INFORMATION RETRIEVAL IN LAW USING A NEURAL NETWORK INTEGRATED WITH HYPERTEXT

T.D. Gedeon* and V. Mital
Department of Computer Science, Brunel University
Uxbridge, Middx. UB8 3PU

Abstract

Hypertext is increasingly finding application in legal practices. One of the most important applications involves the assembly of complex documents, where it is necessary to differentiate between clauses which largely share a common vocabulary.While possessing many advantages over paper-based media, hypertext systems do not adequately cater to information retrieval needs of users. It is essential to provide facilities by which the user can, when desired, break free of the confines of the structure created by hypertext links. However, each of the existing approaches has some inherent limitations. Studies show that full text retrieval gives extremely poor *recall*, although users imagine otherwise. Methods related to subject indexing can suffer from inconsistent interpretations and encoding. Vector retrieval suffers because it does not take into account indirect, multiple connectivities between words and text units. We believe that neural networks provide advantages over existing approaches. We use a network in which text units are linked to significant words found in them, the weight of links being determined through an automatic text analysis based on a normalised word frequency measure. There is no intermediate layer, allowing us to construct the initial state of the network rapidly, and to readily accommodate new documents as they arrive over a period of time: a peculiar requirement of the application. Emphasis is paid to the method in which the user can provide feedback based on the value s/he attaches to the documents retrieved in response to a query.

1. Introduction

The work of a law firm is heavily dependent on being able to appraise and manipulate large quantities of text in a precise manner. Computer support for legal practitioners requires information systems that allow information to be structured in a sophisticated manner, overcoming the constraints imposed by paper-based media. Hypertext is a technology which enables complex informational structures to be built and traversed in a guided manner. Consequently, it is increasingly being used for major applications in law (Yoder and Wettach, 1989; Morrow, Baird and Russel, 1990; Thomas and Mital, 1991). However, progress is not entirely smooth. Part of the problem lies with the nature of hypertext itself; while the peculiar character of the application also contributes to the difficulties.

A widely acknowledged problem with hypertext is that while it allows text to be broken down into small units, the means it provides for users to locate a particular unit are currently not sufficiently powerful (Lesk, 1989). When the number of text units is large, queries achieve "too many hits" and swamp the user with an excess of information. This problem is not dissimilar to that found in full text retrieval (FTR) systems with sizeable collections of documents (Blair and Maron, 1985). It is not possible to create links to express every association of ideas that a user will wish to pursue. If that were attempted, users would be confronted with too many choices at every juncture, without adequate guidance as to which to follow. It is possible to label links and semantically differentiate them (DeRose, 1989; Southam, Mital and Thomas, 1991); however, in a complex domain, links soon proliferate beyond control. The structures created by the links are clearly crucial to successful traversal through hypertextual space; however, the structures should not become a straitjacket.

Currently, research is focusing on providing means to users to break through the structural boundaries of hypertext links and locate information in a manner which is wholly or partly independent of those links. There are a number of ways in which this is being done. While some benefits are obtained, we will see that there are inherent limitations in each of the current methods of providing independent retrieval in conjunction with hypertext. The problem is exacerbated in some kinds of legal applications where clauses on the same subject, largely sharing a common vocabulary, have to be differentiated. In this paper, we present a neural network that is designed to work with hypertext systems in such applications. It will be seen that there are a number of advantages over the current approaches to providing textual retrieval in hypertext.

* Now at The University of New South Wales, Sydney.

CH 3065-0/91/0000-1819 $1.00 © IEEE

2. Current information retrieval in hypertext

Much work is being done on extending hypertext systems to allow users to search the text-base without necessarily having to follow the explicitly provided inter-textual links. A number of approaches have been taken:

(a) FTR facilities have been added to hypertext, often with interfaces allowing users to formulate and pursue complex queries (Coombs, 1990).
(b) Responses indicating the relevancy of text units ("Like" or "Don't Like") are obtained from users. This information is used to restructure the index of key-terms associated with text units (Frisse and Cousins, 1989).
(c) A retrieval mechanism is provided which is based on formulating vectors from the occurrence or absence of words in a text unit. Crouch et al (1990) have extended the basic vector retrieval model (Salton, 1971) and have proposed its integration with hypertext.

The limitations of relying on simple FTR in domains such as law have been extensively studied and noted (Blair and Maron, 1990). Users find it difficult to predict those combinations of words and phrases which occur in most of the relevant documents, but do not occur in most of the non-relevant ones. In fact, "...words and lexical items in general, are a poor approximation to meaning. Even with the addition of thesaurus or a phrasal lexicon, the approximation is not good. ... It is insufficient to treat words as indices irrespective of conceptual relationships between them" (Rau, 1987).

The approach (b) above has been criticised on the grounds that it is subject to the problems of manual indexing - e.g. inconsistent interpretations and encoding by assessers - without having the power of full text indexing (Coombs, 1990).

As for vector retrieval, the performance of systems using this technique does not appear to exceed that achieved by manually adding subject indices and pre-ranking them (Bing, 1989). This is because the vectors are usually formed with only one link between a word and a text unit. This does not deal with the indirect connectivities, such as are taken into account in the neural network-based information retrieval system that we will present in this paper. In a neural network there can be many paths between a text unit and a word: some direct, some through other text units or words. It is postulated that this extra connectivity can discover and capture the semantic significance at a sub-symbolic level.

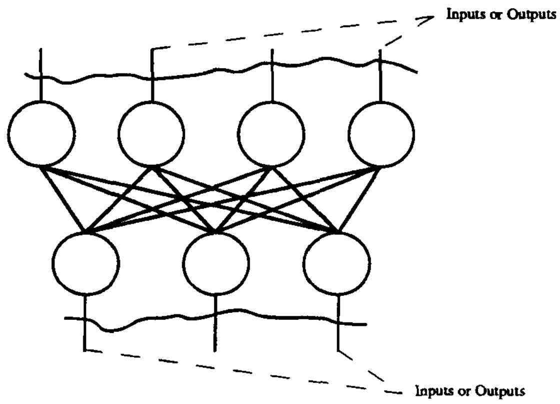

Figure 1 Basic interactive network

3. Neural networks in legal information retrieval

Belew (1987) has done some of the pioneering work in the use of neural networks for information retrieval and has settled for an interactive network of the type illustrated in Figure 1. The network is set up so that each neuron represents either a word or a document (Belew, 1987). Links are weighted in correspondence to a certain word frequency measure derived from automatic analysis of the text. This means that the network is usable, to some extent, immediately on creation - rather than only after extensive training, as would be the case for feed-forward networks (Rumelhart, Hinton and Williams, 1986). Training and learning will change these weights. This is where, in our view, considerable problems are likely to arise when new information is added.

Due to learning during use, the weights of links, which are originally set by relative word frequency measures, are likely to have changed by the time new documents come to be added to the information base. Adding new documents will affect the relative word frequencies of even the existing documents. It is therefore not clear whether the whole network will have to be reinitialised with weights corresponding to new word frequencies, or, whether only the new documents will be connected by such links, while the links of the previously existing documents will keep their status quo in spite of the changed state of play. Either alternative is unsatisfactory. As shall be seen next, we have taken this problem into account when designing our network architecture.

4. An augmented network

The structure of the network that we are using is shown in Figure 2. All words, other than so-called noise words (such as prepositions, articles and function words) are made neurons in the network. All text units are also neurons. Every word has links to each text unit to which it is 'significant'. Significance is judged by a standard technique in automatic analysis of text for information retrieval: forming a numerical measure of the frequency with which a word occurs in a text unit, normalised by (a) the frequency of the word's occurrence in the information-base as a whole, and (b) the size of the particular text unit. Those words which have either too low or too high a word frequency measure are discarded as being either too parochial or too ubiquitous to be of much use in discriminatory retrieval.

Each word is linked to every text unit to which it is significant through a bidirectional link created with a weight corresponding to the frequency measure, these links are called textual-associative (T-A) links. The weight of T-A links is determined by the normalized word frequency measures obtained from the automatic text analysis. Additionally every word is connected to every text unit by latent (L) links which are all initially given a weight of zero. These are also bidirectional, in the sense that weights are the same in both directions.

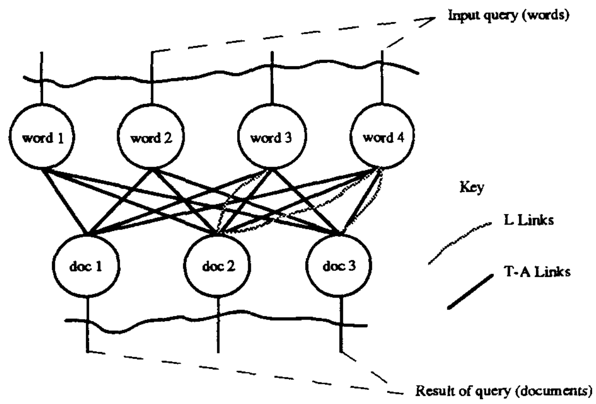

Figure 2 Network designed for growing information-base

Queries are restricted to specifying words which act as concept micro-features. Alternatively, a query may specify at most one text unit. A query may not contain both words and text units because this is of dubious significance in the context of conceptual information retrieval. This is because a text unit may contain, or be relevant to, a number of concepts. If, say, three words and one text unit were to be selected, does it mean that the concept signified by the co-occurrence of the three words, together with all the concepts referred to in the text unit, are part of the query? The reason we allow one text unit alone to be in queries is that then the text units contained in the answer can be thought to be (in some sense) similar to the one in the query in a multidimensional concept space.

The set of retrieved text units itself contains rankings and gradations according to the strength of activation. Consequently, relevance is not an all-or-nothing quality, but is relative.

4.1. Operation of the network

T-A links are always positive as, obviously, there is no meaningful measure of how many times a word does not occur in a text unit. During learning, L links are modified based on user responses to text units supplied as answers to queries. The user may respond as follows: relevant, marginal, and irrelevant. The L links to a text unit from the words forming the query are increased in weight for 'relevant' user responses, and reduced for user assertion that the text unit was 'irrelevant'.

4.2. Exclusionary relation between concept and text unit

The network can learn to assign negative relevance of words to text units (in the context of concepts) because the joint effect of the T-A and L links may become negative. This is significant, as a word may occur in a text unit following an exclusionary declaration about a concept. An example would be a statement in the text that 'we will not consider negligence'. The word negligence does occur, but only so that the concept(s) related to it are immediately excluded. Therefore, this text unit should not be retrieved when the relevant concept is specified in the query. We allow the network to learn about this. We considered including a possible fourth user response during training to explicitly indicate that rather than being 'irrelevant' the text unit has a exclusionary connection to the query concept. However, we believe that this would be perilously close to adopting a symbolic computation approach, with all the attendant problems of exhaustive assignment of symbolic connections (cf. the approach of Rose & Belew, 1989, where some of the links are derived from a conceptual analysis and are similar to those in a semantic network). We would prefer for this information to be discovered at the subsymbolic level. If a word and text unit pair are sufficiently often noted by the user as not relevant, then eventually the network can discover that the concept is excluded even though it is mentioned.

4.3. Vector retrieval is a byproduct:

The first cycle of processing in the network is special, in that the L links are not used. This allows us a baseline of a non-banded or non-discrete valued (continuous) version of vector retrieval (Salton, 1971) as the first set of outputs on the text unit neurons. This is useful as an interim result, and is very fast - important in a practical system. The text unit set produced as an answer to the query can be built up in a number of ways:

(a) The above mentioned interim set of vector retrieval.
(b) The text unit neurons active when the net has settled.
(c) The text unit neurons activated in a short repetitive cycle.
(d) All those above a threshold activation at any time during processing.

Activations of text units and words are limited to the range between 0 and 1. The ceiling could obviously be any arbitrary value, the floor value of 0 is significant in that an activation of 0 implies that that word or text unit neuron is not relevant (is excluded from) the query, and is therefore effectively isolated from influencing the rest of the network. Subsequent incoming signals may of course increase its activation and reinstate it.

4.4. Scaling

Activations coming into neurons are adaptively scaled. This is to guarantee that the network will settle within a reasonably short period, or oscillate in a restrained fashion only. Scaling means that the total energy in the network is measured and every neuron's activation is reduced pro rata to bring the energy to a pre-specified level. Scaling of activations indirectly affects approximately how many text units are to form part of the answer to the query, as well as how many words other than the ones input by way of query are to be considered internally by the network. In other words, scaling affects the scope of the spread of activation. As a simplified example, values of 3 and 10 are chosen for scaling factors of text unit neurons and word neurons respectively. Then the total activation energy is scaled down to accord with these factors. The total could be distributed so that 3 text units and 10 words all have activations of 1, with all the rest 0. This is unlikely, we find that a more wide spread distribution is usual. However, this is not so wide as to contain a lot more than 3 text units and 10 words significantly active. In other words, the spread has been limited to a certain extent. Of course, which text units/words are active will be decided interactively by the network.

4.5. Learning:

As mentioned, the network in its initial state with all L links at 0 still produces useful results. The first cycle produces a vector-retrieval. After the network has settled into a steady state, the result signifies the output of the network's approximation of a concept formed by the query words. This concept may not have been exactly what the user had in mind, as would later be evinced by his responding that some of the output text units are 'irrelevant'.

We can incrementally adjust the network by utilising the user's explicit behaviour. The error or correctness signified by the user's responses is readily interpretable into blaming or rewarding particular parts of the network because there are direct word to text unit relationships. The network is changed by modifying the L links by attaching extra weight or reducing the weight of the connections between the query words and the impugned or approbated text units.

4.6. Discussion

The question which must be answered is why maintain the two sets of links, L and T-A. The major reason is to allow the inclusion of new text units at will into the network without degrading the performance. For this we want to retain the L link information which the network has learnt. Certainly the addition of a number of text units will change the word frequency measure of some connections between words and a particular text unit to the extent that, if T-A links were all we had, the text unit would not be presented to the user although it might be on the basis of the same query using the original text unit set. Nevertheless, the significance of the user selecting a text unit as either 'relevant' or 'irrelevant' to the concept he had in mind when making a query in the smaller set is not diminished and should be carried forward. If a text unit was said by the user to be relevant to a concept derived from the words making up a query it should remain so even though the additions change the word frequencies. Even more importantly - because of the much longer time taken to learn this - if a text unit is learnt to be irrelevant to a concept, it should not become falsely relevant in the light of new text units. However, the relative importances of the L and T-A links when the text unit set changes is a matter which we are still investigating. We are currently looking at the susceptibility of the network in the face of sizable changes to the information base.

4.7. Operation with hypertext

We have implemented the first version of the network on an Apple Macintosh computer using Lightspeed Pascal (Gedeon and Mital, 1991). This version of the network has been integrated to a hypertext system - HyperNotary - which is meant to assist lawyers in the assembly of complex documents, such as wills and pleadings (Southam et al, 1991). HyperNotary is implemented using the popular tool, HyperCard from Apple Computer, Inc.

In HyperNotary, each precedent (previously written document) is stored as a HyperCard stack; each clause or text unit in the precedent is a card in the stack. The neural network has nodes corresponding to each of the cards and to the significant words in the cards. Connectivities (i.e. L and T-A link weights) are maintained. When a precedent is inserted in the information-base, or deleted, the previous word frequency measures are amended in light of the alterations. However, when individual clauses are added to a precedent, we do not recalculate the overall word frequencies until the particular document assembly session is ended. This means that the network does not have the means to directly access the new clauses immediately after they have been added. The advantage is that the session is not interrupted unduly. Our position is that the user is likely to have full knowledge of the newly inserted clauses and will be able to pinpoint them using the other ways of locating clauses, viz., following the hypertext links and searching for their names using HyperCard's own facilities.

5. Conclusions

The connectionist approach to information retrieval has great practical utility. It can be used in conjunction with conventional systems such as fulltext databases, key word retrievers, and hypertext based information systems. We believe that the connectionist approach has considerable advantages over vector retrieval as well as several types of probabilistic retrieval models. There are clearly some applications where conceptual indexing using manual assessment and abstraction of information contained in documents is quite feasible; there, explicit representation of information, such as that based on the object-oriented paradigm is viable (Mital, Stylianou and Johnson, 1991). However, there are other commercial situations where manual assessment is not economically feasible; neural networks, in alliance with automatic text analysis, can prove to be of considerable value.

References

Belew, R.K. (1987) A connectionist approach to conceptual information retrieval, *Proc. First International Conference on Artificial Intelligence and Law*, ACM, 116-125.

Bing, J. (1989) The law of the books and the files: possibilities and problems of legal information retrieval, in: G.P.V. Vandenberghe, Ed., *Advanced Topics of Law and Information Technology*, Hillsdale, NJ, Lawrence Erlbaum, pp.151-182.

Blair, D.C. (1990) *Language And Representation In Information Retrieval*, Amsterdam, Elsevier.

Blair, D.C. & Maron, M.E. (1985) An evaluation of retrieval effectiveness for a full-text document retrieval system, *Communications of the ACM*, **28**, 3, 289-299.

Blair, D.C. & Maron, M.E. (1990) Full-text information retrieval: further analysis and clarification, *Information Processing & Management*, **26**, 3, 437-447.

Coombs, J.H. (1990) Hypertext, full text, and automatic linking, *Proc. ACM SIGIR Conference*, 83-98.

Crouch, C.J.; Crouch, D.B. & Nareddy, K.R. (1990) The automatic generation of extended queries, *Proc. ACM SIGIR Conference*, 369-383.

DeRose, S.J. (1989) Expanding the notion of links, *Proc. Hypertext '89*, 249-257.

Frisse, M.E. & Cousins, S.B. (1989) Information retrieval from hypertext: update on the Dynamic Medical Handbook project, *Proc. Hypertext '89*, 199-212.

Gedeon, T.D. & Mital, V. (1991) An adaptive learning network for information retrieval in a litigation support environment, *Proc. AMSE International Conference on Neural Networks: Methodologies and Applications*, San Diego.

Lesk, M. (1989) What to do when there is too much information, *Proc. Hypertext '89*, 305-318.

Mital, V.; Stylianou, A.K. & Johnson, L. (1991) Conceptual information retrieval in litigation support systems, *Proc. Third International Conference on Artificial Intelligence and Law*, ACM, Oxford.

Morrow, N.M.; Baird, P. & Russell, S. (1990) Legal information in respect of copyright: a hypermedia challenge, *Proc. Online Information 90*, London, 351-359.

Rau, L.F. (1987) Knowledge organization and access in a conceptual information system, *Information Processing and Management*, **23**, 4, 269-283.

Rose, D.E. & Belew, R.K. (1989) Legal information retrieval: a hybrid approach, Proc. First International Conference on Artificial Intelligence and Law, ACM, 138.

Salton, G. (1971) *The SMART Retrieval System - Experiment in Automatic Document Processing*, Englewood Cliffs, Prentice-Hall.

Southam, M.; Mital, V. & Thomas, P. (1991) HyperNotary: Legal document assembly using hypertext, *Proc. 13th BCS Information Retrieval Colloquium*, Lancaster.

Thomas, P. & Mital, V. (1991) Knowledge-based hypertext for legal information retrieval, *Proc. Expersys-91: Applications Des Systemes Experts*, Paris.

Yoder, E. & Wettach, T.C. (1989) Using hypertext in a law firm, *Proc. Hypertext '89*, 159-167.