GML 문서에서 연관규칙 생성 시스템 구현

Implementation of Association Rules Creation System from GML Documents

김의찬* / Eui-Chan Kim
황병연** / Byung-Yeon Hwang

요 약

지리 정보에 대한 관심이 증가되면서 이러한 연구와 활용 분야도 다양해지고 있다. OGC(Open GIS Consortium)에서는 XML(Extensible Markup Language)을 GIS 분야에 도입한 GML(Geography Markup Language)을 개발하였으며 여러 활용 분야에서 GML을 사용하고 계속적으로 연구되고 있다. 본 논문에서는 기존의 XML 문서를 기반으로 연구되었던 데이터 마이닝 방법 중 하나인 연관규칙 기법(Apriori)을 GML 문서들에 사용하여 의미 있는 규칙을 찾아내려 한다.

규칙을 찾는 방법에는 2가지가 있을 수 있다. 하나는 GML 문서에서 내용만을 뽑아내어 그에 따른 규칙을 찾아내는 방법이고, 다른 하나는 GML 문서에서 사용된 태그와 속성을 기반으로 규칙을 찾아내는 방법이다. 본 논문에서는 2가지 방법을 통해 규칙을 찾는 것에 대하여 기술하고 2가지 방법을 적용한 시스템을 보일 것이다.

Abstract

As the increasing interest about geographical information, such researches and applied fields become wide. OGC(Open GIS Consortium) developed GML(Geography Markup Language) which is adopted XML(eXtensible Markup Language) in GIS field. In various applied field, GML is used and studied continuously. This paper try to find out the meaningful rules using Apriori algorithm from GML documents, one of the data mining techniques which is studied based on existing XML documents.

There are two ways to find out the rules. One is the way that find out the related rules as extracting the content in GML documents, the other find out the related rules based on used tags and attributes. This paper describes searching the rules through two ways and shows the system adopted two ways.

주요어: GML, 연관규칙, XML, Apriori, 데이터 마이닝
Keywords: GML, Association Rules, XML, Apriori, Data Mining

■ 본 연구는 2006년도 가톨릭대학교 교비연구비의 지원으로 이루어졌음
■ 논문접수 : 2005. 11. 7 ■ 심사완료 : 2006. 4. 3

* 가톨릭대학교 컴퓨터공학과 박사과정(eckim@catholic.ac.kr)
** 교신저자 가톨릭대학교 컴퓨터정보공학부 교수(byhwang@catholic.ac.kr)

1. 서 론

XML(eXtensible Markup Language)[1]은 W3Consortium(W3C)에서 웹 기반의 구조화된 문서를 기술하는 방법에 대하여 표준화한 언어이다. 기존의 HTML(HyperText Markup Language)처럼 정해져 있는 태그들을 사용하는 것이 아니라 사용자 자신이 태그나 속성을 지정해서 사용하는 언어이다. XML은 확장성, 유연성이라는 특징 외에도 다양한 장점과 여러 가지 기능들을 제공한다. 이러한 점들로 인하여 현재 XML에 대한 연구는 계속적으로 활발하게 이루어지고 있으며 많은 응용분야에서 사용되고 있다.

이렇듯 여러 장점을 가지고 다양한 분야에서 사용되며 응용되고 있는 XML을 OGC(OpenGIS Consortium)에서 GIS 분야에 도입시키려 GML(Geography Markup Language) 사양[2]을 제시하였다.

본 논문에서는 이러한 GML 문서를 통해 데이터 마이닝 기법들 중 하나인 연관규칙 기법을 적용하여 의미 있는 규칙들을 찾아내려 한다. GML 문서로부터 정보를 얻기 위하여 우리는 문서검색을 하게 되고, 그로부터 얻어내는 정보들은 기본적이고 단순한 정보들이 된다. 그러나 단순한 질의 검색을 통해서는 얻어낼 수 없는 정보들이 있는데 이러한 정보들을 함축적이고 암시적인 정보 또는 의미 있는 정보라 한다. 이와 같은 정보를 찾아내는 기법이 데이터 마이닝 기법이다. 데이터 마이닝 기법에는 여러 가지가 있다. 연관규칙(Association Rules), 분류(Classification), 일반화(Generalization), 클러스터링(Clustering) 등 다양하다[3]. 본 논문에서는 여러 데이터 마이닝 기법들 중 연관규칙 기법을 이용하여 GML 문서들로부터 의미 있는 정보를 추출하려고 한다.

연관규칙 기법은 현재 여러 분야에서 응용되며 계속적으로 많이 연구되고 있는 데이터 마이닝 기법이다. 데이터베이스에 저장되어 있는 기본적인 데이터들을 바탕으로 기본 질의문을 통해서는 얻을 수 없는 규칙을 찾아내는 기법이다. 연관규칙 기법을 사용할 때 주로 사용되는 알고리즘으로는 Apriori[4] 알고리즘이 있다. 이 알고리즘은 매우 많이 사용되고 있으며 계속적으로 연구되고 있는 알고리즘으로 본 연구에서도 이 알고리즘을 사용할 것이다.

연관규칙을 사용할 때 일반적인 데이터의 경우 데이터 자체의 값을 사용하여 규칙을 찾아내게 되지만 GML 같은 데이터에서는 태그 사이의 내용뿐만 아니라 태그나 속성도 중요한 의미를 지닐 수 있기 때문에 본 논문에서는 태그 사이의 내용과 더불어 태그와 속성에 관련한 연관규칙도 찾아낼 것이다.

2장에서는 관련연구에 대하여 논의하고, 3장에서는 GML 문서들에 연관규칙을 적용하는 방법에 대하여 기술한다. 4장에서는 3장에서 논의한 내용을 바탕으로 한 실험 예에 대하여 언급하고 5장에서는 시스템 구현 모습에 대하여 기술한다. 마지막으로 6장에서는 본 연구에 대하여 결론을 정리한다.

2. 관련연구

연관규칙은 데이터베이스 내의 단위 트랜잭션에서 빈번하게 발생하는 사건의 유형을 발견하는 것이다. 예를 들어, "전체 고객 중에 빵과 버터, 그리고 우유를 구매한 고객이 10% 이상이고, '빵과 버터'를 구매한 고객의 50%가 우유도 함께 구매한다." 이것이 하나의 발견된 사건의 유형, 즉 하나의 규칙이 된다. 여기서 10%는 연관규칙의 지지도(support)가 되고, 50%는 신뢰도(confidence)가 된다.

연관규칙을 찾는 전체 과정을 간단하게 살펴보면, 전체 데이터베이스에서 먼저 후보 아이템 항목 집합을 찾고, 이 후보 아이템 항목 집합에서 미리 제시된 최소 지지도 값

을 넘는 빈발 항목 집합을 찾아낸다. 빈발 항목 집합을 찾을 때 전체 데이터베이스의 트랜잭션을 반복적으로 검색하면서 조인연산을 계속해서 사용하게 된다. 최종적으로 나오는 빈발항목 아이템 집합에서 최소 신뢰도 값을 넘는 연관규칙을 찾아내게 되는 것이다. 여기서 지지도(S)란, 전체 사건 또는 거래 중에서 어떤 아이템 X와 아이템 Y를 동시에 포함하는 사건 또는 거래가 어느 정도 되는가 하는 것이다. 이것을 식으로 표현하면 다음과 같다.

$$S = \frac{|X \cap Y|}{N}$$

(N은 전체 트랜잭션의 개수)

그리고 신뢰도(C)는 어떤 아이템 X를 포함하는 사건이나 거래 중에서 Y가 포함된 사건이나 거래가 어느 정도인가 하는 것이다. 이것을 식으로 표현하면 다음과 같다.

$$C = \frac{|X \cap Y|}{|X|}$$

지지도를 통해 나온 빈발항목들에서 신뢰도를 통해 최종 연관규칙을 얻어내는 것이다. 대표적인 연관규칙 알고리즘으로는 앞에도 언급한 Apriori 알고리즘이 있다.

기존의 XML 데이터를 가지고 연관규칙을 적용한 연구들이 있다. [5]에서는 XML문서에서 빈발경로를 찾는 연구를 하였고, [6, 7]에서는 연관규칙을 찾기 위한 확장된 XQuery를 제안하였다. [8]에서는 FP (Frequent Pattern)-Growth 알고리즘으로부터 XSD-AR(XML Structural Delta Association Rule)이라 부르는 연관규칙 타입을 제안하였다. 또한 지리정보 데이터베이스로부터 공간(Spatial) 연관규칙을 찾는 연구[9]도 있었는데 본 연구에서는 여러 응용분야에서 사용가능하도록 지리정보를 담고 있는 GML 데이터로부터 연관규칙을 찾아내는 방법을 제안하려 한다.

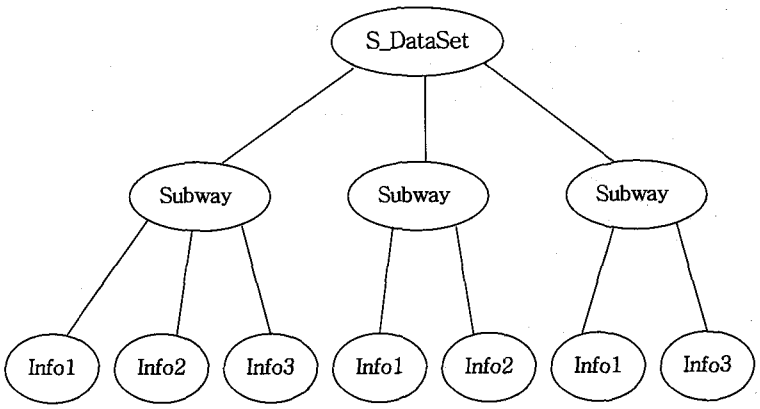
3. GML에서의 연관규칙 추출

본 논문에서 제안하는 방법은 일반 XML 문서가 아닌 GML 문서를 통해서 의미있는 규칙을 찾는 것이다. 이는 많은 GML 문서를 다루는데 있어서 문서들 간의 연관성을 찾거나 문서 내에 있는 지리정보와 관련하여 기본 질의를 통해서는 찾을 수 없는 규칙이나 연관성을 추출하는 것이다. GML 문서의 내용으로부터 연관성 및 규칙을 추출할 수도 있고, GML 문서에서 사용한 태그 및 속성들을 통해서 추출할 수 있다.

3.1 내용으로부터의 추출

GML 문서의 내용으로부터 의미 있는 규칙을 추출하기 위해서는 태그나 속성이 아닌 내용만을 먼저 추출하고 연관규칙 기법을 적용해야 한다. 하나의 문서로 이루어져 있지 않고 여러 문서로 데이터들이 나누어져 있는 경우 내용만을 추출하기 위해서는 GML 문서의 스키마가 동일하여야 한다. 스키마가 동일한 많은 GML 문서로부터 각각의 동일한 태그 내에 있는 내용만을 추출하여 트랜잭션을 구성하여야 연관규칙 기법을 적용할 수 있다.

<그림 1>과 같은 구조를 갖는 문서가 있다고 가정하자. 이것은 지하철 주변의 환경정보를 GML 문서로 작성하였을 때 나타날 수 있는 구조이다. 각 타원들은 태그들을 나타내고 있으며 각각의 Info 내에 지하철 주변 환경 정보 내용이 들어있는 형태이다. 내용으로는 '서점', '분식점', '패스트푸드점', '커피숍', '옷가게' 등 지하철 주변에 어떠한 상점들이 분포해있는지 정보



<그림 1> GML 데이터 구조 예

를 넣을 수 있다. 각 Subway들은 데이터베이스 내에서 트랜잭션이 될 수 있으며 각 트랜잭션의 아이템은 Info 태그 내에 내용을 아이템으로 간주하여 테이블화 할 수 있다. 테이블의 예는 <표 1>과 같다.

여기서 ST는 각 Subway들을 트랜잭션으로 본 것이고 각 Info에 있는 내용들이 환경정보에 들어가게 된다. 테이블의 내용을 이용하여 연관규칙을 찾을 때 좀 더 빠른 수행을 하기 위해서는 [10]에서 사용했던 비트 방식을 이용하여 수행할 수 있다.

<표 1> GML 내용을 테이블화 한 모습

TID	환경 정보(업종)
ST1	분식점, 호프집, 빵집
ST2	분식점, 커피숍, 패스트푸드
ST3	분식점, 호프집, 커피숍, 빵집, 패스트푸드, 서점
ST4	분식점, 패스트푸드, 커피숍, 빵집
ST5	서점, 옷가게, 신발가게, 쥬얼리, 패스트푸드
ST6	커피숍, 패스트푸드, 서점, 옷가게, 쥬얼리

<표 1>에 예시된 테이블을 가지고 연관규칙을 적용하여 구할 때는 Apriori 기법을 사용하여 구하면 된다. 실험 예는 4장에서 살펴보도록 한다.

3.2 태그와 속성으로부터의 추출

3.1절에서는 GML 문서에서 내용만을 이용하여 규칙을 찾아내는 방법을 기술하였다. 이번 절에서는 내용이 아닌 태그와 속성으로부터 빈발 경로를 찾는 방법을 살펴보도록 하겠다.

<그림 1>에 나타난 구조를 예를 들어 보면 다음과 같다. <그림 2>에서는 <그림 1>에 나타난 구조를 정리한 모습이다.

/S_DataSet
/S_DataSet/Subway
/S_DataSet/Subway
/S_DataSet/Subway
/S_DataSet/Subway/Info1
/S_DataSet/Subway/Info2
/S_DataSet/Subway/Info3
/S_DataSet/Subway/Info1
/S_DataSet/Subway/Info2
/S_DataSet/Subway/Info1
/S_DataSet/Subway/Info3

<그림 2> 경로 정리 모습

<그림 2>에 나타난 정보를 이용하여 먼저

빈발 경로를 찾는다. 빈발 경로는 <그림 2>에 나타나 있는 모든 경로를 검색하여 주어진 지지도 임계값보다 큰 경로들만을 찾은 경로이다. 예를 들어 임계값을 3으로 가정하면 <그림 2>에 나타나 있는 경로들 중에 임계값 이상에 해당하는 경로는 '/S_DataSet' 과 '/S_DataSet/Subway', 그리고 '/S_DataSet/Subway/Info1' 이 된다.

빈발 경로를 찾을 때는 서브트리의 개수를 세어서 결정한다. 주의해야 할 점은 서브트리의 개수가 임계값보다 적다고 하더라도 서브트리내의 서브트리까지 계산을 해야 된다는 점이다.

4. 실험 예

4.1 내용을 이용한 예

먼저, GML 문서의 내용을 추출하여 테이블화 하여야 한다. 이번 절에서는 <표 1>의 내용을 이용하여 설명하도록 한다. 최소지지도 임계값은 50%로 가정한다. 전체 Subway 트랜잭션을 검색하여 최소 지지도를 만족하는 업종을 뽑아낸다. 그러면 <표 2>와 같은 빈발 업종을 찾아낼 수 있다. 이 표를 이용하여 후보 업종과 빈발 업종을 반복해서 찾아나가게 된다.

<표 2> 빈발 업종

업 종	지지도
분 식 점	66%
커 피 숍	66%
빵 집	50%
패 스 트 푸 드	83%
서 점	50%

<그림 3>에서는 최소 지지도 임계값에 맞는 빈발 업종을 찾아가는 과정을 보여주고 있다. 최소 신뢰도 임계값을 100%로 가정한다면 다음과 같은 결과가 나오게 된다.

(분식점, 커피숍) → 패스트푸드 (100%)
(분식점, 패스트푸드) → 커피숍(100%)

L2

업 종	지지도
{분식점, 커피숍}	50%
{분식점, 빵집}	50%
{분식점, 패스트푸드}	50%
{커피숍, 패스트푸드}	66%

L3

업 종	지지도
{분식점, 커피숍, 패스트푸드}	50%

<그림 3> 빈발 업종 집합

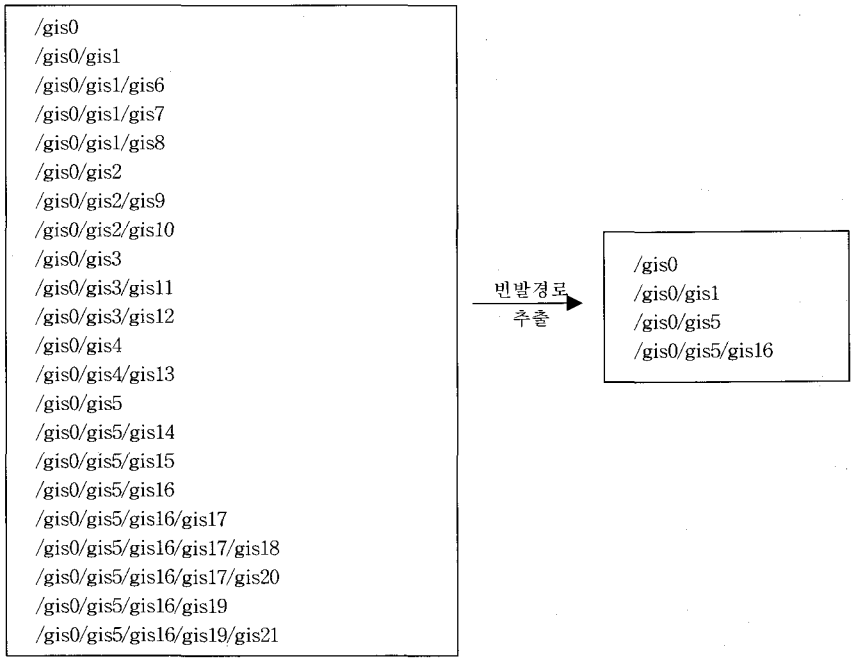
결과를 바탕으로 분식점과 커피숍이 있는 지하철 근처에는 패스트푸드가 있다는 규칙과 분식점과 패스트푸드가 있는 지하철 근처에는 커피숍이 있다는 규칙이 생성된다.

4.2 태그와 속성을 이용한 예

<그림 4>와 같은 경로집합이 있다고 가정하자. 임계값은 3으로 가정하고 임계값 이상의 빈발 경로를 찾으면 다음과 같다.

/gis0/gis1 경로의 서브 트리 개수가 3개이므로 /gis0/gis1이 빈발 경로가 되었고, /gis0/gis5의 서브 트리 개수 또한 3개이므로 빈발 경로가 되었다. 여기서 주의 할 부분은 /gis0/gis5/gis16 경로인데 이 경로에서 gis16 태그의 서브트리는 2개이지만, gis16/gis17/gis18과 gis16/gis17/gis20 그리고 gis16/gis19/gis21 이라는 3개의 서브 트리 경

로를 가지고 있기 때문에 /gis0/gis5/gis16은 임계값 3과 크거나 같게 된다. 따라서 빈발 경로가 되었다. 알고리즘은 <그림 5>와 같다.

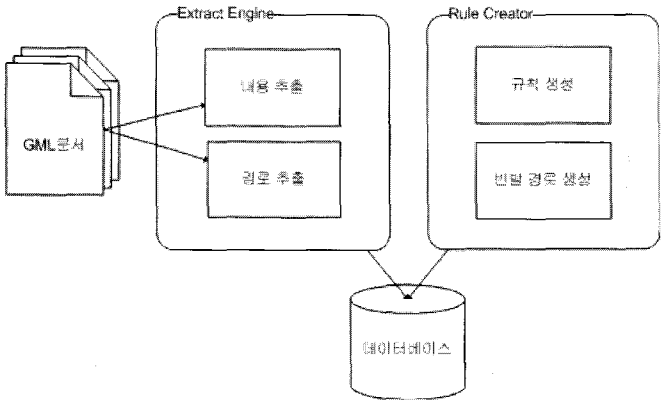


<그림 4> 경로 집합 및 빈발경로 추출

```
1: Input: all_pathList, threshold_value
2: Output: frequent_pathList
3: FindFrequentPathList(pathList, threshold){
4:       count = 0;
5:       for(i = 1; pathList ≠ null; i++){
6:             temp_pathList = pathList_i;
7:          while(all_pathList)
8:          {
9:                if(temp_pathList == each_pathList)
10:                   count++;
11:         }
12:         if(count > threshold_value)
13:               frequent_pathList ∋ temp_pathList;
14:         count = 0;
15:   }
16: }
```

<그림 5> FindFrequentPathList 알고리즘

입력 값은 전체 경로 리스트와 임계값이고, 출력 값은 빈발 경로 집합이다. count 변수는 동일한 경로의 개수를 세기 위한 변수이고 5번째 라인에서 for 문을 이용하여 개수를 구하게 되는데, 6번째 라인에서 pathList_i는 각각의 경로 리스트를 의미하며 이러한 경로들을 temp_pathList에 저장하여 둔다. 7번째 라인의 while문을 이용하여 temp_pathList에 저장되어 있는 경로가 전체 경로 리스트에 몇 번 있는지 확인하게 되고 count 변수를 증가시켜서 개수를 센다. 12번째 라인에서 입력된 임계값과 비교하여 임계값보다 큰 경우에만 frequent_pathList에 temp_pathList에 있는 경로를 포함시키게 된다. 마지막으로 count 변수는 0으로 초기화시키고 다음 경로를 temp_pathList에 넣고 같은 작업을 반복하게 된다.



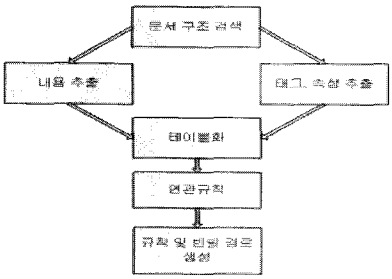
<그림 6> 규칙 생성 시스템 구조도

5. 시스템 구현

5.1 시스템 구조도

GML 문서들로부터 규칙과 빈발경로를 찾아내는 시스템의 구조도는 <그림 6>과 같다. GML 문서를 입력받아 추출 엔진(Extract Engine)을 통해서 내용 및 경로들을 추출하여 데이터베이스에 저장한다. 데이터베이스에 저장되어 있는 내용들과 경로들을 바탕으로 규칙과 빈발경로를 규칙 생성기(Rule Creator)를 통해서 얻어낸다.

규칙 생성 흐름도는 <그림 7>과 같다. GML 문서를 입력받아 문서 구조를 검색한 다음 내용 추출을 하여 연관규칙 기법을 적용하여 규칙을 생성하는 과정과 태그, 속성들을 추출하여 빈발 경로를 찾아내는 과정을 나타내었다.



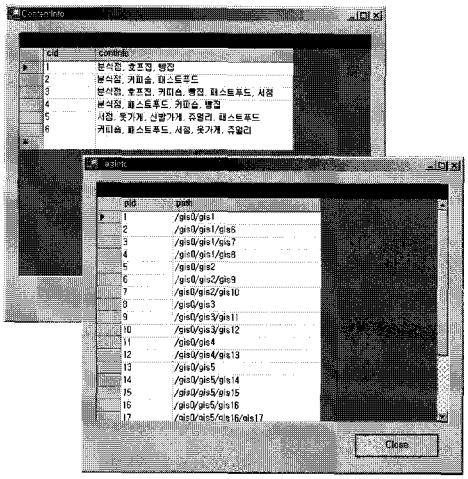
<그림 7> 규칙 생성 흐름도

5.2 시스템 구현 모습

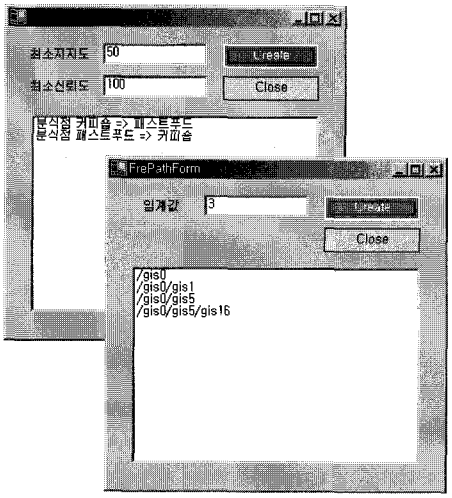
시스템은 두 부분으로 나누어진다. 5.1절에서 설명한 <그림 6>과 같이 추출 엔진 부분과 규칙 생성 부분으로 나누어진다. 추출 엔진 부분은 Java를 이용하여 XRel[11]을 구현하였고 규칙 생성 부분은 C#을 이용하여 구현하였다. 데이터베이스 관리시스템은 MS-SQL을 사용하였다. 추출 엔진을 통해서 내용과 경로를 MS-SQL에 저장하고 규칙 생성기를 통해서 내용에 대한 규칙과 빈발 경로를 추출하도록 하였다. 메인 폼은 간단한 형태로 만들었고 각 메뉴에는 DB와 연결하는 폼, 데이터베이스에 저장되어 있는 내용과 경로를 출력하는 폼, 최소지지도, 최소신뢰도 및 임계값을 입력받아 데이터베이스로부터 규칙과 빈발경로를 생성하는 폼을 열어볼 수 있도록 하였다.

<그림 8>은 추출 엔진을 통해 MS-SQL에 저장된 내용과 경로를 보는 폼 화면이고, <그림 9>는 규칙과 빈발경로를 찾은 화면이다.

<그림 8>에서 보듯이 XRel을 이용하여 내용 및 경로를 추출하였다. 추출한 내용 및 경로들은 MS-SQL에 테이블 형태로 저장하였다. 이것을 이용하여 Apriori 기법을 적용하여 내용에 관한 규칙을 생성하게 된다.



<그림 8> DB 저장 내용 출력 화면



<그림 9> 규칙 및 빈발경로 생성 폼

<그림 9>에서 최소 지지도와 최소 신뢰도를 입력받아 내용에 관한 연관규칙을 찾도록 하였고, 임계값을 입력받아 빈발 경로를 추출할 수 있도록 하였다. 빈발 경로를 추출할 때는 <그림 5>에서 보였던 알고리즘을 이용하여 추출하였다.

본 연구에서 GML 문서를 이용하여 연관규칙을 추출하는 경우 내용에 대한 것과 태그 및 속성에 관한 것으로 나누어 실험하였으며 그 결과의 모습이 <그림 9>에 나타나 있다.

6. 결 론

본 연구에서는 GML 문서에서 연관규칙을 추출하는 방법을 기술하고 시스템을 구현하였다. GML 문서에서 연관규칙을 추출할 때 2가지 방법으로 나누어 볼 수 있는데 하나는 GML 문서 내용을 이용하여 규칙을 찾아낼 수도 있고, 태그와 속성을 이용하여 규칙을 찾아낼 수도 있다.

내용을 이용하여 규칙을 찾는 경우에는 내용들만 추출하여 테이블화 한 다음 각 트랜잭션에 대한 내용 정보들을 바탕으로 기존의 연관규칙 기법을 적용하여 규칙을 찾아내면 된다. 태그와 속성을 이용하는 방법의 경우에는 전체 경로들을 바탕으로 임계값 이상인 빈발 경로들을 찾으면 된다.

내용과 태그, 속성들은 추출 엔진을 통해 데이터베이스에 저장하도록 하였고, 데이터베이스에 저장되어 있는 데이터와 사용자의 임계치 입력값을 통해서 규칙과 빈발 경로를 찾게 된다.

추후 연구 방향으로는 내용들 중에서 각 탐색지점과의 인접 정도를 구분하여 좀 더 자세한 규칙을 찾는 연구와 좀 더 다양한 GML 데이터를 통한 실험으로 다양하고 세밀한 정보 추출이 가능한 시스템 보완작업이 필요하다.

참고문헌

1. W3Consortium, Extensible Markup Language(XML) 1.0, 1998.
2. Open GIS Consortium, Inc., Geography Markup Language Specification (GML) v3.1.1, 2004.

3. M.S. Chen, J. Han, and P.S. Yu, "Data Mining: An Overview from a Database Perspective," IEEE Trans. on Knowledge and Data Engineering, Vol. 8, No. 6, Dec. 1996, pp. 866-883.
4. R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules in Large Databases," Proc. of the 20th International Conf. on Very Large Databases, Santiago de Chile, 1994, pp. 487-499.
5. A. Meisels, M. Orlov and T. Maor, "Discovering Associations in XML Data," Proc. of the 3rd International Conf. on Web Information Systems Engineering, Singapore. Dec. 2002, pp. 178-183.
6. D. Braga, A. Campi, S. Ceri, M. Klemettinen, and P.L. Lanzi, "A Tool for Extracting XML Association Rules," Proc. of the 14th IEEE International Conf. on Tools with Artificial Intelligence, Washington DC, USA, Nov. 2002, pp. 57-64.
7. D. Braga, A. Campi, S. Ceri, M. Klemettinen, and P.L. Lanzi, "Mining Association Rules from XML Data," Proc. of the 4th International Conf. on Data Warehousing and Knowledge Discovery, Aix-en-Provence, France, Sep. 2002, pp. 21-30.
8. L. Chen, S.S. Bhowmick, and L.T. Chia, "Mining Association Rules from Structural Deltas of Historical XML Documents," Proc. of the 8th Pacific-Asia Conf. Sydney, Australia, May. 2004, pp. 452-457.
9. K. Koperski and J. Han, "Discovery of Spatial Association Rules in Geographic Information Databases," Proc. 4th Int'l Symp. on Large Spatial Databases, Maine, Aug. 1995, pp. 47-66.
10. 김의찬, 황병연, "트랜잭션 클러스터링을 이용한 연관규칙 생성," 제 23회 한국정보처리학회 춘계학술대회논문집, 제12권 제1호, 2005, pp. 15-18.
11. M. Yoshikawa, T. Amagasa, T. Shimura, and S. Uemura, "XRel: a path-based approach to storage and retrieval of XML documents using relational databases," ACM Trans. on Internet Technologies, Vol. 1, Issue. 1, 2001, pp. 110-141.

김의찬
1999년 가톨릭대학교 전산학과(이학사)
2001년 가톨릭대학교 대학원 컴퓨터공학과(공학석사)
2006년 가톨릭대학교 대학원 컴퓨터공학과(공학박사 예정)
관심분야 : 데이터 마이닝, 공간 데이터베이스, 전자상거래, XML 데이터베이스

황병연
1986년 서울대학교 컴퓨터공학과(공학사)
1989년 한국과학기술원 전산학과(공학석사)
1994년 한국과학기술원 전산학과(공학박사)
1994년 ~ 현재 가톨릭대학교 컴퓨터정보공학부 교수
1999년 ~ 2000년 University of Minnesota Visiting Scholar
관심분야: XML 데이터베이스, 데이터 마이닝, 지리정보시스템, 정보검색