



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0160789
(43) 공개일자 2022년12월06일

(51) 국제특허분류(Int. Cl.)
G06N 20/20 (2019.01) G06V 10/10 (2022.01)
(52) CPC특허분류
G06N 20/20 (2021.08)
G06V 10/10 (2022.01)
(21) 출원번호 10-2021-0068859
(22) 출원일자 2021년05월28일
심사청구일자 2021년05월28일

(71) 출원인
한국기술교육대학교 산학협력단
충청남도 천안시 동남구 병천면 충절로 1600 (한
국기술교육대학교내)
(72) 발명자
최강선
충청남도 천안시 동남구 용곡2길 43-11 용곡한라
비발디아파트 109동 1704호
정범채
충청북도 청주시 청원구 우암로63번길 37
(74) 대리인
특허법인주원

전체 청구항 수 : 총 5 항

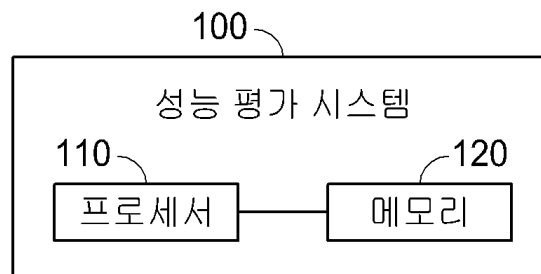
(54) 발명의 명칭 다중 분류 모델의 성능 평가 방법 및 시스템

(57) 요약

본 발명은 다중 분류 모델의 성능 평가 방법 및 시스템에 관한 것이다.

또한, 본 발명은 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와 다중 분류 모델에서 출력된 예측 데이터를 입력받는 입력 단계와, 상기 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고 상기 추출된 점의 위치에 따라 각각의 점에 가중치를 설정하는 가중치 설정 단계와, 실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고 상기 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하는 평가 단계를 구비함으로써, 다중 분류 모델의 성능 평가 시 사람의 주관적 평가와 유사한 평가 결과를 제공할 수 있고, 다중 분류 모델에 대해 보다 정확한 평가 결과를 제공할 수 있다.

대표도 - 도2



(52) CPC특허분류
G06V 30/10 (2022.01)

명세서

청구범위

청구항 1

비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받는 입력 단계;

상기 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고, 상기 추출된 점의 위치에 따라 각각의 점에 가중치를 설정하는 가중치 설정 단계; 및

실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고, 상기 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하는 평가 단계;를 포함하는, 다중 분류 모델의 성능 평가 방법.

청구항 2

제 1항에 있어서,

상기 가중치 설정 단계는,

레벨셋 알고리즘을 이용하여 가중치를 설정하되, 점의 위치가 획의 중앙에 가까울수록 높은 가중치를 설정하는 것을 특징으로 하는, 다중 분류 모델의 성능 평가 방법.

청구항 3

제 1항에 있어서,

상기 평가 단계는,

혼동 행렬(confusion matrix)에 따른 재현율(recall)과 정밀도(precision)의 조화평균을 계산하여 다중 분류 모델의 성능을 평가하는 것을 특징으로 하는, 다중 분류 모델의 성능 평가 방법.

청구항 4

제 3항에 있어서,

상기 평가 단계는,

다중 분류 모델의 성능 평가 이전에, 사전에 설정된 파라미터를 반영하여 혼동 행렬의 참-양성(True Positive, TP), 거짓-양성(False Positive, FP) 및 거짓-음성(False Negative, FN)에 대한 중요도를 서로 다르게 조정하는 것을 특징으로 하는, 다중 분류 모델의 성능 평가 방법.

청구항 5

프로세서; 및

상기 프로세서에 의해 실행 가능한 하나 이상의 인스트럭션을 저장하는 메모리;를 포함하고,

상기 프로세서는, 상기 하나 이상의 인스트럭션을 실행함으로써,

비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받고,

상기 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고, 상기 추출된 점의 위치에 따라 각각의 점에 가중치를 설정하며,

실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고, 상기 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하는, 다중 분류 모델의 성능 평가 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 다중 분류 모델의 성능 평가 방법 및 시스템에 관한 것이다.

배경 기술

[0003] 컴퓨터 그래픽과 컴퓨터비전 분야의 발전과 함께 응용 범위가 넓어짐에 따라 문화 유적 조사와 복원 분야에 3차원 분석 기술이 적극적으로 활용되고 있다.

[0004] 일 예로, 비석에 새겨진 비문을 판독하는 경우, 기존에 주로 사용되던 탁본은 탁본의 숙련도에 따라 결과물이 달라질 수 있고 탁본을 얻기 위해서 비석 표면에 직접 접촉해야 하기 때문에 비석을 훼손할 수 있는 우려가 있지만, 3차원으로 스캔한 비석 데이터를 이용하여 비문을 추출하게 되면, 비석의 표면을 훼손할 우려가 없고, 스캔을 통해 비석에 새겨진 문자와 배경을 고유하게 식별할 수 있으며, 비석에서 획득한 모든 정보를 디지털 데이터로 저장할 수 있다.

[0005] 그리고, 종래에는 다중 분류 모델을 이용하여 비석 데이터로부터 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역을 분류하는 방식으로 비문을 추출하고 있는데, 이에 사용되는 다중 분류 모델이 다양하기 때문에, 각각의 다중 분류 모델에 의한 비문 추출 성능을 객관적으로 비교하고 평가할 필요가 있었다.

[0006] 따라서, 종래에는 비문 추출에 대해 텍스트 추출이나 영상 분할과 유사한 분류 작업으로 간주하여 종래의 성능 평가 지표인 F1 score, 재현율(recall) 및 정밀도(precision)을 활용하였다.

[0007] 그러나, 비석 데이터의 경우에는 일반적인 텍스트나 영상과 다르게 클래스 분포가 편향되어, 글자 획을 나타내는 새김 영역이 배경 영역보다 훨씬 작은 특징을 갖는데, 이런 클래스 불균형에 의해 종래의 성능 평가 지표에 의한 평가 결과가 사람의 주관적 평가와 크게 달라지는 문제점이 있다.

[0008] 예를 들어, 종래의 성능 평가 지표에 의한 평가 결과와 사람의 주관적 평가 간의 차이를 나타내는 도 1을 참조하면, 1-1 내지 3-5에 각각 도시된 3개의 비문에 대해 종래의 성능 평가 지표인 F1 score, 재현율(recall) 및 정밀도(precision)에 의한 평가 결과는, 주관적 평가 기법인 MOS(Mean opinion score)의 평가 결과와 상이한 차이가 있었으며, 예외적으로, 2-1의 비문과 3-2의 비문에 대해서만 유사한 평가 결과가 나타난 것을 확인할 수 있다.

[0009] 참고로, 전술한 MOS는 다중 분류 모델에 의한 비문 추출 성능을 여러 평가자가 각각 점수를 매기고, 여러 평가자의 점수를 평균화한 것이다.

[0010] 이를 해결하고자, 비석 데이터에서 추출한 글자의 품질을 측정할 때 사람의 주관적 요인을 고려하여 주관적 평가와 유사한 결과를 보여주는 추출 비문 인식 지수(Segmented Inscription Recognition Index, SIRI)가 제시되었으나, 종래의 추출 비문 인식 지수는 획의 중심 영역과 경계 영역, 주 영역으로만 나누어 가중치를 설정하기 때문에, 같은 영역 안에서는 획의 두께 변화와 관계없이 동일한 가중치가 설정된다.

[0011] 즉, 종래의 추출 비문 인식 지수는 사람의 주관적 인식에서 글자의 두께가 얇아지거나 두꺼워지는 경우에는 주관적 평가의 차이가 크지 않으나, 획이 사라지거나 잡음이 지저분하게 발생될 때 평가가 매우 낮아진다는 점을 반영하지 못하므로, 최종적인 평가 결과가 여전히 사람의 주관적 평가와 달라질 수 있는 문제점이 있다.

발명의 내용

해결하려는 과제

- [0013] 본 발명은 상기와 같은 문제점을 감안하여 안출된 것으로, 비석 데이터에서 비문을 추출하는 다중 분류 모델의 성능을 평가하되, 사람의 주관적 평가와 유사한 평가 결과를 제공할 수 있는 다중 분류 모델의 성능 평가 방법 및 시스템을 제공하는데 그 목적이 있다.
- [0014] 본 발명의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0016] 상기와 같은 목적을 달성하기 위한 본 발명은 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받는 입력 단계; 상기 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고, 상기 추출된 점의 위치에 따라 각각의 점에 가중치를 설정하는 가중치 설정 단계; 및 실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고, 상기 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하는 평가 단계;를 포함하는, 다중 분류 모델의 성능 평가 방법을 제공한다.
- [0017] 바람직한 실시예에 있어서, 상기 가중치 설정 단계는, 레벨셋 알고리즘을 이용하여 가중치를 설정하되, 점의 위치가 획의 중앙에 가까울수록 높은 가중치를 설정한다.
- [0018] 바람직한 실시예에 있어서, 상기 평가 단계는, 혼동 행렬(confusion matrix)에 따른 재현율(recall)과 정밀도(precision)의 조화평균을 계산하여 다중 분류 모델의 성능을 평가한다.
- [0019] 바람직한 실시예에 있어서, 상기 평가 단계는, 다중 분류 모델의 성능 평가 이전에, 사전에 설정된 파라미터를 반영하여 혼동 행렬의 참-양성(True Positive, TP), 거짓-양성(False Positive, FP) 및 거짓-음성(False Negative, FN)에 대한 중요도를 서로 다르게 조정한다.
- [0021] 또한, 본 발명은 프로세서; 및 상기 프로세서에 의해 실행 가능한 하나 이상의 인스트럭션을 저장하는 메모리;를 포함하고, 상기 프로세서는, 상기 하나 이상의 인스트럭션을 실행함으로써, 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받고, 상기 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고, 상기 추출된 점의 위치에 따라 각각의 점에 가중치를 설정하며, 실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고, 상기 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하는, 다중 분류 모델의 성능 평가 시스템을 제공한다.

발명의 효과

- [0023] 전술한 과제해결 수단에 의해 본 발명은 사전에 설정된 파라미터를 반영하여 혼동 행렬의 참-양성(True Positive, TP), 거짓-양성(False Positive, FP) 및 거짓-음성(False Negative, FN)의 중요도를 조정하고, 중요도가 조정된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가함으로써, 다중 분류 모델의 성능 평가 시 사람의 주관적 평가와 유사한 평가 결과를 제공할 수 있는 효과가 있다.
- [0024] 또한, 본 발명은 새김 영역에 위치하는 점의 위치에 따라 가중치를 설정하고, 설정된 가중치를 실측 데이터와 예측 데이터의 비교 결과에 반영하여 혼동 행렬을 생성함으로써, 다중 분류 모델에 대해 보다 정확한 평가 결과를 제공할 수 있는 효과가 있다.

도면의 간단한 설명

- [0026] 도 1은 종래의 성능 평가 지표에 의한 평가 결과와 사람의 주관적 평가 간의 차이를 나타내는 도면.
- 도 2는 본 발명의 일실시예에 따른 성능 평가 시스템을 설명하기 위한 도면.
- 도 3은 본 발명의 일실시예에 따른 성능 평가 시스템의 프로세서에 의해 수행되는 기능을 설명하기 위한

블록도.

도 4는 가중치가 설정된 새김 영역을 나타내는 도면.

도 5는 혼동 행렬에 따른 참-양성(TP), 거짓-양성(FP), 참-음성(TN) 및 거짓-음성(FN)을 설명하기 위한 도면.

도 6은 본 발명의 일실시예에 따른 성능 평가 방법을 설명하기 위한 도면.

발명을 실시하기 위한 구체적인 내용

- [0027] 하기의 설명에서 본 발명의 특정 상세들이 본 발명의 전반적인 이해를 제공하기 위해 나타나 있는데, 이들 특정 상세들 없이 또한 이들의 변형에 의해서도 본 발명이 용이하게 실시될 수 있다는 것은 이 기술분야에서 통상의 지식을 가진 자에게 자명할 것이다.
- [0028] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도 2 내지 도 6을 참조하여 상세히 설명하되, 본 발명에 따른 동작 및 작용을 이해하는데 필요한 부분을 중심으로 설명한다.
- [0030] 도 2는 본 발명의 일실시예에 따른 성능 평가 시스템을 설명하기 위한 도면이다.
- [0031] 도 2를 참조하면, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 퍼스널 컴퓨터 및 노트북을 포함하는 범용통신단말에 탑재되거나, 해당 범용통신단말에 설치되는 프로그램으로 구현될 수 있다.
- [0032] 이러한, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 비석 데이터로부터 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역을 분류하여 비문을 추출하는 다중 분류 모델의 성능을 평가하기 위한 것으로, 실측(Ground Truth) 데이터와 다중 분류 모델에서 출력된 예측 데이터를 비교하여 다중 분류 모델의 성능을 평가할 수 있다.
- [0033] 특히, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 다중 분류 모델의 성능 평가 시 사람의 주관적 평가와 유사한 평가 결과를 제공할 수 있으며, 종래의 성능 평가 지표인 F1 score, 재현율(recall) 및 정밀도(precision)에 비해 상대적으로 정확한 평가 결과를 제공할 수 있다.
- [0034] 참고로, 진술한 실측 데이터는 사람이 컴퓨터를 이용하여 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링한 것이다.
- [0035] 한편, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 하드웨어 구성으로, 프로세서(110) 및 메모리(120)를 포함하여 구성될 수 있다.
- [0036] 프로세서(110)는 하나 이상으로 포함될 수 있고, 메모리(120)에 저장된 하나 이상의 인스트럭션 또는 프로그램들을 실행할 수 있으며, 아울러, 성능 평가 시스템(100)의 전반적인 동작을 제어할 수도 있다.
- [0037] 메모리(120)는 프로세서(110)에 의해 실행 가능한 하나 이상의 인스트럭션을 저장하거나 프로세서(110)에 의해 실행되는 프로그램들을 저장할 수 있고, 입/출력되는 데이터들을 저장할 수도 있다.
- [0038] 구체적으로, 메모리(120)는 프로세서(110)의 적어도 하나의 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 메모리(120)는 소프트웨어 및/또는 프로그램을 저장할 수 있다. 메모리(120)는 커널, 미들웨어, 어플리케이션 프로그래밍 인터페이스(API), 및/또는 어플리케이션 프로그램 등을 포함할 수 있다. 커널, 미들웨어, 또는 API의 적어도 일부는, 운영 시스템으로 지칭될 수 있다. 커널은, 예를 들면, 다른 프로그램들(예: 미들웨어, API, 또는 어플리케이션 프로그램)에 구현된 동작 또는 기능을 실행하는 데 사용되는 시스템 리소스들(예: 버스, 프로세서(110), 또는 메모리(120) 등)을 제어 또는 관리할 수 있다. 또한, 커널은 미들웨어, API, 또는 어플리케이션 프로그램에서 서버의 개별 구성요소에 접근함으로써, 시스템 리소스들을 제어 또는 관리할 수 있는 인터페이스를 제공할 수 있다.
- [0039] 이하에서는, 본 발명의 일실시예에 따른 성능 평가 시스템(100)에서 프로세서(110)에 의해 수행되는 기능들을 구체적으로 설명한다.
- [0041] 도 3은 본 발명의 일실시예에 따른 성능 평가 시스템의 프로세서에 의해 수행되는 기능을 설명하기 위한 블록도이고, 도 4는 가중치가 설정된 새김 영역을 나타내는 도면이며, 도 5는 혼동 행렬에 따른 참-양성(TP), 거짓-양

성(FP), 참-음성(TN) 및 거짓-음성(FN)을 설명하기 위한 도면이다.

- [0042] 참고로, 도 3에 도시된 블록들의 일부 또는 전부는, 특정 기능을 실행하는 하드웨어 및/또는 소프트웨어 구성들로 구현될 수 있다. 도 3에 도시된 블록들이 수행하는 기능들은, 하나 이상의 마이크로프로세서에 의해 구현되거나, 해당 기능을 위한 회로 구성들에 의해 구현될 수 있다. 도 3에 도시된 블록들의 일부 또는 전부는 프로세서(110)에서 실행되는 다양한 프로그래밍 언어 또는 스크립트 언어로 구성된 소프트웨어 모듈일 수 있다.
- [0043] 먼저, 도 3을 참조하면, 본 발명의 일실시예에 따른 성능 평가 시스템(100)의 프로세서(110)는 입력부(111), 가중치 설정부(112) 및 평가부(113)를 포함할 수 있다.
- [0044] 상기 입력부(111)는 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받는다.
- [0045] 이를 위한, 입력부(111)는 실측 데이터의 입력을 위한 유저 인터페이스(User Interface, UI)와 예측 데이터의 입력을 위한 유저 인터페이스를 제공할 수도 있다.
- [0046] 상기 가중치 설정부(112)는 새김 영역에 위치한 점을 대상으로 가중치를 설정한다.
- [0047] 이러한, 가중치 설정부(112)는 입력부(111)에 입력된 실측 데이터를 기준으로 하여 새김 영역에 대한 정보를 추출하고, 추출된 새김 영역에 위치한 점에 가중치를 설정할 수 있다.
- [0048] 이때, 가중치 설정부(112)는 레벨셋(level set) 알고리즘을 이용하여 새김 영역의 점에 가중치를 설정할 수 있는데, 특히, 새김 영역의 경계에 인접한 점에는 상대적으로 낮은 가중치를 설정하고, 새김 영역의 중앙, 즉, 획의 중앙에 가까울수록 높은 가중치를 설정하도록 구비될 수 있다.
- [0049] 예컨대, 도 4의 (a)에는 새김 영역과 배경 영역으로 구분된 실측 데이터가 도시되어 있는데, 이를 대상으로 가중치를 설정하면, 도 4의 (b)에 도시된 바와 같이, 획의 중앙에는 '8'에 해당하는 가중치가 설정되어 노란 색상으로 표시되고, 새김 영역의 경계에 인접하면 '0' 내지 '4'에 해당하는 가중치가 설정될 수 있다.
- [0050] 또한, 전술한 레벨셋 알고리즘은 새김 영역의 경계를 추적하고 새김 영역의 경계에 위치한 점에서 멀수록 가중치가 증가하게 함으로써, 새김 영역 내에서 획의 중앙에 위치한 점에 상대적으로 높은 가중치를 설정할 수 있다. 참고로, 전술한 레벨셋 알고리즘은 이미 알려진 기술이므로 더 자세한 설명은 생략한다.
- [0051] 상기 평가부(113)는 실측 데이터와 예측 데이터를 비교하여 다중 분류 모델의 성능을 평가한다.
- [0052] 여기서, 평가부(113)는 실측 데이터와 예측 데이터의 비교 결과에 가중치 설정부(112)에 의해 설정된 가중치를 반영하여 혼동 행렬(confusion matrix)을 생성하고, 생성된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가할 수 있다.
- [0053] 또한, 평가부(113)는 실측 데이터와 예측 데이터의 비교 결과에 따라 참-양성(True Positive, TP), 거짓-양성(False Positive, FP), 참-음성(True Negative, TN) 및 거짓-음성(False Negative, FN)으로 구분되는 혼동 행렬을 생성할 수 있다.
- [0054] 참고로, 도 5에 도시된 바와 같이, 참-양성은 글자 획을 올바르게 추정한 것을 나타내고(1), 거짓-양성은 배경을 글자 획으로 잘못 추정한 것을 나타내며(3), 참-음성은 배경을 올바르게 추정한 것을 나타내고(4), 거짓-음성은 글자 획을 배경으로 잘못 추정한 것을 나타낸다(2).
- [0055] 그리고, 평가부(113)는 혼동 행렬이 생성되면, 실험을 통해 사람의 주관적인 평가 결과를 근거로 사전에 설정한 파라미터를 반영하여 참-양성, 거짓-양성 및 거짓-음성에 대한 중요도를 서로 다르게 조정할 수 있다.
- [0056] 또한, 평가부(113)는 아래의 수학적 식 1을 이용하여 참-양성, 거짓-양성 및 거짓-음성에 파라미터를 적용할 수 있다.

[0058] [수학식 1]

$$\begin{cases} TP_w = TP \cdot 1, \\ FP_w = FP \cdot w_{FP}, \\ FN_w = w_{FN} \sum \log(\phi(FN)) \end{cases}$$

[0061] 여기서, TP_w 는 가중된 참-양성 값이고, TP 는 본래의 참-양성 값이며, FP_w 는 가중된 거짓-양성 값이고, FP 는 본래의 거짓-양성 값이며, w_{FP} 는 거짓-양성 값에 적용되는 파라미터이고, FN_w 는 가중된 거짓-음성 값이며, w_{FN} 은 거짓-음성 값에 적용되는 파라미터이고, FN 은 본래의 거짓-음성 값이다.

[0062] 즉, 평가부(113)는 노이즈에 해당하는 거짓-양성이 확과의 거리와 관계없이 사람에 의한 주관적인 평가에 악영향을 주는 점을 반영하여 거짓-양성의 중요도를 조정하고, 확의 끊김에 해당하는 거짓-음성의 경우에도 사람에 의한 주관적인 평가에 악영향을 주므로, 기설정된 파라미터를 적용하여 거짓-음성에 대한 중요도를 조정하도록 구비된다.

[0063] 그리고, 평가부(113)는 중요도를 조정된 혼동 행렬에 기반하여, 재현율(recall)과 정밀도(precision)의 조화평균을 계산하고 계산된 조화평균 값을 다중 분류 모델의 성능을 평가하는 평가 결과로서 출력할 수 있다.

[0064] 또한, 평가부(113)는 아래의 수학식 2를 이용하여 조화평균을 계산할 수 있다.

[0066] [수학식 2]

$$SIRI = \frac{TP_w}{TP_w + \frac{FP_w + FN_w}{2}}$$

[0069] 여기서, SIRI는 다중 분류 모델의 성능에 대한 조화평균 값이고, TP_w 는 가중된 참-양성 값이며, FP_w 는 가중된 거짓-양성 값이고, FN_w 는 가중된 거짓-음성 값이다.

[0070] 즉, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 사람의 주관적인 평가 결과를 근거로 설정된 파라미터를 적용하여 혼동 행렬의 중요도를 조정하고, 중요도가 조정된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가하기 때문에, 다중 분류 모델의 성능 평가 시 사람의 주관적 평가와 유사한 평가 결과를 제공할 수 있다.

[0071] 또한, 본 발명의 일실시예에 따른 성능 평가 시스템(100)은 새김 영역에 위치하는 점의 위치에 따라 가중치를 설정하고, 설정된 가중치를 실측 데이터와 예측 데이터의 비교 결과에 반영하여 혼동 행렬을 생성하기 때문에, 다중 분류 모델에 대한 정확한 평가 결과를 제공할 수 있다.

[0072] 이와 같은, 본 발명은 프로세서(110)가 메모리(120)에 저장된 적어도 하나의 실행 가능한 명령어 또는 적어도 하나의 실행 가능한 프로그램을 실행함으로써 구현될 수 있으며, 후술하는 본 발명의 일실시예에 따른 성능 평가 방법의 경우에도 동일하게, 메모리(120)에 저장된 적어도 하나의 실행 가능한 명령어 또는 적어도 하나의 실행 가능한 프로그램을 프로세서(110)가 실행함으로써 구현될 수 있다.

[0074] 도 3은 본 발명의 일실시예에 따른 성능 평가 방법을 설명하기 위한 도면이다.

[0075] 도 3을 참조하면, 본 발명의 일실시예에 따른 성능 평가 시스템(100)의 프로세서(110)는, 먼저, 비문의 획을 구성하는 새김 영역과 배경을 구성하는 배경 영역으로 라벨링된 실측(Ground Truth) 데이터와, 다중 분류 모델에서 출력된 예측 데이터를 입력받는다(S100).

[0076] 그 다음, 프로세서(110)는 입력된 실측 데이터로부터 새김 영역에 위치하는 점(vertex)에 대한 정보를 추출하고, 추출된 점의 위치에 따라 각각의 점에 가중치를 설정한다(S200).

[0077] 또한, S200 단계에서 프로세서(110)는, 레벨셋 알고리즘을 이용하여 가중치를 설정하되, 점의 위치가 획의 중앙

에 가까울수록 높은 가중치를 설정하도록 구비될 수 있다.

[0078] 그 다음, 프로세서(110)는 실측 데이터와 예측 데이터를 비교하여 다중 분류 모델의 성능을 평가한다(S300).

[0079] 또한, 프로세서(110)는 실측 데이터와 예측 데이터의 비교 결과에 상기 설정된 가중치를 반영하여 혼동 행렬을 생성하고(S310), 사전에 설정된 파라미터를 반영하여 혼동 행렬의 참-양성(True Positive, TP), 거짓-양성(False Positive, FP) 및 거짓-음성(False Negative, FN)에 대한 중요도를 서로 다르게 조정하며(S320), 중요도가 조정된 혼동 행렬을 기초로 다중 분류 모델의 성능을 평가할 수 있다(S330).

[0080] 아울러, S330 단계에서 프로세서(110)는 혼동 행렬(confusion matrix)에 따른 재현율(recall)과 정밀도(precision)의 조화평균을 계산하고, 계산된 조화평균 값을 다중 분류 모델의 성능을 평가하는 평가 결과로서 출력할 수 있다.

[0082] 본 발명의 실시예와 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어로 직접 구현되거나, 하드웨어에 의해 실행되는 소프트웨어 모듈로 구현되거나, 또는 이들의 결합에 의해 구현될 수 있다.

[0083] 소프트웨어 모듈은 RAM(Random Access Memory), ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리(Flash Memory), 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터 판독 가능한 저장 매체에 상주할 수 있다.

[0084] 본 발명의 구성 요소들은 비일시적 저장매체에 저장되고, 프로세서를 포함하는 컴퓨터에서 실행되어 전술한 방법 또는 알고리즘의 단계들을 수행하며, 하드웨어인 컴퓨터와 결합되어 실행되기 위해 컴퓨터 프로그램(또는 애플리케이션)으로 구현되어 컴퓨터 판독 가능한 저장 매체에 저장될 수 있다. 본 발명의 구성 요소들은 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있으며, 이와 유사하게, 실시 예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다.

[0085] 이상에서는 본 발명의 바람직한 실시예를 예시적으로 설명하였으나, 본 발명의 범위는 이와 같은 특정 실시예에만 한정되는 것은 아니며, 특허청구범위에 기재된 범주 내에서 적절하게 변경 가능한 것이다.

부호의 설명

[0087] 100 : 성능 평가 시스템

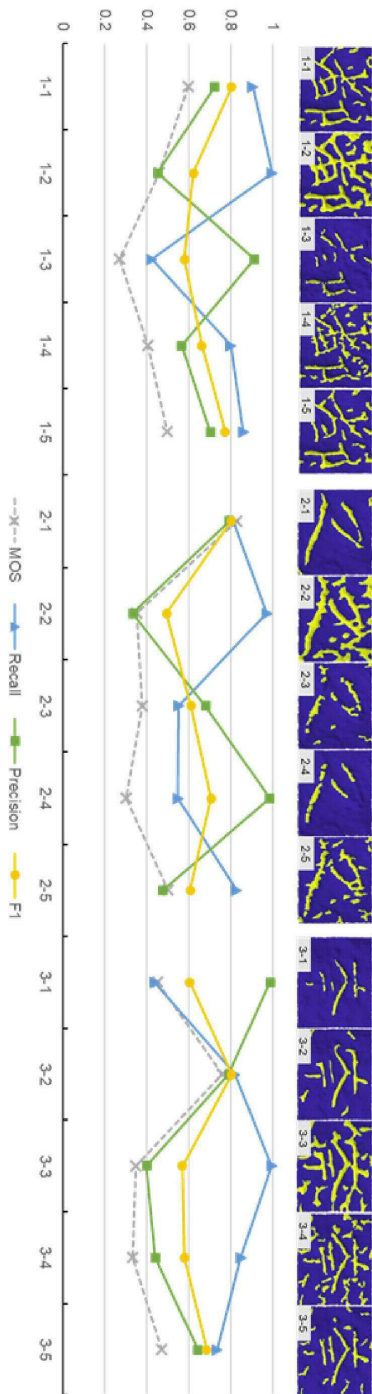
110 : 프로세서 111 : 입력부

112 : 가중치 설정부 113 : 평가부

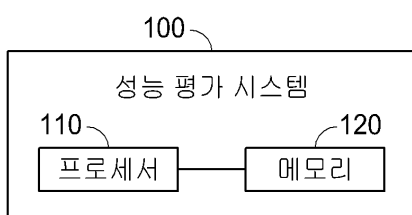
120 : 메모리

도면

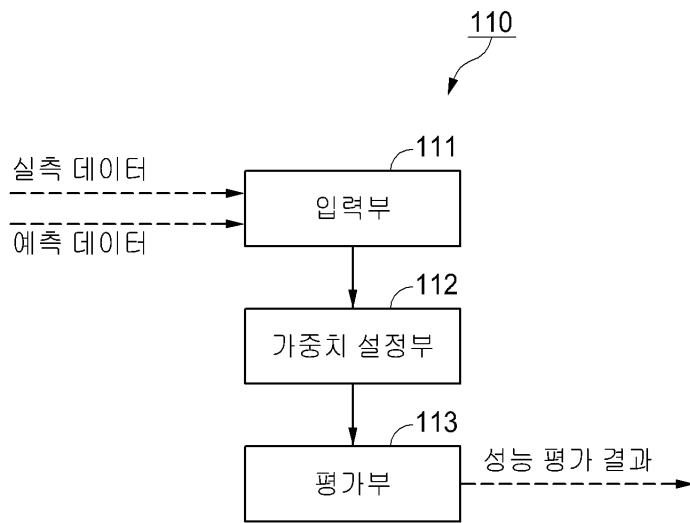
도면1



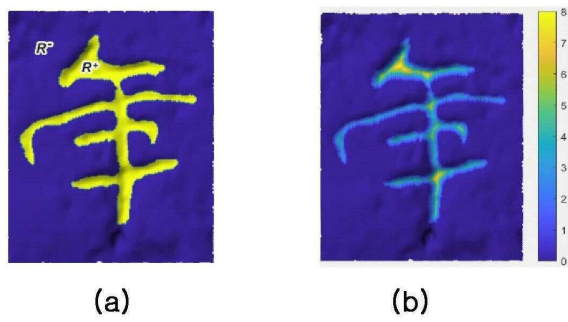
도면2



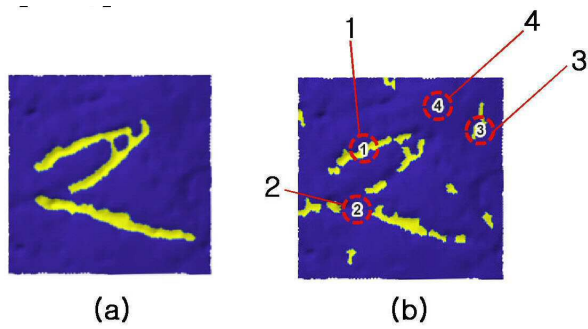
도면3



도면4



도면5



도면6

