Research on Multi-dimensional Data Collaborative Virtual Inventory Based on Markov

Shunkai Sun, Jie Li, Haihua Lu, Qi Xu*, Haobo Cui, Weilin Cao, Yingxiao Chi and Jun Zhou

Ningbo Cigarette Factory, China Tobacco Zhejiang Industrial Co., Ltd., Ningbo 315504, China

*Corresponding author e-mail: Xuq@zjtobacco.com

Abstract. The three-dimensional elevated warehouse is a production equipment, and it is inevitable that it will malfunction during the operation. If it is not handled properly, there will be inconsistencies between information and practice, resulting in inconsistency between inventory information and the actual product. In order to solve the above problems, this project team carried out functional research on various SQL tables of multi-dimensional data collaboration, and constructed a virtual inventory method based on Markov multi-dimensional data collaboration. The behavioral characteristics of the cargo location information in the cargo location table TWMS_LOC and the pallet status table TWMS_PLT are extracted, and the two are compared. If the deviation between the two exceeds the threshold, the status is judged to be no cargo information. Experiments are conducted to test the feasibility and effectiveness of the sentences of the proposed virtual inventory. The results show that the inventory method can better describe the status of the cargo location and information, and effectively realize the virtual inventory at the database level. The research results of this method will completely replace the complicated manual inventory business. It can provide accurate and efficient virtual inventory solutions for similar logistics elevated warehouses.

1. Introduction
At present, academia and business circles have invested a lot of energy in inventory and have achieved fruitful research results. However, the research on database virtual inventory is relatively weak. However, in the overall inventory of the logistics information system, the database is often the most worthy of study. A lot of pallet information and cargo location status are stored in the database. The traditional warehouse management system database is only used for logistics data access. The potential of the data to play a role is far from being explored, and it cannot effectively support the inventory behavior. In particular, it can automatically identify the goods and incoming goods in the designated cargo space according to the needs of the warehouse management personnel. Whether the library time is consistent, so that the demand for smart inventory is very urgent. Therefore, the study of virtual inventory for multi-dimensional data collaboration in database systems is particularly important.

This paper proposes a Markov-based multi-dimensional data collaborative virtual inventory method. This method uses the SQL statements submitted by the user as user behavior characteristics, and uses
Markov to extract and analyze, then judge through the calculated threshold, and finally use the relevant
The experimental data set is tested, and the results show that the system in this paper can effectively
realize the virtual inventory work of the high-bay warehouse.

2. Markov chain
Given a random process $X_t$ if the state of the random process $X_t$ at a certain time $t_0$ is known, the state
of the process $X_t$ at a later time $t (t > t_0)$ has nothing to do with the state of the process before the time $t_0$.
Under the known present conditions, the nature of future changes in the process has nothing to do with
the past, called Markovian or no aftereffect. The random process with such no aftereffect is called
Markov process.

Given a random process $\{X_t, t \in T\}$ if the conditional distribution function of the process exists, and
for any $n$ time $t_i$ in the parameters, $t_i, i=1,2,\ldots,n, t_{i-1}<t_i<\cdots<t_n$

$$P\left\{X_{t_n} \leq x_n \left| X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_{n-1}} = x_{n-1}\right.\right\}$$

$$= P\left\{X_{t_n} = x_n \left| X_{t_{n-1}} = x_{n-1}\right.\right\}$$

The stochastic process $\{X_t, t \in T\}$ is called the Markov process. Here $T$ is the parameter space of the
process, $x_n \in E, E$ is the state space of the process.

3. Virtual Inventory Analysis
In this paper, the virtual inventory is the query behavior of querying the database through multiple tables.
It is identified by querying multiple dimensions of the pallet table, pallet schedule, task table, etc. In the
database system, users mainly use SQL statements Interact with the database management system to
complete information query, modification and deletion operations. By analyzing the execution sequence
of SQL statements, the behavior characteristics of the virtual inventory can be described more
comprehensively. The advantage of analyzing SQL statements is that when the query is processed by
the abnormal operating system, the load of the database virtual inventory is lighter. A new SQL
statement is analyzed and passed some statistical model checks, and only runs in the database
management system when the query is accepted by the abnormal operating system.

In the virtual inventory in this article, the collaborative virtual inventory based on Markov multi-
dimensional data mainly includes two processes, namely the learning process and the operating process.

1) During the learning process, make the warehouse management system run for a period of time
under normal conditions, collect data during normal operation, extract the characteristics of the
TWMS_LOC position table, and establish a normal warehousing system position status.

2) In the process of operation, make the warehouse management system run in the actual environment,
collect the location information in the current pallet status table TWMS_PLT and extract features,
compare the location status with the location information, and observe the status and information
through comparison The degree of deviation of the feature determines whether an abnormality has
occurred.

The abnormal operation of the degree of deviation between the state of the warehouse database and
the information characteristics proposed in this paper mainly includes three stages:

1) Data preprocessing stage. At this stage, the SQL statements submitted by the warehouse
management system to the database, that is, database behavior prediction, are collected and preprocessed.

2) Learning stage. Analyze the characteristics of the cargo location table at this stage to obtain the
normal behavior pattern of the elevated warehouse. The flowchart of the learning phase and the
operation phase is shown in Figure 1. In the learning phase, the data collected during the normal
operation of the elevated library over a period of time, such as SQL statements in the log file, are
extracted from the normal behavior of the elevated library as a historical behavior pattern. Extract the
behavior characteristics of the learning sequence, establish a normal sequence library, use
POSTGRESQL to obtain the Markov chain state set of the learning sequence, and calculate the transition probability matrix of the Markov chain state to describe the normal behavior of the elevated library.

3) Operational stage. Collect operating data of the operated elevated warehouse at this stage. Extract the behavioral features in the operating data as the behavioral features of the currently operated elevated library. Compare the behavior characteristics of the operated elevated warehouse with the historical inventory characteristics to determine whether the current behavior is an abnormal inventory. If the deviation between the two exceeds a certain threshold, it is considered abnormal; if the deviation between the two is within the normal range, the behavior is considered normal.

4. Demonstration of results

4.1. Test indicators
The operation result of an inventory on abnormal inventory is generally measured by three parameters: query rate, error rate and miss rate.

Operation rate, the operation rate refers to the percentage of abnormal data generated by the operation,

\[
\text{Query rate} = \frac{\text{Check the number of goods}}{\text{Total number of cargo spaces}} \times 100\%
\]

False alarm rate, false alarm rate refers to the rate of normal data being mistakenly detected as abnormal data,
Wrong inventory rate = \frac{\text{Actual inventory digits}}{\text{Total number of inventory positions}} * 100\%

The underreporting rate, the underreporting rate refers to the ratio of abnormal data that has not been manipulated,

Missed inventory rate = \frac{\text{Missed inventory digits}}{\text{Total number of inventory positions}} * 100\%

In the signal operation theory, it is used to express the trade-off between the operation rate and the false alarm rate. This article will also use the curve to analyze the performance of the warehouse management system. The point in the upper left corner of the curve represents a 100% operation rate and a 0% false alarm rate. Therefore, the closer the curve is to the upper left corner, the better the performance of the abnormal operation warehouse management system.

4.2. Experimental data

The test data set used in the experiment is a database collection of a clinic, including 39,322 SQL commands such as select, insert, update, and delete. The raw data collected from the database log file is the SQL command line. SQL commands contain a large number of command fields such as operation type, operation object, and table name. In order to improve operation efficiency, SQL statements need to be preprocessed.

The database SQL statement set \langle \text{OP, F, T} \rangle, where \text{OP} is the statement type of the SQL statement, \text{F} is the attribute set involved in the SQL statement, and \text{T} is the table set in the SQL statement.

Each sentence can be classified into a certain library sentence. In actual use, it is considered that as long as the sentence template is the same, the two sentences are of the same type.

SQL 1: select t.*, t.rowid from TWMS_LOC t where T.LOCSTORESTATUS <> 'Free'; SQL 2: select t.*, t.rowid from TWMS_LOC t where T.PRNTNUM in ('L0463_001%', 'L0463_002%', 'L0463_003%', 'L0464_004%', 'L0464_005%', 'L0464_006%', 'L0464_007%', 'L0464_008%'); SQL 3: select t.*, t.rowid from TWMS_LOC t where T.LOCNUM not in (select T.CURRLOC from TWMS_PLT T);

The templates of the 3 sentences are shown in Table 1.

|   | Virtual Inventory SQL Statement |
|---|---------------------------------|
| SQL | 1 | 2 | 3 |
| OP | select | select | select |
| F  | L0463% | L0463% | L0463% |
| T  | TWMS_LOC | TWMS_LOC | TWMS_LOC |

It can be seen from Table 1 that each SQL statement is processed into a statement template, and a sequence of statement templates \langle \text{OP1, F1, T1, OP2, F2, T2, ...} \rangle is generated in chronological order. All learning data and operating data are required. Follow the above process.

4.3. Argument results

This paper selects the data of 3 elevated warehouses for the experiment. If one of the 3 elevated warehouses is selected as the normal elevated warehouse, the other 2 elevated warehouses are abnormal behaviors relative to the selected elevated warehouse. Changes in SQL will cause changes in the behavior pattern characteristics of the warehouse management system. If the changed value exceeds the fluctuations within the normal range, the warehouse management system will recognize it as an abnormal behavior.

After preprocessing, the first 6970 SQL command fields of the inventory statement 1 are used to build PostgreSQL and Markov chains in the learning phase, and the remaining 3988 SQL command fields are used as test data in the operation phase to operate the warehouse management system as a normal behavior false alarm rate. Inventory statement statement 2, inventory statement 3, inventory
statement 4 each have 5000 SQL command fields for abnormal behavior to test the operation rate of the warehouse management system. The parameters of the warehouse management system are set to: N=6, W=7, l(1)=1, l(2)=4, l(3)=8, e(1)=3, e(2)=6, e(3)=9, w=145, δ =1.57.

| SQL | Inventory serials | SQL | Inventory serials |
|-----|------------------|-----|------------------|
| Statement1 | 2353 | Statement3 | 3234 |
| Statement2 | 3955 | Statement4 | 4742 |

Table 2. Number of different sentence database sequences

After experimental calculations, the transition probability matrix of the Markov chain of elevated library 1 is obtained as

\[
P = \begin{pmatrix}
0.024 & 0.246 & 0.009 & 0.345 & 0.125 & 0.251 \\
0.2 & 0.024 & 0.451 & 0 & 0.113 & 0.212 \\
0.091 & 0.083 & 0 & 0.079 & 0.413 & 0.334 \\
0.64 & 0.131 & 0.141 & 0 & 0.088 & 0 \\
0 & 0 & 0.92 & 0.034 & 0.001 & 0.045 \\
0.34 & 0.088 & 0.123 & 0 & 0.27 & 0.179
\end{pmatrix}
\]

5. Conclusion

Aiming at the problem of abnormal behavior of the inventory in the database warehouse management system, SQL statements are used to extract the behavior characteristics of the inventory, and the state of the discrete Markov chain is used to describe the behavior profile of the inventory, and a POSTGRESQL-based Abnormal behavior of the database inventory operates a warehouse management system, which includes a data preprocessing module, a data learning module, and an abnormal operation module. In order to verify the effectiveness of the warehouse management system, the database log data is used for experimental testing. The experimental results show that the state of the discrete Markov chain can be used to describe the behavior profile of the database inventory, and the judgment value curve of normal behavior and the judgment value curve of abnormal behavior can be clearly distinguished, thus proving the effectiveness of the warehouse management system sex. Therefore, the warehouse management system in this paper has certain practical application value in the field of abnormal behavior operation of database inventory.

References

[1] Database PL/SQL Packages and Types Reference [EB / OL], https://docs.oracle.com/cd/E11882_01/appdev.112/e40758/preface.htm
[2] Veronique L. Roger, Alan s. Go, Donald M. Lloyd-Jones, etal. Heart disease and stroke statistics update:a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee[R]. Circulation, 2020, 125:2－220.
[3] Lei Zhang, The research and design of SQL processing in a data-mining system based on MapReduce, Cloud Computing and Intelligence Systems, 2019,301-305.
[4] Yimeng Liu ; Yizhi Wang ; Yi Jin, Research on the improvement of MongoDB Auto-Sharding in cloud environment, Computer Science & Education, 2012, 7th, 851–854.
[5] Chandra, D.G.; Prakash, R., A Study on Cloud Database, Computational Intelligence and Communication Networks, 2019, 4, 513–519.
[6] Wylie, B.; Dunlavy, D., Using NoSQL databases for streaming network analysis, Large Data Analysis and Visualization, 2017,121–124.
[7] F. Chang, et al, Bigtable: a distributed storage system for structured data. OSDI, 2016.
[8] Zhu Wei-ping; Li Ming-xin ;Chen Huan , Using MongoDB to implement textbook management system instead of MySQL, Communication Software and Networks, 2020 , 3th,303-305.