PATTERN CLUSTERING: AN ARTIFICIAL INTELLIGENCE APPROACH

B. Shekar, M. Narasimha Murty and G. Krishna
Department of Computer Science and Automation
Indian Institute of Science
Bangalore 560 012 INDIA

ABSTRACT

The notion of *concept* based on functionality of objects is defined and made use of in the context of pattern clustering. An approach to partition objects using a knowledge base is presented. A different class of concepts called conceptual transformers is proposed and its effects on clustering is looked into.

I INTRODUCTION

Clustering is a process of grouping objects based on a similarity measure. Conventional similarity measures are context-free and gestalf property-independent in nature. A recent development to overcome such problems, is the measure (Michalski 1980): Similarity(A,B)=f(A,B,E,C) where A and B are the objects being compared and E and C are the environment and a set of predefined concepts respectively.

Here we define the term concept from the functional point of view based on the physical properties which imply the function. A knowledge-based approach to partition objects is outlined. An interesting subset of concepts called conceptual transformers, which is an abstraction from real life, is studied in detail. The effects of this type of concepts on the clustering mechanism is looked into.

II CONCEPTS AND CONCEPTUAL TRANSFORMERS

A concept can be recursively defined as follows:(a) A concept is a partial mapping from, either the powerset of the instances of the physical descriptors of objects, or the powerset of user-defined functions, to a domain of labels, (b) A concept is a partial mapping from the powerset of labels to a set of labels.

Observe that the above definition is an extension of "concept' given by Winston (Winston 1984). In the context of clustering, there is a certain category of concepts which cannot be treated the same way in which we view a concept as defined above. This class (Co) of concepts, which in addition to satisfying the above definition will also satisfy the transformer relation R on Co defined by the following first order logic statement:

$$(\exists Pc)\ (\exists T)\ [\ (Pc=0) \wedge (T \in Co) \wedge (\langle Pc, T, t\rangle R \langle Pc', T\rangle)\ ]$$

where Pc is a set of pre-conditions which should be satisfied by a non-empty cluster consisting of one or more objects,'t' is the time required to instantiate the physical descriptors of the object/objects corresponding to the cluster satisfying Pc using the concept T. Pc' is the set of instantiated physical descriptors.

Observe that:(i) The concept T does not directly participate in the concept implied by the cluster, whose physical descriptors have been instantiated by T. (ii) The concept T which is a conceptual transformer with respect to Pc and Pc' does not prevent it from being an ordinary concept in another context. (iii) T will not be a conceptual transformer if the cluster does not satisfy the pre-condition Pc. (iv) R only implies the possibility of Pc' but does not say that instantiation will occur, (v) The usage of the transformer is controlled by the frequency of usage of the respective concept which the transformer has instantiated - a practical aspect. (vi) A transformer may be used to improve the quality of a concept or reinstantiate an aberrated concept. (vii) A concept need not necessarily be an element of a single transformation relation. It can also be an element of the composition of several such relations.

Consider a kitchen knife whose cutting edge has become blunt. A grinder which has the ability to sharpen the blunt edge assumes the role of a conceptual transformer relation, when over a finite period of time, it transforms the blunt edge into a sharp edge. Note that, otherwise the knife should be in good shape, which is a part of the pre-condition Pc. Also observe that the grinder does not participate in the concept of cutting, though the grinder may participate in some other concept. Its frequency of usage is controlled by the deterioration in the sharpness of the edge of the knife. The grinder may be used to either improve the quality of the cutting edge or repair an aberrated edge.

A Cohesion forest is defined as a non-empty collection of N-trees which consists of concepts which have to be necessarily executed to achieve the root concept, and a collection of Q-trees which consists of concepts which improve the quality of the root concept. For a detailed explanation of Cohesion forest refer (Shekar,Murty and Krishna). Consider Figure 1. The branches CD, CE and CF which correspond to Marking, Cutting and Erasing respectively together represent the concept 'Writing'. In the case of Shaving, Looking and Lathering are quality improvement aspects, whereas Cutting is a necessary concept. Hence this is represented as a Q-tree given in Figure 2.

Consider an object, and the functions it is capable of performing. An object can be defined (from the functional point of view) by the instantiations of the physical descriptors necessary to describe the object. One or more of these physical attributes may be essential to perform the function/functions behind the creation of this object. Generally a large subset of the physical attributes may not be relevant to the function (henceforth referred to as primary function) of the object. Consequently, depending on the imagination of the user, these attributes can be made use of in performing certain other functions, (henceforth referred to as secondary functions) which need not necessarily be connected to the primary function/functions.

We illustrate the above with a kitchen knife. The primary function being cutting is the resultant of the edge being sharp, uniform and hard (generally metallic). Due to the usage of a heavy metal in the manufacture, the object posseses a physical attribute namely heaviness. This physical attribute can, and generally is made use of to support several secondary functions including its usage as a paper weight, a hammer (though with a reduction in the quality of hammering),etc.

A natural question that can be raised is the possibility of ternary, tertiary,., functions of an object. Reinstantiation of the values of certain attributes may not affect the primary function of the object under consideration. If we can identify a conceptual transformer which does this reinstantiation and the reinstantiated physcal attribute satisfies a concept, then this object can be said to satisfy a ternary function with a transformer relation. The same argument can be extended to define tertiary,... functions based on the complexity and usage of two or more transformers. Consider a kitchen knife. The tip of the knife can be transformed into a screwdriving edge with the help of a grinder. Thus, the primary function is cutting, secondary functions controlled by heaviness, and the ternary function is screwdriving with the help of a grinder.

I I IFUNCTIQNAL CLUSTERING

Eventhough conjunction can be effectively used in describing objects/classes (Michalski 1980), most of the applications do not warrant such complete descriptions, as functionality of objects plays a vital role in the environmental semantics. Thus, implication which seems to be a better choice for functional clustering of objects is used to generate descriptions like 'cluster => an expression of concepts' where 'cluster' is a group of objects and 'an expression of concepts' consists of concepts defined in the knowledge base connected by logical connectives "A' (AND) and/or ' ® ' (EXCLUSIVE OR). There is no need for having additional connectives in the descriptions of such functional clusters. Note that AND indicates that the concepts can be executed at the same time and EXCLUSIVE OR indicates that only one of the two concepts can be done at the same time but not both. Further details of this approach and an algorithm to perform this clustering is given in (Shekar,et al).

Let us examine the effects of a transformer relation on a set of clusters. The number of concepts at a higher level with respect to the Cohesion forest increases. This can be illustrated by considering the following clusters which have been generated with the Cohesion forest whose simplified form is given in Figures 1 and 2.

Clusterl={Brush,Soap,Water source,Knife,
Pencil,Eraser}
Descriptionl:Clusterl*> Shaving ® Writing
Cluster={Blunt knife}

The description of Cluster2 will contain concepts implied by the knife's secondary functions, which we ignore in the present context. The existence of a grinder in the role of a transformer will give rise to:

Clusterl'={Brush,Soap,Water source,Knife}
Descriptionl':Clusterl=> Shaving
Cluster2'=(Knife,Pencil,Eraser!
Description2':Cluster2'«> Writing

Note that Description2' is at a higher level than Description2, and Descriptionl' is at the same level as Descriptionl with respect to the Cohesion forest. These clusters are shown in Figure 3. Observe that the arrow has a label corresponding to the transformer and the cluster/clusters pointed to by the arrow are bounded by a broken line indicating the fact that these clusters can be executed only after some time specified by the transformer. We call such clusters as virtual clusters. Observe that virtual clusters can exist at

both ends of the arrow. This is shown in Figure 4. Note that if Tr and Tr', the transformers, are the same, then the tranformer may be viewed as a bidirectional conceptual transformer. A screwdriver is an example of a bidirectional transformer.

IV CONCLUSION

We have proposed a functional definition of 'Concept'. Two categories of concepts have been identified (pertinent to clustering): Concepts whose presence is permanently required for the execution of the corresponding function, and conceptual transformers, which do not directly participate in the function. A hierarchical decomposition of functions of objects with the help of transformers, has been looked into. Certain interesting properties of transformers and their effects on cluster configurations leading to virtual clusters have been explained with an example.

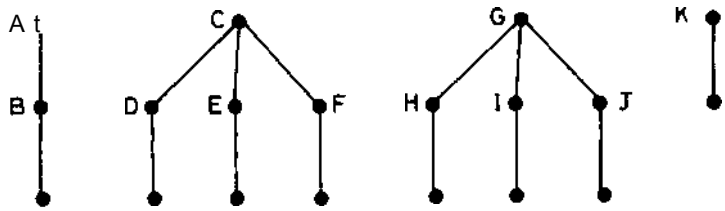
The approach suggested in this paper is different from conventional approaches, as it views objects from the functional angle. The resulting partition is more in the abstract rather than in the physical sense, and such a grouping resembles our mental perception of objects. It should be noted that the above approach is pertinent to man-made objects as they are created from a functional point of view.

REFERENCES

1 Michalski, R. S. "Knowledge Acquisition through Conceptual Clustering: A Theoretical Framework and an Algorithm for Partitioning Data into Conjunctive Concepts." International Journal of Policy Analysis and Information Systems 4:3 (1980) pp. 219

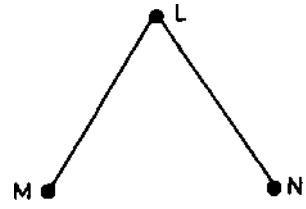
2 Shekar, B., M. N. Murty and G. Krishna "A Knowlege-based Clustering Scheme." Pattern—Recognition Letters (to appear).

3 Winston, P. H. Artificial Intelligence. Addison-Wesley, Reading, 1984.



A = Shaving
B = Cutting
C = Writing
D = Marking
E = Cutting
F = Erasing
G = Lathering
H = Brushing
I = Wetting
J = Soaping
K = Looking

Figure 1 N- trees included in the Cohesion forest



L = Shaving
M = Looking
N = Lathering

Figure 2 Q-trees included in the Cohesion forest

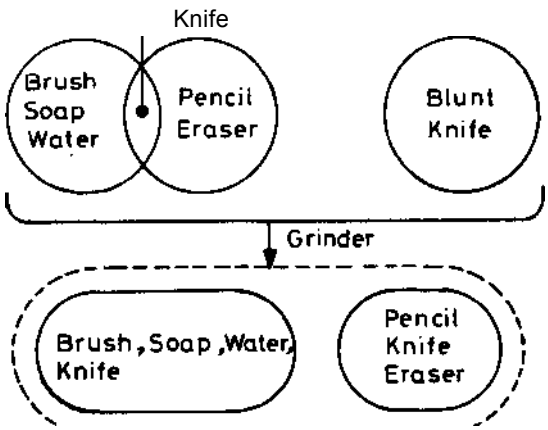


Figure 3

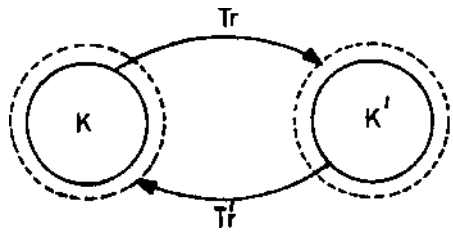


Figure A