



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2011년01월07일  
(11) 등록번호 10-1006116  
(24) 등록일자 2010년12월29일

(51) Int. Cl.

G06F 15/16 (2006.01) G06Q 50/00 (2006.01)

- (21) 출원번호 10-2008-0105820
- (22) 출원일자 2008년10월28일  
심사청구일자 2008년10월28일
- (65) 공개번호 10-2009-0122104
- (43) 공개일자 2009년11월26일
- (30) 우선권주장  
12/125,219 2008년05월22일 미국(US)
- (56) 선행기술조사문헌  
'A time-and-value centric provenance model and architecture for medical event streams', Proc.of ACM (2007.06.11.)\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
한국산업기술평가관리원  
서울 강남구 역삼동 701-7  
인터내셔널 비지네스 머신즈 코포레이션  
미국 10504 뉴욕주 아몬크 뉴오차드 로드
- (72) 발명자  
블라운트 마리온 리  
미국 뉴욕주 퍼트남 카운티 마호팍  
데이비스 존 시드니 2세  
미국 버지니아주 아링톤 카운티 아링톤  
(뒷면에 계속)
- (74) 대리인  
신정건, 김태홍

전체 청구항 수 : 총 10 항

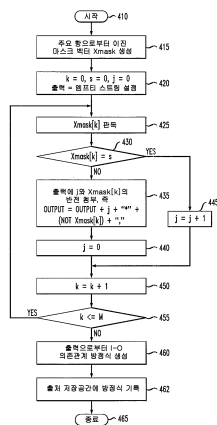
심사관 : 이종익

(54) 데이터 스트림 처리 시스템에서 출처 데이터를 결정하고 확인하는 방법 및 장치

(57) 요약

데이터 스트림 처리 시스템에서 출처 데이터를 결정하고 확인하기 위한 기술이 개시된다. 예를 들어, 복수의 처리 요소를 포함하는 데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법은 다음 단계들을 포함한다. 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소가 획득된다. 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 처리 요소에 대하여 하나 이상의 간격이 계산되며, 상기 간격 중 소정의 간격에 대하여, 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정된다. 다른 방법에서는, 간격이 지정된 다음, 관찰에 기초하여 계산된 간격에 대하여 그 지정된 간격을 비교함으로써 확인된다.

대표도 - 도4



(72) 발명자

**에블링 마리아 렌**

미국 뉴욕주 웨스체스터 카운티 화이트 플레인

**미스라 아찬**

미국 뉴저지주 소머셋 카운티 브릿지워터

**소우 데이비 모우셰**

미국 뉴욕주 웨스체스터 카운티 크로톤 온 허드슨

**왕 민**

미국 뉴욕주 웨스체스터 카운티 코틀란트 매너

---

## 특허청구의 범위

### 청구항 1

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격(interval)을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 포함하고,

상기 하나 이상의 간격은, 새로 관찰된 입력 요소와 대응하는 출력 요소 세트를 사용하여 상기 사용되는 관찰 데이터를 연속적으로 증가시키고, 추가의 관찰 데이터를 이용하도록 상기 간격의 계산을 연속적으로 정제(refine)함으로써 온라인으로 추정되는 것인, 데이터 처리 방법.

### 청구항 2

청구항 1에 있어서,

상기 하나 이상의 간격은 이전에 관찰 및 저장된 입력 요소와 대응하는 출력 요소 세트를 사용하여 오프라인으로 추정되는 것인 데이터 처리 방법.

### 청구항 3

삭제

### 청구항 4

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격(interval)을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 포함하고,

상기 하나 이상의 간격은 오프라인 및 온라인 기술의 조합을 사용하여 추정되는 것인 데이터 처리 방법.

### 청구항 5

청구항 5은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 관찰 데이터는 상기 데이터 스트림 처리 시스템의 정상 동작의 일부로서 외부 프로세스 또는 이벤트에 의해 발생된 실제 입력 및 출력 요소를 포함하는 것인 데이터 처리 방법.

### 청구항 6

청구항 6은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 관찰 데이터는 상기 간격 계산 단계의 일부로서 하나 이상의 프로브 목적으로 발생하는 입력 요소와 그것들의 대응하는 출력 요소를 포함하는 것인 데이터 처리 방법.

**청구항 7**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격(interval)을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 포함하고,

상기 하나 이상의 간격은 상기 데이터 스트림 처리 시스템에 의해 발생된 입력 요소와 그것들의 대응하는 출력 요소 세트에 대해 하나 이상의 통계적 분석 방법을 사용하여 계산되는 것인 데이터 처리 방법.

**청구항 8**

청구항 8은(는) 설정등록료 납부시 포기되었습니다.

청구항 7에 있어서,

상기 통계적 분석 방법 중 하나는 상기 입력 및 출력 요소의 경험적 결합 분포(empirical joint distribution)에 대한 엔트로피 기반의 계산을 사용하는 것인 데이터 처리 방법.

**청구항 9**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격(interval)을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 포함하고,

상기 하나 이상의 간격은 상기 데이터 스트림 처리 시스템에 의해 발생된 입력 요소와 그것들의 대응하는 출력 요소 세트에 대해 하나 이상의 선형 변환을 사용하여 계산되는 것인 데이터 처리 방법.

**청구항 10**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격(interval)을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 포함하고,

상기 간격 계산 단계는 확립된 간격의 길이를 결정하는 지정된 신뢰(confidence) 파라미터를 사용하는 단계를 더 포함하는 것인 데이터 처리 방법.

**청구항 11**

청구항 11은(는) 설정등록료 납부시 포기되었습니다.

청구항 10에 있어서,

상기 신뢰 파라미터는 질의의 일부로서 지정되는 것인 데이터 처리 방법.

**청구항 12**

청구항 12은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 하나 이상의 계산된 간격은 상기 대응하는 출력 요소의 발생에 영향을 미치는 상기 입력 요소에 대해 계산된 하나 이상의 시간 간격을 포함하는 것인 데이터 처리 방법.

**청구항 13**

청구항 13은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 하나 이상의 계산된 간격은 상기 대응하는 출력 요소의 발생에 영향을 미치는 상기 입력 요소에 대해 계산된 하나 이상의 시퀀스 간격을 포함하는 것인 데이터 처리 방법.

**청구항 14**

청구항 14은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 입력 요소와 출력 요소 사이의 연관관계의 관찰을 사용하는 단계는 지정된 과거 히스토리 내에서 관찰된 모든 입력 요소 및 출력 요소를 사용하는 단계를 더 포함하는 것인 데이터 처리 방법.

**청구항 15**

청구항 15은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 입력 요소와 출력 요소 사이의 연관관계의 관찰을 사용하는 단계는 각각의 출력 요소에 대하여 대응하는 입력 요소의 서브세트만 사용하는 단계를 더 포함하고, 상기 서브세트는 외부적으로 지정된 필터링 기준을 사용하여 상기 입력 요소의 총 세트의 필터링을 통하여 획득되는 것인 데이터 처리 방법.

**청구항 16**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법으로서, 상기 시스템은 복수의 처리 요소를 포함하고,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계;

상기 처리 요소에 대하여 하나 이상의 간격을 지정하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 보는 것인, 간격 지정 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격을 계산하고 상기 하나 이상의 지정된 간격과 상기 하나 이상의 계산된 간격을 비교함으로써, 상기 하나 이상의 지정된 간격을 확인(validate)하는 단계를 포함하고,

상기 하나 이상의 간격은, 새로 관찰된 입력 요소와 대응하는 출력 요소 세트를 사용하여 상기 사용되는 관찰 데이터를 연속적으로 증가시키고, 추가의 관찰 데이터를 이용하도록 상기 간격의 계산을 연속적으로 정제(refine)함으로써 온라인으로 추정되는 것인, 데이터 처리 방법.

**청구항 17**

청구항 16에 있어서,

상기 하나 이상의 지정된 간격과 상기 하나 이상의 계산된 간격의 비교 사이의 차이가 지정된 임계치를 초과하는 경우 경고 통지를 발생시키는 단계를 더 포함하는 데이터 처리 방법.

**청구항 18**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 장치로서, 상기 시스템은 복수의 처리 요소를 포함하고, 상기 장치는,

메모리; 및

상기 메모리에 연결된 프로세서를 포함하고,

상기 프로세서는 상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하고 상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격을 계산하도록 구성되고, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되며,

상기 하나 이상의 간격은, 새로 관찰된 입력 요소와 대응하는 출력 요소 세트를 사용하여 상기 사용되는 관찰 데이터를 연속적으로 증가시키고, 추가의 관찰 데이터를 이용하도록 상기 간격의 계산을 연속적으로 정제(refine)함으로써 온라인으로 추정되는 것인, 데이터 처리 장치.

**청구항 19**

청구항 19은(는) 설정등록료 납부시 포기되었습니다.

청구항 1에 있어서,

상기 하나 이상의 계산된 간격은,

대응하는 출력 요소의 발생에 영향을 미치는 입력 요소에 대해 계산된 하나 이상의 시간 간격; 및

대응하는 출력 요소의 발생에 영향을 미치는 입력 요소에 대해 계산된 하나 이상의 시퀀스 간격

중 적어도 하나를 포함하는 것인, 데이터 처리 방법.

**청구항 20**

데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하는 방법을 구현하는 프로그램을 갖는 컴퓨터로 판독가능한 기록 매체에 있어서,

상기 프로그램은 컴퓨터에 의해 실행될 때,

상기 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소를 획득하는 단계; 및

상기 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 상기 처리 요소에 대하여 하나 이상의 간격을 계산하는 단계로서, 상기 간격 중 소정의 간격에 대하여, 상기 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정되는 것인, 간격 계산 단계를 수행하며,

상기 하나 이상의 간격은, 새로 관찰된 입력 요소와 대응하는 출력 요소 세트를 사용하여 상기 사용되는 관찰 데이터를 연속적으로 증가시키고, 추가의 관찰 데이터를 이용하도록 상기 간격의 계산을 연속적으로 정제(refine)함으로써 온라인으로 추정되는 것인, 컴퓨터로 판독가능한 기록 매체.

**명세서**

**발명의 상세한 설명**

**기술분야**

[0001] 본 출원은 이와 함께 동시에 제출되어 그 명칭이 "데이터 스트림 처리 시스템에서 출처 데이터를 유지 및 처리하는 방법 및 장치(METHOD AND APPARATUS FOR MAINTAINING AND PROCESSING PROVENANCE DATA IN DATA STREAM PROCESSING SYSTEM)"인 대리인 문서 번호 YOR920070680US1로서 식별되는 미국 특허 출원과 관련되며, 이의 개시는 여기에 포함된다.

[0002] 본 발명은 데이터 스트림 처리 시스템에 관한 것으로, 보다 상세하게는 이러한 데이터 스트림 처리 시스템에서 출처 데이터(provenance data)를 결정하고 확인(validate)하기 위한 기술에 관한 것이다.

**배경 기술**

[0003] 우리 주변의 수많은 데이터 소스는 특정 애플리케이션에 대하여 상당한 양의 중요한 정보를 포함하는 고용량 스트림을 생성한다. 예시적인 애플리케이션으로는 잠재적인 보안 침해를 검출하기 위해 많은 비디오 공급을 수집(ingest)하는 비디오 감시 애플리케이션이 있다. 다른 예로는 데이터를 분석하여 의학적으로 중요한 이벤트를 식별하고 의료 전문가에게 보고하는 스트림 처리 기반구조로 스트림 데이터를 방출하는 센서로 환자가 둘러싸이게 되는 연속적 건강 모니터링이 있다.

[0004] 이들 애플리케이션의 대부분에 있어서, 시스템에 의해 발생된 매번의 이벤트의 출처를 추적하는 것은 중요하다. 출처에 의해, 시스템에 의한 이벤트의 발생에 대한 기점(origin) 및 정당화(justification)를 의미하는 것이다. 예를 들어, 의료 시스템이 자신의 분석에 기초하여 환자의 약 조제 변경이 필요하다고 제안하는 경우, 이러한 이벤트의 출처는 의료 전문가에게 그 경고의 발생에 사용된 모든 데이터 포인트 및 절차를 알릴 것이다.

[0005] 통상적으로, 이러한 출처 보고는 그들 분석의 설계 중에 개발자에 의해 지정된 데이터를 활용(leveraging)함으로써 수동으로 획득된다.

**발명의 내용**

**해결 하고자하는 과제**

[0006] 본 발명의 원리는 이러한 데이터 스트림 처리 시스템에서 출처 데이터를 결정하고 확인하기 위한 기술을 제공한다.

**과제 해결수단**

[0007] 예를 들어, 본 발명의 하나의 실시예에서, 데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하기 위한 방법을 제공하는데, 시스템은 복수의 처리 요소를 포함하며, 본 방법은 다음 단계들을 포함한다. 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소가 획득된다. 처리 요소의 입력 요소와 출력 요소 사이의 연관관계(association)의 관찰을 나타내는 데이터를 사용하여 처리 요소에 대하여 하나 이상의 간격(interval)이 계산되며, 상기 간격 중 소정의 간격에 대하여, 그 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 결정된다.

[0008] 다른 실시예에서, 데이터 스트림 처리 시스템에 의해 수신된 데이터 스트림과 연관된 데이터를 처리하기 위한 방법을 제공하는데, 시스템은 복수의 처리 요소를 포함하며, 본 방법은 다음 단계들을 포함한다. 복수의 처리 요소 중 적어도 하나의 처리 요소와 연관된 입력 데이터 요소 및 출력 데이터 요소가 획득된다. 처리 요소에 대하여 하나 이상의 간격이 지정되며, 상기 간격 중 소정의 간격에 대하여, 그 소정의 간격 내에 포함된 하나 이상의 특정 입력 요소는 특정 출력 요소에 기여한 것으로 본다. 상기 하나 이상의 지정된 간격은, 처리 요소의 입력 요소와 출력 요소 사이의 연관관계의 관찰을 나타내는 데이터를 사용하여 처리 요소에 대하여 하나 이상의 간격을 계산하고, 하나 이상의 지정된 간격과 하나 이상의 계산된 간격을 비교함으로써, 확인된다.

**효과**

[0009] 유리하게는, 이러한 본 발명의 기술은 데이터 스트림 처리 시스템에서 하나 이상의 처리 요소에 대하여 입력 데이터 요소와 출력 데이터 요소 사이의 의존관계(dependency)를 결정할 수 있다. 즉, 소정의 출력 스트림 요소의 발생에 어느 입력 스트림 요소가 사용되었는지 시스템이 식별할 수 있게 해주는 의존관계 방식식이 결정될 수 있다.

**발명의 실시를 위한 구체적인 내용**

[0010] 본 발명의 이들 및 기타 목적, 특징 및 이점이 첨부 도면과 관련하여 보게 될 본 발명의 예시적인 실시예의 다음의 상세한 설명으로부터 명백하게 될 것이다.

[0011] 예시적인 데이터 스트림 처리 시스템과 관련하여 본 발명의 실시예가 아래에 설명될 것이다. 그러나, 본 발명이 임의의 특정 데이터 처리 시스템에 한정되는 것은 아님을 이해하여야 한다. 그보다 본 발명은 출처 데이터를 결정하고 확인하는 것이 바람직할 임의의 데이터 처리 시스템에 보다 일반적으로 적용 가능하다. 예시적으로 여기에 사용될 때, 용어 "출처"는 "기점" 또는 "소스(source)"로서 정의된다. 따라서, 예를 들어, 출처 데이터(또는 출처 정보)는, 예를 들어 복합 다층 분석을 통하여, 일반적으로 그 데이터가 다른 데이터로부터 유도된 곳인 기점 또는 소스 데이터를 문서화하는 메타 정보이다. 또한, 여기에 사용되는 바와 같은 용어 "경험적(empirical)"은 "경험 또는 실험에 의해 증명 가능하거나 검증 가능한" 것으로서 정의된다. 계산에 전체 입력이 미리 주어지는 경우 그 계산을 "오프라인(offline)"인 것으로 간주한다(예를 들어, <http://www.nist.gov/dads/HTML/offline.html> 참조). 대조적으로, 온라인(on-line) 계산은 계산에 있어서 새로 수신된 입력을 점차적으로 통합하는 것이다.

[0012] 본 발명의 원리의 하나의 예시적인 실제계에서의 실제 적용으로는 건강 관리(healthcare) 영역이 있다. 구체적으로, 건강 관리 영역에서, 입력으로서 EEG(electroencephalogram)나 ECG(electrocardiogram) 데이터의 스트림 뿐만 아니라, 혈압, 몸무게, 또는 혈당측정기 기록과 같은 의료 정보의 스트림을 취하는 스트림 처리 기반구조를 생각해볼 수 있다. 스트림 처리 기반구조는 이들 데이터 스트림에 분석적 알고리즘을 적용하고, 의사나 다른 의료 전문가로부터 주의를 더 끌어야 하는 중요한 의료 이벤트를 식별할 것이다. 이러한 시스템에서는, 스트림의 어느 서브셋이 그 중요한 의료 이벤트를 트리거했는지 그리고 어느 특정 처리 요소가 그 발생에 기여하였는지와 같은, 여기에서는 출처 데이터로서 칭하는 메타 정보를 포함하여, 중요한 의료 이벤트의 발생에 관한 가능한 많은 정보를 의료 전문가에게 주는 것이 중요하다. 스트림 처리 시스템에 의해 발생된 경고를 수신하는 의료 전문가는 그 경고와 관련된 출처 질의를 발행할 수 있으며, 그 출처 질의는 경고의 발생에 기여한 스트림 데이터 항목 및 스트림 처리 노드를 보여주거나 나타내라는 요청이다. 다시, 본 실시예에서 언급된 해결 영역은 건강 관리이지만, 본 발명의 원리는 데이터 스트림의 처리가 적용될 수 있는 임의의 영역에 적용된다는 것을 알아야 한다.

[0013] 구체적으로, 설명될 바와 같이, 본 발명의 예시적인 원리는 스트림 출처 시스템에서 의존관계 방정식의 자동 생성을 제공한다. 또한, 본 시스템에 의해 자동으로 생성되는 유도된 의존관계 방정식에 대하여, 개발자에 의해 지정된 의존관계 방정식을 확인하는 기술을 제시한다. 예시적인 실시예에서, 스트림 처리 시스템에 추가되는 출처 의존관계 유도기(PDD; Provenance Dependency Deriver) 컴포넌트로서, 임의의 소정의 스트림 처리 컴포넌트에 대하여 입력 데이터 요소의 시퀀스(하나 이상의 이산(distinct) 스트림에 속함) 및 출력 데이터 요소의 대응하는 시퀀스를 관찰하는 출처 의존관계 유도기(PDD) 컴포넌트를 제안한다. PDD의 작업은, 블랙 박스로서 처리 컴포넌트를 처리하고, 관찰된 데이터에 기초하여, 대응하는 출력 데이터 요소의 발생에 경험적으로 영향을 미칠 것 같은 입력 데이터의 시간 윈도우(time-window)를 단순히 학습하는 것이다. 따라서, PDD 컴포넌트의 주요 목표는 시불변(time-invariant) 시간적(temporal) 의존관계 함수를 확립하는 것이다. 일반적으로, PDD 컴포넌트는 다음 공식 관계를 확립한다:

[0014] 
$$e_i(t) \leftarrow \bigcup_{j: S_j \text{ is input to } PE_i} \bigcup_{k=1}^{L_j} \{ e \in S_j(t - start_{jk}, t - end_{jk}) \} \tag{1}$$

[0015] 여기에서,  $L_j$ 는 그 값에 따라  $e_i(t)$ 가 좌우되는  $S_j$ 의 값을 정의하는 이산 비결합(disjoint) "시간 간격"의 수이고,  $start_{jk}$ 와  $end_{jk}$ 는 이들 간격의 경계를 정의한다. 용어 "시불변"은 항  $start_{jk}$ 와  $end_{jk}$  그 자체가  $t$ 에 독립적이라는 사실을 의미하는 것이며, 그 결과, 다른 입력 데이터 요소에 대한 유도된 이벤트의 의존성은 각각의 샘플의 특정 타임스탬프나 ID(식별자)에 완전히 독립적으로 간결하게 표현될 수 있다. 여기에서의 주요 목표는 항  $start_{jk}$ 와  $end_{jk}$ 의 값을 추정하려고 하는 것이다. 많은 실제 경우에, 시간적 의존관계 모델은 (각각의 입력 스트림  $S_j$ 에 대하여) 단일 (스트림 당) 간격 항  $\Delta_j$ 의 지정(specification)으로 단순화될 수 있으며, 그리하여 의존관계는 전적으로 입력 이벤트의 가장 최근 윈도우에 대한 것이 되는 것이다:

[0016] 
$$e_i(t) \leftarrow \bigcup_{S_j \text{ is input to } PE_i} \{ e \in S_j(t, t - \Delta_j) \} \tag{2}$$



- [0017] 여기에서,  $\Delta_j$ 는 입력 스트림  $S_j$ 와 연관된 과거 "시간-윈도우"를 나타낸다.
- [0018] 예로써, 이들 항의 값을 확인하기 위하여 여러 다양한 기술 집합 중 하나를 사용할 수 있다. 첫 번째 예시적인 접근법은 주요 컴포넌트 분석을 사용한 입력 데이터의 '선형'(즉, 아핀(affine)) 변환 또는 카루넨 루베(Karhunen-Loeve) 변환(Digital Pictures: Representation, Compression and Standards (Applications of Communications Theory), Arun N. Netravali, Barry G. Haskell, Plenum Press, 1995 참조, 이의 개시는 여기에 포함됨)을 수행하는 것으로 알려진 것으로서 처리 컴포넌트에 적용가능하다. 다른 예시적인 접근법으로는 선형 시스템을 이용하여 본 컴포넌트를 모델링하고 임펄스 응답 또는 주파수 응답 테스트를 수행하여 전달 함수(transfer function)를 산출하는 것이 있다. 이 전달 함수로부터, 입력-출력 의존관계를 얻을 수 있다.
- [0019] 또 다른 예시적인 접근법과 아래에 더 설명되는 접근법은 정보 이론적 구성을 사용하여 입력 출력 상관관계(correlations)를 측정하고 입력 출력 의존관계를 추정하는 것이다. 이 접근법은 컴포넌트에 대해 아무런 가정도 하지 않으며 선형 및 비선형 변환의 일반적인 클래스에 적용한다. 그것은 엔트로피 및 상호 정보와 같은 정보 이론적 개념의 계산 및 경험적 결합 확률(joint probability) 분포 함수의 사용을 수반한다. 예를 들어, 출력  $Y$ 와 점점 길어지는 입력 시퀀스  $X(\cdot)$  사이의 조건부 엔트로피(conditional entropy)의 시퀀스를 계산함으로써 시작할 수 있다. 바꾸어 말하면,  $H(Y(t)); H(Y(t)|X(t)), H(Y(t)|X(t), X(t-1)), H(Y(t)|X(t), X(t-1), X(t-2))\dots$  등을 계산하며,  $H(Y(t))$ 는 랜덤 변수  $Y(t)$ 의 샤논(Shannon) 엔트로피(C. E. Shannon, "A mathematical theory of communication", Bell System Technical Journal, vol. 27, pp. 379-423 및 623-656, July and October, 1948 참조, 이의 개시는 여기에 포함됨)를 나타낸다. 일반적으로, 일부 의존성이 존재하는 한, 추가의 조건화(conditioning)는 조건부 엔트로피를 감소시키게 될 것이다. 따라서, 엔트로피 자체가 수용 가능한 최대 값 아래로 떨어지거나  $H(Y|X, X-1, X-(L+1))$ 가  $H(Y|X, X(t-1), X(t-L))$ 에 비교하여 엔트로피의 감소를 거의 또는 아예 초래하지 않도록 값  $L$ 에 도달할 때까지 조건부 엔트로피(경험적 분포에 기초함)를 계속해서 계산할 수 있다. 그 다음, 의존성이 관찰된 적합한 시간 간격인 것으로  $L$ 을 선언할 수 있다.
- [0020] 이러한 기본 설명에 부가하여, 본 발명의 수많은 다양한 변형예를 나타내는 많은 다른 상세 사항이 존재한다. 본 발명의 하나의 주요 양상은 외부적으로(externally) 지정된 의존관계의 확인(validation)을 수행하는데 PDD 기반의 의존관계 함수를 사용하는 것이다. 구체적으로, 시스템은 경험적으로 유도된 의존관계 값을 스트림 처리 컴포넌트에 대하여 외부적으로(예를 들어, 수동으로) 지정된 값과 비교하여 차이(divergence) 범위를 주목할 수 있다. 차이 정도가 높은 경우, 이는 틀리게 지정된 의존관계 함수이거나 스트림 처리 컴포넌트에서의 실제 처리 로직의 일부 잘못된 작용을 나타내는 것일 수 있다. 많은 상황에서, 잠재적인 장애 또는 이상(anomaly)의 이러한 자동 식별은 확인 또는 조기 장애 검출을 위한 도구로서 매우 중요한 것임을 판명할 수 있다.
- [0021] 상기의 기본 설명은 관찰된 입력 및 출력 값으로부터 시간적 의존관계를 추정하는 것을 언급하지만, 본 발명의 원리는 또한 처리 컴포넌트에 PDD 자체에 의해 발생된 특정 입력 샘플이 공급될 수 있는 상황(시물레이션 또는 테스트 환경에서)을 수용할 수도 있다. 이 경우에, 유도된 의존관계 함수는 종종 보다 정확할 수 있는데, 일련의 또는 일 세트의 입력 요소들이 종종 보다 넓은 샘플 공간을 나타내도록 보다 지능적으로(intelligently) 선택될 수 있기 때문이다. 예를 들어, 의료 스트림 분석 환경에서, 특정 입력 값은 드물게만 또는 심각한 이상(예를 들어, 40 이하의 심박동수 기록) 하에서만 발생할 수 있으며, 이와 같이, 정상 동작 중에는 스트림 처리 시스템이 이러한 입력 상태를 관찰하지 않을 것이다. 그러나, 시물레이션된 값의 사용은 PDD가 훨씬 더 넓은 범위의 동작 파라미터에 걸쳐서 입력-출력 의존관계를 관찰할 수 있게 해준다.
- [0022] 지금까지 본 발명은 의존관계의 기본적인 시간적 모델만 기술하였지만, 본 발명의 기술은 또한, 물론 의존관계의 '모델'이 외부적으로 제공되는 것을 가정하여, 일부 대안의 의존관계 모델을 학습하거나 추정하는데 적용될 수도 있다. 예를 들어, 의존관계가 시간 윈도우에 대한 것이 아니라 "과거 요소의 수"에 대한 것으로 선언되는 경우, 본 발명의 접근법이 여기에도 적용될 수 있다. 간단하게, 적합한 열에 입력 요소  $X$ 의 적합한 시퀀스를 사용하여 도 2에 기재된 바와 같은 경험적 매트릭스를 형성하고, 전과 같은 기술을 적용하여 '시퀀스 간격'  $L$ (시간 간격 대신에)을 결정해야 한다.
- [0023] 상기 설명된 방법은 오프라인(저장된 입력 및 출력 스트림 세트)로)이나 온라인으로 적용될 수 있다. 오프라인 경우에, 본 발명의 원리는 저장된 입력 및 출력 스트림 세트에(적합하게 정의된 히스토리에 걸쳐) 액세스하고 이 세트를 "학습 시퀀스"로서 사용하여 의존관계를 유도하도록 PDD를 제공한다. 온라인 모델에서는, PDD는 인입하는 입력 요소 및 발생된 출력 요소를 연속적으로 관찰한 다음, 점점 더 큰 경험적 매트릭스를 구성함으로써, 경험적 매트릭스가 진화함에 따라 보다 정제(refine)된 의존관계 값의 산출치를 연속적으로 유도할 수 있다.

- [0024] 지금까지 설명된 방법은 블라인드형(blind)으로서, 의존관계 함수(예를 들어, 조건부 엔트로피)가 안정적으로 (지정된 임계치 이하) 남아 있는 것으로 관찰되거나 결과적인 엔트로피가 충분히 작을 때까지 시간 윈도우를 잠재적으로 증가시키며 입력 및 출력 요소의 모든 샘플을 사용하여 경험적 의존관계를 확인하고자 하는 것이다. 그러나, 본 발명의 원리는 또한 경험적 유도 프로세스에 사용되는 선택된 입력 요소 세트를 정제하는 추가의 이전(prior) 지정과 결합될 수도 있다. 예를 들어, 처리 컴포넌트는 일부 외부 지정(예를 들어, 월요일 아침마다 취해지는 심박동수 기록 또는 체육관에 있는 동안에 취해지는 혈압 기록만 이용하는 것을 표시함)을 가질 수 있으며, PDD는 이러한 외부 지정된 필터를 적용하여, 먼저 출력 요소와 잠재적으로 연관될 수 있는 입력 데이터 요소의 벡터 공간을 감소시킬 수 있다. 적합한 최소한의 입력 요소 세트를 결정하는 후속 기술은 그대로이다.
- [0025] 설명된 방법은 또한 시간적 또는 시퀀스 기반의 출처 의존성의 등가 클래스를 구성하는데 적용될 수도 있는데, 개념적으로 등가 클래스는 출력 값 세트(출력 값들 범위의 서브세트)의 모든 구성요소가 그들 대응하는 입력 요소에 대하여 동일하거나 유사한 의존성 함수를 갖게 되는 출력 값 세트로 정의될 수 있다. 이러한 상황은, 예를 들어 처리 컴포넌트로부터의 출력이 출력 값이 40보다 크면 과거 100개의 값에 따라 좌우되지만 출력이 40보다 작으면 과거의 1000개의 샘플에 따라 좌우되는 경우에, 일어난다. 의존관계의 윈도우의 이러한 분할(segmentation)은, 블라인드형으로(출력 범위의 임의의 이전 분할 없이) 또는 출력 세그먼트에 대한 이전 지정의 도움으로(이 경우, 다수의 세그먼트가 하나의 등가 클래스로 결합될 수 있는지의 여부를 판정하는데 본 발명의 원리가 여전히 사용될 수 있음), 행해질 수 있다.
- [0026] 이러한 분할은 출처 의존관계를 더 정제하는 것을 도우며, 모든 출력 값에 대하여 넓은 의존성 윈도우를 정의하는 대신에, 상기 설명된 방법은 출력 값의 특정 범위에 대하여 보다 제한되고 정확한 의존성 윈도우를 잠재적으로 연관시킬 수 있다. 더욱이, 가능한 많은 요소를 하나의 등가 클래스로 형성하는 것은 또한 유도된 출처 의존관계 함수를 저장하는데 필요한 저장 공간을 감소시킨다. 구체적으로, 이 접근법의 간단한 실시예는 먼저 출력 공간을 적합한 구획 세트로 분할하고, 경험적으로 관찰된 출력 요소를 그들 대응하는 구획으로 할당하는 것일 것이다. 그 다음, 상기 설명된 통계적 분석 기술이 각각의 데이터 구획에 대해 적용되어 그것의 시간적 또는 시퀀스 기반의 의존성을 유도할 수 있다. 다수 구획에 대하여 유도된 의존성이 동일한 것(또는, 시간 윈도우의 길이  $\Delta$ 가 서로의  $\pm 2$  내에 있는 것과 같은, 지정된 허용오차 내에 있음)으로 드러나는 경우, 구획은 하나의 등가 구획으로 교체될 수 있으며, 그는 결합된 의존관계 함수에 대응하는 것이다.
- [0027] 이제 예시적인 도면들을 참조하여 다수 사용자 공동(collaborative) 소프트웨어 개발 환경을 지원하는 예시적인 시스템 및 방법이 보다 상세하게 설명될 것이며, 도면에서 동일한 참조 부호는 동일하거나 유사한 구성요소를 나타낸다. 본 발명에 따라 여기에 설명된 시스템 및 방법은 하드웨어, 소프트웨어, 펌웨어, 특수 용도 프로세서, 또는 이들의 조합의 다양한 형태로 구현될 수 있음을 이해하여야 한다. 구체적으로, 하나의 예시적인 실시예에서, 본 발명의 시스템 및 방법은 하나 이상의 프로그램 저장 디바이스(예를 들어, 하드 디스크, 자기 플로피 디스크, RAM, CD ROM, DVD, ROM 및 플래시 메모리) 상에 실제적으로 구현되고 적합한 아키텍처를 포함하는 임의의 디바이스 또는 머신에 의해 실행가능한 프로그램 명령을 포함하는 소프트웨어로 구현된다.
- [0028] 첨부 도면에 도시된 구성요소 시스템 모듈 및 방법 단계들은 소프트웨어로 구현될 수 있기 때문에, 시스템 컴포넌트들 사이의 실제 접속(또는 프로세스 단계들의 흐름)은 애플리케이션이 프로그래밍되는 방식에 따라 상이할 수 있음을 더 이해하여야 한다. 본 명세서에서의 교시에 따라, 관련 기술 분야에서의 통상의 지식을 가진 자라면 본 발명의 이들 및 유사한 구현에 또는 구성을 생각해볼 수 있을 것이다.
- [0029] 도 1은 분석 컴포넌트의 그래프를 도시한다. 그래프는 8개의 처리 요소(PE)(100, 110, 120, 130, 140, 150, 160, 및 170)를 갖는다. 처리 요소는 그것들을 접속시키는 방향 화살표로 표현된 통신 경로를 통하여 접속된다. 각각의 처리 요소가 데이터 스트림에 대한 임의의 계산을 수행할 수 있는 각자의 개별 알고리즘에 따라 스트림을 처리하기 때문에, 처리 요소는 그래프에서의 접속에 따른 다음 처리 요소에 결과적인 출력을 전달한다. 본 발명의 주제를, 각각의 처리 요소에서 어느 입력 스트림 요소가 소정의 출력 스트림 요소의 발생에 사용되었는지 시스템이 식별할 수 있게 해주는 의존관계 방정식을 학습하는 문제로서 예시적으로 설명할 수 있다. 도 1에 도시된 바와 같이, 포트(111, 121, 131, 132, 141, 142, 152, 153, 161, 162, 171, 및 172)를 통해 도달하는 데이터 스트림은 각각의 처리 요소에 입력되고 요소는 대응하는 출력 포트(101, 112, 113, 122, 123, 133, 134, 143, 151, 163, 164, 및 173)를 통해 각각의 처리 요소에 의해 출력된다. 도 1에서, 포트(111, 121 및 132)에 도달하는 데이터 요소는, 출력 포트(101)에 의해 발생된 데이터 요소에 대응하여 동일한 것으로 관찰될 수 있다.
- [0030] 도 6은 시스템에 대한 예시적인 아키텍처를 도시한다. 데이터 소스(605)는 이벤트를 발생시키고, 처리 요소의

분석 그래프(610)에서 디스플레이를 위해 이들 이벤트를 제공한다. 대안으로서, 이벤트는 프로브 발생기(Probe Generator)(615)에 의해 발생할 수 있으며, 프로브 발생기(615)는 구체적으로 의존관계 윈도우의 보다 정확하거나 보다 철저한 유도를 돕는 입력 이벤트를 발생시키는 것이다. 그래프에서 흐르는 이벤트의 모든 스트림의 지속은 저장 관리자(620)에 의해 데이터 저장공간(625)으로 이행된다. 입력-출력 의존관계 방정식을 통계적으로 추정하기 위하여, 경험적 확률 분포가 산출된다. 경험적 분포 산출기(630)는 이들 분포를 산출하도록 데이터 저장공간(620)에 액세스한다. 이들 분포는, 주요 컴포넌트(dominant component) 분석 모듈(635)에 의해, 출력의 발생을 주로 담당하는 컴포넌트의 입력 스트림에서의 위치를 식별하는데 사용된다. 주요 컴포넌트 분석 모듈(635) 내에서, 입력에서의 주요 컴포넌트를 식별하도록 조건부 엔트로피 및 상호 정보 산출치가 경험적 확률 분포로부터 계산되는 것이다. 이들 주요 컴포넌트는 소정의 처리 요소에 대하여 출력의 값을 주로 담당하는 입력의 부분이다.

[0031] 이러한 컴포넌트가 식별되면, I-0(입력-출력) 의존관계 방정식 생성기(640)는 I-0 의존관계 방정식을 생성하고, 그것들을 출처 저장공간(645)에 저장한다. 출처 의존관계 유도기(642)는 본질적으로 주요 컴포넌트 분석 모듈(635) 및 I-0 의존관계 방정식 생성기(640)로 구성된다. 이들 의존관계 방정식은 또한, 출처 모델 확인기(650)에 의해, 통합 개발 환경(655)에서 애플리케이션 개발자에 의해 이루어진 출처 I-0 의존관계 확인의 적합성(compliance)을 확인하도록 사용될 수도 있다.

[0032] 도 3은 모듈(635)의 주요 컴포넌트 분석을 수행하는데 사용되는 방법에 대한 바람직한 실시예를 도시한다. 이 분석은, 도 2에 도시된 포맷으로, 경험적 매트릭스에서의 데이터의 준비(312)로 시작(310)한다. 본질적으로, 스트림은 사이즈 M의 벡터(215, 220, 225, 230, 235)로 벡터화된다. 이들 벡터 각각에, 대응하는 입력 벡터에 대해 발생된 출력  $Y_i$ 에 대응하는 출력 요소(240, 245, 250, 255, 250)가 첨부된다. 데이터가 준비되면, 변수 세트는 주요 컴포넌트 분석을 위하여 초기화된다(315). 관심의 특징 세트 F가 설정된다. 그것의 원소수(cardinality)(처음에는 M)가 변수 K에 저장된다. 주요 특징(dominant feature)의 엠프티(empty) 세트  $F_s$ 가 초기화된다. 마지막으로, 출력 프로세스  $Y_i$ 의 샤논 엔트로피(C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948 참조, 이의 게시는 여기에 포함됨)가 계산되어 변수 pH에 할당된다. 다음 단계(320)에서, F에 속하는 각각의 특징  $X_{t-i}$ 에 대하여, 조건부 엔트로피  $H(Y_i|X_{t-i})$ 가 계산된다. 단계 325에서, 가장 낮은 조건부 엔트로피  $H(Y_i|X_{t-i}, F_s)$ 를 갖는  $X_{t-i}$ 가 F로부터 선택되고  $F_s$ 에 추가된다(330). 다음 단계(335)에서, pH와  $H(Y_i|F_s)$  사이의 차이를 취함으로써 조건부 상호 정보  $I_{cond}$ 가 계산된다.

[0033]  $I_{cond}$ 가 미리 정의된 임계치 mDelta보다 더 큰 경우(340), 현재 특징 세트  $F_s$ 가  $Y_i$ 의 예상에 기여하는 것으로 결론짓는다.  $F_s$ 에 대한  $Y_i$ 의 의존관계가 충분히 높은지 더 테스트하기 위하여,  $H(Y_i|F_s)$ 의 값을 테스트한다(345). 이 값이 임계치 hDelta보다 적은 경우(345), 분석의 출력으로서  $F_s$ 를 반환하고(350), 계산을 종료한다(370). 이 값이 hDelta보다 높은 경우, F로부터  $X_{t-i}$ 를 제거하고, F가 엠프티가 아니면(360), 단계 325에서 계산을 다시 시작한다.  $I_{cond}$ 가 mDelta보다 작은 경우(340),  $X_{t-i}$ 는 F로부터 제거되고(355), 계산은 단계 360에서 다시 시작된다.

[0034] 도 4는 I-0 의존관계 방정식 생성기(640) 내에서 일어나는 계산에 대한 바람직한 실시예의 흐름도를 도시한다. 이 실시예에서, 계산은 주요 컴포넌트 분석 모듈(635)에 의해 계산된 주요 항(dominant term)으로부터 이진 마스크 벡터 Xmask의 생성(415)으로 시작(410)한다. Xmask는 사이즈 M의 이진 벡터이며, 구성요소 Xmask[i]는 X의 위치 i에서의 항이 주요 항이면 1과 같다. 반대로, X의 위치 i에서의 항이 주요 항이 아니면 Xmask[i]는 0과 같다. 다음 단계(420)는 나머지 계산에 필요한 변수를 초기화한다. 이들 변수는 0으로 설정되는 카운터 "k"와 "j", 0으로 설정되는 이진 상태 변수 "s", 및 처음에는 엠프티인 스트링 "출력(output)"이다. 카운터 k는 이 모듈에 의해 처리되는 특징의 수를 추적한다. 결과적으로, 그것의 최대 값은 M이다. j는 Xmask에서 0과 1의 런 렱스(run length)를 추적하는데 사용된다.

[0035] 다음 단계는 Xmask[k]의 값을 관독한다(425). 그 다음, Xmask[k]의 값을 테스트하고, 그것을 s와 비교한다(430). Xmask[k]가 s와 동일하지 않은 경우, 출력 스트링에 "j \* (NOT Xmask[k])"을 첨부하고(435), j를 0으로, s를 Xmask[k]로 설정한다(440). (NOT Xmask[k])는 Xmask[k] 비트의 이진 반전(binary inversion)을 나타낸다. 그 다음, 계산은 단계 450을 실행한다. Xmask[k]가 s와 동일한 경우, j는 1만큼 증분되고(445), 그 다음 계산은 단계 450을 실행한다. 단계 450에서, k는 1만큼 증분된다. 다음 단계에서는, Xmask의 모든 구성요소

를 판독하였는지 체크한다(455). 이는 k가 M보다 작거나 같은지 체크함으로써 행해진다. 답이 '네'이면, 처리할 더 많은 구성요소를 가진 것이며, 계산은 단계 425로 다시 돌아간다. 답이 '아니오'이면, 다음 단계는 Xmask의 런 령스 인코딩인 출력을 사용하여, 그것을 출처 저장 공간에 저장(462)하기 전에, 그것을 I-O 의존관계 방정식으로 변환하는 것이다(460). 계산은 단계 465에서 종료한다.

[0036] 도 5는 출처 모델 확인기(도 6에서의 650) 내에서 일어나는 계산에 대하여 제안된 기술을 도시한다. 계산은 소정의 처리 요소(520)에 대하여 학습된 I-O 의존관계 방정식을 판독한 다음, 개발자 지정된 I-O 의존관계 방정식을 판독함으로써 시작(510)한다. 다음 단계에서는 시스템이 평가한 것과 개발자가 지정한 것 사이의 차이를 평가한다. 그 차이가 엡실론보다 더 큰 경우(550), 임의의 차이 메트릭에 따라, 부적합(non-compliance) 경고가 발행된다(570). 이 경고는 관리자에 의해 소비될 수 있다. 다른 경우에, 그 차이가 엡실론보다 작거나 같은 경우, 적합 메시지를 기록한다(550).

[0037] 마지막으로, 도 7은 본 발명의 기술의 하나 이상의 컴포넌트/단계에 따라 구현될 수 있는 컴퓨터 시스템을 도시한다. 개별 컴포넌트/단계가 하나의 이러한 컴퓨터 시스템 상에서 또는 하나보다 많은 이러한 컴퓨터 시스템 상에서 구현될 수 있음을 더 이해하여야 한다. 분산된 컴퓨팅 시스템 상에서의 구현의 경우, 개별 컴퓨터 시스템 및/또는 디바이스는 적합한 네트워크, 예를 들어 인터넷 또는 월드 와이드 웹을 통하여 접속될 수 있다. 그러나, 시스템은 사설 또는 로컬 네트워크를 통하여 실현될 수 있다. 어떠한 경우든, 본 발명은 임의의 특정 네트워크에 한정되지 않는다.

[0038] 따라서, 도 7에 도시된 컴퓨터 시스템은 여기에 설명된 기능의 전부 또는 일부를 제공할 수 있는 하나 이상의 서버 또는 하나 이상의 기타 처리 디바이스를 나타낼 수 있다. 대안으로서, 도 7은 메인프레임 컴퓨터 시스템을 나타낼 수 있다.

[0039] 컴퓨터 시스템은 일반적으로 프로세서(705), 메모리(710), 입력/출력(I/O) 디바이스(715), 및 컴퓨터 버스(725)나 대안의 접속 구성을 통하여 연결되는 네트워크 인터페이스(720)를 포함할 수 있다.

[0040] 여기에서 사용되는 바와 같은 용어 "프로세서"는, 예를 들어 CPU 및/또는 기타 처리 회로를 포함하는 것과 같은 임의의 처리 디바이스를 포함하도록 의도되는 것을 인식하여야 한다. 또한, 용어 "프로세서"는 하나보다 많은 처리 디바이스를 칭할 수 있고 처리 디바이스와 연관된 다양한 요소가 다른 처리 디바이스에 의해 공유될 수 있음을 이해하여야 한다.

[0041] 여기에서 사용되는 바와 같은 용어 "메모리"는, 예를 들어 RAM, ROM, 고정 메모리 디바이스(예를 들어, 하드 디스크 드라이브), 탈착가능한 메모리 디바이스(예를 들어, 디스켓), 플래시 메모리 등과 같이, 프로세서 또는 CPU와 연관된 메모리를 포함하도록 의도된다. 메모리는 컴퓨터 판독가능한 저장 매체로 간주될 수 있다.

[0042] 또한, 여기에 사용되는 바와 같은 어구 "입력/출력 디바이스" 또는 "I/O 디바이스"는, 예를 들어 처리 유닛에 데이터를 입력하기 위한 하나 이상의 입력 디바이스(예를 들어, 키보드, 마우스 등), 및/또는 처리 유닛과 연관된 결과를 표시하기 위한 하나 이상의 출력 디바이스(예를 들어, 디스플레이 등)를 포함하도록 의도된다.

[0043] 또한, 여기에 사용되는 바와 같은 어구 "네트워크 인터페이스"는, 예를 들어 컴퓨터 시스템이 적합한 통신 프로토콜을 통하여 다른 컴퓨터 시스템과 통신할 수 있게 해주는 하나 이상의 트랜시버를 포함하도록 의도된다.

[0044] 따라서, 여기에서 설명된 방법을 수행하기 위한 명령 또는 코드를 포함하는 소프트웨어 컴포넌트가 연관된 메모리 디바이스(예를 들어, ROM, 고정 또는 탈착가능한 메모리) 중 하나 이상에 저장될 수 있고, 이용될 준비가 되면 (예를 들어, RAM으로) 일부 또는 전체가 로딩되어 CPU에 의해 실행될 수 있다.

[0045] 어떠한 경우든, 여기에서 설명되고 첨부된 도면에 도시된 본 발명의 기술은 하드웨어, 소프트웨어, 또는 이들의 조합의 다양한 형태로 구현될 수 있으며, 예를 들어, 연관된 메모리, 구현 특정 집적 회로(들), 기능 회로 등을 갖는 하나 이상의 동작적으로 프로그래밍된 범용 디지털 컴퓨터로 구현될 수 있음을 인식하여야 한다. 여기에서 제공된 본 발명의 기술에 따라, 당해 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 기술의 다른 구현예들을 생각해볼 수 있을 것이다.

[0046] 본 발명의 예시적인 실시예가 첨부 도면을 참조하여 여기에 설명되었지만, 본 발명은 이들 구체적 실시예에 한정되지 않고, 본 발명의 범위 또는 사상에서 벗어나지 않고서 당해 기술 분야의 숙련자에 의해 다양한 기타 변경 및 수정이 이루어질 수 있음을 이해하여야 한다.

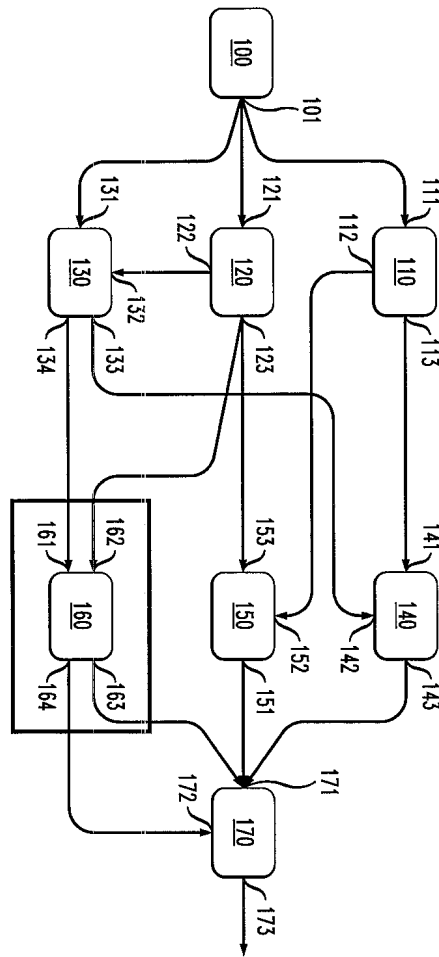
**도면의 간단한 설명**



- [0047] 도 1은 본 발명의 일 실시예에 따른 분석 컴포넌트의 그래프.
- [0048] 도 2는 본 발명의 일 실시예에 따른 데이터 포맷.
- [0049] 도 3은 본 발명의 일 실시예에 따른 주요 컴포넌트 분석을 위한 방법.
- [0050] 도 4는 본 발명의 일 실시예에 따른 입력-출력 의존관계 방정식 생성을 위한 방법.
- [0051] 도 5는 본 발명의 일 실시예에 따른 출처 모델의 확인 방법.
- [0052] 도 6은 본 발명의 일 실시예에 따른 출처 시스템.
- [0053] 도 7은 본 발명의 일 실시예에 따라 본 발명의 기술의 하나 이상의 컴포넌트/단계에 따라 구현될 수 있는 컴퓨터 시스템.

**도면**

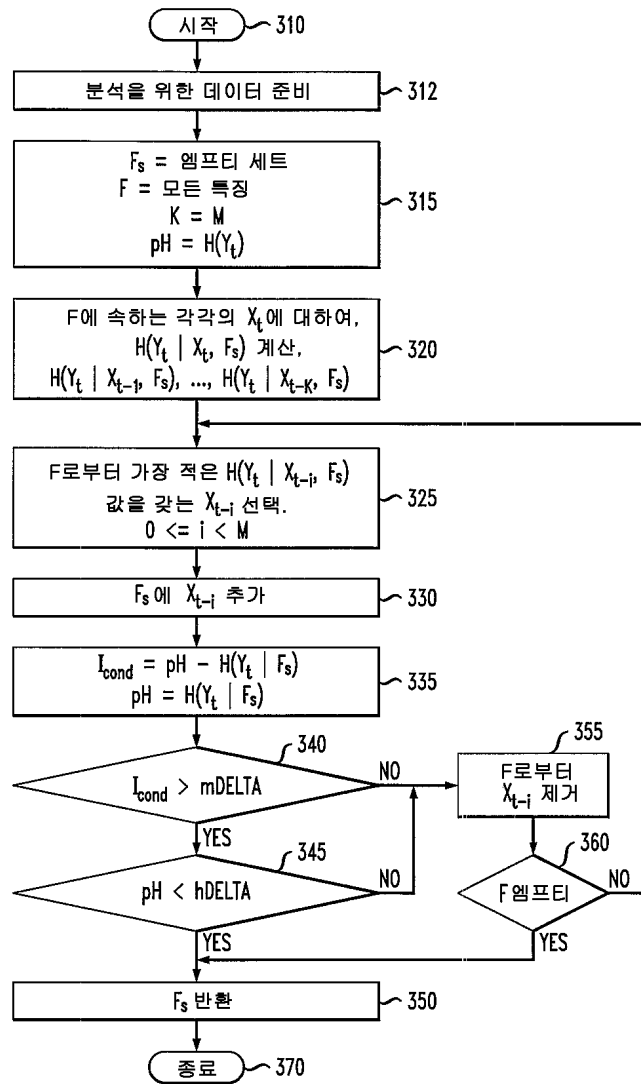
**도면1**



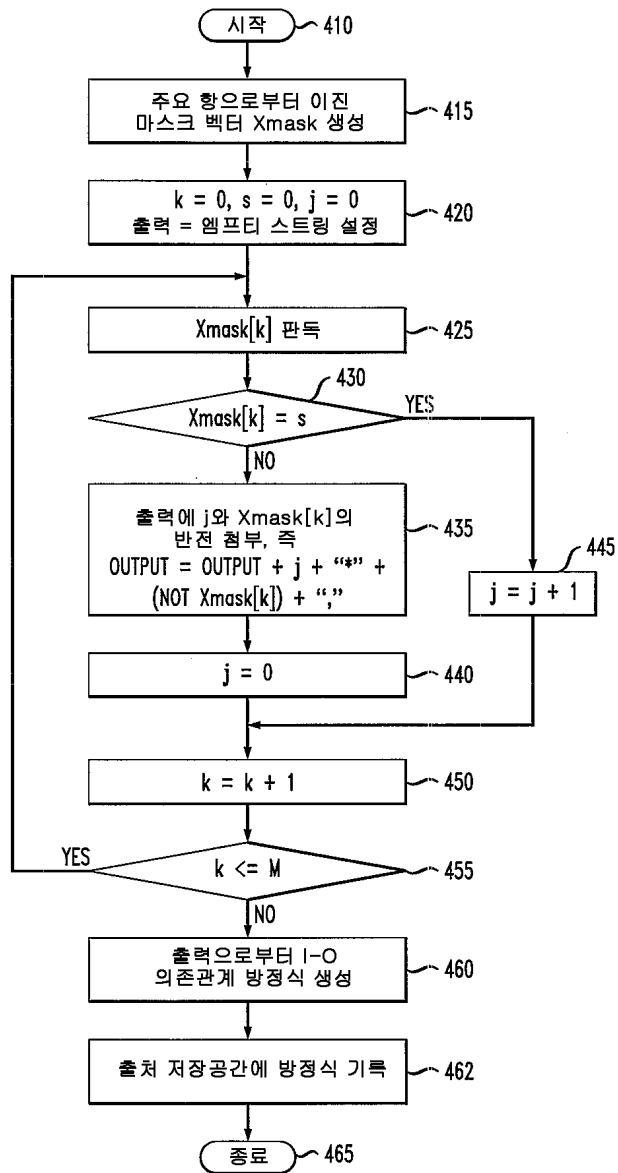
도면2

	입력 스트림	출력 스트림	
210 ~			240 ~
215 ~	$X_{-M+1} X_{-M+2} \dots X_{-5} X_{-4} X_{-3} X_{-2} X_{-1} X_0$	$Y_0$	245 ~
220 ~	$X_{-M+2} X_{-M+3} \dots X_{-4} X_{-3} X_{-2} X_{-1} X_0 X_1$	$Y_1$	250 ~
225 ~	$X_{-M+3} X_{-M+4} \dots X_{-3} X_{-2} X_{-1} X_0 X_1 X_2$	$Y_2$	255 ~
230 ~	$X_{-M+4} X_{-M+5} \dots X_{-2} X_{-1} X_0 X_1 X_2 X_3$	$Y_3$	260 ~
235 ~	...	...	265 ~

도면3

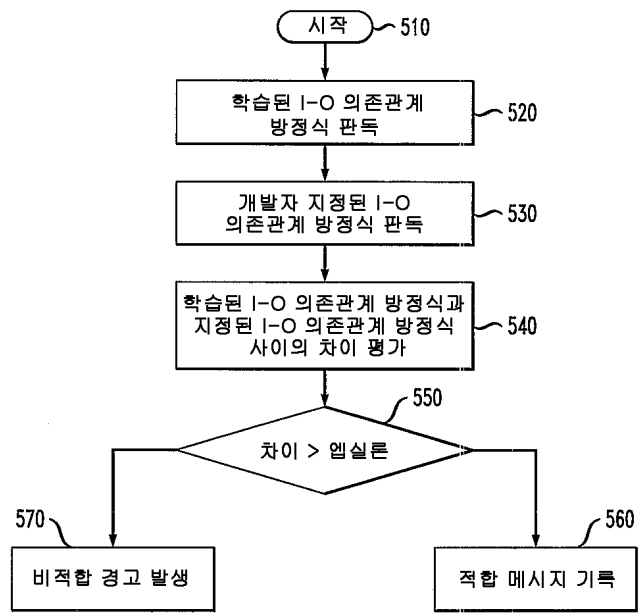


도면4

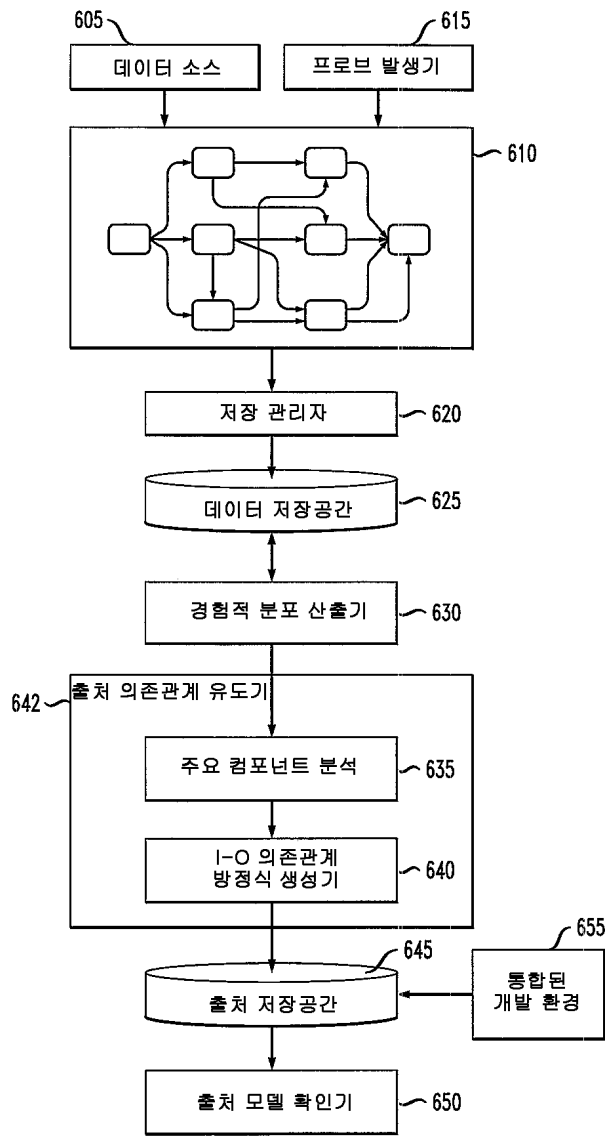




도면5



도면6



도면7

