



An algorithm for categorizing online course resources for vocal music in the context of MOOC teaching and learning

Yuyang Sun^{1,*}

¹ School of Arts, Taishan University, Taian, Shandong, 271000, China

SUMMARY: *The swift increase in the use of MOOC learning, alongside the ever-increasing amount of instructional materials, has posed more difficulty to effectively reach useful information especially when it comes to the quality of teaching. Considering this, the present research presents a classification algorithm applicable to online vocal music course resources. This approach will improve the process of categorizing such materials through extracting important features and using a decision tree classifier. The experimental analysis was done on a vocal music course resources data set, which showed that the approach produces a classification error score of lower than 82 percent and a low F1 error score of less than 81 percent, indicative of uniform results. The method demonstrates an increase in classification accuracy compared to other approaches (4.18 to 6.45). Moreover, the algorithm also scored high in two other datasets. The results of the present study indicate that the suggested methodology is very successful in the process of classifying online vocal music educational materials, which can be greatly beneficial in terms of resource management in online vocal music classes.*

KEYWORDS: *decision tree; feature extraction; classification algorithm; vocal music teaching; course resources*

1 Introduction

Massive Open Online Courses (MOOCs) have played a significant role in higher education since they were introduced into the world [1]. MOOCs are distinguished by numerous resources, extensive availability, high enrollment levels of learners, openness, interactivity and substantial learning independence [2, 3]. Under the framework of the Internet+Education, MOOC-assisted curriculum reform in colleges and universities has slowly been implemented. Coursera has been used by many Chinese institutions which have built upon the benefits of MOOC-based instruction and created curriculum teaching models based on the use of MOOC materials [4-6]. As far as MOOC teaching models are concerned, literature [7] has analyzed MOOC-assisted educational program to assess its instructional efficiency. The research revealed that the model provided valuable learning experiences to participants by conducting a survey of the students involved. Literature [8] systematically described the MOOC teaching model and discussed its suitability in physical education courses by reviewing relevant literature. It was found out that professional theoretical knowledge, skills and movement techniques in physical education could be achieved by students not necessarily in traditional classrooms but also in online learning where similar results can be obtained. Literature [9] evaluated papers on MOOC teaching and learning and pointed to a number of problems in the

*jiaoxuefaZY@163.com

<https://doi.org/10.65102/is2026475>

current research, such as providing too simplistic an explanation of learner-related factors and lack of precision in measuring learning outcomes. Literature [10] suggested a novel teaching model combining traditional English teaching and MOOC teaching resources and demonstrated that this model would be effective in improving classroom performance, enhancing the learning enthusiasm of students and increasing their understanding of autonomous learning.

Educational instruction in vocal music is one of the main focuses of higher education, and with MOOCs, it gives students the chance to discover the vast world of vocal music culture. Such students have access to local specialists and global specialists in vocal music instruction who provide them with information on the topic through their presentations[11-13]. Research [14] talks about incorporating MOOC-based instruction into the curriculum of preschool education majoring in vocal music education in universities. This integration will help make vocal music teaching more creative and effective, meeting the specific needs of such students. One other study [15] analyzes the possibility of using the MOOC model to enhance the effectiveness of teaching, introduce innovative methods of teaching, and lead to the reform of music classes at the university level. The study examines the effects of MOOC-based teaching based on theory and practice. Moreover, literature [16] suggests a hybrid teaching approach to vocal music courses, which is underpinned by a digital resource management system that is explicitly intended to allow delivering vocal music education via MOOCs. Nevertheless, there are still drawbacks, including the restricted accessibility of online materials, wrong intelligent tips, and the inability to combine different types of content. These problems also lead to poor completion rates of courses in online vocal music education. To overcome these barriers, it is essential to use classification algorithms to more efficiently structure and control the online vocal music sources [17-19].

The main aim of the vocal online course resource classification algorithm is to divide the vocal resources into separate categories. It allows personalization of teaching materials that are recommended to users, which eventually increases the quality of learning[20, 21]. The concept of personalized recommendation has become popular and is easily incorporated into the environment of MOOC teaching. Nevertheless, to properly identify suitable music to suit the needs of an individual among a large volume of accessible resources, the personalized music recommendation system should consider the behavior of the user to filter and recommend suitable vocal resources. This procedure guarantees that the learner's requirements will be addressed with customized instruction to achieve better results in online vocal music instruction through technical assistance [22-25]. Literature [26] discussed one of the possible strategies of classifying vocal online course resources by using the MOOC teaching model. The video classification algorithm was developed as a multi-channel convolutional network that leverages deep metric learning and it is shown that this model has extremely high classification rates of vocal online course resources.

An adaptive sliding window mutual information approach is used to select features based on vocal teaching materials to develop a classification algorithm of vocal online course resources. The technique will minimize the dimensionality of the high-dimensional data related to the vocal online course resources by reducing them to low-dimensional space using similar transformation rules. The next one is the application of the decision tree algorithm to form the decision tree classifier. Then these features are passed through such classifiers to finish the process of classification. In order to test the performance of the proposed algorithm, its classification rate and F1 score are measured against a control algorithm. Experimental results are obtained on both RCV1-v2 and AAPD datasets, and the generalization ability of the algorithm to classify vocal music course resources is tested by analyzing Hamming loss, accuracy, recall, and F1 score.

2 Algorithm for categorizing online course resources for vocal music

Effective use of digital teaching materials can help in building an intuitive understanding of courses during lectures, with such advantages as the expansion of the viewpoints of students and the enhancement of their critical thinking skills. The paper is a decision tree-based algorithm that is developed in the framework of MOOC-based education, the purpose of which is to classify online vocal music course resources.

2.1 Course resource feature extraction

As digital vocal teaching materials are increasingly utilized and the number of teaching hours expands, the volume of data related to vocal resources continues to grow. When extracting features from such digital resources, it is essential to account for both the new and historical data to offer a comprehensive solution. This approach is crucial because it ensures that no significant information, which might be hidden within the resources, is overlooked. To achieve this, the adaptive sliding window mutual information method is employed to process both the historical and incremental data, enabling effective feature extraction from online vocal music teaching materials.

The original window data is represented by matrix $X_1 = [x_1, x_2, \dots, x_m]$, while the incremental window data is represented by matrix $X_2 = [x_{m+1}, x_{m+2}, \dots, x_{m+r}]$. The complete dataset from the vocal online course teaching materials is denoted by $X = [X_1, X_2]$. The mutual information matrices for the original and new window data are represented by Z_1 and Z_2 , respectively. Z represents the mutual information matrix for all samples of the vocal online course teaching resources.

In line with the principles of mutual information, the mutual information matrix can be expressed as follows:

$$Z = \frac{1}{m+r} (Z_1 + Z_2) \quad (1)$$

The process of diagonalization employs a unit array to express the formula Z_1 for feature decomposition, which is as follows:

$$I = G_1^T Z_1 G_1 \quad (2)$$

By applying the space defined by G_1 to project onto Z_2 , the following equation is derived:

$$\bar{Z}_2 = G_1^T Z_2 G_1 \quad (3)$$

Combining the equations from Eq. (1) and Eq. (2), we get:

$$G_1^T (Z_1 + Z_2) G_1 = I + \bar{Z} \quad (4)$$

The feature decomposition \bar{Z}_2 formula is then given by:

$$\overline{Z_2} = P_2 \Lambda_2 P_2^T \quad (5)$$

Substituting Eq. (5) into Eq. (4) yields the following expression:

$$P_2^T G_1^T (Z_1 + Z_2) G_1 P_2 = I + \Lambda_2 \quad (6)$$

Such a procedure makes it possible to acquire the outcomes of the feature decomposition of any vocal online course teaching materials.

From Eq. (2), we can infer:

$$G_1 = B_1 \Lambda_1 - \frac{1}{2} \quad (7)$$

In Eq. (7), $\Lambda_1 \in R^{k \times k}$ and $B_1 \in R^{n \times k}$ represent the matrices formed by the first k eigenvalues and the principal component decision matrix of the original vocal music online course resources, respectively.

Through the above process, obtain the eigenvalues Λ_2 and eigenvectors P_2 of the mutual information matrix of the new window data, and $\Lambda_2 = [\mu_1, \mu_2, \dots, \mu_k]$, $P_2 = [\beta_1, \beta_2, \dots, \beta_k]$.

With the aid of these eigenvectors and eigenvalues, the formula to derive the eigenvalues of all vocal online course materials is given below:

$$\Lambda = \frac{1}{m+r} (I + \mu_i) \quad (8)$$

In Eq. (8), m refers to the sample data of historical vocal music online course teaching resources, and r denotes the sample data of new vocal music online course teaching resources.

The formula of the feature vector of vocal music online course teaching materials is given below:

$$P = G_1 \beta_i \quad (9)$$

A principal component decision matrix is formed by using the derived feature vectors. Online vocal music course resources can be mapped on to this matrix and the reduced dimensionality of data results. This step is replicated in subsequent windows to extract features of all vocal music teaching resources.

2.2 Decision Tree Algorithm

Decision tree is one of the basic techniques that may be applied to classification and regression. Its task is to minimize the loss function in the training stage to construct a model, which in turn is employed to predict new data. A popular algorithm used in decision trees is the Classification and Regression Tree (CART) model, which performs both classification and regression. It can offer a conditional probability distribution of a random variable Y given an input random variable X . In the CART method, the decision tree is created as a binary tree in which every internal node has two possible results: yes and no. The left branch represents the values that are categorized as yes, and the right branch represents the values

that are labeled as no. The CART algorithm has two major steps, i.e. tree generation and tree pruning.

2.2.1 Decision Tree Generation

(1) Generation of regression tree

The key process in creating a regression tree involves recursively dividing the independent variable space. As the feature space, or the space of independent variables, is segmented into multiple regions, which are labeled as M , R_1, R_2, \dots, R_M , and so forth, this division enables the model to discover different trends in every region, and each cell R_m corresponds to a specific output value c_m , the regression tree model can be represented in the following

manner:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (10)$$

After setting up the feature space partitioning method, it is possible to find the best output value of every region through the application of the minimization of the square error principle. Where the squared error $\sigma^2 = \sum_{x_i \in R_m} (y_i - f(x_i))^2$. It can be shown that the optimal value c_m

on the cell R_m is the mean of the output y_i corresponding to all input instances x_i on R_m , i.e:

$$c_m = \text{ave}(y_i | x_i \in R_m) \quad (11)$$

Next, the optimal attribute needs to be selected, i.e., the feature space is partitioned. Specifically, given the j th variable x_j and its value s , the feature space can be partitioned into two regions, which can be denoted as respectively:

$$R_1(j, s) = \{x | x_j \leq s\} \text{ And } R_2(j, s) = \{x | x_j > s\} \quad (12)$$

Seek a solution:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (13)$$

Then the corresponding optimal cut-off point s can be matched for a given input variable j such that it satisfies the following equation:

$$c_1 = \text{ave}(y_i | x_i \in R_1(j, s)) \text{ And } c_2 = \text{ave}(y_i | x_i \in R_2(j, s)) \quad (14)$$

Following the steps outlined in the previous paragraph, apply them to all the input variables to find the best splitting variable \hat{j} that creates a pair (\hat{j}, s) and divide the feature space into two parts on an iterative basis. The same is repeated in every region until the predetermined ending conditions are reached.

(2) Generation of classification tree

In contrast to the regression tree, the classification tree uses the Gini index to find the best attributes. It is also useful in determining the ideal binary split point of the variable. To simplify things with regard to classification, suppose there are K classes and the probability that a given sample point is in the k th class is indicated by p_k . This probability distribution can then be used to compute the Gini index as shown below:

$$Gini(p) = \sum_{k=1}^K p_k(1-p_k) = 1 - \sum_{k=1}^K p_k^2 \quad (15)$$

In the case of a binary classification problem, when the probability of a sample being part of one of the classes is p , the Gini index of the probability distribution of the sample can be written as:

$$Gini(p) = 2p(1-p) \quad (16)$$

For a specific sample set, the Gini index can be defined as follows:

$$Gini(D) = 1 - \sum_{k=1}^K \left(\frac{|C_k|}{|D|} \right)^2 \quad (17)$$

where K is the total number of all categories and C_k is the subset of the sample set D consisting of all samples belonging to the k th category. If the set of samples D is divided into two classes according to whether the feature A takes a certain possible value a or not, viz:

$$D_1 = \{(x, y) \in D \mid A(x) = a\}, D_2 = D - D_1 \quad (18)$$

Then the Gini index of the set D is defined under the condition of characteristic A :

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (19)$$

Gini index $Gini(D)$ measures the uncertainty present in sample set D whereas the Gini index $Gini(D, A)$ evaluates the remaining uncertainty of set D which is caused by the division of sample set D according to the values of attribute A . The larger the Gini index, the more uncertain the sample set is.

2.2.2 Pruning decision trees

The process of decision tree pruning helps to reduce overfitting and improve model generalization by eliminating branches that have a small predictive value. The CART pruning algorithm consists of two separate steps. This starts with the leaf nodes and advances towards the root node T_0 , with a series of subtrees $\{T_0, T_1, \dots, T_n\}$ being produced on the way. The next step is to consider each candidate subtree in this sequence and evaluate it using an independent validation sample based on cross-validation to decide what the best level of pruning is.

(1) Subtree Sequence Generation

The first step is to compute the loss function of each subtree:

$$C_\alpha(T) = C(T) + \alpha |T| \quad (20)$$

where T is an arbitrary subtree, $C(T)$ is the error generated by the prediction of the training set, $|T|$ is the number of leaf nodes in the subtree, and $\alpha \geq 0$ is a parameter that can be used to measure the fit of the training set against the complexity of the model. $C_\alpha(T)$ is the overall loss of subtree T with parameter α . For a fixed parameter α , there must exist a subtree that minimizes the loss function $C_\alpha(T)$, denoted as T_α . T_α is optimal in the case where the loss function $C_\alpha(T)$ is minimized, and such an optimal subtree is unique. The optimal subtree T_α is small in the case where α is large and large in the case where α is small. The tree can be pruned recursively, i.e., α is increasing from 0 such that $0 = \alpha_0 < \alpha_1 < \dots < \alpha_n < +\infty$, producing a series of intervals $[\alpha_i, \alpha_{i+1}), i = 0, 1, \dots, n$, the sequence of subtrees obtained by pruning is the same as the interval $\alpha \in [\alpha_i, \alpha_{i+1}), i = 0, 1, \dots, n$ of the sequence of optimal subtrees $\{T_0, T_1, \dots, T_n\}$ corresponds.

Specifically, pruning starts from the overall tree T_0 . For any internal node t of T_0 , the loss function with t as the single node tree is:

$$C_\alpha(t) = C(t) + \alpha \quad (21)$$

The loss function of the subtree T_t with t as the root node is:

$$C_\alpha(T_t) = C(T_t) + \alpha |T_t| \quad (22)$$

When $\alpha = 0$ and α are sufficiently small, there are inequalities:

$$C_\alpha(T_t) < C_\alpha(t) \quad (23)$$

When α increases, there is at a certain α :

$$C_\alpha(T_t) = C_\alpha(t) \quad (24)$$

When α increases again, there is $C_\alpha(T_t) > C_\alpha(t)$. As long as $\alpha = \frac{C(t) - C(T_t)}{|T_t| - 1}$, T_t and t have the same value of the loss function and t has fewer nodes, so t is preferable to T_t , and T_t should be pruned.

For this purpose, for each internal node t in T_0 , compute:

$$g(t) = \frac{C(t) - C(T_t)}{|T_t| - 1} \quad (25)$$

$g(t)$ denotes the degree of reduction of the overall loss function after pruning. The T_t with the smallest $g(t)$ in T_0 is pruned and the resulting subtree is used as T_1 , while the

smallest $g(t)$ is set to α_1 . The T_1 is the optimal subtree for the interval $[\alpha_1, \alpha_2)$. This pruning continues until the root node is obtained. During this process, the value of α is continuously increased to produce new intervals.

(2) Selecting the optimal subtree from the obtained sequence of subtrees

Use the validation set to test the subtree sequence $\{T_0, T_1, \dots, T_n\}$ for the squared error (regression tree) or Gini index (classification tree) of each subtree, where the optimal decision tree is the one with the smallest squared error or Gini index. Since each subtree T_0, T_1, \dots, T_n has its corresponding parameters $\alpha_1, \alpha_2, \dots, \alpha_n$. Therefore, after determining the optimal subtree T_k , the corresponding parameter α_k is also determined, and thus the optimal decision tree T_α is obtained.

2.3 Decision Tree Classifier Construction

Decision tree classifiers are built on features derived out of teaching resource materials that are accessed online and based on online vocal music courses. To construct a decision tree classifier, it starts by examining the values of features of online teaching resource samples and then the different types and groups of teaching resources are recognized and sorted into decision tree structures. Roots as well as leaves in this tree represent teaching resources of vocal music online course. After grouping the leaf nodes, an aggregation algorithm is used to assess them and it is measured in accordance with this particular formula:

$$d_{ij} = \sqrt{\sum_{i=1}^m (x_i - x_j)} \quad (26)$$

where: d_{ij} is the leaf node distance from i to j , m is the historical data of teaching resources of vocal music online course, x_i and x_j are the sample characteristic parameters when the number of teaching resource samples is i and j , respectively.

Let us assume that all parameters are independent when it comes to sorting and storing teaching resources. Two classes are therefore defined as t and j and their distance is computed as:

$$d(t, s) = x_t + x_s \quad (27)$$

where: $d(t, s)$ is the leaf node distance between class t to class s , x_t is the teaching resource sample of class t , and x_s is the teaching resource sample of class s .

The teaching resource examples of online vocal music lessons are paired with their respective location of the domains based on leaf lengths and this mapping serves as the foundation of the decision tree classifiers. The bits of domains are subsequently allocated to the page nodes of the decision-tree of classification so that every leaf node has the correct properties of the online vocal music course teaching resources.

The results of classification are then derived using the Block Coding (BTC) algorithm. This algorithm allows predicting the initial count of classifications and deriving the final classification function. Its exact form is:

$$BTC(R) = -2 \sum_{i=1}^j d_i + m_i \log(N) \quad (28)$$

where: N is the sum of the number of observed parameters, m_i is the number of characteristic leaf nodes, d_i is the distance of the nodes of unclassified resource classes, and R is the number of aggregates.

The classification function gives results on the classification of vocal music online course teaching resources which is the end of the whole classification procedure performed by the decision tree algorithm.

3 Testing and analysis

3.1 Methods of comparison

In analyzing the application effect of the decision tree-based resource classification method for vocal music online courses designed in this paper, a comparison test was carried out. Among them, the classification methods used in the control group were: a resource classification method based on a two-step clustering algorithm, a resource classification method based on an all-encompassing instructional design (UDL) perspective, and a resource classification search method based on a Top-k query algorithm, which are denoted as CM1, CM2, and CM3, respectively.

3.2 Test data

For the test data, this paper takes the resource library built by a vocal music teaching platform itself as the source of test data. In the resource library, it contains the basic resource library and the expansion resource library of the vocal music online course.

(1) The basic resource library of vocal music online course. It contains the following sub-databases: S1.Literature database, including basic knowledge of music, musical notation, text and diagrams.S2.Media database, including all kinds of concerts, recitals, concerts, musicals and other vocal and instrumental audio-visual materials, including five categories of text materials, graphic (image) materials, audio materials, video materials and animation materials.S3.Teaching regulations database, including the disciplinary Teaching regulations database, including subject syllabus, teaching assessment methods and documents. s4. Teaching examples database, including music teachers' teaching cases, excellent lesson plans, music teaching strategies, classroom teaching video recordings, students' excellent assignments and lectures, etc. s5. Vocal Music Teaching Reference Database, including software, courseware, examination question banks and other multimedia materials for vocal music teaching.

(2) Vocal Music Online Course Expansion Resource Library. The content contains the following sub-databases: S6.Music Education Figures Database, containing literature of music masters, representative figures of music genres, vocal composers, performers, and music educators.S7.Music Topics Database, scientific research data on various related topics, doctoral (master's) dissertations, research results of vocal education, important literature, books, etc. S8.Vocal Music Education Comprehensive Database, books, newspapers, electronic journals, patent literature and other data resources. S9. Vocal Music Education Policy Database, including policies, rules, regulations and rules and regulations on music education.

It contains a total of 22,458 online course resources of 9 categories of vocal music, and is

characterized by uneven distribution. In this paper, 2000 of these resources are used as the training set and 1000 resources are used as the test set, and the selection of various types of teaching resources is shown in Figure 1. After preparing the test data according to Fig. 1, four methods are used for classification test respectively.

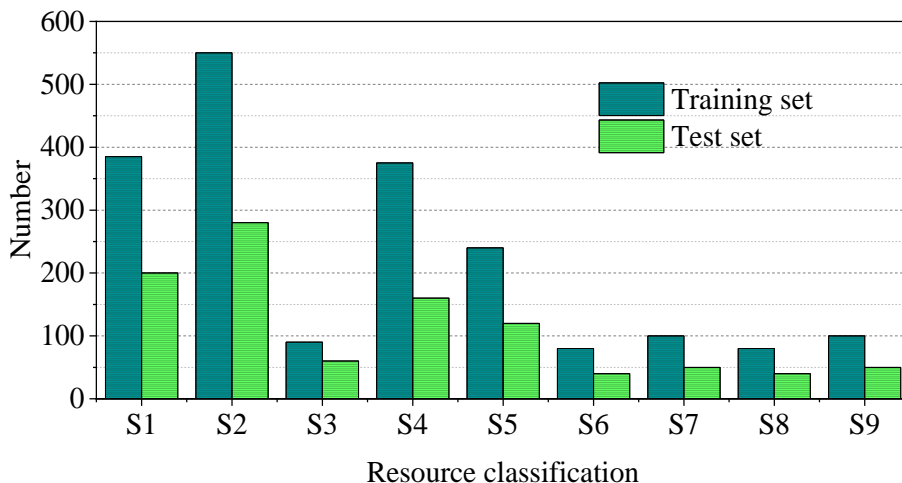


Figure 1: Selection of various teaching resources

The generalization capability of the proposed resource classification algorithm was tested through experiments with the help of the RCV1-v2 and AAPD resources. The RCV1-v2 resource has more than 800,000 news articles that are manually labeled and covers 103 different categories. One or more labels are given to every article. The AAPD resource contains more than 50,000 abstracts of computer science papers divided into 54 categories and all the abstracts are marked with one or more categories. All of them were divided into training and test subsets, in the ratio of 70-30.

3.3 Classification Analysis of Vocal Music Curriculum Resources

3.3.1 Accuracy results

The paper starts by evaluating the accuracy rates of different classification approaches as depicted in Fig. 2. Although the accuracy rates of each of the four methods have a certain level of variability, the specific differences in their classification outcomes are rather distinct. In the case of the CM1 approach, the accuracy of the classification of the different categories of vocal music course resources varied between 72.45% and 83.02%, but there is room to improve this. With the CM2 approach, there are more drastic changes in classification accuracy with the highest value being 83.10 and lowest value being 71.76. The CM3 method also exhibits significant variations where accuracy percentages range between 74.07% and 84.84%. However, the method suggested in this paper proves to be more accurate and stable with the lowest value being 82.91 and the highest value being 87.25. On average, the classification accuracies of the CM1, CM2, CM3 methods, and the method in this paper are 78.87, 79.57, 81.14 and 85.32, respectively, which means that the proposed method is 4.18-6.45 percent better than other methods. The findings indicate that the classification model of online vocal music course resources based on the decision tree algorithm created in this paper can successfully classify different types of resources.

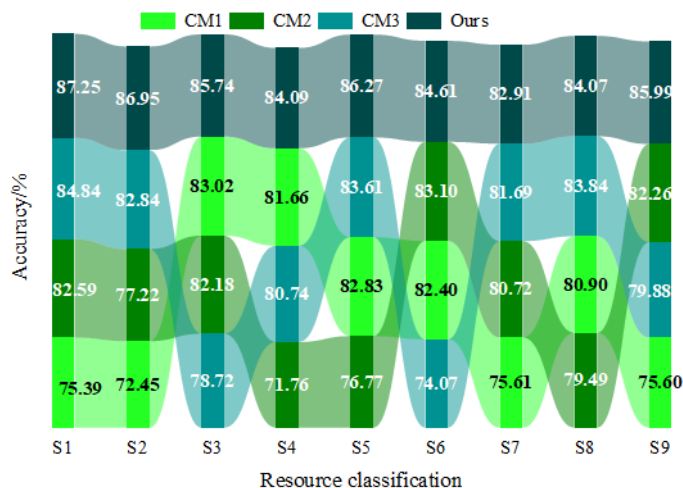


Figure 2: Classification accuracy of different methods

3.3.2 F1 value results

Expanding on these findings, this paper also measures the classification performance of the four methods based on the F1 value as an indicator, as illustrated in Fig. 3. The F1 values of the four methods indicate significant discrepancies between the results of classification of educational resources in terms of various categories of vocal music courses. The F1 values of the CM1 method vary between 72.95% and 83.08% where the lowest value is obtained with music thematic materials (72.95%) and the highest value with comprehensive data in vocal music education (83.08%). The F1 values of the CM2 method are more variable with a minimum of 71.98 percent on music thematic data and a maximum of 83.31 percent on pedagogical regulations data. The F1 values of the CM3 method fall between 71.69 percent and 83.24 percent. The method suggested in this paper, however, has F1 values between 81.25% (when it is applied to teaching reference data in the field of vocal music training) and 86.03% (in the case of teaching regulation data). Of the four methods, the F1 values of the classification of nine types of vocal music course resources using the proposed method are over 80 per cent, indicating that the classification method of online vocal music course resources can be used effectively to classify various types of educational materials.

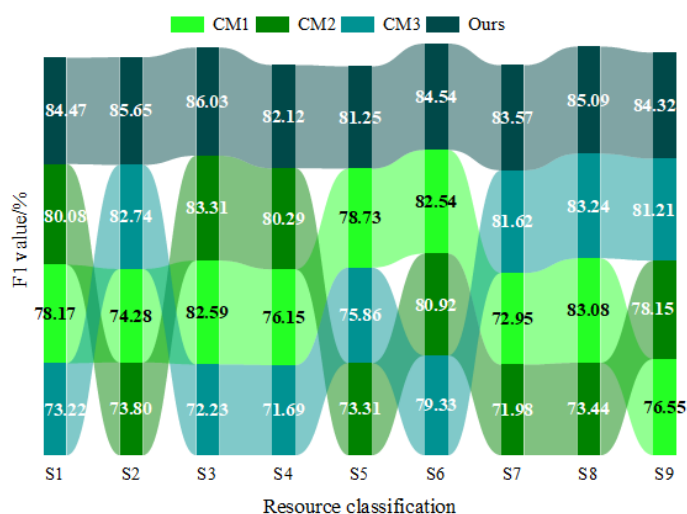


Figure 3: F1 values of different methods

3.4 Analysis of the results of the comparison experiment

The presented section discusses comparison experiments performed on RCV1-v2 and AAPD datasets based on two additional assessment values, namely Hamming Loss (HL) and Recall, in addition to Accuracy and F1 measurements. Hamming Loss is a popular measure in multi-label classification which counts the number of misclassified labels and smaller values are related to better model performance. Recall, conversely, refers to the ratio of positive examples that are identified as such, which indicates the capacity of the model to recognize relevant instances.

The presented table summarizes experiments conducted on RCV1-v2 data and compares the proposed approach with some other ones. As can be seen from the obtained results, the model that was analyzed in the current work performs better than any of the comparison models in all evaluation measures. In particular, the Hamming Loss decreases by 4.71-6.90 percentage points in the training sample and by 3.53-7.87 percentage points in the testing sample. The accuracy, the recall and the F1 scores are higher than 83 percent in the case of the proposed method, which is much higher as compared to the CM1, CM2, and CM3 methods.

Table 1: The results of the RCV1-V2 test set

Models		HL (-)	Accuracy (+)/%	Recall (+)/%	F1 value (+)/%
Training set	CM1	0.0085	82.24	73.78	75.54
	CM2	0.0087	71.55	82.15	76.72
	CM3	0.0086	80.02	79.87	80.76
	Ours	0.0081	85.14	84.09	83.01
Test set	CM1	0.0087	81.32	78.12	81.04
	CM2	0.0085	77.83	80.68	73.87
	CM3	0.0089	76.73	81.68	79.39
	Ours	0.0082	84.42	85.38	84.77

The findings derived using the AAPD dataset (Table 2) are comparable to the ones found based on the RCV1-v2 dataset. The proposed model shows higher performance measured by such measures as Hamming Loss, accuracy, recall, and F1 score. Namely, the Hamming Loss of the proposed approach is 0.0225 in the case of the training set and 0.0238 in the case of testing set, whereas all other methods have Hamming Loss larger than 0.024. In terms of accuracy, recall and F1 score, the CM1, CM2, and CM3 techniques attain scores of 64 percent to 76 percent, and the decision tree-based model presented in this paper is above 76 percent in all the metrics. Such outcomes demonstrate the obvious benefits of the proposed approach particularly in the case of large datasets.

Table 2: The results of the AAPD test set

Models		HL (-)	Accuracy (+)/%	Recall (+)/%	F1 value (+)/%
Training set	CM1	0.0305	64.62	65.29	64.62
	CM2	0.0294	65.09	64.83	65.82
	CM3	0.0249	74.31	75.52	74.73
	Ours	0.0225	78.45	79.57	78.59
Test set	CM1	0.0252	72.64	65.94	71.18
	CM2	0.0271	71.15	68.18	72.52
	CM3	0.0251	67.15	71.41	70.85
	Ours	0.0238	76.06	77.28	77.76

4 Conclusion

The use of digital teaching resources has become an integral part of vocal music education due to its wide range of material offerings, effective analysis options, and the ability to facilitate online interactive teaching, which leads to a better quality and effectiveness of teaching. The present paper discusses a decision-tree-based algorithm that can be used to categorize vocal online course resources. Through the process of selecting relevant attributes of vocal course resources, the suggested algorithm will sort these resources using the extracted information. Comparative experiments are utilized to test the performance of this algorithm, and the main results are as follows:

As far as accuracy is concerned, the suggested algorithm is found to have an average classification rate of 85.32, which is 4.18% to 6.45% greater than the three comparison techniques. On the basis of the F1 value, the classification algorithm of the present paper provides a range of F1 scores of between 81.25% and 86.03% which is much higher than the other methods. Also, the F1 values of all nine types of vocal course resources are constantly higher than 80%. Moreover, two more datasets are tested and it is found that the Hamming loss, accuracy, recall and F1 values of the proposed algorithm are higher than the other methods and therefore have a high generalization capability. The findings indicate that the proposed classification algorithm can be used not only to enhance the accuracy of vocal music online course resources classification, but also to provide opportunities to manage massive amounts of teaching materials in universities.

At present, the creation of knowledge systems and resources related to a particular course is one of the areas of study in the field of education. The future technologies include RDF semantic model, semantic graph-based algorithms of classification, and query algorithms that can be incorporated in the administration of vocal music online course resources.

About the Author

Yuyang Sun was born in Yantai, Shandong Province, China in 1989. He studied at Voronezh State Pedagogical University in Russia from 2015 to 2019 and obtained his doctoral degree in 2019. At present, he works in Mount Taishan College. Published 7 papers and 2 monographs, of which 3 have been included in the Russian BAK. Research direction: Musicology, music education, teaching methods, etc.

References

- [1] Loeckx, J. (2016). Blurring boundaries in education: Context and impact of MOOCs. *International Review of Research in Open and Distributed Learning*, 17(3), 92-121.
- [2] Alhazzani, N. (2020). MOOC's impact on higher education. *Social sciences & humanities open*, 2(1), 100030.
- [3] Liyanagunawardena, T. R., Williams, S., & Adams, A. A. (2014). The impact and reach of MOOCs: a developing countries' perspective. *eLearning Papers*, 38-46.
- [4] Kang, Z., & He, L. (2018). Construction and Practice of SPOC Teaching Mode based on MOOC. *International Journal of Emerging Technologies in Learning*, 13(12).
- [5] Jafari, E., Fathi Vajargah, K., Arefi, M. A. H. B. O. O. B. E. H., & Rezaeizadeh, M.

- (2019). Developing a Mooc-based curriculum model for higher education based on grounded theory: Case Study; Iranian Higher Education. *Technology of Education Journal (TEJ)*, 13(3), 581-593.
- [6] Zhao, H., Li, G., & Feng, W. (2018, August). Research on application of Massive Open Online Course (MOOC) in modern medical education teaching. In 2018 international conference on engineering simulation and intelligent control (ESAIC) (pp. 389-391). IEEE.
- [7] Mellati, M., & Khademi, M. (2020). MOOC-based educational program and interaction in distance education: long life mode of teaching. *Interactive Learning Environments*, 28(8), 1022-1035.
- [8] Chen, J. A., Wu, D., Cheng, J. N., Fu, B., Zhang, L., & Lu, A. M. (2017). Application of MOOC teaching in sports course teaching practice. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(12), 8089-8097.
- [9] Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, 129, 48-60.
- [10] Gao, B. (2019). Highly efficient English MOOC teaching model based on frontline education analysis. *International Journal of Emerging Technologies in Learning (Online)*, 14(6), 138.
- [11] Liu, Z., & Weng, C. (2023). Online Vocal Teaching: The Role of the Traditionally Trained Instructor and the Advantages Offered by the Digital Environment. *Croatian Journal of Education/Hrvatski Časopis za Odgoj i Obrazovanje*, 25(4).
- [12] Yang, G., & Yang, L. (2020). Exploration of vocal music teaching mode from the perspective of the age of artificial intelligence. *International Journal of Frontiers in Engineering Technology*, 2(1), 31-40.
- [13] Zhong, Y. (2025). Exploration on the Reform of Vocal Music Teaching for Music Majors in Colleges and Universities in the New Media Era. *Lecture Notes in Education, Arts, Management and Social Science*, 3(4), 7-11.
- [14] Jie, L., Yodwised, C., & Chalermksanyakorn, T. (2024). Creating the MOOC Teaching Model of Early Childhood Vocal Music Teaching for Teaching First-Year Students at Wenzhou University, Zhejiang Province, China. *Journal of Roi Kaensarn Academi*, 9(11), 1886-1891.
- [15] Liu, G. (2020, March). New age MOOCs teaching in music class of college. In *International Conference on Application of Intelligent Systems in Multi-modal Information Analytics* (pp. 605-610). Cham: Springer International Publishing.
- [16] Jiang, Y. (2022, January). MOOC vocal music blended teaching mode and digital resource platform design in the information age. In *2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City: Volume 2* (pp. 701-707). Singapore: Springer Singapore.
- [17] Ma, X. (2023). The Use of Knowledge Correlation for Classification and Evaluation of

- a Distance Education Teaching Resource Database. *International Journal of Emerging Technologies in Learning*, 18(5).
- [18] Li, Z. (2017). Design of Online Vocal Music Course Based on Azure-PaaS Platform. *International Journal of Emerging Technologies in Learning*, 12(9).
- [19] Baarir, N. F., Bourekkache, S., & Bouchet, F. (2025). Investigating the impact of emotions on learners' success and dropout prediction in a MOOC. *Interactive Learning Environments*, 1-33.
- [20] Tang, Y., & Zeng, X. (2024). Research on vocal music online educational platform based on internet platform. *International Journal of Web Engineering and Technology*, 19(4), 360-378.
- [21] Wang, X. (2022). Design of vocal music teaching system platform for music majors based on artificial intelligence. *Wireless Communications and Mobile Computing*, 2022(1), 5503834.
- [22] Peng, C. (2025). A Personalized Recommendation Method for Vocal Music Teaching Resources Based on Network Topology. *International Journal of High Speed Electronics and Systems*, 2540467.
- [23] Lin, P. H. (2024). Optimization of College Music Teaching Mode Based on Embedded Neural Network. *International Journal of High Speed Electronics and Systems*, 2540112.
- [24] Lin, H., Lin, Y., & Huang, H. (2022, September). Personalized recommendation method of online music teaching resources based on mobile terminal. In *International Conference on Advanced Hybrid Information Processing* (pp. 349-361). Cham: Springer Nature Switzerland.
- [25] Jiang, L. (2024). A fuzzy clustering approach for cloud-based personalized distance music education and resource management. *Soft Computing*, 28(2), 1707-1724.
- [26] Duan, T. (2025). Categorization Algorithm of Vocal Music Online Course Resources Under the Background of MOOC Teaching. *International Journal of High Speed Electronics and Systems*, 2540484.