

# Author-Topic 모델 기반 대본 학습을 통한 비디오 등장 인물 인식

임병권<sup>○</sup> 허민오 장병탁

서울대학교 컴퓨터공학부

bklim@bi.snu.ac.kr, moheo@bi.snu.ac.kr, btzhang@bi.snu.ac.kr

## Recognition of Video Characters by Learning Dialogues Using Author-Topic Models

Byoung-Kwon Lim<sup>○</sup> Min-Oh Heo Byoung-Tak Zhang

School of Computer Science and Engineering

Seoul National University

### 요 약

기계학습 기술이 발달함에 따라 기계학습은 제한된 상황에서 벗어나, 실생활과 비슷한 복잡하고 다양한 상황에서의 학습이 중요한 이슈가 되었다. 본고에서는 현실과 비슷한 상황을 도입하기 위하여 드라마를 사용한다. 드라마 내의 등장인물들은 말투, 어조, 관심주제와 같이 다양한 특성을 내재하고 있다. 등장인물들의 다양한 특성 중 관심주제는 대본 안에 글로 드러나 있으므로 기계학습을 통해 등장인물의 인식에 활용할 수 있다. 최근, 확률그래프모델 분야에서 문서의 주제를 다루는 기법으로 자주 거론되는 토픽 모델 중 하나인 Author-Topic (AT) 모델은 등장인물의 관심주제를 학습하는 데에 적합하다. 본 논문에서는 AT 모델로 대본을 학습하고, 학습된 데이터 분포를 이용하여 장면에 등장하는 인물들을 인식하는 방법을 제시한다. 이 방법의 성능을 측정하기 위해, 미국 TV 드라마 'Friends' 대본 39편을 학습시키고, 장면에 대해 등장인물을 인식하는 실험을 수행하였다. 이 실험을 통해 본고에서 Author-Topic 모델을 이용한 인물 인식 방법이 다수의 인물이 참여한 담화의 인물들을 인식하는데 강점이 있음을 확인할 수 있다.

### 1. 서론

기계학습 기술이 발전함에 따라 기계학습을 좀 더 현실과 가까운 상황에서 학습하는 방법이 연구되고 있다. 최근에는 MMG 플랫폼을 이용하여, 제한된 환경보다 더 현실과 가까운 드라마를 이용하여 좀 더 확장된 기계학습 기술을 연구하고 있다[1]. 드라마 상황하의 등장인물들은 말투, 어조, 대화 주제, 행동 등에 다양한 특성을 내재하고 있다. 특히, 말투와 어조, 행동은 소리와 영상을 통해 주로 전달되지만, 등장인물의 관심 주제는 대본 내에 글의 형태로 드러나 있으므로 계산학적 언어처리 기술로 다룰 수 있으며, 이를 이용하여 인물인식 문제에 적용할 수 있다.

등장인물과 대사를 저자와 글로 치환하면 등장인물 인식 문제를 이제까지 많이 연구된 저자 인식 문제로 변환할 수 있다. 저자 인식 방법은 간단한 계산적 방법으로 출발하여 오래 전부터 연구되었다[2]. 연구가 계속되어 현재는 기계학습 방법도 연구되고 있으며 PCA를 이용한 요소 분석 방법이나, SVM을 이용한 방식의 분류기를 이용한 방법에 대한 연구를 확인할 수 있다[3,4]. 하지만 대부분의 기법들은 인물 개개인을

인식할 뿐 인물집단의 경우엔 인식하지 못한다. AT 모델을 이용하면 인물의 인물이 다수일 경우에도 인식이 가능하며, 성향을 데이터들의 분포로 표현 가능하다.

AT 모델은 LDA(Latent Dirichlet Allocation)의 변형된 모습이다. LDA는 많은 양의 문서를 학습하여, 각 문서의 토픽 분포와 각 토픽의 단어 분포를 알아낼 수 있다. AT 모델은 LDA와 비슷하지만 문서의 토픽분포가 아니라 저자의 토픽 분포와 토픽의 단어 분포를 알아낼 수 있다. 본 연구는 이렇게 토픽의 분포로 표현된 저자의 특징을 이용한다. AT 모델의 특성을 이용해 많은 데이터를 학습하여 저자와 논문의 상관관계를 연구한 사례가 있으며[5,6], AT 모델을 더 발전시켜 인간관계 네트워크를 분석한 연구 사례도 있다[7]. 하지만 위의 연구들과 달리 AT 모델을 데이터 분석에만 사용하는 것이 아니라, 드라마 대본을 통해 학습시킨 후, 학습된 데이터 분포를 이용하여 등장인물 인식 문제에 적용해 볼 것이다.

본 논문은 제시한 방법이 인물인식 문제를 잘 해결하는지 알아보기 위해 여러 편의 대본으로 학습된 AT 모델로 새로운 장면의 대사를 받았을 때 대사를 말한 다수의 등장인물들을 얼마나 정확히 찾아내는지

시험한다. 실제로 실험하기 위해 미국 드라마

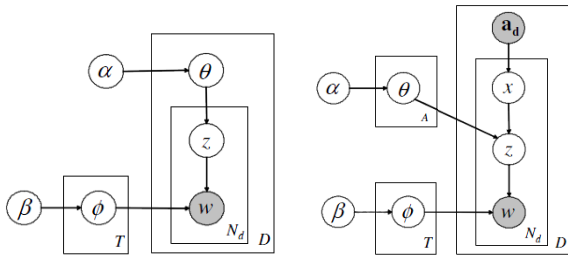


그림 1 LDA(좌)와 Author-topic 모델(우)

‘Friends’의 대본을 이용하였다. 이를 이용한 실험결과는 뛰어난 성능을 보여주지는 못하지만 특징적인 결과를 보여주며 이후에 성능을 향상시키기 위해 더 연구해야 할 방향을 제시하여 준다.

이후 2장에서는 AT 모델과 관련한 관련 연구를 소개하고, 3장에서는 본고에서 수행한 실험에 대하여 설명하며, 4장에서는 실험에 사용될 AT 모델에 대하여 설명한다. 5장에서는 디자인된 실험에 따른 실험 결과를 제시하고 마지막으로 6장에서는 실험 결과의 의미와 향후 방안에 대해 설명한다.

## 2. 문제 정의

대본 학습을 통한 등장인물 인식 문제는 아래와 같이 설명할 수 있다. 주어진 대본을 기계학습 기법을 사용하여 학습하고, 이렇게 학습된 기계학습 기법을 사용하여 인물명이 지워진 새로운 대본의 일부분이 주어졌을 때 그 부분에 참여한 인물들을 맞추는 것이다.

등장인물 인식 문제를 해결하기 위해 AT 모델을 사용하면서 AT 모델의 특성을 확인하기 위해 실험은 두 가지로 나누어 수행하였다. 첫 번째 실험은 학습 데이터에 포함되지 않은 한 부분의 장면의 대본에서 인물명을 모두 지운 상태로 학습된 모델이 장면에서 말하는 인물들을 정확히 인식하는지 테스트해 보는 것이다. 두 번째 실험은 첫 번째 실험의 비교 실험으로 인물명이 지워진 대본의 대사 한 줄이 있을 때 학습된 모델이 그 인물을 인식하는지 테스트해 보는 것이다. 또한 각 실험은 학습데이터의 양을 증가시켰을 때의 변화를 관찰하기 위해 시즌1만 학습시킬 때와 시즌1과 시즌2 모두 학습시킬 때로 나누어 실험하였다.

실험을 위해 미국 드라마 ‘Friends’의 시즌1 대본과 시즌2 대본을 사용하였다. 시즌1과 시즌2 대본은 총 41개의 편으로 구성되어 있으며, 41편의 대본 중 39편은 모델을 학습시키는데 사용할 것이고 2편은 나중에 성능을 측정할 때 사용할 것이다. 39편의 학습에 사용될 대본은 385부분의 장면으로 구성되어 있으며, 2편의 테스트에 사용될 대본은 26부분의 장면으로 구성되어 있다. 수집된 데이터는 전처리 과정을 거치는데, 주요 등장인물 6명을 제외한 다른 인물들은

대사가 너무 적어 실험에 적절하지 않아 삭제하였다.

## 3. 제안 방법

### 3.1 LDA

LDA의 기본적인 아이디어는 문서를 단어의 분포로 이루어져 있는 토픽들의 분포로 나타낸다는 것이며 그림 1 에서 모양을 확인할 수 있다[9,10]. 문서를 토픽의 분포로 나타내기 위해 많은 방법들이 개발되어 있다. 많은 방법들 중에서 가장 많이 쓰이는 방법은 Gibbs sampling 기법이다[8]. Gibbs sampling은 프로그램으로 구현하기가 매우 쉽기 때문에 LDA를 사용하는 경우의 대부분은 이 방법을 사용한다. 주어진 데이터에 LDA를 적용하면 문서와 토픽의 분포, 토픽과 단어의 분포를 얻을 수 있다.

### 3.2 Author-topic 모델

AT 모델은 토픽 모델의 변형이기 때문에 토픽 모델과 매우 유사하며 그림 1 에서 확인할 수 있다. 토픽 모델이 문서의 토픽 분포를 알아내는데 반해 AT 모델은 저자와 토픽의 분포를 알아내는데 있다. AT 모델 또한 Gibbs sampling을 사용하여 분포를 추정할 수 있는데 그렇게 추정된 분포는 식(1)과 같이 표현한다.

$$P(z_i = j, x_i = k | w_i = m, \mathbf{z}_{-i}, \mathbf{x}_{-i}, \mathbf{w}_{-i}) \propto \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta} \frac{C_{kj}^{AT} + \alpha}{\sum_{j'} C_{kj'}^{AT} + T\alpha} \quad (1)$$

$x_i = k$  는  $i$  번째 단어가  $k$  번째 저자라는 것을 나타내며,  $C_{kj}^{AT}$  는  $k$  번째 저자에  $j$  번째 토픽이 나온 횟수를 뜻한다. 위의 식을 통해 우리는 저자와 토픽의 분포인  $\theta$ 와 토픽과 단어의 분포인  $\phi$ 를 알 수 있다.

$$\phi_{mj} = \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta} \quad (2)$$

$$\theta_{kj} = \frac{C_{kj}^{AT} + \alpha}{\sum_{j'} C_{kj'}^{AT} + T\alpha}$$

본 논문에서는 등장인물들을 AT 모델의 일반적인 저자로 간주하고, 한 장면에서 등장인물의 대사를 전체 모은 것을 하나의 문서로 간주하여 실험한다.

### 3.3 추론 방법

본 논문에서는 학습된 모델을 사용하여 새로운

문서가 주어졌을 때 그 문서의 저자를 인식하기 위해 Bayes' theorem을 사용하여 각 저자의 확률을 계산한다. 이 얻어진 확률을 통해 새로운 문서의 저자를 인식할 수 있다.

$$P(a_d | w_{1:N}) = \frac{p(w_{1:N} | a_d)p(a_d)}{p(w_{1:N})} = \frac{p(w_{1:N} | a_d)p(a_d)}{\sum_{a_d} p(w_{1:N} | a_d)p(a_d)} \quad (3)$$

Bayes' theorem을 사용하면 위의 식을 구할 수 있다. 위의 식의  $P(w_{1:N} | a_d)$ 는 sum rule을 사용하면 식(6)과 같이 표현된다.

$$P(w_{1:N} | a_d) = \frac{\sum_{z,x} p(w_{1:N}, z_{1:N}, x_{1:N}, a_d)}{p(a_d)} \quad (4)$$

$$= \frac{\sum_{z,x} p(a_d) \prod_{i=1}^N p(x_i | a_d) p(z_i | x_i) p(w_i | z_i)}{p(a_d)}$$

식(6)을 식(5)에 대입하면 식(7)을 구할 수 있다.

$$P(a_d | w_{1:N}) = \frac{\left( \sum_{z,x} p(a_d) \prod_{i=1}^N p(x_i | a_d) p(z_i | x_i) p(w_i | z_i) \right) p(a_d)}{\sum_{a_d} \left( \sum_{z,x} p(a_d) \prod_{i=1}^N p(x_i | a_d) p(z_i | x_i) p(w_i | z_i) \right) p(a_d)} \quad (5)$$

우리는 식(7)을 통해 각  $a_d$ 의 확률을 구할 수 있으며 각  $a_d$  중 가장 높은 확률을 가지는 저자 집합을 고를 것이다.

$$a_d = \arg \max(p(a_d | w_{1:N})) \quad (6)$$

본 논문에서는 새로운 문서의 저자 집합뿐만 아니라 문서의 일부분의 저자를 인식하는 문제도 함께 다룰 것이다. 문서가 주어졌을 때 저자의 확률은 아래와 같이 표현할 수 있다.

$$P(x | w_{1:N}) = \frac{p(w_{1:N} | x)p(x)}{p(w_{1:N})} = \frac{p(w_{1:N} | x)p(x)}{\sum_x p(w_{1:N} | x)p(x)} \quad (7)$$

식(9)의  $P(w_{1:N} | x)$ 는 식(10)과 같이 나타낼 수 있다.

$$P(w_{1:N} | x) = \frac{\sum_z p(w_{1:N}, z_{1:N}, x)}{p(a_d)} \quad (8)$$

$$= \frac{\sum_x p(x) \prod_{i=1}^N p(z_i | x_i) p(w_i | z_i)}{p(x)}$$

식(10)를 식(9)에 대입하면 아래와 같이 나타낼 수 있다.

$$p(x | w_{1:N}) = \frac{\left( \sum_z \prod_{i=1}^N p(z_i | x_i) p(w_i | z_i) \right) p(x)}{\sum_x \left( \sum_z \prod_{i=1}^N p(z_i | x_i) p(w_i | z_i) \right) p(x)} \quad (9)$$

식(10)을 통해 우리는 각 저자의 확률을 구할 수 있으며, 가장 높은 확률을 가진 저자를 고를 것이다.

$$x = \arg \max p(x | w_{1:N}) \quad (10)$$

#### 4. 실험 결과

AT 모델은 parameter에 따라 분포결과가 달라지기에 높은 성능을 내는 parameter를 찾기 위해 parameter를 바꿔가며 실험하였다. 실험이 두 가지여서 양쪽 모두 높은 성능을 내는 모델을 찾기 위해 아래의 식을 사용하여 결과를 점수화 하였다.

$$P = 10 \times R_1 + R_2 \quad (11)$$

$R_1$ 은 첫 번째 실험에서 맞춘 결과의 개수를 나타내며,

$R_2$ 는 두 번째 실험에서 맞춘 결과의 개수를 나타낸다.

첫 번째 실험에 두 번째 실험보다 더 많은 가중치를 준 것은 첫 번째 실험이 두 번째 실험보다 더 정답을 맞추기 힘들기 때문이다. 학습데이터로 시즌1만 사용한 경우 토픽의 개수가 80개, 반복 횟수가 50번, hyper-parameter  $\beta$ 는 0.06일 때  $P$ 가 가장 높았다.

학습데이터로 시즌1과 시즌2 모두 사용하였을 때는 토픽의 개수 20개 반복 횟수 50번 hyper-parameter  $\beta$ 는 0.01일 때  $P$ 가 가장 높았다. 각 실험의 결과는

아래의 그림 2에서 확인할 수 있다.

데이터의 양이 늘었을 때의 성능향상을 확인하기 위해 실험 결과에서 시즌1 데이터만 사용했을 때와 시즌1과 시즌2 데이터를 모두 사용하였을 때를 비교하여 보았다. 실험1은 적중률이 0.2692에서 0.3077로 올랐고 실험2는 적중률이 0.2073에서 0.2415로 올랐다.

AT 모델의 성능은 무작위로 정답을 선택했을 때와 비교해 보면 얼마만큼의 성능 향상을 가져왔는지 더 알 수 있다. 실험1의 경우 정답은 6명의 인물들의 조합

중 하나이다. 이것은 정답의 개수가 63개임을 뜻하며 확률로 따지면 0.0159이다. AT 모델을 사용한 경우의 결과는 0.3077로 성능에 매우 많이 차이가 난다. 실험2의 정답은 6명의 인물들 중 하나이므로 무작위로 정답을 선택했을 시 0.1667의 확률로 답을 맞춘다. 학습된 모델을 사용하면 0.2415의 적중률을 나타내므로

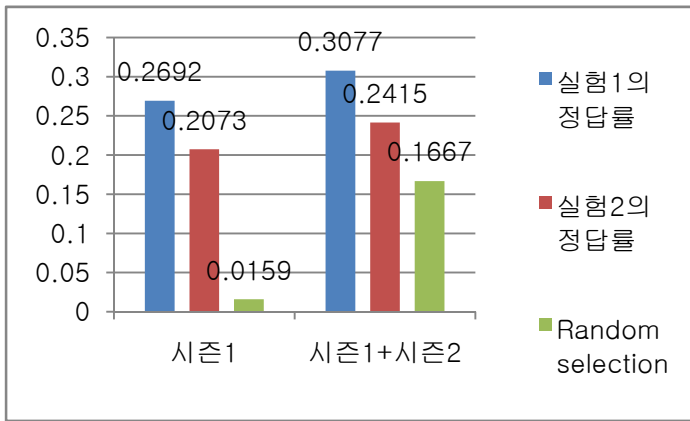


그림 2 실험 결과

실험1만큼은 아니지만 성능향상을 확인할 수 있다. 실험에 따른 성능향상 결과가 다른 이유는 이 실험에서 설계한 AT 모델은 실험1을 염두하고 설계한 모델이며 또한 문제로 들어온 데이터가 실험1에 비해 실험2는 크기가 작기 때문일 것이다.

### 5. 토의

AT 모델을 이용한 인물 인식 실험은 다수의 인물이 생성한 담화가 주어지더라도 참여한 인물들을 인식하는데 강점을 보이는 것을 알 수 있다. 그 이유는 AT 모델은 데이터 학습 과정에서부터 데이터가 복수의 인물로부터 생성되는 것을 가정하기 때문일 것이다.

본고에서 사용한 AT 모델을 이용한 인물 인식 방법은 본 실험의 예처럼 글을 이용한 분야뿐만이 아니라, 음성이나 영상의 분야에도 적용될 수 있으므로 다른 분야에 응용이 가능하다.

아직 실제로 이용할 수 있을 정도의 성능을 관찰할 수는 없었지만, 더 많은 양의 양질의 데이터를 학습에 사용한다면 더 좋은 성능을 얻을 수 있을 것이다.

### 감사의 글

이 논문은 교육과학기술부의 재원으로 국가연구재단의 지원을 받아 수행된 연구(0421-20110032, 지능형 추천 서비스를 위한 인지기반 기계학습 및 추천기술, Videome)이며, 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발) 및 교육과학기술부의 BK21-IT사업에 의해 일부 지원되었음.

### 참고 문헌

[1] Zhang, B.-T. Cognitive learning and the multimodal memory game: Toward human-level machine learning. *IEEE International Joint Conference on Neural Networks*, pp.3261-3267,

2008.  
 [2] T. Mendenhall. The characteristic curves of composition. *Science*, pp.237-249, 1887.  
 [3] S. Argamon, M. Šarić, and S. S. Stein. Style mining of electronic messages for multiple author discrimination. *In Proc. ACM Conference on Knowledge Discovery and Data Mining*, 2003.  
 [4] J. Burrows. Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method. *Clarendon Press, Oxford*, 1987.  
 [5] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, The author-topic model for authors and documents, *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pp.487-494, 2004.  
 [6] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, Probabilistic author-topic models for information discovery, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.306-315, 2004.  
 [7] A. McCallum, A. Corrada-Emmanuel, and X. Wang, The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks: Experiments with Enron and Academic Email, Technical report, UMass CS, 2004.  
 [8] T. L. Griffiths and M. Steyvers, Finding scientific topics, *Proceedings of the National Academy of Sciences of the United States of America*, pp.5228-5235, 2004.  
 [9] M. Steyvers and T. Griffiths, Probabilistic topic models, *Latent Semantic Analysis: A Road to Meaning*, 2006.  
 [10] D. M. Blei, A. Y. Ng, and M. J. Jordan, Latent dirichlet allocation, *Journal of Machine Learning Research*, vol.3, pp.993-1022, 2003.