

## 감정어 추출을 통한 등장인물 성향 가시화 연구

백영태<sup>0</sup>, 박승보<sup>\*</sup>

<sup>0</sup>김포대학교 멀티미디어과

<sup>\*</sup>단국대학교 단국-삼성 모바일연구소

e-mail: hannaek@kimpo.ac.kr<sup>0</sup>, molaal@naver.com<sup>\*</sup>

## Visualization Study of Character Type by Emotion Word Extraction

Yeong Tae Baek<sup>0</sup>, Seung-Bo Park<sup>\*</sup>

<sup>0</sup>Dept. of Multimedia, Kimpo College

<sup>\*</sup>Dankook Samsung Mobile Institute, Dankook University

### ● 요약 ●

본 논문에서는 영화의 등장인물의 성향을 파악하기 위해 시나리오의 대사로부터 감정어를 추출하고, 등장인물의 감정어들을 긍정, 부정, 중립의 3개로 단순화하여 등장인물의 성향을 가시화 시켜주는 방법을 제안한다. 대사로부터 감정어를 추출하기 위해 WordNet 기반의 감정어 추출 방법을 제안한다. WordNet은 단어 간에 상위어와 하위어, 유사어 등의 관계로 연결된 네트워크 구조의 사전이다. 이 네트워크 구조에서 최상위의 감정 항목과의 거리를 계산하여 단어별 감정량을 계산하여 대사를 30 차원의 감정 벡터로 표현한다. 등장인물별로 추출된 감정 벡터를 긍정, 부정, 중립의 3개의 차원으로 단순화 하여 등장인물의 성향을 표현한다.

키워드: 감정어(emotion word), 등장인물(character), 영화(movie), 가시화(visualization)

### I. 서론

영화나 드라마와 같은 스토리 기반의 저작물에서 등장인물은 스토리를 진행하는 핵심 주체이다. 또한 등장인물이 표출하는 감정은 스토리의 갈등이나 심리묘사를 하는 중요한 도구이다. 등장인물의 감정 정보는 등장인물 간의 관계를 묘사하거나 등장인물의 긴장상태 등을 표현하는 정보로써 스토리 기반의 저작물을 분석하는데 매우 유용한 정보이다. 등장인물은 대화를 통해 감정과 관련된 정보를 표출한다. 대화 속의 감정어나 표정변화와 어투와 같은 다양한 정보를 통해 등장인물의 심리 상태나 감정변화가 표출되게 된다. 하지만 표정변화나 어투와 같은 요소들로부터 감정을 파악하기에는 기술적으로 해결해야 할 여러 가지 문제점이 존재한다. 이미지 프로세싱이나 음성 인식과 같은 비정형 데이터로부터 원하는 감정정보를 추출하는 작업이 선행되어야 한다. 원하는 객체나 음성을 추출하고 그로부터 감정요소를 파악하는 것은 기술적 복잡성뿐만 아니라 낮은 정확도가 문제가 된다[1]. 하지만 이에 비해 대화 문장에서 감정어를 추출하는 것은 현실적으로 활용될 수 있는 실현 가능한 접근 방법이다[2]. 대화 속에 존재하는 감정어들을 추출하여 등장인물이 드러내는 감정 상태를 파악하여 스토리 정보로 활용할 수 있을 것이다. 특히 감정어를 이용하여 등장인물의 성향을 파악할 경우 스토리의 갈등 구조를 파악하는 데 많은 도움이

될 것이다. 이를 통해 영화 축약이나 브라우징과 같은 영화 정보 검색에 등장인물 정보를 활용할 수 있을 것이다.

본 논문에서는 영화 시나리오로부터 등장인물의 대화에 존재하는 감정어를 추출하는 방법과 추출된 감정어를 기반으로 등장인물의 성향을 분석하는 방법을 제안한다.

### II. 감정어 기반의 등장인물 성향 분석

#### 1. WordNet 기반의 감정어 추출 방법

WordNet은 단어들을 상위어(hyponym), 하위어(hyponym), 유사어(synonym) 등의 관계로 연결되어진 사전이다[3]. 감정과 관련된 최상위의 용어는 'emotional state', 'emotion', and 'feeling'의 3개이다. 3개의 용어의 하위어들을 추출한 후, 유사한 용어들을 그룹으로 모아서 표 1과 같이 총 30개의 감정항목을 추출하였다. 이 30개의 감정 항목은 감정어가 가질 수 있는 독립적인 감정 차원이므로 감정어는 식 1처럼 30차원의 벡터로 표현할 수 있다.

표 1. 묶어진 30개의 감정항목

Table 1 Emotion categories to be grouped

happiness	liking	dislike	anxiety
pride	expectation	unhappiness	ecstasy
gratitude	humility	apathy	desire
anger	ungratefulness	astonishment	glow
soul	sentiment	shame	gravity
calmness	fearlessness	humor	sympathy
ambivalence	embarrassment	pain	faintness
sensitivity	devastation		

예를 들어 joy의 경우 happiness에 synonym 관계로 연결되어 식 1과 같이 첫 번째 감정항목에만 1이라는 값을 갖는 30차원의 벡터로 표현된다.

$$EV(joy) = \{1, 0, 0, \dots, 0\} \quad (1)$$

감정어는 WordNet에서 상위어와 유사어를 검색하다보면 30개의 감정항목 중에 1개에 도달하게 된다. 그리고 감정항목에 도달하기까지의 거치는 단어수가 경로의 길이가 된다. 경로의 거리를 역수로 취한 것이 기본적인 감정어의 크기가 된다. 또한 도달한 감정항목에 크기값이 적용되어 감정벡터로 표현된다. 경로는 중간에 2~3개로 분기하여 서로 다른 감정항목에 도달할 수 있기 때문에 중간 용어가 갖는 총 의미(sense)들 개수 중에 감정항목에 연결된 의미의 개수로 나누어 주어 여러 감정항목으로 분기되는 감정어는 분기의 비율로 크기를 감소시켜준다. 이러한 개념들이 반영되어 계산되는 감정어 크기의 계산은 식 2와 같이 표현된다.

$$e_i = \frac{1}{D(w \rightarrow EC_i)} \times \frac{m}{t} \quad (2)$$

w: 단어,  $EC_i$ : i번째 감정항목

$D(w \rightarrow EC_i)$ : 단어(w)부터 감정항목( $EC_i$ )까지 거리

m: 감정항목( $EC_i$ )까지 연결된 의미(sense)의 개수

t: 단어(w)의 총 의미 개수

예를 들어 “crazy”라는 단어는 1번째 감정항목인 “happiness”와 10번째 항목인 “anger”에 연결된다. 그리고 감정항목까지의 거리는 4이고, 전체 의미의 개수는 6이고, 각 감정항목에 연결된 의미의 개수는 1이다. 따라서 감정벡터 중 1번째와 10번째인  $e_1$ 과  $e_{10}$ 의 값은 각각 1/24가 된다. 이렇게 계산된 “crazy”의 감정벡터는 식 3과 같이 표현된다.

$$EV(crazy) = \{1/24, 0, 0, 0, 0, 0, 0, 0, 0, 1/24, \dots, 0\} \quad (3)$$

## 2. 감정어 분류를 통한 등장인물 성향파악 방법

WordNet으로부터 얻어진 감정어 항목들을 표 2와 같이 긍정, 부정, 중립의 3개로 분류하였다. 이것을 등장인물의 대화에서 추출된 감정어들에 적용하여 등장인물의 성향이 긍정적인지 부정적인지 중립적인지를 분류하였다. 감정 벡터로 표현된 감정어들을 등장인물에 대하여 합산한 후에 감정 벡터의 값들을 긍정, 부정,

중립의 3가지 항목으로 분류한 후 합산하였다.

표 2. 감정항목의 분류

Table 2. Classification of Emotion Categories

분류	해당 감정항목
긍정	happiness, liking, expectation, ecstasy, gratitude, astonishment, glow, humor, sympathy
부정	dislike, anxiety, ambivalence, unhappiness, embarrassment, apathy, pain, anger, ungratefulness, shame, devastation
중립	humility, desire, faintness, soul, sentiment, gravity, sensitivity, calmness, fearlessness

이것을 그림 1과 같이 긍정, 부정, 중립의 3축의 그래프로 인물의 성향을 표현하였다.

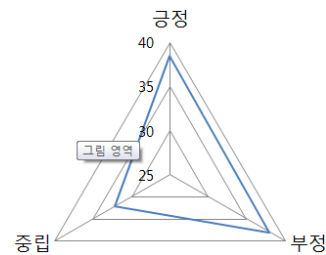


그림 1. 등장인물의 성향 그래프  
Fig. 1. Graph of Character Type

## IV. 결론

등장인물들의 성향을 표현하기 위하여 대사로부터 감정어를 추출하였고, 이 감정어를 긍정적, 부정적, 중립적인 3개로 분류하여 인물의 성향을 가시화 하였다. 비교적 용이하게 감정상태를 추출할 수 있는 WordNet을 이용한 텍스트 기반의 감정어 추출 방법을 제안하였으며, 찾아진 감정벡터를 3개의 축으로 단순화하여 가시적으로 보여줄 수 있는 방법론을 제시하였다.

향후에 감정어 추출과 등장인물 성향 가시화 방법을 이용하여 다수의 영화에서 나타나는 등장인물들의 성향을 가시화 시키는 시스템에 대한 연구를 진행하겠다.

## 참고문헌

- [1] K.-E. Ko, and K.-B. Sim, "Development of Context Awareness Service Inference Method using Multi-Modal Emotion Recognition System," Autumn Conference on Korea Institute of Intelligent System, Vol. 18, No. 2, pp. 261-264, Oct. 2008.
- [2] S.-B. Park, E. You, and J.J. Jung, "Potential Emotion Word in Movie Dialog," Proceedings of the International Conference on IT Convergence and Security 2011, pp. 507-516, Dec. 2011.
- [3] G.A. Miller, "WordNet: A Lexical Database for English," Communications of the ACM, Vol. 38, No. 11, pp. 39-41, 1995.