Research on the Image Annotation Technology for Product Quality and Safety Inspection Data

Xiuli Ning, Ying Li*, Yingcheng Xu and Xiaowei Lu

Quality Management Branch, China National Institute of Standardization, Beijing 10001, China

*Email: 475554762@qq.com

Abstract. In recent years, the vicious events about quality and safety in China have continued to bring serious impacts on people's lives and property. The effective analysis and processing of product quality and safety inspection data will provide intellectual support for the overall improvement in product quality, and the effective control of prominent quality and safety problems in China. Aiming at the phenomenon that there is a large amount of image information in the quality detection data, this paper proposed an image annotation technology based on big data fusion, conducted weight fusion to image similarity and image user similarity, calculated the total similarity of images, and made denoising treatment. The experimental results showed that the method proposed in this study could annotate the quality test data well.

1. Introduction

Along with the development of computer technology, network technology and digital media technology, quality inspection data contains more and more image information. Nevertheless, it is very difficult for relevant departments to search for useful image information and make data analysis. In the light of this phenomenon, this research studied the image annotation technology of quality detection data.

1.1 Characteristics of quality test data

The characteristics of quality test data mainly include:

1) Extensive sources and huge volume: the quality inspection data includes national product quality supervision and random checking, 12315/12365 consumer complaints, WTO/TBT recall notifications, laboratory product testing, product injury and accident, and etc., so the data volume is huge.

2) Different features: quality test data includes text information, image information, audio information, video information, and etc., as well as the structured, semi-structured and unstructured data.

3) Large differences in data quality: the quality of quality test data is uneven. Some data is of high quality, while some data is missing.

### 1.2. Research status of image annotation

The scholars at home and abroad have carried out a large number of studies on image annotation technology. For example, Wang et al. [1] used two-dimensional multi-scale HMM to annotate images, and established Markov chains on multiple scales to express the relationship between multiple scales.

Sanjiv et al. [2] established a general two-layer model by using context information based on CRF, and annotated it by using the relationships between regions, regions and objects, as well as objects and objects in images. Monay et al. [3] combined the annotation keywords and regional features of images for training. Martinez et al. [4] used the local classifier SVM-KNN to automatically select unmarked samples and add them to the training set for image annotation by means of active learning. Wang et al. [5] retrieved a large number of images on the Internet based on the annotation keywords by determining an annotation keyword of the image to be annotated, and calculated the visual feature similarity between the retrieval results and images to be annotated, and then annotated the images. Blei et al. [6] proposed a model based GM-Mixture and GM-LDA for image annotation. He et al. [7] extracted the local features, regional features and global features from images respectively, and annotate images by using multi-scale CRF.

2. Theories and methods

For the collected quality and safety inspection data, the first step is to preprocess the data, including denoising, Chinese word segmentation, stopwords removal, data protocol and data loading processing; the second step is to analyze the image similarity, including attribute similarity calculation and text similarity calculation; the third step is to analyze the image user similarity; then to calculate the total similarity. Finally, noise removal is carried out. The image annotation flow chart based on multi-source big data fusion is shown as figure 1:
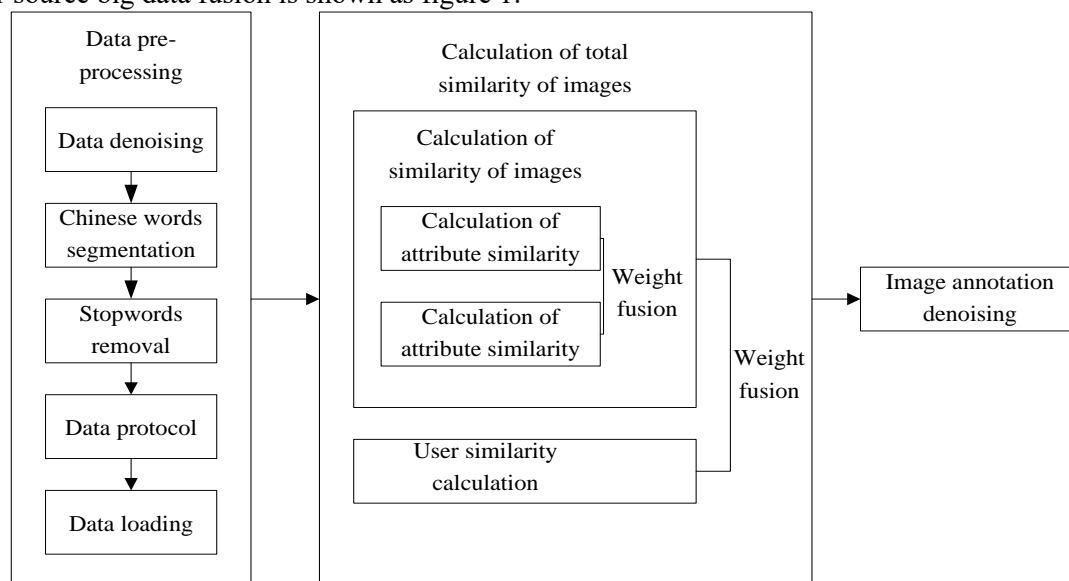
Figure 1. Image annotation flow chart based on multi-source big data fusion

2.1. Data pre-processing

The supervision and random checking data, risk monitoring data, network public opinion data, product quality damage data, notification and recall data, and other data related to product quality collected in this paper was used as the experimental data, in which the related data accompanied with images was chosen to obtain the release time, release location, image, and image-related information, as well as the location information, authentication information, social contact information and other user information. The quality test data was subject to denoising, Chinese word segmentation, stopwords removal, data protocol, and data loading in this paper.

2.1.1. Data denoising. There was a large amount of noise data in the collected quality detection data, so the data shall be denoised first. First of all, the quality inspection data collected in this paper contained a large amount of data without images. Since this paper studied image annotation technology, the

relevant data without images were removed. Secondly, the collected data contained a large number of special symbols, such as "#", "@", "￥", etc., which would affect the accuracy of data analysis; therefore, special symbols must be removed.

2.1.2. Chinese word segmentation. There was no word segmentation marker for the text in the collected data text, but the frequency of adjacent Chinese characters or terms appearing in the corpus could be used to judge whether adjacent Chinese characters or adjacent terms could be combined. In this paper, Chinese words segmentation was conducted with the word segmentation algorithm based on statistics, which could acquire experience information by training a large number of corpus subject to manual word segmentation, convert language knowledge into statistical information, and establish the probability model that could reflect the trust degree of adjacent characters or terms, so as to identify new words and segment sentences into words.

2.1.3. Stopwords removal. Stopwords refer to the words that increase the complexity of data analysis but fail to provide useful information, including auxiliary words, conjunctions, adverbs, etc., in the text. Therefore, these words should be removed before data analysis. For example, "next", "then", "of", "in a word" and so on.

2.1.4. Data protocol. There are two kinds of attribute information in a given data text, namely image information and user information. Image information includes three valid attributes: release time, release site and text content, while user information includes three valid attributes: location information, authentication information and social contact information.

2.1.5. Data loading. The image information and user information subject to de-noising, Chinese word segmentation, stopwords removal and data protocol were stored in the database respectively.

2.2. Image similarity analysis

Image information attributes include release time, release place, author information, text content, and etc., among which, text content is missing or irrelevant, and so on; therefore, the image information was divided into attribute information similarity and text information similarity, of which the image similarity could be obtained by weighting method, with the calculation formula shown as follow:

$$Sim_I(x, y) = \omega_1 Sim_p(x, y) + \omega_2 Sim_T(x, y) \tag{1}$$

Where, $Sim_I(x, y)$ represents the image similarity, $Sim_p(x, y)$ represents the attribute similarity, $Sim_T(x, y)$ represents the text similarity, ω_1 and ω_2 are the weights, and $\omega_1 + \omega_2 = 1$.

2.2.1. Attribute similarity calculation. The attribute information of an image includes the release time, release place and category, etc. In this study, the unit of time granularity is day. The format of place is defined as provincial, municipal and specific address; the categories include supervision and random checking, consumer complaint, recall notification, laboratory product test, product damage, and etc. A complete set of attributes is {time, place, category}, for example, {2018-3-21, Beijing, consumer complaints}. By constructing the network relationship between images and attributes, SimRank is used to analyze the network image and calculate the attribute similarity. When an image satisfied the following three conditions simultaneously, an association relationship with the attribute set could be established:

    1. Time: the time when the text is released should differ within 10 days from the time in the data set;

    2. Place: the place where the text is released belongs to the same province as the place in the data set;

    3. Category: the category information in the text is the same as that in the attribute set.

The construction of bipartite graph is the first step of similarity calculation, and an example is used to illustrate the construction. It is supposed there are four data in the text, as shown in table 1:

Table 1. Image data

No.	Time	Place	Category
1001	2018-5-3	Beijing	Risk monitoring
1002	2018-5-7	Shijiazhuang, Hebei	Supervision and random checking
1003	2018-5-12	Hebei	Supervision and random checking
1004	2018-5-15	Chifeng, Inner Mongolia	Recall notification

As can be seen from the table, the maximum time difference is 13, and there are 2 types of places and 3 types of categories, so there is a total of 13*2*3 attribute sets constructed. Among them, only the numbers 1002 and 1003 satisfy the three conditions for establishing association, so it is considered that the images numbered 1002 and 1003 can be associated with the attribute set.

In order to better analyze the compactness of images and attribute sets, the time and place should be given weights. The higher the weights, the closer relationship between images and attribute sets. The formula for calculating the weight is as follows:

$$W = \omega_1 W_T + \omega_2 W_P \tag{2}$$

Where, W_T represents the weight of time attribute, W_P represents the weight of place attribute, and $\omega_1 + \omega_2 = 1$.

2.2.2. Image text similarity calculation. In the process of image text similarity calculation, image text is a word vector composed of words after data preprocessed, and editing distance algorithm is used to calculate the image text similarity in this study. If the two texts are set to be T_1 and T_2, the similarity calculation formula is:

$$Sim_T(T_1, T_2) = 1 - \frac{Dis(T_1, T_2)}{\max(len(T_1), len(T_2))} \tag{3}$$

Where, $Sim_T(T_1, T_2)$ represents the similarity of T_1 and T_2, and $Dis(T_1, T_2)$ represents the editing distance of T_1 and T_2.

2.3. Image user similarity analysis

User similarity can be calculated by weighting users' location information, authentication information and social information. The formula is as follows:

$$Sim_U(x, y) = \omega_1 Sim_P(x, y) + \omega_2 Sim_A(x, y) + \omega_3 Sim_F(x, y) \tag{4}$$

Where, $Sim_U(x, y)$ is the similarity of users, $Sim_P(x, y)$ is the similarity of place information, $Sim_A(x, y)$ is the similarity of authentication information, $Sim_F(x, y)$ is the similarity of social information, and ω_1, ω_2 and ω_3 are the weights.

2.4. Analysis of total similarity of images

The total similarity of images is obtained by weighting image similarity and user similarity, with the calculation formula as follow:

$$Sim(x,y) = \omega_1 Sim_I(x,y) + \omega_2 Sim_U(x,y) \tag{5}$$

Where, $Sim(x,y)$ represents the total similarity of images, $Sim_I(x,y)$ represents the image similarity, $Sim_U(x,y)$ represents the user similarity, ω_1 and ω_2 are the weights, and $\omega_1 + \omega_2 = 1$.

2.5. Image annotation de-noising
The similarity of the images to be annotated with other images in the database was calculated according to the image similarity calculation method described in section 2.4; when the similarity was greater than the threshold set, the images in the database would be selected, and then an image set would be formed; based on the existing information of each image, the place information, time information, text information, and other annotation information could be obtained. However, there were also some cases of incorrect annotation resulting in a decline in accuracy. Therefore, it is necessary to de-noise the annotation information. TF-IDF (term frequency–inverse document frequency) was employed in this research to remove irrelevant annotation words, with the calculation formula shown as follows:

$$TF - IDF(w_i) = \frac{N_i}{N} I_i \tag{6}$$

Where, N_i is the frequency of occurrence of annotation word w_i in all annotation words, N is the sum of all annotation words, and I_i is the inverse document frequency of annotation word w_i in the corpus.

3. Experiment and result analysis

3.1. Selection of experimental data sets
In this paper, a total of 5,000 pieces of product quality related data, including product quality supervision and random checking, 12315/12365 consumer complaint, WTO/TBT recall notification, laboratory product testing, product injury and accident were collected, and the precision rate and recall rate was used to evaluate the annotation performance. The calculation formula of precision rate and recall rate is as follows:

$$P = \frac{C}{N} \times 100\% \tag{7}$$

$$R = \frac{C}{n} \times 100\% \tag{8}$$

Where, P is the precision rate, C is the number of correct tag words, N is the total number of candidate tag words; R represents the recall rate, and n is the number of correct annotated words.
For the analysis of similarity, the precision rate of similarity was used in this paper for determination. The higher the precision rate of similarity, the more credible the similar images. The precision rate of similarity is defined as follows:

$$Sim\_P = \frac{\sum_{i=1}^{N_r} N_{s1}}{\sum_{i=1}^{N_r} N_{s2}} \tag{9}$$

Where, Sim_P is the precision rate of similarity, N_r is the number of randomly selected images, N_{s1} is the number of images in a similar image set sharing the same content with an image, and N_{s2} is the number of images similar to an image.

3.2. Simulation analysis

As a matter of experience, the weights of image similarity, attribute similarity and image user similarity were assigned respectively; 0.4 and 0.6 were assigned to ω_1 and ω_2 in image similarity, 0.5 and 0.5 were assigned to ω_1 and ω_2 in attribute similarity, and 0.3, 0.3 and 0.4 were assigned to ω_1, ω_2 and ω_3 in image user similarity. Similarity thresholds were set to be 0.5, 0.6, 0.7, 0.8, 0.9, 0.92, 0.94, 0.96 and 0.98, respectively.

Table 2. Comparison table of precision rate and recall rate of different annotation methods

Annotation method	Precision rate	Recall rate
Text annotation	0.578	0.434
Multi-source information text annotation	0.578	0.691
Denoised multi-source information text annotation	0.792	0.632

As could be seen from table 2, the similarity threshold is set at 0.8. When multi-source information text annotation method is used, the precision rate will be reduced due to the introduction of both correct and wrong annotation words, but the recall rate is greatly improved compared with text annotation method. When the text annotation of multi-source information is denoised, the precision rate is greatly improved, and the recall rate is much higher than that of text annotation method. Therefore, the denoised multi-source text annotation method is suitable for image annotation processing of quality control data.
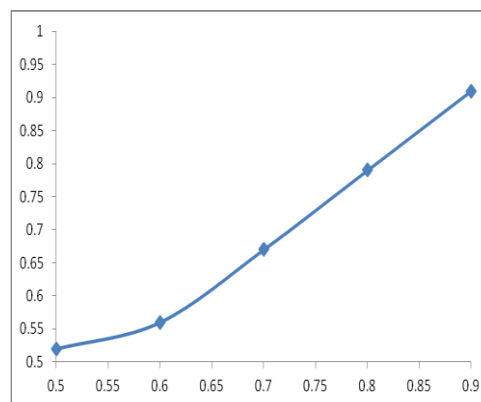
Figure 2. Precision rate when the similarity threshold is 0.5-0.9
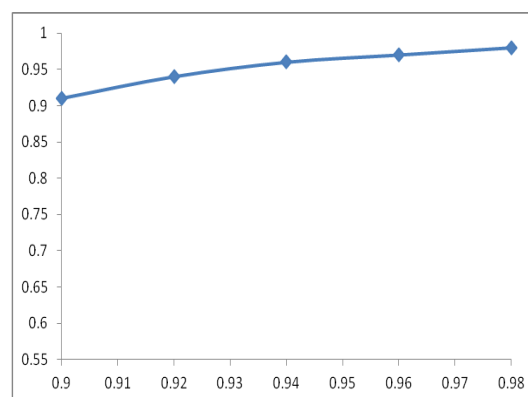
Figure 3. Precision rate when the similarity threshold is 0.9-0.98

It could be seen from figure 2 and figure 3 that, the precision rate increases with the increase of similarity threshold. When the similarity threshold is greater than 0.9, the precision rate reaches a high level and the growth rate gradually slows down.

4. Conclusions

There is a lot of image information in product quality and safety data, which makes it difficult to find useful information from a lot of image information for analysis and processing. The research analyzed the product quality and safety information by using multi-source text information annotation technology, then obtained the total similarity by the weight fusion of each similarity after the calculation of image similarity and user image similarity of each image, and sought the similar image sets through the setting of similarity threshold, so as to annotate the image to be analyzed by using the annotation in the similar image sets. The experimental results show that the image annotation technology based on noise reduction and multi-source big data fusion could effectively annotate the images related to quality inspection, of which both the precision rate and recall rate is better than that of text annotation method.

References

[1]   Wang J, andLi J 2002 Learning-based linguistic indexing of pictures with 2d MHMMs Proceedings of ACM International Conference on Multimedia, pp436-445

[2]   Kumar S ,andHebert M 2005 A Hierarchical field framework for unified context-based classification Proceedings of IEEE International Conference on Computer Vision pp ,1284-1291

[3]   Monay F and,Gatica-Perez D 2003On image auto-annotation with latent space models Proceedings of ACM International Conference on Multimedia pp ,275-278

[4]   Martinez O and,Tsechpenakis G 2008 Integration of active learning in a collaborative CRF Proceedings of IEEE Computer Vision and Pattern Recognition Workshops pp,,1-8

[5]   Wang X J,Zhang L,Jing F,et al 2006AnnoSearch,image auto-anno-tation by search Proceedings of IEEE Conference on Computer Vision and Pattern Recognition pp,1483-1490

[6]   Blei D M and ,Jordan M I 2003Modeling annotated data Proceedings of ACM SIGIR Conference on Research and Development in Informaion Retrieval pp,127-134

[7]   Richard X H and ,Zemel R S 2004Multiscale conditional random fields for image labelling IEEE Conference on Computer Vision and Pattern Recognition pp,695-702