

통계학 교과목으로서의 “데이터마이닝”

한상태1) · 최종후2)

<요약>

최근 각 대학의 통계학과(또는 통계전공)에서는 학부제에 따른 대학교육시스템의 변화에 따라, 새로운 교육체계에 적응하기 위한 여러 각도의 노력을 기울이고 있는 것 같다. 이들 노력 중에 대표적인 것은 학과명칭(또는 전공명칭)의 변경과 새로운 교과과정의 개발이 주된 관심사라 여겨진다. 본 소고에서는 새로운 교과과정의 개발에 있어 많은 대학이 관심을 갖고 있고, 또한 실제적으로 이미 교과과정에 포함시키기로 결정한 “데이터마이닝” 교과목에 대하여, 과연 이 교과목이 통계학 전공자들에게 왜 필요한 것인지의 문제와 만약 필요하다면, 어떠한 내용을 강의해야 하는지를 논의해 보고자 한다.

1. 서론

통계학과의 교과과정에 대한 연구는 통계학회가 설립된 이후 사회의 환경적 변화가 있을 때마다 꾸준히 논의되어 왔다. 특히 컴퓨터의 보급이 현실화 된 80년대 말부터 늘 우리에게 들려오는 소리는 실제적인 문제에 활용할 수 있는 응용관련 통계를 강조해야 한다는(모든 통계학자가 다 그런 것은 아니었지만) 것이었다. 이런 관점에서 많은 대학들이 응용통계 관련 교육을 강화하고자 노력하였고, 이로 말미암아 아직도 타대학과 비교하여 상대적으로 이론적인 통계에 대한 비중이 높은 통계학과는 현 시점에서 보다 큰 위기감을 갖고 있는 것 같다. 결과적으로, 응용통계의 확장에 애쓰신 여러 교수님들의 덕분에 지금은 통계 소프트웨어를 활용한 데이터 분석능력(응용통계의 강화에서 드러난 가장 큰 모습)은 대단히 많이 길러지게 되었다.

‘만족한다는 모습보다는 불안하다고 생각하는 모습이 보다 발전된 모습이라고 하던가?’

최근 들어, 한경수·최숙희(1999)의 ‘정보사회에서의 통계학과 : 위기인가? 기회인가?’의 글을 시작으로, 허명희(1999)의 ‘통계학, 새로운 모습의 탐색’ 및 송재기(1999)의 ‘통계학의 새로운 발전을 기대하며’ 등의 토론 내용이 발표되었다. 또한 이런 시각의 연장선상에서 99년 4월 대전대에서는 정보화 시대를 맞이하여 이에 대응할 수 있는 통계학과의 교과과정을 제안하고 의견교환을 목적으로 통계교육상담연구회 발표회가 있었다. 발표주제는 조신섭(1999)의 ‘정보 관련 통계학과의 교과과정에 대한 제안’을 포함한 5가지였고, 이들 발표주제에 대한 토론이 이어졌다. 이때 발표의 주요 내용은 “정보” 통계와 관련된 강의 경험을 소개하는 것이었는데, 이를 통해 통계관련 학과의 교과과정을 제안하는데 도움이 되길 기대하는 바램도 담겨 있었다. 최근에 이런 움직임들은 예전의 변화모습보다는 훨씬 더 강한 것이었고, 또한 그 어느 때보다 각 대학들의 관심이 컸다고 생각된다.

이와 같은 통계학계의 움직임 속에서 많은 이들의 관심을 끌고 있는 것 중의 하나는 데이터 마이닝(Data Mining)에 관한 것이었다. 실제적으로, 조신섭 교수가 제안한 정보통계 교과과정에는 4학년 2학기에 ‘데이터마이닝’ 과목을 개설할 것을 권고하였고, 대구효성카톨릭대의 경우 1999년 입학한 학생들의 교과과정에도 4학년 2학기에 ‘데이터마이닝과 통계학’이란 교과목이 포

1) (336-795) 충남 아산시 배방면 세출리 산 29-1, 호서대학교 수학과 교수

2) (339-700) 충남 연기군 조치원읍, 고려대학교 정보통계학과 교수 .

함되어 있었다(이상복 교수의 발표내용). 또한 한경수 교수도 통계학과의 데이터베이스 교육을 강조하면서, 데이터마이닝이 현 업계나 학계에서 주목을 받고 있다고 표현하였다. 마지막으로 방통대의 이기재 교수가 소개한 방통대의 교과과정에도 3학년 1학기 교과목으로 '데이터마이닝'을 포함시켰다. 이 이외에도 비공식적으로 접하는 많은 이야기가 대부분 대학의 교과과정 개정에 있어서 데이터마이닝은 크게 거론되고 있다는 것이 학계의 움직임이었다.

학계의 움직임보다는 훨씬 먼저 업계에서는 최근 몇 년간 데이터웨어하우징(Data Warehousing)으로 대변되는 대규모 데이터베이스 구축이 진행되어 왔고, 지금은 충분히 축적된 데이터들을 활용하는 것에 큰 관심이 집중하게 되었다. 사실, 데이터웨어하우스를 구축하지 않으면 정보화 사회에서 기업의 경쟁력과 장래는 없다는 시각에서 데이터베이스 구축에 막대한 예산을 투자하였음에도 불구하고, 실제로는 데이터베이스를 활용하여 기업의 이익을 얻는 일은 제대로 이루어지지 못하였다. 왜냐하면 구축된 데이터베이스를 제대로 활용할 수 있는 준비가 충분하지 못하였기 때문이다. 이런 시점에서, 구축된 데이터베이스를 활용할 수 있는 데이터마이닝 제품들(E-Miner/SAS, Clementine/SPSS, I-Miner/IBM 등)이 출시되었고, 구축된 데이터베이스 활용에 커다란 관심을 둔 기업의 수요가 맞물려 커다란 이슈가 되고 있는 것이다.

이와 같은 학계와 업계의 움직임을 고려하여, 본 논고에서는 데이터마이닝이 통계학과 교과목으로서 왜 필요한지의 문제와 더불어 어떤 교육내용을 담으면 바람직하겠는가를 토의해 보도록 하겠다.

2. 통계학에서 데이터마이닝 교육의 필요성

데이터마이닝은 다소 새로운 분야이고 지금도 그 개념이 발전되어 가는 과정에 있기 때문에 한 마디로 데이터마이닝을 정의한다는 것은 매우 어려운 일이다. 사실, 데이터마이닝은 David(1998) 등이 논의한 바와 같이 통계학, 데이터베이스 테크놀로지(Database Technology), 패턴인식(Pattern Recognition), 기계학습(Machine Learning) 등 매우 다양한 학문과 관련되어 있기 때문이다. 그러나, 많은 연구자들의 공통된 견해를 빌어서 데이터마이닝을 간단히 정의해보면 "대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 함으로써 유용한 지식을 추출하는 일련의 과정들"이라 할 수 있다. 이를 통계적인 관점에서는 Friedman(1997)이 말했듯이 "대용량 데이터에 대한 탐색적 데이터 분석(Exploratory Data Analysis)"이라고 할 수도 있을 것이다. 따라서 통계학 전공자들에게 있어 완전히 새로운 개념은 아닌 것이다. 데이터마이닝에 관한 자세한 내용은 강현철·한상태·최중후 외(1999), 최중후·한상태·강현철(1999) 및 최중후·한상태 외(1998)를 참조하길 바란다.

그러나, 데이터마이닝은 몇 가지 관점에서 기존의 통계학에서 다룬 통계분석과 다른 측면들을 가지고 있는데, 이를 데이터마이닝 통계교육의 필요성과 관련지어 서술하면 다음과 같다.

1) 데이터마이닝의 가장 큰 특징은 대용량의 데이터를 다룬다는 것이다.

실제로 데이터는 유례를 찾을 수 없이 급격히 증가하고 있으며, 수백 기가 바이트의 데이터베이스도 이제 흔한 현상이라 할 수 있다(예를 들면, 미국도매상 Wal Mart의 일일 거래건수 2천만 건 등). 이에 대해 전통적으로 훈련된 통계학자들은 이러한 크기의 데이터를 처리하는 것에 익숙하지 못하다는 현실이 있다. 따라서, 현업에서 요구하고 있는 대용량의 데이터를 다룰 수 있는 능력을 길러주기 위한 교육이 반드시 필요하다는 것이다.

2) 데이터마이닝은 대규모의 관측자료를 다룬다는 것이다.

지금까지의 통계교육에서 다룬 실험자료는 연구를 목적으로 가설검정 등의 구체적인 문제에 답하기 위해서 여러 요인들이 통제된 상황에서 생산되고, 어느 정도는 정제되어 있으며, 정적인

통계학 교과목으로서의 "데이터마이닝"

상태에 있는 현재 자료였다. 그러나 관측자료는 업무활용(상업적 가치가 기준임)을 목적으로, 시간의 흐름에 따라서 비계획적으로 축적되며, 자료분석을 염두에 두고 수집되지 않는 것이 일반적이므로, 정제되어 있지 못하며 동적인 특징을 가지고 있다. 따라서, 기존의 통계교육에서 잘 다루지 않았던, 현장에서의 관측자료에 대한 다양한 분석능력을 길러 줄 필요가 있다는 것이다. 이러한 교육을 통해 학생들은 사회에 진출하기에 앞서 사전에 충분한 적응능력이 길러질 수 있다고 본다.

3) 데이터마이닝은 컴퓨터 중심적 기법(computer-intensive method)이다.

현대의 컴퓨터 중심적 기법(통계학에서 예를 들면, 붓스트랩 등)들은 기존의 기법들로서는 해결하기 곤란한 경우에 있어, 이를 해결하기 위하여 컴퓨터의 강력한 처리속도와 능력을 활용할 수 있도록 해주고 있다. 이런 관점에서, 본 저자들은 허명희(1999)의 주장과 같이 통계교육은 통계학의 정통성은 지키되 컴퓨터의 적극적인 활용을 탐색할 수 있는 교육이 필요하다는 데 전적으로 동의하며, 이를 데이터마이닝 교육에서는 가능케 한다고 생각한다.

4) 데이터마이닝은 경험적 방법(adhockery method)에 근거하고 있다.

많은 데이터마이닝 기법들은 이론적 원리에 기초하여 개발되었다기보다는 경험에 기초하여 개발(예를 들면, 연관성 규칙(Association Rule), 평가(Assessment) 등) 되었다. 이러한 기법들은 그 특성이 수리적으로 밝혀지지 않은 것들이 많다. 따라서, 데이터마이닝 교육을 통하여 통계교육에서의 분석영역 범위를 보다 넓일 필요가 있다는 것이다.

5) 데이터마이닝은 일반화(generalization)에 초점을 두고 있다.

여기서의 일반화는 예측모형이 새로운 자료(new data)에 얼마나 잘 적용되도록 하는 것인가를 의미한다. 따라서 일반화는 데이터마이닝 기법의 비정형성을 어느 정도 해결 또는 보완하여 주는 데 도움을 주고 있다. 추론(inference)이 통계모형들의 초점이라 하면, 일반화는 데이터마이닝의 초점이라 할 수 있다. 그러므로 모집단을 가정하고, 그 모집단에서 iid 표본을 뽑은 상황에서 생각하는 통계학의 기본 모습에 대하여 보다 다양한 통계적 사고능력을 길러 줄 수 있다는 것이다.

6) 데이터마이닝 기법들은 통계학, 컴퓨터과학, 인공지능, 공학 등과 같은 분야에서 개발된 특징을 가지고 있다.

최근 국내의 흐름을 보면, 데이터마이닝을 전산관련 전공자들은 DB 관점으로, 통계 전공자들은 통계분석의 영역으로, 경영학(특히 경영정보) 관련 전공자들은 데이터베이스 마케팅과 거의 동의어로 판단하면서 고객 데이터로부터 마케팅 정보를 이끌어낸다는 관점에서 서로의 영역인 것 같이 커다란 관심을 보이고 있다. 그러나, 실제적으로 데이터마이닝은 분석 part(통계 영역), 데이터 part(전산부서 영역), 적용 part(현업 part: 마케팅부서, 영업부서 등)의 세 분야 사람들이 협력하여 진행되는 것이 사실이다. 따라서, 사회적 요구에 부응하는 다양한 자료에 대한 분석 및 처리 능력을 길러주므로써, 관련 있는 타학문과의 연계성 속에서 통계학의 담당 역할을 새롭게 보여주어야 한다는 측면에서도 데이터마이닝 교육이 필요하다 여겨진다.

3. 데이터마이닝 강의 내용

(1) 교육목표

대용량의 데이터를 분석하기 위해서 필요한 관련 지식과 데이터마이닝의 개념 및 방법론을 익힌다. 또한 실제 데이터를 이용해 각 방법론을 실습하고, 직접 프로젝트를 수행해 봄으로써 현장에서 보여지는 다양한 데이터에 대한 응용력과 적용력을 기른다.

(2) 강의교재 및 참고서적

현재 데이터마이닝에 대하여 국내에서 개발된 교재는 거의 없는 실정이다. 따라서, 현재 우리가 손쉽게 구할 수 있는 교재를 중심으로 구성한다면 다음과 같다.

교재 : SAS Enterprise Miner를 이용한 데이터 마이닝 - 방법론 및 활용 -, 강현철 · 한상태 · 최중후 의 공저, 자유아카데미.

참고서적 :

- R1. SAS Enterprise Miner를 이용한 데이터 마이닝 - 기능과 사용법 -, 최중후 · 한상태 · 강현철 의 공저, 자유아카데미.
- R2. Data Mining Techniques for Marketing, Sales, and Customer Support, Michael J. Berry and Gordon Linoff, John Wiley & Sons, Inc.
- R3. Data Mining Using SAS Enterprise Miner Software, SAS Institute Inc.
- R4. Enterprise Miner Software: Applying Data Mining Techniques, SAS Institute Inc.

관련 웹 사이트를 참고하는 것도 많은 도움이 될 것이다. 그 중 몇 개를 소개하면 다음과 같다.

- W1. <http://www.sas.com>
- W2. <http://www.dni.co.kr>
- W3. <http://www.ibm.com>
- W4. <http://www.spss.com>
- W5. <http://www.data-miners.com>
- W6. <http://www.xore.com>
- W7. <http://www.statserv.com>
- W8. <http://www.datamininglab.com>

(3) 강의내용

1) 데이터마이닝의 정의, 출현배경, 활용분야, 특징, 관련분야, 데이터마이닝 프로세스에 대한 간단한 소개, 데이터마이닝 실제 적용사례(국외, 국내), 데이터마이닝 적용시 문제점 등을 소개하므로써 데이터마이닝에 대한 전반적인 내용을 이해시킨다.

2) 하나의 프로젝트를 진행하는 과정을 통해 데이터마이닝 소프트웨어(E-Miner/SAS를 저자들은 이용하고 있음)의 사용법 및 데이터마이닝의 과정을 알려준다. 이를 통해 데이터마이닝의 절차인 데이터의 생성 및 추출, 탐색 및 변형, 모형화, 평가 등의 과정과 소프트웨어 활용능력에 대한 감을 잡을 수 있으리라 여겨진다.

3) 데이터마이닝의 방법론으로서 우리에게 가장 친숙한 회귀분석(로지스틱 판별분석 포함), 데이터마이닝의 탐색과 모형화라는 두 가지 특성을 모두가지고 있는 의사결정나무분석, 뇌 신경생리학으로부터 영감을 얻어 시작한 신경망 분석(은닉마디(hidden node)라고 불리는 독특한 구성요소 의해서 일반적인 통계모형과 구별되고 있음) 등을 사례분석을 중심으로 교육시키므로써 이들 데이터마이닝 방법론들에 대한 활용능력을 길러준다.

4) 모형 평가의 기본 개념인 오분류표, 이익행렬, 수익 및 이익도표 등과 같은 모형평가와 관련되는 기본지식을 습득시키고 이에 대한 활용을 알아본다. 이를 통해 데이터마이닝에 적용한 여러 모형(예를 들어 회귀분석모형, 의사결정나무분석모형, 신경망모형 등) 들에 대한 상호 비교, 이를 통한 최적 모형의 결정 등을 깨닫게 해 준다.

5) 데이터의 모형화 단계이전에 필요한 데이터의 탐색과 변형(기초 통계, 이상치 및 결측치 확인, 이상치 제거, 관측치 대체, 변수변환을 통한 파생변수의 생성, 중요변수 선택 등의 과정 등) 방법을 알려준다. 또한 추가적으로, 데이터마이닝의 기법으로 군집분석을 어떻게 응용할 수 있는가를 소개해 준다. 이러한 내용의 교육을 통해 학생들에게는 모형구축의 효과를 높일 수 있는 다양한 사고와 능력이 길러질 것이라 여겨진다.

6) 데이터마이닝의 가장 일반적인 작업이라 할 수 있는 연관성 규칙(Association Rules)의 개념을 소개하고 더 나아가 실제적으로 가치 있는 연관성 규칙을 얻기 위한 방법과 사례분석을 통해 얻어낸 연관성을 어떻게 응용할 수 있는지를 알려준다.

(4) 학생들 참여방식

학생들에게 있어 중요한 것은 데이터마이닝에 대한 개념과 지식을 배우는 것일 수도 있다. 그러나 이보다 몇 배 더 중요하다고 생각하는 것은 현업에서 데이터마이닝을 실제로 적용할 수 있는 문제해결능력이 있어야 한다는 것이다. 이런 관점에서 학생들의 수업 참여방식은 다음의 몇 가지 내용을 권고하고 싶다.

1) 충분한 실습을 통해 컴퓨터 활용능력을 높이도록 유도한다.

2) 데이터마이닝 관련 웹 사이트에 있는 정보검색, 각 기업에서 활용하고 있는 데이터마이닝 관련 프로젝트의 현황 파악 등의 정보를 수집하여 정리케 함으로써 정보검색의 중요성과 현장에서의 프로젝트 진행방식 등을 알 수 있도록 유도한다.

3) 과제물을 4회 정도 부과한다. 이 때 처음 2회는 수업에서 소개된 자료를 가지고 배운 내용을 똑같이 따라하는 연습결과를 보고서로 제출케 하고(개별적 보고서), 나머지 2회는 조를 구성(한 조당 4명 정도)하여 프로젝트 형식의 과제를 보고케 한다(정보수집 관련 보고서, 기말 프로젝트 보고서). 이를 통해 팀워크와 커뮤니케이션 능력 향상을 유도한다.

4) 조별 프로젝트에 대한 발표와 토론과정을 갖는다. 이는 이미 다른 통계교과에서도 진행되고 있는 일이라 생각되는데, 이를 통해 듣기와 말하기 등의 능력을 배양시킬 수 있을 것이다.

5) 담당교수의 지도 속에 데이터마이닝 활용 사례집을 만든다(이는 학생들이 느끼는 노력의 결실이라는 뿌듯함과 타대학과의 정보 공유에도 커다란 도움이 될 것이다).

4. 나가며

마지막으로 데이터마이닝과 관련하여 발표된 몇 가지 기사(기사내용의 일부)와 데이터마이닝 관련 논문에 있는 글을 소개하면서 마치도록 하겠다.

- DB 마케팅 시스템 구축에 대한 요구가 급증함에 따라 통계전문가들의 몸값도 덩달아 치솟고 있다. DB 마케팅을 구축하기 위해서는 데이터 분석을 위한 통계 기법이 필수적이지만 이를 지원할만한 전문인력들이 많지 않기 때문이다. 특히 통계학을 전공했다 하더라도 현업에서 방대한 데이터를 다뤄본 경우가 없는 경우가 대부분이기 때문에 이런 경험을 갖춘 인력발굴은 더욱 어려운 상황이다 (제목: 데이터분석 경험에 있는 통계전문가를 찾아라. 정보통신신문 97년 10월 27일자).
- 이제 기업에서는 뭔가 다른 획기적이고 혁신적인 정보시스템을 구축하기보다는 기존 시스템에 있는 데이터와 정보를 어떻게 활용할 것인가를 고민해야 한다. 이를 효과적으로 구현하기 위해서는 기업내부의 자료와 정보를 제대로 분석하고 객관적이고 보편 타당한 평가항목과 통계치출 도출해야 한다 (제목: 기업 내 정보를 보는 올바른 관점. 정보통신신문 99년 3월 1일자)
- 현재 데이터마이닝에 대한 교육과 인력을 양성할 수 있는 전문기관이 거의 없는 실정인데, 정부나 기업 모두 중요성을 인식하고 마이너 인력을 양성하는 과정을 빨리 만들어야 할 것이다. 새로운 데이터들이 초당 증가하고 있는 현실에서 이제는 경쟁의 무기가 데이터마이닝으로 전이되고 있다. (제목: 데이터마이닝의 세계. 뉴스피플, 99년 3월 6일자)
- 국민은행은 지난 해 '폰 뱅킹 등 전자거래 가입이 유망한 고객'이라는 데이터마이닝 작업을 진행했다. 가입 고객을 중 전자거래를 선호할 수 있는 잠재 고객 60만 명을 선정했다. 98년 데이터마이닝 작업이 끝나 현재 그 결과를 바탕으로 고객들을 대상으로 유치 작업에 들어갔다 (제목: 데이터마이닝 어떻게 활용되나. 뉴스피플 1999년 3월 11일자).
- 보험사의 입장에서는 해약할 가능성이 높은 계약 건들만을 골라서 집중적으로 이탈방지노력을 해야만 고객관계관리(CRM)가 이루어 질 것이다. 이러한 것은 소위 "데이터마이닝 과정"을 통해 해결할 수 있다. 결국 DW구축의 진정한 목적과 활용은 기업의 경영목표를 달성하기 위한 정보분석으로 확장되어야 한다는 점을 깨닫게 한다. 데이터마이닝의 과정은 일반적인 통계분석의 과정을 거친다 (제목: 정보분석의 현황과 과제. IT Business 1999년 4월 1일자).

또한, David(1999)는 "Data Mining: Statistics and More?" 에서

통계학의 증진을 위해 우리는 이전의 경험으로부터 배운다(실패를 되풀이하지 말자)는 것이 중요하다. 우리가 그렇게 하지 않으면 통계학(또는 통계학자)은 매우 적절치 못한 미세한 그룹으로 인식

될 것이고, 그들이 적절히 행하는 과학적이며 광범위한 삶에서 근본적인 역할을 못한다는 실제적 위험이 있다. 그리고 통계학자들이 전형적으로 과거에 그들이 관심을 가졌던 문제에만 안주하지 못하고 도록 하고 있다. 데이터마이닝 분야에서 통계학자들이 할 수 있는 매우 실질적인 공헌이 있다. 따라서 데이터마이닝이 근본적으로 통계문제가 되는 것에 관여하는 일에 주저해서는 안된다. 진정으로 중요한 문제들을 해결하는데 있어 통계학이 중요한 발전을 할 실질적인 기회가 주어져 있다. 이들 기회들을 붙잡지 않으면 이것은 통계학자 개인뿐 아니라 학문으로서의 통계학 명성에도 큰 손실이다.

라고 말하고 있다.

현재 국내에서는 SAS사와 SPSS사에서 데이터마이닝 교육이 이루어지고 있다. 또한 IBM에서는 직원들에게 마이닝 교육을 최근에 실시하였다. 이 이외에도 직원들에 대한 마이닝 교육에 관심을 갖고 있는 회사가 많이 있는 것이 현실이다. 이런 사회의 추세를 고려할 때, 이제는 우리 통계학자들도 교재에서 접하는 예제 데이터, 설문조사 자료를 포함하는 실험자료의 분석에 안주했던 모습에서 벗어나 현장의 대용량 실제 데이터를 직접 해결해 줄 수 있는 능력을 길러야 하는 절실함이 있다고 생각된다. 따라서, 그 동안 지속적으로 제기되고 있는 학문을 하는 사람과 현장에서 이를 이용하려는 사람들과의 괴리감 문제를 이제는 해결해야 한다는 시각에서, 정통적인 통계학 교육을 지키면서 사회적 요구에 부응하는 새로운 통계교육의 정체성(identity)을 찾기를 기대해 본다.

참고문헌

- [1] 강현철·한상태·최종후·김은석·김미경 (1999). 「SAS Enterprise Miner를 이용한 데이터 마이닝 -방법론 및 활용-」, 서울: 자유아카데미.
- [2] 송재기 (1999). “토론 : 통계학의 새로운 발전을 기대하며”, <응용통계연구>, 제12권 1호, 305-308.
- [3] 조신섭 (1999). “ ‘정보’ 관련 통계학과 의 교과과정에 대한 제안”, <한국통계학회 통계교육상담연구회 춘계 학술발표회 논문집>, 1-8.
- [4] 최종후·한상태·강현철·김은석·김미경 (1999). 「SAS Enterprise Miner를 이용한 데이터 마이닝 -기능과 사용법-」, 서울: 자유아카데미.
- [5] 최종후·한상태·강현철·김은석 (1998). 「AnswerTree를 이용한 데이터마이닝 의사결정나무분석」, 서울: SPSS 아카데미.
- [6] 한경수·최숙희 (1999). “정보사회에서의 통계학과 : 위기인가? 기회인가?”, <응용통계연구>, 제12권 1호, 295-304.
- [7] 허명희 (1999). “토론 : 통계학, 새로운 모습의 탐색”, <응용통계연구>, 제12권 1호, 309-3313.
- [8] David, J. H. (1998). “Data Mining: Statistics and More?”, *The American Statistician*, 52, 2, 112-118.
- [9] Friedman, J. H. (1997). Data Mining and Statistics: What's the connection?, 2th Symposium on the Interface: Computing Science and Statistics, <http://www.stat.rice.edu/interface97.html>.
- [10] Michael J. A. Berry and Gordon Linoff (1997). *Data Mining Techniques for Marketing, Sales, and Customer Support*. New York: John Wiley & Sons, Inc.
- [11] SAS Institute Inc. (1997). Data Mining Using SAS Enterprise Miner Software.
- [12] SAS Institute Inc. (1998). Enterprise Miner Software: Applying Data Mining Techniques.