The Exploration of Legal Text Corpora with Hierarchical Neural Networks: A Guided Tour in Public International Law

Dieter Merkl[1], Erich Schweighofer[2]

[1] Department of Computer Science, Royal Melbourne Institute of Technology
723 Swanston St., Carlton VIC 3053, Australia
E-mail: dieter@mds.rmit.edu.au

[2] Institute of Public International Law, University of Vienna
Research Center for Computers and Law
Universitaetsstrasse 2, A-1090 Vienna, Austria
E-mail: Erich.Schweighofer@univie.ac.at

Abstract

The classification of feature vectors representing the interpretation of legal documents improves the search for similar or related documents, the interpretation of these documents as well as the navigation within the text corpus.
The need for effective approaches of classification is dramatically increased nowadays due to the advent of massive digital libraries containing free-form legal text documents. What we are looking for are powerful methods for the exploration of such libraries whereby the detection of similarities between groups of documents is the overall goal. In other words, methods that may be used to gain insight in the inherent structure of the various items contained in a text archive are needed.
In this paper we present the results from a case study in legal document classification based on an experimental document archive comprising important treaties in public international law. The core task of classification is performed by a non-standard neural network model with a layered architecture consisting of mutually independent unsupervised neural networks. The distinguished features of this learning architecture is the remarkably fast training time combined with the benefit of explicit cluster representation. The access to legal text archives may be enhanced by guided tours providing the means for convenient voyage in an environment of dynamically classified legal documents.

1 Introduction

During the last years we had to witness an ever increasing flood of written information. This situation was anticipated and referred to as the information crisis in law in [Simitis70] and served as the impetus for the development of legal fulltext information retrieval systems some 25 years ago. The coverage of these pioneers of digital libraries is now quite satisfying but more powerful methods for organising, exploring, and searching collections of textual documents are still needed to deal successfully with that information.
The classical way of dealing with textual information is defined by means of keyword-based document representation. These methods may be enhanced with improved information retrieval systems [Turtle95, Turtle92] or hypertext applications (for an overview see [Di Giorgi94]).
A major drawback is that efficient use still requires profound user experience. Therefore, tools providing assistance for explorative search in text collections based on hypertext interfaces are highly needed [Marchionini88, Nielsen93].

Exploration of document archives may be supported by organising the various documents into taxonomies or hierarchies that are in use by lawyers for centuries. In order to reach such a classification a number of approaches are applicable. Among the oldest and most widely used ones we certainly have to mention statistics, especially cluster analysis. The usage of cluster analysis for document classification has a long tradition in information retrieval research and its specific strength and weaknesses are well explored [Salton89, Willet88].
During the last years renewed interest in artificial neural networks can be observed which is at least partly due to increased computing power available at reasonable prices. In general, there is wide agreement that the application of artificial neural networks is suitable in areas that are characterised by noise, poorly understood intrinsic structure, and changing characteristics.
Each of that is present in legal text classification. The noise is imposed due to the fact that no completely satisfying way to represent legal text documents has been found so far. Second, the poorly understood intrinsic structure is due to the non-existence of an authority knowing the contents of each and every legal document. Finally, the changing characteristics of legal document collections are due to the fact that the collections are regularly updated.

Proceedings of the Int'l Conference of Artificial Intelligence and Law (ICAIL'97). Melbourne, VIC., June 30 - July 3, 1997, ACM Press.

From the wide range of proposed architectures of artificial neural networks we regard the unsupervised models as especially well suited for text classification. This is due to the fact that in a supervised environment one would have to define proper input-output-mappings anew when the text archive changes; and such changes should be expected to happen quite frequently. By input-output-mapping we refer to the manual assignment of documents to classes which, obviously, is only possible when assuming the availability of considerable insight in the structure of the text archive. Contrary to that, in an unsupervised environment it remains the task of the artificial neural network to uncover the structure of the document archive. Hence, the unrealistic assumption of providing proper input-output-mappings is obsolete in an unsupervised environment.

A number of successful applications of unsupervised neural networks to information retrieval have already been reported in literature [Honkela95, Lagus96, Lin91, Merkl95a, Merkl95b, Merkl96]. One of the most distinguished unsupervised neural networks certainly is the self-organising map [Kohonen82]. It is a general unsupervised tool for ordering high-dimensional statistical data in such a way that alike input items are mapped close to each other. In order to use the self-organising map to cluster text documents, we represent the various texts as the histogram of its words and enhance this description by using context-sensitive as well as meta rules. With this data, the artificial neural network performs the classification task in a completely unsupervised fashion.
In this paper we introduce the classification of text documents by means of hierarchically arranged self-organising maps. A hierarchical arrangement has been chosen in order to enable the true establishment of a document taxonomy. Moreover, the hierarchical arrangement leads to remarkably fast training times of the neural network.

The material presented in the remainder of this paper is organised as follows. In Section 2 we provide a brief description of the overall system *KONTERM workstation*. Section 3 we give the details of the neural network we used for document clustering. Section 4 contains an exposition of the highly encouraging training results. In Section 5 we give a review of related work in the field of neural network applications in law and information retrieval. Finally, we provide some conclusions in Section 6.

2 KONTERM Workstation

The aim of the project *KONTERM workstation* is to provide a hybrid application of methods of legal knowledge representation assisting lawyers in their task of managing present high quantities of legal information contained in natural language documents. Besides legal information retrieval and hypertext, a main aim of the *KONTERM workstation* is the automatic analysis of text corpora and the semi-automatic generation of the document description. The document classification task is part of that goal. The documents are segmented into document parts, articles, paragraphs and sentences and are transformed into HTML documents. Legal concepts are represented in a knowledge base of descriptors, probabilistic context-sensitive rules and meta rules. Context-sensitive rules are linguistic templates allowing the recognition of complex concepts in legal documents. The wording of rules is facilitated allowing also probabilistic expressions. Meta rules represent a concept that must be defined as a combination of rules occurring in the same document or section of a document. This method allows the automatic detection of knowledge in legal documents. Thus, the various documents are represented as feature vectors of the form $x=\{t_1, ..., t_m, c_1, ..., c_n, m_1, ..., m_0\}$. The t_i represent terms extracted from the fulltext of the document, the c_i are the context-sensitive rules, and the m_i represent the meta rules associated with the document. Vector space model, cluster analysis and the self-organising map of Kohonen are efficient tools in building the knowledge base. The description of documents is done by matching documents with the knowledge base. This automatic generation of summaries and meta information of the documents is presented in hypertext structure. Hypertext links are generated automatically from concepts to documents, from documents to concepts, from text corpus to documents, from document descriptions to documents etc. The document space can be described using cluster analysis or neural network. Details may be found in [Schweighofer96a, Schweighofer96b].
The analysis of the concept space and the document space by the Kohonen map has been presented in detail in [Schweighofer95]. In this paper we describe the results from document classification by using a non-standard neural network model instead. The architecture of this model comprises a layered arrangement of mutually independent self-organising maps. The major benefits of this new model which justify its utilisation are a substantially reduced training time as compared to self-organising maps as well as a explicit and model inherent cluster separation.
We feel that the approach followed within the *KONTERM workstation* project represents a highly useful form of approximation of the legal language. The already existing vectors for the formalisation of natural language text segments and documents are used as input for the neural network.

The detection of word senses is a central issue of the *KONTERM workstation*. In practice we used the results obtained from statistical cluster analysis although the results achieved with the self-organising maps were slightly better. As the reason we refer to the very long time needed to train the self-organising maps especially when given long document descriptions that are natural in a real working environment. With our recent work within the *KONTERM workstation* project we directed specific focus on the exploration of legal document spaces. More precisely, we are interested in the effects of enhanced document representations as well as in efficient and effective ways of cluster formation.

For further tests, we concentrate on the description of the document space. In practice, there exists high need for efficient methods of clustering similar documents. The document descriptions are produced automatically as well as the vector with weighted indexation of the components.

3 Hierarchical Feature Maps

The key idea of hierarchical feature maps as outlined in [Miikkulainen92, Miikkulainen93] is to apply a hierarchical arrangement of several layers containing two-dimensional self-organising maps [Kohonen82, Kohonen95]. The arrangement may be characterised as having the shape of a pyramid as shown in Figure 1. The hierarchy among the maps is established as follows. For each output unit in one layer of the hierarchy a two-dimensional self-organising map is added to the next layer. The training of each single self-organising map follows the basic self-organising map learning rule.

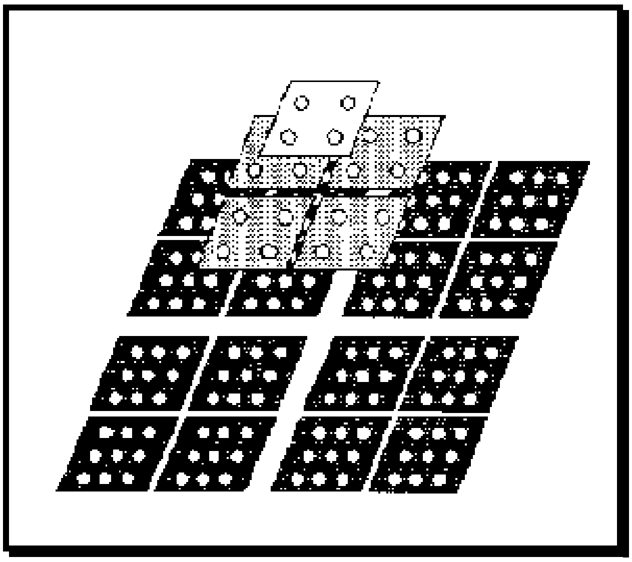

Figure 1: Hierarchical Feature Map

The learning process of self-organising maps can be seen as a generalisation of competitive learning although this is historically incorrect as the self-organising map was presented earlier in literature. The key idea of competitive learning [Rumelhart86] is to adapt the unit with the highest activity level with respect to a randomly selected input pattern in a way to exhibit an even higher activity level with this very input in future. Commonly, the activity level of an output unit is computed as the Euclidean distance between the unit's weight vector and the actual input pattern. Hence, the so-called winning unit, i.e. the winner in short, is the output unit with the smallest distance between the two vectors. Adaptation takes place at each learning step and is performed as a gradual reduction of the difference between the respective components of input and weight vector. The degree of adaptation is guided by a so-called learning-rate, gradually decreasing in the course of time.

As an extension to competitive learning, units in a time-varying and gradually decreasing neighbourhood around the winner are adapted, too. Pragmatically speaking, during the learning steps of self-organising maps a set of units around the actual winner is tuned towards the currently presented input pattern. This learning rule leads to a clustering of highly similar input patterns in closely neighbouring parts of the grid of output units. Thus, the learning process ends up with a topological ordering of the input patterns. One might say that self-organising maps represent a spatially smooth neural version of k-means clustering [Ripley96] where k is equal to the number of output units.

More precisely, the four steps of the learning process may be outlined as given below.

1. Selection of an input $x(t)$.

2. Calculation of the distances between weight vectors and the input vector according to $D_i(t) = || x(t) - m_i(t) ||$. In this expression m_i refers to the weight vector of unit i and $|| \cdot ||$ represents the Euclidean vector norm. As usual in time-discrete notation, t represents the time-stamp of the current learning iteration.
3. Determination of the winner c: $m_c(t)= min_i(D_i(t))$.
4. Adaptation of the weight vectors. In particular we use the following learning rule: $m_i(t+1) = m_i(t) + \varepsilon(t) \cdot \psi_{c,i}(t) \cdot [x(t) - m_i(t)]$. In this formula, $\varepsilon(t)$ represents a time-varying gain term, i.e. learning-rate, decreasing in the course of time. $\psi_{c,i}(t)$ is a time-varying neighbourhood-function taking into account the distance between the winner c and unit i within the output space. The task of the neighbourhood-function is to impose a spatial structure on the amount of weight vector adaptation in that the function computes values in the range of [0, 1], depending on the distance between the unit in question and the winner. By means of this function, units in close vicinity to the winner are provided with a larger value of the neighbourhood-function and thus, are adapted more strongly than units that are farther away from the winner. Please refer to [Merkl95a, Merkl95c] for the exact details of the implementation, especially what concerns the neighbourhood-function $\psi_{c,i}(t)$. We feel that the inclusion of the exact realisation here would require too lengthy a discussion which is not necessary to understand the learning process in general.

Please note that by gradually reducing the learning-rate and neighbourhood-range, the learning process will converge towards a stable state. The stable state is reached when no further changes to the various weight vectors are observed. In practice, however, the learning process may be terminated earlier, namely at the time when no further variation within the process of winner selection is detected. In other words, the training process may be terminated when each input data is mapped repeatedly onto the same unit.

Generally, the training of a hierarchical feature map is performed sequentially from the first layer, i.e. one self-organising map, downwards along the hierarchy. The maps on each layer are trained according to the standard self-organising map learning process. As soon as the first layer map has reached a stable state, training continues with the maps of the second layer. Within the second layer, each map is trained with only that input data assigned to the corresponding unit of the first layer map. Moreover, the length of the input vectors may be reduced by omitting those vector components that are equal in the original input vectors. Such an omission is possible because the corresponding features of the input vector are already represented at the next higher level in the hierarchy. Due to this reduction in size of the input vectors the time needed to train the maps is reduced as well. The training of the second layer is completed when every map has reached a stable state. Analogously, the same training procedure is utilised to train the third and any subsequent layers of self-organising maps.

An interesting property of hierarchical feature maps is the tremendous speed-up as compared to the self-organising map. An explanation that goes beyond the obvious dimension reduction of the input data from one level to the next may be found by an investigation of the general properties of the self-organising learning process. In self-organising maps, the units that are subject to adaptation are selected by using a neighbourhood function. It is common practice that in the beginning of the learning process almost the whole map is affected by the presentation of an input vector. Thus, the map is forced to establish large clusters of similar input data in the beginning of learning. The neighbourhood size decreases gradually during the learning iterations leading to ever finer distinctions within the clusters whereas the overall topology of the cluster arrangement is maintained. However, in the single-level architecture the self-organising process of each cluster interferes with the self-organisation of its topologically neighbouring clusters. Especially units along the boundaries tend to be occasionally modified as belonging to one or another cluster. This interference is one reason for the rather time-consuming self-organising process. Contrary to that, such an interference is dramatically reduced due to the architecture of hierarchical feature maps. The topology of the high-level categories is depicted in the first layer of the hierarchy. Each of its subcategories are then independently organised on separate maps at lower layers within the hierarchy. These maps in turn are free from maintaining the overall structure since this structure is already determined by the architecture of the hierarchical feature map. To conclude, much computational effort is saved due to the fact that the overall structure of the clusters is maintained in terms of the architecture rather than in terms of the learning process. However, the decision on the best size of the various maps as well as on the depth of the hierarchy remains a non-trivial problem requiring some insight in the structure of the underlying document archive.

4 Evaluation: A Guided Tour in Public International Law

The test environment for our approach comprises a text corpus consisting of 100 of the most important treaties in public international law (5 megabytes of text). Text corpus and the automatically produced document description are available at the KONTERM server (`http://www.ifs.univie.ac.at/intlaw/konterm/konterm.htm`). The input to the neural network is represented by feature vectors, each describing the legal interpretation of a particular document. More precisely, a feature vector consists of descriptors, context-sensitive rules, and meta-rules as described in Section 2. In order to reflect the different importance of these three parts of the document description, the specific values of the features are set to 1 for descriptors, to 2 for context-sensitive rules, and to 3 for meta-rules if the respective feature is present in the description of the document at hand. If the feature is not present then a value of 0 is inserted in the respective component of the feature vector. The length of each individual feature vector totals up to 1625 components. These vectors are produced automatically by the *KONTERM workstation*.

The first test phase consisted of a comparison between cluster analysis and the self-organising map [Schweighofer95, Schweighofer96a]. One of the serious shortcomings of cluster analysis is the fact that a high number of documents is not assigned to a particular cluster but rather considered as a cluster on its own. The results can be seen at the KONTERM server. In comparison to that, the Kohonen map gives a much better overview about the document space showing good hills and regions. Taking up the geographical terms used in [Merkl94], by *hill* we refer to a strong concentration of documents with the same (or highly similar) contents whereas a *region* represents a weak relationship between similar documents.

The long training time of the self-organising map is problematic for legal applications. For the training of a self-organising map with input data comparably complex as in the present case study, we would expect the need of about 20 hours or more, depending obviously on the chosen size of the neural network in terms of neurons, on a high-end workstation. Even the description of small in-house collections may require substantial training time.
Contrary to that, one of the striking arguments in favour of the hierarchical feature map is their tremendous speed-up what concerns the training time. Just to give an idea of the speed-up, the training result presented thereafter has been achieved after approximately 20 minutes training time on a SPARC-20 workstation. This confirms the remarks on the efficiency of the hierarchical feature map learning process as given in Section 3.

As another important advantage of the hierarchical feature map we have to mention its inherent cluster segmentation. This behaviour is enabled because of the network's hierarchical architecture with mutually independent neural networks within the various layers. Thus, the hierarchical feature map may provide results that are comparable to the notion of a cluster in cluster analysis. Contrary to that, the various classes are less intuitively observable in self-organising maps because of the lacking border between different classes in the visual representation. We have to note, however, that substantial contemporary research is dedicated to that insufficiency [Cottrell96, Merkl97, Ultsch93].

We performed a series of training sessions with the hierarchical feature map. One representative result is presented thereafter. In particular, we used the following set-up: A small 2x2 network for fast initial classification at the top layer, 4x4 networks that provide enough space for fine-grained classification in layers 2 and 3. Due to space restrictions we cannot present the complete training result here in full detail, we rather have to focus on the discussion of some of the document clusters formed by the neural network.

The result of the training process is the arrangement of the treaties in the document space. The output consists of the number of the row and the column of the respective documents within the hierarchical feature map. The headings of the various clusters were added by the authors during the evaluation of the maps.

Due to the highly limited space in the figures, we cannot present the full titles of the documents. The full titles of the documents are available at our KONTERM server.

Turning now to the classification result, we present the arrangement of treaties within the 2x2 top-layer map in Figure 2. The contents of the fourth cluster is further explored in Figure 3. Finally, we present three clusters of its third-layer-map in Figure 4.

Humanitarian Law Hague & Geneva Conventions Gas Convention	***Human Rights*** UN Covenants, European Convention, American Convention, Declaration on Human Rights, Vienna Declaration
Environment law New treaties and instruments (UNCED), Ozone Convention Geneva Convention	***Roster*** General Other human rights treaties Environment law Space law Economic law Copyright law

Figure 2: Top Layer Map

The 2x2 top layer map lists fine clusters concerning humanitarian law, environment law and human rights. The fourth cluster named *roster* contains a wider range of treaties. The contents of this cluster is further expanded within its corresponding middle layer map as depicted in Figure 3.

Arms Control chem-wepa con-weap un-chart	***Cultural Property*** cult-pro unesco54 gc-pii	***Cultural Heritage*** a_herita valletta	***Law of the Sea (Geneva)*** contshel high-sea terr-sea
Procedural Law icj refug-p	***Diplomatic Law*** consul diplomat	***Genocide*** genocide sea-disp	
Economic Law slavery gatt t-timber wto	***Human Rights*** torture women treaties	***„Old“ Environment Treaties*** nature40 fauna33	***Waste Disposal*** basel-c waste-af
Space Law moon2 space-re	***$ENVIRON_2$***	***$ENVIRON_1$***	***ROSTER***

Figure 3: Second Layer Map of the Roster Cluster

On a closer look at Figure 3 we conclude that fine classes exist concerning the preservation of cultural heritage [0,2], waste disposal [2,3], law of the sea (Geneva) [0,3], environment treaties ([3,2] and [3,1]), „old“ environment treaties [2,2], diplomacy [1,1], and space law [3,0]. The main topics of the other classes are arms control conventions [0,0], protection of cultural property [0,1], procedural law [1,0] and economic law [2,0]. The numbers in brackets indicate the position of the respective documents within the map representation where the count of rows and columns start with the value 0.

For the sake of honesty, however, we have to confess that some inconsistencies might be found in the results. An example is the inclusion of the UN Charter in the cluster of arms control agreements [0,0] but these run-aways can be detected easily looking closely to the classes of next layer.

ENVIRON$_1$	
antark2	1 1
atlantic	0 1
blacksea	0 3
bonn-c	3 0
cites	3 1
env-ia	3 3
genev-p1	2 2
marpol	0 2
mediter	1 3
montreal	1 0
watercou	2 3

ENVIRON$_2$	
artarkt	2 1
bio-wep	0 2
enmod	0 1
moon1	1 3
npt	0 0
nuc-tb	2 0
space-as	3 2
space-li	2 3
spnfzt	3 0
whaling	3 3

ROSTER	
bern-c	3 1
copyr71	2 3
fish58	0 3
icao-pen	2 0
pact2-p1	0 2
paris-k	3 0
patent	3 2
prostitu	0 0
ri-women	1 1
slavery2	0 1
uncsg	2 1

Figure 4: Sample Third Layer Clusters

For the three large classes contained in Figure 3 we give an expanded representation in Figure 4 where we indicate the position of the various treaties within the 4x4 third-layer map. The subclasses gives some information about the interrelations between the documents. The potential of the proper treatment of noisy inputs is also evident. The classes concerning environment law differ in their emphasis on „basic“ environment law (ENVIRON$_1$) or on „mixed“ topics regulating also other areas of international law (ENVIRON$_2$). In the roster class, the documents concerning copyright law are concentrated (subclasses [2,3], [3,0], [3,1], [3,2]) as well as particular human rights treaties ([0,0], [0,1], [0,2], [1,1]).

The result of this classification cycle may be used as the framework of a guided tour in international law through digital libraries. The user defines the framework of interpretation of the text corpus using the tools of descriptors, context-sensitive rules and meta rules. The hierarchical feature map can classify the documents represented as feature vectors without tuning of statistical parameters. The result is a speed-up of the search for similar or related documents. Please note that in another application on reservations to human rights treaties the analysis of the documents was much more detailed than in this test environment.

Given the incorporation of these results in a legal information system allowing hypertext access to the stored documents, as is the case within the *KONTERM workstation* project, the user may now browse through the hierarchy imposed on the various classes of documents where the individual documents might be stored either on local or remote sites. Apart from this homage to user-friendly information access, the time-consuming tasks of text classification and document interrelation can be performed in a highly automated form.

Finally, we want to stress that this approach to classification fits nicely within the current framework of research in data mining [Fayyad96, Mannila97]. Moreover, due to the fast training time this approach is highly promising in a time of huge collections of digital libraries available on the Internet, accessible via the World-Wide-Web.

5 Related Work

Neural networks found some attention for encapsulation of legal knowledge. This might be due to the fact that knowledge-based approaches were awarded only limited success in highly narrow domains. Three main streams of research may be observed.

First, neural networks are trained to represent vague concepts according to some predefined input-output-mapping [Bench-Capon93, Groendijk92, Opdorp91, Philipps89].

Second, neural networks are used to perform a spreading-activation during retrieval as another paradigm to describe the relation between terms and documents or queries [Belew87, Rose89, Rose94]. Comparable work concerning general purpose information retrieval is described in [Wilkinson91, Keane96].

Third, concept or document spaces are described by neural networks adhering to the unsupervised learning paradigm. The paper of [Lin91] marks the first attempt to utilise self-organising maps for information retrieval. Similar to our approach, the authors rely on self-organising maps. In this paper, however, the document representation is made up from 25 manually selected index terms and is thus not really realistic. In [Merkl95a] this line of research is continued, yet this time with full-text indexed documents. In the area of legal information processing, the self-organising map has been used in [Merkl94, Schweighofer95] for exploratory analysis of judicial documents. Among the shortcomings of self-organising maps one certainly has to mention the remarkable computational demands of the learning rule. Possibilities to increase the speed of learning may be found in the learning rule [Merkl95d] and the representation of the documents [Merkl95c]. Only recently, a number of papers have been published on the utilisation of the self-organising map for large-scale document representation [Honkela95] based on the seminal work of [Ritter89] and subsequent classification [Lagus96].

On balance, unsupervised neural networks have proven to be remarkably successful tools for explorative analysis of text archives. A number of studies have shown that unsupervised neural networks are highly capable in uncovering similarities between text documents.

6 Conclusions and Future Work

A classification cycle may be used as the framework of a guided tour in international law through digital libraries. The classification of feature vectors representing the interpretation of a legal document improves the search for similar or related documents, the interpretation of these documents as well as the navigation within the text corpus.

Neural network applications in law have to face the complexity of law requiring a high number of neurons and result consequently in long training times. As in other applications, the time-consuming training-process is commonly regarded as the major obstacle of real-world large-scale neural network applications.

In order to cope with the above mentioned obstacle, we suggest the utilisation of a non-standard neural network model with layered architecture consisting of mutually independent unsupervised learning networks. The effect of this hierarchical structure is a tremendous speed-up as compared to the basic model of self-organising maps. The examples drawn from our document collection on public international law show the feasibility of our approach. The quality of the hierarchical feature map is superior to the self-organising map because of the self-explaining character of the classes and their explicit representation. We are currently working towards an improved document representation to overcome these inconsistencies. Future work will thus concentrate on an evaluation of the effect of using different document representation schemes. Large scale tests in this direction are now possible due to the fast training time.

Possible applications are the mentioned exploration of legal text corpora or a classification-based approach of conceptual information retrieval. This form of document classification may lead to an easy-going description of legal databases providing the basis for hypertext links between the various documents. Then, the access to legal text archives may no longer be restricted by the corset of Boolean logic and Boolean search expressions but rather may be enhanced by guided tours providing the means for convenient voyage in an environment of dynamically classified legal documents.

Acknowledgements

This research project has been partly financed by the Jubiläumsfonds of the Oesterreichische Nationalbank, research project no. 4941.

References

[Belew87]: R. Belew, A Connectionist Approach to Conceptual Information Retrieval, Proc. of the Int. Conference on Artificial Intelligence and Law, 1987

[Bench-Capon93]: T. Bench-Capon, Neural Networks and Open Texture, Proc. of the Int. Conference on Artificial Intelligence and Law, 1993

[Cottrell96]: M. Cottrell and E. de Bodt, A Kohonen map representation to avoid misleading interpretations, Proc. of the European Symposium on Artificial Neural Networks (ESANN'96), Brugge, Belgium, 1996

[Di Giorgi94]: R. M. di Giorgi and R. Nannucci (ed), Hypertext and Hypermedia in the Law, Special Issue, Informatica e Diritto, Edizioni Scientifiche Italiane, 1994

[Fayyad96]: U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining; Menlo Park, CA, AAAI Press, 1996

[Groendijk92]: C. Groendijk, Neural Schemata in Automated Judicial Problem Solving, Proc. of the Int. Conf. JURIX, 1992

[Honkela95]: T. Honkela, V. Pulkki and T. Kohonen, Contextual relations of words in Grimm tales analyzed by self-organizing maps, Proc. of the 5th Int. Conf. on Artificial Neural Networks (ICANN'95), Paris, France, 1995

[Keane96]: S. Keane, V. Ratnaike and R. Wilkinson, Hierarchical News Filtering, Proc. of the Int. Conf. on Practical Aspects of Knowledge Management, Basel, Switzerland, 1996

[Kohonen82]: T. Kohonen, Self-organized formation of topologically correct feature maps, Biological Cybernetics, Vol. 43, 1982

[Kohonen95]: T. Kohonen, Self-organizing maps, Springer-Verlag, Berlin, 1995

[Lagus96]: K. Lagus, T. Honkela, S. Kaski and T. Kohonen, Self-Organizing Maps of Document Collections: A New Approach to Interactive Exploration, Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, 1996

[Lin91]: X. Lin, D. Soergel and G. Marchionini, A self-organizing semantic map for information retrieval, Proc. Int. Conf. on Research and Development in Information Retrieval, Chicago, IL, 1991

[Mannila97]: H. Mannila, Methods and Problems in Data Mining, Proc. of the Int. Conf. on Database Theory, Dephi, Greece, 1997

[Marchionini88]: G. Marchionini and B. Shneiderman, Finding Facts vs. Browsing in Hypertext Systems, IEEE Computer Vol. 21, No. 1, 1988

[Merkl94]: D. Merkl, E. Schweighofer and W. Winiwarter, CONCAT: Connotation Analysis of Thesauri

Based on the Interpretation of Context Meaning, Proc. of the Int. Conference on Database and Expert Systems Applications, Springer, Berlin, 1994

[Merkl95a]: D. Merkl, A connectionist view on document classification, Proc. of the 6th Australasian Database Conf. (ADC'95), Adelaide, Australia, 1995

[Merkl95b]: D. Merkl, Content-based software classification by self-organization, Proc. IEEE Int. Conf. Neural Networks (ICNN'95), Perth, Australia, 1995

[Merkl95c]: D. Merkl, Content-Based Document Classification with Highly Compressed Input Data, Proc. of the 5th Int. Conf. on Artificial Neural Networks (ICANN'95), Paris, France, 1995

[Merkl95d]: D. Merkl, The effect of lateral inhibition on learning speed and precision of a self-organizing map, Proc. Australian Conf. on Neural Networks, Sydney, Australia, 1995

[Merkl96]: D. Merkl and A M. Tjoa, Data Mining in Large Free-Text Document Archives, Proc. of the Int. Symposium on Cooperative Database Systems for Advanced Applications, Kyoto, Japan, 1996

[Merkl97]: D. Merkl and A. Rauber, On the similarity of eagles, hawks, and cows: Visualization of similarity in self-organizing maps, Proc. Int. Workshop Fuzzy-Neuro-Systems'97, Soest, Germany, 1997

[Miikkulainen92]: R. Miikkulainen, Trace Feature Map: A Model of Episodic Associative Memory, Biological Cybernetics, Vol. 66, 1992

[Miikkulainen93]: R. Miikkulainen, Subsymbolic Natural Language Processing: An integrated model of scripts, lexicon, and memory, MIT-Press, Cambridge, MA, 1993

[Nielsen93]: J. Nielsen, Hypertext and Hypermedia, Academic Press Professional, Boston, 1993

[Opdorp91]: G. J. van Opdorp, R. F. Walker, C. Schrickx, C. Groendijk, P. H. van den Berg, Networks at Work: A Connectionist Approach to Non-Deductive Legal Reasoning, Int. Conference on Artificial Intelligence and Law, 1991

[Philipps89]: L. Philipps, Are Legal Decisions Based on the Application of Rules or Prototype Recognition? Legal Science on the Way to Neural Network, Pre-proc. of the 3rd Int. Conf. on Logica, Informatica, Diritto, Florence, 1989

[Ripley96]: B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, 1996

[Ritter89]: H. Ritter and T. Kohonen, Self-Organizing Semantic Maps, Biological Cybernetics, Vol. 61, 1989

[Rose89]: D. E. Rose and R. K. Belew, Legal Information Retrieval: A Hybrid Approach, Proc. of the Int. Conference on Artificial Intelligence and Law, 1989

[Rose94]: D. E. Rose, A Symbolic and Connectionist Approach to Legal Information Retrieval, Lawrence Erlbaum, Hillsdale, 1994

[Rumelhart86] D. E. Rumelhart and D. Zipser, Feature discovery by competitive learning. in: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I: Foundations, D. E. Rumelhart and J. L. McClelland, Eds., Cambridge, MA, MIT-Press, 1986

[Salton89]: G. Salton, Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley, Reading, MA, 1989

[Schweighofer95]: E. Schweighofer, D. Merkl and W. Winiwarter, Information Filtering: The Computation of Similarities in Large Corpora of Legal Texts, Proc. of the Fifth International Conference on Artificial Intelligence and Law, 1995

[Schweighofer96a]: E. Schweighofer, Wissensrepräsentation und automatische Textanalyse im Völker- und Europarecht (Knowledge Representation and Automatic Text Analysis in Public International Law and European Law), Habilitationsschrift (inaugural dissertation), University of Vienna 1996 (in German)

[Schweighofer96b]: E. Schweighofer and D. Scheithauer, The Automatic Generation of Hypertext Links in Legal Documents, Proc. of the Int. Conf. Database and Expert Systems Applications, Springer, Berlin, 1996

[Simitis70]: S. Simitis, Informationskrise des Rechts und Datenverarbeitung (Information Crisis of Law), Müller, Karlsruhe, 1970 (in German)

[Turtle92]: H. R. Turtle and W. B. Croft, A comparison of text retrieval models, Computer Journal, Vol. 35, No. 3, 1992

[Turtle95]: H. Turtle, Text Retrieval in the Legal World, AI & Law, 1995

[Ultsch93]: A. Ultsch, Self-Organizing Neural Networks for Visualisation and Classification, Information and Classification - Concepts, Methods and Applications, O. Opitz, B. Lausen, R. Klar, Eds. Berlin, Springer, 1993

[Wilkinson91]: R. Wilkinson and P. Hingston, Incorporating the vector space model in a neural network used for information retrieval, Proc. of the Int. Conf. on Research and Development in Information Retrieval, Chicago, IL, 1991

[Willet88]: P. Willet, Recend trends in hierarchic document clustering: A critical review, Information Processing & Management, Vol. 34, 1988