

데이터마이닝을 이용한 說問調査 및 分析

박만희* · 채화성** · 신완선***

*(주)CS2B · **(주)UBCare · ***성균관대학교 시스템경영공학부

Questionnaire Survey and Analysis Using Data Mining

Man-Hee Park* · Hwa-Sung Chae** · Wan-Seon Shin***

*CS2B · **UBCare · ***School of Systems Management Engineering, Sungkyunkwan University

Today's database system needs to collect huge amount of questionnaire that results from development of the information technology by the internet, so it has to be administrable. However, there are many difficulties concerned with finding analytic data or useful information in the high capacity-database. Data mining can solve these problems and utilize the database.

Questionnaire analysis that uses data mining has drawn relevant patterns that did not look or was tended to overlook before. These patterns can be applied by a new business rule. The purpose of this research is to analyze the questionnaire results and to present the result that can help to make decision easily with data mining. Recognition and analysis about these techniques of data mining show suitable type of questionnaire survey. This research focus on the form of present composition and the model of suitable questionnaire to analyze the type of it. Also, the comparison between the actual questionnaire result and the conventional statistical analysis is examined.

Keywords : questionnaire survey, data mining, decision tree analysis

1. 서 론

최근 정보기술의 발달로 인터넷을 통한 대량의 설문 결과를 수집하여 데이터베이스에 저장하고 관리할 수 있는 환경이 조성되었다. 그러나 대용량의 데이터베이스로부터 분석적인 데이터나 유용한 정보를 찾아내는 데는 여전히 많은 어려움이 따르고 있다. 이와 같은 문제를 해결할 수 있는 기법 가운데 하나인 데이터마이닝이 많은 분야에서 활용되고 있다. 데이터마이닝은 이전에는 보이지 않거나 간과되는 경향이 있었던 의미 있는 패턴들을 여러 가지 기법들을 이용하여 찾아내는 프로세스이다[12]. 또한, 데이터마이닝은 분석결과를 새로운 비즈니스 규칙으로 적용할 수 있는 예측 모델을 유추하는 데도 이용 가능하다.

본 연구는 설문조사로부터 수집된 데이터를 데이터마

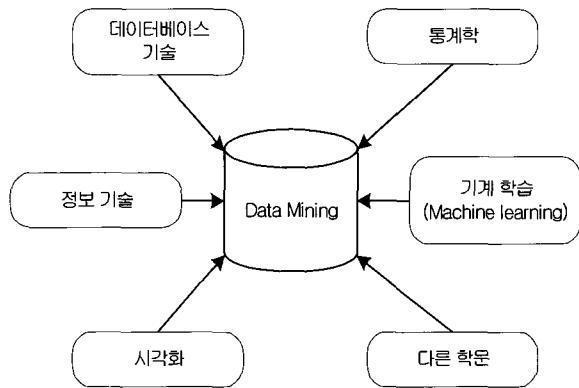
이닝 기법을 적용하여 유용한 분석결과를 도출할 수 있는 방법을 제시하는데 초점을 맞춘다. 여기서 요구되는 방법은 설문조사 결과데이터의 형태와 분석유형을 정확하게 파악하고, 체계적이고 신속하게 결과를 분석하는 것이다. 이를 통하여 설문조사의 적용분야를 확산시키고 전략적인 의사결정 도구로서 역할을 할 수 있다고 판단된다.

본 연구의 목적은 설문조사결과에 데이터마이닝 기법을 적용하여 의사결정에 도움을 줄 수 있는 설문조사의 방법을 제시하는 것이다. 이를 위해 첫째로 데이터마이닝의 기법 및 분석유형에 대하여 알아본다. 둘째, 분석유형에 적합한 설문의 구성 형태 및 모형을 도출한다. 셋째, 실제 조사된 설문결과를 분석유형에 맞추어 분석하고 기존의 통계분석과 비교하여 어떠한 변화가 있는지 살펴본다.

2. 데이터마이닝

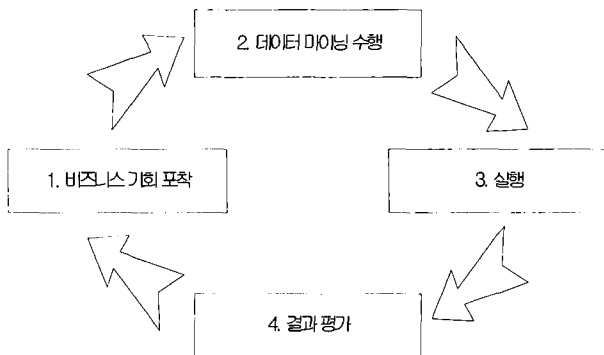
통계학자들은 오랜 동안 통계적으로 의미있는 패턴을 찾아내기 위해 다양한 방법으로 데이터베이스를 분석하여 왔다. 그러나 통계적 분석은 데이터를 요약하는 것에 관련된 것으로 의사결정에 도움을 주는 정보에는 한계가 있었다. 따라서 미처 발견하지 못한 유용한 정보를 찾아내기 위한 강력한 도구가 필요했으며 이러한 과정에서 대량의 데이터로부터 새롭고 의미있는 정보를 추출하여 의사결정에 활용할 수 있는 데이터마이닝이 부각되기 시작했다.

이러한 데이터마이닝은 <그림 1>에서 구성된 바와 같이 데이터베이스기술, 통계학, 정보기술, 기계학습 등 다양한 학문과 연관되어 있으며 데이터로부터 정보를 추출하기 위한 다양한 데이터마이닝 기법을 바탕으로 구성되어 있다.



<그림 24> 다양한 학문의 영향을 받는 데이터마이닝

데이터마이닝은 수억 혹은 수조 바이트의 데이터에 숨어있는 패턴을 발견하는 것이다. 이러한 패턴에 반응해야 하고, 그것을 바탕으로 행동하여야 하며, 궁극적으로는 데이터를 정보로, 정보를 행동으로, 행동을 가치로 변화시켜야만 하는 것이다.



<그림 2> 데이터마이닝의 효과적인 순환구조

이것이 간략한 데이터마이닝의 효과적인 순환과정이다. 데이터마이닝의 효과적인 순환과정은 고객, 시장, 제품에 대한 이해가 증대된 것으로부터 얻게된 지식을 내부의 프로세스에 적용하는 하나의 과정임을 인식하는 것이다. 효과적인 데이터마이닝의 순환구조는 <그림 2>와 같이 크게 비즈니스 기회의 포착, 데이터마이닝의 수행, 실행, 결과평가 등의 4단계로 나뉘어 진다.

데이터마이닝 기법의 알고리즘은 오래 전부터 연구되어 왔다. 과거에는 컴퓨터기술의 한계로 사용되지 못했던 기술들이 현재에 와서 정보기술의 발달과 여러 가지 환경요소의 만족으로 보편화되어 사용되고 있다. 현재 보편적으로 사용하는 기법들 외에도 다수의 새로운 기법들을 적용하고 비교·평가하는 문헌들이 계속 발표되고 있으며 이를 지원하는 도구들도 개발되어 현업에 적용되고 있다. 현재 많이 사용되고 있는 데이터마이닝 기법들의 평가하면 <표 1>과 같으며 여기서 통계란 회귀분석, 상관분석, 판별분석 등과 같은 전통적인 통계기법을 의미한다[11].

<표 1> 데이터마이닝 기법 평가

기 법	고려사항 (A : 좋음, B : 보통, C : 나쁨)				
	(1)	(2)	(3)	(4)	(5)
통 계	B	B	B	B	A
동시발생매트릭스	A	A	B	C	B
K-평균군집화	B	B	B	B	A
의사결정나무	A	B	A	B	B
신 경 망	C	B	A	A	A

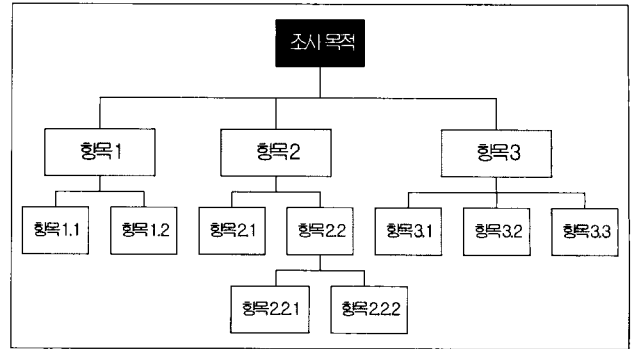
데이터마이닝 기법을 선택하는데 있어 적용한 다섯 가지 고려사항을 간략하게 설명하면 다음과 같다.

- (1) 모형의 설명력 : 모형이 데이터마이닝 결과를 설명하는 능력
- (2) 모형구축의 효율성 : 선택된 기법을 이용하여 모형을 구축하는 데 소요되는 시간과 노력
- (3) 모형의 정확성 : 결과의 신뢰도
- (4) 기법의 보편성 : 다양한 데이터마이닝 작업이나 데이터의 성격에 관계없이 적용할 수 있는 정도
- (5) 기법의 가용성 : 상용화된 S/W 제품의 종류 및 이들이 지원하는 O/S와 하드웨어 플랫폼의 다양성

어떤 데이터마이닝 기법을 적용할 것인가는 어떠한 결론을 도출할 것인가에 대한 문제이므로 각각의 기법이 가진 장·단점을 충분히 고려하여 기법을 선정하여야 한다.

데이터마이닝은 여러 가지 면에서 순수 통계와는 다

르지만 데이터마이닝과 통계간의 중요한 차이점 가운데 하나는 질의의 방향성이다. 우선, 데이터마이닝은 이전에는 보이지 않거나 그렇지 않으면 간과되는 경향이 있던 의미있는 패턴들을 발견하도록 데이터로부터 이끌어내는 프로세스이다. 반면 통계적인 추론은 사람이 가설을 수립하여 그 명제를 증명 또는 반증하기 위하여 통계적 방법을 적용하는 것이다. 또한, 데이터마이닝은 그 목적이 새로운 비즈니스 규칙으로 쉽게 번역될 수 있는 예측 모델을 끌어내는 것이라는 점에서 통계와 다르다.



<그림 4> 설문조사 목적설정과 항목도출

3. 데이터마이닝을 이용한 설문조사 방법

데이터마이닝을 이용한 설문조사 방법의 단계는 아래 <그림 3>과 같이 표현할 수 있다. <그림 3>은 기존의 설문조사 단계와 크게 다르지 않지만 데이터마이닝을 이용하기 위해서 조정되어할 단계별 방법이나 주의사항에는 차이가 있다.



<그림 3> 설문조사의 단계

3.1 목적설정

어떤 목적으로 설문조사를 수행하는지, 수집된 데이터로부터 무엇을 얻을 것인가에 대한 계획을 바탕으로 목적을 설정하여야 한다. 조사목적의 하위나무에 항목을 나열하고 더 세분화된 항목을 다시 아래의 나무에 나열한다. 나무구조로 항목을 표현하면 사전조사를 할 때 혹은 본 조사를 할 때 설문조사 주체자와 조사 대상자에게 전반적인 구조에 대한 이해를 돕는데 도움이 된다. 또한 분석과정에서 분석되어야 할 내용을 손쉽게 정리할 수 있다.

3.2 모집단 정의

조사목적이 어느 특정집단에 국한된 것이라면 모집단을 국한된 집단내부에서 정의하고, 여러 집단에 대한 영향을 도출하기 위한 것이라면 집단의 규모 및 특성에 따라 세부항목으로 분류하여야 한다. 대량의 결과를 수집할 수 있는 상황에서는 집단에 대한 범주를 넓혀 조사를 실시하면 세부 집단별로 유용한 결과를 도출할 수 있다. 또한, 세부집단별로 목적에 어떤 영향을 미치는지 파악하는데 용이하다.

3.3 조사방법 선정

설문조사방법 선정 시 이용할 수 있는 선정기준으로는 조사비용, 조사결과 수량, 조사횟수, 조사시간 등이 있다. 그리고 각각의 설문조사방법에 따른 선정기준별 속성의 특성값을 정리하면 <표 2>와 같다. <표 2>에서 (1)항목은 개별면접조사, (2)항목은 전화조사, (3)항목은 우편조사, (4)항목은 인터넷 조사방법을 나타낸다.

<표 2> 설문조사방법의 속성별 비교

기 준	(1)	(2)	(3)	(4)
① 비용	높음	보통	낮음	매우낮음
② 자료의질 -응답률 -응답자 응답동기 -조사원 편의	높음 높음 보통	보통 높음 낮음	낮음 낮음 없음	낮음 낮음 없음
③ 표본의 질	높음	보통	낮음	낮음
④ 응답내용 - 확인가능성	높음	높음	없음	없음
⑤ 시각적 자료 - 사용가능성	높음	없음	낮음	매우높음
⑥ 속도(시간요소)	낮음	높음	낮음	매우높음
⑦ 조사원 감독의 용이성	낮음	높음	없음	없음
⑧ 익명성	낮음	낮음	높음	높음
⑨ 컴퓨터 활용도 가능	보통	없음	없음	높음
⑩ 응답자의 독해능력 의존성	없음	없음	높음	높음

3.4 표본설계

표본설계는 아래 다섯 가지 사항을 검토하여 조사에 적합한 표본을 설계한다.

- (1) 표본오차는 계산 가능하며 최소화 할 수 있는가?
- (2) 추정량과 그에 대한 분산은 직접 계산 가능한가?
- (3) 가용예산과 인원의 제약조건 하에서 요구정도를 달성할 수 있는가?
- (4) 표본조사를 실행하는 과정은 용이한가?
- (5) 유사한 조사방법을 과거에 실행한 적은 있는가?

3.5 설문지 설계

설문지가 어떻게 설계되는가에 따라 분석기법의 적용 및 결과도출에 영향을 미치게 되므로 조사목적에 맞추어 필요한 정보, 자료의 종류와 측정방법, 분석의 내용과 방법까지 고려하여 신중히 작성되어야 한다. 미리 설문지 작성의 단계를 정해놓고, 단계별로 작성하면서 문제점을 파악하여 적절한 수정과 보완이 이루어져야 한다. 일반적인 설문지 작성과정을 요약하면 다음과 같다.

- ① 필요한 정보의 결정
- ② 개별항목의 내용결정
- ③ 문항형태의 결정
- ④ 개별문항의 완성
- ⑤ 질문의 수와 순서결정
- ⑥ 설문지의 외형결정
- ⑦ 설문지의 사전조사
- ⑧ 설문지의 완성

3.6 조 사

설문지 준비가 완료되면 선택한 조사방법을 이용하여 조사를 실시한다. 조사는 일반적으로 예비조사, 사전조사, 본 조사의 3단계로 구분할 수 있다.

- (1) **예비조사**
 - 효과적인 표본설계를 위해서 관심변수의 분산에 대한 사전정보와 실제 조사의 조사비용에 대한 정보를 얻기 위해서 실시한다.
- (2) **사전조사**
 - 적은 수의 표본을 대상으로 우선 작성된 설문지를 사용하여 샘플조사를 한다.
- (3) **본 조사**
 - 사전조사결과를 바탕으로 설문 내용을 수정하여 설문지를 완성하고 실질적인 자료수집을 실시한다.

3.7 자료준비

자료수집이 완료되면 분석을 실시하는 데 필요한 자료를 정리하여 분석에 적합하도록 사전 처리 작업을 수행하여야 한다. 설문지 설계단계에서 예상치 못한 분석 내용이 추가되거나 자동화된 시스템을 이용하지 않는다면 분석을 위한 자료준비 단계가 반드시 필요하다. 데이터마이닝을 위한 자료준비는 다음의 3가지 고려사항을 준수하여 준비하여야 한다.

- (1) 데이터를 훈련용, 검토용, 평가용 집합으로 구분한다.
- (2) 분석에 이용할 데이터의 양
- (3) 분석을 위해 선택해야 할 변수의 수

3.8 분 석

데이터마이닝을 이용한 분석은 데이터로부터 유용한 정보를 도출하기 위해서 수행되어야 하기 때문에 데이터 변환의 횟수나 어려움을 최소화할 수 있는 데이터마이닝 기법이 선택되어야 한다. 분석에 이용되는 자료의 형태 및 특성에 따라 적용할 수 있는 데이터마이닝 기법은 달라지는데 이를 정리하면 <표 3>과 같다.

분석을 실시하기 위해서는 변수들을 정의하여야 하는데 조사목적, 항목, 문항별로 구성된 그림을 바탕으로 분석할 내용이 무엇인지 결정한다. 그리고 분석할 내용에 필요한 문항들을 선택한 후 문항별로 예측변수와 입력변수를 구분하고 변수구분이 완료되면 데이터마이닝 도구를 이용하여 분석을 실시한다. 분석기법별로 적용 가능한 분석영역에 차이가 있으므로 단일 기법을 넘어선 결과를 도출하기 위해서는 두 개 이상의 기법들을 조합한 분석도 수행할 필요가 있다.

<표 3> 데이터 형태별 설명 및 분석기법

특 성	설 명	분석기법
범주형 변수	- 제한되고 미리 설정된 값으로 집합 특성을 갖는 형태 - 예) 우편번호, 차량 모델, 결혼상태 등	• 시장바구니분석 • 의사결정나무
수치형 변수	- 수치로 설정된 값의 형태 - 예) 나이, 통화지연, 가격, 봉급, 고도, 기온 등	• 의사결정나무 • 신경망분석 • 군집분석 • 기억기반추론
레코드마다 많은 필드가 존재	- 예측변수에 대한 필드와 입력변수에 대한 필드의 숫자가 많은 경우	• 의사결정나무

<표 3> 데이터 형태별 설명 및 분석기법(계속)

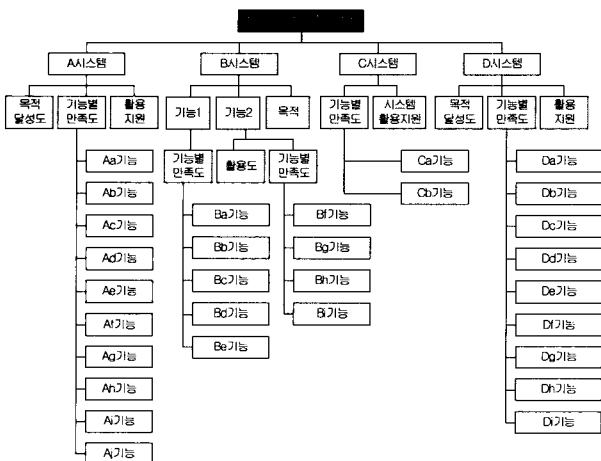
특 성	설 명	분석기법
다양한 종속 변수	- 동일한 입력변수에 따른 예측 변수의 수가 여러 가지인 경우	• 신경망분석
시간순 데이터	- 시간에 따라 변화하는 데이터	• 신경망분석
Free-Form Text	- 극단적인 변수 정보를 포함하는 형태 - 예) 부품번호, 노동시간 등	• 기억기반추론

4. 데이터마이닝을 이용한 설문조사 및 결과 분석 : A사 적용사례

본 설문조사는 고객만족분석시스템(CSAS)(14)을 이용하여 인터넷을 통하여 조사를 실시하고 결과를 분석하였다.

단계 1 : 목적설정

A사에서 직원들의 업무를 지원하기 위해서 사용되고 있는 4가지 정보시스템에 대한 만족도를 조사하고 만족도에 영향을 주는 항목을 도출해낸다. 도출된 항목을 바탕으로 시스템에 대한 평가를 실시하여 문제점을 도출하고 도출된 문제점의 개선방안을 제시하는 것이 조사의 목적이다. 본 연구에서는 조사목적과 관련하여 <그림 5>와 같이 관련 항목들을 도출하였다.



<그림 5> 조사목적에 관련된 항목도출

단계 2 : 모집단 정의

정보시스템을 사용하는 직원들을 대상으로 하는 조사이므로 모집단은 A사의 모든 직원이다.

단계 3 : 조사방법 선정

설문조사 방법은 인터넷을 통한 온라인 설문조사를

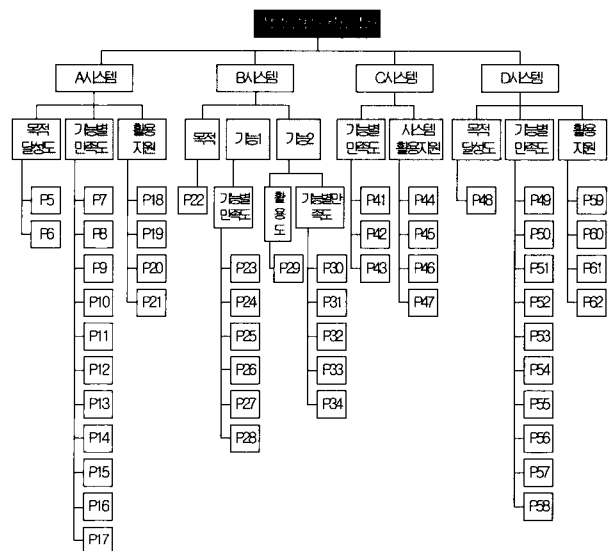
선정하였다.

단계 4 : 표본설계

표본은 A사에서 정보시스템을 사용하고 있는 국내 및 해외 직원을 대상으로 400명을 선정하였다.

단계 5 : 설문지 설계

설문문항들 중 개인신상과 관련된 4개 문항을 제외시킨 후 (단계 1)에서 결정한 항목들을 바탕으로 <그림 6>과 같이 문항들을 나무구조의 구성도로 작성하였다.



<그림 6> 항목에 관련된 문항설계

단계 6 : 조사

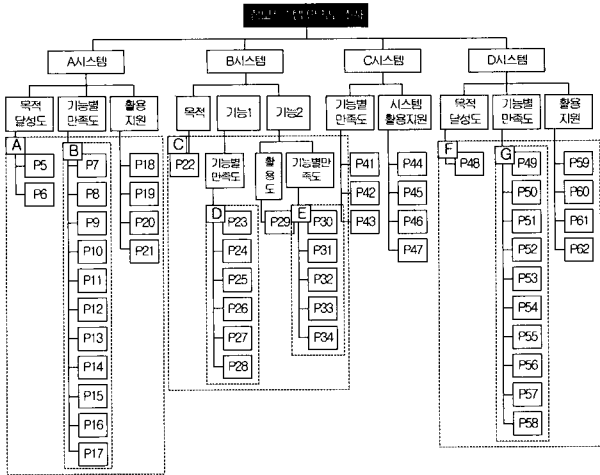
담당자와 충분한 검토를 거친 후 예비조사와 사전조사를 생략하고 본 조사를 실시하였으며 1주일 동안 국내 및 해외에서 인터넷을 통하여 자료를 수집하였다.

단계 7 : 자료준비

설문조사시스템에 수집된 데이터는 관계형 데이터베이스 내에 여러 개의 테이블에 나누어 저장되며 수집된 데이터는 통합된 하나의 테이블로 변형된 후 별도의 데이터베이스에 저장된다.

단계 8 : 분석

본 연구에서 수행한 설문조사의 목적이 영향력 있는 속성이나 항목들을 도출하는데 있으므로 <그림 7>과 같이 7개의 분석범주를 설정하였다. 설정한 7개 분석범주를 바탕으로 의사결정나무 분석기법을 이용하여 결과분석을 실시하였다. 분석도구로는 마이크로소프트 SQL Sever에서 제공하는 Analysis service를 이용하였다



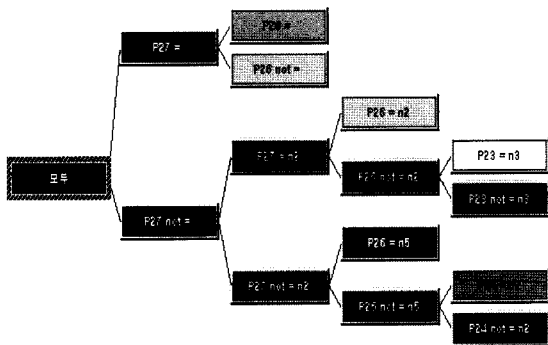
<그림 7> 분석범주 설정

4.1 분석결과

본 연구에서 이용한 분석기법을 바탕으로 수행한 결과분석에서 범주 D의 예측변수에 영향을 주는 입력변수로는 P23, P24, P26, P27로 나타났으며 각각의 변수에 대한 설명은 아래 <표 4>와 같다. 예측변수 P28의 전체 보기항목(n1, n2, n3, n4, n5)에 대한 의사결정나무를 그림으로 표현하면 <그림 8>과 같다.

<표 4> 범주D의 변수 설명

변수명	문항번호	문항설명	보기항목
P23	문항23	Ba기능	매우만족=n1 다소만족=n2 보통=n3 다소불만=n4 매우불만=n5
P24	문항24	Bb기능	"
P26	문항26	Bd기능	"
P27	문항27	Be기능	"
P28	문항28	B시스템 기능I에 대한 전반적인 만족도	"



<그림 8> 예측변수 P28의 보기전체에 대한 의사결정나무

본 연구에서 실시한 분석결과를 바탕으로 범주별로 예측변수에 영향을 주는 입력변수를 정리하면 아래 <표 5>와 같다. 입력변수가 많을 때 본 연구에서 이용한 의사결정나무를 이용하면 예측변수에 영향을 주는 특정 입력변수를 도출할 수 있으며, 도출된 변수를 이용하여 신경망이나 회귀모형 분석을 수행할 수 있다.

<표 5> 예측변수에 영향을 주는 입력변수

범 주	예측변수	영향이 있는 입력변수
A	P6	P8
B	P17	P8
C	P22	P31
D	P28	P23, P24, P26, P27
E	P34	P33, P31
F	P48	P49
G	P58	P50, P52

4.2 통계분석과 결과비교

<표 6> Top-2/Bottom-2 포트폴리오 분석결과

시스템	결 과	
A시스템 (범주 B)	강 점	• Ac기능(P9) • Ad기능(P10) • Ab기능(P8)
	개선점	• Af기능(P12) • Ai기능(P15)
B시스템 기능1 (범주 D)	강 점	• Ba기능(P23) • Bb기능(P24) • Be기능(P27)
	개선점	• Bd기능(P26)
B시스템 기능2 (범주 E)	강 점	• Bf기능(P30) • Bg기능(P31)
	개선점	• Bh기능(P33) • Bg기능(P32)
D시스템 (범주 G)	강 점	• Db기능(P50) • Dc기능(P51)
	개선점	• De기능(P53) • Dh기능(P56)

의사결정나무분석 이외에도 각각의 시스템에 연관성이 있는 기능들을 도출하기 위해서 Top-2/Bottom-2 포트폴리오 분석을 실시하였다. Top-2/Bottom-2 포트폴리오 분석결과는 <표 6>과 같으며, Top-2/Bottom-2 분석결과와 의사결정나무 분석결과와의 비교는 <표 7>과 같다

<표7> Top-2/Bottom-2 분석결과와 의사결정나무 분석결과 비교

시스템 (범주)	Top-2/Bottom-2 분석에서 도출된 기능	의사결정나무분석에서 도출된 기능
A 시스템 (범주 B)	P8, P9, P10, P12, P15	P8
B시스템 기능1 (범주 D)	P23, P24, P26, P27	P23, P24, P26, P27
B시스템 기능2 (범주 E)	P30, P31, P32, P33	P31, P33
D시스템 (범주 G)	P50, P51, P53, P56	P50, P52

위에서 이용한 두 기법간의 차이점을 설명하면 Top-2/Bottom-2 분석은 시스템의 전반적인 만족도를 고려하지 않고 개개의 기능만을 이용하여 분류한 것이고, 의사결정나무는 시스템의 전반적인 만족도를 고려하여 이와 연관성이 있는 항목들을 분류했다는 것이다. 또한 Top-2/Bottom-2는 강점과 개선점을 각각의 기능별로 나타내는 분류를 하였지만 의사결정나무는 전반적인 만족도와 관련된 모든 기능들 중에서 영향력이 있는 기능들에 대한 패턴을 분류하였다. 분류된 패턴을 바탕으로 미래에 대한 예측을 수행할 수 있고 개선방향을 제시할 수 있다.

5. 결론

본 연구에서는 다양한 설문조사 유형에 대해서 데이터마이닝을 활용할 수 있는 단계적 방법을 제시하였다. 설문조사 목적을 설정하는 단계부터 ‘모집단정의’, ‘조사방법 선정’, ‘표본설정’, ‘설문지 설계’, ‘조사’, ‘자료준비’, ‘분석’의 단계에 이르기까지 데이터마이닝을 활용하기 위해 필요한 사항이나 주의해야 할 사항을 제시하였다. 또한, 본 연구에서 제시한 단계적 방법을 이용하여 A사 정보시스템 만족도 설문조사에 적용해 보았다. 분석단계에서 의사결정나무를 이용하여 분석한 결과에서는 개별 시스템별로 전반적인 만족도에 대해 영향을 주는 기능들의 패턴을 분류할 수 있었다. 그리고 통계적 분석방법과 비교한 분석에서 보다 유용한 결과를 도출할 수 있었다.

본 연구에서 제시한 설문조사방법은 4가지 관점에서 효율성 및 유용성을 높일 수 있다. 첫째, 나무구조의 그림을 이용하여 설문조사 목적, 목적과 관련된 항목, 분석할 범주 등을 쉽게 분류할 수 있다. 둘째 초기단계부터 수집자료에 대한 명확한 정의를 통하여 자료준비단계에서 불필요한 노력을 줄일 수 있다. 셋째, 예측변수와 상호 연관성을 가지고 있는 패턴들을 분류할 수 있다. 또

한 분류된 패턴을 바탕으로 미래를 예측하고 문제점을 개선할 수 있는 방법을 모색할 수 있다. 넷째, 반복적으로 설문조사를 수행할 때 효과적인 순환구조를 가지고 있어 지속적인 관리가 가능하며 시간적 변화에 대응하여 항상 새로운 패턴을 발견해낼 수 있다.

본 연구와 관련된 향후 과제로는 첫째, 다양한 분야의 실제 사례에 대해 보다 다양한 데이터마이닝 기법을 적용한 설문조사 분석이 수행되어야 하며, 이를 바탕으로 설문조사의 유형에 따른 가장 효율적인 기법이 제시되어야 하겠다. 둘째, 설문조사의 모든 단계를 자동화하여 분석결과를 제시할 수 있는 시스템을 구현하고 데이터마이닝을 적용한 설문조사시스템을 범용화하는 것이다.

참고문헌

- [1] Abraham Kandel, Mark Last and Horst Bunke, *Data Mining and Computational Intelligence*, Physica Verlag Heidelberg, 2001.
- [2] Dorian Pyle, *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.
- [3] Ian H. Witten and Eibe Frank, *Data Mining*, Morgan Kaufmann Publishers, 2000.
- [4] Krzysztof J. Cios, Witold Pedrycz and Roman W. Swiniarski, *Data Mining Methods for Knowledge Discovery*, Kluwer Academic Publishers, 1998.
- [5] Jesus Mena, *Data Mining Your Website*, Digital Press, 1999.
- [6] Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 2001
- [7] Olivia Parr Rud, *Data Mining Cookbook*, John Wiley & Sons, Inc, 2001
- [8] Ron Kohavi and Foster Provost, *Applications of Data Mining to Electronic Commerce*, Kluwer Academic Publishers, 2001.
- [9] Sholom M. Weiss, Nitin Indurkha, *Predictive Data Mining-A Practical Guide*, Morgan Kaufmann Publishers, 1998.
- [10] 권오주, 이정무, *Microsoft SQL Server 2000 데이터마이닝*, 정보문화사, 2001.
- [11] 장남식, 홍성완, 장재호, *데이터마이닝*, 대청, 2000.
- [12] 허준, 최병주, *클레멘타인을 이용한 데이터마이닝*, 고려정보산업, 2001.
- [13] 홍성완 외, *CRM을 위한 데이터마이닝*, 대청, 2000.
- [14] 노종찬, “고객만족 설문조사 및 분석을 위한 컴퓨터 지원 시스템의 개발”, 석사학위논문, 성균관대학교, pp 29-45, 1999.
- [15] 인터넷사이트, <http://www.metagroup.com>