

다중 공간자료 통합을 위한 확률론적 결정 자료 융합

Probabilistic Decision Fusion for Multi-source Spatial Data Integration

박노욱*¹, 권병두², 지광훈¹

¹한국지질자원연구원 지질자원정보센터 E-mail: nwpark@kigam.re.kr

²서울대학교 지구과학교육과

1. 서론

지구과학 현상의 대부분은 여러 변수가 복합적으로 작용하여 나타나기 때문에 자료의 통합적 분석을 요구하게 된다. 최근 컴퓨터 제반 기술의 발달과 자료 획득 및 처리 기술의 발달로 인해 대용량 자료의 수집, 저장, 분석 및 관리에 많은 관심이 집중되고 있다. 이와 관련하여 컴퓨터 기반의 자료 저장 및 관리에 효과적인 GIS의 중요성이 점차 증대되고 있다. 그러나 기존 GIS의 경우 컴퓨터 기반의 자료 도시에 치우쳐 있기 때문에 고유한 자료 분석 기법을 요구하는 지구과학 분야의 직접적인 적용 연구는 상대적으로 덜 이루어져 왔다. 이러한 배경에서 GIS 기반의 공간자료와 지구과학 자료 고유의 특성을 고려한 분석 방법의 결합에 대한 연구가 1990년대 중반부터 진행되어 왔는데, 이러한 분야는 지구과학, 통계학, 컴퓨터 과학이 결합된 형태이다(Bonham-Carter, 1994).

이 논문에서는 다중 공간자료의 통합 분석을 위해 효율적으로 이용될 수 있는 공간자료 통합 방법 중에서 확률론적 방법에 대해서 논의하고자 한다. 우선 기존 베이지안 이론 기반 확률론적 방법을 변형한 확률비 기반 자료 통합 이론을 소개하고, 개별 자료의 일종의 가중치로 해석할 수 있는 최종 통합 목표에 대한 신뢰값(reliability factor)을 정의하여 이를 통합에 사용하였다. 제안 기법의 적용성을 평가하기 위하여 산사태 취약성 분석을 목적으로 사례연구를 수행하였다.

2. 확률론적 공간통합 방법

(1) 확률비 기반 공간통합

기존 확률론적 방법을 이용한 공간자료 통합 방법은 베이지안 확률론의 틀 안에서 사전 확률, 조건부 확률을 계산한 후에 베이지안 결합 규칙을 이용하여 다중 공간자료를 통합하였다(Chung and Fabbri, 1999; Park et al., 2003). 이때 다양한 자료에 대한 결합 조건부 확률 분포를 함께 정의하는 것이 힘들기 때문에, 조건부 독립의 가정하에 개별 자료에 대한 이변수 확률 분포를 구한 후에 이들의 곱으로 통합을 수행하게 된다. 조건부 독립의 가정은 수학적으로 다루기 편한 형태를 제공하기는 하나, 실제 내부적으로 복잡한 관계를 가지는 자료들에 대해서 증명하기가 쉽지 않은 경우가 많다. 이러한 전통적인 조건부 독립 가정의 확률론적 통합의 대안으로 Journel(2002)은 확률값 자체보다는 확률비(ratio)가 더 안정하다는 점에 착안하여 확률비에 기반한 자료 통합 이론을 제시하였다.

산사태 취약성 분석과 관련하여 이론을 설명하면 다음과 같다. 산사태 발생과 관련된 두 개의 인자를 각각 $E_i (i = 1, 2)$ 라고 하고, 통합 목표 T_p 와 $\overline{T_p}$ 를 “화소 p는 미래의 산사태에 대해서 영향을 받을 것이다.”와 “화소 p는 미래의 산사태에 대해서 영향을 받지 않을 것이다.”고 각각 정의하였을 때, 최종적으로 우리가 원하는 것은 개개 화소에 서의 $P(T_p | E_1, E_2)$ 이다.

이 기법에서는 로지스틱 형태의 확률비를 다음과 같이 정의한다.

$$a = \frac{P(\overline{T_p})}{P(T_p)}, \quad b = \frac{P(\overline{T_p} | E_1)}{P(T_p | E_1)}, \quad c = \frac{P(\overline{T_p} | E_2)}{P(T_p | E_2)} \quad (1)$$

여기서 a는 E_i 가 주어지지 않은 상태에서 통합 목표에 대한 일종의 거리를 나타내는 사전 확률의 비를 의미한다. 그리고 b와 c는 E_1 와 E_2 가 주어졌을 때 통합 목표에 대한 거리를 각각 나타낸다. 만약 이 값이 1보다 크다면 통합 목표와의 관계가 적은 것을 나타낸다. 산사태 발생과 관련된 공간자료들이 산사태 취약성에 대한 중요한 정보를 제공한다고 가정하였을 때, 산사태가 발생한 지역에서의 공간자료의 특성과 산사태가 발생하지 않은 지역에서의 공간자료의 특성은 다르게 나타나야 한다. 로지스틱 형태의 확률비를 이용할 경우, 이러한 차이를 부각시킬수 있는 장점이 있다.

위에서 정의된 a, b, c를 통합하기 위해 다음의 가정을 이용하였다.

$$\frac{x}{b} = \frac{c}{a} \quad (2)$$

여기서 $x = \frac{P(\overline{T_p} | E_1, E_2)}{P(T_p | E_1, E_2)}$ 를 나타낸다. x는 E_1 와 E_2 가 주어졌을 때 통합 목

표에 대한 거리를 나타내며, $\frac{x}{b}$ 는 통합 목표에 대해 E_1 을 이용하였을 경우의 상대적인 기여도의 증가 여부를, $\frac{c}{a}$ 는 E_1 을 이용하지 않았을 경우에 초기 확률에 대한 E_2 의 상대적인 기여도의 증가 여부를 나타낸다. 식 (2)는 E_1 를 이용하기 전이나 후에 E_2 의 상대적인 기여도의 증가 여부는 같다는 가정을 내재적으로 포함하고 있다.

식 (2)를 이용하여 최종적으로 얻고자 하는 결합 조건부 확률 $P(T_p|E_1, E_2)$ 은 식 (3)을 이용하여 얻을 수 있다.

$$P(T_p|E_1, E_2) = \frac{1}{(1+x)} = \frac{a}{(a+bc)} \quad (3)$$

위의 관계식을 m개의 자료로 확장할 경우 다음 관계식을 얻게 된다.

$$P(T_p|E_1, E_2, \dots, E_m) = \frac{1}{(1+x)}$$

$$x = \frac{\prod_{i=1}^m E_i}{a^{m-1}}, E_i = \frac{P(\overline{T_p}|E_i)}{P(T_p|E_i)} \quad (4)$$

(2) 개별 자료의 신뢰도 결정

앞에서 제시한 공간자료를 통합하기 위한 최종적인 관계식 (4)에서는 개별 자료의 통합 목표에 대한 신뢰도 혹은 가중치를 동일하게 주었다. 그러나 공간통합에 사용하는 자료들의 통합 목표에 대한 신뢰도는 다를 수 있기 때문에 이러한 값을 정의하여 통합 과정에 적용할 필요가 있다.

개별 자료 E_i 의 신뢰도를 ω_i 라고 하였을 때, 이를 반영한 공간통합 관계식은 다음과 같이 정의할 수 있다.

$$\frac{x}{a} = \prod_{i=1}^m \left(\frac{E_i}{a} \right)^{\omega_i} \quad (5)$$

식 (5)를 이용할 경우 개별 자료 E_i 의 신뢰도 ω_i 를 어떻게 정의하느냐가 중요한데, 이 논문에서는 통합 목표에 대한 개별 자료의 성공 비율을 계산한 후에 이 값을 개별 자료의 신뢰도로 사용하는 결정 융합 방법을 사용하였다. 다중 센서 원격탐사 자료의 결정 융합 방법에서는 개별 자료에 대한 분류를 수행한 후에 참조 자료와의 비교를 통한 분류 정확도를 계산하고 이를 개별 자료에 대한 신뢰도로 사용하였다. 이 연구에서는 이와 유

사하게 우선 개별 자료에 대한 확률비 기반의 표현을 수행한 후에 상대적인 랭크로 표시한 후에 과거 산사태와의 비교를 통해 0에서 1사이의 값을 가지는 신뢰도 값을 계산하였다.

3. 사례연구

산사태 취약성 분석을 목적으로 1998년 8월 집중 호우로 발생한 산사태로 많은 피해를 입은 충청북도 보은 지역을 대상으로 사례연구를 수행하였다. 과거 산사태 자료는 산사태 발생 이전과 이후의 항공사진을 이용하여 1차적으로 탐지한 후에 현지 조사를 거쳐 최종적으로 결정하였다. 그리고 산사태 발생과 관련된 공간자료로는 사면경사도, 사면방위도, 지질도, 임상종류도, 토양배수도를 사용하였다.

사전 확률은 연구지역의 전체 면적에 대한 과거 산사태 발생자료의 면적의 비로 정의하였으며, 개별자료에 대한 조건부 확률은 자료내의 속성별로 그 안에 포함된 과거 산사태의 면적의 비로 계산하였다. 이러한 과정으로 개별자료에 대한 확률비를 표현하였는데, 개별 자료의 적용 결과는 다음과 같다. 사면경사의 경우 경사각이 25도 이상일 경우에 산사태 발생빈도가 높았으며, 사면방위의 경우 동쪽, 남동쪽, 남쪽, 남서쪽 방위에서 산사태 발생빈도가 높게 나타났다. 이러한 결과는 방위에 따른 일조시간의 차이에 기인한 것으로 판단된다. 임상도의 경우 리기다 소나무와 잣나무에서, 토양 배수도에서는 배수가 매우 양호한 경우에서 산사태 발생빈도가 각각 높게 나타났다. 우리나라에서는 일반적으로 침엽수림에서 더 많은 산사태가 발생한다는 기존 결과와 대체로 일치하였다. 또한 일반적으로 집중 호우시 배수가 잘 될수록 물의 흐름을 조절하여 토양이 물을 많이 포함하여 결국 이로 인해 산사태 우도비가 높아지는데, 용인지역의 경우도 배수가 잘 될수록 산사태 우도비가 뚜렷이 높게 나타났다. 지질의 경우 죽전 화강암에서 산사태 발생 가능성이 높게 나타났다. 개별자료에 대한 기여도를 계산한 결과, 사면경사, 지질도, 임상종류, 토양배수, 사면방위 순으로 나타났다.

이러한 과정으로 계산한 다음 모든 자료를 식 (3), (4), (5)를 이용하여 최종적인 연구지역의 산사태 취약성 분포도를 작성하였다. 작성결과 연구지역의 중앙부와 죽전화강암이 존재하는 연구지역의 중앙 상단부에서 산사태 취약성이 높은 지역이 주로 나타났으며, 지질도상에서 황강리 퇴적암이 존재하는 좌측 부분에서는 상대적으로 취약성이 낮게 나타났다.

4. 결론

이 연구에서는 다중 공간자료의 통합을 목적으로 확률론적 자료 융합 방법을 제안하고 사례연구를 수행하였다. 조건부 확률값을 직접 사용했던 기존 확률론적 자료 융합 방법을 대신하여 확률값 자체보다 유의미한 해석결과를 제공할 수 있는 확률비 기반의 자료 융합 방법을 적용하였다. 그리고 개별 공간자료의 기여도 혹은 가중치 계산을 위하여 랭크 기반의 면적비율을 이용하였다. 제안 기법을 산사태 취약성 분석을 목적으로 사례연구를 수행한 결과, 확률비를 이용함으로써 개별자료내에서의 산사태 발생 특성을 추출할 수 있었으며, 신뢰도 계산을 통하여 개별자료 사이의 가중치를 추출할 수 있었다.

이 연구에서는 제안 기법을 이용하여 산사태 취약성 분포도만을 작성하였는데, 앞으로 미래의 산사태 예측과 관련하여 불확실성 분석을 위한 검증 과정을 수행할 예정이다.

5. 참고문헌

- [1] Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists: modeling with GIS. Pergamon, New York, 398p.
- [2] Chung, C.F. and Fabbri, A.G., 1999, Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing, 65, 1389-1399.
- [3] Journel, A.G., 2002, Combining knowledge from diverse sources: an alternative to traditional data independence hypothesis. Mathematical Geology, 34, 573-596.
- [4] Park, N.-W., Chi, K.-H., Chung, C.F., Kwon, B.-D., 2003, Application of spatial data integration based on the likelihood ratio function and Bayesian rule for landslide hazard mapping. Journal of the Korean Earth Science Society, 24(5), in press.