Volume 11, Number 2 **ISSN 1524-7252**

ACADEMY OF INFORMATION AND MANAGEMENT SCIENCES JOURNAL

Sharad K. Maheshwari
Editor
Hampton University

Editorial and Academy Information
may be found on the Allied Academies web page
www.alliedacademies.org

The *Academy of Information and Management Sciences Journal* is owned and published by the DreamCatchers Group, LLC, and printed by Whitney Press, Inc. Editorial Content is controlled by the Allied Academies, a non-profit association of scholars, whose purpose is to support and encourage research and the sharing and exchange of ideas and insights throughout the world.

Whitney Press, Inc.

Printed by Whitney Press, Inc.
PO Box 1064, Cullowhee, NC 28723
www.whitneypress.com

The *Academy of Information and Management Sciences Journal* is owned and published by the DreamCatchers Group, LLC, PO Box 2689, Cullowhee, NC 28723, USA. Those interested in subscribing to the *Journal*, advertising in the *Journal*, submitting manuscripts to the *Journal,* or otherwise communicating with the *Journal*, should contact the Executive Director of the Allied Academies at info@.alliedacademies.org.

ACADEMY OF INFORMATION AND MANAGEMENT SCIENCES JOURNAL EDITORIAL BOARD MEMBERS

ACADEMY OF INFORMATION AND MANAGEMENT SCIENCES JOURNAL

CONTENTS

LETTER FROM THE EDITOR

We are extremely pleased to present the *Academy of Information and Sciences Journal*, an official journal of the Academy of Information and Management Sciences. The AIMS is an affiliate of the Allied Academies, Inc., a non profit association of scholars whose purpose is to encourage and support the advancement and exchange of knowledge, understanding and teaching throughout the world. The *AIMSJ* is a principal vehicle for achieving the objectives of the organization. The editorial mission of this journal is to advance the knowledge and understanding of information and management science throughout the world. To that end, the journal publishes high quality, theoretical and empirical manuscripts which advance the discipline.

The manuscripts contained in this volume have been double blind refereed. The acceptance rate for manuscripts in this issue, 25%, conforms to our editorial policies.

Our editorial policy is to foster a supportive, mentoring effort on the part of the referees which will result in encouraging and supporting writers. We welcome different viewpoints because in differences we find learning; in differences we develop understanding; in differences we gain knowledge and in differences we develop the discipline into a more comprehensive, less esoteric, and dynamic metier.

The Editorial Policy, background and history of the organization and of the Allied Academies, and calls for conferences are published on the Allied Academies website. Please visit our site and know that we welcome hearing from you at any time.

Sharad K. Maheshwari
Hampton University

www.alliedacademies.org

A COGNITIVE EXPLANATION OF THE CORRECTNESS OF INFORMATION REQUIREMENT SPECIFICATIONS

I-Lin Huang, Langston University

ABSTRACT

The cognitive research in information requirement analysis has identified the knowledge of information analysts, requirement analysis techniques, and problem domains as three important determinants for the correctness of requirement specifications. Various cognitive models have also been proposed on the basis of different cognitive theories in order to explain the important variables for the correctness of requirement specifications. However, these cognitive models are inadequate because they are focused on some rather than all of the three determinants. As a result, two issues related to requirement analysis techniques cannot be accounted for: First, no requirement analysis technique can consistently outperform the others. Second, experienced information analysts use multiple requirement analysis techniques in analyzing complex information systems.

On the basis of the structure building theory of language comprehension, this research proposes a more adequate cognitive model that can not only explicate the interactive relationships among the three determinants for the correctness of requirement specifications, but also provide an explanation for the two unsolved issues related to requirement analysis techniques. In addition, this research also validates the proposed cognitive model by the research evidence from previous cognitive research studies.

INTRODUCTION

It is widely recognized that incorrect requirement specifications are the major cause of system failures (Dardenne, van Lamsweerde, & Fickas, 1993; Davis, 1988; Dorfman, 1990; Greenspan & Mylopoulos, 1982; Scharer, 1981; Standish Group, 1995; Vessey & Conger, 1994). It has been reported that about two thirds of system failures can be attributed to the mistakes made in requirement specifications (Fraser, Kumar, & Vaishnavi, 1991; Shemer, 1987). According to an estimation, incorrect requirement specifications which are not discovered until system implementation may cost fifty to one hundred times more than what would have been required if the errors were discovered during information requirement analysis (Roman, April 1985; Shemer, 1987).

In order to identify the determinants for getting information requirements right, cognitive research has been conducted to study the process of information requirement analysis. One stream of research has investigated the differences in analytical behaviors between novice and expert information analysts. Richer domain knowledge and modeling knowledge have been recognized as the qualities of expert information analysts for better performance in information requirement analysis (Adelson & Soloway, 1985; Allwood, 1986; Koubek et al., 1989; Schenk, Vitalari, & Davis, 1998; Sutcliffe & Maiden, 1990; Vessey & Conger, 1993; Vitalari & Dickson, 1983). Another stream of research has focused on comparing the effectiveness of various requirement analysis techniques in specifying information requirements. However, the research results are contradictory (Kim & Lerch, 1992; Vessey & Cogner, 1994; Yadav, Bravoco, Chatfield, & Raikumar, 1988). No requirement analysis technique has shown consistently better performance than the others (Kim & Lerch, 1992; Poti & Ramesh, 2002; Vessey & Cogner, 1994; Yadav, Bravoco, Chatfield, & Raikumar, 1988). In addition, experienced information analysts use multiple requirement analysis techniques in analyzing complex information systems (Littman, 1989). In order to explain the contradictory results, some researchers have suggested that requirement analysis techniques should be matched to types of problem domains (Fitzgerald, 1996; Jackson, 1994; Vessey & Glass, 1994). Several frameworks have also been proposed to classify requirement analysis techniques on the basis of problem domains (Davis, 1988; Iivari, 1989; Marca & McGowan, 1993; Vessey & Glass, Fall 1994).

RESEARCH QUESTION

In order to identify and explain the important variables for the correctness of requirement specifications, various cognitive models have been built on the basis of different cognitive theories such as normative models, problem solving models, mental models, and comprehension models. However, they are inadequate because they are focused on some rather than all of the three determinants: the knowledge of information analysts, requirement analysis techniques, and problem domains. As a result, at least two issues related to requirement analysis techniques cannot be accounted for: First, no requirement analysis technique can consistently outperform the others. Second, experienced information analysts use multiple requirement analysis techniques in analyzing complex information systems.

Without an adequate model of information requirement analysis, research studies may miss influential variables in viewing and comparing the cognitive processes of information requirement analysis, resulting in erroneous findings. Therefore, the research question for this research is: Is there an adequate cognitive model that can not only explicate the interactive relationships among the three determinants for the correctness of requirement specifications, but also explain the two unsolved issues for requirement analysis techniques?

To answer the research question, this article proposes a cognitive model of information requirement analysis on the basis of the structure building theory of language comprehension

(Gernsbacher, 1990). By modeling information requirement analysis as a process of text comprehension, the cognitive model can not only explicate the interactive relationships among the three important determinants for the correctness of requirement specifications, but also provide an explanation for the two unsolved issues on requirement analysis techniques.

The rest of this article is organized as follows. First, I review the characteristics of information requirement analysis, the cognitive variables for the performance of information requirement analysis, and the cognitive models used in current cognitive research of information requirement analysis. Then, I propose a more adequate cognitive model of information requirement analysis on the basis of the structure building theory of language comprehension. Third, in order to show the adequacy of the proposed cognitive model, I use the cognitive model to provide an explanation on the two unsolved issues related to requirement analysis techniques. Fourth, I validate the proposed cognitive model theoretically by the research findings from the related cognitive research studies. Sixth, I discuss the implications of this research. Finally, I make the conclusion in the final section.

LITERATURE REVIEW

In this section, I first give a brief overview on the characteristics of information requirement analysis. Then I review the cognitive variables important for the correctness of requirement specifications from the cognitive research in information requirement analysis. Finally, I review current cognitive models of information requirement analysis on the basis of their underlying cognitive theories. Due to the similarity of cognitive requirements between the upper-level system design and information requirement analysis, some cognitive models of the upper-level system design are also discussed.

Information Requirement Analysis

Information requirement analysis is the first stage of information systems development. Basically, information requirement analysis can be characterized by three features: the major outputs, the activities, and the roles involved in information requirement analysis. First, the major outputs of information requirement analysis are requirement specifications. Requirement specifications are mainly used as blueprints for creating the intended information systems. In order to support information system development effectively, requirement specifications should reflect an understanding of the intended information system, guide the subsequent design, and serve as a basis for all communications concerning the information system (Shemer, 1987). In order to facilitate an understanding of the target system, requirement specifications need to address user-level systelogical and infological issues. To serve as a communication tool during information system development, requirement specifications should be understandable to naive users, easy to modify, and maintainable. For the purpose of guiding the subsequent design, requirement specifications need to state the required functional and performance characteristics of information systems independent of any actual realization (Davis, 1988; Roman, April 1985).

Second, in order to generate requirement specifications, the activities of information analysts can be categorized into the following four purposes: "(1) identification and documentation of customer and user need, (2) creation of a document that describes the external behavior and its associated constraints which will satisfy those needs, (3) analysis and validation of the requirements document to ensure consistency, completeness, and feasibility, and (4) evolution of needs" (Hsia, Davis, & Kung, November 1993, p. 75).

Finally, there are three roles involved in the process of information requirement analysis: users, information analysts, and system designers (Fraser, Kumar, & Vaishnavi, 1991). Users have information requirements for the information systems under development. System designers have the responsibility of designing the information systems on the basis of the users' information requirements. Information analysts work in between users and system designers. Information analysts collect, understand, and document information requirements from users, and finally pass the requirement specifications to system designers.

Cognitive Variables

The research in the correctness of requirement specifications has been conducted along three dimensions: the knowledge of information analysts, requirement analysis techniques, and problem domains. First, richer domain knowledge and modeling knowledge have been suggested as important factors for better modeling performance of expert information analysts. Domain knowledge is drawn upon by both expert and novice information analysts in specifying information requirements (Sutcliffe & Maiden, 1990; Vessey & Conger, 1993). While understanding problem statements, information analysts use domain knowledge to mentally simulate a scenario of the system behavior in order to test the adequacy of the requirement specifications, to add assumptions to increase the completeness of the requirements, to test internal and external consistency of the requirements, and to abstract, summary, select and highlight important information in the problem statements (Guindon, Krasnar, & Curtis, 1987). Without domain knowledge, even expert information analysts can only specify high-level conceptual models without details (Adelson & Soloway, 1985). With the availability of domain knowledge, novice information can reuse the domain knowledge to achieve almost the same level of completeness of requirement specifications as expert information analysts do (Sutcliffe & Maiden, 1990). On the other hand, modeling knowledge has long been regarded as an important factor to differentiate expert from novice information analysts. Modeling knowledge can be divided into syntactic and semantic parts (Koubek et al., 1989). Syntactic knowledge consists of allowable syntax of a specific modeling language. Semantic knowledge, however, consists of modeling principles which are independent of a particular modeling language (Allwood, 1986). Compared to novice information analysts, expert information analysts with richer semantic knowledge can retrieve and apply more relevant modeling principles, make more critical testing of hypotheses, and finally achieve requirement specifications with better quality (Allwood, 1986; Koubek et al., 1989; Schenk, Vitalari, & Davis, 1998; Vitalari & Dickson, 1983). Modeling knowledge can also be divided into declarative and procedural aspects (Vessey & Conger, 1993). The procedural aspect of a requirement analysis technique is more difficult to learn than the declarative aspect. However, the procedural aspect of modeling knowledge is more important in determining the correctness of requirement specifications (Vessey & Coger, 1993).

Second, research into requirement analysis techniques has focused on comparing the effectiveness of various requirement analysis techniques in specifying information requirements. The purpose of requirement analysis techniques is to provide notation and procedures to help information analysts formalize the domain knowledge of problem domains during the process of information requirement analysis (Sutcliffe & Maiden, 1992). It was found that information analysts who specified information requirements by model-based reasoning based on requirement analysis techniques coud produce more complete solutions than those with partial or no model-based reasoning behavior (Sutcliffe & Maiden, 1992). However, it was also found that novice information analysts had difficulties in identifying important concepts of problem statements with requirement analysis techniques (Batra & Davis, 1992; Batra & Sein, 1994; Sutcliffe & Maiden, 1992). Yadav, Bravoco, Chatfield, & Raikumar (1988) compared the effectiveness of data flow diagrams and IDEF for supporting novice information analysts in specifying information requirements. They found that data flow diagrams were easier to learn and to use. However, neither of them produced significantly better specifications. While object orientation has become a new paradigm for information requirement analysis, there is a debate on which approach, the object orientation or the functional orientation, is a more "natural" way to specify information requirements (Firesmith, 1991; Loy, 1990; Shumate, 1991). The results on the basis of empirical studies are inconclusive. Kim and Lerch (1992) reported that expert information analysts with object-oriented techniques spent less time in analyzing problem domains and developed better understanding of the underlying problem structures than expert information analysts with functional-oriented techniques. However, Vessey and Cogner (1994) found that novice information analysts were better able to apply functional-oriented techniques than to apply object-oriented techniques. In addition, significant learning effects only occurred for functional-oriented techniques.

Finally, the research in the dimension of problem domains argues that the characteristics of problem domains should be the basis for the selection of requirement analysis techniques for information requirement analysis. Littman (1989) conducted several empirical studies to investigate the ways in which expert software designers constructed mental representations of problem domains. He found that expert software designers used multiple mental representations to model problem domains. Littman reported that expert software designers identified several modeling techniques that might be appropriate for a problem domain and then selected one that seemed most appropriate. Vessey and Glass (1994) argued that cognitive fit between problem domains and requirement analysis techniques was important for the effectiveness of information requirement analysis. They suggested that taxonomies of problem domains and taxonomies of requirement analysis techniques were needed to facilitate matching techniques to problem domains. To match requirement analysis techniques to problem domains, requirement analysis techniques have long been classified into functional orientation, data orientation, control orientation, or object orientation (Dorfman, 1990). Marca and McGowan (1993) classified requirement analysis techniques on the basis of the theory of world views in metaphysics. Iivari (1989) added levels of abstraction as another dimension to classify requirement analysis techniques. Sowa and Zachman (1992) provided a framework to categorize requirement analysis techniques on the basis of six dimensions: data, process, network, people, time, and motivation. Opdahl and Sindre (1995) proposed a facet-modeling structure to integrate various requirement analysis techniques. Jackson (November 1994) suggested that future requirement analysis techniques should be more problem-oriented to fit the structures of problem domains rather than solution-oriented.

In this section the cognitive variables determining the performance of information requirement analysis are discussed along three dimensions: requirement analysis techniques, the knowledge of information analysts, and types of problem domains. Research evidence shows that the three dimensions are highly correlated in determining the correctness of requirement specifications. However, the interactive relationships among the three dimensions are still unclear. As a result, there are at least two important issues related to requirement analysis techniques remaining unsolved: First, why no requirement analysis technique can have consistently better performance than others (Kim & Lerch, 1992; Poti & Ramesh, 2002; Vessey & Cogner, 1994; Yadav, Bravoco, Chatfield, & Raikumar, 1988)? Second, why expert information analysts use multiple requirement analysis techniques in analyzing information requirements (Littman, 1989)? On the basis of previous research evidence, we cannot predict the influence of requirement analysis techniques upon the correctness of requirement specifications without knowing first the types of problem domains and the knowledge of information analysts. In order to solve the two issues, I will explore the interaction among the three dimensions through the cognitive models proposed in cognitive research of information requirement analysis.

Cognitive Models

In order to identify and explain important variables for the correctness of requirement specifications, various cognitive models have been proposed on the basis of different cognitive theories. Basically, four approaches have been used to derive cognitive models of information requirement analysis: normative models, problem solving models, mental models, and comprehension models.

First, normative models are referred to as the models of information requirement analysis that are built on the basis of the researchers' experiences or opinions. Normative models are often built for comparing or evaluating requirement analysis techniques. Examples can be found in the research papers of Davis (1988); and Yadav, Bravoco, Chatfield, and Rajkumar (1988). This class of models provides a set of criteria about what should be achieved by good requirement analysis techniques. However, without an understanding of the cognitive behaviors of information analysts, those models provide no guideline for how to support information analysts in understanding problem domains and in specifying correct information requirements.

Second, some researchers believe that information requirement analysis is similar to the cognitive process of problem solving (Malhotra et al., 1980; Schenk, Vitalari, & Davis, 1998; Sutcliffe & Mainden, 1992; Vitalari & Dickson, 1983). They focus on the reasoning processes that information analysts use to analyze information requirements. Specifically, they focus on "the frequency, ordering, and association with analysts' performance of the clues, goals, strategies, heuristics, hypotheses, information, and knowledge manifested in the thought process of the information analysts" (Vitalari & Dickson, 1983, p. 949). Some examples can be found in the research papers of Malhotra et al. (1980); Vitalari and Dickson (1983); and Sutcliffe and Mainden (1992). On the basis of the problem solving paradigm, this class of cognitive models provides a framework for understanding the influence of knowledge and reasoning processes on the correctness of requirement specifications. However, these models cannot identify the interaction among cognitive variables in determining the correctness of requirement specifications. In addition, problem domains have not been identified as a variable in these models.

Third, mental models have also been used to identify important determinants for the correctness of requirement specifications. A mental model is a collection of interconnected autonomous objects (Williams, Hollan, & Stevens, 1983, p.133). According to Williams, Hollan, and Stevens, autonomous objects are mental objects that have their own internal rules to guide their behaviors. The interactions among autonomous objects achieve the task of human reasoning. Some mental models for information requirement analysis can be found in the research papers of Adelson and Soloway (1985); Guindon and Curtis (1988); Guindon, Krasner, and Curtis (1990); and Vessey and Conger (1993). On the basis of the mental models, problem statements, domain knowledge of information analysts, and methodology knowledge of information analysts are identified as three sources of knowledge for specifying information requirements. However, those models provide no theoretical basis for explaining the interactive relationships between problem domains and requirement analysis techniques.

Finally, some researchers view information requirement analysis as a comprehension process. They believe that information requirement analysis is a process of translating the knowledge of problem statements into that of requirement analysis techniques. Therefore, good requirement analysis techniques should be able to make the translation process easy for information analysts. For example, Kim and Lerch (1992) focus the required skills for the translation of information requirements. They believe that object-oriented techniques are better than functional-oriented techniques because the required skill for object oriented techniques (symbolic simulation) is easier than that for functional-oriented techniques (test case). Batra and Sein (1994), on the other hand, suggest that the constructs used by requirement analysis techniques should be close to those of problem domains. By modeling information requirement analysis as a comprehension process, the above two models provide some insights about the required features for good requirement analysis techniques. However, they do not explain the interactive relationships between requirement analysis techniques and problem domains. In addition, the knowledge of information analysts is not included in those models.

In sum, current cognitive models have focused on different aspects of information requirement analysis like knowledge of information analysts, requirement analysis techniques, or modeling behaviors of information analysts. However, due to the lack of an integrated view of information requirement analysis, it is unknown what the interactive relationships between the different cognitive variables are. Even worse, some research studies may reach conflicting or contradictory conclusions because of the negligence of confounding variables.

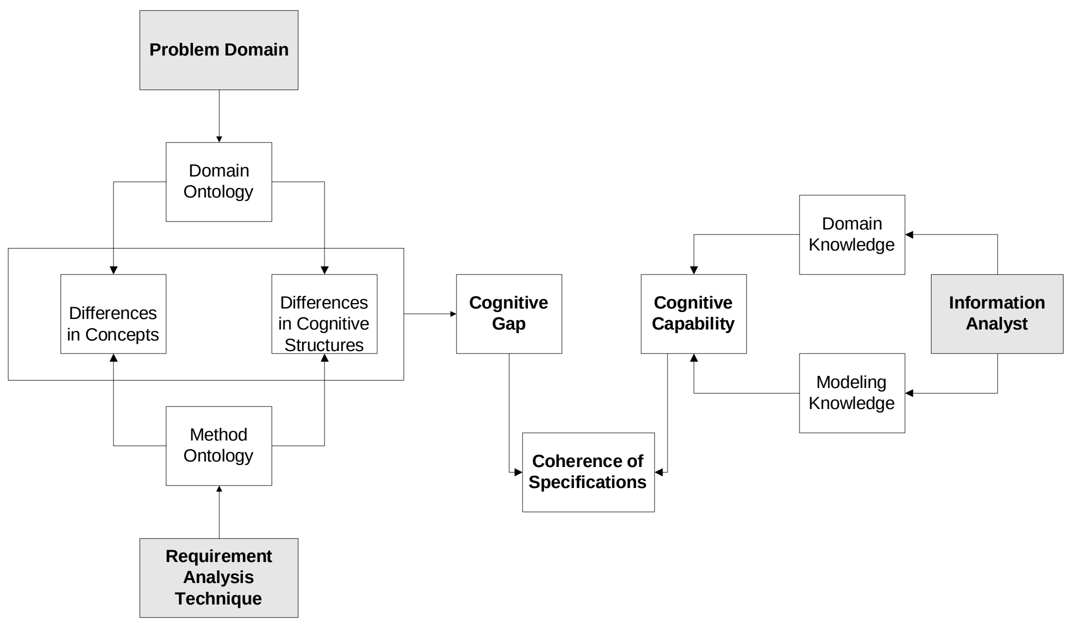
A STRUCTURE BUILDING MODEL FOR INFORMATION REQUIREMENT ANALYSIS

In this section, I will propose a more adequate cognitive model of information requirement analysis that meets two requirements: first, it explicates the interactive relationships among the three determinants for the correctness of requirement specifications; and second, it provides an explanation for the two unsolved issues related to requirement analysis techniques. In the rest of this section I will first discuss the rationale of using text comprehension as a basis for the proposed cognitive model. Then I will discuss the details of the proposed cognitive model.

Information requirement analysis has been recognized as a process of understanding the domains of information systems and then specifying the information requirements for the information systems (Borgida, Greenspan, & Mylopoulos, 1985; Fraser, Kumar, & Vaishnavi, 1991; Huang & Burns, 1997; Yadav, 1983). Empirical studies on the analytical behavior of information analysts also showed a strong association among the activities of gathering information, identifying relevant facts, and conceptual modeling (Batra & Davis, 1992; Sutcliffe & Maiden, 1992). This strong association reflects that information requirement analysis is basically an understanding process.

From the perspective of human cognition, understanding is a process of building a coherent mental representation of the information being comprehended (Gernsbacher, 1990; Graesser, 1995; Kintsch, 1988; Ortony, 1978). On the basis of the structure building model of language comprehension (Gernsbacher, 1990), this article proposes a more adequate cognitive model of information requirement analysis that can explicate the interactive relationships among the knowledge of information analysts, requirement analysis techniques, and problem domains as shown in Figure 1. In addition, the cognitive model can provide a cognitive explanation for the two unsolved issues about requirement analysis techniques.

Figure 1: A structure building model of information requirement analysis



According to the structure building model of information requirement analysis, there are three important features that explain the modeling behavior of information analysts. First, coherence is the goal of information analysts in specifying information requirements. According to Collins Cobuild English Dictionary (1995), coherence is "a state or situation in which all the parts or ideas fit together well so that they form a united whole." While information analysts analyze problem domains on the basis of a particular requirement analysis technique, they may find that the problem statements can not be fitted into the conceptual model very well. Information analysts then perform intensive coherence inferences to resolve the discrepancies. Although correctness may be the goal for information requirement analysis, information analysts believe that understanding is achieved only if the requirement specifications are coherent.

Second, the coherence of a requirement specification can be achieved only when the cognitive capability of the information analyst is big enough to fill the cognitive gap between the problem domain and the requirement analysis technique being used. The cognitive capability of an information analyst can be evaluated by the domain knowledge and modeling knowledge owned by the information analyst. It is well recognized that richer domain knowledge and modeling knowledge are the qualities of expert information analysts for better performance in specifying information requirements. On the other hand, the cognitive gap between a problem domain and a requirement analysis technique can be evaluated by the difference of ontologies used by the problem domain and the requirement analysis technique. An ontology is a conceptual system (Guarino & Giaretta, 1995; Regoczei & Plantinga, 1987; Wand, Monarchi, Parsons, & Woo, 1995) that includes two parts: (1) a set of concepts for describing a problem domain such as entities, relationships, data flows, and agents; and (2) a cognitive structure for organizing the concepts (Marca & McGowan, 1993; Pepper, 1942) such as functional orientation, object orientation, data orientation, and control orientation. The difference in ontologies determines how difficult it is for information analysts to model a problem domain by a particular requirement analysis technique. Requirement analysis techniques act like schemata in text comprehension (Rumelhard, 1980). A schema governs a body of inferences which can help information analysts understand the conceptual structures of problem domains easily (Abelson, 1981). However, if the concepts and cognitive structures between problem domains and requirement analysis techniques are very different, information analysts may abandon the requirement analysis techniques, or even resulting in a failure of comprehension.

Third, understanding is a structure-building process that translates the ontologies of problem domains into those of requirement analysis techniques. Three operations are important for the translation process: mapping, shifting, and integrating. Mapping occurs while an information analyst studies a substructure of a problem domain. The information analyst first selects a requirement analysis technique that is best suited to the substructure of the problem statements. On the basis of modeling knowledge and domain knowledge, the information analyst then maps the substructure of the problem dsomain into the conceptual model of the selected requirement analysis technique. Shifting occurs while an information analyst decides that the selected technique cannot achieve a locally coherent conceptual model for a substructure of a problem domain. In this situation, the information analyst will try other requirement analysis techniques to build a conceptual model with local coherence for the substructure of the problem domain. Finally, integrating is the operation that integrates the conceptual models for the substructures of a problem domain into a conceptual model for the whole problem domain with global coherence.

A COGNITIVE EXPLANATION OF THE TWO UNSOLVED ISSUES ABOUT REQUIREMENT ANALYSIS TECHNIQUES

The cognitive gaps and cognitive capabilities in the proposed cognitive model can explain the two unsolved issues related to requirement analysis techniques: First, why is there no requirement analysis technique can consistently outperform the others? On the basis of the structure building model of information requirement analysis, the best requirement analysis technique should be the one that can not only minimize the cognitive gap by matching the ontology of the requirement analysis technique with the ontology of the particular problem domain, but also maximize the cognitive capabilities by matching the ontology of the requirement analysis technique and the ontology of the information analyst's knowledge. Therefore, the performance of a requirement analysis technique depends on its interactions with the type of problem domain under investigation, and the knowledge of information analysts. As a result, there is no single requirement analysis technique that can outperform other techniques without considering the types of problem domains and the knowledge of information analysts.

Second, why do experienced information analysts use multiple requirement analysis techniques in analyzing complex information systems? This is to say that experienced information analysts are eclectic in terms of their choice of techniques. This phenomenon reflects that in order to achieve high-level coherence of requirement specifications, experts select a particular requirement analysis technique not only based on their experience and knowledge (cognitive capabilities), but also the match between requirement analysis techniques and types of problems domains (cognitive gaps). As a result, experts use different requirement analysis techniques to fit different aspects of a problem domain and to fit their knowledge about the requirement analysis technique and the problem domain.

A THEORETICAL VALIDATION FOR THE STRUCTURE BUILDING MODEL

The proposition of the structure building model of information requirement analysis is based on the findings of empirical research in text comprehension and information requirement analysis. This section will discuss the previous research findings that support the concepts of coherence, cognitive gap and structure building process in the structure building model of information requirement analysis.

First, coherence has been recognized as a measurement for concept comprehension (Murphy & Medin, 1985). Komatsu (1992) suggested that an adequate theory about concept comprehension should be able to explain how coherence is achieved. Gernsbacher (1990) identified four sources of coherence: referential coherence, temporal coherence, location coherence, and causal coherence. He found that sentences with the above four types of coherence were easier to understand. In investigating the inference behavior of readers in text comprehension, Mckoon and Ratcliff (1992) found that inferences for local coherence were the basic inferences that were automatically performed by readers. Zwaan, Graesser, and Magliano (1995) investigated the effects of temporal, spatial, and causal discontinuities on sentence reading times in naturalistic story comprehension. They found that readers simultaneously monitored multiple dimensions of text coherence under a normal reading instruction. In addition, readers' goals would direct readers to allocate cognitive resources to specific dimensions of coherence.

Second, the cognitive gap between a problem domain and a requirement analysis technique is also recognized as a determinant for comprehension. The research in education found that the concepts in science were difficult to learn because of the ontology incompatibility between students' cognition and science (Chi, Scotta, & de Leeuw, 1994). It was believed that students' concepts of physics were mostly matter-based. However, the concepts in the theories of physics were mostly process-based. In an experiment to test users' ability to validate information requirements, Nosek and Ahrens (1986) found that compared to data flow diagrams, task oriented menus, which were believed to be closer to users' cognitive models, were better understood by users. Batra, Hoffer, and

Bostrom (1990) compared the performance of novice analysts in data modeling by relational data models and entity relationship models. They found that entity relationship models, which were believed to be closer to analysts' cognitive models, scored higher in correctness.

Finally, empirical evidence also supports the structure building process of information requirement analysis. Model-based reasoning has been regarded as an important factor for the effectiveness of information requirement analysis (Sutcliffe & Maiden, 1990). Littman (1989) found that expert software designers used multiple mental representations to model a problem domain. In addition, expert software designers engaged in the process of selecting requirement analysis techniques to best fit the problem domain. Wijers and Heijes (1990) found that although expert information analysts had their own preference on requirement analysis techniques, they all had their own strategies to select different techniques in information requirement analysis. Finally, Gentner (1983) suggested that structure mapping was the way that people use to associate domain knowledge to problem domains for comprehension.

IMPLICATIONS OF THIS RESEARCH

This research has developed a cognitive model of information requirement analysis on the basis of the structure building model of language comprehension. This cognitive model contributes to the research in information requirement analysis in four aspects as follows.

First, this cognitive model provides a more adequate theory of information requirement analysis. In addition to explicating the interactive relationships among the knowledge of information analysts, requirement analysis techniques, and problem domains, the cognitive model uses cognitive gaps and cognitive capabilities to explain two important phenomena in information requirement analysis: (1) no requirement analysis technique can consistently outperform the others; and (2) experienced information analysts use multiple requirement analysis techniques in analyzing complex information systems.

Second, this cognitive model provides a basis for empirical validation. On the basis of this cognitive model, empirical research can be conducted to test the influence of the knowledge of information analysts, requirement analysis techniques, problem domains, and the interactive relationships among the three determinants on the correctness of requirement specifications.

Third, the cognitive model provides a framework to view and to compare important aspects of information requirements. According to the cognitive model, the knowledge of information analysts, requirement analysis techniques, and problem domains are interactive in determining the correctness of requirement specifications. Therefore, effective research studies in the cognitive processes of information requirement analysis should consider interactive relationships among these three determinants.

Finally, this cognitive model provides a theoretical basis for the development of computer-aided software engineering (CASE) tools. According to this cognitive model, how to reduce cognitive gaps and how to enhance cognitive capabilities are two important issues for the research in CASE tools.

CONCLUSION

Current research into the process of information requirement analysis has focused on answering the question of what factors influence the performance of information analysts in specifying information requirements. However, this research tries to answer why the knowledge of information analysts, requirement analysis techniques, and problem domains can influence the coherence of information requirement specifications.

On the basis of the theories of text comprehension, coherence is assumed to be the goal of information analysts in specifying information requirements. Cognitive gaps are proposed to measure the fitness between requirement analysis techniques and problem domains. Cognitive capabilities are used to reflect the knowledge of information analysts for transforming the ontologies. On the basis of coherence, cognitive gaps, and cognitive capabilities, a cognitive model of information requirement analysis is then proposed to explicate the interactive relationships among knowledge of information analysts, requirement analysis techniques, and problem domains. In addition, the cognitive model also provides an explanation for the two issues related to the performance of requirement analysis techniques.

In order to reduce the cognitive gaps between problem domains and requirement analysis techniques, three directions for future research in requirement analysis techniques can be identified: First, the ontologies of requirement analysis techniques should be close to those of problem domains. Therefore, domain-specific requirement analysis techniques deserve the attention of future research. Second, requirement analysis techniques should be able to cover and integrate more than one cognitive structure to reduce the need to shift among requirement analysis techniques with different cognitive structures. Third, in order to avoid the cognitive gaps, requirement specification reuse should be a major concern for the research in information requirement analysis.

REFERENCES

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36* (7), 715-729.

Adelson, B., & Soloway, E. (1985). The role of domain experience in software design. *IEEE Transactions On Software Engineering, se-11* (11), 1351-1360.

Allwood, C. M. (1986). Novices on the computer: a review of the literature. *International Journal of Man-Machine Studies, 25*, 633-658.

Batra, D., & Davis, J. G. (1992). Conceptual data modeling in database design: Similarities and differences between expert and novice designers. *International Journal of Man-Machine Studies, 37*, 83-101.

Batra, D., Hoffer, J. A., & Bostrom, R. P. (1990). Comparing representations with relational and EER models. *Communications of the ACM, 33* (2), 126-139.

Batra, D., & Sein, M. K. (1994). Improving conceptual database design through feedback. *International Journal of Human-Computer Studies, 40*, 653-676.

Borgida, A., Greenspan, S., & Mylopoulos, J. (1985, April) Knowledge representation as the basis for requirements specifications. *IEEE Computer*, 82-90.

Chi, M.T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to process: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Collins Cobuild English Dictionary (1995). Harper Collins Publishers.

Dardenne, A., van Lamsweerde, A., & Fickas, S. (1993). Goal-directed requirements acquisition. *Science of Computer Programming, 20*, 3-50.

Davis, A. M. (1988). A comparison of techniques for the specification of external system behavior. *Communications of the ACM, 31*(9), 1098-1115.

Dorfman, M. (1990). System and software requirements engineering. In Thayer, R. H., & Dorfman, M. (eds*.). System and Software Requirements Engineering*, IEEE Computer Society Press. 4-16.

Firesmith, D. (1991). Structured analysis and object-oriented development are not compatible. *Ada Letters, XI* (9), 56-66.

Fitzgerald, B. (1996). Formalized systems development methodologies: A critical perspective. *Information System Journal, 6*, 3-23.

Fraser, M. D., Kumar, K., & Vaishnavi, V. K. (1991). Informal and formal requirements specification languages: Bridging the gap. *IEEE Transaction On Software Engineering, 17*(5), 454-465.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gernsbacher, M. A. (1990). *Language Comprehension As Structure Building*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Graesser, A. C. (1995). Inference generation and the construction of situation models. In Weaver, C. A., Mannes, S. , & Hetcher , C. R. (eds.). *Discourse Comprehension.* Hillsdale, New Jersey: Lawrence Erbaum Associates, Publishers. 117-139.

Greenspan, S. J., & Mylopoulos, J. (1982). Capturing more world knowledge in the requirements specification. *Proceedings of International Conference in Software Engineering*, IEEE Computer Society Press 225-234.

Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases towards a terminological clarification. *Proceedings of the Second Conference on Building and Sharing of Very Large-scale Knowledge Base*, 25-32.

Guindon, R., & Curtis, B. (1988). Control of cognitive process during software design: what tools are needed? *Proceedings of CHI'88*. 263-268.

Guindon, R., Krasner, H., & Curtis, B. (1987). Cognitive process in software design: Activities in early, upstream design. in Bullinger, H. J., & Shackel, B. (eds). *Human-computer Interaction-INTERACT'97*. North-Holland, 383-388.

Hsia, P., Davis, A., & Kung, D. C. (1993). Status Report: Requirements Engineering, *IEEE Software*, 75-79.

Huang, I. L., & Burns, J. R. (1997). Organizational information requirement topology as the basis for organizational information requirement specification reuse. *Proceedings of AIS Americas Conference on Information systems*, 36-37.

Iivari, J. (1989). Levels of abstraction as a conceptual framework for an information system. in Falkenberg, E. D., & Lindgreen, P. (eds*.). Information System Concepts: An In-depth Analysis*. North-Holland Co., 323-349.

Jackson, M. (1994, November). Problems, methods and specialization*.* *IEEE Software*, 57-62.

Kim J., & Lerch, F. J. (1992). Towards a model of cognitive process in logical design: Comparing object-oriented and traditional functional decomposition software methodologies. *Proceedings of CHI'92*, 489-498

Kintsch, W. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Komatsu, L. K. (1992) Recent views of conceptual structure. *Psychological Bulletin, 112*(3), 500-526.

Koubek, R. J., Salvendy, G., Dunsmor, H. E., & Lebold, W. K. (1989). Cognitive issues in the process of software development: review and reappraisal. International *Journal of Man-Machine Studies, 30*, 171-191.

Littman, D. (1989). Constructing expert systems as building mental models. *Knowledge Representation and Organization in Machine Learning*. West German: Springer-Verlag, 88-106.

Loy, P. H. (1990). A comparison of object-oriented and structured development methods. *Software Engineering Notes, 15*(1), 44-48.

Malhotra, A., Thomas, J. C., Carroll, J. M., & Miller, L. A. (1980). Cognitive processes in design. *International Journal of Man-Machine Studies, 12*, 119-140.

Marca, D. A., & McGowan, C. L. (1993). Specification approaches express different world hypotheses. *Proceedings of the Seventh International Workshop on Software Specification and Design*. IEEE Computer Society Press, 214-223.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440-466.

Murphy, G. L. & Medin, D. L. (1985, July). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289-316.

Nosek, J. T., & Ahrens, J. D. (1986). An experiment to test user validation of requirements: Data-flow diagrams vs task-oriented menus. *International Journal of Man-Machine Studies, 25*, 675-684.

Opdahl, A. L., & Sindre, G. (1995). Facet models for problem analysis. *Advanced Information System Engineering*, 54-67.

Ortony, A. (1978). Remembering, understanding, and representation. *Cognitive Science, 2*, 53-69.

Pepper, S. C. (1942). *World Hypotheses: A Study in Evidence*. Berkeley, CA: University Of California Press.

Regoczei, S., & Plantinga, E. P. O. (1987). Creating the domain of discourse: Ontology and inventory. *International Journal of Man-Machine Studies, 27*, 235-250.

Roman, G. (1985, April). A taxonomy of current issues in requirements engineering. *IEEE Computer*, 14-22.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In Spiro, R. Bruce, B., & Brewer, W. (eds.). *Theoretical Issues in Reading Comprehension*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Scharer, L. (1981, April). Pinpointing requirements. *Datamation*, 17-22.

Schenk, K.D., Vitalari, N.P., & Davis, K.S. (1998). Differences between novice and expert system analysts: What do we know and what do we do? *Journal of Management Information Systems, 15*(1), 9-50.

Shemer, I. (1987). System analysis: A systematic analysis of a conceptual model. *Communications of the ACM, 30*(6), 506-512.

Shumate, K. (1991). Structured analysis and object-oriented design are compatible. *Ada Letters, XI*(4), 78-90.

Sowa, J. F., & Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture. *IBM Systems Journal, 31*(3), 590-616.

Standish Group (1995). *Software Chaos*, retrieved November 20, 2004 from http://www.standishgroup.com/chaos.html

Sutcliffe, A., & Maiden, N. (1990). Cognitive studies in software engineering. *Proceedings of the Fifth Conference on Cognitive Ergonomics*, European Association of Cognitive Ergonomics, ECCE-5, 159-171.

Sutcliffe, A., & Maiden, N. (1992). Analysing the novice analyst: cognitive model in software engineering. *International Journal of Man-Machine Studies, 36*, 719-740.

Topi, H., & Ramesh, V. (2002). Human factors research on data modeling: A review of prior research, an extended framework and future research directions, *Journal of Database Management, 13*(2), 3-19.

Vessey, I., & Conger, S. (1993). Learning to specify information requirements: the relationship between application and methodology. *Journal of Management Information Systems, 10* (2), 177-201.

Vessey, I. & Conger, S. A. (1994). Requirement specification: Learning object, process, and data methodologies. *Communications of the ACM, 37*(5), 102-113.

Vessey, I., & Glass, R. L. (Fall 1994). Application-based methodologies: Development by application domain. *Information System Management*, 53-57.

Vitalari, N. P., & Dickson, G. W. (1983). Problem solving for effective system analysis: An experimental exploration. *Communications of the ACM, 26*(11), 948-956.

Wand, Y., Monarchi, D.E., Parsons, J., Woo, C.C. (1995). Theoretical foundations for conceptual modeling in information system development, *Decision Support Systems 15*, 285-304.

Wijers, G. M., & Heijes, H. (1990). Automated support of the modeling process: A view based on experiments with expert information engineers. *Lecture Notes in Computer Science*, Spring-Verlag. 88-106.

Williams, M. D., Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. in Gentner, D., & Stevens, A. L. (eds). *Mental Models*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey. 131-153.

Yadav, S. B. (1983). Determining an organization's information requirements: A state of the art survey. *DATA BASE, 14*(3), 3-20.

Yadav, S. B., Bravoco, R. R., Chatfield, A. T., & Rajkumar, T. M. (1988) Comparison of analysis techniques for information requirement determination. *Communication of the ACM, 31*(9), 1090-1097.

Zwaan, R. A., Graesser, A. C., & Magliano, J. P. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology, 21*(2), 386-397.

DRIVING FACTORS FOR E-COMMERCE: GULF REGION REVIEW

Khalid Al-Rawi, Al-Ain University, UAE
Khaled Sabry, Al-Ain University, UAE
Ahmed Al-Nakeeb, Al-Ain University, UAE

ABSTRACT

Despite the fact that e-commerce is likely to have far-reaching implications for the international market, until now only a few studies have explored e-commerce growth in the Gulf region from an international perspective. This paper reviews and explores e-commerce growth internationally with a particular focus on the Gulf region. It examines some of the challenges e-commerce faces by looking at both local and global factors with a particular focus on the area of technological advancement. Based on the review, a list of driving factors that may contribute to e-commerce evolvement in the region has been drawn up along with recommendations and suggestions for future research.

INTRODUCTION

An early analysis of web-related information economics was given by Benjamin & Wigand (1995), who argued that the Internet has a great potential for efficiency gains along the whole industry value chain, primarily because of transaction cost savings.

The Web not only can help to reduce costs and add value for existing customers, but it also has a potential role in customer acquisition, and in the case of a Web startup, this role is crucial (Bertschek, Fryges & Kaiser, 2006). The Internet is also an important advertising medium (Hofacker, 2001). Drucker (2002) argued that the "truly revolutionary impact of the Internet is e-commerce" (pp.3-4).

E-commerce can be described as "the sale and purchase of goods or services over the Internet"(Chase Paymentech, 2007), the use of IT to allow direct selling and automatic processing of purchases between parties using the internet (Expedite Email Marketing, 2004) or the substitution of information for physical business processes (Weill & Vitale, 2001). An increasing volume of E-commerce research has been conducted for a diverse range of areas in terms of application areas, technological aspects, support and implementation (Ngai & Wat, 2002) as well as the significance of cost aspects in relation to e-commerce development (Cohen & Kallirroi, 2006).

According to Lin, (2007) e-commerce is one of those rare cases where the need to change and emerging technologies come together to re-engineer the way in which business is conducted, improving processes efficiency, allowing for flexibility, working more closely with suppliers, and catering for the needs and expectations of the customers. It allows companies to select the best suppliers regardless of the geographical location and to widen the scope of their market to the global market (Soopramanien, Fildes & Robertson, 2007). Its evolution goes through several stages of growth according to Chan & Swatman (2004). They proposed a four-stages model in relation to B2B E-commerce (table 1): stage 1: growth: initial e-commerce; stage 2: centralised E-commerce; stage 3: looking inward for benefits; and stage 4: global e-commerce.

Table 1 E-commerce four-stages model (Source: Chan and Swatman, 2004)

Stages of growth	Stage 1: Initial e-commerce (early e-commerce)	Stage 2: Centralised e-commerce (ETG)	Stage 3: Looking inward for benefits (barcoding)	Stage 4: Global e-commerce (global e-commerce)
Strategy	Cost-saving	Competitive advantage	Education	Broad e-commerce coverage
Structure	Departmental	Company-wide	Corporate	Corporate
E-commerce technologies used	PC VAN-based EDI	Corporate gateway EDI (direct link, VAN-based EDI, EDI-fax)	Barcoding, EDI	Various, including Internet-based applications
Focus	Suppliers	Customers	Steel industry	Broad coverage of trading partner (emphasis on customers)
Personnel involved	SPPD-supply department	BHP Steel and BHP IT, consultants	BHP Steel e-commerce group	IS council (Steel GM+4 others), BHP IT

The readiness for e-commerce or e-readiness, however, may depends on the "state of play of a country's information and communications technology (ICT) infrastructure and the ability of its consumers, businesses and governments to use ICT to their benefit" (EIU, 2008). E-commerce can be shaped by variety of factors such as the economic, political and social environments in addition to the quality of the ICT infrastructure, however average e-readiness has risen in 2008 in comparison with 2007 (EIU, 2008). Such growth of e-commerce can also be attributed to the unique features of the Internet and the Web such as ubiquity, global reach, universal standards, richness, interactivity, information density, personalization and customization factors (Laudon &Laudon, 2007).

Internationally, many developments have occurred in the last 10 years in the world of e-commerce (IGW, 2007) that impacted on business, government and personal computing. Washington-based Software & Information Industry Association' compiled the top 10 list of developments that have most impacted e-commerce. Google, at the op of the list, used by 30% of Internet users, half of the 6.9 billion online searches conducted by US users in February 2007 were on Google. Followed by Broadband, eBay online auctions, Amazon.com online store, GoogleAdWords Key word advertising, Open standards for the web embodied in HTML (overseen by the WWW Consortium), Wi-Fi Wireless internet, User-generated content (YouTube), iTunes where more than US$2 billion worth of music was sold online or through mobile phones in 2006 and digital sales now account for around 10 % of the music market, and finally BlackBerry communications device which enables new mobile business culture.

Moreover, many companies have invested in technologies that support e-commerce. Cisco invested $4.5 billion in research and development, listening to their customers and investing for the future giving them a competitive advantage and developing innovative solutions that meet customer needs and drive their growth (Cisco, 2007). Cisco provided many companies with solutions to improve their business performance. For example, it provided Partner Infra-Solutions to grocer METRO (Canada) to improve debit/credit transactions, stability, reliability, performance and saving money (Cisco News, 2007). Bill Gates, chairman of MS, stressed the importance of investments in both IT and education for creating economic opportunities and indicated MS commitment for the improvement of technology access and fostering of innovative teaching and learning methods (GulfNews, 2008d).

Historically, e-commerce began in 1995 when Netscape.com (one of the first Internet Portals) started and explored the idea of using the web for sales and advertising, followed by a growth of e-commerce retail sales to a different dimension of doubling and tripling until the dot.com bubble burst in 2001. Since then e-commerce revenues were on the rise and the use of mobile phones for buying goods and services (m-commerce) was also taking off and steadily growing (Laudon & Laudon, 2007). However, the question here is, how far the September 2008 financial market meltdown in the USA, or the global banking crisis will affect this growth, negatively or positively, this will remain to be seen.

According to the Census Bureau of the US Department of Commerce the estimated U.S. retail e-commerce sales for the third quarter of 2007, was $34.7 billion. This represented an increase of 3.6 % from the second quarter of 2007 and 3.4% of the total retail sales ($1,020.4 billion) for the third quarter of 2007 (USCBN, 2007). Despite the fact that the above figure indicates an increase in e-commerce sales, it still represents a small portion of the total sales figure. This indicates, to some extent, that the fundamentals of business have not changed that much, and that most pure-play business models were to a great extent overoptimistic and that the future role of the Internet in business in general and marketing in particular will largely be as part of an integrated model of clicks-and-mortar (Molla & Heeks, 2007), at least until the number of internet users worldwide reaches a different height. On the other hand, in relation to financial aspects, what would be the impact of the financial banking crisis of September 2008 which hit the USA and global banks on technological firms, online banking and consequently on e-commerce growth as a whole? As the CEO of Microsoft put it, nobody knows exactly what's going to happen (GulfNews, 2008k).

From the Gulf and Arab region perspective however, many countries, particularly in the gulf region, have already or are keen to implement e-business and e-commerce strategies to promote economic and social development. E-commerce can create profound changes in the structure of the economy and accelerate social changes, and its adaptation can result in improvements in productivity, particularly, it gives SMEs in the region the ability to penetrate international markets that few years ago used to be difficult to enter due to high transaction costs and other market barriers (UNCTAD, 2002). The emergence of successful industries such as software development and telecom services and adoption of latest technologies in several countries in the region are good evidence of this.

RESEARCH OBJECTIVES AND METHODOLOGY

For such potential benefits to materialize, this paper attempts to demonstrate the need for an action plan to create enabling environment for e-commerce in the region, addressing the main problems and key factors that contribute to e-commerce growth in the region, taking into account key policies to stimulate and maximize the effect of e-commerce on their economies. The aim of this paper is to explore and analyse some of the problems facing e-commerce in the Gulf region as part of a research in e-commerce potential and implementation in the Gulf area. It is of an exploratory nature and the methodology described below is deemed appropriate for the purpose of this paper. A future research will seek to obtain and collect data to test some of the findings of this initial exploration.

The research methodology used in this paper is based on review of available literature and focuses mainly on published literature on e-commerce, presents available statistics, analysis, and critiques. The purpose is to offer an overview of significant literature published on the topic as an initial investigation that may lead to further research and contribution to this field. The review aims to provide an understanding of issues, unresolved questions, difficulties and highlight some factors that may contribute to its success in the region. It tries to answer the questions: what are the problems that face e-commerce in the region? What are the factors that may drive its success in the region?

E-COMMERCE DEVELOPMENT IN THE GULF REGION

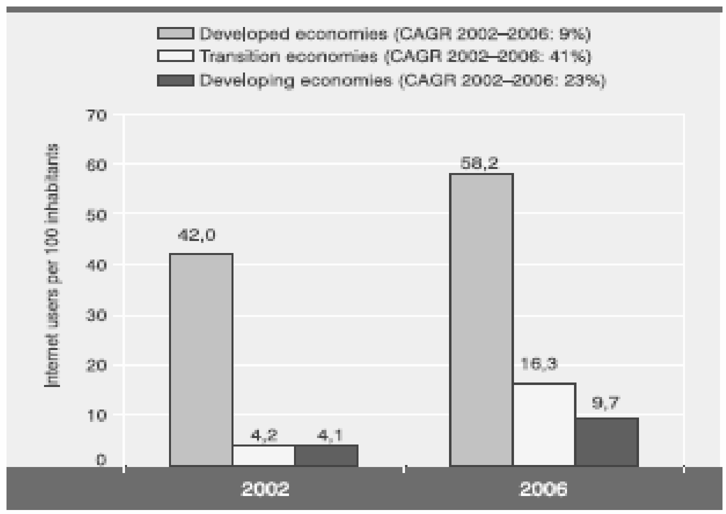
According to DIT's survey on e-commerce in the Arab world only 4% of Internet users in Arab countries made a purchase through the Internet within the span of a year (IAW, 1998). Such low percentage was contributed to the newness of e-commerce in the region at the time, lack of critical mass of Internet users in the Arab world, and the inability to ensure secure transactions. The survey indicated users' reluctance in using their credit cards numbers online due to security issues. The majority of those who made a purchase (92%) indicated that the item they purchased was from companies that were not situated in the Middle East. Similarly, Hoffman & Novak (2000) found that while advertisers were generally keen to use the Web to communicate product information, they were concerned about security/privacy issues. One question might be asked here is, what would be the effect of security problems (such as the 'Card Scams' that occurred in September 2008 and affected UAE banks) on both traders and consumers (GulfNews, 2008j), particularly on online banking? This will remain to be seen.

The above survey also indicated other technological aspects that affected the low level of e-commerce transactions such as slow Internet connections and high cost of building and managing websites. Others believe that the main problem was the lack of understanding of the concept of e-commerce and its integration in the business activities, especially for the low level of e-commerce transactions (Elbeltagi, 2007). According to Numberger & Rennhak (2005), the realization of the importance of e-commerce development is dependent on the widespread use of the available technologies that enables people and organizations across the whole spectrum of social activities to become much more effective and productive, and not necessarily the size of the information and communication technology (ICT) sector itself. Ong (2003) argued that trust is a major issue when it comes to e-commerce, particularly in relation to cross border environment.

In contrast, according to recent statistics, the gap between developing economies and developed economies is narrowing down, but slowly (UNCTAD, 2007). UNCTAD reported that despite the fact that in 2002 developed economies internet penetration was ten times higher than the developing economies; in 2006 this was only 6 times (figure 1).

Despite the world usage of the internet increasing to approx 250% from 2000 to 2007 (IWS, 2007), the Middle Eastern countries usage growth reached as high as 920% growth. However, the percentage of population penetration was 17.4%, remains low in comparison with developed and industrial countries such as Europe (43%), North America (71%) and Australia (73%). In terms of e-commerce, however, there was a noticeable increase in purchasing products and services online. For example, according to the Arab Advisors Group (AAG), Saudi Arabia's Internet users have spent over $3.28bn in B2C e-commerce during 2007, and 48.36% of Internet users in Saudi Arabia have been reported purchasing products and services online and through their mobile handsets over a 12 months period (AMEinfo, 2007d). Also AAG revealed that UAE internet users spent over $1.15b on e-commerce, and that 51.2% of UAE Internet users have purchased products and services online and through their mobile handsets over a 12 months period, and that the majority of UAE e-commerce users make their payments by credit cards, followed by using bank account transfer methods (AAG, 2007). Based on a report by AAG (Zawya, 2008b), B2C e-commerce expenditure exceeded $4.87b in Kuwait, Saudi Arabia, the UAE, and Lebanon, based on major online and face to face surveys. AAG revealed that the total number of e-commerce users in these four countries exceeded 5.1 million users in 2007, and that UAE had the highest penetration at 25.1%, compared with Saudi Arabia (14.3%), Kuwait (10.7%) and Lebanon (1.6%) of the total population.

Figure 1 Internet penetration (Source: UNCTAD, 2007)



On the other hand, there are many activities steadily taking place in the region in the area of ICT implementation and infrastructure. For example, MBC Group, the Dubai-based Middle East Broadcasting Centre, has implemented an innovative network that would help MBC to distribute data, live and pre-recorded video content over a Cisco networking architecture (AmeInfo, 2007a). Another example, is the Kuwaiti Wataniya Telecom teaming up with Cisco to deliver a high-speed connectivity to the Kuwaiti corporate B2B market (AmeInfo, 2007b). Further, JAFZA (Dubai, Jebel Ali Free Zone) appointed Cisco as its technology consultant to formulate its overall technology strategy for its AED 2 billion Convention Centre Complex, in order to further strengthen its position as a global dynamic logistics and business hub that drives the industrial development of Dubai (AmeInfo, 2007c). Furthermore, the UAE is to boost e-government plans to provide 90% of the services online by 2010 (GulfNews, 2008f); the Airlines in the region achieved as high as 95% e-ticketing penetration (GulfNews, 2008g); and an agreement between Microsoft and Emirates group to develop IT innovation lab to develop creative solutions for airline and travel customers around the world (GulfNews, 2008h).

ICT growth rate is also on the increase. The average annual ICT growth rate has reached over 35% in the Middle East region, UAE amongst the highest broadband user penetration rates in the middle east (31.5%) and more than 50% of population has internet access (GulfNews, 2008f). According to GulfNews (2008a), computer shipment including desktops and notebooks increased 35% in 2007 compared with 2006, with UAE as the fastest growing market, followed by Kuwait, Oman and Bahrain. Moreover, major investments in the region have been carried out in enterprise application software (GulfNews, 2008b) reaching a growth of 39% year-on-year increase in spending in 2006, with UAE's ICT sector expected to hit Dh 8.8 billion by 2011 (CzechTrade, 2008).

The UAE government encourages faster e-commerce development in Gulf States (Media Eye, 2007) and the first Gulf e-commerce forum took place in 2007 (MenaFn, 2007). The Gulf PC market value is expected to reach $3.3b in 2008, an increase of approx 27% in comparison with $2.6b in 2007 (GulfNews, 2008i). Further, innovative wireless technologies are taking momentum in the region. RIM (Research in Motion), the designer, manufacturer and marketer of innovative wireless solutions for the worldwide mobile communications market participated in GITEX 2007 in Dubai and promoted one of their products, BlackBerry solution in the region (Zawya, 2008a).

Dubai has made major efforts to establish itself as a hub for e-commerce (LowTax, 2007). In 2002 Dubai Government made significant progress through the e-government project in providing basic services, especially those offered to businesses, on the Internet. Some of the areas that were outlined for achieving improvement were changing management skills, new balance of power, need for IT professionals, and training of civil servants, computer literacy and Internet access. While in January 2007, Dubai International Financial Centre (DIFC) enhanced Data Protection Law through appointing a Data Protection Commissioner for consolidating international best practices already being adhered to by the DIFC. In September 2000 more than a hundred IT companies had been granted licences to operate in Dubai Internet City (DIC) including Microsoft, Oracle and Compaq, with investment of $250 million in the technology, e-commerce and media free zone. By mid-2004, the number of companies had risen to more than 500, reaching almost 1000 by October 2007 (LowTax, 2007). Many Internet and e-commerce applications appeared in Dubai in the last few years (LowTax, 2007) such as: Magrudy's Bookshop, the first book shop in the Middle East to provide secure e-commerce services to a global market over the Internet; Dublin based Misys International Banking Solutions provides online banking solutions to one of Dubai's leading online financial services providers; The Dubai Government encourages e-commerce and broadening the focus of e-government. The Dubai e-Government portal offers a wide range of services for visitors, residents, and businesses, including access to permit and visa renewal services, commerce and business related facilities, and company registration and tax department links; the global logistics firm DHL announced a joint venture with the Dubai customs authorities for the creation of an electronic customs clearance system for streamlining procedures and improve general business efficiency in the United Arab Emirates and beyond.

Economist Intelligence Unit (EIU) and IBM worked together to produce an e-readiness rankings model of the world's largest economies. The model shows that some countries are more e-readiness consistent compared with others. As shown in table 2, the UAE was ranked higher than other gulf and Arab countries, at 33rd position in 2007, but dropping 2 places to 35th in 2008. Saudi was ranked 46 in both 2007 and 2008; while Egypt was ranked 58th in 2007 and gone up 1 place to 57th position in 2008 (EIU, 2008).

Table 2 2008 E-readiness rankings (Selected & adapted from: EIU, 2008)		
2008	**2007**	**Country**
1	2	*USA*
8	7	*UK*
12	13	*Canada*
35	33	*UAE*
46	46	*Saudi Arabia*
53	52	*Jordan*
56	56	*China*
57	58	*Egypt*
67	66	*Algeria*
70	69	*Iran*

FACTORS DRIVING E-COMMERCE IN THE ARAB REGION

According to an AAG report, a wide gap exists between the leading Arab nations and the rest of the world in terms of the online reporting on financial performance, social involvement as well as websites usability. The report pointed out that with the rise of national Internet penetration rates, information availability and transparency, the quality of the corporate online presence, and the usability and interactivity of their websites would be of vital importance (Business Maktoob, 2007). The Gulf region has the potential to play a very important role in the global economy and in contributing to software and technology development, according to Bill Gate (GulfNews, 2008d). However, there is a need for upgrading the region software from 'Gulf 1.0 to Gulf 2.0' and for educational institutions to prepare the students for the knowledge based economy, as stressed by Kito de Boer, Managing Director of MacKinsey in the Middle East (GulfNews, 2008e).

The investigation of the effect of culture including its possible positive or negative contribution on the growth of e-commerce in the region will be of benefit (Ashrafi, Yasin, Czuchry & Al-Hinai, 2007). On the other hand, from the education angle, the constant modification and updating of educational courses will become a vital strategy to ensure high quality of the educational programs developed, with a particular focus on state of the art IT resources at all levels (Fox, 2007; MOHE, 2007). Moreover, there are several factors that contribute to the readiness to participate positively in e-commerce or the digital economy, including ICT hardware and physical infrastructure; ICT level of use; ICT literacy, training and education; set policies and legal issues (Peters, 2005). In terms of IT education, it has become a major priority for the UAE (UAE Interact, 2007), and eventually this will be reflected on the level of use. On the other hand, the application of ICT to areas such as health, government or education can certainly contribute to the ICT literacy and level of use and can, on the long term assist e-commerce development.

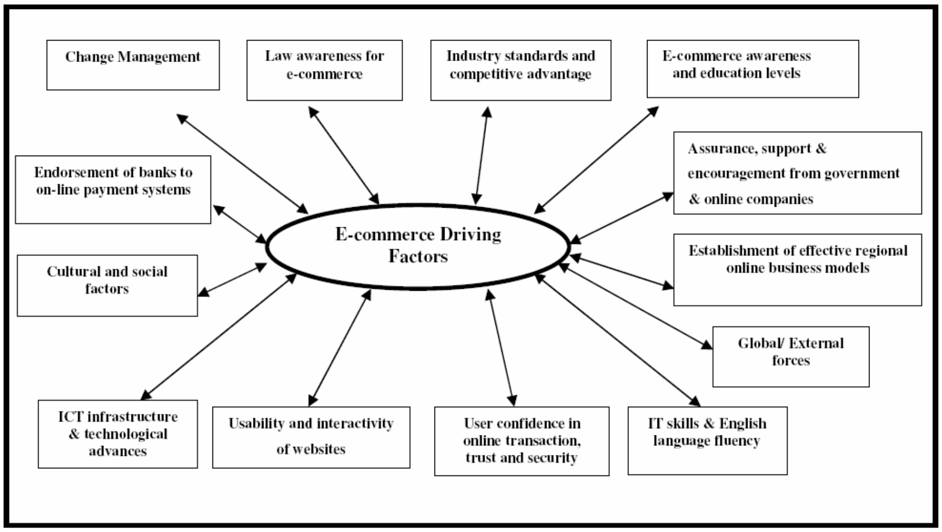
Nitin, William & Roberson (2003) also argued that the key to reducing intangibility surrounding e-commerce is though the creation of tangible cues that the customer can relate to. Many internet users may be less proficient at searching for product information and would generally rely on brands. Therefore, there is also a need to focus on creating a positive and a reliable brand name. Ilfeld & Winer (2001), in exploring the issue of the decision-making process made by consumers' website choices, suggested that high brand awareness is essential for an Internet firm's survival.

Interactive shopping might affect consumer behavior, and consequently might have an effect on retailer and manufacturer revenue (Bradlow & Schmittlein, 2000). They noted that technological advances offered consumers opportunities to search and compare between products and prices in addition to searching for more differentiated products that better fit their needs. Englis & Solomon(2000) explored the decision-making process behind consumer's website choices and the relative effects of the communication channels which aid this process. One of the biggest problems with buying online was the lack of automated inventory and warehousing system (Kaplan & Sawhney, 2000) which hindered the efficiency and speed of meeting the customer needs. The use of Radio Frequency Identification (RFID) technology can deal with these problems and help to rapidly change the way businesses track their inventory. This technology has been used by Walmart, Tesco and U.S. Department of Defense, and already showing benefit (Glover & Bhatt, 2006).

Based on the above literature review, we have developed a list of thirteen driving factors for e-commerce development (figure 2):

E-commerce awareness and education levels
Government support
Law awareness for E-commerce
ICT infrastructure & Technological Advances
Endorsement of Banks to on-line payment systems
Change Management (from traditional approaches to digital approaches)
Establishment of effective regional online business models
Global/ External forces (eg. The September 2008 Global Financial System Turmoil)
IT skills & English language fluency
Cultural and Social factors
User confidence in online transaction, trust and security
Usability and interactivity of websites
Industry standards and competitive advantage

Figure 2 E-Commerce Driving Factors



CONCLUSIONS AND FUTURE WORK

This paper has reviewed and discussed several e-commerce issues with a particular focus on the Gulf region as part of a study in e-commerce potential and implementation in the gulf area. We believe that there are several main driving factors that may contribute to e-commerce evolvement and success in the Gulf region and the Arab World (see Figure 2). These include (but not limited to): e-commerce awareness and education levels; change management; IT skills development and language fluency; ICT infrastructure; law awareness for e-commerce; user confidence in online transaction, trust and security issues; usability and interactivity of websites; assurance and encouragement from government & online companies; establishment of effective regional online business models; social and cultural factors; endorsement of banks to on-line payment systems; global and external factors; industry standards and competitive advantage.

In particular, the Governments' role is essential in promoting e-commerce in the region in terms of ensuring coherent IT and e-commerce policy for consumer protection; secure, transparent, predictable, and enabled environment; support, coordination, collaboration, and cooperation. Traditional companies' strategies should be based on the view that e-commerce does not necessarily replace existing ways of doing business, but enhances them and increases business efficiency and effectiveness, taking into account that its evolution process may take time and may go through several phases or stages of growth. It may, however, require major business processes improvement.

This paper has also raised some questions in relation to the global financial crisis as well as security issues that occurred at the time of writing this review. There is a clear lack of research in the area of e-commerce in the region and consequently, further research is needed to investigate e-commerce development and success factors. This may include case studies of companies and organizations operating in the region as well as internet users in order to investigate the above mentioned factors and possibly additional ones that are closely related to e-commerce implementation, success or failure to guide e-commerce future in the region.

REFERENCES

AAG (2007) Arab Advisor Group. *UAE Internet users spent over US$ 1.15 billion in B2C e-commerce over the past 12 months*. Retrieved November, 2008 from http://www.arabadvisors.com/Pressers/presser-090907.htm

AmeInfo (2007a). *MBC Group deploys Cisco Unified Communications Solution*. Retrieved November, 2008 from http://www.ameinfo.com/139777.html

AmeInfo (2007b). *Wataniya Telecom teams up with Cisco to deliver uninterrupted high-speed connectivity to the Kuwaiti corporate business to business market*. Retrieved November, 2008 from http://www.ameinfo.com/136249.html

AmeInfo (2007c). *Jafza appoints Cisco to build an intelligent infrastructure for its Convention Centre Complex*. Retrieved November, 2008 from http://www.ameinfo.com/133052.html

AmeInfo (2007d). *Saudi Arabia's Internet users spend over $3.28bn in B2C e-commerce during 2007.* Retrieved November, 2008 from http://www.ameinfo.com/143272.html

Ashrafi, R., M. Yasin, A. Czuchry & Y. Al-Hinai (2007). E-commerce practices in the Arabian Gulf GCC business culture: utilization and outcomes patterns. *International Journal of Business Information Systems,* 2(4), 351-371.

Benjamin,S. & T. Wigand (1995). Electronic markets and virtual value chains on the information superhighway. *Sloan Management Review*, winter, 62-27.

Bertschek, I., H. Fryges, & U. Kaiser (2006). B2B or not to be, does B2B E-commerce increase labor productivity. *International journal of the Economics of Business*, 13(3), 387-405.

Bradlow,T. & C. Schmittlen (2000).The little engines that could, modeling the performance of WWW search engines. *Marketing Science*, 19(1), 34-62.

Business Maktoob (2007). *Analysis reveals massive gap between the Arab countries and the industrial nations*. Retrieved November, 2008 from http://business.maktoob.com/reportdetails-20070212072667-Analysis_reveals_massive_gap_between_the_Arab_countries_and_the_industrial_nations.htm

Chan, C. & P. Swatman (2004). B2B E-commerce stages of growth: the strategic imperatives. *Proceedings of the 37th Hawaii International Conference on System Sciences.*

Chase Paymentech (2007). Glossary of electronic payment processing terms. Retrieved November, 2008 from http://www.paymentech.com/glogloe.do

Cisco (2007). *Annual report 2007*. Retrieved January, 2008 from cisco.com/web/about/ac49/ac20/downloads/annualreport/ar2007/pdf/cisco_ar2007_complete.pdf

Cisco News (2007). *Cisco managed solution improves Grocer's debit/credit transactions*. Retrieved November, 2008 from http://newsroom.cisco.com/dlls/partners/success_stories/2007/pss_09-14.html

Cohen, S. & G. Kallirroi (2006). E-Commerce investments from an SME perspective: Costs, benefits and processes. *Electronic Journal of Information Systems Evaluation.* 9(2). 45 -56. Retrieved November, 2008 from http://www.ejise.com/volume-9/v9-iss-2/cohen_and_kallirroi.pdf

Dans, E. & K. Pauwels (2001). Internet marketing the news: Levering brand equity from marketplace to marketspace. *Journal of Brand Management*, 8(4-5), 303-314.

Drucker, P. (2002). *Managing in the next society*. New York: Truman Talley Books.

EIU (2008). Economist Intelligent Unit: *E-readiness rankings 2008, Maintaining momentum* Retrieved November, 2008 from http://www-935.ibm.com/services/us/gbs/bus/pdf/e-readiness_rankings_april_2008_final.pdf

Elbeltagi, I. (2007). E-commerce and globalization: An exploratory study of Egypt. *International Journal Cross Cultural Management*, 14(3), 199-200.

Englis, B. & M. Solomon (2000). Life/Style online: A Web-based methodology for visually-oriented consumer research. *Journal of Interactive Marketing*, 14(1), 2-14.

Expedite Email Marketing, (2004) Internet marketing glossary. Retrieved January, 2008 from http://www.expedite-email-marketing.com

Fox, W. (2007). *The United Arab Emirates: Policy choices shaping the future of public higher education*. Center for Studies in HE. CSHE.13.07.

Glover, B. & H. Bhatt (2006). *RFID Essentials*. O'Reilly.

GulfNews (2008a). *PC shipments into gulf rise 35% in third quarter*. Retrieved November, 2008 from http://archive.gulfnews.com/articles/08/01/08/10180240.html

GulfNews (2008b). *Regional software solutions market will touch $500m*. Retrieved November, 2008 from http://www.gulfnews.com/business/Technology/10184109.html

CzechTrade (2008). *UAE's ICT sector to hit Dh8.8b by 2011*. Retrieved November, 2008 from http://www.czechtradeoffices.com/en/united-arab-emirates/news/uae-s-ict-sector-to-hit-dh8-8b-by-2011-13896/

GulfNews (2008d). *A few minutes with Bill Gates*, Interview, January, 27, 2008.

GulfNews (2008e). *Gulf needs to create 4m jobs in next 20 years*. January, 28, 2008.

GulfNews (2008f). *UAE to boost e-government plan*. January, 28, 2008.

GulfNews (2008g). *Airlines in Mena achieve 80% e-ticketing levels*. January, 28, 2008.

GulfNews (2008h). *Microsoft and Emirates to set up Innovation Lab in Dubai*. January, 30, 2008.

GulfNews (2008i). *Value of Gulf PC market touches $3.3b*. February, 1, 2008.

GulfNews (2008j). *UAE banks restrict ATM usage abroad*. September, 16, 2008.

GulfNews (2008k). *Meltdown will hurt tech firms too*. September, 27, 2008.

Hammond, K. (2001). E-commerce 2000-2010: What expert predict. *Business Strategy Review*, 12(1), 43-50.

Hofacker, F. (2001). *Internet Marketing*. NY: Wiley.

Hoffman, L. & P. Novak (2000). How to acquire customers on the Web. *Harvard Business Review*, May-June, 179-188.

Ilfeld, J. & S. Winer.(2001). *Creating Website traffic*. HAAS School of Business, University of California at Berkeley.

IAW (1998) Internet Arab World. *E-commerce in the Arab World*. Retrieved January, 2008 from http://www.library.cornell.edu/colldev/mideast/intmid1.htm

IGW (2007) InterGovWorld. *The top 10 developments in the world of e-commerce*, 07-12-2007. Retrieved January, 2008 from http://www.intergovworld.com/article/badbf45b0a0104080180bfe1b7c9af9b/pg1.htm

IWS (2007) Internet World Stats. *Internet usage statistics*. Retrieved December, 2007 from www.internetworldstats.com/stats.htm

Kaplan, S. & M. Sawhney (2000). E-Hubs, the new B2B marketplace, *Harvard Business Review*, May-June, 97-103.

Laudon, K. & J. Laudon (2007). *Management Information Systems: Managing the digital firm*. NJ: Prentice Hall.

Lin, H. (2007). The impact of the Website, Quality dimensions on customer satisfaction in the B2C commerce context. *Total Quality Management& Business Excellence*, 18(3), 363-378.

LowTax (2007). *Dubai e-commerce*. Retrieved November, 2008 from http://www.lowtax.net/lowtax/html/dubai/jdbecom.html

Media Eye (2007). *UAE MOE advocates boosted e-commerce sector in Gulf States*. Retrieved January, 2008 from www.mediaeyeme.com

MenaFn (2007). *Bahrain to host Gulf e-commerce Forum Nov. 26*. Retrieved January, 2008 from www.menafn.com

MOHE (2007). Educating the next generation of Emiratis: A master plan for UAE higher education. Office of HE Policy and Planning. In W. Fox (2007). *The United Arab Emirates: Policy choices shaping the future of public higher education*. Center for Studies in HE. CSHE.13.07

Molla, A. & R. Heeks (2007). Exploring e-commerce benefits for businesses in a developing country. *The Information Society*, 23(2), 95-108.

Ngai, E.W.T. & F.K.T. Wat (2002). A literature review and classification of electronic commerce research. *Journal of Information and Management*, 39(5), 415-429.

Nitin, N., J. William & B. Roberson (2003). What really works. *Harvard Business Review*, 81(7), 42-53.

Numberger, S. & C. Rennhak (2005). The future of B2C e-commerce. *Electronic Markets*, 15(3), 269-282.

Ong, C. E. (2003). B2C E-commerce trust in redress mechanism (Cross border issues). *Informing Science Proceeding 2003*. Retrieved November, 2008 from http://www.informingscience.org/proceedings/IS2003Proceedings/docs/018Eang.pdf

Peters, T. (2005). *E-ready for what? E-readiness in developing countries: Current status and prospects toward the millennium development goal*. InfoDev, bridges.org. Retrieved November, 2008 from www.bridges.org/publications/18

Schneider, G. (2002). *Electronic Commerce* (Third Edition). Thomson, pp13-14.

Soopramanien, D., F. Fildes & A. Robertson (2007). Consumer Decision Making, E-commerce and perceived risks. *Applied Economics*, 39(17), 2159-2166.

UAE Interact (2007). *Education Section*. Retrieved December, 2007 from http://www.uaeinteract.com/education/

UNCTAD (2002). *Electronic commerce strategies for development: The basic elements of an enabling environment for e-commerce*. Geneva. Retrieved November, 2008 from http://r0.unctad.org/ecommerce/event_docs/geneva_strategies_issues.pdf

UNCTAD (2007). *Science and Technology for Development: the New Paradigm of ICT*. Information Economy Report 2007-2008. Retrieved November, 2008 from http://www.unctad.org/en/docs/sdteecb20071_en.pdf

USCBN (2007). *Quarterly retail e-commerce sales* (3rd quarter 2007). US Census Bureau News Report. Washington DC. USA

Weill, P. & M. Vitale (2001). Place to Space: migrating to e-business models. Harvard Business School Press. In . E.Turban, J. Lee & D. KIng (2008). *Electronic Commerce 2008: a managerial perspective*. NJ: Pearson Prentice Hall.

Zawya (2008a). *RIM focused on Blackberry growth in the Gulf region*. Retrieved January, 2008 from www.zawya.com

Zawya (2008b). *B2C e-commerce volume exceeded US$ 4.87 billion in Kuwait, Lebanon, Saudi Arabia and UAE in 2007*. Retrieved November, 2008 from http://www.zawya.com/Story.cfm/sidZAWYA20080205052028/secIndustries/pagRetail

AN EVOLUTIONARY METHOD FOR THE MINIMUM TOLL BOOTH PROBLEM: THE METHODOLOGY

Lihui Bai, Valparaiso University
Matthew T. Stamps, University of California, Davis
R. Corban Harwood, Washington State University
Christopher J. Kollmann, Concordia Seminary

ABSTRACT

This paper considers the minimum toll booth problem (MINTB) for determining a tolling strategy in a transportation network that requires the least number of toll locations, and simultaneously causes the most efficient use of the network. The paper develops a methodology for using the genetic algorithm to solve MINTB and presents the algorithm GAMINTB. The proposed method is tested and validated through a computational study with six example networks. Additional numerical test discovers some interesting properties for the proposed method, and provides guidelines for further application of the GAMINTB.

INTRODUCTION

Today, traffic congestion is a pressing issue for society and a major concern for urban planners. The 2007 Urban Mobility Report (Schrank & Lomax, 2007) states that congestion in the U.S. caused 4.2 billion hours of travel delay and 2.9 billion gallons of wasted fuel for a total cost of $78 billion in 2005. Although congestion pricing has been proposed by transportation economists as a means of reducing congestion for over 80 years (see e.g., Pigou (1920) and Beckmann, McGuire & Winston (1956)), it is only recently that this idea has been implemented in practice. Examples include the "Electronic Road Pricing" (formerly "Area Licensing Scheme") program in Singapore implemented in 1975, a toll ring in Bergen, Norway implemented in 1986, and subsequently two toll rings in Oslo and Trondheim, respectively. More recently, London introduced a $5 (now raised to $8) daily fee on cars entering the city center in February 2003. In the United States, the Federal Highway Administration (DeCorla-Souza, 2003) has been funding toll pricing projects in cities such as San Diego, Houston, and Seattle under the Congestion Pricing Pilot Program established by Congress since 1991.

This paper focuses on determining tolls with the primary objective of inducing travelers, who are encouraged only by their own travel costs, to choose routes that would collectively benefit all travelers and use the transportation system resources more efficiently. The secondary objective is to find tolls that require the least number of collection facilities or toll booths. The problem was first introduced in Hearn and Ramana (1998) and was referred to as the minimum toll booth problem (or MINTB). When compared to the marginal social cost pricing (MSCP) advocated by transportation economists, the MINTB solution often requires substantially fewer number of toll booths. For example, the well documented Sioux Falls network (Leblanc, Morlok, & Pierskalla, 1975) with 24 nodes and 76 arcs requires to toll on all 76 arcs under MSCP, but only 32 under the MINTB solution. For the Hull network (Florian & Guélat, 1987) with 501 nodes and 798 arcs, the MINTB solution requires 39 toll booths and this saves 224 compared to MSCP. Considering that operating a manned toll booth costs $180,000 per year (Todd, 2005), MINTB solutions would save approximately eight and forty million dollars per year for Sioux Falls and Hull, respectively. While today's electronic tolling through pre-registered cards, cameras and other systems can be available at a reasonable cost, which may affect the popularity of manned toll booths, safety concerns still exist when drivers have to merge to electronic tolling lanes frequently. Indeed, the latter may pose a major obstacle to the public and the transportation agency's acceptance of the tolling decisions when many toll booths are required in the proposal.

As stated, the MINTB problem was first introduced as a mixed integer linear programming in Hearn and Ramana (1998), where toll pricing problems are defined as finding the optimal tolling strategies that will make the transportation system run most efficiently and will simultaneously optimize a secondary tolling objective. In MINTB, the tolling objective is to minimize the total number of required toll facilities, i.e., toll booths. Other examples of tolling objectives include minimizing the total toll revenues (MINREV) and minimizing the maximum toll on a network (MINMAX). While toll pricing problems such as MINREV and MINMAX are linear and easier to solve for practical networks, efficient solution for MINTB has remained relatively unexplored in the transportation and management science literature.

An investigation of MINTB in Bai's dissertation (Bai, 2004) shows that MINTB is NP-complete. This, from a theoretical perspective, implies that solving MINTB optimally is very difficult. In fact, numerical experiments in Bai (2004) notes that general purpose solvers for mixed-integer programs, such as CPLEX, have been unable to produce optimal solutions for MINTB for a moderate network with 501 nodes and 798 arcs. This confirms the challenge of solving MINTB from an empirical perspective. Thus, solutions to MINTB resort to heuristic methods. For example, Bai (2004) and Bai, Hearn, & Lawphongpanich (2008) solved MINTB using a dynamic slope scaling procedure (DSSP), which is shown to be effective on larger networks such as Sioux Falls, Hull, and Stockholm. She also proposed a DSSP-based neighborhood function and used a simulated annealing method to solve MINTB (Bai, 2004). However, these local search methods may not always reach a global optimal solution.

Genetic algorithms (GA) are well known for their ability to sort through a variety of local optimal solutions until they converge to a global optimal solution (Shepherd & Sumalee, 2004). In the literature, genetic algorithms have been used to solve a variety of toll pricing problems. Zhang (2003) uses a GA to determine the optimal toll locations for a second-best link-based congestion pricing problem, where a simulated annealing method was used to solve for the optimal toll levels. Sumalee (2004) uses a genetic algorithm in conjunction with a branch-tree framework to create a toll set in a cordon-pricing scheme.

Thus far, a genetic algorithm has not been used to solve MINTB and this is the main purpose of the present paper. We propose a genetic algorithm for MINTB (GAMINTB), focusing on the methodology development and validation through small examples. The algorithm first randomly generates binary vectors, each of which represents whether or not a toll booth is used on any arc for a given network. After using a linear program solver to determine feasibility and calculate optimal toll levels, GAMINTB divides all randomly generated toll vectors into two groups: feasible and infeasible groups; and then ranks them according to their feasibility and total number of toll booths. Based on this ranking, the algorithm randomly selects toll vectors to become "parents" where higher ranked toll vectors have higher probability of being selected. The parent toll vectors then reproduce the new generation of toll vectors following an alteration process consisting of "crossover" and "mutation." The new population is again evaluated and ranked, and the process continues until the algorithm reaches the specified number of generations. Finally, the solution with the minimal number of toll booths is returned as the final optimal solution.

For the remainder, the next section provides mathematical formulations for the traffic assignment and the MINTB problems. Then, the genetic algorithm is introduced, followed by preliminary computational results. Finally, we discuss an alternative implementation for future study before offering concluding remarks.

PROBLEM DESCRIPTION

This section formulates the minimum toll booth problem using network notation. The transportation network is represented by sets of arcs and nodes corresponding to roads and intersections, respectively. An origin-destination (O-D) pair (also known as a commodity) is a pair of nodes on which a user (traveler) in the system must begin and end. There is usually more than one path available for each user to travel along for a given commodity. Mathematically, let *G=(N,A)* denote a network with the set of nodes *N* and set of arcs *A*. The set of origin destination pairs is denoted as *K* where *o(k)* and *d(k)* are the origin and destination nodes for the O-D pair *k*. The demand for an O-D pair *k*, denoted as D_k, is simply the number of travelers going from the origin to the destination.

Let $x^k \in R_+^{|A|}$ represent the flow vector for commodity *k*. The vector x^k is feasible if it satisfies $Ux^k = b_k$ where *U* is the node-arc incidence matrix for *G=(N,A)* and $b_k \in R^{|N|}$ is the demand vector defined by

$$(b_k)_i = \begin{cases} D_k & \text{if } i = o(k) \\ -D_k & \text{if } i = d(k) \\ 0 & \text{otherwise} \end{cases}$$

The sum of all flows x^k over all O-D pairs that pass over a given arc *a* is denoted $v_a = \sum_k x_a^k$. It follows, then, that an aggregate flow vector *v* is feasible only if $v = \sum_k x^k$, $Ux^k = b_k$, and $x^k \geq 0$, $\forall k \in K$. Finally, the cost of each path can be measured by the amount of time it takes to travel from the origin to the destination for each user. Let *s(v)* be the travel cost vector for a given flow *v* in the network.

The User Equilibrium Model

The *User Equilibrium* (UE) model is used to describe the behavior of users on a given traffic network when every user chooses the shortest path available (See e.g., Florian and Hearn (1995)). Equivalently, a network is at UE only if the cost of every utilized path for each O-D pair is less than or equal to the cost of every non-utilized path for the respective O-D pair. In mathematical terms, a network *G*(*N,A*) is at UE with $\tilde{v}$ being the equilibrium flow if it satisfies the following condition:

$$s(\tilde{v})^T(v - \tilde{v}) \geq 0, \quad \forall v \in V,$$

where *V* is the set of all feasible aggregate flow vectors for *G*(*N,A*):

$$V = \{v \mid v = \textstyle\sum_k x^k, Ux^k = b_k, x^k \geq 0, \forall k \in K\}.$$

Under UE, each user chooses the path for their own interests. Collectively, UE may not be in the best interest for the system as a whole. If too many travelers select the same path, congestion develops. Typically, the per capita cost of traveling along an arc increases as flow increases, slowing down the network as a whole, and hence, the original, low cost path becomes high cost. An example network is offered later in this section to illustrate the user equilibrium model.

The System Optimal Model

The system optimal model describes the network when it is working as efficiently as possible. If the total travel cost of a given network is minimal, then the network flow, $\bar{v}$, is at

System Optimal (SO). Mathematically, a network $G(N,A)$ is at SO with $\bar{v}$ being the system optimal flow if $\bar{v} \in V$ satisfies

$$s(\bar{v})^T \bar{v} = \min\{s(v)^T v \mid v \in V\}.$$

In the SO model, the network, as a whole, is operating the most efficiently. A network at SO is rarely at UE as well. This is because some travelers may pay higher costs than they would in UE, as SO minimizes the average cost per user.

The Tolled User Equilibrium Model

Imposing tolls on a network adds a monetary component to the cost function in terms of time. The new cost for each arc, composed of time and money, is called the *generalized* cost. Let β_a denote the toll on arc *a*, then the *generalized* cost for arc *a* is expressed as ($s_a(v_a) + \beta_a$). If a traveler has the choice between two paths where Path 1 takes less time than Path 2, but Path 1 also costs money, the traveler may be more likely to take Path 2 depending on how much the toll on Path 1 is. By adding a toll vector β to a network, the user equilibrium, $\tilde{v}$ goes from

$$s(\tilde{v})^T (v - \tilde{v}) \geq 0, \quad \forall v \in V$$

to

$$(s(\tilde{v}_\beta) + \beta)^T (v - \tilde{v}_\beta) \geq 0, \quad \forall v \in V,$$

where the new equilibrium $\tilde{v}_\beta$ is called the *Tolled User Equilibrium* (TUE).

The TUE model describes the network behavior where tolls can be added such that $\tilde{v}_\beta = \bar{v}$, i.e., the tolled user equilibrium reproduces the system optimal flow. In other words, there exist tolls that allow the network to operate so that when every user does what is best for oneself, he/she is also doing what is best for the system.

Toll Pricing Problems

A toll vector is said to be valid if it causes TUE to equal SO. To find the set of all valid toll vectors for a given network, refer to Hearn and Ramana (1998). It states that a toll vector is valid if there exists a ρ vector for each O-D pair *k* satisfying the following:

$$s(\overline{v})+\beta \geq U^T \rho^k,\ \forall k \in K$$
$$(s(\overline{v})+\beta)^T \overline{v} = \sum_k b_k^T \rho^k.$$

These are essentially the Karush-Kuhn Tucker (KKT) conditions that ensure $\tilde{v}_\beta = \overline{v}$. Moreover, from this set of valid toll vectors, a traffic planner can select an optimal toll for any specific objective such as minimizing total revenue or in the case of this paper, minimizing the total number of toll booths.

An Illustrative Example

Figure 1: An Example Network

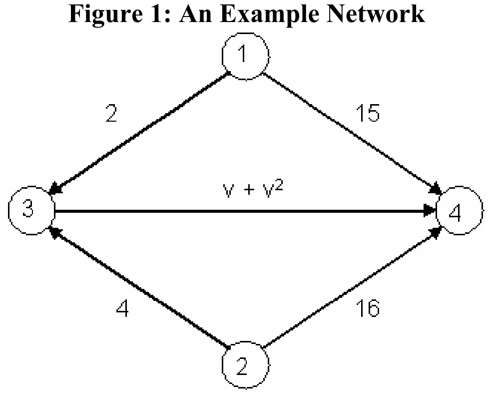
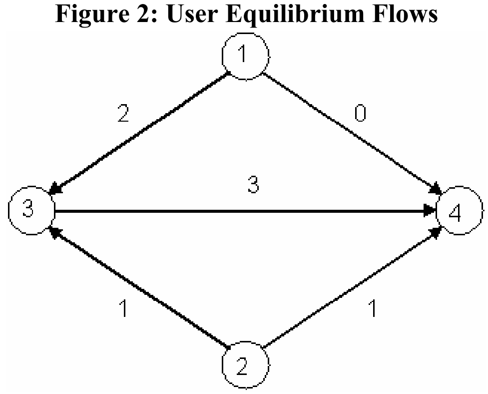


Figure 1 is a 4-node, 5-arc network with the given costs. Note that the cost on the arc connecting nodes 3 and 4 is dependent on the flow. The cost vector for this network is $s(v) = <2,4,16,15,v_5+v_5^2>$. There are two O-D pairs: 1 to 4 and 2 to 4. The demand on each is 2 cars.

Figure 2: User Equilibrium Flows



The network is at UE when v =< 2, 1, 1, 0, 3 >, as illustrated in Figure 2. To verify the flow is in UE, first calculate the new cost vector $s(v)$ =< 2, 4, 16, 15, 12 >. Because, for the first O-D pair, the cost for path 1-4 is 15 and the cost for path 1-3-4 is 14, path 1-4 is not utilized. For the other O-D pair, the cost for both paths is 16, so they are both utilized. Note that the total travel cost for the system is 60 units of time

Figure 3: System Optimal Flows

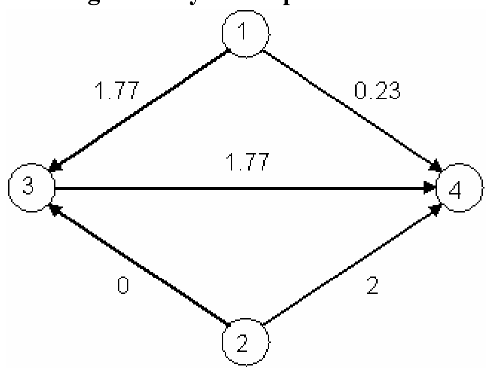
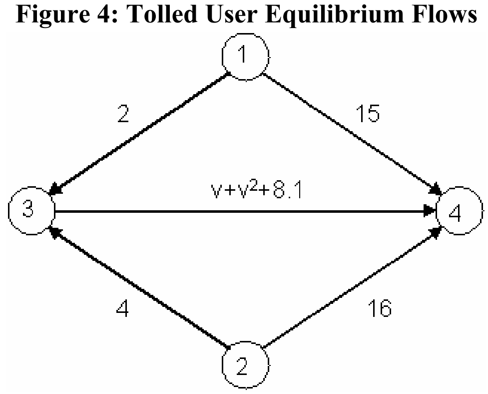


Figure 3 illustrates the network flow at SO. The total travel cost for this flow on the network is approximately 47.67 units of time. This is the minimum travel time for the whole system.

Figure 4: Tolled User Equilibrium Flows



Finally, in Figure 4, a toll of 8.1 has been added to Arc 5. This toll raises the *generalized* costs for paths 1-3-4 and 2-3-4 such that when the network is at SO, the travel cost for path 1-4 is equal to the cost on path 1-3-4, so this O-D pair is at UE. Furthermore, the travel cost on 2-3-4 becomes larger than the travel cost on path 2-4 so this O-D pair is at UE as well. Since both O-D pairs satisfy UE, the new TUE is the same as the SO.

The Minimum Toll Booth Problem (MINTB)

As stated previously, the MINTB objective is to place the fewest number of toll booths on a network so that when the system is at TUE, it is also at SO. Mathematically, let y be a binary vector which corresponds to β so that

$$y_a = \begin{cases} 0 & \text{if } \beta_a = 0 \\ 1 & \text{if } \beta_a > 0 \end{cases}$$

Then, the MINTB problem can be formulated as:

$$\begin{aligned} &\min \sum_{a \in A} y_a \\ &\text{s.t. } s(\bar{v}) + \beta \geq U^T \rho^k, \ \forall k \in K \\ &\quad (s(\bar{v}) + \beta)^T \bar{v} = \sum_k b_k^T \rho^k \\ &\quad 0 \leq \beta_a \leq M y_a, \ \forall a \in A \\ &\quad y_a \in \{0,1\}, \quad \forall a \in A, \end{aligned} \tag{1}$$

where M is an arbitrarily large constant. Note that the first two constraints ensure β is a feasible (valid) toll vector and the third condition states that y_a=1 when β_a>0 and y_a=0 when β_a=0.

The MINTB problem is a mixed integer linear programming problem. In Bai (2004), the problem is shown to be NP-complete by reducing the partitioning problem, an NP-complete problem, to MINTB in polynomial time. Consequently, our focus in this paper is to develop heuristic algorithms to solve MINTB. In particular, we choose to study the genetic algorithm because it is well known for converging to global optimal solutions.

THE GENETIC ALGORITHM FOR MINTB (GAMINTB)

In the past few decades, genetic algorithms have become an increasingly popular heuristic method for problem solving. As one of several evolutionary optimization methods, the genetic algorithm (GA) uses ideas from natural selection to evaluate, breed, and filter through random solutions to obtain an optimal solution after many generations. Most genetic algorithms follow the basic steps described below.

1. *Initialization*: the algorithm begins with randomly generating an initial population of strands, often called chromosomes. Each strand is designed based on certain parameters which can be modified by the user, including the population size and the length of each strand.

2. *Evaluation:* this portion of the algorithm rates each chromosome by assigning it a fitness score, which reflects how well the decoded strand satisfies the constraints and optimizes the objective function. The strands are then ranked based on their respective fitness scores. Note that the means to deal with an infeasible solution include merely accepting it, penalizing it by lowering its fitness score, or completely eliminating it.

3. *Selection:* this process determines which strands from a population will be allowed to reproduce and create the next generation. There are a variety of ways to determine which strands move on, but most involve assigning each strand a probability. This selection probability can be uniform, but it is generally weighted in order to increase the likelihood that the fittest solutions will reproduce.

4. *Alteration:* the most popular methods of alteration are crossover and mutation. The crossover process is designed to take the traits from a set of parents, and pass them on to a new set of (ideally stronger) offspring. Although there are many ways to implement it, crossover typically involves taking a set of alleles from one parent, and switching them with the alleles of the other parent strand. Some forms of crossover produce two offspring, while others elect to produce one child, and allow one of the parent strands to continue on to the next generation. The mutation process, which ensures diversity among the population, usually involves randomly changing some of the numerical values in the offspring.

5. *Termination*: after reproducing, most genetic algorithms are designed to cycle through the evaluation, selection, and alteration using the new and altered strands as the new parent population. Genetic algorithms stop after a specified number of generations.

Next we present the customized genetic algorithm for the MINTB problem (GAMINTB). The algorithm begins by generating a population of size N chromosomes, which are used to represent the presence of a set of toll locations. The strands are of length $|A|$, where $|A|$ is equal to the number of arcs, and they are indexed so that the ith gene corresponds to the ith arc on the network. Each strand, y, is binary, where a 0 represents an arc that does not have a toll booth, and a 1 represents an arc that has a toll booth. For example, $y = (0, 1, 0, 0, 1)$ would represent a 5-arc network with tolls on arcs 2 and 5. The number of chromosomes can be changed in our program by modifying the value of N. The algorithm uses an initialization method comparable to Zhang's (Zhang, 2003) in order to prevent bias among the strands. A random number generator creates an $|A|$ length strand of random numbers between 0 and 1. These numbers are then rounded to the nearest integer (0 or 1). The process is repeated until the number of strands is equal to the specified population size N.

After the initial population of toll locations is produced, it must be evaluated in some way. The y vectors are ranked based on two criteria: feasibility and the number of toll booths. A y vector is feasible if there exist corresponding β and ρ satisfying the constraints in problem (1). For evaluation, the y's are divided into feasible and infeasible groups and then each group is ranked according to the fewest number of toll booths required. For the final rankings, they are regrouped so that feasible solutions are ranked higher than all of the infeasible solutions. To decide on feasibility, the GAMINTB keeps track of βs and ρs for each y. It is important to consider infeasible solutions because some of them may be very similar to the optimal solution, and might only require a slight alteration in order to become feasible. Because of the elaborate alteration process in the GAMINTB, the good qualities of an infeasible solution may be passed down. On the other hand, the ranking process makes the probability very low for infeasible solutions to continue on from one generation to the next.

After the chromosomes are ranked, a set of parents is chosen to produce the next generation. In nature, 'survival of the fittest' typically applies to reproduction. Although the strongest individuals have the best chance of survival, they are not the only ones that reproduce. Rather than settling on a uniform selection probability, GAMINTB bases a strand's selection probability on its ranking (fitness score). It uses a weighted selection probability comparable to the one used by Sumalee (Sumalee, 2004). The weighted probability of selection for a strand of rank i is

$$P(i) = \frac{2((N+1)-i)}{N(N+1)}, \quad (2)$$

where N is the population size. Thus, the likelihood of choosing the strongest strand (ranking index $i = 1$) for the breeding process is $2/(N+1)$, while the probability of choosing the weakest individual (ranking index $i = N$) is $2/(N(N+1))$. After the selection probability has been determined for each strand, the genetic algorithm uses a 'roulette wheel' selection process to determine the parents (See e.g., Sumalee, 2004). Each slot on the roulette wheel is assigned an interval between 0 and 1, and the upper and lower bounds are assigned by cumulatively stacking the selection probabilities in (2). A random number generator is then used to select the strands that will produce the subsequent generation. If the random number is between the bounds of a specified slot, that slot's respective chromosome will be chosen as a parent. The algorithm picks the nearest even integer to (RN) parents, where R is the reproduction rate defined by GAMINTB.

Once the parents are selected, they are randomly paired to create offspring. Since the GA relies on the passing down of good traits from generation to generation, each pair of parents should pass on each shared allele to one child. Because each y chromosome is a binary strand, each child can only receive a 0 or 1 for each unshared parent allele. The GAMINTB objective is to minimize the sum across the strand, so giving each child a 0 for each unshared parent allele seems logical. However, when implementing this crossover method, the GAMINTB converges to a strand of 0's, which is most likely infeasible, and decreases the variety of solutions. Giving the child a 1 in place of the 0 results in equally bad solutions. Since a critical feature of the GAMINTB is randomness, each child is given a random binary number wherever the parents disagree.

For mutation, the GAMINTB selects the nearest integer to (MN) chromosomes randomly from the entire population pool, where M is the rate of mutation. Being that each y chromosome is a binary strand, mutation is merely changing one randomly selected allele from 0 to 1 or 1 to 0 for a chromosome that is selected.

Research done by Ahuja, Orlin, & Tiwari (2000) proposes immigration as a useful tool to adding diversity among the strands. The idea of immigration is to only breed a percentage of the next generation, and fill the remaining slots with randomly generated strings. The GAMINTB incorporates this idea by replacing those strands not selected for reproduction with entirely new strands, following the method used in the initialization phase.

COMPUTATIONAL RESULTS

The purpose of our computational study is to validate GAMINTB through small example networks, as well as to investigate how parameter settings affect the performance of GAMINTB. In this section we present the results of GAMINTB tested for performance and implementational comparisons over six small example networks. The six examples are built upon two basic network structures, the 4-node diamond structure and the 5-node split structure shown in Figures 5 and 6, respectively.

The first three example networks were derivations of the 4-node diamond and the last three were derivations of the 5-node split. The demand for each network was randomly selected, while the cost function vector for each was decided by the Bureau of Public Roads (BPR) cost function $s_a(v_a) = (T_a + 0.15(\frac{v_a}{C_a})^{P_a})$.

Here $s_a(v_a)$ is the cost function, v_a is the variable traffic flow on arc a, T_a is the free travel time constant, C_a is the capacity, and P_a is the delay time factor for arc a. For each network, the constants were randomly selected from uniform distributions as follows: $T_a \in [1, 10]$, $C_a \in [10, 25]$, $P_a \in \{0, 1, 2, 3, 4\}$.

Figure 5: A 4-node 5-arc network

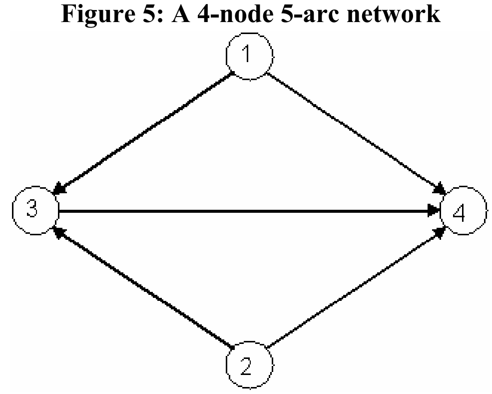
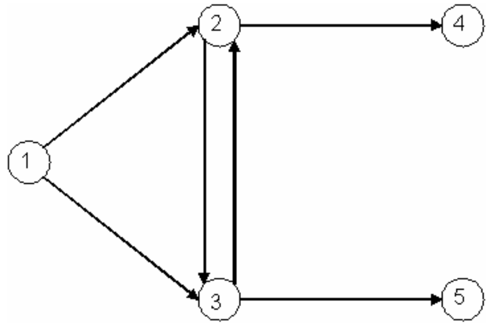


Figure 6: A 5-node 6-arc network



The GAMINTB was programmed using Fortran, and the tests were run on a computer platform with a 2.6 GHz Pentium 4 processor and 512 Mb of RAM. To ensure the GAMINTB converged to the optimal solution, LINGO 7, an NLP solver was used to solve MINTB. For each example network, the GAMINTB converged to the same solution produced by LINGO. In addition, LINGO was also used as the LP solver for GAMINTB to determine the feasibility for binary vector y with respect to constraints in problem (1).

The quality of the GAMINTB is defined as the percentage of the final population that is the optimal solution. It can be used as a measure of both the convergence rate of the GAMINTB and the diversity created within the GAMINTB. The algorithm performed well if the quality fell between 40% and 60%. A solution quality in this range indicates that a clear winner was discovered, yet a sufficient amount of diversity was introduced throughout the algorithm. Note that 100% is not considered to be the best quality because it could indicate that the algorithm only terminates at a local optimum.

We first conduct the "implementational comparison test" to compare the quality of various levels of immigration and mutation, and quality differences between implementing uniform and weighted selection. This test demonstrated the effectiveness of each implementational change according to the GA's quality and aided in the selection of the rate of reproduction, which determined the immigration rate, the percentage of chromosomes to mutate, and whether to implement uniform or weighted probabilities in the selection of parents.

The test was run as follows: for each example network, the GAMINTB was run with mutation alone, immigration alone, and with both; the tested rates of alteration for each scheme were 20%, 30%, and 40%. The initial population size was 100 and the generation number was 20. The selection probability was weighted. To show the effects of mutation and immigration on the convergence rate, the qualities for all GAMINTB runs were recorded and summarized in Table 1.

Table 1 makes it clear that immigration is much more effective in slowing the convergence rate of the GAMINTB. From another perspective, immigration allowed for more variety among optimal solutions into the final population to minimize any fears that the algorithm may prematurely converge to some local optimal solution. The results looked very promising with the standard immigration rate of 30%. On the other hand, mutation was not nearly as effective. It did slow the convergence rate, but only after the mutation rate was raised to an absurdly high level. (A typical mutation rate usually hovers between 2% and 5%). In most cases, the effect of mutation on the genetic algorithm is minimal, thus it is not essential to the success of our genetic algorithm. Finally, utilizing crossover alone led to an extremely high convergence rate for the GAMINTB, which implied that the solution may have converged prematurely.

Table 1: Comparing the Effects of Mutation and Immigration
The Percentage of Optimal Solutions in The Final Generation

Network	Mutation Rate			Immigration Rate			Rate of Both			Neither
	20%	30%	40%	20%	30%	40%	20%	30%	40%	0%
1	82%	70%	62%	61%	30%	31%	56%	11%	3%	0%
2	79%	74%	73%	75%	50%	32%	59%	37%	15%	100%
3	82%	74%	64%	66%	43%	27%	62%	30%	13%	100%
4	80%	65%	64%	61%	37%	10%	43%	22%	8%	0%
5	82%	76%	67%	70%	38%	35%	45%	12%	14%	100%
6	82%	69%	63%	64%	41%	8%	42%	16%	10%	100%

Strong arguments have been made in favor of determining the parents based on a uniform probability or on a weighted probability. Ahuja et al. (1993) argue that biasing selection in favor of the fitter individuals leads to a faster convergence, and that the use of uniform probability provided them with better results (Ahuja et al., 1993). Conversely, Sumaleee (2004) used a weighted probability with a "selection bias," which could focus on choosing the better strands. We tested both selection probability schemes for GAMINTB with the population size and generation number being 100 and 20, respectively. The quality results for GAMINTB (in percentages) are listed in Table 2.

Table 2: Comparing Uniform and Weighted Probability

Network	Uniform	Weighted
1	7%	30%
2	3%	50%
3	11%	43%
4	1%	37%
5	7%	38%
6	5%	41%

Comparing the uniform and weighted probability methods, Table 2 suggests that weighted probability converges at a much higher rate. However, the convergence rate is not high enough to generate much alarm for premature termination. On the other hand, some of the uniform cases experienced rather low convergence rates and were dangerously close to not converging at all. Thus, it can be concluded that GAMINTB is much more effective when using a weighted selection probability. The benefits of uniform probability may have been limited by the fact that the algorithm was only tested on smaller networks. This will be investigated in our future work testing larger networks.

Our second computational test is the "Performance Test." The goal is to demonstrate the effect of population size and generation number on run time and quality. The experiment was conducted as follows: the mutation rate (percentage of population to be mutated with each generation) and immigration rate (percentage of population to be filled by immigration with each generation) were set at zero and thirty percent, respectively and the selection probability was weighted. The first part of the experiment fixed the number of generations at 100 and ran the GAMINTB with population sizes of 500, 1000, 1500, and 2000; the second part of the experiment fixed the population size at 100 and ran the GAMINTB with generation numbers of 500, 1000, 1500, 2000. For each trial, the CPU time and quality of the GAMINTB were recorded. Table 3 displays the CPU time information when varying the population and generation sizes. Additionally, Table 4 provides the quality information when varying the population and generation sizes.

Table 3: CPU time Comparison (Population Size vs. Generation Size)*

Network	Population Size				Number of Generations			
	500	1000	1500	2000	500	1000	1500	2000
1	1.25	4.16	8.83	15.04	0.56	1.16	1.68	2.26
2	1.34	4.37	8.97	15.33	0.64	1.29	1.92	2.55
3	1.38	4.41	8.9	15.05	0.72	1.43	2.15	2.84
4	1.04	3.82	8.06	14.06	0.35	0.7	1.05	1.38
5	1.47	4.71	9.59	16.27	0.74	1.47	2.21	2.92
6	1.45	4.67	9.4	16.06	0.71	1.42	2.14	2.84
*All CPU times are in seconds.								

Examining Table 3, we find that the CPU time grows much more rapidly when increasing the population size than increasing the generation size. Thus, it may be better to increase the number of generations before increasing the population size. On the other hand, Table 4 shows the quality gains from increasing population and generation sizes. Note that in Table 4 after raising the population size above 1000, the quality began to decrease, while increasing the number of generations the quality showed an upward trend for the six example networks. Because of these trends, it is better to increase the number of generations than it is to increase the population size as the numbers get very large.

Table 4: Quality Comparison (Population Size vs. Generation Size)

Network	Population Size				Number of Generations			
	500	1000	1500	2000	500	1000	1500	2000
1	38%	43%	38%	39%	24%	38%	38%	37%
2	41%	43%	41%	43%	35%	34%	35%	50%
3	38%	46%	44%	41%	18%	30%	46%	43%
4	34%	41%	38%	39%	35%	35%	40%	37%
5	21%	24%	15%	18%	14%	10%	19%	15%
6	36%	38%	36%	32%	39%	28%	36%	39%

Overall, as illustrated in Figures 7, 8, 9 and 10, increasing the number of generations increases the total CPU time less rapidly and still produces similar quality results when compared to increasing the population size. Each of these figures demonstrates the following performance data calculated on the six networks. Networks 1 through 6 are marked by ◊, □, Δ, ×, and *, respectively.

Figure 7

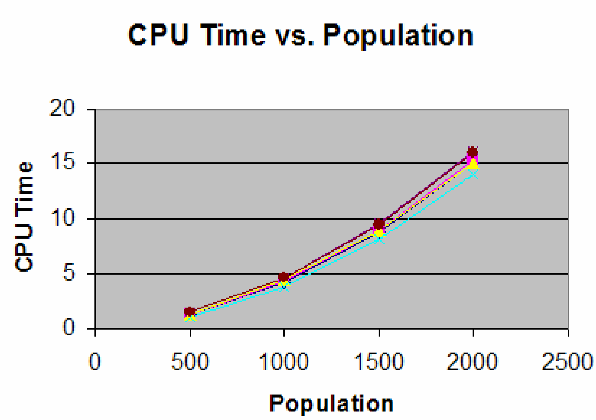


Figure 8

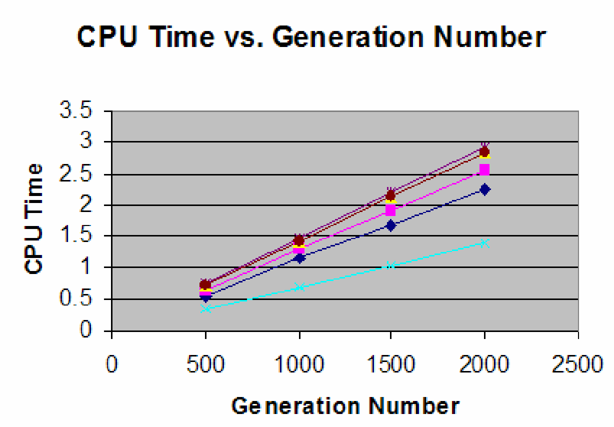


Figure 9

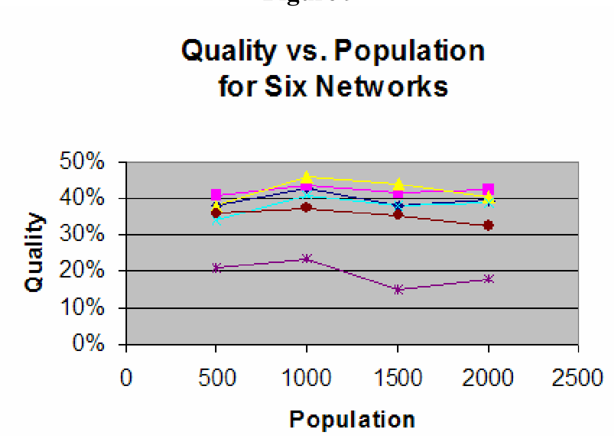
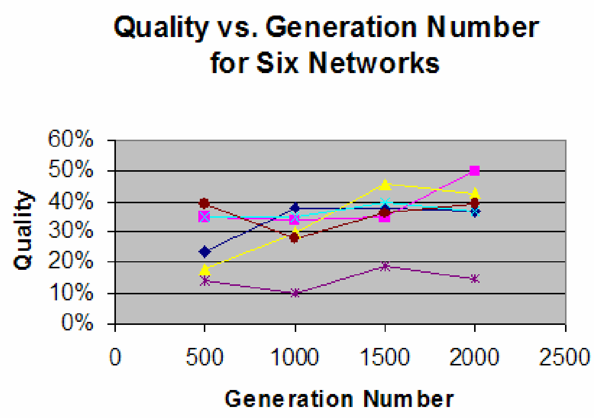


Figure 10



AN ALTERNATIVE IMPLEMENTATION

An alternative approach for GAMINTB without having to check the feasibility for each binary vector y is to relax the first two conditions in problem (1) and add them into the objective function as a penalty. This change will resolve the issue of dividing the y's into feasible and infeasible groups. Instead, they will all be rated by the number of toll booths, but they will also be penalized for not satisfying the toll constraints. To implement this method, the objective function will be composed of three parts. The first is the total number of toll booths, which is our original main objective. The second is adding on a scalar of $\sum_k (U^T \rho^k - (s(\bar{v})+\beta))^2$ if $U^T \rho^k > (s(\bar{v})+\beta)$. This is the penalty for not satisfying the first constraint in (1). The last component of the objective function is adding on a scalar of $((s(\bar{v})+\beta)^T \bar{v} - \sum_k b_k^T \rho^k)^2$, which penalizes *y* for not satisfying the second constraint in (1).

There are issues associated with this method. For example, the penalty method still requires accurate values for β 's and ρ 's, which are difficult for the GAMINTB. Secondly, the determination of appropriate penalty factors may require solving additional nonlinear programs iteratively, which can be very computationally expensive. Many other challenges with this penalty method need adequate exploration and we leave them for future research.

CONCLUSIONS

The purpose of this paper is twofold. One is to apply the genetic algorithm to the MINTB problem and develop a general framework for the methodology. The other is to validate the so developed method through small numerical examples. In response, the paper presents a customized genetic algorithm, GAMINTB, as a heuristic method for solving the minimum toll booth problem; and the algorithm is shown to be effective at solving MINTB for six small networks, all matching the optimal solutions generated by an NLP solver LINGO 7. Furthermore, computational tests suggest the following guideline for using GAMINTB. First, when searching for a better solution, the number of generations should be increased before increasing the population size, due to a lower CPU time and a higher projected convergence rate. Second, the use of immigration is much more effective in promoting diversity than the traditional mutation process. Third, the use of weighed probability in the selection process is better than the uniform probability for solving MINTB. However, whether or not the uniform probability would still be ineffective for larger networks needs further investigation. Finally, other streams of future research include testing and applying GAMINTB for larger networks, and considering the penalty approach for a variation of the GAMINTB.

REFERENCES

Ahuja R.K., Orlin J.B., & Tiwari A. (1993). A Greedy Genetic Algorithm for the Quadratic Assignment Problem. *Computers and Operations Research,* 27, 917-934.

Bai L. (2004). Computational Methods for Toll Pricing Models. Unpublished Ph.D. Dissertation, Department of Industrial and Systems Engineering, University of Florida.

Bai L., Hearn D.W., & Lawphongpanich S. (2008). A Heuristic Method for the Minimum Toll Booth Problem. Working Paper.

Beckmann M., McGuire C., & Winston C. (1956). *Studies in the Economics of Transportation.* Yale University Press, New Haven, CT.

DeCorla-Souza P. (2003). Report to the Transportation Research Board Joint Subcommittee on Pricing. Presented to *82nd Annual Transportation Research Board Meeting 2003.* Washington, D.C.

Florian M., Guélat J., & Spiess H. (1987). An Efficient Implementation of the PARTAN Variant of the Linear Approximation Method for the Network Equilibrium Problem. *Networks,* 17, 319-339.

Florian M., & Hearn D.W. (1995) Network Equilibrium Models and Algorithms. In Ball M.O., Magnanti T.L., Monma C.L., & Nemhauser G.L. (Eds.). *Chapter 6 of Handbooks in Operations Research and Management Science (8): Network Routing.* North-Holland, New York.

Hearn D.W., & Ramana M. (1998). Solving Congestion Toll Pricing Models. In Marcotte P, & Nguyen S. (Eds.), *Equilibrium and Advanced Transportation Modeling,* (pp. 109-124). Kluwer Academic Publishers.

LeBlanc, Morlok E.K., & Pierskalla W.P. (1975). An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. *Transportation Research,* 9, 309-318.

Pigou A.C. (1920). *The Economics of Welfare. MacMillan.* New York.

Schrank D., & Lomax T. (September 2007). Urban Mobility Report 2007. Texas Transportation Institute. from http://mobility.tamu.edu.

Shepherd S., & Sumalee A. (2004). A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems. *Networks and Spatial Economics,* 4, 161-179.

Sumalee A. (2004). Optimal Road User Charging Cordon Design: A Heuristic Optimization Approach. *Computer-Aided Civil and Infrastructure Engineering,* 19, 377-392.

Todd J. (2005). Duke Student Math Aims to Alleviate Tollbooth Lines. *Duke University News and Communications,* June 2005. from http://www.dukenews.duke.edu/2005/06/tollbooths.html.

Zhang X. (2003). Optimal Road Pricing in Transportation Networks. Unpublished Ph.D. Dissertation. Department of Civil Engineering, Hong Kong University of Science and Technology, Hong Kong.

FORECASTING KOREAN STOCK PRICE INDEX (KOSPI) USING BACK PROPAGATION NEURAL NETWORK MODEL, BAYESIAN CHIAO'S MODEL, AND SARIMA MODEL

Kyung Joo Lee, University of Maryland Eastern Shore
Albert Y. Chi, University of Maryland Eastern Shore
Sehwan Yoo, University of Advancing Technology
John Jongdae Jin, University of Maryland-Eastern Shore

ABSTRACT

In this study, we forecast Korean Stock Price Index (KOSPI) using historical weekly KOSPI data and three forecasting models such as back-propagation neural network model (BPNN), a Bayesian Chiao's model (BC), and a seasonal autoregressive integrated moving average model (SARIMA). KOSPI are forecasted over three different periods (i.e., short-term, mid-term, & long-term). The performance of the forecasting models is measured by the forecast accuracy metrics such as absolute forecasting errors and square forecasting errors of each model.

The findings are as follows: First, between BPNN and BC, BPNN performs better than BC for mid term and long term forecasting, while BC performs better than BPNN for the short term forecasting. Second, between BPNN and SARIMA, SARIMA performs better than BPNN for mid term and long term forecasting, while BPNN does better than SARIMA the short term forecasting. Between SARIMA and BC, SARIMA performs better than BC for mid term and long term forecasting, while the other way around is true for the short term forecasting.

In sum, the SARIMA performs best among the three models tested for mid term and long term forecasting, while BC performs best for the short term forecasting.

INTRODUCTION

The ability to forecast the capital market price index is critical to individual investors, institutional investors, and financial analysts. Among many forecasting models for stock prices and market price index, the seasonal autoregressive integrated moving average model (SARIMA) has been one of the most popular forecasting models in capital market studies.

Recently, the neural network model has been frequently used in many capital market studies (e.g., Ansari et. al. (1994), Hamid et. al. (2004), Huang et. al. (2005), Kumar et. al. (2006), Malik et. al. (2006), Stansell et. al. (2004), and Trinkle et. al. (2005)). Major reasons for the neural network model's popularity in capital market forecast are twofold. First, the neural network model is data driven method which learns from sample data and hence does not require any underlying assumptions about the data. Thus, the model is known as a universal functional approximate without severe model misspecification problems due to wrong assumptions (Hornik et. al. (1989)). The model is also outstanding in processing large amount of fuzzy, noisy, and unstructured data. For example, Hutchinson et. al. (1994) examine stock option price data and show that the neural network model is computationally less time consuming and more accurate non-parametric forecasting method, especially when the underlying asset pricing dynamics are unknown or when the pricing equation cannot be solved analytically. Second, stock price data are large, highly complex and hard to model because the pricing dynamics are unknown, which suits the neural network model.

The Bayesian Chiao's model (BC) may be another powerful and practical tool to forecast capital market data for the following reasons. First, the main thoughts of the BC is the dynamic way of combining the prior information (i.e. either from historical datasets or from previous subjective experience) with the current observations, during the process of posterior information. Second, most of the traditional statistics applications are based on the assumptions of independent, identical distributed (i.e. i.i.d.) normal random variables. However, the merit of the BC is to assume independence of variables only. No more identical distributions are needed. Third, the pros of BC can be the dynamic adaptive mechanism of integrating prior knowledge and the current information for accurately predicting the immediate future outcomes. However, the con of this model is its high dependence on the quality of the initial values of the estimates. Further, without constantly absorbing realistic datasets, the long term iterative predictions of this model perform poorly, if merely repeatedly applies the BC. Thus, in general, the BC is more effective in short-term forecasting than it is in long-term forecasting.

Korean stock market is considered more volatile than its US counterpart and hence has fuzzier and unstructured price data, which suits the Neural Network model and the BC well (for the short term forecasting, in particular). Thus, it is meaningful endeavor to examine how well the neural network model and the BC perform in forecasting more volatile Korean market data relative to a conventional SARIMA model which is one of the most popular forecasting models in capital market studies. The purpose of this study is to compare the ability of the neural network model, the BC, and SARIMA model in forecasting Korean Stock Price Index (KOSPI). Weekly data of KOSPI are analyzed in this study.

The remainder of this paper is organized as follows. Sample data and methodology are discussed in the next section, followed by discussions on empirical results in section three. The concluding remarks are presented in the final section.

DATA AND METHODOLOGY

Index Data

The data used in this study are KOSPI for closing prices from the Korean Stock Exchange (KSE) data base. The data series span from 4th January 1999 to 29th May 2006, totaling 390 weeks (89 months) of observations. Although KOSPI data is available since the opening of the Korean Options Exchange for stock price index in July 1997, we exclude two year's data (1997-1998) because the Korean stock market had suffered the severe financial crisis (IMF crisis) during this period.

The data are divided into two sub-periods, one for the estimation and the other for the forecasting. We use four different forecasting periods to examine the potential impact of forecasting horizons on the forecasting accuracy. Forecasting horizons used are 20% (long range), 13% and 8% (mid range), and 6% (short range) of the total number of observations. For the long range forecasting, the first 313 weekly data are used for model identification and estimation, while the remaining 77 weekly data (about 20% of 390 weeks) are reserved for evaluating the performance of SARIMA, BC, and the neural network model. Other forecasting periods were defined in the similar way, resulting in mid range (31-50 weeks ahead) and short range (23 weeks ahead) forecasting horizons.

Neural Network Forecasting

In this study, the back-propagation neural network model (BPNN) is used for time series forecasting. The main reasons for adopting BPNN are twofold. First, BPNN is one of the most popular neural network models in forecasting. Second, BPNN is an efficient way to calculate the partial derivatives of the networks error function with respect to the weights and hence to develop a network model that minimizes the discrepancy between real data and the output of the network model.

BPNN can be trained using the historical data of a time series in order to capture the non-linear characteristics of the specific time series. The model parameters will be adjusted iteratively by a process of minimizing the forecasting errors. Fore time series forecasting, the relationship between output (yt) and the inputs (yt-1 yt-2,…, yt-p) can be described by the following mathematical formulae.

$$y_t = a_o + \sum_{j=1}^{q} a_j f(w_{oj} + \sum_{i=1}^{p} w_{ij} y_{t-i}) + e_t \qquad (1)$$

where a_j (j = 0,1,2,…,q) is a bias on the jth unit, and w_{ij} (i = 0,1,2,…,p; j = 0,1,2.,,,q) is the connection weights between layers of the model, f(·) is the transfer function of the hidden layer, p is the number of input nodes and q is the number of hidden nodes. The BPNN performs a nonlinear functional mapping from the past observation (y_{t-1} $y_{t-2,\ldots}$ y_{t-p}), to the future value (y_t), i.e.,

$$y_t = \varphi(y_{t-1}\ y_{t-2,\ldots}\ y_{t-p}) + et \tag{2}$$

where w is a vector of all parameter and φ is a function determined by the network structure and connection weights.

The NeuroSolutions NBuilder toolbox is used to train data for model developments and test data for forecasting accuracy of the models developed. Here the stop criteria for the supervised training of the networks are specified as follows. The maximum epochs specify how many number of iterations (over the training set) will be done if no other criterion kicks in. The training terminates when one of the following four conditions is met (1) mean square error of the validation set begins to rise, indicating that over fitting might be happening.; (2) when the threshold is less then 0.01, i.e., we are on effectively flat ground; (3) when training time has reached 1000 epochs; (4) the goal (difference between output and target is less than 0.01) has been met. After the training was completed, its epochs and the simulation procedure completed successfully, which indicated the network was trained was predicting the output as desired.

According to the principle of Ockham's razor, the simplest networks topology yielding satisfactory results is used. The networks are created as '2-1-1': that is, two input layers, one hidden layer, and one output layer. The network is also trained using various other topologies such as 2-X-1, while X = 2, 3, 4, and 5. However, the best results are obtained when there is one hidden layer (i.e., X=1).

Bayesian Chiao Forecasting

With respect to the sequence of the logistic data having two possible results (i.e. pass, which means "$X_{m-1} < X_m$"; or failure, which means "$X_{m-1} > X_m$), the trends of the central tendency and deviation can be sequentially adjusted. From BC, the posterior distribution X_{m+1} can be calculated as the following:

"Pass Case": If the predicate value of X_m is "pass", then

$$\mu_{m+1} = \mu_m + (1-c_g)(\sigma_m^2 / sqrt(\sigma_m^2 + a_g^{-2}))(f(D)/(c_g + (1-c_g)F(-D))) \tag{3a}$$

$$\sigma_{m+1}^2 = \sigma_m^2(1-((1-c_g)/(1+a_g^{-2} * \sigma_m^{-2}))(f(D)/A)(((1-c_g)f(D)/A)-D)) \tag{3b}$$

"Failure Case": If the predicate value of X_m is "failure", then

$$\mu_{m+1} = \mu_m - (\sigma_m^2 / sqrt(\sigma_m^2 + a_g^{-2}))(f(D)/F(D)) \tag{4a}$$

$$^{2}{}_{m+1} = \ ^{2}{}_{m} \left(1 - (1 + a_g^{-2} \ {}_m^{-2})^{-1} f(D)\,[\,(f(D)/F(D)) + D\,]/F(D)\right) \tag{4b}$$

Where f(D) = the value of probability density function of X_m = D;

F(D) = the value of cumulative density function of X_m = D;

$D = (b_g - \mu_m)/sqrt(\sigma_m^2 + a_g^{-2})$; $sqrt(\sigma_m^2 + a_g^{-2})$ = the square root of $(\sigma_m^2 + a_g^{-2})$

A = c_g + (1- c_g) F (-D);

a_g = the parameter for the degree of similarity on the attribute values;

b_g = the parameter for the degree of proximity;

c_g = the parameter for the degree of uncertainty in the estimation;

μ_m = the prior central tendency of the observed values of X_m ;

${}_m$ = the prior variance of the observed values of X_m ;

μ_{m+1} = the posterior central tendency of the observed values of X_{m+1} ;

${}_{m+1}$ = the posterior variance of the observed values of X_{m+1} ;

X_m = the prior random variable;

X_{m+1} = the posterior random variable;

Time-series Forecasting

To obtain the KOSPI forecasts from the SARIMA, we adopted the Box and Jenkins' method. Basically, Box and Jenkins' method uses the following three-stage approach to select an appropriate model for the purpose of estimating and forecasting a time-series data.

Identification: we used the SARIMA procedure in SAS statistical software to determine plausible models. The SARIMA procedure uses standard diagnostics such as plots of the series, autocorrelation function (ACF), inverse autocorrelation function, and partial autocorrelation function (PACF).

Estimation: Each of the tentative models is fit and the various coefficient estimates are examined. The estimated models are compared using standard criteria such as Akaike Information Criteria and the significance level of coefficients.

Diagnostic checking: SARIMA procedure is used to check if the residuals from the different models are white noise. The procedure uses diagnostics tests such as ACF, PACF, and Ljung-Box Q-statistics for serial correlation.

Applying these steps, SARIMA (110)(12) was selected as the best-fitting forecasting model for weekly KOSPI data.

Measurement of Forecast Accuracy

Forecast error (FE) is determined by subtracting forecasted value from actual value of KOSPI, and then deflating the difference by the absolute value of actual data as follows:

$$FEt = (At - Ft)/|At| \tag{5}$$

where At = actual value of KOSPI in period t.
Ft = forecasted value of KOSPI in period t.

The reason for using the absolute value of actual KOSPI as deflator is to correct for negative values. Accuracy of a forecast model is measured by sign-neutral forecast error metrics such as absolute forecast errors (AFE) and squared forecast errors (SFE). Since the results are essentially the same, we report only those from using AFE.

RESULTS

Descriptive statistics of forecast accuracy (AFE) from BPNN, BC, and SARIMA using weekly KOSPI price data are presented in Table 1. Those statistics of AFE for four different forecasting horizons such as 77 weeks ahead (long), 50 weeks ahead (upper middle), 31 weeks ahead (lower middle), and 23 weeks ahead (short) are presented in Table 1. Mean, standard deviation, minimum, and maximum of AFE for four different forecasting horizons are also presented in Table 1. The SARIMA provides smallest mean AFE for forecasting horizons of 31 weeks ahead or longer, while BC provides the smallest mean AFE.

This may indicate that, among the three models, SARIMA is the most effective in mid-term and long-term forecasting while BC is the best in short-term forecasting, which is consistent with the findings of Wang et. al. and our prediction.

The mean AFE from the above-mentioned three forecasting models, Kruskal-Wallis χ^2 statistics, and the corresponding p-values are presented in Panel A of Table 2. Kruskal-Wallis χ^2 statistics of all four forecasting horizons are statistically significant at 0.001 significance level, indicating that, overall, forecasting errors from the three models are significantly different each other.

Results from pair wise comparisons between BPNN and BC, between BPNN and SARIMA, and between BC and SARIMA are presented in Panel B of Table 2.

Comparisons between BPNN and SARIMA show that SRIMA produce smaller forecasting errors than BPNN for mid-term and long-term forecasting horizons (i.e., 31 weeks, 50 weeks, & 77 weeks), while BPNN produce smaller forecasting errors than SARIMA for short-term forecasting horizon (i.e., 23 weeks). All differences in forecasting errors between BPNN and SARIMA are statistically significant at the significance level of 0.01. This indicates that SARIMA performs better than BPNN in mid-term and long-term forecasting while BPNN performs better than SARIMA in short-term forecasting.

Table 1: Descriptive Statistics for Forecast Accuracy

Forecasting Horizon	Model	Mean	Std Dev	Min	Quartiles			Max
					25%	50%	75%	
77	BPNN	0.123	0.090	0.001	0.039	0.109	0.207	0.274
	BC	0.224	0.123	0.003	0.110	0.234	0.341	0.398
	SARIMA	0.099	0.063	0.002	0.047	0.102	0.151	0.214
50	BPNN	0.171	0.073	0.022	0.113	0.201	0.227	0.273
	BC	0.202	0.086	0.006	0.135	0.241	0.267	0.313
	SARIMA	0.108	0.052	0.006	0.062	0.123	0.145	0.216
31	BPNN	0.099	0.027	0.062	0.078	0.089	0.127	0.148
	BC	0.093	0.038	0.002	0.072	0.083	0.127	0.156
	SARIMA	0.055	0.026	0.004	0.034	0.056	0.074	0.102
23	BPNN	0.040	0.025	0.004	0.014	0.049	0.065	0.077
	BC	0.031	0.016	0.001	0.019	0.037	0.045	0.059
	SARIMA	0.095	0.066	0.010	0.035	0.084	0.125	0.252

1) Forecast errors are defined as: (At-Ft)/|At|, where At=actual value of KOSPI index, Ft=forecast value of KOSPI index. Forecast accuracy is defined as the mean absolute forecast error (AFE).
2) The number of weeks used to estimate forecasting models is different depending on forecasting horizons. For example, 313 weeks (367 weeks) of data were used to obtain forecasts for the 77 weeks (23 weeks) ahead. Each forecasting horizon represents roughly 20%, 13%, 8% and 6% of total number of observations (390 weeks).
3) BPNN= Back-propagation neural network model without hidden layer. BC = Bayesian Chiao's model. SARIMA= Best-fitted SARIMA model: (110)(12).

Comparisons between BPNN and BC show that BPNN produce smaller forecasting errors than BC for 77 weeks and 50 weeks forecasting horizons. And the differences in forecasting errors between the two methods are statistically significant at the significance level of 0.01. This indicates that BPNN produce more accurate forecasts than BC for these relatively longer-term forecasting. On the other hand, BC produce smaller forecasting errors than BPNN for 31 weeks and 23 weeks forecasting horizons but the differences in forecasting errors between the two models are statistically significant only for 31 weeks forecasting horizon at the significance level of 0.05. This may indicate that BC performs better than BPNN in 31 weeks ahead forecasting but no meaningful conclusion can be drawn for the 23 weeks ahead forecasting.

Comparisons between BC and SARIMA show that SRIMA produce smaller forecasting errors than BC for mid-term and long-term forecasting horizons (i.e., 31 weeks, 50 weeks, & 77 weeks), while BC produce smaller forecasting errors than SARIMA for short-term forecasting horizon (i.e., 23 weeks). All differences in forecasting errors between BC and SARIMA are statistically significant at the significance level of 0.01. This indicates that SARIMA performs better than BC in mid-term and long-term forecasting while BC performs better than SARIMA in short-term forecasting.

Table 2: Comparison of Forecast Accuracies across Forecasting Models

Panel A: Summary Statistics for Forecast Accuracy and Overall Comparisons

Forecasting Horizon (weeks)	Forecasting Models			Kruskal-Wallis 2 stat (p-value)
	BPNN	BC	SARIMA	
77	0.123	0.224	0.099	43.698 (0.001)***
50	0.171	0.202	0.108	33.902 (0.001)***
31	0.099	0.093	0.055	28.500 (0.001)***
23	0.040	0.031	0.095	17.701 (0.001)***

Panel B: Multiple Pair wise Comparisons

Forecasting Horizon (weeks)	BPNN VS. SARIMA	BPNN VS. BC	BC VS. SARIMA
77	0.024 (5.44)***	-0.101 (21.90)***	0.126 (15.65)***
50	0.062 (11.70)***	-0.032 (14.26)***	0.094 (13.63)***
31	0.044 (6.69)***	0.006 (2.20)**	0.037 (4.59)***
23	-0.055 (3.25)***	0.009 (1.18)	-0.064 (5.09)***

1) Forecast errors are defined as: (At-Ft)/|At|, where At=actual value of KOSPI index, Ft=forecast value of KOSPI index. Forecast accuracy is defined as the mean absolute forecast error (AFE).
2) NN= Back-propagation neural network model without hidden layer. BC= Bayesian dynamic adaptive model. SARIMA= Best-fitted SARIMA model: (110)(12).
3) Matched-pair t-tests were used for multiple pair wise comparisons.
***: Significant at <0.01; **: significant at <0.05; *: significant at <0.10;

In sum, SARIMA is the best forecasting model for mid-term and long-term forecasting, while BC is the best forecasting model for short-term forecasting.

CONCLUSIONS

The purpose of this study is to compare the forecasting performance of back-propagation neural network model (BPNN), a Bayesian Chiao's model (BC), and a seasonal autoregressive integrated moving average model (SARIMA) in forecasting weekly Korean Stock Price Index

(KOSPI). Forecasting performance is measured by the forecast accuracy metrics such as absolute forecasting errors (AFE) and square forecasting errors (SFE) of each model.

KOSPI data over the 390 week period extending from January 1999 to May 2006 are analyzed. We find the followings: First, the SARIMA provides most accurate forecasts among the three models tested for mid term and long term forecasting, while BC provides the most accurate forecasts for the short term forecasting. Second, between BPNN and BC, BPNN provides more accurate forecasts than BC for mid term and long term forecasting, while insignificant difference in forecasting errors exists between the two models for the short term forecasting.

These results are robust across different measures of forecast accuracy. Since the accuracy of forecasting values is dependent on the developing process of forecasting models, the results of this study may also be sensitive to the developing process of the BPNN, BC, and SARIMA.

REFERENCES

Ansari, A. M. and Frank C. Lin (1994). Forecasting financial trends in stock markets with neural networks. *Institute of Electrical and Electronics Engineers,* pp188-196.

Chiao, T. (2006) Strategic Quality Competitiveness Management; International Journal of Information Technology and Management, Vol. 5, Nos. 2/3, 2006, pp. 197-211.

Hamid, S. and Z. Iqbal (2004). Using neural networks for forecasting volatility of S & P 500 Index future prices. *Journal of Business Research,* Vol. 57-10, pp 1116.

Hornik, K., M. Stinchcombe, & H. White (1989). Multi-layered feedforward networks are universal approximators. Neural Networks, Vol. 2, pp.359-366.

Huang, W., Y. Nakamori, and S. Wang (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* Vol. 32-10, pp 2513.

Hutchinson, J., W. Andrew, & t. Poggio (1994).A nonparametric approach to pricing and hedging derivative securities via learning networks. Journal of Finance, Vol. 49. pp. 851-889.

Kumar, K. and S. Bhattacharya (2006). Artificial neural network vs. linear discriminant analysis in credit ratings forecast: a comparative study of prediction performances. *Review of Accounting & Finance* Vol. 5-3, pp 216.

Malik, F. and M. Nasereddin (2006). Forecasting output using oil prices: a cascaded artificial neural network approach. *Journal of Economics and Business* Vol. 58, pp 168-180.

Medeiros, M., T. Terasvirta, and G. Rech (2006). Building neural network models for time series: a statistical approach. *Journal of Forecasting Vol. 25-1, pp49.*

Stansell, S. and S. Eakins (2004). Forecasting the direction of change in sector stock indexes: An application of neural networks. *Journal of Asset Management,* Vol. 5-1, pp 37-48.

Tang, y., F. Xu, X. Wan, and Y. Zhang (Working paper). Web-based fuzzy networks for stock prediction. *Georgia State University.*

Trinkle, B. (2005). Forecasting annual excess stock returns via an adaptive network-based fuzzy inference system. *Intelligent System in Accounting, Finance, and Management,* Vol. 13-3, pp 165.

Wang, W., D. Okunbor, and F. Lin (Working paper). Future trend of the shanghai stockmarket. *University of Maryland-Eastern Shore.*

REFINING UNDERLYING DEMAND FUNCTIONS USING INFORMATION FROM RELATED TIME SERIES

Glenn E. Maples, University of Louisiana at Lafayette
Ronald B. Heady, University of Louisiana at Lafayette
Zhiwei Zhu, University of Louisiana at Lafayette

ABSTRACT

This research focused on improving estimations of demand functions using information from a large number of related many-period time series. The key premise is that by improving the data's internal consistency random variation will be reduced in the component time series.

The research starts with the assumption that some of the demand time series are significantly correlated, some only slightly correlated, and some completely uncorrelated. A process is then developed for improving the demand function of any distinct part by using the correlated demands of the other parts. Metrics of effectiveness are proposed and used to test the methodology. The resulting tests demonstrate the effectiveness of the method, referred to as an internally consistent correlations (ICC) method.

Importantly, the research makes no use of a clustering algorithm and no assumption about the "end use" of the data set. Thus, it is expected that the results will apply to a wide array of domains, including forecasting, data mining, signal processing, and process monitoring.

KEY WORDS: Preprocessing Time Series Data, Reducing Time series Noise, Inventory Control, Demand Forecasting, Data Mining, Internally Consistent Correlations.

INTRODUCTION

Modern organizations are using information technology to form new business models by tying customers (CRM), suppliers (SCM), and processes (ERP) to traditional business practices. This expansion of business models offers manifest benefits, including increased efficiencies and opportunities to develop new relationships. However, these advantages come at the price of increased complexity in the decision and operational domains.

Inventory processes are among those most affected by these changes. Alternatives for sourcing, new production methodologies, the increasing use of remanufacturing-all add to the challenges of managing inventory. Efficient administration of inventory, aimed at both improved availability and cost reduction, has become critical to the success of many organizations (Timme, 2003).

One common way to attempt to reduce the complexity of managing inventory is using assemblies. Rather than stocking elemental parts, manufacturers may inventory pre-assembled subcomponents. These assemblies, the basis for modular manufacturing, confer benefits of process flexibility, decreased manufacturing time, increased customization, and adaptive design.

Modularization has become an important thread in recent research. At the firm level, modularization requires different managerial competencies and perhaps requires those people charged with determining standards play a more important role in companies (Baldwin and Clark, 1997). From a strategy perspective, Montreuil and Poulin (2005) have coined the term varietizing to combinatorial strategy of personalization of products requiring the standardization of product design interfaces and modularity. The use of assemblies permits the ability to postpone activities in the manufacturing process, which is a key point in mass customization (Piller, Moeslein, and Stotko, 2004; Su et al, 2005). Return policies, which can be critical to retail competitiveness (Yao, Wu, and Lai, 2005), can be impacted by modularization, as it is a primary determinant of the reuse economics of returned parts. This paper seeks to add to this stream of research in the area of inventory planning and control.

The ultimate goal of this paper is to demonstrate an improved method for determining demand for individual parts by relying on information derived from the demand matrix for all parts. It begins by introducing a conceptual model that shows how important and even unexpected correlations among the demands for parts can occur in a variety of normal manufacturing environments. Then it proposes a process to use these inter-part correlations to extract heretofore-unused information from the historical time series demand data matrix to identify a more correct demand curve of parts. Finally, it provides a series of tests to show the ability of this method to better identify demand in a variety of archetypal situations.

Until the 19th century, factories were craft-centric-each product "off the manufacturing line" was different from the previous one as determined by the skill of the person producing it. In 1797, Eli Whitney proved the concept of interchangeable parts in a musket production line and thus opened the door for standardized parts, modular manufacturing, and mass production.

The use of assemblies allows the decoupling of both design and assembly tasks. At the design level, this means that limiting, improving, or adding new functionality in a product can be done by altering one or more sub-assemblies and meeting established interface requirements. Modularity tends to increase product quality since it can be enforced at the interface level. Design changes can be made in the field (in either a repair or re-design context) at the module level permitting less skilled workers to change products.

In product assembly, assemblages allow tasks to be compartmentalized in both time and distance. Assemblies permit partial production-allowing reduced critical paths and quicker reaction to demand changes. Modularization also allows the vertical disaggregation of production through such techniques as outsourcing and co-location. Finally, assemblies allow manufacturers to limit the number of human interactions with the production process; this is important as these interactions tend to be expensive as well as the source of a disproportionate share of errors.

Assemblies can substantially affect manufacturing complexity and operational control. At first blush, assemblies would simply seem to reduce complexity. After all, the tracking and coordination of a sub-assembly can require much less information and cognitive effort than that required to track its constituent parts. However, as users and manufacturers of MRP and ERP systems are quick to acknowledge, assemblies typically make inventory management and forecasting more difficult (Panisset, 1988).

These difficulties stem largely from the multiplicity of sources of parts-parts in assemblies at various levels of completeness, new parts, reworked parts, reused assemblies, and the possibility of having one or more alternate parts each of which in turn can be found in their own multitude of sources.

The upshot of these complicating factors is that in many real world situations parts have important, partial correlations with other parts. Most manufacturers face these issues; however, they are especially prominent in remanufacturing and build-to-order environments. The addition of intermediate sub-assemblies to the manufacturing process introduces other correlations and even further increases the difficulties in understanding the demand relations between parts.

Partially correlated time series data from inventory data typically have a wide distribution of correlation coefficients. The correlation coefficients are both positive and negative, and include a significant number of cases at the three extremes. That is, correlation coefficients between SKUs will have values near +1, 0, and -1 and all values in between. Many manufacturers have demand data that exhibit this behavior and the resulting component SKU demand patterns can be exceedingly complex. However, because the parts are, in some manner, related to a few end-products there is likely some correlation between them.

This research is focused on improving (reducing the scatter in) correlated demand data. These correlations can be due to the complicating factors described above or from the simple correlations derived from part relationships to final products. For reasons that will be explained shortly, it is assumed that the available data set is large. This means that both the number of time periods for which data are available, and the number of SKUs for which data are available, are large.

These data sets are used in forecasting, data mining, signal processing, production monitoring, and other applications. This research deals with reducing the random fluctuations in the data sets used with those topics, and not the topics themselves. For example, Owehand (2006) suggests a three-fold typology for forecasting: direct, top down, and bottom up. The purpose of this paper is to describe a technique to provide better demand data that might then be used with any of the forecasting approaches identified by Owehand, or as input to algorithms in the other areas mentioned above.

THE METHODOLOGY

This research provides a holistic approach to determining the demand parts. It suggests that an improvement to the demand function for a part can be derived from information from other parts. Specifically, if two or more different SKUs are highly correlated, but at one particular time period their demands do not increase or decrease in a manner consistent with the historical patterns, then it is likely that the demands for that particular time period can be improved by considering the demands of all the other correlated parts. Determining the size of the adjustments is complex and depends on all the other SKU's demand correlations and how their demands behave. This is similar to the issue in financial modeling where the estimation of the variance of prediction error is important in the risk assessment of investment. Several financial modeling approaches have incorporated the concept of an internally consistent correlation as seen in the work of Yan (2001) and an innovation approach to noise reduction in time series analysis as presented by Ozaki and Iino (2001). This paper develops a methodology to make these adjustments and thus improve the internal-consistency of multi-SKU datasets.

To describe the idea behind this research in more abstract terms one could say that the information contained in the entire data set can be used to provide information on the likely behavior of any one part at any one time period. This is done by using SKU-to-SKU correlations in the time domain to increase demand-to-demand correlations in the SKU domain.

Nomenclature

ICC stands for "internally consistent correlations." It can be the name of the method or the name of the resulting data. That is, the ICC method involves the use of intra-SKU correlations to improve data set internal consistency, and the modified data are called ICC data or the ICC results.

- SKU stands for "stock keeping unit" and is used to represent production units at all levels of combination. That is, a particular SKU could be the final consumer product, or any of its assemblies, subassemblies, modules, components, parts, etc., on down to elementary units such as O-rings and bolts.
- N is the number of SKUs being studied. SKUs are identified by either the i or j index and both indices run from 1 to N.
- T is the number of time periods being studied. Time periods are identified by either the s or t index and both indices run from 1 to T.
- D_{it} was the demand for SKU i at time t. It is assumed that the D_{it} values have been “cleaned.” That is, outliers and missing values have been resolved.
- D^{*}_{it} is the result of applying the ICC-based techniques described here to the D_{it} demand data.
- D^{js}_{it} is an intermediate time series useful in deriving certain ICC equations. It provides one estimate of D^{*}_{it}.

- r_{ij} is the Pearson product moment correlation coefficient between SKU_i and SKU_j calculated over time from t = 1 to t = T.
- S_{it} is the value of the theoretical demand (signal) underlying D_{it}. It is unknown except under research conditions where synthetic data are used. In this case, the D_{it} data are generated using the signal value S_{it} and a randomly generated error term. It follows that each data point's value has an associated error $\varepsilon_{it} = S_{it} - D_{it}$.
- $RMSE_i$ is the square root of the mean squared error for SKU_i. The mean squared error is the sum over all t of ε_{it}^2 divided by T. RMSE is the average over some specified set of $RMSE_i$ values and is an important measure of how well the data for the specified set of SKUs conforms to the underlying signal S_{it}.
- SNR_i is the signal-to-noise ratio. It measures how badly the underlying signal is masked by the random fluctuations known as "noise." SNR_i is the standard deviation of S_{it} divided by the standard deviation of ε_{it}. SNR is the average over some specified set of individual SNR_i values.
- Error Terms: The error term, εit, represents all of the random variations that change the demand for SKUi at time t. It is assumed that the errors have a zero mean, are independent of other error terms, and are normally distributed. These assumptions are likely valid when the source of the errors are non-systematic uncorrelated errors such as might occur because of transcription errors, misreadings, random supplier mistakes, and other unexplained happenings. Other errors might be due to poor employee training, lack of cooperation, and so forth. The existence of errors from these and even other sources should be taken for granted and while they do not necessarily meet all of the usual assumptions concerning zero means, independence, and normal distributions it is standard practice to make these assumptions in many aspects of production modeling (see for example, Chen and Lin, 2002).

ICC Processing

Assume SKU_j is a "child" of "parent" SKU_i. In manufacturing terms, SKU_i is higher in the product structure tree. SKU_i might be an assembly whereas SKU_j is a subassembly. If SKU_j and SKU_i are perfectly correlated then, for any t, a change in D_{it} can be predicted by knowing a change in D_{jt}.

Using time periods t = 1 to T for parts i and j provides T pairs of data ($D_{it,}$ D_{jt}). To derive a relationship between D_{jt} and D_{it} one can regress D_{jt} on D_{it} where D_{it} is the independent variable. This regression yields Equation (1) where a_{ij} and b_{ij} are the standard intercept and slope parameters of the linear regression result.

$$D_{jt} = a_{ij} + b_{ij} D_{it} \tag{1}$$

Equation (1) shows no error terms—the assumption that the mean value of each of the error terms is zero causes the error terms to drop out of the regression results. The complexities of the demand time series functions are completely irrelevant. The regression analysis is in the "parts domain," not the "time domain," and thus only cross sectional changes in demand between SKUs i and j are important.

The nature of b_{ij} can be clarified by giving an example. If SKU_i requires, say, 3 of SKU_j and there are no errors in the data, then $b_{ij} = 3$. Similarly, if SKU_i requires 3 of SKU_j and other SKUs also contains some number of SKU_j, then $b_{ij} = 3$ still holds only now b_{ij} is interpreted in the usual multiple regression sense of being the change in the number of SKU_j units for a one unit change in SKU_i given that all other factors are held constant. If the records on SKU_i and SKU_j are error filled or show inconsistent relative usage, then the expected value of b_{ij} will be 3 but any particular value of b_{ij} may differ from 3. Finally, if SKU_j is usually needed three at a time—that is has a cardinality of 3, then a manufacturing company typically would issue a "kit" containing three units of SKU_j. In this case, b_{ij} will equal 1 with units of "kits of SKU_j" per SKU_i. This shows that b_{ij} translates units of measure as well as relative quantities. Rewriting Equation (1) for time s yields Equation (2).

$$D_{js} = a_{ij} + b_{ij} D_{is} \tag{2}$$

Subtracting Equation (2) from Equation (1) yields Equation (3). The superscripts i and s are indices that specify a specific member of the data set. Indices i and s can take on any value from 1 to N and 1 to T, respectively.

$$D_{jt}^{is} = D_{js} + b_{ij}(D_{it} - D_{is}) \tag{3}$$

D^{is}_{jt} is just one estimate of an improved value for D_{jt} that is based on using the demand for part i at time s. Because Equation (3) holds for all s and all i it follows that all TN data points in the demand/time matrix are used to find a better value for each D_{jt} element in the matrix. All of these TN estimates are used to calculate D^*_{jt}. One of these estimates will be $D^{jt}_{jt} = D_{jt}$. So the original value D_{jt} provides one term making up the ICC D^*_{jt}.

The proper weighted average of all possible estimates of D_{jt}, i.e. the D^{is}_{jt} terms, is the ICC demand D^*_{jt}.

$$D_{jt}^{*} \equiv \sum_{i=1}^{N} \sum_{s=1}^{T} W_{ij} D_{jt}^{is} \tag{4}$$

Insert D^{is}_{jt} from Equation (3) into Equation (4) to obtain Equation (5).

$$D_{jt}^{*} \equiv \sum_{i=1}^{N} \sum_{s=1}^{T} W_{ij} (D_{js} + b_{ij}(D_{it} - D_{is})) \tag{5}$$

Now the weights W_{ij} must be determined. W_{ij} reflects the importance or precision of each term in the sum. One cause of uncertainty in D_{it} - D_{is} comes from the fact that demand data from times t and s, may not be comparable because, say, there has been some kind of fundamental shift in the manufacturing operations. This would make the two SKUs demand data be from different distributions and thus not applicable. This problem can be handled by using a time window that includes only data from a constant manufacturing situation. For this study, it is simply assumed that the original dataset only includes times that exhibit demand coming from normal manufacturing conditions.

A second cause of uncertainty in D_{it} - D_{is} comes from the fact that the regression coefficients may fluctuate considerably simply due to random error. Adjusting weights for this kink of uncertainty is more difficult than simply restricting the time range of acceptable data. R_{ij}^2, the coefficient of determination, which is the portion of the total variation in the dependent variable that is explained by its relationship with the independent variable, determines the fit of regression equation. For simple linear regression like that used here, $R_{ij}^2 = r_{ij}^2$ where r_{ij} is the Pearson product moment correlation coefficient. Therefore, using r_{ij}^2 to indicate the merit of a particular D^{is}_{jt} term is reasonable. Other weighting schemes are possible and some will be discussed later, but basing the weights on r_{ij}^2 is intuitive and common so it is used here. This means that the weight for each D^{is}_{jt} term is taken to be some normalized form of r_{ij}^2. The proper normalizing constant is the sum of all correlation pairs yielding Equation (6).

$$W_{ij} = \frac{r_{ij}^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} r_{ij}^2} \qquad (6)$$

Clearly, W_{ij} satisfies the requirements for all weighting functions. That is, each W_{ij} term is greater than or equal to 0, each is less than or equal to 1, and the sum of all W_{ij} terms is 1. Inserting W_{ij} into Equation (5) yields Equation (7), the fundamental equation of ICC data refinement.

$$D^*_{jt} \equiv \sum_{i=1}^{N}\sum_{s=1}^{T} r_{ij}^2 D_{js} + \frac{\sum_{i=1}^{N}\sum_{s=1}^{T} r_{ij}^2 b_{ij}(D_{iT} - D_{iS})}{\sum_{i=1}^{N}\sum_{S=1}^{T} r_{ij}^2} \qquad (7)$$

The degree to which D^*_{jt} is an improvement over D_{jt} depends on the specifics of the problem, but the expectation is that when N or T or the average r_{ij} value is larger there will be more improvement in passing from D_{jt} to D^*_{jt} because more information is available to be extracted from the r_{ij} matrix. In other words, RMSE is expected to decrease and SNR is expected to increase as N increases, T increases, or the average $|r_{ij}|$ increases.

As a check on the plausibility of the Equation (7) one can examine the limiting case where all demand time series are independent ($r_{ij} = 1$ for $i = j$, and $r_{ij} \approx 0$ for $i \neq j$). For $i = j$ the W_{ij} and b_{ij} are one and for $i \neq j$ the weights and slopes are zero. Inserting this special case information into Equation (7) shows that the D^*_{jt} value for SKU_j is unchanged from D_{jt}. Generalizing, one sees that in this special case the entire D^*_{jt} data set is identical to the original D_{jt} data set, as is necessary under the assumption of independence between the time series of SKU_j and the rest of the SKUs. To summarize the methodology just described, the four basic steps are:

1. Obtain the original demand data D_{it} for all SKUs.
2. Resolve irregulars and missing data issues.
3. Calculate the r_{ij} values between every D_{it} and D_{jt} pair.
4. Use Equation (7) to calculate all D^*_{jt}.

Spurious Correlations

The weight terms W_{ij} given by Equation (7) work well for most cases. However, the results of analyses involving many SKUs will have the demands for a few parts incorrectly adjusted simply because random variations sometimes produce series that are correlated. The smaller the ratio of times to number of SKUs (T/N) the more the analysis will be affected by spurious correlations. Therefore, while TN should be large to maximize the information available, T/N should be large to minimize the occurrence of spurious influences.

The problem of spurious correlations can be alleviated in two ways. The most direct way is to impose a cutoff restriction on r_{ij}. That is, if r_{ij} is below some specified value then W_{ij} is set to zero. Above the specific cutoff value the W_{ij} are calculated from Equation (6). This technique works well although it requires some guidance on choosing the proper cutoff level.

One answer is to use about $r_{ij} = 0.4$, equivalent to 16% of the variation in SKU_j being explained by the variation in SKU_i. This value is suggested by the findings of Guilford (1956) whose review of the literature led to a categorization of r_{ij} values. He suggested that correlations in the 0.4 to 0.7 range indicate "moderate correlation; substantial relationship."

A second approach is to choose the cutoff to be the r_{ij} value that corresponds to a given level of acceptable uncertainty that the true r_{ij} value might be zero. This idea, from the hypothesis testing literature, would use the t-test where t is a known function of r_{ij} and N, as is described in Groebner, Shannon, Fry and Smith (2005) or any elementary statistics textbook. A commonly accepted level of uncertainty is 5%, so if N = 30 (the results are insensitive to the exact value of N when N is greater than about 20) this would correspond to a one-tailed r_{ij} value of 0.3. Combining these two estimates suggests that using a cutoff value of about $r_{ij} = 0.35$ is reasonable.

Some researchers might prefer to avoid the use of cutoff values altogether. In this situation the authors have found it useful to use the following equation.

$$W_{ij} = \frac{r_{ij}^4}{\sum_{j=1}^{N}\sum_{i=1}^{N} r_{ij}^4} \qquad (8)$$

This form decreases the weight of slightly correlated SKU pairs far faster than Equation (6). At a r_{ij} value of 0.35 the Equation (6) weight will be 12% of the weight of a perfectly correlated pair, whereas the Equation (8) weight will be only 1.5%. This effectively controls the influence of the existence of many slightly correlated SKU pairs.

Negative Correlations

Negative correlations may be just as important as positive correlations. If either a strong positive or a strong negative correlation exists between SKU_i and SKU_j then either SKU will contribute information about the behavior of the other. However, whether or not one would want to apply an adjustment to the demand for SKU_i based on the values of the negatively correlated SKU_j involves other considerations such as the objectives of the analysis. For example, if the primary purpose is forecasting demand, it would make sense to include information from negatively related SKUs. On the other hand, if one wanted to use the adjusted D_{jt} results to identify similar SKUs for the purpose of building part families then the use of the negative correlations might not be appropriate.

Iteration

Equation (7) can be used in an iterative manner. That is, the first set of D^*_{jt} values can be used as input to Equation (7) just as the D_{jt} values were. This second iteration would provide a second set of D^*_{jt} values. Any number of iterations can follow. Making two passes very frequently, but not always, produces significantly better results than making only one. The improvement made by a third pass is sometimes significant, but rapidly diminishing improvements is common.

Equation (7) should not be iterated many times because the D^*_{jt} values do not eventually become stationary in any but a trivial sense. Instead, after the quick improvements of the first few iterations many more iterations cause the D^*_{jt} values to start decreasing. This is a natural result of the fact that having all D^*_{jt} values approach zero is a singular solution to the equation $D^*_{jt} = f(D^*_{jt})$ where the function is given by Equation (7). There are certain numerical techniques that can be applied to counteract this numerical drift, but the topic is beyond the scope of this research. Fortunately, these techniques are not necessary for the successful application of Equation (7).

TESTING

The required D_{it} test data are generated by specifying a demand function (signal) to generate the S_{it} values, as well as specifying a distribution that governs the error (noise) to generate the ε_{it} terms, and then summing the two terms to form D_{it}. This is repeated for all i = 1 to N and all t = 1 to T. The noise terms are drawn from a normal distribution with a mean of zero and a standard deviation chosen to provide a desired SNR. The demand functions, described in the next section, are chosen to illustrate the ICC method's performance for various prototypical demand patterns.

While generating the test data it is important to identify the various subsets of SKUs that have common characteristics. That is, the demands of one third of the SKUs might be based on one signal, the second third might be based on a different signal, and the final third might purely random. These SKU subsets must be identified so each subset's average improvement can be calculated. Of course, the subset information is not used in the ICC calculations. It is for evaluation purposes only.

Two figures of merit were used. While they are not independent, together they provide familiar information to a broader group of research disciplines. The first method is the fractional change in $RMSE_i$ (the square root of the mean squared error for a specified SKU_i). The error term, ε_{it}, for each SKU and time period is known by construction, so the calculation of $RMSE_i$ follows immediately. Averaging the fractional change in $RMSE_i$ values over all SKUs in a particular subset, as identified in the time series generation stage, provides a more reliable measure of improvement for that subset than examining each SKU's value.

The second figure of merit uses the percent change in SNR_i (signal-to-noise ratio). Formulas for SNR vary somewhat from discipline to discipline but SNR is always the ratio of some measure of the signal strength to a similar measure of the strength of the noise term. Here, SNR is taken to the standard deviation of the signal divided by the standard deviation of the noise. Again, the S_{it} and ε_{it} terms are known by construction so SNR_i is easily calculated. As with $RMSE_i$ data it is advisable to average over all SKUs in a given subset to obtain a more reliable measure of SNR change.

Besides being a useful figure of merit, SNR controls the difficulty of the problem. The relationship between SNR and the potential percentage improvement in either of the figures of merit used here is quite complex. Larger SNR values present an easier to decipher signal, but do not necessarily imply a greater percent improvement. This will be shown in the numerical results.

The question of setting a reasonable initial SNR must be resolved next. Judging the quality of a superb electronic picture or audio signal requires very high SNR values. In these cases, the clarity of the signal is expressed in decibels (db) because the SNR values are so large and vary so greatly that logarithmic measures are required. The relationship between SNR and db is db = 20 $Log_{10}(SNR)$, so SNR = 1 is equivalent to db = 20. When extracting signals from background noise is the goal, as is the case here, then acceptable SNR values are typically much lower. The literature suggests that signals with SNR ratios lower than about 1.0 are challenging; for instance Sen and Carroll (1975) use about 1.0 as the boundary between good and poor signal quality in their research on quality in voice transmission. When the SNR approaches 1.5 the problem of deciphering the signal is considered easy, and when it is 0.5 or less the problem is considered difficult. Below about 0.3, the quality of the incoming signal is generally considered undecipherable except when using highly sophisticated equipment and long sampling times. The result of these considerations is that the test cases used here mostly have initial SNR values of about 1. To provide a range of performance tests one case had a SNR value of approximately 0.3 and another case had a SNR value of approximately 1.5.

Other testing considerations of potential importance include average D_{it} value and the time rate of changes of the D_{it} data. Fortunately, because Equation (7) is based solely on correlations, these factors are completely irrelevant. That is, if one of two SKUs has an average demand value twice the other and/or a slope twice the other, their Pearson correlation coefficient does not change. This constant correlation means these SKUs will continue to improve their D^{*}_{jt} values just as they would if their average values and slopes were identical.

The next two factors that must be specified are T (number of time periods) and N (number of SKUs). These parameters have little affect on the results, except when N/T becomes so small that spurious correlations become a problem, or when NT becomes so large that processing times become excessive. Based on the authors' experience values of N = 50 and T = 30 have been found useful.

Finally, the test cases were designed to demonstrate the ICC method's effectiveness in meeting the following three objectives: the ability to discriminate between real relationships and random ones, the ability to process multiple signals, and the ability to work when most of the demands are unrelated to the demand of interest.

Test Cases

Based on the preceding principles, five test cases were analyzed. Each used these parameters: rij cutoff = 0.35; N = 50 SKUs; T = 30 time periods; Iterations = 1; Error terms generated from a normal [0, 1] distribution; Wij calculated from Equation (6)l and negative correlations not considered.

Case 1 (Saw-on-Sines): There are two subsets of SKUs. The first subset of 25 SKUs have a demand function that is a sine wave passing through about two cycles for T = 30 and with an amplitude varying from +1 to -1. Overlaid (additive to) on the sine wave is a saw-tooth pattern having an amplitude of +0.5 to -0.5 and a period of about 1/6 that of the sine wave. The second subset of 25 SKUs is the same except the sine wave amplitude is twice as much as in the first subset. Both demand functions have a random component added to the signal. The standard deviation of the random components was chosen to make the average SNR values be approximately 0.9 for the first subset of SKUs and 1.5 for the second subset.

Case 2 (Steps): There are three subsets of SKUs. The first subset of 16 SKUs has an abrupt decrease of 3 units after 1/3 of the time T. The second subset of 17 SKUs has an abrupt decrease of 4 units after 2/3 of the time T. The third subset of 17 SKUs is purely random. The average SNR value was set at approximately 1 for the first subset of SKUs and 0.5 for the second subset.

Case 3 (Ramps): There are two subsets of SKUs. The first subset of 25 SKUs has a ramp that increases a maximum of 1 unit and the second subset of 25 SKUs has a ramp that decreases a maximum of 2 units. Both patterns repeat themselves three times for T = 30. The average SNR value was set at 0.33 for the first subset of SKUs and 0.66 for the second subset.

Case 4 (Irregular A): There are two subsets of SKUs. The first subset of 25 SKUs has an underlying pattern that appears irregular and does not repeat, although it does have a general maximum about two times for each T = 30 periods. It is the "signal" for the first subset of SKUs.

The average SNR value was set at approximately 1. The second subset of 25 SKUs had no signal (the data were simply random).

Case 5 (Irregular B): This case is like Case 4 except a different set of random numbers was used to generate an auto correlated demand function for use as the underlying signal. The signal is irregular and does not repeat, although it does have a general maximum about once for each T = 30 periods. The average SNR value was set at approximately 1 for the first subset of SKUs and the second subset had no signal (the data were simply random).

The computer code used to generate these test cases, as well as a computer program that demonstrates the performance of the ICC method, is available for downloading at http://www.ucs.louisiana.edu/~rbh8900/icc.html. If the reader wishes to ICC analyze a data set this demonstration program has the capability of reading a data file (ASCII format) and analyzing its time series data using a wide range of adjustable parameters.

Table 1- Test Results for Five Example Cases

	SKU Subset 1				SKU Subset 2			
Test	Original SNR	Percent Increase in SNR	Original RMSE	Percent Decrease in RMSE	Original SNR	Percent Increase in SNR	Original RMSE	Percent Decrease in RMSE
Case 1	0.93	172%	0.94	63%	1.56	73%	0.97	42%
Case 2	0.95	86%	3.86	46%	0.52	118%	3.72	54%
Case 3	0.3	155%	2.89	61%	0.61	164%	2.82	62%
Case 4	1.07	103%	3.83	51%	NA	NA	NA	NA
Case 5	1.02	51%	3.64	34%	NA	NA	NA	NA

DISCUSSION

The data of Table 1 show that the ICC methodology can achieve remarkable improvements in the signal-to-noise ratios of certain SKUs. One hundred percent increases in average SNR, and 50% decreases in average RMSE, are not uncommon.

Case 1 showed that the processed data retained the presence of a saw-tooth pattern on top of the sine pattern. With an initial SNR of 0.93, the first subset showed an increase of 172% in average SNR and a 63% decrease in average RMSE. With an initial SNR of 1.56, the first subset showed an increase of 73% in average SNR and a 42% decrease in average RMSE.

Cases 2, 4 and 5 included subsets of purely random D_{it} data and all showed that the random data were almost unchanged after processing, as they should be. These cases also showed that the presence of random data did not destroy the improvement shown for the nonrandom D_{it} data. Case 2 used an abruptly changing “step signal” and the processed data successfully retained an abrupt change although it was not as sharp as the original. With an initial SNR of 0.95, the first subset showed an increase of 86% in average SNR and a 46% decrease in average RMSE. With an initial SNR of 0.52, the first subset showed an increase of 118% in average SNR and a 54% decrease in average RMSE. Case 3 used two "ramp signals," and for the SKU subset with an initial SNR of 0.30 there was an increase of 155% in average SNR and a decrease of 61% in average RMSE. The subset with an initial SNR of 0.61 showed an increase of 164% in SNR and a decrease of 62% in average RMSE. Cases 4 and 5 showed that signals with non-repeating irregular patterns improve just as those with simpler patterns. Case 4 had an initial SNR of 1.07 and it increased by an average of 103% while the average RMSE increased by 51%. Case 5 showed that even when only five of 50 SKUs had non-random correlations, the ICC method produced useful results. The initial SNR was 1.02 and it increased 51% on average while there was a decrease of 34% in average RMSE. This suggests that the ICC method will be effective in finding and improving the demand functions of a few related parts mixed in with many dissimilar parts.

These results show that even with very difficult demand functions the ICC method can produce significant improvements. They also show that the percent increase can be very significant even when the average SNR is relatively low. This does not mean that the precise underlying signal has been recovered. With low initial SNR values even large percentage improvements may result in data that still exhibit a great amount of scatter.

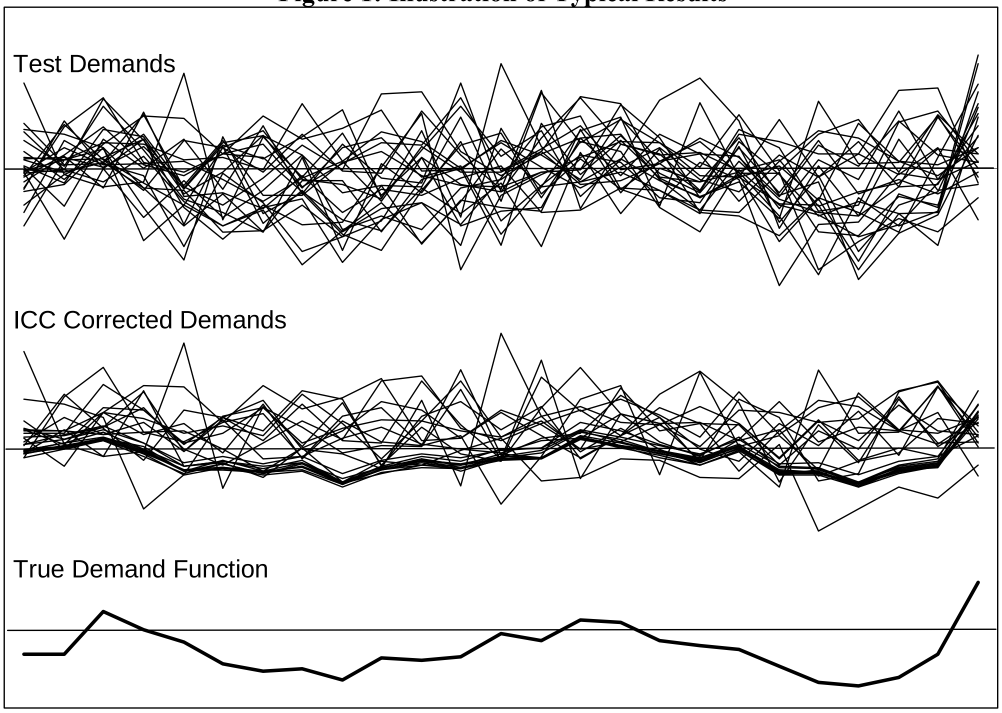
Finally, it should be noted that these results in no way used the information about SKU subsets existing in the data set. If the SKU subsets were known a priori then a simple averaging process would produce good results. The ICC process generated the results in Table 1 without identification of the SKUs.

Figure 1 presents data similar to "Case 4" except that the number of SKUs has been reduced to N = 25 and the number of time periods reduced to T = 25 to make the graph less congested and easier to interpret. Also, each of the 25 lines is offset vertically by a small amount to make interpreting the figure easier.

The top set of 25 lines shows the original data. That is, the D_{it} data points that were generated for 13 random time series and 12 time series with the random components added to an underlying irregular pattern.

The middle set of 25 lines shows the ICC processed data. To accentuate the nature of the ICC process, Equation (7) was applied twice instead of only once as was done with the other test cases. Within the middle group of lines one can see a group of 12 correlated SKU lines arbitrarily plotted near the bottom of the group for easier identification. With each application of Equation (7) the partially correlated data points adjust slightly to better match the correlations that were found in the original data. For comparison, the true signal (i.e., the series used to generate the 12 correlated lines) is shown by the heavier line at the bottom of the figure. While the set of partially correlated lines does not exactly reproduce the signal they are much closer than any of the than the original D_{it} values.

Figure 1: Illustration of Typical Results



CONCLUSIONS AND FUTURE WORK

Demands for any one part may be correlated with demands for other parts for a plethora of reasons: reuse of parts through returns and remanufacturing, inefficient manufacturing processes, defective inventory management policies, simple relationships of parts to final assemblies to name a few. Researchers and practitioners can take advantage of these correlations to develop an internally consistent matrix.

This research provides a new preprocessing methodology that reduces the random components a specified time series data subset by using information from this partially correlated time series data. Synthetic data covering a wide range of conditions demonstrate the methodology significantly reduced the root-mean-squared-error, and increased the signal-to-noise ratio, of the target time series. Often the improvement halves the RMSE and doubles the SNR. In these same tests, not only were target time series improved, but also time series exposed to random data were unaffected.

The technique should be helpful to those doing forecasting, data mining, signal processing, production monitoring, and similar tasks where they can implement a preprocessing step.

The field of numerical analysis is highly developed. This discipline should be studied to find a method of stabilizing the decrease in D^*_{jt} values once the first few iterations have been made. If this is accomplished then Equation (7) can be iterated until all D^*_{jt} values become stationary, thus removing another factor in any testing program.

The cutoff rules discussed previously are adequate for eliminating the effect of numerous, slightly correlated SKUs on the D_{it} values of a given SKU. However, there is no assurance that the guidelines suggested here are optimal. If a theoretical basis can be developed for choosing the proper r_{ij} cutoff value, perhaps one that is determined partially by the nature of the data set, then even more improvement in D^*_{jt} values could be expected.

Equation (7) has been shown to be very effective in improving D_{it} values of related SKUs mixed in with a large number of other D_{it} data. This is an important benefit of the method because group identification does not need to be done prior to the first processing step. However, the method can be made less effective in two circumstances. First, if there are several subsets of SKUs with related, but different demand functions then it is common for Equation (7) to make the similar subsets merge. This can happen when, for instance, there is a small phase shift between the various SKU time series. Although fundamentally different, under these conditions the subsets of SKUs have highly correlated members. Research is needed to find a way to keep such "related but different" demand functions from merging. A related problem occurs when a few large spikes in demand dominate the correlation coefficient value. In this case, the SKUs with similar spikes tend to be treated as one type even if the fine detail between the spikes is significantly different.

REFERENCES

Baldwin, C.Y. and K.B. Clark (1997). Managing in an age of modularity. *Harvard Business Review*, 75(5), 84-93.

Chen, J.M. and C.S. Lin (2002). An optimal replenishment model for inventory items with normally distributed deterioration. *Production Planning & Control*, 13(5), 470-480.

Groebner, D.F., P.W.Shannon, P.C. Fry, and K.D. Smith (2005). *Business statistics – decision making approach (*6th edition). New Jersey: Pearson Prentice Hall.

Guilford, J.P. (1956). *Fundamental statistics in Psychology and Education*. New York: McGraw Hill.

Hayek, P.A. and M.K. Salameh (2001). Production lot sizing with the reworking of imperfect quality items produced. *Production Planning & Control*, 12(6), 584-590.

Montreuil, B. and M. Poulin (2005). Demand and supply network design scope for personalized manufacturing. *Production Planning & Control*, 16(5), 454-469.

Owehand, P. (2006). Forecasting with group seasonality. Retrieved August 1, 2006, from http://alexandria.tue.nl/extra2/200610743.pdf

Ozaki, T. and Iino, M. Iino (2001). An innovation approach to non-Gaussian time series analysis, *Journal of Applied Probability*, 38(A), 78-92.

Panisset, B.D. (1988). MRPII for repair/refurbish industries. *Production and Inventory Management Journal*, 29(4), 12-15.

Piller, F.T., K.S. Moeslein, and M. Christo (2004). Does mass customization pay? An economic approach to evaluate customer integration. *Production Planning & Control*, 15(4), 435-444.

Seifoddini, H. and P.M. Wolfe (1986). Application of the similarity coefficient method in group technology. *IIE Transactions*, 1986, 18(3), 271-277.

Su, J.C.P., Y.H. Chang, and M. Ferguson (2005). Evaluation of postponement structures to accommodate mass customization. *Journal of Operations Management,* 23(3), 305–318.

Sen, T. and J. Carroll (1973). Signal-to-noise ratio as a predictor of speech transmission quality in Audio and Electroacoustics. *IEEE Transactions*, 21(4), 384 – 387.

Timme, S.G. (2003). The real cost of holding inventory, *Supply Chain Management Review.* l7(4), 30.

Yan, H. (2001, July/August). Dynamic models of the term structure, *Financial Analysts Journal*, 60-74.

Yao Z., Y. Wu, and K.K. Lai (2005). Demand uncertainty and manufacturer returns policies for style-good retailing competition. *Production Planning & Control.* 16(7), 691-700.

PROJECT MANAGEMENT SOFTWARE SELECTION USING ANALYTICAL HIERARCHY PROCESS

J. S. Sutterfield, Florida A&M University
Steven Swirsky, Florida A&M University
Christopher Ngassam, Florida A&M University

ABSTRACT

A very large class of problems in management and administration are known as multi-attribute decision problems. Such problems involve decisions that must take into account large numbers of variables and objectives that cannot always be directly quantified. These problems are rendered even more unwieldy to handle by the fact that the variables and objectives are often in conflict. One very popular approach to handling this class of problem is called Analytical Hierarchy Process (AHP). AHP affords a technique for structuring this class of problems so that it can be given a quasi-quantitative structure. This permits handling any number of variables and objectives, quantitative and non-quantitative, compatible and conflicting. One practical problem of the multi-attribute type is the selection of project management software. In this paper there is developed an approach to the use of AHP in the selection of project management software.

INTRODUCTION

Decision Analysis involving multiple variables and objectives that can be quantified is rather commonplace. The methods for solving such problems are rather well established. One simply quantifies with some measure the variables and objectives involved in the problem, then chooses the appropriate solution methodology, obtains the necessary data and calculates an answer. However, with problems involving variables and objectives that cannot be measured, or at best can be only partly measured, the solution approach is not always so clear. This is particularly true when the variables and objectives involve personal preferences. The approach frequently taken with such problems is to simply prioritize the decision considerations and try to choose a solution that maximizes the desired decision quantities at minimum cost. Although this may not be too difficult with a limited number of decision quantities, it can become very difficult when the number of such quantities is large. In addition, the problem becomes vastly more complex when some of the decision quantities are in mutual conflict. Thus, making rational decisions under such circumstances may become extraordinarily difficult.

A number of techniques are available for arriving at decisions having multi-attributes. Unfortunately, most of them require that the attributes be measurable. When the attributes are of a more qualitative nature, the multi-attribute problem becomes much more difficult to handle. Herein lies the value and power of Analytical Hierarchy Process (AHP). With AHP it is possible to give a qualitative type problem a quasi-quantitative structure, and to arrive at decisions by expressing preferences for one attribute over another, and testing whether the preferences are rational consistent.

In this paper, we use AHP to analyze the selection process for project management software (PMS). We approach this problem by treating the PMS features offered by various companies as attributes. We then have a group of PM professionals evaluate the various features as to their importance and desirability. From these evaluations, paired comparison matrices are developed. Next, a set of matrices is developed for each software provider that evaluate just how well each provider's software satisfies each attribute. A consistency ratio is then computed to determine how rigorously rational consistency has been maintained in the analysis. Ordinarily, an AHP analysis would end here, but we extend our analysis by using the Student's "t" test for small sample sizes to determine a confidence interval within which the responses should lie. This analysis fills a void in the literature by demonstrating a perfectly generalized process for solving a practical multi-attribute decision problem, and for arriving at a high level of confidence that a rational decision will have been made.

LITERATURE REVIEW

AHP was originally conceived by Thomas L. Saaty as a structured method for solving problems involving decision variables or decision attributes, at least some of which, are qualitative, and cannot be directly measured (Saaty, 1980). It met with almost immediate acceptance and was applied to a wide range of problems. Very soon it began to be applied to executive decisions involving conflicts in stakeholder requirements and strategic planning (Saaty, 1982; Arbel & Orgler, 1990; Uzoka, 2005). The real power of AHP consists in its use of fairly elementary mathematics to structure complex problems in which decisions involve numerous decision makers, and multiple decision variables. Another facet of the power of the AHP approach consists in its ability to impose a quasi-quantitative character on decision problems in which the decision variables are not necessarily quantitative. The power and versatility of AHP are demonstrated by the wide range of problems to which the approach has been applied. It was used early for such problems as the justification of flexible manufacturing systems (Canada & Sullivan, 1989), and continues to be used in such applications (Chan & Abhary, 1996; Chandra & Kodali, 1998; Albayrakoglu, 1996). It has been used in such widely different applications as business crisis management (Lee & Harrald, (1999) and pavement maintenance (Ramadhan, Wahab & Duffuaa, 1999). Other interesting applications of AHP include the evaluation of personnel during the hiring process (Taylor, Ketcham & Hoffman, 1998), determination of investor suitability in structuring capital investment partnerships (Bolster & Janjigian & Trahan, 1995), apportioning public sector funds where numerous projects usually compete for limited resources (Barbarosoglu & Pinhas, 1995), and determination of real estate underwriting factors in the underwriting industry (Norris & Nelson, 1992). In the areas of accounting and finance, AHP has seen increasing use helping to direct the limited resources of auditors to their most effective and efficient use Baranoff, 1989; Arrington, Hillison & Jensen, 1984) and in the detection of management fraud Webber, 2001; Deshmukh & millet, 1998), and the prediction of bankruptcy (Park & Han, 2002). Thus, AHP is a very powerful, versatile and generalized approach for analyzing multi-attribute decision problems in which the decision considerations do not necessarily need to be directly measurable. Although AHP has been used in a very wide range of applications, this literature search has not disclosed any in which it has been used for project management software selection.

AHP METHODOLOGY

As noted above, the procedure for using Analytical Hierarchy Process is well established. It is a tribute to Dr. Saaty that his original work, done more than a quarter century ago, remains virtually unmodified. Thus, the present work will follow rather closely Dr. Saaty's original approach. This statement involves the following steps:

1) A clear concise description of the decision objective. In practice, this is probably best done by a team of 4 to 6 people who have good knowledge of the objective to be achieved, and who have a stake in arriving at the best possible decision.

2) Identification of those attributes that are to be included in arriving at the desired objective defined in step 1. These attributes are also best identified by a team of 4-6 people and preferably the same team that is used to identify the objective of the analysis.

3) Determination of any sub-attributes upon which an attribute might be based.

4) Identification of a set of alternatives that are thought to achieve, at least partially, the desired objective. We say "at least partially" because probably no alternative will completely provide all desired attributes. However, the alternatives should be selected because of providing some degree of satisfaction to all attributes.

5) Once the attributes are identified, they are entered into a preference matrix, and a preference or importance number is assigned to reflect the preference for/importance of each attribute relative to all others. The strength of

preference/importance is indicated by assigning a preference/importance number according to the following rating scale:

Preference Number for Attribute A over Attribute B	Then Attribute A is ….. over (or to) Attribute B
9	Absolutely more important or preferred
7	Very strongly more important or preferred
5	Strongly more important or preferred
3	Weakly more important or preferred
1	Equally important or preferred

Intermediate degrees of preference for attribute A over attribute B are reflected by assigning even numbered ratings "8", "6", "4" and "2" for one attribute over another. For example, assigning a preference number of "6" would indicate a preference for attribute "A" over attribute "B" between "Strongly more important or preferred" and "Very strongly more important or preferred." Also, the logical principle of inverse inference is used, in that assigning attribute "A" a preference rating of, say "8", over attribute "B", would indicate that attribute "B" were only "1/8" as important as attribute "A." Either such a preference matrix is developed by each participant in the process, or the participants arrive at agreement as to the preference numbers for the attributes.

6) Mathematical operations are then used to normalize the preference matrix and to obtain for it a principal or characteristic vector.

7) Next, participants collectively rate each of the possible approaches/options for satisfying the desired objective(s). This is done along the same lines as shown above for attributes. The same rating scale is used, except here the ratings reflect how well each possible approach/option satisfies each of the attributes. For example, suppose that Alternatives "1" and "2" were being compared as to how well each satisfies Attribute "A" relative to the other. If Alternative "1" were given a preference rating of, say 5, relative to Alternative "2" it would mean that for Attribute "A", Alternative "1" was thought to satisfy it 5 times as well as Alternative "2". Or conversely, Alternative "2" was believed to satisfy Attribute "A" only "1/5" as well as did Alternative "1". Thus, a set of preference matrices is developed, one matrix for each attribute, that rates each alternative as to how well it satisfies the given attribute relative to all other alternatives under consideration.

8) Again, as with the attribute preference matrices, mathematical operations are used to normalize each preference matrix for alternatives, and to obtain a principal or eigenvector for each. When complete, this provides a set of characteristic vectors comprised of one for each attribute. Each of these vectors reflects just how well a given alternative/option satisfies each attribute relative to other alternatives.

9) Then with an eigenvector from "8" above, we cast these into yet another matrix, in which the number of rows equals the number of alternatives/options, and the number of columns equals the decision attributes. This matrix measures how well each alternative/option satisfies each decision attribute.

10) Next, the matrix from "9" above is multiplied by the characteristic vector from "6" above. The result of this multiplication is a weighted rating for each alternative/option indicating how well it satisfies each attribute. That alternative/option with the greatest weighted score is generally the best choice for satisfying the decision attributes, and thus achieving the desired objective.

11) Last, it is necessary to calculate a consistency ratio (C. R.) to ensure that the preference choices in the problem have been made with logical coherence. The procedure for calculating the C. R. may be found in a number of works on AHP (Canada & Sullivan, 1989) and will not be repeated here. According to Saaty, if choices have been made consistently the C. R. should not exceed 0.10. If it does, it then becomes necessary to refine the analysis by having the participants revise their preferences and re-computing all of the above. Thus, the entire process is repeated until the C. R. is equal to or less than 0.10.

12 Having completed steps "1" thru "11", that alternative/option with the greatest weighted score is selected as best satisfying the decision alternatives, and thus achieving the desired objective.

APPLICATION OF AHP METHODOLOGY

The usual practice for developing an AHP analysis is to develop a separate set of matrices for each of the participants, or to have participants brainstorm until they arrive at a group decision for the required matrices. However, in the instant case, time and distance made the usual approach impossible. Consequently, the authors have done the next best thing and have averaged participants' ratings. The average ratings are used in the subsequent analysis, and the sensitivity analysis done on the averaged ratings would be analogous to reconvening the participants in order to refine the ratings and thus arrive at a satisfactory C. R. The results of this are shown in Figures 1 thru 7 below. The codes for the Software Features and Companies (alternatives/options) used in the figures immediately follow:

Software Feature	Code	Company
Multiple Networks Supported	A	1
Simultaneous Multiple User Access	B	2
Full Critical Path	C	3
Time/Cost	D	
Multi-project PERT Chart	E	
Resources Used	F	

ANALYSIS OF RESULTS

As indicated above, the attribute preference matrices for participants were averaged for each attribute, so that the final ratings shown in Fig. 1 are the result of an attribute by attribute average, rather than a matrix by matrix average.

Figure 1: Group averaged attribute averaged preference matrix

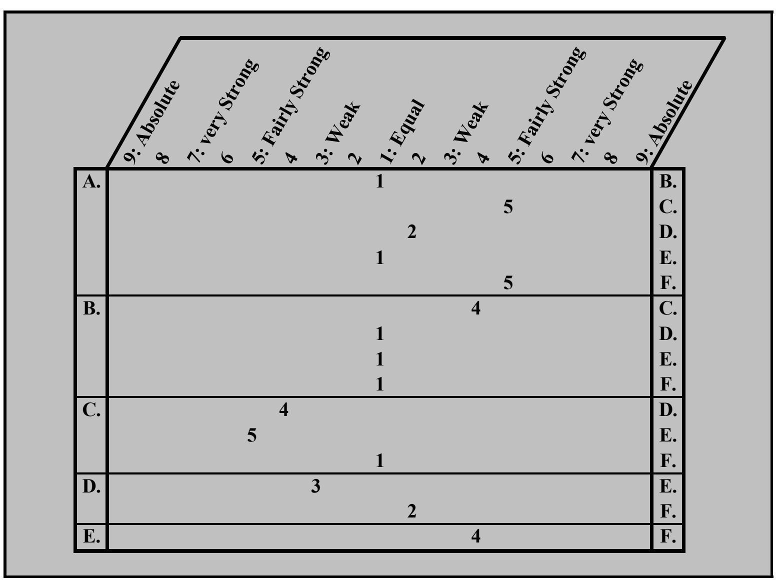


Fig.2 below shows the normalized group preference matrix resulting from normalizing the averaged values in Fig. 1. The characteristic vector for this matrix is shown in the rightmost column.

Fig 2: Normalized group attribute performance preference matrix

							Decimal Equivalents						Row	Row
	A	B	C	D	E	F	A	B	C	D	E	F	Sums	Averages
A.	1.000	1.000	0.200	0.500	1.000	0.200	0.067	0.111	0.069	0.057	0.067	0.051	0.421	0.070
B.	1.000	1.000	0.250	1.000	1.000	1.000	0.067	0.111	0.086	0.113	0.067	0.253	0.697	0.116
C.	5.000	4.000	1.000	4.000	5.000	1.000	0.333	0.444	0.345	0.453	0.333	0.253	2.162	0.360
D.	2.000	1.000	0.250	1.000	3.000	0.500	0.133	0.111	0.086	0.113	0.200	0.127	0.770	0.128
E.	1.000	1.000	0.200	0.333	1.000	0.250	0.067	0.111	0.069	0.038	0.067	0.063	0.414	0.069
F.	5.000	1.000	1.000	2.000	4.000	1.000	0.333	0.111	0.345	0.226	0.267	0.253	1.536	0.256
						Column sums =	1.000	1.000	1.000	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 1.000 & 0.200 & 0.500 & 1.000 & 0.200 \\ 1.000 & 1.000 & 0.250 & 1.000 & 1.000 & 1.000 \\ 5.000 & 4.000 & 1.000 & 4.000 & 5.000 & 1.000 \\ 2.000 & 1.000 & 0.250 & 1.000 & 3.000 & 0.500 \\ 1.000 & 1.000 & 0.200 & 0.333 & 1.000 & 0.250 \\ 5.000 & 1.000 & 1.000 & 2.000 & 4.000 & 1.000 \end{vmatrix} \text{ X } \begin{vmatrix} 0.070 \\ 0.116 \\ 0.360 \\ 0.128 \\ 0.069 \\ 0.256 \end{vmatrix} = \begin{vmatrix} 0.442 \\ 0.729 \\ 2.287 \\ 0.809 \\ 0.434 \\ 1.614 \end{vmatrix}$$

$$D = \begin{vmatrix} 6.3171 & 6.2845 & 6.3528 & 6.320 & 6.285 & 6.305 \end{vmatrix}$$

Lambda Max = 6.3107 C.I. = 0.0621 C.R. = 0.050

As can be seen in the last row of Fig. 2, the C. R. for these averaged attributes is 0.05, a value well within that prescribed by Saaty for logical coherence. This result is very interesting because only one of the participants achieved a C. R. of less than 0.10, this being a C. R. of 0.09. The other six had C. R.s ranging from 0.140 to 0.378. This means that the group as a whole was more consistent in rating the attributes than was each individual. Since the averaged ratings resulted in a C. R. of less than 0.10, it was unnecessary at this point to perform sensitivity analysis on the averaged attribute ratings. However, had sensitivity analysis been necessary, each individual rating would have been adjusted slightly upward or downward until a C. R. of equal to or less than 0.10 were obtained. Since only one of the participants had a C. R. less than 0.10, the averaging process was analogous to calling the participants into caucus and having them arrive at mutually agreeable ratings for all attributes. The conclusion to be drawn from this is that a group of individuals will frequently make more rational choices than will a single individual. As a practical matter, it would seem to be more expeditious to first obtain the individual attribute preference matrices, average them and then call participants together to negotiate more consistent ratings. Alternatively, each individual would have to revise his individual results and have a new C. R. computed. This process would then have to be continued until all participants had individually arrived at C. R.s of equal to or less than 0.10. The former approach of averaging and caucusing employs a Delphi approach to decision making, which has been shown usually to lead to better decision outcomes than decisions made in isolation.

Ordinarily, selection of the options/alternatives (in our case software providers) for satisfying the above attributes would also be made by the primary participants. Each participant would rate each option/alternative as to how well it satisfied each attribute. However, in the present case this part was done by the authors based on an analysis of several project management software products and their common and unique features. Those ratings, as with the attribute preference ratings, were then cast into matrices and C. R.s were calculated for each software provider. The final results for this are shown in Figs. 3 thru 8.

Figure 3: Multiple networks supported

				Decimal equivalents			Row	Row
	P	Q	R	P	Q	R	sums	averages
Company 1	1.000	0.500	4.000	0.308	0.304	0.333	0.945	0.315
Company 2	2.000	1.000	7.000	0.615	0.609	0.583	1.807	0.602
Company 3	0.250	0.143	1.000	0.077	0.087	0.083	0.247	0.082
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 0.500 & 4.000 \\ 2.000 & 1.000 & 7.000 \\ 0.250 & 0.143 & 1.000 \end{vmatrix} \text{ X } \begin{vmatrix} 0.315 \\ 0.602 \\ 0.082 \end{vmatrix} = \begin{vmatrix} 0.946 \\ 1.810 \\ 0.247 \end{vmatrix}$$

$$\left| D \right| = \left| 3.002 \quad 3.004 \quad 3.0008 \right|$$

Lambda max = 3.0023 C.I. = 0.0012 C.R. = 0.002

Figure 4: Simultaneous multiple user access

	P	Q	R	Decimal equivalents P	Decimal equivalents Q	Decimal equivalents R	Row sums	Row averages
Company 1	1.000	1.000	4.000	0.444	0.461	0.364	1.270	0.423
Company 2	1.000	1.000	6.000	0.444	0.461	0.545	1.451	0.484
Company 3	0.250	0.167	1.000	0.111	0.077	0.091	0.279	0.093
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 1.000 & 4.000 \\ 1.000 & 1.000 & 6.000 \\ 0.250 & 0.167 & 1.000 \end{vmatrix} X \begin{vmatrix} 0.423 \\ 0.484 \\ 0.093 \end{vmatrix} = \begin{vmatrix} 1.279 \\ 1.465 \\ 0.280 \end{vmatrix}$$

$$| D | = | 3.023 \quad 3.028 \quad 3.0057 |$$

Lambda max = 3.0189 C.I. = 0.0095 C.R. = 0.019

Figure 5: Full critical path

	P	Q	R	Decimal equivalents P	Decimal equivalents Q	Decimal equivalents R	Row sums	Row averages
Company 1	1.000	9.000	9.000	0.818	0.750	0.857	2.425	0.808
Company 2	0.111	1.000	0.500	0.091	0.083	0.048	0.222	0.074
Company 3	0.111	2.000	1.000	0.091	0.167	0.095	0.353	0.118
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 9.000 & 9.000 \\ 0.111 & 1.000 & 0.500 \\ 0.111 & 2.000 & 1.000 \end{vmatrix} X \begin{vmatrix} 0.808 \\ 0.074 \\ 0.118 \end{vmatrix} = \begin{vmatrix} 2.532 \\ 0.222 \\ 0.355 \end{vmatrix}$$

$$| D | = | 3.132 \quad 3.009 \quad 3.0208 |$$

Lambda max = 3.0539 C.I. = 0.027 C.R. = 0.054

Figure 6: Time/cost trade-off

				Decimal equivalents			Row	Row
	P	Q	R	P	Q	R	sums	averages
Company 1	1.000	0.500	0.333	0.167	0.143	0.182	0.491	0.164
Company 2	2.000	1.000	0.500	0.333	0.286	0.273	0.892	0.297
Company 3	3.000	2.000	1.000	0.500	0.571	0.546	1.617	0.539
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 0.500 & 0.333 \\ 2.000 & 1.000 & 0.500 \\ 3.000 & 2.000 & 1.000 \end{vmatrix} \text{ X } \begin{vmatrix} 0.164 \\ 0.297 \\ 0.539 \end{vmatrix} = \begin{vmatrix} 0.492 \\ 0.894 \\ 1.625 \end{vmatrix}$$

$$|\ D\ | = |\ 3.004 \quad 3.008 \quad 3.0144 \ |$$

Lambda max = 3.0088 C.I. = 0.0044 C.R. = 0.009

Figure 7: Multi-project PERT chart

				Decimal equivalents			Row	Row
	P	Q	R	P	Q	R	sums	averages
Company 1	1.000	1.000	2.000	0.400	0.400	0.400	1.200	0.400
Company 2	1.000	1.000	2.000	0.400	0.400	0.400	1.200	0.400
Company 3	0.500	0.500	1.000	0.200	0.200	0.200	0.600	0.200
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 1.000 & 2.000 \\ 1.000 & 1.000 & 2.000 \\ 0.500 & 0.500 & 1.000 \end{vmatrix} \text{ X } \begin{vmatrix} 0.400 \\ 0.400 \\ 0.200 \end{vmatrix} = \begin{vmatrix} 1.200 \\ 1.200 \\ 0.600 \end{vmatrix}$$

$$|\ D\ | = |\ 3.000 \quad 3.000 \quad 3 \ |$$

Lambda max = 3 C.I. = 0 C.R. = 0.000

Figure 8: Resources used

				Decimal equivalents			Row	Row
	P	Q	R	P	Q	R	sums	averages
Company 1	1.000	0.143	0.200	0.077	0.097	0.048	0.221	0.074
Company 2	7.000	1.000	3.000	0.538	0.678	0.714	1.930	0.643
Company 3	5.000	0.333	1.000	0.385	0.226	0.238	0.848	0.283
			Column totals =	1.000	1.000	1.000		1.000

$$\begin{vmatrix} 1.000 & 0.143 & 0.200 \\ 7.000 & 1.000 & 3.000 \\ 5.000 & 0.333 & 1.000 \end{vmatrix} \quad X \quad \begin{vmatrix} 0.074 \\ 0.643 \\ 0.283 \end{vmatrix} = \begin{vmatrix} 0.222 \\ 2.008 \\ 0.866 \end{vmatrix}$$

$$| \, D \, | = | \, 3.013 \quad 3.121 \quad 3.0628 \, |$$

Lambda max = 3.0657 C.I. = 0.0328 C.R. = 0.066

The results from this part of the process, as can be seen, turned out quite well. Only the C. R. for the third attribute, "Full Critical Path," turned out to have a C. R. greater than 0.10. This C. R. was originally 0.54, a value well above the allowable. Sensitivity analysis was performed on this matrix until the present C. R. of 0.05 was obtained. The remainder of the C.R.s were well below 0.10, as can be seen. Remarkably, the C. R. for "Multi-project PERT Chart" turned out to be 0.000, indicating perfect consistency among the choices. An interesting problem arose for the "Full Critical Path" attribute in that neither Company 2 nor 3 offered this feature. It would be expected that a company would receive a rating of "0" for failure to provide a feature. However, this would have resulted in Company 1 being infinitely preferred over Companies 2 and 3. This would have been inconsistent with the rating scale, which only ranges from "1" to "9". To address this problem, a preference rating of "9" was assigned for Company 1 over Companies 2 and 3. This resulted in ratings of 1/9 or 0.11 being assigned as to the preference for Companies 2 and 3 over Company 1. Thus, the final values for these ratings turned out near "0", though not exactly "0". Once all matrices were assured to have a C. R. of equal to or less than 0.10, their values were cast into a final matrix as shown in Fig. 9 below.

Figure 9: Weighted alternative evaluations

	A	B	C	D	E	F	Weighted Alternative
	0.070	0.116	0.360	0.128	0.069	0.256	
Company 1	0.315	0.423	0.808	0.164	0.400	0.074	0.430
Company 2	0.602	0.484	0.074	0.297	0.400	0.643	0.355
Company 3	0.082	0.093	0.118	0.539	0.200	0.283	0.215
						Total =	1.000

In this matrix, the top row of six numbers will be seen to be the principal or characteristic vector from the attribute preference matrix. Each of the six columns of three numbers under this top row will be seen to be the principal vectors for Companies 1, 2 and 3 as obtained from the analysis described in the immediately preceding paragraph. The rightmost column titled "Weighted Alternative Evaluations" contains the results of this analysis. It will be seen that Company 1 best satisfies the original attributes in that it has a rating of 0.43. Companies 2 and 3 follow in order, with ratings of 0.36 and 0.21, respectively. These ratings are quite stable. Sensitivity analysis was done on these results by varying the ratings in the original attribute preference matrix. However, these results remained virtually unchanged. Company 1 would then be selected as best satisfying the desired attributes.

In order to test for the consistency of the participant attribute preferences, the C. R.s were subjected to a Student's "t" test to obtain a confidence interval. The "t" test is used for applications in which the sample size is smaller than about twenty, the samples reasonably can be assumed to be normally distributed and the distribution of the sample standard deviation is independent of the distribution of the sample values. Having made these assumptions, we obtained the following results:

PM	C.R.
1	0.14
2	0.15
3	0.14
4	0.20
5	0.09
6	0.38
7	0.32
y-bar = 0.20 sample standard deviation = 0.11	

Next a “t” test is performed with six degrees of freedom and a confidence interval of 95%. This yields theoretical lower and upper confidence limits of -0.07 and 0.47. Since a C.R. of zero means perfectly consistent choices, a C.R. of less than zero is impossible. This means that the practical value of the lower confidence limit for this application is zero. Then for a 95% confidence level, all of the C.R. values for our PMs should lie within a range from 0.00 to 0.47. As can be seen from the above C.R.s, all do indeed lie within this confidence interval as obtained from the “t” test. Thus, the responses of all PMs were found to be highly consistent both from the standpoint of the C.R. obtained from the averaged attribute responses, and the confidence interval obtained from the “t” distribution. A decision maker having the foregoing analysis could thus conclude that Software Provider #1 would provide the best software package for the application envisioned.

CONCLUSIONS

The thesis of this paper has been that Analytical Hierarchy Process offers a very powerful, flexible and general approach to decision situations in which the decision variables are not necessarily quantifiable. Although it was not possible to quantify any of the attributes in the above example, it was, nonetheless possible to rationally express preferences for one above another. We say rationally because a decision maker can always express a preference for one attribute over another once a definite objective is established. In this case one preference is more desirable than another in so far as it better satisfies or achieves the objective(s), or vice versa. We have shown in the foregoing analysis that even though none of the C.R.s for a group of respondents may satisfy the 0.1 criterion specified by Saaty, that this obstacle may be overcome by averaging the responses for each attribute. We have demonstrated how to handle the problem in which a given provider does not offer one or more of the desired attributes/features. We have also shown through sensitivity analysis that once the C.R.s for all matrices have been brought to values equal to or less than 0.10 using sensitivity analysis, that the results are remarkably stable. They are relatively insensitive to fairly large changes in the original attribute preferences. We tested the consistency of the responses using the Student’s “t” distribution and found that they can be placed within a 95% confidence interval, which indicates a very good degree of consistency among PM respondents. Finally, in using AHP in the problem of software selection, we have demonstrated the great power and flexibility of this little known process in solving a very wide range of practical decision problems which are otherwise virtually intractable.

REFERENCES

Albayrakoglu, M. Murat, (1996). Justification of New Manufacturing Technology: A Strategic approach Using the Analytical Hierarchy Process, *Production and Inventory Management Journal,* Alexandria: first Quarter 1996, 37(1), 71-77.

Arbel, A. & Y.E. Orgler, (1990). An Application of the AHP to Bank Strategic Planning: The Mergers and Acquisitions Process, *European Journal of Operational Research*, .48(1), 27-37.

Arrington, C.E., W. Hillison & R.E. Jensen, (1984). An Application of Analytical Hierarchy Process to Model Expert Judgments on Analytical Review Procedures, *Journal of Accounting Research,* 22(1), 298-312.

Bagranoff, N.A., (1989). Using an Analytic Hierarchy Approach to Design Internal Control Systems, *Journal of Auditing and EDP*, 37-41.

Barbarosoglu, Gulay, & David Pinhas, (1995). Capital Rationing in the Public Sector Using the Analytical Hierarchy Process, *The Engineering Economist*, Norcross, Summer, 40(4), 315-342

Bolster, Paul J., Vahan Janjigian & Emery A. Trahan, (1995). Determining Investor Suitability Using the Analytical HierarchyProcess, *Financial Analysts Journal*, Charlottesville, July/Aug., 51(4), 63-76

Canada, John R. & William G. Sullivan, (1989). *Economic and Multiattribute Evaluation of Advanced Manufacturing Systems,* Prentice-Hall, Englewood cliffs, New Jersey, Ch. 10, 268 - 279..

Chan, Felix T. S. & K. Abhary, (1996). Design and Evaluation of Automated Cellular Manufacturing Systems with Simulation Modeling and the AHP Approach: A Case Study, *Integrated Manufacturing Systems*, Bradford: 7(6), 39.

Chandra, S. & Rambadu Kodali, (1998). Justification of Just-in-Time Manufacturing systems for Indian Industries, *Integrated Manufacturing Systems*, Bradford: 9(5), 314

Deshmukh, Ashutosh & Ido Millet, (1998). An Analytic Hierarchy Process Approach to Assessing the Risk of Management Fraud, *Journal of Applied Business Research*, 15(1), 87-102.

Lee, Young-Jai & John R. Harrald, (1999). Critical Issue for Business Area Impact Analysis in Business Crisis Management: Analytical Capability, *Disaster Prevention and Management,* Bradford: 8(3), 184.

Norris, Daniel M. & Mark Nelson, (1992). Real Estate Loan Underwriting factors in the Insurance Industry, *Real Estate Finance*, New York, Fall, 9(3), 79.

Park, Cheol-Soo & Ingoo Han, (2002). A Case-based Reasoning with the Feature Weights derived by Analytic Hierarchy Process for Bankruptcy Prediction, *Expert Systems with Applications,* 23(3), 255-264.

Ramadhan, Rezquallah H., Hamad I. Al-Abdul Wahab & Salih O. Duffuaa, (1999). The Use of an Analytical Hierarchy Process in Pavement Maintenance Priority Ranking, *Journal of Quality in Maintenance Engineering,* Bradford: 5(1), 25

Saaty, Thomas L. (1980). *The Analytical Hierarchy Process*, New York: McGraw-Hill.

Saaty, Thomas L.(1982). *Decision Making for Leaders*, Belmont, CA: Wadsworth Publishing Company, Inc.

Taylor, Frank A. III, Allen F. Ketcham & Darvin Hoffman, (1998). Personnel Evaluation with AHP, *Management Decision,* London, 36(10), 679

Uzoka, Faith M.E. (2005). Analytical Hierarchy Process-based system for Strategic Evaluation of Financial Information, *Information Knowledge Systems Management,* 5(1), 49-61.

Webber, Sally. (2001). The Relative Importance of Management Fraud Risk Factors, *Behavioral Research in Accounting*, 1(1), 1-10.

OBJECTS-FIRST VS. STRUCTURES-FIRST APPROACHES TO OO PROGRAMMING EDUCATION: AN EMPIRICAL STUDY

Richard A. Johnson, Missouri State University
Duane R. Moses, Missouri State University

ABSTRACT

Software development is a cornerstone of information technology and computer programming is a key element of software development. Teaching computer programming to undergraduates can be challenging. Adding to the challenge is the near universality of the object-oriented paradigm in computer programming courses. A key question facing programming instructors is whether to take an objects-first approach or a structures-first approach in the classroom. Conventional wisdom seems to favor an objects-first approach, but no hard empirical evidence exists on this question. This study performs a field experiment by having two nearly equivalent sections of introductory Java programming test these two fundamental approaches to OO programming education. The results clearly indicate that students who take an objects-first approach to OO programming outperform those who take a structures-first approach.

INTRODUCTION

The teaching of introductory programming is a foundation in many computer information systems (CIS) and computer science (CSC) curricula. In recent years, virtually all introductory programming courses have shifted from the procedural approach to the object-oriented (OO) approach. Most beginning programming courses appear to be teaching Java , C++, or one of the Visual Studio .NET languages (Visual Basic, C#, or J#) as evidenced by the popularity of various computer programming texts. All of these programming languages are OO, as contrasted with the purely procedural languages of Fortran, Pascal, COBOL, and C.

The basis of any type of computer programming involves the three programming 'structures': sequence (do A, do B, do C, …), selection (if…else decisions), and repetition (while or for loops). The basics of OO programming involve creating classes that serve as templates for instantiating objects in computer memory. Objects have attributes (instance fields) and behaviors (methods) and usually represent things in the real world (such as students, products, or airline reservations). While learning the basics of structured programming (sequence, selection, and repetition) is not always easy for most beginning students of CIS and CSC, it is the general consensus that learning OO programming concepts and techniques may be even more challenging for most students (Sheetz, et al., 1997).

Therefore, one of the most relevant questions regarding how OO programming courses should be taught is whether it is better to teach structured programming concepts first, followed by OO programming concepts, or vice versa. It appears that most authors claim it is better to teach objects first ('early objects') in order to ingrain the student with OO concepts and techniques early and often, thus ensuring greater success in later OO programming (Thramboulidis, 2003; Ragonis & Ben-Ari, 2005). Although the 'objects first' (OF) approach may sound more plausible than a 'structures first' (SF) approach, there appears to be no empirical evidence to support the claim. The purpose of this study is to perform a field experiment to test the claim that OF is superior to SF.

RESEARCH METHOD

The research question driving this study is: What effect does teaching an objects-first approach (vis-à-vis teaching a structures-first approach) have on the performance of introductory programming students in understanding OO concepts and writing OO programs? The hypothesis being tested is that there is no difference in the performance of introductory programming students when provided with an objects-first or a structures-first approach to OO programming.

To test this hypothesis, the authors of this study each taught one section of introductory OO programming (CIS 260) to a section of about 25 students during the Fall 2007 semester at Missouri State University (MSU). These instructors had already been scheduled to teach these sections, so it was not possible for the same instructor to teach both sections due to scheduling conflicts. To minimize the differences in course delivery, the following steps were taken:

1. Both instructors have approximately the same amount of experience (several years) teaching computer programming (and OO programming) in the CIS Department at MSU.
2 The instructors selected two different texts written by the same author (Gaddis, 2008a; Gaddis, 2008b). The only significant difference between the two texts is the ordering of the chapters, one presenting objects and classes in the early chapters while the other doing so in the later chapters. These two texts were designed specifically by Gaddis to support either an OF or a SF approach to teaching OO programming using Java. The reading material, examples, and end-of-chapter problems throughout the texts are essentially the same with the exception of the ordering of the chapters.
3 Both instructors carefully planned the delivery of material (such as chapters covered, assignment problems, and exam questions) to exactly match in both sections (OF and SF) with the only difference being the textbooks used.

4. Both instructors taught their respective sections in the same lab setting using similar lecture/discussion/demonstration techniques while assisting students in writing programs during specified lab time. No graduate assistants were used in the delivery of these courses or in the grading of assignments and exams. The instructors were in total control of these courses.
5. Both instructors communicated with each other frequently during the semester to ensure that they were on track with the plan to deliver the same material in the same manner, with the exception of the OF and SF approaches.
6. Both instructors consulted to design all items (multiple choice and coding exercises) on the performance exams in the courses. The specific criteria used to grade programs submitted by the students were carefully chosen by both instructors together. Students in the different sections were given the same items on exams.
7. Both instructors graded exams taken by the students (to assess performance) together, constantly consulting with each other on how to apply the predetermined criteria to individual programs submitted by students on the exams.

As can be seen, the instructors were extremely careful to try to eliminate any built-in bias in the delivery of the two courses. The only significant variable in the two sections was whether OO programming concepts and techniques were delivered before basic programming structures (OF) or after basic programming structures (SF).

STUDENT BACKGROUND AND DEMOGRAPHICS

Data about the students were also collected during the first week of class to ensure that both groups (OF and SF) were similar in background and ability. Students were given a survey to determine demographic data such as gender, age, college class, major and minor areas of study, background with Java and other programming languages, background in using computers, and desire to learn computer programming. The college GPA and ACT scores were also collected for all students.

PERFORMANCE MEASURES

Three exams were administered to each section of students during the semester (Exam 1, Exam 2, and the final exam). Each exam consisted of 25 multiple-choice questions that covered programming concepts and one programming problem. Students were asked to write a complete Java program on paper on each exam. The instructors thought it would be better to have the students write programs on paper, instead of on a computer, so that credit could be given for code that was close to correct, although points were deducted for incorrect syntax or program logic. (Incorrect syntax on the computer would result in failure to compile and perhaps lead to students ceasing to write additional code.)

While three exams were administered to both the OF and SF sections, only the first and last exams were used in this study because only these two exams were identical for both sections. Exam 1 was the same for both sections because it covered the first two chapters in both texts (basic Java programming involving variables, simple algorithms, and the sequence structure). Exam 2 was not used in this study because it covered different material for each section. During this middle segment of the courses, the OF section covered two chapters on objects and classes while the SF section was covering decisions, loops, and methods. Toward the end of the semester, the OF section was catching up on programming structures while the SF section was catching up on objects and classes. The final exam was identical for both groups, consisting of multiple choice questions on OO concepts and an OO programming problem.

Thus, each exam had a concepts section (assessed via the multiple-choice questions) and a techniques section (assessed via the programming problem). Grading the multiple-choice items was easy but much more care was required when grading the hand-written student programs. As described earlier, the instructors worked together in grading all programs by their respective students to ensure that a consistent grading method was employed. This required very frequent consultations between the instructors during the grading process.

Examples of some of the multiple-choice questions used on Exam 1 are given below:

1. *A runtime error usually the result of*
   a syntax error b) bad data c) a compilation error d) a logical error

2. *What is the value of z after the following statements have been executed?*
   int x = 4, y = 33;
   double z;
   z = (double) (y/x);
   a) 8.0 b) 4 c) 0 d) 8.25

Following is the programming problem on Exam 1:

Write a complete Java application that asks for the user's name and age (in years). The program should then calculate the number of years until the person can retire. Use JOptionPane dialog boxes for input and output. Use a named constant for the retirement age of 65. Add appropriate comments at the beginning of the program. Follow best programming practices and Java programming conventions. (Note: This is an example of a purely structured programming problem, not OO, used on the first exam and represents a fairly simple task.)

Examples of some of the multiple-choice questions used on the final exam are given below:

1. *It is a common practice in OO programming to make all of a class's*
   a) fields private b) fields and methods public c) methods private d) fields public

2. *What is the value of z after the following statements have been executed?*
   Double x = 45678.259;
   DecimalFormat formatter = new DecimalFormat("#,###,##0.00");
   System.out.println(formatter.format(x));
   a) 45,678.3 b) 45,678.26 c) 45678.259 d) 0,045,678.26

Following is the programming problem on the final exam:

Write two Java programs, Book.java and BookApp.java. Use JOptionPane for input and output. The Book class has three instance fields: bookID (integer), bookTitle (String), and bookPrice (double). The Book class has a static field, TAX_RATE = 0.05, which is a named constant. The Book class has two constructors: one has two parameters only (theId and theTitle) while the other has three parameters (theId, theTitle, and thePrice). The Book class has a method called calculateTax() which usues TAX_RATE and bookPrice to return the tax on a book. The Book class has a method called calculateFinalCose(), which calculates the final price of the book to the customer (bookPrice + tax). The Book class has get and set methods for all instance fields. The BookApp class has a while loop that allows the user to continue to create new book objects until indicating that the program should stop. Within the while loop, the BookApp class gets input from the user for bookId, bookTitle, and bookPrice, creates a Book object, calculates the tax using the calculateTax() method and calculates the final cost using the calculateFinalCost() method. Within the while loop, the BookApp class displays all the information about the book object created including the ID, title, price, tax, and final cost. (Note: This is an example of the OO programming problem that all students of both sections, OF and SF, took on a two-hour final exam, and is therefore more involved than the programming problem used on the first exam).

As can be seen, the multiple-choice questions on Exam 1 involve only basic programming concepts and the programming problem on Exam 1 is a straightforward and procedural. However, the final exam multiple-choice questions involve understanding important OO concepts (public and private class members, the new operator, and an object calling a method). The OO programming problem on the final exam involves writing a class used to create objects (containing instance fields, constructors, and methods) and writing an application class that creates objects, calls instance methods, and produces output.

DATA

Table 1 shows results of various demographic data for the OF and SF groups as well as the results from Exam 1 and the final exam. Where appropriate, small-sample T-tests for equivalent means were performed:

Table 1: Means and hypothesis tests for the objects-first (OF) section and the structures-first (SF) section					
		Objects-first	Structures-first	t-value	H_0: $\mu_{OF} = \mu_{SF}$, α=.05
1	# Male	22	20		
2	# Female	3	4		
3	Total students	25	24		
4	Age	20.8	21.0	-0.31	FTR H_0
5	Previous GPA	3.00	2.99	0.12	FTR H_0
6	ACT score	24.1	23.7	0.26	FTR H_0
7	# Freshmen	0	1		
8	# Sophomores	10	14		
9	# Juniors	12	7		
10	# Seniors	3	2		
11	# CIS majors	15	13		
12	# Non-CIS majors	10	11		
13	Previous college computer courses	1.36	1.29	0.23	FTR H_0
14	I am comfortable using computers	4.13*	4.68*		
15	This course is important to my career	3.43*	3.41*		
16	Exam 1-Multiple Choice Items	83.3%	83.2%	0.04	FTR H_0
17	Exam 1-Programming Problem	90.3%	85.9%	1.18	FTR H_0
18	Final Exam-Multiple Choice Items	74.4%	66.6%	1.66	FTR H_0
19	Final Exam-Programming Problem	87.9%	79.0%	2.98	Reject H_0 $\alpha < .005$
*Used a Likert scale of 1-5 for strongly agree to strongly disagree					

DISCUSSION

The data in Table 1 provide several interesting results. Demographically, items 1-15 demonstrate that the two sections of students are extremely similar. The distribution of males and females is nearly identical and the mean ages are statistically the same. Items 5-6 show that the prior abilities of the two groups are also statistically equivalent. Items 7-10 illustrate that the distribution of students by class is nearly the same for both groups while items 11-12 reveal that the distribution of CIS and non-CIS majors is nearly the same. (Statistical tests for equivalent distributions were not performed, but casual inspection largely supports these conclusions.) Items 13-15 point to the fact that both groups have similar backgrounds with computers and motivation for taking introductory Java programming.

The exam results are found in items 16-19. These are the results of the first exam and the final exam for both groups. Recall that both of these exams were identical for both groups of students. Exam 1 covered Java programming basics such as Java syntax, variables, and simple algorithms. The final exam involved object-oriented concepts and programming techniques. The multiple choice items on these exams tested for the understanding of programming concepts while the programming problem tested for knowledge of writing correct code for complete applications. The results show that both groups, OF and SF, were statistically equivalent in all areas except that of actual OO programming. The OF group averaged significantly higher on the programming segment of the final exam.

Items 16-19 in Table 1 do appear to illustrate an interesting trend. Item 16 shows that both groups were virtually equivalent in their understanding of basic introductory Java programming concepts. However, item 17 shows a slight weakness in the SF group. Keep in mind that both groups covered exactly the same material prior to Exam 1. It could be the case that the SF group did have some inherent weaknesses compared to the OF group initially, resulting in a lower mean for the Exam 1 programming segment, although the difference was not statistically significant. The difference between the two groups grew larger by the end of the course when the final exam was administered. The SF group did have a lower mean score on the concepts segment of the final exam, although not statistically significant. But statistical significance was apparent on the programming segment of the final exam.

While these results could be explained by some variation in the delivery of the courses by the two different instructors, such variation was minimized by taking measures to ensure a high degree of consistency. It is much more likely that the early objects approach was instrumental in the higher scores for the OF group on the final exam. If for no other reason, the OF group was exposed to the concepts and techniques of OO programming for several more weeks than the SF group. It could be the case that over time, the SF group could master OO programming as well as the OF group. However, within the short confines of a first course in Java programming, an OF approach could result in greater success for the student.

CONCLUSION

Learning programming is not an easy task for the novice student. Learning OO programming is an even more daunting task. This study compared the performance of two nearly identical groups of introductory programming students. One group studied objects and classes very early in the semester (the objects-first, or OF, group) while the other group studied the basic programming structures (sequence, selection, and repetition) before objects and classes (the structures-first, or SF, group). Both groups took the same first exam (covering only basic Java programming) before they diverged into either the OF or SF approaches. Then both groups took the same final exam which covered full OO development. The OF and SF groups were statistically identical in their performance on the first exam, but the OF group performed significantly better on the programming segment of the final exam. These experimental results point to the possible superiority of an objects first approach to teaching novice programming students, which may lead to higher performance levels in subsequent programming courses and enhanced career opportunities.

REFERENCES

Ragonis, N. & M. Ben-Ari (2005). A Long-Term Investigation of the Comprehension of OOP Concepts by Novices. *Computer Science Education*, 15 (September 2005), 203-221.

Sheetz, S., G. Irwin, D. Tegarden, J. Nelson, & D. Monarchi, D. (1997). Exploring the Difficulties of Learning Object-Oriented Techniques. *Journal of Management Information Systems*, 14 (Fall 1997) 103-131.

Thramboulidis, K., (2003). A Constructivism-Based Approach to Teach Object-Oriented Programming. *Journal of Informatics Education and Research*, 5 (Spring 2003) 1-14.