Defence Research and Development Canada | Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Applying Information Foraging Theory to Military Intelligence Analysis: Project Summary

David J. Bryant
DRDC – Toronto Research Centre

Prepared for: Rachel Remington, Director of Transnational and Regional Intelligence, Canadian Forces Intelligence Command

Defence Research and Development Canada
Scientific Letter
DRDC-RDDC-2018-L050
May 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

Defence Research and Development Canada | Recherche et développement pour la défense Canada

DRDC | RDDC
technologysciencetechnologie

May 2018

DRDC-RDDC-2018-L050

Prepared for: Rachel Remington, Director of Transnational and Regional Intelligence, Canadian Forces Intelligence Command

Scientific Letter

Applying Information Foraging Theory to Military Intelligence Analysis: Project Summary

Background

This Scientific Letter responds to a request from Rachel Remington, Director of Transnational and Regional Intelligence, Canadian Forces Intelligence Command to the author to provide a summary of work completed under the Joint Intelligence Collection and Analysis Capability (JICAC) project to explore the application of Information Foraging Theory (IFT) to the military intelligence domain. Interest in IFT was motivated by the issue of information-overload faced by military intelligence analysts. Due to the rapid rate at which information technologies have evolved, analysts have access to ever-increasing amounts of information but limited time in which to find and process pertinent information for sense-making activities. IFT, as a theory of optimal search in the information space, has the potential to yield useful insights into training and support for analysts.

Project Deliverables

The work perfomed included a literature review examining the basic principles of IFT and their application to information search by intelligence analysts (Bryant, 2014). Another report examined the concept of optimal stopping and described models of several types of information search tasks (Bryant, 2017a). Heuristic stopping rules were identified as potential methods to guide analysts' information search. Research also examined whether humans naturally exhibit adaptive information search behaviour in a simulated intelligence analysis task (Bryant & Li, 2016; Bryant, 2017b). A PowerPoint™ presentation was created as an introduction to IFT in the context of intelligence analysis (Bryant, 2017).

The results of this work demonstrated that IFT addresses the core issues of information search. Specifically, IFT lays out models to explain how an information forager should determine the kinds of information to pursue ("diet" models), how to allocate attention and time to different sources of information ("patch-leaving" models), and how to employ semantic cues to improve the efficacy of searches ("information scent"). The principles of IFT were also found to be well-establised, based on extensive research on optimal foraging theory and studied in the context of both lab-based experimentation and practical, real-world information search tasks. The most significant practical implications of IFT have accrued in the realm of digital information systems, including the World Wide Web (WWW). Here, IFT has inspired improved designs for information systems that reduce costs of information search and led to development of various tools that augment a user's ability to search (e.g., Fu, 2012; Shrinivasan & van Wijk, 2008).

Impact of IFT

There are two levels at which to consider the potential of IFT to promote better support for intelligence analysis: a) a broad level of general principles for effective information foraging, and b) a detailed level of procedures by which information foragers can optimize their information foraging performance within specific task environments.

At the broad level, IFT's description of information foraging as an activity aimed at maximizing obtained value in relation to cost, using models of diet selection, patch-leaving, and scent-following, lead to practical suggestions concerning how to make information foraging easier and more profitable. These suggestions centre on designing information systems in ways that reduce costs associated with search and exploitation (e.g., organizing information in patches by conceptual relatedness rather than physical media). As noted, many such efforts have already been successful.

At the detailed level, IFT provides methods for computing truly optimal information performance but only in situations in which precise and accurate meta-data concerning the information environment and task (e.g., the value of each item of information, times to perform search and exploitation activities, both short- and long-term gain functions) are available, and the forager has the cognitive capability to process these data. Because it is extremely difficult to assess all the necessary meta-data, or provide sufficient information processing support to compute optimal behaviors in real-time, the detailed level of IFT does not provide easy solutions to improving information foraging performance.

Advanced intelligence production paradigms

The value of IFT at the broad level can be seen in the theory's relation to advanced paradigms for intelligence production developed by the United States intelligence community (Roy et al., 2017). Two of these initiatives are of particular interest: Activity-Based Intelligence (ABI) and Object-Based Production (OBP). ABI organizes data collection around activities, interactions, and properties associated with a specific entity, population, or area of interest. A variety of methods are employed to discover correlations and links about the target entity using multiple data sets. OBP organizes data collection around "objects" (people, facilities, units, events) that are tangible in space and time. The goal is to link data from disparate sources in a way that is readily understandable to analysts. Data linked to an object are readily available to anyone interested in that object, which is not the case when raw data are segregated in separate data sets, as that requires an analyst to systematically search for data multiple times and then manually correlate that data.

At a broad level, IFT fits well with the underlying concepts of ABI and OBP. ABI, for example, makes "data neutrality" a central principle, meaning that data from all intelligence domains are correlated early in the process. Thus, ABI reduces the separation of information by type or modality and, instead, focuses on the value of information to underlying events and objects. This establishes a common currency for the evaluation of information value, in line with a fundamental assumption of IFT. ABI also relies on georeferencing to correlate multiple sources of intelligence data across time and space, creating an external information space in which information foraging can take place.

OBP is about organizing data from multiple data sets in a way that is both logical and readily sharable. In this fashion, OBP promotes the organization of data into intuitive patches that are easier to navigate, with relevant information associated with objects rather than a particular intelligence type. Organizing the information space in terms of objects and events makes it easier for analysts to make decisions about the value of information and how long to search in a given patch. According to IFT, a simple information space allows for more effective foraging.

The goals of the advanced paradigms (Roy et al., 2017) also align closely with the principles of IFT. First, these paradigms seek to shift the focus of analysis from documenting what is known—the locations and status of known entities—to discovery of what is unknown. This makes intelligence analysis more explicitly an information foraging activity. Rather than compiling data, an analyst actively searches through data for information to guide hypothesis generation and pattern discovery.

A second goal of these paradigms is to harness the potential of so-called "big data" (Roy et al., 2017). ABI explicitly relies on the integration of large, multi-domain data sets. Indeed, Roy et al. (2017) remark that "With ABI, analysts will "live" within the ocean of data and use naturalistic methods ...", which highlights the similarity of ABI's core concept to the view in IFT of humans as "informavores" constantly seeking and consuming information in a vast information space. If ABI principles are accepted in the intelligence community, the clear implication is that the information space of analysts will become increasingly large, requiring a more optimized foraging process.

Overall, the new paradigms of intelligence production are about changing the traditional intelligence cycle (Roy et al., 2017). These paradigms seek to replace the linear and stovepipe approaches in which information is integrated, often by different analysts than those who collected the data, late in the process, hurting overall understanding. But in ABI methods, analysts have access to raw data not just written summaries.

The ABI and OBP paradigms put greater emphasis on flexibility in data collection, allowing discovery of patterns of activity to drive collection requirements. This leads to a more iterative and adaptive collection process, in line with the principles of IFT. Assessing the value of each piece of data demands that the information be related to the analysis questions and goals of the analyst. Research has demonstrated that an individual's comprehension and retention of information are dramatically enhanced when the individual is actively involved in the search and evaluation of raw data (e.g., Rieh et al, 2016).

The way ahead

IFT describes normative procedures for information collection, which bear on a fundamental problem facing analysts. However, IFT is primarily a descriptive approach to identifying optimal search behaviours within a given information environment, rather than a cognitive model. IFT illuminates the conditions the forager must fulfill to achieve optimized performance; it does not specify the decision processes needed to achieve those conditions. This limits the usefulness of IFT at the detailed level of application.

It may be possible to prescribe optimal behavior in real-time using models of patch-leaving strategies in the context of common information search tasks (Bryant, 2017a). The many obstacles to implementing a system to guide real-time optimal foraging, however, must be carefully considered. First, in addition to substantial computation, determining optimal behavior requires a great deal of meta-data concerning the information environment and forager's foraging history, as well as assumptions concerning the definition of information value that can be difficult to operationalize. Second, although a forager's behavior can often be tracked when using an information system, critical features of the information environment may be undefined. IFT requires a precise specification of environmental factors, such as the distributions of resource items within and across patches, the rates of encounter of information patches, and so on. These are generally not obvious and require work to uncover. In the context of real-time user assistance, the time and effort of describing the information environment—which can radically change as a forager's task requirements and goals change —render environment factors opaque. Finally, although tracking meta-data pertaining to the forager and the information environment is possible in the context of computer-based information systems, such data are not always available. In any event, computer-based information systems, although commonly used and increasingly important to intelligence analysis, do not define the entirety of information gathering. Even when meta-data are available, determining optimal strategies require extensive computation with the aid of computers (Green, 1988). No human can be expected to perform, unassisted, the computations described by IFT to determine in real-time what choices to make in foraging.

A less rigorous, but more tractable, approach to support information foraging is to identify search strategies that are generally optimal and guide analysts to use these in their information gathering activities. Green (1988) demonstrated that the applicability of simple search heuristics depends on the way useful information is distributed across patches in the information space. For example, if the amount of useful information available in a given patch is described by a Poisson distribution, the Fixed-Time heuristic (forage in each patch for a fixed amount of time) will be the optimal decision rule. It is not known, however, how relevant information tends to be distributed across patches within information environments. Although many studies have examined the total amounts of information located throughout the WWW, these do not speak to the distribution of information useful to a given forager's information search. Furthermore, because each information search will have unique goals and context, it is unclear that there is common distribution governing where a forager will find useful information. Nevertheless, this remains an interesting question and, should a general rule of distribution be possible, the use of simple heuristics would make a lot of sense for analysts performing information search.

Of course, it remains an open question as to the extent that analysts can be expected, as a function of their normal information search behavior, to perform such tasks in an optimal, or near-optimal, fashion. The program of experimentation was intended to discover natural information gathering strategies and compare these to optimized models. The program, unfortunately, was not as successful as initially hoped for. Two experiments were completed that provided a very preliminary test of the hypothesis that humans generally conform to principles of optimal search when performing a simulated analysis task (Bryant, 2017b). Limited availability of analsysts as participants prevented a deeper exploration of the question. Research on this issue is sparse in the literature and somewhat contradictory. Nevertheless, indications do suggest that when information systems are designed in ways that make it easier to forage according to IFT principles (e.g., reduce navigation costs, make the patch structure of the information environment more apparent to the user, etc.), then people are more effective and efficient in their information searches (e.g., Athukorala et al., 2014; Wu et al., 2014). A shift in the intelligence analysis process toward the ABI and OBP paradigms would capitalize on this fact.

Although IFT is primarily descriptive, it nevertheless offers principles that can be useful within the overall context of information collection. Thus, at the broad level of application, IFT provides a wealth of concepts to enhance information gathering procedures and systems (Pirolli, 2009). A model that explains peoples' behavior in using information systems can be used to diagnose inefficiencies and develop better, more intuitively applicable, processes. Thus, the bulk of the practical value of IFT so far realized has come from work on the design of digital information systems, including tools for use on the WWW. Numerous researchers have attempted to apply IFT principles to the design of information systems, including interface designs and tools to augment users' ability to search optimally (Bryant, 2014). Some applications, such as the Lumberjack system, have been designed to harness the collection of user meta-data by web sites (web sites record user interactions in some way, such as links clicked, time spent on page, etc.) to predict information goals and generate guidance to the user to assist in locating relevant information (Chi et al., 2001; Chi et al., 2002). User monitoring systems can track the accuracy of navigation links or generate predictions concerning the value of navigation links based on implicit models of the user's information goals. Such IFT-inspired systems have been helpful to designers of information systems in other domains (Athukorala et al., 2014), and there is every reason to expect that development of information systems for intelligence analysts could likewise benefit.

Conclusion

IFT concepts should be brought into the design process associated with the development of information systems under the JICAC project. Systems should be designed with the explicit intent to reduce the costs of information search. Many examples of successful IFT-inspired design exists and these examples can be used to guide system and interface design.

A decision should be made concerning the level, broad versus detailed, at which to pursue further research on IFT. It certainly appears that the greatest impact of IFT on information gathering in the intelligence analysis domain will be realized at the broad level. Organizing information gathering along the lines put forward by emerging advanced intelligence production paradigms (Roy et al., 2017) would be a substantial step towards this goal. People working in the JICAC project should consider additional ways to restructure policies and procedures to facilitate effective information foraging within these paradigms. In contrast, there seems to be limited potential for an IFT-based information system at the detailed level. The technical issues with this approach are large and the potential benefits uncertain.

Prepared by: David J. Bryant (DRDC – Toronto Research Centre).

References

Athukorala, K., Oulasvirta, A., Glowacka, D., Vreeken, J., & Jacucci, G. (2014). Supporting Exploratory Search Through User Modeling. In UMAP Workshops.

Bryant, D.J. (2014). Information Foraging Theory: A framework for intelligence analysis. DRDC – Toronto Research Centre, Scientific Report, DRDC-RDDC-2014-R115.

Bryant, D.J. (2017a). Simple heuristics for guiding information search. DRDC – Toronto Research Centre, Scientific Report, DRDC-RDDC-2017-R119.

Bryant, D.J. (2017b). Experiments on information foraging. DRDC – Toronto Research Centre, Scientific Report, DRDC-RDDC-2017-R159.

Bryant, D.J. (2017). Information Foraging Theory: An Introduction for Intelligence Analysts: Slide deck and notes, DRDC – Toronto Research Centre, Reference Document, DRDC-RDDC-2018-D026.

Bryant, D.J. & Li, A. (2016). INformation FOraging Cognitive Analysis Tool (INFOCAT): An experimental platform for studying information foraging of intelligence analysts. DRDC – Toronto Research Centre, Scientific Report, DRDC-RDDC-2016-R022.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. SIGCHI'01, Seattle, WA.

Chi, E. H., Rosien, A., & Heer, J. (2002, July). LumberJack: Intelligent discovery and analysis of Web user traffic composition. Paper presented at the ACM-SIGKIDD Workshop on Web mining for usage patterns and user profiles, WebKDD, 2002, Edmonton, Canada.

Fu, W. T. (2012). Information Foraging on the Internet. In P. M. Todd, T. T. Hills, and T. W. Robbins (Eds.), Cognitive Search: Evolution, Algorithms, and the Brain (pp. 283–299). Cambridge, MA: MIT Press.

Green, R. F. (1988). Optimal foraging for patchily distributed prey: random search. Technical Report 88–2. University of Minnesota. Department of Mathematics and Statistics.

Pirolli, P. (2009). Information foraging theory: Adaptive interaction with information. New York, NY: Oxford University Press.

Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. Journal of Information Science, 42(1), 19–34.

Roy, J., Bergeron Guyard. A., & Kwantes, P. (2017). High-level assessment of advanced intelligence production paradigms and initiatives (X). DRDC Scientific Letter (DRDC-RDDC-2017-L025). Defence Research and Development Canada.

Shrinivasan, Y. B., & van Wijk, J. J. (2008, April). Supporting the analytical reasoning process in information visualization. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1237–1246). ACM.

Wu, W. C., Kelly, D., & Sud, A. (2014, July). Using information scent and need for cognition to understand online search behavior. In Proceedings of the 37th international ACM SIGIR conference on Research & Development in information retrieval (pp. 557–566). ACM.

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or a tasking agency, is entered in Section 8.) DRDC – Toronto Research Centre PO Box 2000, Toronto Ontario M3M 3B9, Canada	2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Applying Information Foraging Theory to Military Intelligence Analysis: Project Summary

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used. Use semi-colon as delimiter)

Bryant, D. J.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total cited in document.)
May 2018	7	0

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter)

Scientific Letter

8. SPONSORING CENTRE (The name and address of the department project or laboratory sponsoring the research and development.)

DRDC – Toronto Research Centre

PO Box 2000, Toronto Ontario M3M 3B9, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable contract number under which the document was written.)

10a. DRDC DOCUMENT NUMBER	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2018-L050	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

intelligence analysis; information foraging theory

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)