



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2014년12월03일  
(11) 등록번호 10-1465756  
(24) 등록일자 2014년11월20일

(51) 국제특허분류(Int. Cl.)  
G06F 17/00 (2006.01) G06F 17/21 (2006.01)  
G06Q 50/10 (2012.01)  
(21) 출원번호 10-2013-0149153  
(22) 출원일자 2013년12월03일  
심사청구일자 2013년12월03일  
(56) 선행기술조사문헌  
논문1(2013.06)  
KR1020090068803 A

(73) 특허권자  
주식회사 그리핀  
서울특별시 동대문구 경희대로 26 ,511(회기동,경희대학교창업보육센터네오르네상스관지하)  
(72) 발명자  
이철성  
서울특별시 성북구 안암로9가길 73, (안암동5가)조이캐슬 A-402  
(74) 대리인  
조성제

전체 청구항 수 : 총 8 항

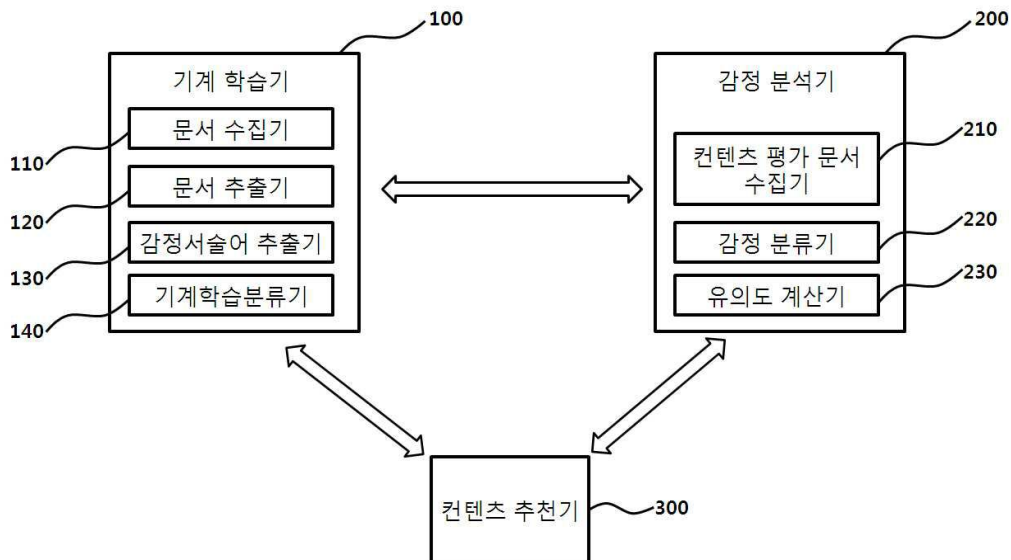
심사관 : 이석형

(54) 발명의 명칭 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법

(57) 요약

본 발명은 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법에 관한 것으로, 문장 수집을 통해 감정 형태소를 사용하여 학습데이터를 생성하고 감정별로 학습데이터를 분류한 기계학습기와, 테스트할 문장을 기계학습기를 통해 분류하여 테스트 문장의 감정 분포를 분석하는 감정 분석기 및 상기 감정 분석기를 통해 분석된 테스트 문장의 작성자에게 유사 느낌의 콘텐츠를 추천하는 콘텐츠 추천기를 포함하는 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법을 제공한다.

대 표 도 - 도1



## 특허청구의 범위

### 청구항 1

문장 수집을 통해 감정 형태소를 사용하여 학습데이터를 생성하고 감정별로 학습데이터를 분류한 기계학습기;  
테스트할 문장을 기계학습기를 통해 분류하여 테스트 문장의 감정 분포를 분석하는 감정 분석기; 및  
상기 감정 분석기를 통해 분석된 테스트 문장의 작성자에게 콘텐츠를 추천하는 콘텐츠 추천기를 포함하며,  
상기 감정 분석기는 온라인 상에서 영화평을 수집하는 콘텐츠 평가 문서 수집기와, 수집된 영화평을 기계학습기를 통해 분류하고, Z-점수를 계산하여 표준화 한 다음 이를 순위별로 산출하는 감정 분류기와, 표준화된 Z-점수 분포를 바탕으로 영화간 유의도를 계산하는 유의도 계산기를 포함하며,  
상기 콘텐츠 추천기는 영화별로 감정에 따른 Z-점수를 저장하며, 사용자의 영화평을 분석하여 감정에 따른 Z-점수를 산출하고, 산출된 Z-점수에 대응하는 영화를 추천하는 것을 특징으로 하는 감정 분석 장치.

### 청구항 2

제1항에 있어서,

상기 기계 학습기는 온라인상의 문서를 포함하는 문서를 수집하는 문서 수집기와, 수집된 문서에서 한글 문서를 추출하는 문서 추출기와, 감정 형태소가 서술어로 사용된 문서를 추출하는 감정서술어 추출기 그리고, 감정서술어 추출기의 결과를 이용하여 학습데이터를 생성 및 학습시키는 기계학습분류기를 포함하는 것을 특징으로 하는 감정 분석 장치.

### 청구항 3

제2항에 있어서,

상기 감정서술어 추출기는 감정형태소가 서술어로 사용된 문장을 추출하고,

상기 기계학습 분류기는 감정을 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension), 생동감(vigor)으로 분류하는 것을 특징으로 하는 감정 분석 장치.

### 청구항 4

제2항에 있어서,

상기 감정서술어 추출기를 이용하여 문장의 서술어에 감정형태소가 있고, 의문문이나 부정어로 사용되는 경우를 제외하고, 다수의 감정형태소가 모두 동일한 분류의 감정으로 사용된 경우만을 상기 학습 데이터로 사용하는 것을 특징으로 하는 감정 분석 장치.

### 청구항 5

삭제

### 청구항 6

제1항에 있어서,

상기 감정 분석기는 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension), 생동감(vigor)의 감정 중 ‘친근감’ 과 ‘분노’, ‘혼란’, ‘피로감’ 이 반대되는 감정으로 사용되고, 영화 평점에서는 친근감이 긍정적인 영향을 미치고, 분노, 혼란, 피로감이 부정적인 영향을 미치는 것을 특징으로 하는 감정 분석 장치.

## 청구항 7

삭제

## 청구항 8

제1항, 제2항, 제3항, 제4항 또는 제6항 중 어느 한 항에 따른 감정 분석 장치를 이용한 영화 추천 방법에 있어서,

한글 문서를 추출하는 단계;

추출된 문서 중 감정 형태소가 서술어로 사용된 문서를 선별하는 단계;

선별된 문서 중 다수의 감정 형태소가 동일한 분류의 감정 형태소로 사용된 사용된 문서를 추출하여 학습데이터를 생성한 후 학습하는 단계;

영화평을 수집하는 단계;

학습된 모델을 바탕으로 수집된 영화평에 나타난 감정 비율을 각 영화별로 산출하는 단계;

산출된 감정 비율을 Z-점수를 통해 정규화하여 영화별로 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension) 및 생동감(vigor)의 감정 분류에 따라 분류하는 단계;

사용자 요청에 따른 감정에 해당하는 영화 리스트를 제공하는 단계를 포함하는 것을 특징으로 하는 감정 분석 장치를 이용한 영화 추천 방법.

## 청구항 9

제8항에 있어서,

상기 학습데이터를 생성하고, 학습하는 단계는,

다항 네이브 베이즈 모델을 사용하여 문장을 형태소 단위로 분리하고, 특정 감정에서 형태소 출현 비율인 점유율을 이용하여 생성하는 것을 특징으로 하는 감정 분석 장치를 이용한 영화 추천 방법.

## 청구항 10

제8항에 있어서,

상기 Z-점수를 이용하여 감정보율을 정규화(standardization)는,

하기 수학적식에 따라 수행되는 것을 특징으로 하는 감정 분석 장치를 이용한 영화 추천 방법.

(수학적식)

$$z(x_{me}) = \frac{x_{me} - \mu_e}{\sigma_e}$$

여기서,  $x_{me}$ 는 영화 m에서 감정 e의 감정보율,  $\mu_e$ 는 감정 e의 평균,  $\sigma_e$ 는 감정 e의 표준편차임.

**명세서**

**기술분야**

[0001] 본 발명은 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법에 관한 것으로, 사용자가 작성한 문서의 분석

을 통해 사용자의 감정을 분석하고, 분석된 감정을 다양한 영역에 사용할 수 있고, 특히, 작성된 영화평을 이용하여 자신이 선호하는 영화와 비슷한 감정의 영화를 찾아주는 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법을 제공한다.

## 배경 기술

- [0002] 기존의 감정 분석은 사용자의 호불호 만을 측정하였다. 특히 영화평 같은 경우에도 추천 또는 비추천 2가지 중 어느 하나만을 선택하도록 하였기 때문에, 최종 결과는 추천 여부에 따른 결과만을 확인할 수 있었다. 따라서, 호불호가 확실하지 않은 감정에 대하여서는 판단이 어려운 단점이 있었다. 즉, 좋거나 싫은 감정 이외의 다른 감정에 대하여서는 그 분석이 어려웠다.
- [0003] 물론, 최근들어 사용자 또는 관람객의 감정을 분석하고 그 결과에 따라 사용자의 상태를 대변하거나, 현 사용자의 감정 상태에 맞는 행위나 서비스를 제공하는 기술에 관한 개발이 활발히 진행중이다.
- [0004] 그러나 기존의 감정 분석을 문장에 대한 단순 코드 부여로 인한 것으로 정확한 감정 상태의 분석이 용이하지 않고, 많은 오류를 포함하고 있다.

## 선행기술문헌

### 특허문헌

- [0005] (특허문헌 0001) (특허 문헌 1) 한국공개특허 제2002-0042248호  
(특허문헌 0002) (특허 문헌 2) 한국공개특허 제2005-0116254호  
(특허문헌 0003) (특허 문헌 3) 한국공개특허 제2009-0027262호  
(특허문헌 0004) (특허 문헌 4) 한국공개특허 제2013-0009360호

## 발명의 내용

### 해결하려는 과제

- [0006] 본 발명은 상술한 문제점을 해결하기 위하여 안출된 것으로서, 사용자 또는 관람객이 남긴 평(글)을 통해 사용자 또는 관람객이 느낀 감정을 분석하되, 좋고 싫음의 감정을 넘어 7단계의 감정으로 분리 평가할 수 있으며, 영화평을 이용하여 감정 비율에 따라 각 영화를 분류하여 그 결과를 사용자에게 추천 정보로 제공할 수 있는 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법을 제공하는 것을 그 목적으로 한다.

### 과제의 해결 수단

- [0007] 본 발명에 따른 문장 수집을 통해 감정 형태소를 사용하여 학습데이터를 생성하고 감정별로 학습데이터를 분류한 기계학습기와, 테스트할 문장을 기계학습기를 통해 분류하여 테스트 문장의 감정 분포를 분석하는 감정 분석기 및 상기 감정 분석기를 통해 분석된 테스트 문장의 작성자에게 유사 느낌의 콘텐츠를 추천하는 콘텐츠 추천기를 포함하는 감정 분석 장치를 제공한다.
- [0008] 상기 기계 학습기는 온라인상의 문서를 포함하는 문서를 수집하는 문서 수집기와, 수집된 문서에서 한글 문서를 추출하는 문서 추출기와, 감정 형태소가 서술어로 사용된 문서를 추출하는 감정서술어 추출기 그리고, 감정서술어 추출기의 결과를 이용하여 학습데이터를 생성 및 학습시키는 기계학습분류기를 포함한다.
- [0009] 상기 감정서술어 추출기는 감정형태소가 서술어로 사용된 문장을 추출하고,
- [0010] 상기 기계학습 분류기는 감정을 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension), 생동감(vigor)으로 분류하는 것을 특징으로 한다.

- [0011] 상기 학습데이터로 감정서술어 추출기를 이용하여 문장의 서술어에 감정형태소가 있고, 의문문이나 부정어로 사용되는 경우를 제외하고, 다수의 감정형태소가 모두 동일한 감정으로 사용된 경우만을 사용하는 것을 특징으로 한다.
- [0012] 상기 감정 분석기는 온라인 상에서 영화평을 수집하는 콘텐츠 평가 문서 수집기와, 수집된 영화평을 기계학습기를 통해 분류하고, Z-점수를 계산하여 표준화 한 다음 이를 순위별로 산출하는 감정 분류기와, 표준화된 Z-점수 분포를 바탕으로 영화간 유의도를 계산하는 유의도 계산기를 포함하는 것을 특징으로 한다.
- [0013] 상기 감정 분석기는 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension), 생동감(vigor)의 감정 중 ‘친근감’ 과 ‘분노’, ‘혼란’, ‘피로감’ 이 반대되는 감정으로 사용되고, 영화 평점에서는 친근감이 긍정적인 영향을 미치고, 분노, 혼란, 피로감이 부정적인 영향을 미치는 것을 특징으로 한다.
- [0014] 또한, 본 발명에 따른 한글 문서를 추출하는 단계와, 추출된 문서 중 감정 형태소가 서술어로 사용된 문서를 선별하는 단계와, 선별된 문서 중 단일 감정의 형태소가 사용된 문서를 추출 후 학습하는 단계 및 학습된 학습 데이터를 이용하여 제공된 문장의 기계적 분석을 통해 7가지의 감정 중 하나의 감정을 선택하는 단계를 포함하는 것을 특징으로 하는 감정 분석 방법을 제공한다.
- [0015] 또한, 본 발명에 따른 감정 분석 방법을 이용한 영화 추천 방법에 있어서, 한글 문서를 추출하는 단계와, 추출된 문서 중 감정 형태소가 서술어로 사용된 문서를 선별하는 단계와, 선별된 문서 중 단일 감정의 형태소가 사용된 문서를 추출하여 학습데이터를 생성한 후 학습하는 단계와, 영화평을 수집하는 단계와, 학습된 모델을 바탕으로 수집된 영화평에 나타난 감정 비율을 각 영화별로 산출하는 단계와, 산출된 감정 비율을 Z-점수를 통해 정규화하여 영화별로 7가지 감정 분류에 따라 분류하는 단계와, 사용자 요청에 따른 감정에 해당하는 영화 리스트를 제공하는 단계를 포함하는 것을 특징으로 하는 감정 분석 방법을 이용한 영화 추천 방법을 제공한다.
- [0016] 상기 학습데이터를 생성하고, 학습하는 단계는, 다항 네이브 베이즈 모델을 사용하여 문장을 형태소 단위로 분리하고, 특정 감정에서 형태소 출현 비율인 점유율을 이용하여 생성하는 것을 특징으로 한다.
- [0017] 상기 Z-점수를 이용하여 감정비율을 정규화(standardization)는, 하기 수학적식에 따라 수행되는 것을 특징으로 한다.

### 발명의 효과

- [0018] 이와 같이 하여 본 발명은 사용자의 감정을 추출하여 7가지 감정으로 분류할 수 있고, 특히, 영화 감상평을 이용하여 관객이 느낀 감정을 분석하고, 분석된 결과를 이용하여 선호하는 영화와 유사한 감정을 느낄 수 있는 영화를 추천할 수 있으며, 영화를 보고 느낀 감정별 순위를 제공할 수 있다.

### 도면의 간단한 설명

- [0019] 도 1은 본 발명의 일 실시예에 따른 감정 분석 장치의 블록도.  
 도 2는 일 실시예에 따른 기계학습 분류기의 블록도.  
 도 3은 일 실시예에 따른 최소 점유율에 따른 기계학습 모델의 분류 정확도의 계산 결과 그래프.  
 도 4는 본 실시예에 따른 영화장르별 Z-점수 평균 계산 정리 그래프.  
 도 5는 본 발명의 일 실시예에 따른 감정 분석 방법 및 영화 추천 방법을 설명하기 위한 흐름도.

### 발명을 실시하기 위한 구체적인 내용

- [0020] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 더욱 상세히 설명하기로 한다. 그러나 본 발명은 이하에서 개시되는 실시예에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이다. 도면상에서 동일 부호는 동일한 요소를 지칭한다.
- [0021] 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능별로 구분한 것에 불과함을 명확히 하고

자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다. 따라서, 본 명세서를 통해 설명되는 각 구성부들의 존재 여부는 기능적으로 해석되어야 할 것이다. 이러한 이유로 본 발명의 감정 분석 장치 및 방법과 이를 이용한 영화 추천 방법의 구성부들의 구성은 본 발명의 목적을 달성할 수 있는 한도 내에서 상이해질 수 있음을 명확히 밝혀둔다.

- [0022] 도 1은 본 발명의 일 실시예에 따른 감정 분석 장치의 블록도이다. 도 2는 일 실시예에 따른 기계학습 분류기의 블록도이다. 도 3은 일 실시예에 따른 최소 점유율에 따른 기계학습 모델의 분류 정확도의 계산 결과 그래프이다. 도 4는 본 실시예에 따른 영화장르별 Z-점수 평균 계산 정리 그래프이다.
- [0023] 도 1 내지 도 4에 도시된 바와 같이 본 실시예에 따른 감정 분석 장치는 문장 수집을 통해 감정 형태소를 사용하여 학습데이터를 생성하고 감정별로 학습데이터를 분류한 기계학습기(100)와, 테스트할 문장을 기계학습기를 통해 분류하여 테스트 문장의 감정 분포를 분석하는 감정 분석기(200)와, 감정 분석기를 통해 분석된 테스트 문장의 작성자에게 유사 느낌의 콘텐츠를 추천하는 콘텐츠 추천기(300)를 포함한다.
- [0024] 상기 기계 학습기(100)는 온라인상의 문서를 포함하는 문서를 수집하는 문서 수집기(110)와, 수집된 문서에서 한글 문서를 추출하는 문서 추출기(120)와, 감정 형태소가 서술어로 사용된 문서를 추출하는 감정서술어 추출기(130) 그리고, 감정서술어 추출기(130)의 결과를 이용하여 학습데이터를 생성 및 학습시키는 기계학습분류기(140)를 포함한다.
- [0025] 문서 수집기(110)는 트위터 상의 문서를 자동으로 수집하는 장치를 사용하는 것이 바람직하다. 본 실시예에서는 트위터 API를 사용하는 것이 효과적이다. 예를 들어, 2011년12월 기준 데이터로 수집시 사용자 약 1천9백만명이고, 트윗은 약 2억개가 된다. 물론, 문서 수집지는 트위터 이외의 다양한 온라인 상의 문서를 수집할 수 있다. 또한, 수집된 문서를 별도의 저장 공간에 저장할 수도 있다.
- [0026] 문서 추출기(120)는 다양한 언어의 트윗 중에서 한글이 포함된 문서를 추출한다. 이는 한국을 기준으로 실시한 것으로, 해당 국가에 따라 문서 추출기에 의해 분류되는 문서의 언어는 가변될 수 있다. 이때, 문서 추출기는 이름, 위치 및 자기 소개에 한글을 사용한 사용자의 문서를 한글 문서로 추출하거나 1000건 이상의 트윗을 작성한 사용자중 한글이 사용된 사용자의 문서를 한글 문서로 추출할 수도 있다. 앞선 문서 수집기(110)의 예에서와 같이 약 2억개의 트윗 중에서 약 32만여개를 추출한다.
- [0027] 감정서술어 추출기(130)는 감정형태소가 서술어로 사용된 문장을 추출한다. 이때, 문장의 서술어에 감정형태소가 있으면, 그 문장의 감정은 감정형태소의 감정으로 결정한다. 이때, 서술어는 종결어미를 갖는 단어를 지칭한다.
- [0028] 본 실시예의 기계학습 분류기(140)는 POMS(Profile of Mood States)이론을 적용한다. 사람의 감정을 7개의 감정, 즉, 분노(anger), 혼란(confusion), 우울(depression), 피로감(fatigue), 친근감(friendliness), 긴장감(tension), 생동감(vigor)으로 분류한다. 이와 같이 7개의 감정을 적용한 기계학습 이론의 클래스(class)로 사용한다. 그리고, 기계학습 분류기(즉, 모델)은 베이지 확률모델(Bayes Probability Model)과 SVM(Support Vector Machine)을 적용할 수 있다. 이때, 베이지 확률모델로는 형태소의 출현 여부에 따라 참, 거짓의 이진값으로 표현하는 이진 독립 나이브 베이지(Binary Independence Naive Bayes) 확률모델과 형태소의 출현 개수로 표현하는 다항 나이브 베이지(Multinomial Naive Bayes) 확률모델을 적용한다. 그리고 SVM의 적용은 데이터 마이닝 툴인 “Rapid Miner 5”를 이용하였고, SVM의 커널함수(kernel function)로 선형(linear)함수, 다항(polynomial)함수, RBF(Radial Basis Function)를 사용하였다.
- [0029] 기계학습분류기(140)는 기계학습 분류 모듈(141)은 물론, 감정 형태소를 생성하는 형태소 모듈(142)과, 학습데이터를 생성하는 학습데이터 모듈(143)을 더 구비할 수 있다.
- [0030] 여기서, 감정 형태소는 감정이 결정된 형태소를 지칭하고, 학습 데이터를 자동으로 생성하는 역할을 한다.
- [0031] 앞서 언급한 바와 같이 본 실시예에서는 POMS(Profile of Mood States)의 요인분석(factor analysis) 결과를 기초로 감정형태소를 생성한다. 총 7개의 감정으로 군집화하여 단어가 내포하는 감정을 결정한다. 7개의 감정은 분노, 당황, 우울, 피로감, 친근감, 긴장감, 생동감으로 구성된다. 형태소 분석기를 적용한 후 품사가 일반동사



(pvg), 정상 형용사(paa), 동작성 명사(ncpa), 상태성 명사(ncpa)인 그램(gram) 형태소를 선별하였다. 이때, 1-gram 형태소 54개만을 선별하고, POMS 감정 형태소로 하기 표 1과 같이 54개의 단어를 추출한다. 즉, 표 1은 POMS로 부터 추출한 감정과 감정 형태소 리스트이다.

표 1

[0032]

| 분노      | 혼란       | 우울       | 피로감      | 친근감     | 긴장감        | 생동감       |
|---------|----------|----------|----------|---------|------------|-----------|
| 분개/ncpa | 당황/ncpa  | 낙심/ncpa  | 기진맥진/ncp | 공감/ncpa | 거북/ncps    | 생기/ncpa   |
| 불쾌/ncps | 당황/ncps  | 불행/ncps  | a        | 다정/ncn  | 겁나/pvg     | 원기왕성/ncps |
| 심술나/pvg | 잇/pvg    | 비참/ncps  | 느긋/ncps  | 다정/ncps | 긴장/ncpa    | s         |
| 앙심품/pvg | 잊어버리/pvg | 슬프/paa   | 느리/paa   | 명석/pvg  | 떨리/pvg     | 쾌활/ncps   |
| 언짢/paa  | 혼란/ncps  | 외롭/paa   | 둔하/paa   | 믿/pvg   | 무섭/paa     | 편안/ncps   |
| 짜증나/pvg |          | 우울/ncps  | 따분/ncps  | 믿어지/pvg | 불안/ncps    | 활기차/paa   |
| 폭발/ncpa |          | 울적/ncps  | 지치/pvg   | 착하/paa  | 안절부절못하/pvg | 활동/ncpa   |
| 화가나/pvg |          | 자포자기/ncp | 피곤/ncps  | 친절/ncps | 조마조자하/paa  | 활발/ncps   |
| 화나/pvg  |          | a        |          |         | 초조/ncps    |           |
|         |          | 후회/paa   |          |         |            |           |

[0033]

POMS의 요인분석 결과로부터 특정 감정을 나타내는 감정형태소를 확보하여 감정형태소를 확장하고, 이를 하기 표 2로 정리하였다.

표 2

[0034]

| 감정형태소     | 유의어     | 출현횟수    | 감정  |
|-----------|---------|---------|-----|
| 화남/pvg    | 성나/pvg  | 억울/ncps | 분노  |
| 당황/ncpa   | 당혹/ncpa | 황당/ncps | 당황  |
| 낙심/ncpa   | 낙담/ncpa | 속상/ncps | 우울  |
| 기진맥진/ncpa | 쓰러지/pvg | 귀찮/paa  | 피로감 |
| 공감/ncpa   | 동감/ncpa | 사랑/ncpa | 친근감 |
| 겁나/pvg    | 아찔/ncps | 두렵/pvg  | 긴장감 |
| 쾌활/ncps   | 명랑/ncps | 재밌/pvg  | 생동감 |

[0035]

앞선 예시에서와 같이 문서 추출기(120)를 통해 추출된 32만개의 한글 트윗을 형태소분석 후 형태소의 출현횟수가 최소 100회 이상인 형태소와 유의어 사전을 이용하여 명확한 감정의 구분이 가능하다고 판단되는 형태소, 이 두 집단을 수집하여 총 322개의 후보 감정 형태소를 분류할 수 있다. 즉, 이는 유의어 사전 후보 선별과 출현횟수 100회 이상 형태소에서 후보 선별을 하고, 후보 형태소로부터 투표방식으로 감정 결정을 한다.

[0036]

후보 감정형태소로부터 선별작업에 의해 감정형태소의 감정이 결정되고 이때의 감정별 감정형태소의 개수는 하기 표 3과 같다.

표 3

[0037]

| 분노 | 혼란 | 우울 | 피로감 | 친근감 | 긴장감 | 생동감 |
|----|----|----|-----|-----|-----|-----|
| 44 | 34 | 54 | 42  | 70  | 35  | 43  |

[0038]

학습데이터 모듈을 학습 데이터를 생성한다. 이때, 학습데이터로는 32만여 개의 한글 트윗(즉, 문장)에서 감정 표현이 명확한 경우만을 사용한다. 즉, 감정서술어 추출기를 이용하여 문장의 서술어에 감정형태소가 있으면, 그 문장의 감정은 감정형태소의 감정으로 결정하였다. 여기서 서술어는 종결어미를 갖는 단어를 지칭한다. 학습데이터로 추출 시에는 의문문이나 부정어로 사용되는 경우(~까요, ~하지 않다 등의 경우)는 제외하는 것이 바람직하다. 또한, 다수의 감정형태소가 하나의 문장에 사용된 경우, 각각의 감정형태소가 모두 동일한 감정으로 사용된 경우만 학습데이터로 이용하였다.

[0039]

이와 같은 방법(부정어, 의문문의 서술어는 제외, 2개 이상의 감정이 다른 감정 형태소가 쓰인 경우 제외)으로 1만 7천여 개의 트윗을 추출하였고, 감정에 따른 트윗의 개수는 표 4와 같다.

표 4

| 분노  | 혼란  | 우울   | 피로감  | 친근감  | 긴장감 | 생동감  |
|-----|-----|------|------|------|-----|------|
| 587 | 205 | 1827 | 1510 | 9126 | 899 | 3043 |

학습시에 트윗(문장)을 형태소 단위로 분리하여 자질로 사용한다. 이때, 형태소 선별 기준은 특정 감정에서의 형태소 출현 비율인 점유율을 이용하는 것이 바람직하다.

점유율(R)은 하기 수학적 식 1과 같이 정의한다.

수학적 식 1

$$R(m_i|E_n) = \frac{\text{Count}(m_i|E_n)}{\text{Count}(E_n)}$$

여기서, E는 감정, m은 형태소를 지칭한다.

상기와 같이 계산된 점유율의 최솟값을 기준으로 출현빈도가 낮은 형태소를 제거한다. 그리고, 이때의 최솟값을 최소점유율(Minimum Occupancy Ratio, MOR)이라 한다.

어떤 형태소가 적어도 하나의 감정에서 최소점유율 이상의 출현빈도를 가지게 되는 경우 자질로 선별하여 분석에 사용한다. 예를 들어 최소점유율이 0.200%일 때 동사 ‘재밌/pvg’은 점유율이 ‘우울’에서 0.198%이지만 ‘생동감’에서 14.2%이므로 자질로 사용하게 된다.

이때, 띄어쓰기 없이 여러 단어가 붙어있는 단어나 맞춤법에 어긋난 단어를 새로운 형태소로 인식하게 되고, 그 결과 새로운 자질로 추가되어 자질의 개수가 필요 이상으로 많아지게 된다. 하지만 이와 같은 비정상적인 형태소는 출현빈도가 낮기 때문에 최소점유율을 사용하여 제거가 가능하다. 따라서, 비정상적인 형태소가 대부분 제거가 되는 최소 점유율 0.2%부터 실험의 분류 정확도에 큰 차이를 보이지않는 1.6%까지 0.2% 단위로 8단계에 나눠서 학습을 하는 것이 효과적이다.

기계학습 분류 모듈(141)은 감정서술어 추출기를 통해 감정서술어가 포함되고, 같은 감정의 형태소가 하나의 문장(트윗)에 사용된 경우에만 학습데이터로 사용하여 학습시킨 분류 모듈을 지칭한다.

기계학습 분류기(140)에 적용된 앞선 이론들의 정확성 확립을 위해 실험을 통한 테스트를 실시하였다.

실험에서는 학습에 필요한 자질의 개수를 줄이면서 그렇지 않은 경우와 비슷한 결과를 보일 수 있을지를 판단하기 위해 2개의 품사군으로 나눠 실험하였다. 모든 품사를 자질로 전부 사용하는 전체품사군과 상태 성명사(ncps), 동작성명사(ncpa), 성상 형용사(paa), 일반 동사(pvg)로 이뤄진 감정품사군, 총 2개의 군으로 나뉘었으며, 감정품사군 자질 개수는 전체품사군 자질 개수의 1/3 정도이다.

테스트 데이터는 연구자 3명에 의한 수작업을 통해 산정하였고, 그 결과는 하기 표 5로 표시된다.

표 5

| 분노 | 혼란 | 우울 | 피로감 | 친근감 | 긴장감 | 생동감 |
|----|----|----|-----|-----|-----|-----|
| 39 | 20 | 26 | 21  | 44  | 30  | 36  |

그리고, 실험에서의 정확도(acc)는 하기 수학적 식 2를 통해 계산한다.



수학식 2

$$acc = \frac{\sum_{c \in E} N_{cc}}{N}$$

[0054]

[0055]

여기서, N은 테스트 전체 개수이고, 감정집합 E = {A, C, D, F, R, T, V}, N<sub>ij</sub>가 정답인 i 인 데이터를 j로 예측한 결과의 개수를 지칭한다.

[0056]

상기 수학식 2의 최소점유율에 따른 기계학습 모델의 분류 정확도를 계산하여 표 6의 결과를 얻었다. 그리고 그 결과그래프는 도 3과 같다.

표 6

[0057]

| 품사군        | 종류                | 학습<br>모델          | 최소점유율(%) |      |      |      |      |      |      |      |
|------------|-------------------|-------------------|----------|------|------|------|------|------|------|------|
|            |                   |                   | 0.2      | 0.4  | 0.6  | 0.8  | 1.0  | 1.2  | 1.4  | 1.8  |
| 비학습        |                   |                   | .204     |      |      |      |      |      |      |      |
| 감정 품<br>사군 | 베이지스<br>확률 모<br>델 | MNB*              | .398     | .421 | .407 | .407 | .403 | .403 | .403 | .407 |
|            |                   | NB**              | .227     | .282 | .236 | .273 | .245 | .232 | .273 | .255 |
|            | SVM               | C-SVC(li<br>near) | .431     | .431 | .421 | .431 | .431 | .426 | .417 | .417 |
|            |                   | C-SVC(po<br>ly)   | .426     | .426 | .426 | .426 | .431 | .431 | .426 | .421 |
|            |                   | C-SVC(rb<br>f)    | .403     | .417 | .421 | .421 | .426 | .431 | .431 | .431 |
| 전체 품<br>사군 | 베이지스<br>확률 모<br>델 | MNB*              | .477     | .477 | .519 | .486 | .500 | .505 | .491 | .500 |
|            |                   | NB**              | .241     | .273 | .301 | .301 | .357 | .333 | .370 | .338 |
|            | SVM               | C-SVC(li<br>near) | .417     | .421 | .394 | .384 | .407 | .384 | .380 | .380 |
|            |                   | C-SVC(po<br>ly)   | .352     | .352 | .361 | .343 | .352 | .352 | .361 | .366 |
|            |                   | C-SVC(rb<br>f)    | .310     | .296 | .301 | .319 | .333 | .329 | .338 | .333 |

[0058]

상기 표와 도면에서와 같이 비교실험대상(baseline)으로는 테스트데이터에서 가장 높은 감정 비율을 차지하는 ‘친근감’ 으roman 분류했을 경우인 비학습모델로 정하였고, 비학습모델의 정확도는 20.4%이다.

[0059]

비학습모델과 기계학습 모델을 비교했을 때 기계학습 모델 모두 비학습모델보다 정확도가 2배 이상 높았다. 그 중 전체 품사를 모두 자질로 사용한 다항 네이브 베이지 모델은 최소점유율에 상관없이 다른 기계학습 모델보다 높은 정확도를 보였다. 그리고 전체품사군을 자질로 한 다항 네이브 베이지 모델은 최소점유율이 0.6%일 때 가장 높은 정확도를 보였고, 이때의 분류 정확도는 51.9%였다.

[0060]

감정품사군만 자질로 선택했을 경우 입력데이터로 사용되는 자질의 개수는 전체품사군을 자질로 했을 경우의 1/3 정도이다. 학습 및 테스트에 사용되는 자질 개수가 적을수록 모델의 학습 및 테스트 속도가 빨라지기 때문에 감정품사군 만으로 전체품사군과 비슷한 정확도를 보인다면 더 효율적인 모델생성이 가능할 것으로 판단하였으나, 감정품사군에서 제외된 다른 품사들 - 일반명사, 조사 등 - 또한 중요한 역할을 하는 것으로 나타났다.

[0061]

여러 기계학습 모델을 적용하여 문서를 주제별로 분류하는 연구에서는 SVM이 베이지 확률모델보다 정확도가 높게 나타났으나, 문서의 감정을 분류하는 본 연구에서는 다항 네이브 베이지 모델이 SVM보다 정확도가 더 높았다. 이는 SVM의 경우 학습된 결과인 결정초평면의 해석이 어렵지만, 트릿의 경우 문서 길이가 140바이트로서 상대적으로 다른 온라인 문서들보다 짧기 때문이다. 즉, 형태소 출현횟수가 대부분 0이거나 1로서 편차가 작기 때문에 결정초평면이 감정 클래스를 정확하게 구분하지 못한 것이다.

- [0062] 마지막으로 SVM과 베이스 확률모델을 비교해 보면 베이스 확률모델에서는 감정품사군만 자질로 사용했을 때보다 전체품사군을 자질로 사용했을 경우 정확도가 더 높았다. 반대로 SVM의 경우에는 감정품사군만 자질로 고려한 경우가 전체품사군을 자질로 고려한 경우보다 정확도가 더 높았다.
- [0063] 상기와 같은 기계 학습기(100)를 이용하여 감정 분석기(200)는 감정 분석을 실시한다. 하기에서 영화평을 이용한 감정 분석을 중심으로 설명한다.
- [0064] 감정 분석기(200)는 영화평 수집기 즉, 컨텐츠 평가 문서 수집기(210)와, 감정 분류기(220) 그리고, 유의도 계산기(230)를 포함한다.
- [0065] 이때, 감정 분석기(200)에 적용된 감정간 특성으로는 ‘친근감’ 과 ‘분노’ , ‘혼란’ , ‘피로감’ 이 반대되는 감정으로 사용한다. 그리고, ‘생동감’ 과 ‘우울’ 이 상반되는 감정으로 사용한다. 그리고, 영화평에 나타난 ‘친근감’ 은 평점에 긍정적인 영향을 미치고, ‘분노’ , ‘혼란’ , ‘피로감’ 은 부정적인 영향을 미치며, ‘생동감’ 과 ‘우울’ 은 영화의 평점에 영향을 미치지 않는다. 그리고, 영화 장르별로, 공포 장르는 ‘긴장감’ , 코미디장르는 ‘생동감’ , 멜로/애정/로맨스 장르는 ‘친근감’ , 범죄장르는 ‘혼란’ 을 주요소로 하는 장르이다.
- [0066] 컨텐츠 평가 문서 수집기(210)는 온라인 상으로 부터 영화 평을 수집한다. 감정 분류기(220)는 수집한 영화평을 기계학습기를 통해 분류하고, Z-점수(Z-Score)를 계산하여 표준화하고, 감정별 순위를 산출한다. 유의도 계산기(230)는 표준화된 감정별 Z 점수 분포를 바탕으로 영화간 코사인 유의도(Cosine similarity)를 계산한다.
- [0067] 본 실시예에서는 컨텐츠 평가 문서 수집기(210)를 통해 영화평을 수집하되, 2008년1월부터 2012년4월까지의 네이버 상의 40자 영화평을 수집하였다. 이때, 영화의 영화평이 1000개 이상 작성된 영화로 제한하였다. 또한 영화의 장르는 ‘네이버 영화’ 에서 제공하는 메타데이터(metadata)를 이용하였다.
- [0068] 그리고, 앞선 실험에서 가장 높은 정확도를 보였던 다항 네이브 베이스 모델을 ‘네이버 40자 영화평’ 에 적용하여 각각의 영화평에 대해 감정을 분류하였고, 이 중 고지전이라는 영화에 대한 결과를 표 7로 정리하였다.

표 7

|           | 분노   | 혼란   | 우울   | 피로감  | 친근감   | 긴장감  | 생동감   |
|-----------|------|------|------|------|-------|------|-------|
| 고지전의 감정비율 | 5.9% | 0.6% | 8.5% | 2.0% | 58.6% | 2.0% | 22.3% |
| 전체 영화평균   | 3.8% | 0.5% | 8.9% | 3.7% | 57.3% | 2.2% | 23.5% |

- [0070] 표 7과 같이 “고지전” 의 결과를 보면 감정별 평균이 각각 달라 어느 감정이 이 영화의 특징으로 나타나는지를 판단하기 어렵다.
- [0071] 이에, Z-점수를 이용하여 감정비율을 정규화(standardization)를 실시하였다. Z-점수에 의한 감정 정규화는 하기 수식 3에 따라 수행되었다.

### 수식 3

$$z(x_{me}) = \frac{x_{me} - \mu_e}{\sigma_e}$$

[0072]

[0073] 여기서,  $x_{me}$ 는 영화 m에서 감정 e의 감정비율,  $\mu_e$ 는 감정 e의 평균,  $\sigma_e$ 는 감정 e의 표준편차이다.

[0074] 상기 고지전에 대한 Z-점수의 적용 결과는 하기 표 8과 같다.

표 8

|           | 분노   | 혼란   | 우울    | 피로감   | 친근감  | 긴장감   | 생동감   |
|-----------|------|------|-------|-------|------|-------|-------|
| 고지전의 Z-점수 | 0.99 | 0.38 | -0.09 | -0.97 | 0.21 | -0.08 | -0.18 |

이와 같이 Z-점수에 따른 정규화 결과를 통해 영화에서 느끼는 감정의 차이를 수치상으로 확인할 수 있다. 예를 들어 '고지전의 '친근감' 이 0.21인데 비해 '분노' 가 0.99로서 '고지전' 관람 후 '친근감' 보다 '분노' 의 감정을 더 많이 느꼈다고 볼 수 있다.

그리고, 약 100편의 영화에 대한 7개 감정의 Z-점수와 영화 평점을 기준으로 분석한 결과, 앞서 언급한 바와 같이 '친근감' 과 '분노' , '혼란' , '피로감' 이 반대되는 감정으로 사용하고, ' '친근감' 은 평점에 긍정적인 영향을 미치고, '분노' , '혼란' , '피로감' 은 부정적인 영향을 미치며, '생동감' 과 '우울' 은 영화의 평점에 영향을 미치지 않는다.

이를 위해 요인 분석을 하였고, 그 분석 결과를 표 9로 정리하였다.

표 9

|     | 요인1(불쾌) | 요인2(슬픔과 감동) |
|-----|---------|-------------|
| 피로감 | 0.887   | -0.01       |
| 평점  | -0.871  | 0.019       |
| 분노  | 0.816   | 0.211       |
| 혼란  | 0.767   | 0.125       |
| 친근감 | -0.705  | 0.405       |
| 긴장감 | 0.487   | 0.164       |
| 생동감 | -0.139  | -0.97       |
| 우울  | 0.038   | 0.766       |

표 8의 결과에서와 같이 주요인과의 요인적재량의 절댓값이 0.7이상일 경우에 그 요인과 밀접한 관계가 있다고 가정할 수 있다. 즉, 요인1과 '분노' , '혼란' , '피로감' 의 요인적재량이 각각 0.887, 0.816, 0.760으로서 0.7이상이기 때문에 이 3개의 감정들과 요인1은 밀접한 관계가 있음을 보였다. 그리고 영화평점과 '친근감' 은 요인적재량이 각각 -0.871, -0.703으로서 -0.7이하이므로 요인1과 상반되는 관계에 있음을 알 수 있다. 이는 영화평에서 '분노' , '혼란' , '피로감' 의 점수가 높으면 낮은 평점을 받을 수 있고, '친근감' 의 점수가 높으면 좋은 평점을 받을 수 있다는 의미이다.

요인2의 경우 '우울' 의 요인적재량이 0.764로서 0.7이상이기 때문에 요인2와 '우울' 은 밀접한 관계이고, '생동감' 의 요인적재량은 -0.971로서 -0.7이하이기 때문에 요인2와 매우 상반되는 관계이다.

따라서, '우울' 과 '생동감' 이 '슬픔과 감동' 요인에서 반대의 감정이고, '분노' , '혼란' , '피로감' , '긴장감' 과 '친근감' 은 '불쾌' 요인에서 서로 반대 감정이다. 그리고, 요인분석결과 '생동감' 과 '우울' 은 평점에 거의 영향을 미치지 않는다.

그리고, 영화를 공포, 액션, 미스터리, 스릴러, 드라마, 전쟁, 범죄,SF, 멜로, 코미디, 모험, 판타지, 가족, 느와르, 애니메이션, 뮤지컬로 16개의 장르로 분류한다. 이때, 10편 미만의 장르는 작은 표본으로 판단에서 고려하지 않는다.

이어서, 장르별 감정의 Z-점수 평균을 계산하여 하기 표 10과 같이 정리하였다.

표 10

| 장르  | 분노    | 혼란    | 우울    | 피로감   | 친근감   | 긴장감   | 생동감   | 평점   |
|-----|-------|-------|-------|-------|-------|-------|-------|------|
| 공포  | 0.41  | 0.55  | -0.14 | 0.63  | -0.52 | 1.93  | -0.53 | 6.86 |
| 코미디 | -0.41 | -0.52 | -0.59 | -0.31 | 0.16  | -0.45 | 0.65  | 8.04 |
| 스릴러 | 0.34  | 0.79  | 0.01  | 0.28  | -0.30 | 0.49  | -0.16 | 7.59 |
| 판타지 | -0.15 | -0.29 | 0.23  | 0.53  | -0.44 | -0.37 | 0.29  | 7.43 |
| 드라마 | -0.07 | -0.08 | 0.26  | -0.32 | 0.28  | -0.18 | -0.23 | 8.21 |

|               |       |       |       |       |       |       |       |      |
|---------------|-------|-------|-------|-------|-------|-------|-------|------|
| 멜로/애정/로<br>맨스 | -0.15 | -0.20 | 0.03  | -0.36 | 0.65  | -0.42 | -0.25 | 7.96 |
| 액션            | 0.05  | 0.07  | -0.32 | 0.37  | -0.41 | -0.08 | 0.47  | 7.43 |
| SF            | 0.21  | 0.46  | -0.17 | 0.49  | -0.34 | -0.14 | 0.23  | 7.37 |
| 범죄            | 0.01  | 0.39  | -0.29 | 0.02  | -0.07 | 0.02  | 0.22  | 8.13 |
| 모험            | -0.05 | -0.13 | -0.03 | 0.50  | -0.39 | -0.30 | 0.36  | 7.36 |
| 가족            | -0.86 | -0.93 | 0.06  | -0.07 | 0.36  | -0.54 | 0.36  | 8.76 |

[0086] 표 10의 결과와 같이 각 장르별 감정의 평균 Z-점수는 ‘분노’, ‘혼란’, ‘피로감’, ‘긴장감’은 공포 장르에서, ‘우울’은 드라마 장르에서, ‘친근감’은 멜로/애정/로맨스 장르에서, ‘생동감’은 코미디 장르에서 가장 높은 값을 보였다. 표 10의 결과를 도 4에와 같이 방사형 그래프로 도식화하였다. 이 그래프를 통해 장르가 공포, 멜로/애정/로맨스, 코미디, 범죄인 경우 하나의 감정으로 장르의 특징이 나타남을 알 수 있다. 장르의 특징을 설명하는 감정은 각각 공포에서 ‘긴장감’, 멜로/애정/로맨스에서 ‘친근감’, 코미디에서 ‘생동감’, 범죄에서 ‘혼란’인 것으로 나타났다. 반면에 나머지 장르의 경우는 여러 감정의 혼합된 형태로 나타났다. 즉, 가족 장르의 경우는 ‘생동감’, ‘친근감’, ‘우울’을 특징으로 하는 장르인 것으로 나타났고, 드라마장르의 경우는 ‘친근감’과 ‘우울’ 감정을 특징으로 갖는 것을 확인할 수 있다.

[0087] 콘텐츠 추천기(300)는 상술한 바와 같이 영화별로 감정에 따른 Z-점수를 저장한다. 그리고, 사용자가 영화평을 남길 경우, 영화평을 분석하여 감정에 따른 Z-점수를 산출한다. 산출된 Z-점수에 대응하는 영화를 추천한다. 이때, 감정 유사도 즉, 점수 유의도를 통해 감정 분포가 비슷한 영화를 추천한다.

[0088] 물론, 영화 추천 방식은 이에 한정되지 않고, 다양한 변형이 가능하다. 예를 들어, 사용자에게 각 감정과 Z-점수를 선택할 수 있게 한다. 이후, 사용자가 선택한, 감정과 Z-점수에 대응하는 영화 리스트를 사용자에게 제공할 수도 있다.

[0089] 그리고, 상술한 동작의 구현을 위해 별도의 웹 서버와 사용자 관리를 위한 관리서버 그리고, 사용자 접속을 위한 접속 단말기 또한 관리를 위한 관리 서버를 구비할 수 있다. 앞서 설명한 감정 분석 장치는 모듈 형태로 상기 관리 서버 또는 웹 서버에 마련될 수도 있다.

[0090] 즉, 사용자가 자신의 단말기를 통해 영화평을 작성하면, 작성된 영화평은 관리 서버의 감정 분석 모듈에 의해 영화평에 대한 감정이 선정된다. 그리고, 선정된 감정결과와 유사한 감정 결과를 갖는 영화를 사용자에게 추천한다. 또는 관리 서버에서 사용자 단말기에 7가지 감정을 제시하고, 사용자는 단말기를 통해 이 감정 중에 자신이 원하는 감정을 선택하는 경우, 관리 서버는 해당 감정에 대응하는 영화를 사용자에게 추천한다. 이때, 최근 개봉 영화를 우선으로 추천할 수 있고, 정확도가 가장 높은 순으로 추천할 수도 있다. 물론, 이외의 다양한 방법과 구조가 가능하다.

[0091] 하기에서는 상술한 감정 분석 방법과 이를 이용한 영화 추천 방법을 설명한다.

[0092] 도 5는 본 발명의 일 실시예에 따른 감정 분석 방법 및 영화 추천 방법을 설명하기 위한 흐름도이다.

[0093] 도 5에 도시된 바와 같이 먼저, 문서 수집기를 통해 다수의 문서를 수집하고, 문서 중에서 한글 문서만을 추출한다(S110). 즉, 한글 문서만을 수집한다.

[0094] 본 실시예에서는 트위터 API를 이용하여 트위터 문서중 한글로 작성된 문서를 1차로 수집하였다.

[0095] 이어서, 한글 문서 중에서 감정 형태소가 서술어로 사용된 문서만을 선별한다(S120).

[0096] 여기서, 감정 형태소는 감정이 결정된 형태소로 수집된 트위터 문서 중에서 출현횟수가 최소 100회 이상의 형태소와 유의어 사전에 의해 구분 가능한 형태소를 후보 감정 형태소로 지칭하였다. 후보 감정 형태소를 분석하여 총 322개의 감정형태소를 확보하였고, 이를 7개의 감정별로 각기 분류하였다.

[0097] 이와 같은 감정 형태소가 문장에서 서술어로 사용된 경우, 그 문장의 감정을 감정형태소가 속한 감정으로 결정하였다. 또한, 다수의 감정 형태소가 하나의 트윗에 사용된 경우, 각 감정형태소가 모두 동일 감정으로 분류된 경우에만 문서를 선별하였다. 즉, 다수의 감정형태소가 서로 다른 감정에 속한 경우에는 문서 선별에서 제외하였다.

[0098] 이어서, 선별된 문서 중에서 단일 감정의 형태소가 사용된 문서만을 추출하여 학습데이터를 생성하고 학습을 실시한다(S130).

[0099] 이때, 학습을 위해 각 트윗을 형태소 단위로 분리하였고, 형태소 선별 기준을 특정 감정에서 형태소 출현 비율인 점유율을 이용하였다. 이때 점유율은 앞서 언급한 수학적 식 1에 따라 산출하였다. 본 실시예에서는 점유율을 0.2% 부터 1.6%까지 0.2% 단위로 8단계로 나누었다. 본 실시예에서는 다항 네이브 베이즈 모델을 사용한다.

[0100] 이어서, 온라인 상으로 각 영화별 영화평들을 수집한다(S140).

[0101] 이어서, 학습된 모델을 바탕으로 수집된 영화평에 나타난 감정 비율을 각 영화별로 산출한다(S150).

[0102] 즉, 앞서 언급한 다항 네이브 베이즈 모델을 통해 분류된 영화평을 각각의 영화별로 감정이 차지하는 비율을 계산한다.

[0103] 이어서, 산출된 감정 비율을 Z-점수를 통해 정규화하여 영화별로 7가지 감정 분류에 따라 분류한다(S160).

[0104] 이를 통해 해당 영화에 대하여 7가지 감정 분류에 따른 결과값을 얻을 수 있다. 이와 같은 결과값은 별도의 저장부에 저장한다.

[0105] 이어서, 사용자 요청에 따른 감정에 해당하는 영화 리스트를 제공한다(S170).

[0106] 사용자가 작성한 영화평에 해당하는 감정과 동일한 감정 분류를 갖는 영화를 제공하거나, 사용자가 요구하는 감정 분류에 해당하는 영화를 제공할 수 있다.

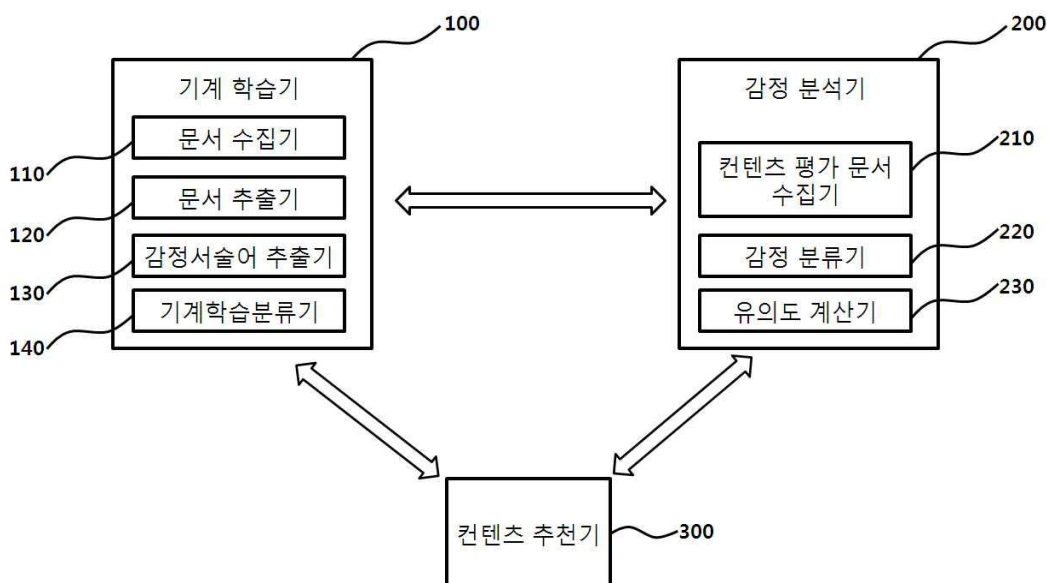
[0107] 상기에서 설명한 본 발명의 기술적 사상은 바람직한 실시예에서 구체적으로 기술되었으나, 상기한 실시예는 그 설명을 위한 것이며 그 제한을 위한 것이 아님을 주의하여야 한다. 또한, 본 발명은 본 발명의 기술 분야의 통상의 전문가라면 본 발명의 기술적 사상의 범위 내에서 다양한 실시예가 가능함을 이해할 수 있을 것이다.

## 부호의 설명

|        |                     |                 |
|--------|---------------------|-----------------|
| [0108] | 100 : 기계 학습기        | 110 : 문서 수집기    |
|        | 120 : 문서 추출기        | 130 : 감정서술어 추출기 |
|        | 140 : 기계학습분류기       | 200 : 감정 분석기    |
|        | 210 : 콘텐츠 평가 문서 수집기 | 220 : 감정 분류기    |
|        | 230 : 유의도 계산기       | 300 : 콘텐츠 추천기   |

## 도면

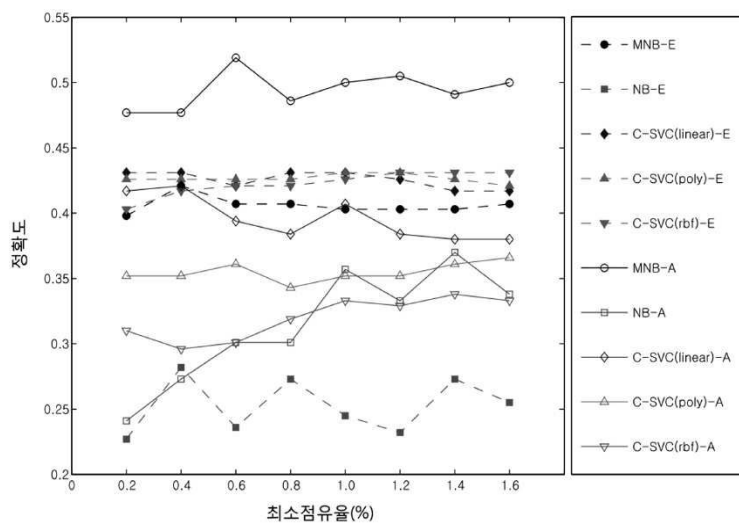
도면1



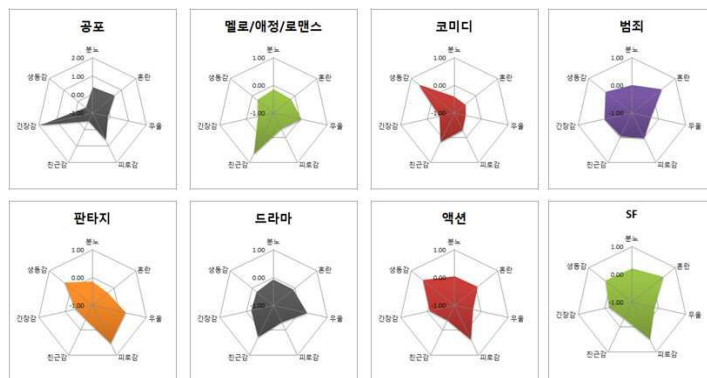
도면2



도면3



도면4





도면5

