

영향지지도에 의한 모호공간관계규칙발굴방법

김 남 혁

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《과학기술의 종합적발전추세와 사회경제발전의 요구에 맞게 새로운 경계과학을 개척하고 발전시키는데 큰 힘을 넣어야 합니다.》(《조선로동당 제7차대회에서 한 중앙위원회사업총화 보고》 단행본 40페이지)

선행연구[1-3]에서는 공간자료기지의 모든 항목들이 지리적현상에 미치는 영향정도가 동등하다는것을 전제로 하여 항목의 빈발성에 따라 규칙발굴을 진행하였다. 그러나 공간자료기지에서 모든 항목들을 따져보면 빈발성이 낮아도 목적하는 공간관계에 미치는 영향정도가 높은 경우가 있고 반대로 빈발성은 높아도 영향정도가 낮은 경우가 있다.

논문에서는 매 항목들의 영향정도를 고려하여 규칙발굴효과를 높일수 있는 모호공간관계규칙발굴방법에 대하여 서술하였다.

1. 이론적기초

1) 항목의 영향지지도결정

항목들의 영향정도를 정량적으로 평가하기 위하여 계층구조분석(AHP)법을 리용하였다.[2] 영향지지도는 AHP법을 리용하여 지표별, 요인별로 목표에 주는 영향정도를 정량화한 다음 그 결과값들과 해당 항목들의 지지도에 의하여 결정한다.

$$MI = 1/N, IS(k) = S(k) \sum_{k=1}^n I(k) / nMI$$

여기서 MI 는 영향정도의 평균값, $IS(k)$ 는 영향지지도, N 은 항목총개수, n 은 해당업무속의 항목개수, $S(k)$ 는 k 항목의 지지도이다.

2) 영향지지도에 의한 모호공간관계규칙발굴

① 자료기지에서부터 공간자료모임을 얻고 그것을 분석하여 잡음자료를 제거한다.

② 아래의 모호성원함수들을 리용하여 모호공간관계의 리산화를 진행하며 규칙발굴을 위한 업무자료기지를 얻는다.

$$\mu(x) = \begin{cases} 1, & x < x_{\min} \\ \frac{x_{\max} - x}{x_{\max} - x_{\min}}, & x_{\min} \leq x \leq x_{\max} \\ 0, & x > x_{\max} \end{cases}, \mu(x) = \begin{cases} 0, & x < x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}}, & x_{\min} \leq x \leq x_{\text{mid}} \\ \frac{x_{\max} - x}{x_{\max} - x_{\min}}, & x_{\text{mid}} \leq x \leq x_{\max} \\ 0, & x > x_{\max} \end{cases}, \mu(x) = \begin{cases} 0, & x < x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}}, & x_{\min} \leq x \leq x_{\max} \\ 1, & x > x_{\max} \end{cases}$$

③ AHP법을 리용하여 영향지지도를 계산한다.

④ 전통적인 규칙발굴산법인 Apriori방법을 개선하여 규칙발굴을 진행한다.

론문에서는 Apriori방법[3]에서 일반적으로 쓰이는 지지도대신에 영향지지도를 적용하여 지리적현상의 본질이 반영된 효과적인 공간관계규칙들이 탐색되도록 하였다.

2. 모형의 응용과 효과성분석

효과성분석을 위하여 영향정도를 고려하지 않고 발굴연산을 진행한다. 다음 영향정도를 고려하여 발굴연산을 진행하여 두 결과를 비교한다. 실험을 위하여 관광봉사지도우에서 러관들의 분포자료를 계측의 방법으로 얻었다.

표 1은 발굴대상의 공간분포상태에 대한 실험자료의 일부이다.

표 1. 공간자료기지의 일부

ID	...	러관 규모	가격	도로와의 거리관계	하천, 호수와의 거리관계	철도와의 거리관계	바다와의 거리관계
1	...	대	비싸다.	1급도로 0.3km	큰 강 0.1km	1.5km	36km
2	...	중	눅다.	2급도로 7.0km	호수 0.2km	12km	0.1km
3	...	소	비싸다.	1급도로 5.5km	작은 강 2.3km	13km	12km
4	...	대	중간.	1급도로 7.8km	호수 1.6km	2.4km	65.8km
5	...	대	비싸다.	1급도로 1.5km	호수 3.2km	0.6km	11.3km
6	...	중	눅다.	2급도로 2.2km	작은 강 1.3km	9.0km	0.2km
7	...	대	중간.	2급도로 0.3km	작은 강 1.6km	8.0km	80km
8	...	대	비싸다.	2급도로 1.4km	작은 강 1.0km	6.4km	58km

성원함수에 의해 모호공간관계의 리산화를 진행한다.(표 2)

표 2. 리산화된 업무자료기지의 일부

ID	...	러관 규모(s)	가격(p)	도로와의 거리관계(d)	하천, 호수와의 거리관계(r)	철도와의 거리관계(c)	바다와의 거리관계(b)
1	...	대(s_1)	비싸다.(p_1)	$d_1(1.0)$	$r_1(1.0)$	$c_1(1.0)$	$b_4(1.0)$
2	...	중(s_2)	눅다.(p_3)	$d_2(0.25),$ $d_3(0.75)$	$r_1(1.0)$	$c_3(0.5),$ $c_4(0.5)$	$b_1(1.0)$
3	...	소(s_3)	비싸다.(p_1)	$d_2(0.63),$ $d_3(0.37)$	$r_1(0.85),$ $r_2(0.15)$	$c_3(0.37)$ $c_4(0.63)$	$b_3(0.5),$ $b_4(0.5)$
4	...	대(s_1)	중간(p_2)	$d_2(0.05),$ $d_3(0.95)$	$r_1(1.0)$	$c_1(0.8),$ $c_2(0.2)$	$b_4(1.0)$
5	...	대(s_1)	비싸다.(p_1)	$d_1(1.0),$	$r_1(0.4),$ $r_2(0.6)$	$c_1(1.0)$	$b_3(0.59)$ $b_4(0.41)$
6	...	중(s_2)	눅다.(p_3)	$d_1(0.9),$ $d_2(0.1)$	$r_1(1.0)$	$c_3(0.88)$ $c_4(0.12)$	$b_1(1.0)$
7	...	대(s_1)	중간(p_2)	$d_1(1.0)$	$r_1(1.0)$	$c_3(1.0)$	$b_4(1.0)$
8	...	대(s_1)	비싸다.(p_1)	$d_1(1.0)$	$r_1(1.0)$	$c_2(0.4),$ $c_3(0.6)$	$b_4(1.0)$

매 항목들의 영향정도를 평가한다.

목표층은 러관수요(수요자는 항목 p, r, b 에 특별히 흥미를 가진다.), 중간층은 매 항목들이 속한다.(p, r, b 는 s, d, c 보다 중요한 항목) 지표층은 매 항목들의 속성이 속한다.

매 항목들의 영향정도는 다음과 같다.

$I(s)=0.104\ 8$, $I(p)=0.177\ 6$, $I(d)=0.150\ 4$, $I(r)=0.208\ 4$, $I(c)=0.150\ 4$, $I(b)=0.208\ 4$

이때 지표층의 매 지표들이 자기 옷층의 요인에 주는 영향정도는 같다.

최소지지도를 0.3, 최소믿음도를 0.8로 설정하여 발굴연산을 진행한다.

이 값들은 규칙들을 발굴하는데 적합한 값을 찾기 위하여 진행한 여러차례의 실험들에 기초하여 선택된것이다.

표 3. 영향지지도를 리용하지 않은 공간관계규칙

No.	규칙	설 명	지지도	믿음도
1	$d1 \rightarrow s1$	려관이 도로와 매우 가까우면 규모가 크다.	0.32	0.89
2	$d1 \rightarrow r1$	려관이 도로와 매우 가까우면 하천, 호수와 매우 가깝다.	0.48	0.80
3	$d1 \rightarrow c1$	려관이 도로와 매우 가까우면 철도와 매우 가깝다.	0.52	0.87
4	$b4 \rightarrow s1$	려관이 바다로부터 매우 멀면 규모가 크다.	0.32	0.80
5	$d1 \rightarrow r1, b4$	려관이 도로와 매우 가까우면 하천, 호수와 매우 가깝고 바다로부터 매우 멀다.	0.32	0.89
6	$d1 \rightarrow c1, b4$	려관이 도로와 매우 가까우면 철도와 매우 가깝고 바다로부터 매우 멀다.	0.32	0.89

표 4. 영향지지도에 의한 공간관계규칙

No.	규칙	설 명	지지도	믿음도
1	$r4 \rightarrow p1$	려관이 바다로부터 매우 멀면 가격이 비싸다.	0.32	0.85
2	$r1 \rightarrow d1$	려관이 하천, 호수로부터 매우 가까우면 도로와 매우 가깝다.	0.52	0.80
3	$d1 \rightarrow c1$	려관이 도로와 매우 가까우면 철도와 매우 가깝다.	0.47	0.87
4	$d1 \rightarrow r1, b4$	려관이 도로와 매우 가까우면 하천, 호수와 매우 가깝고 바다로부터 매우 멀다.	0.36	0.83
5	$d1 \rightarrow c1, b4$	려관이 도로와 매우 가까우면 철도와 매우 가깝고 바다로부터 매우 멀다.	0.33	0.84

표 3, 4에서 보는바와 같이 영향지지도를 리용하지 않고 규칙발굴을 진행한 결과 6개의 규칙이 탐색되었으며 영향지지도에 의한 규칙발굴을 진행한 결과 5개의 규칙이 탐색되었다.

후자에서는 규칙 $\langle d1 \rightarrow s1 \rangle$, $\langle d1 \rightarrow r1 \rangle$, $\langle b4 \rightarrow s1 \rangle$ 이 없고 새로운 규칙 $\langle r4 \rightarrow p1 \rangle$, $\langle r1 \rightarrow d1 \rangle$ 이 있다. 그리고 두 결과에 다 있는 규칙 $\langle d1 \rightarrow c1 \rangle$, $\langle d1 \rightarrow c1, b4 \rangle$, $\langle d1 \rightarrow r1, b4 \rangle$ 는 항목들의 영향정도와 빈발성이 동시에 높은 규칙들이다.

이 규칙들은 려관의 공간분포상태를 보여주는 여러 항목들가운데서 려관가격항목(p), 하천, 호수와의 거리관계항목(r), 바다와의 거리관계항목(b)에 흥미를 가지는 수요자에게 더 쓸모있는 규칙들이다. 이것은 항목들의 영향정도를 고려한 결과 모호관계규칙발굴효과가 보다 높아졌다는것을 알수 있다.

그리고 규칙 $\langle d1 \rightarrow s1 \rangle$, $\langle d1 \rightarrow r1 \rangle$, $\langle b4 \rightarrow s1 \rangle$ 은 빈발성은 높지만 영향정도가 보다 낮은 규칙들이다. 규칙 $\langle b4 \rightarrow p1 \rangle$, $\langle r1 \rightarrow d1 \rangle$ 은 빈발성은 낮으나 영향정도가 보다 높은 규칙들이다.

맺는 말

론문에서는 공간자료들이 지리적현상에 주는 영향정도를 고려한 영향지지도에 의한 모호공간관계규칙발굴방법을 제시하고 실험을 통하여 그 효과성을 검증하였다.

이 방법은 지리적현상을 보다 정확히 반영한 공간관계규칙들을 발굴하기 위한 방법으로서 현실성이 높은 공간지식자료기지를 구축하는데 적용할수 있다.

참고문헌

- [1] B. Sowan et al.; Expert Systems with Applications, 40, 6928, 2013.
- [2] Xiong Ying et al.; Ecological Modeling, 209, 97, 2007.
- [3] 张彦丽; 测绘科学, 34, 2, 233, 2009.

주체105(2016)년 11월 5일 원고접수

Fuzzy Spatial Association Rule Mining Method with Influence Support

Kim Nam Hyok

We newly proposed an approach of spatial association rule mining by considering the influence degree of every item in fuzzy spatial database on the total geosystem and inspected the effectiveness by the experiment. This approach can be expected to be used effectively to build a down-to-earth and reasonable spatial knowledge base.

Key words: fuzzy spatial association, influence support, rule mining