A Data Mining System for Potential Customers Based on One-Class Support Vector Machine

Weijian Mai 1, Fengjie Wu 2*, Fang Li1, Wenjun Luo3, and Xiaoting Mai4

1 School of Electronics and Communications Engineering, Guangzhou University, Guangzhou, Guangdong, 510006, China
2 Laboratory Center of Guangzhou University, Guangzhou University, Guangzhou, Guangdong, 510006, China
3 School of Computer Science and Network Engineering, Guangzhou University, Guangzhou 510006, China
4 School of Life Sciences, Guangzhou University, Guangzhou, Guangdong, 510006, China

*Corresponding author’s e-mail: wufengjie501@gzhu.edu.cn

Abstract. Commodity purchase data is usually severely skewed, which is reflected in the fact that there are far more negative data than positive data. This phenomenon makes it difficult for the binary classification model to obtain satisfactory results. Hence, we transform the binary classification problem into a one-class novelty detection problem. Specifically, this work proposes a potential customer mining system based on the One-Class Support Vector Machine (OCSVM) and demonstrates its effectiveness for classification, prediction, and potential customer mining. This system allows merchants to focus on unpurchased customers with the strongest purchase intentions and to change their purchase decisions with minimal sale costs, which enables merchants to maximize their benefits.

1. Introduction
In reality, there is a common problem with the purchase of commodities, that is, there are far more unpurchased samples than purchase samples. This phenomenon leads to a large gap between the positive and negative sample sizes in the data [1]. Using a binary classification model here will result in an extremely skewed model, making the classification result unreliable. Therefore, we transformed this binary classification problem into a one-class novelty detection problem using the One-Class Support Vector Machine (OCSVM) method [2]. The model performance would be evaluated over the dataset of electric vehicle purchases.

Note that merchants can further promote products for unpurchased customers via sales strategies. However, the effect of the sales strategy will be accompanied by an increase in service costs. Merchants need to make a trade-off between them to successfully promote their products with minimal cost. Our mining system allows merchants to focus on unpurchased customers with the strongest purchase intentions and calculate the required sales costs of making them willing to purchase. In this way, merchants can change purchasing decisions of unpurchased customers with minimal sale costs to maximize the benefits.
2. Data Pre-processing
In this problem, we are given a dataset of electric vehicle purchases. Before creating our model, we pre-process our data by filtering and cleaning, which allows us to obtain more reliable pieces of data and get a more intuitive interpretation of it.

- Data cleaning: we removed the abnormal values such as “NULL” and replaced them with the average value of the feature they belonged to.
- Statistics of data: In the dataset, there is a total of 1959 samples, in which there are 99 samples of Positive (purchase) and 1860 samples of Negative (unpurchased).

3. Methodology

3.1. Commodity purchase factors
Two principal factors \([3]\) affect the purchase decision of commodities: product satisfaction \(F_1\) and personal characteristics \(F_2\). In this work, the personal characteristics include age, work, and family situation, etc. Product satisfaction is the degree of personal satisfaction with different aspects of the product, including comfort, safety, and economy, etc. Given data of all these factors and the final purchase decisions, we propose a novel approach for classification, prediction, and potential customer mining.

3.2. One-Class SVM for Novelty Detection
Since binary classification models cannot achieve satisfactory results for problems with a skewed dataset, we transform this two-class classification problem into a one-class novelty detection problem based on the One-Class SVM (OCSVM) method.

3.2.1. Model Implementation. We utilize the python framework to build the OCSVM model. The hyperplane-based OCSVM algorithm \([4]\) can separate all data points from the zero point in the feature space and maximize the distance from the zero point to the separated hyperplane. This produces a binary function that can obtain the probability density region of the data in the feature space. It returns when in the training data point area, and -1 in other areas.

\[
\min_{\omega, \xi, \rho} \frac{1}{2} ||\omega||^2 + \frac{1}{n} \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad \langle \omega^T \phi(x_i) \rangle > \rho - \xi_i, i = 1, 2, \ldots, n \quad \xi_i \geq 0, i = 1, 2, \ldots, n
\]

In the formula, \(\omega\) and \(\rho\) are the normal vector and intercept of the obtained hyperplane, respectively, \(x_i = [F_1, F_2]\) is the feature vector, \(\xi_i\) is the slack variable, \(\nu \in (0,1)\) is similar to the regularization coefficient of the two-class SVM. \(\phi(x): x \rightarrow F\) is the mapping to the feature space \(F\), so that the dot product can be calculated by the kernel function: \(K(x_i, y_i) = \langle \phi(x_i) \cdot \phi(y_i) \rangle\). The kernel function used in this work is the Radial Basis Function (RBF) \([5]\):

\[
K(x_i, y_i) = \exp\left(-\frac{||x_i - y_i||^2}{2\sigma^2}\right)
\]

\(sgn(x)\) is a step function. Using Lagrange technology and dot product calculation, the decision function is obtained as:

\[
f(x) = sgn\left( \langle \omega^T \phi(x_i) \rangle - \rho \right) = sgn\left( \sum_{i=1}^{n} a_i K(x, x_i) - \rho \right)
\]

This method creates a decision hyperplane with parameters \(\omega, \rho\) in the feature space, which has the largest distance from the zero point in the feature space and separates the zero point from all data points (Figure 1).
3.2.2. Model Training. We used 80% of the negative data (unpurchased) as the training set and 20% as the test set. Note that the training and test set here do not include positive data (purchased). After training, we obtain a classification boundary. Then, we use the positive data as anomalous points for the model to classify and yield the detection accuracy of anomalous points as the model evaluation metric.

3.2.3. Model Evaluation. As shown in Figure 2, white points are training samples (negative), purple points are test samples (negative), yellow points are anomalous samples (positive), and the red closed curve is the learned dividing line, which classifies out-of-line as abnormal points. From the number of training errors 75/1488, the number of test errors 20/371, and the number of outlier detection errors 14/99, it can be concluded that the training accuracy, test accuracy, and novelty detection accuracy rates are 94.96%, 94.61%, and 85.86%, respectively.
3.2.4. Model Robustness. Considering the individual differences in real problems (e.g., some customers gave the product a high satisfaction score but did not buy it in the end), the novelty detection has reached a high accuracy rate and can efficiently separate the regular and abnormal points. However, the data given in this problem is relatively small, making individual differences have a greater impact on the overall accuracy. Therefore, the performance of the model can be further improved by collecting more data. The phenomenon reflects the robustness of our model, which can continuously update the classification curve as the amount of data. Meanwhile, the proposed method will not be troubled by the serious data skewed problem.

3.3. Data Mining System for Potential Customers

3.3.1. Purchase Decision Prediction. After the model is trained, we take the data of 15 customers as input and predict their purchasing decisions. As shown in Figure 3, points within the classification boundary are classified as negative (unpurchased), and points outside the boundary are classified as positive (purchased). Therefore, three customers will make a purchase decision.

![Abnormal Detection](image)

**Figure 3. Purchase Decision Prediction**

3.3.2. Data Mining System for potential customers. For unpurchased customers, merchants can sell products through the improvement of product satisfaction. However, the more the product satisfaction is improved, the more services need to be paid, which increases the cost of the merchants. Merchants want to sell their products successfully with minimal cost. To maximize the benefits for merchants, we developed a data mining system to help merchants focus on the unpurchased customers with the strongest purchase intentions.

Suppose the increment of service degree for the $i^{th}$ customer is $\Delta S_{ser}$, and the increment of satisfaction is $\Delta S_{sat}$. The decision function of OCSVM is $f(x)$, in which the decision value of unpurchased customers is $f(x) = 1$ (regular points), the decision hyperplane is $f(x) = 0$, and $f(x) = -1$ represents purchased customer (abnormal points), which can be described as:
\[ \min \Delta S_{er_i} \rightarrow \min \Delta S_{at_i} \] (5)

Known by:
\[ f(x_i) = 1 \] (6)
\[ x_i = [F_1, F_2] \]

Can be transformed into:
\[ \min \Delta S_{at_i} \rightarrow \min \Delta x_i \rightarrow \min \Delta F_1 \] (7)
\[ \text{s.t. } f(x) = \text{sgn} \left( (\omega^T \phi(x_i)) - \rho \right) = \text{sgn} \left( \sum_{i=1}^{n} a_i K(x,x_i) - \rho \right) < 0 \] (8)

Finally, we obtained the minimal \( \Delta F_1 \) for each unpurchased customer, representing the minimum increment of product satisfaction required for customers to change the purchase decision. After that, we can select the potential customers with the strongest purchase intentions by sorting the \( \Delta F_1 \) value of each customer in ascending order, which is 0.8% (select), 1.3% (select), 3.7% (select), 3.9%, 4.3%, 4.5%, 4.9%, 5.1%, 5.7%, 6.3%, 6.5%, 7.0%, 7.4%, 7.9% and 8.9% for 15 customers, respectively.

Figure 4 shows the visualization results of potential customer selection (points in orange circles). Note that the decision hyperplane obtained from the training section has not changed in this mining process, only the value of \( x_i(\Delta F_1) \) is changed by implementing the mining strategy to make the unpurchased point cross the decision hyperplane (Figure 5).

Our model can be continuously updated as the amount of data to provide more accurate customer mining strategies. For customers with a larger \( \Delta F_1 \) (>6%), more service costs need to be paid, so we suggest abandoning these customers.

In summary, the proposed mining model can effectively select potential customers with the strongest purchase intentions from unpurchased customers, so that merchants can sell products by offering the minimum increment of service, thus maximizing the benefits.

4. Conclusions
Since binary classification models cannot offer satisfactory results for the skewed dataset, we propose the OCSVM model to convert this binary classification problem into a one-class novelty detection problem. Based on the OCSVM model, this work takes the electric vehicle purchases as an example to demonstrate the effectiveness of our model for classification, prediction, and potential customers mining. This system allows merchants to focus on unpurchased customers with the strongest purchase intentions and calculate the required service increment of making them willing to purchase. In this way, merchants
can change the purchase decision of unpurchased customers with minimal sale costs to maximize the benefits. Considering the impact of individual differences on overall performance, the application of our system in small datasets might be limited. However, as the amount of data increases, personal influence is gradually reduced, which will improve the performance of our mining system. Furthermore, more comprehensive information about product purchases also contributes to improving system performance. In future work, it is possible for us to develop a more efficient customer mining system based on massive data and comprehensive purchase information.

Acknowledgments
This work was supported in part by the Climbing Program of Guangdong Province, China, under Grant pdjh2020b0462, and in part by the National College Student Innovation and Entrepreneurship Training Program, China, under Grant S202111078097.

References
[1] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004.
[2] Schölkopf, Bernhard, et al. "Support vector method for novelty detection." NIPS. Vol. 12. 1999.
[3] Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." Wiley interdisciplinary reviews: computational statistics 2.4 (2010): 433-459.
[4] Schölkopf, Bernhard, et al. "Estimating the support of a high-dimensional distribution." Neural computation 13.7 (2001): 1443-1471.
[5] Liu, Quanzhong, et al. "Feature selection for support vector machines with RBF kernel." Artificial Intelligence Review 36.2 (2011): 99-115.