

Improved Latent Semantic Analysis and Support Vector Machine Algorithms for Public Opinion Early Warning of Sudden Security Events: Post-print

Authors: Tian Shihai, Lü Deli

Date: 2017-11-08T00:00:00+00:00

Abstract

Purpose: To address the limitations of existing early warning systems that primarily focus on enterprises and regulatory authorities while neglecting online public opinion, resulting in insufficient early warning capability, lack of transparency and sensitivity, and frequent sudden safety incidents that cannot be promptly handled, this study proposes a novel public opinion early warning model. **Method:** By mining public opinion information through meta-search technology, optimizing sentiment feature orientation weights with a baseline offset, and incorporating correction factors to improve the Latent Semantic Analysis and Support Vector Machine (LSA+SVM) algorithm, a classification-based public opinion early warning model is constructed. **Results:** Using multiple sets of sudden safety incidents as case studies, simulation experiments were conducted using Matlab. The results demonstrate that the proposed model is feasible and responsive, achieving an accuracy of 85.75% at a semantic dimension of 10. **Limitations:** This method is more effective for safety incidents that attract public attention and discussion. **Conclusion:** The improved algorithm is suitable for public opinion early warning and can provide rational recommendations for enterprises and regulatory authorities to take effective early warning measures promptly based on classification results.

Full Text

Preamble

ChinaXiv Collaborative Journal, Issue 2, 2017

Improved Latent Semantic Analysis and Support Vector Machine Algorithm for Early Warning of Public Opinion on Sudden Safety Incidents

Tian Shihai, Lyu Deli
(School of Management, Harbin University of Science and Technology, Harbin
150040, China)

Abstract

[Objective] Existing early warning systems primarily focus on enterprises and regulatory bodies while neglecting online public opinion, resulting in weak early warning capabilities, lack of transparency and sensitivity, and frequent sudden safety issues that cannot be addressed promptly. To address this gap, this study proposes a novel public opinion early warning model. **[Methods]** The model employs meta-search technology to mine public opinion data, optimizes the weight of sentiment feature terms by adding a baseline offset value, and introduces a correction factor to improve the Latent Semantic Analysis and Support Vector Machine (LSA+SVM) algorithm, thereby constructing a public opinion classification and early warning model. **[Results]** Using multiple sudden safety incidents as case studies, Matlab simulation experiments demonstrate the feasibility and rapid response of the proposed model, achieving an accuracy rate of 85.75% when the semantic dimension is set to 10. **[Limitations]** This method proves more effective for safety incidents that attract significant public attention and discussion. **[Conclusions]** The improved algorithm is suitable for public opinion early warning and can provide reasonable recommendations for enterprises and regulatory authorities to take timely and effective measures based on classification results.

Keywords: Latent Semantic Analysis; Support Vector Machine; Public Opinion Early Warning; Sentiment Orientation Analysis

Classification Number: G203

Introduction

Online public opinion is characterized by rapid dissemination, multiple channels, and wide coverage. Its propagation, diffusion, and fermentation regarding hot-button and sudden events play a crucial role in corporate decision-making and management. However, public opinion information is often chaotic, carries strong emotional coloring and noise, and may even threaten corporate survival and development. Therefore, how to properly utilize online public opinion and promptly classify and issue early warnings for enterprise-related information should be a key focus for both businesses and scholars.

Extensive research has been conducted on public opinion early warning both domestically and internationally. Wu et al. [?] constructed a network group behavior model through Agent modeling. Li et al. [?] applied artificial neural networks to predict product safety. Wang [?] analyzed the functions of public opinion intelligence and designed an architecture for public opinion intelligence support systems for emergency response. Papetti et al. [?] proposed

a multi-factor, multi-data-source public opinion early warning model, validating through multiple cases that the new model could maintain accuracy while reducing warning time and source data requirements. Dong et al. [?] analyzed role indicators and identified opinion leaders through subgroup mining to propose recommendations for public opinion mechanisms. Chen et al. [?] effectively predicted public opinion event development trends by establishing opinion interaction mechanisms.

In summary, most existing public opinion early warning research models and predicts warning measures at the macro level. However, the distribution of sentiment feature words is unbalanced relative to the entire model, and semantic dimensions are complex. Therefore, research from the perspective of optimizing semantic dimensions and speed is needed to develop more precise classification methods. This paper establishes a public opinion classification and early warning model by improving Latent Semantic Analysis (LSA) and Support Vector Machine (SVM) algorithms to enhance the accuracy of orientation prediction, improve classification efficiency, and strengthen public opinion situation awareness. This ensures enterprises can take proactive and effective measures before risks escalate further while addressing their own problems, innovating products, and adapting to market requirements based on public opinion feedback.

Online public opinion possesses four special characteristics beyond its dissemination advantages: (1) **Strong influence from national policies and regulations.** Safety standards in control, inspection, and management are updated in real-time with safety incidents and technological progress, representing a key factor for enterprise early warning. (2) **High suddenness with delayed timing and long fermentation periods.** Safety incidents are often triggered by unexpected events, spread rapidly, and attract substantial attention. They typically involve corporate production management systems and industry inspection regimes, leading to prolonged periods of propagation, fermentation, and 沉淀. (3) **Broad audience attention and strong coverage.** Since online content is closely related to daily life and individual safety, the public tends to devote sustained attention until issues are resolved. (4) **Significant impact and damage to enterprises.** Sudden safety incidents often prove fatal to businesses, as exemplified by the “Bawang carcinogen” controversy and the “melamine” incident, underscoring the need for greater resource allocation to crisis early warning.

Based on these characteristics, public opinion classification and early warning models must minimize the semantic dimension of public opinion orientation to enable rapid capture of public sentiment during the initial stages of safety incidents while maintaining excellent combination and classification capabilities to timely rate events and accurately determine their orientation. LSA can eliminate deviations caused by synonyms and polysemes in text analysis, obtain more accurate text vectors, and simplify text vectors to improve computational efficiency. SVM, as a classifier with excellent generalization ability, is widely applied and can be extended to other machine learning problems such as function

fitting [?]. Therefore, this study selects the LSA and SVM algorithm combination to meet classification and early warning requirements, with appropriate improvements to better align with the characteristics of public opinion subjects. The construction of the public opinion classification and early warning model 主要包括以下几个步骤: first, analyzing the public opinion classification and early warning process; second, determining and correcting sentiment feature word weights and improving the LSA+SVM algorithm; and finally, implementing the algorithm model.

2. Construction of Public Opinion Classification and Early Warning Model

2.1 Public Opinion Classification and Early Warning Process

The public opinion classification and early warning process consists of three main stages: information capture, orientation determination, and public opinion classification. Information capture utilizes meta-search technology and Nutch crawlers to perform simple noise reduction, cleaning, and segmentation on captured data to extract sentiment feature words. This study focuses on orientation determination and classification. The public opinion classification and early warning process is illustrated in [Figure 1: see original paper].

2.2 Determination and Improvement of Sentiment Feature Word Weights

Hot keywords related to safety incidents are selected, and the open-source Nutch crawler is used to mine public opinion corpora, obtaining relevant URL lists. Combining standards from the HowNet sentiment analysis lexicon, sentiment orientation feature words are extracted, primarily adjectives, adverbs, and nouns. Sentiment feature term texts are vectorized and stored in the following format: i_a represents the sentiment feature term for public opinion classification; T represents the time when the feature term was obtained; t represents the publication time of the text containing the feature term; r is a binary field indicating whether the URL was reposted; and W represents the importance weight of the webpage source for the feature term. When r is “yes,” W takes the weight value of the sentiment feature term.

Considering that the degree of information source influence and the semantic orientation consequences for related enterprises significantly affect the weight of relevant public opinion feature terms, let the importance degree of sentiment feature term i_a in the public opinion text vector be ik_{tfidf} [?]. The term ik_{tf} represents the frequency of sentiment feature term i_a , where ik_n denotes the occurrence count of feature term i_a . The calculation of ik_{tf} must incorporate the total number of feature terms i_N appearing throughout the text vector. The term idf represents the inverse document frequency of sentiment feature term i_a , i.e., words that appear infrequently in the entire article but have clearly defined

features. Therefore, the inverse of the count of feature term i_a is calculated, and its logarithm is taken:

$$ik_{tfidf} = ik_{tf} \times idf = \frac{ik_n}{i_N} \times \log\left(\frac{1}{idf} + 0.01\right)$$

However, in real-world public opinion texts, long sentences contain numerous adverbs and nouns with more pronounced sentiment orientation, causing feature term weight values to favor long texts and resulting in the log function approaching zero, which loses judgment capability. Additionally, safety incidents commonly use empirical coefficients containing semantics and modality to highlight important feature terms. The temporary release or improvement of national policies and regulations significantly impacts related industries. To solve the log function zero problem, its value is increased by 0.01. To address temporary regulatory policy releases or improvements, a baseline offset value *offset* [?] is added to the main weight value, yielding the public opinion classification sentiment feature term weight formula:

$$ik_{tfidf} = ik_{tf} \times idf = \frac{ik_n}{i_N} \times \log\left(\frac{1}{idf} + 0.01\right) \times \log(offset + 0.01)$$

The public opinion classification sentiment feature term weight formula is used to solve and store weight values for vectorized texts to facilitate subsequent processing of the vector space.

2.3 Vectorization and Classification of Sentiment Feature Words

Public opinion classification sentiment feature words are stored as single text vectors that do not belong to the same concept space, resulting in excessively high dimensionality that requires reduction for combination and classification. The basic process of the sentiment feature word classification method based on the improved LSA+SVM algorithm is shown in [Figure 2: see original paper].

Text preprocessing involves weight value calculation and improvement. LSA segments public opinion classification texts into different local feature spaces through singular value decomposition, avoiding interference from noise terms such as polysemy and synonymy, making the meaning expressed by public opinion sentiment feature words more explicit and perceptible. The vectorized public opinion feature word space vector is decomposed and stored in an $m \times n$ matrix format:

$$A = (a_{ij})_{m \times n}$$

Preliminary processing of the sentiment feature term matrix involves grouping synonyms with high semantic relevance into the same category. Consequently, the probability of different category feature words being synonymous is relatively

low. Thus, matrix A is decomposed into a combination of multiple matrices from different categories:

$$A = USV^T$$

where U and V are the left and right singular vector matrices of AA^T and $A^T A$, respectively, and S is the singular value matrix of A , with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq 0$. Singular Value Decomposition (SVD) is applied to compress the entire USV^T space, obtaining a rank- k matrix:

$$A_k = U_k S_k V_k^T$$

The specific decomposition process is illustrated in [Figure 3: see original paper]. The matrix S_k represents the decomposed basic singular value matrix that has been partitioned into multiple local spaces according to semantic relevance. Feature word similarity relationships are calculated through the inner product of row vectors in matrix A_k :

$$\text{Similarity} = (U_k S_k)(V_k^T) = U_k S_k S_k^T U_k^T$$

The resulting $S_k S_k^T$ represents the inner product relationship between rows i and j , reflecting the similarities and differences between two vectors, where k denotes the dimension after reduction. The new text vector is then sent to the SVM classification module for correlation-based classification.

2.4 Improvement of LSA+SVM Algorithm for Public Opinion Classification

Sudden safety incidents feature long fermentation periods, broad audiences, and significant impacts on corporate reputation, making it difficult for ordinary classifiers to determine sentiment orientation and risk levels. Therefore, a correction factor O_{ia} [?] must be added to the local feature matrix. The correction factor primarily uses the simultaneous appearance of sentiment word f_a and degree adverb g_a in the local feature vector as a benchmark, multiplying their weights to serve as the priority judgment standard for matrix severity:

$$O_{ia} = W_{f_a} \times W_{g_a}$$

In the formula $A_k = U_k S_k V_k^T$, S_k is rearranged using the correction factor O_{ia} and the local matrix obtained after singular value decomposition. Simultaneously, based on the weight values of safety feature words, baseline offset values are added to deviate from the original arrangement pattern and trend. Several singular value linear relationships are simulated as a regression jump curve with correlation, as shown in [Figure 4: see original paper].

When $O_{ia} > 0$, the correlation jump curve appears as in [FIGURE:4(a)], indicating strong positive meaning in the public opinion classification local matrix, typically appearing in comments or articles optimistic about the event or even beneficial to corporate brand image. When $O_{ia} = 0$, the curve shows no obvious jump, generally fluctuating near the horizontal axis, indicating neutral orientation in the local matrix. Such comments tend to narrate facts without clear criticism or support. When $O_{ia} < 0$, the correlation jump curve appears as in [FIGURE:4(b)], indicating strong negative meaning in the local matrix, showing intense reaction to the event and clear criticism of the enterprise. The closer O_{ia} is to 1 or -1, the more severe the sentiment orientation of the feature word.

2.5 Implementation Process of Improved LSA+SVM Algorithm

(1) Training Algorithm Implementation

A large number of public opinion classification training texts are selected to train the improved LSA+SVM algorithm, forming a standard public opinion classification and early warning parameter model to obtain three basic parameters: χ , α , β (penalty function coefficient, linear maximum margin, and kernel function coefficient). The specific training algorithm process [?] is as follows:

Input: Feature word vector set $\{x_1, x_2, \dots, x_n\}$, baseline offset value *offset*

Output: Classification parameter model $\{\chi, \alpha, \beta\}$

The process involves calculating TF-IDF weights with offset, performing SVD decomposition, and optimizing SVM parameters through cross-validation.

(2) Testing Algorithm Implementation

The testing algorithm combines the parameter model and SVM classifier to classify sentiment orientation of new feature term texts. First, texts are divided into two levels based on the correction factor's sign, then further classified into five grades according to weight: Special-grade (S), Heavy-grade (A), Medium-grade (B), Light-grade (C), and Attention-needed (D). Positive public opinion is incorporated into enterprise feedback and innovation information as Grade (P) [?].

Input: Test feature word set to be classified $\{x_1, x_2, \dots, x_n\}$, baseline offset value *offset*

Output: Classification result $\{S, A, B, C, D, P\}$

Through baseline offset and correction factor adjustments combined with extensive text training, the model can more accurately and efficiently assess crisis situations in real-time safety incidents and promptly feed results back to enterprises, achieving the early warning objective.

3. Implementation and Simulation of Public Opinion Classification Early Warning

To ensure accurate public opinion classification, this study examines the practical application of the LSA+SVM algorithm using three sudden safety incidents from different domains: the Mengniu aflatoxin incident (Case 1) representing food safety, the Baidu “Putian System” incident (Case 2) representing internet user security, and the Tianjin Binhai chemical plant leak incident (Case 3) representing production safety. First, “hot words + public opinion terms” are selected based on incident categories, and Python crawlers are designed using meta-search technology to mine these fields across search engine hot news, obtaining over 900 incident-related articles and comment URL lists, as shown in [Figure 5: see original paper].

After simple deduplication and noise reduction of crawled documents, the Institute of Computing Technology, Chinese Academy of Sciences’ ICTCLAS segmentation system and LSA+SVM algorithm are applied to obtain the basic classification parameter model through cross-validation. Algorithm implementation is based on the Windows 7 operating system using Matlab 2012b for simulation. The Radial Basis Function (RBF) kernel is selected for training, with optimal parameter and classification models determined through cross-validation.

Training yields a kernel function coefficient of approximately 0.431 and a penalty function coefficient of 0.424462, resulting in 324 effective feature vectors with negative orientation and 198 effective feature vectors with positive or neutral orientation. Analysis of the final classification jump curve reveals a tendency toward negative correlation, indicating negative public opinion orientation. Comparative experiments are conducted on the classification model under different semantic dimensions, using document orientation accuracy under different parameters as the performance metric:

$$\text{Accuracy} = \frac{PP + NN}{P + N}$$

where P represents the total number of positive documents selected, PP represents documents that were positive at selection ($O_{ia} > 0$) and remain positive after classification, and NN represents documents that remain negative after classification. Three different semantic dimension values are randomly selected: $k = 5, 10, 15$. The results are shown in [Figure 6: see original paper].

The results demonstrate optimal performance at $k = 10$, achieving 87.25% accuracy and efficiently reflecting text relevance characteristics. Too low a dimension leads to result bias, while too high a dimension causes semantic confusion and inaccurate grading [?]. The algorithm implementation results are shown in [Figure 7: see original paper].

The final effective documents and comments are classified according to correction factor values. Positive public opinion is classified as Grade P, while

negative public opinion requiring alerts is divided into five grades equidistantly across the $(-1, 0)$ interval: $S \in [-1, -0.8]$, $A \in [-0.8, -0.6]$, $B \in [-0.6, -0.4]$, $C \in [-0.4, -0.2]$, $D \in [-0.2, 0)$. Considering that the $(0, 0.1)$ interval, though positive, shows unclear orientation, it is assigned to Grade D (attention-needed) [?]. After removing duplicate and training documents, the positive-to-negative document ratios for the three sudden safety incidents are shown in .

TABLE:1 Positive and Negative Document Ratios for Three Sudden Safety Incidents

Sudden Safety Incident	Positive (%)	Negative (%)	Neutral (%)
Baidu “Putian System” Incident	12.3	78.6	9.1
Binhai Chemical Plant Leak Incident	8.7	85.4	5.9
Mengniu Aflatoxin Incident	15.2	72.8	12.0

The positive-negative document ratios provide preliminary insight into public opinion orientation for different sudden safety incidents. Negative and neutral documents are further analyzed to summarize document quantities across different grades, determining the urgency level of each incident and appropriate corporate responses. The grade classification and main public opinion terms are shown in .

TABLE:2 Grade Classification and Main Public Opinion Terms for Three Sudden Safety Incidents

Sudden Safety Incident	Main Public Opinion Terms	Quantity	Grade	Contains Baseline Offset
Baidu “Putian System”	Evil; scandal; collusion; false advertising; desperate struggle; medical ethics deficiency; bottomless greed; fraud; profit-driven murder; cancer; sin; dirty advertising tactics; immoral; indulgence; lack of supervision; inaction	156	S	Y

Sudden Safety Incident	Main Public Opinion Terms	Quantity	Grade	Contains Baseline Offset
Binhai Chemical Plant	Poisoning people; strong protest; life and property unprotected; pungent smell; nausea and dizziness; nowhere to petition; disregard for public safety; huge hidden danger; pollution; unable to eat or sleep; tragic; extremely dangerous; heartbreaking; learn lessons	203	S	Y
Mengniu Aflatoxin	No mercy needed; unprincipled; obviously unconvincing; serious threat to life safety; empty documents; repeated offenses; cost of mistakes too low; major defects; apology insufficient; strongest chemical carcinogen; fragile confidence	89	S	N

(Note: “Y” indicates “Yes,” “N” indicates “No.”)

As shown in , all three sudden safety incidents are classified as Grade S (Special-grade) public opinion, with the Binhai chemical plant explosion and leak incident being the most severe, requiring immediate and high-level attention from enterprises or regulatory authorities.

4. Conclusions and Recommendations

To address the current issues of narrow early warning scope, weak transparency, and untimely response for sudden safety incidents, this study incorporates external stakeholders into network public opinion early warning research. Based on real-time mining of trending events and keywords and considering the characteristics of public opinion classification, baseline offset values for weights are added to improve the LSA+SVM algorithm. The correction factor's positive or negative values enable public opinion orientation determination and early warning classification as follows:

1. When the correction factor is negative within $[-1, -0.4)$, it is classified as Grades S, A, or B based on weight ranking, indicating clear negative orientation with significant impact requiring immediate corporate intervention.
2. When the correction factor falls within $[-0.4, 0.1)$, it is classified as Grades C or D, representing neutral public opinion requiring monitoring and preparation.
3. When the correction factor falls within $[0.1, 1]$, it is classified as Grade P, representing positive public opinion that helps maintain a positive corporate image.

Using LibSVM and Matlab for simulation and accuracy calculation, the algorithm's effectiveness is validated. The results reflect the orientation of public opinion texts and provide accurate alert information for enterprises. The following recommendations are proposed:

Grade S (Special-grade): Attach high importance and respond immediately. Enterprises must dispatch professional PR teams, quickly identify the source of public opinion, and conduct product recalls and compensation to minimize damage to corporate image and demonstrate responsibility.

Grade A (Heavy-grade): Take measures to resolve the crisis. Heavy-grade public opinion requires timely intervention to prevent further diffusion and escalation to special-grade. Enterprises should allocate resources and integrate capabilities to avoid crisis deterioration without harming current interests.

Grade B (Medium-grade): Suppress further diffusion of public opinion information. Closely monitor crisis information status and public opinion direction, guide public opinion appropriately, and activate contingency plans to ensure favorable development.

Grade C (Light-grade): Eliminate interference and respond actively. Provide improvement suggestions to relevant departments, maintain continuous monitoring, and respond promptly to changes.

Grade D (Attention-needed): Conduct routine monitoring. Preliminarily judge public opinion categories, incorporate positive opinions into the enterprise innovation knowledge base, and file negative opinions as potential problem

sources for prevention.

Grade P (Information feedback): Use as feedback suggestions. Since most Grade P information lacks strong emotional coloring or is primarily positive, enterprises can reference this feedback to innovate products, strengthen management, and upgrade services, providing new ideas and opportunities for development.

References

- [1] Wu P, Yang S, Zhang J, et al. Agent-Based Modeling and Simulation of Evolution of Netizen Crowd Behavior in Unexpected Events Public Opinion [J]. *New Technology of Library and Information Service*, 2015(7/8): 65-72.
- [2] Li W, Miao D, Wang W. Two Level Hierarchical Combination Method for Text Classification [J]. *Expert Systems with Applications*, 2011, 38(3): 2030-2039.
- [3] Wang L. Research on Emergency Information Support Based on Network Public Opinion Analysis [J]. *Information Studies: Theory & Application*, 2015, 38(7): 72-75.
- [4] Papetti P, Costa C, Antonucci F, et al. A RFID Web-based Infotracking System for the Artisanal Italian Cheese Quality Trace Ability [J]. *Food Control*, 2012, 27(1): 234-241.
- [5] Dong K, Fu Y, Sun X, et al. Research on Early Warning Mechanism of Enterprise Public Opinion Based on Social Network Analysis—Taking Food Safety Network Public Opinion as an Example [J]. *E-Business Journal*, 2015, 23(8): 54-55, 57.
- [6] Chen F, Chen T. Research on Public Opinion Emergencies Evolution: Based on the Perspective of Opinion Leaders Guiding Role [J]. *Information and Documentation Services*, 2012, 36(2): 23-28.
- [7] Xuan Y, Zhu Q. Research on Tag Semantic Retrieval in Social Tagging System Based on LSA [J]. *Library and Information Service*, 2011, 55(4): 11-14.
- [8] Fan Y, Qin S. Optimizing Decision for Scene Classification Based on Latent Semantic Analysis [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2013, 25(2): 175-182.
- [9] Shang L, Tan Q. Emergency Classification Based on Support Vector Machine [J]. *Journal of Industrial Engineering and Engineering Management*, 2014, 28(1): 119-123.
- [10] Zhang J. A Chinese Keywords Extraction Approach Based on TFIDF and Word Correlation [J]. *Information Science*, 2012, 30(10): 1542-1555.
- [11] Zhang C. Research on Domain-Oriented Public Sentiment Analysis Technologies [D]. Changchun: Jilin University, 2011.

- [12] Gao H, Wang J. A Fuzzy Control Method with Online Self-turning Correction Factor and Its Application [J]. *Microcomputer Information*, 2006, 22(13): 83-84.
- [13] Goñi S M, Oddone S, Segura J A. Prediction of Foods Freezing and Thawing Times: Artificial Neural Networks and Genetic Algorithm Approach [J]. *Journal of Food Engineering*, 2011, 84(1): 164-178.
- [14] Tan G, Liu Z. A Local Latent Semantic Analysis Algorithm Based on Support Vector Machine [J]. *Computer Engineering and Science*, 2016, 38(1): 177-182.
- [15] Sengupta A S, Balaji M S, Krishnan B C. How Customers Cope with Service Failure? A Study Does Brand Reputation and Customer Satisfaction [J]. *Journal of Business Research*, 2015, 68(3): 655-674.
- [16] Zhu G, Qi J. Situation Evaluation of Online Public Opinion on Enterprise Crisis Event [J]. *Information Science*, 2015, 33(6): 48-53.
- [17] Ma N, Liu Y. Multi-Agent Modeling of Public Opinion Evolution Based on SuperNetwork Analysis [J]. *Journal of Systems & Management*, 2015, 24(6): 785-804.

Supporting Data

The supporting data is self-archived by the authors, E-mail: lvdelixx@126.com.

- [1] Tian S, Lyu D. pythonurl1.txt. Python crawler mining code.
- [2] Tian S, Lyu D. wo3url.csv. Mined URL list.
- [3] Tian S, Lyu D. lsasvm.csv. Segmented sentiment orientation feature word matrix.
- [4] Tian S, Lyu D. lsasvmdepart.rdf. Feature word vector matrix calculated by improved LSA+SVM algorithm.
- [5] Tian S, Lyu D. orien.xls. Document orientation accuracy list.

Author Contributions

Tian Shihai: Proposed research ideas, designed research framework, revised final manuscript.

Lyu Deli: Implemented research process, acquired data, conducted experiments, wrote manuscript.

Conflict of Interest Statement

All authors declare no conflict of interest.

Received Date: 2016-08-29

Revised Date: 2016-10-25

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.