

AS/TS : Insight 2를 利用한 統計技法選定諮問시스템 (Artificial Statistician/Tool Selection)

全龍鎭*

要約

다양한 통계기법을 이용하는 연구자들은 不適合한 통계기법을 선정하여 수집된 자료를 분석함에 따라 많은 誤謬를 범하고 있는 실정이며, 자신의 연구목적에 부합되는 통계기법을 선정함에 있어 어느 기법을 선택할 것인가 하는 문제에 직면하게 된다. 본 연구에서 구축한 통계기법 선정을 위한 專門家 시스템(AS/TS)은 수집된 자료의 특성과 연구목적에 적합한 기법을 연구자에게 선정해주는 전문가 시스템이다. 현존하는 대부분의 통계 패키지들은 이용자가 자신의 연구목적에 가장 적합한 통계기법을 선택할 수 있는 지식을 지니고 있는 것으로 전제하고 있기 때문에, AS/TS는 이들 통계 패키지를 이용하기 전에 통계기법만 선택해 주도록 설계되었다. AS/TS는 MS-DOS를 搭載한 16bit 퍼스널 컴퓨터로 INSIGHT 2(shell)를 이용하여 構築하였다.

1. 緒論

지금까지 통계학에 人工知能(artificial intelligence : AI)을 적용한 논문은 다수 있으나, 그 대표적인 것으로 국내에서는 염 봉진교수[1]의 RMEX로, 최적 회귀방정식을 설정하기 위한 전략을 실험자에게 제시해주는 통계적 專門家 시스템(expert system : ES)이 있다. RMEX는 線型 回歸方程式을 세우는데 필요한 통계지식이나 단계들을 적절한 소개하고 설명하도록 설계되었으며, 사전분석으로 다양한 형태의 다변량기법과 기준들을 분류하고 그들의 일반적인 사용상의 지침을 제시하였으며, 회귀모형설정을 위한 독립변수 선택과정과 선택된 결과에 대한 설명기능을 포함하고 있다.

국외에서는 이용자에게 통계분석에 대한 諮問을

제공하기 위해 통계자료분석에 전문가 시스템을 적용하려는 시도가 1977년 Nedler에 의해 처음 제안되었고, 그 후로 Jones(1980)는 컴퓨터를 통계 컨설턴트로 이용하기 위한 방안을 제시하였다. 또한 Chamber (1981, 1982)는 통계전문가 소프트웨어를 개발하기 위한 개념과 잠재성에 대하여 발표하였다. 1983년에는 통계학에서의 전문가 시스템이라는 주제로 런던에서 로얄 통계학회 학술 발표회가 개최되어 다수의 논문이 발표되었으며, 가장 최근에 AT & T 벨 연구소에서는 통계전문가가 아니더라도 회귀분석기법을 이용할 수 있도록 사용방법을 제공하는 회귀전문가(regression expert) 시스템을 구축하였다[10, 11]. 또한 Remus와 Kottmann[1986]의 AIS(artificially intelligent statistician)로, 의사결정자에게 여러가지 상황이 주어질 때 자료수집, 적절한 통계기법 선정,

* 全州又石大學經營學科!

결과해석 등에 인공지능을 적용하는 방안에 대하여 개념적으로 제시하고 있다[19].

본 연구는 각종 연구의 資料處理에 가장 적합한 통계기법을 選定하기 위해 ES 개념을 적용하고자 하는데 초점을 맞추고 있다(artificial statistician / tool selection : AS/TS). 통계기법은 상당히 다양하므로 연구자가 모든 통계기법에 관해 정확히 이해한다는 것은 거의 불가능하기 때문에 연구목적과 자료의 특성에 적합한 통계기법을 選定하기란 대단히 어렵다. 따라서 이러한 문제를 統計專門家(statistics expert : SE)와 같은 수준으로 해결할 수 있는 AS/TS를 구축하여, 不適合한 통계기법선정으로 인한 자료처리 및 해석상의 誤謬를 방지하자는 것이 본 연구의 목적이다.

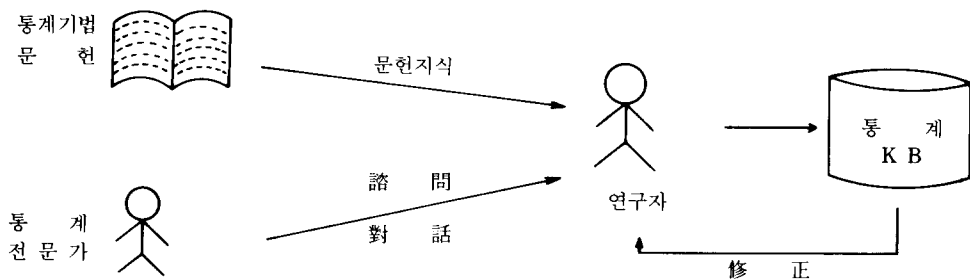
일반적으로 통계기법 선정 절차는 크게 두가지 기준에 의해 결정되는데 1) 수집된 자료의 특성 즉, 名目尺度(nominal scale), 順位척도(ordinal scale), 區間척도(interval scale), 比率척도(ratio scale)와 2) 연구자가 통계분석을 통하여 얻고자 하는 연구목적에 의하여 결정된다. 이와같은 1 단계 통계기법 선정절차가 종료되면 세부적인 전제 조건의 만족 여하에 따라 최종 통계기법이 선정되는 2) 단계 선정절차를 따르게 된다. 2 단계 선정절차는 母集團의 分布 및 獨立性, 標本の 數, 표본의 크기, 구체적인 연구의 목적, 표본자료의 특성(分散度, 母數, 非母數 등) 등에 의해 수행된다. 본 연구에서 통계기법 선정절차를 2 단계로 분리하였는데 그 이유는 深索節次를 단축시킴

으로써 빠른 시간에 推論(inference)하여 결론 導出에 소요되는 시간을 절약하고자 하는 의도가 있기 때문이다.

연구방법은 크게 문헌연구와 AS/TS구축으로 나누어 접근한다. 문헌연구에서는 우선 시스템 구축에 필요한 언어선택 기준, 深索전략 및 ES구축 타당성, 적합성 등에 대하여 고찰한 후 이를 토대로 AS/TS구축 적합성을 검토하고, 이어서 AS/TS구축에 알맞는 表現言語 및 탐색전략 등을 결정한다. 또한 AS/TS구축에 필요한 資源, 구축절차, 각ES기법을 확인 및 결정한 후 이용자가 주위에서 쉽게 접할 수 있는 퍼스널 컴퓨터를 이용하여 AS/TS를 구축하게 된다.

AS/TS는 디스크드라이브 2 대, 主記憶容量이 640 KB로 MS-DOS를 搭載한 16bit 퍼스널 컴퓨터로 生産規則 言語(INSIGHT 2)를 이용하여 구축되었으며, 지식베이스는 에디터(PC-Write)를 이용하여 작성하였다.

통계기법선정 지식은 각 연구자에게 어떤 통계기법을 이용할 것인가에 관한 諮問역할을 하게 되므로 AS/TS는 일종의 자문 시스템이라 할 수 있다. 통계기법선정은 <그림 1>과 같이 1) SE의 전문지식, 2) 통계기법에 관한 전문서적 및 관련 논문 등을 이용함으로써 지식 베이스(knowledge base : KB)를 구축할 수 있다. 통계문헌으로 부터는 기본지식을 얻고 SE와 對話 및 諮問을 통하여 KB를 수정해 간다.



<그림 1> 지식취득 경로

2. 言語와 知識表現

통계기법선택(statistical tool selection : STS)을 위한 지식의 형태를 파악하여 표현에 가장 적합한 기법을 搭載한 ES 구축 언어를 선택해야 한다. 적합한 ES구축 언어의 선택은 ES성공 가능성에 중요한 역할을 하므로 KE는 이용 가능한 언어 중 지식을 가장 잘 표현할 수 있는 언어를 선택해야 한다.

3.1. 知識構造 特性

통계지식에서 기술한 바와 같이 통계기법을 선정하기 위해서는 기법을 사용하는데 따른 전제조건이 요구된다. 모수통계기법을 이용하기 위해서는 모집단의 分布에 관한 가정 등이 요구되며, 모집단의 수에 따라 이용 가능한 통계기법이 결정된다. 자료의 특성에 의해 모수통계기법 또는 비모수 통계 기법을 이용해야 하는가를 결정하게 된다. 이와 같이 STS지식은 통계기법 사용을 위한 조건과 이러한 조건이 만족될 때 이용가능한 통계기법 결정으로 구성된다.

STS지식은 조건(IF, AND, OR)과 결론(THEN, ELSE)으로 구성된 대표적인 형태를 이루고 있다. 條件節에는 통계기법 이용에 필요한 전제조건이 나열되며, 結論部에는 조건절을 만족시키는 구체적인 목표(통계기법)가 위치한다. 이와 같은 지식구조는 파스칼 언어를 이용하여 개발된 生産規則言語 (production rule language : PRL)인 Insight 2에 의해 가장 잘 표현될 수 있다[3, 17].

3.2. 言語選擇

언어선택은 지식표현력과 구축효율성에 대단히 중요하므로 KE는 기존 ES구축 언어를 비교하여 적용영역의 지식에 가장 적합한 언어를 선택해야 한다. 본 절에서는 STS지식을 표현하게 될 PRL의 선택근거와 PRL에 대하여 살펴본다.

3.2.1. 選擇根據

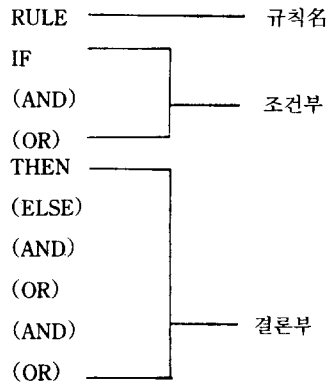
AS/TS의 KB는 STS지식을 표현하기 위한 기본 구조로서 생산 규칙을 이용한다. 생산규칙은 전문가(domain expert : DE)가 어떤 假說을 증명하거나, 주어진 상황으로 부터 결론에 도달하는 과정으로 생각할 수 있다. 이는 인간의 思考過程(thought process)과 매우 유사하게 KB를 구성할 수 있다는 장점이 있다. PRL을 선택한 근거는 다음과 같다.

1. 조건과 결론으로 구성되는 STS를 위한 지식표현에 매우 적합한 언어이다.
2. 컴퓨터 프로그래밍에 관해 충분한 지식이 없더라도 누구나 용이하게 이해할 수 있으므로 전문적인 KE가 아니라도 이용자 스스로 ES를 쉽게 구축할 수 있는 언어이다.
3. 주변에서 쉽게 접할 수 있는 퍼스널 컴퓨터에 적합한 언어이므로 퍼스널 컴퓨터用(PC based) ES를 직접 구축할 수 있다.
4. 대형컴퓨터용 언어를 이용하는 것 보다 비용이 적게들뿐만 아니라 KB 修正과 컴파일을 즉시할 수 있으므로 ES구축 시간이 절약된다.
5. 규칙의 순서와 결론도출은 無關하기 때문에 언제든지 규칙을 쉽게 첨가 시킬 수 있다.

3.2.2. 生産規則 言語

지식을 구성하는 규칙이 因果關係로 표현되면 이는 "IF-THEN"형식으로 模型化될 수 있다. A, B, C로 인해 D가 발생하고 그렇지 않으면 E가 발생한다면, A, B, C, D, E의 관계는 "if A and B and C then D else E"와 같이 생산규칙으로 표현할 수 있다. PRL은 KB를 구성하는 규칙을 표현하기 위해 이와 같이 간단하고 자연스러운 문법을 이용한 언어이다.

PRL을 이용한 규칙은 다양한 형태를 취할 수 있으나 일반적인 형태는 <그림 2>와 같이 규칙명(rule name), 조건부(supporting condition), 결론부(conclusion)로 구분된다.



〈그림 2〉 PRL을 이용한 규칙형태

AS/TS KB의 규칙을 예시하면 다음과 같다.

```

RULE Unqualified data for parameter
IF   Group IS Has not approximately normal
      distribution
AND  Sample size <30
AND  Data IS Interval scaled data
OR   Data IS Ratio scaled data
THEN Data unqualifies for parameter
  
```

위와 같이 모든 규칙은 규칙명, 조건부, 결론부로 구성된다. 각 규칙은 반드시 규칙명이 존재해야 한다.

위 규칙의 규칙명은 :

“Unqualified data for parameter”

조건부는 :

“Group IS Has not approximately normal dist.”

“Sample size <30”

“Data IS Interval scaled data”

“Data IS Ratio scaled data”

결론부는 :

“Data unqualifies for parameter”

이 된다.

위 규칙은 모수통계에 부적합한 자료인가를 판정하기 위한 것으로, 조건부에서 만일(IF), 모집

단이 正規分布를 따르지 않고(AND), 표본크기가 “30” 미만이고(AND), 자료가 구간척도 이거나(OR), 비율척도이면 결론부에서는 조건부가 충족되면 (THEN), 이자료는 모수통계분석에 적합치 않다고 결론을 내리게 된다.

이와 같이 생산규칙언어는 STS 근거를 探索하여 결론인 통계기법에 도달하는 과정을 가장 적절히 표현할 수 있는 언어이다.

3.3. 知識表現

表現方式(paradigm)이 다양하고, 선택기준이 명확치 않기 때문에 ES에서 지식표현(knowledge representation : KR)기법의 선택은 특히 중요하다. 부적합한 KR 기법을 선택에 따라 중요한 지식이 선택한 표현언어로 코드化될 수 없다면 ES 구축은 실패하게 된다.

본 연구의 STS지식은 IF-THEN으로 구성되어 있으므로 PRL 을 이용하여 표현한다.

PRL을 이용해 작성된KB는 大別하여, 목표(goal), 목표를 만족시키기 위해 필요한 규칙(rule), 이용자가 사용하기 용이하도록 정보입력 및 출력을 위한 추가정보 등으로 구성된다. 본 절에서는 목표, 규칙, 추가정보 표현 등에 관하여 기술한다.

3.3.1. 目標表現

추론엔진이 KB의 규칙을 탐색하여 도달하게 되는 목표를 명시할 필요가 있다. 일련의 추론과정들을 통해서 도달 가능하거나 증명할 수 있는 중간 및 최종결론인 목표를 설정한다. PRL은 하나 이상의 목표를 가져야 하는데 이 목표는 KB가 추론가능한 결론이 되거나 KB가 증명할 수 있는 가설이 된다.

각 목표는 목표를 구분하는 일련의 번호가 부여지며, 一次목표(primary goal)는 副목표(subgoal)를 가지며, 부목표 또한 副-副목표(sub-subgoal)로 구성된다. AS/TS의 경우 4 개의 일차 목표를 가지며, 각 목표는 副목표를 갖는다. 〈표 1〉은 AS/TS KB에서 목표부분을 나타내고 있다.

이용자가 연구분석 목표로 일차목표인 “Testing”을 만족시키는 요인을 선택한 후, 부목표에 도달하게 된다. 부목표에는 단일 모수, 單一非母數, 複數모수, 복수非모수, 多數모수, 다수非모수 등이 있다. 이용자로 부터 자료특성, 표본의 수, 모집단의 특성 등에 관한 정보를 입력받아 부목표 중 하나에 도달하게 된다. 부목표에 도달한 후 구체적인 연구목적, 필요한 가정 등을 탐색한 후 최종목표인 통계기법을 선택하게 된다. 이와 같이 목표구조는 최종목표에 도달하기까지의 일련의 순서를 제시한다.

2.	Testing
2.1	Can use one para. test
2.1.1	Select one para. test
2.2	Can use one nonp. test
2.2.1	Select one nonp. test
2.3	Can use two para. test
2.3.1	Select two para.test
2.4	Can use two nonp.test
2.4.1	Select two nonp. test
2.5	Can use multip. test
2.5.1	Select ANOVA
2.6	Can use multinonp. test
2.6.1	select multinonp. test

para : parameter, multip. : multi-parameter
 nonp. : nonparameter, multinonp. : multi-nonparameter
 <표 1>AS/TS KB의 목표(檢定의 경우)

3.2.1. 目標表現

각 규칙을 구성하는 지식이 표현되는 형태에 따라 事實(factual) 정보표현, 數值 (numerical) 정보 표현, 對象-屬性 (objective-attribute-value : OAV) 정보 표현으로 구분될 수 있다. 사실정보 표현은 진실 또는 거짓(true or false)으로 나타낼 수 있는 자료형태로서, 이용자는 “True”와 “False” 중 하나를 택함으로써 추론엔진에 정보를 제공한다.

수치정보 표현은 이용자가 수치로 응답하면 추론엔진은 그 값을 이용하여 KB내의 규칙과 비교

하게 된다. 수치정보를 표현하기 위하여는 대상, 관계 연산자, 수치 값 등이 필요하다. AS/TS KB의 규칙을 예시하면 다음과 같다.

```

RULE Difference between means using Z test
IF Testp2 IS Difference between two group
means
AND Sample size >=30
OR Sum of both sample size >=58
AND Sample IS Independent or no relation
THEN Test tool IS diff. between means using Z
  
```

조건부의 “Sample size >=30”에서 “>=”는 관계연산자로서 컴파일러는 “>=”이 포함된 조건절을 수치정보로 해석하며, 이용자가 수치를 입력하도록 요구한다. 추론엔진은 이용자가 입력한 수치와 “>=” 우측의 수치 “30” 과 비교한다. 입력된 값이 “30” 미만이면 이 조건절은 “거짓”이 되며, “30” 이상이면 “진실”로 처리한다.

사실과 수치표현 외에 조건부를 나타내는 OAV는 대상 (object)과 속성(attribute)을 표현하는 한 쌍으로 구성된 자료 형태이다. AS/TS KB의 규칙을 예시하면 다음과 같다.

```

RULE Difference between means using t test
IF Testp2 IS Difference between two group
means
AND Sample size <30
OR Sum of both sample size <58
AND Grops IS Has approximately normal distribution
AND Sd IS Unknown
AND Sample IS Independent or no relation
AND Variance of group IS Equal
THEN Test tool IS diff. between means using t
  
```

조건절의 “Variance of group IS equal”에서 “IS” 좌측이 대상이 되며, 우측이 특성이 되어

대상-속성 한쌍으로 규칙의 조건부를 표현할 수 있다. 컴파일러는 豫約語(reserved word) "IS"를 이용한 자료형태를 OAV로 인식하게 된다.

사실과 수치표현은 단일 규칙내의 조건부가 진실인가 거짓인가를 판단하기 위한 것이다. 따라서 이용자가 입력한 정보는 단일규칙에만 영향을 미치게 된다. 반면 OAV로 입력된 정보는 多數의 규칙內 결론에 도달하는데 영향을 미친다. AS/TS KB의 규칙을 예시하면 다음과 같다.

RULE Data

IF Data IS Interval scaled data

OR Data IS Ratio scaled data

THEN Data qualifies for nonparameter

!

RULE Nonparameter data

IF Data IS Nominal scaled data

OR Data IS Ordinal scaled data

THEN Data qualifies for nonparameter

!

RULE Unqualified data for parameter

IF Group IS Has not approximately normal distribution

AND Sample size <30

AND Data IS Interval scaled data

OR Data IS Ratio scaled data

THEN Data unqualifies for parameter

위 세 규칙에는 모두 "DATA"라는 대상이 있고 그 특성으로 名目尺度(nominal), 順位척도(ordinal), 區間척도(interval), 比率척도(ratio)가 있다. 4 가지 척도 중 한 특성이 선택되면 모수 통계분석인가, 또는 비모수 통계분석인가를 결정하게 된다. "ordinal"과 "nominal"의 경우는 비모수 통계 분석에 적합한 자료이므로 모수통계기법이 선택된다. 그러나 "Ratio" 또는 "Interval"의 경우라도, 표본크기가 적거나, 정규분포를 따르지

않을 때는 모수통계를 이용할 수 없다. 이와 같이 OAV표현에 의해 입력된 정보로 多數의 규칙이 결정된다.

3.3.3. 追加情報 表現

KE가 이용자에게 추가로 정보를 제공하거나 세부적인 설명을 제시할 때 "EXPAND"文, "DISPLAY"문, "TEST"문 등의 명령어를 사용할 수 있다.

"EXPAND"문은 이용자가 KB내 항목에 대하여 세부적인 설명을 요구할 때 그 항목(목적, 조건절, 결론 등)에 대한 설명을 제시하는데 이용된다. AS/TS KB의 규칙과 "EXPAND"문을 예시하면 다음과 같다.

RULE Reqr.

IF Data qualifies for parameter

AND No. of group >=2

AND Dependent variable is a random variable

THEN Can use regr.

EXPAND Dependent variable is a random variable. The variable being predicted is referred to as the dependent variable. The variable being used to make the prediction is called the independent variable. Independent variables are the causes and dependent variables the effects. A random variable is one whose value is determined by chance processes that are not under the control of the observer.

"EXPAND"문의 경우 이용자는 세부적인 설명을 참조하지 않을 수 있으나 "DISPLAY"문은 이용자의 요구와는 무관하게 KE가 어떤 항목에 대하여 설명하고자 할 때 사용하므로 설명이 자동적으로 나타난다. AS/TS KB의 규칙과 "DISPLAY"文을 예시하면 다음과 같다.

RULE For selecting regression analysis tool

IF Regr. tool IS simple
OR Regr. tool IS multiple
OR Regr. tool IS polynomial
THEN Select regr.
AND DISPLAY regr.
ELSE DISPLAY can't use regr.

DISPAY regr.

You can use REGRESSION ANALYSIS TOOLS Specific regression tool as follow :

You can arrive at the regression equation in order to predict the value of one variables based on another variable using simple regression

BAR Regr. tool IS simple

You can arrive at a multiple regression equation whereby two or more independent variables can be used to predict the value of dependent variable using multiple regression

BAR Regr. tool IS multiple

You can arrive at a polynomial regression equation whereby one or more independent variable using polynomial regression

BAR Regr. tool IS polynomial

위 규칙에서 회귀분석 기법선택 결론에 도달하게 되면 "AND DISPLAY regr."에 의해 KB규칙 뒤에 위치하는 "DISPLAY regr."의 양식으로 결론(통계기법 선택)이 표현된다. 본 연구의 AS/TS의 모든 최종 결론은 이와 같은 형태로 이용자에게 제시된다.

"TEST"문은 KE가 이용자에게 質疑(query)를

하면서 정보를 얻고자 할 때 사용한다. 즉 이용자가 질문의 요지를 쉽게 파악하여 이에 적절한 정보를 제공할 수 있도록 하기 위해 "TEST"문을 사용한다. 이는 추론엔진과 이용자간의 의사소통을 향상시키는 역할을 하게 된다. AS/TS KB의 규칙과 "TEXT"문을 예시 하면 다음과 같다.

RULE Data

IF Data IS Interval scaled data

OR Data IS Ratio scaled data

THEN Data qualifies for parameter

TEXT Data

Select all the data scales which you have obtained or collected.

위 경우 이용자에게 "자료는 ?"과 같이 제시하는 경우와 "수집된 자료의 형태를 선택하십시오"라고 제시할 때 이용자는 후자의 경우 질문의 意圖를 쉽게 파악할 수 있다.

4. 推論戰略 및 實行

일반적으로 탐색전략은 逆方向(backward chaining)추론과 巡방향(forward chaining) 추론으로 구분한다. 역방향추론의 경우 추론엔진은 우선 목표(goal)를 선택한 후, 그 목표를 정당화시키는 조건이나 사실들을 KB에서 찾아 일치되는가를 확인해 나간다. 이 방법은 목표의 수가 적고, 목표를 정당화시키는 조건이나 사실들이 많을 때는 매우 효과적이다. 이 경우 순방향추론을 한다면 모든 조건과 사실들을 탐색한 후에 목표에 도달하므로 추론시간이 길어지고, 탐색결과를 一時 저장하는데 필요한 기억공간도 증가하게 된다.

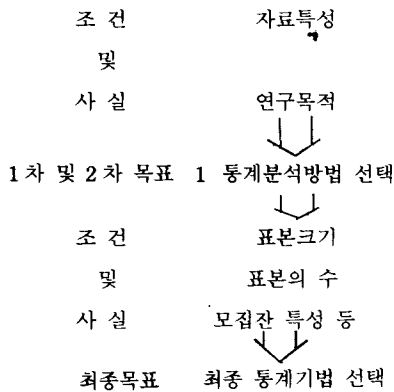
그러나 목표의 수가 많고 목표를 정당화시키는 조건 및 사실이 적은 경우 순방향 추론이 더욱 효과적이다. 순방향의 경우 추론엔진은 이용자로 부터 입력된 정보를 KB의 규칙 및 사실과 일치하는가 탐색하면서 최종목표에 도달하게 된다.

4.1. AS/TS 推論戰略

본 연구의 목표가 되는 통계기법의 수가 많으므로 순방향 추론 전략을 채택하여 최종목표(이

용자에게 가장 적합한 통계기법)에 도달하게 된다. <그림 3>은 AS/TS 추론전략을 제시하고 있다.

AS/TS 추론에서 1차 목표에 도달하기 위해 추론엔진은 KB에서 이 목표를 정당화하는 자료특성 및 연구목적이라는 조건이나 사실을 1차적으로 탐색하게 된다. 1차 목표인 1차 통계기법이 선택된 후 구체적 통계기법인 최종목표에 도달하기 위해 필요한 조건 및 사실(표본 수, 표본크기, 모집단 특성 등)을 탐색한 후 이용자에게 가장 적합한 통계기법을 선택하게 된다



<그림 3>AS/TS 추론전략

4.2. 實行

목표, 규칙, 추가정보 등으로 구성된 AS/TS KB는 原始화일(AS/TS.PRL)이 된다. 이 화일은 컴파일되어 이용자가 대화를 하며 직접 사용할 수 있는 實行화일(AS/TS.KNB)로 된다. 본 절에서는 AS/TS 實行상의 정보입력과 實行 과정을 보인다.

4.2.1. 情報入力 및 實行

실행화일을 로드(load)시키면 처음에 表題(title)가 나타난다. 이 표제에는 AS/TS 제목, 構築者, 개발日字 등이 나타나게 된다. 이 외에도 KE가 이용자에게 제시하고자 하는 정보도 기입하게 된다.

화면 하단에는 “PAGE”, “CONTINUE”, “MENU”, “HELP”, “EXIT” 등이 기능키에 의해 작동되도록 메뉴방식으로 나타난다. <F1키>(PAGE)는 한 화면에 모든정보를 제시할 수 없을 때 다음 화면(페이지)에 제시하는데, 이용자가

다음 화면을 보고자 할 때 사용하게 된다. <F2키>(CONTINUE)는 AS/TS를 계속 실행하고자 할 때 이용한다. <F2키>를 치면 정보입력이 시작된다. <F5키>(MENU)는 프로그램을 로드, 컴파일, 수정할 수 있는 단계로 되돌아 가게 하며, <F6키>(HELP)는 현재 화면상에 나타난 기능키의 역할에 대하여 설명하고 있다. <F7키>(EXIT)는 현재 실행을 중지하고 DOS 상태로 돌아갈 것인가를 확인하는 화면을 나타나게 한다.

AS/TS 사용단계에서 화면하단에 나타난 機能語(function command)에 대하여 이해하지 못하는 경우 <HELP키>를 이용하여 참조할 수 있다. 表題 상태에서 <CONTINUE키>를 선택하면, AS/TS는 최종 통계기법을 선택하기 위해 이용자로 부터 정보를 얻으려고 질문하게 된다. 질문과정에서 이용자는 AS/TS와 對話를 하게 되는데, 이는 크게 대상-속성 質疑, 수치정보 질의, 사실정보 질의 등으로 분리할 수 있다.

4.2.1.1. 對象-屬性 質疑

KE표지가 나타난 후 이용자는 AS/TS에 자신의 연구목적, 자료특성 등에 관한 정보를 제공해야 하는데, OAV질의의 경우 AS/TS는 다수의 항목을 배열한 후 이용자가 이 중 자신의 상황을 가장 잘 나타내는 항목 또는 특성들을 선택하도록 요구한다. <그림 4>은 AS/TS의 연구목적을 나타내고 있는데, 이용자는 자신의 연구목적에 해당되는 항목들을 선택함으로써 AS/TS는 이에 부합되는 분석방법을 일차적으로 선정하게 된다. 위 화면에 나타나지 않은 항목을 선택하고자 할 때는 <F1키>(PAGE)를 이용하면 다음 항목이 계속 나타난다.

이 質疑語의 대상은 연구목적으로

“the general purpose of analysis in your research or study”

이 되며, 이 대상에 대한 속성은 화면상에 이용자가 화살표로 선택하도록 제시되는 항목으로

“Range of values within which the parameter is lie”

“Estimation of parameter”

“Test the hypothesis in your research or study”

“Develop decision rules that minimize error”

“See whether a certain parameter has changed”

“Test difference between two group parameter”

“Determine whether group have identical means”

“Compare simultaneously all the group”

“Find the relationship between sets of observation”

“Measure the degree of relationship between samples”

등이 된다.

STATISTICAL TOOL SELECTION EXPERT SYSTEM

Select all the general purpose of analysis in your research or study

Range of values within which the parameter is lie

Estimation of parameter

Test the hypothesis in your research or syudy

Develop decision rules that minimize error

See whether a certain parameter has changed

Test difference between two group parameter

Determine whether group have identical means

Compare simultaneously all the group

Find the relationship between sets of observation

Measure the degree of relationship between samples

[1PAGE] [2UNKNOWN] [3REPORT] [4EXPAND] [5MENU] [6HELP] [7EXIT]

<그림 4> OAV정보 입력 화면

이 중 연구자는 자신의 연구목적에 가장 적합한 사항을 모두 선택하면 된다. 특성을 선택하는 데는

<방향키> 또는 <스페이스바>를 이용하여 화살표를 이동한 후 <RETURN키>를 친다. 연구에 부합되는 특성(항목)이 또 있으면 화살표를 이용하여 선택한 후 다시<RETURN>를 친다. 이와 같이 반복하여 모든 특성이 선택되면 <F7>(DONE)를 눌러 선택을 마치게 된다. 이와 같은 선택 과정은 규칙표현 중 대상-속성 표현에 의해 나타난다.

4.2.1.2. 數值情報 質疑

수치정보 질의는 사용자가AS/TS의 질문에 대하여 수치값을 입력하도록 한다. 사용자가 수치값을 입력하면 추론엔진은 KE가 KB내에 미리 정한 값과 비교한다. AS/TS가 화면에 나타내는 수치정보 입력화면은 <그림 5>와 같다.

표본크기가 “30” 이상인가 또는 미만인가에 따라 통계기법의 선택이 달라진다. 추정에 있어 “30” 이상이면 Z-통계량을 이용하며, “30” 미만이면 t-통계량 사용여부를 검토하게 된다.

STATISTICAL TOOL SELECTION EXPERT SYSTEM

What is :

Sample size

: _

[] [2UNKNOWN] [3REPORT] [4ESPAND] [5MENU] [6HELP] [7EXIT]

[그림 5] 수치정보 입력 화면

4.2.1.3. 事實情報 質疑

사실정보 질의에서 사용자가 입력할 수 있는 정보는 단지 주어진 항목의 참(true)과 거짓(false) 중 하나를 선택하는 것이다. <그림 6>은 AS/TS의 사실정보 질의 중 하나이다.

STATISTICAL TOOL SELECTION EXPERT SYSTEM

Dependent and independent variables are linearly associated

TRUE FALSE

[] [2UNKNOWN] [3REPORT] [4ESPAND] [5MENU] [6HELP] [7EXIT]

<그림 6>사실정보 입력 화면

- [10] W. A. Gale, *Artificial Intelligence and Statistics*, Addison-Wesley, 1986.
- [11] G. J. Hahn, "More Intelligence Statistical Software and Statistical Expert Systems : Future Directions", *The American Statistician*, Vol.39, No.1(Fabruary 1985), pp.1-16.
- [12] F. Hayes-Roth, "Rule-Based System," *Communications of the ACM*, Vol.28, No.9(September 1985), pp.921-932.
- [13] F. Hayes-Roth, "Knowledge-Based Expert System : Tutorial," *Computer*, Vol.17, No.9(September 1984), pp.11-28.
- [14] S. J. Hong, "Guest Editor's Introduction," *Computer*, Vol.19, No.7(July 1986), pp.12-15
- [15] M. J. R. Keen and G. Williams, "Expert System Shells Come of Age," in M. A. Blamer(ed.) *Research and Development in Expert Systems*, Cambridge, 1985, pp.13-21
- [16] S. Kumara, A. L. Soyster, and R. L. Kashyap, "An Introduction to artificial Intilligence," *IE*, Vol.18, No.12(December 1986), pp.9-20
- [17] *INSIGHT 2 Version 1.0*, Level Five Research, Inc., 1985
- [18] R. H. Michaelsen, D. Michie, and A. Boulanger, "The Technology of Expert Systems," *Byte*, April 1985, pp.303-312.
- [19] W. E. Remus and J. E. Kottimann, "Toward Intelligent Decision Support System : An Artificially Intellgent Statistician", *MIS Quarterly*, Vol.10 No.4(December 1986), pp. 403-418.
- [20] B. A. Thomson and W. A. Thomson, "Inside an Expert System," *Byte*, April 1985, pp.315-330
- [21] H. Schildt, *Advanced Turbo Ploglog : Version 1.1*, McGraw-Hill, 1987.
- [22] D. A. Waterman and F. Hayes-Roth, "An Investigation of Tools for *Building Expert System*," in F. Hayes-Roth, D. A. Waterman, and D. B. Lenat(eds.) *Building Expert Systems*, Addison-Wesley, 1983, pp. 169-215.
- [23] A. P. White, "Inferencies Deficiencies in Rule-Based Expert System," in M. A. Blamer (ed.) *Research and Development in Expert Systems*, Cambridge, 1985, pp.39-50.
- [24] W. A Woods, "Important Issues in Knowledge Representation," *Procedings of the IEEE*, Vol.74, No.10(October 1986),pp.1322-1334.
- [25] N. S. Yaghmai "Expert Systems : A Tutorial," *Journal of the American Society for Information Science*, Vol.35, No.5(September 1984), pp.297-305.

이용자가 참과 거짓을 판단해야 할 사항은

“Dependent and independent variables are linearly associated”

로 獨立變數와 從屬變수가 線形관계라고 판단되면 “TRUE”가 되며, 아니면 “FALSE”로 입력된다. 추론엔진은 이 정보와 일치하는 항목이 KB내에 존재하는가 탐색하게 된다.

위와 같이 OAV정보, 수치정보, 사실정보의 質疑-應答 과정을 통하여 AS/TS는 이용자가 사용할 수 있는 최종 기법을 선택하게 된다.

5. 結論

본 연구에서 구축한 AS/TS는 통계학에 관한 충분한 지식이 없는 이용자를 대상으로, 통계전문가의 조언을 얻어 통계기법을 선정하는 것과 동일한 효과를 얻도록 이용자의 연구목적, 자료 특성 등에 가장 적합한 통계기법을 선정하여 이용자에게 제시하기 위한 ES를 구축하였다. AS/TS는 ES가 통계기법을 선정하는 과정을 그대로 模倣한 프로그램으로 이용자가 제시한 정보를 사용하여 최종목표에 도달하게 된다. AS/TS는 이용자가 입력한 정보와 일치하는 대상을 확인할 때까지 이용자에게 추가적인 정보를 요구하도록 요구하면서 KB를 탐색하여 목표에 도달한다.

본 연구는 연구자들이 이용하게 될 다양한 기법을 다루고 있으므로 통계이론을 심도있게 다루거나 구체적인 모형을 제시하지 않더라도 이용자가 AS/TS를 사용함으로써 연구에 부적합한 통계기법을 選定함에 따라 발생할 수도 있는 자료처리 및 해석상의 誤謬를 제거할 수 있다.

AS/TS는 이용자가 사용하고자 하는 구체적인 통계모형 및 變數設定에 관한 방안은 제시하고 있지 않고, 통계처리 결과에 대한 해석은 각 통계패키지에 따라 出力樣式이 다르기 때문에 연구대상에서 제외하였다. 그러나 AS/TS가 효과적으로 연구자에게 사용되기 위해서는 샘플링用 ES, AS/TS, 통계패키지, 해석用 ES를 링크시키는 것이 필요하다고 판단된다. 이는 다음의 향후 연구되어야 할 분야라고 사료된다.

參 考 文 獻

- [1] 김 영 준, 염 봉 진, “회귀모형 설정을 위한 전문가 시스템,” 추계 인공지능 발표회, 1987년3월, pp.37-53.
- [2] 이 재 규, 서 민 수, “생산계획 시스템,” 추계 인공지능 발표집, 1986년 11월, pp.117-138.
- [3] B. D' Ambrosio, “Insight—A Knowledge System,” *BYTE*, April 1985, pp.345-347.
- [4] R. J. Brachman, S. Amarel, C. Engelman, R. S. Engelmores, and D. E. Wilkins, “What are Expert Systems ?,” in F. Hayes-Roth, D. A. Waterman, and D. B. Lenat(ed.) *Building Expert Systems*, Addison-Wesley, 1983, pp.31-57.
- [5] M. A. Bramer, “Expert systems : The Vision and Reality,” in M. A. Blamer(ed.) *Research and Development in Expert Systems*, Cambridge, 1985, pp.1-12.
- [6] B. G. Buchanan, D. Barsstow, R. Bechtel, J. Bennett, W. Clancey, C. Kulikowski, T. Mitchell, and D. A. Waterman, “Constructing an Expert Systems”, in F. Hayes-Roth, D. A. Waterman, and D. B. Lenat(eds.) *Building Expert Systems*, Addison-Wesley, 1983, pp.31-57.
- [7] E. E. Carter, “Creating A Shell-Based Expert System,” *Computer and Operations Research*, Vol.13, Nos.2/3, pp.325-327.
- [8] N. M. Cooke, J. E. McDonald, “A Formal Methodology for Acquiring and Representing Expert Knowledge”, *Proceedings of the IEEE*, Vol.74, No.10(October 1986), pp.1422-1430.
- [9] T. W. Finin, “Interactive Classification : A Technique for Acquiring and Maintaining Knowledge Bases,” *Proceedings of the IEEE*, Vol.74, NO.10(October 1986), pp.1414-1421.