

Application of SQL in hospital statistics

Xu Lixiang

*Linshu County People's Hospital, Linyi, China
XLX19921214@163.com*

Abstract: *in hospital statistics, due to the difference of data query language, the hospital statistical results are often quite different from the actual data. Therefore, the application of SQL in hospital statistics is proposed. Use java + spark SQL to read the data information to be counted in the hospital, and have an in-depth understanding of common basic tables and fields, so as to facilitate the smooth completion of subsequent statistical work. Using structured query language, the statistical indicators such as postoperative complications, disease ranking and workload of the surgeon are written into SQL statements, and finally the SQL statement integration results are presented in the form of tables. The experimental results show that compared with the two statistical methods proposed in the literature, the statistical error is reduced by 30.25% and 35.61%.*

Keywords: *Structured Query Language; hospital; Statistics; Field; SQL language; Data processing*

1. Introduction

Due to the rapid development of the big data era, how to count the required data in a large number of data resources has become the most important work. Especially the executor of statistical work in grass-roots hospitals [1], because of the particularity of hospitals, the complexity of hospital data is higher. Extracting effective statistical data from complex data has become a major problem faced by statisticians. Usually, all kinds of data required in the hospital are extracted one by one by professional statisticians using the query function of management software [2], and the statistical results are generated in the form of tables. The efficiency of this statistical method is too low. To solve the above problems, this paper proposes an innovative statistical method with SQL as the core. Relying on computer technology to read hospital information, database technology and structured query language are adopted to realize high-quality information statistics and meet different types of hospital data statistics and queries to the greatest extent.

2. Design of hospital data statistics method

2.1. Reading hospital statistics

The basis of data statistics is to read the relevant data in the hospital. In this paper, java + spark SQL is applied to process the data [3]. Use computer software to create spark session service function, and use this as the basis to read data. Usually, the first column in the reading results of hospital data often represents the date. In order to reduce the difficulty of later language writing, the date column is transformed into a time node in milliseconds. The sampling time point of each hospital data has a corresponding time column, and the process data value of each time point is represented in other columns.

In data reading, attention should be paid to conventional statistics such as mean and mean square deviation. Under the action of conventional statistics, the stability of data statistics can be clarified [4]. Because there is no correlation between the conventional statistics and the time series, the meaningless time series can be removed in the conventional statistics, so as to reduce the statistical complexity.

2.2. Familiar with table structure and fields

When designing hospital statistical methods based on SQL language, it is necessary to focus on the analysis of common basic tables and fields. The database table structure in the process of hospital statistics is deeply analyzed to clarify the meaning of different fields in the table, so as to ensure its best effect in the subsequent application of SQL language. The hospital statistical method designed in this

paper mainly aims at the patient hospitalization information table and patient diagnosis information table. The former includes a variety of basic information records such as hospitalization times, admission date and discharge date, while the latter includes various information such as diagnosis type, disease name and ICD code. Take the hospitalization number contained in the above two tables as the associated field, and apply them to hospital statistics at the same time.

2.3. Writing SQL statements

The compilation of SQL statement is the core link in hospital statistics. As a comprehensive database query and program design, it is put into a structured language. Because of its own characteristics, SQL database is not affected by database system differences. Therefore, using SQL language as an auxiliary tool for data input and management can effectively improve the accuracy of hospital statistical results. In the process of key information statistics in many hospitals, the following three aspects need to be considered in writing SQL statements.

The first is to write SQL statements for postoperative complications. Patients' postoperative complications are a necessary statistical content in hospital medical quality monitoring, staff performance appraisal, and hospital grade evaluation. The list of postoperative complications is presented in the form of SQL query statements for the rapid implementation of hospital statistics.

Secondly, the disease ranking is written as SQL query statement, which is an important index to evaluate the performance of the hospital. Disease ranking is the main reference data for the development planning and design of the hospital, and it is also the basis for putting forward the department work plan. By analyzing the disease ranking query process, it is divided into four links. A temporary statistical table is formed according to the data statistics requirements, and then it is introduced into the hospital statistics database by using the category table contained in ICD10 database. Then, the temporarily generated table is combined with the imported category table to generate the current disease ranking information. Finally, save the ranking data obtained from the above operation, and delete the temporary statistical table. In this paper, the sub item method is used to realize the classification of hospital statistical ranking, add the sub item table in the hospital statistical database, and query the required content.

Finally, one of the main contents to complete the SQL statement is the operation workload of doctors at all levels. This index is applied to the monthly performance evaluation of the hospital to count the ranking data of operation volume at different levels. Under the function of SQL statement, the main table and operation table of statistical management database are generated based on the above data to promote the query and statistics.

2.4. Display statistical results

According to the SQL statement written above, obtain the SQL statement integration results according to the hospital statistical requirements. Import the integration statement into the ASP program, and ensure that users can input statistical requirements in the user interface to quickly get statistical results. The statistical results obtained by the user are presented in the form of Excel table. Based on the contents specified in the evaluation standard, judge the data on the home page and output the data information that meets the requirements.

Compared with the traditional hospital statistical methods, statisticians do not need to operate the same work content many times to obtain the required information. The statistical method proposed in this paper is convenient for the hospital statistical work to be completed with high efficiency and high quality in the era of big data.

3. Experiment

In order to observe the application effect of SQL in hospital statistics, experiments were carried out. The Cancer Hospital of Chinese Academy of Medical Sciences was selected as the experimental site. The data required for the experiment are the information of all surgical patients in the hospital from 2016 to 2020. The obtained data include disease diagnosis content, operation information, admission condition and other data. Because the most common complication after surgery is pneumonia, pneumonia is taken as a key statistical item for statistical test. According to the research, postoperative pneumonia refers to the pulmonary infection from the completion of the operation to the discharge of

the patient, and the qualified patient information is selected according to the diagnostic criteria of pneumonia. The experimental data were collected by SQL language and converted into a language that meets the needs of computer calculation, so as to obtain the total number of postoperative pneumonia every year. Because the statistical results often provide data support for the follow-up decision-making of the hospital, the quality of data statistics will have a great impact on the future work of the hospital. According to SPSS software, the statistical results are studied, and the data quality obtained by the statistical methods proposed in this paper is analyzed. At the same time, the hospital statistical methods proposed in literature [1] and literature [2] were used to obtain the number of postoperative pneumonia cases from 2016 to 2020, compare the errors between the three statistical methods and the actual values, and clarify the performance of different statistical methods.

3.1. Experimental results and analysis

Through the above experiment, the statistical result comparison diagram shown in Fig. 1 is generated according to the obtained statistical results.

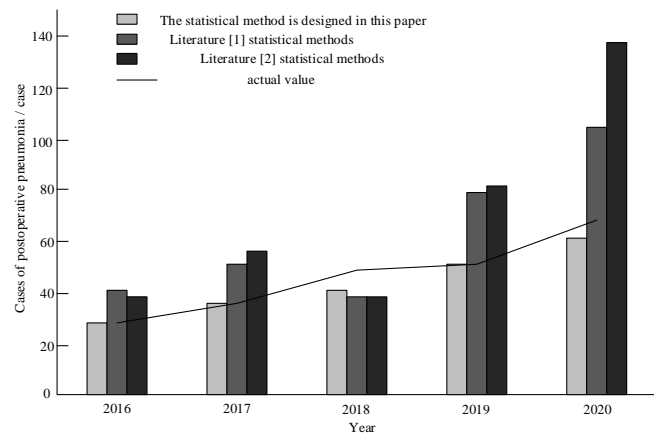


Figure 1: Comparison of statistical results of postoperative pneumonia with different methods

According to the statistical results in Figure 1, it can be concluded that the statistical method proposed in this paper is obviously closer to the real data than the two literature methods. This situation is mainly due to the differences of data query languages of the three statistical methods. This paper uses structured query language to compile relevant data information into a unified language to improve the quality of data statistics. The statistical method proposed in literature [1] takes field query as the core, and the method in literature [2] takes ICD coding query as the core. According to the statistical results, when relying on the field or coding statistical information, some similar diagnostic names, such as "pulmonary infection", "inflammatory pseudotumor", "inflammatory granuloma" and other diagnostic information, will be mistaken for postoperative pneumonia data, resulting in large statistical errors. According to the calculation of the average error of the statistical results, compared with the two statistical methods in literature [1] and literature [2], the average statistical error is reduced by 30.25% and 35.61%.

4. Conclusion

In China's medical and health reform, the development of hospital informatization is an important link of modern management. Due to the increase of the amount of hospital information, the existing methods can not meet the statistical needs. As a widely used and convenient tool, SQL can reduce the learning difficulty of statisticians, improve the quality of data statistics and better meet the needs of hospital statistics.

References

- [1] CHEN Wenjing. Application of digital management in hospital statistics [J]. Management informatization in China, 2018, 21(06):160-161.
- [2] WEI Wenyu. Application analysis of hospital management and statistical information of traditional Chinese medicine hospitals [J]. Continuing medical education in China, 2020, 12(35):103-105.

- [3] CHU Zijun. *Analysis of hospital statistical work mode in the era of big data [J]. Medical information*, 2020,33(21):24-25.
- [4] LAI Fuhu, ZHANG Zheng, XU Qingshi, et al. *Application of "statistical robot" in the field of hospital statistics [J]. Hospital statistics in China*, 2020, 27(05):476-480.