

확률론적 공간 자료 통합 모델을 이용한 산사태 취약성 분석

박노옥*¹⁾, 지광훈¹⁾, 권병두²⁾

1) 한국지질자원연구원 지질자원정보센터

E-mail: nwpark@kigam.re.kr

2) 서울대학교 지구과학교육과

요약: 이 논문에서는 산사태 취약성 분석을 목적으로 확률론적 공간통합의 틀 안에서 범주형 자료와 연속형 자료를 효율적으로 처리할 수 있는 비모수적 우도비 추정 모델과 모수적 예측적 판별 분석 모델을 적용하였다. 적용 모델의 비교를 위해 1998년 여름철 산사태로 많은 피해를 입은 경기도 장흥 지역과 충청북도 보은 지역을 대상으로 사례연구를 수행하였다. 장흥 지역에서는 두 모델이 유사한 예측 능력을 나타내었으나, 보은 지역에서는 모수적 예측적 판별 분석 모델이 상대적으로 높은 예측 능력을 나타내었다. 결론적으로 제안한 두 모델은 산사태 취약성 분석을 위한 연속형 자료 표현에 효율적으로 적용될 수 있으며, 두 모델이 개별적인 연속형 자료 표현의 특성을 가지고 있기 때문에 다른 사례 연구를 통한 검증 작업이 병행되어야 할 것으로 생각된다.

주요어: 산사태 취약성, 공간통합, 우도비, 판별 분석

1. 서론

해마다 여름철에 집중호우나 태풍으로 인한 산사태는 인명과 재산 피해뿐만 아니라 국가 기반시설의 기능에도 피해를 입히고 있으며, 이러한 피해는 계속 증가되고 있는 상황이다. 이러한 산사태 재해는 비록 강우나 지진 등의 요인에 의해 직접적으로 발생하지만, 다른 지구과학 현상과 마찬가지로 하나의 요인에 의해 발생한다기보다는 여러 요인이 복합적으로 작용하여 발생하게 된다. 따라서 산사태 피해를 분석하거나 예방하기 위해서는 여러 요인을 함께 분석할 수 있는 방법론의 개발이 필요한 상황이다.

다양한 지질자료를 효율적으로 통합 분석하기 위한 방법론 개발은 확률/통계, 퍼지 이론, 인공지능 등에 이론적 배경을 두고 1980년대 말부터 진행되어 왔다. 그러나 기

존 산사태 취약성 분석 연구의 경우, 대부분 서로 다른 형태의 자료(범주형, 연속형 자료) 처리와 확률론적 모델을 적용할 때 필요한 가정(조건부 독립, 선형성 등)에 대해서 논의 없이 곧바로 적용하여 왔다.

이 연구에서는 산사태 취약성 분석을 목적으로 Chung and Fabbri(1998)이 제안한 우도비 모델을 기반으로 비모수적(non-parametric) 및 모수적(parametric) 추정 모델을 각각 적용하였다. 연속형 자료의 우도비 계산을 위해 적용한 2가지 모델은 경험적 추정 모델(empirical estimation)과 예측적 판별 분석모델(predictive discriminant analysis model)이다. 1998년 8월 여름 집중호우로 인해 산사태 피해를 입은 경기도 장흥 지역과 충청북도 보은 지역의 사례연구를 통해 모델의 정량적 비교와 적용시 제안점에 대해서 논의하였다.

2. 우도비 기반 공간통합

산사태 취약성 분석과 관련하여 Chung and Fabbri(1998)에 의해 제안된 우도비 기반 공간자료 통합 모델은 산사태 발생지역과 미발생 지역에서의 확률의 비를 계산함으로써 산사태에 대해 취약한 지역과 취약하지 않은 지역의 차이를 정량적으로 부각시킬 수 있다.

보통 산사태 취약성 분석의 경우 지질도, 토양도, 임상도 등의 범주형 자료와 사면고도, 경사, 방위 등의 연속형 자료를 함께 사용하게 된다. 기존 우도비를 기반으로 산사태 취약성 분석을 수행한 경우(Lee and Min, 2001), 연속형 자료를 이진화하거나 몇 개의 클래스를 가지는 범주형 자료로 변환하여 사용하여 왔다. 그러나 이 경우 범주형화 과정에서 정보의 손실이 따르며, 이 과정이 주관적일 수 있다. 따라서 이 연구에서는 범주형 자료와 연속형 자료를 구분하여 처리하고, 특히 연속형 자료의 원래 값을 그대로 사용하고자 하였다.

범주형 자료와 연속형 자료사이에 조건부 독립(conditional independence)을 가정하게 되면 개별적인 자료 처리가 가능하여(Chung and Fabbri, 1998), 최종적인 결합 우도비는 범주형 자료의 우도비와 연속형 자료의 우도비의 곱으로 나타낼 수 있다. 이 연구에서는 범주형 자료의 우도비는 개별 자료의 속성별로 산사태 발생 화소와 미발생 화소의 빈도수를 비교하여 계산하였고, 연속형 자료는 원래 값을 그대로 표현할 수 있도록 빈도수 분포가 아닌 확률 분포를 이용하였다. 이를 위해 비모수적 우도비 추정 모델과 모수적 예측적 판별 분석 모델을 적용하였다.

2.1. 비모수적 우도비 추정 모델

비모수적 추정 모델에서는 특정 확률 분포를 가정하지 않고, 자료 자체의 값으로부터 확률 분포를 추정하게 된다. 비모수적 확률 분포 추정 방법으로는 통계적 패턴 인식 분야에서 평활화 커널 추정(smoothed kernel estimation)과 k-nearest neighbor estimation 등이 대표적으로 사용되고 있는데, 아직까지 지질자료의 표현에 사용된 연구는 거의 없는 상황이다. 이 연구에서는 비모수적 확률 분포 추정을 위해 평활화 커널 추정 방법(smoothed kernel estimation)을 이용하였다. 평활화 커널 방법은 샘플 자료에 커널 함수를 위치시킨 후 커널 함수의 중점으로 경험적 확률 밀도 분포를 계산하는 방법에 해당된다(Parzen, 1962).

2.2. 모수적 예측적 판별 분석 모델

모수적 모델에서는 연속형 자료를 개별적으로 처리하는 것이 아니라 표본(산사태 취약성 분석의 경우 과거 산사태 발생 화소와 나머지 미발생 화소)으로부터 다변량 확률 분포를 가정하는 모수를 추정하여 분석을 수행하게 된다. 모수적 모델을 적용할 때 필요한 모수는 평균과 공분산이며, 표본 평균과 공분산을 모집단의 평균과 공분산의 추정치로 간주하게 된다.

이 논문에서 적용한 예측적 판별 분석 모델은 다변량 정규 분포를 가정하는 최대우도 추정 모델(maximum likelihood estimation model)과는 달리 다변량 student-t 분포를 가정하게 된다(McLachlan, 1992). 최대우도 추정 모델과의 차이는 표본 통계치가 얼마나 모집단 통계치에 근접한지에 대한 불확실성을 고려할 수 있다는 점이다. 표본수가 클 경우, student-t 분포는 정규 분포에 접근하여 두 모델의 결과가 유사하게 나타나지만, 표본수가 작은 경우에는 최대우도 추정 모델의 추정치를 조절할 수 있기 때문에 보다 나은 결과를 나타내는 것으로 알려져 있다(McLachlan, 1992; SpatialModels Inc., 2004).

3. 사례연구

이 연구에서 제안한 우도비 기반 2가지 모델의 실제 적용과 비교를 위해 1998년 8월 여름에 집중호우로 많은 산사태 피해를 입은 경기도 장흥 지역과 충청북도 보은 지역을 대상으로 사례 연구를 수행하였다. 제안 모델의 적용을 위해 Fortran 프로그래밍과 Spatial Prediction Modeling System(SpatialModels Inc., 2004)을 이용하였다. 사

레 연구를 통해 제안한 이론의 실제 적용과정에서의 적용 가능성과 모델의 비교에 대한 논의를 하고자 하였다.

경기도 장흥 지역 사례연구에는 범주형 자료로 임상 종류도, 토양도를, 연속형 자료로 고도, 사면 방위, 사면 방위를 이용하였다. 정량적 비교를 위해 충청북도 보은 지역의 경우에도 동일한 자료를 이용하였다.

두 모델의 유사성의 비교를 위해 정량적 지수로 Chung *et al.*(2002)의 matching rate를 확장한 유사율(similarity rate: SR)을 정의하고 계산하였다. 유사율 계산을 위해, 두 모델의 취약성 분포도에서 상위 취약성 등급 5%별로 특정 누적 등급을 산사태 취약 지역으로, 나머지 비취약지역의 2개의 클래스로 구분을 하였다. 이러한 두 모델별로 구분된 2개의 클래스를 중첩할 경우, 공통 취약 클래스, 공통 비취약 클래스, 한 모델에서는 취약지역으로 구분되지만 다른 모델에서는 비취약지역으로 구분되는 클래스, 반대의 경우 등 모두 4개의 클래스를 얻게 된다. 이로부터 유사율 SR은 두 모델의 공통으로 나타나는 산사태 취약지역의 화소수를 계산한 후에, 이 값을 전체 화소수에서 공통 비취약지역의 화소수를 뺀 값으로 나눠주어 계산하게 된다. 만약 두 모델의 취약성 등급의 위치가 유사할 경우, 공통 취약 지역과 비취약 지역의 화소수가 많이 나타나게 되어 유사율이 높아지게 된다.

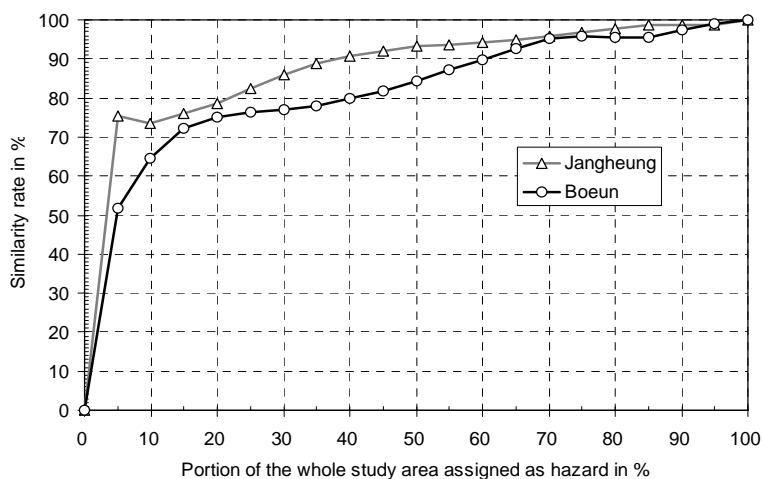


Fig. 1. Similarity rates for the Jangheung and Boeun areas.

장흥과 보은 지역의 유사율을 살펴보면(Fig. 1), 전반적으로 장흥 지역이 보은 지역보다 높게 나타남을 확인할 수 있다. 상위 5%와 10% 취약성 등급을 살펴보면, 장흥 지역에서는 약 75%, 73%, 보은 지역에서는 약 52%, 65%로 각각 나타났다. 이는 장흥 지역에서는 비모수적 경험적 우도비 추정 모델과 모수적 예측적 판별 분석 모델이 유사한 예측 패턴을 보이는 것으로 간주할 수 있다.

각 모델의 예측 능력을 정량적으로 비교하기 위하여, 일종의 검증(validation)으로 과거 산사태를 분할하여 계산하는 예측비율곡선(Chung and Fabbri, 1999)을 작성하였다. 예측비율곡선은 과거 산사태의 분할에 기반하여 계산되는데, 이 연구에서는 임의적 공간 분할 방법을 통해 예측비율곡선을 작성하였다. 과거 산사태 발생위치를 임의 분할을 통해 두 개의 그룹으로 나누어서 한 개의 그룹은 통합에, 나머지 그룹은 검증에 사용하였다. 그리고 반대로 검증용 자료를 통합용 자료로, 그리고 통합용 자료를 검증용 자료로서 사용하였다.

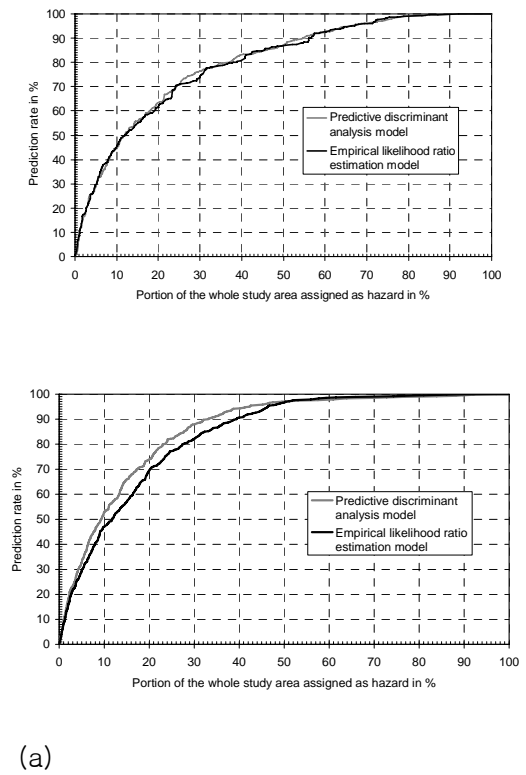


Fig. 2. Prediction rate curves for (a) Jangheung, (b) Boeun.

장흥 지역의 경우(Fig. 2 (a)), 두 모델의 예측 능력이 거의 유사하게 나타났다. 두 모델 모두 상위 10% 취약성 등급에서 약 45%의 예측 능력을 나타내고 있다. 두 모델간의 유사한 예측 능력은 앞 절에서 제시한 유사율과도 연관성이 있다. Fig. 1에서 장흥 지역의 경우 보은 지역에 비해 두 모델간의 유사율이 높게 나타나고 있다. 이러한 두 모델의 통합결과의 유사성에 의해 예측 능력 또한 유사하게 나타난 것으로 판단된다. 모델 구성의 이론적 관점에서 보면, 장흥 지역의 연속형 자료는 비모수적 추정과 모수적 추정을 통해 모두 적절하게 표현될 수 있다고 간주할 수 있다. 그러나 보은 지역의 경우(Fig. 2 (b))에서는 두 모델의 예측 능력의 차이가 나타나고 있다. 비모수적 우도비 추정 모델의 경우, 상위 10%에서 47%의 예측 능력을 나타내는 반면, 모수적 예측적 판별 분석 모델에서는 53%의 예측 능력을 보이고 있으며, 이러한 양상은 상위 50% 취약성 등급까지 계속되고 있다. Fig. 1에서 살펴 본 것처럼, 보은 지역이 장흥 지역에 비해 두 모델의 유사율이 현저하게 떨어지고 있다. 즉 두 모델의 통합 결과의 불일치 부분이 장흥 지역에 비해 많이 나타난다고 해석할 수 있다. 따라서 이러한 상이한 유사율이 최종적으로 교차 검증을 통해 예측 능력에도 반영되었다고 판단된다. 우도비의 이론적 틀안에서 범주형 자료의 우도비는 두 모델에서 모두 동일하게 곱해지기 때문에, 연속형 자료의 영향에 의해 예측 능력이 좌우된다. 따라서 모수적 모델이 비모수적 모델보다 예측 능력이 상대적으로 높은 것으로 나타난 결과는, 자료 표현의 관점에서 보은 지역의 연속형 자료가 모수적 다중 변수 표현으로 더 효과적으로 모델링 될 수 있다는 것으로 해석할 수 있다.

4. 결론

이 논문에서는 산사태 취약성 분석을 위한 기존 GIS 기반 확률론적 공간 통합 모델에서 연속형 자료의 표현을 위해 비모수적 추정 모델과 모수적 예측적 판별 분석 모델을 적용하였다. 경기도 장흥 지역과 충북 보은 지역의 사례 연구 결과, 장흥 지역에서는 두 모델의 유사율이 비슷하게 나왔으며 교차 검증을 통한 예측 능력 비교에서도 유사하게 나왔다. 반면 보은 지역에서는 모수적 예측적 판별 분석 모델의 예측 능력이 비모수적 경험적 우도비 추정 모델보다 전반적으로 높게 나타났다. 비록 동일한 지형자료라 하더라도 연구 지역에 따라 특성이 달라지기 때문에, 보은 지역의 결과만으로 모수적 예측적 판별 분석 모델이 비모수적 경험적 추정 모델보다 적용 가능성이 높다고 결론지을 수는 없다. 따라서 이 연구를 통해 얻어진 결과를 일반화시키기 위해서는 보다 많은 사례 연구를 통한 검증 작업이 수행되어야 할 것으로 판단된다.

5. 사사

이 연구에 많은 조언을 해주신 Geological Survey of Canada의 Dr. C.F. Chung에게 감사드립니다.

6. 참고문헌

- Chung, F.C. and Fabbri, A.G., 1998, Three Bayesian prediction models for landslide hazard. Proceeding of International Association for Mathematical Geology (IAMG'98), 204-211.
- Chung, F.C. and Fabbri, A.G., 1999, Probability prediction models for landslide hazard mapping. Photogrammetric Engineering & Remote Sensing, 65, 1389-1399.
- Chung, F.C., Kojima, H., Fabbri, A.G., 2002, Stability analysis of prediction models for landslide hazard mapping. In Allison, R.J. (ed.) Applied geomorphology: theory and practice. John Wiley & Sons, Inc., London, 3-19.
- Lee, S. and Min, K., 2001, Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, 40, 1095-1113.
- McLachlan, G.J., 1992, Discriminant analysis and statistical pattern recognition. John Wiley & Sons, Inc., New Jersey, 526p.
- Parzen, E., 1962, On the estimation of a probability density function and the mode. Annals of Mathematical Statistics, 33, 1065-1076.
- SpatialModels Inc., 2004, User's guide of spatial prediction modeling system. 108p.