Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp460-465)

A cognitive model based on representations that are spatial functions

JANET AISBETT and GREG GIBBON
School of Design, Communication and Information Technology
The University of Newcastle
University Drive, Callaghan NSW 2308
AUSTRALIA

Abstract: - This paper outlines a cognitive model in which internal representations are spatial functions, and in which the associated process model is governed by distance in psychological space. Motivation for the model comes from the role of similarity judgements in human reasoning, and the apparent ability of humans to create task-dependent features about the concepts used in reasoning. Motivation also comes from the promise that neuroimages might be interpretable in terms of the conceptual tasks in which the person was engaged at the time of imaging. The creation of task-dependent features to aid problem solving is demonstrated in a categorisation task.

Key-Words: - cognitive model, cognitive representation, cortical maps, categorisation, classification

1 Introduction

Lists of numbers or alphanumeric labels are used in many disciplines to represent instances of concepts. For example, commerce uses alphanumeric fields in relational tables to describe employees, products and the like, science uses numeric vectors to describe anything from specimens to experimental conditions to emotional states of mind, and so on. In psychology, cognitive and perceptual modellers routinely use vectors to represent external stimuli and to model the internal representations of these stimuli. In Artificial Intelligence, attribute-value descriptions of things are routinely used in machine learning, as are predicates in logic-based AI.

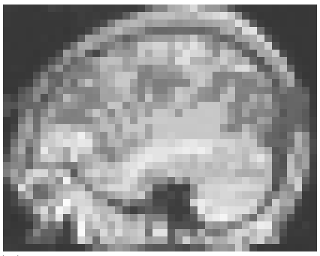
The distributed representations found in artificial neural systems use vectors for input/output, and the representations may be thought of a list of weights. However, the representations are more flexible in that there is not a fixed mapping between weights and characteristics of the things being represented. Such models are traditionally classed as biological models, although they are very simplistic models of biological neural systems in terms of both their structures and their interaction mechanisms.

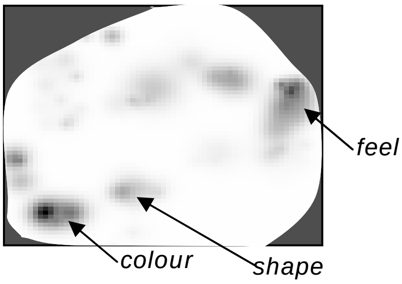
Neural population modelling takes a higher level approach to neural modelling, in attempting to explain the synchronised activity of hundreds, thousands or tens of thousands of neurons. The aggregated activity can be used to explain phenomena seen in neural images. The need to look at aggregated activity to explain even simple perceptual responses has been long recognized (eg. [9]). Patterns of electromagnetic and chemical activity captured in neuroimages such as Fig. 1a are starting to be interpreted in terms of the stimulus presented to the person at the time their brain is imaged [16]. At this stage the interpretations are crude, but as neuroimaging techniques become more sophisticated this area of research is expected to become a key to understanding cognition [13].

This begs the question of whether information that is conventionally represented in alphanumeric form might be better represented by images similar to those that might be captured in the higher resolution, more targeted, neuroimaging of the future. This question is behind our research program. Our long term goal is to determine whether the use of non-pictorial images, such as illustrated in Fig 1b, to represent abstract and concrete concepts might enable automated problem solving approaches which better mimic human approaches.

In previous work, we have developed a formal cognitive model which can be applied when mental representations are modelled using these type of images [2, 5]; and we have applied the model to simulate results of psychological experiments into recognition and categorisation (classification) tasks [5, 6]. Given the current embryonic state of neuroimaging, we have also explored how to simulate the representation of abstract and concrete concepts using analogical but non-pictorial images. This has involved experiments with human subjects [4] as well as theory and simulations [5, 6].

Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp460-465)


(a)



(b)

Fig. 1. (a) fMRI image [16]. (b) Possible use of continuous spatial functions to represent examples of concepts, with values on subsets of the spatial set being related to values of sub-concepts.

In the present paper, Section 2 looks at why a new cognitive modelling approach might be useful, Section 3 provides a fresh look at our model and Section 4 applies it to show how task-specific features can be created when solving a simple problem (in this case, categorisation).

2 A solution to problems with current representational paradigms

The dominant symbolic and neural representational paradigms in AI were criticized by Gärdenfors [9] for their limited ability to represent similarity of concepts, which he claimed was fundamental to human reasoning. Similarity, and the related construct of distance, fit naturally with vector representations.

In such representations, the vector components have a distinguished role, in that they are indecomposable, unlike attributes formed from sets of components (such as colour constituents of hue, intensity and saturation). However, the nature of the dimensions, attributes and features that are used to describe concepts is itself open to debate. Are we born with a set of atomic features (hue, saturation, length, and so on) from which all other features are built?

Braisby and Franks [8] and others have presented cases for the existence of atomic features, arguing that observed plasticity of features results from new groupings or subdivisions of groupings of features, not from the creation of atomic-level features. Vector representations fit naturally with the assumption of atomic features, as do symbolic representations. Distributed ANN representations are more accommodating of plasticity, but in the end assume a finite number of nodes, analogous to vector representations.

Arguments against the existence of atomic features include the ability to learn new categories of things, and the experimentally observed interaction between cognition and perception [14].

The spatial functions described in the next section provide a representational form that caters for both similarity and does not need to assume atomic features. Clearly the notion of magnitude must be represented by such functions. We employ what we have called generalised thermometer coding [6] to do this naturally. Conventional thermometer (analog) coding would represent the magnitudes 1 and 2 on three binary nodes, say, as shown respectively in the left and right sets of 3 circles in Fig 2, using a more redundant and robust representation than the more-familiar place based coding. Thermometer coding can be generalised using a class of continuous functions

$$S(X) \rightarrow [0, 1] \qquad (1)$$

where X might be any continuous connected subset of $\Re^n$, $n \geq 1$. There must be a mapping q from the positive reals into members of the class such that a $\leq$ b implies $q(a)(x) \leq q(b)(x)$ for all $x \in X$. Representations of two values using bars is illustrated in the middle of Fig. 2, and using the spread of Gaussians at the bottom of Fig. 2.

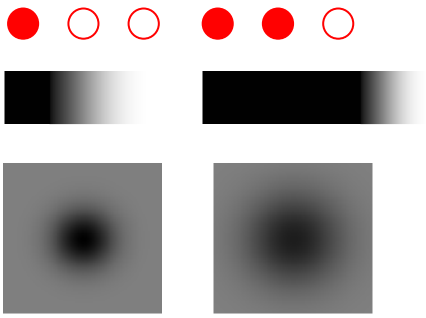


Fig 2. Three different ways of representing a smaller thing (on the left) and a larger one (on the right).

3 A representation based on spatial maps

This section sets out the main structures in the model. These are developed in [5], and are illustrated in Section 4.

3.1 Representations

Cognitive representations of stimuli, memories and so on are members of a family F of (piecewise) differentiable functions

$$f: X \rightarrow (\Re \cup \infty)^n \qquad (2)$$

where $\Re$ denotes the reals, X is a connected bounded subset of $\Re^2$, the point at infinity ∞ represents "not applicable", and n is the maximum number of components from which an object or concept may be cognitively assembled. The point at infinity can be thought of as white noise, and n as a type of capacity (or number of layers) in working memory. The use of layers enables binding of subconcepts. Following researchers such as Freeman [9] at a finer temporal granularity than we are considering here, cortical representations might be waves. If so, layers could be implemented in the brain through frequency or phase. In a computer implementation, layers can be thought of as colour layers.

We assume that external stimuli are represented by functions in F which result from sensory interpretation of those stimuli which is not considered in our cognitive modelling. Likewise, to accommodate human linguistic and motor responses, parts of X must be able to be "read off" by other subsystems which are not modelled.

3.2 Dimensions

Regions in X play the role of dimensions in conventional vector-based modelling. To support this, we allow representations to include some restrictions of members of F to connected subsets of X. That is, valid representations are functions

$$f^*: X^* \rightarrow (\Re \cup \infty)^n \qquad (3)$$

where X^* is a connected subset of X and for some f in F, $f|X^* = f^*$. Fig 3 illustrates.

3.3 Similarity

Similarity is an inverse function of a metric defined via pointwise differences. Specifically, given two instances of concepts represented by functions f and g in F, then the difference between the concepts on the region $A \subset X$ is defined to be

$$d^A(f, g) = \int_A |f(\underline{x}) - g(\underline{x})|\, d\underline{x} \qquad (4)$$

The region of integration A in Eqn. 4 is introduced to cater for the fact that human judgement of the similarity of concepts is a *contextual* judgement. The region A defines the context. Similarity is treated as the negative exponential of distance (Eqn. 5) following psychological convention.

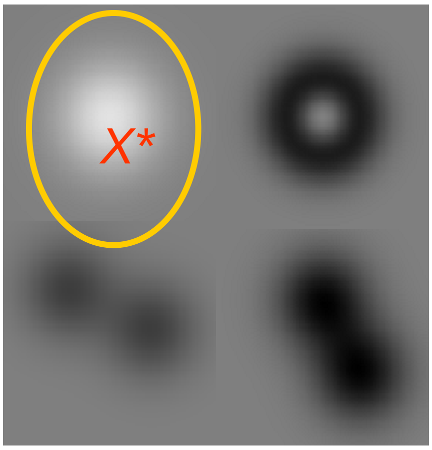


Fig 3. The restriction to a subdomain X (shown here encircled) of a concept represented as a spatial function is a subconcept that is also represented by a spatial function*

$$S(f, g; X^*) = \exp(-\int_{X^*} |f(\underline{x}) - g(\underline{x})|\, d\underline{x}) \qquad (5)$$

As pointed out in [3], such a formulation of similarity and dimensions provides a possible explanation of an experimentally observed phenomenon. In judging similarity of sets of objects that vary on more than one dimension (e.g. colour and size) people appear in some experiments to use a city block metric in which $d = \Sigma d_i$ where d_i is the metric on dimension i. However, they sometimes appear to use Euclidean or higher order Minkowski $d = \{\Sigma d_i^r\}^{1/r}$, $r > 2$ [12, 15]. This can be explained in our framework by the fact that when "dimensions" are defined on disjoint regions our metric is a city block. Yet when "dimensions" are defined as overlapping regions, the integral over the union of the regions is less than the sum of the integrals over the two regions, and so the metric appears to be a higher order Minkowski

3.4 Attention and working memory

Attention and working memory are elusive aspects of cognition, which nevertheless appear in most

Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp460-465)

comprehensive cognitive models including SOAR [11] and ACT-R [7]. The problem with explaining attention is how to do so without invoking a "higher authority" or master controller whose attention then has to be explained. At the same time, the existence of a separate working memory has been questioned, with the suggestion that working memory may be just memories activated *in situ*.

A notional layer of working memory is initially formed about the representation of a new stimulus, by identifying some memories activated by the stimulus as "belonging" to this stimulus. (This might occur in the brain via memories having the same frequency or phase as the stimulus representation.) So working memory consists of activated memories labelled somehow to refer to a stimulus presentation. The magnitude of activation is influenced by the similarity of a memory to the current stimulus in the area of attention on that layer. Activation subsides over time, but later stimuli may activate new memories on this layer. Conservation of energy is assumed, and a layer has a limited life, as working memory has finite capacity.

Attention on each layer varies, and is determined by the difference between the initial stimulus representation and the total activation over all the other layers active at that time. This effectively means that attention is directed to where the currently active memories are most different to the new stimulus. For example if, at the time a stimulus represented by f is received, there are two working memory layers with activation functions g_1 and g_2, then attention will be directed to regions in X where $|f - (g_1 + g_2)/2|$ is greater than some cut-off. With longer time available to examine a stimulus, the cut-off is lowered and regions in X with successively lower values of $|f - (g_1 + g_2)/2|$ are included in the processing.

3.5 Reasoning

The reasoning process is detailed in [5]. While the full process is tedious to expound, the overall pattern is simple and repetitive:

…input $\Rightarrow$ attention setting and memorisation $\Rightarrow$ activation of memories onto layers of working memory $\Rightarrow$ decay of activity $\Rightarrow$ input…

The input is an externally-received stimulus, or, in the absence or repression of such stimuli, the result of combining contents of working memory. Output is assumed to be obtained as total activation on certain regions of X (e.g. a region representing a linguistic symbol).

3.6 Memorisation

Memories are represented as parts of an input function. Formally, a memory of a stimulus f is $\cup_i f|_{A_i}$ where $\cup_i A_i$ is the area in attention at the time of memorization. With longer study time, the area of attention is increased and more of the stimulus representation is memorised.

4 Example

Categorisation (classification) is one of the most fundamental steps in problem solving. This section illustrates the model's performance on a simulated categorisation task in order to demonstrate the model and how regions on the image plane act as task-dependent dimensions.

Fig. 4a depicts simulated input from five stimuli in each of two categories. To make the example easier to follow, we assume the stimuli are presented to the system with sufficient study time for the entire stimulus to be remembered, including the relevant category name. Like exemplar-based [12] or most nearest-neighbour models of categorisation, therefore, we assume here that every example of the category is memorized in full.

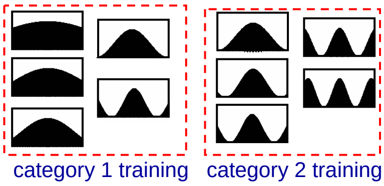
The normalised sum f_i of the activated memories is depicted in Fig. 4b for each of the two categories. These are assumed to be activated on different layers in working memory, following a verbal or visual naming of the category prior to each test stimulus. On presentation of any test stimulus, memories are activated on each layer according to their similarity to the test stimulus in the context of the area in attention on that layer. When a memory is activated, activation in the region corresponding to the relevant category name increases. The test stimulus is categorised according to the category name region with highest total activation.

Prior to the test stimulus being presented, the two layers containing the activated training memories are assumed to be the only contents of working memory. The fact that in practice recent test stimuli may affect classification performance is recognized but rarely taken into account in explaining human categorisation results. (Our model can simulate the retention of memories of recent test stimuli but we have not yet analysed the results against human experimental data.)

Suppose that category 1 had been activated first, with no other active layers; attention on the category 1 layer is therefore determined by the modulus function $|f_1|$. Activation of category 2 then creates attention on the new layer which is the difference $|f_2 - f_1|$. The attention functions are shown in Fig. 4c,

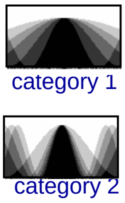
Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp460-465)

and the most prominent "dimensions" or features are isolated in Fig. 4d. The regions shown in black provide the similarity context for the categorisation task. If a longer task time were simulated, the dark grey areas in Fig. 4c might also be inspected.



(a) Training stimulus representations

(b) recalled memories



(c) Attention functions corresponding to layers in (b)

(d) Attention areas derived from (c)

Fig. 4. Stimuli are memorised as in (a), and the categories are recalled as in (b), to form layers 1 and 2 of working memory. Attention is determined by the functions (c), which are formed from the activations functions in (b). The areas in black in (c) are in attention, as shown in (d). If time to make a response was longer, then grey areas in (c) might also receive attention. See text.

Six test stimuli have representations as shown in the first row of Fig 5a. The second row in Fig 5 depicts the restrictions of each of the test stimuli to the area of attention on the category 1 layer, defined by the black regions in the image to the left in Fig 4d. The third row in Fig 5 depicts the restrictions of each of the test stimuli to the area of attention on the category 2 layer, defined by the black regions in the image to the right in Fig 4d. Activation of the memories on a layer is proportional to the similarity of their restriction to the areas shown in Fig 4d to the restrictions of the test stimulus depicted in Fig 5. Comparisons of the total activation on the region of each category name (that is, the total activation of memories stored with that category name) resulting from presentation of a stimulus shows that the first 3 test stimuli in the top row of Fig. 5 will be classified category 1, and the remainder category 2. Note that stimulus 4 appears in the training sets of both categories and is classified as category 2. This is not surprising as this category has two members which are more similar to other category 1 members than is stimulus 4.

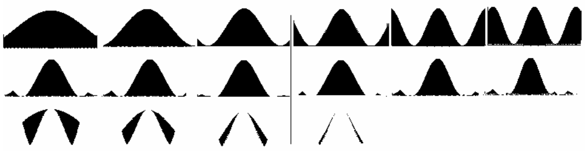


Fig. 5. Top row: representations of 6 test stimuli. Middle row: the restriction of the representations to the area in attention on layer 1. Bottom row: the restriction of the representations to the area in attention on layer 2. The 5th and 6th stimuli restrict to the zero function in the attention area on layer 2.

As a final example of feature formation, consider the two training sets depicted in Fig. 6a. In category 2 there are now "whiskers" of various lengths attached to each of the stimulus representations. These are clearly seen in the accumulated memories depicted in Fig. 6b. The difference function that will direct attention on layer 2 when the accumulated representations of both sets are in working memory is depicted in Fig 6c. This time, the "whiskers" are the prominent region, and these form the distinguishing feature.

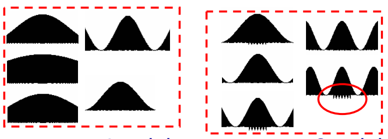
5 Conclusion

We have used a categorisation (classification) task to demonstrate a cognitive model in which representations are essentially real-valued functions on a subset of the plane. The ability to create task dependent features based on connected subsets of the spatial domain contrasts with the finite number of fixed dimensions available in conventional alphanumeric vector modelling.

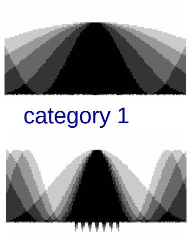
The description of the example given in Section 4 is reminiscent of image classification, although the metric on our space of functions is not one of the usual image metrics, and classification does not proceed by the typical nearest-neighbour, maximum likelihood or distance-to-mean classification approaches. Instead the model uses constructs of working memory and attention. Moreover, the spatial functions we consider are not pictorial, although they will be analogical, in the sense suggested by generalised thermometer coding.

Spatial representations are well established in contexts such as image and geographical systems, and a cognitive process model based on working memory and attention is fairly standard. However, a computational model which uses spatial representations is new. The potential link to cortical imaging and population coding is also interesting. A computational model based on spatial representations is likely to be a useful cognitive modelling tool.

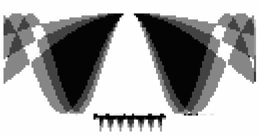
In the introduction, we said we were aiming to investigate if such a model could provide more human-like reasoning. This requires us to try to model experimental results from human decision-making. In current work we are gathering more data from psychological experiments into two-choice decisions (such as recognition, or 2-class categorisations). Because human performance in such tasks is very dependent on the time available, such data forces us to look closely at the dynamics of the representations. We hope to show how cognitive models based on spatial functions can explain further experimental phenomena.



(a) Training stimulus representations; note "whiskers" beneath category 2 members eg as circled.



b) recalled memories



(c) Attention mage on category 2 layer. The "whiskers" are the most prominent feature.

Fig. 6. Creation of a new task dependent feature.

References

[1] Abbott, L. and Dayan, P. The effect of correlated variability on the accuracy of a population code, *Neural Computation*, 1999 Vol. 11, No. 1, pp. 91-101.

[2] Aisbett, J. and Gibbon G. A general formulation of conceptual spaces as a meso-level representation *J. Art. Intelligence*, 2001, Vol. 133, pp. 189 – 232.

[3] Aisbett J. and Gibbon G. Memory and Learning in a Meso Level Reasoning System. *Ninth International Conference on Neural Information Processing*, Singapore Nov 2002. pp. 101-105.

[4] Aisbett J. and Gibbon G. Preserving similarity in representation: a scheme based on images *Proc. Joint Int. Conf. on Cog. Sci.* Sydney 2003.

[5] Aisbett J. and Gibbon G. A cognitive model in which representations are images, *Cognitive Systems Research* 2005, Vol.6, No.4, pp.333-363

[6] Aisbett, J. Townsend J. and Gibbon G. Modelling perceptual discrimination. *8th International Work-Conference on Artificial Neural Networks* 2005 pp. 646-653.

[7] Anderson, J R, Bothell, Byrne, M. D., and Lebiere, C. An integrated theory of the mind. *Psychological Review* 2002.

[8] Braisby, N. and Franks, B.. A creationist myth: Pragmatic combination not feature creation, *Beh. and Brain Sciences*, 2001 Vol. 21, No 1, p19.

[9] Freeman, W. Qualitative Overview of Population Neurodynamics *Neural modeling and Neural Networks*, 1994, 185-215.

[10] Gardenfors, P. *Conceptual Spaces:The Geometry of Thought* MIT Press 2000.

[11] Laird, J., Newell, A. and Rosenbloom, P. Soar: An architecture for general intelligence. *Artificial Intelligence*, 1987 Vol. 33, pp. 1-64.

[12] Nosofsky, R Attention, similarity and the identification-categorization relationship. Journal of Experimental Psychology: General, 1986, Vol.115, No.1, pp.39-57.

[13] Poldrack, R. A. Imaging Brain Plasticity: Conceptual and Methodological Issues— A Theoretical Review, *NeuroImage*, 2000, Vol.12, No. 1, pp.1-13.

[14] Schyns P., Goldstone, R. and Thibaut, J-P. The development of features in object concepts *Beh. and Brain Science*, 1998, Vol. 21, pp. 1-54.

[15] Shepard R. Metric structures in ordinal data. *J. Math. Psychology*, 1966, Vol. 3, pp. 287-315.

[16] Wang, X., Hutchinson, R. and Mitchell, T. Training fMRI Classifiers to Detect Cognitive States across Multiple Human Subjects *Neural Information Processing Systems* 2003.