

GIS에 기초한 광물자원평가모형의 믿음성 검증에 ROC곡선법의 적용

김 연 호

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《과학자, 기술자들은 현실에 튼튼히 발을 붙이고 사회주의건설의 실천이 제기하는 문제들을 연구대상으로 삼고 과학연구사업을 진행하여야 하며 연구성과를 생산에 도입하는 데서 나서는 과학기술적문제들을 책임적으로 풀어야 합니다.》(《김정일선집》 증보판 제15권 492페이지)

최근 GIS에 기초하여 광물자원을 정량적으로 평가하기 위한 많은 모형들과 그것에 기초한 여러가지 방법들이 제기되였다. 이 방법들을 실천에서 적극 리용하기 위하여서는 그것들의 정확성과 믿음성을 정확히 평가하여야 한다.

론문에서는 ROC곡선법을 적용하여 GIS에 기초한 광물자원평가모형의 믿음성을 검증하기 위한 한가지 방법을 제기하였다.

1. ROC곡선법적용의 이론적기초

ROC(Receiver Operating Characteristics)곡선법의 기본원리는 수리통계학에 기초하고있다. ROC곡선법은 무선전파탐지기 수신기의 동작특성을 평가하는데 처음으로 적용되었으며 최근에 정보공학과 의학분야를 비롯하여 여러 분야에 광범히 응용되고있다.[2, 3]

실제로 신호와 잡음을 구별하는 수신기동작특성을 평가하기 위한 ROC곡선의 매 점은 신호를 정확히 구별할 루적확률(감도)을 세로축으로 하고 잡음을 신호로 잘못 판단할 루적확률(특이성)을 가로축으로 하는 자리표계의 점들이다. ROC곡선밀부분의 면적(AUC)은 분류기의 성능을 평가하는 중요한 지표로 된다. AUC값은 0.5~1인데 1에 근사할수록 분류기의 성능이 좋은것으로 평가한다.

선행연구[1]에서는 전망구역평가방법의 믿음성을 검증하면서 실지 존재하는 광상점자료(검증자료)와 평가결과사이의 상관성을 고려하였다. 그러나 론문에서는 ROC곡선법을 리용하여 광상점이 존재하는 구역뿐아니라 광상점이 존재하지 않는 구역의 자료도 검증자료에 포함시켜 평가를 진행하였다.

GIS에 기초한 광물자원평가문제는 전망성이 있는 구역과 그렇지 못한 구역을 분류하는 문제로서 우에서 언급한 신호와 잡음을 구별하는 문제와 원리상 같으므로 ROC곡선법을 적용할수 있다. GIS에 기초한 광물자원평가에서 최종결과는 보통 광물자원분포구역을 그 전망성정도에 따라 분류한 지도형태로 얻어진다. 실제로 전망성이 있다 혹은 없다라는 2개의 값으로 분류하는 경우 전망구역평가결과는 0과 1의 2값지도로 주어진다. 이때 값이 1

이면 유망한 구역으로, 0이면 그렇지 못한 구역으로 평가할수 있다.

이 결과를 실시 광상점이 존재하는 구역과 광상점이 존재하지 않는 구역에 대한 검증자료와 비교하여 표와 같이 종합한다.

표. 믿음성검증을 위한 자료행렬		
평가결과	검증자료	
	1	0
1	b_0	a_0
0	b_1	a_1

표에서 1과 0은 검증자료와 평가결과에서 광상점이 존재하는 경우와 존재하지 않는 경우를 의미한다. b_0 은 실시 광상점이 존재하는 구역을 전망구역으로 정확히 평가한 경우의 수이고 b_1 은 실시 광상점이 존재하는 구역을 전망구역이 아닌것으로 평가한 경우의 수이다. 그리고 a_0 은 광상점이 존재하지 않는 구역을 전망구역으로 평가한 경우의 수이고 a_1 은 광상점이 존재하지 않는 구역을 전망구역이 아니라고 평가한 경우의 수이다.

광물자원전망구역을 정확히 평가할 확률 P_1 과 전망구역이 아닌 구역을 잘못 평가할 확률 P_0 을 다음과 같이 계산할수 있다.

$$P_1 = \frac{b_0}{b_0 + b_1}, \quad P_0 = \frac{a_0}{a_0 + a_1}$$

일반적으로 GIS에 기초한 광물자원평가결과는 0과 1의 2값지도가 아니라 [0, 1]의 연속값으로 이루어진 지도이다. 그러므로 평가결과를 n 개의 등급으로 구분하여 검증자료와 비교한다. 이때 믿음성검증을 위한 자료행렬은 표 1과 같이 2×2 행렬이 아니라 $n \times 2$ 행렬형태로 주어진다. 확률은 2값경우와 같은 방법으로 계산할수 있다.

2. ROC곡선법의 적용

우리는 GIS의 공간분석기능을 리용하여 ROC곡선법에 의한 광물자원전망구역평가의 믿음성을 검증하였다.

ROC곡선법의 적용절차는 다음과 같다.

① 검증자료모임을 작성한다.

검증자료는 이미 알려진 광상점들과 광상점이 분포되지 않은 구역에서 우연적으로 선택한 점들로 구성한다.

우선 이미 작성된 알려진 광상점자료에 GIS의 편집기능을 리용하여 우연적으로 광상점이 분포되지 않은 구역의 점들을 선택하여 추가한다. 다음 GIS공간분석기능들인 근방통계기능과 라스터계산, 재분류기능 등을 리용하여 0과 1로 이루어진 라스터형식의 검증자료모임을 얻는다.

② 연속값형태로 얻어진 광물자원전망구역평가결과를 유한개의 등급으로 구분된 자료로 변환한다.

GIS의 공간분석기능인 재분류기능을 리용하여 [0, 1]의 연속값형태로 얻어진 라스터자료를 유한개의 등급으로 리산화된 라스터자료로 변환한다.

③ 믿음성평가를 위한 자료행렬을 얻는다.

GIS공간분석기능인 라스터계산기능을 리용하여 검증자료모임과 유한개의 등급으로 구분된 전망구역평가자료로부터 표와 같은 형식의 자료행렬을 얻는다.

④ ROC곡선을 작성하고 AUC값을 계산한다.

얻어진 자료행렬로부터 광물자원전망구역을 정확히 평가할 확률(루적확률)과 전망구역이 아닌 구역을 잘못 평가할 확률(루적확률)을 계산하여 ROC곡선을 작성한다. 다음 AUC값을 계산하여 믿음성을 검증한다.

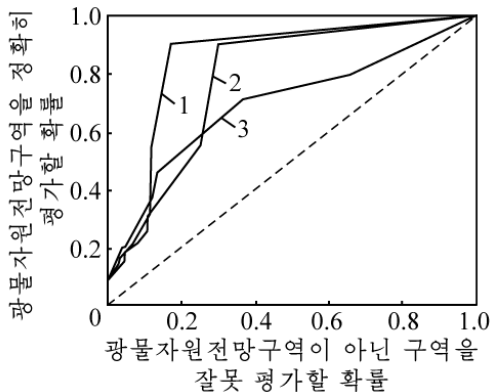


그림. 각이한 모형들의 ROC곡선
1-징후무계모형, 2-믿음도모형,
3-징후리론모형

우리는 이 방법을 선행연구[1]의 자료에 적용하여 그 효과성을 검증하였다.

광상점에 해당하는 5 528개의 화소점과 광상점이 분포되지 않은 구역에서 우연적으로 선택한 9 200개의 화소점을 리용하여 검증자료모임을 얻고 세가지 모형에 기초한 방법들로 얻은 결과를 20개의 등급으로 나눈다. 그리고 MATLAB를 리용하여 ROC곡선을 작도하였다.(그림)

MATLAB의 trapz함수를 리용하여 AUC값을 계산하면 AUC값은 징후무계모형(0.851), 믿음도모형(0.793), 징후리론모형(0.776)으로 가면서 작아진다. 따라서 징후무계모형에 기초한 전망구역평가결과의 믿음성이 가장 높다.

맺 는 말

ROC곡선법에 의하여 얻은 AUC값을 리용하여 GIS에 기초한 광물자원평가모형들의 믿음성을 정량적으로 평가하였다.

참 고 문 헌

- [1] 김연호 등; 지질 및 지리과학, 3, 10, 주체98(2009).
- [2] J. B. Tilbury et al.; IEEE Transactions on Biomedical Engineering, 47, 952, 2000.
- [3] P. Kayastha et al.; Computers & Geosciences, 52, 398, 2013.

주체103(2014)년 8월 5일 원고접수

Application of ROC Curves to Validation of GIS-based Mineral Resource Evaluation Models

Kim Yon Ho

The accuracy of various models for GIS-based mineral resource evaluation was evaluated by ROC curves.

The area under the ROC curves can be used to estimate the quality of the mineral resource evaluation results obtained by predictive models.

Key words: mineral resource, ROC curves