A Trichotomic Approach to Concept Capture and Representation : With its Application to Library Data Mining

Minami, Toshiro
Kyushu Institute of Information Sciences | Kyushu University Library

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

Baba, Kensuke
Research and Development Division, Kyushu University Library

Amano, Eriko
Office for eResource Services, Kyushu University Library

出版情報：Applied Computing and Information Technology, pp.61-75, 2014. Springer International Publishing
バージョン：

A Trichotomic Approach to Approximate Representation of Concepts – With its Application to Library Data Mining –

Toshiro Minami*, Sachio Hirokawa†, Kensuke Baba‡ and Eriko Amano§

*Kyushu Institute of Information Sciences & Kyushu University Library, Japan
Email: minami@kiis.ac.jp
†Research Institute for Information Technology, Kyushu University, Japan
Email: hirokawa@cc.kyushu-u.ac.jp
‡Research and Development Division, Kyushu University Library, Japan
Email: baba.kensuke.060@m.kyushu-u.ac.jp
§Office for eResource Services, Kyushu University Library, Japan
Email: amano.eriko.760@m.kyushu-u.ac.jp

***Abstract*—This paper consists of two parts. Firstly, we propose a method of specifying the concept that is too hard to describe in an exact way, by setting up the relative distances from three key concepts; which we call a trichotomic approach to cencept description in an approximate means. Then we demonstrate its usefulness by applying to library data, a collection of loan records, precisely. We can choose not only the key words but also the patrons or books as the "key concepts" and set the location of the words, patrons, and books according to the "relative distances" from the three key concepts. In such a way, we can compare different types of objects such as patron and books in a uniform framework of trichotomic representation. This framweork of distance can provide with another types of methods for data analysis and finding close objects.**

I. Introduction

In our previous research [1], we proposed a system called ZK (ZeichenblocK, or ZakKi-chou in Japanese; a note-pad) system, which intends to provide with a means to represent the concept which is very hard to describe exactly because it has no appropriate word and expression in the language we are using. Our idea in the paper was to specify the concept not by using words and phrases that describe it but by the location that indicates the relative distances from a couple of key words/expressions, as an approximate description of the concept. One aim of this paper is to investigate further of this approach for specifying concepts in an approximate way.

Onomatopoeia is another good tool to be used to approximately specify the concept which has not appropriate word(s) and expressions. Onomatopoeia is basically a word that mimics a sound made by animals, e.g. cuckoo, neow, honk, or made in nature, e.g. splash, rumble, zoom, and other types of sound and noises. It is also used as a mimesis, which imitates the phenomenon, action, attitude, and something that actually does not directly relate to any kind of sound, but in order to express such things in the similar way to sounds. Japanese language has a lot of such mimesis words and they are very popularly used not only by children but also by adults in their daily conversatiions. For example, the word "mera-mera" describes state of flame that shakes. It is interesting that the word "sheen" describes the state that no sounds heard just like a kind of sound.

By using onomatopoeic/mimesis words as adjective, it is possible to modify the original meaning of the nouns they modify. According to a TV program [2], onomatopoeia has a great power and it is even used in the development of commercial product. For example, a maker tried to develop a part used in a car made of plastics, which looks like to be made of metal. They asked for comments to some testers about the difference of two types of parts; one is made of metal and another made of plastics. The testers described the differnce of them using expressions with onomatopoeia. As the result, the maker was able to find a solution how to produce the plastic part looking exactly like metal.

Both ZK and using of onomatopoeia are common in the sense that they are for describing the concepts based on the key words, they are different in their types that ZK's approach is to describe the concept according to the relative closeness from the keywords, whereas a onomatopoeia word describes the concept by modifing the keyword and show the difference from the keyword. Thus we can say they are somewhat complementary rather than contradictory, so that we can describe a concept in our mind more exactly by mixing them up.

Marketing becomes more and more important for libraries because of the advancement of information age and their patrons' needs has been changing greatly. One of the most important topics for them is obviously to recognize what their patrons are.

In order to capture about the patrons' behaviour, interest areas and others, we have been conducting a series of studies of investigating the patrons' behaviour by analysing the data obtainable by libraries, mainly using the loan records intending to develop tools useful for library marketing.

For example, we defined the profile of a patron by the 10 numbers, vector with 10 dimensions, of borrowed books in terms of the top categories in NDC, from 000 to 900. Then we investigated the width of interest area using the information entropy of the profile vector and the similarity of two patrons by using the cosine similarity of their profile vectors. With such measures we can see not only the characteristic features of patrons but also can see those of faculties by extending the definitions of these measures.

We analyze other data in a similar method of these examples. Firstly we decide what types of behavioural feature to investigate. Then we make up a measuring index or two that would be appropriate to analyze the data in terms of the feature. We calcuate the measure and have a look of the result with a visualization method including statistical ones such as histogram, scatter diagram, and others. We often calcuate the correlation coefficients in order to check the statistical relationship between two indexes. Not only the results of analyses but developing the analysis methods is also very important in our research methodology.

In this paper we investigate the potentiality of using trichotomic approach as a new type of analysis methodology of library data. Dicotomy classifies the targets into two constrasting categories. For example, up vs. down, black vs. white, left vs. right, good vs. evil, and goes on forever. On the other hand, trichotomy has been widely used in western cultures and eastern cultures as well.

For example, the trichotomic phrase “Chi-Jo-Ee (intellect, emotion, volitiion)” describes the three components of our brain’s functions and it is considered that we need to have a good balance of these three in oreder to keep our mind healthy.

The grand chanpions of sumo wrestlers are supposed to fulfill “Shin-Gi-Tai (heart, technique, body),” i.e. mentally tough and stable, have high level technique and phycially strong and healthy. In other words they are expected to be ideal model as a human.

In this paper, we take the idea of trichotomy so that we set up three key concepts/objects for locationing the target according to the relative distances from these three in the triangular form. Also we would like to demonstrate the potentiality and importance of trichotomic approach to data analysis through case studies of applying to library data analysis.

In the case studies we use a collection of loan records and we choose not only the key words but also the patrons or books as the “key concepts” and set the location of the words, patrons, and books according to the “relative distances” from the three key concepts.

In such a way, we can compare different types of objects such as patron and books in a uniform framework of trichotomic representation. This framweork of distance can provide with another types of methods for data analysis and finding close objects.

The rest of this paper is organized as follows: In Section II, we describe the basic idea for specifying concept in an approximate way by giving relative distances based-on the paper [1]. Also we discuss the characteristic features of this approach.

In Section III, we introduce some results of studies for library data analysis we have conducted so far. Also we disucss our approach to data analysis with emphasizing the importance of visualization.

Then in Section IV, we deal with trichotomic approach to data analysis. Our approach in this paper on this issue is to use of triangular representation of objects as well as concepts based on the relative distances from three key concepts/objects. We also show a couple of case studies and demonstrate the importance of trigonomic approach in data analysis.

Finally in Section V, we summarize the discussions we had in this paper and anticipate the research topics in this direction.

II. ZK and Characterization of Concepts According to Relative Distances

III. Library Data Analysis

The underlying data for defining interest range are the circulation records obtained from the Central Library of Kyushu University, Japan, from April 2007 to March 2008, which were used also in [2-4]. The data contain 67,304 records. A record consists of the book ID, book’s NDC (Nippon Decimal Classification) number, call number, borrower’s patron ID (renumbered in considering privacy), affiliation, type, and the timestamps for borrowed and returned dates and times, etc. The number of patrons who borrowed at least one book is 6,118 and the average number of books per patron is about 11. The borrowers affiliate either one of the 12 faculties and other organizations including library. The faculty names used in this paper stand for, AG for Agriculture, DD for Dental, DS for Design, EC for Economy, ED for Education, LA for Law, LT for Letter, MD for Medicine, NC for the special faculty called 21st century program, which is for the students who wish to study a wide variety of fields, O for those in other organizations, PS for Pharmaceutical, SC for Sciences, and TE for Engineering.

First of all we define the concept of the interest area profile of a patron as in the same way as in the papers [3-4]. For the areas of topics, we use the NDC number which is assigned to the books as a part of their bibliographical information. NDC is a decimal classification system localized to Japan of the well-known DDC (Dewey Decimal Classification). The top level categories of NDC consist of the following 10 topic fields; 000 for General Works, 100 for Philosophy and Religion, 200 for History and Geography, 300 for Social Sciences, 400 for Natural Sciences, 500 for Technology (Engineering), 600 for Industry and Commerce, 700 for Arts, 800 for Language, and 900 for Literature.

We define the profile of a patron as a vector with dimension 10, with each element corresponds to one of the 10 top categories of NDC. An element of the vector is the frequency of the borrowed books of the patron which have the corresponding top category numbers of NDC. Thus, for example, if a patron borrows 11 books with NDC number from 100 to 199.99, 12 books from 200 to 299.99, and so on until 19 books from 900 to 999.99, the profile vector of the patron becomes $<$11, 12, 13, 14, 15, 16, 17, 18, 19$>$. We can extend this definition to a group of patrons by just modifying the condition from “borrowed by the patron” to “borrowed by one of the patrons of the group.”

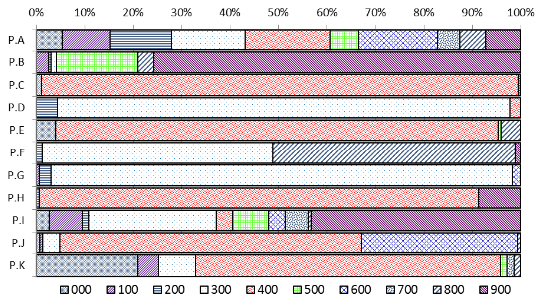

Fig. 1. PatronProf11

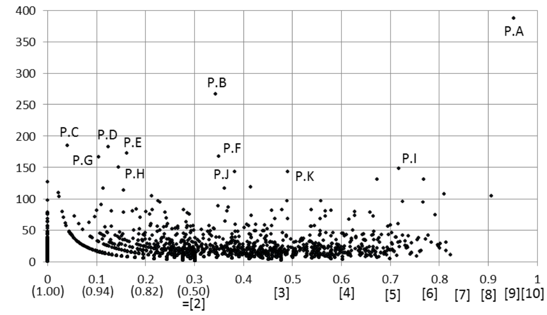

Fig. 2. PatEntFrq

TABLE I. PatPrfTbl

Patron	Range	Strength	Affiliation	Type
P.A	0.95	388	O	O
P.B	0.34	268	LT	D
P.C	0.04	185	SC	B4
P.D	0.12	183	LA	D
P.E	0.16	173	SC	B3
P.F	0.35	168	LA	D
P.G	0.10	167	LA	D
P.H	0.15	150	SC	B4
P.I	0.72	148	O	M
P.J	0.38	143	AG	B3
P.K	0.49	143	SC	M

The interest strength is defined by the number of books, or items, that are borrowed by the patron or the group of the patrons. We define the interest area range, or range size, by the information entropy of the profile by using the ratios of the 10 NDC categories. Let pi = number of the books that belong to the NDC category i divided by the total number of the books, or the strength, of the patron's profile. Then the information entropy of the profile is calculated as the sum of -pi log pi. We use 10 for the base of the logarithmic function in order to make the maximum value to 1 because the number of NDC categories is 10.

Figure 1 shows the interest area profiles of the top 11 patrons according to the number of items, or books in the left graph, and the interest area profiles of the faculties in the right graph. We chose them because firstly they are representative patrons among all the patrons and knowing them is important for library marketing, and secondly because quite a lot of patrons borrow only a couple of books and thus they are not appropriate to use as sample data for developing new methods for profiling the patrons. The top-most patrons from A to K (also called by P.A to P.K) borrows 388, 268, 185, 183, 173, 168, 167, 150, 148, 143, 143 books during one year, respectively.

It is easy to see in the figure that the ratios of books according to the classification number, or topic area, vary from patron to patron. For example, P.A borrows quite a wide area of books with NDC numbers from 000 to 900. On the other hand, P.C borrows mostly with the classification number 400 (Natural Science). Such difference of the width of topic areas indicates a character of the patron in his or her interest range, or curiosity range. Together with the number of the borrowed books, this range can be good measures for characterizing features of a patron, which will be discussed in more detail in the next section.

Table I shows the range, strength, affiliation, and type of the 11 patrons from P.A to P.K. As has been predicted the range of P.A (0.952) is quite high; the highest among 11 patrons. On the other hand P.C has the minimum range value (0.04), who's affiliation is SC and the year 4 undergraduate student (B4).

To have a closer look at the table, there are 4 students with affiliation of SC (Sciences) and 2 of them are B4 (P.C and P.H) and 1 (P.E) is B3 and another one (P.K) is M (Masters). The 3 undergraduate students have very low range values from 0.04 to 0.16. They are very concentrated in learning just like P.C. It is interesting to see that the remaining masters student (P,K) has relatively bigger range value 0.49. He or she borrows the books not only in the natural science field (with NDC 400), but also the books in general topics (with NDC 000), social sciences (with NDC 300) and others as well.

There are 3 Ph.D students with affiliation LA (Law); P.D, P.F, and P.G. The patrons P.D and P.G have similar range values 0.12 and 0.10, whereas P.F has bigger value 0.35. The former 2 students borrow the books with NDC 300 (Social Sciences) mostly, whereas the latter student borrows not only the books of social sciences but also the books with NDC 800 (Language) as many as of 300.

Figure 2 shows the correlation between the range size (x-axis) and the strength (y-axis) for all patrons. The range value 0 means that the patron borrows only one book. The range value is 1 if the patron borrows the books with all the NDC numbers, i.e. from 000 to 900, exactly the same number from each category. The location of the numbers parenthesized with [n] indicate that it is the range value, i.e. entropy, for the case that n categories have equal numbers of books, or $log_{10}n$, which is the maximum value for having values in n categories. The location of the numbers parenthesized with (n) indicate that it is the range value for 2 categories in which one category has the possibility of n and the other has the possibility of $1-n$. In this case, the maximum range value is $log_{10}2 = 0.30$ when n=0.5, i.e. half and half.

The patrons from P.A to P.K are named according to the order of the strength, or the number of borrowed books, so they are located in the upper part of the graph. Patron A (P.A) is located to the right-most and top-most place, which means he or she borrows the books from all the NDC categories with borrowing almost the same number of books each. Furthermore P.A borrows nearly 400 books, which is over 100 books more than the second one, i.e. P.B, who borrows more than 250 books.

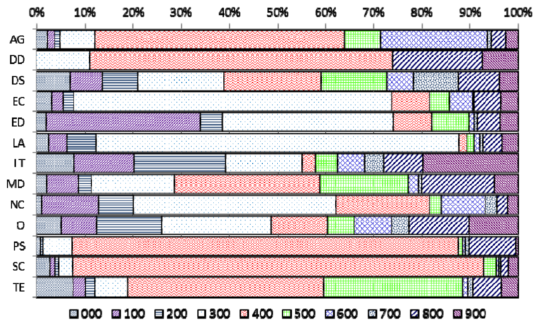

Fig. 3. FacProf

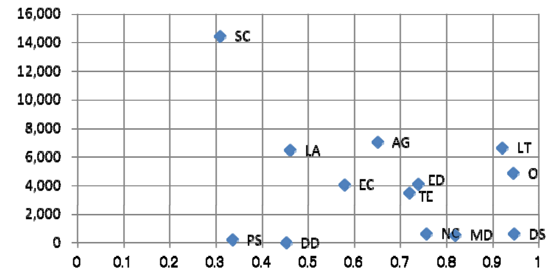

Fig. 4. FacRgSt

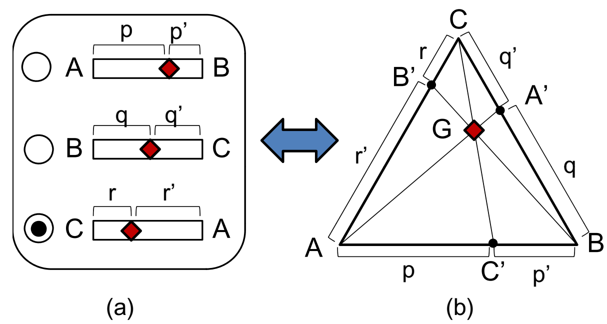

Fig. 5. (ratios-triangle) Two types of Representation for Trichotomic Distance Ratios in: (a) Pairwise Interface and (b) Triangular Interface

The patrons C, D, E, G, and H are located in the left-most part of the graph having the value less than 0.2, which means they borrow books with one category more than 80% of times and other ones less than 20%. Thus they have very limited range of interest.

The patrons B, F, J, and K are located in the range with the range value from 0.3 to 0.5, which means, roughly speaking, they mainly borrow books with 2 or 3 categories.

Figure 3 shows the profiles of faculties. We can see that the faculties PS and SC have a very high top interest area at the NDC 400 (Natural Sciences) and they are similar in this respect. On the other hand the faculties DS, LT, and O have relatively low value in the top interest area and they have a wide range of interest areas. These results are somewhat matching to our intuitive images on these faculties. In this respect, it is interesting to see that DD and MD have relatively wide interest areas.

By observing the right figure in Figure 2, we can see that SC is far away from other faculties in both axes. It has the lowest value in region size and the highest in strength, which mean that patrons in SC borrow the books in natural sciences (NDC 400) mostly and the number of the borrowed books are quite high, which probably means that their places physically locate very close to the library and it is quite easy for them to visit the library and borrow many books. The faculties of PS (Pharmaceutical), DD (Dental), and LA (Law) are located in the left part from the line with the range size 0.5, which means that their patrons also borrows books of their expertise area mainly than other faculties. The reason why the numbers of borrowed books of PS and DD might be that their faculties locate in a different campus from where the library is, and the patrons in PS and DD visit the library in order to get the books they could not find in the libraries of their own campus. LA is, on the other hand, located in the same campus as the library and also the number of the members is larger than that of PS and DD. It is interesting to see that DS (Design) and MD (Medical) are located in the lower right part of the graph where their range size is relatively large. Even though MD locates in the same campus as PS and DD, its range size is far bigger than these two. In order to find the reason of this fact, we investigate more on the patrons' behavior. Anyway in some reason the members of MD visit the library in a different campus in order not to find the books relating to their study in their expertise field but to find books in a wide variety of books. DS locates in a campus of it own, i.e. different campus from that of library and even farther than the campus for MD, PS, and DD. The strength, i.e. the number of borrowed books, is small probably because of this reason. DS is a faculty that relates both to engineering and design, and thus it is easy to guess that their interest range as a whole is wide. However it is still surprising that its range size is larger than any other faculties including O (Other, or unclassified) and that LT (Letter) also has high range size. The members of LT borrow not just the books of literature (NDC 900), but also those in other areas as many as of literature.

IV. Trichotomic Approach to Concept Representation

A. Trichotomic approach to concept representation

B. Triangular representation

In this part of the section, we study representation methods together with interfaces for specifying and showing the positioning of concepts/objects with relative distances from three keys (i.e. keywords, key objects, or three whatevers). Figure 5 shows a sample image of the interface in two types. Let A, B, and C be the keys.

The left part, i.e. (a), shows an interface design for specifying pairwise ratios. As we have three keys, we have three pairs; namely A and B, B and C, and C and A. Even with three pairs, we have only two degrees of freedom. Thus we set up three radio buttons at the left. The selected button indicates that this pair is dependent to other two. In Figure 5, the button for the pair C and A is marked as independent, thus we can specify the ratios for the pairs A and B, and B and C.

For each pair, a slide bar is given and we can specify how the concept/object in mind is located in comparison with two keys. In Figure 5(a), it is given by the location of rather closer to B than A for the pair A and B, with the ratio of $p : p'$ where $p' = 1 - p$, and near the middle for the pair B and C, with $q : q'$ $(q' = 1 - q)$. As the result it is located closer to C than A, with $r : r'$ for the pair C and A.

The right part of Figure 5, i.e. (b), the concept/object is shown as a location in an triangle with three vertices A, B, and C. The location of the point G for the target concept is determined as the intersection point of line segments CC' and AA', where C' and A' are the points that divide the side AB so that $AC' : C'B = p : p'$ and the side BC so that $BA' : A'C = q : q'$ respectively.

The value of r for the ratio for C and A is calculated as follows: By Ceva's Theorem, we have $pqr = p'q'r'$. By using the definition $r' = 1 - r$ we have $pqr = p'q'(1 - r) = p'q' - p'q'r$, so we have $(pq + p'q')r = p'q'$. As the result, we have the following two formulas.

$$r = \frac{p'q'}{pq + p'q'} \qquad r' = \frac{pq}{pq + p'q'}$$

By using these formulats, we can calculate r and r' from p and q and the position marker for the pair CA in Figure 5(a) is displayed according to these values of r and r'.

For example, let $p = 0.7$ and $q = 0.6$, thus $p' = 0.3$ and $q' = 0.4$. Then we have $r = \frac{0.3\times0.4}{0.6\times0.6+0.3\times0.4} = \frac{0.12}{0.54} \approx 0.2$ (Figure 5).

Now we would like to represent G as a linear combination of A, B, and C. That is to find α, β, and γ so that $G = \alpha A + \beta B + \gamma C$, where $0 \le \alpha, \beta, \gamma \le 1$ and $\alpha + \beta + \gamma = 1$.

Let $s = C'G, t = GC$ in Figure 5. By applying the Menelaus's Theorem to the triangle $\triangle CAC'$ and the line BB', we have the equation $r(p+p')s = r'p't$. Thus $s = \frac{r}{r'p'}t$ holds because $p + p' = 1$. From this equation, $s + t = \frac{r+r'p'}{r}t$, and thus we have.

$$\frac{s}{s+t} = \frac{r'p'}{r+r'p'} \qquad \frac{t}{s+t} = \frac{r}{r+r'p'}$$

Since $C' = p'A + pB$,

$$G = \frac{t}{s+t}C' + \frac{s}{s+t}C = \frac{r}{r+r'p'}(p'A + pB) + \frac{r'p'}{r+r'p'}C$$
$$= \frac{1}{r+r'p'}(rp'A + rpB + r'p'C)$$

So we have

$$\alpha = \frac{rp'}{r+r'p'}, \quad \beta = \frac{rp}{r+r'p'}, \quad \gamma = \frac{r'p'}{r+r'p'}$$

From this result we can get the intended result.

$$\alpha : \beta = p' : p, \quad \beta : \gamma = rp : r'p' = \frac{rpq}{q} : \frac{r'p'q'}{q'} = q' : q,$$

and $\gamma : \alpha = r' : r$.

As a special case, we suppose the numbers a, b, and c are assigned to the three keys A, B, and C, respectively. Then we have:

$$p' = \frac{a}{a+b}, \quad p = \frac{b}{a+b}, \quad q' = \frac{b}{b+c}, \quad q = \frac{c}{b+c}, \text{ and}$$

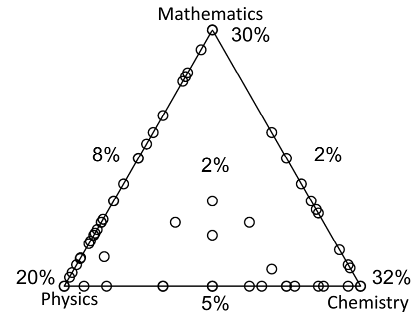

Fig. 6. Scattergram of All Students' Interest Ranges for Physics-Chemistry-Mathematics in Trichotomic Representation

$$r' = \frac{c}{c+a}, \quad r = \frac{a}{c+a}$$

Thus the equation $p'q'r' = \frac{abc}{(a+b)(b+c)(c+a)} = pqr$ is satisfied: i.e. the three lines AA', BB', and CC' meet at the same point G from the Ceva's Theorem.

Because of $r + r'p' = \frac{a}{c+a} + \frac{ca}{(c+a)(a+b)} = \frac{a(a+b+c)}{(a+b)(c+a)}$, we have:

$$\alpha = \frac{rp'}{r+r'p'} = \frac{a}{c+a}\frac{a}{a+b}\frac{(a+b)(c+a)}{a(a+b+c)} = \frac{a}{a+b+c}$$

Similarly, $\beta = \frac{b}{a+b+c}$ and $\gamma = \frac{c}{a+b+c}$. And thus we have the expression for G:

$$G = \frac{aA + bB + cC}{a+b+c}$$

C. Case study 1: Physics-Chemistry-Mathematics Analysis

As the first case study, we investigate the students' interest areas. In order to carry out the analysis we take up physics, chemistry, and mathematics as the three keys for interest areas. We use the NDCs 420, 430, and 410 as the indexes for them because these are for physics, chemistry, and mathematics, respectively. To be more precise, the actual NDC number's range for physics is $420 \le NDC < 430$, and in the same way for chemistry and mathematics. The waits of a patron are taken by the numbers of borrowed books having the corresponding keys; for example a =the number of borrowed books of NDC number 420, or from 420 upto 420.99 in the records for the patron.

Figure 6 shows the scattergram for all students in trichotomic representation together with the percentages according to borrowing patterns. The borrowing patterns are divided into 7 types; 3 types for those students who borrowed books of only one of the three keys, i.e. physics (P), chemistry (C), and mathematics (M), 3 types for those who borrowed only two categories, i.e. PC, CM, or MP, and the rest 1 category for those who borrowed all 3 categories of books, i.e. PCM.

One of the interesting findings is that only 2% of students borrowed these three categories and 83% students borrowed books from only one category. Another one is that among the ratios for only one category, the order is chemistry (32%) $>$ mathmatics (30%) $>$ physics (20%), which is different from our prediction. However for the pairs, physics&mathematics (8%) $>$ physics&chemistry (5%) $>$ chemistry&mathematics

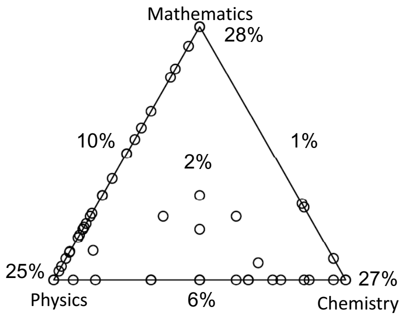

Fig. 7. Scattergram of Faculty of Sciences (SC) Students' Interest Ranges for Physics-Chemistry-Mathematics in Trichotomic Representation

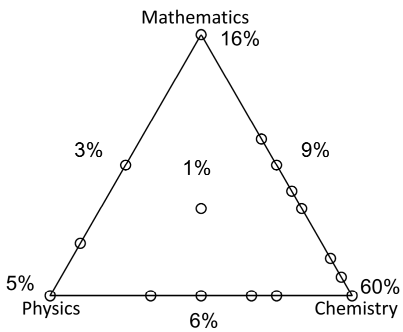

Fig. 8. Scattergram of Faculty of Agriculture (AG) Students' Interest Ranges for Physics-Chemistry-Mathematics in Trichotomic Representation

(2%), and thus the importance order of these three looks like physics > chemistry > mathematics. As a result of such a seemingly contradictory, we may say that chemistry is the most important subject as a single study subject for students as awhole. Thus part of physics's importance comes from that students need to learn physics in order to learn chemistry as it gives theoretical background of chemistry. Similarly, mathematics is important in combination with physics because it gives fundamental concepts and useful tools for analysis for physics.

In order to investigate more about the mutual relationship between physics, chemistry, and mathematics, we focus on the students who are affiliated in Faculty of Sciences (SC).

Figure 7 shows the scattergram for SC. It is obvious to see Figure 6 in comparison with Figure 7 that the students on the side of chemistry&mathematics are much smaller for SC than ALL. Also we find that the ratio for physics (25%) for SC is greater than that for all students (20%) by 5 points and the ratios for chemistry and mathematics are smaller in SC than in ALL. Still in their absolute values, the order of importance is mathematics>chemistry>physics, and thus physics is the least important for students in SC. So we can say that physics is relatively more important for SC students than for students in general.

Faculty of Agriculture (AG) is a constrasting example. Figure 8 shows the scattergram for AG. For AG the order is chemistry (60%) > mathematics (16%) > physics (5%), and the ratio of chemistry is dominatingly large whereas ratio of physics is very small. For the pairs chemistry&mathematics side is the biggest among three pairs.

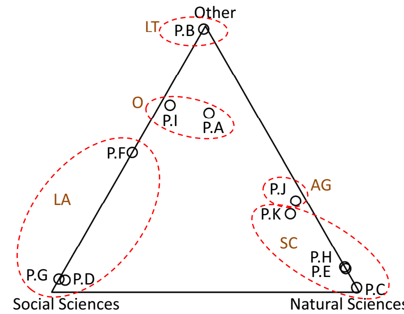

Fig. 9. Scattergram for Top 11 Students in Trichotomic Representation with Social Sciences-Natural Sciences-Other as Keys

D. Case study 2

It is quite popular to divide a person according to his or her major if the major is considered to be science-related or liberal-arts-related. In this section, we take the NDC categories 400 (Natural Sciences) as the representative to science-related subjects and 300 (Social Sciences) as to arts-related subjects. We add up other NDC categories as the third key in order to see the width of interest areas.

We chose the 11 students who appeared in Figure 1, Table I, and Figure 2 as the representatives. Figure 9 shows the scattergram for these students in the trichotomic representation.

It is interesting to see that different from Figure 2, students in the same affiliation are close in Figure 9. Three LA students are located in the left lower area, from vertex for social science to the middle area toward other on the side for social sciences&other. Among these 3 students only P.F is separated as this student is different in his or her interest area range (see Table I.

SC students, on the other hand, are located in the lower area at the right side; from natural science vertex to around the 2:1 point toward the vertex for other. Similar to LA case, one student P.K is separted from other students because he or she has wider interest are range.

The student P.J, who is the only student affiliated in AG, is located very close to P.K. Even though they are different in affiliation and patron type according to Table I, they are very similar in their interest areas.

The 2 students P.A and P.I classified as O (Other) are located in the upper area relatively near the vertex for other subjects. This result is understandable because the students in O might not strongly related to either social or natural sciences.

The only student P.B who affiliated in LT (Letter) is located to the vertex for other. According to Figure 1, he or she is strongly literature-oriented ans should locate far away from both social and natural sciences.

Figure 10 shows the scattergram for 12 faculties plus other (O) in the same keys as in Figure 9. As are expected the faculties considered as liberal-arts-related are located near the left side and those considered as science-related are located near the right side. In the firure are two dashed lines for dividing the whole area into 3 smaller areas; social science oriented, natural science oriented, and intermediate.

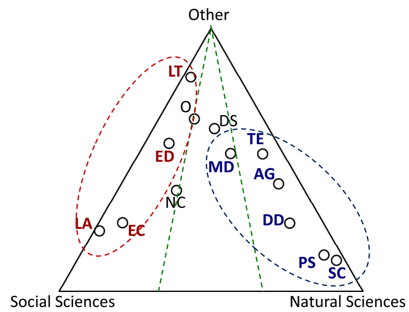

Fig. 10. Scattergram for Faculties in Trichotomic Representation with Social Sciences-Natural Sciences-Other as Keys

It is interesting to see that MD (Medicine) is located in the intermediate area. Even though MD is normally considered as being included in the science-related, it is more liberal-arts oriented than other science-related subjects, probably because they are more expected to be interested in humans not only from biological view point but also from psychological and behavioural view points.

The 2 faculties DS (Faculty of Design, or Faculty of Art and Engineering in Japanese) and NC (Faculty of New Century, or Faculty of 21st Century in Japanese) are difficult to specify if they are liberal arts-related or science-related. In Figure 10, we can see that both of them are located in the intermediate area. As we see further, DS is more science-related than arts-related and NC is located in the left most part of intermediate area and thus we may consider it is almost arts-related.

The other group (O) is also near the border line between arts-related and intermediate areas just like NC. So the students in other group are rather arts-oriented than science-oriented in general.

As we see the hights, or the degree of closeness to the vertex for other, of faculties, LT (Letter) is the highest, which means that the NDC category 900 (Literature) is included in the other subject and thus the students like P.B are located high in the trichotomic representation.

On the other hand, SC (Sciences) and PS (Pharmaceutical) are strongly centered to natual sciences than any other faculties. Among liberal arts-related faculties, LA (Law) and EC (Economy) are most strongly centered to social sciences.

We can recognize also in this case study that trichotomic representation gives new findings about the chracteristic behaviours not only of the 11 students shown in the previous section but also of the faculties as well.

E. Case study 3

V. Conclding Remarks

The conclusion goes here.

Acknowledgment

This work was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 24500318, 2013.

References

[1] T. Minami, H. Sazuka, S. Hirokawa and T. Ohtani, *Living with ZK – An Approach towards Communication with Analogue Messages*, 1998 Second International Conference on Knowledge-Based Intelligent Electronic Systems (KES 1998), Eds. L.C. Jain and R.K. Jain, pp. 369–374, 1998.

[2] NHK, *Close-Up Today: "Pamyu-Pamyu", "Je-Je-Je" – Mystery of Great Propagation of Onomatopoeia –*, 2013. (in Japanese) `http://www.nhk.or.jp/gendai/kiroku/detail02_3362_all.html`