

Environmental Consciousness Data Modeling by Association Rules

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 association rules, decision tree, clustering, neural network and so on. 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 analyze Gyeongnam social indicator survey data using association rule technique for environmental information discovery. We can use to environmental preservation and environmental improvement by association rule outputs.

Keywords : association rule, data mining, environmental information

1. 서론

환경정보화는 최신 정보기술을 이용하여 환경관련 정보를 생산·수집·가공·처리·유통 또는 활용함으로써 환경행정업무의 효율화를 도모함을 그 목적으로 한다. 점차 광역화·다양화·복잡화되고 있는 환경문제를 보다 효과적으로 해결하고 예방하기 위해서는 더욱 광범위한 정보를 요구하게 되었고, 이러한 많은 정보를 적기에 확보하여 환경행정에 활용할 수 있는 체계를 구비하여 환경정책에 필요한 고급적인 정보를 도출하는 업무의 비중이 높아지고 있다(환경부(2004)).

환경데이터에 대해 그 동안 환경관련분야에서는 자료의 수집과 기초적인 분석방법, 다변량 분석방법 등에 중점을 두고 연구가 활발히 이루어지고 있다(이상훈(1995), 이

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

3) Graduate Student, Department of Statistics, Changwon National University, Changwon, Kyungnam, 641-773, Korea

용우(1998), 정상용 등(1998), 최성우와 송형도(2000), 문상기와 우남철(2001), 환경부(2001), 환경부(2002a), 환경부(2002b), 환경부(2002c), 환경부(2002d), 김정태 등(2003), 환경부(2003a), 환경부(2003b) 등). 그러나 데이터의 양이 기하급수적으로 증가하고 있는 오늘날 방대한 양의 데이터베이스(database : DB)에 내재되어 있는 유용한 정보를 탐색하여 의미 있는 지식을 발견하기 위한 연구의 필요성이 대두되고 있으며, 이를 위한 도구가 데이터마이닝(data mining)이다.

데이터마이닝이란 방대한 양의 데이터 속에서 쉽게 드러나지 않는 유용한 정보를 찾아내는 과정으로, 대용량(massive)의 관측 가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 바탕으로 의사결정 등을 위한 정보로 활용하고자 하는 것이다. 데이터마이닝이 적용되는 과정은 탐색(exploration)을 통해 평균, 이상치, 결측치 등을 발견하고 변형(modification)으로 자료를 변환하며, 모형화(modeling)와 모델평가(assessment)의 단계를 거치게 된다.

데이터마이닝 기법으로는 군집분석(cluster analysis), 연결 분석(link analysis), 판별 분석(discrimination analysis) 등과 같은 기존의 통계 분석과 연관성규칙(association rule), 의사결정나무(decision tree) 기법, k-평균 군집방법, 신경망모형(neural network) 등의 분석 기법이 있다. 본 논문에서 적용한 연관성 규칙은 탐색적이며, 비목적성 분석이며, 기존의 데이터를 특별한 변형 없이 계산이 용이하게 사용 가능하다는 장점을 가지고 있으며, 계산 과정이 길고, 반복된 계산이 많으며, 적절한 항목의 결정이 어렵고, 각 항목의 단위에 따른 표준화가 어렵다는 단점을 아울러 가지고 있다. 연관성 규칙은 이러한 단점에도 불구하고 두 항목간의 관계를 명확히 수치화(지지도, 신뢰도, 향상도)함으로써 두 개 이상의 항목간의 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 최소지지도를 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 partitioning 알고리즘을 제안하였으며, Toivonen(1996)은 무작위로 선정된 표본을 가지고 빈발 항목 집합들을 찾은 후에 그 결과를 데이터베이스의 나머지 부분을 가지고 증명하는 표본추출 알고리즘을 제안하였다. 또한 Cheung 등(1996)은 갱신된 데이터베이스에서 이전에 빈발 항목으로 다루어 졌던 항목 집합에 대해서 데이터베이스 스캔 과정을 생략하는 FUP(fast update) 알고리즘에 대한 연구를 하였고, Sergey 등(1997)은 데이터가 데이터베이스 전체에 골고루 퍼져있을 경우 적절한 간격을 이용한 DIC(dynamic itemset counting)의 발전된 알고리즘을 연구하였다. Saygin 등(2001)은 최소지지도와 최대 지지도의 범위를 조정하면서 고려할 항목 집합의 수를 줄이는 방법을 제안하였다.

본 논문에서는 환경 의식자료에 대하여 자료들 간의 상호 관련성을 파악하기 위해 2002년 조사된 경상남도 사회지표 조사 자료를 바탕으로 연관성 규칙을 적용하여 모형을 구축하고 구축된 모형을 분석하고자 한다. 논문의 2절에서는 연관성 규칙에 대하여 기술하고 3절에서는 경상남도 사회 지표조사 자료의 환경관련 설문에 대하여 연관성 규칙의 모형화 방안을 기술한다. 4절에서는 연관성 규칙을 이용한 자료 분석 결과를 기술한 후, 5절에서 결론을 맺는다.

2. 연관성 규칙

연관성 규칙(association rule)은 항목 집합으로 표현된 트랜잭션에서 각 항목간의 연관성을 반영하는 규칙으로서, 기본적으로 미리 결정된 최소지지도 이상의 트랜잭션 지지도를 가지는 항목 집합들의 모든 집합들인 빈발 항목 집합들을 찾아내어 연관규칙을 생성하는 단계로 이루어진다. 질문지조사에서 응답선택지가 “예”와 “아니오”의 이분형으로 구성된 문항 X 와 Y 가 있다고 할 때, 연관성 규칙은 문항 X 를 “예”로 응답하면 문항 Y 도 “예”로 응답한다는 것을 의미하는 것으로 다음과 같이 표현된다.

$$X \Rightarrow Y$$

(if X then Y : 만일 문항 X 를 “예”로 응답하면
문항 Y 도 “예”로 응답한다.)

연관규칙을 평가하는 기준에는 지지도(support), 신뢰도(confidence), 향상도(lift) 등이 있다.

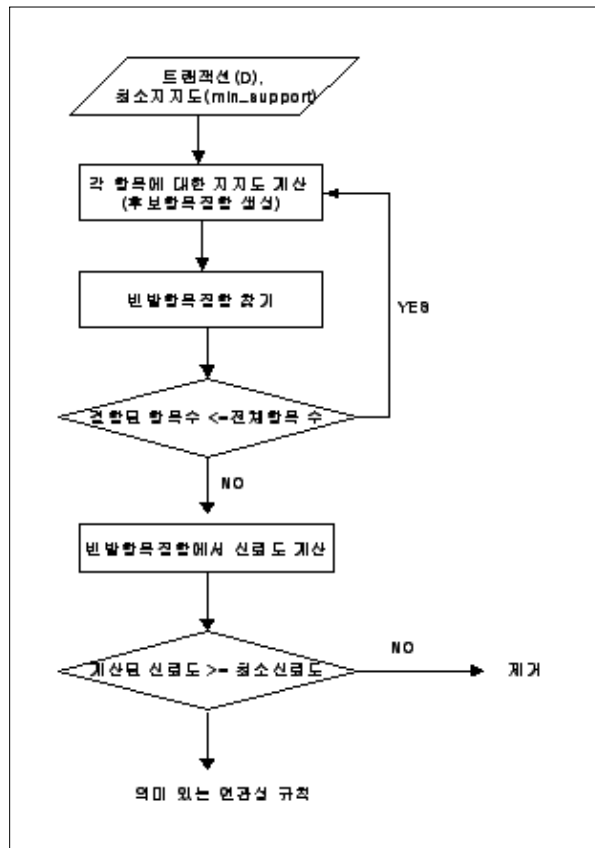
$$\text{지지도} : S_{(X \Rightarrow Y)} = P(X \cap Y)$$

$$\text{신뢰도} : C_{(X \Rightarrow Y)} = P(Y | X) = \frac{P(X \cap Y)}{P(X)}$$

$$\text{향상도} : L_{(X \Rightarrow Y)} = \frac{P(Y | X)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$$

연관성 규칙은 여러 가지 규칙을 생성할 수 있으며, 규칙에서 다루는 데이터의 유형에 따라 불리언 연관성 규칙(boolean association rule)과 정량적 연관성 규칙(quantitative association rule)으로 나누어진다. 규칙에 포함된 데이터의 차원 수에 따라 1차원 연관성 규칙(single-dimensional association rule)과 다중차원 연관성 규칙(multi-dimensional association rule)으로 나누어지며, 규칙에 포함된 요약의 수준에 따라 한 수준 연관성 규칙(single-level association rule)과 다 수준 연관성 규칙(multi-level association rule)으로 분류된다. 또한 일반적인 연관성 규칙 이외에 연관성 규칙의 응용 결과에 따라 순차 연관성 규칙(sequences association rule)과 비 연관성 규칙(dissociation rule)으로 나누어진다. 연관규칙은 Agrawal에 의해 처음 소개된 후, 데이터베이스 검색 회수를 줄이거나, 주기억 장치의 한계를 없애는 등의 발전된 알고리즘들이 발표되어 왔다. 연관규칙을 탐사하는 문제는 기본적으로 미리 결정된 최소지지도 이상의 트랜잭션 지지도를 가지는 항목집합들의 모든 집합들인 빈발항목 집합들(large itemsets)을 찾아내어 연관규칙을 생성하는 단계로 이루어진다. 연관규칙 알고리즘은 데이터베이스에서 개념일반화(concept generalization)에 상관없이 강한 규칙성을 발견하는 데 초점을 둔 Apriori, DHP, Partition, Sampling, FUP, DIC 등이 있다. 이들 중 가장 대표적인 Apriori 알고리즘은 두 단계로 구성되어 있다. 첫 번째 단계는 k -항목집합을 결정하기 위해 $k-1$ 항목 집합을 서로 결합하여 후보항목집합을

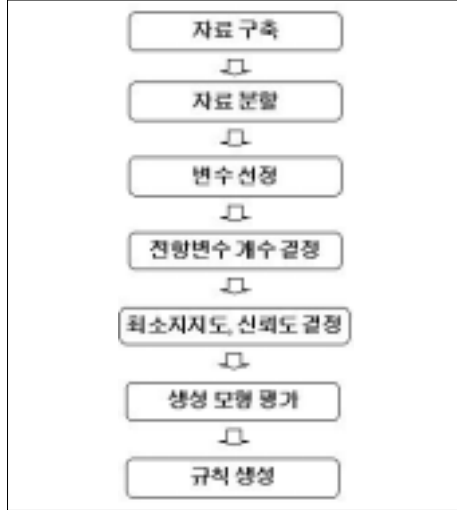
생성하는 단계이며(join step), 그 다음 단계는 생성된 후보 항목 집합을 대상으로 지지도, 신뢰도가 계산되며, 여기서 분석자가 지정한 최소지지도, 최소신뢰도보다 적은 값을 가지는 항목 집합은 가지치기(prune step)되고 전체 항목 집합에 대해 이 과정을 반복하여 최종적으로 강한 연관성을 가진 연관성 규칙을 생성한다. Apriori 알고리즘에서 사용하는 가장 중요한 규칙은 빈도수가 높은 항목의 집합은 집합의 모든 부분 집합도 빈도수가 높다는 사실이다.



<그림 1> Apriori 알고리즘

3. 연관성 규칙 모형화

본 절에서는 2002년 경상남도 사회 지표조사 자료의 환경관련 설문에 대하여 연관성 규칙의 알고리즘 중 Apriori 알고리즘의 모형화 방안에 대하여 기술하고자 한다. 연관성 규칙의 생성을 위하여 SPSS의 Clementine 10.0을 사용하였으며, 연관성 규칙 생성을 위한 모형화 과정은 <그림 2>와 같다.



<그림 2> 연관성 규칙의 모형화

1) 자료 구축

연관성규칙 생성을 위한 첫 번째 단계는 모형화에 적용할 자료를 구축하는 과정이다. 자료 구축 과정은 <표 1>과 같다. 이러한 과정을 거친 후 정제된 데이터베이스를 사용하여 연관성 규칙을 수행하게 된다.

<표 1> 자료 구축

1. 자료 수집	⊙ 2002년 경상남도 사회 지표조사 자료
2. 자료 선정	⊙ 환경관련문항, 인구통계학 속성 문항, 집단구분 문항
3. 자료 정제	⊙ 무응답 등의 결측치 제거
4. 자료 변환	⊙ 모든 문항에 대하여 이분형으로 변환

2) 변수선정

연관성 규칙을 생성하기 위하여 전항변수와 후항변수를 선정한다. 후항변수에는 환경관련 문항을 선정하고 전항변수에는 분석 문항을 선정한다.

3) 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.

4) 전항변수 개수 결정

환경관련 문항에 대하여 분석 문항간의 관련성을 알아보기 위하여 전항변수의 개수

를 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 수를 2로 설정하였다.

5) 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 60으로 결정했을 때 의미 있는 규칙을 발견할 수 있었다.

6) 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.

7) 규칙생성

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

4. 자료 분석

2002년 경남사회지표조사에서 환경관련 문항과 인구통계학 속성 관련 문항, 집단구분 문항을 추출하여 DB를 구축하였으며, 이 문항들을 환경관련 문항 부문, 분석 문항 부문으로 분류하였다. 설문 문항은 다음과 같다.

<표 2> 경남사회지표조사 환경관련 문항

순번	문항	문항 보기
1	f24. 가장 쾌적한 환경	1.풍부한 녹색 공간 2.맑고 깨끗한 물 3.넓은 공지 4.맑은 공기 5. 기타
2	f25. 수돗물의 음용수 적정여부	1.적당하다 2.적당하지 않다 3.상수도 시설이 없어 모르겠다 4.상수도 시설이 없지만, 적당하다고 생각한다 5.상수도 시설이 없으며, 적당하지 않다고 생각한다.
3	f26. 수돗물의 음용수 대책	1.상수도 보호구역 확대 지정 2.상수도 시설의 대폭적 개선 3.상수도 환경감시원 제도의 확대 4.기타
4	f27. 환경오염의 주체	1.기업체 2.일반소비자 및 관광행락객 3.농어민 4.모두 5.모르겠음
5	f28. 쓰레기 분리수거 참여정도	1.잘 참여하고 있다 2.호응하지만 참여정도는 낮다 3.우리 지역은 쓰레기 분리수거제를 실시하지 않고 있다
6	f29. 녹색제품 구입 여부	1.항상 사려고 노력한다 2.가능하면 사려고 노력한다 3.그런 제품을 사 본 적은 있지만, 굳이 사려고 하지 않는다 4.있는 것은 알지만 사본 적이 없다 5.그런 제품이 있는 것도 모른다

<표 3> 경남사회지표조사 분석 문항

순번	문항	문항 보기
1	g30. 지역사회 전반적 평가	1.아주 살기 좋은 곳이다 2.비교적 살기 좋은 곳이다 3.그저 그런 곳이다 4.비교적 살기 나쁜 곳이다 5.아주 살기 나쁜 곳이다
2	g36. 주관적 사회계층	1.상류층 2.중상류층 3.중류층 4.중하류층 5.하류층
3	j1. 연령	() 세
4	j2. 성별	1. 남 2. 여
5	j3. 학력	1. 무학 2. 초졸 3. 중졸 4. 고졸 5. 전문대학재학 6. 전문대졸 7. 대학재학 8. 대졸 9. 대학원이상
6	j6. 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, 기계 조작 및 조립 종사자 10. 단순노무 종사자 11. 군인 12. 가정주부 13. 학생 14. 무직 15. 기타
7	k1. 조사지역	1. 농촌지역 2. 어촌지역 3. 상가지역 4. 주거지역 5.공업지역 6. 기타지역
8	do. 시 군계	1. 시 2. 군

6개의 환경 문항에 대하여 연관성 규칙을 실시한 결과 가장 쾌적한 환경, 수돗물의 음용수 걱정여부, 환경오염의 주체, 녹색제품의 구입 여부의 4개의 항목에 대하여 연관성 규칙이 생성되었다. 연관성 규칙의 생성결과 각 규칙의 신뢰도가 65%~70% 정도로 규칙의 신뢰도가 높다고는 할 수 없다. 세부적인 내용은 다음과 같다.

<표 4> 가장 쾌적한 환경에 대한 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	57.2	0.659	주관적 사회계층 = 중하류층 이하	
2	44.9	0.662	주관적 사회계층 = 중하류층 이하	성별 = 남성
3	45.3	0.68	주관적 사회계층 = 중하류층 이하	조사지역 = 농어촌지역
4	27.4	0.671	주관적 사회계층 = 중하류층 이하	직업 = 가정주부
5	18.7	0.653	주거지 = 시	주관적 사회계층 = 중하류층 이하
6	14	0.668	학력 = 고졸이하	주관적 사회계층 = 중하류층 이하

가장 쾌적한 환경에 대한 연관성 규칙 결과 총 6개의 규칙이 생성되었으며, 주관적 사회계층의 문항에 대하여 중하류층 이하인 집단이 가장 쾌적한 환경의 문항에 대하여 맑고 깨끗한 물과 공기의 응답률이 높은 것으로 나타났다. 또한 이 집단 중에서도 남성인 집단, 조사지역이 농어촌 지역, 가정주부, 주거지가 시인 경우, 학력이 고졸이하인 집단에서 이러한 연관성이 두드러지는 것으로 나타났다.

<표 5> 수돗물의 음용수 적정 여부에 관한 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	50.7	0.656	연령 = 평균이하	
2	23.7	0.653	연령 = 평균이하	성별 = 남성
3	21.1	0.684	연령 = 평균이하	지역사회 전반적인 평가 = 나쁨
4	12.9	0.66	연령 = 평균이하	학력 = 고졸이하
5	40.9	0.662	지역사회 전반적인 평가 = 나쁨	
6	44.8	0.656	거주지 = 시	연령 = 평균이하
7	33.5	0.671	거주지 = 시	지역사회 전반적인 평가 = 나쁨
8	19.5	0.656	성별 = 남성	지역사회 전반적인 평가 = 나쁨

수돗물 음용수 적정 여부에 관한 연관성 규칙 결과 총 8개의 규칙이 생성되었으며, 다. 연령이 평균이하이거나 지역사회에 대한 전반적인 평가가 나쁘다고 생각하는 집단에서 수돗물 음용수 적정 여부에 대해 적당하지 않다고 생각하는 것으로 나타났다. 특히 연령이 평균 이하인 집단에서 성별이 남성, 지역사회 전반적인 평가가 나쁨, 고졸이하 집단, 거주지가 시인 집단에서 수돗물이 음용수로 적절하게 생각하지 않은 것으로 나타났다. 지역사회의 전반적인 평가문항에 대하여 나쁨이라고 응답한 집단에서는 남성과, 거주지가 시인 집단이 수돗물이 음용수로 적절하게 생각하지 않은 것으로 나타났다.

이외에도 <표 6> 및 <표 7>에서 보는 바와 같이 환경오염의 주체에 대해 12개의 규칙이 생성되었으며, 녹색제품 구입 여부에 대해서도 7개의 규칙이 생성되었다.

<표 6> 환경오염의 주체의 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	80.3	67.3	거주지 = 시	
2	16.2	69.4	거주지 = 시	조사지역 = 농어촌지역
3	38.1	68.2	거주지 = 시	성별 = 남
4	44.8	67.9	거주지 = 시	연령 = 평균이하
5	55.2	67.2	거주지 = 시	학력 = 고졸이하
6	50.7	67.8	연령 = 평균이하	
7	10.5	71.6	연령 = 평균이하	조사지역 = 농어촌지역
8	21.1	67.6	연령 = 평균이하	지역사회 전반적인 평가 = 나쁨
9	30	67.4	조사지역 = 농어촌지역	
10	47.2	67.8	성별 = 남	
11	13.9	68.8	성별 = 남	조사지역 = 농어촌지역
12	19.5	68.3	성별 = 남	지역사회 전반적인 평가 = 나쁨

<표 7> 녹색제품 구입 여부에 대한 연관성 규칙

규칙	지지도	신뢰도	전항값1	전항값2
1	57.2	0.654	주관적 사회계층 = 중하류층	
2	24.8	0.7	주관적 사회계층 = 중하류층	지역사회 전반적인 평가 = 나쁨
3	18.7	0.664	주관적 사회계층 = 중하류층	조사지역 = 농어촌지역
4	44.9	0.658	거주지 = 시	주관적 사회계층 = 중하류층
5	45.3	0.665	학력 = 고졸이하	주관적 사회계층 = 중하류층
6	29.1	0.655	학력 = 고졸이하	지역사회 전반적인 평가 = 나쁨
7	12.6	0.671	지역사회 전반적인 평가 = 나쁨	조사지역 = 농어촌지역

5. 결론

본 논문에서는 환경 의식자료에 대하여 데이터마이닝의 기법 중 연관성 규칙의 모형화 방안에 대하여 연구하고 2002년 조사된 경상남도 사회지표 조사의 자료에 대하여 연관성 규칙 기법을 적용하여 모형을 구축하고 구축된 모형을 분석하였다. 연관성 규칙의 모형 생성 및 분석 결과를 통하여 환경관련 문항과 분석 문항에 대한 관련성 여부를 파악 할 수 있었으며, 쉽게 드러나지 않는 유용한 정보를 추출할 수 있었다. 향후 이 정보를 바탕으로 지역 여건 등의 데이터와 연계하여 환경 의식 자료에 대하여 관련성 여부를 다양하게 파악 할 수 있어 분석 결과를 환경개선대책 수립과 환경 정책 결정에 필요한 의사결정 지원에 반영할 수 있을 것으로 사료된다.

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, 한국지하수토양환경학회 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, 한국지하수토양환경학회 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(2002a), 전국 폐기물 발생 현황.
9. 환경부(2002b), 상수도 통계.
10. 환경부(2002c), 하수도 통계.
11. 환경부(2002d), 오수·분뇨 및 축산폐수처리 통계.
12. 환경부(2003a), 환경통계연감.
13. 환경부(2003b), 대기환경연보.
14. 환경부(2004), 환경백서.
15. Agrawal R., Imielinski R., Swami A.(1993), Mining association rules between sets of items in large databases, *Proceeding of the ACM SIGMOD Conference on Management of Data*, Washington, D.C., 207-216.
16. Agrawal R., Srikant R.(1994), Fast algorithms for mining association rules", *Proceeding of the 20th VLDB Conference*, Santiago, Chile. 487-499.
17. Cheung D.W., Han J., Ng V., Fu A.W., Fu Y.(1996), A fast distribution algorithm for mining association rules, *Int's Conf. on Parallel and Distributes Information System*, Miami Beach, Florida, 31-44.
18. Park J.S., Chen M.S., and Philip S.Y.(1995), An effective hash-based algorithms for mining association rules, *Proceeding of ACM SIGMOD Conference on Management of Data*, 175-186.
19. Saygin Y., Vassilios S.V., Clifton C.(2001), Using unknowns to prevent discovery of association rules, *SIGMOD Record*, 30(4), 45-54.
20. Sergey B., Rajeev M., Jeffrey D.U., Shalom T.(1997), Dynamic itemset counting and implication rules for market data, *Proceedings of ACM SIGMOD Conference on Management of Data*. 255-264.
21. Toivonen H.(1996), Sampling large database for association rules, *Proceedings of the 22nd VLDB Conference*, Mumbai(Bombay), India, 134-145.

[2005년 5월 접수, 2005년 6월 채택]