Extraction of the Characteristic Attributes of Student Athletes using a Questionnaire using the Support Vector Machine

Toru Sugihara [*], Soichiro Aihara [†],

Sachio Hirokawa [‡], Takashi Nara [§]

Abstract

Providing special support for student athletes and creating future educational strategies are major issues for many universities. In this study, we created a questionnaire that comprised 77 multiple choice questions. After collecting the responses from 100 student athletes and 141 other students, we analyzed the characteristic attributes of student athletes. We considered the 278 combinations of question items and the response choices as the attributes to represent the students. A student is represented as a 278 dimensional boolean vector. Then, we applied the support vector machine (SVM) machine learning method with feature selection. The result confirmed that it is possible to distinguish between student athletes and other students with a 90% accuracy based on 16 characteristic attributes of the students, such as the following: they spend a lot of time in athletics clubs and not on their studies; they want to work for an economically rich life; they think that it is advantageous to search for jobs or complete their graduation if they have good grades; they have less interests in the international activities in campus life. Students aiming for professional and semiprofessional athletic careers are motivated for studying because of the competitiveness cultivated in their athletic careers.

Keywords: Private University; Questionnaire survey; Student Athletes; Support vector machine

1   Introduction

In the United States, the NCAA (National Collegiate Athletic Association) exists as an organization that supervises over college sports. The NCAA is represented by more than 1000 organizations including universities and supports student athletes. The NCAA has implemented various rules and guidelines. Particularly, the academic rules are strict, and students with low GPA scores, or less than a certain amount of credits are not allowed to participate in games or events [1]. In recent years, a campaign attempting to establish a Japanese version of the NCAA has been active.

The Asahi newspaper and Japanese Association of University Physical Education and Sports conducted a survey in 92 universities regarding special support for student athletes. The results

---

[*] Kanto Gakuin University, Yokohama, Japan
[†] Shibaura Institute of Technology, Saitama, Japan
[‡] Kyushu University, Fukuoka, Japan
[§] Kanto Gakuin University, Yokohama, Japan

showed that 65 universities answered considered that special academic support for student athletes is necessary. Moreover, 34 universities explained that their university had established regulations stipulating that students with poor performances cannot participate in club activities, and 39 universities responded by stating that their universities fostered sociality through communication with the community [2].

The Japanese Association of University Physical Education and Sports investigated large universities and athletic colleges. The results showed that the main expected outcome from universities in the case of club activities was personality growth and leadership training for student athletes (chosen by 96% of respondents), the second highest expected outcome was participation in extracurricular sports in general (chosen by 89% of respondents), and third highest expected outcome was enrichment of university life (chosen by 88% of respondents). According to 70% of the investigated universities, 1it is necessary to support learning for student athletes. Their main answers for the types of support needed were "Consideration of absences when they were at games (59%)," "Timetable organization for student athletes (19%)," "Time schedule consideration for practice time (19%)," and "Opening lessons for student athletes (14%)" [3].

The Waseda University Athlete center has been implementing the Waseda Athlete Program (WAP) since 2014 [4]. For the educational part of the programs, they have implemented educational programs for athletes, career support programs, volunteer/community contribution activities support programs, and international exchange support programs. Also, for the learning support part of the programs, they are implementing learning management systems, selecting of academic advisors, and distributing awards. Rikkyo University conducts leadership training for students in their tennis club [5].

Some small universities have a higher proportion of student athletes than large universities. Small universities do not have sufficient support for student athletes; therefore, we need to know the actual situation of student athletes in such universities. From this information, we can know the importance supporting student athletes.

We conducted a survey on university students to identify the reasons that they are studying at university, their expected career after graduation, the abilities that they have developed, and the manner in which they use each week. Therefore, we aim to reflect the result of these surveys in future educational policies.

In this paper, we answer the following research question by analyzing the responses to the questionnaire we created.

[RQ] What questionnaire responses distinguish student athletes from other students? What types of attributes distinguish student athletes from other students?

2   QUESTIONNAIRE ON UNIVERSITY LIFE

The study was conducted at five universities in June 2015 in the following manner. There were 241 respondents (100 student athletes, 141 other students).

These universities were private universities and did not have a sports department. The total admission capacities of Universities, A, B, C, D, and E were ~10,000, ~500, ~8500, ~11,000, ~1200.

There were 68 students (14 student athletes, 54 other students) at University A, 64 students (42 student athletes, 22 other) at University B, 36 students (36 student athletes, 0 other) at University C, 71 students (6 student athletes, 65 other) at University D, 2 students (2 student athletes, 0 other) at University E (Table1). The entrance examination deviation values of Universities B and E were ~40; those of Universities A and C were ~50, while the examination deviation value of University D was ~60.

There were 65 first year students (19 student athletes, 46 other), 89 second year students (45 student athletes, 44 other, 67 third year students (24 student athletes, 43 other, 17 forth year students (12 student athletes, 5 other (Table2).

<Date> 2015/6
<University> 5 university students
<Number> 241 people (100 student athletes,141 other students)
<Survey method> questionnaire

Table 1. Universities

	student athletes	Other students	total
A University	14	54	68
B University	42	22	64
C University	36	0	36
D University	6	65	71
E University	2	0	2
total	100	141	241

Table 2. School year

	Student athletes	Other students	total
First year	19	46	65
Second year	45	44	89
Third year	24	43	67
Fourth year	12	5	17
total	100	138	238

 (no answer=3)

In the survey, the questions were based on the following parameters: "University/Department/School year/Gender," "Athletic club," "Athlete career," "Reason for studying at university," "Attitudes toward work," "Future outlook," "Weekly Schedule," and "Competence."

We referred to the two factor model of learning motivation theory that Ichikawa proposed [6] and created 12 possible answers for the "Reason for studying at university" question. For the Attitudes toward work question, we used 17 statements created by Benesse. The

answered were classified as follows: "Setting a goal," "Utilitarian learning oriented," "Social self-realization oriented," "Pursuit of private value," "Earn-oriented," "Moratorium oriented," "Comfort-oriented," "Negative self image." For the "Future Outlook" question, we used six statements that Benesse created. These answers were "I have hope in the future," "There is a dream job," "I want to do a big job," "I have clear goals in the future," "I have important things in my career selection," and "I understand my ability and aptitude" [7]. The answers used a four level scale where "very true" was 4 points, "true" was 3 points, "little" was 2 points, and "never" was 1 point.

Therefore, we again referred to statements that Benesse created [8], and created 22 statements for "Competence." These statements also used a four-point scale for gathering answers. Advanced learning level was 4 points, competent learning level was 3 points, little learning level was 2 points, and a poor learning level was 1 point. Additionally, we created 11 sample activities for a typical week schedule,; these activities included reviews, club activities, and watching TV.

We analyzed the items on the questionnaire to distinguish between student athletes and other students using the support vector machine (SVM).

By considering the effects of discrete SVM analysis, the questionnaire in this paper comprises six groups of questions that comprised 5–22 small items. There were 77 items in total. Each item asked for the experience in years, the time they spend, and the degree of empathy. Since most of those responses were numerical values, we could consider the data as 77 dimensional data items and subsequently apply SVM. However, we could not determine the question item that had a strong influence on the prediction score. We introduced 278 dimensional vector representations, used the SVM score of those words for feature selection, and succeeded in improving the prediction performance by 20%. These vector representations gave us an insight into the student athletes.

3 ANALYSIS OF QUESTIONNAIRE BY SUPPORT VECTOR MACHINE

3.1 Questionnaire Response as Attribute

In this paper, we propose a method to convert the responses of questionnaires into vector representations. To apply the SVM method, we represented each student by a 278 dimensional vector whose value was either 1 or 0, i.e., a boolean value. Each component of the vector represented the response of a student. For example, if a student chose "2" for question 2-11, then for the component for 2-11:2 became 1. If another student chose "3" for question 2-11, then the component for 2-11:2 for this student became 0. Thus, we analyzed the students for those 278 attributes using SVM.

For example, consider question Q4 that asks how a student spends time for a week with 11 choices, such as attending classes, reviewing, doing a part time job, and playing games, and asks how long does the student spend performing these activities. The duration of time is listed as follows:

1: 0 h	5: 6-10 h
2: less than 1 h	6: 11–15 h
3: 1–2 h	7: 16–20 h
4: 3–5 h	8: longer than 21 h

Consider the case in which a student responded to the second question of Q4 by selecting "5: 6–10 hours." Then, this information from the student is registered as attribute "Q4-2:5." As another example, question Q2 asks the reason why a student wants to learn at the University. The question had 12 statements and asked the manner in which each statement fit into the four levels as follows:

4:very true,
3:true,
2:somewhat true, and
1: false.

The sixth statement was "I want to make use of what I learned for my future work." If a student selected the choice of "3:true" for the sixth statement then the attribute "Q2-6:3" is registered as his/her information.

In this manner, all the responses of a student are transformed into the attributes in a format of "QX-Y:Z" with numbers for X,Y, and Z. The information of 241 students was then stored as 241 vectors that contained those attributes.

3.2 SVM and Feature Selection

We applied the machine learning method SVM to these data. We used the documents for sports-oriented students as positive documents and the others as negative documents. We used SVMlight [11] for our experiment.

Since we want to know the characteristics of sports-oriented students, we adopted the feature selection method introduced by Sakai and Hirokawa 2012 [9] and Adachi et. al. 2016 [10]. The process comprises the following two steps.

Step 1 Calculation of the SVM score for each word

In the first step, all of the documents are represented as vectors using all words. SVM generates a linear kernel model, and each document d can be predicted if it is positive or negative using the following prediction function:

$$f(d) = \sum_{i=1}^{i=M} v_i * \delta(w_i, d) + b,$$

where v_i = weight (w_i) is the weight of the word wi, and b is a constant, (w_i,d) denotes if the word w_i occurs in d or not, i.e., $(w_i,d) = 1$ if w_i in d else 0. We call the weight (w_i) as the SVM score of the word w_i. Note that weight $(w_i) > 0$ if the word is a characteristic word of a positive document and that weight $(w_i) \leqq 0$ if the word is a characteristic word of the negative document.

Step 2 Vectorization of the top N words

As feature selection, we chose the top N positive words and top N negative words with respect to the SVM score. With those 2*N words, we vectorized the 241 documents and applied SVM. We followed the standard evaluation procedure (five-fold cross-validation) to predict the performance of the proposed method. We randomly separated the 241 documents into five groups and chose documents comprising four groups for training. Then, we applied the model to predict whether a document was positive. The performances, precision, recall, F-measure, and accuracy were calculated compared with the baseline truth. We iterated this experiment five times and calculated the average. We ran this process for N = 1, 2,...,10, 20,..,100, 200.

3.3 Prediction Performance with Feature Selection

Figure 1 displays the prediction performance with respect to the number N of selected features. Note that all the measures, i.e., precision, recall, F-measure, and accuracy, attained the highest score when N = 8. Moreover, this score outperformed the baseline with all words over 20%. This is shown in Table 3 below.
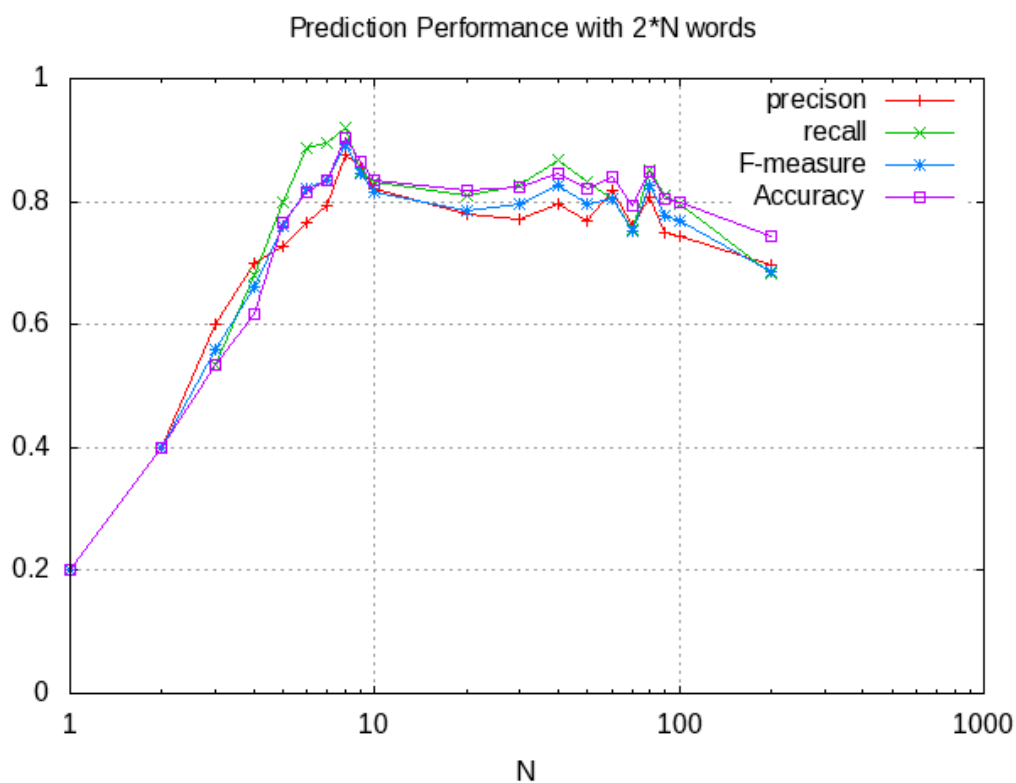

Figure 1. Prediction Performance

Table 3. Prediction performance

	precision	recall	F-measure	accuracy
All Words	0.6959	0.6831	0.6852	0.7425
Best Feature	0.8760	0.9200	0.8925	0.9044

Selection (N = 100)

3.4 Characteristic Features of Student Athletes

In this study, we extracted 16 characteristic words that distinguish between student athletes and other students from the SVM analysis (Tables 4, 5). We described the actual condition of students based on the characteristic words. Tables 4 and 5 list the top eight featured words. A "positive word" is used by student athletes, while a "negative word" is used by other students.

Table4. Positive word

rank	score	count	word	interpretation
1	0.0661	76	Q4-6:1	No partime job
2	0.0524	34	Q4-5:8	over 21 hours on club
3	0.0428	95	Q6-21:2	not gained internationality
4	0.0408	77	Q3-12:4	job for lush life
5	0.0388	38	Q4-2:1	no study
6	0.0374	54	Q4-9:4	TV/DVD 3-5 hours
7	0.0356	12	Q4-5:7	10-20 hours on club
8	0.0347	96	Q2-3:3	study for employment

Table5. Negative word

rank	score	count	word	interpretation
1	-0.0798	59	Q4-5:1	No club
2	-0.0454	110	Q2-9:3	study for employment
3	-0.0360	26	Q6-21:1	not gained internationality
4	-0.0353	86	Q2-3:4	study for employment
5	-0.0309	41	Q4-2:4	3-5 hours for study
6	-0.0297	23	Q6-18:4	gained logical thinking
7	-0.0281	34	Q4-6:4	3-5 hours for part time job
8	-0.0275	77	Q2-4:2	not interested in study

First, regarding the motivation to study at college question, both student athletes and the other students tended to learn with external motives that are more favorable to employment and admission than internal motives, such as intellectual curiosity. That trend is clear for the other student's category. There are many student athletes who answered (Q2-3:3) "I think that it is advantageous for search for jobs or complete graduation if I have good grades: true". Many of the other student category answered (Q2-3:4 very true), (Q2-9:3) "I gain the benefit if I get a good job: true," and (Q2-4:2) "I think that studying itself is interesting even if it is not useful immediately: little."

Second, from the perspective of their careers, student athletes tend to think that work is the means to live an economically rich life. Many student athletes answered (Q3-12:4) "I want to work for an economically rich life."

Third, regarding a one-week schedule student athletes tend to spend much time in club activities. Many student athletes have answered (Q4-5:8, Q4-5:7) "club activities 21 h over: 16–20 h" and (Q4-6:1) "part time job: 0 h," (Q4-2:1) "prepare and review lessons and do homework: 0 h," and (Q4-9:4) "watch TV and DVD: 3–5hours." However, many from the other students category spend no time on club activities (Q4-5:1) "club activities: 0hour," spend 3–5 h preparing and reviewing lessons and doing homework (Q4-2:4), and 3–5 h doing part time jobs (Q4-6:4).

Finally, for self-assessment of competence, many student athletes answered (Q6-21:2) "I acquire an international perspective: little learned." Many other students have answered (Q6-18:4) "I solve the problem in logical: Very learned," and (Q6-21:1) "I acquire an international perspective: never learned."

3.5 Characteristic Features of Student Athletes（Exclude Q4: A Student Spends Weekly Time）

Based the analysis results, we assume that Q 4 is the factor that distinguishes physical student athletes from other students (the manner in which a student spends their time each week). It is natural that student athletes spend a lot of time on club activities and other students spend a lot of time on part time jobs. Therefore, we excluded Q4 from the analysis, and we revealed the characteristics of student athletes further.

Note that all the measures, i.e., precision, recall, F-measure, and accuracy, attained the highest score when N = 8 (Figure 2, Table 6).
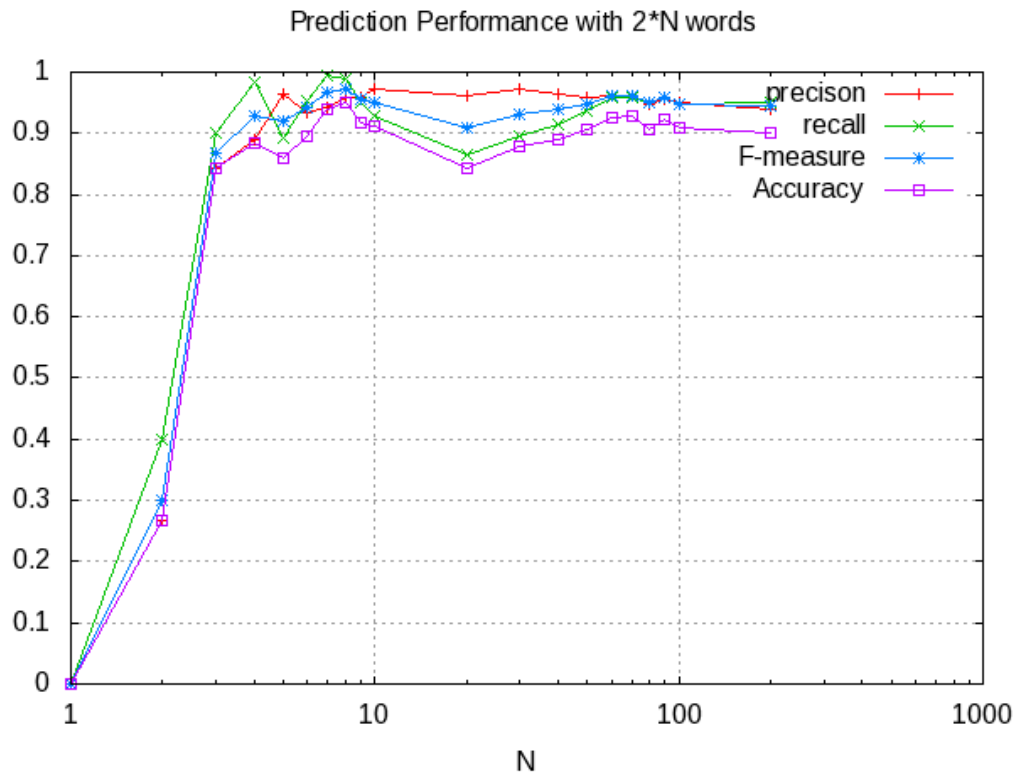
Figure 2. Prediction performance

Table 6. Prediction performance

N	precision	recall	F-measure	accuracy	
1	0.0000	0.0000	0.0000	0.0000	
2	0.2667	0.4000	0.3000	0.2667	
3	0.8429	0.9000	0.8667	0.8429	
4	0.8900	0.9833	0.9274	0.8833	
5	0.9655	0.8922	0.9191	0.8608	
6	0.9346	0.9535	0.9421	0.8944	
7	0.9430	0.9939	0.9672	0.9403	
8	0.9575	0.9901	0.9728	0.9506	Best Feature
9	0.9600	0.9516	0.9549	0.9177	
10	0.9575	0.9901	0.9728	0.9506	

Selection (N = 100)

In this study, we extracted 16 characteristic words that distinguish student athletes and other students based on SVM analysis (shown in Tables 7,8). Positive words showed that the respondents were student athletes, while negative words showed that they were other students.

Table 7. Positive word

rank	score	count	word	interpretation
1	0.2230	96	Q2−3:3	study for employment
2	0.2049	77	Q3−12:4	job for lush life
3	0.1985	132	Q3−8:4	study more about interests
4	0.1887	95	Q6−21:2	not gained internationality
5	0.1840	77	Q2−9:4	study for employment
6	0.1756	102	Q6−11:2	can not listen and speak foreign language
7	0.1738	93	Q6−18:3	gained logical thinking
8	0.1575	77	Q6−4:3	gained the ability to express ideas with figures and numbers

Table 8. Negative word

rank	score	count	word	interpretation
1	−0.2701	96	Q2−6:4	study for employment
2	−0.1795	110	Q2−9:3	study for employment
3	−0.1721	131	Q6−1:3	gained the ability to work with people
4	−0.1579	42	Q2−5:2	not study to think things in many ways
5	−0.1490	114	Q3−7:4	study to be useful in society
6	−0.1429	80	Q6−21:3	gained internationality
7	−0.1352	54	Q5−1:4	have hope in the future
8	−0.1333	23	Q6−18:4	gained logical thinking

We analyzed Tables 7 and .8 and identified the following four trends.

First, with respect to their career perspectives, student athletes tended to think that work is the means to live an economically rich life (Q3-12:4). Other students have hopes for the future (Q5-1:4). Both student athletes and other students answered (Q2-9:3.4) "I gain the benefits, if I get a good job."

Second, regarding learning at the university, student athletes want to learn about what they are interested in (Q3-8:4). Other students want to learn something useful for society (Q2-6:4,Q3-7:4), but they are not interested in thinking about things in many ways (Q2-5:2).

Third, for the self-assessment of competence, student athletes think to gain logical thinking skills (Q6-18:3) and the ability to express their ideas on charts in university life. Other students think about gaining logical thinking skills and cooperatation skills (Q6-1:3).

Finally, regarding internationality, student athletes think not to gain internationality and foreign language ability (Q6-21:2, Q6-11:2). However, other students answered to gain internationality (Q6-21:3).

3.6 Characteristics of Student Athletes Who Intend to Continue Competing at Professional and Semiprofessional Levels After Graduation

We asked 100 student athletes about their career as an athlete after graduation. Nine students are aiming to become professional athletes. There are 19 students who are aiming continue competing at a high level while working as in the industry, while 39 students plan to continue to compete as hobby/entertainment, and 23 students plan to retire from competing after their graduation. The other category was chosen by 10 students.

Table 9 shows the results from analyzing the characteristics of students aiming to become professional or semiprofessional athletes using SVM. The top 5 items are arranged in descending order of the difference, with the SVM scores differing largely by answers to the same question. The biggest difference was Q2-2 "I do not want to lose to rivals, so I study." The second is Q5-1 "I have hopes for the future," the third is "Q6-9 I have confidence and a positive feeling." Professional/semiprofessional oriented students answered "very true (4)," and the student who intends to retire/hobbies answered "little (2)."So the fourth is "Q4-1 Attendance at university classes." The fifth is "Q3-16 I want to live happily now rather than thinking about the future."

Table9. SVM score

rank	max-min	max	min	max	min	Question
1	0.2032	0.0908	−0.1124	4	2	Q2-2 I do not want to lose to rivals, so I study
2	0.1601	0.0946	−0.0655	4	2	Q5-1 I have hope in the future
3	0.1509	0.0991	−0.0518	4	2	Q6-9 I have confidence and positive feeling
4	0.1311	0.0640	−0.0671	6	5	Q4-1 Attendance at university classes
5	0.1301	0.0640	−0.0661	4	2	Q3-16 I want to live happily now rather than thinking about the future

We think that students aiming for professional and semiprofessional athletic careers have led to them being motivated to study based on the competitiveness cultivated during their athletic careers. There must be a very severe barrier to becoming a professional; however,

the students are confident about themselves, are not suspicious about the future, and are optimistic. However, their tendency to not deeply think about their future is owing to their narrow interest as they are driven to compete.

We showed that it is possible to use SVM as a new method in this type of research. We confirmed that it was possible to obtain similar results by applying boolean logistic regression analysis. We plan to describe detailed comparative analysis of the methods in the future.

4   CONCLUSION AND FURTHER STUDY

The results of this survey indicate that the characteristics of student athletes are as follows: (a) they spend a lot of time on athlete clubs and not on study, (b) they want to work for a economically rich life, (c) they think that it is advantageous for to search for jobs or complete graduation if they have good grades, and (d) they have less interest in the international activities of campus life. Also, students aiming for professional and semiprofessional athletic careers are motivated to study owing to the competitiveness cultivated during their athletic careers.

Student athletes spend a lot of time in club activities. Therefore, it is necessary to effectively use the club activity time. We think that construction of learning cycles in club activities is effective; therefore, it is important to educate the club coaches.
In addition, it is important to reform the curriculum based on the characteristics of student athletes.

After considering the results of this study, we need to reconsider the support for student athletes. We suggest the following four policies to support student athletes.

(A) Construction of learning cycles in club activities
Traditionally, many people think that student athletes have developed their personality through sports. They acquire good manners from factors, such as relationships with seniors, teamwork, and leadership, through daily practice. Therefore, they are highly appreciated by companies in many cases. However, the club activities are not an educational program in the regular curriculum.
The WAP textbook shows the importance of setting goals and action plans by incorporating the concept of the PDCA cycle into training [12]. Student athletes think about the purpose of exercise and training, record the result, and reflect on the result of their training; thus, they acquire problem solving and planning skills. We should consider club activities as project-based learning and active learning and incorporate educational viewpoints and methods such as those in a regular curriculum.

(B) Education for club coaches
The coach of club activities is professional in terms of teaching students how to compete, but their role as a "teacher" depends on their individual personality.

It is important for the managers to not only focus on technology but also emphasize on education. Therefore, universities need to hold training seminars for the managers.

(C) Curriculum and career education for student athletes

Universities need to consider establishing departments and courses for student athletes. For student athletes, it is important to educate them that completing their education at the university is useful in society and leads to a fruitful life.

(D) Formulation of educational policies for student athletes and club management policies

It is essential that these policies are formulated by reaching an agreement through discussions within the university. Based on these policies, education for the athlete student and club activities should be managed.

REFERENCES

[1] NCAA, http://www.ncaa.com/(accessed 8th April, 2017)

[2] Asahi Newspaper, (8th May, 2015), Bunbu ryodo shien tesaguri [Support for the study and club activities-Seeking]

[3] Japanese Association of University Physical Education and Sports,supotu/kurabu togososhiki to gakushushien/kyariashien ni kansuru chosakekka daijesuto [Sports/club management organization and study support/Career support survey result digest], 2015

[4]Waseda Athlete Program, https://www.waseda.jp/inst/athletic/wasedasports/program/(accessed 8th April,2017)

[5] Rikkyo University http://cob.rikkyo.ac.jp/blp/2567.html/(accessed 12th April,2017)

[6] Shinichi Ichikawa,Manabu iyoku no shinrigaku [Psychology of motivation to learn], PHP Institute, pp. 46-61, 2001

[7] Benesse Educational Reseach and Development Institute,Gakuseimanzokudo to daigakukyoiku no mondaiten 2007 [Student satisfaction and problems of university education 2007], 2008

[8] Benesse Educational Reseach and Development Institute, Dai2kai daigakusei no gakushu・seikatu jittaichosa hokokusho [Survey on college students learning and life Part 2], 2012

[9] Toshihiko Sakai, Sachio Hirokawa, Feature Words that Classify Problem Sentence in Scientific Article, Proc. iiWAS2012, 2012, pp. 360-367.

[10] Yusuke Adachi, Naoya Onimura, Takanori Yamashita, Sachio Hirokawa, Classification of Imbalanced Documents by Feature Selection, Proc. ICAIT 2017, 2017

[11] SVMlight, http://svmlight.joachims.org

[12] Waseda Athlete Program TEXTBOOK, 2016