

Research on Fire Risk Factors of Cotton in Railway Transportation

Yaxin Hou, Cong Ma, Yidong Li
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China

Abstract—Cotton plays a very important role in the economy of China and wellbeing of millions of people depends upon its good production and utilization in the country. Therefore, the railway transportation of cotton is becoming more and more important. In this paper, we analyze main factors causing cotton fire according to historical information. We first calculate the relative importance of all features for the risk of catching fire with the clustering method. Then, we calculate the risk rate of each main factor. The experiments show that the inside temperature of cabin, and the humidity and inside temperature of cotton pack are the top three factors that may cause fire.

Keywords—railway transportation; cotton; regression analysis; clustering analysis; risk valuation

I. INTRODUCTION

Cotton is the second largest crop after grain and is the main source of income of about 1 billion cotton farmers all over the country. At the same time, cotton is the main raw material for textile industry and is the necessities for everyone's life. Statistics show that there were 33 fire accidents during cotton railway transportation from 2005 to 2011 in China. The fire accidents of cotton railway transportation not only influenced the daily life of the people who live along the railway badly, but also disturbed the normal operation of the railway system and threaten the safety of railway transportation. It will be a certain guiding significance in preventing the occurrence of fire in cotton railway transportation that monitoring the actual transportation environment of cotton and evaluating the risk factors in the cotton railway transportation.

Some research has been done on the fire disaster early warning of cotton logistics warehouse. YIN Tao[1] carried out a series of simulated tests to study the firing mechanism during cotton transportation by railway and put forward conditions for safe cotton transportation by railway. JU Wen-hui[2] analyzed the physical and chemical characteristics of cotton and put forward a kind of mechanism based on Event and Fault Tree Analysis (EFTA) to explore the disaster of cotton logistics warehouses. ZANG Li [3] analyses the current situation of cotton logistics storage, pointing out that the warehouse facilities are too simple and the manual operation mode is backward. ZHANG Jing [4] establishes evaluation index system of cotton transport warehouse by using fuzzy comprehensive analysis method and analytic hierarchy process(AHP).

With the technique development of wireless sensor network(WSN), lots of scholars carried out study on hazardous

materials transportation monitoring with WSN. CAI Liming[5] use WSN, radio frequency identification (RFID), global positioning system (GPS), global system for mobile communication (GSM), geographic information system (GIS) and other modern information technologies to build an intelligent hazmat transportation system, which consists of parameters collection, information management, intelligent alerting and rescue dispatch. Hui Fei[6] designed a hazardous materials transportation monitoring system using WSN, this system monitored ambient situation of the vehicles and the state of the goods and determined the safety status of the transportation.

However, less research has been done on cotton fire safety problems and the disaster mechanism during railway transportation with WSN.

In this paper, WSN are used to collect and monitor the risk factor data in real time. The inside temperature and humidity of cabin are regarded as external risk factors, and temperature and humidity of cotton bags are regarded as internal risk factors. K-means and regression algorithm are used to evaluate correlation between the fire accident and all risk factors. At the same time, some indirect factors are analyzed, and the conclusion is drawn.

II. RELATED WORK

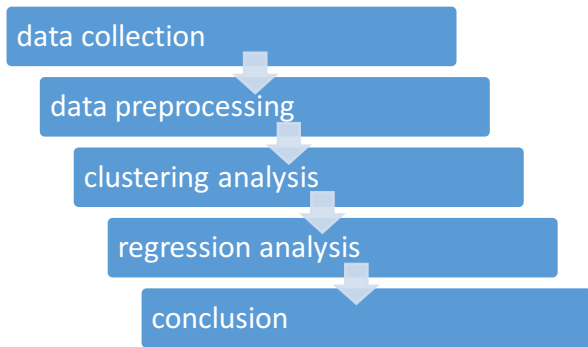
Cotton's major components are cellulose fibrin, protein and fat, etc. These component are all organic macromolecule compound. Cotton's morphological structure is loose and porous and it has large contact area with air. So cotton itself is inflammable, hypergolic and smouldered. When we talk about the fire accident during the cotton transportation, it is influenced by a lot of factors. Inside temperature and humidity of cabin and inside temperature and humidity of cotton bag are four key factors. We set hygographs and thermographs at several place inside the cabin and inside the cotton bags. These hygographs and thermographs collect humidity and temperature data in real time and then transmit these data to the ground system. The ground system establishes an analyzing model and take data it received as target input value and then analyze these data.

The occurrence of fire is closely contracted with weather condition, temperature and humidity inside the vehicle and cotton bag. Conduct in-depth and detailed monitoring, research and analysis are needed. Wireless sensor network in the compartment is used in the cotton railway transportation process to get external temperature and humidity acquisition,

collecting temperature and humidity in the bag of cotton, such data is used as input value. Data classification analysis and classification analysis model are conducted combined with the weather, temperature and humidity and other external information in the process of transportation along, risk is used as the key variable to do data classification analysis and establish a classification analysis model, and use regression analysis method to research the contribution of each characteristic variable to the possibility of the occurring of fire, or we can call it Relative importance.

In the traditional statistics, when there is no correlation or weak correlation between independent variable, the relative importance of the independent variables can be represented by some simple index, for example, square of the standard regression coefficient, square of partial correlation coefficient, square of semi partial correlation coefficient and so on. But obviously, the risk factors during the cotton transportation has correlation with each other. For example, there is correlation between inside temperature and humidity of cabin, as well as the inside temperature and humidity of cabin and cotton bag. So we cannot use the traditional statistical indicators to show the relative weight of the probability of each feature for the occurrence of the fire in the process of cotton transport. This paper firstly use K-means clustering analysis method to analysis multiple features, after removing fully independent or weakly dependent features, the relative importance of hidden danger of the fire in the process of cotton transportation is solved by the method of regression analysis.

For the strong correlation of the features, we establish the relationship model between the features and the probability of catching fire in the process of cotton transportation. We classify the fire hazards according to the diagram based on the output of the relational model. The probability of the risk of fire hidden danger in each kind of cotton is consistent with the trend of the variation of the features. The basic analysis algorithm model is shown in Figure 1.



III. DATA PREPROCESSING

In order to build the mathematical model of the hidden rules of the fire hazard in the process of cotton transportation, First of all we should determine the characteristics of the input variables. In this paper, the main factors that affect the fire risk in the process of cotton transport are mainly studied, including the internal temperature and the humidity of cabin, and the temperature and humidity of the cotton bag.

The data used in this paper is the real-time monitoring data obtained by the vehicle wireless sensor network in the process of cotton transportation. Characteristic variables are defined as follows:

- Temperature in cabin in the process of transportation
- Humidity in cabin in the process of transportation
- Temperature in cotton bag in the process of transportation
- Humidity in cotton bag in the process of transportation

The types of these four characteristics are digital.

Through experimental observation and analysis, fire hazard in cotton transportation is classified as shown in Table 1:

Table.1 Classification of distribution transformer load

Classification	Fire risk probability distribution
No risk	[0,0.25)
Low risk	[0.25,0.5)
Medium risk	[0.5,0.75)
High risk	[0.75,1]

IV. MATHEMATICAL MODELING

A. K-Means algorithm

For a given set of chaotic data, the first step is to carry out a cluster analysis, this step is used to determine the following classification of the training set. Here we use the K-means algorithm

Clustering analysis[7] is one of the important methods in data mining, its goal is to divide the data set into a number of clusters to make the similarity between the data points in the same cluster is as large as possible, and the similarity of the data points among different clusters is as small as possible. One of the most widely used and algorithm based on the idea of a relatively simple division of the K-means algorithm.

The basic principle of K-means algorithm is to select K objects from the N data objects as the initial cluster centers. According to the mean value of each cluster (center object), the distance between each object and the center object is calculated and we divide the corresponding objects according to the minimum distance; to recalculate the mean (central object) of each (the change) clustering; we loop second steps and third steps until each cluster does not change.

The advantages of K-Means clustering algorithm are mainly focused on: 1) Algorithm is fast and simple. 2) For large data sets, there is a high efficiency and scalability. 3) Time complexity is nearly linear, and it is suitable for mining large scale data sets. The time complexity of K-Means clustering algorithm is $O(nkt)$ where n represents the number of data objects, t represents the number of iterations of the algorithm and k represents the number of clusters.

Here is a brief description of the more critical steps in the K-means algorithm.

1) How to determine the value of K

Means K algorithm first select K initial centroid, where K is the user specified parameters, that is, the desired number of clusters. Here we use stability methods to determine the value of K . That is, a data set of 2 resampling produces 2 subsets of data, the same clustering algorithm is used to cluster the 2 data subsets and 2 clustering results with K clustering are generated. The distribution of the similarity of the 2 clustering results is calculated. The 2 clustering results have high similarity, it shows K cluster reflects the stability of the cluster structure and the similarity can be used to estimate the number of clusters. Several K was tested by the second method to find the appropriate K value.

2) The conditions for the algorithm to stop

In general, it stops when the objective function achieves the optimal or the maximum number of iterations. For different distance measures, the objective function is often different. When Euclidean distance is used, the objective function is to minimize the square sum of the distance between the object and the cluster centroid which is shown as follows:

$$\min \sum_{i=1}^K \sum_{x \in C_i} \text{dist}(c_i, x)^2$$

B. Regression analysis algorithm

In practical problems, it is easy to study the effect of a single quantity on a certain type of problem. But when a number of variables act on an unknown at the same time, often the problem is not so simple. At this point, we use the method of regression analysis.

Regression analysis is an analysis method based on a large number of observation data with the use of mathematical statistics method to establish the regression relation function between the dependent variable and the independent variable. In the regression analysis, when the causal relationship of the study involves the dependent variable and the independent variable, it is called a meta regression analysis, when the causal relationship studied concerns dependent variable and two or more than two independent variables, we call it multiple regression analysis[8]. In addition, in regression analysis, linear regression analysis and nonlinear regression analysis are differed based on whether the function expression of the causal relation is linear or nonlinear.

Set Y as the dependent variable, $X_1, X_2 \cdots X_k$ as independent variable, and the relationship between the independent variable and the dependent variable is linear. The multiple linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

Among them $\beta_0, \beta_1 \cdots \beta_k$ are unknown parameter, ϵ is a random variable which is called the error term.

We carry out N sub independent observations on Y and $X_1, X_2 \cdots X_k$ and get N group samples. Through the calculation of equations composed of the N group of samples, we get model parameters β .

The advantages of the regression analysis method are: 1) When using regression analysis method to analyze the multi factor model, the problem can be more convenient. 2) When we use regression model, the only result can be calculated by the standard statistical method if we use the same model and data. But in the form of graphs and tables, interpretation of the relationship between data often vary, the fitting curve draw by different analysts is probably not the same. 3) Regression analysis can accurately measure the degree of correlation between each factor and the degree of regression fitting and improve the effect of prediction equation. One element regression analysis method is suitable for a more influencing factor while multiple regression analysis is applied to multi variable problem.

Theoretically, the relative importance of the factors influencing the fire hazard in the process of cotton transportation can be considered as the isomorphism problem of it. Isomorphism is explained as follows: 1) Classification label can be seen as a model of the given conditions of cotton fire risk rating index. 2) The training set is regarded as a group of fire hazards affecting a given set of factors and the class number. 3) The function can be regarded as the quantitative relationship between the various factors and the number of fire hazards.

V. EXPERIMENTAL ANALYSIS

In this paper, we use the data obtained from the wireless sensor network based on wireless network in May 2010 and August 2010 to analysis the influence factors of fire hazard in the process of cotton transportation and classify the risk index according to analysis with the using of Matlab. Observed data is up to 4477 records. The input characteristic variables are shown in Table 2:

Tab.2 Characteristic variables of experimental analysis

NO.	Characteristic variable
1	Temperature in cabin
2	Humidity in cabin
3	Temperature in cotton bag
4	Humidity in cotton bag

A. Comparison of the contribution of characteristic variables to fire risk

According to the method proposed in this paper, we use cluster analysis and regression analysis to calculate the relative importance of characteristic variables on the occurrence of fire hazard in cotton transportation, the results are shown in Figure 2:

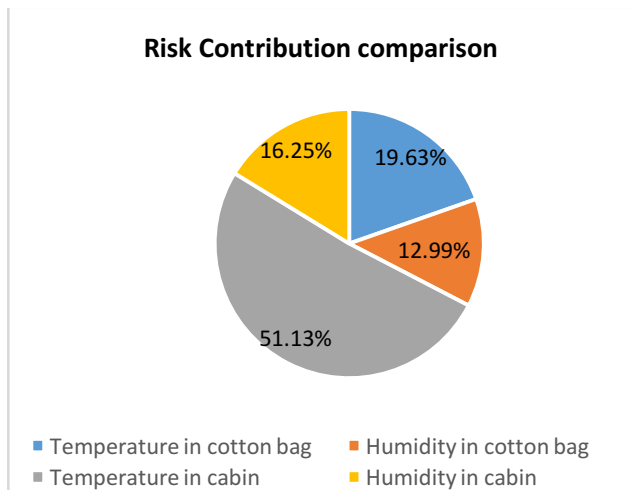


Fig.2 Risk contribution comparison of feature variables

It is not hard to find that the biggest impact on the fire is the temperature in the cabin, next to which is temperature in cotton bag, humidity in the cabin and humidity in cotton bag in the comparison. It is not difficult to understand, cotton is exposed to the outside world for a long time before it is loaded from the ground to the train. Then the external temperature and humidity will directly affect the temperature and humidity of the cotton bag. And in the process of transportation, due to the thermal conductivity of the train itself, outside temperature can more or less affect the cotton, it can be seen that the influence of external temperature is not negligible. In addition, in the case of the packing and transportation, the humidity in the compartment will directly affect the moisture of the cotton bag. For wet cotton, air circulation is bound to be poor, due to the continuous breathing of cotton itself, coupled with the appropriate temperature, it is easy to cause the spontaneous combustion of cotton. Then the risk contribution is clear at a glance based on regression analysis method.

B. Analysis of the relationship among weather, location, speed, moisture regain and characteristic quantity.

According to the monitoring data received from the sensor, when the goods are transported in summer, at this time the weather is hot, the direct embodiment is that the outside temperature is high. Cotton will absorb more heat in the process of loading and unloading, and cotton bag itself also can become wet because of the influence of the surrounding environment factor. So the weather is also an indirect factor that affects the risk of fire.

The temperature and humidity in cabin during the transportation process is shown in Figure 3:

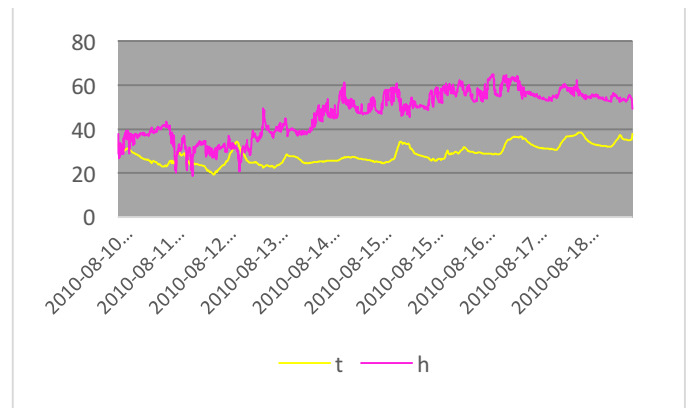


Fig.3 Changes of temperature and humidity along the line

We can see that the temperature is high, the same, the humidity is continuing to go high, under this condition, cotton bag is quite easy to catch fire.

The location of the selection will have an indirect impact on fire risk. In the same season, due to the different latitude and longitude, the daily average temperature, maximum temperature and minimum temperature will vary, this leads to the obvious difference of the external temperature. So it will determine the risk of the fire. For speed, the effect of vehicle speed on the fire risk of cotton can't be ignored, with the high speed of the train, the violent friction between the rail and the rail is bound to generate a huge heat, and part of it will affect the temperature in the car because of the transmission.

Moisture will cause some impact on cotton internal humidity, about seven percent of the cotton moisture regain makes internal humidity increasing, in the airtight compartment, there will be high risk of fire with the respiration of the cotton, as is shown in Figure 4:

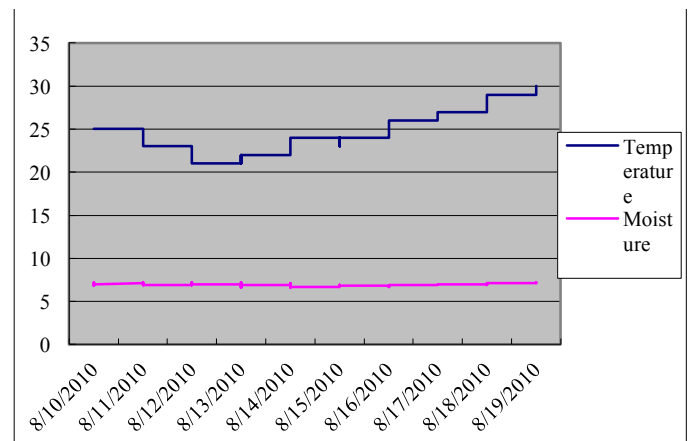


Fig.4 Moisture and temperature chart in summer in the process of transportation

VI. CONCLUSION

From the result of experimental analysis, we can reach the following conclusion:

1) The inside temperature of cabin is the top factor that affects the risk of fire, and its portion is over 50%. In this case,

it is best for the transport of cotton to avoid the high temperature season or to be undertaken in the night.

2) The temperature in cotton bag has great influence on the fire risk of cotton. It is necessary to make a reasonable preservation of cotton before transportation.

3) The effect of the inside humidity of cabin and the humidity of cotton bags on the fire risk is little. And the influence of the humidity of cotton bags is less. Therefore, dehydration treatment of cotton should be done before the transport if possible in order to reduce the risk of spontaneous combustion of cotton due to humidity.

ACKNOWLEDGEMENT

This work is supported by National Science Foundation of China Grant #61672088, Fundamental Research Funds for the Central Universities #2016JBM022.

REFERENCES

[1] Yin Tao, "Study on Simulation Tests of Safe Cotton Transportation by Railway", *Journal of Railway Transport and Economy*, Vol. 38, No 4, pp 89-93, Apr. 2016.

[2] JU Wen-hui, "Study on Fire Risk and Disaster Reducing Factors of Cotton Logistics Warehouse Based on Event and Fault Tree Analysis", *Procedia Engineering*, pp. 418 – 426, 2016.

[3] ZANG Li, "The Study on the Domestic Cotton Logistics", *Journal of Chinese Mark Logistics Management* vol.822, pp. 19, 2015.

[4] ZHANG Jing, "Fire Safety Assessment on National Cotton Logistics Warehouse", *Journal of Fire Protection Technique and Production Information*, pp. 81-84, Aug. 2012.

[5] Liming Cai, Wu Xia, Peng Li, Long Zhang, Jing Liu, "An intelligent transportation system for hazardous materials based on the Internet of Things", *International Conference on Information Technology and Management Innovation (ICITMI)* pp 611-616, 2015.

[6] Hui, Fei, Zhao Xiang-mo, Shi Xin, Yang Lan, "Wireless Sensor Networks for Hazardous Materials Transportation Monitoring: Framework and Implementation", *International Information Institute (Tokyo). Information* Vol. 15, No. 6, pp. 2623-2629, Jun 2012.

[7] Jiawei Han, Kamber Micheline, Jian Pei, *Data Mining: Concepts and Techniques*, third edition[M], China Machine Press, 2012.

[8] Zhang Hengxi, Guo Jilian, Zhu Jiayuan, Yu Jianfei, *Small sample multivariate data analysis method and its application*[M], Northwestern Polytechnical University press, 2002.