Proposal and Evaluation of a Course-Classification-Support System Emphasizing Communication with the Sub-committees Within the Committee of Validation and Examination for Degrees

Kazuteru Miyazaki(✉), Syu Yamaguchi, Rie Mori, Yumiko Yoshikawa, Takanori Saito, and Toshiya Suzuki

National Institution for Academic Degrees and Quality Enhancement of Higher Education, Tokyo, Japan
{teru,yamaguti-s,rmori,yosikawa,tknrsaito,suzuki-t}@niad.ac.jp
http://www.niad.ac.jp

Abstract. The National Institution for Academic Degrees and Quality Enhancement of Higher Education (NIAD-QE) awards academic degrees based on the accumulation of credits. These credits must be classified according to pre-determined criteria for the chosen disciplinary field. This work has been carried out by the sub-committees within the *Committee of Validation and Examination for Degrees* (CVED), whose members should be well-versed in the syllabus of each course to ensure appropriate classification. The number of applicants is increasing every year, and thus, a course classification system supported by information technology is strongly desired. We have proposed the *Course Classification Support system* (CCS) and the *Active Course Classification Support system* (ACCS) for the awarding of degrees in NIAD-QE. On the other hand, in this paper, from the standpoint of emphasizing communication with the sub-committees, we construct a course classification support system using deep learning, which has been developing remarkably in recent years. We also confirm the effectiveness of the proposed method using actual syllabi from two universities.

Keywords: syllabus · course-classification · degree awarding · recommender system · deep learning

1 Introduction

NIAD-QE is an incorporated administrative agency under the Ministry of Education, Culture, Sports, Science and Technology (MEXT). NIAD-QE's missions include contributing to the development of life-long learning society by enabling learners who are not university students to earn academic degrees. Assessing

Y. Kambayashi et al. (Eds.): AICON 2022, LNICST 477, pp. 123–130, 2023.
https://doi.org/10.1007/978-3-031-29126-5_10

the results of various learning provided at the higher education level, NIAD-QE awards Bachelor's degrees to learners who have already acquired sufficient number of academic credits to make sufficient academic achievements in the particular disciplinary field.

In this degree-awarding process, an applicant who wishes to earn a degree from NIAD-QE must classify courses he/she took according to pre-determined criteria in each disciplinary field, being publicized in the guidebook called *Alternative Routes to a Bachelor's Degree* (https://www.niad.ac.jp/media/005/202202/gakushi_annai.pdf). Validity of applicants' course classifications are judged by sub-committees of CVED whose members are university professors in each field. Those judgements are made by sub-committee members based on syllabi: It is examined whether the classification by the applicant is appropriate or not. NIAD-QE awards Bachelor's degrees to about 2,500 applicants each year. One challenge is that the examination of syllabus of each course is significantly time-consuming.

It has been, therefore, desired to reduce the workload in examining syllabi without compromising the accuracy of judgement since syllabi are important material for assessments in academic areas. The focus of our research is to classify individual syllabus based on its content by utilizing information technology for the degree-awarding process at NIAD-QE. For this purpose, we have proposed CCS [6] and ACCS [7]. Following up those attempts, in this paper, we propose a method using the deep learning that emphasizes communication with the sub-committee for better results.

2 Degree-Awarding of NIAD-QE and the Course-Classification-Support System

NIAD-QE's awards academic degrees in two schemes: One is based on the successful completion of a junior college program (or its equivalent) and the accumulation of sufficient credits additionally earned (Scheme I); The other one is based on the successful completion of an NIAD-QE-approved program being provided by an educational institution that has been established legally, yet operates under

Table 1. All disciplinary fields in the degree-awarding by NIAD-QE in 2022

Literature	Commerce	Nutrition
Education	Business Administration	Engineering
Theology	Science	Design
Sociology	Pharmaceutical Sciences	Mercantile Marine Science
Liberal Arts	Nursing	Agriculture
Social Sciences	Health Science	Fisheries
Law	Acupuncture and Moxibustion	Home Economics
Political Science	Oral Health Science	Art
Economics	Judo Therapy	Physical Education

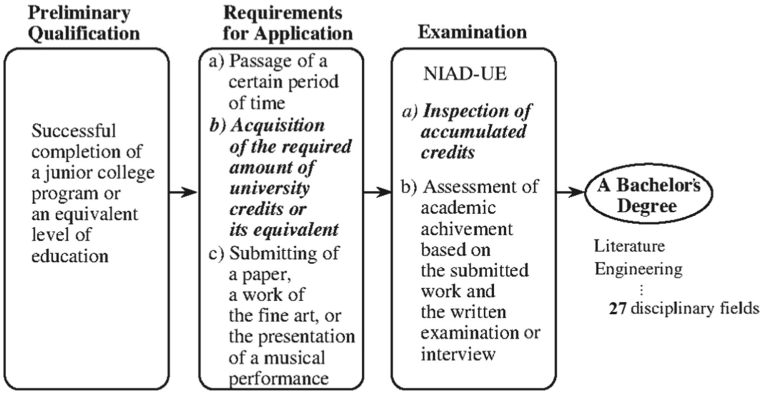

Fig. 1. Scheme for awarding degrees based on credit accumulation (Scheme I).

Table 2. Criteria for the *Information Engineering* sub-field.

Item 1	information engineering basic theory (4+ credits)
Item 2	computer system (4+ credits)
Item 3	information processing (4+ credits)
Item 4	relevant to information
Item 5	exercise and experiment about information engineering
Item 6	related course (4+ credits)
Item 7	others

the jurisdiction of ministries other than MEXT (Scheme II). While a Bachelor's degree can be earned under either of these two schemes, Masters and Doctoral degrees are awarded only under Scheme II. In this paper, we focus on the evaluation process in Scheme I. Figure 1 shows the procedures of application and evaluation for awarding degrees under Scheme I.

Although there exist similar systems as Scheme II in some other countries, the Scheme I of NIAD-QE is a unique system in the world [9,10,12,13]. To assist applicants, NIAD-QE annually publishes the application guidebook for Scheme I. In this paper we refer the 2022 version of it. In 2022, Bachelor's degrees are awarded in 27 disciplinary fields, as shown in Table 1.

An applicant has to demonstrate that he/she has accumulated the appropriate number of credits according to the criteria which are determined in each discipline. Some disciplinary fields are divided into sub-fields. When the field consists of two or more sub-fields, separate credit requirements are specified in each sub-field. Otherwise, the field itself may be considered a sub-field. For example, the *Engineering* field has nine sub-fields as follows: *Mechanical Engineering, Electrical and Electronics Engineering, Information Engineering, Applied Chemistry, Biological Technology, Materials Engineering, Civil Engineering, Architecture, Social Systems Engineering.* Each of these nine fields have its own set of criteria. On the other hand, the field of the *Nursing*, for instance, consists of the only sub-field of *Nursing* with one set of criteria. In this paper, we focus on the field of *Engineering*, particularly the sub-field of *Information Engineering*. The criteria for the sub-field of *Information Engineering* are prescribed as shown in Table 2.

As mentioned above, the course classification by an applicant to fulfill the criteria is not always valid. So, the validity of the course classification by applicants is judged, by the sub-committee of CVED, at the level of the sub-field. Sub-committee members have to take great deal of time to examine the syllabus of each course. Through recent years the number of applicants remains high, so that it is necessary to assist the judgment of the sub-committee by using information technology.

3 The Course-Classification-Support System Using the Deep Learning

3.1 Preparation of Training Data and Previous Methods

In 2002, using the Internet, Miyazaki et al. [6] collected 962 syllabi of the courses related to the *Information Engineering* curriculum offered by 13 universities in Japan. They read these syllabi and classified into the items listed in Table 3.

Table 3. Results of the manual classification.

Item	1	2	3	4	5	6	7
No. of syllabi	108	104	85	54	86	351	48

The papaers [6,7] proposed CCS and ACCS using these data. These systems are of traditional morphological analysis-based methods. It is therefore necessary to perform appropriate pre-processing such as data cleaning in advance, which places a heavy burden on the user.

On the other hand, in this paper, we construct a deep learning system using the syllabus of Items 1 to 6 as training data. It is assumable that one syllabus is related to multiple items. We, therefore, consider formulating the system to process multi-label problems to allow one syllabus to belong to multiple items.

3.2 Relationship with Related Works and Our Approach

The tasks dealt with in this paper require natural language processing. In particular, the problem dealt with in this paper can be regarded as a text classification problem. Various methods such as support vector machines (SVMs) have been proposed to solve classification problems in the machine learning field. Deep learning among them has been attracting attention in recent years, and there are increasing number of researches. Though deep learning is known to be powerful for image classification [3] and game problems [4,8], it has been applied to many classification problems that require natural language processing lately.

The natural language processing technology has made remarkable progress in recent years with the development of distributed representation technologies, such as word2vec [5]. In general, when deep learning is applied to natural language processing, the problem is how to configure the input to the network. There are cases where a sentence is divided into word units and the output value of word2vec is used as the input of the network. On the other hand, we use a more general technique called the Character-level CNN (CLCNN) [17].

With this method, the text is divided into character units, each of which is converted to a character code (e.g., Unicode value) and inputted to the network. As a result, the whole text can be treated as an image. It becomes advantageous to be able to use the results of many research studies on image processing. CLCNNs have been used, for example, in Retty (https://retty.me) where a web service allows users to find good restaurants.

In recent years, transformer [14] have been attracting attention in the field of text processing. Transformer is basically a method of learning the degree of association between words. It cannot be therefore directly applied to CLCNN that processes character by character. It also should be noted that pre-processing such as data cleaning, which has been done in CCS and ACCS, is important, especially in Japanese language processing since learning is performed word by word. Furthermore, from the point of view of the classification problem, existing methods such as SVM can be used instead of CNN. We decided to use CLCNN, which has many achievements as a method of inputting natural language character by character, since the problem of how a computer recognizes natural language also exists in SVM.

3.3 Proposed Method

The network consists of four layers, those are an input layer, a convolution layer, a pooling layer, and a fully connected layer. The network structure is shown in Fig. 2. The batch size and the number of epochs are 100 and 100, respectively. The Unicode value (UTF-8) was used when each character is inputted into the network. The first 500 characters in each syllabus were inputted and the rest of them were deleted. To syllabi with less than 500 characters, 0s were added to make them reach 500. For the same input, convolution was performed using multiple kernel sizes. Here, the kernel size was set to a width of the dimension size of one character and the height of n characters. As the value of n, four kinds of n = 2,3,4,5 were used. Through this, we expect that the convolution results will be similar to those with an n-gram. After passing the convolution results with multiple kernels to the pooling layers, they are passed through a fully connected layer consisting of 64 units to finally obtain 4 outputs (items 1, 2, 3 and 4), corresponding to the presence or absence of each item. In this study, we performed a batch normalization and dropout (the rate is 0.5) processes on the output of the fully connected layer.

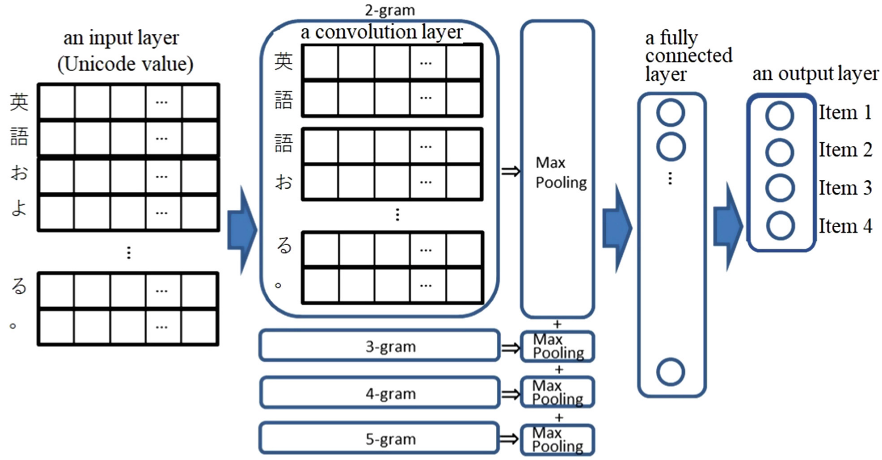

Fig. 2. Network structure.

4 Results and Discussion

Table 4 is the result of practice when the system was applied to the syllabus of courses provided in the field of *Information Engineering* at two national comprehensive universities: X University and Y University. The numbers of syllabi entered are 28 and 60, respectively. The tables show the number of syllabi that have been entered and then classified into the subject categories. In particular, the results of subject items 1 to 4 that are emphasized by the sub-committee are shown. The experiment was performed 10 times from No.1 to 10.

Although there were some variations in each experiment, it was confirmed that the results were close to those of the ACCS. ACCS has a high load of pre-processing such as data cleaning as already mentioned. It is therefore significant that CLCNN, which does not require such pre-processing, obtained results comparable to ACCS. In particular, we are considering application in actual sub-committee of CVED. The fact that the pre-processing can be reduced by deep learning is particularly important in actual operation, since now it takes a lot of preliminary work to carry out detailed preprocessing for all sub-fields.

In addition, when planning the application, it is important to shorten the actual work time. Essentially, our support system should be evaluated by comparing the work time with and without the use of it. But it is impossible to create an environment of accurate comparisons in practice of validation. We, therefore, compared the extent of reduction in the number of courses to be judged. In that sense as well, it is of great significance that we obtained results comparable to ACCS, which has already been confirmed to be more effective than CCS in terms of reducing the number of courses to be validated.

Table 4. Results of X and Y Universities.

	X University					Y University				
No	Item 1	Item 2	Item 3	Item 4	(Total)	Item 1	Item 2	Item 3	Item 4	(Total)
1	9	8	4	13	34	27	24	12	29	92
2	10	17	4	12	43	44	18	17	43	122
3	11	8	3	10	32	44	30	21	38	133
4	9	14	3	10	36	25	25	17	50	117
5	9	8	4	18	39	20	20	20	14	74
6	13	22	3	13	51	38	21	10	43	112
7	10	13	3	8	34	25	19	9	23	76
8	11	20	7	12	50	23	19	20	22	84
9	10	14	3	17	44	28	18	14	31	91
10	9	12	3	14	38	38	22	24	41	125
Average	10.1	13.6	3.7	12.7	40.1	21.2	21.6	16.4	33.4	103
ACCS	9	12	3	6	30	29	17	16	24	86

5 Conclusions

In this paper, we proposed a course-classification-support system using the deep learning. It was confirmed that the method using deep learning proposed in this paper can obtain the same level of performance as the conventionally promising ACCS.

We are planning to use this system in the actual practice of assessment by sub-committees to confirm its effectiveness in near future. It is necessarily to prepare training data first. The courses must be classified according to predetermined criteria. The criteria are determined in each discipline. Some disciplinary fields are divided into sub-fields. It is therefore necessary to prepare a large amount of training data that will be the correct classification for each discipline and sub-fields. However, majority of syllabi published before 2010s are not degitalized but paper-based. It is a barrier when we prepare the training data. We have hired dedicated staff to digitize the syllabi, that is actually judged and whose categories are known in order to overcome this problem.

Furthermore, it is essential to cooperate with the *Degree Validation and Examination Division*, that is responsible for the examination of degrees at NIAD-QE. NIAD-QE awards thousands of degrees each year and have extensive data on those applicants. It takes a lot of effort to extract the data necessary for the course classification support system proposed in this paper. Therefore, it will be important to create a system while taking into consideration the linkage with the huge amout of data archived by the Degree Validation and Examination Division.

Acknoledgement. We would like to thank the staff of the Degree Validation and Examination Division for providing the data for this study.

References

1. Ida, M., Nozawa, T., Yoshikane, F., Miyazaki, K., Kita, H.: Development of syllabus database and its application to comparative analysis of curricula among majors in undergraduate education. Res. Acad. Degrees Univ. Eval. **2**, 85–96 (2005). (in Japanese)
2. Itoh, E., Matsunaga, Y., Yamada, S., Hirokawa, S.: Auto DB construction from Web syllabus. In: The 17th Annual Conference of the Japanese Society for Artificial Intelligence, 1D4-08 (2003). (in Japanese)
3. Le, Q.V.: building high-level features using large scale unsupervised learning. In: Proceedings of the 29th International Conference on Machine Learning, 507–514 (2012)
4. Mnih, V.: Playing Atari with deep reinforcement learning. In: NIPS Deep Learning Workshop 2013 (2013)
5. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)
6. Miyazaki, K., Ida, M., Yoshikane, F., Nozawa, T., Kita, H.: On development of a course classification support system using syllabus data. Comput. Eng. I, 311–318 (2004)
7. Miyazaki, K., Ida, M., Yoshikane, F., Nozawa, T., Kita, H.: Proposal of the active course classification support system to support the classification of courses at the degree-awarding of NIAD-UE. In: Proceedings of 6th International Symposium on Advanced Intelligent Systems, pp. 685–690 (2005)
8. Miyazaki, K.: Exploitation-oriented learning with deep learning introducing profit sharing to a deep Q-network. J. Adv. Comput. Intell. Intelligent Inf. **21**(5), 849–855 (2017)
9. Mori, R.: System of assessment of learning outcomes in regents college: evaluation of learning and credit. Res. Acad. Degrees, **10**, 107–129 (1999). (in Japanese)
10. Mori, R.: The credit transfer system and the validation service at the open university. Res. Acad. Degrees, **17**, 183–198 (2003). (in Japanese)
11. Nozawa, T., Ida, M., Yoshikane, F., Miyazaki, K., Kita, H.: Construction of curriculum analyzing system based on document clustering of syllabus data. J. Inf. Process. **46**(1), 289–300 (2005). (in Japanese)
12. Puirseil, S., Quality assuarance in irish higher education - the higher education and training awards council. Res. Acad. Degrees **15**, 124–140 (2001)
13. Tachi, A.: A study on thomas edison state college, the external degree college established by the state of New Jersey. Res. Acad. Degrees **10**, 73–89 (1999). (in Japanese)
14. Vaswani, A., et al.: Attention is all you need. In: Neural Information Processing Systems (NIPS 2017) (2017)
15. Yamada, S., Itou, E., Hirokawa, S.: A study and analysis of auto collected web syllabus pages. Forum Inf. Technol. 2002 (FIT2002), **32**, 301–302 (2002). (in Japanese)
16. Yamada, S., Matsunaga, Y., Itou, E., Hirokawa, S.: A study of design for intelligent web syllabus crawling agent. Trans. IEICE. D-I, **86**(8), 566–574 (2003). (in Japanese)
17. Zhang, X, Zhao, J., LeCun, Y.: Characterlevel convolutional networks for text classification. arXiv:1509.01626 (2015)