



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0009687
(43) 공개일자 2019년01월29일

(51) 국제특허분류(Int. Cl.)
G06N 99/00 (2019.01)

(52) CPC특허분류
G06N 20/00 (2019.01)

(21) 출원번호 10-2018-0012825

(22) 출원일자 2018년02월01일

심사청구일자 없음

(30) 우선권주장

1020170091728 2017년07월19일 대한민국(KR)

(71) 출원인

한국전자통신연구원

대전광역시 유성구 가정로 218 (가정동)

(72) 발명자

김태환

대전광역시 유성구 지족북로 60, 205동 2405호 (지족동, 한화꿈에그린 2블럭)

김귀훈

대전광역시 서구 만년로 25, 105동 1002호 (만년동, 강변아파트)

(뒷면에 계속)

(74) 대리인

특허법인 무한

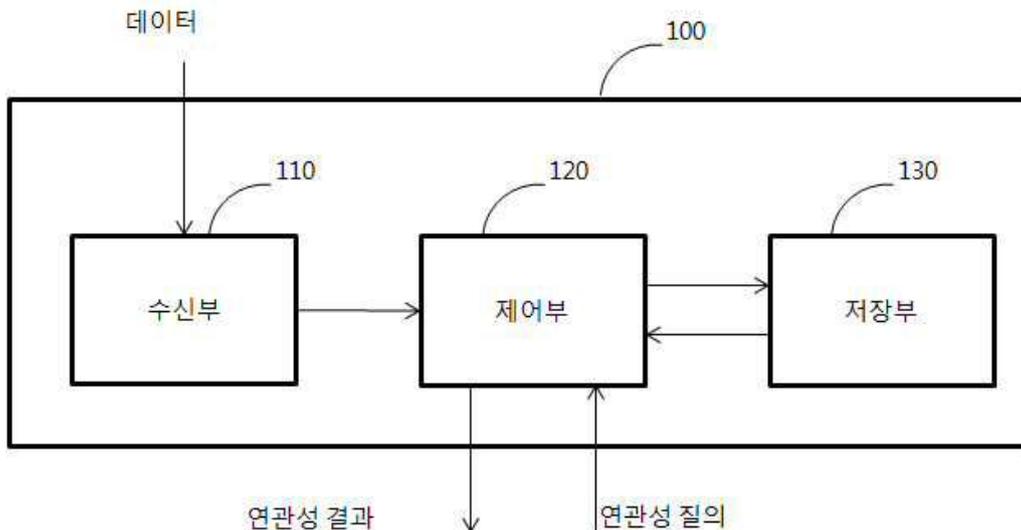
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 시계열 데이터의 연관성 분석 장치 및 방법

(57) 요약

시계열 데이터의 연관성 분석 장치 및 방법이 개시된다. 시계열 데이터의 연관성 분석 방법은 상기 특정 시점에 상기 시계열 데이터 중 제1 데이터와 상기 제1 데이터와의 연관성을 분석하고자 하는 제2 데이터에 대한 2차원 행렬 형태의 연관 필터에 대한 인덱스를 식별하는 단계; 상기 식별된 제1 데이터에 대한 연관 필터의 인덱스와 상기 식별된 제2 데이터에 대한 연관 필터의 인덱스가 서로 중첩되는 위치에서 연관 필터의 값을 추출하는 단계; 및 상기 추출된 연관 필터의 값을 모두 합산하여 상기 제1 데이터와 상기 제2 데이터 간의 연관성으로 분석하는 단계를 포함할 수 있다.

대표도 - 도1



(72) 발명자

김민석

대전광역시 유성구 지족로 362, 303동 1003호 (지족동, 반석마을아파트3단지)

홍용근

대전광역시 유성구 노은동로 219, 306동 1703호 (지족동, 열매마을3단지)

이 발명을 지원한 국가연구개발사업

과제고유번호 CRC-15-05-ETRI

부처명 국가과학기술연구회

연구관리전문기관 국가과학기술연구회

연구사업명 융합연구사업

연구과제명 자가학습형 지식융합 슈퍼브레인 핵심기술개발

기여율 1/1

주관기관 한국전자통신연구원

연구기간 2016.12.01 ~ 2017.11.30

명세서

청구범위

청구항 1

특정 시점에서의 시계열 데이터를 구성하는 데이터들 간 연관성 분석 방법에 있어서,
 상기 특정 시점에 상기 시계열 데이터 중 제1 데이터와 상기 제1 데이터와의 연관성을 분석하고자 하는 제2 데이터에 대한 2차원 행렬 형태의 연관 필터에 대한 인덱스를 식별하는 단계;
 상기 식별된 제1 데이터에 대한 연관 필터의 인덱스와 상기 식별된 제2 데이터에 대한 연관 필터의 인덱스가 서로 중첩되는 위치에서 연관 필터의 값을 추출하는 단계; 및
 상기 추출된 연관 필터의 값을 모두 합산하여 상기 제1 데이터와 상기 제2 데이터 간의 연관성으로 분석하는 단계를 포함하는 시계열 데이터의 연관성 분석 방법.

발명의 설명

기술 분야

[0001] 본 발명은 시계열 데이터의 연관성 분석에 관한 것으로, 보다 구체적으로는 연속적인 트랜잭션(transaction) 혹은 이벤트(event)에 의해 발생하는 시계열 데이터(sequence data)의 연관성을 분석하기 위한 기계학습 방법과 그 장치에 관한 것이다.

배경 기술

[0002] 데이터의 연관성 분석을 위한 종래 기술로는 통계적 기법을 기반으로 데이터의 지지도(Support), 신뢰도(Confidence), 향상도(Lift), IS(Interest-Support) 척도, 교차 지지도(Cross Support) 등을 계산하여 연관성 규칙(association rule)의 평가척도로서 사용하는 연관규칙 분석 방법이 있으며, 대표적으로 아프리오리(Apriori) 알고리즘 등이 있다. 연관성 규칙에 시간 개념을 추가하여 시계열 데이터에 대한 연관성 분석을 위한 방법으로는 순차 연관성 분석 방법(순차 패턴 분석 알고리즘, Agrawal and Srikant, 1995)이 있다. 이러한 연관성 분석 방법은 다량의 데이터에 대한 통계분석을 기반으로 하고 있으며, 장바구니 분석(Market Basket Analysis), 마케팅, 인터넷의 웹 페이지 설계, 부정 탐지(Fraud Detection), 추천 시스템(Recommendation system) 등의 다양한 분야에서 활용되고 있다.

[0003] 그러나 종래 기술들은 분석할 데이터(항목)의 수가 증가하면 복잡도가 크게 증가하고, 반대로 데이터(항목)의 수가 적으면 연관성 규칙을 발견할 수 없는 오류를 가지고 있다. 따라서 상기한 기존의 통계적 기법 기반의 연관성 분석 방법은 다량의 데이터가 확보된 상황에서 후처리(post-processing) 방법으로 실시간 분석을 요구하는 응용 분야에 적용하기 어려운 문제가 존재한다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 시계열 데이터의 실시간 연관성 분석을 요구하는 분야, 분석할 데이터(항목)의 수가 많은 분야 및 데이터(항목)의 수가 적은 분야에 공통으로 적용하여 사용할 수 있는 기계학습 방법과 장치에 관한 것으로 시계열 데이터에 포함된 데이터 간의 상호 연관성을 분석하여 수치로 제공하는 장치 및 방법을 제공한다.

과제의 해결 수단

[0005] 본 발명의 일실시예에 따른 연속되는 데이터로 구성되는 시계열 데이터의 연관성 학습 방법은 상기 데이터의 특성에 따라 상기 데이터를 하나 이상의 필드로 구분하는 단계; 상기 구분된 필드의 구조에 기초하여 상기 시계열 데이터를 구성하는 데이터들 간 2차원 행렬 형태의 연관 필터를 생성하는 단계; 상기 시계열 데이터 중 가장 최

근에 수신된 제1 데이터와 기존에 존재하던 제2 데이터 간의 연관성을 관리하기 위하여 상기 제1 데이터 및 상기 제2 데이터에 대한 상기 연관 필터의 인덱스를 식별하는 단계; 상기 식별된 제1 데이터에 대한 연관 필터의 인덱스와 상기 식별된 제2 데이터에 대한 연관 필터의 인덱스가 중첩되는 위치에 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 업데이트 하는 단계를 포함할 수 있다.

[0006] 상기 생성하는 단계는 상기 데이터들 간 동일 필드에 대한 연관성을 관리하는 메인 연관 필터(Main Relation Filter) 및 서로 다른 데이터 필드에 대한 연관성을 관리하는 횡단 연관 필터(Cross Relation Filter)를 생성하고, 상기 업데이트 하는 단계는 상기 생성된 메인 연관 필터 및 횡단 연관 필터를 이용하여 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 계산할 수 있다.

[0007] 상기 업데이트 하는 단계는 상기 제1 데이터가 수신될 때마다 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 제1 방식에 따라 계산하여 상기 연관 필터에 업데이트 할 수 있다.

[0008] 상기 제1 방식에 따라 업데이트하는 단계는 상기 생성된 시계열 데이터를 구성하는 데이터들 간 연관 필터에 대해 상기 제1 데이터에 대한 연관 필터의 인덱스와 상기 제2 데이터에 대한 연관 필터의 인덱스가 중첩되는 위치를 확인하는 단계; 및 상기 확인된 위치에 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 미리 설정된 값에 따라 증가시켜 업데이트 하는 단계를 포함할 수 있다.

[0009] 상기 업데이트 하는 단계는 특정 시간 주기 또는 특정 데이터 수신 주기 마다 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 제2 방식에 따라 계산하여 상기 연관 필터에 업데이트 할 수 있다.

[0010] 상기 제2 방식에 따라 업데이트 하는 단계는 상기 생성된 시계열 데이터를 구성하는 데이터들 간 연관 필터에 포함된 모든 값을 미리 설정한 값에 따라 감소시켜 업데이트 하는 단계를 포함할 수 있다.

[0011] 본 발명의 일 실시 예에 따른 시계열 데이터의 연관성 분석 방법은 상기 특정 시점에 상기 시계열 데이터 중 제1 데이터와 상기 제1 데이터와의 연관성을 분석하고자 하는 제2 데이터에 대한 2차원 행렬 형태의 연관 필터에 대한 인덱스를 식별하는 단계; 상기 식별된 제1 데이터에 대한 연관 필터의 인덱스와 상기 식별된 제2 데이터에 대한 연관 필터의 인덱스가 서로 중첩되는 위치에서 연관 필터의 값을 추출하는 단계; 및 상기 추출된 연관 필터의 값을 모두 합산하여 상기 제1 데이터와 상기 제2 데이터 간의 연관성으로 분석하는 단계를 포함할 수 있다.

[0012] 상기 제1 데이터와 상기 제2 데이터에 대한 연관 필터가 복수인 경우 상기 추출하는 단계는 상기 복수의 연관 필터 각각에 대해 상기 제1 데이터에 대한 연관 필터의 인덱스와 상기 제2 데이터에 대한 연관 필터의 인덱스가 서로 중첩되는 위치에서 연관 필터의 값을 추출하고, 상기 분석하는 단계는 상기 복수의 연관 필터 각각에 대해 추출된 연관 필터의 값을 모두 합산함으로써 상기 제1 데이터와 상기 제2 데이터 간의 연관성을 계산할 수 있다.

[0013] 상기 연관 필터는 상기 시계열 데이터가 새롭게 수신될 때마다 상기 새롭게 수신된 시계열 데이터와 기존의 시계열 데이터 간의 연관성이 증가되어 업데이트 될 수 있다.

[0014] 상기 연관 필터는 특정 시간 주기 또는 특정 데이터 수신 주기 마다 상기 시계열 데이터에 포함된 데이터들 간의 연관성이 감소되어 업데이트 될 수 있다.

발명의 효과

[0015] 본 발명의 일 실시예에 의하면, 시계열 데이터의 실시간 연관성 분석을 요구하는 분야, 분석할 데이터(항목)의 수가 많은 분야 및 데이터(항목)의 수가 적은 분야에 공통으로 적용하여 사용할 수 있다.

[0016] 본 발명의 일 실시예에 의하면, 시계열 데이터의 패턴을 학습함으로써 특정 시점에서 시계열 데이터에 포함된 데이터 간 상호 연관성을 분석하여 수치로 제공할 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 일 실시예에 따른 시계열 데이터의 연관성 분석 장치에 대한 개념도를 도시한 도면이다.
 도 2는 본 발명의 일 실시예에 따른 시계열 데이터의 연관성을 업데이트 하는 방법을 도시한 도면이다.
 도 3은 본 발명의 일 실시예에 따른 연관 필터의 생성 방법을 도시한 도면이다.
 도 4는 본 발명의 일 실시예에 따른 해시함수를 이용하여 데이터의 연관 필터 내의 인덱스를 계산하는 방법을 도

시한 도면이다.

도 5는 본 발명의 일실시예에 따른 제1 방식에 따라 데이터 간 연관성을 업데이트하는 개념을 도시한 도면이다.

도 6은 본 발명의 일실시예에 따른 제1 방식에 적용되는 연관성을 증가시키는 함수의 예를 도시한 도면이다.

도 7은 본 발명의 일실시예에 따른 제2 방식에 따라 데이터 간 연관성을 업데이트하는 개념을 도시한 도면이다.

도 8은 본 발명의 일실시예에 따른 제2 방식에 적용되는 연관성을 감소시키는 함수의 예를 도시한 도면이다.

도 9는 본 발명의 일실시예에 따른 시계열 데이터의 연관성을 계산하는 방법을 도시한 도면이다.

도 10은 본 발명의 일실시예에 따른 시계열 데이터에 포함된 하나의 특정 데이터와 시계열 데이터 상의 모든 데이터와의 연관성을 계산하는 방법을 도시한 도면이다.

도 11은 본 발명의 일실시예에 따른 데이터와 데이터 간의 연관성을 계산하는 방법을 도시한 도면이다.

도 12는 본 발명의 일실시예에 따른 데이터 집합과 데이터 집합 간의 연관성을 계산하는 방법을 도시한 도면이다.

도 13은 본 발명의 일실시예에 따른 연관성 분석 방법을 지능형 캐싱 시스템에 적용한 예를 도시한 도면이다.

도 14는 본 발명의 일실시예에 따른 연관성 분석 방법을 장바구니 분석 시스템에 적용한 예를 도시한 도면이다.

도 15는 본 발명의 일실시예에 따른 실시간 장바구니 분석 결과의 예를 도시한 도면이다.

도 16은 본 발명의 일실시예에 따른 온라인 추천 시스템을 장바구니 분석 시스템에 적용한 예를 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0018] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0019] 도 1은 본 발명의 일실시예에 따른 시계열 데이터의 연관성 분석 장치에 대한 개념도를 도시한 도면이다.

[0020] 도 1을 참고하면, 본 발명의 연관성 분석 장치(100)는 시간에 따라 연속되는 데이터로 구성되는 시계열 데이터를 수신하는 수신부(110), 수신한 시계열 데이터의 패턴을 학습하여 특정 시점에서 시계열 데이터에 포함된 데이터 상호 간의 연관성을 분석하고 이를 수치로 제공하는 제어부(120) 및 시계열 데이터에 포함된 데이터 상호 간의 연관성을 업데이트하기 위한 모델링 함수를 포함하는 저장부(130)로 구성될 수 있다.

[0021] 구체적으로 수신부(110)는 특정 시간 동안 생성된 시계열 데이터, 혹은 지속적으로 생성 중인 데이터를 시간 순서대로 수신하여 제어부(120)로 전송할 수 있다. 이때, 수신부(110)는 해당 발명의 적용 분야에 대한 시스템 특성과 데이터 종류에 따라 다르게 정의될 수 있다. 예를 들어, 유튜브(Youtube)와 같이 사용자에게 콘텐츠를 제공하는 인터넷 응용 서비스에서 수신부(110)는 실시간으로 발생하는 사용자의 콘텐츠 요청 이벤트 정보를 시간 순서대로 수신하여 제어부(120)로 전송할 수 있다. 또는 장바구니 분석을 위한 매출 관리 시스템에서 수신부(110)는 계산대 시스템(PoS)에서 발생하는 거래 트랜잭션 데이터를 수신하여 제어부(120)로 전송할 수 있다. 또는 네트워크 패킷 간 연관성 분석을 위한 시스템에서 수신부(110)는 실시간으로 수신한 패킷의 헤더 정보를 추출한 데이터를 실시간으로 수신하여 제어부(120)로 전송할 수 있다. 이와 같이 수신부(110)는 해당 발명이 적용, 운영되는 시스템과 모니터링 할 데이터 형태에 따라 적절하게 구현될 수 있다.

[0022] 제어부(120)는 수신된 시계열 데이터에 포함된 데이터 상호 간의 연관성을 저장 및 관리하기 위하여 2차원 행렬 형태로 구성되는 연관 필터를 생성할 수 있으며, 생성된 연관 필터 내에는 시계열 데이터에 포함된 데이터 상호 간의 연관성이 숫자 형태로 저장될 수 있다.

[0023] 이때, 제어부(120)는 수신된 시계열 데이터에 포함된 데이터가 자체 특성에 따라 다차원(multi-dimension) 형태를 가지거나 복수의 필드(field)로 구성되어 있는 경우, 시계열 데이터에 포함된 데이터 간의 연관성을 저장 및 관리하기 위하여 복수의 연관 필터(Correlation filter)를 생성할 수 있다.

[0024] 이때 필요한 경우 제어부(120)는 수신된 시계열 데이터에 포함된 데이터를 복수의 필드로 분할하는 전처리를 수행할 수 있으며, 분할된 복수의 필드 개수에 기초하여 시계열 데이터에 포함된 데이터 상호 간의 연관성을 저장 및 관리하기 위한 복수의 연관 필터들을 생성할 수 있다. 즉, 본 발명의 연관성 분석 장치(100)는 시계열 데이터에 포함된 데이터를 구성하는 하나의 필드에 대해 하나 이상의 연관 필터를 생성할 수 있으며, 생성된 연관

필터들을 통해 필드 단위로 시계열 데이터에 포함된 데이터 간의 연관성을 저장 및 관리함으로써 보다 세밀하게 연관성 분석을 지원할 수 있다.

- [0025] 제어부(120)는 시계열 데이터가 수신되고 필요에 따라 전처리 되면, 수신된 시계열 데이터의 필드와 해당 필드에 대한 필드 값에 따라 수신된 시계열 데이터에 대한 연관 필터의 특정 위치를 결정할 수 있다. 이후 제어부(120)는 결정된 연관 필터의 특정 위치에 수신된 시계열 데이터와 기존에 이미 존재하는 시계열 데이터 간의 연관성을 업데이트 할 수 있다.
- [0026] 구체적으로 제어부(120)는 시계열 데이터가 수신될 때마다 상기 결정된 연관 필터의 특정 위치의 값을 연관성을 증가시키는 함수를 이용하여 업데이트 할 수 있다. 이때, 연관성을 증가시키는 함수는 시계열 데이터를 수신하였을 때 수신한 시계열 데이터와 기존에 이미 존재하는 시계열 데이터들 간의 연관성을 생성하고 증가시키는 함수이다.
- [0027] 또한 제어부(120)는 시계열 데이터 간의 연관성이 저장되어 있는 연관 필터 내의 모든 값을 시간이 지남에 따라 데이터들 간의 연관성을 감소시키는 함수를 이용하여 업데이트 할 수 있다. 즉, 제어부(120)는 연관성을 감소시키는 함수를 이용하여 특정 시간 윈도우(time window) 또는 수신되는 데이터 수신 주기(data receiving interval)에 따라 연관 필터 내에 존재하는 모든 값을 감소시킬 수 있다. 이와 같은 시계열 데이터들 간의 연관성을 증가시키는 함수 및 시계열 데이터들 간의 연관성을 감소시키는 함수와 같은 모델링 함수들은 저장부(130)에 저장될 수 있으며, 연관성 분석 장치(100)가 적용되는 발명에 따라 적절한 모델링 함수가 선택될 수 있다.
- [0028] 이와 같이 시계열 데이터 간의 연관성이 업데이트 된 연관 필터를 이용하여 제어부(120)는 특정 시점에서 시계열 데이터 간의 연관성을 계산할 수 있다. 이때, 제어부(120)는 하나의 시계열 데이터 간의 연관성뿐만 아니라 하나의 시계열 데이터에 대해 지금까지 수신한 모든 시계열 데이터와의 연관성 또는 시계열 데이터의 일부 집합과 시계열 데이터 다른 집합 간의 연관성도 계산할 수 있다.
- [0029] 도 2는 본 발명의 일실시예에 따른 시계열 데이터의 연관성을 업데이트 하는 방법을 도시한 도면이다.
- [0030] 단계(210)에서, 연관성 분석 장치(100)는 시간에 따라 연속되는 데이터로 구성되는 시계열 데이터를 수신할 수 있다.
- [0031] 단계(220)에서, 연관성 분석 장치(100)는 수신된 시계열 데이터의 연관성을 저장 및 관리하기 위하여 $N \times N$ 2차원 행렬 형태로 구성되는 연관 필터를 생성할 수 있다. 이때, 연관성 분석 장치(100)는 수신된 시계열 데이터가 자체 특성에 따라 다차원(multi-dimension) 형태를 가지거나 하나의 시계열 데이터가 복수의 필드(field)로 구성되어 있는 경우, 시계열 데이터 간의 연관성을 저장 및 관리하기 위하여 복수의 연관 필터(Correlation filter)를 생성할 수 있다. 그리고 생성된 연관 필터 내에는 시계열 데이터 상호 간의 연관성이 숫자 형태로 저장될 수 있다.
- [0032] 단계(230)에서, 연관성 분석 장치(100)는 수신된 시계열 데이터의 필드와 해당 필드에 대한 필드 값에 따라 수신된 시계열 데이터에 대한 연관 필터의 특정 위치를 결정할 수 있다.
- [0033] 단계(240)에서, 연관성 분석 장치(100)는 결정된 연관 필터의 특정 위치에 수신된 시계열 데이터와 기존에 이미 존재하는 시계열 데이터 간의 연관성을 업데이트 할 수 있다. 이때, 연관성 분석 장치(100)는 시계열 데이터가 수신될 때마다 상기 결정된 연관 필터의 특정 위치의 값을 연관성을 증가시키는 함수를 이용하여 업데이트 하거나, 시계열 데이터 간의 연관성이 저장되어 있는 연관 필터 내의 모든 값을 시간이 지남에 따라 데이터들 간의 연관성을 감소시키는 함수를 이용하여 업데이트 할 수 있다.
- [0034] 도 3은 본 발명의 일실시예에 따른 연관 필터의 생성 방법을 도시한 도면이다.
- [0035] 본 발명의 연관성 분석 장치(100)는 시계열 데이터에 포함된 데이터 간의 정확한 연관성을 분석하기 위해 데이터를 복수의 필드로 분할하고, 필드 간 연관성을 분석함으로써 데이터 간 세밀한 연관성 분석을 지원할 수 있다. 이를 위해 연관성 분석 장치(100)는 수신된 시계열 데이터에 포함된 데이터의 자체 특성에 복수의 필드를 구분하는 전처리를 수행할 수 있다. 이때, 연관성 분석 장치(100)는 수신된 시계열 데이터에 포함된 데이터 자체가 복수의 필드들로 구성되어 있는 경우에는 해당 데이터의 필드 구성 자체를 활용할 수 있다. 이와는 달리 수신된 시계열 데이터에 포함된 데이터가 단일 필드로 구성되어 있는 경우 연관성 분석 장치(100)는 데이터를 복수의 필드로 분할하는 전처리 없이 진행하거나 단일 필드로 구성된 데이터를 자체적인 규칙을 기반으로 복수의 필드로 분할하는 전처리를 수행 할 수 있다.
- [0036] 예를 들어 유튜브 시스템에서 사용자의 콘텐츠 요청 데이터의 경우, 하나의 요청 데이터를 콘텐츠 제목, 콘텐츠

게시자, 콘텐츠 게시 위치, 추천인 리스트 및 콘텐츠 선호도 등의 필드들로 나누는 전처리를 수행할 수 있다. 또한, 장바구니 분석 시스템의 거래 트랜잭션 데이터의 경우, 매출 품목의 종류별(ex. 식료품, 신선제품, 가전 제품, 생활용품, 화학용품 등)으로 나누어 필드를 구성할 수 있다.

[0037] 연관성 분석 장치(100)는 전처리를 통해 확인된 데이터의 필드 구성을 이용하여 연관 필터를 생성할 수 있다. 연관 필터는 시계열 데이터에 포함된 데이터 간의 연관성을 숫자 형태로 저장 및 관리하는 데이터 구조로 2차원 행렬 형태로 생성될 수 있다. 이때, 생성된 연관 필터는 내부의 데이터가 모두 동일한 초기 값으로 설정될 수 있다.

[0038] 이후 연관성 분석 장치(100)는 생성된 연관 필터에 대해 시계열 데이터에 포함된 데이터가 수신 될 때마다 연관성을 증가시키는 함수를 통해 연관성을 업데이트 하고, 시간이 지남에 따라 시계열 데이터에 포함된 모든 데이터의 연관성을 감소시키는 함수를 통해 연관성을 업데이트 할 수 있다.

[0039] 이때, 생성된 연관 필터의 개수는 연관성 분석 장치(100)가 수신된 시계열 데이터에 포함된 데이터를 어떻게 복수의 필드로 분할하는지에 따라 결정될 수 있다. 즉, 수신된 시계열 데이터에 포함된 데이터의 필드가 도 3의 (a)와 같이 field1의 단일 필드로 구성된 경우, 시계열 데이터에 포함된 Data A와 Data B 간에 하나의 연관 필터가 생성될 수 있다.

[0040] 이와는 달리 도 3의 (b)와 같이 수신된 시계열 데이터에 포함된 데이터의 필드가 field1, field2, field3과 같이 3개로 분할된 경우, 시계열 데이터에 포함된 Data A와 Data B 간에는 복수의 연관 필터가 생성될 수 있다. 구체적으로 연관성 분석 장치(100)는 동일 필드(field1과 field1, field2와 field2, field3과 field3)에 대한 연관성을 저장하는 3개의 메인 연관 필터(Main Relation Filter)(점선)와 Data A와 Data B 간의 서로 다른 필드(field1과 field2, field2와 field3, field1과 field3)에 대한 연관성을 저장하는 3개의 횡단 연관 필터(Cross Relation Filter)(직선) 및 Data A와 Data B 자체를 식별할 수 있는 데이터 식별 필터(Data Identification Filter), 총 7개의 연관 필터를 생성할 수 있다. 이와 같은 연관 필터의 구성은 본 발명의 연관성 분석 장치(100)가 적용되는 도메인 지식에 따라 유연하게 설정될 수 있다.

[0041] 도 4는 본 발명의 일실시예에 따른 해시함수를 이용하여 데이터의 연관 필터 내의 인덱스를 계산하는 방법을 도시한 도면이다.

[0042] 연관성 분석 장치(100)는 수신된 시계열 데이터에 포함된 데이터의 연관성을 연관 필터에 저장 및 관리하기 위해서 해당 데이터에 대한 연관 필터의 특정 위치인 인덱스(index)를 결정할 수 있다. 연관 필터는 해당 데이터의 필드에 따라 관리되기 때문에 연관성 분석 장치(100)는 하기의 식 1을 통해 해당 데이터의 필드 값을 해쉬한 후 연관 필터의 크기로 모듈러 연산을 함으로써 해당 데이터의 필드 값에 대한 인덱스를 계산할 수 있다.

[0043] <식 1>

[0044]
$$\text{index of data "d"} = H(\text{"d"}) \% \text{FilterSize}$$

[0045] 일례도 도 4와 같이 연관 필터의 크기가 1,000 x 1000인 시스템에 "d"라는 데이터가 수신된 경우, 상기 식 1을 통해 연관성 분석 장치(100)는 데이터 "d"의 인덱스 값을 1,000 이하의 특정 값으로 결정할 수 있다. 따라서 데이터 "d"의 인덱스 값으로 3이 나온 경우, 연관성 분석 장치(100)는 연관 필터의 (x, 3)과 (3, y)의 위치에는 데이터 "d"와 다른 데이터 간의 연관성을 저장할 수 있다.

[0046] 도 5는 본 발명의 일실시예에 따른 제1 방식에 따라 데이터 간 연관성을 업데이트하는 개념을 도시한 도면이다.

[0047] 연관성 분석 장치(100)는 시계열 데이터가 수신될 때마다 수신된 제1 데이터와 기존에 이미 존재하는 제2 데이터들 간의 연관성을 업데이트하기 위하여 연관성을 증가시키는 함수를 이용할 수 있다.

[0048] 수신한 제1 데이터의 필드 값에 의하여 연관 필터의 인덱스 x가 결정되면, 연관성 분석 장치(100)는 기존에 이

미 존재하는 제2 데이터들에 대한 연관 필터의 인덱스들 y_i 와의 연관성을 업데이트 하기 위해서 연관성을 증가시

키는 함수를 이용하여 (x, y_i)와 (y_i , x)에 해당하는 연관 필터의 값들을 증가시킬 수 있다. 만약 수신된 제1 데이터가 최초로 수신된 데이터여서 연관성을 맺을 이전 데이터가 없는 경우, 연관성 분석 장치(100)는 제1 데이터에 후속하여 제3 데이터가 수신되었을 때, 제1 데이터와 제3 데이터 간의 연관성을 업데이트 할 수 있다.

- [0049] 예를 들어, 데이터 "d"가 최초로 수신된 경우, 데이터 "d"와 연관성을 맺을 이전 데이터가 없으므로 연관성 업데이트는 수행되지 않을 수 있다. 이후 두 번째 데이터인 "c"가 수신된 경우, 연관성 분석 장치(100)는 최초로 수신된 데이터 "d"와의 연관성을 연관 필터의 인덱스 ($H("d")\%FilterSize$, $H("c")\%FilterSize$)의 위치에 업데이트 할 수 있다. 세 번째 데이터로 "b"가 수신되면 첫 번째 수신된 데이터 "d"와 두 번째 수신된 데이터 "c"와의 연관성을 해당 연관 필터의 인덱스 ($H("d")\%FilterSize$, $H("b")\%FilterSize$)와 ($H("c")\%FilterSize$, $H("b")\%FilterSize$)의 위치에 업데이트 할 수 있다.
- [0050] 이러한 인덱스 결정 방식은 시계열 데이터에 포함된 데이터 간의 연관성뿐만 아니라 연관성의 인과관계도 내포할 수 있다. 다시 말해, 인덱스 ($H("d")\%FilterSize$, $H("c")\%FilterSize$)의 값은 데이터 "e"와 데이터 "c"의 연관성 정보를 내포하고 있을 뿐 아니라 해당 연관성은 데이터 "c"에 의한 데이터 "e"와 데이터 "c" 간의 연관성이라는 것을 내포할 수 있다. 반대로, 인덱스 ($H("c")\%FilterSize$, $H("d")\%FilterSize$)의 값은 데이터 "d"로 인한 데이터 "c"와 데이터 "d"의 연관성이라는 것을 내포할 수 있다.
- [0051] 이때, 업데이트되는 연관 필터의 값은 사전에 정의된 연관성을 증가시키는 함수에 의해 결정될 수 있다. 만약, 데이터 "d"가 이전에 수신한 적이 없어 해당 데이터에 대한 연관 필터의 인덱스에 연관성 값이 없는 경우, 연관성 분석 장치(100)는 0을 초기 값(initial value)으로 설정할 수 있다. 반면, 기존의 데이터 "d"를 수신한 적이 있어 해당 데이터에 대한 연관 필터의 인덱스에 연관성 값들이 이미 존재하는 경우 연관성 분석 장치(100)는 사전에 정의된 연관성을 증가시키는 함수를 이용하여 연관성을 증가시킬 수 있다. 보다 구체적인 연관성을 증가시키는 함수는 다음의 도 6을 통해 설명하도록 한다.
- [0052] 도 6은 본 발명의 일실시예에 따른 제1 방식에 적용되는 연관성을 증가시키는 함수의 예를 도시한 도면이다.
- [0053] 도 6을 참고하면 연관성을 증가시키는 함수들은 모두 최대 값을 1로, 최소 값을 0으로 정의하고 있으며, 시계열 데이터가 수신될 때마다 연관 필터의 해당 인덱스에 대한 연관성 값을 증가시키기 위한 형태를 보여준다. 또한, 연관성 분석 장치(110)가 적용되는 도메인의 상황에 따라 시계열 데이터에 포함된 데이터 간의 연관성을 관리하기 위한 추가 정보가 존재하는 경우 해당 추가 정보를 연관성을 증가시키는 함수의 파라미터로 적용할 수 있다. 예를 들어, 사용자의 콘텐츠 요청 데이터인 경우, 사용자의 멤버십 정보가 추가적으로 데이터에 존재하면, 연관성을 증가시키는 함수에 멤버십 정보를 파라미터로 삽입하여 연관성을 증가시키기 위한 함수의 형태를 변경할 수 있다. 이와 같은 연관성을 증가시키는 함수의 파라미터의 확장성은 멤버십에 따라 연관성 정보에 차등을 두어 관리할 수 있는 방법을 제공할 수 있다.
- [0054] 도 7은 본 발명의 일실시예에 따른 제2 방식에 따라 데이터 간 연관성을 업데이트하는 개념을 도시한 도면이다.
- [0055] 연관성 분석 장치(100)는 특정 시간 주기(time window) 또는 데이터 수신 주기(data receiving interval)에 따라 시계열 데이터에 포함된 데이터들 간의 연관성을 업데이트하기 위하여 연관성을 감소시키는 함수를 이용할 수 있다. 예를 들어 연관성 분석 장치(100)는 연관성을 감소시키는 함수를 30초에 한번씩 호출하거나 데이터 30개를 수신할 때 마다 연관성을 감소시키는 함수를 호출하여 데이터들 간의 연관성을 업데이트할 수 있다. 연관성을 증가시키는 함수가 수신한 데이터를 기준으로 기존의 다른 데이터와의 연관성을 확장하기 위해 연관 필터의 특정 인덱스에 대응하는 값들을 증가시키는 것과는 달리, 연관성을 감소시키는 함수는 0을 제외한 연관 필터 내의 모든 값에 적용되어 연관성을 감소시킬 수 있다.
- [0056] 도 8은 본 발명의 일실시예에 따른 제2 방식에 적용되는 연관성을 감소시키는 함수의 예를 도시한 도면이다.
- [0057] 연관성을 감소시키는 함수는 연관 필터 내의 0을 제외한 모든 값을 정해진 함수 $f(x)$ 에 따라 감소시킨다. 도 8을 참고하면 연관성을 감소시키는 함수들은 모두 최대 값을 1로, 최소 값을 0으로 정의하고 있으며, 연관성 분석 장치(100)는 연관성을 감소시키는 함수가 호출될 때 마다 연관 필터 내의 연관성 값을 연관성을 감소시키는 함수의 형태에 따라 감소시킬 수 있다.
- [0058] 이와 같은 연관성을 증가시키는 함수와 연관성을 감소시키는 함수는 데이터 간 연관성을 확장시키고 감소시키는 역할을 수행하여 연관 필터를 최신으로 업데이트할 수 있다. 이때, 연관성 분석 장치(100)는 적용되는 도메인에 적합하게 연관성을 증가시키는 함수 및 연관성을 감소시키는 함수를 선택하거나 새롭게 정의할 수 있다. 연관성 분석 장치(100)는 이와 같은 모델링 함수들(연관성을 증가시키는 함수 및 연관성을 감소시키는 함수)에 의해 최신으로 업데이트된 연관 필터를 기반으로 특정 시점에서의 시계열 데이터에 포함된 데이터들 간 연관성을 분석할 수 있다.
- [0059] 도 9는 본 발명의 일실시예에 따른 시계열 데이터의 연관성을 계산하는 방법을 도시한 도면이다.

- [0060] 단계(910)에서, 연관성 분석 장치(100)는 특정 시점에 시계열 데이터 중 제1 데이터와 상기 제1 데이터와의 연관성을 분석하고자 하는 제2 데이터에 대한 2차원 행렬 형태의 연관 필터에 대한 인덱스를 식별할 수 있다.
- [0061] 단계(920)에서, 연관성 분석 장치(100)는 식별된 제1 데이터에 대한 연관 필터의 인덱스와 식별된 제2 데이터에 대한 연관 필터의 인덱스가 서로 중첩되는 위치에서 연관 필터의 값을 추출할 수 있다.
- [0062] 단계(930)에서, 연관성 분석 장치(100)는 추출된 연관 필터의 값을 모두 합산하여 제1 데이터와 제2 데이터 간의 연관성을 분석할 수 있다.
- [0063] 도 10은 본 발명의 일실시예에 따른 시계열 데이터에 포함된 하나의 특정 데이터와 시계열 데이터 상의 모든 데이터와의 연관성을 계산하는 방법을 도시한 도면이다.
- [0064] 연관성 분석 장치(100)는 외부로부터 연관성 질의를 수신하면 연관 필터를 기반으로 시계열 데이터에 포함된 데이터 간, 데이터 대 데이터 집단 간, 데이터 집단 대 데이터 집단 간의 연관성 계산하여 결과를 응답할 수 있다. 데이터 간의 연관성을 복수의 필드로 나누어 분석하는 경우 복수의 필터로 나누어 연관성 정보를 관리하기 때문에 연관성 계산시 각 필터에서 계산된 값에 상수(k)를 부여하여 가중치를 부과할 수 있다. 본 발명의 일 실시예에 따른 도 10, 도 11 및 도 12의 예는 이 상수를 모두 1로 설정하고 계산한 것이다.
- [0065] 특정 시점에서 특정 데이터와 현재까지의 수신한 모든 데이터와의 연관성을 계산하는 방법은 다음과 같다. 예를 들어 도 10과 같이 특정 데이터 "d"와 지금까지 연관성 분석 장치(100)가 수신한 모든 데이터와의 연관성을 분석하기 위해 특정 데이터 "d"에 대한 연관 필터의 인덱스를 계산하고 해당 인덱스에 대응하는 모든 행 및 열에 포함된 연관 필터의 값을 합산하여 연관성(Impulse)을 계산할 수 있다.
- [0066] 이때, 도 10의 3열은 특정 데이터 "d"에 의한 모든 데이터와 "d"와의 연관성 값이 포함(x, 3)되어 있고, 3행은 다른 데이터에 의한 다른 모든 데이터와 "d"와의 연관성 값이 포함(y, 3)되어 있다. 그 결과 특정 데이터 "d"에 의한 모든 데이터와 "d"와의 연관성은 3열에 포함된 모든 값을 합산(1.8)함으로써 계산되고, 다른 데이터에 의한 다른 모든 데이터와 "d"와의 연관성은 3행에 포함된 모든 값을 합산(2.9)함으로써 계산될 수 있다. 즉, 연관 필터에 대해 특정 데이터와 이전에 수신한 모든 데이터와의 연관성은 4.7로 계산될 수 있다.
- [0067] 만약 특정 데이터 "d"가 복수의 필드로 분할되어 복수의 연관 필터가 생성되는 경우, 연관성 분석 장치(100)는 각각의 연관 필터에서 계산한 연관성을 모두 합산하여 특정 데이터 "d"와 이전에 수신한 모든 데이터 간의 연관성을 계산할 수 있다. 이때 복수의 연관 필터 각각은 서로 같은 상수의 가중치가 부여되거나 서로 다른 상수의 가중치가 부여될 수 있다.
- [0068] 도 11은 본 발명의 일실시예에 따른 데이터와 데이터 간의 연관성을 계산하는 방법을 도시한 도면이다.
- [0069] 특정 시점에서 특정 데이터와 특정 데이터 간의 연관성을 계산하는 방법은 다음과 같다. 예를 들어 특정 데이터 "d"와 특정 데이터 "a"와의 연관성을 분석하기 위해 각각의 데이터에 대한 연관 필터의 인덱스를 계산하고 해당 인덱스에 대응하는 연관 필터의 값 중 중첩되는 위치의 연관 필터 값을 추출할 수 있다. 도 9와 같이 특정 데이터 "d"에 대한 연관 필터의 인덱스가 3이고 특정 데이터 "a"에 대한 연관 필터의 인덱스가 0이라면, 연관성 분석 장치(100)는 두 인덱스 3과 0에 대응하는 연관 필터의 값 중 중첩되는 위치에 포함된 연관 필터의 값 0.5(3, 0) 및 0.3(0, 3)을 합산하여 두 특정 데이터 사이의 연관성으로 계산할 수 있다. 이때, 0.5(3, 0)는 특정 데이터 "d"에 의한 "d"와 "a"의 연관성을 나타내고, 0.3(0, 3)은 특정 데이터 "a"에 의한 "d"와 "a"의 연관성을 나타낼 수 있다.
- [0070] 만약 데이터 특정 "d"와 특정 데이터 "a"가 복수의 필드로 분할되어 복수의 연관 필터가 생성되는 경우, 연관성 분석 장치(100)는 각각의 연관 필터에서 계산한 연관성을 모두 합산하여 두 특정 데이터 간 연관성을 계산할 수 있다.
- [0071] 도 12는 본 발명의 일실시예에 따른 데이터 집합과 데이터 집합 간의 연관성을 계산하는 방법을 도시한 도면이다.
- [0072] 특정 시점에서 특정 데이터 집합 대 특정 데이터 집합 간의 연관성을 계산하는 방법은 다음과 같다. 예를 들어, 특정 데이터 집합 $X=\{a, b, c\}$ 와 특정 데이터 집합 $Y=\{d, e, f, g\}$ 와의 연관성은 데이터 집합에 포함된 각각의 원소 간 연관성 값을 합산하여 구할 수 있다. 즉, 데이터 집합 X의 원소 데이터 a와 데이터 집합 Y에 포함된 각각의 원소와의 연관성 $\{(a, d), (d, a), (a, e), (e, a), (a, f), (f, a), (a, g), (g, a)\}$ 와 원소 데이터 b와 데이터 집합 Y에 포함된 각각의 원소와의 연관성 $\{(b, d), (d, b), (b, e), (e, b), (b, f), (f, b), (b, g), (g, b)\}$, 원소 데이터 c와 데이터 집합 Y에 포함된 각각의 원소와의 연관성 $\{(c, d), (d, c), (c, e), (e, c),$

(c, f), (f, e), (c, g), (g, c))을 모두 합산한다.

- [0073] 도 12와 같이 데이터 집합 X의 원소 데이터 a에 대한 연관 필터의 인덱스가 3, 원소 데이터 b에 대한 연관 필터의 인덱스가 6 그리고 원소 데이터 c에 대한 연관 필터의 인덱스가 11이라고 가정하자. 그리고, 데이터 집합 Y의 원소 데이터 d에 대한 연관 필터의 인덱스가 2, 원소 데이터 e에 대한 연관 필터의 인덱스가 5, 원소 데이터 f에 대한 연관 필터의 인덱스가 7 그리고 원소 데이터 g에 대한 연관 필터의 인덱스가 9라고 가정하자. 그러면 연관성 분석 장치(100)는 데이터 집합 X의 원소 데이터들과 데이터 집합 Y의 원소 데이터들의 인덱스가 중첩되는 위치(3, 2), (2, 3), (3, 5), (5, 3), (3, 7), (7, 3), (3, 9), (9, 3), (6, 2), (2, 6), (6, 5), (5, 6), (6, 7), (7, 6), (6, 9), (9, 6), (11, 2), (2, 11), (11, 5), (5, 11), (11, 7), (7, 11), (11, 9), (9, 11))에 각각 대응하는 연관 필터의 값을 모두 합산하여 두 특정 데이터 집합 X, Y 간의 연관성으로 계산할 수 있다.
- [0074] 만약 두 특정 데이터 집합 X, Y 내의 원소 데이터들이 복수의 필드로 분할되어 복수의 연관 필터가 생성되는 경우, 연관성 분석 장치(100)는 각각의 연관 필터에서 계산한 연관성을 모두 합산하여 두 특정 데이터 집합 X, Y 간 연관성을 계산할 수 있다.
- [0075] 도 13은 본 발명의 일실시예에 따른 연관성 분석 방법을 지능형 캐싱 시스템에 적용한 예를 도시한 도면이다.
- [0076] 캐싱 시스템(1300)은 임의의 시스템에서 다루는 파일이나 콘텐츠, 네트워크 패킷, 프로세스 등을 빠르고 효율적인 처리를 위해 고성능의 데이터 저장소에 임시적으로 저장하는 시스템을 말한다. 캐싱 시스템(1300)의 효율은 캐싱 정책에 사용되는 알고리즘에 의한 캐싱 적중률(hit ratio)로 측정될 수 있다. 대부분의 캐싱 시스템(1300)에서 캐싱 교환정책으로 사용되는 알고리즘은 LRU(Least Recently Used), LFU(Least Frequently Used), FIFO(First In, First Out)등이 있다.
- [0077] 본 발명의 연관성 분석 장치(100)는 임의의 시스템에서 처리하는 데이터 간의 연관성을 실시간 분석하여 캐싱 교환정책의 알고리즘으로 활용함으로써 캐싱 성능을 향상 시킬 수 있다. 캐싱 시스템(1300)의 데이터 수신부(1310)는 캐싱 대상 즉, 데이터를 수신하면 수신된 데이터를 연관성 분석 장치(100)와 캐싱 정책부(1320)로 전송할 수 있다.
- [0078] 이후 연관성 분석 장치(100)는 데이터를 수신할 때마다 수신된 데이터와 기존에 이미 존재하는 데이터 간의 상호 연관성 정보를 연관 필터에 갱신하여 관리할 수 있다. 캐싱 정책부(1320)는 데이터를 수신하면 연관성 분석 장치(100)로 해당 데이터에 대한 연관성을 질의한다. 연관성 분석 장치(100)는 질의를 수신한 시점까지 수신한 모든 데이터와 해당 데이터의 연관성을 계산하여 캐싱 정책부(1320)로 결과를 전달할 수 있다.
- [0079] 캐싱 정책부(1320)는 저장소(1330)에 저장되어 있는 데이터들의 연관성과 해당 데이터의 연관성을 비교하여 캐싱 정책을 반영한다. 예로, 저장소(1330)에 저장되어 있는 데이터들의 최소 연관성 보다 해당 데이터의 연관성이 높으면 최소 연관성을 가진 데이터를 삭제하고 해당 데이터를 저장소(1330)에 저장한다. 만약, 해당 데이터의 연관성이 저장소에 저장되어 있는 데이터들의 최소 연관성보다 작은 경우 캐싱 정책부(1320)는 해당 데이터를 캐싱하지 않는다.
- [0080] 일례로 데이터 수신부(1310)가 단일 필드로 구성되고, URL 형태의 콘텐츠 이름을 가지는 시계열 데이터를 수신하였다고 가정하자. 그러면, 연관성 분석 장치(100)는 데이터 수신부(1310)로부터 수신된 콘텐츠 이름의 데이터를 3개의 필드(Global routing name segment, Application dependent name segment, Object dependent name segment)로 분할하는 전처리를 수행할 수 있다.
- [0081] 만약 URL 형태의 콘텐츠 이름이 "/com/blogspot/cocinaconsilvia/ almendras.html"이라면 연관성 분석 장치(100)는 아래와 같이 콘텐츠 이름을 3개의 필드로 분할할 수 있다.
- [0082] 1) 필드1(Global routing name segment): /com/blogspot
- [0083] 2) 필드2(Application dependent name segment): /cocinaconsilvia
- [0084] 3) 필드3(Object dependent name segment): /almendras.html
- [0085] 연관성 분석 장치(100)는 전처리를 통해 3개의 필드로 분할된 데이터(콘텐츠 이름)를 미리 생성된 연관 필터에 맵핑할 수 있다. 이때, 상기 데이터를 각각의 연관 필터로 맵핑하기 위한 인덱스 계산 방법은 다음과 같을 수 있다. 이때, N은 연관 필터의 크기를 말한다.
- [0086] 1) 데이터 식별 필터의 인덱스: (H("/com/blogspot/cocinaconsilvia/almendras.html")%N

- [0087] 2) 메인 연관 필터1(필드1과 필드1의 연관성 정보)의 인덱스: $H("/com/blogspot")\%N$
- [0088] 3) 메인 연관 필터2(필드2과 필드2의 연관성 정보)의 인덱스: $H("/cocinaconsilvia")\%N$
- [0089] 4) 메인 연관 필터3(필드3과 필드3의 연관성 정보)의 인덱스: $H("/almendras.html")\%N$
- [0090] 5) 횡단 연관 필터1(필드1과 필드2의 연관성 정보)의 인덱스: $H("/com/blogspot")\%N$
- [0091] 6) 횡단 연관 필터2(필드2과 필드3의 연관성 정보)의 인덱스: $H("/cocinaconsilvia")\%N$
- [0092] 7) 횡단 연관 필터3(필드3과 필드1의 연관성 정보)의 인덱스: $H("/almendras.html")\%N$
- [0093] 이후 연관성 분석 장치(100)는 생성된 연관 필터의 해당 인덱스 값들을 연관성을 증가시키는 함수를 이용하여 증가시킬 수 있다. 연관성 분석 장치(100)는 연관성을 증가시키는 함수를 통해 업데이트된 연관 필터를 이용하여 저장소(1330)에 있는 모든 데이터의 연관성을 다시 계산하여 업데이트 할 수 있다. 만약 저장소(1330)에 있는 데이터들의 최소 연관성 값과 수신된 데이터(콘텐츠 이름)의 연관성을 비교하여 저장소(1330)에 있는 데이터의 최소 연관성 값이 작을 경우, 해당 데이터를 저장소(1330)에서 삭제하고, 새로 수신된 데이터를 저장할 수 있다.
- [0094] 마찬가지로 연관성 분석 장치(100)는 캐싱 시스템(1300)에서 사전에 정의한 주기에 따라 연관성을 감소시키는 함수를 이용하여 모든 연관 필터 내의 값을 업데이트 할 수 있다. 연관성 분석 장치(100)는 연관성을 감소시키는 함수를 통해 업데이트 된 연관 필터를 이용하여 저장소(1330)에 있는 모든 데이터의 연관성을 다시 계산하여 업데이트 할 수 있다.
- [0095] 도 14는 본 발명의 일실시예에 따른 연관성 분석 방법을 장바구니 분석 시스템에 적용한 예를 도시한 도면이다.
- [0096] 장바구니 분석 시스템(1400)은 데이터의 연관성 분석이 가장 많이 활용되고 있는 시스템 중 하나이다. 기존의 연관성 분석은 통계적 기법을 기반으로 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 장바구니 안의 구매 물품 사이의 관계를 분석하고 있다. 주요 통계 지표로 지지도($P(X \cap Y)$), 신뢰도($P(Y|X) = \frac{P(X \cap Y)}{P(X)}$), 향상도($\frac{P(Y|X)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$) 등이 있다. 지지도는 전체 거래 데이터 중 품목 X와 Y를 포함하는 거래에 대한 확률을 의미하고, 신뢰도는 품목 X가 일어난 상황에서 품목 Y가 일어날 확률, 즉 X를 구매한 경우 중에서 항목 Y를 구매한 확률을 의미한다. 향상도는 품목 X를 구매한 경우 그 거래가 품목 Y를 포함하는 경우와 항목 Y가 임의로 구매되는 경우의 비율, 즉 항목 X와 Y의 구매 패턴이 독립적인지를 판단하는 기준으로 사용된다.
- [0097] 하지만 이러한 기존의 연관성 분석은 확보된 데이터의 기간 등과 같은 변수에 따라 분석결과에 큰 영향을 받을 수 있고, 경향 분석 등 실시간으로 연관성 분석을 요구하는 목적에는 사용될 수 없다. 또한, 관찰하고자 하는 품목이 증가하면 계산의 수가 크게 증가하는 단점이 있다.
- [0098] 본 발명의 연관성 분석 장치(100)를 장바구니 분석 시스템(1400)에 적용하여 실시간으로 품목간의 연관성 분석을 할 수 있다. 각 계산대에서 거래가 발생하면 장바구니 데이터를 장바구니 분석 시스템(1400)으로 전송한다. 장바구니 분석 시스템(1400)의 데이터 수신부(1410)는 장바구니 데이터를 수신하여 연관성 분석 장치(100)의 수신부(110)로 전송할 수 있다. 연관성 분석 장치(100)는 장바구니 데이터를 수신할 때마다 연관성을 증가시키는 함수를 이용하여 모든 품목들의 연관성 정보를 업데이트 할 수 있다. 또한, 연관성 분석 장치(100)는 사전에 정의한 주기에 따라 연관성을 감소시키는 함수를 이용하여 모든 품목들의 연관성 정보를 업데이트 할 수 있다.
- [0099] 연관성 분석 장치(100)는 연관성 관리부(1420)로부터 특정 품목에 대한 연관성 질의를 수신하면 질의를 수신한 시점에서 연관성을 계산하여 결과를 다시 연관성 관리부(1420)로 전달할 수 있다. 이때, 연관성 질의는 특정 품목에 대한 전체 거래 품목과의 연관성, 특정 품목 집합에 대한 전체 거래 품목과의 연관성, 특정 품목 집합에 대한 특정 품목 집합과의 연관성, 특정 단일 품목에 대한 특정 단일 품목과의 연관성 등을 질의 할 수 있다.
- [0100] 질의의 예로 도 15는 기저귀와 맥주 간의 실시간 연관성 및 우유와 세탁 세제 간의 실시간 연관성을 질의한 결과를 장바구니 분석 시스템(1400)의 연관성 관리부(1420)에서 출력한 그래프이다. 시간에 흐름에 따라 연관성의 변동을 확인 할 수 있으며 현 시점에서의 연관성을 확인할 수도 있다. 또한, 이러한 실시간 연관성 정보의 이력을 통해 연관성의 주기적 패턴을 분석할 수 있다. 예를 들어, 맥주와 기저귀가 24시간의 주기로 오후 7시에서 8시 사이에 연관성이 증가하는 형태를 보이면, 퇴근 후 아이를 가진 직장인들이 이를 구매하는 것으로 해석될 수

있으며, 이를 판매전략에 활용 할 수 있다.

- [0101] 도 16은 본 발명의 일실시예에 따른 온라인 추천 시스템을 장바구니 분석 시스템에 적용한 예를 도시한 도면이다.
- [0102] 추천 시스템(1610)은 유튜브와 같은 온라인 콘텐츠 서비스나 온라인 쇼핑몰 등에서 대중적으로 사용되고 있다. 기존의 추천 시스템은 사용자의 요청이나 구매 패턴 등을 분석하여 추천 항목을 선정하는데, 일반적으로 데이터가 누적되면 최신 경향을 제대로 반영하지 못하는 문제가 발생한다. 실 예로 유튜브에서 사용자에게 제공하는 동영상 추천 서비스는 시간이 지나 분석 데이터가 누적되면 추천 동영상이 거의 변하지 않는 것을 볼 수 있다.
- [0103] 본 발명의 연관성 분석 장치(100)를 추천 시스템(1610)에 적용하면, 사용자들의 요청 콘텐츠들에 대한 연관성을 실시간으로 분석하여 특정 사용자가 콘텐츠를 요청할 때 해당 콘텐츠와 가장 연관성 있는 콘텐츠를 추천할 수 있다. 서비스 사용자가 온라인 시스템(1600)에 콘텐츠를 요청하면 요청 콘텐츠 수신부(1620)에서 해당 요청을 수신하여 연관성 분석 장치(100)의 수신부(110)에 요청 콘텐츠를 전달한다. 요청 콘텐츠에 대한 추천 콘텐츠를 선정하기 위해 추천 시스템(1610)에서 연관성 분석 장치(100)로 해당 콘텐츠의 연관성을 질의 하면 연관성 분석 장치(100)는 해당 콘텐츠와 가장 연관성이 높은 콘텐츠 항목을 선정하여 응답할 수 있다. 추천 시스템(1610)은 추천 콘텐츠 항목을 수신하면 해당 사용자에게 추천 받은 콘텐츠들을 노출시킬 수 있다.
- [0104] 한편, 본 발명에 따른 방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성되어 마그네틱 저장매체, 광학적 판독매체, 디지털 저장매체 등 다양한 기록 매체로도 구현될 수 있다.
- [0105] 본 명세서에 설명된 각종 기술들의 구현들은 디지털 전자 회로조직으로, 또는 컴퓨터 하드웨어, 펌웨어, 소프트웨어로, 또는 그들의 조합들로 구현될 수 있다. 구현들은 데이터 처리 장치, 예를 들어 프로그램가능 프로세서, 컴퓨터, 또는 다수의 컴퓨터들의 동작에 의한 처리를 위해, 또는 이 동작을 제어하기 위해, 컴퓨터 프로그램 제품, 즉 정보 캐리어, 예를 들어 기계 판독가능 저장 장치(컴퓨터 판독가능 매체) 또는 전파 신호에서 유형적으로 구체화된 컴퓨터 프로그램으로서 구현될 수 있다. 상술한 컴퓨터 프로그램(들)과 같은 컴퓨터 프로그램은 컴파일된 또는 인터프리트된 언어들을 포함하는 임의의 형태의 프로그래밍 언어로 기록될 수 있고, 독립형 프로그램으로서 또는 모듈, 구성요소, 서브루틴, 또는 컴퓨팅 환경에서의 사용에 적절한 다른 유닛으로서 포함하는 임의의 형태로 전개될 수 있다. 컴퓨터 프로그램은 하나의 사이트에서 하나의 컴퓨터 또는 다수의 컴퓨터들 상에서 처리되도록 또는 다수의 사이트들에 걸쳐 분배되고 통신 네트워크에 의해 상호 연결되도록 전개될 수 있다.
- [0106] 컴퓨터 프로그램의 처리에 적절한 프로세서들은 예로서, 범용 및 특수 목적 마이크로프로세서들 둘 다, 및 임의의 종류의 디지털 컴퓨터의 임의의 하나 이상의 프로세서들을 포함한다. 일반적으로, 프로세서는 판독 전용 메모리 또는 랜덤 액세스 메모리 또는 둘 다로부터 명령어들 및 데이터를 수신할 것이다. 컴퓨터의 요소들은 명령어들을 실행하는 적어도 하나의 프로세서 및 명령어들 및 데이터를 저장하는 하나 이상의 메모리 장치들을 포함할 수 있다. 일반적으로, 컴퓨터는 데이터를 저장하는 하나 이상의 대량 저장 장치들, 예를 들어 자기, 자기-광 디스크들, 또는 광 디스크들을 포함할 수 있거나, 이것들로부터 데이터를 수신하거나 이것들에 데이터를 송신하거나 또는 양쪽으로 되도록 결합될 수도 있다. 컴퓨터 프로그램 명령어들 및 데이터를 구체화하는데 적절한 정보 캐리어들은 예로서 반도체 메모리 장치들, 예를 들어, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 롬(ROM, Read Only Memory), 램(RAM, Random Access Memory), 플래시 메모리, EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 등을 포함한다. 프로세서 및 메모리는 특수 목적 논리 회로조직에 의해 보충되거나, 이에 포함될 수 있다.
- [0107] 또한, 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용매체일 수 있고, 컴퓨터 저장매체 및 전송매체를 모두 포함할 수 있다.
- [0108] 본 명세서는 다수의 특정한 구현물의 세부사항들을 포함하지만, 이들은 어떠한 발명이나 청구 가능한 것의 범위에 대해서도 제한적인 것으로서 이해되어서는 안되며, 오히려 특정한 발명의 특정한 실시형태에 특유할 수 있는 특징들에 대한 설명으로서 이해되어야 한다. 개별적인 실시형태의 문맥에서 본 명세서에 기술된 특정한 특징들은 단일 실시형태에서 조합하여 구현될 수도 있다. 반대로, 단일 실시형태의 문맥에서 기술한 다양한 특징들 역시 개별적으로 혹은 어떠한 적절한 하위 조합으로도 복수의 실시형태에서 구현 가능하다. 나아가, 특징들이 특정한 조합으로 동작하고 초기에 그와 같이 청구된 바와 같이 묘사될 수 있지만, 청구된 조합으로부터의 하나 이상의 특징들은 일부 경우에 그 조합으로부터 배제될 수 있으며, 그 청구된 조합은 하위 조합이나 하위 조합의

변형물로 변경될 수 있다.

[0109] 마찬가지로, 특정한 순서로 도면에서 동작들을 묘사하고 있지만, 이는 바람직한 결과를 얻기 위하여 도시된 그 특정한 순서나 순차적인 순서대로 그러한 동작들을 수행하여야 한다거나 모든 도시된 동작들이 수행되어야 하는 것으로 이해되어서는 안 된다. 특정한 경우, 멀티태스킹과 병렬 프로세싱이 유리할 수 있다. 또한, 상술한 실시 형태의 다양한 장치 컴포넌트의 분리는 그러한 분리를 모든 실시형태에서 요구하는 것으로 이해되어서는 안되며, 설명한 프로그램 컴포넌트와 장치들은 일반적으로 단일의 소프트웨어 제품으로 함께 통합되거나 다중 소프트웨어 제품에 패키징 될 수 있다는 점을 이해하여야 한다.

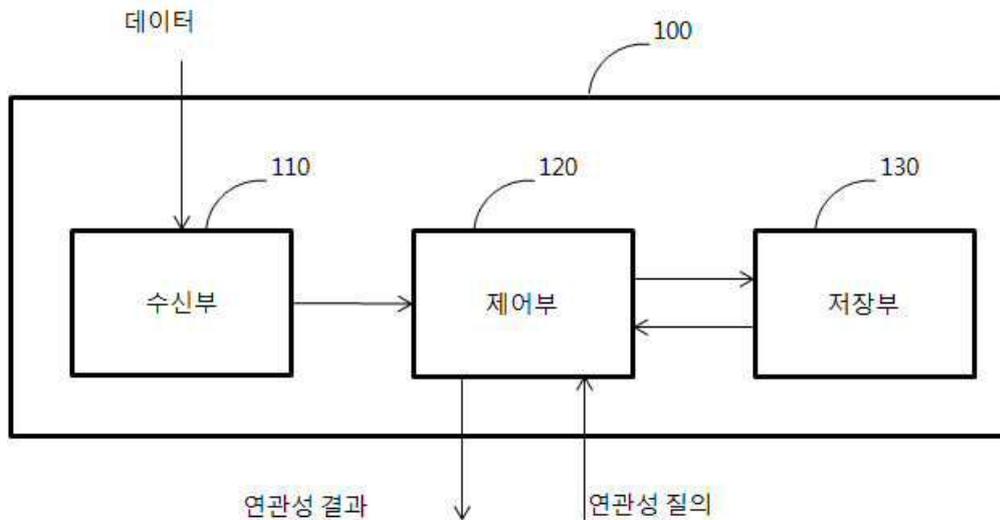
[0110] 한편, 본 명세서와 도면에 개시된 본 발명의 실시 예들은 이해를 돕기 위해 특정 예를 제시한 것에 지나지 않으며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예들 이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 자명한 것이다.

부호의 설명

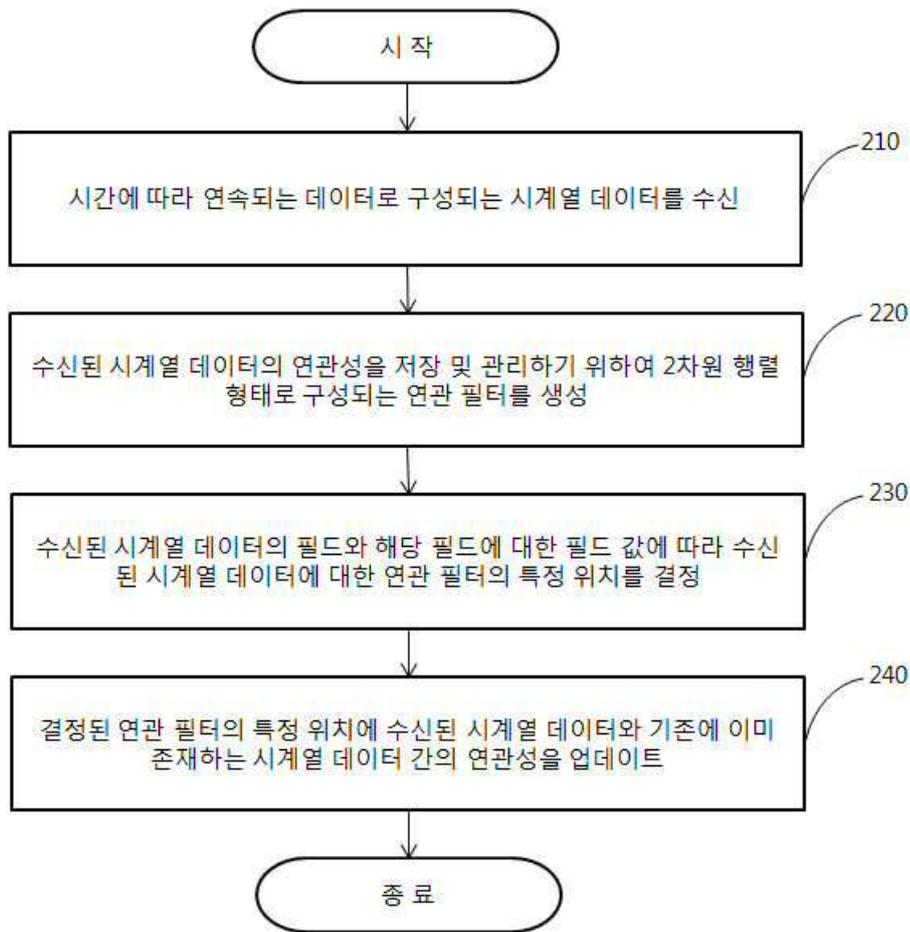
- [0111] 100 : 연관성 분석 장치
- 110 : 수신부
- 120 : 제어부
- 130 : 제어부

도면

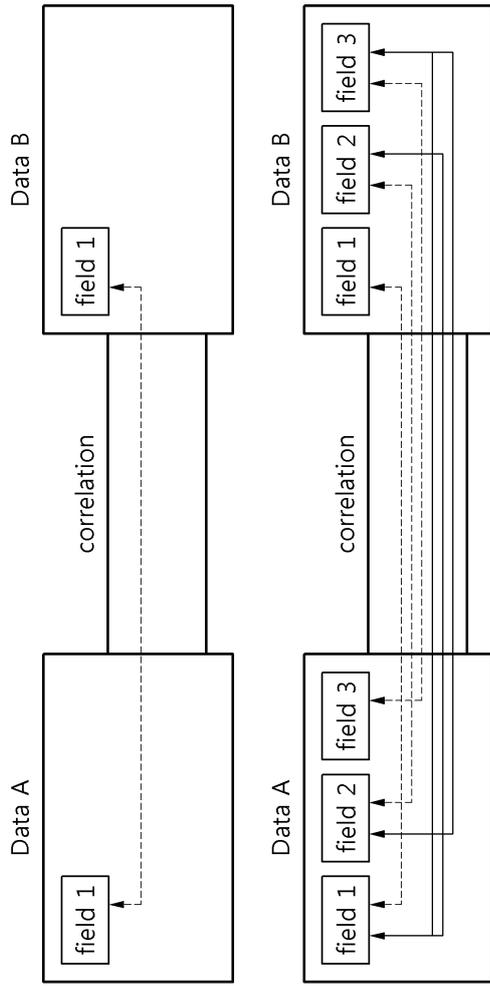
도면1



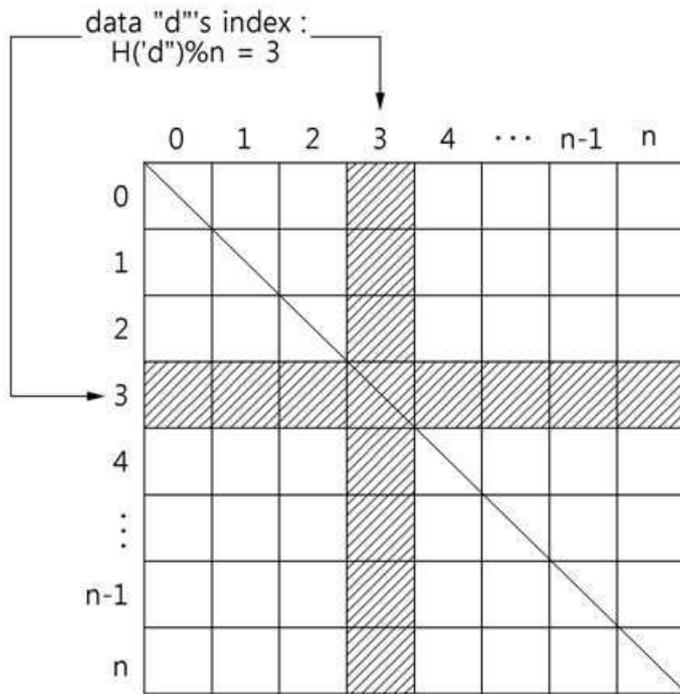
도면2



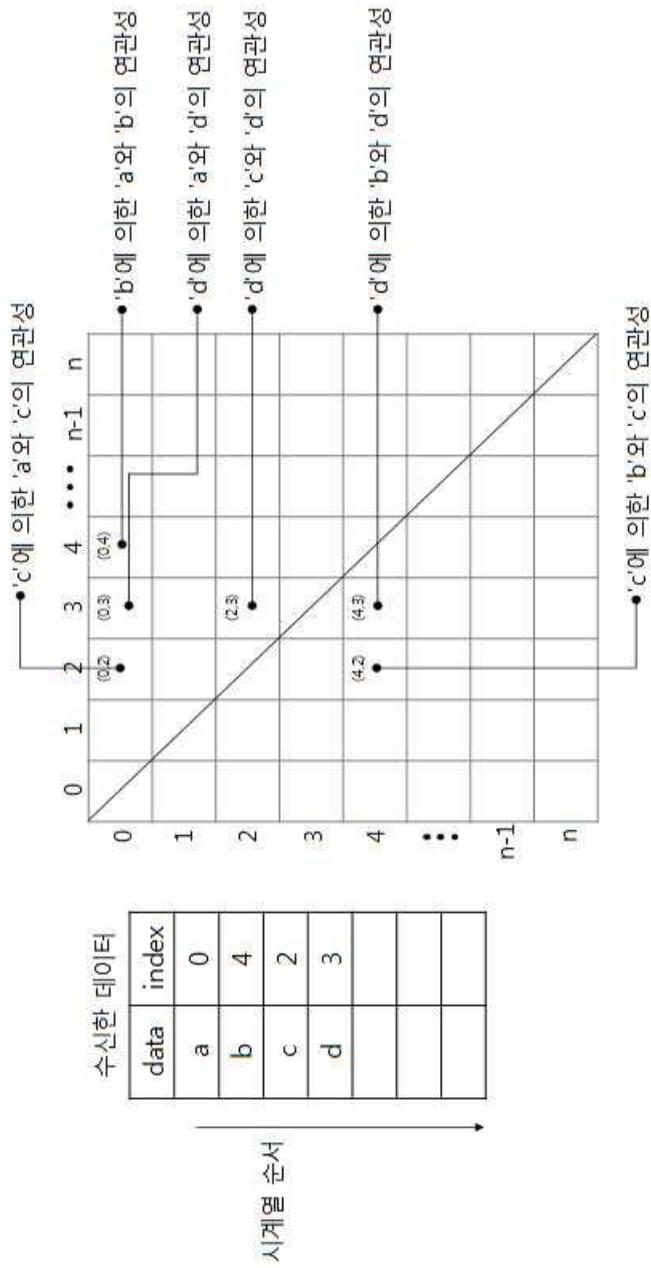
도면3



도면4



도면5

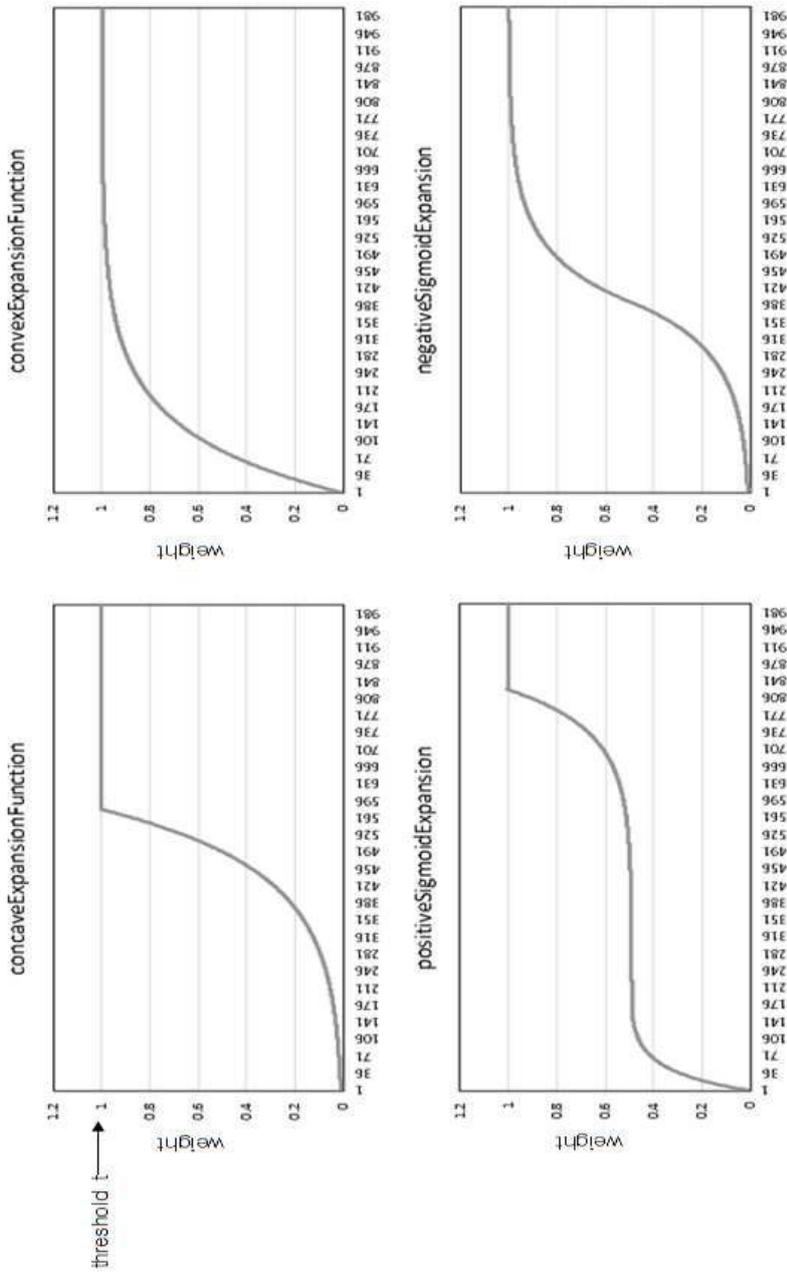


수신한 데이터

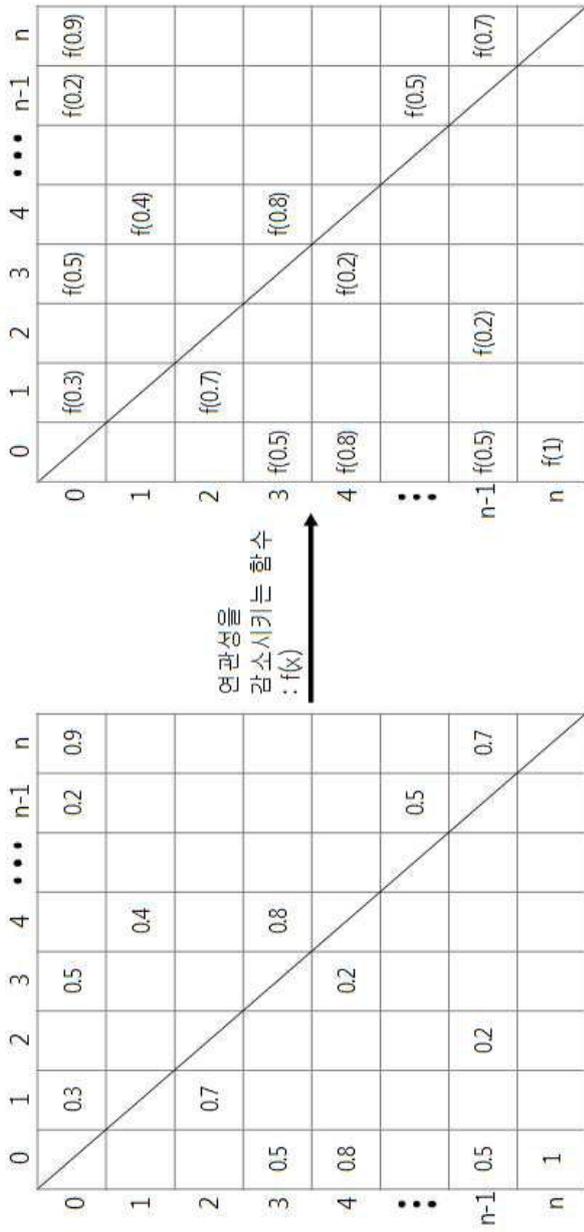
| data | index |
|------|-------|
| a | 0 |
| b | 4 |
| c | 2 |
| d | 3 |
| | |
| | |
| | |

시계열 순서

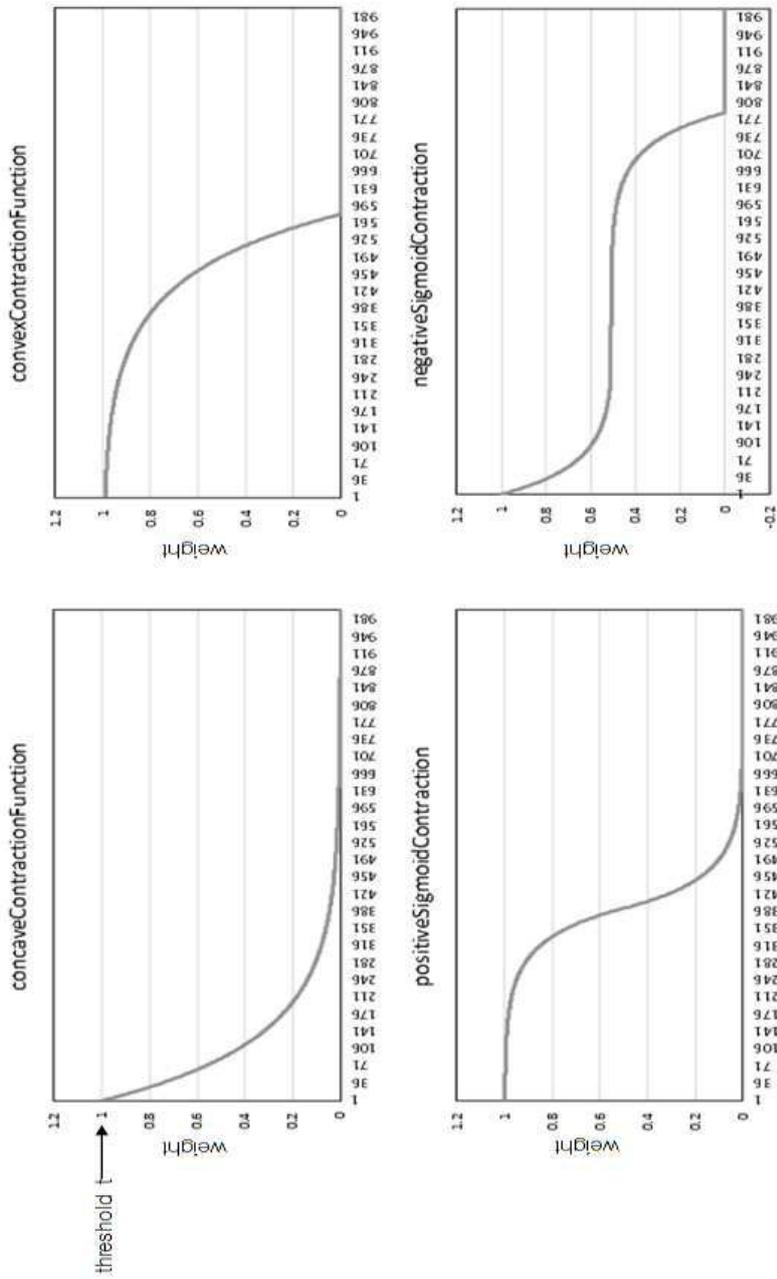
도면6



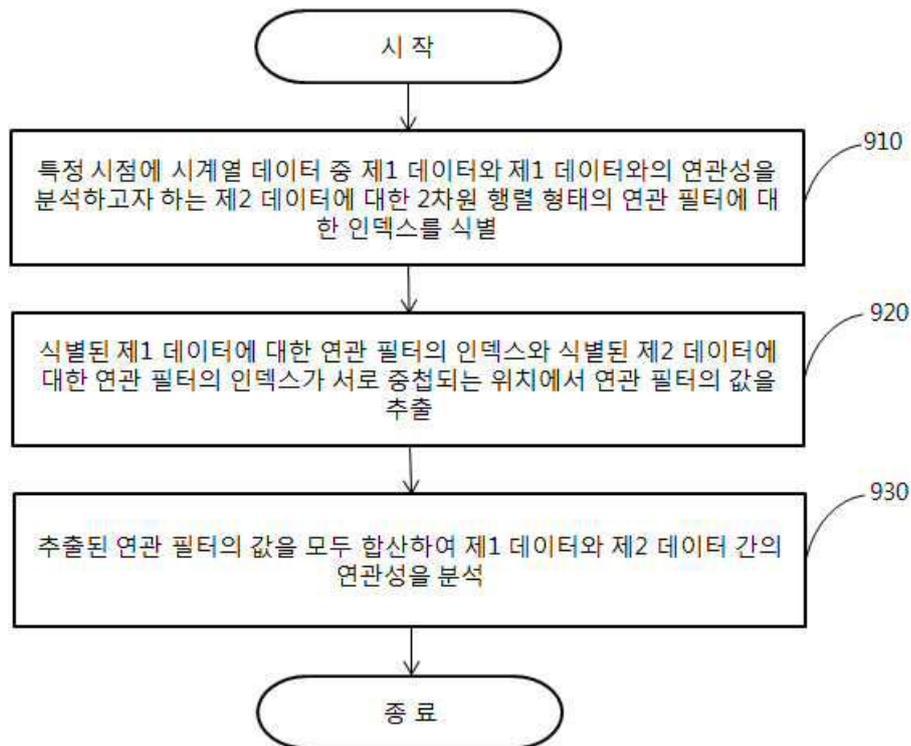
도면7



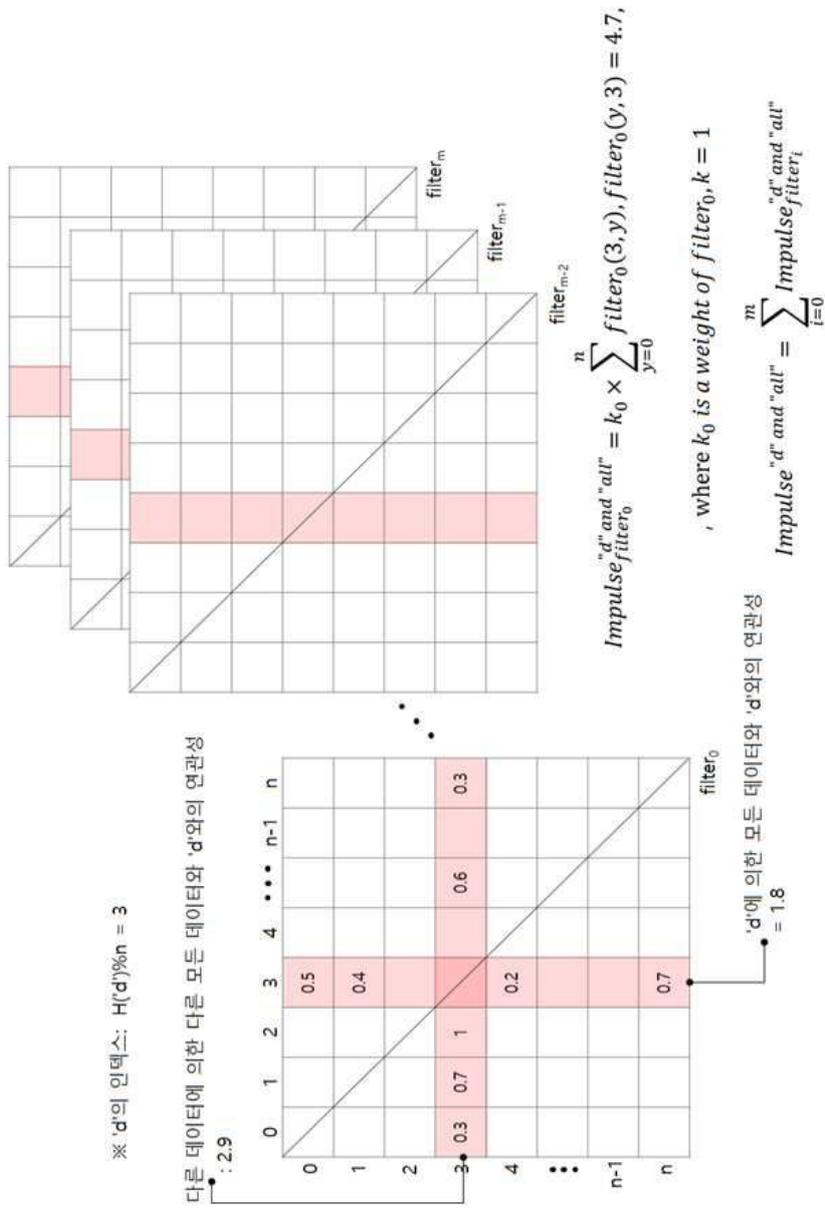
도면8



도면9

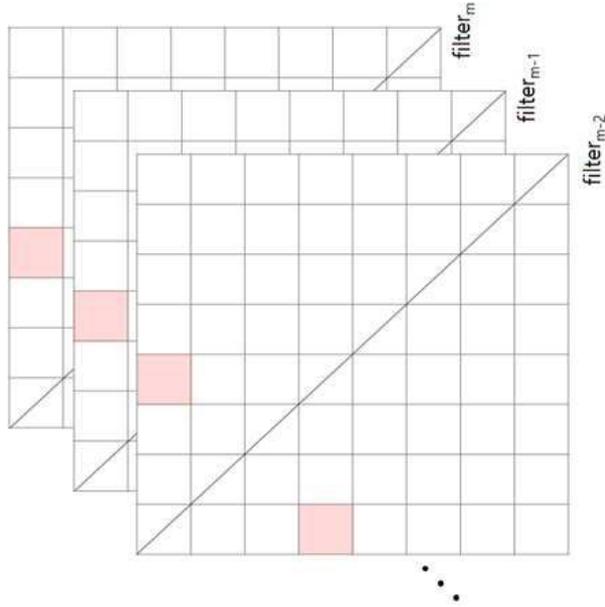
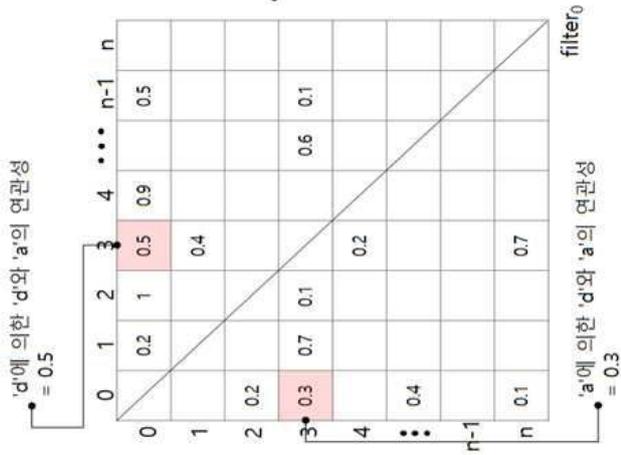


도면10



도면11

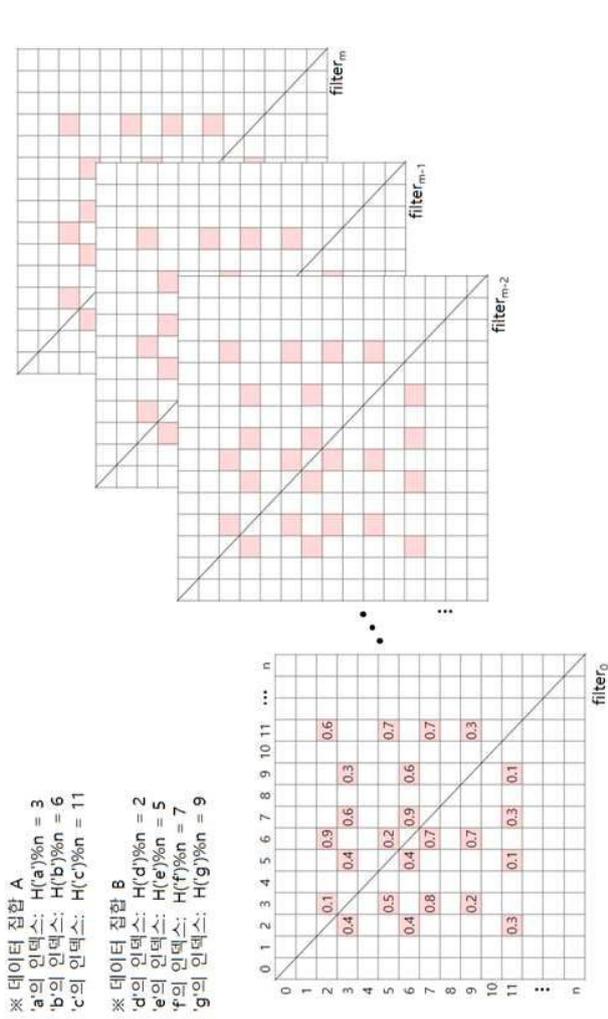
※ 'd'의 인덱스: $H('d')\%n = 3$
 ※ 'a'의 인덱스: $H('a')\%n = 0$



$Impulse_{filter_0}^{d \text{ and } a} = k_0 \times (filter_0(0, 3) + filter_0(3, 0)) = 0.8$
 , where k_0 is a weight of $filter_0, k = 1$

$$Impulse^{d \text{ and } a} = \sum_{i=0}^m Impulse_{filter_i}^{d \text{ and } a}$$

도면12

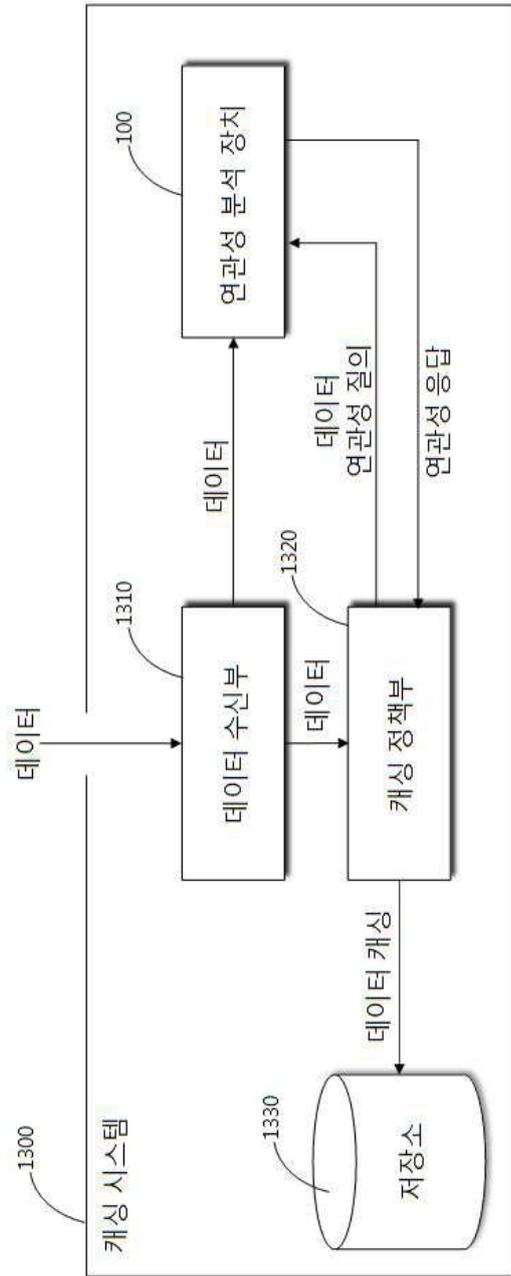


$$Impulse_{x\text{ and }y}^{x\text{ and }y} = k_0 \times \left\{ \begin{array}{l} filter_0(3, 2) + filter_0(2, 3) + filter_0(6, 2) + filter_0(2, 6) + filter_0(11, 2) + filter_0(2, 11) \\ + filter_0(3, 5) + filter_0(5, 3) + filter_0(6, 5) + filter_0(5, 6) + filter_0(11, 5) + filter_0(5, 11) \\ + filter_0(3, 7) + filter_0(7, 3) + filter_0(6, 7) + filter_0(7, 6) + filter_0(11, 7) + filter_0(7, 11) \\ + filter_0(3, 9) + filter_0(9, 3) + filter_0(6, 9) + filter_0(9, 6) + filter_0(11, 9) + filter_0(9, 11) \end{array} \right\} = 11.2$$

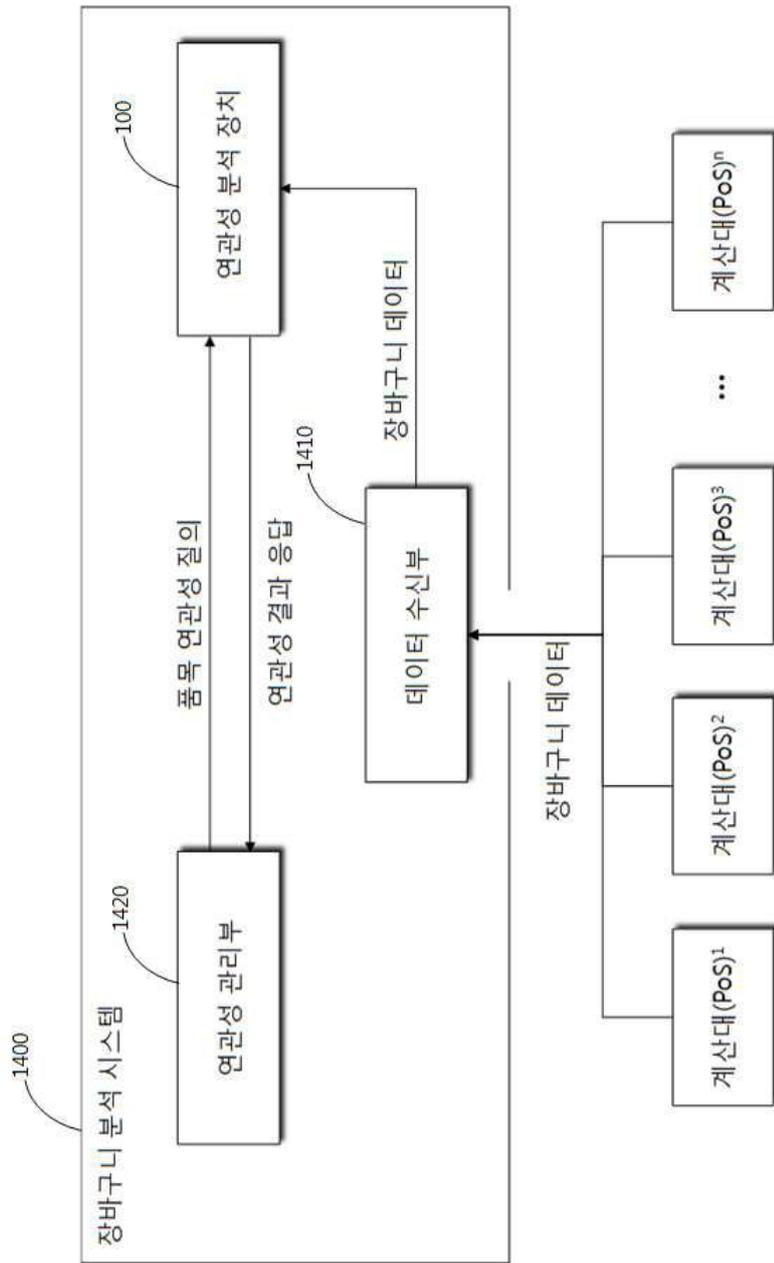
where k_0 is a weight of $filter_0, k = 1$

$$Impulse_{x\text{ and }y}^{x\text{ and }y} = \sum_{i=0}^m Impulse_{filter_i}^{x\text{ and }y}$$

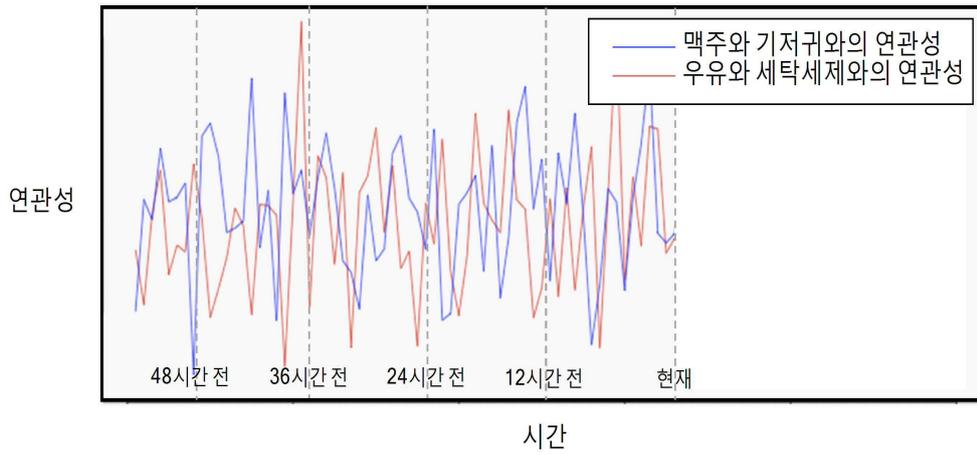
도면13



도면14



도면15



도면16

