

▣ 지역 벡토르지리자료의 완전성 및 위치정확성에 대한 정량적품질평가방법

김준민, 최원일

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《과학자, 기술자들은 사회주의경제발전의 요구에 맞게 인민경제 모든 부문의 생산기술공정과 생산방법, 경영활동을 새로운 과학적도대우에 올려세우는데서 나서는 과학기술적문제를 전망성있게 풀어나가야 하겠습니다.》(《김정일선집》 증보판 제11권 138페이지)

지난 시기에는 공간자료품질에 대하여 고찰할 때 품질요소와 품질부분요소에 대해서만 일반적으로 논의하고 품질요소들을 구체적으로 평가하기 위한 품질기초측정지표와 그 계산공정에 대한 연구는 활발히 진행되지 못하였다.

벡토르지리자료의 품질평가지표는 완전성, 위치정확성, 속성정확성, 시간정확성, 논리적일관성과 같은것이 될수 있다.

본문에서는 ▣지역 벡토르지리자료에 대한 완전성과 위치정확성을 정량적으로 평가하기 위한 방법에 대하여 서술하였다.

1. 벡토르지리자료의 품질평가를 위한 기초공정

벡토르지리자료에 대한 정량적인 품질평가를 진행하자면 다음과 같은 기초공정을 거쳐야 한다.

① 품질평가기준으로 되는 참조자료모임선택

벡토르자료에 대한 자동적인 품질추정은 다음과 같은 몇가지 가정에 기초하여 진행한다.[1, 2]

첫째로, 새 자료모임과 같은 주제이면서 동일한 지리적령역을 포함하는 품질이 알려진 자료모임이 있어야 한다.

둘째로, 새 자료모임이 분포된 령역안에서 자료의 질적특성이 균일하여야 한다.

이런 가정에서 새 자료모임과 령역적으로 겹치면서 품질이 알려진 같은 주제의 자료인 참조자료모임을 찾아야 한다.

② 참조자료모임과 새 자료모임의 공통령역에서 표본자료의 발취

령역적포함이 있는 기정품질자료가 존재하면 두 자료모임의 사검령역에서 자료표본을 발취하고 새 자료모임에 대한 품질지표값을 추정한다.

지리공간자료모임은 대체로 매우 많은 공간자료들을 포함하기때문에 한번에 전체 자료모임을 조사할수 없다. 그러므로 품질지표를 추정하기 위해서는 표본자료를 리용해야 한다.

지리공간자료모임의 표본을 발취하기 위하여 겹침령역을 1개 정규살창망으로 분할한다.

이 정규살창망의 자리표원점은 우연적으로 선택하며 망목의 크기는 목적에 따라 각이하게 설정할수도 있다.

매 항목에 포함된 지리공간자료들중 표본자료의 백분률에 따라 우연적으로 표본자료를 발취한다.

이상과 같이 참조자료모임과 새 자료모임으로부터 표본자료를 얻으면 새 자료모임의 완전성과 위치정확성을 계산한다.

2. 완전성평가

지리공간자료의 완전성은 전체 지물들중 자료모임에 포함된 지물들의 백분률이다. 실례로 주어진 영역에 3개의 대상이 있고 자료모임에 그중 2개가 있다면 이 자료는 66.7%정도 완전하다고 본다. 그런데 지역에 있는 모든 지물들을 다 알수 없으므로 참조자료모임을 리용하여 전체 지물들을 추정한다. 실례로 참조자료모임이 50%의 완전성을 가지고 한 영역에 2개의 참조지물자료들이 포함되어있다면 그 영역에는 4개의 지물이 있다고 볼수 있다.

자료모임이 점 혹은 면자료를 포함한다면 다음의 공식을 리용하여 그 지역에 있는 추정된 지물의 개수와 새 자료모임에 의해 포함된 지물의 개수를 비교하여 새 자료모임의 완전성을 평가한다.

$$C_n = 1 - \left| 1 - \frac{f_n}{f_r} C_r \right| \quad (1)$$

여기서 f_n 은 새 자료모임으로부터 발취된 표본지물의 개수, f_r 는 참조자료모임으로부터 발취된 표본지물의 개수, C_r 는 참조자료모임의 완전성, C_n 은 새 자료모임의 완전성이다.

식 (1)은 참조자료모임이 완전하지 않을수 있다는 사실을 고려하고있다.

점대상인 경우 새 자료모임의 지물의 수가 참조자료모임의 지물의 수보다 작으면 C_n 은 완전성평가에서 루락률을 의미하고 반대로 새 자료모임의 지물의 수가 참조자료모임의 지물의 수보다 크면 C_n 은 완전성평가에서 초과률을 의미한다.

선대상인 경우에는 지물들을 서로 다른 크기로 정의할수 있기때문에 지물의 개수를 리용하여서는 완전성을 평가하기 힘들다. 실례로 도로는 많은 개수의 선지물을 리용하여 표시될수도 있고 적은 개수의 선지물을 리용하여 표시할수도 있다.

두 경우 표시되는 선지물의 개수는 차이나도 도로에 대하여 완전성은 같다. 그러므로 자료모임의 기하학적형태가 선이라면 선의 개수대신 선의 길이에 기초한 공식을 리용한다.

$$C_n = 1 - \left| 1 - \frac{\sum l_n}{\sum l_r} C_r \right| \quad (2)$$

여기서 l_n 은 새 자료모임으로부터 발취된 표본선지물의 길이이고 l_r 는 참조자료모임으로부터 발취된 표본선지물의 길이이다.

새 자료모임의 길이합이 참조자료모임의 길이합보다 작으면 C_n 은 완전성평가에서 루락률을 의미하고 반대로 새 자료모임의 길이합이 참조자료모임의 길이합보다 크면 C_n 은 완전성평가에서 초과률을 의미한다.

완전성추정값은 참조자료모임의 완전성에 대한 정확도에 크게 의존한다.

3. 위치정확성평가

일반적으로 자료모임의 정확성은 세가지 지표 즉 수평정확도, 높이정확도, 정확도경계안에 있는 지물의 비율로 평가한다.

정확도평가를 위한 흐름도는 그림과 같다.

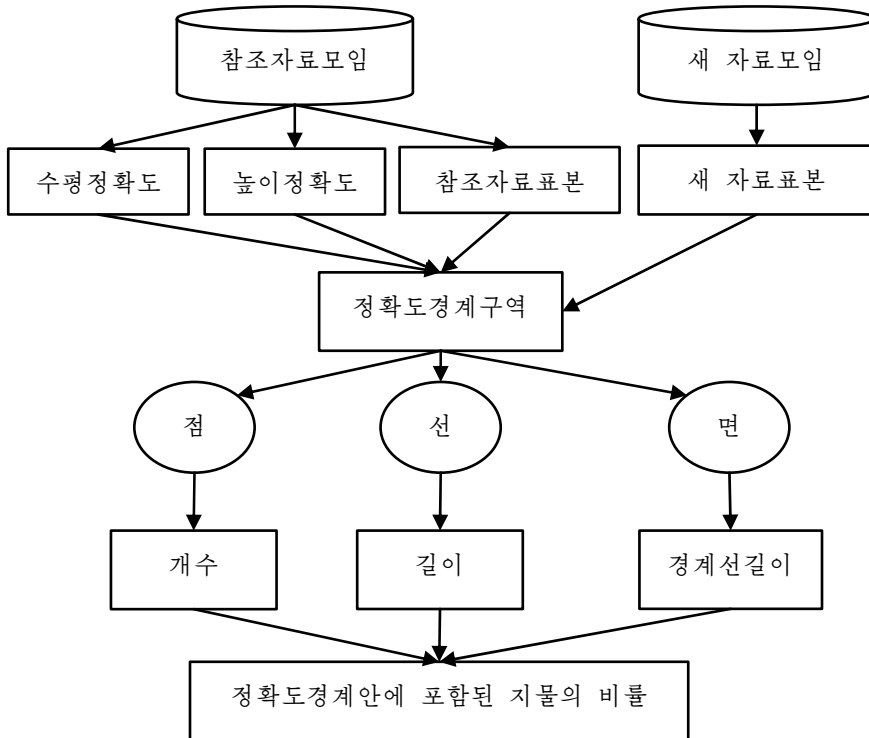


그림. 정확도평가를 위한 흐름도

지물의 실지위치를 정확히 알수 없으므로 근사적으로 참조자료모임으로부터 얻은 지물들의 위치를 리용한다.

정확도평가를 위한 흐름도를 리용하여 새 자료모임에 대한 수평정확도, 높이정확도, 정확도경계안에 포함된 지물의 비율을 계산한다.

먼저 참조자료모임에 대한 수평정확도, 높이정확도, 표본자료를 탐색한다. 그리고 수평정확도와 높이정확도를 리용하여 참조자료모임으로부터 얻은 표본자료를 중심으로 하는 완충구역을 생성한다.

생성된 완충구역에 포함되는 새 자료모임의 표본자료들에 대한 비율을 결정한다.

만일 새 자료모임이 점자료이면 완충구역에 놓이는 점대상의 개수를 계산하고 선자료이면 완충구역에 놓이는 선대상의 총길이를 계산하며 면자료이면 그 완충구역에 놓이는 면대상의 경계선길이를 계산한다.

이 값을 리용하여 새 자료모임에 대한 정확도경계안에 있는 지물의 비율을 계산한다.

새 자료모임의 수평정확도와 높이정확도는 지물들의 주위에서 완충구역을 설정하기 위하여 리용되었으며 참조자료모임의 수평정확도, 높이정확도와 각각 같다.

4. 실험적품질평가

실험적인 검증을 하기 위하여 B지역에서 전체 자료모임의 20%에 대한 품질평가를 진행한다. 품질평가를 위한 자료모임으로서 사용자들이 입력한 철도자료모임을 선정하고 GIS전문가들이 고분해능위성화상에 기초하여 입력한 철도자료모임을 참조자료모임으로 리용하였다.

표본발취를 위한 표본망목망을 구성하고 우연적으로 선택된 망목에 포함된 참조자료와 평가자료를 분석하였다.

먼저 참조자료의 완전성은 1로 하고 기하자료의 완전성을 평가하였다.(표 1)

표 1. 완전성평가결과

No.	새 자료모임의 표본 지물길이의 합/m	참조자료모임의 표본지물 길이의 합/m	불완전성/%	완전성/%
1	91 623.3	91 040.0	초과(0.64)	99.36
2	26 463.4	26 539.7	루락(0.29)	99.71
3	84 045.7	83 591.3	초과(0.54)	99.46
4	50 879.0	50 915.2	루락(0.07)	99.93
5	82 229.7	82 240.7	루락(0.01)	99.99
평균값			0.31	99.69

표 1에서 알수 있는것처럼 해당 철도망자료의 불완전성은 0.31%이고 완전성은 99.69%이다.

다음으로 기하자료의 정확성을 평가하기 위한 정확도경계는 참조자료모임의 정확도를 고려하여 1m의 완충구역으로 설정하고 위치정확성검사를 진행하였다.(표 2)

표 2. 위치정확성평가결과

No.	정확도경계안에 포함된 새 자료모임의 표본지물길이합/m	참조자료모임의 표본지물길이합/m	정확도경계안에 포함된 지물의 비율/%
1	44 283.1	91 040.0	48.6
2	7 870.4	26 539.7	29.7
3	44 830.3	83 591.3	53.6
4	20 327.9	50 915.2	39.9
5	38 271.7	82 240.7	46.5
평균값			43.64

표 2에서 알수 있는것처럼 해당 철도망자료의 정확성은 43.64%이다.

이 방법에서 주의할 점은 참조자료모임의 정확성이 매우 높아야 한다는것이다. 참조자료모임이 정확하지 않다면 새 자료모임의 기하자료에 대한 품질평가지표의 정확한 추정을 할수 없다.

맺 는 말

론문에서 제기한 품질평가공정은 벡토르지리자료에 대한 완전성과 위치정확성에 대한 정량적인 품질평가를 자동적으로 진행할수 있게 한다. 앞으로 벡토르지리자료에 대한 속성정확성, 시간정확성, 논리적일관성에 대한 품질평가를 자동화하기 위한 연구도 심화시켜야 한다.

참 고 문 헌

- [1] Amin Mobasheriet et al.; Geo-spatial Information Science, 21, 3, 234, 2018.
- [2] Alexey Noskov et al.; Big Earth Data, 2, 4, 395, 2018.

주체108(2019)년 10월 5일 원고접수

A Quantitative Approach for Estimating Completeness and Positional Accuracy of Vector Geographic Data in the “ㄴ” Area

Kim Jun Min, Choe Won Il

In this paper we propose a procedure for estimating completeness and positional accuracy on vector geographical data in “ㄴ” area quantitatively.

Keywords: quality evaluation, data quality measure