

핵심어 추출을 이용한 행위자 모델 기반 서사 캐릭터 역할 모델링에 관한 연구

장수지[○], 정해빈, 유민경, 배병철

홍익대학교 게임학부

{klop100418, bean9194, mingyeongyu8}@gmail.com, byuc@hongik.ac.kr

A Study of Actantial Model-based Character Role Modeling in Narrative using Keyword Extraction

Suji Jang[○], HaeBeen Jung, MinKyeong Yu, Byung-Chull Bae

Dept. of Game Software, Hongik University

요 약

본 논문은 텍스트 스토리 데이터에서 핵심어 추출 기법을 사용하여 서사 내 캐릭터 관계를 분류하는 모델에 대한 연구 방법을 기술한다. 이야기 속 전형적인 인물인 행위자의 행동으로 서사를 이해하는 그레마스의 행위자 모델을 기반으로 캐릭터를 행위자 항목별로 분류하여 캐릭터 사이의 관계와 목표를 분석하였으며, 사람 평가자 2명의 분석 결과와 비교하였다. 분석 결과, 대상보다는 주체가 일치도가 높았으며, 주체가 복합어로 표현된 경우보다 단일어로 표현된 경우 일치도가 더 높았다.

1. 서 론

서사학자에 따라 서사(narrative)는 다양하게 정의될 수 있지만, 내용에 해당하는 스토리(story)와 표현에 해당하는 담화(discourse)로 나누어 설명할 수 있다[1]. 여기서 스토리는 인과성이 있는 사건의 시간적 순서에 따른 배열을 의미하며, 담화는 스토리를 전달하기 위한 수단으로서 시점이나 형식 등 서사의 표현 방법을 포함한다. 스토리 연구는 서사의 중요한 요소인 캐릭터의 목표와 의도를 분석하고, 캐릭터의 관계를 이해한다는 점에서 중요하다.

언어학자 그레마스(Greimas)는 캐릭터들 간의 관계를 중심으로 아래 그림 1과 같이 행위자 모델(actantial model) [2]을 통해 서사구조를 설명한다. 행위자(actant)는 이야기 안에서 역할과 관계에 따라 규정되는 전형적인 인물을 뜻하며, 총 6개의 항목으로 분류된다: **주체(subject)**: 이야기의 주인공이자 행위를 이끄는 사람), **대상(object)**: 주체가 궁극적으로 추구하며 결핍을 느끼는 것), **조력자(helper)**: 주체가 대상을 얻기 위한 모험을 도와주는 인물), **적대자(opponent)**: 주체의 모험을 방해하는 인물), **발신자(sender)**: 주체를 모험 속으로 보내는 인물), **수신자(receiver)**: 최종적으로 대상을 얻는 인물). 행위자 모델이 스토리 연구에 쓰인 사례로는 게임 스토리텔링을 위해 플레이어의 아바타와 캐릭터 사이의 관계를 정의할 때 사용되거나[3], 서사 내의 영웅과 악당을 식별하는 스토리 분석 알고리즘[4] 구현 등이 있다.

인공지능을 이용한 서사 관련 연구는 ‘서사 지능(narrative intelligence)’이라는 용어를 통해

정의되며, 서사를 읽고 이해하고 생성하는 것 등을 포함한다[5]. 관련 최신 연구로는 서사를 이해하여 요약해주는 서사 요약 시스템[6]이나 챗봇, 게임에 활용 가능한 인터랙티브 서사 생성[7] 등이 있다. 본 논문은 서사 이론인 그레마스의 행위자 모델을 이용하여 캐릭터의 목표와 관계를 분석하는 캐릭터 관계 분류 기반의 서사 이해 모델에 대해 기술한다.

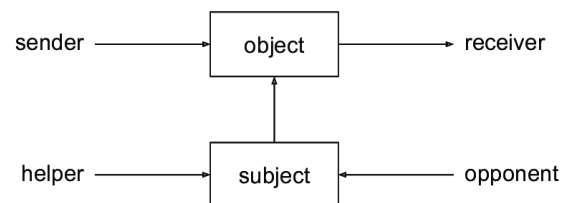


그림 1. 행위자 모델 구조

2. 관련 연구

서사에서 중요한 두 가지 요소는 플롯과 캐릭터라고 할 수 있으며, 플롯은 사건의 인과성에 중점을 둔다는 점에서 중요하다. 예를 들면, 영화 ‘Sunspring’[8]은 일반인들에게 인공지능이 작성한 각본으로 주목을 받았지만, 동시에 개연성이 없고 기괴한 영화라는 평가를 받았다. 이는 플롯의 인과성이 고려되지 않은 채 사건을 생성하고 배열했기 때문이다.

캐릭터의 목표와 의도를 분석하는 연구는 캐릭터 사이의 관계를 분석하는 데 있어 기본적인 연구라 할 수 있다[9]. 캐릭터의 목표 인식 모델에 관한 연구로는

마르코프 논리 네트워크(Markov Logic Networks)[10], Long Short Term Memory 네트워크[11] 등 다양한 방법으로 진행되었다. 그리고 캐릭터 사이의 관계 분석에 대한 연구로는 비지도 학습 모델을 사용해 캐릭터 사이의 관계를 추출하려는 시도가 있었으며[12], CRF(Conditional Random Field) 레이어를 추가한 양방향 LSTM 모델에 오페라 텍스트 대본에서 캐릭터 사이의 관계를 자동으로 추출하는 연구도 있었다[13].

본 논문은 행위자들의 행동으로 서사 구조를 분석하는 그레마스(Greimas)의 행위자 모델(actantial model)을 기반으로 서사 내 캐릭터 사이의 관계를 분류하는 서사 이해 모델의 구축을 제안한다.

3. 서사 이해 모델 구현을 위한 연구 방법

3.1 데이터셋

본 연구에서는 서사 이해 모델의 구축을 위해 ROC 스토리를 이용하여 그레마스 행위자 모델에서 분류할 항목을 추출하였다. ROC 스토리는 5개의 문장으로 구성되며, 풍부한 양의 인과적인 일상적 사건으로 상식 관계를 포착한다.

ROC 스토리는 대체로 인물들이 1명에서 3명 사이로 등장하며, 각각의 인물이 맡게 되는 역할과 목표의 수는 1개 정도이다. 또한, 인물 간의 관계와 목표가 단순하기 때문에 초기 연구로써 행위자 모델을 적용하여 항목으로 분류하기 용이하다.

3.2 자연어 처리를 이용한 행위자 추출 방법

단어의 출현 빈도를 활용하는 것은 기본적인 핵심어 추출 방법들 중 하나이다. 본 논문은 단어의 출현 빈도를 활용한 단어 빈도 분석과 TextRank 기법[14]을 사용하여 텍스트 스토리 데이터에서 행위자 모델의 항목 중 주체와 객체를 중심으로 추출하며, 그림 2와같은 3단계의 방법으로 진행한다. 첫번째 단계는 전처리 과정으로, 먼저 모든 대문자를 소문자로 변환하였다. 다음으로 의미는 없지만 높은 출현 빈도를 보이는 불용어를 제거하였다. 불용어는 기본적으로 NLTK(Natural Language Toolkit)[15]에서 제공하는 Stopword를 사용했으며, I, she, he 등 ROC 스토리에서 주어로 사용될 수 있는 주격 인칭대명사는 제외하였다.

두번째 자연어 처리 (Natural Language Processing) 단계에서는 두 가지 방법을 적용하였다. 첫 번째 방법은 상호참조 해결(Coreference Resolution)[16]이다. 상호참조 해결은 문서 내에서 이미 언급된 개체에 대하여 대명사로 다시 언급된 경우를 상호참조(Coreference)라 할 때, 상호참조가 가능한 멘션(mention)들끼리 하나의 엔티티(entity)로 참조 관계를 만드는 것이다. 전처리된 데이터에 상호참조해결을 사용해 대명사를 명사로 변환하였다. 두 번째 방법은 품사 분석으로, NLTK에서 제공하는 POS

tagging(Part-Of-Speech tagging)을 사용하여 각 토큰에 대한 품사를 태깅하였다.

마지막 세번째 단계는 핵심어 추출 기반 행위자 분류이다. 주체의 경우 명사 혹은 대명사로 태깅된 토큰의 단어 빈도를 카운트하여 가장 많이 카운팅된 토큰을 주체로 분류하였다. 이후, 주체를 제외한 명사 혹은 대명사 토큰과 동사 토큰에 TextRank 알고리즘을 적용하여 가장 높은 랭킹값을 갖는 토큰을 객체로 분류하였다.

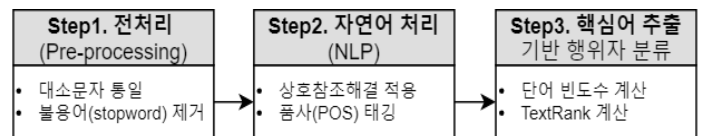


그림 2. 핵심어 추출을 이용한 행위자 추출 3단계 방법

3.3. 행위자 태깅 설계 및 결과

본 태깅 모델은 ROC 스토리와 주체, 대상으로 구성된다. 주체와 대상은 세부 항목으로 단일어와 복합어가 존재하는데, 주체 또는 대상이 한 단어로 명시할 수 있을 경우는 단일어에 태깅하며, 한 단어로 명확하게 명시할 수 없는 경우 복합어에 태깅하도록 하였다. 태깅은 총 2명의 학부 대학생을 대상으로 진행하였다.

표 1. 태깅 데이터 분석

분류	주체(subject)		대상(object)	
	단일어	복합어	단일어	복합어
일치	16	3	9	3
불일치	1		8	
계	20		20	

태깅 결과, 주체(subject)에 대한 태깅은 95% (19/20)의 일치도를 보였으며, 객체(object)에 대한 태깅은 60% (12/20)의 일치도를 보였다 (표 1 참조). 주체(subject)의 경우 20개의 서사 내 주체 중 12개의 주체가 태깅 데이터와 일치하였다(일치도 60%). 12개의 주체는 태깅 데이터에서 두 사람 모두 단일어로 태깅한 16개의 주체 안에 포함된다. 대상(object)은 개의 스토리만 태깅 데이터와 일치하였으며 (일치도 30%), 주체와 마찬가지로 일치한 6개의 대상은 태깅 데이터에서 두 사람 모두 단일어로 태깅한 경우였다.

결과를 정리하면, 명사 한 단어로 명확하게 추출할 수 있는 주체 분류가 대상 분류보다 일치도가 높았다. 또한, 대상 분류의 경우 단일어보다 복합어의 비율이 더 높았으며 태깅 간의 일치도가 낮았다. 이는 대상에 대한

복합어의 기준과 대상이 될 수 있는 행위와 객체의 분류 기준이 명확하게 정의되지 않았기 때문이다.

4. 결론 및 향후 과제

본 논문은 텍스트 스토리 데이터에서 핵심어 추출 기법을 사용하여 서사 내 캐릭터 관계를 분류하는 모델에 대한 연구 방법을 기술하였다. 이야기 속 전형적인 인물인 행위자의 행동으로 서사를 이해하는 그레마스의 행위자 모델을 기반으로 캐릭터를 행위자 항목별로 분류하여 캐릭터 사이의 관계와 목표를 분석하였으며, 서사 속 중심 인물인 주체(subject)와 주체의 목표인 대상(object)에 대한 행위자 분석을 사람의 분석 결과와 비교하였다. 분석 결과, 대상보다는 주체가 일치도가 높았으며, 주체가 복합어로 표현된 경우보다 단일어로 표현된 경우 일치도가 더 높았다.

향후 연구로는 복합어의 모호한 태깅 기준을 확립하여 사람의 태깅 데이터간 일치도를 높이면서 추가적으로 태깅을 진행할 예정이며, 이후 태깅 데이터를 활용하여 언어 모델을 활용한 학습 모델을 개발할 계획이다.

Acknowledgement

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2017R1A2B4010499) and Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2017-0-01772, Development of QA systems for Video Story Understanding to pass the Video Turing Test).

참 고 문 헌

- [1] Genette, Gérard. "Narrative Discourse". 1980.
- [2] Greimas, Algirdas Julien. "Reflections on actantial models." London: Longman Group Ltd, 1996.
- [3] Willumsen, Ea Christina. "Is My Avatar MY Avatar? Character Autonomy and Automated Avatar Actions in Digital Games." Proc. of DiGRA. 2018.
- [4] Schirfe, Henrik. "Searching for narrative structures." AAAI 2002 Spring Symposium. 2002.
- [5] Mateas, Sengers. "Narrative Intelligence." AAAI. 1999.
- [6] Zhang, Renxian, et al. "Coherent narrative summarization with a cognitive model." Computer Speech & Language 35. pp.134-160. 2016.
- [7] Womack, Jon, and William Freeman. "Interactive Narrative Generation Using Location and Genre Specific Context." International Conference on Interactive Digital Storytelling. Springer, Cham, 2019.
- [8] Sharp, Oscar, and R. Goodwin. "Sunspring: A Sci-Fi Short Film Starring Thomas Middleditch." End Cue. Available online: <https://www.youtube.com/watch>. 2016.
- [9] Woloch, Alex. The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel. Princeton University Press, 2009.
- [10] Ha, Eun Young, et al. "Goal recognition with Markov logic networks for player-adaptive games." Seventh Artificial Intelligence and Interactive Digital Entertainment Conference. 2011.
- [11] Min, Wookhee, et al. "Player Goal Recognition in Open-World Digital Games with Long Short-Term Memory Networks." IJCAI. 2016.
- [12] Nijila, M., and M. T. Kala. "Extraction of Relationship Between Characters in Narrative Summaries." 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR). IEEE, 2018.
- [13] Dai, Xinnan, et al. "Opera-oriented character relations extraction for role interaction and behaviour Understanding: a deep learning approach." Behaviour & Information Technology 38.9. pp.900-912. 2019.
- [14] Mihalcea, Rada, and Paul Tarau. "TextRANK: Bringing order into text." Proceedings of the 2004 conference on empirical methods in natural language processing. 2004.
- [15] Steven, Bird, and L. Edward. "NLTK: the natural language toolkit." Proceedings of the COLING/ACL on Interactive presentation sessions. 2006.
- [16] Clark, Kevin, and Christopher D. Manning. "Improving coreference resolution by learning entity-level distributed representations." arXiv preprint arXiv:1606.01323. 2016.