Publication III

Mikko Berg, Jan-Hendrik Schleimer, Jaakko Särelä, and Timo Honkela. 2005. Category learning by formation of regions in conceptual spaces. In: Lorenzo Magnani and Riccardo Dossena (editors). Computing, Philosophy and Cognition. Proceedings of the 2nd European Computing and Philosophy Conference (ECAP 2004). Pavia, Italy. 3-5 June 2004. London, United Kingdom. College Publications. Texts in Philosophy, volume 4, pages 381-396. ISBN 1-904987-24-9.

Category Learning by Formation of Regions in Conceptual Spaces

MIKKO BERG, JAN-HENDRIK SCHLEIMER, JAAKKO SÄRELÄ, AND TIMO HONKELA

ABSTRACT. In this paper, we discuss the issue of conceptualization. The traditional view is that concepts are essentially linguistic. Recently, Gärdenfors has proposed a contradicting view stating that the concepts get associated to language terms, but essentially belong into other domain called conceptual spaces defined by quality dimensions. These dimensions form meaningful representations of the concept domains in hand and they should be formable by mappings from the sensory input and possibly from other more basic quality dimensions as well.

In the space spanned by the quality dimensions, natural concepts form convex regions. The borders of these regions can be hard or soft and can vary according to the context. In the present work, we have decided to code the regions by prototypes, so that instances closest to a particular prototype in the conceptual space form a region. In other words, the regions are defined by the Voronoi tessallations of the prototypes, which later define hard-bordered regions. In the case of soft borders, the prototypes can consist of probabilistic density functions defining graded membership function for each point in the conceptual space.

This paper explores the idea of quality dimensions by trying to realize contextual categorization in such a domain. That is, trying to form prototypes and regions. As addition, the connections to the lower, connectionist level and to the higher, symbolic level are discussed briefly.

1 Introduction

Intelligent systems generalize and compress the complex input they receive through their perceptual organs. This is clearly necessary to survive in a complex and potentially hostile world. Human beings have an exceptional capacity to utilize this process. We often rise from the basic regularities of the world to more abstract interpretations. This makes it possible to exploit

Lorenzo Magnani and Riccardo Dossena, editors, *Computing, Philosophy, and Cognition*, pp. 381–396

even very distant (in time or place) similarities to make effective predictions of the state of the world. Another trait of humans is the capability for complex communication. Probably, to attain robustness, this communication in general takes place using discrete symbols, words.

Until now, a very central issue in artificial intelligence (AI) is arguably the relation between these very central traits: effective modeling of the world (accessed by sensory organs) and effective communication of the relevant parts of these models (language).

The traditional view, formulated by Newell & Simon [Newell *et al.*, 1958] is that we are physical symbol manipulating systems. This is to claim that the models we have of the world, are essentially linguistic. The modern view relies on dynamic systems theory [Kelso, 1995]. It claims that symbols emerge from dynamic interaction processes.

Connectionism is regarded to be a specific version of this dynamic hypothesis [Van Gelder, 1995]. The connectionist paradigm for AI gained popularity, in the early 90's, mainly through the books by the PDP research group [McClelland *et al.*, 1987; Kohonen, 1984]. They argued that human information processing is mainly continuous not discrete. Furthermore, the essential feature of human intelligence is learning, thus making the conceptual system a dynamic process rather than a static one.

One drawback of most of the connectionist algorithms is their distributed knowledge representation, which does not allow explicit interpretations of the inference process. That is why these systems are sometimes referred to as "blackboxes". A famous example is the NetTalk system from Seinowski and Rosenberg, a multi-layerd perceptron capable of reading English texts. The system was trained in a supervised manner with text as input and corresponding phonemes as output. Although achieving an accuracy of 95% the neural network did not extract rules for the decision making, that could be interpreted by linguistic processing. This example underlines the gap between the connectionist models and symbol manipulation systems.

Connectionism can be interpreted as a special case of associationism using ANN (Artificial Neural Networks). Gärdenfors [Gärdenfors, 2000] has presented a new level on top of these neural models trying to reach the symbolic level processes that humans are naturally capable of. The model being functional, Gärdenfors states that conceptual spaces can be seen as a set of attractor points of dynamic systems. Yet, his model retains the possibility of classical symbol manipulation with the three-level-model: 1) connectionism as the lowest, 2) conceptual spaces in between, and 3) classical symbol manipulation as the highest level.

Domains in conceptual spaces are an attempt to give functional and contextual focus for otherwise ambiguous symbolic level. One concept can be evaluated in several domains using different salience weights, where as properties are domain specific. Scale of the particular dimension in a domain is obtained using contrast classes. In another words, the continuous mapping to the subspace is performed within the boundaries of contextual extreme values. For example, what is considered to be (phenomenological) hot for bathing water is merely warm for coffee. In general, different abstractions are created with the corresponding *quality dimensions* having specific metrics.

This article does not try to rescue the idea of quality dimensions from its weaknesses. Most importantly, the satisfactory explanation of how these domains and quality dimensions come about is missing from [Gärdenfors, 2000], Here the dimensions are taken as given, assuming that some of them result from innate biological structures with evolutionary background. This is of course not true to all dimensions that are more abstract and which can be learned.

In conceptual spaces, (natural) concepts are defined as (convex) regions[1]. Voronoi tessellations necessarily result in convex spaces when Euclidean metrics is used. Voronoi tessellation partitions given space based on prototypical attractors. Clustering methods tackle the reverse problem, by defining regions which detect the prototypes.

The nature of a concept in conceptual spaces is

1. prototypical, coding of the structure

2. regional, geometric area instead of points (objects are very narrow concepts, perhaps even points), this makes the concepts vague or fuzzy, which relates to frame theory

In a sense, prototype and frame theory are combined here.

As addition to concept borders, there are also other reasons for modeling vagueness in concept formation and communication. Our dynamic scheme is thought to have three interacting parties: 1) cognitive concepts (including laws of psychology), 2) language and social interaction, and 3) phenomenal common world (including laws of physics). These entities have influence on the prototypes, and their connection is considered to be a source of impreciseness or fuzziness.

Next section explains further what is meant by dynamical hypothesis, followed by two sections discussing how conceptual spaces model extends this, relying first on traditional prototypes and second on convex regions. After that, in section 2 we review two clustering methods, as well as discuss

[1]According to Gärdenfors, natural concepts are the only concepts that can participate in inductive reasoning.

the possibility to combine similar concepts into more general concepts corresponding to larger regions in the conceptual space. Finally, in section 3 we apply these clustering methods to divide a space with color quality dimensions into concepts according to two images differing in characteristics.

1.1 Dynamical hypothesis

As a result from the ability to adapt to the environment and learn from experiences, our concepts change in time. Considering this, it would be implausible to assume nativist perspective for conceptual modeling and use innate rules for all categorization. Instead some rules, namely learning rules, could be used to guide the concept formation process, but not the concepts itself.

Concepts are assumed to emerge through self-organization process guided by top-down (global) and bottom-up (local) influence. The dichotomy results from modeling levels, where complex global behavior emerges from local interaction of simple and homogenous elements [Van Gelder, 1995]. For instance, the limitations of short-term memory could be seen as boundary condition for conscious analyzes of features of an object. In conceptual domain, it would seem natural for the regions to influence the location of prototypes and vice versa, until stable categories are obtained.

The challenge of an emergence theory is to explain the relationship between the chosen levels. By the definition, it is impossible to witness more global phenomena from the local level, but according to microreductionism (weak version of emergentism, [Buchmann, 2001]) top-down constraints are result of bottom-up effects. In fact, there is no level with ontological priority according to constructive reductionism [Kelso, 1995]. However, in this case the relationship between distributed neural level and symbols need not to be merely descriptive, and that is not what the mentioned three-party-interaction scheme implies. The existence of a symbol that groups observations naturally affects perception. This can take form of Categorical Perception to concept borders [Harnad *et al.*, 1991] or paradigm shift [Kuhn, 1996] to entire conceptual system. The effect of symbols becomes more apparent in next section with the notion of prototypicality.

1.2 Prototype theory

Prototype theory was formulated by Rosch and got started from findings relating to typicality (not yet having prototypical structure) among the category members. Findings of Rosch and Mervis [Rosch and Mervis, 1975] emphasized typicality as opposed to all category members representing the category equally. Rosch [Rosch and Lloyd, 1978] found that there are more typical members that are learned faster and serve as cognitive reference. The membership was considered to be graded and it was shown not to result from frequency or familiarity of the particular test items. The correlations with frequencies turned out to be useful in many cases, but not definitive. As an exception, chicken is frequent, but not typical bird. The results of Rosch & al. [Rosch *et al.*, 1976] supported this finding, but only when structural relations between items were held constant.

After that, the characterizing properties were the target of the research. First, Wittgenstein's family resemblance rate was found to describe categories better. There were no explicit definitions, but similarities between individual group members, that could be modeled with locally similar cells. Second, exclusiveness (not total) was also proposed as typicality measure. Then the typicality would not only relate to the features of particular group, but also to the shortage of important features from other groups (contrast category). This is the phenomenon that Gärdenfors' [Gärdenfors, 2000] quality dimensions are explained to obtain their scaling. Contrast categories are difficult to verify empirically, because it would involve all the (other) categories. Third, it was found that broader knowledge structures and top-down processing play their part in this as well. For example functionalities can be inherited to sub-categories [Rosch and Mervis, 1975]. Barsalou [Barsalou, 1985] later repeated the related experiments.

The actual prototype theory was based on one summary representation of all the members, not as commonly misunderstood on the best match. Based on psychological experiments, Strauss [Strauss, 1979] proposed a method, in which features of the prototype should be averaged if their distribution is small and counted distinctively if it is sparse. The counting was explained by subject's interpretation as qualitative differences, not on one continuous axis. There is an analogy to how Gäärdenfors' dimensions evolve from integral, having correlation, to distinct separable dimensions, for example when child learns to separate shape from color. Feature correlations are method for applying prototypes and correlations alone are not sufficient for categorization. In terms of conceptual spaces, after arbitrary mapping, any two points in space can be close to each other. It has been claimed that people use hierarchical clusters. The intermediate groupings effect the typicalities, for example the statement that robin is a typical bird may be overlooking the fact that it is small, chirping, worm- or seed-eating tree bird [Malt and Smith, 1984].

Rosch [Rosch and Lloyd, 1978] describes the vertical dimension of the structure as taxonomy of category relations. There is inclusiveness of subordinate (lower-level) through basic level into superordinate (higher-level). The basic level categories is a topic with much empirical research. Read more from [Rosch and Lloyd, 1978]. The horizontal dimension is segmented structure without clear-cut boundaries. There is only the judgment for clearness of the case, the prototypicality.

There has been the idea of using probabilities to increase the accuracy of categorization and for example Churchland [Churchland, 1989] uses term warranty for uncertainty of chosen prototype. Experiments of Ross and Murphy [Ross and Murphy, 1996] showed that this was not actually accounted and turned the focus on preciseness of categorization.

There should be discussion about to what extent can human cognition be modeled with prototypes or with ANN (Artificial Neural Networks) algorithms. It is argued that theories should be verified using the evidence from psychological research, instead of mere speculations. Some of such attempts to find the limitations of the prototype theory in the past are exemplar effect of context model (started by Medin & Schaffer [Medin and Schaffer, 1978]) and the research on human memory, and different models about the use of background knowledge (e.g. [Murphy and Medin, 1985], read more from "Theory-Theory" in [Laurence and Margolis, 1999])

A vector in ANN model as Roschian prototype represents a summary of all the members of the cell, and not the best match. The prototype theory does not provide any model for the process, representation or learning. It only presents constraints and a possibility to deal with abstractions without any context. One of such constraints or descriptions is that there is correlation structure of the neighbors in nature of family resemblance [Rosch and Lloyd, 1978]. For instance, this is the way in which input of SOM [Kohonen, 1984] map is connected, because it gives emphasis on retaining the local level structure. There is no explicit way to define how SOM creates the model vectors, because the process is a result from heuristic principles. Neither is there any evidence there should be such for prototype theory. For example independent cue model [Medin and Schaffer, 1978] is only one ineffective implementation.

2 Discovering regions in conceptual spaces

Identifying concepts with regions in the space already adds an element of vagueness to the conceptual representation, because it subsumes objects $x \in \mathbb{R}^d$ with a variety of different attributes as one concept.

It is argued that there is a another vague element, namely that objects do not utterly belong to concepts or putting it in probabilistic terms, there are varying probabilities with which different objects are explained by a concept. Then the hard margins of the regions, representing concepts, in the plainly geometric approach make it difficult to incorporate this vagueness. A possible solution is to define a probability distribution in the conceptual space, that itself corresponds to a concept.

Finding the regions can be solved by clustering methods, but it is as well necessary to infer how many clusters are needed, and in an dynamical environment, the decision whether to split or combine regions, respectively concepts, arises. This question can be partly solved by hierarchical clustering methods or moving to Bayesian versions of clustering algorithms that give evidence on the model complexity, e.g. the number of concepts needed.

In the following sections we discuss three methods for finding these regions and defining a vague concept in them. It is assumed that the objects, perceived in nature or encountered in a more abstract way in our mind, are represented in as points in a conceptual space [Gärdenfors, 2000].

2.1 K-means clustering

The *k-means clustering algorithm* [Bishop, 1995] moves a chosen number of k cluster centers, so that they cover the whole data and thereby partitioning it for $i \in [1, k]$ into subsets $\mathcal{S}_i$, defined by their center $\boldsymbol{\mu}_i$ and containing the N_i nearest data points. It does it via minimizing the sum-of-squares error function,

$$E = \sum_{j=1}^{k} \sum_{n \in \mathcal{S}_i} ||\boldsymbol{x}_n - \boldsymbol{\mu}_i||^2 \tag{1}$$

but other distance measures can be used as well. The batch version of the algorithm has an update rule $\Delta\boldsymbol{\mu}_i = \eta(\boldsymbol{x}_n - \boldsymbol{\mu}_i)$ quite similar to that of SOM, only lacking the neighborhood function. With the help of the mean vectors a Voronoi tessellation can be found, as used by Gärdenfors for concept representation.

The defined regions are vague representations of concepts. But if the euclidean distance is used to identify the k nearest neighbors or even a tessellation, than there are hard margin between concepts, which does not seem to be a natural representation.

2.2 Density estimation

As shown in [Bishop, 1995] the k-mean algorithm can be regarded as a limit of the EM optimization of a *Gaussian mixture model* (MOG) with a common variance, when $\sigma^2 \to 0$. In a Gaussian mixture model the probability density of the data $p(\boldsymbol{x}) = \prod_{n=1}^{N} p(\boldsymbol{x}_n)$ is modeled as a weighted sum of Gaussians

$$p(\boldsymbol{x}_n) = \sum_{i=1}^{k} p(\boldsymbol{x}_n|i)p(i). \tag{2}$$

with a soft max prior $p(i) = \frac{\exp(\gamma_i)}{\sum_j \exp(\gamma_j)}$ and $p(\boldsymbol{x}_n|i) \sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma_i)$. The negative log-likelihood of the data

$$- \log(p(\boldsymbol{x})) = - \sum_{n=1}^{N} \log \left\{ \sum_{=1}^{k} p(\boldsymbol{x}_n|i)p(i) \right\} \tag{3}$$

can be used as an error function. Finding the minimum by setting the derivatives for $\boldsymbol{\mu}_i$, σ_i^2 and γ_i to zero and using the the Bayes' theorem to get the corresponding posterior $p(i|\boldsymbol{x}_n) = \frac{p(\boldsymbol{x}_n|i)p(i)}{p(\boldsymbol{x}_n)}$, the following updating rules can be derived

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_n p(i|\boldsymbol{x}_n)\boldsymbol{x}_n}{\sum_n p(i|\boldsymbol{x}_n)} \tag{4}$$

$$\hat{\sigma}_i^2 = \frac{\sum_n p(i|\boldsymbol{x}_n)||\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_i||}{d \sum_n p(i|\boldsymbol{x}_n)} \tag{5}$$

$$\hat{p}(i) = \frac{1}{N} \sum_n p(i|\boldsymbol{x}_n) \tag{6}$$

Due to the nonlinear dependencies in the equation a iterative update scheme is used to solve the problem. Start ing with random initial values for the parameters and then calculating the posterior and the new parameter values. It can be shown that repeating this process will converge to a maximum likelihood solution.

Applying this algorithm to points in a conceptual space results in a probability density function that covers the structure of the points arrangement in the space. This distribution can be identified with a certain concept, where the mean vectors of the Gaussian mixture components are prototype like examples of them. The individual Gaussians can represent more detailed sub-concepts. But still remains the question of how many centers shall be used.

Another unsolved problem is that, when operating the algorithm on every object of the conceptual space one large MOG distribution will result and therefore only one concept. So one has to use the clustering in a hierarchical way. For example first tessellate in a crude way to find different concepts using the k-means algorithm and than find the distributions in the cluster with the help of a Gaussian mixture model.

2.3 Hierarchical clustering

Instead of applying the above mentioned clustering methods repeatedly one can utilize a *hierarchical clustering* in the first place. A possible class of methods are called single linkage algorithms for a detailed description see

[Rohlf, 1982]. These algorithm start by treating every data point as one cluster and than combine the "most similar" according to the used metric. This is done repeatedly using minimum, maximum, the average distance or the distance of the centers of gravity[2] for comparing clusters containing more than one data point, and thereby creating a hierarchical structure.

The lower branches in the hierarchy can be cut away, meeting the concerns of difference only to a certain level of detail. But how is it then that a concept generating process in an intelligent system could find a level that is meaningful to use? There are two answers at hand: (i) just use any detail level for a start, and then, by a process similar to natural selection in living creatures or maximizing the model evidence in AI, it will turn out to be more useful to go into a more detailed version of the concepts or to thin them out and therefore have broader concepts; (ii) in a Bayesian version of the clustering algorithms, in spirit closer to density estimation, it is possible to combine the data likelihood with a prior distribution, representing the anticipation for the number of concepts needed, which can itself result from previous knowledge and experience in the world, and hence get a posterior probability distribution over the needed number of concepts.

2.4 Bayesian mixture model

Deriving concepts from available facts, e.g. sensory data and existing knowledge of the world - in this case represented in conceptual spaces, is an inferential task with statistical properties, resulting from the irregularities in the frequency of the data and the incertitude of the already gained knowledge, respectively.

A mathematical framework for describing statistical inference problems is the Bayesian statistics, where a basic idea is to interpret the probability of an event as the *degree of belief* on the occurrence of that event. Learning the attributes $\boldsymbol{\theta}$ of a model structure $\mathcal{H}$ e.g. the shape and location of the gaussians forming the distribution associated with a concept, is achieved by combining prior knowledge, described by a distribution indicating the believe in certain facts, with new information from data $\boldsymbol{x}$, described by a likelihood of the data given the learned quantity and the model structure. A possibility to calculate the posterior distribution of the attributes, which combines old and new knowledge is given by Bayes' theorem

$$p(\boldsymbol{\theta}|\boldsymbol{x},\mathcal{H}) \quad = \quad \frac{p(\boldsymbol{x}|\boldsymbol{\theta},\mathcal{H})p(\boldsymbol{\theta}|\mathcal{H})}{p(\boldsymbol{x}|\mathcal{H})}, \tag{7}$$

with $p(\boldsymbol{x}|\mathcal{H}) = \int p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{H})d\boldsymbol{\theta}$ being known as the *model evidence*. This

[2] This relates to discussions in prototype theory about which set member, if any (as Roschian prototype theory suggests) should be used as the representative.

integral over all possible parameter values is, for difficult distributions not always solvable, but maximizing it with respect to $\mathcal{H}$ would lead to more optimal model structures.

This calculation of the posterior can be conducted each time new data is available and if the posterior distribution of the former inference step is used as the prior in the next execution of the Bayes' rule, it will lead to an adaptive learning mechanism. An intelligent system acting in a new environment and starting to conceptualize from scratch might in some circumstances not have prior knowledge for the shape of concepts, and therefor the categorization of the new and unknown. Still it is possible to define *non-informative priors*, that do not influence the finding of the posterior for the attributes, but "let the data speak for its self".

As mentioned earlier, one can express the density estimation problem in the bayesian framework (see [Attias, 2000] for a detailed derivation). One advantage is that this treatment allows searching for optimal model structure, e.g. the number of gaussians in the mixture model, whereas this is not feasible in the ML solution (paragraph 2.2) without empirical regularization terms. This is due to the fact that the ML solution from the EM algorithm prefers more complex model structures, that fit better to the data.

The approach in [Attias, 2000] is from the structure of the algorithm related to EM, but utilizes a helpful technique in bayesian inference called variational learning. There the posterior distribution of the parameter, that is often complicated to calculate, due to the difficult integral in (7), is approximated by a distribution with desired properties. In the case where the best model structure should be determined the requirement is that the approximate model evindence needed to optimize the number of gaussian components can be obtained in closed form.

It should be mentioned that there are many other model selection techniques like bootstrapping [Efron and Tibshirani, 1993], cross-validation, Markov-Chain-Monte Carlo sampling and Bayesian Information Criterion (BIC), see [Gelman *et al.*, 2003], which all somehow work in practice, but most of them are theoretically only justified for infinite data sets, whereas concepts can certainly emerge from only few examples.

3 Clustering of color spaces for concepts

As a simple example, the conceptualization of colors in two pictures, originating from a landscape in summer and winter, was studied. Choosing these pictures it can be expected that the process of conceptualization in our model depends on the encountered examples, a peculiarity of concept forming, that can be observed in the real world, e.g. considering various ethnic groups, that divide the color spectrum into differently detailed colors

[Bornstein, 1973; Hardin, 1993].

The color code for the pixel elements of the pictures is the hue-saturation-value color map, which is a intuitive representation for humans. The colors are coded with three numbers, firstly the *hue*, ranging from 0 to 360 degree in a circular arrangement and indicating the color type according to its wavelength, secondly the *saturation* or intensity between 0-100%, telling how grayish the color is and finally the *value* in percentage, that tells the brightness or the spread of wavelength. The hsv color space is redundant because there exists white and black for every color. Therefore, a color spindle instead of the cylinder in HSV model has been suggested [Kamvysselis and Marina, 1999]. It is achieved by reducing the range of the saturation linearly as the intensity approaches 0 or 100%. This modified color code has been used in the experiments and the intervals were scaled to unity.

A representative set of the data points for the summer and winter pictures can be seen in Figures 1 and 2 respectively. The prototypes for the MOG model, i.e., the means of the Gaussians have been marked there with x's as well. As expected, the MOG model has used more resources that is, more prototypes to account for areas having more data points. Observing that they cover the distribution of the color samples quite well, the corresponding colors can be expected to cover the coloring, present in the picture, appropriately. But the results depend completely on how many initial mixture components are chosen.

Thus the clusters given by the EM algorithm were further combined to bigger clusters by the hierarchical linkage algorithm. The resulting colors as well as the hierarchy can be seem in Figsures 1 and 2 for the summer and winter pictures respectively. Now one can see the grouping of different shades of white an brown to a more general concept of the color.

Definite differences in the prototype colors can be seen. While the clusters formed from the summer picture have several shades of green and dark gray, the colors in the winter picture are concentrated in lighter shades of gray and white.

[More inferences of the results are made, when we have the results of the spindle model. Now there are, for example, very dark colors that do not appear to be close to each others. This is due to the significant difference in the hue.]

4 Discussion

We issued some implementation aspects left open by Gärdenfors' Conceptual spaces [Gärdenfors, 2000]. We mainly discussed the formation of the concepts as regions in a given conceptual space. The significance of these results to the understanding of actual implementation of human intelligence

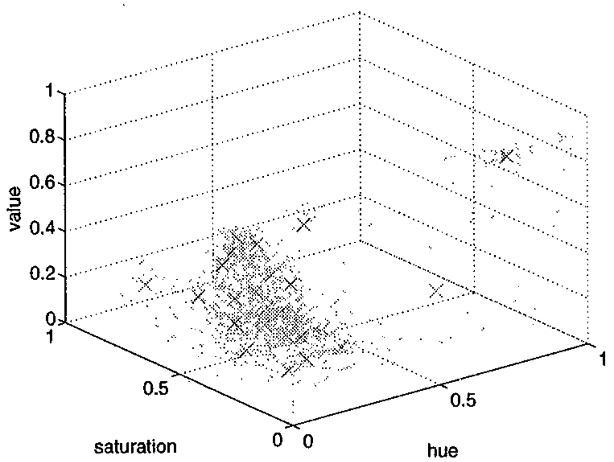

Figure 1. The color vectors of the summer picture with 27 centers for a mixture of Gaussian model after 30 iterations of training with EM algorithm.

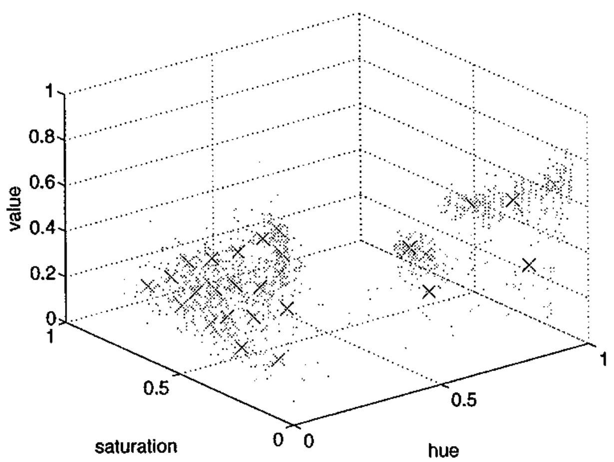

Figure 2. The color vectors of the winter picture with 27 centers for a mixture of Gaussian model after 30 iterations of training with EM algorithm.

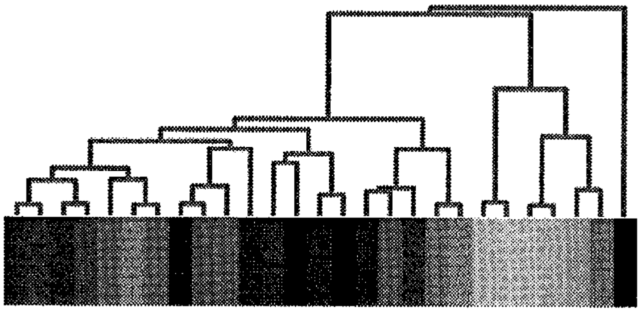

Figure 3. The 27 colors of the summer picture in a dendrogram.

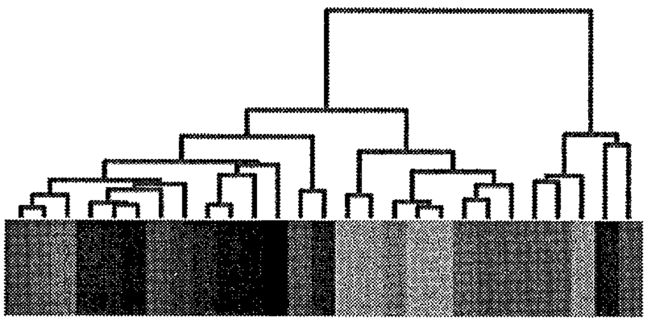

Figure 4. The 27 colors of the winter picture in a dendrogram.

might be questionable or at least modest. However, the central contribution of this paper does not lie therein, but in simulation of the intelligence as an AI project.

We only paid attention to the categorization in already acquired conceptual spaces. We now discuss in brief the connection of the conceptual level to the connectionist, namely the acquisition of the quality dimensions, and the symbolic levels, namely thought processes and language.

A natural way to connect the conceptual level to the basic sensory input level is provided by the connectionist approach. The quality dimensions are determined by the sensory input as well as possibly some other more basic quality dimensions using a flexible nonlinear mapping. However, Gärdenfors usually takes the quality dimensions as given, though clearly this cannot be true for all concepts. The principles guiding the learning are not easy to state, because they should include at least, capacity constraints, generalization of properties and finally, the relevance of different structures in the sensory data for the particular task the concepts are needed for.

Dynamical interaction framework was described as a starting point to explain the emergence of concepts. One possible way to advance into the direction of dynamic systems theories is to have behavioral models with discrete attractor basins (e.g. energy minima) [Cariani, 2001]. Kelso [Kelso, 1995] has studied these extensively and hinted that such basins could be interpret as prototypes. This is significant, because prototype theory itself does not deal with learning or concept formation, but only structure.

Furthermore, to really bridge the conceptual level to the symbolic level, one needs to explain the relation between the acquired concepts and language. We see it plausible to assume that language terms get associated to the regions in the conceptual spaces, that is concepts. Then concepts that get instantiated due to sensory input or voluntary thought processes may trigger the use of language, internally or in a speech act.

Another property of concepts in the influence of natural language, is their context sensitivity. As an example, one could think of the different meanings of hot when going to sauna or having fever. Gärdenfors suggests that by a magnification or scaling of the quality dimensions (see the skin color example on page 119f of [Gärdenfors, 2000]) could amount to this property. In the bayesian framework context sensitivity can be achieved by the use of different priors, that modify the mean and variance of the gaussians to meant the contextual environment.

5 Acknowledgements

We would like to thank all the participants of the course of conceptual modeling at HUT, Autumn 2003, for their valuable input, especially Mr.

Janne Hukkinen and Mr. Karthikesh Raju. We would like to thank the Neural network research centre for providing the infrastrucutre for working on this subject.

BIBLIOGRAPHY

[Attias, 2000] H. Attias. A variational bayesian framework for graphical models. In T. et al Leen, editor, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, Cambridge, MA, 2000.

[Barsalou, 1985] L. Barsalou. Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11:629–654, 1985.

[Bishop *et al.*, 1996] C. M. Bishop, M. Svensen, and C. K. I. Williams. GTM: a principled alternative to the self-organizing map. In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, editors, *Artificial Neural Networks—ICANN 96. 1996 International Conference Proceedings*, pages 165–70. Springer-Verlag, Berlin, Germany, 1996.

[Bishop *et al.*, 1998] Christopher M. Bishop, Markus Svensen, and Christopher K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234, 1998.

[Bishop, 1995] Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[Bornstein, 1973] M.H. Bornstein. Color vision and color naming: A psychological hypothesis of cultural difference. *Psychological Bulletin*, 80:257–285, 1973.

[Buchmann, 2001] M. Buchmann. Emergent properties. In N. Smelser and P. Baltes, editors, *International Encyclopedia of the Social, and Behavioral Sciences*, pages 4424–4428. Elsevier Science, Oxford, 2001.

[Cariani, 2001] P. Cariani. Symbols and dynamics in the brain. *BioSystems*, 60:59–83, 2001.

[Churchland, 1989] P.M. Churchland. *A neurocomputational perspective: The nature of mind and the structure of science.* Cambridge, MA: MIT Press, 1989.

[Efron and Tibshirani, 1993] B. Efron and R. Tibshirani. *An introduction to the bootstrap.* Chapman & Hall, 1993.

[Gärdenfors, 2000] Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought.* MIT Press, 2000.

[Gelman *et al.*, 2003] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis.* Chapman & Hall/CRC, 2 edition, 2003.

[Goldstein, 1999] B. Goldstein. *Sensation and Perception.* International Thomson Publishing Company, 5 edition, 1999.

[Hardin, 1993] C.L. Hardin. *Color for Philosophers.* Indianapolis/Cambridge: Hackett Publishing Company, expanded edition, 1993.

[Harnad *et al.*, 1991] S. Harnad, S.J. Hanson, and J. Lubin. Categorical perception and the evolution of supervised learning in neural nets. *Working Papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*, pages 65–74, 1991.

[Hård and Sivik, 1981] A. Hård and L. Sivik. NCS–natural color system: a Swedish standard for color notation. *Color research and application*, 6:129–138, 1981.

[Kamvysselis and Marina, 1999] M. Kamvysselis and O. Marina. *Imagina: A Cognitive Abstraction Approach to Sketch-Based Image Retrieval.* MIT Press, Cambridge, MA, 1999.

[Kelso, 1995] J. A. S. Kelso. *Dynamic Patterns: The Self-Organization of Brain and Behavior (Complex Adaptive Systems).* Cambridge: MIT press, 1995.

[Kohonen, 1984] T. Kohonen. *Self-Organization and Associative Memory.* Berlin, Heidelberg: Springer, 1984.

[Kohonen, 1995] T. Kohonen. *Self-Organizing Maps*. Berlin: Springer-Verlag, 1995.

[Kuhn, 1996] T.S. Kuhn. *The structure of scientific revolutions (3rd edition)*. Chicago: University of Chicago Press, 1996.

[Lakoff and Johnson, 1999] G. Lakoff and M. Johnson. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. N.Y. : Basic Books publishing, 1999.

[Laurence and Margolis, 1999] S. Laurence and E. Margolis. *Concepts: Core Readings*, chapter 1, pages 3–81. Cambridge, MA: MIT Press, 1999.

[Malt and Smith, 1984] Barbara C. Malt and Edward E. Smith. Correlated properties in natural categories. *Journal of verbal learning and verbal behavior*, 23:250–269, 1984.

[McClelland *et al.*, 1986] J. L. McClelland, D. E. Rumelhart, et al., editors. *Parallel distributed processing: Volume2: Psychological and Biological Models*. Cambridge: MIT press, 1986.

[McClelland *et al.*, 1987] J. L. McClelland, D. E. Rumelhart, et al., editors. *Parallel distributed processing: Volume1: Foundations*. Cambridge: MIT press, 1987.

[Medin and Schaffer, 1978] D. Medin and M. Schaffer. Context theory of classification learning. *Psychological Review*, 85:207–238, 1978.

[Murphy and Medin, 1985] G. Murphy and D. Medin. The role of theories in concept coherence. *Psychological Review*, 92:289–316, 1985.

[Murphy and Wisnewski, 1989] G. Murphy and E. Wisnewski. Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15:572–586, 1989.

[Newell *et al.*, 1958] A. Newell, H.A. Simon, and J.C. Shaw. Elements of a theory of human problem solving. *Psychological Review*, 65:151–166, 1958.

[Rohlf, 1982] F. James Rohlf. Single-link clustering algorithms. In P.R. Krishnaiah and L.N. Kanal, editors, *Handbook of Statistics*, volume 2, pages 267–284. Elsevier Science Publishers, Amsterdam, The Netherlands, 1982.

[Rosch and Lloyd, 1978] E. Rosch and B. Lloyd. *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum, 1978.

[Rosch and Mervis, 1975] E. Rosch and C. Mervis. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

[Rosch *et al.*, 1976] E. Rosch, C. Simpson, and R. Miller. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2:491–502, 1976.

[Ross and Murphy, 1996] B. Ross and G. Murphy. Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25:51–63, 1996.

[Stephens, 1998] M. Stephens. Bayesian analysis of mixture models with an unknown number of components. *Journal of the Royal Statistical Society*, 59:731–792, 1998.

[Strauss, 1979] M. Strauss. Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory*, 5:618–632, 1979.

[Van Gelder, 1995] T. Van Gelder. What might cognition be, if not computation? *Journal of Philosophy*, 92:345–381, 1995.

[Zadeh, 1965] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Mikko Berg, Jan-Hendrik Schleimer, Jaakko Särelä, and Timo Honkela
Neural Network Research Centre
Helsinki University of Technology, Helsinki, Finland
Email: {mikko.berg,schleime,jaakkos,timo.honkela}@hut.fi