

# 영화 시나리오와 영화촬영기법을 이용한 감정 예측 시스템

김진수  
안양대학교 아리교양대학 교수

## Emotion Prediction System using Movie Script and Cinematography

Jinsu Kim  
College of Ari Liberal Arts, Anyang University, Professor

요 약 최근에 다양한 정보로부터 감정을 예측하여 청중에게 감독이 알고자 하는 정보를 빠르게 전달하고자 한다. 또한, 청중은 감독의 의도를 대화 내용에 나타나는 대사뿐만 아니라, 영상내의 다양한 정보인 촬영 기법, 장면의 배경, 배경 음악 등을 통해 비대사 구간에서도 감정의 흐름을 이해하려고 한다.

본 논문에서는 대사와 같은 문맥의 상황뿐만 아니라, 촬영 영상에 담아낸 색상, 음향, 구도, 배치 등에 의해 표현된 정보를 혼합하여 감정을 추출하고자 한다. 즉, 다양한 감정 표현 기법을 대사 구간, 비대사 구간으로 나누어 학습하고 판별하여 영상의 완성도에 기여하고 새로운 변화에 빠르게 적용할 수 있는 감정 예측 시스템을 제안한다. 본 논문에서 제안한 감정 예측 시스템이 변형된 n-gram 방식과 형태소 분석을 적용한 사례와 비교했을 때, 정확도는 약 5.1%, 0.4% 향상되었고, 재현율은 약 4.3%, 1.6% 향상되었다.

주제어 : 감정 예측, 빅데이터, 영화촬영기법, 지지벡터기계, 형태소 분석

**Abstract** Recently, we are trying to predict the emotion from various information and to convey the emotion information that the supervisor wants to inform the audience. In addition, audiences intend to understand the flow of emotions through various information of non-dialogue parts, such as cinematography, scene background, background sound and so on.

In this paper, we propose to extract emotions by mixing not only the context of scripts but also the cinematography information such as color, background sound, composition, arrangement and so on. In other words, we propose an emotional prediction system that learns and distinguishes various emotional expression techniques into dialogue and non-dialogue regions, contributes to the completeness of the movie, and quickly applies them to new changes. The precision of the proposed system is improved by about 5.1% and 0.4%, and the recall is improved by about 4.3% and 1.6%, respectively, when compared with the modified n-gram and morphological analysis.

**Key Words** : Emotion Prediction, Big Data, Cinematography, Support Vector Machine, Morphological Analysis

### 1. 서론

정보통신기술(ICT)의 융합 및 발전에 힘입어 사용되는 원하는 정보를 장소와 시간의 제약에서 벗어나 손쉽게 수집, 가공, 그리고 배포할 수 있게 되었다. 특히 영화,

드라마 등과 같은 콘텐츠 속에 표현된 등장인물의 감정 변화와 전체적인 서사적인 흐름에 따른 감정 흐름을 통해 자신의 취향과 연관된 다양한 콘텐츠를 선택하고 추천 받기를 원한다[1]. 감정 표현은 작가가 청중에게 알리고자 하는 객관적 감정과 청중이 받아들이는 감정은 대

\*Corresponding Author : Jin-Su Kim (kjspace@anyang.ac.kr)

Received October 11, 2018

Accepted December 20, 2018

Revised December 4, 2018

Published December 28, 2018

화 내용에 나타나는 대사뿐만 아니라, 영상에 나타난 다양한 정보인 촬영 기법, 장면의 배경, 배경 음악 등을 통해 비대사 구간에서도 감정의 흐름을 이해하려고 한다. 대사 구간에서는 대사 속에 포함된 어휘를 통해 감정을 쉽게 판별할 수 있지만, 촬영기법이나 화면 배경 등과 같은 부분 속 영상에 나타난 영상 및 음향 정보를 통해 감정을 보다 효과적으로 표현할 수 있다. 즉, 전체적인 감독의 촬영기법에 따라 상이한 감정을 표현하는 경우도 적지 않다. 이러한 대사, 비대사 구간 및 전체적인 영상 표현 기법을 통해 감정을 추출하여 보다 영상물에 대한 이해도를 높이고자 한다.

본 논문에서는 대사와 같은 문맥의 상황뿐만 아니라 촬영 영상에 담아낸 색, 구도, 배치 등에 의해 표현된 정보를 혼합하여 감정을 추출하고자 한다. 이러한 다양한 감정 표현 기법을 대사 구간, 비대사 구간으로 나누어 학습하고 판별하여 영상의 완성도에 기여하고 새로운 변화에 빠르게 적용할 수 있도록 하고자 한다.

## 2. 관련연구

### 2.1 감정 온톨로지

감정을 표현하는 방법은 국가, 종교, 문화 등의 다양한 요인에 따라 다르게 느껴지는 특성을 가지고 있지만, Ekman[2]이 제시한 6가지 기본 감정은 남녀노소, 지역, 문화 등에 상관없이 누구나 일반적으로 공감할 수 있는 기본 감정으로 Anger, Disgust, Fear, Happiness, Sadness, Surprise이다. 특히, 대화에 표현되는 감정은 텍스트뿐만 아니라 장면의 상황까지 고려하여 예측해야 한다. 즉, 카메라 움직임이나 주변 환경 등과 같은 상황정보가 대사 텍스트와 무관한 감정을 느끼게 할 수 있기 때문에, 특정 대사에 의한 문맥적인 내용뿐만 아니라 주변 상황 정보를 혼합하여 감정을 예측해야 한다.

감정 온톨로지(Emotion Ontology)는 기본 감정과 다양한 감정을 유발하는 단어들 간의 연관성뿐만 아니라 촬영기법과 장면의 상황 정보 정보를 기반으로 구축하며 Web Ontology Language인 OWL[3]을 사용하여 감정을 유발하는 단어들의 개념, 관계 등을 정의한다. 감정을 예측하기 위해 각 감정 단어들과 연관 감정 단어들뿐만 아니라 부정어나 은유까지 처리해야 한다.

### 2.2 형태소 분석

색인어 추출을 위한 형태소 분석(Morphological Analysis)은 해당하는 대상의 어절을 최소의 의미를 지닌 형태소 단위로 분석하는 것을 의미한다. 즉, 문단에 나타난 단어로부터 어근, 접두사, 접미사, 품사 등 다양한 언어적 속성을 파악하고 이를 이용하여 형태소를 찾아내거나 처리하는 작업이다.

정보처리를 위한 다양한 오픈 소스 형태소 분석기인 KoNLPy(Korean NLP in Python)[4], MeCab-ko[5], NLTK(Natural Language Toolkit)[6] 등이 개발되었고, 이 중 교육용으로 개발된 자연어 처리 및 문서 분석용 파이프라인 패키지인 NLTK패키지를 적용하고자 한다.

### 2.3 지지벡터머신과 데이터마이닝

일반화 능력이 뛰어나며 지도학습(Supervised Learning) 모델인 지지벡터머신(SVM, Support Vector Machine)은 분류와 회귀 분석 등에 높은 성능을 가진 대표적인 기계 학습 기법이다. SVM은 다양한 커널 함수를 사용하여 입력공간에서는 분류되기 힘든 비선형문제를 특징공간이라는 고차원 공간으로 사상시킴으로써 새로운 공간에서 최대의 마진을 가진 비선형 데이터들의 선형 판별을 수행하며[7], 선형 SVM의 수식은 다음과 같다.

$$\exists \omega, b \text{ 일 때 } \begin{cases} w \cdot x_i + b > 0, & y_i = +1 \\ w \cdot x_i + b < 0 & y_i = -1 \end{cases} \quad (1)$$

$w$ 는 가중치벡터,  $x$ 는 입력벡터,  $b$ 는 기준치이며,  $w$ 와  $b$ 는 학습 데이터로부터 학습되며, 입력데이터  $x_i$ 가 범주에 속하면  $y_i$ 는 +1의 값을 가지며, 속하지 않는다면 -1의 값을 갖게 되어 최종적으로 SVM은 최적의  $w$ 와  $b$ 를 찾는 문제이다.

SVM 학습에 필요한 방대한 자료를 위해 대량의 정형 또는 비정형의 데이터 집합으로부터 가치를 추출하고 결과를 분석하는 기술로 많이 사용되던 데이터 마이닝 기법 중 연관규칙, 순차패턴, 분류, 클러스터링 등을 통해 대량의 실제 데이터로부터 숨겨져 있어 잘 드러나지 않는 유용한 정보들을 추출하고자 한다[8].

### 2.4 영상 감정 표현 요소

영상 속에 적용된 다양한 감정은 감독과 촬영감독의 의도에 의한 영화촬영기법을 통해서도 표현할 수 있다. 영상 속에 촬영 속도가 빠른 고속촬영은 감정 상승 기대

나 세밀한 관찰을 의미하며, 저속촬영은 시간의 압축효과, 빠른 관찰 효과, 감정교란 및 착란효과 등과 같은 의미를 부여한다.

프레임 내에서 활동적인 움직임의 방향에 따라 샷의 역동성이 표현되고 강조되며 다양한 샷은 다음과 같다.

익스트림 롱 샷은 주인공이 갖는 감정의 극단을 역으로 표현하기 위한 도치의 수단임과 동시에 확장된 공간감을 통해 또 다른 영상의 이미지를 구현하거나 암시적 예고를 하는 효과를 줄 수 있다. 하이 앵글 샷은 높은 곳에서 내려다보는 위치에서 촬영함으로써, 조망권 확보, 전지적·객관적 시점, 대상 약화, 심리적 무력감, 박탈감을 강화하는 이미지를 표현할 수 있다. 로우 앵글 샷은 낮은 곳에서 올려다봄으로써 대상의 존재감과 위상을 높이는 이미지, 존엄, 위압감을 줄 수도 있다. 볼 업/다운 샷은 카메라가 고정해 있는 특정 대상을 훑어 올라가거나 내려오는 방법으로, 대상의 면모를 살살이 훑어보게 하거나 시간과 감정의 딜레이 효과를 노리고자 할 때 주로 사용하며 특정 대상의 순간 금금증을 유도했다 풀어 주는 극적 효과를 노릴 수 있다. 달리 인/아웃은 카메라가 고정해 있는 대상을 향해 추적해 들어가거나 그 대상으로부터 빠져나오는 이동효과를 주기 위해 사용하며, 카메라가 특정 대상에게 시선을 몰아감으로써 보는 이의 카메라 시선으로 특정 대상에 대한 감정 유입을 같이 하는 효과를 노리거나 긴장 고조, 이완의 효과를 노릴 때 흔히 사용한다. 액션이나 공포물의 경우, 그 속도를 느리게 하면 그만큼 긴장감이 고조되어 살어나고, 반대로 속도를 빨리하면 긴장감보다는 충격효과를 그만큼 높일 수 있다. 리버스 샷은 대상을 정면에 두고 찍는 것이 아니라, 등 뒤에서 찍는 개념으로 대상을 은밀하게 감추는 효과와 대상들 앞에 펼쳐져 있는 자연이나 공간감을 강조하고자 할 때, 단조로움을 피해 입체적 시각효과를 노릴 때 사용하며, 정적이거나 동적인 2샷에서 주로 사용되고 은밀함, 다정함, 은근함 등의 관계성을 강조한다[9-11].

매 컷마다 인물의 감정을 어떻게 소화하고 전체적인 영상의 분위기를 이끌어갈 것인지는 촬영자의 감각과 역량에 달려 있기 때문에, 이러한 중요성을 간과해서도 안 된다.

모든 드라마와 영화마다 영상의 분위기는 물론 미장센이나 이미지 효과들이 달라지는 이유는 연출자와 촬영자의 호흡을 어떻게 맞추느냐에 따라 달라지기 때문에 다각적인 기법을 통해 다양하게 해석해야 한다.

### 3. 감정 예측 및 추출 시스템

본 논문에서 제안한 감정 예측 시스템의 전체적인 구조는 Fig. 1과 같다. 첫 번째 단계에서는 전처리된 감정 문단들로부터 형태소 분석기(Morphological Analyzer, MORPH)를 통해 후보 키워드를 추출하고, 각 추출된 키워드들과 키워드들 사이의 가중치를 이용하여 대표키워드를 선별하여 감정별 연관-순차 DB를 생성한다. 두 번째 부분에서는 생성된 감정 데이터베이스와 SVM을 이용하여 감정별로 훈련 문단들을 훈련시킨 후, 마지막 부분에서 테스트 문단과 이러한 문단에 상응하는 상황 정보들을 입력으로 하여 표현된 감정을 SVM을 통해 예측한다.

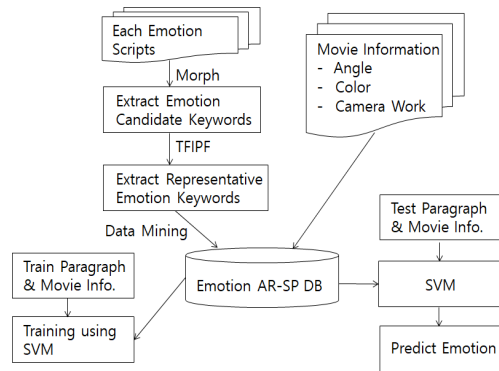


Fig. 1. Emotion Extract System Architecture

#### 3.1 감정 키워드 및 연관 DB 구축

감정 예측을 위한 기본적인 감정 키워드를 추출하기 위해 형태소 분석기를 사용하였다. 영화 시나리오에 나타난 대화나 상황 기술은 맞춤법, 띄어쓰기 등이 비교적 양호하기 때문에 형태소 분석기를 이용하여 감정에 영향을 많이 주는 동사, 명사, 그리고 부정어 등을 추출하고 감정 예측을 위한 후보 감정 키워드들을 구축하였다. 이때, 감정 키워드 추출을 위한 형태소 분석기로 NLTK(Natural Language Toolkit)를 사용한다.

Ekman이 제시한 6가지 기본 감정과 6가지 감정 편별의 최소 한계값보다 적을 경우, 중립 감정(Neutral)으로 표현하여 7가지 감정으로 분류한다.

감정 집합의 훈련 문장들을 이용하여 각 감정집합의 후보 키워드를 각각 추출한다. 이때, 감정 대표 키워드를 추출하기 위해 단어 빈도와 역문서 빈도의 곱을 가중치로 사용하는 TF-IDF(Term Frequency - Inverse

Document Frequency)[12]를 변형하여 TF-IPF(Term Frequency - Inverse Paragraph Frequency)를 사용한다 [13]. 어휘 척도인 단어 빈도와 역문단 빈도수의 곱인 TF-IPF의 수식은 다음과 같다.

$$IPF_k = \log_2 \frac{N}{PF_k} + 1 \quad (2)$$

$$TF_k \cdot IPF_k \\ = TF_k \cdot [\log_2(N) - \log_2(PF) + 1] \quad (3)$$

이 때, TF-IPF의 문단(paragraph)은 장면을 설명하는 상황 묘사 부분, 각 장면에서 캐릭터들 간의 대화에서 발생하는 일련의 대사들을 하나의 문단으로 설정하고 실험 집합에 필요한 문단들을 형태소 분석을 통해 가중치가 높은 감정 후보 키워드들을 생성한다.

TF-IPF를 통해 추출한 키워드 후보들 집합으로부터 각 감정별 키워드를 선별하여 상위 N개의 감정 대표 키워드들의 집합을 생성하고, 감정 대표 키워드들 간의 연관도와 순차 패턴의 데이터 마이닝 기법을 적용하기 위해 각 감정 키워드별 연관-순차 패턴 정보인 연관 DB를 생성한다.

Table.1에서는 기본 감정 표현에 많은 영향을 미치는 대표 후보 키워드들의 일부를 보여준다.

Table 1. Extracted Emotion Keywords

Emotion	Emotion Keywords
Anger	displeased, ireful, blow-up, pissed, rageful, outrage, deception, traitor, hatred, ...
Disgust	icky, fulsome, rancid, loath foul, secretion, stench, filth, ...
Fear	allfired, awful, awesome, appalling, confinement, danger, terror, horror, explosive, afraid, scared, frighten, ...
Joy	happy, luck, blessing, delight, amusing, beamy, carefree, cheerful, ...
Sadness	sorrowful, tearful, dolorous, grieved, heavy hearted, lachrymose, pain, lone, funeral, ...
Surprise	amazed, alarmed, frightened, large eyed, startled, scream, fainting, breaking news, sudden stop, ...

### 3.2 대사 패턴 기반의 감정 생성

등장인물의 평온함, 기쁨, 화남, 슬픔, 놀람, 공포 등과 같은 다양한 감정에 대한 표준화 데이터를 구축하기 위해 시나리오에 나타난 대량의 대사를 기반으로 감정을 예측하는 대사 패턴 기반의 감정 생성 모델을 구축한다. 감정은 자극에 따른 반응을 통해 인물간의 관계의 변화

를 감지하며 이러한 관계를 통해 표현되는 감정의 순차적인 변화를 통해 패턴을 찾아내는 모델이 도출되어야 한다. 이러한 감정의 변화를 대사를 기반으로 1차적으로 감정을 컴퓨터가 해석할 수 있도록 표현하는 것이 대사 기반 감정 생성 모델이며 감정 키워드와 연관DB 구축을 통해 수행한다.

### 3.3 비대사 패턴 기반 감정 생성

시나리오에 나타난 등장인물들 간의 대화 내용뿐만 아니라, 플롯에 나타난 비대사 구간 혹은 영상에 나타난 물건, 모양, 음향 등과 같은 다양한 상황 정보를 가지고도 감정을 추출하는데 참고할 수 있다. 즉, 영상은 침묵의 언어라고 하듯이 영상에 표현된 색상, 샷 형태, 각도, 이미지의 조명과 톤 등도 감정을 추출하는 하나의 요인이 될 수 있다.

감독의 계획에 따라 색상에 통해 감정의 일부를 보여 줄 수 있다. 예를 들어, 빨강은 열정, 흥분, 타오르는 생명의 상징, 과격, 경고, 주의 등을 표현하고, 주황은 빨강에 비해 안정적, 열정감, 고귀함, 노랑보다 더 따뜻함, 황홀감, 여성의 색, 성숙의 극치, 사치와 변화를 꿈꾸는 상징의 색이다. 노랑은 빛에 가까운 색으로, 밝음, 따뜻함, 순수, 보호 본능, 상상력 등을 수반하며, 초록은 자연의 색으로 감정의 안식, 성장의 완속미 등의 안정을 준다. 파랑은 빨강과 대척점의 색으로 이중성, 평화, 광활함, 희망, 다른 한편으로는 절망과 고독, 우울의 색이다.

비대사적 패턴에 배경 음향을 가지고도 감정을 전개하거나 고조시키는 요인이 존재한다. 이러한 배경 음향으로부터 감정의 흐름을 표현하기 위한 패턴을 추출하여 감정 추출의 한 인자로 사용하여 강화하고자 한다. 배경 음향과 감정들 간의 상관관계를 수집하고 장면들 사이의 미묘한 감정의 변화되는 요인을 찾아 배경 음향 패턴 기반의 감정 생성 모델을 구축한다.

## 4. 실험 및 분석

감정 훈련 문서를 수집하기 위해 IMSDB(The Internet Movie Script Database)[14]라는 웹 사이트에 존재하는 1194개의 영화 스크립트 중 100개 스크립트를 선별하여 텍스트파일로 변환 및 정제과정을 통해 XML 파일로 변환하였다. 이러한 감정 훈련 문서들로부터 추출된 감정-순차 DB와 대표 키워드들을 이용하여 SVM에 훈련한다.



본 논문에서 제안한 카메라 촬영 기법, 색상, 음향 정보 등의 상황정보를 추가하여 텍스트에 나타나지 않은 감정 예측에 영향을 주었다.

## 5. 결론 및 향후 연구

본 논문에서 제안한 시스템은 대사구간의 각 감정 문단으로부터 형태소 분석을 통해 대표 키워드들을 추출하고, 비대사구간의 다양한 정보를 취합하여 감정을 예측하기 위해 대표 키워드의 가중치와 데이터 마이닝 기법을 혼합하여 변형된 n-gram방법, 형태소분석만을 사용한 방법에 비해 정밀도는 각각 약 5.1%, 0.4% 향상되었고, 재현율은 약 4.3%, 1.6% 정도 향상되었다.

향후 연구에서는 대사구간 및 비대사구간 외에 다양한 정보를 혼합한 더욱 효율적인 기법을 통해 감정 예측 성능을 높일 수 있도록 연구를 해야 하며, 훈련된 키워드들 사이의 상호 위치에 따른 연관성까지도 고려하여 가중치를 부여하고 감정을 예측해야 할 것이다.

## REFERENCES

- [1] O S. Kim & S. W. Lee (2015). A Movie Recommendation Method based on Emotion Ontology, *Journal of Korea Multimedia Society* 18(9), 1068-1082.
- [2] Scherer, Klaus R & Paul Ekman. (2014). *Approaches to Emotion*, Psychology Press.
- [3] D.L. McGuinness & F.V. Harmelen. (2004). *OWL Web Ontology Language Overview*, W3C Recommendation.
- [4] *KoNLPy*, <http://konlpy.org>
- [5] *mecab-ko*, <https://bitbucket.org/eunjeon/mecab-ko>
- [6] *NLTK*, <https://www.nltk.org/>
- [7] Bernhard E Boser, Isabelle M Guyon, & Vladimir N Vapnik. (1992). A Training Algorithm for Optimal Margin Classifiers, *Proc. The Fifth Annual Workshop on Computational Learning Theory*, 144-152.
- [8] J. W. Han. (2011). *Data Mining: Concepts and Techniques*, Morgan Kaufmann.
- [9] David Stump. (2014). *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows*, Routledge.
- [10] Blain Brown. (2016). *Cinematography: Theory and Practice: Image Making for Cinematographers and Directors*, Routledge.
- [11] Harry Box. (2013). *Set Lighting Technician's Handbook:*

*Film Lighting Equipment, Practice, and Electrical Distribution*, Routledge.

- [12] Manning, C. D, Raghavan, P & Schutze, H. (2009), *Introduction to Information Retrieval*, Cambridge University Press.
- [13] J. S. Kim. (2014). Emotion Prediction of Document using Paragraph Analysis, *Journal of Digital Convergence*, 12(12), 249-255.  
DOI : 10.14400/JDC.2014.12.12.249
- [14] *IMSDb*, <http://www.imsdb.com>
- [15] R Baeza-Yates & B Ribeiro-Neto. (2004). *Modern Information Retrieval: The Concepts and Technology behind Search*, ACM Press New York.

김진수(Kim, Jinsu)

[정회원]



- 1998년 2월 : 인천대학교 전자계산공학과 (공학사)
- 2001년 8월 : 인하대학교 컴퓨터공학과 (공학석사)
- 2010년 2월 : 인하대학교 정보공학과 (공학박사)
- 2011년 3월 ~ 현재 : 안양대학교 아리교양대학 조교수
- 관심분야 : 데이터마이닝, 정보검색, 융복합, WSN, 빅데이터
- E-Mail : kjspace@anyang.ac.kr