2A1-IOS-3b-2

Students Capability Growth Trajectory Mining

Thongchai Kaewkiriya[*1, *2] Ryosuke Saga[*2] Takeo Ichinotsubo[*2] Shunsuke Aoki[*2]
Woyuan Niu[*2] Hiroshi Tsuji[*2]

Thai-Nichi Institute of Technology[*1] Osaka Prefecture University[*2]

In order to find patterns of a student's growth, this paper proposes a new method on trajectory mining and illustrates its case study. First, this paper introduces the basic terms (such as capability space and capability structure) and the idea on trajectory mining. The capability space spans the possible student states and the capability structure expresses dependency on growth. Then it reviews the algorithm for data mining based on ISM (Interpretive structural modeling) and SPM (Sequential Pattern Mining). Designing questionnaire for collecting student's growth, this paper shows the example of finding by trajectory mining. The case study includes 8 capabilities (English skill, Japanese skill, presentation skill, programming skill and so on) and 60 subjects which includes both undergraduate students and graduate students. Using the collected responses, the proposed algorithm constructs 28 capability structures.

1. Introduction

Data analysis including data mining has contributed to a variety of field such as, education and business [Tsuji 2013]. One example is to analyze the growth of the organization. It can be assessed which direction the organization should grow. For education data mining has been applied in several cases such as learner's recommendation system [Romero 2008].

The consideration of the growth for student's capability by analysis of log file is promising because log file includes a lot of information on each user's profile. To get suitable knowledge from log file, the data structure is important To design data structure, let us review the standard concept of capability maturity model integration (CMMI). This model assessed the quality of the software development [Paulk 1995] . The level of capability grows up from initial to optimizing in CMMI. Each capability is considered by being divided into 5 levels [Chrissis 2003]; level 1 is the lowest and level 5 is the highest. This research allows us to analyze growth trajectory. For example, it used for personal e-learning [Nakamura 2010] for science teacher in sharing knowledge base on content [Sakoda 2009].In addition, the formal concept of trajectory mining on the capability space was implemented as an information system named SPICE [Ichinotsubo 2012]. And an application for software outsourcing clients showed the availability of the proposed concept.

In order to show how to find the pattern of a student's growth, this paper is organized as follows; Section 2 consists of 5 sections. The first reviews capability of a student which is the base for discussion. Second reviews basic idea of CMMI and SPICE. Third, introduces term definitions. Fourth, reviews ISM algorithm for finding capability structure extraction. Fifth, we also review SPM algorithm. Section 3 will propose visualization path of growth trajectory. Section 4 will propose capability structure extraction from TNI, which consists of 3 parts. First part describes overview of sample. Second part shows design of questionnaire. Third parts shows what kinds of patterns are found.

Contact: Thongchai Kaewkiriya, Thai-Nichi Institute of Technology, 1771/1 Pattanakan Rd, Suanluang Bangkok Thailand, +66835450622, thongchai@tni.ac.th

2. Background

2.1 Student capability

In order to assess student capability, we need some factors to help, such as student's personal profile and to analysis history log to assess. In this research, the meaning of student capability is the level of student's capability. For consideration from the trend of the growth for each student's log will analyze into a group or individual. This research divided into 8 capabilities which is Japanese skill, English skill, Java programming skill; Study planning, Presentation skill, E-mail, Web skill, Database skill and 60 subjects which includes both undergraduate students and graduate students.

2.2 Basic idea for analysis

In this research, we use basic idea of CMMI which is prepared for evaluating software organizations by using the capability and maturity. Both maturity and capability will be consistent. Also, the trend of growth will be reflected in quality. CMMI has been known that maturity of capability grows up from level 1 (initial) to level 5 (optimizing). Basic idea of CMMI is shown in Figure 1.

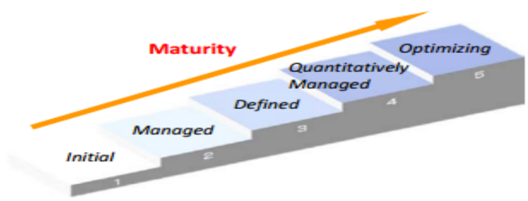


Figure 1. Capability Maturity Model Integration

In our previous research, CMMI was embedded SPICE (Spiral Enhancement Capability Support System) which is a learning support system. SPICE is able to store the growth log of students and collecting all the information of state from student. In order to store log for analysis the growth trajectory and capability structure, this paper uses SPICE.

SPICE will also help users to share various experiences while a supervisor will help to adjust student's capability more appropriate. The previous research has already studied about SPICE system, such as how to make SPICE more efficient [Nakamura 2012] for users. It creates growth path to be easy readable and not to be complicated. In addition, we also presented the creation of visualization path [Ichinotsubo 2012] to analyze easily. The structure of SPICE is shown in Figure 2. Many functions have been developed to enhance the capability of SPICE by introduction a growth trajectory [Ichinotsubo 2012].

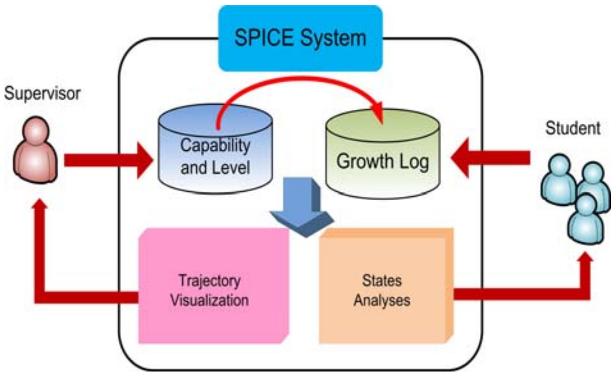


Figure 2. Structure of SPICE

2.3 Term definition

Let us define capability space and capability structure, by setting capability $C_i : C_1, C_2, C_3, \ldots C_n$. In Figure 3, Capability C_i is growth Y axis and capability C_j is growth X axis. Each has 5 levels which are used basic idea from CMMI: l_{i1}, l_{i2}, l_{i3}, l_{i4}, l_{i5}. Unit U_i which is a growing object with includes both organizations and individuals. For example, student in Japanese class has an objective to improve their own skills, such as speaking skill, listening skill, writing skill, etc. We will consider growth units. Each capability should be evaluated in one of 5 levels; poor, average, good, very good, excellent. In determination each level's detail will depend on the appropriate supervisor.

On the other hand, we assume that there are more than three capabilities for one subject domain. For example, there are four capabilities in the English language study domain (speaking, listening, reading, and writing) where a unit is a student or learner.

Let us define the capability space. The basic idea arises from the comparison of knowledge space theory [Paul 1985] [Masahiro 2005] for knowledge acquisition. Capability space is a set of potential capability states. And capability subspace is a set of state vectors that consist of some capability axes. We take 2 axes C_i and C_j for example in Figure 3 $X_i(t)$ is a capability state for unit U_i at $X_i\ (1) = (l_{11}, l_{21}, l_{31} \ldots\ l_{n1})$. We suppose units only grow at one capability state for one change from state (l_{i1}, l_{j1}) to (l_{i5}, l_{j5}).

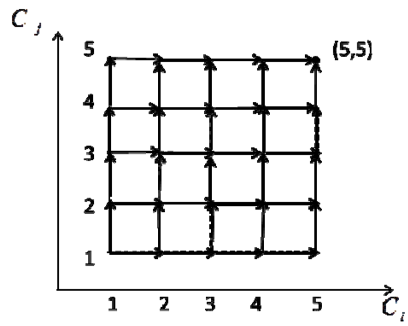
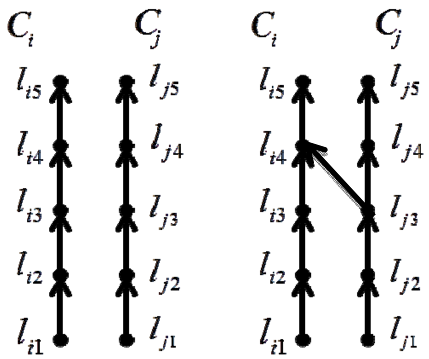


Figure 3. Two axes based Capability Space

Capability structure is structure of relation at least 2 capabilities. Figure 4 presents two capabilities in C_i and C_j. Sometimes levels l_{i5} in Figure 4 (a) can be achieved without any condition. However, achieving levels l_{i5} in Figure 4 (b) may sometimes require another condition. For example, l_{i5} should be achieved before a unit achieves l_{i4} in other words; l_{i3} and l_{j3} are prerequisites for achieving l_{i4}.



(a) (b)

Figure 4. Example of capability structure

2.4 Interpretive Structure Modeling

ISM is sophisticated model methodology presented by [Warfield 1974]. ISM algorithm supports to determine order and setting on the complication for relationships among elements. Then ISM could be a computer-aided method to developing graphical representations for system structure [Warfield 1982]. The concept is to use experts' experience to decompose an elaborated system into various elements and create a multilevel structural framework. It analyzes an application domain of element and resolves these in a graphical representation of their directed relationships and hierarchical levels. The ISM procedure is explained as follows;

- Step 1 is to identify elements which are interrelated to the problem parts. Elements should be identified by surveying or researching.
- Step 2 is to define the contextual relationship between the elements which needs to be examined.
- Step 3 is to define structural self-interaction matrix (SSIM) of elements which refers the relationship among elements of the system.

- Step 4 defines reachability matrix (RM).
- Step 5 defines partitioning of the reachability matrix.
- Step 6 gets the basic of the relationship as stated above in the matrix reachability. Then, drawing a graph and remove a transitive connection.
- Step 7 produces a converted result into a form of ISM model by changing the element nodes and statements.

2.5 Sequential Pattern Mining

Sequential Pattern mining [Philippe 2013] is one of a famous algorithm used in data mining and knowledge discovery. It helps finding the frequent sequences and the relationship between item sets within large database. SPM is an extraction method of frequent pattern in order to consider the frequency the happening between items. It is supported from IBM research institute [Agrawal 2005]

The concept of algorithm SPM in our research is the application of sequence's principle [Aoki 2013]. For example one sequence begins with the symbol "⟨" and end with another symbol "⟩". Each sequence consists of item ik (k=1,2,…,l) and has arranged as series. Moreover, items which happened at the same time will are represented as symbol "( )"

For example, item I={a, b, c, d, e} has now appeared in the sequence. If they are generated when item "a" begins, "b", "d", and "e" are concurrent then the representation is as follows; ⟨a (b d e) ⟩.

One feature of pattern mining is shown in Table I (a). From the example, under the condition that minimum support is 40% for 2 sequences (Sequence ID: 1 and 4), item "20" occurs behind item "10". It is said that ⟨10, 20⟩ satisfies the minimum support. This is a sequential pattern. Likewise, a frequent patterns are shown in Table I (b). The minimum support can be changed by an analyst. The number of the extracted sequences can be changed by changing minimum support.

Table I Example of sequence and minimum support

ID	Sequence
1	⟨10 (20, 40, 50)⟩
2	⟨10 30⟩
3	⟨(10, 30, 50)⟩
4	⟨10 (20, 50) 30⟩
5	⟨30⟩

(a)

Sequence	Minimum support
⟨10, 20⟩	40%
⟨10 (20,50)⟩	40%
⟨10, 50⟩	60%

(b)

3. Visualization of path growth trajectory

An example of visualization is shown in the type of vector. Each path can reach the destination or goal depending on individual's capability level. The current state can be plotted in capability space. We can visualize path by using data of student's growth log. Also, we can assign capability level in each capability. For example, capability of Japanese reading level could be in "poor", "average", "good", "great" and "perfect"). The example is shown in Figure 5. Figure 5 also illustrates the growth trajectory.

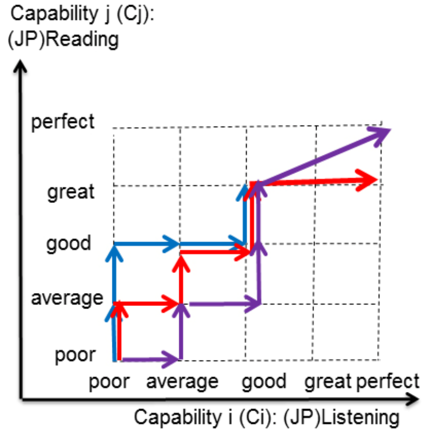


Figure 5. Example of visualization path and growth trajectory

4. Capability structure extraction from TNI

4.1 Overview of sample

For group sample, we corrected some data on eight capabilities from 60 students of Thai-Nichi Institute of Technology. The defined capabilities are shown in Table II.

Table II: The detail of capabilities

No.	Detail of capabilities
1	Japanese skill
2	English skill
3	Java programming skill
4	Study planning ability
5	Presentation skill (Power point)
6	E-mail skill
7	Web skill
8	Database skill

Capability structure extraction consists of three steps. First step is to record growth. This step will collect data by questionnaire from students. In detail, students will collect data of achievement year in each capability's level (See Figure 6).

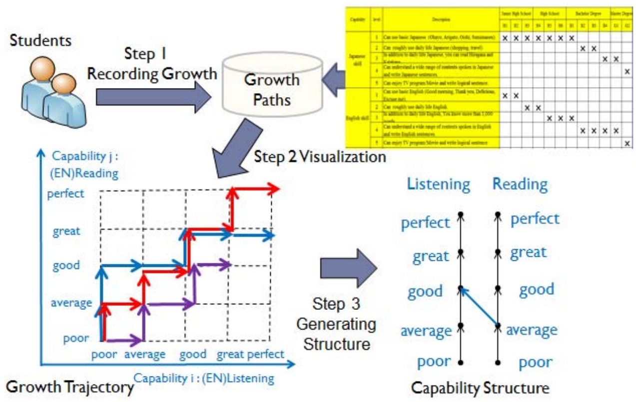


Figure 6. Step of capability structure extraction

Second step is to create visualization path which traverses capability state. Each path can reach the destination depending on individual's capability level. Third step is the process of generating structure. We call it capability structure extraction. For a part of algorithm in extraction, we use ISM Interpretive Structural Modeling and SPM Sequence Pattern Mining introduced before.

4.2 Subjects and Design of questionnaire

As case study for the proposed method, we use sample of students from Thai-Nichi Institute of Technology. The samples are as follows; undergraduate year 3-4 students from the information technology, undergraduate year 3 students from business Japanese, undergraduate year 2 from computer engineer and graduate level from IT. Total are 60 students.

We define eight capabilities (English skill, Japanese skill, presentation skills, Java skills, study planning, web skills, e-mail skills and database skills). Each capability is assigned capability levels. Due to the sample group's institute has been taught in Japanese language which is a mandatory course. Figure 7 shows an example of questionnaire and Figure 8 shows an example of capability levels.

Capability	level	Description	Junior High School			High School			Bachelor Degree				Master Degree	
			H1	H2	H3	H4	H5	H6	B1	B2	B3	B4	G1	G2
Japanese skill	1	Can use basic Japanese. (Ohayo, Arigato, Oishi, Sumimasen).	×	×	×	×	×	×	×					
	2	Can roughly use daily life Japanese.(shopping, travel)								×	×			
	3	In addition to daily life Japanese, you can read Hiragana and Katakana.										×	×	
	4	Can understand a wide range of contexts spoken in Japanese and write Japanese sentences.												×
	5	Can enjoy TV program/Movie and write logical sentence.												
English skill	1	Can use basic English (Good morning, Thank you, Delicious, Excuse me).	×	×										
	2	Can roughly use daily life English.			×	×								
	3	In addition to daily life English, You know more than 1,000 words.					×	×	×					
	4	Can understand a wide range of contexts spoken in English and write English sentences.								×	×	×	×	
	5	Can enjoy TV program/Movie and write logical sentence												×

Figure 7. Example of questionnaire

Japanese skill	
Level	Explanation
1	Can use basic Japanese. (Ohayo, Arigato, Oishi, Sumimasen).
2	Can roughly use daily life Japanese. (shopping, travel)
3	In addition to daily life Japanese, you can read Hiragana and Katakana.
4	Can understand a wide range of contexts spoken in Japanese and write Japanese sentences.
5	Can enjoy TV program/Movie and write logical sentence.

Figure 8. Example of explanation levels

4.3 Finding

The notice of the test, we divided in to 4 groups. For group one, we consider the difference between undergraduate students and graduate students. For group two, we consider the difference between male students and female students. For group three, we consider the difference between the relationships of each capability structure. For example, there is no dependency capability list and strong dependency capability pair. For the last group, we will place emphasis on the strong consistency's pair and inconsistency's pair.

4.4.1 Difference between undergraduate group and graduate group.

Figure 9 and 10 shows the sample of capability space and capability structure which has a relationship between English capability and Japanese capability. From the example we consider the different between undergraduate and graduate student group.

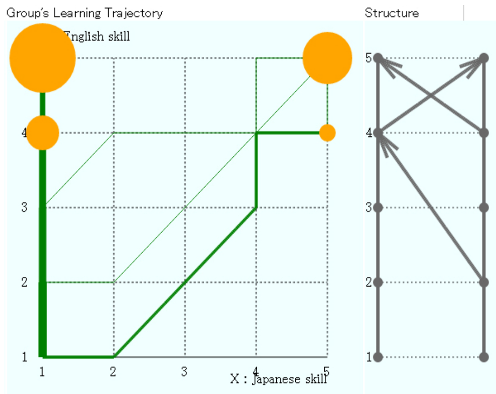


Figure 9. Graduate students group

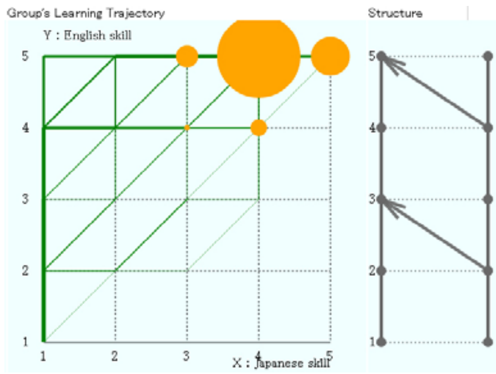


Figure 10. Undergraduate students group

Graduate students: Stronger English skills than Japanese skill, because some graduate students have no Japanese skills. Generally, students that graduate from TNI continue studying graduate program at TNI. Students must have Japanese skill.

Undergraduate students: Stronger English skills than Japanese skills (little), because every students have Japanese skills. Average number of undergraduate students has stronger English and Japanese skills than graduate students. The difference between English skills and Japanese skills for graduate students has more than undergraduate students.

For capability structure of graduate group, Japanese level 2 is necessary for English level 4 and English level 4 is necessary for Japanese level 5. Also, Japanese level 4 is necessary for English level 5, as well. Capability structure of undergraduate group, Japanese level 2 is necessary for English level 3. Moreover, Japanese level 4 is necessary for English level 5. Therefore, we conclude that Japanese level 4 is necessary for English level 5. Both groups are the same, which mean that students will be able to achieve English level 5 must achieve Japanese level 4 first.

4.4.2 Difference between male students and female student.

Figure 11 and 12 shows the sample of capability space and capability structure which will show the difference between English capability and Japanese capability. From the example, we consider the difference between male and female students group.

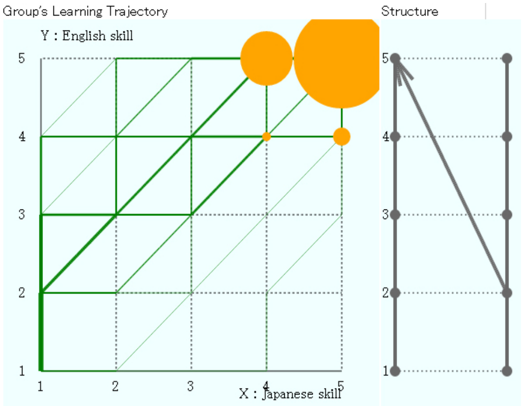


Figure 11. Female student group

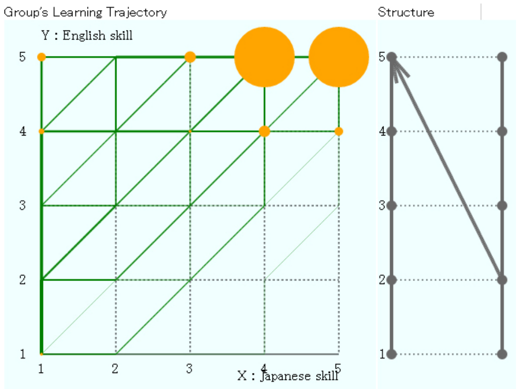


Figure 12. Male student group

Average number of male student in English skills level 5 and Japanese skills level 4 is similar to English skills level 5 and Japanese skills level 5. Then Japanese skills level 2 is necessary for English skills level 5 in both group. The average number of female students in English skills and Japanese skills are stronger than average number of male students in English skills and Japanese skills. For capability structure of female and male group Japanese skill level 2 is necessary for English skill level 5. So, we conclude that Japanese skill level 2 is necessary for English skill level 5. Both groups are the same. It means that students who can achieve English skill level 5 must achieve Japanese skill level 2 first. Consideration of the advisor to teach student maybe emphasis on students who achieve Japanese skill level 2 first in order to have a change for achieving English skill level 5 onward.

4.4.3 No dependency capability list.

This section will be emphasis with no dependency capability list. No dependency capability means any relationship between 2 capabilities C_i and C_j which capability structure pair as shown in Figure 13.

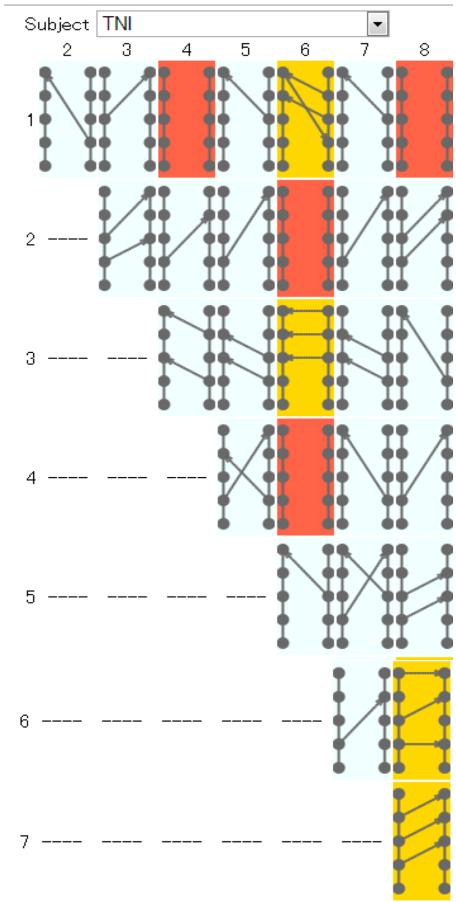


Figure 13. Capability structure of ISM

Figure 13, there are 4 pairs for no dependency consist of 1-4, 1-8, 2-6, 4-6. Pair of 1-4 which means Japanese skill is not related with study planning ability. Pair of 1-8 means Japanese skill is not related with database skill. Pair of 2-6 means English skill is not related with e-mail skill. Pair of 4-6 means study planning is not relate with e-mail skill. Therefore, we consider getting the result for processing. Advisor may consider to emphasis capability structure as relationship which will give a chance to student to get more advantages from studying.

4.4.4 Strong dependency capability pairs

Figure 13; consider the sample groups which can be defined as following; if the number of relationship is equal 1 path, it is defined as fair. If the number of relationship is equal 2 paths, it is defined as good. If the number of relationship is more or equal 3 paths, it is defined as strong.

A pair of capability is in the level of fair total 12 pairs.

A pair of capability is in the level of good total 7pairs.

A pair of capability is in the level of strong total 4 pairs.

Consider level of strong, first pair between levels of Japanese skill (C_i) with e-mail skill (C_j) l_{j2} is necessary l_{i5}, l_{j3} is necessary l_{i4}, l_{j4} is necessary l_{i5}. Consider level of the strongest, second pair between study plans (C_i) with e-mail skill (C_j) $l_{j3,4,5}$ is necessary $l_{i3,4,5}$. Consider level of strongest, third pair between e-mail skill (C_i) with database skill (C_j) li2 is necessary l_{j2}, l_{j3} is necessary l_{j4}, l_{i5} is necessary l_{j5}. Consider level of strongest, the last pair between web skill (C_i) with database skill (C_j) l_{j2} is necessary l_{i5}, l_{j3} is necessary l_{i4}, l_{j4} is necessary l_{i5}. Consider the use of strongest dependency, normally is the responsibility of advisor. Advisor may choose strongest pair in the same group with the relationship to use as e-mail skill with database skill and web skill with database skill.

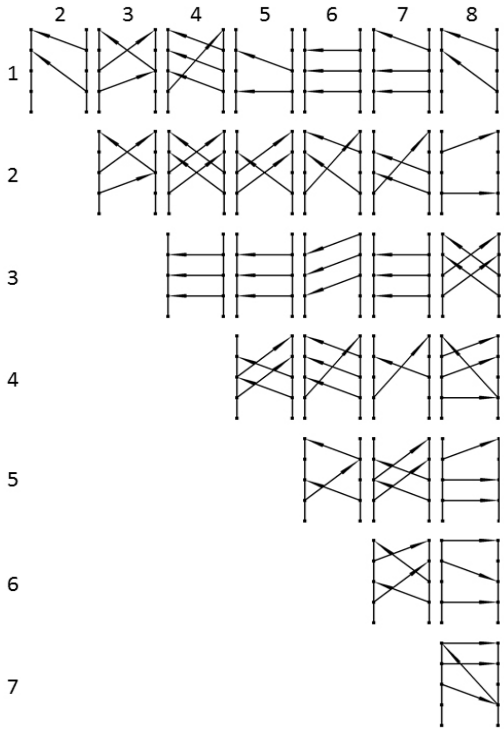


Figure 14. Capability structure of SPM.

Consideration algorithm SPM, we use minimum support as 90%. The strongest dependency consist of 9 pairs 1-4, 2-4, 3-8, 4-5, 4-6, 4-8, 5-7, 6-7, 7-8. The number of relationship is equal 4 paths. Then strong dependency consist of 14 pairs 1-3, 1-6, 1-7, 2-3, 2-5, 2-6, 2-7, 3-4, 3-5, 3-6, 3-7, 5-6, 5-8 and 6-8. The number of relationship is equal 3 paths.

So, the total of strong dependency will be 23 pairs. We compare the result of ISM algorithm and SPM algorithm (minimum support 90%). The result of ISM for strong dependency is 4 pairs. The result of ISM for no dependency is 4 pairs. The result of SPM for strong dependency is 23 pairs. The result of SPM for no dependency is none. Therefore, the result of SPM algorithm is better than ISM algorithm.

4.4.5 Consideration of consistency pair and inconsistency pair

In this section we find consistency and inconsistency. The consistency can be defined like this;

If the number of path is larger than 15 it is define as strong.

If the number of path is less than 15 it is define as inconsistent. We consider the capabilities consistency pair from a group of 60 samples. Each capability path will have the thickness of each path, which means the value of student's sample. For example, the thickness of path is very thick means a lot of student's value in that path. The thickness of path is less thick means lower student's value.

Figure 15 capabilities consistency pair consists of 5 pairs. First is presentation skill and study planning. Second is Web skill and English skill. Third is Web skill and presentation skill. Forth is Database skill and e-mail skill. Last is Database skill and Web skill. Other pairs will have a trend in inconsistent pairs which spread more into each levels and value of student in each path will is determined not to meet the criteria for more than15 paths.

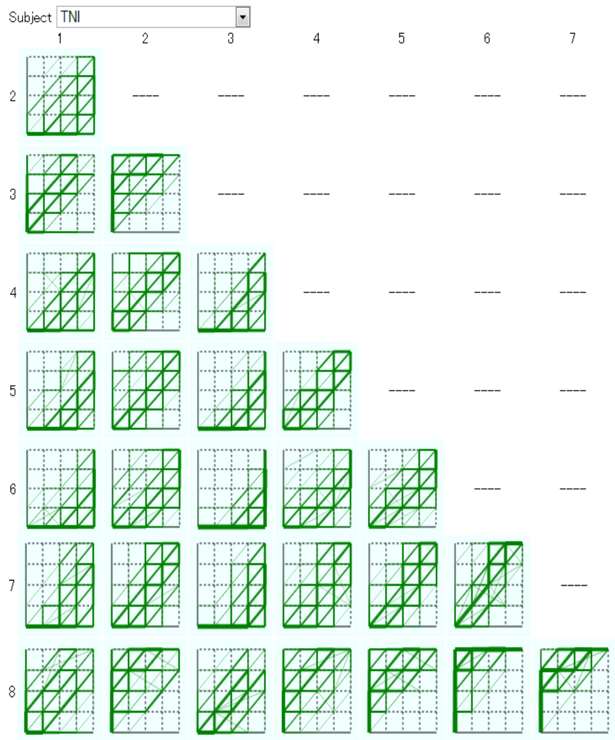


Figure 15. All group paths of capability.

5. Discussion

We have presented the method of trajectory mining and its case study. The objective of this research is to find the patterns of student's growth. We used the sample of 60 subjects. The findings for the case study are as follows(1) it showed the difference between undergraduate group and graduate group. We concluded that Japanese level 4 is necessary for English level 5. Both groups are the same, which means that students can achieve English level 5 have to achieve Japanese level 4 first. (2)it showed the difference between male students and female student. We concluded that Japanese skill level 2 is necessary for English Skill level 5. Both groups are the same, which means that students can English level 5 have to achieve Japanese level 2 first. (3) it also considers about strong dependency. From the sample group, we have the value of strong dependency as 4 pairs. Each pair has number of relationship more than 3 paths. Therefore, there are 5 pairs of strong consistency which supervisor can consider for teaching or advising students.

6. Conclusion

This research aims to find the patterns of student growth based on trajectory mining. First we introduced the previous work. In addition, we reviewed about CMMI, as well. Second, we reviewed the background of the research (basic idea, term definition, ISM and SPM). Then, we introduced a visualization path of growth trajectory. Further, we proposed capability structure extraction from TNI. Finally using the volumes of growth logs, the proposed algorithm had constructed 28 capability structures.

This research proposed a method for finding pattern base on trajectory mining only. It doesn’t design the form of finding message for the analysis of automatic trial result. Our next task, will be present the finding message and create a planning generator in order to advise the students accordingly.

7. Acknowledgement

The authors would like to sincere thanks to students of TNI who responded to our survey. This works was partially supported by Japanese KAKENHI-C (2350049). Also, the authors would like to sincerely acknowledge Mr. Mauricio LETELIER for checking the draft.

References

[Tsuji 2013] H. Tsuji and R. Saga, Trajectory Mining on Capability Space Its Concept and Potential Application, Proc. of 46th Hawaii International Conference on System Sciences HICSS-46, 2013.

[Romero 2008] C. Romero, S. Ventura and E. Garcia, Data mining in course management systems: Moodle case study and tutorial Science Direct, Vol. 51, No. 1, 2008.

[Paulk 1995] M. Paulk, C. Weber, B. Curtis and M. Chrissis, Software Engineering Institute: The Capability Maturity Model: Guidelines for Improving the Software Process, Addison-Wesley, 1995.

[Chrissis 2003] M. Chrissis and M. Shrum, CMMI: Guidelines for Process Integration and Product Improvement, Addison-Wesley, 2003.

[Nakamura 2010] Y. Nakamura and H.Tsuji, Spiral Capability Enhancement Support System, Proc. of International Symposium on Aware Computing, 2010.

[Sakoda 2009] M. Sakoda, Y. Wada, H. Tsuji and K. Seta, Social Network Service with Maturity Level for Science Teachers, SMC/IEEE, San Antonio, 2009.

[Ichinotsubo 2012] T. Ichinotsubo, Develop a Prototype System for supporting Growth Using Visualization of the Growth History, Graduation Thesis of School of Engineering at Osaka Prefecture University, 2012.

[Nakamura 2011] Y. Nakamura, H. Tsuji, K. Seta and K. Hashimoto, Visualization of Learner's State and Learning Paths with Knowledge Structures: A. Konig et al. (Eds.), 2011.

[Warfield 1974] J. Warfield, Developing interconnected matrices in structural modeling, IEEE Transcript on Systems, Men and Cybernetics, Vol. 4, No. 1, 1974.

[Rick 2007] G. Rick and L. Ningwei, Using Interpretive Structural Modeling to Identify and Quantify Interactive Risks, Call Paper program, ASTIN Co, 2007.

[Nakamura 2012] Y. Nakamura, W. Niu, T. Ichinotsubo, R. Saga and H. Tsuji, Capability Structure Extraction from Growth Trajectory, The paper of Technical meeting on Information System, IS-12-048, 2012.

[Paul 1985] D. Paul and F. Jean, Spaces for the assessment of knowledge, International Journal of Man-Machine Studies, 1985.

[Masahiro 2005] I. Masahiro, Analysis of Information Table by Rough Set Theory, System, control and information: The Institute of Systems, Control and Information Engineers, Vol. 49, No. 5, 2005.

[Warfield 1982] J. Warfield, Interpretive Structural modeling, In Olsen, S.A. (ed), Group planning and problem solving method in engineering management, John Wiley and Sons, Inc., 1982.

[Philippe 2013] F. Philippe, SPMF, A sequential Pattern Mining Framework, http://www.philippe-fournier-viger.com/spmf/ (Accessed on 10-02-2013)

[Agrawal 2005] R. Agrawal and R. Srikant, Mining Sequential Patterns, International conference on data engineering, (ICDE ’95), 2005.

[Aoki 2013] S. Aoki, Dependence extraction from growth trajectory using sequential pattern, Bachelor Thesis of School of Engineering at Osaka Prefecture University, 2013.