

의료 통계 분석 및 시각화 시스템*

이돈수⁰, 최수미
세종대학교 컴퓨터공학부
spike25@empal.com⁰, smchoi@sejong.ac.kr

A System for Statistical Analysis and Visualization in Medicine

Don-Soo Lee⁰, Soo-Mi Choi
School of Computer Engineering, Sejong University

요 약

임상 및 실험 결과를 대외적으로 공인 받기 위해서는 통계적 검정 절차를 거치는 것이 일반적이다. 하지만 통계적 전문 지식이 부족한 사용자가 통계 소프트웨어를 배우는 데는 시간이 많이 걸리며, 결과 해석에도 어려움이 많은 실정이다. 데이터의 특성과 성질에 맞추어 통계법이 선택되어야 하는데, 통계지식이 부족한 초보자들은 가장 일반적인 분석법을 적용시키곤 한다. 이와 같은 방식의 통계분석은 잘못된 결과로 이어질 수 있기 때문에 올바른 분석법을 가이드 해주는 기능이 필요하다. 또한 통계분석법의 적합성을 평가하는데 있어 오차와 잔차의 등분산성 가정이 유용하게 쓰여질 수 있다. 본 연구에서는 사용자에게 올바른 분석법을 제시하는 비주얼 가이드 인터페이스와 잔차를 3D Glyph를 이용하여 보여주는 불확실성 시각화 방법을 사용하였다. 분석법 적용에서 나타나는 불확실한 데이터의 시각화는 의사결정에 도움을 줄 수 있다.

1. 서 론

임상에서 환자를 대하면서 많은 시간과 노력을 기울여 얻어진 관찰 결과 혹은 실험을 통해서 얻어진 데이터의 의미를 정확하게 분석하는 것은 매우 중요하다. 그러나 전문적인 통계적 지식의 부족으로 실험이 잘못 분석되거나 결과 해석에 어려움을 겪는 경우가 종종 발생한다[1]. 또한 통계 분석 자체에 내재된 불확실성으로 인하여 의사결정이 어려울 때가 종종 있다[2]. 그러므로 이러한 의료 데이터의 통계분석을 쉽게 할 수 있도록 가이드 해주는 비주얼 인터페이스와 불확실한 데이터의 시각화 방법에 대한 연구가 필요하다.

이와 관련된 연구를 살펴보면, MedCalcTM[3]는 임상 의학에 특화된 통계소프트웨어로서 조작이 간편한 인터페이스와 의학 분야에 자주 사용되는 분석법을 제공한다. 그러나 적절한 통계 분석법을 선택할 수 있도록 가이드 해주는 기능을 제공하지는 않고 있다. ViStaTM[4]는 초보자부터 전문가를 대상으로 만들어진 통계시스템으로, 분석과정을 시각화하여 보여준다. 특히 WorkMap 과 GuideMap 을 제공하여 데이터 분석과정과 올바른 분석을 하도록 도와주는 기능을 제공한다. dBSTATTM[5]는 통계지식이 부족한 일반인을 위해 설계된 통계 소프트웨어로서 통계 마법사 기능을 통해 간편한 분석을 할 수 있도록 하였다. 그러나 분석법의 적절성 평가를 해주는 기능은 없다.

본 연구에서는 사용자가 실수를 하더라도 경고 메시지를 통하여 미연에 잘못된 분석을 방지하고, 올바른 분석으로 이끄는 비주얼 가이드 인터페이스를 제공한다. 또한 전체적인 통계분석과정을 이해하기 쉬운 아이콘으로 구성하여 시각화하는 아이콘 트리맵을 제공함으로써 결과해석을 돕고자 한다. 그리고 잔차를 3D Glyph를 이용하여 시각화하는 방법을 개발하여 분석법의 적절성 평가에 도움을 주고자 한다.

2. 의료 통계 분석 및 시각화 시스템 설계

2.1 시스템의 구조 및 특징

개발된 시스템의 주요 특징은 다음과 같다. 첫째, 의료분야에서 자주 사용되는 통계 분석법을 제공한다. 둘째, 올바른 통계분석으로 유도하는 비주얼 가이드를 제공한다. 셋째, 통계 분석시 나타나는 오차와 잔차와 같은 불확실한 데이터를 3D Glyph를 이용하여 시각화함으로써 의사결정에 도움을 주고자 한다. 시스템의 전체 구조도는 그림 1 과 같이 나타난다. 데이터가 입력되면 통계 분석에서 기본적인 분석이 수행된다. 이때 비주얼 가이드는 데이터의 적합도 평가 모듈의 검사를 거쳐 부적절성이 발견되면 경고 메시지를 출력하여 실수를 방지한다. 불확실성 시각화 과정은 분석법별 오차 계산 모듈을 통해 분석법의 적절성을 평가하여 의사결정을 돕는다. 과정이 끝나면 결과해석 모듈에서 텍스트, 2D, 3D 그래프, SAS 코드를 제공한다.

*본 연구는 보건복지부 보건의료기술 연구개발 사업에 의해 지원되었습니다. (02-PJI-PG3-51312-0003).

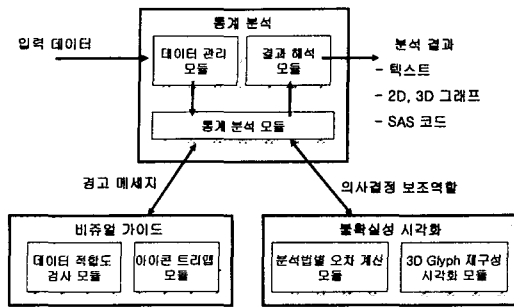


그림 1. 시스템 구조도

2.2 비주얼 가이드

비주얼 가이드는 적합도 검사 모듈과 아이콘 트리맵 모듈로 구성된다. 적합도 검사 모듈은 입력된 데이터의 무결성, 데이터형, 분석법적합도 검사로 구성되어 있다. 만일 이들 검사에서 부적절 하다고 판단되면 사용자에게 경고 메시지를 출력 하고, 올바른 분석법을 제시한다. 아이콘 트리맵은 통계분석의 전체적인 과정을 알기 쉬운 아이콘으로 구성하여 보여줌으로써 통계지식이 부족한 사용자가 분석과정을 이해하는데 도움을 줄 수 있다.

2.3 3D Glyph를 이용한 인터랙티브 데이터 시각화

Glyph란 위치, 형태, 색, 투명도와 같은 속성을 갖는 그래픽 객체를 의미한다. 이러한 속성들은 데이터의 여러 차원을 동시에 표현함으로써 다변량 데이터 시각화에 많이 사용되어진다[6]. 그림 2는 평균, 분산, 구간의 통계적 의미의 정보를 3D Glyph로 보여주는 예이다. 이러한 Glyph 객체들은 시점 변환이나 수차 변환과 같은 인터랙티브한 조작이 가능하여 데이터 분석이 용이하다.



그림 2. 3D Tukey Glyph

통계 분석법에 데이터를 적용시키는데 산출되는 잔차는 분석법의 적절성을 평가하기 위해 필요한 정규성, 등분산성 평가를 위해 유용하게 쓰여질 수 있다[2]. 본 연구에서는 잔차를 통계적 범주의 불확실한 데이터[7]로 보고 사용자가 현재의 잔차 모델과 일반적으로 잘 알려

진 잘못된 모델과 비교 검토해 봄으로써 분석법의 신뢰도를 평가하는 기능을 제공한다. 잔차 데이터는 3D Glyph를 이용하여 사용자가 쉽게 이해할 수 있도록 한다.

3. 구현 및 결과

개발된 시스템은 Windows 2000 OS를 탑재한 펜티엄 866Mhz 급 PC에서 비주얼 C++ 언어를 이용하여 구현하였다.

다음은 비주얼 가이드가 쌍을 이룬 자료의 T 검정에서 적용되는 예를 보여준다. 그림 3의 그래프에서 보여지듯이 입력 정규분포 가정을 만족하지 않는 데이터 B의 경우 분석을 진행하면, 그림 4와 같이 경고 메시지를 주고 비모수적 통계 방법을 적용하라고 가이드 해준다.

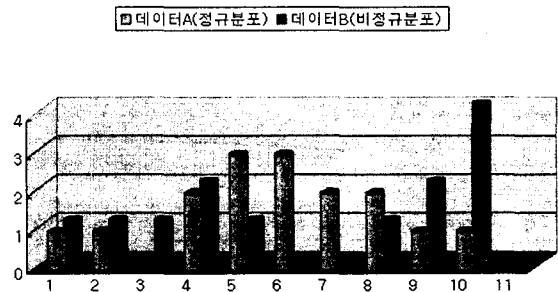


그림 3. 정규분포 성질이 다른 데이터A, B 그래프

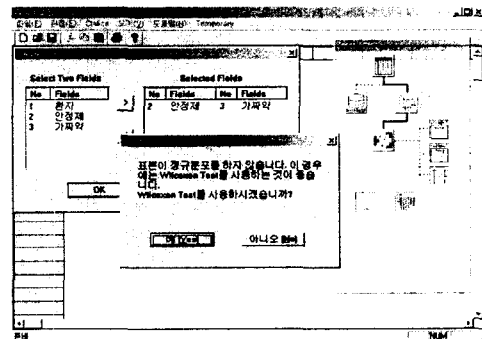


그림 4. 적합한 통계 분석법 가이드 화면

오른편에 보여지는 아이콘 트리맵은 내부적으로 적합도 검사의 수행 결과, 현재 진행되고 있는 경로, 본 시스템에서 제안하는 경로를 아이콘과 다른 색으로 보여줌으로써 분석 과정을 쉽게 이해할 수 있도록 해준다. 빨간색은 지나온 경로를 나타내고, 파란색은 현재 진행되는 경로, 흰색은 선택으로 분기되는 것을 나타낸다.

입력된 데이터와 분석법을 적용시켜 산출되는 잔차를 검토하면 분석법의 적절성을 평가할 수 있다. 단순선형

회귀분석에서 잔차는 식(1)에서 오차항 ε_i 으로 나타나며 식(2)와 같이 계산된다.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

$$= E(Y_i) + \varepsilon_i \quad (i = 1, \dots, n)$$

$$\varepsilon_i = Y_i - E(Y_i) \quad (2)$$

따라서 회귀모형이 적절하다면 잔차는 등분산성의 가정을 반영해야 한다. 본 연구에서는 의사결정에 도움을 주기 위해 분석법의 불확실성을 사용자가 평가할 수 있도록 하였다. 그림 5는 단순선형회귀 분석이 끝난 후 잔차를 알기 쉬운 3D Glyph를 이용하여 보여주고 있다.

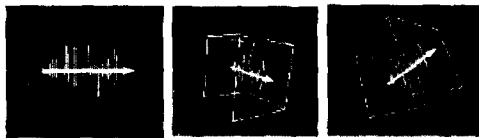
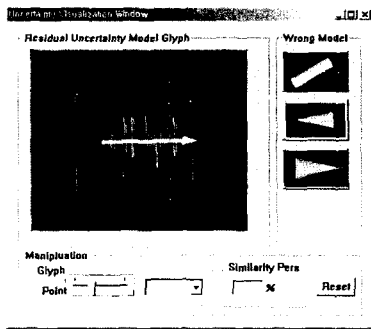


그림 5. 3D Glyph를 이용한 잔차의 시각화

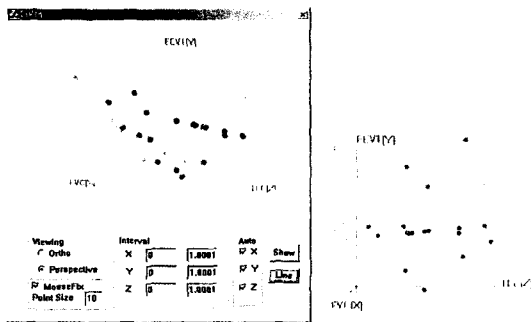


그림 6. 3D 산점도 그래프 화면

일반적으로 통계분석에는 2D그래프가 사용되는데, 3D

그래프가 이해하기 쉬운 경우가 있다. 그림 6은 다중선형회귀분석의 3D 산점도 그래프를 보여주는데, 2D에서는 보이지 않는 3차원의 데이터 연관 관계를 볼 수 있다. 본 연구에서 개발된 시스템과 기존의 시스템들의 특징에 따라 비교하여 정리해 보면 표 1과 같다.

표 1. 시스템 특징 비교표

소프트웨어 특징	MedCalc	ViSta	dBSTAT	Our System
적절한 통계 분석 가이드	데이터형 검사	Guidemap	통계 마법사	비주얼 가이드 경고 메시지
분석과정 시각화	텍스트	Workmap	제공없음	아이콘 트리맵
불확실성 시각화	제공없음	제공없음	제공없음	Glyph 시각화
그래프 형식	2D 형식	2D, 3D	2D	2D, 3D 인터랙티브

4. 결론 및 향후연구

본 연구에서는 의료데이터의 통계분석을 위하여 의학 통계에서 많이 사용되는 통계분석법을 구현하였고, 쉽고 정확한 사용을 위하여 비주얼 가이드 인터페이스를 연구 개발 하였다. 그리고 불확실한 데이터를 3D Glyph를 이용하여 시각화하는 방법을 개발하여 의사결정에 도움을 줄 수 있었다. 향후에는 잔차 모델과 잘못된 모델을 자동으로 비교해주는 연구와, 보다 많은 분석법에 불확실성 시각화 방법을 적용하는 것이 필요하다.

5. 참고 문헌

- [1] 김성권, "SAS와 통계", 대선, 2000.
- [2] 송해양, 김동재 공저, "통계학", 청문각, 2002.
- [3] Frank Schoonjans, "MedCals", www.medcalc.be.
- [4] Forrest W. Young, Carla M. Bann, "ViSta: A Visual Statistic System.", Statistical Computing Environment for Social Research, pp 959-998, 1993.
- [5] 김수녕, "윈도우용 통계소프트", 탐진, 2000.
- [6] Alex T. Pang, "Approaches to Uncertainty visualization", 1996.
- [7] Soo-Mi Choi, Don-Soo Lee, et al., "Interactive Visualization of Diagnostic Data from Cardiac Images using 3D Glyphs", ISMDA, 2003.
- [8] Vladimir Batagelj, Andrej Marvar, "Visualization of The Multivariate Data Using 3D and VR Presentation vlado.fmf.uni-lj.si/vrml/paris.97, 1997.
- [9] 유대근, 권영식 공저, "통계분석을 위한 SPSS WIN 8.0", 기한재, 1999.
- [10] Leland Wilkinson, "Presentation Graphics", SPS S Ins. 233 South Wacker, Chicago, IL 60606.
- [11] Colin Ware, "Information Visualization", Morgan Kaufmann Publishers, 2000.