Service Process Improvement Based on Exceptional Pattern Analysis

Bing Li Shuo Pan

School of Information Technology & Management Engineering
University of International Business and Economic
Chaoyang District, Beijing, China (100029)
E-mail: lb0501@126.com, shuopan@126.com

Abstract. The continual improvement of service process is one of the most important factors for improving service process management level. But it is very difficult to identify the real causes of service fault executing service process improvement. Aiming at the problem, this paper proposed a service fault recognition & improvement method based on association rules mining. The method provides a new approach for enterprises to improve effectively service process fault.

1. Service Fault Analysis Based on Association Rules

In the present complicated and fierce competitive circumstance, it is increasingly important to improve the management of service process so as to improve enterprise's service effectively and quickly satisfy the customers' needs. Therefore, many scholars, domestic or foreign, have done lots of research in this area and attained abundant achievements. For instance, many companies and scholars have studied the improvement of service process in terms of 6σ management [1,2,3] and then use method DMAIC to achieve the optimization and improvement of service process. However, these methods need to be measured, analyzed and improved by the professional statistical means and specialists, which is comparatively difficult because of professional obstacles. Moreover, sometimes the real reasons for faults generated in service are too complicated to be improved, or even, be identified by these methods related above. In order to realize the dynamic optimization of service process, Yushun Fan and Fabio Casati[4,5] have proposed the framework of a real-time business process performance management(RTBPPM) and a toolbox of business process intelligence(BPI). Both of them have the common features as the real-time monitor of service process, which is efficient and convenient as well. J.Leon Zhao and Edward Stohr[6] have proposed a method of service time assessment based on workflow management and have put forward control strategy of service process. Besides, Cheung.C.F[7,8] has proposed multi-perspective knowledge-based system(MPKBS), which integrates case inference technique and time sequence model, predicts service quality, navigates automatically, and assesses service performance by analyzing customers' rights or conditions of customers and staffs. Although these service-oriented implementations and optimizations are effective, we still lack a good method to improve the service. It is undoubtedly beneficial for enterprise's management to find a better method to improve services based on the research mentioned above. Considering this demand, this paper proposes the method of service fault recognition and improvement which utilizes the historical records of service process, analyzes exceptional past events, discovers and fixes the imperfect links, and elevates the management level of service process.

Service faults are always the effects of several causes. The way to analyze service faults is to scan mass historical logs about service faults. Owing to masses of historical information generated in enterprise service process, one discovers that it is difficult to find out the causes of faults both by traditional cause-effect method and 6σ management. Therefore, we will use association analysis belonging to data mining technique to analyze the association between the service faults and their causes.

1.1 Basic Concept

Association analysis, proposed by Agrawal from IBM Almaden Research Center, America, in 1993, is a simple but practical method in data mining. From mass business records, we use it to get association rules which show us the interesting associations among items.

Here are some explanations about the related concepts. Let $I = \{i_1, i_2, \cdots, i_m\}$ be the set of m different transaction databases. We assume that task-related data D is the set of transactions in database and every transaction T is the set of items, which satisfies $T \in I$. The corresponding identifier of every transaction is TID. An association rule is a consequence of implication such as $A \Rightarrow B$, therein A, B are attributes(items) of the transaction record respectively, as well as $A \cap B = \Phi$.

Usually, Support and Confidence are used to measure an association rule. For an association whose form likes $A \Rightarrow B$, (and "=>" means "lead to") Support is the percentage of the transactions in which both item A and item B occur in all transactions, recorded as $\sup(A \Rightarrow B) = p(A \cup B)$.Confidence is the conditional probability which B occurs in these transactions in which A occurs, and recorded as $conf(A \Rightarrow B) = p(B \mid A)$. Generally, only these association rules above both minimum threshold of Support ($\min_\sup$) and minimum threshold of Confidence ($\min_conf$), can be regarded as meaningful.

The algorithm "Apriori" by Rakesh Agrawal is the core algorithm of association rules mining; though many improved algorithms have followed, they are all based on "Apriori". It contains two parts: first, find out every combination of data items which satisfies the $\min_\sup$, called frequent itemset, and second, generate the rules which satisfy $\min_conf$ from every frequent itemset.

However, in the process of generating rules, many lengthy, unintelligible, and redundant rules and information will be mined out and reduce complexity of algorithm. In order to solve these problems mentioned above, R.Srikant, Robertor J and Bayardo J[9-11] proposed association mining methods based on constraints. Aimed at different problems, many scholars, at present, are researching on association mining based on constraints from different aspects. Based on this train of thought, this paper will try to bring constraint-based association mining to identify the causes for service faults.

Association Analysis of Service Fault Based on Constraint

When enterprises analyze service faults, the “possible cause---fault” records in historical logs can be utilized to find out association rules underlying general faults and their causes. When these rules’ Support and Confidence reach or surpass the minimum threshold, we consider them as strong associations which, correspondingly, reflect the obvious weak links in service process. Therefore, by analyzing the rules afterwards, finding out the weak links, and improving the main process, we can effectively elevate service process management.

In order to reduce the complexity of algorithm and decrease the generation of needless or incomprehensible rules, we brought in some antecedent constraints (service fault reasons) according to features of a certain service, and then a consequent-fixed, antecedent-constrained association rules mining was formed. Besides, we added some evaluations about rules’ Novelty to keep the incremental mining process run better.

First, in order to sharply decrease the number of generated rules, specialist could preliminarily sift the reason set $A = \{reason_i \mid i \in N\}$ which causes a certain kind of faults, to ascertain the reasons which have little influence with the fault and create the following constraints for exclusive combination from the reasons and fault.

We introduced the following constraint:

$$reason_i \Rightarrow !result_j$$

And $reason_i$ ------ The reason i in rules antecedents

$result_j$ ------ The kind of fault j in rules consequents

This is a concise constraint, which can effectively decrease the number of frequent 1-itemset in the mining process. When we applied in the first scanning of historical service records, it indirectly decreased the number of frequent n-itemsets and increased the generation of rules, and of course the thresholds are decided by special application and real mining work.

Secondly, although there are many reasons leading to faults in service process, only a few frequently occurring factors contribute to the discovery of the faults’ origin. So, we can proceed to add the following constraint in order to avoid the redundancy of antecedent rules.

We used the following constraint:

$$count(antecedent) \le \max_antecedent$$

And $count(antecedent)$ ------ the number of rule antecedents

$\max_antecedent$ ------the maximum number of antecedents

This is an anti-monotone constraint, which prevents the generation of long-scheme rules, reduces the times of database scanning and effectively shortens the runtime.

Thirdly, according to the consequent-fixed rules, some complicate rules (the one has longer antecedent) barely offer any useful information, for example:

$$B_1, B_2 \Rightarrow C \qquad (confidence = 80\%) \qquad \text{(Rule 1)}$$

$$B_1 \Rightarrow C \qquad (confidence = 85\%) \qquad \text{(Rule 2)}$$

Rule 1 means that there is 80% probability of generating fault C, when B_1 and B_2 occur at the same time. We generally regard Rule 1 as a meaningful one, because it reflects the universality of reasons generating fault C. But comparing with Rule 2, the meaning of Rule 1 becomes dull.

Robertor and Bayardor J from IBM Almaden Research Center proposed the concept of Improvement [11] in the research of constraint-based rule mining in large, dense database. The Improvement of a rule is, in the condition of rule consequent fixed, the minimum difference between the Confidence of one rule and its sub-rule's Confidence, for instance, according to $C \Rightarrow A$, its improvement is: $imp(C \Rightarrow A) = \min\left(\forall C' \subset C, conf(C \Rightarrow A) - conf(C' \Rightarrow A)\right)$, and due to the characteristics of dense database, the Improvement defined in that paper should be above 0. Here, we introduced Improvement to eliminate some rules which are not so meaningful.

Finally, since association rules include prior knowledge, we need to compare the new-generated rules with the old rules to estimate its Novelty. Because the whole mining process is an incremental process, we should find out the variances between the new and the old, and we consider the new ones with large variances as the novel, the new-mined knowledge which deserves our attention and preservation. Aimed at the analysis of service fault, on account of fixed rule consequents, we mainly pay attention to the difference of antecedents between the new rules and the old.

Assume S_1 represents the set of new-generated rules and S_2 the set of old rules. The number of rules in S_1, S_2 are $count(S_1)$ and $count(S_2)$ comparatively. w_{ij} is used to measure the difference between Rule r_i in S_2 and Rule r_j' in S_1, which is actually the Novelty. To estimate Novelty of rules in S_1 is, in fact, to find the Cartesian product $S_1 \times S_2$.

w_{ij} is calculated by the formula as follows:

$$w_{ij} = \begin{cases} 0 & r_i \in S_2, \text{ or } r_i \text{ and } r_j' \text{ have different consequents.} \\ \sum_{h=1}^{m} f_h \cdot diff_h & \text{other conditions}, h = 1,2,\cdots,m \text{ is the number of attributes in a rule antecedent.} \end{cases} \quad \textbf{(1)}$$

Therein, f_h represents the weight value of h, $diff_h$ stands for the different degree between attributes h in r_i and its counterpart in r_j' according to the same language variable. For one attribute, we could disperse its value into several intervals marked with (an) ordinal numbers, in which the minimum is 1 and the maximal is the number of intervals. We define the difference of two continuous intervals as 1.

The novelty of Rule r_i, comparing with S_2, is the sum of all w_{ij}.

$$W_i = \sum_{j=1}^{count(S_2)} w_{ij} \qquad \textbf{(2)}$$

In this way, we can find all the novelty of rules in S_1. In practical, for rules which have the same consequent, we can set a minimum threshold θ according to the practical situation. When a rule satisfies $W_i > \theta$, we regard it as a novel one, and preserve it for following analysis, or else, we delete this rule. In this way, we can decrease the number of rules generation once more.

Based on the theory above, we designed the improved algorithm of service faults analysis based on association mining ("IApriori") as follows:

Algorithm：IApriori

input： min_sup (minimum Support), min_conf (minimum Confidence), max_antecedent(maximum number of antecedent), exclusive constraint, min_imp(minimum Improvement), all the novelty thresholds of rules , database D

output： All rules which satisfy the constraint including min_sup, min_conf, max_antecedent_count , min_imp, θ

```
begin:
Put all transactions in D into temporary table D';
    if (consequent is frequent) then
        add consequent to L1;
    else
        return;
 L1= L1+find_frequent_1-itemsets(D'); //generates frequent 1-itemset
 delete T where not contain L1 ;
 while (LK-1≠φ and k<= Antecedent+1)
{ CK=apripri_gen( LK-1,min_sup);//generates K-itemset
   LK=subset(CK,D');//generates frequent K-itemset
gen_rules(k, LK, LK-1) //generates rules with fixed consequent}
for each ri in new rules table
    θi =novelty(ri)   // calculate novelty
```

```
  if θi > θi'      //estimate whether its novelty satisfies the novelty
threshold of old rules which has same consequent with it

  output the rule ri ; return;

procedure gen_rules(k, LK, LK-1)
for each c in LK
  if consequent∈c
   {conf=sup(c)/sup(c-consequent);
        conf>=min_conf  then
           {  max_subconf=0;
               for all (k-1)-subset s in (c-consequent)
                 if （conf(s)> max_subconf ）then max_subconf=conf(s);
                     if (conf-max_subconf)≥ min_imp then
                         output the rule c − consequent ⇒ consequent with
                         confidence=conf and support=sup(c);
            }
     }
return;
```

2. Improvement of Service Fault Based on Constraint

According to a series of analysis above, we used the following steps to realize service process improvement based on association analysis.

Step 1:

After a certain period of system running, decision center analyzes the log of exceptional events which are stored in data warehouse and then create tables of exceptional events and reasons (Table 1). The reasons of exceptional events include internal reason (such as employee's illness), external reason (such as delayed stock of component) and improper service model and etc.

Step 2:

In order to find out the degree of frequency that exceptional events and its corresponding reasons occur at the same time, we use "IApriori" to mine out rules formed as follows:

$$causes(X,"R1") \wedge \cdots \wedge causes(X,"Ri") \Rightarrow occurs(X,"Ej")$$

$$[\text{support} = \text{x\%}, \text{confidence} = \text{y\%}]$$

Step 3:

According to the generated rules above, companies can improve bottlenecks and improper process path to optimize the service process.

Table 1. Transaction Data of Exceptional Event Reasons

Exceptional event ID	Reason ID list
E100	R1，R2，R3，……
E200	R2，R5，…….
…….	……

3 Algorithm Analysis

According to the algorithm we proposed, we have done an experiment to validate it. With IBM data generator, we generated simulate historical logs for test, showed in Table 2. Experimental circumstance as follows: Pentium 1.4 PC, Memory 512M, Windows XP.

Table 2. Testing Parameter

Record number	Transaction number	Item number	Minimum Support	Minimum Confidence
10080条	860个	125个	0.1	0.1

As test data, we analyzed it with “IApriori” and brought several simulate constraints into our experiment. As a result, concise constraint had sharply decreased the rules number and runtime, and the influences of improvement and Novelty are also obvious. The average runtime of the association rule algorithm gets reduced only from 189 sec to 107 sec, which is less than half, and we can see it had not been reduced comparatively like generated rule number. So maybe it is important for us to advance the relativity between runtime and rule number in the future. We compared our algorithm with the traditional “Apriori”. The comparison results are shown in Table3:

Table 3. Comparison of Algorithm

	Test times	Average rules number	Average runtime
Traditional Apriori	15	152	189.454 s
IApriori	15	19	107.117s

4. Conclusion

In this paper, to realize the service fault recognition & improvement, we proposed a service fault recognition & improvement method based on association rules mining, and it could become a new approach to correct the faults generated in enterprises service process effectively. In the future, we will use practical data to validate our method further, and then try to improve it and make it work better in practical business.

Reference

1. Liu yi-qun. The fifth special topic of BMG’s Six Sigma: Six Sigma process design. China quality. 2003 No.12：26
2. Li xiao-fen, Sima tisheng. Sigma management & supply chain management. Market Modernization. 2004 Vol.12：29-30
3. Chen jing, Du lanying. Six Sigma Management in Logistics Enterprise. LOGISTICS TECHNOLOGY. 2004 Vol.4：15-17
4. Yushun Fan, Xinxin Bai. Real-Time Business Process Performance Management. E-Commerce Technology for Dynamic E-Business, 2004. IEEE International Conference on, 13-15,Sept.2004:341 – 344
5. Fabio Casati, Umesh Dayal, Mehmet Sayal, and Ming-Chien Shan. Business Process Intelligence. http://www.hpl.hp.com/techreports/, 2002.
6. J.Leon Zhao, Edward Stohr. Temporal Workflow Management in a Claim Handling System, WACC’99 2/99 Sancisco, CA, USA: 187 – 195
7. Cheung, C.F.,Lee, W.B., Wang, W.M., Chu, K.F.,To, S.. A multi-perspective knowledge-based system for customer service management. Expert Systems with Applications Volume: 24, Issue: 4, May, 2003:457-470
8. Ludwig, H, Analysis framework of complex service performance for electronic commerce. Database and Expert Systems Applications, 1998. Proceedings. Ninth International Workshop on , 26-28 Aug. 1998:638 – 643
9. R.Srikant, Q.Vu. Mining Association rules with Item Constraints [A]. Proc. Of the Third Int’l Conf. On Knowledge Discovery in Databases and Data Mining [C]. 1997:67-73.
10. R.T.Ng, V.S.Lakshmanan, J.Han, A.Pang. Exploratory mining and pruning optimizations of constrained association rules [A]. Proc. Of the 1998 AACM-SIGMOD Int’l Conf. On Management of Data [C]. 1998. 13-24.
11. Robertor J.Bayardo J, R. Agrawal. Constraint-based rule mining in large, dense database [A]. proc. Of the 15th int’l Conf. On Data Engineering [C]. 1999. 188-197.
12. Wang Fu-dong, Li Bing, Xue Jin-song, Zhu Yun-long. Constraint-based Association Rule Mining in CRM, COMPUTER INTEGRATED MANUFACTURING SYSTEMS. 2004 Vol.10 No.04 ： 465-470