Foraging for Information Resources in Cyberspace: Intelligent Foraging Agent in a Distributed Network

George E. Mobus and Manuel Aparicio IV

Abstract

Recent advances in autonomous robotic control may be applied to the problem of designing intelligent, mobile agents for cyberspace. This paper examines the field of behavior-based reactive systems (BBRS) as used in robotics that may be applicable in constructing intelligent agents. The problem of search in an unstructured and dynamic, distributed environment has been explored in a physical robot. Two major aspects -- the use of chaos to generate stochastic path selection, and a unique associative learning mechanism that bounds the search space on subsequent runs -- have been shown to be an efficient strategy for the robot in exploring a nonstationary environment. The role of chaos is to produce novel, yet constrained paths. The robot learns from experience that certain associations will increase its chances of finding mission-critical events, while others will inhibit it. On subsequent forays into the environment, the robot uses this knowledge to improve its discovery process. This paper describes the application of these mechanisms to an intelligent agent in search of information resources in a distributed, dynamic, and unstructured computer environment. We present a search for keywords, where the agent learns the node/directory/filename cues that will increase its chances of finding documents containing those keywords.

1.0 Introduction

Animals that search for their food or other resource have been observed to follow a seemingly random course until the search object is detected. This search-and-acquire behavior is termed "foraging", and it has been studied considerably for a wide variety of species, including humans. Koshland [4] has studied a stochastic search mechanism in chemotactic bacteria, with implications that stochastic mechanisms underlie many animal search behaviors. It is an open issue as to what form of indeterminacy the underlying process follows. In the case of bacteria in a relatively food-rich environment, the search may indeed be governed by a purely random or white noise process, but other possibilities exist. The process may involve colored noise or a 1/f distribution. Another line of research suggests that an ordered process such as a Markov process may govern the dynamics of observable behaviors.

It may well be that every form of stochastic process is represented based on the density of the resource in the search space. What is clear in all of the examples of foraging that we have studied is that some form of indeterminancy is prevalent while the organism is not being stimulated by a specific semantic gradient. This suggests that a stochastic decision generator is ubiquitous in nature. It appears to have been conserved across the phylogenetic spectrum, and so must be extremely fundamental. In this paper, we introduce an approach to the search for and acquisition of information in a distributed, dynamic, and unstructured network environment, that is based on the biological process of foraging. There are a number of elements in information search and retrieval that are held in common with foraging behavior. This represents a specific implementation of a BBRS-based approach to knowbotic agents, but it may also have more general applicability to search problems in computer science. The method is called edge-of-chaos (EOC) Search, for reasons given below.

1.1 Finding Information in Cyberspace

"Cyberspace" is a term that has caught on in both the popular press and in technical work. Loosely, it refers to the emerging environment within a computer network such as Internet; an environment of communication and storage. Information is created, stored, and retrieved at an unprecedented level of efficiency.

We use the term "agent" here to refer to a software component -- a task running in a distributed environment -- used to facilitate the access and interactions of a user with resources in the cyberspace. An "autonomous agent" is a task that makes judgement decisons without direct supervision of the user, and a "mobile autonomous agent" is a task that can be run on any platform in the network and has the ability to port itself across the network for the purpose of obtaining resources in a remote node.

We are interested in the task of a mobile autonomous agent, that is given a basic description of what constitutes information to its user, and then searches through cyberspace for instances of that information. We make a basic set of assumptions about the distribution, dynamics, and cost of information in the space. We then advance a model of foraging that appears to meet the requirements generated from these assumptions. We describe the basic formalism that can be used to analyze the model and report on some initial results.

1.2 Structured vs. Unstructured Locations

The problem of search in a computational setting is often characterized in a static, structured framework. For example, one might be searching for a keyword in a list of keywords from an abstract. The abstract is indexed to an article, and the problem of finding information is that of finding the indices to all articles whose abstracts contain the appropriate keywords. Nothing changes from one search to the next, so if one search has produced a set of indices on one set of keywords, that set may be retained for future inclusion in a superset if any new keywords are added to the original search criteria. The organization of the environment is highly structured. There is a given set of data structures and procedures that may be routinely followed to acquire the sought information.

In contrast, picture an environment such as Internet. At numerous sites, individuals are producing information that might be useful to many other individuals across the net. Assuming that this information is not particularly proprietary (for example, putting text files of recently published articles in an archive), it would be valuable if others could have read access to it. The conditions of wanting information and creating information are completely asynchronous events. Furthermore, there is no established channel for notifying prospective users of the existence of such information. The events are dynamic -files may be moved to different subdirectories in seemingly random fashion; the environment is unstructured in that directory paths and names follow no global convention.

While this would at first seem to preclude the environment ever being understood, just because it is unstructured, dynamic and indeterminate does not necessarily mean that it is totally without order. The field of chaos theory provides some possible insights into how such an environment can, in fact, be organized, even if seemingly random. The key is in understanding the role of probablistic causality in such systems. For example, document authors often give files names that are relevant to their content. Hence, if one learns that certain strings used in filenames have been associated with the contents of interest, even when that string is not part of a keyword, one may use the occurrence of that string in some other file name as a possible indicator that the file contains desired information.

Several mechanisms for coordinating information producers with information users have arisen in the Internet environment. Usenet is often used to announce the existence of anonymous ftp sites; Gopher is rapidly becoming a popular system for browsing remote locations; World Wide Web (WWW) provides hypertext access to many valuable resources. The problem with these mechanisms is that they are formal methods and require a high level of structure (standards) imposed on both producers and con-When a producer knows that sumers. someone needs the information he/she has produced and the consumers know where to look to find that information, everything works well.

But what of the unstructured, uncoordinated occurrences that doubtless go on in the Internet environment? In the vernacular of a phrase voiced by someone in Usenet, "So many ftp sites to check out, so little time...". It is very likely that there are many more producers and consumers on the net than will ever be brought together by formal, structured mechanisms. In short, the net is simply too dynamic to accommodate everyone.

With this background, the notion of an autonomous mobile agent with the ability to search through a huge network and report back with potential, if not actual, information sources has been advanced. Such a notion seems a natural outgrowth of an individual's efforts to browse in Internet. However, it is not so clear that limitless browsing will be permitted in the information superhighway of the not-too-distant future. The reason is that all information transactions, even just the search itself, have real costs associated with To date, these costs have been them. absorbed by various funding means for Internet users, but this is largely an historical artifact of Internet's origins and is changing.

1.3 Search Costs

In reality, someone is paying for network access and interactions. In the future, as commercial enterprises invade the net, the reality of these costs will begin to strike everyone directly.

Thus, any agent that will operate in the cyberspatial environment must do so with certain cost constraints added to its operation. In this way, the search for information in cyberspace will become, very much, a problem in optimization.

Given the dynamics of information creation, its indeterminate distribution in the network, and the costs that are associated with finding it, we are motivated to find models of search and acquisition that may accommodate this type of environment. The foraging of food by some species of animals in the biological domain may provide such a model.

1.4 Biologically Motivated Search: An Application in Autonomous Agents

Foraging is a behavior that seems well matched to the exigencies of a dynamic, uncertain environment where the cost of search is explicit. An animal must find a food source in a finite time at a cost of energy expended that is less than the energy extracted from the food.

In the last decade there has been an increasing interest in biologically motivated computational processes that solve problems resembling those solved by animals. The field of "soft" computing, a term coined by Lotfi Zadeh, embodies attempts to emulate natural processes such as evolution (genetic algorithms), brain-like information processing (neural networks), and several other life-like processes (artificial life). The purpose of this endeavor is to understand the computational properties of these processes in order, on at least one hand, to solve complex problems that are not tractable in conventional algorithmic computations.

We are similarly motivated to use foraging behavior as a model of searching in the kind of environment envisioned above. Thus we look at autonomous mobile agents that forage the cyberspatial environment for information "food". They are given a budget that constrains their search. But also, they are given the ability to learn the causally linked attributes of objects in cyberspace that provide cues to the occurrence of the sought information. Our agent is able to learn the dynamics of information production and storage, and conduct efficient searches for future occurrences of that information.

1.5 Intelligent Agents, Knowbots, and Foraging

In a sense, the notion of an intelligent agent is a generalization of the notion of a mobile autonomous robot. In fact, the starting point for the current effort was the successful demonstration of such a robot's foraging, searching, and learning to associate sensory-based cues with the occurrence of a sought object [6]. In each subsequent search, the robot improved its ability to find the sought object when the latter was being moved about in a stochastic manner.

The logical extension of a robot into a cyberspatial environment is what we call a knowbot. The specific purpose of a knowbot is to search in cyberspace for a desired informational resource within a specified time and budget (measured in terms of numbers of node jumps and sensory probes, if not directly monetary). The method of search to be employed is foraging.

2.0 Edge-of-chaos Search

Langton has described a "zone" of behavior on a suitably chosen parameter space between an ordered, periodic dynamic and a chaotic dynamic characterized by strange attractors and sensitivity to initial conditions [5]. This intermediate phase between order and chaos he calls the "edge of chaos" and suggests that it is where life is organized. This region has a number of interesting properties that lend credence to such an argument. Namely, it harbors behaviors that are complex, stochastic, and yet locally ordered. Further, Langton argues that this zone is dominated by the dynamics of information processes. Hence, one might further argue that such a regime would possess important computational properties.

In a similar vein, Bak [1] describes the dynamics of "self-organized criticality" as a form of weak or quasi-chaos. This dynamic, too, is characterized by elements of nonpredictability of detail, yet having some inherent structure. Several lines of research suggest that a transition zone between fullblown chaotic behavior and strictly organized periodic behavior has the qualitatively right mixture of novelty and organization to warrant its investigation as a fundamental organizing principle in the study of life.

We speculate that, in general, a good source of stochastic behavior for a search mechanism is a chaotic or quasi-chaotic one. Under assumptions that a suitable and controllable parameter is available to the searcher, one can easily picture a search mechanism that is tunable between the extremes of a fixed-point attractor and one that fills the phase space.

2.1 Chaotic Neural Oscillator

In the course of constructing a neural network circuit that could function as an oscillator that would emulate the actions of a central pattern generator (CPG) for motion control, Mobus and Fisher [7] discovered that the circuit's output did not follow a strictly periodic course. The output was sinsoidal in form, but the peak amplitude and cycle length of the wave form were erratic (Figure 1). We undertook a cursory analysis of the data to see if it showed some of the signatures of chaos. Analysis of over ten thousand data points was done for attractor reconstruction and spectral distribution. Figure 2 shows a typical x-y plot of an attractor reconstructed with a tau of 10. Clearly, from the reconstruction it would appear that if the attractor is real, then it is also strange.

The purpose of the CPG circuit in the robot brain is to cause the robot to weave back and forth as it moves in a generally forward direction. This causes the robot to sample a wider swatch of space than would be possible with simple forward motion. The introduction of a chaotic component into the generated sine wave turned out to be a serendipitous find. First, it showed that the linear adaptrode model, when used in a recurrent circuit, could generate highly nonlinear behavior. Second, the stochastic component introduces noise into the search mechanism in a manner that may be analogous to simulated annealing in gradient descent searches. This latter factor may have significant implications in the general search problem domain when the environment is nonstationary and, for example, the energy surface of the domain is constantly shifting. In such a situation, escaping local minima would be an ongoing problem.

2.2 Graph Search

In order to formalize the notion of a chaotically driven search, we are beginning to frame the problem in the context of graph theory. This approach not only establishes a theoretical basis for analyzing the method; it is also directly applicable to the problem of foraging in a distributed network, since the latter may be characterized as a graph.



Figure 1. Central pattern generator (CPG) oscillatory output

This work is just starting, so in this paper we report only some preliminary results in formulation and relate this to the problem of foraging for information in a computer network. The description below was developed partly in collaboration with Arun Jagota at Memphis State University.

2.2.1 Description

Consider a graph, G = (V, E) where V is the set of vertices and E is the set of weighted edges. G is bi-directed, with all pairs of edges having the same weight, and there are no negative weights on the edges in E. We investigate the problem of the single-pair shortest path (SPSP) as an optimization problem in G. This problem can be solved by a singlesource shortest path algorithm in O($|V|^2 +$ |E|) time, and no asymptotically faster method is known.

The SPSP problem can also be reformulated as a search problem on integer-valued weighted edges by transforming the graph with "extra" verticies such that each new edge has a unit weight. The search can then be conducted in a breadth-first manner in $O(|V^*| + |E^*|)$ time, with the recognition that the number of verticies and edges in the transformed graph is a function of the size of weights in the non-transformed graph. Thus the solution from a breadth-first search is not expected, in general, to be better than the single-source algorithm.

As with the traditional form of this problem, a search consists of starting from a distinguished vertex, s, and proceeds by some method until a destination vertex containing the sought object is found. In a static formulation of this problem, the destination vertex is fixed. We are interested, however, in finding an object on a vertex where the object moves among some subset of vertices, V', by a stochastic dynamic between runs. The movement of the object from vertex to vertex is slow compared to the solution time. What distinguishes this formulation is that the objective is to minimize the sum of shortest paths over a set of runs. We seek to bound its performance between a pure stochastic search (on a randomized graph) that is known to have the worst-case performance, and an algorithmic solution such as Dijkstra's algorithm on the sequence of graphs.

One further consideration in this problem needs to be characterized. The stochastic process that moves the search object among the subset of verticies as described above is not necessarily purely random. On the contrary, we assume that there exists an underlying causality that proscribes associative relations between observables. Furthermore,



Figure 2. Attractor reconstruction from free-running oscillators

the laws of causality require a specific temporal ordering among observables such that one observable reliably (in the sense of correlation) precedes another observable. In terms of our graph formulation, we consider a specific subgraph of G, G', which establishes a relationship between objects that are observable and that would be found in verticies nearby the target vertex, which contains the sought object. This subgraph may itself be generated by some organized, stochastic process, in which case the relations are statistical. By the same sort of dynamic, in each run, G' is fitted onto some vertex subset in G. Thus the search object moves around in G along with some other objects

that are related causally to the search object but are not themselves sought. Figure 3 shows a sequence of simplified graphs in which the search object and its causally associated objects are moving about.

Our purpose is to investigate the use of chaos as a search tool, compared to the random walk or white noise processes used in pure stochastic search. Further, we wish to investigate the use of associative learning in improving the performance of the search over the set of graphs. We conjecture that the edge-of-chaos search method with associative learning will be shown to provide performance improvements over algorithmic solutions.



Figure 3. Graph in which search object is moving about

2.2.2 Algorithm

Each vertex in G represents a spatial coordinate. The edges incident on a vertex, v, are labeled according to a global coordinate system. For example, each edge could be labeled by degrees from an arbitrary north. In the simple Euclidian interpretation, each edge would be labeled at one vertex in complement to its label at the adjacent vertex. This labeling would be part of a network initialization, so would not affect the time complexity of subsequent searches.

At initialization, a vertex is chosen as home. This will be the start vertex for a set of subsequent searches. Given an integer k, we will iterate the search k times and sum the path lengths on each search. The search method with the lowest total path sum wins the competition. In the following pseudocode, we assume that all vertices are reachable from s.

```
1 Sum_of_paths <- 0
2 For i <- 1 to k do
    In each search, starting at s, set
3
    the path_counter to 0
4
    do forever
      check the local vertex for the search object
      if found exit the loop
      else select some edge to explore via
      procedure M and traverse it
      incrementing the path counter
8
    end do loop
    Sum_of_paths <- Sum_of_paths + path_counter</pre>
10 end for loop
11 return Sum_of_paths for procedure M
```

The actual search in the inner loop (lines 4-8) could be replaced by a suitably modified breadth-first search or Dijkstra's algorithm wherein the loop terminates on finding the search object. Clearly, the method of line 7 is the focus of this approach.

In a more general vein, the number of sequential searches, k, could be set to infinity, in which case the relevant problem is to show that the path cost per search converges to an optimal value.

2.3 Internal Representation and Learning

The quasi-chaotic oscillator driving this search can be driven to a focused search under the influence of learning. For purposes of this paper, which is to introduce these considerations rather than completely explicate them, we will only briefly discuss the issues of representation and learning. A much more detailed account is given in Mobus [6], but Figure 5 and Figure 4 demonstrate the search for a moving object in physical space, without and with associative learning, respectively. Once an associative cue is found, the agent is shown to follow the cue to the mission-critical event.

A number of approaches to representing causal relations have been developed [2,6,8]. A key issue in this problem is the constraints placed on causal encoding by the multiresolution temporal attributes of nondeterministic causality. Briefly, these constraints involve the phasic relationship between episodic and sporadic events. Time must be explicitly represented in the system in order to capture these constraints in the context of on-line computation. Few attempts have been made to incorporate these requirements in strictly symbolic representation systems.

A neural representation using Adaptrode encoding has been shown to capture both the statistical properties the one-way and temporal ordering constraint of causality [6]. A new neural network architecture that does not use distributed representation as it is normally interpreted [3]. is employed to capture representations of the causal associations found in the input vector stream. A hierarchy of these networks can represent levels of more abstract relationships and, in fact, conveys the expressive power of conceptual dependency.

2.4 Discussion

The problem of foraging for food has provided a biological motivation for a more generalized search problem. We are beginning to formulate this problem in terms of a graph theoretic formalism that will make it accessible to analysis and application. Couched in the language of optimization problems, we see the relationship with animals that need to find food before their store of energy is exhausted. Intuitively, we can see that the success of a simple breadth-first search would be subject to the statistical properties of the underlying dynamic, which, if the process is nonstationary, would be unreliable as a source of information. Contrarily, a pure random search, while guaranteed to find an



Figure 4. Edge-of-chaos search without an associative cue

object, even a sparsely distributed one, will do so only in the limit in infinite time. Neither of these is of practical value.

We have conjectured that foraging consists of a stochastic process that is organized yet novel. Further, the process is controllable in the sense that a driving parameter that is under the control of the searcher can be used to vary the intensity of indeterminism in behavior. Thus the search may run the gamut from purely ordered to purely random. Coupled with an appropriate learning mechanism that can exploit causally ordered relations between objects having semantic value and others that do not, but can be used as predictors and/or measures of a gradient, this search mechanism may prove to be optimal for the class of dynamic search problems posed.



Figure 5. Edge-of-chaos search with an associative cue

3.0 Example of a Foraging Knowbotic Agent

In this section we provide a short example of a foraging agent that is being developed at the IBM Westlake Lab. The objective of the agent is to search files in a network, looking for occurrences of one or more keywords in text files. The keywords are themselves indicators of the information content of the file. The agent's job is to record the path to any file that contains these keywords. Additionally, however, the agent must learn the associative quality of substrings in the directory/filename structure that might be used as cues to other files of interest.

As an example, consider a file such as this document. It is not uncommon for researchers to name paper files such as:

mobus.aparicio.foraging.txt

The name of the file is related to its contents. Let's say a potential information consumer is interested in autonomous agents, and has supplied her agent with the keywords "autonomous" and "agent". Suppose that the agent comes across the above named file, and through the quasi-chaotic selector process decides to search it for the occurrence of either or both of these keywords. The file will be read, using the same chaotic selector to sample words or phrases. Since both authors seem to like using the word "agent", it is likely that the agent will find at least one occurrence of that word.

Such a finding constitutes a reward stimulus or reinforcement to the Adaptrode learning algorithm. But what is the agent learning? Prior to reading the file, the agent bound the substrings "mobus", "aparicio" and "foraging" to neural representations [6]. Each of these is active over the course of sampling the file. With the occurrence of the reinforcement signal from finding "agent", the connection weights to the bound neurons become potentiated, leading to a semipermanent memory trace of the association. The path leading to this file is explicitly recorded (like a fact) for later retrieval, but this is not the associative aspect we are interested in.

On returning to home base, the agent reports to its user that the keyword was encountered, and regurgitates the path. The user can then use a Gopher or anonymous ftp to check out the file. If the file is worthy of retrieval, the user rewards her agent, and this leads to a long-term reinforcement of the bound neural representations.

Now on subsequent forays into cyberspace, the agent happens on the filename "mobus.edge.of.chaos.search.txt". It will recognize the substring "mobus" as being associated with the keyword "agent". This may be a strong inducement for the agent to conduct a more thorough examination of the file to see that it does, indeed, contain the keyword "agent". On returning home, the agent reports finding another file that may be of interest to the user.

The key here is that the association caused the agent to switch from a purely stochastic search to a more directed search. The same principle applies to directory name substrings. For example, should the agent encounter a directory such as "mobus/chaos/papers", it would suspect that more information on agents would be found there. Instead of stochastically picking subdirectories to search, it would preferentially search this particular subdirectory.

Thus an agent would be able to evolve its own set of heuristics for guiding future searches in the course of conducting a more stochastic search. Contrarywise, if either the short-term reinforcement from finding the keyword in the searched file or the longerterm reinforcement from the user's finding the file of value did not obtain, the binding of the conditionable substrings to the unconditioned substring, "agent", would decay, and the agent would no longer be compelled to more exhaustively search a file with the name "mobus.really.good.stuff.txt" just because the substring "mobus" was encountered.

3.1 Conclusions

In this paper we have presented some preliminary findings with respect to the application of a biologically motivated search strategy to the field of autonomous mobile agents. Further motivation for this approach is garnered from the successes of using edgeof-chaos search in a mobile robot. We have suggested that the latter may be generalized in the framework of cyberspatial characterization to knowbotics, and have presented some of the relevant issues involved in formulating this mapping.

A central project in this thesis will be a full formalization of the edge-of-chaos search mechanism in a graph theoretic framework. In this, we will attempt to formalize the notion of foraging for resources in a stochastic, dynamic, and unstructured (at the the micro level) environment. The applications of such an approach go beyond a foraging agent, since the problem of search is ubiquitous in computational problems.

The work reported above is, however, preliminary and exploratory. It remains to be shown that the successes enjoyed in the physical space of a robot will transfer readily to the logical space of a knowbot.

Acknowledgements

This work was spoorted by a grant from the IBM Westlake Lab, Software Solutions. The authors would like to thank Frank McKiel, Susan Harris, and Jim Nicol for sponsorship of this project, and Tzvi Raz and Linda Gaines for management of the CAS program between Westlake and the University of North Texas.¹

About the Authors

George Mobus recently received his PhD in Computer Science at the University of North Texas in Denton, TX. Dr. Mobus, while earning his doctorate, has been a fulltime Lecturer in the department and has taught systems, data structures, software engineering, and advanced programming courses. His research interests have focused on a new, neurally inspired learning mechanism called the Adaptrode. This research has included the construction of a mobile, autonomous robot that learns by experience. Dr. Mobus is the author or co-author of 15 technical papers and articles in refereed conferences and journals. His Internet address is mobus@ponder.cs.unt.edu.

Manuel Aparicio IV is an Advisory Programmer at IBM whose work combines his long interests in both neurocomputing and human factors. As a National Science Foundation Minority Fellow, he received his PhD in Psychology from the University of South Florida in 1988. His dissertation work on the computational neurophysiology of learning was published in the book, Motivation, Emotion, and Goal Direction in Neural Networks, and he was recently co-editor of the book, Knowledge Representation and Inference in Neural Networks. He recently was part of the development team for the IntelliAgent project, IBM's first commercial His Internet address agent. is aparici@msnvm1.vnet.ibm.com.

¹ The IBM contact for this paper is Manuel Aparicio, IBM Westlake Lab, 40-D2-02, 1 E. Kirkwood Blvd., Roanoke, TX 7626-0001.

References

- Bak, P., and Chen, K., "Self-organized criticality", *Scientific American*, January, 1991, pp. 46-53.
- [2] Galambos, J. A., Ableson, R. P., and Black, J. B., *Knowledge Structures*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.
- [3] Hinton, G. E., McClelland, J. L., and Rumelhart, D. E., "Distributed Representations", in *Parallel Distributed Processing*, Rumelhart, D. E., McClelland, J. L. (eds), Cambridge, MA: The MIT Press, 1986.
- [4] Koshland, D.E., Bacterial Chemotaxis As A Model Behavioral System, New York: Raven Press, 1980.

- [5] Langton, C. G., "Introduction", Artificial Life II: SFI Studies in the Sciences of Complexity, Proceedings Volume X, 1992, pp. 3-23.
- [6] Mobus, G.E., "Toward a theory for learning and representing causal inferences in neural networks", in *Neural Networks for Knowledge Representation and Inference*, Levine, D. L., and Aparicio, M. (eds), Hillsdale, NJ: Lawerence Earlbaum Associates, 1994.
- [7] Mobus, G.E., and Fisher, P.S., "A mobile autonomous robot for research in intelligent control", Technical Report CRPDC-93-12, Center for Research in Parallel and Distributed Computing, University of North Texas, Denton, 1993.
- [8] Pazzani, M. J., Creating A Memory of Causal Relationships, Hillsdale, NJ: Lawrence Erlbaum Associates, 1990.