

**안전 소프트웨어의 신뢰도 정량 평가 BBN을 위한
전문가 지식추출 지침**

A Guide on the Elicitation of Expert Knowledge in
Constructing BBN for Quantitative Safety Assessment of
Safety-Critical Software

KAERI

제 출 문

한국원자력연구소장 귀하

본 보고서를 2003 연도 “정지/저출력 및 디지털 계통의 위험도 평가기술 개발” 과제의 기술보고서로 제출합니다.

2003. 8.

부서명 : 종합안전평부

주 저 자 : 엄홍섭

공 저 자 : 강현국

장승철

하재주

요 약 문

본 보고서에서는 전문가의 주관적 판단이나 정성적 자료로부터 왜곡과 편향을 최소화하여 확률적 방법으로 표현된 자료를 추출하는 방법론을 기술하였다. 본 보고서의 주된 활용 목적은 Bayesian Belief Nets(BBN) 기술을 활용하여 안전 소프트웨어의 신뢰도를 정량 평가 하는 것이다. 이 과정에서 필연적으로 전문가의 지식을 추출하는 과정이 포함되는데, 특히 BBN의 노드확률테이블을 작성할 때와 노드에 입력되는 값을 정할 때는 전문가의 주관적 판단이나 정성적 자료를 확률적 표현으로 변환해야 하는 경우가 많고 여러 가지 바이어스나 인적 오류가 도입될 가능성이 높다. 이런 바이어스나 오류들의 발생을 사전에 방지하거나 또는 이미 전문가에 의하여 작성된 자료에 포함된 바이어스와 오류를 제거할 수 있는 방안을 제공하는 것이 본 보고서의 목적이다.

기술보고서의 주요 내용은 다음과 같다.

- o 바이어스/오류의 종류와 내용

전문가의 지식 추출 과정 특히 확률의 추출 작업에서 나타날 수 있는 여러 가지의 바이어스(bias, 편향)와 오류의 내용과 사례

- o 전문가의 지식 추출 절차 가이드

전문가로부터 지식을 추출하는 과정에서 바이어스의 도입을 방지할 수 있는 절차와 기법

- o BBN에 사용된 전문가의 지식(확률 값) 사례

디지털 시스템의 신뢰도/안전성 평가용 BBN 모델에 사용된 전문가의 확률 값 사례

SUMMARY

This report describes the methodology which could elicit probabilistic representation from the experts' knowledge or qualitative data

It is necessary to elicit expert's knowledge while we quantitatively assess the reliability of safety critical software using Bayesian Belief Nets(BBNs). Especially in composing the node probability table and in making out the input data for BBN model, experts' qualitative judgment or qualitative data should be converted into probabilistic representation. This conversion process is vulnerable to bias or error. The purpose of the report is to provide the guideline to avoid the occurrence of this kinds of bias/error or to eliminate them which is included in the existing data prepared by experts.

The contents of the report are:

- o The types and the explanation of bias and error
The types of bias and error which might be occur in the process of eliciting the expert's knowledge.
- o The procedure of expert's judgment elicitation.
The process and techniques to avoid bias and error in eliciting the expert's judgments.
- o The examples of expert's knowledge appeared in the BBNs
The examples of expert's knowledge (probability values) appeared in the BBNs for assessing the safety of digital systems

목 차

제출문	1
요약문	2
목차	4
1. 서론	7
1.1 BBN을 활용한 안전소프트웨어의 신뢰도 정량 평가	7
1.2 전문가의 지식과 판단을 활용한 BBN 모델링	7
1.3 바이어스 방지/발견/제거 기법과 절차	8
1.4 가이드의 구성	8
2. 전문가의 지식 추출 작업에서 발생하는 바이어스와 오류	10
2.1 인지적 바이어스/오류	12
2.1.1 대표성	12
2.1.2 이용의 용이성(회상/상상의 용이성)[5]	18
2.1.3 조정 및 고정(Adjustment and Anchoring)	20
2.1.4 기타 바이어스	22
2.2 동기적 바이어스(Motivational bias)	27
2.2.1 사회적 압력	27
2.2.2 잘못된 해석(Misinterpretation)	27
2.2.3 잘못된 진술(Misrepresentation)	28
2.2.4 희망적 사고(Wishful thinking)	28
3. 전문가의 지식 추출 절차	29
3.1 작업 1 : 질문 영역과 질문의 선정	29
3.1.1 질문 선정에 포함된 작업 단계	29
3.1.2 권고 또는 외부 전문가의 도움이 필요한 단계의 결정	30
3.1.3 선택된 질문을 위한 체크리스트	32
3.1.4 일반적으로 발생하는 문제점들과 그 해결 방안	33
3.2 작업 2 : 질문의 정련 (Refining the questions)	34
3.2.1 질문들의 구조화(structuring)가 필요한 이유	34
3.2.2 질문의 구조화 기법들	35
3.2.3 전문가의 선정이 질문의 정련보다 선행되어야 하는 경우	40
3.2.4 일반적으로 발생하는 문제점들과 그 해결 방안	40
3.3 작업 3 : 전문가(들)의 선정과 동기부여	42
3.3.1 전문가의 답을 구하는 적용 분야	42

3.3.2	전문가의 문제 해결 과정에 대한 자료가 필요한 적용 분야	46
3.3.3	일반적으로 발생하는 문제점과 그들의 해결방안	48
3.4	작업 4 : 추출 구성요소(components, or building blocks)의 선택	50
3.4.1	필요한 구성요소의 결정-체크리스트	50
3.4.2	각 구성요소로부터의 선정	51
3.4.3	일반적으로 발생하는 문제점과 해결 방안	63
3.5	작업 5 : 추출의 설계와 맞춤 (Designing and Tailoring the Elicitation)	64
3.5.1	추출 설계상의 고려 사항	64
3.5.2	일반적으로 발생하는 문제점과 해결 방안	65
3.6	작업 6 : 추출 작업의 연습과 조직 내 교육	67
3.6.1	추출 작업의 연습	67
3.6.2	조직 내부 (사내) 교육	70
3.6.3	일반적으로 발생하는 문제점과 해결 방안	71
3.7	작업 7 : 추출 작업의 지휘/지도/관리 (Conducting the elicitation)	73
3.7.1	일정 관리	73
3.7.2	추출 작업의 set up과 수행(지휘)	75
3.7.3	일반적으로 발생하는 문제점과 해결 방안	79
4.	신뢰도/안전성 평가용 BBN 모델에 사용된 전문가의 확률 값 설정 사례	84
4.1	BBN for M-ADS system and DO-178B Standard	84
4.2	Review of software design document using BBN	106
4.3	기타 BBN의 노드확률테이블	113
4.3.1	외부 개발 계층제어 시스템의 평가 BBN	113
4.3.2	컴퓨터 시스템 명세 문서의 안전성 평가용 BBN	113
5.	요약 및 결론	115
5.1	전문가의 판단에 내재된 바이어스의 형태와 내용	115
5.2	전문가의 지식 추출 절차	115
5.3	기존 BBN 연구에서 나타난 전문가 지식 추출 사례	116
5.4	추후 연구 내용	116
	참고문헌	117

<표 차례>

표 1 인지적 바이어스/오류의 형태	10
표 2 동기적 바이어스/오류의 형태	11
표 3 Cognitive bias	11
표 4 Motivational bias	12
표 5 복잡성에 따른 구조화 기법의 필요성	35
표 6 Sheman Kent의 등급 척도	57

<그림 차례>

그림 1 M-ADS 신뢰도 평가용 상위레벨 BBN 그래프	84
그림 2 소프트웨어 설계 문서 평가 BBN 그래프	106

1. 서론

1.1 BBN을 활용한 안전 소프트웨어의 신뢰도 정량 평가

원전 안전계통에 사용되는 디지털 시스템의 확률론적 안전성평가(Probabilistic Safety Assessment: PSA)를 위해서는 안전 소프트웨어에의 신뢰도에 대한 고려가 필수적이다. 현재 사용되고 있는 소프트웨어의 신뢰도 평가 방법들은 (i)시험 및 신뢰도 성장 모델에 근거한 직접적 평가 방법들, (ii)개발 과정의 품질이나 설계의 다양성 그리고 정형적 방법론에 근거한 간접적 평가 방법들, (iii)신뢰도에 영향을 주는 여러 증거들을 종합한 방법들로 구분되는데 이들 기법들은 모두 단독적으로는 안전 소프트웨어의 평가에 사용하는데 모두 한계성 또는 현실적 제약점을 가지고 있다는 것이 현재 관련 학계의 정설이고 원자력 규제기관의 인허가 입장이다[1]. 이에 따라 PSA와 같은 현실적인 요구에 대응하기 위한 연구가 다양하게 진행 중인데 그 중 하나로 Bayesian Belief Nets(BBN)를 이용하는 방안이 제안되고 있다[16].

BBN 방법은 아직까지는 실용적인 문제의 해결에 적용이 많지 않은 새로운 방법론이지만, 기존의 안전 소프트웨어의 평가 시 핵심적 역할을 하는 전문가의 판단과정을 정형적으로 나타낼 수 있고 또 신뢰도 평가 시 고려되는 정성적 판단 증거를 포함한 다양한 증거들을 한 모델에 결합시켜 그 결과를 정량화 할 수 있다는 특징이 있어 다른 방법론에 비하여 안전 소프트웨어의 정량적 신뢰도 평가에 활용 가능성이 높다[1][17].

예를 들면, 소프트웨어의 시험 결과 값을 단독으로 사용하는 것에 비해서 시험 결과 외에도 요구명세서의 적절성, 완벽성, 개발 과정, 방법론, 시험 과정 그리고 확인 및 검증 활동에 대한 전문가의 판단(평가)을 정량화하여 이를 평가 모델에 반영할 수 있다면 더 바람직한 결과를 얻을 수 있는데 BBN에서는 이것이 가능하다.

한편 BBN 기술이 실용적 방안이 되기 위해서는 아직 해결해야 할 문제들이 있다. 그 중에서 가장 논점이 되어있는 것은 전문가의 주관적, 정성적 지식을 정확하게 정량화(확률 값으로 변환)하는 문제이고 이것을 해결하기 위해 현재 연구가 진행 중이다[18][19].

1.2 전문가의 지식과 판단을 활용한 BBN 모델링

일반적으로 전문가의 판단이 필요한 경우는 다음과 같다.

- o 새롭거나, 희귀하거나, 복잡하거나, 잘 이해되고 있지 않은 현상에 대한 추정치(estimated)가 필요한 경우

- o 미래의 사건을 예상해야 할 경우
- o 기존 자료를 종합하거나 해석해야 할 경우
- o 전문가의 문제 해결 과정 또는 그룹의 의사결정 과정을 알아야 할 경우
- o 특정 분야에서 현재까지 알려진 것, 모르는 것, 그리고 알아야 할 것이 무엇인지를 결정해야 할 경우.

BBN은 일반적으로 불확실성 하에서의 의사 결정에 주로 사용되는 기술인데 대부분의 BBN 적용 문제에서는 위에서 언급된 여러 가지 경우가 포함되고 있으며, 안전 소프트웨어의 정량적 신뢰도 평가 문제도 마찬가지이다. 안전 소프트웨어의 신뢰도를 정량적으로 평가하는 것은 불확실성이 많이 존재하고 또 시험(testing)이나 신뢰도 성장 모델(Reliability growth model)과 같은 기존의 정량적 신뢰도 평가 기법들로는 불충분하므로 기존의 규제 기반은 정성적이고 결정론적(deterministic) 방법에 의존하고 있는데 BBN은 이런 정성적이고 결정론적인 신뢰도 평가 방안들을 확률 형태로 정량화 하는데 사용되는 기법이다. 따라서 BBN 모델링을 할 때 많은 부분을 전문가의 지식과 판단 그리고 문제 해결 방식에 의존한다. BBN으로 모델링 할 때 일반적으로 다음과 같이 3가지 부분에서 전문가의 지식과 판단이 필요하다.

- (a) BBN 그래프(graph or topology) 작성
- (b) 노드확률테이블(node probability table) 작성
- (c) 노드(변수)의 입력 값 작성: 정성적 증거(evidence)값의 정량화(확률 값으로 변환)

1.3 바이어스 방지/발견/제거 기법과 절차

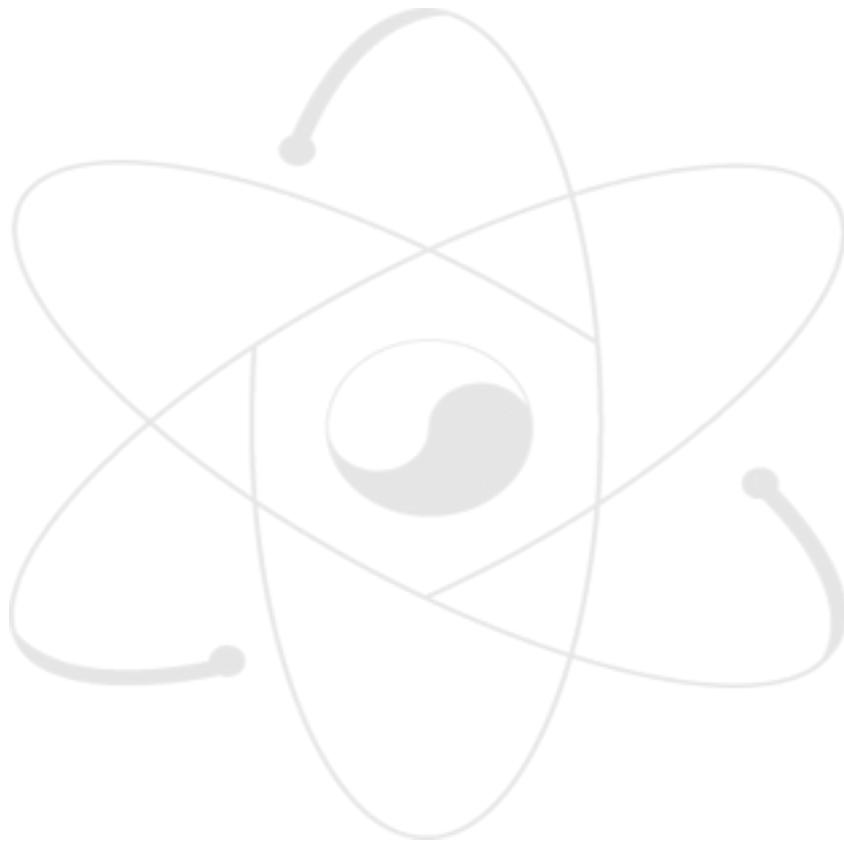
전문가의 판단에 내재된 바이어스를 처리하는 방법들은 드물고 아직 기술적으로 초기단계이다. 판단(judgment)에 대한 연구는 전문가의 바이어스 존재에 대한 주의를 이끌어 냈지만 지식 추출 과정에서 발생하는 바이어스를 처리하는 문제에 대해서는 아직 효과적인 방안이 나오지 않고 있다. 따라서 현재까지는 이 분야의 전문가들은 그들이 일상적으로 접하는 바이어스에 대한 그들 고유의 처리 방법을 개발해서 사용하고 있는 실정이다.

1.4 가이드의 구성

제 2장에서는 전문가의 지식을 추출하는 과정에서 발생할 수 있는 오류와 바이어스에 대한 개요와 여러 분야(지식 공학, 심리학 등)에서 수집된 사례를 수집 정리하였다.

제 3장에서는 지식 추출 과정에서 바이어스의 도입을 미연에 방지할 수 있는 절차와 기법들을 기술하였다.

제 4장에서는 디지털 시스템의 안전성과 신뢰성 평가용 BBN 모델에 사용된 전문가의 확률 값 사례(노드확률테이블, 노드 입력 값)를 수집 정리하였다.



2. 전문가의 지식 추출 작업에서 발생하는 바이어스와 오류

확률 판단에 영향을 주는 공통적 문제점(바이어스와 오류)은 다양하며 이러한 문제점에 대해 충분히 주의를 기울이는 것이 전문가로부터 확률 값을 추출하거나 또는 추출된 확률을 조정하는데 중요하다. 바이어스는 인지적(cognitive)인 것과 동기적(motivational)인 것으로 구분할 수 있는데 이들은 다시 표-1, 표-2와 같이 세부적으로 나눌 수 있다. 표-3과 표-4는 표-1과 표-2의 영문 표기이다.

표 1. 인지적 바이어스/오류의 형태

오류/바이어스 형태	대분류	소분류
인지적 바이어스 및 오류	대표성(전형)	결과의 사전확률에 대한 무감각
		표본 크기에 대한 무감각
		가능성에 대한 오인
		예측성에 대한 무감각
		타당성에 대한 착각
		회귀에 대한 오인
	이용의 용이성	사례의 검색 용이성에 따른 바이어스
		검색 집합의 효율성에 따른 바이어스
		상상의 용이성에 의한 바이어스
		상호 관계의 착각
	조정과 고착	불충분한 조정
		결합사건과 분리사건의 평가에서 나타나는 편중
		주관적인 확률분산의 평가에서 나타나는 고착
	기타	결합 오류
		분산, 공분산, 상관성을 평가할 때의 어려움
		보수적 경향
		과신
		원인과 진단상의 추론에 관련된 오류들
		확실성의 부인

표 2. 동기적 바이어스/오류의 형태

오류/바이어스 형태	대분류	소분류
동기적 바이어스	전문가에 의한 바이어스	사회적 압력
		희망적 사고
	지식 엔지니어 또는 인터뷰 담당자에 의한 바이어스	잘못된 표현
		잘못된 해석

표 3. Cognitive bias

Types of bias	Group	Subgroup
Cognitive biases	Representativeness	Insensitivity to prior probability of outcomes
		Insensitivity to sample size
		Misconceptions of chance
		Insensitivity to predictability
		The illusion of validity
		Misconceptions of regression
	Availability	Biases due to the retrievability of instances
		Biases due to the effectiveness of a search set
		Biases of imaginability
		Illusory correlation
	Adjustment and anchoring	Insufficient adjustment
		Biases in the evaluation of conjunctive and disjunctive events
		Anchoring in the assessment of subjective probability distributions
	Others	Conjunction fallacy
		Difficulties in assessing variance, covariance and correlation
		Conservatism
		Overconfidence
		Fallacies associated with causal and diagnostic reasoning
		Denial of certainty

표 4. Motivational bias

Types of bias	Group	Subgroup
Motivational bias	bias caused by experts	Social pressure
		Wishful thinking
	bias caused by interviewer or knowledge engineer	Misinterpretation
		Misrepresentation

2.1 인지적 바이어스/오류

인지적 바이어스는 대표성(representativeness)을 비롯해서 표-1에 나타난 여러 가지 바이어스의 총칭으로 이것은 인간의 마음 활동에 그 원인이 있다.

2.1.1 대표성

확률을 판단할 때 흔히 일어나는 아래 목록의 오류들을 총칭하는 것이다.

- o 결과의 사전 확률에 대한 무감각(Insensitivity to prior probability of outcomes)
- o 표본 크기에 대한 무감각(Insensitivity to sample size)
- o 가능성에 대한 오인(Misconceptions of chance)
- o 예측성에 대한 무감각(Insensitivity to predictability)
- o 타당성에 대한 착각(The illusion of validity)
- o 회귀에 대한 잘못된 생각(Misconceptions of regression)

2.1.1.1 결과의 사전 확률에 대한 무감각(Insensitivity to prior probability of outcomes) (또는 기본비율(base-rate)의 무시 오류)

주관적 확률 평가에서 가장 일반적이고 공통적으로 나타나며 또한 치명적 오류 중의 하나인데 전문가의 판단을 포함하여 다양하고 넓은 환경에서 발생한다. 이 오류를 인식하는 법을 배우는 가장 좋은 방법은 많은 사례를 연구하는 것이다.

[사례]

다음과 같이 어떤 사람에 대한 설명이 주어졌다.

“(영국에서의 경우) 매우 스포츠맨처럼 보이는 젊은 남자인데 경주용 자동차를 가

지고 있고 매력적인 금발 여자 친구가 있다”

그리고 위의 설명을 참조로 다음과 같은 질문이 주어졌다.

“위의 사람이 일류 전문 축구선수일 가능성과 간호사일 가능성 중 어느 것이 더 높은가?”

위의 질문에 “전문 축구선수”라고 대답했다면 이 오류에 빠진 것이 된다. 즉, 위의 어떤 사람에 대한 설명이 어떤 특정 부류(직업)의 사람에 대한 전형적인 (stereotypical) 이미지에 보다 더 잘 어울렸기 때문에 서로 다른 직업들의 기본(기저)비율(base-rate) 빈도를 무시한 실수가 발생한 것이다. 실제로 영국에는 약 400명 정도의 일류 축구선수가 있고 수 천 명의 남자 간호원이 있다. 따라서 위의 설명 외에 다른 정보가 없는 한 문제의 남자는 간호원일 가능성이 더 높은 것이다.

사람들이 이와 같이 사전 확률을 무시하고 대표성에 의해 확률을 평가한다는 가설에 대하여 Kahneman과 Tversky가 실험을 수행하였는데 그 내용은 다음과 같다[2].

[사례]

(실험)

(a) 피험자들에게 전문가 집단 100명(모두 엔지니어와 변호사들로 구성되었다)에서 임의로 추출된 몇 명의 개성(성격)이 서술된 내용이 주어졌다. 그리고 피험자들에게 주어진 각 서술(개인의 개성을 기술한)에 대해 그 사람이 엔지니어와 변호사 중 어느 부류에 속할 가능성이 많은지 평가하도록 했다.

(b) 실험은 (i)엔지니어 70명과 변호사 30명으로 이루어진 그룹에서 샘플링 한 경우와 (ii)엔지니어 30명과 변호사 70명으로 이루어진 그룹에서 샘플링 한 경우 두 가지로 수행되었다.

(실험 결과)

상식적으로는 어떤 특정 서술 내용(개성을 기술한)이 변호사보다 엔지니어에 속할 확률이 (ii)의 경우보다 (i)의 경우가 더 높게 나와야 되지만, 실험 결과는 두 경우 모두 같은 확률 판단을 했다. 즉, 실험 대상자들은 명백하게 각 서술 내용이 두 직업의 전형적 형태의 대표성 정도에 얼마나 어울리는가에 의거하여 평가를 했고, 각 부류(직업)의 확률에 대해서는 거의 주의를 하지 않거나 무시했던 것이다. 각 개인에 대한 어떠한 서술도 제공하지 않고 단순히 그 사람이 엔지니어에 속할 확률을 질문 받았을 때는 피험자들은 (i)의 경우 0.7, (ii)의 경우 0.3이라는 정확한 답을 제시했다. 그러나 피험자들에게 유용하지 않은 정보를 제공했을 경우 응답은 (i), (ii)

의 경우 모두 0.5였다.

이 실험을 근거로 Kahneman과 Tversky는 다음과 같이 결론을 내렸다.

“특정한 증거가 주어지지 않은 경우에는 사전(prior) 확률이 적절하게 사용되지만, 가치 없는(관련이 없는) 증거가 주어지면 사전 확률은 무시된다.”

[사례]

(문제)

어떤 심장병이 1000명에 한 명 비율로 유행하고 있다. 이 병을 검사하는 방법은 5%의 거짓 양성 반응(false positive rate 5%)을 가지고 있다. 그리고 이 검사 방법은 병이 걸린 사람은 모두 정확하게 진단한다고 한다. 이럴 때 임의로 선택된 사람이 양성 반응을 나타내었는데 그 사람이 실제로 병에 걸렸을 확률은?

이 질문은 Harvard Medical University의 학생과 스태프 60명에게 주어졌다.

응답자의 절반 정도가 95%로 대답

답의 평균값은 56%

정답은 2%(11명)

(설명)

10,000명의 모집단이 있다고 하면 1%인 10명이 실제로 병에 걸려있다. 10,000명 모두를 검사하면 실제로 병에 걸린 10명 외에 500명(5% false positive rate)이 추가로 양성 반응이 나타난다. 따라서 양성 반응이 나타난 510명 중 실제로 병에 걸린 사람은 10명이므로 약 2%가 정답. 95%와 같이 높은 값을 생각한 사람은 병에 걸리는 확률이 매우 낮은 것을 간과했기 때문이다. 상대적으로 거짓 양성 반응 비율은 매우 높다.

(베이스 정리에 의한 설명)

A: event "병에 걸린 사람"

B : event "양성 반응"

구하고자 하는 것은 $p(A|B)$, 그리고 $p(A|B) = 1 - p(\text{NOT } A|B)$

$$p(\text{NOT } A|B) = \frac{p(B|A) * p(A)}{p(B)} = \frac{p(B|A) * p(A)}{p(B|A) * P(A) + p(B|\bar{A}) * p(\bar{A})}$$

$$p(A) = 0.001$$

$$p(\text{NOT } A) = 0.999$$

$$p(B|\text{NOT } A) = 0.05$$

$$p(B|A) = 1$$

그러므로

$$p(\text{NOT } A|B) = \frac{0.05 * 0.999}{0.001 + 0.05 * 0.999} = 0.9804$$

따라서

$$p(A|B) \text{는 약 } 0.02(2\%)$$

[사례] [3]

(상황)

- (a) 유방암의 최종적인 진단 방법은 생체검사법(biopsy. 완전 마취상태에서 가슴 부위의 의심스러운 부분을 제거해내는 외과적 수술 방법)이다.
- (b) 생체검사법은 어려운 수술이므로 유방 x선 사진(mammogram)에 의해서 여부를 판정한다.
- (c) 가슴이 단단해지고 아픔을 호소하는 여성 중에서 악성 암을 가지고 있을 빈도는 1/100이다.
- (d) 유방 x선 사진의 정확도 90%. 암에 걸렸을 때 x선 사진이 양성으로 나타나는 확률이 90%이고 암에 걸리지 않았을 경우 x선 사진이 음성으로 나타나는 확률이 90%라는 것을 의미한다.
- (e) 가슴이 아픈 환자의 x선 사진이 양성으로 나타났을 경우 의사는 x선 사진 결과에 근거한 암의 조건부 확률을 추정해서 생체검사법 시행 여부를 결정해야 한다.

(실험 결과)

인터뷰한 의사의 95%가 위의 상황에서 가슴이 아픈 환자의 x선 사진이 양성으로 나타났을 경우 그 환자가 암에 걸렸을 확률을 75%로 대답했다.

(정답)

약 8%

$$p(C|+) = \frac{p(+|C)p(C)}{p(+)}$$

$$p(C) = 1/100,$$

$$p(+|C) = 0.9,$$

$$p(+) = p(+|C)p(C) + p(+|\text{not } C)p(\text{not } C)$$

$$= (.9)(.01) + (.1)(.99) = 0.108$$

따라서, $p(C|+) = 0.0833$

전문적 기계(시뮬레이터)의 비경험 운전자와 경험 운전자를 대상으로 한 실험에서도 유사한 결과가 나타났다. 더욱이 일반적 지식(상식)에 관련된 사항이나 기술적 지식에 관련된 사항 두 가지 모두의 경우에서도 두 그룹 간에 현저한 차이가 없는 것으로 나타났다[4].

2.1.1.2 표본 크기에 대한 무감각

대표성에 크게 영향을 받는 인간의 사고는 종종 표본의 크기를 무시하게 만들어서 확률을 추출할 때 오류를 일으키게 한다.

[사례]

(실험)

95명의 피험자(subject: 피술자, 실험의 대상자)들에게 다음과 같은 질문을 하였다.

(a) 어떤 도시에 두 개의 병원이 있다. 큰 병원에서는 매일 약 45명의 아기가 태어나고 작은 병원에서는 매일 약 15명의 아기가 태어난다. 일반적으로 태어난 아기의 50%는 남자아이이다. 그러나, 정확한 비율은 매일 매일 변한다(어떤 날은 50%보다 높고 어떤 날은 50%보다 낮다).

(b) 위와 같은 상황에서 두 병원은 1년 동안 남자아이의 탄생 비율이 60%가 넘는 날들을 기록했다. 그러면 어느 병원이 그런 날을 더 많이 기록했을까?

(실험 결과)

95명의 피험자 중 21명은 큰 병원, 21명은 작은 병원이 그런 날이 더 많다고 응답했고, 그리고 53명은 두 병원이 같은 수의 그런 날을 기록했을 것이라고 응답했다.

정답 : 작은 병원에서 그런 날이 더 많이 기록되었을 가능성이 높다.

2.1.1.3 기회와 무작위성에 대한 잘못된 인식(Misperception of chance and randomness)

이것은 부분적 무작위성에 집착되어 나타나는 오류로서 “도박사의 오류” 또는 “작은 수의 법칙”이라고 알려져 있다. 동전던지기의 예를 들면, 동전의 앞면이 몇 번 연속해서 나온 다음에는 뒷면이 나올 가능성이 크다고 생각하는 것이 여기에 속한다. 일반적으로 동전던지기 실험과 같은 무작위 과정에서는 H(동전 앞면)-T(동전 뒷면)-H-T-T-H와 같은 순서가 H-H-H-T-T-T와 같은 순서보다 더 무작위성 개념에 가깝고 따라서 무작위 실험에서 일어날 가능성이 높다고 생각하게 되는 오류이다.

2.1.1.4 예측성에 대한 무감각(Insensitivity to predictability)

사람들은 미래의 주가가 상품의 수요 또는 축구 시합의 결과와 같은 예측을 할 경우 주어진 설명이나 묘사의 좋고 나쁨에 많은 영향을 받는다. 예를 들어 어떤 회사의 설명 자료가 주어지고 그 회사의 미래 수익에 대한 예측을 할 경우, 설명 자료가 훌륭하면 미래 수익을 높게 평가하게 되는 것이다. 그러나 이런 설명의 좋고 나쁨은 그 설명의 신뢰성이나 정확한 예측을 가능하게 해 주는 정도에 의하여 영향을 받지 않는다. 따라서 사람들이 설명의 좋고 나쁨에 의해서만 예측을 하게 된다면 그 예측은 증거의 신뢰성과 예측의 정확성에 무감각한 그런 예측이 될 것이다.

[사례]

피험자들에게 교생들의 수업 실습에 대한 보고서를 제공하고 그 다음 일부 피험자들에게는 그 교생들의 수업의 질에 대한 평가를 하도록 하고, 다른 피험자들에게는 그 교생들의 5년 후의 순위를 예측하도록 했다.

그 결과 두 가지 상황에서의 피험자들의 응답은 동일했다. 말하자면, 미래의 척도에 대한 예상(교생들의 5년 후 성공 정도)이 그 예상의 기초가 된 정보(현재의 실습 품질)에 대한 평가와 동일했다.

2.1.1.5 타당성에 대한 착각(The illusion of validity)

예측 결과와 입력 정보가 서로 잘 어울려서 생기는 정당하지 않는 확신을 타당성에 대한 착각이라고 부를 수 있다. 이러한 착각은 사람들이 예측의 정확성을 제한하는 요인들을 잘 알고 있는 경우에도 여전히 존재한다.

입력 자료 패턴의 내부적 일관성은 이런 입력 자료에 기초한 예측에 있어서 사람들의 신뢰성을 높이는 중요한 결정요인이다. 예를 들어, 사람들은 어떤 학생의 마지막

학년 성적을 예측할 때, 1차 연도 성적이 A와 C로 된 학생보다 모두 B인 학생의 성적 예측에 보다 확신을 가진다.

매우 일관성 있는 패턴은 종종 입력 변수들의 밀접하게 연관되어 있을 때 나타난다. 따라서 사람들은 이런 연관된 입력 변수에 기초한 예측에 대하여 보다 높은 확신을 가진다. 그러나 통계학에 의하면 입력 변수들이 서로 연관되어 있을 때보다 서로 독립적일 때 보다 정확한 예측 결과를 얻을 수 있다. 그래서 입력 자료들 사이의 연관성은 사람들의 확신을 증가시키지만 정확성은 감소시킨다.

2.1.1.6 회귀에 대한 잘못된 생각(Misconceptions of regression)

사람들은 일상생활에서 평균값으로의 회귀에 대한 많은 사례를 접하게 되지만 이런 현상에 대한 정확한 직관을 개발하지 못하고 있는 것이 일반적 상황이다.

[사례]

비행 훈련 교사들이 자신들의 지도 경험을 토론하면서 나온 이야기로, 생도들에게 멋진 착륙이라고 칭찬을 하면 그 다음 착륙은 대체로 나쁘게 되고, 나쁜 착륙이라고 비난을 하면 그 다음 착륙은 많이 좋아진다고 한다. 따라서 교사들은 말에 의한 보상은 소용이 없지만 말에 의한 벌은 효과적이라는 결론을 내렸다.(이것은 심리학의 학설에 위배되는 것이다) 이 결론은 잘못되었으며 그 이유는 평균으로의 회귀 때문이다.

특히 교육에 있어서 회귀 효과를 이해하지 못하면 벌의 효과는 과대평가하게 되고 보상의 효과는 과소평가하게 되는 수가 많다.

2.1.2 이용의 용이성(회상/상상의 용이성)[5]

용이성(Availability)은 확률을 판단할 때 발생하는 사례의 검색 용이성(retrievability of instances), 상관성의 착각(illusory correlation), 탐색 집합의 효율성에 기인한 바이어스들(biases due to the effectiveness of a search set), 그리고 상상력의 바이어스(biases of imaginability)과 같은 범주의 오류들을 총체적으로 기술하는 용어이다.

2.1.2.1 사례의 검색 용이성(retrievability of instances)

어떤 사람이 막 자동차 사고를 목격했다면 그 순간 자동차 사고가 일어나는 사건에 대한 그의 주관적 확률은 일시적으로 증가할 것이다. 그 사건이 인상적이어서 더 기억에 남기 때문이다. Tversky와 Kahneman의 실험[5]에서 유명 인사들의 이름이

나열된 목록을 보여주며 그 목록에 남성이 여성보다 더 많은지를 질문했는데 모두 그렇다고 응답했다. 그 목록의 남성들이 여성들보다 더 잘 알려져 있었을 지도 모르지만, 그 목록에 나열된 남성과 여성의 수는 같은 수였다. 이 경우에서 남성의 이름을 보다 쉽게 상기시키는 친밀성이 판단의 오류를 일으키는 원인이 되었다.

2.1.2.2 탐색 집합의 효율성(Biases due to the effectiveness of a search set)

[사례]

질문: 세 글자 이상으로 이루어진 영어 단어 중에서 “r”을 포함하는 단어를 임의로 골랐을 때 “r”이 첫째 자리에 있을 가능성과 셋째 자리에 있을 가능성을 비교하면 어느 것이 더 높은가?

정답: 세 번째 자리에 있는 단어가 더 많다.

오류: 대부분의 사람들은 첫 번째 자리에 있다고 생각한다.

원인: 단어를 찾을 때 세 번째 자리보다 첫 번째 자리로 찾는 것이 더 쉽기 때문이다.

2.1.2.3 상상력의 바이어스들(Biases of imaginability)

어떤 부류(class)의 크기를 추정하도록 질문 받은 경우(예: 일년 동안 교통사고로 죽은 사람의 수 등) 피험자들은 그 부류의 구체적 사례가 기억으로부터 회상되는 용이성에 따라 추산된 값에 근거하는 경향이 있다. 따라서 주어진 사건이 일어날 빈도는 그 사건의 구체적 사례를 얼마나 쉽게 회상 할 수 있는가에 의해 추산된다.

[사례]

피험자들에게 주어진 질문 :

10명으로 이루어진 그룹에서 2명과 8명으로 구성 할 수 있는 위원회의 수는?

응답결과:

2명으로 이루어진 위원회의 수: 중앙값(median) 70개

8명으로 이루어진 위원회의 수: 중앙값 20개

정답 : 두 가지 경우 모두 45개

바이어스의 원인 : 2명으로 구성된 위원회를 상상하는 것이 8명으로 구성된 위원회를 상상하는 것보다 쉽고 또 10명을 2명으로 구성된 위원회로

분류하는 것이 8명으로 구성된 위원회로 분류하는 것보다 쉽기 때문이다. 따라서 2명으로 구성된 위원회의 수가 피험자들에게 보다 상상하기 용이했고(more available) 이로 인하여 바이어스가 발생한 것이다.

어떤 부류의 빈도를 평가할 때, 그 부류의 구체적 사례가 기억에 저장되어 있지 않고 단지 어떤 규칙에 의해서만 해야 할 경우에는 일반적으로 그 부류의 구체적 사례를 몇 가지 상상한 다음 이들 구체적 사례가 만들어지는 용이성에 근거하여 빈도나 확률을 평가하게 된다. 이 경우 빈도는 구체적 사례를 구축하는데 필요한 상상력에 의해 평가된 것이다.

2.1.2.4 상호관계(상관성)의 착각(Illusory correlation)

두 사건이 동시에 일어날 가능성(co-occurrence)이 그들 두 사건 사이의 연상의 강도(strength)에 의해 판단되는 오류이다. 즉 두 사건이 연상적으로 강하게 연결되어 있을 경우 실제보다 더 자주 일어날 것으로 결론을 내리기 쉽다는 것이다. 주관적 확률 평가자는 종종 두 사건 사이의 상관성에 근거한 결합 확률이나 조건부 확률을 질문 받게 되므로 이런 오류에 대해 주의가 필요하다.

2.1.3 조정 및 고정(Adjustment and Anchoring)

2.1.3.1 불충분한 조정

확률을 평가할 때 어떤 초기 값을 정한 다음 그 수치로부터 조정하거나 수정해서 최종 값을 구하는 일이 종종 일어난다. 그리고 이 초기 값으로부터의 조정은 불충분한 경우가 대부분이다.

[사례]

(실험)

고등학생으로 이루어진 두 개의 그룹에게 다음 식의 계산을 하도록 5초의 시간이 주어졌다.

그룹 A: $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2$ 그룹 B: $2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$

시험 결과, A그룹의 중앙값은 2,250, B그룹의 중앙값은 512가 나왔다. 그리고 정답은 두 식 모두 40,320이다. 이 실험 결과에 대한 Kahneman과 Tversky의 분석은 다음과 같다.

“주어진 짧은 시간 내에 추산을 하기 위해서 먼저 식의 앞부분에 있는 몇 개항의 곱셈을 수행하고 여기에서 나온 값을 기초로 최종 답을 보외추정(extrapolation) 하였는데 그 보외추정이 불충분하게 된 결과이다. 그리고 두 그룹의 답이 다른 것은 초기 계산에서 나온 값(고착 값, anchoring value)의 차이를 반영한 것이다.”

2.1.3.2 결합사건과 분리사건의 평가에서 나타나는 편중

[사례] [6]

(실험 상황)

실험에는 다음의 3가지 사건의 주어졌다.

- (1) 단순 사건(simple event) : 50개의 빨간 돌과 50개의 하얀 돌이 들어있는 가방에서 빨간 돌 한 개를 꺼내는 것
- (2) 결합 사건: 90개의 빨간 돌과 10개의 하얀 돌이 들어있는 가방에서 빨간 돌을 7번 연속적으로 꺼내는 것(with replacement)
- (3) 분리사건: 10개의 빨간 돌과 90개의 하얀 돌이 들어있는 가방에서 7번 연속해서 돌을 꺼낼 때 적어도 한 개의 빨간 돌이 있는 것(with replacement)

그리고 피험자들에게 두 개의 사건 중 하나에 내기를 걸게 하였는데 (1)번과 (2)번, 그리고 (1)번과 (3)번이 각각 제시되었다.

(실험결과)

대부분의 피험자는 (1)의 단순 사건(확률: 0.5)보다는 (2)의 결합 사건(확률: 0.48)을 선택했고 또 (3)의 분리사건(확률: 0.52)보다는 (1)의 단순 사건을 선택했다.

따라서 피험자들은 모두 확률이 적은 쪽에 내기를 걸었다.

도박의 선택과 확률 판단에 대한 연구에 따르면 사람들은 결합 사건의 확률을 과대 평가하는 경향이 있고, 분리 사건의 확률은 과소평가하는 경향이 있다. 이런 바이어스는 고정 효과로 설명된다.

복합 사건의 평가에 나타나는 바이어스는 특히 계획을 세우는 상황에서 현저하다. 신제품의 개발과 같은 어떤 계획의 성공적인 완료는 결합적 특성이 있어서 연속된 사건들이 발생되어야 한다. 이 사건들 하나 하나는 가능성이 높지만 사건들의 수가 많아지면 전체의 성공 확률은 매우 낮아질 수 있다. 결합 사건의 확률을 과대 평가하는 일반적인 경향은 계획의 성공 가능성이나 또는 정해진 시간 내에 과제의 완

료 가능성 평가에 있어서 보장되지 않는 낙관주의로 이끌 수 있다.

반대로 분리 사건의 구조는 위험의 평가에서 많이 나타난다. 원자로나 인간의 몸과 같은 복잡한 시스템은 중요한 구성 부분이 하나만 고장이 나도 기능 장애가 생긴다. 각 구성 부품의 고장 가능성은 매우 낮을지라도 이런 구성 부품의 수가 많으면 전체의 고장 확률은 높을 수 있다. 하지만 고정 성향 때문에 사람들은 복잡한 시스템의 고장 확률을 과소평가하려는 경향이 있다.

그래서 고정 바이어스의 방향은 사건의 구조에서 추론할 수 있는데, 결합 사건의 체인 형태 구조는 과도평가로, 분리 사건의 갈매기 형태 구조는 과소평가로 종종 유도된다.

2.1.3.4 주관적인 확률분포의 평가에서 나타나는 고정

전문가들은 주관적인 확률 분포를 평가하는데 있어서 과도하게 신뢰도 구간을 좁히는 것으로 나타났다. 이 바이어스는 초보자나 숙련자 모두에게서 나타났고 또 적절한 점수 규칙을 도입해도 제거되지 않았는데 이 효과는 고정(anchoring)에 기인한다.

예를 들어, 다우 존스의 평균값의 X_{90} 을 예측할 경우, 사람들은 일반적으로 다우 존스 평균에 대한 최선의 견적을 생각한 다음 이 값을 위쪽으로 확장시켜 조정한다. 만약 이 조정이 충분하지 못하면 X_{90} 값은 적절하게 설정되지 못할 것이다. 이와 유사한 조정이 X_{10} 값을 정할 때도 일어나는데 최선의 견적 값에서 아래쪽으로 확장시키는 조정이 불충분하게 되는 것이다. 결과적으로 X_{10} 과 X_{90} 사이의 신뢰도 구간이 너무 좁게 설정되고 평가된 확률 분포는 너무 타이트하게 되어 버린다.

2.1.4 기타 바이어스

2.1.4.1 결합 오류(Conjunction fallacy)

두 사건 A와 B가 동시에 일어나는 확률 $p(A \text{ and } B)$ 는 각 사건의 확률 $p(A)$ 나 $p(B)$ 보다 클 수 없다. 그러나 두 사건이 결합된 서술이 각 구성 요소 하나에 대한 서술보다 어떤 부류의 특징을 보다 더 잘 나타내는 것처럼 보이는 경우가 종종 있으며, 이로 인하여 확률을 판단할 때 $p(A \text{ and } B)$ 를 $p(A)$ 또는 $p(B)$ 보다 크게 추산하는 경우가 생긴다.

[사례]

- a. 빌은 34살이며 지능은 높지만 상상력이 부족하다. 그리고 충동적이며 일상생활

은 무미건조하다. 학교에서는 수학을 잘 했으며 사회계열이나 인문계열의 과목은 잘 못했다.

b. 위에서 주어진 정보에 근거하면, 다음 문장 중 어느 것이 더 가능성이 높을까?

(i) 빌은 회계사이고 취미로 재즈 연주를 한다.

(ii) 빌은 취미로 재즈 연주를 한다.

통계학을 공부한 사람들을 대상으로 한 실험에서 피험자의 92%가 (i)을 선택하였다.

2.1.4.2 분산, 공분산, 상관성을 평가할 때의 어려움 (Difficulties in assessing variance, covariance and correlation)

일반적으로 사람들이 통계적 분산을 추산할 때 많은 어려움을 겪는 것으로 알려져 있다. 분산을 추정하면서 실제로는 분산계수(표준편차/평균)를 추정하는 것 등이 그 예이다.

[사례]

1.8 ~ 2.1 미터 높이의 나무로 이루어진 숲의 상부를 보면 매우 매끄럽게 보인다. 반면 연필이나 전화기 그리고 기타 잡동사니가 널려있는 책상 위를 보면 무척 울퉁불퉁하고 고르지 못하게 보인다. 그러나 숲의 상부가 책상 위보다 분산(variance)이 크다. 여기에서 사람들에게 직접적으로 그리고 크게 영향을 미치는 것은 관심의 대상이 되는 객체의 높이(평균 높이)에 대비되는 상대적 변동성(분산 계수, coefficient of the variation)이 되어 추정의 오류를 일으키게 된다.

공분산의 경우, 관심의 대상이 되는 두 개의 사건 X와 Y를 생각할 때 X사건의 “발생 함”과 “발생하지 않음”을 고려하고 이들 각각에 대하여 다시 Y사건의 “발생 함”과 “발생하지 않음”을 고려하여야 한다. 즉 사건 X,Y의 유무가 모두 고려된 “2 x 2 테이블”의 도식에서 나올 수 있는 모든 결과를 고려해야 하지만 많은 경우에 사람들은 그들의 판단 과정에서 이 테이블에 있는 4개의 셀(cell) 중 하나 또는 그 이상을 무시하는 것으로 나타났다.

자료 또는 사건의 상관성을 인식하는데 있어서는 상관계수가 0.2-0.3의 경우는 거의 인식하지 못하고, 0.6 이하에서는 상관성을 인식하는데 매우 어려움을 겪었으며 0.6-0.8에서는 과소평가 되었다. 그리고 0.85 이상의 경우에만 일관되게 인식 가능한 것으로 나타났다.

2.1.4.3 보수적 경향(Consevatism)

어떤 정보(또는 사건)에 근거하여 자신의 의견(또는 믿음)을 변경할 때 이들 정보를 처리하는 인지적 보수성에 기인한 오류를 말한다.

[사례]

두 개의 가방이 있다. 한 가방에는 30개의 흰 공과 10개의 검은 공이 있고 다른 가방에는 30개의 검은 공과 10개의 흰 공이 들어있다. 임의로 하나의 가방을 선택하고 이 선택된 가방에서 임의로 한번에 하나씩 다섯 번(모두 5개의 공) 복원 추출(selection with replacement)하였는데 4개의 흰 공과 한 개의 검은 공이 나왔다. 그러면 공을 꺼낸 가방이 흰 공이 많이 들어있는 가방일 확률은?

이 문제에 대한 전형적인 응답은 0.7에서 0.8 사이였고 정답은 0.96이다.

A: 임의변수 “선택된 가방” 따라서 $A=\{a_1, a_2\}$ 이고 a_1 은 흰 공이 많이 들어있는 가방, a_2 는 검은 공이 많이 들어있는 가방이다. 그리고 $p(a_1) = p(a_2) = 0.5$ 이다.

B: 사건 “5번 선택에서 4개의 흰 공과 1개의 검은 공이 나온 사건”

따라서 문제에서 구하고자 하는 것은 $p(a_1|B)$ 이다. 베이스 정리에 의해;

$$p(a_1|B) = \frac{p(B|a_1)*p(a_1)}{p(B|a_1)*p(a_1) + p(B|a_2)*p(a_2)}$$

$$p(B|a_1) = \frac{405}{1024}, \quad p(B|a_2) = \frac{15}{1024}$$

따라서 $p(a_1|B) = 0.964$

베이스 정리는 어떤 증거에 입각하여 의견(믿음)을 수정하는 정형적 규칙이다. 그러나 인간은 일반적으로 베이스 정리에서 제시하는 것보다 더 작게 그들의 확률을 수정한다. 즉 정보의 갱신에 보수적인 것이다.

이러한 보수적 경향의 주된 이유는 자료의 잘못된 결합(misaggregation)이다. 각각의 자료에 대해서는 정확하게 감지하고 또 그것의 특징적 의미(diagnostic meaning)를 잘 알고 있지만 의견(믿음)을 수정할 때 그 자료의 특징적 의미와 다른 자료의 특징적 의미를 적절히 결합하지 못하기 때문이다.

2.1.4.4 과신(Overconfidence)

어떤 사건에 대한 평가 또는 질문에 대한 응답에는 확신의 정도가 수반되어 있다.

예를 들면 다음과 같은 질문이 있을 경우;

“다음 두 도시 중 어느 도시의 인구가 더 많을 까? (1) Hyderabad (2) Islamabad”

(1) 또는 (2)를 응답했을 때 그 사람은 자기의 응답에 대한 확신의 정도를 가지고 있다. 확신의 정도가 50%정도라면 그의 대답은 추측의 정도이고 100%라면 절대적으로 확신하고 있는 것 등이다.

과신에 대한 20 여 년에 걸친 연구결과에서 증명된 것은 거의 모든 경우에 있어서 실험 대상자가 자신의 응답에 대해 가지고 있는 확신의 정도 값은 실제의 정확한 답보다 높은 것으로 나타났다. 즉, 자신의 답에 대해 100%의 확신을 가지고 있는 경우 그것이 정확한 답이 될 상대적 빈도는 80%정도였고, 답에 대해 90%정도의 확신을 가지고 있는 경우에 정확한 답의 상대적 빈도는 75%, 그리고 답에 대해 80%정도의 확신을 가지고 있는 경우에 정확한 답의 상대적 빈도는 65% 등이다.

이러한 과신은 증거를 축적하는 과정에서도 나타난다. 어떤 사건의 평가나 예측을 하기 위해 그에 관련된 증거나 정보를 수집하는 경우, 예측의 정확도가 한계에 다다른 어떤 지점이 자료수집 과정의 한 곳에 존재한다. 그럼에도 불구하고 평가자 또는 예측자가 얻어진 자료로부터 도출하는 결론에 대한 확신의 정도 값은 자료가 점점 더 많이 수집됨에 따라 한계치를 넘어 더 올라가게 된다. 그래서 자료수집 과정의 최종 단계에 이르러서는 대부분의 판단은 과신의 바이어스를 지니게 된다.

2.1.4.5 원인파 진단상의 추론에 관련된 오류들 (Fallacies associated with causal and diagnostic reasoning)

사람들은 일반적으로 주위의 사건들을 결합된(응집된, coherent) 형태로 해석하는 경향이 있으며 원인-결과 관계의 도식에 의해서 사건들을 조직화하므로 서 이런 목표를 달성한다. 조건부 확률 $p(X|D)$ 에서 사건 X와 증거(또는 진단 증상)D는 동일한 정보량을 가진 것으로 취급되지만, 심리적(정신적)으로는 원인이 되는 자료가 같은 크기의 정보량을 가진 다른 자료보다 훨씬 더 크게 판단에 영향을 미친다. 즉, 인과 관계 도식을 야기하는 종류의 자료가 존재하면 다른 종류의 자료는 무시되거나 아주 적게 고려되는 것이다.

o 원인->결과 추론이 결과->원인 추론보다 확신도가 높아진다

[사례]

다음의 예는 원인에서 결과로 연결되는 추론이 결과에서 원인으로 연결되는 추론보다 더 확신도가 높게 되는 경우이다.

다음의 세 가지 사건 중에서 어느 것이 가장 가능성이 높을까?

- (a) 만약 어머니의 눈이 푸르다면 그녀의 딸의 눈도 푸르다.
- (b) 만약 딸의 눈이 푸르다면 그녀의 어머니 눈도 푸르다.
- (c) 위의 두 사건은 같은 크기의 가능성을 가지고 있다.

실험 결과 165명의 피험자 중 75명이 (c)를 선택하였고 69명이 (a)를 선택하였으며 단지 21명만이 (b)를 선택하였는데 이것은 추론의 비대칭성을 보여주고 있다

일반적으로 딸이 그녀의 어머니를 닮았다고 생각하는 것이 그 반대의 경우보다 생각하기 더 쉬우며 또 딸의 속성들을 그녀의 어머니에 기인한 것으로 생각하는 것이 일반적 사고과정이지만 확률의 관점에서는 두 가지 경우가 동일한 가능성을 가지고 있다. 이 실험의 결론은 원인이 되는 자료가 결과 자료의 확률 값에 미치는 영향이 증상자료(diagnostic data)가 원인 자료의 확률 값에 미치는 영향보다 크다는 것이다. 그리고 그 이유는 사람들이 원인으로부터 결과를 추론하는 방식이 결과로부터 원인을 추론하는 방식보다 더 자연적이기 때문이다.

- o 어떤 자료가 원인과 증상의 두 가지 성질을 모두 가지고 있을 때, 이 자료를 가지고 조건부 확률 판단을 하게 되는 경우 자료의 원인적 성질에 더 많은 가중치가 주어진다.
- o 어떤 새로운 증거(또는 자료)에 의해서 추론을 할 때, 이 새로운 증거를 이미 머리 속에 존재하는 원인->결과 모델에 융화시키는 방식이 기존의 모델을 진단 추론 모델(결과->원인 추론)로 바꾸는 것보다 보다 용이하고 사고과정이 자연스럽다. 따라서 어떤 사건들 사이의 설명이 기존 개념을 부분적으로 그리고 최소한으로 바꾸어 가능하다면, 일반적으로 사람들은 그들의 의견을 급격하게 수정하지 않게 된다.

2.1.4.6 확실성의 부인 또는 제어

불확실성은 평가자 자신에게서 기인하는 것이 아니고 평가하고자 하는 대상(사건 또는 상황)이 가진 고유 속성이라고 생각하는데서 기인하는 오류이다. 즉 평가자가 실제로는 제어할 수 없는 상황이나 사건인데도 불구하고 그것들에 대하여 영향을 미칠 수 있다고 생각을 하고 이로 인하여 왜곡된 확률 평가가 이루어지는 것이다.

[사례] [7]

- (a) 두 개의 그룹 A와 B가 구성되었는데 A그룹은 26명 B그룹은 27명이고 모두 뉴욕에서 근무하는 사무직 근로자이다.

(b) 두 그룹 모두에게 \$1 복권을 살 기회를 주었다(모두 53개의 복권이 상자에 들어있고 그 중의 하나가 당첨복권). 나중에 추첨에 의해서 당첨 복권으로 결정되면 \$50의 상금을 받는다.

(c) A그룹의 사람들에게는 자기가 살 복권을 선택할 수 있는 기회를 주었고 B그룹의 사람들에게는 선택 기회를 주지 않고 그냥 나누어주었다.

(d) 추첨 직전에 두 그룹의 사람들에게 그들이 가진 복권을 얼마에 되팔 수 있는지 질문했다.

위의 실험결과, A그룹의 재 매각 금액은 \$8.67(중앙값)이고 B그룹의 재매각 금액은 \$1.96(중앙값)이었다.

2.2 동기적 바이어스(Motivational bias)

동기적 바이어스는 사회적 압력(social pressure), 잘못된 해석(misinterpretation), 잘못된 진술(misrepresentation), 희망적 사고(wishful thinking)를 총칭하며 이것은 전문가의 감정적인 필요와 욕구에 그 원인이 있다.

2.2.1 사회적 압력

사회적 압력은 전문가가 인정받고 싶고 또 긍정적인(전망이 밝은) 쪽에 서고 싶다는 욕망으로 인하여 자신의 생각을 변경하는 것을 말하며 이는 의식적이거나 무의식적일 수 있다. 사회적 압력은 인터뷰 담당자나 다른 전문가와 함께 있는 상황으로부터 생길 수도 있고 다른 사람들의 반응을 마음속에서 평가하는 것으로부터 생길 수도 있다.

사회적 압력이 여러 전문가들과 얼굴을 맞댄 상황에서 생기면 이것을 “집단 사고(group think)”라 칭한다. 집단 사고는 개인이 집단의 판단 또는 그 집단에서 존경받는 누군가의 판단에 맞추기 위해 자신의 생각을 변경할 때 발생한다. 집단 사고는 특히 다음과 같은 경우에 문제가 된다.

- o 그룹의 멤버들이 전에 함께 일했을 경우
- o 그룹 멤버들이 서로 상충된 의견이 나오는 것에 대해 불편함을 느낄 경우
- o 그룹에 지배적인 리더가 있는 경우

2.2.2 잘못된 해석(Misinterpretation)

잘못된 해석은 전문가의 생각이 추출 기법이나 문서화의 결과로 인해서 바뀌는 것이다. 이런 현상은 자주 발생하고 있으나 많은 주의를 끌지 못하고 있었다. 잘못된

해석은 추출 작업이 전문가의 관점이 아니라 인터뷰 담당자의 관점에서 행해질 경우나 응답 모드(response mode)의 결과로 일어난다. 예를 들어, 만약 전문가가 그의 판단을 특정 응답 모드로 적절하게 변환할 수 없다면 잘못된 해석이 일어날 것이다. 일반적으로 각 분야의 전문가들은 확률 분포, 순위, 백분위수의 응답 모드에서 특히 어려움을 겪는 것으로 나타났다.

2.2.3 잘못된 진술(Misrepresentation)

잘못된 진술은 자료를 분석하거나 또는 모델링하는 것으로 인해 전문가의 생각이나 답변이 변경되는 것이다. 전문가의 자료를 모델링하는 사람이 컴퓨터 프로그램의 입력 자료를 만들거나 또는 분석을 위해서 자료의 모습이나 내용에 암묵의 가정을 만드는 일이 종종 있는데 이 경우 misrepresentation이 발생할 수 있다. 예를 들어, 분석가는 전문가의 자료가 정규분포라고 가정할지도 모른다. 이 가정이 정당하지 못하다면 전문가의 자료는 잘못 전달되는 것이다.

2.2.4 희망적 사고(Wishful thinking)

희망적 사고는 전문가의 바람이나 또는 그가 질문 받은 분야의 참여가 그의 응답에 영향을 미치는 것이다. 예를 들어, 사람들은 종종 그들의 주어진 시간 내에 성취할 수 있는 것에 대해서는 희망적인 생각 때문에 과도하게 낙관적으로 판단한다. 희망적 사고의 효과가 가장 큰 경우는 당사자가 개인적으로 연관되어 있거나 또는 그의 응답으로 인하여 어떤 이익을 얻을 수 있는 경우이다. 예를 들면, 한 전문가가 여러 회사의 서비스 상태를 평가하게 되었을 때, 대상 회사들 중 한 회사가 그에게 좋은 평가에 대한 보답을 제공한다면 희망적 사고가 일어날 수 있고 따라서 전문가의 판단은 왜곡될 수 있다.

3. 전문가의 지식 추출 절차

2장에서 기술된 바이어스나 오류들은 전문가의 지식추출 과정에서 흔히 나타나는 것들이나 현재 이들은 완전하게 방지하는 체계적인 기법이나 방법론은 없고, 다만 일부 특정 분야나 케이스에 적용 가능한 기법이나 지침이 얼마간 있는 실정이다. 따라서 어떤 문제 해결을 위해 전문가의 지식추출이 필요한 경우 대부분은 지식추출 전문가의 개인적인 경험이나 know-how에 의존하고 있는데 소프트웨어의 안전성/ 신뢰성 평가를 비롯한 소프트웨어 공학 분야에서의 현황도 마찬가지이다.

본 장에서는 안전 소프트웨어의 신뢰도와 안전성 평가용 BBN을 구축하기 위하여 전문가의 지식을 추출할 때 사용 가능한 체계적인 전문가 지식추출 절차를 기술하였다. 작성된 절차는 NUREG/CR-5424 “Eliciting and Analyzing Expert Judgment”[10]의 내용 중 Part II “Elicitation Procedures”를 기본으로 하였으며 소프트웨어 공학의 특성을 고려하여 작성하였다. 이 절차는 크게 7개의 작업으로 구성되어 있으며 각 작업에는 상세 단계와 단계 별 수행 내용이 들어있다.

3.1 작업 1 : 질문 영역과 질문의 선정(Selection the question Areas and questions)

전문가에게 물어볼 질문을 선정하는 과정은 점차적으로 진화하는 형태를 가지는데 질문 목록이 진화할수록 선정 과정은 명확해진다. 질문을 선정하는 작업에는 다음의 3 단계가 있다.

- o 단계-1: 과제의 목적 또는 목표 정의
(defining the project's purpose or goals)
- o 단계-2: 질문 영역의 선정(selecting the general question areas)
- o 단계-3: 질문 작성/확인(identifying the questions)

3.1.1 질문 선정에 포함된 작업 단계

단계-1 : 과제의 목적 또는 목표 정의(defining the project's purpose or goals)

과제의 목표는 그 과제가 성취해야 할 어떤 것을 의미한다. 대상 전문가는 작업할 과제의 목표에 대하여 명확하게 알지 못하는 경우가 종종 있으며 이로 인하여 세부적 내용을 다룰 때 문제가 생길 여지가 있다. 따라서 이를 피하기 위하여 첫 단계로 과제의 목표를 가능한 명확하게 인지하여야 한다.

단계-2 : 질문 영역의 선정(selecting the general question areas)

질문 영역은 조사의 대상이 되는 특정 이슈들이다. 이들 질문 영역은 (i) 과제의 목

표 (ii)고객의 요구사항 그리고 (iii)그 주제에 대하여 전문가의 판단을 수집할 때 발생하는 실용적 문제들과 같은 정보를 고려해서 개발된다. 일반적으로 처음 만들어진 여러 개의 질문 영역 중에서 최종적으로는 몇 개만이 최종 질문 영역으로 선택된다.

단계-3 : 질문 작성/확인(Identifying the questions)

질문은 질문 영역 내에서 전문가가 대답해야 할 구체적이고 상세한 사항이다. 질문이 적절하게 만들어진 것을 판정하는 방법은 그 질문이 전문가가 대답할 수 있을 정도로 충분히 구체적이고 명확한가를 보는 것이다. 만약 전문가가 만들어진 질문을 적절히 다룰 수 없다면 그 질문은 아마도 질문이 아니라 질문 영역에 가까운 것일 수 있다.

전문가의 지식을 추출 작업을 수행할 때 위의 3단계 중 현재 어느 상태에 있는지를 평가할 필요가 있다. 일반적으로 과제의 목표 또는 질문 분야가 이미 정의된 경우가 많다.

3.1.2 권고 또는 외부 전문가의 도움이 필요한 단계의 결정

질문 영역의 선정, 질문의 확인, 그리고 질문을 정련하는 데는 전문가가 필요하다. 여기에 필요한 전문가가 나중에 질문에 대답할 전문가(자문 전문가)가 아니라면 문제가 없지만 만약 그 질문들을 대답할 전문가(외부 전문가)라면 지식추출 과정 전반에 걸쳐 영향을 미치므로 주의가 필요하다. 나중에 질문에 대답해야 할 전문가를 질문 작성 단계에 활용할 때는 다음의 세 가지 옵션이 가능하다.

(1) 질문 분야의 선정, 질문의 확인, 질문의 정련과 같은 과정에 전혀 참여하지 않고 다만 최종 지식추출 작업 직전에 최소한으로 관여하는 경우

이 경우는 자문 전문가와 지식 추출 작업자가 공동으로 질문 작성에 관련된 작업을 수행하고 모든 작업이 끝난 질문들을 외부 전문가에 보여준 다음 그들의 반응(질문의 의미 해석 등)을 받아 최종적으로 질문들을 정련하는데 활용하는 방안이다.

o 장점:

- 지식추출 담당 작업자가 질문의 개발 과정을 관리할 수 있다.
- 질문들이 최종적으로 완성될 때까지는 외부 전문가를 선정할 필요가 없다.

o 단점

- 외부 전문가들에 대한 동기부여가 미진할 가능성이 있다. 일반적으로 질문의

개발 초기부터 참여하는 것이 그 질문에 응답하는데 더 많은 열성을 가지게 된다.

- 외부 전문가들이 질문 개발 초기부터 참여한 경우에 비하여 질문에 응답할 때 보다 많은 어려움을 겪을 가능성이 높다.
- 질문의 공정성에 대한 의심이 있을 가능성이 있다. 즉, 만들어진 질문들이 외부 전문가들을 특정 방향으로 유도하려고 한다는 인상을 주기 쉽다.

(2) 질문 분야의 선정, 질문의 확인, 질문의 정련과 같은 과정에 부분적으로 관여하는 경우

지식 추출 담당자가 질문의 선정, 확인, 정련을 한 다음 외부 전문가가 그것들을 수정(추가, 삭제, 개정 등)하는 방안이다.

o 장점

- 지식추출 담당자가 질문 개발 시 충족해야 할 기준들을 명시할 수 있다.

o 단점

- 질문의 공정성에 대한 의심이 있을 가능성이 있다. 즉, 만들어진 질문들이 외부 전문가들을 특정 방향으로 유도하려고 한다는 인상을 주기 쉽다.

(3) 질문 분야의 선정, 질문의 확인, 질문의 정련과 같은 과정에 전적으로 참여하는 경우

질문 개발의 초기부터 외부 전문가들이 직접 참여하는 것인데, 이 방안은 질문의 개발에 외부 전문가들을 충분히 활용할 수 있는 경우에 적합하다. 질문의 개발에는 많은 회의와 공동 협의가 필요하므로 외부 전문가는 위치적으로 가깝고 또 같은 조직에 속해 있어야 이것이 가능하다. 다만, 지리적으로나 조직상으로 떨어져 있더라도 정기적으로 공동 작업을 수행 할 수 있는 여건이라면 이 방안의 채택도 가능하다.

이 방안은 또한 외부 전문가의 협력이나 검토가 그 과제의 성공을 위해서 필수적인 상황인 경우에 채택된다. 예를 들면 외부 전문가가 극소수라서 그들의 협력이나 검토가 필수 불가결하게 되는 경우 등이다.

o 장점

- 외부 전문가가 질문 개발 초기부터 적극적으로 참여하는 경우 그 질문들에 대한 응답에 보다 적극성을 보이게 된다.

- 외부 전문가들은 만들어진 질문들을 자신들의 노력의 산물이라고 생각하기 때문에 그 과제의 결과에 대하여 보다 적극적인 관심을 가지게 된다.

o 단점

- 지식추출 담당자가 질문의 개발에 관하여 완전하고 직접적인 관리를 할 수 없다.
- 그룹으로 구성된 외부 전문가들과 공동으로 작업하게 되면 질문들을 개발하는 데 있어서 특별한 기술들을 필요로 하게 된다.
- 다른 2개의 옵션에 비하여 이 방안은 보다 어려운 계획을 필요로 한다.

3.1.3 선택된 질문을 위한 체크리스트

3.1.1에서부터 기술된 단계를 거쳐 질문들이 만들어진 다음에는 그 질문들의 적절성에 대한 평가가 필요하다. 그 목적을 위한 체크리스트는 다음과 같다.

(1) 질문들은 과제의 목표를 충족하는데 필요한 자료를 제공할 것인가?

종종 만들어진 질문들은 과제에서 요구하는 내용을 충분히 도출하지 못하게 되는 경우가 발생한다. 이를 방지하기 위해서는 과제가 요구하는 전문가 판단의 형태, 응답, 문제 해결, 기타 필요한 자료를 충분히 검토해야 한다.

(2) 질문들은 과제의 범위 내에서 가능한 것들인가?

여기에서 특히 고려해야 할 것으로는 시간과 예산 그리고 기타 필요한 지원 조건 여부이다.

o 시간

과제에서 정해진 시간에 적합한 질문의 개수가 고려되어야 한다. 어느 정도 상세한 응답을 요구하는지, 또는 단순히 응답만 요구하는 지 아니면 응답에 더하여 추가 자료를 보태어야 하는지 등을 적절히 고려해야 한다.

o 예산

자문 전문가와 외부 전문가를 사용할 때 필요한 예산이 고려되어야 한다.

o 기타 지원 조건 여부

전문가의 가용성에 대한 검토가 필요하다. 즉, 필요한 전문가가 위치적으로 근접한 위치에 있는가, 또는 떨어져 있을 경우 정기적인 공동 작업이 가능한가 등이다.

(3) 전문가의 판단 자료가 큰 어려움이 없이 얻어질 수 있는가?

질문의 응답에 필요한 자료가 공개 자료인가 아니면 특정 기관이나 개인의 등급(기밀)자료인가 등에 대한 검토가 필요하다. 예를 들면, 질문들이 특정 기관에서 외부 노출을 꺼리는 자료에 관련된 응답을 요구하는 경우 지식추출 작업자는 이에 대한 충분한 방안을 마련해야 한다.

3.1.4 일반적으로 발생하는 문제점들과 그 해결 방안

(1) 과제의 목표와 수집되어야 할 정보 그리고 질문 분야가 명확하게 주어지지 않았을 경우

과제의 초기에는 그 목표나 수집되어야 할 정보의 내용 그리고 질문 분야가 구체적이고 상세하지 않은 경우가 많이 있다. 이 경우에는 관련자들과의 토의를 통해 어느 부분에서 문제가 불분명해지는 지 확인해야 한다. 그리고 작업을 계속하기 위해 필요한 정보를 관련자들로부터 추출해야 한다. 이 경우 인터뷰 방법이 자주 사용된다.

(2) 질문 분야로부터 도출된 질문들이 여전히 애매하고 범위가 넓은 경우

질문이 애매하고 광범위하게 만들어졌다는 징후는 자문 전문가나 외부 전문가가 질문들을 보다 상세하게 나누어야 된다고나 또는 질문에 응답하기 위하여 추가적인 정보를 요구하는 경우이다.

이를 해결하기 위해서는 어떤 전문가에게 그 질문에 응답하도록 요청한 다음 그 전문가가 그 질문을 해결하기 위해 질문을 분해하는 방법을 참고하여 보다 세분화된 질문을 작성한다. 또 하나의 방안은 전문가에게 어떻게 하면 그 질문에 잘 응답할 수 있는 지 자문을 받는 것이다.

(3) 허용된 시간에 비해 너무 많은 질문이 선택된 경우

허용된 시간 내에 대답할 수 없을 정도로 많은 개수의 질문이 선정되는 것은 종종 일어나는 어려움이다. 이런 어려움의 초기 징후는 전문가들이 사람들이 질문의 개수와 가능한 시간의 양에 대해 자발적으로 토론하려고 하지 않는 것이다. 그리고 지식 추출에 경험이 있는 사람들일지라도 특정 질문 개수에 응답하는데 필요한 시간을 과소 추정하기 쉽다.

이를 해결하기 위해서는 자문 전문가에게 현재 주어진 추출 과정과 응답 자료의 상세 정도가 고려된 상황에서 작업을 완료하기 위해서는 어느 정도의 시간이 필요할 것인지 추정 해보도록 하는 것이다. 만약 이렇게 만들어진 추정치에 의해서 질문의 개수가 너무 많다는 것이 나타난다면 다음과 같은 방법을 고려해야 한다.

- 각 질문 분야별로 질문의 수를 줄임
- 질문 분야의 수를 줄임
- 보다 적은 전문가를 선정
- 추출 방법을 보다 간단하고 빠른 것으로 수정
- 보다 덜 상세한 응답 자료를 요구
- 가능하다면 계획상의 시간을 연장

일반적으로, 개별 전문가로부터 지식을 추출하는 것이 전문가 그룹으로 지식을 추출하는 것보다 시간이 많이 소요된다. 왜냐하면, 상세한 자료를 필요로 할 경우 개별 인터뷰 등이 필요하기 때문이다. 또 문제 해결에 대한 자료를 구하는 것이 단순히 응답을 구하는 것보다 시간이 많이 걸리고 문제 해결의 양이나 상세 정도가 깊을수록 많은 시간을 요한다. 예를 들면, 전문가의 일반적 해석을 얻는데 필요한 시간은 그 해석에 관련된 정의, 추론, 가정, 참고사항, 계산 등을 포함한 자료를 얻는데 필요한 시간보다 훨씬 적다.

3.2 작업 2 : 질문의 정련 (Refining the questions)

단계-1에서 만들어진 질문들을 정련하는 목적은 인간의 인식 한계를 고려하고 또 전문가가 쉽게 이해하고 처리할 수 있는 질문들로 만드는 것이다. 전문가의 인식에 부정적으로 영향을 미치는 요인들을 최소화 하여 보다 높은 품질의 전문가 판단을 이끌어 낼 수 있다. 이러한 작업에는 (i) 질문을 이해하는데 필요한 정보를 제시하는 방법 (ii) 정보를 순서화 (ordering)하는 방법 (iii) 질문들을 보다 쉽게 이해할 수 있도록 나누는 방법이 있다.

3.2.1 질문들의 구조화(structuring)가 필요한 이유

질문의 정련은 구조화를 통해 이루어진다. 질문들을 구조화한다는 것은 조직적이고 제어된 방법으로 질문하는 것인데 이의 목표는 양질의 자료를 얻기 위한 것이다. 질문들을 구조화하는 방법으로는 규칙적인 방법으로 정보를 제공하고 보다 쉽게 대답할 수 있는 부분으로 나누고 그림이나 수학적 방법으로 이들을 표현하고 키워드를 정의하는 것 등이 있다.

질문의 구조화를 통해서 다음과 같은 이점이 생긴다.

- o 전문가의 주의를 그가 제공해야 하는 것에 집중시킨다.
- o 문제를 풀어야 할 인식적 부담을 경감시킨다.
- o 질문의 범위를 명확히 하여 전문가가 그것을 다르게 해석하지 않도록 한다.

o 전문가가 보다 쉽게 받아들일 수 있게 만든다.

어떤 구조화 기술이 필요하고 어느 정도 필요한지는 질문의 성질에 따라 달라지는데 특히 질문의 복잡성에 의해 결정된다. 여기서 복잡성이란 질문을 풀기 위해 요구되는 정보의 양과 질문에 대한 답을 검증하는 방안이 있는지 여부를 의미한다. 아래의 표는 질문의 복잡성에 따른 구조화 기법의 필요성 정도를 요약한 것이다.

표 5. 복잡성에 따른 구조화 기법의 필요성

질문의 복잡도	질문의 정보 양	질문의 분해	질문의 표현	질문의 어법
단순	낮음	낮음	낮음	높음
복잡	높음	높음	높음	높음

3.2.2 질문의 구조화 기법들

3.2.2.1 질문에 관련된 정보의 제시(표현, presentation)

질문을 정교하게 만들수록 그에 대한 더 많은 정보가 필요하고 또 그런 정보의 제시를 계획하는 데는 많은 시간을 요하게 된다. 정보의 제시는 세 가지 관점으로 나누어진다.

- (i) 정보의 형태
- (ii) 정보 제시의 최적 순서 결정
- (iii) 과제 관련자와 전문가의 역할

(1) 정보의 형태

필요한 질문 정보의 형태에는 배경정보, 가정, 그리고 정의(definition)가 있다.

o 배경 정보

전문가들이 질문에 응답하기 위해서는 그 질문에 관련된 배경 정보가 필요한 경우가 대부분이다. 예를 들면, 전문가가 어떤 실험 결과를 예측하는 질문에 응답하기 위해서는 그 실험의 과정과 초기조건, 경계 조건 등이 필요하다.

o 가정

전문가에게 질문의 응답을 요구할 때에는 그 질문에 내포된 가정을 명확히 하는 것이 필요하다. 특히 그 질문이 복잡한 것일수록 가정에 대한 정보 제공이 중요하다. 가정에 대한 충분한 정보가 제공되지 않는 경우 전문가는 종종 자신들이 생각한 가정에 근거하여 응답을 하게 되고 이것은 종종 질문의 의미를 전혀 다르게 만들기도

한다.

o 정의

질문에 사용된 많은 단어나 구절들은 그 질문의 작성자와 응답자 사이에서 서로 다르게 해석되는 경우가 종종 발생한다. 이런 가능성이 있는 단어나 구절들에 대하여 충분한 고려가 있어야 한다.

(2) 정보 제시의 최적 순서 결정

위에서 언급된 정보들을 전문가에게 제시할 때 그 순서가 중요하다. 정보의 제시 순서를 정하는 하나의 방법은 그 정보의 논리적 흐름을 고려하는 것이다. 그리고 그 흐름을 검사하는 방법 중 하나는 자문 전문가에게 그 질문을 풀게 하고 그 과정을 따라가는 것이다. 만약 전문가가 그 문제를 해결하는 도중의 어떤 지점에서 제시되지 않은 다른 정보를 요구하게 되면 그 곳에 필요한 정보를 추가하는 방식이다.

정보의 제시 순서 결정시 고려해야 할 또 다른 점은 인간의 어떻게 정보를 이해하고 기억해 내는가 하는 점이다. 일반적으로 사람들은 새로운 정보를 접하면 그 정보들이 그들 자신의 사전 지식이나 정신 모델의 영역에 속할 때 보다 용이하게 수용하는 것으로 알려져 있다.

문제해결 과정이 복잡하고 동적인 경우에 전문가들은 그 질문에 관련된 정보를 잊기 쉽다. 이를 해결하기 위해서 제안된 방법 중 하나는 질문의 해석에 중요한 정보는 그 질문 속에 포함시키는 것이다. 예를 들면, “x가 일어난 조건에서 y가 일어날 확률은 얼마인가?” 하는 것 등이다. 만약 가정이나 정의 그리고 다른 정보가 복잡하고 양이 많아 질문의 일부분으로 포함시킬 수 없는 경우에는 질문의 바로 전에 이들을 배치하는 것이 필요하다.

사람들이 정보를 이해하고 수용하는 방법에 대한 위의 여러 가지 점을 고려해 보면 정보의 제시 순서는 일반적으로 계층구조를 가지게 된다. 즉 일반적인 정보에서 구체적이고 상세한 정보의 순서가 된다거나 또는 그 반대의 순서가 된다.

3.2.2.2 질문의 분해

질문의 구조화를 위한 또 다른 전형적인 방법은 질문의 분해이다. 즉 정보 처리의 부담을 줄이고 또 정확성을 높이기 위해서 질문들을 나누는 것인데 그 예는 다음과 같으며 분해 된 질문에 응답한 사람들이 보다 정확한 판단을 내렸다.

(질문 분해의 예)

분해 되지 않은 질문: 1970년도에 미국에 거주하고 있는 가구의 수는?

분해 된 질문: 1970년도의 미국 인구의 수는? 그리고 한 가구 당 평균 가족의 수는?

질문 분해 시 고려 사항들은 다음과 같다.

- (a) 질문 분해 시 가장 먼저 고려해야 할 것은 질문의 복잡성 정도이다. 복잡한 내용의 질문일 경우에는 분해가 거의 필수적이며 위해도 분석(risk analysis) 등이 그 대표적 예이다.
- (b) 구해야 할 자료의 상세 정도도 질문의 분해 시 고려해야 할 사항 중의 하나이다. 상세한 레벨의 자료를 구하고자 한다면 그에 상응하는 정도로 질문을 분해해야 한다. 예를 들면, 문제 해결의 과정에 대한 자료를 구하고자 한다면 그에 대한 질문은 분석에 필요한 자료를 구하는 질문보다 상세해야 한다. 일반적으로 이 경우 상세 정도를 한 레벨 더 높여야 한다.
- (c) 질문의 분해 작업을 계획할 때 외부 전문가의 참여 여부도 하나의 고려 사항이다. 외부 전문가가 이 작업에 어느 시점에서 어느 정도까지 참여할 지를 고려해야 하는데 만약 참여하게 된다면 가능한 이른 시점에 참여하는 것이 바람직하다.

3.2.2.3 질문의 표현(Representation)

질문의 분해가 완료되면 그 상세 정도에 관계없이 그들을 표현하는 것이 효과적이다. 여기서의 표현은 질문들에 내포된 여러 요인들이나 그들 상호간의 관계를 보여주는 다이어그램이나 수식들을 말한다. 예를 들면, 원자로의 PSA에서 사고 순서들이 다이어그램(사건 수목, 고장수목 등)으로 나타내지고 그 각각에 대한 결과가 표시되는 것 등이다.

질문의 표현은 다음의 목적을 위해 사용된다.

- o 외부 전문가의 판단 가이드 제공
- o 전문가가 해답에 도달한 과정을 문서화
- o 추후 유사한 문제의 해결을 위한 가이드 제공
- o 전문가의 판단이 어떤 방식으로 진행되고 분석되었는지에 대한 가이드 제공

질문의 표현 시 고려해야 할 사항들은 질문의 분해 시 고려 사항들과 유사하다. 가

장 중요한 점은 표현의 구성이 질문의 분해 내용, 분석 계획, 그리고 전문가가 문제를 해결하는 방법과 호환성이 있어야 한다는 점이다. 만약 기존의 표현 기법들이 적합하지 않다면 해당 경우에 맞는 특수한 기법을 만들 필요가 있다. 자주 사용되는 몇 개의 표현 기법들에는 다음과 같은 것들이 있다.

(a) 사고 순서 표현 기법

PSA에서 자주 사용되는 기법으로 원인과 그에 대응하는 결과 관계에 따라 표현하는 기법이다. 이 기법은 원자력분야의 위해도/신뢰도 분석에 많이 사용되며 물리적 현상을 모델링하는 일반적 기법으로 인정되고 있다. 이 기법에 대한 상세 내용은 “PRA Procedure Guide NUREC/CR-2300, U.S. NRC 1983”에 기술되어 있다.

(b) 시간적 연결 기법

이 기법은 표현하고자 하는 내용의 구성요소들을 시간적으로 연결하는 것이며, 각 마일스톤이 정해진 계획에 따라 반드시 완수되어야 하는 기술적 프로그램의 관리를 위해 종종 사용된다. 이 표현 기법은 각 활동이나 마일스톤이 어떻게 병행하여 달성되어야 하는지 그리고 어떤 순서로 달성되어야 하는지를 보여준다. PERT가 이런 기법의 한 예이다.

(c) 대체 안 평가 기법

이 기법은 표현 내용의 구성요소들을 대안으로 보고 특정 속성의 관점에서 그들을 평가하는 기법이다. Saaty의 Analytic Hierarchy Process(AHP)는 여기에 속하는 가장 보편적 방법이다. AHP는 의사결정자나 전문가가 여러 개의 대안이나 제품으로부터 합의된 기준에 가장 적합한 한 가지 안이나 제품을 선택하는데 많이 사용된다.

(d) 특수 기법

해당 과제만을 위하여 특별하게 고안된 기법으로 기존의 다른 기법들이 해당 과제에 적합하지 않을 경우 사용된다. 따라서 이런 기법들은 한 가지 경우에는 적합하지만 다른 경우에 적용할 수 없는 경우가 대부분이다.

3.2.2.4 질문의 어법(phrasing)

질문의 구조화를 위한 또 하나의 요소는 질문의 어법(phrase, or wording)으로 이것은 질문의 어법과 전문가의 응답 모드를 말한다. 신중한 질문 어법은 전문가가 문제를 제대로 이해하는 것을 도와주고 또 편견에 빠지거나 부당하게 영향을 받는 것

을 최소화한다. 이에 대한 한 가지 예는 다음과 같다.

(질문 어법의 예-1)

질문A: X사건이 Y번 일어날 것이라고 생각합니까?

질문B: X사건이 Y번 일어날 것이라고 생각합니까, 또는 안 일어날 것으로 생각합니까?

여기서 질문B가 보다 균형 있는 질문이다. 왜냐하면 이 질문이 모든 가능성에 대하여 언급하고 있기 때문이다.

(질문 어법의 예-2)

전문가에게 두 개의 곡선(하나는 컴퓨터 코드로 작성된 것이고 다른 하나는 실험에 의하여 작성된 것)을 제시하고 그 곡선에 변형(diverge)이 있는 곳을 확인하는 실험이다.

질문A: 곡선이 변형되었다고 생각하는 곳에 표시를 하시오.

질문B: 만약 곡선에 변형이 있다고 생각하면 그곳에 표시를 하시오.

여기서 질문 A는 전문가들에게 곡선에 변형이 확실히 있다는 인상을 깊게 주는 어법이기 때문에 잘못된 것이며 따라서 질문 B가 보다 적절한 어법을 가지고 있다.

페인[8]에 의하면 모든 질문의 어법은 바이어스를 가지고 있으며 현실적으로 가능한 최선의 방법은 동일한 정도의 바이어스를 가지게 하는 것이라고 한다.

불명확한 질문 어법으로부터 발생하는 또 다른 문제는 전문가들이 질문을 다르게 해석하여 전혀 다른 응답을 하는 것이다. 위의 예-1의 실험에서 전문가들이 변형(diverge)에 대하여 각기 다른 해석을 하였다.

질문의 명쾌함에 관련된 또 하나의 요인은 질문의 길이이다. 페인(Payne)에 의하면 한 문장에 포함된 단어의 수가 25개를 넘으면 이해력이 감소하는 것으로 나타났다. 따라서 문장은 가능하면 짧게 하는 것이 효과적이다.

o 질문의 어법에서 고려할 사항

질문의 어법에서 고려할 사항 중 가장 중요한 것은 명쾌성과 바이어스이다. 명쾌성은 지식추출 담당자와 자문 전문가의 검토를 통해 향상시킬 수 있다. 반면, 질문의 어법에서 바이어스를 제거하기 위한 손쉬운 방법은 없다. 가장 좋은 전략은 이 문제에 충분한 주의를 기울이고 질문 구절을 세밀히 검사하는 것이다.

질문의 어법에 있어서 또 하나의 고려 사항은 복수의 전문가들에게 공통적으로 이

해될 수 있는 것을 작성하는 일이다. 어법(분해나 표현도 마찬가지이다)이 단지 몇 명의 자문 전문가에 의하여 검토되고 시험되었다면 그것은 왜곡된 어법을 포함할 가능성이 많다. 그리고 단지 한 사람의 자문 전문가만이 이 작업에 활용되었다면 어법에 왜곡이 있을 가능성은 훨씬 더 높아진다.

3.2.3 전문가의 선정이 질문의 정련보다 선행되어야 하는 경우

전문가의 선정과 그들의 동기부여가 질문의 정련 단계보다 선행되어야 하는 경우는 다음의 세 가지이다.

- (i) 과제의 목적이 전문가의 문제 해결과정을 파악하기 위한 것이거나 또는 추후의 의사결정에 대한 가이드로 제공하기 위한 경우
- (ii) 전문가가 질문을 수락하지 않을 징조가 있을 경우
- (iii) 외부 검토자가 질문의 선정과 어법에 바이어스가 있을지도 모른다는 우려를 표명한 경우

만약 위의 경우가 발생하면 외부 전문가를 질문의 정련 작업에 가능한 초기부터 참여시키는 것이 필수적이다. 그렇지 않을 경우 지식 추출 과정에서 심각한 문제가 발생한다.

3.2.4 일반적으로 발생하는 문제점들과 그 해결 방안

(1) 질문을 정련 할 때 외부 전문가로부터의 정보가 불충분 한 경우

정보의 불충분에 대한 징후는 질문들을 검토한 자문 전문가의 반응에서 발견될 수 있다. 그들이 질문에 응답하는데 혼란을 느끼거나 또는 현재의 질문 형식으로는 적절한 응답을 할 수 없는 것이 그 징후이다.

질문을 검토하는데 자문 전문가가 활용되지 않았을 경우, 이런 문제점은 외부 전문가가 그 질문을 처음 접했을 때 나타나게 되며, 그러면 외부 전문가는 보다 많은 정보를 요구하거나 또는 질문의 분해, 표현, 어법에 대하여 비판을 제기할 것이고 심지어는 질문의 응답을 거부할 가능성도 있다. 어떠한 경우든 이런 문제는 과제의 진행에 심각한 영향을 주어 신뢰를 잃게 되거나 또는 일정에 맞추어 작업을 끝내지 못하는 사태가 된다.

이런 사태가 발생하는 경우는 종종 너무 촉박한 일정에 기인한다. 지식추출 담당자는 일정에 맞추기 위해 전문가의 자문에 소요되는 시간을 줄이거나 또는 자신이 직접 질문의 구조화 작업을 수행해버리기 때문이다.

이러한 심각한 상황을 회피하는 가장 좋은 방법은 외부 전문가를 질문의 정련 작업 초기에 활용하는 것이다.

특히 아래의 세 경우는 반드시 그렇게 해야 한다.

- (i) 과제의 목적이 전문가의 문제 해결과정을 파악하기 위한 것이거나 또는 추후의 의사결정에 대한 가이드로 제공하기 위한 경우
- (ii) 전문가가 질문을 수락하지 않을 징조가 있을 경우
- (iii) 외부 검토자가 질문의 선정과 어법에 바이어스가 있을지도 모른다는 우려를 표명한 경우

(2) 질문의 분해 정도가 너무 복잡하고 상세하게 된 경우

이 경우는 질문의 분해와 표현이 과제가 필요로 하는 내용에 비해 과도하게 상세하게 된 경우이다. 필요한 정도보다 한 레벨 더 상세하게 자료를 모으는 것이 일반적이지만 이 경우는 그것을 초과했을 때 일어난다.

질문의 분해 작업에 경험이 없는 사람들은 완전하게 작업을 하려는 경향이 있고 또 질문들을 대답하기 쉽게 만들려고 한다. 이것들은 질문의 분해 목적이며 긍정적 측면이 있지만 과도할 경우 부작용이 발생하게 된다.

이를 해결하기 위해서는 작업자는 한두 번 정도 과도한 질문 분해 작업을 직접 해 볼 필요가 있다. 그로부터 자신이나 다른 사람의 이런 과도한 분해 경향을 인식하는 것이다.

만약 다른 사람들(예를 들면 전문가들)이 이런 경향을 보인다면 이에 대한 두 가지 대처 방법이 있다.

- (a) 그들로 하여금 작업을 그대로 계속하게 해서 그들 스스로 과도한 질문 분해에 따르는 실제적 문제점을 깨닫게 한다.
- (b) 민감도 분석 등을 통하여 어느 시점에 그들이 과도한 질문 분해를 하였는지 보여준다.

첫 번째 방법이 작업자의 전문성을 키우는 좋은 방안이나 시간이 많이 소요된다는 단점이 있다.

(3) 질문이 잘못 정의되었거나 또는 서로 다르게 해석될 가능성이 있는 경우

이것은 위의 경우와는 반대인데 촉박한 과제 일정 등에 기인하여 질문의 구조화 작업이 제대로 되지 못했을 때나 또는 인간의 인지적 한계 때문에 발생한다.

우리가 명쾌하다고 생각하는 질문들도 사실은 그렇게 명확하지 않은 경우가 많다. 왜냐하면 사람들은 자신도 모르는 사이에 문장에는 없는 가정과 정의를 하면서 읽기 때문이다. 만약 자문 전문가가 질문들에 대하여 파일럿 테스트를 하게 된다면 그들의 응답 상태로부터 이 경우에 대한 징후를 발견할 수 있다. 예를 들면 자문 전문가가 추가로 정보를 요구하거나 또는 가정을 추가하므로써 질문의 껍을 채우는 것 등이다.

이러한 상황을 피하기 위해서는 자문 전문가로부터 입력을 구해야 한다. 만약 최악의 경우가 발생하여 제대로 구조화되지 못한 질문들이 외부 전문가에게 주어지고 그로부터 지식추출 작업이 수행된다면 다음의 두 가지 대처 방안이 가능하다.

- (a) 전문가들을 모두 모아 질문의 정련(정제) 작업을 새로 하는 것이다. 그리고 새로 정련(정제)이 된 질문들을 가지고 작업을 진행시킨다. 가능하다면 이것이 최선의 방법인데, 왜냐하면 모든 전문가들은 동일한 질문들에 대하여 대답을 하게 되고 그들의 응답들은 합리적으로 합쳐질 수 있기 때문이다.
- (b) 각각의 전문가들이 질문에 응답하는데 필요한 가정이나 정의를 별도로 만드는 방안이다.

3.3 작업 3 : 전문가(들)의 선정과 동기부여 (Selecting and motivation the experts)

전문가로부터 지식을 추출할 때 두 가지 형태가 있다. 첫 번째 형태는 전문가로부터 답을 구하는 것이고 두 번째 형태는 전문가의 문제 해결 과정에 대한 자료를 구하는 것이다. 이들 두 가지 형태는 서로 다르기 때문에 각기 서로 다른 접근방법을 선택해야 한다.

예를 들면, 전문가의 답을 구하는 경우에는 일반적으로 4명에서 40명 정도의 전문가로부터 정량적 형태의 답을 구하고 전문가의 문제 해결 과정에 관한 자료를 구하는 경우에는 몇 명의 전문가에게 집중적으로 초점을 맞추는 방법을 사용하는 것이다.

따라서 첫 번째 형태의 경우에는 전문가의 다양성과 그들의 판단을 정량화 할 수 있는 능력 등을 고려해서 전문가를 선택하는 것이 좋고 두 번째 형태의 경우에는 전문가의 문제해결 과정을 추출하는 작업에 자발적으로 시간과 노력을 제공할 의사가 있는지 여부를 기준으로 선택하는 것이 좋다.

3.3.1 전문가의 답을 구하는 적용 분야

대부분의 응용분야, 특히 위해도/신뢰도 및 의사결정 분석 분야에서는 전문가의 해

법을 구하는 것이 주목적이고 전문가의 해법은 주로 확률과 같은 정량적 형태로 요구된다. 따라서 전문가는 해당 분야에 대한 전문 지식뿐 아니라 확률과 같은 응답 형태에 대한 지식도 충분히 갖추어야 한다. 전문가의 답을 구하는 적용 분야에서 전문가의 선정 시 고려해야 할 사항들을 보면 다음과 같다.

3.3.1.1. 전문가의 자격

전문가는 해당 분야의 전문 지식을 충분히 가지고 있어야 하며 또한 지식을 추출하고자 하는 상세 정도(level of detail)를 충족할 수 있어야 한다. 예를 들면 여러 가지 서로 다른 원자로 형태에 대한 전문가는 특정 파이프 파열에 대한 확률을 구하는데 필요한 전문가라고 볼 수 없다. 유사한 예로서, 특정 파이프 전문가는 냉각제 상실 사고(LOCA: Loss of Coolant Accident)의 상대적 빈도에 대하여는 잘 모르는 경우가 있는 것 등이다.

3.3.1.2 요구되는 전문 지식

지식 추출과 관련하여 요구되는 전문 지식에는 두 가지 형태가 있다. 하나는 질문의 대상이 되는 분야에 대한 전문 지식이고 다른 하나는 응답 모드와 관련된 전문 지식이다. 여기서 응답 모드는 전문가가 그의 판단을 제시하는 형식(예를 들면, 확률, 가능성, 연속 축척(scale), 순위나 등급, 쌍 비교)을 의미한다.

이 전문가의 판단을 구하는 작업에서 이 두 가지 중 어느 하나의 전문성을 결여한 사람을 지식 추출의 전문가로 활용할 경우에 응답 자료의 품질 저하와 같은 심각한 문제를 일으킨 것으로 나타나 있다[9].

여기서 가장 주의를 기울여야 할 사항은 해당 분야의 전문가가 항상 응답모드에 관련된 전문지식을 갖추고 있다고 생각하는 것인데 실제로 그렇지 않은 경우가 대부분이다.

일반적으로 위에서 언급된 두 가지 형태의 전문 지식을 모두 갖춘 전문가(예를 들면, 확률 추정에 경험이 있는 파이프 전문가)가 그렇지 않은 전문가(확률 추정의 경험이 없는 파이프 전문가)보다 지식 추출 작업에 더 적합하지만 현실적으로 현실에서는 이런 최적의 전문가를 선정하지 못할 경우가 종종 있다. 이에 대한 두 가지 대처 방안은 다음과 같다.

- (a) 전문가가 응답모드에 익숙하지 못한 경우 그들이 문제해결에 사용하는 모드(결정론적 방식)를 그대로 사용하도록 하는 것.
- (b) 응답 모드에 친숙해지도록 훈련 등을 시행

3.3.1.3 유명한 전문가를 선택해야 하는 경우

특정한 경우에는 위에서 언급된 전문성 외에 전문가의 다른 특성이 필요한 경우가 있는데 전문가의 명성이 여기에 속한다. 해당 분야와 대중에게 잘 알려져 있고 존경을 받는 전문가를 선정하게 되면 그 연구는 보다 높은 신뢰성을 얻을 수 있다. 그 예로는 미국의 식량과 자원을 예측하는 연구에서는 저명한 전문가들(Club of Rome, Limits to Growth)들을 채용하므로써 좋은 반응을 얻었다. 따라서 수행하려는 과제가 논쟁의 소지가 있는 것으로 판단되면 위에서 언급된 전문성(해당분야에 대한 전문성, 응답 모드에 관련된 전문성 등)은 물론이고 명성이 있는 전문가를 선정하는 것을 고려해 보아야 한다.

저명한 전문가를 채용하므로써 생기는 또 하나의 좋은 점은 그것이 다른 전문가들이 그 과제에 참여하는 동기부여가 되어 추가로 전문가를 구하는 일이 쉽게 된다는 점이다.

3.3.1.4 전문가 선정에 있어서 추가로 고려해야 할 사항

o 복수의 다양한 전문가

일반적으로 복수의 다양한 전문가를 선정해서 문제가 다각도로 완전하게 검토되는 것이 바람직하다. 다양한 전문가들이 모이면 한가지 문제를 다양한 측면에서 접근하여 다양한 방법으로 풀어내기 때문이다.

참여자의 다양성은 한 명의 전문가로부터 정보를 얻을 때 발생할 수 있는 여러 가지 바이어스를 최소화 시킬 수 있다. 특히 미래의 사건이나 상황을 예측하는 경우에는 전문가의 다양성이 매우 중요하다.

지식의 수집(acquisition) 작업에서는 복수의 전문가를 선정하는 것이 권장되고 있는데 그 이유는 결합된 추정치가 한 개인의 추정치보다는 실제 해답에 가까울 가능성이 높은 것으로 알려져 있기 때문이다.(Bryson, and Waller, 1985)

o 선정할 전문가의 수

선정할 전문가의 수는 추출 기법에 따라 달라진다. 예를 들면, 일대일 면접이 포함된 경우라면 5명에서 9명의 전문가가 적합하다. 5명보다 적은 전문가를 선정하면 충분한 다양성이나 추론에 대한 정보를 얻지 못하게 되는 경우가 많다.

3.3.1.5 전문가 선정 방안(계획)

일반적인 전문가 선정 방안은 전문가가 다른 전문가를 지정하는 방식이다. 특정화

된 많은 전문 분야에서는 전문가들은 서로를 잘 알고 있어 해당 연구에 적합한 다른 전문가를 쉽고 적절하게 추천할 수 있다. 이 방안에서 발생할 수 있는 문제점은 같은 관점을 가진 전문가들이 선정될 수 있다는 점이다. 따라서 이를 피하기 위해서는 또 다른 방안이 보충되어야 하는데, 예를 들면, 서로 다른 기관(학교, 연구소, 기업연구소, 정부기관 등)에서 각기 같은 수의 전문가를 선정하는 것 등이다.

3.3.1.6 전문가를 참여시키기 위한 동기 부여

전문가에게 동기를 부여하는 첫 단계는 전문가의 관점에서 제안된 과제에 대하여 생각해 보는 것이다. 즉 전문가에게 동기 부여가 될만한 요인(예를 들면, 기여나 인정 등)들을 극대화하고 그들이 참여를 꺼리게 되는 요인들(예를 들면, 과도한 시간 소요 등)을 최소화하는 것이다. 예를 들어, 위해도 평가 분야에서 동기 부여가 될 수 있는 두 가지 점을 보면 그 분야에 공헌을 할 수 있다는 점과 인정받을 기회가 생긴다는 점이다.

만약 선정 대상이 된 전문가들에게 그들이 수행하고자 하는 과제에 반드시 필요한 입력력을 줄 수 있다고 말해 준다면 그들은 보다 적극적으로 참여하게 된다.

일반적으로 전문가들은 그들이 과정을 제어할 수 있다면 그 과제에 대하여 보다 호감을 가지고 또 사용하고자 하는 기법이나 구하고자하는 결론에 대하여 적극적으로 지원하는 경향이 있다. 이런 경향은 “만약 내가 그것을 할 수 있다면 그것은 좋은 것이다”라는 인간의 잠재적 믿음에서 비롯된다. 또한 대상 전문가들이 그 과제에 참여하여 작업을 수행하게 되면 그들에게 어떤 이득이 생기거나 명성에 도움이 된다고 생각하게 되면 보다 적극적으로 참여하게 된다.

대상 전문가가 과제 참여를 꺼리게 만드는 요인들을 최소화하는 데는 주의가 필요하다. 일반적으로 이런 요인들 중에서 많은 시간을 빼앗기는 것이 가장 큰 요인이다. 이 부정적 요인을 적절히 처리하기 위해서는 작업 수행에 필요한 시간을 최소화하고 또 참여자에게 중요한 역할을 맡기는 것과 같이 다른 긍정적 방안을 통하여 수행해야 한다. 그리고 이런 방안들이 효과가 없으면 보수를 제공하는 방안을 고려해야 한다.

o 보수(pay)를 통한 동기 부여

일반적으로 보수를 통한 동기부여는 다른 어떤 방안도 가능하지 않을 경우나 또는 전문가가 아주 많은 시간과 노력을 투입해야만 하는 경우에 사용하는 것이 좋다. 왜냐하면 전문가가 과제의 본질적 관점에 관심을 가지고 참여했을 경우 보다 좋은 결과를 얻을 수 있기 때문이다. 보수를 제공해야 하는 전문가는 어떤 경우이든 마

지막 방안으로 고려해야 한다. 이 방안의 단점은 비용이 들며, 연구 결과를 왜곡시킬 가능성이 높고 과제에 대한 참여자의 관점에 예상하지 못한 영향을 줄 수 있기 때문이다.

보수와 관련된 한 가지 예외는 여행비용과 숙박비이다. 전문가들은 여행비용이나 숙박비와 같이 실제로 작업에 필요한 비용에 대해서는 그것을 그들의 시간과 노력에 대한 보수라고 생각하지 않고 있기 때문에 이런 비용은 반드시 지불해야 한다.

- 과제의 본질적 관점에 대한 커뮤니케이션을 통한 동기 부여

과제의 본질적 관점이 전문가에게 어떤 방식으로 전달되는가 하는 것이 전문가의 참여 여부에 중대한 영향을 미친다. 따라서 전문가에게 처음 참여를 요청할 때는 간단한 메모(1 페이지 이내)를 준비하는 것이 좋다. 이 메모는 전화를 통하거나 또는 면접 시 대본으로 사용한다.

대부분의 사람들은 메일보다는 사람을 통해서 전달된 요구 사항에 대해 보다 응답을 잘하는 경향이 있으므로, 직접 전문가를 만나서 이 메모를 전달하는 것이 바람직하다.

커뮤니케이션 이론에서 요약된 지침에 따르면 다음과 같은 정보의 아이템은 커뮤니케이션이 필요한 것으로 되어있다.

- 전문가와 접촉하게 된 이유
- 누가 그 연구를 지휘하고 후원하는가.
- 연구에 소요되는 시간, 자원, 기간, 시작 시기 등
- 어떻게 본인이 선정되었으며, 누가 추천했는지,
- 연구에서 수행되어야 할 내용
- 전문가의 판단은 익명으로 관리되는 지, 만약 그렇다면 어떤 방법으로 관리되는지.
- 예상 결과물과 그에 대한 접근 여부.
- 강제적 참여인지 또는 자원 형식인지에 대한 정보

3.3.2 전문가의 문제 해결 과정에 대한 자료가 필요한 적용 분야

문제 해결의 과정에 대한 자료를 구하는 경우가 보다 일반적으로 되고 있다. 응용 분야가 다음의 어느 한 가지에 속한 경우에는 특히 해결 과정에 대한 자료가 필요하다.

- (a) 문제가 현재 어떤 방법으로 해결되고 있는지를 결정하고 새로운 표준들을 정

립해야 할 경우

(b) 초보자의 문제 해결 방법을 평가하고 또 초보자의 교육을 위해 자료를 사용할 경우

(c) 지식기반 시스템을 구축하기 위하여 자료를 수집할 경우

위와 같은 경우에도 전문가의 답을 모으는 것이 종종 발생하나 그것은 주목적이 아니고 문제해결 과정에 대한 정보의 일부로서 수행된다.

3.3.2.1 전문가가 갖추어야 할 자격

문제 해결에 대한 자료를 구하는 작업에서는 단순히 답을 구하는 작업보다 더 많은 전문가의 자질이 요구된다. 이들 전문가는 문제 해결에 매우 능숙하여야 하고 그들의 문제 해결 과정을 기술하는데 명료해야 하며 어렵고 시간이 걸리는 작업에 자발적으로 동참할 수 있어야 한다.

일반적으로 이런 전문가는 어느 분야에서든지 매우 드물다. 전문가가 되는 과정에서 전문가의 기본적 생각 과정은 자동화되고 무의식화 되기 때문에 이를 명시적으로 나타내는 것이 어렵기 때문이다.

3.3.2.2. Method-driven selection

종종 인터뷰 담당자나 지식 엔지니어가 전문가를 선택할 수 없는 경우가 있다.(예를 들면 자금을 지원하는 조직에서 전문가를 지명하는 경우). 이때는 그 전문가에게 맞도록 지식 추출 방법을 만들어서 그 전문가가 그에 대한 응답을 할 수 있도록 해야 한다.

만약 적합한 전문가들이 충분히 있는 경우라면 만들어진 지식 추출 방법에 적합한 전문가만을 선정하는 것도 가능하며 이때 선정의 대상이 되는 전문가는 작업에 참여하고자 하는 자발성과 계획된 지식추출 기법을 감당할 수 있는 능력이 있는 사람이 된다.

이 경우 최종적인 전문가 선정 방안은 각 지식추출 기법을 각 전문가에게 시험적으로 시도하여 어떤 전문가/기법 조합이 가장 효과적인지 검토한 다음 이루어지는 것이 보통이다.

3.3.2.3 전문가에 대한 동기 부여

전문가의 답을 구하는 응용분야에서와 마찬가지로 여기에서도 전문가의 참여 동기를 유발하는 요인은 극대화하고 참여 동기를 위축시키는 요인은 최소화해야 한다.

문제해결 분야에서 특히 전문가의 참여 동기를 위축시키는 것은 작업에 필요한 많은 시간문제이다. 어떤 경우에는 전문가의 수를 늘려 각 전문가의 시간 부담을 줄이는 방안도 가능하지만 한 전문가의 문제해결 과정을 깊이 있게 추출해야 한다거나 전문가의 수가 적다면 이 방안도 구현하기 어렵다.

두 번째로 전문가의 참여를 위축시키는 요인은 두려움이다. 단순히 답을 구하는 응용분야와는 달리 이 분야는 문제 해결의 과정을 모두 들어내야 하므로 전문가들은 자신들의 그런 내부적 과정을 드러내기 꺼리며 또 자신들이 사용하는 과정이 틀렸을지도 모른다는 두려움을 가진다. 이러한 두려움을 완화시키는 방법은 과제의 목적을 적절히 설명해주고 또 그 결과물의 유용성을 인지시키는 것이다. 그에 더하여 그들에게 이런 분야(문제해결 과정을 밝히는 분야)에서 완전한 답이란 있을 수 없으며 또한 본 작업은 전문가에 대한 시험이 아니고 그들의 판단과정을 배워 유용하게 사용하는 것이 주목적이라는 것을 말해 주어 두려움을 완화시키는 것이다.

이 분야에서 세 번째로 전문가의 참여를 위축시키는 요인은 그들이 자신들의 문제 해결 과정을 적절히 명료하게 표현할 수 없을지도 모른다는 우려이다. 이러한 우려를 완화시키는 방법은 대부분의 전문가들이 동일한 문제점을 가지고 있다는 점과 그것을 해결하기 위한 기법들이 있다는 것을 설명해 주는 것이다.

문제해결 과정에 대한 자료를 구하는 작업은 단순히 답을 구하는 작업에 비해 위에서 언급된 추가적인 동기 위축 요인들이 있지만 동시에 동기유발 요인도 가지고 있다. 즉, 새롭고 색다른 작업에 참여할 기회를 제공한다는 점인데 과학자나 엔지니어에게는 이점이 매우 매력적인 요인이 된다. 중요한 것은 전문가가 작업기간 동안 지속적으로 흥미를 가질 수 있도록 해야 하는 점이다.

3.3.3 일반적으로 발생하는 문제점과 그들의 해결방안

(1) 전문가들이 작업에 참여하기를 원치 않는 경우

대상으로 선정된 전문가들 중 1/3에서 3/4정도가 참여를 수락하는 것이 일반적이다. 만약 참여를 수락한 전문가가 1/3미만이라면 전문가의 동기부여에 문제점이 있는 것으로 볼 수 있다.

이를 해결하기 위해서는 처음 접촉했던 전문가들 중 한사람에게 그의 참여거부 원인(요인)을 요청해 보는 것이 좋다. 대상자로는 평소에 알고 지내던 사람이거나 또는 솔직한 사람이 적합하다. 사적으로 이런 전문가에게 그가 참여를 거부한 실제 원인에 대하여 들어보는 것이다. 만약 그렇게 할 수 없는 경우에는 자문 전문가에게 물어보는 것을 고려해야 한다. 다만 자문 전문가는 과제에 직접 참여하고 있기

때문에 외부 전문가와는 관점이 다르다는 것을 인지해야 한다.

위의 방안들이 적합하지 않거나 불가능한 경우라면 이미 작성된 전문가의 동기 유발 요인과 위축 요인들을 다시 세밀하게 검토해야 한다. 동기 유발 요인들이 위축 요인들을 극복할 정도로 강하게 나타나지 않았거나 또는 생각지 못한 위축 요인들이 있을 가능성이 높기 때문이다.

대부분의 전문가들에게 작용하는 한 가지 동기 위축 요인은 그들의 노력과 판단이 실제 사용되지 않을 것이라는 믿음이다. 이 경우 전문가들은 이미 상위 층에서 결정이 이미 내려져 있으며 그들은 단지 전시용으로만 이용된다고 생각하는 것이다. 전문가들의 그런 생각(사실이든 아니든)에 대해서는 그 점을 솔직하게 분명히 밝히는 것이 바람직하며, 이를 위해 과제 관련자/관리자(상위층 등)에게 이에 대한 도움을 받는 것이 도움이 된다.

(2) 비전문가를 포함해서 모든 사람이 참여하기를 원하는 경우

너무 많은 사람들이 참여하고자 하는 것도 문제가 된다. 이때는 참여자를 그 분야별 전문성에 따라 그룹으로 나눈 다음 모든 참여자에게 응답을 받는다. 그리고 당초 전문가로 지정된 그룹의 응답은 분석을 위한 자료로 사용하고 그렇지 않은 다른 그룹의 응답 자료는 보조적 자료로서 기록만 해 둔다. 이렇게 함으로서 분석에 사용되지 않은 자료를 제공한 전문가 그룹은 그들의 관점이나 의견을 제시할 수 있는 기회를 가질 수 있고 또 다른 그룹에서 중요한 정보를 간과하지 않았나에 대한 검증 역할을 수행할 수 있기 때문에 만족하게 생각하며, 분석에 사용될 자료를 제공한 그룹은 그들의 전문성이 인정되었고 또 그들이 응답한 자료가 과제의 입력으로 사용되었다는 점에서 만족한다.

(3) 전문가가 너무 바빠서 실제로 필요한 정도의 깊이까지 작업에 참여할 수 없는 경우

일반적으로 이 어려움은 문제 해결 과정에 대한 자료를 구하는 경우에 더 심각하게 나타나며 그 징후는 위에서 언급된 전문가가 참여를 거부하는 경우와 비슷하다.

이를 해결하기 위한 방안도 유사하며 커뮤니케이션을 잘 활용해야 한다. 최상의 전문가는 그렇지 못한 전문가들보다 바쁜 것이 일반적이며 따라서 이들을 활용하고자 한다면 보다 강도가 높고 색다른 동기 유발 요인들을 찾아 그것을 제시해야 한다. 해당 전문가에게 어떻게 하면 참여할 것인지 직접 물어보는 것도 한 방법이다.

(4) 전문가를 선택한 시스템에 비평이 있을 경우

전문가 선정의 신뢰성에 대한 비평이 지식 추출 작업이 끝난 다음에 나온다면 그에 대한 대처 방안은 마땅한 것이 없다. 다만 비평을 방어하고 설명하기 위한 방법으로 선정 방안을 기록해 두는 것이 고작이다. 이를 위해 선정 기준이나 선정 기준이 만들어진 사유, 그리고 참여를 요청했던 전문가들과 실제로 참여한 전문가들에 대한 기록을 반드시 만들어 두어야 한다.

전문가 선정을 하는 과정에서 이에 대한 비평이 발생하면 두 가지 대처 방안이 있다.

- (i) 전문가 선정 방안을 다시 만들어서 그 분야의 최고 전문가들을 포함시키는 것이다. 만약 그 최고 전문가들이 참여를 수락하면 진짜 전문가가 참여하지 않았다는 일반적 비평들은 없어질 것이다. 만약 최고 전문가가 full-time 참여가 어려운 경우에는 부분적 참여를 유도해야 한다. 그리고 최악의 경우로 최고 전문가들이 참여하지 못하게 되는 일이 일어나더라도 최소한 그들에게 참여를 요청했다는 것은 인정받을 수 있게 된다.
- (ii) 선정 방안에 비평을 제기한 사람을 새로운 선정 방안 작성작업에 참여시켜 작업의 어려움을 인식하도록 하고 또 그들을 다른 전문가를 선정하는데 활용하는 것이다.

3.4 작업 4 : 추출 구성요소(components, or building blocks)의 선택 (Selecting the components of elicitation)

추출작업에 필요한 구성요소는 기본적으로 다음과 같다.

- o 추출 환경
- o 응답 모드
- o 문제해결 과정의 추출
- o 전문가 응답의 결합
- o 문서화

추출 작업 계획 단계에서는 현재의 작업 특성에 적합한 구성요소를 결정하고 그 구성요소의 추출에 가장 적합한 기법들을 결정해야 한다. 기법의 선정에 있어서 일반적 고려 사항은 사람들은 새로운 기법보다는 기존의 기법들을 선호하고 또 기존의 인정받은 기법이 그들의 작업 결과에 대한 신뢰도를 높인다는 믿음을 가지고 있다는 점이다.

3.4.1 필요한 구성요소의 결정-체크리스트

(1) 추출 환경

전문가의 판단을 모아야할 필요가 있을 경우에는 추출 환경이 필요하다. 전문가와 작업 수집자가 어떻게 만나고 또 어떻게 전문가의 판단이 획득되어야 하는가에 대하여는 작업의 특성에 관계없이 어느 정도 단계가 필요하다.

(2) 응답 모드

전문가의 응답이 특정 형태(형식)로 작성되어야 하거나 또는 전문가가 자신들의 판단을 특정 형태로 변환해야 하는 경우에는 응답모드/분산측정이 필요하다.

(3) 문제해결 과정의 추출

문제해결 과정의 추출은 (i) 과제의 목표가 지식의 획득이거나 (ii) 전문가가 어떻게 그들의 결론에 도달했는가에 대한 관심이 있을 경우, 그리고 iii) 전문가의 응답들을 결합할 경우에 필요하다.

만약 전문가들이 문제를 다르게 해석했다면 그들의 응답을 결합하는 것은 사과와 오렌지를 섞는 것과 유사하다. 그들의 문제 해결과정에 대한 정보를 사용하면 이런 실수를 방지할 수 있다.

(4) 전문가 응답의 결합

복수의 전문가가 선정되고(따라서 그들의 의견은 종종 상충된다) 하나의 결론이 필요할 경우에는 전문가 응답의 결합이 필요하다. 예를 들면, 위해도 분석에서 여러 전문가의 추정치를 결합하여 모델에 사용할 하나의 추정치를 결정하는 경우 등이다.

(5) 문서화

전문가의 답이나 문제해결 과정에 대한 영구적(또는 반영구적) 기록이 필요한 경우이다.

3.4.2 각 구성요소로부터의 선정

(1) 추출 환경으로부터 선정

전문가의 판단을 추출하는 환경(또는 방법)은 (i) 전문가들이 함께 모여 토의를 통하는 방법과 (ii) 인터뷰 담당자와 전문가가 일대일로 만나 면접을 통해 수행하는 방법과 (iii) 전문가를 인터뷰 담당자나 다른 전문가들로부터 물리적으로 고립시킨 후 그들의 판단 자료를 메일이나 전화를 통하여 커뮤니케이션을 수행하는 방법의

세 가지가 있다

실제의 추출 작업에서는 이들 세 가지 방법 중 어느 한가지만을 단독으로 사용하는 경우는 드물고 해당 작업의 특성에 따라 두 가지 이상이 겸용된다.

a. Interactive group

o 기법 개요

이 방법은 전문가들과 세션 조정자(또는 인터뷰 담당자)가 모두 모여 회의를 하면서 작업을 수행한다.

o 장점

다른 두 가지 방법에 비하여 보다 정확한 자료를 구할 수 있고 얻어지는 자료의 양도 많다. 이것은 전문가들이 그들의 생각을 교환함으로써 만들어지는 시너지 효과에 기인한다.

o 단점

그룹 사고에 관련된 바이어스의 잠재적 가능성이 있고 회합 일정의 관리나 복수의 전문가를 관리해야 하는 어려움이 있다.

b. Delphi

o 기법 개요

Delphi 기법은 전문가들이 상호간 또는 중재자와 교신을 하지 않고 단독으로 그들의 의견을 제시한다. 중재자는 이들 의견들을 모아 익명으로 다시 전문가들에게 배포하면 전문가들은 그것을 참고하여 자신들의 의견을 갱신하게 된다. 필요하다면 합의에 이를 때까지 이런 작업이 반복된다. 이 기법은 상호 교신에 의한 바이어스를 줄이기 위해 개발되었다.

o 장점

그룹 다이내믹스로부터 야기되는 바이어스를 피하기 위해 고안되었다.(그러나 그 효과에 대하여는 이론이 있다.)

o 단점

얻어지는 자료의 양에 제한이 있으며, 그룹 상호간의 교신을 통한 시너지 효과를 기대할 수 없고 또 다른 방법에 비하여 시간이 가장 많이 소용된다.(전자 메일을 통한 자료 취합/배부의 경우에는 소요 시간을 많이 줄일 수 있다.)

c. 개별 면접

o 기법 개요

인터뷰 담당자가 전문가와 일대일 면접을 통하여 자료를 구하는 방법으로 가장 융통성 있게 구조화시킬 수 있다. 구조화되지 않은 인터뷰는 인터뷰 담당자가 자료수집 목표나 질문들을 사전에 개략하지 않은 경우에 발생한다.

o 장점

상세한 자료를 얻을 수 있는 가장 좋은 방법이며 전문가의 문제 해결 과정에 대한 자료를 구할 때 주 방법으로 사용된다. 그룹 다이내믹스로부터 발생하는 바이어스들을 피할 수 있고 추후에 수학적 결합기법들을 통해 자료를 통합 할 수 있다.

o 단점

시간이 많이 소요되며 전문가들 간의 의견 교환을 통한 시너지 효과가 없다.

(2) 응답모드/분산 측정으로부터 선정

가. 응답모드

a. 물리적 양의 추정(또는 평가)

o 개요

전문가가 기술적 질문에 대하여 물리적 양(온도, 시간, 압력, 부피)에 대한 평가를 응답하는 것이다. 이 방법은 가장 많이 사용되며 흔히 다른 응답모드(확률 평가 등)와 병행되어 사용되어진다.

o 장점

물리적 과정을 측정하기 어렵거나 불완전하게 이해된 질문에 대한 응답 형태로는 가장 편리하고 융통성있는 형식이다. 전문가들은 이런 물리적 양의 추정이나 평가를 응답 모드로 사용할 때는 이것이 그들의 작업에서 자주 사용되는 것이기 때문에 거의 어려움을 느끼지 않으며 따라서 이 응답 모드에 대하여는 특별한 교육이 필요 없다.

b. 확률 추정(또는 평가)

o 개요

전문가가 질문에 대한 응답으로 확률 평가치를 제시하는 것이며 주로 어떤 사건이 발생할 가능성을 예측하는데 사용된다. 동일 질문에 대한 복수 전문가의 확률 평가

치 응답은 확률 분포를 구성하기 위해 사용될 때도 있다.

o 장점

의사결정 분석에서 가장 흔히 사용되는 방법이다. 확률 기반 응답모드를 사용하는 가장 큰 이유 중의 하나는 이를 기반으로 한 추출과 분석 기법들이 많이 확립되어 있기 때문이다. 또한 확률 기반 평가는 모델링과 분석에 가장 편리한 형식이다.

o 단점

대부분의 전문가는 확률 평가에 익숙하지 못하고 따라서 이런 형식의 응답모드 사용을 꺼려하는 경향이 있다. (확률 평가에 대한 교육 등으로 이런 현상을 완화시킬 수 있다.) 그리고 이 방법을 정확하게 적용하려면 시간이 매우 많이 소요되고 이것은 전문가를 지치게 만들 우려가 있다.

c. Odds ratio

o 개요

이것은 전문가가 질문에 대하여 “전체 N번의 시도 중 x번 가능성이 있다”와 같은 형태로 응답하는 것이다. 예를 들면 어떤 특정 사건이 일어날 가능성은 1000번 중 1번이다“ 등으로 대답하는 것이다. 발생 빈도가 희소한 물리적 사건의 평가에 가장 흔하게 사용된다.

o 장점

희소 사건의 빈도나 가능성을 평가하는 가장 편리한 형식이다. 대부분의 사람들은 희소 사건을 생각하는데 있어서 확률(0.001)보다는 이 방식(1 in 1000)이 더 쉽다고 생각한다.

o 단점

전체의 시도(trials) 횟수가 너무 적게 설정되면 그것은 전문가의 판단에 부정적 영향을 미치고 그 결과로 전문가는 적절한 평가를 수행하지 못할 수가 있다.

d. 확률 분산

o 개요

여기에서의 확률 분산은 광범위한 의미로 사용되며 추정을 위한 가능한 값들의 집합과 그리고 그에 연관된 각 값들의 발생 가능성 또는 확률을 의미하며 각 값들의 집합은 최소/최대 절대 값을 포함해야 한다. 확률분포함수(probability density

function: pdf)가 이 형식에서 가장 흔히 사용된다.

o 장점

앞에서 기술된 확률 평가와 동일하게 의사결정 분석에서 가장 흔히 사용되는 방법이다. 확률 기반 응답모드를 사용하는 가장 큰 이유 중의 하나는 이를 기반으로 한 추출과 분석 기법들이 많이 확립되어 있기 때문이다. 또한 확률 기반 평가는 모델링과 분석에 가장 편리한 형식이다.

o 단점

대부분의 전문가들은 확률 평가를 잘 하지 못하며 따라서 이런 응답모드의 사용에 거부감이나 저항감을 느낀다. 또한 전문가가 확률 분산의 개념을 충분히 이해하지 못하는 경우가 종종 있으며 이 경우 별도의 교육이 필요하다. 정확히 수행하기 위해서는 많은 시간을 요한다는 점과 그에 따라 전문가가 피로해 진다는 점은 확률 값 평가와 동일하다.

e. 연속 척도(scale)

o 개요

연속 척도는 선형 또는 로그 스케일로 된 값을 가진 연속선이다. 선의 시작과 끝에 표시되는 값은 극단 값으로 문자나 숫자로 표시되고 그 사이의 척도 값은 정수, 문자로 표현된 범주, 확률, 등급, 또는 물리적 양의 측정치로 표시된다. 전문가는 질문에 대하여 이 연속 척도의 어느 부분에 표시하는 형태로 자신의 응답을 제시한다. 척도에 표시되는 라벨들은 어떤 물리적 양의 측정치가 사용되지 않았다면 분명하게 정의되도록 주의해야 한다.

o 장점

사용 방법에 대한 교육이 거의 필요 없고 분석을 위해서 숫자나 연속변수로 쉽게 변환 할 수 있다. 대부분의 사람들은 이 연속 척도를 사용하여 정확한 값을 평가할 수 있다.

o 단점

특정한 용도를 위하여 연속선형 척도를 개발하는데 많은 시간과 노력이 필요하다. 그리고 라벨의 표현 방법에 기인하는 바이어스를 줄이기 위해 주의가 필요하고 이들 라벨에 대한 명확하고 분명한 정의를 만드는데도 주의가 요구된다. 로그 척도가 사용된 경우에는 이에 대한 교육이 필요하다.

f. 쌍 비교(pairwise comparisons)

o 개요

쌍 비교는 전문가로 하여금 개체나 사건 또는 기준의 집합들을 평가할 때 한 번에 단지 2개 만 비교하게 하여 그 순위를 정하게 하는 방법이다. 비교의 기준은 중요성, 발생의 가능성, 어떤 특성의 소유 등이 될 수 있다. N개의 개체가 있는 경우 모든 가능한 쌍 비교를 위해서는 $n(n-1)/2$ 번의 비교가 필요하다.

o 장점

대부분의 사람들은 쌍 비교를 사용함으로써 적절한 평가를 할 수 있는데 그것은 한 번에 단지 2개만을 고려하기 때문이다. 즉, 그들의 정보처리 능력을 초과함으로써 생기는 오류를 피할 수 있다는 점이다. Saaty의 Analytic Hierarchical Process(AHP)와 같은 방법은 전문가가 제시한 응답들에 대한 수학적 일관성을 검증하는 수단을 제공한다. 이 방법은 간단한 브리핑만으로도 설명이 가능하므로 전문가들이 사용하기 쉽고 또 정성적 자료에 대한 정량적 분석을 가능하게 해 준다.

o 단점

비교 가능한 모든 조합을 추출하는데 많은 시간이 요구되고 또 자료의 상대적 관계만을 알 수 있다. 그리고 상대적 비교를 절대 관계로 변환하기 위해서는 기본 척도(baseline scale)나 값이 필요하다.

g. 등급화(ranks or ratings)

o 개요

등급화는 문제가 되는 객체나 사건 또는 값에 대하여 전문가가 생각한 특정 순서대로 숫자나 설명을 배당하는 것이다. 등급은 정수의 오름차순이나 내림차순 또는 순서를 나타내는 설명(좋음, 보통, 나쁨 등)으로 표시된다. 예를 들면, 전문가가 가장 중요하다고 생각하는 항목에는 “1”을 그리고 가장 덜 중요하다고 생각하는 항목에는 “5”를 부여하는 것이다.

등급 라벨이 전문가마다 다르게 해석할 여지가 있는 경우에는 숫자와 단어를 동시에 사용하는 것이 바람직하다. 예를 들면, 정성적인 용어 “작은 기회”는 전문가에 따라 1%에서 40%까지 의미하는 것으로 나타났다.(ref: Keeney 1989). 따라서 등급에 대한 전문가들의 해석을 일관성 있게 유지하기 위해서는 가능하다면 숫자와 정성적 설명을 동시에 사용해야 하며 아래의 예와 같다.

[숫자와 정성적 설명을 동시에 사용한 등급 라벨의 예]

- 3 - Completely related, approximately 80% related
- 2 - Significantly related
- 1 - Slightly related, approximately less than 20%
- 0 - Not at all related

전문가들이 동일한 등급을 서로 다르게 해석할 가능성을 줄이기 위해 Sherman Kent는 다음과 같은 등급 척도를 개발하였다.

표 6. Sherman Kent의 등급 척도

Order of Likelihood	Synonyms	Chances in 10	Percent
Nearly Certain	Virtually certain		99
	We are convinced	9	
	Highly probable		
	Highly likely	8	80
Probable	Likely		
	We believe		
	We estimate	7	
	Chances are good		
Even Chance	It is probable	6	60
	Chances slightly better than even		
	Chances about even	5	
Improbable	Chances slightly less than even	4	40
	Probably not		
	Unlikely	3	
Nearly Impossible	We believe not	2	20
	Almost impossible		
	Only a slight chance	1	
	Highly doubtful		10

o 장점

전문가들이 사용하기 쉽고 따라서 교육이 거의 필요 없다. 등급화 방법은 정성적 정보를 다루어야 하는 적용분야나 또는 가능한 답이 매우 제한된 적용분야에 적합하다.

o 단점

사람들은 7(±2)개 이상의 것을 한번에 처리하는데 어려움을 느낀다(Miller, 1956).

따라서 전문가들이 많은 아이템을 비교하고 그들을 등급화 하는 것은 어렵고 부정확할 가능성이 많다. 7개 이상을 비교할 때는 보다 많은 전문가를 활용하거나 또는 적은 전문가들에게 보다 많은 시간을 할당해야 한다.

h. 베이시안 갱신(Bayesian updating)

o 개요

베이시안 갱신은 서로 다른 근원의 정보를 결합함으로써 추정치를 갱신하는 과정이다. 이 방법으로 측정된 자료와 전문가의 판단 자료를 결합할 수도 있고 또 한 전문가가 서로 다른 시기에 제출한 동일한 문제에 대한 응답을 결합하거나 여러 전문가들의 응답을 결합할 수도 있다. 이 방법은 또한 자료의 결합 기법으로도 사용된다.

o 장점

다양한 형태의 자료를 결합하는 편리한 수단을 제공하며 자료의 조건부 성질(conditional nature)을 고려할 수 있다.

o 단점

각 자료의 분포에 대한 가정을 필요로 한다. 또 가정된 분포의 파라미터에 대하여 전문가의 추가적 평가가 필요할 때도 있다.

나. 이산(분산, dispersion) 측정으로부터 선정

a. 범위(Ranges)

o 개요

범위는 추정 값이 존재하는 곳으로부터 양쪽으로 동일한 일정 간격을 나타내는 두 개의 값이다. 일반적으로 전문가들에게 절대 최소/최대값을 요구하는 경우가 많다. 예를 들면, 한 전문가가 어떤 사건의 확률 값을 0.001로 평가했다면 그에 더하여 그 사건의 최소 값과 최대 값을 요청하게 되고 그 요청에 대한 응답은 최소 0.001에서 최대 0.005등이 될 수 있다.

o 장점

범위는 쉽게 추출 할 수 있다. 또한 이것은 분석 목적을 위한 이산 측도로서 유효하다.

o 단점

일반적으로 사람들은 최대 값과 최소 값의 추정에 정확하지 못하다. 즉 최대 값은 과소 추정하고 최소 값은 과대 추정하는 경우가 자주 일어난다. 그리고 범위는 분석 목적에 적합하게 정의될 수 없는 경우가 종종 있다.

b. 백분위수(Percentiles)

백분위수는 전문가의 추정치가 특정한 분포 값으로 지정되는 경우인데 이 분포 값은 미리 지명된 조각으로 나누어져 있다. 확신도 구간 추정은 두 개의 다른 백분위수를 추정하는 것과 동일한 경우이다. 예를 들어, 전문가가 어떤 사건의 실패 확률을 0.01이라고 추정하고 5번째 백분위수와 95번째 백분위수를 각각 0.0001과 0.01로 추정했다면 이 구간(0.0001, 0.01)은 90% coverage 구간에서 전문가의 추정 값이 된다.

o 장점

여러 가지 분석과 모델링 기법에 편리하게 적용할 수 있다.

o 단점

인간은 일반적으로 특정 백분위수를 추정하는 것을 잘 하지 못한다. 따라서 교육이 필요한 경우가 많다.

c. 분산/표준 편차(Variances, standard deviations)

고전 통계에서 사용되는 분산이나 표준 편차의 사용을 의미한다.

o 장점

기존의 여러 가지 분석과 모델 기법에 적용하기 용이하다.

o 단점

일반적으로 사람들은 분산이나 표준 편차를 사용한 측정이나 평가에 익숙하지 않고 부정확하다. 이 응답 모드를 사용할 경우에는 해당 전문가가 통계에 익숙하지 않다면 교육을 해야 한다.

(3) 문제해결 과정 추출을 위한 방법으로부터 선정

문제해결 과정을 추출하는 방법으로는 verbal protocol, verbal probe, ethnographic technique이 가장 쉽고 간단한 방법이다. 다른 방법도 있으나 이들 세 가지가 전문가에게 적용하기 쉽고 또 바이어스에 대하여 내성을 가장 많이 가지고 있다.

a. Verbal protocol

Verbal protocol은 전문가로 하여금 그들의 문제해결 과정을 말하도록(think aloud) 하는 것으로 심리학에 기초한 방법이다.

o 장점

동기적 바이어스가 개입되는 것을 피할 수 있다. 왜냐하면 인터뷰 담당자가 전문가에게 질문을 하는 것이 아니기 때문에 인터뷰 담당자의 유도가 개입될 여지가 없기 때문이다.

o 단점

한번에 한 명의 전문가에게만 사용되어야 한다. 따라서 그룹 환경이나 또는 전문가/인터뷰 담당자가 전화나 메일로 커뮤니케이션을 할 때는 사용할 수 없다. 그리고 이 방법은 시간이 매우 많이 소요된다. 전문가는 적어도 두 번은 그의 생각을 언어로 나타내어야 하고 또 그들의 생각을 말로 잘 표현할 수 없는 경우도 많기 때문이다.

b. Verbal probe

Verbal probe는 마감 질문으로 특정 시점에서 특정 방법으로 수행된다. 추출 작업에 사용되는 verbal probe는 전문가가 해결에 도달한 직후 주로 사용되고 그가 방금 해결한 그 문제에만 초점을 맞춘다. 예를 들면, 전문가가 어떤 문제를 해결하고 응답을 제시했다면 verbal probe는 왜 그 응답이 나왔는가를 알기 위해 사용된다.

o 장점

문제 해결에 사용된 전문가의 추론에 대한 일반적 정보를 얻는 빠른 수단이다. 개별 전문가에게나 또는 그룹 환경 모두에서 사용 가능하다.

o 단점

Verbal probe는 전문가와의 일대일 상황에서 그 전문가가 말로 응답할 수 있는 경우에 가장 유효한 방법이고 서면에 의한 응답 방식은 부적절하다. 그리고 verbal protocol에 비하여 동기적 바이어스가 개입될 가능성이 조금 높다.

c. Ethnographic technique

이 방법은 전문가의 응답을 질문으로 바꾸어 다시 질문하는 형태이다. 예를 들어, 전문가가 망아지 사육 프로그램은 망아지가 몸무게를 유지하지 못하는 경우에는 프로그램의 조정이 필요하다고 말했다면;

인터뷰 담당자: 몸무게를 유지하지 못하는 경우에 그렇다는 말인가요?

전문가: 예. 그 시점에서 망아지는 몸무게를 유지해야 합니다.

인터뷰 담당자: 그 시점에서요?

전문가: 예. 6개월이 된 그 시점에서는 매월 1.5에서 2파운드 정도는 몸무게가 늘어야 합니다.

o 장점

전문가의 문제 해결 과정에 대하여 매우 상세한 자료를 얻을 수 있고 또 전문가에 응답에 기초한 질문이기 때문에 다른 질문 방법에 비하여 상대적으로 바이어스의 개입 소지가 적다. 이 방법은 또한 인터뷰 담당자가 전문가의 용어를 제대로 이해했는가를 점검하는 방법도 된다.

o 단점

일반적으로 시간이 많이 걸리기 때문에 그룹 환경에서는 적합하지 않고 개별 전문가에게 단지 몇 개의 정의만을 추출하는데 적용 가능하다. 또 전문가가 문제를 푸는 도중에는 이 방법을 사용할 수 없는데 그것은 전문가의 주의를 흐뜨리기 때문이다.

(4) 결합 형태의 선정

과제의 필요에 의하여 복수의 전문가 또는 전문가의 서로 다른 답으로부터 하나의 답을 구해야 하는 경우가 종종 있다. 이를 위한 기본적 방법은 I) 전문가들로 하여금 하나의 합의된 해답에 도달할 때까지 계속 작업을 하게 하는 것과 ii) 전문가들의 응답을 수학적으로 결합하여 하나의 해답으로 만드는 방법이다.

a. 행동적 집성(Behavioral aggregation)

행동적 집성 방법은 전문가의 합의에 의존하는 방식으로 합의는 종종 추출 세션 중에 이루어진다. 그룹 환경에서는 중재자가 전문가들에게 하나의 합의안을 제시할 것을 요청하고 이것은 전문가들 사이에서 설득과 양보의 방식으로 결정된다. Delphi 방법에서는 하나의 해답에 도달할 때까지 계속적으로 전문가 응답을 교환하는 방식으로 진행된다.

o 장점

세션 진행 중에 하나의 합의된 답을 구할 수 있고 또 합의된 응답이기 때문에 전문가의 익명성을 보호받을 수 있다. 그리고 이 방법은 전문가들로 하여금 응답 결과를 지지하게 만드는 효과가 있다.

o 단점

특정 추출 환경 및 바이어스 방지 방안과 함께 사용되어야 하므로 사전의 세밀한 계획이 필요하다. 이 방법은 특정한 경우에는 “누구도 실제로는 그렇게 생각하지 않지만 부지중에 묵인하게 되는” 그룹 사고 상황을 만들 수도 있다. 이 방법은 또한 그룹 사고가 무의식적 동의 방식을 허용하지 않는다면 많은 시간을 요하고 또 다른 의견 제시를 억압하는 상황이 되기 때문에 올바른 답을 찾을 가능성을 낮춘다.

b. 수학적 집성(Mathematical aggregation)

이 방법은 복수 전문가의 응답을 하나의 평가 값 또는 하나의 추정 분포로 만들기 위해 수학적 수단을 사용하는 것이다. 예를 들면 전문가 각각에게 동일한 가중치를 준다던가 또는 전문가의 전문성이나 자료의 중요성에 따라 서로 다른 가중치를 주는 방식이다.

o 장점

행위적 집성 방식보다는 계획이 늦어도 된다. 또 서로 다른 수학적 방식을 연속적으로 적용 가능하다.(반면 행위적 집성 방식은 한번만 가능하다.)

o 단점

다른 모든 집성 방식과 마찬가지로 전문가 응답들 사이의 상이성과 그 상이성의 이유를 모호하게 만든다. 그리고 어떤 경우에는 이 방식으로 사용하여 만든 하나의 응답이 관련된 모든 전문가가 거부하는 답이 되는 경우도 종종 발생한다. 예를 들면, 전문가들이 어떤 물리적 과정을 기술하는 서로 다른 분포 상태들을 응답했다면 이들을 수학적으로 집성한 결과는 현실에서는 일어날 수 없는 현상을 기술하는 내용이 될 수도 있다.

(5) 문서화 방법으로부터 선정

전문가의 판단을 기록하는 옵션은 두 가지가 있다: (i) 전문가의 응답만을 기록. (ii) 전문가의 응답과 문제해결 과정에 대한 정보를 기록.

(i) Answer-only 문서화

전문가의 대답 또는 해결 방안만을 기록하는 것이다.

o 장점

가장 빠르고 쉬운 문서화 방식이다.

o 단점

나중에는 어떻게 그 해답에 도달하게 되었는지에 대한 상세한 정보를 재구축 할 수 없다.

(ii) 응답 및 문제해결 문서화

전문가의 응답과 그에 대한 해결 과정을 모두 기록하는 방식이고 그 상세 정도는 상황에 따라 달라진다.

o 장점

전문가의 판단이나 추출 과정에 대한 비평이 있을 경우 방어용으로 사용 가능하다. 또 판단을 갱신하거나 수정하는데 필요한 자료를 제공한다.

o 단점

기록된 자료 때문에 비평받을 소지가 발생한다. 즉 기록하지 않았으면 그냥 넘어갈 수 있었는데 기록함으로써 문제를 야기한 경우이다. 그러나 외부 검토자들의 최근의 경향은 기록이 빠지거나 불충분한 작업에 대하여 보다 비판적이다. 따라서 전문가들이 어떻게 해답에 도달했는가에 대하여 기록을 하고 그 중 특정 문제에 대하여 비평을 받는 것이 이런 정보를 기록하지 않은데 따른 정보의 불충분으로 비평을 받는 것보다 낫다.

3.4.3 일반적으로 발생하는 문제점과 해결 방안

(1) 선정하고자 하는 기법들에 대한 자료가 희소하거나 또는 내용이 서로 상충될 때

추출 작업을 위한 구성요소들을 선택할 때 지침으로 사용하기 위해 여러 가지 관련 문헌을 참고하게 된다. 특히 특정 적용분야나 환경에서 어느 방법이 가장 적합한가를 알기 위해 문헌을 조사하게 되는데 이들 방법들을 비교한 문헌은 매우 드물고 또 어떤 방법에 대해서는 서로 상충되는 견해를 나타내는 수도 있다. 이런 현상은 개별 면접, Delphi 방법, interactive group methods, 응답 모드, 자료의 결합 방식에 대한 기법들을 조사할 때 자주 나타난다.

주어진 환경에 가장 적합한 추출 구성요소를 선정하는 데는 다음의 두 단계가 유용하다.

(a) 1단계

기법과 관련된 문헌을 조사한다. 최근에 지식 수집(knowledge acquisition) 분야에

서 서로 다른 추출기법에 대한 비교와 평가에 대하여 새로운 시도가 진행되고 있다.

(b) 2단계

가능한 추출 구성요소들을 나열한 다음 주어진 환경에서의 그들의 장단점을 분석한다. 그리고 이 목록을 추출 구성요소 선정을 위한 의사결정 지원 목적으로 사용한다. 각 기법의 강점과 약점에 대한 정보의 취합은 문헌들에 나타난 것뿐만 아니라 작업자 자신의 관점에서 생각한 것도 포함시켜야 한다. 그리고 비교의 기준으로는 저자가 명명한 이름이 아니라 그 기법의 정의를 사용해야 한다.

이 목록은 추후에 기법 선정의 정당성을 증명하기 위해 사용되어 질 수 있으며 최종 보고서 등에 포함시킬 수 있다. 보고서에는 특정 방법을 선정한 논리적 이론적 근거를 기술해야 하고 표준적 방법을 변경하여 사용하였다면 그 이유를 설명해 두어야 한다.

3.5 작업 5 : 추출의 설계와 맞춤 (Designing and Tailoring the Elicitation)

추출 구성요소들과 그에 대한 기법들이 결정되면 다음 단계는 이들을 특정 분야에 적합하도록 맞추어야 한다.

3.5.1 추출 설계상의 고려 사항

(1) 전문가들을 소집하거나 또는 개별적으로 면접시 발생하는 비용

전문가의 판단을 얻기 위한 기본적 추출 환경이 결정되었다 할지라도 다른 기법들을 추가하거나 보완할 필요가 있다. 그렇게 함으로서 보다 빠르고 경제적으로 추출 작업을 수행 할 수 있기 때문이다. 특히 interactive group 이 기본 추출 환경으로 결정된 경우에는 보다 비용이 적게 드는 Delphi 또는 개별 면접과 같은 방법을 병행하는 것이 바람직하다. Interactive group은 모든 전문가와 과제 관련자가 함께 모여야 하므로 많은 비용(시간과 예산)이 들고 또 현실적으로 일주일에 한 번 이상 모든 전문가를 소집하여 회의를 하는 것은 어렵다. 따라서 이런 회의는 반드시 필요한 경우에만 하도록 하고 그 외에는 보다 비용이 적게 드는 Delphi나 개별 면접을 활용해야 한다.

전문가들을 모두 소집하여 회의를 해야 하는 경우는 다음과 같다.

- o 복잡한 응답모드와 분산추정을 사용할 경우
- o 추출 과정에서 그룹토의의 시너지 효과가 필요할 경우

- 전문가들을 다른 업무로부터 해방시켜 현재의 작업에 주의를 집중하게 할 필요가 있을 경우

한편 Interactive group을 사용할 경우 발생하는 비용은 다음과 같은 것들이 있다.

- 전문가의 여행/숙식비용
- 회의실과 회의 진행에 필요한 물건 등(전문가가 깨어있도록 제공하는 커피 등)
- 회의 녹화 비용
- 기타 행정/관리비용(타이핑, 복사, 회의 통보, 관련자료 준비 등)

(2) 추출 과정의 구조화

구조화는 추출 과정에 제어/관리를 부과하는 것을 의미한다. 여기에는 앞에서 기술된 질문의 구조화 외에도 (i) 추출을 가이드하기 위한 사전 설계된 질문 세트, (ii) 전문가들 간에 특정 종류의 커뮤니케이션만 허용, 그리고 (iii) 전문가에게 응답 모드 중 하나만 사용하게 요구하는 것 등이 포함된다.

(3) 바이어스의 처리

바이어스에 주의를 기울여야 하는 이유는 (i) 바이어스가 발생하면 결과물의 품질이 저하될 가능성이 높고 (ii) 편견에 대한 우려(과제 검토자들 사이에서 주로 발생한다)가 과제의 신뢰도에 영향을 미치기 때문이다.

(4) 추출 전/후의 문서화

문서화에 포함되어야 할 내용은 다음과 같다.

- 문제의 설명과 그 마지막 버전
- 전문가 이력
- 전문가의 자료를 구한 방법
- 전문가의 응답

(5) 질문의 제시

3.5.2 일반적으로 발생하는 문제점과 해결 방안

(1) 추출 계획에 충분한 시간이 할당되지 않은 경우

가장 보편적으로 발생하는 문제점으로 사람들이 추출 작업을 수행하는데 있어서 계획 단계를 불필요하게 생각하는데서 기인한다. 일정에 계획단계가 빠져 있다면 이런 문제가 발생할 소지가 많다.

계획이 없더라도 그 결과가 심각하지 않을 특수한 경우를 제외하고는 계획은 필수적이며 일정을 조정해서라도 계획을 수립하기 위해 필요한 시간을 할당하여야 한다.

계획 단계를 생략 또는 간단히 처리할 수 있는 경우는;

- (i) 단지 몇 명의 전문가들과 직접 면접을 수행하는 경우이다. 이 경우 그룹 미팅을 위한 준비나 또는 자료의 수집과 재 배부(Delphi 방법 사용시)와 같이 시간이 소요되고 복잡한 작업이 필요 없기 때문에 계획은 매우 간단해지므로 생략 가능하다.
- (ii) 구조화되지 않은 기법들을 사용할 경우에는 계획이 불필요하다. 구조화를 잘 하려면 많은 계획 시간이 필요하다. 그러나 구조화된 기법들은 비록 계획단계에서 많은 시간을 요하기는 하지만 추출 과 분석의 효용성을 높여준다는 것을 기억해야 한다.

(2) 바이어스 가능성의 무시

과제 스태프들의 무지로 인하여 바이어스의 발생 가능성이 종종 경시된다. 그들은 그것의 존재 자체를 모르거나 또는 의도적으로 무시하기도 한다. 두 가지 경우 모두 자주 일어나는 현상인데 그것은 전문가의 판단에 있어서 바이어스에 대한 주제가 일반적으로 언급된 적이 없고 또 그것을 다루는데 익숙해 있지 않기 때문이다.

그러나 전문가의 판단에 대한 외부 검토자의 비평은 늘어가는 추세에 있다. 일반적으로 이들 외부 검토자들은 인터뷰 담당자가 전문가들을 리드함으로써 발생하게 되는 바이어스들에 대하여 민감하다. 예를 들면, NUREC-1150 작성의 경우 전문가 판단 사용 방법을 검토한 어떤 위원회에서는 전문가들로 하여금 과제관련자가 제안한 문제들에 근거하여 작업을 하도록 한 방안에 대해 비판하였다. 그리고 전문가들이 독립적으로 그들 자신의 문제를 만들고 정련 하는 것을 허용하도록 주장하였다.

외부 검토자들로부터 자주 제기되는 또 하나의 비평은 group think이다. 여기서 발생할 수 있는 우려는 일부 전문가들이 과도하게 그들의 의견을 주장하거나 또는 의도적으로 자신의 의견을 표현하지 않고 다른 사람들의 의견을 수용해버리는 것이다.

이해의 상충도 바이어스의 원인중 하나이다. 이런 바이어스가 발생하면 전문가의 판단에는 그들이 잠재적으로 원하는 것이 반영되게 되고 극단적 경우에는 전문가가 보수를 받고 지원해야 하는 위치나 입장을 반영하는 판단이 되어버리는 경우도 생

긴다.

바이어스 문제를 대처하는 일반적 방안은 실제 추출 작업을 진행하기에 앞서 설계 단계에서 본 장에서 기술된 바이어스에 관련된 내용들을 반영하는 것이다. 또 다른 부분적 접근 방법은 위의 세 가지 바이어스(인터뷰 담당자로부터의 사회적 압력, 그룹 사고, 희망적 사고)에 대하여 초점을 맞추는 방안이다. 그리고 마지막 방안으로 바이어스의 발생 가능성을 고려하지 않고, 추출 작업 후에 바이어스의 존재 여부를 확인 할 수 있을 정도의 충분한 자료를 구한 다음 바이어스를 시험하는 여러 가지 기법들을 사용하여 모아진 자료를 검증하는 것이다. 이 방법은 바이어스를 전혀 고려하지 않는 방법에 비하여 한 가지 장점이 있다. 즉, 나중에 외부 검토자가 자료의 바이어스에 대하여 의문을 제기하면 바이어스에 관련된 자료의 분석 결과를 제시 할 수 있다는 점이다.

3.6 작업 6 : 추출 작업의 연습과 조직 내 교육 (Practicing the elicitation and training the in-house personnel)

추출 작업에 최종적으로 들어가기 전에 추출 설계단계에서 작성된 모든 내용을 마지막으로 점검하고 연습과 파일럿 테스트 등을 통하여 실제 추출작업 과정에서 생길 수 있는 문제점들을 확인하는 것이 필요하다.

3.6.1 추출 작업의 연습

(1) 연습해야 할 내용

추출 작업을 수행하기 전에 반드시 연습해야 할 최소한의 내용은 다음과 같다.

a. 전문가에게 추출 과정을 소개하는 것

전문가에게 추출 과정을 소개해야 하는 것은 다음의 필요성 때문이다.

- o 전문가들을 앞으로 해야 할 단계와 일정에 친숙하게 만든다.
- o 그들이 접하게 될 문제들과 친숙하게 만든다.
- o 그들이 사용하게 될 응답모드와 친숙하게 만든다.
- o 작업 중 그들에 빠지기 쉬운 바이어스에 대해 인식하게 한다.

만약 전문가들이 추출 과정의 개발에 깊이 참여하지 않았다면 그들이 앞으로 무엇을 해야 하는 지에 대하여 간략하고도 명료한 브리핑이 반드시 필요하다.

b. 추출 절차

추출에 대한 계획이 잘 수립되었다 하더라도 몇몇 전문가를 선정하여 시험팀을 구

성해 시험해 봄으로서 많은 추가적인 내용을 확인할 수 있다. 예를 들면, 필요한 추출 시간이 예상보다 길어지거나 또는 전문가들이 피로해지는 것 등이다.

c. 문서화 절차

문서화에 대한 파일럿 테스트는 문서의 형식이나 관련 지원 문제를 어떻게 개정해야 하는지에 대한 자료를 제공해 준다. 예를 들면 인터뷰 담당자가 인터뷰를 수행하면서 동시에 문서화가 어렵다는 것 등을 보여줄 수 있다.

d. 전문가의 응답을 수학적으로 결합할 경우 이에 대한 연습

전문가의 응답들을 수학적으로 결합하는 것은 정형적 절차이기는 하지만 이를 위해 선택된 방법들 중에는 추출된 자료에 적용하기에는 적합하지 못한 것도 있을 수 있다. 이를 위해 선택된 방법들을 시험해 볼 필요가 있다.

e. 연습 결과로 얻어진 자료가 사용하고자 하는 모델에 적합한 지 검토

파일럿 테스트로부터 얻어진 자료를 가지고 사용하고자 하는 모델에 적용해 볼 필요가 있다. 종종 이 단계에서 전문가가 응답한 자료의 형태와 모델이 요구하는 자료의 형태에 불연속이 나타난다.

(2) 파일럿 테스트에 필요한 내용

파일럿 테스트가 필요한 절차는 절차 중에서 전문가가 참여하는 부분이다. 그러나 실제 문제에서는 파일럿 테스트가 시간과 노력을 많이 필요하기 때문에 일부분에 대하여서만(샘플링) 수행되는 것이 일반적 상황이다. 다음의 항목들은 파일럿 테스트의 대상으로 제안된 항목들이다.

a. 문제와 응답에 대한 전문가의 이해

문제가 모두 작성된 후 전문가에게 응답을 구하기 위하여 그 문제들이 제시되는 경우와 전문가가 문제의 개발에 참여한 경우의 두 가지 경우의 두 가지 경우에서 파일럿 테스트가 수행될 수 있다. 첫 번째 경우 수행되는 파일럿 테스트에서 얻어지는 정보는 전문가가 그 문제들을 어떻게 해석하는가에 대한 것이고 두 번째 경우 수행되는 파일럿 테스트에서 얻어지는 정보는 전문가들이 그들이 풀어야 할 기본 문제 영역을 잘 이해하고 있는가를 검토한 다음 문제들을 정련하기 위한 것들이다.

b. 전문가의 응답 모드 사용 상태

이것은 전문가들이 응답 모드의 사용 방법을 이해하고 있는가를 검토하기 위한 것이다. 전문가들의 응답은 논리적/수학적 기준에 따라 검토되는데, 예를 들면 모든

상호 배타적 사건들의 확률의 합은 1이 되어야 한다는 것 등이다.

c. 추출 절차에 대한 전문가의 이해

전문가가 선택된 추출 절차에 대하여 잘 이해하고 있는지 검토하기 위한 것이다. 예를 들면, “verbal protocol” 기법이 선택되었다면 전문가가 “think aloud”에 대하여 잘 알고 있는지 그리고 그것을 사용할 수 있는지를 점검해 보아야 한다.

d. 전문가의 특정 문서화 형식 사용 상태

일반적으로 전문가들은 요청 받은 것만큼 완전하게 그들의 판단에 대한 문서화를 제공하지 않는다. 파일럿 테스트를 통하여 전문가들이 문서화에 대한 지시 내용을 충분히 이해하고 있는지 점검하고 이를 피드백 해야 한다.

(3) 파일럿 테스트 방법

a. 전문가 표본 선택과 사람 수

파일럿 테스트를 위하여 전문가 표본을 선택할 때에는 추출 작업을 수행할 전문가들을 대표하는 표본을 선택하는 것이 중요하다. 예를 들면, 추출 작업에 참여할 전문가들을 정부기관, 연구소, 대학, 산업체에서 추렸다면 파일럿 테스트의 표본 전문가들도 이들 기관들에서 골고루 선택하여야 한다.

일반적으로 추출 방법을 개발에 참여한 전문가는 전문가 표본에 포함시키지 않는 것이 바람직하다. 또한 자문 전문가들도 파일럿 테스트를 위한 전문가 표본으로는 적합하지 않다. 이들은 추출 방법의 개발에 참여했기 때문에 새로운 관점으로 시험에 접근하기 어렵기 때문이다.

표본 전문가의 수는 추출 작업에 참여할 전문가의 수에 따라 달라지는데 일반적으로 5명 인내가 적합한데 이것은 추출 작업에 참여할 전문가의 수가 50을 넘는 경우가 거의 없기 때문이다. 이처럼 적은 수의 전문가로 파일럿 테스트를 수행하므로 테스트는 철저하게 수행해야 한다. 즉, 표본으로 선택된 전문가들로부터 가능하면 많은 정보를 얻어내어 이를 피드백 해야 한다.

파일럿 테스트에는 두 가지가 있다. 하나는 철저한 파일럿 테스트(intensive pilot test)이고 다른 하나는 한정된 파일럿 테스트(limited pilot test)이다. 철저한 파일럿 테스트는 Meyer(1986)에 의하여 개발되었는데 전문가의 이해(understanding)를 추적하기 위한 것이며 구조화되고 깊이 있는 인터뷰와 관찰이 요구된다. 한정된 파일럿 테스트는 철저한 파일럿 테스트가 수행되고 거기서 나온 문제들이 질문에 반영된 다음에 하는 것이 가장 바람직하다.

b. 파일럿 테스트 순서

파일럿 테스트의 순서는 일반적으로 다음과 같다.

- (i) 철저한 파일럿 테스트를 수행한다.
- (ii) 테스트에서 나온 결과들을 테스트 항목(질문과 절차 등)에 반영하여 갱신한다.
- (iii) 한정된 파일럿 테스트를 병행하여 추출 과정의 처음부터 끝까지를 연습한다.

3.6.2 조직 내부 (사내) 교육

사내 과제 관련자는 전문가의 판단 분야에 익숙하지 못한 경우가 많으며 따라서 추출 방법론에 대한 교육이 필요하다. 전문가의 판단 추출에 경험이 있는 사람일지라도 그들이 특정 방법론에 익숙하지 못하다면 교육이 필요하다. 교육은 그들에게 방향/방침을 제공하고 그들이 일에 익숙해 질 때까지 연습할 수 있는 기회를 제공한다.

교육이 필요한 분야는 연습이 필요한 분야와 동일하며 아래와 같다.

- o 전문가에게 추출 과정을 소개하는 것
- o 추출 절차
- o 문서화 절차
- o 전문가의 응답을 수학적으로 결합할 경우 이에 대한 연습
- o 연습 결과로 얻어진 자료가 사용하고자 하는 모델에 적합한 지 검토

교육으로 성취해야 할 숙련의 정도는 과제에서 요구하는 목표와 개인의 욕구에 따라 달라진다. 예를 들면, 전문가가 그들의 판단을 제공하기를 꺼려하는 경우에는 자료 수집자는 필요한 판단 자료를 얻기 위해 보다 높은 정도의 숙련성을 습득할 수 있는 교육이 필요하다.

(1) 교육이 필요한 시점

사내 참여자의 교육은 특히 다음의 경우에 필요하다.

- o 추출 작업을 수행할 사람과 계획한 사람이 다른 경우
- o 추출 작업과 문서화를 수행할 사람들이 그들의 작업에 불편함을 느낄 경우
- o 두 명 이상 또는 팀으로 전문가의 지식 추출 작업을 수행하고 문서화 할 경우
- o 전문가의 지식 추출 과정에 두 명 이상의 사내 인원이 참여하게 될 경우

(2) 교육 방법

교육이 필요한 5가지 아이টে에 대한 교육 방법은 다음과 같다.

o 전문가에게 추출 과정을 소개하는 것

소개 담당자가 한 명의 청중(계획된 추출 방법론에 익숙하지 않은 사람이 좋다) 앞에서 예행연습을 한다.

o 추출 절차 및 문서화 절차

이 두 가지는 다른 분야(a와 c 분야)에 비해 복잡하여 교육하기 어렵다. 일반적으로 권장되는 교육 방법은 다음의 세 가지이다.

(i) 교육 대상자에게 추출 절차에 대한 교육을 한 다음 그로 하여금 제한된 파일럿 테스트를 연습으로 해보게 하는 것이다. 많은 양의 파일럿 테스트 표본과 많은 시간을 필요로 하지 않는다는 장점이 있다.

(ii) 경험이 있는 자료 수집가가 제한된 파일럿 테스트를 먼저 수행하고 피 교육자는 그것을 관찰 한 다음 마지막 테스트를 피 교육자가 직접 수행하는 방법이다. 장점으로서는 피 교육자가 숙련된 추출 과정을 관찰하므로 서 많은 것을 배울 수 있다는 점이고 단점으로는 많은 수의 시험을 필요로 한다는 점이다.

(iii) 피 교육자가 한정된 파일럿 테스트나 그것의 비디오 테이프를 관찰한 다음 서로 역할(인터뷰 담당자/전문가)을 담당하여 연습을 하는 방법이다. 장점으로서는 파일럿 테스트 회수의 제한이 없고 또 전문가의 역할을 가상적으로 하는데서 추후 작업에 도움이 되는 여러 가지 이익을 얻을 수 있는 점이다.

o 전문가의 응답을 수학적으로 결합할 경우 이에 대한 연습 또는 연습 결과로 얻어진 자료가 사용하고자 하는 모델에 적합한 지 검토

제한된 파일럿 테스트로부터 얻어진 자료를 피 교육자에게 준 다음 그 자료들을 사용하여 작업을 수행하도록 한 다음 발생하는 문제점들에 대하여 보고를 하도록 하는 방법이다.

3.6.3 일반적으로 발생하는 문제점과 해결 방안

(1) 파일럿 테스트 결과 전문가가 응답모드에서 어려움을 느끼는 경우

파일럿 테스트를 수행하는 과정에서 전문가가 응답모드 부분에서 멈추거나 또는 혼란을 느끼거나 또는 이해할 수 없다는 코멘트가 나오면 응답 모드 부분에 문제가 있다고 볼 수 있다.

이럴 경우에는 전문가에게 응답 모드 부분에 대한 브리핑을 다시 하는 것을 고려해 보아야 한다. 이런 어려움이 발생하는 주원인은 응답 모드에 대하여 당초 실시한 브리핑이나 설명문서가 불충분했기 때문이다. 새로운 브리핑을 새로 표본 선출된 전문가에게 실시한 다음 철저한 파일럿 테스트를 다시 수행해 보는 것이 바람직하다. 만약 이렇게 했는데도 표본 선출된 전문가가 응답 모드에 어려움을 느낀다면 그 응답 모드의 사용을 재 고려해 보고 다른 응답 모드를 찾아보아야 한다.

(2) 파일럿 테스트 결과 추출작업에 필요한 시간이 당초 계획 시간보다 더 필요한 것으로 나타날 경우

추출을 위한 시간이 당초 할당했던 시간을 초과하는 것은 종종 일어나는 일이다. 이것은 정해진 시간 내에 사람들이 할 수 있는 일의 양을 과대 추정하거나 불확실성을 과소 추정하기 때문에 생긴다. 제한된 파일럿 테스트를 수행하므로써 이런 어려움을 발견할 수 있다.

이에 대한 대처 방안은 제한된 파일럿 테스트를 통하여 얻어진 자료(어떤 단계에서 어느 정도 시간이 소요되었는지 등)를 검토한 다음 소요되는 시간을 줄이기 위해서 단계별로 조치를 강구하고 다시 파일럿 테스트를 수행해서 그 소요 시간을 측정하는 것이다. 다른 방안으로는 일정을 재조정하는 것이 있는데 첫 번째 방법에 비하여 고려해야 할 사항이 많지 않으므로 현실적으로 많이 채용된다.

(3) 내부 관련자가 교육에 저항감을 가질 경우

사내 작업 관련자가 교육에 저항감을 느끼는 이유에는 다음과 같은 것들이 있다.

- 추출 작업을 그들 작업의 일부로 생각하지 않고 과외의 일로 생각한다.
- 추출 작업을 수행하기 위한 자격이 부족하다고 생각한다.
- 만약 추출작업이 실패할 경우 비난받을 것을 두려워한다.
- 그들이 계획에 참여하지 않은 일을 하는 것을 싫어한다.

각 개인들이 교육을 거부하는 이유들을 찾아 그 각각의 이유에 대한 해결을 시도해야 한다. 전체 중 극히 일부만이 이런 저항감을 보일 경우에는 그 소수 집단에 배한 관심(배려)이 문제를 해결할 수도 있다.

(4) 문서화 방식(구조)이 과제의 다른 목적에 적합하지 않을 경우

문서화 방식이 파일럿 테스트되지 않았다면 이런 문제점은 추출 작업이 다 끝난 다음에 종종 발견되고 작업 전체의 성과에 중대한 악영향을 미친다.

이런 문제가 파일럿 테스트 전 또는 테스트 결과로 발견되었다면 적합한 새로운 문

서화 방식을 개발하여 사용하는 것으로 해결된다. 그러나 이런 문제가 추출 작업이 끝난 다음에 발견되면 구해진 자료들을 적합한 형태로 변환하는 한정된 해결 방안 밖에 없는데 이런 변환작업에는 그 자료들을 잘못 해석하거나 또는 검증되지 않은 가정을 하거나 하는 문제점을 내포하고 있다.

(5) 전문가의 문제 해결에 대한 문서화가 다른 방식으로 되었거나 또는 상세 정도가 다를 경우

이런 문서의 일관성 결여는 복수의 과제 관련자나 전문가들이 문서화를 수행한 경우나 또는 서로 다른 정도의 복잡성을 가진 질문들이 문서화에 포함되었을 경우 종종 나타난다.

추출 연습 과정에서 이런 문제점이 발견되었다면 예행연습을 통하여 이를 해결할 수 있는데 문서의 일관성을 향상시키는 방안은 다음과 같은 것들이 있다.

- (i) 문서화 지침을 더욱 엄격하게 한다.
- (ii) 문서화에 대하여 추가적 교육을 실시한다.
- (iii) 자격 있는 사람을 지정하여 그로 하여금 모든 기록 자료를 검토하게 한다.

추출 작업이 모드 끝난 다음 이런 문제점이 발견되었다면(전형적으로 일어나는 경우이다) 이의 해결 방안은 다음의 두 가지로 한정된다고 볼 수 있다.

- (i) 갭(gap)을 채우기 위해 추출을 재구성하는 방법이다. 이것은 빠진 정보를 메우기 위해 추출된 모든 정보를 재처리하는 작업을 수반한다. 더구나 필요한 정보가 빠져있을 경우에는 전문가를 재 접촉하거나 또는 그들로 하여금 재 작업을 하도록 요청해야할 경우까지도 발생한다. 이 방법은 너무 어려워서 잘 시도되지 않는다.
- (ii) 두 번째 방법은 수집된 자료를 조절하는 것인데 이것은 자료를 보다 일반적 레벨로 만들거나 또는 자료의 정밀성을 저하하는 결과를 초래한다. 예를 들면, 어떤 전문가의 문제해결에 대한 노트는 아주 상세하고 다른 전문가의 노트는 그렇지 못하다면 덜 상세한 노트를 공통 분모로 사용해야 한다.

3.7 작업 7 : 추출 작업의 지휘/지도/관리 (Conducting the elicitation)

추출 작업을 위한 계획과 연습 그리고 파일럿 테스트가 끝나면 실제의 추출 작업이 수행된다. 이 단계는 일정 관리에 관련된 내용과 실제 작업의 수행으로 나누어진다.

3.7.1 일정 관리

일정 관리에서 가장 중요한 것은 선정된 전문가들과 좋은 관계를 유지하고 또 매끄럽게 작업이 진행되도록 하는 것이다. 추출 환경이 메일 조사, Delphi, 전화 인터뷰 등으로 정해진 경우에는 다음의 회합 계획 수립, 회합 일정 확인에 대한 내용은 생략 가능하다.

3.7.1.1 회합 계획 수립

개별 면접이나 interactive group 환경에서는 일반적으로 다음과 같은 단계가 필요하다.

- 자신의 소개와 과제에서의 담당 업무(역할)
- 현재 통화(또는 협의)의 가능여부 문의
- 인터뷰 미팅의 일정 계획을 위한 자신의 계획 설명
- 작업 시간이 얼마나 걸리는지에 대한 설명
- 전문가에게 그가 작업 가능한 일자와 시간을 문의
- 전문가가 제시한 일자, 시간, 장소에 대한 확인
- 미팅 약속 확인
- 감사 인사.

3.7.1.2 회합 일정의 확인

회합 일정의 확인은 일반적으로 다음과 같다.

- 회합 일정의 확인

자신을 소개하고 앞에서 정한 약속의 세부적 내용(일자, 시간, 장소, 작업 내용 등)에 대하여 확인. 확인은 약속 일자보다 2일전 정도가 좋으며 만약 전문가가 그 작업을 위하여 여행을 하여야 한다면 여행 출발 2일전 정도가 바람직하다.

- 전문가의 사정으로 당초의 약속 변경이 필요할 때.

한 전문가에게 사정이 발생하여 회합 일정을 변경해야 할 경우에는 다음의 사항들을 고려해야 한다.

- 일정 변경이 다른 전문가들에게 영향을 주는지 여부
- 회합이 비디오 녹음 가능한지 여부
- 회합 최소에 관련된 비용이 발생하는지 여부
- 다른 전문가와 접촉을 해야 할 경우 시간 내에 가능한지 여부

단지 한 전문가만이 회합 참석이 어려운 경우 그에게 회합 녹음 테이프를 보여 준

후 나중에 개별 면접 등을 할 수 있다.

복수의 전문가가 참석 불가능한 경우에는 일정을 다시 잡아야 한다.

3.7.2 추출 작업의 set up과 수행(지휘)

3.7.2.1 추출의 계획(setup) 지침

추출 작업의 setup은 필요한 자료를 준비하는 것과 회의실을 준비하는 것 등을 의미한다. 이런 준비 사항들은 추출 작업의 핵심적 요소는 아니지만 전문가들이 작업을 편하게 하고 그들을 보다 쾌적한 환경에서 일 할 수 있게 해 주며 이것은 작업을 효율성과 작업 결과의 품질에 영향을 미친다.

일반적으로 각 추출 환경에 따른 준비 사항들은 다음과 같다.

a. 개별 면접

- o 전문가의 이름
- o 과제에 대한 간략한 개요
- o 문서화 양식, 질문 주제, 그리고 기타 양식
- o 질문과 배경 정보에 대한 사본
- o 예비 필기도구
- o 작업에 필요한 기자재, 도구(녹음기, 컴퓨터, 전원 코드 등)

b. Delphi 환경

- o 질문서의 표지
- o 질문서 사본
- o 전문가들의 이름, 주소(메일 주소 포함), 전화 번호
- o (가능하다면) 전문가들의 명단 공개(전문가의 매니저나 또는 전자게시판 등에)

c. Interactive group situation

- o 그룹 중재자/인터뷰 담당자에게 필요한 자료 준비
- o 녹음 도구
- o 전문가들의 명패
- o 작업 일정
- o 전문가들이 사용할 배경 정보, 질문서, 문서화 양식의 사본
- o 휴식 시간 및 필요 용품들
- o 회합이 하루 이상 걸리는 경우에는 숙박에 관련된 자료와 여행 등 전문가들의

편의와 휴식/오락을 위한 자료들

- o 예비 필기도구

3.7.2.2 추출의 수행 지침

일반적으로 추출 작업은 다음의 단계로 나누어진다.

- o 전문가에게 추출 작업을 소개
- o 자료를 수집하고 기록
- o 추출 작업 중 바이어스의 감시

그러나 실제로 추출 작업을 수행할 때는 위의 일반적 단계를 참조하는 것 뿐 아니라 전 단계의 연습과 파일럿 테스트에서 얻어진 결과들을 충분히 활용하는 것이 바람직하다.

(1) 전문가에게 추출 작업을 소개

a. 개별 면접(individual interview)

- o 인터뷰 담당자의 소개
- o 전문가의 직업적 배경에 대한 질문부터 시작
- o 전문가에게 표본 질문 제시
- o 전문가에게 바이어스에 대한 브리핑
- o 전문가에게 한 세트의 질문을 준 다음 구두로 반복해보도록 함
- o 전문가에게 작업의 시작을 알림

b. Interactive group situation

- o 관련 자료의 배포
- o 회의 조정자/인터뷰 담당자, 과제 참여자, 참석 전문가들의 소개
- o 과제의 목표, 일정, 그리고 일반적 추출 과정에 대한 소개
- o 전문가에게 표본 질문을 주고 연습
- o 전문가에게 바이어스에 대한 브리핑
- o 전문가의 의문사항에 대한 응답 및 작업의 시작

c. Delphi situation

Delphi 방법 전 과정이 메일에 의해서 진행된다면 조정자 또는 인터뷰 담당자의 소개는 필요가 없고 관련 자료와 간단한 커버 레터를 배포하면 된다. 관련 자료의 내용은 개별 면접에서 기술된 내용과 유사하다.

(2) 전문가의 판단 자료를 수집하고 기록

a. 개별 면접, group interactive, Delphi 방법 사용

일반적으로 이 단계는 연습과 파일럿 테스트에서 나온 결과들을 참조하여 수행되어야 한다. 만약 파일럿 테스트가 수행되지 않았다면 뒤의 “일반적으로 발생하는 문제점과 해결 방안”을 참조하여 발생하는 문제점들을 해결해야 한다.

b. 문제 해결 자료의 추출을 위한 세 가지 기법의 사용

문제 해결 과정에 대한 자료를 얻기 위해서는 verbal protocol, verbal probe, ethnographic 기법이 개별 면접과 병행하여 자주 사용된다. 그리고 verbal protocol 과 ethnographic 기법은 전문가들이 문제를 어떻게 풀었는가에 대한 간략한 정보를 얻기 위하여 interactive group situation에서 사용된다.

o Verbal protocol

Verbal protocol 기법을 사용 시 고려해야 할 사항들은 다음과 같다.

- 인터뷰 담당자의 위치는 전문가의 바로 뒤 좌측(전문가가 오른손으로 필기를 할 경우)
- 전문가가 평소 작업하는 환경과 가능한 유사하도록 문제를 설계
- 전문가가 문제를 해결하기 위해 사용할 그래프, 테이블, 공식 등의 사본 준비
- 전문가나 기록 도구에만 의존하지 말고 노트를 할 것
- 어떻게 문제를 푸느냐에 대하여 가정적으로 이야기하는 것이 아니라 문제의 work through를 해야 한다는 것을 전문가에게 강조 할 것
- 전문가에게 문제 해결을 요청할 때 “thinking aloud”에 대한 중요성을 강조할 것.
- 전문가들은 종종 “thinking aloud” 하는 것을 멈추는 경향이 있다. 이 경우 그에 대한 즉각적 조언을 할 것. 조언은 동일한 형태를 반복하지 말고 다양한 형태를 사용할 것

o Verbal probe

Verbal probe 기법 사용 시 고려해야 할 사항들은 다음과 같다.

- 개별 면접에서 verbal probe를 사용할 경우에는 verbal protocol과 동일한 set up을 사용한다.
- 그룹 미팅에서 복수의 전문가들에게 이 기법을 사용할 경우에는 그룹 중재자가 미리 전문가들에게 그들의 문제 해결 과정에 대한 이유를 제시해 달라고

요청한다. 여기서 주의할 점은 일반적으로 사람들은 그들의 추론을 문서화 형식으로는 잘 표현하지 못하기 때문에 반드시 말로 설명하도록 요청해야 한다는 점이다.

- Verbal probe는 과거시제로 표현 할 것.
- 중복 반응에 대한 점검을 할 것

o Ethnographic 기법

Ethnographic 기법을 사용 시 고려해야 할 사항들은 다음과 같다.

- Ethnographic 기법이 어떻게 설정되었나 하는 것은 주로 같이 사용된 다른 기법이나 추출 환경에 의하여 결정되므로 같이 사용된 기법이나 환경의 set up을 따른다.
- Ethnographic 질문의 길이는 서로 다르게 해서 전문가가 지루해지지 않게 해야 한다.
- 이 기법이 전문가의 문제 해결 과정에 대한 자료를 얻기 위해 사용되었다면 질문 시기는 전문가가 문제 해결을 마친 시점으로 계획해야 한다.

(3) 추출 작업 중 발생하는 바이어스의 감시와 조정

전문가의 판단 자료를 수집하고 기록하는 과정에서 바이어스가 개입되기 쉽다. 바이어스를 처리하는 단계는 일반적으로 다음과 같다.

- a. 추출 계획 단계에서 바이어스에 대한 발생 예기 및 대처 방안 수립
- b. 전문가로 하여금 바이어스 개입의 잠재성을 인식하게 하고 추출 절차에 친숙하도록 함.
- c. 바이어스 발생의 감시
- d. 바이어스 발생 시 실시간 조정
- e. 바이어스가 발생한 자료에 대하여 분석 수행

(a) 단계는 본 지침서의 “추출 계획의 수립” 부분에 기술되어 있고 b 단계는 본 장의 앞부분에 기술되어 있다. 그리고 e 단계는 3.7.3의 (4)항에 기술된 분석 방법들이 주로 사용되며 추출 작업 진행 시 해당되는 c와 d 단계는 다음과 같다.

c 단계: 바이어스 발생의 감시

추출 작업 진행 중 바이어스들이 발생할 때 그것의 발생을 알리는 신호가 있다. 숙련된 인터뷰 담당자나 관찰자들은 이런 신호를 감지할 수 있다. 일반적으로 바이어스의 감시에는 전문가가 그의 생각이나 응답을 말로 표현하는 것을 필요로 하는데

그렇지 않은 경우 바이어스의 감시가 매우 어렵게 된다.

d 단계: 바이어스 발생 시 실시간 조정

이 단계는 바이어스를 관리하는데 있어서 가장 어려운 단계인데 잘못 수행할 경우 결과나 추후의 분석을 혼란스럽게 만들 소지가 많기 때문이다. 인터뷰 담당자는 사전에 조정 시점을 결정해야 하는데 이런 조정이 자료가 수집되는 조건을 변경시키기 때문이다.

만약 분석할 자료가 수집되기 전에 바이어스 발생 가능 조건이 수정되었다면 문제가 없다. 그러나 바이어스가 발생하는 도중 단계와 그것이 수정된 단계의 두 가지 단계에서 자료가 수집되어 그것이 합쳐졌다면 이 자료는 바이어스의 존재를 시험할 수 없게 된다.

3.7.3 일반적으로 발생하는 문제점과 해결 방안

(1) 전문가가 추출 과정에 거부감을 나타내거나 또는 불확실성 하에서의 판단을 거부할 경우

불확실성 관점에서 사고(생각)하는데 익숙하지 않은 전문가들이 포함된 interactive group 상황에서 이런 경우가 종종 발생하는데 이것의 징후는 전문가들이 그들의 판단 제시를 꺼리는 것이다. 그들은 자기들끼리 수군거리거나 또는 추출 절차를 공개적으로 비판하고 그들의 자료 제공을 거절하거나 한다. 이렇게 그의 판단 제시를 꺼리는 전문가에게 의견을 강요해서는 안 되는데 그 이유는 (i) 그가 그 분야의 전문가가 아닐지도 모르고 (ii) 그에게 의견을 강요하면 그는 부정적으로 될 뿐만 아니라 과제 전체에 대한 그의 관점에 부정적 영향을 미치게 되기 때문이다.

이 문제에 대한 해결 방안은 전문가가 의견 제시를 거부하는 이유를 밝히는데 있다. 이런 경향을 보이는 전문가와 개별적으로 접촉하여 (유쾌한 방법으로) 왜 그가 판단 제시를 거부하는 지 그 이유를 물어보아야 한다. 이때 반드시 개별적으로 접촉해야 하는데 그 이유는 그룹 상황에서는 그가 자신의 실제 이유를 밝히지 않을 가능성이 높고 또 그의 저항감을 더 강하게 만들 소지가 있기 때문이다.

일반적으로 이 상황에서 발생하는 전문가들의 저항 이유로는 다음과 같은 것들이 있다.

- o 전문가들이 자신이 개발했던 추출 과정에 대한 의심 또는 불안이 있을 때
- o 전문가들이 그들의 자료 사용 용도를 신뢰하지 않을 때
- o 문제에 관련된 새로운 자료가 있는데 그 자료를 자신들의 판단 결정에 사용

하기에 곤란하다고 생각할 때.

- o 그들의 자료가 잘못 해석되거나 또는 범위에서 벗어나거나 또는 부당하게 비평받을지도 모른다는 두려움이 있을 때
- o 그 주제에 관하여 주어진 높은 불확실성 하에서는 신뢰할 만한 추정치를 제공하지 못한다고 생각할 때

이런 어려움을 해결하는 수단은 그 어려움에 잠재한 원인을 찾는 데 있다.

a의 경우에는 그들에게 질문의 재 정련을 허용하고 또 현재의 추출 방법에 이르게 된 설계상의 고려사항과 제한사항들을 설명한다. 이렇게 함으로서 추출 방법이 잘못되지 않았다는 것을 그들에게 납득시키는데 도움이 되고 또 그들의 요구를 모두 받아들여야 하는 곤란한 경우를 피할 수 있다. 예를 들면 그들이 요구하는 변경사항에 대하여 그것이 과제의 요건에 맞지 않는다는 것 등을 지적해 낼 수 있다.

b의 경우에는 과제 보고서의 개요를 보여주거나 또는 그들 판단을 최종적으로 사용할 사람이나 조직에 대한 브리핑을 한다.

c의 경우에는 추출 작업 전에 전문가들이 그 자료를 공람 할 수 있게 하고 또 그들의 판단 자료는 잠정적인 것이고 언제든지 변경할 수 있으며 어느 합의된 시점에서 최종적 자료를 작성한다는 것을 설명한다.

d의 경우에는 전문가들에게 추출 작업의 완료 후 그들의 판단에 대한 문서화 기록을 검토할 수 있다는 것을 설명해 주고 또 그들의 판단이 어떤 방식으로 보고서나 모델에 표현되는지에 대하여 개요를 보여준다.

e의 경우에는 그 분야의 높은 불확실성이 바로 그들 전문가의 판단을 구해야 하는 이유 중의 하나라고 설명하고 추출 연습(과정)이 어떻게 그들의 판단 자료에 대한 신뢰성을 높이는데 도움을 줄 것인지를 요약해 제시한다.

이들 방법이 모두 실패하면 추출 과정의 많은 부분(가능하다면)에서 전문가와 일대일로 작업하는 것을 고려해야 한다. 예를 들면 그룹 환경보다는 개별 면접 등을 통해 전문가의 자료를 구하는 것인데 이렇게 하면 한번에 한 명의 비협조적인 전문가와 논의를 할 수 있고 그들의 개별적 요구 사항에 대하여 보다 융통성 있게 대처하는 것이 가능하다.

(2) 전문가들이 그들이 무엇을 해야 하고 또 어떻게 그것을 해야 하는가에 대하여 혼란을 느끼는 경우

전문가들이 추출 절차를 따라 작업할 때 그 속도가 매우 느리거나 또는 당황함을

보이는 경우가 있다. 이런 현상은 전문가들에게 표본 질문들과 주어진 응답모드로 작업을 하게 했을 때 확인 가능하다. 그리고 철저한 파일럿 테스트를 수행하였다면 그로부터 추출의 어떤 부분이 혼란을 야기하는가에 대한 경고를 얻을 수 있다.

이를 해결하기 위한 첫 단계는 전문가의 그런 반응의 원인을 확인하는 것이다. 그 원인은 추출을 위한 일반 절차로부터 생겨날 수도 있고, 그들이 사용하는 응답모드로부터 생길 수도 있으며 또는 어떤 지시사항에 대한 혼란일 수도 있다. 이런 전문가의 혼란에 관련된 특수한 문제 중의 하나는 종종 그들이 무엇에 대하여 혼란을 느끼는 것인지 적절하게 표현 할 수 없다는 점이다. 혼란의 원인을 확인하는 방법 중 하나는 제시된 정보를 검토해서 전문가에게 어느 부분이 불명확한지 지적하도록 요청하는 것이다. 그리고 그것이 전문가의 이해 결여에서 생긴 혼란이 아니라 표현에 있어서 명쾌함이 결여되어 생긴 문제라는 것을 전문가에게 전달해 주는 것이다.

추출 작업 중 전문가에게 혼란이 발생하면 그것이 해결될 때까지 전문가와 함께 작업하는 것이 필요하다. 그렇게 하지 않으면 심각한 결과를 초래하기 때문이다. 예를 들면 전문가가 혼란을 느끼는데도 불구하고 추출 작업을 강행한 경우 그 전문가는 추후에 추출 방법을 비평했고 또 그 자료의 정당성에 대하여 의문을 제기한 경우가 있다.

(3) 전문가가 최종 응답이 문서화되지 않은 경우

대부분의 작업에서 공통적으로 발생하는 이 어려움의 주원인은 추출이 점진적으로 진행되기 때문이다. 질문의 어법은 그룹 작업 중 점진적으로 개량되고 언제 최종적 질문이 완성될 지는 미리 알기 어렵다. 따라서 질문의 최종 형태나 그에 대한 최종 해법을 기록하지 못한 사건은 그 당시에는 종종 발견되지 못한다. 이런 문제점을 해결하기 위해서는 작업에 관련된 사람들이 그들의 기억이 생생할 때 기록들을 검토해야 한다. 그리고 인터뷰 담당자는 그룹이나 개별 전문가에게 쉽게 고칠 수 있을 때 최종 형태에 대한 검토를 요청하는 것이 바람직하다.

추출 작업 후 바로 이런 문제점이 발견되었다면 그룹이나 전문가 그리고 인터뷰 담당자가 공동으로 기억을 되살려 그들의 문서를 갱신할 수 있다. 만약 추출 작업 후 시간이 많이 지나서 이런 문제점이 발견되면 인터뷰 담당자는 녹음 기록을 활용하여 기록을 갱신해야 한다. 나중의 방안은 시간이 매우 많이 소요되고 또 기억이 분명하지 않은 경우가 있으므로 매우 어렵다.

(4) 추출 과정 중 바이어스가 발생하였을지도 모르나 그것이 감시되지 않은 경우

바이어스는 추출의 설계 단계나 수행 단계에서는 일반적으로 고려되지 않는 경우가

많다. 따라서 대부분의 경우 바이어스의 가능성은 추출의 마지막 단계에서 고려되는데 그것은 추출 결과에 대해 바이어스를 제기하는 일이 자주 일어나기 때문이다. 바이어스에 대한 의문 제기가 이처럼 나중에 발생하면 그것을 감지하기 위한 방법은 초기에 바이어스를 고려했을 때에 비하여 매우 제한적이다. 그러나 추출 작업에서 충분한 자료가 구해졌다면 특정 바이어스에 대하여는 그것의 존재를 시험하는 것은 가능하다.

전문가의 자료에 존재하는 바이어스를 분석하는 방법에는 다음과 같은 것들이 있다.

- o 입도(granularity)를 사용하는 방법
- o 상관성의 가설 근원(hypothesized sources of correlation)을 사용하는 방법
- o 상관성 분석을 사용하는 방법
- o 다변량 분석(multivariate analysis)을 사용하는 방법
- o 모의기법을 사용하는 방법
- o 추출기법을 사용하는 방법
- o 가정을 사용하는 방법

과제의 외부 관계자들이 가장 염려하는 바이어스는 일반적으로 다음의 세 가지이다.

- (i) 인터뷰 담당자가 전문가의 의견을 의도적으로 유도하므로서 발생한 바이어스
- (ii) 그룹 사고 환경에서 한 전문가가 다른 전문가를 어떤 방향으로 유도하므로서 발생한 바이어스
- (iii) 전문가 사이에서 이해의 상충이나 또는 전문가의 판단에 영향을 미치는 바람이나 이해로부터 발생한 바이어스

일반적 원칙은 모든 경우에서 추출된 자료에 대한 바이어스 분석을 수행하고 그에 대한 결과를 과제의 보고서에 포함시키는 것이다.

(5) 전문가들의 판단 자료 사이에 차이가 있을 경우

전문가들의 판단 자료 사이에 차이가 있다는 사실은 종종 문제점으로 인식되고 있지만 실제로는 문제가 아닐 수도 있다. 일부의 사람들은 전문가들 사이의 의견 불일치를 추출 과정의 실패 표시로 보는데, 이는 추출 방법론이 하나의 올바른 답을 만들어 내지 못했거나 또는 문제를 해결하는 수단을 제공하지 못했다고 생각하기 때문이다. 그러나 전문가들이 문제를 서로 다른 방법으로 풀고 또 그들이 문제를 해결하는 방법이 그들의 해답에 영향을 미치는 것은 비합리적인 것이 아니다. 다만

전문가들 사이의 상이한 응답이 문제가 되는 경우는 그 자료들이 지식기반 시스템 용 프로그램에 사용되거나 또는 수학적으로 결합해야 하는 경우처럼 반드시 결합되어야 할 때다.

전문가들의 응답 자료 사이에서 차이가 발생하면 먼저 그것이 실제로 문제인지 아닌지를 검토해야 한다. 즉 그 상이점이 (i) 추출 과정의 취약성에서 기인한 것인지 아니면 (ii) 전문가의 전문성이 가지는 고유 특성인지를 확인해야 한다. 이를 확인하기 위해서는 전문가가 문제를 해결한 과정에 대한 자료가 필요하다.

만약 전문가에게 제시된 문제의 정의나 가정이 엄밀하지 못하게 되어있었다면 전문가들은 그 문제를 명확하게 하기 위해 자신들이 만든 정의나 가정을 사용하였을 것이고 이것은 문제를 본래와는 다른 것으로 만든다. 그 결과로 전문가들은 서로 다른 해결 방법을 사용하게 되고 따라서 서로 다른 답을 제공하게 된다. 이 경우에는 전문가 응답의 차이가 문제가 된다. 한편 전문가들이 그들이 문제를 해결할 때 사용한 가정이나 정의가 동일하다면 응답의 차이는 전문가들이 문제를 해결하는 방법이 다른 것에서 기인한 것으로 볼 수 있고 이것은 앞서 예시한 경우를 제외하고는 실질적인 문제로 볼 수 없다.

(i)의 경우가 발생하면 그에 대한 효과적 대처 방안은 없으나 한 가지 방안은 본래와는 달라진 질문에 대한 응답을 별도로 추려내는 것이다. 그리고 다른 방안은 다른 가정이나 정의를 사용한 전문가와 재 접촉해서 합의된 정의와 가정을 사용한 새로운 응답을 요청하는 것이다.

(ii)의 경우에는 일반적으로 문제가 되지 않으나 외부의 관련자들이 전문가 응답의 차이점을 부정적으로 보는 문제가 있을 수 있다. 이 때는 이런 전문가 응답의 차이점이 전문가의 응답에 내재한 불확실성과 다양성에 기인한 당연한 현상이라는 것을 외부 관련자들에게 납득시켜야 한다. 그리고 그들에게 이들 차이점이 수행된 추출 방법론에 기인한 것이 아니라는 것을 보여주는 증거들을 제공해야 한다.

4. 디지털 시스템의 신뢰도/안전성 평가용 BBN 모델에 사용된 전문가의 확률 값 설정 사례

4.1 BBN for M-ADS system and DO-178B Standard

이 BBN은 소프트웨어 안전 표준의 요건인 DO-178B[11]을 BBN으로 변환하는 가능성을 조사하고 소프트웨어의 안전성 측정 방법들을 확률 값으로 변환해서 PSA에 사용될 수 있는지 가능성을 검토하기 위해서 만들어졌다. 이 연구는 OECD HALDEN reactor project의 일부로 수행되었으며 상세 BBN 그래프와 계산 및 분석 결과는 “Estimating dependability of programmable systems using bayesian belief nets”[12]에 상세하게 기술되어 있다.

작성된 BBN의 상위 레벨 그래프는 다음 그림 1과 같다.

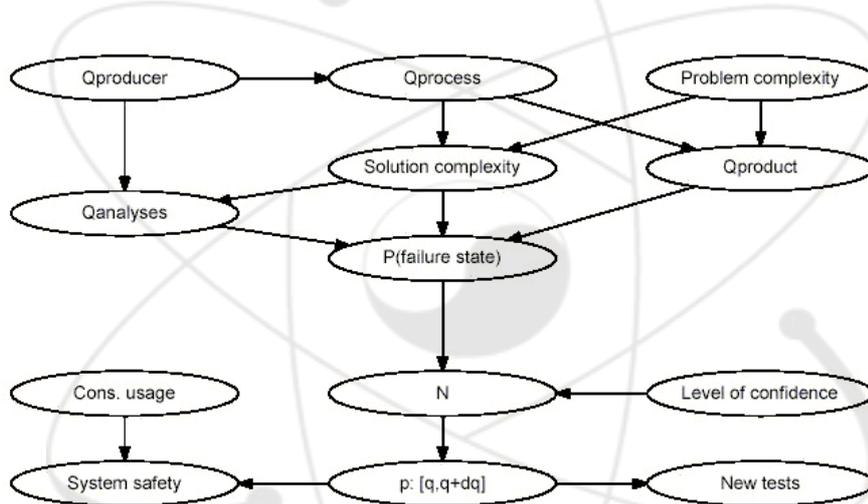


그림 1 M-ADS 신뢰도 평가용 상위레벨 BBN 그래프

노드확률테이블은 소프트웨어의 신뢰도 평가, DO-178B 표준, 그리고 대상 시스템에 대하여 전문성을 가진 전문가들에 의하여 작성되었지만, 전문가에 대한 교정(calibration)이나 평가는 이루어지지 않았다.

(1) 노드 변수 상태(states)

Node variables	States of variables		
Quality-control-documents	none	partly	completed
QA-standards	none	generic	detailed
Failure-in-other-products	>50%	10% - 50%	<10%
Number-of-products	<10	10 - 100	>100
Usage-time	100 h	100 -10000 h	>10000 h
Documentation	bad	acceptable	excellent
Quality-measures	<0.1	0.1 - 0.8	>0.8
Quality-control	strict	lousy	
QA-policy	bad	acceptable	excellent
Producer's-pedigree	low	medium	high
System-quality	low	medium	high
Development-quality	low	medium	high
User-experience	low	medium	high
Producer-quality	low	medium	high

(2) 각 노드 별 prior 값

Producer Quality	
Low	0.25
Medium	0.5
High	0.25

User Experience	
Low	0.3
Medium	0.4
High	0.3

Producer Pedigree given the Producer Quality			
	Producer Quality		
	low	medium	high
low	0.9	0.2	0.0910
medium	0.1	0.6	0.1818
high	0.0	0.2	0.7272

Development Quality given the Producer Quality			
	Producer Quality		
	low	medium	high
low	0.9	0.2	0.0910
medium	0.1	0.6	0.1818
high	0.0	0.2	0.7272

QA policy given the Producer Quality			
	Producer Quality		
	low	medium	high
low	0.9	0.2	0.0910
medium	0.1	0.6	0.1818
high	0.0	0.2	0.7272

System Quality given the Producer Quality and the Development Quality									
	Producer Quality								
	low			medium			high		
	Development Quality								
	low	medium	high	low	medium	high	low	medium	high
low	1.0	0.9	0.7	0.9	0.2	0.1	0.7	0.2	0.0
medium	0.0	0.1	0.2	0.1	0.6	0.2	0.2	0.6	0.1
high	0.0	0.0	0.1	0.0	0.2	0.7	0.1	0.2	0.9

Usage Time given the User experience			
	User Experience		
	low	medium	high
<100 h	0.8	0.1	0.1
100-10000h	0.1	0.8	0.1
<10000 h	0.1	0.1	0.8

Number of Products given the User experience			
	User Experience		
	low	medium	high
<100 h	0.8	0.1	0.1
100-10000h	0.1	0.8	0.1
<10000 h	0.1	0.1	0.8

Failures in other Products given the Producer's Pedigree and the Use Experience									
	Producer's Pedigree								
	low			medium			high		
	User Experience								
	low	medium	high	low	medium	high	low	medium	high
low	0.3	0.2	0.0	0.9	0.2	0.1	0.9	0.2	0.0
medium	0.4	0.2	0.1	0.1	0.6	0.2	0.1	0.6	0.1
high	0.3	0.6	0.9	0.0	0.2	0.7	0.0	0.2	0.9

Quality measures given the System Quality			
	System Quality		
	low	medium	high
< 0.1	0.8	0.1	0.1
0.1 - 0.8	0.1	0.8	0.1
> 0.8	0.1	0.1	0.8

Documentation given the Development Quality			
	Development Quality		
	low	medium	high
bad	0.9	0.2	0.0910
acceptable	0.1	0.6	0.1818
excellent	0.0	0.2	0.7272

Quality Control given the QA Policy			
	QA Policy		
	low	medium	high
strict	0.1	0.5	0.8
lousy	0.9	0.5	0.2

Quality Control Document given the Quality control		
	Quality Control	
	low	lousy
none	0.0	0.5
partly	0.2	0.5
completed	0.8	0.5

QA Standards given the QA Policy			
	QA Policy		
	bad	acceptable	excellent
none	0.5	0.3	0.1
generic	0.3	0.5	0.2
detailed	0.2	0.2	0.7

Quality of producer (Quality_Producer)(Labelled Chance Node)

5good	.06
4	.44
3	.45
2	.05
1bad	.01

Problem complexity (Problem_complexity)(Labelled Chance Node)

low	.167
medium	.5
high	.333

Consequences and Usage (Conce_Usage) (Labelled Chance Node)

High 1	1
--------	---

Level of confidence (c) (Interval Chance Node)

<0, .9]	.167
<.9, .92]	.167
<.92, .94]	.167
<.94, .96]	.167
<.96, .98]	.167
<.98, 1]	.167

(4)The higher-level network

Quality of producer	5 good	4	3	2	1 bad
Quality of process	good	.9	.8	.5	.1
	bad	.1	.2	.5	.9

Problem complexity	low		medium		high	
Quality of process	good	bad	good	bad	good	bad
Quality of product	good	.9	.3	.6	.4	.3
	bad	.1	.7	.4	.6	.7

Solution complexity	low					medium					high				
Quality of producer	good	4	3	2	bad	good	4	3	2	bad	good	4	3	2	bad
Quality of analyses	good	.99	.95	.9	.7	.3	.95	.9	.7	.3	.1	.8	.7	.5	.3
	bad	.01	.05	.1	.3	.7	.05	.1	.3	.7	.9	.2	.3	.5	.7

Problem complexity	low		medium		high	
Quality of process	good	bad	good	bad	good	bad
Solution complexity	low	.9	.7	.1	.1	.01
	medium	.1	.2	.8	.6	.19
	high	0	.1	.1	.3	.8

P(system in failure state) (Failure_rate) (Interval Chance Node)

Quality of product	good			bad		
Solution complexity	low	medium	high	low	medium	high

Quality of analyses	good	bad										
<0, .000001]	01	01	0	0	0	0	0	0	0	0	0	0
<.000001, .00001]	.7	.5	01	01	0	0	0	0	0	0	0	0
<.00001, .0001]	25	.4	.7	.5	01	01	01	01	0	0	0	0
<.0001, .001]	.04	.09	.25	.4	.7	.5	.7	.5	01	01	0	0
<.001, .01]	0	0	04	09	25	.4	.25	.4	.7	.5	.1	0
<.01, .1]	0	0	0	0	04	09	04	09	25	.4	.8	.1
<.1, 1]	0	0	0	0	0	0	0	0	04	.09	.1	.9

N (N) (Interval Chance Node)

Expression $\max(0, \log(1 - c) / \log(1 - \text{Failure_rate}) - 1)$
--

N has the states:

- [0, 10]
- <10, 100]
- <100, 1000]
- <1000, 10000]
- <10000, 100000]
- <100000, 1000000]
- <1000000, Inf>

True failure rate p: <q, q + dq] (true_failure_rate) (Interval Chance Node)

Expression $\text{Beta}(1, N + 2)$

The node has the states:

- [0, .000001]
- <.000001, .00001]
- <.00001, .0001]
- <.0001, .001]
- <.001, .01]
- <.01, .1]
- <.1, 1]

New failure free tests (New_tests) (Interval Chance Node)

Expression $\text{Exponential}(1 / \text{true_failure_rate})$

The node has the states:

- [0, 10]
- <10, 100]
- <100, 1000]
- <1000, 10000]
- <10000, 100000]
- <100000, 1000000]
- <1000000, Inf>

System safety (System_safety) (Labelled Chance Node)

Consequences and Usage	High 1						
True failure rate p : <q, q + dq]	[0,.000001]	<.000001, .000001]	<.000001, .0001]	<.0001, .001]	<.001, .01]	<.01, .1]	<.1, 1]
lev1	1	1	1	1	1	1	1

(5) The lower-level network, the quality of the producer

Quality of producer	5 good	4	3	2	1	bad
sw planning	good	.95	.9	.5	.1	.05
	bad	.05	.1	.5	.9	.95

Qproducer_A1 - sw planning		good	bad
sw req. standards	yes	.8	.2
	no	.2	.8

Qproducer_A1 - sw planning		good	bad
sw design standards	yes	.9	.1
	no	.1	.9

Qproducer_A1 - sw planning		good	bad
sw code standards	yes	.95	.05
	no	.05	.95

Quality of producer	5 good	4	3	2	1 bad
sw development process	good	.99	.95	.5	.05
	bad	.01	.05	.5	.95

Qproducer_A2 - sw development process		good	bad
Are other products made by the producer reliable and of high quality	yes	.95	.1
	no	.05	.9

Qproducer_A2 - sw development process		good	bad
producer experience	yes	.9	.2
	no	.1	.8

Quality of producer	5 good	4	3	2	1 bad
---------------------	--------	---	---	---	-------

vv of sw req. process	good	.9	.8	.5	.2	.1	
	bad	.1	.2	.5	.8	.9	
Quality of producer		5 good	4	3	2	1 bad	
vv of sw design process	good	.9	.8	.5	.2	.1	
	bad	.1	.2	.5	.8	.9	
Quality of producer		5 good	4	3	2	1 bad	
vv of sw coding & integration process	good	.95	.9	.5	.1	.05	
	bad	.05	.1	.5	.9	.95	
Quality of producer		5 good	4	3	2	1 bad	
vv of verification process results	good	.99	.95	.5	.05	.01	
	bad	.01	.05	.5	.95	.99	
Qproducer_A7 - vv of verification process results					good	bad	
Has the producer sufficient and qualified resources for V&V?					yes	.9	.1
					no	.1	.9
Qproducer_A7 - vv of verification process results					good	bad	
Has the producer an acceptable V&V philosophy?					yes	.9	.3
					no	.1	.7
Qproducer_A7 - vv of verification process results					good	bad	
Has the producer long practice in software V&V activities?					yes	.9	.3
					no	.1	.7
Quality of producer		5 good	4	3	2	1 bad	
sw CMP	good	.99	.95	.5	.2	.1	
	bad	.01	.05	.5	.8	.9	
Quality of producer		5 good	4	3	2	1 bad	
SQA Process	good	.99	.95	.5	.2	.1	
	bad	.01	.05	.5	.8	.9	
Quality of producer		5 good	4	3	2	1 bad	
certification liaison process	good	.9	.8	.5	.2	.1	
	bad	.1	.2	.5	.8	.9	

(6) The lower-level network, the quality of the process

Quality of process				good	bad	
sw planning				good	.95	.3
				bad	.05	.7
Qprocess_A1 - sw planning				good	bad	
Is the sw development plan complete?				yes	.7	.3
				no	.3	.7
Qprocess_A1 - sw planning				good	bad	

Is the sw verification plan complete?	yes	.8	.2
	no	.2	.8
Qprocess_A1 - sw planning		good	bad
Are there acceptable configuration management plans for all phases?	yes	.9	.1
	no	.1	.9
Qprocess_A1 - sw planning		good	bad
SQA-Records	yes	.6	.4
	no	.4	.6
Quality of process		good	bad
sw development process	good	.99	.1
	bad	.01	.9
Qprocess_A2 - sw development process		good	bad
Are all system functional requirements, safety requirements and auxiliary requirements specified in the software specification?	yes	.9	.05
	no	.1	.95
Qprocess_A2 - sw development process		good	bad
design description	yes	.8	.1
	no	.2	.9
Quality of process		good	bad
vv of sw req. process	good	.9	.1
	bad	.1	.9
Qprocess_A3 - vv of sw req. process		good	bad
Are the high level requirement specifications correct unique and consistent?	yes	.99	.2
	no	.01	.8
Qprocess_A3 - vv of sw req. process		good	bad
Is tracability between the high level requirement and the final product facilitated?	yes	.9	.1
	no	.1	.9
Quality of process		good	bad
vv of sw design process	good	.9	.3
	bad	.1	.7
Qprocess_A4- vv of sw design process		good	bad
Are the design documents a correct and consistent refinement of the high level requirements?	yes	.99	.2
	no	.01	.8
Qprocess_A4- vv of sw design process		good	bad
Is it a clear correspondence between each item of the high level requirements and corresponding elements of the design?	yes	.9	.1
	no	.1	.9
Qprocess_A4- vv of sw design process		good	bad
Has an analysis shown that the software architecture does not contain any internal inconsistencies?	yes	.95	.1
	no	.05	.9

Quality of process		good	bad
vv of sw coding & integration process	good	.8	.2
	bad	.2	.8
Qprocess_A5- vv of sw coding & integration process		good	bad
Is there a clear correspondence between each element of the design and corresponding code modules?	yes	.99	.2
	no	.01	.8
Qprocess_A5- vv of sw coding & integration process		good	bad
Is the code a correct and consistent refinement of the low-level requirements given in design documents?	yes	.9	.1
	no	.1	.9
Quality of process		good	bad
testing integration process	good	.99	.2
	bad	.01	.8
Qprocess_A6 - testing integration process		good	bad
sw verification cases and procedures	yes	.9	.1
	no	.1	.9
Qprocess_A6 - testing integration process		good	bad
sw verification results	yes	.99	.2
	no	.01	.8
Quality of process		good	bad
sw CMP good	good	.9	.1
	bad	.1	.9
Qprocess_A8 - sw CMP		good	bad
Are all tools used to produce the software, in all life cycle phases, identified, controlled and retrievable?	yes	.9	.1
	no	.1	.9
Qprocess_A8 - sw CMP		good	bad
Is a software product baseline established and placed in a Software Configuration Index?	yes	.9	.1
	no	.1	.9
Quality of process		good	bad
SQA Process	good	.7	.05
	bad	.3	.95
Qprocess_A9 - SQA Process		good	bad
Is the software quality assurance process properly performed and recorded?	yes	.95	.05
	no	.05	.95
Quality of process		good	bad
certification liaison process	good	.9	.1
	bad	.1	.9

(7) The lower-level network, the quality of the product

Quality of product		good	bad
sw planning	good	.75	.2

	bad	.25	.8
Qproduct_A1 - sw planning		good	bad
Is the system made according to a plan which includes software certification aspects?	yes	.7	.3
	no	.3	.7
Qproduct_A1 - sw planning		good	bad
Is the system made according to a plan which includes software verification aspects?	yes	.9	.1
	no	.1	.9
Quality of product		good	bad
sw development process	good	.8	.2
	bad	.2	.8
Qproduct_A2 - sw development process		good	bad
Does the source code possess the following properties?	yes	1	0
	no	0	1
Qproduct_A2 - sw development process		good	bad
Is the executable object code correct?	yes	1	0
	no	0	1
Quality of product		good	bad
vv of sw req. process	good	.85	.05
	bad	.15	.95
Qproduct_A3 - vv of sw req. process		good	bad
Do the software high level requirements comply with system requirements?	yes	.99	.01
	no	.01	.99
Qproduct_A3 - vv of sw req. process		good	bad
Is the high-level requirement specification correct, unique and consistent?	yes	.95	.05
	no	.05	.95
Qproduct_A3 - vv of sw req. process		good	bad
Has the executable code been verified on the target computer?	yes	.99	.01
	no	.01	.99
Quality of product		good	bad
vv of sw design process	good	.9	.05
	bad	.1	.95
Qproduct_A4- vv of sw design process		good	bad
Does the low-level requirements (design documents) comply with the high level requirements?	yes	.99	.01
	no	.01	.99
Qproduct_A4 - vv of sw design process		good	bad
Are the low-level requirements accurate and consistent?	yes	.95	.05
	no	.05	.95
Qproduct_A4 - vv of sw design process		good	bad

Are there no conflicts between the low-level requirements and the hardware/software features of the target computer?	yes	.99	.01
	no	.01	.99
Qproduct_A4 - vv of sw design process		good	bad
Have all algorithms used in the program been verified with respect to accuracy and correctness?	yes	.95	.05
	no	.05	.95
Qproduct_A4 - vv of sw design process		good	bad
Is the software architecture compatible with high level requirements?	yes	.9	.1
	no	.1	.9
Qproduct_A4 - vv of sw design process		good	bad
Has an analysis shown that the software architecture does not contain any internal inconsistencies?	yes	.9	.1
	no	.1	.9
Qproduct_A4 - vv of sw design process		good	bad
Is the software partitioning integrity confirmed?	yes	.9	.1
	no	.1	.9
Quality of product		good	bad
vv of sw coding & integration process	good	.95	.05
	bad	.05	.95
Qproduct_A5 - vv of sw coding & integration process		good	bad
Is the code a correct and consistent refinement of the lowlevel requirements given in design documents?	yes	.99	.01
	no	.01	.99
Qproduct_A5- vv of sw coding & integration process		good	bad
Does the source code match the data- and control flow defined in the software architecture?	yes	.8	.4
	no	.2	.6
Qproduct_A5- vv of sw coding & integration process		good	bad
Is there a clear correspondence between each element of the design and corresponding code modules?	yes	.7	.4
	no	.3	.6
Qproduct_A5- vv of sw coding & integration process		good	bad
Is it verified that the code is accurate and consistent, also including the attributes	yes	.95	.05
	no	.05	.95
Quality of product		good	bad
testing integration process	good	.99	.01
	bad	.01	.99
Qproduct_A6 - testing integration process		good	bad
sw verification results	yes	.9	.1
	no	.1	.9
Quality of product		good	bad
vv of verification process results	good	.99	.01
	bad	.01	.99

Qproduct_A7 - vv of verification process results		good	bad
Is the test performed correctly?	yes	.9	.2
	no	.1	.8
Qproduct_A7 - vv of verification process results		good	bad
Are all test results correct, or if not, are the discrepancies explained?	yes	.9	.5
	no	.1	.5
Qproduct_A7 - vv of verification process results		good	bad
Does the test cover high- and low level requirements, and all software structures, decisions and branches	yes	.9	.1
	no	.1	.9
Qproduct_A7 - vv of verification process results		good	bad
Have all specified functions and safety actions been tested, and has fault injection been used to test the robustness of the system?	yes	.9	.01
	no	.1	.99
Quality of product		good	bad
sw CMP	good	.8	.5
	bad	.2	.5
Qproduct_A8 - sw CMP		good	bad
Are all software configuration management activities recorded in the Software Configuration Records?	yes	.9	.1
	no	.1	.9
Quality of product		good	bad
SQA Process	good	.7	.05
	bad	.3	.95
Qproduct_A9 - SQA Process		good	bad
Is the software quality assurance process properly performed and recorded?	yes	.9	.1
	no	.1	.9
Qproduct_A9 - SQA Process		good	bad
Does the software development conform to the SQA requirements?	yes	.9	.1
	no	.1	.9
Quality of product		good	bad
certification liaison process	good	.9	.1
	bad	.1	.9
Quality of product		good	bad
quality of COTS	good	.9	.1
	bad	.1	.9
Quality of product		good	bad
mmi_aspects	good	.9	.2
	bad	.1	.8

(8) The lower-level network, the quality of the analysis

Quality of analyses		good	bad
sw planning	good	.6	.5
	bad	.4	.5
Qanalysis_A1 - sw planning		good	bad
Does overall software quality assurance plans exist for all phases?	yes	.6	.4
	no	.4	.6
A1-11.5 Do overall software quality assurance plans exist for all phases?		yes	no
Qualification routines for software systems	yes	.9	.1
	no	.1	.9
A1-11.5 Do overall software quality assurance plans exist for all phases?		yes	no
QA document for managing software projects	yes	.9	.1
	no	.1	.9
A1-11.5 Do overall software quality assurance plans exist for all phases?		yes	no
Qualification of vendors?	yes	.9	.1
	no	.1	.9
A1-11.5 Do overall software quality assurance plans exist for all phases?		yes	no
QA in project management?	yes	.9	.1
	no	.1	.9
Qanalysis_A1 - sw planning		good	bad
Is the software verification plan satisfactory?	yes	.8	.2
	no	.2	.8
Qanalysis_A1 - sw planning		good	bad
Is the recording of all software verification activities complete?	yes	.9	.1
	no	.1	.9
Qanalysis_A1 - sw planning		good	bad
Is the recording of all software quality assurance activities complete?	yes	.9	.1
	no	.1	.9
Quality of analysis		good	bad
sw development process	good	.6	.5
	bad	.4	.5
Qanalysis_A2 - sw development process		good	bad
Is tracability between the high level requirement and the final product facilitated?	yes	.9	.05
	no	.1	.95
Qanalysis_A2 - sw development process		good	bad
Does the source code possess the following properties?:	yes	1	0
	no	0	1

A2-11.11 Does the source code possess the following properties?:		yes	no
Correct implementation of all low-level requirements?	yes	.9	.1
	no	.1	.9
A2-11.11 Does the source code possess the following properties?:		yes	no
Consistency	yes	.9	.1
	no	.1	.9
A2-11.11 Does the source code possess the following properties?:		yes	no
Verifiability	yes	.9	.1
	no	.1	.9
A2-11.11 Does the source code possess the following properties?:		yes	no
Tracability	yes	.9	.1
	no	.1	.9
Qanalysis_A2 - sw development process		good	bad
Is the executable object code correct?	yes	1	0
	no	0	1
A2-11.12 Is the executable object code correct?		yes	no
Is the translation of the source code into executable code made in a trustworthy way?	yes	.9	.1
	no	.1	.9
A2-11.12 Is the executable object code correct?		yes	no
Is the high-level requirement specification correct, unique and consistent?	yes	.9	.1
	no	.1	.9
Quality of analysis		good	bad
vv of sw req. process	good	.9	.1
	bad	.1	.9
Qanalysis_A3 - vv of sw req. process		good	bad
Is the hl req. accurate and consistent	yes	.9	.1
	no	.1	.9
Qanalysis_A3 - vv of sw req. process		good	bad
Is the high-level requirement verifiable	yes	.99	.1
	no	.01	.9
Quality of analysis		good	bad
vv of sw design process	good	.9	.1
	bad	.1	.9
Qanalysis_A4 - vv of sw design process		good	bad
Are the low-level requirements accurate and	yes	.95	.05
	no	.05	.95

consistent?			
Qanalysis_A4 - vv of sw design process		good	bad
Are there no conflicts between the low-level requirements and the hardware/software features of the target computer?	yes	.95	.05
	no	.05	.95
Qanalysis_A4 - vv of sw design process		good	bad
Is the high-level requirement verifiable	yes	.99	.1
	no	.1	.99
Qanalysis_A4 - vv of sw design process		good	bad
Is the software architecture compatible with high level requirements?	yes	.9	.1
	no	.1	.9
Qanalysis_A4 - vv of sw design process		good	bad
Has an analysis shown that the software architecture does not contain any internal inconsistencies?	yes	.95	.05
	no	.05	.95
Qanalysis_A4 - vv of sw design process		good	bad
Are there any conflicts between the software architecture and the hardware/software features of the target computer?	yes	.95	.05
	no	.05	.95
Qanalysis_A4 - vv of sw design process		good	bad
Is the software architecture verifiable?	yes	.95	.05
	no	.05	.95
Quality of analyses		good	bad
vv of sw coding & integration process	good	.9	.1
	bad	.1	.9
Qanalysis_A5 - vv of sw coding & integration process		good	bad
Does the source code match the data- and control flow defined in the software architecture?	yes	.95	.05
	no	.05	.95
Qanalysis_A5 - vv of sw coding & integration process		good	bad
Is the source code verifiable?	yes	.99	.01
	no	.01	.99
Qanalysis_A5 - vv of sw coding & integration process		good	bad
Does the code conform to the coding standard?	yes	.9	.1
	no	.1	.9
Qanalysis_A5 - vv of sw coding & integration process		good	bad
Is there a clear correspondence between each element of the design and corresponding code modules?	yes	.95	.05
	no	.05	.95
Quality of analyses		good	bad
testing integration process	good	.99	.1
	bad	.01	.9
Qanalysis_A6 - testing integration process		good	bad

Which types of tests have been performed on the system?	yes	.9	.1
	no	.1	.9
Quality of analyses		good	bad
vv of verification process results	good	.99	.1
	bad	.01	.9
Qanalysis_A7 - vv of verification process results		good	bad
Is the test performed correctly?	yes	.9	.2
	no	.1	.8
Qanalysis_A7 - vv of verification process results		good	bad
Are all test results correct, or if not, are the discrepancies explained?	yes	.9	.5
	no	.1	.5
Qanalysis_A7 - vv of verification process results		good	bad
Does the test cover High and low-level requirements, software structures, decisions and branches?	yes	.9	.1
	no	.1	.9
Qanalysis_A7 - vv of verification process results		good	bad
Have all specified functions and safety actions been tested, and has fault injection been used to test the robustness of the system	yes	.9	.01
	no	.1	.99
Quality of analyses		good	bad
sw CMP	good	.7	.05
	bad	.3	.95
Qanalysis_A8 - sw CMP		good	bad
Are all problems revealed during the software development reported?	yes	.9	.05
	no	.1	.95
Quality of analyses		good	bad
SQA Process	good	.7	.05
	bad	.3	.95
Qanalysis_A9 - SQA Process		good	bad
Is the software quality assurance process properly performed and recorded?	yes	.7	.05
	no	.3	.95
Quality of analyses		good	bad
Has the development been made in co-ordination with, and with feedback from, the certification authorities?	yes	.7	.1
	no	.3	.9
Quality of analyses		good	bad
Have Human-System Interaction aspects been properly considered during the development of the system?	good	.9	.1
	bad	.1	.9
Quality of analyses		good	bad

How good is the analyses team performance?	good	.95	.1
	bad	.05	.9

(9) Observations on the Quality of the Producer: "yes" and "no"

A1-11.6 sw req. standards	.9	.1
A1-11.7 sw design standards	.95	.05
A1-11.8 sw code standards	.95	.05

A2 Are other products made by the producer reliable and of high quality (pedigree_obs)	1	0
A2 producer experience (experience_obs)	.5	.5
A3_5_obs Qproducer_A3 - vv of sw req. process	.9	.1
A4_12_obs Qproducer_A4 - vv of sw design process	.9	.1
A5_5_obs Qproducer_A5- vv of sw coding & integration process	.7	.3
A7_resources_obs A7: Has the producer sufficient and qualified resources for V&V?	.9	.1
A7_philophy_obs A7: Has the producer an acceptable V&V philosophy?	.9	.1
A7_practice_obs A7: Has the producer long practice in software V&V activities?	.95	.05
Qproducer_A8 - sw CMP (A8 Are all software configuration management activities recorded in the Software Configuration Records?)	1	0
Qproducer_A9 - SQA Process A9 Is the software quality assurance process properly performed and recorded?	.95	.05
Qproducer_A10 - certification liaison process A10 cert_feed_obs (A10_obs)	.9	.1

Observations on the Quality of the Development Process	yes	no
A1-11.2 Is the sw development plan complete?	1	0
A1-11.3 Is the sw verification plan complete?	.7	.3
A1-11.4 Are there acceptable configuration management plans for all phases?	1	0
A1-11.5 SQA-Records	.7	.3
A2-11.10 design description	.8	.2
A2-11.9 Are all system functional requirements, safety requirements and auxiliary requirements specified in the software specification?	.6	.4
A3-2 Are the high-level requirement specification correct, unique and consistent?	.9	.1
A3-6 Is traceability between the high level requirement and the final product facilitated?	.85	.15

A4 -2 Is the design documents a correct and consistent refinement of the high level requirements?	.9	.1
A4 -6 Is there a clear correspondence between each item of the high level requirements and corresponding elements of the design?	.8	.2
A4 -9 Has an analysis shown that the software architecture does not contain any internal inconsistencies?	.85	.15
A5-1 Is the code a correct and consistent refinement of the low-level requirements given in design documents?	.9	.1
A5-5 Is there a clear correspondence between each element of the design and corresponding code modules?	.95	.05
A6 11-13 sw verification cases and procedures	.95	.05
A6 11-14 sw verification results	.95	.05
A8 11-16 Is a software product baseline established and placed in a Software Configuration Index?	.99	.01
A8 11-15 Are all tools used to produce the software, in all life cycle phases, identified, controlled and retrievable?	.99	.01
A9 SQA_obs Is the software quality assurance process properly performed and recorded?	.8	.2
Qprocess_A10 - certification liaison process A10 Has the development been made in coordination with, and with feedback from, the certification authorities?	1	0

Observations on the Quality of the Product	yes	no
A1-11.1 Is the system made according to a plan which includes software certification aspects?	.7	.3
A1-11.3 Is the system made according to a plan which includes software verification aspects?	.8	.2
A2-11.11 Does the source code possess the following properties?	.7	.3
A2 11-12 Is the executable object code correct?	.9	.1
A3-1 Do the software high level requirements comply with system requirements?	1	0
A3-2 Is the high-level requirement specification correct, unique and consistent?	1	0
A3-3 Has the executable code been verified on the target computer?	1	0
A3-7 Have all algorithms used in the program been verified with respect to accuracy and correctness?	.95	.05
A4-1 Do the low-level requirements (design documents) comply with the high level requirements?	.5	.5
A4-2 Are the low-level requirements accurate and consistent?	.8	.2
A4-3 Are there no conflict between the low-level requirements and the hardware/software features of the target computer?	1	0
A4-7 Have all algorithms used in the program been verified	.95	.05

with respect to accuracy and correctness?		
A4-8 Is the software architecture compatible with high level requirements?	.85	.15
A4-9 Has an analysis shown that the software architecture does not contain any internal inconsistencies?	1	0
A4-13 Is the software partitioning integrity confirmed?	0	1
A5-1 Is the code a correct and consistent refinement of the low-level requirements given in design documents?	.9	.1
A5-2 Does the source code match the data- and control flow defined in the software architecture?	.9	.1
A5-5 Is there a clear correspondence between each element of the design and corresponding code modules?	.95	.05
A5-6 Is there verified that the code is accurate and consistent?	.9	.1
A6 11-14 sw verification results	.95	.05
A7_1_obs Is the test performed correctly?	1	0
A7_2_obs Are all test results correct, or if not, are the discrepancies explained?	.95	.05
A7_3_obs Test completeness: Have all specified functions been tested? safety actions been tested? fault injection been used to test the robustness of the system sw verification cases and procedures?	.85	.15
A7_4_obs Does the test cover .	.85	.15
A8 C11-18 Are all software configuration management activities recorded in the Software Configuration Records?	1	0
A9_obs_1 Does the software development conform to the SQA requirements?	.95	.05
A9_obs_2 Is the software quality assurance process properly performed and recorded?	.6	.4
Qproduct_A10 - Has the development been made in co-ordination with, and with feedback from, the certification authorities?	.5	.5
cots_obs Qproduct: If the quality of all COTS used in the system has been evaluated, what is the result?	1	0
hsi_obs Have Human-System Interaction aspects been properly considered during the development of the system?	1	0

Observations on the Quality of the Analysis	yes	no
A1-11.3 Is the software verification plan satisfactory?	.7	.3
a11_5_q1_obs Qualification routines for software systems	.9	.1
a11_5_q2_obs QA document for managing software projects	1	0
a11_5_q3_obs Qualification of vendors	.9	.1
a11_5_q4_obs QA in project management	.9	.1
A11_11_q1_obs Correct implementation of all low-level requirements	.98	.02
A11_11_q2_obs Consistency	.98	.02

A11_11_q3_obs Verifiability	1	0
A11_11_q4_obs Traceability	.95	.05
A11_12_q1_obs Is the translation of the source code into executable code made in a trustworthy way?	1	0
A11_12_q2_obs Is the high-level requirement specification correct unique and consistent?	1	0
A1-11.14 Is the recording of all software verification activities complete	1	0
A1-11.19 Is the recording of all software quality assurance activities complete?	1	0
A2-11.9 Is tracability between the high level requirement and the final product facilitated?	1	0
A3-2 Is the high-level req. accurate and consistent	.98	.02
A3-4 Is the high-level requirement verifiable	1	0
A4-2 Are the low-level requirements accurate and consistent?	.95	.05
A4-3 Are there no conflicts between the low-level requirements and the hardware/software features of the target computer?	1	0
A4-4 Is the high-level requirement verifiable	1	0
A4-8 Is the software architecture compatible with high level requirements?	.98	.02
A4-9 Has an analysis shown that the software architecture does not contain any internal inconsistencies?	.8	.2
A4-10 Are there any conflicts between the software architecture and the hardware/software features of the target computer?	1	0
A4-11 Is the software architecture verifiable?	1	0
A5-2 Does the source code match the data- and control flow defined in the software architecture?	.9	.1
A5-3 Is the source code verifiable?	1	0
A5-4 Does the code conform to the coding standard?	1	0
A5-5 Is there a clear correspondence between each element of the design and corresponding code modules?	.9	.1
A6 11-13 All types of tests have been performed on the system?	.95	.05
A7: Have all specified functions been tested? Have all specified safety actions been tested? Has fault injection been used to test the robustness of the system sw verification cases and procedures?	.85	.15
A7_1: Is the test performed correctly?	1	0
A7_2: Are all test results correct, or if not, are the discrepancies explained?	.95	.05
A7 3-8: Does the test cover all requirements and structural attributes?	.85	.15
A8 11-17 Are all problems revealed during the software development reported?	1	0

Is the software quality assurance process properly performed and recorded?	.95	.05
A10 Has the development been made in co-ordination with, and with feedback from, the certification authorities?	.9	.1
HMI_obs Have Human-System Interaction aspects been properly considered during the development of the system?	1	0
Analyses_team_obs How good is the analyses team performance?	.9	.1



4.2 Review of software design document using BBN[13]

안전 소프트웨어를 평가할 때 품질에 관련된 여러 증거들을 토대로 결론을 내리는 것은 쉽지 않고 또 승인 기준을 만족하는 충분한 증거들이 모아졌다는 시점을 판단하는 것도 어렵다. 일반적으로 이런 평가는 정성적으로 이루어지며 체크리스트를 많이 사용하는데 체크리스트의 사용에서 생기는 문제점은 다음과 같다.

o 아무리 잘 개발된 소프트웨어라도 체크리스트 상에서 몇 가지 불충분한 점수를 받는 항목이 있을 수 있다. 체크리스트는 이런 몇 가지 부정적 증거에도 불구하고 전체 결론이 긍정적이 된 것에 대한 적절한 설명을 해주지 못한다.

o 체크리스트는 어느 정도까지 검사를 해야 소프트웨어가 합리적인 신뢰성을 확보했는지 결정하는데 도움을 줄 수 없다.

o 체크리스트로는 소프트웨어 공학에서 사용되는 측정법들과 같은 서로 다른 종류의 정보에 대해 고려하기 힘들다.

이런 체크리스트를 이용한 안전 소프트웨어의 승인 문제를 해결하기 위하여 BBN을 이용한 연구[13]가 수행되었으며 작성된 BBN 그래프는 다음과 같다.

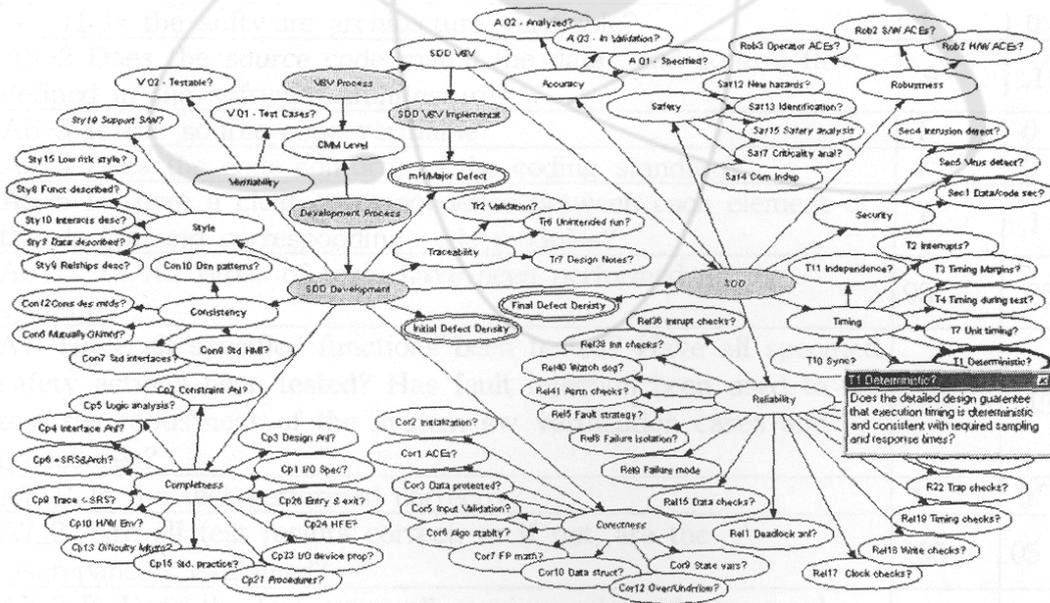


그림 2 소프트웨어 설계 문서 평가 BBN 그래프

그림 2에 나타난 BBN의 노드확률테이블은 다음과 같다.

(1) Requirements Phase Model

Name: System Requirements Safety

Correct	0.99
Incorrect	0.01

Name: Requirements Safety

System Requirements Safety	Correct		Incorrect	
Final Requirements	Good	Poor	Good	Poor
Safe	1	0	0	0
Not Safe	0	0	0	0

Name: Initial Requirements

Development Process	Good	Poor
Good	0.99	0.9
Poor	0.01	0.1

Name: Development Process R

Good	0.9
Poor	0.1

Name: CMM Level R

V&V	Good		Poor	
Development Process	Good	Poor	Good	Poor
< Level 3	0.01	0.95	0.95	0.99
Level 3	0.99	0.05	0.05	0.01

Name: V&V R

Good	0.9
Poor	0.1

Name: Final Requirements

Initial Requirements	Good				Poor			
Development Process	Good		Poor		Good		Poor	
V&V	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Good	0.98	0.89	0.69	0.49	0.89	0.81	0.63	0.45
Poor	0.02	0.11	0.31	0.51	0.11	0.19	0.37	0.55

(2) Architectural Design Phase Model

Name: Initial Architecture

Development Process	Good		Poor	
Good	0.99		0.9	
Poor	0.01		0.1	

Name: Development Process A

Good	0.9
Poor	0.1

Name: CMM Level A

V&V	Good		Poor	
Development Process	Good		Poor	
< Level 3	0.01	0.95	0.95	0.99
Level 3	0.99	0.05	0.05	0.01

Name: V&V A

Good	0.9
Poor	0.1

Name: Final Architecture

Initial Architecture	Good				Poor			
Development Process	Good		Poor		Good		Poor	
V&V	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Good	0.98	0.89	0.69	0.49	0.89	0.81	0.63	0.45
Poor	0.02	0.11	0.31	0.51	0.11	0.19	0.37	0.55

Name: Complexity A

Low	0.33
Medium	0.33
High	0.33

Name: Architecture Safety

Requirements Safety	Correct		Incorrect	
Final Architecture	Good	Poor	Good	Poor
Safe	1	0	0	0
Not Safe	0	0	0	0

(3) Detailed Design Phase Model

Name: Initial Design

Development Process	Good	Poor
Good	0.99	0.9
Poor	0.01	0.1

Name: Development Process D

Good	0.9
Poor	0.1

Name: CMM Level D

V&V	Good		Poor	
Development Process	Good	Poor	Good	Poor
< Level 3	0.01	0.95	0.95	0.99
Level 3	0.99	0.05	0.05	0.01

Name: V&V D

Good	0.9
Poor	0.1

Name: Final Design

Initial Design	Good		Poor	
Development Process	Good	Poor	Good	Poor

V&V	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Good	0.98	0.89	0.69	0.49	0.89	0.81	0.63	0.45
Poor	0.02	0.11	0.31	0.51	0.11	0.19	0.37	0.55

Name: Complexity D

Low	0.33
Medium	0.33
High	0.33

Name: Design Safety

Architecture Safety	Correct		Incorrect	
Final Design	Good	Poor	Good	Poor
Safe	1	0	00	
Not Safe	0	0	00	

(4) Implementation Phase Model

Name: Initial Code

Development Process	Good	Poor
Good	0.99	0.9
Poor	0.01	0.1

Name: Development Process C

Good	0.9
Poor	0.1

Name: CMM Level C

V&V	Good		Poor	
Development Process	Good	Poor	Good	Poor
< Level 3	0.01	0.95	0.95	0.99
Level 3	0.99	0.05	0.05	0.01

Name: V&V C

Good	0.9
------	-----

Poor	0.1
------	-----

Name: Final Code

Initial Architecture	Good				Poor			
Development Process	Good		Poor		Good		Poor	
V&V	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Good	0.98	0.89	0.69	0.49	0.89	0.81	0.63	0.45
Poor	0.02	0.11	0.31	0.51	0.11	0.19	0.37	0.55

Name: Complexity C

Low	0.33
Medium	0.33
High	0.33

Name: Code Safety

Requirements Safety	Correct		Incorrect	
Final Architecture	Good	Poor	Good	Poor
Safe	1	0	00	
Not Safe	0	0	00	

(5) Anomaly Resolution Model

Name: Initial Revisions

Development Process	Good	Poor
Good	0.99	0.9
Poor	0.01	0.1

Name: Development Process AR

Good	0.9
Poor	0.1

Name: CMM Level AR

V&V	Good		Poor	
Development Process	Good	Poor	Good	Poor

< Level 3	0.01	0.95	0.95	0.99
Level 3	0.99	0.05	0.050.01	0.01

Name: V&V AR

Good	0.9
Poor	0.1

Name: Final Revisions

Initial Revisions	Good				Poor			
Development Process	Good		Good		Good		Poor	
V&V	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Good	0.98	0.89	0.69	0.49	0.89	0.81	0.63	0.45
Poor	0.02	0.11	0.31	0.51	0.11	0.19	0.37	0.55

Name: Complexity AR

Low	0.33
Medium	0.33
High	0.33

Name: Anomaly Resolution

Testing	Correct		Incorrect	
Final Revisions	Good	Poor	Good	Poor
Good	1	0	0	0
Poor	0	0	0	0

(6) Validation Testing Model

Name: Testing

Good	0.9
Poor	0.1

(7) Final Product Model

Name: Validated Software Safety

--	--	--

Code Safety	Safe		Not Safe	
Anomaly Resolution	Good	Poor	Good	Poor
Safe	1	1	0.9	0
Not Safe	0	0	0.1	1

Name: Reliability

Validated Software Safety	Safe	Not Safe
< 0.99	0.05	0.5
0.99 0.999	0.15	0.3
0.999 0.9999	0.5	0.15
> 0.9999	0.3	0.05

4.3 기타 BBN의 노드 확률 테이블(node probability table)

4.3.1 외부 개발 계측제어 시스템의 평가 BBN[14]

Edf 내부 평가 전문가들이 외부에서 개발된 계측제어 시스템을 평가하고 승인하기 위하여 BBN을 사용하였다. 이 BBN의 노드 확률 테이블(node probability table)은 다음과 같은 기준으로 작성되었다.

- o 모든 노드는 2개의 상태 공간을 가진다.
(예) good or poor, acceptable or not acceptable, yes or no etc.
- o 모든 노드는 "object x is y"와 같은 문장과 일치시킨다.
(예) Requirements specification is suitable
- o 노드 상태에 부여할 확률 값은 {0, 0.25, 0.5, 0.75, 1}의 집합에서 선택한다. 이유는 이보다 더 상세한 값에 대한 정당화가 어려웠기 때문이다. 또한 이들 값은 다음과 같은 정성적 판단 기준을 표현한다.
{impossible, improbable, probable, quite probable, certain}

4.3.2 컴퓨터 시스템 명세 문서의 안전성 평가용 BBN[15]

컴퓨터 시스템의 검증에 관련된 문제를 해결하고 확인을 위한 설계 패러다임을 제공하기 위해 수행된 과제 "Deva : design process"의 일부분으로 컴퓨터 시스템 명세 문서의 안전성 평가가 BBN을 이용하여 수행되었다. 이 BBN의 노드 확률 테이블은 독립된 평가 전문가가 그의 전문 영역의 지식과 경험을 바탕으로 직접 작성하였는데 이 방법은 BBN의 규모가 작고 노드 확률 테이블(node probability table)의 차원이 작았기 때문에 가능하였다. 일부 노드의 노드 확률 테이블(node probability

table)은 선형 보간법(linear interpolation)을 사용하여 테이블 값을 채운 다음 전문가의 검토를 받는 방식으로 수행되었다. 이렇게 작성된 노드확률테이블(node probability table) 중 "Licensee Verification Results" 노드의 노드확률테이블(node probability table)는 다음과 같다.

"Licensee Verification Results" 노드의 노드확률테이블(node probability table)

Licensee verification quality		Low				
Adequacy of comp. system spec.		Awful	Unsatisfac.	OK	Good	Wanderful
Licensee verification results	0 issues	0.1	0.22	0.9	0.9425	0.95
	A few issues	0.2	0.182	0.08	0.046	0.04
	Many issues	0.7	0.598	0.02	0.015	0.01

Licensee verification quality		Medium				
Adequacy of comp. system spec.		Awful	Unsatisfact	OK	Good	Wandeful
Licensee verification results	0 issues	0.05	0.065	0.15	0.7875	0.9
	A few issues	0.15	0.2475	0.8	0.1965	0.09
	Many issues	0.8	0.6875	0.05	0.016	0.01

Licensee verification quality		High				
Adequacy of comp. system spec.		Awful	Unsatisfact	OK	Good	Wandeful
Licensee verification results	0 issues	0.02	0.0245	0.05	0.6875	0.8
	A few issues	0.08	0.203	0.9	0.2965	0.19
	Many issues	0.9	0.7725	0.05	0.016	0.01

5. 요약 및 결론

BBN은 안전 소프트웨어의 신뢰도를 정량 평가하기 위하여 최근에 많이 활용되고 있는 기술이다. BBN을 구축할 때 가장 어려운 점은 노드확률테이블의 작성과 같은 정성 데이터의 정량화(확률 값 설정)이며 여기에는 전문가의 지식추출 과정이 필연적으로 포함된다. 이 과정에서 바이어스와 오류가 추출 자료에 도입되고 이를 회피하기 위한 체계적 지식 추출 시스템이 요구되나 현재로서는 일반적으로 인정되는 기법이나 방법론이 없는 실정이다. 본 보고서에서는 지식추출 과정에서 일반적으로 발생하는 바이어스와 오류의 종류, 소프트웨어 공학의 특성이 고려된 지식 추출 절차 그리고 기존 BBN 연구에서 나타난 전문가의 지식추출 사례(노드확률테이블 위주)를 정리하였고 이들 자료는 원전 안전소프트웨어의 신뢰도와 안전성 평가용 BBN 연구에 활용될 예정이다.

5.1 전문가의 판단에 내재된 바이어스의 형태와 내용

전문가의 판단에 내재된 바이어스를 처리하는 방법들은 드물고 아직 기술적으로 초기단계이다. 판단(judgment)에 대한 연구는 전문가의 바이어스 존재에 대한 주의를 이끌어 냈지만 지식 추출 과정에서 발생하는 바이어스를 처리하는 문제에 대해서는 아직 효과적인 방안이 나오지 않고 있다. 따라서 현재까지는 이 분야의 전문가들은 그들이 일상적으로 접하는 바이어스에 대한 그들 고유의 처리 방법을 개발해서 사용하고 있는 실정이다. 본 보고서에서는 심리학과 지식 공학 분야에서 연구된 바이어스의 형태와 부분적인 방지 방법에 대하여 정리하였으며 이 자료는 소프트웨어 평가 전문가 또는 지식 추출 담당자가 관련된 작업을 수행할 때 교육 또는 보조 자료로 활용할 예정이다.

5.2 전문가의 지식 추출 절차

BBN은 일반적으로 불확실성이 존재하는 문제의 해결에 사용되는데 이때 가장 어려운 점은 전문가의 주관적, 정성적 지식을 정확하게 정량화(확률 값으로 변환)하는 것이다. BBN으로 모델링 할 때 일반적으로 다음과 같이 3가지 부분에서 전문가의 지식과 판단이 필요하다.

- (a) BBN 그래프(graph or topology) 작성
- (b) 노드확률테이블(node probability table) 작성
- (c) 노드(변수)의 입력 값 작성: 정성적 증거(evidence)값의 정량화(확률 값으로 변환) 경우

본 보고서에 기술된 전문가의 지식추출 절차와 기법은 일반적인 문제 해결의 경우를 대상으로 작성되었으나 소프트웨어의 평가를 위한 BBN 구축 작업에도 효과가 있을 것으로 판단된다. 그리고 이 절차와 기법은 추후 실제 소프트웨어 평가 문제에 적용하고 그 경험을 바탕으로 동 분야의 전문적 가이드로 보완할 계획이다.

5.3 기존 BBN 연구에서 나타난 전문가 지식 추출 사례

디지털 시스템의 신뢰도와 안전성 평가를 위해 BBN을 사용한 연구에서 나타난 전문가의 지식 추출 사례를 노드확률테이블 위주로 정리하였다. 디지털 시스템의 평가에 BBN을 활용한 사례가 아직 많지 않은 관계로 여기에 수집된 자료가 일반적으로 인정되기에는 어려우나 이들은 모두 각 분야의 전문가와 지식 공학자들에 의하여 실제 케이스로 작성된 자료이므로 앞으로의 소프트웨어의 신뢰도와 안전성 평가를 위한 BBN 연구에 참조가 될 것으로 보인다.

5.4 추후 연구 내용

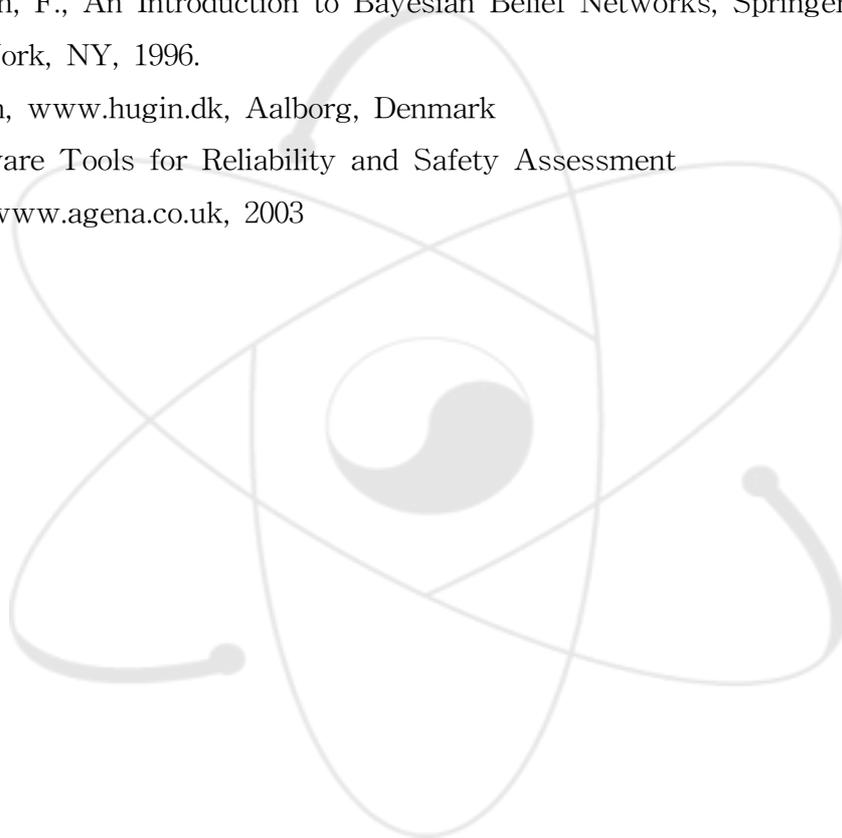
안전 소프트웨어의 신뢰도를 정량 평가하는 BBN을 구축하기 위하여 추가적으로 필요한 연구 내용은 다음과 같다.

- o 전문가에 의해 작성된 자료 또는 기 수집된 자료에서 바이어스를 발견하고 제거하는 기법
- o 제 3장의 전문가 지식 추출 절차를 안전 소프트웨어의 신뢰도, 안전성, 품질 평가에 전문적으로 사용할 수 있도록 구체화
- o 소프트웨어와 디지털 시스템의 신뢰도 정량 평가용 BBN 모델에 사용된 확률값의 추가적인 수집 및 분석. 여기에는 세부적으로 (i) 노드확률테이블(node probability table)의 조건부 확률 값, (ii) root node의 사전 확률 값, (iii) 노드에 입력되는 증거(evidence)의 확률 값이 있다.

참고문헌

- [1] 엄홍섭, 강현국, 성태용, 원전 안전 소프트웨어의 정량적 신뢰도 평가를 위한 Bayesian Belief Nets 기술 분석, KAERI/AR-594/2001, 한국원자력연구소, 2001
- [2] Kahneman, D. & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, 80, 237-251.
- [3] Eddy D., "Probabilistic reasoning in clinical medicine," *Judgement under uncertainty*, Cambridge University Press, Cambridge, 1982
- [4] Thys, W., "Fault Management," Ph.D. dissertation, Delft University of Technology, Delft, 1987.
- [5] Tversky, A., and Kahneman, D., "Availability: A Heuristic for Judging Frequency and Probability," *Judgement under Uncertainty*, Cambridge University Press, 1982
- [6] Bar-Hillel, M. On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 1973, 9, 396-406.
- [7] Langer, E., "The illusion of control," *The Journal of Personality and Social Psychology*, vol. 32, 1975.
- [8] Payne, S. *The Art of Asking Questions*, Princeton University Press, 1951
- [9] Hogarth, R. Cognitive Process and the Assessment of Subjective Probability Distribution, *Journal of the American Statistical Association*, 70, (350), 1975
- [10] Meyer M.A., *Elicitin and Analyzing Expert Judgment*, NUREC/CR-5424, 1990
- [11] RTCA. *Software considerations in airborne systems and equipment certification*, DO-178B, Requirements ad Technical Concepts for Aeronautics, 1992.
- [12] Bjorn Axel Gran and Gustav Dahll, *Estimating dependability of programmable systems using bayesian belief nets*, OECD HALDEN rector project, HWR-627, 2000 May
- [13] Gary Johnson, *Bayesian Belief Networks based Review of Software Design Documents*, International Topical Meeting on Nuclear Instrumentation, Control, and Human-Machine Technologies(NPIC & HMIT 2000), Washington DC, Nov. 2000

- [14] Marc Bouissou, Assessment of a Safety-Critical System Including Software: A Bayesian Belief Network for Evidence Sources, Reliability and Maintainability Symposium, 1999
- [15] B. Littlewood and L. Strigini, Examination of Bayesian Belief Networks for Safety Assessment of Nuclear Computer-based Systems, ESPRIT DeVa Project 20072, 1998
- [16] Neil, M., et al, "Applying Bayesian Belief Networks to System Dependability Assessments," Proceedings of Safety Critical Systems Club, February 1996.
- [17] Jensen, F., An Introduction to Bayesian Belief Networks, Springer-Verlag, New York, NY, 1996.
- [18] Hugin, www.hugin.dk, Aalborg, Denmark
- [19] Software Tools for Reliability and Safety Assessment
<http://www.agenaco.uk>, 2003



서 지 정 보 양 식

수행기관보고서번호	위탁기관보고서번호	표준보고서번호	INIS 주제코드
KAERI/TR-xxxx/2003			
제목 / 부제	안전 소프트웨어의 신뢰도 정량 평가 BBN을 위한 전문가의 지식추출 가이드		
연구책임자 및 부서명 (주저자)	엄홍섭 (종합안전평가부)		
연구자 및 부서명	강현국 (종합안전평가부), 장승철(종합안전평가부) 하재주 (종합안전평가부)		
출판지	대전	발행기관	KAERI
페이지	121 p.	도표	있음(○), 없음()
발행년	2003. 8.		
크기	21×29.7cm		
참고사항			
비밀여부	공개(○), 대외비(), _ 급비밀	보고서종류	기술보고서
연구위탁기관		계약 번호	
초록	<p>Bayesian Belief Nets(BBN) 기술을 활용하여 안전 소프트웨어의 신뢰도를 정량 평가할 때 필연적으로 전문가의 지식을 추출하는 과정이 포함된다. 특히 BBN의 노드확률테이블을 작성할 때와 노드에 입력되는 값을 정할 때는 전문가의 주관적 판단이나 정성적 자료를 확률적 표현으로 변환해야 하는 경우가 많은데 이 과정에서 여러 가지 바이어스나 인적 오류가 도입될 가능성이 높다. 이런 바이어스나 오류들의 발생을 사전에 방지하거나 또는 이미 전문가에 의하여 작성된 자료에 포함된 바이어스와 오류를 제거할 수 있는 방안을 제공하는 것이 본 가이드의 목적이다. 기술보고서의 주요 내용은 다음과 같다.</p> <ul style="list-style-type: none"> ○ 전문가의 지식 추출 과정 특히 확률의 추출 작업에서 나타날 수 있는 여러 가지의 바이어스(bias, 편견)와 오류의 내용과 사례 ○ 전문가로부터 지식을 추출하는 과정에서 바이어스의 도입을 방지할 수 있는 절차와 기법 ○ 디지털시스템의 신뢰도/안전성 평가용 BBN 모델에 사용된 전문가의 확률 값 사례 		
주제명 키워드 (10단어내외)	안전 소프트웨어, 신뢰도, Bayesian Belief Nets, BBN, 전문가 지식 추출, 바이어스		

BIBLIOGRAPHIC INFORMATION SHEET					
Performing Org. Report No.		Sponsoring Org. Report No.		Standard Report No.	INIS Subject Code
KAERI/TR-xxxx/2003					
Title / Subtitle		A Guide on the Elicitation of Expert Knowledge in Constructing BBN for Quantitative Reliability Assessment of Safety Critical Software			
Project Manager and Department		H.S. Eom (Integrated Safety Assessment Division)			
Researcher and Department		H.G. Kang (ISA), S. C. Chang(ISA), J. J, Ha(ISA)			
Publication Place	Daejon	Publisher	KAERI	Publication Date	2003. 8.
Page	121 p.	Ill. & Tab.	Yes(<input type="radio"/>), No (<input type="checkbox"/>)	Size	21× 29.7cm
Note					
Classified	Open(<input type="radio"/>),Restricted(<input type="checkbox"/>),- __ Class Document		Report Type	Technical Report	
Sponsoring Org.				Contract No.	
Abstract(15-20 Lines)		<p>It is necessary to elicit expert's knowledge while we quantitatively assess the reliability of safety critical software using bayesian belief nets. Especially in composing the node probability table and in making out the input data for BBN model, expert's qualitative judgments or qualitative data should be converted into probabilistic representation. This conversion process has vulnerability which is easy to introduce bias or error to the data we use. The purpose of the report is to provide the guideline to avoid the occurrence of this kinds of bias/error or to eliminate them which is included in the existing data prepared by experts. The contents of the report are as follows:</p> <ul style="list-style-type: none"> o The types of bias and error which might be occur in the process of eliciting the expert's knowledge. o The process and techniques to avoid bias and error in eliciting the expert's judgments. o The examples of expert's knowledge(probability values) appeared in the BBNs for assessing the reliability of digital system. 			
Subject Keywords (About 10 words)		Safety critical software, Bayesian Belief Nets, BBN, Elicitation of expert's knowledge, bias			