Data Analysis and Research of Lithium-Ion Battery Based on Data Mining Technology

Zhi Yang and Jianhua Wu
Institute of Rail Transportation, Jinan University, No.206, Qianshan Road, Zhuhai City, Guangdong Province, China
Email: 3417213148@qq.com; tjhwu@jnu.edu.cn

Abstract. With the development of consumer electronics and energy storage systems, the market demand of lithium-ion batteries is rapidly increasing. To improve the application efficiency of lithium-ion battery, it is necessary to investigate the influence of battery parameters on the available capacity of the battery. In this paper, the data mining technology is used to study and analyze the parameter data of lithium-ion battery, aiming at exploring relationships among multi-parameters and capacity in battery charge and discharge processes, and Python language is applied to realize this end. The proposed data mining technology for lithium-ion battery includes the cleaning and discretization of lithium-ion battery data, the correlation analysis of lithium battery parameters using association rule Apriori algorithm, and the visual processing of the relationship between charge and discharge time and battery capacity.

1. Introduction
With the rapid development of modern society and the continuous progress of industry, traditional energy sources such as coal, oil, and natural gas are increasingly exhausted. People urgently need renewable energy to replace the use of traditional energy. As an idea candidate of renewable energy storage devices, lithium-ion batteries are becoming more and more popular because of their advantages of environmental friendliness, ecological protection [1], high energy density [2], so it needs more scientific research and technical applications to enable it to develop rapidly.

At present, the production of lithium batteries in the world is mainly concentrated in Japan, South Korea and China [3]. The technological innovation of lithium battery abroad includes two aspects: further increasing energy density and improving battery safety performance. In the international field, with more advanced technology and patents of lithium battery in China, the core competitiveness of technology is getting stronger, and the gap between China and Japan, South Korea in the field of lithium battery is constantly narrowing [4].

Data mining techniques are more and more frequently applied on numerical data to discover new knowledge, especially providing insight to the hidden information of data. The research on the data mining of lithium-ion battery has not been sufficiently explored yet. Based on the parameter information of voltage, current, power, capacity, energy and temperature of lithium-ion battery, this paper excavates and analyzes the data of lithium-ion battery in the processes of charge and discharge, and explore the relationship between the parameters in the lithium battery database.

The remainder of this paper is organized as follows: Section 2 introduces the proposed scheme of lithium-ion battery data mining. Section 3 describes the detailed techniques of battery data mining.
including data preprocessing and implementation of Apriori algorithm. Section 4 demonstrates the experimental results and discussions, followed by the conclusion drawn in Section 5.

2. Proposed Scheme of Lithium-Ion Battery Data Mining

The proposed lithium-ion battery data mining scheme is mainly divided into three steps:

1. Preprocess the data in the database, and use the way of data cleaning to remove dirty data in the database. Then, use the way of data transformation to discretize the data so that the data can be transformed into a format that is easy to handle.

2. Obtain the confidence and support degrees among the parameters of lithium-ion battery by using the Apriori algorithm of association rules, and arrange the results into a format which is easy to analyze. Select the association rules needed by the research for data analysis.

3. Analyze and visualize the relationship between charge and discharge time and capacity in the database, and analyze the influence of other parameters on the capacity of lithium battery. The overall design diagram of lithium-ion battery data mining is shown in figure 1.

![Figure 1. Overall design of data mining for lithium battery.](image)

3. Data Mining

3.1. Data Preprocessing of Lithium-Ion Battery

Data preprocessing refers to the necessary processing of data in the database before data mining, so as to meet the standards required by using data mining to analyze data [5]. Data preprocessing includes four parts: data cleaning, data integration, data transformation, and data specification [6]. In this paper, data cleaning and data transformation are employed to process the data of lithium-ion battery, which makes it convenient for data mining and improves the accuracy of data mining results.

3.1.1. Data Cleaning. The main ways of data cleaning can be divided into two ways: filling the missing data in the database and eliminating the noise data in the database. In the process of data cleaning, we need to pay attention to the dirty data in the database. First of all, we should analyze the causes and existing forms of the dirty data in the database. After that, we should use the existing data mining algorithms to clean up the dirty data in the database. After the data cleaning is completed, the data in the database can meet the data requirements of lithium-ion battery analysis. In the data mining of lithium-ion battery, it is found that there is a null value in the database, so the code “if len(str(sf.cell_value (i9)) > 0:” is added to the code of the data mining algorithm to detect and delete the null value in the database.

3.1.2. Data Transformation. In the data analysis, the data form in the database may not be suitable for the mining task and algorithm. Therefore, it is necessary to transform the data in the database. Data
transformation mainly includes four ways: simple function transformation, data specification, data discretization, and attribute construction [7].

In the data mining of lithium-ion battery, it is found that due to the strong continuity of various parameters of lithium-ion battery, the results of data mining cannot meet the requirements of lithium-ion battery research. The step time in lithium-ion battery data recorded by a battery charger is in the form of “hours: minutes: seconds”, which is not easy for data mining. Accordingly, it is necessary to convert the step time in minutes. Besides, most of the data of the five parameters, including battery voltage, current, power, capacity, and energy, retain three digits after the decimal point, so that the interval between the data is too narrow to easily analyze. Therefore, it requires the data discretization method to discretize the parameters of the database according to an isometric interval.

As one of the ways of data transformation, data discretization refers to segmenting continuous data into discrete intervals [8]. The principle of segmentation is based on equal distance, equal frequency or optimization. The data discretization method adopted in this paper is the distance interval method, which discretizes each column of data in the database according to an isometric interval or custom interval. It has the advantages of flexibility and maintaining the original data distribution.

3.2. Implementation of Apriori Algorithm

Python language is employed for data mining, and the programming software is python3.7.0. The python modules including numpy library, pandas library, and matplotlib library are used in this work. After completing the data preprocessing, the data format has met the requirements of data mining. Figure 2 shows the process of generating frequent itemsets by applying the association rule algorithm. It continues to connect frequent itemsets with other frequent itemsets and filter out itemsets that are greater than the minimum support until there is no itemset greater than the minimum support in a certain dimension [9]. Then it outputs all the frequent itemsets and their support information, and the support information only retains the two digits after the decimal point.

```python
def apriori(dataSet, minSupport):
    C1 = createC1(dataSet) # Build the initial candidate itemset C1
    D = [set(x) for x in dataSet]
    L, supportData = scanD(D, C1, minSupport)
    L = [L] # Each itemset in the original L contains an element
    k = 2 # The newly generated itemset should contain 2 elements, so k=2
    while len(L[k-2]) > 0:
        Ck = aprioriGen(L[k-2], k)
        Lk = scanLk(D, Ck, supportData, minSupport)
        prunedLk = pruning(Lk, L[k-2], supportData)
        L.append(Lk)
        k += 1 # The number of elements in the newly generated itemset should continue to increase
    print('Support information for all candidate sets: ', end='')
    print(list(set.union(*list(set.union(*itemsets))))
    return L, supportData # Returns a list of all frequent itemsets that meet the criteria and support information for all candidate itemsets
```

![Figure 2](image-url) Output all frequent itemsets.

Figure 3 is the code that generates the confidence rule. The confidence is defined by the code “conf = supportData [freqSet] / supportData [freqSet-conseq]”. Comparing the frequent itemsets with the set minimum confidence, it outputs the confidence rules that are greater than the minimum confidence.

4. Experimental Results and Discussions

When the minimum support is set as 0.01, the resulting support value retains only two digits after the decimal point. Figure 4 shows the support rules for code outputs, which are listed from large to small. As can be seen from the figure, the item sets with the greatest support are “voltage 3.4” and “current 30”. Since the constant current charging and constant voltage charging are employed in the charging process of lithium-ion battery, it can be considered that the charging current of lithium-ion battery in the constant current charging stage is stable at about 30A. When the charging voltage changes to 3.4V, the voltage growth slows down gradually, until it reaches a steady state and enters the constant voltage charging stage.
Figure 3. Generate confidence rules.

Figure 4. Support rules for output.

When the minimum confidence is set as 0.3, the resulting confidence value retains only two digits after the decimal point. The output confidence rules listed in figure 5 are also from large to small.

From the output results, we can see that the parameter relationship of lithium-ion battery is not related one by one, and multiple parameters will influence each other.

Figure 5. Confidence rules for output.

The confidence rules among the parameters of lithium battery are analyzed to investigate the relationship between the voltage and other parameters of lithium-ion battery in the processes of charge and discharge on the capacity of lithium battery.

4.1. Association Rules between Capacity and Step Time

A total of 693 confidence rules are obtained between capacity and step time, among which there are 364 confidence rules of “frozenset ({'capacity'})"-> “frozenset ({'step time'})” and 329 confidence rules of “frozenset ({'step time'})"-> “frozenset ({'capacity'})”. Some association rules between capacity and step time are shown in figure 6.

Figure 6. Partial association rules of capacity and step time.

From figure 6, it can be seen that the association rules with “frozenset ({'capacity'})"-> “frozenset ({'step time'})” have high confidences, among which 355 confidence rules have values of 1.0, while...
only 9 confidence rules have values less than 1.0 and all greater than 0.3. Therefore, the value of capacity corresponds to the value of step time one by one, and we can find out the relationship between capacity and step time from the association rules of “frozenset ({'capacity'}) -> frozenset ({'step time'})”. For example, the association rule of “frozenset ({'capacity: 45.2'}) -> frozenset ({'step time: 90.0'})” can be interpreted as that when the charging time is 90 minutes, the capacity of lithium-ion battery is 45.2 Ah. From the output confidence rule, we can draw the following conclusion: during the lithium-ion battery charging process, at the 30th minute, the battery capacity is 15.2 Ah which is 26.39% of the rated capacity; at the 60th minute, the battery capacity is 30.3 Ah which is 52.26% of the rated capacity; and at the 90th minute, the battery capacity is 45.2 Ah which is 78.13% of the rated capacity. According to the charge protocol of lithium-ion battery, the charging current decreases sharply in the later stage of charging, so the change value of capacity of lithium battery also shrinks sharply in the same charging time.

4.2. Association Rules Between Capacity and Voltage

There are a total of 749 confidence rules in which the output confidence between the capacity and the voltage is greater than 0.3. Among them, there are 508 confidence rules of “frozenset ({'capacity'}) -> frozenset ({'voltage'})” and 241 rules of “frozenset ({'voltage'}) -> frozenset ({'capacity'})”. Figure 7 shows some association rules between battery capacity and voltage.

From the output confidence rule, we can see that the confidence value of all association rules of: frozenset ({'capacity'}) -> frozenset ({'voltage'}) is 1.0, which can be considered that the capacity is a constantly changing value in the whole charging process of lithium-ion battery. Besides, the confidence of “frozenset ({'voltage'}) -> frozenset ({'capacity'})” is less than 1.0, which can be considered that if the capacity is within a certain value, the voltage is constant. For example, the rule “frozenset ({'voltage: 3.409'}) -> frozenset ({'capacity: 52.528'}) conf: 0.33” means that when the voltage is 3.409 V, there are three different capacity values. When the confidence value of an association rule is less than 0.01, it is considered that the lithium-ion battery has reached the stage of constant voltage charging. Figure 8 shows some association rules whose confidence values are less than 0.01 in some association rules between voltage and frozenset capacity. It indicates the battery is in the stage of constant voltage charging process, and its voltage is 3.65 V.

In the stage of constant current charging, there is a positive correlation between voltage and capacity, the capacity increases from 0 to 57.518 Ah, and the voltage increases from 2.706 V to 3.65 V. In the stage of constant voltage charging, the capacity of lithium-ion battery continues to rise until it is full, and the voltage remains constant. From the output association rules, we can see that there is a strict positive correlation between capacity and voltage, and there is no outlier, which are in agreement with the experimental results.

The capacity of lithium-ion battery increases gradually when the battery is charged at constant current in the early stage of lithium battery charging, and the current decreases rapidly in the later stage of constant voltage charging, but the battery capacity is still increasing. Therefore, there is a
negative correlation between battery charging current and capacity in the later stage of charging. However, in the stage of constant current charging, both battery voltage and capacity are increasing, and there is a positive correlation between capacity and voltage.

4.3. Visual Processing of Charge and Discharge Time and Capacity
The visual processing step is to visualize the relationship between the charge or discharge step time and battery capacity, which is used to intuitively analyze the percentage of battery charge or discharge capacity to the total capacity. The relationships between step time and capacity in the cases of charge and discharge are shown in figures 9-10, respectively. From the figures, it can be seen that there is a strict linear relationship between step time and charge or discharge capacity, but the growth rate of capacity decreases significantly at the end of the charging curve.

![Figure 9. Diagram in the case of charging.](image1.png)

![Figure 10. Diagram in the case of discharge.](image2.png)

During the charging process of the lithium battery, the capacity of the lithium battery is 26.7% of the rated capacity after 30 minutes of charging, 52.8% of the rated capacity after 60 minutes of charging, and 77.6% of the rated capacity after 90 minutes of charging. Compared with the association rule analysis of the relationship between charging time and capacity of lithium battery, we can see that the maximum error of both sides is 1.16% when the charging time is 60 minutes, so it can be considered that the error between the two sides is small, our conclusion is reliable. We can get the formula of charging time and capacity of lithium battery through visual analysis chart: capacity (AH) = 5/3 * step time (min). From figure 10, we can see that the change of the capacity of lithium battery in the discharge process is similar to that in the charging process. We can get the formula of discharge time and capacity of lithium battery through visual analysis diagram: capacity (AH) = 60-5/3 * step time (min).

The capacity of the last part of the charging process rises more slowly than before because the lithium battery changes from constant current charging to constant voltage charging after 1 hour and 55 minutes. In the case of constant voltage charging, the current decreases sharply, and the current decreases from 30A to 3A within two minutes. The decrease of current leads to slower capacity growth. In the constant current stage of lithium battery charging, the voltage changes very little, and the temperature changes between 25 and 28 degrees Celsius, while the visualization chart shows that there is a strict linear relationship between the step time and the charging capacity. therefore, we can think that the temperature in this range has no great effect on the capacity.

Because the capacity of lithium battery has a strict linear relationship with the charge and discharge time, we can see that the lithium battery analyzed in this paper is a relatively new battery. If we can collect some data of depleted battery for comparative analysis, we can further analyze and study the loss and aging degree of lithium battery.

5. Conclusion
Aiming at developing data mining techniques for lithium-ion battery, this paper establishes an effective mining process to analyze the potential correlations among the parameters of lithium-ion
battery, such as the step time of charging and discharging, voltage, current, and capacity by using the association rule of Apriori algorithm. The relationships between the step time and battery capacity are visualized for easy analysis of battery charging and discharging processes. The research on lithium-ion battery by data mining can express the data information more clearly and effectively, which reveals the relationships among the data and could help users to understand the charge and discharge performance of lithium-ion battery more clearly, thus improving the application efficiency of lithium-ion battery in practice. However, due to the limitation of battery test data, the relationships among more parameters of lithium-ion battery, such as battery power capability and ambient temperature, have not been sufficiently explored. These will be investigated in our future work.

References
[1] Liu Y J, Sun G Y, Cai X H, Yang F, Ma C, Xue M and Tao X Y 2020 Nanostructured strategies towards boosting organic lithium-ion batteries Journal of Energy Chemistry 54.
[2] Chemali E, Kollmeyer P J, Preindl M and Emadi A 2018 State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach Journal of Power Sources 400.
[3] Xue H B, Wu W L and Cheng X Y 2018 Research on the foreign science and technology cooperation mode of China’s lithium battery industry Energy Research and Utilization 50-53. (in Chinese)
[4] Chen M 2019 Research on the Development Strategy of ND Company: An Enterprise of Lithium Battery Powered by New Energy (Zhejiang University of Technology) pp 7-9 (in Chinese).
[5] Konstantinou N and N W Paton 2020 Feedback driven improvement of data preparation pipelines Information Systems 92.
[6] Zhang Y and Li X L 2020 Data preprocessing of floating car based on DBSCAN algorithm Jiangxi Science 293-297+319 (in Chinese).
[7] Zhao R 2019 Research on the Application of Inconsistency in Data Discretization and Classification (Dalian Maritime University) pp 2-5 (in Chinese).
[8] Dhalmahapatra K, Shingade R and Maiti J 2020 An innovative integrated modelling of safety data using multiple correspondence analysis and fuzzy discretization techniques Safety Science 130.
[9] Zhou Y W 2020 Design and implementation of book recommendation management system based on improved Apriori algorithm Intelligent Information Management 75-87.