조사자료 데이터베이스의 허위 잠재 가능성 분류군 탐지

변 루 나[†] · 한 정 혜[††]

요 약

인간이 자료를 생산하여 구축하는 조사자료 데이터베이스는 응답자나 조사자의 고의 또는 실수로 인해 비표본오차는 언제든지 발생할 수 있고 그에 따른 오차를 찾아내는 조사관리는 시간적 비용적 기술적으로 대단히 어렵다. 조사관리를 시의성 있게 수리적이고 체계적으로 찾아내는 일이 결코 쉽지 않기 때문에, 지금까지는 단순히 조사항목 연관성 불일치 또는 임의로 선택한 현장을 방문하여 착오 자료 등을 찾아냄으로써 조사관리 하는 것에 불과하였다. 이에 본 연구에서는 비표본오차 중에서 응답자나 조사원의 허위응답과 허위조사를 예방할 수 있는 잠재 가능성을 분류하는 휴리스틱한 방법을 제시하고자 한다. 먼저 일정한 기간마다 지속적으로 실시되는 조사를 대상으로 질적, 양적 자료의 구성에 관계없는 이항반응 자료로 변환하여 허위일 가능성이 있는 패턴을 찾아보았다. 그리고 조사구의 지리적 위치도 고려하여 최종 허위응답과 허위조사 잠재 가능성 분류군을 탐지하였다. 분석결과 허위조사의 경우를 정확히 탐지하였으며, 허위조사 잠재 가능성 분류군에 대한 특징적인 지식을 얻을 수 있었다. 본 연구결과는 비표본오차를 보다 정확하고 시의성 있게 관리할 수 있는 조사관리 방법론을 제공함으로써, 조사자료 데이터베이스 품질을 높일 수 있는 가능성과 의의를 가진다.

키워드 : 조사자료 데이터베이스, 비표본오차, 허위조사, 허위응답, 조사관리

The Detection of Unreliable Data in Survey Database

Luna Byon[†] · Jeong Hye Han[††]

ABSTRACT

The Non-Sampling Error can happen any time by means of the intended or unintended error by the interviewer or respondent, but it is very difficult to find the error in survey database because it can hardly be computed mathematically and systematically. Until now, we have found it accidentally through the simple relation between the items or through the inspection from the random field. Therefore we introduced an heuristic methodology that can detect the interviewer's error by statistical decision-making or data mining techniques with a case study. It will be helpful so as to improve the statistical quality and provide efficient field management for the supervisor.

Key Words : Survey Database, Non-sampling Error, Unreliable Interview, Unreliable Response, Field Management

1. 서 론

전자상거래 트랜잭션 데이터나 웹 로그 등과 같은 대부분의 데이터베이스는 자료생산부터 저장까지 컴퓨터가 수행하여 얻어짐으로써, 네트워크의 장애나 컴퓨터의 오작동을 제외하고는 완벽한 자료를 생산한다. 그러나 국가의 경제, 사회문제 등 효율적인 의사결정을 하기 위하여 인간이 직접 자료를 생산하여 구축하는 조사자료 데이터베이스는 인간의 성실성에 따라 자료의 정확성이 의존하게 된다. 현재 조사자료 데이터베이스의 생산은 노트북이나 PDA와 같은 CAPI(Computer Aided Personal Interview)방식으로 조사자가 현장에서 온라인으로 전송하는 방식 등을 도입하고 있다.

이와 같은 조사자료 데이터베이스를 토대로 각종 경제, 사회통계 등이 얻어지므로, 올바른 정책수립에 성공하기까지 연결되는 데에는 무엇보다도 응답자의 성실한 응답과 조사자의 정확한 자료 입력이 요구된다. 현재는 주로 자료의 범위(Range)와 같은 간단한 유효성검사 프로그램을 활용하여 오류입력 방지가 전부이다. 그러나 응답자과 조사자의 상호작용을 통해서 생산되는 조사자료 데이터베이스에는 당사자들의 내적 요인과 외적 요인들에 의하여 허위응답 및 허위조사의 가능성이 충분히 내재되어 있을 수 있다. 따라서 조사관리자의 현지 방문 실사를 통해서 조사자의 자료입력 사항을 확인 및 점검함으로써, 조사자료 데이터베이스를 모니터링 하는 것이다.

현재 조사관리자는 자신의 경험이나 직관에 의존하여 조사관리를 하고 있으며, 조사관리 업무 지식의 축적이나 업무 인수인계가 되고 있지 않아 장기적인 품질향상에 대한 전략이나 대비도 없는 상태이다. 뿐만 아니라 조사관리자를 통한 조사관리 방법론은 응답자와 조사자라는 인적 자원,

† 정 회 원 : 통계청 통계기획국 조사관리과
†† 종신회원 : 청주교육대학교 컴퓨터교육과 조교수
논문접수 : 2004년 12월 31일, 심사완료 : 2005년 4월 20일

다양한 조사항목의 형태, 행정 시스템 외에도 시간과 비용적인 문제가 수반되어 아직까지 과학적인 방법으로의 접근이 매우 어려운 상태이다. 따라서 허위자료가 포함된다면 정책수립의 근본을 흔드는 치명적인 영향을 줄 수 있다는 중요성을 생각해볼 때 과학적인 조사관리 탐지알고리즘이 시급히 요구된다고 하겠다.

농어촌 통계조사와 같이 조사자가 면접조사를 토대로 구축되어지는 조사 데이터의 경우는 '111…111'과 같은 단순 허위패턴은 없다고 가정할 수 있다. 왜냐하면 그런 단순 허위패턴은 육안으로도 비교적 탐지가 손쉬우며, 이미 전산 프로그래밍에서 필터링을 하고 있으며, 일단 발각될 경우 조사자를 추적하여 인사 조치를 취할 수 있기 때문에 그런 경우가 발생할 가능성은 거의 없다고 하겠다. 따라서 본 논문에서는 조사자료 데이터베이스의 품질을 높이기 위해 지능적인 허위응답이나 허위조사의 잠재적 가능성 집단을 데이터마이닝(Data Mining)을 활용하여 탐지함으로써, 이 집단을 우선적으로 조사관리 하도록 의사결정을 지원하고자 한다. 즉, 허위일 가능성이 있는 자료를 미리 선별하여 조사관리를 함으로써 조사자료 데이터베이스 기반의 통계품질 향상 기여에 그 목적이 있다.

본 연구내용은 2장에서는 관련연구 및 허위조사의 현행 한계점을 중심으로 살펴보고, 3장에서는 허위조사 또는 허위응답 탐지를 위한 사례분석을 위한 실험 설계와 패턴분석 알고리즘을 제시한다. 그리고 4장에서는 분석결과 및 평가를 서술하였다. 마지막으로 5장에서는 본 알고리즘이 조사자료 데이터베이스 품질에 기여할 수 있는 실용성과 향후 연구 과제를 설명하였다.

2. 관련 연구

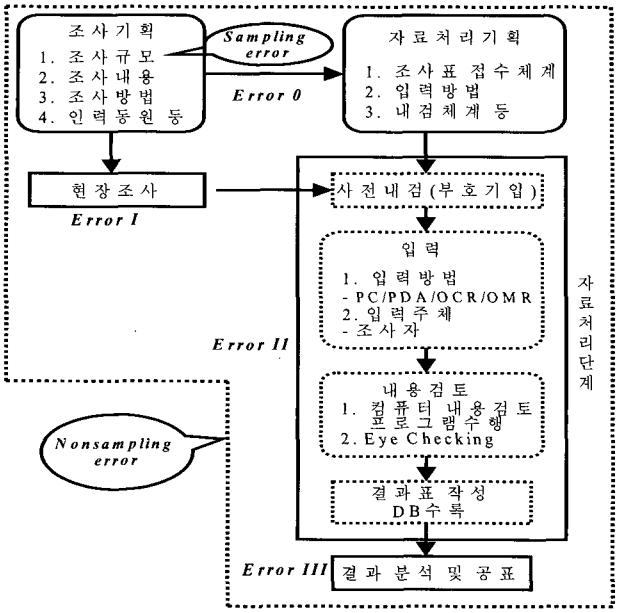
2.1 조사자료 데이터베이스

일반적으로 조사자료 데이터베이스는 (그림 1)과 같이 조사와 자료처리 기획으로, 조사규모, 조사항목, 조사방법, 입력방법, 내용검토체계, 조사표 접수체계 등의 계획을 수립하고 이에 따라 조사가 끝나게 되면 자료처리단계에 접어든다. 자료처리단계는 크게 「사전내용검토 및 부호기입단계」, 「입력」, 「컴퓨터 프로그램을 통한 내용검토」를 거쳐 「결과표작성 및 DB수록」 이용자들의 「통계 이용」으로 나뉜다.

조사자료 데이터베이스는 표본오차(sampling error)와 비표본오차(non-sampling error)로 구성된 우연오차(variable error)를 갖는다. 표본오차는 여러 표본이론을 통해서 어느 정도 과학적으로 관리가 이루어지고 있는 반면, 비표본오차는 표본오차 이외에 조사의 기획단계에서부터 최종 보고서의 공표에 이르기까지 모든 과정에서 부주의나 실수, 또는 알 수 없는 원인 등으로 발생되는 오차로서 표본조사나 전수조사의 모두에서 발생이 가능하다.

2.2 허위조사와 허위응답

비표본오차 중 부주의나 실수로 인한 조사 자료의 입력이



(그림 1) 조사자료 DB의 통계생산 프로세스

나 고의적인 허위가 존재할 수 있다. 허위응답이란 응답자가 조사자가 정의하고 기획한 대로 응답하지 않거나 정확하게 대답하지 않고, 실수나 고의로 왜곡, 누락 또는 거짓으로 조사자료 데이터베이스에 입력하는 경우를 일컫는다. 또한 허위조사는 조사 기획자가 정의하고 기획한 대로 정확하게 조사대상자를 면접하여 조사하지 않음으로써, 실제적인 통계사실을 실수나 고의로 왜곡, 누락 또는 거짓으로 조사자료 데이터베이스에 입력하는 경우를 말한다.

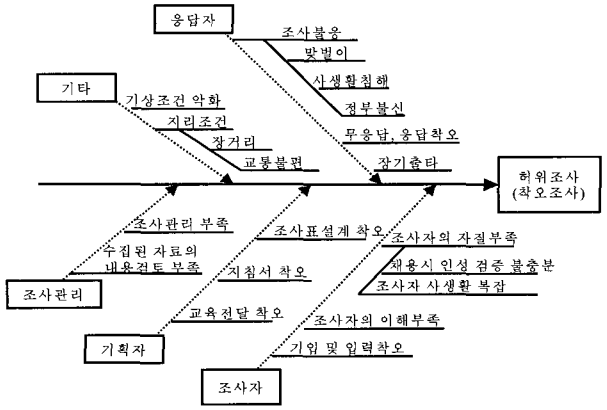
비표본오차의 최소화를 위해 많은 연구가 활발히 진행되고 있는데[1], 이는 통계의 질적 향상을 위하여 모집단의 선택, 표본설계, 조사표 설계, 인터뷰 기법 등으로부터 비표본오차를 줄이기 위한 관리방안을 제시하고 있다. 그러나 이러한 연구들은 허위조사 탐지로 비표본오차를 최소화하기 위한 조사관리를 위한 방법론에 대해서는 언급하고 있지 않으며, 경험적인 해결법으로 지식관리시스템 구축에 대한 제안 등도 제시하지 못하고 있다. 왜냐하면 허위조사(자료)의 발생 요인은 많은 인적, 지리적, 사회적, 그 외에 다양한 원인에 의해서 존재할 수 있으므로, 그 원인을 찾아 제거하거나 허위여부를 탐지하기 위한 과학적 접근이 매우 어렵기 때문이다.

따라서 본 연구에서는 먼저 허위조사에 영향을 미친다고 생각되는 요인들을 품질개선 활동에서 널리 사용되고 있는 이시가와 다이어그램(Ishikawa Diagram)을 활용하여[2], 조사관리의 경험과 지식을 토대로 (그림 2)와 같이 정리하였다.

2.3 허위 탐지 유사연구

본 절에서는 허위조사 탐지를 위한 지식의 관리와 실증분석 사례 연구가 어떻게 이루어져야 하는가에 대하여 유사 연구들을 살펴봄으로써 해결방법을 모색하고자 한다.

먼저 유사한 토픽관련 연구로는 인터넷에서 웹서버에 대



(그림 2) 허위조사에 대한 특성요인도

한 침입 및 범죄로 인한 비정상 행위 등의 피해를 데이터마이닝 기법을 이용하여 탐지하는 알고리즘이다[3, 4]. 그러나 이러한 방법들은 컴퓨터가 생성하는 자료 데이터를 기반으로 하는 것으로, 조사자료 데이터베이스에 적용할 수 없다. 또 다른 유사 연구로는 보험사기 적발모형과 같은 연구 등이 있으나[5, 6], 이는 보험사고 피해자의 인간관계도, 사고유형 등을 통해서 반복적인 보험사고 피해 금을 받아가는 것을 탐지하는 것이다. 따라서 보험사기 적발을 위하여 피해자와 가해자 간의 트리 구조를 바탕으로 자료의 시각화를 제공하는 첨단 소프트웨어의 활용하여 사기 적발탐지를 하고 있다. 그 외에도 기업 내 사원부정의 탐지 연구[7]도 있는데, 이는 단순 사례유형의 카이스퀘어(χ^2)분석만을 하였으며 향후 사내 부정에 대한 예측까지는 다루지 못하는 한계를 갖고 있다.

이와 같이 본 논문의 허위 탐지와 유사한 토픽의 기존 연구들이 일부 있기는 하지만, 며, 주로 보험사기 적발 모형과 관련된 기초연구가 있는 수준이다. [1]에서 언급 했듯 조사자에 의한 인위적 또는 무의식적 오류들의 종류와 탐지의 필요성을 있으나, 조사자료 데이터베이스에 대한 방법론상의 부재와 적용의 어려움으로 인하여 사례연구나 예측방법에 대한 연구는 전무한 실정이다.

따라서 본 연구에서는 보험사기 적발모형, 데이터 마이닝 기법, 지리적 계량화, 통계품질 요소를 관리하는 기법인 관리도와 특성요인도 기법[1] 등을 복합적으로 고려하여, 조사자료 데이터베이스에 대한 허위탐지 가능성을 제안해보고자 한다.

3. 허위조사 및 허위응답 탐지 프로세스

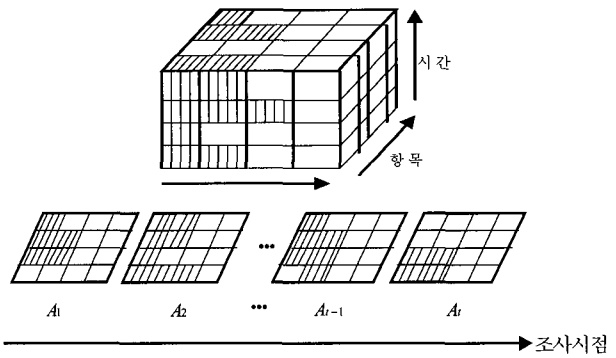
3.1 실험 설계

본 절에서는 허위 가능성을 갖는 자료를 탐지하기 위해 전국에 약300여명의 조사자에 의해 2002년 10월부터 2004년 6월까지 월별로 수집되고 있는 A조사를 실험대상으로 하였다. 조사 자료의 구성은 320가구로 레코드별 144개의 조사항목으로 구성되어 있으며 매일 매일의 가계수지 내용을 기록한 내용이다. 144개의 항목에 대한 계량적 자료이다. 36명의 조사자를 임의 추출한 후, 사전정보(priori information)를 이용하여 성실군 18명, 임의군 17명, 불성실군 1명(적발사례) 실험데이터 셋을 세팅하여 본 논문의 방법론을 통한 각 집단별 탐지결과를 비교했다.

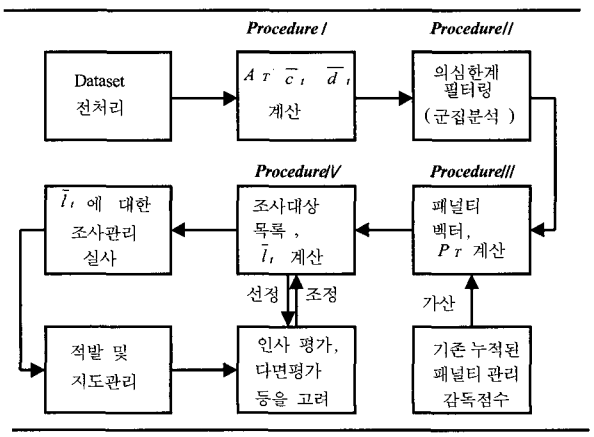
3.2 조사관리를 위한 패턴분석의 알고리즘

본 절에서는 허위자료 탐지를 위한 패턴분석 알고리즘을 제시하고자 한다. 실험데이터 셋, $DS=(d_{ijk})$, I=조사구, j=조사항목, k=시간은 조사시점(21개월), 조사구(36개), 조사항목(144개)으로 3차원(Cubic)이다. (그림 3)은 각 시점별로 조사구와 조사항목으로 음영부분은 조사자의 인사이동으로 조사자의 변동이 있는 경우이다.

각 시점별로 조사구와 조사항목으로 구분되는 자료에 대하여 기존의 시계열분석 모형 등을 적용하기는 적당치 않다. 그 이유는 허위조사의 가능성을 갖는 패턴의 탐지가 목적이지, 이 자료를 모형화하여 추정 값을 예측하는데 관심을 두는 것이 아니며, 또한 조사자의 인사이동을 고려하여야 하기 때문이다. 따라서 이 실험데이터를 시점별로 나눈 후, 조사항목은 조사시점별로 계절적 요인이 유사하다는 가정 하에서, 관리기법을 활용한 허위조사 패턴 프로세스는 (그림 4)와 같다.



(그림 3) 실험 데이터 셋



(그림 4) 탐지 프로세스

```
Procedure Ⅰ : preprocessed Dataset
Input : D_original, time, num_res, num_item
Output : A_T, c_T, d_T
c_0 = 0
1  for t = 1 ,···, time,
2  for i = 1,···, num_res
3  for j = 1,···, num_item
4  If {d_ij | d_ij ∈ D_original, d_ij = null), then
                a_ij = 0  else a_ij = 1
5  end
6  end
7  A_t = get(a_ij)
8  c_t = count(a_ij) for all i, j
9  d_t = c_t − c_t−1
10 end
```

```
Procedure Ⅱ : Calculation of clustering
Input : time , num_res, d_t  ε
Output : clusteres_sets, cl_max
1  for t = 1 ···, time
2  for i = 1 ···, time, num_res
3  for j = 1,···, num_item
4       clustering (t_i)
5  end
6  end
7  end
8  cl_max = number(cluster)
```

```
Procedure Ⅲ : Calculation of panalty
Input : cl_max, clustered_sets, etc_p,
        num_base
Output : p_T
1  for t = 1 ···, time
2  for i = 1 ···, time, num_res
3  for j = 1,···, num_item
4       p_i = p_i + etc_p_i
5  end
6  end
```

```
Procedure Ⅳ : Calculation of l_t
Input : p_T, UCL, LCL, num_res
Output : l_T
1  for i = 1 ···
2      If {LCL ≤ p_i ≤ UCL), then l_i = 0
                            else  l_i = 1
3  end
```

(그림 5) 탐지 단계별 알고리즘

먼저 실험데이터 셋은 응답자의 조사결과에 따라 계량 데이터가 들어있거나 해당이 없는 경우 널(null)값을 갖고 있으므로, 전처리(pre processing)을 통해 이진(binary)으로 변환한 A_t를 얻는다. A_t로부터 각 조사항목에 대하여 합을 구하여, 전체 144개 조사항목 중 각 응답에 해당하는 항목의 개수를 시점별로 얻는다. i번째 조사구에서 응답 항목수의 합을 c_t라 하고 주기를 12개월로 잡아 전년대비 항목 갯수의 차 벡터 D_t를 구한다. 이 벡터를 대상으로 각 조사구에 대하여 계층적 군집분석(hierarchical cluster), R(range)-Chart등을 실시하여, 특별히 차이가 크거나 작은 조사구를 추출해 낸다. 이 조사구의 조사자에 대하여 다른 여러 조사관리 지식항목과 조사위치(지리적 접근 편의성)를 고려한 패널티 벡터 P_t를 계산한다. 최종 계산된 벡터 P_t의 분포를 고려하여 관리 한계값을 초과하는 조사자를 추출하여, 조사관리 감사 대상으로 최종 분류한다. 이와 같은 단계별 알고리즘을 의사코드로 나타내면 (그림 5)와 같다.

위 (그림 5)의 num_res과 num_item은 조사구수와 항목수를 의미하며, 단계Ⅱ에서의 $\overline{d_t}$은 전년대비 동일 조사구내의 응답항목수의 차, ε은 조사관리자가 정한 허용한계, $clusteres_sets$은 군집에 따라 분류된 셋, cl_max는 군집의 개수를 뜻한다. 단계Ⅲ에서 etc_p은 각 기관내의 평가기준에 의한 다른 여러 종류의 패널티 벡터를 뜻하며, 마지막 과정의 l_T은 최종 허위탐지 가능성이 의심되는 조사구 리스트 벡터이며, UCL과 LCL은 상하선의 허용한계로서 조사관리자가 사전에 정하는 값 이다.

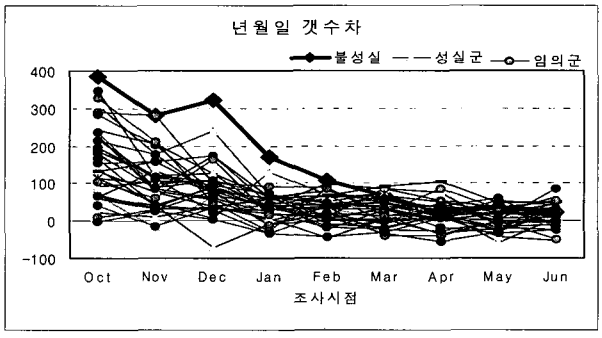
4. 사례분석 및 활용

4.1 분석 프로세스 적용

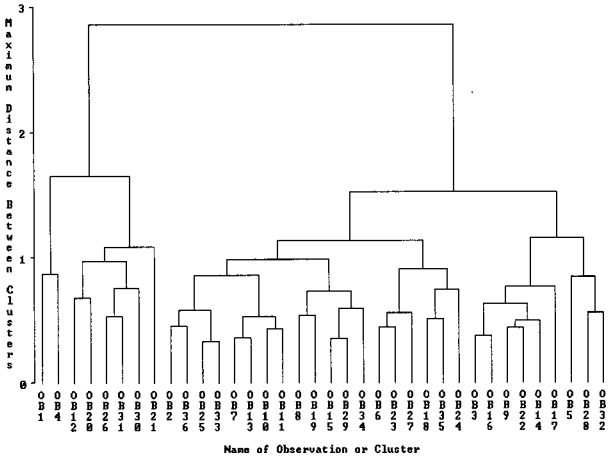
본 절에서는 허위자료 탐지를 위한 단계별 알고리즘에 따라 실증분석을 실시하였다. 먼저 36명의 조사자에 대하여 응답한 조사구의 값별로 이진행렬에 대한 합산 값을 대상으로 년 주기별로 D_t를 계산하였다. (그림 6)은 9개 조사시점별로 36개 조사자에 대한 결과이다.

이 때, 허위조사로 지적받은 조사자의 조사구가 가장 위에 위치하였음을 가시적으로도 분명히 나타나고 있어, 본 계량법의 단순성에 비해 정확함을 가시적으로 보여주고 있다. 또한, 이 D_t에 대해 완전연결법(complete linkage method)을 이용하여 조사구별 군집분석의 결과 나타난 수직 덴드로그램은 (그림 7)과 같다.

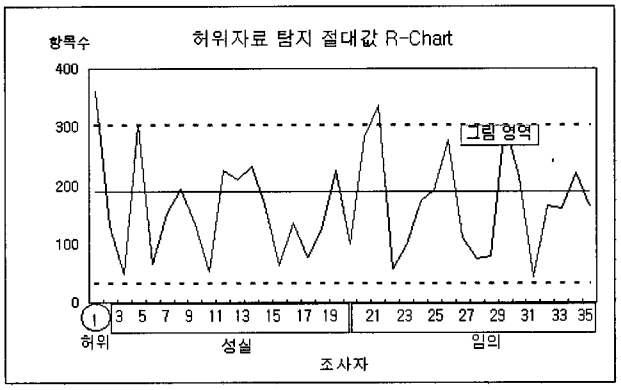
36개 조사구의 조사항목 수를 전년 동월과 비교하여 그 차 벡터를 토대로 조사자별로 변동의 폭을 점검하기 위하여 R-Chart를 작성하였다. 다음 (그림 8)의 그래프는 전년 동월 대비 차이의 절대 값을 취한 결과이다. 이 경우 만약 전년 동월 대비 차이가 거의 없다면 허위자료를 작성하는 조사자를 탐지할수 없는 단점이 있다.



(그림 6) 조사항목별 전년 동월대비 차 분포

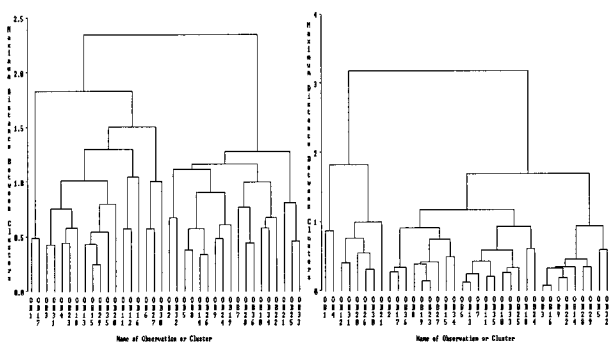


(그림 7) 조사구별 덴드로그램



(그림 8) R-Chart관리도(n=9일 때 d3=0.184)

(그림 8)에서 조사자 1번, 21번, 30번은 관리한계를 벗어나며, 관리한계 내에는 들지만 3번, 4번, 5번, 9번, 14번, 22번, 26번, 32번도 관리대상에는 포함해야한다고 판단된다. 그러나 이런 조사구별로 이루어진 군집데이터에는 시험조사기간의 영향, 인사이동으로 인한 변동으로 단일 조사자가 아니라 두 명 이상의 조사자들이 포함된 경우도 있다. 보통 인사이동에 따른 업무분장이 1월에 있으며 시험조사도 12월에 끝나므로, 1월을 전후로 덴드로그램을 (그림 9)와 같이 구해보면 실제로 현격한 차이가 있음을 반증한다.



(그림 9) 조사자 덴드로그램

〈표 1〉 허위조사 의심군 분류 결과

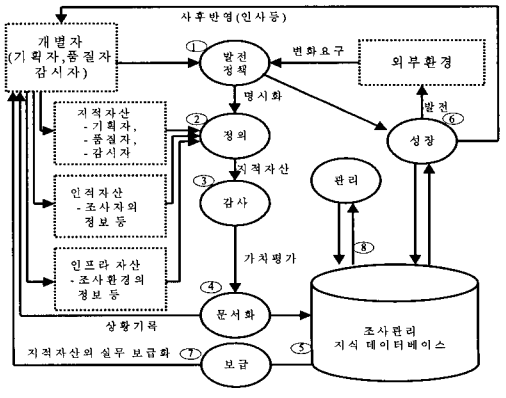
실험군	조사구	지리적 특징 (소요시간)	최종분류
불성실군	X1	약1시간	허위
임의군	X20	약 40분	의심 조사구
	X30	1시간20분	
	X27	1시간 30분	
성실군	X5	30분	
임의군	X32	약40분	
성실군	X4	약40분	
임의군	X28	약40분	
	X26	1시간	
성실군	X12	약40분	

따라서 시점별 그래프, 군집분석, R-Chart 등을 종합하여 조사관리 대상 조사구를 1차적으로 선별한 후 이를 대상으로 해당 조사구를 담당하는 조사자의 인사이동 날짜(1월 기준), 지리적 특수성, 조사자에 대한 평소 근무 스타일 등을 종합적으로 고려하여 최종 패널티 벡터값 P_t을 계산하여, 다음 <표 1>과 같이 최종 허위조사 의심군을 선별하여 집중 조사관리 대상으로 분류하였다.

4.2 잠재 의심 분류군의 패턴 지식관리

<표 1>을 보면 허위조사로 감사에 지적된 1번 조사구에 대하여 전반적인 패널티 점수가 아주 높게 부여가 되어 본 결과가 정확함을 알 수 있다. 사전 성실군 조사구 4번, 5번, 12번의 경우 조사자들의 인사이동을 고려할 때 전임자들의 시험조사결과로 허위가능성이 탐지된 것이었다. 그리고 나머지는 임의군에서 허위조사 의심 조사구으로 추정되어 나온 결과로, 철저한 사후 조사관리를 하여 구체적인 상황 등을 종합적으로 확인해야할 필요가 있다.

본 연구 결과를 바탕으로 사후 조사관리를 하여 얻어지는 조사업무 경험이 없는 신규직원이거나 해당 업무경험이 없는 경우, 지리적으로 먼 경우, 나이가 많은 경우, 장거리 출퇴근자인 경우 등과 같은 구체적인 지식은 다음 (그림 10)과 같이 조사관리 지식시스템에 집적하여 활용하여야 하겠다.



(그림 10) 효율적인 조사관리 지식 운영 프로세스

5. 결 론

농어촌 통계조사와 같이 조사항목이 대량이며 조사자가 면접조사를 토대로 구축되어지는 조사 데이터의 경우는 조사의 착오(허위, 탁상조사)를 시의 적절하게 탐지하거나 통제하기가 어려운 것이 현실이다. 특히 조사 자료를 토대로 조사자를 바로 역 추적이 가능하므로, '111…111'과 같은 단순 허위패턴이 발생할 가능성은 거의 없는 반면, 조사자의 개인 상황이나 지리적 위치나 비용에 따라 탐지가 어려운 지능적인 허위조사의 가능성은 존재할 수 있다.

따라서 본 논문에서는 조사자료 데이터베이스의 품질을 높이기 위해 지능적인 허위응답이나 허위조사의 잠재적 가능성 집단을 데이터 마이닝을 활용하여 탐지함으로써, 이 집단을 우선적으로 조사관리 하도록 의사결정을 지원하여 통계품질을 꾀하였다. 즉, 다양한 분석을 통하여 허위조사를 정확히 탐지해내고 허위조사 가능이 있는 분류군을 제시함으로써, 조사관리의 효율성을 높였으며 업무태만 등에 대한 예방적 효과도 기대할 수 있게 하였다.

향후 연구로 다른 데이터 마이닝 기법이나 품질관리 기법들의 응용을 통해서 보다 지능적이고 고의적이거나 실수로 인한 허위조사자나 허위응답자의 탐지를 위한 다양한 사례 분석 방법론을 개발해야 할 것이다.

참 고 문 헌

[1] Paul P. B., Robert M. G, Lars E. L., Nancy A. M and Seymour Sudman, "Measurement Errors in Surveys", JOHN WILEY & SONS, INC., 1991.

[2] 박영택, '공공행정부문 Single PPM 품질혁신', Single PPM 품질혁신추진본부, 2000.

[3] 박광진, 유황빈, "데이터마이닝 기법을 이용한 비정상행위 탐지 방법 연구", 정보보호학회논문지, 제13권 제2호, pp.99-106, 2003.

[4] 박정호, 오상현, 이원석, "데이터베이스 시스템에서 연관 규칙 탐사 기법을 이용한 비정상 행위 탐지", 정보처리학회논문지C 제9-C 제6호, pp.831-840, 2002.

[5] Belhadji E.B., G. dionne and F. Tarkhani, "A Model for the Detection of Insurance Fraud," The Geneva Papers on Risk and Insurance, Vol.25, No.4, pp.517-538, 2000.

[6] Danzon, Patricia, "The Frequency and Severoty of Medical Malpractice Claims," Journal of Law and Economics, Vol. 27, pp.116-142, 1984.

[7] 김영태, "사원부정의 특성에 관한 상호관련성 분석", 충청회계학연구 제2권 제1호, pp204-217, 1995. 12.

변 루 나

e-mail : lnbyon@hanmail.net

1992년 충북대학교 컴퓨터공학과(석사)

2004년 충북대학교 통계학과 전산통계 (박사)

2000년~현재 통계청 조사관리과

관심분야 : 전산통계, 데이터마이닝, 전자상거래

한 정 혜

e-mail : hanjh@cje.ac.kr

1998년 충북대학교 전자계산학과(박사)

1998년~1999년 연세대학교 산업시스템 공학과 포닥연구원

1999년~2001년 행정자치부 국가전문행정연수원 통계연수부 전산교육 전임교수

2001년~현재 청주교육대학교 컴퓨터교육과 조교수

관심분야 : 멀티미디어, 인간과 로봇 상호작용, 데이터마이닝