

Development of Discernment Analysis System by Graphical User Interface¹⁾

Kyung-Joon Cha²⁾ · Young-Jae Shin³⁾ · Yong-Koun Lee⁴⁾

Abstract

우리는 다양한 자료에서 유의미한 정보를 파악하기 위한 방법으로 다변량 분석 방법 중에서 정준판별분석, 로지스틱, 다층퍼셉트론 그리고 의사결정나무를 사용자 편의를 극대화하고 사용이 간단한 비주얼 베이직 6.0을 이용하여 개발하였다.

Keywords : 정준판별분석, 로지스틱, 다층퍼셉트론, 의사결정나무

1. 머리말

최근 급증하는 데이터와 컴퓨터의 발전으로 정보화 사회로 빠르게 진행되고 있다. 이러한 정보화 사회에서 어떠한 방법으로 다량의 정보를 효과적으로 다루고 활용하느냐에 따라 상이한 수많은 결과들이 도출된다. 따라서 복잡하고 다양한 자료를 통계적 절차에 따라 처리 및 분석이 이루어져야 하며, 이는 정확한 결과 도출이 가능하게 한다. 현재 통계 분석을 위해 국내에서 많이 사용되는 SAS, SPSS, MINITAB 등은 많은 모듈로 구성되어 있고 다양한 통계 분석이 가능하다. 그러나 프로그램 사용법과 결과를 바르게 해석하기 위해 통계적 이론의 학습과 부단한 노력이 필요하다. 더불어 유사한 방법론들이 하나의 진행 과정에서 수행되지 않기 때문에 비교 분석이 용이하지 않다.

최근까지 다양한 자료에서 유의미한 정보를 구성하는 방법 중 많은 분야에서 활용되는 방법으로 다변량 분석이 있다. 따라서 다변량 분석과 관련하여 시스템을 구현하는 연구가 다양한 각도로 이루어졌다. 서혜선 등(1999)은 허명희(1999)에 의해 연

1) 이 논문은 한양대학교 일반연구비 지원으로 연구되었음.

2) First Author : Professor, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.
E-mail : kjcha@hanyang.ac.kr

3) Mse, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

4) Ph.D student, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

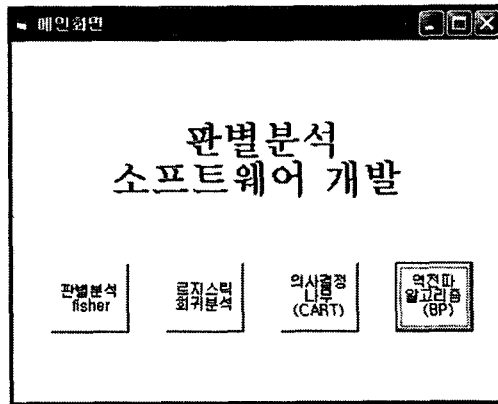
구된 다변량 수량화분석을 시스템화 하였으며, 현기홍, 최용석(2000)은 이원표 자료 행렬의 행과 열을 그래프에 나타내어 관계와 패턴을 분석하는 다변량 그래픽 분석 기법인 행렬도를 구현하였고, 한상태 등(2001)은 SAS/AF(application frame)와 SCL(screen control language)을 이용하여 SAS의 명령어 방식의 분석 과정을 메뉴 방식으로 제공하여 충분한 활용이 어려웠던 다변량 기법을 일반 사용자들도 쉽게 사용할 수 있는 다변량 분석 시스템을 개발하였다.

본 연구에서는 다변량 분석 중 이미 알려진 상호배반적인 몇 개의 집단에 속하는 다변량 관측치로부터 서로의 중복을 최소화하면서 구분할 수 있는 함수를 추정하는 판별분석을 비주얼 베이직 6.0(visual basic 6.0)을 사용하여 사용자가 편리하게 활용 가능한 GUI(graphical user interface) 시스템을 개발하였다. 개발에 사용되는 판별분석 방법으로 고전적인 Fisher의 정준판별분석, 의학 분야 중 질병 진단에 많이 활용되는 로지스틱 판별분석, 기계학습 방법인 신경망 중 다층 퍼셉트론 그리고 결과 해석이 용이한 의사결정나무를 이용하였다.

2. 판별분석 시스템의 구성

판별분석 시스템은 비주얼 베이직을 사용하여 GUI에서 누구나 쉽게 클릭만으로 사용할 수 있도록 하는데 중점을 두고 구축하였으며, 전체 시스템은 메뉴방식으로 구성되었다. 즉, 분석을 위한 절차 과정에 필요한 자료의 처리 및 분석에 대한 내용 보다 최종적인 모형이나 결론 도출에 활용할 수 있게 구성하였다.

본 연구에서 개발한 다변량 분석 시스템은 <그림 1>처럼 Fisher의 정준판별분석, 로지스틱 판별분석, 다층 퍼셉트론 중 역전파 알고리즘 그리고 의사결정나무 중 CART 등 다양한 판별분석을 수행할 수 있도록 되어 있다. 또한 각 분석방법은 텍스트 형식으로 저장된 자료를 분석 과정에서 직접 입력함으로써 간단하게 시스템에 적용시킬 수 있고, 정해진 분석 방법별로 결과가 도출되어 오분류로 최종적인 결과를 출력 할 수 있도록 되어 있다.

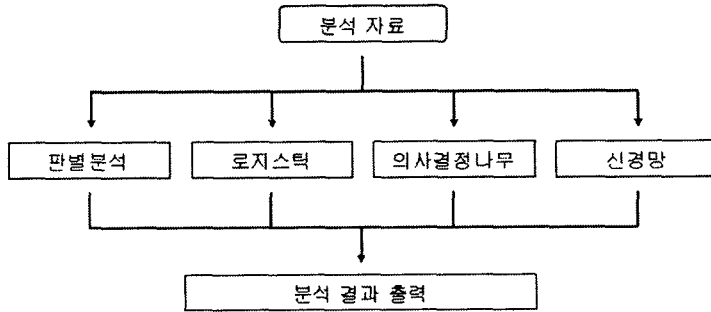


<그림 42> 판별분석 시스템 초기화면

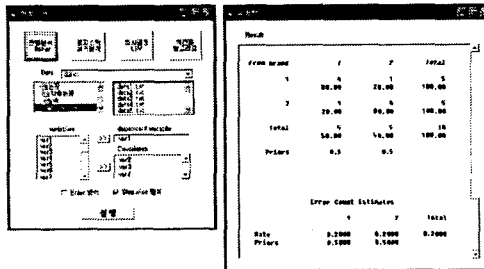
3. 판별분석 시스템의 실행

본 연구에서 개발한 판별분석 시스템의 결과 비교 분석을 위해 활용한 자료는 일반적으로 통계적 방법론의 효율성 평가에 다양하게 사용되는 붓꽃자료이다. 붓꽃자료는 3가지 종류(versicolor, setosa, virginica)의 붓꽃에 대해 꽃의 종류(species), 꽃받침 너비(petal width), 꽃받침 길이(petal length), 꽃잎 너비(sepal width) 그리고 꽃잎 길이(sepal length)를 측정한 것이다. 꽃의 각 종류마다 50개씩의 자료로 전체 150개의 자료이다.

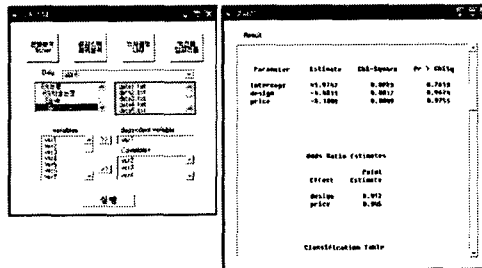
전체적인 분절 절차는 <그림 2>와 같은 자료처리 순서를 갖는다. 입력된 자료는 4가지 판별분석 방법에 모두 적용이 가능하며 각 방법에 따른 옵션의 선택과 결과의 도출은 다르게 나타나게 된다. <그림 3> ~ <그림 6>은 각 방법의 실행 화면과 결과창의 화면이다. 방법별로 기본적인 절차는 분석할 자료를 우선 선택하고 이와 함께 부분적인 옵션의 선택이 필요하면 실행을 하면 결과가 도출된다.



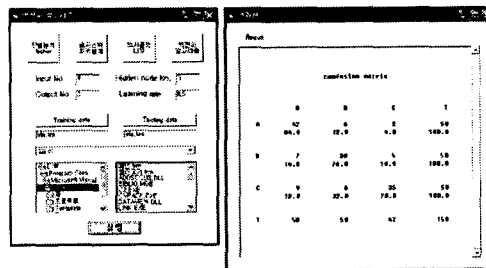
<그림 43> 판별분석 시스템 자료처리 순서 및 결과 출력



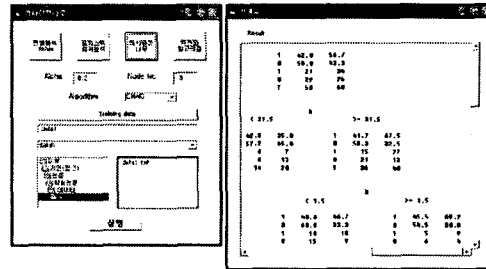
<그림 44> 정준판별분석 실행



<그림 45> 로지스틱 실행



<그림 46> 다층퍼셉트론 실행



<그림 47> 의사결정나무 실행

4. 결론

본 연구에서는 다양한 자료에서 유의미한 정보를 판단하는 다변량 분석 방법 중에서 판별분석을 비주얼 베이직 6.0을 사용하여 개발하였다. 개발에 사용된 분석 방법은 정준상관분석, 로지스틱, 다층퍼셉트론 그리고 의사결정나무가 이용되었다. 본 연구는 다음과 같은 특징을 갖고 있다.

첫째, 다양한 통계 분석 모듈에서 제공하는 판별분석 방법은 사용법과 결과를 빠르게 해석하기 위해 통계적 이론의 학습과 부단한 노력이 필요하나, 본 연구에서 개발한 시스템은 판별분석만을 위한 간단한 절차를 갖는다.

둘째, 유사한 판별분석 간의 비교를 통하여 전체적인 상황에서 적절한 판별분석 방법론을 선택가능하게 하였다.

셋째, 비주얼 베이직을 사용하여 GUI에서 누구나 쉽게 클릭만으로 사용할 수 있게 하여 분석을 위한 절차 및 분석에 대한 내용보다 최종적인 모형이나 결론 도출에 활용할 수는 특징을 갖는다.

참고문헌

1. 서혜선, 김미경, 허명희 (1998). SAS AF/SCL로 구현한 다변량 수향화 시스템, *한국분류학회*, 제 3권, 1-11.
2. 허명희 (1999). *사회과학을 위한 다변량자료분석*, 서울 : 자유아카데미.
3. 현기홍, 최용석 (2000). 행렬도 시스템(Biplots System)의 개발, *응용통계연구*, 13권 2호, 297-306.
4. 한상태, 강현철, 이성진, 장명석, 이덕기, 유동균 (2001). Development of Multivariate Analysis System by Using SAS/AF and SCL, *The Korean Communications in Statistics*, 8, 507-514.