

Fraud Detection of Maintenance Services Agents: A case study in Mobile Phone Industry

Helan Liang¹, Jie Zhang² and Xiuli Wu²

1Department of Logistics Engineering, University of Science and Technology Beijing,
Beijing, 100083, China

2Department of Mechanism Engineering, University of Science and Technology Beijing,
Beijing, 100083, China

Abstract. In this paper, In order to keep away from maintenance agents' frauds, a new method base on market segmentation and outlier detection is proposed for one mobile phone manufacturer who has 120 maintenance agents nationwide for providing after-sales maintenance services. Firstly, agents are segmented by a two-stage cluster algorithm according to their maintenance characters. Secondly, an outlier detection method for maintenance records analysis is proposed to mine out suspicious agents. While the method is validated with data from this mobile phone manufacturer, the underlying approach is more broadly applicable for industries with maintenance services agents.

Keywords: Market segmentation, outlier detection, fraud, data mining

1. Introduction

Agents for providing agile after-sales maintenance services is widely applied in industries with large sales and wide sales areas, such as automotive, household appliances etc.. [1] It is significant for the manufacturers to take measures for supervising the agents who are always aiming at earning more income even with frauds. in this paper, we propose a new method for fraud mining of maintenance service agents base on market segmentation and outlier detection and take the mobile phone industry as an example to show how the method is used. The paper is structured as follows. Section 2 described the general maintenance process and problems for agents management with the specific case of a mobile phone manufacturer. Section 3 describes the segmentation algorithm and the corresponding segment results of the specific case. Section 4 the outlier detection method is proposed and process of maintenance records analysis are reported. Finally, conclusions in Section 5.

2. Problem Description

Service agents system is that service providers (service station) who are appointed by the equipment providers (headquarters) are in charge of the regional equipment testing or replacement, so as to get the reward from the equipment providers [2]. As to the mobile phone manufacturer we investigated in this paper, it locates in ShenZhen, China and has 120 maintenance agents nationwide.

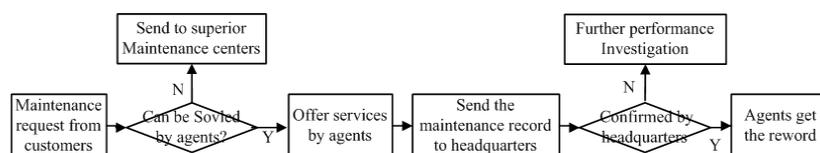


Fig.1 maintenance process description

Considering the agents locate all around the country, and the consumption of spare parts has close relations with regional factors, we can divide the agents into segments by the factors which influence the

maintenance behaviors. Also such similarity of maintenance characters will be reflected in the maintenance records as well, so to do the maintenance records analysis base on segments should be more reasonable. According to this, in this paper, firstly the segmatation theory[3-5] is used to divide service agents into segments with similar maintenance characteristics. Further, an outlier detection method was used for maintenance records analysis so as to mine out the suspicious agents within each maintenance segment.

3. Segmentation of Maintenance Agents

3.1. Segmentation criteria definition

Nowadays, the mobile phone manufacturer has grouped its agents geographically for easy management, such as agents of northeast areas, southern areas etc. However, such segments by one dimension criterion can not comprehensively reflect the similarity of maintenance behaviors. In order to subdivide the agents more precisely, this paper designs the segmentation criteria and segment algorithm according to the factors that impact both the materials consumption and the maintenance behaviors of various regions.

Tab.1 Segmentation criteria setting

Segment criteria	optional Values
Rank of the agent	Regional agent; District agent
Logistics conditions	Good, fair, poor
Distance from superior center	Near, fair, far
Climate patterns	monsoon climate, continental climate, high altitude mountain
geographical Distribution	Coast, Northeast, Northwest, midlands
Working condition of local network	good, fair, poor

As mobile phones are exposed in air, water vapor and dust in the air and regional temperature may influence the lifetime of components. Especially components which are highly sensitive to temperature, humidity, dust are likely to have higher consumption. Besides, the working condition of mobile phones has also been influenced by the local network. Typically, we found agents located in basin areas such as Chengdu where maintenance services referring to signal receiving are much more. Once the product is work out, the availability of spare parts may influence the customers' maintenance decisions. On one hand, The more convenience the regional logistics conditions, such as there are railways, highways and other efficient logistics facilities, the better the availability of spare parts. On the other hand, the nearer the location of the agent from the superior maintenance center can help for spare parts' agile replenishment. Besides, the availability of spare parts will also be influenced by rank of agents whose response time and spare parts varies according to its ranks. Sum up, in this paper, we define six criteria for segmentation and their optional values are shown in Table 1.

3.2 . Segmentation algorithm

The two-stage method base on hierarchical agglomerative clustering and K-means algorithm is designed in this paper. Considering that take the current geographical segmentation of agents as starting split point can not only keep the similarity characters, but also reduce the number of iterations. In summary, the algorithm is proposed as follows:

Step1: Take the geographical segmentation of agents as the initial segments, and calculate the respective centers of each segment.

Step2: calculate the Ward's distance and combine the two clusters which have the minimum distance. Calculate the condensation coefficient of the p -th iteration Del_p .

Step3: Repeat Step2 until all clusters are merged into one. Select the maximum Del_p and the corresponding number of clusters and clustering results of the p th iteration are used as the initial number and cluster centers of K-means algorithm

Step4: re-divide the agents by K-means algorithm.

3.3 Segmentation results and verification

According to the method mentioned, 120 agents of the mobile manufacturer are segmented. Firstly, set the specific values respectively of maintenance agents according to the criteria mentioned above. Segmentation results by the two-stage algorithm are as Table2 shown.

Tab.2 Segmentation Result

Segments	features of each segment
segment 1	thirty-one agents mainly located in coastal cities in south and east China with monsoon climate, convenience logistics and perfect network system
segment 2	forty-four agents mainly located in second-tier or third-tier cities in midlands with dry and continental climate and fair urban logistics.
segment 3	twenty-six agents mainly allocated in third-tier cities in north China which have poor urban logistics and most of them have continental climate
segment 4	nineteen agents mainly allocated in core cities in northwest and northeast areas with convenience traffic, and most of them are district maintenance agents as well

4. Outlier Detection

Outliers in a data set are commonly defined as individuals that are substantially different from the rest of the data. Such irregularities can indicate an error in the data, or abnormal behavior of the underlying system[6]. According to our problem, agents within each segments should have similar maintenance characteristics while those extremely different are likely with frauds. in order to find out such agents within each segment, an outlier detection method is proposed as follows:

4.1. Data preprocess

Firstly, calculate the consumption value of each component according to the maintenance records, and components which have larger consumption values are selected as A-type components. Secondly, calculate the consumption rate of each A-type components of each agent, after that, the mean and standard deviation of each A-type components in every maintenance segment can be obtained. Finally, create the standardized criterion matrix by Eq (1).

$$z_{ij} = \frac{x_{ij} - f_{jk}}{\sigma_{jk}} \quad i \in \{1,2,\dots,120\}, j \in \{1,2,\dots,8\}, k=1\dots 4 \quad (1)$$

Whereas, x_{ij} , z_{ij} are the original value and that after standardized of the j th criterion of the i th agent respectively. f_{jk} , σ_{jk} are the average consumption rate and standard deviation of j in the k th maintenance segment respectively.

4.2 .Outlier detection algorithm

Notice that when executing the hierarchical agglomerative clustering, clusters increase slowly or object combined latter are caused by its inconsistent with the majority of other objects [7-8], in this paper, the clustering coefficient is used for determining whether the agent is an outlier. The process is as follow:

Step1: execute the data preprocess. Calculate the Euclidean distance matrix. $k=1; p=1;$

Step2: Take each agent in the k th segments as one cluster, calculate the distance between clusters by Ward's method, and combine the two with minimum Ward's distance. Calculate the coefficient $Del_{k,p}=(D_{k,p+1} - D_{k,p})/ D_{k,p}$. whereas, $D_{k,p+1}$, $D_{k,p}$ are the squared residuals by Ward's method before and after the p th iteration.

Step3: Repeat Step2 till all agents are group into one cluster. Trace back from the final clusters, if $Del_{k,p}$ is larger than the threshold, divide into the original two clusters. If the original cluster is composed of one agent, put the agent into the outlier list.

Step4: $p=p-1$; repeat Step4, till $p=1$;

Step5: $k=k+1$; repeat Step2~Step4, till all segments are iterated and output the outlier list.

4.3 Results and model validation

Base on the outlier detection mentioned above, 143,841 maintenance records received by the mobile phone manufacturer in the first three quarters of 2011 are used and totally 6 outlier are mined out. Calculate the average deviations ratio of consumption rates of A-type components of the outlier agents from the

segment it belongs to, and results are as Fig.2 shown. We can see consumption rates of the outlier agents are much higher than its segment's average, which shows that results by the outlier method is reasonable.

Further, compare the method we proposed with that the company used at present which calculate average consumptions rate of agents without considering the maintenance segmentation. From Fig.3 We can see that consumption rate of most agents in segment 1 are higher than average while most of those in segment 3 and 4 are lower, so most suspicious agents found out will come from segment 1 while outliers in segments 3 and 4 can not be mined out. Besides, how many outliers should be mined out is not easy to be answered by traditional method. However, by the outlier method, it can automatically search out the outliers according to the clustering pedigree of agents and changes of clustering coefficient, which is more reasonable for fraud detection.

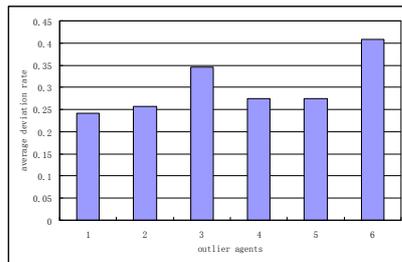


Fig.2 deviations of the outlier agents

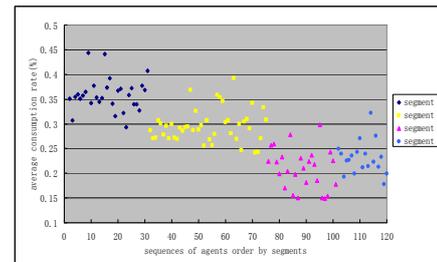


Fig.3 average consumptions rate of agents without considering segmentation

5. Conclusion

In this paper, a new method base on market segmentation and outlier detection is proposed to mine out the maintenance agents with frauds for one mobile phone manufacturer. Firstly, Its 120 maintenance agents are divided into four segments by a two-stage cluster algorithm according to six criteria which may impact regional maintenance behaviors. Secondly, Take its 143,841 maintenance records as original input and six suspicious agents are mined out by the outlier detection algorithm. Comparing the method we proposed with that the company used at present, we find that method base on maintenance segmentation and outlier detection is more reasonable. As maintenance agents are widely applied in industries with large sales and wide sales areas, the underlying approach can be more broadly used for other industries with maintenance services agents.

6. References

- [1]Liu Yingzi. How to Manage the Suppliers Engaged in Maintenance Service[J] .Logistics Technology, 2005,4:23-25.
- [2]Kotler, P. Marketing Managment: Analysis, planning, Implementation and Control[J]. 9th Ed,Prentice Hall, Inc., NewYork,NY.
- [3] R.J. Kuo, L.M. Ho, C.M. Hu. Integration of self-organizing feature map and K-means algorithm for market segmentation[J]. Computers & Operations Research. 2002,29:1475-1493.
- [4]Chen Ying,Wu Zhiwei. The Market Segmentation Research on Consumers' Behavioral Character of China's Commercial Bank[J]. HJournal of Financial ResearchH, 2007,3:71-85.
- [5] Liu Xiangxi, Huang Rizheng. Application of DataMing on Market Segmentation[J]. Industries forum, 2003:485-507.
- [6] Breunig M.Kriegel H -P.Ng R.Sander J.LOF:Identifying density-based local outliers[C].Proc of the 2000 ACM SIGMOD Int'l Confon Management of Data.Dallas:ACM Press,2000:93-104.
- [7] Angiulli F,Basta S,Pizzuti C. Distance-based detection and prediction of outliers[J].IEEE Trans on Knowledge and Data Engineering,2006,18(2):145-160.
- [8] Qiu Sulin, Wang Lizhen. Modified k-means algorithm based on Ward' s method and application[J].Computer Engineering and Applications,2008:44(31):169-172.