

질적변수에 대한 계량화를 통한 사면붕괴 예측모형[†]

(Prediction Modeling through Quantification for
Qualitative Variables)

나종화*, 유혜경**, 남은미***, 조완섭****

(Jong-Hwa Na, Hye-Kyung Yu, Eun-Mi Nam, Wan-Sup Cho)

요 약 본 논문에서는 수량화 방법과 AHP(Analytic Hierarchy Process) 기법을 사용하여 산사태 발생에 대한 통계적 예측모형을 구축하는데 목적이 있다. 수량화(Quantification) 방법은 질적변수에 수량을 부여하는 통계적 방법으로, 기 조사된 자료에 기반하여 분석을 수행하는 방법이다. 본 논문에서는 서구의 다변량분석 기법인 정준상관분석의 결과를 토대로 수량화 과정을 구체적으로 제안한다. 데이터에 기반한 수량화 방법과는 달리 AHP 기법은 일종의 다기준 의사결정을 위해 사용되는 기법으로, 설문자료에 기반한 분석법이다. 실제자료에 대한 분석으로 산사태 발생여부를 측정된 자료(한국지질자원연구원 제공)와 전문가 설문을 통해 수집된 자료를 이용하였다. 이들 자료에 대해 수량화 분석과 AHP분석을 통해 산사태 발생여부를 예측할 수 있는 두 종류의 평가표와 함께 로지스틱 회귀를 통한 통계적 예측모형을 개발하였으며, 두 모형간의 성능비교와 안정성 평가를 수행하였다.

핵심주제어 : 수량화Ⅱ 방법, 정준상관분석, AHP기법, 로지스틱 회귀모형

Abstract The purpose of this paper is to provide the statistical models for landslide prediction through quantification and AHP methods. Quantification method is a statistical method of providing quantity to qualitative variables by analyzing the observed data. In this paper, we suggest the quantification process based on the results of canonical correlation analysis. In contrast with the quantification method which is based on given data, the AHP(Analytic Hierarchy Process) technique is a kind of method based on questionnaire data which is usually taken from professionals. We analyze both the real data(provided from KIGAM) and questionnaire data collected from professionals of various related area. We developed two kinds of evaluation table which provide the scores of landslide possibility and the logistic model providing the probability of occurring landslide. Finally we compare the performance and evaluate the stability of the suggested two models.

Key Words : quantification method II, canonical correlation analysis, analytic hierarchy process, logistic regression

[†] 본 연구는 교육과학기술부와 한국산업기술진흥원의 지역혁신
인력양성사업으로 수행된 연구결과임

* 충북대학교 정보통계학과 교수

** 충북대학교 정보통계학과 박사과정

*** 식품의약품안전청(KFDA)

**** 충북대학교 경영정보학과 교수

1. 서론

본 연구에서는 산사태 발생에 영향을 미치는 주요 자연변인의 상대적 중요도를 점수화하여 현장에서 적용할 수 있는 산사태 위험도를 측정할 수 있는 등급표를 통계적 기법을 통해 구축하는 것을 그 목적으로 한다. 분석에 사용된 방법으로는 수량화II 방법과 AHP 기법이 각각 사용되었다.

수량화II 방법은 1940년대 말 일본의 Hayashi 박사에 의해 개발된 방법으로 질적인 자료를 분석의 목적에 맞도록 최적의 수량이나 점수를 주어 양적인 자료로 변환하는 분석법으로,^[2] 서구의 통계이론인 정준상관분석과 밀접하게 연관되어 있다.^[3] 반면 전문가의 설문에 기초한 AHP 분석은 1970년대 말 Satty에 의해 처음 소개된 방법으로,^{[8],[9]} 실제자료에 대한 조사가 미흡하거나 전문가의 견해가 상대적으로 중요하게 작용하는 문제에 효과적인 방법이다.^{[4],[5]}

본 논문의 2장에서는 수량화II 분석의 과정을 설명하고, 산사태 자료에 대한 소개와 함께 분석결과(평가표)를 제시하였다. 3장에서는 AHP 분석을 위한 전문가 설문의 구성과 AHP 분석 절차 및 분석결과(평가표)를 제시하였다. 4장에서는 제시된 두 가지 평가표의 성능비교를 수행하고, 평가표점수를 통한 산사태 발생확률에 대한 예측모형을 구축하였으며, 5장은 결론으로 구성되었다. 본 논문에서 수량화분석은 SAS 9.1로, AHP기법은 R프로그램을 이용해 각각 분석을 수행하였다.

II. 정준상관분석을 이용한 수량화방법

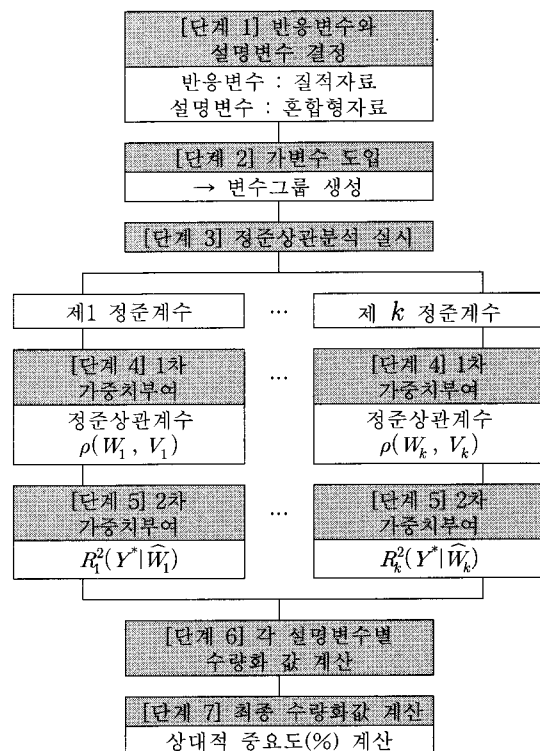
2-1. 정준상관분석 이론

정준상관분석(canonical correlation analysis)은 여러 변수들을 두 변수집단인 반응변수집단과 설명변수집단으로 분류한 후, 두 변수집단 사이의 상관구조를 가장 잘 설명할 수 있는 변수들의 선형결합(linear combination)을 찾는 것이다. 즉, 변수집단에서 상관이 가장 큰 선형결

합을 찾는 것이다. 이 때 변수들이 범주형인 경우는 가변수화하여 활용할 수 있다. 선형결합의 짝의 수의 크기가 가장 작은 변수집단의 변수수와 일치될 때까지 계속된다. 정준상관은 두 변수군 사이의 연관성의 정도를 나타낸다.^{[1],[6]}

2-2. 수량화II 방법 적용 단계

정준상관분석을 이용한 수량화II 분석 과정은 아래의 <그림 1>과 같다.



<그림 1> 수량화II 분석 과정

위 그림에서 정준계수는 정준상관분석을 실시하여 구한다. 다음의 선형결합식

$$W = aY,$$

$$V = b'X = b_1X_1 + b_2X_2 + \dots + b_qX_q$$

에서 제1 정준변수쌍이라 불리는 $(\widehat{W}_1, \widehat{V}_1)$ 은 W 와 V 간의 단순상관계수 $\rho(W_1, V_1)$ 을 최대화시키는 변수집단내의 선형결합으로 각각의 분산

이 단위분산이 되도록 구성한다. 다음으로 제2 정준변수쌍 ($\widehat{W}_2, \widehat{V}_2$)는 ($\widehat{W}_1, \widehat{V}_1$)과는 독립이며 역시 단순상관계수가 최대가 되도록 하는 단위 분산을 갖는 선형결합을 구한다. 이 과정을 반복하여 제k 정준변수쌍 ($\widehat{W}_k, \widehat{V}_k$)을 구한다. 1차 가중치가 부여된 정준계수에 2차 가중치 $R_i^2(Y^*|\widehat{W}_i)$ 값을 곱하여 각 설명변수별 정준계수에 2차 가중치를 부여한다. 여기서, 2차 가중치 $R_i^2(Y^*|\widehat{W}_i)$ 는 종속변수집단 Y^* 의 전체 표본분산 p 에 대해 i 번째 정준변수 \widehat{W}_i 에 의하여 설명되어지는 비율로 식으로 표현하면 다음과 같다.

$$R_i^2(Y^*|\widehat{W}_i) = \sum_{k=1}^p r^2(\widehat{W}_i, Y_k^*)/p$$

여기서 Y^* 는 변수집단 Y 를 표준화한 것을 의미한다.

2-3. 수량화II 방법을 이용한 산사태 자료 분석

2-3-1. 분석 자료 소개

본 논문에서는 지질별 산사태 발생에 대한 자료(제공 : 한국지질자원연구원)를 이용해 산사태 발생 위험도 산정을 위한 평가표를 개발하고자 한다. 자료의 수는 310개이며 분석에 사용된 변수는 <표 1>과 같다.

<표 1> 산사태 자료 변수 설명

구분	변수	설명
반응변수	산사태 발생 여부	1:발생/ 0:미발생
설명변수	Lithology (*)	암석종류
	Elevation	지형고도
	Slope Angle	사면경사
	Specific Gravity	비중
	Moisture Content	함수율
	Void ratio	간극비
	Porosity	공극률
	Saturation	습윤도
	Wet Density	전체밀도
	Sat_Density	포화밀도
	Dry Density	건조밀도
	USCS (*)	입도분포
	Permeability	투수계수

2-3-2. 분석 결과

산사태 자료에 대해 수량화II 분석을 수행한 결과를 이용하여 모든 변수를 상대적인 크기로 나타내고 반응변수에 미치는 중요도를 계산한 후 관련 전문가의 의견을 고려하여 산사태 발생에 영향을 미치는 중요 변수를 <표 2>와 같이 6개의 변수로 결정하였다. 선정된 6개 변수 중 Slope angle(사면경사)이 반응변수 산사태 발생 여부에 가장 많은 영향을 미치는 것으로 나타났다.

<표 2> 중요 변수 선정 결과

변수	점수 (중요도)	순위
Lithology	10(9.7)%	5
Elevation(m)	15(14.4)%	2
Slope angle(°)	40(41.2)%	1
Porosity(%)	12(12.2)%	4
Dry density(t/m³)	8(8.4)%	6
Permeability(cm/s)	15(14.2)%	3
합계	100%	

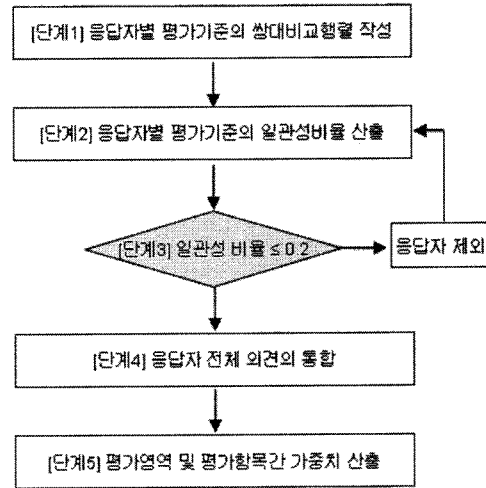
최종 선정된 6개 변수를 수량화II 분석을 이용해 <표 3>과 같이 산사태 발생위험도를 수치적으로 알아볼 수 있는 산사태 발생 위험도 평가표를 개발하였다.

평가표를 만들기 위해서 각 변수별로 분위수 정보를 이용해 산사태 발생에 미치는 영향을 3~5등급으로 분류하였다. 평가표는 100점 만점의 표로써 특정 지점의 산사태 발생위험도를 수치적으로 평가할 수 있게 해준다. 수량화 분석을 통해 제시된 평가표는 AHP의 기초자료로 활용하였다.

<표 3> 산사태 발생 위험도 평가표(수량화II)

관찰항목	세부등급	구분	점수
Lithology	Granite	1	10
	Gneiss	2	9
	Mudstone	3	6
Elevation(m)	431 ~ 785	1	15
	20 ~ 225	2	10
	226 ~ 430	3	5
Slope angle(°)	27.1 이상	1	40
	23.1 ~ 27.0	2	30
	19.1 ~ 23.0	3	20
	15.1 ~ 19.0	4	10

Porosity(%)	15.0이하	5	0
	80.1이상	1	12
	75.1~80.0	2	9
	67.1~75.0	3	6
	46.1~67.0	4	3
Dry Density (g/cm3)	46.0이하	5	0
	0.80~1.18	1	8
	1.19~1.31	2	6
	1.32~1.38	3	4
	1.39~1.47	4	2
Permeability (cm/s)	1.48이상	5	0
	$6.0 \times 10^{-3} \sim 1.0 \times 10^{-2}$	1	15
	$4.5 \times 10^{-3} \sim 6.0 \times 10^{-3}$	2	11
	$2.6 \times 10^{-3} \sim 4.5 \times 10^{-3}$	3	7
	$1.1 \times 10^{-3} \sim 2.5 \times 10^{-3}$	4	3
	1×10^{-3} 이하	5	0



<그림 2> AHP 분석 절차

III. AHP기법을 이용한 수량화 방법

3-1. AHP기법 이론 및 적용단계

다기준 의사결정 기법의 하나인 AHP기법은 의사결정을 위한 방법으로 Saaty(1980)에 의해 개발되었다. 이 기법은 최종 대안의 선택 또는 종합적 목표(overall goal)를 달성하기 위해 고려해야 할 여러 기준들(criteria)의 상대적 가중치와 각 기준에 있어서 대안들(alternatives)의 상대적 가중치를 결정 한 뒤 이들을 곱한 평점의 합을 비교하여 대안들 간의 종합적 우선순위를 평가하는 방법이다. 본 연구에서는 AHP 분석의 중간과정에 사용되는 기준(중요항목)과 대안(세부항목)들에 대해 AHP기법의 상대적 가중치결정법을 이용하고자 한다.

전문가 설문 의 응답자료를 이용한 AHP기법의 적용은 회수된 설문 에 대해 일반적으로 다음 <그림 2>와 같은 단계를 통해 이루어진다.

3-2. AHP기법을 이용한 산사태 자료 분석

3-2-1. 분석 자료 소개

AHP기법 분석을 수행하기위해 앞서 수행한 수량화Ⅱ 분석의 결과를 이용한다. 즉, 수량화Ⅱ 분석에서 선정된 6개의 중요항목에 대해 전문가들의 설문을 아래의 <표 4>와 같이 개발하였다. 다만 수량화 분석에서 선정되었던 Dry Density

항목 대신 이와 유사한 Water Contents가 사용되었다. 설문의 방법은 두 항목간의 비교를 모든 조합에 대해 실시하는 쌍대비교를 이용하였으며, 항목별 평가척도는 9점 척도를 이용하였다. 평가항목의 세부등급들 간의 설문도 이와 동일한 방식으로 구성하였다.

<표 4> 평가항목들 간 쌍대비교 설문지 (예시)

평가항목	Lithology	Elevation	Slope Angle	Porosity	Water Contents	Permeability
평가구분	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
질문	매우 중요 (5점)	(4점)	중요 (3점)	(2점)	동등 (1점)	...
Factor1이 Factor2보다 상대적으로 얼마나 중요하다고 생각하십니까?						
Factor1이 Factor3보다 상대적으로 얼마나 중요하다고 생각하십니까?						
⋮						
Factor5이 Factor6보다 상대적으로 얼마나 중요하다고 생각하십니까?						

3-2-2. 분석 결과

설문에 참여한 전문가는 총 9명으로 구성되었으며, 참여자 전원이 설문 에 응답하였다. AHP 가중치 산정법에 따라 중요항목에 대한 가중치를 선정하고, 세부항목에 대해서도 동일한 분석을 수행한 결과를 요약하면 <표 5>와 같다. AHP 분석은 최근 통계분석도구로 많이 이용되

는 R(<http://www.r-project.org>)을 이용하였다.

<표 5> 산사태 발생 위험도 평가표(AHP)

관찰항목	평가구분			
	가중치 (점수)	세부등급	가중치	점수
Lithology	0.110 (10)	Gneiss	0.384	10
		Mudstone	0.292	9
		Granite	0.324	6
Elevation (m)	0.092 (10)	431이상	0.478	10
		226-430	0.361	9
		225이하	0.161	6
Slope angle (°)	0.310 (32)	27.1이상	0.391	32
		23.1~27.0	0.247	24
		19.1~23.0	0.155	16
		15.1~19.0	0.113	8
		15.0이하	0.094	0
Porosity (%)	0.115 (12)	80.1이상	0.385	12
		75.1~80.0	0.249	9
		67.1~75.0	0.156	6
		46.1~67.0	0.119	3
		46.0이하	0.090	0
Water Contents (%)	0.168 (16)	40.0이상	0.423	16
		30.0~39.9	0.256	12
		20.0~29.9	0.158	8
		10.0~19.9	0.096	4
		9.0이하	0.067	0
Permeabil ity (cm/s)	0.204 (20)	6.0*10 ⁻³ ~ 1.0*10 ⁻²	0.266	20
		4.5*10 ⁻³ ~ 6.0*10 ⁻³	0.223	15
		2.6*10 ⁻³ ~ 4.5*10 ⁻³	0.180	10
		1.1*10 ⁻³ ~ 2.5*10 ⁻³	0.149	5
		1*10 ⁻³ 이하	0.180	0

IV. 성능평가 및 예측모형

이 절에서는 본 연구에서 제시된 두 종류의 산사태 위험도 평가표의 성능을 평가하고, 나아가 평가점수에 기초한 산사태 발생 예측모형을 개발한다. 평가표의 성능 비교를 위해서는 수량화Ⅱ 분석에 사용되었던 실제 관측자료 중 일부 ($n = 112$)가 사용되었다.

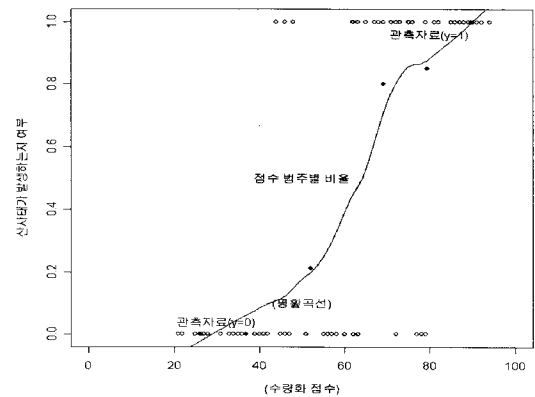
4-1. 수량화 평가표의 성능 및 예측모형

아래의 <표 6>은 실제 자료를 <표 3>의 수량화 평가표에 적용한 결과를 요약한 것이다.

<표 6> 수량화 점수에 따른 산사태 발생비율

수량화점수	미발생	발생	발생비율
0-10	-	-	-
11-20	-	-	-
21-30	10	-	-
31-40	19	-	-
41-50	8	3	27.3
51-60	9	-	-
61-70	4	7	63.6
71-80	4	19	82.6
81-90	-	23	100.0
91-100	-	6	100.0
합계	54	58	

위 결과를 그림으로 표현하면 <그림 3>과 같다. 그림 내의 곡선은 비율의 변화를 쉽게 파악하기 위한 평활곡선이다.



<그림 3> 수량화 점수와 산사태 발생비율

위 그림으로부터 수량화 점수가 높을수록 산사태의 발생비율이 커짐을 알 수 있다. 또한 수량화 점수에 따른 산사태 발생확률의 예측을 위해 로지스틱회귀모형을 적합한 결과는 다음과 같다.

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -9.0937 + 0.1460 \cdot \text{수량화점수}$$

즉, 산사태 자료에서 점수가 1점 늘어날 때 마다 산사태가 발생할 오르는 1.157배(=e^{0.1460}), 약 15.7%만큼씩 증가한다.

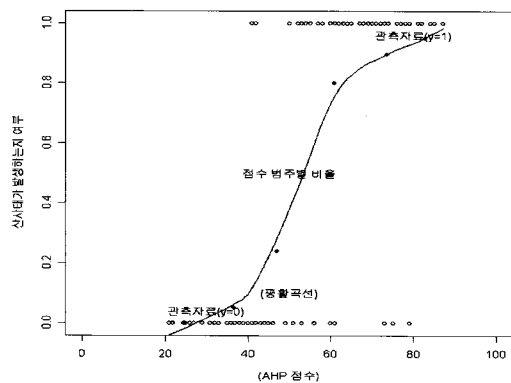
4-2. AHP 평가표의 성능 및 예측모형

아래의 <표 7>은 실제 자료를 <표 5>의 AHP 평가표에 적용한 결과를 요약한 것이다.

<표 7> AHP 점수에 따른 산사태 발생비율

AHP점수	미발생	발생	발생비율
0-10	-	-	-
11-20	-	-	-
21-30	13	-	-
31-40	16	-	-
41-50	15	3	16.7
51-60	5	8	61.5
61-70	2	15	88.2
71-80	3	16	84.2
81-90	-	16	100.0
91-100	-	-	-
합계	54	58	

위 결과를 그림으로 표현하면 <그림 4>와 같다.



<그림 4> AHP 점수와 산사태 발생비율

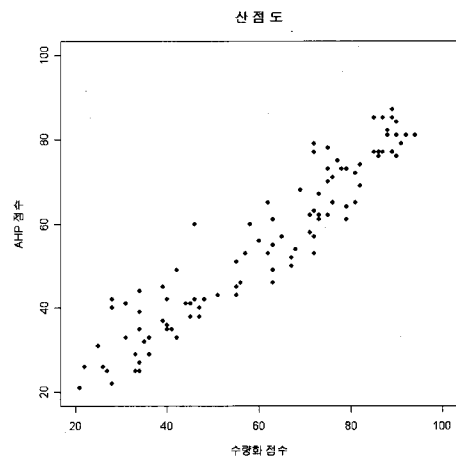
수량화 점수에 따른 산사태 발생확률의 예측을 위해 로지스틱회귀모형을 적합한 결과는 다음과 같다.

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -8.1976 + 0.1489 \cdot \text{AHP점수}$$

즉, 산사태 자료에서 점수가 1점 늘어날 때 마다 산사태가 발생할 오르는 1.161배(=e^{0.1489}), 약 16.1%만큼씩 증가한다.

4-3. 수량화II 방법과 AHP기법의 비교

수량화 점수와 AHP 점수 간의 상관분석을 통해 두 평가표의 일관성의 정도를 살펴보기로



<그림 5> 수량화 점수와 AHP점수 산점도

한다. 앞 절에서 사용된 자료를 두 평가표에 적용한 결과를 산점도로 나타내면 <그림 5>와 같다. 이 그림으로부터 두 평가표에 의한 점수 간에는 뚜렷한 상관이 나타나고 있다.

이로부터 두 평가점수간의 상관분석을 수행한 결과는 <표 8>과 같다. 두 평가점수 간의 상관계수는 0.952로 매우 높은 상관이 존재함을 알 수 있다. 이는 두 평가표에 의한 평가점수가 매우 유사한 값을 제공하여 일관성의 측면에서 우수함을 나타낸다.

<표 8> 수량화II와 AHP 점수간 상관분석

Pearson 상관계수	수량화 점수	AHP 점수
수량화 점수	1	.952**
AHP 점수	.952**	1

** : 유의수준 0.01에서 유의함.

V. 결 론

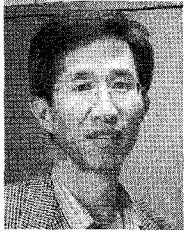
본 논문에서는 수량화Ⅱ 방법과 AHP 기법을 이용하여 산사태 발생에 영향을 미치는 주요변인들의 중요도를 점수로 산정한 산사태 발생위험도 평가표를 개발하고, 이로부터 구해진 평가점수를 통해 산사태 발생확률의 예측을 위한 통계적 모형을 구축하였다.

측정 데이터의 수량화 분석과 전문가 설문에 대한 AHP 분석에 의한 두 종류의 평가표가 제시되었다. 실제자료에 대해 제시된 평가표를 적용한 결과, 두 평가점수 간에는 매우 높은 상관관계가 존재(상관계수=0.952)하는 것으로 나타나 두 방법이 일관성 있는 분석임을 확인 할 수 있었다. 또한 산사태 발생 위험도 점수에 따른 산사태 발생 위험 확률을 통계적으로 모형화하여, 평가점수에 따른 산사태발생 예측확률을 추정할 수 있게 하였다.

본 논문을 통해 지금까지 산사태 발생여부의 예측을 경험적·주관적 평가 방법에서 자료에 기반한 과학적 평가 방법과 함께 여러 전문가의 의견을 종합한 객관적인 평가 방법을 동시에 제시하였다는 점에 의의를 둘 수 있다.

참 고 문 헌

- [1] 김기영·문권순·전명식, "SAS 상관분석", 자유아카데미, pp. 37-88, 1992.
- [2] 이용구·김용희, "질적자료의 수량화 분석방법에 관한 비교 연구", 통계논문집, Vol.-No.1, pp. 35-36, 1994.
- [3] 허명희, "수량화방법 I, II, III, IV", 자유아카데미, pp. 1-4, 31-57, 1992.
- [4] 황규승, "엑셀을 활용한 경영과학의 이해", 학현사, pp. 208-218, 2003.
- [5] Belton, V., Stewart, T.J., "Multiple Criteria Decision Analysis", An Integrated Approach. Kluwer Academic Publishers Group, pp. 372, 2002.
- [6] David R. Hardoon, Sandor Szedmak and John Shawe-Taylor, "Canonical correlation analysis ; An overview with application to learning methods" Department of Computer Science Royal Holloway, University of London, May 28, pp. 3-5, 2003.
- [7] Saaty, T. L., "The Analytic Hierarchy Process", McGraw-Hill, 1980.
- [8] Saaty, T. L., "How to make a decision : The Analytic Hierarchy Process", European Journal of Operational Research 48, pp. 9-26, 1990.
- [9] Saaty, T. L., "Rank from comparisons and from ratings in the analytic hierarchy/network processes", European Journal of Operational Research 168, pp. 557-570, 2006.



나 종 화 (Jong-Hwa Na)

- 정회원
- 서울대학교 계산통계학과 이학박사
- 공군사관학교 전산통계학과 전임강사
- 서울대학교 통계연구소 특별연구원
- (미) PSU 방문교수
- 현 충북대학교 정보통계학과 교수
- 관심분야 : 전산통계, 데이터마이닝, 수리통계



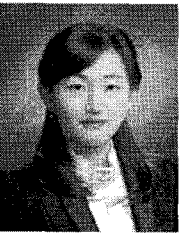
조 완 섭 (Wan-Sup Cho)

- 한국과학기술원 공학박사(전산학)
- 한국전자통신연구원 (ETRI) 연구원
- 미국 Univ. of Florida Post- Doc. 연구원
- 충북대학교 경영정보학과 부교수
- 관심분야 : Data Warehouse & OLAP, CRM, DB, BIT



유 혜 경 (Hye-Kyung Yu)

- 2005년 2월 : 충북대학교 통계학과 (이학학사)
- 2007년 2월 : 충북대학교 통계학과 (이학석사)
- 2009년 2월 ~ 현재: 충북대학교 정보통계학 (박사과정)
- 관심분야 : 데이터마이닝, 전산통계, 자료분석



남 은 미 (Eun-Mi Nam)

- 2007년 2월 : 충북대학교 통계학과 (이학학사)
- 2009년 8월 : 충북대학교 통계학과 (이학석사)
- 2009년 10월 ~ 현재: 식품의약품안정청
- 관심분야 : 데이터마이닝, 전산통계, 자료분석