

멀티모달 자질을 활용한 다중 화자 대화 속 인물 식별

한기종^{0,1}, 최성호², 신기연¹, 장병탁², 최기선¹

¹ 한국과학기술원, 시맨틱웹첨단연구센터
{han0ah, nuclear852, kschoi}@kaist.ac.kr

² 서울대학교, 바이오지능 연구실
{shchoi, btzhang}@bi.snu.ac.kr

Character Identification on Multiparty Dialogues using Multimodal Features

Kijong Han^{0,1}, Seong-Ho Choi², Giyeon Shin¹, Byoung-Tak Zhang², Key-Sun Choi¹

¹ KAIST, Semantic Web Research Center

² Seoul National University, Biointelligence Lab

요약

다중 화자 대화 속 인물 식별이란 여러 등장인물이 나오는 대본에서 ‘그녀’, ‘아버지’ 등 인물을 지칭하는 명사 또는 명사구가 실제 어떤 인물을 나타내는지 파악하는 문제이다. 대본 자연어 데이터만을 입력으로 하는 대화 속 인물 식별 문제는 드라마 대본에 대해서 데이터가 구축되었고 이를 기반으로 여러 연구가 진행되었다. 그러나, 사람도 다중 화자 대화의 문장만 보고는 인물을 지칭하는 명사 또는 명사구가 실제 어떤 인물인지 파악하기 어려운 경우가 있다. 이에 본 논문에서는 발화가 되는 시점의 영상 장면 정보를 추가적으로 활용하여 인물 식별의 성능을 높이는 방법을 제시한다. 또한 기존 대화 속 인물 식별 연구들은 미리 정의된 인물을 대상으로 분류하는 형태로 접근해왔다. 이는 학습에 사용되지 않았던 인물이 나오는 임의의 다른 드라마 대본이나 대화 등에 바로 적용될 수 없다. 이에 본 논문에서는 영상 정보는 활용하되, 한번 학습하면 임의의 대본에 적용될 수 있도록 사전 인물 정보를 사용하지 않는 상호참조해결 기반의 인물 식별 방법도 제시한다.

주제어: 인물 식별, 멀티모달, 개체 연결, 상호참조해결

1. 서론

다중 화자 대화 속 인물 식별(Character Identification)이란 드라마 대본 등 여러 등장인물이 나오는 대본에서 인물 멘션(Mention)들을 실제 인물과 연결해 주는 문제이다[1]. 인물 멘션이란 ‘그녀’ 또는 ‘아버지’와 같이 인물을 지칭하는 명사 또는 명사구이다. 즉 이러한 명사 또는 명사구들이 실제 어떤 인물을 나타내는지 파악하는 문제라고 볼 수 있다. 대화 문장 속에서 인물 멘션이 의미하는 실제 인물을 파악하는 문제는 질의 응답, 요약과 같은 고차원 자연어처리 문제를 해결하기 위해 중요한 단계이다 [2]. 인물 식별 문제는 미국 드라마 프렌즈(Friends) 두 시즌(Season) 47개 화(Episode)에 대해서 데이터가 구축되었고[1,2], SemEval 2018 Task4[3]로 공개되어 다수의 연구팀이 참여하는 등 여러 관련 연구가 진행되어 왔다[1-5].

이 문제는 자연어처리에서 개체 연결(Entity Linking)[6] 및 상호참조해결(Coreference Resolution)[7,8] 문제와 비슷한 형태를 가진다. 그러나 연결 대상이 되는 개체(Entity)가 지식베이스 개체가 아닌 대화 속 등장인물이기에 지식베이스 정보를 활용할 수 없고, 기존 상호참조해결 문제처럼 같은 개체를 나타내는 멘션(Mention)들끼리 묶어주는 것 뿐만 아니라 실제 인물과 연결까지 해주어야 된다는 점이 다르다. 또한 이러한 다중 화자 대화/대본은 은유, 수사적 표현 등이 많은 구어



그림 1 영상 정보 없이 대본 정보만으로는 대화 속 인물 식별이 어려운 예시

체로 구성되며 문맥 변화가 빠르게 일어나기에 이 문제를 어렵게 만든다[1,2].

기존 연구들이 다양한 기법을 제시하고[2-5] 성능을 어느정도 올렸지만, 이 연구들은 자연어 대본만을 입력으로 사용했다. 그러나 사람도 대본 문장만을 보고 인물 멘션이 누구를 지칭하는지 식별을 하기 어려운 경우들이 존재한다. 그림 1의 예시와 같이 어떤 여성이 잠을 자고 있고, 사람들이 이 여성을 대상으로 대화를 나누고 있는 영상 정보가 없다면 그림 1의 대본 만으로 ‘그녀’가 이 여성을 지칭한다는 것을 알아내기 힘들다. 이에 본 연구에서는 대본 속 해당 문장이 발화되는 시점의 영상 장면 정보까지 활용하는 방법을 제시한다. 이 방법은 다

IDs	단어	화자	발화 시간	영상 정보 파일	인물 ID
/friends-s01e01 0 0	He PRP (TOP(S(NP*)	he - - Monica_Geller *	55422 59256	00005.pickle	(284)
/friends-s01e01 0 1	's VBZ (VP*	be - - Monica_Geller *	55422 59256	00005.pickle	-
/friends-s01e01 0 2	just RB (ADVP*)	just - - Monica_Geller *	55422 59256	00005.pickle	-
/friends-s01e01 0 3	some DT (NP(NP*	some - - Monica_Geller *	55422 59256	00005.pickle	-
/friends-s01e01 0 4	guy NN *)	guy - - Monica_Geller *	55422 59256	00005.pickle	(284)
/friends-s01e01 0 5	I PRP (SBAR(S(NP*)	I - - Monica_Geller *	55422 59256	00005.pickle	(248)
/friends-s01e01 0 6	work VBP (VP*	work - - Monica_Geller *	55422 59256	00005.pickle	-

그림 2 입력 및 학습 데이터 예시

양한 평가 지표에서 최대 4.74%의 성능향상을 이루었다.

또한 기존 연구들은 주로 미리 주어진 등장 인물을 대상으로 인물 멘션들을 분류하는 형태로 접근해 왔다 [2,4,5]. 이러한 방식으로 학습된 모델들은 학습에 사용한 프렌즈(Friends) 대본이 아닌 다른 드라마나 임의의 대본에서는 작동하지 않는다. 새로운 대본에 나오는 새로운 등장인물을 분류하도록 다시 학습되어야 하기 때문이다. 따라서 이 문제는 상호참조해결을 주 접근법으로 한 후, 최종 인물 연결은 상호참조해결에서 나온 멘션 그룹들을 등장 인물들과 이름 문자열 일치 수준 연결 등의 학습이 필요 없는 방식으로 접근해야 임의의 대본에 적용될 수 있다. [2]에서 상호참조해결 기반의 접근법도 제시하기는 했지만 비교적 성능이 낮다. 이에 본 논문에서는 [2]의 연구보다 진보된 상호참조해결 방법론을 사용하고, 영상 정보까지 결합하여 임의의 대본에 적용될 수 있는 상호참조해결 기반 인물 식별도 어느정도 의미 있는 성능을 낼 수 있음을 보인다.

2. 관련 연구

인물 식별(Character Identification) 문제는 처음 [1]에서 구체적으로 정의되었고 프렌즈 2시즌(Seson)에 대해서 데이터가 구축되었다. [2]에서는 이 데이터의 오류를 정제하고 이 문제를 해결하는 방식을 제시하였다. 응집하는 CNN(Agglomerative CNN) 구조 기반의 상호참조해결과 개체 연결 모듈을 결합하여 학습하는 방식이다. [3]에서는 그 동안의 연구결과와 데이터를 정제하고 평가 지표를 공식화 하여 SemEval 2018 Task4로 공개되었다.

이 Task에 여러 연구팀이 참여하였고, [4,5] 연구가 우수한 성능을 보였다. [4] 연구는 이 문제를 다중 분류 형태로 정의하고 Bi-LSTM 기반으로 접근했으며 각 인물의 정보를 벡터로 표현하고 있는 행렬인 Entity Library란 요소를 도입한 방법을 제시했다. [5] 연구는 이 문제를 시퀀스 레이블링(Sequence Labeling) 형태로 정의하고 RNN Encoder-Decoder 기반의 구조를 사용하였다. RNN 계열 모델들이 이 문제와 같이 문서 단위의 긴 입력에서의 정보 손실로 인한 성능 저하 문제를 해결하고자 포지션 인코딩(Position Encoding)과 셀프 매칭 네트워크(Self-matching network)를 결합한 구조를 제시했다. 본 연구는 이 기존연구들과 다르게 영상정보까지 사용하여 성능을 올렸으며 [4] 연구의 방법론 기반에 영상정보를 결합하였다.

영상 또는 이미지와 자연어 정보를 결합해서 입력으로 활용하는 멀티모달 형태의 문제에서는 영상 장면 또는 이미지 자질을 추출하기 위해 ResNet[11] 구조가 많이 사용된다[9,10]. 본 연구에서도 이 기법을 사용하였다.

상호참조해결은 최근 종단간 Bi-LSTM 기반의 모델을 사용한 연구가 우수한 성능을 보여준다[7,8]. [7]에서 명사구의 중심어를 어텐션(Attention)으로 학습하고 기본적으로 모든 명사구 후보와 모든 후보 쌍 간의 상호참조 해결 가능성을 고려하는 형태의 뉴럴넷 구조를 제시하였다. [8]에서는 [7]구조 기반에서 상호참조해결의 일관성을 위해 고차 추론(Higher-order Inference)을 하고 연산량을 좀 더 효율적으로 줄이는 방식을 제시하였다. 본 연구에서는 약간 변형된 [8] 연구의 상호참조해결 구조를 기반으로 하되 영상 정보까지 결합하여 인물 식별을 하는 방법을 제시한다.

3. 입출력 정의 및 데이터 전처리

3.1 입출력 정의

이 문제는 토큰화 및 자연어 처리가 된 대본, 대본의 해당 문장이 발화된 영상 속 시간 정보, 해당 시간에 해당하는 영상 전처리 파일 정보, 인물 멘션의 위치를 입력으로 받는다. 출력으로는 각 인물 멘션들이 나타내는 실제 인물 ID를 출력한다. 입력은 CoNLL 형식으로 되어 있으며 예시는 그림 2에 나와 있다. 그림 2의 예시 데이터는 “Monica_Geller” 라는 화자가 “He’s just some guy I work” 라는 문장을 발화하였고 해당 발화가 영상의 55422ms 부터 59256ms 구간에 나온다는 의미이다. 또한 이 시점의 영상 정보는 00005.pickle 파일에 저장되어 있다는 의미이다. 그리고 이 문장 속에 “He”, “Guy”, “I” 라는 인물 멘션들이 각각 284, 284, 248 번의 ID를 가지는 인물을 나타낸다. 학습 데이터에는 이 인물 ID까지 구체적으로 명시되어 있고, 테스트 데이터에서는 위치만 표현되고 이 아이디어가 어떤 값인지는 모델이 출력을 해야한다.

이 데이터 형태에서 시간 정보 및 영상 전처리 파일 정보를 제외하고는 SemEval2018 Task4[3]에서 공개된 데이터를 활용했다. 발화 시간 정보를 이 데이터에 부착하기 위해 SemEval 데이터 대본과 자막 데이터 정보를 동기화 하였고 이를 3.2절에 설명하였다. 영상 정보 전처리 파일에 대해서는 3.3절에 설명하였다.

3.2 대본과 자막 정보 동기화

자막 데이터는 시간정보를 가지고 있지만 대본 데이터는 시간정보가 없다. SemEval2018 Task4[3] 데이터는 대본을 기준으로 구축되었기 때문에 시간 정보를 부착하기 위해 자막 데이터와 동기화하는 작업을 하였다. 대본 데이터의 문장과 자막 데이터의 문장이 온전히 일치하지 않았기 때문에 이 작업이 필요하였다.

먼저 자막 대사 하나와 대본 대사 하나를 기본 단위로 하여, 각 대본 대사에 대해서 토큰 유사도가 일정 값 이상인 후보군을 추출하였다. 이후 최대한 많은 대본 대사에 자막과 매칭될 수 있도록 최장 증가 수열 알고리즘을 통해서 각 대본 대사에 0개 또는 1개의 자막을 매칭시켜서 시간정보를 부착하였다.

3.3 영상 전처리

발화가 되는 구간의 영상 자질을 표현하기 위해 먼저 해당 영상 구간을 초당 24개의 이미지를 가지는 각 장면(Frame)을 나누었다. 각 장면 이미지에 대하여 ImageNet으로 미리 학습된(Pre-trained) ResNet-152[11] 네트워크를 통과시킨 후, 최종 분류 레이어 직전 Hidden Layer인 2048 크기의 장면의 벡터값으로 추출한다. 이 장면에 대한 각각의 벡터값들을 평균 낸 것을 해당 영상 구간에 포함되는 모든 장면의 벡터값을 평균 을 나타내는 입력 자질 벡터로 사용하였다.

4. 접근 방식

4.1 미리 정의된 등장 인물을 대상으로 분류하는 방식

본 논문의 첫번째 접근 방식 중에 하나인 미리 정의된 인물 대상으로 분류하는 방식은 SemEval 2018 Task4[3]에서 우수한 성능을 보인 연구 중 하나인 AMORE-UPF 구조[5]를 그대로 사용하되, 추가적으로 영상 정보를 활용할 수 있는 파이프라인을 결합하였다. 이 내용은 그림 3에서 확인할 수 있다.

[5] 연구의 모델을 간략히 요약하자면 다음과 같다. 먼저 입력 단어 멘션(그림 3에서 He)을 word2vec[12]을 이용하여 벡터로 표현하고 해당 멘션의 발화자(그림 3에서 Monica)를 랜덤 하게 벡터로 표현 한 후 이 두 벡터를 합쳐서 입력으로 사용한다. 이 입력을 Bi-LSTM 을 통과시킨다. 그 다음으로 인물 식별을 하고자 하는 i번째 멘션의 Bi-LSTM 출력 벡터를 Linear Layer를 통과 시킨다 이 벡터를 e_i 라고 하자.

그림 3에서 Entity Library 행렬 $E \in \mathbb{R}^{N \times k}$ 는 N 개의 분류 대상이 되는 등장 인물 개체(Entity)를 각 k 차원의 벡터로 표현한 행렬이다. 이 E 는 학습과정에서 업데이트 된다. 이 때 멘션을 표현하는 e_i 와 각 인물 특징을 표현하는 Entity Library 행렬 E 의 각각의 벡터와 코사인 유사도(Cosine Similarity)를 계산한다. 이 값을 Softmax 레이어를 통과시켜 해당 멘션이 나타내는 인물이 누구인지 분류하는 방식이다.

그림 3에서와 같이 우리는 여기에서 멘션의 최종 표현인 e_i 벡터가 영상 정보도 반영하도록 수정하였다. ResNet-152[11]을 통과한 영상 정보 벡터를 ReLU 활성화함

수를 가진 출 Linear Layer를 통과 시킨 벡터 v_i 를 e_i 를 계산하는 데에 입력으로 들어가는 h_i 와 결합시켰다. 멘션이 각 인물을 나타낼 확률값을 가지고 있는 최종 출력 벡터 o_i 는 다음 수식은 다과 같이 표현될 수 있다.

$$o_i = \text{softmax}(\text{cosine}(E, (W_o(h_i || v_i) + b)))$$

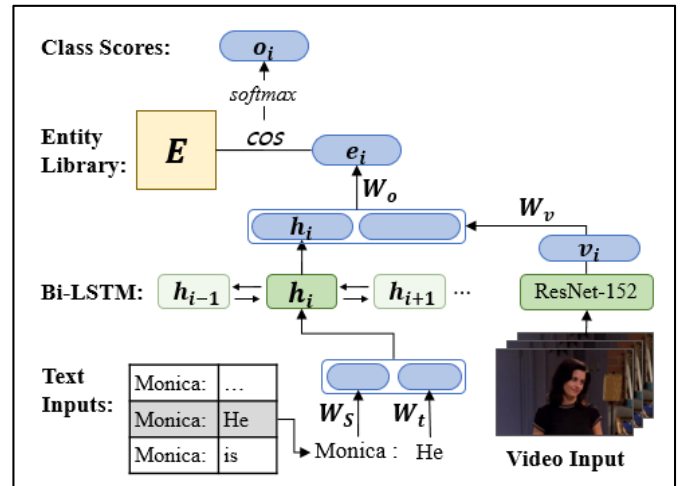


그림 3 미리 정의된 등장 인물을 대상으로 분류하는 인물 식별 방식 알고리즘 개요

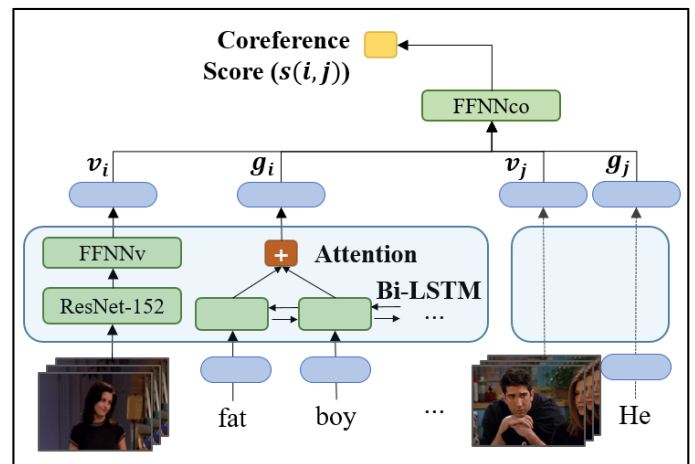


그림 4 상호참조해결 기반 인물 식별 알고리즘 개요

4.2 상호참조 해결 기반 방식

4.2.1 상호참조 해결 알고리즘

본 논문에서는 미리 정의된 인물 정보 없이 한번 학습 후 임의의 대본에 적용될 수 있는 상호참조해결 기반의 방식도 제안한다. 먼저 상호참조해결 알고리즘은 [8] 모델을 기반으로 하였다.

[8] 모델을 간단히 설명하면 다음과 같다. 먼저 입력 단어 들을 Glove[13], ELMo[14] 임베딩과 Bi-LSTM 레이어를 통해 표현한다. 각 멘션이 여러 단어로 이루어져 있을 경우는 중심어 등 중요한 단어에 좀 더 가중치를 주기 위하여 이를 Attention으로 학습 하고, Attention 레이어를 통과한 i번째의 멘션 표현 벡터 g_i 를 생성해낸다. 그리고 상호참조인지 알고자 하는 두 멘션 벡터 표현을

표 1 각 방법별 인물 식별 성능

방법	주요 6명 인물 + 나머지 대상		전체 인물 대상	
	정확도	F1-Score	정확도	F1-Score
SemEval 2018 방법들	Zuma-AR	46.85	44.68	33.06
	KNU-CI	85.10	86.00	69.49
	Kampfpujdding	75.37	73.51	59.45
	AMORE-UPF	77.23	79.36	74.72
<i>AMORE-UPF (+영상 정보)</i>		<i>81.97(+4.74)</i>	<i>83.82(+4.46)</i>	<i>76.49(+1.77)</i>
상호참조해결 기반 접근법	대본 정보만 활용	78.06	77.95	67.48
	+영상 정보	<i>79.99(+1.93)</i>	<i>79.64(+1.69)</i>	<i>70.52(+3.04)</i>
				<i>44.25(+2.14)</i>

Fully Connected Layer를 통해 멘션 i 가 멘션 j 의 선행사일 점수 $s(i, j)$ 를 계산한다. 현재 멘션에 대해 앞의 모든 선행하는 멘션에 대해 이 $s(i, j)$ 를 포함하는 수식을 통해 계산되는 최종 상호참조 점수(Coreference Score)를 비교하여 가장 점수가 높은 멘션과 상호참조라고 보고 그 멘션과 같은 대상을 나타낸다고 그룹화를 해준다.

이 모델은 모든 단어 조합과, 모든 멘션 조합을 기본적으로 연산 대상으로 한다. 이렇게 되면 연산량이 많기 때문에 멘션이 될 가능성이 적은 단어 조합은 점수를 구하여 제거하는 과정을 거친다. 그러나 본 인물 식별 문제는 입력 정의에서 대상 멘션의 범위는 주어지기 때문에 우리는 이 과정에서 점수대로 멘션을 선택하는 과정을 제거하고 입력에 주어진 멘션 후보만 선정하는 과정을 추가하였다.

또한 두 멘션이 상호참조인지 계산할 때 각 멘션이 발화된 시점의 영상 정보를 반영하여 판단할 수 있도록 수정하였다. 먼저 멘션 i 가 발화된 시점의 ResNet을 통과한 영상정보 벡터를 그림 4와 같이 Fully Connected Layer(그림 4에서 $FFNN_v$)를 통과 시킨 벡터 v_i 를 생성하였다. 그 후 이 벡터 v_i 를 선행사 점수 $s(i, j)$ 를 계산하는 Fully Connected Layer(그림 4에서 $FFNN_{co}$)에 입력으로 추가적으로 사용하도록 하였다. 수정된 선행사 점수 $s(i, j)$ 계산을 수식으로 표현하면 아래와 같다.

$$s(i, j) = w_a \cdot FFNN_{co}([v_i, v_j, g_i, g_j, g_i \circ g_j, \phi(i, j)])$$

v_i, v_j 는 각각 i 번째와 j 번째 멘션이 발화된 시점의 영상 정보 벡터 표현이며, g_i, g_j 는 [8]논문에서도 사용된 위에 설명한 i 번째와 j 번째 멘션의 벡터 표현이다. \circ 는 벡터 요소별 곱(element-wise multiplication)이며 콤마(,)는 벡터 결합(Concatenate)을 나타낸다. $\phi(i, j)$ 나머지 벡터 값들은 기존 [8] 연구에서 사용된 두 멘션의 발화자가 일치하는지 정보와 두 멘션 간의 문서 상에서 거리를 나타내는 Feature 벡터이다.

4.2.2 상호참조 해결 결과를 실제 인물과 연결

인물 인식 문제가 상호참조해결과 다른 점은 상호참조해결에서 처럼 같은 대상을 나타내는 멘션들끼리 묶어주는 것 뿐만 아니라 묶인 그룹을 실제 인물과 연결해

주어야 한다. 우리는 이 과정을 [2] 논문에서 나온 규칙 기반 방식을 그대로 사용하였다. 이 과정을 요약하면 다음과 같다. 먼저 그룹 내의 각 멘션들을 다음의 우선순위로 실제 인물과 매칭한다. 1번 우선순위는 1인칭 대명사일 경우 화자인 인물과 바로 매칭하고, 2번 우선순위는 전체 이름 문자열 일치, 3번 우선순위는 이름만(성이 아닌 이름) 문자열 일치이다. 이렇게 그룹 내의 각 멘션들의 실제 인물과 매칭이 끝나면 그룹 내에서 가장 많이 매칭된 인물을 해당 그룹을 나타내는 인물이라고 선택한다.

5. 실험 및 결과

5.1 데이터 및 평가지표

데이터는 영상 정보를 추가적으로 활용했다는 것 외에 SemEval 2018 Task4[3]와 같은 데이터와 학습/테스트 데이터 구분을 사용하였다. SemEval 2018 Task4에서 드라마 한 화(Episode)를 여러 개로 나눈 씬(Scene) 단위를 문서로 잡은 기준을 사용하여 비교하였다. 총 448개의 Scene이 있으며 학습데이터는 374개의 Scene (약 14,000여 문장, 13,280 개의 멘션) 평가 데이터는 74개의 Scene(약 2,500여 문장, 2,429개의 멘션)으로 구성되어 있다.

평가로는 크게 두가지 기준이 사용된다. 먼저 주요 6명 인물을 식별하고 나머지는 “Others” 태그로 통일하여 분류하는 평가 방식이다. 다른 하나는 학습데이터와 평가 데이터에 각각 최소 한 번 이상 등장한 78명의 인물을 전체 인물이라 정의하고 이를 대상으로 분류하는 방식이다. 평가 지표로는 정확도(Accuracy)와 각 인물별로 F1 Score를 Macro Average한 Macro-F1이 사용된다

5.2 실험 결과

5.2.1 영상 정보 활용을 통한 성능 향상

표1에서 보듯이 먼저 영상정보를 사용한 것이 성능을 확실히 올렸음을 알 수 있다. 위 표에서 우리가 기반 모델로 삼은 AMORE-UPF[5]와 우리가 이 모델에 추가적으로 영상정보를 결합한 모델(AMORE-UPF+(영상 정보))을 비교했을 때 영상정보 활용 방법이 모든 지표에서 +1.77 ~ +4.74 사이의 성능 향상을 이루었음을 볼 수 있다. 상호참조해결 기반 접근법에서도 영상 정보를 추가적으로

활용한 것이 모든 지표에서 +1.69 ~ +3.04 정도의 성능 향상을 보여준다.

5.2.2 임의의 대본에 적용가능한 상호참조해결 기반 방식의 성능

두번째로 임의의 대본에 적용될 수 있는 상호참조모델 기반의 접근법으로도 어느정도 준수한 성능을 낼 수 있다는 것을 보였다. 인물에 대한 사전정보를 학습에 사용하지 않기 때문에 높은 성능의 인물 대상 분류 방식 기법들 보다는 성능이 낮지만, 주요 6명 인물에 대해서는 약 80%의 정확도, 전체 인물에 대해서도 약 70% 정도의 준수한 정확도를 보여준다. 상호참조해결 알고리즘을 잘 설계하고 영상 자질을 잘 사용하면 상호참조해결 기반 인물 인식 방법도 충분히 실용적으로 사용될 수 있을 것이다.

6. 결론

본 연구에서는 다중 화자 대화 속에서 인물을 나타내는 명사 또는 명사구들이 실제 어떤 인물을 나타내는지 식별하는 인물 식별(Character Identification) 문제에서 영상 정보를 추가적으로 활용하는 방법을 제시하고 이 방법이 성능을 올릴 수 있음을 보였다.

또한 미리 정의된 인물을 대상으로 분류하는 기존의 접근법과 다르게 한번 학습하면 새로운 인물이 나오는 임의의 대본에 적용될 수 있는 상호참조해결기반의 접근법으로도 영상정보를 활용하여 준수한 성능을 낼 수 있음을 보였다.

사사

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-01780, 비디오 이해를 위한 이벤트-상황 지식체계 학습 및 이벤트인식/관계추론 기술 개발)

참고문헌

- [1] Chen, Yu-Hsin, and Jinho D. Choi, "Character identification on multiparty conversation: Identifying mentions of characters in tv shows." Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue.
- [2] Chen, Henry Y., Ethan Zhou, and Jinho D. Choi, "Robust Coreference Resolution and Entity Linking on Dialogues: Character Identification on TV Show Transcripts." CoNLL. 2017.
- [3] Choi, Jinho D., and Henry Y. Chen. "SemEval 2018 Task 4: Character Identification on Multiparty Dialogues." SemEval. 2018.
- [4] Cheoneum Park, Heejun Song, and Changki Lee, "KNU CI System at SemEval-2018 Task4: Character Identification by Solving Sequence-Labeling Problem." SemEval. 2018.
- [5] Aina, Laura, et al. "AMORE-UPF at SemEval-2018 Task 4: BiLSTM with Entity Library." SemEval. 2018.
- [6] Avirup Sil, Gourab Kundu, Radu Florian, and Wael Hamza, "Neural cross-lingual entity linking.", AAAI, 2018.
- [7] Kenton Lee, Luheng He, Mike Lewis, and Luke S.Zettlemoyer, "End-to-end neural coreference resolution.", EMNLP, 2017.
- [8] Kenton Lee, Luheng He, and Luke Zettlemoyer, "Higher-order Coreference Resolution with Coarse-to-fine Inference", NAACL-HLT, 2018
- [9] Akira Fukui, Dong Huk Park, Daylen Yang, Anna Rohrbach, Trevor Darrell, Marcus Rohrbach, "Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding". EMNLP. 2016.
- [10] Kyung-Min Kim, Min-Oh Heo, Seong-Ho Choi, Byoung-Tak Zhang, "DeepStory: Video Story QA by Deep Embedded Memory Networks.", IJCAI, 2017.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", CVPR. 2016.
- [12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, Jeff Dean, "Distributed representations of words and phrases and their compositionality.", NIPS, 2013.
- [13] Pennington, Jeffrey, Richard Socher, and Christopher Manning, "Glove: Global vectors for word representation.", EMNLP, 2014.
- [14] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer, "Deep contextualized word representations", NAACL-HLT, 2018