

Chris Thornton

School of Cognitive Sciences
University of Sussex
Brighton, England

ABSTRACT

Bundy, Silver and Plummet (1985) provide an analysis of the Focussing algorithm and the Classification algorithm in the case where the description space consists of a set of relation trees. This paper discusses an extension to their analysis in which the description space is construed as a geometric space. Under this construal the behaviour of both the Focussing algorithm and the Classification algorithm is analysed in terms of the construction of hypercuboids. This analysis leads to a number of observations: (i) that a distinction can be made between a strong and a weak version of the disjunctive-concept problem; (ii) that certain solutions to the disjunctive-concept problem can be shown to exploit what are, in effect, distance functions over the description space and (iii). that the Classification algorithm is only capable of learning a subset of the possible disjunctive concepts in any given domain.

1. Introduction

In the interpretation of the Focussing algorithm presented by Bundy et al. (ibid.) the description space consists of a set of relation trees and execution of the algorithm involves the manipulation of upper markers (which collectively define the most general form of the concept being learned) and lower markers (which collectively define the most specific form of the concept being learned). The form of this marker-manipulation is reasonably simple: following the presentation of a new positive instance, one or more lower markers may be raised so that the most specific form of the concept covers the new instance; following the presentation of a new negative instance, an upper marker may be lowered so that the most general form of the concept no longer covers the new instance.

In an alternative interpretation (Thornton 1987) the description space is a geometric space and execution of the algorithm involves the shrinking of an outer hypercuboid (which defines the most general form of the concept being learned) and the expansion of an inner hypercuboid (which defines the most specific form of the concept being learned). Hypercuboid-manipulations correspond to marker-manipulations in the obvious way: following presentation of a new positive instance, the inner hypercuboid may be expanded so as to enclose the point which corresponds to the new instance; following presentation of a new negative instance, the outer hypercuboid may be shrunk so as to exclude the point which corresponds to the new instance.

To help clarify this alternative interpretation two descriptions of the behaviour of the Focussing algorithm will be contrasted. The first of these descriptions will invoke the standard, marker-oriented interpretation: the second will

invoke the alternative, hypercuboid-oriented interpretation.

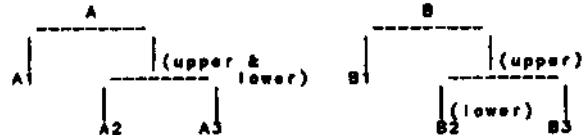


Fig. 1

Suppose that a description space consists of two trees called "A" and "B" whose structure is as depicted in Fig. 1. (Instances in this description space are simply conjunctions of tip nodes taken from trees A and B; e.g. A2 & B3.) The Focussing algorithm is presented with a sequence of four instances from this description space: <A2 & B2. A3 & B2. A2 & B1. A1 & B2>. The first two elements are positive instances of the concept to be learnt and the second two are negative instances. Presentation of the first positive instance (i.e. initialisation) causes the placing of upper markers at the root nodes of both trees and lower markers at the appropriate tip nodes (A2 & B2). Presentation of the second positive instance (A3 & B2) causes the lower marker in tree A to be raised to the parent of A2. Presentation of the first negative instance (A2 & B1) causes the upper marker in tree B to be lowered and presentation of the second negative instance (A1 & B2) causes the upper marker in tree A to be lowered. The positions of markers at this point are as depicted in Fig. 1.

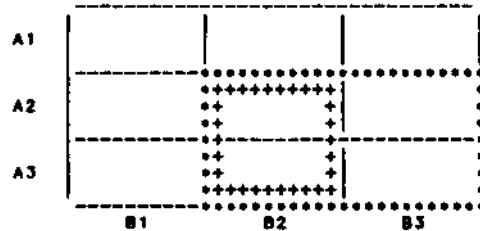


Fig. 2

The hypercuboid-oriented interpretation of the behaviour of the algorithm in this case is as follows. We say that the description space has two dimensions called "A" and "B" which range over the values {A1, A2, A3} and {B1, B2, B3} respectively. Presentation of the first positive instance (initialisation) triggers the construction of an inner hypercuboid enclosing the A2/B2 cell and an outer hypercuboid enclosing the entire space. Presentation of the second positive instance causes the inner hypercuboid to be expanded so as to include the A3/B2 cell. Presentation of the two negative instances cause the outer hypercuboid to be shrunk to exclude both the A1/B2 and A2/B1 cells. The structure of

the inner and outer hypercuboids at this point is as shown in Figure 2. (The outer hypercuboid is drawn using "x" characters: the inner hypercuboid is drawn using "v" characters.)

Note that the sequence of instances described does not enable the Focussing algorithm to construct a fully-specified concept: the most general form of the defined concept is not identical to the most specific form. Under the standard interpretation the evidence for this is the fact that one set of markers do not coincide (i.e. one of the trees is not firmed-up). In the hypercuboid-oriented interpretation the evidence is the fact that the hypercuboids are not identical.

2. The disjunctive-concept problem

Bundy et al. show that there are sequences of positive and negative instances which, when presented to the Focussing algorithm, will cause upper markers to be moved below lower markers or lower markers to be moved above upper markers. Such an event corresponds to the emergence of an inconsistency and the failure of the algorithm. This problem is frequently referred to as the disjunctive-concept problem on the grounds that such sequences may be associated with disjunctive concepts.

An inconsistency (i.e. the inversion of an upper and a lower marker) is described, under the hypercuboid-oriented interpretation, as the situation which arises when an inner hypercuboid boundary extends beyond an outer hypercuboid boundary. Viewing inconsistencies in this way leads to the observation that (i) there are precisely two situations which will cause inconsistencies to arise and (ii) that these situations will always lead to inconsistencies regardless of the order in which instances are presented.

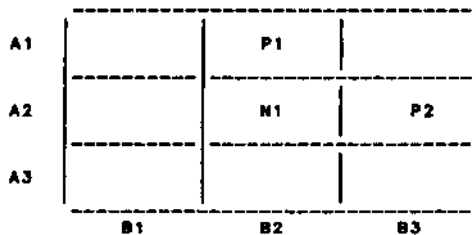


Fig. 3

Consider Fig. 3 which shows the geometric form of the previously defined description space. Positive instances are depicted as "Ps": negative instances are depicted as "Ns". Two positive instances are depicted (P1 and P2) and one negative instance (N1).

Note that as things stand it is impossible to construct a hypercuboid which encloses all the Ps and excludes all the Ns: clearly, any hypercuboid which encloses all the Ps must extend beyond a hypercuboid which excludes all the Ns. Translated into the standard interpretation, this observation says that, in any set of (upper and lower) markers which both cover all the positive instances and exclude all the negative instances, there will be at least one lower marker which is above an upper marker. Presentation of the corresponding set of instances will therefore inevitably lead to an inconsistency regardless of the presentation schedule.

If the only positive instance was P2, a valid hypercuboid (i.e. one which encloses all the Ps and excludes all the Ns) could be constructed, but not by the Focussing algorithm. The reason is very simple. The Focussing algorithm actually manipulates markers not hypercuboid boundaries and its boundary-manipulation capabilities are therefore constrained. It can locate boundaries at positions X and Y in some dimension D if and only if the values of D falling between X and Y are descended from a unique node in a tree of the underlying tree-based description space. The values in the singleton set {B3} are not descended from a unique node in the tree-based description space, therefore hypercuboid boundaries cannot be positioned so as to enclose them. Since any valid hypercuboid for {P2, N1} must have such boundaries, we can infer that the Focussing algorithm cannot construct a valid hypercuboid for this set of instances.

The Focussing algorithm then can fail either because (i) the valid hypercuboid does not exist or because (ii) it cannot be constructed due to constraints on boundary manipulations. In either of these situations failure will occur regardless of presentation schedules.

A useful distinction can therefore be made between two different versions of the disjunctive-concept problem: a strong version corresponding to situation (i) and a weak version corresponding to situation (ii). Drawing this distinction leads to the observation that techniques such as the "least-disjunction" procedure (Utgoff 1985) and "tree-hacking" (Bundy et al. 1985) which, in effect, alter the set of possible boundary positions by introducing novel disjunctions into the underlying description space, can alleviate the weak version of the disjunctive-concept problem but not the strong version.

3. Assignment heuristics

A modification of the Focussing algorithm which is capable of learning disjunctive-concepts is associated with Mitchell et al. (1983). This technique (called Shell Creation by Bundy et al. 1985) involves the construction of what are, in effect, multiple sets of markers (i.e. "rule-shells"). Under the hypercuboid-oriented interpretation this technique corresponds to the construction of sets of inner and outer hypercuboids (Thornton 1987).

Note that, in the case where multiple inner and outer hypercuboids have been constructed, the presentation of a new instance does not have deterministic consequences. If the new instance is a positive instance, then an inner hypercuboid may have to be expanded to enclose it. If the new instance is a negative instance, then an outer hypercuboid may have to be shrunk to exclude it. The question is (in both cases): which particular hypercuboid should be chosen?

Bundy et al. have noted that there appears to be no infallible way of choosing how to assign new instances to existing hypercuboids/rule-shells (1985, p. 170). The conclusion seems to be that assignments must be carried out using a heuristic.

There are obviously many possible assignment heuristics that might be used; however we can imagine a completely general heuristic which computes preferences for assignments on the basis of explicit features of the data (i.e. which does not employ domain-knowledge). It can be shown that the preferences computed by this heuristic must satisfy certain criteria. (We consider the case where there are two defined "points" in the description space named P1 and P2.)

We can assume that the completely general heuristic (like any other heuristic) must behave systematically. Therefore (i) the computed preference for assigning an instance occupying P1 to a hypercuboid enclosing P2 must be identically equal to the computed preference for assigning an instance occupying P2 to a hypercuboid enclosing P1: (ii) the computed preference for assigning an instance occupying P1 to a hypercuboid enclosing P1 must have a maximum value, since the limiting case is the assignment of an instance to a single-cell hypercuboid which already encloses it.

In addition, we can infer that preferences generated by the heuristic must be intrinsically coherent. This means that the computed preference for assigning an instance occupying P1 to a hypercuboid enclosing P2 must not be outweighed by the computed preference for assigning P1 first to some intermediate hypercuboid, and then on to the one enclosing P2. If computed preferences were of this sort, then we would be able to increase our level of preference for any particular assignment simply by devising ever more elaborate assignment schedules.

Any heuristic which satisfies these criteria must exploit what is, in effect, a valid distance function over the description space. Presented with any two points in the description space, the heuristic must be capable of generating a distinct level of preference for assigning one of the two points to a hypercuboid enclosing the other. This level of preference, interpreted as a measure of distance, satisfies all the axioms for a mathematical distance function, including triangle inequality, reflexivity and symmetry.

4. The Classification algorithm

Bundy et al. (ibid.) note that the Classification algorithm, associated with Quinlan (1983) and Hunt (1966), is capable of dealing flawlessly with disjunctive concepts. In their presentation of this algorithm, the description space is an attribute-value space and the action of the algorithm involves (an iteration of) the selection of an attribute A followed by the splitting of the instances (which are all given in advance) into sets, such that all the instances in a given set have the same value of A. This process terminates when every set of instances contains either uniformly positive or uniformly negative instances. The structure of attribute choices forms a decision tree (i.e. disjunctive rule) for the concept underlying the instances.

In the hypercuboid-oriented interpretation of this algorithm (Thornton 1987), the description space is a geometric space consisting of K dimensions each of which ranges over the possible values of a single attribute. Instances correspond to points in the normal way. Initialisation of the algorithm corresponds to the construction of a single hypercuboid which encloses all the points in the description space. This hypercuboid is then subdivided into K hypercuboids, where K is the number of values of the chosen attribute. Any of the K hypercuboids which do not contain uniform instances are then further subdivided. The division process continues until all hypercuboids contain instances which are either

uniformly positive or uniformly negative.

Note that a hypercuboid constructed by the Classification algorithm will have boundaries which, for each dimension D, enclose either a single value of D or the complete range of D. In other words, the boundaries in any one dimension of a hypercuboid formed by the Classification algorithm must, by definition, be either maximally close or maximally distant.

If we consider the entire set of possible hypercuboids which can be constructed in any given geometric space of a reasonable size, and compare it with the set of hypercuboids which can be constructed in the same space using boundaries which are of the described form, we will conclude that the size of the former set is overwhelmingly greater than the size of the latter set. The implication is that the Classification algorithm can only construct a small proportion of the possible hypercuboids in any given space and must therefore be assumed to be only capable, in general, of constructing a small proportion of the possible disjunctive concepts for a given domain.

5. Summary

We have shown that under the hypercuboid-based interpretations of the Focussing algorithm and the Classification algorithm a number of possibilities emerge: (i) a simple account of the disjunctive-concept problem can be presented; (ii) a distinction can be made between a strong and a weak version of the problem; (iii) techniques such as tree-hacking can be shown to alleviate only the weak version; (iv) solutions to the problem which involve the construction of multiple hypercuboids and the exploitation of completely general assignment heuristics can be shown to exploit what are, in effect, valid distance functions over the description space; (v) the Classification algorithm can be shown to be capable of forming only a subset of the possible disjunctive concepts for a given domain.

ACKNOWLEDGEMENTS

I would like to thank Allan Ramsay, Guy Scott, Steve Draper, Claire O'Malley and Andrew Law for providing extremely helpful comments on this paper.

REFERENCES

- Bundy, A., Silver, B. and Plummer, D.. "An Analytical Comparison of Some Rule-Learning Programs" *Artificial Intelligence*, 27, 1985, pp. 137-181.
- Hunt, E.B., Marin, J. and Stone, P.J.. *Experiments in Induction*, Academic Press, New York, 1966.
- Mitchell, T.M., Utgoff, P.E. & Banerji, R. "Learning by Experimentation: acquiring and modifying problem-solving heuristics" in R.S. Michalski, J.G. Carbonell and T.M. Mitchell (eds.) *Machine Learning*, Tioga, Palo Alto, CA, 1983, pp. 163-190.
- Quinlan, J.R.. "Learning Efficient Classification Procedures and their Application to Chess End Games" in R.S. Michalski, J.G. Carbonell and T.M. Mitchell, (eds.) *Machine Learning*, Tioga, Palo Alto, CA, 1983, pp. 463-482.
- Thornton, C.J. "Hypercuboid-formation behaviour of two learning algorithms." CSRP 067, The University of Sussex, Cognitive Studies Programme 1987.
- Utgoff, P.E. "Shift of Bias for Concept Learning" in R.S. Michalski, J.G. Carbonell and T.M. Mitchell (eds.) *Machine Learning: Vol 2*, Morgan Kaufman, Los Altos, 1986, pp. 107-148.