Toward a Learning of Object Models Using Analogical Objects and Verbal Instruction

Norihiro Abe, Fumihide Itoh and Saburo Tsuji
Department of Control Engineering
Faculty of Engineering Science
Osaka University
Toyonaka Osaka JAPAN*

Abstract

In this paper an attempt of analogical learning method by verbalism is shown in order to create a model for an identification of unknown objects. When we expect a computer to recognize objects, the models of them must be given to it, however there are cases where some objects may not be matched to the models or there is no model with which object is compared. At that time, this system can augment or create new descriptions by making use of explicit verbal instructions.

1. Introduction

We have reported the story understanding system which uses both linguistic and pictorial information in order to resolve the meaning of given sentences and images[1]. By this research, we could have belief that correct meaning of given sentences is obtained if the relations among noun phrases, which correspond to objects in the images, consistent with the relations observed among objects in the picture.

The fact that this identification of objects and interpretation of the given sentences supplements each other simplifies both the detection of objects and disamibiguation of word sense or prepositional groups. In spite of these effects, this formalism has a defect that it requires additional knowledge sources from the system, which are the meaning of verbs and the model of objects that will appear in the images. All of models of objects or actors that are supposed to appear in the picture must be given to our system in order to achieve its purposes. But it is not easy for us to store all of such necessary models in the computer. If a person who does not know well about the details of this system wants to interact with this system, he will give up using this system, as he knows nothing of the representation of models in the computer. To make matters worse, there are quite many variations in real objects which we will encounter in the real world. For example,we can see various type of houses. In the traditional AI system, a generic model is utilized to identify such class of various objects. But it is not easy for such a system to discriminate idiosyncrasy of various objects. Fig.1 shows a part of sample story used to experiment its story understanding capability[1]. Even if the system is supposed to be given a generic model (for example, BOGLE) that represents both OBAQ and OJIRO, the system will not be able to discriminate them. The system needs some proper models for OBAQ and OJIRO. But if a new character which has some similar points to OBAQ and OJIRO apperes in the story, some modification to this BOGLE model is required. Even if we could give the system some suitable generic models, it is not easy to augment the description of generic models so as to represent all possible features that various objects in the class will have. Thus generalization process could not be accomplised in advance, but should be achieved through the experience.

In order to realize this type of learning, there are two important problems to be solved. First is an explanation capability. Unless a capability to convey one's obscure points to his partner is given to the system, it is difficult for the system to obtain good instructions from its partner. Although it is needless to say that a facility to interpret a natural language is required from the system, the facility to state a level of its understanding on objects is also inevitable. Concerning to this point we have already reported it in [2], then it will be abbreviated here.

Second is that from what kind of levels of knowledge state the system should start its learning process. Should an initial state of knowledge be given in forms of an inner representation or be explained in some natural language? We select the former approach by just the following reason. We think it quite difficult to give a clear view to unknown object without referring models. So we restrict a class of objects learned by our system to the group objects in which the system can obtain clear views concerning to their similarity through the comparison with similar examples.

But the assumption is not required that examples should be different in only one or two points at most from the unknown object. Many discrepancies between the object and its models are permitted to exist because such differences can be explained explicitly in the language by a teacher. And through a cognition of analogical or discrepant points of objects belonging to the same concept class, a generalization process is invoked that creates a common concept to them.

2. Concept of Analogy

When we think that there is an analogical point among things, we have already known the reason for an existance of the analogy. On some occasion, the analogy means a similarity in a shape or a color or a coincidence of a location, and in another case it implies a similarity in substructures of the objects. Throughout our life, we acquire a way how to find a similarity among many things. A matching mechanism we have uses some intrinsic attributes of objects when they are compared with those objects. Our pattern matcher does not examine descriptions of objects in a uniform way like the traditional one for the abstract learning, but it must properly change its estimation on similarity according to the objects.

Considering the pattern matcher like this, an another question arises that a representation of objects should be changed according to objects. It may be true that there are their own representations for each class of objects, say birds, vehicles or houses. In fact, some part of our knowledge can be described in forms of procedures, and others can be represented as in tables or graphs. Though we also believe that each class of objects should be described in their own representations, it is quite difficult to compare things which are described in

Fig.1 A portion of the story.

different ways. From the reason like this, we have determined a unique form of representation as reported in [1,2], which can be supposed to be applicable to all objects. By these settlements, it becomes possible for the system to make correspondences between the descriptions of objects and the expressions of sentence. Still more, it can be expected that a dialog between the system and T progresses smoothly and that the teacher can infere the method used in the interpretation of language and in that of the description of objects that the system owns.

3. Description of models

The model description in this paper is the same one shown in the paper[2], then we would like abbreviate details here. Fig. 2 shows a frame model of OBAQ, where the IMAGE slot needs an explanation for it is newly introduced slot in this paper. It has a pointer to an instance image of OBAQ, and by tracing this pointer the system can get a real image of OBAQ. This is necessary for compensating an insufficient parts of the representation of the system. A concrete example will be shown later.

4. Basic strategy for learning

Our learning method does not require a forced arrangement of samples, but starts its learning from seeing an example, however it needs the existence of frame representations of models which are used in comparison with an unkown object.

Then it tries to generate a model for the unknown object by referring to an analogical model and using a teacher's indication, and simultaneously it augments the concept trees of objects. At that time, the first key for a detection of analogy is assumed to be in contiguous relations between subparts and locations of subparts of objects. The mathematical models on analogy extraction utilize abstract relations between geometrical figures, however it is too abstract to obtain the same result as we shall reach . When we are told that a unknown object is similar to a certain object among various points of view we usually expect that many substructures having similar features will be found in the same location as the referred object. Of course, there are many examples that a resemblance in a location is not useful but prevents the program from achieving a correct detection of analogy. At that case, the teacher should explicitly tell the program to ignore that method and to use other methods such as similarity in relations or shapes or colors of objects. As we usually employ these variety way to detect analogy and record these experiences into our memory, we can easily decide what method should be used to compare things. We can not say that the program has learned something until these mechanisms recording a standards to compare things into memory are realized in the program.

5. Algorithm of learning

Let S(CI*) and S(CO*) be a group of parts whose RELATION is CIN and COUT, respectively. And S(C), S(C*), S(CIN), S(COUT) are defined as follows.

$S(C) = S(CI*) \cup S(CO*)$, $S(C*) = S(CI*) \cap S(CO*)$

$S(CIN) = S(CI*) - S(C*)$, $S(COUT) = S(CO*) - S(C*)$

Then a strategy for finding a candidate part of object part described by model is the following. Try the following procedure by setting S to S(COUT), S(CIN), S(C*), S(IN) in this order. Let a region including parts

```
OJIRO
 AKO        $VAL        BOGLE
 CLASS      $VAL        INSTANCE
 SUBPART    $VAL        J-BODY
 IMAGE      $VAL        OBAQ1
 SEX        $VAL        MAN
 REASON     $VAL        GIVEN

 J-BODY
 AKO        $VAL        BODY
 CLASS      $VAL        INSTANCE
 FIGURE     PART        OJIRO
            RELATION    IN
            POSITION    ((**) **)
 SHAPE      $VAL        REGION
 SUBPART    $VAL        (J-MOUTH J-EYE
                        J-HAIR J-HAND)
 COLOR      $VAL        WHITE

 J-MOUTH
 AKO        $VAL        MOUTH
 CLASS      $VAL        INSTANCE
 FIGURE     PART        J-BODY
            RELATION    IN
            POSITION    ((C) C)
 SHAPE      $VAL        REGION
 SUBPART    $VAL        J-LIP
 COLOR      $VAL        PINK

 J-HAIR
 AKO        $VAL        HAIR
 CLASS      $VAL        INSTANCE
 FIGURE     PART        J-BODY
            RELATION    COUT
            POSITION    ((C) U)
 SHAPE      $VAL         BRANCH
 SUBBRANCH  $VAL         (H1 NIL H2 NIL
                              H3 NIL)
 COLOR      $VAL         BLACK
 NUMBER     $VAL         THREE

 J-EYE
 AKO        $VAL        EYE
 CLASS      $VAL        INSTANCE
 FIGURE     PART        J-BODY
            RELATION    IN
            POSITION    ((*) U)
 SHAPE      $VAL        REGION
 SUBPART    $VAL        (J-R-EYE J-L-EYE)
 COLOR      $VAL        WHITE
 NUMBER     $VAL        TWO
 CONCEPT    $VAL        T
```

Fig.2 A frame of OJIRO copied from that of OBAQ.

in S be L. Then try (1) at first by finding out elements to which the case (1) is applicable. Next try (2) and then try (3), (4), (5) in this order in the same way to the case (1) (see Fig.3)

(1) one-to-one correspondence case: Unless this correspondence is denied by a teacher, it is accepted and delete x from S and M, delete y from O, where x and y is the part shown in Fig. 3, respectively and M, O means the set of model parts and object parts. If denied, this pair is recorded in NPL(Not Pair List), and put x to the last of S in order to test it again in (5).

(2) one-to-many or many-to-one case: Unless one reliable correspondence between x and y can be found, postpone the decision of x, and put it into PLIST.

(3) many-to-many case: By utilizing relational constrains on their locations among them, select a consistent combination of correspondences of them. If some are left unmatched, they must be put to the last of S.

(4) no correspondence: put x to the last of S.

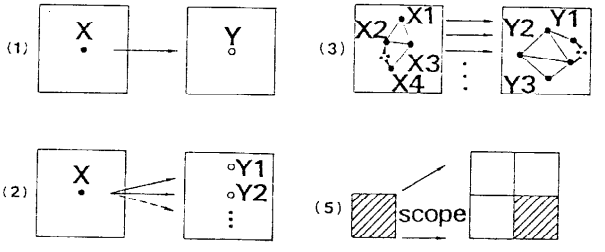

Fig.3 Correspondence between model and object.

(5) After all above cases are tested, extend a scope for the search of plausible parts into a region including current neighbour region. When no extension is allowed and PLIST ∩ M ≠ Ø then ask the teacher what each element in PLIST ∩ M is. In the above process, when some correspondences recorded in NPL are detected later, it is clear that they are not regarded as candidates. If some parts of the model or the object are left unknown even after these processes, system must ask them of the teacher.

6. Picture Processing

It is necessary to find regions in the given image which seem to correspond to the model parts. This is independently done from the identification process of parts; that is, for the given image an extraction process of lines and regions is conducted. At first, all line segments are found and next all regions enclosed by some of these line segments are detected. Then relations such as (C)IN/(C)OUT are computed for these regions. Using a result of this preprocessing of the image, a rough search of the object is tried in a top-down fashion which looks for a group of regions that satisfies the contiguous relation (that is CIN or COUT). Let call it as a rough model of the object. For example, the program will seek for the three regions, because a top level contiguous relations holds among Q-BODY, Q-L-HAND and Q-R-HAND. If this method fails to find a candidate for the object to be identified, the object detection will be attempted by removing one of the contiguous regions to the main part of model, say the Q-L-HAND.

7. Dialog

```
C: What is ojiro?
T: OJIRO IS OBAQ'S BROTHER.
C: Does ojiro resemble obaq?
T: YES. BUT OJIRO HAS JUST ONE HAIR. (a)
C: A color of his body is not white.
It is yellow. Is it true ? (note that
a color of OJIRO'S BODY is yellow) (b)
T: YES.
C: I have found his body.  I have found
his hair. I have found his left hand.
I have found his right hand.          (c)
I think I could find his mouth, but it
is different from his model.
T: WHAT ARE YOU DOING?
C: I have found a red branch to the left
of his right hand as a candidate of his
mouth. But it is different from his model.
Do you need another explanation of the
location of his mouth ?               (d)
T: YES.
C: Do you want an explanation in absolute
coordinate ?
T: NO.
C: It is to the right of his left hand in
the image. Do you want another expla-
nation in absolute coordinate?
T: YES.
C: It is in the center of the middle part
of his body in the image.
T: EXPLAIN THE DIFFERENCE BETWEEN THE
MODEL AND OBJECT.
C: On what part do you want an explanation?
T: MOUTH.
C: The mouth of the model is different
from that of the model in the points of
shape, subpart and color.
Do you need more explanation?
T: YES.
C: The shape of the model is a pink
region, and that of the object is a red
branch. The mouth of the model has a lip,
but that of the object has not subpart.
T: OJIRO'S MOUTH IS A RED LINE.       (e)
C: I have found his mouth.  I think I could
find his right eye, but it is different from
his model.
T: WHAT ARE YOU DOING?
C: I have found a black region in the
center of the middle part of his body as
a candidate of his right eye.  But it is
different from his model.
T: EXPLAIN OJIRO'S RIGHT EYE.
C: Which do you mean the model or object?
T: MODEL.
C: It is a white region which is above
his mouth in the model. Do you need ano-
ther explanation of his mouth?
T: NO.
T: DOES IT HAVE SUBPARTS?
C: Yes, his right eye has a pupil in the
model.
T: OJIRO'S RIGHT EYE IS A BLACK REGION.
C: I have found his right eye. I have found
his left eye. Finally, I have found ojiro.
T: EXPLAIN OJIRO.
C: He is a kind of bogle.
T: EXPLAIN OJIRO'S STRUCTURE.
C: Ojiro is a yellow region and his body
has a hair, a mouth, eyes and hands.
```

8. Examples

We would like to explaine how the dialog shown in the above has been obtained. By (a) the program copies OBAQ model and constructs a frame tree consisting from a BOGLE frame and the OBAQ frame is stored as an instance of the BOGLE frame. But it records in its STM that there is just one hair as the teacher said so. In this case, a serious problem will occurs if C does not know what a hair is, however C can recognize what a hair means, as it is stored as a part of OBAQ (At present, a verbal definition of a new object is not considered). Next, C looks for a candidate region of OJIRO using the copied model. As mentioned previously, it tries to find the rough model from the second frame of Fig.1. In this frame, OBAQ, TABLE, APPLE, CLOCK and OJIRO are drawn, but as the first three objects have been found in the first frame, C tries to find them before a detection of OJIRO. Then there is a possibility that C will looking for CLOCK and OJIRO as a candidate of OJIRO. In this case, CLOCK cannot clearly be matched to the rough model, therefore C succeeds in the detection of plausible regions of OJIRO. But regrettably a color of the region (yellow) which seems to be OJIRO'S body(J-BODY) being not different from that of the model(white), C cannot believe its tentative conclusion. This causes a complain shown in (b) and by accepting T's agreement C can believe its correctness and T can also think c to be in a right state. (Here, there is another problem about how a contradiction should be resolved when T's belief does not agree with that of C.) Consequently, C changes value of COLOR in J-BODY into YELLOW.

Next, C tries a verification of J-HAIR which is the first member of S(COUT), where S(COUT) = (J-HAIR, J-HAND)

As C can be aware of the fact that J-HAIR is a hair by its AKO slot and that there is a note on the hair in STM, it can know that OJIRO'S hair cannot be recognized only by referring to the copied model. Since the just one alteration in the number of hairs is recorded there, C thinks their location to be same as the model specification. As J-HAIR has one-to-one correspondence with J1 (see Fig. 4), the system believe this one as far as the teacher does not deny it, which can be found in the ((C)U) part of J-BODY. It ends the verification of J-HAIR by storing (J1 NIL) into SUBB slot. In a similar way to this, C begins to identify J-HAND, however C can be aware of that it should look for J-R-HAND and J-L-HAND, as there is a CONCEPT slot in J-HAND, which signifies that this frame is used not to represent graphical relations but to represent conceptual relations between frames. So C succeeds in the identification of them beause of a perfect match in their locations, colors and

substructures. The result of this steps is reported in (c). Now there being no parts in the model which belong to S(COUT), elements in S(CIN), S(C*) must be checked but there is none in them. Consequently, the identification process proceeds to S(IN) and C starts a verification of J-MOUTH, where S(IN) = (J-EYE, J-MOUTH). But as there is no possibility for the case (1), J-MOUTH and P4, P5, J2 have one-to-many correspondences. At this step it is impossible for the system to decide which one has the best correpondence with J-MOUTH, we get PLIST = (J-MUOTH). For J-EYE, which implies that J-R-EYE and J-L-EYE, there is no candidate in ((L)U), ((R)U) of the object. Consequently the scope of the search must be extended to its neighbour region. This leads the search process to the step for finding J-R-EYE in that scope shown in Fig. 5. As relational constraint between J-R-EYE and J-MOUTH does not contradict with that between P4 and J2, J-R-EYE⟷P4, J-Mouth⟷J2 are obtained. Then other properties are tested for verification of its decision. But regrettably, discrepancies are found for both his mouth and eyes. The candidate for his mouth is a line segment, whereas the model says that it is a region and that it has a substructure. Similarly the candidate for his right eye is a black region, but its model description says that it is a white region with a substructure. C complains about their disagreements in the order of their discovery.

Therefore it at first complains of his mouth as shown in (d). Given teacher's instruction on a shape of mouth, C is convinced of his decision and add a new slot SUBBRANCH in place of SUBPART and records (J2 NIL) there becase it has found that his mouth is not a region but a line segment. Here instead of the instruction (e), teacher can say that C should believe the given image correct. In that case, C suppose its decision to be right and does the same thing as the above. The difference between these two cases is the latter has a high risk in the correctness of its conclusion.

Next, C complains about the discrepancies his eyes have. When this is resolved by a conversation, it is clear that J-L-EYE corresponds to P5. And now nothing is stated about his left-eye after an instruction on his right eye has been given to it, because they have the same properties concerning to both their models and object parts. In case where one of them is not same, a question is asked by C about that difference .

Let consider the reverse case: learning of OBAQ from the model of OJIRO. At first R1,R2 and R3 will become candidates for Q-BODY because inner regions are not considered as candidates. But it is easily seen that R2 is the best candidate for Q-BODY. Consequently Q-R-HAND⟷R1, Q-L-HAND⟷R2 will follow (⟷ means a correspondence). Next, K1, K2 and K3 will have candidacy for Q-HAIR but the system cannot decide which one is best, PLIST is set to (Q-HAIR). But this problem is not solved by extending the scope, therefore it must ask the teach what they are. With this question it knows that OBAQ has three hairs. Now parts corresponding to Q-R-EYE, Q-L-EYE and Q-MOUTH must be found in the object, however it is quite difficult to do that by the given description of OJIRO alone: for all of them locate in ((C)C) of J-BODY and only one part R8 locates in the corresponding ((C)C) region of

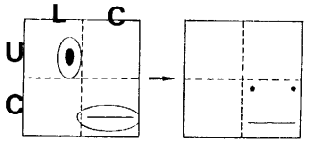

Fig.5 A scope for finding J-R-EYE, J-MOUTH.

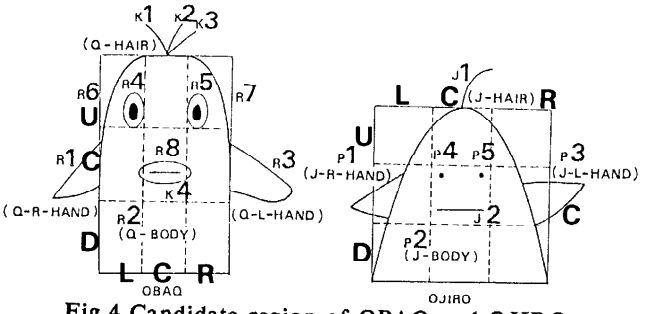

Fig.4 Candidate region of OBAQ and OJIRO.

R2. If R8 is supposed to be Q-R(L)-EYE, their color, substructure contradict each other. And it also difficult to regard R8 as Q-MOUTH for Q-MOUTH is a branch but R8 is a region. As a result of this uncertainty, the scope must be extended. Then a result shown in Fig. 6 is obtained, Q-R-EYE⟷R4, Q-MOUTH⟷R8 follow if relational constraints of them are known. Regrettably these are not computed from the descriptions recorded in the frame representation because the precise locational relation between Q-MOUTH and Q-R-EYE can not be obtained. But by tracing a pointer stored in IMAGE slot of the copied model of OBAQ, it is easy to get the locational relation between Q-R-EYE and Q-MOUTH and compute Q-R-EYE⟷R4, Q-MOUTH ⟷R8. Howevers their color and substructures do not coincide each other, they must be asked. At last a correspondence between Q-L-HAND and R6 is easily obtained.

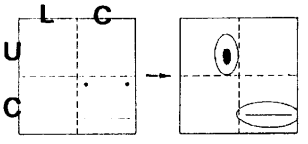

Fig.6 A scope for finding Q-R-EYE, Q-MOUTH.

Now we would like to give another example to show the current capability of this system: Example is given in Fig.7. Let suppose we are given a model for (A) and we must teach the system to model (B). Through the same process shown above, ROOF1⟷R2 and WALL1⟷R1 are obtained in this order. Next R6 is hypothesized to correspond to DOOR1 due to its CIN relation and its location. As this is denied by the teacher(it is clear that it is not a door), NPL is set to ((DOOR1 R6)). Then the scope for DOOR1 is extended and this time R4 and R6 become candidates for it. By referring to NPL, one-to-one correspondence between DOOR1 and R4 is obtained, however, their locations are different each other. This discrepancy must be resolved by instructions. As the consequence of this it can know that R4 is a door and that there is a possibilty the locational constraint on doors does not necessarily succeed, by comparing the location of R4 and DOOR1. This is recorded in its memory and it should be used later after the system has experienced more examples about a class of these objects. The same thing is done for windows because the position R5, R6 do not coincide with that of WINDOW1 or R3 (R3 is found in the region predicted by WINDOW1 but R5 and R6 cannot be recognized until they are told). If the teacher explicitly says that the

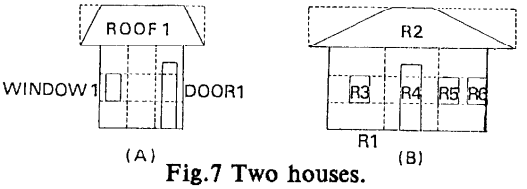

Fig.7 Two houses.

locational constraint will not be valid for identifying doors or windows, this fact must be recorded into their frame descriptions. But this has not been implemented in the present system.

As easily understood from the above example, the current representation has many weak points. One of them is that it cannot discriminate cases shown in Fig. 8. In this figure, P is contiguous to the side, while Q is contiguous to the bottom although their location is in ((L) D) of some part. If the system can discriminate them, it can easily infer that R4 will correspond to DOOR1 without regarding R6 as DOOR1.

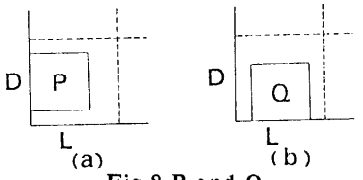

Fig.8 P and Q.

9. Use of Generic Frames

As mentioned in 8., OBAQ frame causes BOGLE frame to be generated as a generic one, and OJIRO frame is obtained through learning process. Naturally this generic frame should be able to represent all of its instances including of OJIRO frame, and common properties to them should be stored in suitable slots of a parent frame. At present our program just makes frame trees in which OJIRO and OBAQ frame are children of BOGLE.

A reason for this is that there is a danger of global rearrangements of frame trees. In the first example, we at first believe OBAQ frame to be an instance frame but it may turn out that it is not an instance when other examples of OBAQ not matched to his frame appear in image, because there are many variations in his shape as he can wink or move his eyes or open his mouth. After program have experienced these example, it should make a general concept of OBAQ and arrange frame trees by erasing unnecessary instances about him. It is not sufficient to record such possibilities in forms of AND or OR combinations of descriptors because arbitrary combinations of variations in its substructures are not necessarily allowed. The rearrangement of frame trees is a difficult problem, considering the possibility man is apt to fail in giving correct instructions.

Though there are some incomplete points in the construction of frame trees, program can use a portion of them to identify subparts of the object to be learned. For example, suppose that we would like to teach by referring to OJIRO a character Q-KO, who resembles to him very much except for her eyes and her eyes are rather similar to OBAQ's. In the course of identification of her, if OBAQ frame is not stored, program will complain about her eyes as well as in the learning of OJIRO from OBAQ. However it can use OBAQ's eyes in the recognition of her eyes by tracing its AKO link and finding OBAQ frame, after a failure in the matching of her eyes to OJIRO's. Of cause, it does not do that without teacher's permission, but will ask for his approval.

10. Conclusion

A new attempt toward a verbal modeling of objects has been shown in this paper, however there are many incomplete points concerning to the learning method taken in this research. The basic strategies for finding analogies among things have been given to the system as the known fact, but they should be obtained by itself in the course of learning which needs more examples than that experimented in this paper. To accomplish our purposes we must implement many programs including language system which understands ellipsis , anahora and gives us good explanations on the structure of objects and the reason why such structural descriptions have been obtained. And if a fatal error is detected after an acquisition of some models, it must be corrected by considering the history in the construction of models. This error correcting process has a close relation to frame trees.

References

1) N.Abe, I.Soga and S.Tsuji: A Plot Understanding System on Reference to both Image and Language, 7th-IJCAI, p.77 (1981)
2) N. Abe and S.Tsuji: A Learning of Object Structures by Verbalism, COLING-82, (1982)
3) P.H.Winston: Learning Structural Description from Examples, Ph.D.Th., MIT (1975)
4) P.H.Winston: Learning by Creating and Justifying Transfer Frames, Artif. Intell., 10,2, p.147 (1978)
5) P.H.Winston: Learning and Reasoning by Analogy, CACM, 23, 12, p.689 (1980)
6) T.G. Dietterich and R.S.Michalski: Inductive Learning of Structural Descriptions, Artificial Intelligence,16,p.257(1981)
7) S.A.Vere: Inductive Learning of Relational Productions, Pattern- Directed Inference Systems (Academic Press, 1978)