



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0121506
(43) 공개일자 2022년09월01일

(51) 국제특허분류(Int. Cl.)

G06F 16/35 (2019.01) G06F 16/33 (2019.01)
G06F 16/93 (2019.01) G06K 9/62 (2022.01)
G06N 20/00 (2019.01)

(52) CPC특허분류

G06F 16/35 (2019.01)
G06F 16/3334 (2019.01)

(21) 출원번호 10-2021-0025666

(22) 출원일자 2021년02월25일

심사청구일자 2021년02월25일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

김재영

서울특별시 구로구 구로동로 148, 본관3층 영상의학연구소 (구로동, 고려대부속구로병원)

전은태

경기도 안산시 단원구 화정천서로 193, 303호 (초지동)

정진만

서울특별시 강서구 강서로 348, 132동 501호 (내발산동, 우장산힐스테이트)

(74) 대리인

특허법인 무한

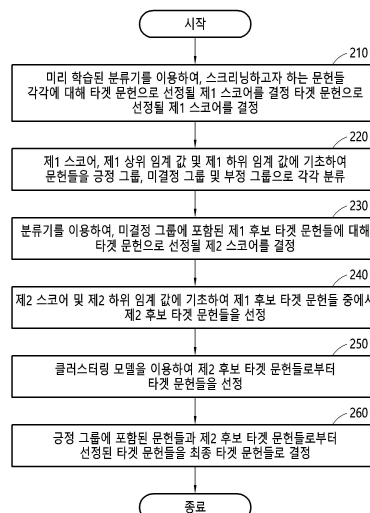
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 문헌 스크리닝 방법 및 장치

(57) 요약

문헌 스크리닝 방법 및 장치가 개시된다. 문헌 스크리닝 방법은 미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하는 단계, 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류하는 단계, 분류기를 이용하여, 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 타겟 문헌으로 선정될 제2 스코어를 결정하는 단계, 제2 스코어 및 제2 하위 임계 값에 기초하여 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하는 단계, 클러스터링 모델을 이용하여 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하는 단계 및 긍정 그룹에 포함된 문헌들과 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하는 단계를 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류

G06F 16/93 (2019.01)

G06K 9/6218 (2022.01)

G06N 20/00 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	9991007421
과제번호	202013C13-02
부처명	다부처
과제관리(전문)기관명	(재단)범부처전주기의료기기연구개발사업단
연구사업명	범부처전주기의료기기연구개발사업(R&D)
연구과제명	(참여1)노인성 신경질환 모니터링을 위한 다중영상 기반 키오스크 시스템
기 여 율	1/1
과제수행기관명	고려대학교 산학협력단
연구기간	2020.09.01 ~ 2021.02.28

명세서

청구범위

청구항 1

문헌 스크리닝 방법에 있어서,

미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하는 단계;

상기 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 상기 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류하는 단계;

상기 분류기를 이용하여, 상기 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 상기 타겟 문헌으로 선정될 제2 스코어를 결정하는 단계;

상기 제2 스코어 및 제2 하위 임계 값에 기초하여 상기 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하는 단계;

클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하는 단계; 및

상기 긍정 그룹에 포함된 문헌들과 상기 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하는 단계를 포함하는 문헌 스크리닝 방법.

청구항 2

제1항에 있어서,

상기 분류기는,

기존에 스크리닝된 문헌들 및 상기 부정 그룹에 포함된 문헌들 중 적어도 하나에 기초하여 학습된 것을 특징으로 하는 문헌 스크리닝 방법.

청구항 3

제1항에 있어서,

상기 제1 스코어를 결정하는 단계는,

상기 분류기에 리브 원 아웃 교차 검증(leave-one-out cross-validation)을 적용하는 단계;

상기 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치(keyword weighting)를 적용하는 단계; 및

상기 키워드 가중치가 적용된 분류기를 이용하여 상기 제1 스코어를 산출하는 단계를 포함하는 문헌 스크리닝 방법.

청구항 4

제1항에 있어서,

상기 제1 상위 임계 값 및 상기 제1 하위 임계 값은,

상기 분류기에 적용된 리브 원 아웃 교차 검증의 검증 결과에 기초하여 결정되는 것을 특징으로 하는 문헌 스크리닝 방법.

청구항 5

제1항에 있어서,

상기 제2 후보 타겟 문헌들을 선정하는 단계는,

상기 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용하는 단계; 및

상기 리브 원 아웃 교차 검증의 검증 결과에 기초하여 상기 제2 하위 임계 값을 결정하는 단계를 포함하는 문헌 스크리닝 방법.

청구항 6

제1항에 있어서,

상기 클러스터링 모델은,

기존에 스크리닝된 문헌들 및 상기 제2 후보 타겟 문헌들에 기초하여 학습되고, 메트릭 K-평균 클러스터링 모델(metric K-means clustering model)인 것을 특징으로 하는 문헌 스크리닝 방법.

청구항 7

제1항에 있어서,

상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하는 단계는,

상기 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하는 단계; 및

상기 클러스터링의 결과에 기초하여 상기 제2 후보 타겟 문헌들로부터 상기 타겟 문헌들을 선정하는 단계를 포함하는 문헌 스크리닝 방법.

청구항 8

제1항에 있어서,

상기 분류기는,

최대 엔트로피 분류기(maximum entropy classifier)인 것을 특징으로 하는 문헌 스크리닝 방법.

청구항 9

제1항 내지 제8항 중 어느 하나의 항의 방법을 실행시키기 위한 인스트럭션들을 저장하는 컴퓨터 판독 가능한 저장 매체.

청구항 10

문헌 스크리닝 장치에 있어서,

메모리 및 프로세서를 포함하고,

상기 메모리는 상기 프로세서에 의해 실행 가능한 인스트럭션들(instructions)을 저장하고,

상기 인스트럭션들이 상기 프로세서에 의해 실행될 때, 상기 프로세서는 상기 문헌 스크리닝 장치가,

미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하고,

상기 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 상기 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류하고,

상기 분류기를 이용하여, 상기 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 상기 타겟 문헌으로 선정될 제2 스코어를 결정하고,

상기 제2 스코어 및 제2 하위 임계 값에 기초하여 상기 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하고,

클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하고,

상기 긍정 그룹에 포함된 문헌들과 상기 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하도록 상기 문헌 스크리닝 장치를 제어하는 문헌 스크리닝 장치.

청구항 11

제10항에 있어서,

상기 분류기는,

기존에 스크리닝된 문헌들 및 상기 부정 그룹에 포함된 문헌들 중 적어도 하나에 기초하여 학습된 것을 특징으로 하는 문헌 스크리닝 장치.

청구항 12

제10항에 있어서,

상기 프로세서는 상기 문헌 스크리닝 장치가,

상기 분류기에 리브 원 아웃 교차 검증(leave-one-out cross-validation)을 적용하고,

상기 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치(keyword weighting)를 적용하고,

상기 키워드 가중치가 적용된 분류기를 이용하여 상기 제1 스코어를 산출하도록 상기 문헌 스크리닝 장치를 제어하는 문헌 스크리닝 장치.

청구항 13

제10항에 있어서,

상기 제1 상위 임계 값 및 상기 제1 하위 임계 값은,

상기 분류기에 적용된 리브 원 아웃 교차 검증의 검증 결과에 기초하여 결정되는 것을 특징으로 하는 문헌 스크리닝 장치.

청구항 14

제10항에 있어서,

상기 프로세서는 상기 문헌 스크리닝 장치가,

상기 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용하고,

상기 리브 원 아웃 교차 검증의 검증 결과에 기초하여 상기 제2 하위 임계 값을 결정하도록 상기 문헌 스크리닝 장치를 제어하는 문헌 스크리닝 장치.

청구항 15

제10항에 있어서,

상기 클러스터링 모델은,

기존에 스크리닝된 문헌들 및 상기 제2 후보 타겟 문헌들에 기초하여 학습되고, 메트릭 K-평균 클러스터링 모델(metric K-means clustering model)인 것을 특징으로 하는 문헌 스크리닝 장치.

청구항 16

제10항에 있어서,

상기 프로세서는 상기 문헌 스크리닝 장치가,

상기 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하고,

상기 클러스터링의 결과에 기초하여 상기 제2 후보 타겟 문헌들로부터 상기 타겟 문헌들을 선정하도록 상기 문헌 스크리닝 장치를 제어하는 문헌 스크리닝 장치.

청구항 17

제10항에 있어서,

상기 분류기는,

최대 엔트로피 분류기(maximum entropy classifier)인 것을 특징으로 하는 문헌 스크리닝 장치.

발명의 설명

기술 분야

[0001] 아래 실시예들은 문헌 스크리닝 기술에 관한 것이다.

배경 기술

[0002] 메타 분석을 포함하는 체계적 문헌 고찰은 연구 질문에 맞는 문헌 검색, 선정/배제 기준에 따른 연구 설정, 자료 추출, 연구의 비뚤림 위험 평가, 자료 합성 및 결론 도출을 포함하는 다양한 과정이 필요하다. 또한, 체계적 문헌 고찰은 최소 두 명의 연구자가 문헌 스크리닝, 문헌의 최종 선택, 자료 추출 및 질 평가 과정에 참여하여 합의를 유도해야 하므로 연구자의 많은 시간과 노력을 필요로 한다. 최근에 다양한 연구가 수행되고, 연구와 관련된 문헌의 수가 기하급수적으로 증가하고 있어, 연구자들이 문헌 스크리닝을 위해 많은 시간과 노력을 들이고 있다. 문헌 스크리닝은 연구의 결론 도출에 영향을 줄 수 있어 정확성이 요구되는 과정이다.

[0003] 근거 중심의 보건의료 연구 분야에서 메타분석을 포함하는 체계적 문헌 고찰은 최상위 근거가 된다. 또한, 임상진료지침 및 최근 의료기술평가 분야에서 체계적 문헌 고찰의 결과물은 인증을 위한 근거 자료가 되어 체계적 문헌 고찰의 활용도가 점차 증가하고 있다. 근거 중심 보건의료 연구 분야에서 체계적 문헌 고찰(메타분석 포함)이 주요 연구 방법으로 자리매김하고 있지만, 표준적 방법론의 부재, 체계적 문헌 고찰 간의 변이성, 보고 내용의 다양성 등의 문제점이 있다.

[0004] 최근에는 관련 논문 및 연구의 급격한 증가로 관련 분야에 대한 체계적 문헌 고찰(메타분석 포함)의 시행주기가 점차 빨라지고 있다. 국내의 경우에는 특히 의료기술평가 방법 전문가 풀(pool)이 적을 뿐만 아니라, 의료기술평가 방법론에 대한 역사가 깊지 않아 체계적 문헌 고찰의 방법론 발전이 필요하다. 따라서, 연구 결과의 일관성을 증대시키고, 연구 결과가 국제적 표준에 부합하도록 하며 국제적 표준이 될 수 있는 연구 방법의 개발이 필요하다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0005] 일 실시예에 따른 문헌 스크리닝 방법은 미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하는 단계; 상기 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 상기 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류하는 단계; 상기 분류기를 이용하여, 상기 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 상기 타겟 문헌으로 선정될 제2 스코어를 결정하는 단계; 상기 제2 스코어 및 제2 하위 임계 값에 기초하여 상기 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하는 단계; 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하는 단계; 및 상기 긍정 그룹에 포함된 문헌들과 상기 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하는 단계를 포함할 수 있다.

[0006] 상기 분류기는, 기존에 스크리닝된 문헌들 및 상기 부정 그룹에 포함된 문헌들 중 적어도 하나에 기초하여 학습된 것을 특징으로 할 수 있다.

[0007] 상기 제1 스코어를 결정하는 단계는, 상기 분류기에 리브 원 아웃 교차 검증(leave-one-out cross-validation)을 적용하는 단계; 상기 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치(Keyword weighting)를 적용하는 단계; 및 상기 키워드 가중치가 적용된 분류기를 이용하여 상기 제1 스코어를 산출하는 단계를 포함할 수 있다.

[0008] 상기 제1 상위 임계 값 및 상기 제1 하위 임계 값은, 상기 분류기에 적용된 리브 원 아웃 교차 검증의 검증 결

과에 기초하여 결정되는 것을 특징으로 할 수 있다.

- [0009] 상기 제2 후보 타겟 문헌들을 선정하는 단계는, 상기 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용하는 단계; 및 상기 리브 원 아웃 교차 검증의 검증 결과에 기초하여 상기 제2 하위 임계 값을 결정하는 단계를 포함할 수 있다.
- [0010] 상기 클러스터링 모델은, 기존에 스크리닝된 문헌들 및 상기 제2 후보 타겟 문헌들에 기초하여 학습되고, 메트릭 K-평균 클러스터링 모델(metric K-means clustering model)인 것을 특징으로 할 수 있다.
- [0011] 상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하는 단계는, 상기 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하는 단계; 및 상기 클러스터링의 결과에 기초하여 상기 제2 후보 타겟 문헌들로부터 상기 타겟 문헌들을 선정하는 단계를 포함할 수 있다.
- [0012] 상기 분류기는, 최대 엔트로피 분류기(maximum entropy classifier)인 것을 특징으로 할 수 있다.
- [0013] 일 실시예에 따른 문헌 스크리닝 장치는 메모리 및 프로세서를 포함하고, 상기 메모리는 상기 프로세서에 의해 실행 가능한 인스트럭션들(instructions)을 저장하고, 상기 인스트럭션들이 상기 프로세서에 의해 실행될 때, 상기 프로세서는 상기 문헌 스크리닝 장치가, 미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하고, 상기 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 상기 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류하고, 상기 분류기를 이용하여, 상기 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 상기 타겟 문헌으로 선정될 제2 스코어를 결정하고, 상기 제2 스코어 및 제2 하위 임계 값에 기초하여 상기 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하고, 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하고, 상기 긍정 그룹에 포함된 문헌들과 상기 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하도록 상기 문헌 스크리닝 장치를 제어할 수 있다.
- [0014] 상기 분류기는, 기존에 스크리닝된 문헌들 및 상기 부정 그룹에 포함된 문헌들 중 적어도 하나에 기초하여 학습된 것을 특징으로 할 수 있다.
- [0015] 상기 프로세서는 상기 문헌 스크리닝 장치가, 상기 분류기에 리브 원 아웃 교차 검증(leave-one-out cross-validation)을 적용하고, 상기 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치(Keyword weighting)를 적용하고, 상기 키워드 가중치가 적용된 분류기를 이용하여 상기 제1 스코어를 산출하도록 상기 문헌 스크리닝 장치를 제어할 수 있다.
- [0016] 상기 제1 상위 임계 값 및 상기 제1 하위 임계 값은, 상기 분류기에 적용된 리브 원 아웃 교차 검증의 검증 결과에 기초하여 결정되는 것을 특징으로 할 수 있다.
- [0017] 상기 프로세서는 상기 문헌 스크리닝 장치가, 상기 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용하고, 상기 리브 원 아웃 교차 검증의 검증 결과에 기초하여 상기 제2 하위 임계 값을 결정하도록 상기 문헌 스크리닝 장치를 제어할 수 있다.
- [0018] 상기 클러스터링 모델은, 기존에 스크리닝된 문헌들 및 상기 제2 후보 타겟 문헌들에 기초하여 학습되고, 메트릭 K-평균 클러스터링 모델(metric K-means clustering model)인 것을 특징으로 할 수 있다.
- [0019] 상기 프로세서는 상기 문헌 스크리닝 장치가, 상기 클러스터링 모델을 이용하여 상기 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하고, 상기 클러스터링의 결과에 기초하여 상기 제2 후보 타겟 문헌들로부터 상기 타겟 문헌들을 선정하도록 상기 문헌 스크리닝 장치를 제어할 수 있다.
- [0020] 상기 분류기는, 최대 엔트로피 분류기(maximum entropy classifier)인 것을 특징으로 할 수 있다.

발명의 효과

- [0021] 일 실시예에 따르면 문헌 고찰 시에 방대한 자료에서 적절한 참고 자료를 찾기 위한 시간을 최소화할 수 있다.
- [0022] 일 실시예에 따르면 텍스트 마이닝 기법을 통해 문헌 검색 과정을 간소화할 수 있다.
- [0023] 일 실시예에 따르면 연구에 대한 질 좋은 참고 자료를 통해 연구의 일관성을 증대시킬 수 있다.
- [0024] 일 실시예에 따르면 문헌 검색 시 기계 학습을 이용해서 정확성을 높일 수 있다.
- [0025] 일 실시예에 따르면 종래에 사람이 해야 하는 스크리닝 과정을 자동화하여 문헌 고찰 과정에서 필요로 하는 시

간과 노력을 줄일 수 있다.

- [0026] 일 실시예에 따르면 체계적 문헌 고찰(메타분석 포함)의 질적 및 양적 향상을 이룰 수 있다.
- [0027] 일 실시예에 따르면 문헌 및 연구 결과가 기하급수적으로 늘어나고 있는 요즘 관심 문헌의 검색이 간단해지므로 의료 정보의 대중화에 기여할 수 있다.
- [0028] 일 실시예에 따르면 문헌 고찰 및 메타 분석이 용이해지므로, 의학 문헌에 대한 접근성이 향상되면서 관련 문헌의 통합에 의한 결론 도출이 용이해지면서 의학 발전에 기여할 수 있다.
- [0029] 일 실시예에 따르면 기존의 텍스트 마이닝 기법에 의한 방법이나 소프트웨어보다 기계 학습을 통해 향상된 문헌 스크리닝 시스템을 통해 문헌 고찰 및 메타 분석을 자동화해주는 소프트웨어의 기반을 마련할 수 있다.

도면의 간단한 설명

- [0030] 도 1은 일 실시예에 따른 문헌 스크리닝 시스템의 전체적인 개요를 도시하는 도면이다.
- 도 2는 일 실시예에 따른 문헌 스크리닝 방법을 설명하기 위한 흐름도이다.
- 도 3은 일 실시예에 따른 최종 타겟 문헌들을 결정하기 위한 도면이다.
- 도 4는 일 실시예에 따른 분류기의 학습 과정을 설명하기 위한 도면이다.
- 도 5는 일 실시예에 따른 분류기에 의해 분류된 문헌들의 그룹을 설명하기 위한 도면이다.
- 도 6은 일 실시예에 따른 분류기의 학습 과정을 설명하기 위한 도면이다.
- 도 7은 일 실시예에 따른 제2 후보 타겟 문헌들을 설명하기 위한 도면이다.
- 도 8은 일 실시예에 따른 클러스터링 모델의 학습 과정을 설명하기 위한 도면이다.
- 도 9는 일 실시예에 따른 제2 후보 타겟 문헌들의 클러스터링 과정을 설명하기 위한 도면이다.
- 도 10은 일 실시예에 따른 최종 타겟 문헌들을 결정하는 과정을 설명하기 위한 도면이다.
- 도 11 및 도 12는 일 실시예에 따른 각 스코어에 기초하여 선정된 문헌들의 수를 나타내는 그래프들이다.
- 도 13은 일 실시예에 따른 문헌 스크리닝 장치의 구성을 도시하는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0031] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0032] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0033] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.
- [0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0035] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0036] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면

부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

- [0037] 도 1은 일 실시예에 따른 문헌 스크리닝 시스템의 전체적인 개요를 도시하는 도면이다.
- [0038] 본 명세서에서 설명하는 문헌 스크리닝 시스템은 기계 학습 알고리즘을 이용하여 체계적 문헌 고찰 및 메타 분석 과정 중 문헌 스크리닝 과정을 수행할 수 있다. 문헌 스크리닝 시스템은 방대한 양의 문헌들로부터 연구와 관련된 관심 문헌 또는 최종 타겟 문헌들을 선정할 수 있다.
- [0039] 문헌 스크리닝 시스템은 체계적 문헌 고찰과 메타 분석을 하기 위해서 특정 주제 및 핵심단어와 관련된 문헌 검색, 문헌 스크리닝, 최종 문헌 선택, 자료 추출, 자료 합성 및 결론 도출을 위한 일련의 과정 중에 문헌 스크리닝을 위한 기계 학습 알고리즘에 관한 것일 수 있다. 문헌 스크리닝 시스템은 기계 학습 알고리즘을 이용하여 종래에 관련 연구자들의 노력과 시간을 가장 많이 필요로 하며 개인 간 오차가 가장 크게 발생할 수 있는 과정인 문헌 스크리닝을 효율적으로 개선하거나 더 나아가 위 과정을 대체할 수 있다.
- [0040] 문헌 스크리닝 시스템은 원시 데이터의 일부를 학습 데이터로 사용하여 이차 미세 학습을 통해 학습을 효율적으로 수행하도록 할 수 있는 이중 스크리닝 계획(Double screening scheme)과 비지도 학습의 한계를 보완하기 위한 준-지도 클러스터링(Semi-supervised clustering)이 적용될 수 있다.
- [0041] 문헌 스크리닝 시스템은 지도 학습으로 학습된 지도 학습 모델인 분류기를 이용하여 문헌들을 복수의 그룹들로 분류하고, 분류 결과 미결정 그룹으로 분류된 문헌들에 대해 클러스터링을 수행하여 그룹으로의 분류와 클러스터링의 결과에 기초하여 최종 타겟 문헌들을 선정할 수 있다.
- [0042] 문헌 스크리닝 시스템은 문헌들에 대해 1차 스크리닝, 2차 스크리닝 및 문헌 선정 과정을 통해 문헌 스크리닝 방법을 제공할 수 있다. 문헌 스크리닝 시스템은 1차 스크리닝의 결과로 문헌들을 그룹으로 분류할 수 있고, 2차 스크리닝의 결과로 후보 타겟 문헌들을 선별할 수 있다. 문헌 스크리닝 시스템은 문헌 선정의 결과로 최종 타겟 문헌들을 결정할 수 있다. 문헌 스크리닝 시스템은 1차 스크리닝 및 2차 스크리닝을, 분류기를 이용하여 수행할 수 있고, 문헌 선정 과정을 클러스터링 모델을 이용하여 수행할 수 있다.
- [0043] 도 1을 참조하면 문헌 스크리닝 장치(120)는 문헌들(110)을 입력받아 최종 타겟 문헌들(130)을 선정할 수 있다. 문헌들(110)은 통합 데이터베이스에 포함된 문헌들 중 중복이 제거된 문헌들일 수 있다. 최종 타겟 문헌들(130)은 문헌 스크리닝의 목적이 되는 연구와 관련된 문헌들로 선정된 문헌들일 수 있다. 문헌 스크리닝 장치(120)는 학습된 분류기를 이용하여 스크리닝을 수행하고 학습된 클러스터링 모델을 이용하여 스크리닝된 문헌들에 대해 클러스터링을 수행하고 최종 타겟 문헌들(130)을 결정할 수 있다.
- [0044] 도 2는 일 실시예에 따른 문헌 스크리닝 방법을 설명하기 위한 흐름도이다.
- [0045] 도 2를 참조하면 단계(210)에서 문헌 스크리닝 장치는 미리 학습된 분류기를 이용하여, 스크리닝하고자 하는 문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정할 수 있다. 여기서 분류기는 기존에 스크리닝된 문헌들에 기초하여 학습될 수 있고, 예를 들어 최대 엔트로피 분류기(maximum entropy classifier)일 수 있다. 분류기는 기존에 스크리닝된 문헌들을 학습 데이터로 하여 지도 학습될 수 있다.
- [0046] 문헌 스크리닝 장치는 분류기에 리브 원 아웃 교차 검증(leave-one-out cross-validation)을 적용하고, 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치(Keyword weighting)를 적용할 수 있다. 문헌 스크리닝 장치는 키워드 가중치가 적용된 분류기를 이용하여 제1 스코어를 산출할 수 있다.
- [0047] 리브 원 아웃 교차 검증은 필요 표지(labeling) 데이터를 줄이고 모델 학습 효율을 극대화함으로써 사용자가 문헌 스크리닝에 참여야 하는 노력을 최소화시킬 수 있다. 또한, 리브 원 아웃 교차 검증은 표지 데이터에서의 포함 확률을 확인하여 해당 데이터를 재표지하거나 가중화에 사용된 주요 단어를 업데이트함으로써 학습 데이터의 질과 모델 성능을 높일 수 있다.
- [0048] 분류기는 키워드 가중치 과정에서, 연구 주제어, 연구 설계 관련 단어 등 주요 단어 가중화를 수행하여, 사용자의 사전 지식을 활용할 수 있게 하여 강건(Robust)한 결과를 얻을 수 있다.
- [0049] 단계(220)에서 문헌 스크리닝 장치는 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 문헌들을 긍정 그룹(positive group), 미결정 그룹(non-decision group) 및 부정 그룹(negative group)으로 각각 분류할 수 있다. 실시예에 따라 부정 그룹은 제1 부정 그룹 및 제2 부정 그룹으로도 분류될 수 있다. 여기서, 제1 상위 임계 값 및 제1 하위 임계 값은 분류기에 적용된 리브 원 아웃 교차 검증의 검증 결과에 기초하여 결정될 수

있다. 예를 들어 제1 스코어가 제1 상위 임계 값 이상인 문헌들은 긍정 그룹으로 분류될 수 있고, 제1 상위 임계 값 및 제1 하위 임계 값 사이인 문헌들은 미결정 그룹으로 분류되고, 제1 하위 임계 값 이하인 문헌들은 부정 그룹으로 분류될 수 있다.

- [0050] 문헌 스크리닝 장치는 단계(230)에서 분류기를 이용하여, 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 타겟 문헌으로 선정될 제2 스코어를 결정할 수 있다. 분류기는 기존에 스크리닝된 문헌들 및 부정 그룹에 포함된 문헌들 중 적어도 하나에 기초하여 학습될 수 있다.
- [0051] 단계(240)에서 문헌 스크리닝 장치는 제2 스코어 및 제2 하위 임계 값에 기초하여 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정할 수 있다. 문헌 스크리닝 장치는 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용할 수 있다. 문헌 스크리닝 장치는 리브 원 아웃 교차 검증의 검증 결과에 기초하여 제2 하위 임계 값을 결정할 수 있다.
- [0052] 단계(250)에서 문헌 스크리닝 장치는 클러스터링 모델을 이용하여 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정할 수 있다. 클러스터링 모델은 기존에 스크리닝된 문헌들 및 제2 후보 타겟 문헌들에 기초하여 학습되고, 메트릭 K-평균 클러스터링 모델(metric K-means clustering model)일 수 있다. 클러스터링 모델에 기존의 비지도 학습의 문제점인 불확실성을 절감시키고 원하는 결과를 얻을 수 있게 해주는 준-지도 클러스터링이 적용될 수 있다. 문헌 스크리닝 장치는 클러스터링 모델을 이용하여 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하고, 클러스터링의 결과에 기초하여 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정할 수 있다.
- [0053] 단계(260)에서 문헌 스크리닝 장치는 긍정 그룹에 포함된 문헌들과 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정할 수 있다.
- [0054] 도 3은 일 실시예에 따른 최종 타겟 문헌들을 결정하기 위한 도면이다.
- [0055] 도 3을 참조하면 분류기(320)는 1차 스크리닝 과정에서 통합 데이터베이스에 포함된 전체 문헌들(305)에서 추출된 중복이 제거된 문헌들(310)을 복수의 그룹들로 분류할 수 있다. 분류기(320)는 중복이 제거된 문헌들(310)을 긍정 그룹(325), 미결정 그룹(330) 및 부정 그룹(335)을 포함하는 그룹들로 분류할 수 있다. 여기서 긍정 그룹(325)에 포함된 문헌들은 최종 타겟 문헌들(355)로 선정될 수 있다.
- [0056] 분류기(320)는 2차 스크리닝 과정에서 미결정 그룹(330)에 포함된 문헌들을 제1 후보 타겟 문헌들로 선정하고 제1 후보 타겟 문헌들로부터 제2 후보 타겟 문헌들(340)을 선정할 수 있다. 제2 후보 타겟 문헌들(340)은 미결정 그룹(330)에 대해 스크리닝이 수행된 문헌들이라고도 지칭될 수 있다.
- [0057] 클러스터링 모델(345)은 문헌 선정 과정에서 제2 후보 타겟 문헌들(340)에 대해 클러스터링을 수행할 수 있고, 제2 후보 타겟 문헌들(340)에 포함된 클러스터 중 타겟 문헌들(350)을 선정할 수 있다. 문헌 스크리닝 장치는 긍정 그룹(325)에 포함된 문헌들 및 타겟 문헌들(350)을 최종 타겟 문헌들(355)로 결정할 수 있다.
- [0058] 도 4는 일 실시예에 따른 분류기의 학습 과정을 설명하기 위한 도면이다.
- [0059] 도 4를 참조하면 전체 문헌들(410) 중에서 미리 정해진 기준에 기초하여 기존에 스크리닝된 문헌들(420)이 선정될 수 있다. 기존에 스크리닝된 문헌들(420)은 사용자들이 직접 문헌을 스크리닝을 수행함으로써 스크리닝된 문헌들일 수 있다. 분류기(430)는 기존에 스크리닝된 문헌들(420)을 학습 데이터로써 지도 학습될 수 있다.
- [0060] 도 5는 일 실시예에 따른 분류기에 의해 분류된 문헌들의 그룹을 설명하기 위한 도면이다.
- [0061] 도 5를 참조하면 분류기에 키워드 가중치와 리브 원 아웃 교차 검증이 적용되어 상위 임계 값과 하위 임계 값이 결정될 수 있다. 또한, 분류기는 문헌들(510) 각각에 대해 타겟 문헌으로 선정될 스코어를 결정할 수 있다. 문헌들(510)은 각각 결정된 스코어, 상위 임계 값과 하위 임계 값에 기초하여 복수의 그룹들로 분류될 수 있다. 도 5의 실시예에서 문헌들(510)은 네 개의 그룹들로 분류될 수 있다. 문헌들(510)은 긍정 그룹(520), 미결정 그룹(530), 제1 부정 그룹(540) 및 제2 부정 그룹(550)을 포함할 수 있다. 이 중 긍정 그룹(520)에 포함된 문헌들은 모두 최종 타겟 문헌들로 결정될 수 있고, 미결정 그룹(530)에 포함된 문헌들은 2차 스크리닝 과정을 통해 최종 타겟 문헌들로 선정될 수 있다.
- [0062] 도 6은 일 실시예에 따른 분류기의 학습 과정을 설명하기 위한 도면이다.
- [0063] 1차 스크리닝 과정에서 상위 임계 값 및 하위 임계 값에 기초하여 문헌들은 긍정 그룹, 부정 그룹, 제1 부정 그룹 및 제2 부정 그룹으로 분류될 수 있다. 도 6을 참조하면 제2 부정 그룹(630)은 제1분류된 그룹들 중 제2 부정 그룹에 대응될 수 있다. 분류기(640)는 전체 문헌들(610) 중에서 미리 정해진 기준에 기초하여 기존에 스

리닝된 문헌들(620)과 제2 부정 그룹(630)으로 분류된 문헌들에 기초하여 학습될 수 있다.

- [0064] 도 7은 일 실시예에 따른 제2 후보 타겟 문헌들을 설명하기 위한 도면이다.
- [0065] 분류기는 1차 스크리닝 과정의 결과로 문헌들 중에서 미결정 그룹으로 분류된 문헌들에 대해 2차 스크리닝 과정에서 제2 후보 타겟 문헌들을 선정할 수 있다. 여기서 미결정 그룹에 포함된 문헌들은 제1 후보 타겟 문헌들이라고도 지칭될 수 있다. 미결정 그룹은 타겟 문헌으로 선정될 문헌들인지 아닌지 여부가 결정되지 않은 문헌들을 포함할 수 있다. 분류기는 미결정 그룹에 포함된 문헌들을, 문헌들 각각에 대해 결정된 타겟 문헌으로 선정될 스코어 및 하위 임계 값에 기초하여 제2 후보 타겟 문헌들로 분류할 수 있다.
- [0066] 도 8은 일 실시예에 따른 클러스터링 모델의 학습 과정을 설명하기 위한 도면이다.
- [0067] 도 8을 참조하면 클러스터링 모델(860)은 제2 후보 타겟 문헌들(850)과, 기존에 스크리닝된 문헌들(830) 중 특히 선정될 문헌들(positive cases)(820)에 기초하여 학습될 수 있다. 선정될 문헌들(820)은 기존에 스크리닝된 문헌들(830) 중에서 미리 정해진 기준에 기초하여 타겟 문헌으로 선정될 수 있는 문헌들을 포함할 수 있다.
- [0068] 도 9는 일 실시예에 따른 제2 후보 타겟 문헌들의 클러스터링 과정을 설명하기 위한 도면이다.
- [0069] 도 9를 참조하면, 분류기(910)는 제2 후보 타겟 문헌들에 대해 클러스터링을 수행할 수 있다. 문헌 스크리닝 장치는 문헌들이 클러스터링 된 클러스터들 중에서 선정된 클러스터에 포함된 문헌들을 타겟 문헌들로 결정할 수 있다.
- [0070] 도 10은 일 실시예에 따른 최종 타겟 문헌들을 결정하는 과정을 설명하기 위한 도면이다. 도 10을 참조하면, 문헌 스크리닝 장치는 제2 후보 타겟 문헌들(1010)이 클러스터링 된 클러스터들 중에서 선정된 클러스터에 포함된 문헌들과, 긍정 그룹(1020)에 포함된 문헌들을 최종 타겟 문헌들(1030)로 결정할 수 있다.
- [0071] 도 11 및 도 12는 일 실시예에 따른 각 스코어에 기초하여 선정된 문헌들의 수를 나타내는 그래프들이다.
- [0072] 도 11은 모든 문헌들 중 중복이 제거된 로우 문헌들(raw articles) 3862개에 대하여, 각 문헌들이 타겟 문헌으로 선정될 스코어의 구간 별로 선정된 논문 개수를 시각화한 그래프일 수 있다. 여기서 스코어는 본 명세서에서 설명하는 제1 스코어에 대응할 수 있다. 도 11을 참조하면, 스코어가 0.1보다 큰 문헌들은 3246개이고, 스코어가 0.2보다 큰 문헌들은 2627개일 수 있다. 또한, 스코어가 0.3보다 큰 문헌들은 2090개이고, 스코어가 0.4보다 큰 문헌들은 1593개일 수 있다.
- [0073] 도 12는 1차 스크리닝의 결과로 미결정 그룹에 포함된 문헌들 375개에 대하여 각 문헌들이 타겟 문헌으로 선정될 스코어의 구간 별로 선정된 논문 개수를 시각화한 그래프일 수 있다. 여기서 스코어는 본 명세서에서 설명하는 제2 스코어에 대응할 수 있다. 도 12를 참조하면, 스코어가 0.1보다 큰 문헌들은 374개이고, 스코어가 0.2보다 큰 문헌들은 365개일 수 있다. 또한, 스코어가 0.8보다 큰 문헌들은 48개이고 스코어가 0.9보다 큰 문헌들은 12개일 수 있다.
- [0074] 도 13은 일 실시예에 따른 문헌 스크리닝 장치의 구성을 도시하는 도면이다.
- [0075] 도 13을 참조하면, 문헌 스크리닝 장치(1300)는 프로세서(1310), 메모리(1320)를 포함할 수 있다. 실시예에 따라, 문헌 스크리닝 장치(1300)는 데이터베이스(1330)를 더 포함할 수도 있다. 문헌 스크리닝 장치(1300)는 본 명세서에서 설명된 문헌 스크리닝 장치에 대응할 수 있다.
- [0076] 메모리(1320)는 프로세서(1310)에 연결되고, 프로세서(1310)에 의해 실행가능한 인스트럭션들, 프로세서(1310)가 연산할 데이터 또는 프로세서(1310)에 의해 처리된 데이터를 저장할 수 있다. 메모리(1320)는 비일시적인 컴퓨터 판독가능 매체, 예컨대 고속 랜덤 액세스 메모리 및/또는 비휘발성 컴퓨터 판독가능 저장 매체(예컨대, 하나 이상의 디스크 저장 장치, 플래쉬 메모리 장치, 또는 기타 비휘발성 솔리드 스테이트 메모리 장치)를 포함할 수 있다.
- [0077] 데이터베이스(1330)는 문헌 스크리닝 장치(1300)가 문헌 스크리닝 방법을 수행하는 데 필요한 데이터를 저장할 수 있다. 예를 들어, 데이터베이스(1330)는 학습된 분류기 및 클러스터링 모델의 파라미터를 저장할 수 있다.
- [0078] 프로세서(1310)는 문헌 스크리닝 장치(1300) 내에서 실행하기 위한 기능 및 인스트럭션들을 실행하고, 문헌 스크리닝 장치(1300)의 전체적인 동작을 제어한다. 프로세서(1310)는 본 명세서에서 설명된 문헌 스크리닝 장치의 동작과 관련된 하나 이상의 동작을 수행할 수 있다.
- [0079] 예를 들어 프로세서(1310)는 문헌 스크리닝 장치(1300)가 미리 학습된 분류기를 이용하여, 스크리닝하고자 하는

문헌들 각각에 대해 타겟 문헌으로 선정될 제1 스코어를 결정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다. 프로세서(1310)는 문헌 스크리닝 장치(1300)가 분류기에 리브 원 아웃 교차 검증을 적용하고, 리브 원 아웃 교차 검증이 적용된 분류기에 키워드 가중치를 적용하고, 키워드 가중치가 적용된 분류기를 이용하여 제1 스코어를 산출하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다.

[0080] 프로세서(1310)는 문헌 스크리닝 장치(1300)가 제1 스코어, 제1 상위 임계 값 및 제1 하위 임계 값에 기초하여 문헌들을 긍정 그룹, 미결정 그룹 및 부정 그룹으로 각각 분류하고, 분류기를 이용하여, 미결정 그룹에 포함된 제1 후보 타겟 문헌들에 대해 타겟 문헌으로 선정될 제2 스코어를 결정하고, 제2 스코어 및 제2 하위 임계 값에 기초하여 제1 후보 타겟 문헌들 중에서 제2 후보 타겟 문헌들을 선정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다.

[0081] 프로세서(1310)는 문헌 스크리닝 장치(1300)가 분류기에 리브 원 아웃 교차 검증 및 키워드 가중치를 적용하고, 리브 원 아웃 교차 검증의 검증 결과에 기초하여 제2 하위 임계 값을 결정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다.

[0082] 프로세서(1310)는 문헌 스크리닝 장치(1300)가 클러스터링 모델을 이용하여 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다. 프로세서(1310)는 문헌 스크리닝 장치(1300)가 클러스터링 모델을 이용하여 제2 후보 타겟 문헌들에 대해 클러스터링을 수행하고, 클러스터링의 결과에 기초하여 제2 후보 타겟 문헌들로부터 타겟 문헌들을 선정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다.

[0083] 프로세서(1310)는 문헌 스크리닝 장치(1300)가 긍정 그룹에 포함된 문헌들과 제2 후보 타겟 문헌들로부터 선정된 타겟 문헌들을 최종 타겟 문헌들로 결정하도록 문헌 스크리닝 장치(1300)를 제어할 수 있다.

[0084] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0085] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0086] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0087] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0088] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0089] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

부호의 설명

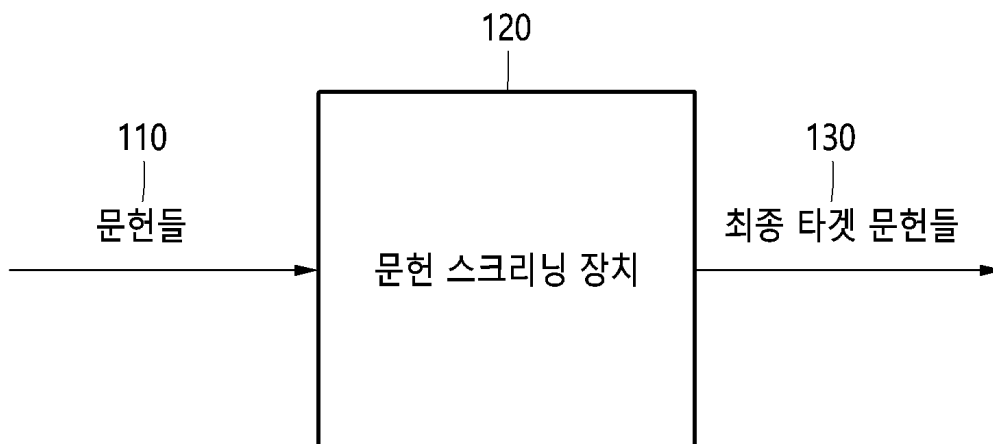
[0090] 120, 1300: 문헌 스크리닝 장치 320, 430, 640: 분류기

345, 860, 910: 클러스터링 모델 1310: 프로세서

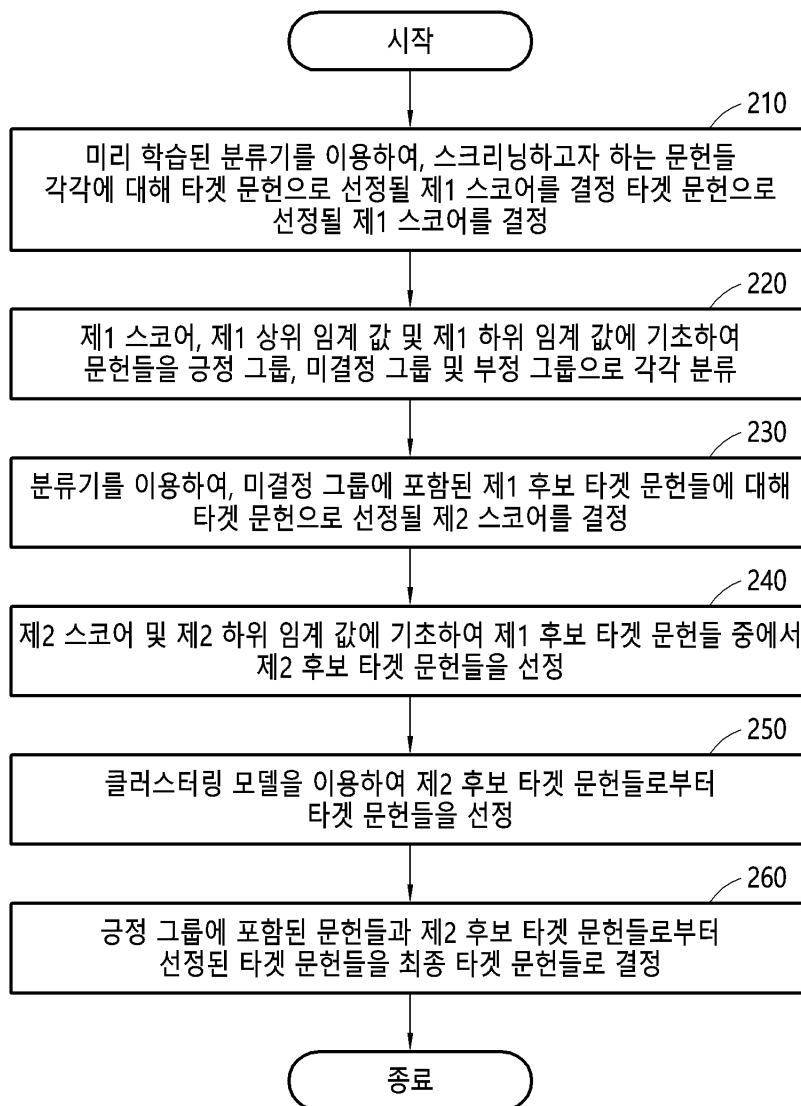
1320: 메모리 1330: 데이터베이스

도면

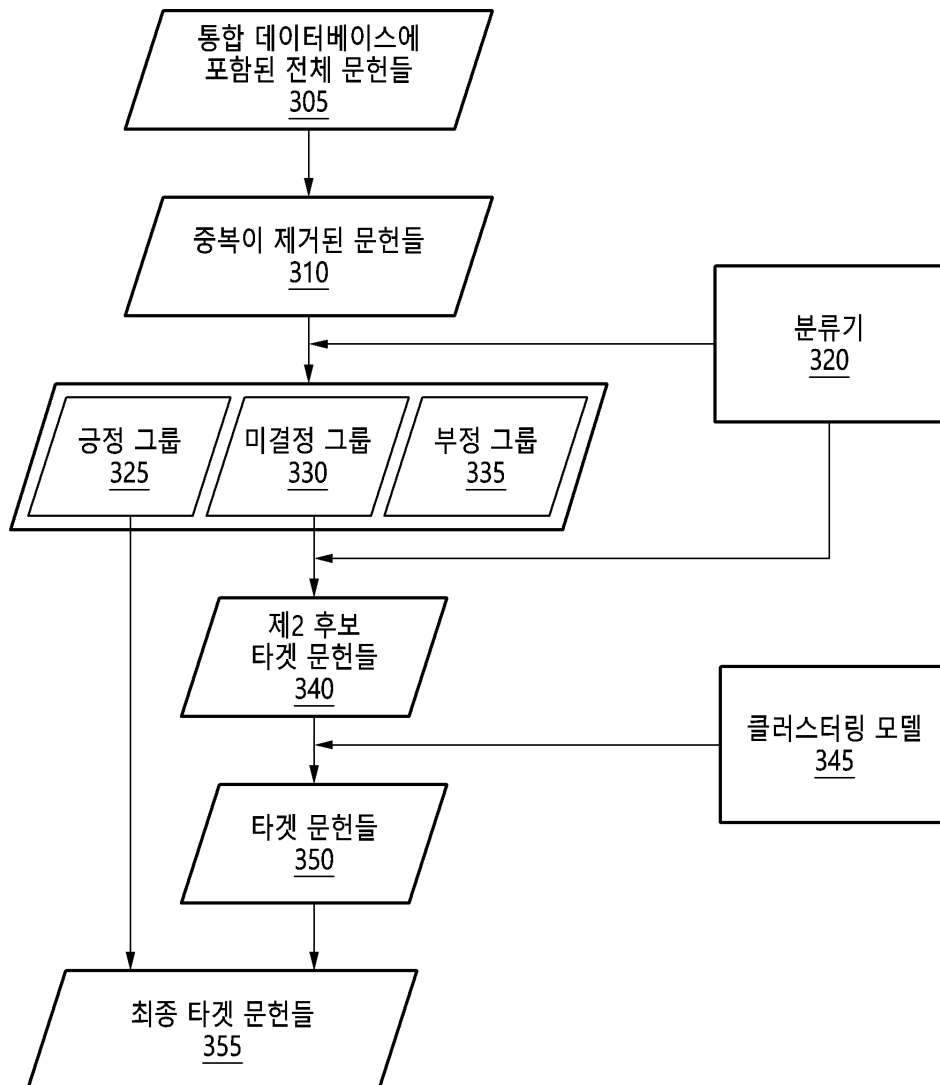
도면1



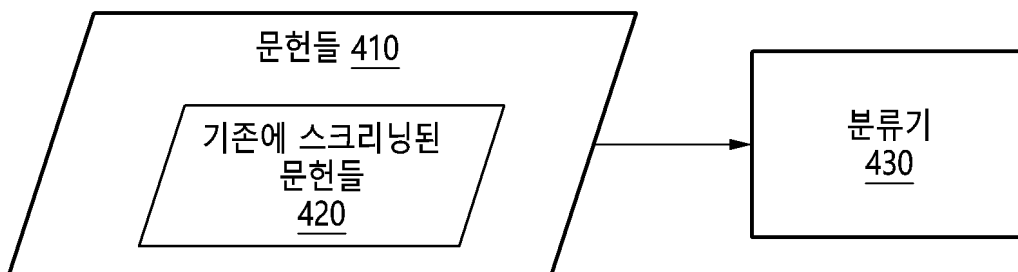
도면2



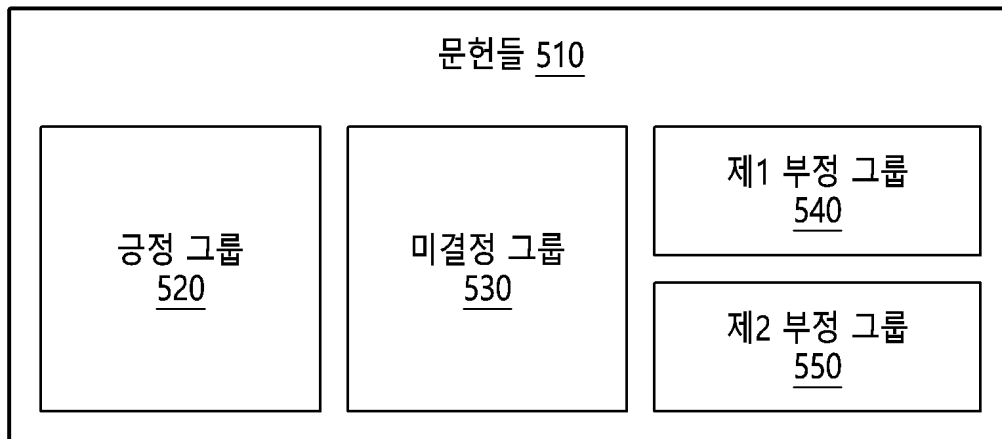
도면3



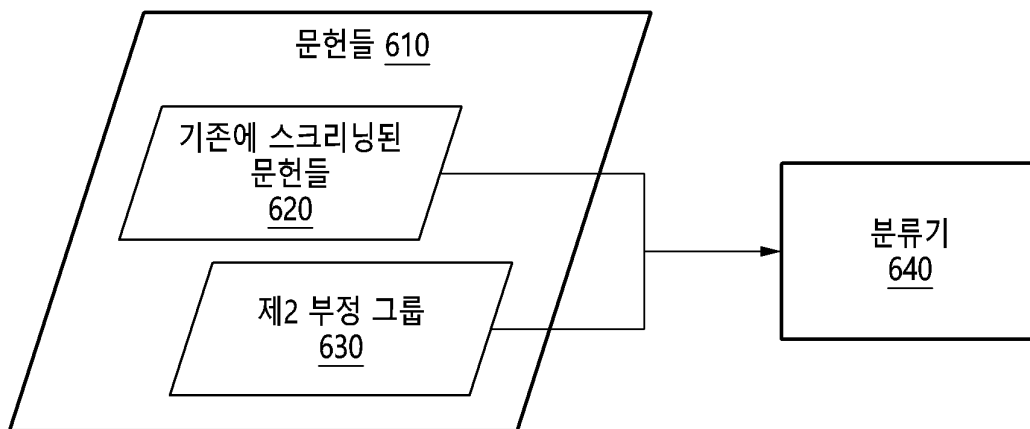
도면4



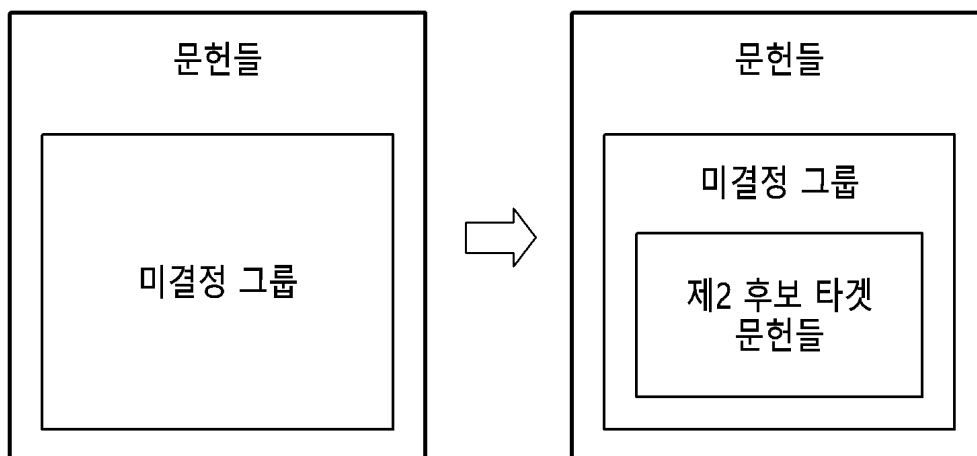
도면5



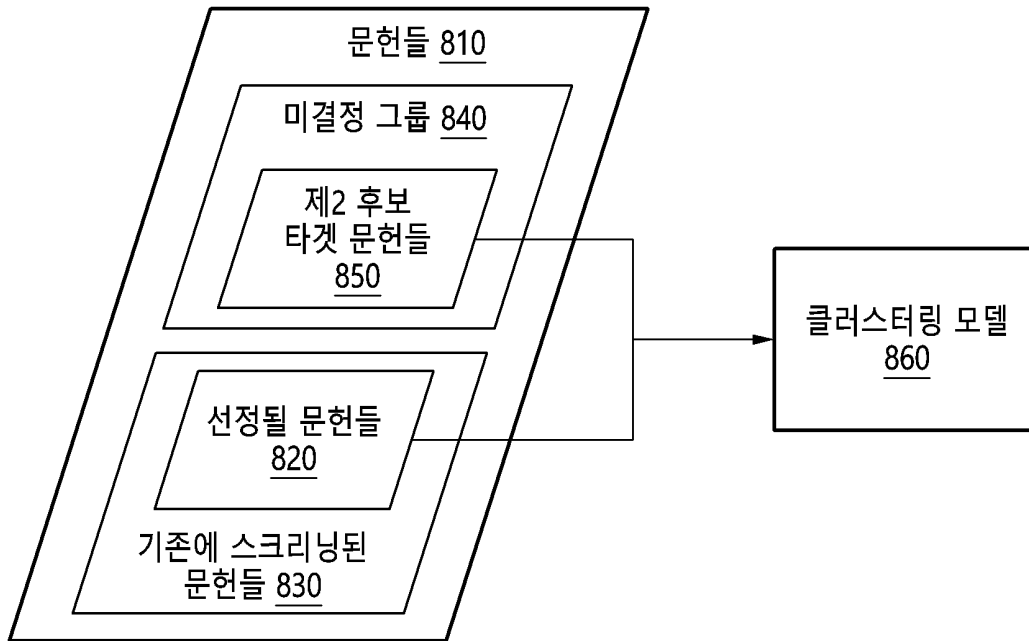
도면6



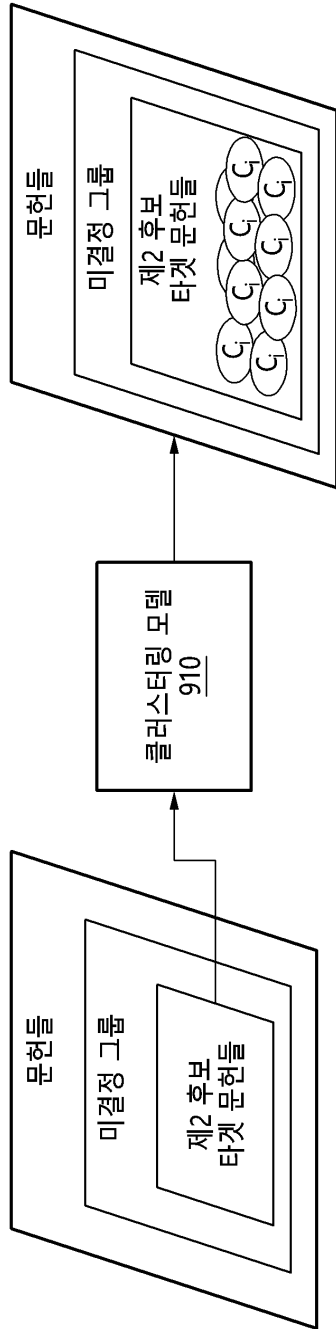
도면7



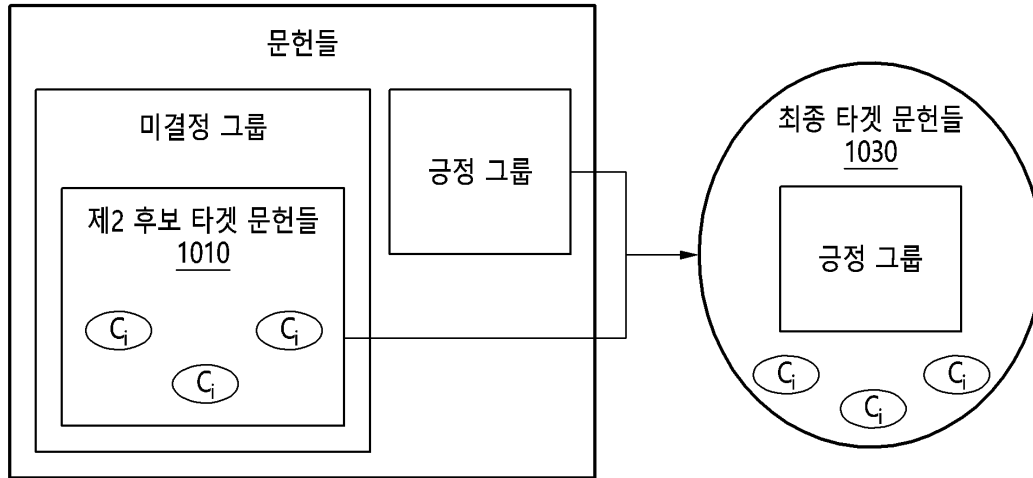
도면8



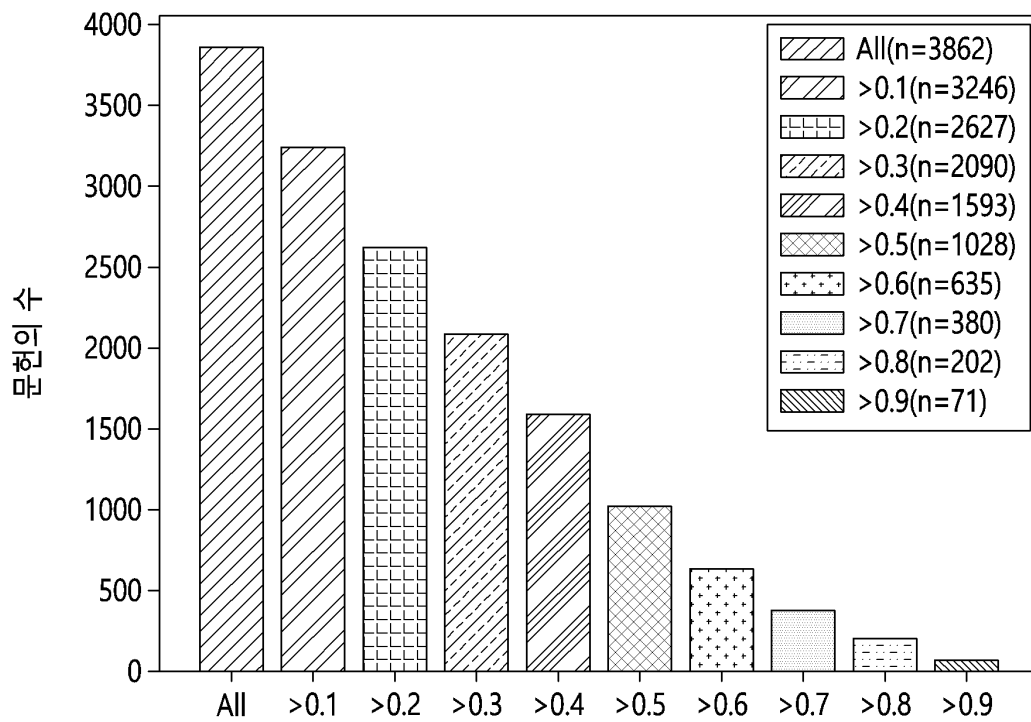
도면9



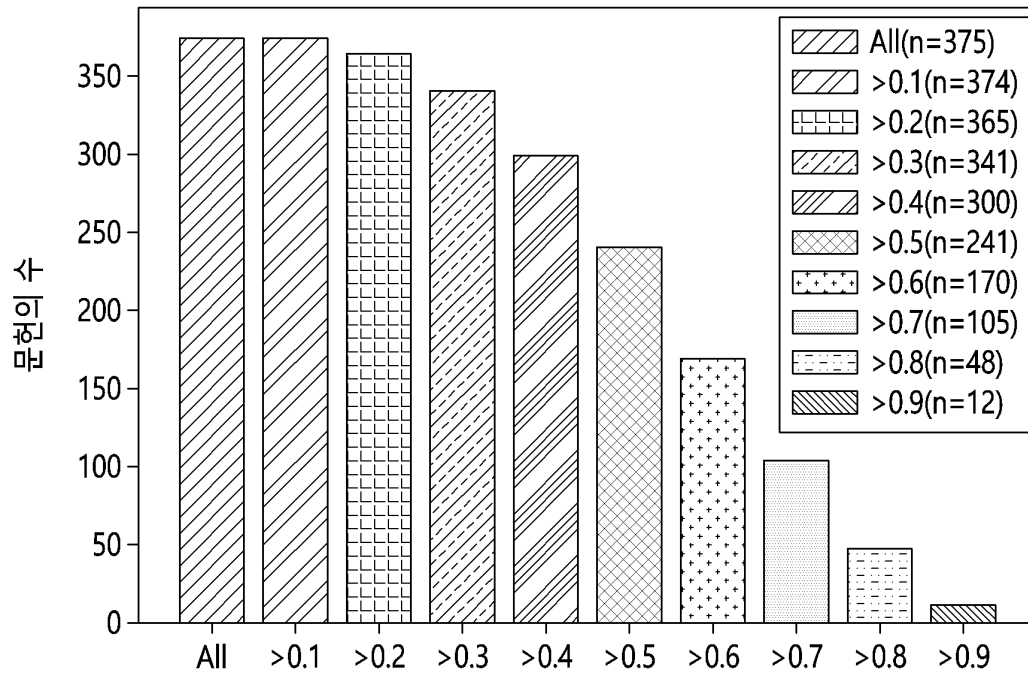
도면10



도면11



도면12



도면13

