

# 지식발견(KDD)을 응용한 지역개발계획수립 지원 프로세서의 설계

한상진\*, 김호석, 김성희, 배해영  
인하대학교 전자계산공학과

[jinsting@lycos.co.kr](mailto:jinsting@lycos.co.kr), {bluesnow, laufer}@dmlab.inha.ac.kr, [hybae@inha.ac.kr](mailto:hybae@inha.ac.kr)

## A Design of Region-Development Plan Support Processor Using Knowledge Discovery in Database.

Sang-Jin Han\*, Ho-Suk Kim, Hee-Sung Kim, Hae-Young Bae  
Dept. of Computer Science & Engineering, Inha University

### 요 약

최근 정보기술의 가속적인 발전과 인터넷의 급속한 보급으로 인하여 우리는 다양하고 방대한 양의 지역정보를 접하고 이용하고 있다. 그러나 지역개발사업을 추진하는데 있어서 계획수립이 차지하는 중요성이 매우 큼에도 불구하고 지역을 대표하는 객관적이고 유용한 정보를 찾아내어 지역개발계획수립에 활용하는 예는 거의 없었다. 이에 여러 곳에 산재되어있는 지역정보들을 통합하여 관리하고 이러한 대량의 지역 데이터들로부터 지역을 특징지을 수 있는 보다 현실적이고 유용한 정보를 추출하거나 생성하여 지역정보 분석에 활용하는 방법이 필요하게 되었다.

본 논문에서는 지역개발계획을 수립하는데 있어서 방대한 양의 데이터로부터 유용한 정보를 추출하고 발견하는 지식발견(KDD : Knowledge Discovery in Database)[1] 프로세서의 전체과정에 지역개발계획 수립 목적에 맞추어 지역개발이론에 기초한 지역정보 분석과정을 삽입함으로써 보다 합리적이고 현실적인 지역개발계획이 수립되도록 지원할 수 있는 프로세서를 설계한다.

### 1. 서론

정보기술의 빠른 발전은 엄청난 양의 데이터를 전자적으로 수집하고 보관하는 기능을 가능케 하였다. 그러나 그러한 대량의 가공되지 않은 데이터는 그 자체로 별 의미가 없다. 데이터의 양이 오늘날과 같이 방대하지 않았던 과거에는 소수의 전문가들이 통계기법이나 질의를 통해 데이터를 분석하고 요약된 결과를 보고서 형식으로 제공하곤 하였으나, 이러한 접근법은 데이터의 양이 증가하고 문제의 영역이 확대되고 복잡해짐에 따라 효율성이 떨어지며, 분석을 통해 얻을 수 있는 정보의 품질도 기대하기 어렵다. 이에 데이터의 홍수와 정보의 빈곤이라는 환경에 처한 지금, 데이터를 여과하고, 분석하며, 결과를 해석하는 데이터 분석방안에 높은 관심을 갖게 되었으며, 지식발견(Knowledge Discovery in Database), 정보발견(information discovery), 정보수확(information harvesting) 등의 이름으로 대량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 활용하는 기법들이 발전 되어지고 있다. 마찬가지로 지역개발계획을 수립하는데 있어서 대량의 지역 데이터들로부터 보다 객관적이고 현실적인 정보들을 추출하여 지역정보 분석에 활용할 필요성이 있다.

본 논문은 앞에서 언급한 기법들 중 지식발견(KDD) 프로세서의 전체과정을 단계별로 응용하고 지역정보 분석과정

을 삽입함으로써 지역개발계획을 수립하는 담당자로 하여금 지역에 관련된 다양한 형태의 자료들을 데이터베이스로 구축하고 구축된 데이터로부터 보다 현실적이고 객관적인 지역지표를 생성하여 지역정보 분석에 활용할 수 있도록 함으로써 지역별 차별화 없이 일률적으로 각종 개발사업을 추진하는 것을 지양하고, 보다 유연하고 현실성 있는 개발계획을 수립할 수 있도록 하는 프로세서를 설계한다.

본 논문의 구성은 다음과 같다. 제 2 장 관련 연구에서는 본 연구의 기본이 되는 데이터웨어하우스와 지식발견(KDD) 프로세서의 대해서 간략하게 소개한다. 제 3 장에서는 개발계획수립지원을 위한 전체적인 프로세서 모형에 대해서 설명 하고, 제 4 장에서는 지식발견(KDD) 프로세서의 전체과정을 단계별로 응용한 지역개발계획 수립 지원 프로세서의 구조를 설계한다. 마지막으로 제 5 장에서 결론 및 향후 과제를 논한다.

### 2. 관련 연구

#### 2.1 데이터웨어하우스(Data Warehouse)

데이터 웨어하우스는 1980년대 중반 IBM 이 자신이 하드웨어를 판매하기 위해 처음으로 도입했던 개념으로, IBM 은 인포메이션 웨어하우스(Information Warehouse)라는 용어를 사용하였다. 이후 이 개념은 많은 하드웨어, 소프트웨어

및 톨 공급 업체들에 의해 이론적, 현실적으로 성장하였으며, 1980년대 후반 Inmon 이 데이터 접근 전략으로 데이터 웨어하우스 개념을 사용함으로써 많은 관심과 집중을 받게 되었다.

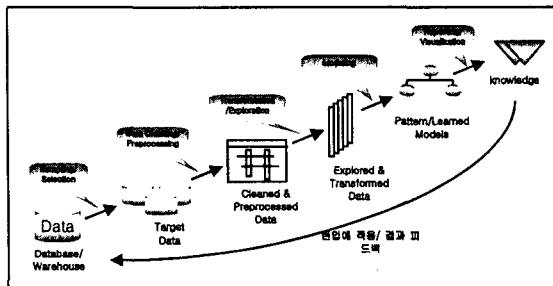
가장 대표적인 정의는 Inmon(1992)[7]의 것으로 그는 데이터 웨어하우스를 '기업의 의사 결정 과정을 지원하기 위한 주제 중심 적이고 통합적이며 시간성을 가지는 비휘발성 자료의 집합'으로 정의하고 있다. 또한 Kelly(1994)[8]는 전사적 데이터 웨어하우스를 '기업 내의 의사 결정 지원 어플리케이션들을 위한 정보기반을 제공하는 하나의 통합된 데이터 저장공간'으로 정의하고 있다.

2.2 지식발견(Knowledge Discovery in Database)

지식발견(KDD)은 1990년대 초반부터 대량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 의사결정에 활용하는 작업으로 소개되어 왔는데 1995년 캐나다 몬트리올에서 개최된 지식발견과 데이터마이닝에 관한 국제 학술대회(The first international conference on knowledge discovery & data mining)에서 지식발견은 데이터로부터 유용한 정보를 발견하는 프로세스의 전과정이라고 정의되었다.

일반적으로 지식발견 프로세스는 다음의 여섯 단계로 구성된다. 최종적으로 필요로 하는 정보를 정의하고 이를 위해 요구되는 원천 데이터의 성격과 소재, 그리고 충실도를 파악하는 준비작업을 거친 후 시작한다[1].

- 1 단계 : 데이터의 선택(Data Selection)
- 2 단계 : 데이터의 정제(Data Cleaning)
- 3 단계 : 데이터의 보완(Data Enrichment)
- 4 단계 : 데이터의 변환(Data Transformation)
- 5 단계 : 데이터마이닝 기법 선택 및 적용 (Selection & Application of Data Mining Techniques)
- 6 단계 : 모형의 평가(Model Evaluation)

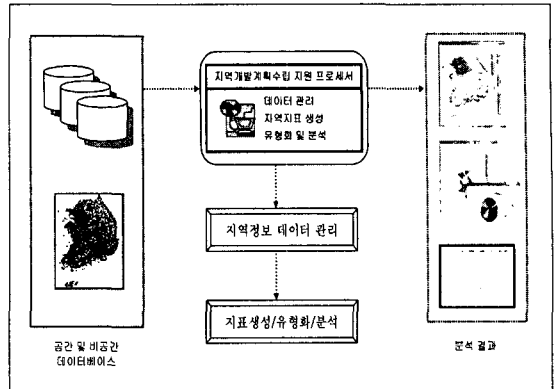


[그림 1] 지식발견(KDD) 프로세스의 단계

3. 지역개발계획수립 지원 프로세스의 전체 모형

지역개발계획 지원 프로세스는 기존에 연구되어지고 있는 지식발견(KDD) 프로세스의 데이터마이닝기법 선택 및 적용 단계에 지역개발이론에 기초한 다양한 지역정보분석 과정을 삽입하고 지역정보나 분석결과에 대한 데이터 시각화를 통해서 보다 현실적이고 객관적인 지역정보를 활용하여 지역개발계획을 수립할 수 있도록 지원할 수 있는 프로세스이다.

다음의 [그림 2]는 지역개발계획수립 지원 프로세스의 전체적인 모형을 보여준다

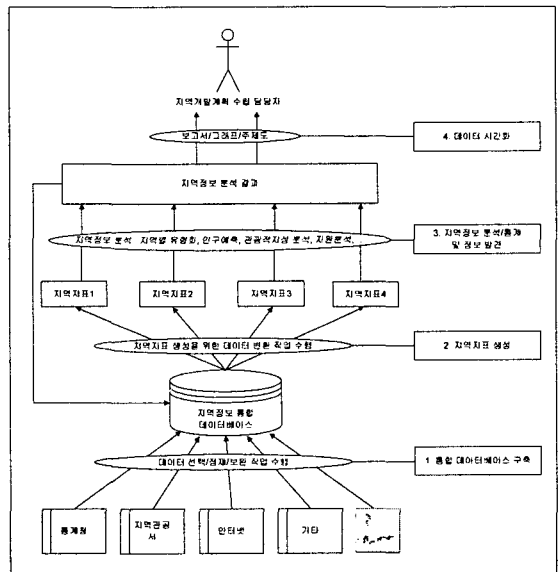


[그림 2] 지역개발계획수립 지원 프로세서의 모형

본 논문에서 제안하는 지역개발계획수립 지원 프로세스는 지식발견(KDD) 프로세스의 6 단계를 응용하여 지역별로 산재되어 있는 다양한 형태의 기초자료들을 한데 모아 데이터베이스로 구축하여 관리하고, 구축된 데이터베이스를 이용하여 사회변화에 유연하게 대처할 수 있도록 사용자 질의에 의해서 지역의 특성을 나타낼 수 있는 현실적이고 객관적인 지역지표를 생성한다. 생성된 지역지표들은 지역개발이론[5] 등에 입각하여 자원분석, 상관분석, 지역별 유형화, 인구예측 등의 분석 및 통계 과정을 거쳐 보고서나 그래프, 주제도 등의 가시적인 결과물로 사용자에게 보여줌으로써 보다 효율적이고 현실성 있는 지역개발계획이 수립될 수 있도록 의사결정자들에게 정보를 제공한다.

4. 지역개발계획수립 지원 프로세서의 구조

지역개발계획수립 지원 프로세스는 [그림 3]과 같이 크게 4 단계로 구성된다.



[그림 3] 지역개발계획수립 지원 프로세서의 구조

4.1 통합 데이터베이스 구축

지역정보 통합 데이터베이스 구축단계는 지역개발계획수립 지원을 위해 활용 되어지는 지역의 기초자료들로 통계청이나 인터넷, 지역기관 등을 통해서 취득할 수 있는 text, excel, html 파일의 지역자료들을 변환하여 데이터베이스로 구축한다. 구축된 모든 데이터는 행자부코드를 기본키로 생성되며, 지식발전(KDD)프로세서의 데이터의 정제, 보완단계를 적용하여 데이터의 충실도와 정확도를 증가시킨다.

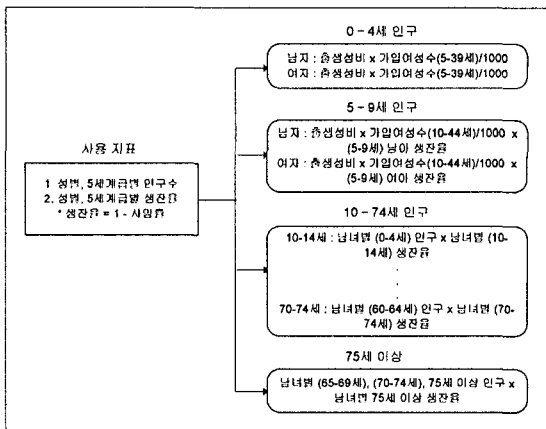
4.2 지역지표 생성

지역지표는 지역정보 분석에 이용되는 객관적이고 과학적인 기준자료로서 지역개발이론에 따르면 관광개발권, 영농권, 수계권등의 지역구분 지표와 비교지역특성을 나타내는 대외여건 지표, 소득이나 토지현황, 환경수준 등을 나타내는 지역특성 지표로 구분 지을 수 있다. 지역지표 생성은 구축된 기초지역정보 통합 데이터베이스로부터 파생항목을 만들거나 항목의 값을 그룹화시키는 작업을 통하여 생성한다. 예를 들어 지역의 입야면적을 전체면적으로 나누어 준 '입야율'이 파생항목의 예가 될 수 있다. 또한, 변화하는 사회 환경에 맞추어 사용자가 유동적으로 생성할 수 있도록 통합 데이터베이스에 대한 사용자의 질의를 처리하는 기능이 포함된다.

4.3 지역정보 분석 및 정보 발전

지표분석은 지역에 대한 생산성 이론, 부족자원 이론, 수출기반 이론, 제도 이론 등의 지역개발이론을 적용하여 지역에 대한 자원평가, 인구예측, 상관분석, GRDP 분석 등을 수행한다. 또한 지역을 일정한 단위로 나누어 등급을 매길 수 있는 지역별 유형화 작업을 통해서 유사지역을 분류하고 정보를 비교할 수 있게 한다. 지역지표에 대한 분석을 통하여 보다 현실적이고 객관적으로 지역에 대한 현황을 파악할 수 있으며, 지역의 특성에 맞는 개발계획을 수립할 수 있다.

다음의 [그림 4]는 지표분석의 한 예로 사용 지표와 코호트 생산률법에 의한 5 세간격별 인구예측 방법을 나타낸다.

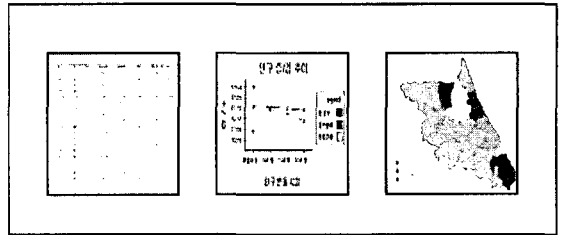


[그림 4] 5 세간격별 인구예측

4.4 데이터의 시각화

지역의 기초정보나 지역지표, 분석결과 등을 보고서나 그래프, 행정경계 공간데이터와 연계된 주제도 등으로 시각화함으로써 지역개발계획수립 지원 프로세서의 효율성과 활용성을 증가시킨다.

다음의 [그림 5]는 지역정보 데이터 시각화의 예이다.



[그림 5] 데이터의 시각화

5. 결론 및 향후 과제

본 논문에서는 지식발전(KDD) 프로세서의 전체과정을 응용하고 지역개발이론에 기초한 지역정보 분석 단계와 데이터 시각화 단계를 삽입하여 보다 효율적이고 현실적인 지역개발계획수립 지원 프로세서를 설계하였다. 지역개발계획수립 지원 프로세서는 지역의 기초정보들을 데이터베이스로 구축하고 지역을 대표할 수 있는 지역지표를 생성하여 분석에 활용할 수 있는 의사결정 지원 프로세서라고 할 수 있다.

향후 과제로는 대량의 데이터로부터 잠재적으로 유용한 정보를 추출하는 기법들에 대한 연구가 꾸준히 진행되고 있으므로 이러한 연구들을 지속적으로 적용시켜야 한다. 또한 행정경계 공간데이터를 이용한 단순한 주제도의 생성뿐만 아니라, 공간데이터의 분석을 통한 지역지표의 생성에 대한 연구를 통하여 제한한 프로세서의 성능과 활용도를 증가시킬 수 있어야 한다.

참고문헌

- [1] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, KDD-96, AAAI Press, 1996
- [2] Brachman, R., and Anand, T., The process of Knowledge Discovery in Database: A Human Centered Approach. In AKDDM, Cambridge, MA: AAAI/MIT Press, 1996
- [3] Fayyad U., Piatetsky-Shapiro G, SmythP., "From Data Mining to Knowledge Discovery in Database", AI Magazine, 1996
- [4] Tandem Computer Incorporated, "Knowledge Discovery through Data Mining", White Paper, 1996
- [5] 박종화, 윤대식, 이종열, "지역개발론", 박영사, 1996
- [6] 정지용, 임상봉, "지역사회개발학", 서울대학교 출판부, 1998
- [7] Inmon, W. H., Building the Data Warehouse, 2nd ed. John Wiley & Sons 1996
- [8] Kelly, S., Data Warehousing : the router to mass customization, John Wiley & Sons, pp.184-186, 1996
- [9] Oracle, Data Warehousing Concept Paper, 한국 오라클, pp.1, 1997
- [10] 조재희, 박성진, "데이터웨어하우징과 OLAP", 대청, 1996