



# Insider Trading Detection Algorithm in Industrial Chain Based on Logistics Time Interval Characteristics

Fulin Chen<sup>1</sup>, Kai Di<sup>2</sup>(✉), Hansi Tao<sup>2</sup>, Yuanshuang Jiang<sup>2</sup>, and Pan Li<sup>1</sup>

<sup>1</sup> School of Cyber Science and Engineering, Southeast University, Nanjing 211189, Jiangsu, China

<sup>2</sup> School of Computer Science and Engineering, Southeast University, Nanjing 211189, Jiangsu, China

dikai@seu.edu.cn

**Abstract.** Insider trading behavior is becoming increasingly prevalent with the rapid development of the industrial chain. Insider trading refers to the illegal behavior of conducting insider trading by obtaining insider information. The existing insider trading detection methods of industrial chain do not consider the problems of inefficient industrial chain data characteristics and long trading time span, resulting in poor algorithm effect. Therefore, in order to solve the above problems, this paper proposes an algorithm for detecting insider trading in the industrial chain based on logistics time interval characteristics. Firstly, aiming at the problem of inefficiency of industrial chain data characteristics, this algorithm proposes a logistics index construction method for describing the whole process of insider trading behavior; Secondly, aiming at the problem of long time span of transaction, a dynamic sliding window method is proposed; Finally, the isolation forest algorithm is improved to identify the abnormal data. Verified under the real data set, the results show that compared to using the isolation forest methods, the F1 value of the insider trading behavior detection problem of the industry chain can be improved by 20.68% by using the logistics time interval feature.

**Keywords:** Industrial chain · Insider trading detection · Anomaly detection

## 1 Introduction

With the development of the global supply chain, the safe and stable operation of the industrial chain plays a vital role in the economic development of a country. At present, the environment of the industrial chain trading market is complex, and there are various trading risks [1]. Insider trading behavior is one of the risks. Insider trading refers

This work was supported by the National Key Research and Development Program of China (No. 2022YFB3304400), the National Natural Science Foundation of China (Nos. 62076060, 61932007), the Key Research and Development Program of Jiangsu Province of China (No. BE2022157), and the Defense Industrial Technology Development Program (JCKY2021214B002).

to traders directly or indirectly using inside information to buy and sell commodities and obtain improper economic benefits. Inside information refers to the non-public information obtained by the staff of some financial institutions or regulatory departments of the industrial chain due to the convenience of their position or position. Since insider information can affect the price in the trading market, users who have access to insider information can spread it to others [2], thereby indirectly obtaining excess profits. This behavior affects the fairness of the industrial chain trading market and threatens the healthy development of the industrial chain trading market. In the supervision process of insider trading, it shows the characteristics of strong concealment of insider trading [3]. Therefore, this paper designs an insider trading behavior detection algorithm of industrial chain based on the characteristics of logistics time interval to mine the hidden insider trading behavior.

At present, the detection of insider trading behavior of industrial chain usually uses the regularized abnormal indicators of dealer trading patterns and the data-driven model methods [4] to identify the abnormal behavior of dealers. For the methods of regularized abnormal indicators, due to the characteristics of physical delivery of the industrial chain, commodities can be traded not only online but also offline, so the whole process of circulation of commodities cannot be supervised, which leads to the inapplicable income measurement indicators based on informed trading. For the data-driven model methods, the direct use of logistics features has a high dimension and the model learning is difficult, which makes the detection effect of insider trading behavior not good. Therefore, in order to help regulators efficiently supervise the insider trading behavior in the industrial chain trading market, this paper studied the problem of insider trading behavior detection based on logistics characteristics in the industrial chain trading market.

The main contributions of this paper are summarized as follows:

- 1) Firstly, aiming at the problem of low efficiency of industrial chain data characteristics, the algorithm proposed a construction method of logistics indicators to describe the whole process of insider trading behavior according to the three dimensions of own trading mode, commodity trading mode and dealer mode;
- 2) Secondly, aiming at the problem of long time span of transaction, a method of using dynamic sliding window is proposed to extract the logistics characteristics within the time interval, and then judge whether the time interval is abnormal;
- 3) Finally, the isolation forest algorithm is improved to identify the abnormal data in the abnormal window. The experimental results show that compared to using the isolation forest methods, the F1 value of the algorithm in this paper is increased by **20.68%** in identifying the insider trading behavior of traders in the industrial chain trading market.

## 2 Related Work

In terms of methods, insider trading detection in the industrial chain market can be divided into model-based and data-driven insider trading detection methods [4].

For the model-based insider trading detection methods, Fama et al. proposed the event study method to measure the normal returns and abnormal returns before and after a certain event [5], which can be used to judge insider trading. Easley et al. mathematically

described the trading process of informed and uninformed people, and proposed a model to estimate the probability of informed trading, which is used to estimate the probability of whether a transaction is an informed trader [6]. Mienna proposed a statistical model based on long time series data sets [7] in view of the complex trading strategies of insider traders and the difficulty of calculating additional returns in traditional econometrics models. Cline et al. considered the partial observability of insider trading and proposed a bivariate probability model to detect the behavior of illegal insider trading [8].

For data-driven insider trading detection methods, Deng et al. proposed a Gradient boosted decision tree (GBDT) based approaches for insider trading detection with differential evolution (DE) for parameter initialization [9, 10]. Esen et al. proposed a clustering-based insider trading detection method, which takes the outlier value of trading behavior as the suspicion degree of insider trading behavior through K-means and hierarchical clustering method and verifies it through the event study method [11]. Islam proposed the method of using Long Short-Term Memory network (LSTM) in deep learning to learn the structured and unstructured features of illegal insider trading events [12]. Seth et al. proposed a multi-stage insider trading detection method including deep neural network, consensus model and statistical methods to identify illegal insider trading behaviors through event analysis and detection of unstructured and structured data [13]. Lauer et al. proposed to build a training data set based on news events before insider trading events and proposed an augmentation method to expand the size of the data set, and used XGBoost to predict insider trading events [14].

However, the above methods do not consider the problems of inefficiency of industrial chain data characteristics and long transaction time span, so that the previous index modeling cannot be applied, resulting in poor algorithm effect. Therefore, this paper designs an insider trading behavior detection algorithm of industrial chain based on the characteristics of logistics time interval to mine the hidden insider trading behavior.

### 3 Problem Formulation and Analysis

In this section, the problem of insider trading detection in the industrial chain trading market is formally defined. The trading of industrial chains is different from other electronic transactions, which involves the transportation of goods logistics and has a large time span, so it is necessary to combine the characteristics of logistics to detect insider trading behavior. Next, the insider trading detection problem based on the characteristics of logistics time interval in the industrial chain is defined in detail.

Suppose there are a number of dealers that have traded during the time interval  $[t_a, t_b)$ , and we use sets  $A = \{a_1, \dots, a_i, \dots, a_N\}$  to represent these trading accounts. A trader  $a_i \in A$ , trading in  $M$  types of commodities, we define it as  $C^i = \{c_1^i, \dots, c_j^i, \dots, c_M^i\}$ . Among them, where  $K$  transactions on the  $j$ -th commodity are represented as  $R_j^i = \{r_{j,1}^i, \dots, r_{j,k}^i, \dots, r_{j,K}^i\}$ , then the  $k$ -th transaction can be represented as  $r_{j,k}^i = \langle s_{j,k}^i, e_{j,k}^i, p_{j,k}^i, v_{j,k}^i, f_{j,k}^i \rangle$ , where  $f_{j,k}^i = 1$  represents the selling transaction behavior and  $f_{j,k}^i = -1$  represents the buying transaction behavior.

The problem of insider trading behavior detection [15] is defined as: judging whether the trading behavior  $r_{j,k}^i$  in the trading behavior data  $R_j^i$  of dealers is insider trading behavior. The variables involved are listed in Table 1.

**Table 1.** Variable description

variables	Description
$a_i$	Trading account $a_i$
$r_{j,k}^i$	The the $k$ -th transaction made by trading account $a_i$ on the trading behavior of $j$ -th commodity categories
$s_{j,k}^i$	The start time of the $k$ -th transaction of the transaction account $a_i$ on the $j$ -th commodity
$p_{j,k}^i$	Commodity price of the $k$ -th transaction of the transaction account $a_i$ on the $j$ -th commodity
$v_{j,k}^i$	The number of items of the $k$ -th transaction of the transaction account $a_i$ on the $j$ -th commodity
$e_{j,k}^i$	The completion time of the $k$ -th transaction of the transaction account $a_i$ on the $j$ -th commodity
$L$	Sliding window time length
$H$	Detecting sequence length

## 4 Algorithm Design

### 4.1 Interval Logistics Index Construction Algorithm

The main function of interval logistics characteristic index construction is to use indicators [16] under different dimensions to describe insider trading behavior and normal trading behavior in the industrial chain trading market. The specific steps of interval logistics characteristic index construction algorithm are as follows:

**Using Sliding Window to Divide the Logistics Behaviour.** The logistics behavior of the trading account  $a_i$  that detects the time interval of  $[t_a, t_b)$  on the  $j$ -th commodity is divided into  $N$  segments using the unit time interval length  $T$ , where  $T$  is expressed as the Eq. 1:

$$T = \frac{t_b - t_a}{N} \quad (1)$$

Therefore, the time interval sequence obtained by partitioning is expressed in Eq. 2:

$$S = \{ [t_a, t_a + T), [t_a + T, t_a + 2T), \dots, [t_a + (N - 1)T, t_b) \} \quad (2)$$

According to the characteristics that the logistics interval will span multiple units of time  $T$ , a sliding window of length  $L$  is used to extract the logistics characteristics of this logistics behavior, and the time interval of the sliding window is expressed as Eq. 3:

$$SS = \{ [t_a, t_a + L), [t_a + T, t_a + T + L), \dots, [t_a + NT - L, t_b) \} \quad (3)$$

According to different types of logistics behaviors of traders, logistics behaviors related to logistics characteristics in the sliding window within  $[t_x, t_x+L)$  can be divided into buying logistics behaviors and selling logistics behaviors, in which logistics behaviors entering the warehouse are represented by  $B_{t_x}$  in formula 4, and logistics behaviors exiting the warehouse are represented by  $S_{t_x}$  in Eqs. 4 and 5.

$$B_{j,t_x}^i = \{ (t_{j,k}^i, e_{j,k}^i, p_{j,k}^i, v_{j,k}^i, f_{j,k}^i) | t_x \leq e_{j,k}^i < t_x + L, f_{j,k}^i = 1 \} \quad (4)$$

$$S_{j,t_x}^i = \{ (t_{j,k}^i, e_{j,k}^i, p_{j,k}^i, v_{j,k}^i, f_{j,k}^i) | t_x \leq s_{j,k}^i < t_x + L, f_{j,k}^i = -1 \} \quad (5)$$

### Calculating Logistics Indicators of Incoming Warehouse and Outgoing Warehouse.

The total value of goods corresponding to logistics behaviors of incoming warehouse and outgoing warehouse are expressed as Eqs. 6 and 7:

$$U_{j,t_x}^i = \sum_{R_{j,k}^i \in B_{t_x}} p_{j,k}^i * v_{j,k}^i \quad (6)$$

$$V_{j,t_x}^i = \sum_{R_{j,k}^i \in S_{t_x}} p_{j,k}^i * v_{j,k}^i \quad (7)$$

For different types of logistics behavior, the contributions to the logistics index are different. Therefore, the commodity value of logistics behaviors of incoming warehouse and outgoing warehouse accounting for their total logistics behaviors are defined as Eqs. 8 and 9, respectively.

$$\alpha_{j,t_x}^i = \frac{U_{j,t_x}^i}{U_{j,t_x}^i + V_{j,t_x}^i} \quad (8)$$

$$\beta_{j,t_x}^i = \frac{V_{j,t_x}^i}{U_{j,t_x}^i + V_{j,t_x}^i} \quad (9)$$

**Calculating the Characteristic Indicators of Three Dimension.** Next, in order to better describe the indicators proposed, we have made the following definitions:

**Definition 1.** Total value of goods in interval: it represents the total value of goods in and out of warehouse on the logistics characteristics within the sliding window, and the outlier of the timing relationship of the trader's logistics behavior on this kind of goods. The logistics behavior in this time zone, the total value of the goods that are shipped out and shipped into the warehouse are expressed as Eq. 10:

$$O_{j,t_x}^i = U_{j,t_x}^i + V_{j,t_x}^i \quad (10)$$

In the formula, the first item represents the total value of all goods entering the warehouse in the time range of  $[t_x, t_{x+L})$ , which is expressed as the commodity price  $p_{j,k}^i$  multiplied by the quantity  $v_{j,k}^i$  of all purchasing logistics behaviors, and the second item represents the total value of all goods exiting the warehouse.

**Definition 2.** Ratio of interval commodity trading to all traded commodities: it represents the ratio of the value of commodities in the sliding window to the sum of the value of commodities traded by all users in the sliding window and is expressed as Eq. 11:

$$P_{j,t_x}^i = \alpha_{j,t_x}^i \frac{U_{j,t_x}^i}{\sum_j U_{j,t_x}^i} + \beta_{j,t_x}^i \frac{V_{j,t_x}^i}{\sum_j V_{j,t_x}^i} \quad (11)$$

In the formula,  $\alpha_{j,t_x}^i$  and  $\beta_{j,t_x}^i$  represent the proportion of logistics activities of incoming warehouse and outgoing warehouse respectively; In the first term, the numerator represents the total value of the goods  $j$  of the incoming warehouse in the time range  $[t_x, t_{x+L})$ , and the denominator represents the sum of the total value of all the goods of the incoming warehouse in the time range. In the second term, the numerator represents the total value of the goods  $j$  of the outgoing warehouse in the time range  $[t_x, t_{x+L})$ , and the denominator represents the sum of the total value of all the goods of the outgoing warehouse in the time range.

**Definition 3.** The ratio of interval commodity trading to all dealers: it represents the ratio of the sum of commodity values of the logistics behavior of the dealer and all dealers on this commodity, which is expressed as Eq. 12.

$$Q_{j,t_x}^i = \alpha_{j,t_x}^i \frac{U_{j,t_x}^i}{\sum_i U_{j,t_x}^i} + \beta_{j,t_x}^i \frac{V_{j,t_x}^i}{\sum_i V_{j,t_x}^i} \quad (12)$$

In the formula, the numerator represents the total value of goods stored in the warehouse within the time range. The denominator represents the sum of the total value of the inventory of all the traders who traded in the commodity during the time period.

## 4.2 Interval Anomaly Detection Algorithm for Logistics Characteristics

The main function of abnormal interval detection based on interval logistics characteristic indexes is to detect abnormal interval according to the logistics characteristic indexes and the surrounding normal indexes. The specific steps of the algorithm are given in Algorithm 1, and we describe its detailed process as follows:

1) Construct index sequence composed of sliding window and surrounding time interval. For the  $r$ -th index  $F_{j,r,t_x}^i$  of the  $j$ -th commodity logistics of the  $i$ -th trader in the time range  $[t_x, t_{x+L}]$ , the logistics index within the detection time interval  $H$  is expressed as Eq. 13:

$$FS_{j,r,t_x}^i = \left\{ F_{j,r,t_x - \frac{H}{2}L}^i, F_{j,r,t_x - \frac{H}{2}(L-1)}^i, \dots, F_{j,r,t_x}^i, \dots, F_{j,r,t_x + \frac{H}{2}(L-1)}^i, F_{j,r,t_x + \frac{H}{2}}^i \right\} \quad (13)$$

When detecting whether each sliding window is abnormal with the surrounding sliding window, the length of  $L$  time interval is used to select in turn, so as to ensure

that the sliding window to be detected does not overlap. During sequence exploration, it is necessary to ensure that it does not exceed the boundaries of the original indicator sequence [17].

2) Calculate the anomaly index corresponding to the sliding window. For the  $r$ -th index outlier of the  $j$ -th commodity logistics of the  $i$ -th trader in the time range  $[t_x, t_x+L]$  is expressed as Eq. 14:

$$S_{j,r,t_x}^i = \frac{F_{j,r,t_x}^i - \text{median}(FS_{j,r,t_x}^i)}{MAD_{FS_{j,r,t_x}^i}} \quad (14)$$

The function of *median* represents the Median of samples in the time series, and represents the Median Absolute Deviation (MAD) of samples [18].

3) Determine whether the sliding window is abnormal. For each feature, its outliers are calculated, and its total outliers are expressed as Eq. 15.  $\rho_r$  represents the parameters of this indicator:

$$S_{j,t_x}^i = \sum_r \rho_r S_{j,r,t_x}^i \quad (15)$$

Whether the interval is abnormal is judged according to the relationship with threshold  $\delta_s$ . The robustness of the anomaly detection algorithm is enhanced by weighted fusion of abnormal degrees of multiple dimensional indicators [19].

### 4.3 Abnormal Logistics Feature Detection Based on Isolation Forest

In this paper, the method based on isolation forest [20] is used to perform anomaly detection on the data in the detected anomaly interval, which can improve the effectiveness of the whole anomaly detection algorithm and enhance the interpretability of the algorithm [21]. The specific steps of the abnormal logistics feature detection algorithm based on isolation forest are given in Algorithm 2, and we present its specific description as follows:

- 1) The newly constructed feature vector for each transaction is expressed as  $\langle p_{j,k}^i, v_{j,k}^i, f_{j,k}^i, An_{j,k}^i \rangle$ ;
- 2) Mark the transaction  $R_{j,k}^i = \langle s_{j,k}^i, e_{j,k}^i, p_{j,k}^i, v_{j,k}^i, f_{j,k}^i \rangle$  to determine whether it is within the  $m$ -th anomaly interval  $[t_s^m, t_e^m]$ , and if so, it is marked as:

$$An_{j,k}^i = \begin{cases} 1, & t_s \leq s_{j,k}^i < t_e f_{j,k}^i = 1 \text{ or } t_s \leq e_{j,k}^i < t_e f_{j,k}^i = -1 \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

The feature vector is input into the isolation forest algorithm to determine whether each logistic behavior of the user is abnormal.

## 5 Experiments

### 5.1 Experimental Environment

The experiments were run on a PC with windows10 operating system, the Processor was AMD Ryzen 7 3800X 8-Core Processor 3.89 GHz. The experimental code was implemented using Python 3.8.13.

The experimental data comes from the real industry chain data. Due to the risk of leakage of logistics data for traders' privacy, therefore, all the data provided were desensitized according to the industry chain electronic transaction sensitive information desensitization and encryption regulations. All traded commodity names and dealer names in this paper are represented using a post-desensitized string.

### 5.2 Performance Indices and Benchmark Algorithms

**Experimental Indicators.** For classification problems, the following indicators are usually considered [21]:

- 1) Accuracy rate: all predicted dealer associations are consistent with the actual user associations.
- 2) Precision rate: the proportion of actual transactions that are abnormal among all transactions that are predicted to be abnormal.
- 3) Recall rate: the proportion of actual transactions predicted to be abnormal among all transactions that are abnormal.
- 4) F1 value: Considering the above two indicators.
- 5) Running time.

**Comparison Algorithm.** We have chosen the following three comparative algorithms:

- 1) Using the Isolation Forest algorithm (IForest) for detection [20], randomly selecting features from the original features to construct an isolated tree, and calculating the outliers of the sample through the path length of the sample. The shorter the path length, the greater the outlier value;
- 2) The Histogram Based Outlier Score (HBOS) method [22] is used to divide each dimension of the data into intervals, and the outlier value of the sample is calculated through the density of the interval where the sample is located. The lower the density, the greater the outlier value;
- 3) The anomaly detection method of K Nearest Neighbor (KNN) [23] is used to calculate the outliers by calculating the distance between the sample and the surrounding points. The larger the distance, the larger the outlier.

**Parameter Settings.** The effectiveness of the algorithm was verified by randomly sampling the original data set from 50% to 100% to obtain different sizes of data sets. 50 abnormal transactions are generated per day, where the number of items in each abnormal transaction is set to 100. Sequence detection length  $H = 6$ .



**Table 2.** Experimental index results under different transaction sample numbers

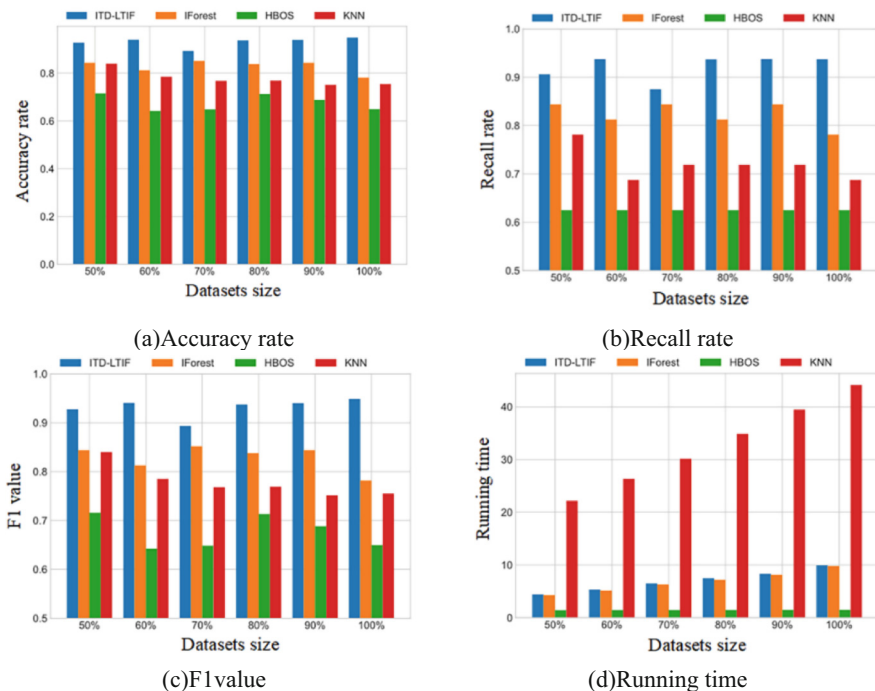
Datasets size %	Algorithm name	Accuracy rate	Recall rate	F1 value	Running time/second
50%	ITD-LTIF	<b>0.9277</b>	<b>0.9062</b>	<b>0.9168</b>	4.419
	IForest	0.8438	0.8438	0.8438	4.291
	HBOS	0.7158	0.6250	0.6673	<b>1.454</b>
	KNN	0.8401	0.7812	0.8096	22.185
70%	ITD-LTIF	<b>0.8934</b>	<b>0.8750</b>	<b>0.8841</b>	6.484
	IForest	0.8517	0.8438	0.8477	6.311
	HBOS	0.6485	0.6250	0.6365	<b>1.473</b>
	KNN	0.7682	0.7188	0.7427	30.172
100%	ITD-LTIF	<b>0.9488</b>	<b>0.9373</b>	<b>0.9430</b>	9.956
	IForest	0.7817	0.7812	0.7814	9.817
	HBOS	0.6498	0.6250	0.6372	<b>1.503</b>
	KNN	0.7550	0.6875	0.7197	44.129

### 5.3 Experimental Result

As shown in Table 2 and Fig. 1(a)(b)(c), the insider trading detection method based on time interval characteristics proposed in this paper is superior to other comparison algorithms in terms of precision, recall rate and F1 value on transaction data sets of different sizes. Compared with the isolation forest algorithm, the proposed algorithm improves the accuracy index by **9.94%**, **15.74%**, **4.89%**, **11.84%**, **11.37%**, **21.37%** respectively. In terms of the recall rate, the proposed algorithm improves by **7.39%**, **15.38%**, **3.69%**, **15.34%**, **11.12%**, **19.98%** respectively; In terms of F1 value, the algorithm proposed in this paper improves by **8.65%**, **15.55%**, **4.29%**, **13.59%**, **11.24%**, **20.68%** respectively.

As shown in Table 2 and Fig. 1(d), the proposed insider trading detection algorithm based on logistics time interval characteristics is not significantly different from the isolation forest method in terms of time, and the algorithm execution time shows an increasing trend with the increase of transaction datasets size. The histogram based anomaly detection algorithm takes the least time, and with the increase of the size of the transaction data set, the execution time of the algorithm has little effect. The anomaly detection method based on k-nearest neighbor method takes the longest time and the execution time of the algorithm increases significantly with the increase of the size of the transaction data set.

**Results Analysis.** In terms of precision, recall rate and F1 index, the insider trading detection algorithm based on logistics time interval characteristics proposed in this paper is superior to other comparison algorithms, but there is no obvious change in the trend, which is due to the randomness of sample sampling, and different sizes of transaction data sets are different in the distribution of normal trading behavior. As for the running time of the algorithm, the insider trading detection algorithm based on the characteristics



**Fig. 1.** Variation of experimental indicators with the proportion of datasets size

of logistics time interval proposed in this paper has little difference. This is because although the algorithm proposed in this paper needs more time to detect the additional time interval, the isolation forest method has a higher time dimension and requires more time. Therefore, the running time of the two algorithms is not greatly different. Histogram-based anomaly detection takes the least time due to its simple detection method and fast computation, so the increase of its dataset size is not enough to have a noticeable effect on it. The k-nearest neighbors algorithm takes the longest time, because it takes a lot of time to calculate the distance between its K neighboring points. With the increase of the proportion of the datasets size, the algorithm needs to spend more running time on transaction detection.

## 6 Conclusions

This paper studies the detection of insider trading behavior based on logistics characteristics in the industrial chain trading market, and proposes an insider trading detection algorithm based on logistics time interval characteristics. Firstly, according to the characteristics of long logistics time span, the dynamic sliding window is used to extract the logistics transaction behavior in the time interval. Secondly, three characteristics under the dimension of own trading mode, commodity trading mode and dealer mode are used to describe. Then, after the abnormal logistics time intervals are identified, the overlapping abnormal time intervals are optimized into non-overlapping abnormal intervals

by using the method of reducing the sliding window to improve the accuracy of the algorithm. Finally, the method based on isolation forest was used to judge the abnormal transaction behavior of the extracted effective logistics features, so as to improve the effect of the algorithm. Through the experiments on the logistics data sets of real trading behavior of dealers, the results show that the proposed algorithm improves the F1 value by 20.68% compared with the direct using of isolation forest.

## References

1. Angelopoulos, J., Sahoo, S., Visvikis, I.D.: Commodity and transportation economic market interactions revisited: new evidence from a dynamic factor model. *Transp. Res. Part E: Logist. Transp. Rev.* **133**, 101836 (2020)
2. Baklarz, A., Bogusz, J., Martysz, C.: Models of Propagation of Inside Information. *Acta Physica Polonica A* **138**(1) (2020)
3. Adams, B.J., Perry, T., Mahoney, C.: The challenges of detection and enforcement of insider trading. *J. Bus. Ethics* **153**(2), 375–388 (2018)
4. Hilal, W., Gadsden, S.A., Yawney, J.: Financial fraud: a review of anomaly detection techniques and recent advances. *Expert Syst. Appl.* **193**, 116429 (2022)
5. Fama, E.F., Fisher, L., Jensen, M.C., et al.: The adjustment of stock prices to new information. *Int. Econ. Rev.* **10**(1), 1–21 (1969)
6. Easley, D., Kiefer, N.M., O'hara, M., et al.: Liquidity, information, and infrequently traded stocks. *J. Finance* **51**(4), 1405–1436 (1996)
7. Minenna, M.: Insider trading, abnormal return and preferential information: supervising through a probabilistic model. *J. Bank. Finance* **27**(1), 59–86 (2003)
8. Cline, B.N., Posylnaya, V.V.: Illegal insider trading: commission and sec detection. *J. Corp. Finan.* **58**, 247–269 (2019)
9. Deng, S., Wang, C., Wang, M., et al.: A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market. *Appl. Soft Comput.* **83**, 105652 (2019)
10. Deng, S., Wang, C., Fu, Z., et al.: An intelligent system for insider trading identification in Chinese security market. *Comput. Econ.* **57**(2), 593–616 (2021)
11. Esen, M.F., Bilgic, E., Basdas, U.: How to detect illegal corporate insider trading? A data mining approach for detecting suspicious insider transactions. *Intell. Syst. Account. Finan. Manage.* **26**(2), 60–70 (2019)
12. Islam, S.R., Khaled Ghafoor, S., Eberle, W.: Mining illegal insider trading of stocks: a proactive approach. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 1397–1406. IEEE, Seattle (2018)
13. Seth, T., Chaudhary, V.: A predictive analytics framework for insider trading events. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 218–225. IEEE, Atlanta (2020)
14. Lauer, F., Arbex Valle, C.: Detecting and predicting evidences of insider trading in the Brazilian market. In: Dong, Y., Ifrim, G., Mladenčić, D., Saunders, C., Van Hoecke, S. (eds.) *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track. ECML PKDD 2020. LNCS*, vol. 12461, pp. 241–256. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-67670-4\\_15](https://doi.org/10.1007/978-3-030-67670-4_15)
15. Donoho, S.: Early detection of insider trading in option markets. In: *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 420–429. Association for Computing Machinery, New York (2004)
16. Tangwongsan, K., Hirzel, M., Schneider, S., et al.: General incremental sliding-window aggregation. *Proc. VLDB Endowment* **8**(7), 702–713 (2015)

17. Blázquez-García, A., Conde, A., Mori, U., et al.: A review on outlier/Anomaly detection in time series data. *ACM Comput. Surv.* **54**(3), 56:1–56:33 (2021)
18. Howell, D.C.: Median Absolute Deviation. *Encyclopedia of Statistics in Behavioral Science*. Wiley, New York (2005)
19. Sun, H., He, Q., Liao, K., et al.: Fast anomaly detection in multiple multi-dimensional data streams. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 1218–1223 (2019)
20. Ounacer, S., Bour, H.A.E., Oubrahim, Y., et al.: Using isolation forest in anomaly detection: the case of credit card transactions. *Periodicals Eng. Nat. Sci.* **6**(2), 394–400 (2018)
21. Han, S., Hu, X., Huang, H., et al.: ADBench: anomaly detection benchmark. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2022)
22. Kalaycı, İ., Ercan, T.: Anomaly detection in wireless sensor networks data by using histogram based outlier score method. In: 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1–6 (2018)
23. Ying, S., Wang, B., Wang, L., et al.: An improved KNN-based efficient log anomaly detection method with automatically labeled samples. *ACM Trans. Knowl. Disc. Data* **15**(3), 34:1–34:22 (2021)