En Route to Data Mining in Legal Text Corpora: Clustering, Neural Computation, and International Treaties

Dieter Merkl

Department of Computer Science
Multimedia Database Systems Group
Royal Melbourne Institute of Technology
723 Swanston Street
Carlton, VIC 3053, Australia
dieter@mds.rmit.edu.au

Erich Schweighofer

Institute of Public International Law
Research Center for Computers and Law
University of Vienna
Universitätsstraße 2
A-1090 Vienna, Austria
Erich.Schweighofer@univie.ac.at

Abstract

The huge amount of data in legal information systems requires a new generation of techniques and tools to assist lawyers in analyzing data and finding critical nuggets of useful knowledge. A promising approach for data mining in legal text corpora is classification. What we are looking for are powerful methods for the exploration of such libraries whereby the detection of similarities between documents is the overall goal. These methods may be used to gain insight in the inherent structure of the various items contained in a text archive. 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 essentials of our approach are the usage of a vector space document representation and the utilization of an unsupervised artificial neural network for document classification.

1 Introduction

During the last years we witnessed an ever increasing flood of written information. This situation was anticipated and referred to as the information crisis in law in [30] 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 organizing, exploring, and searching collections of textual documents are still needed to deal successfully with that huge amount of information.

Within the legal community, the traditional method of manual analysis and interpretation is still widely used. The classical way of dealing with digital textual information is defined by means of a keyword-based document representation. These methods may be enhanced with improved information retrieval systems [31, 32] or hypertext applications [5]. 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 [14, 20].

Exploration of document archives may be supported by organizing the various documents into taxonomies or hierarchies that have been used by lawyers for centuries. In order to reach such a classification a number of approaches is 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 [26, 35].

The process of finding useful patterns in data is nowadays referred to as knowledge discovery in databases (KDD) or data mining [6]. Data mining is the process of applying specific algorithms for extracting patterns (models) from data. In the spirit of [7, 13] we regard clustering as one of the essential techniques during a data mining process in order to enable the discovery of useful data patterns.

Due to the fact that the various documents comprising the text archive do not lend themselves to immediate analysis some pre-processing with intellectual input is necessary [3, 29]. This process of data preparation, data selection, data cleaning and the incorporation of appropriate prior knowledge ensures the quality of the derived knowledge.

The area of data mining in legal text corpora is characterized by noise, poorly understood intrinsic structure, and changing characteristics. 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. In general, there is wide agreement that the application of artificial neural networks is suitable in such areas. Increased computing power available at reasonable prices has led to a renewed interest in artificial neural networks. 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 being able to provide proper input-output-mappings is obsolete in an unsupervised environment.

A number of successful applications of unsupervised neural networks to information retrieval has already been reported in literature [11, 12, 15, 16]. One of the most distinguished unsupervised neural network certainly is the self-organizing map [9]. 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-organizing 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.

The material presented in the remainder of this paper is organized as follows. In Section 2 we provide a brief description of the overall system *KONTERM workstation*. In 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 *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 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 to detect 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 $x = (t_1, \ldots, t_m, c_1, \ldots, c_n, m_1, \ldots, m_o)^T$. 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-organizing 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 [27, 28]. In this paper we describe the results from document classification by using a neural network model. We feel that the approach followed within the *KONTERM workstation* project represents a highly useful form of approximation of the legal document. The already existing vectors for the formalization 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 *KONTERM workstation*. In practice we used the results obtained from statistical cluster analysis although the results achieved with the self-organizing maps were slightly better. As the reason we refer to the very long time needed to train the self-organizing 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 Self-Organizing Feature Maps

Within our approach we utilize an artificial neural network adhering to the unsupervised learning paradigm, namely self-organizing feature maps [9, 10]. The architecture of self-organizing feature maps consists of a layer of input units and a grid of output units. In the case of our application we used a two-dimensional plane of output units. Each output unit is connected to its topological neighbors and is assigned a so-called weight vector which is of the same dimension as the input data.

The learning process of self-organizing maps can be seen as a generalization of competitive learning although this is historically incorrect as the self-organizing map was presented earlier in literature. The key idea of competitive learning [25] is to adapt the unit to 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 the 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 neighborhood around the winner are adapted, too. Pragmatically speaking, during the learning steps of self-organizing 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 neighboring 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-organizing maps represent a spatially smooth neural version of k-means clustering where k is equal to the number of output units [22].

The crucial steps of the learning process can be described as follows.

1. Random selection of one input vector $x(t)$.

2. Selection of the winning unit c by using the Euclidean distance measure: $||m_c - x|| \leq ||m_j - x||$, for all output units j. In this formula m_c (m_j) denotes the weight vector assigned to output unit c (j).

3. Adaptation of the weight vectors m_j in the neighborhood of the winning unit c at learning iteration t: $m_i(t+1) = m_i(t) + \epsilon(t) \cdot h_{c,i}(t) \cdot [x - m_i(t)]$. The strength of the adaptation is determined with respect to a so-called learning rate $\epsilon(t)$ which starts with an initial value in the range of [0, 1] and decreases gradually during the learning process to 0. The scalar function $h_{c,i}(t)$ determines the amount of adaptation dependent on the neighborhood relation between the winning unit c and unit i which is currently under consideration. Generally, the weight vectors of units which are in close neighborhood to the winning unit are adapted more strongly than weight vectors which are assigned to units that are far away from the winning unit. A convenient and widely used implementation is marked by a Gaussian. This so-called neighborhood function has to guarantee that at the end of the learning process only the weight vector which is assigned to the winning unit is adapted. Obviously, with these two restrictions on the learning rate and the neighborhood function the learning process will terminate.

The outcome of the learning process of self-organizing feature maps results in a clustering of related input data in topologically near areas within the grid of output units. The repetition of this adaptation during the numerous presentations of input vectors makes the formation of areas possible which consist of output units specialized to regularities in the feature vectors of the various input data.

4 Data Mining 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. Text corpus and the automatically produced document description are available at the KONTERM WWW-server[1]. The input to the neural network is represented by feature vectors, each describing the contents 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 with respect to increasing precision in formalizing the legal language, 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 *KONTERM workstation*.

The test phase consisted of a comparison between cluster analysis and the self-organizing map. The results of the cluster analysis are available from the KONTERM WWW-server. 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.

In comparison to that, the Kohonen map gives a much better overview of the document space showing good hills and regions as shown in Figure 1. Taking up the geographical terms used in [19], 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.

Please note, the graphical representation contains as many entries as there are output units in the artificial neural network. Thus, every entry corresponds to exactly one unit of the self-organizing feature map. Each entry is further assigned either a short abbreviation of the document title or a dot. A list can be found at the KONTERM WWW server. Due to the limited space in the figures the abbreviation of only one document is shown even in the case where more than one text segment is assigned to an output unit. The other text segments are given as footnotes.

The documents of the collection cover the following main topics: Copyright law (UR), Diplomacy law (DK), Environment law (UM), Humanitarian law, Geneva conventions (GR), Hague conventions (HR), Human Rights (MR), Law of the sea (SR), Law of the United Nations (VN), Nuclear weapons (AW), Preservation of cultural heritage (AE), Space law (WR).

The map as depicted in Figure 1 shows good hills concerning the Geneva (GR) and Hague conventions (HR), law of the sea (SR), copyright law (UR), law of the United Nations (VN), preservation of the cultural heritage (AE), nuclear weapons (AW), diplomacy law (DK) and space law (WR). Good regions are formed for documents concerning human rights (HR) and environment law (UM). The various classes are less intuitively observable in self- organizing maps because of the lacking border between different classes in the visual representation. However, substantial contemporary research is dedicated to that insufficiency [4, 17, 18, 33].

We have to note that the long training time of the self-organizing map remains problematic for legal applications. For the training of a self-organizing map with input data comparably complex as in the present case study, we have noticed 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. At present, we are working with some extensions of the feature map in order to overcome this deficiency.

The result of this process of data mining may be used as the framework of a guided tour in international law through digital libraries. 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 manner.

5 Related Work

Connectionist models 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. We may distinguish between three distinct classes of neural network applications in law. First, neural networks are trained to represent vague concepts according to some predefined input-output-mapping [2, 8, 34, 21]. Second, spreading-activation during retrieval is used as another paradigm to describe the relation between terms and documents or

[1] http://www.ifs.univie.ac.at/intlaw/konterm/konterm.htm

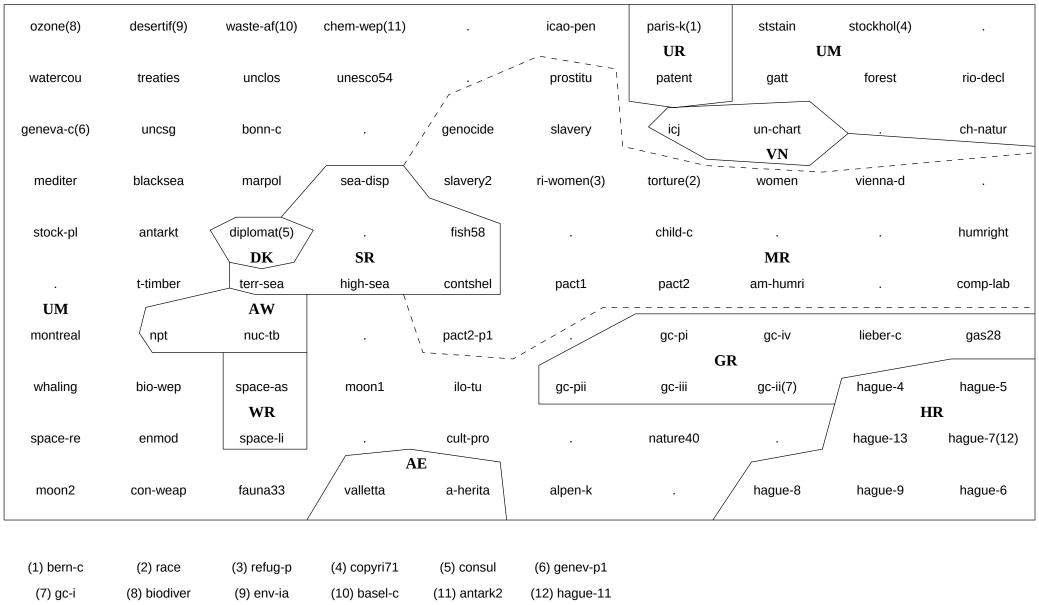

Figure 1: Self-organizing map of public international law treaties

queries [1, 24, 23]. Third, concept or document spaces are described by neural networks adhering to the unsupervised learning paradigm.

The paper of [12] marks the first attempt to utilize self-organizing maps for information retrieval. Similar to our approach, the authors rely on self-organizing maps. In this paper, however, the document representation is made up from 25 manually selected index terms and is thus not really realistic. In [15] this line of research is continued, yet this time with full-text indexed documents. In the area of legal information processing, the self-organizing map has been used in [19, 29] for exploratory analysis of judicial documents.

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 of uncovering similarities between text documents.

6 Conclusions and Future Work

Data mining in huge legal information systems may provide the necessary overview of the contents of text corpora. We suggest the utilization of an unsupervised training rule. The examples drawn from our document collection on public international law show the feasibility of our approach.

Future work will thus concentrate on a speed-up of the network and an explicit representation of classes. Possible applications are the exploration of 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 tight corset of Boolean logic and Boolean search expressions but rather may be enhanced by guided tours providing the means of convenient voyage in an environment of dynamically classified legal documents.

Acknowledgments

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

References

[1] R. K. Belew. A connectionist approach to conceptual information retrieval. In *Int. Conference on Artificial Intelligence and Law*, 1987.

[2] T. Bench-Capon. Neural networks and open texture. In *Int. Conference on Artificial Intelligence and Law*, 1993.

[3] K. W. Church and R. L. Mercer. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19(1), 1993.

[4] M. Cottrell and E. De Bodt. A Kohonen map representation to avoid misleading interpretations. In *European Symposium on Artificial Neural Networks*, 1996.

[5] R. M. DiGiorgi and R. Nannucci, editors. *Hypertext and Hypermedia in the Law.* Special Issue, Informatica e Dritto, Edizioni Scientifiche Italiane, 1994.

[6] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 1996.

[7] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining.* AAAI Press, Menlo Park, CA, 1996.

[8] C. Groendijk. Neural schemata in automated judicial problem solving. In *Legal Knowledge-Based Systems - Information Technology and Law (JURIX'92)*, 1992.

[9] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 1982.

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

[11] K. Lagus, T. Honkela, S. Kaski, and T. Kohonen. Self-organizing maps of document collections: A new approach to interactive exploration. In *Int. Conference on Knowledge Discovery and Data Mining*, 1996.

[12] X. Lin, D. Soergel, and G. Marchionini. A self-organizing semantic map for information retrieval. In *Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, 1991.

[13] H. Mannila. Methods and problems in data mining. In *Int. Conference on Database Theory*, Delphi, Greece, 1997.

[14] G. Marchionini and B. Shneiderman. Finding facts vs. browsing in hypertext systems. *IEEE Computer*, 21(1), 1988.

[15] D. Merkl. A connectionist view on document classification. In *Australasian Database Conference*, 1995.

[16] D. Merkl. Lessons learned in text document classification. In *Workshop on Self-Organizing Maps*, 1997.

[17] D. Merkl. Exploration of document collections with self-organizing maps: A novel approach to similarity visualization. In *European Symposium on Principles of Data Mining and Knowledge Discovery*, 1997.

[18] D. Merkl and A. Rauber. Alternative ways for cluster visualization in self-organizing maps. In *Workshop on Self-Organizing Maps*, 1997.

[19] D. Merkl, E. Schweighofer, and W. Winiwarter. CONCAT: Connotation analysis of thesauri based on the interpretation of context meaning. In *Int. Conference on Database and Expert Systems Applications*, 1994.

[20] J. Nielsen. *Hypertext and Hypermedia.* Academic Press, Boston, 1993.

[21] L. Philipps. Are legal decisions based on the application of rules or prototype recognition? Legal science on the way to neural network. In *Pre-Proceedings Int. Conference on Logica, Informatica, Dritto*, 1989.

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

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

[24] D. E. Rose and R. K. Belew. Legal information retrieval: A hybrid approach. In *Int. Conference on Artificial Intelligence and Law*, 1989.

[25] D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing: Explorations in the microstructure of cognition.* MIT-Press, Cambridge, MA., 1986.

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

[27] 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.

[28] E. Schweighofer and D. Scheithauer. The automatic generation of hypertext links in legal documents. In *Int. Conference on Database and Expert Systems Applications*, 1996.

[29] E. Schweighofer, W. Winiwarter, and D. Merkl. Information filtering: The computation of similarities in large corpora of legal texts. In *Int. Conference on Artificial Intelligence and Law*, 1995.

[30] S. Simitis. *Informationskrise des Rechts und Datenverarbeitung (Information Crisis in Law and Data Processing).* Müller, Karlsruhe, 1970.

[31] H. R. Turtle. Text retrieval in the legal world. *Artificial Intelligence & Law*, 3(1-2), 1995.

[32] H. R. Turtle and W. B. Croft. A comparison of text retrieval models. *Computer Journal*, 35, 1992.

[33] A. Ultsch. Self-organizing neural networks for visualization and classification. In *Information and Classification – Concepts, Methods, and Applications.* Springer-Verlag, Berlin, 1993.

[34] G. J. van Opdorp, R. F. Walker, C. Schrickx, C. Groendijk, and P. H. van den Berg. Networks at work: A connectionist approach to non-deductive legal reasoning. In *Int. Conference on Artificial Intelligence and Law*, 1991.

[35] P. Willet. Recend trends in hierarchic document clustering: A critical review. *Information Processing & Management*, 24, 1988.

En Route to Data Mining in Legal Text Corpora:
Clustering, Neural Computation, and International Treaties
Dieter Merkl and Erich Schweighofer
In *Proceedings of the Int'l Workshop on Database and Expert Systems Applications.* Toulouse, France, Sept 1–2, 1997, IEEE CS Press, pp 465–470.