

연관규칙 마이닝을 이용한 한국 신노년층의 생활 만족도에 관한 연구

임 영 희 · 이 종 욱 · 박 대 희(고려대학교 컴퓨터정보학과)
장 진 경(숙명여자대학교 가정아동복지학부)

I. 서론

한국은 세계에서 그 유래가 없는 급속한 고령화를 경험하고 있다. 지난 2000년 65세 이상 노인인구 집단이 7%를 넘어서 고령화 사회에 진입한 이후 2006년 현재 9.5%로 고령화 속도가 매우 빠르게 진행되고 있다(통계청, 2006). 이렇게 급증하는 노인인구에 대한 부양문제는 ‘연금재정 파탄과 삶의 질 저하’, ‘국가경쟁력 저하’ 등의 결과를 초래할 ‘시한폭탄’, ‘재앙’ 등 부정적으로 표현되어 왔다(한경혜, 윤성은, 2007). 그러나 최근 한국에서도, 육체적으로나 정신적으로 건강한 노인이 증가하고 있고, 경제적으로 자녀에게 전적으로 의존하지 않고 연금, 저축, 부동산 수입 등 자산으로 경제력을 확보한 노인이 증가하고 있다(박영순, 2006). 이들은 전통적인 노인에게 비해 개방적이고 모험적인 성향을 가지고 있으며, 건강과 사회적 삶에 대한 관심이 높은 것으로 나타났다(Mathur, 1998). 기존의 노인층이 노년기를 ‘인생의 종말기’로 보는 반면, 신노년층은 이 시기를 ‘자기실현의 기회’, ‘제 3의 인생’이라 여긴다.

현재까지 한국 노인의 삶의 질에 관한 연구들은 부양 대상으로서의 노인들의 삶의 질에 대해 연구되어 왔지만, 신노년층이라 불리우는 향상된 교육수준과 높은 경제력을 갖고 있는 중산층 이상의 노인들의 삶의 질에 관한 연구는 거의 전무한 실정이다. 따라서 본 연구에서는 신노년층의 생활 만족도에 대한 특성을 파악하고자 한다.

일반적으로 사회과학분야에서 하나의 현상을 파악하기 위해 사용하는 양적연구 방법의 하나인 설문조사 분석은 1) 해당 변인에 대한 통계적인 분포를 분석하거나 2) 두 변인간의 연관성을 분석하거나 3) 연구모형을 확립한 후, 변인들 간의 인과관계에 대한 가설들을 검증하는 형태로 이루어져왔다. 그러나 하나의 변인에 대한 빈도 분석은 분포나 중앙경향치, 산포도 등의 기초적인 통계 정보만을 제공하며, 두 변인간의 연관성 분석만으로는 다양한 변인 간의 복합적인 상관성을 찾아내기가 쉽지 않다. 또한, 연구 모형에서 설정된 가설을 검증하는 방식은 이론 및 선행연구를 바탕으로 가설된 연구 모형이 전체 데이터의 속성을 잘 반영하고 있는지를 검증하는 것이며, 가설의 적절성과 유효성에 의해 분석 결과가 좌우된다는 문제점을 안고 있다. 따라서 본 논문에서는 사전에 가설을 설정하고 이를 검증하는 방식에서 벗어나, 데이터 마이닝 기법을 이용하여 사전지식 없이 데이터 내에 내재된

어 있는 유용한 지식을 찾고자 한다. 이를 위해 데이터 마이닝의 연관 규칙 분석을 통해 각 항목 간에 내재되어 있는 연관관계를 분석하고, 이를 이용하여 신노년층의 생활만족도에 대한 이론 모형을 제시하고자 한다. 이와 함께 다차원 연관규칙(association rule) 분석을 통해 다양한 항목간의 복합적인 상관성을 도출하고자 한다.

데이터마이닝은 대용량의 데이터에 내재되어 있는 유용한 지식을 찾아내는 기법으로서, 고객관계 관리(CRM: Customer Relationship Management), 금융, 서비스 통신업 등의 다양한 분야에 활용되고 있으며, 최근에는 설문 분석에서도 효과적으로 이용되고 있다(김완섭, 이수원, 2006; 이정수, 김교정, 2003; 김영숙, 정국인, 박소라, 2008; 김혜숙, 문양세, 노희영, 김진호, 2006; Nayak, Buys, Lovie-Kitchin, 2006; Pitt, Nayak, 2007). 본 연구에서는 데이터 마이닝의 대표적인 분석 기법인 연관규칙 분석을 통해 신노년층의 생활만족도에 영향을 미치는 가족지지, 수행능력, 욕구충족, 생산적 활동 간에 내재되어 있는 연관관계를 찾고자 한다. 본 연구에서는 WEKA 3.5.7을 이용하여 연관규칙 분석을 수행하였다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터 마이닝의 연관규칙 분석에 대해 간략히 소개하고, 3장에서는 분석의 대상이 되는 신노년층의 설문 자료와 연관규칙 분석 결과에 대해 기술한다. 4장에서는 결론 및 향후 연구과제에 대해 논한다.

II. 데이터 마이닝의 연관규칙 분석

데이터 마이닝은 대용량의 데이터로부터 내재되어 있는 유용한 지식을 찾아내는 데이터 분석 기법으로, 기존의 통계 분석이나 OLAP(On-Line Analytical Processing), 데이터베이스의 질의어로는 찾아내기 어려운 숨겨진 지식들을 추출하기 위하여 사용된다. 기존의 방법들은 일반적으로 가설이 있는 상황에서 통계 분석을 통해 지식을 검증하는 형태를 취하므로, 분석자가 가설을 찾지 못할 경우 데이터 내에 숨겨져 있는 유용한 지식도 찾아낼 수가 없다. 반면, 데이터 마이닝의 경우에는 가설을 알지 못하는 상황에서도 다양한 분석 기법을 통해 내재되어 있는 유용한 지식을 찾아낼 수 있다는 장점을 가지고 있다(김완섭, 이수원, 2006). 대표적인 데이터 마이닝의 분석 방법으로는 분류(classification) 분석, 군집(clustering) 분석, 연관규칙 분석 등이 있다.

본 연구에서 사용되는 연관규칙 분석이란 데이터 내에 존재하는 각 항목들 간의 연관관계를 찾아내는 작업이다. 예를 들면, 연관규칙 마이닝을 통해 소매점에서 기저귀를 구매하는 고객들은 맥주도 동시에 구매한다는 연관 규칙을 찾아낼 수 있으며, 연관 규칙은 다음과 같이 표현된다. “기저귀 ⇒ 맥주[10% of support, 80% of confidence]”. 여기서 10%의 지지도(support)라는 것은 주어진 소매점의 전체 구매내역 데이터 중, 10%가 기저귀와 맥주를 동시에 구매한다는 것을 의미하며, 80%의 신뢰도(confidence)라는 것은 기저귀를 구매한 고객 중, 80%가 맥주를 함께 구매한다는 것을 의미한다.

1. 연관규칙의 정의

먼저 항목들의 집합 $I = \{I_1, I_2, \dots, I_m\}$ 와 각각의 트랜잭션 T 에 대해 $T \subseteq I$ 의 관계를 가진 항목들의 집합이 있을 때, 각각의 트랜잭션 T 는 고유한 트랜잭션 구분자(transaction identifier)를 갖는다.

A 를 항목들의 집합이라고 하면, 트랜잭션 T 가 필요충분조건으로 $A \subseteq T$ 를 만족하는 경우에만 트랜잭션 T 가 항목 A 를 포함한다고 한다. 여기서 $A \subseteq I, B \subseteq I, A \cap B = \emptyset$ 을 만족하는 경우, 연관규칙은 $R:A \Rightarrow B$ 와 같이 조건 명제의 형태로 표현되며, A 를 규칙의 조건부(antecedent), B 를 결과부(consequent)라 한다. 추출된 연관규칙의 평가 기준으로는 지지도(support)와 신뢰도(confidence)를 사용한다. 지지도란 해당 규칙이 전체 데이터 집합에 얼마나 자주 나타나는가를 의미하며, 확률 $P(A \cup B)$ 를 계산함으로써 얻을 수 있다. 신뢰도란 규칙의 조건부가 일치할 경우, 해당 결과부가 도출될 확률을 의미하며, 조건부 확률 $P(B | A)$ 를 계산함으로써 얻을 수 있다.

$$\text{support}(A \Rightarrow B) = P(A \cup B)$$

$$\text{confidence}(A \Rightarrow B) = P(B | A)$$

따라서 연관규칙이 데이터의 집합에 대해 적절해지려면, 충분한 지지도와 신뢰도를 가져야 한다.

2. 연관규칙 탐사의 기본적인 접근 방식

주어진 데이터의 집합에서 추출되는 연관규칙은 사용자가 정의한 최소 지지도와 최소 신뢰도 이상의 값을 가져야 한다. 연관규칙 마이닝은 기본적으로 다음의 두 단계로 구성 된다:

- 1단계 : 사용자가 정의한 최소 지지를 만족하는 데이터 항목의 집합(빈발 항목 집합: frequent item set)을 탐사하는 단계이다. 이 단계에서는 각각의 데이터 항목에 대한 지지를 계산하여 미리 정의된 최소 지지를 만족하는 데이터 항목 집합들을 추출해 낸다.
- 2단계 : 연관 규칙을 생성하는 단계이다. 이 단계에서는 1단계에서 탐사된 빈발 항목 집합을 이용하며, 데이터의 부분 집합에서 생성된 규칙 중 사용자가 정의한 최소 신뢰도를 만족하는 규칙들을 최종 규칙으로 추출해 낸다.

본 연구에서는 신노년층의 설문 분석에 연관규칙 분석을 사용함으로써, 가족지지, 수행능력, 욕구 충족, 생산적 활동이 신노년층의 생활만족도에 미치는 영향 뿐 아니라 각 항목간의 복합적인 연관성까지 찾아내고자 한다.

III. 연관규칙 분석을 이용한 신노년층의 설문 분석

1) 신노년층 설문 데이터

본 연구의 조사대상은 서울시 및 수도권에 거주하고 있는 60세 이상의 남녀노인으로, 한국의 신노년층을 대표할 수 있는 중산층 이상의 노인을 대상으로 하였다. 중산층의 범위를 어떻게 설정할 것인가는 논란의 대상이 되고 있지만, 본 연구에서는 기존 연구를 바탕으로 다음의 변수들을 고려하여 중산층 이상의 노인을 선별하였다. 1) 거주형태는 자가이거나 최소한 전세 독체에 거주하여야 한다. 2) 은퇴 전 직업 또는 현직의 직업적 특성이 전문직·사무직·관리직이거나 직업 내 지위가 과장 이상이어야 한다. 3) 스스로를 중산층이라고 여기는 중산층 귀속의식(소속감)을 가지고 있어야 한다.

본 연구의 예비조사는 2006년 7월에서 8월에, 본 조사는 2006년 12월에서 2007년 1월에 걸쳐 실시하였으며, 총 206명의 응답자 중, 무응답 항목이 없는 190명의 설문 데이터를 최종 분석 자료로 사용하였다.

2) 주요 변수의 측정도구

한국 신노년층의 가족지지, 수행능력, 욕구충족, 생산적 활동, 삶의 질 등 5가지 주요 변수들에 대한 측정도구의 내용은 다음과 같다

가족지지는 지지의 제공자에 따라 배우자와 자녀가 주는 정서적 지지 및 정보적 지지의 문항으로 구성하였다. 정보적 지지와 정서적 지지 문항은 Seeman등(1996)의 연구에서 활용된 것을 성혜영(2005)이 한국어로 번안한 문항을 사용하였다. 정서적 지지는 사랑과 보호를 받고 있다는 느낌과 자신의 고민을 잘 들어 준다는 느낌을 말하여, 정보적 지지는 필요한 일에 대해 구체적인 도움을 받거나 정보를 제공받는 것을 말한다. 각 문항은 5점 척도로 '전혀 그렇지 않다' 1점, '매우 그렇다' 5점을 부여하여 점수가 높을수록 배우자, 자녀로부터의 지지가 높은 것을 의미한다.

수행능력은 일상생활을 영위하는데 필요한 신체적 기능과 인지 기능으로 구분하여 문항을 구성하였다. 본 연구에서는 성혜영(2005)이 Row와 Kahn의 성공적 노화모델 검증에 사용한 문항을 수정·보완하여 사용하였다. 각 문항은 5점 척도로 '전혀 아니다' 1점, '매우 그렇다' 5점을 부여하여 점수가 높을수록 수행능력이 떨어지고 있다고 인식하고 있음을 의미한다.

욕구충족은 중상층 노인이 삶의 영역들에 대해 부여하는 상대적 중요성을 파악할 수 있도록 하였다. 욕구는 생존 욕구, 사회적 욕구, 자아 욕구, 자아성취 욕구의 4가지로 구분된다. 욕구에 해당하는 문항은 이수애(2005)의 Maslow 욕구이론을 참조하여 개발한 문항을 사용하였다. 각 문항은 5점 척도로 '전혀 아니다' 1점, '매우 그렇다' 5점을 부여하여 점수가 높을수록 생존, 사회적, 자아, 자아성취 욕구가 충족되고 있음을 의미한다. 각 문항 중 부정적으로 되어 있는 문항(총 11문항)은 역코딩하여 사용하였다.

생산적 활동은 유급노동 뿐만 아니라 자원봉사활동이나 단체(모임)활동도 포함된다. 본 연구에서는 유급노동보다는 무급노동에 초점을 맞추어 자원봉사, 학습활동, 단체 활동으로 구분하여 각 활동 영역에 얼마나 참여하는지를 '전혀 하지 않는다'에서 '매일 한다'까지 6점 척도로 구성하였다.

생활만족도는 현재 자신의 삶에 대한 주관적인 만족 정도를 의미하며, 본 연구에서는 Diener, Emmons, Larsen과 Criffin(1985)이 개발한 SWLS(satisfaction with life scale)을 성혜영(2005)이 한국어로 번안한 것을 사용하였다. 5점 척도로 '전혀 아니다' 1점, '매우 그렇다' 5점을 부여하여 점수가 높을수록 생활만족도가 높음을 의미한다.

연관규칙 분석을 위해 가족지지, 수행능력, 욕구충족, 생산적 활동, 생활만족도 등의 변수들은 소속 문항의 평균값을 계산한 후, 세 개의 범위로 나누어 만족, 보통, 불만족으로 레이블링 하였다.

3) 조사 대상자의 인구학적 특성

조사대상자의 인구학적 특성은 다음 <표 1>과 같다. 전체 조사대상자 206명 중 '남자노인'(53.9%)이 '여자노인'(46.1%)보다 많았으며, 교육수준은 '대졸'이 46.3%로 가장 많았다. 결혼 상태는 대다수가 '초혼'(88.3%)이었으며, 종교는 '기독교'(51.0%), '불교'(15.0%)순으로 종교가 있는 경우가 없는

경우보다 더 많은 것으로 나타났다. 자녀수는 ‘2명’(50.5%), ‘3명’(26.5%)순으로 2~3명의 자녀를 두고 있었으며, 현재 ‘부부만 산다’(55.9%), ‘미혼자녀와 동거한다’(26.0%)의 가족형태가 대다수(81.9%)로, 아들 부부 또는 딸 부부와 함께 사는 3세대 가족형태가 매우 낮음을 알 수 있다. 현재 연령대는 90.8%가 60대이며, 은퇴 이전 또는 현재 직업은 초등학교 교사 65점을 기준으로 그 이상의 직업을 가졌거나, 현재 갖고 있는 것으로 나타났다. 현재 주관적인 계층의식을 살펴본 결과, 70.8%가 자신을 ‘중산층’이라고 여기고 있는 것으로 나타났다.

<표 1> 조사대상자의 사회인구학적인 특성

변 인	구 분	N(%)	변 인	구 분	N(%)	
성 별	남	111(53.9)	동거가족	혼자 산다	9(4.4)	
	여	95(46.1)		부부만 산다	114(55.9)	
교육수준	초졸 이하	2(1.0)		아들부부와 동거	18(8.8)	
	중 졸	9(4.4)		딸부부와 동거	5(2.5)	
	고 졸	49(24.1)		미혼자녀와 동거	53(26.0)	
	대 졸	94(46.3)		손자녀와 동거	2(1.0)	
	대학원 졸	48(23.6)		기 타	3(1.5)	
결혼상태	초 혼	182(88.3)		연 령	60~69세	187(90.8)
	사별 후 독신	18(8.7)			70~79세	15(7.3)
	사별 후 재혼	3(1.5)			80세 이상	4(1.9)
	이혼 후 독신	1(.5)		직업지위 ¹⁾	65~74점	22(24.7)
	이혼 후 재혼	2(1.0)	75~84점		62(69.7)	
		85 이상	5(5.6)			
종 교	없 다	40(20.0)	주관적 계층의식	상	12(5.8)	
	개신교(기독교)	102(51.0)		상 중	23(11.2)	
	천주교	28(14.0)		하	21(10.2)	
	불 교	30(15.0)		중 중	상	73(35.4)
자녀수	0명	2(1.0)			중	55(26.7)
	1명	21(10.5)			하	18(8.7)
	2명	101(50.5)	하 층 ²⁾	상	2(1.0)	
	3명	53(26.5)		중	1(.5)	
	4명 이상	21(11.4)		하	1(.5)	

<표 2>는 기초 통계 자료로서, 각 항목들에 대한 응답의 평균값을 조사한 것이다. 평균값을 살펴보면, 한국의 신노년층은 가족지지, 수행능력, 욕구 충족, 생활만족도에서 평균적으로 높은 만족도를 보이며, 생산적 활동의 경우에는 만족도가 다소 떨어짐을 알 수 있다. 이는 자원봉사와 같은 단체 활동에 대해 소극적인 한국 노인들의 특성이 신노년층에서도 역시 반영되고 있음을 보여준다.

- 1) 대표적인 직업의 직업지위(사회경제적 지위점수)를 보면 경찰관 67.60, 컴퓨터 프로그래머 72.99, 초등학교 교사 65점, 중학교 이상 교사 73.99점, 초등학교 교장 76.21점, 의사 78.35점, 대기업 사장 78.60점, 변호사 80.65점 등이다.
- 2) 계층 귀속의식에서 하층이라고 인식하였지만, ‘직업지위’가 65점 이상이므로 본 연구 대상자에 포함시키기로 하였다.

<표 2> 각 항목별 응답자의 평균

	남	여	전 체
가족 지지	3.94	4.03	3.98
수행 능력	3.26	2.95	3.11
육구 충족	3.76	3.75	3.76
생산적 활동	2.91	3.51	3.19
생활 만족도	3.49	3.35	3.43

4) 연관규칙 분석 결과

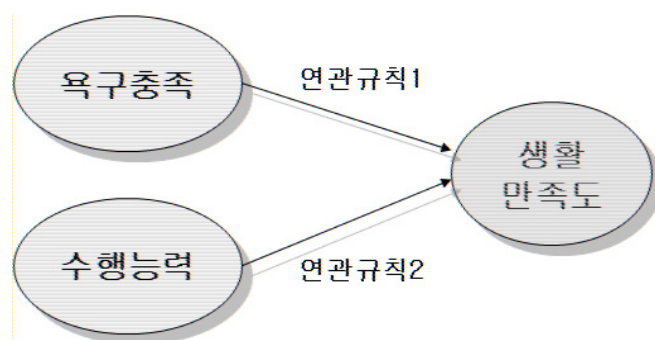
본 연구에서는 가족 지지, 육구충족, 생산적 활동, 수행 능력과 생활 만족도 간의 연관관계 추출하는 것을 목표로 하였다. 연관규칙 분석은 WEKA 3.5.7을 이용하였고, 최소 지지도는 10%, 최소 신뢰도는 60%로 하였다.

실험 결과 도출된 변인들 간의 1차원 연관규칙³⁾은 <표 3>, <표 4>와 같다. 연관규칙 조건부와 결론부의 괄호 안의 수치는 조사 대상자 중 몇 명이 해당 조건을 지지하는 가를 의미한다. 예를 들어 연관규칙 1을 살펴보면, “육구 충족이 높다”라고 응답한 127명 중, “생활만족도가 높다”라고 응답한 조사 대상자가 90명이라는 의미이다. 따라서 신뢰도는(90/127) * 100% = 71%가 된다.

<표 3> 연관규칙 분석을 통해 추출된 1차원 연관규칙들(1)

규칙번호	규칙 내용	신뢰도
1	육구충족이 높으면(127), 생활만족도가 높다(90)	71%
2	수행능력이 높으면(57), 생활만족도가 높다(50)	88%

연관규칙 분석을 통해 추출된 <표 3>의 1차원 규칙들을 살펴보면, 육구충족과 수행능력만이 신노년층의 생활만족도에 직접적인 영향을 미치는 변인으로 나타났다(<그림 1>)



<그림 1> 신노년층의 생활만족도에 대한 이론 모형 1

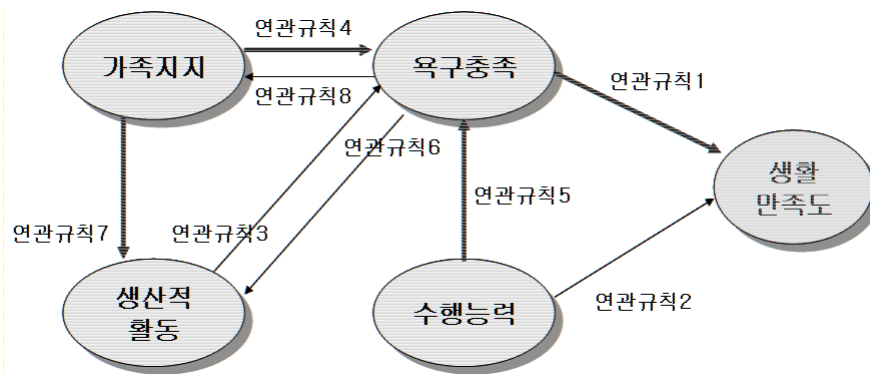
3) 연관규칙에서의 차원이란 조건부에 있는 변수의 개수를 의미한다. 즉, 1차원 연관규칙은 "A ⇒ B"이며, 2차원 연관규칙은 "A and B ⇒ C"의 형태를 갖는다.

<표 4>의 1차원 규칙들은 신노년층의 생활만족도에 직접적인 영향을 미치지 못하지만, 수행능력과 욕구 충족을 매개변인으로 한 가족지지, 생산적활동의 연관성에 관한 규칙들로서, 4개의 변인들 간의 다양한 연관성을 보여준다.

<표 4> 연관규칙 분석을 통해 추출된 1차원 연관규칙들(2)

규칙번호	규칙 내용	신뢰도
3	생산적 활동이 보통이면(133), 욕구충족이 높다(94)	71%
4	가족 지지가 높으면(165), 욕구충족이 높다(119)	72%
5	수행능력이 보통 이상이면(145), 욕구충족이 높다(103)	71%
6	욕구충족이 높으면(127), 생산적 활동이 보통이다(94)	74%
7	가족지지가 높으면(165), 생산적 활동이 보통이다(74)	72%
8	욕구충족이 높으면(127), 가족지지가 높다(119)	94%

<그림 2>는 연관규칙 분석 결과 도출된 변수들 간의 1차원 연관규칙들을 종합하여 만든 이론 모형이다.



<그림 2> 신노년층의 생활만족도에 대한 이론 모형 2

위에서 제시된 이론 모형 2에서, 욕구충족이 생활만족도에 정적(+)인 영향을 미친다는 연관규칙 1번, 가족지지와 수행능력이 욕구 충족에 정적(+)인 영향을 미친다는 연관규칙 4번과 5번, 그리고 가족지지가 생산적활동에 정적(+)인 영향을 미친다는 연관규칙 7번은 기존의 연구 결과(이수애, 2005; 장석영, 2002)와 일치한다. 따라서 데이터 마이닝의 연관규칙을 이용하면, 선행 연구들에서 찾아낸 변수들의 연관관계(연관규칙 1, 4, 5, 7) 뿐만 아니라, 변수들 간의 새로운 연관관계(연관규칙 2, 3, 6, 8)도 찾아낼 수 있다. 이는 선행 연구나 전문가의 지식에 의해 이론 모형을 설정한 후, 통계적인 분석 도구를 이용하여 가설을 검증하는 전통적인 방식에서 벗어나, 사전지식 없이 데이터 자체에 내재되어 있는 변인들 간의 상관성을 찾아낸 후, 이로부터 이론 모형을 도출하는 새로운 형태의 방법론이라 할 수 있다.

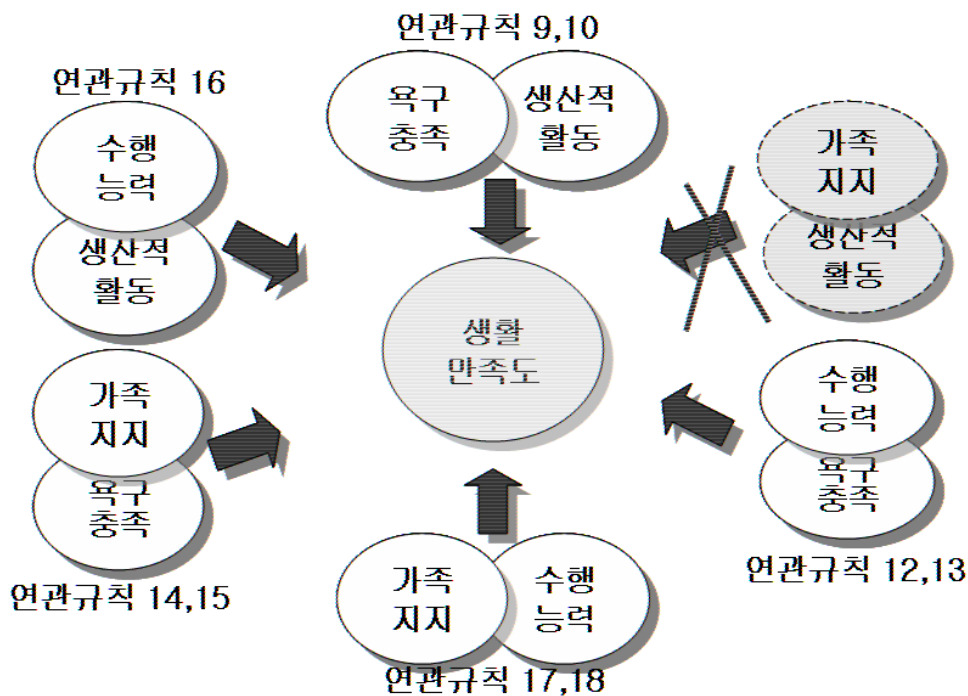
데이터 마이닝의 연관규칙 분석을 이용할 경우, 두 변수간의 연관관계 뿐 아니라, 여러 항목간의 다양한 연관관계도 찾아낼 수 있다. 다음의 <표 5>는 연관규칙 분석을 통해 도출된 2 차원 연관규칙의 예이다.

<표 5> 연관규칙 분석을 통해 추출된 2 차원 규칙들

규칙번호	규칙 내용	신뢰도
9	욕구충족이 높고, 생산적활동이 보통이면(94), 생활만족도가 높다(70)	74%
10	욕구충족이 보통이고, 생산적활동이 보통이면(39), 생활만족도가 보통이다(25)	64%
11	수행능력이 높고, 욕구충족이 높으면(50), 생활만족도가 높다(49)	98%
12	수행능력이 보통이고, 욕구충족이 높으면(53), 생활만족도가 높다(33)	62%
13	수행능력이 보통이고, 욕구충족이 보통이면(35), 생활만족도가 보통이다(24)	69%
14	가족지지가 높고, 욕구충족이 높으면(119), 생활만족도가 높다(86)	72%
15	가족지지가 높고, 욕구충족이 보통이면(46), 생활만족도가 보통이다(28)	61%
16	수행능력이 높고, 생산적 활동이 보통이면(45), 생활만족도가 높다(41)	91%
17	가족지지가 높고, 수행능력이 낮으면(37), 생활만족도가 높다(27)	68%
18	가족지지가 높고, 수행능력이 높으면(54), 생활만족도가 높다(49)	91%

<표 5>의 연관규칙 17번을 살펴보면, 생활만족도에 직접적인 영향을 미치는 수행능력(그림 1 참조)이 낮더라도, 가족지지가 높으면 생활만족도가 높음을 알 수 있다. 따라서 가족지지와 생산적 활동이 신노년층의 생활만족도에 직접적인 영향을 미치지 못하는 못하지만, 다른 변인들과 함께 생활만족도에 간접적인 영향을 미침을 알 수 있다. 연관규칙 12번과 13번을 보면, 같은 수행능력을 갖고 있더라도 욕구충족의 정도에 따라 생활만족도가 달라진다는 사실을 알 수 있다.

도출된 2차원 연관규칙들에서 주목할 만한 사실은, 신노년층의 생활만족도에 직접적인 영향을 미치지 않는 것으로 나타난 가족지지와 생산적활동을 조건부로 갖는 연관규칙이 도출되지 않았다는 점이다. 이는 간접변인인 가족지지와 수행능력만으로는 신노년층의 생활만족도를 결정할 수 없다는 것을 의미한다. 신노년층의 생활만족도에 영향을 미치는 두 변인들의 조합을 살펴보면, <그림 3>과 같다.



<그림 3> 두 변인들의 조합과 생활만족도와의 연관성

연관규칙 분석의 특성상 3, 4차원의 고차원 연관규칙들도 추출이 가능하다. 즉, 연관규칙 분석을 통해 신노년층의 생활만족도를 좌우하는 여러 항목간의 내재되어 있는 다양한 연관성을 찾아낼 수 있다. 다음의 <표 6>은 연관규칙 분석을 통해 추출된 3, 4차원 규칙들이다. 다양한 형태의 3차원 규칙들이 추출되었으나, 지면 관계상 생략하였다. 또한, 100%의 신뢰도를 보이는 연관규칙 14는 우리가 직관적으로 쉽게 이해할 수 있는 규칙임을 알 수 있다.

<표 6> 연관규칙 분석을 통해 추출된 3, 4차원 연관규칙들

차원	규칙번호	규칙 내용	신뢰도
3차원	19	가족 지지가 높고, 수행능력이 높고, 욕구충족이 높으면(49), 생활만족도가 높다(48).	98%
	20	가족지지가 높고, 수행능력이 높고, 생산적 활동이 보통이면(43), 생활만족도가 높다(40).	93%
4차원	21	가족지지가 높고, 수행능력이 높고, 욕구충족이 높고, 생산적 활동이 많으면(40), 생활만족도가 높다(40)	100%

IV. 결론

통계적인 방법을 이용한 기존의 설문 분석 방법은 사전 지식이나 선행연구를 통해 이론 모형을 정립하고, 통계적인 틀들을 이용하여 이론 모형에서 설정한 가설을 검증하는 방식으로 이루어져 왔다. 따라서 연구자가 유의미한 가설을 설정하지 못하거나, 이론 모형의 정립이 잘못되어 있을 경우, 올바른 분석 결과를 얻을 수 없다는 문제점을 안고 있다. 또한 다양한 항목간의 연관성을 파악하는데도 어려움이 있다. 본 논문에서는 데이터 마이닝의 연관규칙 분석을 이용하여 가족지지, 수행능력, 욕구충족, 생산적 활동과 신노년층의 생활만족도와와의 상관성을 도출하고, 그에 따른 이론 모형을 제시하였다. 그 결과, 선행 연구들에 의해 검증된 연관규칙뿐 아니라 내재되어 있던 새로운 연관규칙들을 찾아낼 수 있었다. 이는 통계적인 기법을 이용하는 전통적인 설문 분석과는 차별화되는 새로운 형태의 방법론으로서, 그 의의가 크다 할 수 있다.

또한 다차원 연관관계 분석을 통해 한국 신노년층의 생활만족도에 영향을 미치는 여러 항목간의 다양한 상관관계를 도출할 수 있었다.

본 논문은 신노년층의 생활만족도에 관한 연관관계만을 찾았으나, 향후 삶의 질을 측정하는 중요한 변수인 자기효능감, 우울감과 다른 항목들 간의 연관관계에 대한 연구도 수행할 예정이다. 또한 분류분석이나 회귀분석 등 다양한 데이터 마이닝 기법을 통한 한국 신노년층의 삶의 질에 관한 지속적인 연구를 수행할 예정이다.

[참고문헌]

- 한경혜 · 윤성은(2007), 대중매체에서의 신노년 담론 분석: 신문매체를 중심으로, 한국노년학회, Vol. 27, No. 2, pp. 299-322, 2007.
- 박영순(2006), 노인소비자의 주요 소비자 문제와 상담사례, 한국가정관리학회·한국가족자원경영학회 2006년 춘계학술대회 자료집, pp. 87-105, 2006.
- 김완섭 · 이수원(2006), 데이터마이닝을 이용한 설문조사의 심층 분석, 공학교육연구, Vol. 9, No. 4호, pp. 71-82, 2006.
- 이정수 · 김교정(2003), 설문 데이터를 위한 다차원 연관 규칙 마이닝, 한국지능정보시스템학회 학술대회논문집, pp. 395-399, 2003.
- 김영숙 · 정국인 · 박소라(2008), 재가노인 사례관리의 욕구사정 정확도 향상을 위한 욕구추출 알고리즘 개발: 데이터 마이닝 분석기법을 활용하여, 한국사회복지학, Vol. 60, No. 1, pp. 187-209, 2008.
- 김혜숙, 문양세, 노희영, 김진호(2006), 연관규칙을 이용한 방학 중의 생활습관 분석, 한국정보과학회 가을 학술발표논문집, Vol. 33, No. 2(A), 2006.
- R. Nayak, L. Buys and J. Lovie-Kitchin(2006), Data Mining In Conceptualising Active Ageing, Proc. of 5th Australasian conference on Data mining and analytics, Vol. 61, pp. 39-45, 2006.
- E. Pitt and R. Nayak(2007), The Use of Various Data Mining and Feature Selection Methods in the Analysis of a Population Survey Dataset, Proc. of the 2nd International Workshop on Integrating AI and data mining, Vol. 84, pp. 83-93, 2007.
- 성혜영(2005), 성공적 노화 모델 연구 : Rowe와 kahn의 구성 요소를 활용하여, 성균관대학교 대학원, 2005.
- 이수애(2005), 주관적 삶의 질 설명모형에 관한 연구, 동신대학교 대학원, 2004.
- 한형수(2002), 한국사회 도시노인의 삶의 질에 관한 연구, 고려대학교 대학원, 2002.
- 장석영(2002), 중산층 노인의 생활실태에 관한 연구: 일산 아파트 지역 노인들을 중심으로, 연세대학교 교육대학원, 2002.

연관규칙 마이닝을 이용한 한국 신노년층의 생활 만족도에 관한 연구

박 정 윤(중앙대학교 가족복지학과 부교수)

최근 한국사회는 저출산·고령화로 인해 인구구조가 빠르게 변화하고 있으며 이에 대처하기 위하여 학문적으로, 실천적으로 여러 영역에서 학제적인 연구와 대안들 마련되고 있다. 특히 고령사회로의 전환으로 인해서는 노인인구 비율이 증가하게 되고 그러면서 노인관련 이슈들 - 노화, 노인생활 전반(가족, 경제, 보건, 부양, 문제-에 관심이 급증하고 친고령사회조성을 위한 연구들이 호라발하게 실시되고 있다.

이즈음에 본 연구의 주제는 노인대상이라는 점에서 시의적절한 연구라고 보여진다. 본 연구에서 도입한 신노년층이라는 용어는 지금까지의 노인연구와는 다른 관점이라고 할 수 있다. 뿐만 아니라 연구방법도 “연관규칙 마이닝”라는 새로운 기법을 적용하여 노인기의 가족생활에 대한 새로운 시각에서의 관련변인들의 관계를 밝히는 점에서 매우 의의가 크다고 하겠다. 앞으로 이렇게 새로운 논의와 연구들이 지속되어야 할 필요가 있다고 보여진다.

또한, 본 논문은 노인의 삶의 질이나 생활만족도를 향상시키기 위한 변인을 밝히는데 설문조사를 통한 연구모형이나 가설의 귀무여부 검증하는 단선적 분석이 아니라 관련변인들의 역학관계를 연관규칙 마이닝이라는 연구방법을 통해서 노인 생활만족도에 영향을 주는 변인간의 상호작용을 구체적으로 분석하였다. 따라서 본 연구에서 밝혀진 변인의 통제나 조절 정도에 따라서 노인의 행복한 노년기 생활을 유지할 수 있을지를 예측할 수 있다는 점에서 앞으로 많은 시사점을 갖는다고 할 수 있다.

한편, 본 연구결과에서 신노년층의 생활만족도에 영향을 주는 변인- 가족지지, 수행능력, 욕구충족, 생산적 활동- 4가지의 관련성에서 흥미로운 것은 수행능력과 욕구충족변인이 직접적인 영향을 주고 가족지지와 생산적 활동 변인은 간접적 영향변인이라는 점이다. 즉, 신노년층은 개인적 욕구충족이나 생활능력에 대해서 긍정적으로 지각하는 것이 중요하다는 점이 밝혀졌으므로 노인의 생활만족도 증진을 위해서는 개인적 만족감 증진을 위한 노력을 우선시해야 한다. 또, 자신의 수행능력지각이 비슷한 경우도 욕구충족에 따라 생활만족도가 다르며, 간접변인 2가지 역시 유사한 상관관계를 보임으로 노인 개개인의 생활만족도가 다양한 변인의 인지나 만족에 영향을 받음으로 특정 변수 하나(예를 들어 가족지지, 수행능력향상 등)의 획일적인 변화보다는 여러 상황이나 측면을 고려하는 것이 필요함을 보여준다는 점도 매우 의미있는 결과라고 할 수 있다.

이러한 연구의 의의점에도 불구하고 본 연구가 보다 나은 연구가 되기 위해서 몇가지 보완점을 토론자가 제시하고자 한다.

첫째, 연구방법에서 용어정의 및 측정도구의 보완이 필요하다. 신노년층 개념 정의의 근거가 제시되지 않아 연구대상에 대한 부연설명이 필요하다. 특히 소득계층, 직업군, 거주지역으로만 신노년층으로 명명하는 것은 설득력이 부족함으로 이에 대한 객관적 근거를 제시해야 한다. 또, 독립변인인 가족지지, 수행능력, 욕구충족, 생산적 활동을 선정하게 된 근거를 제시하고, 각 독립변인의 문항수, 문항내용, 신뢰도, 선행연구에서 인용한 경우라면 인용자료 등을 명시해야 한다.

셋째, 연구결과를 다양한 측면으로 해석할 필요가 있다. 독립변인의 연관성 정도에 따라서 생활만족도가 다른 점에 대한 구체적인 분석을 시도함으로써 새로운 연구방법으로 분석한 의미를 한층 더 부각시킬 수 있다. 아울러 기술통계 분석도 상세하게 제시하여 연관관계 결과를 예측하는데 자료가 될 것으로 사료된다.

끝으로 본 연구와 같은 새로운 연구방법을 도입하여 기존의 노인기의 삶의 질 향상 또는 생활만족도, 결혼만족도 등에 관련된 변인간의 연관성을 밝히는 연구를 지속적으로 해야 한다. 또, 노인분야 뿐만 아니라 가족생활 전반의 가족관계 내의 현상을 규명하는데 관련된 변인간의 연관성을 밝힘으로써 이를 관계강화 또는 향상시킬 방안을 제시하는 것을 제언한다.