시추공 정보의 클러스터링 기법을 이용한 지반 분석 시스템의 개발

이 규 병[1] 김 유 성
인하대학교 전자계산공학과
g1981770@inhavision.inha.ac.kr yskim@inha.ac.kr

Development of a Subsurface Exploration Analysis System Using Clustering Technique on Bore-hole Information

Kyu-Byoung Lee[1] Yoo-Sung Kim
Dept. of Computer Science & Engineering, Inha University

요 약

본 논문은 지반조사 정보 시스템이 관리하고 있는 시추공의 지층 구성 정보를 바탕으로 시추공을 클러스터링하여 지반의 구성을 분석하는 시스템의 설계 및 구현을 소개한다. 지반조사 정보시스템은 지반 시추공의 현장조사 및 실내시험을 통해 얻는 시추공 정보를 체계적으로 관리하여 정보의 재활용을 돕기 위한 데이터베이스 시스템이다. 본 논문에서 제안된 지반 분석 시스템은 지반조사 데이터베이스의 시추공 정보를 이용하여 지반이 가지고 있는 특성 정보를 추출하여 유사한 특성을 갖는 시추공의 집합으로 클러스터링하여 사용자에게 지반 구성 정보를 제공하는 시스템이다. 본 시스템에서 사용된 클러스터링 기법은 지반조사 데이타가 갖는 지반의 구성 요소 및 각 지층의 구성비를 통해 유사한 특성을 지니는 지반들을 그룹화하고, 그룹된 지반들의 특성을 찾는다. 그룹된 지반들의 분석을 통해 찾아진 유사한 특성은 지도상에 가시화함으로서 사용자에게 해당 지형에 대한 지반 특성을 제공한다.

1. 서 론

구조물을 설계하고 시공하기 위하여 구조물을 지지할 지반의 성질을 조사해야 한다. 지반조사 데이타는 지구를 구성하고 있는 지반에 관련된 데이타로서 비균질하고 비정형화된 특성을 갖는다. 매년 많은 지반조사의 결과로 대량의 지반 조사 자료들이 발생한다. 그러나, 조사 자료들은 보고서의 형태로 보관되거나 소멸되어 차후 인근 지역 혹은 유사 구조물의 시공을 위한 자료로 활용되지 못하고 있다. 최근에 이르러, 지반조사 자료의 재활용을 위해 지반조사 데이타의 중요성을 인식하고, 데이타베이스를 구축하는 실정이다[1].

지반조사 데이타는 구조물을 시공하거나 설계하기 위해 이용될 뿐만 아니라, 지층의 구성이나 토질의 종류 등 지반을 구별하는 정보로 이용될 수 있다. 즉, 지반조사 데이타는 실제 지반을 뚫어서 현장실험, 실내실험, 육안관찰 등을 통해서 얻어진 결과로서 지질도나 토양도([2,3])보다 현장의 특성을 정확하게 반영한다. 지반조사 데이타에서 비균질한 지반 특성을 고려하여 서로 구별되는 특징을 추출하면, 현장의 특성을 정확하게 반영한 지반 분석이 이루어지게 된다. 분석된 자료들은 다른 지반조사방법 및 범위 설정시 이용되거나, 실시되는 지반조사량을 낮추게 하고, 현장의 구성 지반의 유추를 가능케 한다.

본 연구에서는 방대한 자료를 갖는 지반조사 데이타에 자료들 간의 특성을 파악하기 위해 시추공 정보의 클러스터링 기법을 이용하여 유사한 지반끼리 그룹화하고, 그룹 지어진 지반의 특성을 알아내어 이를 가시화함으로써 지형, 지질 등 지반에 관한 유용한 정보를 제공할 수 있는 지반 분석 시스템을 설계 및 구현 한다. 지반 분석 시스템은 지반조사 데이타 특성을 고려하면서도 대용량의 데이타에 효율적으로 클러스터링이 이루어지도록하여, 지반조사 데이타에서 얻을 수 있는 정보를 극대화한다.

본 논문의 구성은 다음과 같다. 2절은 관련 연구로서 지반조사와 본 시스템에서 사용되는 지반조사 정보시스템에 관해 알아보고, 3절은 지반 분석 시스템과 지반조사 데이타에 사용된 클러스터링 방법과 클러스터링으로 얻어진 특성의 표현에 관해 설명한다. 마지막으로 4절에서 결론 및 향후연구 계획으로 본 논문을 맺는다.

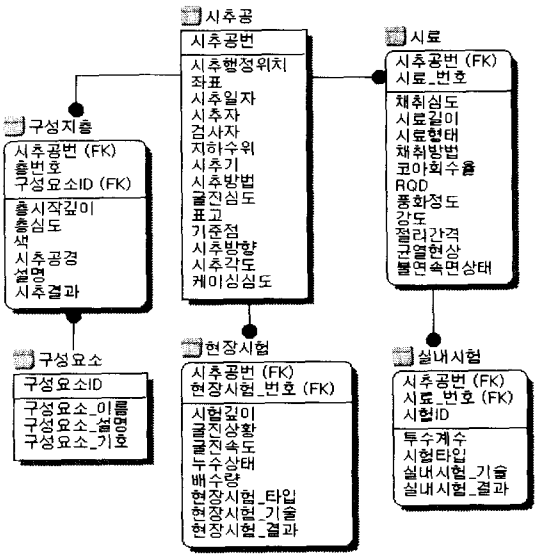
2. 관련 연구

2.1 지반조사 정보 시스템

지반은 구조물의 기초로서, 암석지반(암반)을 포함하는 것, 혹은 토목구조물을 설치하거나 토목공사에서 굴착의 대상이 되는 지구 표면층 부분을 일컫는다[1]. 지반조사는 구조물을 설계하고 시공하기 위하여 구조물을 지지할 지반의 성질을 조사하는 작업으로 기능적인 구조물을 가급적 안전하면서도 경제적으로 건설하며 유지 관리하는데 필요한 기초자료, 즉 지반의 특성 및 지층의 상태, 침하량 산정에 필요한 기본자료 등을 준비하는데 사용된다[1,4].

지반조사로부터 얻어진 자료는 체계적이며 유기적인 관리 및 제공 서비스를 하기 위해 지반조사 정보시스템을 구축하여 사용하고 있다. 한국건설기술연구원([1,5])에서 제안한 표준 지반조사 양식을 기반으로 한 지반조사 정보시스템의 지반조사 데이타베이스의 E-R 다이어그램은 [그림 1]과 같다. 구축된 데이타베이스에는 시추공, 구성지층, 구성요소, 시료, 현장시험, 실내시험의 테이블이 있고, 특히 구성지층과 구성요소 테이블에는 지반을 구성하는 각 지층과 깊이, 지층의 토질 등의 항목으

로 이루어져 있어, 지반을 분류하는데 필요한 정보로써 사용하게 된다. 시료, 현장시험, 실내시험 테이블의 자료들은 해당 지반의 실험결과로 세부적인 성질을 알 수 있는 자료가 된다.



[그림 1] 지반조사 데이타베이스의 E-R 다이어그램

지반조사 정보시스템은 지층의 구성이나 토질의 종류 등을 표시하는 정보를 체계적으로 관리하며 이러한 정보들은 현장의 특성을 정확하게 반영하기 때문에 이 데이타들을 유용하게 활용할 수 있는 분석 시스템이 필요하지만, 현재까지 개발된 지반 분석 시스템은 없는 상태이다.

본 연구에서는 한국건설기술연구원([1]) 지반조사 정보시스템의 데이타를 이용하여 지반 분석 시스템을 개발한다.

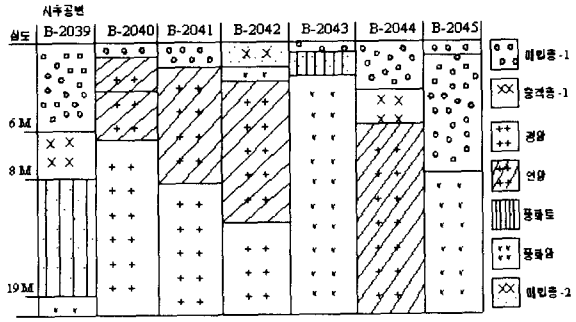
3. 시추공 클러스터링 기법을 이용한 지반 분석 시스템

3.1 지반 분석 시스템의 개요

시추공 정보의 클러스터링 기법을 이용한 지반 분석 시스템은 지반이 가지고 있는 특성 정보를 추출하여 사용자에게 지반, 지형 등 유용한 정보를 제공하기 위한 시스템으로 지반이 가지고 있는 특성을 추출하기 위해서는 지반조사 정보 시스템의 구성지층과 지층의 구성비 데이타를 이용한다.

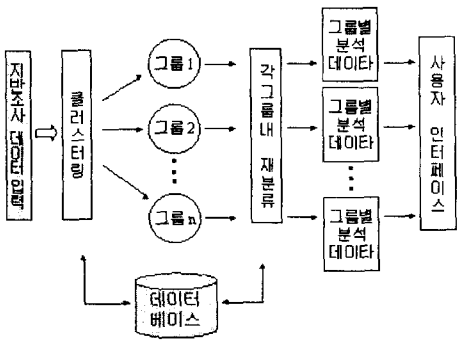
지반조사에 의해 알 수 있는 구성지층은 [그림 2]와 같다. 각 지반의 구성지층은 [그림 2]와 같이 서로 다른 지층으로 이루어져 있고, 동일한 지층 구성을 갖더라도 구성비가 다를 수 있다. 비균질하고 비정형화된 지반에서 특성을 찾는다면 사용자에게 지반, 지형 등 유용한 정보로 사용될 수 있다.

[그림 2]에서 시추공번 B-2040, B-2041, B-2042는 서로 유사한 지층으로 이루어진 것을 알 수 있다. 유사한 지층으로 구성된 지반들을 그룹화하면 그룹이 나타내는 유사한 특징을 찾을 수 있다. 본 시스템은 지층들의 특성을 찾아내고자 시추공 정보의 지층구성 정보 클러스터링 기법을 이용해 유사한 지층들로 이루어진 지반으로 분류하여 그 특성을 찾고, 이 특성을 이해하기 쉽고, 다차원적인 판단을 할 수 있는 정보로 제공한다.



[그림 2] 지반을 구성하는 지층의 예

본 연구에서 개발한 시스템의 구성도는 [그림 3]과 같다. 지반조사 데이타들을 분류하기 위한 클러스터링 단계와 재분류 단계를 거쳐 재분류된 데이타들은 구별되는 특성들로 도출하게 된다. 이 특성들은 사용자 인터페이스 부분에서 지도상에 표현하여 사용자에게 지반, 지형 등의 유용한 정보로서 가시적인 해석을 가능하게 한다. 시추공 클러스터링 기법을 이용한 지반 분석 시스템은 데이타베이스 관리 시스템으로 Oracle8, 프로그래밍 언어로 비주얼 C++를 사용하여 WindowsNT4.0 환경하에서 구현되었다



[그림 3] 지반 분석 시스템 구성도

3.2 시추공 정보를 이용한 클러스터링 기법

지층구성 정보를 클러스터링 기법을 이용한 지반 분석 시스템에서는 유사한 지층으로 구성된 지반의 특성을 추출하기 위해 클러스터링 기법([6])을 사용한다. 본 시스템에서는 <알고리즘 1>과 같은 클러스터링 알고리즘을 이용하여 지반 특성이 서로 구별되도록 그룹화한다.

<알고리즘 1>은 대용량 데이타에 효과적인 K-Means 클러스터링 알고리즘([6,7,8])을 변경하여 지반조사 데이타베이스에 구성된 시추공 정보에서 지반을 구성하고 있는 지층의 종류와 각 지층의 구성비에 따라 유사한 것끼리 분류한다. 비균질한 지반을 서로 구별할 수 있는 특성을 추출하기 위해서는 지반이 구성되어 있는 지층 종류와 구성비를 통해 이 지층들로 이루어진 지반을 대표할 수 있는지 여부를 판단한다. 지반을 구성하고 있는 지층의 구성비는 지반을 이루고 있는 지층이 그 지반을 대표할 수 있는 만큼의 비율을 가지고 있는지를 판단하는 기준이 되며, 사용자의 입력값(지반의 깊이)에 따라 구성비는 상이

하게 나타나게 된다.

<알고리즘 1> 시추공 정보 클러스터링 알고리즘

step 1. 대표 오브젝트의 개수 k를 입력 받는다.
step 2. 구성비를 산출하기 위한 지반의 깊이를 입력 받는다.
step 3. 랜덤방법으로 대표 오브젝트의 개수 k만큼 각 지반 데이터(오브젝트)에서 대표 오브젝트를 구한다.
step 4. 각 대표 오브젝트와 나머지 오브젝트의 구성 지층 비교
  If(대표 오브젝트의 구성 지층 = 나머지 오브젝트의 구성 지층)
    If(대표 오브젝트의 지층 구성비 = 나머지 오브젝트의 지층 구성비)
      then 대표 오브젝트와 동일한 클러스터에 추가
step 5. 구성된 각 클러스터의 대표 구성지층과 지층 구성비를 이용해 클러스터끼리 비교. 동일하다면, 클러스터간 통합
step 6. 더 이상 통합 불가능하면, 중지

시추공 정보 클러스터링의 결과로 지반조사 데이타들은 동일한 지층들로 구성된 지반들의 데이타들은 한 클러스터에 그룹화하여 데이타베이스에 저장하고, 재분류 단계의 입력으로 사용한다.

그룹으로 분류된 지반조사 데이타는 지반을 구성하는 동일한 종류의 지층들로 분류되었지만, 지층의 구성비율이 서로 상이하다면 유사한 지반이라고 판단하기 어렵다. 따라서, 유사한 지층의 구성비를 갖는 그룹으로 재분류하기 위해 <알고리즘 2>를 이용한다.

<알고리즘 2> 재분류를 위한 알고리즘

step 1. 각 그룹의 대표 오브젝트의 지층 구성비 계산
step 2. 대표 오브젝트의 지층 구성비와 다른 오브젝트들간의 지층 구성비를 비교
  If(대표 오브젝트의 지층 구성비 = 나머지 오브젝트의 지층 구성비)
    then 그 그룹내에서 소 그룹들로 재분류
step 3. 각 소 그룹들간의 지층 구성비 비교
  구성비가 동일하다면, 소 그룹으로 통합
step 4. 더 이상 통합이 불가능하면, 중지

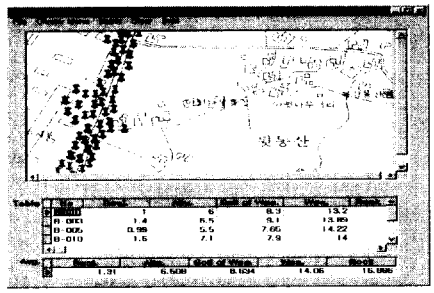
데이타베이스에 저장되어 있는 클러스터링의 결과를 이용하여 지반조사 데이타들은 동일한 지층구성과 지반을 구성하는 지층들이 유사한 구성비를 갖는 데이타들끼리 분류된다. 재분류된 지반조사 데이타들은 지층구성, 지층구성비가 유사한 것으로 서로 구별되는 특성을 갖게 되는데, 이 특성은 다양한 비균질한 지반들간에 구별할 수 있는 특성이 된다.

3.3 지반 분석 결과의 가시화

시추공 클러스터링을 이용한 지반 분석 시스템은 지반조사 데이타의 분석 정보를 가시화하여 사용자가 쉽게 다차원적인 판단이 가능하도록 한다. 가시화된 분석 정보는 클러스터링의 결과를 보여주는 지도와 해당 클러스터([6])내의 데이타 값들로 구성되어 있다. 지반 분석 시스템을 통한 테스트용 지반조사 데이타의 분석의 예는 [그림 4]와 같다. 클러스터링의 결과는 지도상에 구별되는 기호로 나타내고 기호에 해당되는 클러스터내의 지반의 내용과 클러스터가 갖는 지반의 평균 지층구성비 정보를 나타낸다.

지반 분석 시스템은 사용자가 원하는 지반의 깊이를 선택하여 그 깊이까지 구성된 지층의 데이타를 이용해 클러스터링을 시작하게 된다. 지반에 대한 깊이의 선택은 사용자가 원하는 깊이까지만 구성되어 있는 지층들간의 특성들로 분류함으로써 사용자가 의도한 깊이까지의 지반 정보를 보여주게 된다.

서로 구별되는 지반조사 데이타의 특성을 가시화해 주기 위해서 지반조사 데이타가 가지고 있는 위치정보를 이용하여 지도상에 기호로 나타내게되고 유사한 지반들간의 지리적 정보, 전체 지형에 대한 지반 정보까지 사용자가 부가적으로 얻게 함으로써, 사용자에게 다차원적인 분석을 가능하게 한다.



[그림 4] 지반조사 데이타의 지반 분석의 예

4. 결론 및 향후 연구

본 연구에서 개발한 시추공 정보의 클러스터링을 이용한 지반 분석 시스템은 지반조사 정보시스템의 지층 구성 정보를 이용하여 지반이 가지는 특성을 추출하여 이 특성을 유용한 정보로 활용한다. 추출된 특성은 비균질한 지반들이 유사한 지반들로 그룹되어진 그룹의 특성으로 이 특성을 가시화해 줌으로서 사용자가 유사한 지반들의 지리적 위치나, 지역 지형의 지반구조의 특성을 판단할 수 있게 함으로써 기존에 사용되어진 지질도나 토질도보다 유용한 역할을 하게 된다.

본 연구에서는 방대한 지반조사 데이타에 자료들 간의 특성을 파악하기 위해 데이타 마이닝에서 사용되는 클러스터링 기법을 지반조사 데이타에 적용하였다. 시추공 정보의 클러스터링 기법은 지반조사 데이타 특성을 고려하면서도 대용량의 데이타에 효율적으로 클러스터링이 이루어지도록 하여 지반조사 데이타에서 얻을 수 있는 정보를 극대화한다.

향후 연구방향은 본 시스템이 지반 구성과 구성비를 이용해 단순히 그룹화 하여 정보를 보여주는 것에 그치지 않고, 지층구성과 구성비 외에 다른 인자를 포함하여 지반이 갖는 특성들의 소실을 최소화하는 효과적인 클러스터링 방법을 이용하고, 데이타 마이닝 기법을 추가하여 지층간의 연관 관계를 제시할 수 있는 시스템을 연구하고 설계한다.

참고문헌

[1] 한국건설기술연구원, 구조물의 안전성과 경제성을 높이기 위한 지반조사 기술개발 및 D/B 구축, 과학기술부, 1999.
[2] 권영식외 5인, 지형분석, 교학연구사, 1988.
[3] 양승영, 강필종, 야외지질학, 형설출판사, 1989.
[4] 과학기술편집부, 토질조사의 계획과 적용, 과학기술, 1998
[5] 임종석, 지반조사 상세편람, 엔지니어즈, 1996.
[6] Ian H. Witten and Eibe Frank, Data Mining, Morgan Faufmann, 1999.
[7] Alexander Hinneburg and Daniel A. Keim, "Clustering Techniques for Large Data Set-From the Past to the Future," In Proceeding of KDD, 1999.
[8] Raymond T. Ng and Jiawei Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," In Proceeding of VLDB, 1994.