



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0130153
(43) 공개일자 2020년11월18일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) G06N 3/08 (2006.01)
(52) CPC특허분류
G06N 20/00 (2019.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0053662
(22) 출원일자 2020년05월06일
심사청구일자 없음
(30) 우선권주장
10 2019 206 621.6 2019년05월08일 독일(DE)

(71) 출원인
로베르트 보쉬 게엠베하
독일 데-70442 슈투트가르트 포스트파흐 30 02 20
(72) 발명자
파일 토마스
독일 71272 렌닝엔 안나-토이러-슈트라쎄 14
(74) 대리인
양영준, 노대웅

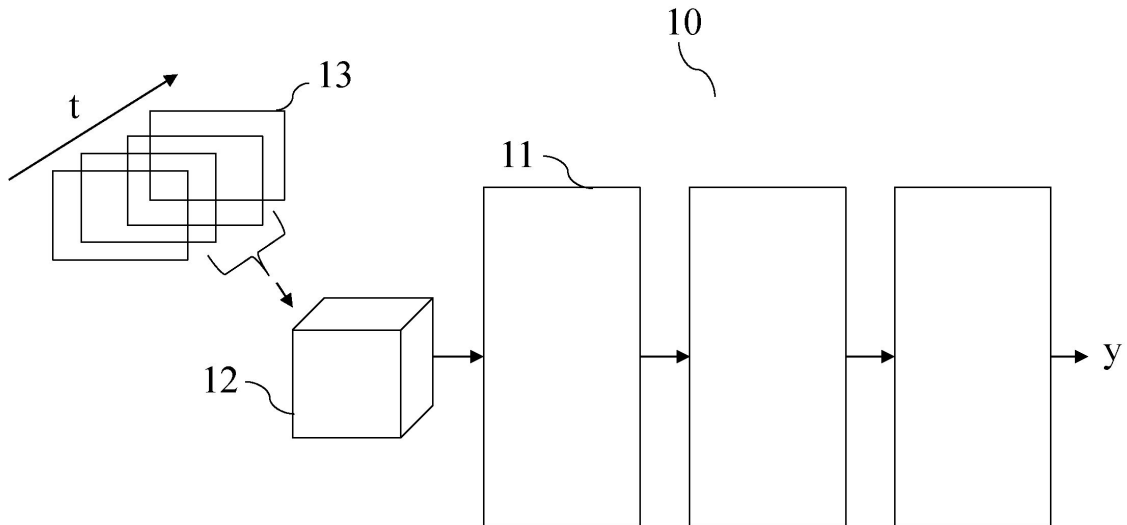
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 기계 학습 시스템, 그리고 기계 학습 시스템을 생성하기 위한 방법, 컴퓨터 프로그램 및 장치

(57) 요약

본 발명은, 하나 또는 복수의 시간 필터(21)를 포함하는 기계 학습 시스템(10)에 관한 것이다. 이미지의 시간 시퀀스를 포함하는 입력 변수가 기계 학습 시스템(10)에 의해 필터(21)를 사용하여 처리된다. 기계 학습 시스템(10)은, 모두 이미지의 각각 동일한 좌표에 또는 중간 결과의 각각 동일한 좌표에 놓여 있는 일련의 픽셀에 필터(21)를 적용하도록 설계된다. 필터(21)의 필터 계수는 양자화된다. 본 발명은 또한, 기계 학습 시스템(10)을 생성하기 위한 방법, 컴퓨터 프로그램, 및 장치와도 관련이 있다.

대표도



명세서

청구범위

청구항 1

하나 이상의 필터(21)를 포함하는 기계 학습 시스템(10), 특히 컨볼루션 층을 갖는 심층 신경망으로서,
기계 학습 시스템(10)은, 필터(21)를 사용하여 입력 변수(12)에 따라 출력 변수를 결정하도록 설계되며,
입력 변수(12)는 이미지의 시간 시퀀스를 포함하며,
기계 학습 시스템(10)은, 모두 기계 학습 시스템(10)의 이미지의 각각 동일한 좌표에 또는 중간 결과의 각각 동일한 좌표에 놓여 있는 일련의 픽셀에 필터(21)를 적용하도록 설계되며,
필터(21)의 필터 계수가 양자화되는, 기계 학습 시스템(10).

청구항 2

제1항에 있어서, 필터(21)가 다차원적이며, 이 경우, 기계 학습 시스템(10)이, 모두 이미지의 각각 동일한 좌표에 또는 중간 결과의 각각 동일한 좌표에 놓여 있는 일련의 섹션에 필터(21)를 적용하도록 설계되는, 기계 학습 시스템(10).

청구항 3

제2항에 있어서, 기계 학습 시스템(10)은 또한, 이미지 내부에서 각각 상이한 좌표에 또는 중간 결과 내부에서 각각 상이한 좌표에 놓여 있는 일련의 또 다른 섹션에 필터(21)를 추가로 적용하도록 설계되는, 기계 학습 시스템(10).

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 필터 계수 외에 중간 결과도 양자화되는, 기계 학습 시스템(10).

청구항 5

제4항에 있어서, 필터 계수 및/또는 중간 결과가 이진 또는 삼진인, 기계 학습 시스템(10).

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 필터(21)는, 필터(21)를 사용하는 경우에 이미지 또는 중간 결과의 개별 픽셀을 생략하도록 설계되는, 기계 학습 시스템(10).

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 기계 학습 시스템(10)은 필터(21) 직후에 배치된 풀링 연산(pooling operation)을 포함하는, 기계 학습 시스템(10).

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 이미지는 사전 설정 가능한 순서에 따라 연속적으로 배열되며,
필터(21)는, 각각 동일한 좌표에 놓여 있는 픽셀 또는 섹션의 시퀀스에 적용되고, 상기 시퀀스의 픽셀 또는 섹션이 상기 순서에 따라 배열되는, 기계 학습 시스템(10).

청구항 9

제1항 내지 제8항 중 어느 한 항에 따른 기계 학습 시스템(10)을 작동시키기 위한 방법으로서, 다음의 단계들, 즉:

이미지의 시퀀스를 기계 학습 시스템(10)의 입력 변수로 그룹핑하는 단계,

입력 변수에 따라 필터(21)를 사용해서 출력 변수를 결정하는 단계
를 포함하는, 기계 학습 시스템의 작동 방법.

청구항 10

제9항에 있어서, 이미지가 각각 바로 후속하는 시점에 연속적으로 검출되며,

그 다음에, 출력 변수가 결정되었다면, 검출된 새로운 이미지가 입력 변수에 부가되고, 입력 변수의 포함된 이미지들 중 하나, 특히 시간상 가장 오랫동안 포함된 이미지가 상기 입력 변수로부터 제거되며,

그에 이어서 또 다른 출력 변수가 필터를 사용해서 입력 변수에 따라 결정되는, 기계 학습 시스템의 작동 방법.

청구항 11

제1항 내지 제8항 중 어느 한 항에 따른 기계 학습 시스템(10)의 학습 방법으로서, 다음 단계들, 즉:

적어도 복수의 이미지 시퀀스 또는 관련 트레이닝 출력 변수를 포함하는 트레이닝 데이터를 제공하는 단계,

각각 하나 이상의 출력 변수를 각각의 이미지 시퀀스로 결정하는 단계,

결정된 출력 변수 및 트레이닝 데이터에 따라 비용 함수를 결정하는 단계,

비용 함수에 따라 기계 학습 시스템(10)의 파라미터를 조정함으로써, 사전 설정 가능한 기준과 관련하여 비용 함수가 최적화되는 단계, 및

기계 학습 시스템(10)의 파라미터, 특히 필터 계수를 양자화하는 단계

를 포함하는, 기계 학습 시스템의 학습 방법.

청구항 12

컴퓨터를 이용해서 명령을 실행할 때, 상기 컴퓨터로 하여금 제9항 내지 제11항 중 어느 한 항에 따른 방법을 실시하도록 설계된 명령을 포함하는 컴퓨터 프로그램.

청구항 13

제12항에 따른 컴퓨터 프로그램이 저장되어 있는 기계 판독 가능한 메모리 요소.

발명의 설명

기술 분야

[0001] 본 발명은, 양자화된 파라미터들을 가진 기계 학습 시스템에 관한 것이다. 또한, 본 발명은, 기계 학습 시스템을 생성하기 위한 방법 및 컴퓨터 프로그램, 및 장치에 관한 것이다.

배경 기술

[0002] 미공개 문서 DE 10 2017 218 889.8호는, 하나 또는 복수의 출력 변수에 대한 내부 프로세싱 체인에 의해 하나 또는 복수의 입력 변수를 처리하도록 설계된 인공 지능 모듈을 개시하고 있다. 내부 처리 체인은 하나 또는 복수의 파라미터에 의해 확정된다. 하나 이상의 통계 분포에 따라 파라미터를 결정하도록 설계된 분포 모듈이 제공된다.

[0003] 미공개 문서 DE 10 2018 216 471.1호는 양자화된 신경망을 개시하고 있다.

[0004] 예컨대 Hubara 외 공저, "Quantized neural networks: Training neural networks with low precision weights and activations"(The Journal of Machine Learning Research, 2017년, 18권, 1호, 페이지 6869 ~ 6898쪽)에 개시된 바와 같이, 매우 강력한 양자화에 의해 양자화된 신경망을 생성할 수 있다.

[0005] RNN(Recurrent Neural Networks: 순환 신경망)은, 센서 데이터의 일 시간 차원 내에 포함되어 있는 정보도 처리할 수 있다는 장점이 있다. 하지만 RNN의 단점은, 이 순환 신경망이 강력하게 양자화될 수 없다는 것이다.

[0006] 그러나 모바일 컴퓨팅 유닛 상에서 RNN을 더욱 효율적으로 작동시킬 수 있기 위해서는, RNN의 파라미터를 양

자화하는 것이 바람직하다.

[0007] 독립 청구항 1에 따르면, RNN 대신에 시간 컨볼루션을 갖는 신경망을 사용하는 것이 제안된다. 이는, 계산이 복잡한 컨볼루션에도 불구하고, 이들 신경망이 놀랍게도 이들 신경망의 파라미터의 양자화 시, 전술한 RNN의 장점을 유지하면서, 필적하는 RNN보다 더 에너지 절약적일 수 있다는 장점을 갖는다.

발명의 내용

[0008] 제1 양태에서는, 시계열(time series)을 처리하기 위한, 특히 메모리 효율적이고 에너지 효율적인 기계 학습 시스템이 제안된다.

[0009] 청구항 1에 따른 특징들을 갖는 기계 학습 시스템, 특히 신경망은 하나 이상의 필터를 포함하며, 이 경우 적어도 필터의 필터 계수가 양자화된다. 필터 계수 외에 기계 학습 시스템의 또 다른 파라미터가 양자화되는 것도 고려해볼 수 있다. 대안적으로 또는 추가로, 적어도 활성화가 양자화될 수 있다. 활성화의 개수가 기계 학습 시스템의 파라미터(예컨대 가중치)의 개수를 초과하는 경우, 활성화의 양자화가 유용할 수 있다.

[0010] 기계 학습 시스템은, 필터를 사용하여 입력 변수에 따라 출력 변수를 결정하도록 설계된다. 입력 변수는 이미지의 시간 시퀀스를 포함한다. 대안적으로, 시간에 걸쳐 연속적으로 검출된 복수의 변수, 예컨대 1차원 센서 데이터가 사용될 수 있다. 기계 학습 시스템은 또한, 모두 이미지의 각각 동일한 좌표에 또는 기계 학습 시스템의 중간 결과의, 특히 중간 처리된 이미지의 각각 동일한 좌표에 놓여 있는 일련의 픽셀에 필터를 적용하도록 설계된다. 이미지 대신 검출된 변수가 사용되는 경우, 각각 동일한 위치에 놓여 있는 검출된 변수 또는 중간 결과의 데이터 포인트에 필터가 적용된다. 중간 결과는 예를 들어 신경망의 층들(layers)의 활성화이다.

[0011] 또한, 기계 학습 시스템의 필터 계수 및 바람직하게는 또 다른 파라미터가 각각 사전 설정 가능한 양자화된 값 범위로부터의 값을 갖는 것도 고려할 수 있다. 사전 설정 가능한 양자화된 값 범위는, 사전 설정된 개수의 양자화 레벨에 의해 그리고/또는 사전 설정 가능한 개수의 비트에 의해 정의될 수 있다.

[0012] 양자화란, 필터 계수가 십진 표현에서 양자화된 값만을, 바람직하게는 부동 소수점, 고정 소수점 또는 정수 형식으로 양자화된 값만을 취할 수 있다는 것으로 이해될 수 있다. 바람직한 방식으로, 필터 계수는 최대 32비트로, 특히 최대 16비트 또는 8비트로 그리고 바람직한 방식으로는 4비트 또는 1비트로 양자화된다. 양자화는 나아가, 필터 계수가 복수의 사전 설정 가능한 양자화 레벨의 가장 가까이 놓여 있는 양자화 레벨로 반올림되거나 반내림된 것으로 이해될 수도 있다. 이 경우, 필터 계수는 각각 복수의 양자화 레벨의 하나의 값을 가지며, 이 경우 상기 값은 메모리 내에 저장된다. 이 경우, 양자화 레벨이 메모리 효율적인 방식으로 저장될 수 있는 것이 바람직하다. 양자화 레벨의 수는, 양자화의 목적으로 사용된 비트의 수에 좌우된다. 대안적으로는, 학습된 기계 학습 시스템이 양자화될 수 있으며, 이와 같은 시스템에서는 그의 파라미터가 학습 시 최대 해상도에서 매칭된 후에 양자화되었다.

[0013] 사전 설정 가능한 양자화된 값 범위는 하나의 리스트일 수도 있다. 필터 계수가 감소된 해상도를 갖는 것을 고려할 수 있으며, 이 경우 해상도는 양자화에 의해 감소되었다.

[0014] 해상도는, 얼마나 많은 상이한, 특히 양자화된 가능한 값들에 의해 각각 파라미터 및/또는 중간 결과가 나타나게 되는지를 특성화할 수 있다.

[0015] 필터의 적용이란, 모두 동일한 좌표를 갖는 데이터 포인트 또는 섹션으로써, 필터가 수학적으로 컨볼루션되는 것으로 이해될 수 있다.

[0016] 제1 양태에 따른 기계 학습 시스템은, 양자화를 통해 필터 계수를 저장하기 위한 비트가 덜 요구됨에 따라, 메모리 공간이 덜 요구된다는 장점이 있다. 이로 인해, 하드웨어 내에서 기계 학습 시스템의 보다 콤팩트한 디자인을 가능하게 하는 더 작은 메모리가 사용될 수 있다. 파라미터 및 활성화의 저장 및 관독을 위한 에너지 소비가 비트의 수에 비례하고, 곱셈은 적어도 파라미터 및 활성화의 해상도(비트 수)의 제공이기 때문에, 특히 양자화를 통한 비트 수의 감소는 특히 컴퓨터 리소스 효율적이다. 기계 학습 시스템이 소프트웨어로 구현된 경우에도 이러한 장점들이 작용하는데, 그 이유는 더 적은 메모리가 요구되기 때문이다.

[0017] 예컨대 이미지와 같은 고해상도 입력 변수 외에 필터가 에너지 소비에 가장 큰 영향을 미치기 때문에, 필터의 양자화가 유리하다. 그 이유는, 필터의 양자화의 결과로 나타나는 컴퓨터 리소스 안전 처리에 의해, 기계 학습 시스템이 고정식 용례(예컨대 컴퓨팅 센터)를 위해서뿐만 아니라 이동식 용례(예컨대 모바일 데이터 처리 시스템 또는 임베디드 시스템)를 위해서도 사용 가능하기 때문이다. 이와 같은 방식으로 양자화된 기계 학습 시

템은, 센서 가까이에서 또는 센서 자체에서 측정된 데이터를 처리하기에 특히 적합하며, 이로써 측정된 데이터는 압축되어 보다 대역폭 효율적 방식으로 전송될 수 있다.

- [0018] 이미지는 라이다 이미지, 레이더 이미지, 초음파 이미지 또는 카메라 이미지일 수 있다.
- [0019] 입력 변수는, 복수의 검출된 변수 또는 이미지의 시퀀스를 포함하는 텐서(tensor)이다. 검출된 변수들 또는 이미지들은 텐서의 제1 차원을 따라 배열된다. 그 다음에, 필터가 적어도 제1 차원을 따라 적용된다. 그렇기 때문에, 시간적 순서를 특성화하는 제1 차원을 따른 필터의 적용을 시간 컨볼루션이라고 지칭한다.
- [0020] 이와 같은 기계 학습 시스템의 장점은, 시간 축을 따라서 포함된 정보들이 시간 필터링을 통해 함께 처리된다는 것이다. 이로 인해, 상기 기계 학습 시스템은 반복적인 특성을 갖는다. 하지만 일반적으로 강력하게 양자화될 수 없는 반복적 링크지는 요구되지 않는다. 그럼에도, 양자화된 필터를 통해, RNN을 이용하는 것보다 더 효율적인 입력 변수 시퀀스의 처리가 달성된다.
- [0021] 더욱이, 필터는 다차원적이고, 기계 학습 시스템은 나아가, 모두 이미지의 각각 동일한 좌표 또는 중간 결과의 각각 동일한 좌표에 놓여 있는 일련의 섹션에 필터를 적용하도록 설계된다.
- [0022] 하나의 섹션은, 검출된 변수, 이미지, 또는 중간 결과의 개별 좌표에 놓여 있는 복수의 데이터 포인트 또는 픽셀을 포함한다. 섹션은 1차원/다차원일 수 있고, 이에 상응하게 필터의 차원 수가 선택되어야 한다. 이로 인해, 모두 제1 차원을 따라서, 특히 제1 차원에 평행하게, 놓인 복수의 픽셀 또는 데이터 포인트가 추가로 필터링 시 제1 차원을 따라 고려된다는 사실에 주목한다.
- [0023] 예를 들어, 이 섹션은 하나의 선 또는 직사각형일 수 있다. 이미지 내 섹션들의 상이한 정렬, 예컨대 수직, 수평, 또는 대각선의 정렬이 고려될 수 있다.
- [0024] 더 나아가, 기계 학습 시스템은 또한, 이미지 내 각각 상이한 좌표들 또는 중간 결과 내 각각 상이한 좌표들에 놓인 일련의 또 다른 섹션에 필터를 추가로 적용하도록 설계된다.
- [0025] 상기 제안에서의 장점은, 필터의 추가 적용을 이용하여 각각 검출된 변수를 따라 추가의 정보가 출력 변수로 유입된다는 것이다. 다시 말해, 필터는 텐서의 또 다른 차원에 적용된다. 이를 시간적 및 공간적 컨볼루션이라고 지칭할 수 있다.
- [0026] 더 나아가, 필터 계수 외에, 기계 학습 시스템의 중간 결과, 특히 활성화도 양자화된다.
- [0027] 또한, 중간 결과가 사전 설정 가능한 양자화된 값 범위 또는 사전 설정 가능한 또 다른 양자화된 값 범위로부터의 각각 하나의 값을 갖는 것도 고려할 수 있다.
- [0028] 더 나아가, 필터 계수 및/또는 중간 결과가 이진 또는 삼진이다. 이는, 개별 필터 계수 및/또는 중간 결과가 최대 2개 또는 3개의 상이한 값을 취할 수 있다는 것으로 이해된다. 바람직하게, 이들 상이한 값은 최대 1개 또는 2개의 비트에 의해 표현된다. 이상적으로는 1개의 비트에 의해 표현되는데, 그 이유는 이 경우 특히 계산 효율적이고 메모리 효율적인 이진 신경망이 존재하기 때문이다.
- [0029] 더 나아가, 특히 시간 차원을 따르는 필터링 시, 필터가 개별 데이터 포인트를 생략한다. 이는, 소수의 필터 계수를 이용해서 필터의 더 큰 시야가 달성될 수 있고, 이로써 더 나아가 과거까지 살펴볼 수 있다는 장점이 있다.
- [0030] 더 나아가, 기계 학습 시스템은 필터 직후에 배치된 풀링 연산(pooling operation)을 포함한다.
- [0031] 풀링 연산이란, 필터링된 개별 데이터 포인트 또는 필터링된 픽셀이 무시되는, 특히 폐기되는 것으로 이해된다. 바람직하게는, 클러스터 내부에서 0에 가까운 값을 갖는 데이터 포인트가 무시되는 로컬 풀링(local pooling)이 사용된다. 대안적으로 평균 풀링(average pooling)이 사용될 수 있다.
- [0032] 더 나아가서는, 이미지 또는 검출된 변수가 사전 설정 가능한 순서에 따라 연속적으로, 특히 이미지 또는 검출된 변수의 검출된 개별 시점에 따라 배열된다. 바람직하게 상기 배열은 검출 시점 이후 오름차순/내림차순으로 수행된다.
- [0033] 주지할 점은, 기계 학습 시스템은 오직 하드웨어로만 실현될 수도 있고, 소프트웨어로도 또는 소프트웨어와 하드웨어의 혼합으로도 실현될 수 있고, 이로 인해 파라미터를 저장하기 위해 각각 더 적은 메모리가 제공될 수 있어서, 본 발명의 제1 양태에 따른 기계 학습 시스템이 초소형 기술 시스템에도 설치 및 사용될 수 있다는 점이다.

- [0034] 본 발명의 제2 양태에서는, 특히 본 발명의 제1 양태의 기계 학습 시스템을 작동시키기 위한, 컴퓨터로 구현된 방법이 제안된다. 이 방법은 다음과 같은 단계들을 포함한다: 복수의 검출된 변수를 기계 학습 시스템의 입력 변수로 그룹핑하는 단계. 이어서 입력 변수에 따라 필터를 사용해서 출력 변수를 결정하는 단계가 후속된다. 이 단계에서 기계 학습 시스템의 필터가 사용되며, 특히 입력 변수 또는 중간 결과의 시간에 따른 필터링이 실시된다.
- [0035] 본 발명의 제2 양태의 경우, 이미지가 각각 바로 후속하는 시점에 연속적으로 검출된다. 그 다음에, 출력 변수가 결정되었다면, 검출된 새로운 이미지가 입력 변수에 부가되고, 포함된 이미지들 중 하나, 특히 시간상 가장 오랫동안 포함된 이미지가 상기 입력 변수로부터 제거된다.
- [0036] 바람직하게, 입력 변수의 이미지는 텐서의 제1 차원을 따라 하나의 위치만큼 이동되며, 이 경우 비워진 지점에 검출된 새로운 이미지가 부가되고, 입력 변수의 검출된 가장 오래된 이미지가 텐서로부터 제거된다.
- [0037] 더 나아가, 일련의 시점들 중 개별 변수가 검출된 시점에 변수, 특히 이미지가 각각 할당되며, 이 경우 각각의 시점에서 새로운 변수가 검출되고, 직전의 시점에 검출된 변수 외에 새로 검출된 변수가 입력 변수에 부가되며, 가장 오래전의 과거 시점에 할당된 검출된 변수가 입력 변수로부터 제거된다.
- [0038] 본 발명의 또 다른 일 양태에서는, 제1 양태의 기계 학습 시스템을 학습하기 위한 방법이 제안된다. 학습은, 적어도 트레이닝 데이터를 제공하는 단계를 포함한다. 트레이닝 데이터는, 적어도 검출된 변수의 복수의 이미지 시퀀스 또는 시계열을 포함한다. 그 후에, 각각 하나 이상의 출력 변수를 각각의 이미지 시퀀스 또는 시계열로 결정하는 단계가 이어진다. 그 후에, 결정된 출력 변수 및 트레이닝 데이터에 따라 비용 함수를 결정하는 단계가 이어진다. 이어서, 비용 함수에 따라 기계 학습 시스템의 파라미터를 조정함으로써, 결과적으로 사전 설정 가능한 기준과 관련하여 비용 함수가 최적화되는 단계가 뒤따른다. 그 후에, 파라미터, 특히 필터 계수의 양자화가 이어진다.
- [0039] 바람직하게, 양자화는 처음부터 사용되고, 이 양자화를 고려하여 기계 학습 시스템이 학습된다.
- [0040] 또 다른 일 양태에서는, 제1 양태 또는 제2 양태에 따라 생성된 기계 학습 시스템이, 일 센서의 검출된 센서 변수, 특히 이미지에 따라, 추후 제어 유닛을 이용해서 제어 변수를 결정하기 위해 사용될 출력 변수를 결정할 수 있다.
- [0041] 제어 변수는 기술 시스템의 액추에이터를 제어하기 위해 사용될 수 있다. 기술 시스템은, 예컨대 적어도 부분적으로 자율적인 기계, 적어도 부분적으로 자율적인 차량, 로봇, 공구, 작업 기계 또는 드론과 같은 비행 물체일 수 있다. 입력 변수는, 예를 들어 검출된 센서 데이터에 따라 결정될 수 있고, 기계 학습 시스템에 제공될 수 있다. 센서 데이터는, 예컨대 기술 시스템의 카메라와 같은 센서에 의해 검출될 수 있고, 대안적으로 외부로부터 수신될 수 있다.
- [0042] 또 다른 일 양태에서는, 컴퓨터 프로그램이 제안된다. 컴퓨터 프로그램은, 전술한 본 발명의 제2 양태의 방법들 중 하나를 실시하도록 설계된다. 컴퓨터 프로그램은, 컴퓨터 프로그램이 컴퓨터상에서 실행될 때 컴퓨터로 하여금 모두 자체 단계들을 갖는 상기 방법들 중 하나를 실시하게 하는 명령을 포함한다. 또한, 컴퓨터 프로그램이 저장되어 있는 기계 판독 가능 메모리 모듈도 제안된다.
- [0043] 전술된 양태들의 실시예들이 첨부 도면부에 도시되어 있고, 이하의 명세서에서 더욱 상세하게 설명된다.

도면의 간단한 설명

- [0044] 도 1은 각각 시계열을 처리하기 위한 2개의 상이한 신경망에 대해, 양자화 해상도에 걸쳐 예시적으로 도시된 에너지 소비의 개략도이다.
- 도 2는 시간 컨볼루션을 위한 필터를 포함하는 심층 신경망의 개략도이다.
- 도 3은 시간 컨볼루션의 개략도이다.
- 도 4는 시간 컨볼루션을 위한 필터를 포함하는 신경망을 작동시키기 위한 방법의 개략적 흐름도이다.
- 도 5는 로봇에 대한 시간 컨볼루션을 위한 필터를 포함하는, 신경망 사용의 개략도를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0045] 도 1은, 시간 필터링(NN-temp)을 갖는 신경망의 측정된 에너지 소비(E) 및 양자화 해상도(Q)에 대한 RNN의 측정

된 에너지 소비(E)가 도시되어 있는 그래프의 개략도를 보여준다. 2개의 신경망이 동일한 데이터 세트 상에서 유사한 결과를 얻는 점에 주의하였다. y축은, 그래프의 원점에서 가장 낮으며 원점까지의 거리가 증가할수록 상승하는 에너지 소비(E)를 지시한다. x축 상에는, 그래프의 원점의 방향으로 증가하는 양자화 해상도(Q)가 기입되어 있다. 즉, 원점에 가까울수록, 양자화 해상도, 특히 파라미터를 형성하기 위한 비트의 수가 더 높다. 하기에서는, 높은 양자화 해상도에 대한 시간 필터링(NN-temp)을 갖는 신경망이 동일한 양자화 해상도에서 필적하는 RNN보다 확연히 더 큰 에너지 소비를 갖는 예시적 경우가 고려된다.

[0046] 두 신경망의 파라미터의 해상도가 감소하면, 상기 두 신경망의 에너지 소비가 각각 감소한다. 시간 필터링(NN-temp)을 갖는 신경망의 최소 양자화 해상도를 RNN의 최소 양자화 해상도와 비교해보면, 두 신경망의 결과의 품질이 실질적으로 변하지 않고 유지되는 조건하에, 시간 필터링(NN-temp)을 갖는 신경망의 최소 양자화 해상도가 RNN의 최소 양자화 해상도보다 훨씬 더 낮은 점이 눈에 띈다. 이는, 시간 필터링(NN-temp)을 갖는 신경망이 강한 양자화로 인해 RNN보다 에너지를 덜 소비하도록 한다.

[0047] 도 2는, 기계 학습 시스템(10), 특히 심층 신경망의 개략도를 보여준다. 기계 학습 시스템(10)은 사전 설정 가능한 순서로 서로 연결되어 있는 복수의 층(11)을 포함한다. 기계 학습 시스템(10)은 입력 변수(12)에 따라 출력 변수(y)를 결정한다.

[0048] 기계 학습 시스템(10)은, 분류, 다시 말해 이미지들 중 하나 또는 가장 최근에 검출된 이미지들 중 하나, 즉, 입력 변수(12)의 분할을 출력 변수(y)로서 출력할 수 있다. 대안적으로 출력 변수(y)는 회귀(regression)일 수 있다.

[0049] 상기 실시예에서 입력 변수(12)는 텐서이다. 텐서는 연속으로 검출된 복수의 이미지(13)를 포함한다. 이 경우, 이미지들은 텐서의 제1 차원을 따라 배열된다.

[0050] 층들은 완전 연결 층(fully connected layers)일 수 있으며, 이 경우 층들 중 적어도 하나는 시간 필터링을 수행하도록 설계된다. 기계 학습 시스템을 통한 입력 변수(12)의 전파 시, 시간 필터링이 사용된다. 시간 필터링은 이하 도 3에서 더욱 상세하게 설명된다.

[0051] 도 3은, 시간 컨볼루션, 달리 표현하면 시간 필터링의 개략도를 보여준다.

[0052] 도 3에 도시된 텐서(20)는, 기계 학습 시스템(10)의 입력 변수(12) 또는 중간 변수일 수 있다. 중간 변수는 예를 들어 층들(11) 중 하나의 출력 변수이다.

[0053] 텐서(20) 아래의 좌측에는 좌표계(23)가 개략적으로 도시되어 있다. 이 좌표계(23)는 예를 들어 3개의 차원(h, t, w)을 포함하고, 이들 차원을 따라 텐서(20)가 연장된다. 이미지(13)가 "시간" 차원(t)을 따라 텐서(20) 내에서 연이어 배열되는 점을 주지한다. 차원 h 및 w는 예를 들어 이미지(13)가 연장되는 공간적 차원들이다.

[0054] 필터(21)를 이용해서 이제 텐서(20)가 시간 필터링된다. 이는, 필터가 텐서(20)를 따라 "시간" 차원(t)을 따라 필터링을 수행하는 경우에 발생한다. 시간 컨볼루션의 한 예시적인 파형, 특히 텐서(20)를 갖는 필터(21)의 시간 컨볼루션의 예시적인 파형이, 파선 화살표(22)에 따라 개략적으로 도시되어 있다.

[0055] 시간 필터링은 대안적으로 추가 필터들(21a, 21b, 21c) 중 하나를 사용하여 수행될 수 있다. 이들 추가 필터는, 이들이 상이한 복수의 차원을 갖는다는 점에서 상호 구별된다. 예를 들어, 추가 필터(21c)는 3차원 필터이다. 텐서(20)를 갖는 3차원 필터(21c)의 컨볼루션의 가능한 일 파형이 3차원 필터(21c) 옆에 파선 화살표(22b)에 의해 예시적으로 도시되어 있다. 컨볼루션이 파선 화살표(22b)에 상응하게 구현되는 경우, 텐서(20)를 갖는 3차원 필터(21c)의 필터링은 시간 필터링일뿐만 아니라 공간 필터링이기도 하다는 점을 주지한다.

[0056] 도 4는, 기계 학습 시스템(10)을 작동시키기 위한 방법의 개략적인 흐름도를 보여준다.

[0057] 제1 실시예에서, 본원 방법은 단계 "S40"으로 시작한다. 본 단계에서는, 입력 변수(12)가 기계 학습 시스템(10)에 제공된다.

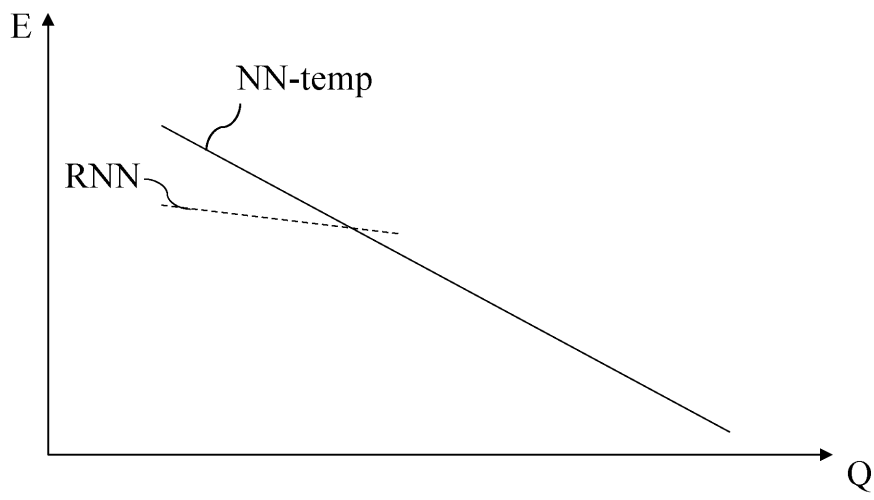
[0058] 후속하는 단계 "S41"에서는, 제공된 입력 변수(12)가 기계 학습 시스템(10)에 의해 전파된다. 전파 시, 기계 학습 시스템(10)의 입력 변수(12) 또는 중간 변수가 필터(21)에 의해 시간 필터링된다(도 3에 대한 설명 참조). 필터는 양자화된 필터 계수를 갖는다. 바람직하게, 기계 학습 시스템(10)의 모든 또 다른 파라미터도 마찬가지로 양자화된다.

[0059] 후속하는 단계 "S42"에서는 기계 학습 시스템의 출력 변수가 출력된다.

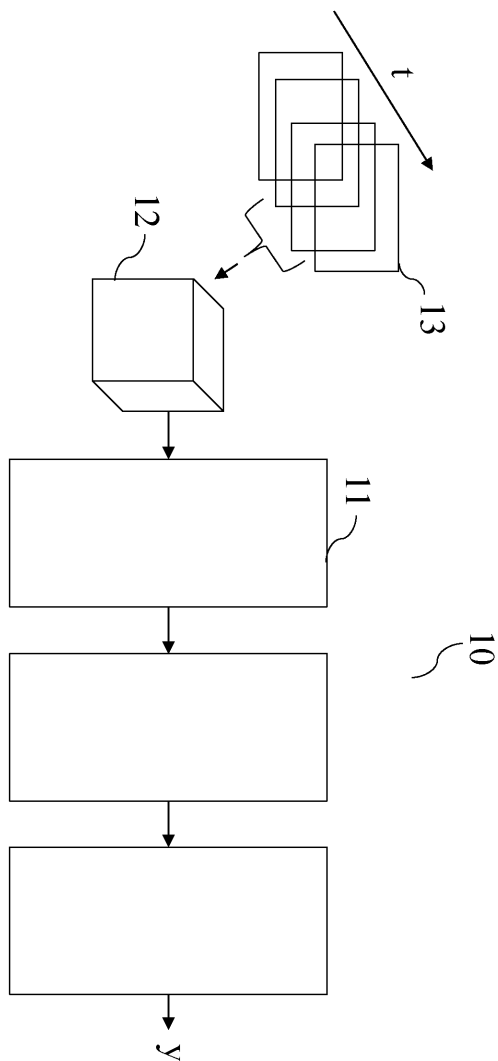
- [0060] 이어서, 후속 단계 "S43"에서는 새로운 이미지가 검출되어 입력 변수(12)에 부가되고, 입력 변수(12)의 한 포함 된 이미지가 입력 변수(12)로부터 제거된다.
- [0061] 이어서 본원 방법의 모든 단계 "S40" 내지 "S43"가 반복적으로 실시된다.
- [0062] 기계 학습 시스템(10)을 작동시키기 위한 방법의 또 다른 일 실시예에서는, 단계 "S40" 내지 "S43"이 기계 학습 시스템(10)을 학습하기 위해 사용된다. 학습 시, 단계 "S43" 이후에 추가로, 기계 학습 시스템의 파라미터가 비용 함수(loss function)에 따라 최적화된다. 이 경우, 비용 함수는, 기계 학습 시스템(10)의 결정된 출력 변수와 각각 사용된 입력 변수의 관련 트레이닝 출력 변수 간의 편차를 특성화한다. 출력 변수의 결정 시, 이미 양자화된 파라미터가 사용될 수 있다. 기계 학습 시스템의 파라미터를 통한 비용 함수의 최적화는, 예컨대 "(시간 기반) 오차역전파법[Backpropagation (Through Time); BPTT]"과 같은 예를 들어 경사 하강법(gradient descent method)을 이용해서 수행될 수 있다. 이 경우, 경사 하강법은, 파라미터에 따라 비용 함수를 최소화하거나 최대화하기 위하여 사용된다. 파라미터가 최적화된 후에는, 파라미터, 특히 필터 계수가 양자화된다. 바람직하게는 파라미터를 최적화하기 위해 적어도 32비트의 파라미터 해상도가 사용된다. 이 경우, 양자화 시 해상도가 감소하는데, 예를 들어 8비트로 또는 심지어 1비트까지 감소한다.
- [0063] 기계 학습 시스템(10)을 작동시키기 위한 방법의 제1 실시예의 일 개선예에서는, 기계 학습 시스템(10)의 출력 변수에 따라 적어도 부분 자율 로봇을 제어하기 위하여, 단계 "S43"의 종료 후에 선택적으로 상기 출력 변수가 사용될 수 있다. 적어도 부분 자율적인 로봇은 이하 도 5에서 예시적으로 설명된다.
- [0064] 도 5는, 제1 실시예에서 적어도 부분 자율적인 차량(100)에 의해 제공되어 있는, 적어도 부분 자율적인 로봇의 개략도를 보여준다. 또 다른 실시예에서는, 적어도 부분 자율적인 로봇이 정비 로봇, 조립 로봇 또는 고정식 생산 로봇, 대안적으로는 드론과 같은 자율 비행 물체일 수 있다.
- [0065] 적어도 부분 자율적인 차량(100)은 검출 유닛(30)을 포함할 수 있다. 검출 유닛(30)은 예컨대 차량(100)의 주변을 검출하는 카메라일 수 있다. 검출 유닛(30)은 기계 학습 시스템(10)과 연결될 수 있다. 기계 학습 시스템(10)은, 제공된 입력 변수, 예컨대 검출 유닛(30)에 의해 제공된 입력 변수에 따라, 그리고 기계 학습 시스템(10)의 복수의 파라미터에 따라 출력 변수를 결정한다. 출력 변수는 제어 유닛(40)으로 전달될 수 있다.
- [0066] 제어 유닛(40)은, 기계 학습 시스템(10)의 출력 변수에 따라 액추에이터를 제어하며, 바람직하게는, 차량(100)이 비충돌 조작(collision-free maneuver)을 실행하도록 액추에이터를 제어한다. 제1 실시예에서, 액추에이터는 차량(100)의 엔진 또는 브레이크 시스템일 수 있다.
- [0067] 또 다른 일 실시예에서, 부분 자율 로봇은 공구, 작업 기계 또는 생산 로봇일 수 있다. 가공품의 재료가 기계 학습 시스템(10)을 이용하여 분류될 수 있다. 이 경우, 액추에이터는 예컨대 연삭 헤드를 작동시키는 모터일 수 있다.
- [0068] 또한, 차량(100), 특히 부분 자율 로봇은 컴퓨팅 유닛(50) 및 기계 관독 가능 메모리 요소(60)를 포함한다. 메모리 요소(60)에는, 컴퓨팅 유닛(50)에서 명령이 실행되면 컴퓨팅 유닛(50)에 의해 기계 학습 시스템(10)이 작동되게 하는 명령을 포함하는 컴퓨터 프로그램이 저장될 수 있다.

도면

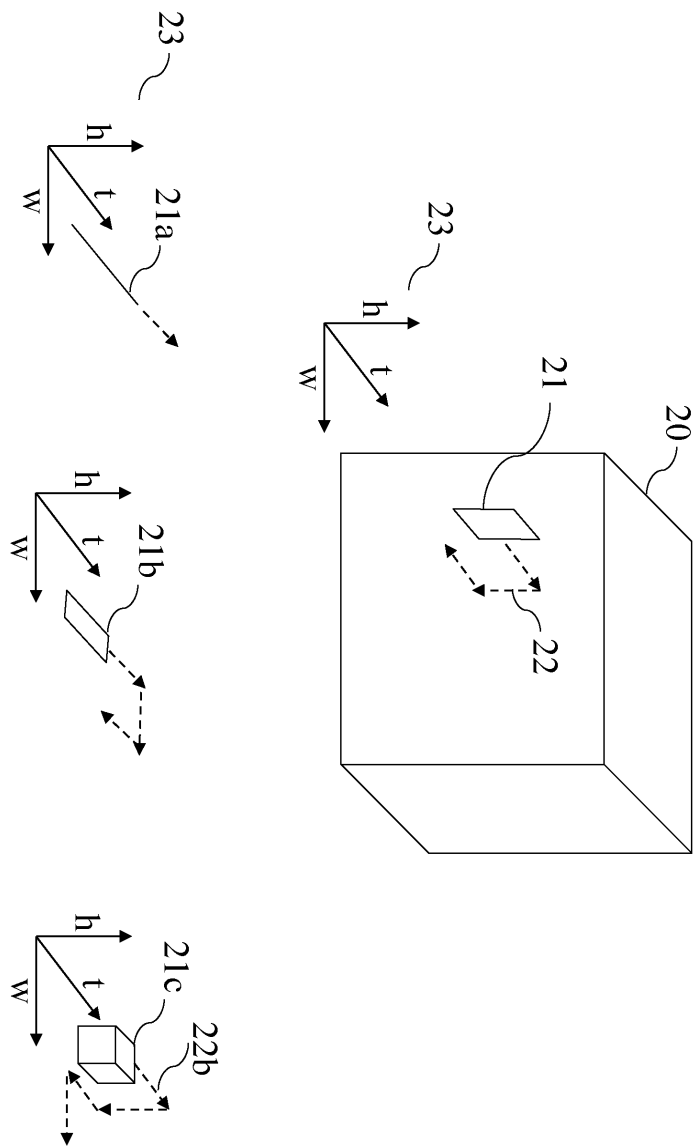
도면1



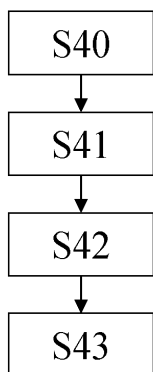
도면2



도면3



도면4



도면5

