|  |
| --- |
| 내러티브 아크 기반 5단계 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 24장본) [1]을 내러티브 아크(탄생–고조–전환–하강–귀환)의 5단계로 분할한 뒤, klt2023[2] 형태소 분석기를 이용해 명사 어휘를 추출하고, *465-EmoNouns\_KcBERT* 감정사전을 매핑하여 단계별·전체 감정 분포를 계산했다.  주요 결과는 ①슬픔–분노가 서사 전반의 축을 형성하며, ② 서사의 중반(4단계)에서 분노와 기쁨이 역전(angry-cathartic shift)되고, ③ 귀환 단계에서 슬픔이 재상승하지만 분노가 잔존하는 정략적 지표를 확인했다. 본 연구는 고전소설 정서학적 독해에 감정 어휘 벡터를 적용한 방법론을 제안한다. |
|  |

**1. 서론**

조선 후기 작자미상의 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 아크 모델은 고전 서사의 상승–위기–해결 구조를 파악하는 데 적합하다.

**2. 데이터 및 방법론**

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

각 단계의 분할에 대해서는

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

의 5단계로 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 1:1로 맵핑한다.

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

표 1. 진행 절차 요약

|  |  |
| --- | --- |
| **절차** | **상세** |
| 데이터 준비 | 《홍길동전》(경판 24장본)을 를 ①탄생·소외, ②갈등 고조, ③공직 갈등, ④적, ⑤귀환 단계로 분할 |
| 형태소 분석 | Konlp.klt2023분석기 사용. 명사만 추출. |
| 감정 맵핑 | *465-EmoNouns\_KcBERT* 사전의 8대 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 단계별 감정 분석 결과**

TBD

표 2.

탄생–고조–전환–하강–귀환

**3.1. 1단계 : 탄생**

**3.2. 2단계 : 고조**

**3.3. 3단계 : 전환**

**3.4. 4단계 : 하강**

**3.5. 5단계 : 귀환**

**4. 통합 감정선 및 추세**

**5. 논의**

**4. 결론**

**참고문헌**

[1] 홍길동전 (경판24장본)

(http://www.davincimap.co.kr/davBase/Source/davSource.jsp?Job=Body&SourID=SOUR001001)

[2] KoNLTK(Korean Natural Language Toolkit) – KoNLP

(https://konlp.readthedocs.io/en/latest/?badge=latest)

[3] 조단비, 이현영, 정원섭, 강승식, “부분 단어 토큰화 기법을 이용한 뉴스 기사 정치적 편향성 자동 분류 및 어휘 분석,” 정보처리학회논문지: 소프트웨 어 및 데이터 공학, Vol.10(1), pp.1-8, 2021.

[4] Gaydhani, A., V. Doma, S. Kendre, and L. Bhagwat, “Detecting hate speech and offensive language on twitter using machine learning: an n-gram and tfidf based approach,” in arXiv:1809. 08651, 2018.

[5] Gitari, N., Z. Zuping, H. Damien, and J. Long, “A lexicon-based approach for hate speech detection,”International Journal of Multimedia and Ubiquitous Engineering, Vol.10(4), pp.215-230, 2015.

[6] Davidson, T., D. Warmsley, M. Macy, and I. Weber, “Automated hate speech detection and the problem of offensive language,” in Proceedings of the 11th International AAAI Conference on Web and Social Media, pp.512-515, 2017.

[7] Alshalan, R., and H. A. Khalifa, “Hate speech detection in Saudi twittersphere: A deep learning approach,” in Proceedings of the 5th Arabic Natural Language Processing Workshop, pp.12-23, 2020.

[8] Rizwan, H., M. Haroon, and A. Karim, “Hate-speech and offensive language detection on Roman urdu,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Langauge Processing, pp.2512-2522, 2020.

[9] Moon, J., W. I. Cho, and J. Lee, “BEEP! Korean corpus of online news comments for toxic speech detection,” in arXiv:2005.12503, 2020.

[10] Mikolov, T., I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Advances in Neural Information Processing Systems, Vol.26, pp.3111-3119, 2013.

[11] Bojanowski. P., E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” Transactions of the Association for Computational Linguistics, Vol.5, pp.135-146, 2017.

[12] 조단비, 워드 임베딩과 딥러닝 기법을 이용한 혐오표현 탐지, 국민대학교 석사학위 논문, 2021.