AN APPROACH FOR DETECTION OF LOGICAL INCONSISTENCY

RINKI YADAV

UNIVERSITI SAINS MALAYSIA

2012

AN APPROACH FOR DETECTION OF LOGICAL INCONSISTENCY

by

Rinki Yadav

Thesis submitted in fulfilment of the requirements

for the degree of

Master of Science

February 2012

Declaration

Name : **Rinki Yadav**

Matric No : **P-COM0012/08(R)**

School : **School of Computer Sciences**

Thesis Title : **AN APPROACH FOR DETECTION OF LOGICAL INCONSISTENCY**

I hereby declare that this thesis I have submitted to SCHOOL OF COMPUTER SCEINCES is my own work. I have stated all references used for the completion of my thesis.

I conscientiously believe and agree that the university can take disciplinary actions against me should my thesis be found to be the work or idea of other persons. I am aware that the degree awarded will be forfeited in the event of plagiarism.

Student's Signature : ……………………………………….................

Dated : …………………………………….

IC No./ Passport No : …G3254669……

Acknowledgements

Numerous people have helped me to complete this thesis. I would like to express my gratitude to those individuals for their wonderful support and assistance during the course of this work.

First and foremost, I would like to thank my supervisor Dr. Vincent Khoo, who has been a strong and supportive supervisor to me throughout my studies at University Science Malaysia, giving me great freedom to pursue independent work. He kept me focussed, carefully listened to my problems, and has been a flexible but a strong advocate to me. In reviewing my writing, he offered painstaking comments, whilst respecting my voice. What kept me moving constantly, were his amazingly insightful comments with detailed attention to my arguments, showing the ways of dramatically improving them.

A special thank goes to my husband Prabhakar, for his unconditional support, patience and love which was always there for me, and also to my son Atharv for his amazing adventures, which kept me inspired every day. I would like to thank Lim Chia Yean for being a great companion in the lab and always being ready to provide her support. Special thanks to Christine Chua for helping in editing my thesis.

Finally, in recollecting the memories of past few years, I am delighted to see the sheer scale of excitement, frustration that this research project has brought me, and hopeful that I would have an opportunity to share this wealth of experience and knowledge with others.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AHP	Analytic Hierarchy Process
CI	Consistency Index
CM	Consistency Measure
CR	Consistency Ratio
DM	Decision Maker
DR	Direct Rating
FAHP	Fuzzy Analytic Hierarchy Process
HCI	Human Computer Interface
IDOT	Inconsistency Detection On Triads
MCDM	Multi Criteria Decision Making
PA	Point Allocation
PCM	Pairwise Comparison Matrix
RI	Random Index
TCR	Triad Consistency Ratio

Suatu Pendekatan untuk Pengesanan Ketidaktekalan Logikal

ABSTRAK

Respon dalam-talian terhadap soal selidik pembuatan keputusan pelbagai kriteria sering terdapat ketidaktekalan yang mungkin disebabkan oleh interpretasi soalan yang tidak tekal atau minda responden yang sering berubah-ubah. Sekiranya ketidaktekalan respon ini tidak dikesan dan tiada tindakan pembetulan yang sesuai diambil, maka hasil soal selidik sering kali tidak tepat. Motivasi utama penyelidikan ini adalah untuk meminimumkan ketidaktekalan hasil yang tidak diingini melalui sokongan pengesanan ketidaktekalan yang berunsur dinamik. Akan tetapi, dalam kebanyakan soal selidik dalam-talian, didapati bahawa sokongan pengesanan ketidaktekalan adalah agak terlewat jika tahap ketidaktekalan ini hanya ditonjolkan pada peringkat akhir soal selidik. Ia sepatutnya ditonjolkan sebaik sahaja respon yang tidak tekal dikesani.

Dalam usaha menyelesaikan masalah yang sebegini, penyelidikan ini dijalankan untuk mengesan respon yang tidak tekal sebaik sahaja ia timbul. Sebagai tambahan, teknik perbandingan tiga serangkai berasaskan tujuh aturan logik juga dicadangkan. Dalam proses penyelidikan, didapati bahawa tatasusunan soalan yang secara rawak tidak menggalakkan pengesanan respon yang tidak tekal dengan efisien. Justeru, aturan pembentukan tiga serangkai optimum diaplikasikan. Kemudiannya, didapati bahawa pengesanan boleh dimulakan sebaik sahaja (n-1) respon daripada sejumlah n(n-1)/2 respon telah diterima. Pada (n-1) respon berikutnya, pengesanan ketidaktekalan logikal akan dilakukan secara berlanjutan.

Keputusan penilaian penyelidikan ini telah menunjukkan bahawa pendekatan untuk pengesanan ketidaktekalan logikal dengan pembentukan tiga serangkai

(IDOT) adalah betul and berkesan. Di samping itu, keputusan penilaian juga menunjukkan bahawa penggunaan semua tujuh aturan logik dalam pengesanan ketidaktekalan logikal adalah wajib and diperlukan. Selain itu, perbandingan antara IDOT dengan pendekatan untuk pengesanan ketidaktekalan yang lain juga menunjukkan keberkesanan IDOT dalam meningkatkan kadar konsistensi dengan penurunan kadar tidak tekal dari 16.4% ke 6.7%. Justeru, dipercayai bahawa IDOT untuk penyusunan khas soalan adalah diperlukan untuk mengesan respon yang tidak tekal dengan lebih baik dan lebih awal.

An Approach for Detection of Logical Inconsistency

ABSTRACT

Online responses to multi-criteria decision making questionnaires often contain inconsistencies, due to possible inconsistent interpretations of the questions or the erratic mental states of the respondents. If such inconsistent responses remain undetected and no appropriate corrective actions are taken, the questionnaire outcomes will be less reliable. The main research motivation was to minimize such inconsistent responses by detecting the logically inconsistent responses dynamically. The conventional approaches compute a consistency ratio to quantitatively specify the level of inconsistent responses. However, in most online questionnaires, it is too late to highlight the inconsistency level at the end. They should be flagged out as soon as each inconsistent response surfaces.

In order to resolve such a problem, this research has been conducted to detect logically inconsistent responses as soon as they arise. Furthermore, a triad comparison technique based on seven logical rules is proposed. In the course of the research, it was found that random arrays of questions did not promote efficient detection of inconsistent responses. As such, the optimal triad formation rules are applied. It was then found that detection could start as early as after receiving (n-1) responses, among a total of $n(n-1)/2$ responses. Subsequent to the $(n-1)^{th}$ response, logical inconsistency detection would then be performed.

The evaluation results show the correctness and effectiveness of the proposed approach for inconsistency detection on triads (IDOT). The results further show the necessity of all seven rules in logical inconsistency detection. The comparison of IDOT with other conventional detection approaches shows the effectiveness of IDOT in terms of improved consistency ratio, which is reduced to 6.7% from 16.4%. It is believed that IDOT for specially sequenced questions is necessary to better and early detect inconsistent responses.

CHAPTER 1

INTRODUCTION

1.1 Research Overview

As more and more people use the internet for business transactions, for buying and marketing products online, online rating systems play a significant role in making decisions. These rating systems, often called "reputation systems" or "recommendation systems" are used for rating a product or a service. Thus, these systems allow users to know the opinion of others and let others know their views on the various products or services. This can help people in making decisions about what to buy and which product to trust. For example, eBay.com users can read the feedbacks given by others before making their own decisions. Other sites such as Amazon.com allow users to express their opinion on products and allow others to respond to those reviews. There are other recommendation systems whereby users make comparative judgments before viewing the opinions of others on the product. In all type of recommendation systems, to express the views or to know others, there are set of questions that need to be answered. These online questionnaires can either be unstructured or structured (Tovey, 2003).

For unstructured questionnaires, the respondents provide ratings through Direct Rating (DR) or Point Allocation (PA). However, these rating methods have their own drawbacks as each respondent tends to give more weights when using PA than when using DR. The reverse is also true for the least important attribute(s). In contrast, structured questionnaires enable users to compare all the attributes and better express their opinions. They have been found to be better than the DR/PA method (Tovey, 2003). There are various popular tools available today for helping users in decision making using structured questionnaires like Expert Choice, Make It Rational, Decision Labs, etc. These tools make use of comparative judgment in decision making. A widely used technique is the Analytic Hierarchy Process (AHP) (Saaty, 1980). These AHP tools decide on the ranking/selection of choices based on the responses to the questionnaires. When using these tools, the fatigue factor needs to be considered if the number of attributes is large. For *n* attributes, direct rating may require just *n* comparisons whereas for comparative judgment, ${}_{n}C_{2}$ comparisons need to be done. There are other limitations that lead to inconsistency. The presence of inconsistency in comparison judgments will lead to incorrect priority order with significant frequency (Lipovetsky and Conklin, 2002). Even after the existence of AHP for the last three decades and extensive research on the issue of inconsistency, there are still open questions in inconsistency detection. For example, should the inconsistency detection rule depend on the number of criteria? What should be the criteria to declare the responses as inconsistent? (Ishizaka and Labib, 2011)

In this research, some of the open issues highlighted in (Ishizaka and Labis, 2011) have been addressed. An approach is proposed to help in detecting inconsistency as early as possible. This early detection would remove the need to re-answer the questionnaires as required in conventional techniques if the responses are found to be inconsistent. Comparative questions will be reviewed and rephrased to make them more easily understandable. The rephrased questionnaires would aid in detecting logically inconsistent responses at two stages. Respondents with better understanding of the questions can better externalize their thoughts, thus minimizing the rate of inconsistency. This research work has been targeted on online questionnaires.

1.2 Research Background

Inconsistency is a major issue in prioritizing selection criteria. A simple description of inconsistency is giving contradictory answers during the comparison of two criteria at one time. This type of comparison is done in structured questionnaires, where different criteria are compared among themselves. The inconsistency may arise either due to wrong selection of criteria when comparing the relative importance of criteria or during the comparison of degree of importance of criteria. For example, if 'criterion C_1 is more important than criterion C_2' and 'C_2 is more important than C_3', then 'C_1 should be more important than C_3'. If it is 'C_3 should be more important than C_1' instead, then it is considered as inconsistent. Similarly, inconsistency may arise when criteria are compared numerically. For e.g., suppose C_1 is 2 times more important than C_2, C_2 is 5 times more important than C_3, then C_1 should be 10 times more important than C_3. However, the scale chosen for such comparison allows selection of numerical values among 1 to 9 only. So if a respondent chooses 9 instead of 10, it leads to inconsistency. Such inconsistencies lead to the wrong selection of alternatives.

In the conventional approaches, inconsistency is only measured quantitatively based on the value of the Consistency Ratio (CR) (Saaty, 1980), which is computed after taking into account all the responses. Researchers have tried to solve the inconsistency issue by attempting different detection methods. In conventional method, respondent has to re-answer the questionnaire till the consistency is achieved in terms of CR. In other consistency improvement techniques, sometimes all the responses are altered without the respondents' knowledge or sometimes responses are changed without knowing the exact responses that led to inconsistency. Most of the methods changed the responses on the scale to make the CR within a prescribed threshold of 10% (Saaty, 1980). Little work has been done for detection of logical inconsistency in the responses.

An approach has been proposed for early detection of logical inconsistencies in the responses, which would help respondents make consistent logical responses for the entire set of questions.

1.3 Motivation

This research seeks to detect logical inconsistencies in online questionnaires for prioritizing criteria. Inconsistency is mainly caused by variations in the mental state of the respondents and it often leads to the undesired selection of alternatives. The motivation of this research is to find a mechanism that can detect logical inconsistencies separately, involving minimum changes to responses, so that the system is consistent. Even if the response needs to be changed, the changes made should be known to the respondent.

1.4 Problem Statement

Contradictory responses may be due to respondents not being familiar with the process, fatigue from comparing a large number of criteria, or their mental inability in responding to the subject. The inconsistency in comparative responses produces incorrect selection of alternatives (Lipovetski and Conklin, 2002). The pairwise matrix fails to achieve the required consistency ratio in many cases. Therefore, an approach that helps detect logical inconsistencies but does not depend on consistency ratio should be designed. Inconsistency in responses must be detected dynamically to help respondents make consistent judgments in all the questions, before computing the weights to prioritize the criteria. This research will focus on answering the following question: How to early detect logical inconsistent responses in online questionnaires?

1.5 Research Questions

a. How to detect logical inconsistent responses in online questionnaires?
b. How to detect logical inconsistent online responses as early as possible?
c. How early can inconsistency detection start?
d. What rules could be used to detect logical inconsistent online responses and how to apply them against the responses?
e. How to generate the sequence of questions to facilitate dynamic inconsistency detection?

1.6 Research Objectives

The research objectives of the proposed work are:

1. To enable the users to respond more consistently to online questions by detecting logically inconsistent response dynamically.

2. To start the detection after getting initial (n-1) responses.

1.7 Research Scope

This research is about the detection of inconsistent responses in the relative importance between two criteria. In this research, a typical question is split into two parts. Part (a) indicates the relative importance, while part (b) indicates the degree of relative importance between two criteria.

The adopted rule-based approach is applied to the answers among three (3) questions which contain loosely-coupled parts (a) and parts (b). However, in the evaluation of the proposed approach, sample answers are generated for triads containing tightly-coupled parts (a) and parts (b).

The focus of this research is only on the detection of logical consistency in the relative importance between two criteria, and not on the detection of inconsistency in the degree of relative importance between two criteria.

1.8 Contributions

The main contributions of the research are as follows:

a) An approach for early detection of inconsistency after (n-1) responses dynamically.

b) A set of rules for logical inconsistency detection of relative importance between two criteria.

c) A technique for the sequencing of questions based on optimal triad formation rules and identification of minimum number of triads for inconsistency detection.

Other than the above contributions, a sub contribution of this research is a basic display approach for verification before validation of logical inconsistency in the degree of relative importance between two criteria.

1.9 Limitations

This research is based on the mathematical foundation of AHP. This logical inconsistency detection approach assumes that the initially answered (n-1) questions are consistent. However, if they are not, this might increase the level of inconsistency instead.

1.10 Thesis Organizations

The layout of the thesis is organized as shown in Figure 1.1. Chapter 1 briefly outlines the research overview, background of the criteria prioritization problem, motivations, problem statement, objective of the research, and contributions. Chapter 2 covers the literature review of the research problem background. Chapter 3 introduces the research methodology. Chapter 4 provides a detailed explanation of the proposed approach on early detection of logical inconsistency. Chapter 5 presents the evaluation and results of the early inconsistency detection on triads. Finally, a summary of the thesis and the main conclusions of the study are presented in Chapter 6, along with recommendations for future work.

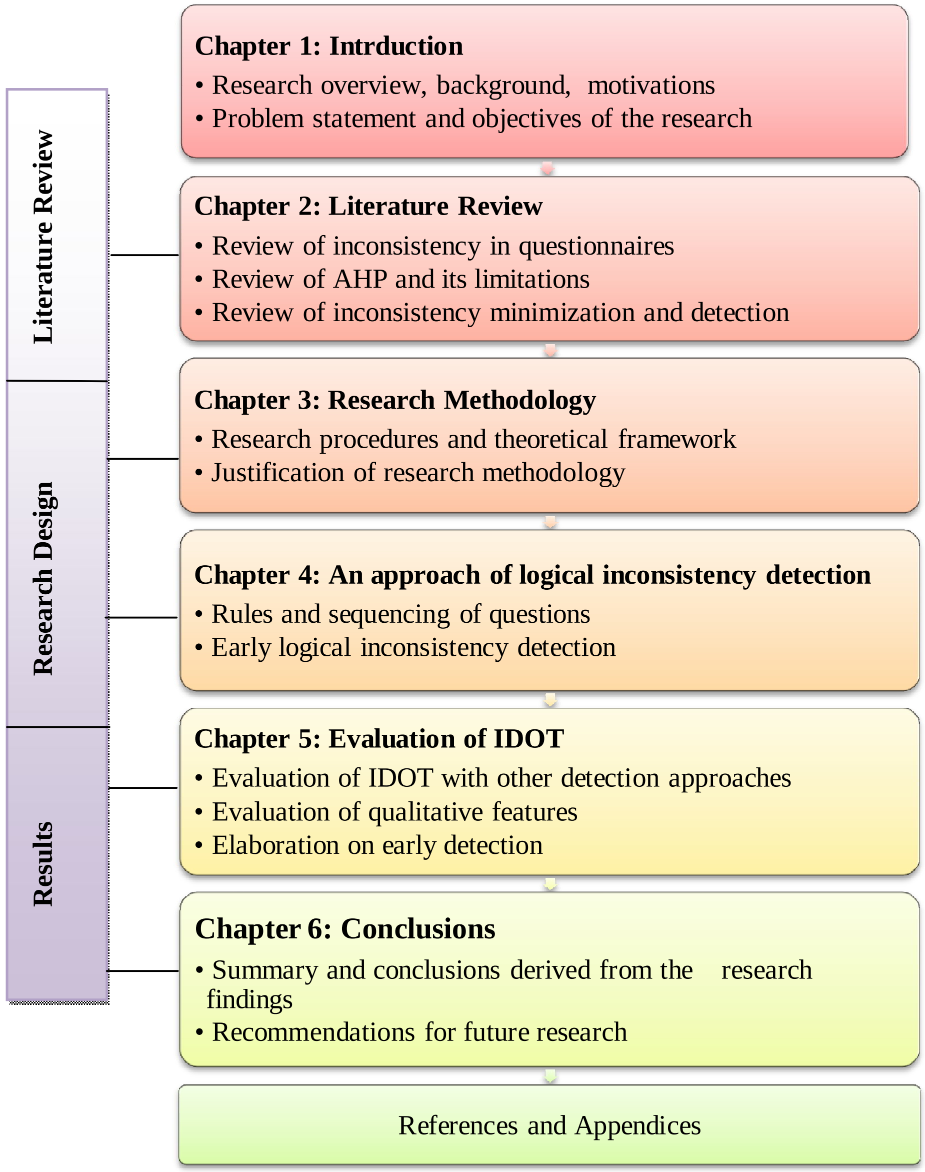

Figure 1.1: Layout of the thesis

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Detection of logical inconsistency in online rating systems is the main motivation of the work. Therefore, in Section 2.2, the review starts with different types of online rating mechanisms for multi-criteria attribute evaluation and their inherent drawbacks. Section 2.3 discusses the different techniques of multi-criteria decision making (MCDM), followed by a thorough review of the Analytic Hierarchy Process (AHP) and its shortcomings in Section 2.4.

In Section 2.5, inconsistency as a major issue in the AHP process (Bozoki and Rapcsak, 2008, Lamata and Pelaez, 2002) and its detection as an important aspect of getting consistent responses is discussed. Section 2.6 analyzes the role of Likert scale in achieving consistency and conclusions are drawn based on the study. The known types of inconsistencies in pairwise comparisons are classified in Section 2.7, while Section 2.8 discusses in detail the computation process of Consistency Ratio (CR) and how it has been interpreted by other researchers. The main drawbacks of CR computation are also reviewed in this section. Section 2.9 discusses qualitative inconsistency and shows the various types that might exist. The various consistency improvement approaches, which include optimization and direct adjustment techniques, are reviewed in Section 2.10. In Section 2.11, various techniques that can be used for evaluation of prioritization techniques are described. Finally, Section 2.12 summarizes the research work reviewed. Table 2.1 shows the flow of the work reviewed in the chapter.

Table 2.1: Review and analysis of the research problem

Features	Past Research
Environment	Manual and online questionnaire
Mode of detection	Detection at the end of a questionnaire
Question layout	Single question layout
Basis of detection	Matrices for all responses
Nature of detection	Quantitative and qualitative detection

2.2 Inconsistency in Online Questionnaires

Decision making is a complex process for managing an organization's knowledge assets. Decision making makes use of the knowledge and experience of an individual in the organization. This is especially true when decision making has to be done on multiple criteria, requiring individuals to rank a set of alternatives according to their relative importance. There are different mechanisms to rank a set of alternatives like Direct Rating (DR) and Point Allocation (PA). These two weight elicitation methods are very popular but have their own respective drawbacks. When using DR, a decision maker tends to give more weights to the more important attributes and less weights to the less important attributes (Shirland et al., 2003). This tends to produce weights that are linear when sorted by size. While using PA, they tend to produce non-linear weights (Bottomlay et al., 2000). Although many attempts have been made to identify a reliable weight elicitation method, research indicates that there is little consistency in their results (Eckenrode, 1965, Johanna and Koele, 1995, Doyle et al., 1997, Poyhonen and Hamalainen, 2001). These unstructured methods are popular because they are easy to use in multi-attribute weighting, but the low consistency rate in their results show that they are not reliable for multi-attribute weight elicitation. Researchers have investigated a number of novel decision making procedures in generating weight attributes like regression analysis, multidimensional scaling (Caroll, 1972), Logit (Chapman and Staelin, 1982), and Analytic Hierarchy Process (AHP) (Armacost and Hosseini, 1994). Other linear programming techniques used to identify group conflicting preferences have been developed by (Choi, 2001), while (Jian, 1999) describes an integrated approach that combines subjective and objective inputs to generate rankings.

Several studies investigate the effect of comparative judgments versus DR or PA systems. They found that comparative judgment is more important than DR under certain circumstances (Tovey, 2003). It appears that direct evaluations based on self-judgment lead to systematic "overweighting" of unimportant attributes and "underweighting" of important ones (Butler et al., 2000). This comparative judgment comes under structured type rating systems. The more widely used decision making tools have structured questionnaires. They are discussed in detail in the following sections.

2.3 Multi-Criteria Decision Making (MCDM) and its Tools

The existing decision making methods can either be considered as single objective or multiple objectives. In the most practical situation, the problem is to deal with multiple criteria for making prioritization and selection. MCDM has long been used for prioritization of criteria. These methods are classified under different groups by different authors (Pomerol and Barba-Romero, 2000). MCDM methods can be broadly classified under four categories:

- Ordinal methods
- Weighting methods
- Outranking methods
- Additive utility based methods

The existing decision making methods are shown in Figure 2.1. Ordinal methods derive ranking for the set of alternatives by aggregating the individual pre-orders with respect to the criteria, while in the weighting method, the final outcome largely depends on weights assigned to the criteria. Although results obtained by these methods are not reliable, they are simple to apply and very popular among Decision Makers (DM) in the real world (Jenssen, 2001).

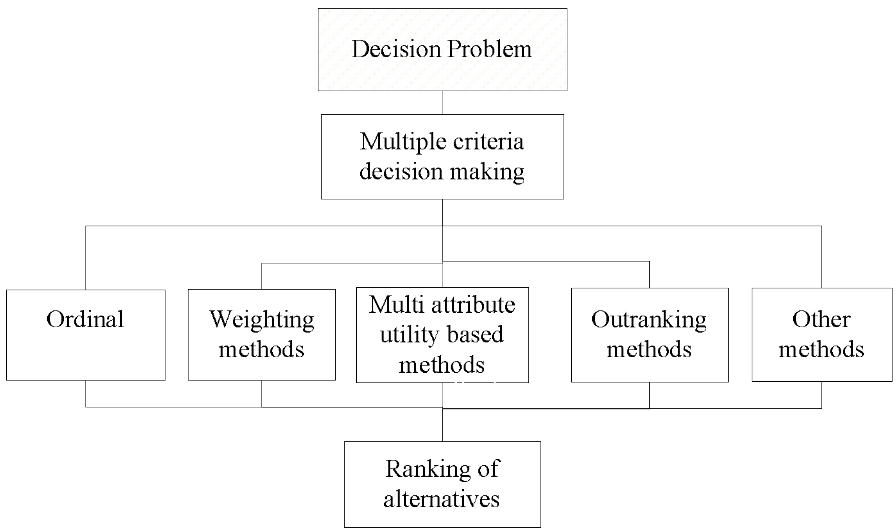

Figure 2.1: Decision making methods for criteria selection and prioritization

The outranking methods differ among other aspects, in the way that each method formalizes the above concepts. Two popular methods used are Elimination Et Choix Traduisant la REalite' (ELECTRE) (Roy, 1968, Roy, 1996) and Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) (Jenssen, 2001). These methods involve pairwise comparison of criteria and aggregation of preferences to each criterion.

The additive utility-based methods deal with the utility of the criteria to evaluate the alternatives. These utility functions put together multiple criteria into a single criterion by making use of the subjective information provided on a set of actions and then using multi criteria evaluations of these actions. The main techniques under additive utility-based methods are Multi Attribute Utility Theory (MAUT) (Keeny and Raiffa, 1976), Simple Multi-Attribute Rating Technique (SMART) (Von and Edwards, 1986), Utility Theory Additive (UTA) (Jacquet and Siskos, 1982), and Analytic Hierarchy Process (AHP) (Saaty, 1980). The major advantage obtained using these utility functions is that they can be used for probabilistic outcomes (Zoints, 1990). The process of using these functions is considered long and difficult, and questions that DM has to answer are not easy to understand, which leads to inconsistency in responses. However, the results provided by these utility functions are reliable for complete ranking of the alternatives, though it is difficult to implement compared to the outranking methods (Pomerol and Barba-Romero, 2000).

2.4 Analytic Hierarchy Process (AHP)

AHP process supports a strong theoretical interpretation based on the theory of graphs (Harker and Vargas, 1987) and a hierarchical model which is a central part of this methodology (Ho, 2008). AHP is based on setting up the hierarchy of criteria when structuring the problem and effectively tackles weight evaluation, a major issue of MCDM. AHP is widely used as a tool in various application areas for selection and ranking (Sipahi and Timor, 2010, Vidal et al., 2010, Wang et al., 2010)

2.4.1 AHP Model

AHP is a systematic procedure for representing the criteria of any problem hierarchically. It organizes the basic rationality by breaking down a complex problem into smaller constituent parts and then guides DMs through a series of pairwise comparisons of judgments to express the relative importance of criteria in the hierarchy. AHP uses these judgments to derive the priorities and find alternative solutions. The process consists of the following steps:

1. Constructing a decision hierarchy by breaking down the decision problem.
2. Performing pairwise comparisons of the decision criteria.
3. Estimating the weights of the decision criteria.
4. Aggregating the relative weights of the decision criteria to provide a priority list for the decision elements.

In the hierarchical representation, each node represents the main criteria which can have sub-criteria in the lower immediate nodes to be prioritized. Each relationship is weighted according to the influence strength of an alternative or criterion at the same level. The elements at one level are influenced by elements of the node just above it. This influence is distributed from the top with the main objective having a value of one. This value is further divided among other elements of the nodes until the last level in the hierarchy. The degree of influence is measured on a nine-point scale and the final outcome is in terms of weights assigned to each of the alternative. The 1 to 9 scale is used as shown in Table 2.2.

Table 2.2: AHP 9-point ratio scale

Intensity of relative importance	**Definition**
1	Equally important
3	Moderately important
5	Strongly important
7	Very Strongly Important
9	Absolutely Important
2,4,6,8	Intermediate values between the two adjacent judgments
Reciprocal of above non-zero values	If a criterion has one of the above values (e.g. 3) compared with a second criterion, then the second criterion has the reciprocal value (1/3) when compared to the first.

The second step involves pairwise comparison of the decision elements. Each criterion is compared with other criteria and the result of comparison is placed in a pairwise comparison matrix. Pairwise comparisons are the fundamental step in the evaluation of weights. In the next step, the Eigenvalue method is used to estimate the relative weights of each alternative. The consistency of pairwise evaluations is checked, and if the judgments are not acceptably consistent, the respondents are asked to revise the judgments and redo the questionnaire.

In the last step of AHP, the relative weights of various levels are aggregated. The output is composite weights, from which ranking of the alternatives is done. The method also accepts a certain degree of inconsistency and acceptable level of inconsistency is defined in terms of a ratio. The complete method is available in the user-friendly software package called Expert Choice, which is widely used in decision making. Different commercial tools available for decision making are shown in Table 2.3.

Table 2.3: Tools for decision making

Name	**Technique Used**
Expert Choice	AHP
Make It Rational	AHP
Decision Lens	AHP
Decision Lab 2000	PROMETHEE
Hiview and Equity	Weighted sum
Criterium Decision Plus	AHP + SMART
HIPER 3+	AHP
Logical Decision	AHP + Smarter + Tradeoff
Super Decisions	AHP + ANP
Telelogic Focal Point	AHP

The main characteristics of the tools are briefed in Table 2.4. The criteria for comparison are algorithm used, application area, platform used, and year of development.

Table 2.4: Comparison of different tools available for decision making

Tool	Algorithms	Application area	Platform	Year
Make It Rational	AHP	Business management, Service industry	Web-based and tool-based	2010
Telelogic Focalpoint	AHP	Portfolio management	Web-based	2008
Decision lens	AHP and ANP	Marketing, Finance, Performance strategy	Web-based and tool-based	2007
Making Decisions in Integration of Automotive Software and Electronics	ATAM and AHP	Automotive software	Only method or technique proposed	2007
Super Decisions	ANP	Marketing, Medical, Political, Social	Tool-based	2000
Logical Decisions	AHP, Smarter, tradeoff and other algorithms	Business evaluation, weapon evaluation, airplane systems	Tool-based	1991

2.4.2 AHP Questionnaire

One of the primary inputs after the criteria have been identified for the AHP model is the response to a comparative judgment questionnaire. In this questionnaire, strengths of criteria are compared among themselves on the scale. A sample questionnaire layout is shown in Figure 2.2. In this layout, the respondent has to mark his degree of preference on either sides of the scale while comparing two criteria. The criteria appear randomly on both sides of the linguistic scale as shown in Table 2.2 for comparison. For an uninformed DM, this random appearance of criteria may appear confusing and lead to incorrect choices. The DM compares all the criteria with respect to each other using the 9-point scale.

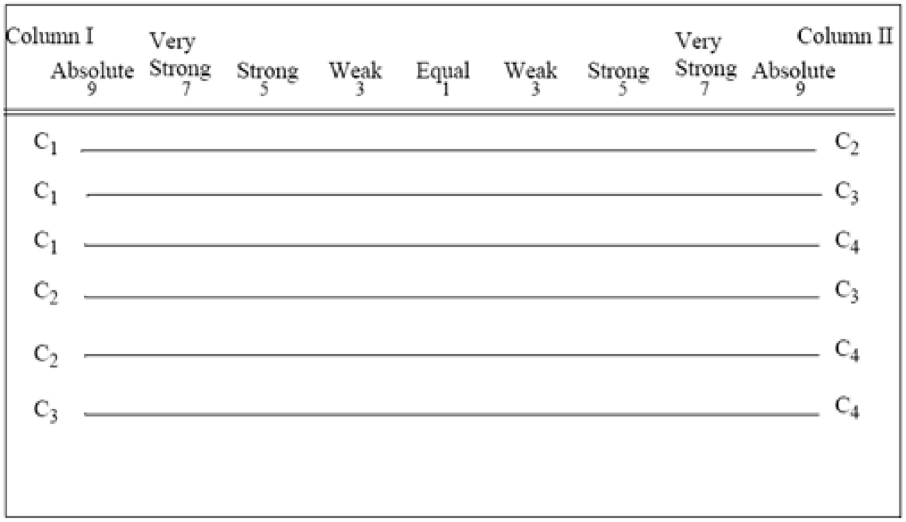

Figure 2.2: Saaty's original questionnaire layout (Saaty, 1980)

This method is good when the respondent is certain about his choices and for comparison of strength. However, for situations whereby the respondent is not certain, for example, when he does not have any experience on the subject, AHP fails to deliver. In spite of its popularity and availability of the tool for last 30 years, this method is often criticized for not being able to map the decision makers perception to a number (Chen, 2009).

The conventional AHP questionnaire is not well suited for uninformed respondent. (Temesi, 2010) discussed the relationship between the consistency of pairwise comparison matrix and the consistency of the DM. In multi-criteria decision making problems using pairwise comparison matrices, it is crucial to distinguish the type of DM. Informed and uninformed DMs should be tackled differently to get their real preferences and obtain an error free pairwise comparison matrix. Interactive questioning procedures are recommended to reach that goal.

AHP is well suited for some of the problems since the approach is qualitative and easier to implement. It is easier to validate the output of AHP application. Since

AHP has an inherent issue of inconsistency, it cannot be directly applied to solve decision making problems. In order to eliminate the drawbacks of AHP, this issue of inconsistency in responses has to be minimized.

2.4.3 Limitations of AHP

In AHP, the elements of a problem are compared in pairs with respect to their importance ("weight" or "intensity"). When elements in a set of a problem are compared with each other, a square matrix is produced. Elements of a lower triangle of the matrix are the inverse of elements of an upper triangle i.e. $a_{ij} = 1/a_{ji}$.

a) The need for a Scale of Comparison

It is often observed that there exists a scale for each underlying problem. In this case, the comparison judgments are expressed as ratios on a scale. For example, if 'C_1 is 5 times more important than C_2' and 'C_1 is 1/2 times less important than C_3', then 'C_2 should be 5/2 times more important than C_3'. However, the value 2.5 is not on the ratio scale of 1-9. Some approximation is needed, leading to some additions in inconsistency. To resolve this problem, several scales have been proposed by different researchers but Saaty's 1-9 linear scale (Saaty, 1980) remains the most popular. Other scales like (Harker and Vargas, 1987) evaluated a quadratic and a root square scale but were in favour of Saaty's 1–9 scale. However, they have the view that Saaty's scale can be altered to match the needs of individuals. They could not prove the superiority of their case by one simple example. (Lootsma, 1993) proposes a scale based on geometric mean value and claimed it to be better than the 1-9 linear scale. (Salo and Hamalainen, 1997) pointed out that the 1-9 scale would give local weights that are not evenly distributed, causing insensitivity especially when the choices have very much less differences. They proposed a scale whereby local weights are more evenly spread out over the range of [0.1, 0.9]. (Ma and Zheng, 1991) proposed another scale whereby the inverse elements *x* of the scale 1/*x* are linear instead of the *x* in the Saaty scale. (Donegan et al., 1992) proposed an asymptotic scale that avoids the boundary problem. The possibility of integrating negative values into the scale has also been explored (Millet and Schoner, 2005). However, in the end, Saaty's linear scale remains the most popular and (Saaty, 1980, Saaty, 1990) advocates it as the best scale to represent weight ratios, though it has the inherent issue of adding inconsistency.

b) Definition of Consistency Ratio

It is emphasized that consistency ratio is related to Saaty's scale. The structuring process in AHP specifies that items to be compared should be within one order of magnitude (Bozóki et al., 2010). This helps to avoid inaccuracy associated with cognitive overload as well as relationships that are beyond the 1–9 scale (Murphy, 1993). The widely used definition is Saaty's consistency ratio (Tovey, 2003) and is defined as CR = CI/RI, where $CI = \frac{\lambda_{\max} - n}{n-1}$, $\lambda_{\max}$ is the largest Eigenvalue of the matrix , CI is consistency index and RI is the random index. CI can be used to measure inconsistency only when it is benchmarked to determine the magnitude of the deviation from consistency. RI is a ratio obtained by generating the matrix of order *n*. 1000 random matrices on the scale of 1-9 are randomly generated for order of matrix 3 to 10 and then CI is calculated for each case. RI is defined as the average of CI for each order. RI is taken from the Table 2.5.

Table 2.5: Random consistency index adapted from (Saaty, 1980)

n	3	4	5	6	7	8	9	10
RI	0.58	0.90	0.12	0.24	0.32	0.41	0.45	0.49

Based on the value of CR, the condition can be classified as consistent or inconsistent. Using Saaty's original "rule of thumb", the pairwise comparison matrix is deemed to be inconsistent only if CR > 0.10. If only two criteria (or alternatives) are present, inconsistency is always zero, since the DM gives only one importance ratio. The major drawback of Saaty's inconsistency definition seems to be the 10% rule of thumb for any order of matrix and this definition is the most widely accepted rule (Bozóki et al., 2010). Its consistency definition has some drawbacks. (Koczkodaj, 1993) believes that this 10% definition of consistency ratio has been a major weakness of AHP. Another weakness is related to the location of inconsistency. Since an Eigenvalue is a global characteristic of a matrix, by examining it, we cannot say which response has contributed to the increase of inconsistency (Bozóki et al., 2010). However, Gower's plot (Li and Ma, 2007) is able to pinpoint the outliers in the responses which have contributed to the inconsistency.

Though consistency ratio provides the measure of inconsistency but its definition which relates 10% rule doesn't seem to be appropriate and does not work well for all order of matrices. There should be different threshold for different order of matrix or altogether new threshold should be defined which correctly defines the inconsistency for different order.

2.5 Inconsistency Detection

In real-life decision problems, pairwise comparisons are rarely consistent (Saaty, 1994, Bozoki and Rapcsak, 2008, Keri, 2010, Temesi, 2006). Nevertheless, a DM needs to maintain a level of consistency in the judgments. Inconsistent judgments may lead to senseless decisions. Judgment consistency in the pairwise methods is measured using the consistency ratio (Tovey, 2003). Inconsistency measures have been proposed by various researchers (Aguaron and Moreno-Jimenez, 2003, Bozoki et al., 2011, Saaty, 1980, Salo and Hamalainen, 1997) but the major problem is the interpretation of these definitions. Saaty's 10% rule of thumb has been widely accepted as a measure of inconsistency. (Salo and Hamalainen, 1997) defines their consistency measure in the closed interval [0, 1], with an increase in values indicating a decrease inconsistency. The measure in (Aguaron and Moreno-Jimenez, 2003) is Geometric Consistency Index, rather than consistency index and is applied in situation where row averaging method is applied. Saaty applied Eigenvector method to compute the CI. Except at the end points of the interval [0, 1], the measure definition is not clearly interpretable. Furthermore, because the thresholds associated with these measures are based on "rules of thumb" and/or randomly generated matrices in some situations, these measures do not appear to be appropriate. For example, in a study that applied AHP to elicit subjective probabilities from human experts, (Monti and Carenini, 2000) highlights that the manifest inconsistency showed by the expert's assessments based on deferent elicitation techniques provided us with the evidence that the 0.10 value for CR was not appropriate.

While techniques that employ pairwise comparisons and use ratio scale to map the human preferences to a number have several advantages over PA and DR, they also have two major shortcomings. First, as the number of criteria increases in pairwise comparisons, it starts to produce in conflicting choices and lack of transitivity (Flynn et al., 1990). Second, defining inconsistency through an inconsistency index is not a sufficient criterion for describing inconsistency (Bozóki et al., 2010) but there is a need for an alternate mechanism to address the other type of inconsistency that exists in it. To better understand the limitations of conventional inconsistency detection approach, an analysis is conducted and described in the next section.

2.6 Analysis of Conventional Inconsistency Detection

Saaty's conventional approach fails to achieve consistency in term of CR below 10% and is mainly due to two reasons.

- Limitation of Likert scale.
- Wrong selection of degree of relative importance.

Research has been done on the limitations of the Likert scale by defining different scales of degree of relative importance or by introducing fuzzy logic to better capture the mind of DM. The focus of this analysis is to minimize the impact of wrong selection of degree and then observe the impact of Likert scale on inconsistency detection. To minimize the wrong selection of degree of relative importance between two criteria, a method of determining missing responses (Harkar, 1987) is applied and pairwise matrix is filled completely. CR is computed for n = 2 to 15, where n is the number of criteria. The following sub-sections describe the geometric mean method for determining the missing responses and analyze the impact of Likert scale on inconsistency detection.

2.6.1 Geometric Mean Method

One of the methods in estimating the missing comparisons is the Geometric Mean method (Harker, 1987). Let a_{ij} be introduced to denote the missing comparison value in the i^{th} row and j^{th} column. For a perfectly consistent case $a_{ij} = a_{ik} . a_{kj}$, where a_{ik} and a_{kj} are known initial comparisons.

The above formula can only be true if the matrix is perfectly consistent. The combination of a_{ik} and a_{kj}, is called an elementary path (of length 2) connecting the missing comparison of items i and j (Harker, 1987). It is important that such connecting paths comprise a pair of known comparison values. If one of the elements in the pair is missing, the Geometric Mean method cannot be applied.

The above formula will accurately calculate the comparison value in a perfectly consistent matrix. In an inconsistent matrix, we should consider calculating a_{ij} using more than one elementary path. The formulation involves multiplying all possible elementary paths between i and j, followed by taking the q^{th} root (where q is the number of all possible paths). Note that an elementary path does not always consist of two elements. In a matrix of size n the number of elements in an elementary path can be from 2 to (n −1). Thus, formula can be extended to include these additional elements. CP_r is a connecting path with r +1 elements. The parameter r, called the connecting path index, will define the number of elements in the connecting path (Harker, 1987).

$$CP_r : a_{ij} = a_{ik1}*a_{k1k2}*a_{k2k3}*a_{k3k4}\ldots\ldots a_{krj}$$

The following formula provides the general geometric mean estimation:

$$a_{ij} = \sqrt[q]{\prod_{r=1}^{q} CP_r}$$

where CP_r is a connecting path with r +1 elements, r is the connecting path index, and q is the number of all possible connecting paths for $1 \leq r \leq n - 2$.

2.6.2 Observations and Conclusions

100 random questionnaires were generated in Matlab for n = 2 to 15 and responses were generated using the geometric mean method. CR is computed and shown in Table 2.6 for each case.

Table 2.6: Results of quantitative analysis

No.	Number of Criteria	Consistency Ratio (%)
1	2	0.00
2	3	0.01
3	4	0.94
4	5	2.45
5	6	3.12
6	7	3.97
7	8	4.91
8	9	5.31
9	10	6.76
10	11	7.16
11	12	8.99
12	13	9.78
13	14	10.25
14	15	11.45

The conclusions drawn from analysis of conventional inconsistency detection approach:

1. It is not possible to generate matrices with CR less than 10% as value of n increases. This situation arises due to limitations of Likert scale.
2. For situations, CR<10%, but responses to the questionnaire are inconsistent, there is a need to find an alternate way to detect logical inconsistent responses. This type of inconsistency occurs because of wrong selection of criteria.

2.7 Classification of Inconsistency among the Criteria

Based on different works of researchers in inconsistency detection, it can be broadly classified as:

1. Quantitative inconsistency
2. Qualitative inconsistency

A quantitative inconsistency can further be classified as:

a. **Moderate Quantitative Inconsistency**: It is defined as one in which the overall CR<10%.
b. **Strong Quantitative Inconsistency**: It is defined as one in which the property $a_{ij}.a_{jk} = a_{ik}$ is not met while responding to the questions on scale.

A qualitative inconsistency can similarly be classified as:

a. **Moderate Qualitative Inconsistency**: It is defined if $a_{ij} > 1$ which implies $a_{ik} > a_{jk}$ for any $i, j, k \in \{1,2,3...n\}$.
b. **Strong Qualitative Inconsistency**: It is defined if $a_{ij} \geq 1$ which implies $a_{ik} \geq a_{jk}$ for any $i, j, k \in \{1,2,3...n\}$.

Figure 2.3 shows the classification of inconsistencies.

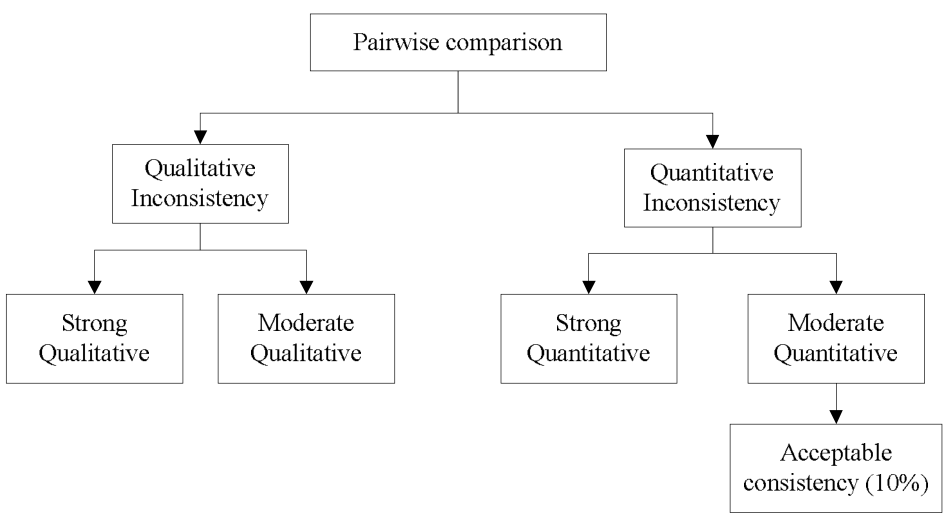

Figure 2.3: Classification of inconsistency