

속성기반 감성분석의 연구 동향 및 방향성

조선희[○] 최동희 강재우[†]

고려대학교

{ehcho8564, choidonghee, kangj}@korea.ac.kr

Issues on Aspect-based Sentiment Analysis: A Survey

Seonhee Cho[○] Donghee Choi Jaewoo Kang[†]

Korea University

요 약

속성기반 감성분석(Asspect-based Sentiment Analysis)은 문자 데이터에서 특정 속성의 감성 극성에 대해 분석하는 연구 분야이다. 문장 단위로 분석했던 기존 감성분석과는 달리 더욱 세분화된 정보를 추출할 수 있다는 점에서 많은 관심을 받고 있다. 최근에는 속성기반 감성분석의 기존 벤치마크 데이터에서 높은 성능을 내는 것에 집중한 연구 뿐만 아니라, 기존 접근법 혹은 데이터셋의 한계점을 짚고 새로운 벤치마크 데이터셋을 제시하는 연구들도 등장하고 있다. 본 연구에서는 속성기반 감성분석 분야에서의 최신 연구동향을 살펴보고 SOTA(state-of-the-art) 모델과 새롭게 제시되는 데이터셋에 대해 살펴보고, 기존 모델들이 여러 데이터셋에서 어떠한 성능을 보이는지 결과를 분석해보고자 한다.

1. 서 론

최근, 인터넷과 소셜미디어의 사용이 활발해지면서 글에 드러나는 작성자의 주관적인 태도와 감성을 분석하는 감성 분석(sentiment analysis)이 많은 관심을 받고 있다. 그 중 속성기반 감성분석(aspect-based sentiment analysis)은 더욱 세분화된 감성 분석으로 문장에서 언급된 특정 속성에 대한 감성을 파악하는 연구이다. 예컨대, ‘배우들은 조금 아쉬웠지만, 만족스러운 영화!’라는 문장이 있다면 ‘배우’와 ‘영화’라는 두 가지 속성에 대해서, 각각 ‘부정’, ‘긍정’이라고 평가할 수 있을 것이다. 본 논문에서는 속성기반 감성분석 연구의 최신 동향을 살펴보고, 기존 SOTA(state-of-the-art)모델들이 여러 새로운 벤치마크 데이터에서 어떠한 경향성을 띄고, 얼마나 좋은 성능을 보이는지 살펴보고자 한다.¹

2. 연구 동향

속성기반 감성분석에서의 최근 연구 동향은 다음과 같이 두 가지로 나뉜다. 1) 기존 벤치마크 데이터셋[1]에서 높은 성능을 보이는 모델 기반 연구와 2) 기존 데이터셋이나 문제 정의의 한계점을 분석하고 그 문제를 다룬 새로운 데이터셋을 제시하는 연구이다. 2장에서는 각 흐름에 맞는 최신 연구들에 대하여

살펴보고, 이후 3장에서 각 SOTA 모델들의 성능을 실험한 결과를 알아볼 것이다.

2.1 모델 기반 연구

초반의 속성기반 감성분석 연구에는 다양한 머신러닝 기반 방법론들이 많이 사용되었는데, 나이브 베이즈, 로지스틱 회귀, 결정트리, 서포트 벡터머신 등이 대표적이다. 그러나 딥러닝 기반의 인공신경망 모델들이 더욱 좋은 성능을 내기 시작하면서, 위와 같은 방법들은 최근 학계에서는 잘 쓰이지 않고 있다.

딥러닝 기반 모델들은 태스크에 맞는 특징을 모델 설계자가 설정하는 것이 아닌, 데이터 그 자체를 통해 좋은 성능을 내는 점에서 각광받고 있다. 순환신경망(RNN) 기반 모델[2]들은 데이터를 순차적으로 입력 받기 때문에 어순이 중요한 자연어 데이터를 분석하는 것에 적합하고, 합성곱 신경망(CNN) 모델[3]은 더 적은 매개변수를 사용함에도 문장 구조를 잘 잡아낸다는 장점이 있다. 그래프 기반(GNN) 모델[4]은 문장 내 단어들의 의존 관계(dependency)를 그래프로 나타내고, 감성 분석에 그 정보를 사용하기 때문에 복잡한 통사 구조의 문장에 더욱 효과적이다. 최근에는 어텐션 매커니즘 기반의 모델[5][6][7]들이 가장 우수한 성능을 보인다. 어텐션 매커니즘은 문장 내 모든 단어 쌍의 관계성을 분석하기 때문에 멀리 떨어져 있는 단어들의 관계도 파악할 수 있다. 이와 더불어 신경망 모델과 어텐션 매커니즘을 함께 사용하는 연구[8][9]도 제시되었다.

BERT[10]와 같이 대규모 데이터를 이용해 미리 학습을 시킨 뒤 다운스트림 태스크에 적용하는 사전학습 언어모델(pretrained language model; PLM) 역시 활발하게 사용되고 있다. 일반 코퍼스 혹은 다른 도메인의 큰 코퍼스로 미리 학습시킨 BERT기반

¹ 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이며 (NRF-2020R1A2C3010638) 국가초고성능컴퓨팅센터로부터 초고성능컴퓨팅 자원과 기술지원을 받아 수행된 연구성과임 (KSC-2019-CRE-0186). 또한, 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT명품 인재양성사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01819).

[†] 교신저자

모델[11]들과, 감성 분석 분야의 코퍼스로 사전학습을 시킨 언어모델 연구[12]들이 제시되었다

2.2 데이터셋 기반 연구

이처럼 속성기반 감성분석 연구에서 여러 딥러닝 모델들이 아주 높은 성능을 내기 시작하면서, 모델의 성능을 더 높이기 보다는 기존 문제 정의, 접근법의 문제점을 찾고 그를 해결한 새로운 벤치마크 데이터를 제공하는 연구 흐름이 등장하였다.

[13]은 기존 벤치마크 데이터에서 대부분의 예제들이 오직 하나의 속성만을 포함하거나, 속성들이 모두 같은 감성을 띠는 점에서 사실상 문장 단위의 감성분석과 크게 다르지 않다는 점을 지적하였다. 더욱 정교한 속성기반 감성 분석을 위하여 [13]은 모든 예제가 두 개 이상의, 서로 다른 감성을 띄는 속성을 포함하도록 한 새로운 MAMS 데이터셋을 제시하였다. 이 때 기존 벤치마크 데이터인 SemEval-2014 restaurant review 데이터셋과 같은 코퍼스에서 추출하여 데이터의 분포를 동일하게 맞춰주었다.

[14]에서도 “aspect robustness”라고 명명한 비슷한 문제를 지적하였다. 이는 한 문장 안에 여러 속성들이 있을 때, 타겟 속성과 그 외의 속성들의 감성 극성이 다른 경우, 기존 문장에서 타겟 속성의 감성이 달라진 경우 등에서도 모델이 타겟 속성에 대한 분석을 잘 할 수 있는지를 의미한다. 단순히 문장 내에 다수의 속성이 포함되는 것에만 집중한 MAMS와는 다르게, [14]는 각 속성들이 서로 다른 감성 극성을 띄는지도 확인하였다.

[15]에서는 기존 감성분석 연구들이 한정적인 데이터만을 사용한다는 점을 지적하였다. 보다 넓은 범위에서의 데이터를 학습, 평가에 사용하기 위하여 Yelp, Amazon, SST[16], Opinions[17]의 데이터를 속성기반 감성분석에 사용할 수 있도록 가공한 YASO 데이터셋을 제공하였다. 이 때 데이터의 출처에 따라서 도메인이 다른데, YASO 데이터 내에서 도메인에 대한 정보는 알 수 없어 인도메인(in-domain) 성능이 아니라, 새로운 도메인(cross-domain)에서의 성능을 측정할 수 있다는 장점이 있다.

그 외에 속성기반 감성분석에 대한 새로운 문제 정의 주창한 연구들도 있다. [18]에서는 특정 속성에 대한 감성이 어떻게 정해지는지 파악하는 것도 중요함을 지적하였다. 따라서 문장 내 속성에 대하여, 그에 대한 의견을 나타내는 감성어(opinion term)를 찾아야 하는 TOWE라는 데이터셋을 제시하였다. 나아가, [19]에서는 TOWE와 SemEval 데이터를 모두 사용하여 문장에서 (속성, 감성어, 감성)으로 구성된 쌍을 추출해야 완전한 속성기반 감성분석이 가능하다고 주장하였다.

3. 실험

지금까지 속성기반 감성분석 분야의 SOTA 모델들과 새로이 제시된 데이터셋에 대해 소개하였다. 3장에서는 SOTA 모델들이 각 데이터셋에서 어떠한 성능을

보이는지를 정리하였다.

3.1 실험 환경

베이스라인 모델은 앞서 제시하였던 [2], [3], [4], [5], [6], [7], [8], [10], [11]를 사용하였고, 성능을 측정할 데이터로는 [1], [13], [14] 데이터셋을 사용하였다. YASO 데이터[15]의 경우, 아직 해당 데이터를 활용한 실험이 많지 않을 뿐 아니라, 인도메인 성능을 측정하기 애매하고, Twitter 데이터의 경우 데이터의 분포가 여타 데이터와 다르다는 점에서 제외하였다. 성능이 기재된 논문은 성능 값의 위 첨자로 표시하였다. #는 [11]을, +는 [13]을, *는 [14]를 참조하였음을 의미한다. 따라서 정확한 실험 세팅은 해당 논문을 참고하기 바란다. 여러 논문에서 서로 다른 성능이 제시되었을 때, 참조한 모든 연구에서의 성능을 평균낸 값으로 기재하였다. 메트릭은 정확도(accuracy) 기준이며, 각 실험 결과는 표 1에 기재하였다.

표 1. 속성기반 감성분석 실험 결과

모델	SemEval 2014		MAMS	ARTS	
	Laptop	Rest.		Laptop	Rest.
[2]	68.03*	78.88**	76.16**	22.57*	30.18*
[3]	65.67*	77.12**	71.81**	10.34*	13.12*
[4]	72.41*	77.86*	76.95*	19.91*	24.73*
[5]	64.42*	78.07**	67.54**	16.93*	21.52*
[6]	-	80.98*	66.72*	-	-
[7]	-	78.6*	76.6*	-	-
[8]	67.55*	75.98*	71.10**	9.87*	14.64*
[10]	76.44**	80.96***	82.22*	50.94*	54.82*
[11]	78.3***	85.83***	85.10*	53.29*	59.29*

4. 실험 결과

기존 벤치마크 데이터인 SemEval에서의 성능과 비교하였을 때, 대부분의 모델이 ARTS, YASO에서 더 낮은 성능을 보였다. 그러나 SemEval에서보다 MAMS에서 더 높은 성능을 보이는 경우가 다수 있는데, 그럼에도 ARTS에서는 상당히 낮은 성능을 보인다. 이는 MAMS에서 SemEval로 학습한 모델들이 각 속성에 대한 분석을 이용하기보다 문장의 전반적인 감성이나 여러 속성이 공유하는 감성들에 치우치는 추론을 하는 것으로 유추할 수 있다. 즉, [14]에서 제시한 문제가 실제 실험 결과에서도 드러난 것이다.

BERT를 사용하지 않은 모델은 YASO에서 더 낮은 성능을 보이는 반면, BERT 기반의 모델들은 준수한, 혹은 더 높은 성능을 보인다. 이는 PLM이 큰 코퍼스로 학습하는 과정에서 다른 도메인에 대한 지식을 익히는 것이 가능하고, 때문에 cross-domain 환경에서도 충분히 좋은 성능을 보이는 것이라 해석할 수 있다.

또한, 여러 베이스라인 모델들의 성능을 직접 실험하여 도출한 논문들이 일부 있는데, 같은 모델과

같은 데이터셋임에도 세팅에 따라 상당한 차이의 성능을 발표하는 경우가 종종 있었다. 따라서 여러 데이터셋에서의 각 모델들의 성능을 비교하고자 한다면 현재까지 발표된 성능들을 모두 모아 평균내어 기재한 표 1의 결과가 참고할만한 자료가 될 수 있을 것이다.

5. 결 론

본 논문에서는 속성기반 감성분석 분야의 최근 연구 경향과 기존 SOTA모델들이 새로운 벤치마크 데이터에서 어떠한 성능을 보이는지에 대해 알아보았다. 최근 연구 경향은 크게 1) 기존 데이터에서의 성능을 높이는 것에 집중한 연구와 2) 기존 연구 방식의 한계점을 짚고 새로운 데이터셋과 연구 방법을 제시하는 연구들로 나뉜다. 1)에 치중했던 기존 흐름과는 다른 새로운 패러다임이 제시됐고, 기존 벤치마크 데이터에서 우수한 성능을 보인 모델들이 2)에서 제시한 문제들에 대해서는 잘 작동하지 않았다는 것은 주목할 만하다. 향후 연구들에서 이러한 문제를 어떻게 해결할지가 기대되는 바이다.

참고 문헌

- [1] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Eryigit, G. "SemEval-2014 Task 4: Aspect Based Sentiment Analysis." In International workshop on semantic evaluation. 2014
- [2] Tang, D., Qin, B., Feng, X., & Liu, T. (2015). Effective LSTMs for target-dependent sentiment classification. *arXiv preprint arXiv:1512.01100*.
- [3] Xue, W., & Li, T. Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*. 2018.
- [4] Zhang, C., Li, Q., & Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477*. 2019.
- [5] Tang, D., Qin, B., & Liu, T. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*. 2016.
- [6] Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*. 2019.
- [7] Ma, D., Li, S., Zhang, X., & Wang, H. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*. 2017.
- [8] Wang, Y., Huang, M., Zhu, X., & Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing. 2016.
- [9] Huang, B., Ou, Y., & Carley, K. M. Aspect level sentiment classification with attention-over-attention neural networks. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. 2018.
- [10] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [11] Xu, H., Liu, B., Shu, L., & Yu, P. S. BERT post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*. 2019.
- [12] Ke, P., Ji, H., Liu, S., Zhu, X., & Huang, M. Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis. EMNLP. 2020.
- [13] Jiang, Q., Chen, L., Xu, R., Ao, X., & Yang, M. A challenge dataset and effective models for aspect-based sentiment analysis. In Proceedings of the 2019 EMNLP-IJCNLP. 2019.
- [14] Xing, X., Jin, Z., Jin, D., Wang, B., Zhang, Q., & Huang, X. Tasty Burgers, Soggy Fries: Probing Aspect Robustness in Aspect-Based Sentiment Analysis. *arXiv preprint arXiv:2009.07964*. 2020.
- [15] Orbach, M., Toledo-Ronen, O., Spector, A., Aharonov, R., Katz, Y., & Slonim, N. YASO: A New Benchmark for Targeted Sentiment Analysis. *arXiv preprint arXiv:2012.14541*. 2020.
- [16] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.
- [17] Ganesan, K., Zhai, C., & Han, J. Opinosis: A graph based approach to abstractive summarization of highly redundant opinions. 2010.
- [18] Fan, Z., Wu, Z., Dai, X., Huang, S., & Chen, J. Target-oriented opinion words extraction with target-fused neural sequence labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, pp. 2509-2518. 2019.
- [19] Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., & Si, L. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 05, pp. 8600-8607. 2020.