Credit Rating System for Power Consumers Based on Integrated Payment Platform

Yan SHI*, Wen-jie LI and Bao-feng XU
East Inner Mongolia Electric Power Co., Ltd. Power, Research Institute, Hohhot, Inner Mongolia, China

*Corresponding author

Keywords: Integrated payment platform, Consumer rating, Credit rating.

Abstract. With the continuous deepening of power reform, power companies have gradually entered the era of market economy, and the pressure on business operations has gradually increased. However, power companies have not been effective in dealing with malicious defaults, evading electricity charges and stealing electricity. How to effectively avoid the risk of arrears brought by customers and improve the economic efficiency of enterprises has always been a problem faced by power companies. This paper mainly studies the credit rating system for power consumer based on the integrated payment platform, and proposes an algorithm model, which combines similar SQL statements with the platform, and uses the big data software framework and data processing technology to quickly construct user images. It also assesses the key technologies in consumer rating and its partitioning and evaluation, and the results will be showed, which laid the foundation for further research.

Preface

With the continuous deepening of power reform, power companies have gradually entered the era of market economy, and business pressures have gradually increased. However, power companies have no effective preventive methods for malicious defaults, evasion of electricity charges and electric leakage. How to effectively avoid the risk of arrears brought by customers and improve the economic efficiency of enterprises has always been a problem faced by power companies. In order to better classify the user's consumption habits and behavior habits, and extract all kinds of effective information of customers on the basis of big data for analysis and visualization. It provide guarantee for precision marketing, and provide scientific guidance for making decisions. According to natural attributes, behavioral data, channel data and data classification, the author uses big data software framework and data processing technology to quickly construct user portraits, user rating, credit rating and evaluation of power users conduct an evaluation and demonstrate the results of the implementation[1].

Data Processing

Customer Data Collection

Carrying out customer portraits requires analysis and classification of individual customers, which is based on the foundation of big data. Big data technology means that we can group and image customers in an automated and scaled manner through the screening, integration and analysis of a large amount of information, so as to obtain conclusive information that can guide decision-making. With the advancement of the construction of power integration payment platform, the integrated payment platform system has been widely used in daily marketing operations, and the basic data is used as an information carrier in the construction of integrated payment platform system. From the customer's business expansion registration to the electricity bill accounting summary statistics, from the customer payment to the accounting department's electricity costs, every customer and electricity bill is recycled into a piece of data stored in the system. The data of customer personal information,
electricity information and payment information is interwoven into the network. With the advancement of time, it is increasing, so this is the “big data of power supply marketing system.” In addition to the integrated payment platform database, the business department can obtain customer electricity behavior information through market research, on-site electricity inspection, and simple statistical analysis of business data as a database.

Data Preprocessing

Customer clustering is an important analytical tool for customer behavior analysis. Customer clustering is the aggregation of a large number of customers into different classes. Customers in each category have similar attributes, while customers in the same category have different attributes [2]. Careful and practical customer clustering is of great benefit to the company's business strategy. The SQL statement classification algorithm is a self-organizing clustering algorithm for analyzing customer behavior. This method makes the data easier to visualize, and it highlights the characteristics of attracting people's interest. The number of cluster centers is automatically generated from the data.

Indexing and Selection

SQL-like SELECT statements mainly include the following function calls:

- loc - index row data through row labels
- iloc - get row data by line number
- ix - combine the first two hybrid indexes
- Importcopy

```python
df_titanic_tmp = copy.deepcopy(df_titanic.iloc[0:5])  # Select the first five columns as an experiment
df_titanic_tmp.index = ['a', 'b', 'c', 'd', 'e']  # Modify the index name
df_titanic_tmp.columns

df_titanic_tmp.rename(columns={'PassengerId': 'haha'}, inplace=True)  # Modify column name

# Select all rows of a row
df_titanic_tmp.loc['a']
df_titanic_tmp.iloc[1]
df_titanic_tmp.ix['a']
df_titanic_tmp.ix[1]

# Select all rows of a column
df_titanic_tmp['Survived']
df_titanic_tmp[['Survived', 'Pclass']]
df_titanic_tmp.loc[:, ['Survived']]
df_titanic_tmp.iloc[:, 1]
df_titanic_tmp.ix[:, ['Survived']]
df_titanic_tmp.ix[:, 1]

# Select some rows of certain columns
df_titanic_tmp.loc['a': 'c', 'Survived': 'Name']
df_titanic_tmp.iloc[0:2, 1:3]
df_titanic_tmp.ix['a': 'c', 'Survived': 'Name']
df_titanic_tmp.ix[0:2, 1:3]
```

De-duplication

```python
df_titanic_tmp.drop_duplicates(subset=['Survived'], keep='first', inplace=False).iloc[:, 1]
```

subset: makes distinct for the selected column, defaults to all columns;
keep: The value option {'first', 'last', False}, retains the first, last, or eliminates all of the duplicate elements;
inplace: defaults to False, returns a new dataframe; if True, returns the original dataframe after deduplication.

Grouping

```python
Importnumpy

df_titanic_tmp.groupby('Survived').size()
df_titanic_tmp.groupby('Survived').count()
df_titanic_tmp.groupby('Survived').agg({'Sex': np.sum})
df_titanic_tmp.groupby('Survived').agg({'Sex': np.max})
df_titanic_tmp.groupby('Survived').agg({'Sex': pd.Series.nunique})
```
**Combination of Data Frames**

Similar to the JOIN mode in SQL, you can implement left join, right join, inner join, full join, and so on.

(1) Horizontal merger

```python
df_titanic_tmp1 = copy.deepcopy(df_titanic_tmp)
df_titanic_tmp1.rename(columns={'Name': 'Name1'}, inplace=True)
pd.merge(df_titanic_tmp, df_titanic_tmp1, how='left', left_on='Name', right_on='Name1')
```

(2) Vertical merger

```python
pd.concat([df_titanic_tmp, df_titanic_tmp1])
```

**Data Binning (Discrete)**

```r
sbdeep <- function(data, parts, xiaoz) {
  parts <- parts # divide into 10 bins
  xiaoz <- xiaoz # minimum value
  value <- quantile(data, probs = eq(0, 1, 1/parts))
  # divided into 4 segments by data, and the step size is 1/4.
  number <- mapply(function(x) {
    for (i in 1:(parts - 1)) {
      if (x >= (value[i] - xiaoz) & x < value[i + 1]) {
        return(i)
      }
    }
    if (x + xiaoz > value[parts]) {
      return(parts)
    }
    return(-1)
  }, data)
  # labeling L1 L2 L3 L4
  return(list(degree = paste("L", number, sep = ""),
              degreevalue = number,
              Value = table(value), number = table(number)))
}

# Convert continuous variables into ordered variables

```r
train_data["Action_bin"] = binning(train_data["Action"], cut_points) 
Print(pd.value_counts(train_data["Action_bin"], sort = False))
```

**Build a Data Warehouse**

The companies excavate the customer's behavior data which is the subject data, and use other business system data as information supplement and improving. The user numbers in the marketing business application system represent users. In the process of data analysis and mining, the information received by the integrated payment management platform, and the acquisition system and GIS system is associated with the account number. Thereby, it can supplement the user's natural attributes and obtain the customer management data, establish data tables and constraints according to the theme, as well as collect and analyze user attribute, payment behavior and data warehouse. The specific process is as follows:

![Diagram](image)

**Figure 1.** The process of building a build data warehouse.

Therefore, careful and practical customer clustering is very beneficial to business strategy. The customer intelligence analysis algorithm based on swarm intelligence is a self-organizing clustering analysis of customer behavior on the basis of a simple model of ant colony cooperation and ant nest classification. Algorithm, which makes the data easier to visualize, highlights interesting features. The number of cluster centers is automatically generated from the data. The main idea of the cluster intelligence algorithm based on swarm intelligence is to randomly place the objects to be clustered in
a two-dimensional grid environment. Each object has a random initial position, and each ant can move on the grid and measure the group similarity of the current object in the local environment, so through the probability of converting the group similarity into the probability of moving the object by the probability conversion function, it picks up or drops the object with this probability. Ant colony joint actions result in objects belonging to the same category that can be clustered together in the same spatial region \(^2\).

According to the analysis and research in the previous two sections, the customer behavior analysis model based on big data is designed in detail. The specific objectives can be summarized as follows:

1. Customer's personal attributes: such as name (replaced by number), gender number, family area, income (stable income and unsteady income), voltage level, electricity consumption, etc.
2. Customer payment habits: whether the arrears are paid in time, the time of payment, and the payment method, such as business hall payment, manual collection, etc.

The behavior and attributes of payment have obvious characteristics compared to other industries, and the power company has a stable customer information storage platform, which is the integrated payment platform database, so the data can be collected in the most convenient way. The combination of customer attributes and payment attributes is applied to the cluster analysis of data mining. It tries to import a large amount of customer information from an existing database into data mining software.

Meanwhile, it uses selected data mining algorithms to classify customers, so as to achieve customer classification, and obtain some parameters that are critical to customer classification. Finally, the experimental results of data mining are analyzed one by one, and the characteristics of each type of customer group, such as the amount of payment and the payment method, are summarized and further developing targeted customer electricity usage behavior.

Using data mining technology and the different analysis of each application model according to the data results of the mining, the customer classification can be accurately and concretely. The focus of customer classification work in this cluster mining is to analyze the data, and then use the relevant customer personal attributes to describe and locate the classified customer group. The data mining process is shown below:

![Data mining process diagram](image)

By analyzing the user's analysis of the payment records, the normal payment users are divided into: all settled before the date of the liquidated damages, all settled after the liquidated damages date. The arrears users are divided into: occasional arrears, arrears Larger; often arrears but not large.

**User Credit Rating Analysis**

Based on the above analysis results, the following credit-making credit rules are formed: rating of normal credit user: users who have settled all the electricity bills after each issue will be credited according to the number of payment of liquidated damages;

Affrated user credit rating: The user level is divided according to the number of arrears in a period of time, and is divided into three levels from 1-3, with 1 being the lowest and 3 being the highest;

According to the credit rating and the number of reminders in a period of time, the credit rating is divided into three categories, namely, A, B, and C. Among them, A is the best credit, C is the worst.

Through 3 user levels and 3 credit levels, a total of 9 levels cover all users. According to the level of different customers, the next step will be to gradually withdraw the power user risk warning system.
and customer VIP mechanism to better serve the power customer. In order to be more in line with the actual business, the credit rules can be customized according to the actual situation of each power supply unit.

![Diagram of system management and analysis for user credit rating](image)

Figure 3. Structure diagram of the system of management and analysis for user credit rating.

In order to verify the validity of the theoretical model constructed in this paper, it is verified and analyzed for the real data of electricity users in a county in Mengdong. According to the analysis of the behavior of the county's company paying electricity bills, as shown in Figure 4a, in a county nearly 400,000 households in the past three years of payment behavior records, 68.83% of the electricity bill were paid on the day of informing payment; In the record, 99.79% of the records were all paid within three weeks from the power supplier's notice of the payment of electricity, and only 0.21% of the records were postponed until three weeks to complete the payment, as shown in Figure 4b.

![Data analysis diagram](image)

Figure 4. Data analysis diagram.

According to the aforementioned model, the key step of the system of management and analysis for user's credit rating is established in the environment of big data:

![Diagram of system management and analysis for user credit rating](image)

The system studied by the above method has the following advantages:
1. The model is simple, retaining the serviceability and stability of the original system;
2. Preprocessing big data can eliminate various types of data noise and eliminate data duplication, and overcome sensitivity to data;
3. It is possible to analyze data from various dimensions, calculate the user credit rating, eliminate the problem of user brushing, and improve fairness;
4. Use a variety of different algorithms to speed up data processing;
Conclusion

The system has modeled and analyzed 5 million users in Mengdong area, and divided the user level and credit rating of Mengdong area as shown in the following figure:

![User distribution in different credit rating.](image)

In 2018, the company launched a new modern service system. With the deepening of the power system reform, the opening of the incremental market, and the rapid development of the “big cloud” technology, marketing has become the forefront of the company's business, which is the “frontier position” that takes customers as the center and meet the people's needs. Under the new situation, further strengthening the basic management of marketing is a solid foundation for transforming the marketing profession into a new service system. The establishment of credit evaluation index system is a process of continuous improvement. In this process, we should learn from the advanced experience of other industries such as banks. Power supply enterprises pay more attention to customer credit evaluation, risk warning prevention and diversified marketing services. The next step requires more follow-up work to continue to build a marketing decision model, and provide a more intelligent and objective user evaluation system for marketing.

References


