

의사결정지원을 위한 지식표현 및 확률추론

김 성 식(한국교원대학교)

의사결정 문제가 의사결정자의 취향에 좌우되는 불확실한 상황에서 의사결정은 하나가 아닌 일련의 관련된 결정들로 구성된다. 규칙기반 전문가 시스템은 그와같은 의사결정 문제를 표현할 수 없다. 이 논문은 그와같은 문제들을 해결하기 위하여 모델기반 지식표현과 확률추론을 결합하여 자문시스템 IDPI를 제시하고 그와같은 방법을 사용하여 대학생들을 위한 진로자문 시스템을 구현한다.

Decision making in uncertain situations that preferences of decision maker is important consists of a series of related decisions. Rule-based expert system can not represent such a complex decision problems. For these decision problems, this paper suggests a new methods IDPI(Influence Diagram-based Probabilistic Inference) which combines model-based knowledge representation and probabilistic inference, and implements a career counsellor for the university students using the combined methods.

1. 서론

전산학의 궁극적인 목표는 일상적인 인간활동의 대부분을 컴퓨터가 대신하고 인간은 보다 질이 높은 문화생활, 여유롭고 풍요로운 삶을 영위하도록 하는 데에 있다고 말할 수 있다. 특히 인공지능에서 가정하는 컴퓨터는 스스로가 지식을 축적 및 창조하고 상황을 판단하며 의사결정을 내리고 합리적인 활동을 수행하는 지능적인 컴퓨터이다.

인간의 활동 중에서 가장 어렵고 중요한 것이 의사결정이다. 인간과 마찬가지로 인공지능 컴퓨터에게도 가장 중요한 활동이 의사결정 문제이다. 인공지능 컴퓨터는 합리적인 활동을 수행하는 컴퓨터이며 합리적인 활동은 합리적인 의사결정을 전제로 가능한 것이다. 이 논문은 그

와 같은 결정을 위한 지식표현 및 확률추론 방법을 제시한다.

컴퓨터를 이용하여 개인이나 조직의 의사결정을 지원, 의사결정의 질을 높이기 위한 노력은 오랫동안 전산학에서도 관심의 대상이 되어 왔다. 종래에는 데이터베이스를 이용한 의사결정 지원 시스템이 사용되었고, 최근에는 특정분야의 전문지식을 저장하고 추론기법을 활용한 전문가 시스템들이 개발되어 성공적으로 활용되고 있다. 그러나 전문가시스템은 불확실성이 개입될 경우에는 적용 불가능하다. 불확실성이란 의사결정의 결과가 결정된 활동이 수행되는 상황에 좌우된다는 뜻이다. 또한 전문가시스템은 의사결정자의 특성이나 신념을 표현할 수 없다.

불확실하고 복잡한 상황에서 의사결정은 하나가 아닌 일련의 관련된 결정들로 구성된다. 각각의 결정은 그 구조와 표현방법이 다를 수도

있다. 따라서 다양한 의사결정 상황을 표현하고 논리적 추론을 수행할 수 있는 새로운 의사결정 시스템을 설계할 필요가 있다. 이 논문은 그와 같은 문제들을 해결하기 위하여 모델기준 지식 표현과 확률추론을 결합하는 영향흐름도를 통한 확률추론 IDPI (Influence Diagram-based Probabilistic Inference)를 제시하고 그와같은 방법을 사용하여 대학생들을 위한 진로자문 시스템을 구현한다.

2. 전문가 시스템의 취약점

전통적인 의사결정지원 시스템은 관리자의 합리적인 의사결정을 지원하기 위하여 조직에 부착된 컴퓨터 시스템과 결합된 데이터베이스 시스템을 의미하며, 이와같은 시스템은 대규모의 정보를 보관하고 문제에 대한 기초자료의 제공이 그 임무이었다[1]. 따라서 주요 역할은 보관하고 있는 자료의 수정, 정리 또는 통계치나 보고서의 산출에 불과하였다. 이와같이 전통적인 의사결정 지원 시스템은 최초의 설계목적에 따라 필요한 자료를 산출하는 프로그램을 만들어 두고 의사결정에 활용하는 것이었다. 그러나 이와같은 방법은 단순한 자료의 모집에 불과하며 새로운 문제가 발생할 경우에는 아무 쓸모가 없고 그 분야의 전문가에게만 필요하다는 단점이 있다[6].

전통적인 의사결정 지원 시스템의 이와같은 단점을 극복하기 위하여 등장한 것이 전문가 시스템이다. 전문가 시스템은 지능적인 활동을 수행하는 인공지능 프로그램의 일종이며, 화학분석, 광물분석, 의학진단 등의 분야에 활용되고 있으며, 자료의 해석에서 고장분석, 질병진단, 복잡한 구조물의 분석 및 일련의 작업계획 등 다양한 분야에 사용되고 있다[2].

전문가 시스템은 일련의 방대한 지식을 규칙에 따라 분류하여 지식베이스를 구축하고 다양한 형태의 질의에 응답 자문하여 준다. 이와같이 지식의 분류와 자문을 통한 의사결정 지원

시스템의 대표적인 사례는 MYCIN 전문가 시스템이다. MYCIN은 질병의 유형을 정밀하게 분류하여 놓고 증상에 따른 질병의 가능성을 확실히와 함께 규정하고, 환자의 증상에 대한 정보가 들어오면 질병을 진단하고 적절한 투약방법을 추천하는 의사결정 지원방법이다[2].

전문가 시스템은 데이터베이스를 이용하는 전통적 의사결정 지원방법에 비하여 여러가지 장점을 가지고 있다. 첫째로 전통적인 의사결정 지원방법은 용도중심이므로 작업 처리 절차에 따라 데이터를 저장하고 있으나 전문가 시스템의 지식베이스는 IF THEN 형식에 따라 문장형태로 저장되어 있으며 가감 삭제에 아무 제한이 없다. 즉 외부 프로그램이나 처리절차에 대하여 완전히 독립적이다[1].

전문가 시스템의 또다른 장점은 시스템 자신의 결정이나 추론활동에 대하여 스스로 설명할 수 있는 기능을 가지고 있다는 점이다. 이와같은 기능은 사용자가 전문가 시스템의 도움을 받아 의사결정을 하는데에 확실감을 주는 훌륭한 기능이다. 전문가시스템은 전문가의 많은 경험과 특수한 지식이 필요한 분야에 있어 매우 효과적인 의사결정 지원 시스템이며, 인공지능의 성공적인 활용을 보여준 대표적인 사례이다.

그러나 널리 알려진 바와같이 전문가 시스템은 지식베이스에 없는 새로운 유형의 증상을 가지고 있는 경우에는 이를 처리하지 못하며, 상호배타적인 대안과 복수원인을 올바르게 표현하지 못한다는 단점이 있다[5].

전문가 시스템의 보다 중요한 취약점은 의사결정자를 무시하고 있다는 점이다. 예를들면, MYCIN을 사용하는 의사결정자가 자문결과 제시된 약을 사먹고 나타나는 부작용을 보다 중요시하고 있다든가 전문가 시스템이 제시하는 방법이 아닌 다른 방안을 생각하고 있는 경우에는 적합하지 않으며, 의사결정자를 전문가 시스템의 복잡한 규칙을 사용하도록 강요한다는 단점이 있다.

이와같은 문제들을 해결하기 위하여 본 논문

은 모델베이스 지식표현과 확률추론에 의한 대안비교 방법을 제안한다. 그러나 이 논문에서 주장하는 모델베이스 지식표현이나 확률추론은 전문가시스템의 결점을 보완하는 전문가시스템의 일종으로 간주할 수 있다.

3. 모델베이스 문제표현

3.1 영향흐름도에 의한 문제표현

영향흐름도는 오래전부터 의사결정 문제를 그래프 형태로 표현함으로써 사건들 사이의 관계를 시각적으로 설명하는 도구로 사용되었다[9]. 예를들면 어느 도매상에서 제조비용 C에 따라 어떤 제품의 가격을 P로 결정하면 판매량이 Q 일때 소득은 R이 된다는 관계를 영향흐름도로 표현하면 그림 1과 같다.

그림 1에서 가격 P는 의사결정 노드이며 제조비용 C와 판매량 Q는 기회노드이다. 즉, 가격은 제조회사에서 임의로 선택할 수 있는 사항이나 제조비용이나 판매량은 일종의 불확실성이며 확률로 표시된다. 이 확률중에서 제조비용은 이 문제를 분석할 당시의 경제구조나 유통구조를 분석하여 결정해야 하며 판매량의 확률은 제조회사의 의사결정에 따라 결정된 가격에 좌우된다.

또한, 수익 R은 이 의사결정 문제에서 가장 중요한 사항으로 가치노드이나 여기에서는 가격과 판매량에 따라 자동적으로 결정된다. 이와같이 불확실한 의사결정 문제는 영향흐름도에 의하여 간명하게 표시된다. 영향흐름도는 모든 원인을 하나의 결과 즉, 가치노드로 모으기 때문에 대안의 비교를 명확하게 할 수 있다[13].

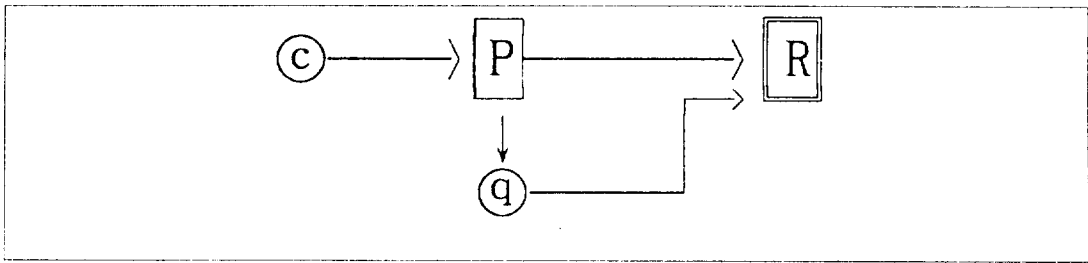


그림 1. 영향흐름도에 의한 의사결정 문제의 표현

3.2 영향흐름도의 작성

의사결정 문제는 그림 2와 같은 과정을 거쳐 영향흐름도가 작성된다. 즉, 1) 문제의 구체화, 2) 문제해결 모델구성, 3) 명료화, 4) 모델 재구성 등의 과정을 거쳐 영향흐름도를 변형시키며 의사결정자의 마음에 드는 모델이 형성되면 최종적인 영향흐름도로 규정한다.

문제의 구체화는 대안과 대안의 결과를 탐색하면서 각 대안별로 중요성이나 기대치 등을 규정하는 단계이며, 문제해결 모델구성은 구체화된 문제의 대안과 결과에 대한 예측을 토대로 가상세계의 모델을 구성하는 단계이다.

문제의 명료화는 구성된 모델을 변형하여 문제해결 모델을 재구성하는 것으로 필요에 따라

문제의 구체화와 모델구성 단계를 반복해야 한다. 이와같은 과정을 거쳐 의사결정자가 마음에 드는 최종적인 영향흐름도를 얻는다.

영향흐름도에 의한 의사결정 문제의 표현은 영향흐름도의 구성 및 변형에 의하여 동적으로 불확실성의 문제를 취급한다는 장점이 있다. 따라서 의사결정문제를 표현하는 방법으로는 영향흐름도 방법을 사용하고 완성된 영향흐름도를 신뢰망으로 변형시켜 이 신뢰망에 확률추론을 적용하여 대안을 비교하는 방법을 사용하면 훌륭한 의사결정 지원방법이 될 수 있다. 이와같은 방법을 이 논문에서는 영향흐름도에 기초한 확률추론 IDPI(Influence Diagram based Probabilistic Inference)라고 부르고 있다. IDPI는 영

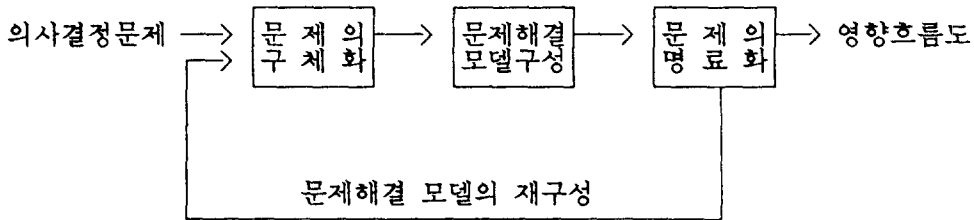


그림 2. 영향흐름도의 작성

영향흐름도에 기초한 의사결정 문제의 모델링과 구성된 모델을 신뢰망으로 변형하여 확률추론을 수행하는 두가지 작업에 중점을 둔다.

3.3 모델베이스 문제표현의 구성요소

영향흐름도를 사용하는 모델베이스 의사결정은 문제를 명확히 규명하는 과정이 문제풀이의 일부가 된다. 문제의 명확한 규명을 위하여 필요한 사항들은 다음과 같다.

- 1) 주어진 사실들과 그 사실들로부터 추론할 수 있는 사실들
- 2) 의사결정자가 선택할 수 있는 몇개의 처리 가능한 대안들
- 3) 대안수행에 영향을 주는 요소들
- 4) 중요요소들을 고려한 대안의 수행결과
- 5) 대안의 수행결과에 대한 확률분포

의사결정은 크게 두가지 방식으로 수행될 수 있다. 하나는 "어떤 조건들이 만족하면 어떤 행동을 반드시 수행해야 한다" 라는 경험적 방식이고, 다른 하나는 추론을 통하여 최선의 대안을 선택하는 의사결정론적 의사결정이다. 영향흐름도에 의한 문제표현 단계는 경험적 의사결정에 따라 모델을 단순화하는 과정이다.

IDPI의 모델베이스 의사결정의 초기단계는 영향흐름도로 표현된 모델을 새로운 모델로 재구성하는 모델평가와 재구성 작업을 반복 수행함으로써 문제를 명확히 규정하는 것이다. 이와 같은 방식은 종래의 추론방식이 주어진 모델에 의의를 제기하지 않고 결론을 이끌어내는 데 비

하여 의사결정 문제를 표현하는 모델 자체에 대한 평가 및 재구성으로 문제를 해결하려는 방식이다. 이와같은 방법은 사용자의 기호에 맞추어 동일한 문제를 여러가지 형태의 모델로 표현할 수 있다는 장점이 있다.

규칙기준 전문가시스템과 모델기준 의사결정 시스템 중 어느것을 선택하느냐는 문제는 주어진 의사결정 문제의 특성에 따라 달라질 수 있다. 의사결정 문제와 관련된 요인들을 살펴보면 1) 상황의 복잡성, 2) 대안의 다양성, 3) 대안 선정결과의 불확실성, 4) 대안의 상호관련성, 5) 대안결과에 대한 선호도의 크기, 6) 결정결과 나타날 현상중에 치명적인 요소의 유무 등이 있다.

의사결정 문제 중 의사결정과정이 최초로 의도한 대로 진행되는 단순한 경우에는 규칙기준 전문가 시스템을 사용하고 결과에 대한 불확실성이 크고 결정의 결과에 대한 이해관계가 중요한 문제는 모델중심 의사결정 시스템을 사용하는 것이 좋을 것이다. 모델중심 의사결정 시스템은 의사결정 문제를 영향흐름도로 표현하는 과정에서 사용자의 의도에 따라 적절한 타협이 이루어지므로 규범적인 최적화가 아니라 절차적인 최적화이며 제한된 합리모형이다[14].

3. 신뢰망을 이용한 확률추론

3.1 신뢰망과 영향흐름도의 결합

신뢰망(belief net)은 일련의 변수 또는 노드들

과 변수중 단말 노드들에 대한 한계 확률분포, 변수들 간의 방향성 링크들에 대한 조건부 확률 분포로 구성된다. 신뢰망의 모든 노드는 취할 수 있는 값이 상호배타적이며 그 값을 가질 확률분포가 사전에 또는 사후에 주어진다.

네트워크 형태로 모델을 구성하고 확률정보를 통하여 추론을 수행하는 방법으로 널리 사용되는 신뢰망과 영향흐름도는 다음과 같은 차이가 있다[9]. 영향흐름도는 신뢰망의 각 노드를 확률 노드 또는 기회노드라고 부르고 여기에 결정노드와 가치노드를 추가한 것이다[13]. 신뢰망과 영향흐름도의 주된 차이점은 신뢰망에는 의사결정이나 선택노드가 없다는 점과 영향흐름도는 모든 원인을 하나의 결과 즉, 가치노드로 모으기 때문에 대안의 비교를 명확하게 할 수 있다는 점이다.

따라서 영향흐름도의 형태로 표현된 의사결정 문제에 신뢰망을 이용한 확률추론을 적용하는 방법을 택한다. 즉, 영향흐름도의 결정노드를 신뢰망의 기회노드로 표현하면서도 사용자의 선택에 좌우되는 노드로 규정하고, 모든 확률이 최종적인 노드에 누적되도록하여 최종노드를 가치노드화하여 확률추론을 적용하는 방법을 사용한다. 이와같은 방법은 네트워크를 따라 상호배타적인 정보를 자연스럽게 표현할 수 있고, 전문

가의 의견과 의사결정자의 의견을 동시에 수용할 수 있다는 점에서 전문가 시스템의 단점을 보완한다. 신뢰망을 이용한 확률추론에서는 전문가와 사용자로부터 관찰되는 정보에 따라 신뢰망의 노드와 링크에 확률을 부여하고 구하고자 하는 정보를 매트릭스 곱셈으로 손쉽게 구할 수 있다[그림 3].

3.2 상징적 확률추론

신뢰망을 이용한 최초의 확률추론 방법은 Pearl에 의하여 제시되었다[10]. 그는 신뢰망의 각 노드에 Lambda와 Pi 분포를 추가하여 입력 데이터에 대한 인과적 추론과 진단적 추론을 병행 전개하는 방법을 사용하였다[8]. Pearl이 제시한 확률추론 방법들은 전술한 확률추론의 장점을 모두 가지고 있으나, 데이터 중심이며 질의가 발생할 때마다 질의와 관계없는 모든 Lambda, Pi 값들도 수정하므로 비능률적이다 [12].

이와같은 수치적 접근 방법은 대안비교의 방법으로는 우수하나 deKleer의 ATMS[3]나 Sherlock 진단시스템[4]과 같이 복수의 해답을 구하고 추론의 결과를 누적적으로 활용할 수 없다. 스텐포드 대학의 Shachter 교수에 의하여 개발된 상징적 확률추론(SPI : Symbolic

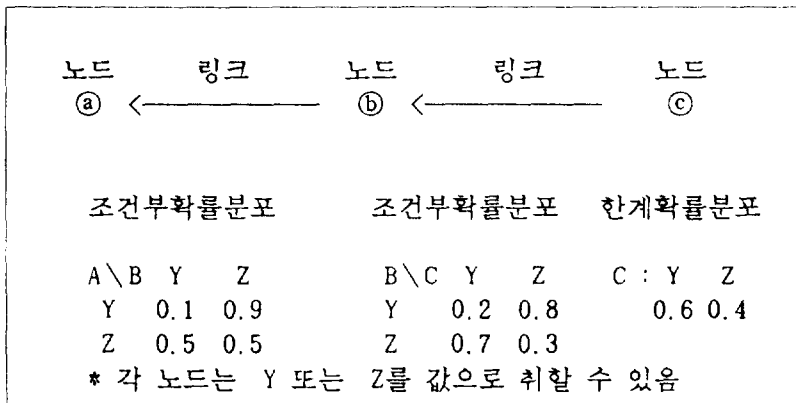


그림 3. 신뢰망에 확률부여 방법

Probabilistic Inference)은

- 1) 신뢰망의 각 노드에 상징식을 저장해 두고,
- 2) 관찰의 결과는 상징식에 추가해 두며,
- 3) 질의가 발생하면 상징식들을 수치화하여 구하고자 하는 확률을 계산하는 방법이다[12]. 상징적 확률추론은 관찰의 결과를 누적시켜 빠른 시간내에 적은 수의 정보를 가지고 명확한 답변을 제시할 수 있다.

상징적 확률추론은 신뢰망을 분할하여 트리구조 형태로 변환하고 관찰의 결과에 따라 상징식

을 추가 또는 수정하고, 질의에 따라 확률을 계산하는 3가지 단계를 거치는데 가장 중요한 것은 첫단계인 신뢰망의 분할과 분할된 신뢰망을 상징식으로 표현하는 것이다. 신뢰망의 분할은 확률계산에 소요되는 변수들의 수를 줄임으로써 수행시간을 크게 단축시킨다.

그림 4는 SPI를 설명하기 위한 신뢰망 사례이며, 그림 5는 그림 4를신뢰망의 분할에 의하여 트리구조 형태로 변형한 것이다. 그림 4에서 각 노드의 한계확률은 다음 공식으로 표현된다.

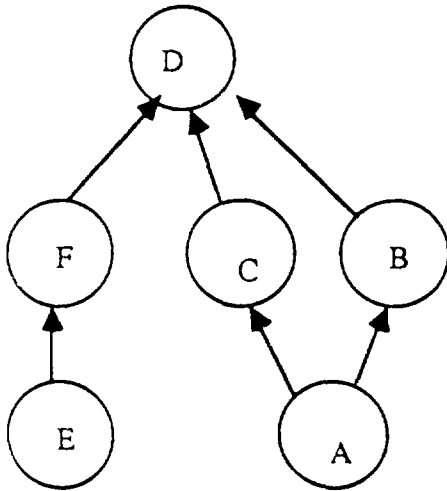


그림 4 간단한 신뢰망 사례

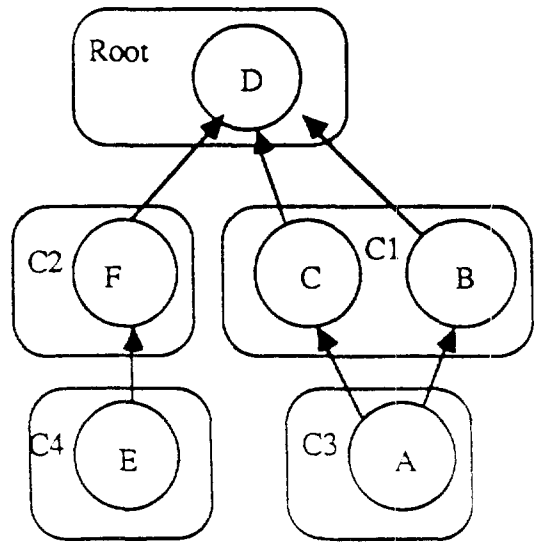


그림 5 신뢰망 사례의 분할

$$P(A) = P(A)$$

$$P(B) = P(B|A)P(A)$$

$$P(C) = P(C|A)P(A)$$

$$P(D)=$$

$$P(D|B\&C\&E)P(F|E)P(E)P(C|A)P(B|A)P(A)$$

$$P(E) = P(E)$$

$$P(F) = P(F|E)P(E)$$

SPI 알고리즘은 첫단계로 그림 4의 신뢰망을 그림 5와 같이 분할하여 트리구조로 변환시키는 것이다. 이와같이 신뢰망을 분할하여 변환된 그림 5와 같은 신뢰망에서 각 노드의 한계확률은 그 노드의 조건부확률과 직계부모 노드의 한계확률만으로 표현하며 다음과 같이 단순화된 상징식을 사용한다.

$$\text{exp}(A) = A$$

$$\text{exp}(B) = (B|A)A$$

$$\text{exp}(C) = (C|A)A$$

$$\text{exp}(D) = (D|BCF)BCF$$

$$\text{exp}(E) = E$$

$$\text{exp}(F) = (F|E)E$$

여기에서 exp는 expression을 뜻하고 (|)는 조건부확률을 뜻한다. 예를들면 한계확률 P(D)는 다음과 같은 상징식의 전개로 얻을 수 있다.

$$\text{exp}(D) = (D|BCF)BCF$$

여기에서 (D|BCF)는 조건부 확률로 그림 5의 루트 노드가 가지고 있으며, BCF는 자손노드 C1에서 BC, C2에서 F의 확률을 받아온다. C1은 BC를 확장하여 다음 공식을 얻으며 이때 (B|A)(C|A)는 C1이 갖고 있다.

$\text{union}((B|A)A), ((C|A)A) = (B|A)(C|A)A$
 C2는 F를 확장하여 (F|E)E를 얻으며 (F|E)는 C2에 갖고 있다. C3에서 A의 한계확률 C4에서 E의 한계확률을 갖고 있으므로 구하고자 하는 확률을 모두 계산할 수 있다. 확률값을 구하기 위한 실행시간은 확률공식에 나타나는 변수의 수가 증가함에 따라 지수적으로 증가하므로 이와같이 분할 방식을 통하여 확률변수의 수를 줄이는 것이 중요하다. 예를들면 그림 4에서 P(D)는 6개의 변수가 있으나 그림 5에서 exp(D)는 4

개의 변수만 있게 된다.

상징적 확률추론의 분할 알고리즘은 다음과 같은 조건을 만족해야 한다. 1) A가 B의 부모노드이면 A는 B와 다른 분할에 속해야 한다. 2) A가 B의 부모노드이면 분할 트리에서 A는 B의 자손분할에 속해야 한다. 3) 각 분할은 오직 하나의 부모분할만을 갖는다. 이와같은 제한조건은 분할의 결과 트리구조를 형성하게 되고 원래의 그래프에서 루트노드는 단말노드가 되고 분할트리는 원래 그래프를 뒤집어 놓은 모습이 된다. 분할트리에서 루트는 하나이어야 한다.

3.3 신뢰망용 이용한 확률추론 프로그래밍

그림 6은 그림 3의 신뢰망과 확률분포 사례를 프로그램으로 표현하고 조회하는 방법을 나타낸 것이다. 노드 B의 한계확률은 노드 C의 한계확률 (0.6 0.4)과 링크 B←C의 조건부확률 ((Y 0.2 0.8) (Z 0.7 0.3))의 매트릭스 곱셈으로 구하여 (0.4 0.6)이 됨을 알 수 있고, 노드 A의 한계확률은 앞에서 구한 노드 B의 한계확률 (0.4 0.6)과 링크 A←B의 조건부확률 ((Y 0.1 0.9) (Z 0.5 0.5))의 매트릭스 곱셈으로 구하여 (0.34 0.66)이 됨을 알 수 있다.

신뢰망에서 루트노드는 상징적 확률추론에서 단말노드가 되고, 단말노드는 루트노드가 된다. 네트워크 형태인 신뢰망을 사용하기 간편하도록 트리구조로 바꾸는 과정에서 링크의 방향이 바뀌기 때문이다.

4. 진로지도 프로그램

4.1 영향흐름도에 의한 진로지도 문제표현

의사결정 이론가에 따르면[7] 의사결정자는 의사결정을 행하기 전에 먼저 의사결정의 결과 얻게 될 이익과 손실을 분야별 항목별로 대차대조표(Balance Sheet)에 기록하고 그 결과를 토대로 의사결정을 행하는 것이 결정의 질을 높일 수 있다고 주장한다. 그러나 이와같은 방법을 카운셀러의 도움으로 수행하는 것은 위험한 방

```

:: Define the nodes and values of the nodes
  (var (A Y) (A Z))
  (var (B Y) (B Z))
  (var (C Y) (C Z))
:: Define the marginals.
  (m C (0.6 0.4))
:: Define the conditionals.
  (conds (A (B)) : A ← B
    ((Y) 0.1 0.9)
    ((Z) 0.5 0.5))
  (conds (B (C)) : B ← C
    ((Y) 0.2 0.8)
    ((Z) 0.7 0.3))
:: Query the values of node B
  (observe (b))
  (p ((b))) = (0.4 0.6)
:: Query the values of node A
  (observe (a))
  (p ((a))) = (0.34 0.66)

```

그림 6. 신뢰망을 이용한 확률추론 프로그래밍

법이다. 카운셀러의 주관에 좌우되어 특정항목을 지나치게 강조하거나 소홀히 함으로 인한 오류가 개입되기 때문이다[7]. 따라서 대학생들이 졸업후에 어떠한 분야에 직업을 가지고 일할 것인가를 선택하는 진로결정 문제를 지도하여 주는 진로지도(Career Counseller) 프로그램을 개발할 필요성이 있다.

진로지도 프로그램에 사용되는 영향흐름도는 모델베이스 지식표현 방법으로 구성된다. 영향흐름도의 구성요소에는 사실, 대안, 고려요소, 확률분포 등이 있다. 설계자는 먼저 개략적인 구성요소들을 가지고 있어야 한다. 그러나 의사결정자가 고려하고 있는 요소들을 수용할 수 있어야 한다. 즉, 사용자와 상호작용적으로 영향흐름도를 구성하되 사용자를 안내할 수 있는 수준의 지식들은 미리 가지고 있어야 한다.

여기에서는 진로지도 문제를 대학 4학년 학생이 졸업후 선택할 직업을 결정하기 위하여 고려

되어야 할 요소로서 1) 수입, 난이도, 자유시간 등 의사결정자의 개인적이익과, 2) 지위, 부모의 후원, 소속집단, 친구 등 사회적이익, 3) 가족적 배경, 학문적 배경, 생활철학(공헌도, 정당성, 장래성) 등 생활상의 이익 등 3가지 요소를 설계자가 제공하고 그이외에 사용자가 생각하는 중요요소는 모델평가 및 재구성 작업을 통하여 추가하는 것으로 하였다. 이와같은 사실들의 발굴 작업과 동시에 사실들을 분류하면 완성되는 영향흐름도를 통한 확률추론의 수행시간을 단축시킬 수 있다.

다음에는 의사결정 대안의 발굴이다. 사실들의 수집과 마찬가지로 의사결정 대안도 시스템이 제공하는 대안과 사용자가 생각하고 있는 대안을 수용하는 데, 중요한 것은 너무 많지 않아야 한다는 것이다. 진로지도 문제에서는 선택할 수 있는 대안 즉 직업이 취할 수 있는 값을 '대학', '정부', '산업계', '개인사업' 등으로

정하였다.

영향흐름도에 필요한 정보는 각 대안들이 취할 수 있는 값의 종류 그리고 각 값들이 갖는 확률분포이다. 대안이 취할 수 있는 4가지 값들은 상호배타적이며 이들에 대하여 영향을 미치는 의사결정자의 개인적배경, 사회적배경, 가족적배경, 학문적배경 및 생활철학 등을 중심으로 하나의 조건부 확률분포를 부여하여야 한다.

이와같은 정보들은 사용자와 대화를 통하여 어떻게 영향흐름도를 구성하는가에 따라서 결정된다. 즉, 그것이 결정노드이면 사용자가 대화식으로 부여하고 확률노드이면 사전에 지식전문가나 설문조사 또는 통계조사를 통하여 객관적으로 부여한다. 이 논문에서는 편의상 주관적으로 부여하였다. 어떤 노드를 결정노드로 할 것인가 기회노드로 할 것인가는 사용자가 결정한다.

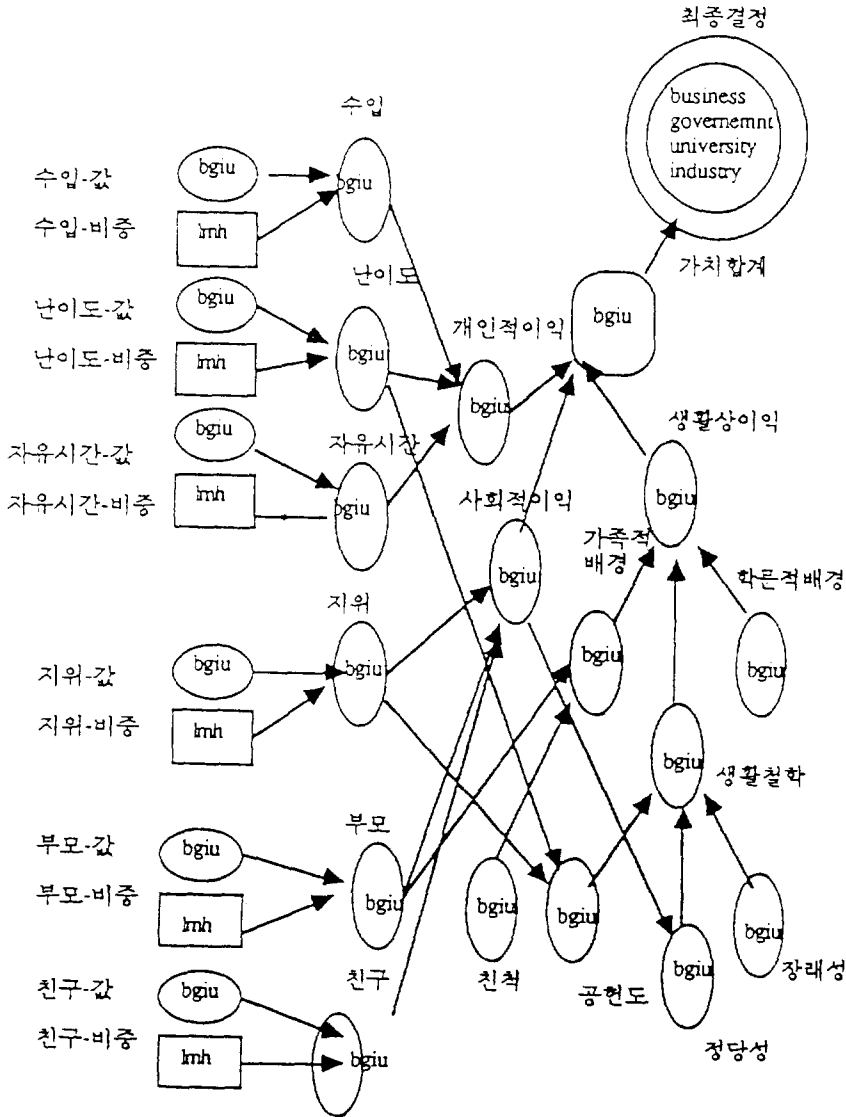


그림 7. 진로지도 문제의 표현 (신뢰망 형태)

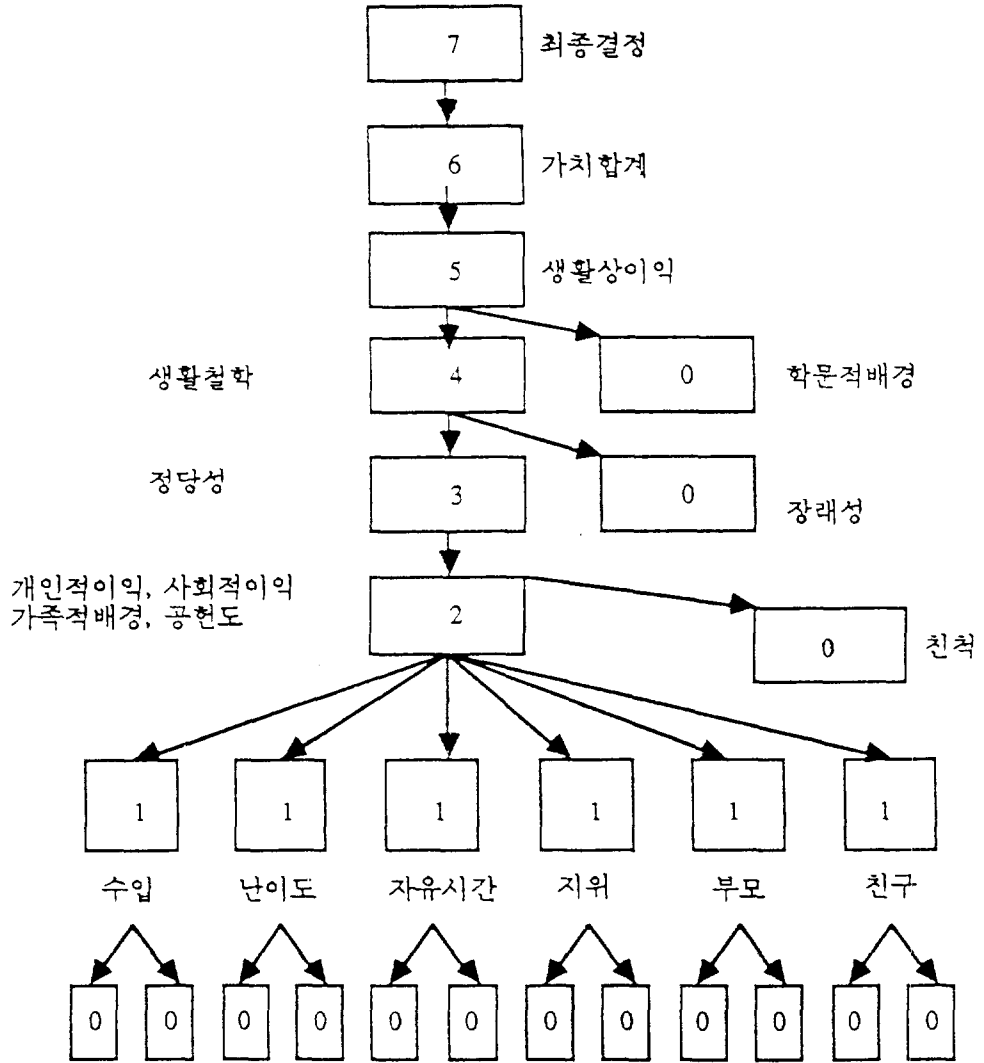


그림 8. 진로지도 문제의 표현 (트리 형태)

그림 7은 영향흐름도의 형태로 진로지도 문제를 표현한 것이며, 그림 8은 이와같이 표현한 영향흐름도를 트리형태로 분할한 것이다. 그림 7에서 결정노드는 직사각형, 가치노드는 완만한 직사각형, 확률노드는 원형, 단말노드는 이중원으로 표현하였다.

개인적이익과 사회적이익의 구성요소들은 본인 스스로 그 가중치를 결정하게 하는 결정노드로 하였다. 결국 모든 요소들의 영향이 가치노드인 이익의 합계로 누적되어 최종 의사결정을 하는데 대안을 비교 가능하게 된다. 선택 가능한 대안의 수는 개인사업(b), 정부(g), 산업(i), 대학(u) 등 4가지 분야로 제한하고 가중치는 상, 중, 하(low, middle, high)로 구분하여 선택하도록 하였다.

진로지도 문제풀이에 사용되는 입력은 그림 7에서 변수들과 변수들이 취할 수 있는 값, 그리고 한계 확률분포와 조건부 확률분포이다. 여기에서는 편의상 대부분의 자료를 사전에 입력하고 사용자는 결정노드에 해당되는 변수들에 대하여만 자료를 입력시킬 수 있게 하였다. 즉, 사용자는 개인적 소득과 사회적 소득의 구성요소들에 대한 가중치를 선택할 수 있는 것이다. 출력은 본인이 희망하는 영향흐름도내에 어떠한 한계확률도 찾아볼 수 있다.

4.2 확률추론

부록에는 진로지도 문제의 신뢰망에 확률을 부여하는 프로그램과 프로그램 수행의 일부 입, 출력 과정을 수록하였다. 사용자의 직업선택 선호도는 개인사업, 정부기관, 산업체, 대학의 순서로 확률이 표시되어 있고 출력은 개인적이익, 사회적이익, 생활상이익, 총이익에 의한 직업선택 등의 확률을 표시한다. 산출 결과를 살펴보면 이 사용자는 최종적으로 CAREER-CHOICE의 값이 (개인사업 정부기관 산업체 대학)의 순서로 (0.24 0.23 0.22 0.28)의 확률을 나타내고 있으므로 대학분야에 진출함이 적절함을 알 수 있다. 프로그램의 출력 결과에서 (개인사

업 정부기관 산업체 대학)의 선택 대안들이 사용자가 가중치의 값을 선택함에 따라서 경쟁적으로 증감되고 있는 것을 볼 수 있다.

5. 결론

본 논문은 영향흐름도를 이용한 모델베이스 문제표현 방법과 신뢰망을 이용한 확률추론을 결합하여 새로운 의사결정 지원방법 IDPI를 제시하고, IDPI를 사용하여 대학생들의 진로지도 프로그램을 사례구현한 것이다.

영향흐름도에는 의사결정노드가 있어서 사용자의 의견을 수렴할 수 있고, 또한 모든 원인을 가치노드로 모으기 때문에 대안의 비교를 명확하게 할 수 있다는 장점이 있다. 따라서 IDPI는 의사결정 문제를 영향흐름도의 형태로 표현하고 여기에 신뢰망을 이용한 확률추론을 적용한다. 즉, 영향흐름도의 결정노드를 신뢰망의 기회노드로 표현하면서도 사용자의 선택을 허용하고, 모든 확률을 최종적인 노드에 누적시켜 최종노드를 가치노드화하는 방법을 사용한다.

IDPI는 전문가의 의견과 의사결정자의 의견을 동시에 수용할 수 있고 네트워크를 따라 상호배타적인 정보를 자연스럽게 표현할 수 있다는 점에서 전문가 시스템의 단점을 보완한다. 사례구현에 사용된 언어는 Allegro Franz Lisp 이고, 사용된 시스템은 SUN 3.0/160이다. 이 논문에서 주장하는 모델베이스 지식표현과 확률추론은 MYCIN 유형의 전문가시스템의 결점을 보완하는 전문가시스템 개발도구로 사용될 수 있다.

참 고 문 헌

- [1] Breese, J.S., "Knowledge Representation and Inference in Intelligent Decision System", *T.R. of Rockwell International, Palo Alto*, 1987
- [2] Buchanan, B. and Shortliffe, E., *Rule-based Expert System: The MYCIN Experiment of the Stanford Heuristic Programming Project*, Addison-Wesley, 1984
- [3] DeKleer J., "An Assumption-based TMS", *Artificial Intelligence Vol. 28*(1986), PP. 127 - 162
- [4] DeKleer J. and Williams B., "Diagnosis with behavioral modes", *Proceedings of American Association of AI.*(1989), PP 1324-1330
- [5] Heckerman D.E., Horvitz E.J., "On the Expressiveness of Rule-based Systems for Reasoning with Uncertainty", *Proceedings of AAAI*, 1987.
- [6] Holtzman S., *Intelligent Decision Systems*, Addison-Wesley, 1989
- [7] Janis I, Mann L., *Decision Making*, Yale Univ. Press, 1977
- [8] Kim, Jin Hyung, "CONVINCE: A Conversational Inference Consolidation Engine", *UCLA Ph.D Dissertation*, 1983
- [9] Oliver, R.M., & Smith, J.q., *Influence Diagram, Belief Nets and Decision Analysis*, John Wiley & Sons, CA, 1991
- [10] Pearl J., "Fusion, Propagation, & Structuring in Belief Networks", *Artificial Intelligence Vol. 29*, 1986
- [11] Pearl J., *Probilistic Reasoning in Intelligent Systems*, Morgan Kaufman, CA, USA, 1988
- [12] Shachter R, D'ambrosio B, and Del Favero B, "Symbolic Probabilistic Inference in Belief Networks", *Proceedings of AAAI.*(1989)
- [13] Shachter, R.D., "Probabilistic Inference and Influence Diagrams", *Operations Research*, Vol. 36, No. 4(1988), PP. 589 - 64
- [14] Simon, H.A., *The Science of The Artifical*, MIT Press, 1981
- [15] 김성식, 황종선, "상징적 확률추론과 의사결정지원", 「한국정보과학회 춘계학술발표논문집」, 18권, 1호(1991), pp 97-100, 1991
- [16] 김성식, 황종선, "합리적 활동을 수행하는 에이전트 이론모형", 「한국정보과학회 논문지」, 제 19권 제 2호(1992), pp 187-199, 1992
- [17] 김성식, "확률추론을 이용한 학생지도", 「한국정보과학회 전산교육연구회 추계학술발표논문집」, pp 163-180, 1992
- [18] 김성식, "인공지능 멀티미디어시대의 컴퓨터교육", 「한국정보과학회 전산교육연구회 춘계학술심포지움」, pp 31 - 45, 1993

부록 1. 영향흐름도에 확률부여 방법

```

:: Variables => 수입의 중요도는 상,중,하의 어느 하나를 취하고,
::              수입의 값은 사업,정부,기업체,대학 중에서 하나를 택함
::              수입의 중요도와 값이 결합하여 수입을 결정하며,
::              의사결정자는 사업,정부,기업체,대학 중에서 하나를 택함.
(var (income-weight low) (income-weight middle) (income-weight high))
(var (income-value business-best)      (income-value government-best)
    (income-value industry-best)      (income-value university-best))
(var (income business-best)            (income government-best)
    (income industry-best)            (income university-best))

:: Marginal Probabilities => 한계확률은 모든 루트노드에 부여함

:: 가치의 보편적 기준   b   g   i   u
(m income-value   (0.4  0.1  0.3  0.2)) ;수입은 사업,기업체,대학,정부 순
(m difficulty-value (0.1  0.4  0.2  0.3)) ;난이도는 정부,대학,기업체,사업 순
(m freetime-value  (0.2  0.3  0.1  0.4)) ;자유는 대학, 정보, 사업, 기업체순
(m status-value    (0.3  0.2  0.4  0.1)) ;지위는 기업체,사업,정부,대학 순
(m community-value (0.2  0.3  0.2  0.3)) ;동료집단은 정부와 기업체가 유리
(m parents-value   (0.3  0.2  0.3  0.2)) ;부모집단은 사업과 대학에 유리함

:: 동일한 확률부여      b   g   i   u
(m relatives        (0.25 0.25 0.25 0.25) ; 이들은 개인에 따라 다르며
(m future           (0.25 0.25 0.25 0.25) ; 보편적 기준이 없음

:: 중요도에 따른 확률   l   m   h
(m income-weight    (0.2  0.3  0.5)) ; 중요도는 낮음, 중간, 높음으로 정하여
(m difficulty-weight (0.2  0.3  0.5)) ; 20%,30%,50%의 중요도를 정함(평균: 33%)

:: Conditional Probabilities => 조건부 확률은 모든 링크에 부여함

(conds (income (income-value income-weight))
  ((business-best low)      0.55 0.15 0.15 0.15) ;상,중,하 선택에 따라
  ((business-best middle)  0.70 0.10 0.10 0.10) ;확률은 85%, 70%,
  ((business-best high)    0.85 0.05 0.05 0.05) ;55%로 부여함
  ((government-best low)   0.15 0.55 0.15 0.15)
  ((government-best middle) 0.10 0.70 0.10 0.10)
  ((government-best high)  0.05 0.85 0.05 0.05)

```

((industry-best low)	0.15 0.15 0.55 0.15)
((industry-best middle)	0.10 0.10 0.70 0.10)
((industry-best high)	0.05 0.05 0.85 0.05)
((university-best low)	0.15 0.15 0.15 0.55)
((university-best middle)	0.10 0.10 0.10 0.70)
((university-best high)	0.05 0.05 0.05 0.85))

부록 2. 프로그램의 실행결과

- * 직업별 효용등 모든 자료가 저자에 의하여 주관적으로 부여 되었으나 진로 선택에 대한 확률값이 모든 분야가 0.25에 접근하도록 하는 등 프로그램 실험에는 하자가 없음. 실제 사용시에는 전문가 시스템을 개발할 때와 마찬가지로 설문조사, 전문가상담 등으로 자료수집해야 함.

초기의 확률 :	개인사업	정부기관	산업.기업	대학.연구	
(p (U-SELF-UTILITY))	= #	(0.27770376	0.22316751	0.2543034	0.24269235)
(p (SOCIAL-UTILITY))	= #	(0.22273192	0.27726817	0.23265523	0.26001674)
(p (U-LIFE-UTILITY))	= #	(0.24838649	0.24683507	0.24384996	0.24552739)
(p (CAREER-CHOICES))	= #	(0.24696182	0.24641979	0.23599353	0.24866216)

1) 이용자의 중요도에 대한 선택이 영향을 줌

이용자 개인의 수입에 대한 중요도 선택후(여기에서는 높음) 확률:

	개인사업	정부기관	산업.기업	대학.연구	
(observe (INCOME-WEIGHT HIGH))					
(p (U-SELF-UTILITY))	= #	(0.28348792	0.21738337	0.25630748	0.24086276)
(p (SOCIAL-UTILITY))	= #	(0.22273192	0.27726817	0.23265523	0.26001674)
(p (U-LIFE-UTILITY))	= #	(0.24838649	0.24683507	0.24384996	0.24552739)
(p (CAREER-CHOICES))	= #	(0.24851176	0.2449239	0.2366032	0.24825618)

이용자 개인의 업무난이도에 대한 중요도 선택후(여기에서는 높음) 확률:

	개인사업	정부기관	산업.기업	대학.연구	
(observe (DIFFICULTY-WEIGHT MIDDLE))					
(p (U-SELF-UTILITY))	= #	(0.28261855	0.2181735	0.25388378	0.2433824)
(p (SOCIAL-UTILITY))	= #	(0.22273192	0.27726817	0.23265523	0.26001674)
(p (U-LIFE-UTILITY))	= #	(0.248314	0.24689794	0.24365973	0.2457169)
(p (CAREER-CHOICES))	= #	(0.24826212	0.2451493	0.23593698	0.24892935)

이용자 개인의 자유시간에 대한 중요도 선택후(여기에서는 높음) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (FREETIME-WEIGHT HIGH))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)
 (p (SOCIAL-UTILITY)) = #(0.22273192 0.27726817 0.23265523 0.26001674)
 (p (U-LIFE-UTILITY)) = #(0.248314 0.24689794 0.24365973 0.2457169)
 (p (CAREER-CHOICES)) = #(0.24782392 0.245725 0.23456265 0.25036752)

이용자 개인의 지위에 대한 중요도 선택후(여기에서는 높음) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (STATUS-WEIGHT LOW))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)
 (p (SOCIAL-UTILITY)) = #(0.2334739 0.26652616 0.23670089 0.25667557)
 (p (U-LIFE-UTILITY)) = #(0.24973135 0.24586603 0.24428162 0.24557851)
 (p (CAREER-CHOICES)) = #(0.2512664 0.24266326 0.23585355 0.24963687)

이용자 개인의 동료에 대한 중요도 선택후(여기에서는 높음) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (COMMUNITY-WEIGHT HIGH))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)
 (p (SOCIAL-UTILITY)) = #(0.23155735 0.2684707 0.23472835 0.25855076)
 (p (U-LIFE-UTILITY)) = #(0.24965706 0.2459072 0.2442048 0.24561822)
 (p (CAREER-CHOICES)) = #(0.2507147 0.24318905 0.23530106 0.25013855)

이용자 개인의 부모에 대한 중요도 선택후(여기에서는 높음) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (PARENTS-WEIGHT HIGH))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)
 (p (SOCIAL-UTILITY)) = #(0.23346165 0.26653844 0.23647162 0.25666994)
 (p (U-LIFE-UTILITY)) = #(0.25025758 0.24523188 0.2447862 0.24498528)
 (p (CAREER-CHOICES)) = #(0.25139183 0.24246871 0.23589982 0.24945307)

의사결정 자료(현재확률) : 개인사업 정부기관 산업.기업 대학.연구

(0.25) (0.24) (0.24) (0.25)

2) 이용자의 배경이 진로선택에 영향을 줌

이용자 개인의 진로선택에 관한 학문적배경 선택후(여기서는 대학) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (ACADEMY UNIVERSITY-BEST))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)

(p (SOCIAL-UTILITY)) = #(0.23346165 0.26653844 0.23647162 0.25666994)

(p (U-LIFE-UTILITY)) = #(0.18248522 0.17789064 0.18274143 0.44601938)

(p (CAREER-CHOICES)) = #(0.23343097.0.22481243 0.22134984 0.30457032)

이용자 개인의 진로선택에 관한 친척배경 선택후(여기서는 사업) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (RELATIVES BUSINESS-BEST))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)

(p (SOCIAL-UTILITY)) = #(0.23346165 0.26653844 0.23647162 0.25666994)

(p (U-LIFE-UTILITY)) = #(0.23533991 0.1644458 0.16234984 0.42700112)

(p (CAREER-CHOICES)) = #(0.24789593 0.22111619 0.2153519 0.30014876)

이용자 개인의 진로선택에 관한 장래성배경 선택후(여기서는 산업) 확률:

개인사업 정부기관 산업.기업 대학.연구

(observe (ADVANCEMENT INDUSTRY-BEST))

(p (U-SELF-UTILITY)) = #(0.28078762 0.22017244 0.24828881 0.24907944)

(p (SOCIAL-UTILITY)) = #(0.23346165 0.26653844 0.23647162 0.25666994)

(p (U-LIFE-UTILITY)) = #(0.21986143 0.1470367 0.21349302 0.392725)

(p (CAREER-CHOICES)) = #(0.24120693 0.21228243 0.22642459 0.28777933)

의사결정 자료(최종확률) : 개인사업 정부기관 산업.기업 대학.연구

(0.24) (0.21) (0.23) (0.29)