

위성 영상 분류 기법 선정을 위한 의사 결정 지원 시스템

Decision support system on selection of classification method for remote sensing imagery

황보주원*, 유기윤**, 김용일***

Ju Won Hwangbo, Kiyun Yu, Yong Il Kim

* 서울대학교 지구환경시스템공학부 석사과정 (juwon2@snu.ac.kr)

** 서울대학교 지구환경시스템공학부 조교수 (kiyun@plaza.snu.ac.kr)

*** 서울대학교 지구환경시스템공학부 부교수 (yik@plaza.snu.ac.kr)

요 약

본 연구에서는 사례기반추론(case-based reasoning)을 기본으로 하여 실무자의 분류 기법 또는 분류 구조 결정을 돕는 의사 결정 지원 시스템의 모델을 제시한다. 주요한 네가지 고려 항목은 자료종류(dataset), 위치(location), 기후(climate), 그리고 분류항목(class)이며 사용자는 이들 네 항목에 대해 적합한 값을 선택하게 된다. 본 시스템은 색인화(indexing) 규칙에 따라 관계형 데이터베이스에 저장된 사례들을 추출하여 제시하며 사용자는 그 중 가장 높은 일치도를 보인 사례들을 참고할 수 있다. 본 연구에서는 위계 구조를 통해 다양한 분류 조건을 스크린 상에서 선택할 수 있게 함으로써 사용자가 이에 내재된 논리를 분류 구조의 설계에 반영할 수 있게 한다. 또한 Statistics 기능을 통해 여러 사례의 항목당 분포를 사용자가 검토할 수 있게 함으로써 가장 적합한 사례를 의사 결정 지원 시스템과의 피드백을 통해 찾아낼 수 있게 해준다. 이밖에 분류 조건을 변화 시켜가면서 상황의 변화를 참고할 수 있도록 Navigation 기능을 고안하였다.

주요 용어 : 사례기반추론, 위성 영상, 분류, 의사 결정 지원 시스템

1. 서론

전통적인 토지 피복 분류 방식은 위성으로부터 얻은 영상에 통계적인 알고리즘인 감도 분류 및 미감도 분류의 패턴 인식 기법을 적용하는 것이다(Richards, 1992). 그 후 기존의 통계적 방법의 한계를 극복하기 위해 퍼지(fuzzy) 이론을 도입한 신경망(neural network) 알고리즘, decision tree 기법 등이 개발되었다. 이렇게 다양한 분류 알고리즘은 데이터의 종류 또는 분류 클래스에 따라 다른 성과를

보임이 알려져 있다. 분광 밴드의 종류, 부가 정보의 유무, 대상 지역 등의 세부 사항도 분류 성과에 영향을 준다. 그러므로 인공위성 영상으로 토지 피복을 분류하려는 사용자는 영상이나 분류 기법을 선택하고 분류 대상을 그룹화하는 데 있어 이에 관해 몇 가지 복잡한 결정을 내려야 한다.

지금까지 다양한 알고리즘을 사용하거나 서로 다른 목적에 따라 분류를 수행하고 성능을 평가한 연구는 수없이 이루어져 왔으나, 이는 전문가가 아니면 바로 응

용하기가 어렵고 단편적이라는 한계가 있다. 이로 인해 실무자가 인공위성 영상을 이용한 분류 작업을 수행하고자 할 때 영상이나 기법을 선택함에 있어 어려움이 있어 왔다. 따라서 본 연구에서는 실무자의 분류 기법 또는 분류 클래스 결정을 돕는 의사 결정 지원 시스템의 모델을 제시하고자 한다. 이 시스템은 분류를 수행하기 위해 고려해야 할 사항을 단계별로 제시하여 주고, 조건에 맞는 이전 사례들과 그 문맥(context)을 제공함으로써 사용자가 분류 구조(scheme)를 설계하는 과정을 지원한다.

2. 사례기반 추론

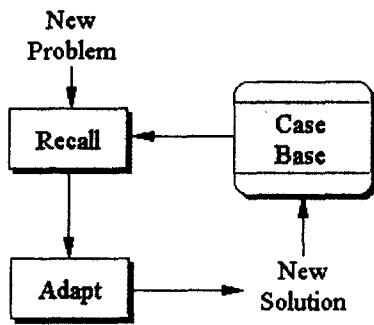


그림 1' 사례기반 추론의 기본 개념

법관이 지난 사건에 대한 판례를 통해 당면한 사건에 대한 판단을 내리는 것처럼 인간은 과거의 사례나 경험에 비추어 현재의 문제를 인식하고 유형화하여 해결책을 내놓을 수 있다. 즉, 현재의 문제가 과거에 나타났던 사례나 경험과 일치하는지 여부를 살피고 일치할 경우 과거의 해결책에 비추어 답을 유추하는 것이다. 사례기반 추론은 이러한 인간의 지적 활동을 모델화한 것으로, 과거문제로부터 얻은 경험이나 지식을 사례 데이터베이스로 구축하여 어떠한 상황이나 문제가 발생하면

기존의 사례 데이터베이스에서 가장 일치하거나 유사한 사례를 선택하여 그 사례가 제시하는 해결책으로 현재의 문제에 대한 답을 제시한다(그림 1).

사례기반 추론은 문제와 해답 사이의 관계를 밝혀내는 것이 아니라 사례 중심의 지식을 사용하여 문제를 해결하고자 한다. 즉, 사용자로 하여금 과거의 사례를 바탕으로 새로운 해답을 발견할 수 있게 하는 것이다. 과거에 어떤 것이 만족스럽게 작동했다는 사실은 그와 유사한 새로운 상황에서 그 방법을 적용할 수 있게 하는 충분한 정당성을 제공하여 준다(Kolodner, 1993). 사례들을 이용함으로써 해결책을 찾는 과정의 효율성을 높일 수 있고, 과거에 일어났던 문제들을 고찰함으로써 잠재적인 실패는 예방될 수 있다. 이렇게 사례기반 추론을 통해 단축한 시간은 문제를 해결하는 실제적인 과정에 투자할 수 있다.

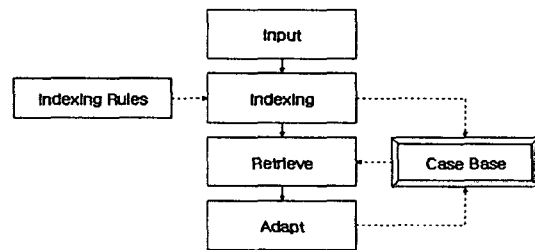


그림 2 사례기반 추론의 과정

3. 의사 결정 지원 시스템의 구성

3.1 색인

분류 사례를 구성하는 요소에는 여러 가지가 있을 수 있겠으나, 본 시스템에서는 사용자가 과거의 분류 상황을 검색하기 위하여 필요하다고 생각되는 4개의 특성(feature)을 주된 색인으로 한다. 이들은 각각 dataset, location, climate, class이다.

3.1.1 Dataset

위성 영상은 수십 기가 이용되고 있으며 보편적으로 쓰이고 있는 것만 해도 십여개에 달한다. 본 시스템에서는 많이 쓰이고 있는 15개의 위성을 색인화 하였으며, 이밖의 위성은 '기타' 항목에 위치하도록 하였다.

3.1.2 Location

본 시스템에서는 위치 확인의 복잡성을 최소화하고 사용자의 편의성을 증가시키기 위해 사례가 발생한 위치를 대륙 또는 국가로 색인화하였다. 여기서 상위 feature는 대륙을 의미하며 그에 따른 sub-feature는 국가가 된다. 분류의 의사 결정에 도움이 되는 사례를 찾아내기 위해서 다음에 기술될 기후와 분류 클래스를 복합적으로 고려하도록 한다.

3.1.3 Climate

사용자가 특정 사례를 검색하고자 할 경우 기후에 대해서도 고려할 수 있도록 한다. 사실상 climate 항목은 location 항목과 어느 정도 중복되는 개념이라고 할 수 있다. 하지만 한 나라 안에도 여러 기후대가 존재할 수 있고 식생의 경우 기후 조건이 무엇보다 토지피복의 구성에 영향을 주기 때문에 기후대를 색인화하였다.

3.1.4 Class

본 시스템의 클래스 색인 구조는 the United States Geological Survey scheme을 기본적으로 따랐다. 단, USGS 시스템은 미국이라는 국가 내의 토지 이용을 체계화하는 데 목적이 있으므로 몇 가지 항목에 대하여 수정을 가하였다. 우선 Level I의 3번 항목인 방목장(rangeland)은 지구 전체적으로 볼 때 하나의 상위 항목으로 분류될 만큼 비중 면에서 중요하지 않다.

따라서 방목장의 경우 2번의 농작물(vegetation) 항목의 목초지(pasture)과 통합된 개념으로 본다. 또한 USGS Level I 8번의 툰드라(tundra)의 경우도 원래는 수목의 생육 한계를 넘어선 한대 지방의 초원을 의미하므로, 그대로 초원(grass) 항목의 하위 feature로 넣어 주었다.

3.2 의사 결정 지원 시스템의 구조

본 의사 결정 지원 시스템의 구조는 그림 3과 같이 표현될 수 있다. 시스템 내의 주요한 두 부분은 사례 베이스와 모델 베이스이다. 사용자는 user interface를 통해 시스템에 접근하게 된다. User interface는 스크린 상에 나타나는 선택창 및 결과창으로 이루어진다. 사례 베이스는 앞 절에서 제시된 구조로 색인화된 entity와 attribute를 가진다. 모델 베이스는 사례 베이스의 검색 알고리즘과 함께 결과 및 통계를 보여주는 기능을 보유한다.

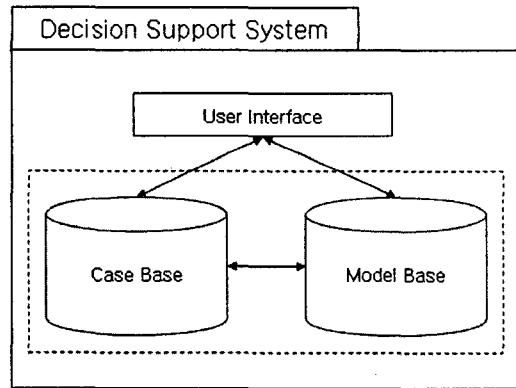


그림 3 의사 결정 지원 시스템 모델

3.2.1 사례베이스의 구성

사례베이스는 그림 4과 같은 entity와 attribute로 구성되어 있다. idCase는 사례의 일련 번호이며 noClass는 분류한 클래스의 개수이다. 원전을 참고할 수 있도록 referenceText 속성도 추가하였다.

Dataset entity는 idDataset과 nameDataset, 속성 두 개가 있다. 예를 들어 Landsat 영상은 idDataset=1이고 nameDataset="Landsat"이 된다. Location은 위치 정보(idLocation), 국가 이름(nameNation)으로 구성되어 있고, Climate는 idClimate 속성을 가진다. Class는 idClass와 nameClass 및 레벨별 id를 가지는데, 그림 9에 표현되어 있다. Method 항목에는 분류에 사용한 기법 이름(nameMethod)과 결과 정확도(overallAccuracy), 그리고 분류 과정을 서술한 내용인 methodText를 부여한다. 여기서 dataset, climate, class, method 항목은 사례당 1개 이상의 값을 가질 수 있다.

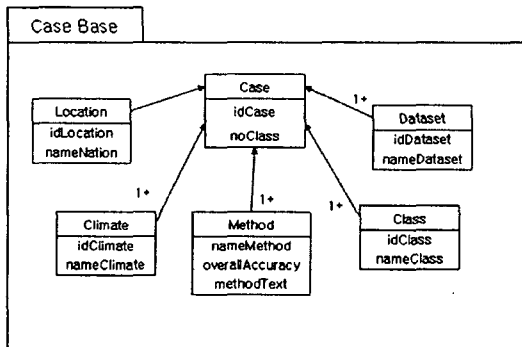


그림 4 사례베이스의 구성(Entity-Attribute)

3.3 Case Retrieval 모드

Case retrieval은 조건을 선택하는 것으로 시작된다. 즉, 사례의 색인을 확인하는 일이다. 본 의사 결정 지원 시스템의 첫 화면은 그림 5의 구성을 지닌다. 첫 화면에서는 dataset, location과 climate, class의 세 항목의 조건을 선택하는 화면으로 들어갈 수 있도록 선택 버튼을 제공한다. 본 시스템에서는 사용자의 이해를 돕기 위해 이러한 선택 버튼에는 도움말(help) 기능을 추가하도록 한다. 즉, 마우스 버튼을 갖다 대거나 마우스 오른쪽 버튼을 클릭

했었을 때 선택 항목에 대한 간단한 안내가 화면에 표시되도록 하는 것이다. 또한, 사례 도출에 필요한 조건들을 선택하고 난 후 다시 첫 화면으로 돌아왔을 때 선택한 항목을 한눈에 볼 수 있도록 오른쪽 박스에 표시한다.

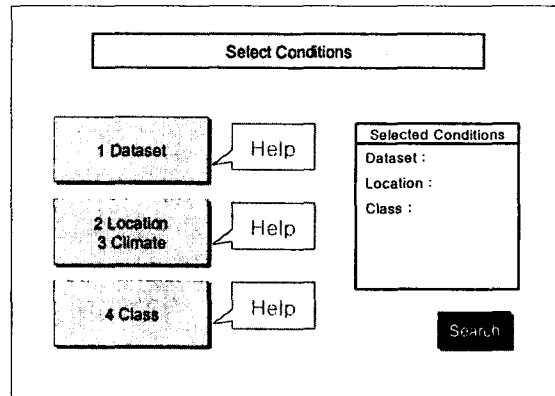


그림 5 Case Retrieval User-Interface

GUI 환경에서 선택된 색인은 idDataset, idLocation, idClimate, idClass의 네 가지 항목으로 입력되어 model base 내에서 매치 과정을 거친다.

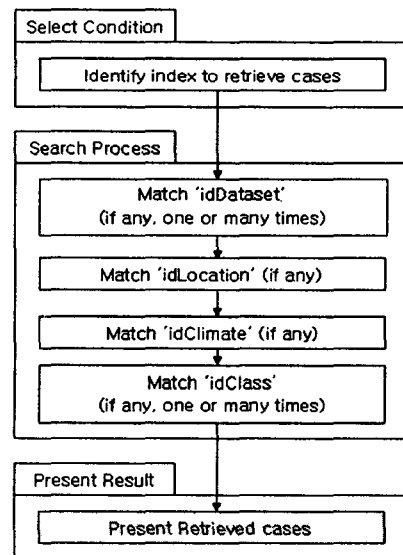


그림 6 매치 과정(match process)

이렇게 해서 검색을 마치면 결과 relational table이 나타나도록 한다. 앞의 case retrieval 과정을 통해 사용자가 지정한 색인과 일치되는 사례들을 제시하는 것이다. 표에는 dataset, 클래스 개수, 분류 기법, 정확도가 나타난다. 이들은 사례 베이스에서 하나의 사례를 구성하는 요소로서 각각 nameDataset, noClass, nameMethod, 그리고 overall Accuracy 값을 의미한다.

사례의 보다 상세한 정보는 텍스트 화면으로 제시한다. Relational table에서 사용자가 관심있는 사례의 행(row)을 선택하였을 때 표현한다. 여기에 수록되는 내용은 idCase, nameDataset, nameNation, nameClimate, nameClass, methodText(분류 방법), 그리고 referenceText(출처)이다.

3.4 Statistics 모드

검색 결과로 많은 수의 사례가 제시되었다면, 이들을 통계적으로 분석함으로써 사용자는 원하는 문제에 대한 좀더 상세한 사례를 찾거나 성공 사례의 분포에 대해 알아 보는 것이 필요하다. 본 시스템에서는 relational table 내에서 사용자는 정렬(sort) 기능과 선택(selection) 및 저장(save) 기능을 사용할 수 있도록 한다.

정렬(sort) 기능은 relational table 내에서 사용자가 사례들을 열별 값(클래스 개수, 정확도)의 크기에 따라 열거되도록 순서를 바꾸는 것을 말한다. Dataset이나 분류 기법과 같이 순서를 따질 수 없는 속성의 경우에는 빈번한 순서에 따라 종류별로 제시하도록 한다. 예를 들어, 분류 기법에 의한 정렬을 택한다면 가장 많이 쓰인 MLC 분류기법을 적용한 사례들이 제시되고, 그 다음으로 쓰인 ANN 기법의 사례들이 나열되는 식이다.

다음으로 statistics 모드에서 제공하는 것은 선택(selection) 기능이다. 정렬 기능을 이용하든지 아니면 사용자가 훑어보든가 해서 통계를 낼 사례 집합을 지정할 수 있다. 이 기능은 이미 case retrieval로 검색한 사례들 가운데서 특정한 집합의 분포도를 알고 싶을 경우 유용하다. 선택한 사례는 저장(save) 기능으로 나중에 다시 불러들일 수 있도록 한다.

Statistics 모드에서는 dataset과 location에 대해서 단순히 사례의 빈도수를 제공하여 준다. 한편 클래스 항목에 대해서는 USGS Classification Scheme을 응용한 분포도를 제시한다.

3.5 Navigation 모드

Navigation 모드는 case retrieval과 statistics 기능을 유기적으로 결합시켜 놓은 것이다. 활용 방법은 다음과 같다. 우선 좌측 상단의 조건 선택 창에서 검색하고자 하는 색인을 지정한다. 그리고 나서 조건 선택 창 우측 하단에 위치한 search 버튼을 눌러 사례베이스에서 case retrieval을 수행한다. 그 결과는 좌측 하단에 relational table로 나타나게 된다. 이전의 statistics GUI에서와 마찬가지로, 정렬 및 선택 기능을 이용하여 우측에 위치한 통계 분석을 수행한다.

4. 결론

본 분류 기법 선정 의사 결정 지원 시스템은 그림 7과 같이 종합적으로 표현될 수 있다. 우선 사례 검색에 필요한 색인을 지정하게 된다. 지정한 색인을 바탕으로 case retrieval을 수행한 결과는 좌측 하단(statistics 모드)과 같이 표시된다. 사용자는 이 화면에서 relational table 상에서 나타나는 사례들을 정렬 또는 선택하면서

사례들의 분포를 dataset, location, method, 그리고 class 그래프 상에서 확인할 수 있다. 만약 도출된 사례가 너무 많거나 적다면, 우측 하단(navigation 모드)의 GUI를 이용해 적절한 개수의 사례가 도출되도록 조정한다. 본 의사 결정 지원 시스템은 위성 영상을 이용하여 토지 피복 분류를 수행하고자 할 때 고려할 수

있는 많은 항목들을 GUI 화면 상에서 검토하여 사용자가 스스로 초점을 맞추어 갈 수 있도록 도와준다. 따라서 표준화된 이용 방법을 정의할 수는 없다. 여기에 제시한 의사모델 베이스를 제공함으로써 사용자와 시스템이 유기적으로 진화해가면서 가장 적절한 해답을 찾도록 하는 데에 의의가 있다.

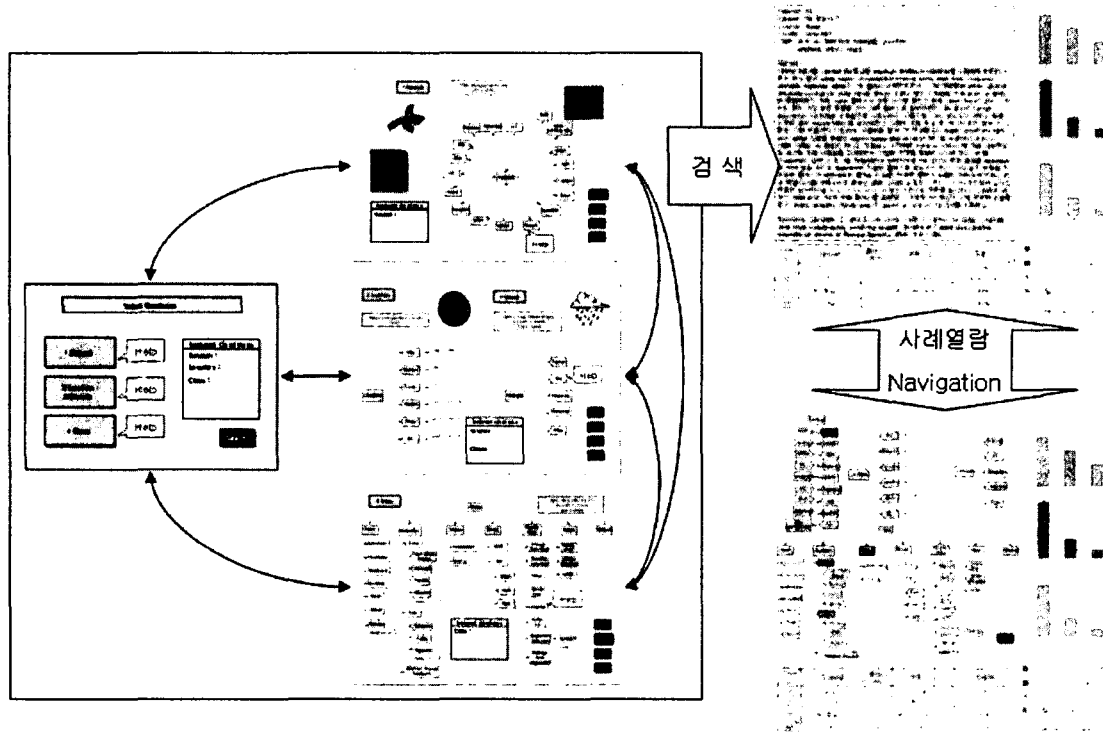


그림 7 분류 기법 선정 의사 결정 지원 시스템의 이용

참고 문헌

이재규 외 공저, 1996. 전문가시스템, 원리와 개발, 범영사.
 Ashley, K.D., 1990. Modelling legal argument: Reasoning with cases and hypotheticals, MIT Press, MA.
 Foody, G.M., and M.K. Arora, 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network, International

Journal of Remote Sensing, 18:799-810.
 Kolodner, J.L., 1993. Case-Based Reasoning, Morgan Kaufmann Publishers, Inc. CA.
 Richards, J.A., 1992. Remote Sensing Digital Image Analysis, Springer-Verlag, Cambridge, UK.