

Association Rule of Gyeongnam Social Indicator Survey Data for Environmental Information

Hee-Chang Park¹⁾ · Kwang-Hyun Cho²⁾

Abstract

Data mining is the method to find useful information for large amounts of data in database. It is used to find hidden knowledge by massive data, unexpectedly pattern, relation to new rule. The methods of data mining are decision tree, association rules, clustering, neural network and so on.

We analyze Gyeongnam social indicator survey data by 2001 using association rule technique for environment information. Association rule mining searches for interesting relationships among items in a given large data set. Association rules are frequently used by retail stores to assist in marketing, advertising, floor placement, and inventory control. There are three primary quality measures for association rule, support and confidence and lift.

We can use to environmental preservation and environmental improvement by association rule outputs

keywords : association rule, data mining, environmental information

1. 서론

변화하는 역사적 흐름 속에서 우리가 처해 있는 사회적 상태를 종합적이고 집약적으로 나타냄으로써 사회구성원들의 삶의 질을 전반적으로 파악하고 사회변화를 포착할 수 있는 척도가 바로 사회지표이다. 그동안 환경부문의 사회지표조사 데이터에 대해 자료의 수집과 기초적인 분석방법, 다변량 분석방법 등에 중점을 두고 연구가 활발히 이루어지고 있다(이상훈(1995), 이용우(1998), 정상용 등(1998), 최성우 등(2000), 문상기와 우남철(2001), 환경부(2001), 환경부(2002), 김정태 등(2003), 환경부(2003a),

1) First Author : Professor, Department of Statistics, Changwon National University, Changwon, Gyeongnam, 641-773, Korea
E-mail : hcpark@changwon.ac.kr

2) Graduate Student, Department of Statistics, Changwon National University, Changwon, Gyeongnam, 641-773, Korea
E-mail : cho1023@changwon.ac.kr

환경부(2003b) 등). 그러나 현재까지 환경분야에서는 방대한 양의 데이터베이스에서 잘 드러나지 않는 유용한 정보를 탐색하는 연구가 수행되지 않고 있다. 데이터의 양이 기하급수적으로 증가하고 있는 오늘날 거대한 데이터베이스에 내재되어 있는 유용한 정보를 탐색하여 의미 있는 지식을 발견하기 위한 연구의 필요성이 대두되고 있으며, 이를 위한 도구가 데이터마이닝이다.

본 연구는 경남지역 환경정보 분석을 위한 데이터마이닝 기법 중 연관성규칙기법을 제안함으로써 환경 정책을 수립하는데 있어서 필요한 의사결정 지원시스템 역할을 하는데 그 목적이 있다. 연관성 규칙은 탐색적이며, 비목적성 분석이며, 기존의 데이터를 특별한 변형 없이 계산이 용이하게 사용 가능하다는 장점을 가지고 있으며, 계산 과정이 길고, 반복된 계산이 많으며, 적절한 품목의 결정이 어렵고, 각 품목의 단위에 따른 표준화가 어렵다는 단점을 아울러 가지고 있다. 연관성 규칙은 이러한 단점에도 불구하고 두 품목간의 관계를 명확히 수치화함으로써 두 개 이상의 품목간의 관련성을 표시하여 주기 때문에 현업에서 많이 활용되고 있다.

연관성 규칙은 Agrawal 등(1993)에 의해 처음 소개된 이후, Agrawal 등(1994)은 후보 항목 집합을 구성하고, 발생 빈도수를 계산하고 난 후에 사용자가 정의한 최소 지지도를 기초로 빈발 항목 집합들을 결정하는 Apriori, AprioriTid 알고리즘을 제안하였다. Park 등(1995)은 데이터베이스를 중복되지 않는 크기로 분할하고 한번에 한 개의 분할 영역만을 고려하여 그 안에서 빈발 항목 집합을 생성하는 분할 알고리즘을 제안하였으며, Toivonen(1996)은 무작위로 선정된 표본을 가지고 빈발 항목 집합들을 찾은 후에 그 결과를 데이터베이스의 나머지 부분을 가지고 증명하는 샘플링 알고리즘을 제안하였다.

본 논문에서는 2001년 조사된 경상남도 사회지표조사 자료의 환경관련 설문에 대하여 내재되어 있는 정보를 추출하기 위해 연관성 규칙을 적용하였다. 본 논문의 2절에서는 자료 구축 및 자료 탐색에 대하여 기술하고 3절에서는 연관성 규칙의 분석 단계에 대하여 기술한다. 4절에서는 연관성 규칙을 이용한 자료 분석 결과를 기술한 후, 5절에서 결론을 맺는다.

2. 자료 구축 및 탐색

2.1 자료의 구조

경남사회지표조사의 자료구조는 크게 일반사항(인구통계학적 문항)과 도민의식조사 부분으로 나누어져 있다. 일반사항은 조사응답자의 연령, 성별, 학력, 가구주와의 관계, 결혼 유무, 직업으로 구성되어 있으며, 도민의식조사부분은 소득·소비, 보건·체육, 주택, 환경·교통, 사회, 정보화, 문화·여가의 총 7개 분야로 구성되어 있다. 조사대상 모집단은 경남 20개 시·군에 거주하는 17세 이상의 모든 도민과 가구이며, 표본 수는 400개 조사구의 10,000가구로 결정하였다. 표본의 배분은 인구 및 가구 비율에 따라 시·군별로 층화 추출하였고, 해당 시군에서의 표본수를 다시 인구와 가구 비율로 읍·면·동별로 층화하여 배분하였다. 표본수의 결정 후, 2000년 인구주택 총조사에서 사용된 조사구를 이용하여 표본조사구를 계통추출법을 사용하여 추출하였고, 추출된 각 표본조사구에서 2000년 인구주택 총 조사에서 사용된 가구명부를 기초로 하여, 25

개의 표본 가구를 다시 계통추출 하였다.

2.2 자료의 선정

2001년 경남사회지표조사에서 환경관련 문항과 인구통계학 속성 관련 문항, 집단구분문항을 추출하여 DB를 구축하였으며, 이 문항들을 환경관련 문항 부문, 인구통계학적 문항 부문, 구분 문항 부문으로 분류하였다. 설문 문항은 다음과 같다.

<표 2.1> 경남사회지표조사 환경관련 문항

순번	문항	응답 선택지
1	지역의 상수도 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
2	지역의 하수도 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
3	지역의 소음진동 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
4	지역의 악취 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
5	지역의 대기 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
6	지역의 토양 오염도	1.아주심함 2.심함 3.보통 4.좋음 5.아주좋음
7	쓰레기 종량제 봉투의 사용 만족도	1.아주만족 2.만족 3.보통 4.불만 5.아주불만

<표 2.2> 경남사회지표조사 인구통계학적 문항

순번	문항	응답 선택지
1	연령	()세
2	성별	1. 남자 2. 여자
3	학력	1. 무학 2. 초졸 3. 중졸 4. 고졸 5. 전문대학재학 6. 전문대졸 7. 대학재학 8. 대졸 9. 대학원이상
4	가구주와의 관계	1. 본인 2. 부부 3. 부모 4. 자녀 5. 형제 6. 친지 7. 동거인 8. 기타
5	결혼유무	1. 미혼 2. 유배우 3. 사별 4. 이혼
6	직업	1. 의회의원, 고위임직원 및 관리자 2. 전문가 3. 기술공 및 준전문가 4. 사무종사자 5. 서비스 종사자 6. 판매 종사자 7. 농업, 임원 및 어업 숙련 종사자 8. 기능원 및 관련 기능 종사자 9. 장치, 기계조작 및 조립 종사자 10. 단순노무 종사자 11. 군인 12. 가정주부 13. 학생 14. 무직 15. 기타
7	조사지역	1. 농촌지역 2. 어촌지역 3. 상가지역 4. 주거지역 5.공업지역 6. 기타지역
8	설문지 작성자	1. 가구주 2. 배우자 3. 가구원
9	시,군명	1. 창원시 2. 마산시 3. 진주시 4. 진해시 5. 통영시 6. 사천시 7. 김해시 8. 밀양시 9. 거제시 10. 양산시 11. 의령군 12. 함안군 13. 창녕군 14. 고성군 15. 남해군 16. 하동군 17. 산청군 18. 함양군 19. 거창군 20. 합천군
10	월평균소득	()원
11	월평균저축	()원
12	거주 주택 소유형태	1. 자가 2. 전체 3. 보증부 월세 4. 월세 5. 기타
13	거주 주거형태	1. 단독주택 2. 아파트 3. 연립주택 4. 다세대주택 5. 기타
14	주택 사용방수	()개
15	주택 면적	()m ²

<표 2.3> 경남사회지표조사의 집단 구분

순번	문항	응답 선택지
1	향후 2-3년 이후 가구소득	1. 많이나아짐 2. 조금좋아짐 3. 전과동일 4. 조금나빠짐 5. 많이나빠짐
2	의료시설과 서비스 만족도	1. 매우만족 2. 대체로 만족 3. 보통 4. 대체로 불만 5. 매우불만
3	주관적 건강상태	1. 아주 건강 2. 건강 3. 보통 4. 허약 5. 아주 허약
4	거주지 주택규모의 만족도	1. 매우 불만 2. 불만 3. 보통 4. 조금 만족 5. 매우 만족
5	거주지 일조통풍의 만족도	1. 매우 불만 2. 불만 3. 보통 4. 조금 만족 5. 매우 만족
6	거주지 상하수도시설의 만족도	1. 매우 불만 2. 불만 3. 보통 4. 조금 만족 5. 매우 만족
7	주관적 사회계층	1. 상류층 2. 중상류층 3. 중류층 4. 중하류층 5. 하류층
8	자치단체의 업무 인지 방법	1. 대중매체 2. 홍보자료 3. 인터넷 4. 반사회및통반장 5. 주위사람 6. 기타
9	자치단체의 주민여론 반영도	1. 아주 잘 반영 2. 잘 반영 3. 보통 4. 잘반영 안됨 5. 전혀 반영 안됨.
10	시책추진의 우선적 처리사항	1. 물가 2. 주택 3. 실업 4. 지역경제개발 5. 노사분규 6. 환경개선 7. 사회복지시설 8. 사회안정 9. 문화여가시설 10. 보건위생 11. 기타

자료의 정제는 데이터의 충실도와 분석 결과의 정확성을 기하기 위해 결측치를 분석대상에서 제외시킨 후 새로운 DB를 구축하는 과정으로 본 조사자료에서는 이들을 처리한 후 정제된 DB를 구축하였다.

2.3 자료의 변환

사회지표조사의 자료에 대하여 환경 관련 문항에 대하여 분석 결과의 해석을 용이하게 하기 위해 '쓰레기 종량제 봉투의 사용 만족도' 문항은 1. 아주불만 2. 불만 3. 보통 4.만족 5. 아주만족으로 변환하였다. 구분문항에 대하여 '향후 2-3년 이후 가구소득' 문항은 1. 많이 나빠짐 2. 조금 나빠짐 3. 전과동일 4. 조금 좋아짐 5. 많이 좋아짐으로 변환하였으며, '의료시설과 서비스 만족도', '주관적 건강상태', '주관적 사회계층', '자치단체의 주민여론 반영도' 등의 문항은 '향후 2-3년 이후 가구소득' 문항과 마찬가지로 응답선택지를 1→5, 2→4, 4→2, 5→1로 변환하였다. 그 다음에는 연관성 규칙을 적용하기 위하여 <표 2.4>, <표 2.5>, <표 2.6>과 같이 2차 변환을 실시하였다.

<표 2.4> 환경관련문항 자료변환

순번	문항	문항 보기	변경내용
1	지역의 상수도 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
2	지역의 하수도 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
3	지역의 소음진동 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
4	지역의 악취 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
5	지역의 대기 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
6	지역의 토양 오염도	1. 좋음 2. 보통이하	1•2•3 → 2, 4•5 → 1
7	경제발전과 환경보존 선호도	1. 경제개발 2. 환경보존	1•2 → 1, 3•4 → 2
8	환경오염의 해결주체	1. 행정당국 2. 그외	1 → 1, 2•3•4 → 2
9	쓰레기 종량제 봉투의 사용 만족도	1. 만족 2. 보통이하	1•2•3 → 2, 4•5 → 1

<표 2.5> 인구통계학 문항자료 변환

순번	문항	문항 보기	변경
1	연령	1. 평균초과 2. 평균이하	42.49 이하 → 2 , 42.49 초과 → 1
2	성별	1. 여 2. 남	1 → 2 , 2 → 1
3	학력	1. 대재이상 2. 고졸이하	1,2,3,4 → 2 , 5,6,7,8,9 → 1
4	가구주와의 관계	1. 그외 2. 본인	1 → 2 , 2,3,4,5,6,7,8, → 1
5	결혼유무	1. 그외 2. 미혼	1 → 2 , 2,3,4 → 1
6	직업	1. 그외 2. 가정주부	13 → 2 , 1,2,3,4,5,6,7,8,9,10,11,13,14,15 → 1
7	조사지역	1. 그외 2. 농어촌	1,2 → 2 , 3,4,5,6, → 1
8	설문지 작성자	1. 그외 2. 가구주	1 → 2 , 2,3 → 1
9	시, 군명	1. 군 2. 시	1,2,3,4,5,6,7,8,9,10 → 2 , 11,12,13,14,15,16,17,18,19,20 → 1
10	월평균소득	1. 평균초과 2. 평균이하	143.51 이하 → 2 , 143.51 초과 → 1
11	월평균지출	1. 평균초과 2. 평균이하	31.36 이하 → 2 , 31.36 초과 → 1
12	거주 주택 소유형태	1. 그외 2. 자가	1 → 2 , 2,3,4,5 → 1
13	거주 주거형태	1. 그외 2. 단독주택	1 → 2 , 2,3,4,5 → 1
14	주택 사용방수	1. 평균초과 2. 평균이하	2.91 이하 → 2 , 2.91 초과 → 1
15	주택 면적	1. 평균초과 2. 평균이하	22.76 이하 → 2 , 22.76 초과 → 1

<표 2.6> 구분문항 자료변환

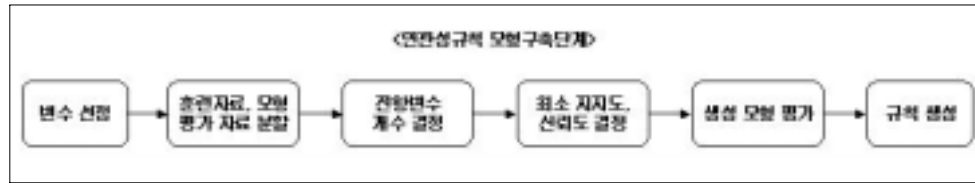
순번	문항	문항 보기	변경
1	항후 2-3년 이후 가구소득	1.보통이상 2.나빠짐	1,2 → 2 , 3,4,5 → 1
2	의료시설과 서비스 만족도	1.불만 2. 불만	1,2 → 2 , 3,4,5 → 1
3	주관적 건강상태	1.보통이상 2. 좋지않음	1,2 → 2 , 3,4,5 → 1
4	거주지 주택규모의 만족도	1.보통이상 2. 불만	1,2 → 2 , 3,4,5 → 1
5	거주지 일조통풍의 만족도	1.보통이상 2. 불만	1,2 → 2 , 3,4,5 → 1
6	거주지 상하수도시설의 만족도	1.보통이상 2. 불만	1,2 → 2 , 3,4,5 → 1
7	주관적 사회계층	1.중산층이상 2.중산층미만	1,2 → 2 , 3,4,5 → 1
8	자치단체의 업무 인지 방법	1. 그외 2. 언론매체 및 온라인	1,2,3 → 2 , 4,5,6 → 1
9	자치단체의 주민여론 반영도	1.보통이상 2. 불만	1,2 → 2 , 3,4,5 → 1
10	시책추진의 우선적 처리사항	1.그외 2. 환경개선	6 → 2 , 1,2,3,4,5,7,8,9,10,11 → 1

2.4 자료의 탐색

경남 사회 지표조사에서 구축된 환경 관련 문항의 기초 통계량 등을 토대로 하여 자료 탐색을 한 결과 일차원적인 정보를 얻을 수 있었다. 그러나 이 정보만으로는 데이터의 속성이나 문항간의 관련성을 설명하기에는 부족하다. 그러므로 환경관련 데이터에 내재되어 있는 정보를 추출하기 위하여 연관성 규칙을 적용한다.

3. 연관성 규칙 모형화

연관성 규칙 생성 단계는 다음과 같다. 본 논문에서는 연관성 규칙 적용을 위하여 SPSS의 Clementine 10.0을 사용하였다.



〈그림 3.1〉 연관성 규칙 분석 단계

[단계 1] 변수선정

연관성 규칙을 생성하기 위하여 전항변수와 후항변수를 선정한다. 후항변수에는 환경관련 문항을 선정하고 전항변수에는 인구통계학적 문항과 구분문항을 선정한다.

[단계 2] 자료 분할

일반적인 연관성규칙은 훈련자료와 모형평가 자료로 분할하여 규칙을 생성하지 않으나 본 연구의 목적이 연관성 규칙 모형의 개발이고 연관성 규칙에 의해 생성된 모형의 평가 또한 중요하기 때문에 훈련자료와 모형평가 자료를 각각 1/2, 1/2로 분할한다.

[단계 3] 전항변수 개수 결정

환경관련 문항에 대하여 인구통계학 문항, 구분문항간의 관련성을 알아보기 위하여 전항변수의 개수를 결정해야 한다. 전항변수의 개수가 적으면 연관성 규칙 모형의 정확도는 증가하나 의미 있는 규칙을 찾아내지 못할 수 있고, 전항변수의 개수가 많으면 많은 규칙을 생성하나 연관성 규칙 모형의 정확도가 감소할 수 있다. 연관성 규칙의 신뢰도와 규칙 생성을 고려하여 전항변수의 개수를 2로 설정하였다.

[단계 4] 최소지지도, 신뢰도 결정

연관성 규칙 생성에 있어 최소지지도와 신뢰도를 결정해야 한다. 최소지지도와 신뢰도를 낮게 결정하면 연관성 규칙의 생성이 많아지나 의미 없는 규칙이 생성될 수 있고, 최소지지도와 신뢰도를 높게 하면 의미 있는 규칙을 찾아내지 못하게 되는 경우도 있다. 이에 최소지지도와 신뢰도를 조절하여 규칙을 생성한 결과 최소지지도 10, 신뢰도를 80으로 결정했을 때 의미 있는 규칙을 발견할 수 있었다.

[단계 5] 생성모형 평가

훈련 자료에 의한 연관성 규칙 모형에 대하여 모형이 제대로 구축되었는지 평가하기 위하여 모형 평가 자료에 의하여 모형을 평가한다. 모형평가 시 정확도가 낮게 나타나면 모형을 다시 생성해야 한다.

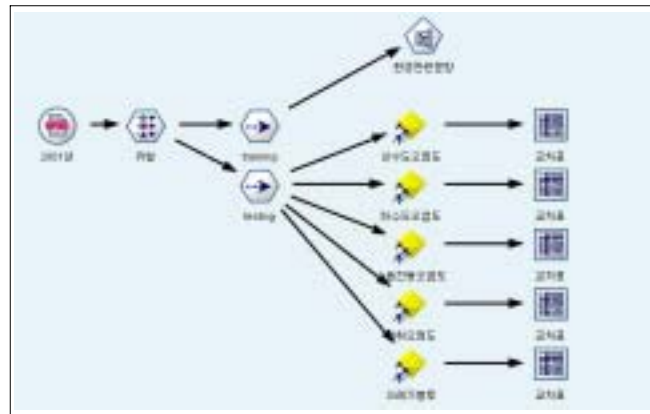
[단계 6] 규칙생성

모형 평가에서 모형이 적합하다고 판단되면 환경관련 문항 각각에 대하여 연관성 규칙을 생성한다.

4. 연관성 규칙의 적용 결과

4.1 연관성 규칙 생성 스트림

연관성규칙 생성 스트림은 <그림 4.1>과 같다.



<그림 4.1> 연관성 규칙 생성 스트림

4.2 연관성 규칙 생성

연관성 규칙에서는 각각의 환경 문항에 대하여 인구통계학적 문항과 구분 문항간의 관련성을 알아보기 위하여 분석을 실시한다. 각 문항에 대한 분석 결과 총 5문항에 대하여 연관성 규칙이 발견되었다. 결과는 다음과 같다.

(1) 상수도 오염도의 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	15.9	88.0	거주지 상하수도 시설의 만족도 = 불만	
2	13.3	89.9	거주지 상하수도 시설의 만족도 = 불만	시군 = 시
3	12.1	89.8	거주지 상하수도 시설의 만족도 = 불만	자치단체의 업무인지방법 = 언론매체 및 온라인
4	11.1	89.0	거주지 상하수도 시설의 만족도 = 불만	주관적 사회계층 = 중산층 미만
5	12.1	87.1	거주지 상하수도 시설의 만족도 = 불만	학력 = 고졸 이하
6	10.7	86.7	거주지 상하수도 시설의 만족도 = 불만	월평균 저축액 = 평균 이하
7	15.3	87.6	거주지 상하수도 시설의 만족도 = 불만	시책추진의 우선적처리사항 = 환경개선
8	11.7	85.7	거주지 일조통풍의 만족도 = 불만	연령 = 평균이하

상수도 오염도의 연관성 규칙 결과 총 8개의 규칙이 생성되었다. 거주지 상하수도 시설의 만족도가 불만인 집단은 상수도 오염도도 보통이하라고 생각하는 것으로 나타났다. 이들 중에서는 시에 사는 집단, 자치단체의 업무인지 방법이 언론 매체나 인터넷인 집단, 주관적 사회계층이 중산층 미만인 집단, 학력이 고졸 이하인 집단, 월평균

저축액이 평균 이하인 집단, 그리고 시책추진의 우선적 처리사항이 환경개선인 집단은 상수도 오염도가 보통이하라고 생각하였다. 또한 거주지 일조통풍의 만족도가 불만이고 연령이 평균 이하인 집단은 상수도 오염도가 보통이하라고 생각하는 것으로 나타났다.

(2) 하수도 오염도의 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	15.9	92.4	거주지 상하수도 시설의 만족도 = 불만	
2	13.3	93.2	거주지 상하수도 시설의 만족도 = 불만	시군 = 시
3	12.1	94.2	거주지 상하수도 시설의 만족도 = 불만	자치단체의 업무인지방법 = 언론매체 및 온라인
4	10	93.8	거주지 상하수도 시설의 만족도 = 불만	연령 = 평균이하
5	15.3	92.2	거주지 상하수도 시설의 만족도 = 불만	시책추진의 우선적 처리사항 = 환경개선
6	12.1	91.9	거주지 상하수도 시설의 만족도 = 불만	학력 = 고졸 이하
7	11.1	92.2	거주지 상하수도 시설의 만족도 = 불만	주관적 사회계층 = 중산층 미만
8	17.2	87.2	거주지 일조통풍의 만족도 = 불만	
9	11.7	89.9	거주지 일조통풍의 만족도 = 불만	연령 = 평균이하
10	13.4	88.7	거주지 일조통풍의 만족도 = 불만	자치단체의 업무인지방법 = 언론매체 및 온라인
11	14.7	88.5	거주지 일조통풍의 만족도 = 불만	시군 = 시
12	16.4	86.9	거주지 일조통풍의 만족도 = 불만	시책추진의 우선적 처리사항 = 환경개선
13	10.2	86.4	거주지 일조통풍의 만족도 = 불만	거주지 주택 규모의 만족도 = 불만
14	12.8	56.1	거주지 일조통풍의 만족도 = 불만	주관적 사회계층 = 중산층 미만
15	12.8	85.8	거주지 일조통풍의 만족도 = 불만	학력 = 고졸 이하
16	10.2	86.4	월평균 소득 = 평균 미만	월평균 저축액 = 평균이하
17	19.3	86.3	성별 = 남	자치단체의 업무인지방법 = 언론매체 및 온라인
18	14.9	86.2	성별 = 남	학력 = 고졸 이하
19	15.1	56.1	연령 = 평균이하	학력 = 고졸 이하

하수도 오염도의 연관성 규칙 결과 총 19개의 규칙이 생성되었다. 먼저 거주지 상하수도 시설의 만족도가 불만인 집단은 하수도 오염도가 보통이하로 생각하는 것으로 나타났다. 이들 중에서는 시에 사는 집단, 자치단체의 업무인지 방법이 언론매체 및 온라인인 집단, 연령이 평균이하인 집단, 시책추진의 우선적 처리사항이 환경개선인 집단, 학력이 고졸이하인 집단, 그리고 주관적 사회계층이 중산층 미만인 집단은 하수도 오염도가 보통이하라고 생각한다. 또한 거주지 일조통풍의 만족도가 불만인 집단은 하수도 오염도가 보통이하라고 생각하였다. 이들 중에서는 연령이 평균이하인 집단, 자치단체의 업무인지 방법이 언론매체 및 온라인인 집단, 시에 사는 집단, 시책추진의 우선적 처리사항이 환경개선인 집단, 거주지 주택 규모의 만족도가 불만인 집단, 주관적 사회계층이 중산층 미만인 집단, 학력이 고졸이하인 집단은 하수도 오염도가 보통이하로 생각하는 것으로 나타났다. 그리고 월평균 소득이 평균 미만이고 월평균 저축액이 평균 미만인 집단, 성별이 남자이고 자치단체의 업무인지 방법이 언론매체 및 온라인인 집단과 학력이 고졸이하인 집단, 연령이 평균 이하이고 학력이 고졸이하인 집단은 하수도 오염도가 보통이하로 생각하였다.

(3) 소음진동 오염도의 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	11.7	87.4	거주지 일조통풍의 만족도 = 불만	연령 = 평균이하

소음진동 오염도의 연관성 규칙 결과 총 1개의 규칙이 생성되었다. 거주지 일조통풍의 만족도가 불만이고 연령이 평균이하인 집단은 소음진동 오염도가 보통이하라고 생각하는 것으로 나타났다.

(4) 악취 오염도의 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	10	86.6	거주지 상하수도 시설의 만족도 = 불만	연령 = 평균이하
2	11.1	85.4	거주지 상하수도 시설의 만족도 = 불만	주관적 사회계층 = 중산층 미만
3	17.2	86.1	거주지 일조통풍의 만족도 = 불만	
4	11.7	89.6	거주지 일조통풍의 만족도 = 불만	연령 = 평균이하
5	16.4	86.1	거주지 일조통풍의 만족도 = 불만	시책추진의 우선적 처리사상 = 환경개선
6	10.2	85.4	거주지 일조통풍의 만족도 = 불만	거주지 주택 규모의 만족도 = 불만
7	12.8	87.7	거주지 일조통풍의 만족도 = 불만	주관적 사회계층 = 중산층 미만
8	12.8	86.4	거주지 일조통풍의 만족도 = 불만	학력 = 고졸 이하
9	11.6	85.9	거주지 일조통풍의 만족도 = 불만	월평균 저축액 = 평균이하
10	13.4	86.6	거주지 일조통풍의 만족도 = 불만	자치단체의 업무인지방방법 = 언론매체 및 온라인

악취 오염도의 연관성 규칙 결과 총 10개의 규칙이 생성되었다. 거주지 상하수도 시설의 만족도가 불만인 집단 중에서 연령이 평균이하인 집단과 주관적 사회계층이 중산층 미만인 집단은 악취 오염도가 보통이하로 생각하는 것으로 나타났다. 또한 거주지 일조통풍의 만족도가 불만인 집단은 악취 오염도도 심하다고 생각하였다. 이들 중에서 연령이 평균이하인 집단, 시책추진의 우선적 처리사항이 환경개선인 집단, 거주지 주택 규모의 만족도가 불만인 집단, 주관적 사회계층이 중산층 미만인 집단, 학력이 고졸이하인 집단, 월평균 저축액이 평균이하인 집단, 그리고 자치단체의 업무인 지방방법이 언론매체 및 온라인인 집단은 악취 오염도가 보통이하라고 생각하였다.

(5) 쓰레기 종량제 봉투의 만족도에 대한 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	15.9	95.8	거주지 상하수도 시설의 만족도 = 불만	
2	13.3	96.8	거주지 상하수도 시설의 만족도 = 불만	시군 = 시
3	12.1	96.2	거주지 상하수도 시설의 만족도 = 불만	자치단체의 업무인지방방법 = 언론매체 및 온라인
4	10	97.4	거주지 상하수도 시설의 만족도 = 불만	연령 = 평균이하
5	15.3	95.7	거주지 상하수도 시설의 만족도 = 불만	시책추진의 우선적 처리사항 = 환경개선
6	12.1	95.7	거주지 상하수도 시설의 만족도 = 불만	학력 = 고졸 이하
7	11.1	96.4	거주지 상하수도 시설의 만족도 = 불만	주관적 사회계층 = 중산층 미만
8	10.7	95.1	거주지 상하수도 시설의 만족도 = 불만	월평균 저축액 = 평균이하
9	11.7	95.1	거주지 일조통풍의 만족도 = 불만	연령 = 평균이하
10	23.2	95.9	의료시설과 서비스 만족도 = 불만	연령 = 평균이하
11	34.1	95.4	의료시설과 서비스 만족도 = 불만	시군 = 시
12	15.9	95.6	의료시설과 서비스 만족도 = 불만	자치단체의 주민여론 반영도 = 불만
13	11.0	95.8	의료시설과 서비스 만족도 = 불만	거주지 주택규모의 만족도 = 불만

쓰레기 종량제 봉투의 만족도의 연관성 규칙 결과 총 13개의 규칙이 생성되었다. 거주지 상하수도 시설의 만족도가 불만인 집단은 쓰레기 종량제 봉투의 만족도가 보통이하라고 생각하고 있는 것으로 나타났다. 이들 중에서는 시에 사는 집단, 자치단체의 업무인지방방법이 언론매체 및 온라인인 집단, 연령이 평균이하인 집단, 시책추진의 우선적 처리사항이 환경개선인 집단, 학력이 고졸 이하인 집단, 주관적 사회계층이 중산층 미만인 집단, 그리고 월평균 저축액이 평균이하인 집단은 쓰레기 종량제 봉투의 만족도가 보통이하라고 생각하였다. 또한 거주지 일조통풍의 만족도가 불만이고 연령이 평균이하인 집단은 쓰레기 종량제 봉투의 만족도가 보통이하라고 생각하였다. 의료시설과 서비스 만족도가 불만인 집단 중에서는 연령이 평균이하인 집단, 시에 사는 집단, 자치단체의 주민여론 반영도가 불만인 집단, 그리고 거주지 주택규모의 만족도가 불만인 집단은 쓰레기 종량제 봉투의 만족도가 보통이하로 생각하는 것으로 나타났다.

5. 결론

환경분야에서는 방대한 양의 데이터베이스에 내재되어 있는 정보를 탐색하여 의미 있는 지식을 발견하기 위한 연구의 필요성이 대두되고 있다. 이를 위해 본 논문에서는 2001년 조사된 경상남도 사회지표조사 자료의 환경관련 실문에 대하여 자료의 구축 및 탐색을 통하여 인구통계학적 문항뿐만 아니라 환경문항과 관련성이 있을 것으로 추정되는 구분문항들을 추출하여 연관성 규칙을 적용하였다. 환경관련 문항에 대하여 연관성 규칙을 적용한 결과, 쉽게 드러나지 않는 유용한 정보를 추출할 수 있었다. 향후 본 논문의 결과를 바탕으로 지역 여건 등의 데이터와 연계함으로써 환경 정보화를 통하여 환경개선대책 수립과 환경 정책 결정에 필요한 의사결정 지원 등 효율적인 환경행정의 수행과 환경 정책 수립에 기여할 수 있을 것이다.

참고문헌

1. 김정태, 정진도, 김광석(2003), 여름철 충청남도 서북부 지역에서의 대기오염물질 농도 분포특성에 관한 연구, *대한환경공학회 2003 추계학술발표회 논문집*, 1326-1328
2. 문상기, 우남철(2001), 통계분석을 이용한 지하수위 변동 특성 분류, *한국지하수도양환경학회 01 추계학술발표회논문집*, 2001권, 155-159
3. 이상훈(1995), 수질자료의 추세분석을 위한 비모수적 통계검정에 관한 연구, *환경영향평가*, 제4권 제2호, 93-103
4. 이용우(1998), 폐기물 배출량의 지역간 차이에 관한 분석, *대한지리학회* 33권 2호, 209-224
5. 정상용, 강동환, 심병완(1998), 부산지역 지하수의 수질오염 특성, *한국지하수도양환경학회 98 공동심포지엄 및 추계학술발표회 논문집*, 1998권, 86-92

6. 최성우, 송형도(2000), 다변량 통계분석법을 이용한 대구지역 부유분진의 오염원 기여도 추정, *한국환경위생학회지*, 제26권 제4호, 1-8
7. 환경부(2001), *전국폐기물통계조사*.
8. 환경부(2002), *전국폐기물발생현황*.
9. 환경부(2003a), *환경통계연감*.
10. 환경부(2003b), *대기환경연보*.
11. 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.
12. Agrawal R., Srikant R.(1994), Fast algorithms for mining association rules, *In Proc. of the 20th VLDB Conference, Santiago, Chile*.
13. Park J.S., Chen M.S., and Philip S.Y.(1995), An effective hash-based algorithms for mining association rules, *In Proc. of ACM SIGMOD Conference on Management of Data*.
14. Toivonen H.(1996), Sampling Large Database for Association Rules, *Proceedings of the 22nd VLDB Conference Mumbai(Bombay), India*.

[2004년 11월 접수, 2005년 2월 채택]