

온톨로지와 군집분석을 이용한 지하공간 정보모델 개발

이상훈^{1*}

Development of Subsurface Spatial Information Model with Cluster Analysis and Ontology Model

Sang-Hoon LEE^{1*}

요 약

지하공간 개발의 증가에 따라 지층단면도 등 다양한 형태로 제공되는 지하공간 정보모델의 신뢰성이 요구되고 있다. 그러나 지반은 근본적으로 불확실하며, 이를 표현하는 정보모델도 자료부족, 해석표준 부재 등의 비통계적 요인과 외부환경 변수라는 통계적 요인으로 불확실성을 가진다. 따라서, 현재의 모델 생성은 고도로 훈련된 전문가에 의해 이뤄지고 있다. 본 연구는 지반공학 전문가의 경험과 지식에서 시맨틱을 추출하고, 이를 온톨로지 모델과 정보량으로 정량화하였다. 정량화한 온톨로지 모델은 군집분석의 클러스터간 거리계산에 적용하여 시맨틱을 고려한 군집분석 방법론을 제안하였다. 본 제안 방법을 실험지역에 적용한 결과 기존 K-Means 방법에 비해 전문가의 해석과 유사한 결과를 도출하였으며, 수작업으로는 어려운 대용량 데이터를 손쉽게 처리하고 3차원 GIS로 가시화가 가능하였다. 본 연구를 통해 지반공학 전문가의 도움 없이도, 그 경험을 고려하면서 대량의 지반정보 데이터를 효과적으로 처리하여 신뢰성 있는 지하공간 정보모델을 생성할 수 있을 것이다.

주요어 : 지하공간 정보모델, 군집분석, 시맨틱, 온톨로지, 3D-GIS

ABSTRACT

With development of the earth's subsurface space, the need for a reliable subsurface spatial model such as a cross-section, boring log is increasing. However, the ground mass was essentially uncertain. To generate model was uncertain because of the shortage of data and the absence of geotechnical interpretation standard(non-statistical uncertainty) as well as field environment variables(statistical uncertainty). Therefore, the current interpretation of the data and the generation of the model were

2010년 11월 3일 접수 Received on November 3, 2010 / 2010년 12월 3일 수정 Revised on December 3, 2010 / 2010년 12월 6일 심사완료 Accepted on December 6, 2010

1 한국건설기술연구원 U-국토연구실 Ubiquitous Land Implementation Division, Korea Institute of Construction Technology

* 연락처자 E-mail : sanghoon@kict.re.kr

accomplished by a highly trained experts. In this study, a geotechnical ontology model was developed using the current expert experience and knowledge, and the information content was calculated in the ontology hierarchy. After the relative distance between the information contents in the ontology model was combined with the distance between cluster centers, a cluster analysis that considered the geotechnical semantics was performed. In a comparative test of the proposed method, k-means method, and expert's interpretation, the proposed method is most similar to expert's interpretation, and can be 3D-GIS visualization through easily handling massive data. We expect that the proposed method is able to generate the more reasonable subsurface spatial information model without geotechnical experts' help.

KEYWORDS : *Subsurface Spatial Information Model, Cluster Analysis, Semantics, Ontology, 3D-GIS*

서 론

지하공간은 인간의 생존과 개발에 필수적이다. 최근 세계적인 도시화로 인한 기존 토지의 부족은 지하공간의 이용을 더욱 가속화시키고 있다. 효과적인 지하공간의 개발을 위해서는 지반공학적 판단과 해석이 필요하며, 이를 위해 시추주상도, 지층단면도, 보간 기법을 이용한 지층도 등 다양한 형태의 지하공간모델이 작성되고 있다. 이러한 정보모델의 신뢰성을 높이는 것은 굴착행위가 이뤄지는 모든 건설공사에서 비용의 감소와 효과적인 사업운영에 필수적이다.

일반적으로 지하공간모델의 생성은 데이터 수집에서 시작된다. 정의된 위치에서 수행된 시추조사 결과를 통해 구조물의 기초가 설치될 암반선을 찾고, 시추와 현장시험을 통해 얻어진 각 토층의 관찰 및 실험정보를 통해 지층의 구성과 강도와 같은 지반공학적 성격을 조사한다. 지반공학자는 조사된 정보의 해석을 통해 시추주상도 혹은 단면도와 같은 지하공간모델을 작성하고, 설계 엔지니어는 이를 구조물 설계에 이용한다. 그러나, 지반은 근본적으로 불확실성을 가지며, 앞서 언급한 지하공간모델 생성과정도 매우 불확실하다. 지하공간모델에 대한 불확실성의 원인을

세 가지로 요약해 볼 수 있다. 첫째, 측정된 자료의 부족(scarcity)로 인해 발생한다. 조사된 시추공만으로 해당 지역을 보간하여 지하공간모델을 생성하기에는 수치공 수가 부족한 경우가 많다. 둘째, 지층을 구분하는 표준방법이 없다. 일반적으로 지층의 구분은 지반공학자의 경험에 의해 이뤄진다. 이는 비통계적인 불확실성(non-statistical uncertainty)을 가져온다. 셋째, 지반은 하중 증가, 지하수위 상승과 같은 외부 요인에 의존하여 물리적인 속성변화가 발생한다. 예를 들어 지하수위의 증가는 토사의 강도를 감소시킨다. 이는 통계적인 불확실성(statistical uncertainty)을 가져온다. 이러한 이유로 인하여 일반적으로 지반의 해석과 지하공간모델의 생성은 고도로 훈련된 전문가를 통해 이뤄지고 있다.

Houlding(1994)은 불확실성의 원인을 위치 혹은 측정기계의 오류에 따른 조사에러의 가능성과 전문가의 경험 혹은 이전지식의 불확실성에 의한 해석에러의 가능성으로 구분하였다. 앞서 언급한 통계학적 및 비통계학적 불확실성은 해석에러에 해당하며, 본 연구에서는 해석의 에러를 최소화하는 것을 주요한 목표로 한다. 기존의 지층의 인식과 지하공간모델의 생성은 여러 전문가 그룹의 해석을 통해 이뤄져, 많은 비용과 시간이 소요

되었다. 본 연구는 비통계학적 불확실성을 해소하기 위하여 전문가의 경험과 지식을 시맨틱 기술의 하나인 온톨로지 방법론으로 표현하고, 통계학적 불확실성 해소를 위하여 데이터마이닝 방법론의 하나인 군집분석 방법론을 적용하였다. 또한, 두 방법의 합리적인 통합을 위하여 온톨로지의 유사도를 정보량으로 정량화하고, 이를 데이터 군집분석 시 거리계산에 적용하였다. 본 연구를 통해 일반사용자는 전문가 그룹의 도움 없이도 전문가의 경험과 관련한 주관적인(subjective) 특성을 고려하면서, 많은 양의 지반공학 데이터로부터 자동적으로 지하공간모형을 생성할 수 있다.

1. 연구방법

본 연구의 진행과정은 그림 1과 같다. 먼저 지반공학 데이터의 해석과 데이터마이닝에 관련한 연구분석을 수행하고, 시추주상도, 현장 및 실내시험 결과를 수집하여 지반정보 DB를 구축하였다. 지반공학자의 데이터 분류 및 해석에 관한 경험과 지식을 바탕으로 시



FIGURE 1. 연구흐름도

맨틱을 추출하고, 이를 온톨로지 언어인 OWL(Web Ontology Language)로 기술하였다. 또한, 온톨로지 모델의 객체 사이의 유사도를 정보량으로 계산하여 정량화하였다. 계산된 정보량을 군집분석 시 군집 여부를 판단하기 위해 작성되는 거리에 포함시켜, 시맨틱을 고려한 군집분석 알고리즘을 설계하였다. 알고리즘을 구축된 지반정보DB에 적용하여 지반공학적 지층을 구분하고, 이를 지형도와 DEM과 결합하여 3D-GIS기반의 지하공간모형을 개발하였다. 마지막으로 기존의 단순 군집분석과 전문가를 통한 결과와 비교 및 분석을 실시하였다.

2. 관련연구

지반, 지질분야의 데이터를 분석하기 위한 데이터마이닝 기법의 하나로 다양한 군집분석 기법이 적용되었다. 데이터마이닝은 데이터에 저장된 많은 데이터로부터 다양한 종류의 쿼리를 수행하는 가장 형식적인 방법이다 (Thuraisingham, 2000). 적절한 정보를 추출하고 주요한 패턴이나 트렌드를 찾아낼 수 있기 때문에, 지하공간 정보모형을 생성하기 위해서는 데이터마이닝이 필수적이다. 데이터마이닝을 통해 조사된 지반정보 샘플의 분류와 숨겨진 지층구조의 결정, 그리고 비이상적인 지층패턴의 인식 및 구조예측이 가능하게 된다. 기존 연구는 지하공간의 수직적인 프로파일을 위한 군집연구가 주로 수행되었다. 유사한 속성의 토사의 영역을 분류하거나, 현장 시험 결과의 군집분석을 통해 토사의 등방성(homogeneity)을 찾아 지층을 구분하였다(이상훈 2008; Hegazy and Mayne, 2002; Balbi and Saboy, 2004). 또한, Zhou and Maerz(2003)는 시추공 데이터의 비연속성을 확인하기 위하여 다변량 군집분석기법을 적용하였다. 이러한 통계적인 데이터마이닝은 상대적으로 간단한 방법으로 많은 변수의 지층을 인식할 수 있는 장점이 있으나, 얻어진 결과의 품질은 수집된 자료의 품질이나 측정주

기에 주로 의존한다는 단점이 있다.

통계적인 데이터마이닝 이외에도 Shahin *et al.*(2001)은 인공지능을 이용하여 토질분류, 지 지력, 침하분석에 적용하였으며, Das and Basudhar(2004)와 Juang *et al.*(2001)은 각각 현장시험의 하나인 SPT(Standard Penetration Test), CPT(Cone Penetration Test) 결과를 인공신경망에 적용하여 대상지반의 분류 및 특성 규명에 관한 연구를 수행하였다. 이외에도 Green *et al.*(2002)은 시추조사 자료를 3차원 GIS에 적용하여 석회암 지하공간의 지도를 작성하였다.

군집분석으로 대표되는 통계적인 방법 혹은 인공지능 방법은 데이터의 불확실성을 최소화 하면서 전문가의 개입 없는 자동적인 방법론 구현을 목표로 한다. 그러나, 두 방법 모두 데이터의 수나 품질에 의존하기 때문에, 비통계학적 불확실성을 해소하면서 좀 더 정확한 결과를 얻기 위해서는 전문가의 지식 혹은 경험에 의존해야 한다.

본 연구는 간단하며 자동적으로 이뤄지는 군집분석의 장점과 함께, 정성적인 전문가의 경험과 지식을 반영한 지층 분류 및 지하공간 모델 생성 방법론을 제안하였다.

지반정보 군집분석 및 시맨틱

1. 지반정보DB와 지반공학적 해석

지하공간을 이루는 토사와 암반은 복잡하며 불확실한 성질을 가진다. 또한, 지반의 이질성은 지반물질 속성의 공간적 변화에 의해 발생한다. 예를 들어 암반에서 암의 강도변화나 점토에서 공극의 변화는 각각 암반과 점토가 이질성을 가지게 한다. 이러한 비균질성(inhomogeneity)과 비등방성(anisotropy)은 지반물질에 새로운 경계를 만들게 된다. 그러나, 불확실한 지반물질에서 이러한 경계를 결정하기 위해 비균질성 여부를 판단하고, 조정하는 것은 많은 교육과 훈련이 필요하다. 따라서, 지반의 해석 및 모델 생성은 표준적인

방법보다는 지반공학자의 경험과 엔지니어링적 판단에 의존하고 있는 실정이다.

현재, 지하공간모델을 좀 더 정확하고 자세히 표현하기 위해 시추주상도, 지층단면도, 지질도, 구조지질도, DEM 등 다양한 자료가 이용된다. 이중에서 일반적으로 건설공사 및 환경조사에서는 시추조사, 실내시험, 현장시험 정보 순으로 그 이용도가 높은 것으로 조사되었다(PEER, 2004). 본 연구에서는 지하공간 모델을 위한 대상 자료로 지반정보DB의 시추조사, 실내시험, 현장시험의 결과를 이용하였다. 지반정보DB에서는 지층의 형상 및 관련 특성을 기술한 시추주상도와 SPT, CPT 등의 현장시험, 그리고 샘플채취 후 실내시험을 수행하여 지반정수를 기록한 시험결과(기초물리 특성, 입도시험, 전단시험, 투수시험, 등)를 포함한다. 또한, 해당 대상지의 위성영상, 지형도 등에 시추 위치정보를 매핑하여, GIS-DB로 구축하였다. 이러한 정보는 웹GIS를 통해 지도상에서 시추공의 위치검색과 텍스트 기반의 속성검색으로 정보를 열람할 수 있다(지반정보DB포털; 이상훈과 장용구, 2007).

본 연구의 지하공간모델은 지반정보DB 중에서 통일분류법(USCS, Unified Soil Classification System)에 의한 지층분류와 지반의 상대적인 강도를 나타내는 SPT정보를 주요한 대상으로 하였다. 먼저, USCS는 지반공학이나 지질분야에서 주로 이용되는 정성적인 토사 분류법으로 흙의 텍스처와 입도크기를 나타낸다. 본 분류법은 지반공학자가 두 개의 기호조합(예:SM)으로 표현한다. 이에 비해 SPT 실험결과는 수치로 표현되는 정량적 데이터이다. SPT는 "중량 64kg의 추를 76cm의 높이에서 자유낙하시켜 시험용 샘플러를 지반에 30cm관입시키는 타격회수"로 정의된다(박인식 역, 1995). SPT의 주요한 목적은 모래나 자갈의 상대밀도의 정도를 측정하는 것이다. 이 방법은 간단하고 저렴하기 때문에 널리 퍼져 있으며, 대략적으로 토사의 강도정수를 추정할 수 있어 지반조건 결정에 좋은 가이드가 된다.

일반적으로 지반공학자는 USCS 분류시스템과 SPT 결과분석을 통해 지반물질의 속성을 정의하고, 정의된 속성을 기초로 해당 지점의 지반물질을 인근지점과 유사성에 따라 하나의 지반공학적 지층단위를 생성한다. 그러나, 분류시스템의 결과만으로는 유사한 지반물질의 경계, 즉 지층을 찾아내기 어렵기 때문에 전문가의 지식에 의한 해석이 이뤄지거나, 통계학적인 군집분석으로 이를 해결하기도 한다.

2. 지반정보 군집분석

군집분석은 데이터를 유사한 특성의 그룹으로 나누는데 이용되는 통계학적 기법이다. 특히, 이전에 알려지지 않은 변수를 그룹화하여 지하공간의 수직적인 프로파일을 작성하거나, 유사한 파라미터를 가진 토사를 분류하는데 유용하다. 그러나, 기존의 군집분석은 데이터의 정량성을 강조하여 여러가지 다른 관점을 반영하지 못하는 문제가 있다(Macskassy *et al.*, 1997). 그리고, 군집 대상이 다양한 차원의 속성을 내포하고 있을 경우, 거리를 찾아내는 과정이 고차원 공간에서 이뤄지기 때문에 군집분석이 어려워지는 문제가 있다(Beyer *et al.*, 1999). 마지막으로 군집분석은 본질적으로 분류된 군집을 설명할 수 없다면 의미가 없다는 약점이 있다. 기존 K-Means 군집분석 혹은 세련된 군집분석 방법도 주관성(subjectivity)과 설명능력(explain-ability)이 부족한 문제가 발생한다. 본 연구는 기존 군집분석의 문제를 극복하기 위하여 지반공학자의 경험과 지식에서 시맨틱을 추출하고, 이를 활용하여 다양한 관점이 포함되는 군집분석 방법을 제안하였다.

3. 지반의 특성분류와 시맨틱 추출

지반 엔지니어는 건축물 혹은 기초 지지에 관련된 공학적 속성에 따라 토질 분류하였다. 현대의 공학적 분류시스템은 현장 관측에서부

터 토사의 공학적 특성과 행위의 기초적인 예측에 적용할 수 있도록 설계되었다. 현재 가장 일반적인 토질분류법은 토사를 조립토(예: 모래, 자갈), 세립토(예: 실트, 점토), 그리고 고유기질토(예: Peat)로 구분한 USCS분류시스템이며, 도로건설을 위한 토사와 토사혼합물의 분류를 위한 AASHTO 토사분류법이 있다. 이외에도 토양자원 보전을 위해 토사의 내부적인 속성(Soil Morphology), 행위, 기원 등에 의한 분류가 있다. 대표적인 분류법으로 "USDA Soil Taxonomy", "World Reference Base for Soil Resource" 등이 있다. 그림 2는 각 토사분류법일 비교한 것이다.

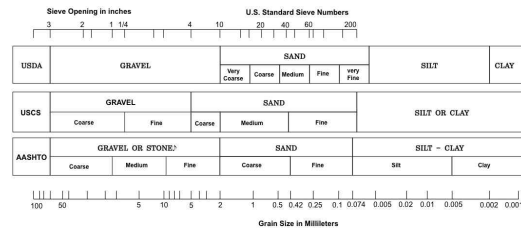


FIGURE 2. USDA, USCS, AASHTO 토사분류법 비교

그림 2의 비교에서 보듯이 각 분류법은 공통적으로 체분석 시험을 통해 자갈, 모래, 실트, 점토를 구분하고 있지만, 그 범주는 다르다. 따라서 기초적인 토사의 분류에서도 목적을 고려하여 전문가의 판단이 수반되어야 한다. 본 연구에서는 이러한 토사분류법의 분류와 각 개체의 관계를 고려하여 지반분류시스템에 관한 시맨틱을 온톨로지 모델로 표현하였다.

두 번째로 SPT를 통해 얻어진 지반의 상대적인 강도지수인 N값은 다양한 연구자들의 경험에 의해 토사와의 상관관계가 제안되었다. 표 1은 Terzaghi와 Peck은 점토를 N값이 30이하에서 연경도를 6단계로 구분하였으며, 표 2는 Peck과 Meyerhof가 모래를 N값이 50이하에서 상대밀도를 5단계로 구분한 것이다(박인식 역, 1995).

TABLE 1. 점토의 N값/압축강도와 연경도 비교

연경도	매우 약함	약함	중간	강함	매우 강함	견고함
N값	0-2	2-4	4-8	8-15	15-30	30 이상
압축강도	0.25 이하	0.25-0.5	0.5-1.0	1.0-2.0	2.0-4.0	4.0 이상

TABLE 2. 모래의 N값, 상대밀도, 마찰각 비교

N값	상대밀도	내부마찰각	
		Peck	Meyerhof
0-4	very loose : 0.0-0.2	28.5이하	30 이하
4-10	loose : 0.2-0.4	28.5-30	30-35
10-30	medium : 0.4-0.6	30-36	35-40
30-50	dense : 0.6-0.8	36-41	40-45
50이상	very dense : 0.8-1.0	41이상	45이상

이렇게 토사분류와 SPT의 N값은 다양한 연구자들의 경험에 의해 그 특성을 분류하였다. 그러나, 단순히 동일한 USCS분류값을 가진 토사도 지하수위에 따른 함수비 변화 혹은 지중압력 등 외부영향에 따라 서로 상이한 특성과 행동을 지니게 된다. 따라서, 정확한 토사의 분류나 지반의 비균질성으로 지층을 구분하기 위해서는 USCS 토사분류 뿐만 아니라, 강도특성을 나타내는 SPT의 N값 등 다양한 경험적 분류가 고려되어야 한다.

4.지반분류시스템 온톨로지와 정보량 계산

온톨로지는 “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 표현할 수 있는 대표적인 시맨틱 기술의 하나이다. 본 연구에서는 앞서 언급한 토사 및 강도 분류 등에 관한 지반엔지니어의 경험과 지식을 온톨로지 모델로 표현하여, 기존의 분류에서 표현하지 못한 지반 도메인에 내재된 시맨틱을 표현하였다. 온톨로지 언어는 W3C에서 표준으로 채택한 OWL을 적용하였다. 이는 기존의 RDF(Resource Description Framework)보다 표현성이 뛰어나며, 단순한 정보 표현이외에 서술논리를 지원하여 추론도 가능하다. 본 연구에서는 추론이 가능한 OWL-DL을 채택하였다. 온톨로지 모델의 작성은 이상훈과 장병욱(2009)이 활용한 방법으로 UML툴로 UML-Profiles을 작성하고, 이를 OWL로 변환하여 개념과 관계를 설정하였다. 여기에 온톨로지 편집툴인 Protege3.4를 이용하여 오류수정과 인스턴스를 입력하였다. 그림 3과 그림 4는 Protege-OWLviz로 작성된 토사분류 시스템과 지층강도 시스템의 온톨로지 모델이다.

군집분석에서 온톨로지를 이용하기 위해서는 온톨로지의 개념 사이의 거리를 계산해야 한다. 온톨로지에서 개념 사이에는 보유한 기능에 따라 가중치가 필요하며, 이를 정보량(IC, Information Content)으로 표현할 수 있다(Resnik, 1995). 일반적으로 자식개념은 부모개념에 비해 특정기능에 관한 세부적인 기능을 제공한다. 개념이 상세할수록 자식개념과 부모개념 사이에 많은 개념이 정의되어 밀집도가 높아진다. 정보량은 상세화정도를 나타내며, 온톨로지에서 자식 개념으로 갈수록 높은 값을 가지게 된다. 정보량을 구하기 위해서는 먼저 식 (1)에 따라 빈번도 확률(FP, Frequency Probability)을 계산하고, 식 (2)를 통해 정보량을 계산한다. 그림 3와 4는 온톨로지에서 각 개념별로 계산된 정보량을 표시하였다. 구해진 정보량은 식 (3)을 통해 각 개념사이의 거리로 계산할 수 있다(Lin, 1998).

$$FP = \frac{\sum Count(n)}{\sum Count(n) \text{ of Parent Concept}} \quad (1)$$

$$IC = -\log_{10}(FP) \quad (2)$$

$$\text{sim}(c_i, c_j) = \frac{2 \times IC(c_i \cap c_j)}{IC(c_i) + IC(c_j)} \quad (3)$$

이 방법은 개념사이에 유사도를 예지 (Edge)의 수로 측정하는 방법이 아니라, 노드 가중치 차이로 거리를 측정할 수 있다. 공통 부모개념이 깊을수록 두 개념의 유사도가 높아지며, 공통부모 개념으로 부터 두 개념이 멀리 떨어져 있을 수록 두 개념의 유사도는 낮아지게 된다. 식 (3)의 유사도 측정값은 정규화되어 0과 1사이의 값을 가진다.

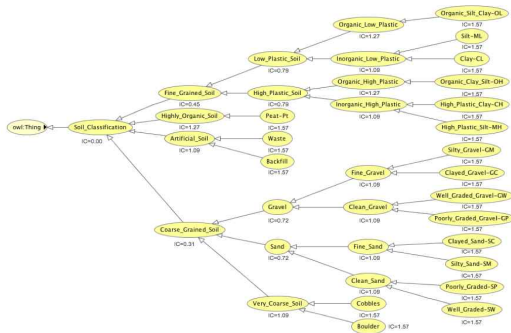


FIGURE 3. 토사분류시스템 온톨로지 모델과 정보량

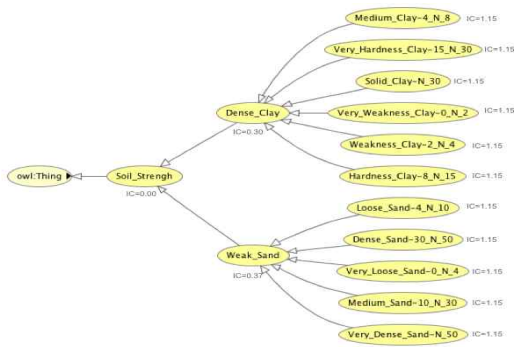


FIGURE 4. 토사강도시스템 온톨로지 모델과 정보량

온톨로지를 활용한 군집분석

1. 개요

그림 5와 같이 온톨로지를 활용한 군집분석은 시추조사를 통해 얻어지는 지층의 분류 및 속성정보, SPT 현장시험을 통해 얻어지는 상대적인 지반 강도정보 N값, 그리고 실내시험으로 얻어지는 토사의 상대밀도 등 물성값을 입력자료로 이용한다. 또한, 앞 절에서 작성된 토질분류 및 지반강도와 관련한 시멘틱을 나타내는 지반공학 온톨로지를 입력한다. 여기서 온톨로지 모델 내 개념들 사이의 거리차이 계산은 추론엔진인 Pellet2.0을 통해 이뤄지게 된다. 각 입력자료는 초기 클러스터와 온톨로지 클러스터와의 거리 비교를 통해 군집분석이 수행된다.

분할적인 군집분석 방법에는 K-Means와 ISODATA 알고리즘이 대표적이다. K-Means는 계층적 클러스터링처럼 초기에 병합이 부적절할때 패턴을 다시 재배열할 없다는 문제가 있다. ISODATA 알고리즘은 K-Means와 같이 K개의 중심을 가지지만, 클러스터 개수를 K로 한정하지 않아서, 초기 병합의 문제점을 해결할 수 있다(Tou and Gonzalez, 1974). ISODATA 알고리즘에서는 클러스터의 평균과 표준편차를 계산하여 클러스터를 분리, 병합, 삭제한다. 이를 통해 가장 적합한 클러스터의 수를 산정하거나 자유자재로 클러스터의 개수를 변환할 수 있다. 또한, 대용량의 데이터의 경우 수렴속도가 높다는 장점을

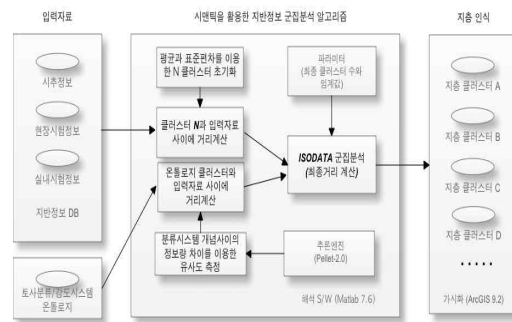


FIGURE 5. 온톨로지를 활용한 군집분석

가진다. 본 연구에서는 ISODATA 군집방법을 기초로 온톨로지 모델의 정보량과 결합하였다.

2. 온톨로지기반 군집화 과정

본 연구에서는 기존 분할적 군집분석 방법에 추가적으로 정보량을 이용하여 온톨로지 클러스터 중심과의 거리 차를 총합하여 계산한다. 클러스터는 대상 변수에 대한 클러스터 중심뿐만 아니라, 대상 온톨로지에 대한 클러스터 중심을 하나 더 가지게 된다. 따라서, 일반 군집화 과정은 $dist_{init}(X, x)$ 로 클러스터 중심과 미지값(x)와의 유클리드 거리를 통해 구해지지만, 온톨로지 클러스터의 중심 X 와 x 사이의 거리를 계산하는 $dist_{ont}(X, x)$ 를 포함하여, 식 (4)와 같이 최종적인 거리, 즉 유사도를 계산한다. 여기서 $dist_{ont}(X, x)$ 는 그림 3과 4에서 제시한 각 개념과 정보량을 기초로 입력값이 속하는 개념과 중심 개념 사이에 유사도 차이를 정보량으로 환산한다. 유사도 차이는 Pellet추론엔진으로 입력된 개념의 부모 개념을 추론하여 식 (3)으로 계산된다.

$$dist_{TOTAL}(X,x) = \alpha \times dist_{INT}(X,x) + (1-\alpha) \times dist_{ONT}(X,x) \quad (4)$$

또한, 사용자는 α 를 통해 각 거리의 비중을 고려할 수 있다. 예를 들어 대상지에 대한 전문가의 특정한 경험 혹은 지식이 존재하여 온톨로지 모델로 이를 제공할 경우, 가중치 α 를 줄여 온톨로지에 대한 비중을 감소시킬 수 있다. 또한, 자료의 분산이 크거나 정성적인 자료가 부족하여 적합한 온톨로지 모델을 작성하기 어려운 경우 가중치 α 를 높여 입력 자료와 파라미터 사이의 상대적인 거리에 대한 비중을 증가시킬 수 있다.

적용 및 분석

1. 실험적용

개발된 방법론은 그림 6의 실험지역을 대상

으로 적용하였다. 지반정보DB로부터 얻어진 부산광역시 명지동 인근의 “명지-녹산공업단지개발(1992년)”의 27개 시추공과 413회 SPT시험결과, “명지주거단지개발 (2002년)”의 33개 시추공과 1366회 SPT시험결과, “명지대교건설(2005년)”의 55개 시추공과 2,369회 SPT시험결과로 각각 다른 시기의 총 115개의 시추공과 4,148개의 SPT시험결과를 주요한 대상으로 하였다. 본 지역은 낙동강 하구의 델타지역으로 주로 백악기 화성암류가 분포하며, 지형은 해발고도가 높지 않으나, 주위의 산사면은 경사가 가파른 편이다. 또한, 공업단지 및 주거단지 개발이 오랜기간 동안 이뤄져 수십년에 걸쳐 많은 지반정보 데이터가 생산되었으며, 수작업으로 처리하여 많은 시간과 비용이 소요되었다.



FIGURE 6. 부산 명지동 실험지역 위성영상(IKONOS, 2007)

온톨로지 모델을 활용한 군집분석과 지하공간모델 가시화를 위하여 범용적인 수치해석 프로그램인 Matlab 7.6과 ArcGIS 9.2 ArcSCENE을 이용하였다. 지반정보DB로부터 실험지역의 데이터 중 먼저 시추공정보에서 지층 위치값($X, Y, From_depth, To_depth$)와 USCS분류값(USCS), 지하수위(WL), 표고(EL)를 불러오고, SPT 현장시험정보에서 SPT시험의 위치값(X, Y, Z)과 타격회수 N 값(N), 관입깊이(Depth)를 불러온다. 그리고, 그림 3과 그림 4의 토사분류 및 강도에 관한 온톨로지 모델을 정보량과 함께 입력한다. SPT시험이 수직으로 1.5m씩 동일한 간격으

로 조사되기 때문에 SPT의 N값을 기준 단위로 하여 USCS분류값을 추가한다. 또한, 개발한 온톨로지 모델에 해당하는 정보량을 추가하여 군집분석 알고리즘이 수행된다. 먼저 기존 군집분석에 따라 N값에 대한 클러스터 중심을 찾고, 할당된 정보량(IC)에 따른 온톨로지 클러스터 중심을 찾아 식 (4)에 따라 대상값의 최종적인 유사도 거리를 계산하여 가장 가까운 클러스터로 대상 값이 할당된다. 본 실험에서는 데이터만을 이용한 군집분석과 작성한 온톨로지의 영향을 동일하게 설정하기 위하여 가중치 α 를 0.5로 하였다. 군집분석 결과는 x,y,z 좌표형태로 구성되어 DEM과 수치지형도와 손쉽게 결합되어 그리드형태의 3차원 지하공간모델을 구축하였다. 이 결과는 그림 7과 같이 3차원 GIS로 가시화하였다. 해당지역은 맵핑지인 관계로 수치지도가 제대로 제작되지 못하여 위성영상으로부터 추출된 DEM을 이용하여 지하공간모델을 생성하였다.

2. 결과분석

결과의 분석을 위하여 비통계학적 불확실성 해소를 위해 이용되는 지반공학 전문가의 지층해석과 자동적인 군집분석에 의한 지층해석, 그리고 군집분석에 전문가 지식을 결합한 제안 방법에 대한 결과를 지층 단면도 형태로 그림 8과 같이 비교하였다.

그림 8의 첫 번째, 두 번째 그래프에서 알

수 있듯이, 심도 0-25m에서는 제안 방법이 K-Means에 의한 방법에 비해 전문가 해석과 거의 일치하였다. 또한, 심도 30-40m 사이 지층의 K-Means에 대한 결과는 SPT N값이 급격하게 증가 등의 원인으로 풍화암(Weather Rock, WR)으로 분류되었다. 그러나 제안 방법의 결과는 실트(Silt, SM)로 분류되며, 이는 지반공학 전문가의 결과와 동일한 것으로 40m이하 층이 다시 실트 층임을 볼 때 제안한 방법이 옳음을 알 수 있다. 즉, 단순히 값만을 가지고 군집을 분석하는 K-Means 기법은 대상지층이 실트층임에도 불구하고, 강도값이 풍화암층과 유사하여 이를 풍화암 군집으로 병합시키는 오류를 발생시킨 것이다. 실제로 지중압력 혹은 지하수 등으로 인해 압밀(Consolidation) 현상에 의해 실트층이 높은 강도를 가질 수 있는 현상을 간과한 것이다. 이외에도 K-Means기법은 40~50m 사이를 점토와 모래의 복합토층 혹은 이상치(Outlier)로 분류하였으나, 제안 방법은 전문가 해석과 유사하게 두 가지 지층으로 구분하였다.

지층 단면도에 대한 비교 이외에 명지주거단지 내에 존재하며 연약지반처리를 수행한 BH001에서 BH020까지의 20개 시추공을 비교분석한 결과, 전문가 해석결과에 대한 K-Means 방법의 RMSE는 0.8563이고, 전문가 해석결과에 대한 제안방법의 RMSE는 0.2469로, 제안 방법

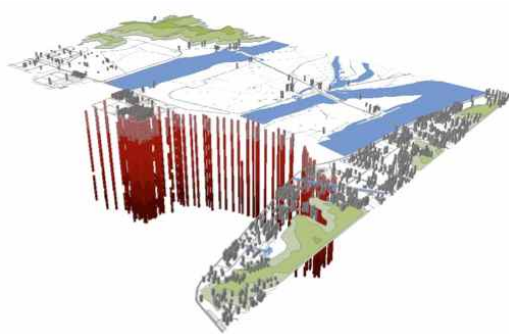


FIGURE 7. 지하공간모델 3차원 가시화(ArcSCENE)

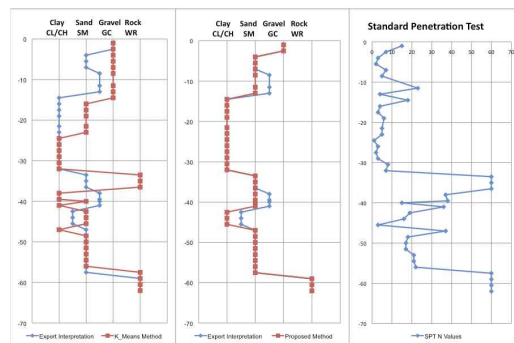


FIGURE 8. 지층분류에 대한 전문가, K-Means, 제안방법의 비교(SPT 강도정보 참조)

이 기존 통계적인 K-Means방법에 비해 정밀하게 군집분석을 수행함을 알 수 있었다.


K-Means와 같은 기존 군집분석 방법은 대용량의 데이터를 자동적으로 분류할 수 있다는 장점이 있지만, 렌즈(lense)지역을 제대로 표현하지 못하고 동일한 N값 혹은 USCS 분류의 토사라도 지하수위 등으로 인해 발생하는 이질적인 특성을 구별하지 못하는 단점이 있다(그림 8의 심도 30-40m 지점). 그러나, 제한한 시맨틱을 활용한 군집분석의 경우에는 지반공학자의 지식을 온톨로지와 정보량으로 표현하여, 지반공학자가 해석한 결과와 가장 근접한 결과를 보인다. 또한, 단순한 값에 의한 통계적인 군집 분류가 아니라, 표현된 시맨틱으로 군집분석의 결과를 설명할 수 있는 장점을 지닌다.

결 론

지하공간 개발의 증가에 따라 지층단면도 등 다양한 형태의 지하공간모델의 신뢰성에 대한 요구가 증가되었다. 지하공간모델의 생성은 시추조사와 각 토층에 해당하는 실험을 통해 지층의 구성과 지반공학적 성격이 조사된다. 조사된 정보를 통해 지층단면도 혹은 시추주상도 형태의 지하공간모델이 작성된다. 그러나, 이러한 모델의 생성에는 자료의 부족뿐만 아니라, 지층 해석에 관한 표준방법이 없어 발생하는 비통계적 불확실성과 하중, 지하수위 변화 등 환경에 따른 물리적인 속성변화에 따른 통계학적 불확실성이 존재한다. 따라서, 기존의 지하공간모델의 작성은 고도로 훈련된 전문가를 통해 이뤄져 많은 시간과 비용이 소요되었다.

본 연구는 비통계학적 불확실성을 해소하기 위하여 전문가의 경험과 지식을 대표적인 시맨틱 기술의 하나인 온톨로지로 표현하였다. 또한, 통계학적 불확실성 해소와 대용량 자료 처리에 손쉬운 군집분석을 적용하였다. 두 방법론을 유기적으로 통합하기 위하여 온톨로지 모델의 유사도를 정보량으로 정량화하여 군집

분석의 유사도 거리계산 시 총합하는 시맨틱을 활용한 군집분석 알고리즘을 제시하였다. 본 방법론을 활용하여 실험지역 지반의 지층을 구분하고, 지형도, 위성영상 등과 결합하여 3D-GIS기반 지하공간모델을 구축하였다. 제안한 방법은 기존 K-Means 방법보다 전문가 해석과 유사한 결과를 가지며, 불확실성 혹은 시맨틱을 손쉽게 표현할 수 있는 장점이 있다. 또한, 전문가 해석에 비해 대용량 데이터 처리, 3D-GIS 표현, 그리고 다른 GIS자료와도 통합분석이 손쉬운 장점을 가진다.

향후에는 개발된 지반분류 및 강도에 관한 온톨로지 모델을 좀 더 상세화하고, 온톨로지 모델 내 개념사이의 관계를 확률로 표현하여 시맨틱 유사도를 좀 더 보강할 계획이다. 이를 통해 좀더 정밀하고 설명 가능한 군집분석 방법론으로 발전시킬 것이다. 

참고문헌

- 박인식 역. Roy E. Hunt. 1995. 지반조사핸드북(Geotechnical Engineering Investigation Manual). 엔지니어스. 76-77쪽.
- 이상훈. 2008. 클러스터링과 지구통계학 기법을 이용한 지하공간정보 모델 생성시스템 개발. 한국지리정보학회지 11(4):64-75.
- 이상훈, 장병욱. 2009. 시맨틱 공유를 위한 MDA기반 지하공간정보 온톨로지 모델 개발. 지형공간정보학회지 17(1):121-129.
- 이상훈, 장용구. 2007. 지반정보 분석 및 평가를 위한 웹기반 지리공간정보 시스템 개발. 한국지리정보학회지 10(4):142-152.
- 지반정보DB포털 <http://www.geoinfo.or.kr>
- Balbi, D.J. and F. Saboy. 2004. CPTU-soil profile interpretation based on similarity concept. Proc. 2nd International Site Characterization Conference. Porto, Portugal.

- Beyer, K., J. Goldstein, R. Ramakrishnan and U. Shaft. 1999. When is 'nearest neighbor' meaningful. Proc. of ICDT-1999. Jerusalem, Israel. pp.217-235.
- Das, S.K. and P.K. Basudhar. 2004. Site characterization of an alluvial site using artificial neural network. International Workshop on Risk Assessment in Site Characterization and Geotechnical Design, Bangalore.
- Green, A.J., J.W. Marken, C.E. Alexander and C.S. Alexander. 2002. Karst unit mapping using geographic information system technology. Mower Country, Minnesota, USA, Environmental Geology 42(5):457-461.
- Gruber, T. 1992. A Translation Approach to Portable Ontology Specification. Knowledge Systems Laboratory, Stanford University, Stanford, CA, Technical Report KSL, 92pp.
- Hegazy, Y.A. and P.W. Mayne. 2002. Objective site characterization using clustering of piezocone data, Journal of Geotechnical and Geoenvironmental Engineering 128(12):986-996.
- Houlding, S.W. 1994. Uncertainty, Sampling Control and Risk Assessment. In: Houlding, S.W. (ed.), 3D Geoscience Modeling-Computer Techniques for Geological Characterization, Berlin: Springer-Verlag, pp.185-200.
- Juang, C.H., T. Jiang and R.A. Christopher. 2001. Three dimensional site characterization: neural network approach. Geotechnique 51(9):799-809.
- Lin, D. 1998. An information-theoretic definition of similarity. Proc. of the 15th International Conference on Machine Learning.
- Macskassy, S.A., A.B. Banerjee, B.D. Davison and H. Hirsh. 1998. Human performance on clustering web pages: a preliminary study. Proc. of KDD-1998, New York, NY, USA, pp.264-268.
- PEER(Pacific Earthquake Engineering Research). 2004. Archiving and Web Dissemination of Geotechnical Data : Development of a Pilot Geotechnical Virtual Data Center. Consortium of Organizations for Strong-Motion Observation Systems.
- Resnik, P. 1995. Using information content to evaluate semantic similarity in a taxonomy. International Joint Conference for Artificial Intelligence 1995. pp.448-453.
- Shahin, M.A., M.B. Jaksa and H.R. Maier. 2001. Artificial neural network applications in geotechnical engineering. Australian Geomechanics 36(1):49-62.
- Thuraisingham, B. 2000. Data Mining Technologies, Techniques, Tools, and Trends. John Wiley & Sons, Inc.
- Tou and Gonzalez. 1974. ISODATA (Iterative Self-Organizing Data Analysis) Technique.
- Zhou, W. and N.H. Maerz. 2003. Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Computers & Geosciences 28:827-839. 