

Publication Date: 10 August 2024

Archs Sci. (2024) Volume 74, Issue S1 Pages 73-78, Paper ID 2024s111.
<https://doi.org/10.62227/as/74s111>

The Integrated Teaching Platform of Innovation, Entrepreneurship and Moral Education in Colleges via Multimedia Network

Zijin Li¹ and Weijie Zhao^{1,*}

¹Pingdingshan University, Pingdingshan 467000, Henan, China.

Corresponding authors: (e-mail: keyanzwj@163.com).

Abstract A teaching platform community for innovation, entrepreneurship, and moral education is being developed in colleges and universities against the backdrop of a multimedia network. This platform will aid in the formation of a multi-subject collaborative education mechanism and the joint cultivation of innovative and entrepreneurial talents, thereby promoting the high-quality development of innovation and entrepreneurship education in colleges and universities. Given the widespread use of neural networks, we address the utility of this technology in cross-modal multimedia retrieval from the following three perspectives using a depth-limited Boltzmann machine pre-trained sparse neural network model. The integration of moral education into the teaching platform allows for the realisation of quantitative and qualitative evaluation from theory to practise. Experiments demonstrate that our model can perform significantly better than a single model's baseline performance.

Index Terms neural network, integration of moral education, teaching, platform construction

1. Introduction

To support the overall transformation of higher education, colleges and universities should further their teaching of innovation and entrepreneurship. Building a scientific and technical consortium with participation from big and medium-sized technology businesses, small and medium-sized enterprises, colleges, and universities would also help to boost the leading and leading role of innovative enterprises [1], [2]. In order to carry out innovation and entrepreneurship education, foreign nations currently have experience developing communities based on the combination of production and education. For instance, American colleges and business firms collaborate to create the Innovation and Entrepreneurship Accelerator, a unique organisation that can coordinate the interaction between the university and the business and foster close cooperation between the two [3], [4]. College instructors and students can empower instructors and students by participating in the innovation and entrepreneurship accelerator. Local colleges and universities are developing their programmes in innovation and entrepreneurship education in the context of the new era, which can also be seen as a key component of the national innovation-driven development strategy [5].

Currently, innovation and entrepreneurship education in local colleges and universities has advanced favourably thanks to the aggressive promotion of numerous national initiatives,

but there are still some open issues that require attention. There is a lack of innovative and entrepreneurial project incubation platforms, there is a lack of integration between entrepreneurship education and professional education, etc. This theoretical and practical aspect of the integration of production and education means that it must follow the path of integrated development and common development with other elements, i.e., community development, in order to achieve practical results in innovation and entrepreneurship education in colleges and universities and fully play its role in the national innovation-driven development strategy [6], [7]. Building a community of innovation, entrepreneurship, and education in nearby colleges and universities for the integration of industry and education is helpful for achieving the effective connection between the supply and demand of social talent training [8].

Even though the existing subspace learning-based cross-modal multimedia retrieval techniques have had some success, there are still numerous issues: Semantic information, a high-level abstract idea from human comprehension, is not used by unsupervised retrieval techniques, but it is crucial for enhancing Unsupervised approaches cannot adequately describe the high-level semantic connection of multimodal data due to the model's strong class-discriminative capabilities; supervised retrieval methods, on the other hand, can increase retrieval accuracy without taking any unlabeled data into account

[9]. Consequently, it is impossible to accurately model the global semantic structural information of multimodal data; To further increase retrieval accuracy, semi-supervised retrieval techniques make use of the latent semantic information in unlabeled material. But these techniques When different modal data are treated independently and combined in a way that is complementary to cross-media semantic analysis, retrieval accuracy can be increased [10]. The deep neural network can be thought of as a multi-layer nonlinear projection, which is different from a single projection matrix. In addition, the traditional cross-modal multimedia retrieval method based on subspace learning uses a pair of projection matrices to project the underlying features of the image and text into the common subspace of the two [10].

The three primary contributions of this study are as follows:

- Using deep learning to build a platform for instruction that integrates innovation, entrepreneurship, and moral education at colleges and institutions against a backdrop of multimedia networks.
- This research investigates the use of neural network technology in cross-modal multimedia retrieval from three perspectives using a depth-restricted Boltzmann machine pre-trained sparse neural network model.
- Research and analysis of the integrated teaching platform of innovation, entrepreneurship, and moral education at colleges and universities using our method has shown that it is effective and suitable, as demonstrated by experiments.

II. Methods

A. Collaborative deep semantic learning framework

This section explains the cross-modal multimedia retrieval approach based on cooperative deep semantic learning, including implementation details and a detailed flow diagram [11]. Figure 1 displays the CR-CDSL frame diagram. First, utilising the features and semantic labels of the retrieved multimodal data, two complimentary deep neural network models are built and trained. The underlying characteristics of the picture and text samples are projected into their shared semantic representation space using the two network models. Create weak semantic labels for unlabeled images and text, then input the labelled training examples and the newly generated weak semantic labels into the neural network model to retrain the network.

Two complimentary deep neural network models are created through collaborative deep semantic learning in order to jointly produce weak semantic labels for unlabeled samples. The algorithm flow is as follows: First, initialize the predicted label vector matrix US of unlabeled examples as an all-zero matrix; second, Input all image and text underlying features and their label vector matrix LS and initialized US into two deep neural network models and start training; then, input unlabeled image samples UI and text samples UT into trained neural network models and obtain their top-level outputs IO and TO ; based on the computed IO and TO , the synergistic

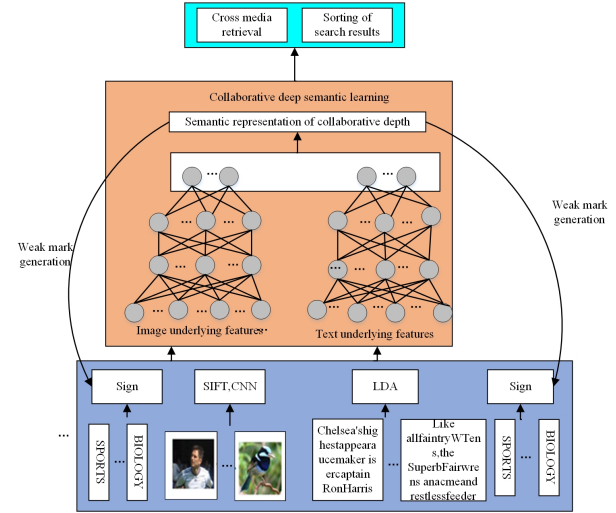


Figure 1: Framework of collaborative deep semantic learning

semantic representation US of unlabeled examples can be Eq. (1)

$$S_U = (O_I - O_T) * (O_I - O_T). \quad (1)$$

Export US , and we get weak semantic markers for unlabeled examples. It is then processed as follows: first convert US into a probability distribution representation (as shown in Eq. (2)), then, for each column in it, set the maximum value of the elements in is to 1 and the rest to 0.

$$p_{ij} = \frac{\exp(s_{ij})}{\sum_{k=1}^c \exp(s_{ik})}, s_{ij} = p_{ij}. \quad (2)$$

With the newly generated weak semantic tokens, I and T and the updated S are input again into the deep neural network model and trained again. The above process of weak semantic token generation and retraining can be iterated multiple times. When the final iteration is over, the final outputs IO and TO of the deep neural network corresponding to the image and text, respectively, can be obtained, they are regarded as the final shared semantic space. Now, the underlying features of the image and text have been projected to the common sub space, the projection looks like this:

$$M_I : I_U^{u \times p} \rightarrow O_I^{u \times c}. \quad (3)$$

$$M_T : T_U^{u \times q} \rightarrow O_T^{u \times c}. \quad (4)$$

The similarity between picture samples and text samples can be directly assessed once the model obtains the common subspace of images and texts, and the samples that are most similar to the query samples are fed back to the user.

B. Network model structure

The output of the deep neural network model translates the underlying features of images and texts to the semantic space [12]. The underlying picture and text features serve as the input, and the semantic space serves as the output. As a result, the network model's input layer's number of neurons represents the dimension of the underlying features of texts and images, the output layer's number of neurons represents the dimension of semantic space, and the hidden layer's number of neurons is set to adhere to the principle of decreasing layer by layer. As a result, Figures 2, 3, 4 and 5 demonstrate, respectively, the topologies of the deep neural network models employed in the tests on the Wikipedia dataset, NUS-WIDE dataset, Wikipedia-CNN dataset, and INRIA-Web search-CNN dataset. The experiment trains an image network model with a structure of 128-70-30-10 on the Wikipedia dataset, and a text network model with a structure of 10-10-10-10 on the NUS-WIDE dataset. It also trains an image network model with a structure of 4096-1000-500-100-10 on the Wikipedia-CNN dataset, and a text network model with a structure of 100-50-20-10 on the Wikipedia-CNN dataset.

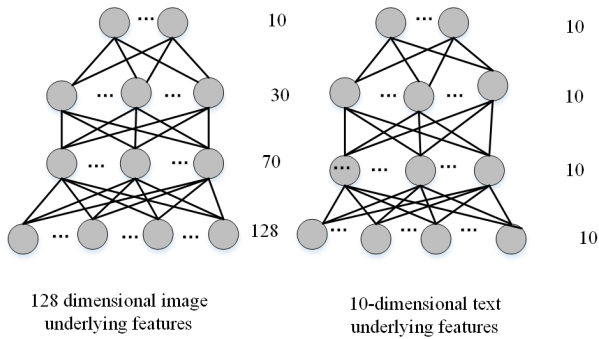


Figure 2: Network structure on Wikipedia dataset

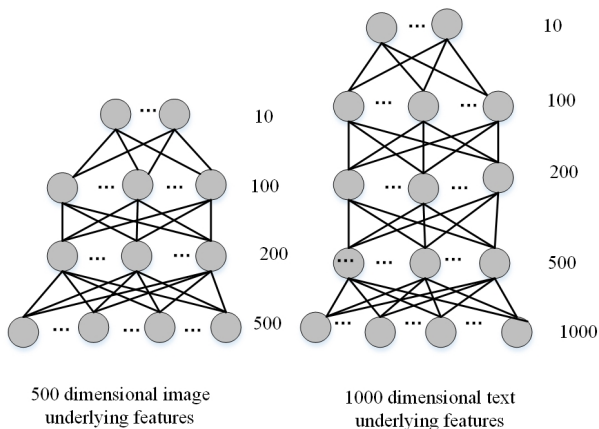


Figure 3: Network structure on nus-wide dataset

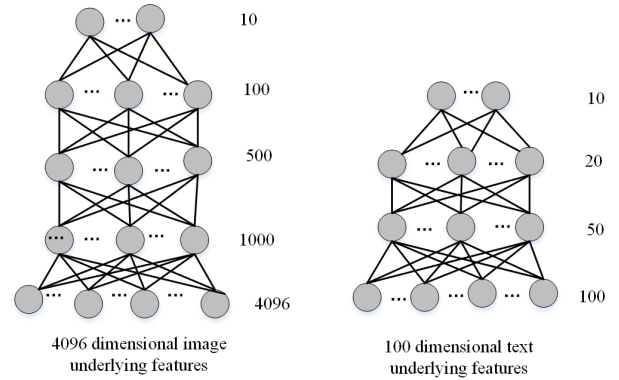


Figure 4: Network structure on Wikipedia CNN dataset

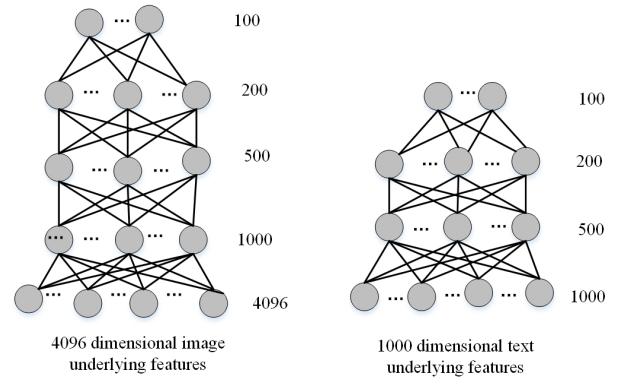


Figure 5: Network structure on INRIA web search CNN dataset

C. Cross-modal multimedia semantic matching based on deep neural network

The underlying properties of the image and text are mapped to their shared semantic space before the image and text are mutually examined (as seen in Figure 6).

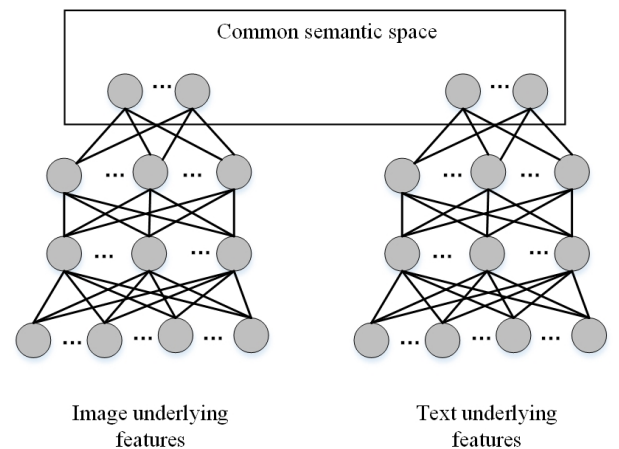


Figure 6: Cross modal multimedia semantic matching diagram based on deep neural network

After projecting the underlying features of the image and

Input	Train the underlying features and labels of images and texts, and test the underlying features of images and texts
Output	Cross modal multimedia retrieval results
1)	Create image network N_I and text network N_T
2)	Input the bottom features and marks of the training image into N_I and train the network, and input the bottom features and marks of the training text into N_T and train the network
3)	Input the bottom features of the test image into the trained network N_I , and input the bottom features of the test text into the trained network N_T
4)	Obtain the top-level output of network N_I and N_T , which is the common semantic space of image and text

Table 1: Cross modal multimedia semantic matching algorithm based on deep neural network

Method	Distance function	12T	T2I	Avg
MRCR-RSNN	L1	21.74%	21.03%	21.39%
MRCR-RSNN	L2	20.72%	21.17%	20.95%
MRCR-RSNN	LKL	22.33%	21.32%	21.81%
MRCR-RSNN	NC	25.34%	21.47%	23.42%
MRCR-RSNN	CC	30.02%	21.35%	25.69%

Table 2: Retrieval map values obtained by macruran applying five distance functions on Wikipedia dataset

Data set	Task	CCA	PLS	BLM	MRCR-NN	MRCR-SNN	MRCR-RSNN
Wikipedia	12T	18.23	23.76	25.63	26.54	28.12	30.02
	T2I	20.93	17.24	18.36	19.37	20.19	21.35
	Avg	19.63	20.51	21.22	22.95	24.15	25.69
Nus-wide	12t	28.74	27.53	29.77	29.86	30.67	31.22
	T2I	28.41	26.63	28.11	28.61	28.73	29.96
	Avg	28.57	27.08	28.96	28.94	29.25	29.72
Wikipedia-CNN	12t	22.63	25.06	26.63	30.82	32.72	34.22
	T2I	24.64	25.24	26.63	27.36	29.21	31.93
	Avg	23.65	25.15	26.48	29.09	30.96	33.07

Table 3: Map values retrieved by macruran, CCA, pls and BLM on Wikipedia dataset, nus-wide dataset, and Wikipedia CNN dataset

text into the semantic space of the two, a query text is obtained, and the retrieval model can obtain its semantic representation π_T^k . Most Relevant Images:

$$D(T, I) = \text{distance}(\pi_T^k, \pi_I^k). \quad (5)$$

Similarly, the same is true for image query related text. The cross-modal multimedia semantic matching algorithm based on deep neural network is shown in Table 1.

III. Experiments

A. Evaluation indicators

Centred Correlation, Normalised Correlation, KL (Libeler) divergence, L2, and L1 are some of the current distance functions used in cross-modal multimedia retrieval [13]. The experiment applies 5 distance functions to the Wikipedia dataset using the MRCR-RSNN approach, and the resulting MAP values are displayed in Table 2 to help determine which distance function is the most appropriate. The best outcomes are obtained at the CC distance. CC distance is therefore used as the distance metric function in subsequent studies.

B. Quantitative evaluation

On the Wikipedia dataset, NUS-WIDE dataset, and Wikipedia-CNN dataset, the tests compare the aforementioned approach with a number of different unsupervised cross-modal multimedia retrieval methods, and the retrieved MAP values are displayed in Table 3. According to experimental findings, MRCR-RSNN greatly enhances retrieval performance and produces superior results on the Wikipedia dataset, NUS-WIDE dataset, and Wikipedia-CNN dataset.

The Wikipedia dataset was created to address issues with cross-modal multimedia retrieval. Figure 7 illustrates how the mentioned methods—MRCR-NN, MRCR-SNN, and MRCR-RSNN—as well as unsupervised comparison techniques—CCA, PLS, and BLM—are used in the Histograms of MAP values for each class retrieved from the Wikipedia dataset for three different tasks of image retrieval text, text retrieval image, and average retrieval performance [14], [15]. Figure 8 displays the retrieval precision-recall curves for both text and picture retrieval from the Wikipedia dataset using the aforementioned method and a comparison method. These findings demonstrate that the described methods outperform comparison methods in most semantic categories and precision-recall, with MRCR-RSNN having the best retrieval performance.

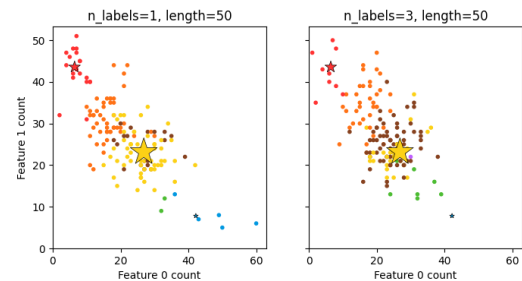


Figure 7: Histogram of map values of each class retrieved by macron, macruran, CCA, pls and BLM on Wikipedia dataset by image retrieval text (top), text retrieval image (middle) and average retrieval performance (bottom)

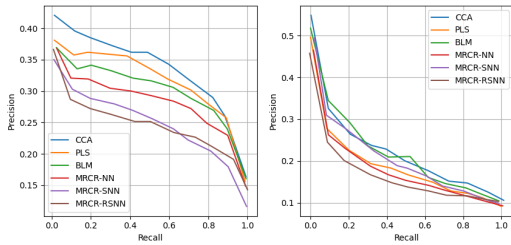


Figure 8: Quasi curvature recall curves of macron, macros, macruran, CCA, pls and BLM for image retrieval text (left) and text retrieval image (right) on Wikipedia dataset

A lot of training examples are present in the NUS-WIDE dataset, which is used in experiments to assess retrieval performance. Similar to Figure 8 and 9 displays the methods MRCR-NN, MRCR-SNN, MRCR-RSNN as well as unsupervised comparison methods CCA and PLS for three distinct tasks: average retrieval performance, text-image retrieval, and picture retrieval. Each type of MAP value found in the NUS-WIDE dataset is represented by a BLM histogram. Figure 10 displays the retrieval accuracy-recall curves for both text and picture retrieval from the NUS-WIDE dataset using the aforementioned method and the comparison method. These findings demonstrate that, for the majority of semantic categories as well as precision-recall, the aforementioned methods outperform the comparison methods in terms of retrieval performance.

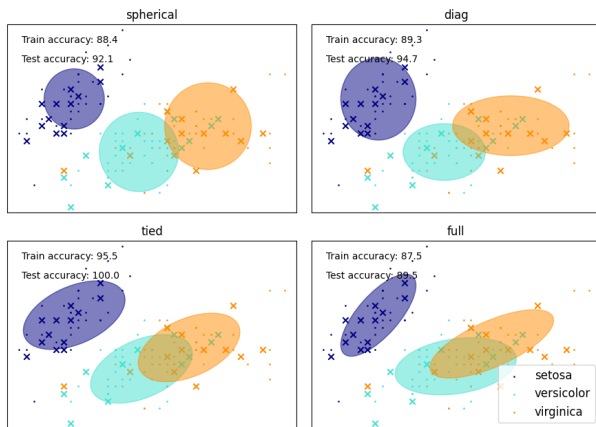


Figure 9: Histogram of map values of each class retrieved by macruran, CCA, pls and BLM on the nus-wadi dataset for image retrieval text (top), text retrieval image (middle) and average retrieval performance (bottom)

Convolutional Neural Networks (CNN), which can automatically extract image visual features without requiring human involvement, have recently been shown to have stronger feature representation in the field of computer vision [16]. Therefore, using the Wikipedia-CNN dataset, the trials will additionally test the aforementioned approaches. The Wikipedia-CNN dataset's histograms of per-class MAP values

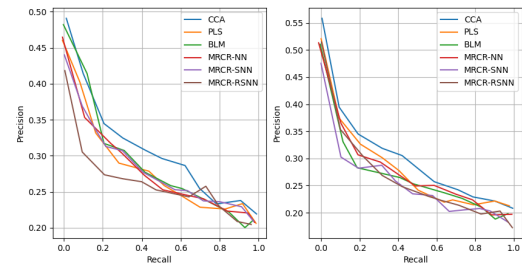


Figure 10: Recall of accuracy of mercer NN, mercer SNN, mercer RSNN, CCA, Pls and BLM in image retrieval text (left) and text retrieval image (right) on the NUS range dataset

for the three distinct tasks of extracting text from images, extracting images from text, and average retrieval performance for the aforementioned and comparable approaches are shown in Figure 11. Figure 12 displays the retrieval precision-recall curves for both text and picture retrieval from the Wikipedia-CNN dataset using the aforementioned method and a comparison method. The results once more demonstrate that, for the majority of semantic categories as well as precision-recall, the described methods outperform the competing methods in terms of retrieval performance.

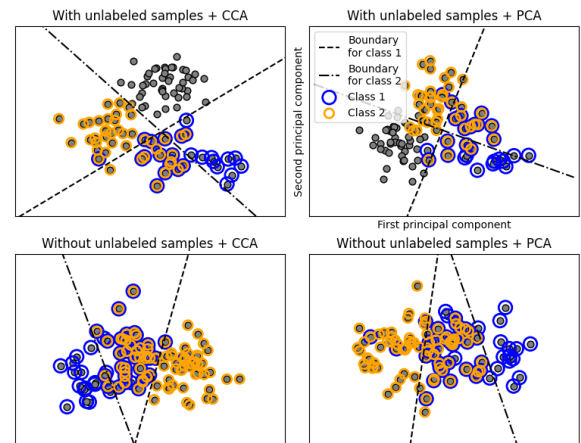


Figure 11: Mapping values (top), text retrieval images (middle) and average retrieval performance (bottom) of each class retrieved by mercer NN, mercer SNN, mercer RSNN, CCA, pls and BLM on the Wikipedia CNN dataset

IV. Conclusion

In this study, a neural network-based cross-modal multimedia model for modal reconstruction is proposed. To translate the underlying image data into the text feature space, we employ a deep neural network model. In particular, the test images' underlying features are fed into the trained network model after it has been trained using the training images and the text's underlying features, and the network's top-level output is then used as the text feature space. By doing this, the image's underlying features are immediately projected into

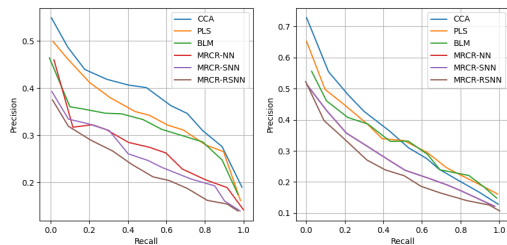


Figure 12: Accuracy recall curves of mercer NN, mercer SNN, mercer RSNN, CCA, pls and BLM in image retrieval texts (left) and (right) on Wikipedia CNN dataset

the feature space of the text, which is followed by the cross-media retrieval between the image and the text. In addition, we integrate the model with the university's teaching platform for the integration of innovation, entrepreneurship, and moral education to achieve quantitative and qualitative evaluation from theory to practise.

V. Funding

There is no funding support for this study.

References

- [1] Baer, M., Hoegl, M., & Weiss, M. (2022). The human side of innovation management: bridging the divide between the fields of innovation management and organizational behavior. *Journal of Product Innovation Management*, 39(3), 283-291.
- [2] Yz, A., & Hz, B. (2022). Research on the quality evaluation of innovation and entrepreneurship education of college students based on extenics. *Procedia Computer Science*, 199, 605-612.
- [3] He, K., Yao, L., Zhang, J., Li, Y., & Li, C. (2021). Construction of genealogical knowledge graphs from obituaries: multitask neural network extraction system. *Journal of Medical Internet Research*, 23(8), e25670.
- [4] Xu, Y., Tao, Y., Zhang, C., Xie, M., Li, W., & Tai, J. (2022). Review of digital economy research in China: A framework analysis based on bibliometrics. *Computational Intelligence and Neuroscience*, 2022, Article ID 2427034, 11 pages.
- [5] Huang, Y. (2021). Research on the application of computer technology in innovation and entrepreneurship education. *Journal of Physics: Conference Series*, 1744(4), 042095.
- [6] Li, Y. (2021). Research on the construction of college English mixed teaching model based on modern educational technology and computer technology. *Journal of Physics: Conference Series*, 1915(2), 022091.
- [7] Costa, R. D., Dias, A., Cascais, E., Pereira, L., & Gonçalves, R. (2022). Lifestyle entrepreneurship innovation and self-efficacy: exploring the direct and indirect effects of marshaling. *International Journal of Tourism Research*, 24(3), 443-455.
- [8] Grave, W. D., Cantillon, P., & Dornan, T. (2022). The social construction of teacher and learner identities in medicine and surgery. *Medical Education*, 56(6), 614-624.
- [9] Wang, M., Rasoolimanesh, S. M., & Kunasekaran, P. (2022). A review of social entrepreneurship research in tourism: knowledge map, operational experiences, and roadmaps. *Journal of Sustainable Tourism*, 30(8), 1777-1798.
- [10] Liang, C., Liu, J., Xue, X., Anwar, F., & Cao, L. (2022). Countermeasures to achieve high-quality development of dairy industry in Xinjiang: A case study of Tianrun Dairy Industry. *Asian Agricultural Research*, 14(5), 5.
- [11] Maier, D. (2022). The use of wood waste from construction and demolition to produce sustainable bioenergy—a bibliometric review of the literature. *International Journal of Energy Research*, 46(9), 11640-11658.
- [12] Dobrova, V. Y., Popov, O. S., Shtrimaitis, O. V., Andreeva, O. O., & Proskurnia, O. M. (2022). Joint task force core competency framework adoption process at a national level: a survey of Ukrainian-based clinical research

professionals. *Therapeutic Innovation & Regulatory Science*, 56(5), 814-821.

- [13] Mavilia, R., & Pisani, R. (2022). Blockchain for agricultural sector: the case of South Africa. *African Journal of Science, Technology, Innovation and Development*, 14(3), 845-851.
- [14] Lines, B. C., Kakarapalli, R., & Nguyen, P. H. D. (2022). Does best value procurement cost more than low-bid? A total project cost perspective. *International Journal of Construction Education and Research*, 18(1), 85-100.
- [15] Liu, J., Gong, E., & Wang, X. (2022). Economic benefits of construction waste recycling enterprises under tax incentive policies. *Environmental Science and Pollution Research*, 29(9), 12574-12588.
- [16] Loi, M., Fayolle, A., van Gelderen, M., Riot, E., Refai, D., Higgins, D., ... & Cavarretta, F. (2022). Entrepreneurship education at the crossroads: challenging taken-for-granted assumptions and opening new perspectives. *Journal of Management Inquiry*, 31(2), 123-134.

...