Title	The theory of network analysis of teaching
Sub Title	
Author	佐伯, 胖(Saeki, Yutaka) 浦, 昭二(Ura, Shoji)
Publisher	慶応義塾大学藤原記念工学部
Publication year	1966
Jtitle	Proceedings of the Fujihara Memorial Faculty of Engineering Keio University (慶応義塾大学藤原記念工学部研究報告). Vol.19, No.73 (1966.) ,p.77(27)- 103(53)
JaLC DOI	
Abstract	In this article, a system has been developed, through which optimal teaching course can automatically be provided. The system has been designated as "Network Analysis of Teaching (NAT)." This report is a brief statement of the theoretical study in NAT. Here, the authors attempt to construct the general procedure for analyzing a given subject into elementary items and arranging them in an optimal sequence of teaching. Models of various patterns are proposed here, by which the optimal teaching sequences may justly be deduced. NAT has been developed principally for the purpose of the proper application of computer to the study of educational processes.
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001004-00190073-0027

The Theory of Network Analysis of Teaching

(Received April 26, 1966)

Yutaka SAYEKI*

Shoji URA**

Abstract

In this article, a system has been developed, through which optimal teaching course can automatically be provided. The system has been designated as "Network Analysis of Teaching (NAT)." This report is a brief statement of the theoretical study in NAT. Here, the authors attempt to construct the general procedure for analyzing a given subject into elementary items and arranging them in an optimal sequence of teaching. Models of various patterns are proposed here, by which the optimal teaching sequences may justly be deduced. NAT has been developed principally for the purpose of the proper application of computer to the study of educational processes.

I. Introduction

Today, computers are widely used in the field of teaching machines. Most of them have various branching systems according to student's responses in frames and scores in pre-tests, IQ tests, etc., all of which are stored in a computer. However, we should not be over-dependent upon such statistical data, in deciding a course of teaching.

A programmed sequence of teaching may be said to be a "hypothesis" concerning teaching processes. Teachers or programmers of teaching program make the hypothesis that a student who has learned such and such items can easily master the item concerned. Teachers should induce this hypothesis from the various analyses of the given subject material rather than from statistical data of students' responses to frames.

Whether or not a student should branch here or there should be decided mainly from the logical relations or from the context of the items in the subject concerned. First, teachers should carefully pick out all the items necessary for students' understanding of the subject. Second, teachers must inquire and analyze exhaustively all the logical relations among the items picked out. If some students cannot pass a certain frame, teachers should re-examine the hypothesis itself concerning the logical

*佐 伯 胖, Graduate student, Doctor Course, Graduate School of Engineering, Keio University.

**浦 昭 二, Professor, Faculty of Engineering, Keio University.

relations among the items related to the frame. Letting a student branch off in case of an erroneous response may tend to prevent the re-examination or re-organization by teachers of the contextual sequence of frames.

Moreover, there must exist a more general hypothesis in teachers' minds, when they decide teaching sequences. Teachers must have adopted some educational principles in deciding a teaching sequence. These principles might be inferred from various educational experiences and various psychological theories. So, if some number of students cannot pass a certain frame, teachers must examine also the educational principles themselves by which they have judged that the sequence adopted is more proper than any other feasible sequence.

Unfortunately, today's computer-based teaching machines[1),2),3)] are not intended to facilitate these kinds of examination of teachers' hypotheses which should be constructed through contextual and logical analysis of subject materials and through the educational principles adopted to decide the optimality of teaching sequences.

Besides, it is a common belief that we should make better use of students' errors, in order to develop heuristic or creative thinking and insight ability. This is one of the reasons why many researchers are trying to construct complicated branching systems with computer equipment. However, this kind of belief has little basic theoretical ground. Before adopting this common belief, we should inquire what the heuristic is, or what the creative thinking, insight ability, etc., are. For this purpose, we must also construct the general methodology to analyze subject materials, even if the subject is heuristic, and also we must clarify, in operational terms, educational principles which decide the optimality of teaching sequences.

Therefore, the author believes that computer techniques should be developed more for the purpose of promoting the research of teachers in the educational analyses as mentioned above, than for the purpose of making complicated branching systems.

The Network Analysis of Teaching (NAT) has been developed for this purpose by the author. The basic idea of the NAT theory has already been disclosed in his previous report.[4)] However, since then, the author has made a few experiments controlled by a computer, and has changed some aspects of his theory. For ex-

1) N. A. Crowder, Automatic Tutoring by Intrinsic Programming. In A. A. Lumsdaine and R. Glaser (eds.), Teaching Machines and Programmed Learning. Washington D. C. National Education Association. 1960.
2) D. L. Bitzer, PLATO: An Electronic Teaching Device. Urbana, Ill. Univer. of Ill., Coordinated Science Laboratory. 1963.
3) L. M. Stolurow, A Model and Cybernetic System for Research on the Teaching-Learning Process. In Programmed Learning, Vol. 2, No. 3, pp. 138–157. Sweet and Maxwell, London. 1965.
4) P. Y. Sayeki, Network Analysis of Teaching. In the Science of Learning (English Edition), Vol. 2, No. 1. The Center for the Science of Learning, Keio University, Tokyo. 1966.

ample, the author adds to the attempt to set an evaluation measure of the optimality of teaching sequences, and modifies the model of "familiarity" laying stress upon students' learning history in the course of teaching. This report deals with the NAT theory at its present stage, being a condensed statement of the master's dissertation which was submitted to the Department of Administration Engineering of Keio University by the author. In his master's dissertation, the author dealt also with the practical applications using computers, which are not included in this report.

We hope that the NAT theory will solve many important problems concerning the theories of educational processes.

II. Analysis of educational goal

In general, it is often convenient to analyze teachers' goals in teaching a given subject, by the following three systems A, B and C. By this analysis, teachers can find the elementary items necessary and sufficient for introducing students into the precise concepts of the subject.

System A

System A may be called a hypothetico-deductive system. It is expressed in short in the following proposition:

If X, then Y. (X; hypothesis, Y; conclusion.)

This system is constructed only by deductive procedures to prove a theorem. As long as in this system, we have no concern with any terminology in the process of proofs. We are concerned only with whether the propositions deduced from a hypothesis are true or false. We may call this system "closed", since the logical elements are assumed to have been defined and clarified in the other Systems, and we need not inquire why we postulate this or that hypothesis, or what the meaning of the conclusion is, or what can be inferred from the fact that if X, then Y.

Method to find System A

We may find all the nodes in System A by inquiring what the statement is that can answer deductively the question "Why?" Then, we should explain all these questions by statements which can be expressed in the proposition that if X, then Y.

System B

System B may be called a semantic and heuristic system. The reason why we call it "semantic" is that it is concerned with the meaning of the terms used in System A. We must inquire as to the meaning of the hypothesis and conclusion. We must also explain by well-known or pre-defined terms what may be indicated from the fact that if X, then Y. Definitions of terms used in System A are included System B. All statements in System B are only inductive, or analogical explanations. Applicational instances of the concepts are included in System B as long as they seem to facilitate understanding of the concepts.

Another part of System B is "heuristic". A heuristic method (or simply, heuristic) is a process that may facilitate solving problems in a specific class but offers no guarantee of doing so.[5]

In proving System A we often use some heuristic methods besides the simple substitution and detachment methods. If these heuristics are well formulated and can be used in other fields, we put them in System B. It must be clear that they are empirically applicable to broader fields even though we need not prove theoretically their applicability.

Though substitution and detatchment methods are, of course, heuristic, perhaps we need not include them in System B, since they are so familiar to us in theory proving problems.

Some kind of heuristics are often buried quite implicitly in our minds. For example, in most mathematics, we develop a concept to a more and more generalized or unified one. The generalization or the extension of concepts may be said to be the "heuristic way of thinking." We often fail to find such heuristics explicitly. Let us take another example. We should always ask ourselves in the procedure of reasoning whether a statement is a necessary condition or a sufficient one. This examination is an important heuristic way of thinking in theoretical fields.

Method to find System B

1) B from A and A's neighborhood

First, we should clarify the concept of neighborhood of a given subject. A neighborhood is a set of subjects which we do not intend to teach at this time but which we taught just before or we shall teach just after a given subject.

Now, we can pick out System B items from the relations between System A and its neighborhood by the following inquiries.

(i) We must find the statements introducing an Aporia (*ἀπορία*)[6] from the concepts already well-known or regarded to have been introduced before teaching the subject. Here, an Aporia means a problem situation where students feel a sense of deadlock and seek for some new information to solve it. We must ask

5) A. Newell, J. C. Show and H. A. Simon, Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristics. In E. A. Feigenbaum and Feldman (eds.), Computers and Thought, pp 113–114. McGraw-Hill, N. Y. 1963.

6) *ἀπορία* embarrassment, doubt, new difficulty, (Greek-English Dictionary, compiled by Karl Feyerabend, second edition, Berlin-Schöneberg, 1910.)

ourselves how you lead to the Aporia by gathering the facts.

(ii) We must find the statements introducing a hypothesis which seem to solve the above Aporia. They are introduced by inductive or analogical procedures from well-known facts or pre-defined concepts.

(iii) We must find the facts which are taught and which will also lead to another Aporia in the next subject.

2) B used in A

We must find the heuristics used in System A. For instance, if we use some techniques or some empirical rules to solve the theory-proving problem in System A, we must clarify these techniques and rules and regard them as System B. We need not include simple logical rules, *e.g.* substitution and detatchment methods, since they are so familiar to us that we can use them without being taught intentionally.

3) Heuristics between A and "open" fields

(i) We must define every term used in System A by clear and well-known terminology.

(ii) We must explain the meaning of the conclusion of System A by the terms or examples from various other fields. We must also ask ourselves what field we can apply it to. The field we use to explain the concept must be quite another field.

(iii) We must explain when we can apply this concept. We must concentrate students' minds on the conditions when we apply the theorem.

System C

System C may be called an "algorithmic" system. It concerns only the algorithm[7] solving the practical problems of the subject. Here, an algorithm is a process sequenced by simple operations which guarantees the solution of all problems in a specific class, if the problems have a solution.

Method to find System C

1) First we must regard ourselves as a computer capable only of simple operations, such as addition, subtraction, multiplication, division, judgement of a condition, etc.

2) Then we must inquire of ourselves what kind of problems in the subject are solvable by this computer, and also how to solve them. If we can find a process sequenced by simple operations which guarantees the solution in a specific class, that is the algorithm for the subject, and also System C.

3) However, there is a subject well-known as an algorithm in itself. For example,

---

7) A. Newell, J. C. Show and H. A. Simon, Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristics. In E. A. Feigenbaum and Feldman (eds.), Computers and Thought, pp 113–114.

the Laplace Transformation is well-known as an algorithm, though its solvability cannot be proved even in the higher mathematics.

4) If often happens that an algorithm is buried in our mind. An expert can solve a problem quite fluently because he has an algorithm implicitly. In some cases, one of the differences between an expert and a beginner is that an expert has a good discriminant algorithm for classifying problems into several patterns. In other words, at first glance, an expert recognizes the pattern of the given subject and then applies a proper heuristic or an algorithm to solve the problem. For example, discrimination between letter A and B can be made by examining whether it has a peak or not, though it is not realized explicitly in our minds.

Note:

Note that in these three analyses, A, B and C, we need not concern ourselves whether we can really teach this or that item to students. We decide it after analyzing a network which will be shown in the next chapter.

III. Nodes, networks and blocks

Condition of items to be necessary and sufficient

In the previous chapter, we have established a method to find elementary items necessary and sufficient for understanding a given subject. The condition of items to be necessary and sufficient is quite important to make a good analysis. We must clarify their meanings.

That an item is necessary for the subject means that a teacher needs the item to lead a student to the understanding of the subject. Items that are sufficient for the subject are those which, if a teacher uses a proper teaching method he needs no other information than those.

Similarly, for any item, we can say whether is necessary and sufficient. In this case, we need those items and no other items to lead a student to the understanding of an item, as long as we use a proper teaching process. So, in examining each item, we must inquire of ourselves whether or not we need other information to induce or deduce that item.

Nodes and networks

After analyzing Systems A, B and C, items must be necessary and sufficient for each other. Then we may call those items nodes, since they will make a network showing all relations among them. Let us take an example of teaching the following proof:

A1. Prove that if we are given $a^m = a \cdot a \cdots a$, we can get the following exponential laws, where m, n are any positive integer.

$$a^m \cdot a^n = a^{m+n} \qquad a^m \cdot b^m = (ab)^m$$

$$a^m \div a^n = a^{m-n} (m > n) \qquad b^m / a^m = (b/a)^m$$
$$(a^m)^n = a^{mn}$$

We can prove this proposition by the following hypothetico-deductive items. (Table 1.)

Table 1. Necessary and Sufficient Nodes for Proving A1.

Node	Elementary Item	Proof	Necessary and Sufficient Nodes
A6	$a^m = a \cdot a \cdots a$ where $m > o$, int.	Assumed to be mastered.	
A7	$a^m \cdot a^n = a^{m+n}$ where $m, n > o$, int.	$a_m = a \cdot a \cdots a$ $\therefore a^m \cdot a^n = \overbrace{(a \cdot a \cdots a)}^{m} \overbrace{(a \cdot a \cdots a)}^{n}$ $\therefore a^m \cdot a^n = a^{m+n}$	A6
A8	$a^m \div a^n = a^{m-n}$ $m > n > o$, int.	$a^m \div a^n = \frac{a^m}{a^n} = \frac{\overbrace{a \cdot a \cdots a}^{m}}{\underbrace{a \cdot a \cdots a}_{n}} = a^{m-n}$	A9 A6 A10
A9	$a \div b = \frac{a}{b} = a \times \frac{1}{b}$	Assumed to be mastered	
A10	$\frac{ac}{ab} = \frac{c}{b}$ $(ab \neq 0)$	Assumed to be mastered	
A11	$(a^m)^n = a^{mn}$ where $m, n > 0$, int	$a^m = a \cdot a \cdots a$ $\therefore (a^m)^n = (a^m)(a^m) \cdots (a^m)$ $= (a \cdots a)(a \cdots a) \cdots (a \cdots a)$ $= a^{mn}$	A6
A12	$(ab)^m = a^m b^m$ where $m > 0$, int.	$(ab)^m = (ab)(ab) \cdots (ab)$ $= (a \cdots a)(b \cdots b)$ $= a^m b^m$	A6
A13	$\left(\frac{b}{a}\right)^m = \frac{b^m}{a^m}$ where $m > 0$, int.	$\left(\frac{b}{a}\right)^m = \left(\frac{b}{a}\right)\left(\frac{b}{a}\right) \cdots \left(\frac{b}{a}\right)$ $= \frac{bb \cdots b}{aa \cdots a} = \frac{b^m}{a^m}$	A6 A14
A14	$\frac{b}{a} \cdot \frac{d}{c} = \frac{bd}{ac}$ $(ac \neq 0)$	Assumed to be mastered.	

Connecting these nodes, we may get the following network. (Fig. 1)

We can easily get a sub-network for System A. Similarly we can get sub-networks for Systems B and C. Finally, we may combine the sub-networks into one network. However, it often happens that we should divide System C items into several blocks.

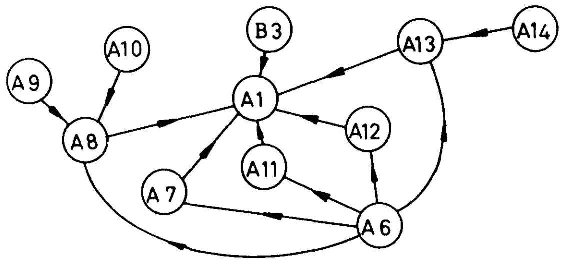


Fig. 1. The network for proving A1.

Blocking of System C

If we want to teach students only an algorithm or an operation to solve problems, we need not necessarily use items in System A or B. We can teach an algorithm only by System C. However, we have adopted the educational view that we should not introduce any concept without reason. An algorithm includes no proof, *i.e.* reason, regarding its own solvability or applicability. Proofs should have been included in System A. So, we make it a rule to introduce an algorithm after its solvability is proved. If its solvability cannot be proved even in a higher concept, there must be some kind of reasoning, *e.g.* induction or analogy, before introducing the algorithm. Then it is clarified that before introducing the algorithm we must use some items of Systems A and B.

However, another problem occurs. Should we introduce an algorithm every time after its solvability is proved, or should we introduce a set of algorithm after their solvability and applicability are proved?

This is the kind of problem teachers must have decided before teaching. Then we must divide System C into several blocks.

In a block, a set of algorithms are taught successively. But between blocks, we must insert some of the reasoning in System A or B.

Let us take an example of teaching the following algorithms to get the value of any given exponential expression by the exponential laws.

C 1. If you find a in the expression, transform a into a^1. $(a \to a^1)$

C 2. $a^{m_1} \cdot a^{m_2} \cdots a^{m_n} \to a^{m_1+m_2+\cdots+m_n}$

C 3. $(a^m)^n \to a^{mn}$

C 4. $a^m\, b^m\, c^m \cdots f^m \to (abc \cdots f)^m$

C 5. $\div a^n \to a^{-n}$

C 6. $1/a^n \to a^{-n}$

C 7. $(a/b)^m \to a^m\, b^{-m}$

C 8. $a^0 \to 1$

C 9. $a^{-n} \to 1/a^n \ (n>0)$

C10. $a^m \to a \cdot a \cdots a$

In this example, we may block out System C for teaching the process of applying the exponential laws, as follows:

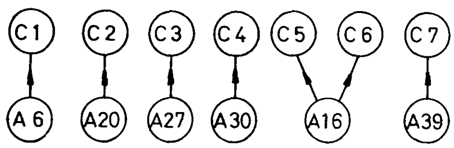
Block 1: C1, C2, C3, C4, C5, C6, C7.

Block 2: C8, C9, C10.

(Block 1 is the methods to simplify an exponential expression, while Block 2 is the methods to get the value of a simplified exponential expression.)

How to decide the optimal sequence in a block

In System C, nodes have no logical relation to each other, since they are all elementary algorithms. They are logically related only to nodes of other Systems. (Fig. 2.)



A20: If $a^0=1$, $a^{-n}=1/a^{-n}$, then $a^m \cdot a^n=a^{m+n}$ where m, n are any integer.

A 6: $a^m=a \cdot a \cdots a$ where $m>0$, int.

A16: If $a^0=1$, $m=0$ in $a^m \div a^n=a^{m-n}$, it is necessary that $a^{-n}=1/a^n$.

A27: If $a^0=1$, $a^{-n}=1/a^n$, then $(a^m)^n=a^{mn}$ when m, n are any int.

A30: If $a^0=1$, $a^{-n}=1/a^n$, then $a^m b^m=(ab)^m$ when m, n are any int.

A39: If $a^{-n}=1/a^n$, $a^m b^m=(ab)^m$, $(a^m)^n=a^{mn}$, for any int. of m, n, then we can deduce that $(a/b)^m=a^m/b^m$ for any int. m.

Fig. 2. Logical relations between nodes of System C and those of System A.

Then, if we want to teach nodes in a System C block, we can never decide the optimal sequence by logical contexts in the block. In our example here, we can teach Block 1 only after A6, A20, A27, A30, A16, A39 have been taught. (Fig 3.)

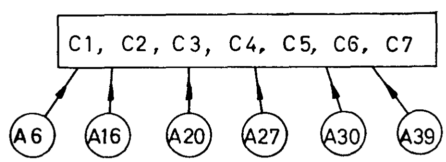


Fig. 3. The relation between a block and nodes.

So, we may introduce the following two scales to decide the sequence in a block.

Scale of importance

We may generally say that an item is important when it is used frequently in the real world. In most cases, we would sequence items of teaching by an order of importance. So in our theory, we apply an ordinal scale of importance to all nodes in a block according to a teacher's judgement. If values on the scale are all different we can decide the sequence by their order of values. However, we may put the same value on different nodes. In such a case, we must apply the following scale.

Scale of simplicity

We may generally say that an item is simple when the number of operational elements in its algorithm are very few. (Algorithmic items always consist only of operational elements.) We apply ordinal scale or simplicity to the nodes which have the same value on the importance scale.

If we cannot find the optimal sequence according to the above two scales, we can teach the nodes at random which have the same value on a simplicity scale.

There may be another approach to decide the sequence in a block. We may consider values of scales as weights of nodes. In this case, we apply the above two scales independently on all nodes in a block, multiplying two values (weights) of a node, ordering the nodes according to the product values of the two different scales.

Then we can find the optimal sequence of nodes in a block. However, it is more important to decide the sequence of all nodes and blocks in a network.

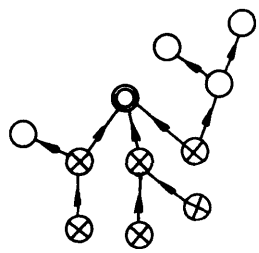
IV. How to decide the optimal sequence

Terminology

At first we must define the terminology used in our theory.

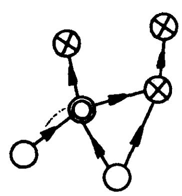
1) An *input-node* is a node which necessarily precedes and leads to the node concerned.

◎ The node concerned
⊗ Input-node



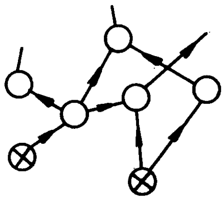
2) An *output-node* is a node which cannot be learned if the node concerned is not learned.

◎ The node concerned
⊗ Output-node



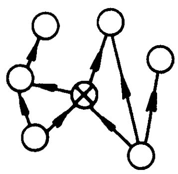
3) A *beginning-node* is a node which is assumed to have been mastered before teaching the subject in the network, *i.e.* which has no input-nodes in the original network.

⊗ Beginning node



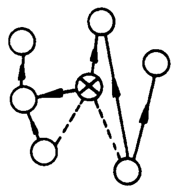
4) A *learned-node* is a node which has been learned, and whose input-nodes have been used and are no longer necessary to lead to the concept of the node concerned. All input paths which connect input nodes and the node concerned must be considered as disconnected because we assume that if we teach a node, students will not forget it. A network must be revised whenever a node is learned. An original network consists of nodes which are not learned. A learned network consists only of learned-nodes.

Before learning



⊗ The node to be learned

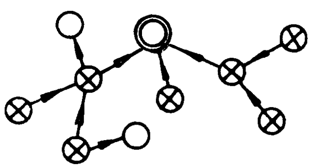
After learning



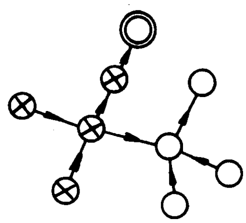
---- Line disconnected
⊗ The node learned

5) An *input-number* is the number of input nodes of the node concerned.

Input-number=7.



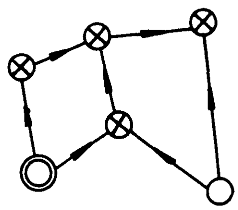
Input-number=4.



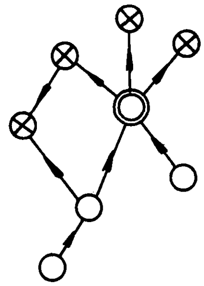
◎ The node concerned ⊗ Input node

6) An *output-number* is the number of output-nodes of the node concerned.

Output-number=4.



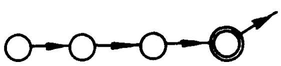
Output-number=4.



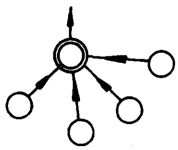
◎ The node concerned ⊗ Output-node

7) An *input-level* is the number of simultaneous steps which lead to the node concerned. (So, beginning nodes or learned nodes are all zero-input-level nodes.)

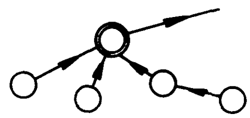
Input-level=3.



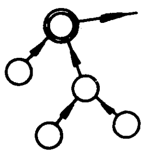
Input-level=1.



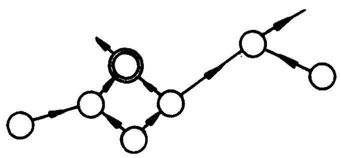
Input-level=2.



Input-level=2.



Input-level=3.



◎ The node concerned

We can easily get an input-level by the following formula:

Input-level=Input-numbers−(Beginning node number+learned node number)+1

Pattern models for the optimal sequence of nodes

Using the above several terms, we shall introduce pattern models which automatically decide the optimal sequence of nodes in a network. For this purpose, we shall at first examine criteria which shall be applied when an experienced teacher wants to decide on an optimal sequence of items in teaching. We may find the following criteria.

1′) Easiness of an item to be learned.
2′) Familiarity of the concepts used.
3′) Applicability.
4′) Importance.
5′) Teachability.

1′) Easiness of an item to be learned

When an experienced teacher comes to the place where he must choose one of the items to teach, he may examine whether or not it is easy for a student to learn it. We may generally say that an item is easy to learn when the number of concepts used in explaining it are only a few.

Then, we can translate the criterion of easiness into the following pattern model expressed by our network terminology.

An *easy node* is the node whose number of input-nodes is minimum, and whose input-level is minimum.

2′) Familiarity

It will often happen that an item easy in the above sense is still difficult for a student to understand. One of the factors to be considered is the familiarity of an item. When the concepts used in explaining an item are quite unfamiliar, a student will find it difficult to understand. We may justly say that an item is familiar when many of the concepts used in explaining it have been taught recently.

So, the familiarity should be decided by the whole history of the student's learning, that is, the familiarity may be effected not only by what items a student has learned but also by when he learned each of them. Therefore, we have made it a rule to record a history of learning by the following method.

First, we have defined *the familiarity indicator*. When we decide a node to teach, we put the familiarity indicator increased by 1 to all its immediate output-lines which are emitted directly from the node. The familiarity indicator of unlearned nodes are all 1.

For example, in the following figure (Fig. 4), if we decide to teach A5 first of all, its familiarity indicator must be changed to 2. If we decide next to teach A3, its familiarity indicator must be 3.

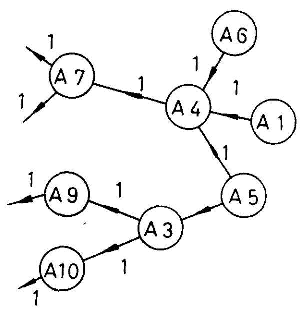


Fig. 4.

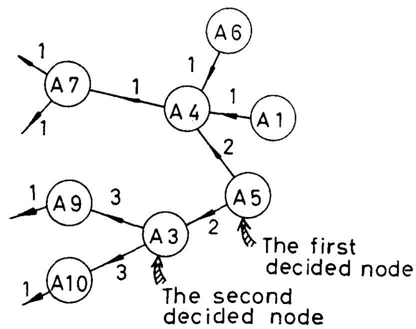


Fig. 5

Then we can define the familiarity as follows:

The *familiarity* is the sum of familiarity indicators of the immediate input-line of the node concerned. For example, the familiarity of A5 equals to 1, that of A3 changes in 2, that of A4 changes in 4(=2+1+1), that or A9 changes to 3. (Fig. 5.) Then the most familiar node is the node whose familiarity is maximum.

3′) Applicability

Another factor of explainability to be considered in choosing a new node is the easiness of a node to lead to the next-nodes, *i.e.* the applicability. When we cannot teach a true concept of an item by one node, we can explain it by applying its concept, developing or changing slightly its meaning, at the next node. For example, the concept of extension in mathematics may be difficult to be understood in only one instruction. But a student may understand it when he studies several instances of its application or a slight development of the concept introduced later. All its applications or developments are to be included in System B of the subject.

Then we can translate applicability into the following pattern model.

An *applicable node* is the node whose majority of its next nodes are easy or familiar nodes. The number of these kind of next nodes must be maximum.

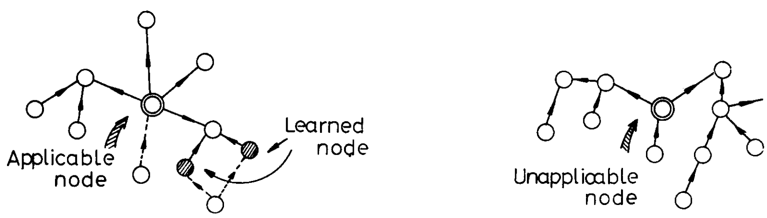


Fig. 6. Applicable nodes.

4′) Importance

We may often say that some nodes are important and others are trivial, in deciding what to teach. Then we must consider the importance of each item. We may generally say that an item is important when we cannot teach most of the items in the subject without it.

However, we may say that an item is important because it is quite useful in the actual world, or indispensable when a student goes into a higher concept, though in the field of the concerned subject, it may seem to be trivial. In such a case, the higher concept must have been included in the network, since in analyzing the three Systems, A, B and C, we need not be concerned with whether an item is a lower or a higher concept. We must include everything related to the subject in a network.

We may say an item or a concept is important from an *a priori* imperative without any reason. But even in such a case, we must teach its importance concretely, *i.e.* we must explain its importance by various instances or applications or development. And also we must include these materials in a network.

Then the only thing we have to do is to examine the "closed" field of a network, as long as we analyze the subject precisely and completely, and make a good network of Systems A, B and C.

We can find an important node quite easily by the following pattern model.

An *important node* is the node whose number of all output-nodes is maximum.

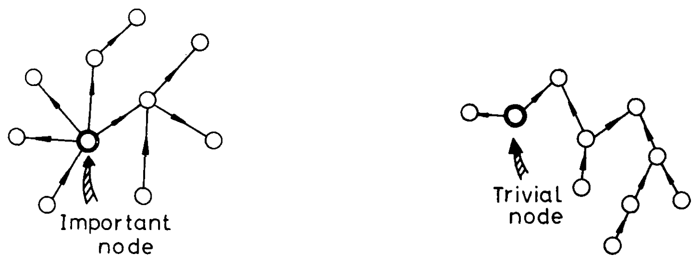


Fig. 7. Important nodes

5′) Teachability

Even if an item is easy, familiar, applicable or important, we cannot teach it actually unless it is teachable. We may say that an item is teachable when we can explain it exclusively by concepts which have been learned or which are assumed to have been mastered before introducing the subject concerned.

Then we can translate teachability into the following pattern model.

A *teachable node* is the node whose input-level is zero. We include a beginning-node in a teachable node though it is not necessary to teach and is assumed to be mastered.

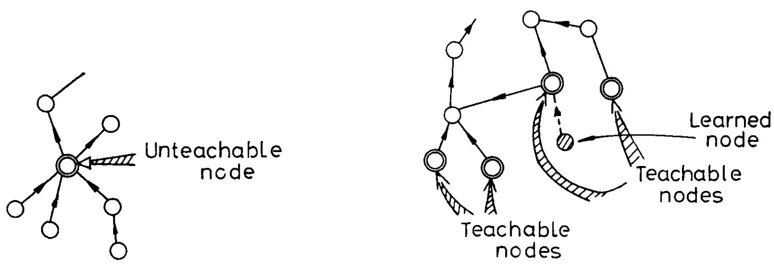


Fig. 8. Teachable nodes

How to produce the optimal sequence of nodes

Now, we must apply the criteria, easiness, familiarity, applicability, importance and teachability, to decide the optimal node to be selected, in the course of teaching. However, there are various ways of thinking as to which criterion we should take first among them. We may classify them into the following three classes according to different demands.

1) Teacher's demands
    i) Importance
    ii) Applicability
2) Student's demands
    i) Easiness
    ii) Familiarity
3) A logical demand
    i) Teachability
    (we may say that we cannot logically teach a node unless it is teachable.)

First, we shall examine which demand we should take precedently between teacher's and students'. Then, we shall decide how to find an optimal teaching course by connecting these criteria.

Let us take an example to teach the following moral proposition.

"If you have a seat in a street-car or in a bus, you should look to see whether an old man or woman, or a mother carrying a baby in her arms, is standing nearby. Then, if you find one of them, you should offer your seat to him or her, with a smile."

Now, let us analyze and pick up all the necessary and sufficient nodes through the analysis of the three systems, A, B and C.

System A

We may find all the nodes in System A by inquiring what the statement is that can answer deductively the question "Why"? Then, we should explain all these questions by statements which can be expressed in the proposition that if X, then Y.

Why should we look to see for an old man or woman or a mother carrying a baby in her arms in a street car or in a bus on the condition that we have a seat and that we are healthy enough?

Answer:

A1) If you have a seat in a street car or in a bus, and if you are healthy enough, you should look out to see whether an old man or a woman, or a mother carrying a baby in her arms, or someone under a similar situation, is standing nearby, since if you find one of them you should offer your seat to him or her.

Why should we offer our seat if we find an old man or woman or a mother carrying a baby in her arms.

Answer:

A2) If you can offer your seat to someone, it is your duty to offer it to an old man or woman, or a mother carrying a baby in her arms, if one of them is standing nearby.

Why should we select an old man or woman, or a mother carrying a baby in her arms among others?

Answer:

A3) An old man or woman, or a mother carrying a baby in her arms is apt to be tired and feel it difficult to keep standing in a street-car or in a bus.

Why should we offer a seat to someone?

Answer:

A4) If you ought to be kind to someone, while having a seat in a street-car or in a bus, you can be kind by offering your seat to the one who finds it most unpleasant to keep standing in a street-car or in a bus.

Why are we justly said to be kind to someone if we offer our seat to someone who finds it most unpleasant to keep standing in a street-car or in a bus?

Answer:

A5) If a person who finds it most unpleasant to keep standing in a street-car or in a bus, is offered a seat by chance, his pain or unpleasant feeling will be much relieved.

Why should we relieve someone from his pain or unpleasant situation?

Answer:

A6) If you relieve a person from his pain or unpleasant situation you may well be said to be kind.

Why should we smile when we are offering a seat?

Answer:

A7) You should smile when you are doing a kind act to someone, if you should be kind to someone.

A8) If you should be kind to someone you should give him a pleasant feeling as much as possible.

A9) A smile usually gives people a pleasant feeling.

System B

Definition

B1) To be kind to someone, or do a kind act to someone, is:

1′) to relieve someone from his difficulty or unpleasant situation,

2′) to give him a pleasant feeling.

Postulate

B2) Be kind to all people around you as much as possible.

Neighborhood

B3) By the same reason, if you find a person physically handicapped standing in a street-car or in a bus, while you have a seat, you should offer him your seat, since a physically handicapped person finds it difficult to keep standing in a moving vehicle.

System C

Algorithm

C1) How to offer your seat to someone. (refer to the usual manners.)

Network

Now, we can construct a network connecting all these nodes. (Fig. 9)

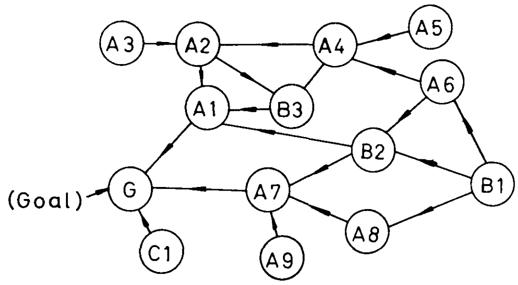


Fig. 9. A network for teaching a moral proposition

Decision model from teacher's demands

A teacher generally demands to teach fundamental and important nodes first even in case that the concepts regarding them may be rather unfamiliar to a student. So, we may construct a decision model based principally on the teacher's demands. This decision model first chooses the nodes whose importance (=output-number) is maximum.

Now, let us see the following fact. If we postulate that the most important node is *not* teachable, it must have some input nodes since an unteachable node is the node whose input-level is greater than 1. Since the output number of those input nodes is greater than that of the node concerned, there must be another node whose output number is greater than that of the most important node. This is a contradiction.

Since, as we have proved above, the most important node is always "teachable", we can decide the most "important" node to teach, as long as we can unify the node only by maximizing output-numbers. However, it often happens that several nodes have the same maximum output-number. In such a case, we may apply another criterion from the teacher's demands, *i.e.* applicability, to the nodes which have already been found to have the maximum output-number. If we can find a unified node whose applicability is maximum, we can teach it since it is also the most important node and therefore teachable. If there are still several nodes which maximize both importance and applicability, we should apply familiarity and easiness successively until we can get the unified node to teach.

If we cannot still find the unified node to teach even by these criteria, we may choose at random one of those nodes which have passed through the above four criteria. Then we can actually find the unified node to be taught in any case. (Fig. 10)

In our example, we can decide the optimal sequence of nodes principally from the teacher's demands.

(1) First, we must choose B1 since its output number is maximum. (The most important node.) Then, the familiarity indicators on its immediate output-lines must be changed to 2.

(2) Next, we must choose B2 since its output-number is maximum. The familiarity indicators on its immediate output-line must be changed to 3.

(3) We have the two nodes, A6 and A5, which maximize the output-number, and their applicability is also the same. However, the familiarity is larger in A6 than

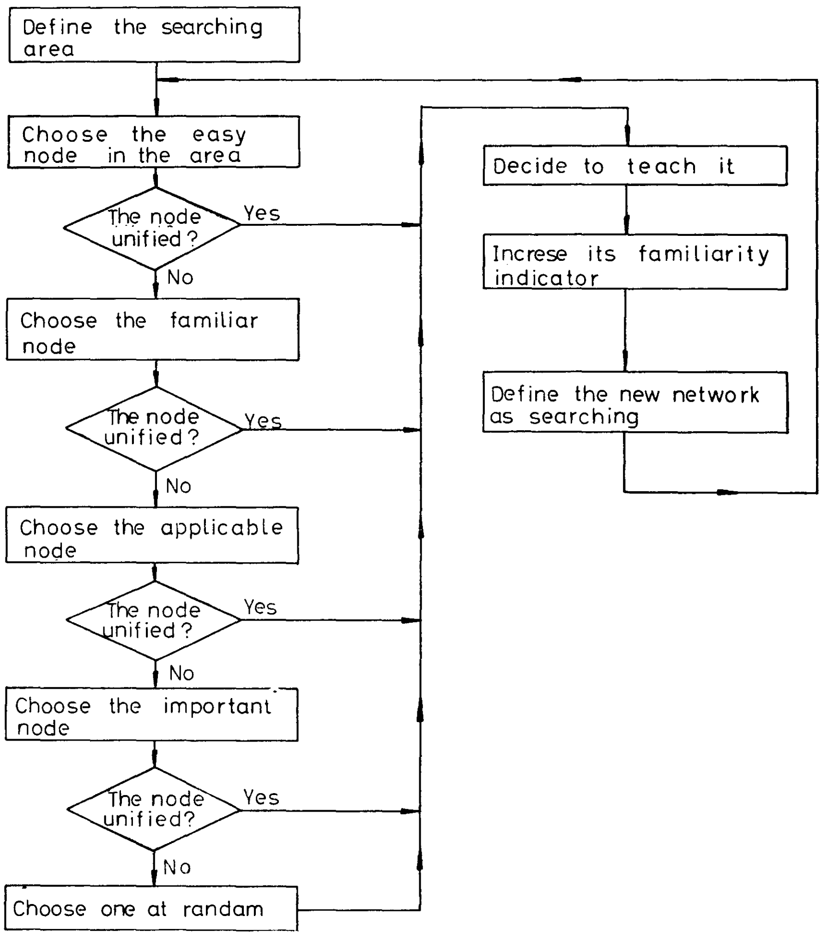


Fig. 10. A flow chart for deciding the optimal sequence of nodes by the decision model from teacher's demands.

in A5. So, we must choose A6.

And so on·····

In this way, we can get the following optimal teaching sequence of nodes. (Table 2)

Table 2. The optimal sequence from teacher's demands model.

Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Node	B1*	B2	A6	A5*	A4	A3*	A2	B3	A8	A9*	A7	A1	C1*	G
Importance	9	7	5	5	4	4	3	2	2	2	1	1	1	0
Familiarity	1	2	5	1	12	1	13	14	2	1	24	20	1	39
Input number	0	1	2	0	4	0	6	7	1	0	4	8	0	13

* A beginning node.

Average Familiarity $= \frac{136}{14} = 9.70.$

Decision model from the student's demands

Now, let us examine another decision model principally from the student's demands. In this case we must apply the four criteria by the following priority order. (Fig. 11)

1. Easiness to be learned
2. Familiarity
3. Applicability
4. Importance

Then, we can get the following table which shows the optimal sequence of nodes and also shows some other characteristics of this teaching sequence. (Table 3)

Table 3. The optimal sequence from the student's demands.

Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Node	B1	A8	A9	B2	A7	A6	A5	A4	A3	A2	B3	A1	C1	Goal
Importance	9	2	2	7	1	5	5	4	4	3	2	1	1	0
Familiarity	1	2	1	2	12	2	1	15	1	19	20	28	1	33
Input-number	0	1	0	1	1	2	0	4	0	6	7	8	0	13

Average familiarity $= \frac{138}{14} = 9.86.$

Comparison and evaluation of the two decision models

From the above tables, Tables 2 and 3, we can induce various characteristics. For example, an input number may indicate the complexity of an item. Importance may indicate the basic fundamentarity of an item. Familiarity may indicate the expected students' motivational strength of each node.

Then, one of the criteria that may indicate the validity of a program is the

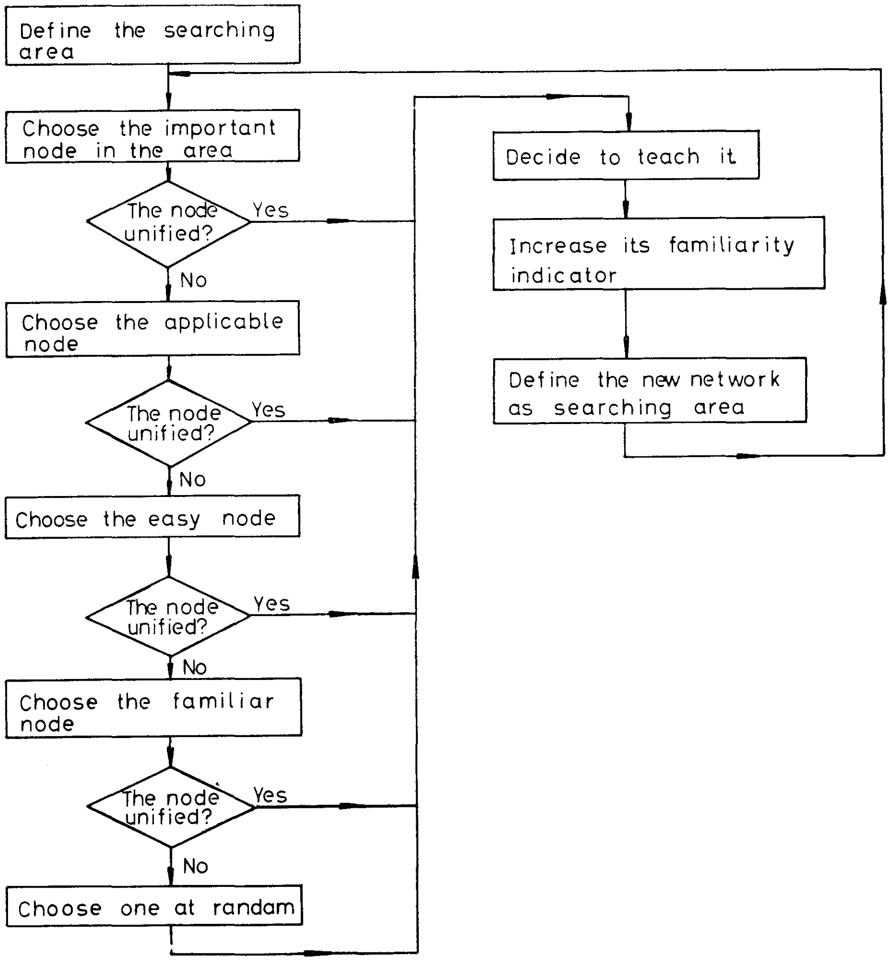


Fig. 11. A flow chart for deciding the optimal sequence of nodes by the decision model from the student's demands.

average familiarity (A.F.). If a program has high value of A.F., it supposedly has strong chains of association among nodes on the whole. We may judge from two A.F.s (9.70, 9.86) that the program from the teacher's demands is slightly more valid than that from the student's demands.

The following graphs, (Fig. 12, Fig. 13) show the relation between familiarity and input-number. Ideally speaking, familiarity should be lineary increased with input-numbers, since complex concepts should be learned with more strong association chains. Judging from this, we may again expect that the program from the students' demands is more valid than that of the teacher's demands.

Then, we may justly say that the decision model derived from the student's demands may be more suitable for students to learn.

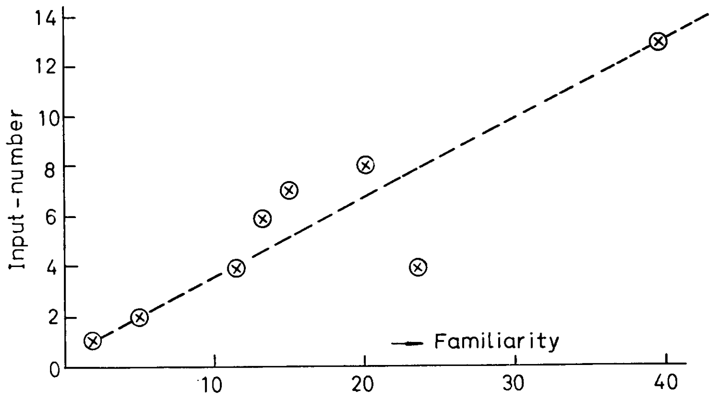


Fig. 12. The teacher's demands model

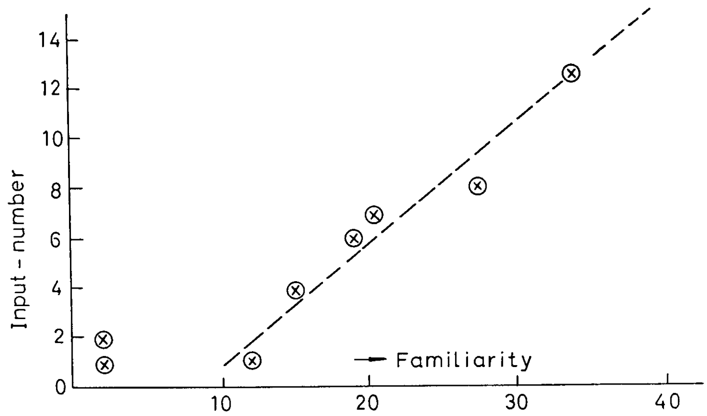


Fig. 13. The student's demands model

Consequently, we may get the following optimal sequence which has been proved to be most adaptive.

1. To be kind to someone, or a kind act for someone, is:
   1′) to relieve someone from his difficulty or unpleasant situation,
   2′) to give him a pleasant feeling.
2. If you should be kind to someone you should give him a pleasant feeling as much as possible.
3. A smile usually gives people a pleasant feeling.
4. Be kind to all people around you as much as possible.
5. You should smile when you are doing a kind act to someone, if you should be kind to someone.
6. If you relieve a person from his pain or unpleasant situation you may well be said to be kind.
7. If a person who finds it most unpleasant to keep standing in a street-car or in

a bus, is offered a seat by chance, his pain or unpleasant feeling will be much relieved.

8. If you ought to be kind to someone, while having a seat in a street-car or in a bus, you can be kind by offering your seat to the one who finds it most unpleasant to keep standing in a street-car or in a bus.

9. An old man or woman, or a mother carrying a baby in her arms is apt to be tired and feel it difficult to keep standing in a street-car or in a bus.

10. If you can offer your seat to someone, it is your duty to offer it to an old man or woman, or a mother carrying a baby in her arms, if one of them is standing nearby.

11. By the same reason, if you find a person physically handicapped standing in a street-car or in a bus, while you have a seat, you should offer him your seat, since a physically handicapped person also finds it difficult to keep standing in a moving vehicle.

12. If you have a seat in a street-car or in a bus, and if you are healthy enough, you should look to see whether an old man or woman, or a mother carrying a baby in her arms, or someone under a similar situation, is standing nearby, since if you find one of them, you should offer your seat to him or her.

13. How to offer your seat to someone. (Refer to the usual manners.)

14. The goal attained.

V. Adaptability to individual differences

Before applying this program to students who may have initial knowledge individually different, we must give them a pre-test.

Items which are to be tested in a pre-test should be chosen from items of large output-numbers, since a node which has a large output-number may be justly said to be greatly effective in the course of the program. In our example, we must include the following items in the pre-test.

1. B2 (output-number is 7)
2. A6 (5)
3. A5 (5)
4. A4 (4)
5. A3 (4)

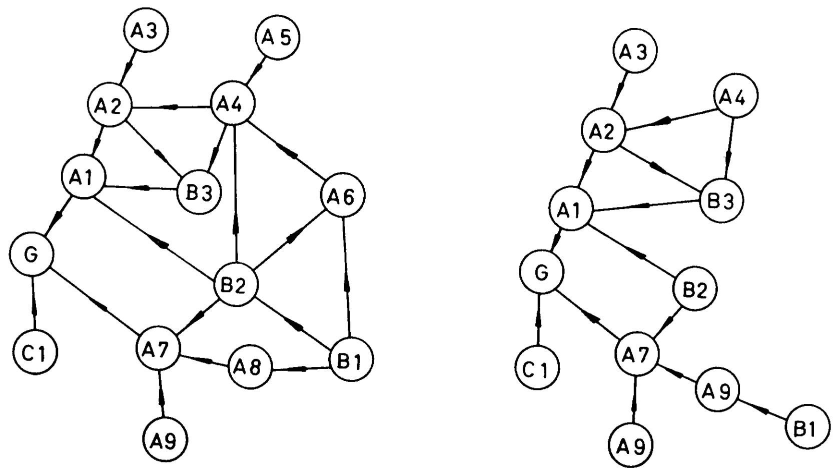
Besides, if a student has no knowledge on beginning nodes, he should not be applied to this program, since a beginning node is the one assumed to have been mastered before the student is introduced to the program. So, in our example here, we must add the following beginning nodes to the pre-test and select the proper students qualified for this program.

6. B1
7. A5
8. A9
9. C1

If a student cannot pass the pre-test of beginning nodes, he must get some other instructions or programs before being introduced to this program. He must not branch off in the course of this program, since branching off may destroy the logical and mental context of this program.

If a student can pass the other pre-test of the nodes which have large output-numbers, we must provide him with a program which has been slightly changed by the following procedure.

For example, if a student has already mastered the items, B2, A6, A4, A5, the network for him must be changed by disconnecting input-lines of these nodes. (Fig. 14). Then we must apply the decision model to the changed network to produce an optimal sequence of his own.



Original network

Changed network for a student who has mastered B2, A6, A4 and A5.

Fig. 14.

This procedure of changing the network may easily be done by computer equipment. So, if computer equipment is available, we can provide the optimal teaching sequences for any individual student.

VI. Conclusion

Summarizing this network theory, *i.e.* Network Analysis of Teaching, we may point out the following two aspects.

1) A methodological aspect

NAT may clarify the whole structure of educational processes which have hitherto been rather vaguely defined. We can see the operational meaning of what teaching is, and find pattern models for teachers' criteria in deciding the optimal teaching course. Moreover, we can clarify precisely what the most important concept is in a given subject, and find out exhaustively all necessary and sufficient concepts for it by analyzing it by the three Systems, A, B and C. Then, we may construct a network connecting these items or nodes.

2) A practical aspect

If we can get a whole network of a given subject and choose a proper decision model producing the optimal sequence of nodes, *i.e.* items, we can teach the subject to students of any level. For example, a student may have already mastered some of the nodes in a network. In this case, we have only to regard them as learned-nodes, disconnect them from their input-nodes, and then apply the criteria to the network. This can be done quite automatically by a computer. For this purpose, students should take a pre-test before learning, putting their results into a computer which will produce the optimal sequences of learning for each student.

It may happen that another node of higher concept must be added to a given network, or another network must be connected to it. We can also easily do this by only examining their input-nodes.

We often meet the case where we need logical training rather than algorithmic training. In such a case, we may apply the first criterion of importance to System A only, rather than the whole network. According to various teaching situations, we can freely choose the initial area to which we should apply several criteria to find out the most valuable node.

It may be obvious that NAT is useful in some sense, and that it raises important suggestions concerning the theories of educational processes. However, there are many problems which should be clarified in the further research.

Some of them may be as follows:

1′) Verification of the decision model.

2′) Establishment of the proper evaluation measures.

3′) Necessity for developing computer techniques.

4′) Construction of the more adaptive pattern models.

1′) Verification of the decision model.

We have adopted the decision model from the teacher's demands as the most valid model. However, we must examine whether it can be suitable for any kind of subject. We can construct various other models to decide the optimal sequence only if a proper measure for the evaluation of a model has been established.

2′) Establishment of the proper evaluation measure

We have evaluated a program by A.F. (Average Familiarity) and by some other graphs. However, it seems to us that we can construct the more proper measures than these. For this purpose, we must examine further what kind of criteria is valid when we justly evaluate the excellence of the sequence of a program. It is true that we cannot verify the validity without a number of experiments. However, we may need further research for the characteristics of the criteria themselves. For this purpose, we must study further in mathematics and statistics, and also psychology and education.

3′) Necessity for developing computer techniques

NAT requires a computer which has a large storage size. For example, in the present stage, we cannot compute the optimal sequence if the number of items (or nodes) is larger than fifty, by TOSBAC-3400 (written in FORTRAN language). However, if we write originally a computer program by the machine language, and if we devise properly to use an external large storage equipment, and also if a a computer of a larger inner storage size is available, perhaps we can compute the optimal sequence for hundreds of items. So, we are now developing various computer techniques for the NAT System.

4′) Construction of the more adaptive pattern model

We have constructed the pattern models, importance, familiarity, applicabibity, easiness, teachability, etc. However, we shall be able to make more suitable pattern models in the future. For example, we may define a space (likely, a metric space,) whose coordinates are those of item characteristics, *e.g.* entropy, absolute familiarity, etc. These characteristics, must be defined operationally and quite independently from relative characteristics of students, subject materials, teachers, teaching methods, etc.

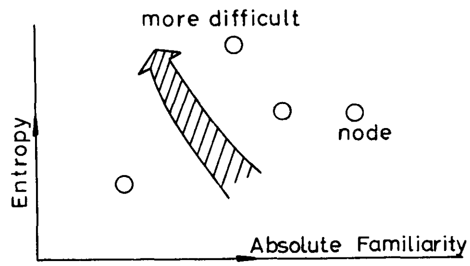
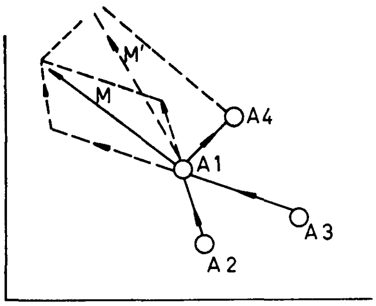


Fig. 15. A space where each node might be located.

Then we can locate all nodes of a network in this space. (Fig. 15). Then we may connect them according to their respective logical relations. If we could construct this space, we would define a motivational vector which indicates the direction and magnitude of student's motivation. For example, in Fig. 16, if a student has learned A1 whose input-nodes are A2 and A3, he may be expected to have a motivational vector (M), in which A2 and A3 are composed by the parallelogram law. Then, he may have an inertia or motivation in the direction of vector M. So, he may like to be taught in this direction.

We should have inquired much more before we could discuss these concepts clearly. At the present stage, we cannot deny that these are only conjectures. We need still deeper understanding of psychology and we must get the co-operation of various psychologists and educators for the true development of NAT.



M: The motivational vector expected in case that a student has learned A1.

M': The motivational vector expected in case that a student has learned A4 after A1.

Fig. 16.

Acknowledgement

We are greatly indebted to Professor M. Murai[8] for his valuable suggestions from his educational standpoint.

8) 村 井 実 Professor and President of the Center for the Science of Learning in Keio University.