

인공지능기초응용 II

7 주차 과제

인공지능응용
K2025029 금동환

내러티브 아크 기반 5 단계 분할을 통한 《홍길동전》 텍스트 형태소 및 감정선 분석

금동환
국민대학교 소프트웨어융합대학원
e-mail: dhkeum9886@gmail.com

Morphological and Sentiment Analysis of 《The Story of Hong Gildong》 Text via a Five-Stage Narrative Arc Division

Donghwan Keum
Kookmin University

요 약

《홍길동전》(경판 24 장본) 원문[1]을 내러티브 아크의 5 단계(발단-상승-위기-절정-하강)로 분할한 뒤, klt2023 형태소 분석기[2]로 추출한 명사에 465-EmoNouns_KcBERT 감정 사전[3]을 매핑하여 단계별·전체 감정 분포를 정량화하였다. 총 151 개 감정 명사 중 슬픔 39 회(25.8 %), 분노 26 회(17.2 %), 기쁨 17 회(11.3 %)가 3 대 정서를 형성했다. 슬픔·분노가 서사 전반의 축을 이루고, 기쁨은 초반에 소폭 출현했다가 절정기(의적 활동 시작)에서 급등 후 최종 하강 단계에서 다시 감소하여 서사의 카타르시스 구조와 일치함을 보여준다. 이러한 계량 지표는 《홍길동전》의 정서 흐름을 시각화하며, 고전 서사 연구에 감정-언어 분석이 유효함을 입증한다.

1. 서론

조선 후기 작가미상의 영웅소설 《홍길동전》은 신분제 모순을 고발하고 의적의 영웅담을 제시한다. 기존 연구가 주로 주제·인물·신분제 담론에 집중했다면, 본 연구는 텍스트 내 감정선을 정량화해 서사 구조와 정서 변화를 연계한다. 내러티브 아크의 단계 모델은 고전 서사의 구조를 파악하는 데 적합하므로 이를 채택하여 단계 별 감정의 분포를 정량화 하여 분석에 활용한다.

2. 데이터 및 방법론

연구 목적으로 경판 24 장본 텍스트 전체를 UTF-8 로 정규화하여, 내러티브 아크 기법을 기반으로 5 단계(1. 발단 2. 상승 3. 위기 4. 절정 5. 하강)로 전문을 분할한다. 각 단계의 분할에 대해서는

1. 발단 : 길동의 탄생과 신분에 대한 인식
2. 상승 : 형제들 간의 불평등과, 관직 진출의 좌절
3. 위기 : 공직 생활과 부조리한 명령
4. 절정 : 출가 및 의적 집단 결정, 자신만의 정의 실현
5. 하강 : 호국 활동, 관직 제수 후 귀환

의 5 단계로 나눈 후 별도의 txt 파일로 분할하여 단계 별 분석이 가능하도록 했다.

klt2023 분석기를 사용하여 단계 별 txt 파일의 명사를 추출하여 추출된 명사와 감정을 465-EmoNouns_KcBERT 기반으로 정합되면 명사만을 1:1 로 매핑한다.

매핑된 데이터를 collections 를 통해 정렬하여 matplotlib 을 이용해 시각화 한다.

표 1. 진행 절차 요약

절차	상세
데이터 준비	《홍길동전》(경판 24 장본)를 ①발단, ②상승, ③위기, ④절정, ⑤하강 단계로 분할
형태소 분석	KoNLPy klt2023 분석기사용. 명사만 추출.
감정 매핑	465-EmoNouns_KcBERT 사전의 8 대 기본 감정(기쁨·슬픔·분노·사랑·놀람·불안·혐오·희망)에 정합되는 명사만 채택. 신뢰도는 고려 하지 않음.
정량 지표 계산	단계별 감정 빈도, 전체 빈도 대비 상대비율을 데이터화. Python + collections + matplotlib 을 통해 수행 및 시각화

이상의 절차를 통해 단계별 감정선 데이터를 구축하고, 감정선 분석을 진행하였다.

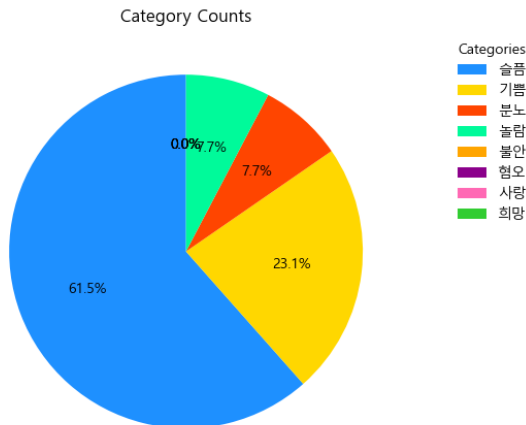
3. 내러티브 아크 단계별 감정 분석 결과

정량화된 지표는 표 2로 정리되었다.
각 단계별 데이터의 분석을 진행하였다.

3.1. 1 단계 : 발단

1 단계에서는 총 13 개의 감정 명사가 추출되었는데, 그중 8 개(61.5 %)가 슬픔 범주에 속한다. ‘마음(6 회)’과 ‘생각(2 회)’이 슬픔을 대표하며, 출생 차별로 생긴 내적 고통이 서사의 주된 부분을 형성한다. 반면 기쁨은 ‘꿈’과 ‘흥’이 3 회(23.1 %) 나타나 아직은 극의 분위기가 슬프고 분노로만 가득 차 있지 않음을 알 수 있다.

그림 1. 1 단계 감정 분포 파이차트

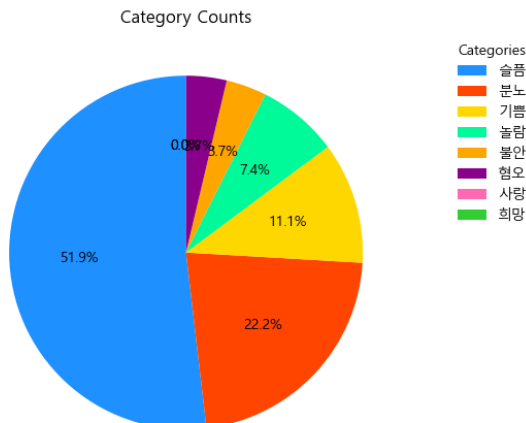


3.2. 2 단계 : 상승

2 단계로 넘어가면 감정 명사 총수는 27 개로 배 이상 늘어나고, 슬픔은 여전히 과반(14 회 · 51.9 %)을 차지하지만 분노가 6 회(22.2 %)로 급등한다. 상위 어휘 ‘한(4 회)’과 ‘원망(2 회)’이 초란의 모략과 자객 특재 사건을 인하여 등장하고 있음을 알 수 있다.

소원·편안 같은 기쁨(3 회 · 11.1 %)은 비중이 점차 줄어들며 슬픔이 전반을 차지하고 분노 정서의 비율이 점점 상승하며 극의 갈등이 점점 극대화 되어가고 있음을 알 수 있다.

그림 2. 2 단계 감정 분포 파이차트

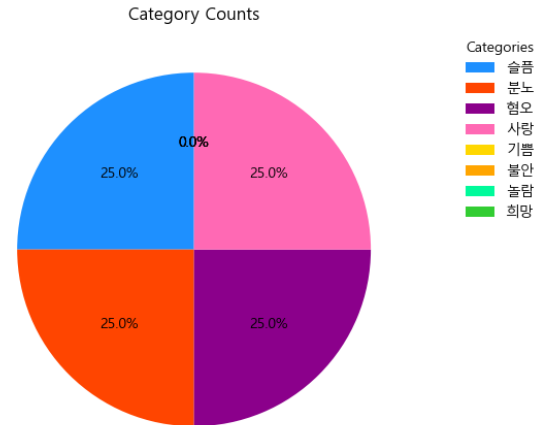


3.3. 3 단계 : 위기

3 단계는 감정 빈도가 단 4 개로 급감한다. 슬픔·분노·혐오·사랑이 각 1 회(25 %)만 출현한다.

이는 길동이 공직에 잠시 몸을 담거나 방황하는 구간으로, 사실상 정서적 공백이 형성된다.

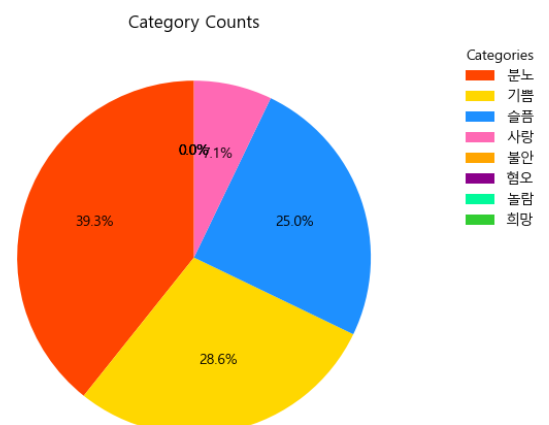
그림 3. 3 단계 감정 분포 파이차트



3.4. 4 단계 : 절정

4 단계에서는 분위기가 돌변한다. 감정 명사는 28 개로 다시 치솟고, 분노는 11 회(39.3 %)로 정점에 도달하며 탐관오리를 향한 길동의 감정이 명확히 드러난다. 동시에 기쁨이 8 회(28.6 %)로 급등하는데, 이는 활빈당 활동의 성공과 백성 구출이 가져온 카타르시스를 반영한다. 그러나 슬픔도 7 회(25 %)로 여전히 높아, 영웅이 사회적 의무를 수행하면서도 개인적 한을 완전히 해소하지 못했음을 보여 준다.

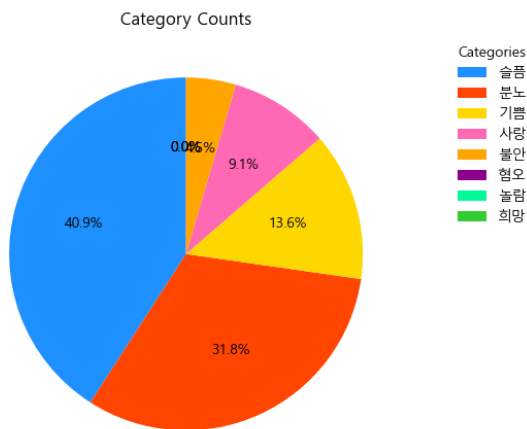
그림 4. 4 단계 감정 분포 파이차트



3.5. 5 단계 : 하강

5 단계에 이르면 슬픔이 9 회(40.9 %)로 다시 1 위가 되고, 분노가 7 회(31.8 %)로 뒤를 잇는다. 울도국 건국 이후에도 슬픔과 의분이 상당 부분 잔존한다. 기쁨(3 회 · 13.6 %)과 사랑(2 회 · 9.1 %)은 존재하지만, 길동의 정서적인 한이 완전히 사라지지 않았음을 알 수 있다.

그림 5. 5 단계 감정 분포 파이차트



3.6. 해석

정량지표로 분석한 정서의 흐름은 슬픔, 분노, 기쁨이 극의 서사에 따라 등장하며, 마지막 단계에서도 슬픔과 분노가 상당 부분 등장하며 길동의 한이 완전히 해결되지 않은 미완으로 남았음을 입증한다.

표 2. 내러티브 단계 별 감정 정량화 데이터

단계	데이터
1 발단	슬픔(8), 기쁨(3), 분노(1), 놀람(1)
2 상승	슬픔(14), 분노(6), 기쁨(3), 놀람(2), 불안(1), 혐오(1)
3 위기	슬픔(1), 분노(1), 혐오(1), 사랑(1)
4 절정	분노(11), 기쁨(8), 슬픔(7), 사랑(2)
5 하강	슬픔(9), 분노(7), 기쁨(3), 사랑(2), 불안(1)

표 3. 내러티브 단계 별 감정 단어 정량화 데이터

단계	데이터
1 발단	마음(6), 꿈(2), 생각(2), 위신(1), 적막(1)
2 상승	마음(9), 생각(5), 한(4), 설움(2), 소원(2), 원망(2), 안심(1), 편안(1), 동정(1)
3 위기	호의(1), 마음(1), 동정(1), 원(1)
4 절정	신(10), 생각(6), 소원(2), 감복(1), 동요(1), 위엄(1), 꿈(1), 애석(1), 죽음(1), 감동(1), 마음(1), 순종(1), 한(1)
5 하강	마음(4), 생각(4), 한(2), 신(2), 태평(2), 탄복(2), 편안(1), 호의(1), 측은(1), 흥악(1), 애(1), 안심(1)

4. 통합 감정선 및 추세

그림 6 에는 8 대 감정의 단계별 비중을 한눈에 보여주는 꺾은선이 제시되어 있다. 가장 많이 분포하는 슬픔, 분노, 기쁨 정서의 데이터를 직관적으로 종합하면 다음과 같은 흐름이 알 수 있다.

4.1. 슬픔

슬픔(파랑)은 1 단계 62 %에서 시작해 2 단계 52 %로 완만히 감소하다가 3 단계에서 25 %로 급락한다. 그러나 하강기인 5 단계에서 41 %로 다시 뛰어올라 ‘시작-종결을 감싸는 기본 정서’임을 확인시킨다. 급락 지점(3 단계)은 정서적 공백 상태임을 확인한다.

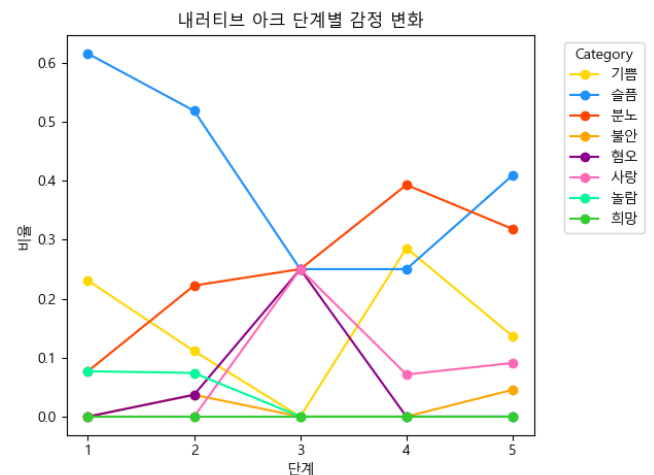
4.2. 분노

분노(빨강)는 정반대 궤적을 그린다. 1 단계 8 %에 불과하던 값이 2 단계 22 %로 두 배 이상 뛰고, 의적 전환 직후인 4 단계에 39 %로 정점에 도달한다. 마지막 5 단계에서도 32 %를 유지해, 카타르시스 이후에도 ‘잔여 의분’이 상당 부분 남아 있음을 시사한다.

4.3. 기쁨

기쁨(노랑)은 1 단계 23 % → 2 단계 11 % → 3 단계 0 %로 미끄러지다가, 활빈당 활약이 시작되는 4 단계에 29 %로 폭등한 뒤 5 단계 14 %로 반감된다. 즉 의적 성공-보상 구간에서만 일시적 낙관이 분출된다

그림 6. 내러티브 아크 단계별 감정 변화



5. 결론

본 연구는 《홍길동전》 전체 본문을 내러티브 아크 5 단계로 분할한 뒤, klt2023 형태소 분석과 465-EmoNouns KcBERT 사전을 결합해 감정 명사를 정량화하였다. 단계별·전체 빈도와 비율을 기반으로 슬픔·분노·기쁨이 만들어 내는 데이터의 분포와 3 단계에서의 정서 공백 이후 이어지는 4 단계에서의 폭발적 감정 데이터를 수치로 확인할 수 있었다. 이는 형태소 기반 감정어 분석이 서사의 정서 흐름을 가시화함으로써, 이야기 흐름의 숨은 패턴과 잔여 감정까지 선명하게 드러냈다. 앞으로도 동일 방법론을 적용하면 서사 전반에서 플롯·정서 상관 관계를 체계적으로 비교·평가에 활용이 가능할 것으로 기대된다.

참고문헌

- [1] 홍길동전 (경관 24 장본)
(<http://www.davincimap.co.kr/davBase/Source/davSource.jsp?Job=Body&SourID=SOUR001001>)
- [2] KoNLPy: Korean NLP Toolkit
(<https://konlp.readthedocs.io/en/latest/?badge=latest>)
- [3] 강승식, “감정명사 465 개, 심리동사 456 개”, 2025
(<https://ecampus.kookmin.ac.kr/mod/ubfile/view.php?id=1708195>)
- [4] RYAN L. BOYD, KATE G BLACKBURN, JAMES W PENNEBAKER, “The narrative arc: Revealing core narrative structures through text analysis”, 2020
(<https://www.science.org/doi/full/10.1126/sciadv.a2196>)
- [5] 강남오, 김재호. (2021). 옛한글 형태소 사전 구축을 위한 기법 연구. 한국지식정보기술학회 논문지, 16(6), 1305-1313. 10.34163/jkits.2021.16.6.018
- [6] 신현보, 김혜진. (2019). 주경로 분석과 연관어 네트워크 분석을 통한 ‘구전(WoM)’ 관련 연구동향 분석. 지능정보연구, 25(3), 179-200.
- [7] Emily Bishop, “Finding the arc: A story about a search for narratives”, 2012
(<https://www.jbe-platform.com/content/journals/10.1075/ni.22.2.09bjs>)
- [8] 강승식. 다층 형태론과 한국어 형태소 분석 모델. 한국정보과학회 언어공학연구회 학술발표 논문집, 1992
- [9] 조단비, 이현영, 강승식. (2020-07-02). SentencePiece 토큰라이저와 부분단어 임베딩 기법을 이용한 비형식적 텍스트의 감정 분석. 한국정보과학회 학술발표논문집, 서울.
- [10] 원혜진, 이현영, 강승식. (2020-07-02). 대규모 텍스트 분석을 위한 한국어 형태소 분석기의 실행 성능 비교. 한국정보과학회 학술발표논문집, 서울.
- [11] OpenAI. (2025) ChatGPT (version o3) [Large language model]. <https://chatgpt.com>