An Application of Formal Concept Analysis to Neural Decoding

Dominik Endres[1], Peter Földiák[1], and Uta Priss[2]

[1]School of Psychology, University of St. Andrews, {dme2,pf2}@st-andrews.ac.uk
[2]School of Computing, Napier University, u.priss@napier.ac.uk

Abstract. This paper proposes a novel application of Formal Concept Analysis (FCA) to neural decoding: the semantic relationships between the neural representations of large sets of stimuli are explored using concept lattices. In particular, the effects of neural code sparsity are modelled using the lattices. An exact Bayesian approach is employed to construct the formal context needed by FCA. This method is explained using an example of neurophysiological data from the high-level visual cortical area STSa. Prominent features of the resulting concept lattices are discussed, including indications for a product-of-experts code in real neurons.

1 Introduction

Mammalian brains consist of billions of neurons, each capable of independent electrical activity. From an information-theoretic perspective, the patterns of activation of these neurons can be understood as the codewords comprising the neural code. The neural code describes which pattern of activity corresponds to what information item. We are interested in the (high-level) visual system, where such items may indicate the presence of a stimulus object or the value of some stimulus attribute, assuming that each time this item is represented the neural activity pattern will be the same or at least similar. *Neural decoding* is the attempt to reconstruct the stimulus from the observed pattern of activation in a given population of neurons [1,2,3,4]. Popular decoding quality measures, such as Fisher's linear discriminant [5] or mutual information [6] capture how accurately a stimulus can be determined from a neural activity pattern (e.g., [4]). While these measures are certainly useful, they provide little information about the structure of the neural code, which is what we are concerned with here. Furthermore, we would also like to elucidate how this structure relates to the represented information items, i.e. we are interested in the semantic aspects of the neural code.

To explore the relationship between the representations of related items, Földiák [7] demonstrated that a sparse neural code can be interpreted as a graph (a kind of "semantic net"). Each codeword can then be represented as a set of active units (a subset of all units). The codewords can now be partially ordered under set inclusion: codeword $A \leq$ codeword B iff the set of active neurons

of A is a subset of the active neurons of B. This ordering relation is capable of capturing semantic relationships between the represented information items. There is a duality between the information items and the sets representing them: a more general class corresponds to a smaller subset of active neurons, and more specific items are represented by larger sets [7]. Additionally, storing codewords as sets is especially efficient for sparse codes, i.e. codes with a low activity ratio (i.e. few active units in each codeword). These findings by Foldiak [7] did not employ the terminology and tools of Formal Concept Analysis (FCA) [8,9]. But because this duality is a Galois connection, it is of interest to apply FCA to such data. The resulting concept lattices are an interesting representation of the relationships implicit in the code.

We would also like to be able to represent how the relationship between sets of active neurons translates into the corresponding relationship between the encoded stimuli. In our application, the stimuli are the formal objects, and the neurons are the formal attributes. The FCA approach exploits the duality of extensional and intensional descriptions and allows to visually explore the data in lattice diagrams. FCA has shown to be useful for data exploration and knowledge discovery in numerous applications in a variety of fields [10,11].

This paper does not include an introduction to FCA because FCA is well described in the literature (e.g., [9]). We use the phrase *reduced labelling* to refer to line diagrams of concept lattices which have labels only attached to object concepts and attribute concepts. As a reminder, an object concept is the smallest (w.r.t. the conceptual ordering in a concept lattice) concept of whose extent the object is a member. Analogously, an attribute concept is the largest concept of whose intent the attribute is a member. *Full labelling* refers to line diagrams of concept lattices where concepts are depicted with their full extent and intent.

We provide more details on sparse coding in section 2 and demonstrate how the sparseness (or denseness) of the neural code affects the structure of the concept lattice in section 3. Section 4 describes the generative classifier model which we use to build the formal context from the responses of neurons in the high-level visual cortex of monkeys. Finally, we discuss the concept lattices so obtained in section 5.

2 Sparse coding

One feature of neural codes which has attracted a considerable amount of interest is its *sparseness*. As detailed in [12], sparse coding is to be distinguished from local and dense distributed coding. At one extreme of low average activity ratio are local codes, in which each item is represented by a separate neuron or a small set of neurons. This way there is no overlap between the representations of any two items in the sense that no neuron takes part in the representation of more than one item. An analogy might be the coding of characters on a computer keyboard (without the Shift and Control keys), where each key encodes a single character. It should be noted that locality of coding does not necessarily imply that only one neuron encodes an item, it only says that the neurons are highly

selective, corresponding to single significant items of the environment (e.g. a "grandmother cell" - a hypothetical neuron that has the exact and only purpose to be activated when someone sees, hears or thinks about their grandmother).

The other extreme (approximate average activity ratio of 0.5) corresponds to dense, or *holographic* coding. Here, an information item is represented by the combination of activities of all neurons. For N binary neurons this implies a representational capacity of 2^N. Given the billions of neurons in a human brain, 2^N is beyond astronomical. As the number of neurons in the brain (or even just in a single cortical area, such as primary visual cortex) is substantially higher than the number of receptor cells (e.g. in the retina), the representational capacity of a dense code in the brain is much greater than what we can experience in a lifetime (the factor of the number of moments in a lifetime adds the requirement of only about 40 extra neurons). Therefore the greatest part of this capacity is redundant.

Sparse codes (small average activity ratio) are a favourable compromise between dense and local codes ([13]). The small representational capacity of local codes can be remedied with a modest fraction of active units per pattern because representational capacity grows exponentially with average activity ratio (for small average activity ratios). Thus, distinct items are much less likely to interfere when represented simultaneously. Furthermore, it is much more likely that a single layer network can learn to generate a target output if the input has a sparse representation. This is due to the higher proportion of mappings being implementable by a linear discriminant function. Learning in single layer networks is therefore simpler, faster and substantially more plausible in terms of biological implementation. By controlling sparseness, the amount of redundancy necessary for fault tolerance can be chosen. With the right choice of code, a relatively small amount of redundancy can lead to highly fault-tolerant decoding. For instance, the failure of a small number of units may not make the representation undecodable. Moreover, a sparse distributed code can represent values at higher accuracy than a local code. Such distributed coding is often referred to as coarse coding.

3 Concept lattices of local, sparse and dense codes

In the case of a binary neural code, the sparseness of a codeword is inversely related to the fraction of active neurons. The average sparseness across all codewords is the sparseness of the code [13,12]. Sparse codes, i.e. codes where this fraction is low, are found interesting for a variety of reasons: they offer a good compromise between encoding capacity, ease of decoding and robustness [14]; they seem to be employed in the mammalian visual processing system [15]; and they are well suited to representing the visual environment we live in [16,17]. It is also possible to define sparseness for graded or even continuous-valued responses (see e.g. [18,4,12]). To study what structural effects different levels of sparseness would have on a neural code, we generated random codes, i.e. each of 10 stimuli was associated with randomly drawn responses of 10 neurons, subject to

the constraints that the code be perfectly decodable and that the sparseness of each codeword was equal to the sparseness of the code. Fig.1 shows the contexts (represented as cross-tables) and the concept lattices of a local code (activity ratio 0.1), a sparse code (activity ratio 0.2) and a dense code (activity ratio 0.5). In a local code, the response patters to different stimuli have no overlapping activations, hence the lattice representing this code is an anti-chain with top and bottom element added. Each concept in the anti-chain introduces (at least) one stimulus and (at least) one neuron. In contrast, a dense code results in a larger number of concepts which introduce neither a stimulus nor a neuron. The lattice of the dense code also contains substantially longer chains between the top and bottom nodes than in the case of sparse and local codes.

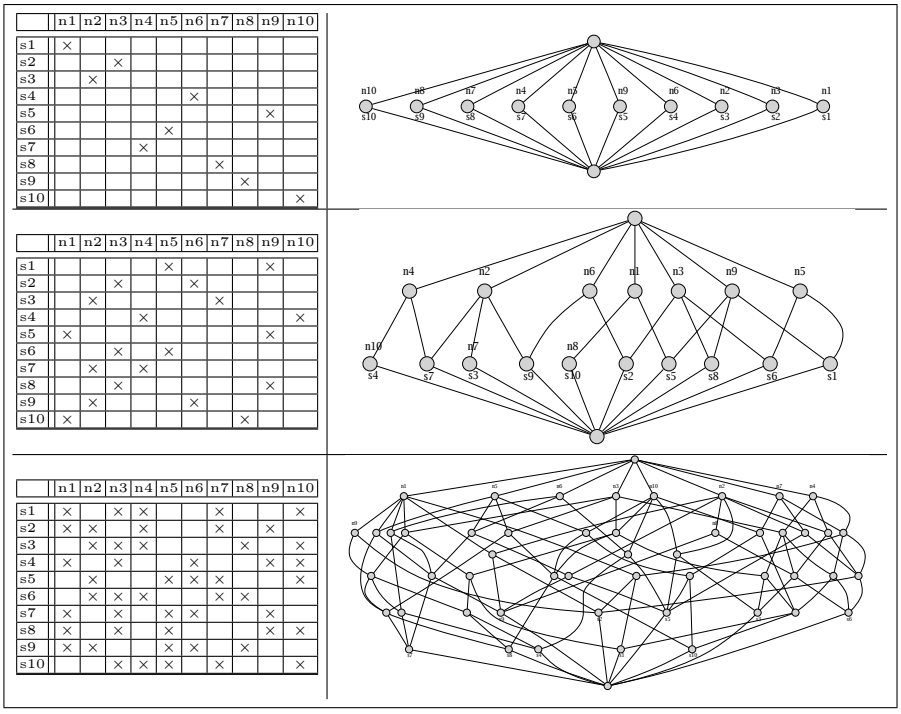

Fig. 1. Contexts (represented as cross-tables) and concept lattices for a local, sparse and dense random neural code. Each context was built out of the responses of 10 (hypothetical) neurons (n1, ..., n10) to 10 stimuli (s1, ..., s10). Each node represents a concept.

The most obvious differences between the lattices is the total number of concepts. A dense code, even for a small number of stimuli, will give rise to a large number of concepts, because the neuron sets representing the stimuli are very

probably going to have non-empty intersections. These intersections are potentially the intents of concepts which are larger than those concepts that introduce the stimuli. Hence, the latter are found towards the bottom of the lattice. This implies that they have large intents, which is of course a consequence of the density of the code. Determining these intents thus requires the observation of a large number of neurons, which is unappealing from a decoding perspective. The local code does not have this drawback, but is hampered by a small encoding capacity (maximal number of concepts with non-empty extents): the concept lattice in fig.1 is the largest one which can be constructed for a local code comprised of 10 binary neurons. Which of the above structures is most appropriate depends on the conceptual structure of the environment to be encoded and the appropriate sparseness that can be selected.

4 Building a formal context from responses of high-level visual neurons

To explore whether FCA is a suitable tool for interpreting real neural codes, we constructed formal contexts from the responses of high-level visual cortical cells in area STSa (part of the temporal lobe) of monkeys. Characterising the responses of these cells is a difficult task. They exhibit complex nonlinearities and invariances which make it impossible to apply linear techniques, such as reverse correlation [19], that were shown to be useful in understanding the responses of neurons in early visual areas [20,21]. The concept lattices obtained by FCA might enable us to display and browse these invariances: if the response of a subset of cells indicates the presence of an invariant feature in a stimulus, then all stimuli having this feature should form the extent of a concept whose intent is given by the responding cells.

4.1 Physiological data

The data were obtained through [22], where the experimental details can be found. Briefly, spike trains were obtained from single neurons within the upper and lower banks of the superior temporal sulcus (STSa) of an awake and behaving monkey (*Macaca mulatta*) via standard extracellular recording techniques [23]. During the recordings, the monkey had to perform a fixation task. This area contains cells which are responsive to faces. Extracellular voltage fluctuations were measured, and the stereotypical action potentials (i.e. 'spikes') of the neuron were detected and their temporal sequence was recorded resulting in a 'spike train'. These spike trains were turned into distinct samples, each of which contained the spikes from -300 ms before to 600 ms after the stimulus onset with a temporal resolution of 1 ms. The stimulus set consisted of 1704 images, containing colour and black and white views of human and monkey head and body, animals, fruits, natural outdoor scenes, abstract drawings and cartoons. Stimuli were presented for 55 ms each without inter-stimulus gaps in random

sequences. While this rapid serial visual presentation (RSVP) paradigm complicates the task of extracting stimulus-related information from the spike trains, it has the advantage of allowing for the testing of a large number of stimuli. A given cell was tested on a subset of 600 or 1200 of these stimuli, each stimulus was presented between 1-15 times.

The data were previously analysed with respect to the stimulus selectivity of individual cells only. Previous neural population decoding studies were aimed at identifying stimulus lables (e.g. [2,3]) only. This paper presents the first analysis of the semantic structure of neural data with FCA.

4.2 Bayesian thresholding

In order to apply FCA, we extracted a binary attribute from the raw spike trains. We could use many-valued attributes to describe the neural response, but we will employ a simple binary thresholding as a starting point. This binary attribute should be as informative about the stimulus as possible, to allow for the construction of meaningful concepts. We do this by Bayesian thresholding, as detailed below. This procedure also avails us of a null hypothesis H_0 ="the responses contain no information about the stimuli".

A standard way of obtaining binary responses from neurons is thresholding the spike count within a certain time window. This is a relatively straightforward task, if the stimuli are presented well separated in time and a large number of trials per stimulus are available. Then latencies and response offsets are often clearly discernible and thus choosing the time window is not too difficult. However, under RSVP conditions with few trial per stimulus, response separation becomes more tricky, as the responses to subsequent stimuli will tend to follow each other without an intermediate return to baseline activity. Moreover, neural responses tend to be rather noisy. We will therefore employ a simplified version of the generative Bayesian Bin classification algorithm (BBCa) [24], which was shown to perform well on RSVP data [25].

BBCa was designed for the purpose of inferring stimulus labels g from a continuous-valued, scalar measure z of a neural response. The range of z is divided into a number of contiguous bins. Within each bin, the observation model for the g is a Bernoulli scheme with G types and with a Dirichlet prior over its parameters. It is shown in [24] that one can iterate/integrate over all possible bin boundary configurations efficiently, thus making exact Bayesian inference feasible. Moreover, the marginal likelihood (or model evidence) becomes thus available, which can be used to infer the posterior distribution over all spike counting windows. We make two simplifications to BBCa: 1) z is discrete, because we are counting spikes and 2) we use models with only 1 bin boundary Z_0 in the range r of z, i.e.

$$P(g = l_i | z = z_i) = \begin{cases} p_{l_i} \text{ if } z_i \leq Z_0 \\ q_{l_i} \text{ otherwise} \end{cases} \tag{1}$$

$$\sum_g p_g = 1 \ , \ \sum_g q_g = 1 \tag{2}$$

$$p(p_0, \ldots, p_G) = \frac{\Gamma\left(\sum_g \alpha_g\right)}{\prod_g \Gamma(\alpha_g)} \prod_g p_g^{\alpha_g - 1} \tag{3}$$

$$p(q_0, \ldots, q_G) = \frac{\Gamma\left(\sum_g \beta_g\right)}{\prod_g \Gamma(\beta_g)} \prod_g q_g^{\beta_g - 1} \tag{4}$$

$$p(Z_0) = \frac{1}{|r|}. \tag{5}$$

We have no a priori preferences for any stimulus label, thus we choose $\forall g : \alpha_g = \beta_g = 1$. Since the Dirichlet priors on the p_g and q_g are conjugate to the likelihood of the data (eqn.(1)), the posteriors can be computed in closed form. Further details of the posterior computation after observing a set of stimulus-response pairs (l_i, z_i) are analogous to [24].

The bin membership (higher bin = stimulus has attribute) of a given neural response can then serve as the binary attribute required for FCA, since BBCa weighs bin configurations by their classification (i.e. stimulus label decoding) performance. We proceed in a straight Bayesian fashion: since the bin membership is the only variable we are interested in, all other parameters (counting window size and position, class membership probabilities, bin boundaries) are marginalised. This minimises the risk of spurious results due to "contrived" information (i.e. choices of parameters) made at some stage of the inference process. Afterwards, the probability that the response belongs to the upper bin is thresholded at a probability of 0.5, i.e. if the probability is larger than 0.5, then there will be a cross in the context. Instead of this simple binarisation, other methods of conceptual scaling could be used.

Since BBCa yields exact model evidences, it can also be used for model comparison. Running the algorithm with no bin boundaries in the range of z effectively yields the probability of the data given the "null hypothesis" H_0: z does not contain any information about g. We can then compare it against the alternative hypothesis described above (i.e. the information which bin z is in tells us something about g) to determine whether the cell has responded at all.

4.3 Cell selection

The experimental data consisted of recordings from 26 cells. To minimise the risk that the computed neural responses were a result of random fluctuations, we excluded a cell if 1) H_0 was more probable than 10^{-6} or 2) the posterior standard deviations of the counting window parameters were larger than 20 ms, indicating large uncertainties about the response timing. Cells which did not respond above the threshold included all cells excluded by the above criteria (except one). Furthermore, since not all cells were tested on all stimuli, we also had to select tuples of subsets of cells and stimuli such that all cells in a tuple

were tested on all stimuli. Incidentally, this selection can also be accomplished with FCA, by determining the concepts of a context with gIm ="stimulus g was tested on cell m" and selecting those with a large number of stimuli $\times$ number of cells. One of these cell and stimulus subset pairs (16 cells, 310 stimuli) was selected for further exemplary analysis, but the lattices computed from the other subset pairs displayed similar features.

5 Results

To analyse the neural code, the thresholded neural responses were used to build stimulus-by-cell-response contexts. We performed FCA on these with ColibriConcepts[1], created stimulus image montages[2] and plotted the lattices[3]. In these graphs, the images represent the formal objects. The top of the frame around each concept image contains the concept number and the list of cells in the intent (which, unfortunately, may be difficult to see in the printed version of the graphs. Moreover, the list is truncated if more than 6 cells are in the intent.).

Fig.2 shows a lattice which has an emphasis on "face" and "head" concepts. The concepts introducing human and cartoon faces (i.e. with extents consisting of general "face" images) tend to be higher up in the lattice and their intents tend to be small. In contrast, the lower concepts introduce mostly single monkey faces (and faces of the monkey's caregivers), with the bottom concepts having intents of ≥ 7 cells. We may interpret this as an indication that the neural code has a higher "resolution" for faces of conspecifics (and other "important" faces) than for faces in general, i.e. other monkeys are represented in greater detail in a monkey's brain than humans or cartoons. This feature can be observed in most lattices we generated. Thus, monkey STSa cells are not just responsive to faces in general, but to specific subclasses, such as monkey faces, in particular.

Fig.3 shows a subgraph from a lattice with full labelling. Full labelling is of interest in these applications because viewing the full extent simultaneously gives an impression of "what this concept is about". The concepts in the left half of the graph are face concepts, whereas the extents of the concepts in the right half also contain a number of non-face stimuli. Most of the latter have something "roundish" about them. The bottom concept, being subordinate to both the "round" and the "face" concepts, contains a stimulus with both characteristics, which points towards a product-of-experts (PoE) encoding [26]. In PoE, each 'expert' can be thought of as an attribute (or attribute combination) of the represented item. These experts are expected to correspond to meaningful aspects of the information items. Several examples of this kind can be found in the other graphs of the complete concept lattices, which cannot be included in this paper.

[1] available at http://code.google.com/p/colibri-concepts/
[2] via ImageMagick, available at http://www.imagemagick.org
[3] with Graphviz, available at http://www.graphviz.org

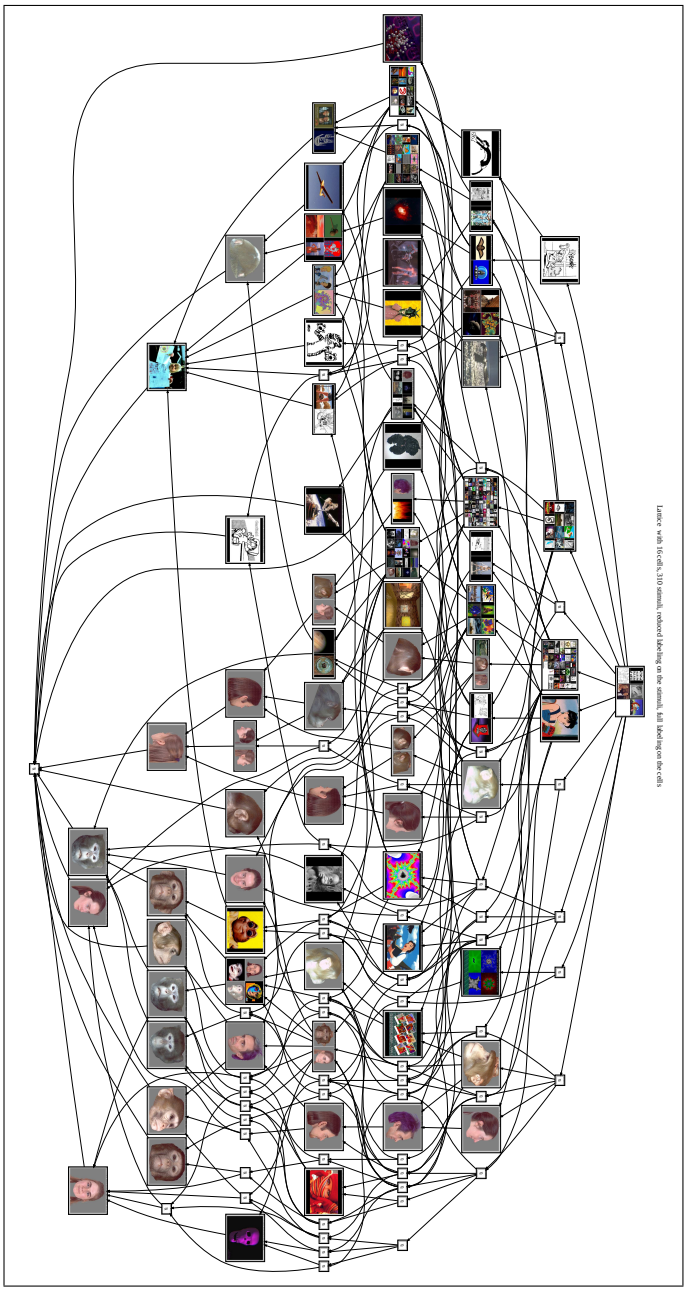

Fig. 2. A lattice with reduced labelling on the stimuli, i.e. stimuli are only shown in their object concepts. The $\emptyset$ indicates that an extent is the intersection of the parent concept extents, i.e. no new stimuli were introduced by this concept.

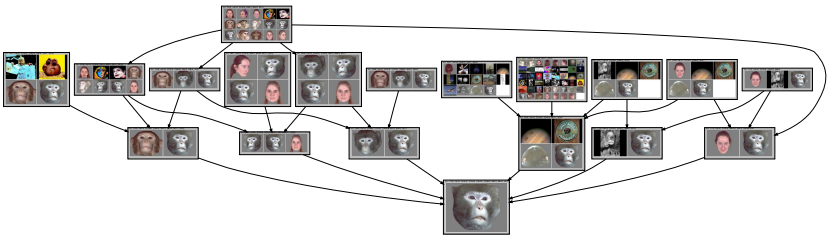

Fig. 3. A subgraph of a lattice with full labelling. The concepts on the right side are not exclusively "face" concepts, but most members of their extents have something "roundish" about them.

6 Conclusion

We demonstrated the potential usefulness of FCA for the exploration and interpretation of neural codes. This technique is feasible even for high-level visual codes, where linear decoding methods [20,21] fail, and it provides qualitative information about the structure of the code which goes beyond stimulus label decoding [1,2,3,4]. The semantic structure of neural data has previously been analysed with tree-based clustering methods [27]. Imposing a tree structure on the data may be inappropriate for neural data that reflects a more general semantic structure, as supported by our results.

Clearly, however, our application of FCA for this analysis is still in its infancy. It would be very interesting to repeat the analysis presented here on data obtained from simultaneous multi-cell recordings, to elucidate whether the conceptual structures derived by FCA are used for decoding by real brains. On a larger scale than single neurons, FCA could also be employed to study the relationships in fMRI data [28].

Acknowledgements D. Endres was supported by MRC fellowship G0501319.

References

1. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. Science **233**(4771) (1986) 1416–1419
2. Földiák, P.: The 'Ideal Homunculus': statistical inference from neural population responses. In Eeckmann, F., Bower, J., eds.: Computation and Neural Systems. Kluwer Academic Publishers, Norwell, MA (1993) 55–60
3. Oram, M., Földiák, P., Perrett, D., Sengpiel, F.: The 'Ideal Homunculus': decoding neural population signals. Trends In Neurosciences **21** (June 1998) 259–265
4. Quiroga, R.Q., Reddy, L., Koch, C., Fried, I.: Decoding Visual Inputs From Multiple Neurons in the Human Temporal Lobe. J Neurophysiol **98**(4) (2007) 1997–2007
5. Duda, O., Hart, P., Stork, D.: Pattern classification. John Wiley & Sons, New York, Chichester (2001)

6. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, New York (1991)
7. Földiák, P.: Sparse neural representation for semantic indexing. In: XIII Conference of the European Society of Cognitive Psychology (ESCOP-2003). (2003) http://www.st-andrews.ac.uk/∼pf2/escopill2.pdf.
8. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In Rival, I., ed.: Ordered sets. Reidel, Dordrecht-Boston (1982) 445–470
9. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical foundations. Springer (1999)
10. Ganter, B., Stumme, G., Wille, R., eds.: Formal Concept Analysis, Foundations and Applications. Volume 3626 of Lecture Notes in Computer Science. Springer (2005)
11. Priss, U.: Formal concept analysis in information science. Annual Review of Information Science and Technology **40** (2006) 521–543
12. Földiák, P., Endres, D.: Sparse coding. Scholarpedia **3**(1) (2008) 2984 http://www.scholarpedia.org/article/Sparse_coding.
13. Földiák, P.: Sparse coding in the primate cortex. In Arbib, M.A., ed.: The Handbook of Brain Theory and Neural Networks. second edn. MIT Press (2002) 1064–1068
14. Földiák, P.: Forming sparse representations by local anti-Hebbian learning. Biological Cybernetics **64** (1990) 165–170
15. Olshausen, B.A., Field, D.J., Pelah, A.: Sparse coding with an overcomplete basis set: a strategy employed by V1. Vision Res. **37**(23) (1997) 3311–3325
16. Olshausen, B.: Learning Linear, Sparse, Factorial Codes. Technical Report AIM 1580 (1996)
17. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. Annual Review of Neuroscience **24** (2001) 1193–1216
18. Rolls, E., Treves, A.: The relative advantages of sparse versus distributed encoding for neuronal networks in the brain. Network **1** (1990) 407–421
19. Dayan, P., Abbott, L.: Theoretical Neuroscience. MIT Press, London, Cambridge (2001)
20. Jones, J., Palmer, L.A.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology **58**(6) (1987) 1233–1258
21. Ringach, D.L.: Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. Journal of Neurophysiology **88** (2002) 455–463
22. Földiák, P., Xiao, D., Keysers, C., Edwards, R., Perrett, D.: Rapid serial visual presentation for the determination of neural selectivity in area STSa. Progress in Brain Research (2004) 107–116
23. Oram, M.W., Perrett, D.I.: Time course of neural responses discriminating different views of the face and head. Journal of Neurophysiology **68**(1) (1992) 70–84
24. Endres, D., Földiák, P.: Exact Bayesian bin classification: a fast alternative to bayesian classification and its application to neural response analysis. Journal of Computational Neuroscience **24**(1) (2008) 24–35 DOI: 10.1007/s10827-007-0039-5.
25. Endres, D.: Bayesian and Information-Theoretic Tools for Neuroscience. PhD thesis, School of Psychology, University of St. Andrews, U.K. (2006) http://hdl.handle.net/10023/162.
26. Hinton, G.: Products of experts. In: Ninth International Conference on Artificial Neural Networks ICANN 99. Number 470 in ICANN (1999)

27. Kiani, R., Esteky, H., Mirpour, K., Tanaka, K.: Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. Journal of Neurophysiology **97**(6) (April 2007) 4296–4309
28. Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L.: Identifying natural images from human brain activity. Nature **452** (2008) 352–255 http://dx.doi.org/10.1038/nature06713.