



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0053578
(43) 공개일자 2009년05월27일

(51) Int. Cl.

G06Q 50/00 (2006.01)

(21) 출원번호 10-2007-0120469

(22) 출원일자 2007년11월23일

심사청구일자 2007년11월23일

(71) 출원인

세종대학교산학협력단

서울 광진구 군자동 98 세종대학교

(72) 발명자

백성욱

서울 서초구 잠원동 대림아파트 8-1305

김충곤

서울 관악구 신림12동 748-26 두영빌라 301호

전체 청구항 수 : 총 3 항

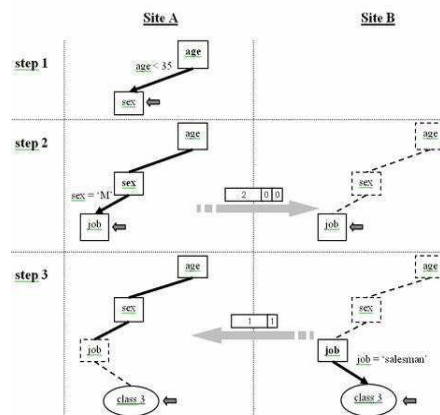
(54) 분산환경 기반 정보예측 시스템

(57) 요약

본 발명은 데이터마이닝 기술을 통해 생성된 예측모델을 이용하여 정보를 예측하고자 하는 상황에서, 분석에 필요한 데이터가 부족할 때 분석에 필요한 데이터를 다른 사이트에서 실데이터의 이동 없이 가상으로 공유하여 정보를 예측할 수 있는 시스템이다.

본 기술을 활용하면 불완전한 정보를 가진 사이트에서도 다른 사이트의 정보를 활용하여 예측모델을 검색하는 것이 가능하다.

대표도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 10643

부처명 서울시

연구사업명 신기술 연구개발 지원사업

연구과제명 유비쿼터스 컴퓨터 환경을 위한 분산형 데이터마이닝 기술

주관기관 세종대학교 산학협력단

연구기간 2005년12월1일 ~ 2007년11월30일

특허청구의 범위

청구항 1

에이전트와 미디어이터를 구비한 분산환경에서 정보예측모델로부터 정보를 예측하는 방법에 있어서

에이전트들은 미디어이터에 접속함으로써 가상으로 정보를 공유할 준비를 하는 제 1단계와;

사용자가 불완전한 데이터를 가지고 미디어이터로 예측모델을 요청하는 제 2단계와;

미디어이터가 지식기반 예측모델에 사용될 DB속성들의 일부를 포함하는 에이전트가 있다면 가상공유 정보기반 의사결정에 참여하도록 하는 제 3단계와

미디어이터는 예측모델(의사결정나무)이 없는 에이전트에게 예측모델을 전송하는 제 4단계와;

예측모델(의사결정나무)의 첫 번째(그 후 다음 번째) 속성을 포함한 에이전트는(선택 에이전트) 그 속성에 해당되는 데이터 값을 기반으로 의사결정나무의 구조를 따라 의사결정 과정을 진행하고 그 외의 참여 에이전트들은 수행된 에이전트의 수행 결과를 기다리는 제 5단계와;

선택 에이전트는 의사결정 과정에서 생성된 의사결정나무 경로 정보를 미디어이터로 보내는 제 6단계와;

6단계에서 받은 의사결정나무 경로 정보를 선택 에이전트 외의 모든 에이전트에게 전송하는 제 7단계와;

전달된 의사결정나무 경로정보를 통해 의사결정을 수행을 진행하여 선택 에이전트와 동일한 상태로 되는 제 8단계와;

에이전트는 의사결정 과정이 추가로 남아 있는지를 판단하고 단계 3부터 의사결정 과정이 추가로 없을 때까지 반복하는 제 9단계;및

모든 에이전트가 미디어이터로 종료 신호를 보내고 시스템을 종료하는 제 10단계를 포함하는 분산환경에서 정보를 예측하는 방법.

청구항 2

청구항 1에 있어서, 상기 제 6단계에 있는 정보예측모델로부터 정보를 검색 시 분산된 정보를 가상으로 공유하는데 있어서 실데이터를 공유하지 않고 의사결정 경로 정보만 공유하는 방법

청구항 3

각 에이전트가 정보예측을 위한 의사결정나무를 최종적으로 만들 때 자신이 의사결정 진행에 필요한 속성 정보를 가지고 있을 때는 분기를 하여 의사결정을 진행하고 부분적인 의사결정 경로 정보를 다른 에이전트에 전달하여 최종적으로 불완전한 정보를 가진 사이트에서 의사결정을 완전히 진행하여 정보를 예측할 수 있는 방법.

명 세 서

발명의 상세한 설명

기술 분야

<1> 본 발명은 각 사이트에 있는 정보를 가상으로 공유하여 정보를 예측 할 수 있는 방법에 관한 것이다.

배경 기술

<2> 일반적으로 정보예측을 하기 위해 필요한 정보들은 여러 사이트에 저장되어 있다. 예를 들어 국민의 특성에 대한 정보모델로부터 정보예측을 하기 위해서는 연봉, 나이, 성별, 소득, 소비, 입원 여부 등 국민(개인적) 특성에 대한 많은 데이터가 필요하다. 정보예측에 필요한 정보들을 알기 위해서는 개인이 속한 회사에 저장되어 있는 DB의 정보, 카드회사의 DB에 저장되어 있는 이용정보, 병원의 이용기록 등이 필요하다. 개개인의 정보는 그 특성상 매우 개인적이며 타인에게 알려질 경우 악용될 소지가 있기 때문에 정보이동에 보안이 매우 중요하다.

<3> 종래기술은 중앙집중식으로 필요한 데이터를 한 곳에 모아 정보를 예측해야 하기 때문에 위에 언급한 것과 같이 개인의 프라이버시를 침해할 요소가 크다.

발명의 내용

해결 하고자하는 과제

<4> 종래의 기술을 이용한 정보예측 시스템은 정보 이동의 한계로 인하여 다양한 분야의 정보를 적극적으로 활용하기 매우 어렵다.

과제 해결수단

<5> 본 발명을 통해 필요한 데이터를 실데이터의 이동 없이 의사결정나무 경로 정보 공유만으로 예측모델을 검색하는 데에 있다.

효 과

<6> 이상 설명한 바와 같이 본 발명에 의하면, 기존의 정보예측 시스템의 문제점인 실데이터의 이동에 따른 개인정보 유출 위험과 그에 따라 기존의 정보예측 시스템에서는 서로 정보 교환을 할 수 없어서 정보예측을 위해 정보를 제공하지 못했던 각 분야의 방대한 정보를 활용함으로써 실질적이고 유용한 정보예측을 할 수 있다는 특징이 있다.

<7> 도 4에 따르면 정보예측모델로부터 의사결정을 진행할 때 기존 중앙집중식 정보예측 시스템을 이용하면 정보예측모델에 필요한 DB속성을 모두 가지고 있지 않기 때문에 E, F, G 에 관련된 정보는 예측하기가 어렵다. 반면 분산기반 정보예측 시스템을 이용하면 필요한 정보를 다른 사이트로부터 가상으로 공유할 수 있기 때문에 E, F, G에 관련된 정보를 예측함으로써 더욱 정확한 예측이 가능해 진다.

발명의 실시를 위한 구체적인 내용

<8> 본 발명은 정보예측에 필요한 데이터를 실 데이터의 이동 없이 의사결정나무 경로 정보 공유를 통하여 정보를 예측 할 수 있는 알고리즘을 구현한 것이다. 본 알고리즘에 의하면 한 사이트에 필요한 정보를 모두 통합하여 정보를 예측 했을 때와 동일한 결과를 얻을 수 있다.

<9> 본 발명은 미디에이터-에이전트 기반의 시스템으로서, 모바일 기기나 유비쿼터스 환경과 같은 분산 환경에서도 가능하다.

<10> 이하, 본 발명의 실시 예에 대한 구성 및 그 작용을 첨부한 도면을 참조하면서 상세히 설명하기로 한다.

<11> 도면 1은 본 발명의 에이전트-미디에이터 기반 알고리즘의 흐름을 나타낸다. 분산기반 정보예측 시스템의 알고리즘은 다음과 같다.

<12>

<13> 1단계 에이전트들은 미디에이터에 접속함으로써 가상으로 정보를 공유할 준비를 함.

<14> 2단계 사용자가 불완전한 데이터를 가지고 미디에이터로 예측모델을 요청함

<15> 3단계 미디에이터가 지식기반 예측모델에 사용될 DB속성들의 일부를 포함하는 에이전트가 있다면 가상 공유 정보기반 의사결정에 참여하도록 함.

<16> 4단계 미디에이터는 예측모델(의사결정나무)이 없는 에이전트에게 예측모델을 전송함

<17> 5단계 예측모델(의사결정나무)의 첫 번째(그 후 다음 번째) 속성을 포함한 에이전트는(선택 에이전트) 그 속성에 해당되는 데이터 값을 기반으로 의사결정나무의 구조를 따라 의사결정 과정을 진행하고 그 외의 참여 에이전트들은 수행된 에이전트의 수행 결과를 기다림.

<18> 6단계 선택 에이전트는 의사결정 과정에서 생성된 의사결정나무 경로 정보를 미디에이터로 보냄

<19> 7단계 6단계에서 받은 의사결정나무 경로 정보를 선택 에이전트 외의 모든 에이전트에게 전송함.

<20> 8단계 전달된 의사결정나무 경로정보를 통해 의사결정을 수행을 진행하여 선택 에이전트와 동일한 상

태로 됨.

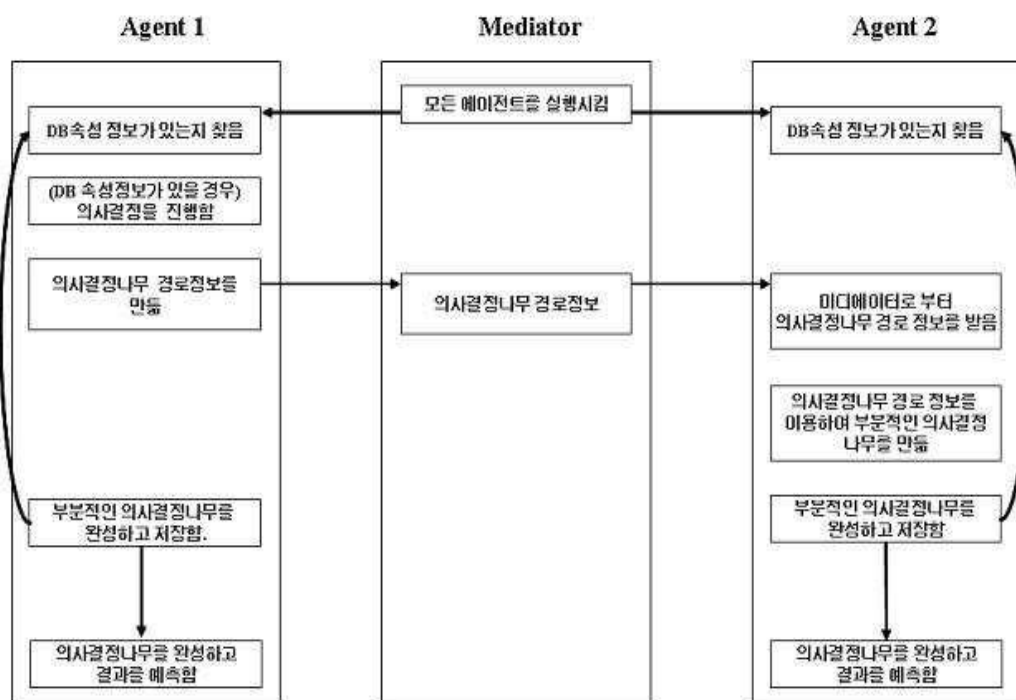
- <21> 9단계 에이전트는 의사결정 과정이 추가로 남아 있는지를 판단하고 단계 3부터 의사결정 과정이 추가로 없을 때까지 반복함.
- <22> 10단계 모든 에이전트가 미디어이터로 종료 신호를 보내고 시스템을 종료함.
- <23> 도 2는 분산되어 있는 데이터를 나타낸 표와 데이터마이닝을 통해 완성된 정보예측 모델이다. 정보예측 모델 검색에 필요한 정보는 일반적으로 분산되어 있는 모습을 나타낸다.
- <24> 도 3은 필요한 데이터를 가상으로 공유하여 정보를 예측하는 과정이다. 총 2개의 사이트라고 가정 했을 때, 사이트A에는 ssn, age, sex, degree DB속성 정보가 있고, 사이트B에는 ssn, job, salary DB속성 정보가 있다.
- <25> 도 2에 의하면 첫 번째 필요한 age DB속성정보는 사이트 A에 존재 하므로 A사이트에 있는 정보를 기준으로 분기를 한다(step1). 이와 같이 두 번째 필요한 DB속성 sex DB 속성 정보도 A사이트에 존재하므로 추가로 분기를 한다(Step2). 세 번째 필요한 job 필드는 B사이트에 존재하므로 지금까지 의사결정나무 경로 정보를 B사이트로 보낸다. 총 분기 횟수를 헤더에 넣고, 그 다음에는 의사결정나무 경로 정보를 보낸다. 이 때 왼쪽으로 분기한 정보는 0, 오른쪽으로 분기한 정보는 1로 표현한다.(21010에서 2는 총 분기 횟수, 0은 분기 정보를 나타냄). B사이트에서는 받은 의사결정나무 경로 정보를 기반으로 부분적인 의사결정을 진행하고 job DB속성을 이용하여 그 다음 분기를 한다. job DB속성을 통해서 최종 의사결정 진행을 완료하고 Class 3을 얻는다. 그리고 의사결정나무 경로 정보를 다시 A사이트로 보내어 A사이트 역시 최종 의사결정 진행을 완료하고 Class 3을 얻는다.

도면의 간단한 설명

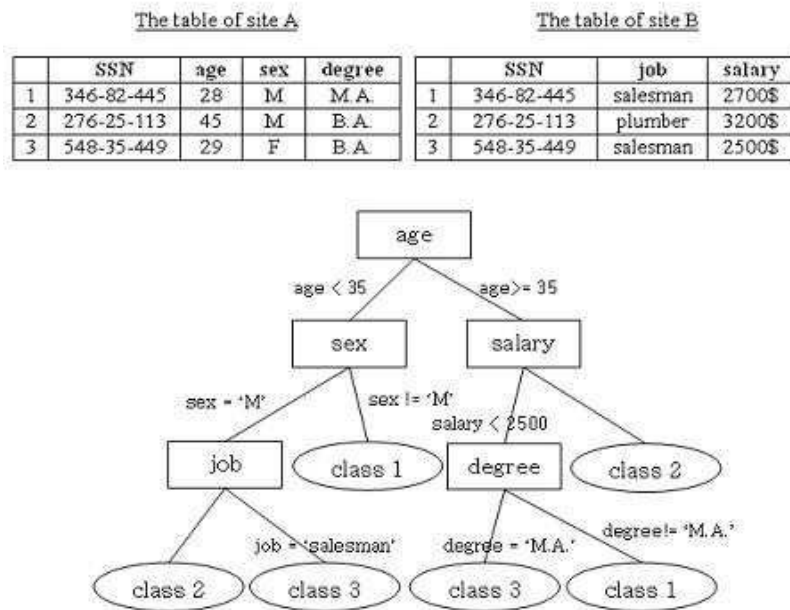
- <26> 도 1 은 분산환경 기반 정보예측 시스템 알고리즘의 흐름도,
- <27> 도 2 은 분산환경에서 불완전한 정보와 정보예측모델 구성도,
- <28> 도 3 는 분산환경에서 의사결정나무 형태의 정보예측모델로부터 정보를 예측하는 알고리즘의 예시도, 그리고
- <29> 도 4 는 기존의 중앙집중식 정보예측과 분산기반 정보예측 시스템의 비교도 이다.

도면

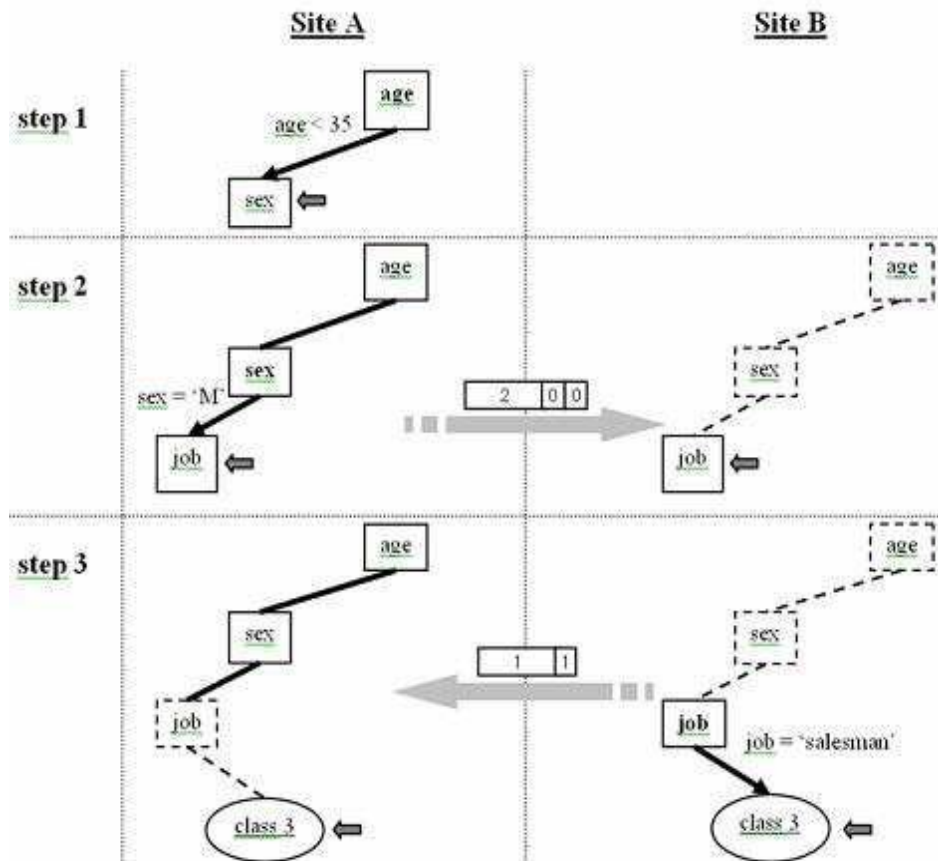
도면1



도면2



도면3



도면4

