Non-accidental Features in Learning

Richard Mann and Allan Jepson*
Department of Computer Science
University of Toronto
Toronto, Canada, M5S 1A4

Abstract

We consider the application of symbolic learning in natural domains. Given a conceptualization of the domain in terms of a particular set of predicates, we consider how knowledge about the domain can constrain learning. In particular, by specifying a set of predicates which are not allowed to vary within a category we restrict the set of possible categories. We suggest that such constraints arise naturally in perception if we distinguish the set of "singular" predicates which describe degenerate configurations of the features (measure-zero events). This is related to the notion of *non-accidental features* in vision. We illustrate the approach by categorizing motion sequences in a simplified visual domain.

Introduction

In this paper we consider the application of symbolic learning in natural domains. Typically, symbolic learning approaches (cf. Winston, 1975) have assumed a learning system which operates independently of the feature extraction process and treats all features identically. We suggest that in order to achieve levels of performance comparable to humans, learning systems must use additional domain-dependent knowledge about the distribution of the features.

An example of the use of domain knowledge in learning is shown in Figure 1. Feldman (1992b) observed that when shown a single positive example of the concept in Figure 1(a), subjects generalize to produce additional examples as shown in Figure 1(b). In generalizing the example subjects implicitly assume that the placement of the dot *on the line* is an important feature, while the exact *position* of the dot (one third of the way up from the end of the line) is not. What is important here is that there is a degenerate configuration (measure-zero event) of the features (the coincidence of the dot and the line in the image) that provides a *reliable inference*[1] to a familiar world regularity (contact between a dot and a line). It is the nature of this world regularity that determines the distribution of the various features within the category. Similar observations have been made in the computer vision literature where so called *non-accidental features* (parallelism, cotermination, symmetry, etc.) are used to make inferences about about world structure (Witkin & Tanenbaum, 1983; Jepson & Richards, 1992).

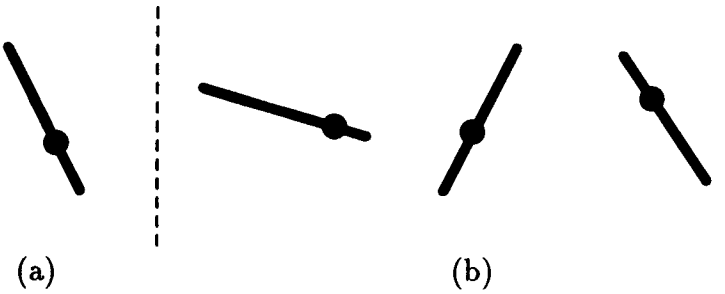

Figure 1: The use of domain-dependent knowledge in categorization. Subjects shown the single positive example in (a) generated further examples as shown in (b). (after Feldman,1992b).

In a learning system without any knowledge of the domain, however, features are treated identically and additional examples are required to learn the correct categories[2]. We suggest that rather than providing additional examples, assumptions about world structure can be presented explicitly to the learning system in the form of constraints on the features within categories. We propose a simple modification to the standard algorithms for learning conjunctive concepts to include this information. To illustrate the ideas we will focus on a concrete example of categorizing motion from visual input.

Qualitative Description of Motion

The description of motion has been studied by many researchers. The description of motion events was first studied by linguists (Miller, 1972) who attempted to characterize various motion verbs (throwing, catching, etc.) based on their underlying properties such as the cause and the direction of the motion. The problem of recognizing motion events directly from image sequences was first studied by Badler (1975), and later developed for more specialized domains such as biomedical images (Tsotsos, 1980) and traffic scenes (Neumann, 1984). The description of

*Richard Mann was supported by an NSERC postgraduate fellowship. Allan Jepson was supported by CIAR, IRIS and NSERC Canada. We would like to thank Sven Dickenson, Dale Schuurmans and Chris Williams for helpful comments on this work. Correspondence should be addressed to the first author (Email: mann@vis.toronto.edu).

[1](Jepson & Richards, 1992).

[2]For example, to learn the dot-on-line concept above, two additional training examples are required: one positive example to show that the position of the dot along the line can vary, and one negative example to show that the dot must not move off the line.

motion is also the subject of much research in the area of qualitative reasoning (Bobrow, 1985). All of these studies assume a well defined hierarchy of motion events specific to the domain. In this paper we consider how such natural categories may be learned from visual input. In particular, we consider how domain knowledge will constrain the categories formed.

To illustrate the problem consider a simple domain containing a ball and a table. We use an *interval based* representation (McDermott, 1980) consisting of a series of "motion eras" described by a set of *predicates* (binary valued features) indicating the motion of the ball as shown in Figure 2. The description is *qualitative* in the sense that each assignment of the predicates specifies a "primitive" category that represents a range of possible observations.

T: touches table
X: moves horizontally
L: moves left
R: moves right
Y: moves vertically
U: moves up
D: moves down
G: gravitational motion (ie., accelerate downward at 9.8m/s^2)

Figure 2: Predicates used to describe motion of ball.

Based on consistency of the predicate assignments (and assuming the table is horizontal) we can rule out most combinations, but are still left with 18 possible eras as shown in Figure 3. Let us consider a world where we observe only eight of the possible motion eras as shown in Figure 3(a). These observations would arise, for example, in a domain where the only source of support is from the table, and where the ball is a "beanbag" which does not bounce when it hits the table. Once given a set of relevant predicates, the task of the machine learning system is to decide how the primitive categories should be grouped into different classes.

Natural Constraints on Learning

The problem of grouping the observations into one or more classes is one of *unsupervised learning*, or *clustering*. Following the approach of Michalski & Stepp (1983) we can cluster the observations into a set of one or more *conjunctive concepts* where the observed motions are treated as positive examples and the unobserved cases as negative examples. We define a *categorization* $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ as a set of conjunctive concepts with each $C_i \in [T, F, *]^n$ specifying a *generalization* of one or more of the primitive categories. A categorization is a *covering* if each observed case can be generalized to one or more of the concepts, and none of the unobserved cases are covered[3]. We typically choose the most general set of concepts given by the preference,

[3]Note that we allow the concepts to cover inconsistent assignments of the predicates as well as the observed cases—we assume that in describing the categorization the user specifies the consistency constraints as well as the covering.

$TXL\overline{YG}$ rolling(L)	$TXR\overline{YG}$ rolling(R)	$T\overline{XYG}$ sitting	$\overline{TX}YDG$ falling
$\overline{T}XLYUG$ proj(LU)	$\overline{T}XRYUG$ proj(RU)	$\overline{T}XLYDG$ proj(LD)	$\overline{T}XLYDG$ proj(RD)

(a)

$\overline{T}XLYU\overline{G}$ flying(LU)	$\overline{T}XRYU\overline{G}$ flying(RU)	$\overline{T}XLYD\overline{G}$ flying(LD)	$\overline{T}XRYD\overline{G}$ flying(RD)
$\overline{TX}YU\overline{G}$ flyvert(U)	$\overline{TX}YD\overline{G}$ flyvert(D)	$\overline{T}XL\overline{YG}$ flyhoriz(L)	$\overline{T}XR\overline{YG}$ flyhoriz(R)
$\overline{TX}YUG$ bouncing	$\overline{TXYG}$ floating		

(b)

Figure 3: The primitive categories of motion eras given by the consistent assignments of predicates. (a) Observed cases. (b) Remaining unobserved cases.

Ordering (Parsimony): A partition $\mathcal{C}'$ is preferred over $\mathcal{C}$ only if each concept in $\mathcal{C}$ is covered by a single concept in $\mathcal{C}'$ and there are fewer concepts in $\mathcal{C}'$.

For the example considered here, possible generalizations of the positive examples are shown in Figure 4. Choosing the most general cover gives the categories {*ontable*, *projectile*, *freefall*} shown at the top of the figure.

At this point we note that blindly applying the learning algorithm to obtain the maximally general covers does not form reasonable categories. For example, while it seems reasonable to group leftward and rightward *rolling* in the same category, we may not wish to group *sitting* and *rolling*. That is, because *sitting* is really a degenerate case of *rolling* (in which the horizontal velocity is absent) it is assumed that a different underlying process is involved and these events should be categorized differently. To implement this constraint we distinguish the set of "singular" predicates which describe degenerate configurations of the feature space (ie., measure-zero events) and stipulate that the categories must not generalize these predicates,

Constraint (Singular predicates): A covering $\mathcal{C}$ is disallowed if it generalizes one or more predicates which describe degenerate configurations of the feature space.

In our example, the singular predicates correspond to T, X, Y, G which describe degenerate configurations of position, horizontal and vertical velocities, and acceleration respectively. Applying the constraint yields a unique preferred covering with the categories {*rolling*, *sitting*, *projectile*, *falling*}, as shown by the boxes in Figure 4. This corresponds to our intuitive notion of motion categories.

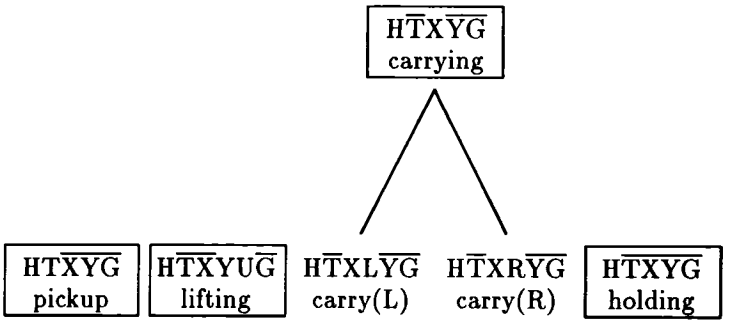

Figure 5: New eras observed in a new context where a robot hand touches the ball (designated by predicate H). Note that due to the singular predicate, H, these observations are categorized independently of those in Figure 4.

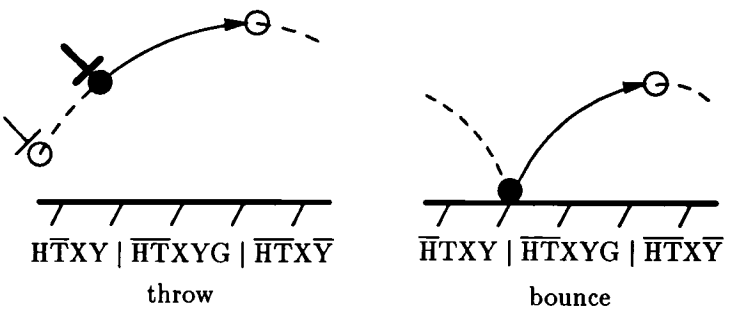

Figure 6: Using information at the interval boundaries (beginning and end points) to disambiguate between the *throw* and *bounce* events. The values of the predicates at the beginning and end of the intervals are shown at the left and right of the eras respectively.

vertical alignment of the bug and the ball, the single category *drop-ball-on-bug* would describe this event. In this situation the singular predicate *bug-under-hand* would allow us to *predict* the dropping event. Note that in each of these examples the system must be given new, sometimes non-obvious features. This process is similar to the idea of noticing "suspicious coincidences" in features (Barlow, 1990) and the idea of detecting "unexplained regularities" in *scientific discovery* (Langley et al, 1987).

Describing "events" by changes in motion

As noted by Badler (1975) we consider any *change* in motion as an "event". For example, as shown in Figure 7, the change between the era where the hand contacts the ball and where the ball is free falling define a *dropping* or *throwing* event. Using the clustering principles described above, these events, as well as others such as bouncing, hitting, etc., could be categorized. We do not develop these ideas here, but the representation would be similar to that proposed by Borchardt (1990) where a set of templates is used to describe all possible *transitions* between different motion states or eras[4]. The main difference, however, is that rather than merely storing a series of cases (or applying some domain-independent clustering heuristic), our

[4] A similar idea of storing a series of "visual cases" to describe motion events (eg., the collision between different objects) was proposed by Narayanan (1992)

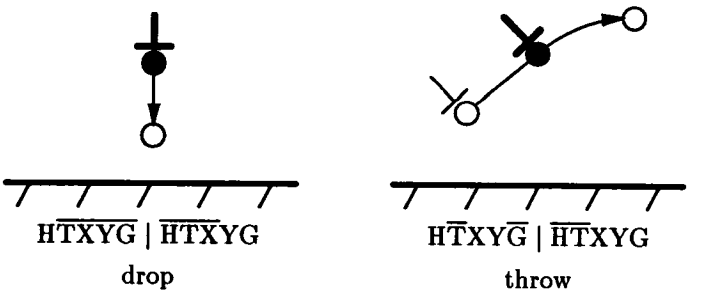

Figure 7: Classification of "events" based on the transition between motion eras. The values of the predicates before and after the transition are shown.

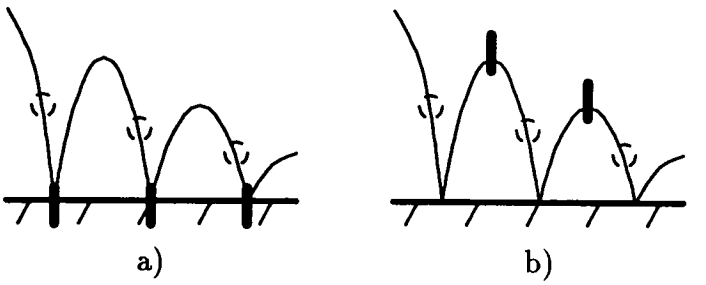

Figure 8: Segmenting motion sequences. According to Rubin (1986) the motion is usually segmented at the points of contact with the table as in (a). (b) shows another possible segmentation.

approach is to cluster events based on the particular set of singular predicates relevant to the domain.

A more difficult question, however, concerns how to define events in the first place. Consider the bouncing ball shown in Figure 8. According to the set of features presented in this paper the sequence would be segmented at both the bounces and the peaks of the path since the direction of motion changes in each case. According to experiments performed by Rubin (1986), however, subjects generally segment the motion only at the bounces as shown in Figure 8(a). Rubin explains this by noting that the bounces have a sudden change of direction which implies a force impulse, while the peaks follow a smooth path with no implied change in forces.

Discussion

In this paper we have proposed a general principle for constraining learning in natural domains based on simple assumptions about the set of predicates we use to conceptualize the domain. As noted by Mitchell (1980), we can view our singularity constraint on categories as a *bias* in learning. We can also view this approach as specifying a very weak form of *domain theory* to the learning algorithm. Similar ideas of restricting generalization have been proposed by other authors (cf. Soloway, 1978) but have typically been based on ad hoc rules. Here we attempt to justify our biases based on prior assumptions about the events in the world (ie., our choice of singular predicates). This idea is related to the notion of "non-accidental features" in vision. Witkin & Tanenbaum (1983) noted that non-accidental image features such as parallelism and coter-

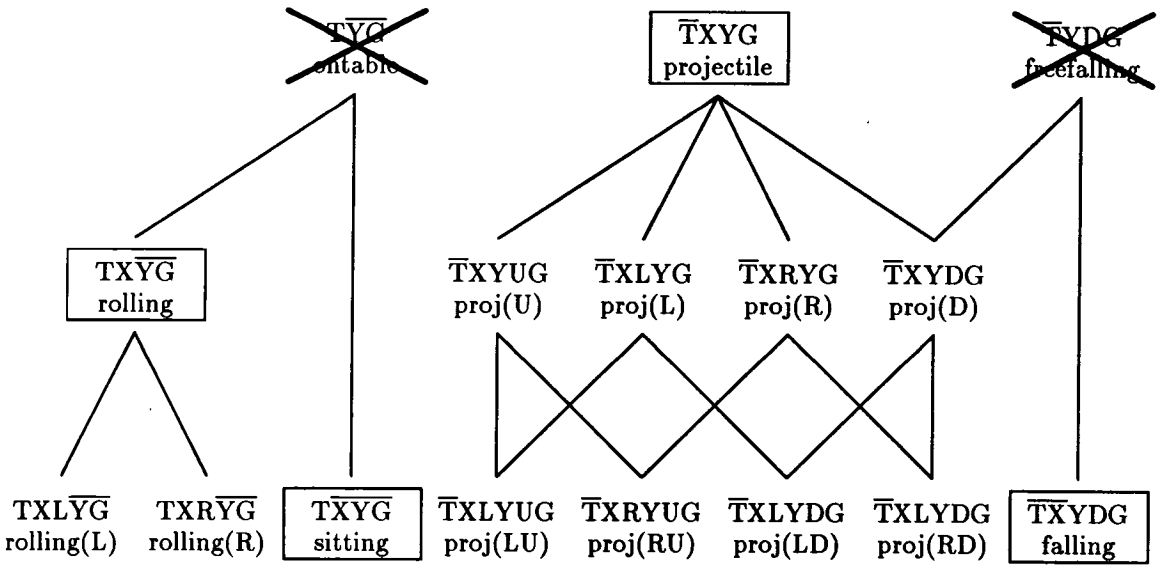

Figure 4: Generalization of observed motion eras (bottom row) to conjunctive concepts. The X's mark generalizations prevented by the constraint discussed in the paper. The boxes show the preferred categories.

Singular Predicates and Qualitative Probabilities

The singular predicate constraint has a natural interpretation if we view each concept as specifying a qualitative probability distribution over the domain. A predicate is *singular* with respect to a particular domain if it (or its negation) describes a measure-zero subset of the feature space which occurs with non-zero probability. The singular predicate constraint simply means that we don't allow categories which contain singularities.

As described by Jepson & Richards (1992), we note that the set of singular predicates is a result of the types of regularities that exist in the world. For example, the observation of G (acceleration at 9.8 m/s^2) is due to the fact that there is gravity in the world, the observation of $\overline{Y}$ (zero vertical velocity) is due to the fact the objects can be supported by the table, etc. Note that other features, such as a particular non-zero value of horizontal velocity, are not singular predicates since there are no corresponding regularities in the world that occur with a non-zero probability.

An important unanswered question concerns how to determine the set of singular predicates for a particular domain in the first place. While there are in principle an infinite number of possible singularities, we can restrict our search to plausible ones based on some general knowledge about the domain (ie., the types of objects we expect in the world). This view is supported by psychological experiments for the motion domain where Rubin (1986) demonstrates that there are only a small number of perceptually distinguishable motion events (eg., starts, stops, accelerations, etc.) that are used to form categories. A more general question concerns which of the remaining (non-singular) predicates are relevant to forming categories. While there are no general criteria for determining relevance of particular features to a class (see Watanabe, 1985 for a discussion) it may be possible to provide some specific principles to guide learning in particular domains. For example, as described by Pazzani (1990) we may use assumptions about causality (eg., temporal precedence, no action at a distance) to rule out irrelevant features in learning. Evidence for similar principles (eg., object persistence, spatiotemporal continuity, support relations) has been presented in the developmental psychology literature (see Spelke, 1988 for a review).

Extending Learning

In the previous section we showed how domain knowledge about the predicates could be used to constrain the categories formed. We now sketch an approach to learning in more complex domains by considering several examples.

Extending the context by adding new features

Consider a system which observes a robot hand as well as the ball and table. Suppose that when the hand does not contact the ball we observe the same set of eras as above, but when the hand contacts the ball we observe the new eras shown in Figure 5. According to the principles described above we would form the new categories {*pickup*, *lifting*, *carrying*, *holding*} distinguished by the presence of the hand. According to the singularity constraint, each new assignment of the singular predicates (hand contact in this case) defines a new *context* for categorization. As another example, consider motion eras due to the two events, *throwing* and *bouncing* as shown in Figure 6. In order to differentiate between the two events, the system must consider the boundaries of the motion eras. The fact that the throw era starts with hand contact while the bounce era starts with table contact serve to differentiate the events. Finally, an interesting special case occurs when a new context yields a single category. For example, imagine a system which observes a *dropping* event (ie., transition from *holding* to *falling*) whenever a bug moving on the table comes directly under the hand holding the ball. If the system had a singular predicate describing the

mination between line segments provide strong cues for world structure. Feldman (1992a, 1992b) used a similar argument to justify the use of such regularities in categorization. A formal analysis of non-accidental features and perceptual inference is given by Bennett et al (1989) and Jepson & Richards (1992). The idea of a qualitative description of probability distributions in terms of singularity structure is similar to that proposed by Bennett et al (1991).

With respect to the domain of motion sequences, much remains to be done. As noted in the last example, in order to classify motion sequences more finely, some form of *state information* will be required. While a full representation of state and temporal constraints (for example, using the *situation calculus* described by McCarthy & Hayes, 1969) may be required in general, a partial representation may be sufficient. As noted by Rubin (1986) some new set of features based on the underlying dynamics (forces, mass, gravity, etc.) may be used. While our approach may be extended in this direction, it appears that to describe more complex motion events (eg., throwing, catching, etc.) we will need more abstract features to describe physical relations such as *contact*, *attachment*, and *support* (see Siskind, 1992 for a discussion). We are currently investigating ways to extend the learning framework in this way.

References

Badler N (1975), *Temporal scene analysis: Conceptual descriptions of object movements*, PhD thesis, University of Toronto.

Barlow H (1990), Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception, *Vision Research*, 30:1561–1571.

Bennett B, Hoffman D & Prakesh C (1989), *Observer mechanics*, London: Academic Press.

Bennett B, Hoffman D & Richards W (1991), *Reasoning under uncertainty: Lebesgue logic and order logic*, preprint.

Borchardt G C (1990), *Transition space*, AI Memo 1238, MIT.

Bobrow D G (1985), *Qualitative reasoning about physical systems*, Cambridge, MA: MIT Press.

Feldman J (1992a), Constructing perceptual categories, *CVPR 92*.

Feldman J (1992b), *Perceptual categories and world regularities*, PhD thesis, MIT, Department of Brain and Cognitive Sciences.

Jepson A & Richards W (1992), What makes a good feature?, In *Spatial vision in humans and robots*, Harris L & Jenkin M (eds.), Cambridge: Cambridge University Press.

Langley P, Simon H A, Bradshaw G L & Zytkow J M (1987), *Scientific discovery: Computational explorations of the creative process*, Cambridge, MA: MIT Press.

McDermott (1982) A temporal logic for reasoning about processes and plans, *Cognitive Science* 6:101–155.

McCarthy J & Hayes P J (1969), Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, Meltzer B & Michie D (eds.), Edinburgh: Edinburgh University Press.

Michalski R S & Stepp R E (1983), Automated construction of classifications: conceptual clustering versus numerical taxonomy, *IEEE Trans. on Pat. Anal. & Mach. Intell.*, 5(4):396–410.

Miller G (1972), English verbs of motion: A case study in semantics and lexical memory, In *Coding processes and human memory*, Melton & Martin (eds.), Washington: V H Winston & Sons.

Mitchell T M (1980), *The need for biases in learning generalizations*, Technical Report CBM-TR-117, New Brunswick, NJ: Rutgers University, Department of Computer Science.

Narayanan H (1992), *Imagery, Diagrams, and Visual Reasoning*, PhD thesis, Ohio State University.

Neumann B (1984), Natural language description of time-varying scenes, In *Advances in natural language processes*, Waltz D (ed.), Hillsdale, NJ: Lawrence Erlbaum.

Pazzani M J (1991), A computational theory of learning causal relationships, *Cognitive Science* 15:401–424.

Rubin J (1986), *Categories of visual motion*, PhD thesis, MIT, Department of Brain and Cognitive Sciences.

Soloway E M (1978), *Learning = Interpretation + Generalization: A case study in knowledge-directed learning*, COINS-TR-78-13, University of Massachusetts at Amherst.

Siskind J M (1992), *Naive physics, event perception, lexical semantics and language acquisition*, PhD thesis, MIT, Department of Electrical Engineering and Computer Science.

Spelke E S (1988), The origins of physical knowledge, In *Thought without language*, Weiskrantz L (ed.), Clarendon Press.

Tsotsos J K, Mylopoulos J, Covey H D & Zucker S W (1980), A framework for visual motion understanding, *IEEE Trans. on Pat. Anal. & Mach. Intell.*, 2(6):563–573.

Watanabe S (1985), *Pattern recognition: human and mechanical*, Wiley.

Winston P H (1975), Learning structural descriptions from examples, in *The psychology of computer vision*, Winston P H (ed.), New York: McGraw-Hill.

Witkin A & Tanenbaum J M (1983), On the role of structure in vision, In *Human and machine vision*, Beck J, Hope B & Rosenfeld A (eds.), New York: Academic Press.