

A Kind Of Binary Coding Based Classification Method For Composite Information And Its Application

zhangchong, wangcai, caojingyao

Chengdu Neusoft University, China

*zhangchong@nsu.edu.cn

Keywords: binary coding, classification, composite information

Abstract. A new classification method based on binary coding is proposed for composite information consisting of a variety of basic information types, which takes the composition of composite information into account at the initial stage of design, so it has high efficiency and good scalability in search operation, and it is applied to a practical project successfully.

1 Introduction

Classification of goods is the most common thing in life, such as household appliances, clothing, food, etc. This is a typical classification mainly based on the nature of the article to differentiate from each other to facilitate management. The basic data types in programming languages, such as integral type, floating-point type, String, Boolean type, are a common classification in information data, which aims to improve the efficiency of computer storage and data search. But not all of the data can be uniquely divided into a classification, it may belongs to a number of different classifications. For this kind of composite data processing is relatively tricky, it will get a different efficiency according to the different classification design.

2 Binary coding based classification

There's such a requirement in the project I am involved in development - Sichuan Youth Football Information Management system. The provincial education department needs to hold a number of different levels of juvenile football matches, such as primary school football matches, high school football matches, and so on. The notice needs to be sent to the corresponding schools before a match begins. Schools according to the general classification, can be divided into primary schools, junior high schools, and so on, but often not so. For example, full secondary school, which both includes junior and senior high schools. As for the nine-year compulsory school, both includes primary and junior high schools, and for the 12 years consistent school, it even includes all the primary, junior and senior high school stages. In this case, how does we use a single field to represent a school type? A simple approach is to use a different code representation for each different type of school. But this is not an ideal design, since it does not take full advantage of the fact that different schools contain certain school stages, which will bring some complexity in later data operations. For example, to launch a province's junior high school football game, it is necessary to notice all the schools which contains the junior high school stage, and it also makes the search operation complicated. So, how to design a good classification for the school types which can bring high efficiency and easy operation for the future data retrieval? In view of this, this paper proposes a binary coding based classification method to classify this kind of composite information.

For the first step, we list all the school types as follows:

Table 1 school types

Serial NO.	School Type
1	Primary school
2	Junior high school
3	Senior high school
4	Full secondary school
5	Nine year compulsory school
6	12 Years Consistent School

We analyze the characteristics of these school types, although there are a variety of types, but in fact, all the school types are composite information combined by the primary, junior and senior high schools these three basic school stages. Therefore, we define these three stages as basic school types, assigning a different code for each type, which is represented by binary encoding, as follows:

Table 2 Basic school types

Classification	Binary School Type Code	Corresponding Decimal Encoding	Data type
Primary school	1	1	Integer
Junior high school	10	2	Integer
Senior high school	100	4	Integer

For other type of schools, because they are a combination of basic school types, the code can be determined by the sum of the basic school type codes which they contain, as follows:

Table 3 Composite school types

Classification	Binary School Type Code	Corresponding Decimal Encoding	Notes
Full secondary school	110	6	contains junior and senior high schools, with code $2+4=6$
Nine year compulsory school	011	3	contains primary and junior high schools, with code $1+2=3$
12 Years Consistent School	111	7	contains primary, junior and senior high schools, with code $1+2+4=7$

Then we get a complete list of school types, as follows:

Table 4 All school types

Classification	Binary School Type Code	Corresponding Decimal encoding	Notes
Primary	1	1	
Junior high	10	2	
Senior high	100	4	
Full secondary	110	6	contains junior and senior high schools, with code $2+4=6$
Nine year compulsory school	011	3	contains primary and junior high school, with code $1+2=3$
12 Years Consistent School	111	7	contains primary, junior and senior high schools, with code $1+2+4=7$

Since we use the power series of 2 to represent the basic school types, it can ensure that the composite school type code that is derived from any sum of the 2 power series will not be the same as the basic school type code, so all the school type codes are unique and different.

3 Classification Search

When the School's classification code is defined, it does not show any advantages compared to giving a random allocation of an integer for the school type code, since they have no substantial difference. But when we execute a data query operation, the advantage of binary coding based classification design begins to play a role. To Search all junior high Schools for example, we only need to use the school type code and junior high School type code to perform a binary bit AND operation. If the result is equal to the junior high school type code value, it is the school that qualified. In the Oracle database, the SQL statement is as follows:

```
SELECT Bitand (6,2) from DUAL;
```

Bitand () is a built-in function provided by the database Oracle that receives two integer parameters, it converts them to binary numbers, and executes them by bit AND operation, and finally gets the bit AND result value. In the example given above, the school type is 6, which indicates a full secondary school, when it does an AND Operation with the junior high School type Code 2, as follows:

```
0110B and 0010B =0010B;
```

The result is 0010B, or decimal value 2, equals to the Junior high school type code. It shows that full secondary school is one of the type we want. It contains Junior high school stage, which is consistent with the fact. If the query condition itself contains more than one school stages, such as querying all high school students (including junior and senior high school), then we only need to modify the school type value as input parameter to 6 , and if a school and its execution of the bitwise AND operation result is also 6 , that is the qualified school. From the example given above we can see, no matter how complex the composition of the school is, the computing formula or function keeps the same, which greatly simplifies the query operation of the database.

4 Conclusions

The binary coding based classification method proposed in this article is ideal for such a situation that the basic information types (generally within 10 or more) are not so many, but the combination of the basic information is very complicated. Because of its unique design in terms of classification code, it has an innate advantage in search for this kind of information. Even if we introduce new basic information types, no changes is needed in the previous design, we also can get the information we want which brings a good scalability. In addition, the bit operation based on binary has greatly improved the speed and efficiency of database query operation. However, if there are too many types of basic information, the classification code values represented by the power of 2 will grow very quickly, and the method is less applicable, so you need to analyze the original information to extract as few basic features as possible. The classification method has been applied to Sichuan youth Football Information Management System project, and it did a good job in searching for different types of schools from the database.

Acknowledgment

This work is supported by Sichuan province Youth Football Information Management system project (No. 16WT0216), all support is gratefully acknowledged.

References

- [1] Cao Jianjun, Shinchun, Wu Jianming, Shang, Peng. An incomplete record classification detection method based on bitwise operations [J]. *Systems Engineering and electronic technology*, 2010,1(6): 2489-2492.
- [2] Wang Yuyi. *Information Classification and encoding* [M]. National Defense Industry Press, 2003, 39-43.
- [3] Gufahi, Lippman, Lai Luyan. Research on information-sharing information classification coding standard system and implementation process [J]. *Intelligence Magazine*, 2008, 27 (4): 25-28.
- [4] Wu Junjie. Encoding of information: four-bit binary encoding [J]. *China Information Technology Education*, 2014 (5): 77-79.
- [5] Zhang Dongyang, Li Shi, liwenbin. Research on a novel binary coding theory [J]. *Dimension Computer Information*, 2010, 26, 195-196.