

Analysis of Simultaneous Activities on the Time Use Survey Using Data Mining

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 simultaneous activities. But if we use the association rule, we can found the ratio of particular activities at the same time. And we 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, 동시행동

1. 서론

동일한 자료를 이용하여 기존과 다른 분석을 통하여 새로운 결과를 얻는 것은 매우 중요한 일이다. 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 미비하고, 분석에 대한 다양한 연구가 부족하다. 따라서 본 논문에서는 데이터마이닝을 이용하여 자료의 가치를 높이고자 한다.

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.

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

데이터마이닝으로의 발전단계를 보면 1960년대 데이터 수집(Data Collection)에서 시작하여 1980년대에 데이터베이스와 같이 데이터 접근(Data Access)이 이루어졌고, 1990년대에 데이터웨어하우징(Data Warehousing)의 등장으로 2000년대에 데이터마이닝이 이루어졌다고 볼 수 있다. 이제 데이터마이닝은 기업 등에서만 사용하는 것이 아니고, 공학적 연구 및 인간 생활에 활용할 시기가 왔다고 한다. 본 논문은 이에 주안점을 두고 인간 생활사에서의 분석에서 데이터마이닝의 접근을 시도하는 것이다.

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 30여개국 이상에서 생활시간조사가 실시되고 있다.

그 동안 우리나라에서도 시간활용에 대한 연구와 조사가 전문가와 연구기관의 관심에 의해 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 살펴보고, 3장에서는 데이터마이닝에서의 기법 중에 하나인 연관성 규칙에 대하여 살펴보고, 4장에서는 분석의 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 개인 특성 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 행동을 한 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. 연관성분석

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 보험요율에 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 들

어있는 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 주행동과 더불어 어떤 행위를 하였는가 하는 빈도(Count)이다. 빈도를 기반으로 연관정도를 정량화하기 위해서 다음 3가지 기준을 고려한다.

관심 있는 규칙($A \rightarrow B$)에 대하여,

Support : $P(A \& B)$ = 전체 행위 중 A와 B를 동시에 행하는 비율은 어느 정도인가?

전체 행위수

Confidence : $P(B | A)$ = A를 행하는 사람들 중에서 B를 같이 행하는 비율은 어느 정도인가?

A를 동시에 행하는 수

Lift : $P(B | A) / P(B)$ = 임의로 B를 행하는 경우에 비해 A와의 관계가 고려되어 행하는 경우의 비?

연관성규칙에서는 Confidence의 값이 크면 좋지만 Confidence가 높은 연관규칙이 최상은 아니다. A와 B의 기본적으로 행위의 비율이 어느 정도 수준이 되어야 의미가 있을 것이다. 또한, $A \rightarrow B$ 가 높은 Confidence를 주지만, 원래 B행위를 행하는 기본 확률보다 커야 A를 고려해서 B를 생각하는 것이 의미가 있을 것이다. 즉, Lift는 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. 그리고 행동에 대한 소분류가 125개나 되기 때문에 결과에서는 Support와 Confidence를 기준으로 각각 상위 5개를 실었다. <표 1>은 매 정시의 동시행동에서 Support에 대한 결과이다.

<표 1> 매 정시 동시행동의 Support

시간	주행동과 동시행동	Support(%)	Lift	시간	주행동과 동시행동	Support(%)	Lift
0시	TV 시청, 간식과 음료	0.13	9.34	12시	가족 친척 외 교체, 가족외와 식사	1.34	1.48
	유흥(술, 춤, 노래방), 가족 친척 외 교체	0.11	48.93		라디오, 주업	0.40	1.67
	TV 시청, 가족친척과 교제	0.08	6.42		TV 시청, 가족과 식사	0.32	1.16
	라디오, 학교 외에서의 스스로 학습	0.08	11.03		가족 친척 외 교체, 간식과 음료	0.23	3.01
	독서, 라디오	0.07	25.46		가족친척과 교체, 가족과 식사	0.23	4.72
1시	유흥(술, 춤, 노래방) , 가족 친척외 교체	0.09	68.58	13시	가족 친척 외 교체, 가족외와 식사	1.42	1.35
	독서, 라디오	0.07	45.71		TV 시청, 가족과 식사	0.75	1.54
	라디오, 주업	0.07	10.54		가족 친척 외 교체, 간식과 음료	0.65	2.86
	CD, Tape 등 음악듣기, 학교 외에서의 스스로 학습	0.05	24.40		라디오, 주업	0.50	1.53
	주업, CD, Tape 등 음악듣기	0.04	8.18		가족친척과 교체, 가족과 식사	0.46	3.61
2시	라디오, 주업	0.05	20.85	14시	라디오, 주업	0.73	1.58
	CD, Tape 등 음악듣기, 주업	0.03	20.43		가족 친척 외 교체, 간식과 음료	0.53	3.24
	유흥(술, 춤, 노래방), 가족 친척외 교체	0.03	89.73		TV 시청, 가족과 식사	0.31	1.78
	독서, 라디오	0.02	83.92		TV 시청, 간식과 음료	0.27	1.32
	라디오, 학교 외에서의 스스로 학습	0.01	49.73		가족 친척 외 교체, 가족외와 식사	0.20	1.06
3시	라디오, 주업	0.03	24.63	15시	라디오, 주업	0.75	1.75
	CD, Tape 등 음악듣기, 주업	0.03	27.32		가족 친척 외 교체, 간식과 음료	0.64	2.54
	유흥(술, 춤, 노래방), 가족 친척외 교체	0.01	214.85		TV 시청, 간식과 음료	0.30	1.09
	주업, 간식과 음료	0.01	7.69		가족친척과 교체, 간식과 음료	0.14	2.18
	담배 피우기, 주업	0.01	16.81		TV 시청, 아이에게 책 읽어주기, 놀아주기	0.12	1.75
4시	라디오, 주업	0.04	26.99	16시	가족 친척 외 교체, 간식과 음료	0.69	2.45
	CD, Tape 등 음악듣기, 주업	0.02	23.85		라디오, 주업	0.63	1.47
	가족 친척 외 교체, 주업	0.01	8.73		가족친척과 교체, 간식과 음료	0.19	2.61
	간식과 음료, 주업	0.01	7.81		유흥(술, 춤, 노래방), 가족 친척외 교체	0.14	4.72
	출퇴근 위해 이동, 담배피우기	0.01	43.88		TV 시청, 아이에게 책 읽어주기, 놀아주기	0.13	1.71

<표 1> 매 정시 동시행동의 Support(계속)

시간	주행동과 동시행동	Support(%)	Lift	시간	주행동과 동시행동	Support(%)	Lift
5시	라디오, 주업	0.05	11.61	17시	라디오, 주업	0.53	1.44
	라디오, 식사준비	0.02	8.78		가족 친척 외 교제, 간식과 음료	0.41	2.49
	담배 피우기, 개인위생	0.02	7.56		TV 시청, 간식과 음료	0.32	1.06
	식사준비, TV 시청	0.02	6.84		TV 시청, 아이에게 책 읽어주기, 놀아주기	0.16	1.40
	라디오, 출퇴근 위해 이동	0.01	9.39		TV 시청, 가족친척과 교제	0.15	1.02
6시	TV 시청, 식사준비	0.21	1.19	18시	TV 시청, 가족과 식사	0.81	1.32
	TV 시청, 가족과 식사	0.17	5.07		TV 시청, 간식과 음료	0.30	1.07
	라디오, 식사준비	0.09	3.10		가족 친척 외 교제, 가족 외와 식사	0.29	1.73
	TV 시청, 가족 외와 식사	0.08	3.53		라디오, 주업	0.27	1.48
	출퇴근 위해 이동, 라디오	0.08	8.23		출퇴근 위해 이동, 라디오	0.22	3.57
7시	TV 시청, 가족과 식사	2.22	2.20	19시	TV 시청, 가족과 식사	4.03	1.23
	가족친척과 교제, 가족과 식사	0.45	5.96		가족친척과 교제, 가족과 식사	0.96	2.52
	TV 시청, 가족 외와 식사	0.41	1.84		TV 시청, 간식과 음료	0.62	1.38
	출퇴근 위해 이동, 라디오	0.30	5.53		가족 친척 외 교제, 가족외와 식사	0.49	2.53
	TV 시청, 옷 갈아입기(외모관리)	0.25	1.01		유흥(술, 춤, 노래방), 가족 친척외 교제	0.28	7.41
8시	TV 시청, 가족과 식사	2.35	1.83	20시	TV 시청, 간식과 음료	1.74	1.33
	출퇴근 위해 이동, 라디오	0.58	4.12		가족친척과 교제, 가족과 식사	0.65	1.80
	TV 시청, 가족 외와 식사	0.57	1.71		유흥(술, 춤, 노래방), 가족 친척 외 교제	0.44	7.53
	가족친척과 교제, 가족과 식사	0.31	3.73		TV 시청, 아이에게 책 읽어주기, 놀아주기	0.37	1.10
	TV 시청, 간식과 음료	0.29	1.56		가족 친척 외 교제, 간식과 음료	0.27	1.86
9시	TV 시청, 가족과 식사	0.93	2.06	21시	TV 시청, 간식과 음료	2.30	1.25
	라디오, 주업	0.58	1.55		TV 시청, 가족친척과 교제	1.50	1.02
	TV 시청, 간식과 음료	0.50	1.23		유흥(술, 춤, 노래방), 가족 친척외 교제	0.54	8.45
	TV 시청, 가족 외와 식사	0.34	1.48		TV 시청, 아이에게 책 읽어주기, 놀아주기	0.46	1.04
	가족 친척 외 교제, 간식과 음료	0.26	4.07		가족 친척 외 교제, 간식과 음료	0.25	1.67

<표 1> 매 정시 동시행동의 Support(계속)

시간	주행동과 동시행동	Support(%)	Lift	시간	주행동과 동시행동	Support(%)	Lift
10시	라디오, 주업	0.77	1.56	22시	TV 시청, 가족친척과 교체	1.11	1.14
	가족 친척 외 교체, 간식과 음료	0.46	3.11		TV 시청, 간식과 음료	1.09	1.31
	TV 시청, 가족과 식사	0.33	2.69		유흥(술, 춤, 노래방), 가족 친척 외 교체	0.45	11.02
	TV 시청, 가족 외와 식사	0.21	2.15		아이에게 책 읽어주기, 놀아주기, TV 시청	0.25	1.19
	TV 시청, 아이에게 책 읽어주기, 놀아주기	0.17	2.06		간식과 음료, 가족친척과 교체	0.24	2.50
11시	라디오, 주업	0.76	1.42	23시	TV 시청, 가족친척과 교체	0.47	1.68
	가족 친척 외 교체, 간식과 음료	0.47	3.37		TV 시청, 간식과 음료	0.38	1.92
	TV 시청, 간식과 음료	0.27	1.15		유흥(술, 춤, 노래방), 가족 친척외 교체	0.30	22.24
	TV 시청, 아이에게 책 읽어주기, 놀아주기	0.16	1.98		라디오, 학교 외에서의 스스로 학습	0.18	6.84
	TV 시청, 가족 외와 식사	0.15	2.18		독서, 라디오	0.14	8.09

<표 1>에서와 같이 전체 응답자 42,973명 중에서 0시(오후 12시)에 TV 시청과 간식과 음료를 먹는 동시행동은 0.13% 별로 많지는 않으나, TV 시청을 하면서 간식과 음료를 먹는 비율이 전체 응답자가 간식과 음료를 먹는 비율에 비해 9.34배나 많이 나타나, TV 시청이라는 행동과 간식과 음료를 먹는 행동은 높은 관계가 있다고 볼 수 있다.

오전 7시, 8시에는 TV를 시청하면서 가족과의 식사를 동시에 하는 사람은 각각 2.22%, 2.35%이며 이때의 Lift는 각각 2.20과 1.83으로 양의 관계가 있음이 나타났다. 즉, 일반적으로 알려진 바와 같이 식사 중에 TV를 시청하는 사람이 많은 것으로 분석할 수 있다. 이와 더불어 출근을 위해 이동 중에 라디오를 청취하는 비율도 0.30%와 0.58%로 비교적 상대적으로 높은 비율을 보이고 있다. 또한 이동 중에 라디오 청취는 Lift가 5.53과 4.12로 높은 관계를 가지고 있다. 그러나 퇴근시간은 분산되어 다른 동시행동에 비하여 상대적으로 낮게 나타났다.

일반적인 업무시간인 오전 10시와 11시에는 라디오를 들으면서 주업을 하는 비율이 0.77%와 0.76%로 Lift는 1.56과 1.42로 나타났으며, 마찬가지로 오후 업무시간인 2시에서 4시 사이에서도 주업과 라디오 청취의 동시 행동은 가장 높은 비율을 차지하고 있다.

본격적인 여가 시간인 18시 이후에는 TV 시청과 동시에 일어나는 여러 행동들이 높은 비율을 보이고 있으며, 가족 친척 외와의 교체 즉, 직장 동료 등과의 교체와 유흥(술, 춤, 노래방)의 동시 행동이 점차 높은 비율을 보이고 있다. 23시에 가족 친척 외와의 교체와 유흥(술, 춤, 노래방)의 동시 행동의 Lift는 22.24로 아주 높은 상관관계를 보이고 있다.

<표 2> 매 정시·동시행동의 Confidence

시간	Rule	Confidence(%)	Lift	시간	Rule	Confidence(%)	Lift
0시	재봉, 뜨개질 → TV 시청	50.00	12.94	12시	가족 친척외 교제 → 가족외와 식사	26.55	1.48
	옷정리 → TV 시청	50.00	12.94		라디오 → 주업	25.44	1.67
	잡지 → CD, Tape 등 음악듣기	50.00	141.35		국가 및 지역행사 관련 → 간식과 음료	25.00	16.45
	걷기, 산책 → 가족 친척외 교제	50.00	105.32		유홍(술, 춤, 노래방) → 가족 친척외 교제	22.50	4.46
	의료서비스 받기 → TV 시청	50.00	12.94		가족친척과 교제 → 가족과 식사	20.91	4.72
1시	걷기, 산책 → 담배피우기	50.00	934.11	13시	국가 및 지역행사 관련 → CD, Tape 등 음악듣기	50.00	50.79
	다름질, 바느질, 의류손질 → TV 시청	50.00	33.41		기타 관람 및 행사참여관련 행동 → 가족 친척외 교제	30.00	4.51
	의료서비스 받기 → TV 시청	50.00	33.41		가족친척과 교제 → 가족과 식사	28.14	3.61
	다름질, 바느질, 의류손질 → 가족 친척과 교제	50.00	228.56		라디오 → 주업	27.70	1.53
	가족과 식사 → TV 시청	40.00	26.73		가족 친척외 교제 → 가족외와 식사	21.41	1.35
2시	가정관리 관련 이동 → 라디오	100.00	671.39	14시	국가 및 지역행사 관련 → 라디오	50.00	26.54
	생활시간조사 설문지 기입 → CD, Tape 등 음악듣기	50.00	438.46		라디오 → 주업	38.57	1.58
	기타 여가 관련 활동 → 라디오	40.00	268.56		의류수선, 세탁서비스 받기 → 가족 친척외 교제	33.33	6.20
	라디오 → 주업	31.25	20.85		다름질, 바느질, 의류손질 → TV 시청	20.00	2.94
	CD, Tape 등 음악듣기 → 주업	30.61	20.43		기타 자원봉사(헌혈, 상담) → 독서	20.00	20.45
3시	가정관리 관련 이동 → 라디오	100.00	998.79	15시	라디오 → 주업	40.60	1.75
	잡지 → CD, Tape 등 음악듣기	100.00	1193.0		유홍(술, 춤, 노래방) → 가족 친척외 교제	28.93	4.83
	세탁 및 세탁물 널기 → TV 시청	50.00	363.97		기타 개인유지 → CD, Tape 등 음악듣기	20.00	24.61
	CD, Tape 등 음악듣기 → 주업	36.11	27.30		기타 일관련 행동 → 가족 친척외 교제	16.67	2.78
	가족 보살피기 관련 이동 → 독서	33.33	386.92		재봉, 뜨개질 → TV 시청	16.39	2.54

<표 2> 매 정시 동시행동의 Confidence(계속)

시간	Rule	Confidence(%)	Lift	시간	Rule	Confidence(%)	Lift
4시	컴퓨터 관련 학습 → 라디오	100.00	976.09	16시	라디오 → 주업	35.03	1.47
	아동 노인 장애인 관련 → 라디오	100.00	976.09		유홍(술, 춤, 노래방) → 가족 친척의 교제	28.05	4.72
	기타 여가 관련 활동 → CD, Tape 등 음악듣기	50.00	766.93		다름질, 바느질, 의류손질 → TV 시청	17.31	2.45
	기타 여가 관련 활동 → 라디오	50.00	488.05		가족과 식사 → TV 시청	16.81	2.38
	드라이브, 소풍, 관광 및 구경 → CD, Tape 등 음악듣기	50.00	766.93		자가 치료 → TV 시청	12.20	1.73
5시	아동 노인 장애인 관련 → 라디오	100.00	394.02	17시	가정계획(저축, 가족회의) → 담배피우기	50.00	91.77
	다름질, 바느질, 의류손질 → 라디오	100.00	394.02		라디오 → 주업	34.24	1.44
	컴퓨터 관련 학습 → 라디오	33.33	131.34		유홍(술, 춤, 노래방) → 가족 친척의 교제	23.61	4.39
	놀이(아이들 놀이, 바둑, 장기 등) → TV 시청	28.57	109.56		전시회 박물관 → 가족친척과 교제	20.00	13.44
	CD, Tape 등 음악듣기 → 주업	27.27	15.00		선생님과 상담 및 학교 방문 → 간식과 음료	20.00	6.54
6시	기타 여가 관련 활동 → 자가 치료	50.00	976.09	18시	라디오 → 주업	24.89	1.48
	가계부정리 → TV 시청	50.00	15.07		재봉, 뜨개질 → TV 시청	23.53	1.61
	숙제 및 공부 봐주기 → 방, 물품 정리	50.00	57.11		유홍(술, 춤, 노래방) → 가족 친척의 교제	22.88	6.13
	유홍(술, 춤, 노래방) → 가족 친척의 교제	22.22	149.13		라디오 → 출퇴근 위해 이동	20.25	3.56
	가족친척과 교제 → 가족과 식사	20.69	20.47		가족과 식사 → TV 시청	19.32	1.32
7시	드라이브, 소풍, 관광 및 구경 → 라디오	100.00	76.69	19시	재봉, 뜨개질 → TV 시청	58.82	2.75
	가족친척과 교제 → 가족과 식사	59.57	5.96		가족친척과 교제 → 가족과 식사	38.47	2.51
	비디오 → 가족과 식사	50.00	5.00		기타 개인유지 → TV 시청	33.33	1.56
	이미용관련 서비스 받기 → TV 시청	50.00	4.95		다름질, 바느질, 의류손질 → TV 시청	30.30	1.42
	가정계획(저축,가족회의) → 가족외와 식사	50.00	22.46		스포츠 경기 관람 → 간식과 음료	30.00	14.38

<표 2> 매 정시 동시행동의 Confidence(계속)

시간	Rule	Confidence(%)	Lift	시간	Rule	Confidence(%)	Lift
8시	기타 가사일 → 라디오	66.67	37.38	20시	재봉, 뜨개질 → TV 시청	64.71	1.97
	재봉, 뜨개질 → TV 시청	55.56	4.30		간식과 음료 → TV 시청	43.52	1.33
	비디오 → 설거지, 식후 정리	50.00	15.50		아이에게 책 읽어주기, 놀아주기 → TV 시청	36.10	1.10
	비디오 → 옷 갈아입기(외모관리)	50.00	17.84		취미활동 → TV 시청	36.76	1.12
	가족친척과 교제 → 가족과 식사	37.18	3.73		담배피우기 → TV 시청	34.86	1.06
9시	재봉, 뜨개질 → TV 시청	57.14	4.07	21시	아동 노인 장애인 관련 → 옷정리	100.00	298.25
	라디오 → 주업	34.64	1.55		재봉, 뜨개질 → TV 시청	64.44	1.56
	취미관련 강습 → 간식과 음료	33.33	11.52		간식과 음료 → TV 시청	51.52	1.25
	기타 일관련 행동 → TV 시청	33.33	2.37		자가 소비를 위한 농림어업일 → TV 시청	42.86	1.04
	가족과 식사 → TV 시청	28.89	2.06		아이에게 책 읽어주기, 놀아주기 → TV 시청	42.83	1.04
10시	가계부정리 → TV 시청	50.00	5.06	22시	재봉, 뜨개질 → TV 시청	62.75	2.16
	라디오 → 주업	38.60	1.56		기타 개인유지 → TV 시청	42.86	1.48
	가족과 식사 → TV 시청	26.63	2.69		간식과 음료 → TV 시청	38.04	1.31
	재봉, 뜨개질 → TV 시청	25.00	2.53		취미활동 → TV 시청	36.59	1.26
	가정계획(저축,가족회의) → TV 시청	25.00	2.53		아이에게 책 읽어주기, 놀아주기 → TV 시청	34.52	1.19
11시	연극, 콘서트 → 간식과 음료	50.00	17.06	23시	재봉, 뜨개질 → TV 시청	52.27	3.42
	가정계획(저축,가족회의) → 간식과 음료	50.00	17.06		취미활동 → TV 시청	41.54	2.72
	라디오 → 주업	38.86	1.42		이웃 및 친분있는 사람 돌기 → TV 시청	37.50	2.46
	유흥(술, 춤, 노래방) → 가족 친척외 교제	36.11	7.66		다름질, 바느질, 의류손질 → TV 시청	36.36	2.38
	국가 및 지역행사관련 → CD, Tape 등 음악듣기	33.33	35.35		일관련 물품구입 → CD, Tape 등 음악듣기	33.33	39.01

<표 2>는 매 정시의 동시행동에서 Confidence에 대한 상위 5개의 결과이다.

<표 2>에서와 같이 0시에 재봉, 뜨개질을 하는 사람의 50%는 TV 시청을 동시에 하고 있다는 법칙을 볼 수 있다. 이때의 Lift는 12.94로 높은 상관관계를 보이고 있다. 또한 잡지들

본다면 CD, Tape 등 음악듣기를 하는 경우는 Lift가 141.35로 아주 높은 상관관계를 보이고 있다.

6시에서 8시 사이에는 아침 식사시간으로 가족, 친척과의 교제 즉, 가족, 친척과의 대화 등과 같은 교제를 하는 사람들은 아침식사를 하는 비율이 높게 나타났다. 여기서 다른 법칙들이 상대적으로 높은 Confidence를 보이고 있는 것은 특정 행동을 하는 사람들의 수는 관계없이 집중적으로 동일한 다른 행동을 하면 Confidence가 높게 나타나기 때문이다.

10시에서 12시 사이와 14시에서 18시 사이의 주업무 시간에는 라디오를 듣는 사람이 주업을 행하고 있는 비율이 높게 나타났으며, 12시에 가족, 친척 외와 교제를 하는 사람이 식사를 하는 확률은 26.55%로 식사라는 생리적인 현상도 다른 이와 교제로서 여기고 있다.

여가시간인 18시 이후에는 재봉, 뜨개질하는 사람들은 반드시 TV를 시청하는 확률이 상대적으로 높게 나타나고 있다. 이는 다른 행위를 하는 사람에 비해 집중적으로 TV를 보기 때문이다.

5. 결론

생활시간조사에서 응답자들은 동시행동에 대하여 실제적 행동에 비하여 낮은 응답율을 보이고 있는 것으로 추정된다. 이는 설문지 작성이 일과 후에 한꺼번에 기입을 하기에 그날의 행동을 기억에 의하여 작성하기 때문이다. 그럼에도 불구하고 동시행동에 대하여 데이터마이닝을 이용하여 분석한 결과 다음의 사실을 확인 할 수 있었다.

support의 결과로서 TV 시청을 하면서 가족과의 식사를 동시에 하는 비율이 다른 동시행동에 비하여 높게 나타났으며, 출퇴근을 위한 이동 중에 라디오를 많이 청취하며, 주업과 라디오 청취가 동시 행동으로 높은 비율을 차지하고 있었다. 또한 여가 시간에 직장 동료 등과의 교제와 유흥(술, 춤, 노래방)의 동시 행동이 높은 비율을 보이고 있었다.

confidence의 결과에서와 같이 잡지를 볼 때는 CD, Tape 등 음악듣기를 하는 비율이 높으며, 6시에서 8시 사이에는 아침 식사시간으로 가족, 친척과의 교제 즉, 가족, 친척과의 대화 등과 같은 교제를 하는 사람들은 아침식사를 하는 비율이 높게 나타났다. 전체적으로 대중매체의 이용 중에서 TV 시청이 동시 행동으로 높은 비율을 차지하고 있다. 이는 <표 2>의 confidence에서 확연히 나타나고 있다.

통계청의 분석에서는 동시행동에 대하여 일반적인 분석만 행하였으나 본 연구에서는 데이터마이닝의 연관성 규칙을 이용한 결과 특정행동을 동시에 한 비율과 특정 행위를 하였을 때 다른 행위를 수행할 확률을 구할 수 있었다.

본 연구를 통하여 동시행동에 대한 분석 전에는 쉽게 나타나지 않은 개인별 번호의 부여와 응답자의 무성의한 응답과 같은 비표본오차가 있음을 볼 수 있었다. 향후의 조사에서는 행동분류에 대한 보다 많은 연구와 조사시 응답자에 대한 성의있는 응답에 보다 많은 신경을 기울여야 한다는 것을 얻은 것은 예상치 않은 결과라고 볼 수 있다.

향후의 연구과제로는 소분류에 의한 분석은 모든 행위를 125개나 세분화하여 나누었기 때문에 support와 confidence가 낮게 나오는 경향이 있었다. 따라서 분류를 중분류 등과 같이 적게 사용하여 분석하고, 성별의 특성과 요일별 특성을 고려한다면 피부에 와 닫는 결과를 산출 할 것으로 생각된다. 또한 24시간 전체를 사용하기 보다는 연구 목적에 맞는 계층과 시간을 사용한다면 데이터마이닝 본연의 목적인 예기치 않은 결과를 산출할 수 있으리라 생각된다.

참고문헌

1. 강현철, 한상태, 최중후, 김차용, 김은석, 김미경 (1999). *데이터마이닝-방법론 및 활용*, 자유아카데미.
2. 김희재, 남기성 (2002). *한국노동자의 생활시간*, 자유아카데미.
3. 손애리 (2001). 「생활시간조사」 개발현황과 전망, 제7회 『통계의 날』 기념 세미나 발표집, 1-12.
4. 통계청 (2000). *1999 생활시간조사보고서*, 통계청.
5. 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.
6. 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.
7. Bing L., Wynne H., Yiming M. (1999). *Mining Association Rules with Multiple minimum Supports*, proceedings of ACM KDD-99.
6. Jiawei Han, Micheline Kamber (2001). *Data Mining Concepts and Techniques*, *Morgan Kaufmann Publishers*.
8. 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.
9. Saygin Y., Vassilios S.V., Clifton C. (2002). *Using Unknowns to Prevent Discovery of Association Rules*, 2002 Conference on Research Issues in Data Engineering.