

## 공간 데이터 분석을 위한 공간 연관 규칙 탐사 시스템의 설계 및 구현

안찬민\*, 이윤석\*\*, 이주홍\*\*\*, 박상호\*\*\*\*

### Design and Implementation of Spatial Association Rule Discovery System for Spatial Data Analysis

Chan-Min Ahn\*, Yun-Seok Lee\*\*, Ju-Hong Lee\*\*\*, Sang-Ho Park\*\*\*\*

#### 요약

최근 공간 정보들을 효과적으로 이용할 수 있는 기술에 대한 연구가 활발하게 이루어지고 있다. 효율적인 지식 탐사를 위해 다양한 기존의 데이터 마이닝 방법들이 확장되어 공간 데이터 마이닝에 사용되고 있다. 그러나 기존의 공간 연관 규칙 탐사 시스템들은 프레디캣 간의 연산을 통해 규칙을 발견함에 따라 질의 결과에 다양한 비공간 속성들을 반영하지 못하는 문제점을 가지고 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 공간 데이터베이스에서 사용되는 질의를 확장하고, 위상정보에 따른 데이터를 구성한 후 비공간 객체 속성간의 연관 규칙을 발견하는 시스템을 제안한다. 특히 지리 정보 시스템에 적용 가능한 모델을 구현하였다. 이렇게 구현된 시스템은 사용 중인 공간 데이터베이스를 확장하므로 이식성이 뛰어나고, 공간 속성뿐만 아니라 다양한 비공간 속성을 고려함으로써 좀 더 실생활에 유용한 공간 연관 규칙을 발견할 수 있다.

#### Abstract

Recently, the study about the technology which effectively manage spatial information is actively conducted. For the effective knowledge inquiry, various extended data mining methods are applied in spatial data mining. However, former spatial association rule system appears the problem that does not reflect various non-spatial property along the inquiries because it searches the rule from the calculation among predicates. To resolve the problem, present study suggests the system that extends the inquiries using in spatial database, searches the association rule among non-spatial object property after setting the data based on space information. Especially, the model which is applicable to geographical information system is embodied. Embodied system with this method enables to search more useful spatial association rule in real life since it shows high migration property with extended spatial database and considers spatial property and various non-spatial property.

▶ Keyword : 공간 연관 규칙 탐사 시스템(Spatial Association Rule Discovery System), 공간 데이터 마이닝(Spatial Data Mining), 공간 데이터 분석(Spatial Data Analysis), 지리 정보 시스템(Geographic Information System).

• 제1저자 : 안찬민 • 교신저자 : 이주홍

• 접수일 : 2005.11.30, 심사완료일 : 2006.02.01

\*, \*\* 인하대학교 컴퓨터·정보공학과, \*\*\* 인하대학교 컴퓨터공학부 부교수, \*\*\*\* 인하대학교 컴퓨터·정보공학과

## I. 서론

현재 우리가 살고 있는 공간 속에는 수많은 데이터들이 있다. 이러한 데이터는 환경이 변화에 따라 자연적으로 그 양도 빠르게 증가하고 있다. 최근 공간 데이터들을 효과적으로 이용할 수 있는 기술에 대한 연구가 활발하게 이루어지고 있다. 이러한 기술들은 지리 정보 시스템이나 지식 베이스 시스템에 유용하게 사용될 수 있다.[1][2] 그러나 방대한 양의 공간 데이터로부터 사용자가 원하는 정보를 추출하기 위해서는 많은 비용을 필요로 한다. 이러한 문제를 해결하기 위해서 공간 데이터 마이닝 기법들을 이용한다. 공간 데이터 마이닝은 공간 데이터베이스에 저장된 자료들로부터 공간적 속성이 고려된 유용한 지식들을 발견하는 것이다[4][5]. 이러한 지식 발견은 공간 데이터를 이해하고, 공간 데이터와 비공간 데이터간의 본질적인 관계를 찾아내며, 추약된 방법으로 데이터 규칙성을 나타내는 중요한 역할을 하게 된다.

공간 데이터 마이닝 기법에 관한 연구는 크게 공간 데이터의 일반화된 특성을 추출하는 일반화(generalization) 기법, 공간 데이터간의 연관 관계를 추출하는 연관(association) 기법, 공간 객체들 간의 유사성에 의해 그룹화 하는 클러스터링(clustering) 기법, 통계적 방식에 근거한 기법 등으로 분류할 수 있다. 이 중에서도 연관 기법은 주어진 데이터 집합에서 흥미로운 관계를 발견하는 작업으로 공간 지식 발견에 매우 중요한 기법이라 할 수 있다.

기존의 공간 연관 규칙 탐사 시스템들은 프레디컷(predicate) 간의 연산을 통해 연관 규칙을 발견한다. 그러나 각 시스템의 최종 질의 결과를 보면 공간 속성에 대한 분류는 건물이나 도로와 같이 큰 범주에 그칠 뿐 건물의 속성이나 도로의 교통량 등과 같은 세부적인 비공간 속성은 나타나지 않는다. 기존 시스템은 사전에 비공간 속성에 대한 프레디컷이 정의 되어 있지 않으면, 질의 결과에 반영되지 못한다는 문제점이 있다.

본 논문에서는 이러한 문제점을 해결하기 위해 공간 데이터 베이스에서 사용되는 질의를 확장하고, 위상정보에 따른 데이터를 구성한 후 비공간 객체 속성간의 연관 규칙을 발견하는 시스템을 제안한다. 특히 지리 정보 시스템에 적용 가능한 모델을 구현하였다.

## II. 관련 연구

### 2.1 GeoMiner

J. Han이 제안한 GeoMiner[4][6] 시스템은 관계형 데이터 마이닝 시스템인 DBMiner 시스템을 공간으로 확장시킨 시스템이다. 비공간 데이터들은 DBMiner 시스템에서 비공간 데이터 마이닝에 대한 함수들을 통해 처리되고, 공간 데이터들과 비공간 데이터와 공간 데이터 사이의 관계에 대한 처리는 GeoMiner 전용 함수를 통해 이루어진다. 공간 데이터 마이닝 질의 언어는 DBMiner에서 사용하는 DMQL(Data Mining Query Language)을 확장한 GMQL(Geo Mining Query Language)을 사용한다.

GeoMiner는 객체들 간의 위상관계에 따른 프레디컷을 사용하여 공간 연관 규칙을 발견한다. GeoMiner의 공간 연관 규칙 질의 처리 과정은 [그림 1]과 같다[6].

```

Step 1 : Task_relevant_DB :=
        extract_task_relevant_objects(SDB,RDB);
Step 2 : Coarse_predicate_DB :=
        SAM_coarse_computation(Task_relevant_DB);
Step 3 : Frequent_coarse_predicate_DB :=
        filtering_with_minimum_support(Coarse_
                                     predicate_DB);
Step 4 : Fine_predicate_DB :=
        refined_spatial_computation(Frequent_coarse_
                                     predicate_DB);
Step 5 : Find_frequent_predicates_
        and_mine_rules(Fine_predicate_DB);
    
```

그림 1. GeoMiner의 공간 연관 규칙 처리 과정  
Fig. 1 Mining Spatial Association Rule Process of GeoMiner

### 2.2 INGENS

D.Malerba가 제안한 INGENS(Inductive Geographic Information System)[7] 시스템은 지리 정보 시스템(Geographic Information System)과 데이터 마이닝, 그리고

객체 지향 데이터베이스 시스템(OODBMS)을 접목시켰다. 기존의 지리 정보 시스템 응용 프로그램들이 데이터의 수집과 저장, 표현에 초점을 둔 반면, INGENS는 데이터 마이닝 기법의 사용과 공간 데이터의 해석 작업에 초점을 두고 개발되었다.

INGENS는 SPADA(Spatial Pattern Discovery Algorithm)(8) 알고리즘을 사용하여 공간 연관 규칙을 발견한다. 공간 연관 규칙 마이닝은 다음과 같은 논리적 구성을 따른다.

**Given**

- a spatial database (SDB),
- a set of reference objects S,
- some task-relevant geographic layers R<sub>k</sub>, together with spatial hierarchies defined on them,
- two thresholds for each level l in the spatial hierarchies, minsup(l) and minconf(l)

**Find**

- strong multiple-level spatial association rules.

**2.3 SPIN!**

지리 정보 시스템과 데이터 마이닝을 접목시킨 또 다른 시스템으로 M. May가 제안한 SPIN!(9) 시스템이 있다.

SPIN! 시스템의 일반적인 구조는 (그림 2)와 같다. 유연성과 확장성이 높은 Java 2 Enterprise Edition (J2EE)을 기반으로 구현되었으며, 클라이언트/서버 구조로 이루어져 있다. 서버측 컴포넌트는 Enterprise Java Bean(EJB)을 기반으로 구현되었다.

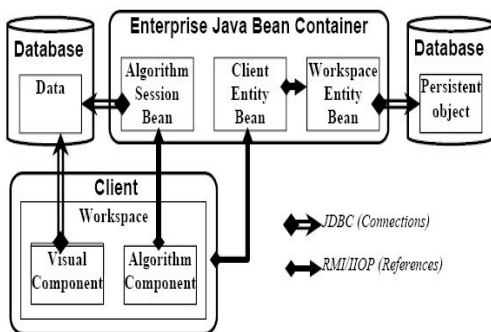


그림 2. SPIN! 시스템 구조  
Fig. 2 SPIN! System Architecture

SPIN! 시스템에서 공간 연관 규칙은 INGENS 시스템과 마찬가지로 SPADA 알고리즘을 사용한다. 하지만 INGENS 시스템이 객체 지향 데이터베이스 기반의 구조로 구성되어 있는 것에 비해, SPIN! 시스템은 관계형 데이터베이스를 기반으로 구성되어 있다는 차이점이 있다.

**2.4 기존 시스템의 문제점**

SPIN! 시스템들은 거리와 위상 관계에 대한 공간 프레딕트 계층(hierarchy)을 구성하여 공간 연관 규칙을 발견한다. 그러나 각 시스템의 최종 질의 결과를 보면 공간 속성에 대한 분류는 건물이나 도로와 같이 큰 범주에 그칠 뿐 건물의 속성이나 도로의 교통량 등과 같은 세부적인 비공간 속성은 나타나지 않는다. 기존 시스템은 사전에 비공간 속성에 대한 프레딕트가 정의되어 있지 않으면, 질의 결과에 반영되지 못한다는 문제점이 있다. 따라서 객체들 간의 지리적 위상 관계보다 각 객체의 속성들도 고려한 종합적인 연관 규칙 탐사 방법이 요구된다.

**III. 공간 연관 규칙 탐사 시스템의 설계 및 구현**

기존의 공간 연관 규칙 탐사 시스템들은 프레딕트 간의 연산을 통해 규칙을 발견함에 따라 질의 결과에 다양한 비공간 속성들을 반영하지 못하는 문제점을 가지고 있다.

본 논문에서는 이러한 문제점을 해결하기 위해 객체의 속성에 대한 연산을 통해 위상정보에 따른 데이터를 구성한 후 비공간 객체 속성간의 연관 규칙을 발견하는 시스템을 제안한다. 제안하는 시스템은 GMS를 기반으로 SMQL(3)을 사용하여 공간 연관 규칙을 발견한다.

GMS는 공간 DBMS로서 다중 사용자를 지원하며 공간 데이터 및 비공간 데이터를 효율적으로 저장, 관리할 수 있다.

**3.1 SMQL**

SMQL은 SIMS(Spatial Information Management System)에서 사용하기 위한 공간 데이터 마이닝 질의 언어이다(3). SMQL은 공간 데이터 마이닝 기법 중에서 Association,

Classification, Clustering, Trend Analysis를 지원한다. 그 중 Association의 BNF는 [그림 3]과 같다.

### 3.2 질의 처리 과정 및 결과

Geominer에서 사용된 방법은 각 객체들의 거리 정보를 계산하여 거리에 대한 정보를 비공간 데이터로 저장한 데이터 베이스에서 연관 규칙 알고리즘을 적용하여 연관 규칙을 찾는다. 본 논문에서 제안한 공간 연관 규칙 탐사 시스템은 SMQL을 사용하여 연관 규칙을 발견한다. SMQL은 GMS를 기반으로 GMS에서 얻어낸 데이터를 마이닝하기 위한 언어이다. SMQL로 작성된 질의는 파서를 통해 GMS에서 사용 가능한 질의로 변경된다. 변경된 질의는 GMS로부터 질의에 주어진 거리에 포함된 데이터를 추출하고, 이 데이터를 이용하여 연관규칙을 발견한다. 예를 통해 질의 처리 과정을 살펴보면 다음과 같다. 질의는 역을 중심으로 500미터 이내의 범위에서 공간 연관 규칙을 찾으려 하였다. 이를 SMQL로 작성하면 [그림 4]와 같다.

```

hierarchy_description ::=
    description_statement (',' hierarchy_description)
description_statement ::=
    hierarchy_name FOR attribute_name
hierarchy_name ::= hierarchies WITH attribute
hierarchies ::= string
attribute ::= attribute_name (',' attribute_name )
threshold_specification ::=
    threshold_description THRESHOLD number
    { AND threshold_specification }
threshold_description ::=
    DISTINCT_VALUE | DESIRABLE_RULE |
    SUPPORT | CONFIDENCE
    
```

그림 3. SMQL에서의 Association BNF  
Fig. 3 Association BNF of SMQL

우선 기준이 될 OID A를 찾는다. 질의에서 Building 테이블의 BusinessType="역"인 모든 OID가 검색된다. 공간 연관 규칙을 찾고자 하는 기준의 역할은 검색 목표를 분명히 하는 효과를 갖는다. 두 번째 단계에서는 기준이 될 A를 기점으로 "이렇게 연관된" 정보를 찾을 것인지 전체적인 집합 B를 구한다. 질의에서 보면 역에서 거리가 500 이내의 범위에 속한 모든 객체들의 OID가 검색된다. 세 번째 단계에서는 이렇게 얻은 데이터 집합 B를 기준으로 규칙을 발견하고자 목표했던 데이터 집합을 구한다. 만약 질의에 Using hierarchy 구문이 명시되어 있으면, 구해진 데이터 집합은 개념 계층을 이용하여 구문에 지정된 level의 값으로 대체된다. 네 번째 단계에서 Apriori 알고리즘을 적용하여 연관 규칙을 발견한다.

질의 수행 결과는 [그림 5]와 같다. 그러나 발견된 규칙의 support가 낮다. 좀 더 정확하고, 일반화된 규칙을 발견하기 위해 BusinessTree라는 계층 테이블을 이용하여 연관 규칙을 찾으려 하였다.

```

MINE ASSOCIATION AS Station
FOR T2.*
FORM Building_T1, Building_T2
WHERE T1.BusinessType = "역"
      AND DISTANCE(T1,T2,500)
SET SUPPORT THRESHOLD 0.2 AND
CONFIDENCE THRESHOLD 0.6;
    
```

그림 4. SMQL 질의 예제  
Fig. 4 SMQL Query Example

```

SMQL_query ::= MINE rule_header |
RETRIEVE TASK RELEVANT DATA
      AS task_name
      { USING HIERARCHY hierarchy_description }
      { USING TASK RELEVANT DATA task_name }
      { FOR analysis_standards }
FROM table_list
      { WHERE conditions }
      { SET threshold_specification }

rule_header ::=
    association_header | characteristic_header
association_header ::=
    ASSOCIATION (AS pattern_literal)
task_name ::= literal
    
```

```

Is_a(X,역) & Near(X,Y) & BusinessType(Y,이발소)
-> OfficePurchasePrice(Y,low) (13.8, 50.8)
Is_a(X,역) & Near(X,Y) & Floor(Y,2)
& BusinessType(Y,맥주)
-> OfficePurchasePrice(Y,high) (13.2, 54.7)
Is_a(X,역) & Near(X,Y) & BusinessType(Y,맥주)
& OfficePurchasePrice(Y,high)
-> Earn(Y,high) (12.8, 51.5)
Is_a(X,역) & Near(X,Y) & Visitor(Y,many)
& BusinessType(Y,소주)
-> OfficePurchasePrice(Y,high) (13.8, 50.8)
Is_a(X,역) & Near(X,Y) & Floor(Y,1)
& BusinessType(Y,여성캐주얼)
-> OfficePurchasePrice(Y,high) (13.9, 50.6)
    
```

그림 5. 질의 결과  
Fig. 5 Query Result

Hierarchy를 이용한 질의는 (그림 6)과 같다. 이 경우 질의 처리 과정 중 세 번째 단계에서 BusinessType의 값을 BusinessTree의 Level1의 값으로 대체시킨다. (그림 7)은 hierarchy가 적용된 질의 결과를 나타낸다.

```

MINE ASSOCIATION AS Station
USING HIERARCHY BusinessTree
WITH level1 FOR BusinessType
FOR T2.*
FORM Building_T1, Building_T2
WHERE T1.BusinessType = "역"
AND DISTANCE(T1,T2,500)
SET SUPPORT THRESHOLD 0.3 AND
CONFIDENCE THRESHOLD 0.5;
    
```

그림 6. Hierarchy를 이용한 SMQL 질의 예제  
Fig. 6 SMQL Query Example using Hierarchy

```

Is_a(X,역) & Near(X,Y) & Floor(Y,1)
& BusinessType(Y,약류)
-> OfficePurchasePrice(Y,high) (33.9, 55.1)
Is_a(X,역) & Near(X,Y) & Floor(Y,2)
& OfficePurchasePrice(Y,high)
-> BusinessType(Y,주점) (31.2, 60.4)
    
```

그림 7. Hierarchy가 적용된 질의 결과  
Fig. 7 Query Result

발견된 규칙을 풀이하면 다음과 같다. "역 부근의 1층에 있는 의류점은 평당 가격이 높고 support는 33.9%, confidence는 55.1%이다" "역 부근의 2층에 위치한 업종

중 평당 가격이 높은 곳은 주점인 경우가 많고 support는 31.2%, confidence는 60.4%이다". 이를 종합하여 보면, 대체적으로 역 부근에서는 주점과 의류점이 호환이라는 것을 알 수 있다.

종종 동일한 조건절을 가진 질의를 반복해서 수행해야 하는 경우가 발생한다. 예로 규칙이 발견되지 않아 질의 목표를 수정해야 하는 경우나 support와 confidence의 수치를 변경해야 하는 경우를 들 수 있다. 이와 같은 경우 작업 관련 데이터를 사용하여 질의 처리 시간을 단축시킬 수 있다. (그림 8)은 인구가 많은 건물 근처의 건물들에 대한 연관 규칙을 찾는 질의 예제이다.

```

RETRIEVE TASK RELEVANT DATA AS pop
FOR T2.attr, T2.age, T2.price, T2.pop
FORM Building_T1, Building_T2
WHERE T1.pop > 200
AND T2.oid = T1.oid
AND DISTANCE(T1,T2,100)
SET SUPPORT THRESHOLD 0.3 AND
CONFIDENCE THRESHOLD 0.5;
    
```

그림 8. 반복 작업을 위한 SMQL 질의 예제  
Fig. 8 SMQL Query Example for repetition

이 경우 질의 처리 과정 중 두 번째 단계까지 검색된 OID는 pop 파일로 저장된다. 질의 결과 규칙이 발견되지 않아 (그림 8)의 질의를 인구가 많은 건물 근처의 도로들에 대한 연관 규칙을 찾는 것으로 목표를 변경하였다. 수정된 질의는 (그림 9)와 같다. 이전 질의에서 처리된 pop 파일을 사용하여 연관 규칙을 찾기 때문에 질의 처리 시간을 상당히 줄일 수 있다. 질의 결과는 (그림 10)과 같다.

```

MINE ASSOCIATION
USING TASK RELEVANT DATA pop
FOR T2.traffic, T2.parking, T2.tra_acc, T2.park_acc
FORM Building_T1, Building_T2
SET SUPPORT THRESHOLD 0.3 AND
CONFIDENCE THRESHOLD 0.5;
    
```

그림 9. 작업 관련 데이터를 사용한 질의 예제  
Fig. 9 Query Example using Task-Relevant Data

```

ls_a(X,300) & Near(X,Y) & Tra_acc(Y,8)
    → Park_acc(Y1) (30.8, 50.2)
ls_a(X,500) & Near(X,Y) & Park_acc(Y,3)
    → Parking(Y,2) (31.2, 60.4)
ls_a(X,300) & Near(X,Y) & Parking(Y,2)
    → Tra_acc(Y,4) (31.2, 60.4)
    
```

그림 10. 질의 결과  
Fig. 10 Query Result

### 3.3 시스템 구조

시스템의 구성은 (그림 11)과 같다. 사용자는 사용자 인터페이스를 통해 질의를 입력한다. 질의어 파서(Parser)는 입력된 질의를 분석하여 파싱 트리(Parsing Tree)를 구성한다. 파싱 트리의 값들은 연관 규칙 발견을 위한 각 모듈로 전달된다.

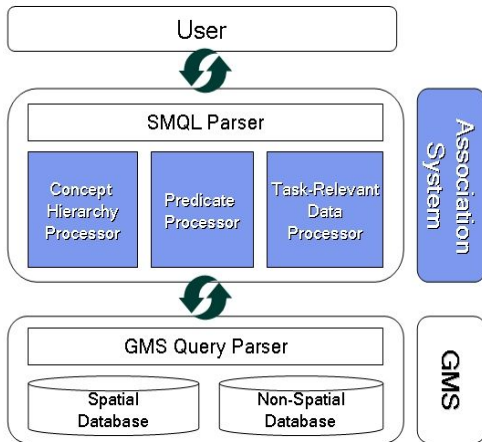


그림 11. 공간 연관 규칙 시스템의 구조  
Fig. 11 Spatial Association Rule Discovery System Architecture

작업 관련 데이터(Task-Relevant Data) 처리 모듈은 파싱 트리로 구성된 질의를 GMS에 사용가능한 질의로 변경하여 조건에 적합한 공간 및 비공간 데이터를 얻는다. 개념 계층(Concept hierarchy) 처리 모듈은 SMQL을 사용하여 새로운 계층 테이블을 생성한다. 또한 연관 규칙을 발견하기 위한 데이터 수집과정에서 수집된 데이터를 질의에 주어진 level의 값으로 대체한다. 프레딕트 처리 모듈은 발견된 규칙에 대한 해석이 쉽도록 미리 정의된 프레딕트를 사용하여 속성을 설명한다.

### 3.4 작업 관련 데이터 처리 모듈

작업 관련 데이터 처리 모듈은 질의에 주어진 조건에 따른 공간 및 비공간 데이터들을 GMS로부터 얻는다. 각 단계에서 생성되는 데이터는 중복되는 데이터를 제거하고 미리 정의된 파일로 저장된다. 만약 사용자가 RETRIEVE TASK RELEVANT DATA 구문을 이용하여 질의하였다면, 미리 정의된 이름이 아닌 사용자가 지정한 이름의 파일로 저장된다. 이렇게 저장된 데이터는 질의의 반복 수행에 이용될 수 있다. 작업 관련 데이터 처리 모듈은 다음과 같은 절차를 따른다.

#### ▶(그림 9)의 질의와 같이 USING TASK RELEVANT DATA 구문을 사용하였을 경우

1. 파일에 저장된 OID의 나머지 속성데이터를 검색한다.
2. (그림 9)의 예제에서 pop 파일에 저장된 OID의 나머지 속성데이터(traffic, parking, tra\_acc, park\_acc)를 검색한다.

#### ▶(그림 8)의 질의와 같이 RETRIEVE TASK RELEVANT DATA 구문을 사용하였을 경우

1. WHERE절의 조건에서 기준이 되는 OID를 검색한다.
2. 기준이 되는 OID와 조건절의 DISTANCE를 계산한다.
3. 계산 결과 검색된 OID를 중복을 제거하고 지정된 파일명으로 저장한다.
4. (그림 8)의 예제에서 pop > 200 인 OID를 기준으로 거리가 100 이내인 객체들의 OID가 검색된다. 검색 결과에서 중복되는 OID를 제거하고 pop 파일로 저장된다.

#### ▶(그림 4)와 같이 그 이외의 질의인 경우

1. WHERE절의 조건에서 기준이 되는 OID를 검색한다.
2. 기준이 되는 OID와 조건절의 DISTANCE를 계산한다.
3. 계산 결과 검색된 OID의 나머지 속성데이터를 검색한다.

(그림 4)의 예제에서 BusinessType="역" 인 OID를 기준으로 거리가 500 이내인 객체들의 OID가 검색된다. 검색된 OID의 나머지 모든 속성데이터를 검색한다.

### 3.5 개념 계층 처리 모듈

데이터베이스에 저장된 객체를 연관 규칙에 효율적으로 사용하기 위해 계층적 구조로 정리할 필요가 있다. 개념 계

층(Concept hierarchy)은 하위 개념들의 집합으로부터, 보다 일반적인 상위 개념들로의 순차적인 매핑을 정의한다[10]. 개념 계층구조는 원시 데이터를 보다 일반적인 상위 추상 단계에서 다룰 수 있다는 점에서 유용하다.

개념 계층 처리 모듈은 계층 테이블을 이용하여 속성 값을 상위 또는 하위 개념 단계의 값으로 대체한다. 개념 계층 처리 모듈은 다음과 같은 절차를 따른다.

▶(그림 6)의 질의와 같이 USING HIERARCHY 구문을 사용하였을 경우

1. 질의에 주어진 hierarchy table과 level을 얻는다.
- 2.Task-relevant data 처리 모듈로부터 결과 레코드들을 얻어온다.
- 3.얻어온 결과 레코드에서 값을 변경시킬 속성이 숫자인지 문자인지 판별한다.
- 4.숫자일 경우 hierarchy table에서 숫자 값이 속하는 필드의 튜플을 찾는다.  
문자일 경우 hierarchy table에서 문자 값과 동일한 필드의 튜플을 찾는다.
- 5.해당 튜플에서 질의에 지정된 level의 값을 얻는다.
- 6.결과 레코드에 해당 속성의 값을 얻은 level의 값으로 대체시킨다.

(그림 6)의 예제에서 BusinessTree 라는 hierarchy table과 변경할 레벨인 level1을 얻는다. Task-relevant data 처리 모듈로부터 얻어온 레코드에서 Business Type 속성에 대해 숫자 값인지 문자 값인지 판별한다. 판별된 값과 BusinessTree 테이블의 필드 값이 일치하면 해당 튜플에서 level1 필드의 값을 얻는다. 기존의 BusinessType 속성 값을 얻어온 level1의 값으로 대체시킨다.

### 3.6 프레디킷 처리 모듈

프레디킷 처리 모듈은 발견된 규칙에 대한 해석이 쉽도록 미리 정의된 프레디킷을 사용하여 속성을 설명한다. 공간 데이터베이스에 저장된 객체를 이용한 공간 연관 규칙 발견 모듈은 객체의 출현 빈도에 따른 연관성을 보여줄 뿐 구체적으로 무엇을 의미하는지 의미가 모호해지는 경우가 있다. 이를 위해 공간 및 비공간 정보에 대한 프레디킷을 처리하는 모듈을 설계하였다. 프레디킷은 질의를 사용하여 미리 정의할 수 있다.

프레디킷 처리 모듈은 다음과 같은 절차를 따른다.

- 찾고자 하는 값이 비공간 정보인지, 공간정보인지 판별한다.
- 비공간 정보인 경우, predicate 테이블에서 해당 속성 값을 찾고 대응되는 predicate로 변환하여 결과로 출력한다.
- 공간 정보인 경우, 거리 값 대신 질의에서 주어진 predicate 값을 결과로 출력한다.
- Predicate가 정의되어 있지 않으면 테이블의 속성 값을 출력한다.

## IV. 결론 및 향후 과제

효율적인 지식 탐사를 위해 다양한 기존의 데이터 마이닝 방법들이 확장되어 공간 데이터 마이닝에 사용되고 있다. 그러나 기존의 공간 연관 규칙 탐사 시스템들은 프레디킷의 연산을 통해 규칙을 발견한다. 그렇기 때문에 사전에 비공간 속성에 대한 프레디킷이 정의 되어 있지 않으면 질의 결과에 다양한 비공간 속성들을 반영하지 못하는 문제점을 가지고 있다.

본 논문에서는 이러한 문제점을 해결하기 위한 시스템을 제안하였다. 공간 데이터베이스에서 사용되는 질의를 확장하고, GMS를 기반으로 시스템을 구현하여 위상관계 뿐만 아니라 다양한 비공간 속성이 고려된 질의 결과를 보였다.

향후 지도 데이터의 상세화와 다양한 공간 및 비공간 정보를 추가하여 직선거리뿐만 아닌 실세계에 더욱 근접한 정보를 분석할 수 있도록 개발되어야 할 것이다.

## Acknowledgment

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성.지원사업의 연구결과로 수행되었음.

## 참고문헌

- [1] 오석, "지리 정보 추출의 자동화 알고리즘. 한국컴퓨터 정보학회 논문지 5권 4호", pp 21-27, 2000.
- [2] .조성훈, 안동규, 김재홍, "CEO의 효율적/유효적 의사결정을 위한 경영성과 데이터마이닝 시스템의 구축. 한국컴퓨터정보학회 논문지 5권 4호", pp 41-47, 2000.
- [3] 박선, 박상호, 안찬민, 이윤석, 이주홍, "SIMS를 위한 공간 데이터 마이닝 질의 언어. 한국정보과학회 춘계 학술발표논문집, 제 31권 제1호, pp 70-72, 2003.
- [4] J. Han, K. Koperski and N. Stefanovic, "GeoMiner : A system prototype for spatial data mining", In Proc. of the ACM SIGMOD. pp 560-563, 1997.
- [5] K. Koperski and J. Han, "Discovery of Spatial Association Rules in Geographic Information Databases", In Proc. of the SSD. pp 47-66, 1995.
- [6] K. Koperski, "A Progressive Refinement Approach to Spatial Data Mining", Ph.D. Thesis, Simon Fraser University, 1999.
- [7] D. Malerba, F. Esposito, A .Lanza and F. A. Lisi, "Discovering Geographic Knowledge: The INGENS System", In Proc. of the ISFIS. pp. 40-48, 2000.
- [8] D. Malerba and F. A. Lisi, "An ILP method for spatial association rule mining", Working notes of the First Workshop on Multi -Relational Data Mining, 2001.
- [9] M. May, "SPIN! an Integrated Spatial Knowledge Discovery Platform", Leopold, E. (ed.) Fachgruppentreffen Maschinelles Lernen der Gesellschaft für Informatik, Sankt Augustin pp 18-20, GMD Report, 2000.
- [10] J. Han and M. Kamber, "Data Mining, Concepts and Techniques", Morgan Kaufmann Publishers, 2001.

## 저자 소개



### 안 찬 민

2003년 2월 인하대학교  
컴퓨터공학과 학사  
2003년 ~ 현재 : 인하대학교  
대학원 박사과정  
〈관심분야〉 데이터마이닝, 알고리즘,  
프로그래밍어 및 설계



### 이 윤 석

2003년 8월 한서대학교  
컴퓨터정보학과 학사  
2003년 ~ 현재 : 인하대학교  
대학원 석사과정  
〈관심분야〉 데이터마이닝,  
데이터베이스



### 이 주 홍

2001년 2월 한국 과학 기술원  
컴퓨터 공학 박사  
2002년 ~ 현재 : 인하대학교  
컴퓨터공학부 부교수  
〈관심분야〉 데이터마이닝,  
데이터베이스, 정보검색,  
신경망, 기계학습



### 박 상 호

2002년 2월 인하대학교  
컴퓨터공학과 학사  
2004년 2월 인하대학교 대학원  
컴퓨터정보공학과 석사  
2004년 ~ 현재 : 인하대학교  
대학원 박사과정  
〈관심분야〉 데이터마이닝,  
데이터웨어하우스