이분적 터널 암반 분류를 위한 정성적 자료의 지구 통계학적 연구 - Ⅰ. 이론-

A Geostatistical Study Using Qualitative Information for Tunnel Rock Binary Classification - Ⅰ. Theory -

유 광 호*
You, Kwang-Ho

Abstract

In this paper, the incorporation of qualitative(or soft) data, such as outputs of geophysical tests or construction experience which has so far been cumulated, was discussed for rock classsification. Geostatistics was used for this research since the parameters for the design of tunnels are spatially correlated. In particular, indicator kriging technique, which is one of non-parametric approaches, was used. As a selection criteria for an optimal classification, the cost of errors was adopted and the binary classes were only considered for rock classification. In future, incorporating an appreciable amount of available qualitative data will be necessary in tunnelling projects in which quantitative data are scarce. In this respect, this research is of great significance.

요 지

본 논문에서는 암반 분류를 위해 물리탐사 결과나 그동안 축적된 시공경험 등의 정성적 자료의 사용을 고려하였다. 터널 설계를 위한 요소(parameter)들이 공간적 상관관계를 갖기 때문에 지구통계학(Geostatistics)을 이용하였으며, 특히, 비모수적(non-parametric)방법 중의 하나인 지시크리깅(indicator kriging) 기법을 사용했다. 최적 분류를 위한 선택 기준으로는 오차에 대응하는 비용(the cost of errors)을 사용했으며, 암반분류는 이분적 분류에 한정하였다. 앞으로, 정량적 데이타가 절대적으로 부족한 터널공사등에서 비교적 많은 양이 존재하는 정성적 데이타의 이용은 절실하며, 이러한 점에서 본 연구가 가지는 의미는 크다.

1. 서 론

터널 시공에 있어서 직접 측정한 정량적 데이타는 극히 적으며, 이는 보링이나 샘플링이 비싸기 때문이다. 반면에 상당히 많은 양의 정성적 데이타, 예를 들면, 전문가의 견해, 비슷한 지

* 정회원, 삼성건설 기술연구소 선임연구원

반에서의 경험등이 존재한다. 따라서, 보다 지반의 상태를 잘 이해하기 위해서는 정량적 데이타 뿐만 아니라 정성적 데이타의 이용도 고려되어야 한다고 생각된다.

터널 시공 또는 설계 회사는, 모든 가능한 데이타로부터 특정 시공 지역을 역학적 거동이 비교적 비슷한 세부지역들로 분류하고, 그 지역내의 물성치, 예를 들면, RQD값, 탄성계수, 및 RMR값등을 알고 싶어한다. Solow[14](1986)와 Knudsen과 Fritz[10](1989)는 한지역을 두 개의 세부지역으로 분류하는 문제를 지시크리깅(indicator kriging)을 사용하여 접근하였다. 각 점에서 특정 경계값을 초과할 확률을 계산하고, 그 값들을 이용하여 경계면을 구하였다. 하지만, 어느 방법도 정성적 데이타의 이용은 고려하지 않았다.

한 지역을 분할하는 일반적인 방법은 그 지역을 격자형(grid)으로 세분하고, 모든 자료를 이용하여 각 점을 평가하고, 평가된 각 점을 근거로 그 지역을 분류하는 것이다. 하지만 피할 수 없는 평가오차(estimation error)로 말미암아 몇몇 점들은 잘못 분류되게 된다. 따라서, 분류는 실제상태에 대한 최적의 예측이어야 하는데, 최적의 예측은 선택기준(selection criterion)이나 경제적 근거에 바탕을 두어 정의되어야 한다. 특히, 한 특정 지역을 균일한 세부 영역으로 나눌 때, 선택기준으로, 오차에 대응하는 비용(the cost of errors)을 최소화 하는 것이 가장 효과적인 것으로 알려져 있다(Aspie와 Barnes[3], 1990).

지시크리깅은 한 변수의 실제값은 모르지만, 그 값이 특정 경계값보다 적거나 크다는 것을 알 때에 효과적인 기법이다(Dubrule과 Kostov[6], 1986; Journel[7], 1986; Omre[13]. 1987), 정성적 데이타의 신뢰도를 정량화하기 위해 0과 1사이의 모든 값을 갖는 소프트 지시자(soft indicator)의 사용이 Kulkarni[11](1984)와 Journel[7](1986)에 의해 제안되었다.

따라서, 본 논문에서는 정량적 데이타는 물론 정성적 데이타도 이용 가능한 최적의 이분적(binary) 암반 분류법을 지시크리깅 기법과 오차에 대응하는 비용의 선택기준을 이용하여 제시하고자 한다.

2. 지반공학 데이타의 분류

지반공학 데이타는 신뢰도(degree of certainty)에 따라 크게 하드(hard)데이타와 소프트(soft)데이타로 나눌 수 있다. 또한, 데이타가 주는 정보의 크기에 따라 점(point)데이타와 지역(areal)데이타로도 나눌 수 있다.

하드데이타는 한 특정 점에서 정확히 측정된 데이타이다. 약간의 측정오차가 포함될 수 있지만, 이 데이타에 내재된 불확실성은 보통 무시된다. 직접 측정한 RQD값, 일축압축강도값 등이 이 범주에 속한다.

많은 사람들(Bardossy 등[5], 1988; Dubrule과 Kostov[6], 1986; Journel[7], 1986 등)이 하드데이타가 부족한 경우를 위해 정성적 데이타의 이용을 제안했다.

이와같은 정성적 데이타는 소프트데이타라고 명명된다. 불연속면 간격(discontinuity spacing)의 측정치로 부터 계산된 RQD값, 전문가의 추정값등이 이 부류에 속한다.

소프트데이타를 이용하기 위해서는, 우선 소프트데이타의 신뢰도가 정량화되어야 한다. 하지만, 이들 데이타는 신뢰 구간이나 확률분포에 의해 정량화될 수 있다. 또한, 소프트데이타는 보다 많은 정보나 경험이 축적될수록 점점 더 하드데이타에 근접하게 된다. 다시말해, 하드데이타는 소프트데이타의 극한 형태라고 볼 수 있다.

3. 불확실성의 확률적 평가

실제의 많은 문제에 있어서, 미지값에 대한 포괄적(global)확률 분포 보다는 오히려 국부적(local) 확률 분포가 요구된다. 예를 들면, 터널 시추조사에 있어서 암반 분류 지표인 RMR의 포괄적 확률 분포는 그 암반 상태에 대한 개략

적인 설명을 줄 뿐이다. 반면, RMR값에 대한 국부적 확률 분포는 상세한 터널 설계를 위한 충분한 정보를 제공한다.

공간적으로 상호 상관관계를 갖는 변수(a spatially correlated variable)를 위한 여러가지 평가 기법이 지구통계학(Geostatistics) 문헌에 소개되어 있는데, 크게 모수적(parametric) 방법과 비모수적(non-parametric) 방법으로 나눌 수 있다(Alli 등[1], 1990). 모수적 방법에서는 값에 대한 불확실성은 그 분산값(estimation variance)에 의해 정량화될 수 있지만, 기본 확률분포(underlying probability distribution)를 가정해야 하기 때문에, 공간적으로 상호 상관관계를 갖는 변수에는 타당하지 않다(Journel[9], 1989, p. 22), 반면에, 비모수적 방법은 기본 확률분포를 가정할 필요가 없다. 이 방법은 각 점에서의 지시사변화(indicator transformation)에 근거를 두고 있으며, 최종적으로 누적밀도함수를 준다.

3.1 비모수적(Non-parametric) 방법

한 점 x에서의 미지값 $Z(x)$를 주변의 n개의 알려진 값 $z(x_i)$, $i=1, 2,\cdots,n$ 으로 부터 평가할 때, 이 평가에 관한 불확실도를 생각해 보자. 가장 쉬운 방법은 이미 알고 있는 n개의 데이타의 분포(distribution)를 이용하는 것이다. 즉, 미지값 $Z(x)$가 한개의 주어진 경계값 z보다 작거나 같을 확률은 주어진 n개의 데이타가 그 경계값보다 작거나 같을 비율로 정해진다.

$$\mathrm{Prob}[Z(x) \leqslant z \mid n \text{ 기지데이타}] \approx Z(x_i) \leqslant z,\ i=1,2\cdots, n\text{의 비율} \tag{1}$$

다음과 같이, 지시데이타(indicator data)를 정의함으로써,

$$i(x_i) = \begin{cases} 0, \text{ 만약 } z(x_i) > z \\ 1, \text{ 만약 } z(x_i) \leqslant z \end{cases} \tag{2}$$

식(1)은 다음과 같이 된다.

$$\mathrm{Prob}[Z(x) \leqslant z \mid \text{기지 데이타}] \approx \frac{1}{n}\sum_{j=1}^{n} i(x_j) \tag{3}$$

식(3)의 왼쪽 항은 조건 누적밀도함수 F(x|n 기지 데이타)이며, 이는 지시데이타의 동일 가중 평균값(equally weighted average)으로 구해진다. 이와 같은 계산은 $Z(x)$와 $z(x_i)$들이 각각 서로 상관관계가 없을 때는 통계적으로 정당화될 수 있다. 하지만, Journel[8](1988)에 의하면, 장소에 따라 값들의 상관관계가 변할 때는, 다음과 같은 비균등 가중 평균값이 보다 타당하다.

$$F(x|n \text{ 기지 데이타}) \approx \mathrm{Prob}[Z(x) \leqslant z|n \text{ 기지 데이타}] \approx \frac{1}{n}\sum_{j=1}^{n} w(x_j)\, i(x_j) \tag{4}$$

여기서, n과 $w(x_j)$는 미지의 데이타 수와 특정 경계값 z에서의 구해야 할 가중치이다.

따라서, 미지값에 대한 불확실도는 평가오차(estimation error)의 형태가 아닌, 확률분포의 형태로 나타내진다.

3.2 누적밀도함수의 평가

지반공학에서 일반적으로 사용되는 변수들, 예를들면, RQD, 압축응력값, RMR값등은 공간적으로 상당한 상관관계를 보인다(Vanmarcke[15], 1978: Baecher[4], 1978). 다시 말하면, 가까운 샘플은 멀리 떨어져 있는 샘플 보다 더욱 유사한 경향을 보인다.

따라서, 이 변수들은 공간적 상관관계를 고려한 통계과정(stochastic process)를 통해 처리되어야 한다. 미지값과 주변 데이타 사이의 상관관계를 정의하기 위해서는, 이들 값 사이의 공간상관관계 측정(proximity measure)이 필요하다.

여러가지 알려진 공간상관관계 측정중에, 베리오그램(variogram) 측정은 지구통계학에서 폭넓게 사용되고 있다. 이 베리오그램 측정은 변수값과 그 위치의 함수이다. 실험적 지시베리오그램(indicator variogram) 함수 $2\gamma_I(h)$는 단순히 거리가 대략 h 만큼 떨어진 n(h)개의 데이타 쌍들의 차이의 제곱에 대한 평균이다.

$$2\gamma_I(h) = \frac{1}{n(h)}\sum_{j=1}^{n(h)} [i(x_j) - i(x_j+h)]^2 \tag{5}$$

식 (4)에서 n(h)는 대략 거리가 h만큼 떨어진 데이타의 수 이다.

일단, 공간상관관계 측정이 선택되면, 일반적인 내삽기법(interpolation algorithm), 예를 들면, 보통크리깅(ordinary kriging)등이 식(4)의 미지 가중치를 구하기 위해 사용될 수 있다. 각 격자점에서 각각의 지시크리깅 방정식의 해를 구함으로써, 사후누적밀도함수(posterior cumulative density function)값이 계산되어, 개략적이긴 하지만 확률분포를 나타낸다.

4. 오차에 대응하는 비용(The Cost of Errors)

한 터널 중심선이 N개의 격자점으로 세분화되었다고 하자. 각 점은 계산된 값에 따라 두개의 등급(class) 중 하나로 분류된다. 예를 들면, RMR값 50을 기준으로 해서 계산된 값이 50 보다 크면 질이 좋은 암반(등급 1), 그렇지 않으면, 질이 나쁜 암반(등급 2)으로 분류된다고 하자.

만약, 한 점에서 암반이 과소 평가된다면, 이 점은 설계시 과다 보강될 것이며, 이는 최적화 설계라는 점에서 볼 때, 필요 이상의 보강재를 사용했기 때문에, 추가비용이 낭비된 것으로 간주될 수 있다. 이 추가비용은 분류시 생기는 오차 때문에 생기는 비용 (cost of errors)이다. 반면, 이 점이 과대 평가되면, 설계시 과소 보강되어 터널의 붕괴등에 따른 공기 지연이나 인재등의 손실을 야기시킬 수 있다. 이것 또한 추가비용을 초래하게 된다. 하지만, 이 점이 옳게 분류되면, 추가비용은 없을 것이다.(Aspie[2], 1989, p.19).

터널상의 한점에서 그 참값을 z(x), 평가된 값을 $\hat{z}(x)$라 하면, 그 분류오차에 의한 비용 계산은 다음의 손실함수(loss function) $\$_e$에 의해 정의된다.

$$\$_e = S_{ij} \text{ 만약 } z(x) \in C_i, \text{ 그러나 } z(x) \in C_j, \quad i, j = 1 \text{ 또는 } 2 \tag{6}$$

식 (6)에서, C_i는 등급의 종류를 나타내며, S_{ij}의 첫번째 아래첨자는 평가된 등급을, 두번째 아래첨자는 실제 등급을 나타낸다. 따라서, 4가지의 가능한 경우가 있으며, 그중 $\$_{11}$과 S_{22}는 영(zero)이 된다.

다음의 식 (7)과 같이 참값과 계산값에 대한 지시함수(indicator function)를 정의하면, 식 (6)은 식 (8)과 같이 된다.

$$I(z(x) \in C_j) = \begin{cases} 1, \text{ 만약 계산값이 } C_j\text{에 속하면} \\ 0, \text{ 만약 그렇지 않으면} \end{cases} \tag{7}$$

$$I(z(x) \in C_i) = \begin{cases} 1, \text{ 만약 참값이 } C_i\text{에 속하면} \\ 0, \text{ 만약 그렇지 않으면} \end{cases}$$

식 (7)에서 첨자 i와 j는 1과 2의 값을 취한다.

$$\$_e = \sum_{i=1}^{2} \sum_{j=1}^{2} \$_{ij} * I(\hat{z}(x) \in C_i) * I(z(x) \in C_j) \tag{8}$$

한 점에서, 두 개의 다른 오차에 대응하는 비용이 계산된다. 만약 한 점에서 등급이 i로 평가되면, 식 (8)에서 $I(\hat{z}(x) \in C_i)$은 식(7)의 정의에 의해서 1이 되기 때문에, 이 경우의 오차에 대응하는 비용은 다음과 같이 된다.

$$S_e | z(x) \in C_i = \sum_j \$_{ij} * I(z(x) \in C_j), \; i = 1,2 \tag{9}$$

오차에 대응하는 비용은 z(x)가 미지수이므로, 무작위 변수(random variable)이며, 이와 같은 변수를 처리하는 표준 방법은 그것의 기대값(expected value)을 구하는 것이다(Aspie[2], 1989, p.22). 또한, 두 개의 가능한 등급 중에 최적 등급은 최소의 오차에 대응하는 비용의 기대값을 주는 등급이 그 점에서 최적 등급으로 선택되어야 한다. 따라서, "min"과 기대값 연산자 "E"를 사용하여 최소 기대값은 다음과 같이 정의될 수 있다.

$$\$_e^* = \min_i \{E [\sum_{j=1}^{2} \$_{ij} * I(z(x) \in C_j]\}, \; i = 1,2 \tag{10}$$

합(sum)에 대한 기대값은 단순히 각각의 기대값들의 합이므로 (Lindgren[12], 1976, p.107), 식 (10)은 다음과 같이 된다.

$$\$_e^* = \min_i \{\sum_{j=1}^{2} E[\$_{ij} * I(z(x) \in C_j)]\}, \; i = 1,2 \tag{11}$$

$\$_{ij}$는 정해진 (not random)값이므로,

$$\$_e^* = \min_i \{ \sum_{j=1}^{2} \$_{ij} * E[I(z(x) \in C_j)] \}, i=1,2 \quad (12)$$

식 (12)에서 지시함수의 기대값, $E[I(z(x) \in C_j)]$은 $Pr[z(x) \in C_j | \text{기지의 데이타}]$ 이기 때문에, 한 점에서의 오차에 대응하는 비용의 기대값은

$$\$_e^* = \min_i \{ \sum_{j=1}^{2} \$_{ij} * P_r[z(x) \in C_j | \text{기지의 데이타}] \}, i=1,2 \quad (13)$$

이 된다. 식 (13)에서 $Pr[z(x) \in C_j | \text{기지의 데이타}]$은 기지의 데이타가 주어진 상태에서 참값이 등급 j에 속할 조건부 확률을 의미한다. 다시 말해서, 식 (13)은 오차에 대응하는 비용의 기대값을 계산하기 위해서는 단지 조건부 확률 만이 필요하다는 것을 의미하며, 이는 어떠한 확률분포에도 관계 없이 사실이다.

그러나, 식 (13)은 한 점에서의 기대값을 준다. 터널 전 구간에서의 기대값은 전 구간을 N개의 격자점으로 세분화하고, 각 점에서의 기대값을 구하여, 모두 더함으로써 얻어질 수 있다. 또한, 전술한 바와 같이, 각 점에서 최적 등급이 정해지며, 따라서, 터널 전 구간에 걸쳐 암반 분류가 가능하게 된다.

5. 결론 및 맺음말

본 논문에서는 정량적 데이타는 물론 정성적 데이타도 이용 가능한 최적의 이분적 암반분류 방법을 제시하였다. 이는 정성적 데이타가 극히 한정되어 있는 터널공사등에 도움이 될 것으로 생각되며, 다음과 같은 결론을 얻을 수 있었다.

1) 정성적 데이타가 정량적인 데이타가 부족한 터널 시공등에 암반 분류를 위해 체계적으로 사용될 수 있다.
2) 선택기준의 하나로써 오차에 대응하는 비용(the cost of errors)이 암반분류를 위해 효과적으로 사용될 수 있다.

앞으로, 본 논문에서의 제시된 암반분류 방법을 이용한 응용예는 별도의 논문을 통해 발표할 예정이다. 추후의 연구과제의 하나로써, 추가 조사가 필요할 경우, 어떻게 효율적으로 조사가 행해질 수 있는가에 대한 연구가 가능할 것으로 생각된다.

또한, 본 논문에서는 이분적(binary) 분류에 대해 언급하였다. 하지만, 보다 일반적인 분류를 위해서는 다분적 분류(multiple classification)에 대한 연구가 이루어져야 한다고 생각된다.

참 고 문 헌

1. Alli, M.M., Nowatzki, E.A. & Myers, D.E., (1990), "Probabilistic Analysis of Collapsing Soil by Indicator Kriging," Mathematical Geology, Vol. 22, No. 1, pp.15−38.
2. Aspie, D. (1989), The Cost of Classification Errors in Geologic Site Characterization, M.S. Thesis, University of Minnesota, 138p.
3. Aspie, D. & Barnes, R.J., (1990), "Infill−Sampling Design and the Cost of Classification Errors, "Mathematical Geology, Vol. 22, No. 8, pp.915−932.
4. Baecher, G.B.(1978), "Analyzing Exploration Strategies," in: Site Characterization & Exploration, C.H. Dowding(ed.), Proceedings, Specialty Workshop, Evanston, Illinois, pp. 220−246.
5. Bardossy, A., Bogardi, I. & Kelly, W.E.,(1988), "Imprecise (fuzzy) Information in Geostatistics, "Mathematical Geology, Vol. 20, No. 4, pp.287−311.
6. Dubrule, O. & Kostov C., (1986), "An Interpolation Method Taking Into Account Inequality Constraints: I. Methodology," Mathematical Geology, Vol. 18, No. 1, pp.33−51.
7. Journel, A.G.(1986), "Constrained Interpolation and Qualitative Information−The Soft Kriging Approach," Mathematical Geology, Vol. 18, No. 3, pp.269−286.
8. Journel, A.G(1988), "Non−parametric Geostatistics for Risk and Additional Sampling Assessment," in: Principles of Environmental Sampling, Larry Keith(ed.), American Chemistry Society, pp.27−45.

9. Journel, A.G.(1989), Fundamentals of Geostatistics in Five Lessons, Short Course Presented at the 28th International Geological Congress, American Geophysical Union, Washington, D.C. 40p.
10. Knudsen, H.P. & Fritz, C.B., (1989), "Using IK to Determine Pod Boundaries in a Karst Hosted Lignite Deposit," in: Application of Computers and Operations Research in the Mineral Industry, Alfred Weiss(ed.), 21st. International Symposium, pp.227－236.
11. Kulkarni, R.B.(1984), "Bayesian Kriging in Geotechnical Problems," in: Geostatistics for Natural Resources Charcterization. Part 2, G. Verly et al. (eds.), D. Reidel Publishing Company, pp.775－786.
12. Lindgren, B.W.(1976), Statistical Theory, 3rd. edition, Macmillan Publishing, New York, 614p.
13. Omre, H. (1987), "Bayesian Kriging－Merging Observations and Qualified Guesses in Kriging," Mathematical Geology, Vol. 19, No. 1, pp.25－39.
14. Solow, A.R.(1986), "Mapping by Simple Indicator Kriging," Mathematical Geology, Vol. 18, No. 3, pp.335－352.
15. Vanmarcke, E.H.(1978), "Probabilistic Characterization of Soil Profiles," in: Site Characterization & Exploration, C.H. Dowding (ed.), Proceedings, Specialty Workshop, Evanston, Illinois, pp.199－216.

(접수일자 1993. 4. 28)