

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0142705
(43) 공개일자 2022년10월24일(51) 국제특허분류(Int. Cl.)
G06F 40/30 (2020.01) G06F 40/268 (2020.01)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06F 40/30 (2020.01)
G06F 40/268 (2020.01)
(21) 출원번호 10-2021-0049078
(22) 출원일자 2021년04월15일
심사청구일자 2021년04월15일(71) 출원인
연세대학교 원주산학협력단
강원도 원주시 흥업면 연세대길 1
(72) 발명자
조수행
서울특별시 송파구 잠실로 88, 102동 2101호
(74) 대리인
김보정

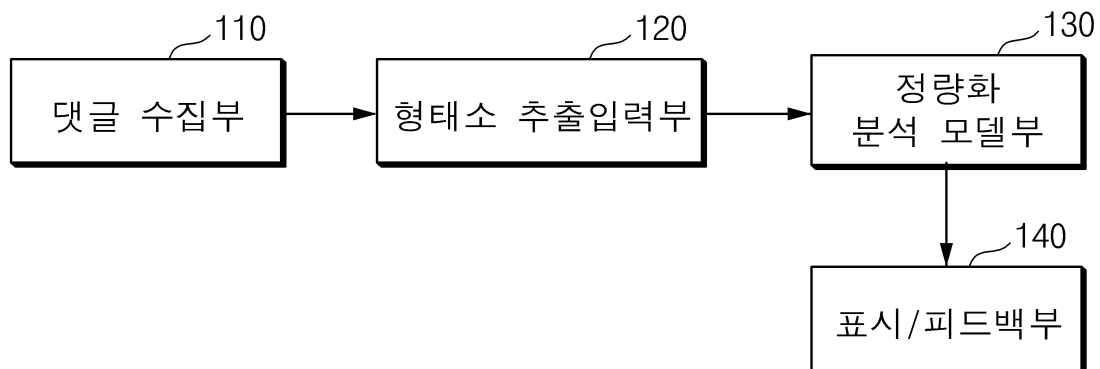
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법

(57) 요약

초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법이 개시된다. 본 발명의 댓글 판단 방법은 실시간 댓글을 수집하여 형태소들을 추출하여 정량화 감정분석 모델에서 분석하여 정량화 감정분석 결과를 획득하여 이를 표시하거나 방송주체에게 피드백하게 구성함으로써, 초고속 분석이 가능한 인공지능을 활용하는 본 발명의 기술을 통해 실시간 댓글을 분석하여 시각화하거나 정량화하여 표시할 수 있기 때문에 만족도를 효과적으로 분석하고 관리하며 실시간 피드백으로 즉각 활용할 수 있는 효과가 있다.

대표도 - 도4



(52) CPC특허분류
G06N 20/00 (2021.08)

명세서

청구범위

청구항 1

장치에서 실시간으로 방송되고 있는 댓글을 분석하여 실시간으로 피드백해주는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법에 있어서,

(a)게시글에 포함되는 댓글을 수집하는 단계

(b)상기 댓글에서 형태소들을 추출하고, 추출된 형태소들 중 적어도 하나를 정량화 감정분석 모델로 입력하는 단계;

(c)상기 정량화 분석 모델로부터 초고속 분석이 가능한 인공지능을 활용하여 감정분석 결과를 획득하는 단계;및

(d)상기 감정분석 결과를 표시하거나 또는 상기 장치의 방송주체에게 획득된 상기 감정분석 결과를 실시간으로 피드백하는 단계;

를 포함하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 2

청구항 1에 있어서,

상기 (a)단계는

상기 장치의 게시글에 포함되는 복수 개의 댓글을 수집하는 단계를 포함하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 3

청구항 2에 있어서,

상기 (b)단계는

상기 댓글에서 은어, 약어, 그리고 축약된 표현들을 모두 형태소로 인식하여 추출하고, 상기 복수 개의 댓글 전체에서 형태소를 추출하는 것을 특징으로 하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 4

청구항 3에 있어서,

상기 (b)단계는

상기 형태소들 중 추출된 명사;

상기 형태소들 중 추출된 은어, 약어, 그리고 축약된 표현;및

상기 형태소들 중 추출된 조사, 동사, 형용사, 특수 문자, 알파벳 대문자, 소문자 및 이모티콘 중 적어도 하나를 정량화 분석 모델로 입력하는 단계

를 포함하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 5

청구항 4에 있어서,

상기 (d)단계는

상기 감정 분석 모델의 출력 값을 미리 정해진 하나 이상의 기준 값과 비교하여 댓글의 감정의 종류를 분류하여 표시하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 6

청구항 4에 있어서,

상기 (d)단계는

상기 정량화 분석 모델 결과를 모델의 출력값을 일정 간격의 시간동안 실시간으로 입력되는 문장형 댓글로 나타나는 만족도를 정량화하여 시간별로 긍정과 중립 그리고 부정의 추이를 시각화하여 표시하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 7

청구항 6에 있어서,

상기 (d)단계는

상기 실시간 댓글을 긍정/부정 혹은 5점 만점 등으로 정량화하여 표시하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

청구항 8

청구항 4에 있어서,

상기 (d)단계는

상기 댓글의 감정의 종류 각각에 대응하는 확률을 계산하고, 계산된 확률에 기초하여 상기 댓글의 감정 점수를 계산하는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법.

발명의 설명

기술 분야

[0001] 본 발명은 실시간 댓글 판단방법에 관한 것으로, 구체적으로 초고속 분석이 가능한 인공지능을 활용하여 실시간 댓글을 긍정/부정 혹은 5점 만점 등으로 정량화하여 만족도를 효과적으로 분석하고 관리하며 실시간 피드백으로 즉각 활용할 수 있는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법에 관한 것이다.

배경 기술

[0002] 소셜 네트워크 서비스(SNS)는 온라인상에서 개인의 일상생활을 비롯한 관심사 및 특정 주제에 대한 토론 등 다양한 정보를 상호 공유하면서 인적 네트워크를 형성할 수 있도록 하는 1인 커뮤니티이다.

[0003] 그리고 SNS 이외에도 모바일 메신저 등 데이터 기반의 공유 서비스는 사용자가 사진, 영상, 글 등의 게시물을 등록하면 사용자와 관계망을 구축한 친구, 선배, 가족 등의 지인들은 게시물 하단에 댓글이나 간단한 이모티콘을 남김으로써 안부나 자신의 의견을 교류할 수 있다.

[0004] 이와 같이 현재 그 수가 늘어가고 있는 유튜브, 아프리카 TV, 홈쇼핑, 공연이나 스포츠 경기 나 다양한 라이브 방송 플랫폼 등에서 시청자나 청취자 등이 실시간으로 올리는 댓글은 그 양과 속도로 방송의 주체자가 관리하기 매우 어렵고 만족도와 관련된 피드백을 실시간으로 활용하기 어렵다.

[0005] 더구나 인기 방송의 경우에는 실시간 확인이 불가능할 정도로 많은 댓글이 달리기 때문에 빠른 스크롤로 방송인이나 함께 보고 있는 시청자조차 읽을 수 없다.

[0006] 즉 긍정적인 댓글이 더 많은 지 부정적인 것이 많은 지 알기가 어렵다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) KR 등록특허공보 제10-2078627호(2020.02.12)

발명의 내용

해결하려는 과제

[0008] 따라서 초고속 분석이 가능한 인공지능을 활용하는 본 발명의 기술을 통해 실시간 댓글을 긍정/부정 혹은 5점 만점 등으로 정량화하여 표시할 수 있는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 이러한 과제를 해결하기 위한 본 발명의 장치에서 실시간으로 방송되고 있는 댓글을 분석하여 실시간으로 피드백해주는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법은, (a)게시글에 포함되는 댓글을 수집하는 단계, (b)상기 댓글에서 형태소들을 추출하고, 상기 형태소들 중 적어도 하나를 정량화 감정분석 모델로 입력하는 단계, (c)상기 정량화 감정분석 모델로부터 초고속 분석이 가능한 인공지능을 활용하여 정량화 감정분석 결과를 획득하는 단계 및 (d)상기 장치의 방송 주체에게 획득된 상기 정량화 감정분석 결과를 실시간으로 피드백하거나 표시하는 단계를 포함하게 구성함으로써 달성될 수 있다.

[0010] 또한, (a)단계는 상기 장치의 게시글에 포함되는 복수 개의 댓글을 수집하는 단계를 포함하고, 상기 (b)단계는 상기 댓글에서 은어, 약어, 그리고 축약된 표현들을 모두 형태소로 인식하여 추출하고, 상기 복수 개의 댓글 전체에서 형태소를 추출한다.

[0011] 또한, (c)단계는 상기 형태소들 중 추출된 명사, 상기 형태소들 중 추출된 은어, 약어, 그리고 축약된 표현, 및 상기 형태소들 중 추출된 조사, 동사, 형용사, 특수 문자, 알파벳 대문자, 소문자 및 이모티콘 중 적어도 하나를 정량화 분석 모델로 입력하는 단계를 포함할 수 있다.

[0012] 그리고 상기 (d)단계는 상기 감정 분석 모델의 출력 값을 미리 정해진 하나 이상의 기준 값과 비교하여 댓글의 감정의 종류를 분류하여 표시하거나, 일정 간격의 시간동안 실시간으로 입력되는 문장형 댓글로 나타나는 만족도를 정량화하여 시간별로 긍정과 중립 그리고 부정의 추이를 시각화하여 표시할 수 있다.

발명의 효과

[0013] 따라서 본 발명의 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법에 의하면 초고속 분석이 가능한 인공지능을 활용하는 본 발명의 기술을 통해 실시간 댓글을 시각화하거나 긍정/부정 혹은 5점 만점 등으로 정량화하여 표시할 수 있기 때문에 만족도를 효과적으로 분석하고 관리하며 실시간 피드백으로 즉각 활용할 수 있는 효과가 있다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 일실시예에 의한 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법을 설명하기 위한 흐름도,

도 2는 본 발명의 일실시예에 의한 정량화 감정분석 모델을 생성하는 방법을 설명하기 위한 도면,

도 3은 본 발명의 일실시예에 의한 정량화 분석 모델의 일례를 도시한 도면,

도 4는 본 발명의 일실시예에 의한 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법을 구현하기 위한 주

요 구성도,

도 5는 정량화 감정분석 모델의 출력 값을 긍정, 부정 또는 중립으로 표시한 도면,

도 6은 인공지능 모델을 이용한 댓글 분석의 소요시간을 예시한 도면,

그리고

도 7은 인공지능 모델을 이용한 댓글 분석의 일 실행예를 예시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0015] 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정 해석되지 아니하며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다.
- [0016] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "장치" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 및/또는 소프트웨어의 결합으로 구현될 수 있다.
- [0017] 명세서 전체에서 "및/또는"의 용어는 하나 이상의 관련 항목으로부터 제시 가능한 모든 조합을 포함하는 것으로 이해되어야 한다. 예를 들어, "제1 항목, 제2 항목 및/또는 제3 항목"의 의미는 제1, 제2 또는 제3 항목뿐만 아니라 제1, 제2 또는 제3 항목들 중 2개 이상으로부터 제시될 수 있는 모든 항목의 조합을 의미한다.
- [0018] 명세서 전체에서 각 단계들에 있어 식별부호(예를 들어, a, b, c, ...)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 한정하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0019] 이하, 도면을 참고하여 본 발명의 일 실시예에 대하여 설명한다.
- [0020] 도 1은 본 발명의 일 실시예에 의한 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법을 설명하기 위한 흐름도로서, 도시된 바와 같이, 본 발명의 댓글 판단 방법은 실시간으로 초고속 분석 인공지능을 이용한 댓글 정량화 감정분석 방법에 대해 프로그래밍된 정량화 감정분석 모델을 통해 수행하고, 이를 표시하거나 또는 피드백하는 것을 특징으로 한다.
- [0021] 다시 말하면, 본 발명은 소정 장치에서 실시간으로 방송되고 있는 댓글을 분석하여 실시간으로 피드백해주는 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법에 관한 것으로, 먼저 게시글에 포함되는 댓글을 수집하는 단계(S110)와 단계 S110에서 수집된 댓글에서 형태소들을 추출하고, 추출된 형태소들 중 적어도 하나를 정량화 감정분석 모델로 입력하는 단계(S120), 상기 정량화 감정분석 모델로부터 초고속 분석이 가능한 인공지능을 활용하여 정량화 감정분석 결과를 획득하는 단계(S130) 및 단계 S130에서 획득된 정량화 감정분석 결과를 표시하고 상기 장치의 방송주체에게 실시간으로 피드백하는 단계(S140)를 포함한다.
- [0022] 이를 위하여 본 발명은 도 4의 구성도에서와 같이, 게시글에서 댓글을 수집하는 댓글수집부(110)와 댓글수집부(110)에서 수집한 댓글에서 형태소를 추출하는 형태소 추출입력부(120)와, 형태소 추출입력부(120)에서 추출된 형태소를 정량화 감정분석 모델을 통하여 분석하는 정량화 분석 모델부(130), 그리고 분석된 결과를 시각적으로 표시하거나 장치의 방송주체에게 실시간으로 피드백하는 표시부(140)를 통하여 이루어지게 할 수 있다.
- [0023] 이러한 댓글수집부(110), 형태소 추출입력부(120), 정량화 분석 모델부(130), 표시/피드백부(140)는 PC와 스마트폰과 같은 하나의 단말기에서 이루어지게 할 수 있다.
- [0024] 즉, 게시글에 포함되는 댓글을 수집하는 단계(S110)는 댓글수집부(110)에서 장치에서 실시간으로 방송되고 있는 게시글에서 댓글을 수집하도록 동작한다.
- [0025] 댓글수집부(110)는 장치에서 게시글에 포함되는 댓글을 수집하는 장치로, 온라인상에서 개인의 일상생활을 비롯한 관심사 및 특정 주제에 대한 토론, 사용자가 사진, 영상, 글 등의 게시물을 등록하면 사용자와 관계망을 구축한 친구, 선배, 가족 등의 지인들은 게시물 하단에 댓글이나 간단한 이모티콘을 남기는 다양한 정보로부터 댓글을 수집한다.

- [0026] 또한, 다양한 사이트의 게시글에 대한 댓글을 수집할 수도 있다. 예를 들어, 소셜 네트워크 서비스(SNS), 뉴스 기사, 블로그, 예약 사이트, 쇼핑몰 등의 사이트 내에 게시된 특정 게시글에 달린 하나 혹은 그 이상의 댓글을 수집할 수 있다.
- [0027] 댓글을 수집할 시, 게시글이 게시된 사이트에 대한 정보를 함께 수집할 수 있는데 예를 들어, SNS, 뉴스 사이트, 블로그, 예약 사이트, 쇼핑몰 등 중 게시글의 출처에 대한 정보가 수집될 수 있다. 게시글의 주제 또는 카테고리에 대한 정보, 예를 들어, 정치, 시사, 영화, 연예, 스포츠, 게임, 여행 등의 카테고리에 대한 정보도 수집될 수 있다. 게시글의 콘텐츠 유형에 대한 정보, 텍스트, 사진, 동영상 등의 정보가 수집될 수도 있다.
- [0028] 또한, 댓글수집부(110)는 장치의 게시글에 포함되는 하나 이상의 복수 개의 댓글을 수집하여 형태소 추출입력부(120)에서 복수 개의 댓글전체에 대해서 형태소들을 추출하게 할 수 있다.
- [0029] 형태소 추출입력부(120)는 댓글수집부(110)에서 수집된 댓글에 대하여 형태소들을 추출하여 정량화 감정분석 모델부(130)에서 인공지능을 이용한 댓글 감정 분석이 이루어지게 한다.
- [0030] 이러한 형태소 추출부(120)는 댓글에서 은어, 약어, 그리고 축약된 표현들을 모두 형태소로 인식하여 추출하고, 상기 복수 개의 댓글 전체에서 형태소를 추출할 수도 있으며, 추출된 형태소는 정량화 분석 모델부(130)로 입력한다.
- [0031] 참고로 축약된 언어는 "짱이다, 팡이다, ^^, ㄸ" 등을 예로 들 수 있다.
- [0032] 또한, 형태소 추출입력부(120)는 형태소들 중 추출된 명사와 추출된 조사, 동사, 형용사, 특수 문자, 알파벳 대문자, 소문자 및 이모티콘 중 적어도 하나를 정량화 분석모델부(130)로 분석을 위하여 입력한다.
- [0033] 정량화 분석 모델부(130)는 형태소 추출입력부(120)에서 추출된 형태소들을 인공지능으로 분석하여 감정 분석 결과인 정량화 감정분석 모델을 획득하도록 동작한다.
- [0034] 정량화 분석 모델부(130)는 정량화 감정분석 모델의 출력 값을 미리 정해진 하나 이상의 기준 값과 비교하여 댓글의 감정의 종류를 분류하여 표시/피드백부(140)에 표시할 수도 있고, 일정 간격의 시간동안 실시간으로 입력되는 댓글을 정량화하여 시간별로 긍정과 중립 그리고 부정의 추이를 시각화하여 표시/피드백부(140)에 표시할 수도 있고, 실시간 댓글을 긍정/부정 혹은 5점 만점 등으로 정량화하여 표시/피드백부(140)에 표시할 수도 있다.
- [0035] 도 5의 정량화 감정분석 모델의 출력 값을 긍정, 부정 또는 중립으로 표시한 도면을 참고하면, 일정 간격의 설정된 시간 동안 입력되는 댓글을 정량화해서 시간별로 그 추이를 보여주는 것으로, 긍정과 중립 그리고 부정에 대한 댓글이 표시되어 있고, 가중치는 동그라미의 갯수로 표시하였다.
- [0036] 도면에서는 설정된 시간동안 부정의 댓글이 많았음을 나타낸다.
- [0037] 도면을 참고하면, 설정된 시간이 경과되면 다시 설정된 시간동안 댓글을 분석하여 긍정과 중립 그리고 부정의 가중치로 표시하는 것을 반복하여 실시간으로 설정된 시간내에서의 감정분석결과를 스트리밍 형식으로 표시하는 것이다.
- [0038] 즉, 본 발명은 등록된 댓글을 분석하는 것이 아니라 실시간으로 게시되는 댓글을 분석하여 실시간으로 피드백을 해주는 시스템이다.
- [0039] 도 6의 인공지능 모델을 이용한 댓글 분석의 소요시간을 예시한 도면을 참고하면, 본 발명은 실시간 분석이 실제로 가능한 기술임을 알 수 있다.
- [0040] 도 6은 현재 최신 컴퓨터 (i9-9000, Nvidia GTX2080 super GPU)를 이용하여 본 발명의 프로그래밍한 인공지능 모델을 이용할 경우 5383천건의 문장의 형태소 분석 등을 통한 감성분석은 512 ms걸리므로 한 건당 100 ms밖에 걸리지 않은 실행예가 표시한 것이다.
- [0041] 또한, 정량화 분석 모델부(130)는 댓글의 감정의 종류 각각에 대응하는 확률을 계산하고, 계산된 확률에 기초하여 상기 댓글의 만족도 점수를 계산하여 표시부(140)에 표시할 수도 있다.
- [0042] 도 7의 인공지능 모델을 이용한 댓글 분석의 일 실행예를 예시한 도면을 참고하면, 각 댓글에 감정의 종류가 표현되어 있고 이를 확률로 계산된 것이 예시되어 있다.
- [0043] 예를 들면, 댓글이 "너무 재미있었고 유익한 수업이며 학생들을 잘 챙겨주는 모습이 좋았다."라고 게시되면,

이는 99.21%의 긍정리뷰로 추측한다는 내용으로 표시함으로써, 방송 주체가 실시간으로 댓글의 내용을 이해할 수 있도록 하는 것이다.

- [0044] 이러한 댓글 정량화 분석방법과 표시방법은 아래에서 상세히 설명한다.
- [0045] 다시 도 1을 참고하면, 단계 S110에서 댓글이 수집되면, 형태소 추출부(120)는 댓글 수집부(110)에서 수집된 댓글의 텍스트를 정제하여 형태소들을 추출하고, 추출된 형태소들 중 적어도 하나를 정량화 분석 모델부(130)로 입력한다(S120).
- [0046] 댓글은 텍스트 이외의 이모티콘의 형태, 은어, 약어, 그리고 축약된 표현 등이 수집될 수 있다.
- [0047] 일례로 수집된 댓글은 미리 정해진 방법으로 정제될 수 있다. 텍스트가 정제되면서 명사, 조사, 동사, 형용사, 특수 문자, 알파벳의 대소문자 등 중 적어도 하나 이상으로 분류될 수 있다.
- [0048] 미리 정해진 기준에 따라 형태소들 중 적어도 하나를 정량화 분석 모델부(130)로 입력되면, 정량화 분석 모델부(130)에서는 기설정된 정량화 감정분석 모델을 이용하여 감정을 분석하도록 한다.
- [0049] 일례로 정량화 감정분석 모델은 게시글의 내용을 추출하여 게시글의 분석을 위해 로지스틱 회귀 분석을 통한 지도 학습 방법을 수행하여 해당 댓글의 내용을 학습하고 딥 러닝 등의 기계 학습의 학습 모델을 기반으로 신규 댓글의 내용의 감정을 회귀식의 결과를 기준으로 댓글이 긍정인지 부정인지 또는 그 이외의 감정을 가지는지를 판단하기 위한 모델이다.
- [0050] 정량화 감정분석 모델을 생성하는 방식에 대해서는 아래에서 설명하기로 한다.
- [0051] 또한, 분류된 형태소들 중 명사만 입력할 수도 있고, 또는 명사를 비롯한 조사, 동사, 형용사, 특수 문자, 알파벳의 대소문자 및 이모티콘 중 적어도 하나가 입력될 수도 있다.
- [0052] 단계 S140에서는 정량화 감정분석 모델로부터 정량화 감정분석 결과를 획득하여 표시하거나 방송주체에게 실시간으로 피드백한다.
- [0053] 일례로 정량화 감정분석 결과를 감정의 종류 및 감정 점수 중 적어도 하나를 포함할 수 있다. 감정의 종류로는 긍정, 부정의 감정을 포함하되 치우치지 않은 중립의 감정이 포함될 수 있고, 긍정 및 부정의 감정이 세분화된 정량화 감정분석 결과를 획득할 수 있다. 정량화 분석 모델은 정량화 감정분석 결과를 수치로 출력할 수 있다.
- [0054] 예를 들어, 긍정 75%, 부정 25%의 수치가 출력된 경우, 상기 수치에 의해 감정 종류를 긍정으로 출력할 수 있고, 감정 점수로 75점이 출력될 수 있고 또는 별점이나 미리 정해진 기준에 따라 상, 중, 하로 표시될 수도 있다.
- [0055] 일례로 명사가 정량화 감정분석 모델에 입력되면, 정량화 감정분석 모델은 댓글에 포함되어 있는 명사들을 기준으로 댓글의 내용이 긍정인지 부정인지 또는 다른 감정을 포함하는지 판단할 수 있으며, 추가적으로, 감정의 종류에 대응하는 정도를 수치적으로 점수화하여 나타낼 수도 있다.
- [0056] 이하, 도면을 참고하여 정량화 감정분석 모델에서 정량화 감정분석 결과를 도출하는 방법에 대하여 설명한다.
- [0057] 도 2는 본 발명의 일실시예에 의한 정량화 분석 모델을 생성하는 방법을 설명하기 위한 도면으로, 도시된 바와 같이 정량화 감정분석 모델은 상술한 바와 같이 로지스틱 회귀 분석을 통한 지도 학습 방법을 수행하여 댓글의 내용을 학습함으로써 생성될 수 있다.
- [0058] 이를 위하여 다양한 사이트에 게시된 게시글 각각으로부터 댓글을 수집하고, 각 댓글을 이용하여 다양한 학습 데이터(210)들을 생성할 수 있다. 학습 데이터(210)는 댓글 및 댓글의 출처(게시글 및 사이트)에 대한 정보, 댓글 게시자의 정보 등을 포함하되, 정량화 감정분석 모델의 학습 데이터(210)에 대응하는 학습 결과물(230)이 출력되는 학습 데이터(210)가 선별적으로 입력될 수 있다.
- [0059] 학습 데이터(210)는 선형회귀의 기계 학습인 인공지능학습(220)을 통해 학습 결과물(230)을 출력하고, 이러한 학습 결과는 딥 러닝 기반의 감정 분석 모델인 학습결과물(230)로 생성될 수 있다. 학습 결과물(230)은 댓글에 대응하는 감정 종류 및/또는 감정 점수를 포함할 수 있다. 실시예와 같이, 특정 시점마다 댓글을 추출하여 저장하고 해당 댓글에 대한 기계 학습을 통해 정량화 감정분석 모델(240)로 생성할 수 있다.
- [0060] 또한, 정량화 감정분석 모델(240)은 학습 시 댓글의 게시글을 게시하는 사이트를 소비하는 사용자의 성향에 기초하여 또는 주 사용자의 연령대, 성별 등을 고려하여 특정 단어 또는 특정 품사에 대해서 가중치를 적용하여

학습될 수 있다.

- [0061] 예를 들어, 뉴스 기사를 소비하는 주 연령대에 따라 또는 쇼핑물을 이용하는 성별에 따라 가중치가 달리 적용되어 학습될 수 있다.
- [0062] 또는, 감정분석 모델로 입력이 결정될 시, 입력하고자 하는 형태소들에 대해서 가중치가 설정될 수 있다.
- [0063] 이하에서는 인공지능 기반의 정량화 감정분석 모델을 이용하는 실시예에 대해 설명하도록 한다. 정량화 감정분석 모델이 학습된 입력에 따라 정량화 감정분석 모델은 달리 생성될 수 있음은 물론이다.
- [0064] 도 3은 본 발명의 일실시예에 의한 정량화 감정분석 모델의 일례를 도시한 도면으로, 도시된 바와 같이, 수집된 댓글은 앞서 설명한 바와 같이 특정 시점에 사이트에 게시된 게시글에 달린 댓글들이 수집된 것으로 감정분석 모델로 입력되기 위한 사이트 정보, 게시글 정보 등을 포함할 수 있다.
- [0065] 정량화 감정분석 모델은 댓글의 게시글을 게시하는 사이트를 소비하는 사용자의 성향에 기초하여 또는 주 사용자의 연령대, 성별 등을 고려하여 특정 단어 또는 특정 품사에 대해서 가중치를 적용하여 기계 학습된 모델에 해당할 수 있다.
- [0066] 정량화 감정분석 모델은 댓글의 전처리를 통해 정제된 명사만을 입력으로 받을 수도 있고, 명사 이외의 조사, 동사, 형용사, 특수 문자 알파벳 대소문자 및 이모티콘, 은어, 약어 그리고 축약된 표현 등 중 적어도 하나의 키워드를 명사와 더불어 정량화 감정분석 모델로 입력될 수 있다.
- [0067] 예를 들어, 댓글에 포함되는 명사 각각과 명사에 연결된 조사를 포함하여 문법적 관계를 입력으로 받는 실시예, 명사들과 댓글 전체의 동사를 받는 실시예, 명사들과 댓글에 포함된 특수 문자들을 받는 실시예 등이 가능하며, 각각의 실시예를 위한 정량화 감정분석 모델은 달리 생성될 수 있다.
- [0068] 상술한 바와 같이 정량화 감정분석 모델의 출력으로 감정 종류 및/또는 감정 점수가 포함될 수 있다.
- [0069] 예를 들어, 정량화 감정분석 모델은 입력에 대응하는 감정들에 대한 확률을 수치적으로 출력할 수 있다. '긍정' 및 '부정'의 두 극단적인 감정 종류에 대한 확률을 각각 계산하거나, '희', '로', '애', '락' 등의 감정들 각각에 대한 확률을 계산할 수 있으며, 각각의 출력을 위해 정량화 감정분석 모델은 다른 형태로 학습될 수 있다.
- [0070] 감정 종류는 정량화 감정분석 모델에 미리 정해진 기준 값에 따라 회귀선을 기준으로 긍정, 중립, 부정으로 분리될 수 있다. 일례로, 특정 댓글에 대한 감정 분석 모델의 학습 결과가 '긍정' 72%, '부정' 25%으로 나타난 경우, 감정 종류는 '긍정'으로 판단될 가능성이 높다. 더불어, 감정 분석 모델에서 학습된 확률에 따라서 감정 점수가 결정될 수 있으며, 상기의 일례의 경우, '긍정 72점'의 점수가 결정되거나, '긍정'의 확률과 '부정'의 확률의 차이 또는 다른 계산법에 의해 점수가 계산될 수도 있다.
- [0071] 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, 컨트롤러, 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, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 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.
- [0072] 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, 가상장치(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 하나 이상의 컴퓨터 판독 가능 기록 매체

에 저장될 수 있다.

[0073] 본 발명의 일실시예에 의한 초고속 분석 인공지능을 이용한 실시간 댓글 판단 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 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

[0074] 롬 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0075]

[0076] 이상에서 본 발명은 기재된 구체예에 대하여 상세히 설명되었지만 본 발명의 기술사상 범위 내에서 다양한 변형 및 수정이 가능함은 당업자에게 있어서 명백한 것이며, 이러한 변형 및 수정이 첨부된 특허 청구범위에 속함은 당연한 것이다.

부호의 설명

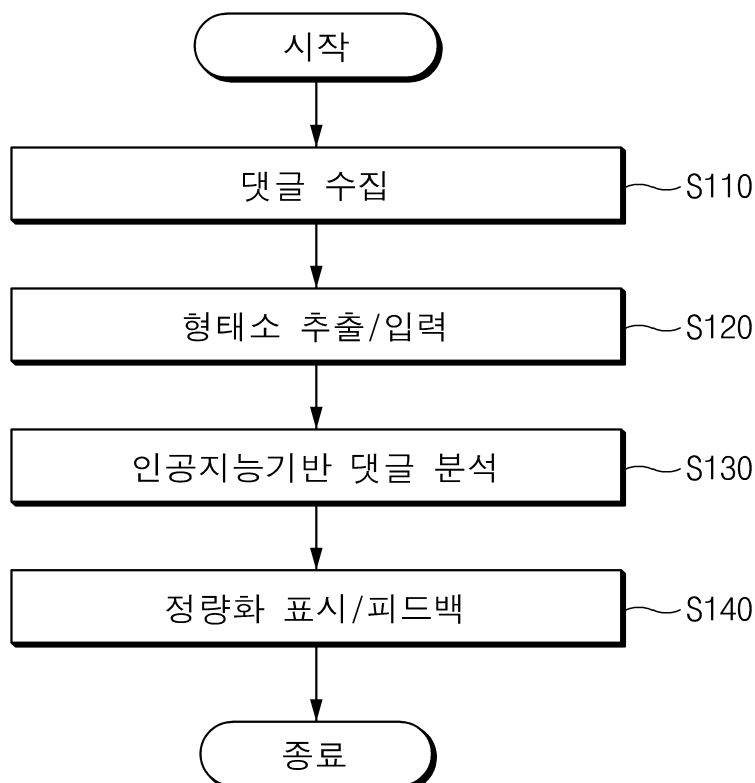
[0077]

110 : 댓글수집부 120 : 형태소추출입력부

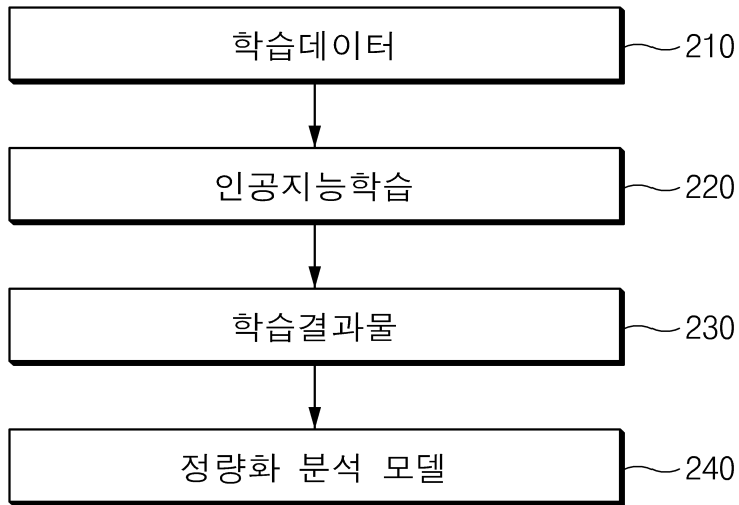
130 : 정량화 모델 분석부 140 : 표시/피드백부

도면

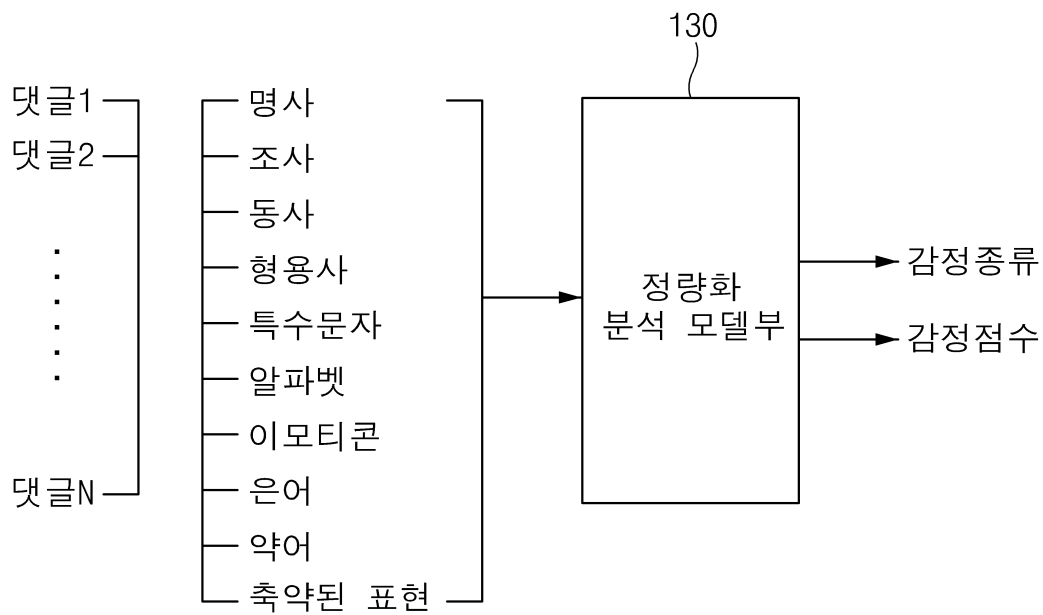
도면1



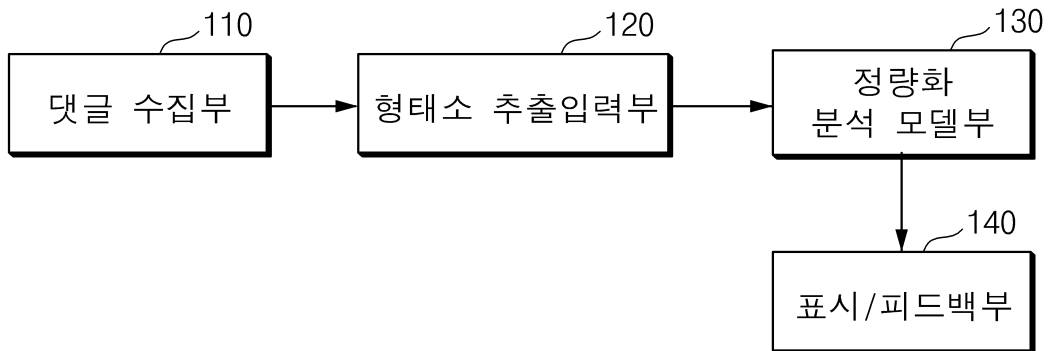
도면2



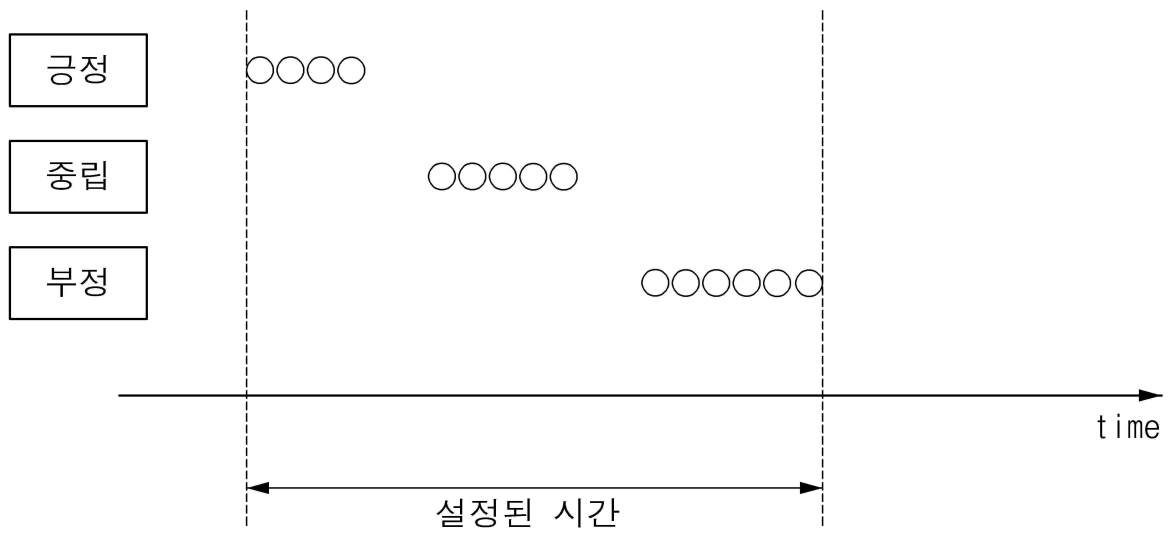
도면3



도면4



도면5



도면6

In [20]: len(X_test)

Out[20]: 5383

```
In [21]: %%time
# 인공지능 모델로 예측 후 100점만점 5점만점 긍정/부정 분류
test_results=loaded_model.predict(X_test)
test_results_100 = [round(test_results[i][0]*100,0) for i in range(len(test_results))]
test_results_5 = [int((round(test_results[i][0]*100,0)-1)/20)+1 for i in range(len(test_results))]

pos_neg = []
for i in range(len(test_results)):
    if (test_results[i][0]>0.5):
        pos_neg.append('Positive')
    else:
        pos_neg.append('Negative')
```

Wall time: 512 ms

도면7

```
In [21]: predict_pos_neg_cv(['너무 재미있었고 유익한 수업이며 학생들을 잘 챙겨주는 모습이 좋았다'])
predict_pos_neg_cv(['그냥 나가버리고 싶었다'])
predict_pos_neg_cv(['다시 듣고 싶다'])
predict_pos_neg_cv(['다시는 듣고 싶지 않아요'])
predict_pos_neg_cv(['학생들과의 허물없는 소통이 너무 좋았습니다. 한 학기동안 수고하셨습니다!'])
predict_pos_neg_cv(['광이었어요'])
predict_pos_neg_cv(['짱이에요'])
predict_pos_neg_cv(['정말 재미없고 지루했다'])
predict_pos_neg_cv(['강의를 이해하기 쉬웠고, 수업 참여도에 대해 상대적으로 자율성을 보장하여 학생들로 하여금 더 편안한 환경을 갖게 하였다
정 리뷰로 추천합니다 ^^
```

'너무 재미있었고 유익한 수업이며 학생들을 잘 챙겨주는 모습이 좋았다'는 99.21% 확률: 긍정 리뷰로 추천합니다 ^^

'그냥 나가버리고 싶었다'는 83.20% 확률: 부정 리뷰로 추천합니다 ^^

'다시 듣고 싶다'는 87.48% 확률: 긍정 리뷰로 추천합니다 ^^

'다시는 듣고 싶지 않아요'는 80.24% 확률: 부정 리뷰로 추천합니다 ^^

'학생들과의 허물없는 소통이 너무 좋았습니다. 한 학기동안 수고하셨습니다!'는 78.60% 확률: 긍정 리뷰로 추천합니다 ^^

'광이었어요'는 92.97% 확률: 부정 리뷰로 추천합니다 ^^

'짱이에요'는 95.79% 확률: 긍정 리뷰로 추천합니다 ^^

'정말 재미없고 지루했다'는 97.23% 확률: 부정 리뷰로 추천합니다 ^^

'강의를 이해하기 쉬웠고, 수업 참여도에 대해 상대적으로 자율성을 보장하여 학생들로 하여금 더 편안한 환경을 갖게 하였다.'는 94.05% 확률: 긍정 리뷰로 추천합니다 ^^