KNOWLEDGE CAPTURE TO SUPPORT INFORMATION FLOW MANAGEMENT IN COMPLEX SYSTEMS

Edward Szczerbicki
School of Engineering
The University of Newcastle
Newcastle NSW 2308, Australia
E-mail: Edward.Szczerbicki@newcastle.edu.au

KEYWORDS
Information management, Knowledge capture, Information flow modelling.

ABSTRACT

Systems become increasingly complex. Their decomposition into smaller units is the usual way to overcome the problem of complexity. This has historically led to the development of atomized structures consisting of a limited number of autonomous subsystems that decide about their own information input and output requirements, i.e. can be characterized by what is called an information closure. Autonomous subsystems can still be interrelated and embedded in larger systems, as autonomy and independence are not equivalent concepts. These ideas are gaining very strong interest in both academia and industry, and the atomized approach to information flow modelling and evaluation is an idea whose time has certainly come. This presentation discusses some modelling and evaluation issues, and challenges existing in the exciting area of knowledge capture for information flow managemen support for autonomous subsystems.

INTRODUCTION

Managing complex systems that function in changing and uncertain information-rich environments requires greater understanding and knowledge about the role of information in systems operation. To gain this understanding, an approach is needed that could be used to model and evaluate information flow in different situations. Such an approach is presented in this paper.

In fact, this paper goes well beyond the above in proposing an approach considering important practical issues of information flow, i.e. delays, incompleteness, imprecision and loss in value. The current practice of dealing with such issues are mostly when problems are detected and reactively. This situation may not be desirable and definitely be a major drawback for systems that more and more rely on the timeliness and quality of information for their operation. The proposed approach, in this respect, would greatly enhance the understanding of the various factors that influence the quality of information to the benefit of better decisions in adequate time, which in fact is the core of the philosophy behind any information system development.

System decomposition into smaller units is the usual way to overcome the problem of complexity. This has historically led to the development of atomised structures consisting of a limited number of *autonomous subsystems/agents* that decide about their own information input and output requirements, i.e. can be characterised by what is called an *information closure*. Autonomous subsystems/agents can still be interrelated and embedded in larger systems, as autonomy and independence are not equivalent concepts. These ideas are recently gaining very strong interest in both academia and industry, and the atomised approach to systems modelling, design and development is an idea whose time has certainly come (Morimoto 2001, Tharumarajah 1999, Prakken 2000, O'Grady 19999). The issues discussed in this paper will focus on information flow for autonomous subsystems/agents.

In a real-world industrial context, autonomous subsystems/agents consist of groups of people and/or machines tied by the flow of information both within a given subsystem and between this subsystem and its external environment (Szczerbicki 1996a). We will briefly present a modelling approach that could be used to evaluate such an information flow. The suggested approach allows for the evaluation of an information flow to be performed for different types of external and internal environments of a given system. The approach also accommodates the question of uncertain and imprecise information flow modelling.

ANALYTICAL MODELLING APPROACH

An autonomous subsystem/agent is usually functioning in the external environment which determines the decision-making process. Its knowledge could be described by the following:

Proceedings 18th European Simulation Multiconference
Graham Horton (c) SCS Europe, 2004
ISBN 3-936150-35-4 (book) / ISBN 3-936150-36-2 (CD)

(i) characteristic of the external environment (relationship between variables describing the environment and its dynamics),
(ii) characteristic of the internal environment, i.e. the relationship between the actions of the members of an agent,
(iii) the range of information about variables describing external environment.

The formal representation of the above knowledge is presented in this Section. For the knowledge extraction purposes, a general approach is needed that captures the whole of the behaviour of an agent. Such an approach, based on correlation between information and energy, is very briefly outlined next. Certain features implemented in previous research presented in (Szczerbicki 2000, 2002a, 2002b) are included for the sake of completeness.

Let A represent the set of possible actions which can be undertaken by the members of an agent, Z the set of corresponding consequences, and X random variables describing the actual state of the external environment. It can be assumed that:

$$z=f(a,x) \tag{1}$$

as the particular consequence (z) depends usually on an action (a) undertaken in the particular state of the environment (x). On the other hand, the decision about particular action depends on information that is available about the state of the environment. If ß stands for the decision function, we have

$$a=ß(d) \tag{2}$$

where d represents information.

For general description of the function f(a,x) let us consider certain correlation between information, action, and energy. Its theory is relatively young, but is has already been pointed out that in certain situations energy can be replaced by information and vice versa (matsumoto 19999, Bogdan 2000). This replacement is of statistical character and according to it for certain amount of information, say C_1, certain task can be performed using E_1 energy (Szczerbicki 1996a). Then, for a given C_1 there exists the best way (action A_{opt}) to fulfill the job, i.e. the action which uses E_1 energy. Actions different than A_{opt} result in more energy consumption. The above concept was presented in detail in (Szczerbicki 1996a, 2003) and used to arrive at the best decision functions $ß_i$. It can be shown that for n-person agent we have (Szczerbicki 1996a, 2003):

$$ß_i(d_i)+\sum_{j\neq i} qE[ß_j(d_j)\mid d_i=E(b_i\mid d_i) \tag{3}$$

where i, j = 1, 2, ..., n.

Formalization of agent decision making process expressed by (3) is a tool necessary for modelling and evaluation of information flow in an autonomous system. Information flow connects agent members with the external environment described by random variables X. The connection is represented by information structure. This structure is modelled by matrix C in which $c_{ij}=1$ if the ith member has obtained (either by observation or communication) information about the jth variable X realization (if $c_{ij}=0$ he/she has not got it). The ith variable X realization can be observed only by the ith member of the agent. He/she can be informed about other realizations only when communication (information exchange) inside the agent is organized. The value of information structure defined above is given by the following (Szczerbicki 2000):

$$VC=\min E[f(a, X)\mid C0]-\min E[f(a, X)\mid C], \tag{4}$$

where min $E[f(a, X)\mid C0]$ represents the utility of information structure C0 in which $c_{ij}=0$ for each i and j. Using (3) the VC can be represented by:

$$VC=E[b^T ß]. \tag{5}$$

With the modelling tools given by (3) and (5) one can easily extract knowledge about autonomous systems functioning in various decision situations. In Table 1 some samples of such a knowledge are specified for static environments [11]. This knowledge is easily codified and can be used in control, command, and management of autonomous systems. Similar rules can be easily captured for decision situations involving dynamic environments (please see [11] for details).

Table 1: Production rules describing agents functioning in static environment

RULE 11
IF an external environment of an autonomous agent is static,
AND it is described by random variables,
THEN the value of an information structure that represents the flow of information between the agent and its environment depends on interaction between agent members, correlation between random variables, and their variance.

RULE 12
IF an external environment of an autonomous agent is static,
AND it is described by a random variable,

THEN the value of information about this variable realization is proportional to the value of its variance.

RULE 13
IF an external environment of an autonomous agent is static,
AND it is described by random variables,
THEN full information has the value that is always greater or equal to the value of any other information structure.

RULE 14
IF an external environment of an autonomous agent is static,
AND there is no interaction in the internal environment,
THEN it is enough to restrict the information flow only to observation; organizing an information exchange does not improve the value of a resulting information structure.

RULE 15
IF an external environment of an autonomous agent is static,
AND there is an interaction in the internal environment,
AND the relationship between variables describing the external environment is of statistical character,
THEN information structure should include observation and communication.

RULE 16
IF an external environment of an autonomous agent is static,
AND the relationship between variables describing the external environment is given by function dependence,
THEN communication between agent members does not affect the value of information structure; information flow should be restricted to observation.

RULE 17
IF an external environment of an autonomous agent is static,
AND interaction in the internal environment is of substitute character,
THEN positive correlation in the external environment is preferred.

RULE 18
IF an external environment of an autonomous agent is static,
AND interaction in the internal environment is of complementary character,
THEN negative correlation in the external environment is preferred.

RULE 19
IF an external environment of an autonomous agent is static,
AND the relationship between variables describing the external environment is given by function dependence,
AND there is an interaction in the internal environment,
THEN it is easier to improve the value of information flow for small agents than for larger agents.

RULE 20
IF an external environment of an autonomous agent is static,
AND the relationship between variables describing the external environment is of statistical character,
AND there is no interaction in the internal environment,
THEN efficiency of an information flow increases with increasing n.

RULE 21
IF an external environment of an autonomous agent is static,
AND the relationship between variables describing the external environment is of statistical character,
AND there is an interaction in the internal environment,
AND there is no communication between agent members,
THEN the increase in the value of information structure decreases with increasing n.

RULE 22
IF an external environment of an autonomous agent is static,
AND the relationship between variables describing the external environment is of statistical character,
AND there is an interaction in the internal environment,
AND there is communication between agent members,
THEN efficiency of an information flow increases with increasing n.

RULE 23
IF an external environment of an autonomous agent is static,
AND there is an interaction in the internal environment,
THEN the losses caused by incomplete information increase with decreasing

correlation in the external environment.

SOFT MODELLING APPROACH

A formal quantitative model as presented in the previous Section can be helpful in creation of knowledge connected with an information flow evaluation in autonomous systems. Because of its complexity the model cannot be used for analysis and evaluation of an information flow in all possible decision situations. Qualitative modeling and reasoning, on the other hand, are areas of Artificial Intelligence (AI) that focus on reasoning about the behaviour of real life complex systems without relying on numbers. In the development of an information structure for a given system, Qualitative Reasoning (QR) tools can play a role similar to that of traditional analysis based on the mathematical model.

Next, some non-quantitative tools are discussed for addressing the problem of knowledge acquisition for an autonomous subsystem/agent in various decision situations.

Connectionist systems

Problem solving tasks, such as information structure development, may be considered pattern classification tasks. The system analyst learns mappings between input patterns, consisting of characteristics of system's external and internal environment, and output patterns, consisting of information structures to apply to these characteristics. Thus, neural networks (neural-based expert systems) offer a promising solution for automating the learning process of the analyst.

As we already know, systems analyst, while developing an information structure for a given system, transforms certain characteristics of a system into recommendations concerning the flow of information. These characteristics represent the input for the system and their full description (for both static and dynamic environments) includes 5 parameters: correlation in the external environment (*r*), dynamics (*t*), interaction in the internal environment (*q*), delay (*d*), and type of the process describing the external environment (*w*). Output consists of the following decisions (recommendations): (i) observation (or sensoring) should be present, and (ii) exchange of information should be present. An input portion together with an output portion of the data represents a training pair. The training pairs were used to train a 5-10-2 neural network (Szczerbicki 1996b).

The target values for each output node were normalised in such a way that the maximum target for each node received a value of 0.75 and the minimum target for each node received a value of 0.25. The training values for each input node were identically normalised. The learning rate and momentum term of 0.9 were used in the network. The network was trained using error back propagation procedure with a training tolerance of 5%. The network was considered trained if, for all training pairs and output nodes, |(desired output - actual output)/(desired output)| < tolerance.

After training, additional characteristics of a system were generated for use by the network. Five sets of characteristics were submitted to the network. In response, the network suggested five information flow recommendations. In all cases the recommendations agree with the IF ... AND ... THEN rules presented in Table 1.

Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decisionmaking knowledge from the supplied data (Safavian and Landgrebe 1991). Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (**S**), each belonging to one of the classes $\mathbf{C}_1$, $\mathbf{C}_2$, ..., $\mathbf{C}_k$ is as follows (Quinlan 1990):

Step 1. If all the objects in **S** belong to the same class, for example $\mathbf{C}_i$, the decision tree for **S** consists of a leaf labelled with this class.

Step 2. Otherwise, let T be some test with possible outcomes $O_1, O_2, ..., O_n$. Each object in **S** has one outcome for T so the test partitions **S** into subsets $\mathbf{S}_1$, $\mathbf{S}_2$, ... $\mathbf{S}_n$ where each object in $\mathbf{S}_i$ has outcome O_i for T. T becomes the root of the decision tree and for each outcome O_i we build a subsidiary decision tree by invoking the same procedure recursively on the set $\mathbf{S}_i$.

The above procedure is applied to training sets. The training sets are delivered from the analysis based on the quantitative model presented earlier.

Suppose, for illustration purposes, that we are interested in decision making situations involving static environment only. For such cases the following rules can be delivered using decision tree classifiers (Szczerbicki 1996b):

Rule 1
IF an external environment of a system is static
AND it is described by random variables

AND there is no interaction in the internal environment
THEN communication (exchange of information) between system elements is not necessary

Rule 2
IF an external environment of a system is static
AND it is described by random variables
AND there is interaction in the internal environment
AND the relationship between variables describing the external environment is of statistical character
THEN exchange of information between system elements should be organised

Rule 3
IF an external environment of a system is static
AND it is described by random variables
AND there is interaction in the internal environment
AND the relationship between variables describing the external environment is given by function dependence
THEN exchange of information between system elements is not necessary

The use of decision trees is simple and as effective as the analysis based on a rigorous mathematical model (the production rules formulated above are the same as the rules based on quantitative modelling given in Table 1).

Signed directed graphs

A directed graph, or digraph, is a graph in which all edges are directed (Chartrand and Oellermann 1993). A signed digraph is a digraph with either + or - associated with each edge. SDG nodes are chosen as variables relevant to or representative of the problem that is studied. There is an edge from variable A to variable B if a change in A has a significant direct effect on B. The sign of the edge is + if an increase in A leads to an increase in B, and a decrease in A leads to a decrease in B. The sign is - if the effect is opposite; an increase in A leads to a decrease in B, and a decrease in A leads to an increase in B.

According to the mathematical model, information flow depends on the following state parameters: delay of information (d), amount of information (a), dynamics in the external environment (w), variance in the external environment (s), and interaction in the internal environment (q). The above parameters influence the loss in the value of information caused by delay (L1), the loss in the value of information caused by incompleteness (L2), and total loss (LV). Based on relationships and dependencies described by mathematical model, the SDG can be developed and then simplified. Two principles are used for the simplification process. The first one is the principle of removal of intermediate nodes and the other one is the simplification of positive feedback loop.

The following logic rules can be written for the model after simplification (Szczerbicki 2002b):

SDG Rule 1:
IF [d=+] .and. p[dLV]
THEN it is a possible solution pattern for a positive change in d

SDG Rule 2:
IF [a=+] .and. n[aLV]
THEN it is a possible solution pattern for a positive change in a

SDG Rule 3:
IF[w=+] .and. p[wLV]
.and. p[wd]
.and. p[dLV]
THEN it is a possible solution pattern for a positive change in w

Using the above logic rules, the qualitative behaviour of the SDG model can be found. It is easy to notice that the corresponding qualitative states (consistent patterns) for the parameters of our interest are given as follows:

(i) solution pattern for a positive change in d

d	a	w	LV
+	0	0	+

(ii) solution pattern for a positive change in w

d	a	w	LV
0	+	0	-

(iii) solution pattern for a positive change in a

d	a	w	LV
+	0	+	+

The above results of qualitative simulation are again the same as quantitative information flow modelling and evaluation. For example, they depict the adverse character of two contrary information attributes, i.e. delay and incompleteness. They also show clearly the effects of increasing dynamics in the external environment. More generally, the results show that as far as the analysis of overall directions of a system behaviour is concerned the simple qualitative model can be sufficient at a minimum level of complexity.

CONCLUSION

This paper tries to signal some of the emerging challenges and opportunities in the area of information flow modelling and simulation based on formal mathematical modelling platform. It also includes the preliminary results of some non-quantitative procedures applied in the process of knowledge acquisition for information management. The procedures show the potential for use in reasoning and retrieval of knowledge describing the flow of information between a system and its external environment as well as within a system. It was shown that the techniques applied are able to provide general knowledge about system functioning in static and dynamic external environments. The techniques presented illustrate the ease and appropriateness of such methods for dealing with implicit knowledge and also provide a model for extension into other expert domains.

REFERENCES

Bogdan R.J. 2000. *Minding Minds*. MIT Press, Boston.

Chartrand, G. and O.R. Oellermann. 1993. *Applied and algorithmic graph theory*. McGraw-Hill, New York.

Matsumot, G. 1999. "Brain Computing", *Artificial Life and Robotics*, 3, 24-27.

Morimoto T. 2001. "Application for communication agent". In *Knowledge-based Intelligent Information Engineering Systems and Allied Technologies*, N. Baba et al (Eds). IOS Press, Amsterdam, 1619-1621.

O'Grady P. 1999. *The Age of Modularity*. Adams and Steele, Iowa City.

Prakken B. 2000. *Information, Organisation and Information Systems Design*. Kluwer, Boston.

Quinlan, J.R. 1990. "Decision trees and decision making". *IEEE Transactions on Systems, Man, and Cybernetics*. 20, 339-346.

Safavian, S. and Landgrebe D. 1991. "A survey of decision tree classifier methodology". *IEEE Transactions on Systems, Man, and Cybernetics.* 21, 660-668.

Szczerbicki E. 1996a. "External environment of an autonomous manufacturing agent: dynamics and representation". *International Journal of Systems Science,* 27, 1211-1218.

Szczerbicki E. 2000. "Simulation modelling for complex production systems". *Cybernetics and Systems: An International Journal*. 31, 333-351.

Szczerbicki E. 2002a. "Intelligent integration for autonomous manufacturing systems". *International Journal of Knowledge-Based Intelligent Engineering Systems*. 6, 214-219.

Szczerbicki E. (2002b). Soft modelling support for information management. Cybernetics and Systems: An International Journal. 33, 413-426.

Szczerbicki E. 2003. "Hard and Soft Modelling Based Knowledge Capture for Information Flow Management". In *Knowledge Based Information Retrieval and Filterin.,* Abramowicz, W. (Ed), Kluwer Academic Publishers, Boston, MA, 75-100.

Szczerbicki E. 1996b. "Decision trees and neural networks for reasoning and knowledge acquisition for autonomous agents". *International Journal of Systems Science.* 27, 233-239.

Tharumarajah A. 1999. "A self-organising model for scheduling distributed autonomous manufacturing agents". *Cybernetics and Systems: An International Journal,* 29, 461-480.

AUTHOR BIOGRAPHY

EDWARD SZCZERBICKI (MSc, PhD, DSc) has had very extensive experience in the area of intelligent systems development, information theory, and knowledge and information management over an uninterrupted 25 year period which he spent in the top systems research centres and universities in the USA, UK, Germany, Poland, and Australia. In this area he contributed to the understanding of information phenomena and knowledge management in complex systems operating in changing environments characterised by informational uncertainties. He has published a total of 200 refereed papers, over 110 of which appeared in refereed journals and the rest in international conference proceedings. With his papers published in the beginning of the nineties in *IEEE Transactions on Systems, Man, and Cybernetics* and *International Journal of Systems Science* E Szczerbicki become one of the first scientists to develop autonomous systems based on information flow for the purposes of intelligent modelling and simulation decision support. This was his unique contribution to the emerging cross-disciplinary research field of *intelligent systems* for which information has a *value*, is treated as the main *resource*, and is the base for *intelligent* decision making. E. Szczerbicki is Associated Editor, Guest Editor or a Member of the Editorial Board for seven international journals in the general area of systems science and knowledge based systems.