COLING 82, J. Horecký (ed.)
North-Holland Publishing Company

A learning of object structures by verbalism

Norihiro Abe, Saburo Tsuji
Faculty of Engineering Science
Osaka University
Toyonaka Osaka JAPAN

In this paper an attempt of learning by verbalism is shown in order to create the models 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 being explicitly taught using 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. In this research, we have believe that a correct meaning of the 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 images.

The fact that this identification of objects and the 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, 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 models in a computer. If a person who does not know well about the details of this system wants to interact with it, he will give up to use the 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 objects. But it is not easy for such a system to discriminate idiosyncrasy of varous objects. Fig.1 shows a part of sample story used to experiment its story understanding capability. 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 model for OBAQ and OJIRO. But if a new character which has some similar points to OBAQ and OJIRO apperes in the story, some modifications to the BOGLE model are required. Thus generalization process could not be acomplised in advance, but should be achieved through experiance.

When we are asked to do some task, we are usually given informations concerning to the objects of that task and their processing method. In case where we encounter some unkown objects in the course of the task, we can construct a more generic model including them together with a creation of instance models for those individuals by demanding an explanation to a person who knows well about those objects. In this real situation, it cannot be expected that a learning process proceeds successfully like the experiment studied by Winston, as the assumption fails of success that the samples can be arranged conveniently for the learning. We usually augment our knowledge by explicitly being taught about missing or insufficient parts of the known models.

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.

Second is a point 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 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 to models. So we restrict a class of objects learned by our system to the group of objects of which the system can obtain clear views concerning to their conditions through the comparison with their similar example.

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 explicity in the language by a teacher. And through a cognition of analogical or discrepant points of objects belonging to the same conceptual class, a generalization process is invoked that creates a common concept to them.

2. Description for Object

The model description used in this paper is the same one shown in the paper[1] except for the usage of the frame representation to describe relations among subparts of the model. Let explan using an example. Fig.2 shows the OBAQ, who is an actor of the sample story shown in Fig.1. To describe location of subparts of this model, its main part is enclosed by a rectangle as shown in Fig.2. Then this rectangle is devided into 9 subregions and the location of its subparts is described in terms of these subregions. When some of these subparts has also subparts, they are hierarchically described in the similar way. And the relations between these subparts is represented using the frame structures. The frame structures corresponding to the OBAQ model is given in Fig.3 (this figure shows a hypothetical model of OJIRO obtained from the copy of OBAQ frame.)

3. Frame Representation

The slot **AKO** means a well-known relation A-KIND-OF, and the **CLASS** indicates whether the frame is gneric or instance frame. If the frame is generic, then it has two slot, **GENE** recording its lower class of generic frames and **INST** recording its instance frames. The **FIG** slot represents a pictorial reration to its parent frame. This slot means that the part corresponding to this frame is a subpart of the frame stored in the **PART** and that it can be found by looking for the region designated in **POS**. And the facet **DIR** describes a relation which this part has to its parent. There are three relations concerning to the **DIR** as shown in Fig.2 and concering to the **POS**, many combinations of subregions are permitted which can be expressed with the symbols, L,C,R and U,C,D. Especially the symbols *, ** are used to designate the locations shown in Fig.4. The slot **SHAP** represent whether the part corresponding to this frame is a region(**REG**) or a branch(**BRA**). The **SUBP** slot records its subparts and their locations of or relations to this part are described in three facets as shown above. Especially when the **SHAP** condition is **BRA**, this frame has a **SUBB** slot instead of

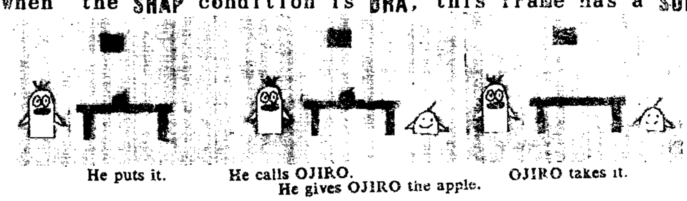

Fig 1.
A sample story

SUBP slot and a branch structure is recorded here. An example of a branch is shown in Fig.5. The **COL** slot records a color of this part and a slot **CONCEPT** means that this frame is prepared for the conceptual consistency of frames and not for pictorial relation. In addition to these slots, there are several slots, **WAKE**, **SEX**, **NUM** and so on. These are prepaired to generate a sentence for stating a reason why this frame is required or an explanation about why discrepancies between an object and its model can be found out in its matching process.

4. Basic Strategy of Learning

The system tries to generate a model for the unknown object by referring to an analogical model and using a teacher's instruction, 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 locations of subparts of objects. When we are told that an 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 refered object. Of course, there are many examples that resemblance in a location is not useful but prevents the program from achieving a correct detection of analogy. At that case, the teacher should explicilty tell the program to ignore

```
OJIRO                                    J-BODY
  AKO       $VAL   BOGLE                   AKO     $VAL   BODY
  CLASS     $VAL   INSTANCE                CLASS   $VAL   INSTANCE
  SUBP      $VAL   J-BODY                  FIC     PART   OJIRO
  WAKE      $VAL   GIVEN                           DIR    IN
  SEX       $VAL   MAN                             POS    ((**) **)
                                           SHAP    $VAL   REC
J-HAND                                     SUBP    $VAL   (J-MOUTH J-EYE J-HAIR J-HAND)
  AKO       $VAL   HAND                    COL     $VAL   WHITE
  CLASS     $VAL   INSTANCE
  FIG       PART   J-BODY                J-HAIR
            DIR    (OR (COUT) (CIN))       AKO     $VAL   HAIR
            POS    ((*) C)                 CLASS   $VAL   INSTANCE
  SHAP      $VAL   REC                     FIC     PART   J-BODY
  SUBP      $VAL   (J-R-HAND J-L-HAND)             DIR    COUT
  COL       $VAL   WHITE                           POS    ((C) U)
  NUMB      $VAL   TWO                     SHAP    $VAL   BRA
  CONCEPT   $VAL   T                       SUBB    $VAL   (L1 NIL L2 NIL L3 NIL)
                                           COL     $VAL   BLACK
J-EYE                                      NUMB    $VAL   THREE
  AKO       $VAL   EYE
  CLASS     $VAL   INSTANCE              J-MOUTH
  FIG       PART   J-BODY                  AKO     $VAL   MOUTH
            DIR    IN                      CLASS   $VAL   INSTANCE
            POS    ((*) U)                 FIC     PART   J-BODY
  SHAP      $VAL   REC                             DIR    IN
  SUBP      $VAL   (J-R-EYE J+L-EYE)               POS    ((**) C)
  NUMB      $VAL   TWO                     SHAP    $VAL   REC
  COL       $VAL   WHITE                   SUBP    $VAL   J-LIP
  CONCEPT   $VAL   T                       COL     $VAL   PINK

J-LIP                       J-R-EYE                        J-R-HAND
  AKO    $VAL  LIP            AKO    $VAL  RICHT-EYE         AKO    $VAL  RIGHT-HAND
  CLASS  $VAL  INSTANCE       CLASS  $VAL  INSTANCE          CLASS  $VAL  INSTANCE
  FIC    PART  J-MOUTH        FIC    PART  J-BODY            FIG    PART  J-BODY
         DIR   IN                    DIR   IN                       DIR   COUT
         POS   ((**) C)              POS   ((L) U)                  POS   ((L) C)
  SHAP   $VAL  BRA            SHAP   $VAL  REC               SHAP   $VAL  REC
  SUBB   $VAL  (L4 NIL)       SUBP   $VAL  J-R-PUPIL         COL    $VAL  WHITE
  COL    $VAL  RED            COL    $VAL  WHITE.
```

Fig 3. Frame copied from OBAQ frame.

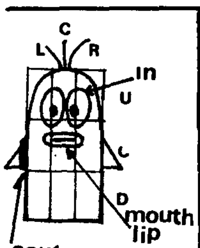

Fig 2. OBAQ

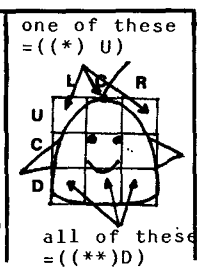

Fig 4. *, ** and OJIRO

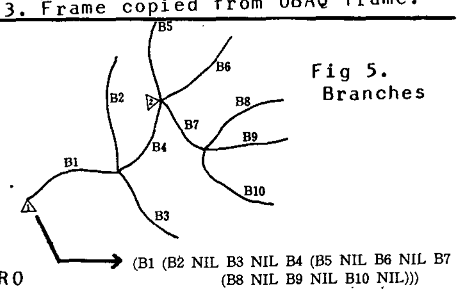

Fig 5. Branches

this standard 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 experinces 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 untill these mechanisms recording a standards to compar things into memory are realized in the program.

5. Scenario

C: What is ojiro?
T: OJIRO IS OBAQ'S BROTHER.
C: Does ojiro resemble obaq?
T: YES. BUT OJIRO HAS JUST ONE HAIR. (1)
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) (2)
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. But I cannot find his mouth. (3)
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 ? (4)
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 explanation in absolute coordinate? (5)
T: YES.
C: It is in the center of the middle part of his body in the image. (6)
T: EXPLAIN THE DIFFERENCE BETWEEN THE MODEL AND OBJECT.
C: On what part do you want an explanation?
T: MOUTH. (7)
C: The mouth of the model is different from that of the model in the points of location, shape, subpart and color. Do you need more explanation? (8)
T: YES.
C: The mouth is all the middle part of his body, and that of the object is in the center of the middle patr of his body. 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. (9)
C: I have found his mouth. But I cannot find his right-eye.
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. (10)
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 want another explanation?
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.

6. Example

We would like to explain how the dialog shown in 5. has been obtained. By (1) the program generate OIJRO model from OBAQ model shown in Fig.3 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. Next, C looks for a candidate region of OJIRO using the copied model. In the second frame, OBAQ, TABLE, APPLE, CLOCK and OJIRO are drawn, but as the first four objects have been appeared in the first frame, in this case C can find OJIRO by looking for new objects. But regrettably a color of the region (yellow) which seems to be OJ-IRO'S body(J-BODY) being different from that of the model(white), this cause a complaint shown in (2) and by accepting a T's agreement C can believe its correctness and T can also think C in a right state. Consequently, C changes value of **COL** 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, and can find a line in the ((C)U) part of J-BODY. It ends the verification of J-HAIR by storing (H_1 NIL). into **SUBB** slot in place of (L_1 NIL L_2 NIL L_3 NIL). 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. So C succeeds in the identification of them because of a perfect match in their locations, colors and substructures.

The result of this steps is reported in (3). Next, the identification process proceeds to S_{IN} and C starts a verification of J-MOUTH, where S_{IN}={J-MOUTH, J-EYE}. As the locational constraint for this part is ((**)C), which means that it occupies ((L)C), ((C),C) and ((R),C) of J-BODY, the check is attempted whether just one candidate can be found for each of these 3 subregions. In this case, nothing is found for ((L)C) and ((R)C) but several parts are found in ((C)C) of J-BODY. So this process is suspended and identification of other parts (J-R-EYE and J-L-EYE) is attempted, but the same ambiguity as the above occurs and this causes the identification steps to be suspended. Consequently, for each one of these 3 parts, their results are just same each other; there are 3 parts in the ((C)C) of J-BODY and they are candiades for J-MOUTH, J-R-EYE and J-L-EYE. Then C avails of the relational constraint on locations of them in order to clarify their correspondences as far as possible. It infers that J-MOUTH probably locates in a lower position than J-EYE, because the location of J-MOUTH is ((**)C) and that of J-R-EYE and J-L-EYE is ((L)U) and ((R)U) respectively (in this example note that the location of J-EYE, ((**)U) can be also available). And it is also decidable if which black region corresponds to J-L(R)-EYE using the relation between ((L)U) and ((R)U). By this assumption on availabilty of the relational constraints, C can discover one possible correspondence between the model and object. Then other properties are tested. 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. Similary the candidate for his left(right) eye is a black region,but its model description is that it is a white region with a substructure. At the present state of program, any estimation on which is more plausible is not realized regarding to the accordance of these properties, C simply complains about their disagreements in the order of their discovery.

Therefore it at first complains of his mouth as shown in (4). Given teacher's instruction on a shape of mouth, C is convinced of his decision and add a new slot **SUBB** in place of **SUBP** and records (H_2 NIL) into it becase it has found that his mouth is not a region but a line segment. Here instead of the instruction (9), T can say that C should be 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 that the latter has a high risk in the correctness of its conclusion.

Next, C complains about the discrepancies of his eyes. Note here that nothing is stated about his left-eye once an instruction on his right-eye is 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 about the difference by C.

7. Use of Generic Frames

As mentioned in 4., OBAQ frame causes BOGLE frame to be generated as a generic one, and OJIRO frame is obtained through learning process. At present our program just makes frame trees in which OJIRO and OBAQ frame are child of BOGLE.

A reason for this is partly due to a lack of condsideration how simple pictorial descriptions can be compiled from various types of deviations in slot values. An another reason is that there is a danger of partial rearrangements of frames trees. In the 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 not matched to this frame appears in image, because there are many varietions in his shape as he can wink or move his eyes or open his mouth. After program have experinced these example, it should make a general concept of OBAQ and arrange frame trees by erasing unnecessary instances about him.

As a more important problem, strategies to discover cues for finding analogy between subparts must be stored in some slots of their model; that is, the locational constraint is a useful cue for human, animals and so on, but is not adequate for doors and windors of houses.

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 a character Q-KO by referring to OJIRO. Let suppose that Q-KO resembles to him very much except for her eyes but that they 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.

8. Explanation Capability

It is necessary for teacher to be given sufficient explanations about the level of knowledge the computer has attained. Unless the computer can tell him what it is looking for, what it has already found, what sort of descrepancies it suffers from, he cannot give proper instructions leading the computer to a satisfastory state. There are many sentence generating and explanation systems, however an explanation system like this research has not been investigated in the point that our system tries to give its partner an explanation or pictorial features of objects to be modeled by translating sentences not from the case frame of sentences but from frames corresponding to the pictorial models. Naturally such an explanation is on locations, shapes, colors and relations that models or objects have, and must be given in the forms that the partner can easily understand what the system knows. For this purpose, the explanation on locations is first attempted using the referred things in the dialog, and is finally given in an absolude coordinate based on the 9 subregions if there is no reference or the reference stack becomes empty. (4),(5),(6) in the Scenario shows this mechanism. The next important thing is that a partner may expect a detail explanation for something, but expect just a simple one for others. Regrettably the present sate of our program cannot detect his demand like this or resolve ambiguous points of his question, then it must ask him about his require as shown in (7). In this case, there are also many things to be explained, however the points are only stated by the program and the detail explanation is left to the partner as in (8). We believe this method proper because of easiness of explanations.

The comparison between things are listed in above of (9) in order to clarify their differencies. If more detail on the lip is needed, the partner can ask the system about it. On account of limited space, though we cannot state a sufficient discussion, there are many problems to be improved on how the system should grasp partner's intention or requirments. They must be solved for giving simple explanation to the partner.

Reference

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)P.H.Winston: Learning Structual Description from Examples, Ph.D.Th., MIT (1975)

3)P.H.Winston: Learning by Creating and Justifying Transfer Frames, Artif. Intell., 10,2, p.147 (1978)

4)P.H.Winston: Learning and Reasoning by Analogy, CACM, 23, 12, p.689 (1980)

5)J.W.Weiner: BLAH, A System which Explains its Reasoning, Artif. Intell., 15, 1, p.19 (1980)