

데이터 마이닝을 이용한 자연재난 경고

유준석*, 최원길*, 김응모*
*성균관대학교 정보통신공학부
e-mail:scv82nim@hanmail.net

Natural Disaster Alarming using Data Mining

Joon-Suk Ryu*, Won-Gil Choi*, Ung-Mo Kim*
*School of Information and Communication Engineering, Sungkyunkwan
University

요 약

지구 온난화와 도시화 등으로 인해 자연 재난은 매년 그 강도는 증가하고 있고 그 결과 재난으로 인한 피해도 증가하고 있다. 우리나라 역시 매년 자연재난으로 많은 피해를 입고 있다. 이러한 재난을 지역의 특성을 이용하여 분류한 후 특성이 유사한 지역을 모아 패턴을 찾게 되며 패턴은 데이터 마이닝을 이용해 찾게 된다. 데이터 마이닝이란 여러 가지 데이터 안에 의미 있는 패턴을 찾아내는 기술이며 여기서 찾게 되는 패턴은 지금까지 발생한 재난의 발생 직전까지의 패턴이 된다. 이렇게 찾아낸 패턴은 현재 연속적으로 변하는 환경의 패턴 비교되며 패턴이 유사할 경우 경고하여 재난이 발생하기 전 먼저 예측하고 대비하는 방법을 제시한다.

1. 서론

최근 지구는 지구온난화의 영향으로 전 세계에 걸쳐 자연 재난이 증가하고 있을 뿐만 아니라 그로 인한 피해 또한 급격히 늘고 있는 상황이다. 우리나라의 경우 도시화, 산업화에 따라 재난으로 생기는 피해가 대규모화 되는 동시에 재난위험요인 또한 증가되고 있고 정부의 지속적인 수해 방지 및 물 관리 대책 추진에도 불구하고 해마다 자연재난에 의한 막대한 재산 피해와 인명 피해를 입고 있다[1]. 이렇게 매년 증가하는 재난의 강도로 인해 현대 사회는 지금까지의 재난 이후의 복구 및 보상위주의 재난관리 패러다임을 미래형 패러다임으로의 전환이 요구되고 있는 시점이다[2]. 본 논문에서는 지역별로 공간을 나눈 후 그 지역에 해당하는 데이터를 얻어 지역별 재난 예측, 경고를 함으로서 재난으로 인한 피해를 최소화하는 방법을 제안한다. 본 논문은 다음과 같이 구성되어 있다. 2장은 GIS, Bio Ad Hoc네트워크, 지진계, 기상장비와 마이닝 기법에 대해 간략하게 설명한다. 3장에서는 데이터마이닝을 통한 예측 및 경고 방식에 대해 기술한 후, 마지막 4장에서는 결론을 맺고자 한다.

2. 관련 연구

1. 데이터 마이닝 기법 (Sequence Pattern Mining)

연속 규칙은 연관 규칙에 시간 관련 정보가 포함된 형태로서, 시간상에 순차적으로 나타나는 사건이나 거래의 종속 관계를 찾아내는 작업이다[7].

본 논문에서 사용되는 DynamicSome 데이터마이닝 기법은 ApriorSome algorithm과는 다르게 step이라는 변수를 이용하여 특정한 길이의 sequence만을 산출한다. 또한 Backward 단계에서 만들어지던 후보 sequence가 intermediate 단계에서 생성된다. DynamicSome은 initialization 단계, forward 단계, intermediate 단계로 3단계로 구성 되어있으며 마이닝 전 데이터를 알기 쉬운 연속된 숫자로 바꿔준다[7][그림 2].

환경변화	Mapped TO
지진파변화	1
라돈방출	2
전기저항값	3
동물이동	4
지진활동변화	5

그림 1 데이터 Mapping

1.1 initialization 단계

첫 단계에서는 step을 포함한 길이까지의 모든 sequence를 산출한다. 즉 step이 3일 경우 길이가 1인 sequence부터 길이가 3인 sequence까지 산출을 하게 된다 [그림 2].

1.2 Forward 단계

이 단계에서는 step의 배수 길이의 sequence만을 산출한다. initialization 단계에서는 Apriori 후보 생성 알고리

즘을 사용하지만 DynamicSome에서는 on-the-fly라는 방식의 후보 생성 알고리즘을 사용한다. 이 방법은 otf라는 함수에 길이가 k인 sequence, 길이가 j인 sequence와 어떠한 sequence의 집합인 c를 인자 값으로 넘겼을 경우 j+k 길이의 sequence중 c안에 포함되어 있는 sequence를 반환한다. 예를 들면 길이 3의 sequence는 길이 1과 2의 sequence를 조인하여 만들 수 있으며 길이 6의 sequence는 길이 3의 sequence를 조인하여 만들 수 있다[그림 2][그림 3].

1.3 Intermediate 단계

이 단계는 aprioriSome algorithm의 Backward단계와 후보를 생성한다는 점을 제외하곤 모두 동일하다. Forward단계에서 산출되지 않은 길이의 sequence를 찾아내는 단계이다. 예를 들면 길이가 3,6,9인 sequence를 forward 단계에서 산출했고 길이가 9인 sequence에 아무 것도 남아 있지 않다면 sequence길이가 7을 산출한 후 8을 산출하여 Maximum을 찾아낸다[그림 2].

Initialization 단계

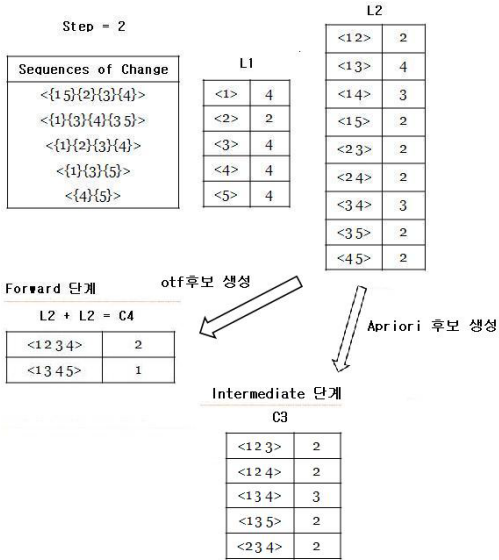


그림 2 DynamicSome의 마이닝 과정

1.4 On-The-fly 후보 생성

otf는 인자로 길이가 k인 sequence와 길이가 j인 sequence, sequence의 집합인 c를 받는다. 그 후 j-sequence와 k-sequence가 c에 모두 포함되어있고 서로 오버랩 되지 않았을 경우 (k+j)-sequence가 후보 sequence가 된다. 또한 조인할 때 X2.end < X2.start (X2는 길이가 2인 sequence를 나타낸다.)를 만족 시킬 경우에 조인을 한다[그림 3].

X2.X2와 <{1} {2} {3} {7} {4}>을 인자로 otfg가 호출 된경우

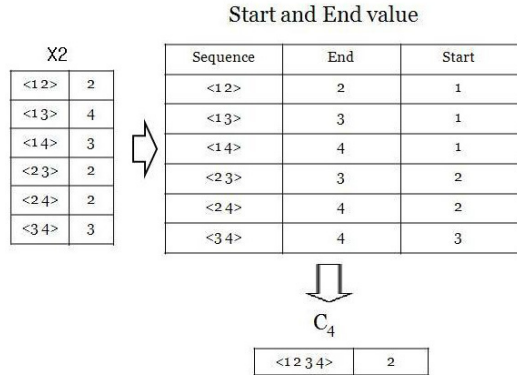


그림 3 OTF 후보생성 과정

2. GIS

Geographic Information System 는 지구상에서 발생하는 공간현상의 제반현상들의 위치, 속성정보를 결합하여 컴퓨터를 통해 입력 저장하고 검색, 갱신 등 정보를 관리하고 처리 및 분석을 통하여 사용자에게 원하는 정보를 제공하는 정보시스템이며 의사결정지원체계이다[3]. 최초 GIS는 자원관리를 목적으로 개발 되었고 고가의 컴퓨터 비용으로 인해 정부에 한정되어 사용 되어졌지만 컴퓨터 기술의 발달과 그래픽 처리기술의 발달로 공공기관에 의해 GIS는 발전하게 되었고 현재 많은 나라에서 여러 가지 분야에 적용하여 사용되고 있다[5]. 본 논문에서는 GIS에 저장된 정보를 이용하여 해당 지역의 특성 등을 파악한다.

3. Bio Ad Hoc 네트워크

Ad Hoc 네트워크는 기존의 네트워크와 같이 네트워크 인프라가 구축된 상태에서 통신을 수행하는 것이 아닌 인프라가 존재하지 않은 상태에서 각 단말들 상호간의 라우팅으로 데이터 송·수신 등의 통신 기능을 수행할 수 있는 형태의 네트워크를 말한다[6]. 이러한 Ad Hoc 센서를 동물에 탑재하여 Bio Ad Hoc 네트워크가 구성된다[4]. 이 분야는 현재 많은 연구가 진행되어 왔으며 저비용 및 효율적으로 전송하기 위한 연구도 계속 되어가는 중이다. 이 논문에서는 이러한 Bio Ad Hoc을 이용해 동물들의 재난이 일어나기 직전 변화를 측정하여 재난에 대한 동물들의 행동데이터를 측정한다.

4. 지진계 및 기상 장비

기상장비 및 지진계는 현재 한국의 기상청에서 사용 되는 장비들로 지진계는 지진이 있을 경우 지진의 움직임, 세기와 위치 등을 기록하는 기계로 우리나라에는 7 가지의 종류가 존재하고 기상 장비는 지구의 풍향 및 온도 등을 측정하는 장비로 현재 27가지가 사용되고 있다[3].

3. 데이터 마이닝을 통한 예측과 경고 및 데이터 공유

본 장은 일반 적인 데이터 마이닝과 매우 유사한 방식으로 진행되며 데이터베이스는 hierarchy 방식으로 되어 있다. 최하위 노드는 지역은 넓이에 따라 주, 시, 읍 등으로 나눌 수 있고 루트 노드는 모든 데이터를 가지고 있는 데이터베이스와 Data Warehouse이다[그림 4].

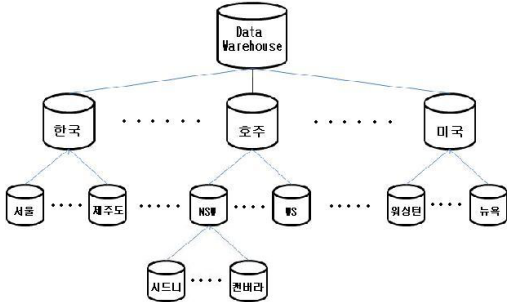


그림 4 데이터베이스 전체구조

3.1 데이터 수집

데이터는 Bio Ad Hoc, 지진계와 기상 장비를 통해서 설정해 놓은 변화가 있을 시에만 해당 지역의 데이터베이스에 입력되며 시간도 함께 입력한다. GIS는 지역의 특징 정보를 가지므로 지역의 특성이 변하지 않는다면 업데이트가 필요하지 않고 모든 데이터는 라이프 타임을 갖으며 라이프타임이 다하면 저장되어있던 데이터는 삭제하여 데이터 저장 공간의 낭비를 방지한다.

3.2 상위 노드로의 데이터 전송

데이터는 재난이 발생할 경우에만 그 상위의 데이터베이스에 전송되어 진다. 전송 시에는 재난이 발생당시 존재하는 모든 데이터를 상위 노드로 전송하게 된다.

3.3 데이터 패턴 마이닝

최상위 노드는 하위 노드로부터 받은 데이터를 재난의 종류별로 나눈 다음 재난의 종류에 따라 재난에 가장 많은 영향을 주는 지역 환경을 가지고 데이터를 Clustering 한다[그림 5]. 그 후 환경 변화를 계산하기 쉽게 숫자로 mapping한 후 연속 규칙 마이닝 기법의 하나인 Dynamic Algorithm을 통해 패턴을 찾는다[7][그림 1][그림 2]. [그림 6]은 [그림 1]의 매핑을 사용한 [그림 2]의 결과로 나온 패턴이다. 그 결과 지진은 발생 전 {지진과변화, 라돈방출, 전기 저항 값, 동물이동}라는 패턴을 갖는 것을 찾게 된다.

지진, {지진파변화, 라돈가스 방출, ..., 동물이동}, GIS, 기후정보
 ↓ 하위노드로부터 들어온 데이터

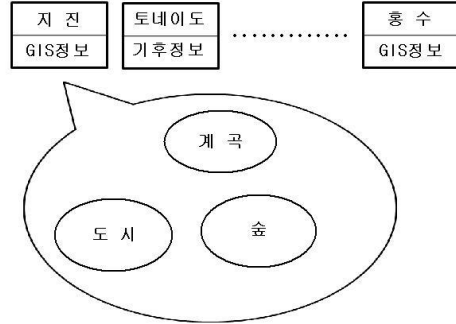


그림 5 Clustering을 이용한 지역 분류

3.4 하위 노드로의 마이닝된 패턴 전송

하위 노드는 상위 노드에 저장되어 있는 축적된 패턴들 중 자신의 GIS정보와 유사한 정보의 패턴을 가지고 온다. 이렇게 상위 노드에서 가져온 정보를 이용하여 재난을 예측하고 경고를 하게 된다.

3.5 재난 예측

데이터 분석은 데이터의 변화가 있을 때만 실행하며 변화가 있는 데이터는 자신이 속해있는 재난 패턴을 찾는다 [그림 7]. 그리고 데이터를 상위 노드에서 받아온 패턴과 비교하여 유사도를 측정하며 유사도의 값에 따라 경보를 올린다[그림 8].

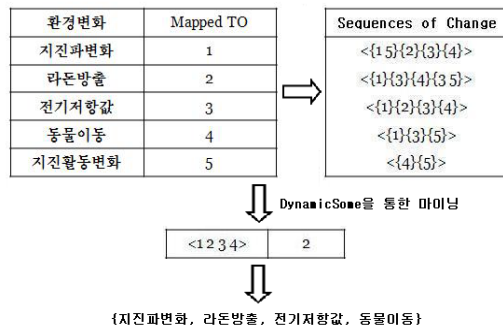


그림 6 마이닝을 통한 패턴 추출

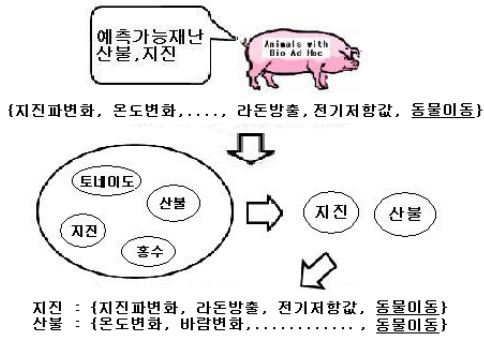


그림 7 변화가 속해있는 재난 찾기

[5] 과학기술정책관리연구소, “GIS 기술개발 연구기획사업”, 과학기술처
 [6] 임지형, 이병길, 김현근, 정교일, 양대현 “유비쿼터스 및 Ad Hoc 네트워크 망에서의 정보보호 분석”
 [7] Rakesh Agrawa and, Ramakrishnan Srikant, “Mining Sequential Patterns”,

4. 결과

본 논문에서는 증가하는 자연 재난으로 인한 인명피해와 재산 피해를 데이터 마이닝 기법을 통해 최소화하는 방법을 제안하였다. 유사한 지역들의 재난 데이터를 마이닝하여 재난이 발생되기 전 생기는 패턴을 찾아낸 후 이 패턴과 현재의 패턴을 비교하여 예측하고 경고를 함으로서 지금까지의 재난 이후의 복구 및 보상위주의 재난관리 시스템에서 실시간 재난 경고 시스템으로 변화가 가능하게 된다. 향후 연구로는 현재 국한된 지역을 좀 더 큰 범위로 넓히기 위한 연구와 최소의 데이터로 재난을 예측할 수 있는 연구가 필요하다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었으며(IITA-2008-C1090-0801-0028), 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 지식경제부의 유비쿼터스 컴퓨팅 및 네트워크 원천 기반기술 개발사업의 08B3-B1-10M 과제로 지원된 것임.

5. 참고문헌

[1] 이재은, 양기근, “지속 가능한 재난 관리의 효율화 방안 연구”, 현대사회와 행정(2005. 4.) 제 15권 제 1호, 105-135.
 [2] 이재은, 김검훈, 류상일, “미래 사회의 환경변화와 재난관리시스템 발전 전략”, 현대사회와 행정 (2005, 12) 제 15권 제 3호, 53-83
 [3] 기상청 홈페이지, <http://www.gis.co.kr/>
 [4] 이동은, 이구연, “바이오 애드 혹 네트워크를 이용한 재난 조기 경고 시스템 설계 및 분석”, 2006년도 한국정보과학회 가을 학술발표논문집 Vol. 33, No. 2(A)