

## 웹 기반의 통계 프로그램의 유형 분석과 설계 방안 연구

정남철\*

### 요약

본 논문에서는 통계 어플리케이션을 웹 서버에서 내려받기 하여 클라이언트에 저장, 설치하고 stand alone 형태로 운영되도록 개발된 DAVIS에 대하여 개발 기법과 구현된 형태를 연구하고, 서버 기반의 통계 프로그램과 클라이언트 기반의 통계 프로그램에 대하여 고찰한다. 그리고 이들 유형에 대한 장단점을 파악하여 좀 더 발전된 통계학습시스템의 설계 방안을 제안한다. 이 시스템은 클라이언트 요청에 의하여 클라이언트에서 어플리케이션이 실행되고, 통계 데이터는 데이터베이스 서버에서 로드하거나 사용자에 의해 클라이언트에서 입력하는 형태로 설계되어 통계 분석을 수행토록 한다.

## Design Scheme and Analysis of Web-Based Statistics Program Types

Nam-Cheol Jung\*

### Abstract

In this paper, I discuss the development techniques and implemented type of DAVIS, a standalone JAVA-based application, by which statistics application can be downloaded from a web server, stored and installed to a client. Also I discuss both the server-based application and the client-based application that analyze statistics data. Considering the strengths and weaknesses of types discussed previously, I propose an advanced learning system of statistics that can execute an application in a client by client request. And data used in this system are loaded from a database server or input into a client by users.

**Keywords :** 통계프로그램, 시각화도구, 통계학습시스템, 회귀계수, DAVIS, applet, FEDF, p-p Plot

### 1. 서론

정보화 사회로 진입하면서 통계 프로그램들이 개인용 컴퓨터에 설치되어 손쉽게 이용이 가능하게 되었으며, 그래픽 기능을 이용하여 처리 결과를 쉽게 이해할 수 있게 되었다. 이와 관련된 연구는 Tierney, Weisberg[13], 허문열[8], Cleverland[5] 등이 많은 공헌하였다. 특히 Tierney는 통계 프로그램을 손쉽게 개발할 수 있는 객체 지향 프로그래밍 언어인 Xlisp\_Stat을 개발하였으며[10], 개발자들은 이 언어의 객체지향의 모듈을 통해 통계 분석 및 통계 동적 프로

그램을 쉽게 구사할 수 있다.

초고속 인터넷의 발달과 개인용 컴퓨터의 성능이 향상되면서 통계 데이터를 실시간으로 분석할 수 있고, 결과를 빠른 시간에 그래픽으로 살펴볼 수 있게 되었다. 이러한 결과로 인터넷상에서 통계와 관련된 정보교환[7]과 통계 소프트웨어의 활용에 대한 관심이 날로 높아지고 있다. 인터넷에서 통계의 학문적 연구는 주로 학습용으로 개발되거나 간단한 통계 분석을 할 수 있도록 구현되고 있다.

인터넷에서 통계 프로그램의 공유는 저작권 문제와 처리 속도의 지연을 야기한다. 그래서 많은 통계학자들과 프로그램 개발자들은 통계 분석의 목적에 맞게 공개용 통계 프로그램을 부분적으로 구현하여 인터넷상에서 활용할 수 있도록 연구하고 있다.

본 논문에서는 stand alone 형태로 개발된

※ 제일저자(First Author) : 정남철  
접수일자:2007년04월25일, 심사완료:2007년06월21일  
\* 동남보건대학 웹컨텐츠과  
ncjung@dongnam.ac.kr

DAVIS[9]에 대하여 개발 기법과 구현된 형태를 고찰하고, 통계 프로그램을 웹상의 서버에서 실행되는 서버 기반의 통계 프로그램과 클라이언트에서 실행되는 클라이언트 기반의 통계 프로그램에 대하여 연구한다. 그리고 이들 유형에 대한 장단점을 파악하여 좀 더 발전된 형태의 통계 프로그램의 설계 방향을 제안한다.

## 2. 관련 연구 : DAVIS

### 2.1 DAVIS의 개요

허문열과 송광열은 JAVA 기반의 데이터 시각화 시스템인 DAVIS를 개발하여 발표하였다[11]. DAVIS는 웹상에서 다운로드 받아 클라이언트에 저장, 설치하여 독립적으로 운영할 수 있는 통계 시각화 프로그램이다. 특히 이 시스템의 어떤 시각화 도구 상에서 데이터의 한 부분을 마우스로 브러싱(brushing)하여 데이터 구조를 시각적으로 탐색해 볼 수 있도록 설계하였다. 이렇게 브러싱된 여러 점의 데이터는 동적으로 다른 시각화 도구에 연결할 수 있다. 또한 DAVIS는 실시간으로 분류된 데이터의 결과를 시각화 도구에 전이함으로써, 수학적인 군집(clustering) 방법의 결과를 시각적으로 탐색할 수 있도록 구현하였다.

### 2.2 시각화 도구

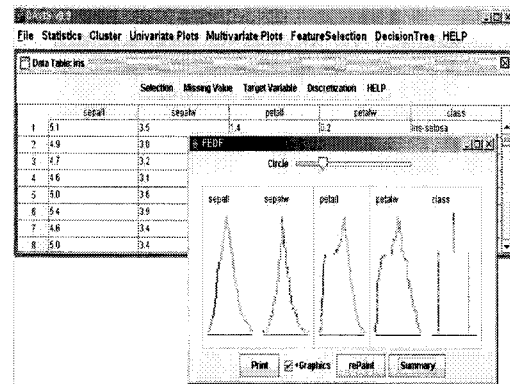
DAVIS를 위한 시각화 도구로는 히스토그램, 상자그림(boxplot), FEDF(flipped empirical distribution function[8]), 수평좌표(parallel coordinates), 덴드로그램(dendrogram), 스캐터플롯(scatterplot matrix), 그랜드 투어(Grand Touring), 그리고 추적 그랜드 투어(tracking Grand Tour[9]) 등을 포함하고 있다.

FEDF는 허문열에 의해 제안된 또 다른 형태의 EDF이다. FEDF는 다음과 같이 정의 된다.

$$\begin{cases} EDF(x) & \text{if } x \leq median \\ 1 - EDF(x) & \text{otherwise} \end{cases}$$

FEDF를 사용하면, 경사진 분포의 왜도와 첨도에 관련된 정보를 그래픽으로 얻을 수 있다. 그림들은 각각의 데이터 점들의 위치를 제공하

기 때문에 이 도구는 관심 있는 점들의 값들을 동적으로 다른 그림에 연결시키는 효과적인 도구로 제공된다. (그림 1)은 UCI 데이터베이스에서 주어진 Fisher의 iris 데이터[14]를 사용하여 FEDF를 구현한 모습이다.



(그림 1) DAVIS 중 FEDF 시각화 도구

수평좌표는 d-차원의 데이터의 점들을 표현하기 위해 그들을 데이터 변수에 따라 d 개의 수평축을 그린다. 각 데이터의 점은 그것의 각 축에 따라 각 변수의 값을 위치로 표시하게 되고, 이 점들은 선으로 결합된다. 이 그림은 데이터 시각화와 데이터 마이닝으로 사용하는데 많은 관심을 보여 왔다[11][12].

덴드로그램은 다차원 데이터를 표시하는 또 다른 도구로 군집(cluster)의 계층적 구조를 보여주는 그림이다. 이 그림의 형태는 트리 구조를 가지며, 가장 위층의 모든 관찰 값들은 1개의 군집에 속하고 가장 하층의 각각의 관찰 값들은 각각 다른 군집에 속하도록 표시한다. 그리고 유사한 속성을 가진 원소들은 서로 근접하게 배치한다.

스캐터플롯은 시각화를 위해 가장 널리 사용되는 도구이다. Becker, Cleveland, Wilks는 데이터 분석을 위해 동적 그래픽으로 스캐터플롯을 널리 사용했다[4]. SAS, S-PLUS, 그리고 SPSS를 포함한 대부분의 통계 패키지들은 그래픽 도구를 구현하는 수단으로 이 도구를 사용하고 있다.

그랜드 투어는 Asimov에 의해 개발된 다차원 데이터를 시각화하기 위한 도구이다[3]. 이 도구

는 임의로 생산된 일련의 평면들을 삽입한 평면상에 데이터의 점들을 연속적으로 투영한 것으로 그랜드 투어에 삽입한 평면상에서 투영된 데이터의 점들을 관찰할 수 있다. 허문열과 김기열은 그랜드 투어를 수정하여 투어하고 있는 동안에 관찰자가 데이터의 점들이 지나간 흔적을 관찰할 수 있는 추적 그랜드 투어(TGT)를 제시하였다. TGT는 작은 크기의 데이터에 대하여 데이터 구조를 탐색하는데 특히 유용하다.

### 2.3 DAVIS의 구현

DAVIS는 풀다운(pull-down) 메뉴로 구성하며, 데이터 메뉴에서, 데이터들은 파일로부터 읽어오거나 혹은 어떤 편집기로부터 아스키(ASCII) 형태로 가져 온(복사하여 붙이기) 데이터가 로드될 때, 2차원 테이블 형태로 데이터를 관찰할 수 있도록 데이터 컴포넌트를 제공한다. 데이터 컴포넌트는 테이블에서 변수를 선택하거나, 관찰 값의 일부 데이터를 선택하거나, 단순임의 샘플링(simple random sampling)으로 일부 데이터를 채취하는 메뉴 버튼을 제공하며, 테이블에서 변수 행을 클릭하면 오름차순 또는 내림차순으로 특정 변수에 대하여 데이터를 정렬할 수 있도록 구현하였다.

그 다음 메뉴로 3개의 컴포넌트 : 군집, 통계, 시각화 도구 등을 제공한다. DAVIS는 어떤 단계에서든 군집 분석을 행할 수 있도록 구성되어 있다. 군집 후에, 결과는 즉각적으로 오픈된 모든 도구에 전이된다. 또한 마우스를 사용해서 어떤 시각화 도구에서 데이터의 일부를 선택, 삭제, 포커싱할 수 있고, 이 조작의 결과는 즉각적으로 다른 도구에 전이되도록 구현하였다.

### 2.4 DAVIS의 장단점

허문열에 의해 개발된 DAVIS는 데이터의 구조를 잘 나타낼 수 있는 시각화 도구로서 통계 교육에 활용할 수 있는 좋은 콘텐츠이며, 누구든지 웹상에서 다운로드하여 클라이언트에 설치하여 사용하기 쉽도록 구현하였다. 웹 기반으로 보기는 어렵지만, 웹이 정보를 교환하고 공유를 목적으로 한다는 측면을 고려하면, 웹을 이용하여 다운로드하여 활용할 수 있다는 것으로 웹의 한 형태로 분류해도 될 것으로 생각된다. 개방형 프로그램인 DAVIS의 실행은 클라이언트에서 독립

적으로 동작하는 방식을 취하고 있어 서버에는 전혀 부담을 주지 않아 처리 속도 및 서버 자원의 절약 등에서 효율적인 시스템이다. 또한 사용자가 마우스를 통해 쉽게 이용할 수 있도록 사용자 인터페이스를 잘 설계하였다.

다만, 데이터 형식이 Weka에서 사용하는 arff를 따르고 있어 이 형식을 잘 이해하지 못하는 사용자들은 데이터 작성에 어려움이 있다.

## 3. 서버 기반의 통계 프로그램

### 3.1 개념

이 유형은 통계 프로그램을 서버에 설치하고, 응용 프로그램(또는 CGI 프로그램)에 의해 통계 프로그램을 서버에서 실행하고, 그 결과를 클라이언트에 전송하여 통계 분석이나 e\_러닝에 이용하는 형태이다. 이 유형은 제한 사항이 많이 있으나, Friendly는 SAS와 같은 통계 패키지를 서버에 설치하고 서버 측에서 이를 처리하여 클라이언트에 제공하는 방식은 다음과 같은 경우에 유리하다고 주장하였다[6].

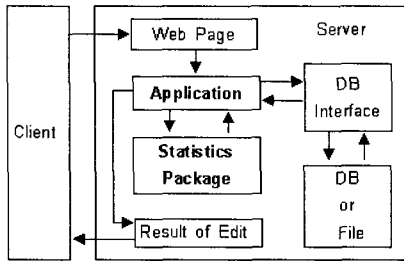
- 클라이언트에 되돌려지는 결과가 하나의 표나 그래픽과 같은 1개의 파일일 경우
- 서버에 있는 데이터 셋에 근거하여 결과를 산출할 경우
- 서버가 많은 프로세스를 수용할 수 있을 때

서버에 통계 프로그램을 설치하여 웹상에서 통계 분석을 수행할 경우에 인터페이스를 통해 초보자들이 쉽게 접근할 수 있도록 개발할 수 있으며, 의도하는 대로 처리 결과를 클라이언트에 전송하여 브라우저에 나타낼 수 있다.

### 3.2 서버 인터페이스

서버 기반의 통계 프로그램에서 통계 처리를 위해서는 웹 서버와 통계 프로그램 간에 인터페이스를 할 수 있는 어플리케이션이 필요하다. 또 자료를 데이터베이스에서 로딩하여 사용할 경우에도 데이터베이스와 인터페이스할 수 있는 어플리케이션이 필요하다. 통계 프로그램이 웹 서버에 설치되어 있는 경우에 서버 인터페이스 프로그램으로 CGI나 ASP 등을 사용하여 운영할

수 있다. CGI 프로그램은 웹 서버의 본래의 처리 기능뿐만 아니라 새로운 처리 기능을 수행할 때 이용한다. (그림 2)는 서버 기반의 통계 프로그램을 이용할 경우에 처리 흐름도와 인터페이스를 표시한 것이다.



(그림 2) 서버 기반의 통계 프로그램 흐름도

서버 인터페이스를 위한 CGI 프로그램은 클라이언트에서 서버에 통계 분석을 요청할 때마다 서버 안에 새로운 프로세스가 생성하므로 서버에게 많은 부담을 주게 된다. 이 대안으로 ASP를 이용할 수 있는데, ASP는 클라이언트로부터 요청이 들어올 때마다 새로운 프로세스를 생성하는 CGI와는 달리 MS-Windows의 스레드(Thread) 개념인 프로세스 공유 기술로 이를 처리하는 방식을 채택하므로 서버의 부담이 경감된다. 그러나 ASP를 이용하여 빠르게 통계 분석 및 처리할지라도 많은 사용자가 요청할 경우에 통계 프로그램을 공유하므로 속도가 저하되는 문제점을 가지고 있다.

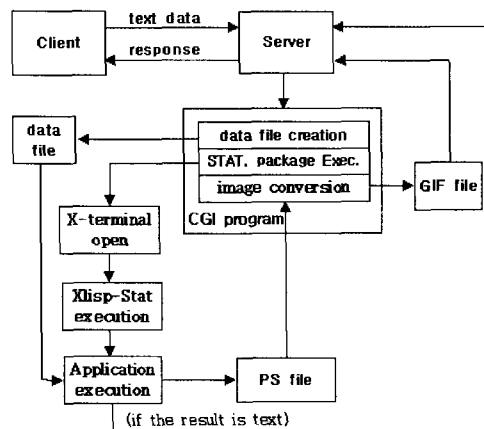
### 3.3 통계 프로그램의 저작권 문제

웹 서버에 설치된 상용의 통계 프로그램을 불특정의 사용자에게 이용을 허용하는 것은 저작권 침해 논란의 여지가 있다. 그러나 저작권 문제를 해결하기 위한 한 방법으로, 통계 프로그램을 JAVA나 ASP 등을 이용하여 어플리케이션으로 개발하여 셰어웨어로 사용하게 할 경우에는 가능하다. 이는 주로 웹상에서 통계 교육용 콘텐츠나 간단한 통계 분석 및 데이터 시각화 등에 활용할 수 있다.

저작권 문제를 해결하기 위한 사례로 InterStat을 들 수 있다[1]. InterStat은 통계 프

로그램인 Xlisp-Stat을 서버에 설치하고 통계 분석 및 처리를 위한 어플리케이션 프로그램을 개발하여 사용자에게 서비스하는 통계학습 시스템이다. Xlisp-Stat은 Tierney가 개발하여 셰어웨어로 공개한 통계 프로그램이다[10].

InterStat에서는 Xlisp-Stat과 서버 간에 인터페이스하는 CGI 프로그램을 통해 Xlisp-Stat을 실행하고, 클라이언트의 통계 분석 요청에 서비스하는 형태를 채택하고 있다. (그림 3)은 InterStat에서 CGI 프로그램의 프로세스이며, 이 시스템의 플랫폼은 리눅스이다.



(그림 3) InterStat의 CGI 프로세스

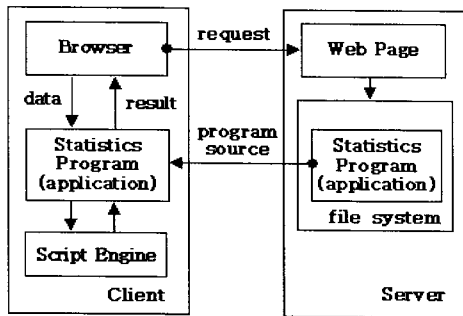
## 4. 클라이언트 기반의 통계 프로그램

### 4.1 통계 분석 흐름

일반적으로 서버-클라이언트 모델에서 통계 프로그램(여기서는 통계 분석을 위한 어플리케이션을 말한다)은 서버의 가상 디렉토리에 존재하며, 클라이언트에서 통계분석을 요청할 경우에 통계 프로그램이 클라이언트로 전달되어 클라이언트에서 제공된 데이터를 이용하여 실행하게 된다. 여기서 통계 프로그램은 상용의 통계 소프트웨어가 아니라, 통계 분석이나 시각화할 수 있는 응용 프로그램을 의미하며, 스크립트 프로그램이나 자바 애플릿에 의해 구현될 수 있다. (그림 4)는 클라이언트 기반의 통계 프로그램 처

리 흐름도이다.

통계 프로그램과 클라이언트간의 인터페이스는 API 개념을 이용한 NCSA Mosaic CCI나 Netscape Client API(NCAPI), DDE, OLE 등을 이용하여 어플리케이션을 실행하게 된다. 클라이언트에서 사용자가 통계 분석을 위해 통계분석 웹 페이지를 요청하면, 서버는 요청된 페이지에 해당하는 통계 분석 어플리케이션인 소스나 바이트 코드를 클라이언트에 전송한다. 클라이언트는 전송되어 온 통계 분석 어플리케이션을 스크립트 엔진 또는 자바 가상 머신(JAVA Virtual Machine : JVM)에 의해 번역하여 실행한 후에 그 결과를 브라우저에 출력한다.



(그림 4) 클라이언트 기반의 통계프로그램 처리 흐름도

#### 4.2 처리 속도의 향상

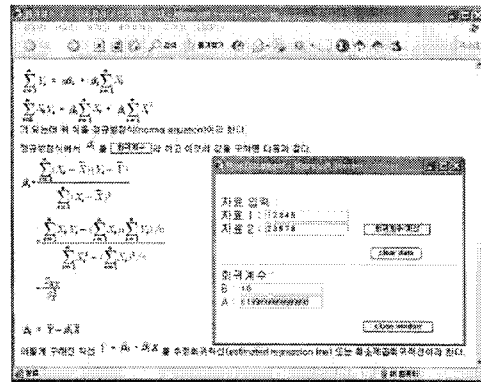
클라이언트 기반의 통계 프로그램의 처리 속도는 서버 기반의 통계 프로그램의 처리 속도보다 매우 우수하다. 사용자의 요청에 의해 전송된 프로그램 소스는 클라이언트에서 번역, 실행하여 분석 결과를 브라우저에 표시하면 되므로 서버 입장에서는 처리 부담이 줄어들고, 사용자는 분석 결과를 빠르게 확인할 수 있다.

이 유형의 문제점으로 스크립트로 개발된 통계 분석 어플리케이션의 경우에 클라이언트로 전송된 소스를 사용자가 볼 수 있으므로 프로그램이 공개되는 단점을 가지고 있다.

#### 4.3 간단한 구현의 예

본 논문에서는 통계 교육용으로 활용하기 위해 회귀계수를 구하는 통계 프로그램을 예제로

간단하게 구현하였다. 사용자는 회귀계수를 구하기 위해 데이터를 일정한 입력 폼(HTML 폼)에 입력하고 “회귀계수” 버튼을 누르면 이벤트가 발생되어 어플리케이션을 클라이언트로 전송한다. 그 다음 전송된 통계 어플리케이션은 클라이언트에서 실행되고 그 결과 값을 브라우저에 출력한다. (그림 5)는 통계학습시스템에서 클라이언트 기반으로 통계량을 계산하는 예이다. 이 시스템에서의 통계량 계산 프로그램은 자바스크립트 언어를 사용하여 구현하였다.



(그림 5) 통계학습시스템에서 클라이언트 기반의 통계량 계산

### 5. DB 자료를 활용한 통계 프로그램의 설계 방안

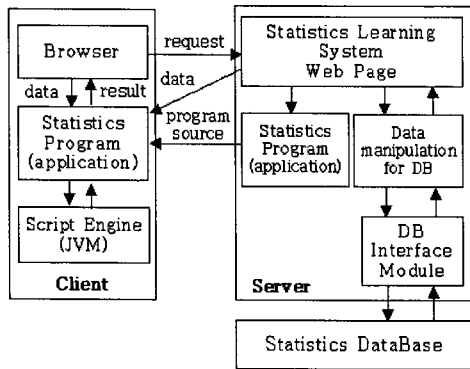
지금까지 웹상에서 통계 분석을 위한 여러 유형을 고찰하였다. 각 유형별로 장단점을 가지고 있어 어떤 유형의 통계 분석 형태가 바람직하다고 할 수 없다. 따라서 이곳에서는 앞에서 언급한 클라이언트 기반의 통계 프로그램 유형에 데이터베이스의 자료를 이용할 수 있는 통계학습시스템을 제안한다.

#### 5.1 시스템 구성

제안 시스템은 (그림 6)과 같이 제어 흐름으로 수행되며, 서버는 통계학습시스템을 구성하는 웹 페이지, 통계분석 Applet 프로그램, 데이터베이스 데이터를 조작하는 프로그램, DB를 인터페이스하는 모듈(ASP 프로그램), 그리고 데이터베

이스 서버로 구성된다. 클라이언트는 브라우저와 스크립트엔진(또는 JVM)로 구성된다.

사용자는 통계학습시스템에서 학습하고자 하는 단원에서 통계 분석 모듈을 클릭하면, 애플릿으로 되어 있는 통계 응용프로그램이 클라이언트로 전달된다. 전송된 애플릿이 실행되면, 애플릿 창에 데이터를 입력하고 분석 버튼을 클릭한다. 입력된 데이터는 애플릿 프로그램에 의해 통계분석을 하여 결과를 보여 준다. 데이터베이스의 자료를 이용할 경우에는 통계학습시스템의 웹 페이지에 마련된 “데이터베이스 이용하기” 링크를 클릭하여 입력, 수정, 삭제 등의 조작을 할 수 있으며, 데이터베이스의 자료를 조회하여 데이터를 포커싱(focusing), 군집(clustering), identify 등의 조작으로 통계 프로그램에 이용할 수 있다.



(그림 6) DB 자료를 활용한 클라이언트 기반의 통계 프로그램 흐름도

서버-클라이언트 환경에서 본 시스템의 데이터베이스 인터페이스 모듈은 ASP 프로그램으로 작성하였다. 통계 응용프로그램은 자바 언어로 작성하여 서버에 바이트코드 상태로 탑재하였으며, 클라이언트로 전송되어 클라이언트에서 실행할 수 있도록 애플릿으로 만들어 서버의 부담을 경감시켰다. 데이터베이스는 SQL Server 2000을 이용하였다.

### 5.2 사용자 인터페이스 설계

메뉴를 이용한 윈도우 형태를 통해 메인 화면에서 모든 작업을 수행할 수 있도록 풀다운 메

뉴(pulldown menu)로 설계한다. 메뉴에는 데이터, 계산, 통계, 그래프, 도움말 등으로 구성하며 각 메인 메뉴에는 서브메뉴로 구성한다. 이 시스템은 사용자의 마우스에 의한 작업으로 대부분을 조작할 수 있는 GUI 방식을 채택한다.

### 5.3 데이터 입력

데이터 메뉴에서는 텍스트 입력 방식을 이용하여 사용자가 직접 데이터를 입력할 수 있으며, 여러 조작을 통해 서버의 데이터베이스 자료를 자료 입력창에 가져올 수 있도록 설계한다. 물론 데이터베이스의 자료를 이용하지 않을 경우에는 사용자가 입력한 자료를 이용하여 클라이언트에서 통계 분석을 할 수 있다.

### 5.4 구현 예

본 논문에서는 제안한 시스템 중 통계 프로그램을 서버에 탑재하고 데이터를 클라이언트에서 입력하여 통계분석을 하는 클라이언트 기반의 통계 프로그램 방식으로 한 예를 구현한다.

예를 들어, p-p plot(Probability-Probability plot)은 자료가 어떤 특정 분포에 적합한지를 판단하는 여러 방법 중의 하나이다. 일반적으로 누적분포함수는 S형 곡선으로 나타나게 되는데, p-p plot은 가정된 분포에 의한 좌표 변환에 의해 S형 곡선을 만들고 이를 그려 보는 방법이다. 그러므로 자료가 가정 분포에 따른다면 직선상에 가깝게 나타날 것이며, 가정된 분포를 따르지 않는다면 직선으로부터 크게 이탈하게 된다.

n개의 표본을  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 이라 가정된 분포의 누적분포함수를  $F(x)$ 라 하고, 한 자료를 크기 순으로 정렬하여 순서 통계량을 구하여  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 으로 표시하면,

$$E[F(X_{(i)})] = \frac{i}{(n+1)}, i = 1, 2, \dots, n$$

이 성립한다. 이러한 결과를 이용하여 n개의 순서쌍을 다음과 같이 만들어 이를 그린 그림이 p-p plot이다.

$$\left(F(x_{(1)}), \frac{1}{n+1}\right), \left(F(x_{(2)}), \frac{2}{n+1}\right), \dots, \left(F(x_{(n)}), \frac{n}{n+1}\right)$$

그러므로 p-p plot은 (0,1)을 표시할 수 있는 정사각형 위에 자료들이 가정된 분포를 만족한다면 45도 각도의 직선상에 그려진다.

(그림 7)은 p-p plot을 그리기 위한 알고리즘이다[2].

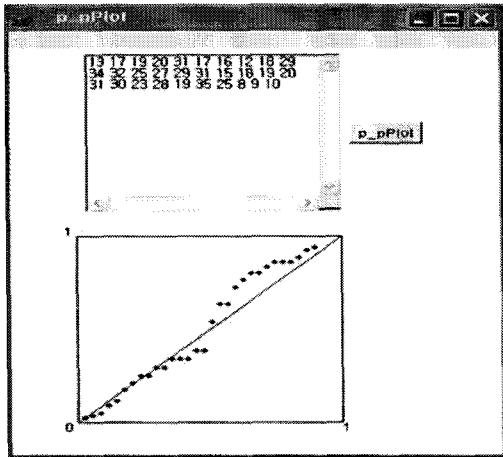
1. 입력된 자료로부터 평균( $\mu$ )과 분산( $\sigma^2$ )의 추정값을 구한다.
2. 자료를 순서 정렬한다.
3. 1에서 구한 추정값으로 자료를 표준화시킨다.  

$$z_{(i)} = \frac{x_{(i)} - \mu}{\sigma}$$
 (\*  $\mu$ 와  $\sigma$ 는 평균과 분산의 추정값임)
4. 정규분포로부터 누적확률을 구한다.  

$$p_{(i)} = \Pr(Z < z_{(i)})$$
5. 가로축에  $\frac{i}{n+1}$ , 세로축에  $p_i$ 를 값을 갖는 좌표의 점을 표현하고 그 점들과 (0,0)부터 (1,1)의 점을 연결한 직선과 비교한다.

(그림 7) p-p plot을 그리기 위한 알고리즘

(그림 8)은 사용자가 입력창을 통해 자료를 입력, 수정, 삭제 등의 조작으로 앞에서 언급한 p-p Plot 알고리즘을 통해 구현한 결과이다.



(그림 8) 애플릿으로 구현한 p-p Plot

## 6. 결론 및 고찰

웹상에서 통계 프로그램을 이용하는 방법에 대하여 조사하고 검토하여 보았다. 웹상의 특정 사이트에서 통계 프로그램을 다운로드하여 설치하고 실행할 수 있는 DAVIS는 클라이언트에서 독립적으로 동작하는 방식이므로 서버에 부담을 주지 않아 처리 속도 및 서버의 자원 절약 등에서 효율적인 시스템이다. 그러나 데이터 형식이 Weka에서 사용하는 arff를 따르고 있어 데이터 투입에 있어 약간의 어려움이 있다.

서버에서 실행하는 통계 프로그램의 유형은 인터페이스 프로그램만 잘 개발하면 데이터베이스 및 다른 프로그램 등을 쉽게 이용할 수 있는 장점이 있지만, 상용인 경우에 저작권 등에 문제점이 생길 수 있다. 상용의 소프트웨어가 아니라도 서버에는 부담을 주게 되어 서버 운영에 여러 가지 문제가 생길 수 있다.

통계 프로그램은 서버에 두고 클라이언트의 요청에 의하여 해당 프로그램 소스를 클라이언트로 전송하여 클라이언트에서 실행하는 형태는 처리 속도가 빠르고 서버에 부담을 주지 않는 반면에, 프로그램 소스가 공개되는 단점이 있다.

그러므로 본 연구에서는 서버에 있는 프로그램이 클라이언트로 전송되어 실행되는 애플릿을 이용하여 통계를 분석할 수 있고, 통계 데이터는 서버의 데이터베이스나 사용자가 제공하는 것을 이용할 수 있는 시스템을 제안하였다. 본 시스템은 애플릿을 이용하였기 때문에 프로그램에 대한 보안이 유지되며, 데이터베이스의 대량의 자료를 이용할 수 있다.

현재까지는 통계에 전반에 걸쳐 통계학습시스템을 개발 중에 있으며, 시스템이 완성될 시점에 웹상에 구축할 예정이다.

### 참고문헌

- [1] 정남철, "인터넷에서 통계학습시스템 설계 및 구현", 박사학위논문, 성균관대학교, 1996.
- [2] 최병수 외 2, 통계계산 자바프로그램, 자유아카데미, 서울, 2003.
- [3] Asimov, D., "The grand tour: A tool for viewing multidimensional data", SIAM Journal of Scientific and Statistical Computations, 제6권, 제1호, pp. 128~143, 19

85.

[4] Becker, Richard A., Cleveland, William S., and Wilks, Allan R., Dynamic Graphics for data Analysis, Wadsworth & Brooks/Cole, New York, 1988.

[5] Cleveland, W.S. and McGill, R., Dynamic Graphic for Statistics, Chapman & Hall, New York, 1997.

[6] Friendly, M., "Running SAS on the Web", <http://pascal.math.youku.ca/SCS/Online/sasweb/>, 1995.

[7] Helberg, C., "Statistics on the Web", <http://www.rds.u.wisc.edu/~helberg/statistics.html>, 1996

[8] Huh, Moon Yul, "Exploring Multidimensional Data with FEDF", Journal of Computational and Graphical Statistics, 제4권, 제4호, pp. 335~343, 1995.

[9] Hur, Moon Yul and Song, K.Y., "DAVIS: A Java-based Data Visualization System", Computational Statistics, vol. 17-3, pp. 411-423, 2002.

[10] Tierney, Luke, LISP-STAT : Objected Oriented Environment for Statistical Computing and Dynamic Graphics, Wiley, New York, 1990.

[11] Wegman, E. J., Luo, Q., "High Dimensional Clustering using Parallel Coordinates and the Grand Tour", Computing Science and Statistics, vol.28, pp. 61-368, 1997.

[12] Wegman, Edward J., "Data Mining and Visualization", Bulletin of the International Statistical Institute, ISI, Proceedings Book 3, pp. 223-226, Helsinki, Finland, 1999

[13] Weisberg, S. and Cook, R.D., An Introduction to regression Graphics, John Wiley & Sons, Inc, New York, 1994.

[14] <http://www.ics.uci.edu/~mlearn/MLRepository.html>



### 정 남 철

1977년 : 경인교육대학교 졸업  
 1983년 : 광운대학교 전자계산학과  
 1987년 : 성균관대학교 경영대학원  
 정보처리학(경영학석사)  
 1996년 : 성균관대학교 대학원 전  
 산통계학전공(경제학박사)

1983년~1989년 : 수협중앙회 전자계산소  
 1989년~1992년 : 세계일보 전산정보국  
 1992년~현 재 : 동남보건대학 웹컨텐츠과 부교수  
 관심분야 : 교육용컨텐츠, 인터넷방송, 멀티미디어  
 응용, 전자상거래시스템, 통계학습시스템