

Application of Data Mining on Simultaneous Activities on the Time Use Survey

Ki Seong Nam¹⁾, Hee Jea Kim²⁾

Abstract

This Paper analyzed simultaneous activities of the time use survey by Korea National Statistical Office to use data mining's association rule. The survey of National Statistical Office in 1999 considered general analysis for main activities like that personal care(eating), employment and study, leisure, travel by purpose. But if we use the association rule, we can found the ratio of simultaneous activities at the same time. And also we can found the probability that another activities practise if we act one particular activity. Using this association rule of data mining we can do more developed and analytical sociological study.

Keywords : 데이터마이닝, 연관성규칙, 생활시간, 동시 행동

1. 서론

인간들의 생활을 이해하는데 기본이 되는 중요한 요소의 하나는 바로 시간이다. 시간은 단지 정태적인 현상만을 나타내는 것이 아니라 과거와 미래의 삶이 포함된 것이다. 그러나 현대사회에서의 시간은 노동과 여가를 극단적으로 분리시킬 뿐만 아니라 그 속에서의 생활을 아주 세분화시키고 있다. 현대인들이 생활시간을 어떻게 배분하고 있는가를 파악하는 것은 바로 그들의 생활형태를 분석하는 것이다. 하루 24시간은 노동생활시간과 소비생활시간으로 대별되고, 이외에도 다양한 활동이 행해지고 있다. 이러한 활동들은 동일한 시간에 두 가지 이상의 행동을 하고 있는 경우도 많이 있다. 본 논문에서 주된 활동을 토대로 분석한 생활시간의 배분을 바탕으로 특정 시간대에 동시적으로 행해진 행동(동시행동simultaneous activities)을 데이터마이닝의 분석을 통하여 분석하고자 한다.

1. 경북 경주시 석장동 707번지, 동국대학교 정보통계학과 강사
E-mail : ksnam@dongguk.ac.kr

2. 부산광역시 금정구 장전동 산30번지 부산대학교, 사회조사연구소 연구원
E-mail : khjjh1@hanafos.com

일반적으로 사람들은 동일한 시간에 두 가지의 행동을 하는 경우가 많이 있다. 예를 들어 라디오를 청취하면서 독서를 하거나, 식사를 하면서 TV를 시청을 하는 예이다. 만약 식사를 할 때는 반드시 TV를 시청하는 사람이 많다면, 상식적인 이야기지만 식사시간에 혐오스러운 방송을 한다면 시청률이 떨어질 것이다. 데이터마이닝에서의 연관성 분석을 이용한다면 이와 같은 결과를 얻을 수 있다.

데이터마이닝은 축적된 정보기술의 발달과 비즈니스적 요구에 따라 시장에 등장하게 되었다. 정보화 시대의 도래와 더불어 정보기술은 가속적 발전을 가져왔으며, 전통적인 전문가시스템에는 한계성을 보이고 있다. 현재는 데이터의 홍수 속에 정보의 빈곤을 느끼기 때문에 데이터마이닝의 필요성이 더욱 피부에 와 닿는 것이다. 데이터마이닝은 많은 정보들 중에서 소중한 정보를 캐어내는 좋은 분석 방법이며, 자료의 방대한 양이 있을 때에만 의미를 가진다. 데이터마이닝은 기업 등에서만 사용하는 것이 아니고, 공학적 연구 및 인간 생활에 활용할 시기가 왔다고 한다. 본 논문은 이에 주안점을 두고 생활시간 분석에서 데이터마이닝의 방법을 시도하는 것이다

동일한 자료를 이용하여 이전과 다른 분석 방법을 사용하여 새로운 결과를 얻는 것은 매우 중요한 일이다. 물론 주어진 자료에서 분석 방법의 미비로 인하여 소중한 정보를 찾아내지 못하는 것은 자료의 손실이라 할 수 있다. 현재 통계청에서 조사된 국민생활시간조사는 소중한 자료는 준비되어 있는데 반하여 조사에 대한 소개가 부족하고, 분석에 대한 다양한 연구가 부족하다. 따라서 본 논문에서는 데이터마이닝 기법 중의 하나인 연관성규칙을 이용하여 자료의 가치를 높이고자 한다.

국민생활시간조사에 대한 분석에 앞서 시간활용연구에 관한 논문을 살펴보면 1913년 미국에서 출판된 “일하는 남자들은 시간을 어떻게 보내는가?”(1913)가 발표된 후로 미국(1920, Univ. of Cornell), 러시아(1924), 영국(1938, BBC) 등과 같이 각국의 정부 및 연구소 차원에서 다양한 조사가 이루어졌다. 일본의 경우 1969년부터는 NHK에 의해, 1976년부터는 통계국에 의해 대규모 조사가 정기적으로 실시되고 있다. 캐나다는 1986년 전국규모의 조사를 처음으로 실시한 이후 1998년까지 3차례 실시하였고, 네덜란드는 1975년 이후 1998년까지 5차례, 노르웨이는 1970년 이후 1990년까지 3차례 실시하는 등 현재 30여 개국 이상에서 생활시간조사가 실시되고 있다.

그 동안 우리나라에서도 시간활용에 대한 연구와 조사가 전문가와 연구기관의 관심에 의해서 이루어져 왔는데, 1981년 KBS에 의해 시간활용을 어떻게 하는지에 대한 조사가 「국민생활시간조사」라는 명칭으로 처음으로 실시된 이래, 1995년까지 6차례 실시되었다. 또한 농촌진흥청에서 농가주부의 생활을 분석하기 위한 시간조사를 1979년 이후 실시하고 있고, 그 이외 연구자들에 의해 소규모 연구조사가 지속적으로 실시되어 왔다. 전자의 경우는 조사목적이 각 개인의 인구학적 특성별 시간대별, 요일별 대중매체 이용실태 및 여가생활의 파악에 주어져 있었던 반면, 후자의 연구들은 주로 주부의 생활시간구조를 파악하기 위한 것과 기본자료를 통한 주부의 가사노동가치를 측정하는 목적 등 가정학적 접근이 주류를 이루었다.

기존 시간연구들이 특정 연구 목적을 위한 특화된 소규모 조사이었으며, 전국적인 규모의 시간조사의 필요성이 대두되고 특히 정책적 활용도의 인식이 증가함에 따라, 통계청에서는 1997년 처음으로 시간에 관한 조사를 개발하기로 결정하고, 1999년 처음으로 전국 10세 이상 약42,972명을 대상으로 조사를 실시하였다.

이 조사를 이용한 학문적 연구는 제7회 『통계의 날』 기념 세미나에서 손애리(2000)를 비롯한 여섯 편의 논문이 발표된 적이 있으며, 개인적 연구로 김희재, 남기성

(2002)은 단행본의 책으로 한국노동자의 생활시간을 발표한 적이 있다.

본 논문은 1장에서 데이터마이닝 및 생활시간조사를 소개하고, 2장에서는 실제 데이터로 사용한 한국의 생활시간조사에 대하여 살펴보았다. 3장에서는 데이터마이닝 기법의 하나인 연관성 규칙에 대해 고찰하고, 4장에서는 식사, 일 또는 학업, 교제 및 여가활동, 이동에 따른 동시행동의 분석 결과를 살펴보고, 마지막으로 5장에서는 결론을 다루기로 한다.

2. 생활시간조사

생활시간조사는 통계법에 의한 지정통계(승인번호 : 제10152호)로 1999년 처음으로 실시된 조사이며 앞으로 5년 주기로 통계청에서 조사를 실시할 예정이다. 우리나라 국민들이 주어진 하루 24시간을 어떤 형태로 보내고 있는지를 조사하여 국민의 평균적인 생활 방식과 삶의 질을 파악하고, 주어진 시간자원을 효율적으로 활용하는데 필요한 기초자료 제공한다. 또 무보수 가사노동에 소요된 시간과 그 외 다양한 생활시간을 파악하여 국민계정에 가계부문 생산을 반영하기 위한 연구와 각종 노동, 복지, 문화, 교육, 교통 관련 정책 수립 및 학문적 연구활동에 기초자료 제공하고 있다.

조사대상으로는 조사기간을 기준으로 전국에 거주하는 만 10세이상 가구원을 대상으로 하며, 조사표 배부일부터 회수일까지 5일간 가구의 조사대상자 전원이 부재중인 경우나 조사표 작성이 불가능하다고 인정되는 경우는 조사대상에서 제외된다. 또한 제외 가구원으로는 친척방문, 출장, 병원 입원 등의 사유로 조사표 배부일부터 회수일까지 5일간 부재중인 가구원, 심신장애, 치매 등의 사유로 본인 스스로 시간일지 작성이 불가능한 가구원이다.

조사항목으로는 가구관련 사항(주택의 종류, 입주형태, 주거전용면적, 자가용 보유, 미취학 아동 보육 여부), 개인관련 사항(가구주와의 관계, 성별, 연령, 교육정도, 혼인상태 등의 개인특성 및 경제활동상태, 휴무제도, 시간부족느낌 등), 시간일지(주행동, 동시행동, 장소(집 안, 밖), 이동수단) 등이다.

조사방법으로는 가구관련사항, 개인관련사항은 조사담당자가 대상가구를 방문하여 가구원과 면담을 통하여 조사하는 면접 타기식으로 조사하며, 시간사용은 시간일지(time-diary)를 조사대상가구에 배포하여 대상가구원이 2일(48시간) 동안 행동과 시간을 직접 기입하는 자기기입식으로 조사된다.

시간일지의 행동분류체계는 응답자가 행동한 목적에 따라 행동을 9개 대분류, 51개 중분류, 125개 소분류로 분류되어 있다.

추정항목으로는 평균시간, 행위자비율, 행위자평균시간, 10분 간격별 행위자비율이 있으며, 주행동은 10분 간격의 시간대에 한가지 행동을 한 경우 그 행동이며, 같은 시간대에 두가지 이상의 행동을 한 경우 가장 길게 한 행동이나 응답자가 주관적으로 주된 행동이라고 판단한 행동이고, 동시행동은 같은 시간대에 두 가지 이상의 행동을 한 경우 주행동 이외의 행동으로 두개의 행동이 병행되는게 논리적으로 가능한 경우에만 동시행동으로 인정하고 있다.

3. 연관성규칙

데이터마이닝(Data Mining)의 사전적인 의미는 채굴(Mine)한다는 뜻으로 ‘자료를 거대한 더미 속에서 가치 있는 무언가를 캐낸다’라는 것이다. 즉, 방대한 양의 데이터로부터 쉽게 드러나지 않는 유용한 정보들을 추출하는 과정을 뜻한다. 여기서 정보는 묵시적이고 잘 알려져 있지 않지만 잠재적으로 활용가치가 있는 정보를 말한다. 예를 들면 “창문이 많은 집에 도둑이 많다”고 한다. 어쩌면 너무나도 당연한 말처럼 들릴 수도 있지만, 누구나 이러한 사실이 보험 요율에 적용될 수 있으리라고 생각하지는 않았을 것이다. 실제 영국의 한 보험회사는 이러한 사실을 이용하여 차별적인 보험요율을 적용함으로써, 보다 효과적인 정책을 수립하게 되었고, 이로 인해 기업의 수익증대에 커다란 공헌을 하게 되었다고 한다.

창문과 도둑의 관계를 찾아내는 것과 같이 데이터마이닝이란 기업 등에서 보유하고 있는 거래자료, 고객자료, 상품자료, 마케팅 활동의 피드백 자료와 기타 외부자료를 포함하여 사용 가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 실제 경영의 의사결정 등을 위한 정보로 활용하고자 하는 것이다.

연관성규칙(association rule)은 데이터마이닝 분석 기법의 하나로 데이터 안에 존재하는 항목간의 종속관계를 찾아내는 작업을 의미하는 것으로 마케팅에서는 손님의 장바구니에 들어있는 품목간의 관계 즉, 동시에 구매될 가능성이 큰 상품들을 찾아낸다는 의미에서 장바구니분석(market basket analysis)이라고 한다.

본 연구에서의 연관성규칙에서 얻어지는 결과는 If A, then B ($A \rightarrow B$) 와 같은 형식으로 표현된다. 이는 “행위 A를 하는 경우에는 행위 B도 실행한다.” 라고 해석된다. 연관규칙은 구체적인 행위가 언급되므로, 이해가 쉽고 명쾌하며 다른 분야로의 적용이 용이하다. 즉, 유용한 규칙을 찾게 된다면 상당히 실질적이고 효과적인 정보를 얻을 수 있게 되는 것이다.

유용한 연관규칙을 이끌어내기 위해서는 먼저, 어떠한 행위들이 어느 정도의 연관성을 가지고 있는지를 측정해야 할 것이다. 측정의 기본은 특정시간에 주 행동과 더불어 어떤 행위를 하였는가 하는 빈도(Count)이다. 빈도를 기반으로 연관정도를 정량화하기 위해서 지지도(support), 신뢰도(confidence), 향상도(lift)의 3가지 기준을 고려한다.

지지도는 다음 식과 같이 표현된다.

$$\text{Support} = P(A \cap B)$$

관심 있는 규칙($A \rightarrow B$)에 대하여, ‘전체 행위 중 A와 B를 동시에 행하는 비율은 어느 정도인가?’를 의미하며, 지지도가 높다는 의미는 전체 행위에서 A와 B를 동시에 행동하는 비율이 높다는 뜻으로 두 가지의 행위는 행동과학적 측면에 중요한 의미를 가진다는 뜻이다.

신뢰도는 다음 식과 같이 표현된다.

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

신뢰도는 ‘A를 행하는 사람들 중에서 B를 같이 행하는 비율은 어느 정도인가?’를 의미하며, 신뢰도가 높다는 것은 B 행위는 A 행위에 종속되는 관계에 있기에 A 행위를 하는 사람들은 B 행위를 하는 비율이 높다는 의미이다. 기본적으로 A와 B의 행위의 비율이 어느 정도 수준이 되어야 의미가 있을 것이나, 지지도가 낮은 것일지라도 신뢰도가 높다는 것은 두 행위의 관련성이 높다는 의미이기에 행동과학적 측면에서는 신뢰도가 가장 중요하다고 할 수 있다.

향상도는 다음과 같이 표현된다.

$$\text{Lift} = P(B | A) / P(B)$$

향상도는 ‘임의로 B를 행하는 경우에 비해 A와의 관계가 고려되어 행하는 경우의 비?’를 의미하며, 신뢰도와 전체 사람들 중에서 B 행위를 한 사람들과의 비율이기에 신뢰도를 보조하는 역할을 하는 것이다. 연관성규칙에서는 신뢰도의 값이 크면 좋지만 향상도가 낮다면 의미는 반감되는 것이다. 즉, $A \rightarrow B$ 가 높은 신뢰도를 주지만, 원래 B행위를 행하는 기본 확률보다 커야 A를 고려해서 B를 생각하는 것이 의미가 있을 것이다. 즉, 향상도는 1보다 작으면 음의 연관관계, 1이면 독립에 가까운 사건, 1보다 크면 양의 연관관계를 가진다고 볼 수 있다.

4. 분석결과

본 분석에서는 통계청에서 1999년 조사된 국민생활시간조사에서 첫째 날에 해당하는 42,973명의 자료에서 동시행동 부분을 사용하였으며, 사용 소프트웨어는 SAS Ver 8.1 Enterpriser Miner를 사용하였다. 분석 대상 변수로는 개인유지(식사), 일 또는 학업, 교제 및 여가활동, 이동의 4가지를 분석하였다. 즉, 24시간 중에서 주행동으로 위의 4가지 종류의 행동을 가장 많이 행하는 몇 가지의 시간대를 선택하여 동일시간에 행하는 동시행동을 연관성 규칙을 이용하여 분석하였다. 행동에 대한 분류는 소분류로서 모두 125개의 행동을 사용하였다.

4.1 식사

사람들은 일반적으로 하루에 3번의 식사를 하고 있다. 3번의 식사는 사람들의 생활스타일에 따라 식사시간도 차이가 난다. 1999년 생활시간조사 결과에서의 매 시간대의 행위자 비율에 따르면 아침 식사는 오전 8시(12.9%), 점심은 오전 12시(22.9%), 오후 1시(23.8%), 그리고 저녁식사는 오후 7시(20.0%), 오후 8시(14.2%)에 하는 것으로 나타났다. <표 1>은 매 시간대에서 식사의 행위자 비율이 높은 시간대에 행하는 동시행동에 대한 연관성 규칙을 이용한 분석 내용이다.

〈표 1〉 식사시간에 행하는 동시행동 분석

시간	A 행동	A 행동의 행위자 비율	B 행동	향상도	지지도 (%)	신뢰도 (%)
08:00	가족과의 식사	10.2	가족 친척과의 교제	3.73	0.31	3.08
			TV	1.83	2.35	23.57
	그 외의 식사	2.7	TV	1.71	0.57	22.02
12:00	가족과의 식사	3.8	가족 친척과의 교제	4.72	0.23	5.09
			TV	1.16	0.32	7.25
	그 외의 식사	19.1	그 외 사람과의 교제	1.48	1.34	7.45
13:00	가족과의 식사	7.0	가족 친척과의 교제	3.61	0.46	5.88
			TV	1.54	0.75	9.67
	그 외의 식사	16.8	그 외 사람과의 교제	1.35	1.42	8.98
			CD, Tape 등 음악듣기	1.12	0.17	1.1
19:00	가족과의 식사	14.4	가족 친척과의 교제	2.52	0.96	6.27
			TV	1.23	4.03	26.33
	그 외의 식사	5.6	그 외 사람과의 교제	2.53	0.49	9.05
20:00	가족과의 식사	10.3	가족 친척과의 교제	1.8	0.65	6.29
			그외의 식사	3.9	그외 사람과의 교제	1.77

〈표 1〉에서와 같이 아침 식사에 해당하는 오전 8시와 저녁식사에 해당하는 오후 7시와 8시에는 가족과의 식사가 높은 비율을 차지하며, 점심식사(오전 12시와 오후 1시)는 가족 외의 사람과의 식사가 높은 비율을 차지하고 있다. 아침식사 시간에 행하는 동시행동으로는 ‘가족·친척과의 교제’와 ‘TV 시청’을 하는 것으로 나타났으며, ‘오전 8시에 식사를 하면 가족, 친척과의 교제를 한다’는 규칙이 성립하는 신뢰도는 3.08%이고, 지지도는 0.31%로서 두 가지 행동을 동시에 행하는 확률은 0.31%이며, 이때의 향상도는 3.73으로 양의 관계를 가지고 있으며, 임의로 ‘가족, 친척과의 교제’를 하는 것에 비해 3.73배나 된다. 그리고 ‘가족과의 식사를 하면 TV를 시청한다’는 규칙에서는 신뢰도는 23.57%이며, 지지도는 2.35% 교제에 비하여 높은 비율을 보이고 있으며, 이때의 향상도는 1.83으로 양의 관계를 보이고 있다. 즉, ‘오전 8시에 가족과 식사를 하면 TV를 시청한다’는 경우가 가족과 식사를 하는 전체 사람의 23.57%이다.

점심 식사에 해당하는 12시와 오후 1시에는 ‘가족 외의 식사’를 하는 비율이 높으며, 이에 해당하는 동시행동은 ‘가족 외 사람과의 교제’가 높은 비율을 차지하며, 이때의 지지도는 각각 1.34%와 1.42%이며, 신뢰도는 7.45%와 8.98%이며, 향상도는 1.48과 1.35로 나타났다.

저녁식사에서는 오후 7시에는 ‘TV시청’과의 신뢰도가 26.33%이나, 오후 8시에는 신뢰도가 나타나지 않고, ‘가족친척과의 교제’에서 신뢰도가 오후 7시보다 높은 6.29%이다.

식사에서 나온 몇 가지 규칙 중에서 신뢰도가 가장 높은 것은 ‘오후 7시에 가족, 친척과 식사를 하면 TV를 시청한다’로서 26.33%이다.

4.2 일과 학습

사람들이 수면 다음으로 가장 많은 시간을 할애하고 있는 것은 일과 학습이다. <표 2>는 일과 학습시간을 가장 많이 행하는 시간대에 동시에 행하는 동시행동에 대한 분석이다.

<표 2> 일과 학습시간에 행하는 동시행동 분석

시간	A 행동	A 행동의 행위자 비율	B 행동	향상도	지지도 (%)	신뢰도 (%)
11:00	주업	29.0	라디오	1.42	0.76	2.78
	부업	0.2	CD, Tape 등 음악듣기	1.12	0.00	1.05
	집에 가져와서 일함	0.1	세탁 및 세탁물 널기	1.15	0.00	1.79
			교제관련 전화 통화	1.64	0.00	1.79
			CD, Tape 등 음악듣기	1.89	0.00	1.79
	그외 일관련 행동	0.2	라디오	1.96	0.01	3.85
	무급가족종사일	1.3	라디오	2.1	0.06	4.12
기타 일관련 행동	0.0	그 외 사람과의 교제	3.54	0.00	16.67	
16:00	주업	25.8	라디오	1.47	0.63	2.63
	집에 가져와서 일함	25.8	라디오	2.84	0.01	5.08
	무급가족종사일	0.1	TV	1.07	0.10	7.55
			라디오	2.21	0.05	3.95
	구직활동	0.1	간식과 음료	1.2	0.00	5.71
담배피우기			8.96	0.00	5.71	
11:00	학교외에서의 스스로 학습	0.5	CD, Tape 등 음악듣기	6.63	0.03	6.25
	학교외에서의 기타 학습 관련	0.0	간식과 음료	1.18	0.00	3.45
			라디오	1.76	0.00	3.45
학습관련 물품 구입	0.0	그외 사람과의 교제	7.08	0.00	33.33	
16:00	학교에서의 스스로 학습	1.2	CD, Tape 등 음악듣기	1.59	0.01	1.32
	학교행사	0.1	라디오	1.27	0.00	2.27
	학교외에서의 스스로 학습	1.2	CD, Tape 등 음악듣기	5.11	0.06	4.24

<표 2>에서 특징적인 것은 ‘오전 11시에 기타 일 관련 행동을 하면 가족, 친척외의 사람과 교제를 한다’에서 신뢰도가 16.67%이며, 이 때의 향상도는 3.54로 나타났다. 그러나 동시에 행동을 하는 비율은 낮아 지지도는 0.00%로 나타났다. ‘오전 11시에 주업을 하면 라디오를 듣는다’는 지지도는 0.76%로 상대적으로 높으나, 신뢰도는 2.78%로 높지는 않았다. 이는 주업을 하는 동안에 동시에 행하는 행동의 종류가 많기 때문으로 추정되며, 오후 4시에도 비슷한 현상이 나타난다.

학습시간에는 학생이란 연령의 특성으로 인하여 ‘라디오, CD, Tape 등 음악듣기’ 등이 상대적으로 높은 신뢰도를 보이고 있다. 지지도는 낮지만 ‘오전 11시에 학습관련 물품을 구입하면 가족, 친척 외의 사람과 교제를 한다’에서 신뢰도가 33.33%로 나타났다. 이는 학생들이 물품을 구입할 때 혼자 가기보다는 여러 명이 같이 가기 때문이다.

4.3 교제 및 여가활동

교제 및 여가활동은 일반적으로 일과시간 이후에 이루어지기 때문에 오후 7시, 8시, 9시에 행동을 취하는 사람을 대상으로 동시행동을 분석하였으며, 행동에 대한 종류가 많아서 규칙도 많이 나왔으나 지면 관계로 지지도가 0.1% 미만에 대해서는 생략하였다. <표 3>은 교제 및 여가활동시간에 행하는 동시행동 분석의 결과이다.

<표 3>에서와 같이 ‘오후 7시에 가족, 친척과의 교제를 하면 가족과 식사를 한다’에서 신뢰도가 38.47%로 높은 비율을 보이며, 이때의 향상도는 2.52이다. 이는 <표 1>에서의 ‘오후 7시에 가족과의 식사를 하면 가족, 친척과의 교제를 한다’와 지지도와 향상도는 수식의 특성 상 같게 나타난다. 그러나 신뢰도는 무엇이 조건부로 되는가에 따라 다르게 나타난다.

‘오후 8시에 라디오를 들으면 주업을 한다’에서 신뢰도가 19.81%로 높은 것은 자영업자들이 항상 라디오를 틀어놓고 일을 하기 때문으로 추정된다. ‘오후 8시에 담배를 피우면 TV를 시청한다’에서 신뢰도가 34.86%로 높게 나타난 것은 특이한 현상으로 여겨진다. 그러나 이 때의 향상도가 1.06으로 낮기 때문에 담배를 피우는 소수의 사람들이 오후 8시에 일반인들이 가장 많이 행하는 행동(23.3%)인 TV를 보기의 행위가 집중되기 때문이다. 따라서 특이한 현상이라 규정할 수는 없다. 마찬가지로 ‘오후 9시에 가족, 친척 외와 교제를 하면 TV를 시청한다’의 신뢰도가 42.12%로 아주 높게 나타났다. 이때의 향상도는 1.02로 낮게 나온 것도 동일한 현상이다. 다른 행동을 행한 일반인이 TV를 시청하는 비율에 비하여 높지 않기 때문이다.

‘오후 9시에 유흥(술, 춤, 노래)을 하면 가족, 친척 외의 사람과 교제 중이다’에서는 신뢰도가 28.33%로 다른 시간에서도 비슷한 현상으로 지지도는 0.54로 별로 높지 않지만, 향상도가 8.45로 아주 높게 나타났다. ‘오후 9시에 그 외 사람과 교제를 하면 유흥(술, 춤, 노래)을 한다’에서도 신뢰도가 16.24%로 비교적 높게 나타난 것은 교제의 장소로 유흥주점이 일반화 된 것으로 예상할 수 있다.

<표 3> 교제 및 여가활동시간에 행하는 동시행동 분석

시간	A 행동	A 행동의 행위자 비율	B 행동	향상도	지지도 (%)	신뢰도 (%)
19:00	가족 친척과의 교제	1.0	가족과의 식사	2.52	0.96	38.47
			간식과 음료	2.01	0.10	4.2
	그 외 사람과의 교제	2.5	그 외의 식사	2.53	0.49	13.66
			간식과 음료	3.18	0.24	6.64
			유흥(술, 춤, 노래방)	7.41	0.28	7.94
	TV	12.8	가족과의 식사	1.23	4.03	18.84
			간식과 음료	1.38	0.61	2.87
	라디오	0.1	주업	2.16	0.24	22.58
			출·퇴근	4.92	0.27	25.38
	CD, Tape 등 음악듣기	0.0	주업	1.25	0.10	13.06
유흥(술, 춤, 노래방)	0.0	그 외 사람과의 교제	7.41	0.28	26.52	
20:00	가족 친척과의 교제	1.7	가족과의 식사	1.8	0.65	18.55
			간식과 음료	1.81	0.25	7.25
	그 외 사람과의 교제	2.8	그외의 식사	1.77	0.24	6.56
			간식과 음료	1.86	0.27	7.45
			유흥(술, 춤, 노래방)	7.53	0.44	11.97
	TV	23.3	간식과 음료	1.33	1.74	5.31
			아이에게 책 읽어주기, 놀아주기	1.1	0.37	1.14
	라디오	0.1	주업	2.54	0.19	19.81
			학교 외에서의 스스로 학습	3.87	0.11	11.22
			출·퇴근	5.11	0.15	15.04
CD, Tape 등 음악듣기	0.2	주업	1.59	0.10	12.39	
		학교 외에서의 스스로 학습	5.09	0.12	14.75	
유흥(술, 춤, 노래방)	1.7	그 외 사람과의 교제	7.53	0.44	27.53	
담배피우기	0.2	TV	1.06	0.14	34.86	
21:00	그 외 사람과의 교제	2.4	가족과의 식사	1.91	0.19	5.43
			간식과 음료	1.29	0.2	5.76
			TV	1.02	1.5	42.12
			유흥(술, 춤, 노래방)	8.45	0.54	16.24
	TV	34.9	간식과 음료	1.25	2.29	5.55
			아이에게 책 읽어주기, 놀아주기	1.04	0.46	1.11
			가족 친척과의 교제	1.02	1.5	3.63
	라디오	0.2	주업	2.6	0.17	15.6
			학교 외에서의 스스로 학습	3.86	0.16	14.95
			출·퇴근	5.04	0.11	10.33
CD, Tape 등 음악듣기	0.1	학교 외에서의 스스로 학습	5.5	0.17	21.28	
유흥(술, 춤, 노래방)	2.0	그 외 사람과의 교제	8.45	0.54	28.33	

4.4 이동

이동시간에 대한 분석은 출근에 해당하는 오전 8시와 퇴근 시간에 해당하는 오후 6시, 7시, 8시에 이동의 행동을 취하는 사람을 대상으로 동시행동을 분석하였으며, 동시행동에 대한 종류가 많아서 규칙도 많이 나왔으나, 지면 관계로 지지도가 0.01% 미만에 대해서는 생략하였다. <표 4>은 이동시간에 행하는 동시행동 분석의 결과이다.

<표 4> 이동시간에 행하는 동시행동 분석

시간	A 행동	A 행동의 행위자 비율	B 행동	향상도	지지도 (%)	신뢰도 (%)
08:00	개인유지관련 이동	0.3	그 외 사람과의 교체	1.3	0.01	2.27
			라디오	1.27	0.01	2.27
	출·퇴근	9.6	라디오	4.12	0.58	7.34
			CD, Tape 등 음악듣기	2.51	0.14	1.76
	그 외 일관련 이동	0.7	라디오	1.72	0.02	3.06
	통학 및 학습관련 이동	5.4	그 외 사람과의 교체	1.68	0.15	2.92
CD, Tape 등 음악듣기			2.4	0.09	1.69	
가족 보살피기 관련 이동	0.6	가족 친척과의 교체	3.38	0.02	2.79	
08:00	교제활동 관련 이동	1.4	라디오	3.58	0.04	6.37
			낮잠, 졸음	1.22	0.01	1.08
			가족 친척과의 교체	4.58	0.03	3.78
			라디오	4.4	0.07	7.84
			CD, Tape 등 음악듣기	1.92	0.01	1.35
	그 외 여가활동 관련 이동	0.7	낮잠, 졸음	1.62	0.01	1.43
			그 외 사람과의 교체	2.05	0.02	3.57
			라디오	2	0.02	3.57
CD, Tape 등 음악듣기	2.03	0.01	1.43			
18:00	출·퇴근	6.0	라디오	3.57	0.22	3.93
	그 외 일관련 이동	1.1	라디오	3.86	0.04	4.26
			담배피우기	3.14	0.01	1.18
	통학 및 학습관련 이동	1.9	라디오	1.18	0.02	1.3
			CD, Tape 등 음악듣기	1.85	0.02	1.3
	가족 보살피기 관련 이동	0.3	가족 친척과의 교체	1.49	0.01	2.05
	교제활동 관련 이동	2.7	라디오	1.46	0.05	1.61
			CD, Tape 등 음악듣기	2.86	0.06	2.01
그 외 여가활동 관련 이동	1.3	라디오	1.36	0.02	1.51	
		CD, Tape 등 음악듣기	1.9	0.02	1.34	

<표 4> 이동시간에 행하는 동시행동 분석(계속)

시간	A 행동	A 행동의 행위자 비율	B 행동	향상도	지지도 (%)	신뢰도 (%)
19:00	개인유지관련 이동	0.6	그 외 사람과의 교체	1.01	0.02	3.63
	출·퇴근	5.9	라디오	4.92	0.27	5.32
			CD, Tape 등 음악듣기	1.79	0.07	1.31
	그 외 일관련 이동	0.7	라디오	2.75	0.02	2.97
	통학 및 학습관련 이동	1.1	라디오	1.27	0.01	1.38
			CD, Tape 등 음악듣기	4.72	0.03	3.45
	가정관리 관련 이동	0.8	라디오	2.06	0.02	2.23
교제활동 관련 이동	2.3	라디오	2.74	0.07	2.96	
		CD, Tape 등 음악듣기	1.62	0.03	1.18	
그 외 여가활동 관련 이동	1.3	라디오	1.54	0.02	1.67	
20:00	개인유지관련 이동	0.4	라디오	1.8	0.01	1.75
	출·퇴근	3.5	라디오	5.11	0.15	4.98
			CD, Tape 등 음악듣기	2.61	0.06	2.06
	통학 및 학습관련 이동	0.9	라디오	2.11	0.02	2.06
			CD, Tape 등 음악듣기	2.98	0.02	2.35
	가족 보살피기 관련 이동	0.3	가족 친척과의 교체	1.86	0.02	6.5
	교제활동 관련 이동	3.4	낮잠, 졸음	2.45	0.03	1.12
라디오			3.12	0.07	3.04	
CD, Tape 등 음악듣기			1.54	0.03	1.22	
그 외 여가활동 관련 이동	1.1	라디오	2.12	0.02	2.07	
		CD, Tape 등 음악듣기	1.83	0.02	1.45	

<표 4>에서와 같이 4 가지의 시간대 공통으로 ‘출퇴근 시간에 라디오를 듣는다’의 신뢰도가 다른 동시행동에 비하여 높게 나타나고 있다. 이는 빠른 이동을 위하여 자가 운전자들이 교통정보를 많이 하는 ‘라디오’를 듣고 있는 것으로 예상된다. 이는 향상도에서도 비교적 높게 나타나 일반적인 행동에 비하여 이동 중에 라디오를 많이 청취하는 것으로 볼 수 있다. 또한 ‘CD나 Tape’보다는 이동 중에 아직은 ‘라디오’가 일반화되어 있었다. 그러나 오후 7시와 8시에서의 통학 및 학습관련 이동에서는 ‘라디오’에 비하여 ‘CD나 Tape 등의 음악 듣기’가 높은 것은 학생들은 대중교통을 이용하기 때문에 ‘라디오’보다는 손쉽게 접하기 때문이다.

‘오후 6시에 그외 일 관련 이동을 하면 담배를 피운다’에서 신뢰도는 1.18%로 별로

높지 않지만 향상도가 3.14로 높은 것은 다른 행위를 할 때보다 이동 중에 담배를 많이 피우기 때문이다.

5. 결론

보통의 사람들이 하루 24시간 중에서 많은 비중을 차지하는 개인유지(식사), 일 또는 학업, 교제 및 여가활동, 이동이란 주된 행동을 특히 많이 행하는 시간대에 4 가지의 행동과 동시에 행하는 동시행동을 연관성 규칙을 이용하여 분석하였다.

생활시간조사 결과 응답자들은 동시행동에 대하여 실제적 행동에 비해 낮은 응답률을 보이고 있는 것으로 추정된다. 이는 설문지 작성이 일과 후에 한꺼번에 기입을 하기에 그날의 행동을 기억에 의하여 작성하기 때문이다. 상대적으로 낮은 응답률에도 불구하고 동시행동에 대하여 연관성 규칙을 이용하여 분석한 결과 다음의 사실을 확인 할 수 있었다.

첫째, 식사에서 나온 몇 가지 규칙 중에서 신뢰도가 가장 높은 것은 ‘오후 7시에 가족, 친척과 식사를 하면 TV를 시청한다’로서 26.33%이다. 이때의 지지도는 4.03%이며, 지지도는 1.23이었다.

둘째, 일과 학습관련 시간에서 특징적인 것은 ‘오전 11시에 기타 일 관련 행동을 하면 가족, 친척 외의 사람과 교제를 한다’에서 신뢰도가 16.67%이며, 이 때의 향상도는 3.54로 나타났으며, ‘오전 11시에 학습관련 물품을 구입하면 가족, 친척 외의 사람과 교제를 한다’에서 신뢰도가 33.335로 나타났다. 이는 학생들이 물품을 구입할 때 혼자 가기보다는 여러 명이 같이 가기 때문이다.

셋째, 교제 및 여가활동에서는 ‘오후 7시에 가족, 친척과의 교제를 하면 가족과 식사를 한다’에서 신뢰도가 38.47%로 높은 비율을 보이며, 이때의 향상도는 2.52이다. ‘오후 9시에 유흥(술, 춤, 노래)을 하면 가족, 친척 외의 사람과 교제 중이다’에서는 신뢰도가 28.33%로 높게 나타난 것은 다른 시간에서도 비슷한 현상으로 지지도는 0.54로 별로 높지 않지만, 향상도가 8.45로 아주 높게 나타났다. ‘오후 9시에 그 외 사람과 교제를 하면 유흥(술, 춤, 노래)을 한다’에서도 신뢰도가 16.24%로 비교적 높게 나타난 것은 교제의 장소로 유흥주점이 일반화 된 것으로 예상할 수 있다.

넷째, 이동시간에서는 네 시간대 공통으로 ‘출퇴근 시간에 라디오를 듣는다’의 신뢰도가 다른 동시행동에 비하여 높게 나타나고 있다. 이는 빠른 이동을 위하여 자가 운전자들이 교통정보를 많이 하는 라디오를 듣고 있는 것으로 예상된다. 이는 향상도에서도 비교적 높게 나타나 일반적인 행동에 비하여 이동 중에 라디오를 많이 청취하는 것으로 볼 수 있다.

통계청의 분석에서는 동시행동에 대하여 분석을 따로 하지 않았으나 본 연구에서는 데이터마이닝의 연관성 규칙을 이용한 결과 특정행동을 동시에 한 비율과 특정 행위를 하였을 때 다른 행위를 수행할 확률을 구할 수 있었다. 그러나 본 연구에서는 동시행동에 대하여 많은 빈도가 나타나지 않아 지지도가 특히 낮은 한계점이 있다. 이는 조사시에 동시행동에 대하여 보다 신경을 기울인다면 동시행동에 대한 표준 분석법으로 연관성 규칙이 적용되리라 생각된다.

향후의 연구과제로는 소분류에 의한 분석은 모든 행위를 125개나 세분화하여 나누었기 때문에 지지도와 신뢰도가 낮게 나오는 경향이 있었다. 따라서 분류를 중분류

등과 같이 적게 사용하여 분석하고, 성별의 특성과 요일별 특성을 고려한다면 피부에 와 닿는 결과를 산출 할 것으로 생각된다.

참고문헌

1. 강현철, 한상태, 최종후, 김차용, 김은석, 김미경 (1999). *데이터마이닝-방법론 및 활용-*, 자유아카데미.
2. 김희재, 남기성 (2002). *한국노동자의 생활시간*, 자유아카데미.
3. 남기성, 김희재 (2003). Analysis of Simultaneous Activities on the Time Use Survey Using Data Mining, *2003 한국데이터 정보과학회 춘계학술발표회 논문집*.
4. 손애리 (2001). 「생활시간조사」 개발현황과 전망, 제7회 『통계의 날』 기념 세미나 발표집, 1-12.
5. 통계청 (2000). *1999 생활시간조사보고서*, 통계청.
6. Agrawal R., Imielinski R., Swami A. (1993). Mining association rules between sets of items in large databases, *In Proc. of the ACM SIGMOD Conference on Management of Data*, Washington, D.C.
7. Agrawal R., Mannila H., Srikant R., Toivonen H., and Verkamo A. (1996). *Fast discovery of association rules*, *Advances in Knowledge Discovery and Data Mining*, Cambridge MA: AAAI Press/MIT Press.
8. Bing L., Wynne H., Yiming M. (1999). Mining Association Rules with Multiple minimum Supports, *Proceedings of ACM KDD-99*.
9. Jiawei Han, Micheline Kamber (2001). *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers.
10. Silverstein C., Brin S., Motwani R. (1998). Beyond Market Baskets: Generalizing Association Rules to Dependence Rules, *Data Mining and Knowledge Discovery*, No.2, P 39-68.
11. Saygin Y., Vassilios S.V., Clifton C. (2002). Using Unknowns to Prevent Discovery of Association Rules, *2002 Conference on Research Issues in Data Engineering*.

[2003년 3월 접수, 2003년 7월 채택]