

Research Article

A New Method for Feature Extraction and Classification of Single-Stranded DNA Based on Collaborative Filter

Bingyong Yan ¹, Haixu Cui ¹, Haitao Fu,² Jiale Zhou,^{1,3} and Huifeng Wang ¹

¹Key Laboratory of Advanced Control & Optimization for Chemical Process of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

²School of Electronic Information Engineering, Beijing Institute of Technology, Beijing 100081, China

³Key Laboratory for Advanced Materials & Department of Chemistry, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Bingyong Yan; byyan@ecust.edu.cn and Huifeng Wang; whuifeng@ecust.edu.cn

Received 20 March 2020; Accepted 11 May 2020; Published 21 July 2020

Guest Editor: Hou-Sheng Su

Copyright © 2020 Bingyong Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The traditional support vector machine algorithm is not enough to classify single-stranded DNA molecules, so this paper proposes an improved threshold extraction algorithm based on collaborative filter for the classification of single-stranded DNA. Firstly, according to the different characteristic curves of the blocking current signals formed by the four bases (A, T, C, and G) that make up DNA molecules crossing the nanopore, the collaborative filter feature extraction algorithm with improved threshold is proposed. Then, the feature information is reconstructed and sent to the SVM classifier for training. Finally, the unfiltered, collaborative filter, improved threshold collaborative filter, and Bessel filter data are, respectively, extracted and sent to the SVM classifier for classification and comparison research. The experimental results show that the improved collaborative filter algorithm has higher accuracy in single-stranded DNA molecular classification.

1. Introduction

In recent years, nanochannel technology has developed into an indispensable tool for single molecule experiments, which provides a new way for high sensitive detection of single molecules and the study of weak interaction between single molecules. This technology is widely used in DNA single molecule sequencing, protein structure analysis, and early diagnosis of major diseases. Nanochannel technology is mainly used to analyze the weak blocking current signal generated by the unknown molecule that passes through the nanopore and to study the information of biogenetics and life science. Compared with the traditional detection technology, it has the characteristics of simple operation, clear structure, and fast detection speed, so it is called the most promising third generation DNA sequencing technology [1–4].

Due to the huge amount of data of blocking current generated by the molecules to be measured crossing the

nanopore, the traditional data analysis and processing methods are far from meeting the requirements of DNA sequencing. Therefore, support vector machine and other auxiliary research tools have undoubtedly become one of the powerful tools for analyzing single-stranded DNA data [5].

At present, many researchers have applied SVM in bioinformatics recognition [6, 7]. For example, Balachandran et al. [8] used the SVM model to predict in vitro phage virus proteins. Zhao et al. [9] used the SVM model to recognize amino acids. Zhong et al. [10] used SVM as a base classifier to recognize miRNA precursors. Zhou et al. [11] used the SVM model to recognize the DNA sequences of analytes such as *Bacillus subtilis*. Kumar et al. [12] used SVM to classify RNA-binding and nonbinding proteins. Dai [13] used SVM to classify imbalanced protein data. Through the above research and analysis, we can see that the classification rate using traditional SVM for classification is difficult to improve. In order to further improve the recognition rate, Tabard-Cossa et al. [14] and Kowalczyk et al.

[15], respectively, studied the synthesis of enhanced nanopores, the mechanism of noise generation, and the noise model of nanopores. Dekker [16] and Goto et al. [17] designed low-noise I-V conversion sensor methods to denoise nanopores. These methods can improve the signal-to-noise ratio of the blocking current, and the accuracy of the recognition is improved to a certain extent. However, because the collected blocking current signal is a very weak picoampere signal, most of the research on denoising of blocking current signal is only based on the analysis of external physical conditions, while there is little research on the specific blocking current signal itself.

Considering the existing research problems, this paper proposes a new collaborative filter classification method based on improved threshold. The basic idea is to use a fixed force between a single base and a nanopore [9], while the force between adjacent bases is uncertain, so the fluctuation of blocking current signal value is in a small range, but the block current signal generated by the same base through the nanopore shows a certain similarity in the whole signal [18, 19]. Therefore, based on the self-similar structure of the nanopore blocking current signal in the entire time domain, the collaborative filter algorithm was first used to analyze the grouped signals. By introducing the compensation factor, an improved threshold selection algorithm was proposed to extract the characteristics of the signal. Then, the processed data are reconstructed and sent to the SVM for training. Finally, the above algorithm was used to analyze the blocking current signals generated by A_{14} and CA_3 single-stranded DNA molecules through the nanopore.

2. Introduction to Improved Collaborative Filter Algorithm

Considering the similarity of blocking current signals with the same base in the entire blocking current signal, the new feature extraction and classification method proposed in this paper are shown in Figure 1.

The first step is to use the blocking current signal generated from the DNA molecule through the nanopore channel as raw data.

The second step is to find out the similar blocks of the raw data and divide the most similar n blocks into a group with a certain threshold.

The third step is to coprocess the n groups. At this time, each group is a matrix. First, the n group matrix is subjected to two-dimensional discrete transformation, respectively, and then processed by introducing improved thresholds to filter out noise. Finally, two-dimensional discrete inverse transformation is used to reconstruct the raw signal. The reconstructed signal is the filtered signal with obvious characteristics.

In the fourth step, the current blocking curves after reconstruction of the characteristics of the two DNA single-stranded molecules A_{14} and CA_3 processed in the first three steps are labeled and mixed and then sent to the SVM for training, and the classification results are analyzed.

The details of each functional module are described below.

2.1. Grouping of Signals. Figure 2 is a schematic diagram of grouping similar blocks. Each grouping block with similar characteristics is grouped for collaborative processing to reveal the characteristics of noise coverage and provides guarantee for SVM classification.

The selected blocking current signal is marked as R , a reference segment is first selected from R as D , and the comparative segment L from R is then selected without repeating. And Euclidean distance is used to judge the similarity between D and L [20]:

$$d_{(D,L)} = \sqrt{\sum_{i=1}^n (D_i - L_i)^2}, \quad i = 1, 2, 3, \dots, n, \quad (1)$$

where i is the selected i -th segment.

Then, it is normalized [21]:

$$\bar{d}_{(D,L)} = \frac{\sqrt{\sum_{i=1}^n (D_i - L_i)^2}}{w}, \quad i = 1, 2, 3, \dots, n, \quad (2)$$

where w is the width of the selected reference segment. $\bar{d}_{(D,L)}$ is smaller, with higher similarity between D and L .

Then, the fixed reference segment D is selected and searched in the entire area of the blocking current signal length l ($l \gg w$). At the same time, L moves across the entire segment R in steps of k and obtains m segments with the smallest distance from the reference segment to form group (D). And it is saved to a two-dimensional array of m rows and w columns, group (D) $_{m \times w}$.

Finally, the reference segment D traverses the entire blocking current signal in steps of k and records groups formed by different reference segments.

2.2. Collaborative Processing. For the n groups generated in Section 3.1, this section uses collaborative filter to perform filter processing on each group of signals in order to be able to extract the characteristic information of the grouped signals.

Collaboration: each segment in each group is traversed through the entire blocking current signal, and each group contains the information of other groups, so this process can be regarded as a "collaborative" process.

Collaborative filter consists of three steps:

The first step is the two-dimensional discrete transformation of groups, and each group forms a matrix.

The second step performs threshold processing on each group matrix to filter the noise information in the raw data.

The third step is to transform the two-dimensional discrete inverse transformation on the matrix after the threshold processing in the second step and reconstruct signal with obvious characteristic information.

Each step is explained in detail as follows:

(1) Two-dimensional discrete transformation of the group:

$$G(D) = \text{dct2}(\text{group}(D)), \quad (3)$$

where $\text{dct2}(\cdot)$ is the two-dimensional discrete cosine transform.

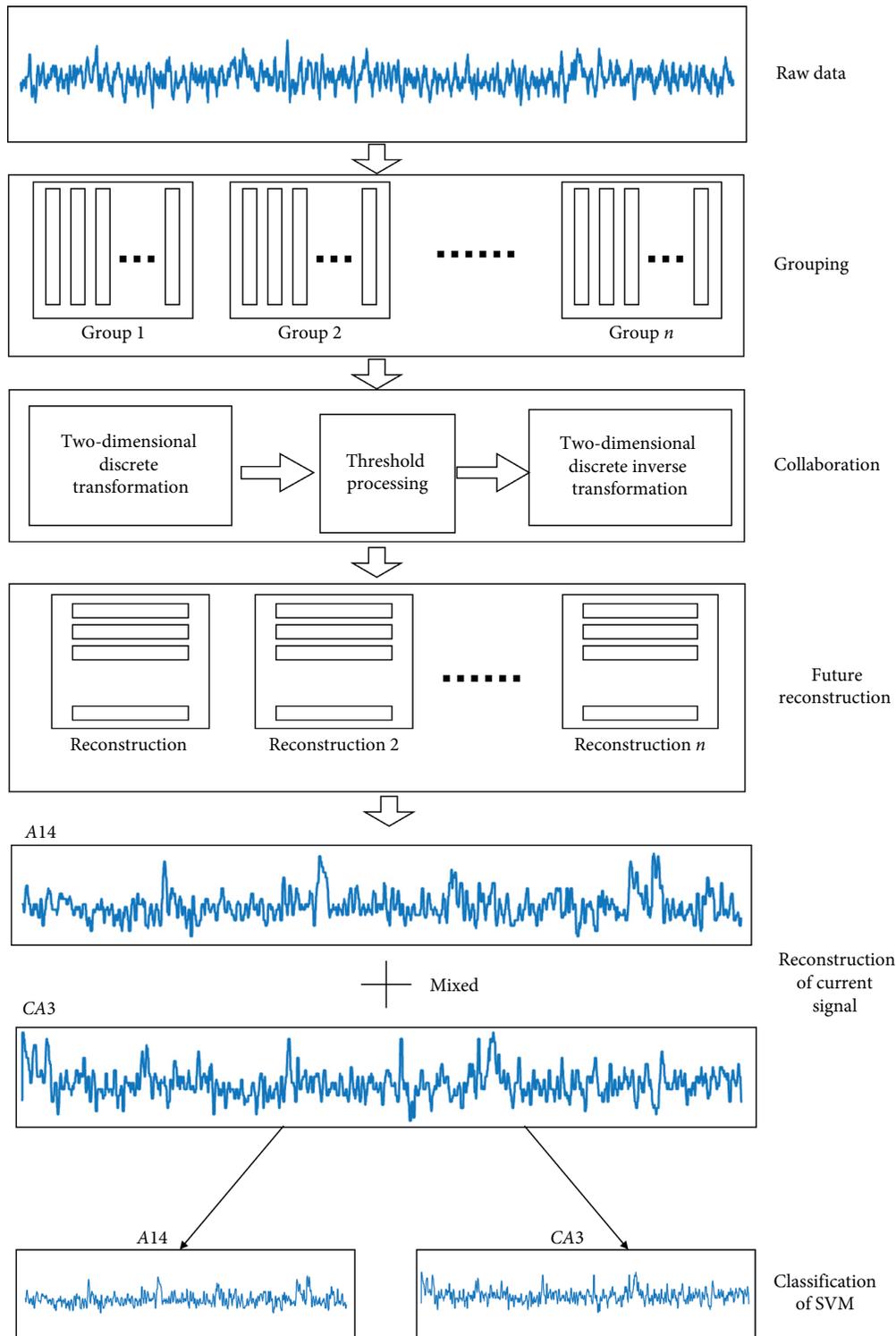


FIGURE 1: Flowchart of collaborative filter algorithm.

(2) A threshold value is selected for each group as λ and threshold noise reduction is performed in the transform domain. Coefficients smaller than the threshold

value are set to zero to attenuate noise, and coefficients larger than the threshold value are retained. This paper uses the hard threshold method, which is defined as

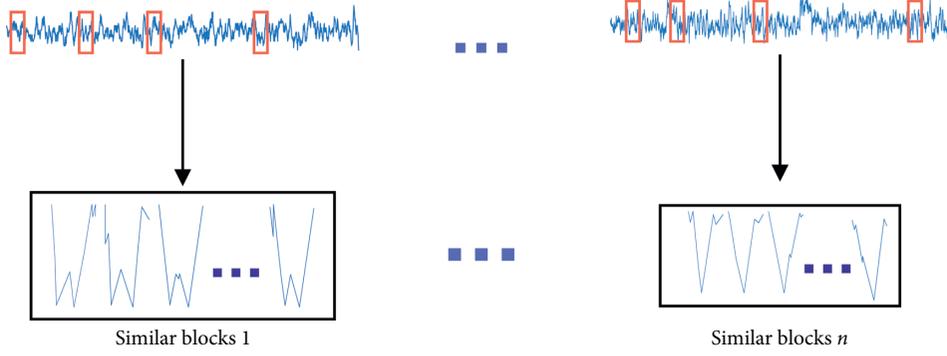


FIGURE 2: Similar block grouping.

$$T_H(g) = \begin{cases} g, & |g| < \lambda, \\ 0, & |g| \leq \lambda, \end{cases} \quad (4)$$

where λ is based on the threshold denoising method of Donoho, which is approximately optimal in the sense of mean square error, and at the same time can ensure that the reconstructed signal has the smoothness of the raw signal. The definition of threshold by VisuShrink proposed by Donoho and Johnstone is [22] as follows:

$$\lambda = \delta^* \sqrt{2 \log^*(m^*w)}, \quad (5)$$

where δ is the noise standard deviation of the raw signal.

Since the noise of the blocking current signal of DNA passing through the nanopore is unknown, this paper uses the absolute deviation of the median of the coefficient matrix $G(D)$ to estimate [18]:

$$\text{MAD} = \text{median}(|D_i - \text{median}(D_i)|), \quad i = 1, 2, 3, \dots, n, \quad (6)$$

where MAD is median absolute deviation, $\text{median}(\cdot)$. D_i is the element in the coefficient matrix $G(D)$. The estimated noise standard is defined as

$$\hat{\delta} = \frac{\text{MAD}}{k}, \quad (7)$$

where k is the scale factor constant, which is generally selected as 0.6745 [23].

- (3) After thresholding the transform coefficient matrix, the grouped filter results are obtained by two-dimensional discrete inverse cosine transform, as follows:

$$\text{group}(D) = \text{idct2}\{T_H(g)\}, \quad (8)$$

where $\text{idct2}(\cdot)$ is the two-dimensional discrete cosine transform.

Through three steps, the n groups in 2.1 can be processed collaboratively, and finally, n group signals with noise removed can be obtained.

2.3. Improved Threshold. In the case of actual measurement, the additional noise changes caused by slight environmental differences and the small changes in hardware circuit components and reference ground can cause signal drift.

Although the data can be filtered using a collaborative filter to reduce noise interference, if the input signal has drift and contains nonzero mean noise interference, it will lead to the deviation of the final results of data processing. Therefore, it is necessary to compensate the drift of system.

This section improves the threshold value in the second step of collaborative filter data processing in Section 2.2 and introduces threshold compensation factor λ_c to compensate the drift of system.

The improved threshold is defined as

$$\lambda = \delta^* \sqrt{2 \log^*(m^*w) + \lambda_c}, \quad (9)$$

where when the circuit is at zero input signal, the output value of the acquisition circuit at this time is R_0 . The data processing methods in Sections 2.1 and 2.2 are used to obtain the threshold λ_i , $i = 1, 2, \dots, n$ without input.

Compensation factor is defined as

$$\lambda_c = \frac{1}{n} \sum_{j=1}^n \lambda_j^i. \quad (10)$$

2.4. Feature Extraction of Signals. Because the traditional SVM-based method is used for classification, feature information of the raw data is drowned in noise, resulting in the unsatisfactory classification effect of SVM. In this paper, the raw signal is processed by the collaborative filter method and then reconstructing data with obvious characteristics.

The features of each group are displayed from the submerged noise, and the reconstructed obvious feature structure provides guarantee for the accuracy of the SVM classification.

Each group is composed of the m most similar to the original reference segment, so there is overlap between these m contrast segments. That is to say, a single point exists in multiple segments at the same time, so the reconstruction of the signal is to arithmetically average these m similar segments [24] to obtain the final output:

$$\text{Result}(i) = \frac{\sum_{i=1}^m \text{group}(D)_{i^*w}}{m}, \quad (11)$$

where $\text{group}(D)_{i^*w}$ is the row vector in the group.

The characteristic reconstructed blocking current signal is

$$\text{signal} = \text{result}\{\text{result}(1), \text{result}(2), \dots, \text{result}(n)\}. \quad (12)$$

3. SVM Classification and Recognition Based on Improved Feature Extraction

3.1. Experimental Data. The block current signals generated by the two single-stranded DNA molecules A_{14} and CA_3 to be recognized as they pass through the nanopore are shown in Figures 3 and 4, respectively.

The baseline current is 70.00 pA (both 800,000 sampling points).

3.2. Parameter Selection. This paper mainly uses signal-noise ratio and root mean squared error [25] as the evaluation criteria to determine the data based on cooperative filter and feature reconstruction and passes these two standards to determine the parameters that the collaborative filter algorithm needs to determine, that is, compare the moving step size k of the segmented segments with the number of sampling points included in each segment.

The definition of SNR used in this paper is

$$\text{SNR} = 10 \log\left(\frac{S}{N}\right) = 10 \log\left(\frac{\sum_{i=1}^N S^2(i)}{\sum_{i=1}^N (S_f(i) - S(i))^2}\right). \quad (13)$$

The definition of RMSE is

$$\text{RMSE} = \frac{\sum_{i=1}^N (S_f(i) - S(i))^2}{N}, \quad (14)$$

where $S(i)$ is the value of the initial input signal, $S_f(i)$ is the value of the output signal after collaborative filter, and N is the total length of the input signal. The larger the signal-to-noise ratio, the smaller the root mean square error, and the stronger the desiccation ability.

(a) Moving step of grouped fragments

It can be seen from the curve trend in Figure 5 that when the width of the segment is 50, the SNR is the largest, the RMSE is the smallest, and the denoising

effect is the best, so the moving step of the segment is 30.

(b) The width of the fragment

It can be seen from the curve trend in Figure 6 that when the moving step size of the fixed segment is 50, the SNR is the largest, the RMSE is the smallest, and the denoising effect is the best, so the width of the segment is 10.

Through the analysis of the above experimental data, it can be concluded that the collaborative filter algorithm has the best data processing effect when the moving step length of the segment is 30 and the width of the segment is 10. At this time, the SNR is 49.77 and the RMSE is 0.16.

3.3. Comparison of Filter Results. The parameters determined in Section 3.2 are segment moving step 30 and segment width 10. In order to highlight the performance of the algorithm, this section will compare it with the Bessel filter algorithm [26].

Figures 7 and 8 compare the data processing effect without and with improved threshold collaborative filter algorithm. Figure 7 shows the entire DNA molecule fragment, and Figure 8 shows a portion of the DNA molecule fragment. From Figures 7 and 8, it can be seen from the overall and partial filter results that the improved threshold collaborative filter algorithm (SNR = 49.77 and RMSE = 0.16) has a significantly better effect on data processing than the without improved threshold collaborative filter algorithm (SNR = 38.62 and RMSE = 0.89).

Figures 9 and 10 compare the data processing effect without improved threshold collaborative filter algorithm and Bessel algorithm. Figure 9 shows the entire DNA molecule fragment, and Figure 10 shows a portion of the DNA molecule fragment. From Figures 6 and 10, it can be seen from the overall and partial filter that the effect of the collaborative filter algorithm (SNR = 38.62 and RMSE = 0.89) on data processing is similar to that of Bessel filter (SNR = 38.62 and RMSE = 0.89).

Figures 11 and 12 compare the data processing effect of improved threshold collaborative filter algorithm and Bessel algorithm. Figure 11 shows the entire DNA molecule fragment, and Figure 12 shows a portion of the DNA molecule fragment. From Figures 11 and 12, it can be seen from the overall and partial filter that the effect of the improved threshold collaborative filter algorithm (SNR = 49.77 and RMSE = 0.16) on data processing is significantly better than that of Bessel filter on data processing (SNR = 39.71 and RMSE = 0.82).

From the above comparison results, it can be concluded that the denoising effect of the improved threshold collaborative filter algorithm is significantly better than without the improved threshold collaborative filter algorithm and Bessel Filter algorithm. Therefore, after the improved threshold collaborative filter algorithm is used to process the raw data, the characteristic information of the data is more obvious.

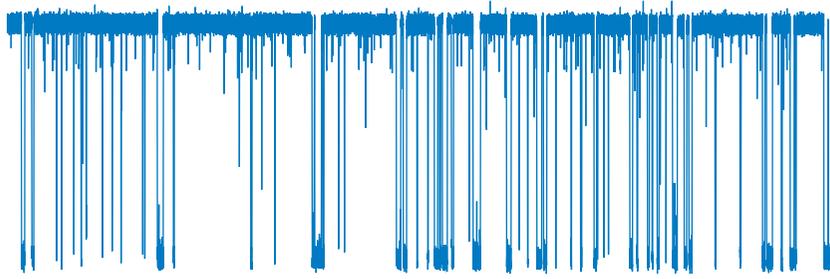
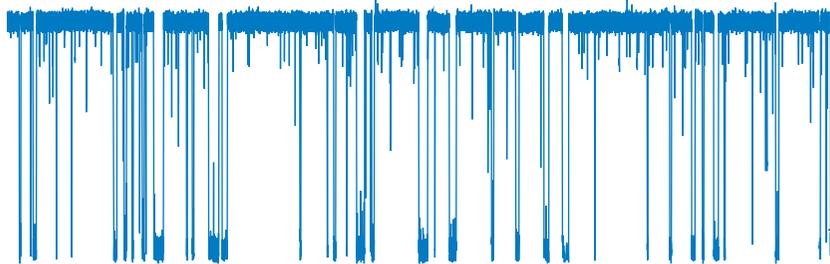
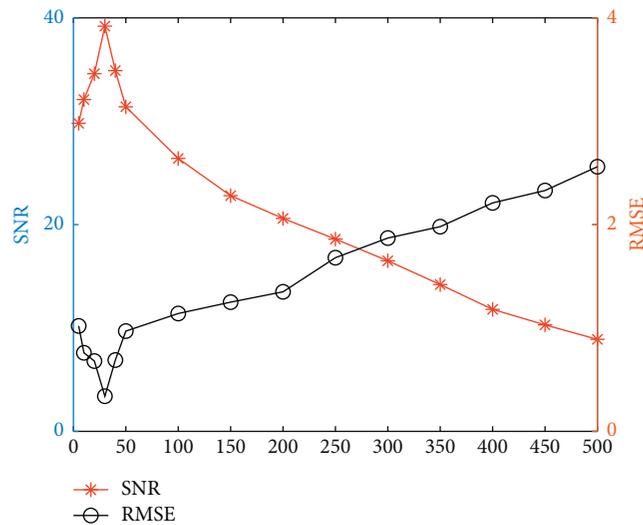
FIGURE 3: A_{14} raw data signal.FIGURE 4: CA_3 raw data signal.

FIGURE 5: Change chart of step size.

3.4. Comparison of Classification Results. In order to verify the effectiveness of the algorithm proposed in this paper, SVM classification algorithm [27] is used to classify and study the collaborative filter, Bessel Filter, and improved threshold collaborative filter.

The blocking current sampling points of CA_3 and A_{14} molecules are 15822 and 16628, respectively. 70% of the reconstructed datasets are used for training models and 30% for testing the effect of model recognition and classification.

Table 1 shows the classification accuracy of raw data, Bessel Filter data, collaborative filter data, and collaborative filter data with improved threshold using SVM. According to the classification accuracy in the table, it can be seen that the classification effect of collaborative filter without improved thresholds is similar to that of the Bessel filter by about 77%, while the classification effect of collaborative filter algorithm with improved thresholds is up to 95.88% better than the other two algorithms.

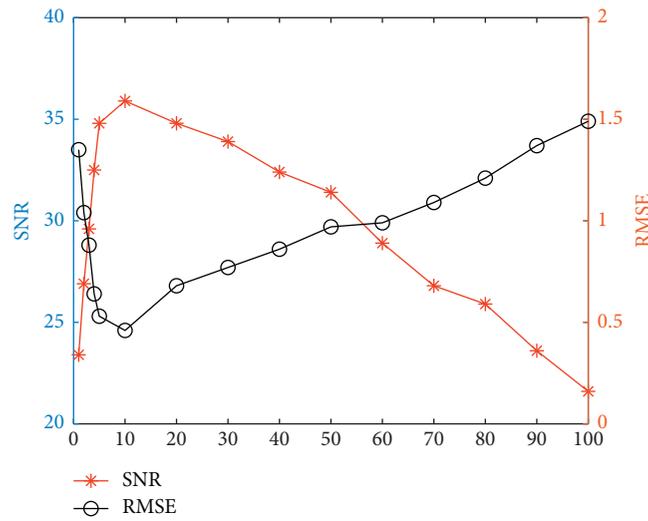


FIGURE 6: Change chart of width.

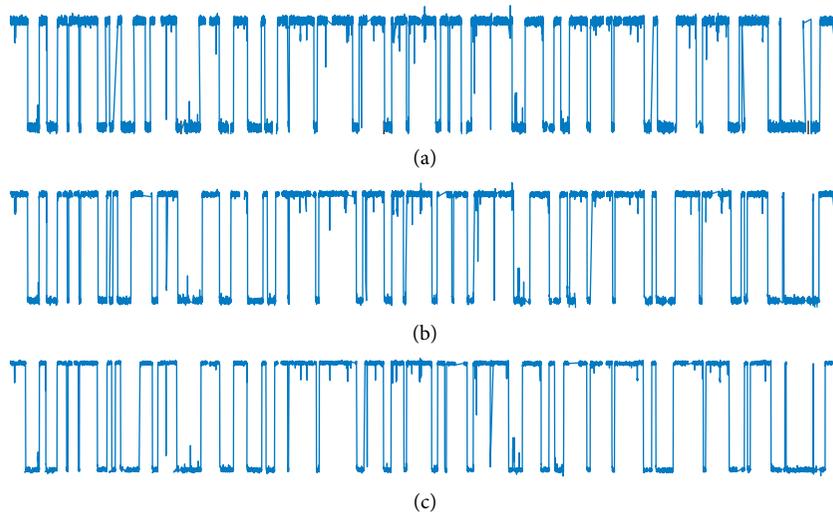


FIGURE 7: Comparison diagram with and without improved threshold collaborative filter: (a) A_{14} raw data; (b) collaborative filter; (c) improved threshold collaborative filter.

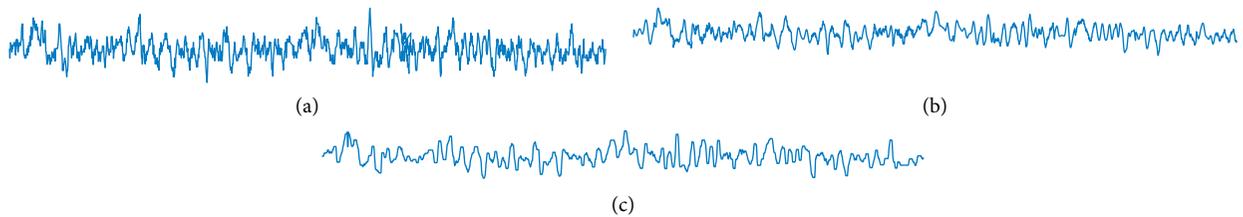


FIGURE 8: Partial diagram of collaborative filter with and without improved threshold collaborative filter: (a) A_{14} raw data; (b) collaborative filter; (c) improved threshold collaborative filter.

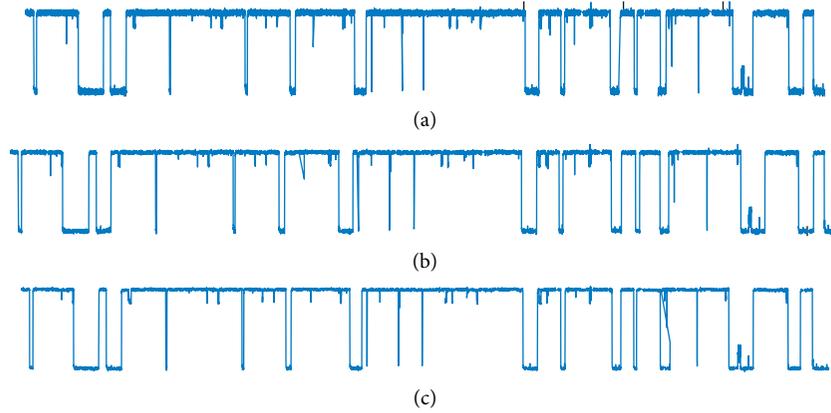


FIGURE 9: Comparison of Bessel and collaborative filters (A_{14} overall): (a) A_{14} raw data; (b) Bessel filter; (c) collaborative filter.

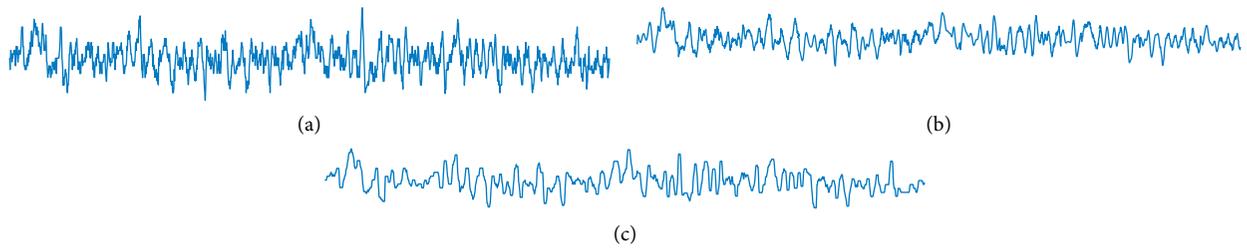


FIGURE 10: Comparison of Bessel and collaborative filters (A_{14} part): (a) A_{14} raw data; (b) Bessel filter; (c) collaborative filter.

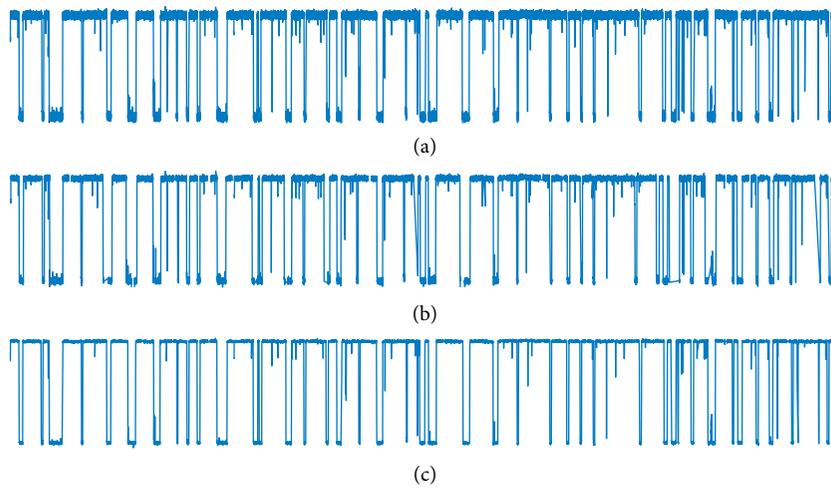


FIGURE 11: Comparison of Bessel and improved threshold collaborative filters (A_{14} overall): (a) A_{14} raw data; (b) Bessel filter; (c) improved threshold collaborative filter.

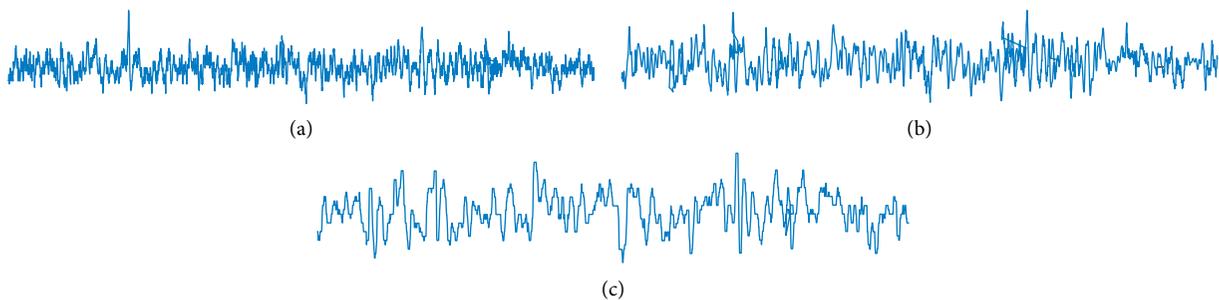


FIGURE 12: Comparison of Bessel and improved threshold collaborative filters (A_{14} part): (a) A_{14} raw data; (b) Bessel filter; (c) improved threshold collaborative filter.

TABLE 1: SVM classification results.

Filter algorithm	Classification accuracy (%)
Raw data	53.53
Bessel filter	77.05
Collaborative filter	76.38
Collaborative filter with improved threshold	95.88

4. Conclusion

Due to the large amount of environmental noise and the instrument's own noise mixed in the raw sampling data, it is difficult to obtain the feature information of the raw data only by SVM for classification, resulting in low classification accuracy. Therefore, in consideration of signal drift caused by various noises, based on the premise of grouping, thresholding, and reconstruction of the raw data based on the collaborative filter algorithm, this paper improves the threshold value selected during the thresholding process in the collaborative algorithm and introduces the threshold drift compensation factor to compensate for signal drift to compensate for the effects of noise.

Then, the raw data are processed using a collaborative filter algorithm with improved thresholds to obtain data groups with obvious feature information, and data groups with obvious feature information are used for data reconstruction. Then, the data processing effect of improved threshold collaborative filter is compared with the data processing effect of the unimproved collaborative filter and Bessel filter. The data processing effect of improved threshold collaborative filter is significantly better than the other two data processing methods.

Finally, the data processed by the three data processing methods are sent to SVM for training, and the classification accuracy of the data processed by the improved threshold collaborative filter algorithm is obviously better than the other two data processing methods.

Data Availability

The nanopore current data belong to the School of Chemical Engineering and Molecular Engineering of East China University of Science and Technology, which belongs to the school cooperative relationship. Since the School of Chemical Engineering and Molecular Engineering still needs to apply this dataset to other biological research, the experimental dataset of this paper is not public.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Major Scientific Research Instrument Development Project (no. 21327807), National Natural Science Youth Fund (no. 51407078), and National Natural Science Foundation of China (no. 61773165).

References

- [1] H. Su, M. Long, and Z. Zeng, "Controllability of two-time-scale discrete-time multiagent systems," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1440–1449, 2020.
- [2] H. Su, J. Zhang, and X. Chen, "A stochastic sampling mechanism for time-varying formation of multiagent systems with multiple leaders and communication delays," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3699–3707, 2019.
- [3] C. Cao, "Application of third generation sequencing technology to microbial research," *Microbiology*, vol. 43, no. 10, pp. 2269–2276, 2016.
- [4] S. Ambardar and M. Gowda, "High-resolution full-length HLA typing method using third generation (Pac-Bio SMRT) sequencing technology," *Methods in Molecular Biology*, vol. 1802, pp. 135–153, 2018.
- [5] M. Jain and M. Akeson, *High-Coverage Long Read DNA Sequencing with the Oxford Nanopore MinION*, UC Santa Cruz Electronic Theses and Dissertations, 2017.
- [6] X. Wang and H. Su, "Self-triggered leader-following consensus of multi-agent systems with input time delay," *Neurocomputing*, vol. 330, pp. 70–77, 2019.
- [7] P. Dixit and G. I. Prajapati, "Machine learning in bioinformatics: a novel approach for DNA sequencing," in *Proceedings of the 2015 Fifth International Conference on Advanced Computing & Communication Technologies (ACCT)*, pp. 41–47, Haryana, India, February 2015.
- [8] M. Balachandran, T. H. Shin, and L. Gwang, "PVP-SVM: sequence-based prediction of phage virion proteins using a support vector machine," *Frontiers in Microbiology*, vol. 9, p. 476, 2018.
- [9] Y. Zhao, B. Ashcroft, P. Zhang et al., "Single molecule spectroscopy of amino acids and peptides by recognition tunneling," *Nature Nanotechnology*, vol. 9, no. 6, pp. 466–473, 2014.
- [10] L. Zhong, J. T. L. Wang, D. Wen, and B. A. Shapiro, "Pre-miRNA classification via combinatorial feature mining and boosting," in *Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, PA, USA, October 2012.
- [11] Q. Zhou, Q. Jiang, and W. Dan, "A new method for classification in DNA sequence," in *Proceedings of the 2011 6th International Conference on Computer Science & Education (ICCSE)*, Singapore, August 2011.
- [12] M. Kumar, G. P. S. Gromiha, and G. P. S. Raghava, "SVM based prediction of RNA-binding proteins using binding residues and evolutionary information," *Journal of Molecular Recognition*, vol. 24, no. 2, pp. 303–313, 2011.
- [13] H.-L. Dai, "Imbalanced protein data classification using ensemble FTM-SVM," *IEEE Transactions on NanoBioscience*, vol. 14, no. 4, pp. 350–359, 2015.
- [14] V. Tabard-Cossa, M. D. Trivedi, and A. N. N. MarzaliJetha, "Noise analysis and reduction in solid-state nanopores," *Nanotechnology*, vol. 18, no. 30, Article ID 305505, 2007.
- [15] S. W. Kowalczyk, A. Y. Grosberg, and Y. C. Rabin, "Modeling the conductance and DNA blockade of solid-state nanopores," *Nanotechnology*, vol. 22, no. 31, Article ID 315101, 2011.
- [16] J. Dekker, W. B. K. Pedrotti, and W. B. Dunbar, "An area-efficient low-noise CMOS DNA detection sensor for multi-channel nanopore applications," *Sensors and Actuators B: Chemical*, vol. 176, pp. 1051–1055, 2013.
- [17] Y. Goto, I. Yanagi, and K. Matsui, "Integrated solid-state nanopore platform for nanopore fabrication via dielectric

- breakdown, DNA-speed deceleration and noise reduction,” *Scientific Reports*, vol. 6, no. 1, Article ID 31324, 2016.
- [18] Y. Liu and H. Su, “Containment control of second-order multi-agent systems via intermittent sampled position data communication,” *Applied Mathematics and Computation*, vol. 362, Article ID 124522, 2019.
- [19] Y. Liu and H. Su, “Some necessary and sufficient conditions for containment of second-order multi-agent systems with sampled position data,” *Neurocomputing*, vol. 378, pp. 228–237, 2020.
- [20] J. M. Smith, D. T. Lee, and J. S. Liebman, “An $O(n \log n)$ heuristic for steiner minimal tree problems on the euclidean metric,” *Networks*, vol. 11, no. 1, pp. 23–39, 2010.
- [21] K. Dabov, V. A. Foi, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [22] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [23] Howell D. C, Median Absolute Deviation, 2008.
- [24] Zhaoyi, “Digital filtering arithmetic average method and weighted average method,” *Instrumentation Technology*, no. 4, p. 41, 2001.
- [25] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature,” *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [26] P. T. Trinh, R. Brossier, L. Métivier, J. Virieux, and P. Wellington, “Bessel smoothing filter for spectral-element mesh,” *Geophysical Journal International*, vol. 209, no. 3, pp. 1489–1512, 2017.
- [27] G. M. Foody and A. Mathur, “Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification,” *Remote Sensing of Environment*, vol. 93, no. 1-2, pp. 107–117, 2004.