

An Index Navigator for Understanding and Expressing User's Coherent Interest

Yukio OHSAWA* and Masahiko YACHIDA

Dept. Systems and Human Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560
Japan

Abstract

It is important to guide a user to interesting documents in a large-scale document-database. However, when the user is not an expert of the area of interest, it is difficult for him/her to name precise keywords in which he/she is interested, nor to select areas of his/her own interest. This paper presents an Index Navigator which clarifies what areas the user is interested in, what keywords he/she should investigate, and what documents concern his/her interest. A tough problem for such a system is to understand interesting areas for the user, among other areas-sets which can explain his/her behavior. Our Index Navigator employs an inference method called Cost-based Cooperation of Multi-abducers (CCMA), for understanding user's interest from the history of the user's expression of interest in incomplete keywords, even if the changing speed of the user's interest is totally unknown. With this device, the Index Navigator guides the user to really important areas, keywords and documents.

1 Introduction

Recently, very large amounts of text have been stored in several types of media such as CD-ROMs, electronic libraries, and the Internet. In such situations, a user desiring important documents often loses his or her way because there are too many information resources.

In order to cope with this problem, methods have been presented to make a match between the user's interest and the documents. A major strategy is to represent the user's interest by a keyword-vector, in which each keyword is assigned a weight based on the frequency it was read or written by the user [Krulwich, 1995], or by the sheer keywords requested by the user. This vector is used to match keyword-vectors of documents in hand. Unfortunately, especially when the user is not familiar with the domain of the desired document, this matching becomes extremely difficult because of the discrepancy

between the terms in the domain and the vocabulary of the user [Furnas et al, 1987]. Also, the user may fail to name precise keywords in which he/she is interested, because of the poor expressiveness due to haste.

In order to cope with this problem, we aim at improving the user's ability to express his/her interest. We propose a method to *infer* the areas of the user's interest from *incomplete* (poor) keywords he/she requested. Additional keywords are obtained from these areas (classifying the interest to areas is beneficial itself too, e.g. for searching documents in a library classified by areas, or investigating what academic society the user should subscribe). By seeing additional keywords, the user learns to express his/her own interest, which is useful when the user writes, reads, or searches documents from a larger set of documents in electronic libraries or the Internet.

A tough problem in this inference, however, is again the lack in the keywords requested by the user - areas cannot be uniquely determined from too few keywords. An effective approach to obtain unique and precise areas of the user's interest is to consider the history of requested keywords, i.e. inferring the user's current interest to be coherent with his/her recent interests. However, changes in the user's interest are difficult to understand. For example, a user who has been interested in using syntactic rules for machine translation may, one day, face a problem that sentences cannot be translated without considering semantics. He may change his interest from syntax to semantics. Such a change comes from newly encountered situations, unable to be expressed in the user's model acquired from the user's past behaviors [Armstrong et al, 1995; Krulwich, 1995]. In [Balanovic and Shoham, 1995], the user's model is revised dynamically by changing weights of words based on their frequency in documents the user accessed recently. However, frequencies of incomplete keywords compose incomplete information about the change in the user's interest, e.g. shifting from the use of "Indonesian" to "English" does not mean that interest in "translation" from Indonesian to English disappeared.

In this paper, we present an *Index Navigator* which clarifies to a user what *areas*, *keywords*, and *documents* match his/her interest, based on the history of the keywords the user selected. For understanding the user in-

*Email: osawa@sys.es.osaka-u.ac.jp

terest transitions, the Index Navigator uses an inference method called *Cost-based Cooperation of Multi-abducers* (CCMA) [Ohsawa and Yachida 1997]. CCMA obtains a coherent explanation of user's history of selected keywords, to classify the user's current interest to areas. Since CCMA understands an event-sequence even if the changing speed is totally unknown, our Index Navigator allows for unpredictable changes in the user's interest. The Index Navigator presents areas, keywords, and documents matching the user's interest, so that the user learns to express his/her own interest.

2 The Outline of Index Navigator

In this paper, we present an Index Navigator, which supports a user who searches documents matching his/her interest, but cannot express his/her interest fully by keywords nor content areas.

A user of the Index Navigator selects keywords, i.e., index terms he/she can request, by clicking on a given *Input-Index List* (the largest sub-window in Figure 1) or by entering keywords in the upper left window in Figure 1. Each time the user selects a set of keywords, a set of areas are inferred corresponding to the user's current interest, from the history of selected keywords, by the *Interest Tracer*. These inferred areas are listed in the *Areas List*, in the middle right of Figure 1. Keywords expressing the user's interest are listed in the *Output-Index List*, the upper right of Fig.1. At the same time, documents corresponding to the user's interest are displayed in the *Document List*, at the bottom of Fig.1.

In the remainder of this paper, we first sketch the mechanism of the Index Navigator in Section 3 where the subsystems of the Index Navigator, each executing inference to present the Areas List, the Output-Index List, and the Document List, are stated. Then, we evaluate the performance of the Index Navigator experimentally in Section 4 and discuss about the results. We make conclusional remarks in Section 5.

3 The Mechanism of Index Navigator

The Index Navigator is composed of three subsystems. First, the *Interest Tracer* fills in the Areas List. Then, the Output-Index List and the Document List are filled in, in the second and the third subsystems respectively. The following subsections describe the details of these subsystems.

3.1 Interest Tracer

In order to consider changes in user's interest, we employ an inference system called *Cost-based Cooperation of Multi-abducers* (CCMA), which explains an event-sequence by hypotheses (corresponding to the causes), even if the changes in causal events occur at a totally unknown speed. In this subsection, let us show why and how we use CCMA for the Interest Tracer.

Coherent explanation of selected keywords

The history of the user's behaviors can be regarded as a coherent sequence of events (caused by the user's coher-

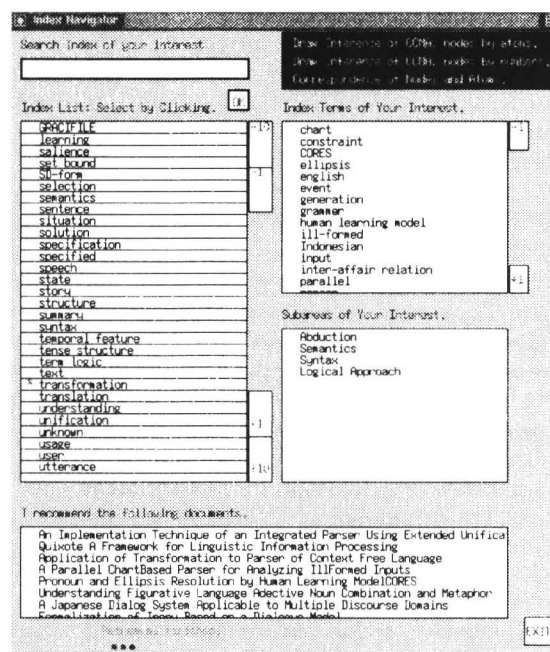


Figure 1: A window of the Index Navigator.

ent interest), which has been considered in many works. In probabilistic models, *transition probabilities* represent the coherence among events at nearby moments or places [Kashino et al, 1995; Chen et al, 1994]. However, a transition probability, i.e., the occurrence probability of an event after a fixed or unfixed interval from a preceding event cannot be determined if the situation changes at a totally unknown speed, though it may work to obtain probable areas of a user or a document given without history [Fung and Favero, 1995].

Coherence among sequential events has been formalized in [Ng and Mooney, 1990], where *Explanatory Coherence* (EC, hereafter) is defined by the number of common hypotheses shared by the solution hypotheses-sets of *abduction* (*hypothetical reasoning*). In abduction, causes of observed events (goals) are inferred as a solution hypotheses-set [Poole et al, 1987; Poole, 1993]. Abduction on EC obtains the solution hypotheses-sets of maximal explanatory coherence. EC is useful even if the speed of changes in events are unknown, because it does not rely on any predefined parameter of changing speed, like transition probabilities. However, EC cannot evaluate an explanation of observed events, if they do not have common causes, i.e. if the situation changes radically. On the other hand, in cost-based abduction (CBA, hereafter), a weight is assigned to each element hypothesis and the best solution hypotheses-set is defined as the hypotheses-set with the lowest cost, where cost is the sum of weights of element hypotheses in the solution hypotheses-set. The solution hypotheses-set of CBA is meaningful in that it corresponds to the MAP (maximum a-posteriori) explanation of one goal [Char-

niak and Shimony 1994]. However, CBA does not consider the coherence in the sequence.

For understanding an event-sequence changing in an unknown speed, both of these two criteria, i.e. the MAP explanation in CBA and the coherence considered in EC, are merged in CCMA.

Process and Semantics of CCMA

The fundamental mechanism and algorithm of CCMA was presented very recently [Ohsawa and Yachida 1997](in Japanese), as a new distributed reasoning framework. However, CCMA has been applied only to *non-real world* toy problems, e.g., artificially given varying fault-states (bad connections which is sometimes naturally fixed by unpredictable external forces) in electronic circuits such as small full-adders. This restriction in the previous applications of CCMA came from the fact that the semantics of CCMA was yet under construction and all the efforts have been devoted to analyzing "why and when CCMA works" by solving artificial (i.e. intentionally produced) problems where too complex elements in the *real-world* to take into account are not included (more detailed experimental data than [Ohsawa and Yachida 1997] about the analysis of CCMA on artificial problems will appear in future in English, which shows that CCMA catches state transitions better than previous distributed reasoning methods e.g. FA/C [Carver and Lesser, 1995]).

With our later progress in formulating the semantics of CCMA (as in Eq.(3) and Eq.(4) below), we came to deal with unpredictable changes in the *real world* such as the will of users of document search engines. The Interest Tracer is the first application of CCMA to a real world problem: understanding the changing interest of users - changing in unpredictable speed moved by complex elements in the real world (emotional occasions or accidentally encountered situations) which is hardly included in artificial problems such as ones we considered so far.

Here, let us introduce reader to CCMA's fundamental algorithm and the updated formulation of CCMA semantics.

In CCMA, multiple abduction systems (*abducers*, hereafter) of CBA are assigned to goals (events observed in an event-sequence), and each abducer obtains an optimal (lowest cost) solution hypotheses-set to explain its goal. Each abducer iterates cycles of revising weights of hypotheses according to the explanations of recent events by recent abducers and then explaining its own goal again, until the solution hypotheses-sets converge to a coherent explanation of the sequential events.

Formally, the problem of CCMA is described as finding solution-sets $h_j (j = 1, 2, \dots, N)$, which satisfies the following constraints in Eq.(1) where N is the length of the events-sequence, i.e., the number of observed events. Here, knowledge is composed of *background knowledge* Σ (knowledge given as truth in any environment) and hypothetical knowledge H (a set of element hypotheses which may be false and may contradict others). G_j is the

j th event in the sequence. The solution to be searched is a set of hypotheses-sets $h_j \in 2^H$ which, if combined with Σ , supports G_j for each $j (j = 1, 2, \dots, N)$ without contradiction (ϕ denotes the empty clause).

$$\begin{aligned} h_j &\in 2^H, \\ h_j \cup \Sigma &\vdash G_j, \\ h_j \cup \Sigma &\not\vdash \phi. \end{aligned} \quad (1)$$

We deal with Σ composed of causal rules expressed in Horn clauses as Eq.(2), which means that y is true if all of $x_1, x_2, \dots, x_n$ are true, where variables $x_1, x_2, \dots, x_n$ and y are atoms each of which corresponds to an event occurrence. When applied to our Index Navigator, x is true if x , for example, means "semantics" and the user is really interested in semantics.

$$y : -x_1, x_2, \dots, x_n. \quad (2)$$

Also, a probabilistic causality can be described by assigning a conditional atom x_{n+1} representing an event of probability $p(x_{n+1})$. The *weight* of x_i is defined by $weight(x_i)$ equal to $-\log p(x_i)$, by which the solution of CBA corresponds to the MAP hypotheses as we stated above.

For example, in Figure 2, leaves denote the given hypotheses (content areas of the user's interest) H and other nodes denote goals (selected keywords). Edges in Figure 2 denote the causal links between areas and keywords, e.g., the left-most graph means a Horn clause

$$GPSG : -syntax, p_1,$$

where p_1 is an atom representing an event of probability $p(p_1)$, i.e., (the number of documents including "GPSG") / (the number of documents in the area of syntax analysis) (note: GPSG is a grammatical analysis method). Also, the probabilities of "GPSG" and "syntax" are (the number of documents including "GPSG") / (D := the number of documents in the database) and (the number of documents in the area of "syntax") / D , respectively. This implies that we approximate the interest of a user by the interest of the society, i.e. the number of accepted documents. Weights of hypotheses are defined on these probabilities. In Figure 2, "GPSG," "transformation," "understanding," and "translation" depict keywords that the user selected sequentially at time t (G_1), $t + dt$ (G_2), $t + 2dt$ (G_3), and $t + 3dt$ (G_4) respectively. Bold-face numbers depicted below the nodes in Figure 2 are the weights of corresponding hypotheses (virtual weights for a simple illustration). As a measure of coherence, the solution hypotheses-set for G_i is selected by minimizing C_{1j} in Eq.(3).

$$C_{1j} = \sum_{\eta} weight(\eta, s.t., \eta \in h_j, \eta \notin h_{j+1} \cup h_{j-1}), \quad (3)$$

i.e. hypotheses *not believed* in adjacent abducers are *least believed*. Here we have the notion of coherence in EC, where hypotheses *believed* in adjacent abducers are *strongly believed*, and also the notion of lowest-cost solution employed in CBA. If minimizing C_{1j} is not enough

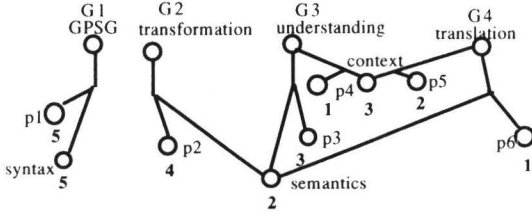


Figure 2: A knowledge-base for CCMA. Nodes at the bottom (i.e. leaves) depict areas, and nodes at the top depict requested keywords.

to define a unique solution, we should focus attention to more strongly believed hypotheses by minimizing

$$C_{2j} = \sum_{\eta} \text{weight}(\eta, s.t., \eta \in h_j, \eta \notin h_{j+1} \cap h_{j-1}). \quad (4)$$

In CCMA, the minimization of C_{1j} and C_{2j} are realized by distributed reasoning. Here, the j th abducer is the abducer which explains the j th goal G_j from Σ and H by the lowest-cost solution hypotheses-set h_j . The solution hypotheses-set h_j is a set of atoms in H whose cost, defined in Eq.(5), is the lowest of all the feasible explanations of G_j . Here, η_{ij} is the i th element hypothesis in H in the j th abducer, whose weight is w_{ij} .

$$\text{cost}_j = \sum_{j|\eta_{ij} \in h_j} w_{ij}. \quad (5)$$

Variables appearing in the following procedure of CCMA, are defined as, k : the count of distributed abduction cycles of the multiple abducers, t_{ij} : the truth value of η_{ij} defined as 1 for true, 0 for false. The item in Table 1 for $\{t_{i(j-1)} + t_{i(j+1)}\}$, (current) state of η_{ij} denotes the next value of w_{ij} , i.e., “ w_{ij} ” in Table 1 means that w_{ij} stays with the current value and “ w_{oij} ” denotes that w_{ij} changes into w_{oij} , the value of w_{ij} given initially when k was 0. Refractory periods are explained later.

The procedure of CCMA

1. $k = 0$.
2. $k = k + 1$. Obtain the lowest cost solution h_j for G_j by the j th abducer. Put η_{ij} to an *absolute refractory period* in the k th cycle if t_{ij} came to be 1 in the k th cycle and to a *relative refractory period* if t_{ij} came to be 0 in the k th cycle but was 1 in the $(k - 1)$ th. Otherwise put η_{ij} to a *reactive period*.
3. Change w_{ij} as shown in Table 1. **Stop** if none of w_{ij} for $j = 1$ to N is changed here. Otherwise Go to 2.

Briefly, changing a weight means a belief revision in the corresponding hypothesis. w_{ij} is set to 0 in the $k - 1$ th cycle which is a reactive period, if the corresponding hypotheses in adjacent abducers (i.e. $\eta_{i(j-1)}$ and $\eta_{i(j+1)}$) are inferred as true, by which C_{1j} in Eq.(3) is minimized locally in the j th abducer in the k th cycle.

Refractory periods, in which weights are not reduced so easily as in a reactive period, prevents CCMA from

$t_{i(j+1)} + t_{i(j-1)}$	0	1	2
State of η_{ij}			
reactive period	w_{ij}	0	0
relative refractory period	w_{ij}	w_{ij}	0
absolute refractory period	w_{oij}	w_{ij}	w_{ij}

Table 1: Change of w_{ij} , for the state of η_{ij} and value of $t_{i(j-1)} + t_{i(j+1)}$.

limitless cycles. For example, Figure 3 shows the process of CCMA for Figure 2. Each graph in Figure 3 is a subtree of Figure 2, extracted in a top-down manner beginning with each goal. In Cycle 1, abducers infer that the user is interested in $\{h_1: \text{syntax}, h_2: \text{semantics}, h_3: \text{context}, h_4: \text{semantics}\}$ because these are the lowest-cost hypotheses-sets in abducers 1, 2, 3 and 4 (for G_1, G_2, G_3 and G_4) respectively. Messages are sent as the dashed arrows in Figure 3, which change the weights of the hypotheses at the arrowheads, e.g., the arrow from “semantics” in abducer 4 to “semantics” in abducer 3 in Cycle 1 reduces the weight of the latter and enhances the belief in the interest in “semantics” of abducer 3 in Cycle 2. In Cycle 3, the procedure stops because the weights do not change any more. The result is that the user is coherently interested in “semantics”, after the change from “syntax” to “semantics” at time $t + dt$. Thus, the coherent interest of the user and the radical change in the interest are obtained. Generally, the propagation of the weight-changing message from a hypothesis successfully stops at the abducer whose goal cannot be supported by the same hypothesis. However, suppose that “context” of G_3 receives a message from “context” of G_4 in Cycle 2, it becomes true again in Cycle 3 as well as in Cycle 1, so that abducers return to the same states as in Cycle 1 - hence limitless cycles.

Refractory periods extinguish the causes of limitless cycles (not completely but mostly) by two main effects. First, in an *absolute refractory period*, η_{ij} totally ignores the belief of the adjacent abducers in $\eta_{i(j+1)}$ and $\eta_{i(j-1)}$, which may be due to a previous message from η_{ij} . Second, in a *relative refractory period* of η_{ij} , η_{ij} is believed more strongly if it is believed by both of its neighboring abducers, than if it is believed by only one of them. This helps focusing the solution to the one of the minimal C_{2j} (see Eq.(4)) and avoiding too many hypotheses exchanged for one observation, as “semantics” and “context” in Figure 3.

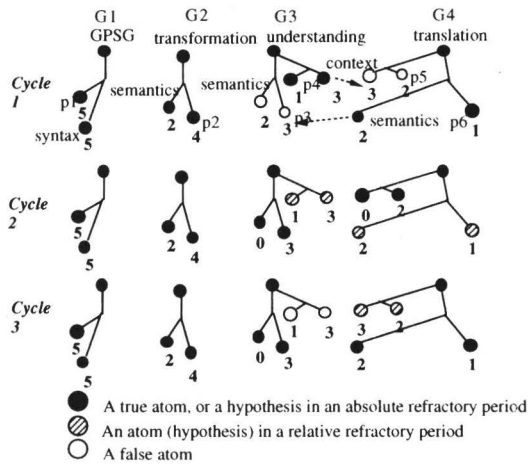


Figure 3: CCMA process for the sequence of *G1*, *G2*, *G3*, and *G4*.

3.2 Generating the Output-Index List and the Document List

Keywords and documents relative to the intersection of areas in the Areas List are presented in the Output-Index List and the Document List respectively. For example, for the Horn-clause in Eq.(6), *keyword0* is obtained if *area1* and *area2* are in the Areas List, with an assigned priority of p_0 . The priority of *keyword1* becomes $p_1 \cdot p(\text{area3})$ in this case, which is normally less than p_0 (if p_0 is not far less than p_1 and $p(\text{area3})$ is far less than 1). Keywords in some area in the Areas List are sorted by priority values and listed in the Output-Index List,

$$\begin{aligned} \text{keyword0} &: -\text{area1}, \text{area2}, p_0. \\ \text{keyword1} &: -\text{area1}, \text{area3}, p_1. \end{aligned} \quad (6)$$

A document in the Document List is selected similarly. The only difference is that another knowledge-base including rules as in Eq.(7) is used, where the probability of p_0 is given by $1/(\text{number of documents in } \text{area1} \cap \text{area2})$.

$$\text{document0} : -\text{area1}, \text{area2}, p_0. \quad (7)$$

4 The Performance of the Index Navigator

The system is coded in C++ and the interface is in X-window, on a Pentium 166MHz computer with 64M RAM running Linux. In the experiment here, the Index Navigator is used with a knowledge base acquired from the data-base in the CD-ROM of the Journal of Japanese Society of Artificial Intelligence, in HTML format, in which 741 papers are stored, given keywords, and classified by areas. Ten students in the engineering field, who know nothing about the mechanism of CCMA, used this system for one week concentrating on retrieving interesting documents. These students were assigned the

task of presenting reports at the end of the week, about the documents they read.

Four windows of Index Navigators, similar to Figure 1, were displayed, each employing an Interest Tracer realized by CCMA, CBA, EC, and Bayesian Networks. Users could select favorite windows without being bound by prejudice on inference methods, because these windows were marked simply "A", "B", "C" and "D" respectively (though the keywords selected are shared by all four methods, the user selected keywords by clicking on the Input Index List in his/her favorite window).

During the week, the user's selection of keywords and the inference results (presented lists) were recorded automatically. Also, in order to construct a Bayesian network to be compared with CCMA, probabilities were calculated for any transition from the interest in one area to another at the next time a user selects keywords (areas of the user's interest at a time were determined by the areas in the Areas List in the window he/she selected).

4.1 Evaluation of the Performance of the Index Navigator

The usefulness of the Interest Tracer by CCMA

Figure 4 shows the changes in the number of users who clicked on the Input Index List in each of the four windows for each three hours (the periodical increase and decrease are due to days and nights).

Multiple different ways to interpret the results in Figure 4 may be possible. First of all, the curves depict that the number of users of every method decreased with date to be none finally, which may seem to imply that all the tested methods were judged to be useless at last. However, this is not the case because the tested students were assigned the task of presenting a report about their retrieved documents. That is, they had to spend time reading the retrieved documents *carefully* and for finishing their reports in the last few days, rather than retrieving new documents and learning to express their interest in keywords. Secondly, those methods whose users came to be none early may seem to be so sophisticated that they satisfied users quickly, because users of one method did not increase with the decrease of other methods, according to Figure 4. However, this is neither the case – according to the detailed data obtained by monitoring users' behaviors, users tried all the methods (windows A, B, C, and D) in the early days of the week and quit to use useless ones – it was very rare that a user suddenly began to use methods new to him/her (in other words, users seldom stopped their tasks on the early days by being satisfied with the information they obtained). Thus, we interpret a method to be the more useful, the longer used it was.

The performance of Bayesian network grew first due to its learning, but CCMA overwhelmed this afterward. According to the users' comments, the inferred areas by Bayesian network rapidly went out of their interests. For example, when a Bayesian network was used and a user's interest seldom stayed in "semantics" for a long time on the first day, it was inferred that his interest changed to

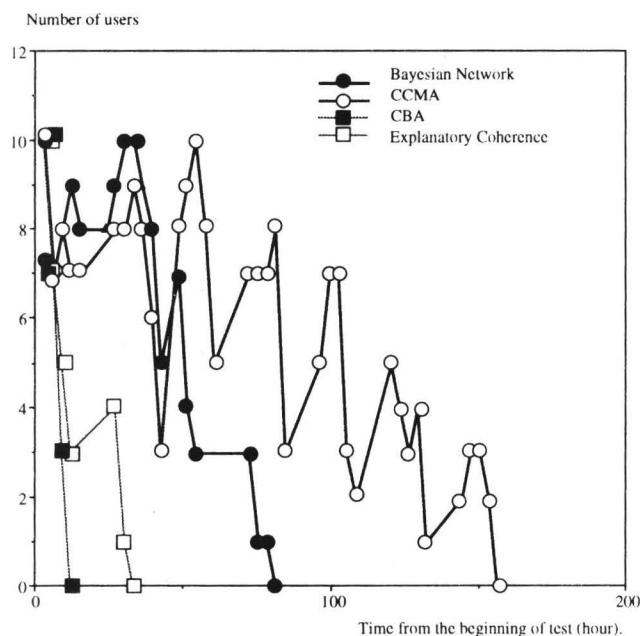


Figure 4: Number of users of Index Navigator employing variable methods

area “syntax” even when he selected “translation” after “understanding”, i.e. even though he was still interested in “semantics” on the second day. This is because the transition probabilities learned on the first day were used even though they no longer worked - the changing speed of user’s interest varies on many factors, e.g., whether he is trying to catch the overall perspective of a field or to understand one concept deeply. On the other hand, by CCMA this user was inferred to be interested in the area “semantics” throughout the second day. Also, his interest was inferred as in “syntax” area when he selected “translation” after “GPSG”, and this was a correct understanding on the first day. The performance of CBA (employing a CBA reasoner for one keywords-set selected at each moment independently of other moments, in the Interest Tracer) was the worst, because CBA does not consider the coherence at all. Finally, *too coherent* history of areas came out by EC (Explanatory Coherence). Whenever EC is used, areas which *are* now not, but *were* in the user’s interest previously, came into the Areas List. This is especially troublesome when the user’s interest changes from one area to another, because areas of his/her previous interest also appear in the areas-list and thus decreases the extracted keywords and documents.

The keywords in the Output-Index List worked

Although the Index Navigator itself cannot be applied to the entire Internet yet, because not all the pages are classified by areas, the keywords learned by using the Index Navigator can be used as terms to search for interesting pages using existing search engines.

The pages extracted by the search engine TITAN (<http://isserv.tas.ntt.jp/chisho/titan.html>) after

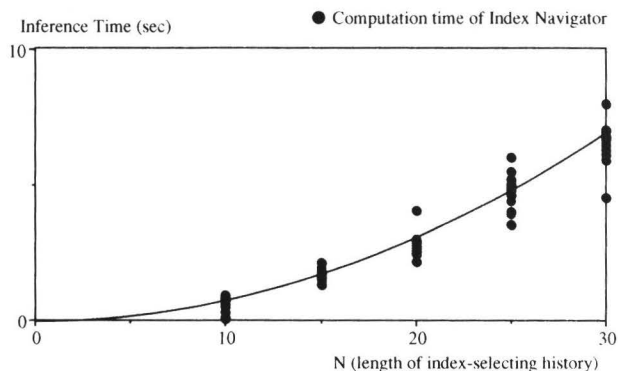


Figure 5: The time performance of CCMA.

the testing week drastically came to fit the user’s interest. Only 7.2% accessible pages (pages except those already closed or forbidden to access) in TITAN’s output matched the tested users’ interest before the testing week, which grew to 92.2% after the week.

The speed of the Interest Tracer by CCMA

The most time-consuming component in the Index Navigator is the Interest Tracer. Here, we sped up the abducers in the CCMA of the Interest Tracer by the *Networked Bubble Propagation* (NBP) method, a high-speed CBA method. NBP obtains a near lowest cost solution of CBA within a polynomial-time of $O(|H|^2)$, where $|H|$ is the number of hypothesis in H [Ohsawa and Ishizuka 1997]. We employed this approximate solution method, because multiple abducers compensate for inaccuracy (difference from the exact optimal solution) in each other. The CCMA calculation time is approximately $O(N^2 \cdot |H|^2)$, since each of N abducers takes $O(|H|^2)$ in one cycle, and N cycles run before all the messages propagate over N abducers where N is the length of the sequence.

The time taken for $N = 1$ to 30 of the Interest Tracer by CCMA is shown in Figure 5. The data fits $O(N^2)$, corresponding to the above anticipation of $O(N^2|H|^2)$ considering that $|H|$ is independent of N . The current Index Navigator considers keywords selected at 20 recent moments, to make the system respond within 5 second.

5 Conclusions

We presented an Index Navigator, which provides a user with keywords that express his/her interest. A user who learns to express his/her own interest in keywords by the Index Navigator becomes able to search for important documents in larger information sources. The successful results show that user interests are understood correctly, even if the user interest changes radically. CCMA, an inference method for understanding changing event-sequences, achieves good inference for this purpose, and also good time-performance, i.e. polynomial against the size of knowledge and the length of

the sequence.

The Index Navigator is applied to a CD-ROM currently and users are indirectly aided to use the Internet, if they will, by learning to express their interests. In order to make the Index Navigator useful as a tool to directly search documents from a larger document collection e.g. an electronic library or the Internet, the inference speed of CCMA will be an important issue. In order to apply the Index Navigator to such large-scale document collections, our next plan is to re-install the distributed inference of CCMA on a parallel computer, which will further improve the inference-speed.

Acknowledgments

We appreciate Prof. Mitsuru Ishizuka (the University of Tokyo) and Prof. Toru Ishida (Kyoto University) for helpful comments. Also, we express special thanks to Assoc. Prof. Seiji Yamada (Tokyo Institute of Technology), Assoc. Prof. Takashi Washio (Osaka University) and Dr. Tomohiro Yamaguchi for constructive discussions.

References

- [Armstrong et al, 1995] Armstrong,R., Freitag,D., Joachims,T. and Mitchell,T., WebWatcher: A Learning Apprentice for the World Wide Web, *AAAI Spring Sympo. Series on Information Gathering from Distributed, Heterogeneous Environments*, 1995¹.
- [Balabanovic and Shoham, 1995] Balabanovic,M., Shoham,Y., Learning Information Retrieval Agents: Experiments with Automated Web Browsing, *AAAI Spring Sympo. Series on Information Gathering from Distributed, Heterogeneous Environments*, 1995.
- [Carver and Lesser, 1995] Carver, N., Lesser, V., The DRESUN Tested for Research in FA/C Distributed Situation Assessment, *Proc. Int'l Conf. Multiagent Systems*, 33-40, 1995.
- [Charniak and Shimony 1994] Charniak,E., Shimony, S.E., Cost-based abduction and MAP explanation, *Artif.Intell.* 66(2): 345-374, 1994.
- [Chen et al, 1994] Chen, M.Y., Kundu, A. and Zou, J., Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type Stochastic Network, *IEEE Trans. PAMI*, 16, 481-496, 1994.
- [Fung and Favero, 1995] Fung,R., Favero,B.D., Applying Bayesian Networks to Information Retrieval, *Com. ACM*, 38(3): 42-48, 1995.
- [Furnas et al, 1987] Furnas,G.W., Landauer,T.K., Gomez, L.M., and Dumais, S.T., The Vocabulary

Problem in Human-System Communication, *Com. ACM*, 30(11): 964-971, 1987.

[Kashino et al, 1995] Kashino,K., Nakadai,K., Kinoshita,T. and Tanaka,H., Organization of Hierarchical Perceptual Sounds: Music Scene Analysis with Autonomous Processing Modules and a Quantitative Information Integration Mechanism, *Proc.IJCAI'95*, 158-164, 1995.

[Krulwich, 1995] Krulwich,B., Learning User Interests Across Heterogeneous Document Databases, *AAAI Spring Sympo. Series on Information Gathering from Distributed, Heterogeneous Environments*, 1995.

[Ng and Mooney, 1990] Ng,H., Mooney,R., On the Role of Coherence in Abductive Explanation, *Proc. AAAI'90*, 337-342, 1990.

[Ohsawa and Ishizuka 1997] Ohsawa,Y., Ishizuka,M., Networked Bubble Propagation: a Polynomial-time Hypothetical Reasoning Method for Computing Near-optimal Solutions, *Artif. Intell.* July 1997.

[Ohsawa and Yachida 1997] Ohsawa,Y., Yachida,M., Understanding Sequential Events by Cost-based Cooperation of Multiple Abducers, *J. Japanese Soc.AI*, 12(4), July 1997 (in Japanese).

[Poole et al, 1987] Poole,D., Aleliunas,R. and Goebel,R., Theorist : A Logical Reasoning System for Defaults and Diagnosis, in *the Knowledge Frontier : Essays in the Knowledge Representation*, Springer-Verlag, 1987.

[Poole, 1993] Poole,D., Probabilistic Horn Abduction and Bayesian Networks, *Artif. Intell.* 64(1): 81-130, 1993.

¹ Papers referred here from *AAAI Spring Sympo. Series on Information Gathering from Distributed, Heterogeneous Environments* are available from <http://www.isi.edu/sims/knoblock/sss95/proceedings.html>