

공간 연관규칙의 추출과 적용*

- 도시성장 예측모델을 사례로 -

조성휘** · 박수홍***

Extraction and Application of Spatial Association Rules*: A Case Study for Urban Growth Modeling

Seong hwi Cho** and Soo hong Park***

요약 : 최근 GIS 분야에서는 GIS와 동적 프로세스 모델링에 기반한 셀룰라 오토마타(CA)의 결합을 통한 공간 모델링 연구가 활발히 진행되고 있다. 그러나 선행 연구에서의 CA 모델링은 해결하고자 하는 문제의 핵심이 되는 규칙을 찾아내는 기능은 제공하지 못하고 있다. 본 연구에서는 GIS와 지식 발견 기법을 이용하여 공간 연관 규칙을 추출하는 방법론을 제시하고 실제 사례에 적용하였다. 이러한 방법론의 제시는 CA 모델링의 기능을 더욱 향상시킬 것으로 기대된다. 또한 본 연구에서 사례로 적용한 도시지역 확장 외에 다양한 문제에 대한 활용이 가능할 것으로 판단된다.

주요어 : 셀룰라 오토마타, 지리정보시스템, AOI

Abstract : Recently spatial modeling that combined GIS and Cellular Automata(CA) which are based on dynamic process modeling has been discussed and investigated. However, CA-based spatial modeling in previous research only provides the general modeling framework and environment, but lacks of providing simulation or transition rules for modeling. This study aims to propose a methodology for extracting spatial relation rules using GIS and Knowledge Discovery in Database(KDD) methods. This new methodology has great potentials to improve CA-based spatial modeling and is expected to be applied into several examples including urban growth simulation modeling.

Key Words : Cellular Automata(CA), GIS, Attribute-Oriented Induction(AOI)

1. 서론

1) 연구배경과 목적

지리정보의 획득과 처리기술이 급속히 발전함에 따라 다양한 형태의 방대한 공간 데이터들이 수집되고 있으며 이를 보다 효율적으로 관리하고 이용하기 위해 데이터베이스로 구축되고 있다. 사람들

의 관심사는 이제 구축된 데이터베이스의 관리 차원에 국한되지 않고 이를 이용하여 적극적으로 공간적인 문제를 해결하고 예측하는 방향으로 그 범위를 넓혀가고 있다. 이러한 요구에 맞추어 볼 때 GIS는 데이터의 저장, 관리, 표현에 있어서 강력한 기능을 가지고 있지만 상대적으로 분석 및 모델링 기능에 있어서는 공간분석의 단순성, 시뮬레이션 최적화와 수행시간 등에 대한 제약사항을 가지고

* 이 논문은 한국과학재단 목적기초연구(2000-2-51300-001-3) 지원에 의하여 연구되었음.

** 인하대학교 지리정보공학과 석사과정(Master Student, Dept. of Geoinformatic Engineering, Inha University) imcho1@nate.com

*** 인하대학교 지리정보공학과 조교수(Assistant Professor, Dept. of Geoinformatic Engineering, Inha University) shpark@inha.ac.kr

있다(정재준, 2001).

이에 GIS가 가진 한계성을 극복하려는 목적으로 GIS와 각종 모델의 통합에 대한 연구 등이 진행되어 왔으며 특히 동적 프로세스 모델링을 위한 셀룰라 오토마타(Cellular Automata, CA)와의 통합에 대한 연구가 진행되어 왔다(Park and Wagner, 1997). 하지만 선행 연구들에서 제안된 CA 모델은 해결하고자 하는 문제의 핵심인 변화규칙을 스스로 정의하지 못하는 한계를 가지고 있다.

본 연구는 GIS가 가진 공간연산 기능과 일반적인 지식 발견(Knowledge Discovery in Database, KDD) 기법 중 하나인 AOI(Attribute-Oriented Induction)를 이용하여 공간 연관규칙 추출에 적용 가능한 방법론을 제시하는 것을 목적으로 한다. 부차적으로는 이 방법론을 도시성장 모델링에 적용하여 방법론의 타당성과 활용 가능성을 입증하고자 한다.

2) 연구범위와 방법

본 연구에서는 공간 연관규칙을 추출하기 위하여 우선 물리적인 도시지역의 확장(도시성장)이라는 문제에 초점을 맞추었다. 또한 수도권 내에 위치한 대도시 중 경기도 수원시(121,093,951.3m², 2003년 1월 1일 기준)를 실험을 위한 공간적 범위로 선정하고 1960년대부터 1990년대까지 약 30년간 10년 단위로 구축된 다양한 지형자료를 이용하여 공간 연관규칙을 추출하고 모형에 적용하였다.

공간 연관규칙의 추출은 두 가지 과정을 통해 수행하였다. 우선 GIS가 제공하는 공간연산 기능을 이용하여 공간적 패턴을 추출하고 다음으로 지식 발견 기법인 AOI를 이용하여 최종적으로 일련의 규칙들을 추출하였다. 추출된 규칙은 모형에 적용할 수 있는 알고리즘으로 작성하였다.

모형은 CA 기반의 시뮬레이터인 Cellular Automata System과 상용 GIS 소프트웨어인 ArcGrid가 결합된 통합시스템을 사용하여 구현하였다. Cellular Automata System은 Cellang이라는 셀룰라 언어를 제공하여 작성된 알고리즘을 쉽게 반영할 수 있도록 한다(박수홍, 1997). 모형을 이용한 시뮬레이션 결과는 Clarke의 UGM 결과와 비교·분석하였다.

2. 연구관련 기초이론

1) 셀룰라 오토마타

셀룰라 오토마타는 1950년대 초 John von Neuman과 그의 동료 Stanislaw Ulman에 의해 발표된 이론 및 기법으로 물리학, 화학, 생물학, 수학, 컴퓨터 공학 등을 포함한 여러 자연과학 및 공학 분야에서 다양한 동적 현상들을 연구하는데 사용되어 왔다. 셀룰라 오토마타의 일반적인 특성은 무질서 시스템(chaotic system)과 인공생명(artificial life)을 연구하는데 있어 주요한 수단으로 간주되고 있으며 동적이고 복잡한 자연 현상들을 시뮬레이션 하는데 유용한 도구로 사용되어 왔다.

셀룰라 오토마타 모델은 크게 네 가지의 기본 요소로 구성된다. 첫 번째 요소는 셀룰라 공간(cellular space)으로 무한한 다차원의 공간을 동일한 크기의 셀로 분할한 공간을 의미하며 여러 가지의 정규 다각형으로 분할이 가능하다. 보통 삼각형, 사각형, 육각형이 많이 사용된다. 두 번째 구성 요소는 지역 상태(local state)로 이는 주어진 시간에서 각 셀의 상태, 구체적으로는 셀이 가지는 값(주로 숫자로 표현)이다. 세 번째 구성요소는 주변셀(neighborhood)이다. 주변셀은 중심셀(focus cell) 주위에 인접하고 있는 셀들의 집합을 말하는데 보통 거리, 방향, 혹은 각도에 의해 결정된다. 셀룰라 오토마타의 주변셀은 영상처리나 공간필터(spatial filter)의 마스크(mask)나 윈도우(window)와 개념적으로 매우 유사하다. 셀룰라 오토마타에서 주변셀의 구성은 매우 자유롭기 때문에 다양한 형태나 크기의 구성이 가능하며 다차원 혹은 매시기별 다른 주변셀을 구성할 수 있다. 셀룰라 오토마타의 마지막 구성요소는 변화규칙(transition rule)이다. 변화규칙은 각 셀이 매 시간별로 어떻게 변화할 지를 규정하는데 여기에는 주변셀들의 구성과 위치가 정의되어 있다. 변화규칙은 모든 셀룰라 공간에 동일하게 적용되는 공통규칙(universal rule)으로 지역적인 특성을 지니고 있다. 보통 변화규칙은 특정한 컴퓨터 언어로 기술되거나 셀룰라 오토마타 패키지에서 제공되는 언어로 기술된다(강영욱·박수홍, 2000).

$$S^{t+1} = f(S^t, N)$$

일반적인 셀룰라 오토마타의 원리는 위의 식과 같다. S는 셀룰라 오토마타가 갖는 모든 가능한 지역 상태의 집합을 의미하며, N은 각 셀이 갖는 주변셀들로 시간 t부터 t+1까지의 변화를 정의한 변화규칙 함수 f의 입력값으로 사용된다(Li and Yeh, 2000). 매 시간 각 셀이 갖는 지역상태는 변화규칙에 의해 새로운 상태로 변화하게 되고 이 상태는 다시 그 다음 번 단계의 초기상태로 입력되는 재귀적 방식을 취한다.

2) AOI 기법

지식 발견은 데이터 내에 암시적(implicit) 혹은 잠재적으로 내재되어 있는 유용한 정보를 추출하는 기법이다. 기존의 데이터베이스가 사람이 직접 분석 가능한 크기와 양을 초과함에 따라 데이터베이스에서 자동적으로 필요한 지식을 추출하기 위한 요구가 발생하였다. 특히 1990년대에 들어 데이터베이스와 기계학습(machine learning) 연구자들에 의해 주된 연구 분야로 자리 잡게 되었다. 그 중 관계형 데이터베이스에서 지식 발견을 위해 제안된 기법 중 하나가 AOI이다(Han et al, 1992; Han and Kamber, 2001). AOI는 데이터의 상향식 요약 방식으로 각각의 속성이 원하는 수준으로 일반화될 때까지 불필요한 속성의 제거, 튜플(tuple)의 통합을 반복해 나가는 방법이며 일련의 튜플로 데이터가 가진 특징 또는 규칙이 요약된다.

(1) AOI의 기본요소

일반적으로 하나의 데이터베이스 내에 저장된 데이터의 양은 방대하지만 특정 문제와 관련된 데이터는 그 중 일부에 지나지 않는다. 예를 들어 과학을 전공하는 대학원생들의 특징을 찾기 위해서는 그 대학 또는 대학 내의 전체 학생에 대한 데이터베이스가 아니라, 과학을 전공하는 대학원생들의 데이터만 필요한 것과 같다. 관련 데이터는 몇 개의 릴레이션(relation) 이하로 표현된다.

배경 지식은 일반화 과정을 조정하기 위해 필수적이며 개념 계층(concept hierarchy)으로 표현한다. 개념 계층은 개념의 분류를 통해 속성값의 범위를 몇 개의 서로 다른 계층으로 조직화한 것을 의미하며 일반화된 상위단계에서부터 세부적인 하위단계까지로(general-to-specific) 구성된다. 최상위 단

계는 값을 갖지 않으며(null description) 예약어 "ANY"로 표현한다. 최하위 단계는 데이터베이스에서 속성이 갖는 실제값들로 구성된다. 개념 계층은 일반화된 개념을 통해 추출된 규칙을 표현할 수 있도록 하며 사용자가 원하는 단순하고 명시적인(explicit) 형식으로 나타내도록 한다. 표 1은 어느 대학의 학생 관련 데이터베이스의 개념 계층을 표현한 예이다. 예를 들어 {biology, chemistry, computing, ... , physics} C science라는 형식은 science가 속성 Major의 속성값인 biology, chemistry, computing 등의 상위 개념임을 의미한다.

표 1과 같은 개념 계층은 다시 개념 트리(concept tree)의 형태로 표현할 수 있다. 개념 트리는 개념들 간의 계층 또는 포함관계를 더 쉽게 알 수 있도록 한다. 그림 1은 속성 Status의 개념 트리

표 1. 데이터베이스의 개념 계층 테이블

{ biology, chemistry, computing, ... , physics } C science
{ literature, music, ... , painting } C art
{ science, art } C ANY (Major)
{ freshman, sophomore, junior, senior } C undergraduate
{ M.A., M.S., Ph.D. } C graduate
{ undergraduate, graduate } C ANY (Status)
{ Burnaby, ... , Vancouver, Victoria } C British Columbia
{ Calgary, ... , Edmonton, Lethbridge } C Alberta
{ Hamilton, Toronto, ... , Waterloo } C Ontario
{ Bombay, ... , New Delhi } C India
{ Beijing, Nanjing, ... , Shanghai } C China
{ China, India, Germany, ... , Switzerland } C foreign
{ Alberta, British Columbia, ... , Ontario } C Canada
{ foreign, Canada } C ANY (Birth_Place)
{ 0.0 - 1.99 } C poor
{ 2.0 - 2.99 } C average
{ 3.0 - 3.49 } C good
{ 3.5 - 4.0 } C excellent
{ poor, average, good, excellent } C ANY (Grade)

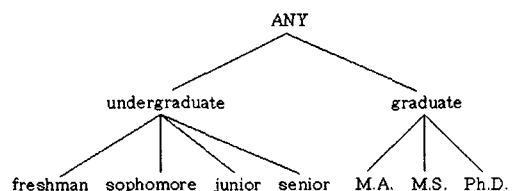


그림 1. 속성 Status의 개념 트리

로서 속성값 {freshman, sophomore, junior, senior}, {M.A., M.S., Ph.D.}은 최하위 노드로 표현되고 상위 개념인 undergraduate, graduate는 상위노드로 표현된다. 또한 일반화의 최상위 단계, 즉 최상위 노드는 예약어 "ANY"로 표현된다.

뷰(view)의 논리적인 관점에서 볼 때, 릴레이션에 포함되는 각각의 튜플은 논리곱(conjunctive) 형태의 논리식으로 표현된다. 다시 릴레이션은 이와 같은 논리곱 형태의 논리식들이 논리합(disjunctive)의 형태로 결합된 하나의 집합으로 표현된다. 그러므로 규칙 추출을 위한 데이터와 추출된 규칙 모두 관계식의 형태(relational form) 또는 일차술어해석(first-order predicate calculus)으로 나타낼 수 있다.

(2) AOI의 과정

AOI의 과정은 다음 일련의 단계를 통해 수행된다. 대학원생들의 패턴을 파악하는 문제를 예로 수행 과정을 기술하도록 한다. 표 2는 university 데이터베이스에 포함된 student라는 릴레이션을 나타낸다. 규칙 추출의 목적이 대학원생들의 특징을 파악하는데 있으므로 표 1로 표현된 개념 계층을 이용하여 속성 Name, Major, Birth_Place, Grade를 포함한 대학원생들의 데이터만을 수집하도록 한다.

구하고자 하는 대학원생, 즉 "graduate"가 속성 Status의 개념 계층에서는 최하위 계층이 아니므로 최하위 계층에 속하는 속성값 {M.S., M.A., Ph.D.}를 가진 튜플들을 선택한다. 속성 Status는 이미

표 3. AOI를 위한 초기 릴레이션

Name	Status	Major	Birth_Place	Grade
Anderson	history	Vancouver	3.5	1
Gupta	math	Bombay	3.3	1
Liu	biology	Shanghai	3.4	1
...
Monk	computing	Victoria	3.8	1

"graduate"라는 제약조건을 통해 더 이상 일반화가 되지 않으므로 이를 제외한 속성 Name, Major, Birth_Place, Grade를 포함하도록 한다. 그 결과 표 3과 같은 초기 릴레이션(initial relation)을 얻을 수 있다. 초기 릴레이션에는 Vote라는 특별한 속성이 추가되는데 각 튜플에 대하여 1로 초기값이 설정된다.

일반화는 정의된 개념 계층을 이용하여 속성을 상위 단계의 개념으로 일반화하고 이와 동시에 속성 제거와 튜플의 통합을 통해 초기 릴레이션을 최소화시키는 단계이다. 현재 해당 속성의 값이 상위 단계로 일반화가 가능하다면 상위 계층의 값으로 대체한다. 속성을 상위 단계로 일반화하는 과정에서 해당 속성이 갖는 값들이 동일한 튜플은 통합이 되며 통합된 수만큼 속성 Vote의 값이 증가한다. 해당 속성이 다양한 값을 갖고 있음에도 더 이상 상위 단계로 일반화가 되지 않을 경우에는 속성을 제거한다. 표 4는 일반화 과정의 알고리즘으로 상위 개념으로의 일반화와 튜플의 통합, 속성의 제거를 반복적으로 수행하는 과정을 나타내고

표 2. student 릴레이션

Name	Status	Major	Birth_Place	Grade
Anderson	M.A.	history	Vancouver	3.5
Bach	junior	math	Calgary	3.7
Carlton	junior	liberal arts	Edmonton	2.6
Gupta	Ph.D.	math	Bombay	3.3
Hart	sophomore	chemistry	Richmond	2.7
Jackson	senior	computing	Victoria	3.5
Liu	Ph.D.	biology	Shanghai	3.4
...
Meyer	sophomore	music	Burnaby	3.0
Monk	Ph.D.	computing	Victoria	3.8
Wise	freshman	literature	Toronto	3.9

표 4. 일반화 과정의 알고리즘

```

begin {basic attribute-oriented induction}
  for each attribute Ai (1 ≤ i ≤ n, where n=# of attribute)
    in the generalized relation GR do
      while #_of_distinct_values_in_Ai > threshold do {
        if no higher level concept in the concept
          hierarchy table for Ai
        then remove Ai
        else substitute for the values of the Ai's
          by its corresponding minimal generalized concept;
          merge identical tuples }
      while #_of_tuples in GR > threshold do {
        selectively generalized attributes;
        merge identical tuples }
  end
  
```

있다. 첫 번째 while-loop의 조건부에서는 속성 A_i 의 값의 개수를 임계치(threshold)와 비교한다. 속성 값의 개수가 임계치보다 큰 경우, 즉 일반화가 가능한 경우 개념 계층과 비교하여 상위 계층이 존재하지 않으면(then 부분) 속성 A_i 를 제거하고 상위 계층이 존재하면(else 부분) 상위 계층의 값으로 대체하고 튜플을 통합한다. 모든 속성에 대하여 같은 과정을 반복하여 수행한다. 최종적으로 통합된 튜플의 개수를 설정한 임계치와 비교하여 임계치보다 클 경우 튜플의 통합과정을 추가로 수행한다. 알고리즘의 수행을 통해 초기 릴레이션은 상위 개념을 갖고 있지 않은 속성 Name을 제거하고 개념 계층을 이용하여 속성 Major, Birth_Place, Grade를 일반화한다.

표 5는 초기 릴레이션에 알고리즘을 적용하여 속성 Major, Birth_Place, Grade를 일반화한 것이다. 일반화가 불가능한 속성 Name을 제거한 릴레이션에서 속성 Major, Birth_Place, Grade의 값을 각각의 상위 계층인 {art, science}, {B.C., Ontario, India, China, ...}, {poor, average, good, excellent}로 대체한다. 각 튜플의 속성 Major, Birth_Place, Grade의 값을 비교하여 동일한 경우 통합하고 속성 Vote의 값은 각각의 튜플이 가진 값의 합으로 대체한다. 속성 Birth_Place의 경우 상위 계층이 존재하므로 상위 계층의 값인 {foreign, Canada}로 대체하여 일반화한다.

서로 다른 튜플 간에 하나의 속성을 제외하고 다른 모든 속성(속성 Vote는 제외)의 값이 같으며 그 하나의 속성을 집합으로 표현이 가능한 경우가 있다고 하자. 표 5의 경우 첫 번째와 두 번째 튜플의 속성 Major 값을 예로 들면 표 6의 {art, science}와 같이 하나의 집합으로 표현이 가능하다. 개념 계층을 적용하면 이 값은 "ANY"와 같으므로 논리식으로 표현시에는 생략된다.

최종적으로 일반화가 완료된 릴레이션은 하나의

표 5. 일반화된 릴레이션

Major	Birth_Place	Grade	Vote
art	Canada	excellent	35
science	Canada	excellent	40
science	foreign	good	25

표 6. 단순화된 릴레이션

Major	Birth_Place	Grade	Vote
{art, science}	Canada	excellent	75
science	foreign	good	25

튜플을 가질 경우 논리곱 형태의 논리식으로 변환된다. 다수의 튜플들인 경우에는 논리합을 이용하여 결합된 형태로 변환된다. 다음의 식은 단순화 단계까지 완료된 릴레이션을 논리식의 형태로 변환한 것이다. 표 6의 첫 번째와 두 번째 튜플은 각각 논리곱의 형태로 표현하며 이 두 식을 논리합의 형태로 결합하였다. 논리식은 초기 릴레이션으로 요약으로서 대학원생 중 75%가 Canada 출신으로 성적은 excellent이며 나머지 25%는 science를 전공하는 외국학생으로 성적은 good임을 나타내고 있다.

$$\forall (X)graduate \rightarrow (Birth_Place(x) \in Canada \wedge Grade(x) \in excellent)[75\%] \vee (Major(x) \in science \wedge Birth_Place(x) \in foreign \wedge Grade(x) \in good)[25\%]$$

3. 공간 연관규칙 추출

1) 데이터 구축

본 연구에서는 공간 연관규칙 추출을 물리적인 도시지역의 확장이라는 문제에 적용하고 수도권 내에 위치한 대도시 중 경기도 수원시를 연구의 공간적인 범위로 설정하였다. 또한 모델을 단순화하기 위하여 구축할 데이터의 항목을 도시지역, 도로, 배제지역(수계와 개발제한구역 포함), 경사도의 네 가지 항목으로 설정하였다.

우선 대상지역에 대하여 활용 가능한 다양한 축척의 지형도 및 주제도, 수치지도를 수집하였다. 수집된 여러 유형의 공간 데이터를 대상지역에 대하여 전체 시기(약 30년간)에 걸쳐 일관성을 유지하고 연속적인 형태의 데이터로 구축하였다. 항목별 공간 데이터의 입력은 종이 형태로 발간된 지형도와 주제도의 경우에 수동적인 디지털라이징(digitizing) 방법과 스캐닝(scanning) 및 벡터화 방법(vectorizing)의 병행에 의해 이루어졌으며, 수치지형도와 기타 주제도는 데이터의 변환과정을 통

해 수행하였다. 연구에서 수집된 데이터의 주요 내용은 표 7과 같으며 대상 시기는 1960년, 1970년, 1980년, 1990년 총 네 시기이다.

농림지역과 기타지역을 비도시화 지역으로 통합하였고 교통망은 도시화에 실질적인 영향을 미치는 2차선 이상의 도로를 이용하고 철도와 고속도로는 제외하였다. 수계망은 폴리곤 형태와 라인 형태의 데이터에서 폴리곤 형태의 데이터만 이용하였다. 개발제한구역은 1970년대부터 대도시의 시가지가 무제한적으로 확산되는 것을 방지하기 위해 설치된 지역이므로 1970년대부터 데이터를 구축하였다. 경사도는 수치지형도에서 등고선 레이어만 추출한 후 TIN(Triangular Irregular Network)으로 보간하고 퍼센트 형식의 DEM(Digital Elevation Model)으로 변형하여 이용하였다.

CA 모델에 적용하기 위해서 구축된 데이터를 항목별로 격자 형태의 데이터(grid data)로 변환하였다. 격자의 크기는 위성 영상의 크기와 급격한 토지이용 변화를 표현할 수 있는 최대의 크기인 100m로 설정하였다. 그림 2는 대상지역인 수원시의 1990년대 도시지역, 도로, 경사도, 배제지역 데이터로서 격자 형태의 데이터로 최종 변환하였다. 데이터 구축과 변환 과정은 ESRI사의 ArcCatalog와 ArcMap, ArcToolbox를 이용하였다.

2) 공간 연관규칙 추출 과정

(1) 공간연산 적용

GIS의 중요한 기능 중의 하나는 새로운 패턴을 인식하는 것이다(김계현, 2000). 본 연구에서 추출하고자 하는 규칙은 일정 크기와 모양을 가진 지역 내에서 발견되는 국지적인(local) 패턴을 의미한다. 공간적으로 인접한 주변 셀에 바탕을 둔 근린

(focal) 연산과도 매우 유사하다.

도시지역의 확장과 관련된 규칙을 찾기 위해 우선적으로 네 시기의 도시지역 레이어를 중첩하여 각 시기별(1960-1970, 1970-1980, 1980-1990)로 신규 생성된 도시셀들을 추출하였다. 그림 3은 1970년대 신규 생성된 도시셀을 추출하는 과정이다. 1970년대 도시지역과 1960년대 도시지역의 중첩연산(-)



그림 3. 시기별 신규 도시셀 추출

표 7. 연구에서 수집된 데이터의 종류와 시기

구분	세부내용	시기
지형도	1:50,000 지형도	1960, 1970, 1980, 1990년대
	1:250,000 지세도	1961, 1965, 1973, 1976, 1986, 1991년
주제도	1:50,000 개발제한 구역도	1972, 1981, 1995년
	1:25,000 토지이용현황도	1974년
	1:50,000 도시계획 총괄도	1994년
	1:200,000 도로망도	1997년
수치 지형도	1:5,000 수치지형도	1998년
	1:25,000 수치지형도	
수치 데이터	경기도 전역의 DEM (30M)	1990년대
위성	LANDSAT MSS	1972, 1979, 1981년
영상	LANDSAT TM	1985, 1988, 1992, 1996, 2000년

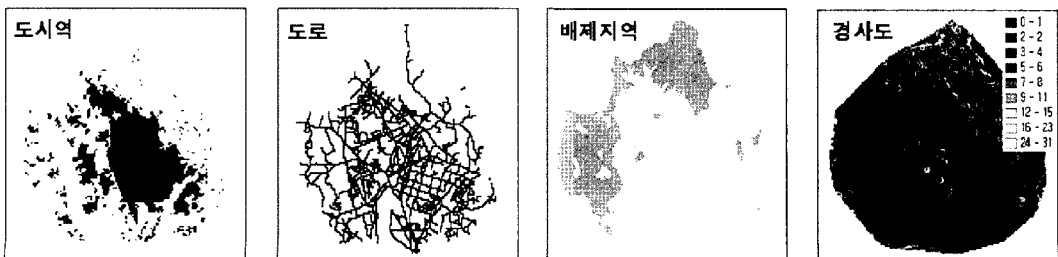


그림 2. 구축된 데이터 - 1990년대 수원시

을 통해 1970년대 신규 생성된 도시지역 레이어를 생성하였다. 1980년대와 1990년대 생성된 신규 도시지역 레이어 또한 동일한 방법을 이용하여 생성하였다.

각 시기별 신규 도시셀들은 이전 시기에서 비도시화 지역에 속하므로 이전과 당해년대의 도시지역 및 도로 레이어와 중첩하여 주변지역의 상태를 조사하였다. 비도시셀 주변에 위치한 도시지역 및 도로는 향후 도시셀로 변환하는데 있어 직접적인 영향을 미치는 요인이며 공간 연관규칙의 핵심이다. 1970년대의 경우 신규 도시지역 레이어와 1960년대 도시지역과 도로, 1970년대 도시지역과 도로 레이어를 각각 중첩하여 주변도시셀 및 도로셀 레이어를 생성한다. 즉 하나의 신규 도시지역 셀을 중심으로 두 시기의 주변도시셀과 도로셀의 개수를 구하는 것이다. 하나의 신규 도시지역 셀에 대하여 두 시기의 주변셀들을 파악하는 것은 이전 시기만을 참조할 경우 생기는 초기조건에 의한 영향을 최소화하기 위한 것이다. 해당 셀이 비도시지역인 이전년대에는 주변셀이 존재하지 않으므로 주변에 의한 영향을 파악할 수 없다.

비도시화 지역은 주변에 위치한 기존의 도시지역과 도로셀에 의해 영향을 받아 도시로 변환된다. 일반적으로 CA에서는 대상 셀에 영향을 미치는 주변의 셀들을 가리켜 인접(neighborhood)이라 하며, Moore의 인접 개념은 GIS에 사용되는 공간필터의 마스크나 윈도우와 개념적으로 매우 유사하다(박수홍, 1997). 많은 경우에 규칙적인 주변지역의 설정보다는 불규칙적이고 다양한 형태의 경계 설정이 요구되나 본 연구에서는 일정한 공간상의 패턴을 추출하기 위해 동일한 형태와 크기의 윈도우를 적용하였다(그림 4).

주변도시지역과 도로에 대한 분석결과는 각각 레이어로 생성되며 배제지역 및 경사도 레이어와의 중첩을 통하여 도시로의 성장이 가능한 셀의 후보를 선정하게 된다. 또한 각 시기별로 선정된 셀들을 하나로 통합하여 대상지역의 시공간 데이터를 하나의 집합 또는 테이블의 형태로 표현하였다.

(2) 통계적 분석

통계적 분석은 AOI를 적용하기에 앞서 제약조건과 임계치를 설정하기 위한 과정이다. 제약조건은 규칙추출에 필요한 데이터를 선택하는데 사용되며, 임계치는 일반화하기에 적절한 속성값을 가진 튜플을 선택하거나 튜플의 통합 정도를 결정하기 위한 기준으로 사용된다. 본 과정에서는 도시지역, 도로, 경사도, 배제지역의 데이터 중 경사도의 제약조건을 설정하였다.

경사도의 경우 튜플을 선정하기 위한 제약조건으로써 사용되며 향후 도시지역으로 확장이 가능한 비도시셀의 허용 경사도를 의미한다. 경사도의 통계치를 분석한 결과 대부분의 값이 평균값 2.36을 중심으로 3배수 표준편차(3σ, 약 99.74%) 범위 이내에 위치하고 있으므로 경사도의 제약조건을 11이하로 결정하였다(그림 5).

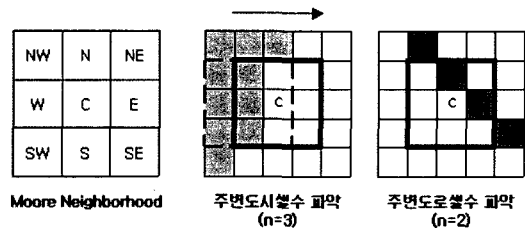


그림 4. Moving Window를 이용한 주변지역 분석

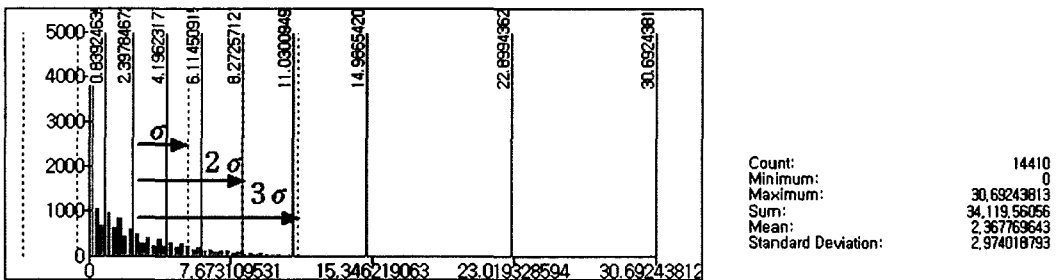


그림 5. 입력데이터가 갖는 경사도의 통계치

(3) AOI를 이용한 규칙 추출

공간연산을 적용한 결과 기존의 입력 데이터인 도시지역, 도로, 배제지역, 경사도 외에 신규 생성 도시지역과 주변도시셀 수, 주변도로셀 수 레이어가 추가되었다. 동일한 위치에 있는 셀을 중심으로 표현하면 각각의 레이어는 해당 셀에 대한 속성이며 레이어가 갖는 값들은 속성값에 해당된다.

릴레이션 구성 후, AOI의 적용을 위하여 우선 배경지식의 정의와 관련 데이터를 수집하여 초기 릴레이션을 구성하는 단계를 수행하였다. 배경지식의 경우 데이터의 특성에 따라 개념 계층의 정의가 달리 이루어지므로 도시지역의 확장이라는 문제에 맞추어 정의하였다. 문제를 "도시로 성장한 비도시셀의 변화규칙 파악"이라는 구체적인 내용으로 설정하고 속성의 특성(표 8)에 맞추어 개념 계층을 표 9와 같이 정의하였다. 주변도로셀 수의 경우 속성의 특성 분류시 일반화 대상으로 분류되었지만 도시지역 확장의 유형으로 볼 때 도로에 의한 성장은 도시지역에 부가적으로 일어난다는

표 8. 속성의 특성 분류

속성	특성	AOI 수행시 적용방법
도시지역	비도시지역만 필요	도시지역에 해당하는 튜플 삭제 후 속성 제거
도로	불필요	속성 제거
배제지역	제약조건	설정된 임계치를 적용하여 튜플 삭제
경사도	제약조건	설정된 임계치를 적용하여 튜플 삭제
신규 생성 도시지역	신규 생성 도시지역만 필요	신규 생성 도시지역이 아닌 튜플 삭제 후 속성 제거
{주변도시셀 수, 주변도로셀 수}	일반화 대상	개념 계층을 적용하여 일반화 단계 수행

표 9. 개념 계층 정의

{ 0, 1, 2 } ⊂ unchangeable
{ 3, 4, 5, 6, 7, 8 } ⊂ changeable
{ unchangeable, changeable } ⊂ ANY (주변도시셀 수)
{ 0 - 11 } ⊂ changeable
{ 12 - 100 } ⊂ unchangeable
{ unchangeable, changeable } ⊂ ANY (경사도)

특징을 가지고 있다. 주변도로셀 수는 주변도시셀 수와의 조합을 통해 적용하는 것을 원칙으로 하여 개념 계층의 정의에서는 제외하였다.

초기 릴레이션은 도시지역, 도로 속성에 대한 제약조건을 통해 튜플을 삭제하고 속성을 제거하여 표 10과 같이 초기화되었다. 다시 배제지역, 경사도, 신규 생성 도시지역의 제약조건 적용을 통해 튜플을 삭제하고 주변도시셀 수와 주변도로셀 수의 조합별로 튜플을 통합하면 일반화와 단순화가 완료된 릴레이션이 생성된다(표 11).

최종결과로 나온 릴레이션은 다음과 같이 논리식으로의 표현이 가능하다. 각각의 튜플은 논리곱의 형태로 각 속성을 결합하여 표현하였으며 릴레이션은 논리합의 형태로 튜플을 결합하여 표현하고 있다.

대용량의 데이터베이스에서는 많은 연관 관계들이 존재하지만 일부는 거의 발생하지 않거나 의미가 없는 경우도 있다. 빈번하게 발생하는 패턴에 초점을 맞추기 위해 지지도(support)와 신뢰도(confidence) 개념을 사용하였다. 즉 발견된 규칙들은 정의된 지지도와 신뢰도를 최소한으로 만족하는 값들에서 선택하였다. 지지도는 A→B로의 규칙이 전체 튜플에서 만족된 비율을 말하며 신뢰도는 A가 발생했을 때를 전제로 A→B로의 규칙이 만족하는 비율을 의미한다(조영아, 1999; Han and Kamber, 2001).

$$\text{지지도}(A \rightarrow B) = \frac{A \text{와 } B \text{를 모두 포함한 튜플}}{\text{전체 튜플}}$$

$$\text{신뢰도}(A \rightarrow B) = \frac{A \text{와 } B \text{를 모두 포함한 튜플}}{A \text{를 포함한 튜플}}$$

표 10. 초기 릴레이션

배제지역	경사도	신규 생성 도시지역	주변도시셀 수	주변도로셀 수	Count
Y	3	Y	1	0	1
N	1	Y	3	0	1
N	1	Y	3	0	1
N	2	N	3	2	1
...
N	2	Y	7	2	1
N	1	Y	8	8	1

튜플의 개수 : 6864

표 11. 단순화된 릴레이션

주변도시셀 수	주변도로셀 수	Count
8	4	274
8	3	268
3	0	237
4	0	129
5	0	98
4	4	90
4	3	80
5	4	71
3	{1,2,3,4,5,6,7,8}	346
4	{1,2,5,6,7,8}	176
5	{1,2,3,5,6,7,8}	237
6	{0,1,2,3,4,5,6,7,8}	286
7	{0,1,2,3,4,5,6,7,8}	300
8	{1,2,3,4,7,8}	770

튜플의 개수 : 14

- V(X)비도시셀 → (주변도시셀 수(x) ∈ 3 ∧ 주변도로셀수(x) ∈ 0) [3.4%]
- V(주변도시셀 수(x) ∈ 4 ∧ 주변도로셀수(x) ∈ 0) [1.9%]
- V(주변도시셀 수(x) ∈ 4 ∧ 주변도로셀수(x) ∈ 3) [1.2%]
- V(주변도시셀 수(x) ∈ 4 ∧ 주변도로셀수(x) ∈ 4) [1.3%]
- V(주변도시셀 수(x) ∈ 5 ∧ 주변도로셀수(x) ∈ 0) [1.4%]
- V(주변도시셀 수(x) ∈ 5 ∧ 주변도로셀수(x) ∈ 4) [1.0%]
- V(주변도시셀 수(x) ∈ 8 ∧ 주변도로셀수(x) ∈ 3) [3.9%]
- V(주변도시셀 수(x) ∈ 8 ∧ 주변도로셀수(x) ∈ 4) [4.0%]
- V(주변도시셀 수(x) ∈ 3 ∧ 주변도로셀수(x) ∈ {1,2,3,4,5,6,7,8}) [5.0%]
- V(주변도시셀 수(x) ∈ 4 ∧ 주변도로셀수(x) ∈ {1,2,5,6,7,8}) [2.6%]
- V(주변도시셀 수(x) ∈ 5 ∧ 주변도로셀수(x) ∈ {1,2,3,5,6,7,8}) [3.5%]
- V(주변도시셀 수(x) ∈ 6) [4.2%]
- V(주변도시셀 수(x) ∈ 7) [4.4%]
- V(주변도시셀 수(x) ∈ 3 ∧ 주변도로셀수(x) ∈ {1,2,5,6,7,8}) [11.2%]

표 12는 각 튜플이 갖는 지지도로서 각 튜플의 속성 Count의 값을 Count의 총합인 6864로 나눈 결과이다. 일반적으로 규칙으로 선정되기 위해서는 높은 지지도의 값을 나타내어야 하는데 주변도시셀 수와 주변도로셀 수에 대한 조합이 다양하므로 결과적으로 튜플의 통합이 완전하게 이루어지지 못하여 그 값이 높지 않음을 볼 수 있다. 이는 주변도시셀과 도로셀 수의 영향은 속성 Count를 통해 각각이 비도시셀에 미치는 영향의 정도로만 파악할 수 있을 뿐 개별적인 현상 또는 패턴을 일반화할 경우 정보의 손실이 크다는 한계를 가지고

표 12. 각 조합이 갖는 지지도 (단위: %)

도로 도시	0	1	2	3	4	5	6	7	8
3	3.45	0.55	0.74	1.39	1.22	0.71	0.34	0.29	0.10
4	1.87	0.46	0.58	1.16	1.31	0.58	0.53	0.24	0.14
5	1.42	0.37	0.43	0.85	1.03	0.75	0.53	0.29	0.18
6	0.85	0.32	0.20	0.50	0.71	0.53	0.49	0.32	0.20
7	0.69	0.21	0.17	0.67	0.72	0.59	0.78	0.36	0.13
8	2.69	1.57	1.64	3.90	3.99	3.27	2.81	1.26	0.64

표 13. 각 조합이 갖는 신뢰도 (단위: %)

도로 도시	0	1	2	3	4	5	6	7	8
3	39.10	6.27	8.41	15.84	13.86	8.08	3.96	3.30	1.15
4	27.15	6.73	8.42	16.84	18.94	8.42	7.78	3.57	2.10
5	24.13	6.40	7.38	14.53	17.48	12.80	9.11	4.92	3.20
6	20.62	7.69	4.89	12.23	17.13	12.93	11.88	7.69	4.89
7	16.00	5.00	4.00	15.33	16.66	13.66	18.00	8.33	3.00
8	12.35	7.21	7.54	17.90	18.30	15.03	12.89	5.81	2.93

있기 때문이다.

의미있는 규칙을 발견하기 위해서는 사용자에 의한 임계값 설정이 요구되는데 알고자 하는 규칙의 정도에 따라 임계값은 임의적일 수 있다. 가장 적절한 규칙이라 함은 도시지역의 확장을 정확도 평가 기준에 따라 가장 높은 값으로 나타내는 것을 의미한다. 우선 지지도의 임계치를 1%, 신뢰도의 임계치를 16%으로 하였으며 임계치를 만족하지 못하는 조합의 경우 각 행별로 평균을 구하였다(표 13). 일반적으로 AOI 과정에서는 최종적으로 일반화된 튜플의 개수를 2-8개 수준으로 허용하고 있다(Han and Kamber, 2001). 임계치의 설정은 다양한 지역 상태를 반영하기 위하여 최대 8개의 튜플로 요약될 수 있는 값으로 설정하였다.

이미 릴레이션의 일반화를 통해 규칙을 추출하였으나 지지도와 신뢰도를 각각의 튜플에 대하여 계산한 것은 규칙이 갖는 확률의 임계값을 설정하기 위해서이다. 본 연구에서 사용한 CA 모형은 확률모형에 근거하고 있으며, 주변도로가 주변도시에 대하여 종속적인 점을 감안하여 주변도시셀 수에 대한 주변도로셀 수의 조건부 확률을 임계값으로 사용하였다.

표 14는 최종적으로 추출된 연관규칙의 알고리즘으로서 제약조건과 변화규칙으로 구성되었다. 알고리즘으로 표현된 공간 연관규칙은 모든 셀에 대하여 동일하게 적용되는 것을 원칙으로 한다. 규칙의 개략적인 내용은 다음과 같다. 우선 대상셀이 비도시셀인지 여부를 검사하고 허용 경사도와 배제지역 적용 여부를 검사한다. 모든 제약조건을 만족하는 비도시셀에 대하여 주변도시셀과 주변도로셀을 파악하여 각 조건에 해당하는 확률값의 임계치와 난수와의 비교를 통해 도시셀로의 변환 여부를 결정한다. 모든 셀에 대하여 규칙 적용이 완료되면 특정 시간 t에 대한 규칙 적용이 끝나게 되며

표 14. 추출된 공간 연관규칙의 알고리즘

```

begin { 공간 연관규칙(도시성장규칙) }
// (1<= i <= m(n), m(n)=대상지역 셀 전체의 행(열))
for each 셀i, j
  If 대상셀 = 도시화가 가능한 비도시셀 then
    If 경사도 >=0 & 경사도 <= 11 then
      If not 배제지역 then

        // 임계치를 만족하는 조합들의 confidence 이용
        // u:도시셀 수, r:도로셀 수, md:난수
        규칙 1: (u=3 & r=0) & Int(39,10) >= md then
          셀i, j=도시
        규칙 2: (u=4 & r=0) & Int(27,15) >= md then
          셀i, j=도시
        규칙 3: (u=4 & r=3) & Int(16,84) >= md then
          셀i, j=도시
        규칙 4: (u=4 & r=4) & Int(18,94) >= md then
          셀i, j=도시
        규칙 5: (u=5 & r=0) & Int(24,13) >= md then
          셀i, j=도시
        규칙 6: (u=5 & r=4) & Int(17,48) >= md then
          셀i, j=도시
        규칙 7: (u=8 & r=3) & Int(17,90) >= md then
          셀i, j=도시
        규칙 8: (u=8 & r=4) & Int(18,30) >= md then
          셀i, j=도시

        // 임계치를 만족하지 못하는 조합의 경우
        for each #_of_ 주변도시셀
          If  $\sum(\text{confidence} * \text{support}) / \sum \text{support} >= \text{난수}$ 
            then 셀ij 도시셀
end

```

변환을 마친 데이터는 다시 시간 t+1의 입력데이터로 사용된다.

4. 모형구현을 통한 검증 및 분석

1) 모형구현과 시뮬레이션 수행

추출된 공간연관 규칙은 CA-GIS 통합시스템을 이용하여 모형으로 구현하였다. 최종적으로 추출된 규칙(표 14)은 Cellang의 문법에 맞추어 스크립트로 작성하였다. 규칙의 정확도를 평가할 목적으로 1960년대 데이터를 이용하여 1990년까지 도시지역 확장에 대한 시뮬레이션을 수행하였다. 입력데이터로는 도시지역, 도로, 경사도, 배제지역을 사용하였으며 변화규칙을 적용하여 조건을 만족하는 비도시셀을 도시셀로 변환하였다. 시뮬레이션 1회는 1960년대부터 1990년까지의 변화에 해당하며 1년당 1회의 비율로 반복주기를 적용하여 시뮬레이션 1회당 30회의 반복을 수행하였다. CA는 난수를 이용한 확률(stochastic) 모형이므로 난수로 인한 영향을 최소화하기 위하여 시뮬레이션을 100회 반복하였다. 시뮬레이션 결과로는 규칙에 의해 확장된 도시지역의 이미지가 생성되며 입력데이터와 동일한 포맷으로 되어있다. 각각의 시뮬레이션 결과에 대하여 정확도를 평가하고 이 값들을 평균하여 사용하였다.

2) 정확도 평가

본 연구에서는 모형을 통해 도시지역 확장을 분석한 결과의 정확도 평가를 위해 Lee-Sallee shape index를 도입하였다. Lee-Sallee shape index는 과거 자료로 입력된 도시지역과 모델에서 예측한 도시지역의 레이어를 중첩하고 도시지역의 합집합과 교집합을 셀 단위로 계산하여 합집합으로 교집합을 나누어 계산한다. 과거자료로 입력된 값과 모델에서 예측된 값이 완전히 일치되기 위해서는 Lee-Sallee 측정값이 1이 되어야 하며 값이 0에 가까울수록 입력된 값과 예측치 간의 일치도가 낮음을 나타낸다(Clarke, et al, 1996; 강영욱·박수홍, 2000).

$$\text{Lee - Sallee shape index} = \frac{A \cap B}{A \cup B}$$

3) UGM(Urban Growth Model)과의 비교 · 분석

(1) UGM 개요

USGS가 수행한 Gigalopolis 프로젝트의 핵심이 되는 UGM은 UC Santa Barbara의 Keith C. Clarke 교수에 의해 개발되었다. UGM은 과거로부터 데이터를 입력한 시점까지의 도시지역의 성장을 규칙화하고 이를 토대로 향후 도시지역의 확대를 공간적으로 예측하는 모델로서 래스터 형태의 데이터를 사용하며 셀룰라 오토마타에 기초를 둔 모델이다.

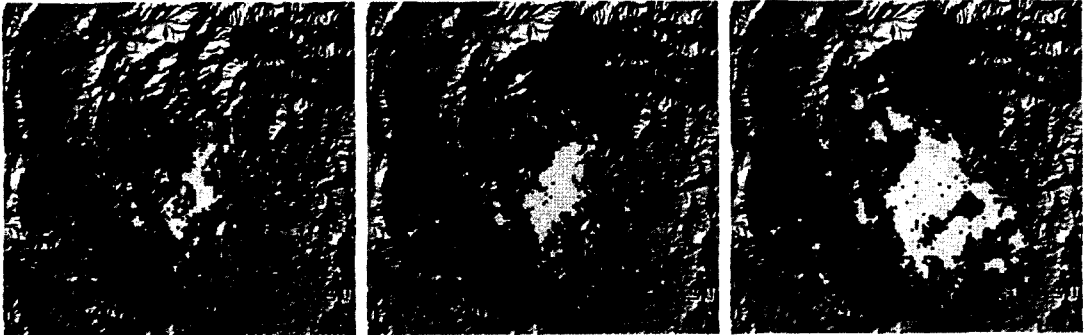
UGM에서의 도시성장은 도시지역, 즉 인공구조물에 의해 토지피복이 형성된 지역의 물리적 확장이라는 측면에서 본 연구와 맥락을 같이한다. 도시성장은 시뮬레이션하려는 지역을 격자단위의 셀로 나누며, 도시가 성장하는 규칙은 셀 단위의 지리적 공간에 동일적으로 적용된다. 도시성장을 분석하거나 예측하기 위해, 즉 이 모델을 수행하기 위해 필요한 데이터는 도시지역, 도로, 경사도, 배제지역으

로 본 연구에서 사용되는 입력데이터와 동일하다. 기본적으로 미래의 도시성장에 대한 예측은 과거에서 현재까지의 도시성장이 이루어진 자료와 모델수행에 입력된 자료를 비교하여 모델 수행에 필요한 입력계수들의 값을 정한 후, 이 값을 계수로서 입력하여 수행한다. UGM 모델링에서는 모델의 보정이 잘 되었는가를 판단하기 위한 방법으로 이미 알려진 현재를 예측하기 위해 과거의 공간자료 및 통계치를 사용하고 있으며, Lee-Sallee shape index를 이용하여 정확도를 평가한다.

(2) 결과 비교 · 분석

앞서 언급한 Lee-Sallee shape index를 이용하여 본 연구결과와 UGM 결과를 비교 · 분석하였다. 시뮬레이션 결과 중 실제 데이터와의 비교가 가능한 1970, 1980, 1990년대의 결과를 비교하였으며 결과 이미지는 그림 6과 같다. 우선 성장 정도를 비교하면 1970년대부터 두 결과 간의 뚜렷한 차이를 볼 수 있다. UGM의 경우 해당시기의 실제 도시지역

본 연구결과



UGM 결과



■ 기준년도 도시역 ■ 시뮬레이션 도시역 두 도시역의 공동부분 ■ 배제지역

그림 6. 시기별 도시지역 확장 분석결과 비교

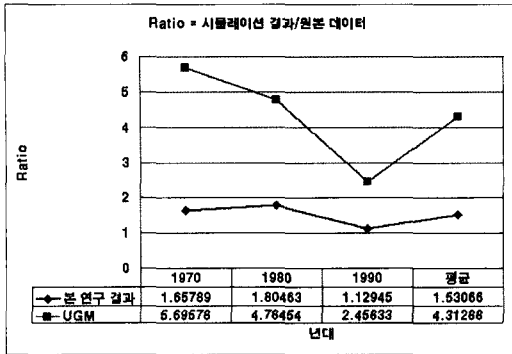


그림 7. 시기별 Ratio 비교

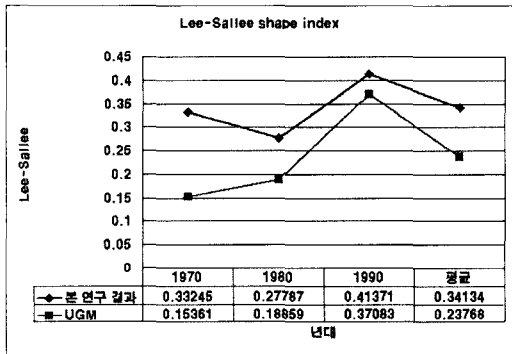


그림 8. 시기별 Lee-Sallee shape index 비교

과 비교하여 확장이 과도하게 이루어지고 있음을 볼 수 있다. 그림 7의 Ratio(시뮬레이션된 도시지역 셀 수/실제 도시지역셀 수)값을 보면 1970년대 최대 5.6배, 평균 4.3배의 과성장을 보이고 있다. 본 연구결과와 경우 1980년대 최대 1.8배, 평균 1.5배로 양적인 비교에 있어서는 상대적으로 실제 데이터와 가까운 비율을 보이고 있음을 알 수 있다. 형태의 일치도를 평가하는 Lee-Sallee shape index를 적용한 결과 또한 두 결과 간에 차이점을 볼 수 있다. UGM의 경우 성장이 과도하게 이루어진 1970년대의 값은 0.15로 가장 낮고 평균적인 값에 있어서도 0.23 정도이다. 본 연구결과 역시 과성장이 가장 심한 1980년대의 값이 0.27로 가장 낮고 평균값은 0.34를 나타낸다(그림 8).

UGM은 본 연구와 도시의 물리적 확장에 초점을 맞추었다는 점, 입력 데이터가 동일하다는 점에서 공통점을 갖는다. 하지만 UGM이 대상지역 또는 데이터에 상관없이 범용적으로 사용할 수 있는

정형적인 규칙 또는 모형을 가지고 최적화 과정을 수행하는 경우라면, 본 연구는 대상지역의 데이터에서 규칙을 직접 추출해내고 이에 따른 유연성 있는 규칙을 적용한다는 점에서 상대적인 차이점을 가지고 있다.

5. 결론

본 연구는 CA 모델링을 위한 공간 연관규칙 추출을 주목적으로 하였으며 래스터 GIS에서 사용 가능한 공간연산 기법들과 지식 발견 기법의 하나인 AOI를 이용하여 CA를 기반으로 적용 가능한 규칙들을 추출하였다. 추출된 규칙은 도시지역의 물리적 확장이라는 사례에 적용하였는데, 실제 경기도 수원시를 대상으로 1960년대부터 1990년대까지의 도시지역 확장을 시뮬레이션하였다. 또한 UGM 결과와의 비교·분석을 통해 추출된 규칙의 적용 가능성을 제시할 수 있었다.

제시된 공간 연관규칙 추출 방법론은 모델링의 핵심이 되는 변화규칙을 주어진 데이터를 이용하여 스스로 정의할 수 있도록 하였다. 추출된 규칙을 이용한 시뮬레이션 결과와 타 모형의 결과를 비교하였을 때, 정확도 평가 기준인 Lee-Sallee shape index의 값이 최대 0.41, 평균 0.34로 비교대상인 UGM의 최대 0.37, 평균 0.23보다 상대적으로 높은 값을 나타내었다. 또한 실제 데이터와의 양적인 비율을 비교하였을 경우 평균 1:1.53인 반면, UGM 결과는 평균 1:4.31의 과성장을 나타내었다. 이는 연구에서 제시한 평가 기준에 근거하였을 경우 상대적으로 실제 현상을 잘 반영하고 있으며 실제 현상에 대하여 적용 가능함을 확인할 수 있었다.

하지만 다음 사항들과 같이 보완할 점이 요구된다. 첫째, 규칙을 추출하는데 사용한 지식 발견 기법은 사용자가 배경 지식 및 임계치에 대한 정의를 해야 하는 일종의 감독분류 기법에 해당된다. 이는 해결하고자 하는 문제에 대한 전문적인 지식을 요구하며 배경지식 및 임계치의 설정에 따라 규칙이 달라질 수 있다는 가능성을 가지고 있다. 둘째, 본 연구에서 적용한 공간연산 기법, 즉 주변셀 분석을 위한 윈도우의 형태 및 크기가 단순하

다는 점이다. CA에서는 모델링 하고자 하는 문제에 따라 다양한 형태 및 크기의 이웃 또는 윈도우를 적용한다. 본 연구의 사례인 도시지역 확장에 있어서도 실제 모든 지역이 동일한 크기의 주변에 의해 영향을 받지 않는다는 것에 대한 고려가 미흡했다. 위의 원인들은 정확도에 영향을 미쳐서 Lee-Sallee shape index가 최대 0.41을 넘지 못하는 결과를 초래한 것으로 판단된다. 이러한 문제를 해결하기 위해 다양한 공간연산 기법의 적용과 배경 지식 및 임계치 설정을 위한 객관적인 기준 마련에 대한 보완 연구가 요구된다.

본 연구에서는 도시지역의 물리적 확장이라는 사례에 국한되어 공간 연관규칙을 추출하였는데 향후 산불확산 모델 및 기타 공간 확산모델 등을 포함한 다양한 동적 공간 모델을 보다 효과적으로 구현하고 실행할 수 있으리라 기대된다.

文 獻

- 강영옥 · 박수홍, 2000, "서울대도시지역 도시성장 예측에 관한 연구," *대한지리학회지*, 35(4), 621-639.
- 김계현, 2000, GIS 개론, 대영사.
- 박수홍 · 주용진 · 신윤희, 2002, "도시성장 예측 모델 개발을 위한 시공간 데이터베이스의 구축," *지리학연구*, 36(4), 313-326.
- 박수홍, 2001, "CA-GIS 통합시스템의 설계와 구현," *한국GIS학회지*, 9(2), 185-206.
- _____, 1997, "CA-GIS 통합시스템을 이용한 GIS 연산의 구현," *한국GIS학회지*, 5(1), 99-113.
- 정재준, 2001, 도시권의 도시성장 분석 및 예측을 위한 셀룰라 오토마타 모델링, 서울대학교 박사학위논문.
- 조영아, 1999, Arc/View를 이용한 광주 · 전남지역 공간 연관 규칙 탐사, 전남대학교 석사학위논문.
- Clarke, K.C., Hoppen, S. and Gaydos, L.J., 1996, Method and techniques for rigorous calibration of a cellular automaton model of urban growth, *Proceedings, Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling*, January 21-25th, Santa Fe, NM.
- Han, J., and Kamber, M., 2001, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers.
- Han, J., Cai, Y. and Cercone, N., 1992, Knowledge discovery in databases: an attribute-oriented approach, *Proceedings of the 18th VLDB Conference*.
- Li, X. and Yeh, A.G., 2000, Modeling sustainable urban development by the integration of constrained cellular automata and GIS, *International Journal of Geographical Information Sciences*, 14(2), 131-152.
- Park, S.H and Wagner, D.F., 1997, Incorporating cellular automata simulators as analytical engines in GIS, *Transactions in GIS*, 2, 213-231.
- Sui, D.Z., 1998, GIS-based urban modeling: practices, problem, and prospects, *International Journal of Geographical Information Sciences*, 12(7), 651-671.
- Tomlin, D., 1990, *Geographic Information Systems and Cartographic Modeling*, Prentice Hall.
- <http://www.suwon.ne.kr>
- <http://staff.vbi.vt.edu>
- <http://www.ncgia.uscb.edu>

최초투고일 04. 02. 19
최종접수일 04. 06. 04