

시맨틱 기술과 베이지안 네트워크를 이용한 산사태 취약성 분석 Landslide Susceptibility Analysis Using Bayesian Network and Semantic Technology

이상훈*

Lee, Sang Hoon

要 旨

비탈면 혹은 절성토지의 파괴로 사람과 재산에 심각한 피해를 입히기 때문에 미리 산사태 취약성 분석을 수행하여 개발 혹은 자연재해로부터 위험을 대비하는 것이 필요하다. 기존의 산사태 취약성 분석은 휴리스틱, 통계학적, 결정론적 혹은 확률론적 방법을 통해 이뤄졌다. 그러나, 적은 현장정보 등으로 분석의 신뢰도가 떨어지거나, 전문가의 경험과 지식을 기존 정량적인 해석모델에 반영하기 어려웠다. 본 연구는 산사태 취약성 분석에 대한 전문가 지식과 공간입력자료의 시맨틱을 추출하여 온톨로지 모델을 구축하고, 이를 베이지안 네트워크에 반영하여 확률적인 산사태 모델링을 제안하였다. 기존에 전문가 수작업으로 이뤄지던 베이지안 네트워크의 구조 생성을 온톨로지 모델의 지식추론으로 자동화하고, 현장정보뿐만 아니라 전문가 지식을 모델링에 반영하여 조건부 산사태 발생확률분포를 작성하였다. 이 결과를 GIS에 적용하여 산사태 취약성 지도를 작성하였다. 검증을 위해 충남 홍성일원의 오서산 지역에 적용한 결과 기존 산사태 발생흔적과 86.5% 일치하였다. 본 연구를 통해 일반 사용자도 전문가 도움 없이도 광역적인 산사태 취약성 분석이 가능하리라 기대된다.

핵심용어 : 산사태 취약성 분석, 시맨틱, 온톨로지, 베이지안 네트워크, GIS, DL추론

Abstract

The collapse of a slope or cut embankment brings much damage to life and property. Accordingly, it is very important to analyze the spatial distribution by calculating the landslide susceptibility in the estimation of the risk of landslide occurrence. The heuristic, statistic, deterministic, and probabilistic methods have been introduced to make landslide susceptibility maps. In many cases, however, the reliability is low due to insufficient field data, and the qualitative experience and knowledge of experts could not be combined with the quantitative mechanical-analysis model in the existing methods. In this paper, new modeling method for a probabilistic landslide susceptibility analysis combined Bayesian Network with ontology model about experts' knowledge and spatial data was proposed. The ontology model, which was made using the reasoning engine, was automatically converted into the Bayesian Network structure. Through conditional probabilistic reasoning using the created Bayesian Network, landslide susceptibility with uncertainty was analyzed, and the results were described in maps, using GIS. The developed Bayesian Network was then applied to the test-site to verify its effect, and the result corresponded to the landslide traces boundary at 86.5% accuracy. We expect that general users will be able to make a landslide susceptibility analysis over a wide area without experts' help.

Keywords : Landslide Susceptibility analysis, Semantic, Ontology, Bayesian Network, GIS, DL-Reasoning

1. 서 론

태풍 혹은 집중호우, 지진 등으로 인한 비탈면 혹은 절성토지의 붕괴는 많은 인적 물적 피해를 발생시키는 주요한 자연재해이다. 특히, 최근의 도시 팽창으로 전

원공간이 개발됨에 따라 비탈면에 발생하는 산사태의 발생빈도가 증가되어 더 많은 인명피해와 재산피해가 우려되고 있다. 따라서, 미리 산사태 취약성을 산정하여 그 공간적 분포특성을 GIS로 분석하는 것은 산사태 위험을 판단하는데 중요한 역할을 한다. 현재 산사태

2010년 10월 20일 접수, 2010년 11월 19일 채택

* 정회원 · 한국건설기술연구원 U-국토연구실 전임연구원/팀장(sanghoon@kict.re.kr)

취약성 지도 작성은 휴리스틱, 통계학적, 결정론적 혹은 확률론적 방법 등이 소개되었다. 그러나, 현장 데이터가 부족하여 신뢰도가 떨어지거나, 전문가에 의해 주관적인 해석만 이뤄지는 한계가 있었다. 특히, 기존의 방법들에서는 정성적인 전문가의 경험과 지식을 정량적인 역학적 해석모델과 통합하여 반영할 수 없었다.

최근, 위성영상, 항공사진, LIDAR 등 해상도가 높은 공간정보가 생산되면서 GIS를 이용하여 넓은 지역을 빠르게 분석할 수 있다. GIS를 이용하면, 산사태 취약성 지도 작성을 부분적으로 자동화하고 현장조사를 최소화할 수 있다(Mejia-Navarro and Wohl, 1994). 그러나, 광역적인 범위에서 지형, 경사, 토질, 식생 등 산사태의 원인요소를 정량화하고 일반화시키기 어렵기 때문에, 주로 전문가가 개입되거나 통계기법이 이용되고 있다. 또한, 산사태의 원인 요소로 입력되는 지반의 물성과 공간정보는 근본적으로 불확실성을 내포한다. 이로 인하여 산사태 취약성 분석이 더욱 복잡하고 부정확하게 되지만, 현재는 거의 무시되고 있다.

본 연구에서는 전문가의 경험과 지식을 반영하면서 지반물성, 공간정보 등 입력자료의 불확실성을 고려할 수 있도록 시맨틱 기술과 확률적 추론모델인 베이시안 네트워크(Bayesian Network)를 결합하여 산사태 취약성 분석을 구현하였다. 먼저, 산사태의 취약성 분석에서 전문가의 경험과 지식을 정량적으로 활용하기 위하여 시맨틱 표현기법의 하나인 온톨로지 모델을 구축하고, 추론엔진을 이용하여 작성된 온톨로지 모델을 베이시안 네트워크의 구조로 자동 변환하였다. 또한, 베이시안 네트워크의 조건부확률분포를 이용하여 산사태 원인요소의 불확실성을 확률로 표현하여 산사태의 취약성을 분석하였다. 본 연구를 통해 광역적인 대상 지역에 대해 일반 사용자도 전문가 도움없이도 정량적인 산사태 취약성 지도를 작성할 수 있을 것이다.

1.1 연구방법

본 연구의 진행과정은 그림1과 같다. 먼저, 기존 산사태 취약성 분석 방법을 조사, 분석하였다. 기존 결정론적 방법에서 원인요소 및 역학적 해석과정 등 산사태 취약성 분석에 관한 지식을 기초로 시맨틱을 추출하고, 산사태 위험성 분석과정을 표현하는 온톨로지 레이어 모델을 OWL-DL(Web Ontology Language-Description Logic)로 작성하였다. 개발된 온톨로지 모델은 추론엔진인 JENA를 통해 베이시안 네트워크의 그래프 모델인 방향성 비순환 그래프로 변환하였다. 두번째, 지형, 경사, 토질, 식생 등의 원인요소의 불확실성을 고려하여 베이시안 네트워크 그래프의 변수로 이용하고, 조건

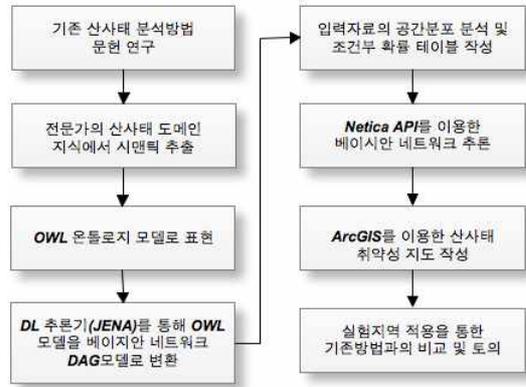


그림 1. 연구방법

부확률 테이블(CPT, Conditional Probabilistic Table)을 통해 각 변수 사이에 의존성을 표현하였다. 세번째, 베이시안 네트워크 엔진인 Netica-API(Norsys, 2009)를 이용하여 학습과 추론을 통해 산사태 취약성을 분석하고, GIS를 통해 지도화하였다. 또한, 검증에 위해 실험 지역을 대상으로 제안된 방법에 따라 산사태 취약성 지도를 작성하여 실제 산사태 흔적과 비교 및 검증을 실시하였다.

1.2 관련연구

산사태 취약성 분석에는 휴리스틱 방법, 통계학적 방법, 결정론적 방법 혹은 확률론적 방법이 있다. 먼저 휴리스틱 방법은 전문가의 경험과 지식을 기초로 산사태 가능성 요인을 정리하고, 중요도에 따라 다른 가중치를 할당한다. 사면의 경사와 지형의 형태를 기초로 산사태 취약성의 경중을 할당한다(Barredo et al., 2000). 그러나, 전문가 지식은 특정지역에 한정되는 문제가 있기 때문에 GIS와 함께 넓은 지역의 취약성을 분석하기에는 한계가 있다.

통계학적 방법은 산사태의 발생빈도와 지형, 지질, 식생 등의 요인들 사이에 상관관계를 통해 발생가능성을 추정하는 기법이다. 특히, GIS와 함께 간접적으로 산사태의 취약성을 매핑하기에 적합하다(Van Westen, 1993). 최근에는 퍼지집합, 인공신경망이나 Logistic Regression을 이용하여 지형, 토지피복, 토질 등의 변수를 예측하기도 한다(Lu and Rosenbaum, 2003). 그러나, 통계학적 방법은 산사태 발생요인을 간략화하는 경향이 있다. 산사태 종류에 따라 각각의 원인요소가 개별적으로 분석되어야 함에도 불구하고, 연구지역의 가정을 다른 지역까지 일반화시킨다. 또한, 시간변화를 무시하기 때문에 지하수위 상승, 토지이용변화와 같은

조건변화를 분석에 적용할 수 없다.

결정론적 방법은 무한사면모델을 기초로 역학적인 해석을 통해 안전율을 계산하는 정량적인 방법이다. 여러 연구자들은 산사태에 주요한 발생인자의 하나인 강우영향을 고려하기 위해 GIS, 수문모델, 그리고 사면안정모델을 이용하여 공극수압의 변화에 따른 산사태 위험성을 분석하였다(Gritzner et al., 2001; Chen & Lee, 2003). 결정론적 방법은 역학적 기법을 적용하여 실제 지반 물질의 파괴 매커니즘을 물리적으로 해석하여 경험적, 통계적인 방법보다 합리적인 방법으로 인정되고 있다. 그러나, 결정론적인 방법은 현장 혹은 실내 시험을 통해 얻어지는 상세한 입력자료를 요구하기 때문에, 넓은 지역보다는 좁은 지역에 적용할 수 밖에 없다는 단점을 가진다.

확률론적 방법은 결정론적 방법의 안전율을 과과확률로 나타낸 것이다. 이 방법은 점착력, 내부마찰각, 전단강도와 같은 지반 물성치의 변이와 사면에서 발생하는 공극수압의 공간적, 시간적 변화와 같은 불확실성을 고려할 수 있다. 산사태 취약성 해석에 불확실성을 고려하기 위하여 Chung and Fabbri(1999)는 베이지안 조건부 확률을 이용하였고 Suzen and Doyuran(2003)는 Weight of Evidence를 이용하였다. 그러나, 확률계산을 위해서는 결정론적인 방법에 비해 더 많은 수의 정보가 필요하며, 초기 지반의 물성치가 정확하지 않을 경우 불확실성의 신뢰도가 떨어지는 단점이 있다. 따라서, 현실적으로 자료가 부족하기 때문에 전문가 의사결정과 같은 주관적 확률평가를 통해 신뢰도를 보완하게 된다.

여러 연구자들은 전문가의 지식으로 확률평가를 보완하기 위해, 불확실한 지식 및 정보들의 확률적 인과관계를 표현하는 베이지안 네트워크를 적용하였다. 베이지안 네트워크는 컴퓨터의 기계학습을 통한 상황예측 및 추론 등 다양한 분야에서 사용되었다(Pearl, 1988). 산사태 취약성 분석에서 베이지안 네트워크를 이용할 경우, 다양하며 불확실한 정보와 그 연관성을 고려하면서 확률적인 사건예측이 가능하다. 또한, 베이지안 네트워크는 공간정보의 표현과 모델링에도 적용할 수 있다. Stassopoulou et al.(1998)은 베이지안 네트워크로 지중해 연안의 산불발생지역의 사막화 위험성을 추론하였으며, Kalacska et al.(2005)는 베이지안 네트워크를 이용하여 위성영상에서 LAI(Leaf Area Index)를 계산하였다.

2. 산사태 온톨로지 모델 설계

2.1 온톨로지 레이어의 구성

온톨로지는 “An ontology is a formal, explicit

specification of a shared conceptualization of a domain of interest”로 정의할 수 있다(Gruber, 1992). 온톨로지는 해당 도메인에서 용어와 개념 혹은 의미를 관계와 함께 묘사하여 도메인의 지식을 효과적으로 표현할 수 있다. 따라서, 온톨로지는 시스템 사이에서 도메인 지식을 공유하는데 유용하게 쓰이고 있다(Chandrasekaran, et al., 1999). 본 연구에서는 산사태 취약성 분석에 활용되는 지형, 토질, 지질, 수리/수문, 강우 등 다양한 도메인에 대한 전문가의 경험과 지식을 온톨로지로 표현한다. 온톨로지를 통해 무한사면안정 모델과 같이 정량적인 지식뿐만 아니라, 전문가의 경험에 따른 분류체계 등에 내재된 시맨틱(semantics)을 공유할 수 있다.

온톨로지는 특정한 업무 혹은 관점에 따라 표현범위가 달라지기 때문에 상위, 도메인, 업무, 응용 온톨로지를 레이어 형태로 구성할 경우 해당 도메인을 좀더 상세하게 표현할 수 있다. 본 연구에서는 산사태 취약성 분석을 위해 다양한 도메인이 결합되는 점을 고려하여 온톨로지를 세분화하여 레이어 형태로 구성하였다. 특히, 기존 산사태 분석시 전문가에 의해 수작업으로 정의되던 일반개념, 측정단위, 공간모델에 관한 온톨로지는 이상훈과 장병욱(2008)이 제시한 온톨로지 레이어 모델을 반영하여 그림 2와 같이 레이어를 구성하였다.

온톨로지 언어는 W3C에서 표준으로 채택한 OWL을 적용하였다. 이는 기존의 RDF 혹은 RDFS보다 표현성이 뛰어나며, 단순한 정보 표현 이외에 서술논리를 지원하여 추론할 수 있기 때문이다. OWL-DL은 추론에 필요한 사실을 표현하기 위하여 Axiom과 클래스간

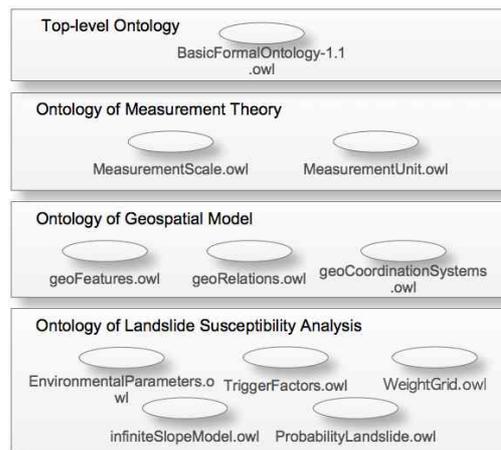


그림 2. 산사태 취약성 분석을 위한 온톨로지 레이어 모델

의 관계와 속성을 정의하여 새로운 관계를 만들 위하여 Constructor를 이용한다. 본 연구에서는 온톨로지 언어로 OWL-DL, 편집툴로는 Protégé 3.4를 이용하였다.

2.2 산사태 온톨로지

산사태는 “Landslide is movement of a mass of rock, earth or debris down a slope”로 정의된다 (Cruden and Varnes, 1996). 다양한 원인이 조합되는 산사태는 크게 내적원인과 외적원인으로 나눌 수 있다. 내적 원인은 사면의 지형, 지질, 토질, 토지이용 및 식생상태 등으로 토층의 전단강도를 감소시켜 파괴를 유발한다. 외적원인은 강우, 절토, 침식, 지하수위 변동 등으로 사면의 전단응력을 증가시켜 산사태를 유발한다. 이렇게 산사태는 토층에서 발생하는 불균등한 힘의 결과이다. 본 연구에서는 힘의 불균형을 역학적으로 해석하기 위하여 무한사면안정모델을 적용하였다.

무한사면안정모델은 무한히 긴사면에서 외력과 저항력에 대한 한계평형식으로부터 유도된다. 활동면의 길이 활동면의 깊이에 비하여 충분히 긴 사면의 안정계산에 이용되며, 흙의 성질과 지하수위 조건은 전체 사면에 걸쳐 변하지 않는 것으로 가정한다. 그림 3과 같이 지하수위면이 파괴면 위로 mz되는 깊이에 있고, 사면에 평행한 방향으로 정상침투가 일어난다고 가정한다. 이러한 가정을 바탕으로 무한사면 안전율(Fs)은 식 (1)과 같다. 여기서, c'는 흙의 점착력, m은 토층심도에 의한 포화된 토층심도의 비, γ, γ', γsat는 각각 흙의 습윤, 수중, 포화단위중량을, z는 수직하 흙의 깊이, β는 사면의 경사, 그리고 ϕ'는 흙의 유효 내부마찰각이다.

$$F_s = \frac{\tau_f}{\tau} = \frac{c' + [(1 - m)\gamma + m\gamma'] \times z \cos^2\beta}{\{[(1 - m)\gamma + \gamma_{sat}] \times z \sin\beta \cos\beta\}} \tan\phi' \quad (1)$$

본 연구에서는 무한사면안정모델을 기초로 산사태의 발생원인과 해석과정을 분류하고, 각 개념과 관계속성

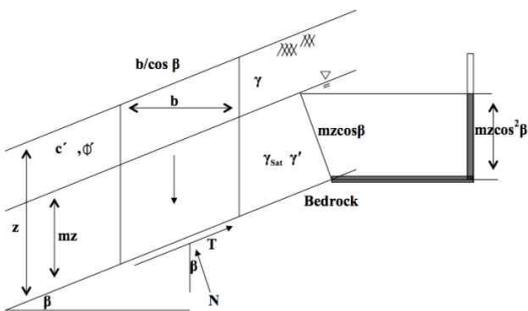
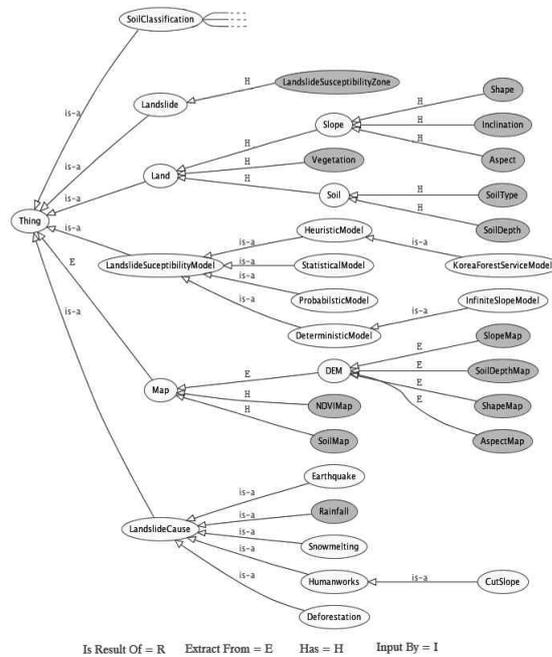


그림 3. 무한사면안정해석 모델



Is Result Of = R Extract From = E Has = H Input By = I

그림 4. 산사태 OWL-DL 모델
일부발췌(Protégé-OWLviz로 작성)

을 정의하였다. 그리고, 산사태 발생의 내적원인, 외적원인, 그리고 해석모델에 관한 지식을 OWL-DL의 온톨로지 모델을 통해 그림 2와 같이 environmental Parameters.owl, triggerFactors.owl, infiniteSlope Model.owl에 표현하였다. 최종적으로 그림 4와 같이 설계된 온톨로지 레이어 모델을 이용하여 산사태 취약성 분석을 수행하는 대상범위 및 특정 시간에 대한 인스턴스를 생성하였다.

3. 베이지안 네트워크 구조 추론

3.1 산사태 취약성 분석을 위한 베이지안 네트워크 복잡한 시스템도 모듈로 나누어 개념과 연결자로 이뤄진 그래프로 나타낼 수 있으며, 이를 통해 복잡성과 불확실성을 쉽게 설명할 수 있다(De Vel et al., 2006). 베이지안 네트워크는 그래프 이론과 확률 이론을 결합한 그래프 모델이다(Pearl, 1986). 베이지안 네트워크는 빈도주의적 확률 관점이 아닌, 사전/사후 확률을 조합한 베이지안 이론을 기초로 하기 때문에, 산사태 취약성 분석 시 입력자료가 일부 없거나 불확실할 경우에도 추론이 가능하다. 즉, 사면, 토질, 지질 등 관측 데이터 뿐만 아니라, 전문가의 지식과 경험과 같은 시맨틱을 변수의 하나로 조합하여 발생확률을 계산할 수 있다.

이러한 장점으로 인하여 베이스안 네트워크는 최근 복잡한 도메인에서 불확실성을 다루는 대표적인 확률모델로 각광받고 있다(Lee and Abbott, 2003).

베이스안 네트워크는 방향성 비순환 그래프를 이용하여 복잡한 시스템을 모듈화 한다. DAG에서 노드는 도메인의 확률변수를 나타내고 노드간의 연결자인 아크는 변수간의 확률적 의존관계를 나타낸다(Charniak 1991). 산사태 취약성 분석에서 사면의 경사와 같은 원인이 요소가 노드에 해당하며, 산사태 붕괴에 영향을 주는 임계치 이상의 강우량이 아크로서 노드 사이에 독립성을 나타낸다. 베이스안 네트워크에서 부모노드와 자식노드의 값은 조건부 확률분포를 가진다. 예를 들어 “강우량이 증가할 확률이 P=65%일 때 사면이 파괴될 확률 P=60%이다”가 이에 해당한다.

이러한 베이스안 네트워크 모델을 도메인에 적용하기 위해서는 해당 도메인에 따라 적절한 변수를 선정하고, 변수 사이에 원인관계를 기초로 확률분포를 할당해야 한다. 그러나, 이러한 작업은 매우 복잡하고 어렵기 때문에 전문가의 경험 혹은 지식에 의존하고 있는 실정이다.

3.2 온톨로지 모델의 베이스안 네트워크 변환

수작업으로 이뤄지는 베이스안 네트워크의 구조 작성의 어려움을 해결하기 위해 K2(Cooper and Herskovits, 1991), MDL(Lam and Bacches, 1993), CAMML(Wallace et al.,1996) 등이 개발되었다. 그러나 이러한 알고리즘은 구조를 너무 단순화하고 많은 양의 입력자료를 요구할 뿐만 아니라, 복잡한 다중 노드의 경우에는 부적합하였다. 무엇보다도 기존의 지식을 이용할 수가 없었다. 최근 바이오인포매틱스 혹은 통신망 관리에서 기존의 지식을 이용하여 베이스안 네트워크의 구조를 생성하기 위한 방법으로 온톨로지를 이용한 방법이 제안되었다(Helsper and der Gaag, 2002; Devitt, et al., 2006).

온톨로지는 전문가의 경험과 지식을 구조적으로 표현하기 위하여 해당 도메인을 개념과 관계로 나누어 결정론적으로 구성한다. 베이스안 네트워크는 관심 도메인에서 개념 혹은 변수 사이의 의존성 혹은 독립성을 확률론적으로 표현한다. 현재까지 결정론적인 온톨로지 방법론과 확률론적인 베이스안 네트워크 방법론은 기계학습이나 의사결정에 독립적으로 이용되어 왔다. 본 연구에서는 두 방법론을 결합하여 전문가의 수작업에 의존하던 베이스안 네트워크의 구조를 일반 사용자도 손쉽게 구성할 수 있으며, 전문가도 베이스안 네트워크 이론을 모두 이해하지 않고도 그 구조를 작성할

수 있다. 또한, 온톨로지가 표현하는 지식을 이용하기 때문에 적은 입력자료로도 베이스안 네트워크를 구축할 수 있다. 그 작성과정은 1) 베이스안 온톨로지 작성, 2) 노드 속성값 정의, 3) 노드 사이에 관계 정의, 4) OWL추론을 통한 베이스안 네트워크 모델 작성으로 이뤄진다.

3.2.1 베이스안 네트워크 온톨로지 작성

베이스안 네트워크의 구조를 추출하기 위해서는 관심 도메인의 온톨로지를 작성하고, 온톨로지서 문제 해결을 위한 원인관계를 추출해야 한다. 온톨로지서 개념의 속성은 베이스안 네트워크 노드에서 속성의 모임을 나타낸다. 속성의 모임은 노드명, 조건부확률테이블, 상태명(비연속노드에서 입력될 수 있는 변수의 값), 레벨(연속노드에서 입력될 수 있는 변수의 범위)을 포함한다. 벡터형태의 공간정보와 같이 연속변수를 위한 범위설정이나 초기 확률값 설정은 베이스안 네트워크 온톨로지의 개념 속성을 설정하여 조정할 수 있다. 이러한 온톨로지서 개념 및 개념 사이에 상속관계를 이용하여 베이스안 네트워크의 노드, 속성, 그리고 아크를 정의할 수 있다. 본 연구에서는 앞서 작성한 온톨로지를 기초로 베이스안 네트워크를 추출하기 위한 온톨로지 모델을 제안하였다.

3.2.2 노드 속성값 정의

베이스안 네트워크 노드의 속성은 Name, Kind, State name, Value(or Interval)을 포함한다. 작성한 온

```

### http://owl.geoinfo.or.kr/land#Inclination
:Inclination rdf:type owl:Class ;
    owl:equivalentClass [ rdf:type owl:Class ;
        owl:oneOf ( :very_steep
            :flat
            :very_flat
            :peak
            :medium_steep
            :steep
        )
    ];
    rdfs:subClassOf :Slope .

### http://owl.geoinfo.or.kr/land#very_steep
:very_steep rdf:type owl:DatatypeProperty ,
    owl:FunctionalProperty ;
    rdfs:domain [ rdf:type owl:Class ;
        owl:intersectionOf ( [ rdf:type owl:Restriction ;
            owl:onProperty :Angle ;
            owl:someValuesFrom [ rdf:type rdfs:Datatype ;
                owl:onDatatype xsd:integer ;
                owl:withRestrictions ( [ xsd:minInclusive 40 ]
            )
        ]
        [ rdf:type owl:Restriction ;
            owl:onProperty :Angle ;
            owl:someValuesFrom [ rdf:type rdfs:Datatype ;
                owl:onDatatype xsd:integer ;
                owl:withRestrictions ( [ xsd:maxExclusive 60 ]
            )
        ]
    )
    ] .
    
```

그림 5. 베이스안 네트워크 온톨로지의 제약조건 일부발췌

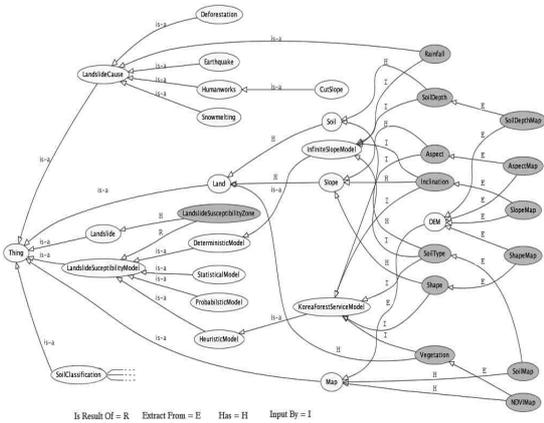


그림 6. 베이지안 네트워크 온톨로지 추론후 일부발췌

톨로지에서 클래스와 데이터타입 속성은 베이지안 네트워크 각 노드의 속성을 표현하는데 활용된다. 예를 들어 경사(Inclination)노드의 상태는 네 가지 종류의 상태(e.g. peak, very steep, steep, medium steep, flat, very flat)으로 분류할 수 있으며, 그림 5와 같이 equivalentClass와 DatatypeProperty을 OWL언어의 제약조건으로 표현된다.

3.2.3 노드 사이에 관계 정의

베이지안 네트워크에서 노드와 노드 사이에 아크(arcs)는 온톨로지 개념 사이의 관계로 정의된다. 예를 들어 산사태는 하나 이상의 원인을 가지며, 원인과 산사태 사이에는 객체속성(e.g. Is-A, Has, isResultOf, extractFrom)과 같은 원인관계가 성립된다. 온톨로지에 표현된 이러한 객체속성의 관계는 베이지안 네트워크에서 산사태 결과노드와 원인노드 사이에 아크로 변환된다. 또한, 추론엔진을 통해 인스턴스(실제 입력값)를 확인하여 그 무결성을 검증한다.

3.2.4 OWL-DL 추론을 통한 베이지안 네트워크 모델 작성

도메인과 베이지안 온톨로지는 OWL모델 형태로 작성되어, OWL-DL모델로부터 JENA추론엔진을 통해 자동적으로 아크가 생성된다. 추론엔진은 기 정의된 객체관계, 규칙과 현장입력자료를 체크하여 노드사이의 관계를 그림 6과 같이 추론이 이뤄지며, 이를 기초로 최종적으로 베이지안 네트워크의 DAG가 작성된다. 여기서 현장입력자료의 존재 여부 등에 따라 최종 베이지안 네트워크의 노드가 삭제되거나, 관계가 재설정 되기도 한다.

4. 산사태 취약성 분석

4.1 연구지역 및 자료 전처리

연구지역은 그림 7과 같이 충남 홍성군 일원의 오대산 일원이다. 이 지역은 넓은 범위의 지도변수와 다양한 산사태 가능성을 가진 인접지역으로 이뤄져 선정되었다.

이 지역의 성층구조는 여러 개의 다른 화강암, 층적층, 변성암의 기원을 가진다. 계곡은 인접한 모암으로부터 퇴적, 층적층으로 이뤄지며, 여러 개의 단층선도 명확히 나타나고 있다. 암반은 물리적, 화학적 풍화를

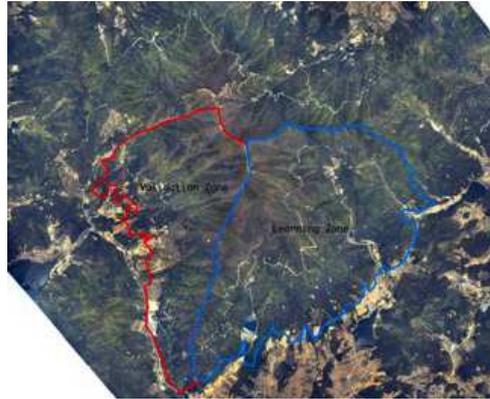


그림 7. 연구지역 항공사진과 베이지안 네트워크 학습/검증범위

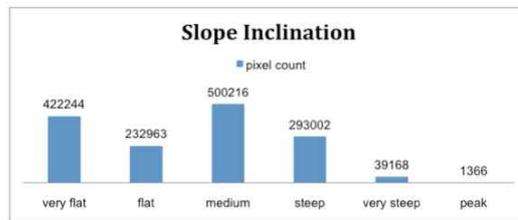
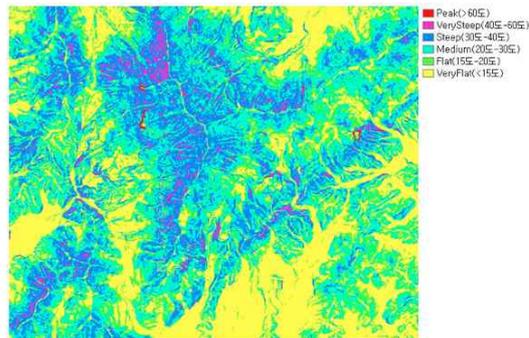


그림 8. 연구지역 사면경사 재분류 및 분포확률

받았다. 모암의 다양성으로 인해 다양한 비등방적인 토사를 확인하였으며, 토사는 봄과 여름의 더위와 습도 등 환경적 조건으로 인해 화학적 풍화가 발생되었다.

연구지역의 지형은 다양하게 이뤄져 있다. 급격한 산사태 언덕과 평탄한 충적지 등이 존재하며, 표고는 130에서 1800으로 다양하다. 사면의 경사가 넓은 범위에서 존재하지만, 사면의 경사가 1/3인 지역은 18-35%에 해당한다. 언덕을 가로지르는 몇 개의 임도가 있으며, 국내의 기후특성에 따라 평균 여름에 강우량은 100mm에 해당하며, 강설량은 50mm이다.

본 연구에서 제안한 베이지안 네트워크 모델에 입력하기 위한 입력요소로 사면경사, 향, 식생지수(NDVI), 토사종류, 토층깊이, 강우량에 대한 자료를 확보하였다. 확보한 자료는 기존의 경험지식 혹은 설계지침에 제안된 분류체계에 따라 재분류하고, 연구지역에 대한 지도화 및 그 분포확률을 계산하였다. 예를 들어 그림 8과 같이 사면경사는 FHWA(미)의 토사종류별 사면안정성과 국토해양부의 절토사면 안정기준을 반영하여 6단계와 각각의 경사범위를 설정하고, 5m 셀로 구분하여 지도화하였다.

4.2 베이지안 네트워크의 학습 및 검증

베이지안 네트워크 학습은 관측데이터로부터 조건부 확률을 업데이트 과정(Jensen, 2001)으로, 산사태 취약

성 분석과 같은 전문가 분류에 적용할 수 있다. 주요한 학습 알고리즘에는 Counting, EM(Expectation-Maximization), Gradient Descent 등이 있다(Netica, 2006). 본 연구에서는 불확실한 자료학습이 가능하고 실용적인 EM 학습알고리즘을 채택하였다.

학습 데이터는 그림 7의 오른쪽 지역을 선정하였고 관측 테이블에 산사태 위험이 높은 위치와 원인을 목록화하였다. 학습이 완료되어 작성된 최종적 베이지안 네트워크 모델은 그림 7의 왼쪽 지역으로 검증하였다. 학습과 검증에 활용한 자료는 산림청(2004)에서 '04-'05년 동안 구축한 국가 산사태 위험지도(1/25,000, 709도엽)를 활용하였다. 이 지도는 사면길이, 경사, 위치, 기반암, 지질 등을 현장 조사하고, 전문가가 부여한 점수를 기초로 경험적인 산사태 취약성 분석을 실시하였다. 특히 산사태 발생현황은 현장조사를 통해 검증이 이뤄졌다.

제안된 베이지안 네트워크 모델은 학습지역을 대상으로 4개 상태와 18개 아크, 그리고 총 590개 CPT가 Netica-API에 적용되어 학습이 이뤄졌다. 최종적으로 그림 9와 같이 산사태 취약성에 대한 네가지 상태 클래스가 21.8%, 37.2%, 18.1%, 22.8%로 사후확률(posterior probability)이 설정되었다.

학습된 베이지안 네트워크 모델은 검증 지역을 대상으로 적용한 결과, 그림 10과 표 1과 같이 산사태 발생

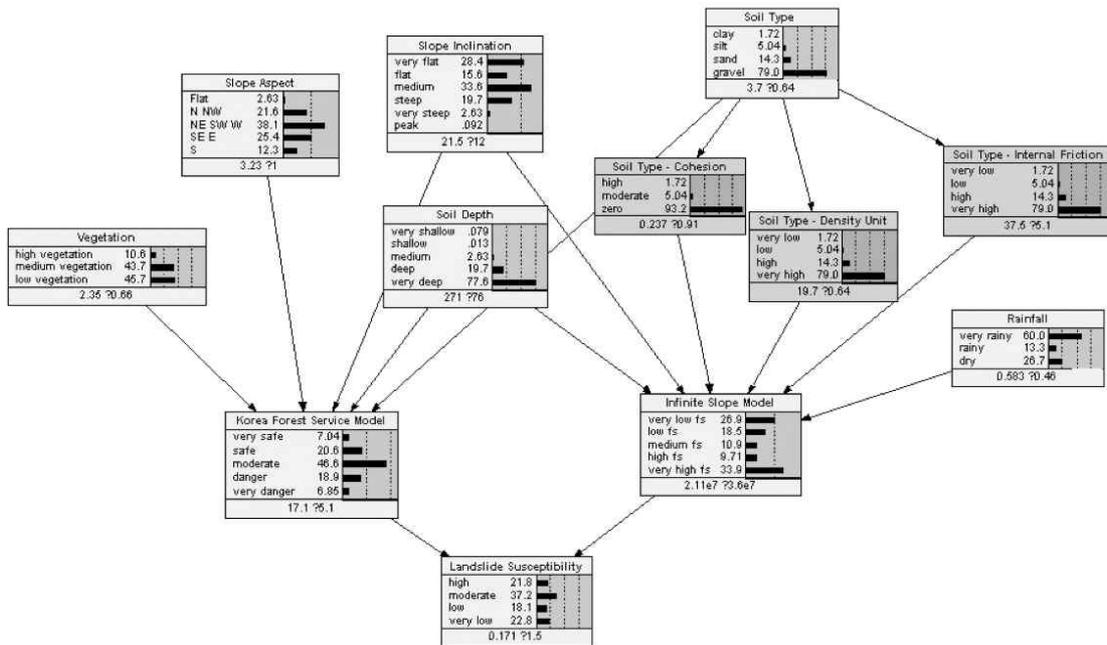


그림 9. 학습된 최종 베이지안 네트워크 모델과 사후확률 분포

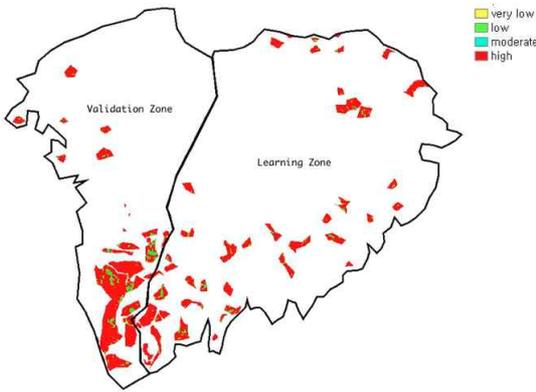


그림 10. 산사태 발생지역에 대한 베이지안 네트워크 학습결과

표 1. 학습/검증 지역의 산사태 취약성 비교

단계	위험 등급	셀 수		영역범위(%)	
		학습 지역	검증 지역	학습 지역	검증 지역
1	very low	2,519	1,328	0.9	0.8
2	low	88,879	50,092	33.2	31.4
3	moderate	62,448	42,054	23.3	26.3
4	High	114,120	66,230	42.6	41.5
합계		267,966	159,704	100.0	100.0

단계	위험 등급	산사태 셀 수		산사태 영역범위(%)	
		학습 지역	검증 지역	학습 지역	검증 지역
1	very low	1	1	0	0
2	low	1,685	2,115	11.3	13.3
3	moderate	32	28	0.2	0.2
4	High	13,188	13,712	88.5	86.5
합계		14,906	15,856	100.0	100.0

범위가 위험지역으로 추론된 지역과 86.5% 일치함을 알 수 있었다.

또한, 제안된 모델에 대한 예측성능의 민감도를 평가하기 위하여 계산된 RMSE(Root Mean Square Error)는 0.432(cut-off 값 0.5)로 유의함을 알 수 있다.

5. 결론

산사태 취약성 지도 작성에 휴리스틱, 통계적, 결정론적, 확률적 방법 등이 소개되었다. 그러나, 현장데이터가 부족하여 신뢰도가 떨어지거나 전문가에 의해 주관적 해석에 의존하는 등의 문제가 있었다. 본 연구는 전문가의 경험과 지식을 반영하면서 지반물성, 공간정보 등 입력자료의 불확실성을 고려할 수 있도록 시멘틱

기술과 베이지안 네트워크를 결합하여 산사태 취약성 분석을 수행하였다.

전문가의 경험과 지식을 온톨로지 모델로 표현하고, DL 추론기법을 통해 베이지안 네트워크의 구조를 자동 생성하였다. 또한, 연구지역에 대한 GIS 입력자료를 기초로 생성한 베이지안 네트워크를 학습시키고, 검증을 통해 유의한 베이지안 네트워크 모델을 제시하였다. 본 연구는 전문가의 경험과 지식을 시멘틱 기술로 베이지안 네트워크에 반영하여 일반 사용자도 전문가 도움없이 손쉽게 광역한 지역의 산사태 취약성 분석이 가능할 것이다.

참고문헌

1. 산림청, 2004, "산사태위험지관리시스템", <http://sansatai.forest.go.kr>.
2. 이상훈, 장병욱, 2009, "시멘틱 공유를 위한 MDA기반 지하공간정보 온톨로지 모델 개발", 지형공간정보학회, 제17권 제1호, pp.121-129.
3. Barredo, J.I., Benavides, A., Hervas, J., Van Westen, C.J., 2000, "Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin", ITC Journal, Issue 1, 2000, pp.9-23.
4. Chandrasekaran, B., Josephson, J.R. and Benjamins, V.R., 1999, "What are ontologies, and why do we need them", IEEE Intelligent Systems, Vol.14, No.1, pp.20-26.
5. Charniak, E, 1991, "Bayesian networks without tears", AI Magazine, Vol. 12(4), pp.50-63.
6. Chen H, Lee CF, 2003, "A dynamic model for rainfall-induced landslides on natural slopes". Geomorphology vol. 51, pp.269-288.
7. Chung, C.F., Fabbri, A.G., van Westen, C.J., 1995, "Multivariate regression analysis for landslide hazard zonation", Geographical Information Systems in Assessing Natural Hazards, pp.107-133.
8. Cooper, G. and Herskovits, E., 1991, "A bayesian method for constructing bayesian belief networks from databases", In Proceedings of Uncertainty in Artificial Intelligence.
9. Chung, C.F. and Fabbri, A.G., 1999, "Probabilistic prediction models for landslide hazard mapping", Photogrammetric Engineering & Remote Sensing, Vol.65, No.12, pp.1389-1399.
10. Cruden, D. M. and Varnes, D. J., 1996, "Landslide types and processes", Landslides: Investigation and Mitigation, TRB Special Report, 247, National

- Academy Press, Washington, pp.36-75.
11. De Vel, O, Liu, N., Caelli, T. and Caetano, T., 2006, "An Embedded Bayesian Network Hidden Markov Model for Digital Forensics", IEEE Intelligence and Security Informatics Conference (ISI 2006), pp.459-465.
 12. Devitt, A., Danev, B., Matusikova, K., 2006, "Ontology-driven Automatic Construction of Bayesian Networks for Telecommunication Network Management", 2nd Int. Workshop : Formal Ontologies Meet Industry(FOMI 2006).
 13. Gritzner ML, Marcus WA, Aspinall R, Custer SG, 2001, "Assessing landslide potential using GIS, soil wetness modelling and topographic attributes", *Geomorphology*, vol. 37, pp.149-165.
 14. Gruber, T., 1992, "A translation approach to portable ontology specification", Knowledge Systems Laboratory, Stanford University, Stanford, CA, Technical Report KSL 92-71.
 15. Helsper, E. and der Gaag, L. C. Van, 2002, "Building bayesian networks through ontologies", In Proceedings of the 15th European Conference on Artificial Intelligence, pp.680-684.
 16. Jensen, F.V., 2001, "Bayesian Networks and Decision Graphs", Springer, New York.
 17. Kalacska, M., Sanchez-Azofeifa, G.A., Caelli, T., Rivard, B.; Boerlage, B., 2005, "Estimating leaf area index from satellite imagery using Bayesian networks", *Geoscience and Remote Sensing, IEEE Transactions on* Vol. 43, Issue 8, pp.1866-1873.
 18. Lam, W. and Bacchus, F., 1993, "Learning bayesian belief networks: An approach based on the mdl principle", *Computational Intelligence*, Vol. 10, pp.269-293.
 19. Lee, S.M. and Abbott, P. A., 2003, "Bayesian networks for knowledge discovery in large datasets", *Journal of Biomedical Informatics*, Vol.36, pp.389-399.
 20. Lu P, and Rosenbaum MS, 2003, "Artificial Neural Networks and Grey Systems for the Prediction of Slope Stability", *National Hazards* Vol. 30, pp.383-398.
 21. Mejia-Navarro M, Wohl EE.,1994, "Geological hazard and risk evaluation using GIS : methodology and model applied to Medellin", *Bulletin of the Association of Engineering Geologists* 31(4), pp.459-481.
 22. Netica-API, Norsys System Corp., <http://www.norsys.com/>
 23. Pearl, J .,1986, "Fusion, propagation and structuring in belief networks", *Artificial Intelligence*, Vol. 29, pp.241-288.
 24. Protégé, Stanford Center for Biomedical informatics Research, <http://protege.stanford.edu>
 25. Jena, Hewlett-Packard Development Company, <http://jena.sourceforge.net>
 26. Stassopoulou, A., Petrou, M., Kittler, J., 1998, "Application of a Bayesian network in a GIS based decision making system", *International Journal of Geographical Information Science* 12 (1), pp.23-45.
 27. Suzen, M.L. and Doyuran, V., 2003, "A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate", *Environmental Geology*.
 28. Van Westen, C.J. ,1993, "Application of Geographic Information Systems to Landslide Hazard Zonation", Ph-D Dissertation Technical University Delft. ITC-Publication Number 15, ITC, Enschede, The Netherlands.
 29. Wallace, C. S., Korb, K., and Dai, H., 1996, "Causal discovery via mml", *The 13th International Conference on Machine Learning*, pp.516-524.