Information Engineering Express
International Institute of Applied Informatics
2016, Vol.2, No.1, 97 – 106

Superior Factors to Distinguish Students to Be Cared in Introductory Programming Education

Hiromitsu Shimakawa, Dinh Thi Dong Phuong *

Abstract

Every student has its own motivation and learning strategies, which conform a learning status of the student. Appropriate supervision according to the learning status contributes to improvement of the learning of each student. Many of existing works try to figure out learning status directly from observable learning behavior. This paper proposes to utilize internal factors consisting of learning motivation and strategies, to distinguish learning status of students. It presents a way to derive the internal factors from records collected from their usual learning behavior, using the similarity of students over successive years. The experiment results indicates the strong possibility of the distinction from learning behavior. It implies the feasibility of immediate distinction of learning status of students, which enables efficient allocation of teaching power on the spot.

Keywords: contextual inquiry, learning status, learning behavior, motivation and learning strategies, persona

1 Introduction

In the introductory programming course, students are required to acquire skills to write codes, understanding various abstract concepts. In the programming education in high educational institutes like universities, a few teachers supervise large number of students. The environment makes it difficult to provide sufficient supervision for students suffering from difficulties in programming. If students fail to get supervision at the difficulties, they would get desperate, because they are at a loss as to what to do. The lack of supervision on the timing easily deprives their ambitions to learn programming.

Once students lose their ambitions to learn programming, its recovery is quite difficult, because the programming is a hard subject demanding to understand many abstract concepts. During programming education courses, teachers should monitor whether students maintain their ambitions to learn programming. However, the monitoring is infeasible, because teachers are too few to pay attention to whole students. We need a tool to automate the monitoring of learning status of students during the course.

Though many institutes conduct interviews or questionnaires for students to examine the learning status, we cannot expect them to bring a practical solution. It would costs

* College of Information and Science, Ritsumeikan University, Shiga, Japan

tremendously if we conducted interviews for all students. Repeated questionnaires during a course annoy students. Students tired of many questionnaires submit no serious answers, which prevents teachers from getting truthful data. It is necessary to know learning status of students from their usual learning activities without annoying them and at low costs.

Experienced teachers could know learning status of students from fine records of their learning activities. Nevertheless, they should focus on a few students who are likely to lose learning ambitions for programming. A way to find current students similar to past students who have lost learning ambitions would bring a promising solution to predict learning status of current students. The similarity calculated with records of learning activities of both students enables teachers to focus on students with signs of losing ambitions. Conventional works value learning behavior observable from outside as data to be collected from their usual learning activities without annoying them[6][10].

This paper propose a method to distinguish learning status of each student in programming education, to give supervision appropriate to the student. The method records learning behavior, which is external factors. It distinguishes the learning status of each student, using motivation and learning strategies as internal factors. It is necessary to distinguish learning status of each student immediately, because it changes all the time.

The method aims at timely distinction using 2 kinds of association. One is association of external factors with internal factors, while the other is that of internal factors with each student. Based on the similarity of students over years, the paper presents superior factors effective for the immediate distinction.

The paper figures out internal factor values of freshmen in actual programming courses in a university. In an experiment, it compares the values in successive years. A educational institute like a university usually admits students at a specific range in the degree of intelligence. When students of an institute are compared in different years, the proposed method finds many similar students in terms of internal factors. It suggests that internal factors make it possible to immediately distinguish learning status of current students with characteristic factors of past students.

The paper explains how to figure out of the 2 kinds of association. It also shows the experimental results.

2 external and internal factors

Since the number of teaching staff is small compared with that of student, it is infeasible for teaching staff to take care of all students equally. To realize effective programming courses, the teaching staff should distinguish students who can learn by themselves from those who should be cared by the teaching staff. This paper refers to the former as *active students*, while the latter as *passive students*. The attitude of a student for programming is referred to as *learning status* of the student in the paper.

2.1 Learning behavior - external factors

Students take various kinds of learning behavior under a given learning environment. It includes how long they engage in studying of a specific subject, when they submit home works, what kind of supplemental material they refer to, and so on. Since the learning behavior is observable from the outside, the paper refers to it as external factors.

In the conventional methods[6][8][10], external factors have been adopted as measures to know learning status of students. Observing external factors, instructors try to direct as many as students with course settings, aiming that students acquire programming skills. The instructional design methods such as [3] and [12] address the issue. The ARCS model proposed by Keller[7] is also considered to lead students to participation in learning with course settings which brings attention, relevance, confidence, and satisfaction.

2.2 Motivation and learning strategies - internal factors

The distinction of passive students from active students depends on their motivation and learning strategies. Psychologist researchers state "motivation is an internal state or condition (sometimes described as a need, desire, or want) that serves to activate or energize behavior and give it direction"[9]. Learning strategy is a person's approach to learning and using information to understand information and solve problems[15][13].

P.R.Pintrich at al. summaries motivation factors and learning strategy factors with detail explanations in MSLQ [11]. The motivation factors assess students' goals and value beliefs for a course. They consist of intrinsic, extrinsic, task value, control belief, self-efficacy, and affective components. The learning strategy ones regard students' use of different cognitive and metacognitive strategies, and student management of different resources. They contain rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time and study environment, effort regulation, peer learning, and help seeking.

Motivation expresses internal state and desire of students to the learning. Learning strategies expresses strategies to the learning in cognitive and metacognitive way and to resource management. Both of motivation and learning strategies come from inside of students. We refer to motivation and learning strategies as internal factors.

2.3 Learning behavior affected with course settings

In general, the course settings greatly affect behavior, motivation, strategies of student learning, and consequently their understanding. For example, students worried about the credits would mind their current ranking in all students. Under specific course settings, students with similar motivation and learning strategies are considered to take several characteristic learning behavior which can be regarded qualitatively identical. Since impacts of different course settings are strongly related with some of internal factors of students, the characteristics are considered to appear in specific kinds of learning behavior, not in all ones. Since course settings changing every week give different stimuli to students, learning behavior of students varies every week. However, under specific course settings in one week, students who have same motivation and learning strategies are expected to take one of a few kinds of characteristic learning behavior. Since course settings are uniform with all students every week, we can categorize students with internal factors, through observation of their learning behavior.

3 Distinction of learning status

3.1 Immediate distinction with internal factors

In higher educational institutes like universities, courses of similar settings are repeated for several years. The repetition produces several student groups, each of which consists of students similar in learning. Since teachers in a programming course know stereotyped learning of students in past courses, they can infer the learning status of current students from that of past students. The basic idea of the paper founds on this knowledge. Suppose a student in a specific week of the current course. From students in the same week of a past course, let us pick up students similar to the student in the current course. If the past students are suffering from learning in the week, we regard the current student also have high probability to suffer from learning in the week. The paper propose a method to extract passive students from current students based on their similarity to past students who need cares from teaching staff. The power of the teaching staff is assigned to the passive students, leaving active ones study by themselves, which achieve efficient teaching.

The method collects the learning behavior of students as their portfolios. To know the motivation and learning strategies which affect learning behavior of students, the method applies the contextual inquiry[1] to past students. The result of the contextual inquiry is analyzed from the viewpoints of motivation and learning strategies of students. Based on the strength of motivation and learning strategies, past students are classified into personas[2]. Founding on the personas, the method identifies superior factors to distinguish learning status of past students. Assuming the similarity between past and current students, the method derives the learning status of current students.

The method has been applied in the introductory C programming course in Ritsumeikan University. More than 500 hundred freshmen take the course as a compulsory subject.

3.2 Contextual inquiry

Programming teachers sometimes know what behavior comes from motivation and leaning strategy of students. However, the knowledge of the teachers may not explain all behavior caused by an internal factor. For example, some teachers may believe that the intrinsic motivation factor would emerge as long learning time and many submissions of advanced assignments. In fact, the intrinsic motivation results in not only long learning time and many submissions of advanced assignments, but also high score and many submissions of easy assignments. It is also difficult to estimate behavior factors of students from their internal factors in a quantity fashion only from observation of behavior.

To obtain a clear image of learning behaviors in specific conditions, a questionnaire would not be a good method, because its questions are predetermined with assumptions of questionnaire maker. Learners who give answers to the questionnaire may have no experiences for conditions assumed in questions. It prevents collection of truthful information. The contextual inquiry [1] lets us know what behaviors learners take in actual conditions. To know the behavior emerging from internal factors of the students, the proposed method conducts the contextual inquiry. Two students having finished the programming course interview with each other about their behavior and motivation of programming learning. To get students prepared for the interview, the method provides lectures to teach them how to conduct the contextual inquiry before the interviews. The lectures explain the purpose of interviews is to know how they behave when they gain or lose their vigor to study programming. The lecture emphasizes the procedure of the contextual inquiry obliges interviewers to delve into the details of the learning behaviors mentioned in the answers. In the contextual inquiry, a scenario is described to reflect all facts obtained in the interview.

Following is a part of an interview of Donald (interviewer) and Daisy (interviewee).

Donald What makes you inclined to programming?

Daisy I regard programming as a challenge. When I have solved a tough assignment, I obtain a strong sense of achievement.

Donald When you face a tough one, you might sometimes find no way to solve it. Have you ever run into such a situation?

Daisy Yes, many times. In such a case, I will repeat to check sample codes in the textbook. Sometime I try easier assignments in the same section.

Donald What will you do, if you cannot get anything from them?

Daisy I will search Web pages explaining similar matters using the Internet. I do not prefer to be supervised by TAs, because I feel lost in the challenge.

Donald But, they give you hints, even an encouragement sometimes.

Daisy I am not pleased, even if they encourage me. but, when they give high grades to my codes, I get satisfactions.

The interview reveals Daisy has intrinsic motivation. She is challenging because she tries to solve tough assignments. When she faces difficulties, she tries to overcome them by herself, checking sample codes, and easier assignments, and searching the Web. She does not prefer supervisions nor hints from TA. She has extrinsic motivation, because she is strongly satisfied with high score. She also has self-efficacy. She gets strong sense of achievement when she finishes tough assignments. The scenarios of students bring about internal factors along with external factors resulting from the internal ones.

The proposed method qualifies internal factors of a student from his secenario obtained in the contextual inquiry. If a student mentions a topic associated with a specific internal factor in his scenario, the method provides the student with a point for the internal factor. Suppose a student mentions he would ask his friend immediately when he encounter an assignment which is hard for his to solve. In this case, the value for the help seeking is increased one point. After the qualification, each student has a vector composed by points representing the strength of the 8 kinds of the internal factors.

3.3 Persona

The method addresses a group of students similar in motivation and learning strategies with a persona[2]. The personas are deeply related with learning status of students. In the method presented in the paper, the learning behavior of students is investigated with the contextual inquiry[1]. The method analyzes results of the contextual inquiry, from the view points of 8 kinds of internal factors, 5 kinds of motivation and 3 kinds of learning strategies to classify students into several personas.

Table 1: personas

Persona	Characteristics
Industrious persona	has a strong will to improve programming ability.
Easy-going persona	studies programming as far as it is a fun.
Demanding persona	wants a preferable learning environment prepared.
Obliged persona	engages in programming under a pressure all the time.
Unwilling persona	hates programming learning, because it is too hard.

Empirically, 5 kinds of personas shown in table.1 are found in the programing course of Ritsumeikan University. From student requirements revealed in the contextual inquiry, personas are divided into 2 categories in terms of the learning status: active personas and passive personas. Supervisors should leave the former study freely to encourage their spontaneous learning, while the supervisors should attend the latter to prevent them from loosing willingness for learning. In the Ritsumeikan University, industrious and easy-going personas are included in the active personas, while the passive personas consists of the remaining.

3.4 Behavior factor values to be analyzed

Web servers facilitate us to collect learning operation logs from students. The logs express interactions of learners with the servers through Web pages. They express the student learning behaviors. From the logs, we can figure out real values of external factors. Table 2 shows an example of external factors and their values in a week:

Table 2: Example of external factors

Symbol	Explanation	Sample value
lt	total length of periods student logs in to server in a week	325(min)
lc	times that student logs in to server	5 (times)
s_{adv}	the number of submitted assignments of advaced level	3 (assignments)
s_{all}	the number of all submitted assignments	8 (assignments)
es	the number of submitted assignments before last day	6 (assignments)
ls	the number of submitted assignments on last day	2 (assignments)
s_t	total score	76 (point)
cl	total number of any button clicked on exercise site	533 (times)
s_{adv}	score of assignments of advaced level	28 (point)
v_p	the nubmber of times student visit progress page	3 (times)

The learning time of a student in a week can be calculated by sum of his/her login sessions into the Web site in the week. The login times is the number of times the student logins. The advanced submissions represent the number of advanced assignments the student submits. There is a deadline for submission of assignments every week. The last day submission shows the number of assignments the student submits in the last day. The total clicks are the number of clicks of the student into items inside the pages of the Web site in the week. The score indicates the points the teachers of TAs provides to the students after evaluation. Values of all factors are normalized into a uniform range [0,1].

3.5 Association using vectors

Students having finished the programming course have real vectors, which are composed of external factors. Each external factor in the real vector of a student is retrieved from his portfolio during the course. Once the contextual inquiry is applied to the students, they have another kind of real vectors, which consists of internal factors. In addition to that, a persona each student belongs to is determined through the analysis of his scenario. Founding on the real vectors and personas, the method figures out 2 kinds of association.

Let us first focus on the association of personas with internal factors. The method calculates the mean m and the standard deviation σ for each element of internal factor values. Some elements have large values, while other have small values. To treat them in a fair way, element value v is converted into Z-score, z, which is calculated with $z = 50.0 + \frac{(v-m)}{\sigma} \times 10.0$. The method calculates the Z-score for all internal factor for every student. For all students belonging to each persona, the method figures out a vector consisting of the mean of Z-scores for every internal factor. Along with the mean, the standard deviation is also calculated for every internal factor. The vector composed of the mean of Z-scores represents the internal factors of the persona. Especially, small standard deviation indicates most students in the persona have similar value for the internal factor. The internal factor is characteristic to the persona.

Next, we consider the association of internal factors with external factors. Using the internal factor vectors of students, we can pick up students who have high values for a specific internal factor. For every internal factor, the method calculates the mean of external factor values of students picked up. The vector composed of the mean values corresponds to each internal factor. It represents the external factor values of students strong in the internal factor. In other words, the vector indicates learning behavior of students with strong values for the internal factor.

4 Figuring out actual associations

We collected learning behavior of students in the introductory C programming course in 2012. We conducted the contextual inquiry for 42 students under their consent. In the experiment, we focus on 8 kinds of internal factors; they are intrinsic, extrinsic, task value, self-efficacy, and affective component for motivation, while effort regulation, help seeking, and metacognitive component for learning strategies[11].

4.1 Association of personas with internal factors

Two teachers read scenarios obtained in the contextual inquiry. To remove personal bias of the teachers, they individually classified students into 5 kinds of personas explained in table.1. When their decision are not coincident, they made an agreement for the decision, exchanging their opinions. The number of students who are demanding, easy going, industrious, obliged, and unwilling is 5, 10, 15, 6, and 6, respectively.

Table 3: Association of personas with internal factors

persona	intrinsic	extrinsic	task value	self efficacy	affective	effort reg.	help seeking	meta recog.
demanding	42.33	53.32	41.10	**55.76**	**49.65**	**30.11**	54.54	**33.78**
	4.89	7.33	8.28	4.57	4.57	2.99	12.57	2.44
easy going	**52.71**	52.10	55.15	53.93	47.82	48.43	**49.04**	59.42
	4.77	6.58	9.08	4.92	5.63	4.89	4.77	5.60
industrious	59.83	55.76	54.54	56.98	**47.21**	54.54	47.62	54.95
	5.84	5.98	7.66	5.46	4.35	5.81	7.35	11.07
obliged	44.77	52.91	46.80	**36.63**	**55.96**	**60.03**	48.84	36.63
	6.11	12.55	11.51	2.88	4.20	3.05	13.95	6.75
unwilling	**39.68**	46.80	44.77	32.55	**54.95**	**36.63**	51.89	**33.57**
	2.28	5.76	6.11	4.99	2.88	4.55	6.52	2.28

For each student, Z-score was derived for each of the internal factors. For each persona, a mean vector was figured out with the average of the Z-scores, as it is illustrated in table.3. In the table, the upper figure in each cell indicates the mean value of Z-scores for all students belonging to the persona, while the lower one shows their standard deviation. We ordered whole cells in the increasing order of the standard deviation. A cell has a bold figure, if its standard deviation belongs to smallest one third in the whole cells. Each row of the mean values in the table is referred to as the *weight vector* of the persona for internal factors, because it indicates what weight the persona has for every internal factor. It shows the characteristics of the persona. Especially, the bold value is peculiar to the persona.

4.2 Association of internal factors with external factors

For all students, the method has calculated the mean m and the standard deviation σ of the values for every internal factor. For every internal factor, students whose values are larger than $m+\sigma$ are picked up, to gather top 16% students in term of the strength of the internal factor. Table.4 presents the mean of external factor value of those top 16% students. Let e be the mean of all student values for a specific external factor. The cells have bold figures if the values in the cells are either larger than $1.1 \times e$ or smaller than $0.9 \times e$. Those external factors are specific to the top students in terms of the strength of the internal factor.

Let us change the way to see the table. The column vector of table.4 represents how much a specific external factor is affected by each internal factor. The method refers to each column of the table as the *gene vector* of the external factor for the internal factors.

Table 4: Association of internal factors with external factors

lt	lc	s_{adv}	s_{ad}	s_{all}	es	ls	s_t	cl	s_{adv}	v_p
intrinsic	390.88	**14.33**	1.78	6.85	**4.49**	**2.36**	78.60	231.04	11.00	6.20
exrinsic	389.23	11.79	1.84	6.97	**3.91**	**3.06**	83.94	219.01	11.84	6.05
task value	439.24	**13.46**	1.94	7.11	**3.46**	3.65	**87.44**	**245.60**	**12.81**	6.27
self efficacy	332.52	12.09	1.83	6.92	**3.22**	3.70	79.69	**179.04**	11.30	5.89
affective	355.82	**10.58**	**1.59**	6.42	**2.09**	**4.52**	71.51	**204.09**	9.86	5.90
effort reg.	402.24	**13.60**	1.87	7.03	**2.08**	**4.96**	83.46	**242.87**	10.57	**7.21**
help seeking	357.17	10.87	1.90	7.00	**1.67**	**5.33**	**67.90**	**264.43**	**12.37**	**9.17**
meta recog.	357.38	10.76	1.94	7.18	**3.35**	3.83	84.67	**179.37**	**12.21**	**5.38**

4.3 Combination of two association

The gene vectors present the association of external factors with internal factors. They show characteristic internal factor values for each external factor. When learning behavior is observed from a specific student, we can know internal factors which is dominant in the student, comparing the learning behavior with the gene vectors for all external factors. On the other hand, the weight vectors present the association of personas with internal factors. Combining the weight vectors with the gene vectors, we would be able to know the personas of individual students from their learning behavior. We can distinguish passive personas from others.

5 Discussion

5.1 Preciseness in distinction

In addition to 2012, we also conducted the contextual inquiry for students in 2013. To examine their internal factors, we randomly picked up 15 students. Through the analysis of their scenarios, we obtain a real vector of internal factor values for each of the students. Their personas are determined from the scenario contents by judgement of teachers. The weight vector for each persona in 2012 represents the characteristics of the persona in the year. For each of the 15 students in 2013, his persona was estimated using the cosine similarity of his real vector to each of the weight vectors in 2012. We regarded he belongs to the persona which has the highest cosine similarity. Since personas of the 15 students are judged from the scenario contents, we examine the coincidence of the estimated results with the judged ones. For the category consisting of the 5 kinds of personas, the coincidence is 8 out of 15. However, the coincidence is 14 out of 15, when we adopt the category of an active persona and a passive persona. The result reveals that the proposed method has enough power to discriminate passive students, if we know real vectors consisting of internal factor values for individual students. Note that, to achieve it, we should conduct the contextual inquiry followed by scenario analysis.

5.2 Ability for immediate distinction

In the contextual inquiry, many students mentioned they change their learning status during they take the programming course. To effectively utilize small number of teaching staff, we should discriminate passive students from active students on the spot. Since teaching contents in a week constitute one meaningful group, we need to discriminate the persona of every student, once in a week. However, the contextual inquiry needs too much time and effort to conduct every week.

Let us denote a weight vector of persona p as $f_p = (w_{1p}, w_{2p}, \cdots, w_{np})$, where n is the number of internal factors. We can regard persona p gives an effect on i-th internal factor with weight w_{ip}. We can also denote a gene vector of external factor b as $g_b = (q_{1b}, q_{2b}, \cdots, q_{nb})$. It represents the external factor value of a specific student is q_{jb}, if a student has only j-th internal factor. Let us consider the dot product $f_p \cdot g_q$. We can regard the external factor of the persona p appears as a result of accumulated effects of $w_{ip} \times q_{ib}$, where $i = 1, 2, \cdots n$.

If we have a way to handle the dot product of the weight vectors and the gene vectors for all learning behavior and students in every week, we can figure out correct element values for gene vectors and weight vectors, which enables us to discriminate passive students immediately after the exercise every week. The Non-negative Matrix Factorization (NMF)[14] is a promising way to handle the dot product of multiple gene vectors and weight vectors. The research team is now engaging in the formulation of the discrimination method of students using NMF[4].

6 Conclusion

Some students can learn by themselves, while others need to be attended by teaching staff. The latter are passive students. It is essential to focus teaching staff power on passive students to teach programming to far more students than teachers. To achieve it, we should distinguish passive students from others. Using the contextual inquiry, the proposed method identifies the personas of individual students. At the same time, it examines internal factors and external factors of them. The method figures out the association of personas with internal factors as well as the association of internal factors with external factors. Combining these association, the method distinguishes passive students. The paper examines the ability of the method. It also discusses the next step for the distinction of passive students immediate after every week exercise.

References

[1] H. Beyer and K. Holtzblatt, "Contextual Design: Defining Customer-Centered Systems," Morgan Kaufmann, 1998.

[2] A.Cooper, "The Inmates are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity," Pearson, 2004.

[3] W.Dick, "The Systematic Design of Instruction," Plentice Hall, 2008.

[4] Dinh Thi Dong Phuong, and Hiromtisu Shimakawa, "Superior Factors to Predict Learning Status," Proc. of LTLE2015, pp.307-312.

[5] Renaud Gaujoux, and Cathal Seoighe, "A flexible R package for nonnegative matrix factorization," BMC Bioinformatics, Vol.11, No.1, p. 367. 2010.

[6] S. Graf, C. Ives, N. Rahman, and A. Ferri, "AAT: A Tool for Accessing and Analyzing Students Behavior Data in Learning Systems," Proc. of LAK, pp.174-179, 2011.

[7] John M Keller, "Motivational Design for Learning and Performance: The ARCS Model Approach," Springer, 2010.

[8] P.Kinnunen, and L.Malmi, "Why Students Drop Out CS1 Course?," Proc.of ICERE6, pp.997-108, 2006.

[9] P.Kleinginna, Jr., and A.Kleinginna, "A categorized list of motivation definitions with suggestions for a consensual definition," Motivation and Emotion, 5, pp.263-291, 1981

[10] MM. McGill, "Learning to program with personal robots: Influences on student motivation," ACM Trans. Computing Education, Vol.12, No.1, Article no.4, 2012.

[11] P.R.Pintrich, D.A.F.Smith, T.Garcia, and W.J.McKeachie, "A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)," Nat'l Center for Research to Improve Post secondary Teaching and Learning, 1991.

[12] Robert A. Reiser, "Trends and Issues in Instructional Design and Technology," Pearson, 2012.

[13] D.H.Schunk, and B.J.Zimmerman, Self-regulated learning: From teaching to self-reflective practice, Guilford Press, 1998.

[14] T.Segaran, "Programming Collective Intelligence," O'reilly, 2007

[15] R.E.Slavin, "Educational psychology: Theory and Practice," 10th, Edition, Pearson, 2006.