

Study on Dangerous Signal Mining of Dangerous Goods Transport Vehicles

Shengpu Li and Xiaohui Wang

*Collage of Computer Science and Technology, Pingdingshan University,
Pingdingshan, Henan, China
Lsp1519@163.com, wxh197@163.com*

Abstract

Dangerous signal mining method of dangerous goods transport vehicles and how to improve the accuracy of the mining are studied in the paper. There are many dangerous conditions such as high temperature, high pressure and other dangerous condition caused by the influence of the external factors such as friction in the transportation process of dangerous goods. If people can discover these dangerous condition signals as soon as possible, they can deal with them as soon as possible, which can effectively reduce the transport accidents and loss. The dangerous condition signals change as the interference of different nonlinear random mutation is present. The traditional signal mining methods have not a good solution to properly capture abnormal signal nonlinear changes, which causing flammable signal detection in the transport of dangerous goods inaccurate. The paper proposes a dangerous signal mining model based on the characteristics association mining algorithm for dangerous goods transport vehicles. First, the paper studies Cluster analysis theory and present the model for dangerous signal mining for dangerous goods transport vehicles. Second, the paper present wavelet transform model for the data extraction from transport vehicle status signal feature, and present a new mining model based the characteristics association data mining algorithm for dangerous signal mining, which can improve the accuracy of the mining. Finally, the experiment was done and shown that the new mining model for dangerous goods transport vehicles dangerous signal mining, can greatly improve the accuracy of the excavation.

Keywords: *Dangerous goods, Transport vehicles, Dangerous signal mining, Data mining*

1. Introduction

With the rapid development of logistics industry, different industries put forward higher requirements of logistics transportation [1]. Transport of dangerous goods, has always been the hot spot problem of the logistics transportation industry [2]. In the course of dangerous goods transport, there are maybe present lots of dangerous conditions, for example, if temperatures arise and burning occurs, there will be a wide range of explosion, which brings to public security a great threat [3]. There are many dangerous conditions such as high temperature, high pressure and other dangerous condition caused by the influence of the external factors such as friction in the transportation process of dangerous goods. If people can find these dangerous condition signals as soon as possible, they can deal with them as soon as possible, which can effectively reduce the transport accidents and loss. So, how to effectively mine the dangerous signal of dangerous goods vehicles transport and improve its accuracy has become a core problem need to study in the logistics transportation field, which has attracted the extensive attention of many experts [4].

At present, the dangerous signal mining method of dangerous goods vehicles transport mainly includes the signal Mining method based on k-means clustering algorithm, the signal Mining method based on the ant colony algorithm and the signal Mining method based on relational clustering algorithm [5-7]. Among above, the signal mining method based on relational clustering algorithm is the most commonly used [8].

In the process of vehicles transportation, dangerous goods are affected by lots of external factors, such as abnormal changes in temperature, the external turbulence friction, *etc.*, in dangerous conditions, with the changes of outside interference, nonlinear random mutation occurred to the high risk signal such as high temperature and high pressure. The traditional signal mining methods have not a good solution to properly capture abnormal signal nonlinear changes, which causing flammable signal detection in the transport of dangerous goods inaccurate.

In order to avoid the disadvantages of the traditional algorithm, this paper proposes a mining algorithm based on the characteristics association mining algorithm used in the dangerous goods transport vehicles high-risk signal mining. Using the method of wavelet transform, dangerous goods transport vehicle status signal feature extracted for mining high-risk signal in the next step. Finally, high risk signal mining succeeds by calculating correlation between the signal characteristics of dangerous goods transport vehicles status. The algorithm presented in this paper for dangerous goods transport vehicle signal mining, high-risk signal can be excavated from a large number of transport vehicle status signal, the actual state of dangerous goods transport vehicles is obtained, which can ensure the transport vehicles to make the right decisions and reduce the occurrence of dangerous accidents. The dangerous signal mining method of dangerous goods vehicles transport has a very huge space to research, it attaches more attention by a growing number of experts and scholars [9-10].

2. The Dangerous Signal Mining Model for Dangerous Goods Transport Vehicles

In this section, we first briefly introduce the concept of Cluster analysis. Then, we research k-means algorithm used in Cluster analysis. Next, we present the model for dangerous signal mining for dangerous goods transport vehicles.

2.1. Cluster Analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results [11].

Typical cluster models include:

- Connectivity models: for example hierarchical clustering builds models based on distance connectivity.
- Centroid models: for example the k-means algorithm represents each cluster by a single mean vector.

- Distribution models: clusters are modeled using statistical distributions, such as multivariate normal distributions used by the Expectation-maximization algorithm.
- Density models: for example DBSCAN and OPTICS defines clusters as connected dense regions in the data space.
- Subspace models: in Biclustering (also known as Co-clustering or two-mode-clustering), clusters are modeled with both cluster members and relevant attributes.
- Group models: some algorithms do not provide a refined model for their results and just provide the grouping information.
- Graph-based models: a clique, *i.e.*, a subset of nodes in a graph such that every two nodes in the subset are connected by an edge can be considered as a prototypical form of cluster. Relaxations of the complete connectivity requirement (a fraction of the edges can be missing) are known as quasi-cliques.

2.2. Centroid-based Clustering

In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k , k -means clustering gives a formal definition as an optimization problem. The optimization problem itself is known to be NP-hard, and thus the common approach is to search only for approximate solutions. A particularly well known approximative method is Lloyd's algorithm [12], often actually referred to as "k-means algorithm".

"k-means algorithm": Given a set of observations (x_1, x_2, \dots, x_n) , where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k sets ($k \leq n$) $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS)[13]:

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (1)$$

where μ_i is the mean of points in S_i .

Given an initial set of k means $m_1(1), \dots, m_k(1)$ (see below), the algorithm proceeds by alternating between two steps:[14]

- Assignment step: Assign each observation to the cluster whose mean yields the least within-cluster sum of squares (WCSS). Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean. (Mathematically, this means partitioning the observations according to the Voronoi diagram generated by the means).

$$S_i^{(t)} = \left\{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \quad \forall j, 1 \leq j \leq k \right\} \quad (2)$$

- where each x_p is assigned to exactly one $S_i^{(t)}$, even if it could be assigned to two or more of them.
- Update step: Calculate the new means to be the centroids of the observations in the new clusters.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (3)$$

Since the arithmetic mean is a least-squares estimator, this also minimizes the within-cluster sum of squares (WCSS) objective.

The algorithm has converged when the assignments no longer change. Since both steps optimize the WCSS objective, and there only exists a finite number of such partitioning, the algorithm must converge to a (local) optimum.

2.3. The Model of High-risk Signal Detection

As the transportation cause static electricity and shaking and then may well resulting explosion. Some special equipment installed in the dangerous goods transport vehicles can collect lots of real-time signals. Furthermore, the state signals within the transport of dangerous goods vehicles are effectively processed and mined to automatically alarm in time, if there is high-risk signal. The principle as follows:

Dangerous goods transport vehicle status signal set can be expressed in mathematical formulas $V = \{v_1, v_2, \dots, v_q\}$, v_l is the l th status signal collected in the process of the transport of dangerous goods. The attributes of all the element in the dangerous goods transport vehicle status signal set is $K = \{k_1, k_2, \dots, k_p\}$, and then $v_l[k_m]$ represents the value of attribute k_m of the l th status signal. According to the associated cluster theory, the dangerous goods vehicles state signals effectively classified by its properties. If the state parameter of the dangerous goods vehicles in the process of transportation is $e, q, p, e = 1$, clustering centers in the procedure of clustering are $B_{(d)} = (b_1, b_2, \dots, b_e)$, state parameter of the dangerous goods vehicles can be effectively updated by using the following equation.

$$y_{kl} = \frac{1}{\sum_{m=1}^e \left[\frac{f_{kl}}{f_{km}} \right]^{\frac{2}{p-1}}} \quad f_{kl} \neq 1$$

$$y_{kl} = 0 \quad e_{kl} = 0, l \neq m$$

$$y_{kl} = 1 \quad e_{kl} = 0 \quad (4)$$

The mean value of all dangerous goods transport vehicle status signal can be calculated by equation (5).

$$b_l = \frac{\sum_{k=1}^q y_{kl}}{\sum_{k=1}^q y_{lm}} \quad (5)$$

Then compare b_d with $b_{(d+1)}$.if the data can satisfy the conditions in the equation (6), the initial clustering of the transport of dangerous goods vehicles state signal is completed, otherwise, it need to continue to perform cluster analysis.

$$|b_d - b_{(d+1)}| \leq \varphi \quad (6)$$

If the high risk signal set of all the dangerous goods vehicles is Z , the amount of all status signals is q , the number of corresponding properties is p , attribute set of the dangerous goods vehicles status signals is $D = \{d_1, d_2, \dots, d_r\}$, attribute membership values of the dangerous goods vehicles status signals can be calculated according to the equation (7).

$$H(D) = \sum_{l=1}^q \prod \omega(d_k, z_l) \quad (7)$$

In the equation (4), q is the amount of all status signals, ω is membership of corresponding attribute, $\omega(d_k, z_l)$ represents membership of attribute d_k of the l th status

signal. The correlation between dangerous goods vehicles status signals can be calculated according to the equation (8).

$$y_{kl} = \frac{\exp\left(\frac{-\beta}{\kappa}\right)}{\sum_{k=1}^P \exp\left(\frac{-\beta}{\kappa}\right)} \quad (8)$$

If δ is correlation threshold, the equation (9) can complete high-risk signal mining of the dangerous goods vehicles.

$$\begin{cases} e_k > \delta, & \text{the signal is dangerous signal} \\ e_k \leq \delta, & \text{the signal is not dangerous signal} \end{cases} \quad (9)$$

The method based on the above elaboration can calculate membership of the dangerous goods vehicles status signals, and complete high-risk signal mining of the dangerous goods vehicles. But, in the process of vehicles transportation, dangerous goods are affected by lots of external factors, such as abnormal changes in temperature, the external turbulence friction, etc. In dangerous conditions, with the changes of outside interference, nonlinear random mutation occurred to the high risk signal such as high temperature and high pressure. The traditional signal mining methods have not a good solution to properly capture abnormal signal nonlinear changes, which causing flammable signal detection in the transport of dangerous goods inaccurate.

3. Optimization of the Dangerous Signal Mining Model

The traditional signal mining method based on relational clustering algorithm have no good solution for signal's abnormal nonlinear mutation, which lead to the dangerous signal mining in the dangerous goods transport is not accurate. So, this section proposes a mining algorithm based on the characteristics association clustering mining algorithm which can be used in the dangerous goods transport vehicles dangerous signal mining.

In this section, we first present wavelet transform model for the data extraction from transport vehicle status signal feature. We next present a new mining model based the characteristics association data mining algorithm for high-risk signal mining.

3.1 The Data Extraction from Transport Vehicle Status Signal Feature

The wavelet transform can provide us with the frequency of the signals and the time associated to those frequencies, making it very convenient for its application in numerous fields. For instance, signal processing of accelerations for gait analysis [15], for fault detection [16], for design of low power pacemakers and also in ultra-wideband (UWB) wireless communications [17].

A function $\varphi \in L^2(R)$ is called an orthonormal wavelet if it can be used to define a Hilbert basis, that is a complete orthonormal system, for the Hilbert space $L^2(R)$ of square integrable functions [18].

The Hilbert basis is constructed as the family of functions $\{\varphi_{jk} : j, k \in Z\}$ by means of dyadic translations and dilations of φ ,

$$\varphi_{jk}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k) \quad (10)$$

for integers $j, k \in Z$.

This family is an orthonormal system if it is orthonormal under the standard inner product $\langle f, g \rangle = \int_{-\infty}^{\infty} \int (x) \overline{g(x)} dx$ on $L^2(R)$

$$\langle \varphi_{jk}, \varphi_{lm} \rangle = \delta_{jl} \delta_{km} \quad (11)$$

where δ_{jl} is the Kronecker delta.

Completeness is satisfied if every function $h \in L^2(R)$ may be expanded in the basis as

$$h(x) = \sum_{j,k=-\infty}^{\infty} c_{jk} \varphi_{jk}(x) \quad (12)$$

with convergence of the series understood to be convergence in norm. Such a representation of a function f is known as a wavelet series. This implies that an orthonormal wavelet is self-dual.

Using wavelet transform, dangerous goods transport vehicle status signal feature extracted for mining high-risk signal in the next step, the detailed steps as follows.

$h(v)(h(v) \in N^2(T))$ represents the dangerous goods vehicles status signals, signal processed with wavelet transform according to the equation (13).

$$Y(c, d) \leq h(v), \zeta_{c,d}(v) \geq \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} h(v) \zeta_{c,d}\left(\frac{v-d}{c}\right) dv \quad (13)$$

In the equation, $c = 2^l$, $d = 2^l$, $(l, m) \in B^2$. According to wavelet frequency analysis theory, second wavelet transform is used in extraction of dangerous goods vehicles state signal, the state signal decomposition results are obtained as follows:

$$\begin{cases} e_{l+1}(q) = \sum_{m \in B} j(m-2q)e_l(m) \\ f_{l+1}(q) = \sum_{m \in B} i(m-2q)e_l(m) \end{cases} \quad (14)$$

The following equations can be used to normalize dangerous goods vehicles state signal characteristics.

$$\begin{aligned} z' &= \frac{z - \bar{z}}{u} \\ u &= \sqrt{\frac{\sum_{k=1}^Q (z_k - \bar{z})^2}{Q}} \\ \bar{z} &= \frac{\sum_{k=1}^Q z_k}{Q} \end{aligned} \quad (15)$$

In the above equations, Q is the number of vectors of dangerous goods vehicles state signal characteristics. z'_k is the characteristic vector normalization processing results.

According to the above method, status signal characteristic can be excavated from a large number of transport vehicle status signal, and then provide the basis for transport vehicle high-risk signal mining.

3.2 Characteristics Association Data Mining to enable better High-risk Signal Mining

In dangerous goods vehicle high-risk signal mining process, the high-risk signal mining need to calculate correlation coefficient between different transportation vehicles state signal characteristics. The detailed content as follows.

If P is dimension of transportation vehicles state signal characteristics Matrix, Q is the number of transportation vehicles state signal characteristics, $Z_k = (z_{k1}, z_{k2}, \dots, z_{kp})$ describes the location of signal sampling points. The equation (16) can calculate correlation between transportation vehicles state signal and the corresponding characteristic parameters.

$$\begin{aligned} x_{kl}(v+1) &= \zeta \cdot x_{kl}(v) + e_1 \cdot t_1 \cdot (r_{kl}(v) - z_{kl}(v)) + e_2 \cdot t_2 \cdot (r_{il}(v) - z_{kl}(v)) \\ z_{kl}(v+1) &= z_{kl}(v) + x_{kl}(v+1) \end{aligned} \quad (16)$$

The condition that Signal feature correlation needs to accord with can be obtained by equation (17).

$$\mu = \zeta - \frac{Q - X_{\max}}{\sqrt{e_1 - e_2}} \times iter_{\max} \quad (17)$$

The above content explains signal mining process. First, using wavelet transform, status signal characteristic can be excavated to provide the basis for transport vehicle high-risk signal mining. Then, calculate correlation coefficient between different transportation vehicles state signal characteristics to complete high risk signal mining.

4. Experiment Results Analysis

The effectiveness and performance of the proposed algorithms are verified by experiment and comparative analysis. The experiment need to gather a lot of dangerous goods transport vehicle status signal as sample data. The path of the transport of dangerous goods vehicles is in Figure 1.



Figure 1. The Path of the Transport of Dangerous Goods Vehicles

Set high risk signal spectrum of signal mining as Figure 2.

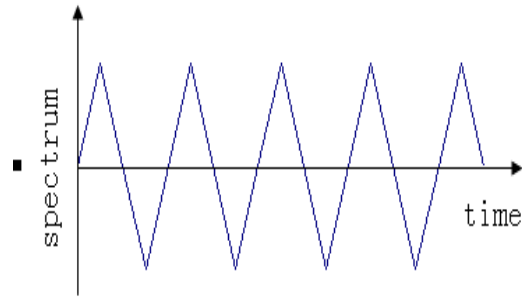


Figure 2. The Spectrum of Signal Sample

In the experiment, high-risk signal mining for dangerous goods vehicles were done respectively using the traditional algorithm and the proposed algorithm. The results can be described in Figure 3, 4.

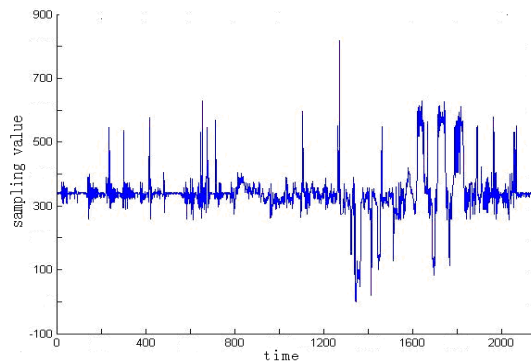


Figure 3. The High Risk Signals Mined by Traditional Algorithm

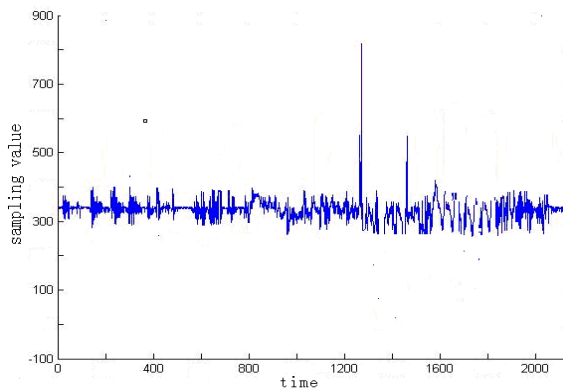


Figure 4. The High Risk Signals Mined by Proposed Algorithm

Compare the three pictures, the signal spectrum result of the algorithm proposed in this paper for dangerous goods transport vehicles high-risk signal mining is more close to the actual sample, which shows the superiority of the algorithm proposed in this paper.

In the experiment, the time cost of two algorithms is shown in Figure 5.

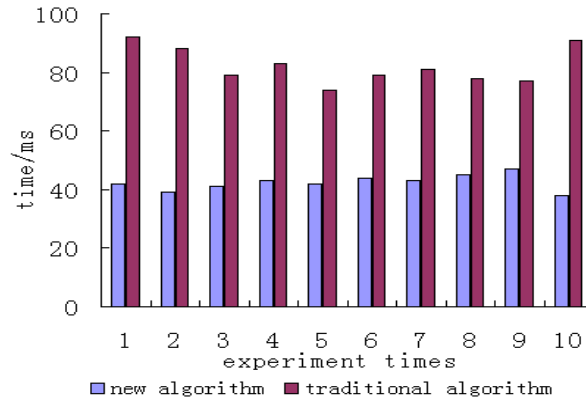


Figure 5. Time Comparison of Two Algorithms

Combining with different time's experiments, time comparison of two algorithms can be expressed in Table 1.

Table 1. Time Comparison of Two Algorithms

Times of experiments	Time cost (ms)	
	Ttraditional algorithm	Proposed algorithm
1	92	42
2	88	39
3	79	41
4	83	43
5	74	42
6	79	44
7	81	43
8	78	45
9	77	47
10	91	38

Table 1 shows that the time of the algorithm presented in this paper for dangerous goods transport vehicles high-risk signal mining, shorter than traditional algorithm. In the experiment, respectively using the traditional algorithm and the proposed algorithm for dangerous goods transport vehicles high-risk signal mining, mining accuracy can be expressed in Figure 6.

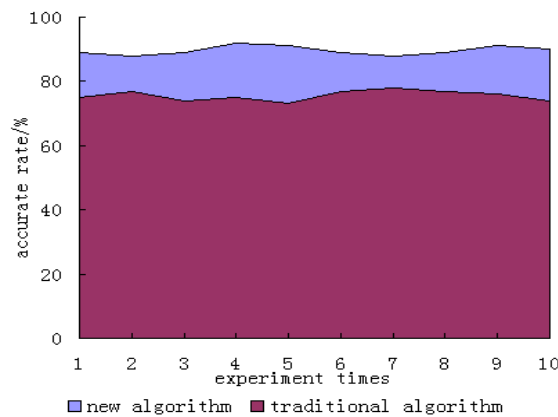


Figure 6. Mining Accuracy Comparisons of Two Algorithms

Mining accuracy comparisons of two algorithms can also be described in the Table 2.

Table 2. Mining Accuracy Comparisons of Two Algorithms

Times of experiments	Accuracy (%)	
	Traditional algorithm	Proposed algorithm
1	75	89
2	77	88
3	74	89
4	75	92
5	73	91
6	77	89
7	78	88
8	77	89
9	76	91
10	74	90

All the experiments and simulations show that the algorithm proposed in this paper for dangerous goods transport vehicles high-risk signal mining is faster and more accurate than traditional algorithm.

5. Conclusion

Dangerous goods in the process of transportation vehicles, by the environment temperature changes, the influence of the external turbulence external factors such as friction, dangerous conditions such as high temperature, and high pressure dangerous signal as the interference of different nonlinear random mutation is present. The traditional signal detection method for this kind of abnormal signal nonlinear changes have not a good solution, unable to properly capture, causing dangerous signal detection in the transport of dangerous goods is not allowed. The paper proposed the algorithm based on characteristics association clustering for dangerous goods transport vehicles high-risk signal mining. First, using wavelet transform, status signal characteristic can be excavated to provide the basis for transport vehicle high-risk signal mining. Then, calculate correlation coefficient between different transportation vehicles state signal characteristics to complete high risk signal mining. The experiment and simulation shows that applying the algorithm proposed in this paper in dangerous goods transport vehicles high-risk signal mining may help transport vehicles promptly discover the hidden safety trouble and adopt correct measures to reduce the accident loss and accident happening rate.

References

- [1]. H. Y. Liu, Q. Q. Yuan, *etc.*, "Survey of Parallel Algorithms for Data Mining", *Electronic science and technology*, vol. 19, no. 1, (2006), pp. 69-72, 77.
- [2]. S. L. Kou and Q. S. Cai, "Research on Feature-Selection in Chinese Text Classification", *computer simulation*, vol. 24, no. 3, (2007), pp. 289-291.
- [3]. W. Li and P. Y. Liu, "Research on anti- spam e- mail filtering algorithm based on Bayesian", *Computer Engineering and Applications*, vol. 43, no. 23, (2007), pp. 174-177.
- [4]. Y. Hu and W. M. Wang, "Energy Optimization of SIFT Algorithm for Wireless Sensor Networks", *Bulletin of Science and Technology*, vol. 9, (2013), pp. 98-101.
- [5]. D. L. Zhao, W. K. Liang, *etc.*, "On sub-band energy feature extraction method of vibration fault signals of hydroelectric sets", *Journal of Hydroelectric Engineering*, vol. 23, no. 6, (2004), pp. 116-119.
- [6]. B. Zhang and F. Chen, "An Immune System Multi-objective Optimization Algorithm for Analysis of Military Communication Network Planning", *National Defense Science & Technology*, vol. 3, (2012), pp. 27-31.
- [7]. K. F. Su, C. H. Ou and H. C. Jiau, "Localization with Mobile Anchor Point s in Wireless Sensor Networks", *IEEE Transactions on Vehicular Technology*, (2005) October 15-19, Kyoto, Japan.
- [8]. J. W. Tao, "A Distributed Web Log Mining System Based on Mobile Agent Technology", *Computer Simulation*, vol. 23, no. 10, (2006), pp. 109-113.

- [9]. W. Zhao and P. L. He, "Research on Data Preprocessing Technology in Web Log Mining", *Computer Applications*, vol. 23, no. 5, (2003), pp. 62-67.
- [10]. L. Chen and L. C. Jiao, "International study on Internet/Web data mining with the state of art and advances", *Journal of Xidian University*. Vol. 28, no. 1, (2011), pp. 2-3.
- [11]. "Cluster analysis", [OL][2014-05-22].http://en.wikipedia.org/wiki/Cluster_analysis
- [12]. S. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, vol. 28, (2012), pp. 129-137.
- [13]. K-means clustering [OL][2014-05-22].http://en.wikipedia.org/wiki/K-means_clustering.
- [14]. D. MacKay, "In Information Theory", *Inference and Learning Algorithms*, Edited H. S. Nalwa, Cambridge University Press, London, vol. 20, (2003), pp. 284-292.
- [15]. R. E. Smalley and J. Nanosci, "Novel method for stride length estimation with body area network accelerometers", *Computer simulation*, vol. 3, no. 75, (2009).
- [16]. J. Liu, "Shannon wavelet spectrum analysis on truncated vibration signals for machine incipient fault detection", *Measurement Science and Technology*, vol. 23, no. 5, (2012), pp. 1-11.
- [17]. A. N. Akansu, W. A. Serdijn and I. W. Selesnick, "Emerging applications of wavelets", *A review, Physical Communication*, Elsevier, vol. 3, (2010), pp. 1-18.
- [18]. "Wavelets transform", [OL][2014-05-25].http://en.wikipedia.org/wiki/Wavelet_transform.

