

인터넷 설문조사의 방법론적인 문제점과 데이터마이닝 기법을 활용한 개인화된 인터넷설문조사 시스템의 구축

김광용^{*†}, 김기수^{**}

^{*}충실대학교 경상대학 경영학부, ^{**}KT&G

Methodological Issues in Internet Survey and Development of Personalized Internet Survey System Using Data Mining Techniques

Gim, Gwangyong^{*†}, Kim, Gisoo^{**}

^{*}Division of Business Administration, Soongsil University, ^{**}KT&G

Key Word : 인터넷 설문조사, 설문 품질, 사용자 만족, 사례기반추론, 웹 설문 디자인

ABSTRACT

The purpose of this research is to summarize the methodological issues in internet survey and to suggest personalized internet survey system using data mining technique for enhancing the survey quality of internet survey as well as utilizing the benefit of interactive multimedia factors of internet survey. The data mining technique used in this paper is Case Based Reasoning for adopting individual design preference affecting survey quality.

For achieving the research purpose, two surveys, pre & post survey, were performed. Pre survey was done for implementing CBR database to find individual index affecting survey quality and post survey was used for measuring the performance of personalized internet survey system. The result shows that the survey quality of personalized web survey system is better than generalized web survey system.

1. 서론

이 논문은 1999년도 한국학술진흥재단의 연구비에 의하여 지원되었음(KRF-1999-041-C00338)

† 교신저자 gygim@ssu.ac.kr

최근 인터넷 사용자의 급속한 증가와 더불어 새로운 리서치도구로서의 인터넷 설문

조사가 많이 시행되고 있다. 그러나 인터넷 설문조사가 적은 비용, 편리성, 짧은 설문기간 등의 여러 장점을 갖고 있지만, 전통적인 설문조사방법과 매우 다른 매체특성을 가지고 있으며, 그 역사도 매우 짧고, 새로운 리서치 도구로서의 이론적, 방법론적 문제에 대한 연구가 부족한 상황 하에서 인터넷 설문조사를 검증 없이 활용하는 것은 많은 편이가 발생할 수 있다[4, 6, 7, 9, 21].

현재까지 인터넷 설문조사에 관한 연구는 주로 인터넷 설문조사와 기존 설문방법(주로 전화 설문이나 우편 설문)과의 비교분석에 많은 연구가 이루어져왔다[5, 11, 12, 16, 17, 19, 22]. 그러나 기존연구는 인터넷 설문조사의 여러 가지 방법론적 문제점에 관한 종합적인 정리가 안된 상황에서 단편적인 주제에 관하여 실증적, 탐색적 연구가 이루어져 왔으며, 특히 최근에는 인터넷 설문조사의 강점 중의 하나인 다양한 설문디자인이 설문 품질에 미치는 영향에 관한 연구가 시도되고 있는 상황이다. 따라서 본 연구는 인터넷 설문조사의 방법론적인 문제점들을 정리하고, 문제점 중의 하나인 무응답오류(nonresponse error)와 측정오류(measurement error)를 줄여 설문품질(survey quality)을 높이는 방법 중의 하나로 설문응답자의 설문지 디자인 선호도에 따른 개인화된 설문조사 구축 방법을 제시하고, 그 효과를 실증적으로 검증하고자 한다. 인터넷 설문 디자인에 관한 국내외의 연구에서 설문 배열, 형태, 길이, 기술적 활용, 색상, 그래픽, 동영상 등이 설문 응답률과 설문품질에 영향을 미치고 있다는 것을 실증적으로 보여주고는 있으나[4, 8, 10, 13, 15, 20, 23], 이러한 결과를 구체적으로 어떻게 활용하여 시스템화 하는가에 대한 연구는 현재까지 이루어지고 있지

않다.

본 연구는 이러한 선행연구의 실증적 결론을 기반으로 인터넷 설문조사의 가장 큰 장점으로 거론되는 설문 응답자와의 상호작용을 최대한 살린 “개인화된 인터넷 설문조사 시스템(personalized internet survey system)”을 구축하고 실증적으로 검증하였다. 인터넷 설문 응답자의 개인특성, 디자인 선호도 및 설문추적자료, 응답률 및 응답 품질을 데이터베이스화하고, 이러한 데이터베이스를 기초로 데이터마이닝 기법 중의 하나인 사례기반추론(case based reasoning) 기법을 활용하여 “개인화된 인터넷 설문조사 시스템”을 설계하고 구축하여, 그 실증적 유효성을 분석하였다. 특히 추가된 응답자의 특성이 계속 데이터베이스화 함으로써 자동적으로 설문 품질을 높혀 가는 “개인화된 인터넷 설문조사 시스템”은 기존 우편이나 전화설문에서는 구현하기 힘든 것으로서, 인터넷 설문조사의 장점을 살린 지능화된 인터넷 설문시스템(intelligent internet survey system)의 구축이라 할 수 있다.

본 연구는 제 2장에서 인터넷 설문조사의 방법론적인 문제점, 인터넷 설문디자인의 기존 연구들에 관하여 정리하였고, 제3장에서 사례기반추론 데이터베이스 구축을 위한 사전조사와 설계방법, 사후조사 등의 연구모형을 설명하였고, 제 4장에서 개인화된 인터넷 설문조사 시스템의 효과를 실증적으로 검증하였다. 마지막으로 제5장에서 본 연구가 갖는 이론적, 실증적 의의와 향후 연구과제에 대하여 토의하였다.

2. 선행연구

2.1 인터넷 설문조사의 방법론적 문제점

인터넷 설문조사는 기존의 설문지를 하이퍼텍스트(Hypertext) 형태로 사이버공간에 위치시킴으로써, 인터넷을 이용하는 사람들을 대상으로 공개되어 있는 설문조사방법을 말한다. 인터넷 설문조사는 전통적인 설문방법에 비교하여 방법론측면에서 상당히 유사하나, 인터넷이란 매체환경의 특성을 이용하여 “즉각적인 상호작용”이 가능하다는 점과 멀티미디어적인 디자인 요소가 사용된다는 데 가장 큰 차이가 있다[14].

이러한 인터넷 설문조사의 설문품질과 관련되어 발생하는 방법론적인 문제점은 일반 설문과 유사하게 설문조사 시 발생하는 오류(error)와 관련된 것으로, 대표성 오류(Coverage error), 표본추출 오류(Sampling error), 무응답 오류(Non-response error), 측정 오류(Measurement error)로 나누어 볼 수 있다[4, 7, 21].

대표성 오류(Coverage Error)는 현재 전통적 설문조사 방법과 비교하여 인터넷 설문조사에서 가장 많이 비판을 받는 것으로 연구하고자 하는 목표 모집단(Target Population)과 표본추출에 앞서 정의되는 틀 모집단(frame population)의 불일치로 인하여 나타난다. 이러한 오류는 인터넷 사용자들이 계속 증가하고 있고, 인터넷 이용 환경이 끊임없이 변하여 목표모집단의 정의가 아직까지 정착화 되어있지 않는 것이 가장 큰 이유이며, 향후 인터넷 이용자의 행태에 관한 완벽한 정리가 되기 전까지는 제한된 목표 모집단(예를 들어 인터넷 서점 이용자)하에서의 조사가 이루어질 것으로 보인다. 현재까지는 많은 연구들이 인터넷 설문

조사와 일반 설문조사의 응답자 분포 등을 비교분석하여 이러한 대표성 오류에서 벗어날 수 있는지에 관한 실증적 연구를 많이 시도하고 있으나, 향후 대표성 오류와 관련되어 수행되어야 할 인터넷 설문조사 연구 과제로는 목표모집단을 보다 포괄적으로 정의할 수 있는 방법(예를 들어 한국의 성인 남자)이나 아주 제한된 목표모집단의 정의(예를 들어 인터넷 사용자)를 할 수 있는 방법 등을 찾아내는 등 보다 체계적인 이론적, 실증적 연구가 필요하다 하겠다.

표본추출 오류(Sampling Error)는 틀 모집단의 모든 응답자가 측정되지 않는다는 오류에서 발생하는 것으로 근본적으로 현재 인터넷 설문조사가 확률 기반의 표본추출을 할 수 없다는 점(아주 제한된 범위의 웹 사용자 등은 예외)에서 인터넷 설문조사의 큰 문제점으로 지적받고 있는 분야이다. 근본적으로 대표성 오류가 해결되지 않으면 개선되기 힘든 오류로 인터넷 설문조사의 큰 표본크기가 좀더 유용한 설문 품질을 보장한다든지, 아니면 표본크기가 표본오차의 전부로 해석되는 것은 잘 못된 것이라고 할 수 있다.

무응답 오류(Non-response Error)는 설문 응답자들이 응답을 할 수 없거나 의도적으로 하지 않는 오류로 응답자와 비응답자간의 차이로 인하여 발생하는 오류로 볼 수 있다. 현재까지 이루어진 실증적 연구는 일반적으로 인터넷 설문이 우편설문보다 높은 비응답률을 보여주고 있으며, 우편설문에서 응답률을 높이기 위하여 시도되는 다양한 방법 등이 인터넷 설문에서는 아직까지 그 실증적 유효성을 보이고 있지 못하는 것으로 밝혀졌다. 또한 이러한 인터넷 설문은 낮은 응답률은 인터넷 접속 환경 등의 기술

적 이유, 설문디자인의 유저인터페이스 미숙, 높은 인터넷 접속 비용, 조직의 인터넷 사용감시, 스팸메일 등에 관한 인터넷이용자의 거부감등으로 설명되고 있다. 최근에는 인터넷 패널 등을 이용하여 무응답 오류를 줄이고자 하는 노력을 많이 하고 있지만 여전히 인터넷 패널의 설문 학습효과로 인한 설문 품질에 관한 이슈제기가 되고 있어 앞으로 많은 연구가 필요하다 하겠다. 본 연구에서는 설문 디자인의 개선을 통한 무응답률을 줄이는 방법을 고려하였다.

마지막으로 본 연구와 가장 관련이 깊은 분야인 측정 오류(Measurement Error)로 측정하고자 하는 것과 측정된 것의 차이를 말하는 것이다. 이미 전통적인 설문에서 연구된 것을 기초로 인터넷 설문에서도 같은 영향이 있는지에 관한 연구가 이루어지고 있는 분야로 인터넷 설문에서는 추가로 인터넷 이용자의 브라우저 환경, 이용자 정보 기술 관련 선호도, 하드웨어의 다양성 등을 고려하여야 한다. 특히 인터넷 설문에서는 다양한 형태의 설문지 디자인(배열, 색상, 소리, 이미지, 동영상)이 가능하여 전통적인 설문방법보다 설문디자인이 측정오류와 많이 관련되어 있다는 것을 실증적으로 보여주고 있다.

2.2 인터넷 설문지 디자인에 관한 연구

인터넷 설문은 텍스트(Text) 중심의 기존 설문과는 달리 네트워크 중심의 환경과 멀티미디어 요소를 고려한 설문이 가능하므로 이러한 새로운 컨셉이 설문지에 도입되어질 때, 설문품질과 응답률에 어떠한 영향을 주며, 영향을 준다면 어떠한 구성요인이 크게

작용할 것인가에 관련하여 최근 몇 가지 연구가 시행되고 있으나, 짧은 역사와 웹 설문조사에 대한 기존 조사전문가들의 접근상의 한계로 말미암아 실제적으로 다양한 연구는 실시되지 못하고 있는 실정이다.

이러한 인터넷 설문의 고유 특성은 <표-1>에서 보는 바와 같이, 설문길이와 구성, Respondent Friendly 설문지의 설계를 위한 최신 웹 기술의 적용, 시각적 자극요소의 이용, 응답 양식 및 카테고리의 구성 등 응답자들의 자발적인 참여를 유도하여 응답률을 높이고 설문 품질을 향상시키기 위한 연구가 진행되고 있다. 특히 Couper et al.(2001)은 설문 응답 형태에서 Pull down 방식과 Radio Button 방식을 실증적으로 비교 연구한 결과 Radio Button 방식이 Pull down 방식보다 높은 설문 완성률을 보인 것으로 나타났다. 또한 김광용 & 김기수(1999b) 연구에서 응답의 배열이 가로방식보다 세로방식에서 보다 높은 설문의 완성률과 일관성이 있는 것으로 나타났다.

3. 개인화된 인터넷 설문조사 시스템의 설계 및 구현

본 연구는 사례기반추론 방법을 이용하기 위하여 필요한 Index 변수 추출 및 데이터베이스 구축을 위하여 인터넷 설문을 이용한 1차 조사를 실시하고, 이렇게 구축된 데이터베이스를 이용하여 사례기반추론을 이용한 개인화된 설문조사 시스템을 구축한 후, 2차 조사를 통하여 구축된 개인화된 설문조사 시스템의 효과성을 검증하는 3단계로 연구를 진행하였다.

3.1. 1차 조사

3.1.1. 개인화된 설문 시스템 구축을 위한 잠정 변수들

개인화된 웹 설문을 구현하는데 있어 영향력 있는 변수를 선정하기 위해 기존 관련 연구에서 “컴퓨터 및 웹의 이용”에 차이를 나타내고 있는 개인특성 변수들을 가능한 최대로 추출하여 1차 설문조사에서 응답의 품질에 영향을 미치고 있는 변수를 분석하였다[7, 9, 21]. 또한 이미 기술한 인터넷 설

문 디자인에 관한 연구를 종합하여, 디자인 구성요소에 따라 개인별 선호와 성과에 차이가 있다는 변수들을 최대한 정리하여 개인화 변수로 이용하였다. 이를 통해 1차 사전설문에 사용되어진 전체 “개인화 변수”는 <표-2>와 같다. 이 변수는 1차 조사 후, 데이터의 질과 응답률 등 종속변수에 유의한 차이를 나타내는 변수만을 선정하여 개인화된 설문조사 시스템의 추론을 위한 사례기 반추론의 인덱스로 사용되어졌다.

<표 1> 인터넷 설문지 디자인에 관한 선행연구

구 분	연구 내 용
설문길이 와 구성	<ul style="list-style-type: none"> • 설문이 길어질수록 표본 수가 줄어듦[4]. • 속도에 최적화를 위해 Single 스크린은 부적당[4] • 스크린 당 1 Page를 선호함[2, 23] • 모니터해상도에 따라 응답률이 다름[4]
신기술의 적 용	<ul style="list-style-type: none"> • Adaptive Questioning & User Tracking 기능활용은 응답자의 인지적 부담을 경감시킴[8, 13] • Plain과 Fancy 웹 설문의 응답률/작성률의 비교결과 차이 없음[10]. • Pop-Up Survey는 응답률을 높일 수 있음[15]. • Helper Applications & Plug-ins 활용가능[8]
시각적 요소들	<ul style="list-style-type: none"> • 가상현실의, 적용가능성[14] • 멀티미디어의 이용가능성[13] • Fancy버전의 웹 설문에 Brightness color, Using Table, Animation의 사용결과 일반설문이 보다 우수함[10] • Logo사용은 응답률에 부정적이나 작성률에는 긍정적임[23] • 그래픽의 사용은 응답률을 증가시킴[20] • 설문응답의 진척도를 제시할 것[8]
응답형태	<ul style="list-style-type: none"> • Open-Ended Question 사용의 신중한 고려[10] • Question Box(Drop down, Radio Button)소개[2, 15] • Survey type form 소개[10, 13] • 응답의 가로 및 세로 배열형태의 차이[1]
개인화	<ul style="list-style-type: none"> • Personalized E-Mail Invitation[18] • Personalize Survey는 Security가 중요[15] • Respondent Friendly Design의 한 방법임[10]

<표 2> 개인화된 설문조사 시스템의 추론을 위해 선정된 잠정 변수들

	잠정적으로 선정된 변수들
웹 설문지 디자인 요소	<ul style="list-style-type: none"> ▪ Interface Mode(Pull-down, Radio) ▪ 응답의 배열형태(수평, 수직) ▪ 설문에 대한 자세한 설명 ▪ 설문을 돕기 위한 그림 사용 ▪ 설문을 돕기 위한 동영상 사용 ▪ Open-ended Question사용 ▪ Forced-Question사용 ▪ 실시간 통계/진척도 제시 기능 사용
개인 특성	<ul style="list-style-type: none"> ▪ 연령 ▪ 교육수준 ▪ 성별 ▪ Web/Computer 사용경험 ▪ 직업 ▪ 거주지역
응답자 반응	<ul style="list-style-type: none"> ▪ Perceived Ease of Use ▪ 응답자 태도(Respondent Attitude) ▪ 웹 설문조사 이용 (Web Survey Usage)

또한 본 연구에서는 응답자의 응답률, 작성률, 일관성은 다음과 같이 정의하여 사용하였다. 응답률은 설문 페이지를 본 후 설문에 응한 사람의 비율로 이메일 초청을 받은 Pre-selected와 웹 광고를 보고 스스로 찾아온 self-selected로 나누어 정의하였다 [8, 9].

$$\text{Pre-Selected 응답률} = \frac{\text{타당한 응답수}}{\text{발송 Email 수} - (\text{주소불명} + \text{부적절응답})} \times 100$$

$$\text{Self-Selected 응답률} = \frac{\text{타당한 응답자수}}{\text{설문첫페이지 클릭수} - (\text{부적절응답})} \times 100$$

작성률은 작성한 설문 문항 중 응답한 문항의 비율을 말하며, 일관성 비율은 동일한 설문을 물었을 때 같은 결과가 나오는 지를 측정하기 위한 것으로 응답의 신뢰성을 측정하는 척도로 볼 수 있다. 본 연구에서는 설문응답자가 말하지 않아도 알 수 있는 Cookie 정보를 추출하여 응답자에게 질문 (Browser, OS, 나이, 설문작성 시간) 함으로써 일관성을 추출하였다.

$$\text{작성률} = \frac{\text{작성한 설문문항}}{\text{총 설문문항수}} \times 100,$$

$$\text{일관성비율} = \frac{\text{일관된 응답수}}{\text{전체일관성 측정문항수}} \times 100$$

3.1.2. 개인화된 설문 시스템 구축을 위한 변수 추출

총 16일간 3,445명이 응답한 1차 조사 결과를 분석하여 보면 먼저 응답형태에 관한 선호도 분석결과에서는 대부분 어떤 형태이든 상관없다는 의견이 지배적이었지만, Pull-Down보다는 Radio버튼을 선호하는 경향이 두드러지게 나타났다. 다음으로 응답의 배열형태에 관한 선호는 가로보다는 세로배열을 보다 선호하는 것으로 조사되었다. 그러나, 실제로 세로배열은 설문지의 스크롤을 길게 만드는 단점이 있다는 것을 생각하면, 실제로 이러한 선호가 응답의 질에 긍정적인 영향을 가진다고는 말할 수 없을 것이기 때문에, 이에 대한 구체적인 검증은 2차 설문조사를 통해서 살펴보고자 한다.

최종적으로 1차 사전조사의 응답의 질을 살펴보면, 작성률은 98%, 그러나 설문응답의 일관성 측면은 84%로 약간 낮은 비율을 보이고 있다. 한편, 응답의 질(작성률, 일관

<표 3>개인화 변수선정 결과

	인구통계 관련 변수(6)	인터넷이용 관련변수(5)
작성률	연령, 직업, 거주지역	설문참여경로, 설문참여목적
일관성	연령, 성별, 최종학력, 직업, 월수입, 거주지역	설문참여경로, 설문참여목적, 컴퓨터이용, 접속장소, 인터넷사용

<표 4> 설문지 유형 선호도

웹 설문지 종류	응답형태	배열형태	빈도	%
pwss1.asp	Pull-Down	Row	556	16.1
pwss2.asp	Pull-Down	Column	704	20.4
pwss3.asp	Radio	Row	893	25.9
pwss4.asp	Radio	Column	1,292	37.5

성)의 차이를 나타내는 변수를 ANOVA와 T-Test를 통해 선정하였는데, 선정된 변수를 정리하면 <표 3>과 같다. 총 11개의 변수가 작성률과 일관성의 차이를 설명하고 있는데, 이는 사후조사에서의 사례기반추론의 인덱스(Index)가 되어 사례기반 추론을 위한 선발 질문지(Screening Questionnaire)의 질문내용으로 구성되어 데이터베이스 내에서 동일한 특징을 가지는 사전조사 응답자를 탐색하는 데 이용되어진다.

3.1.3 설문지 형태에 따른 선호도 분석결과

본 연구에서 설문지의 유형을 결정하는 디자인 요소는 사용자 인터페이스에 관련한 문헌연구와 웹 설문지 디자인 구성요소를 기준으로 앞에서 살펴본 바와 같이 8가지 웹 디자인에 관련한 요소를 추출하였으나, 실제적으로 이를 조합한 경우의 수를 구현하여 웹 설문을 만드는 것은 시간과 비용 및 응용측면에서 비효율적이기 때문에, 본 연구에서는 웹 설문지의 차별적인 구성을 이루는데 있어 다소간의 논란의 여지가 있

는 항목들 제외하고, 실험에 있어 명확한 디자인의 차이를 보이는 응답형태(Pull-Down, Radio)와 배열형태(Row, Column)의 조합에 따른 4가지 설문지 유형을 개인화를 위한 설문지로 단순화하여, 비교하였다. <표 4>는 1차 사전조사를 통해 결정된 4가지 웹 설문지의 응답 및 배열형태의 구성과 각각의 선호빈도를 나타낸 것으로서, pwss4.asp 설문지의 경우, 즉 Radio button을 사용하고 세로배열의 응답 항목을 가진 설문지디자인을 가장 많은 응답자가 선호하는 설문지의 형태로(37.5%) 나타났다.

그러나 <표 5>를 보면, 앞서 pwss3.asp의 경우, 전체 선호도 측면에서는 2번째로 높았지만, 전반적인 응답의 질(작성률과 일관성)은 가장 우수한 설문방식으로 나타나고 있어, 다양한 설문방식을 개인적으로 배분하여 응답자의 선호를 만족시키고, 응답의 질을 높일 수 있는 개인화된 설문시스템의 구축이 필요하다는 당위성을 제시하고 있는 것을 알 수 있다.

〈표 5〉 설문지 선호유형에 따른 설문의 질

		pwss1.asp	pwss2.asp	pwss3.asp	pwss4.asp	F	Sig.
작성률	M	97.83	97.85	98.4	98.6	2.175	.089(*)
	SD	10.04	10.1	7.0	6.2		
일관성	M	82.9	82.68	95.3	84.1	3.027	.028(**)
	SD	20.07	19.0	18.3	18.8		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.2. 사례기반추론을 이용한 개인화된 설문조사 시스템의 설계

3.2.1 사례기반 추론의 특징과 처리과정

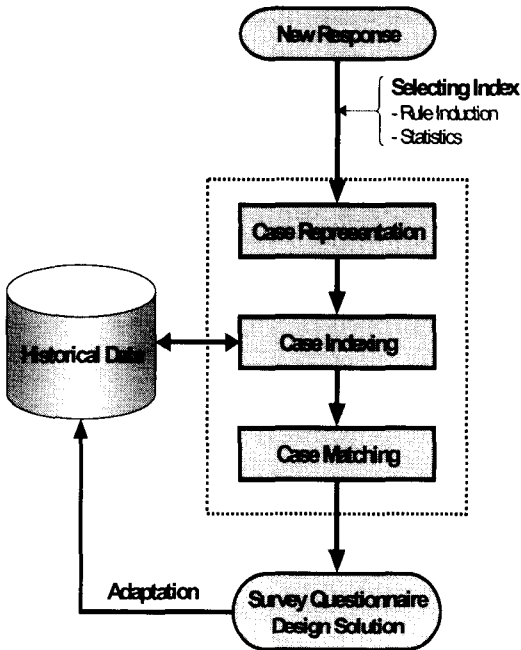
본 연구에서는 개인별 특성에 따라 응답률과 설문 품질을 향상시킬 수 있는 웹 설문지를 추천해 내기 위해서 “사례기반추론(Case-Based Reasoning, 이하 CBR이라 함.)”을 적용하였는데, CBR은 새로운 요구에 대응하는 과거의 해답을 채택하거나, 과거의 사례를 이용하여 새로운 상황을 설명하거나, 과거의 사례로 새 해답을 평가하거나, 또는 새로운 상황을 이해하기 위해서나 새로운 문제에 대한 적당한 해답을 만들기 위해 선례로부터 추정하는 것을 말한다. 특히, 문제해결에 필요한 일정한 규칙을 찾기 힘든 영역과, 특히, 과거의 경험으로부터 효과적인 의사결정을 이끌어낼 수 있는 경우에 매우 효과적인 문제해결 방법론으로 알려져 있다[3]. 따라서, CBR은 문제영역의 특성에 따라 경험과 지식이 문제해결에 유용한 거의 모든 분야에서 활발히 이용되고 있다.

CBR의 추론과정은, 먼저 주어진 먼저 새로운 사례 또는 문제가 주어졌을 때, 사례집합(Case Base)에서 일정한 유사성 척도에

부합되는 과거사례를 추출(Retrieve)한 후, 추출된 사례를 재사용(Reuse)하여 해결에 이용한다. 만약 추출된 사례에 의한 해답이 새로운 문제해결에 적합하지 않으면, 이를 수정(Revise)하여 새로운 해답을 제시하고 이를 다시 새로운 사례로 저장(Retain)하는 과정으로 순환되어진다[3].

3.2.2 개인화된 웹 설문시스템 (Personalized Web Survey System) 추론과정

〈그림 1〉은 본 연구에서 사용한 CBR을 통한 개인화된 웹설문 시스템(이하 PWSS로 칭함)의 추론과정을 설명한 것이다. 1차 사전 설문을 통해 응답의 질과 응답률, 개인별 웹 설문지 디자인에 대한 선호, 인구통계 특징 등 사용자를 특정 짓는 변수들이 사례 베이스(Historical Database)에 저장되어진다. 이 때, 각 사례는 선호하는 웹 설문 디자인 요인에 따라 구성될 수 있는 웹 설문지 디자인의 경우의 수만큼 웹 설문지를 작성하게 되고, 각 사례별로 특정 웹 설문지가 결정되어진다. 본 연구의 경우 응답 형태(pull down & radio button)와 배열형태(가로 & 세로)의 조합인 2×2의 4가지 경우의 수가 있다.



$$DIS_{ab} = \text{Min}(\sqrt{\sum_{i=1}^n W_i \times (f_{ai} - f_{bi})^2})$$

- n : 사례의 수
- W_i : 사례의 가중치
- f_{ai}, f_{bi} : 참조사례, 테스트 사례

<그림 1> CBR을 통한 PWSS의 추론과정

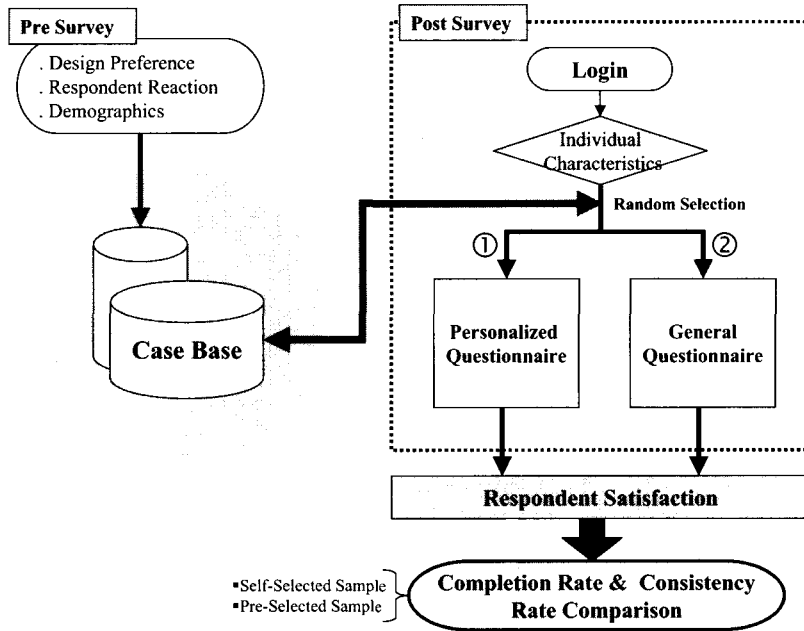
2차 조사에서는 1차 조사에서 선정된 Indexing 변수를 질문하여 DB에 입력되게 하고, 가중치(W_i)와 인덱스의 일치정도($(f_{ai} - f_{bi})^2$)를 곱하여 총합점수를 계산하여(위 공식참조), 가장 높은 유사도(또는 가까운 거리에 있는)를 갖는 사례 1개를 추출(1-nearest neighbor)하여, 그 사례에 저장되어 있는 설문지 디자인의 Prototype을 응답자에게 제시하는 과정으로 이루어진다.

이와 동시에 응답자의 입력결과는 다시 사용된 설문지유형의 설문 품질의 평균으로 대체되는 “Adaptation”과정을 통해 다음 응답자에 Matching에 이용된다.

3.3. 2차 조사

3.3.1 연구모형

<그림 2>에서 보듯이 2차 설문조사는 CBR을 이용하여 구축된 개인화된 설문조사 시스템(PWSS)의 성과측정을 위한 것으로 총 619명을 대상으로 설문을 시행하였다. 설문내용은 사례의 추출을 위한 Indexing 변수를 질문을 하여 PWSS나 GWSS로 배정한 후 본 내용의 설문은 포탈(Portals)사이트의 사용동기에 관련하여 TAM (Technology Acceptance Model) 측정항목들로 구성되어 있다. 먼저 설문조사가 실시되는 웹 페이지에 로그인하게 되면, 응답자는 개인특성변수에 대한 질문에 답하게 되고, Java Script의 Random Selection Option을 통해 CBR 추론을 이용한 개인화된 웹 설문 방식(PWSS: Personalized Web Survey System, ①)을 실시할 것인가, 또는 일반적인 웹 설문방식(GWSS: General Web Survey System, ②)으로 보내질 것인가가 무작위로 결정하게 된다. 물론 2개의 설문 질문 내용은 동일하다. 최종 설문결과 PWSS와 GWSS간의 설문만족도와 응답의 질(완성률과 일관성), 응답률을 비교하게 된다. 또한 응답자들의 유형은 패널회원에게 E-Mail을 통하여 참여를 권유한 경우와 (Pre-Selected), 일반 배너광고나 검색엔진을 통해 참여한 Self-Selected 표본에 따라 추가적인 분석을 시도하였다.



<그림 2> 연구모형

3.3.2 개인화된 웹 설문지 설계

1차 사전조사가 종료된 후, 자료가 저장된 Historical DB에는 응답의 질에 관련한 2개의 필드(완성률과 일관성)와 설문지 디자인 구성요소를 고려한 웹 설문지의 표준 형태에 관련한 필드 1개가 더 추가되어진다. 설문지의 표준형태 필드에는 응답자의 웹 설문지 선호에 따른 웹 설문지의 경우의 수, 즉 <그림 3>에서 보듯이 “버튼(Pull-down, Radio)×배열(Row, Column)”= 4가지의 웹 설문지의 형태가 개인별로 할당되어 지는데, 이러한 웹 설문지 형태들은 사후조사에서 인덱스로 사용할 설문문항에 응답할 경우, 사전조사에서의 응답결과와 가장 유사한 선호를 갖는 응답자중에서 가장 응답의 질이 높았던 응답자의 웹 설문지 디자인 선호형태를 추출하여 응답자에게 뿌려줌으로써, 실제로 개인화된 설문조사시스템이

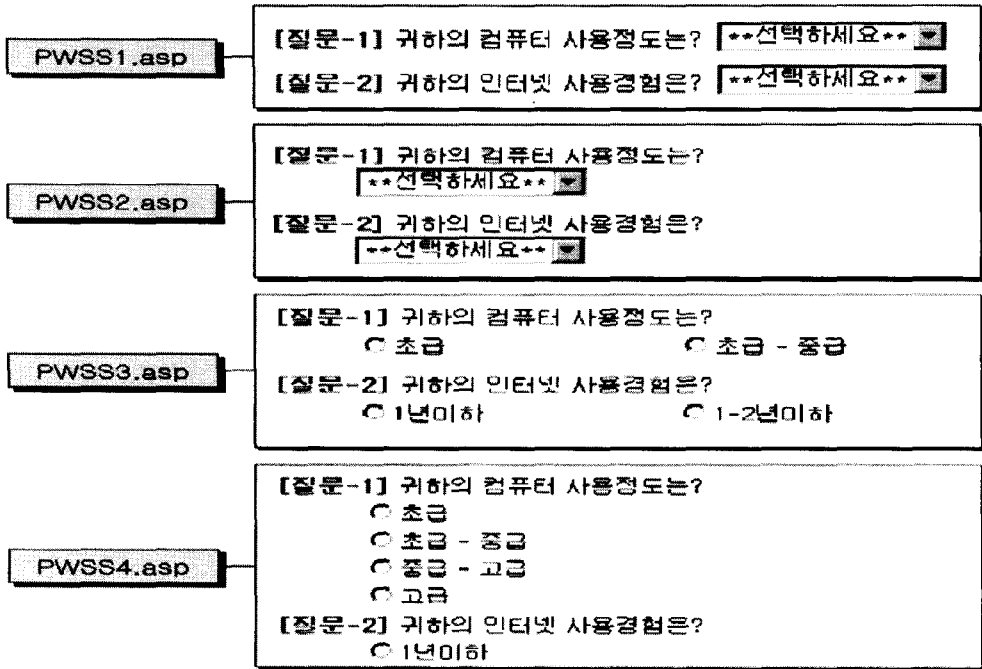
응답의 질을 높일 수 있는지를 검증하는데 이용된다.

4. 분석 결과

4.1. PWSS 대 GWSS의 비교분석 결과

4.1.1 응답률 비교

<표 6>은 사전조사 사례 DB에 근거하여 설문지를 배분하는 CBR을 이용한 개인화된 웹 설문모드(PWSS)와 일반설문모드(GWSS)간의 응답률을 비교한 것으로, Page View는 각각의 개별 설문지 모드의 노출빈도를 가리킨다. 응답률은 앞에서 설명한 바와 같이, Page view를 기준으로, 실제 Submit를 누른 응답자 중에서, 부적절한 응답을 뺀 나머지가 실제 응답률로 계산되



<그림 3> 개인화된 설문 4가지 유형의 예

어진다. 아래에서 보는 바와 같이, 전체 응답률은 71%로서, 노출대비 응답률은 전통적인 설문조사 방법에 비해 매우 높게 나타나고 있다. 한편, PWSS와 GWSS간의 응답률 측면에서 PWSS는 76%로서 GWSS의 63%보다 상대적으로 높은 응답률을 보이는 것을 알 수 있다. 특히 Pull-Down 방식을 사용한 pwss1.asp와 pwss2.asp 보다 Radio Button을 사용한 pwss3.asp와 pwss4.asp가 높은 응답율을 보여 보다 효과적인 인터페

이스임을 알 수 있다.

4.1.2 PWSS의 작성률 및 일관성 비교

<표 7>은 PWSS의 4가지 방식간의 작성률, 일관성의 평균차이를 비교한 결과로서, 사전조사에서의 PWSS설문 종류별 선호도에 따른 작성률과 일관성에서 pwss3.asp와 pwss4.asp가 우수한 방식으로 조사되었던 것과 비슷한 결과를 나타내고 있으나, PWSS의 각 종류별 작성률, 일관성은 통계

<표 6> PWSS와 GWSS의 응답률 비교

	PWSS					GWSS	전체
	pwss1.asp	pwss2.asp	pwss3.asp	pwss4.asp	소계		
Page view	114	122	144	163	543	411	1,086
Submit	66	76	130	142	414	265	828
응답률	58%	62%	90%	87%	76%	63%	71%

<표-7> PWSS 종류별 작성률과 일관성의 비교

		pwss1.asp	pwss2.asp	pwss3.asp	pwss4.asp	f-value	p-value
작성률	M	99.28	98.82	99.35	99.28	1.391	0.245(n.s)
	SD	1.66	2.03	1.89	1.73		
일관성	M	85.48	88.03	86.65	85.45	0.411	0.745(n.s)
	SD	17.85	13.27	20.6	15.42		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

<표-8> PWSS와 GWSS의 작성률 및 일관성

속 성		PWSS	GWSS	f-value	df	t-value	p-value
작성률	M	99.21	97.64	23.963	617	2.505	0.013(**)
	SD	1.83	12.08				

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

속 성		PWSS	GWSS	f-value	df	t-value	p-value
일관성	M	86.3	82.8	5.408	617	2.244	0.025(**)
	SD	17.18	21.24				

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

적 유의한 차이는 없는 것으로 조사되었다. 이는 각각의 설문종류가 개별응답자에게 가장 높은 작성률과 일관성을 얻도록 최적화되어 배분되었기 때문인 것으로 판단되며, 사례기반추론을 활용한 웹 설문조사시스템이 응답의 질을 높히는 방향으로 각 설문이 개개인의 특성에 맞게 잘 배분하였다는 것을 알 수 있다.

4.1.3 PWSS와 GWSS의 작성률과 일관성 비교

<표 8>은 PWSS와 GWSS의 일관성 및 작성률을 비교한 것으로서, 2가지 응답의 질이라는 측정척도에 대해 모두 PWSS가 우세한 방식으로 나타나고 있다. 특히, PWSS의 작성률은 응답의 편차가 상대적으로 낮아 안정적인 작성율을 보이고 있음을

알 수 있다. 또한 일관성의 측면에 있어서도 95% 신뢰수준에서 차이를 보이고 있어 본 연구에서 제안하는 CBR 기법을 활용한 개인화된 웹 설문시스템은 일반 웹 설문시스템보다 작성률, 일관성, 응답율 등에서 우수함을 알 수 있다.

웹 설문지의 고도화 가능성을 있다는 전제로 Dillman et al.(1998)은 다양한 고도화 방법을 제시하고 일반설문 방식과 비교하였으나, 실제 큰 차이가 없어 클라이언트 시스템을 고려한 단순한 웹 설문지 인터페이스를 권고한 바 있다. 그러나 인터넷의 보급, 확산과 기술향상을 고려하여 볼 때, 본 데이터마이닝 기법을 활용한 개인화된 설문조사시스템이 설문응답의 질을 높인다는 실증적 연구결과는 의미있는 연구결과를 보이고 있다. 이러한 이론적 활용방안이 실용적

<표 9> 표본추출방법에 따른 비교

속 성		Pre-Selected	Self-Selected	f-value	df	t-value	p-value
만족도	M	3.21	3.13	0.416	511	1.196	0.232(n.s)
	SD	0.71	0.74				
작성률	M	98.72	98.85	0.05	572	-0.241	0.809(n.s)
	SD	6.91	6.05				
일관성	M	86.21	83.81	0.505	535	1.605	0.109(n.s)
	SD	18.41	18.06				

*p<0.1, **p<0.05, ***p<0.01

인 방향으로 이용되기 위하여서는 시스템의 안정성과 처리속도가 매우 주요한 이슈가 되는데, 실증적 분석결과 3445개의 데이터베이스를 검색하는 시간과 유사성 (Euclidean Distance로 측정)을 계산하여 웹 설문지 방식을 제시하여 설문지를 전송하는 속도에 있어 일반 설문지 방식과 큰 차이가 없었다는 것이다. 물론 앞으로 데이터베이스가 매우 커져 검색 하는데 점점 시간이 많이 들 것으로 보이지만 이 문제는 향후 대용량 데이터베이스가 구축되어 실용화 되는 것을 가정하여 보다 깊이 있는 추가 연구가 필요할 것으로 보인다.

4.2. 표본추출 방법에 따른 비교분석

<표 9>는 연구자가 이메일을 통하여 초청한 응답자의 Pre-selected와 배너광고, 경품사이트, 검색엔진을 통해 참여한 Self-selected의 경우를 나누어 비교, 분석한 결과이다. 만족도와 작성률 측면에서는 통계적으로 유의한 차이가 존재하지 않으나 일관성부분에서는 Pre-selected가 보다 높은 것으로 나타났다. 이 것은 연구자의 초청을 받은 응답자의 경우 가능한 호의적이

고 성실하게 답하려고 하는 Social Desirability Bias가 작용한 것을 알 수 있다. 또한 Self-selected의 경우 경품을 받고자 하는 목적으로 설문참여를 하게 되어 실제 작성률은 높으나 성실도가 상대적으로 낮았다는 것을 알 수 있다.

5. 결론 및 향후 연구과제

본 연구의 목적은 전통적인 설문방식에서 응답의 질을 통제했던 인터뷰어(Interviewer)의 역할을 인터넷 설문조사에서는 사례기반추론이라는 데이터마이닝 기법으로 대체하여 구현해보자는 것이었다. 이는 기존의 고객의 인지적 정보를 수집하는데 있어 오프라인 설문방식보다 웹 기반의 상호작용성과 멀티미디어적 장점을 최대한 활용하여 더욱 친근한 인터페이스를 구현함으로써, 무응답 오류와 측정 오류를 줄여 설문 품질을 높이고자 하는 연구목적에 가치고 출발하였다

연구 분석 결과 <표 10>에서 보듯이 사례기반 추론을 이용한 PWSS가 응답률, 작성률, 일관성 측면에서 모두 GWSS에 비하

여 우수한 것으로 나타났다. 그러나 설문만족도에서는 통계적인 유의한 차이는 없었으며, 표본 추출방식에 있어서도 일관성만이 사전 이메일초청자인 Pre_selected가 배너광고나 검색엔진을 보고 온 Self-Selected 보다 우수한 것으로 나타났을 뿐 다른 큰 차이는 없는 것으로 나타났다.

<표 10> 연구결과의 요약

	PWSS	GWSS
응답률	✓	
작성률	✓	
일관성	✓	
만족도	n.s	n.s

Pre-Selected	Self-Selected
n.s	n.s
✓	n.s
n.s	n.s

이러한 결과를 종합적으로 분석하여 보면, 개인에 따른 설문지 디자인 요소의 선호가 존재하고, 이러한 것을 응답자의 선호와 프로파일에 근거하여 적절히 배분할 경우, 응답율과 일관성의 통제가 가능하며, 응답률을 높일 수 있다는 것을 알 수 있다. Couper(2000)와 Dillman(2002)은 인터넷 설문조사에서 다양한 디자인 요소를 고려함으로써 무응답오류와 측정오류를 상당부분 줄일 수 있으며, 더 나아가 대표성 오류도 개선할 수 있다고 하였다. 그런 점에서 본 연구는 보다 높은 응답률과 응답의 질을 확보하기 위하여 인터넷 설문지의 디자인 요소가 영향을 주고 있다는 것을 실증적으로 검증하였다. 특히 사례기반 추론이라는 데이

터마이닝 기법을 활용하여 개인화된 지능형 인터넷 설문조사 시스템을 설계함으로써, 인터넷 설문 디자인의 영향을 어떻게 시스템화 시키고 끊임없는 학습과정을 통하여 고객 개개의 특성에 맞는 설문지를 만들어 가는가에 관한 이론적, 실증적 토대를 제시하였다고 할 수 있다.

본 연구는 설문내용, 설문길이, 설문기간, 설문형태(layout), 보상 및 유인, 프로모션, 응답자 상호작용 등을 통제하였는데 이러한 다양한 요인들에 관한 상호간 영향도 측정할 필요가 있다. 기존 웹 설문조사에 관한 실증적 연구를 살펴보면 이러한 요인들이 설문응답의 질에 영향을 미치는 것으로 나타나고 있지만 본 연구에서는 연구의 범위가 너무 확대되어 체계적으로 고려할 수가 없었다. 향후 이러한 요인들을 포함한 보다 정교한 실험설계를 통하여 표준 인터넷 설문지 디자인에 관한 원리 정립이 되었으면 하는 기대를 한다.

그 외 인터넷 설문의 장점을 더욱 살려, 그래픽 또는 동영상 사운드를 포함한 인터넷 설문에 관한 보다 심도깊은 분석을 통하여 개인화된 설문조사 시스템의 보다 완벽한 구축을 기대하며, 사례기반추론이외의 다양한 데이터마이닝 기법의 활용에 관한 연구도 기대된다. 특히 대용량 데이터가 구축되었을 경우 사례기반추론의 실용화에 어떤 영향이 있는지에 관한 실증적 연구도 필요하다고 본다.

참고 문헌

- [1] 김광용 & 김기수 (1999a), "다양한 인터넷 설문 방식 비교와 설문만족에 관한 실증

- 연구”, 「한국경영정보학회 춘계학술논문집」, pp. 181-190.
- [2] 김광용 & 김기수 (1999b), “웹 설문지 디자인과 작성률, 일관성에 관한 연구” 「한국 분류학회 학술논문집」, pp. 35-43
- [3] Aamodt, A & Plaza, E, (1994), “Case-Based Reasoning: Foundational issues, Methodological Variations, and System Approaches”, *AI Communications*, Vol. 7, No. 1. pp. 39-59
- [4] Batagelj, Z & Vehovar, V, (1988), “Technological and Methodological Issues In WWW Surveys”, *AAPOR98, Software and Methods for conducting Internet survys*, St.Louis,
- [5] Batagelj, Z. Vehovar, V. Lozar, K, (1999), “Web vs. telephone surveys : An empirical study of mode effects and non-response behaviour”, *JSM99 conference*, Baltimore, Maryland.
- [6] Beniger, J. R. (1998), “Presidential address : Survey and market research confront their futures on the World Wide Web”, *Public Opinion Quarterly*, Vol. 62, No. 3, pp. 329-354
- [7] Couper, M. P. (2000), “Web Surveys: review of issues and approaches”, *Public Opinion Quarterly*, Vol. 64, No. 3, pp. 464-294
- [8] Couper, M. P., Traugott, M. W, and Lamias, M. J. (2001), “Web Survey Design and Administration”, *Public Opinion Quarterly*, Vol. 65, No. 2. pp. 230-253
- [9] Dillman, D. A., (2002), “Navigating the rapid of change: some observations on survey methodology in the early 21 century”, *Public Opinion Quarterly*, Vol. 66, No. 3. pp. 479-494
- [10] Dillman, D.A, Tortora, R.D., & Bowker, D., (1998), “Influence of Plain Vs. Fancy Design on Response Rates For Follow-up Web Surveys.”, *Proceeding of Survey Methods Section, Annual Meetings of the American Statistical association, dallass, Texas.*
- [11] Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinly, L. (2001), “Insuaring sample equivalence and paper and pensil assessment”, *Computers in Human Behavior*, Vol. 17, No. 2, pp. 339-346
- [12] Faight, K. S., Green, K. W, and Whitten, D. (2004), “Doing survey research on the internet”, *Journal of Computer Information System*, Vol. 5, No. 1, pp. 26-34
- [13] Feinberg, M. J. & Johnson, D., (1998), “Designing and Developing Surveys on WWW Sites.”, *Communication of ACM*, pp.38-42.
- [14] Furrer O. & Sudharshan D.(2001), “Internet marketing research_ opportunities and problems,” *Qualitative Market Research: An International Journal*, Volume 4 . Number 3 . 2001 . pp. 123-129
- [15] Johnson, M. J. (2001), “Digitizing Consumer Research,” *Journal of Consumer Research*, Vol. 28, pp. 232-258

- [16] Klassen, R. D. & Jacobs, J. (2001), "Experimental Comparison of Web, Electronic and mail survey in operations management", *Journal of Operations Management*, Vol. 19, No. 4, pp. 713-718
- [17] Knapp, H. & Kirk, S. A. (2003), "Using pencil and paper, Internet and touch tone phones for self administered surveys: does methodology matter?", *Computers in Human Behavior*, Vol. 19, No. 1, pp. 117-134
- [18] Krotki K. & Dennis M., (2001), "Probability Based Survey Research on the Internet," *AAPOR conference paper*
- [19] Hayslett, M. M. & Wildemuth, B. M. (2004), "Pixels or Pensils? The relative effectiveness of web-based versus paper surveys", *Library & Information Science Research*, Vol. 26, No. 1, pp. 73-93
- [20] Porter, S. R. & Whitcomb, M. E. (2003), "The Impact of Contact type on Web Survey Response Rates", *Public Opinion Quarterly*, Vol. 65, No. 4, pp. 579-588
- [21] Schaefer & Dillman, "Development of a Standard E-Mail Methodology : Results of an Experiment", *Public Opinion Quarterly*, 62, 1998
- [22] Schillewaert, N, (1998), "Non-Probability Sampling for WWW Surveys : A Comparison of methods", *The Journal of Market Research Society*, vol. 40, No. 4, pp. 307-322.
- [23] Vehovar, V., Batagelj Z., Lozar K., & Zaletel M, (199), "Design Aspect of the Web Questionnaires", *the 10th Annual Advanced Techniques(ART) Forum(AMA)*.
-