

## 학업성취도에 대한 대입전형 요인들의 영향력 분석<sup>†</sup>

조장식<sup>1</sup>

<sup>1</sup>경성대학교 정보통계학과

접수 2010년 5월 29일, 수정 2010년 7월 12일, 게재확정 2010년 7월 16일

### 요약

본 논문에서는 부산 소재 K 대학교 신입생들의 학업성취도에 대해 신입생의 특성변수를 포함한 전형관련 변수들에 대한 영향력 분석을 연구한다. 이를 위해 모수적인 방법인 다중회귀분석과 비모수적인 방법인 의사결정나무 분석을 통하여 학업성취도에 대한 전형관련 변수들에 대한 주효과와 상호작용효과를 각각 분석하였다.

주요용어: 데이터마이닝, 의사결정나무모형, 학업성취도, 회귀분석.

### 1. 서론

대부분의 대학들은 21세기의 지식기반사회를 선도해 나갈 유능한 인재를 양성하기 위한 특성화되고 전문화된 방식으로 각 대학별 교육목표에 부합하는 신입생들을 선발하기 위해 다양한 전형방법을 개발하고 있다. 따라서 신입생 선발을 위한 전형방법이 적절하지 못하면 각 대학별 교육목표에 부합하는 우수학생을 선발하기 어렵게 된다. 이런 관점에서 볼 때, 실제 자료를 이용하여 각 전형방법의 타당성을 검증하기 위한 심도있는 연구가 필요하다.

허명희 (1996)는 입학선발자료와 대학학업성취도의 관련성을 분석하였으며, 한상태 (2000)는 대학입시 평가를 위한 통계적 방법에 대해 소개한 바 있다. 또한 최국렬 (2000)은 일반화선형모형을 이용하여 대학학업성취도에 미치는 요인을 분석하였고, 이종승 (2001)은 신입생의 입학성적과 대학성적간의 관계 분석을 하였다. 부재울 (2007a)은 대학입시 제도에서 전형자료들의 당락에 대한 영향력과 학업성취도에 예측력 변화에 대한 분석을 시도하였고, 또한 그는 (2007b) 학생부 반영방법을 중심으로 대학입시 전형 방법에 대한 심도있는 분석을 하였다. 그리고 조장식과 조길호 (2010)는 적성전형과 학업성취도와의 관련성을 보고한 바 있다.

본 논문에서는 대입 전형방법별 전형요소들이 학업성취도를 나타내는 평균평점에 대한 영향력을 분석하고자 한다. 이를 위해 모수적인 방법인 회귀분석을 통해서 각 전형방법별 독립변수들에 대한 주효과를 분석하고자 한다. 그러나 회귀분석 방법은 엄격한 가정 뿐만아니라, 다양한 범주형 독립변수들로 구성된 경우 이들에 대한 상호작용효과를 분석하는 데는 다소 한계가 있다. 따라서 본 논문에서는 이를 보완할 방법으로 비모수적인 방법인 의사결정나무 분석을 통하여 독립변수들에 대한 다양한 상호작용효과를 분석하고자 한다. 의사결정나무분석에 대한 연구로는 Park과 Song (2002)은 전공이탈자에 대한 예측모형을 구축하였으며, Kim 등 (2007)은 대입 지원자의 이탈자에 대한 예측모형을 구축하였다.

<sup>†</sup> 이 논문은 2010학년도 경성대학교 학술연구비지원에 의해 연구되었음.

<sup>1</sup> (608-736) 부산광역시 남구 대연동 110-1번지, 경성대학교 이과대학 정보통계학과, 교수.  
E-mail: jscho@ks.ac.kr

본 논문의 구성은 다음과 같다. 제 2장에서는 분석자료에 대한 소개와 간단한 기술통계를 보이고, 제 3장에서는 모수적인 방법인 회귀분석을 통해 독립변수들에 대한 주효과를 분석한다. 그리고 제 4장에서는 의사결정나무분석을 통해 독립변수들에 대한 상호작용효과를 분석하고, 마지막으로 제 5장에서는 결론을 제시한다.

## 2. 분석자료

본 연구에서 분석대상자는 2009학년도 부산소재 K 대학교에 수능전형, 인문계고전형, 일반전형 및 적성전형으로 입학한 신입생 중 1학기를 이수한 학생만을 대상으로 하였다. 분석에 사용된 총 자료 수는 2,565개이며, 사용된 변수들은 평균평점에 영향을 미칠 것으로 생각되는 개인특성변수를 포함한 입학관련 변수들로써, 각 변수들에 대한 설명은 다음 표 2.1과 같다.

표 2.1 변수 설명

변수명	항목
구분	전형방법
전형	수능전형=1, 인문계고전형=2, 일반전형=3, 적성전형=4
	평균평점
	2009학년도 1학기 평균평점
관련	언어영역
	언어영역의 백분위 점수
	수리영역
	수리영역의 백분위 점수
	외국어영역
	외국어영역의 백분위 점수
변수	내신성적1
	고교 1학년 내신성적 환산점수
	내신성적2
	고교 2학년의 내신성적 환산점수
개인	성별
	남자=1, 여자=0*
	계열
	예체능계열=1, 인문계열=2, 자연계열=0*
특성	재수여부
	재수=1, 재수안함=0*
변수	지역
	부산=1, 울산/경남=2, 기타=0*

\*: 회귀분석에서 기준범주를 의미함

위의 표 2.1에서 구분변수인 전형방법별로 평균평점을 종속변수로, 그 외 변수를 독립변수로 하여 분석하고자 한다.

아래 표 2.2는 전형방법별 교차표를 제시한 것이다. 표 2.2에서 알 수 있듯이, 일반전형이 40.9%로 가장 많고, 그 다음으로 인문계고전형 (28.7%), 수능전형 (15.6%), 적성전형 (14.9%)의 순으로 나타났다.

표 2.2 전형방법별 교차표 (단위: 명, %)

변수	구분	전형방법				전체
		수능	인문계고	일반	적성	
성별	남	15.4	29.6	37.7	17.3	1,260
	여	15.7	27.8	43.9	12.6	1,305
계열	예체능	5.8	13.9	78.0	2.3	395
	인문	15.0	30.1	38.5	16.4	1,284
	자연	20.8	33.2	27.8	18.3	886
지역	부산	17.3	25.8	41.5	15.4	2,071
	울산/경남	6.7	43.3	37.7	12.3	390
	기타	14.4	31.7	39.4	14.4	104
재수여부	재수안함	13.6	31.0	39.4	15.9	2,330
	재수	34.9	5.5	54.9	4.7	235
전체		15.6	28.7	40.9	14.9	2,565

아래 표 2.3은 평균평점에 대한 집단별 평균을 비교한 결과이다.

표 2.3 평균평점의 평균비교

변수	구분	N	평균	표준편차	F-값
성별	남	1,260	2.5450	1.0192	179.510***
	여	1,305	3.0414	0.8523	
계열	예체능	395	2.8624	0.9359	5.399***
	인문	1,284	2.8363	0.9388	
	자연	886	2.7124	1.0234	
지역	부산	2,071	2.7895	0.9754	0.426
	울산/경남	390	2.8232	0.9492	
	기타	104	2.8605	0.9464	
재수여부	재수안함	2,330	2.7787	0.9596	9.626***
	재수	235	2.9844	1.0523	
	수능	399	2.9091	0.9682	
전형방법	인문계고	736	2.8777	0.8736	58.622***
	일반	1048	2.9126	0.9419	
	적성	382	2.2109	1.0181	
합계		2,565	2.7975	0.9701	

\*, \*\*, \*\*\*는 각각 유의수준 0.1, 0.05, 0.01에서 통계적으로 유의함.

위의 표 2.3에서 알 수 있듯이, 지역을 제외한 성별, 계열, 재수여부, 전형방법에서 평균평점의 차이는 통계적으로 유의하게 나타났다. 특히 전형방법에서는 적성전형으로 입학한 학생들의 평균평점이 다른 전형으로 입학한 학생들의 평점보다 크게 낮음을 알 수 있다.

### 3. 회귀분석 결과

이 절에서는 각 전형방법별로 평균평점에 영향을 주는 전형관련 변수들에 대한 주효과를 분석하기 위해 표 2.1에서 제시한 변수들을 사용하여 다중회귀분석을 실시하였다. 이에 대한 결과는 아래 표 3.1과 같다.

표 3.1 전형방법별 회귀분석결과

독립변수	수능전형		인문계고 전형		일반전형		적성전형	
	$\beta$	s.e.	$\beta$	s.e.	$\beta$	s.e.	$\beta$	s.e.
상수	2.761***	0.4137	-20.993***	4.1979	-17.641***	3.232	-25.571***	8.3614
언어영역	0.005	0.0031	-0.002	0.0017	0.000	0.0016	0.001	0.0036
수리영역	0.006**	0.0027	0.005***	0.0015	0.004**	0.0015	0.004	0.0032
외국어영역	0.003	0.0034	0.008***	0.0019	0.007***	0.0018	0.004	0.0044
내신성적1	-	-	-0.050**	0.0195	-0.016	0.0182	-0.002	0.0359
내신성적2	-	-	0.043***	0.0059	0.032***	0.0054	0.038***	0.0110
남자	-0.567***	0.0967	-0.421***	0.0169	-0.400***	0.0586	-0.386***	0.1272
예체능계열	-0.197	0.2167	-0.023	0.1224	-0.070	0.0907	-0.207	0.4079
인문계열	-0.255**	0.1182	-0.193**	0.0784	0.347***	0.0774	-0.117	0.1554
재수안함	-0.178	0.1161	-0.300	0.2268	1.380***	0.2023	-1.382**	0.6132
부산	-0.154	0.2433	0.136	0.1494	-0.166	0.1414	0.018	0.2003
울산/경남	0.131	0.2996	0.235	0.1539	-0.162	0.1567	-	-
<i>n</i>	399		736		1,048		211	
<i>R</i> <sup>2</sup>	0.101		0.171		0.123		0.182	
<i>F</i>	4.857***		13.552***		13.205***		4.449***	

위의 표 3.1의 결과를 살펴보면 다음과 같다. 먼저 수능전형 (내신성적은 전형요소에 포함되지 않음)에서는 수리영역의 점수가 통계적으로 유의하게 평균평점에 영향을 미치며, 수리영역의 점수가 높을수

록 평균평점이 높아짐을 알 수 있다. 또한 여자에 비해서 남자가 0.567 정도가 통계적으로 유의하게 낮게 나타났으며, 자연계열에 비해서 인문계열이 0.255 정도 통계적으로 유의하게 낮게 나타났음을 알 수 있다.

다음으로 인문계고 전형은 수리영역, 외국어영역, 내신성적1, 내신성적2가 통계적으로 유의하게 평균평점에 영향을 미치며, 내신성적1을 제외한 점수들이 높아질수록 평균평점이 높아짐을 알 수 있다. 그러나 내신성적1은 오히려 평균평점에 음의 영향을 미치는 것으로 나타났다. 이런 결과는 인문계고 전형에서만 통계적으로 유의하게 나타났으며, 대입을 준비하는 수험생 입장에서 내신성적을 고교 2-3학년 때 적극적으로 관리하는 측면도 있겠지만, 내신성적을 반영하는 방법의 차이에 의해서 나타난 결과로 보여진다. 또한 여자에 비해서 남자가, 자연계열에 비해서 인문계열이 통계적으로 유의하게 평균평점이 낮은 것으로 나타났다.

다음으로, 일반전형은 수리영역, 외국어영역, 내신성적2가 통계적으로 유의하게 평균평점에 영향을 미치며, 이들의 점수가 높아질수록 평균평점이 높아짐을 알 수 있다. 또한 여자에 비해서 남자가 통계적으로 유의하게 낮으며, 자연계열에 비해서 인문계열이 통계적으로 유의하게 평균평점이 높은 것으로 나타났다. 또한 재수한 학생에 비해서 하지 않은 학생이 유의하게 높게 나타났다.

마지막으로 적성전형에서는 내신성적2가 높을수록 통계적으로 유의하게 평균평점이 높게 나타났으며, 여자에 비해서 남자가, 재수한 학생에 비해서 하지 않은 학생이 유의하게 평균평점이 낮게 나타났다.

#### 4. 의사결정나무 분석

실증연구에서 많이 사용하는 모수적 (parametric) 방법인 회귀분석은 독립변수들에 대한 주효과를 분석하는데는 효과적이지만, 상호작용효과를 분석하는데 한계가 있다. 따라서 이 절에서는 학업성취도에 영향을 주는 독립변수들의 상호작용효과를 분석하기 위해 비모수적 (non-parametric) 방법인 의사결정나무 (decision tree) 분석을 이용하고자 한다. 그러나 의사결정나무분석은 회귀분석과 달리 독립변수가 종속변수에 미치는 주효과를 파악할 수 없다는 한계점을 지니고 있으나, 독립변수들의 상호작용효과가 종속변수에 미치는 영향을 파악한다는 점에서 회귀분석과는 관점의 차이가 있으며, 회귀분석 결과를 보완할 수 있는 방법이 될 수 있다. 의사결정나무분석에 관한 연구로는 이용구과 허준 (2009)과 유영재와 박희창 (2009) 등이 있다.

의사결정나무분석에서 사용되는 변수들은 효과적인 상호작용효과 분석을 위해서 전형관련 변수들 (연속형 변수)인 평균평점, 수리영역, 언어영역, 외국어영역, 내신성적1, 내신성적2는 각각 상위 33%, 상위 66%를 기준으로 ‘상’, ‘중’, ‘하’로 분류하였다. 의사결정나무 분석에 사용된 변수들을 정리한 결과 아래 표 4.1과 같다.

의사결정나무분석을 위해 지니지수 (Gini index)를 분리기준으로 사용하였으며, 이지분리 (binary split)를 수행하는 CART (Classification And Regression Trees; Breiman 등, 1984) 알고리즘을 사용하였다. 정지규칙으로는 최대나무깊이 (maximum tree depth=5), 최소 케이스 수 (minimum number of cases: parent node = 20, child node=1)로 설정하였으며, 가지치기 (pruning)를 병행하였다. 의사결정나무 분석을 한 결과 아래 그림 4.1과 같다.

표 4.1 변수 설명

변수명	설명	항목
평점수준	2009년 1학기 평균평점	1=하, 2=중, 3=상*
전형방법	전형방법	1=수능, 2=인문계고, 3=일반, 4=적성
수리수준	수리영역백분위 점수수준	1=하, 2=중, 3=상
언어수준	언어영역백분위 점수수준	1=하, 2=중, 3=상
외국어영역	외국어영역 백분위 점수수준	1=하, 2=중, 3=상
내신수준1	1학년 내신성적 수준	1=하, 2=중, 3=상
내신수준2	2-3학년 내신성적 수준	1=하, 2=중, 3=상
재수여부	재수여부	1=재수안함, 2=재수함
성별	성별	1=남, 2=여
지역	거주 지역	부산=1, 울산/경남=2, 기타=3
계열	학과 계열	1=예체능, 2=인문, 3=자연

\*: 목표점수를 의미함

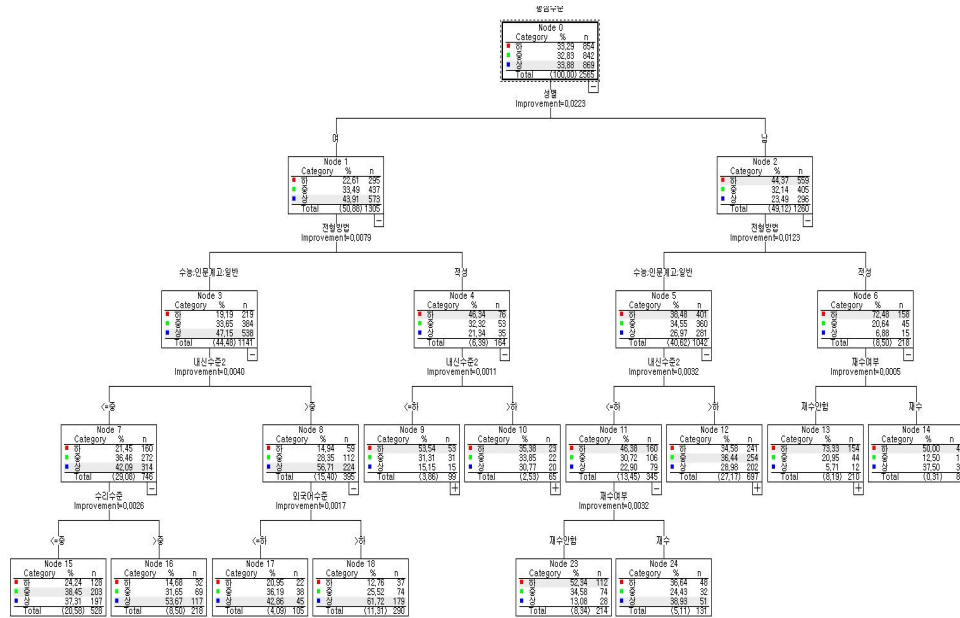


그림 4.1 의사결정나무분석 결과

먼저 목표점수를 기준으로 평점수준에 가장 많은 영향을 주는 중요변수는 성별, 전형방법, 내신수준2, 재수여부, 수리수준, 외국어수준 등의 순으로 나타났다. 먼저 남자 (23.49%)에 비해 여자 (43.91%)가 평점수준이 더 높으며, 남자와 여자 모두 적성전형에 비해 수능, 인문계고, 일반전형에서 평점수준이 높은 것으로 나타났다. 즉, 남자 (여자의 경우 적성전형 6.88% (21.34%)보다 수능, 인문계고, 일반전형 26.97% (47.15%)이 높게 나타났다. 또한 여자의 경우 수능, 인문계고, 일반전형과 적성전형 모두에서 내신수준2가 높을수록 평점수준이 높게 나타났으며, 남자의 경우 수능, 인문계고, 일반전형에서는 내신수준2가 높을수록 평점수준이 높게 나타났으며, 적성전형에서는 재수를 하지 않은 학생에 비해서 재수한 학생들의 평점수준이 높게 나타났다. 이와 같은 결과는 어느 정도 회귀분석의 결과와도 유사함을 알 수 있다.

이상의 결과를 이익도표를 통해서 정리한 것이 표 4.2이다. 먼저, Index (%)의 값이 가장 큰 노드는 182.2%인 18번 노드 (여자이면서 수능, 인문계고, 일반전형으로 입학하고 내신수준2가 ‘상’이고 외국어수준이 ‘중’이상인 집단)인데, 이 노드에 속하는 신입생들의 평점수준이 ‘상’인 비율이 전체 신입생들의 평점수준이 ‘상’인 비율보다 1.82배나 더 높음을 의미한다. 같은 방법으로, Index (%)의 값이 가장 작은 노드는 16.8%인 13번 노드 (남자이고 적성전형으로 입학하고 재수를 하지 않은 집단)인데, 이 노드에 속하는 신입생들은 평점수준이 ‘상’인 비율이 전체 신입생들의 평점수준이 ‘상’인 비율에 비해서 0.16배에 불과하다는 것을 의미한다. 또한 Index (%)의 값을 기준으로 볼 때, 16, 17, 24번 (23, 9, 12번) 노드 등에 속하는 신입생 집단의 평점수준이 ‘상’인 비율이 전체 신입생의 평점수준이 ‘상’인 비율보다 높은 (낮은) 수준을 보여주고 있다.

표 4.2 이익도표 결과

Nodes	Node: n	Node: %	Gain: n	Gain(%)	Resp(%)	index(%)
18	290	11.3	179	20.6	61.7	182.2
16	218	8.5	117	13.5	53.7	158.4
17	105	4.1	45	5.2	42.9	126.5
24	131	5.1	51	5.9	38.9	114.9
14	8	0.3	3	0.3	37.5	110.6
15	528	20.6	197	22.7	37.3	110.1
10	65	2.5	20	2.3	30.8	90.9
12	697	27.2	202	23.2	29.0	85.5
9	99	3.9	15	1.7	15.2	44.8
23	214	8.3	28	3.2	13.1	38.6
13	210	8.2	12	1.4	5.7	16.8

주:Nodes=의사결정나무에서 노드 (집단)번호, Node:n=해당 노드의 응답자 수,  
Node:%= 해당 노드의 응답자 수/전체 응답자 수, Gain: n = 해당 노드에서  
목표범주 수, Gain (%)=(해당 노드에서 목표범주 수)/(전체 목표범주의 수) × 100,  
Resp (%) = (해당 노드에서 목표범주 수)/(해당 노드의 응답자 수) × 100,  
Index (%) = Resp (%) / (전체 목표범주의 비율) × 100.

## 5. 결론

본 논문에서는 대입 전형방법별 전형요소들이 학업성취도를 나타내는 평균평점에 미치는 영향력을 분석하기 위해 모수적인 방법인 다중회귀분석을 통해서 주효과를 분석하였다. 또한 독립변수들의 상호작용 효과를 분석하기 위해 비모수적인 방법인 의사결정나무 분석을 사용 하였다. 먼저 회귀분석결과 전형방법별로는 수능전형, 인문계고 전형, 일반전형은 유사한 결과를 보였으나, 적성전형에서는 다른 결과를 보였다. 수능전형, 인문계고 전형, 일반전형에서는 수리영역, 외국어 영역, 내신성적2에서 평균평점에 유의한 영향을 미쳤으며, 성별과 계열에서 평균평점에 차이가 있는 것으로 나타났다. 또한 의사결정나무분석 결과에서는 여자이면서 수능, 인문계고, 일반전형으로 입학하고 내신수준2가 ‘상’이고 외국어수준이 ‘중’이상인 신입생 집단은 평점수준 ‘상’의 비율이 전체 신입생의 평점수준 ‘상’의 비율보다 특히 높았으며, 남자이고 적성전형으로 입학하고 재수를 하지 않은 신입생들은 평점수준 ‘상’의 비율이 전체 신입생들의 평점수준 ‘상’의 비율보다 특히 낮았다.

한편, 실제로 신입생들의 학업성취도를 나타내는 평균평점에 영향을 주는 요인들은 매우 많으나, 본 연구에서는 자료수집이 가능한 제한된 자료에 기초해서 분석된 결과이므로, 연구결과를 일반화시키기에 한계가 있다는 점을 밝힌다.

## 참고문헌

- 부재울 (2007a). 2008학년도 대학입시 제도에서 전형자료들의 당락에 대한 영향력과 학업성취도 예측력 변화에 대한 분석. <교육평가연구>, **20**, 57-79.
- 부재울 (2007b). 2008학년도 대학입시 전형방법에 대한 연구: 학생부 반영방법을 중심으로. <초등교육연구>, **18**, 275-292.
- 유영재, 박희창 (2009). 피에치피와 웨카를 이용한 데이터마이닝 도구의 설계 및 구현. <한국데이터정보과학회지>, **20**, 425-433.
- 이용구, 허준 (2009). 생명보험사의 개인연금 보험예측 사례를 통해서 본 의사결정나무 분석의 설명변수 축소에 관한 비교 연구. <한국데이터정보과학회지>, **20**, 179-190.
- 이종승 (2001). 충남대학교 신입생의 입학성적과 대학성적 간의 관계분석. <학생생활연구>, **28**, 1-22.
- 조장식, 조길호 (2010). 고전검사이론에 의한 적성전형 시험의 문항분석. <한국자료분석학회지>, **12**, 271-282.
- 최국렬 (2000). GLM을 이용한 대학학업성취도 분석. <응용통계연구>, **13**, 11-18.
- 한상태 (2000). 대학입시평가를 위한 통계적 분석방법. <한국자료분석학회지>, **2**, 25-36.
- 허명희 (1996). 입학선발지표와 대학학업성취도의 관련성. <교육학연구>, **34**, 299-309.
- Kim, K. I., Kim, S. H., Kim, E. Y., Kim, H., Yang, J. W. and Cho, J. S. (2007). Churn analysis for the first successful candidates in the entrance examination for K university. *Journal of Korean Data & Information Science Society*, **18**, 1-10.
- Park, C. and Song, G. M. (2002). Analysis of students leaving their majors using decision tree. *Journal of Korean Data & Information Science Society*, **13**, 157-165.

## The influence analysis of admission variables on academic achievements<sup>†</sup>

Jang Sik Cho<sup>1</sup>

<sup>1</sup>Department of Informational Statistics, Kyungsoong University

### Abstract

In this paper, we study the influence analysis of admission variables including their characteristics on academic achievements of freshmen at K university in Busan. First, multiple regression analysis is used to examine the main effects of admission variables including students' characteristics on the academic achievements. Also, Decision tree analysis is used to examine the interaction effects for the admission variables on the academic achievements. The results of this paper may be helpful to K university in designing effective admissions strategies for recruiting students.

*Keywords:* Academic achievements, admission, decision tree analysis, multiple regression analysis.

---

<sup>†</sup> This research was supported by Kyungsoong University Research Grants in 2010.

<sup>1</sup> Professor, Department of Informational Statistics, Kyungsoong University, Busan 608-736, Korea.  
E-mail: jscho@ks.ac.kr