Computers in Industry 92 (2017) 43–49

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Research on outlier detection algorithm for express logistics

Yan Zhang[a,b], Zheyi Li[a], Jiefeng Wen[a], Yiwei Ma[a], Oh-kyoung Kwon[b,*]

[a] Chongqing University of Post and Telecommunication, Chongqing 400065, China
[b] Graduate School of Logistics, INHA University, Incheon 402751, Korea

ARTICLE INFO

Article history:
Received 14 September 2016
Received in revised form 2 May 2017
Accepted 23 May 2017
Available online xxx

Keywords:
Express logistics
Multi-attribute
Outlier detection

ABSTRACT

To detect the problems of time delay, path error and destination error in express logistics process effectively, a novel outlier detection algorithm for express logistics is proposed in this paper. To test the detection results, the express logistics system operating model is built to test the detection results. Experiment results show that the proposed algorithm is well applied to the express logistics data with multi-attribute characteristics, and can work well in detecting the abnormal conditions of express logistics.

1. Introduction

The joint distribution [1–3] is the development direction of the modern scale express logistics distribution mode [4–6]. The problem of heavy traffic and complex route [7,8] will be happened because of its large-scale load distribution mode [9] and no distinguish between shippers and commodity [10]. This will lead to delays in distribution time, the distribution path error and the destination error in some degree. Therefore, it is a very important part of modern large-scale express logistics distribution mode that how to find out the abnormal distribution problems quickly and feedback to the enterprises for dynamic processing in time.

At present, the outlier detection technology for express logistics is still in the initial stage, because most of the outlier detection technique researches are mainly focused on the simple and structured dataset [11]. It is mainly include: Lee et al. [12] proposed the TRAOD algorithm based on classification and detection framework; Knorr et al. [13–15] proposed an algorithm based on extraction global features of trajectory; Li et al. [16–18] proposed a method based on the classifier, and so on. Outlier detection methods above are all trajectory outlier detection methods of line segment with certain shape. However, express logistics abnormalities include distribution time delay, path error and destination error, and so on. At the same time, the trajectory of express logistics is directed line segment that consists of a series of distribution centers in accordance with time and distribution level. It has multi-attribute characteristics of time, direction, position coordinate, level of city, city point affiliation and so on compared with the traditional trajectory. Thus existing outlier detection methods [19] are unable to meet the demand of express logistics distribution.

Therefore, this paper takes the express logistics trajectory dataset with multi-attribute characteristics as the research object to find the anomalies of express logistics, and proposes an express logistics outlier detection method to apply to the express logistics.

2. Problem description

To illustrate what is express logistics outlier detection and solve the express logistics anomaly detection problem effectively, we firstly give a definition to the express logistics abnormality. This paper mainly studies the three anomalies in express delivery, including distribution time delay, distribution path error and destination error. The three anomalies are as shown in Fig. 1.

Detailed definitions as follows:

Definition 1. Distribution time delay. When a parcel is in the process of distribution, the normal time-consuming should be fixed in the same section under normal circumstances. But if the time-consuming in a distribution path sub-segment did sharp increase suddenly than the normal time-consuming, it means the distribution time delay problem occurs.

Definition 2. Distribution path error. When a parcel is in the process of distribution, there is an established distribution path of the parcel according to the re-planning. If the distribution path did not follow the established path, it means the path error problem may occur. There are two cases that can lead to this result: (1) As the amount of goods of a distribution center or a logistics center (A or D shown in Fig. 1) is too large, the system

* Corresponding author.
E-mail address: scm@inha.ac.kr (O.-k. Kwon).

http://dx.doi.org/10.1016/j.compind.2017.05.003
0166-3615/

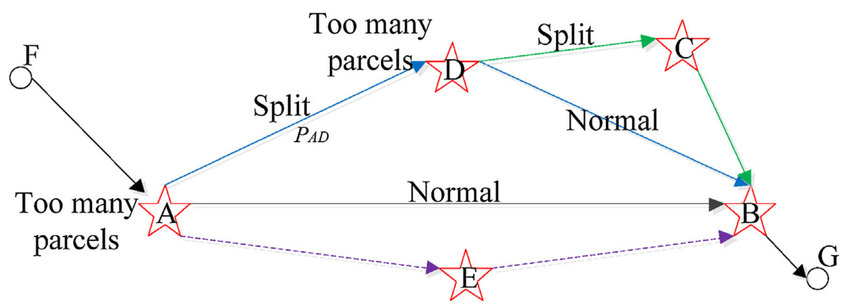

Fig. 1. The distribution schematic.

will plan a diversion path (A-E-B shown in Fig. 1) based on certain rules. (2) The distribution path has been wrong and the path error problem occurs. In the first case, detecting whether the actual split distribution scheme (A-D-B shown in Fig. 1) is consisted with the best split distribution scheme(The distance is the shortest and it will not lead to the backlog of parcels. A-E-B shown in Fig. 1). If the actual split distribution scheme is different with the best split distribution scheme, the path error problem occurs.

Definition 3. Destination error. When the distribution is finished, detecting whether the attributes (position coordinates, serial number, subordinate level, etc.) of the parcel actually reached center is consistent with the destinations. If the actual distribution destination was inconsistent with the destination on the order, it means the destination error problem occurs.

3. Proposed algorithm in detail

In order to find out the abnormal distribution problems quickly, we propose a novel express logistics outlier detection method to apply to the express logistics after depth analysis of the modern scale express logistics distribution mode.

3.1. Algorithm realization process

3.1.1. Distribution trajectory data multi-attribute representation

Extracting the distribution center's position coordinates, serial number, subordinate level, and the time of the parcel arrives at distribution center, the volume of the parcels in the distribution center, and the max transport capacity of the line which between two adjacent distribution centers, to represent the distribution path. A sub-segment of distribution path can be expressed as follows:

$$p < x_s, y_s, x_e, y_e, num_s, num_e, grad_s, grad_e, tim_s, time_e, gamt_{se}, ctr_{se}, flag >$$

where: x_s, y_s represent the abscissa and the ordinate of the starting distribution center on the path sub-segment respectively; x_e, y_e represent the abscissa and the ordinate of the ending distribution center on the path sub-segment respectively; num_s, num_e represent the serial number of the starting and ending distribution center on the path sub-segment respectively; $grad_s$, $grad_e$ represent the level of the starting and ending distribution center on the path sub-segment respectively; tim_s, $time_e$ represent the time of the parcel arrives at the starting and ending distribution center respectively; $gamt_{se}$ represents the total amount of parcels from the starting distribution center to the ending distribution center (It means that $gamt_{se}$ is the amount of all the parcels through the line of the path sub-segment); ctr_{se} represents the maxi transport capacity of the line that between the starting distribution center and the ending distribution center (It means that ctr_{se} is the biggest capacity of the starting distribution center delivery the parcel to the ending distribution center one day); *flag* symbolize whether the parcel has reached the termination (It means that the parcel will not be distributed).

Annotation: The starting point and the ending point above-mentioned do not refer to the source or the destination point of the order, but to the two adjacent distribution centers that form a complete distribution path. As shown in Fig. 1, it is assumed that a complete distribution path is A-D-C-B, and then it is made up of three sub-segments path (A-D, D-C, C-B). The two extreme points of these sub-segments path are the starting and ending distribution centers respectively. For example: the point D and point C are the starting and ending distribution centers of the sub-segment D-C respectively.

As shown in Fig. 1, it is assumed that a parcel should be distributed from A to B, so its complete distribution path is A-D-C-B. And a sub- segment of the distribution path p_{AD}(ie. A-D) can be expressed as follows:

$$p_{AD} < x_A, y_A, x_D, y_D, num_A, num_D, grad_A, grad_D, tim_A, tim_D, gamt_{AD}, ctr_{AD}, flag >$$

3.1.2. The process of outlier detection

3.1.2.1. Time outlier detection. Assuming that the normal distribution time-consuming from one distribution center to an adjacent center is *Time_t*(unit: day), if the actual distribution time-consuming t was longer than *Time_t*, then the distribution time-consuming is abnormal. For example, assuming that the parcel g should be sent to D from A, then a sub-segment of its distribution path can be expressed as p_{AD} that shown in Part 1, and the corresponding distribution time-consuming is t_{AD} (unit: day). If $t_{AD} > Time\_t$, we set the outlier degree of distribution time as$O_{tim} = \frac{t_{AD} - Time_t}{Time_t}$. Otherwise, set the outlier degree of distribution time as $O_{tim} = 0$. Where, t_{AD} and *Time_t* are calculated by the following equations:

$$t_{AD} = tim_D - tim_A \quad (1)$$

$$Time\_t = \overline{t_{AD}} + 3 \cdot \mu_{AD} \quad (2)$$

Where $\overline{t_{AD}}$ and μ_{AD} are defined as follows:

$$\overline{t_{AD}} = \frac{1}{n} \sum_{i=1}^{n} t_{ADi} \quad (3)$$

$$t_{ADi} = t_{Di} - t_{Ai} \quad (4)$$