

On the Development of Statistical Software using Microsoft COM Technology on the Internet Environment

Kyung-Soo Han¹⁾ Seok-Ki Kim²⁾ Sook-Hee Choi³⁾

Abstract

Widespread of internet causes rapid changes in many areas of statistics such as statistical computation and education. In this paper, the efficient methods for statistical software development on the internet environment are introduced. Also the developed examples about Monte-Carlo permutation tests for two-way contingency tables using Microsoft COM (Component Object Model) technology are presented.

1. 서 론

지난 20여년 동안 컴퓨터 하드웨어 및 소프트웨어 기술의 빠른 발전은 통계학 교육과 연구에 많은 변화를 가져왔다. 이론 위주의 통계 교육을 보완하기 위하여 많은 대학들은 강의실에서 컴퓨터를 이용한 통계 자료 분석 과목을 설정하고 있다. 또한 통계 비전공자들을 위한 교양 통계 교육의 중요성과 수학적 이해에 바탕을 둔 통계 교육의 어려움을 인식한 통계 교육용 소프트웨어들이 개발되고 있다. 뿐만 아니라 컴퓨터는 통계 이론 연구 결과의 검증에 필요한 모의실험이나 계산집약적(Computational Intensive) 통계이론을 연구할 때 필수적인 도구로 인식되고 있다.

사용하기 쉬우면서도 정보 교환이 빠른 인터넷의 확산은 통계 교육과 연구를 포함한 많은 통계 관련 부분에 변화를 주고 있으며 인터넷 환경을 효과적으로 활용하기 위한 다양한 시도가 이루어지고 있다. Symanzik와 Klinke(1998)는 통계학 분야에서 인터넷 기술의 사용 실태와 가능성을 요약 페이지, 저널, 소프트웨어, 교육 등의 분야별로 소개하였다. 또한 그들은 통계 소프트웨어의 배포와 통계 교육에서 인터넷 활용의 유용성을 강조하고 있다. West, Ogden과 Rossini(1998)는 인터넷에서의 통계 소프트웨어 활용 방안을 논의하였으며, CGI(Common Gateway Interface) 기법과 Java를 사용하여 개발한 통계 자료 분석 예제들을 GASP(URL:GASP)에서 찾아 볼 수 있다.

West와 Piegorsch(1996)는 Fortran 또는 S-Plus를 사용하여 통계 계산 함수나 소프트웨어를 제작할 경우에 통계 비전공자가 연구 결과물을 쉽게 활용하기 어려운 단점이 있다고 지적한다. 또한 그들은 통계 전공자라 하더라도 연구자의 개발 환경과 사용 환경이 다르다면 문제가 발생할 수 있다고 주장하면서 가능한 해결책으로 인터넷의 활용을 제시하였다. 그러나 인터넷을 활용하는 기술은 매우 빠르게 변화하고 있기 때문에 새로운 기술을 습득하여 활용하기란 그리 쉽지 않다.

본 논문은 인터넷 환경에서 소프트웨어 객체 조립 기술을 통계 소프트웨어 개발에 활용하는 기

1) Associate Professor, Division of Mathematics and Statistical Informatics, Chonbuk National University, Chonbuk, 561-756, Korea

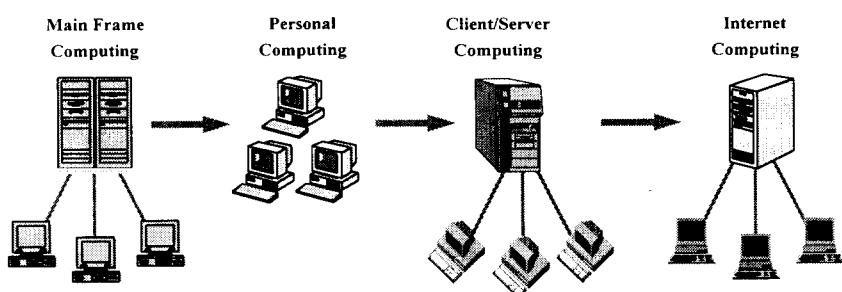
2) Researcher, Nursing Information Engineering Center, Kunsan College of Nursing, Kunsan, Chonbuk, 573-440, Korea

3) Professor, Dept. of Computer Science and Statistics, Woosuk University, Chonbuk, 565-701, Korea

법들을 비교하여 설명하고 예제 소프트웨어를 개발하는 과정을 소개한다. 2절은 지난 20여년 동안 컴퓨터 하드웨어 환경 변화에 따른 통계 계산 환경 변화를 다룬다. 컴퓨터 하드웨어의 급속한 발전과 비교하여 '소프트웨어 위기(Software Crisis)' 인식에서 나타난 객체지향 프로그래밍(Object-Oriented Programming) 기법들의 개발에 따른 통계 소프트웨어 개발 방법의 변화는 3절에서 다루어 진다. 4절은 마이크로소프트사의 소프트웨어 객체 조립 기술인 COM(Component Object Model)을 이용하여 2원 분할표의 소표본 검증법에 필요한 소프트웨어를 개발하는 과정을 간단히 소개한다.

2. 통계 계산 환경의 변화

통계 계산 환경의 변화는 컴퓨터 하드웨어 환경의 변화로 설명할 수 있으며 이러한 환경의 변화는 불과 20여년 만에 [그림 1]과 같은 변천 과정을 거쳤다. 이러한 변화를 주시하여야 하는 이유는 통계학의 모든 부분에서 컴퓨터를 배제할 수 없기 때문이다. 특히 통계 계산이나 교육 부분은 변화하는 환경과 기술 위에서 수행되는 소프트웨어에 의존한다. 따라서 변화하는 환경에 대한 인식과 변화할 환경에 대한 준비가 필요하다.



[그림 1] 컴퓨터 사용 환경의 변화

1970년대 이전에는 컴퓨터 하드웨어를 제조한 업체 중심의 독점적 기술들이 모든 컴퓨터 사용 환경을 결정하였다. 이는 IBM 메인프레임을 구입하면 IBM 하드웨어에 맞는 소프트웨어를 구입해서 사용해야 함을 뜻한다. 모든 자원을 메인프레임에 의존하는 터미널을 이용해야 했다. 따라서 메인프레임에 설치된 SAS, SPSS 등과 같은 소프트웨어를 모든 사용자가 같이 사용하기 때문에 사용자 수가 증가하거나 제공하고자 하는 서비스 프로그램이 증가할 때마다 성능을 증대 시켜야만 하는 단점이 있었으나 유지보수가 쉽다는 장점도 있었다.

1980년대 후반부터 하드웨어가 급격히 발전하면서 컴퓨터의 보급이 확산되고 개인용 컴퓨터에서 사용할 수 있는 소프트웨어가 기하급수적으로 증가하게 되어 개인용 컴퓨터 시대가 열렸다. 1992년 Microsoft사의 Windows 3.1의 출시로 그래픽 사용자 환경(GUI; Graphic User Interface)이 도입되기 시작하였고, Windows 95가 보편화되면서 개인용 컴퓨터의 사용자 환경이 급속도로 그래픽 환경으로 변화하게 되었다. 고성능 마이크로 프로세서가 개발되면서 메인프레임에서나 사용이 가능하던 통계 소프트웨어들이 개인용 컴퓨터에서 사용할 수 있게 되었다. 또한 개인용 컴퓨터 환경에 적합한 새로운 통계 소프트웨어의 개발도 많아져서 많은 통계 처리가 개인용 컴퓨터에서 이루어졌다.

간편한 네트워크 설정을 지원하는 Windows 95의 보편화로 값싼 네트워크 환경 구축이 가능하게 되어 서버의 빠른 계산 능력을 개인용 컴퓨터에서 이용할 수 있는 길이 열렸다. 네트워크에 연결되지 않은 컴퓨터는 모든 계산을 자신의 컴퓨터에서만 실행해야 하기 때문에 통계 소프트웨어의 계산 능력이나 속도는 사용자가 보유한 컴퓨터의 성능에 좌우된다. 따라서 빠른 통계 계산을 위해서는 고성능 컴퓨터가 생산될 때마다 자신이 보유한 컴퓨터의 성능을 향상시켜야만 했으며 방대한 양의 데이터나 통계 계산이 복잡한 문제는 저 성능의 개인용 컴퓨터에서 처리가 곤란하여 UNIX를 운영체제로 사용하는 월스테이션과 같은 고성능 컴퓨터가 필요했다.

사회 환경의 급속한 발전에 따라 처리해야 할 자료의 양이 방대해져서 자료 수집과 처리, 수집된 자료의 분석 결과를 제공하기 위하여 고성능 서버와 기존의 개인용 컴퓨터를 클라이언트로 사용하는 Client/Server 환경이 나타났다. Client/Server 환경에서는 클라이언트의 다양한 하드웨어와 운영체제에 적합한 소프트웨어를 개발하여 사용해야만 한다. 이러한 환경은 기존의 소프트웨어 개발 방법과 사용 환경에 변화를 주지 않고 계산 능력의 효율을 높일 수 있다는 장점을 가지는 반면에 소프트웨어 개발과 유지보수가 어렵고 비용이 많이 든다는 단점이 있다.

1995년 Netscape의 웹 브라우저인 Navigator가 등장하면서 네트워크 환경은 급속도로 인터넷 환경으로 변화하였다. 클라이언트에 웹 브라우저만 있으면 웹 서버가 제공하는 모든 서비스를 이용할 수 있게 된 것이다. 인터넷을 기반으로 한 계산 환경의 특징은 클라이언트의 웹 브라우저는 웹 서버 접속과 관련된 간단한 인터페이스만을 구현하고 고성능의 서버는 다양한 데이터를 처리하거나 복잡한 계산을 수행하는 공간에 제약받지 않는 역할 분담이라 할 수 있다.

이와 같은 계산 환경의 극심한 변화를 통계계산 분야도 같이 겪어 왔다. 통계자료분석이 컴퓨터가 없이는 가능하지 않고, 교육 환경도 멀티미디어 환경으로 급격히 변화하고 있기 때문에 통계학의 각 부분에 컴퓨터를 활용하기 위해서는 현재의 인터넷을 기반으로 한 계산 환경을 예의 주시할 필요가 있다. 인터넷 환경에서 통계 계산이나 교육을 위하여 CGI나 Java를 사용한 소프트웨어의 개발 노력이 활발하게 이루어지고 있는 것이 대표적인 예라 할 수 있다.

3. 통계 소프트웨어 개발 방법의 변화

컴퓨터 하드웨어의 급격한 발전에 비하여 소프트웨어의 발전은 너무나 느려서 '소프트웨어 위기 (Software Crisis)'가 발생하였다. 소프트웨어 개발이 하드웨어의 급격한 발전에 비하여 늦어지는 이유를 컴퓨터 전문가들은 기존에 만들어진 소프트웨어 모듈을 재사용하지 않고 중복 개발하기 때문이라고 지적한다. 이전에 개발한 소프트웨어 모듈을 재사용하기 위하여 C++와 같은 객체지향 프로그래밍(OOP; Object-Oriented Programming) 언어가 탄생했다. 그러나 C++에서의 소프트웨어 모듈의 재사용성은 프로그램 원시코드의 재사용을 의미한다. 따라서 원시코드를 구할 수 없으면 재사용이 가능하지 않다. 또한 객체지향 언어가 아닌 Fortran이나 Basic과 같은 언어로 개발된 모듈은 재사용하기 어렵다. 문제는 현존하는 방대한 양의 통계계산 알고리즘들이 구조적 언어를 사용하여 구현되어 있다는 것이다.

재사용성 문제를 해결하기 위하여 Microsoft 사는 1995년 분산된 컴퓨터 환경에서 다양한 언어로 개발된 2진 코드를 재사용 할 수 있는 COM(Component Object Model)을 제안하였다. COM은 하나의 응용 소프트웨어를 개발할 때 이미 개발된 다양한 형태의 소프트웨어 모듈을 블록을 쌓듯이 조립하여 빠르고 쉽게 개발할 수 있게 한다. 이러한 기법은 사용하는 각각의 모듈이 이미 검증

된 것이므로 안정적이고 효율적이라 할 수 있다.

Visual Basic을 이용하여 통계자료 분석용 소프트웨어를 만들 때 COM 기술을 이용하면 무척 간단히 개발할 수 있다. 데이터 입력 부분은 Visual Basic에서 제공하는 FlexGrid 컴포넌트를 이용하고 그래프는 Chart 컴포넌트를 이용한다. 통계계산 부분은 Excel에서 제공하는 통계 함수를 이용하면 단지 몇 시간만에 간단한 기능을 가지는 통계자료분석 소프트웨어를 만들 수 있다. 이와 같은 방법으로 통계 소프트웨어를 개발하면 기초적인 지식만 가지고도 빠른 시간 안에 자신이 만든 통계 소프트웨어를 이용하여 통계자료를 처리할 수 있을 것이다. 또한 피교육자가 직접 소프트웨어를 제작하는 과정을 통하여 통계 이론을 보다 효과적으로 습득할 수 있을 것이다.

이러한 기법을 네트워크 중심의 다수의 고성능 서버가 존재하는 분산 컴퓨팅(Distributed Computing) 환경에 적용하면 통계 기법을 획기적으로 보급할 수 있다. 통계학자들이 제안하는 대부분의 통계 분석 기법들은 사용자로부터 데이터를 입력받아 통계 분석 결과를 제공하기 때문에 필수적으로 프로그램이 필요하다. 따라서 많은 통계학자들이 Fortran, S-Plus, SAS, Matlab 등을 이용하여 분석한 결과를 공개한다. 그러나 통계 비전공자가 통계학자들이 분석이나 제시한 이론의 검증을 위하여 개발한 소프트웨어 모듈을 사용하기 위해서는 저자가 개발한 소프트웨어 환경을 알아야 이용할 수 있기 때문에 응용분야에 통계 기법을 전파하기는 매우 어렵다.

여기서 하나의 대안으로 떠오르는 것이 바로 인터넷 기반의 계산 환경이다. 사용자는 복잡한 통계전문 소프트웨어를 이용하지 않고도 그들의 데이터를 웹 브라우저에서 입력하여 바로 통계분석 결과를 얻을 수 있고 저자의 분석 결과를 직접 확인할 수 있다. 물론 기존에 널리 알려진 CGI 기법을 이용하여 개발할 수도 있다. CGI 기법은 기존에 널리 알려진 언어를 사용하기 때문에 쉽게 배울 수 있다는 장점을 가지고 있으나 여러 사용자들에게 동시에 서비스하기 위하여 같은 프로그램이 서버 컴퓨터에 서비스하고자 하는 수량만큼 로드 되어야 한다. 따라서 과부화가 발생되면 서버가 다운될 우려가 있다. 또한 모듈의 재사용이 어려워 서버에 과중한 부담을 주게 되어 결과적으로 고성능 서버의 낭비 요인이 된다.

Java로 개발하면 운영 환경에 무관하다는 장점을 가진다. 그러나 Java로 개발된 모듈은 사용자의 컴퓨터에 설치된 후 사용되므로 수행을 위한 부가적인 시간이 요구되며, 사용자가 보유한 컴퓨터의 성능이 좋지 않다면 계산 성능이 문제되는 단점이 있다. 한경수, 안정용, 강윤비(1998)는 Java를 이용하여 통계학 개념을 교육할 수 있는 프로그램(URL:CyberStat)을 개발하였으며, West와 Ogden(1997)은 Java를 사용하여 통계자료분석과 그래프를 제공하는 프로그램(URL:WebStat)을 개발하였다. 그러나 다른 언어로 개발된 그래프나 수치 계산 라이브러리 등을 사용하기 어려워 개발 시간이 오래 걸리는 단점이 있다. [표 1]에서 볼 수 있듯이 컴포넌트 기술이 통계 계산 모듈을 보다 효과적으로 재사용할 수 있는 인터넷 기술이라 할 수 있다.

[표 1] CGI, Java, 컴포넌트 기술의 비교

	CGI	Java	컴포넌트 기술
플랫폼 의존도	의존적	독립적	기술에 따라 다름
수행 위치	웹 서버	클라이언트	조절 가능
재사용성	소스의 재사용	클래스의 재사용	객체의 재사용
과부화 가능성	사용자 · 서비스 증가시	없음	없음

앞으로의 계산 환경은 사용자의 컴퓨터가 고성능이 아니더라도 네트워크에 연결된 여러 고성능 서버 컴퓨터들을 이용하여 작업을 빠른 시간에 효율적으로 마칠 수 있는 분산 컴퓨팅이 될 것으로 많은 컴퓨터 전문가들은 예측한다. 분산 컴퓨팅 환경에서 소프트웨어 개발 방법은 다음 3가지 기법이 대표적이다.

1) Microsoft COM

Microsoft가 제안한 객체기반 기술인 OLE(Object Linkage and Embedding)를 분산 컴퓨팅 환경에 적합하도록 확장시킨 것으로 Windows 환경에서 개발되어 현재 UNIX 환경에서도 운용 가능하도록 개발 중에 있다. Microsoft Windows 자체가 COM 컴포넌트들로 구성된 하나의 응용 프로그램으로 생각할 수 있을 정도로 많은 부분들이 COM을 지원하는 객체 형태로 이루어져 있다. COM의 특징은 COM 객체의 위치가 클라이언트 컴퓨터, 로컬 서버 또는 원격 서버에 관계없이 동일한 참조 환경을 제공하는 위치 투명성(location transparency)의 특성을 가진다는 것이다.

2) OMG CORBA(Object Management Group Common Object Request Broker Architecture)

CORBA는 하드웨어에 독립적인 소프트웨어의 개발과 보급을 목적으로 대학, 연구소, 산업체 등으로 구성된 비영리 컨소시엄인 OMG에서 1989년 제안하였다. CORBA는 벤더 중심적인 소프트웨어 개발 환경을 극복하기 위한 분산 객체 기술로 각 벤더들의 UNIX 환경에 의존하지 않고 독립적으로 수행할 수 있는 소프트웨어 개발에 목적을 두고 있다. Bell 연구소의 Computing and Mathematical Sciences Research Division(URL:CMSRD)에서는 S, Xlisp-Stat, Matlab 등 기존에 통계 계산을 위해 사용되던 소프트웨어의 데이터 오브젝트를 CORBA 환경에서 사용하기 위한 연구와 데이터 기반의 분산 컴퓨팅에 관한 연구가 진행되고 있다.

3) Sun Microsystems JavaBeans

Sun Microsystems가 Java로 개발된 컴포넌트의 재사용을 목적으로 1996년 제안한 기술로 Java를 사용하여 개발하고 있기 때문에 Java의 장점인 플랫폼에 중립성과 단순성을 그대로 상속 받았다. 엄밀히 말해서 JavaBeans는 분산 객체 컴포넌트는 아니지만 네트워크와 상관없이 지역적으로 빈을 조작하고 통신하므로 분산 네트워크 환경을 활용한 컴포넌트 기술로 볼 수 있다. 또한 JavaBeans-CORBA 브리지를 통해 CORBA와 호환되며, Microsoft의 인터넷 컴포넌트인 ActiveX 컴포넌트를 사용할 수 있도록 지원하고 있다.

[표 2] COM, CORBA, JavaBeans의 비교

	COM	CORBA	JavaBeans
제 안	Microsoft	OMG	Sun Microsystems
특 징	2진 코드의 재사용	분산 객체 기술의 표준	플랫폼 중립적
서버 운영체제	Windows, UNIX	UNIX	Windows, UNIX
사용언어	Mixed Language	IDL, C, C++, Smalltalk	Java
관련자료	http://www.microsoft.com/com	http://www.omg.org/corba	http://java.sun.com/beans

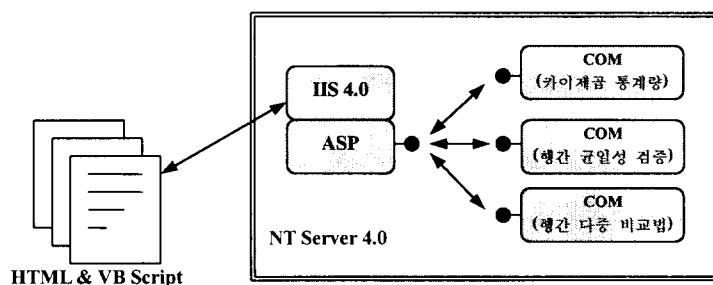
앞에서 언급한 기법들의 특징을 간단히 정리하면 [표 2]와 같다. 이들 기법들 중에서 어떤 기법이 가장 우수한 분산 객체 기술이고, 미래의 분산 컴퓨팅 환경을 좌우하는 기술이 될지 아무도 예측할 수 없다. 다만 중요한 것은 이들 기법들이 서로의 호환을 위하여 노력을 계속하고 있고 기능의 통합도 준비중에 있다는 사실이다. 따라서 이러한 기법들을 활용한 통계 소프트웨어를 개발하여 통계 교육에 이용하거나 통계 기법을 응용 분야에 활용하는 것이 인터넷 환경과 분산 컴퓨팅 환경에 대비하여 요구된다고 할 수 있다.

4. 개발 예제

통계 자료를 분석하려는 사용자가 웹 브라우저에서 데이터만 입력하면 고성능 서버가 통계 계산을 처리한 후 웹 브라우저에 결과를 되돌려주는 분산 객체 프로그래밍 기법을 혀명회(1997)의 논문 "2원 분할표의 소표본 검증법"에 적용하였다. 혀명회(1997)는 임의 2원 분할표 생성을 위한 몬테칼로 알고리즘에 기초하여 소표본 카이제곱 검증 방법을 제안하였다. 그러나 제안된 기법을 활용하기 위해서는 컴퓨터 프로그램이 필요한데 통계학이나 컴퓨터의 비전문인이 프로그래밍 하기는 매우 어려운 문제이다. 저자가 프로그래밍한 SAS/IML 프로그램으로 처리하고 싶어도 SAS가 없거나 SAS를 다루어 본적이 없다면 이 또한 무용지물이다.

본 논문에서 개발된 프로그램에서는 누구나 웹 브라우저에서 자신의 자료만 입력하면 2원 분할표의 검증 결과를 얻을 수 있다. 사용자는 자신의 컴퓨터에 사용할 줄 모르는 새로운 통계 자료 분석용 소프트웨어를 설치할 필요가 없다. 어떤 종류의 컴퓨터라도 웹 서버에 접속할 수 있는 브라우저만 있으면 모든 준비는 충분하다. 사용자의 컴퓨터가 성능이 좋지 않더라도 서버에서 모든 계산이 이루어지므로 고성능 서버의 빠른 통계계산 이점을 누릴 수 있다.

개발한 예제의 구조는 [그림 2]와 같이 통계 계산에 필요한 소프트웨어 모듈들을 COM 컴포넌트로 제작하여 NT Server에 설치하였다. 인터넷 사용자가 통계 계산을 요구하면 IIS(Internet Information Server)와 연동되어 있는 ASP(Active Server Page)가 요청한 계산에 필요한 COM 컴포넌트를 호출하여 계산을 수행하고, 결과를 HTML 형태의 문서로 변환하여 사용자의 웹 브라우저에 나타낸다.



[그림 2] 개발된 프로그램의 구조

향후 추가적인 계산의 필요나 계산 알고리즘의 변화로 COM 컴포넌트의 변환이 필요할 경우 해당 컴포넌트를 교체하거나 새로운 컴포넌트를 제작하여 NT 서버에 탑재할 수 있다. 즉 사용자 인

터페이스와 계산 수행 부분의 분리로 프로그램의 변경이나 추가가 용이하도록 구성하였다.

사용한 프로그램 개발 도구는 [표 3]과 같으며 개발 및 운영환경은 [표 4]와 같다. 다양한 개발 도구를 사용할 때의 단점은 각 도구들의 사용법을 배우는 것이 힘들다는 것이지만 한 번 익숙해지면 각 소프트웨어의 장점을 살려 소프트웨어 개발을 쉽고 빠르게 할 수 있다는 장점이 있다.

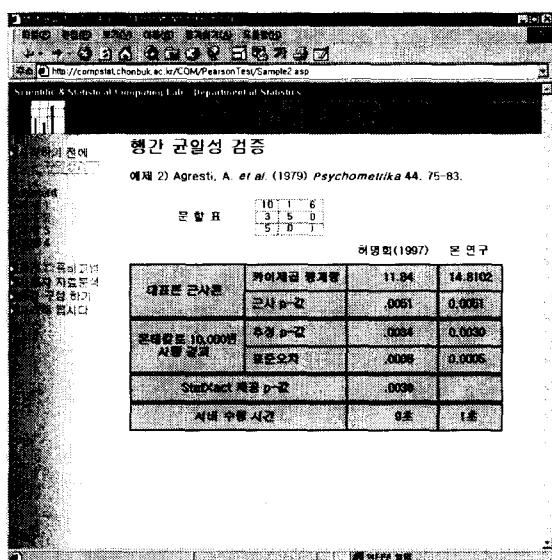
[표 3] 프로그램 개발에 사용된 도구

개발 도구	담당 부분
Digital Visual Fortran & IMSL Library	통계 계산 알고리즘 작성
Microsoft Visual C++ (ATL)	COM 인터페이스 작성
Microsoft Visual Basic Script	웹 페이지와 COM의 인터페이스 담당
Microsoft FrontPage 98	웹 페이지 디자인

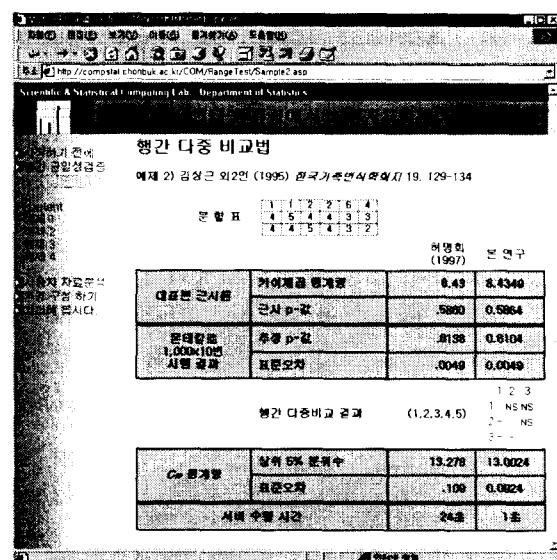
[표 4] 프로그램 개발 및 운영 환경

제품명	담당부분
Microsoft Windows NT Server 4.0	COM 운영을 위한 서버의 운영체제
Microsoft Internet Information Server 4.0	웹 서버의 운영체제
ASP(Active Server Pages)	웹 서버와 COM의 연동

NT Server 4.0에서 운영되는 웹 서버인 IIS가 사용할 COM을 제작하기 위하여 Microsoft Visual C++가 제공하는 ATL(Active Template Library) 프로그래밍 기법을 이용하였다. DLL(Dynamic Link Library) 형태로 개발된 COM을 사용하면 복수 사용자가 동시에 서버에 접속하더라도 응용 프로그램은 서버의 메모리 공간에 한번만 로드 되므로 사용자 수가 증가하더라도 서버가 과부화 되지 않아 서버가 다운될 우려가 없다.



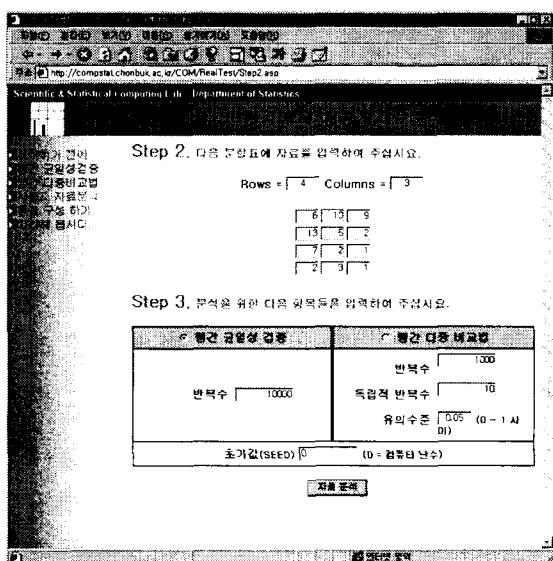
[그림 3] 행간 균일성 검증 예제



[그림 4] 행간 다중 비교법 예제

웹 페이지 개발은 Visual Basic Script와 FrontPage98을 사용하였다. ASP는 사용자가 자료분석을 위한 기초 데이터를 입력할 때 웹 서버가 동적으로 웹 페이지를 생성하는 데 이용된다.

[그림 3]과 [그림 4]는 허명희(1997)에서 사용한 행간 균일성 검정 사례 2와, 행간 다중 비교법 사례 2를 적용한 예제로 그림에서의 결과 화면은 실시간으로 계산된 결과를 제공하기 때문에 폐이지를 생신할 때마다 값이 변한다.



[그림 5] 사용자 입력 자료의 분석을 위한 설정 화면

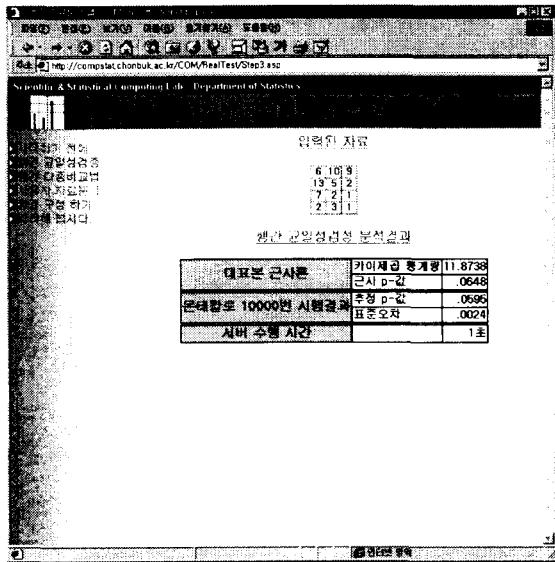
다른 2원 분할표 데이터를 사용하여 분석하려면 데이터 입력을 위한 웹 페이지에 데이터를 입력하면 실시간으로 분석할 수 있어서 사용법은 매우 간단하다. [그림 5]에서와 같이 분석 방법으로 허명희(1997)에서 제시한 분석 방법 중에서 행간 균일성 검정과 행간 다중 비교법을, 분석에 필요한 요소로 시뮬레이션 반복 횟수, 난수 발생기의 초기값, 유의수준을 사용자가 설정할 수 있다.

사용자가 입력한 자료에 대한 행간 균일성 검정의 분석 결과는 [그림 6]과 같으며, 행간 다중 비교법의 분석 결과는 [그림 7]과 같다.

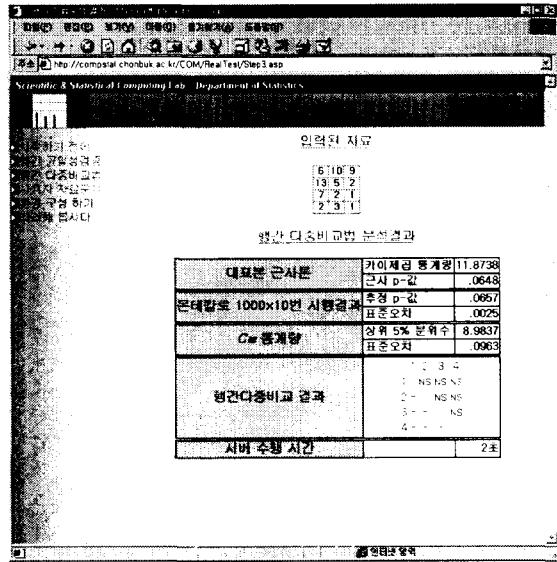
본 논문에서 사용한 방법이 COM을 활용한 가장 효과적인 소프트웨어 제작 방법이라고 단언할 수는 없다. 그러나 개발과정을 소개하는 이유는 컴포넌트들을 조합하는 방법만 알면 분산객체 프로그래밍이 생각보다는 쉽다는 것을 보여주기 위함이다.

개발에 앞서 먼저 COM과 웹을 운영할 서버가 준비되어져야 하며 다음 항목들이 설치되어 있어야 한다.

- Microsoft Windows NT Server 4.0
- NT Server Service Pack 3 or 4
- NT Option Pack



[그림 6] 사용자 입력 자료의 행간 균일성
검정 분석 결과



[그림 7] 사용자 입력 자료의 행간 다중
비교법 분석 결과

Option Pack을 설치하는데 있어서 주의할 사항은 설치 옵션에서 반드시 IIS와 FrontPage98 Server Extension을 선택해야 한다.

소프트웨어 개발 과정을 단계별로 간략히 설명하면 다음과 같다.

1. Digital Visual Fortran과 IMSL을 사용하여 통계 계산에 사용할 함수를 작성한다.
2. COM과의 통신을 위한 인터페이스 부분을 1에서 작성한 Fortran 코드에 삽입한다.
3. Microsoft Visual C++의 ATL COM AppWizard를 사용하여 COM을 작성하기 위한 프로젝트를 생성한다.
4. ATL Object를 생성하고, 2에서 작성한 컴포넌트를 프로젝트에 삽입한다.
5. COM을 호출할 클라이언트와의 인터페이스를 위하여 Member Variable, Method, Property를 정의한다.
6. COM 컴포넌트를 생성을 위해 프로젝트를 build한 후 생성된 DLL 파일을 서버에 등록한다.
7. FrontPage 98을 사용하여 HTML 페이지를 디자인한다.
8. VB Script를 사용하여 COM과 통신하기 위한 프로그램을 HTML 페이지에 삽입한다.

5. 결 론

본 논문에서는 인터넷 환경과 분산 컴퓨팅 환경에 적합한 통계 소프트웨어의 개발 방법을 논의하였고, 허명희(1997) 논문 중 일부를 Microsoft의 분산 컴퓨팅 모델인 COM 기술을 사용하여 개발하였다. 개발된 소프트웨어는 Netscape Navigator 또는 Internet Explorer가 설치되어 있고 인터넷을 사용할 수 있는 컴퓨터 환경이라면 추가적인 소프트웨어의 설치 없이 사용자가 입력한 자료를 실

시간으로 분석하여 결과를 제공할 수 있도록 구성하였다.

Symanzik, Klinke(1998)는 인터넷 기술들의 역사를 돌이켜 볼 때 현재의 기술 중 극소수만이 5년 내지 10년 동안 여전히 영향을 주는 기술로 남아있을 것이라고 예측한다. 즉 현재 주목받는 기술이 몇 년 뒤에는 무용지물이 될 수도 있다는 뜻이다. 본 논문에서 사용한 COM 기술이 분산 컴퓨팅 환경에 가장 적합한 기술이라고 단언할 수는 없다. 그러나 현재까지 개발된 통계 계산 소프트웨어나 구현된 알고리즘을 변화된 환경에서도 특별한 수정 없이 사용 가능하게 할 수 있는 개발 방법과 기술의 습득은 앞으로도 중요한 문제로 남아 있을 것이다.

대용량 데이터베이스의 통계분석과 permutation test, bootstrap, MCMC 등과 같이 계산 집약적인 통계적 방법들은 통계 계산을 위해 다중 프로세서(multiple processor) 또는 병렬 프로세서(parallel processor)를 사용한 서버 환경과 분산 컴퓨터 환경이 머지않아 도입될 전망이다. 따라서 통계 소프트웨어 개발에 있어서 이러한 컴퓨터 환경 변화에 대한 사전 준비가 필요하다.

본 논문에서 개발된 소프트웨어는 전북대학교 수학·통계정보과학부 Scientific & Statistical Computing Laboratory 홈페이지(<http://compstat.chonbuk.ac.kr>)에서 자유롭게 이용할 수 있고 개발에 관한 보다 상세한 정보를 얻을 수 있다.

References

- [1] 한경수, 안정용, 강윤비 (1998). 통계학 교육을 위한 전자교재의 활용, *한국통계학회 응용통계 연구*, 제11권 1호, 5-12.
- [2] 허명희 (1997). 2원 분할표의 소표본 검증법, *한국통계학회 응용통계연구*, 제10권 2호, 339-352.
- [3] Symanzik, J. and Klinke, S. (1998). Current Internet Technology - Blessing or Curse ?, *30th Symposium on the Interface : Computing Science and Statistics*, <http://www.galaxy.gmu.edu/~symanzik>.
- [4] West, R.W. and Ogden, R.T. (1997). Statistical Analysis with WebStat, a Java applet for the World Wide Web, *Journal of Statistical Software*, On-line Journal at <http://www.stat.ucla.edu/journals/jss/v02/i03>
- [5] West, R.W., Ogden, R.T. and Rossini, A.J. (1998). Statistical Tools on the World Wide Web, *The American Statistician*, Vol. 52, No. 3, 257-262.
- [6] West, R.W. and Piegorsch, W.W. (1996). Interactive Statistics on the Internet: Applications in Environmental Biology, *University of South Carolina Department of Statistics Technical Report*. No. 187. 62P10-4

URL References

- URL:CMSRD : <http://cm.bell-labs.com/>
- URL:CyberStat : <http://compstat.chonbuk.ac.kr/cyberstat/>
- URL:GASP : <http://www.stat.sc.edu/rsrch/gasp/>
- URL:Webstat : <http://www.stat.sc.edu/~west/webstat/>