(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0047862
(43) 공개일자 2024년04월12일

(51) 국제특허분류(Int. Cl.)
G06F 18/00 (2023.01)
(52) CPC특허분류
G06F 18/241 (2023.01)
G06F 18/214 (2023.01)
(21) 출원번호 10-2022-0127446
(22) 출원일자 2022년10월05일
심사청구일자 2022년10월05일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
이정혜
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
강혜원
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
김재호
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
(74) 대리인
특허법인 수

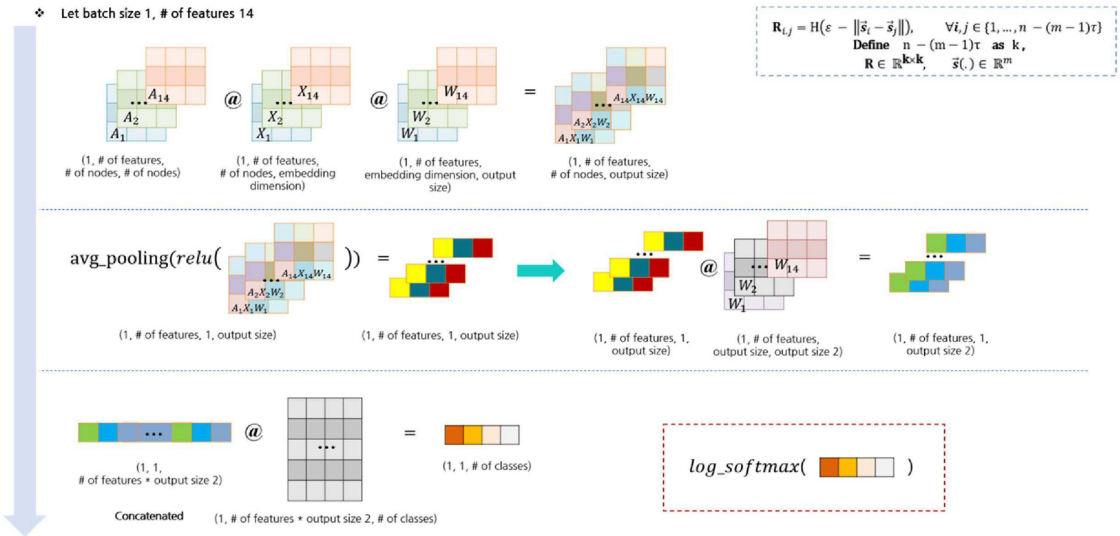
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **시계열 데이터 분류를 위한 방법 및 이를 이용한 센싱 장치**

(57) 요 약

본 발명에 따른 시계열 데이터 분류 방법은, 다변량 시계열 데이터를 획득하는 제1 단계, 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성하는 제2 단계, 생성된 RP를 GCN 알고리즘에 적용시키는 제3 단계 및 GCN 알고리즘을 거쳐 나온 임베딩 벡터들을 바탕으로 분류 태스크를 시행하는 제4 단계를 포함한다.

대 표 도 - 도7



이 발명을 지원한 국가연구개발사업
과제고유번호 1711165204
과제번호 2020R1C1C1011063
부처명 과학기술정보통신부(P71)
과제관리(전문)기관명 한국연구재단(ABX3869)
연구사업명 개인기초연구(과기정통부)(1621110150155120012343O1)
연구과제명 딥러닝 기반의 개인정보보호가 보장되는 인공지능 연합학습 플랫폼 개발
기 여 율 35/100
과제수행기관명 울산과학기술원
연구기간 2022.03.01 ~ 2023.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1485018735
과제번호 ARQ202101403002
부처명 환경부(P23)
과제관리(전문)기관명 한국환경산업기술원(ABX2873)
연구사업명 환경보건 디지털 조사 기반 구축 기술개발사업(B050)
연구과제명 GIS 기반 빅데이터 및 기계학습을 활용한 지역단위 환경보건 감시체계 평가 시스템 개발
기 여 율 50/100
과제수행기관명 울산과학기술원
연구기간 2022.01.01 ~ 2022.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711152473
과제번호 2020-0-01336-003
부처명 과학기술정보통신부(P71)
과제관리(전문)기관명 정보통신기획평가원(ACS6921)
연구사업명 정보통신방송혁신인재양성(1625564130133210021333O4)
연구과제명 인공지능대학원지원(울산과학기술원)
기 여 율 15/100
과제수행기관명 울산과학기술원
연구기간 2022.01.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

다변량 시계열 데이터를 획득하는 제1 단계;

상기 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성하는 제2 단계;

상기 생성된 RP를 GCN 알고리즘에 적용시키는 제3 단계; 및

상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들을 바탕으로 분류 태스크를 시행하는 제4 단계;

를 포함하는 것을 특징으로 하는 시계열 데이터 분류 방법.

청구항 2

제1 항에 있어서,

상기 제2 단계에서, 상기 획득되는 RP의 수는 상기 다변량 시계열 데이터의 변수의 수에 대응되도록 생성되는 것을 특징으로 하는 시계열 데이터 분류 방법.

청구항 3

제1 항에 있어서,

상기 제4 단계에서, 상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들은 풀리-커넥디트-레이어를 적용하여 분류 태스크를 시행하는 것을 특징으로 하는 시계열 데이터 분류 방법.

청구항 4

제1 항에 있어서,

상기 제3 단계에서, 생성된 RP를 인접행렬로 변경하고 및 상기 시계열 데이터로부터 노드 피쳐를 생성하여, 상기 인접 행렬 및 상기 노드 피쳐를 상기 GCN 알고리즘에 적용시키는 것을 특징으로 하는 시계열 데이터 분류 방법.

청구항 5

다변량 시계열 데이터를 획득하는 제1 단계;

상기 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성하는 제2 단계;

상기 생성된 RP를 GCN 알고리즘에 적용시키는 제3 단계; 및

상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들을 바탕으로 분류 태스크를 시행하는 제4 단계;

를 포함하는 것을 특징으로 하는 시계열 데이터 분류 방법을 이용한 센싱 장치.

청구항 6

제5 항에 있어서,

상기 제2 단계에서, 상기 획득되는 RP의 수는 상기 다변량 시계열 데이터의 변수의 수에 대응되도록 생성되는 것을 특징으로 하는 센싱 장치.

청구항 7

제5 항에 있어서,

상기 제4 단계에서, 상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들은 풀리-커넥디트-레이어를 적용하여 분류 태스크를 시행하는 것을 특징으로 하는 센싱 장치.

청구항 8

제5 항에 있어서,

상기 제3 단계에서, 생성된 RP를 인접행렬로 변경하고 및 상기 시계열 데이터로부터 노드 피쳐를 생성하여, 상기 인접 행렬 및 상기 노드 피쳐를 상기 GCN 알고리즘에 적용시키는 것을 특징으로 하는 센싱 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 실시간으로 수집되는 시계열 데이터를 분류하기 위한 방법 및 이를 이용한 센싱 장치에 관한 것이다.

배 경 기 술

[0003] 시계열 데이터 분류 방법은 매우 다양한 분야에서 활용된다. 예를 들어, 생체 신호(Biomedical signal)을 다루는 분야에서는 EEG(뇌파), ECG(심전도) 데이터를 가지고 환자의 상태를 정상 또는 비정상으로 분류하기 위한 시계열 분류 모델 연구가 활발히 일어나고 있다.

[0004] 또한, 사이버 보안(Cyber security) 분야에서는 실시간 네트워크 패킷 데이터를 통해서 해당 네트워크 패킷이 악의적 공격에 대한 것인지, 또는 무해한 것인 것인지 분류하는 모델이 개발되고 있다.

[0005] 또한, 사용자 행동 인식(Human Activity Recognition) 분야는 실시간 센서 데이터를 통해 사용자의 움직임을 인식하고 분류하는데, 예를 들어, 사용자가 스마트 워치를 착용하고 있을 때, 인공지능 모델을 통해 사용자가 스쿼트, 수영, 걷기, 뛰기 등의 활동 중에 어떤 운동 자세를 취하고 있는지를 분류할 수 있는 연구가 진행되고 있다.

[0006] 이를 위해 기존에는 의사 결정나무(Decision tree), 로지스틱 리그레션(logistic regression) 등의 전통적인 머신러닝 알고리즘, 립러닝을 활용하여 시계열 분류를 하는 방법을 이용하여 왔다.

[0007] 그러나 머신 러닝은 딥 러닝의 이점을 살리지 못하고 시계열 도메인에 따라 피처(feature) 선택이나 추출 기법이 잘 통하지 않는 문제가 있다.

[0008] 한편, 최근에는 시계열 데이터를 RP(recurrence plot)와 같은 이미지 형태로 변환한 후 이를 CNN 에 적용하는 연구가 개발되고 있다. 이러한 방법은 다른 이미지 방법론보다 좋은 결과를 낸다는 보고가 있다. 그러나 이러한 이미징 방법은 연산 비용(computation cost)이 많이 든다는 단점이 존재한다.

발명의 내용

해결하려는 과제

[0010] 본 발명의 목적은 시계열 데이터의 분류 성능을 높이면서도 정보 손실이 적은 시계열 데이터 분류 방법을 제공

하는 것이다.

[0011] 또한 본 발명의 목적은 멀리 있는 데이터 포인트 간의 관계성도 정확히 학습할 수 있는 시계열 데이터 분류 방법을 제공하는 것이다.

과제의 해결 수단

[0013] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 시계열 데이터 분류 방법은, 다변량 시계열 데이터를 획득하는 제1 단계; 상기 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성하는 제2 단계; 상기 생성된 RP를 GCN 알고리즘에 적용시키는 제3 단계; 및 상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들을 바탕으로 분류 태스크를 시행하는 제4 단계;를 포함한다.

[0014] 일 실시예에서, 상기 제2 단계에서, 상기 획득되는 RP의 수는 상기 다변량 시계열 데이터의 변수의 수에 대응되도록 생성된다.

[0015] 일 실시예에서, 상기 제4 단계에서, 상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들은 풀리-커넥디트-레이어를 적용하여 분류 태스크를 시행한다.

[0016] 일 실시예에서, 상기 제3 단계에서, 생성된 RP를 인접행렬로 변경하고 및 상기 시계열 데이터로부터 노드 피처를 생성하여, 상기 인접 행렬 및 상기 노드 피처를 상기 GCN 알고리즘에 적용시킨다.

[0017] 본 발명의 다른 실시 형태에 따른 센싱 장치는, 다변량 시계열 데이터를 획득하는 제1 단계; 상기 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성하는 제2 단계; 상기 생성된 RP를 GCN 알고리즘에 적용시키는 제3 단계; 및 상기 GCN 알고리즘을 거쳐 나온 임베딩 벡터들을 바탕으로 분류 태스크를 시행하는 제4 단계; 를 포함하는 것을 특징으로 하는 시계열 데이터 분류 방법을 이용한다.

발명의 효과

[0019] 이와 같이 본 발명에 따르면, 다음과 같은 효과가 있다.

[0020] 본 발명에 따른 시계열 데이터 분류 방법은 RP 기반에 GCN을 적용하여 시계열 데이터의 분류 성능을 높이면서도 정보 손실이 적은 효과가 있다.

[0021] 또한 본 발명에 따른 시계열 데이터 분류 방법 및 센싱 장치는 사진과 달리 시계열 데이터의 경우 멀리 떨어진 데이터 포인트 간의 상관성도 제대로 잡아 멀리 있는 데이터 포인트 간의 관계성도 정확히 학습할 수 있는 효과가 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명에 따른 시계열 데이터 분류 방법을 개략적으로 나타낸 흐름도이다.

도 2는 시계열 데이터로부터 RP를 생성하여 이미지화 하는 예를 나타낸다.

도 3 및 도 4는 RP를 생성하는 예를 나타낸다.

도 5 및 도 6은 본 발명에 따른 시계열 데이터 분류 방법에서 RP를 생성하고 이를 GCN에 적용하는 모습을 예시적으로 도시화한 것이다.

도 7은 본 발명에 따른 시계열 데이터 분류 방법에서의 RP를 GCN 알고리즘에 적용하는 예를 도시적으로 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0024] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있

는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0025] 이하, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 하기 위하여, 본 발명의 바람직한 실시예들에 관하여 첨부된 도면을 참조하여 상세히 설명하기로 한다.

[0026] 도 1은 본 발명에 따른 시계열 데이터 분류 방법을 개략적으로 나타낸 흐름도이다.

[0027] 도 1을 참조하면 본 발명의 일 실시예에 따른 시계열 데이터 분류 방법은 다변량 시계열 데이터 획득 단계(S110), RP 생성 단계(S120), GCN 알고리즘 적용 단계(S130) 및 풀리 커넥티드 레이어를 적용하여 분류 태스크 적용 단계(S140)를 포함한다.

[0028] 먼저 본 발명에 따른 다변량 시계열 데이터 분류 방법은 분류할 대상이 되는 다변량 시계열 데이터를 획득 한다(S110). 일 실시예에서, 센싱 장치(미도시)에서 주기적으로 측정되는 복수의 센서로부터 복수의 M개 시계열 데이터를 획득한다.

[0029] 그런 다음 다변량 시계열 데이터로부터 RP(Recurrence Plot)을 생성한다(S120).

[0030] RP(Recurrence plot)는 시계열 데이터를 이미지로 표현하는 방법 중의 하나이다. 에크만 등이 만든 RP는 시계열을 생성하는 비선형 동적 시스템의 대표적인 현상인 주기성 또는 불규칙한 순환성과 같은 반복 패턴을 시각화하는 도구이다. RP는 원시 시계열 데이터에서 새로 생성된 벡터 간의 쌍별 관계에 초점을 맞춰 데이터의 숨겨진 패턴을 포착할 수 있다.

[0031] 도 2는 시계열 데이터로부터 RP를 생성하여 이미지화하는 예를 나타낸다.

[0032] 도 2를 참조하면, RP는 시계열 데이터를 m 차원의 공간 궤적에, 그리고 각 공간 궤적에 위치한 데이터 포인트 간의 거리를 이용하여 시계열 데이터를 이미지 이미지화한 것이다.

[0033] 도 3 및 도 4는 RP를 생성하는 예를 나타낸다.

[0034] 도 3을 참조하면, 좌측에 표시바와 같이 1부터 10까지의 타임 스텝과 이에 대응하는 센서 데이터를 포함하는 시계열 데이터를 획득하면, 이를 3개의 칸들로 이루어진 3차원 공간 벡터를 생성하게 된다. 이때, 새로운 데이터를 생성하기 위해서는 임의의 타임 스텝에 대응되는 시계열 데이터를 골라야 한다. 즉, 몇 개의 타임 스텝에 대응되는 시계열 데이터를 고를 것인지, 몇 칸을 띄워서 타임 스텝을 고를 것인지를 사용자가 임의로 골라야 한다.

[0035] 도 3에 도시된 우측 3차원 벡터 값들은 3개의 타임 스텝을 3칸 띄우고 골라서 그에 대응되는 시계열 데이터 데이터를 고른 것이다. 새롭게 고른 데이터 포인트들로 이루어진 집합은 각각 새로운 차원의 데이터 포인트로 간주된다. 도 3에서는 총 4개의 벡터, 즉 데이터 포인트가 만들어진다. 즉, 도 3의 좌측에서 1변량 데이터가 10개 있던 것을 우측과 같이 3차원 데이터 포인트가 4개 있는 것으로 변경된다.

[0036] 그런 다음 도 4에 도시된 바와 같이 도 3에서 구한 3차원 데이터 포인트 4개를 각각 페어-와이즈(pair-wise) 하게 거리를 구한다. 도 4는 각 거리 값을 나타낸 것이며, 이는 1과 0으로 이루어진 행렬이 된다.

[0037] 이 단계에서는, 변수 별로 RP를 생성하게 되므로, 만약 변수가 5개라면, 5 장의 RP 이미지가 생성된다. 즉, 상기 획득되는 RP의 수는 상기 다변량 시계열 데이터의 변수의 수에 대응되도록 생성된다.

[0038] 다시 도 3 및 도 4를 참조하면, 변수 별로 RP를 생성하는데, RP를 생성하는 과정은 기존의 공지의 방법을 이용할 수 있다. 다만, 이 RP를 해석하는 관점을 달리할 수 있다. 즉, 본 발명에 따른 시계열 데이터 분류 방법의 RP는 새롭게 만들어낸 데이터포인트 간의 Pair-wise 한 거리값으로 이루어진 행렬을 나타낸다.

[0039] 이러한 RP는 그래프(인접행렬 ; adjacency matrix)로 해석될 수 있다. 또한, 이 그래프는 노드(node)와 엣지(edge)로 이루어져 있다. 노드 간의 거리가 일정 이상 가까우면 엣지를 통해 두 노드가 연결될 수 있다. 그리고 노드 간의 거리가 멀면 연결이 되지 않는다.

[0040] 도 3 및 도 4에 도시된 예시를 보면, 새롭게 구한 세 개의 타임 스탭으로 이루어진 벡터들을 노드(node) 라고

볼 수 있다. 또한, RP를 이루는 값들은 각 노드 간의 거리로 해석할 수 있으므로, 값이 1이면 노드 간의 연결이 일어난 상태, 값이 0이면 노드 간의 연결성이 없다고 해석할 수 있다.

[0041] GCN 알고리즘을 사용하기 위해서는 그래프(인접행렬) 뿐 아니라, 각 노드 별로 부여되는 추가적인 정보(node feature)가 필요하다. 각 노드는 세 개의 타임 스탭으로 이루어진 벡터이다. 그러므로, 각 노드들은 세 개의 타임 스탭에 대응되는 센서값으로 이루어진 벡터 값과 대응된다고 볼 수 있다. 따라서, 세 개의 타임 스탭에 대응되는 센서값으로 이루어진 벡터를 추가적인 정보인 노드 피쳐(node feature)라고 간주할 수 있다.

[0042] 한편, 이러한 RP는 기존에 흔히 CNN과 함께 쓰여, 시계열 분류 성능을 높여왔다. 그러나 이러한 RP 기반 CNN 모델에는 다음과 같은 한계점이 존재한다.

[0043] 먼저 CNN 기반 모델에는 정보를 상당히 많이 잃을 수 있다. 원래 빼곡히 숫자로 채워져 있던 일변량 시계열 데이터를 RP로 변환하면 0과 1만 남게 되기 때문에 이 과정에서 너무 많은 정보를 잃게 된다.

[0044] 또한, 기존의 CNN 기반 모델에서는, 멀리 있는 데이터 포인트 간의 관계성을 제대로 학습하기 힘들다는 한계가 존재한다.

[0045] CNN은 본래 '사진'과 같은 데이터의 특징을 학습하기 위해 개발된 알고리즘으로, 이웃한 데이터포인트 간의 상관성을 잘 학습하는데 그 장점이 있다. 그러나, 시계열 데이터는 '사진'과 달리 멀리 떨어져 있는 데이터포인트 간의 상관성도 유의해야 한다. 태생이 사진이 아닌 시계열 데이터가 변환된 RP 역시, 멀리 떨어져 있는 데이터 포인트 간의 상관성도 볼 필요가 있는데, CNN을 RP에 바로 적용해 버린다면, 이웃한 데이터포인트 간의 상관성 위주로 학습하기 때문에, 시계열 데이터 기반의 RP의 특징을 제대로 잡아내기 어려울 수 있기 때문이다.

[0046] 이러한 단점을 해결하기 위해 본 발명에서는 생성된 RP를 GCN 알고리즘에 적용시키는 단계를 적용한다(S130).

[0047] 본 발명에서는 RP를 시간 단계 사이의 인접 행렬로 간주하여 그래프 컨볼루션 네트워크(GCN)가 CNN을 대체할 수 있게 한다. GCN을 통해 노드 특징 매트릭스에 해당하는 RP(인접 행렬)와 원시 시계열 데이터를 모두 GCN 레이어 훈련에 사용할 수 있어 '대량 데이터 손실'을 억제할 수 있게 하는 것이다.

[0048] 아울러, GCN은 먼 시간 단계 사이의 유사성과 관계를 고려할 수 있게 한다.

[0049] 도 5 및 도 6은 본 발명에 따른 시계열 데이터 분류 방법에서 RP를 생성하고 이를 GCN에 적용하는 모습을 예시적으로 도시화한 것이다.

[0050] 도 5를 참조하면, 본 발명은 인접 행렬로 RP를 생성하고 원시 시계열 데이터로부터 노드 피쳐를 만든다. 그리고 나서 생성된 RP를 인접행렬로 변경하고 및 상기 시계열 데이터로부터 노드 피쳐를 생성하여, 상기 인접 행렬 및 상기 노드 피쳐를 상기 GCN 알고리즘에 적용시킨다.

[0051] 예를 들어, 도 6의 구간 1과 같이, 샘플의 각 형상에 대한 RP plot(인접 행렬)과 이에 대응하는 노드 피쳐 매트릭스(node feature matrix)를 생성한 후, 인접 행렬 및 노드 피쳐를 상기 GCN 알고리즘에 적용하는 것이다.

[0052] 그리고 도 6의 구간 2와 같이, 각 인접 행렬을 GCN을 통해 임베딩 벡터를 추출하고, 분류를 위해 최종 완전 연결 레이어의 피쳐에서 임베딩을 연결한다. 즉, RP별로 GCN 알고리즘을 적용시켜 나온 임베딩을 최종적으로 이어서(concatenate) 풀리 커넥티드 레이어(fully-connected layer)를 통해 분류 태스크를 시행한다(S140).

[0053] 도 7은 본 발명에 따른 시계열 데이터 분류 방법에서의 RP를 GCN 알고리즘에 적용하는 예를 도시적으로 나타낸다.

[0054] 도 5 및 도 6에 도시된 바와 같이, RP를 인접행렬 즉, 그래프 및 노드 피쳐(Node feature) 로 바꾸고, 도 7에서와 같이 GCN 알고리즘을 적용하게 된다.

[0055] 도 7의 예에서 배치 크기(batch size)는 1로 하고, 피쳐의 개수는 14로 한 예를 나타내었다. 각 변수 별로 RP 그래프 및 노드 피쳐 로 변환하여, GCN 연산을 적용하면, 각각의 RP 및 노드 피쳐는 하나의 벡터로 변환된다. 그리고 나서 각 변수를 GCN 알고리즘을 적용하여 만든 이 벡터들을 이어붙여서, 풀리 커넥티드 레이어(fully-connected layer)를 적용함으로써, 분류(classification) 태스크를 수행할 수 있게 된다.

[0056] 본 발명에 따른 시계열 데이터 분류 방법은 생체 신호 분류 장치, 사이버 데이터 패킷 분류 장치, 사용자 행동 인식 장치 등 다양한 센싱 장치에 적용될 수 있을 것이다.

[0057] 시계열 분류 문제를 풀기 위한 모델 성능 검증을 위해 사용하는 공인 데이터로 ArabigDigits, JapaneseVowels, CharacterTrajectories가 있으며, 본 발명에서 제안된 RP 기반의 GCN 모델의 성능과 기존 모델 대비 성능을 비

교하였다.

[0058] 비교 실험을 위해 GRU, RP 기반 CNN 의 성능도 함께 검증하였다.

표 1

[0059]

	GRU	RP 플롯 + CNN	RP 플롯 + GCN	타임 스텝 길이	샘플 수
F1	0.39	0.9	0.94	152	2858
Accuracy avg.	43.54%	90.17%	93.74%		
파라미터 수	1175495	920660	1185044		

표 2

[0060]

	GRU	RP 플롯 + CNN	RP 플롯 + GCN	타임 스텝 길이	샘플 수
F1	0.99	0.89	0.94	30	8800
Accuracy avg.	99.16%	88.78%	93.74%		
파라미터 수	10021386	10051210	8534666		

표 3

[0061]

	GRU	RP 플롯 + CNN	RP 플롯 + GCN	타임 스텝 길이	샘플 수
F1	0.91	0.34	0.93	10	640
Accuracy avg.	92.93%	30.91%	93.36%		
파라미터 수	6590209	6569451	6697737		

[0063] 표 1은 CharacterTrajectories 공인 데이터를 이용한 결과이며, 표 2는 ArabicDigits 공인 데이터를 이용한 결과이고, 표 3은 JapaneseVowels 공인 데이터를 이용한 결과이다.

[0064] 표 1 내지 표 3을 참조하면 RP 기반 GCN이 가장 좋은 성능을 보임을 확인할 수 있다.

[0065] 본 발명에 따른 시계열 데이터 분류 방법은 RP 기반에 GCN을 적용하여 시계열 데이터의 분류 성능을 높이면서도 정보 손실이 적은 효과가 있다.

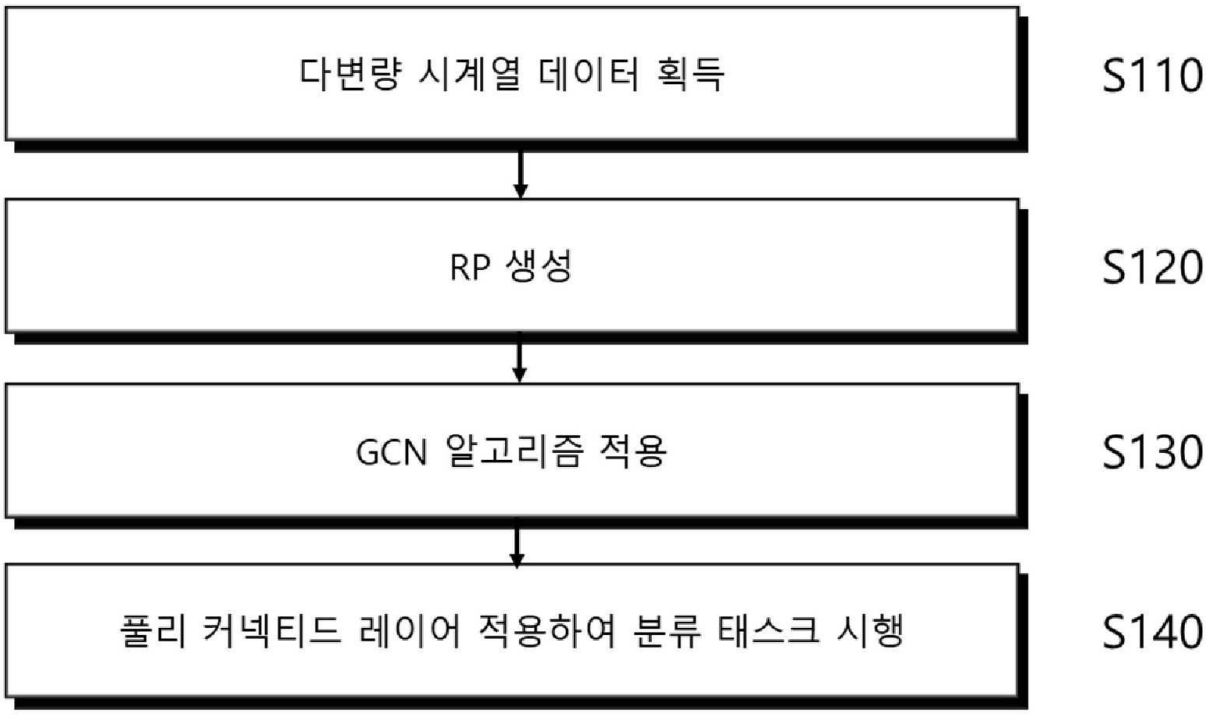
[0066] 또한 본 발명에 따른 시계열 데이터 분류 방법 및 센싱 장치는 사진과 달리 시계열 데이터의 경우 멀리 떨어진 데이터 포인트 간의 상관성도 제대로 잡아 멀리 있는 데이터 포인트 간의 관계성도 정확히 학습할 수 있는 효과가 있다.

[0067] 이상에서 본 발명이 구체적인 구성요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나, 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명이 상기 실시예들에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형을 꾀할 수 있다.

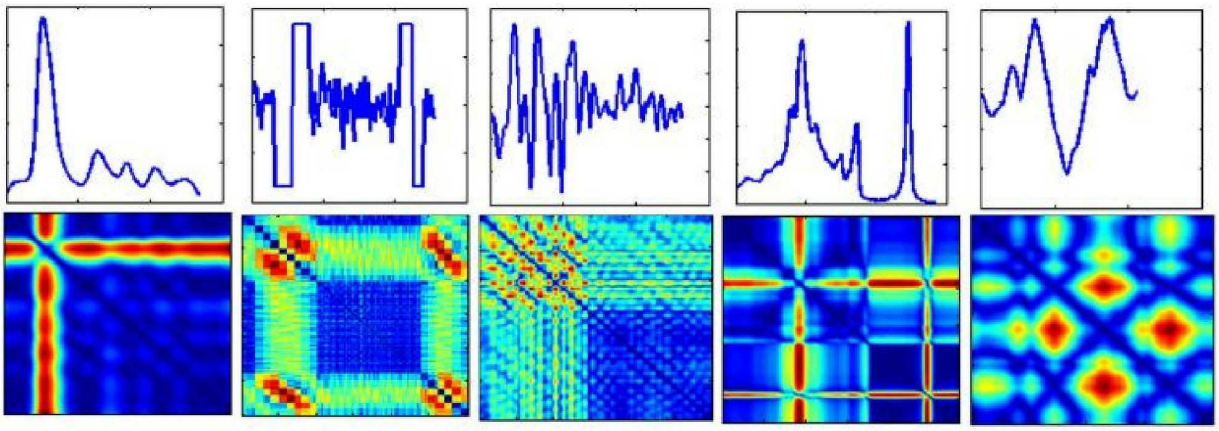
[0068] 따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등하게 또는 등가적으로 변형된 모든 것들은 본 발명의 사상의 범주에 속한다고 할 것이다.

도면

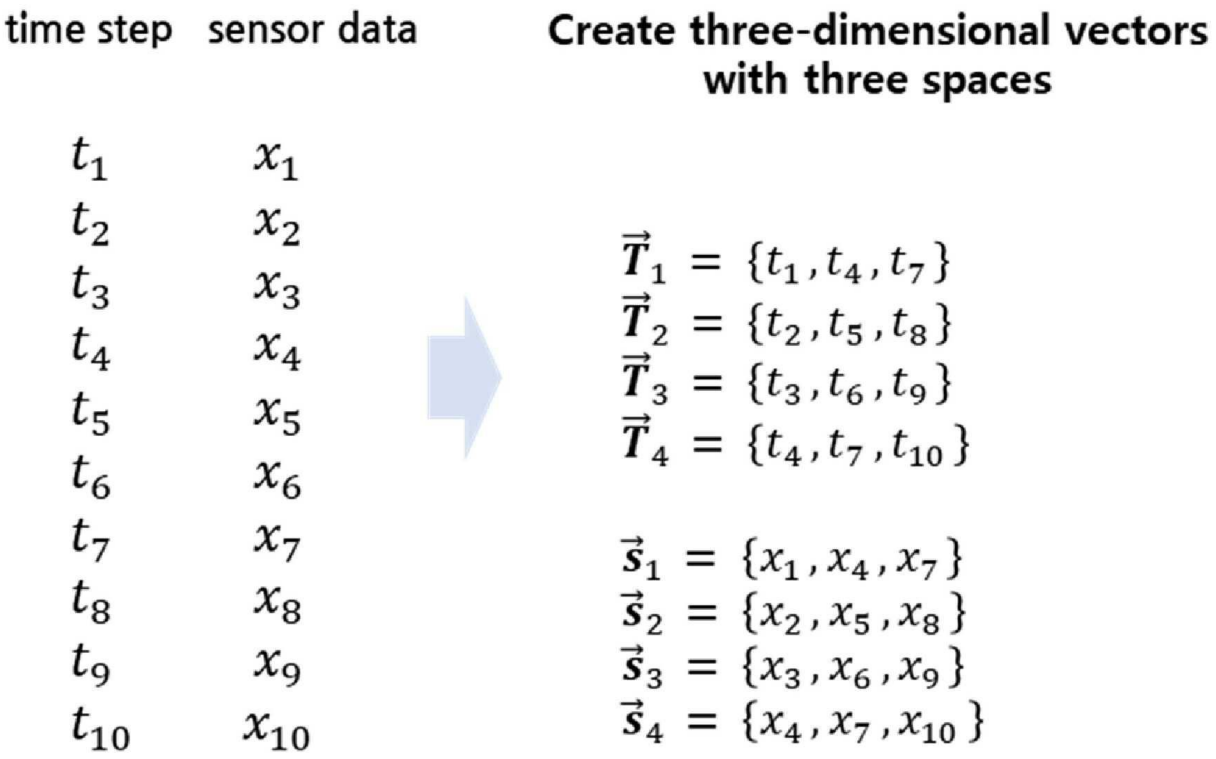
도면1


다변량 시계열 데이터 획득
S110
RP 생성
S120
GCN 알고리즘 적용
S130
풀리 커넥티드 레이어 적용하여 분류 태스크 시행
S140

도면2



도면3



도면4

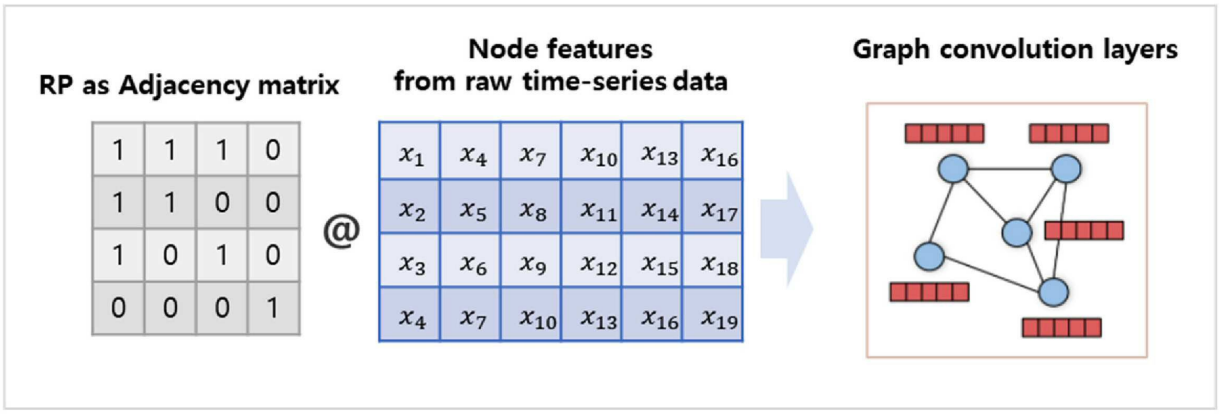
	$\{t_1, t_4, t_7\}$	$\{t_2, t_5, t_8\}$	$\{t_3, t_6, t_9\}$	$\{t_4, t_7, t_{10}\}$
$\{t_1, t_4, t_7\}$	$H(\|\vec{s}_1, \vec{s}_1\|)$	$H(\|\vec{s}_1, \vec{s}_2\|)$	$H(\|\vec{s}_1, \vec{s}_3\|)$	$H(\|\vec{s}_1, \vec{s}_4\|)$
$\{t_2, t_5, t_8\}$	$H(\|\vec{s}_2, \vec{s}_1\|)$	$H(\|\vec{s}_2, \vec{s}_2\|)$	$H(\|\vec{s}_2, \vec{s}_3\|)$	$H(\|\vec{s}_2, \vec{s}_4\|)$
$\{t_3, t_6, t_9\}$	$H(\|\vec{s}_3, \vec{s}_1\|)$	$H(\|\vec{s}_3, \vec{s}_2\|)$	$H(\|\vec{s}_3, \vec{s}_3\|)$	$H(\|\vec{s}_3, \vec{s}_4\|)$
	$H(\|\vec{s}_4, \vec{s}_1\|)$	$H(\|\vec{s}_4, \vec{s}_2\|)$	$H(\|\vec{s}_4, \vec{s}_3\|)$	$H(\|\vec{s}_4, \vec{s}_4\|)$

Recurrence plot; $\mathbf{R} \in \mathbb{R}^{4\times4}$

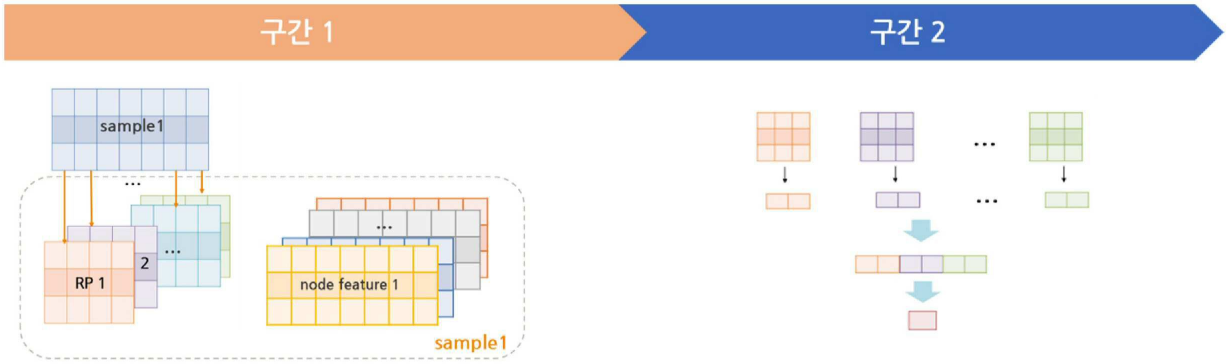
x_1	x_4	x_7
x_2	x_5	x_8
x_3	x_6	x_9
x_4	x_7	x_{10}

Feature matrix

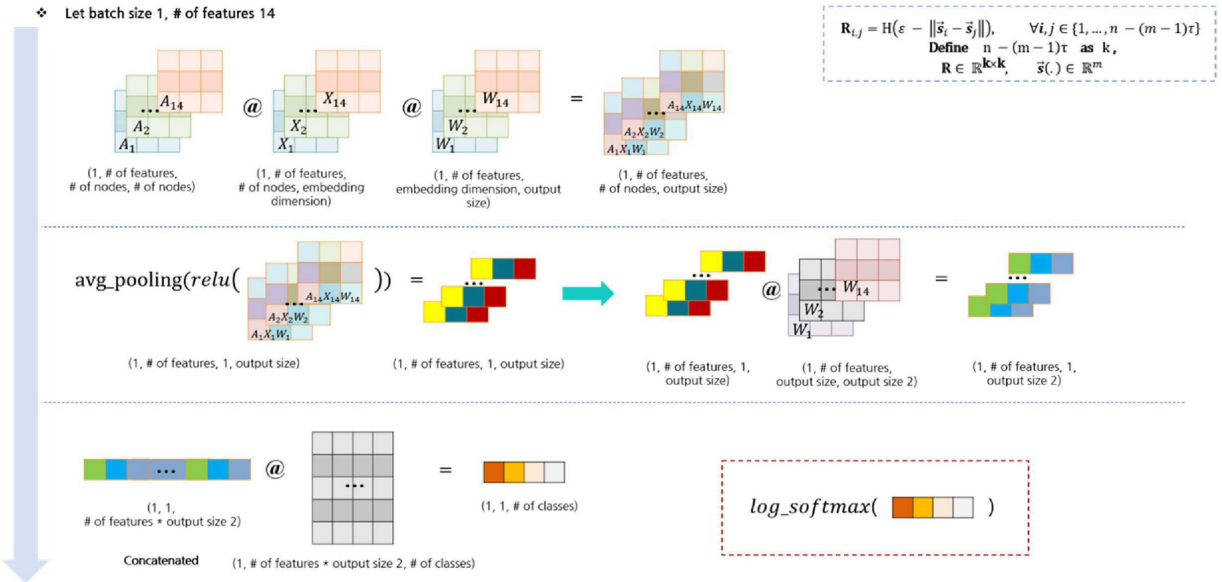
도면5


RP as Adjacency matrix
Node features from raw time-series data
Graph convolution layers
1 1 1 0
1 1 0 0
1 0 1 0
0 0 0 1
@
x_1 x_4 x_7 x_{10} x_{13} x_{16}
x_2 x_5 x_8 x_{11} x_{14} x_{17}
x_3 x_6 x_9 x_{12} x_{15} x_{18}
x_4 x_7 x_{10} x_{13} x_{16} x_{19}

도면6


구간 1
구간 2
sample1
RP 1
2
node feature 1
sample1

도면7


Let batch size 1, # of features 14
$\mathbf{R}_{i,j} = \mathrm{H}(\varepsilon - \|\vec{s}_i - \vec{s}_j\|), \quad \forall i,j \in \{1,\dots,n-(m-1)\tau\}$
Define $n-(m-1)\tau$ as k,
$\mathbf{R} \in \mathbb{R}^{k\times k}$, $\vec{s}(.) \in \mathbb{R}^m$
(1, # of features, # of nodes, # of nodes)
(1, # of features, # of nodes, embedding dimension)
(1, # of features, embedding dimension, output size)
(1, # of features, # of nodes, output size)
avg_pooling(relu( ))
(1, # of features, 1, output size)
(1, # of features, 1, output size)
(1, # of features, 1, output size)
(1, # of features, output size, output size 2)
(1, # of features, 1, output size 2)
(1, 1, # of features * output size 2)
Concatenated
(1, 1, # of classes)
(1, # of features * output size 2, # of classes)
log_softmax( )