Application, Analysis and Evaluation of Neural Networks-Based Interaction Information Retrieval

Sandor Dominich

Department of Computer Science, University of Veszprem, 8200 Veszprém, Egyetem u. 10, Hungary, E-mail: dominich@dcs.vein.hu; Department of Computing and Information Technology, Buckinghamshire Chilterns University College, Queen Alexandra Road, High Wycombe, HP11 2JZ, United Kingdom, E-mail: sdomin01@bcuc.ac.uk

Abstract

Connectionist (artificial neural networks) based views for adaptive clustering in Information Retrieval have proved to be viable approaches, and have yielded a number of models and techniques. However there has never been any exhaustive and methodical — i.e., theoretical, formal, practical, simulation- and user-based — evaluation of such a retrieval method and system. The aim of the paper is therefore just this. The paper suggests an adaptive clustering technique for information retrieval based on the Interaction Information Retrieval (I^2R) method. Theoretical (formal) as well as simulation results as regards computational complexity of this method are presented and discussed. Evaluations of retrieval effectiveness are also given using standard test collections. Two applications were designed, developed and implemented based on this method. Their relevance effectiveness is evaluated by experiments carried out with human subjects. The results obtained are promising, and they show that a retrieval technique based on adaptive clustering using a connectionist approach proves useful when retrieving from homogeneous documents, and when emphasis is on high precision.

1 Introduction

The application of soft computing techniques to information retrieval has aimed at enhancing retrieval performance by offering techniques to capture aspects that could hardly be modelled by other means numerically, which is important because only this can yield implementable systems. For example, fuzzy set theory allows for expressing the inherent vagueness encountered in the relation between terms and documents, and fuzzy logic makes it possible to express retrieval conditions by means of formulas in the weighted Boolean model [Kraft, Bordogna, and Pasi, et al., 1998].

Another soft computing technique is based on artificial neural networks. Connectionism and learning allow to model relations between documents as well as their changes [Cunningham et al., 1997]. The documents are interconnected (connections follow from, e.g., common meaning), and these links are not static: they can change, for example in an interaction with the user' query. Since the documents are interconnected some of them may be more strongly linked to each other thus forming groups, which yields the important issue of clustering.

Clustering is a well-known technique applied in Information Retrieval (IR). It is typically used to group documents (to be searched), in which case the result is a — usually disjoint — set of document groups called clusters, each cluster containing — in some sense — similar documents. Several clustering methods and techniques have been proposed so far, such as, for example, based on similarity measures [van Rijsbergen,

1979; Salton and McGill, 1983], neighborhoods [Voorhees, 1985], hierarchies [Lebowitz, 1987; Willett, 1988; Fisher and McKusick, 1989; Crawford et al., 1991; Tanaka et al., 1999], on matrix theory (diagonalisation, singular value decomposition [Deerwester et al., 1990]).

Retrieval is performed based on a cluster representative which may but need not be one of the cluster members. If the particular retrieval method used associates a cluster representative to a query, then every member of that cluster is returned in response to the query. This view of retrieval is based on the well-known cluster hypothesis according to which closely associated documents tend to be relevant to the same request. It is commonly agreed that — *a priori*, fixed — clustering should be stable under growth, description and ordering. Recent research reveals a sound mathematical background for this type of clustering as well as for its evaluation [Hearst and Pederson, 1996; Mather, 2000].

As somewhat opposed to fixed clustering, adaptive clustering (i.e., a clustering in which the cluster structure is being developed in the presence of the query) has proved to be a viable approach to IR [Belew, 1989; Rose, 1994; Johnson et al., 1994, 1996; Shaw et al., 1997; Mobasher et al., 1998]. Retrieval is then viewed similar to that in fixed clustering: those documents are said to be retrieved which form the same cluster ('nearest' to query).

One way of conceiving adaptive clustering is to adopt a connectionist-based (semantic networks, neural networks) view [Cohen and Kjeldsen, 1987; Belew, 1989; Doszkocs et al., 1990; Chen, 1994, 1995]. As yet there have not been any exhaustive and methodical — both theoretical and practical — evaluations as to the computational complexity as well as standard and practical effectiveness of such retrieval method and practical systems (applications).

Thus the aim of this paper is just this.

Following the connectionist line, the present paper proposes a retrieval model using adaptive clustering (with neural networks) based on the Interaction Information Retrieval (I^2R) paradigm, and reports on a complex evaluation, analysis and application.

2 Associative Interaction Information Retrieval

The Interaction-based paradigm of *IR* (*I^2R*) was first suggested in [Dominich, 1994; van Rijsbergen, 1996] based on the concept of interaction according to the Copenhagen Interpretation in Quantum Mechanics (query: measuring apparatus, documents: observed system, retrieval: measurement).

The documents are represented as a flexibly interconnected network of objects (or artificial neurons). The interconnections are adjusted each time a new object (e.g., a document) is fed into the network. The query interacts with the other objects, i.e., it is treated like any other object: it is interconnected with the already interconnected other objects. Thus, on the one hand, new connections will develop (between the object-query and the other objects), and on the other hand, some of the existing connections can change. Retrieval is

defined as recalled memories, i.e., those documents are said to be retrieved which belong to reverberative circles triggered by a spreading of activation started at the object-query. The reverberative circles correspond to clusters, which are not fixed as they develop in the presence of the query. This model will be referred to as Associative Interaction Information Retrieval (AI^{2}R). The idea of flexible, multiple and mutual interconnections from I^2R also appear and are investigated in [Salton, Allan and Singhal, 1996; Salton, Singhal, Mitra and Buckley, 1997; Pearce and Nicholas, 1996; Carrick and Watters, 1997; Liu, 1997; Mock and Vemuri, 1997; Dominich, 1997, 2001].

Any object (or document) o_i, $i = 1, 2, ..., M$, is assigned a set of identifiers (e.g., keywords, index terms) t_{ik}, $k = 1, 2, ..., n_i$. There are weighted and directed links between any pair (o_i, o_j), $i \neq j$, of objects. The one is the ratio between the number f_{ijp} of occurrences of term t_{jp} in object o_i, and the length n_i of o_i, i.e. total number of terms in o_i:

$$w_{ijp} = \frac{f_{ijp}}{n_i}, \quad p = 1, ..., n_j \tag{1}$$

Because w_{ijp} is analogous to the probability with which object o_i 'offers' t_{jp} (or equivalently with which t_{jp} is extracted from o_i when being in o_j), the corresponding link may be viewed as being directed from object o_i towards object o_j.

The other weight, w_{ikj}, is the inverse document frequency. If f_{ikj} denotes the number of occurrences of term t_{ik} in o_j, and df_{ik} is the number of documents in which t_{ik} occurs, then:

$$w_{ikj} = f_{ikj} \log \frac{2M}{df_{ik}} \tag{2}$$

Because w_{ikj} is a measure of how much content of object o_j is 'seen' (or 'mirrored' back) by term t_{ik}, the corresponding link may be viewed as being directed from o_i towards o_j. The other two connections — in the opposite direction — have the same meaning as above: w_{jik} corresponds to w_{ijp}, while w_{jpi} corresponds to w_{ikj} (Figure 1). Figure 2 shows a simple example.

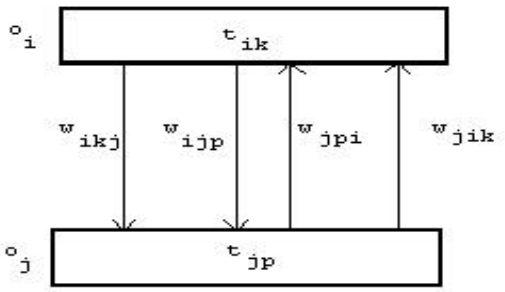

Figure 1. Connections between object pairs.

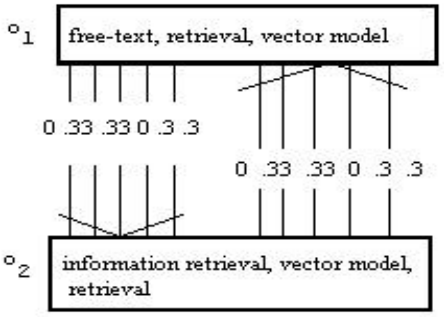

Figure 2. Weighted connections.

The process of query answering is performed in two phases: (a) Interaction (Figure 3). The query Q is incorporated first into the network.

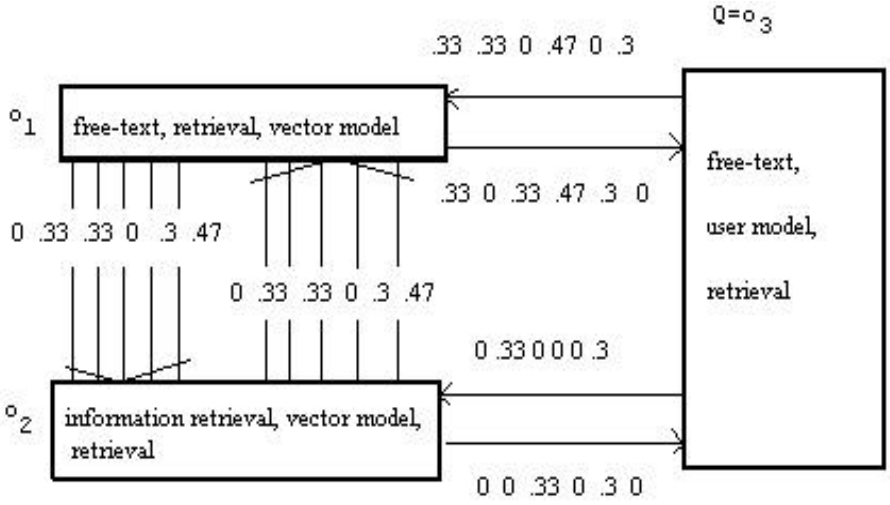

Figure 3. Interaction between query and documents: new connections between query Q and object-documents o_1 and o_2 are developed, and there is a changed link between o_1 and o_2 (0.47 instead of 0.3). The activation starts at Q, and spreads over to o_1 from which to o_2, and then back to o_1 (o_1 and o_2 form a reverberative circle).

(b) Retrieval. A spreading of activation takes place according to a winner takes all strategy. The activation is initiated at the query, say o_j, and spreads over along the strongest connection thus passing on to another neuron, and so on. The strength of the connection between any pair (o_i, o_j), $i \neq$ i, of objects, and thus between the query and another object o_i is defined as follows:

$$\sum_{p=1}^{n_j} w_{jpi} + \sum_{k=1}^{n_i} w_{jik} \qquad (3)$$

This summation is made possible by the meaning associated to w_{jpi} and w_{jik} (formulas 1, 2); each represents a measure of the extent to which the query, represented by o_j, 'identifies' — the content of — o_i. After a number of steps the spreading of activation reaches an object already affected, i.e., a reverberative circle is formed: if

continued the activation would spread in a circle of objects. This is analogous to a local memory recalled by the query (Figure 3). The reverberative circle can be interpreted as an adaptive cluster associated to the query when this is present. Those objects are said to be retrieved in response to the query which belong to the same reverberative circle, and they are ranked in the order of maximal activation, i.e., in the order in which they are traversed. The same objects may not form the same cluster for a different query.

3 Theoretical Evaluation of the AI^{2}R Method

As it could be seen in part 2 an algorithm which implements the method should compute a huge number of weights. Thus the question of tractability and hence complexity of such a computation arises, and it is answered in the following two theorems.

THEOREM 1. The complexity of weights computation is polynomial.

Proof. The computation of weights includes three phases: (i) the computation of individual weights (formulas 1 and 2), (ii) the computation of sums (formula 3), and (iii) the application of winner-take-all strategy, i.e., finding the strongest (maximal) connection. As it can be seen from the formulas (1) and (2) there are $2(n_i + n_j)$ number of weights between every pair (o_i, o_j), $i \neq j$, of which there are ${}^{M}C_2$ (combinations of M taken by 2), hence $2(n_i + n_j)\ {}^{M}C_2 = O(NM^2)$, where O denotes 'big-Oh', is an upper bound for weights computation, where $N = \max_{i,j}(n_i, n_j)$, i.e., the largest of object lengths. The computation of the sums of weights (formula 3) between a given object o_i and all the other objects o_j, of which there are $M - 1$, takes time $(n_i + n_j)(M - 1)$, and thus an upper bound for the computation of all sums in the network is $(n_i + n_j)(M - 1)^2 = O(NM^2)$ because i can vary, too, at most $M - 1$ times. An upper bound to find the strongest connection from a given object and all the others is $O(M)$ (finding the maximum froma sequence of numbers), and thus an upper bound for the selection of all strongest connections in the entire network is $O(M)(M - 1) = O(M^2)$. Hence an overall upper bound for weights computation (all three phases (i)-(iii), see above) is $O(NM^2) + O(NM^2) + O(M^2) = O(NM^2) = O(K^3)$, where $K = \max(N, M)$. ■

In other words the computation of weights is tractable. Once the weights have been calculated, the complexity of the retrieval process itself is given by:

THEOREM 2. The retrieval process takes polynomial time.

Proof. The spreading of activation starts at o_q representing the query, and means finding $\max_i w_{iq}$, $i = 1, ..., M - 1$, i.e., finding the maximum of all the weights linking o_q with all the other objects of which there are $M - 1$ (where $w_{iq} = \sum_p w_{qpi} + \sum_k w_{qik}$, see formula 3). Finding this maximum has complexity $O(M)$. Let $o_{m'}$ denote the winner object, i.e., the object to which the activation spreads, and let L denote a list keeping all the winner objects. It should be checked whether m' has already been a winner or not. This is accomplished by checking whether m' is in L or not: if it is we have a reverberative circle, and we stop, but if it is not in L it is written into

L in the next available location. Checking L once takes $O(\text{length}(L)) = O(M)$. Because the spreading of activation is carried out at most M times, checking L takes $M \cdot O(M) = O(M^2)$ time. If all the weights are unique there only is one reverberative circle, but if there are objects, say o_i ($i = j_1, ..., j_k$), from which there are more than one (i.e., multiple), say n_i, maximal weights the number of reverberative circles will be $\prod_i n_i = O(n^k)$, where $n = \max_i n_i$, and thus an upper bound for the overall complexity of retrieval is $O(n^k) \cdot O(M^2) = O(n^k \cdot M^2) = O((\max(n^k, M^2))^2)$ ■ Theorem 2 tells us that the retrieval process itself is a tractable operation. These results mean that the computations involved have polynomial complexity, and thus the method is tractable.

After weights summations there are $M - 1$ links (weights) — $s_1, ..., s_{i-1}, s_{i+1}, ..., s_M$ — from an object o_i to all the other objects. Because i varies from 1 to M there are at most $M(M - 1) = O(M^2)$ links to be evaluated in all (in a search). Depending on the multiplicity (i.e., unique, double, triple maximum, or higher) of the maximum of the sequence $s_1, ..., s_{i-1}, s_{i+1}, ..., s_M$ the number of reverberative circles can increase. The number of retrieved objects depends on two factors: (a) the number of the reverberative circles, and (b) the number of objects a reverberative circle contains. In order to render the influence of the first parameter simulations were carried out using a C program written for this purpose. M was taken 100000, and different number of sequences of weights were generated at random. In each of these cases the maximum and its multiplicity was determined (Table 2).

Table 2. Simulation of the multiplicity of maxima. (In 985 sequences out of 1000 sequences there is a unique maximum, in 14 cases there is double maximum, and in 3 cases there is triple maximum)

	Multiplicity of maxima proportions			
Number of generated sequences	**1**	**2**	**3**	**4**
1 000	985	14	1	0
2 500	2469	30	1	0
10 000	9532	439	26	2

Drawing the empirical density function yields a curve represented by the thinner line in Figure 4. The value of the empirical density function on every interval $\Delta x = (0, 1), (1, 2), (2, 3), (3, 4)$, is calculated using the usual ratio

$$\frac{number\_of\_values\_within\_interval}{interval\_length \times total\_number\_of\_values} \qquad (4)$$

for each of the three cases, and then the corresponding values are averaged.

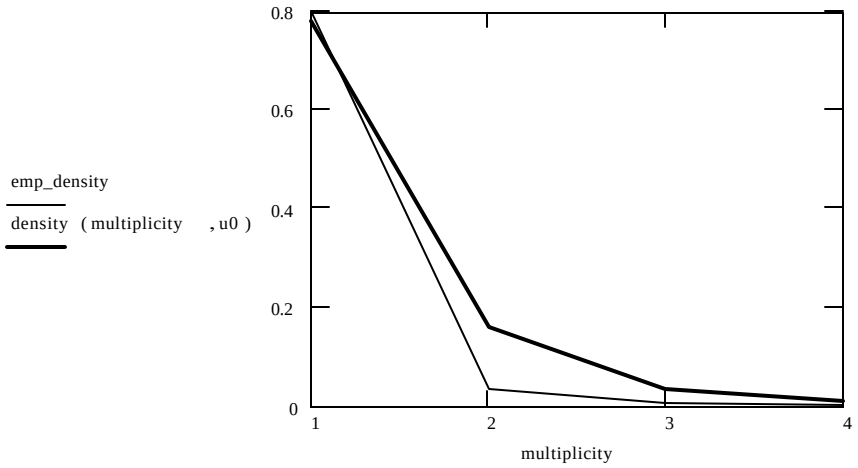

Figure 4. Empirical and estimated density functions for the multiplicity of maxima.

It can be approximated by, e.g., the function

$$f(x) = u^2 e^{-u^{0.7}x} \tag{5}$$

After curve fitting this becomes

$$f(x) = 3.864e^{-1.605x} \tag{6}$$

and thus the probability to have multiple (e.g., 2 or 3) maximum in a random sequence s_1, ..., s_{i-1}, s_{i+1}, ..., s_M of weights is estimated to be

$$\int_2^3 3.864\, e^{-1.605x} dx = 0.078 \tag{7}$$

The simulation results show that there always are a few multiple maxima, and their proportion is not high.

4 Standard Effectiveness Using Test Collections

In order to evaluate the standard retrieval effectiveness of the AI^{2}R technique this was implemented in C++, and tested on the ADI and MEDLINE standard test collections.

Table 3. Statistics for the ADI test collection.

Subject Area	Information Science
Type	Homogeneous
No. of Documents	82
No. of Queries	35
No. of Terms	736
Mean no. terms/Document	11
Mean no. of Terms/Query	6

Table 4. Statistics for the MEDLINE test collection.

Subject Area	Medical Sciences
Type	Homogeneous
No. of Documents	1033
No. of Queries	30
No. of Terms	5732

Mean no. terms/Document	55
Mean no. of Terms/Query	9

Index terms were obtained automatically using a standard technique (stoplist and stemming). The standard 11-point recall–precision plots are shown in Figure 5 and 6. These show, for comparison purposes, the results of correlated search (CS) for the same ADI test collection [Bodner and Song, 1996], and of SMART [Deerwester et al., 1996].

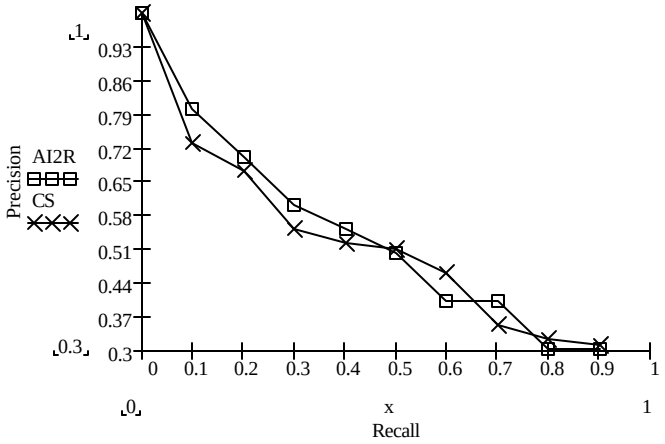

Figure 5. Recall–precision plot of AI^{2}R for the ADI test collection.

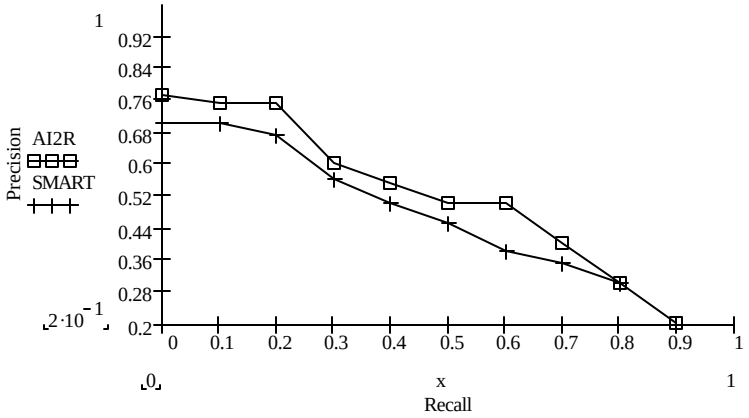

Figure 6. Recall–precision plot of AI^{2}R or the MED test collection.

The average precision at seen relevant documents is 0.55 in the AI^{2}R search, whereas 0.43 using SMART search. If we take, as usual, the SMART results as a reference, the AI^{2}R outperforms it by about 20%, and it compares well to CS.

5 Applications

5.1 i²r Interaction Information Retrieval

An application was developed which makes it possible to search the M.Sc. theses written (mostly in the Hungarian language) in the School of Technical Informatics at the University of Veszprém, Hungary, http://dcs.vein.hu/CIR/i2rapp/index.html). The application (written in a combination of C, Visual Basic, CGI, MathCAD) consists of several modules, which are described briefly.

Object Editor. This module makes it possible to create/edit the objects. An object consists of two distinct physical files stored on disk (easier creation, update and maintenance): one stores the effective data (text, etc.) of a document, while the other contains a sorted list (to speed up weights computation and search) of associated index terms.

Object Base Editor. This module makes it possible to create and modify (add or exclude objects) object bases containing object-documents. An object base corresponds to a network of connected documents.

Validation Module. This module carries out formal and consistency validations (e.g., character handling in query, existence and correct format of object base, time out, over- and underflow, etc.).

Search Module. This module is used online on the World Wide Web. It consists of (i) a series of user interfaces which are seen by the user, and (ii) of a set of search programs which carry out the retrieval itself.

Figure 7. Query formulation and object base selection.

Figure 8. Search results.

Figure 7 shows the page on which the user can enter the query, and can select the object base to be searched Figure 8 shows those two pages which contain the answer to the query. The first page (top one) contains a list of hits showing the major parameters of the documents found (title, author, year). If the user wishes to find out more about a specific document he/she can click on the respective list element, and a next page is shown on the screen (bottom one), from which he/she is pointed to further links (University Library).
Only the non-zero links are effectively stored, and this yields a considerably less storage capacity. We have also seen that the retrieval is computation demanding. Using a numeric encoding the time needed for the computation of connection strengths was reduced. A list is compiled containing each index term exactly once. Every index term is assigned a unique numeric value for two reasons: the comparison of numbers is much — more than ten times — faster than that of strings. This result was obtained by carrying out about a hundred thousand comparisons on equal integer, not equal integer, equal real, not equal real numbers, as well as equal and not equal strings using fast string comparison algorithms. This technique makes it possible to have a compact representation of the object base for retrieval purposes.

5.2 i^{2}rMeta

The I^{2}R Metasearch is a Web meta-search engine which was developed as another application of the AI^{2}R method (http://www.dcs.vein.hu/CIR/i2rmeta/i2rmeta.cgi).

The interface (Figure 9). This is a module written in PERL, and operates as follows. It accepts the user's query which it sends to four Web search engines (Altavista, Google, Webcrawler and Northernlight) as an HTTP request. The first twenty hits of the hit list returned by each of these search engines are taken into account. The corresponding web — HTML — pages are downloaded and processed: the tags are removed, the words are identified, stop listed and (Porter-) stemmed. Also an index is build with the thus obtained words and the corresponding URLs (to be shown as final answer). Thus every web page will generate an object for the AI^{2}R module. To gain speed the pages are downloaded in parallel (Parallel User Agent). Also a time limit is set so as to avoid waiting in vane for a non-responding server.

The AI^{2}R module (Figure 10). This module is written in C, and implements the AI^{2}R method. It interconnects the objects established by the Interface module as well as the query (which is treated as another object), establishes the reverberative circle(s), and passes the corresponding URLs (index table) to the Interface module which writes them out on the screen (for the user) in the order in which they appeared in the circle. To gain speed every word is assigned a unique numeric identifier, which are sorted. Thus weights computation is speeded up (point 5.1), and the weights can be calculated in just one pass through the sorted identifiers. Only those reverberative circles are established which are born following the maximal weights.

However, a control variable makes it possible to experiment (in the future) with circles born as second maximal weights.

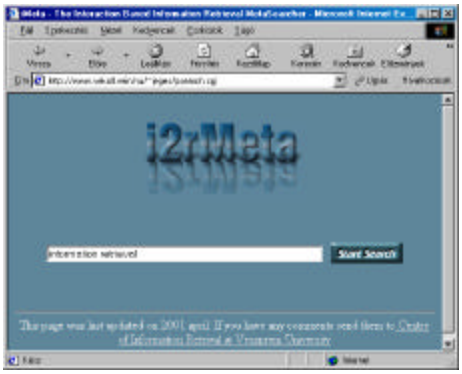

Figure 9. i2rMeta Interface.

Figure 10. i2rMeta search results.

6 Evaluations of the Applications

6.1 Evaluation by Users

The i2r Interaction Information Retrieval application was evaluated by its typical users, i.e., undergraduate students. They were asked to peform searches, and to qualify the returned documents as to how relevant they found them on a scale of four values as follows: not satisfied = 1, too few relevant documents = 2, satisfied = 3, very satisfied = 4. The results are shown in Figure 11.

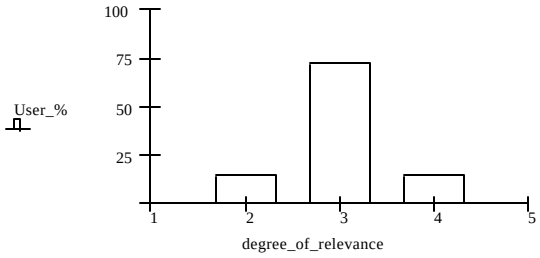

Figure 11. User satisfaction with relevance in AI^{2}R.

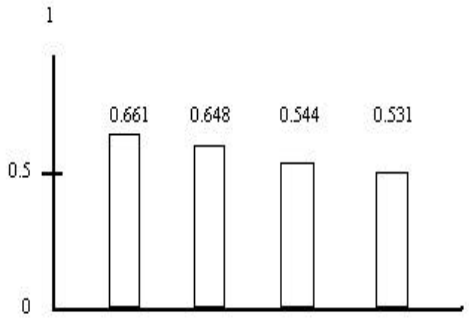

Figure 12. Average values for first five/ten precision.

The results show that almost 75% of the users were satisfied with the returned documents. 13% were very satisfied, and 12% thought that there were too few relevant documents in the answers.

These practical evaluations have confirmed what could be expected based on the results of the simulation and test collections, namely that the users are satisfied with the retrieval application based on AI^{2}R which favours high precision and low recall, and if we were to qualify the application we could say that it performs at an effectiveness level of about 75%.

The i2rMeta application was also evaluated by users. A number of fourty — undergraduate and postgarduate — users, coming from a variety of fields (computer science, chemistry, electrical engineering, tourism, theatre, philology, enviromnetal sciences) were asked to perform searches using i2rMeta. The queries were their choice, and they varied (e.g., Western digital specification, horned cattle, beggarly, country wedding, tesco, lyrid, meteors, wireless markup language, mp3, regular expression in perl, HPLC, perl tutorial, driver, development kit, Nokia Mobile Phones, Search Engines). Every user performed four searches (thus 120 searches were performed in all), and asked to give the total number of hits returend by i2rMeta as well as the number of hits they considered relevant. Thus precision could be estimated in each case, whose average value of this is 0.66.

6.2 First Five/Ten Precision

The i2r Interaction Information Retrieval application was evaluated for its ability to put relevant pages within the first five and ten links returned (this being the quality a typical undergraduate user would expect in our setting). Experiments were carried out based on a method suggested by [Leighton and Srivastava, 1999]. Relevance categories were established first, before evaluating any links. Separate searches were performed for each query. Each of the thus returned lists (HTML pages) was evaluated (the persons the information needs originated from were not involved in these evaluations, which is not an uncommon practice as this would be hardly feasible): placement in a relevance category, calculation of numeric precision. The categories are as

follows: (i) inactive links: file not found, forbidden, server not responding, (ii) duplicate links: the same link as an earlier one (e.g, duplicate theses titles due to multiple authors), (iii) category zero: irrelevant thesis, (iv) category one: relevant hit (either technically, i.e., the thesis title contains the search expression, and/or the thesis is judged to be relevant due to its content). A document is either in a category or not in the category. In these tests precision was calculated, which is preferred in undergraduate context. Other measures of user satisfaction (such as screen layout, etc.) were not incorporated.

First five precision. The classes and their weights for the first five precision case are as follows: (a) Class 1: contains the first two links, and has weight 10, and (b) Class 2: contains the next three links, and has weight 5. A weighted sum is computed as follows: $Links_{1,2} \times 10 + Links_{3,4,5} \times 5$. In order to obtain a final value for precision the above weighted sum should be divided by the total weighted sum corresponding to the case when there are five answers. This is expressed as $2 \times 10 + 3 \times 5 = 35$. If there are fewer links than five then 35 is decreased by 5 for every missing link. Thus the formula for the denominator is as follows: $35 - missing\_links \times 5$. Hence the final formula for the computation of the final value of precision is as follows:

$$\frac{Links_{1,2} \times 10 + Links_{3,4,5} \times 5}{35 - missing\_links \times 5} \tag{8}$$

First ten precision. The classes and their weights for the first ten precision case are as follows: (a) Class 1: contains the first two links, and has weight 20, (b) Class 2: contains the next three links, and has weight 17, and (c) Class 3: contains the last five links, and has weight 10. Taking into account the above, and following the line for the first five precision case the final formula in this case is as follows:

$$\frac{Links_{1,2} \times 20 + Links_{3,4,5} \times 17 + Links_{6\text{-}10} \times 10}{141 - missing\_links \times 10} \tag{9}$$

In the case of duplicate links the above formulae were used twice each: once when a duplicate link was counted as being just one link (without penalty), and once when one of the duplicate links was counted as being irrelevant (with penalty). Figure 12 shows the averages for the first five/ten precision values.

The average for the first five precision is 0.661 without penalty and 0.648 with penalty. The average for the first ten precision is 0.544 without penalty and 0.531 with penalty. The mean of these four values is 0.596. Precision effectiveness was much affected by irrelevant links the proportion of which was about 20%. As the proportion of duplicate links is relatively low in this application (it is not typical for a thesis to have several authors) their existence does not have a considerable affect on precision.

7 Discussion

Just like in, e.g., the Hopfield network the AI^{2}R network is a single-layered interconnected network, but unlike in the Hopfield net, where nodes are activated in parallel and then relaxed until a stable state is reached, in the AI^{2}R network nodes are sequentially activated according to a winner takes all strategy. In principle, convergence is represented by the stable state in the Hopfield net, and by the reverberative circle in the AI^{2}R network.

The suggested retrieval process, i.e., finding the reverberative circles, reminds of the Longest Circle problem which is known to be NP-complete. (The Longest Circle problem reads as follows: Given an *n*-node undirected graph, and a positive integer *k*. Does the graph contain a simple cycle having at least *k* nodes?) However, the two problems are not equivalent: first of all because of Theorem 2, and on the other hand because retrieval means precise traversal (follow the maximum) of the nodes starting from a fixed one (the query). The parallel or similarity might be caused by the following: the graph obtained after (i) having performed retrieval, (ii) having identified the circles, (iii) having kept only the links corresponding to maximums (but unweighted), allows for formulating the Longest Circle — but this is no longer the initial retrieval problem.

The average precision values were 0.555 for ADI, and 0.51 for MED, hence the mean precision given by the standard tests is (0.555 + 0.51) / 2 = 0.553. The AI^{2}R method was implemented in two applications whose precisions were evaluated using two series of experiments: with human subjects, and first five/ten prescision. The mean of the results obtained for the first five/ten precision is (0.661 + 0.648 + 0.544 + 0.531) / 4 = 0.596. The results of user evaluations can be combined into the following numeric value: $0 \times 0 + 0.333 \times 0.12 + 0.666 \times 0.75 + 1 \times 0.13 = 0.669$, which, averaged with the mean precision value for i2rMeta, gives (0.669 + 0.66) / 2 = 0.665. Table 5 summarizes the average results.

Series of experiments	**Mean precision**
Standard tests	0.553
First five/ten precision	0.596
User evaluations	0.665

Table 5. Summary of the series of experiments to evaluate the effectiveness of *I^{2}R*.

The evaluations based on the standard tests and the first five/ten precision method are practically close to each other: both operate under controlled conditions with well defined parameters. Surprisingly enough, the users' perception of precision was higher, and this might not have been totally expected based solely on the results of the other two series of experiments. These values reflect that there is not a direct and necessary logical implication or equivalence between the results of standard evaluations and those based on real human subjects .

8 Conclusions

An adaptive clustering technique for information retrieval was suggested based on the interaction IR method using neural networks.

It was shown that the complexity of the computations involved is polynomial, hence the AI^2R method is tractable. An estimation was given as to the probability to have multiple reverberative circles, and this was found to be 0.07-0.08.

Standard test collections were used to evaluate the classical effectiveness of the AI^2R method. The results show that it is useful when high precision is favoured at low to middle recall values.

Two applications based on the AI^2R method were also designed and presented briefly. Their precisions were evaluated by experiments carried out with human subjects, and using a first five/ten precision method. The results of these series of experiments show that both applications meet very well users' satisfaction.

In all, the precision level expected for such a retrieval method and system is in the area 50%–70%.

These series of theoretical research, tests and experiments constitute an exhaustive and methodical evaluation of a retrieval technique based on adaptive clustering using a connectionist approach. At the same time they perhaps constitute a line and methodology recommended for a complete research and analysis of any retrieval method and system.

Acknowledgements

The author would like to thank the National Science Foundation, Hungary (OTKA T 030194), for financially supporting this research, and also E. Jeges, A. Nagy, and A. Skrop for their help in developing the application and carrying out the tests and simulations.

References

Baeza-Yates, R., and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*, Addison Wesley.

Belew, R. K. (1989). Adaptive information retrieval: Using a connectionist representation to retrieve and learn about documents. *Proceedings of the 12th ACM SIGIR '89*, pp. 11-20, Cambridge, MA, ACM Press.

Bodner, R., and Song, F. (1996). Knowledge-based approaches to query expansion in information retrieval. In McCalla, G. (ed.) *Advances in Artificial Intelligence*, Springer, 146-158.

Carrick, C. and Watters, C. (1997) Automatic Association of News Items. *Information Processing and Management*. 33(5): 615-632.

Chen, H., Hsu, P., Orwig, R., Hoopes, L., and Nunamaker, J.F. (1994). Automatic concept classification of text from electronic meetings. *Communications of the ACM*, **37**(10): 56-73.

Chen, H. (1995). Machine Learning for information retrieval: Neural networks, symbolic learning and genetic algorithms. *Journal of the American Society for Information Science*, **46**: 194-216.

Cohen, P., and Kjeldson, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management*, **23**: 255-268.

Crawford, S.L., Fung, R., Appelbaum, L.A., and Tong, R.M. (1991). Classification trees for information retrieval. *Proceedings of the 8th Workshop on Machine Learning*, Morgan Kaufmann, 245-249.

Cunningham S.J., Holmes G., Littin J., Beale R., and Witten I.H. (1997). Applying connectionist models to information retrieval. In: S. Amari, and N. Kasobov (eds.*) Brain-Like Computing and Intelligent Information Systems*, pp 435-457. Springer-Verlag.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, **41**: 391-407.

Dominich, S. (1994) Interaction Information Retrieval. *Journal of Documentation*, 50(3): 197-212.

Dominich, S. (1997). The Interaction-based Information Retrieval Paradigm. In (Kent, A., ed.) *Encyclopedia of Library and Information Science*, Vol. 59, Suppl. 22, Marcel Dekker, Inc., New York Basel Hong Kong, 218-236.

Dominich, S. (2001). *Mathematical Foundations of Information Retrieval*. Kluwer Academic Publishers, Dordrecht, Boston, London.

Doszkocs, T., Reggia, J., and Lin, X. (1990). Connectionist models and information retrieval. *Annual Review of Information Science & Technology*, **25**: 209-260.

Fisher, D.H., and McKusick, K.B. (1989). An empirical comparison of ID3 and back-propagation. *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, Detroit, MI, 788-793.

Hearst, M. A., and Pederson, J. O. (1996). Reexamining the cluster hypothesis: Scatter/Gather on retrieval results. *Proceedings of the 19th Annual International ACM SIGIR Conference*, Zurich.

http://dcs.vein.hu/CIR. *AI²R Application.*

Johnson, A., Fotouhi, F., and Goel, N. (1994). Adaptive clustering of scientific data. *Proceedings of the 13th IEEE International Phoenix Conference on Computers and Communication*, Tempe, Arizona, pp. 241-247.

Johson, A., and Fotouhi, F. (1996). Adaptive clustering of hypermedia documents. *Information Systems*, **21**: 549-473.

Kraft, D.H., Bordogna, P. and Pasi, G. (1998). Fuzzy Set Techniques in Information Retrieval. *In*: Didier, D. and Prade, H. (eds.) *Handbook of Fuzzy Sets and Possibility Theory. Approximate Reasoning and Fuzzy Infomation Systems*. Kluwer Academic Publishers, AA Dordrecht, The Netherlands, Chp. 8.

Lebowitz, M. (1987). Concept learning in a rich input domain: Generalization-based memory. In Carbonell, J.G., Michalski, R.S., and Mitchell, T.M. (eds.) *Machine Learning, An Artificial Intelligence Approach, Vol. II.*, Morgan Kaufmann, 193-214.

Leighton, H. V., and Srivastava, J. (1999). First Twenty Precision among World Wide Web Search Services (Search Engines). *Journal of the American Society for Information Science*, **50**(10): 870-881.

Liu, G.Z. (1997) Semantic Vector Space Model: Implementation and Evaluation. *Journal of the American Society for Information Science*, 48(5): 395-417.

Mather, L. A. (2000). A Linear Algebra Measure of Cluster Quality. *Journal of the American Society for Information Science*, **51**: 602-613.

Mobasher, B., Cooley, R., and Srivastava, J. (1998). Creating Adaptive Web Sites Through Usage-Based Clustering of URLs. Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange,

Mock, K.J. and Vemuri, V.R. (1997) Information Filtering via Hill Climbing, Wordnet and Index Patterns. *Information Processing and Management*. 33(5): 633-644.

Pearce, C. and Nicolas, C. (1996) TELLTALE: Experiments in a Dynamic Hypertext Environment for Degraded and Multilingual Data. *Journal of the American Society for Information Science*, 47(4): 263-275.

Rose, D. E. (1994). *A symbolic and connectionist approach to legal information retrieval*. Hillsdale, NJ, Erlbaum.

Salton, G., Allan, J. and Singhall, A. (1996) Automatic Text Decomposition and Structuring. *Information Processing and Management*. 32(2): 127-138.

Salton, G., and McGill, M. (1983). *Introdcution to Modern Information Retrieval*. McGraw Hill, New York.

Salton, G., Singhall, A., Mitra, M. and Buckley, C. (1997) Automatic Text Structuring and Summarization. *Information Processing and Management*. 33(2): 193-207.

Sebastiani, F. (1994) A probabilistic termonological logic for information retrieval. *ACM SIGIR 17th International Conference on Research and Development in Information Retrieval*, Dublin, Ireland, Springer, London, 122-130.

Shaw, W., Burgiu, R., and Howell, P. (1997). Performance standards and evaluations in IR test collections: Cluster-based retrieval models. *Information Processing and Management*, **33**: 1-14.

Tanaka, H., Kumano, T., Uratani, N., and Ehara, T. (1999). An efficient document clustering algorithm and its application to a document browser. *Information Processing and Management*, **35**(4): 541-557.

van Rijsbergen, C. J. (1979). *Information Retrieval*. Butterworth, London.

van Rijsbergen, C. J. (1996). Quantum Logic and Information Retrieval. *Proceedings of Workshop on Logical and Uncertainty Models in Information Retrieval*, University of Glasgow, Glasgow, July, 1-2.

Vorhees, E. (1985). The cluster hypothesis revisited. *SIGIR*, 188-196.

Willett, P. (1988). Recent trends in hierarchic document clustering. *Information Processing and Management*, **24**: 577-597.

Yu, C.T., Suen, C., Lam, K., and Siu, M.K. (1985). Adaptive record clustering. *ACM Transactions on Database Systems*, **10**(2): 180-204.