Sparse neural representation for semantic indexing

Peter Földiák

School of Psychology, University of St Andrews, St Andrews KY16 9JP, Scotland, U.K.

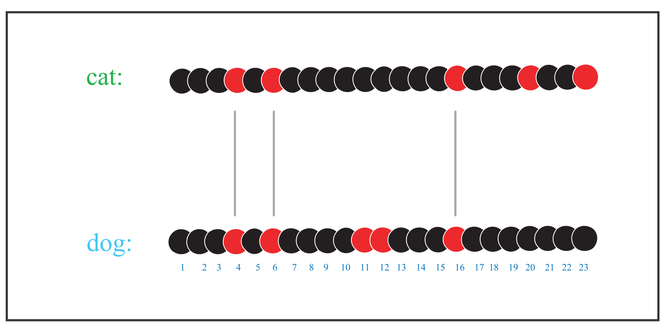
Peter.Foldiak@st-andrews.ac.uk http://psy.st-and.ac.uk/foldiak

Abstract

1. Traditional models think of concepts as nodes and associations as links between nodes. Distributed neural representations suggest a model of relationships between concepts where explicit links are not even necessary. Similarity and multiple inheritance can be expressed as vector similarity. High dimensional multiple categorisation is implicit in the neural representation itself.
2. Sparse representations (low fraction of active neurons) are found at different levels of the sensory system, and can be assumed to be present in semantic representations. It is efficient to consider a sparse activity vector as a mathematical set: the set of active units.
3. Semantic relationships can be represented implicitly by different kinds of overlap between these ***sets*** of 'features', rather than with direct links. Set algebra can be used to combine representations for expressing and indexing new concepts or items. The advantages of neural representations can be rescued for practical applications where neural network learning would be currently unfeasible.
4. Efficient and precise ***'concept' search*** can be based on sets.

Most of today's catalogues and directories have a ***tree*** structure where each item is assigned to a single category corresponding to a branch of the tree. Traditional cognitive models of the relationships between concepts are usually also captured in ***graphs*** (often also trees, e.g. semantic nets), where the nodes are the concepts and the links express relationships. The rich and multifaceted structure of relationships between concepts is not captured well by these models. This results in poor retrieval performance even in computerised systems. Distributed neural representations studied in both artificial and biological neural networks, however, suggest a different model.

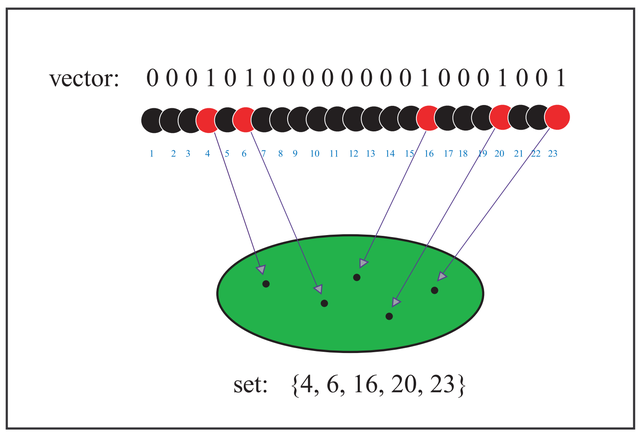
Neural representations consist of a large set of neurons, only some of which are active when representing an item. For each distinguishable item, a different pattern of activity can be observed. This pattern can be described by a ***vector***, whose components correspond to the activities of the neurons. In a high-level, semantic neural representation, two similar items would be represented as two similar, i.e. highly overlapping patterns of activities. But neural representations encode more than just the amount of similarity between items. As each neuron, or subset of neurons, encodes one particular aspect or feature of the represented item, the pattern of activity corresponding to two items also reveals the ***nature*** of the relationship between the items. Note that the relationships between items are represented implicitly in the relationships of the patterns of activity, and need not be made as explicit links as in the graph models.



Representations for "cat" and "dog". The amount of similarity is represented by amount of overlap (3 of the 5 active units). The common active units (4, 6, 16) correspond to "pet", and describe the nature of the relationship.

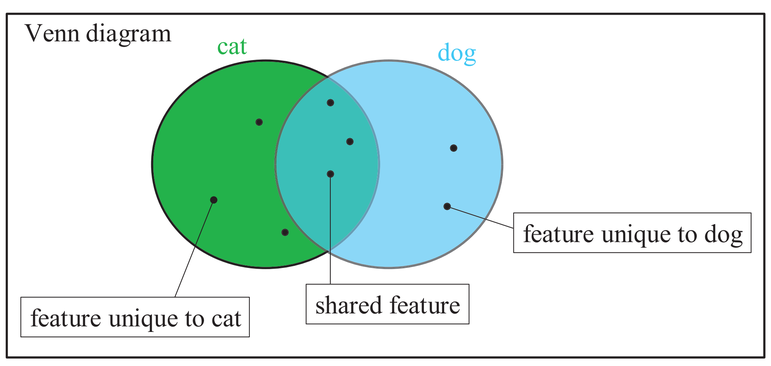
An important property of neural representation is ***sparseness***, i.e. the average fraction of active neurons. Neural representations in the sensory system have been found to be sparse experimentally, and theoretical models explicitly maximising sparseness result in reasonable fits to data from primary visual cortex.

Assuming that the neural activity is binary (inactive/active or 0/1), the activity vectors can be described equivalently a ***set*** of active units. If the representation is sparse, i.e. the fraction of active units is small, this is also an efficient (short) description of the vector. (This is also known as the sparse representation of a vector).



A binary activity vector can be considered equivalently as a ***set*** of active units. If units (or subsets of them) stand for features, this is equivalent to a representation of an item by a set of features (or 'microfeatures').

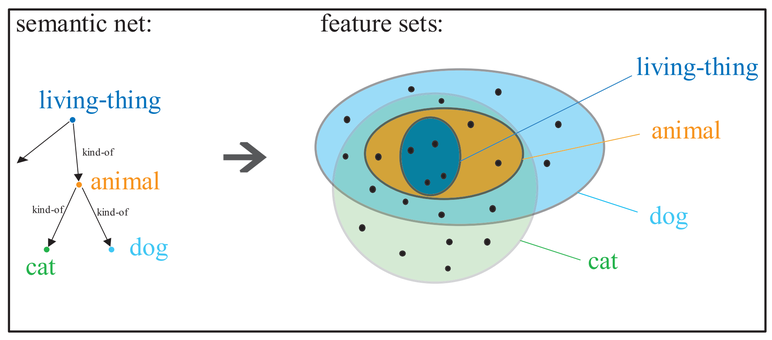
If items are represented as mathematical sets, their relationships can be described as overlapping sets and visualised with Venn diagrams. Venn diagrams show sets as closed curves and their elements as items inside the curves graphically.



Venn diagrams show the relationships between feature sets.

Inheritance ('kind-of') relationships between items map into subset relationships between feature sets. The more specific concept has all the features of the more general concept, in addition to the unique features specific to it. The more general item is therefore represented as a subset in the feature-set representation. Further generalisation corresponds to nested subsets. Any tree hierarchy (such as a semantic net) can therefore be converted to a set of feature-sets. The features need not have names themselves (i.e. they can be 'microfeatures'); they only impose relationships between named concepts that contain them.

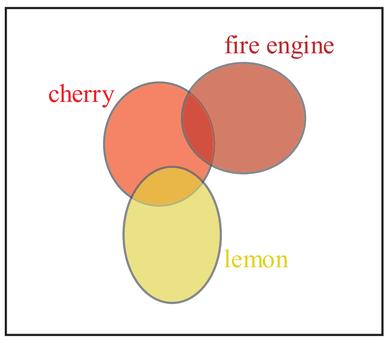
Hierarchical inheritance relationships can be mapped into feature sets. "animal" is a subset of the "dog" feature set, "living-thing" is a further subset of "animal".



Note the duality between the 'sets of things' and the 'set of features' representations. In the traditional 'sets of things' model each concept is a set of the things that belong to that concept, so 'animal' is a subset of 'living-thing' (as only some living things are animals). In the 'sets of features' model, the most general concept corresponds to the smallest set, with each specialisation adding more features to form supersets.

One of the drawbacks of semantic nets is the difficulty of inference. As representing all possible (combinatorially large number of) associations explicitly is impractical, 'spreading activation' was proposed to work out indirect relationships. The assumption was that association is transitive, so "if assoc(A,B) and assoc(B,C) then assoc(A,C)". While this is sometimes true (e.g. "assoc(tomato,cherry) and assoc(cherry,fire engine) then assoc(tomato, fire engine)" - because they are all red), it is not generally true: assoc(cherry, fire engine) and assoc(cherry, lemon) but NOT assoc(fire engine, lemon). This latter is not transitive because the association between cherry and fire engine is along a different ***aspect*** (colour) than that between cherry and lemon (function). Spreading activation fails in such cases.

Feature sets deal with transitivity appropriately. They represent not just association, but also the nature of the association. The association between cherry and fire engine, and that between cherry and lemon, correctly, does ***not*** imply that lemon and fire engine are associated. The overlap between the first pair involves colour-related features, while that between the second pair function-related features.



If concepts are represented as sets, set operations (set algebra) can be used to combine concepts to
- define new concepts,
- express queries.

operations with sets

- e.g., A = {e, f, g}; B = {g, h, i, j}
- A and B are sets; e, f, g, h, i, j are elements
- set union: $A \cup B = \{e, f, g, h, i, j\}$
- set intersection: $A \cap B = \{g\}$
- set difference: $A \setminus B = \{e, f\}$
- set expressions: e.g., $(A \cap B) \setminus (C \cup D)$

Oversimplified example:

- definitions:

```
thing := !
animal := thing ∪ !
...
penguin := (bird \ fly) ∪
                (antarctic \ place)
```

- query:

```
cold ∪ animal
    → penguin
```

Comments:
- "!" is a set with a single newly generated feature
- animal is a set of two features (one in thing, one new)

- penguin contains the bird features except the flying related features, plus the antarctic specific features (but not place features in antarctic, as penguin is not a place)

- question: "What is an animal related to cold?"
- answer: penguin !

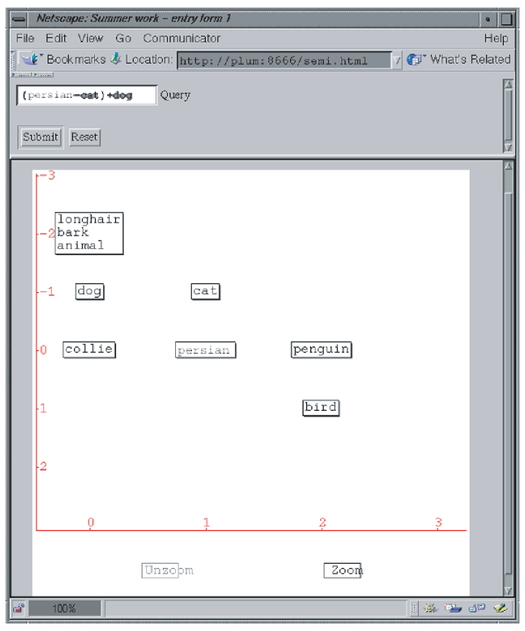
The query performs a search for the smallest stored sets of maximal overlap with the query set efficiently.

The result can be displayed graphically in 2D. The best match is close to origin of coordinate system (0,0). Vertical axis is generality/specificity axis (more general concepts above, more specific below). Horizontal axis is dissimilarity (more similar is placed closer to the left).

Example shows ***analogy***:
"What is the equivalent of Persian in dogs?"

query:
```
persian \ cat ∪ dog
```

answer:
```
Collie
```



Acknowledgements

Peter Földiák was supported by University of St Andrews, the Albert Szent-Györgyi Fellowship (Hungary), and a grant from the N.S.F. (U.S.A.)