LEARNING PERCEPTUALLY-GROUNDED SEMANTICS IN THE L_0 PROJECT

Terry Regier*
International Computer Science Institute
1947 Center Street, Berkeley, CA, 94704
(415) 642-4274 x 184
regier@cogsci.Berkeley.EDU

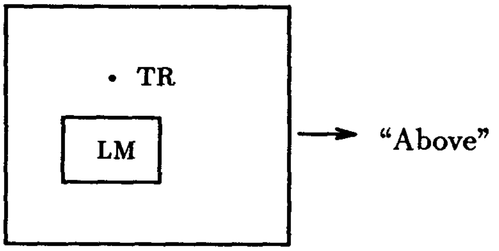

Figure 1: Learning to Associate Scenes with Spatial Terms

ABSTRACT

A method is presented for acquiring perceptually-grounded semantics for spatial terms in a simple visual domain, as a part of the L_0 miniature language acquisition project. Two central problems in this learning task are (a) ensuring that the terms learned generalize well, so that they can be accurately applied to new scenes, and (b) learning in the absence of explicit negative evidence. Solutions to these two problems are presented, and the results discussed.

1 Introduction

The L_0 language learning project at the International Computer Science Institute [Feldman *et al.*, 1990; Weber and Stolcke, 1990] seeks to provide an account of language acquisition in the semantic domain of spatial relations between geometrical objects. Within this domain, the work reported here addresses the subtask of learning to associate scenes, containing several simple objects, with terms to describe the spatial relations among the objects in the scenes. This is illustrated in Figure 1.

For each scene, the learning system is supplied with an indication of which object is the reference object (we call this object the *landmark*, or LM), and which object is the one being located relative to the reference object (this is the *trajector*, or TR). The system is also supplied with a *single* spatial term that describes the spatial relation portrayed in the scene. It is to learn to associate *all* applicable terms to novel scenes.

The TR is restricted to be a single point for the time being; current work is directed at addressing the more general case of an arbitrarily shaped TR.

Another aspect of the task is that learning must take place in the absence of explicit negative instances. This condition is imposed so that the conditions under which learning takes place will be similar in this respect to those under which children learn.

Given this, there are two central problems in the subtask as stated:

- Ensuring that the learning will generalize to scenes which were not a part of the training set. This means that the region in which a TR will be considered "above" a LM may have to change size, shape, and position when a novel LM is presented.
- Learning without explicit negative evidence.

This paper presents solutions to both of these problems. It begins with a general discussion of each of the two problems and their solutions. Results of training are then presented. Then, implementation details are discussed. And finally, some conclusions are presented.

2 Generalization and Parameterized Regions

2.1 The Problem

The problem of learning whether a particular point lies in a given region of space is a foundational one, with several widely-known "classic" solutions [Minsky and Papert, 1988; Rumelhart and McClelland, 1986]. The task at hand is very similar to this problem, since learning when "above" is an appropriate description of the spatial relation between a LM and a point TR really amounts to learning what the extent of the region "above" a LM is.

However, there is an important difference from the classic problem. We are interested here in learning whether or not a given point (the TR) lies in a region (say "above", "in") which is itself *located relative to* a LM. Thus, the shape, size, and position of the region are dependent on the shape, size, and position of the current LM. For example, the area "above" a small triangle toward the top of the visual field will differ in shape, size,

*Supported through the International Computer Science Institute.

and position from the area "above" a large circle in the middle of the visual field.

2.2 Parameterized Regions

Part of the solution to this problem lies in the use of *parameterized regions.* Rather than learn a fixed region of space, the system learns a region which is parameterized by several features of the LM, and is thus dependent on them.

The LM features used are the location of the center of mass, and the locations of the four corners of the smallest rectangle enclosing the LM (the LM's "bounding-box"). Learning takes place relative to these five "key points".

Consider Figure 2. The figure in (a) shows a region in 2-space learned using the intersection of three half-planes, as might be done using an ordinary perceptron. In (b), we see the same region, but learned ***relative to the five key points of an LM.*** This means simply that the lines which define the half-planes have been constrained to pass through the key points of the LM. The method by which this is done is covered in Section 5. Further details can be found in [Regier, 1990].

The critical point here is that now that this region has been learned relative to the LM key points, ***it will change position and size*** when the LM key points change. This is illustrated in (c). Thus, the region is parameterized by the LM key points.

2.3 Combining Representations

While the use of parameterized regions solves much of the problem of generalizability across LMs, it is not sufficient by itself. Two objects could have identical key points, and yet differ in actual shape. Since part of the definition of "above" is that the TR is not in the interior of the LM, and since the shape of the interior of the LM cannot be derived from the key points alone, the key points are an underspecification of the LM for our purposes.

The complete LM specification includes a bitmap of the interior of the LM, the "LM interior map". This is simply a bitmap representation of the LM, with those bits set which fall in the interior of the object. As we shall see in greater detail in Section 5, this representation is used together with parameterized regions in learning the perceptual grounding for spatial term semantics. This bitmap representation helps in the case mentioned above, since although the triangle and square will have identical key points, their LM interior maps will differ. In particular, since part of the learned "definition" of a point being above a LM should be that it may not be in the interior of the LM, that would account for the difference in shape of the regions located above the square and above the triangle.

Parameterized regions and the bitmap representation, when used together, provide the system with the ability to generalize across LMs. We shall see examples of this after a presentation of the second major problem to be tackled.

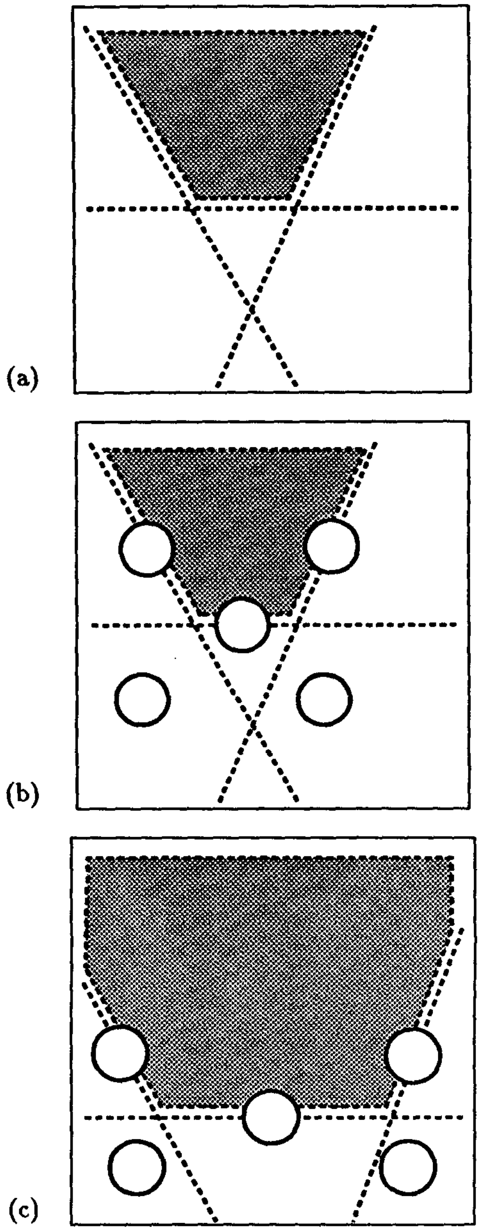

Figure 2: Parameterized Regions

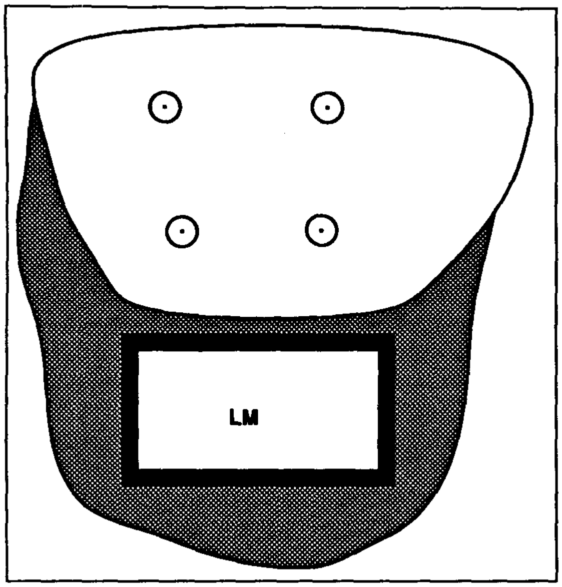

Figure 3: Learning "Above" Without Negative Instances

3 Learning Without Explicit Negative Evidence

3.1 The Problem

Researchers in child language acquisition have often observed that the child learns language apparently without the benefit of negative evidence [Braine, 1971; Bowerman, 1983; Pinker, 1989]. While these researchers have focused on the "no negative evidence" problem as it relates to the acquisition of grammar, the problem is a general one, and appears in several different aspects of language acquisition. In particular, it surfaces in the context of the learning of the semantics of lexemes for spatial relations. The methods used to solve the problem here are of general applicability, however, and are not restricted to this particular domain.

The problem is best illustrated by example. Consider Figure 3. Given the landmark (labeled "LM"), the task is to learn the concept "above". We have been given four positive instances, marked as small dotted circles in the figure, and no negative instances. The problem is that we want to generalize so that we can recognize new instances of "above" when they are presented, but since there are no negative instances, it is not clear where the boundaries of the region "above" the LM should be. One possible generalization is the white region containing the four instances. Another possibility is the union of that white region with the dark region surrounding the LM. Yet another is the union of the light and dark regions with the interior of the LM. And yet another is the correct one, which is not closed at the top. In the absence of negative examples, we have no obvious reason to prefer one of these generalizations over the others.

One possible approach would be to take the smallest region that encompasses all the positive instances. It should be clear, however, that this will always lead to closed regions, which are incorrect characterizations of such spatial concepts as "above" and "outside". Thus, this cannot be the answer.

And yet, humans do learn these concepts, apparently in the absence of negative instances. The following sections indicate how that learning might take place.

3.2 A Possible Solution and its Drawbacks

One solution to the "no negative evidence" problem which suggests itself is to take every *positive* instance for one concept to be an *implicit negative* instance for all other spatial concepts being learned. There are problems with this approach, as we shall see, but they are surmountable.

There are related ideas present in the child language literature, which support the work presented here. [Markman, 1987] posits a "principle of mutual exclusivity" for object naming, whereby a child assumes that each object may only have one name. This is to be viewed more as a learning strategy than as a hard-and-fast rule: clearly, a given object may have many names (an office chair, a chair, a piece of furniture, etc.). The method being suggested really amounts to a principle of mutual exclusivity for spatial relation terms: since each spatial relation can only have one name, we take a positive instance of one to be an implicit negative instance for all others.

In a related vein, [Johnston and Slobin, 1979] note that in a study of children learning locative terms in English, Italian, Serbo-Croatian, and Turkish, terms were learned more quickly when there was little or no synonymy among terms. They point out that children seem to prefer a one-to-one meaning-to-morpheme mapping; this is similar to, although not quite the same as, the mutual exclusivity notion put forth here.[1]

In linguistics, the notion that the meaning of a given word is partly defined by the meanings of other words in the language is a central idea of structuralism. This has been recently reiterated by [MacWhinney, 1989]: "the semantic range of words is determined by the particular contrasts in which they are involved". This is consonant with the view taken here, in that contrasting words will serve as implicit negative instances to help define the boundaries of applicability of a given spatial term.

There is a problem with mutual exclusivity, however. Using it as a method for generating implicit negative instances can yield many *false negatives* in the training set, i.e. implicit negatives which really should be positives.

Consider the following set of terms, which are the ones learned by the system described here:

- above
- below
- on
- off

[1] They are not quite the same since a difference in meaning need not correspond to a difference in actual reference. When we call a given object both a "chair" and a "throne", these are different meanings, and this would thus be consistent with a one-to-one meaning-to-morpheme mapping. It would not be consistent with the principle of mutual exclusivity, however.

- inside
- outside
- to the left of
- to the right of

If we apply mutual exclusivity here, the problem of false negatives arises. For example, not all positive instances of "outside" are accurate negative instances for "above", and indeed *all* positive instances of "above" should in fact be *positive* instances of "outside", and are instead taken as negatives, under mutual exclusivity.

"Outside" is a term that is particularly badly affected by this problem of false implicit negatives: all of the spatial terms listed above except for "in" (and "outside" itself, of course) will supply false negatives to the training set for "outside".

The severity of this problem is illustrated in Figure 4. In these figures, which represent training data for the spatial concept "outside", we have tall, rectangular landmarks, and training points[2] relative to the landmarks. Positive training points (instances) are marked with circles, while negative instances are marked with X's. In (a), the negative instances were placed there by the teacher, showing exactly where the region *not* outside the landmark is. This gives us a "clean" training set, but the use of teacher-supplied explicit negative instances is precisely what we are trying to get away from. In (b), the negative instances shown were derived from positive instances for the other spatial terms listed above, through the principle of mutual exclusivity. Thus, this is the sort of training data we are going to have to use. Note that in (b) there are many false negative instances among the positives, to say nothing of the positions which have been marked as both positive and negative.

This issue of false implicit negatives is the central problem with mutual exclusivity.

3.3 Salvaging Mutual Exclusivity

The basic idea used here, in salvaging the idea of mutual exclusivity, is to treat positive instances and implicit negative instances differently during training:

> Implicit negatives are viewed as supplying only *weak* negative evidence.

The intuition behind this is as follows: since the implicit negatives are arrived at through the application of a fallible heuristic rule (mutual exclusivity), they should count for less than the positive instances, which are all assumed to be correct. Clearly, the implicit negatives should not be seen as supplying excessively weak negative evidence, or we revert to the original problem of learning in the (virtual) absence of negative instances. But equally clearly, the training set noise supplied by false negatives is quite severe, as seen in the figure above. So this approach is to be seen as a compromise, so that we can use implicit negative evidence without being overwhelmed by the noise it introduces in the training sets for the various spatial concepts.

The details of this method, and its implementation under back-propagation, are covered in Section 5. However,

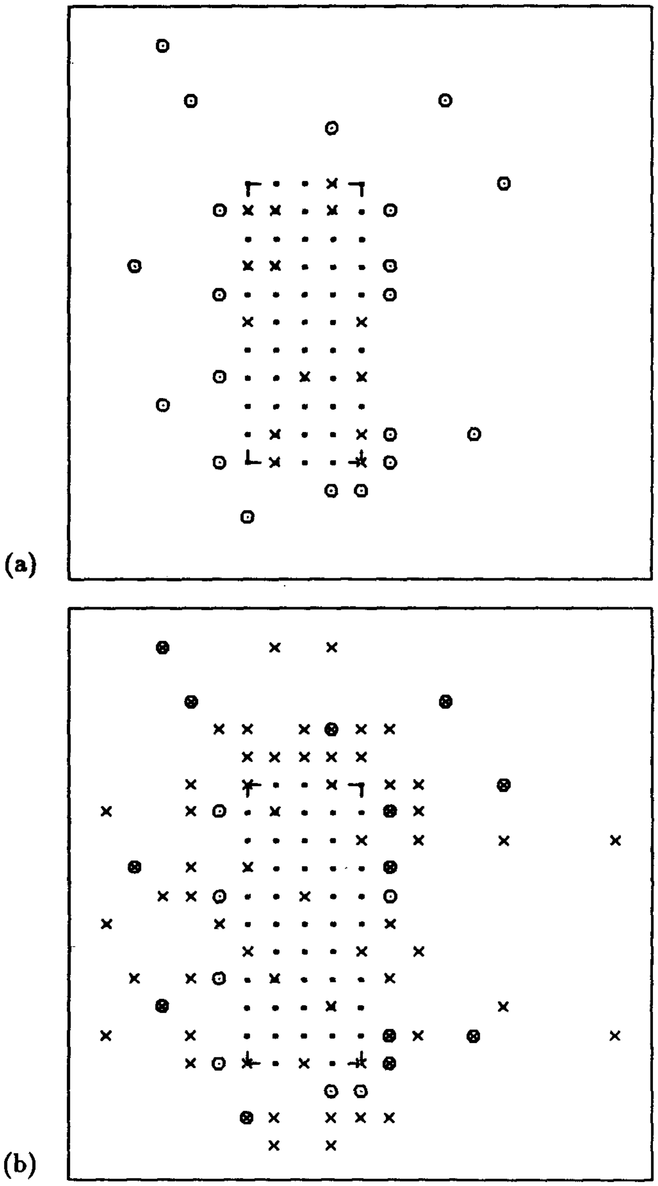

Figure 4: Ideal and Realistic Training Sets for "Outside"

[2] I.e. trajectors consisting of a single point each

this is a very general solution to the "no negative evidence" problem, and can be understood independently of the actual implementation details. Any learning method which allows for weakening of evidence should be able to make use of it. In addition, it could serve as a means for addressing the "no negative evidence" problem in other domains. For example, a method analogous to the one suggested here could be used for object naming, the domain for which Markman suggested mutual exclusivity. This would be necessary if the problem of false implicit negatives is as serious in that domain as it is in this one.

4 Results

This section presents the results of training.

Figure 5 shows the results of learning the spatial term "outside", first without negative instances, then using implicit negatives obtained through mutual exclusivity, but without weakening the evidence given by these, and finally with the negative evidence weakened.

The landmark in each of these figures is a triangle. The system was trained using only rectangular landmarks.

The size of the black circles indicates the appropriateness, as judged by the trained system, of using the term "outside" to refer to a particular position, relative to the LM shown. Clearly, the concept is learned best when implicit negative evidence is weakened, as in (c). When no negatives at all are used, the system overgeneralizes, and considers even the interior of the LM to be "outside" (as in (a)). When mutual exclusivity is used, but the evidence from implicit negatives is not weakened, the concept is learned very poorly, as the noise from the false implicit negatives hinders the learning of the concept (as in (b)). Having *all* implicit negatives supply only weak negative evidence greatly alleviates the problem of false implicit negatives in the training set, while still enabling us to learn without using explicit, teacher-supplied negative instances.

It should be noted that in general, when using mutual exclusivity without weakening the evidence given by implicit negatives, the results are not always identical with those shown in Figure 5(b), but are always of approximately the same quality.

Regarding the issue of generalizability across LMs, two points of interest are that:

- The system had not been trained on an LM in exactly this position.
- The system had never been trained on a triangle of any sort.

Thus, the system generalizes well to new LMs, and learns in the absence of explicit negative instances, as desired. All eight concepts were learned successfully, and exhibited similar generalization to new LMs.

5 Details

The system described in this section learns perceptually-grounded semantics for spatial terms using the

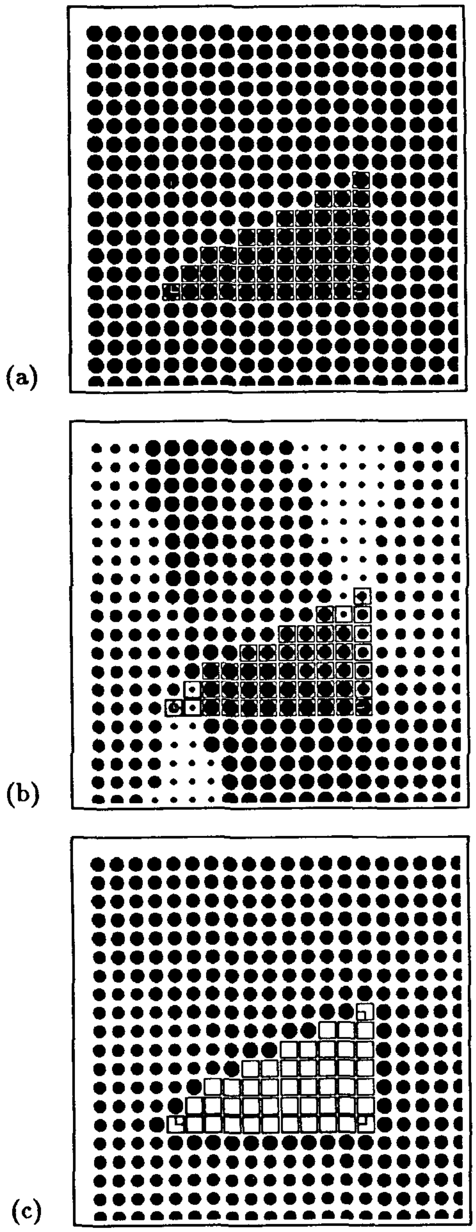

Figure 5: "Outside" without Negatives, and with Strong and Weak Implicit Negatives

quickprop[3] algorithm [Fahlman, 1988], a variant on back-propagation [Rumelhart and McClelland, 1986].

This presentation begins with an exposition of the representation used, and then moves on to the specific network architecture, and the basic ideas embodied in it. The weakening of evidence from implicit negative instances is then discussed.

5.1 Representation of the LM and TR

As mentioned above, the representation scheme for the LM comprises the following:

- A bitmap in which those pixels corresponding to the *interior* of the LM are the only ones set.
- The x, y coordinates of several "key points" of the LM, where x and y each vary between 0.0 and 1.0, and indicate the location of the point in question as a fraction of the width or height of the image. The key points currently being used are the center of mass (CoM) of the LM, and the four corners of the LM's bounding box (UL: upper left, UR: upper right, LL: lower left, LR: lower right).

The (punctate) TR is specified by the x, y coordinates of the point.

The activation of an output node of the system, once trained for a particular spatial concept, represents the appropriateness of using the spatial term in describing the TR's location, relative to the LM.

5.2 Architecture

Figure 6 presents the architecture of the system. The eight spatial terms mentioned above are learned simultaneously, and they share hidden-layer representations.

5.2.1 Receptive Fields

Consider the right-hand part of the network, which receives input from the LM interior map. Each of the three nodes in the cluster labeled "I" (for interior) has a receptive field of five pixels.

When a TR location is specified, the values of the five neighboring locations shown in the LM interior map, centered on the current TR location, are copied up to the five input nodes. The weights on the links between these five nodes and the three nodes labeled "I" in the layer above define the receptive fields learned. When the TR position changes, five new LM interior map pixels will be "viewed" by the receptive fields formed. This allows the system to detect the LM interior (or a border between interior and exterior) at a given point and to bring that to bear if that is a relevant semantic feature for the set of spatial terms being learned.

5.2.2 Parameterized Regions

The remainder of the network is dedicated to computing parameterized regions. Recall that a parameterized region is much the same as any other region which might be learned by a perceptron, except that the lines which define the relevant half-planes are constrained to go through specific points. In this case, these are the key points of the LM.

A simple two-input perceptron unit defines a line in the x, y plane, and selects a half-plane on one side of it. Let w_x and w_y refer to the weights on the links from the x and y inputs to the perceptron unit. In general, if the unit's function is a simple threshold, the equation for such a line will be

$$xw_x + yw_y = 0, \qquad (1)$$

i.e. the net input to the perceptron unit will be

$$net_{in} = xw_x + yw_y. \qquad (2)$$

Note that this line always passes through the origin: (0,0).

If we want to force the line to pass through a particular point (x_t, y_t) in the plane, we simply shift the entire coordinate system so that the origin is now at (x_t, y_t). This is trivially done by adjusting the input values such that the net input to the unit is now

$$net_{in} = (x - x_t)w_x + (y - y_t)w_y. \qquad (3)$$

Given this, we can easily force lines to pass through the key points of an LM, as discussed above, by setting (x_t, y_t) appropriately for each key point. Once the system has learned, the regions will be parameterized by the coordinates of the key points, so that the spatial concepts will be independent of the size and position of any particular LM.

Now consider the left-hand part of the network. This accepts as input the x, y coordinates of the TR location and the LM key points, and the layer above the input layer performs the appropriate subtractions, in line with equation 3. Now each of the nodes in the layer above that is viewing the TR in a different coordinate system, shifted by the amount specified by the LM key points. Note that in the BB cluster there is one node for each corner of the LM's bounding-box, while the CoM cluster has three nodes dedicated to the LM's center of mass (and thus three lines passing through the center of mass). This results in the computation, and through weight updates, the learning, of a parameterized region.

Of course, the hidden nodes (labeled "I") that receive input from the LM interior map are also in this hidden layer. Thus, receptive fields and parameterized regions are learned together, and both may contribute to the learned semantics of each spatial term. Further details can be found in [Regier, 1990].

5.3 Implementing "Weakened" Mutual Exclusivity

Now that the basic architecture and representations have been covered, we present the means by which the evidence from implicit negative instances is weakened. It is assumed that training sets have been constructed using mutual exclusivity as a guiding principle, such that each negative instance in the training set for a given spatial term results from a positive instance for some other term.

[3] Quickprop gets its name from its ability to quickly converge on a solution. In most cases, it exhibits faster convergence than that obtained using conjugate gradient methods [Fahlman, 1990].

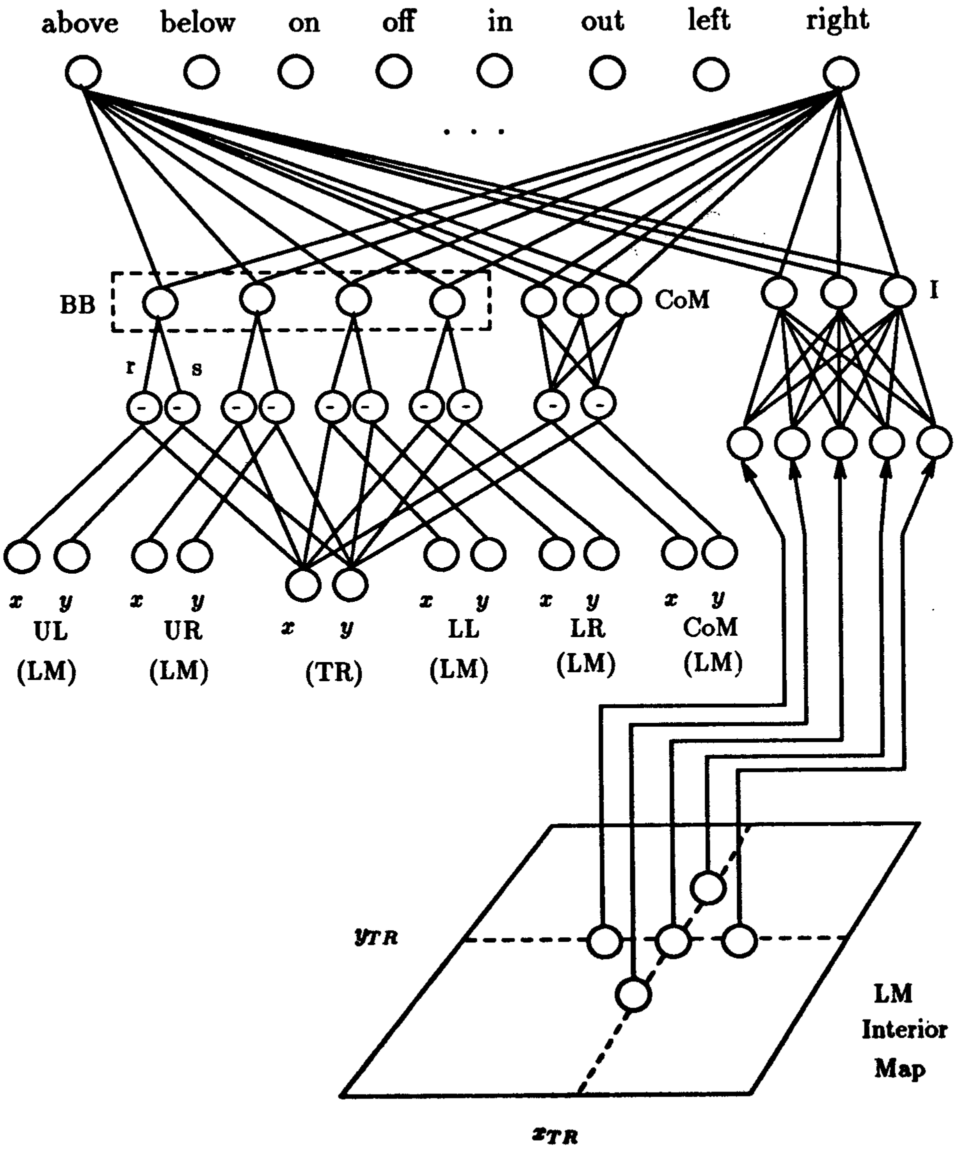

Figure 6: Network Architecture

- Evidence from implicit negative instances is weakened simply by attenuating the error caused by these implicit negatives.
- Thus, an implicit negative instance which yields an error of a given magnitude will contribute less to the weight changes in the network than will a positive instance of the same error magnitude.

This is done as follows:

Referring back to Figure 6, note that output nodes have been allocated for each of the spatial terms to be learned. For a network such as this, the usual error term in back-propagation is

$$E = \frac{1}{2}\sum_{j,p}(t_{j,p} - o_{j,p})^2 \tag{4}$$

where j indexes over output nodes, and p indexes over input patterns.

We modify this by dividing the error at each output node by some number $\beta_{j,p}$, dependent on both the node and the current input pattern.

$$E = \frac{1}{2}\sum_{j,p}(\frac{t_{j,p} - o_{j,p}}{\beta_{j,p}})^2 \tag{5}$$

The general idea is that for positive instances of some spatial term, $\beta_{j,p}$ will be 1.0, so that the error is not attenuated. For an implicit negative instance of a term, however, $\beta_{j,p}$ will be some value *Atten*, which corresponds to the amount by which the error signals from implicit negatives are to be attenuated.

Assume that we are currently viewing input pattern p, a positive instance of "above". Then the target value for the "above" node will be 1.0, while the target values for all others will be 0.0, as they are implicit negatives. Here, $\beta_{above,p} = 1.0$, and $\beta_{i,p} = Atten, \forall i \neq above$.

The value $Atten = 32.0$ was used successfully in the experiments reported here.

6 Conclusion

The system presented here learns perceptually-grounded semantics for the core senses of eight English prepositions, successfully generalizing to scenes involving landmarks to which the system had not been previously exposed. Moreover, the principle of mutual exclusivity is successfully used to allow learning without explicit negative instances, despite the false negatives in the resulting training sets.

Current research is directed at extending this work to the case of arbitrarily shaped trajectors, and to handling polysemy. Work is also being directed toward the learning of non-English spatial systems.

References

[Bowerman, 1983] Melissa Bowerman, "How Do Children Avoid Constructing an Overly General Grammar in the Absence of Feedback about What is Not a Sentence?," In *Papers and Reports on Child Language Development.* Stanford University, 1983.

[Braine, 1971] M. Braine, "On Two Types of Models of the Internalization of Grammars," In D. Slobin, editor, *The Ontogenesis of Grammar.* Academic Press, 1971.

[Fahlman, 1988] Scott Fahlman, "Faster-Learning Variations on Back Propagation: An Empirical Study," In *Proceedings of the 1988 Connectionist Models Summer School.* Morgan Kaufmann, 1988.

[Fahlman, 1990] Scott Fahlman, (personal communication), 1990.

[Feldman *et al.*, 1990] J. Feldman, G. Lakoff, A. Stolcke, and S. Weber, "Miniature Language Acquisition: A Touchstone for Cognitive Science," Technical Report TR-90-009, International Computer Science Institute, Berkeley, CA, 1990, also in the Proceedings of the 12th Annual Conference of the Cognitive Science Society, pp. 686–693.

[Johnston and Slobin, 1979] Judith Johnston and Dan Slobin, "The Development of Locative Expressions in English, Italian, Serbo-Croatian and Turkish," *Journal of Child Language*, 6:529–545, 1979.

[MacWhinney, 1989] Brian MacWhinney, "Competition and Lexical Categorization," In *Linguistic Categorization*, number 61 in Current Issues in Linguistic Theory. John Benjamins Publishing Co., Amsterdam and Philadelphia, 1989.

[Markman, 1987] Ellen M. Markman, "How Children Constrain the Possible Meanings of Words," In *Concepts and conceptual development: Ecological and intellectual factors in categorization.* Cambridge University Press, 1987.

[Minsky and Papert, 1988] Marvin Minsky and Seymour Papert, *Perceptrons (Expanded Edition)*, MIT Press, 1988.

[Pinker, 1989] Steven Pinker, *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press, 1989.

[Regier, 1990] Terry Regier, "Learning Spatial Terms Without Explicit Negative Evidence," Technical Report 57, International Computer Science Institute, Berkeley, California, November 1990.

[Rumelhart and McClelland, 1986] David Rumelhart and James McClelland, *Parallel Distributed Proccessing: Explorations in the microstructure of cognition*, MIT Press, 1986.

[Weber and Stolcke, 1990] Susan Hollbach Weber and Andreas Stolcke, "L_0: A Testbed for Miniature Language Acquisition," Technical Report TR-90-010, International Computer Science Institute, Berkeley, CA, 1990.