



An improved accurate classification method for online education resources based on support vector machine (SVM): Algorithm and experiment

Zhi Quan¹ · Luoxi Pu²

Received: 22 August 2022 / Accepted: 5 December 2022 / Published online: 15 December 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the face of surging online education around the globe, it seems quite necessary and helpful for learners and teachers to have the plethora of online resources well sorted out beforehand. To some extent, the efficiency and accuracy of resource search and retrieval may determine the quality and influence of online education. In this research, based on the methodological framework of design science, the support vector machine (SVM) algorithm is chosen to optimise the design of an accurate resource classifier. The aim is to improve the unsatisfactory classification effect of traditional classification methods for online education resources, so that online learners can enjoy more accurate and convenient access to education resources they are seeking out of many more. For the purpose of performance evaluation, the proposed SVM-based classifier was compared with two other classification methods based on multiple neutral networks and deep learning respectively. Upon collection and pre-processing of online materials, the features of educational resources were extracted and output in the form of feature vectors. By calculating the similarity between the extracted feature vectors and the standard vectors of the set type, we obtained the classification results of online education resources for each of the three classifiers. It was found that, compared with those of the two traditional classification methods, the precision ratio and the recall ratio of the proposed classifier improved by 3.26% and 2.01% respectively. In the meantime, the proposed SVM-based classifier seems to more advantageous in performance balance with better *F* measurement.

Keywords Support vector machine (SVM) · Design science · Online education resources · Accurate classification

✉ Zhi Quan
quanzhi@swufe.edu.cn

¹ Southwestern University of Finance and Economics, Chengdu, China

² UCSI University, Kuala Lumpur, Malaysia

1 Introduction

In the past decades, online education has become a global phenomenon by penetrating into virtually all countries in the world with the expansion of the Internet infrastructure (Chan et al., 2022; Dhawan, 2020; McCarty et al., 2006; Rye, 2014). This trend is particularly the case for China, with perhaps the largest population of learners on the Earth. According to a 2018 report, up to 144 million people had taken up online education as of June 2017 in China (Yang & Du, 2018); the number surged amazingly to over 300 million (including millions of teachers) in 2020 after the breakout of the COVID-19 pandemic, when online delivery was almost the only choice for both educators and students then (Li, 2020b). This world's largest ICT-based teaching experiment and reform have been continuing in the post-epidemic era, offering a wealth of diversified education resources through the Internet (Li, 2020a). Online education, receiving unprecedented and ever-increasing attention from educators, students and the public, has become integral to the whole educational system.

The abundant materials provided by all kinds of online education platforms bring both benefits and challenges to users. On the one hand, teachers and students are enabled to employ online courses to complete pedagogical processes. On the other hand, it tends to be rather time-consuming to retrieve the right materials that one is seeking out of a huge amount of online resources. To some extent, the efficiency of resource search and retrieval may determine the quality and influence of online education. Thus, enriched content entails enhanced classification methods for effectively and efficient delivery of online education, otherwise many learners may get lost and bored, and even daunted, when wading through massive data of online materials.

In the practice of classifying online education resources, there seem to be problems of incomplete coverage of fields and unscientific categorisation, among others. In this research, an accurate classification method based on support vector machine (SVM) is proposed in order to improve the utilization of online education resources. In a properly built model for autonomous learner users, SVM has the potential to help enhance resources classification and allocation of online education resources. Delving into the optimisation of SVM is hence of significant educational and research value. A brief review of relevant literature is set out below, followed by the proposed algorithm and results of a comparison experiment to evaluate the effectiveness and efficiency of the SVM-based classification algorithm.

2 Literature review

Classification can be seen as a process to summarise the features and classification rules based on sample data sets of existing resources and to establish rules of discrimination, so that new resources can be categorised according to such established rules (see Diederich, 2008). To classify the ever-growing body of

online education resources may be a task beyond human capacities. We may take China as an example: years ago, it was planned that by 2020, there would be over 3,000 national-level courses available online (Yang & Du, 2018); the amount of online courses offered by all levels of education providers may become even multi-fold of that now. In addition, each course may include multiple formats of materials: texts, images, audios, videos and so on, which exacerbates the complexity and difficulty of data search and retrieval. In this case, it seems a good solution to train machines to analyse such large and complex datasets (Shalev-Shwartz & Ben-David, 2014), and document classification has been a traditional task that machine learning can deal with satisfactorily (Mehryar et al., 2018).

In history, international research on the classification methods of education resources had an early start. Methods based on word frequency statistics and factor analysis have been widely used in email, information retrieval, and so on (see Joachims, 2002). Other traditional methods of more complex resource classification may have to rely on multiple neural network integration and deep learning (Lam et al., 2012). However, some problems, such as incomplete coverage of fields and unscientific categorisation, can be identified in practice, which ultimately affects the quality of their application. Thus, the support vector machine (SVM) is adopted as the basis for the proposed classifier in this research, since this algorithm may help solve the extremum problem in the traditional methods (Hamel, 2009).

SVM is an established method in natural language process (NLP). It is one of the most significant kernel-based methods of machine learning and one of the most popular supervised learning algorithms, widely used for classification and regression analysis (Steinwart & Christmann, 2008). A classic definition is set out as follows (Cristianini & Shawe-Taylor, 2000, p. 7):

Support Vector Machines (SVM) are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimisation theory that implements a learning bias derived from statistical learning theory. This learning strategy introduced by Vapnik and co-workers is a principled and very powerful method that in the few years since its introduction has already outperformed most other systems in a wide variety of applications.

Given the agreed-on outperformance of SVM over other methods, it is not surprising that SVM has been widely used in research of various fields, including but limited to financial analysis, medical analysis, biology, so on and so forth (Ma & Guo, 2014; Murty & Raghava, 2016; Suykens et al., 2015; Wang, 2005).

Here is a basic explanation and illustration of the fundamental principle of SVM for classification, as shown in Fig. 1. The central task of SVM is to create a hyperplane between data sets to indicate which class an item probably belongs to. The challenge is to train the machine to understand structure from data and map with the right class label. For the best result, the hyperplane has the largest distance to the nearest training data points of any class. Thus, the classification based on the SVM algorithm can be considered as a hyperplane made up of

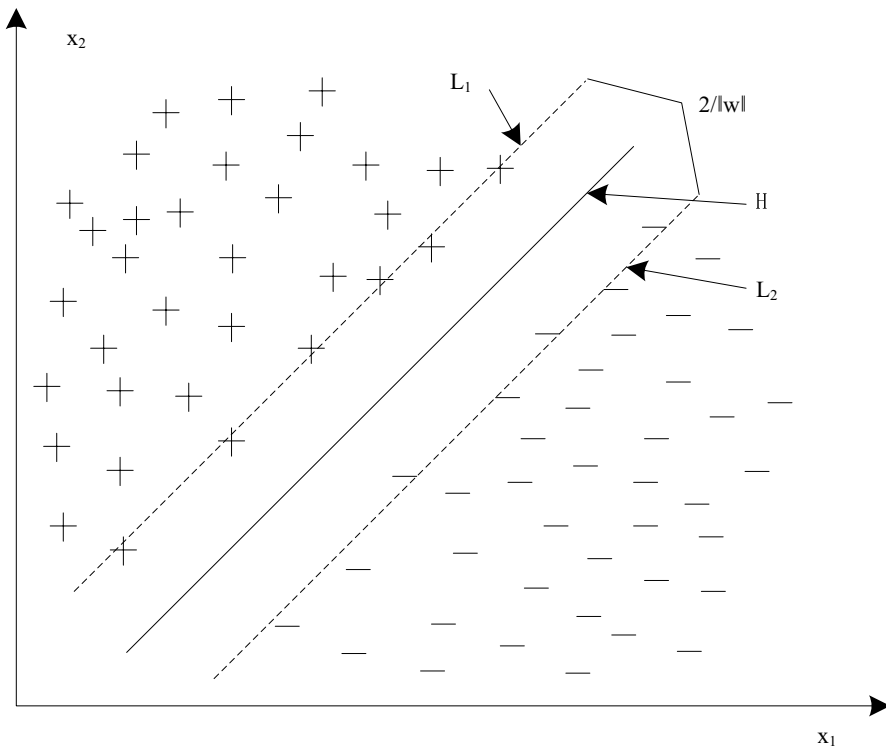


Fig. 1 The principle of classification using the SVM algorithm

multiple separate heterogeneous data samples, to which a solution can be worked out. Through the calculation of the maximum spacing of heterogeneous samples, the category of the target sample can be determined, and upon further processing the classification of all samples can be completed.

The signs "+" and "-" in Fig. 1 represent distributed and negative data samples respectively. It can be seen in the graph that, H points to the centre of the adjacent edge between the two types of samples, serving as the isolation barrier for different types of samples; L_1 and L_2 are located on the verge of the two types of samples, and the distance between them shows the classification margin. During the process of solving the maximum value, if the dividing line H can completely separate positive and negative samples like a watershed, H is then optimal (Liu et al., 2019). Experimental results show that, if the classification samples are distributed in multi-dimensional space, on the assumption that there is hyperplane H -value completely separated with the most significant classification interval, and that there are two types of classification – 'positive' and 'negative', then the classification interval is the largest, which can also be used to predict other data of the same classification. Deng et al. (2013) provide more comprehensive and extended elaboration on the mathematic representation of SVM.

Certainly SVM is not perfect from its inception; it has been embracing continuous improvement with researchers' effort. Wiering and Schomaker (2015) point out

the limitations of standard SVM: a shallow model with a single layer of support vector coefficients, and overreliance on inflexible kernel functions; instead, they propose a transition from the single-layer SVM to the multi-layer SVM with deep architectures. There have also been proposals to integrate SVM with other established algorithms (e.g. Stoean & Stoean, 2014). In a word, iterative progress can make SVM a refined tool for both research and practice. Although it may not be the most novel or popular one, there seems to be much room to optimise this method based on previous and existing endeavours.

3 Methodology

This research follows the methodological framework of design science research (DSR) or design research. DSR has been widely used in IT engineering for decades (Gregor, 2021; Hevner et al., 2004; Peppers et al., 2007), and it features intentional creation and development of artifacts that serve human purposes (Dresch et al., 2015), aiming to “change the state-of-the-world through the introduction of novel artifacts” (Vaishnavi & Kuechler, 2008, p. 18). Apart from tangible and concrete solutions to identified problems in human activities, DSR also seeks new understanding through dynamic interaction of artifacts and knowledge.

This research aims to develop an optimised classifier with higher classification accuracy of online education resources and gain new knowledge on how to best classify massive resources for best delivery of online education. The ultimate goal of the proposed method is to categorise the disciplines, specialisation and chapters of such resources based on the content. However, different education resources rely on various storage formats, including texts, videos, web pages, images, etc. It is helpful to target online education resources in different multimedia formats and determine the specific types via content analysis and feature analysis. Prior to the design work, it is necessary to identify and set rules for online resources of different disciplines as the reference standards for classification.

3.1 Automatic collection

Appropriate methods of automatic collection and processing of online education resources are entailed. The first step is to collect the target online education resources on the Internet using web crawlers. Compared with traditional methods of data collection, web crawlers can start to search webpages on the Internet according to pre-set URLs and extract the links in the pages. Then new links can be retrieved, and education resources can be downloaded automatically. The initial collection of education resources involves data including texts, videos and webpages, and the main module of web crawlers supports webpage data download and data parsing.

In the actual collection process, the action is to initiate the downloader module of web crawlers according to set parameters and read the first URL. According to the results, online education resources are searched on the Internet. Comparison with data in the local resource library is then conducted upon identification of the online

resource location, so as to check if the online resource has already been stored in the local repository. If it is already included, there is no need to repeatedly download it, otherwise it will be downloaded and stored in the local repository. During the comparison of education resources, apart from the resource names, the size and update date/time should also be compared, to ensure that the collected resources are of the latest versions. Upon the completion of a round of collecting resource data, the corresponding URLs are stored in the list of assessed resources. The next step is to parse the newly downloaded data using the parser in the crawlers and extract the URL information. If the extracted URL is existing, it is deleted immediately, otherwise it is stored in the list of resources to access. Upon the completion of resource downloading following the above procedure, another URL is extracted and the above procedure is cycled. All the online education resources are collected from the Internet, and collected data are stored separately according to different storage types.

3.2 Processing of collected resources

It is helpful to clarify the processing of online resources in text, image and video formats. The purpose of processing text-format educational resources is to extract the target data and convert it to row format. The units of text data are words, phrases, paragraphs and so on, and text information exactly consists of such units of natural language. In the process of text feature representation, it is a major step to extract the noun phrases and proper nouns, like names of people and places in the texts. There might be a large number of characters irrelevant to the central ideas of the texts, such as numbers, links, punctuation marks, and stop words. In order to reduce the complexity of processing text data, it is helpful to have the semantic vocabulary set of highly conceptual texts as the textual feature set and replace original text with the textual features. To ensure the effective information content in the textual feature vector, pre-processing text data is an important procedure worthy of further research. The next step is to filter the stop words in the original texts. Stop words can be divided into two kinds: undistinguishable words and function words. Undistinguishable words refer to high frequency words in almost all types of texts. Function words include pronouns, participles, and so on. Upon establishment of the stop word list, the matched words will be deleted, while the unmatched will be retained in the keyword list.

3.2.1 Preprocessing image-format education resources

Pre-processing image-format educational resources takes two steps: the one is screening based on image quality, and the other is unified processing of image formats. Simple filtering comes first during the quality-based screening of images, where median filtering processor is employed. If the output image resolution is higher than 75%, it will be retained, otherwise the image will be removed. The unified processing of image formats can be divided into two aspects: the unification of image storage formats and image colours. It is set that the storage format of educational resources should be JPG and the colour space should be RGB.

3.2.2 Pre-processing video-format education resources

To guarantee the classification efficiency of educational resources in video formats, it is necessary to mine valuable information frame-by-frame in the video resources, which can be deemed as a video-format bag of words. Figure 2 illustrates the text mining process of video-format educational resources.

Mining of image information in video-format educational resources can be conducted in the same way. Then video-format educational resources can be pre-processed following the above methods for text-format and image-format educational resources.

3.3 Extracting features online education resources

This section addresses document frequency, information gain and word frequency. Firstly, document frequency refers to the number of documents in the collection that contain a term. The larger the number is, the more frequently the certain term appears in documents, and the more the feature words contribute to classification. This can be used as an important criterion for classification. After word filtering, the dimensions of text vector are reduced, with little impact on the classification accuracy. The extraction of feature words can help reduce the dimensions of vector space and amount of calculation, which can indirectly improve the efficiency and accuracy of text classification.

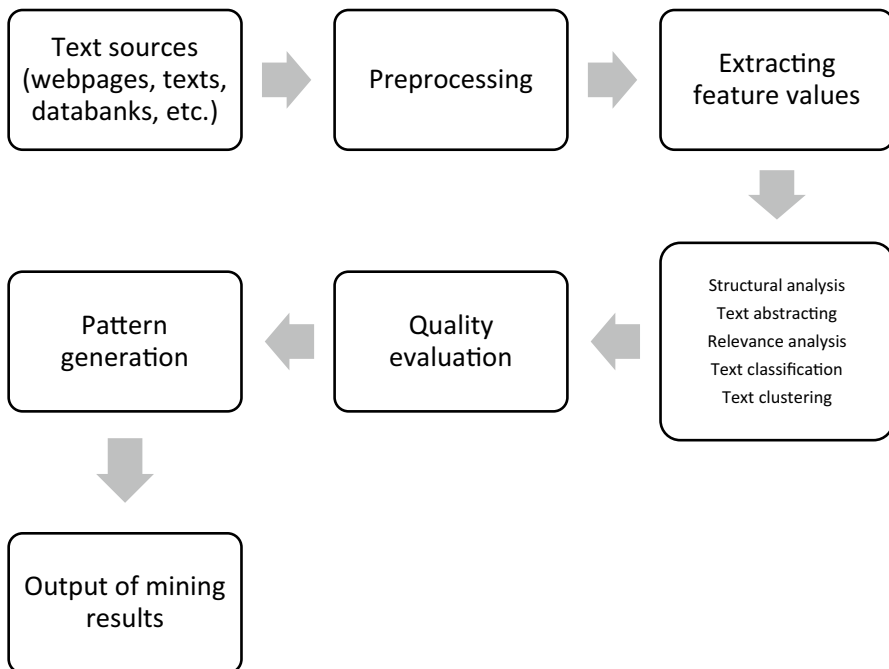


Fig. 2 Flow chart of text mining of video education resources

Secondly, information gain of textual resources refers to the changed amount of entropy with the creation of texts, which is an important part of text classification. The features of information gain extracted from online educational texts can be expressed as:

$$\begin{aligned} \mu_{IG}(w) = & P(W) \sum_{i=1}^m P(C_i|W) \lg P(C_i|W) - \sum_{i=1}^m P(C_i) \lg P(C_i) \\ & + P(\bar{W}) \sum_{i=1}^m P(C_i|\bar{W}) \end{aligned} \quad (1)$$

In Eq. (1), C_i and W stand for class variables and features respectively. The variables $P(C_i|W)$ and $P(\bar{W})$ are the probabilities of the text falling into the category C_i under two conditions: including or excluding the feature W ; $P(C_i|\bar{W})$ is the conditional probability of the text falling into the category C when W is not included.

Thirdly, word frequency refers to the frequency of target words in textual educational resources, which can be calculated as follows:

$$\mu_w(t, \vec{d}) = \frac{tf(t, \vec{d}) \times \lg\left(\frac{N}{n_t} + 0.01\right)}{\sqrt{\sum_{t \in \vec{d}} \left[tf(t, \vec{d}) \times \lg\left(\frac{N}{n_t} + 0.01\right)\right]^2}} \quad (2)$$

In Eq. (2), $\mu_w(t, \vec{d})$ refers to the weight of the word t in the text $\vec{d}$; the variable $tf(t, \vec{d})$ is the word frequency; N and n_t stand for the total numbers of training texts and those found to include the word t .

3.4 Building an accurate classifier of resources using the SVM algorithm

Based on the fundamental SVM illustration in Fig. 1, an optimised accurate classifier can be built with the following steps. We assume that, as per the set of training texts of online educational resources, the initial input is (x_n, y_n) , and then the division of the hyperplane H can be shown using the linear equation in Eq. (3) in the two-dimensional space.

$$u^T x + b = 0 \quad (3)$$

In Eq. (3), u and b are the normal vector and offset value of the linear equation respectively. The samples on L_1 and L_2 in Fig. 1 are defined as the support vector, and $\frac{2}{\|u\|}$ represents the classification margin. Then the maximization problem of online educational resource classification can be converted to a corresponding dual problem: with constraints set, the task is to find out the maximum value of the dual function.

$$\begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \end{cases} \quad (4)$$

The vector α in Eq. (4) is the language multiplier corresponding to the samples in the training data. With the solution of the vector α , the specific values of the parameters of u and b in the optimal hyperplane function. The kernel function of null SVM is $K(x_i, x_j)$, based on which the function of an accurate classifier for non-linear resources can be developed as:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \quad (5)$$

Lastly, the iterative training procedure of the proposed SVM-based optimised classifier is shown in Fig. 3.

3.5 Implementing classification of online educational resources

This step is to have the automatically collected and pre-processed online educational resources as the entries and load them chronologically into the SVM-based classifier. According to the features of the loaded resources, with the classifier serving as the running environment, the degree of similarity between the feature vector extracted and the features of the target categories can be calculated following this equation:

$$Sim(D_0, D_i) = \sum_{k=1}^p \mu_{ki} \times \mu_{kj} \quad (6)$$

In Eq. (6), μ_{ki} and μ_{kj} are the standard feature vector of a certain category D_0 and the comprehensive feature vector extracted from loaded samples. If we compare the results from Eq. (6) with the pre-set similarity threshold, the category with calculated value above the threshold can be used to label the specific educational resource. If there are more than one category that can meet the criterion, the one with the highest similarity will be used as the final result of classification (Beyene et al., 2020).

4 Results from a comparison experiment

A comparison experiment was designed to test the classification accuracy of the proposed SVM-based accurate classifier of online educational resources. The counterparts in the experiment were the resource classification method based on fusion of multiple neural networks and the method based on deep learning (Lam et al., 2012). In order to reduce the impact of independent variables on the results, the online education platform and raw samples of resources were identical.

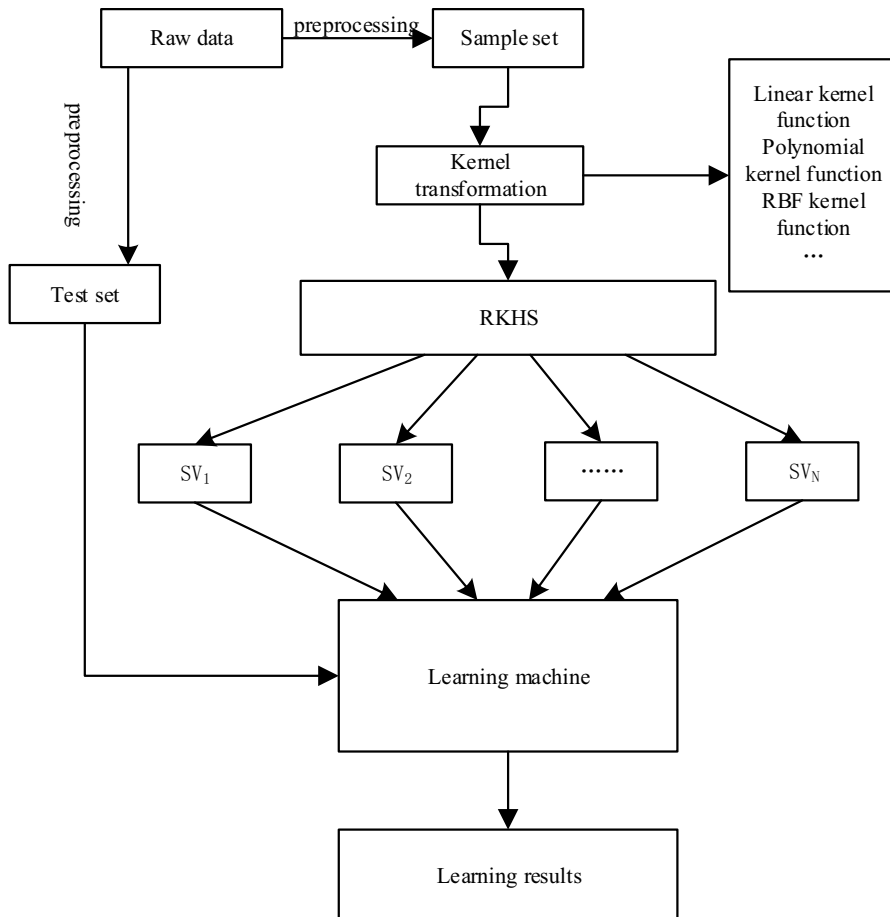


Fig. 3 The iterative training procedure of the proposed resource classifier

4.1 Configuration of the online education platform

The experiment used the online education platform of a university located in the southwestern region of China as the environment. The platform consists of several client sides for students and teachers, and one server and one database. All the resources of the online education platform are stored in the database. In the experiment, the three classification methods (the SVM-based one, and the ones based on fusion of multiple neural networks and deep learning) were translated into codes and embedded into the online education platform. Figure 4 illustrates the configuration of the classification methods in the experimental environment.

The two experimental comparison methods can also be imported and configured following the same way. In order to guarantee the independence of the three

Crossvalidate Parameters...

General

SVM type: C_SVC

Kernel function type: Linear

Multi-class SVM type: 1-v-1

Parameter

Penalty factor C: 100

Degree: 2

nu: 1.0

Gamma: 0.5

Epsilon: 0.1

Coefficient: 0

Cache Size: 40

Ping criteria: 0.001

☐ Enable probability estimate

☒ Enable shrinking

☐ Enable Weight

Sample<label:weight>: 4:100 2:50 7:25

Cross validate: 2

OK Cancel

Fig. 4 Configuration of the classification methods of online education resources

classification methods, parallel running was selected to implement the calling and switching among the different methods.

The samples of online educational resources used in this experiment were taken from two sources: the database of the education platform of the above-mentioned university, and the database of the University of California, Irvine. Several experiments were conducted on multiple data sets from the above database, to find more accurate results of classification. Upon statistical processing, the size of the sample online educational resources in this experiment amounted to 254.65 GB, all for the discipline of mathematics. The types of resources included texts, tables, images, videos, audios and so on.

After that, three indicators: precision ratio, recall ratio and F measurement, were set as the indicators to evaluate classification results. Recall ratio refers to the proportion of correctly categorised documents after classification, while precision ratio refers to the ratio between the number of correctly categorised documents and the number of documents expected in that category. The results of precision ratio and recall ratio are shown as P and R . The quantitative results of the two indicators can be expressed as:

$$\begin{cases} P = \frac{TP}{TP+FP} \times 100\% \\ R = \frac{TP}{TP+FN} \times 100\% \end{cases} \quad (7)$$

In Eq. (7), TP refers to the amount of correctly categorised resources, while FP is the amount of incorrectly categorised resources, and FN refers to the amount of unclassified resources that fall into the category. The last indicator F measurement is used to measure the balance between precision ratio and recall ratio, which can be expressed as:

$$F_{\beta}(P, R) = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (8)$$

In Eq. (8), β is the adjustment parameter. In general, the higher the F measurement is, the more balanced the precision ratio and recall ratio are.

4.2 Result analysis of the comparison experiment

As designed, the different classification methods were imported into the experiment environment. Upon debugging it was ensured that the methods could run in the experiment environment. Classification results of the proposed SVM-based classifier, together with those of the other two methods, can be worked out following the same procedure. Then we compared the results with the quantitative results of the set indicators, as shown in Table 1.

Putting the data in Table 1 into Eq. (7), we can work out the average precision ratio and recall ratio of the three classification methods, respectively as: 94.01% and 96.29%; 95.51% and 97.26%; 98.02% and 98.86%. It can be seen that the proposed SVM-based classification method achieved some improvement in both precision and recall ratios. If we input the results of precision ratio and recall ratio into Eq. (8) and determine the parameter β as 1, the quantitative results of F measurement on different data sets can be worked out, as shown in Fig. 5.

Table 1 Testing results of classification performance of online education resources

Times of experiments		1	2	3	4	5
Classifier based on fusion of multiple neural networks	TP/GB	46.27	43.44	45.62	42.77	44.01
	FP/GB	2.81	2.79	2.94	2.85	2.76
	FN/GB	1.85	1.67	1.57	1.72	1.74
Classifier based on deep learning	TP/GB	49.35	48.62	47.17	48.04	49.22
	FP/GB	2.27	2.35	2.19	2.31	2.28
	FN/GB	1.44	1.29	1.35	1.41	1.33
Proposed SVM-based classifier	TP/GB	51.86	61.39	55.82	54.07	62.08
	FP/GB	1.22	1.05	1.16	1.21	1.07
	FN/GB	0.59	0.67	0.82	0.64	0.55

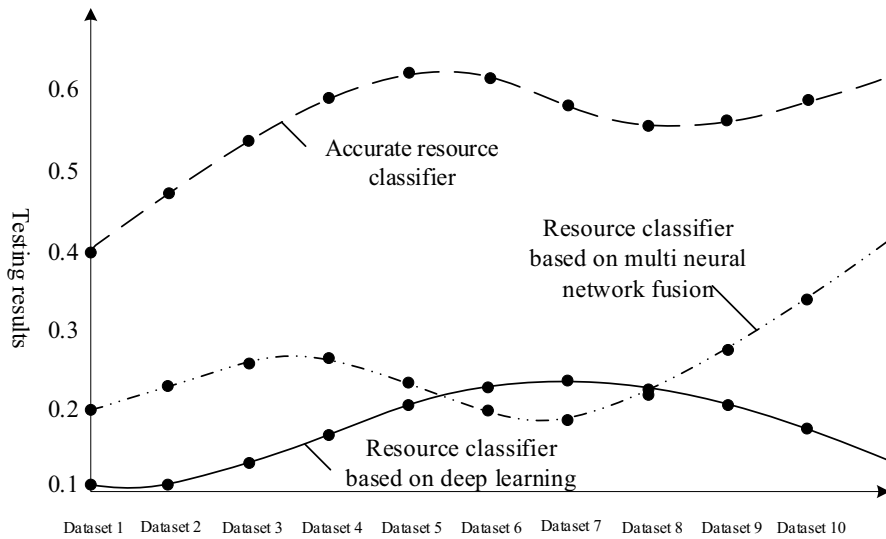


Fig. 5 Testing results of F measurement

As can be seen from Fig. 5, the F measurement result of the proposed SVM-based accurate classification method for online educational resources is always higher in all data sets than those of the other two methods. In other words, the classification performance of the proposed classifier tends to be more balanced. It implies that, the proposed accurate classifier of online educational resources may achieve the least exclusion of possible samples to the utmost degree of classification accuracy.

5 Conclusions

In a nutshell, the SVM-based classifier designed in this research can achieve slight improvement in both precision ratio and recall ratio of resource classification, when compared with the traditional methods. The classifier can reconcile the two seemingly conflicting indicators to a large extent and produce more balanced output with more accurate and inclusive results of classification. That is, by using this SVM-based accurate classifier, users may be provided with some more resources put into the correct categories at a time. It is believed that easier, more decentralised and engaging access to online resources is the trend of ICT-enhanced education (Fox, 2011). As the designed classifier can enhance the classification results of online education resources, it will thus indirectly improve the retrieval efficiency and usability of such resources.

The limitations of this research may invite further effort on this topic. Firstly, all the samples are of mathematics only. If the algorithm is applied to other disciplines distinctively different from mathematics, e.g., visual arts, literature, the testing results may differ. In the meantime, online materials for cross-disciplinary subjects, such as behavioural finance and computational linguistics, are not considered

in this experiment. Secondly, no more major methods apart from SVM are involved in this research, while the synergy of hybrid algorithms may help further improve classification results, e.g., combining SVM with deep learning techniques. All in all, optimal results of classification entail ongoing and iterative optimisation of accurate classifiers, including this SVM-based one. To achieve this goal, programming integrating multiple methods can be tested on materials from multiple disciplines to identify the best solution and best practice.

Data availability The data supporting the findings of this study are available upon reasonable request.

Declarations

Conflicts of interest/Competing interests No potential conflict of interest was reported by the authors.

References

- Beyene, W. M., Mekonnen, A. T., & Giannoumis, G. A. (2020). Inclusion, access, and accessibility of educational resources in higher education institutions: Exploring the Ethiopian context. *International Journal of Inclusive Education*, 1–17. <https://doi.org/10.1080/13603116.2020.1817580>
- Chan, R. Y., Bista, K., & Allen, R. M. (Eds.). (2022). *Online teaching and learning in higher education during COVID-19: International perspectives and experiences*. Routledge.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- Deng, N., Tian, Y., & Zhang, C. (2013). *Support vector machines: Optimization based theory, algorithms, and extensions*. CRC Press.
- Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Diederich, J. (Ed.). (2008). *Rule extraction from support vector machines*. Springer.
- Dresch, A., Lacerda, D. P., & Antunes, J. A. V., Jr. (2015). *Design science research: A method for science and technology advancement*. Springer.
- Fox, R. (2011). Technological practice and change in education. In R. Kwan, C. McNaught, P. Tsang, F. L. Wang, & K. C. Li (Eds.), *Enhancing learning through technology - Education unplugged: Mobile technologies and Web 2.0 (International Conference, ICT 2011 Proceedings)* (pp. 1–7). Springer.
- Gregor, S. (2021). Reflections on the practice of design science in information systems. In S. Aier, P. Rohner, & J. Schelp (Eds.), *Engineering the transformation of the enterprise: A design science research perspective* (pp. 101–113). Springer.
- Hamel, L. H. (2009). *Knowledge discovery with support vector machines*. Wiley.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Joachims, T. (Ed.). (2002). *Learning to classify text using support vector machines*. Springer.
- Lam, H. K., Ling, S. H., & Nguyen, H. T. (Eds.). (2012). *Computational intelligence and its applications evolutionary computation, fuzzy logic, neural network and support vector machine techniques*. Imperial College Press.
- Li, C. (2020a). Online education in the post-epidemic Era. *International Journal of Social Science and Education Research*, 3(10), 332–335. [https://doi.org/10.6918/IJOSSE.202010_3\(10\).0052](https://doi.org/10.6918/IJOSSE.202010_3(10).0052)
- Li, Y. (2020b). Booming online education. *China Today*, pp. 56–59. Retrieved from http://www.china-today.com.cn/ctenglish/2018/s1/2020b06/t2020b0605_800208979.html
- Liu, M., Liu, X., & Li, S. (2019). Research on the resource planning procedure of education cloud platforms: Take enterprises of digital education resources as an example [(in Chinese) Jiao Yu Yun Ping Tai Zi Yuan Gui Hua Liu Cheng Yan Jiu: Yi Shu Zi Jiao Yu Qi Ye Wei Li]. *E-Education Research*, 40(11), 78–85.
- Ma, Y., & Guo, G. (Eds.). (2014). *Support vector machines applications*. Springer.

- McCarty, S., Ibrahim, B., Sedunov, B., & Sharma, R. (2006). Global online education. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *The international handbook of virtual learning environments* (pp. 723–787). Springer.
- Mehryar, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning* (2nd ed.). The MIT Press.
- Murty, M. N., & Raghava, R. (2016). *Support vector machines and perceptrons: Learning, optimization, classification, and application to social networks*. Springer.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Rye, S. A. (2014). The educational space of global online higher education. *Geoforum*, 51, 6–14. <https://doi.org/10.1016/j.geoforum.2013.09.011>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer.
- Stoean, C., & Stoean, R. (2014). *Support vector machines and evolutionary algorithms for classification: Single or together*. Springer.
- Suykens, J. A. K., Signoretto, M., & Argyriou, A. (Eds.). (2015). *Regularization, optimization, kernels, and support vector machines*. CRC Press.
- Vaishnavi, V. K., & Kuechler, W., Jr. (2008). *Design science research methods and patterns: Innovating information and communication technology*. Auerbach Publications.
- Wang, L. (Ed.). (2005). *Support vector machines: Theory and applications*. Springer.
- Wiering, M. A., & Schomaker, L. (2015). Multi-layer support vector machines. In J. A. K. Suykens, M. Signoretto, & A. Argyriou (Eds.), *Regularization, optimization, kernels, and support vector machines* (pp. 457–476). CRC Press.
- Yang, C., & Du, M. (2018). Booming online education brings chances, challenges in China. *People's Daily*. Retrieved from http://www.chinadaily.com.cn/edu/2018-02/08/content_35667621.htm

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.