Research Article

Research on Online Rapid Sorting Method of Waste Textiles Based on Near-Infrared Spectroscopy and Generative Adversity Network

Jinquan Hu,1,2 Huihua Yang,1,3 Guoliang Zhao,4 and Ruizhi Zhou5

1School of Artificial Intelligence, Beijing University of Posts and Telecommunications, 10 Xitucheng Road, Haidian District, Beijing, China
2School of Electronic Engineering and Automation Chemistry, Guilin University of Electronic Technology, No. 1 Jinji Road, Qixing District, Guilin, Guangxi, China
3School of Computer Science and Information Security, Guilin University of Electronic Technology, No. 1 Jinji Road, Qixing District, Guilin, Guangxi, China
4Beijing Institute of Fashion Technology, Beijing 100029, China
5Shanghai Jicai Environmental Protection Technology Co., Ltd., Shanghai, China

Correspondence should be addressed to Huihua Yang; jxhjq@guet.edu.cn

Received 3 March 2022; Revised 9 April 2022; Accepted 23 April 2022; Published 14 May 2022

Abstract

In this paper, aiming at the application of online rapid sorting of waste textiles, a large number of effective high-content blending data are generated by using generative adversity network to deeply mine the combination relationship of blending spectra, and a BEGAN-RBF-SVM classification model is constructed by compensating the imbalance of negative samples in the data set. Various experiments show that the model can effectively extract the spectrum of pure textile samples. The classification model has high robustness and high speed, reaches the performance of similar products in the world, and has a broad application market.

1. Introduction

With the rapid development of economy and the improvement of people’s daily consumption level, China consumes about 60 million tons of textiles made of various fiber raw materials every year on average [1, 2]. According to the Annual Report on Comprehensive Utilization of Resources in China published by the National Development and Reform Commission, China produces more than 25 million tons of waste textiles every year. The comprehensive utilization of waste textiles is less than 5 million tons, and the recycling rate is less than 20%. Therefore, the recycling of waste textiles is very necessary. However, due to the great difference in the component content of waste textiles, it is very difficult to screen the raw materials in the process of recycling. At present, sorting is still mainly done manually, with slow speed and low recognition rate. It cannot meet the requirements of industrial-scale production. Therefore, online efficient identification and automatic sorting at the recycling site is an important prerequisite and guarantee for the recycling of waste textiles.

The research of artificial neural network is an important branch in the field of artificial intelligence. Since Hinton put forward the deep neural network in 2006, the neural network has made great achievements in the field of information processing such as voice, image, and text [3–5]. Researchers in major universities and research institutions all over the world have vigorously carried out relevant research and proposed a variety of deep neural networks, which has greatly expanded the application of deep neural network [6, 7]. Modern deep neural network model mainly includes three basic structures: sequence-to-sequence neural network, convolution neural network, and adversarial generation network [8, 9].
Near-infrared spectroscopy (NIR) is a comprehensive technology that integrates spectral measurement technology, chemometrics, and computer technology. It is the best choice to achieve real-time, rapid, and nondestructive measurement [10]. Its online technology has been successfully applied in the fields of medicine, petrochemical, food, and agriculture, and textile testing is beginning to move from the laboratory stage to the industrial application stage [11]. The industrial application of the near-infrared spectroscopy analysis method for waste textiles is different from the laboratory mainly because of the diversity of recovered samples. For the industrial scene, the spectrum of textiles is affected by many factors such as dyeing, weaving, processing technology, damage degree, complex blending composition, and so on [12], so it is difficult to establish a complete data set. Field analysis results are much lower than laboratory results. In this paper, the deep learning method is used to deeply mine the potential distribution information of the original spectral data, and the generative adversary network is used to construct the virtual data set to compensate for the imbalance of the original data set. The test results show that the method can effectively improve the completeness of the data set and the accuracy of the model.

2. Experimental Part

2.1. Experimental Instruments. In this paper, the noncontact diffuse reflectance near-infrared spectrometer DA7440 of Pertern Company was used as the spectrum acquisition equipment, the spectral range was 950–1650 nm, the resolution was 5 nm, the number of scans was 100, and the time to generate the spectrum was less than 700 millisecond.

2.2. Source of Samples. All samples, collection equipment, and experimental sites are provided by Shanghai JICAI Environmental Protection Technology Co., Ltd., and the chemical analysis results (labels) of the samples are provided by the Beijing Institute of Fashion Technology. Since the most valuable part of the recycled waste textiles is the pure textiles of various materials, the sample set is shown in Table 1.

The pure material samples and the blended samples in the table are divided into a training set and a test set according to different proportions and are used for training the model. The field test set is used for the final performance verification and evaluation of the model. The field test included 3 pieces of pure acrylic, 48 pieces of pure cotton, 47 pieces of pure polyester, 41 pieces of wool, and 86 pieces of blended yarn.

2.3. Sample Set Characteristics. Recycled waste textiles are usually used and idle for several years, with various varieties and colors, unknown processing conditions, different degrees of damage, different thicknesses, and random spectral acquisition positions, and are easily affected by buttons and zippers, which lead to different degrees of spectral changes and affect the classification performance of the model.

Table 1: Sample type and quantity.

| Serial number | Species                     | Quantity |
|---------------|----------------------------|----------|
| 1             | Pure acrylic fiber         | 75       |
| 2             | Pure cotton                | 226      |
| 3             | Pure polyester             | 240      |
| 4             | Pure wool                  | 473      |
| 5             | Blending                   | 100      |
| 6             | Pure viscose fiber         | 1        |
| 7             | Pure silk                  | 1        |
| 8             | Field test set             | 225      |

Figure 1 shows a raw near-infrared spectrum of a portion of a textile. Figure 1(a) shows two spectra of pure acrylic, pure cotton, pure polyester, and pure wool samples, respectively. It can be seen that different kinds of pure materials have their own characteristics, and there are great differences between them. It is easy to distinguish different pure materials. However, it also shows that it is difficult to obtain good classification performance if the pure material samples are used as the positive and negative samples of the training model, respectively. Figure 1(b) is the original spectrum of some pure wool, which shows that pure wool samples have some common characteristics, but some samples show obvious other characteristics because the sample label does not contain the color, process, and other information of the sample. It is not possible to determine the specific reasons for the different characteristics of the same material sample, but the model must be able to accurately extract the pure material samples affected by various factors. Figure 1(c) shows the spectra of some pure wool and wool-containing blended samples. It can be seen that different contents, different material blends, and blending processes will have different degrees of impact on the spectra of samples. High content of blended samples and some special blended samples (such as viscose) will seriously affect the classification accuracy of the model, and these samples will be better as the negative samples of the training model. But in fact, there are many kinds of blended fabrics, and it takes a lot of time and economic cost to get more blended samples. Figure 1(d) shows the spectra of some pure cotton samples with different colors. It can be seen that different color processes will have a great impact on the spectrum of the sample. Of course, in addition to the different colors of these samples, it is uncertain whether other processing processes are the same, so it is uncertain whether the impact is entirely caused by dyeing, but it can be explained that for waste textiles, many factors will cause great changes in the spectrum of samples, thereby increasing the classification difficulty of the model.

In general, the near-infrared spectra of waste textiles have obvious species characteristics, but they are also affected by many factors, and the samples that cause classification errors are mainly some high-content blends and some special process blends. Collecting more data sets is the best way to improve the accuracy of the model, but it takes a lot of time and economic costs. It is very difficult for the industry to work after obtaining a complete data set, and we must deal with these problems as much as possible from the algorithm.
3. Algorithm Structure

The research on the application of near-infrared spectroscopy in textile classification and quantitative analysis has been widely carried out in various research institutions and universities, and many mature algorithms have been accumulated. However, most of the algorithms are aimed at some specific applications or specific types of processing methods, and there are few comprehensive application methods for all waste textiles. There is still a large research gap in the field of industrial application. Because the composition information of textiles needs to be determined by chemical methods, the cost of time and economy is relatively high, resulting in a small number of data sets and unbalanced distribution, which greatly affects the performance of the model. One way to deal with data imbalances is to create virtual data sets. The generative adversarial network (GAN) [13–15] has an excellent performance in this respect. In this paper, the virtual blending data set is constructed by using the generative antagonistic network structure to supplement and improve the blending data, especially the high content blending data, so as to provide the accuracy of classification and other performances.

The GAN was first created by Ian Goodfellow in 2014 and consists of a generator (G) and a discriminator (D), both of which are multilayer perceptual networks, as shown in Figure 2. The generator fabricates data from a random noise $Z$ input, and the discriminator is responsible for determining whether the data are true and giving a score.

The objective function of the GAN model is represented by the following:
\[
\min_G \max_D \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_{noise}} [\log (1 - D(G(z)))].
\]

(1)

The log \([D(x)]\) in the first item represents the result of the discriminator’s judgment of the real data, and the second item represents the synthesis and judgment of the data. The generators and discriminators are trained alternately through a two-sided game between the maximum and minimum values until the Nash balance is reached. Given any generator \(G\), the training criteria for discriminator \(D\) are to maximize \(V(D, G)\). Therefore, assuming that the generator is fixed, the maximized objective function can be represented by the following formula:

\[
V(D, G) = \int_x p_{data}(x) \log(D(x)) + p_g(x) \log(1 - D(x)) dx.
\]

(2)

The formula for calculating the optimal discriminator is

\[
D^*_G(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}.
\]

(3)

The optimal discriminator is substituted into the objective function \(V(D, G)\) to obtain the objective function of \(G\):

\[
C(G) = \mathbb{E}_{x \sim p_{data}} \left[ \log \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \right] + \mathbb{E}_{z \sim p_z} \left[ \log \frac{p_g(z)}{p_{data}(x) + p_g(z)} \right]
\]

(4)

When \(p_g = p_{data}\), \(D = 0.5\), the discriminator cannot distinguish between real samples and generated samples, and the generator perfectly replicates the data generation process.

In this paper, Boundary Equilibrium GAN (BEGAN) is used. The optimization objective of BEGAN is slightly different from that of the original GAN. BEGAN requires that the reconstruction error distribution of the two should be as close as possible. The reconstruction error is calculated by

\[
L(v) = |v - D(v)|^\gamma,
\]

(5)

where \(v\) represents the input data of the discriminator; \(D(v)\) represents the output of the discriminator \(D\), and \(\eta\) is the norm selection parameter.

BEGAN uses proportional control theory during training to keep the balance between the generator and the discriminator during training, namely

\[
\gamma = \frac{E[L(G(z))] - L(G(x))}{E[L(x)]}.
\]

(6)

When training, you can choose \(\gamma\) to adjust the generation effect. Proportional control theory makes the training stable and fast and greatly improves the convergence speed of the network.

The network structure of the system is shown in Figure 3. The whole network structure is composed of a generator \(G\), a classifier \(C\), and a discriminator \(D\), wherein the generator \(G\) and the discriminator \(D\) are multilayer perceptrons, and the classifier \(C\) is a pretrained Gaussian kernel support vector machine. The inputs \(I_1\) and \(I_2\) of the generator \(G\) are two randomly selected pure material spectra, and \(Z\) is a random number. \(I_{pred}\) is the output of the generator. It is the mixture spectrum of \(I_1\) and \(I_2\) generated by simulation, and the mixture ratio is randomly selected between 50% and 99%. The discriminator uses the error between the output \(I_{pred}\) of the Wasserstein distance metric generator and the mixed material spectrum \(I_3\) and ends the training when the two are balanced. The optimization objectives of the discriminator and the generator are

\[
L_D = L(I_3) - k_\lambda L(I_{pred}),
\]

\[
L_G = L(I_{pred}),
\]

\[
k_{i+1} = k_i + \lambda_k \left[ \gamma L(x) - L(G(z)) \right].
\]

(7)

The classifier \(C\) is trained with the pure material spectrum \(I_4\) as a positive sample in advance so that it has the ability to correctly classify the pure material spectrum. The classifier uses the output \(I_{pred}\) of the generator as a negative sample to ensure the diversity of \(I_{pred}\) in the training process of the antagonistic network.

4. Experiment and Analysis

In order to test the effectiveness of the algorithm, several experiments are carried out. The first experiment is the spectrum generation experiment, which is used to verify that the generative antagonistic network can effectively generate the mixed sample spectrum. The second experiment is the classification model training and testing experiment, which compares the performance of the classification model with other classification models. The third experiment is a field test experiment, in which the classification model is deployed in the sorting system, and the operators test the accuracy and classification time on the spot.

4.1. Spectrum Generation Experiment. Experimental environment: CPU: Intel Core i9, memory: 32 GB, GPU: Nvidia GT1080, operating system: WIN10, software: PYTHON.

Experimental method: After the training of the generated antagonistic network, take several outputs of the generated network, record the mixing parameters, and draw the spectrum as shown in Figure 4.
Figure 4(a) shows part of the generated spectrum and part of the actual spectrum. There is no connection between them, but it can be seen that they all reflect the peak characteristics of the main components. Figure 4(b) shows two selected spectra with close mixing ratios, in which the first and third spectra are generated spectra, and the second and fourth spectra are actual spectra. It can be seen that the generated spectrum not only reflects the peak characteristics of the main components but also reflects the diversity, which can enrich the spectral samples of blended materials and improve the classification performance.

4.2. Comparison Experiment of Classifier Performance. The experimental environment is the same as above. Experimental method: The data set shown in Table 1 was used for the experimental data set, and the first derivative of SG and SNV was used for the spectral preprocessing method. Firstly, all pure material spectrum samples are used to train RBF-SVM, then all blended samples are used to train BEGAN, and then 1000 50%–99% random component mixed spectra are generated by the trained generator as negative samples. Finally, all the pure material spectra, blended spectra, and the generated spectra were used to synthesize the data set, which was divided into training set and test set according to the ratio of 9:1, and RBF-SVM was trained and tested by grid optimization method. The comparison algorithms include linear SVM, extreme learning machine, KNN, and decision tree. The training set and test set are allocated by leave-one-out method, and the average value is obtained by repeating 50 times. Considering the imbalance of positive and negative samples, this paper uses sensitivity and

| Performance | CLASS | BEGAN-RBF-SVM | Linear-SVM | ELM | KNN | Precision Tree | One-class SVM |
|-------------|-------|---------------|------------|-----|-----|----------------|---------------|
| Sensitivity | Acrylic | 0.97 | 0.91 | 0.90 | 0.94 | 0.88 | 0.92 |
|             | Cotton | 0.94 | 0.89 | 0.89 | 0.91 | 0.85 | 0.88 |
|             | Poly | 0.94 | 0.90 | 0.91 | 0.92 | 0.83 | 0.90 |
|             | Wool | 0.93 | 0.85 | 0.83 | 0.87 | 0.82 | 0.87 |
| Specificity | Acrylic | 0.01 | 0.13 | 0.12 | 0.08 | 0.16 | 0.15 |
|             | Cotton | 0.05 | 0.09 | 0.13 | 0.05 | 0.11 | 0.10 |
|             | Poly | 0.06 | 0.16 | 0.15 | 0.04 | 0.23 | 0.17 |

Figure 4: Generated spectrogram.

Figure 5: Schematic diagram of sorting line.
specificity to measure the performance of each classifier, and the experimental results are shown in Table 2:

It can be seen from the above table that the performance of various conventional machine learning algorithms is relatively close, among which the effect of KNN is slightly better, but its decision time for a single sample is much longer than that of other algorithms, and the test time on the local machine is more than 300 ms per sample, which is not suitable for on-site rapid measurement. It should be noted that although the algorithm in this paper has obvious advantages, it can only show that the algorithm can effectively classify but cannot fully explain its superiority over other algorithms because they do not use the same training set and test set. It should also be noted that the specificity for acrylic is 0 because there is only one sample with more than 50% acrylic in the blend sample.

The one-class SVM algorithm in the above table is a classification algorithm designed for unbalanced data sets. This method only uses the original data sets for training and testing. It can be seen from the results in the table that the performance of this method is close to that of other traditional methods and slightly inferior to that of the algorithm in this paper.

4.3. Classifier Field Accuracy Test Experiment. Experimental environment: The sorting system uses a desktop computer as the main controller, the CPU is Intel Core i7, the memory is 16 GB, and there is no GPU. The computer is connected to the near-infrared spectrometer DA7440 of Pertern Company through the network interface, and it takes 0.9 seconds to collect one spectrum and generate a CSV file. The algorithm is deployed in the WINDOWS application, and the trained model is used to complete the classification by reading the data in the CSV file.

Experimental method: 225 pieces of waste clothes with tags are provided on-site and placed on the sorting line one by one by the operator. The sorting line is equipped with a weighing sensor as the starting valve for starting a sorting. The schematic diagram of the sorting line is shown in Figure 5. When the weighing sensor detects the textile, it will delay 200 ms to start the spectrometer to collect the spectrum, and then control the turntable system to put the textile into the corresponding box after spectrum preprocessing and classifier judgment.

Because each part works in parallel, the average processing time of each textile is less than 1.2 seconds and meets the sorting requirements of 2000 pieces per minute. When 225 test samples were sorted, the operator manually counts the sorting results according to the sorting results of each sample and the tag on it and repeats three times. The test results are shown in the following table.

In the above table, TP represents the positive sample of correct prediction, that is, the high-value sample to be extracted, which should be as close as possible to the actual number. FP indicates the positive sample of error prediction, that is, the blended sample is judged to be a pure material sample, which is a misjudgment and will reduce the purity of the product. At present, the purity of recycled waste textiles in the market is 93%–95%. FN indicates the positive sample of wrong prediction, that is, the pure material sample is judged as the blended sample, which is a missed judgment and will reduce the benefit of the product. The average recall rate in the above table basically meets the benefit requirements of industrial production.

It should be noted that the test results in the above table are better than the results of Experiment 2, which is not consistent with the general industrial situation. This is because in most of the recycled waste blended samples, the proportion of each content is in the range of 3/7, 2/8, or 4/6, and the proportion of 90% or even more than 95% is less. In the training process of the model, many high-content blended samples are generated as negative samples, so the model generally does not misjudge the low-content blended samples. The reason for missing judgment may be that the position of spectrum acquisition is affected by other factors such as too thin textile, zipper or button, and special dyeing process, which causes great changes in the spectrum.
In particular, the dataset was manually sorted by several workers before being sent to the laboratory for chemical composition analysis. Then, according to the results of the chemical composition analysis, the sorting accuracy is between 50% and 80%, which is related to the proficiency of workers and the sorting speed. After the chemical analysis of the sample, the component label is nailed on the collar. Considering the memory ability of workers, the comparison test with manual sorting will not be carried out after the sorting equipment is adopted.

4.4. Field Batch Pure Material Spectrum Extraction Experiment. The experimental environment is the same as above.

Experimental method: The pure material samples on the spot were mixed into 3000 unknown samples, then sorted by the sorting line, and then, the operator manually searched and compared the classification results of these pure material samples and repeated three times, and the test results are shown in the following table.

Since the situation of 3000 unknown samples is unknown, it is impossible to judge whether there is a misjudgment (blended yarn is judged as pure yarn), and only the TP and FN values of pure yarn samples with tag labels are counted. Compared with the results in Table 3, the results in Table 4 do not decrease significantly, indicating that the system model has high robustness and is fully suitable for industrial application.

5. Conclusion

In this paper, aiming at the current situation of waste textile recycling at home and abroad, this paper puts forward an online rapid sorting application method based on near-infrared spectroscopy. Recycling enterprises are limited by the scale of economic investment, it is difficult to build a complete textile near-infrared spectrum analysis data set, and the effect of using traditional near-infrared spectrum analysis methods is poor. In this paper, the combination relationship of the blending spectrum is deeply mined by generating countermeasure network, a large number of effective high content blending data are generated, the negative sample imbalance of the data set is compensated, and the BEGAN-RBF-SVM classification model is constructed. Various experiments show that the model can effectively extract the spectrum of pure textile samples. The classification model has high robustness and fast speed, which reaches the performance of similar international products and has a broad application market.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by National Key R&D Program of China (Grant No. 2018AAA0102600) and Guangxi Key Laboratory of Automatic Detection Technology and Instrument Foundation.

References

[1] Textile Research, “Researchers from Qingdao university report on findings in textile research (rapid identification of plant- and chemical-dyed cotton fabrics using the near-infrared technique),” Journal of Technology, pp. 127–132, 2020.

[2] Z. Li, “Application of research and development of intelligent wall insulation materials based on recycling of waste textiles in civil engineering health monitoring,” Advances in Civil Engineering, vol. 2021, Article ID 5585829, 10 pages, 2021.

[3] R. De, A. Chakraborty, and R. Sarkar, “Document image binarization using dual discriminator generative adversarial networks,” IEEE Signal Processing Letters, vol. 27, pp. 1090–1094, 2020.

[4] C. Wang, C. Xu, X. Yao, and D. Tao, “Evolutionary generative adversarial networks,” IEEE Transactions on Evolutionary Computation, vol. 23, no. 6, pp. 921–934, 2019.
[5] A. Oluwasanmi, M. U. Aftab, A. Shokanbi, J. Jackson, B. Kumeda, and Z. Qin, “Attentively conditioned generative adversarial network for semantic segmentation,” IEEE Access, vol. 8, Article ID 31733, 2020.

[6] D. M. M. Yacout and M. S. Hassouna, “Identifying potential environmental impacts of waste handling strategies in textile industry,” Environmental Monitoring and Assessment, vol. 188, no. 8, p. 445, 2016.

[7] P.-C. Su, Y.-H. Hsueh, M.-T. Ke, J.-J. Chen, and P.-C. Lai, “Noncontact ECG monitoring by capacitive coupling of textiles in a chair,” Journal of Healthcare Engineering, vol. 2021, Article ID 6698567, 8 pages, 2021.

[8] H. Hong, X. Li, and M. Wang, “GANE: a generative adversarial network embedding,” IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 7, pp. 2325–2335, 2020.

[9] C. Rong, X. Zhang, and Y. Lin, “Feature-improving generative adversarial network for face frontalization,” IEEE Access, vol. 8, Article ID 68842, 2020.

[10] Nanotechnology-Nanoparticles, “Recent Findings from Yazd University Has Provided New Information about Nanoparticles (Comfort and reflectance properties of viscose/polyester blend fabric printed by vat/disperse dyes in visible/near-infrared region),” Technology News Focus, vol. 2020, pp. 356–359, 2020.

[11] M. Li, G. Han, W. Jiang et al., “Rapid identification of plant- and chemical-dyed cotton fabrics using the near-infrared technique,” Textile Research Journal, vol. 90, no. 19-20, 2020.

[12] Textile Research, “Researchers at Qingdao university report new data on textile research (rapid identification of fibers from different waste fabrics using the near-infrared spectroscopy technique),” Technology News Focus, pp. 231–233, 2019.

[13] X. Yu, S. Wang, Y. Gao, and Z. Bao, “Orthogonal design-guided preparation of multi-level porous-activated carbon by pyrolysis of waste polyester textiles,” Environmental Science and Pollution Research, vol. 25, no. 30, Article ID 30567, 2018.

[14] X. Li, Z. Jin, Yingjie Qiu, J. Jiang, and X. Jin, “Study on the hyper-spectral image characteristics of liquid stains in waste textiles,” in Proceedings of the 2021 Fourteenth International Symposium on Computational Intelligence and Design (ISCID), pp. 117–120, Hangzhou, China, December 2021.

[15] M. Lee and J. Seok, “Controllable generative adversarial network,” IEEE Access, vol. 7, Article ID 28158, 2019.