Eng Int Syst (2022) 1: 35–37

Engineering
Intelligent Systems

Massive Real-Time Data Mining Algorithm for a Multimedia Database

Jiaju Gong* and Qin Wu

Electronic and Information Technology, Jiangmen Polytechnic, Jiangmen 529030, China

The traditional classification mining algorithm for massive data has a complicated calculation process, poor real-time performance and low accuracy. This paper presents a real-time data mining algorithm for a multimedia database. The wavelet de-noising method is used to de-noise the data in the multimedia database to reduce the interference of the background area to the feature extraction of the multimedia data. The effective area extraction algorithm based on the center of mass is used to extract the effective area and is combined with the SIFT algorithm and LBP algorithm to extract the data features from the multimedia database. Experimental results show that the algorithm is accurate, reliable, has a high real-time mining ability and a practical ability.

Keywords: Multimedia; Database; Massive; Real-time; Data Classification; Mining

1. INTRODUCTION

With the rapid development of information technology, more and more multimedia data can now be obtained from the Internet, digital libraries and digital publications, and demand for information is moving towards diversification and integration. For multimedia databases, a large number of studies in the past have focused on the research of content-based information retrieval, and to a certain extent solved the problems of information search and information resources (Nguyen et al., 2015; Alfaro et al., 2016).

Isazadeh proposed an association rule mining algorithm based on similarity (Jiang et al., 2017). The multimedia data to be excavated was divided into several parts, and the clustering of data was completed according to differently colored histograms. At the same time, the final category labels were obtained. According to the previous time series, the multimedia data were arranged and replaced with the label of its group. Label sequences were obtained, and the data mining method was used to mine the relevant information. However, this method has low timeliness and accuracy (Om et al., 2019; Alizah et al., 2019; Sudarsan et al., 2019).

Aiming at the drawbacks of the above algorithm, a new algorithm for massive real-time data mining for a multimedia database is proposed. The proposed algorithm can effectively extract the features of multimedia data in a complex environment with less features, high accuracy and real-time performance.

2. MASSIVE REAL-TIME DATA CLASSIFICATION MINING ALGORITHM FOR A MULTIMEDIA DATABASE

2.1 Feature Extraction of Multimedia Data

2.1.1 LBP Algorithm for Feature Extraction

LBP (Local Binary Pattern) is an operator used to describe the local texture features of images. It has the obvious advantages

*Corresponding Author e-mail: goden2019@163.com

Figure 1 The original data set.

of rotation invariance and gray invariance. The LBP algorithm takes the gray value of the central pixel in the multimedia data as a threshold and compares it with the gray value of the adjacent pixel. The local texture feature is described by the obtained binary code (Isazadeh et al., 2016; Bianchi et al., 2017).

Assuming that $x_k = i + \lambda_1, y_k = j + \lambda_2$, wherein, i and j are non-negative integers, and λ_1, λ_1 are floating-point numbers in the range of [0, 1], the pixel value $l(i+\lambda_1, j+\lambda_2)$ of this point can be determined by several pixel values corresponding to the original co-ordinate, and the equation is described as follows:

$$\begin{aligned} l(i+\lambda_1, j+\lambda_2) &= (1-\lambda_1)(1-\lambda_2)l(i, j) \\ &+ (1-\lambda_1)\lambda_2 l(i, j+1) + \lambda_1(1-\lambda_2)l(i+1, j) \\ &+ \lambda_1\lambda_2 l(i+1, j+1) \end{aligned} \tag{1}$$

Where, $l(i, j)$ is used to describe the pixel value of original multimedia data point (i, j).

Assuming that the gray value of center pixel in the multimedia data is l_c and the gray values of q sampling points are l respectively, the LBP feature value near the center pixel can be obtained by the following equation:

$$P_{q,r} = \sum_{i=1}^{q} 2^{i-1} W(l_i - l_c) \tag{2}$$

In which,

$$W(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

2.2 Optimization of The Algorithm

The multimedia data is complex and the features of one algorithm are not enough to extract all the necessary information. The SIFT algorithm has poor performance on the feature extraction of the multimedia data with light invariance, while the LBP algorithm has light invariance and the gridding feature extraction, which can effectively make up for the shortcomings of the SIFT algorithm (Mencía et al., 2016).

2.3 Classification Mining With Cascade AdaBoost Classifier

This section refers to the open source cascade AdaBoost algorithm in the OpenCV library for the massive real-time data classification mining of a multimedia database. This algorithm is the most mainstream algorithm in the field of massive real-time data classification mining for a multimedia database (Chen et al., 2016).

It is assumed that the cascade classifier is composed of strong classifiers of n layers, and the false detection rate is usually $Z = \prod_{i=1}^{n} z_i$. Among them, z_i is the inherent false detection rate of the i-th strong classifier. Another important parameter of the cascade classifier is that the pass detection rate is $B = \prod_{i=1}^{n} b_i$. Among them, b_i is inherent pass detection rate of the i-th strong classifier.

3. EXPERIMENT ANALYSIS

In order to validate the effectiveness of the algorithm, a set of 2D multimedia data with elevation values is used for validation. The selected spatial data set is described in Figure 1, which contains 1200 data points. The proposed algorithm, association rules mining algorithm and Gaussian mixture algorithm are utilized to mine the multimedia data described in Figure 1, and the obtained results are described in Figure 2.

Analysis of Figure 2 shows that the results of the association rule mining algorithm significantly reduce the processing of multimedia data points, significantly reducing the time complexity. The mining results of the Gaussian mixture mining algorithm are similar to that before mining, and the effect is not significant. However, the proposed algorithm reduces the processing of meaningless points so that the result of mining is closer to the linear structure, which shows that the classification mining result of the proposed algorithm is accurate and validates the effectiveness of the proposed algorithm.

4. CONCLUSION

This paper focuses on the massive real-time data classification mining for a multimedia database to extract the features of multimedia data. The comparison experiment shows that

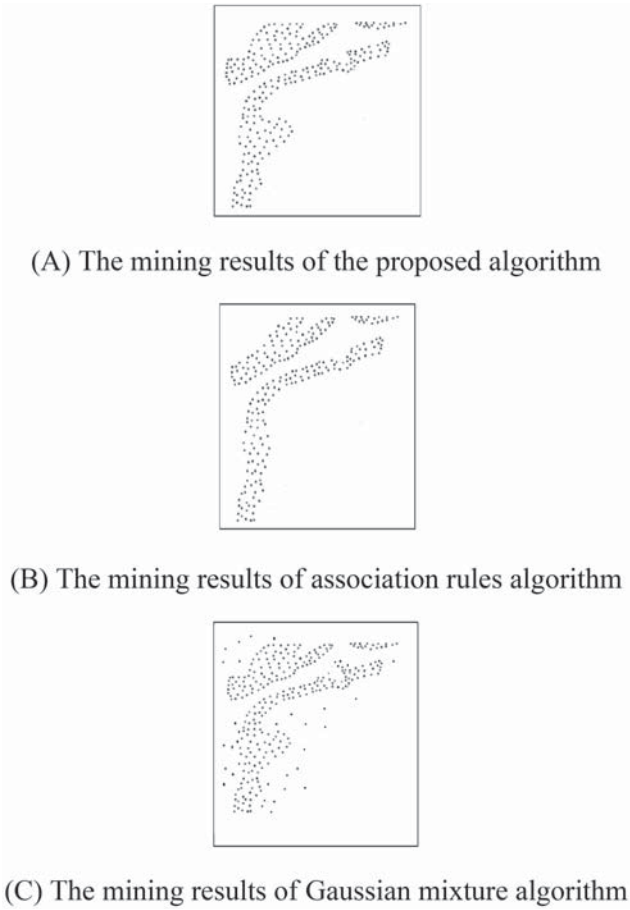

Figure 2 The mining results of three kinds of algorithms.

the research method is more accurate and effective than the traditional method, which provides a favorable basis for related research. On the basis of this, the classification mining of multimedia data is realized by cascade classifier, which improves the shortcomings of the traditional algorithm in real-time and accuracy. Innovative, valuable results are collected, but this is only a small part of the results that can be achieved by the multimedia data mining. The future of multimedia databases will be combined with a variety of disciplines and a variety of information carriers such as the emerging Local-Based Service (LBS), mobile device-based services, sensor networks, etc., resulting in many new scenarios.

ACKNOWLEDGEMENT

Research and application of one belt, one road background and Chinese educational gameplay mechanism (No.: 2018jc02023).

REFERENCES

1. Alfaro, C., Cano-Montero, J., Gómez, J., et al. A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, 2016, 1: 197–213.
2. Alizah A., Tan L. L., Ng K. T., Noraini I., Siti Zarikh S. A. B. (2019). Transforming Public Libraries into Digital Knowledge Dissemination Centre in Supporting Lifelong Blended Learning Programmes for Rural Youths. *Acta Informatica Malaysia*, 3(1): 16–20.
3. Bianchi, F.M., Maiorino, E., Livi, L., et al. An agent-based algorithm exploiting multiple local dissimilarities for clusters mining and knowledge discovery. *Soft Computing*, 2017, 5: 1347–1369.
4. Chen, Z.Y., Fan, Z.P., Sun, M. A multi-kernel support tensor machine for classification with multi-type multi-way data and an application to cross-selling recommendations. *European Journal of Operational Research*, 2016, 1: 110–120.
5. Isazadeh, A., Mahan, F., Pedrycz, W. MFlexDT: Multi-flexible fuzzy decision tree for data stream classification. *Soft Computing*, 2016, 9: 3719–3733.
6. Jiang, C., Liu, Y., Ding, Y., et al. Capturing helpful reviews from social media for product quality improvement: A multi-class classification approach. *International Journal of Production Research*, 2017, 55: 3528–3541.
7. Mencía, E.L., Janssen, F. Learning rules for multi-label classification: A stacking and a separate-and-conquer approach. *Machine Learning*, 2016, 1: 1–50.
8. Nguyen, T., Nahavandi, S., Creighton, D., et al. Mass spectrometry cancer data classification using wavelets and genetic algorithm. *Febs Letters*, 2015, 24: 3879–3886.
9. Om P. S., Gaurav K., Mukesh K. (2019). Role of Taguchi and Grey Relational Method in Optimization of Machining Parameters of Different Materials: A Review. *Acta Electronica Malaysia*, 3(1): 19–22.
10. Sudarsan B., Neepa B., Kartick C. M. (2019). Parallel and Distributed Association Rule Mining Algorithms: A Recent Survey. *Information Management and Computer Science*, 2(1): 15–24.