

Learning to Understand - General Aspects of Using Self-Organizing Maps in Natural Language Processing

Timo Honkela
Helsinki University of Technology
Neural Networks Research Centre
P.O.Box 2200, FIN-02015 HUT, Finland
tel: +358 0 451 3275, fax: +358 0 451 3277
e-mail: Timo.Honkela@hut.fi
WWW: <http://nucleus.hut.fi/~tho/>

Abstract

The Self-Organizing Map (SOM) is an artificial neural network model based on unsupervised learning. In this paper, the use of the SOM in natural language processing is considered. The main emphasis is on natural features of natural language including contextuality of interpretation, and the communicative and social aspects of natural language learning and usage. The SOM is introduced as a general method for the analysis and visualization of complex, multidimensional input data. The approach of how to process natural language input is presented. Some epistemological underpinnings are outlined, including the creation of emergent and implicit categories by SOM, intersubjectivity and relativity of interpretation, and the relation between discrete symbols and continuous variables. Finally, the use of SOM as a component in an anticipatory system is presented, and the relation between anticipation and self-organization is discussed.

Keywords: natural language processing, self-organizing maps, semantics, epistemology, neural networks.

1. Introduction

Traditionally the formal study of language has centered around structural and static aspects. The automatic analysis and generation of syntactic structures has mainly been based on explicit, hand-written, and symbolic representations. In semantics the main focus has been on propositional structures, quantifiers, connectives, and other phenomena that match well the apparatus of predicate logic. This paper aims at widening the scope to include many more natural features of natural language including contextuality of interpretation, and the communicative and social aspects of natural language learning and usage. The principles of formalizing specific aspects of these phenomena are considered in this paper including the following:

- In traditional study of language fixed categorizations are normally used. Lexical items, words and phrases, are positioned into categories such as verbs and nouns, and these categories are used in abstract rules, e.g., of the type "S \rightarrow NP VP", i.e., a sentence consists of a nominal phrase and a verb phrase. It may seem that the abstract rules are precise, but when they are applied, discrepancies exist between the rules and the actual use of language. A rule may be incorrect in various ways. For instance, a rule may be overly general and should be refined. Refining the rule may be based on adding extra

restrictions on its use, or creating more fine grained categories that divide the feature space into smaller areas. When this refinement process is continued into the extreme it may appear that each word has a category of its own. At least, it seems that a natural grammar has a fractal structure.

- The use and interpretation of language is adaptive and context-sensitive. One can, of course, find the most usual patterns and make definitions based on them, but in the actual discussions and writings, words are often used creatively based on the particular situation. The well-known ambiguity "problem" highlights the context-sensitivity: there may be multiple interpretations for a word or a phrase but in the context the desired interpretation can be understood. Most often human listeners or readers do not even notice the potential alternative readings of distinct words. The preceding text and the overall context supports an anticipatory process that blocks effectively incorrect interpretations.
- The context-sensitivity of interpretation is also relevant when one considers the more fine-grained structure of semantic and pragmatic level. The traditional, logic-based ontology of natural language interpretation is based on the idea that the world consists of distinct objects, their properties, and the relationships between the objects. Such a view neglects the fact that the propositional level of sentences does not have a simple one-to-one counterpart in the reality. The reality is highly complex, apparently high-dimensional in the perceptual level, changing, consisting of non-linear and continuous processes. Thus, studying the epistemological level having basically "names" and "objects" as their referents, may be considered to be far too simplistic. One should, for instance, take into account the relation between discrete symbols and the continuous spaces which the symbols refer to.
- Human understanding of natural language is based on the long individual experience. Inevitably the differences in personal histories cause differences in the way humans interpret natural language expressions. In the light of the previous discussion, it should be clear how approaching this kind of phenomenon is difficult using the apparatus of symbolic logic without considering more refined mathematical tools of algebra. The subjectivity of interpretation is apparent when thoughtfully. The communication is enabled by the intersubjectivity based on the learning process in which the interpretations are adapted to match well enough so that meaningful exchange of thoughts becomes possible. Possibility of fine-grained differences in interpretation become understandable when continuous variables and spaces are considered as the counterparts for the distinct symbols that are used in communication. For instance, if person A has a prototypical idea of a specific color having the value $[0.234 \ 0.004 \ 0.678]$ in a color coding scheme, and for the person B the corresponding vector is $[0.232 \ 0.002 \ 0.677]$, it is clear that the communication based on the symbol is successful in spite of the small difference, error, in the interpretation. Actually, it may be still more fruitful to consider a reference relation as a distribution rather than a relation between a symbol and a numerical value or vector. Such an approach enriches strongly the possibility to study fine-grained phenomena of natural language interpretation as opposed to the model theoretical approach.

In the following, Kohonen's Self-Organizing Maps (SOMs) (Kohonen, 1982, 1995) are introduced. The SOMs may provide a sound basis for modeling the general underlying principles of natural language learning and interpretation. The motivation for such a claim is presented in the rest of the paper.

2. Self-Organizing Maps

The SOM is a widely-used artificial neural network model in which learning is unsupervised: no a priori classifications for the input examples are needed. In comparison, the backpropagation algorithm, for instance, requires the examples to consist of input-output pairs. The network architecture of the SOM consists of a set of laterally interacting adaptive processing elements, nodes, usually arranged as a two-dimensional grid called the map. All the map nodes are connected to a common set of inputs. Any activity pattern on the input gives rise to excitation of some local group of map nodes. After learning, the spatial positions of the excited groups specify a mapping of the input onto the map. The learning process is based on similarity comparisons in a continuous space. The result is a system that map similar inputs close to each other in the resulting map. The input may be highly complex multidimensional data like in the real-life applications in speech recognition, image analysis, and process monitoring. (Kohonen, 1995)

2.1. Learning Process in SOM

Assume that some sample data sets have to be mapped onto the array depicted in Fig. 1; a sample set is described by a real vector $x(t) \in \mathbb{R}^n$ where t is the index of the sample, or the discrete-time coordinate. In setting up the neural-network model called the Self-Organizing Map, we first assign to each unit in the array a parameter vector $m_i(t) \in \mathbb{R}^n$ called the codebook vector, which has the same number of elements as the input vector $x(t)$. The initial values of the parameters, components of $m_i(t)$, can be selected at random.

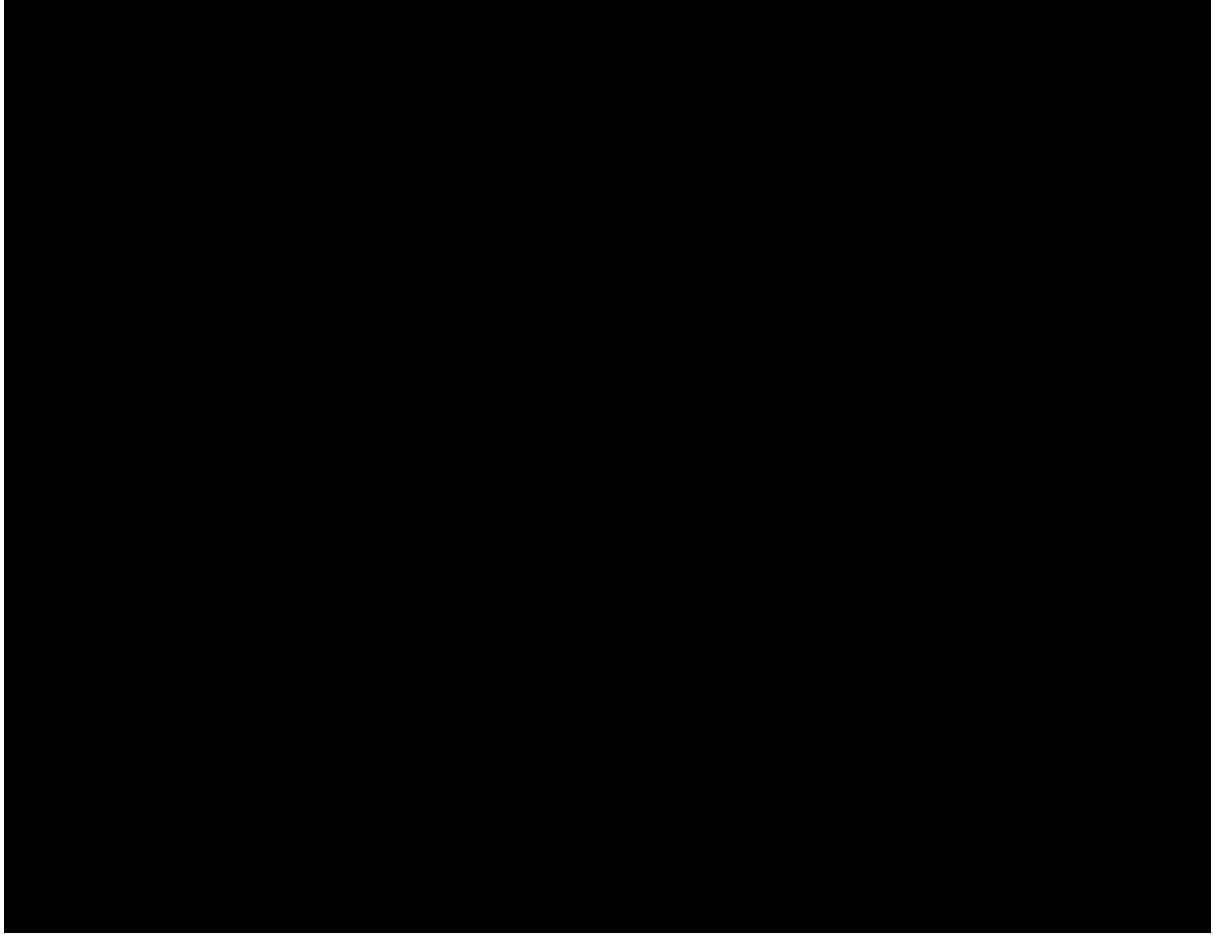


Fig. 1. The basic architecture of the self-organizing map.

The "image" of an input item on the map is defined to be in the location, the $m_i(t)$ of which matches best with $x(t)$ in some metric. The self-organizing algorithm that creates the ordered mapping can be described as a repetition of the following basic tasks:

1. An input vector $x(t)$ (like one row of Table 1) is compared with all the codebook vectors $m_i(t)$. The best-matching unit on the map, i.e., the unit where the parameter vector is most similar to the input vector in some metric, called the winner, is identified.
2. The codebook vectors of the winner and a number of its neighboring units in the array are changed incrementally according to the learning principle specified below.

The basic idea in the SOM is that, for each input sample vector $x(t)$, the parameters of the winner and units in its neighborhood are changed closer to $x(t)$. For different $x(t)$ these changes may be contradictory, but the net outcome in the process is that ordered values for the $m_i(t)$ are finally obtained over the array. If the number of input vectors is not large compared with the number of codebook vectors (map units), the set of input vectors must be presented reiteratively many times. As mentioned above, the codebook vectors may initially have random values, but they can also be selected in an ordered way. Adaptation of the codebook vectors in the learning process takes place according to the following equation:

$$m_i(t+1) = m_i(t) + \alpha(t)[x(t) - m_i(t)] \text{ for each } i \in N_c(t),$$

where t is the discrete-time index of the variables, the factor $\alpha(t) \in [0,1]$ is a scalar that defines the relative size of the learning step, and $N_c(t)$ specifies the neighborhood around the winner in the map array. At the beginning of the learning process the radius of the neighborhood is fairly large, but it shrinks during learning. This ensures that the global order is obtained already at the beginning, whereas towards the end, as the radius gets smaller, the local corrections of the codebook vectors in the map will be more specific. The factor $\alpha(t)$ decreases during learning. Details about the selection of parameters, variants of the map, and thousands of application examples can be found in (Kohonen, 1995). A recent work describing the SOM in data analysis and exploration is (Kaski, 1997).

2.2. Introductory example: from a table to a map

As a very simple illustrative example we demonstrate how different languages can be identified from their written forms by the SOM. Consider Table 1 that is an excerpt from a larger table used in the experiment. It gives the relative frequencies of characters in a small text corpus, selected from each of the languages considered.

Table 1. Part of a table containing relative frequencies of characters.

Language	'a'	'b'	'c'	'd'	'e'	'f'	'g'	'h'	...
English	0.078	0.016	0.019	0.053	0.128	0.018	0.024	0.078	...
Estonian	0.118	0.008	0.000	0.055	0.127	0.002	0.016	0.019	...
Finnish	0.119	0.000	0.000	0.007	0.083	0.000	0.000	0.022	...
French	0.069	0.005	0.035	0.042	0.139	0.010	0.015	0.009	...
Hungarian	0.088	0.020	0.009	0.023	0.102	0.007	0.034	0.021	...
...	

Mutual similarities of the different languages are very hard to deduce from this kind of a table. However, when the row vectors of the table are applied as inputs to a self-organizing map, an outcome of the learning process is that statistically similar vectors (rows) become mapped close to each other in the display. In Fig. 2, similar statistics of characters in different languages are reflected in their "images" being correspondingly close on the map: the nearer the "images" of two languages (like *Danish* and *Norwegian*) are, the more similarities they have in their written features.

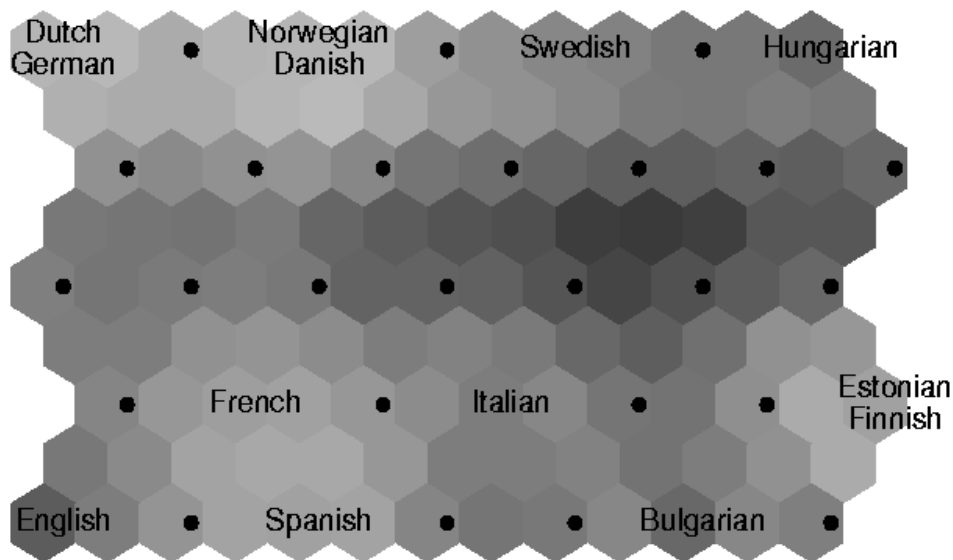


Fig. 2. A self-organized hexagonal map based on the relative frequencies of characters in some languages. The map tends to preserve the topology of the original input space.

The comparison was based on the character set which was used in the original material obtained from the Internet. The aim was not to provide a detailed linguistic analysis and, thus, the input material was randomly selected. Simplifying transliterations of many accented characters had been used the original texts. The phonetic resemblances were not considered. However, although only such a set of crude features derived from a small corpus was used as input data, some basic relationships can be seen on the map. The Romance and Germanic languages fall into areas of their own. Finnish and Estonian have been positioned into the same map node. The dark color denotes a larger distance between the codebook vectors. Thus, one can also find meaningful clustering structure on the map while the previously mentioned areas are separated by the "borderlines".

3. Maps of Natural Language

In languages we use signs and conventions to present a kind of terrain. Like maps, language can open up new worlds: we can innovate through language. We can work out problems, understand and communicate complex ideas by unpacking them in words. But like maps, language also distorts. In using language, we necessarily reduce and group and select. This is exactly what self-organizing maps also do: generalize and organize information, as well speech signal, pictorial images as symbolic input if the context is present.

In most practical applications of the SOM, the input to the map algorithm is derived from some measurements, usually after preprocessing. In such cases, the input vectors are supposed to have metric relations. Interpretation of languages, in their written form, is based on the processing of sequences of discrete symbols. To create a map of discrete symbols that occur within the sentences, each symbol must be presented in the due context.

3.1. Basic Principles of Using SOM in Natural Language Interpretation

The Self-Organizing Map, SOM (Kohonen, 1995) is well suited to serve as the central processing element in modeling natural language interpretation because of the following reasons:

- The SOM algorithm modifies its internal presentation, i.e., the map node vectors according to the external input which enables the adaptation.
- The SOM is able to process natural language input to form "semantic maps" (see Ritter and Kohonen, 1989). Natural language interpretation using the SOM has further been examined, e.g., in (Miikkulainen, 1993; Scholtes, 1993; Honkela, 1995).
- Symbols and continuous variables may be combined in the input, and they are associated by the SOM (see, e.g., Honkela, 1991). Continuous variables may be quantized, and a symbolic interpretation can be given for each section in the possibly very high-dimensional space of perceptual variables.
- Because the SOM is based on unsupervised learning, processing external input without any prior classifications is possible (Kohonen, 1995). The map is an "individual" model of the environment and of the relation between the expressions of the language and the environment.
- The SOM enables creating a model of the relation between the environment and the expressions of the language used by the others. In addition, generalizations of this relations can be formed (Honkela, 1993).

3.2. Creating Maps of Words

It has earlier been shown that the SOM can be applied to the analysis and visualization of contextual roles of words, i.e., similarities in their usage in short contexts formed of adjacent words (Ritter and Kohonen, 1989). In the unsupervised formation of the map of words, each input x consists of an encoded word and its averaged context. Each word in the vocabulary is encoded as a n -dimensional random vector. In our experiments (e.g., Honkela et al. 1995, Kaski et al. 1996, Lagus et al. 1996) n has usually been 90. A more straightforward way to encode each word would be to reserve one component for each distinct word but then, especially with a large vocabulary, the dimensionality of the vector would be computationally intractable. A mathematical analysis of the dimensionality reduction based on random encoding is presented by Ritter and Kohonen (1989).

The basic steps in forming a map of words are listed in the following (see also Fig. 3):

1. Create a unique random vector for each word in the vocabulary.
2. Find all the instances of each word to be mapped in the input text collection. Such words will be called key words in the following. Calculate the average over the contexts of each key word. The random codes formed in step 1 are used in the calculation. The context may consist of, e.g., the preceding and the succeeding word, or some other window over the context. As a result each key word is associated with a contextual fingerprint. If the original random encoding for a single word is, for instance, 90-dimensional, the resulting vector of key word with context is 270-dimensional if one neighboring word from both sides is included. The classifying information is in the context. Therefore the key word part of the vector is multiplied by a small scalar, ϵ . In our experiments ϵ has usually been 0.2.
3. Each vector formed in step 2 is input to the SOM. The resulting map is labeled after the training process by inputting the input vectors once again and by naming the best-matching neurons according to the key word part of the input vector.

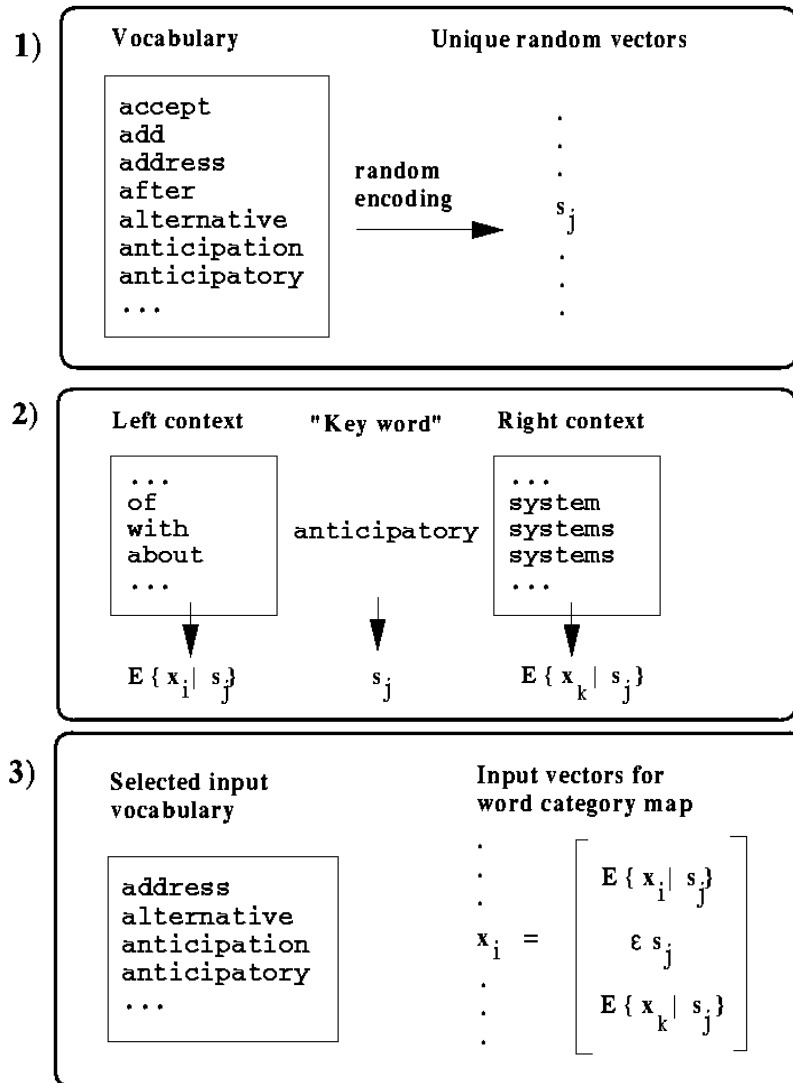


Fig. 3. Creation of input vectors for the word category map.

3.3. Comparisons of Artificial Neural Networks Models

Artificial neural-network models may be classified according to their specific features, such as structure, dynamics, and type of adaptation. The nature and source of the input material and its preprocessing should also be specified. The input data material may be symbolic, numerical, or a combination of the both. The speech signal is a typical example of continuous-valued, non-symbolic input material. On the other hand, written texts are symbolic, although they may be transformed into vectorial form in some way. One method of such a transformation was presented in the previous chapter.

The appearance of natural-language expressions, written or spoken, is sequential. This inherent property raises the question of how to handle time. Elman (1991) presents different ways of representing time. One possibility is to concatenate subsequent "moments" into single input vectors like in the current study. Another possibility is to use networks with recurrent connections to make them operate sequentially. The learning principle of the SOM, however, is not suitable for such recursive operations. On the other hand, the self-organizing map is the only neural-network architecture in which the spatial order between representations

emerges automatically. Such a self-organized order is also one of the main properties of the neural structures in the brain. Neurophysiological studies supporting the existence of SOM like processing principles in the brain have been reviewed in Kohonen (1995).

The basic learning strategies of adaptive systems can be categorized into supervised, reinforced, and unsupervised. In supervised learning, the system is given input-output pairs: for each input there must also exist the "right answer" to be enforced at the output. The system then learns these input-output pairs. The task is not trivial, however, and after the learning period the network is also able to deal with inputs that were not present in the learning phase. This property ensues from the generalizing capabilities of the network. The drawback of supervised learning is the need for correct output in each input example. In some cases, obtaining the output for each input case is a very laborious task especially if a large source material is used.

Whereas supervised learning models are suitable for classification, unsupervised learning can be used for abstraction. The self-organizing map, considered in this article, enables autonomous processing of linguistic input. The map forms a structured statistical description of the material, allowing very nonlinear dependencies between the items. Application of the self-organizing maps to natural language processing has been described, e.g., by Ritter and Kohonen (1989), Scholtes (1993), Miikkulainen (1993), Honkela et al. (1995).

4. Epistemological Considerations

In the following, the problem areas presented in the introduction and the methodological tools provided by the self-organizing maps and the underlying principles, are tied together.

4.1. Emergent Categories

Conceptually interrelated words tend to fall into the same or neighboring node in the word category map (see, e.g., Kaski et al., 1996; Kohonen et al., 1996). Fig. 4 shows the results of a study in which texts from the Usenet newsgroup sci.lang were used as input for the SOM (Lagus et al., 1996). The overall organization of a word category map reflects the syntactic categorization of the words. In the study by Honkela, Pulkki, and Kohonen (1995) the input for the map was the English translation of Grimm fairy tales. In the resulting map, in which 150 most common words of the tales were included, the verbs formed an area of their own in the top of the map whereas the nouns could be found in the opposite corner. The modal verbs were in one area. Semantically oriented ordering could also be found: for instance, the inanimate and animate nouns formed separate clusters.

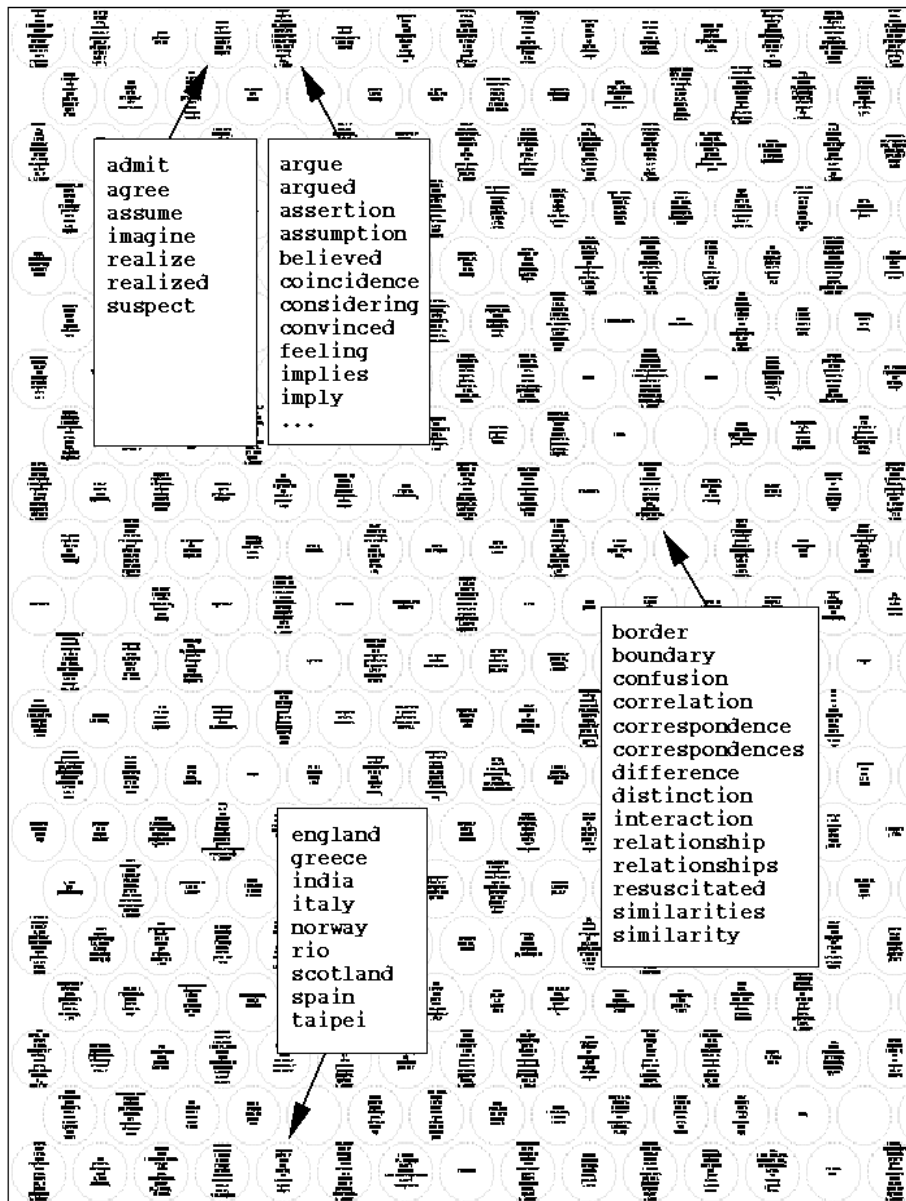


Fig. 4. An illustration of a word category map based on the texts of the Usenet newsgroup sci.lang. The map consists of 15 x 21 nodes most of which contain several words.

An important consideration is that in the experiments the input for the SOM did not contain any predetermined classifications. The results indicate that the text input as such, with the statistical properties of the contextual relations, is sufficient for automatic creation of meaningful implicit categories. The categories emerge during the learning process. The symbol grounding task would, of course, be more realistic and complete if it is were possible to provide also other modalities such as pictorial images as a part of the context information.

4.2. Intersubjectivity and Relativity: Relation Between Discrete and Continuous

Subjectivity is inherent in human natural language interpretation. The nature and the level of the subjectivity has been subject to several debates. For instance, the Chomskian tradition of linguistics as well as the philosophy of language based on predicate logic seem clearly to undermine the subjective component of language processing. In them, the relation between

the "names" of the language and the "objects" or "entities" may be taken as granted and to be unproblematic.

Consider now that we are about to denote an interval of a single continuous parameter using a limited number of symbols. These symbols are then used in the communication between two subjects (human or artificial). In a trivial case two subjects would have same denotations for the symbols, i.e. the limits of the intervals corresponding to each symbol would be identical. If the "experience" of the subjects is acquired from differing sources, the conceptualization may very well differ. This kind of difference in the density pattern is illustrated in Fig. 4a. In Fig 4b, the interval from x_0 to x_4 is divided into smaller intervals according to the patterns. The first subject uses two symbols, A and B, whereas the second subject also utilizes a third one, namely C for the interval between x_1 and x_3 . Thus, if the context (the parameter value in this simplified illustration of Fig. 5) is not present, a communicated symbol may lead to an erroneous interpretation.

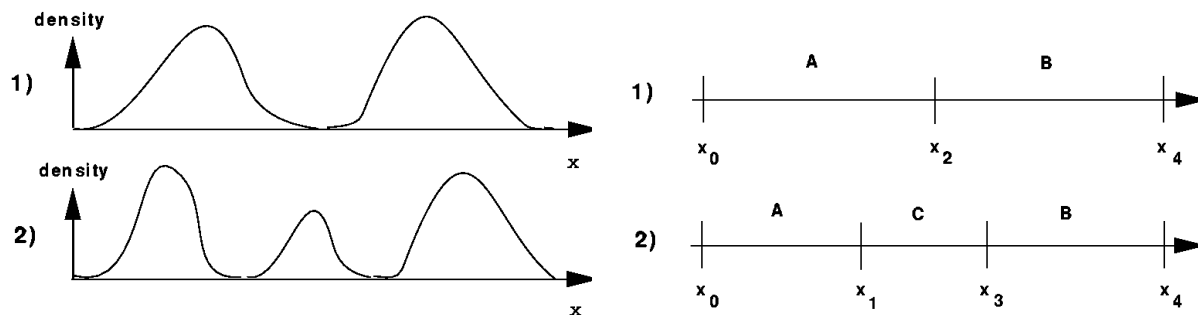


Fig. 5. Illustration of symbols associated to a continuous parameter by two subjects (A and B). In the left, the varying density patterns show the motivation for different conceptualizations of the same phenomenon. The conceptualization, i.e., naming the space, is presented in the right.

One may then ask how to deal with this kind of discrepancies. A propositional level is not sufficient. The key idea is to provide the means for a system to associate continuous-valued parameter spaces to sets of symbols, and furthermore, to "be aware" of the differences in this association and to learn those differences explicitly. These kinds of abilities are especially required by highly autonomous systems that need to communicate using an open set of symbols or constructs of a natural language. This kind of association of set of symbols and a set of continuous parameters is a natural extension or modification of the word category maps (see Honkela, 1993; Honkela and Vepsäläinen, 1991). An augmented input consists of three main parts: the encoded symbol, the context which is the parameter vector in this case, and identification of the utterer or source of the symbol being used. The map nodes associate symbols with the continuous parameters. One node corresponds to an area in the multidimensional space, i.e., a Voronoi tessellation determined by the codebook vector associated with the map node and its neighboring nodes. The relation is one-to-many: one symbol is associated with infinite number of points.

In this kind of mapping, the error (cf. Rosen, 1985), or a kind of relativity is a necessity in communication. One can define the exact reference of a symbol in a continuous space only to a limit that is restricted by several issues, for instance, the limited time available for communication in which the level of intersubjectivity is raised. A common source of context is often not available either. Von Foerster (1972b) has outlined the very basic epistemological

questions that are closely related to the topics of the present discussion. He states, among other things, that by keeping track of the computational pathways for establishing equivalence, "objects" and "events" emerge as consequences of branches of computation which are identified as the processes of abstraction and memorization. In the realm of symbolic logic the invariance and change are paradoxical: "the distinct being the same", and "the same being distinct". In a model that includes both the symbolic description as well as the continuous counterpart, there is no paradox, and the relationship may be computed, e.g., by the self-organizing map.

The previously presented framework also provides a means to consider the relationship between language and thoughts. In the case of colors, one may hypothesize that the perceptual input in the human experience is overwhelming when compared with the symbolic descriptions. Thus, the "color map" is based on the physical properties of the input. On the other hand, abstract concepts are based on the cultural "agreements" and they are communicated symbolically so that the relation to external, physical counterparts is less straightforward. A natural result would be that such concepts are much more prone to subjective differences based on the cultural environment. Even if the original perceptual input is available but it is constantly associated with a systematic classifying symbolic input, the result deviates strongly compared with the case in which the latter information is not available. Von Foerster (1972a) has described the phenomenon and its consequences in the following way: "We seem to be brought up in a world seen through descriptions by others rather than through our own perceptions. This has the consequence that instead of using language as a tool with which to express thoughts and experience, we accept language as a tool that determines our thoughts and experience." In linguistics this kind of idea is referred to as the Sapir-Whorf hypothesis.

4.3. Anticipatory Systems

Rosen (1985) has described anticipatory behavior to be one in which a change of state in the present occurs as a function of some predicted future state. In other words, an anticipatory system contains a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to a later instant.

Music involves the expectation and anticipation of situations on the one hand, and confirmation or disconfirmation of them on the other. Kaipainen (1994) has studied the use of SOMs in modeling musical perception. In addition to the basic SOM, Kaipainen uses a list of lateral connections that record the transition probabilities from one map node to another. The model is based on the specific use of the SOM in which the time dynamics of a process are characterized by the trajectory path on the map. This aspect has been important already in the first application area of the SOMs, namely speech recognition, and more recently in process monitoring. The model of musical perception was tested in three modes called "Gibsonian", "autistic", and "Neisserian". The Gibsonian and autistic are the two extremes regarding the use of anticipation recorded in the trajectory memory: the first model was designed so that it did not use the trajectory information at all. The result was that continuity from one variation to another could not be maintained. On the other extreme, the autistic model was parameterized so that it developed a deterministically strong schematic drive. It began to use its internal representational states, eventually becoming ignorant of the input flow of musical patterns. The intermediate model, denoted as Neisserian, was the one that performed best in musical terms. It was open to the input having at the same time an internal schematic drive, anticipation, which intentionally actualized musical situations rather than just recognizing them as given.

When the relationship between Rosen's formulations and Kohonen's self-organizing maps is considered the following quote may be of interest (Rosen, 1985): "Briefly, we believe that one of the primary functions of the mind is precisely to *organize* percepts. That is, the mind is not merely a passive receiver of perceptual images, but rather takes an active role in processing them and ultimately in responding to them through effector mechanism. The *organization* of percepts means precisely the establishment of relations between them. But we then must admit such relations reflect the properties of the active mind as much as they do the percepts which the mind organizes." It seems that the SOM concretizes this idea. In general, Rosen's point of view may be characterized as physical and biological whereas Kohonen's main results are related to computational, epistemological, neurophysiological, and cognitive aspects. Many basic issues are interrelated, though, including those of error, order and disorder, similarity, and encoding.

5. Conclusions

The SOM-based approach seems to be well in line with the systemic and holistic principles widely adopted in the cybernetic research community while it concentrates on the interaction between processing elements, studies the effects of interactions, is especially suited to study of nonlinear phenomena with mutually dependent variables, and the validation of the results is based on comparison of the behavior of the model with reality.

In natural language processing new promising application areas are arising. It may be concluded that we are gradually learning to understand how we learn to understand.

References

- Elman Jeffrey (1991). Finding Structures in Time. *Cognitive Science*, 16, pp. 96-132.
- von Foerster Heinz (1972a). Perception of the Future and the Future of Perception. *Instructional Science* 1, 1, R.W. Smith, and G.F. Brieske (eds.), Elsevier/North-Holland, New York/Amsterdam, pp. 31-43. (Also appeared in von Foerster, H.: *Observing Systems*, Intersystems Publications, Seaside, CA, 1981, pp. 189-204.)
- von Foerster Heinz (1972b). Notes on an Epistemology for Living Things. BCL Report No. 9.3, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 22 p. (Also appeared in von Foerster, H.: *Observing Systems*, Intersystems Publications, Seaside, CA, 1981, pp. 258-271.)
- Honkela Timo and Vepsäläinen Ari M. (1991). Interpreting Imprecise Expressions: Experiments with Kohonen's Self-Organizing Maps and Associative Memory. *Artificial Neural Networks*, T. Kohonen and K. Mäkisara (eds.), vol. I, 897-902.
- Honkela Timo (1993). Neural Nets that Discuss: A General Model of Communication Based on Self-Organizing Maps. *Proc. ICANN'93, Int. Conf. on Artificial Neural Networks*, S. Gielen and B. Kappen, Springer, London, 408-411.
- Honkela Timo, Pulkki Ville, and Kohonen Teuvo (1995). Contextual relations of words in Grimm tales analyzed by self-organizing map. In F. Fogelman-Soulie and P. Gallinari (eds.) *ICANN-95, Proceedings of International Conference on Artificial Neural Networks*, vol. 2, pp. 3-7. EC2 et Cie, Paris.

Honkela Timo, Kaski Samuel, Lagus Krista, and Kohonen Teuvo. Newsgroup Exploration with WEBSOM Method and Browsing Interface. Report A32, Helsinki University of Technology, Laboratory of Computer and Information Science, January, 1996.

Honkela Timo (1997). Self-Organizing Maps of Words for Natural Language Processing Applications. Proceedings of Soft Computing '97, in print, September 17-19, 1997, 7 p.

Honkela Timo (1997). Emerging categories and adaptive prototypes: Self-organizing maps for cognitive linguistics. Extended abstract, accepted to be presented in the International Cognitive Linguistics Conference, Amsterdam, July 14-19, 1997.

Kaipainen Mauri (1994). Dynamics of Musical Knowledge Ecology - Knowing-What and Knowing-How in the World of Sounds. PhD thesis, University of Helsinki, Helsinki, Finland, Acta Musicologica Fennica 19.

Kaski Samuel (1997). Data Exploration Using Self-Organizing Maps. Dr.Tech thesis. Helsinki University of Technology, Espoo, Finland, Acta Polytechnica Scandinavica, no. 82.

Kaski Samuel, Honkela Timo, Lagus Krista, and Kohonen Teuvo (1996). Creating an order in digital libraries with self-organizing maps. Proceedings of WCNN'96, World Congress on Neural Networks, Lawrence Erlbaum and INNS Press, Mahwah, NJ, pp. 814-817.

Kohonen Teuvo (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, pp. 59-69.

Kohonen Teuvo (1995). Self-Organizing Maps. Springer-Verlag.

Kohonen Teuvo, Kaski Samuel, Lagus Krista, and Honkela Timo (1996). Very large two-level SOM for the browsing of newsgroups. Proceedings of ICANN'96, International Conference on Artificial Neural Networks.

Lagus Krista, Honkela Timo, Kaski Samuel, and Kohonen Teuvo (1996). Self-organizing maps of document collections: a new approach to interactive exploration. E. Simoudis, J. Han, and U. Fayyad (eds.), Proceedings of the Second International Conference on Knowledge Discovery & Data Mining, AAAI Press, Menlo Park, CA, pp. 238-243.

Miikkulainen Risto (1993). Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory. MIT Press, Cambridge, MA.

Ritter Helge and Kohonen Teuvo (1989). Self-organizing semantic maps. Biological Cybernetics, vol. 61, no. 4, pp. 241-254.

Rosen Robert (1985). Anticipatory Systems. Pergamon Press.

Scholtes Jan C. (1993). Neural Networks in Natural Language Processing and Information Retrieval. PhD thesis, University of Amsterdam, Amsterdam.