

Application of Transaction Mining Based on FP-Table Algorithm in Mobile Electricity Market

Chuncheng Gao¹, Yong Dai¹, Minghai Jiao²

¹Nari Group Corporation/State Grid Electric Power Research Institute, Nanjing, China

²Computing Center, Northeastern University, Shenyang, China

Email: gaochuncheng@sgepri.sgcc.com.cn, daiyong1@sgepri.sgcc.com.cn, mhjiao@cc.neu.edu.cn

Received 27 May 2015; accepted 11 July 2015; published 14 July 2015

Abstract

Electricity market trade based on mobile intelligent device will extend the volume of transaction. For the massive and various trading data, transaction mining algorithm is very useful to find the relationship of correlative elements such as trade price and power capacity, and it always occurs between the power users and power generation enterprises. The novel FP-Table algorithm is proposed in this paper to solve the massive transaction mining problem. The FP-Table algorithm integrates the Hash table into FP-Growth algorithm, using two-dimension table saving frequency count of item pair, then mining the frequency items of electricity transactions efficiently. Application of mobile transaction mining is proved to be high efficiency and high value by performance experiment results.

Keywords

FP-Table Algorithm, Electricity Trade, Transaction Mining, Mobile Electricity Market

1. Introduction

Electric power trade is the core business of electricity market transaction [1]. It is organized by Power Market Trade Center for electric market members such as power users, power generation enterprises and so on, trading on web business platform according to real power supply and demand. Nowadays, the electric power trade has extended the various types in China, including power generation trading, direct trading between power users and power generation enterprises, external power supply exchange, external and internal province exchange.

With the application of web 2.0 theory and technology, the mobile device is useful in social network based on internet website; information dissemination is so fast for every online user that mobile technology promotes user involved in the network activity, typical application software such as Facebook, Twitter, Tencent QQ, Dianping, and so on. Not only is the public message pushed to the users, but also the emotion information as product recommendation and dealing preference selection is gotten by the online users. So the application of mobile electricity market based on mobile internet device is high value for power trading users.

As to the massive transaction data mining problem, there has some data mining theory and method research on the business transaction. Agrawal *et al.* [2] present an efficient algorithm that generates all significant association rules among items in the database, in which each transaction consists of items purchased by a customer in a visit. Smyth *et al.* [3] propose data-driven evolution algorithm for data mining motivated by real world data sets, and solve the classical problems in data analysis involving multivariate data including classification, regression, clustering, and density estimation. From the theory, Meo [4] presents a new model to evaluate dependencies in data mining problems. Meanwhile, the well-known concept of the association rule is replaced by the new definition of dependence value, which is a single real number uniquely associated with a given item set. As the point of rules optimization, Lakshmanan [5] proposes an architecture for supporting constraint-based, human-centered and exploratory mining of various kinds of rules including associations, introduces the notion of constrained frequent set queries (CFQs), and develops effective pruning optimizations for CFQs with 1-variable (1-var) constraints. In the next year, Lu *et al.* [6] extend the scope of mining association rules from traditional single-dimensional intra-transaction associations, to multidimensional inter-transaction associations. On the other algorithm hand, high quality algorithm is also applied for efficient mining frequency items in trading datasets. Mabroukeh and Ezeife [7] present taxonomy of sequential pattern-mining techniques in the literature with web usage mining as an application, introducing a enhancing understanding taxonomy for classifying sequential pattern-mining algorithms based on important key features. For incremental mining association rules, an effective sliding-window filtering algorithm is researched by Lee *et al.* [8]. In recent years, massive data mining problem is researched hot by researchers, Patil *et al.* [9] give contribution to improved apriori algorithm by hiding sensitive association rules which are generated by applying improved apriori algorithm on supermarket database. Singh and Agarwal [10] propose a new optimized algorithm and to compare its performance with the existing data mining algorithms. Prasanna and Seetha [11] present a method for generating association rules from large high dimensional data, which can obtain more rapid computing speed and sententious rules. And then, probability-based algorithm, one of an incremental algorithm, is researched by Ariya and Kreesuradej [12], which applies the principle of Bernoulli trial to predict expected frequent item sets for reducing collected border item sets and a number of times to rescan the original database.

The rest of this paper is organized as follows. Section 2 describes the details of the transaction mining based on FP-Table algorithm, which includes definition of two-dimension transaction pattern table and FP-Table. Transaction mining process based on FP-Table algorithm is presented in Section 3. Then the application of transaction mining in mobile electricity market is executed in Section 4.

2. Transaction Mining Based on FP-Table Algorithm

2.1. Definition of Two-Dimension Transaction Pattern Table

Definition 1 The two-dimension transaction table is a data set items associate frequency pattern table, which is made of item rows and item columns, saving the combination count information for a transaction database. A transaction is a collection of multiple items in database. And the arbitrary combination of two items in the transaction is called basic associate frequency pattern unit.

From a transaction of database, a sequence of items is considered as a queue. So the each previous item and the following each item in the item queue constitutes the rows and columns of two-dimensional table, the value between relevant row and column in the table is the frequency count. Since a trade database is composed of many transactions, it also includes all the items of transaction expressing the rows and columns of two-dimensional table, showing all the transaction combination counts between every two items in trade database.

For the simple example, a transaction T_1 is a sequence of items collection {abcef}, the two-dimension table of frequency item pattern count is created in **Figure 1**. The row is made of item a, b, c, e, f, and the column is also made of item a, b, c, e, f. The frequency count value of each two item combination of {a, b}, {a, c}, {a, e}, {a, f}, {b, c}, {b, e}, {b, f}, {c, e}, {c, f}, {e, f} is recorded as 1 in the table, others is recorded as zero in the table. The number (Num) 0, 1, 2, 3, 4 is marked as sequent item a, b, c, e, f.

2.2. Definition of Data Mining FP-Table

Definition 2 Hash-T is a hash table, which combines the candidate item frequency pattern with hash table from transaction database.

		a	b	c	e	f
	SN	0	1	2	3	4
a	0	0	1	1	1	1
b	1	0	0	1	1	1
c	2	0	0	0	1	1
e	3	0	0	0	0	1
f	4	0	0	0	0	0

Figure 1. Example of a transaction pattern table.

The candidate items are the whole transaction items in the trade database. Firstly, the database is thoroughly scanned, and the candidate items are counted by frequency, as is called Support degree (Sup). Then all the items are sorted by the sequence. At the same time, the hash table is used for exactly defining candidate items sequence by Sequence Number (SN) and Item Name (Name). The example of Hash-T, from a transaction database including three transaction items, is as **Figure 2**.

From **Figure 2**, transaction database is composed as three transactions, including {abcef}, {acg} and {ei}. All the items are scanned and the frequency items are counted, then the Hash-T is constituted by the title Num, SN, Name and Sup, noting as number, address, item name and support degree.

Definition 3 FP-Table is a combination table based on improved FP-Growth algorithm, integrating sorted Hash-T and two-dimension table.

Given **Figure 2** example, firstly, a two dimension table is created by frequency count of all the item pairs in the transaction database. Then Hash-T is sorted by support degree, so the sequence of candidate items is a, c, e, b, f, g and i. And the items of a, c, e are kept as mining items according to the support degree value of 2 (the support degree value is defined as great and equal 2). At last, the two dimension table is converted as FP-Table, which is shown as **Figure 3**.

Lemma 1 The item pair in the FP-Table is the minimal combination unit. And the frequency combination item depends on the frequency count of item.

Proof Firstly, given a transaction including two items $\{A_1A_2\}$, the frequency item and the frequency count are scanned, noted as $\langle A_1 A_2: 1 \rangle$;

Then, given two transaction including collection as $\{A_1A_2\}$, $\{A_1A_2A_3\}$, the frequency items and the frequency count are also scanned, note as $\langle A_1 A_2: 2 \rangle$, $\langle A_1 A_3: 1 \rangle$, $\langle A_2 A_3: 1 \rangle$, $\langle A_1 A_2 A_3: 1 \rangle$. Above frequency count value of combination item $\langle A_1 A_2 A_3: 1 \rangle$ is 1, just due to the minimal frequency count value of item pair $\langle A_1 A_3: 1 \rangle$ and $\langle A_2 A_3: 1 \rangle$ is 1;

Next, given n transaction including collection as $\{A_1A_2\}$, $\{A_1A_2A_3\}$, ..., $\{A_1A_2A_3 \dots A_n\}$, the frequency items are scanned and noted as $\langle A_1 A_2: n \rangle$, $\langle A_1 A_3: n-1 \rangle$, ..., $\langle A_1 A_n: 1 \rangle$, $\langle A_2 A_3: n-1 \rangle$, ..., $\langle A_2 A_n: 1 \rangle$, ..., $\langle A_1 A_2 A_3: n-1 \rangle$, ..., $\langle A_1 A_2 A_3 A_4: n-2 \rangle$, ..., $\langle A_1 A_2 A_3 \dots A_{n-1} A_n: 1 \rangle$. The frequency count value of combination item of $\langle A_1 A_2 A_3 \dots A_{n-1} A_n: 1 \rangle$ is 1, due to the minimal frequency count value of item pair $\langle A_1 A_2: n \rangle$, $\langle A_1 A_3: n-1 \rangle$, $\langle A_1 A_4: n-2 \rangle$, ..., $\langle A_1 A_n: 1 \rangle$ is 1.

3. Transaction Mining Based on FP-Table Algorithm

3.1. FP-Table Creating Algorithm

FP-Table algorithm is implemented as two steps. The first step is the creation process of FP-Table, and the next step is the mining process of FP-Table.

Building FP-Table Algorithm: given a trade database, once the transaction database is scanned, the candidate frequency items set is set into hash set table Hash-T, including SN, Name and Sup. At the same time, the frequency counts between all item pairs are recorded in two-dimension table. Then, the sort process is implemented in Hash-T according to the descending order of support degree. The items, which support degree is less than threshold value, are removed from Hash-T, and the item number is reordered again, Hash-T is updated as result. So the FP-Table is created by Hash-T index of items, and the transaction frequency count value of index item in Hash-T is set into FP-Table, obtaining from two-dimension table. The algorithm is shown as **Figure 4**.

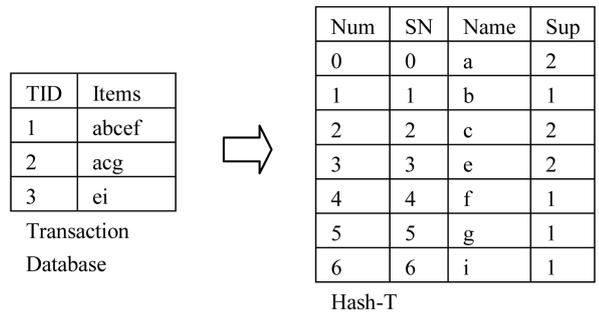


Figure 2. Example of Hash-T table.

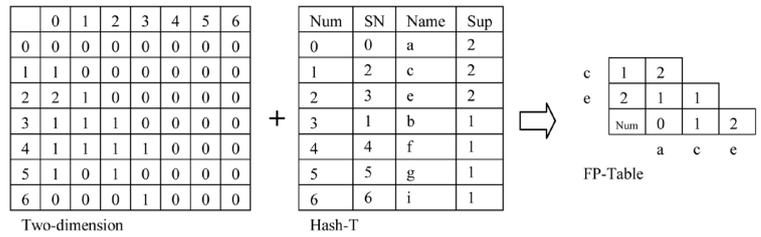


Figure 3. Example of FP-Table.

```

Input: trade database;
Output: FP-Table;
for (i =1, i++, i < sizes of database) {
Scanning the transaction Ti of trade database;
if ( item in Hash-T is null)
    {item is added in Hash-T table; count=1;}
else {
    count++;}
Creating two-dimension table, count value is recorded between row and column.}
Sort is implemented according to descending support degree;
if ( Sup < threshold ) {
item is removed from Hash-T;}
for (j=0; i++; j< sizes of Hash-T item array){
item frequency count = count in two-dimension table;}
    
```

Figure 4. Building FP-Table algorithm.

3.2. Mining FP-Table Algorithm

Mining FP-Table Algorithm: the combination items and frequency counts are given according to the Hash-T item index from FP-Table. Firstly, the item pairs and support degree are created, and then more items are combined with a set of transaction according to the lemma 1. After all the frequency items are combined recursively, the mining process of FP-Table is completed. The algorithm is shown as **Figure 5**.

4. Application of Transaction Mining in Mobile Electricity Market

4.1. Transaction Database in Electricity Market

Given transaction mining in electricity market, the transaction database is proposed for electricity transaction as a case study. The items of electricity transaction are derived from mobile electricity trade system platform. Items are classified as power generation trade (a), direct trading between power users and power generation enterprises (b), electricity price (c), inter provincial Trading (d), bilateral negotiation (e), centralized matchmaking (f), medium and long-term transactions (g), month transactions (i), year transactions (j), and delivery of electricity trading (l). The Hash-T table and two-dimension table are built as **Figure 6**.

4.2. Frequency Item Pattern Mining in Electricity Market

For the Hash-T table, all the items are sorted by the support degree (Sup). Then the items which support that degree is less than threshold are removed from Hash-T table. Next, the FP-Table is constituted as the following step: first, the sorted items are recorded in the dynamic array; second, the frequency count of item in Hash-T is selected and filled into FP-table. Since the FP-Table is composed by row and column, the frequency count in FP-Table is expressed as associated relationship among item pairs. But the sparse data value of frequency count in FP-Table is the majority; the dynamic array technology is used to deal with the effective count value. The FP-Table is built as **Figure 7**.

From the FP-Table, the frequency item in electricity market is recursively proceeded to pattern mining according to lemma 1. For example, as to item g in sorted Hash-T and FP-Table, the frequency count value between item pairs of g and a is 5, the frequency count value between item pairs of g and c is 5, and the frequency count value between item pairs of g and e is 4. So the frequency item pattern mining process is described as the following: once the frequency count value of g and e is 4, all the frequency count of item combination with e is only 4, that is, the frequency pattern of g has {a g: 5}, {c g: 5}, {e g: 4}, {a c g: 5}, {a e g: 4}, {c e g: 4}, {a c e g: 4} according to the pattern count condition of <a: 5, c: 5, e: 4>.

```

Input: FP-Table and minimal support degree;
Output: all frequency items collection;
if (i==2){
  all item pair from FP-Table;}
if (i>2){
  all combination items according to lemma 1 from FP-Table;}
    
```

Figure 5. Mining FP-Table algorithm.

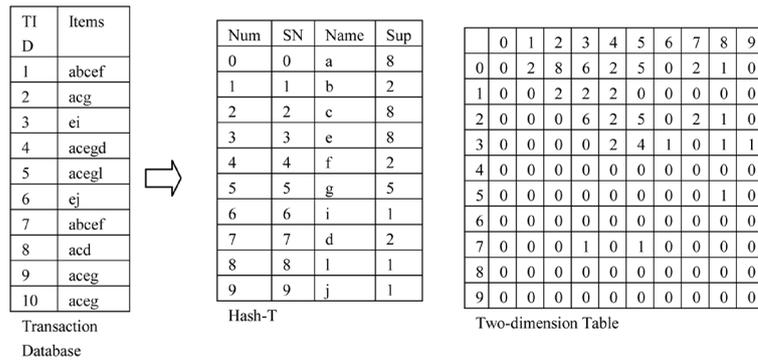


Figure 6. Hash-T table and two-dimension table of transaction database in electricity.

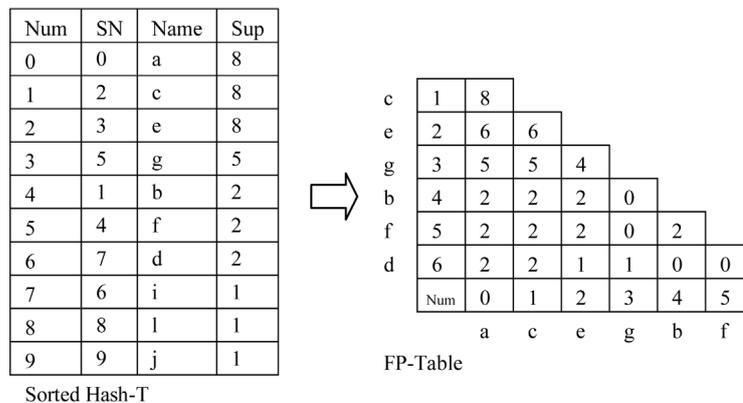


Figure 7. Sorted items in Hash-T table and built FP-Table.

All the items of frequency pattern in electricity trade database are proceeded by the above method. And the experiment proves that the algorithm is efficient and robust.

Acknowledgements

Gao CC thanks the science and technology research project sponsored by China State Grid Corp under grant number DZN17201400039. The algorithm research and technology application is supported by the project. Then the transaction mining application effect will feed back the experiment in the paper.

References

- [1] Deb, R.K., Hsue, L.L., Albert, R. and Christian J.E. (2001) Multi-Market Modeling of Regional Transmission Organization Functions. *The Electricity Journal*, **14**, 39-54. [http://dx.doi.org/10.1016/S1040-6190\(01\)00174-9](http://dx.doi.org/10.1016/S1040-6190(01)00174-9)
- [2] Agrawal, R., Imieliński, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, New York, June 1993, 207-216. <http://dx.doi.org/10.1145/170035.170072>
- [3] Smyth, P., Pregibon, D. and Faloutsos, P. (2002) Data-Driven Evolution of Data Mining Algorithms. *Communications of the ACM*, **45**, 33-37. <http://dx.doi.org/10.1145/545151.545175>
- [4] Smyth, R. (2000) Theory of Dependence Values. *ACM Transactions on Database Systems*, **25**, 380-406. <http://dx.doi.org/10.1145/363951.363956>
- [5] Lakshmanan, L.V.S., Ng, R., Han, P. and Pang P. (1999) Optimization of Constrained Frequent Set Queries with 2-Variable Constraints. *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, Philadelphia, May 1999, 157-168. <http://dx.doi.org/10.1145/304182.304196>
- [6] Lu, H., Feng, L. and Han, J. (2000) Beyond Intratransaction Association Analysis: Mining Multidimensional Intertransaction Association Rules. *ACM Transactions on Information Systems*, **18**, 423-454. <http://dx.doi.org/10.1145/358108.358114>
- [7] Mabroukeh, N.R. and Ezeife, C.I. (2010) A Taxonomy of Sequential Pattern Mining Algorithms. *ACM Computing Surveys*, **43**, 1-41. <http://dx.doi.org/10.1145/1824795.1824798>
- [8] Lee, C.H., Lin, C.R. and Chen, M.S. (2001) Sliding-Window Filtering: An Efficient Algorithm for Incremental Mining. *Proceedings of the Tenth international conference on Information and Knowledge Management*, Atlanta, October 2001, 263-270. <http://dx.doi.org/10.1145/502585.502630>
- [9] Patil, S.P. and Patewar, T.M. (2012) A Novel Approach for Efficient Mining and Hiding of Sensitive Association Rule. *Proceedings of the 2012 Nirma University International Conference on Engineering*, Ahmedabad, 6-8 December 2012, 1-6. <http://dx.doi.org/10.1109/nuicone.2012.6493184>
- [10] Singh, A. and Agarwal, J. (2014) Proposed Algorithm for Frequent Item Set Generation. *Proceedings of the 2014 Seventh International Conference on Contemporary Computing*, Noida, 7-9 August 2014, 160-165. <http://dx.doi.org/10.1109/ic3.2014.6897166>
- [11] Prasanna, K. and Seetha, M. (2012) Mining High Dimensional Association Rules by Generating Large Frequent K-Dimension Set. *Proceedings of the 2012 International Conference on Data Science & Engineering*, Cochin, 18-20 July, 2012, 58-63. <http://dx.doi.org/10.1109/ICDSE.2012.6282304>
- [12] Ariya, A. and Kreesuradej, W. (2013) Probability-Based Incremental Association Rule Discovery Using the Normal Approximation. *Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration*, San Francisco, 14-16 August 2013, 432-439.