

# 사회 연결망 분석을 이용한 복수전공 유입 및 유출 분석<sup>†</sup>

조장식<sup>1</sup>

<sup>1</sup>경성대학교 정보통계학과

접수 2012년 5월 30일, 수정 2012년 6월 27일, 게재확정 2012년 7월 2일

## 요약

각 대학마다 복수전공에 대한 선택의 폭이 넓어짐에 따라 학과별 복수전공자의 유입과 유출에 심한 편중현상이 나타나고 있다. 이런 현상의 결과로 특정 학과에서는 과도한 복수전공자의 유입(유출)으로 여러 가지 문제가 발생하고 있다. 따라서 본 논문에서는 사회연결망 분석과 의사결정나무 분석을 이용하여 학과별 복수전공자들의 유입과 유출에 대한 특성을 분석하였다. 분석방법으로 데이터 마이닝의 한 기법인 의사결정나무 모형을 활용하였으며, 분석결과에 대한 적절한 함의를 찾기 위해서 이지분리를 하는 CART 알고리즘을 사용하였다. 분석결과에 따르면, 복수전공 유입에 영향을 미치는 특성으로는 학과별 수능성적이 가장 많은 영향을 미치며, 그 다음으로 계열, 강의평가점수, 취업률의 순서로 나타났다. 한편 복수전공 유출에 영향을 미치는 특성으로는 계열이 가장 많은 영향을 미치며, 그 다음으로 수능성적, 취업률, 강의평가점수의 순서로 나타났다.

주요용어: 데이터마이닝, 복수전공, 사회연결망분석, 상호작용효과, 연결정도 중심성, 의사결정나무분석.

## 1. 서론

많은 대학에서 복수전공에 대한 선택의 폭을 넓힘으로써, 복수전공을 신청하는 재학생들이 해마다 증가하는 경향이 있다. 특정 학과에서 해당 학과 학생들이 타 학과로 복수전공을 하는 학생을 ‘복수전공 유출’이라 하고, 타 학과 학생들이 특정학과로 복수전공을 하기 위해서 신청한 학생을 ‘복수전공 유입’이라고 하자. 그러나 학과별로 복수전공자들의 유입과 유출에 심한 편중현상이 나타남으로써, 특정학과에서는 과도한 복수전공자들로 인한 강의실 배정, 분반, 추가적인 실험실습비 집행 등의 문제가 발생하고 있다. 따라서 학과별 복수전공자의 유입과 유출에 대한 심도있는 분석이 요구된다.

한편, 사회연결망분석이란 사회구조를 개체와 개체를 연결하는 링크로 구성되는 연결망으로 도식화하고 이들 간의 관계(상호작용)를 계량화하는 분석방법이다(Wasserman와 Faust, 1995). 방법론 측면에서 사회연결망 분석이 기존의 다른 이론들과의 차이점은 관심의 출발점이 ‘관계’에 있다는 것이다(Scott, 2000). 즉, 기존의 분석방법에서는 속성 데이터는 개체가 고유하게 가지고 있는 속성을 기초로 한 데이터로서, 개인적인 속성에 관련된 양적 데이터는 변수분석을 통해 데이터 분석이 이루어진다. 그러나 사회연결망분석에서 다루는 관계 데이터는 개인과 다른 개인(또는 조직)이 만드는 연결인 관계를 기초로 한 데이터로써, 연결망 분석을 통하여 데이터 분석이 이루어진다. 따라서 사회연결망 분석을 사용하면 다른 분석도구로는 파악하기 어려운 개체간의 상호작용 패턴 및 구조에

<sup>†</sup> 이 논문은 2012학년도 경성대학교 학술연구비지원에 의해 연구되었음.

<sup>1</sup> (608-736) 부산광역시 남구 수영로 309 번지, 경성대학교 이과대학 정보통계학과, 교수.  
E-mail: jscho@ks.ac.kr

관련된 정보를 얻을 수 있다. 사회연결망분석과 관련된 자세한 내용으로는 Kim (2007), Son (2010), Huh (2010) 등이 있으며, 특히 Choi 등 (2011)과 Choi와 Lee (2011)는 사회연결망 분석을 축구경기와 학제적 연구의 가능성에 적용 하였다.

본 연구에서는 모 대학교의 학과별 복수전공자들의 유입과 유출에 대한 자료를 이용하여 학과별 복수전공의 유입과 유출에 대한 연결망 구조를 분석하고자 한다. 또한 복수전공의 유입과 유출에 영향을 미치는 주요 변수를 분석하기 위해 독립변수들의 2차, 3차 이상의 상호작용효과를 분석하고자 한다. 이를 위해 데이터마이닝의 한 기법인 의사결정나무 (decision tree) 모형을 이용하고자 한다. 이와 관련된 보다 자세한 내용은 Kass (1980), Breiman 등 (1984), Biggs 등 (1991), Loh와 Shih (1997)와 SPSS Inc. (1988) 등이 있다.

본 연구의 구성은 다음과 같다. 제 2절에서는 사회연결망 분석의 중심성 척도인 연결성, 근접성, 매개성 척도에 대한 개념과 분석 데이터를 소개하였다. 제 3절에서는 사회연결망 분석을 이용하여 연결정도 중심성 지표에 대해 독립변수별 복수전공자의 유입과 유출을 비교 하였다. 제 4절에서는 의사결정나무분석을 통하여 연결정도 중심성 지표에 어떤 속성변수들이 유의하게 영향을 주는지를 분석하였다. 마지막으로 제 5장에서는 본 연구에 대한 결론을 제시하였다.

## 2. 사회연결망 분석의 개념

사회연결망분석은 수많은 개체들과 그 개체들 사이의 관계망 내에서 개체들이 얼마나 중요한 역할을 하는지 또는 연결 관계가 무엇을 의미하는지 등, 개체들 사이에 관련성을 찾기 위한 사회과학적 또는 통계학적 방법론이라고 할 수 있다. 사회연결망 분석에서 대표적인 척도는 중심성 (centrality) 이다.

중심성은 연결망 속에서 개체에 의해 지속되는 관계의 수를 계산함으로써 측정되는데, 중심성은 권력의 영향력이라는 개념과 연결되어 가장 많이 쓰이는 척도 가운데 하나이다. 구조적 위치에서 연결망의 중심부에 위치할수록 많은 교류와 다양한 상호작용을 할 수 있다는 가능성을 내포한다는 것을 의미한다. 특정 개체나 개인의 연결망 구조에서 구조적 위치가 중요한데, 중심성을 파악함으로써 이 구조적 위치를 계량화할 수 있다. 중심성은 한 개체가 얼마나 많은 다른 개체들과 연결되어 있는 가로 측정할 수도 있고, 한 개체가 다른 모든 개체들에 도달하려면 몇 단계나 필요한지로 측정할 수도 있다.

중심성을 측정하는 방법에는 연결정도 중심성, 근접 중심성, 그리고 매개 중심성이라는 세 가지 유형이 있다. 그러나 본 연구에서는 학과별 복수전공자들의 유입과 유출에 대한 연결망 구조를 파악하는 것이 목적이므로 중심성 척도 중에서 연결정도 중심성을 중심으로 분석하고자 한다.

연결정도 중심성은 다른 개체들과의 연결정도를 중심으로 보는 개념이다. 이는 한 개체에 연결된 다른 개체의 수로 측정이 된다.  $Z_{ij}$  는  $i$  개체로부터  $j$  개체로의 관계를 의미하며, 연결정도 중심성은 전체 연결 수에서 각 개체의 내향 연결정도 (indegree)와 외향 연결정도 (outdegree)의 비율로 다음과 같이 측정된다.

$$C_i = \sum_{j=1}^n (Z_{ij} + Z_{ji}) / \sum_{i=1}^n \sum_{j=1}^n Z_{ij}. \quad (2.1)$$

여기서  $Z_{ij}$ 의 값은 관계가 있을 경우 1, 없을 경우는 0의 값을 취한다.

본 연구에서 사용한 자료는 모 대학교에서 2011년 2월에 졸업한 졸업생 중에서 학과별로 복수전공을 이수한 학생들만을 대상으로 하였다. 그리고 학과별 복수전공의 유입과 유출에 대해서 사회연결망

분석을 위해서는 학과가 개체에 해당되며, 학과별 복수전공자의 유입과 유출이 개체간의 관계를 의미한다. 대학 전체 학과구조를 개체와 개체를 연결하는 링크로 구성되는 연결망으로 도식화하고 이들 간의 관계 (상호작용)를 계량화하여 분석하고자 한다. 본 연구에서 사용한 소프트웨어로는 Uncinet 홈페이지 (<http://www.analytictech.com/ucinet/>)에서 제공하는 Ucinet 6.0과 Minitab 16.0을 이용하여 분석하였다.

### 3. 사회연결망 분석

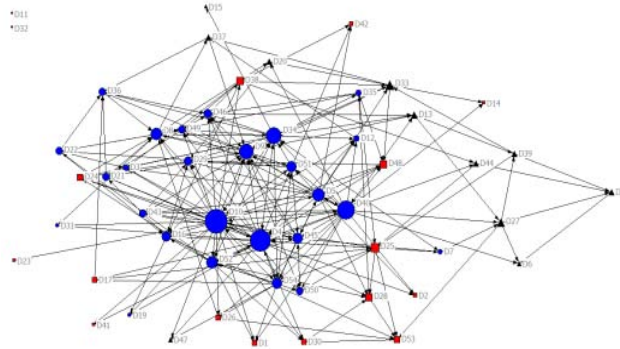
이 절에서는 복수전공의 유입과 유출에 대한 학과별 연결망 구조를 중심성 척도 중 연결정도 중심성을 중심으로 분석한다. 이를 위해 학과별 복수전공의 유입과 유출에 영향을 미치는 속성변수들로 계열, 평균 수능성적, 평균 강의평가점수, 취업률 등을 고려하였다. 여기서 복수전공의 유입과 유출은 중심성 척도에서 내향 연결정도와 외향 연결정도를 각각 의미한다. 해석의 용이성을 위해 수능성적, 강의평가점수, 취업률 등 연속형 변수에 대해서는 학과별 상위 25%에 해당하는 학과를 ‘상’, 하위 25%에 해당하는 학과를 ‘하’, 그 외에 해당하는 학과를 ‘중’으로 분류하였다. 각 변수별 복수전공의 유입과 유출에 대한 연결정도 중심성의 기술통계는 아래 Table 3.1과 같다.

**Table 3.1** Descriptive statistics for inflow and outflow of double majors

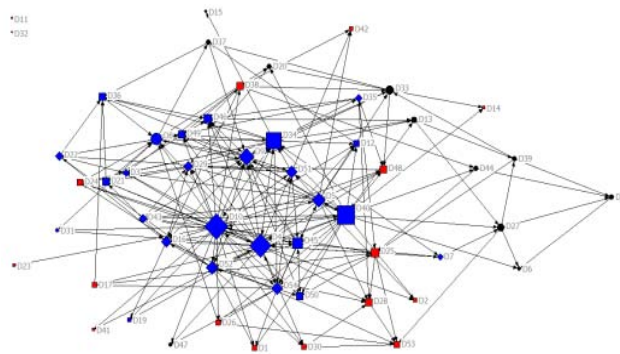
variables	categories	N	inflow		outflow	
			mean	s.d.	mean	s.d.
department category	humanity	26	1.78	2.31	1.51	1.20
	nature	16	0.22	0.34	0.55	0.35
	art	11	0.26	0.25	0.43	0.20
SAT	high	14	2.29	2.80	1.31	0.91
	middle	26	0.67	1.11	0.91	0.76
	low	13	0.25	0.24	0.83	1.43
course evaluation	high	13	0.31	0.37	0.76	0.70
	middle	26	1.37	2.29	1.17	1.18
	low	14	0.94	1.38	0.89	0.86
employment rate	high	13	0.80	1.19	0.83	0.70
	middle	26	1.38	2.29	1.04	0.87
	low	14	0.46	0.88	1.07	1.43
total		53	0.99	1.79	0.99	1.00

위의 Table 3.1의 결과로부터 다음과 같은 결과를 알 수 있다. 먼저, 계열별 (department category)로는 인문계열 (humanity)은 자연계열 (nature)과 예체능계열 (art)에 비해서 상대적으로 복수전공의 유입 (inflow)과 유출 (outflow)이 활발하게 이루어지고 있음을 알 수 있다. 그리고 유입이 유출보다 상대적으로 연결정도 중심성이 높음을 알 수 있다. 다음으로 수능성적 (SAT)은 ‘상 (high)’ 그룹에 속하는 학과들이 그렇지 않은 학과들에 비해서 상대적으로 복수전공의 유입과 유출이 활발하게 진행되고 있으며, 특히 유입이 유출보다 연결정도 중심성이 높음을 알 수 있다. 강의평가점수 (course evaluation)는 ‘중 (middle)’ 그룹에 속하는 학과일수록 상대적으로 복수전공의 유입과 유출이 활발하게 진행되고 있으며, 특히 유출이 유입에 비해서 높게 나타났다. 한편 취업률 (employment rate)은 낮은 (low) 학과일수록 유출이 상대적으로 높음을 알 수 있다. 다음으로 계열, 수능성적, 강의평가점수, 취업률과 같은 속성변수를 사용하여 복수전공 유입과 유출에 대한 사회연결망 분석을 하고자 한다. 아래 Figure 3.1 - Figure 3.4는 계열, 수능성적, 강의평가점수, 취업률 등 속성변수 각각에 따른 학과별 복수전공의 유입과 유출에 대한 사회연결망 구조를 나타낸 것이다. 여기서 노드의 크기는 연결정도 중심성의 크기의 정도를 의미하며, 노드의 모양 (또는 색깔)은 각 속성변수에 대한 구분

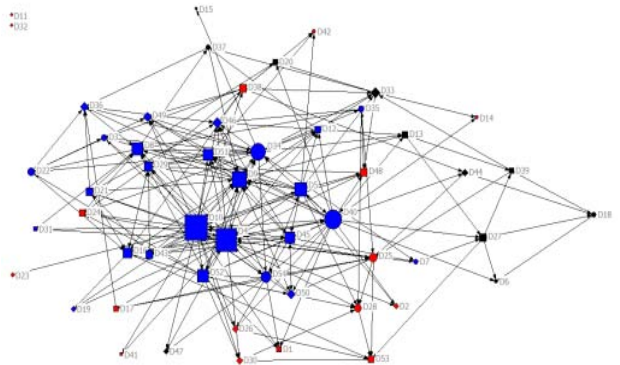
항목을 나타낸다. 그리고 그래프는 용수철 사영 (spring embedding) 방법으로 작성되었으며, 중심 위치에 있을수록 복수전공의 유입과 유출이 활발하게 진행되는 것을 의미한다.



**Figure 3.1** Inflow and outflow of double majors for department category (◇:high, □:middle, ○:low)



**Figure 3.2** Inflow and outflow of double majors for SAT score levels (◇:high, □:middle, ○:low)



**Figure 3.3** Inflow and outflow of double majors for course evaluation score levels (◇:high, □:middle, ○:low)

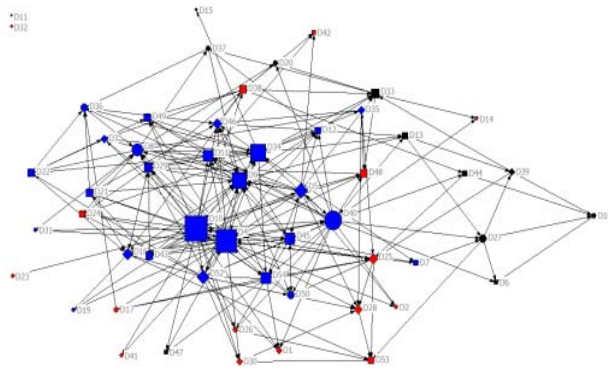


Figure 3.4 Inflow and outflow of double majors for employment rate (◇:high, □:middle, ○:low)

Figure 3.1-Figure 3.4의 결과에 따르면, 계열별 복수전공의 유입과 유출에 대한 연결망 구조에서 인문계열이 자연계열과 예체능 계열에 비해서 상대적으로 중심에 위치하고 있으며, 노드의 크기가 커서 복수전공의 유입과 유출이 활발한 것으로 보인다. 그리고 수능성적별 복수전공의 유입과 유출에 대한 연결망 구조에서, 수능성적이 ‘상’과 ‘중’인 학과들이 상대적으로 연결망 구조에서 중심에 위치하고 있고 노드의 크기가 커서 이들 학과 간에 복수전공의 유입과 유출이 활발한 것으로 나타났다. 다음으로 강의평가점수별 복수전공의 유입과 유출에 대한 연결망 구조에서, 강의평가점수가 ‘중’인 학과들이 상대적으로 중심위치에 있어서 이들 학과 간에 복수전공의 유입과 유출이 활발한 것으로 나타났으며, 취업률별 복수전공의 유입과 유출에 대한 연결망 구조에서는 취업률이 ‘중’과 ‘상’인 학과들 간에 복수전공의 유입과 유출이 활발한 것으로 나타났다.

#### 4. 의사결정나무 분석

3절에서는 계열, 수능성적, 강의평가점수, 취업률 각각에 대해서 복수전공의 유입과 유출에 대한 분석을 했다. 그러나 이러한 결과는 다른 속성변수들의 상호작용에 의해서 나타나는 복수전공의 유입과 유출의 특성을 분석하기에는 한계가 있다. 따라서 이 절에서는 복수전공의 유입과 유출에 대한 학과별 연결망 구조의 연결정도 중심성을 여러 속성변수들의 상호작용 효과를 중심으로 분석하고자 한다.

한편, 실증연구에서 많이 사용하는 회귀모형은 독립변수들에 대한 주효과를 분석할 경우에는 효과적이지만, 독립변수들의 2차, 3차 이상의 상호작용효과를 분석하는 것은 한계가 있었다. 따라서 이 절에서는 복수전공의 유입과 유출에 대한 연결정도 중심성에 영향을 주는 독립변수들의 고차 상호작용효과를 분석하기 위해 비모수적 방법인 의사결정나무 (decision tree) 모형을 이용하고자 한다. 그러나 의사결정나무분석은 독립변수가 종속변수에 미치는 주효과 (main effect)를 파악할 수 없다는 한계점을 지니고 있으나, 독립변수들의 고차 상호작용효과가 종속변수에 미치는 영향을 파악한다는 점에서 주효과를 주로 분석 대상으로 하는 회귀분석 결과를 보완할 수 있는 방법이 될 수 있다. 의사결정나무 분석을 활용한 연구로는 Cho (2010)와 Cho (2011) 등이 있다.

따라서 이 절에서는 의사결정나무 분석을 이용하여 복수전공의 유입과 유출에 대한 연결정도 중심성을 분석하고자 한다. 사용된 독립변수는 3절에서 사용한 계열, 수능성적, 강의평가점수 및 취업률 등 속성변수와 동일하다. 의사결정나무분석을 수행하기 위해 지니지수 (Gini index)를 분리기준으로 사

용하였으며, 이진분리를 수행하는 CART (Classification And Regression Trees; Breiman 등, 1984) 알고리즘을 사용하였다. 정지규칙으로는 최대나무깊이 (maximum tree depth=3), 최소 케이스 수 (minimum number of cases; parent node = 10, child node=1)로 설정하였으며, 가지치기 (pruning)를 병행하였다.

복수전공 유입과 유출에 대한 의사결정나무분석 결과는 아래 Figure 4.1과 Figure 4.2에 제시하였다.

먼저, 복수전공 유입에 대한 의사결정나무분석의 결과는 아래 Figure 4.1과 같다. Figure 4.1의 결과에 따르면, 복수전공의 유입에 가장 영향을 많이 미치는 변수로는 학과별 수능성적이고, 그 다음으로 계열, 강의평가점수, 취업률의 순으로 나타났다. 먼저 학과의 수능성적이 '상'인 그룹은 2.291로 수능성적 '중'과 '하'인 그룹 0.529보다 복수전공의 유입이 높게 나타났다. 그리고 수능성적이 '중'과 '하'인 학과들 중에서는 인문계열이 1.169로 자연계열과 예체능계열의 0.236보다 복수전공 유입이 높게 나타났다. 또한 이들 자연계열과 예체능 계열 중에서 학과 취업률이 '상'인 학과 그룹이 0.286으로 취업률이 '중'과 '하'인 학과 그룹의 0.218 보다 복수전공의 유입이 상대적으로 높게 나타났다.

한편 수능성적이 '상'인 학과 그룹 중에서 강의평가점수가 '하'인 학과 그룹이 0.755인 반면, 강의평가점수가 '중' 또는 '상'인 학과 그룹은 3.146으로 복수전공의 유입이 매우 높음을 알 수 있다.

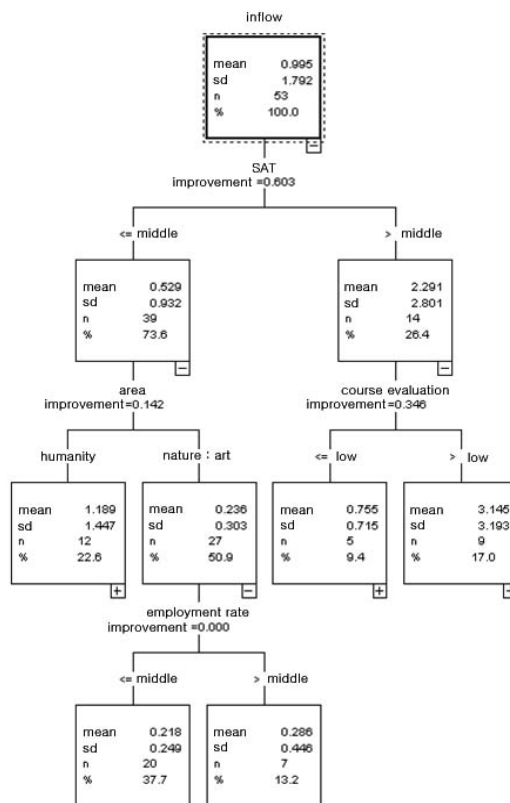


Figure 4.1 Decision tree analysis for inflow of double majors

다음으로, 복수전공 유출에 대한 의사결정나무분석 결과는 아래 Figure 4.2와 같다. Figure 4.2의 결과에 따르면, 복수전공 유출에 가장 많은 영향을 미치는 변수로는 계열이며, 그 다음으로는 수능성적, 취업률, 강의평가점수의 순으로 나타났다. 먼저 계열이 인문계열인 경우 1.506으로 자연계열과 예체능계열의 0.502보다 높게 나타났다. 특히 인문계열 중에서는 수능성적이 ‘하’인 그룹인 경우 3.007로서, 수능성적이 ‘중’과 ‘상’인 그룹의 경우 1.381보다 높게 나타나서 복수전공 유출이 상대적으로 높은 것으로 나타났다. 그리고 수능성적이 ‘중’과 ‘상’ 그룹 중에서는 취업률이 ‘중’과 ‘하’인 그룹이 1.448로 취업률이 ‘상’인 그룹의 1.180보다 높게 나타나서 복수전공 유출이 높은 것으로 나타났다. 한편 계열이 자연계열과 예체능계열 중에서는 자연계열이 0.553으로 예체능계열 0.429보다 다소 높게 나타나서 복수전공 유출이 상대적으로 높음을 알 수 있다. 특히 자연계열 중에서는 강의평가점수가 ‘하’인 학과 그룹이 0.825로 강의평가점수가 ‘중’과 ‘상’인 학과그룹의 0.490보다 상대적으로 높게 나타나서 복수전공의 유출이 높음을 알 수 있다.

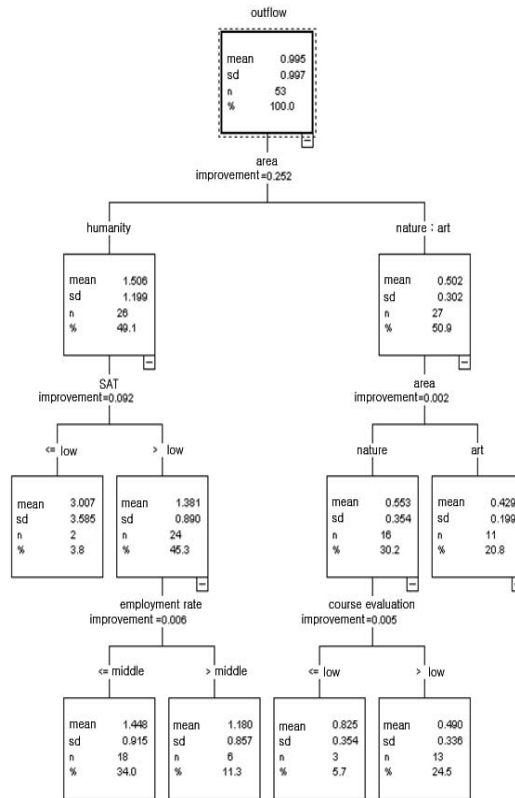


Figure 4.2 Decision tree analysis for outflow of double majors

## 5. 결론

본 논문에서는 사회연결망 분석과 의사결정나무분석을 이용하여 학과별 복수전공의 유입과 유출에 대한 특성을 분석하여 아래와 같은 결과를 얻었다.

먼저 복수전공 유입에 영향을 미치는 특성으로는 수능성적이 가장 많은 영향을 미쳤으며, 그 다음으로 계열, 강의평가점수, 취업률의 순서로 나타났다. 수능성적이 ‘상’이고 강의평가점수가 ‘중’과 ‘상’인 경우 3.145로 복수전공 유입이 특히 높음을 알 수 있었다. 그리고 수능성적이 ‘중’과 ‘하’이면서 자연 및 예체능계열이고, 취업률이 ‘중’과 ‘하’인 학과그룹이 0.218로 특히 복수전공 유입이 낮음을 알 수 있었다.

다음으로 복수전공 유출에 영향을 미치는 특성으로는 계열이 가장 많은 영향을 미쳤으며, 그 다음으로 수능성적, 취업률, 강의평가점수의 순서로 나타났다. 인문계열이면서 수능성적이 ‘하’인 학과그룹이 3.007로 복수전공 유출이 특히 높으며, 자연계열이면서 강의평가점수가 ‘중’과 ‘상’인 경우 0.336으로 복수전공 유출이 특히 낮음을 알 수 있었다.

이상의 결과로부터 복수전공의 유입과 유출에 대한 분석결과를 바탕으로 해당 부서에서 복수전공자의 편중현상으로 나타날 수 있는 문제점을 해결하는데 기초자료로 활용될 수 있다. 또한 본 연구에서는 학과별로 수집 가능한 변수들만을 분석대상으로 했다는 측면에서, 복수전공의 선택여부에 영향을 미치는 다양한 중요한 변수가 누락 되었다. 따라서 본 연구의 결과를 지나치게 일반화하는 것은 한계가 있다. 따라서 복수전공의 유입과 유출에 영향을 미치는 다양한 자료를 바탕으로 심도 있는 분석은 향후의 과제로 남긴다.

## 참고문헌

- Bigg, D., de Ville, B. and Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, **18**, 49-62.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and regression trees*, Wadsworth, Belmont.
- Cho, J. S. (2010). The influence analysis of admission variables on academic achievements. *Journal of the Korean Data & Information Science Society*, **21**, 729-736.
- Cho, J. S. (2011). Determinants of job finding using student's characteristic information. *Journal of the Korean Data & Information Science Society*, **22**, 849-856.
- Choi, K. H. and Lee, Y. W. (2011). The deduction of objective linguistic information using statistical methods- The grouping of the possibility of interdisciplinary research. *Journal of the Korean Data & Information Science Society*, **22**, 49-55.
- Choi, S., Kang, C., Choi, H. and Kang, B. (2011). Social network analysis for a soccer game. *Journal of the Korean Data & Information Science Society*, **22**, 1053-1063.
- Huh, M. H. (2010). *Introduction to social network analysis using R*, Freedom Academy, Seoul.
- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, **29**, 119-127.
- Kim, Y. H. (2007). *Social network analysis*, Pakyoungsa, Seoul.
- Loh, W. and Shih, Y. (1997). Split selection methods for classification trees. *Statistica Sinica*, Taiwan.
- Scott, J. (2000). *Social network analysis : A handbook*, Sage Publications, London.
- Son, D. W. (2010). *Social network analysis*, Kyungmoon Publishers, Seoul.
- SPSS Inc. (1988). *AnswerTree 1.0 user's guide*, SPSS Inc., Chicago.
- Wasserman, S. and Faust, K. (1995). *Social network analysis : Method and applications*, Cambridge University Press, Cambridge.



## Inflow and outflow analysis of double majors using social network analysis<sup>†</sup>

Jang Sik Cho<sup>1</sup>

Received 30 May 2012, revised 27 June 2012, accepted 2 July 2012

### Abstract

Recently, the number of students who get double majors has tended to increase in many universities. As results, many problems occur because immoderate inflow of double-major students is concentrated in a specific popular department. In this paper, we study the characteristic of inflow and outflow of double majors using social network analysis and decision tree analysis. According to the results, SAT score affected the inflow of double majors the most. Additionally, department category, course evaluation score, employment rate also affected the inflow of double majors in the order named. On the other hand, department category affected the outflow of double majors the most. Additionally, SAT score, employment rate, course evaluation score also affected the outflow of double majors in the order named.

*Keywords:* Centrality, data mining, decision tree analysis, degree, double majors, interaction effect, social network analysis.

---

<sup>†</sup> This research was supported by Kyungsung University Research Grants in 2012.

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