Journal of Network Intelligence ©2024 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **9**, Number **4**, November **2024**

Intelligent Classification of Large-scale Book Information Based on Extreme Learning Machine

Peng Li
Shandong University Library, Jinan 250100, P. R. China
jnlp2023@163.com

Kun Liu*
Shandong University Library, Jinan 250100, P. R. China
liukun@sdu.edu.cn

Huan Jiang
College of Sciences
Trinity University of Asia, 275 E. Rodriguez Sr. Blvd., Manila, Philippines
ja4904@163.com

*Corresponding author: Kun Liu
Received January 11, 2024, revised April 22, 2024, accepted July 29, 2024.

Abstract. *Focusing on the problems of low classification precision and poor generalization of existing classification algorithms when dealing with large-scale book data, an intelligent classification algorithm on the ground of Extreme Learning Machine (ELM) for large-scale book information is studied. Firstly, the book information dataset is processed in batches to reduce the data dimension and the data complexity, and at the same time, a multi-layer autoencoder is used for unsupervised coding of the data to realize the deep feature extraction, so as to achieve the purpose of optimizing the traditional extreme learning machine algorithm. Based on the above optimized ELM algorithm, this paper introduces a new stream regularization idea into the ELM, which can not only maintain the geometric structure of the input book features, but also correct the distance between the sample points through the book category information, so as to realize the intelligent classification of large-scale book information when mapping the book features. The experimental outcome indicates that the accuracy, precision and recall of the algorithm offered in this article are 92.49%, 93.51% and 91.72%, respectively, with good categorization performance.*
Keywords: Extreme learning machine; book classification; unsupervised coding; stream form regularization; batch processing

1. **Introduction.** With the advent of the knowledge economy era, the book publishing industry continues to increase the circulation of books with each passing day, and the demand for book purchases also grows rapidly [1, 2, 3], resulting in tens of thousands of books in the library's collection, which is huge in number. The manual classification of book cataloging in the work of libraries is time-consuming and laborious, and if the book classification work is carried out manually, it is not only more complex, costly and inefficient, but also the accuracy can not be absolutely guaranteed. As artificial intelligence continuously developing, the employment of computer-based machine learning technology to realize a variety of automated tasks has become the mainstream and trend in various fields, which can greatly improve the efficiency of traditional manual operations [4,

5]. The use of advanced artificial intelligence technology to solve the informatization and classification automation of library resources has become the research focus and primary task of library expert system technology [6].

1.1. **Related Work.** The current research on automatic book classification is mainly to introduce the results achieved by machine learning in the field of text classification into the field of book classification. Maron and Kuhns [7] proposed to represent the original text with index words, and determine the categories according to the probability of the index words appearing in the classes, and used this method for book information retrieval. Salton et al. [8] suggested Vector Space Model (VSM) and used it to represent the text. Andreasen and Grove [9] proposed a factor analysis method to classify books and periodicals. Drucker et al. [10] proposed a simple linear classifier based on self-correcting class weight vectors with user feedback. Melucci [11] discussed the technical details of the implementation of text categorization and tested it on their own dataset. Meng et al. [12] proposed a tree-like structure for text classification and improved the classification accuracy. Cortes and Vapnik [13] proposed Support Vector Machine (SVM) on the ground of statistics, and the SVM algorithm is an optimization problem to find the convex quadratic programming. Later, Mikolo et al. [14] proposed word vector model word2vec based on deep learning. Cichosz [15] proposed Glove (Global Vectors for Word Representation, Glove) which takes into account the global information of words while preserving the local information of the words by using the co-occurrence matrix of the words. Moon et al. [16] utilized the Generative Pre-Training Word Vector Model GPT (Generative Pre-Training) to find the optimum of a convex quadratic programming problem and encode book information, but the second half of the information was lost. Gonzalez et al. [17] utilized the bidirectional tranformer as an encoder to fully consider the information in the front and back directions of the word. However, when the data size of the book information is large, the above machine learning algorithm has the problem of insufficient feature representation in the classification process.

In addition to the above machine learning algorithms, as a class of Extreme Learning Machine (ELM) [18, 19, 20] on the ground of the structure of single obscured level feed-forward neural network, which has quick solution speed and excellent generalization capability, it shows excellent performance in classification tasks. Xin et al. [21] applied Regularized Extreme Learning Machine (RELM) to the problem of book information classification. Lei et al. [22] proposed a book classification model on the ground of ELM, but ELM simply superimposes the loss of the model during the back propagation process. Chen et al. [23] offered a new feature selection algorithm combined with Extreme Learning Machine to classify book information, but it could not achieve accurate classification. Although ELM has achieved some good classification results in text classification, ELM presents a nonlinear geometric structure by randomly mapping the input text features into the feature space of ELM in the process of mapping, which affects the classification performance of book information.

1.2. **Motivation and contribution.** Aiming at the shortcomings of the existing book information classification algorithms, this paper researches a large-scale book information intelligent classification method on the ground of ELM.

(1) The traditional ELM algorithm is optimized, and a multilayer autoencoder is used to encode the data without supervision to realize the deep feature extraction, and the new output weight formula is derived to construct a streaming classifier with inherited factors to realize the model solving.

(2) Based on the above optimized ELM algorithm, we introduce a new streamform regularization idea into the ELM, which can correct the distance between sample points

through the category information of books when mapping book features, and realize intelligent classification of large-scale book information. The experimental outcome indicates that the method designed in this article is with higher classification performance and efficiency compared with the comparison algorithm.

2. Theoretical analysis.

2.1. **Extreme learning machine.** Extreme Learning Machine (ELM) is a Single-hidden Layer Feed-forward Neural (SLFN) network structure, consisting of stimulus layer, obscured level and output layer, which can avoid the shortcomings of traditional neural networks such as slow learning and easier to fall into the local minimum [24]. ELM's structure is shown in Figure 1.

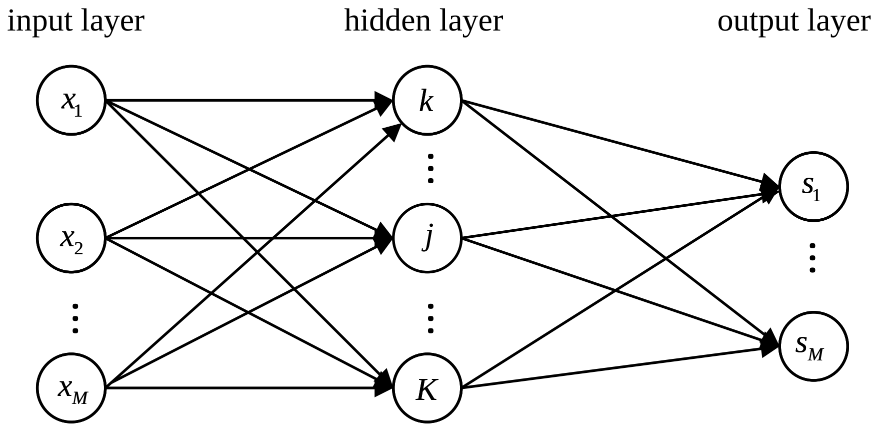

Figure 1. The structure of ELM network

Set amount of samples $\{(x_i, s_i)\}_{i=1}^{M}$ to be M, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}]^T \in \mathbb{R}^m$ is the m-dimensional stimulus value of the i-th sample and $s_i \in \mathbb{R}$ is its output value. Assuming that the amount of nodes in the intermediate obscured (hidden) layer is K and the activation operation is $g(x)$, the result of the ELM is explicited as follows.

$$f_i = \sum_{j=1}^{K} \alpha_j g\left(\vec{v}_j \cdot \vec{x}_i + a_j\right), \quad j = 1, 2, \ldots, M \tag{1}$$

where $\vec{v}_j = [v_{j1}, v_{j2}, \ldots, v_{jm}]$ is the weight among the stimulus node and the j-th implicit level node, α_j is the weight among the output node and the j-th implicit level node, and a_j is the corresponding threshold value of the implicit level node.

In order to obtain the prime network parameters $\vec{v}$ and α such that the error among the network output and the true value tends to zero, i.e., $\sum_{i=1}^{M} f_i - s_i \approx 0$, there is:

$$\sum_{j=1}^{K} \alpha_j g\left(\vec{v}_j \cdot \vec{x}_i + a_j\right) = s_i, \quad i = 1, 2, \ldots, M \tag{2}$$

Its matrix form can be expressed as $W\alpha = S$.

$$W(\vec{v}_1, \ldots, \vec{v}_K, a_1, \ldots, a_K, \vec{x}_1, \ldots, \vec{x}_M) = \begin{bmatrix} g\left(\vec{v}_1 \cdot \vec{x}_1 + a_1\right) & \ldots & g\left(\vec{v}_K \cdot \vec{x}_1 + a_K\right) \\ \vdots & & \vdots \\ g\left(\vec{v}_1 \cdot \vec{x}_M + a_1\right) & \ldots & g\left(\vec{v}_K \cdot \vec{x}_M + a_K\right) \end{bmatrix}_{M \times K} \tag{3}$$

where W is the implicit layer output value, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_K]^T$, $S = [s_1, s_2, \ldots, s_M]^T$.

The output weight matrix is obtained through the computation of the least squares solution:

$$\|W\hat{\alpha} - S\| = \min_{\alpha} \|W\alpha - S\| \tag{4}$$

Finally, the least squares solution $\hat{\alpha}$ is obtained as follows:

$$\hat{\alpha} = W^{\dagger} S \tag{5}$$

where $W^{\dagger}$ is the Moore-Penrose generalized inverse of W.

2.2. **Streamform regularization.** Streamform regularization mainly utilizes the idea of streamform assumption [24], which is a technology applied in the field of data stream processing, aiming at improving and standardizing the throughput and performance when processing data streams. If the sample points x_p and x_q are close, then their labels y_p and y_q should also be close to each other. Thus, the objective function of streamform regularization can be expressed as follows:

$$\text{Min} : \frac{1}{2} \sum_{pq} \|v_{pq} y_p - y_q\|_2^2 \tag{6}$$

where v_{pq} represents the similarity between samples x_p and x_q. When calculating the similarity between the sample points, we first need to adopt the K-nearest neighbor way to discover the k points that are closest to the sample points as their neighbor nodes. When calculating the distance between the sample points, a supervised distance measure is used, i.e., a distance calculation method that takes into account the sample category information, which is shown as follows:

$$D(x_p, x_q) = \begin{cases} \sqrt{1 - e^{-\frac{d^2(x_p, x_q)}{\beta}}}, & x_p = x_q \\ \sqrt{e^{-\frac{d^2(x_p, x_q)}{\beta}}} - \vartheta, & x_p \neq x_q \end{cases} \tag{7}$$

where β and ϑ are specified constants; $d(x_p, x_q)$ is the Euclidean distance ignoring the category information of the sample points, i.e., $d(x_p, x_q) = |x_p - x_q|$; $D(x_p, x_q)$ are the distances after combining the category information of the sample points.

Parameter β prevents the rapid increase of $d(x_p, x_q)$ in the exponential function, and parameter β has a more significant moderating effect when $d(x_p, x_q)$ is larger. The parameter ϑ $0 \leq \vartheta \leq 1$ is a constant factor that regulates the distance.

Then select the k samples that are closest to sample point x_p as their neighbor nodes and calculate the similarity v_{pq} between sample point x_p and its neighbor node x_q. When calculating the similarity, a Gaussian kernel function is used to calculate the similarity, and only the nodes that belong to the same category and are neighbors to each other are calculated. The similarity is shown as follows:

$$v_{pq} = \begin{cases} \exp\left(-\frac{\|x_p - x_q\|_2^2}{\mu}\right), & e(x_p, x_q) = 1 \text{ and } C(x_p) = C(x_q) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where μ is the similarity factor, $e(x_p, x_q) = 1$ denotes that samples x_p and x_q are neighboring nodes to each other, $C(x_p) = C(x_q)$ denotes that samples x_p and x_q belong to the same category. The similarity matrix V between all the samples is obtained and V is computed as a symmetric matrix. The equation is shown below.

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1M} \\ \vdots & \ddots & \vdots \\ v_{M1} & \cdots & v_{MM} \end{bmatrix}_{M \times M} \tag{9}$$

3. **Optimization of extreme learning machine algorithm based on large scale data.** To address the issue of large-scale data, such as high memory energy consumption, low classification accuracy and poor generalization, we propose to optimize the traditional ELM algorithm. Firstly, the data set is processed in batches to reduce the data dimension and input complexity; then the multilayer auto-coder structure is adopted to unsupervised coding of each batch of data to achieve deep feature extraction; finally, the stream regularization idea is used to construct a stream classifier with inheritance factors to maintain the integrity of the data and to enhance the generalization performance of the algorithm.

Let N training samples (x_l, s_l), $l = 1, 2, \ldots, M$, where x_l is the l-th sample and s_l is the object vector of x_l.

(1) Data batching. Calculation of the output weight α_{l+1} of the current sample data relies on the output weight α_l of the previous batch of samples, while the update of the obscured level output matrix W of every batch of sample data is independent of each other. Therefore, we first parallelize W, then establish the functional relationship between batches, and then serially update α to enhance the inheritance relationship between sample points and reduce the size of input data.

(2) Unsupervised encoder. To achieve effective input features, sparse constraints are added to the optimization objective. Different from the traditional ELM, in this paper, l_1-paradigm penalty is used instead of l_2-paradigm penalty in the optimization objective function, in which the l_1-paradigm based auto-coding training process is shown as follows:

$$\begin{cases} L_\alpha = \arg\min B_j\alpha_j - W_j^2 + \mu\alpha_i \\ B_j = \vartheta(W_j \cdot b_j + \sigma_j) \end{cases} \tag{10}$$

where $\|\cdot\|_1$ denotes the l_1-paradigm; μ is a nonnegative constant; B_j is the j-th layer autoencoder random hidden layer output; b_j and σ_j are the orthogonal random weights and bias, separately.

(3) Stream Regularization. To precisely reflect the flow relationship between sample data and solve the issue of insufficient learning of ELM for large-scale datasets, this paper introduces the constraint of stream regularization, which preserves the topological relationship between data points in the new feature space, and obtains the optimal mapping through optimization $\sum_{i,j}^{M} \|x_i - x_j\|^2 V$, i.e., there exists:

$$\sum_{i,j}^{M} \|x_i - x_j\|^2 V = \sum_{i,j}^{M} (x_i x_i^T - 2x_i x_j^T + x_j x_j^T) V_c = 2\,\mathrm{tr}\left(X(D_c - V_c)X^T\right) = 2\,\mathrm{tr}\left(K_c X^T\right) \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of the matrix; $D_c = \sum_{i,j}^{M}(i, j)$ is the diagonal matrix; K_c is a symmetric semipositive definite matrix; and V_c is the matrix of edge weights between the sample points x_i and x_j.

(4) Batch supervised classifier. When solving the generalized eigenvalue problem of ELM, for the purpose of guaranteeing the completeness of the data, this article introduces the inheritance factor, and constructs the batch inheritance learning paradigm by establishing the output weight relationship of each batch, then the optimization function is expressed as follows:

$$\min L = \frac{1}{2}\|\alpha\|^2 + \frac{D}{2}\sum_{i=1}^{M}\epsilon_i^2 + \frac{\lambda}{2}\|\alpha - \nu\|^2 \quad \text{s.t.} \quad w(x_i)\alpha = s_i - \epsilon_i \tag{12}$$

where D is the regularized least-mean-square parameter, ϵ_i is the output error, α is the weight vector, and ν is the initial output weight obtained from the first batch of data; λ is a successor factor to regulate the ratio of output weights between batches; and $i = 1, 2, \dots, M$.

To further enhance the training speed and processing efficiency, in the classification decision stage, the original ELM classifier is replaced by a popular classifier containing inheritance factors as the output, and the convex differentiable objective function can be constructed as follows with reference to Equation (12).

$$\min \frac{1}{2}\|\alpha\|^2 + \frac{D}{2}\|S - W\alpha\|^2 + \frac{\lambda}{2}\|\alpha - h\|^2 + \frac{\lambda}{2}\operatorname{tr}(\alpha^T W^T K W \alpha) \tag{13}$$

4. Research on intelligent classification of large-scale book information based on extreme learning machine.

4.1. Book information representation. Aiming at the problem that the current book information classification algorithm cannot realize the accurate classification of large-scale data, based on the above optimized ELM algorithm, this chapter introduces a new idea of streamform regularization into the ELM, and intelligently classifies the large-scale book information through the streamform regularization ELM for the purpose of enhancing the performance of book classification.

The book information classification model based on stream regularization ELM consists of three parts: (1) book information representation: using subject words as the features of book information dataset and representing the information dataset by vector space model; (2) feature extraction: adopting Singular Value Decomposition (SVD) to perform feature extraction on book features, the original high level book features are extracted into the book information dataset. feature extraction, mapping original high-dimensional features to their low-dimensional subspace; (3) book information classification: using stream regularization ELM to learn book classification model from large-scale book information; specific process is implied in Figure 2.

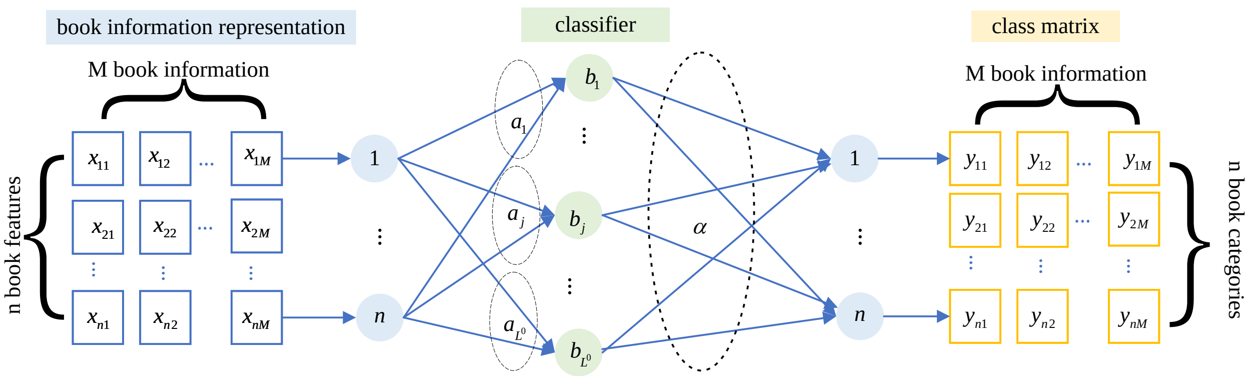

Figure 2. The flow chart of classification algorithm is designed in this paper

Book information dataset $Z = \{(x_i, y_i), i = 1, 2, \dots, M\}$ has M books, x_i and y_i represent the i-th book and its category respectively, and book x_i can be denoted as $(s_1, s_2, \dots, s_j)^T$, where s_l represents the l-th feature word in the dataset. In this paper, the feature words in books are represented by TF-IDF weights [25] to reflect the semantic differences between different books in the dataset. The term frequency of feature word s_l in x_i is shown in Equation (14).

$$tf(s_l, x_i) = \frac{\text{num}(s_l)}{\text{count}(x_i)} \tag{14}$$

where $\text{num}(s_l)$ denotes the number of times s_l shows in the book and $\text{count}(x_i)$ expresses the amount of all topic words in x_i. The inverse document frequency indicates the importance of the feature word in that information dataset. The inverse document frequency $\text{idf}(s_l)$ of feature word s_l is denoted as follows.

$$\text{idf}(s_l) = \log \frac{M}{\text{df}(s_l)} \tag{15}$$

where M denotes the amount of texts in the book information dataset, and $\text{df}(s_l)$ is the amount of texts containing the feature word s_l. TF-IDF is used as the product of item frequency and inverse text frequency as indicated in Equation (16).

$$\text{tfidf}(s_l, x_i) = tf(s_l, x_i) \times \text{idf}(s_l) \tag{16}$$

To eliminate the effect of book document length on the TF-IDF weights, a normalization operation is performed.

$$x_{il} = \frac{\text{tfidf}(s_l, x_i)}{\sqrt{\sum_i \text{tfidf}(s_l, x_i)^2}} \tag{17}$$

where x_{il} represents the TF-IDF weight of the l-th feature word in the i-th book after normalization. Then book x_i can be denoted as vector $[x_{i1}, x_{i2}, \ldots, x_{il}]^T \in R^l$. Multiple documents in the book dataset will form a two-dimensional matrix X. The rows and columns of the two-dimensional matrix represent the amount of documents and the whole amount of feature words in the book information dataset, separately.

4.2. **Book information feature extraction.** The book information data still has high dimensionality after the above processing, so this paper adopts the singular value decomposition [26] to extract the book features in order to reduce the book features. Firstly, the singular value decomposition of book information matrix X is carried out, i.e.:$X = V\Sigma U^T$ where V and U represent the two orthogonal matrices of book matrix X, respectively. The n-rank optimal approximation matrix of book information matrix X can be expressed as $X_n = V_n \times \Sigma_n \times U_n^T$, where V_n denotes the first n columns of matrix V, U_n^T denotes the first n rows corresponding to matrix U^T, and $\Sigma_n = diag(\delta_1, \delta_2, \ldots, \delta_n)$ denotes the diagonal matrix containing the first n largest singular values. x_i is the text matrix after feature extraction, which contains the most important n textual features of book matrix X. Similarly, the i-th book is represented after singular value decomposition as follows.

$$\hat{x}_k = x_i^T V_n \Sigma_n^{-1} \tag{18}$$

The features of book x_i are reduced from 1 to n. Next, the book matrix $X = (x_1, x_2, \ldots, x_M)^T$ after the dimensionality reduction process is used as an input to the streaming regularized extreme learning machine for training the classifier.

4.3. **An intelligent classification method for large-scale book information based on extreme learning machine.** On the ground of the above analysis and optimized extreme learning machine algorithm, the following mathematical description of the large-scale book information dataset after the above processing is carried out: book information dataset $Z = \{(x_i, y_i), i = 1, 2, \ldots, M\}$ has N books (x_i, y_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ represents the i-th book information after preprocessing, x_{in} represents the TF-IDF weight of the n-th feature word in the information of the i-th book, n is the number of feature words after dimensionality reduction processing, $y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^T \in R^m$ represents the category to which the i-th book belongs, m denotes the amount of categories in book dataset Z. The amount of nodes in the obscured level L_0 and the activation operation $g(x)$ of the streaming regularized ELM are set first, and then the parameters of the streaming regularized ELM are randomly set: weights $a_j \in R^n$, biases $b_j \in R^m$, $j = 1, 2, \ldots, L_0$. The streaming regularized ELM model is shown as follows:

$$\sum_{j}^{L_0} \alpha_j g(x_i) = \sum_{j}^{L_0} \alpha_j g(a_j x_i + b_j) = y_i \tag{19}$$

where $a_j = [a_{j1}, a_{j2}, \ldots, a_{jn}]^T$, $j = 1, 2, \ldots, L_0$, denote the weight vectors among the input layer and the j-th obscured level node of the stream regularized extreme learning machine, b_j denotes the bias of the j-th obscured level node, and $\alpha_j = [\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jm}]^T$ denotes the weight vector between the j-th obscured level node and the output layer. After mathematical transformation, the ELM model can also be explicited in the form of a matrix: $W\alpha = Y$. Where W is the output matrix of the hidden layer, expressed as follows.

$$W = \begin{bmatrix} g(a_1, b_1, x_1) & g(a_2, b_2, x_1) & \cdots & g(a_{L_0}, b_{L^0}, x_1) \\ g(a_1, b_1, x_2) & g(a_2, b_2, x_2) & \cdots & g(a_{L^0}, b_{L^0}, x_2) \\ \vdots & \vdots & \ddots & \vdots \\ g(a_1, b_1, x_M) & g(a_2, b_2, x_M) & \cdots & g(a_{L^0}, b_{L^0}, x_M) \end{bmatrix}_{M \times L^0} \tag{20}$$

where α is the weight matrix between the obscured and output levels.

$$\alpha = [\alpha_1^T \ldots \alpha_{L_0}^T]^{L_0 \times m} \tag{21}$$

To increase the classification accuracy, the training error (empirical risk) and the quadratic norm of the output matrix (structural risk) are minimized at the same time to obtain the weight matrix between the hidden layer and the output layer, which can be obtained as Equation (22).

$$\min ||W\alpha - Y||_2^2 + \frac{D_1}{2}||\alpha||_2^2 \tag{22}$$

where D_1 is the regularization factor and $||\cdot||_2^2$ denotes the biparadigm number. To address the issue of nonlinear mapping of book information features in the training process of ELM by randomly setting the weights a_j among the stimulus level and the obscured level and the bias b_j of the nodes in the obscured level, this article adopts the objective operation Equation (6) of the stream regularization idea into the regularized limit learning machine, so as to make it select the neighboring points that can maintain the original structure and make the input book information maintain its original geometric structure in the process of mapping.

$$\min \frac{1}{2}||W\alpha - Y||_2^2 + \frac{D_1}{2}||\alpha||_2^2 + \frac{D_2}{2}\,\mathrm{tr}(Y^T L Y) \tag{23}$$

where D_1 is the regularization factor of the ELM to regulate the empirical and structural risks in the extreme learning machine, and D_2 is the regularization factor in the flow regularization idea. By derivation of the objective function Equation (23), the following equation is achieved.

$$W^T(W\alpha - Y) + D_1\alpha + D_2W^TLW\alpha = 0 \tag{24}$$

The matrix α among the obscured and output levels of the streaming regularized limit learning machine is calculated through the least squares method as follows.

$$\alpha = (W^TW + D_1I + D_2W^TLW)^{-1}W^TY \tag{25}$$

5. **Performance testing and analysis.**

5.1. **Analysis of downscaling results.**

To estimate the performance of the method designed in this article, 27808 titles of 10 categories are randomly selected from a university library, containing 25391 feature words, 14192 training set samples and 5759 test set samples. The HRELM [21], BPNNB [27], ADTR [28] and the classification algorithm OURS designed in this paper are trained and tested respectively. OURS algorithm is set with the amount of nodes in the obscured layer of 9,000, the regularization factor D_1 of 1,000, and the regularization factor D_2 of 0.01. The experimental platform configured with Windows 10, 64-bit operating system, Core(TM) i5-10500 CPU @ 3.10GHz, 8 GB RAM, and the simulation experiment environment is Matlab R2022a.

For the purpose of verifying the effectiveness of the stream regularization strategy in the large-scale book information classification problem of extreme learning machine, OURS algorithm and traditional ELM classification algorithm are firstly compared on a university library dataset. Table 1 implies the accuracy of the OURS algorithm and the traditional ELM classification algorithm for each category. It also indicates that the accuracy of OURS algorithm is better than that of ELM in all categories, with an average accuracy of 92.42%, while the average accuracy of ELM is 83.84%, and the average accuracy of OURS in all categories is significantly higher than that of ELM.

Table 1. The accuracy of OURS and ELM on a university library dataset

Classes	**ELM**	**OURS**	**Classes**	**ELM**	**OURS**
life	87.19%	93.27%	drama	78.39%	91.12%
music	79.32%	91.64%	language	84.16%	95.29%
math	81.52%	94.37%	sfr	88.19%	90.61%
project	89.13%	93.16%	stype-metal	82.74%	92.15%
skate	83.58%	92.39%	tapioca	84.19%	90.18%

Table 2 gives the classification accuracy, precision, recall and F1 metric of OURS and traditional ELM algorithms. From the table, it can be seen that the accuracy rate of OURS is 92.49%, precision rate is 93.51%, recall rate is 91.72% and F1 metric is 92.61%, while the accuracy rate of ELM is 83.74%, precision rate is 84.19%, recall rate is 81.05% and F1 metric is 82.59%. are all significantly higher than ELM.

Table 2. The classification performance of OURS and ELM

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-measure**
OURS	92.49%	93.51%	91.72%	92.61%
ELM	83.74%	84.19%	81.05%	82.59%

Through the above outcome, we can see that the classification performance of OURS achieves the best results, which indicates that OURS has better classification performance and stability of classification. Compared with traditional ELM classification method, the classification performance of OURS is significantly improved after the introduction of streamform regularization in the ELM. It proves the previous hypothesis that streamform regularization can make the input book information features maintain the original book feature structure after mapping. Moreover, the new stream regularization idea prioritizes the sample points with the same category in the process of constructing the Laplace matrix, especially when calculating the distance between samples is the same, OURS will prioritize the proximity points with the same category, which can further improve the classification performance.

5.2. **Comparative experiments on classification results of multiple algorithms.** Under different training sizes, the results of the four experiments of HRELM classification, BPNNB classification, ADTR classification and OURS classification are compared, and the comparison results are implied in Figure 3. It can be seen that the classification accuracy of the four classification algorithms increases with the increase in the amount of training data. When the training data size reaches 1300, the classification accuracy of all four classification algorithms is above 80%. However, under the condition of the same training data size, the accuracy of OURS classification algorithm has the highest mean value of 90.2%, followed by HRELM classification algorithm, and BPNNB classification and ADTR classification have the worst accuracy.

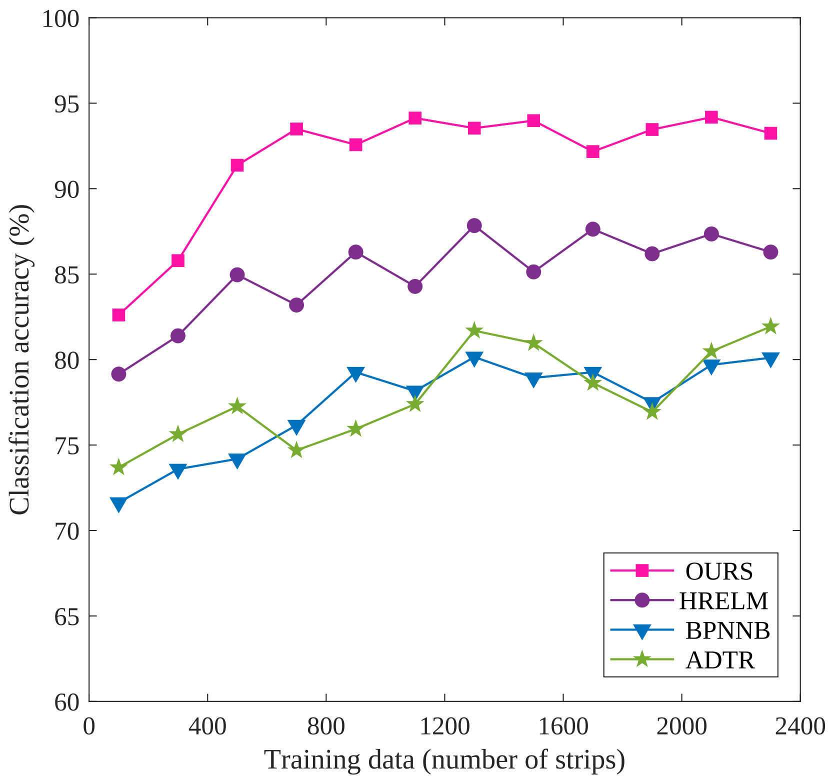

Figure 3. Comparison of classification accuracy of different classification algorithms

Figure 4 indicates the accuracy, precision, recall and F1 metric of OURS and the other three algorithms, and the F1 metric, as a comprehensive evaluation criterion combining precision and recall, can reflect the performance of the algorithms in a comprehensive way. From the figure, it can be seen that the classification performance of OURS algorithm is better than that of HRELM classification, BPNNB classification and ADTR

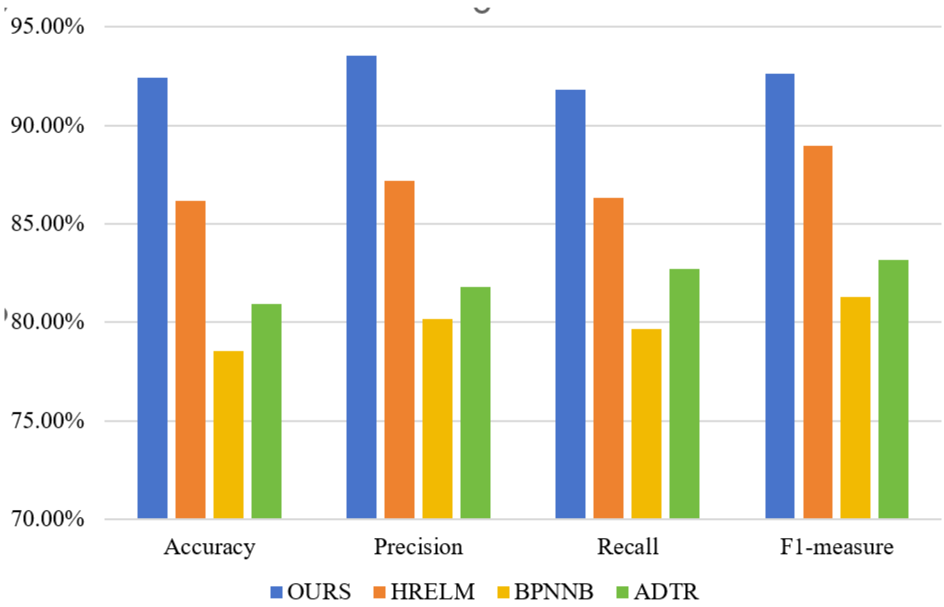

Figure 4. The accuracy, precision, recall and F1 metric of different algorithms

classification, which further illustrates the effectiveness of the stream regularization idea introduced in the extreme learning machine. Moreover, after the introduction of stream regularization, in the process of constructing the Laplace matrix of OURS, the category information of the books can correct the Euclidean distance between the sample points, which further improves the classification performance of OURS through the category information. Although HRELM algorithm introduces the limit learning machine into book data classification, it does not optimize the RELM, which makes the classification effect poor, while BPNNB algorithm and ADTR algorithm are classification algorithms constructed on the basis of BP neural network and decision tree, respectively, which have the disadvantages of slow transmission learning and easier to fall into the local minima, and thus the classification effect is the worst.

6. **Conclusion.** Aiming at the issue of low accuracy of existing book information classification algorithms, this paper studies a large-scale intelligent classification method of book information based on ELM. Firstly, the book information data set is processed in batches, and the data is unsupervised coded by multilayer autoencoder to realize the optimization of traditional ELM algorithm. Then, based on the optimized ELM algorithm, this paper introduces a new stream regularization idea into the extreme learning machine, which can not only maintain the geometric structure of the input information features, but also correct the distance between the sample points by the category information of the book, so as to realize the intelligent classification of large-scale book information. The experimental outcome indicates that the designed method has higher classification performance and efficiency compared with the comparative algorithms.

REFERENCES

[1] D. Cullen and A. Bell, "The Social Text and Networked Knowledge: New Modes of Scholarly Book Publishing in Indigenous Studies," *Journal of Scholarly Publishing*, vol. 49, no. 2, pp. 193-212, 2018.
[2] A. Martin, S. Bartscher-Finzer, and A. Richards, "The Motivations of Book Publishers in Germany and Australia: An Empirical Comparison," *Publishing Research Quarterly*, vol. 34, pp. 69-88, 2018.

[3] M. P. Satija and D. Martínez-Ávila, "Features, Functions and Components of a Library Classification System in the LIS tradition for the e-Environment," *Journal of Information Science Theory and Practice*, vol. 3, no. 4, pp. 62-77, 2015.
[4] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, no. 1, pp. 3-24, 2007.
[5] A. A. Soofi and A. A. Awan, "Classification techniques in machine learning: applications and issues," *Journal of Basic & Applied Sciences*, vol. 13, no. 1, pp. 459-465, 2017.
[6] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 273-292, 2019.
[7] M. E. Maron and J. L. Kuhns, "Probabilistic Indexing and Information Retrieval," *Journal of the ACM*, vol. 7, no. 3, pp. 216-244, 1960.
[8] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613-620, 1975.
[9] N. C. Andreasen and W. M. Grove, "The classification of depression: traditional versus mathematical approaches," *The American Journal of Psychiatry*, vol. 139, no. 1, pp. 45-52, 1982.
[10] H. Drucker, B. Shahrary, and D. C. Gibbon, "Support vector machines: relevance feedback and information retrieval," *Information Processing & Management*, vol. 38, no. 3, pp. 305-323, 2002.
[11] M. Melucci, "A basis for information retrieval in context," *ACM Transactions on Information Systems*, vol. 26, no. 3, pp. 1-41, 2008.
[12] J. Meng, H. Lin, and Y. Yu, "A two-stage feature selection method for text categorization," *Computers & Mathematics with Applications*, vol. 62, no. 7, pp. 2793-2800, 2011.
[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111-3119, 2013.
[15] P. Cichosz, "A case study in text mining of discussion forum posts: classification with bag of words and global vectors," *International Journal of Applied Mathematics and Computer Science*, vol. 28, no. 4, pp. 787-801, 2018.
[16] J. H. Moon, H. Lee, W. Shin, Y.-H. Kim, and E. Choi, "Multi-modal understanding and generation for medical images and text via vision-language pre-training," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 6070-6080, 2022.
[17] J. A. Gonzalez, L.-F. Hurtado, and F. Pla, "TWilBert: Pre-trained deep bidirectional transformers for Spanish Twitter," *Neurocomputing*, vol. 426, pp. 58-69, 2021.
[18] L. Oneto, F. Bisio, E. Cambria, and D. Anguita, "Statistical learning theory and ELM for big social data analysis," *IEEE Computational Intelligence Magazine*, vol. 11, no. 3, pp. 45-55, 2016.
[19] Y. Zhao, G. Wang, Y. Yin, Y. Li, and Z. Wang, "Improving ELM-based microarray data classification by diversified sequence features selection," *Neural Computing and Applications*, vol. 27, no. 1, pp. 155-166, 2016.
[20] R. C. Deo, M. K. Tiwari, J. F. Adamowski, and J. M. Quilty, "Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model," *Stochastic Environmental Research and Risk Assessment*, vol. 31, pp. 1211-1240, 2017.
[21] J. Xin, Z. Wang, L. Qu, and G. Wang, "Elastic extreme learning machine for big data classification," *Neurocomputing*, vol. 149, pp. 464-471, 2015.
[22] Y. Lei, L. Cen, X. Chen, and Y. Xie, "A hybrid regularization semi-supervised extreme learning machine method and its application," *IEEE Access*, vol. 7, pp. 30102-30111, 2019.
[23] Y. Chen, X. Xie, T. Zhang, J. Bai, and M. Hou, "A deep residual compensation extreme learning machine and applications," *Journal of Forecasting*, vol. 39, no. 6, pp. 986-999, 2020.
[24] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121-2133, 2022.
[25] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, 40, 2019.
[26] E. K. Wang, S. P. Xu, C.-M. Chen, and N. Kumar, "Neural-Architecture-Search-Based Multiobjective Cognitive Automation System," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2918-2925, 2021.
[27] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Review*, vol. 35, no. 4, pp. 551-566, 1993.

[28] I. P. B. D. Purwanta, C. K. Adi, and N. P. N. P. Dewi, "Backpropagation Neural Network for Book Classification Using the Image Cover," *International Journal of Applied Sciences and Smart Technologies*, vol. 2, no. 2, pp. 179-196, 2020.

[29] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130, 2015.