

統計專門家시스템을 위한 統計處理過程의 工學的 接近 研究

車 弘 濬*

요 약

본 연구는 통계처리과정을 공학적인 접근으로한 인공지능화를 시도하려는 개념화에 있는데, 이는 전산학에서의 소프트웨어공학과 같이 통계지식공학으로 확대 적용하고, 더불어 그 관계를 일반화 해보려는 것이다. 그러므로 방법적인 도출은 통계전문가시스템 설계를 하는데 불변성으로 명확하게 나타내려는 것이 아니라 발전적으로 요구되는 편의를 위해서 유연한 대처로 했으며, 이를 위한 지식의 표현 방안을 확실히 구성했다. 따라서 이를 확대 적용할 수 있도록 공학적 처리 모형을 개념화 하여서 이러한 문제의 공학적인 접근 방안을 제안했다.

1. 통계처리의 공학적 접근 배경

인간은 어떤 대상이든 이를 분석하고 연구하는 통계적 방법을 추구하고, 이를 계속하여 발전시켜 왔으며, 특히 수리적 분석과 모형에 의한 실천적 응용을 비약적으로 발전시켜 타 학문분야에 까지 영향을 미치게 하였다.

따라서, 오늘에 이르른 통계학은 교육학, 심리학, 의학 및 공학연구에 그 응용도가 더 넓어지고 있는 반면, 컴퓨터의 발달에 부응하여 인간과 기계계(Man-machine Interface)를 개선하려는 많은 연구가 진행되어, 통계처리를 패키지(Package)화 함과 동시에 전문가시스템(Expert system)으로 발전시키게 되었다.

특히 개인용 컴퓨터(Personal computer)의 등장과 함께 컴퓨터에의 입·출력 조작기술의 진보와 기억장치의 대용량화는 자료를 모아서 정리해야 할 부담이 보다더 가볍고 자유로와 졌으며, 프로세서의 고속화 기술은 통계처리시간을 단축시켰을 뿐만 아니라 처리비용의 저렴화 실현을 이루었으며, 컴퓨터 시스템기술 발달은 기호처리를 가능화 시킴으로서, 앞으로의 통계처리 지식과 기술 및 언어를 지원하는 분야에 대해서 공학적인 접근을 필연적이게 했다.

실제 이 시점에서, 방대한 자료를 수집하여 정보를 추출하는 기상관리 시스템 통계처리는 컴퓨터의 네트워크 시스템에 의한 포집집단(Data acquisition field)의 선택에 구속받지 않고 자유롭게 의미적인 정보추출을 가능하게 하고 있으며, 감염증 진단과 진료를 위해서 의

*강원대학교 전자계산학과 부교수, 강원도 춘천시 효자1동 산192-1.

료기술의 지식기반(Knowledge rule-based)으로 두고서 통계처리를 응용하는 MYCIN시스템도 있지만, 앞으로 공학적인 자연어이해시스템(Natural language understanding system)과 기호신호에 의해 처리되는 하이퍼카드(Hyper card) 시스템 방식으로, 인공지능화 시킨 방법이 실현화될 경우, 통계처리의 정보나 처리방법을 컴퓨터와 대화(Interactive)로서 이루어질 수 있는 전문가시스템이 출현되어질 것이다.

그러나, 이러한 궁극적인 목적을 위한 통계전문가시스템은 통계 지식정보처리를 위한 사용자와의 대화라던가, 자료를 입력 할때마다 갖게되는 노작의 부담을 덜게하는 기능을 컴퓨터에 실현시키는데 있으므로, 사용자의 의도를 적절히 이해하는 과정을 컴퓨터 대화기능으로 도입 하여야 하는데, 그런것이 어떠한 이론과 도구를 이용하여 실현할 수 있을 것인가? 하는 문제가 제기된다.

따라서 “이해란 무엇인가?”라는 보편적인 문제로부터, “사용자가 의도하는 실제적인 자료를 계산기적-전자계산학적-으로 어떻게 표현시킬 것인가?”, “복합적인 지식원(Knowledge source)을 어떻게 소프트웨어적으로 도구화할 것인가?” 등에 대하여 새로운 접근법이 연구되어져야 할 것이며, 그 필요성이 더욱 요구되어 진다.

아직까지 통계처리를 컴퓨터에 의존하여 처리한다는 것이 이용이라는 범주에서 크게 벗어나지 못한 상태이므로 전문가시스템화(Expert system)의 새로운 시도라는 관점에서 볼 때 통계전문가시스템 연구의 학문적 규명과 시스템공학적 접근은 정보화 사회의 실현을 앞둔 시대적, 사회적, 학문적인 요청인 것이다.

2. 통계전문가시스템의 목표

통계전문가시스템(Statistical Expert System)이란 표현은 Portier와 Lai(1983)가 “분석 결정을 위한 통계전문가시스템”이란 논문발표 이후 Hand(1984)가 통계전문가시스템의 설계 방안을, Hahn(1985)은 지능적인 소프트웨어의 발전이 가져오게 될 미래의 지능통계전문가시스템에서 여러번 언급 되었지만 이렇다할 정의가 없는 실정이다.

그러므로, 이의 설명을 요약한다면 “전자계산학의 기술분야로서의 인공지능(Artificial intelligence)을 응용하여, 통계지식을 기반으로 하는 지식공학(Knowledge engineering)으로 완성시킨 시스템”이라 할수 있다.

바꾸어 말하면, 실생활에서 발생하는 자료를 수집·가공하여 통계적인 처리를 하는 컴퓨터시스템의 이론과 기술개발을 목적으로 하는 컴퓨터 응용공학이라 할 수 있으며, 궁극적으로는 일반인들이 통계처리를 할려고 할 때 1차적으로 통계지식에 대한 결핍에서 오는 두려움과 2차적으로는 어떠한 통계기법을 적용할 것인가? 또한, 그 결과를 해석하고 어떻게 이용할 것인가? 하는 문제를 해결하여 일반대중화를 실현할 수 있게 하려는 서비스를 목표하는 시스템이라 할 수 있다.

따라서, 통계전문가시스템의 목표는 순수한 전문가시스템공학적 측면으로 Data Base의 통계지식을 기반으로 한 지적인 검색지원 시스템으로서

- 사용자와 대화식의 통계처리(Statistics Processing)기능
- 사용자의 통계처리 지식과 언어 지원 시스템 기능
- 사용처리 편의를 위한 도구적인 시스템 기능

에 대한 이론의 정립에 따른 그 기술을 개발함으로써, 특별히 전문가적인 도움없이 간단히 컴퓨터시스템을 다루어지며, 실제적인 통계처리 작업을 할 수 있도록 실용성을 갖게하는 인간과 기계계의 Human Interface 구축에 있게 된다.

3. 접근연구에 대한 영역의 구성

통계전문가시스템의 연구는 인접 영역의 학문과의 연계성이 수반되어야 하는데 개략적인 영역은 다음과 같이 구분할 수 있다.

(1) 통계학적 영역

통계란 용어가 지니고 있는 내용은 수치적 자료 혹은 수량적 자료를 의미하며, 이에 대하여 그 집단의 개별적인 특성을 인식케하는 통계적 방법을 생각하게 되는데, 이같은 통계적 방법은 사회 현상이나 자연현상 전반에 걸친 집단과 집단간의 관계를 파악 분석하고, 거기서 한 법칙성의 존재여부를 논의하는데 아주 중요한 분석도구가 되는 분야이다.

(2) 지식공학적 영역

컴퓨터라는 인공지능정보처리 기계계에 통계처리 과정의 모형을 구축하는데 있어서 통계처리 구조를 설명하고, 통계처리 규칙의 가설에 따라 처리모형을 논리적인 접근방법으로 해명하고, 지식공학적인 관점에서는 인공지능기법을 활용하여 실제적인 통계처리과정을 공학적 문제로 해결할 수 있는 실질적인 공학시스템 개발을 중심 과제로 삼는 분야이다. 즉, 이것은 전문가 통계처리시스템(Expert Statistic Processing System)의 구현을 목적하는 것으로, 실제로 전문가시스템이 특징적인 집약적지식의 범주(Knowledge-intensive Domain)에서 문제를 풀도록하고, 이 방법의 확산을 꾀하므로(attempt to clone an expert), 이러한 지식을 컴퓨터에 축적(Acquisition)하고, 통계를 요구하는 사용자의 의도를 잘 나타내어질 수 있도록, 처리방법을 표현(Guide representation)해 줌으로서, 요구자가 의도했던 결과가 표출(Manipulation)되도록 하는 지식공학의 분야이다.

(3) 소프트웨어 공학적 영역

통계처리 지식을 컴퓨터에 부여하는 작업으로서, 통계처리 지식의 표현과 정보처리 이용의 문제들까지 동시에 해결할 수 있도록 하는 프로그래밍 하려는 방법론적인 규칙을 다루는 것이다. 즉 통계처리과정에서 통계지식 표현형식과 정보처리방법 및 이용에 대한 이해를 간편하고 쉽게 하기 위해서 소프트웨어 공학적 연구에 앞서서 당연히 검토되어야 하는 기본적인 과제이지만, 이러한 문제를 처리성이 높게 인공지능화 기능을 활용하여 능동적으로 스스로 처리할 수 있게 프로시듀어팩키지 프로그램(Procedure Package Program)의 작성과 새로운 통계처리 언어의 구현을 의미한다.

일반적으로 언어는 문자, 음성음운, 구문, 의미, 이용체제로 구성되는데, 그와 같이 소프트웨어공학에서도 각 구성요소에 대한 규범이론의 확립과 함께 언어현상을 설명하는데 복수의 문법이론이 수월하게 표현되고, 실현될 수 있는 컴퓨터의 범용성을 이용하여 쉽게 다룰 수 있으므로, 지식공학적인 분야에서 다듬어진 이같은 문제들을 통계처리 실행과정에서 사용자들이 이해의 어려움이나, 활용상의 복잡함 등을 쉽게 해결해줄 수 있게하는 분야이다.

(4) 시스템 공학적 영역

시스템 공학적이라는 것은 학문적인 관점에서 서술되어있는 통계처리를 시스템처리과정 모형의 체계와 실제로 시스템개발 작업을 공학적으로 해결하는 것이다. 즉, 이론과 통계처리 구조의 실재적 심리성(Psychological reality)-인간공학적으로 부담을 느끼는-과 지식공학적인 측면에서 통계처리의 이론적 내용을 파악 이해하고, 처리과정의 체계가 실제적으로 운용되는 시스템개발 작업을 수행하고 인간과 기계계의 관계를 공학적으로 해결하는 것이다.

따라서, 이러한 통계처리를 시스템론적 요소(Component)로 볼때에 하드웨어(Hardware), 소프트웨어(Software), 어플리케이션웨어(Application ware)로 대별 하지만, 이 요소에 의한 시스템의 구성(Organization)은 전형적인 분할제어관리(Devide and Conquer)를 갖는 형태로 단계를 갖추어서, 방법론적 이론(Methodeloge), 제어관리 모델(Model), 알고리즘(Algorithm), 구현화작업(Implementation)으로 발전시켜 통계정보를 요구하는 사용자에게 처리과정의 기능이 쉽고, 편리하며, 효과적으로 요구자의 의도를 표현할 수 있도록 표현범위(Representation domain)를 경제성, 완전성, 일반성을 갖추어 사용자로 하여금 부담을 느낄 수 없도록 자연스럽게 운영관리되며, 만족시키게 하는 인간공학이 포함되는 분야이다.

4. 통계전문가시스템의 방법론

통계전문가시스템은 통계학 지식과 통계처리 지식이 컴퓨터시스템에서 실질적으로 정보처리되는 과정에서 일어나는 상태와의 공학적인 상호관계를 어떤 수단을 써서 모형화 할 것인가가 쟁점이 된다. 즉, 전문가 시스템에서도 초기 인공지능화에 앞서 지식정보처리를 수식화의 알고리즘을 통하여 프로그램 개발에 의존하였던 것처럼, 전문가시스템에서 처리하는 통계처리도 응용문제를 풀기 위해서 수학적 이론을 이용하여, 문제의 해를 구하고 관계를 해석하는 처리과정을 프로그래밍으로 조작(Operation)되게 하여야 한다.

그러므로, 통계문제를 공학적인 방법으로 만족하게 처리하려는 문제점은, 통계처리하고자 하는 정보의 표현 형식과 처리된 그 결과와 해석을 컴퓨터시스템에 의하여, 조작되게 작업하는 규칙이 필요하다는 것이다. 즉, 컴퓨터시스템 매체를 통하여 정보의 입출력이 이루어져야 하며, 이를 통한 의사 전달이 이루어지는 형식표현 요구와 규정이 따라야 하기 때문에, 이와같은 형식화는 양적인 관계를 기술하는 수식이 아니라 술어 논리식(Predicate Logic)같은 기호적인 시스템으로 표현되어야 한다.

그리고, 이러한 방법론적인 접근 방법은 목적 범위(Task Domain)에서 처리와 변화가 이

루어지는 경우와, 여러가지 다양한 표현(Various representation) 범위에서 처리와 변화가 이루어지는 경우로 구분시키면 된다.

그러나, 이러한 방법론적인 접근이 시도될 때는 반드시 정보관리시스템에서 보여주는 단순성(Less complex), 용이성(Easier to follow), 능동성(More properly the Multiple Representation) 및 확장성(Possible extensions)이 검토되어야 한다.

따라서, 통계처리를 사용자와 시스템구성요소간의 연계되는 관계로 볼 때, 통계처리의 지식표현(Knowledge representation)은 하드웨어, 소프트웨어, 어플리케이션웨어가 <그림 1>에서와 같이 통계처리의 공학적 지식표현은 형식화(Formalization), 추론기능은 처리와 표현(Processing and representation)과 해석(Interpreter), 사용자 요구에 대한 입출력 기능으로 구성시킬 수 있다.

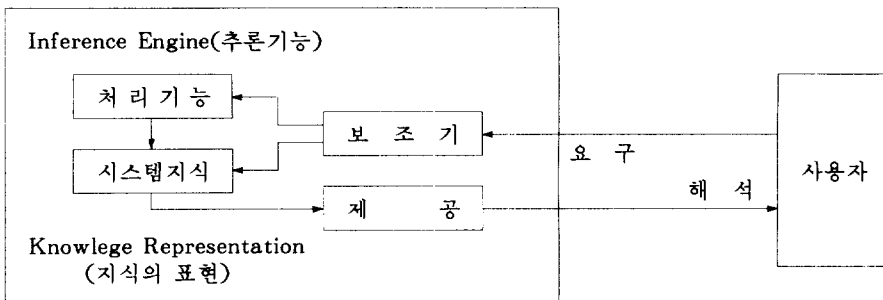


그림 1 통계전문가시스템의 지식표현 구성

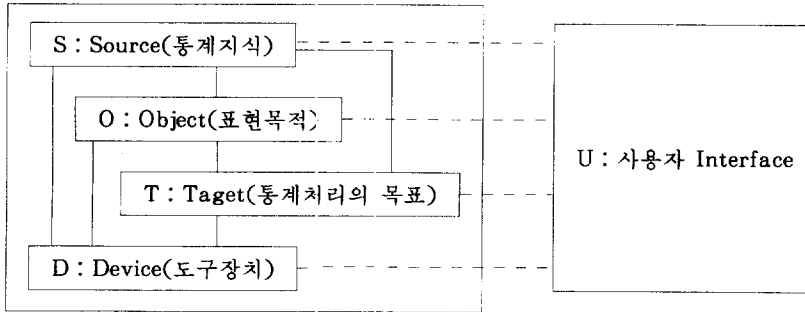
그리고, 형식화(Formalization) 단계에서 통계처리 구조를 이루는 요소를 명시적으로 정보화 함으로써, 일정한 형식을 갖는 의미적인 기호 표현과 그 표현에 대한 해석을 전제로하여 조작·규칙(원리)으로 정형화 되어야 한다.

특히 이 단계에서 사용자의 요구를 최초로 접하게 되므로 시스템도움자(user's apprentice)가 정형화된 조작(Operation)에 따라 변환(Conversion)이나 정화(Adapter)가 충분히 이루어질 수 있도록 다양한 형식이 필요하다.

처리와 표현과 해석의 단계에서 형식화된 기호표현을 자료기반(Data base)과 지식기반(Knowledge base)에 두고 실제로 추론기능(Inference Engine)으로 운영되어서 해답을 얻는 것으로, 통계지식은 확률론적처리 방법에 따르는 처리기능과 논리구조적 시스템 지식이 상호 보완검색(Retrieval)되어 다양한 형태의 해석 모형으로 설정된 자료기반에서 지식기반에 따른 추론기능을 발휘하게 한다.

이러한 과정에서 사용자가 요구한 문제가 형식화단계에서 처리된후, 요구한 사용자가 이를 인식할 수 있도록 입출력장치에 표현되어 제공받게 되는것은 시스템 이용에 대한 효율성으로 평가받는 단계가 되겠지만, 이 단계에서는 무엇보다 기계계와 인간과의 의사전달(Interface)의 도구(Tool)화가 개입되어야 하므로, 물리적인 시스템운영조작(Operation)을 행하므로써 이루어지는 부분에 관한 공학적 문제로 미루어야 한다.

이같이 형식화, 처리와 표현과 해석, 입출력 과정은 통계의 공학적 처리의 이론정립과 불가분의 관계가 있으나, Rich(1988)의 프로그램 사용자들을 위한 학습구조개념에 적용한다면 실제적인 통계의 공학적 과정은 통계지식(Source), 형식화된 처리와 표현목적(Object), 이를 근본으로 한 통계처리의 목표(Target), 또한 이들로부터 얻어진 정보와 자료를 해석하는 장치도구(Device)가 <그림 2>와 같이 사용자에게 대해서 유기적인 구조를 가져야 된다.



* 주. _____ 내적인 공학관계
 사용자의 요구

<그림 2> 통계전문가시스템의 처리구조

이러한 유기적인 처리구조는 사용자 요구에 응해서 즉각적인 표현으로 실행 되어지는 시스템으로 형성되어야 하므로, 이 사용자 요구가 포괄할 범위(Task domain)는 과제의 편집(Text editing), 그래픽 표현(Graphic specification), 고정처리(Post processing), 화상(Image), 화일(Filing), 능동적인 지식의 습득(Dynamic reading)이 서로 연계되어 구성해야 할 것이다.

따라서, 사용자 요구를 Ω 라 할 때 이 Ω 를 요구자에 대한 정의로 구분한다면, 첫째 표현 상태가 재수정(Editing)이나 복구가 자유스럽게 할 수 있도록 항상 유지보수(Maintenance) 되는 현상(Π), 둘째 사용자가 규정에 의한 방법으로 상호관계를 정의하는 것(Θ), 세째 첫번째의 현상(Π)에서 얻어진 결과를 표현(Hard copy)하도록 변환시켜 주는 것(Γ), 네째로 표현현상(Π)에 대하여 사용자의 대화구정(Θ)이 사용자에게 대하여 관련지워줄 수 있는 융통성(Versa)(Δ)으로 구분되며, 이들은 각각 Ω 에 대하여

$$\Omega = [\Pi, \Theta, \Gamma, \Delta]$$

로 서술된다.

5. 통계전문가시스템의 공학적인 과제

전문가통계처리를 지식처리시스템과의 공학적인 접근방법으로 수행하고자할 때의 기본적인

인 처리 단계는 <그림 3>과 같이 신호처리되는 자료의 처리단계와, 자료단계에서 얻어진 신호를 기호화 처리과정으로서 정보처리 단계와, 이를 다시 지식 기반을 두고 처리되는 지식 처리 단계로 구분할 수 있다. 이와 같은 단계적 처리 과정은 기계적인 방법으로 추출한 데이터를 점차적으로 고수준의 추상화 과정을 위한 형식 조작이 되는데, 이때 각 단계에서의 규칙성의 발견과 시스템의 규정화 문제가 주된 과제로 제기 된다.

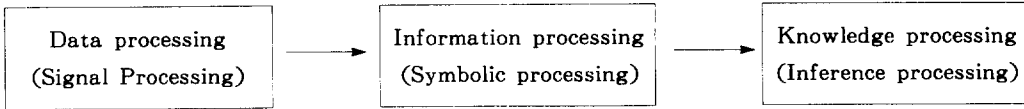


그림 3 통계전문가시스템의 공학적 처리단계

따라서, 지금까지 언급한 통계전문가시스템의 공학적과제를 요약해서 나열한다면 다음과 같다.

- 통계의 표기 시스템 규정에서
 - 전문용어, Thesaurus
 - 해석의 언어적 추론법
 - 지식에 의한 해석법
 - 입출력 기호화와 표기 표현법
- 통계 Text처리기술 개선 및 개발로서
 - 논리, 수리적 알고리즘 개선
 - 통계 Text 압축, 검색
 - 통계 Text editor 개발
- 인간과 컴퓨터와의 대화시스템으로
 - CAI(Computer aided instruction)
 - Software
 - RISC(Reduce instruction system component)
 - 통계지능 Software
- 컴퓨터 시스템 개발로
 - I/O device 시스템
 - 자동 자료 수집 · 가공 시스템
 - 실시간 검색 시스템
 - 전문가 통계 D.B 시스템

6. 통계전문가시스템에 대한 제언

통계란 그 어휘 자체가 계산해서 다루어야 한다는 말이므로 사람들이 개입한다는 뜻이며,

인공적이란 말로 표현할 수 있다. 즉, 통계란 말은 오랜 역사속에서 사용되어온 소프트웨어를 뜻하며, 그것을 이용 했을 때 편리했던 도구가 발달되어있지 못했던 것이라 말할 수 있다.

그러나, 현재는 컴퓨터시스템으로 이를 대체할 수 있으므로 당연한 공학적 발전을 정립하지 않을 수 없다고 본다. 본 논문이 제언하는 결론적 결과는 통계학이 당면한 한 과제가 있다면, 공학적 영역으로의 확대 발전하기 위해서 전문가통계처리의 공학적 과제의 연구가 반드시 검토되어 추진되므로써, 전문적인 학문의 한 분야로서 더욱더 발전될 수 있도록 공학적 접근모형을 제언한다.

참 고 문 헌

- (1) Buchanan, B.G. and E.H. Shortliffe(1985), "Rule-Based Expert Systems", Addison-Wesley Publishing Co.
- (2) Chen, P. and M.A. Harrison(1988), "Multiple representation document development", IEEE Computer, vol.21, no.1, 15-31
- (3) Gale, W.A.(1986), Artificial Intelligence and Statistics, Addison-Wesley Publishing Co.
- (4) Ghosh, S.P.(1987), "Data Enginerring for Intelligent Inference ; Statistical Data", Proceedings of the Third International Conference on Data Engineering, Los Angeles, U.S.A., 281
- (5) Hahn G.J.(1985), "The Intelligent Statistical Expert System : Future Directions." The American Statistician, vol.39, 12-14
- (6) Hand, D.J.(1984), "Statistical Expert System : Design", The American Statistician, vol.38, 351-369
- (7) Harkong, L. and F.R. Hickma(1985), "Expert System Techniques : An Application in Statistic", Cambridge University Press
- (8) Jida, J. and J.Lemaire(1986), "Expert System and Data Analysis Package Management", COMPSTAT. Proceedings in Computational Statistics. 7th Symposium, Rome, Italy, 251-258
- (9) Klahr, P. and Donald A. Waterman(1986), "Expert Systems", Addison-Wesley Publishing Co.
- (10) Portier, K.M. and P.Y. Lai(1983), "A Statistical Expert System for Analysis Determination", Proceedings of the ASA Statistical Computing Section, 309-311
- (11) Rowley, D.(1986), "Statistical Approches to the Generation of Rules for Expert System", Second International Expert Systems Conference, London, England, 369-375
- (12) Rich, C. and Richard C. Waters(1988), "The Programmer's Apprentice : a Research Overview", IEEE Computer, vol.21, no.11, 11-25
- (13) Schoen, E. and G. Smithm, G. Buchanan(1988), "Desing of Knowledge-based System with a Knowledge-based Assistant", IEE Trans. Software Engineering, 1771-1791

Engineering approach of Statistics Processing for the Statistical Expert System

HONG JUN TCHA*

Abstract

Engineering approach of statistics processing is defined for the statistical expert system. First, the engineering approach requirement are conceptualized by using an artificial intelligence in statistics, with the extensions being additional statistical knowledge engineering such as software engineering, optional relationships, and the generalization abstraction. The methodology produces statistical expert system designs that are not only accurate representations of reality but also enough to accommodate future processing requirements. It also representations of knowledge that must be constructed, using the extended engineering processing model conceptualization and proposed engineering approach of the problem.

*Dept. of Computer Science, Kang-Weon National University, Kangweon-do, Chuncheon-si, Hyoja-dong.