THE IEEE Computational Intelligence BULLETIN

IEEE Computer Society

IEEE Computer Society
Technical Committee
on Computational Intelligence

December 2002 Vol. 1 No. 1

Communications

Profile

Research Outlook

Feature Articles

Book Review

Announcements

IEEE Computer Society Technical Committee on Computational Intelligence (TCCI)

The Technical Committee on Computational Intelligence (TCCI) of the IEEE Computer Society deals with tools and systems using biologically and linguistically motivated computational paradigms such as artificial neural networks, fuzzy logic, evolutionary optimization, rough sets, data mining, Web intelligence, intelligent agent technology, parallel and distributed information processing, and virtual reality.

If you are a member of the IEEE Computer Society, you may join the TCCI without cost. Just fill out the form at http://computer.org/tcsignup/.

The IEEE Computational Intelligence Bulletin

Aims and Scope

The IEEE Computational Intelligence Bulletin is the official publication of the Technical Committee on Computational Intelligence (TCCI) of the IEEE Computer Society, which is published twice a year in both hardcopies and electronic copies. The contents of the Bulletin include (but may not be limited to):

1. Letters and Communications of the TCCI Executive Committee
2. Feature Articles
3. R & D Profiles (R & D organizations, interview profiles on individuals, and projects etc.)
4. Book Reviews
5. News, Reports, and Announcements (TCCI sponsored or important/related activities)

Materials suitable for publication at the IEEE Computational Intelligence Bulletin should be sent directly to the Associate Editors of respective sections.

Technical or survey articles are subject to peer reviews, and their scope may include the theories, methods, tools, techniques, systems, and experiences for/in developing and applying biologically and linguistically motivated computational paradigms, such as artificial neural networks, fuzzy logic, evolutionary optimization, rough sets, and self-organization in the research and application domains, such as data mining, Web intelligence, intelligent agent technology, parallel and distributed information processing, and virtual reality.

Message from the TCCI Chair

It is my pleasure to welcome you to the first issue of the IEEE Computational Intelligence Bulletin, the official publication of the IEEE Computer Society Technical Committee on Computational Intelligence (TCCI).

The TCCI is a new technical committee of the IEEE Computer Society. I was elected to chair the Task Force on Virtual Intelligence (TFVI) in April 2002. On May 8, 2002, at an IEEE Computer Society Technical Activities Board (TAB) meeting in Portland, Oregon, I presented a proposal to move the TFVI to a technical committee (TC) and change the TC name from Virtual Intelligence to Computational Intelligence (CI). The proposal was unanimously accepted by the TAB meeting, and formally approved by the IEEE Computer Society Board of Governors in July 2002.

Since the approval of the TCCI, I have formed an Executive Committee to manage TCCI activities with the following members.

- Benjamin W. Wah (University of Illinois, Urbana-Champaign, USA) is the past chair of the TFVI. He will provide advice for the Chair on TCCI awards and interactions with sister societies as well as on other important issues of the TCCI activities.
- Nick J. Cercone (Dalhousie University, Canada) looks after student affairs. He is looking at the possibilities of (1) providing student grants to help defray the cost of attending TCCI conferences, (2) encouraging student poster sessions with a supplement to the TCCI conference proceedings, (3) arranging for industry to meet students, and (4) promoting the interaction between TCCI conference invited speakers and graduate students.
- Gusz Eiben (Vrije Universiteit Amsterdam, The Netherlands) takes care of curriculum issues, including the coverage of CI degree programs at universities. Professor Eiben believes typical CI components would involve evolutionary computing, neurocomputing, fuzzy computing, machine learning and data mining, adaptive and intelligent agents, and self-organizing systems.
- Vipin Kumar (University of Minnesota, USA) deals with publication matters. Our publication activities will involve possible special issues on Computational Intelligence related topics in IEEE Transactions.
- Jiming Liu (Hong Kong Baptist University, China) has been appointed as the Editor for the IEEE Computational Intelligence Bulletin. He has set up an editorial board and a plan for the publication of this Bulletin. The Bulletin will cover news and announcements of the TCCI activities, feature articles, book reviews and other Computational Intelligence relevant items.
- Ning Zhong (Maebashi Institute of Technology, Japan) is the Vice Chair for conferences and membership activities. The TCCI currently sponsors the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), and co-sponsors the IEEE International Conference on Data Mining (ICDM), the IEEE/WIC International Conference on Web Intelligence (WI), and the IEEE/WIC International Conference on Intelligence Agent Technology (IAT).

We have a standing invitation to join the TCCI for faculty, students, researchers and application developers from different Computational Intelligence related areas, such as artificial neural networks, fuzzy logic, evolutionary optimization, rough sets, data mining, Web intelligence, intelligent agent technology, parallel and distributed information processing, and virtual reality. You can visit the IEEE Computer Society's website (http://computer.org/tcsignup/) and fill out the TC membership form there.

I believe Computational Intelligence can play a very important role in the IEEE Computer Society's activities, and the Executive Committee will do our best to promote more activities sponsored by the TCCI and advance the TCCI's role both within the IEEE Computer Society and in collaboration with related sister societies. The TCCI has been planning on a number of exciting activities, as indicated in the above Executive Committee members' roles, and the participation and contributions from our TCCI members are essential. If you have any suggestions on any of the above activities or on other possible and worthwhile activities, please let me know.

I hope you will enjoy reading this Bulletin. If you need any more information about TCCI, please visit our website at http://www.cs.uvm.edu/∼xwu/tcci/index.shtml

Xindong Wu
(University of Vermont, USA)
Chair, TCCI -
IEEE Computer Society Technical Committee on Computational Intelligence

Message from the Editor-in-Chief

The 21st century will continue to be computation-centric; social and economic development will be benchmarked by how intelligently the power of computing can be utilized and extended. As the official publication of the IEEE Computer Society Technical Committee on Computational Intelligence, the IEEE Computational Intelligence Bulletin will provide a new gateway for researchers and practitioners to have direct access to the latest information on advanced research, industrial development, and professional activities in the major technical areas of Computational Intelligence, to share ideas and experiences, and to stimulate dialogues on emerging or challenging issues.

This first issue of the Bulletin presents an excellent coverage of some of the most challenging problems for today's as well as future Computational Intelligence community. The issue contains two feature articles: one on Web-log mining for request prediction and another on context-based problem solving in intelligent agents, an article on the coming Bison project in self-organizing P2P systems, and a R&D profile report on a newly established Cork Constraint Computation Centre. Last but not the least, it provides a book review of *Blondie24, Playing at the Edge of AI* and a highlight of several major upcoming events.

This debut of the first issue of the IEEE Computational Intelligence Bulletin shows the result of months of collaborative hard work by all the members of the Editorial Board. I would like to express my gratitude to their timely and professional efforts in working closely with contributing authors on the contents and presentations.

I hope that you will indeed enjoy reading the IEEE Computational Intelligence Bulletin, and find it informative, insightful, and inspiring.

Jiming Liu
(Hong Kong Baptist University, Hong Kong)
Editor-in-Chief
IEEE Computational Intelligence Bulletin

Cork Constraint Computation Centre

Making Hard Choices Easier

www.4c.ucc.ie

What Is Constraint Computation?

Difficult problems can offer too many choices, many of which are incompatible, few of which are optimal. The Cork Constraint Computation Centre (4C) develops the basic science that will make it easier for computers to help us make these choices.

Some examples of constraints:

- The meeting must start at 6:30.
- The separation between the soldermasks and nets should be at least 0.15mm.
- This model only comes in blue and green.
- This cable will not handle that much traffic.
- These sequences should align optimally.
- John prefers not to work on weekends.
- The demand will probably be for more than 5 thousand units in August.

Some examples of constraint satisfaction or optimization problems:

- Schedule these employees to cover all the shifts.
- Optimize the productivity of this manufacturing process.
- Configure this product to meet my needs.
- Find any violations of these design criteria.
- Optimize the use of this satellite camera.
- Align these amino acid sequences.

Constraints arise in design and configuration, planning and scheduling, diagnosis and testing, and in many other contexts. Constraint programming can solve problems in telecommunications, internet commerce, electronics, bioinformatics, transportation, network management, supply chain management, and many other fields.

Constraint computation has seen fundamental scientific advances, e.g. in understanding the relationship between problem structure and problem complexity. Constraint technology has demonstrated its commercial value.

Research Agenda

We apply advances in artificial intelligence and other disciplines to make constraint programming more powerful, more practical and easier to use. The work is centred in the field of artificial intelligence, but embedded in the broader constraint programming community. The centre conducts basic research in areas vital to the next generation of constraint technology.

Specifically, the Centre seeks advances in:

Automation:

The process of modeling domain knowledge, tailoring heuristics, and exploring alternatives must become more automated. Specific topics include acquisition, validation, optimisation, learning, and explanation. Progress can be made here by abstracting our experience with specific applications.

Applications:

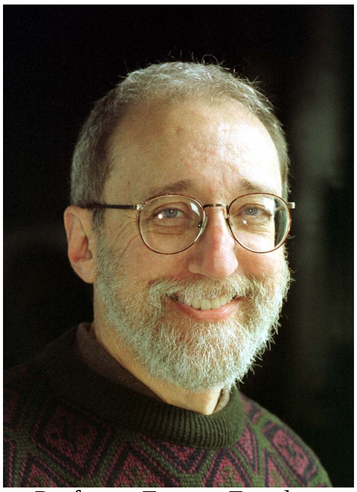
Applications will motivate and validate advances in automation. Application domains may include bioinformatics, configuration, computer and telecommunications networks, design, electronic commerce, planning and scheduling.

Adaptability:

In the real world we are confronted with uncertainty and change, with probabilities and preferences, with failures and tradeoffs, with collaborators and competitors. New applications and new contexts, e.g. interactive internet applications, present new challenges. Constraint technology must be further enriched to cope better with these challenges.

History

The Cork Constraint Computation Centre (4C) was established in October of 2001 when Professor Eugene C. Freuder received a Science Foundation Ireland Principal Investigator Award and moved his research lab from the University of New Hampshire in the U.S. to merge with the Constraint Processing Group in the University College Cork (UCC) Department of Computer Science, headed by Professor James Bowen. Professor Freuder became the Director of 4C and Professor Bowen the Co-Scientific Director.



Professor Eugene Freuder

Professor Jim Bowen

In 2002 Dr. Toby Walsh received a grant from Science Foundation Ireland and became the Deputy Director of 4C. 4C is attracting additional funding from government and industry, and will soon have three dozen academic, research, administrative and student staff members.

Professor Toby Walsh

Human Capital

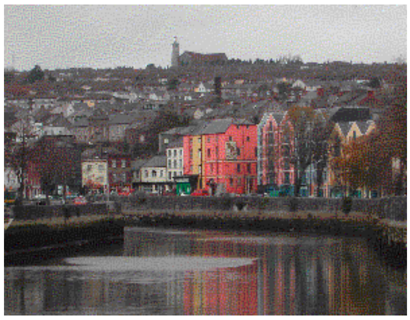
4C started with five students in the fall of 2001 and expects to have seventeen in the fall of 2002. Eight of these will be supported by Enterprise Ireland grants and one by the Irish Embark Initiative. Undergraduates also began gaining experience at 4C starting in the summer of 2002.

4C staff were involved in organizing programs for young scientists in the European artificial intelligence community and the international constraint programming community. Gene Freuder was invited to speak to both groups of young scientists. Chris Beck and Toby Walsh each gave tutorials at the American Association for Artificial Intelligence Conference in Edmonton, Canada in 2002.

Technology Transfer

4C receives support from Xerox Corporation and cadcoevolution. 4C has already visited or been visited by Bouygues, ILOG, LPA, RaidTec, Synopsis and Kinematik. Jim Bowen and Barry O'Sullivan were two of the founders of Suntas Technologies. Gene Freuder is on the Technical Advisory Boards of ILOG and Celcorp and Senior Technical Advisor of Ecora. Dr. Jim Little, who has extensive industry experience, is the 4C External Liason Officer.

Members of 4C have worked with many companies in the past including Oracle, Concentra, Nokia, Trilogy, ILOG, Lucent, Calico, Candle, Cabletron/Aprisma, Xerox, Ecora, British Telecommunications, Digital Equipment Corporation, Imperial Chemical Industries, IBM, Westinghouse, Zuken-Redac, Celestica, International Computers Ltd., Frequentis, Baan, and Grundfos.



International Outreach

4C is bringing international meetings to Ireland. In 2002 4C hosted a joint workshop of the European Research Consortium for Informatics and Mathematics Working Group on Constraints and the European Network of Excellence in Computational Logic area on Constraint and Logic Programming. Barry O'Sullivan, Toby Walsh, and Gene Freuder were among the Organizers. In 2003 the International Conference on Principles and Practice of Constraint Programming will come to Ireland, with Jim Bowen as Conference Chair, in 2004 the International Joint Conference on Automated Reasoning with Toby Walsh as Conference Chair.

In its first year: 4C members presented papers and helped organize workshops at scientific meetings in France, Cyprus, Canada, Italy, and the United States. Gene Freuder and Toby Walsh were Invited Speakers at scientific meetings in France, Cyprus, Austria, and Italy. 4C hosted visitors from England, France, Germany, Italy, Morocco, Sweden, Turkey, and the United States. Research collaborations were pursued with scientists in the U.S., U.K., Sweden, Germany, France, Spain, and Italy.

Contact Information

Director:
Professor Gene Freuder
E-mail: e.freuder@4c.ucc.ie

External Liaison Officer:
Dr. James Little
Phone: +353-(0)21-4255410
E-mail: j.little@4c.ucc.ie
Post: 4C, UCC, Cork, Ireland

Website: www.4c.ucc.ie

The BISON Project

Alberto Montresor * Ozälp Babaoglŭ *

Introduction

Modern distributed information systems are gaining an increasing importance in our every day lives. As access to networked applications become omnipresent through PC's, hand-held and wireless devices, more and more economical, social and cultural transactions are becoming dependent on the reliability and availability of distributed applications.

As a consequence of the increasing demands placed by users upon networked environments, the complexity of modern distributed systems has reached a level that puts them beyond our ability to deploy, manage and keep functioning correctly through traditional techniques. Part of the problem is due to the sheer size that these systems may reach, with millions of users and interconnected devices. The other aspect of the problem is due to the extremely complex interactions that may result among components even when their numbers are modest. Our current understanding of these systems is such that minor perturbations (e.g., a software upgrade, a failure) in some remote corner of the system will often have unforeseen, and at times catastrophic, global repercussions. In addition to being fragile, many situations (e.g., adding/removing components) arising from their highly dynamic environments require manual intervention to keep information systems functioning.

In order to deal with the scale and dynamism that characterize modern distributed systems, we believe that a paradigm shift is required that includes self-organization, adaptation and resilience as intrinsic properties rather than as afterthought. For this reason, we have started BISON (*Biology-Inspired techniques for Self Organization in dynamic Networks*), an international project partially funded by the European Commission. BISON draws inspiration from natural and biological processes to devise appropriate techniques and tools to achieve the proposed paradigm shift and to enable the construction of robust and self-organizing distributed systems for deployment in highly dynamic modern network environments.

*Department of Computer Science, University of Bologna, Mura Anteo Zamboni 7, 40127 Bologna (Italy), Email: {babaoglu,montresor}@CS.UniBO.IT

Overview of the BISON Project

Nature and biology have been a rich source of inspiration for computer scientists. *Genetic algorithms* [4], *neural networks* [7] and *simulated annealing* [8] are examples where different natural processes have been mimicked algorithmically to successfully solve important practical problems. More recently, *social insects* [2] have been added to this list, inspiring biomimetic algorithms to solve combinatorial optimization problems. *Immune networks* [9] represent a recent frontier in this area.

These processes are examples of *complex adaptive systems (CAS)* [6] that arise in a variety of biological, social and economical phenomena. In the CAS framework, a system consists of large numbers of *autonomous agents* that individually have very simple behavior and that interact with each other in very simple ways. CAS are characterized by total lack of centralized coordination. Despite the simplicity of their components, CAS typically exhibit what is called *emergent behavior* that is surprisingly complex [6]. Furthermore, the collective behavior of a well-tuned CAS is highly adaptive to changing environmental conditions or unforeseen scenarios, is robust to deviant behavior (failures) and is self-organizing towards desirable configurations.

Parallels between CAS and advanced information systems are immediate. BISON will exploit this fact to explore the possibility of using ideas and techniques derived from CAS to enable the construction of robust, self-organizing and self-repairing information systems as ensembles of autonomous agents that mimic the behavior of some natural or biological process. In our opinion, the application of CAS will enable developers to meet the challenges arising in dynamic network settings and to obtain global properties like resilience, scalability and adaptability, without explicitly programming them into the individual agents. This represents a radical shift from traditional algorithmic techniques to that of obtaining the desired system properties as a result of emergent behavior that often involves evolution, adaptation, or learning.

The dynamic network architectures that will be explicitly dealt with in the project are *peer-to-peer* (P2P) and *grid* computing systems [11, 3], as well as *ad-hoc networks* (AHN) [5]. P2P systems are distributed systems based on the concept of resource sharing by direct exchange between *peer* nodes, in the sense that all nodes in the system have equal role and responsibility [11]. Exchanged resources include content, as in popular P2P document sharing applications, and CPU cycles or storage capacity, as in computational and storage grid systems. P2P systems exclude any form of centralized structure, requiring control to be completely decentralized. In AHN, heterogeneous populations of mobile, wireless devices cooperate on specific tasks, exchanging information or simply interacting informally to relay information between themselves and the fixed network [12]. Communication in AHN is based on multi-hop routing among mobile nodes. Multi-hop routing offers numerous benefits: it extends the range of a base station; it allows power saving; and it allows wireless communication, without the use of base stations, between users located within a limited distance of one another.

We have chosen these domains for their practical importance in future distributed computing technologies as well as their potential for benefiting from our results. P2P/Grid systems can be seen as dynamic networking at the application level, while AHN results from dynamic networking at the system level. In both cases, the topology of the system typically changes rapidly due to nodes voluntarily joining or leaving the network, due to involuntary events such as crashes and network partitions, or due to frequently changing interconnection patterns.

The use of CAS techniques derived from nature in the context of information systems is not new. Numerous studies have abstracted principles from biological systems and applied them to network-related problems, primarily routing. However, much of the current work in this area can be characterized as *harvesting* — combing through nature, looking for a biological system or process that appears to have some interesting properties, and applying it to a technological problem by modifying and adapting it through an enlightened trial-and-error process. The result is a CAS that has been empirically obtained and that appears to solve a technological problem, but without any scientific explanation of why.

BISON proposes to take exploitation of CAS for solving technological problems beyond the harvesting phase. We will study a small number of biology-inspired CAS, applied to the technological niche of dynamic networks, with the aim of elucidating principles or regularities in their behavior. In other words, BISON seeks to develop a rigorous understanding of why a given CAS does or does not perform well for a given technological problem. A systematic study of the rules governing good performance of CAS offers a bottom-up opportunity to build more general understanding of the rules for CAS behavior. The ultimate goal of the BISON project is then the ability to synthesize a CAS that will perform well in solving a given technological task based on the accumulated understanding of its regularities when applied to different tasks. In addition to this ambitious overall objective, BISON has more concrete objectives to obtain robust, self-organizing and self-repairing solutions to important problems that arise in dynamic networks at both the system layer and the application layer. Here we outline these objectives.

Status of the Project and Participants

The BISON project will officially begin on January, 2003 and will last for three years. Given the interdisciplinary nature of the problem ahead, the BISON consortium brings together experts from a wide range of areas including core disciplines (physics, mathematics, biology) and "user" disciplines (information systems, telecommunication industry). The Department of Computer Science of the University of Bologna (Italy) brings its experience in the development of fault-tolerant distributed systems and, more recently, in the development of P2P systems Telenor (Norway's leading telecommunication company) conducts research towards nature-inspired techniques for network monitoring and surveillance, and for traffic control and network reconfiguration. The Department of Methods of Innovative Computing, University of Dresden (Germany) is an interdisciplinary institution dedicated to the modeling of biological systems and to the development of algorithms for bioinformatics and operates at the interface of computer science, mathematics and biology. The Dalle Molle Institute for Artificial Intelligence (IDSIA, Switzerland) has a long experience in the definition of combinatorial optimization algorithms based on *swarm intelligence* and *ant colony optimization*. The Santa Fe Institute (Santa Fe, USA) is considered a top-level center for studies in complexity sciences, and its participation will be uniquely valuable in achieving

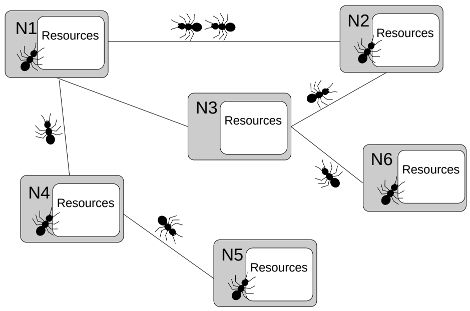


Figure 1. An Anthill System

the objectives of BISON.

Anthill

Despite the fact that the Bison project has not yet officially started, in order to pursue these ideas we have developed *Anthill*[1], a novel framework for P2P application development based on ideas such as multi-agent systems and evolutionary programming borrowed from CAS [4]. The goals of Anthill are to provide an environment that simplifies the design and deployment of P2P systems based on these paradigms, and to provide a "testbed" for studying and experimenting with CAS-based P2P systems in order to understand their properties and evaluate their performance. An Anthill system is composed of a collection of interconnected *nests*. Each nest is a peer entity that makes its storage and computational resources available to swarms of *ants* – autonomous agents that travel across the network trying to satisfy user requests. During their life, ants interact with services provided by visited nests, such as storage management and ant scheduling.

Having developed a first Anthill prototype, we are now in the process of testing the viability of our ideas regarding P2P as CAS by developing common P2P applications over it [1, 10]. For example, Messor is a load-balancing application for grid computing [10] aimed at supporting the concurrent execution of highly-parallel, time-intensive computations, in which the workload may be decomposed into a large number of independent jobs. The computational power offered by a network of Anthill nests is exploited by Messor by assigning a set of jobs comprising a computation to a dispersed set of nests. To determine how to balance the load among the computing nodes, Messor uses an algorithm inspired by the following observations. Several species of ants are known to group objects (e.g., dead corpses) in their environment into piles so as to clean up their nests. Observing this behavior, one could be mislead into thinking that the cleanup operation is being coordinated by some "leader" ant. It is possible to describe an artificial ant exhibiting this very same behavior in a simulated environment, following three simple rules: (i) wanders around randomly, until it encounters an object; (ii) if it was carrying an object, it drops the object and continues to wander randomly; (iii) if it was not carrying an object, it picks up the object and continues to wander. Despite their simplicity, a colony of these "unintelligent" ants is able to group objects into large clusters, independent of their initial distribution in the environment.

It is possible to consider a simple variant (the inverse) of the above artificial ant that drops an object that it may be carrying only after having wandered about randomly "for a while" without encountering other objects. Colonies of such ants try to disperse objects uniformly over their environment rather than clustering them into piles. As such, they could form the basis for a distributed load balancing algorithm.

Figure 2 illustrate how the load balancing process performed by Messor evolves over time. The results were obtained in a network of 100 idle nests, initially connected to form a ring (for visualization reasons). Initially, 10,000 jobs are generated in a single node. The different histograms depict the load observed in all the nests (x-axis) after 0, 5, 10, 15, 20, and 50 iterations of the algorithm. At each iteration, a set of 20 ants perform a single step, i.e. they move from one node to another, possibly moving jobs from overloaded nests to underloaded nests. As the figure illustrates, only 15-20 iterations are required to transfer jobs to all other nodes in the network, and after 50 iterations, the load is perfectly balanced. The first iterations are spent exploring the neighborhood in the network. After a few iterations, new connections are created and used to transfer jobs to remote parts of the network.

References

Additional information can be found at the project web site, `http://www.cs.unibo.it/bison`. Information about the Anthill toolkit can be found at `http://www.cs.unibo.it/anthill`.

References

[1] Ö. Babaoğlu, H. Meling, and A. Montresor. Anthill: A Framework for the Development of Agent-Based Peer-to-Peer Systems. In *Proceedings of the 22th International Conference on Distributed Computing Systems (ICDCS)*, Vienna, Austria, July 2002.

[2] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999.

[3] I. Foster and C. Kesselman, editors. *The Grid: Blueprint for a Future Computing Infrastructure,*. Morgan Kaufmann, 1999.

[4] D. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[5] Z. Haas and M. Pearlman. The zone routing protocol (ZRP) for ad-hoc networks (InternetDraft). Mobile Ad-hoc Network (MANET) Working Group, Aug. 1998.

[6] J. Holland. *Emergence: from Chaos to Order*. Oxford University Press, 1998.

[7] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the Natl. Acad. Sci. USA*, 79:2554–2558, 1982.

[8] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[9] E. Klarreich. Inspired by Immunity. *Nature*, 415:468–470, Jan. 2002.

[10] A. Montresor, H. Meling, and O. Babaoğlu. Messor: Load-Balancing through a Swarm of Autonomous Agents. In *Proceedings of the 1st Workshop on Agent and Peer-to-Peer Systems*, Bologna, Italy, July 2002.

[11] A. Oram, editor. *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*. O'Reilly & Associates, Mar. 2001.

[12] C. Toh. *Ad Hoc Mobile Wireless Networks: Protocols and Systems*. Prentice Hall, 2002.

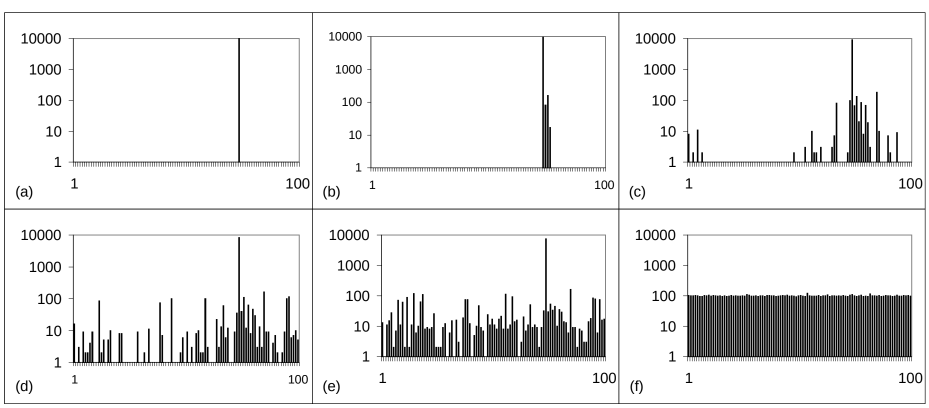


Figure 2. Load distribution after 0, 5, 10, 15, 20, 50 iterations.

Web-log Mining for Quantitative Temporal-Event Prediction

Qiang Yang[1], Hui Wang[1] and Wei Zhang[2]

Abstract— **The web log data embed much of web users' browsing behavior. From the web logs, one can discover patterns that predict the users' future requests based on their current behavior. These web data are very complex due to their large size and sequential nature. In the past, researchers have proposed different methods to predict what pages will be visited next based on their present visit patterns. In this paper, we extend this work to discover patterns that can predict when these web page accesses will occur. Our method is based on a novel extension of association rule classification method. We extend the traditional association rules by including the temporal information explicitly in each rule, and reason about the confidence of each prediction in terms of its temporal region. We compare two different methods for temporal event prediction, demonstrate the effectiveness of our methods empirically on realistic web logs, and explore the tradeoff between prediction accuracy and data mining time for our models.**

Index Terms— **Web Log Mining, Quantitative Predictions for Web Accesses.**

I. Introduction

THE rapid expansion of the World Wide Web has created an unprecedented opportunity to disseminate and gather information online. As more data are becoming available, there is much need to study web-user behaviors to better serve the users and increase the value of enterprises. One important data source for this study is the web-log data that traces the user's web browsing. In this paper, we study prediction models that predict the user's next requests as well as when the requests are likely to happen, based on the web-log data. The result of accurate prediction can be used for recommending products to the customer, suggesting useful links, presending, pre-fetching and caching web pages for reducing access latency [11], [18].

An important class of data mining problems is mining sequential association rules from web log data. The web-log data consists of sequences of URLs requested by different clients bearing different IP addresses. Association rules can be used to decide the next likely web page requests based on significant statistical correlations. In the past, sequential association rules [3], [2] have been used to capture the co-occurrence of buying different items in a supermarket shopping. Episodes were designed to capture significant patterns from sequences of events [8]. However, these models were not designed for the prediction task, because they do not specify how to select among multiple predictions for a given observation. The works by [6], [17] considered using association rules for prediction by selecting rules based on confidence measures, but they did not consider the classifiers for sequential data. In the network system area, n-gram or path based rules have been proposed for capturing long paths that occur frequently [12], [16], but the researchers in these areas did not study the models in the context of association rules, and offered no comparison with other potential prediction models in a systematic way. As a result, it remains an open question how to construct association rules that predict not only what is likely to happen next given the current observed events, but when these events will occur.

In this paper, we present a quantitative model for temporal event prediction on the web. By quantitative we mean the ability to predict a time interval in which the next web page visits will occur. For example, after observing a web user visiting pages A and B in a row, our system might predict that page C is most likely to be the next page to be visited by the user, and that C is most likely to be visited within 10 to 20 seconds from the current time. Our approach is to extend the traditional association rules by including additional constraints and representations. In web-access prediction, previously used representations often state that if access to pages A, B and C are observed, then D will be predicted to occur next. We extend this representation by including, on the right hand side of each rule, a probable temporal region [t1, t2] in which D will happen, and a confidence estimate on when D will occur. In addition, we place the restriction that the left-hand-side of the rule A, B and C must occur next to each other and in that order; in essence, A, B, C is a substring. The right hand side to be predicted corresponds to a web page access that occurs frequently enough and falls into the specified window with high probability.

The main contribution of the work is the time-accuracy tradeoff between two different methods for web log mining. The first method is based on a minimal-temporal-region heuristic. This method has been studied in AI in the past [19], [20], where an effective solution has been proposed for assembly-line event sequences. Here we generalize this method to web logs, taking into account the special properties in the web logs. Our generalization allows the left-hand-sides of rules to be greater than one, which enables more accurate predictions. Our second method is aimed at achieving efficiency while sacrificing accuracy slightly; it is based on the computation of confidence intervals of normal distributions

1. Department of Computer Science, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong, China,(qyang,whui)@cs.ust.hk
2. The Boeing Company, Seattle, WA USA, Wei.zhang@boeing.com

in temporal events. The minimal-temporal-region method is shown to be more accurate but takes longer in the learning phase. On the other hand, the confidence-interval-based method is more efficient to mine, but is less accurate. Our study offers the web-log mining system designers a choice in algorithms according to the needs in their application domains.

The paper is organized as follows. Section II discusses rule-representation methods. Section III discusses region representation methods. Section IV presents the experimental results. Section V discusses related work. Section VI concludes the paper with a discussion of future work.

II. Rule Representation and Selection

A. Web Logs and User Sessions

Consider the Web log data from a NASA Web server shown in Table I. Typically, these web server logs contain millions of records, where each record refers to a visit by a user to a certain web page served by a web server. This data set contains one month worth of all HTTP requests to the NASA Kennedy Space Center WWW server in Florida. The log was collected from 00:00:00 August 1, 1995 through 23:59:59 August 31, 1995. In this period there were 1,569,898 requests. There are a total of 72,151 unique IP addresses, forming a total of 119,838 sessions. A total of 2,926 unique pages were requested.

TABLE I
Example Web Log

kgtyk4.kj.yamagata-u.ac.jp - - [01/Aug/1995:00:00:17 -0400] ”GET / HTTP/1.0” 200 7280
kgtyk4.kj.yamagata-u.ac.jp - - [01/Aug/1995:00:00:18 -0400] ”GET /images/ksclogo-medium.gif HTTP/1.0” 200 5866
d0ucr6.fnal.gov - - [01/Aug/1995:00:00:19 -0400] ”GET /history/apollo/apollo-16/ Apollo-16.html HTTP/1.0” 200

Given a web log, the first step is to clean the raw data. We filter out documents that are not requested directly by users. These are image requests or CSS requests in the log that are retrieved automatically after accessing requests to a document page containing links to these files and some half-baked requests. Their existence will not help us to do the comparison among all the different methods.

We consider web log data as a sequence of distinct web pages, where subsequences, such as user sessions can be observed by unusually long gaps between consecutive requests. For example, assume that the web log consists of the following user visit sequence: (A (by user 1), B (by user 2), C (by user 2), D (by user 3), E (by user 1)) (we use ”()” to denote a sequence of web accesses in this paper). This sequence can be divided into user sessions according to IP address: Session 1 (by user 1): (A, E); Session 2 (by user 2): (B, C); Session 3 (by user 3): (D), where each user session corresponds to a user IP address.

In deciding on the boundary of the sessions, we studied the time interval distribution of successive accesses by all users, and used a heuristic splitting method for a new session. We will present this method in detail in Section V.

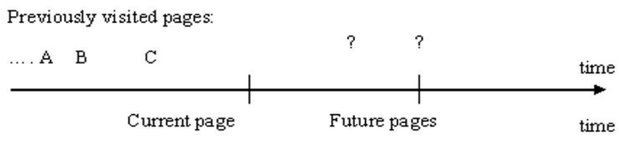


Fig. 1. Moving Window Illustration

To capture the sequential and time-limited nature of prediction, we define two windows. The first one is called *antecedent window*, which holds all visited pages within a given number of user requests and up to a current instant in time. A second window, called the *consequent window*, holds all future visited pages within a number of user requests from the current time instant. In subsequent discussions, we will refer to the antecedent window as W_1, and the consequent window as W_2. Intuitively, a certain pattern of web pages already occurring in an antecedent window could be used to determine which documents are going to occur in the consequent window. Fig 1 shows an example of a moving window.

The moving windows define a table in which data mining can occur. Each row of the table corresponds to the URL's captured by each pair of moving windows. The number of columns in the table corresponds to the sizes of the moving windows. Table II shows an example of such a table corresponding to the sequence (A, B, C, A, C, D, G), where the size of W_1 is three and the size of W_2 is one. In this table, under W_1, A1, A2 and A3 denote the locations of the last three objects requested in the antecedent window, and ”Prediction” and ”Time Interval” are the objects and predicted time interval in the consequent window.

TABLE II
A portion of the Log Table extracted by a moving window pair of size [3, 1]

W1			W2	
A1	**A2**	**A3**	**Prediction**	**Time Interval**
A	B	C	A	[10 min, 20 min]
B	C	A	C	[5 min, 15 min]
C	A	C	D	[3 min, 4 min]

B. Prediction Rule Representation

We now discuss how to extract rules of the form $LHS \rightarrow RHS$ from the session table. Our different methods will extract rules based on different criteria for selecting the LHS. However, we restrict the RHS in the following way. Let U1, U2, ... Un be the candidate URL's for the RHS that can be predicted based on the same LHS. We build a rule $LHS \rightarrow \langle Uk, [t1, t2] \rangle (supp, conf)$ where the URL Uk occurs most frequently in the rows of the table among all Ui's in the set U1, U2, ... Un. [t1, t2] is the region determined by the data mining algorithm in which the event Uk is most likely to occur, and *supp* and *conf* are the support and confidence for such occurrence, respectively, in Equations 2 and 2 below.

The rule representation we use is known as the *latest-substring rules*. These rules not only take into account the order and adjacency information, but also the *recency* information about the LHS string. In this representation, only the substrings ending in the current time (which corresponds to the end of the window W_1) qualifies to be the LHS of a rule. These are also known as hybrid n-gram rules in some literature [12], [16]. For example, Table III shows the latest-substring rules example.

TABLE III
LATEST-SUBSTRING RULES. T1, T2 AND T3 ARE TIME INTERVALS

W1	W2	Latest substring Rules
A, B, C	D	⟨A, B, C⟩ → ⟨D, T1⟩, ⟨B, C⟩ → ⟨D, T2⟩, ⟨C¿ → ⟨D, T3⟩

Viewed from another angle, latest-substring rules could also be considered as the union of Nth-order Markov models [9], where N covers different orders up to the length of W_1. Therefore, it is more general than the N-gram models or Nth-order Markov models. However, through our other experiments, we have found out that the Markov models' performance drops when N exceeds a certain threshold, but the latest-substring method that considers multiple Nth-order models for different N demonstrates a monotonically increasing precision curve.

For any given set of rules, we also have an option to add a default rule that captures all cases where no rule in the rule set applies; when no LHS of all rules apply to a given observed sequence of URL's, the default rule always applies. For example, a default rule can simply be the most frequently requested page in the training web log.

For each rule of the form LHS→ RHS, we define the *support* and *confidence* as follows

$$labeleq2sup = \frac{count(LHS, RHS)}{count(Table)} \quad (1)$$

$$conf = \frac{\sup(LHS, RHS)}{\sup(LHS)} \quad (2)$$

In the equations above, the function *count*(*Table*) returns the number of rows in the log table, and *count*(*LHS*) returns the number of rows in the log table that W_1 is a certain LHS.

$$sup(LHS) = \frac{count(LHS)}{count(Table)} \quad (3)$$

C. Rule-Selection Methods

In classification, Liu et a. [6] extended association rules to build confidence-based classifiers. In this section, we extend their work further by including rankings on temporal intervals. Our goal is to output the best guess on a class based on a given observation. In different rule-representation methods, each observation (or case) where the LHS matches the case can give rise to more than one rule. Therefore, we need a way to select among all rules that apply. In a certain way, the rule-selection method compresses the rule set; if a rule is never applied, then it is removed from the rule set. The end result is that we will have a smaller rule set with higher quality. In addition to the extracted rules, we also define a default rule, whose RHS is the most popular page in the training web log and the LHS is the empty set. When no other rules apply, the default rule is automatically applied.

For a given set of rules and a given rule-selection method, the above rule set defines a classifier. With the classifier, we can make a prediction for any given case. For a test case that consists of a sequence of web page visits, the prediction for the next page visit is correct if the RHS of the selected rule occurs in window W_2. For N different test cases, let C be the number of correct predictions. Then the precision of the classifier is defined as

$$precision = \frac{C}{N} \quad (4)$$

Our rule selection method is called the most-confident selection method. It always chooses a rule with the highest confidence among all the applicable association rules. A tie is broken by choosing a rule with a longer LHS. For example, suppose that for a testing case and antecedent window of size four, an observed sequence is (A, B, C, D). Suppose that using the most-confident rule selection method, we can find three rules which can be applied to this example, including:

Rule 1: (A, B, C, D) → E, T1 with confidence 30%
Rule 2: (C, D) → F, T2 with confidence 60%
Rule 3: (D) → G, T3 with confidence 50%.

In this case, the confidence values of rule 1, rule 2 and rule 3 are 30%, 60% and 50%, respectively. Since Rule 2 has the highest confidence, the most-confident selection method will choose Rule 2, and predict F.

The rationale of most-confident selection is that the testing data will share the same characteristics as the training data that we built our classifier on. Thus, if a rule has higher confidence in the training data, then this rule will also show a higher precision in the testing data. As we will see, this assumption is not always correct, as it can lead to overfitting rules. However, this problem can be solved by introducing a filtering step, which removes all rules for which the support value is below a threshold. In our experiment, we used a support threshold of value 10.

Note that a rule may have different regions with different confidence values. Each region is associated with a different confidence value. In addition, we allow the RHS of a rule to predict more than one URL, in the decreasing value of the confidence strength. For example, we might have a rule whose LHS is "A, B, C" and whose RHS is {(D, [t1, t2], conf1, supp1), (F, [t3, t4], conf2, supp2)}. In this case, our prediction algorithm can predict up to n events that might occur in the future, including D, F, etc. This is known as the n-best method. For this method, if one of the predicted URL occurs in the corresponding range, then a hit is registered towards final precision calculation.

We performed several experiments to show the effects of n-best prediction. Figures 2 and 3 show the trend as n increases. The three dotted lines correspond to three region-selection methods, which we will explain in detail in Section III. It is clear from the figures that, when we set n to be two, the precision is already high enough. When n is greater than two, there is not much improvement. This result tells us that typically each association rule needs only consider the top two best predictions among all possible predictions.

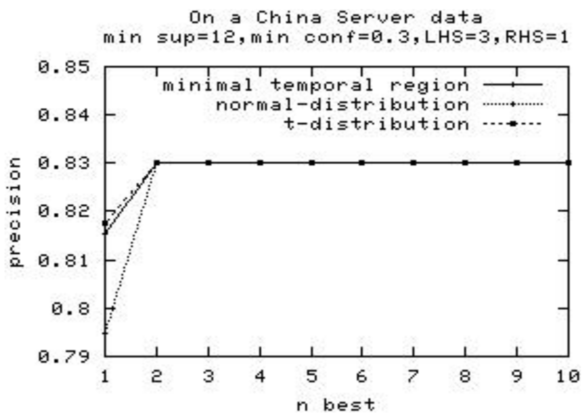


Fig. 2. Precision as the n increases on a China Web data

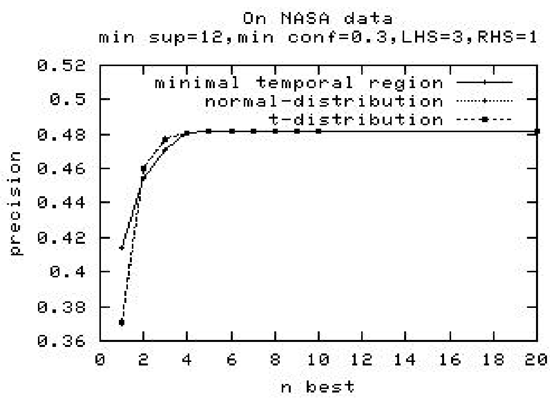


Fig. 3. Precision as the n increases on NASA data

III. Temporal Region Representation Methods

Now, we will describe how to choose a certain temporal region for rule construction. We have two families of region selection methods: a confidence interval based method and a minimal region selection method.

A. *Confidence Interval Method*

Consider a prediction: when (A, B, C) occurs, what is the next event that is likely to occur? Furthermore, if we decide that D is most likely to occur next, when will D occur? We are interested in computing a time interval [t_1, t_2] meaning that D is likely to occur in the future between t_1 and t_2 time scope. We also would like to place a high level of confidence on this interval prediction; for example, we might choose a 95% confidence. In order to make the prediction, we will collect all the association rules of the form: LHS→ ⟨RHS, [t_1, t_2]⟩ (supp, conf) from the training data. The task in this section is how to get [t_1, t_2] that is both accurate and narrow to be useful.

Our method is to compute the set of time lags in which D occurs after A, B and C occurs. This collection of time points is called the lag-set. For an example, a rule:

(A, B, C)→ ⟨D, [t_1, t_2] ⟩ (supp, conf)

has a lag-set {4, 9, 20, 22, 31, 39, 39, 39, 40, 41, 41, 42, 43, 45, 53, 61}

This means that when A, B, C are observed to occur next to each other, D occurred at 4 seconds after C, 9 seconds after C, and 20, 22, 31, 39 seconds after the occurrence of C. Note that we have three '39"s here, denoting that 'D' occurred three times at 39 seconds after C. *Supp* and *conf* are this rule's support and confidence information.

A naïve time interval for this rule is to choose [t_1, t_2] to be: [4, 61], corresponding to the first and last time points of the lag set. However, we could do much better. From the lag-set such as the one above, we can draw an occurrence density curve. If the lag-set is large enough, we may expect the curve to demonstrate the standard normal distribution as shown in Fig 4. Thus, we can use the normal distribution formulas to choose an interval [t_1, t_2].

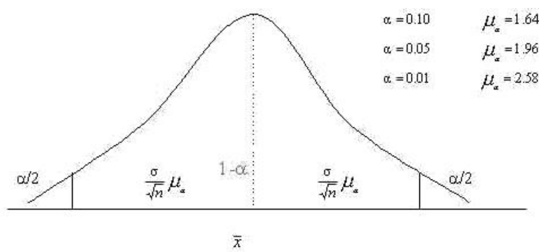


Fig. 4. Standard Normal Distribution

In classical statistics theory, for a large data set, we use formula (5) to measure the confidence interval.

$$[\bar{x} - \frac{\sigma}{\sqrt{n}}\mu_\partial, \bar{x} + \frac{\sigma}{\sqrt{n}}\mu_\partial] \quad (5)$$

where $\bar{x}$ is the mean, σ^2 is the standard deviation, μ_∂ follows the normal distribution table. n is the number of examples in the training data that supports this interval.

For the example above, if n=16, $\bar{x} = 35.56$, $\sigma = 15.6$, then for the confidence level 1-0.05=95% ($\partial = 0.05$), $\mu_\partial = 1.96$. Hence, the temporal region will be [28.21,42.91].

For a small data set, we use the t-distribution formulas instead, and changing σ^2 to s^2(s^2 is variance deviation), and μ_∂ to $t_\partial(n-1)$; the latter $(n-1)$ is a number in the t-distribution table equal to $t_\partial(n-1)$. The interval is chosen according to formula (6):

$$[\bar{x} - \frac{s}{\sqrt{n}} t_\partial, \bar{x} + \frac{s}{\sqrt{n}} t_\partial] \quad (6)$$

As an example, let the confidence level be 1-0.05 = 95%, $t_\partial(n-1) = t_\partial(15)$ = 1.753. Thus the region is [28.96, 42.16].

B. Minimal Temporal Region Selection

The confidence-interval-based method presented above chooses an interval based on the confidence region. However, it does not express our wish to find a temporal region that is as narrow as possible while covering as many training cases as possible. A minimal temporal region method was proposed in [19] and [20]. However, in these works it is required that each rule's LHS has a size of one. In this section, we extend this method to include association rules whose LHS can be greater than one.

A minimal temporal region is the smallest time interval that covers all the values in a subset of a lag set. Consider an example, suppose that a rule: (A, B, C) $\rightarrow$ D has a lag set {0,17,62,87). This will result in 10 temporal regions: [0,0], [0,17], [0,62], [0,87], [17,17], [17,62], [17,87], [62,87] and [87,87]. Our aim is to choose a temporal region from the above with the smallest scope and covers all occurrences. We use a heuristic in formula (7) to obtain a score for each of these regions:

$$Score = W_1 * Accp + W_2 * Rng + W_3 * Cov \quad (7)$$

Intuitively, this formula is trying to balance three factors: high accuracy, short range of the time interval and large coverage. The region with the highest score will be chosen. The definitions for Accp, Rng, Cov are as follows:

1. Prediction Accuracy (Accp): this factor computes the percentage of cases that a target event occurs in the time region over all cases that a condition event occurs.

2. Range (Rng): While Accp reward large regions (their values increase monotonously as the size of a temporal region grows), Rng is a factor encouraging smaller regions, is defined as 1-Intv(r)/(MaxLag-MinLag+1), where Intv(r) is the region size of rule r.

3.Coverage (Cov): This computes the rate of cases covered by a rule over all cases that are covered by the same condition-target pair but with the full search scope defined by MinLag and MaxLag. We denote the latter as AllCntScp. Then the Cov is AllCnt/AllCntScp.

The weights W_1, W_2 and W_3 express their relative importance. In general, they can be learned using linear regression method. We set them to one in our experiments.

IV. Experimental Results

In the last section, we presented two methods for temporal region computation. The first method is a confidence-based method, based on the assumption that the event distribution follows a normal distribution. The advantage of this method is that it requires one scan of the time points in the web log, resulting in linear time complexity in computation. However, the normal-distribution assumption is quite a strong one. The minimal temporal region heuristic, on the other hand, does not make this assumption. Instead, it looks for a good trade off point between the coverage, size and accuracy of the time points in the web log. The price to pay is that it involves more computation.

In this section, we will explore the relative merits of these two methods in detail. The tradeoff between accuracy and computation time studied here corresponds to the main contribution of this paper.

A. Experimental Setup

Our goal is to select a best rule representation with the region. We employ 3 realistic data sets: NASA, EPA and a new data of Web Server located at China. EPA log was collected from 23:53:25 on Tuesday, August 29 1995 to 23:53:07, August 30 1995, about 4.8M. After removing some irregular logs, we have 2225 unique visiting IP address, and 4149 unique pages are requested and 17933 requests. The NASA data is described in Section 2. We also used a more recent dataset from a penpal-service portal site located in Beijing. It was collected from 00:00:00 Jan 22, 2002 to 21:12:44 Jan 22, 2002, with a size of about 7.8M. After data cleaning, we have 270 unique IP address, 1000 unique web pages and 9688 requests. In this experiment, requests on the same CGI with different parameters are considered as different pages. For example: "/htbin/wais.pl?STS-59" and "/htbin/wais.pl?IMAX" are two pages in our system.

To obtain user sessions, we use a heuristic user-session splitting method. The heuristic is to calculate the mean of the gaps between two consecutive requests in the web logs. For each next page request, if the time gap is larger than a constant number of the ancestral mean time gap , we consider the request as starting a new session. For example, we use 70 as the constant factor in subsequent experiments.

In our experiment, we split all the sessions into training set and testing set by splitting the data into two equal parts and then construct the association rules from the training data. We restrict the LHS sizes to be no larger than three and RHS size to be one. Our filtering method removed all rules for which the confidence is less than a minimal confidence and the support is less than a minimal support. The minimum confidence and support values are used as variables in our tests in the following sections to test their effectiveness.

B. Comparison on Precision

Table IV provides a comparison of precision of the naive temporal-region selection based methods. When a default rule is used, where the default rule is defined as the most popular page in the web log, the default prediction is made whenever no rule whose LHS matches the current observation in the test data.

The 'Precision with default rule' is defined as:

$$Precision = \frac{C_{all}}{N_{test\_cases}} \quad (8)$$

The set of N_{test_cases} test cases is a section of the web log where each test case corresponds to a user access to a web page. C_{all} corresponds to the set of all correct guesses according to our prediction; this set is also known as all the correct "hits".

The 'Precision without default rule' is defined as:

$$Precision = \frac{C_{without\_default\_rule}}{N_{predicted\_cases}} \quad (9)$$

TABLE IV

PRECISION AS n IN N-BEST INCREASES FOR NASA DATA

Confidence-interval Based	n-best	Precision with default rule	Precision without default rule
	n=1	0.35411	0.380741
	n=2	0.42339	0.457589
Minimal temporal region selection		Precision with default rule	Precision without default rule
	n=1	0.38232	0.41987
	n=2	0.41630	0.45866

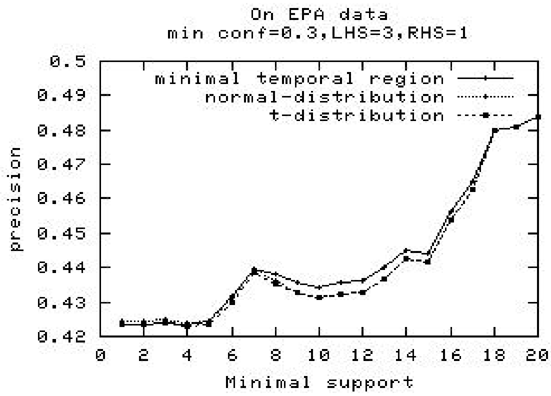
In this test, all data from the NASA log are used, with 50From these results, we conclude that both methods give similar accuracy results, with the minimal temporal-region selection giving slightly better performance when $n = 1$.

C. Comparing Temporal-Region Selection Methods

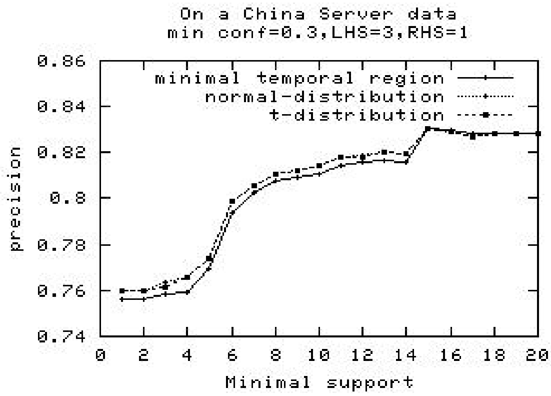
In this section, we will compare the performance of three region-selection methods. For brevity, we only consider prediction without the default rule.

Fig 5a-c show the prediction precision as the minimal support changes. In this test, minimal temporal region is better than the standard normal distribution and t-distribution, especially for large datasets (NASA data). This can be explained by the fact that the standard normal/t distribution methods prefer the largest regions around the mean of each event occurrence. However, the minimal temporal region also prefers small regions by using the factor Rng. Therefore the minimal temporal region method makes a balance between accuracy and narrow time regions. For the China web data, these methods have the similar performance.

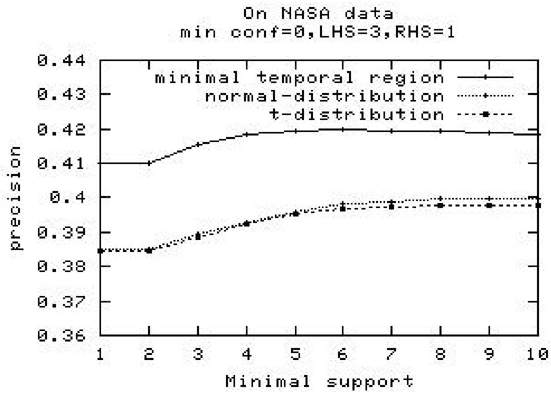
We now consider the performance in terms of accuracy as the minimal confidence increases. The results are shown in Fig 6a-c. The minimal temporal region method is a little better than the other two for NASA data, but is comparable for the other two datasets. The general trend is that the precision will



a. Precision as the Minimal Support Increases on EPA data



b. Precision as the Minimal Support Increases on A China Web data



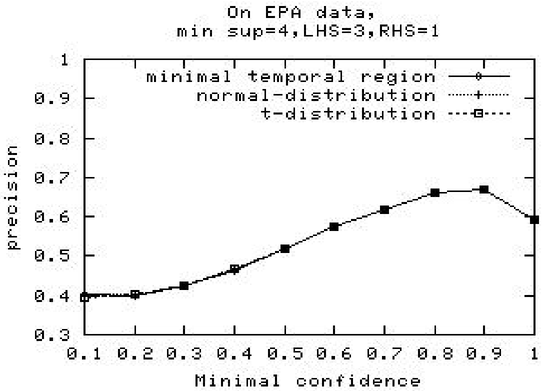
c. Precision as the Minimal Support Increases on NASA data

Fig. 5. Precision results on three data sets

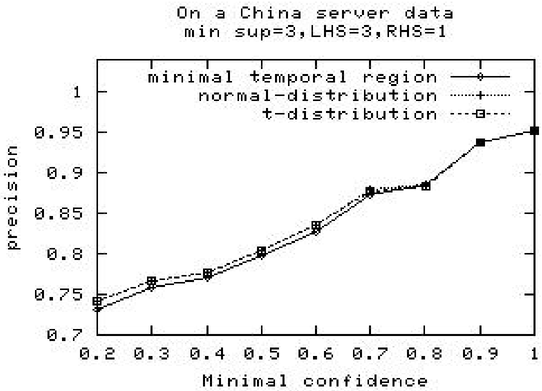
increase as the minimal confidence increases. The decrease in EPA data when minimal confidence is one

We next varied the size of LHS. We set min_conf=0.6, min_sup=15 for the China data and min_conf=0.3, min_sup=12 for the NASA data. The results are in Fig 7a and 7b. As the LHS size increases, the prediction precision first increases, then decreases, especially for China web data. This is because as larger LHS rules are admitted, more overfitting rules are also admitted. These rules typically have high confidence. Thus, there is a decrease in precision when LHS past 3.

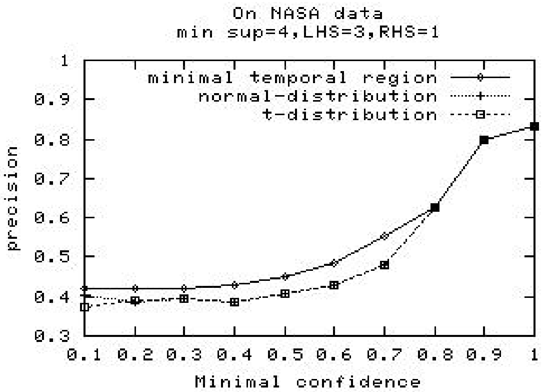
We also tallied the number of rules in our rule sets with different sized LHS's. This allows us to show what proportion of the predictions benefited from rules of different lengths. The results are shown in Table V. The n% is the defined as



a. Precision as the Minimal Confidence increases on EPA data

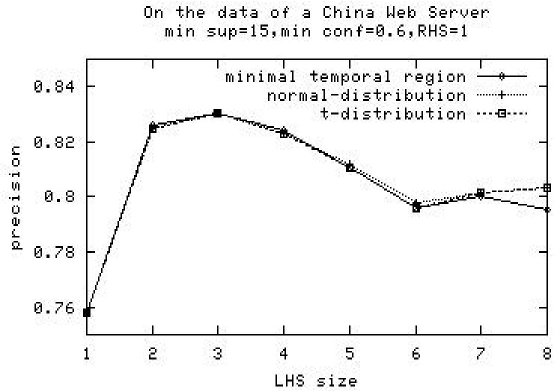


b. Precision as the Minimal Confidence Increases on a China Web data

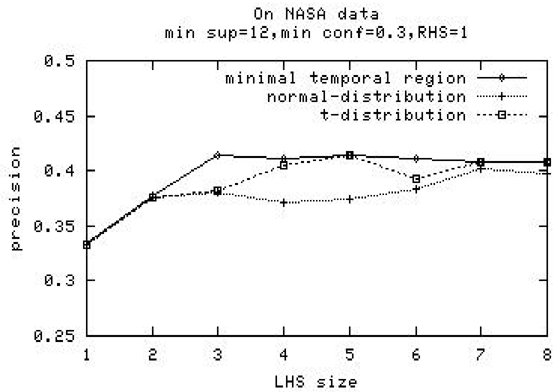


c. Precision as the Minimal Confidence Increases on NASA data

Fig. 6. Precision results on three data sets



a. Precision as the LHS Size Increases on a China web Server data (set min_conf=0.6, min_sup=15)



b. Precision as the LHS Size Increases on NASA data (set min_conf=0.3, min_sup=12)

Fig. 7. Precision results as the LHS Size Increases

the proportion of LHS size over all the predicted instances. From Table V, we can see that for all methods, the majority of rules are still length-one rules. However, there is a significant number of length-two rules as well.

Finally, the time complexity of the confidence-based model in our web-log mining problem is only linear in the length of the web logs. For each antecedent window W_1, the number of consecutive substrings that end with the end of the window is of size $|W_1|$. Thus, the number of LHS's that need be examined is only $O(|W_1| * |WebLog|)$. In contrast, the minimal temporal-region selection method takes $O(|Weblog|^2)$ time to mine in the worst case, since for each event E which occurs N times, it computes all $O(N^2)$ time intervals before selecting a best one according to formula (7).

V. Related Work

Much recent research activity in sequence prediction falls into the research areas of data mining and computer networks. In the data mining area, most algorithms are designed to deal with a database consisting of a collection of transactions (see [13] for example). These records store the transaction data in applications such as market-basket analysis. The focus of research has been how to perform efficient and accurate association and classification calculations.

In data mining area, general classification algorithms [13] were designed to deal with transaction-like data. Such data has a different format from the sequential data, where the concept of an attribute has to be carefully considered. As shown in this paper, these algorithms can be used to build the prediction models by applying a 'moving-window' algorithm across the whole web log sequence, such that the transactions appearing together in the same window can be regarded as a record in transaction data.

Association is another extensively studied topic in data mining. Association rules [3] were proposed to capture the co-occurrence of buying different items in a supermarket shopping. It is natural to use association rule generation to

TABLE V
DISTRIBUTION OF THE LHS SIZES FOR RULES USED IN THE PREDICTION
(MIN_SUP=10, MIN_CONF=0.3, N-BEST=1, LHS¡=3, RHS=1)

		LHS=1	LHS=2	LHS=3
A China Server	Naive	78.5%	17.2%	4.3%
	Minimal temporal region	78.5%	17.2%	4.3%
	Standard Normal distribution	65%	22.9%	12.1%
	t-distribution	65%	22.9%	12.1%
NASA	Naive	35.6%	40%	24.4%
	Minimal temporal region	39.5%	39.6%	20.9%
	Standard Normal distribution	36.1%	39.8%	24.1%
	t-distribution	35.6%	39.4%	25.0%

relate pages that are most often referenced together in a single server session [14], [5]. However, correlation discovery is not sufficient to build a prediction model, because they do not consider the sequential nature of knowledge embedded in web logs. In data mining area, [4], [2] proposed sequential association mining algorithms, but these are designed for discovery of frequent sequential transaction itemsets. They cannot be applied directly for sequence prediction without first being converted to classifiers. [6], [17] considered using association rules for prediction and classification, which achieved observable improvement on accuracy of classification models, but they did not consider sequential data either.

In the network area, researchers have used Markov models and N-grams [16], [12], [15] to construct sequential classifiers. Markov models and N^{th}-order Markov models when parameterized by a length of N, are essentially represent the same functional structure as N-grams. Generally speaking, these systems analyze the past access history on the web server, maps the sequential access information in N consecutive cell series called N-grams, and then builds prediction models. [9] proposed several different ways to build N-gram based models, and empirically compared their performance on real-world web log data. [16], [17] performed empirical studies on the tradeoffs between precision and applicability of different N-gram models, showing that longer N-gram models can make more accurate prediction than shorter ones at the expense of lower coverage. [16] proposed an intuitive way to build the model from multiple N-grams and select the best prediction by applying a smoothing or 'cascading' model, which prefers longer n-gram models. [15] proposed a small variant version of the longest match method by defining a threshold to go down a certain sequential path. [12] suggested a way to make predictions based on K^{th}-order Markov models.

Researchers in Machine Learning [19], [20] have studied the temporal region learning to find event patterns represented in the form of temporal orders and time. Heuristic methods are studied to select the best rules to be applied. However, these methods have only been designed to discover rules for which the left-hand-side has size one, and are tested on artificially designed event sequences that are of small scale. In this paper, we extend the representation to include larger sized rules, and test the rule based prediction results on realistic and large-scale data sets.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we studied different association-rule based temporal region prediction methods for web request prediction. We studied three different methods, the naïve method, the confidence interval based methods and the minimal temporal region method for the prediction. Our conclusion is that the confidence interval based methods and the minimal temporal region methods perform similarly, with the latter being a little better in precision. Our method represents a novel extension of the association rule based classification method for large-sized sequential data.

In the future, we plan to explore more on the relationship between temporal region prediction and other types of classification. We will also try to integrate the different methods. We believe that the confidence interval based method can indeed be enhanced by factors such as the range and coverage factors used in the temporal region prediction methods.

REFERENCES

[1] M.Arlitt, R.Friedrich, L.Cherkasova, J.Dilley, and T.Jin. Evaluating Content Management Techniques for Web Proxy Caches. *HP Technical Report*, Palo Alto, Apr. 1999.
[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A.I. Verkamo. Fast discovery of association rules. In *Advances in knowledge discovery and data mining.* 307-328, AAAI/MIT Press, 1996
[3] R. Agrawal and R. Srikant. Fast algorithm for mining association rules. *Proceedings of the Twentieth International Conference on Very Large Databases.* 1994. pp 487-499
[4] R. Agrawal and R. Srikant. Minging Sequential Patterns. In *Proc. of Int'l Conference on Data Engineering,* Taipei, Taiwan, 1995
[5] E. Cohen, B. Krishnamurthy and J. Rexford, Efficient algorithms for predicting requests to web servers. In *Proceedings of the IEEE INFOCOM '99 Conference*, 1999.
[6] B. Liu, W. Hsu, and Y. Ma. Integrating Classification and Association Mining. In *Proc. of the Fourth Int'l Conf. on Knowledge Discovery and Data Mining (KDD-98)*, New York, 1998, pp. 80-86.
[7] L. Oreiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and regression trees. Wadsworth, Belmont,CA, 1984
[8] H. Mannila, H. Toivonen, and I. Verkamo. *Discovering frequent episodes in sequences.* In Proceedings of the First Int'l Conference on Knowledge Discovery and Data Mining (KDD'95), Montreal, Canada, August 1995. AAAI Press. pp. 210 – 215.
[9] A. E. Nicholson, I. Zukerman, and D. W. Albrecht. A Decision-theoretic Approach for Pre-sending Information on the WWW. In *PRICAI'98 - Proc. the Fifth Pacific Rim Int'l Conf. on Artificial Intelligence*, Singapore, page 575-586.
[10] J. Pei, J.W. Han, B. Mortazavi-asl and H. Zhu, Mining Access Patterns efficiently from Web Logs, In *Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD 2000*: 396-407. 2000
[11] M. Perkowitz and O. Etzioni. Adaptive web sites: Concept and case study. In *Artificial Intelligence*, volume 118 (1-2), pages 245–275, 2001.
[12] J.Pitkow and P.Pirolli. Mining Longest Repeating Subsequences to Predict World Wide Web Surfing. In *Second USENIX Symposium on Internet Technologies and Systems*, Boulder, C0, 1999.
[13] J.R. Quinlan, C4.5: programs for machine learning, Morgan Kaufmann, 1993
[14] J. Srivastava, R. Cooley, M. Deshpande and P. Tan, Web usage mining: discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1(2):12–23, 2000.

[15] S. Schechter, M. Krishnan and M. D. Smith, Using path profiles to predict HTTP requests. In *7th International World Wide Web Conference,* pages 457–467, Brisbane, Qld., Australia, April 1998.
[16] Z. Su, Q. Yang and HJ Zhang, A Prediction System for Multimedia Pre-fetching in Internet. In *ACM Multimedia Conference 2000*, 2000
[17] Q. Yang, H. Zhang and I. T. Li. Mining Web Logs for Prediction Models in WWW Caching and Prefetching . In The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'01, August 2001 San Francisco, California, USA.
[18] Qiang Yang and Henry Haining Zhang. Integrating Web Prefetching and Caching Using Prediction Models. In *World Wide Web Journal. Kluwer Academic Publishers.* Vol. 4 No. 4, 2001. Pages 299-321.
[19] W. Zhang, A region-based learning approach to discovering temporal structures in data, *In Proceedings of the International Conference on Machine Learning.* 1999
[20] W. Zhang, Some Improvement on Event-Sequence Temporal Region Methods. In *Proceedings of the European Conference on Machine Learning.* PP 446-458. 2000

Exploring Computational Mechanism for Contexts

Leiguang Gong[1]

Abstract--- **Human problem solving is done within a context, which constrains the solution space. There have been growing interests in computational modeling of contextual knowledge representation and context-based problem solving. There are different approaches to exploring the concept of context and corresponding computational mechanism for representing and utilizing the contextual model, both formal and empirical based on different interpretations of the notion. This paper explores and discusses some aspects of the computational mechanism for contexts, and proposes a goal-directed framework for context-based problem solving. In this model, a context is defined in terms of an object and its relationships with other objects. Every context is centered at such an object, and cannot exist without it. A problem-solving task (e.g. find an object) with this framework can thus be defined as a process of determining a context or a sequence of contexts in which a solution path (e.g. steps for finding the object) can be decided. A set of operations is specified for context manipulation. The questions the author tries to answer are: what is context, how can contexts be formed and manipulated, and how can contexts be used in reasoning and problem solving.**

I. INTRODUCTION

Strictly speaking, all human problem solving is done within a context, which constrains the solution space. Even universally applicable solutions are conditioned by the only universe we know of. Philosophers and psychologists have studied the concept of context for quite sometime [1,6,10,14,19]. However, serious interest and research effort in the scientific community of machine intelligence have only been demonstrated in the last ten to fifteen years. Most research in machine intelligence still assumes context is defined a-priori by the investigator and is external to the computational problem solving systems, though the importance of representing and using contextual knowledge or information in a explicit and systematic way has been recognized by more and more researchers and demonstrated by increasing number of theoretical works and application systems [3,5]. In the area of computational intelligence, there are many different approaches to exploring the concept of context and the corresponding computational mechanism for representing and utilizing the contextual model. Like any other subjects in the field of artificial intelligence, there are both formal and empirical approaches. Research on formalizing contexts [4,13,17,18] has been primarily concerned with the locality of knowledge, i.e. relativity of trueness in knowledge bases, and its applications in knowledge representation and integration. The most well known example of such application is the Cyc system [15]. The empirical approaches, that are often more closely associated with real-world applications, focus on domain-specific architectures and techniques for representing and using contextual knowledge and information [5,7,11,16,20]. Many industrial applications of contexts such as context-aware computing and CRM systems [9] are based on empirical approaches. Research emphases for those applications are acquisition and representation of (user/customer) contextual information and the design of inference rules that encode the context-based inference steps. Various model, design architectures, and techniques have been proposed and developed [5,11,20].

The functional nature of contexts is its filtering, constraining power. It is no surprise that many personalization-based applications such as CRM, information delivery, HCI and etc. are built based on the concept and related techniques. Cyc and others use the concept to structure knowledge bases to maintain the consistency and trueness. However, contexts in the context of problem solving, particularly in real world physical contexts related applications such as navigations, HCI, robots, etc. are the subject's perception of its environment and surroundings. The task to the subject is to achieve the intended (e.g. find an object) or unintended goal (e.g. get out of the unexpected dangerous situation). The challenge is to find the right solutions and find them quickly. From a computational point of view, such problem solving process can be formulated as a process of context transformation. For a given context and a goal, the subject constructs an initial solution space from the knowledge memory, and then transforms the given context stepwise into a context in which the final goal is directly reachable. The criteria for guiding or controlling the transformation are goal-directed and domain-specific. Additional knowledge and

1. T. J. Watson Research Center, IBM, Hawthorne, NY 10532, USA, leiguang@us.ibm.com

information may be acquired either from memory or from other subjects or external environment as the transformation proceeds.

This article will focus on the discussion of some basic issues and concepts for goal-directed context-based problem solving. Specific examples (both realistic and hypothetic) will be used in the discussion. We propose a generic computational framework for contexts. The proposed model is not intended to explore and interpret the full semantics and the mechanism of contexts in human cognition and problem solving, rather to try to model some important aspects of computational mechanism of contexts. The questions we try to answer are: what is context, how can contexts be formed and manipulated, and how can contexts be used in reasoning and problem solving.

II. WHAT IS CONTEXT

Context is one of most frequently used word [3]. In modern language, the word context is defined in dictionaries (www.dictionary.com) bearing two basic meanings: (1) "To knit or bind together; to unite closely" and (2) "That which surrounds, and gives meaning to, something else." Although there are many other interpretations, the basic meanings are same. But in computational modeling of contexts, the understanding and interpretation of the notion has become increasingly diverse and various. All kinds of definitions and treatments of the notion can be found from the research literature in the field [2,3,7,20]. However, no matter how different these definitions are most of them fall into either (1) or (2). From a computing point of view, these two definitions specify two different computational mechanisms. If we looked at closely how the notion of context is actually used in both human communication and computational systems, two patterns of usage stand out: "in the context of X" (CO) and "the context for X" (CF), where X stands for any physical or conceptual entity and event (e.g. a game, the human history, AI research, etc.). For instance, by saying "in the context of yesterday's baseball game", we generally mean a space of all things and events within the ball game. By saying the context for yesterday's baseball game, we generally mean a collection of things and events externally related to the ball game. These two patterns of using contexts represent two different, though complimentary, views and treatments of the notion. Most (if not all) works in contextual modeling either intentionally or unintentionally reflect one of the two views. The pattern of CO views a context as a space of everything grouped or contained by a thing X (e.g. a game). Computational systems with this view try to answer the general question, "what can be said and done given the X". On the other hand, the pattern of CF views context as a space of things externally related to and referenced by X. Computational systems with this view try to answer a question of "what is the contextual space for X, and how can this space be used to help understand X". In computational terms, a common semantic element of the two different views and treatments of context is the notion of space, i.e. a space of information structured from either external sources, or from internal sources, or from both. In CO, such spaces are formed (knitted) through partitioning of larger information or knowledge spaces. In CF, such spaces are the results of growing about reference objects. CO models the mechanism defined in (1) and CF models the mechanism defined in (2). CO and CF represent the two general approaches to contextual modeling in machine intelligence. Because of our interest in CF modeling, we will only give a computational definition based on (2).

We assume that a meaningful context does not exist without a reference object. In other words, a context is always defined in reference to an object, i.e. centered at an object. The object can be anything from a physical entity to an abstract concept such as a person, a house, a word or sentence, an idea, a feeling, an event and so on. A context can be in general defined as a space of objects or entities, and relationships/associations among them centered at a reference object. we adopted a simpler and more quantifiable definition.

A *context* is the collection of an object and its relationships with other objects including these objects. The object is called target object o or reference object. To formulate this more concisely, $c = (o, R, B)$, where R is the set of all the relationships the target object has with all the objects in B. An *object* is a physical or conceptual entity with a unique identity. An object with more than one subcomponents or parts is called compound object. An object is atomic if it does not have subcomponents. A table leg can be treated as an atomic object, whereas tables as a compound object. A *relationship* is a mutually connected bond between two objects (e.g. *is-a, contained-in, married-to*, etc.).

III. APPROACHES TO CONTEXT

3.1 General Approaches

As we indicated in the previous section that there are two general approaches to contextual modeling: CO and CF.

- The CO approach can be expressed as following: Given a context C (a set of unspecified or partially specified assumptions), what then can be inferred within C? C may be recursively defined by one or more smaller contexts or sub-contexts.
- The CF approach can be specified as:

Given an object of interest O, what context or a sequence of contexts can be formed, such that a task about object O can be accomplished?

Most formal approaches are CO approaches, since they treat the notion of context as a mechanism of structuring (knitting) knowledge in terms of logical partitions, or simply described as ist(c,p) (asserting that the proposition p is true in the context of c). The significance of research in contextual knowledge representation using CO has been demonstrated in building very large-scale knowledge bases [15]. In CF approach, context is modeled as a space of objects and their interrelationships in the vicinity of an (or a group of) object (or called target object or reference object). The basis of this model is the referential relationship or association between the target object and its surroundings. Such mechanism specifies a neighborhood of things and relationships anchored by a or a group of target or reference objects either recorded from the world external to the problem solver or internally in its memory [14]. This object of reference plays a central role in both human and machine context-based problem solving. It in general refers to the goal of the problem solving. For instance when we talk about the context for an English word, usually understanding the meaning of the word is the goal of a human or machine reader. In medical imaging diagnoses the determination of abnormal findings is usually made in the context of specific human body parts (e.g. brain) or biological subsystems (e.g. digestive system). Furthermore, in a complex problem solving, to reach a final goal may require the fulfillments of some intermediate goals. The realization of each intermediate goal may in turn depend on its own context. For instance, to understand the word in the above example may require a clear understanding of another word in a statement as part of the context for the first word. The context for understanding the second word is usually different from the context for the first word. Generally speaking, the CO approach is rather static in nature, for it must serve the purpose of logical partition of a knowledge base, though the partitions can be gradually expended and revised by various means of learning. But contexts may not be formed or changed during a particular session of problem solving. On the other hand, CF approach is a rather dynamic mechanism, since it has to be formed, by the human or machine problem solver, during the process of a particular problem solving cycle. As a matter of fact, the two seemly very different treatments to contexts represent two closely interrelated aspects of context-based problem solving. For instance, the task of a goal-directed intelligent agent upon entering a situation/environment is to timely form an internal representation of the situation (i.e. context) in such a way that irrelevant information can be quickly filtered out, and useful information preserved. In the process of context formation and other related operations, the agent may recall the similar experience it had before along with the contextually structured knowledge for dealing with the situation if such knowledge is available internally. The experience learned from this encounter will be remembered and processed with other collected lessons internally by the agent's learning mechanism, one of key learning results will be the expansion or revision of the contextual partitions of the internal knowledge base.

In context-based problem solving, frequently a goal or a task is directly related to an object. This is particularly true for image understanding and robot navigation, where the main goal is to localize a target object. In CRM and many online customer service systems [20], the goal is to understand the customer's needs. Applications of this kind require the recognition and understanding of the target objects in terms of their contexts. The question then is what is the context for a target object? How can such a context be formed, represented and used? If the localization or understanding of the target object cannot be accomplished directly, what context manipulations can be done to help go through necessary intermediate steps?

We propose an object-centered computational model of contexts for machine intelligent problem solving. In this model, a context is defined in terms of an object and its relationships with other objects. This object is called the reference object or the target object. Every context is centered at such an object, and cannot exist without it. A problem-solving task (e.g. find an object) can thus be defined as a process of determining a context or a sequence of contexts in which a solution path (e.g. steps of finding the object) can be decided. A set of operations or operators is specified in this framework. The actual rules (criteria) for deciding specific contexts are problem-specific. The proposed framework is intended to explore some aspects of the very elusive and complex concept of context in the context of machine intelligent problem solving, and to facilitate systematic design of context-based modeling and problem solving systems.

3.2 A Framework for Goal-Directed Context-based Problem Solving

Generally speaking, the usage of contexts by intelligent systems is supported by the assumption that a context or a set of contexts may either contain the solution to a problem or may provide sufficient supporting data or information. This assumption is quite valid for most context-based problem-solving tasks. A general computational framework of context for machine intelligent problem solving is illustrated in Figure 1. The context engine (the middle block) works with and supports a problem solving or planning system by forming or representing an initial context, revising or changing (shifting, growing, and shrinking) the context according to the current goal. The initial context may be acquired through various types of sensors such as cameras and microphone, and/or from other interaction channels.

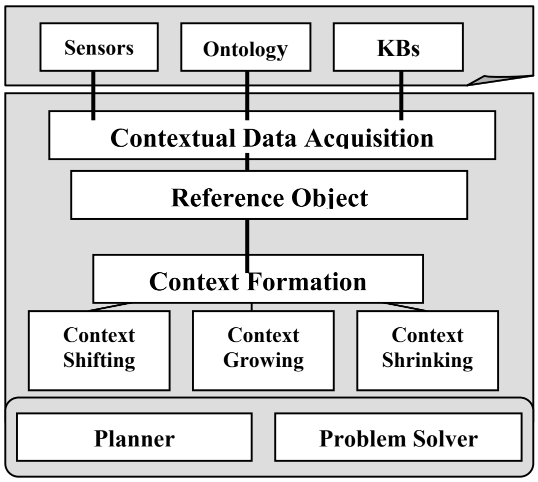


Fig. 1. A framework for context-based problem solving

In this framework, once a reference or goal object is determined, the formation of a context or a set (sequence) of contexts can be carried out either interactively with the external environment (e.g. continuous sensor data or a human user), or internally as a planning process [12] if the task is fixed at the beginning (e.g. analyze a CT scan for brain abnormalities) and prior knowledge about the domain is available (e.g. a semantic net model of the human brain in knowledge bases (KBs)). The question is for a given specific task, what contextual information is needed and how such desired context(s) can be determined and formed. We will explore these issues in the next section followed by some application examples.

3.3 Context Manipulations

An initial context for a specific problem-solving task can be characterized as being either too broad, too narrow, irrelevant, or just right. If the initial context is too broad, the problem solver may have difficulties focusing on most relevant and valuable information. If it is too narrow, then most relevant and valuable information may be missing. If the initial context presented to the problem solver has nothing to do with the given task, the context is irrelevant or wrong. However, in some situations to achieve the final goal may require a strategy of achieving some intermediate goals using different contexts, though the initial context is relevant to the task. So it is fundamentally important for a context-based intelligent system to be able to correctly evaluate the initial context. Context evaluation is still an open research problem, and will not be further discussed here.

The following context operations are specified: context shifting, context growing, and context shrinking, defined as following:

3.3.1 Context Formation

Given a world of objects and relationships $w\{O, R)$, a context c_i can be formed by a formation function f, i.e. context $c_i = f(w) = (o_i, R_i, B_i)$ where o_i is the reference or target object, R_i is the set of relationships o_i has with the set of objects B_i. For instance, in image understanding, a vision system requested to detect a specified object such as a boat from a satellite picture, may first form an initial contextual model about the boat and its relationship with other possible background objects. In this case, w is all the objects and their relationships recorded from the entire visual field in the photo, o_i is boat, B_i is the entire background, and R_i is the relationship *in*. In CRM applications, an intelligent system must be able to form, from all available customer data w, an initial representation of a customer o_i profile (i.e. contextual model of the customer), such that valuable information about the customer (such as the customer's financial interest) can be retained and used for inferring the system's recommendations or suggestions for product or service purchasing. B_i in this case may contain things like family members, properties, education background, etc. R_i may include relations like *has, owns, interested-in*, etc. In highly controlled and constrained applications, such contextual formation can be done in an optimal way, i.e. no irrelevant or redundant information will be collected and represented. In less controlled real world environments, application systems like robot navigation and active vision systems may not, at the beginning, have a clear idea about the boundary of the required context. The initial context formation most likely is an approximation. The degree of relevancy of the approximation is determined by both the perception power and the prior knowledge about the environments of the systems.

3.3.2 Context Shrinking/Pruning

For a given context $c_i = (o_i, R_i, B_i)$, a *context-shrinking* operator is defined as: $hop(c_i) = c_j$ where $c_j = (o_i, R_j, B_j)$, and $R_j \subset R_i$ and $B_j \subseteq B_i$. The new context is pruned in that it at least has fewer relationships. Please note that the operator implies that when an object has no relationship with the reference object, the object is automatically removed from the context. There may be multiple relationships between the object and the reference object. Pruning one relationship may not necessarily result in the object being removed from the context. In the previous vision system example, the initial context background may be too big to focus on. The vision system may need to reduce the context to the point (e.g. river) that the target object can be quickly and uniquely identified. In the above CRM example, the initial context formation may contain too much information about the customer. Some kind of filtering is needed to prune the space. Context shrinking or pruning is a filtering and focusing mechanism. A fundamental aspect of human intelligence is the ability to filter away irrelevant or relatively unimportant contextual information, and to concentrate on the contextual space within which the solutions or the goals can be found or achieved efficiently. The question of what and how much should be pruned is an open and challenging one, and can only be answered in the context of a specific system and the context of the problems the system is asked to solve.

3.3.3 Context Growing

For a given context $c_i = (o_i, R_i, B_i)$, a *context-growing* operator *gop* when applied to a context creates a revised context with more relationships. It is defined as: $gop(c_i) = c_j$, where $c_j = (o_i, R_j, B_j)$, and $R_i \subset R_j$ and $B_i \subseteq B_j$. This operator is the inverse of the shrinking operator. Sometimes, for whatever reason, an initial context may not be broad enough to cover needed information, or too much prune is done to the previous context. The context thus needs to be expanded to the point that sufficient information is available for the system to reach the goal. For instance, the vision system in the above example may fail to localize the target object with the initial contextual information. It may be due to its initial contextual model of the perceived world is too narrow and does not contain the target object. Or it may because the initial field of view is too small to cover the target object. In either case, the initial context has to grow until the necessary information is present. The same question asked for context shrinking can be raised here too.

3.3.4 Context Shifting

For a given context $c_i = (o_i, R_i, B_i)$, *a context-shifting* operator *sop* when applied to a context creates a new context that at least has a new target object. It is defined as: $sop(c_i) = c_j$, where $c_j = (o_j, R_j, B_j)$ and $o_i \neq o_j$. Context shifting is signified by the change of focus, i.e. reference object. Many real world applications involve complex problems, where a goal can only be achieved stepwise through the realization of some intermediate sub-goals. To achieve these sub-goals, the initial context has to be transformed (shift) into a sequence of different contexts within which the solution paths to the sub-goals can be found. In vision example, if the detection of the target object boat requires the detection of another object *river*, the context for localizing *river* has to be formed. In situations (e.g. robot navigation) where unexpected events occur, the system is forced to change its priorities and shift of context must be made.

4 EXAMPLES

In image understanding, the goal usually involves detection or recognition of some object(s) or interpretation of some relationship(s) between objects in spatial contexts. This class of problems is particularly difficult to solve, when the initial visual context is very large and there are many objects similar to the target object to be recognized in the scene. For instance, it would be very difficult and time-consuming to look for a particular highway from a large satellite image of the earth without a spatial context small enough to constrain the analysis and recognition process.

A spatial context consists of a set of objects and the relationships between them. In imaging terms, a spatial context is the whole or part of an image space, i.e. an array of pixels or voxels within which a target objects is likely to be detected. Thus, the earth in a satellite image defines a spatial context as an approximately round-shaped region within which major landmarks such as oceans and mountains can be located. These landmarks can be used to further localize other less significant geographical objects, such as rivers, forests or deserts. Such landmarks, including the earth itself are reference objects, which provide anchor points for defining specific contexts. To establish the presence of a reference object frequently requires the presence of yet another reference object. Such a sequence of reference objects represents an abstract plan for detecting the target object from the given scene or image. The order in which contexts are established specifies a sequence of focal regions from which the chain of reference objects can be localized. We [11] report the experiments with MR scans of human brain. The task is to detect multiple sclerosis (MS) lesions from the image. In this problem, the initial context for the task is the entire image space with both the brain structure and the background. Our image analysis system determines in order to localize the MS lesions more accurately the initial spatial context needs to be reduced in a series steps. The system, according to its priori knowledge about the brain anatomy, generates a processing and analysis plan specifying a sequence of target or reference objects and their relationships (i.e. contexts). The nested containment of different anatomical and pathological (MS lesion) structures is a simple *contained-in* relationship and is used in the formation of a sequence of contexts (plan) for recognition. According to the analysis plan (detect brain first, then white matter, and finally MS lesion) generated by the planner [12], the system shift from the initial context (specified by the MS lesion and its *contained-in* relationship with the entire image) to a new context (specified by the new target object brain and its *contained-in* relationship with the image). Once the brain structure is segmented, the context shrinking operation is performed to eliminate the background region and the region of skull. Next, the context specified by the white matter and its relationship of *contained-in* the brain become present. The process continues until a smallest context (i.e. lesion is *contained-in* the white matter) is obtained.

Some more complex tasks require the problem solver not only to locate objects, but also to understand and interpret the relationships between the objects of interest. Assume we have a digital secretary for handling the incoming telephone calls for a company executive. The decision to any incoming call the secretary has to make are assumed to be either "put call through" or "take the message". Apparently such decision-making very much depends on the secretary's understanding and interpretation of the relationship between the caller and the called i.e. the executive. However, the importance of the relationship is all relative and context-based. The initial contextual information our digital secretary has is the caller and the executive. They are related at the moment by a caller-called relationship. Apparently such a context (centered-at the executive) does not provide enough information for it to make a appropriate decision. The context needs to be expanded (by *gop* operation). So the information that the caller married to the executive (relationship) is added to the initial context. At this moment, the decision of "put through" seems to be the right one, but still the secretary is not sure. It then adds the fact that executive is currently *in* a (important) *meeting*. Now it looks like that the secretary should just take a message from the caller, but it recognizes that not enough is known about the caller's present context. The secretary then shifts the context from executive-centered to the caller-centered, and collects that the caller (the executive's spouse) is in a medical emergence room. The secretary put the call through immediately to the executive.

V. DISCUSSIONS AND FUTURE WORK

Context is a very elusive concept. While research on contextual modeling and problem solving has been intensified in recent years, a coherent computational theory about context has yet to be developed. This article proposes a computational framework for context-based reasoning and problem solving. The framework is based on the definition of context as a space of objects and their relationships centered at a reference or target object. The work reported here is still limited in several ways. More studies are needed to understand the extent of generality of this model and its potential and limitation in broad range of application domains. Some important issues such as the evaluation of the goodness of a context and the comparison of contexts are yet to be investigated.

REFERENCES

[1] D. Annis, "A Contextual theory of epistemic justification", Amaerican philosophical Quarterly, 15: 213-29, 1978.

[2] M. Benerecetti, P. Bouquest, and C. Ghidini, "On the dimensions of context dependence: partiality, approaximation, and perspective", Proceedings of CONTEXT 2001, pp. 59-72, 2001.

[3] P. Brezillon, "Context in problem solving: a survey", The Knowledge Engineering Review, 14(1):1-34, 1999.

[4] S. Buvac, "Quantificational logic of context". In Proceedings of the Eleventh National Conference on Artificial Intelligence, AAAI, Menlo Park, Ca. (AAAI-93), Washington, DC, 1993:412 -419.

[5] G. Chen and D. Kotz, "*A Survey of context-aware mobile computing research*". Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College.

[6] S. Cohen, "Knowledge, context, and social standards", Synthese, 73: 3-26, 1973.

[7] A. Dey and G. Abowd, "Toward a better understanding of context and context-awareness", *Proceedings of the CHI 2000 Workshop on The What, Who, Where, When, and How of Context-Awareness*, The Hague, Netherlands, April 2000.

[8] IV. Ezhkova, "Knowledge formation through context formalization", Journal of Computers and Artificial Intelligence, 8(4): 305-22.

[9] J. Fink and A. Kosba, "A review and analysis of commercial user modeling servers for personalization and world wide web", Journal of User Modeling and User-Adapted Interaction, Kluwer Academic Publishers, 10: 209-48, 2000.

[10] A. Goldman, "Psychology and philosophical analysis", Proceedings of the Aristotelian Society, pp. 195-209, 1989.

[11] L. Gong, and CA. Kulikowski, "Defining spatial contexts for focused image analysis", in Proc. of IEEE SYSTEMS, MAN, AND CYBERNETICS 2001. Arizona, pp. 1184-87.

[12] L. Gong, and C. Kulikowski, "Composition of image analysis processes through object-centered hierarchical planning", IEEE Transaction on PAMI, Vol. 17, No.10, pp. 997-1009, 1995.

[13] R. Guha, "Contexts: a formalization and some applications", PhD thesis, Stanford Univ. 1995.

[14] F. Heylighen, "Advantages and limitations of formal expression", Foundations of Science 4(1), 1999.

[15] D.B. Lenat, "Cyc: a large-scale investment in knowledge infrastructure." Communications of the ACM 38, No.11, 1995.

[16] C.L Lin, P. Burlina and R. Chellappa. "Global activity detection: contextual recognition of regular object configurations". Technical Report, CAR-TR-748, Center for Automation Research, Univ. of Maryland, Dec. 1994.

[17] J. McCarthy, "Notes on formalizing contexts", Proceedings of 13th IJCAI, Vol. 1, pp. 555-60.

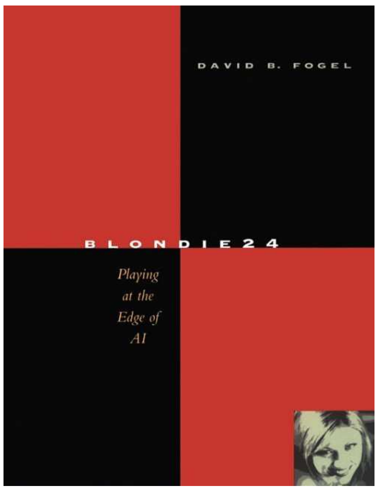
[18] J. Sowa, "Representing and reasoning about contexts", Proceedings of AAAI-92 workshop on prepositional knowledge representation, Stanford, pp. 133-42. 1992.

[19] N. Vassallo, "Contexts and philosophical problems of knowledge", Proceedings of CONTEXT 2001, pp. 353-66, 2001.

[20] T. Winograd, "Architectures for context", *Human-Computer Interaction*, 16:2-3, 2001.

Blondie24, Playing at the Edge of AI

By David B. Fogel, Morgan Kaufmann Publishers, 2002, ISBN 1-55860-783-8



Reviewed by Eric Harley[1]

Computer science has failed abysmally at producing machines which display intelligence. According to Fogel, the last 50 years of effort in artificial intelligence have been on the wrong track, leaving us no closer to the goal than when we started. The wrong track has been the attempt to make the computer imitate our behavior. Scientists striving to build an artificial intelligence load the computer with knowledge on a chosen topic, along with an algorithm to do the associated task. The hope is that the computer will equal or surpass our intelligence for that subject. This approach dates back to the Turing Test, which Fogel points out has been misquoted and misinterpreted almost from day one. Misquoted or not, the Turing Test has a computer pretend to be human, and this paradigm became a signpost saying that the road to artificial intelligence is through mimicry of human behavior. According to Fogel, however, that path leads only to an illusion of intelligence — for example the kind of wooden intelligence exhibited by Deep Blue.

If imitation of human intelligence has not led to machine intelligence, then what will? What is intelligence? Fogel is critical in general of researchers in this field for skirting this last question. He believes that if the field of artificial intelligence had started with a proper definition of intelligence then there would have been a better chance of creating it. That is common sense — knowing what you are trying to build is crucial. Fogel therefore gives the following definition of intelligence.

> Intelligence is the capacity of a decision-making system to adapt its behavior to meet goals in a range of environments.

Fogel looks to nature for an example of intelligence. He describes the intricate and seemingly clever behavior of a certain species of wasp. However, he points out that an individual wasp is fixed in its behavior. An experiment by Jean Henri Fabré described in the Notes section of the book seems to make this clear. The wasp, Fogel says, is like the 'proverbial robot', an automaton with no adaptive behavior. Therefore the individual wasp is not intelligent. However, Fogel considers the species of wasp to be intelligent as a group or system, since it evolves to meet a changing environment. The species is what has learned the intricate behavior. The point is made that in general, intelligence requires a reservoir of knowledge and a mechanism to adjust that knowledge in the face of changes in the environment. The holder of the knowledge depends on the system — for a species it is the genetic pool, for a social group it is the culture, and for an individual it is the brain.

According to Fogel, a definition is neither right nor wrong, but rather useful or not useful. Perhaps that is true when one is making up a new definition, or defining a term for use in a local environment. The above definition of intelligence is useful, since it leads to an algorithm for creating an intelligent machine. The idea is to mimic evolution. First you create a population of machines, each with a reservoir of knowledge which could initially be basically garbage. Then you repeat a two-step process of variation and selection to cause the

[1]Ryerson University, Canada, eharley@scs.ryerson.ca

knowledge to evolve in some direction. The variation step duplicates the machines and applies some random changes to the knowledge in the daughter machines. The selection step tests the machines against some criterion defined in relation to a goal and removes the weaker half of the population. The expectation is that after some number of generations, the resulting machines will be excellent at achieving the desired goal.

To test his idea, Fogel decides to use the algorithm to create a program that can teach itself to play a game without being told ahead of time what are the characteristics of good moves or even what are the features to consider when assessing moves. This would be a learning machine, adapting to an environment that demands ever-increasing levels of skill. Fogel chooses the game of checkers, rather than chess, since checkers like chess involves very little if any luck, and because checkers is a simpler game, more appropriate for his modest computing resources (a 400 MHz Pentium II computer). He chooses a simulated neural network to be the reservoir of knowledge. The neural network is set up to be the evaluator of board positions. In essence, it is a mathematical function whose inputs represent the 32 squares on the checker board, each with a value denoting what piece if any is on the square. The knowledge is captured in the weights of the connections and the thresholds. Some of the history and theory of neural networks is described, so that the reader has a fairly clear idea of what these terms mean and how they relate to concepts such as 'hill-climbing' and 'local maxima'. Variation is carried out by a program which duplicates the neural networks and applies some constrained random variation to the weights. Selection is performed by a program which uses the neural networks as board evaluators in competitions within the population of neural networks. Each neural network plays at least five games, and the neural networks with the best scores live on for the next stage of variation, while the poorer performers are removed from the population. The hope is that after many generations of variation and selection the best neural network will be an expert at checkers (i.e., have a rating above 2000).

Clearly, the above experiment requires a lot of programming, which is undertaken by Kumar Chellapilla, a graduate student with an interest in evolutionary algorithms. While Chellapilla sets up the program, Fogel studies the history of programs that play checkers. One chapter of this book is devoted to Arthur Samuel, a pioneer in artificial intelligence and in the world of checkers programs. Samuel, working at IBM in the 1950's created a checkers program which to some extent taught itself. Samuel constructed its evaluation function from a number of features of positions in checkers which human experts over the years have determined to be important. Samuel then allowed the program to play against itself and adjust the weights of these features to improve its play. The challenge that Samuel left open to future researchers was to design a program that could invent its own features, i.e., one that could learn the game from scratch without human advice. This is one of the goals of the experiments in this book, so it is named Samuel's challenge. Another famous pioneer of computer science, Allan Newell, expressed the view that it is very unlikely that win/lose information is sufficient to allow a game-playing program to improve its performance. This statement contrasts with Fogel's experimental design. The selection step in his evolutionary algorithm is applied after the neural networks have been used to play at least five games. The neural networks of course are given no feedback regarding which moves were good or even which games resulted in a win. Thus, Fogel sets his experiment historically in terms of the *Samuel-Newell challenge*. It is an attempt to do what Samuel wanted to do (make a program that learns features by itself) in a way that Newell thought would not work (using only win/lose feedback).

I do not want to go into the results of Fogel's experiments, since this comprises much of the excitement of the book. He and Chellapilla test the evolved programs on an internet checkers club, and these tests are described with humor and anticipation, as well as a scientific eye. I will just say that Blondie24, the most evolutionarily-advanced of his programs, attains a level which is truly amazing, considering that 'she' is given no knowledge beyond the basic rules and piece count.

In conclusion, let us consider what Fogel has achieved, besides writing a very entertaining and informative book on artificial intelligence. Did he meet the Samuel's challenge? I think the answer here is clearly yes, since the evolutionary program that he designed learned to play checkers at a very high level without input of human expert knowledge. The Notes state that some reviewers objected that his program may simply be using the piece count. He rules this out by showing that the program is much better than any program that bases its decisions simply on the differential and worth of pieces. What attributes or features of checkers did the program discover to be important? No one knows, since that information is buried in the array of weights, but the book presents some anecdotal evidence that mobility might be one of the discovered features. Did he meet Newell's challenge? Without a doubt, since there is no feedback to the program after moves. The program definitely learned to play expert checkers via selection based on the results of five or more games.

Did Fogel create artificial intelligence? I think the answer again is yes. The evolutionary program has a built-in goal to make its neural networks better checkers players. It has a self-created environment, which is the checkers tournament. It makes decisions regarding which neural networks to throw out and which to keep for future generations. It adapts, since the neural networks continually get better under this selective environment. An interesting point is that the end product which looks intelligent is Blondie, yet she is not in fact the intelligence. Like the individual wasp, Blondie is fixed in her responses. If she played a million games, she would not be an iota smarter. In this sense, she is like Deep Blue. In retrospect, if you consider not just Deep Blue but also the program which adjusted the weights of features in Deep Blue's evaluation function based on many master-level games, then that program is also an example of artificial intelligence.

Is Fogel on the right track for creating the kind of general artificial intelligence symbolized in the computer HAL of *2001, A Space Odyssey*? This we do not know. Fogel says he has no idea how to build such a creative computer. I wonder if he has good enough definition of intelligence, or if perhaps his interpretation is not quite right. His definition refers to a decision-making system, and Fogel seems to be quite liberal in his interpretation of what that can be. He says it can be a whole species of wasp, yet this does not fit well with our intuition. Changes in the behavior or anatomy of the wasp presumably arise because of variations in the genetic pool and selection by the environment for the individuals that are better suited for survival. Thus, the species evolves and adapts to the environment, but where are the decisions being made? Neither the species nor the environment seems to be involved in any decision-making. If the wind blows down all but one well-rooted, hardy tree, which goes on to make a forest, then we may say there was selection and adaptation, but not a decision. Therefore, I do not think a species of tree or a species of wasp is intelligent, based on Fogel's definition. Perhaps a better example of intelligence would be an ape that decides to use a stick to probe for termites, or a human, who can adapt her behavior to any number of new challenges.

Fogel believes that the future of artificial intelligence lies in machines that can learn, since only they will be able to solve the problems for which we do not have solutions. Furthermore, he thinks that the way to achieve this goal is by copying the process of evolution, which already has proved successful in generating intelligence. If Fogel is right, then probably what his evolutionary algorithm should be selecting for is the property we define as intelligence, or a concomitant (maybe a sense of humor =)). Time will tell if his definition of intelligence is useful enough to lead him or others to the proper experiment.

Related Conferences & Career Opportunities

TCCI Sponsored Conferences

WI 2003
The 2003 IEEE/WIC International Conference on Web Intelligence
Beijing, China
Oct 13-17, 2003
http://www.comp.hkbu.edu.hk/WI03/
Submission Deadline: March 20, 2003

Web Intelligence (WI) is a new direction for scientific research and development that explores the fundamental roles as well as practical impacts of Artificial Intelligence (AI) and advanced Information Technology (IT) on the next generation of Web-empowered products, systems, services, and activities. It is the key and the most urgent research field of IT in the era of World Wide Web and agent intelligence. The IEEE/WIC 2003 joint conferences are sponsored and organized by IEEE Computer Society Technical Committee on Computational Intelligence (TCCI) and by Web Intelligence Consortium (WIC), an international organization dedicated to the promotion of world-wide scientific research and industrial development in the era of Web and agent intelligence.

The technical issues to be addressed include, but not limited to: Intelligent Web-Based Business, Knowledge Networks and Management, Ubiquitous Computing and Social Intelligence, Intelligent Human-Web Interaction, Web Information Management, Web Information Retrieval, Web Agents, Web Mining and Farming, Emerging Web Technology.

———————-

IAT 2003
The 2003 IEEE/WIC International Conference on Intelligent Agent Technology
Beijing, China
Oct 13-17, 2003
http://www.comp.hkbu.edu.hk/IAT03/
Submission Deadline: March 20, 2003

The 2003 IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003) is a high-quality, high-impact agent conference, which is jointly held with the 2003 IEEE/WIC International Conference on Web Intelligence (WI 2003). The IEEE/WIC 2003 joint conferences are sponsored and organized by IEEE Computer Society Technical Committee on Computational Intelligence (TCCI) and by Web Intelligence Consortium (WIC), an international organization dedicated to the promotion of world-wide scientific research and industrial development in the era of Web and agent intelligence.

The upcoming meeting in this conference series follows the great success of IAT-99 held in Hong Kong in 1999 (http://www.comp.hkbu.edu.hk/IAT99/) and IAT-01 held in Maebashi City, Japan in 2001 (http://kis.maebashi-it.ac.jp/iat01/). The aim of IAT 2003 is to bring together researchers and practitioners from diverse fields, such as computer science, information technology, business, education, human factors, systems engineering, and robotics to (1) examine the design principles and performance characteristics of various approaches in intelligent agent technology, and (2) increase the cross-fertilization of ideas on the development of autonomous agents and multi-agent systems among different domains. The technical issues to be addressed include, but not limited to: Applications, Computational Models, Architecture, and Infrastructure, Autonomy-Oriented Computation (AOC) Paradigm, Learning and Self-Adapting Agents, Data and Knowledge Management Agents, and Distributed Intelligence.

———————-

ICDM'03
The Third IEEE International Conference on Data Mining
Melbourne, Florida, USA
November 19-22, 2003
http://www.cs.uvm.edu/~xwu/icdm-03.html
Submission Deadline: June 10, 2003

The 2003 IEEE International Conference on Data Mining (IEEE ICDM '03) provides a leading international forum for the sharing of original research results and practical development experiences among researchers and application developers from different data mining related areas such as machine learning, automated scientific discovery, statistics, pattern recognition, knowledge acquisition, soft computing, databases and data warehousing, data visualization, and knowledge-based systems. The conference seeks solutions to challenging problems facing the development of data mining systems, and shapes future directions of research by promoting high quality, novel and daring research findings. As an important part of the conference, the workshops program will focus on new research challenges and initiatives, and the tutorial program will cover emerging data mining technologies and the state-of-the-art of data mining developments.

Topics related to the design, analysis and implementation of data mining theory, systems and applications are of interest. These include, but are not limited to the following areas: foundations of data mining, data mining algorithms and methods in traditional areas (such as classification, clustering, probabilistic modeling, and association analysis), and in new areas, mining text and semi-structured data, and mining temporal, spatial and multimedia data, data and knowledge representation for data mining, complexity, efficiency, and scalability issues in data mining, data pre-processing, data reduction, feature selection and feature transformation, post-processing of data mining results, statistics and probability in large-scale data mining, soft computing (including neural networks, fuzzy logic, evolutionary computation, and rough sets) and uncertainty management for data mining, integration of data warehousing, OLAP and data mining, human-machine interaction and visualization in data mining, and visual data mining, high performance and distributed data mining, machine learning, pattern recognition and scientific discovery, quality assessment and interestingness metrics of data mining results, process-centric data mining and models of data mining process, security, privacy and social impact of data mining, and data mining applications in electronic commerce, bioinformatics, computer security, Web intelligence, intelligent learning database systems, finance, marketing, healthcare, telecommunications, and other fields.

——————————-

ICTAI'02
The Fourteenth IEEE International Conference on Tools with Artificial Intelligence
Washington D.C., USA
November 4-6, 2002
http://www.nvc.cs.vt.edu:8086/

The annual ICTAI conference is an international forum for the exchange of ideas relating to artificial intelligence (AI) among academia, industry, and government agencies. It fosters the creation and transfer of such ideas, and promotes their cross-fertilization over all AI application domains and AI paradigms through a unifying theme: AI Tools. ICTAI focuses on both theory and developing, implementing, and evaluating theoretical and applied frameworks that may serve as tools for developing intelligent systems and pursuing AI applications.

We invite paper submissions that include but are not limited to the following topics: AI algorithms, AI in software engineering, cognitive modeling, collaborative software agents, fuzzy logic and reasoning under uncertainty, genetic algorithms, intelligent tutoring/training systems, intelligent internet agents, intelligent interface agents, environmental applications, AI in data mining, intelligent information retrieval, AI in database, knowledge sharing, machine learning, natural language processing and speech understanding, neural networks, planning and scheduling, qualitative reasoning, vision and image processing, AI in multimedia systems, hybrid intelligent systems, and AI in real-time applications.

Other Computational Intelligence Conferences

AAMAS'03
The Second International Joint Conference on Autonomous Agents and Multi-Agent Systems
Melbourne, Australia
July 14-18, 2003
http://www.aamas-conference.org

Agents are one of the most prominent and attractive technologies in computer science at the beginning of the new milennium. The technologies, methods, and theories of agents and multiagent systems are currently contributing to many diverse domains such as information retrieval, user interfaces, electronic commerce, robotics, computer mediated collaboration, computer games, education and training, ubiquitous computing, and social simulation. They not only are a very promising technology, but are also emerging as a new way of thinking, a conceptual paradigm for analyzing problems and for designing systems, for dealing with complexity, distribution, and interactivity, while providing a new perspective on computing and intelligence. The AAMAS conferences aim to bring together the world's researchers active in this important, vibrant, and rapidly growing field.

The AAMAS conference series was initiated in 2002 as a merger of three highly respected individual conferences (ICMAS, AGENTS, and ATAL). The aim of the joint conference is to provide a single, high-profile, internationally renowned forum for research in the theory and practice of autonomous agents and multiagent systems.

——————————-

IDEAL'03
The Fourth International Conference on Intelligent Data Engineering and Automated Learning
Hong Kong, China
March 21-23, 2003
http://www.comp.hkbu.edu.hk/~ideal03/
Submission Deadline: Dec 31, 2002

As a biennial conference, IDEAL was firstly launched from Hong Kong in 1998, which is dedicated to emerging and challenging topics in intelligent data analysis and engineering and their associated learning paradigms. Following the past highly successful IDEAL'98, IDEAL'00 and IDEAL'02, IDEAL will become an annual conference starting from 2003. The fourth conference will aim to provide a forum for researchers and engineers from academia and industry to meet and to exchange ideas on the latest developments in data engineering and learning, and to share their successes. IDEAL is an ideal forum for revealing and developing the latest theoretical advances and practical applications in intelligent data engineering and automated learning.

Topics of interest include statistical learning, data mining, agents, bioinformatics, multimedia information and financial engineering

---

AWIC'03
The 2003 Atlantic Web Intelligence Conference
Madrid, Spain
5-7 May 2003
http://nova.ls.fi.upm.es/~ernes/webconference/AWIC03.htm
Submission Deadline: Jan 10, 2003

The 2003 Atlantic Web Intelligence Conference brings together scientists, engineers, computer users, and students to exchange and share their experiences, new ideas, and research results about all aspects (theory, applications and tools) of Artificial Intelligence Techniques applied to Web Based Systems, and discuss the practical challenges encountered and the solutions adopted.

The conference will cover a broad set of Artificial Intelligence Techniques, such as (but not restricted to): agents, artificial life, case based reasoning, fuzzy logic, genetic algorithms, genetic programming , knowledge based systems, multivalued logic, neural networks, ontologies, probabilistic and evidential Reasoning, reinforcement learning, rough sets, and semantic networks; applied to the corresponding problem(s) of a Web based system such as (but not restricted to): art of web design, automatic Web content design and production, automatic Web page generation, conversational systems, customer/user profiling and clustering, electronic marketplaces, e-mail and SMS filtering and production, information retrieval, negotiation systems, price dynamics and pricing algorithms, purchasing behaviour, recommender systems, remembrance systems, security, privacy, integrity and trust, virtual hosts and virtual agents, Web browsing and exploration, Web merchandising, marketing and publishing, Web personalisation, and Web-log mining.

IJCAI'03
The Eighteenth International Joint Conference on Artificial Intelligence
Acapulco, Mexico
August 9-15, 2003
http://www.ijcai-03.org/1024/index.html
Submission Deadline: Jan 14, 2003

The IJCAI-03 Program Committee invites submissions of full technical papers for IJCAI-03, to be held in Acapulco, Mexico, August 9-15, 2003. Submissions are invited on substantial, original, and previously unpublished research on all aspects of Artificial Intelligence. Refer to the web site for the list of related keywords.

ISMIS 2003
The Fourteenth International Symposium on Methodologies for Intelligent Systems
Maebashi TERRSA, Maebashi City, Japan
Oct 28-31, 2003
http://www.wi-lab.com/ismis03/
Submission Deadline: March 10, 2003

This Symposium is intended to attract individuals who are actively engaged both in theoretical and practical aspects of intelligent systems. The goal is to provide a platform for a useful exchange between theoreticians and practitioners, and to foster the cross-fertilization of ideas in the following areas: active media human-computer interaction, autonomic and evolutionary computation, intelligent agent technology, intelligent information retrieval, intelligent information systems, knowledge representation and integration, knowledge discovery and data mining, logic for artificial intelligence, soft computing, and Web intelligence. In addition, we solicit papers dealing with Applications of Intelligent Systems in complex/novel domains, e.g. human genome, global change, manufacturing, health care, etc.

INMIC 2002
International Multi-topic Conference
Karachi, Pakistan
Dec 27-28, 2002
http://ewh.ieee.org/r10/karachi/inmic2002/

INMIC (International Multi-topic Conference) has been annually organized by the IEEE Pakistan for last many years. In 2002, the IEEE Karachi is sponsoring the conference with KIIT (Karachi Institute of Information Technology) as a Co-Sponsor with the Technical Co-sponsorships of Computer Society & Communications Society Karachi Chapters. Scope and areas of the conference include: databases, Internet Systems & Security, visualization and graphics, knowledge applications, agent technologies, e-commerce, distributed systems, electronic data interchange, software engineering, human computer interaction, pattern recognition, speech recognition & synthesis, computer vision, high performance computing, numerical computing, digital signal processing, wireless communications, bio informatics.

Career Opportunities

Dorothean Professor in Computer Science at the University of Vermont
http://www.cs.uvm.edu/jobs/cssearch.shtml
Apply by January 20, 2003

——————————-

Assistant / Associate Professors in Computer Science at Hong Kong Baptist University
http://www.comp.hkbu.edu.hk/temp/Ad-2003-COMP-AP.txt
Apply by February 28, 2003

——————————-

Two tenure-track Assistant / Associate Professor positions in Computer Science at the University of Central Arkansas
http://www.cs.uca.edu/ad.htm
Application review will begin on December 1, 2002 and continue until positions are filled.

IEEE Computer Society
1730 Massachusetts Ave, NW
Washington, D.C. 20036-1903

Non-profit Org.
U.S. Postage
PAID
Silver Spring, MD
Permit 1398