Platform Construction

In the era of big data, the use of connected big data can prompt the effective integration of disordered data, providing decision support for the development of the truck industry. Visual data technology can present abstract data in a visual way, Simplify and facilitate complex data, so visual data technology has been effectively developed in the truck industry.

3.1.1. Visualization of big data for vehicle monitoring

Data of vehicle monitoring can be uploaded by T-BOX to the driving big data monitoring center. If the car company has already built a platform, the data from the original platform can be forwarded, verified, parsed and monitored in a unified manner[12].

Eventually, the monitoring results are visualized and monitored by monitoring departments (personnel), such as truck companies and state regulators, respectively. At the same time, the diagnosis and test results will be forwarded to the user's mobile phone/car phone to remind the user of driving safety, as shown in Figure 8.

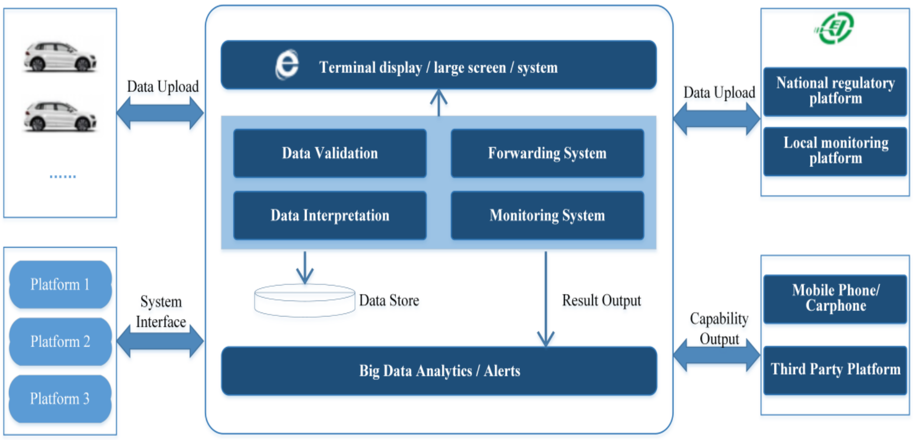

Fig. 8. Establishing a unified monitoring center

3.1.2. Visualization of big data in the vehicle market

Existing truck market data sources are partially within the truck companies, such as order flow data, etc. Part of it lies outside the truck industry, such as dealers, third-party web pages (programs), etc. The data sources are more complex and the data analysis is more difficult. The visualization platform can simplify and facilitate the complex and massive data, and can effectively use the collected information and data for different interaction scenarios. At the same time, the raw data generated from databases, cloud data, and spreadsheets can be vividly presented to the analyst in the form of images, Enables more efficient analysis of data and information in a short period of time, Improve the efficiency and accuracy of data analysis, so as to achieve the prediction of the development of the truck industry and better grasp the development direction of the industry [13].

Through visual analysis of the collected data, we can calculate and compare year-on-year, market share, sales volume and other related elements. Present complex data to the analyst in a simple way through dynamic charts to better analyze it[14]. In addition, the results of the data analysis can be displayed directly through a visualization platform, corresponding data or elements can be downloaded and exported in the visualization platform, help relevant decision makers with data support and reduce unnecessary time consumption. The visual data platform is always in dynamic balance and can be changed according to the data collection in a timely manner to ensure that the analysis results are timely and effective, Able to grasp the development direction of the truck industry in a short period of time.

In addition to sales volume, several dimensions such as production and inventory can be used to monitor the overall development of the truck market and provide data support to leaders for the development of the truck industry.

3.1.3. Visualization of vehicle policy data

Policy data on trucks has a huge impact on the truck industry, from the government website and other channels to obtain and classify the policies that have a greater impact on new energy, intelligent network connection, and car purchase subsidies, analyze the favorable models and companies so as to dig out the current policy encouragement direction.

3.2 Truck Industry Development Decision Platform Construction

Truck industry development decision building needs to be based on the collection, analysis, monitoring and early warning of connected big data, By conducting an industry-wide analysis, it can effectively provide decision support for the development of the truck industry, enabling leaders to synthesize information from all aspects and make the most appropriate decisions in a short period of time. The construction of truck industry development decisions can be presented by means of a large data screen.

The large data screen can integrate and analyze the massive data of the truck industry to ensure the scientific and effective decision-making of the truck industry development. The large data screen is displayed in a dynamic way during the visualization presentation to ensure the timeliness, reliability and accuracy of the data of the truck industry development.

3.3 Mobile Data-side Monitoring Construction

Although the large data screen can show visual data information to decision makers in a comprehensive, timely and effective manner, it is difficult for the large data screen to deliver the development of the truck industry to the practitioners in a timely and effective manner, practitioners have difficulty making timely judgments about the direction of industry development. The use of mobile data monitoring App can effectively solve such problems.

Mobile App is a visualization tool that has emerged in recent years and is supported by the industry in terms of portability, ease of operation and experience. The establishment of a mobile data monitoring app for the truck industry can effectively improve the grasp of the industry's development direction. Generally speaking, the content of the data monitoring app is presented in the form of data reports, and the mobile terminal makes it as simple as possible while ensuring the amount of information, For example, in the core part of the mobile terminal, it can focus on directly displaying the information content related to the truck market such as sales volume, ranking, year-on-year, chain, market share, etc. to ensure that industry practitioners can grasp the dynamics of the industry development in a timely and effective manner.

4. Conclusion

This paper combines the data characteristics of the truck industry, and elaborates on the perspectives of data sources, data integration technology, and big data platform construction for the truck industry. The following construction points of the connected big data platform for the truck industry were derived:

(1)There are different data processing ideas for industry data, consumer preference data, truck market data and policy data, and appropriate collection and cleaning methods are adopted according to local conditions, and integrated on the platform through the architecture of "Lake Warehouse as One".

(2)ID-Mapping technology, as the enterprise's unified code for data generation objects, plays a role in guaranteeing the data circulation of different databases and data systems, and has great significance in the acquisition and integration of data outside the enterprise.

(3)The platform is based on the collected and cleaned data, and establishes a three-layer architecture of processing-storage-service, applying mathematical model algorithms, and the final generated results need to be provided to the regulator or individual users in the form of visualization.

References

[1] Cui Xin. Research on the application of vehicle profiling technology based on big traffic data. Information Technology and Informatization,2022(01):126-130.

[2] Malikov A. A., Kozlovskii V. N.,and Vasin S. A.. Tools for Timely Response to Customers' Quality Requirements in the Auto Industry. Russian Engineering Research, 2022, 42(3): 295-300.

[3] Miguel Pablo Martínez de, DePablosHeredero Carmen, Montes Jose Luis, et al. Impact of Dynamic Capabilities on Customer Satisfaction through Digital Transformation in the Automotive Sector.Sustainability, 2022, 14(8): 4772-4772.

[4] Huang Yajuan. The use of big data analysis technology in the new energy vehicle industry. Auto Time,2020(07):69-70.

[5] Bai Xuesen. Research on new energy vehicle operation evaluation based on big data analysis. Xiamen Science & Technology,2022(03):35-38.

[6] Giacosa Elisa, Culasso Francesca, Crocco Edoardo. Customer agility in the modern automotive sector: how lead management shapes agile digital companies.Technological Forecasting & Social Change, 2022(175):121362.

[7] Anita P, Gabriella T. Covid-19 and transformational megatrends in the European automotive industry: Evidence from business decisions with a Central and Eastern European focus. ENTREPRENEURIAL BUSINESS AND ECONOMICS REVIEW, 2021, 9(4): 19-33.

[8] Martin K. Automation, digitalization, and changes in occupational structures in the automobile industry in Germany, Japan, and the United States: a brief history from the early 1990s until 2018. Industrial and Corporate Change, 2021, 30(3): 499-535.

[9] Lüthje B. Going digital, going green: changing production networks in the automotive industry in China. International Journal of Automotive Technology and Management, 2021, 21(1-2): 121-136.

[10] Sadia R,Lingyun Y,Junxiu T, et al. Exploring Effective Relationships Between Visual-Audio Channels in Data Visualization. Journal of Visualization, 2023, 26(4): 937-956.

[11] Thoralf R,Alexander K,Sebastian B, et al. An Information System Supporting Insurance Use Cases by Automated Anomaly Detection. Big Data and Cognitive Computing,2022,7(1): 4.

[12] C. T,O. O,C. E, et al. A Digital Maintenance Practice Framework for Circular Production of Automotive Parts. IFAC-PapersOnLine, 2020, 53(3): 19-24.

[13] Nir G,Eliya L,Polina L, et al. Use of a Novel Three-dimensional Head-mounted Digital Visualization Platform in Corneal Endothelial Transplantation.. Ophthalmology and therapy,2022, 12(1): 625–631.

[14] Pelle Anita, Tabajdi Gabriella. Covid-19 and transformational megatrends in the European automotive industry: Evidence from business decisions with a Central and Eastern European focus. ENTREPRENEURIAL BUSINESS AND ECONOMICS REVIEW, 2021, 9(4): 19-33.