



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년12월04일
 (11) 등록번호 10-1469136
 (24) 등록일자 2014년11월28일

(51) 국제특허분류(Int. Cl.)
 G06F 19/00 (2011.01)
 (21) 출원번호 10-2014-0076414
 (22) 출원일자 2014년06월23일
 심사청구일자 2014년06월23일
 (56) 선행기술조사문헌
 KR1020120128407 A

(73) 특허권자
 한국지질자원연구원
 대전광역시 유성구 과학로 124 (가정동)
 (72) 발명자
 연영광
 세종특별자치시 나리로 38, 712-2604(한솔동, 첫
 마을아파트7단지)
 조성준
 대전광역시 유성구 엑스포로 448, 206-503(
 전민동, 엑스포아파트)
 한종규
 대전광역시 유성구 은구비남로 55, 707-2101(지족
 동, 열매마을7단지)
 (74) 대리인
 김정수

전체 청구항 수 : 총 4 항

심사관 : 이복현

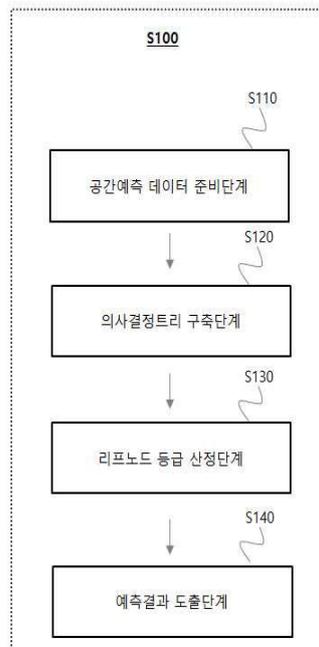
(54) 발명의 명칭 의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법 및 이를 기록한 기록매체

(57) 요약

의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법 및 이를 기록한 기록 매체를 개시한다.

상기 의사결정 트리에서 리프노드 등급 결정방법은 공간예측에 필요한 훈련 데이터셋을 이용하여 하나의 루트 노드와 더불어 부 노드와 복수개의 자 노드 관계를 갖는 의사결정트리 형태로의 학습단계 (S110); 상기 학습을 (뒷면에 계속)

대표도 - 도1



마친 의사결정 트리로부터 리프노드 등급 결정단계(S120)를 포함하고, 상기 학습단계(S110)에서 의사결정트리의 각 노드는 훈련데이터의 클래스 분포 개수를 저장하고 이를 저장하기 위한 구조를 포함하며, 상기(S120)의 리프노드 등급 결정단계에서는 루트노드로부터 리프노드까지 일련의 노드들을 조합에 각 노드에 저장된 클래스 분포수를 이용하여 리프노드 등급을 결정하는 것을 특징으로 한다.

이와 같은 본 발명은 다른 리프노드 등급방법들보다 별도의 최적의 파라미터 값을 찾는 과정을 필요치 않으며, 부 노드의 등급을 자 노드 등급결정과정에서 다시 이용하기 때문에 동일한 부 노드로부터 분리된 자 노드간에 보상하는 것을 특징으로 한다.

이 발명을 지원한 국가연구개발사업

과제고유번호	GP2013-014
부처명	미래창조과학부
연구관리전문기관	산업기술연구회
연구사업명	주요사업-부처임무형
연구과제명	광화대 멀티스케일 3D 지질모델링 구현
기여율	1/1
주관기관	한국지질자원연구원
연구기간	2013.01.01 ~ 2015.12.31

특허청구의 범위

청구항 1

의사결정 트리가 학습할 데이터인 예측 대상이 되는 공간사건 발생 레이어와 공간사건 발생 유발환경 레이어들에 대하여, 위치 대응할 수 있도록 동일한 범위를 갖고 기하 보정하는 공간예측 데이터 준비단계(S110);

상기 공간예측 데이터를 훈련 데이터로 이용하여 하나의 루트노드와 더불어 부 노드와 복수 개의 자 노드 관계를 갖는 의사결정트리를 구축하는 의사결정트리 구축단계(S120);

구축된 상기 의사결정트리를 이용하여 루트노드로부터 계산된 등급 계산 결과를 자 노드에 반영하며 전체 리프노드에 이르기까지 각 노드에 포함된 사건 및 미사건 발생 클래스의 개수를 이용하여 등급을 계산하는 리프노드 등급 산정단계(S130); 및

상기 리프노드의 등급값(우선순위)을 공간 좌표상에서 매핑하는 예측결과 도출단계(S140);를 포함하고,

각 노드는 리프노드 등급 산정을 위해 훈련데이터가 분류된 사건발생과 사건 미발생 클래스 개수 정보를 저장하는 것을 특징으로 하는 컴퓨터에 의해 수행되는 의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법.

청구항 2

제1항에 있어서,

상기 훈련 데이터로 이용하여 하나의 루트노드와 더불어 부 노드와 복수개의 자 노드 관계를 갖는 의사결정트리는 자 노드(R(node.child)) 및 부 노드(R(node.parent))로 구성되며, 상기 자 노드(R(node.child))와 상기 부 노드(R(node.parent))는 식 4로 정의되는 것을 특징으로 하는 컴퓨터에 의해 수행되는 의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법.

[식 4]

$$R(\text{node.parent}) \leq R(\text{node.child})$$

여기서, 부 노드(R(node.parent))와 자 노드(R(node.child))의 등급이 같은 경우 자 노드에 공간사건 발생 클래스가 포함되어져 있지 않으며, 부 노드보다 자 노드의 등급이 클 때 자 노드에 사건 발생 클래스가 존재한 경우이다.

청구항 3

제1항에 있어서

상기 리프노드 등급 산정단계(S130)는,

아래에 기재된 식 5를 이용하여 수행하는 것을 특징으로 하는 컴퓨터에 의해 수행되는 의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법.

[식 5]

$$R(\text{node.child}) = R(\text{node.parent}) + \frac{n_{\text{event}}}{(n_{\text{event}} + n_{\text{non-event}}) \times d}$$

여기서, d는 트리에서의 노드의 깊이를 나타낸다.

청구항 4

제1항 내지 제3항 중 어느 하나의 항에 기재된 컴퓨터에 의해 수행되는 의사결정트리에서 공간 예측을 위한 리프노드 등급 결정 방법이 기록된 컴퓨터 판독 가능한 기록 매체.

명세서

기술분야

[0001] 본 발명은 공간 예측에 관한 것으로, 보다 상세하게는 데이터마이닝 분류기법들 중 하나인 의사결정트리에서 공간예측을 위한 리프노드 등급 결정 방법 및 이를 기록한 기록매체에 관한 것이다.

배경기술

[0002] 의사결정트리는 트레이닝 결과가 규칙형태로 직관적으로 변환가능하기 때문에 다양한 예측분야에서 예측결과와 원인분석 등에 이용되고 있으며, 정확도와 속도를 보장하기 때문에 예측분야에서 성공적으로 응용되고 있다.

[0003] 일반적인 트리기반 알고리즘은 트리성장 및 트리전지과정으로 구성된다. 트리알고리즘으로부터 구축되는 트리구조의 차이는 주로 트리 성장과정에서 이용되는 선택 속성기준, 트리전지기준으로부터 비롯된다.

[0004] 훈련데이터를 통해 의사결정 트리가 구축된 후, 루트노드부터 리프노드까지 일련의 노드 조합으로 규칙변환이 가능하다. 훈련데이터는 트리 구축과정에서 하나의 루트노드로부터 분기되어 최종적으로 리프노드에 분포하게 되며, 트리로부터 유도된 규칙간의 등급(혹은 우선순위)은 리프노드에 포함된 클래스 분포를 이용하여 계산된다. 이는 리프노드의 전체 클래스에 대한 다수 클래스의 비율로 계산되며, 소수의 클래스는 오분류로 간주된다. 공간예측 응용에서는 이와 달리 전체 클래스에 대한 예측 대상이 되는 타깃 클래스의 비율로 응용할 수 있다.

[0005] 의사결정 트리는 노드분류기준에 의해 순수한 분류를 지향하기 때문에 궁극적으로 리프노드의 사건클래스의 구성비는 0 또는 1로 수렴하게 되어, 이와 같은 리프노드 등급결정이 용이하지 못하다. 기존의 리프노드 등급방법들은 이러한 문제를 해결하기 위해 고안되었으며 주로 비공간 데이터에서 예측 응용을 위해 사용되었다.

[0006] 예측 정확도는 리프노드 등급결정 방법 따라 달라질 수 있다. Laplace estimate는 리프노드의 등급계산을 위해 리프노드에 포함된 사건클래스의 빈도를 라플라스 수정을 응용하여 기존 확률 추정문제를 완화하기 위해 사용되며, 공간예측응용을 위해 아래의 식1 과 같이 수정할 수 있다.

[0007] [식 1]

$$R(node) = \frac{n_{event} + 1}{n_{event} + n_{non_event} + c}$$

[0008] 여기서, c는 전체데이터 셋에서 클래스 수이다.
 [0009]

[0010] 또 다른 리프노드 평가방법인 M-estimate는 사건클래스에 대한 선행 확률을 이용한다. 공간예측응용에서는 파라미터 상수 b, m에 대하여 b가 사건발생에 대한 선행 확률일 때, 식 2과 같이 수정 정의된다.

[0011] [식 2]

$$R(node) = \frac{n_{event} + bm}{n_{event} + n_{non_event} + m}$$

[0012] m-branch 기법은 M-estimate의 변형으로 공간사건 예측을 위해 식 3과 같이 정의된다. 여기서, m은 노드의 깊이, 클래스 에 포함된 샘플의 수이다. 레이블이 정해지지 않는 샘플이다.
 [0013]

[0014] [식 3]

$$R(node.chld) = \frac{n_{event} + mR(node.parent)}{n_{event} + n_{non_event} + m}$$

[0015] 여기서, 파라미터 m은 $M+(d-1)/d \times M\sqrt{N}$ 으로 계산된다.
 [0016]

[0017] 기존 제시된 리프노드 등급방법들(Laplace estimate, m-branch 및 M-estimate)은 공간 다중클래스 분류를 위해 제시되었기 때문에, 앞서 정의된 식 1부터 식 3까지와 같이 공간예측 응용을 위해 전체 클래스에 대한 공간사건

발생 클래스 비 형태로의 수정을 통해 상대적인 발생 가능 정도를 표현할 수 있다.

선행기술문헌

특허문헌

[0018] (특허문헌 0001) 대한민국 특허등록번호 제10-1264223호(발명의 명칭: 의사결정 트리를 이용한 공간사건 예측분 석 방법)

비특허문헌

[0019] (비특허문헌 0001) [1] learning", In Proc. of the European Conf. on Artificial Intelligence, pp.147-149, 1990.

(비특허문헌 0002) [2] B. Zadrozny, and C. Elkan, "Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers", In Proc. 18th Int. Conf. on Machine Learning, pp. 609-616, 2001.

(비특허문헌 0003) [3] C. Ferri, P.A. Flach, and J. Hernandez-Orallo, "Improving the AUC of Probabilistic Estimation Trees", In Proc. of the 14th European Conf. on Machine Learning, pp. 121-132, 2003.

(비특허문헌 0004) [4] D.M. Cruden, and D.J. Varnes, "Landslide types and process. In:Turner, A.K. and R.L. Schuster(eds.), Landslides: investigation and mitigation. Transportation Research Board, Special Report, vol 247, pp. 36-75. 1996

(비특허문헌 0005) [5] F.J. Provost, and P. Domingos, "Tree Induction for Probability-based Ranking", Machine Learning, Kluwer Academic Publisher, vol. 52, pp. 199-215, 2003.

(비특허문헌 0006) [6] J. Cussents, "Bayes and psudo-bayes estimates of conditional probabilities and their reliabilities", Proceedings of European Conference on Machine Learning, pp. 136-152. 1993.

(비특허문헌 0007) [7] J.R. Quinlan, "C4.5: programs for machine learning", Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

(비특허문헌 0008) [8] R. Dikau, L. Schrott, D. Brunnsden, and M.L. Ibsen, "Landslide recognition: Identification, Movement and Causes", John Wiley & Sons: Chichester, UK, 1996, pp. 122-136.

(비특허문헌 0009) [9] T. Brandenburger, and A. Furth, "Cumulative Gains Model Quality Metric", Advances in Decision Sciences, doi.org/10.1155/2009/868215, 2009.

(비특허문헌 0010) [10] W.R. Tobler, "A computer movie simulating urban growth in the Detroit region", Economic Geographer, vol. 46, pp. 234-40, 1970.

(비특허문헌 0011) [11] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. MacLachlan, A. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand, and D. Steinberg, "Top 10 algorithms in data mining", Knowledge and Information Systems, vol. 14, pp. 1-37, 2008..

(비특허문헌 0012) [12] Y.K. Yeon, J.K. Han, and K.H. Ryu, "Landalide susceptiblity maping in Injae, Korea, using a decision tree", Engineering Geology, vol. 116, pp. 274-282, 2010.

발명의 내용

해결하려는 과제

[0020] 본 발명이 해결하고자 하는 과제는 의사결정트리에서 공간예측을 목적으로, 훈련데이터로부터 학습한 트리의 리

프노드를 상대적인 등급을 효율적으로 결정하기 위한 것을 목적으로 한다.

과제의 해결 수단

[0021] 상기 과제를 해결하기 위한 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정방법은 공간예측을 위한 훈련 데이터셋을 이용하여 하나의 루트노드와 더불어 부 노드와 복수개의 자 노드 관계를 갖는 의사결정트리 형태로의 학습단계 (S110); 상기 학습을 마친 의사결정 트리로부터 리프노드 등급 결정단계(S120)를 포함하고, 상기 학습단계(S110)에서 의사결정트리의 각 노드는 훈련데이터의 클래스 분포 개수를 저장하고 이를 저장하기 위한 구조를 포함하며, 상기(S120)의 리프노드 등급 결정단계에서는 루트노드로부터 리프노드까지 일련의 노드들을 조합에 각 노드에 저장된 클래스 분포수를 이용하여 리프노드 등급을 결정하는 것을 특징으로 한다.

[0022] 따라서 상기 훈련된 의사결정 트리를 구성하는 각 노드마다 등급이 결정되며, 부 노드와 자 노드간의 등급은 차는 다음 식 4로 정의되는 것을 특징으로 한다.

[0023] [식 4]

[0024]
$$R(\text{node.parent}) \leq R(\text{node.child})$$

[0025] 여기서, 부 노드(R(node.parent))와 자 노드(R(node.child))의 등급이 같은 때 자 노드에 공간사건 발생 클래스가 포함되어져 있지 않으며, 자 노드에 사건 발생 클래스가 존재할 때 부 노드보다 자 노드의 등급이 큰 경우를 나타낸다.

[0026] 상기 리프노드 등급결정 수행단계(S130)는 아래에 기재된 식 5을 이용하여 수행하는 단계인 것을 특징으로 한다.

[0027] [식 5]

[0028]
$$R(\text{node.child}) = R(\text{node.parent}) + \frac{n_{\text{event}}}{(n_{\text{event}} + n_{\text{non-event}}) \times d}$$

[0029] 여기서, d는 트리에서의 노드의 깊이를 나타낸다.

[0030] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시 예에 따른 의사결정트리를 이용한 공간사건 예측분석용 리프노드 등급 결정방법이 기록된 컴퓨터 판독 가능한 기록 매체를 제공한다.

발명의 효과

[0031] 따라서, 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정방법은 기존의 방법들을 이용하여 야기되는 문제점들을 개선하는데 사용할 수 있다.

[0032] 의사결정 트리는 트리구축을 위한 다양한 파라미터들이 있으며 파라미터들에 의해 트리 구조가 달라질 수 있다. 기존의 리프노드에 포함된 클래스 분포를 이용한 방법들은 이러한 트리 구조에 민감한 예측결과를 도출할 수 있다. 본 발명은 전체 트리노드의 클래스 분포를 이용하기 때문에 기존 예측 결과보다 덜 민감한 결과를 도출한다.

[0033] 공간데이터의 특징은 비교적 대용량 데이터이기 때문에 의사결정 트리에서 대용량의 공간 데이터를 이용하여 모델을 만들기 위해서는 많은 비용을 요구한다. 따라서 리프노드 등급 결정과정을 위한 별도의 사용자 정의의 파라미터는 반복적인 트리구축을 위한 비용이 소요되는 문제가 있다. 본 발명은 트리구조에 대해 민감한 결과를 도출하지 않기 때문에 별도의 파라미터가 없다. 따라서 트리구축에 따른 비용이 높지 않다.

[0034] 리프노드 등급 평가방법들은 규칙에 대한 상대적인 우선순위를 평가하기 위한 방들이기 때문에 규칙 적용범위가 반영이 되지 못하다. 즉 규칙에 대한 적용범위가 등급 결정에 반영되지 못했다. 본 발명의 규칙 평가방법은 부

노드의 평가결과를 이용하기 때문에 규칙 적용 범위가 효과적으로 반영되는 특징이 있다.

[0035] 따라서 본 발명에서 제안하는 의사결정트리에서 공간예측을 위한 리프노드 평가방법은 기존에 보였던 다양한 문제점 들을 해결하며 효과적으로 공간예측을 수행할 수 있는 이점을 갖는다.

도면의 간단한 설명

[0036] 도 1은 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정 방법을 설명하기 위한 흐름도이다.

도 2는 공간사건으로부터 훈련된 의사결정 트리와 공간좌표 상에 표현된 공간사건 발생위치도이며, 의사결정 트리에서 숫자와 함께 표시한 리프노드들은 공간사건 발생 위치도에서 번호와 함께 구분한 직사각형 영역들과 각각 대응 한다.

도 3은 본 발명에서 제시한 리프노드 등급 결정 방법과 기존의 방법들(Laplace, M-branch, m-estimate)을 이용하여 얻어진 결과들의 누적이득차트를 나타낸 도이다.

도 4는 리프노드 등급들 간의 정량적인 성능 비교를 위한 AUC를 나타낸 도이다.

도 5는 본 발명에 대한 성능을 평가하기 위해 공간사건 데이터를 두 개의 서브 그룹(사각형(□)으로 묘사된 LandslideSetA, 원(○)으로 묘사된 LandslideSetB)으로 나누어 교차 검증한 결과의 도면이다.

발명을 실시하기 위한 구체적인 내용

[0037] 이하, 본 발명의 바람직한 실시 예의 상세한 설명은 첨부된 도면들을 참조하여 설명할 것이다. 본 발명을 설명함에 있어서, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다.

[0038] 본 발명의 개념에 따른 실시 예는 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있으므로 특정 실시 예들을 도면에 예시하고 본 명세서 또는 출원에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시 예를 특정한 개시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0039] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0040] 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0041] 이하, 도면을 참조하여, 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정방법을 보다 상세하게 설명하도록 한다.

[0042] 도 1은 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정 방법을 설명하기 위한 흐름도이다.

[0043] 도 2는 공간사건으로부터 훈련된 의사결정 트리와 공간좌표 상에 표현된 공간사건 발생위치도이며, 의사결정 트리에서 숫자와 함께 표시한 리프노드들은 공간사건 발생 위치도에서 번호와 함께 구분한 직사각형 영역들과 각각 대응 한다.

[0044] 도 3은 본 발명에서 제시한 리프노드 등급 결정 방법과 기존의 방법들(Laplace, M-branch, m-estimate)을 이용하여 얻어진 결과들의 누적이득차트를 나타낸 도이다.

[0045] 도 4는 리프노드 등급들 간의 정량적인 성능 비교를 위한 AUC를 나타낸 도이다.

[0046] 도 5는 본 발명에 대한 성능을 평가하기 위해 공간사건 데이터를 두 개의 서브 그룹(사각형(□)으로 묘사된 LandslideSetA, 원(○)으로 묘사된 LandslideSetB)으로 나누어 교차 검증한 결과의 도면이다.

[0047] 먼저, 도 1에 도시된 바와 같이, 본 발명의 실시 예에 따른 의사결정트리에서 공간 예측을 위한 리프노드 순위 결정 방법(S100)은 공간예측 데이터 준비단계(S110), 의사결정트리 구축단계(S120), 리프노드 등급 산정단계(S130) 및 예측결과 도출단계(S140)를 포함한다.

[0048] 상기 공간예측 데이터 준비단계(S110)는 의사결정 트리가 학습할 데이터인 예측 대상이 되는 공간사건 발생 레이어와 공간사건 발생 유발환경 레이어들에 대하여, 위치 대응할 수 있도록 동일한 범위를 갖고 기하 보정하는 단계일 수 있다.

[0049] 상기 의사결정트리 구축단계(S120)는 상기 공간예측 데이터를 훈련 데이터로 이용하여 하나의 루트노드와 더불어 부 노드와 복수개의 자 노드 관계를 갖는 의사결정트리를 구축하며, 각 노드는 리프노드 등급 산정을 위해 훈련데이터가 분류된 사건발생과 사건 미발생 클래스 개수 정보를 저장하는 구조를 취할 수 있다.

[0050] 상기 리프노드 등급 산정단계(S130)는 상기 학습과정을 통해 구축된 의사결정트리를 이용하여 루트노드로부터 계산된 등급 계산 결과를 자 노드에 반영되며 전체 리프노드에 이르기까지 각 노드에 포함된 사건 및 미사건 발생 클래스의 개수를 이용하여 등급을 계산할 수 있다.

[0051] 상기 예측결과 도출단계(S140)는 상기 리프노드의 등급값(우선순위)을 공간 좌표상에 매핑하는 단계이다.

[0052] 여기서, 상기 리프노드 등급 산정단계(S130)는 상기 0과 1로의 극단적인 등급 값으로의 계산결과를 방지하기 위해 부 노드로부터 등급값을 상속받아 동일 부 노드에서 분기된 자 노드의 등급 값과 상호 보상되는 것을 특징으로 한다. 이러한 특징은 상기 자 노드(R(node.child))와 상기 부 노드(R(node.parent))간의 등급 계산에서 식 4와 같이 정의된다.

[0053] [식 4]

$$R(\text{node.parent}) \leq R(\text{node.child})$$

[0054]

[0055] 여기서, 부 노드(R(node.parent))와 자 노드(R(node.child))의 등급이 같은 때 자 노드에 공간사건 발생 클래스가 포함되어 있지 않으며, 자 노드에 사건 발생 클래스가 존재할 때 부 노드보다 자 노드의 등급이 크다.

[0056] 예를 들어, 도 2를 참조하면, 좌측의 의사결정 트리는 우측에 묘사된 공간사건 발생위치로부터 유도 되었다고 할 때, 의사결정 트리는 루트노드는 플러스(+)로 묘사된 사건발생 클래스 개수와 마이너스(-)로 묘사된 사건 미발생 클래스 개수를 포함한다. 이후 트리는 분류 기준에 따라 자 노드들을 형성하게 된다. 이러한 클래스 분포를 포함한 트리의 구축결과를 이용하여 리프노드 등급을 산정하게 되며 이때, 본 발명에서 제시하는 리프노드 등급단계(S130)는 아래에 기재된 식 5를 이용하여 수행한다.

[0057] [식 5]

$$R(\text{node.child}) = R(\text{node.parent}) + \frac{n_{\text{event}}}{(n_{\text{event}} + n_{\text{non-event}}) \times d}$$

[0058]

[0059] 여기서, d는 트리에서의 노드의 깊이를 나타낸다.

[0060] 식 5를 통해 도 2의 의사결정 트리의 리프노드 등급 결과는 R(leaf1) = 0.008, R(leaf2) = 0.017, R(leaf3) = 0.024 그리고 R(leaf4) = 0.019로 계산된다. 계산된 각 리프노드 등급들은 도 2의 리프노드 번호와 대응되는 직사각형에 매핑되어 상대적인 등급 값을 통해 공간사건 예측결과가 만들어 진다.

[0061] 이하에서는 본 발명에서 제시하는 의사결정트리를 이용한 공간사건 예측분석용 리프노드 순위 결정 방법을 산사

태 공간사건에 적용시킨 실험 예를 설명하도록 한다.

[0062] 먼저, 산사태는 전형적인 공간사건으로 주로 집중강우, 지진으로 인해 기인되지만 지형, 지질(토양), 임상과 같은 환경적 조건에 의해 발생여부 혹은 규모가 달라질 수 있다. 따라서 산사태 공간사건의 예측은 과거에 발생된 위치의 유사한 조건을 통해 알려지지 않은 미래사건의 예측을 할 수 있다.

[0063] 본 실험에 적용하기 위한 자원들은 다음과 같다.

[0064] 먼저, 해당 지역은 강원도 지역의 인제읍과 북면 사이에 지역이며, 산사태 위치들은 항공이미지 및 삼각불규칙망(triangulated irregular network)을 분석하여 추출된 산사태 위치들일 수 있다. 참고로, 삼각불규칙망은 공간을 불규칙한 삼각형으로 분할하여 생성된 일종의 공간자료 구조, 지형의 경사, 주향, 체적, 표면길이, 단면도의 생성, 하천, 선의 생성, 능선 추출 및 기사도 분석 등을 포함한 지표면 및 지형분석 TIN 자료에 사용되는 망일 수 있다.

[0065] 지형도(topographic map) 및 산림도(forest map)는 산사태 관련 요인들을 추출하는 데 사용되는 맵 자료들일 수 있다. 지형적 요인들(Topographical factors)(예를 들어, 측면, 곡률, 산마루 및 경사지도)는 산사태 발생과 관련된 지형학적 요인들이다.

[0066] 산림도 요인들(나이, 밀도, 직경 및 형태)은 산사태 발생을 막는데 도움을 주는 초목의 상태를 반영한다. 토양 요인들(배열구조, 배수, 재질 및 두께)은 산사태 규모에 영향을 미치는 요인들이다.

[0067] [표 1]

Map Source	Thematic Layer	Scale
<i>Airborne Image</i>	Landslide	0.4m
<i>Topographic map</i>	Aspect	1:25,000
	Curvature	
	Ridge	
	Slope	
<i>Forest map</i>	Age	1:25,000
	Density	
	Diameter	
	Type	
<i>Soil map</i>	Texture	1:25,000
	Drainage	
	Material	
	Thickness	

converted into 5 × 5 m float-type raster images of 1,387,870

[0068] 표 1은 공간사건들의 구성요소들 및 환경 데이터셋을 나타낸 표이다.
 [0069]

[0070]
 [0071] 도 5를 참조하면, 공간 사건위치들이 두 그룹 즉 사각형(□)으로 묘사된 LandslideSetA, 원(○)으로 묘사된 LandslideSetB이 있을 때, 도 5(a)는 LandslideSetA로 말두고 LandslideSetB를 이용하여 예측모델을 평가했으며, (b)는 이와 반대로 평가하여 두 결과의 평균을 산출하였다.

[0072] 즉, 도 4를 참조하면, 산사태 공간사건 데이터를 두 개의 서브 그룹으로 나누어 교차 검증한 결과 86.09%로 높은 예측 평가 결과를 보였다. 이와 같은 본 발명은 다른 리프노드 등급방법들보다 별도의 최적의 파라미터 값을 찾는 과정을 필요치 않으면서 높은 성능을 보인다. 이러한 원인은 단순 리프노드의 클래스 분포를 이용하는 것이 아니라 부 노드의 등급 결과를 자 노드 등급계산과정에서 반영하기 때문에 극단적인 등급결과 도출을 피할 수 있도록 하기 때문이다.

[0073] 본 발명에서 제시하는 의사결정트리를 이용한 공간사건 예측분석용 리프노드 순위 결정 방법은 다양한 컴퓨터

수단을 통하여 판독 가능한 프로그램 형태로 구현되어 컴퓨터로 판독 가능한 기록매체에 기록될 수 있다. 여기서, 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0074] 예컨대 기록매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광 기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함한다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드 뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함할 수 있다. 이러한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0075] 따라서, 본 발명의 실시 예에 따른 의사결정트리에서 공간예측을 위한 리프노드 등급 결정방법에서는 일반적으로 분류기의 특성에 따라 사건 클래스에 대한 신뢰도로 간주될 수 있는 등급값을 트레이닝 데이터로부터 유도하거나, 확률을 추정할 경우, 추정된 클래스의 분균형이 매우 큰 데이터세트에서의 사건발생에 대한 추정확률이 일반적으로 매우 낮기 때문에 상대적 등급으로 고려하여야 하는 문제점을 해결하기 위하여, 부 노드의 등급값을 자 노드 등급 계산과정에 반영하도록 하여 하나의 부 노드로부터 분리된 자 노드간의 급격한 등급 변화를 줄일 수 있다. 따라서 예측성 성능을 향상시킬 수 있다는 이점을 제공한다.

[0076] 이전의 방법들은 획일적 리프노드의 클래스 분포를 이용하여 분산으로 등급값을 추정하여, 불균형이 매우 큰 데이터세트에는 적합하지 않았으며, 최적의 파라미터 값을 요구하는 부가적인 과정이 필수적이었다.

[0077] 또한, 이전의 리프노드 우선순위 결정 방법들은 오직 비공간 응용들에서만 고려되어 왔다.

[0078] 반면에 본 발명에서 제안하는 의사결정트리를 이용한 공간사건 예측분석용 리프노드 순위 결정 방법은 변수들 값을 찾기 위한 추가 절차가 필요 없다는 이점을 갖는다.

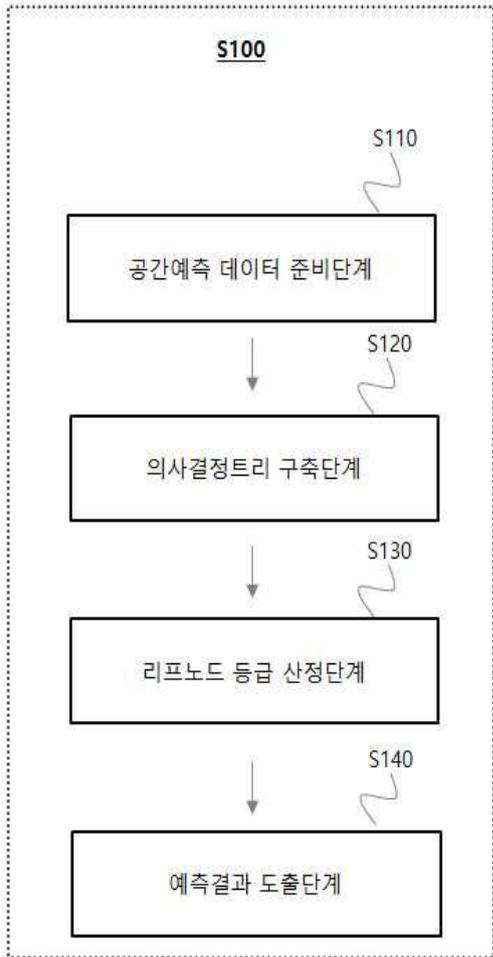
[0079] 이상 본 발명이 양호한 실시 예와 관련하여 설명되었으나, 본 발명의 기술 분야에 속하는 자들은 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에 다양한 변경 및 수정을 용이하게 실시할 수 있을 것이다. 그러므로 개시된 실시 예는 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 하고, 본 발명의 진정한 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

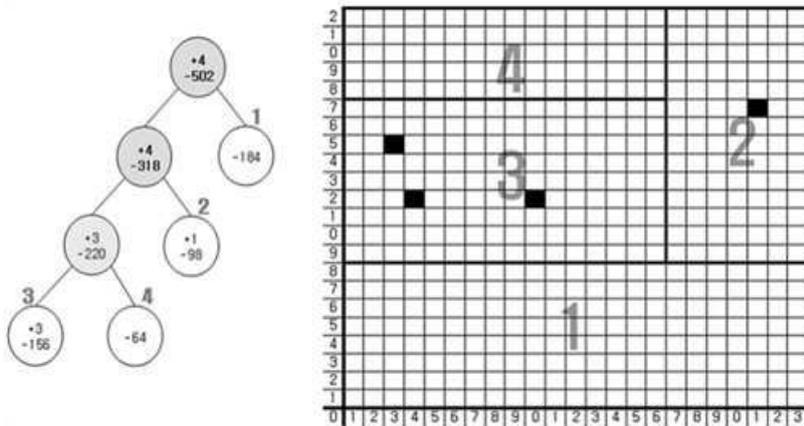
- [0080] S100: 리프노드 등급 결정 방법
- S110: 공간예측 데이터 준비단계 S120: 의사결정트리 구축단계
- S130: 리프노드 등급 산정단계 S140: 예측결과 도출단계

도면

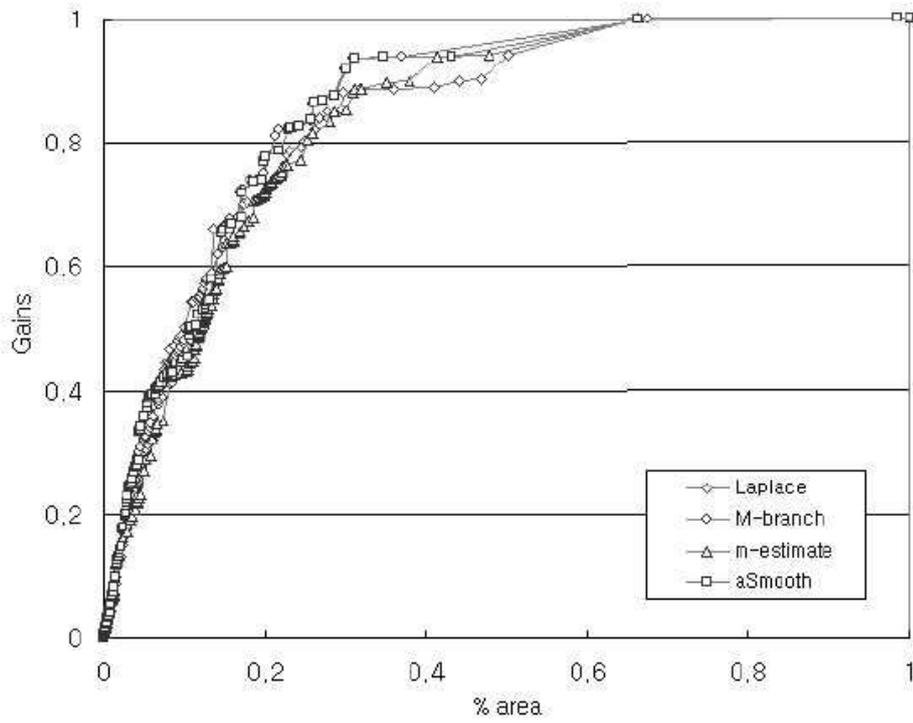
도면1



도면2



도면3

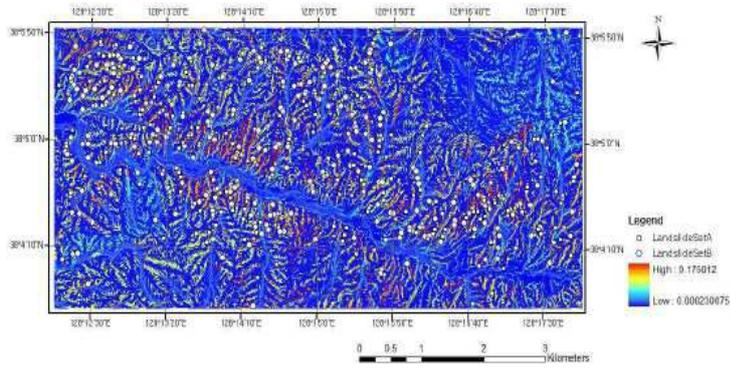


도면4

리프노드 평가방법	AUC
M-estimate	84.39
Laplace	84.72
m-branch	86.07
본 발명	86.09

도면5

a



b

