A Robust Framework for Targeted Sentiment Analysis

CP3106

Xiang Pan

Department of Computer Science
National University of Singapore

22 April 2020

Outline

Abstract

1. Differences between Sentiment Analysis and Targeted Sentiment Analysis(TSA)
2. The problems of existed TSA model.
3. Generic Framework for TSA problem.
   * Auxiliary training methods
4. Robust Test Dataset for TSA
   * Adversarial Training
   * Data Augmentation

Sentiment Analysis

Sentiment Analysis:Process of finding out extracting experiences and emotions from the given dataset.

Different Tasks of Sentiment Analysis

Task	Dataset	Coherence	Source	Collection	Target Structure	Example application domain
TG-ABSA	SemEval 2014	Strong	Online Review	Crawling	Aspect (Entity)	Product, service, movie, Apps
TN-ABSA	Twitter	Weak	Twitter	Filtering	Entity	Event, people, organization
T-ABSA	Sentihood, Baby Care	Moderate	Forum	Crawling	(Entity, Aspect)	product, service

Figure: Different Task of Sentiment Analysis

Aspect Based Sentiment Analysis

Subtask of ABSA

1. Subtask 1:Aspect term extraction
2. Subtask 2:Aspect term polarity
3. Subtask 3:Aspect category detection
4. Subtask 4:Aspect category polarity

The subtask 2, aspect term can be seen as a target, which can be formatted to targeted sentiment analysis.[**?**]

Targeted Sentiment Analysis

Task Setting

"I loved their fajitas" $\rightarrow$ fajitas: positive
"I hated their fajitas, but their salads were great" $\rightarrow$ fajitas: negative, salads: positive
"The fajitas are their first plate" $\rightarrow$ fajitas: neutral
"The fajitas were great to taste, but not to see" $\rightarrow$ fajitas: conflict

Formal Definition

$$Sentence = \{w_1, w_2, w_3, w_4, ..[t_1, t_2, ..., t_m]...., w_n\} \tag{1}$$

$$Target\ Text = \{t_1, t_2, ..., t_m\} \tag{2}$$

$$polarity = \{positive, neutral, negative\} \tag{3}$$

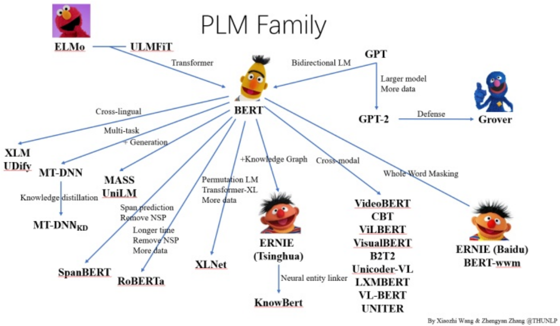

Figure: Pre-trained Language Models Family

Text Classification

BERT-Based Methods

To utilize bert, we have several direct ways:

1. Using BERT embeddings as the input of a sequence
2. Fine-tuned BERT by [CLS] classification token.

How to fine-tune BERT for text classification.[?]

1. Various fine tuning methods
   1.1 Within-task pre-training
   1.2 In-domain pre-training
   1.3 Cross-domain pre-training
2. Different learning rates are used for different layers of BERT

Table: Previous Work on three typical datasets

	Models	Twitter		Restaurant		Laptop	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
RNN baselines	TD-LSTM	0.7080	0.6900	0.7563	-	0.6813	-
	ATAE-LSTM	-	-	0.7720	-	0.6870	-
	IAN	-	-	0.7860	-	0.7210	-
	RAM	0.6936	0.6730	0.8023	0.7080	0.7449	0.7135
Non-RNN baselines	Feature-based SVM	0.6340	0.6330	0.8016	-	0.7049	-
	Rec-NN	0.6630	0.6590	-	-	-	-
	MemNet	0.6850	0.6691	0.7816	0.6583	0.7033	0.6409
AEN-BERT	AEN-GloVe	0.7283	0.6981	0.8098	0.7214	0.7351	0.6904
	BERT-SPC	0.7355	0.7214	0.8446	0.7698	0.7899	0.7503
	AEN-BERT	0.7471	0.7313	0.8312	0.7376	0.7993	0.7631

Previous Work Problems

1. Modeling the data from different aspects, which makes the model not general enough for further improvements.
2. Targeted-Dependent

Framework

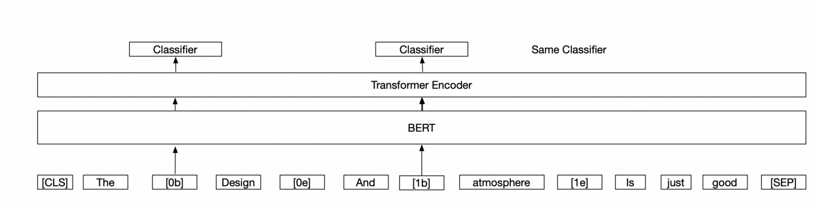

Figure: Our framework for TSA

Auxiliary Training

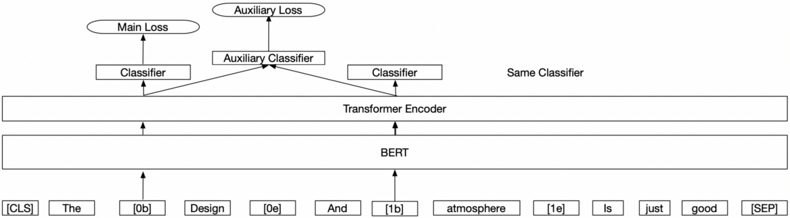

Figure: auxiliary training

Visualization

There BERT may have some dependency on the target tokens.

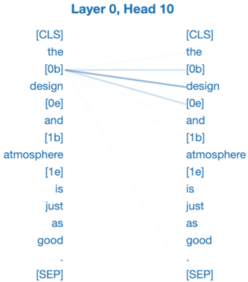

Figure: visualization of the BERT Bottom layer

Robust Test

For explicitly test the robustness of model towards unseen target, we proposed the robustness dataset.

1. Removing the test samples with the same or similar target in training set.
2. Re-split the data set.

Adversarial Training

$$r_{adv} = -\epsilon \frac{g}{||g||_2} \tag{4}$$

$$g = \nabla_x \log \ p(y|x; \hat{\theta}) \tag{5}$$

and $p_{adv} = \epsilon$ is the size of the perturbations.

$$loss_{aux} = -\log p(y|x + r_{adv}; \theta) \tag{6}$$

The total training loss is:

$$loss(main\ target) = loss_{main}(begin\_token) + p_{aux} * loss_{aux}(main\ target) + loss_{adv} \tag{7}$$

Data Augmentation

Replace the target words using wordnet synonyms.

$augment(target\ words) = synonyms$

For example:

augment('laptop')='laptop computer'

augment('staff')='faculty'

Table: Test results on three typical data sets.

	Models	Twitter		Restaurant		Laptop	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
RNN baselines	TD-LSTM	0.7080	0.6900	0.7563	-	0.6813	-
	ATAE-LSTM	-	-	0.7720	-	0.6870	-
	IAN	-	-	0.7860	-	0.7210	-
	RAM	0.6936	0.6730	0.8023	0.7080	0.7449	0.7135
Non-RNN baselines	Feature-based SVM	0.6340	0.6330	0.8016	-	0.7049	-
	Rec-NN	0.6630	0.6590	-	-	-	-
	MemNet	0.6850	0.6691	0.7816	0.6583	0.7033	0.6409
AEN-BERT	BERT	0.7464	0.7304	0.8128	0.6979	0.7654	0.7283
	BERT-SPC	0.7355	0.7214	0.8446	0.7698	0.7899	0.7503
	AEN-BERT	0.7471	0.7313	0.8312	0.7376	0.7993	0.7631
BERT-PT	BERT-PT	-	-	0.8495	0.7696	0.7807	0.7508
TD-BERT	TD-BERT	0.7669	0.7428	0.8510	0.7835	0.7807	0.7508
	TD-BERT-QA-CON	0.7731	0.7440	0.8456	0.7961	0.7842	0.7437
Our	Framework	0.7673	0.7451	0.843	0.780	0.7633	0.7291
	Framework+aux1.0	-	-	**0.8554**	**0.7962**	**0.7806**	**0.7523**
	Framework+aux0.1	-	-	0.8464	0.788	0.7759	0.7370

Table: Test results on three domain bert

		Twitter		Restaurant		Laptop	
Domain	Method	acc	f1	acc	f1	acc	f1
joint	BERT-ADA Joint	-	-	0.8635	0.7889	0.7896	0.7418
rest	BERT-ADA Rest	-	-	0.8714	0.8005	0.7860	0.7409
lapt	BERT-ADA Lapt	-	-	0.8551	0.7809	0.7919	0.7418
	Framework	0.7673	0.7451	0.843	0.7808	0.7633	0.7291
	aux 1	-	-	0.8554	0.7962	0.7806	0.7523
	aux 0.1	-	-	0.8464	0.7883	0.7759	0.737
joint	Framework	-	-	0.8491	0.7697	0.7821	0.7421
	aux 1	-	-	0.8768	0.8153	0.7978	0.7506
	aux 0.1	-	-	0.8625	0.8017	0.7915	0.745
rest	Framework	-	-	0.86911	0.8061	0.7931	0.7461
	aux 1	-	-	0.8786	0.8139	0.7978	0.7571
	aux 0.1	-	-	0.87589	0.81696	0.7978	0.7538
lapt	Framework	-	-	0.8562	0.7928	0.79	0.7477
	aux 1	-	-	0.8786	0.8139	0.8088	0.7627
	aux 0.1	-	-	0.87589	**0.81696**	0.8025	**0.7653**

Rmoving the Test samples with seen targets

For simplicity, we remove those test samples with the same or similar targets in the training set.

Table: Statistics of re-split dataset.

dataset	stat-type	train	test
twitter	size	6248	619
	target-number	104	358
restaurant	size	3608	393
	target-number	606	304
laptop	size	2328	282
	target-number	461	232

Table: Test results on three re-split(removing) data sets.

	Models	Twitter		Restaurant		Laptop	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
	BERT-SPC	0.7052	0.6898	0.7857	0.6570	0.7524	0.6875
Our	Framework	0.7197	0.7054	0.8420	0.7740	0.7696	0.7282
	Framework+adv1	0.7673	0.7513	0.8545	**0.7954**	0.7821	0.7420
	Framework+aux1	-	-	0.8420	0.7735	0.7774	0.7466
	Framework+adv1+aux1	-	-	0.8500	0.7742	0.7884	**0.7466**

Resplit the Datasets

For comparable to original datasets, we combine the training set and test set together. Then we re-split the datasets by the target.

Table: Test Results on the re-split datasets

Model	Methods	Twitters		Restaurants		Laptops	
		acc	f1	acc	f1	acc	f1
BERT-SPC	BERT-SPC	0.6162	0.6132	0.7551	0.6242	0.7247	0.683
	aug	0.6224	0.616	0.7645	0.6502	0.7154	0.645
Framework	Framework	0.6702	0.6583	0.7701	0.6426	0.736	0.7165
	aug	0.6778	0.6668	0.7757	0.6805	0.7566	0.7332
	aux1			0.7888	**0.7061**	0.7453	0.7309
	adv1	0.6759	**0.6709**	0.7925	0.6773	0.7566	0.7334
	aug+adv1	0.6479	0.6293	0.7888	0.6881	0.7753	**0.7574**
	aux1+adv1			0.7664	0.6809	0.7528	0.7318
	aug+aux1+adv1			0.7477	0.643	0.7303	0.7158

Conclusion

1. We propose a general framework and auxiliary training method for targeted sentiment analysis. Our results on TSA task is reach or achieve better performance on original datasets.
2. We propose a robust test dataset for unseen targets.
3. Our framework is less target-dependent and more robust towards unseen targets.
4. We implemented two methods to enhance the robustness based on our framework.

Future Work

1. Try to combine three methods for robustness.
2. Data Augmentation for auxiliary training.

Reference I

[1] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar.
SemEval-2014 task 4: Aspect based sentiment analysis.
In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August 2014. Association for Computational Linguistics.

[2] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang.
How to fine-tune bert for text classification?, 2019.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
arXiv:1810.04805 [cs], May 2019.

[4] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu.
ERNIE: Enhanced language representation with informative entities.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, Florence, Italy, July 2019. Association for Computational Linguistics.

[5] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu.
A Survey on Knowledge Graphs: Representation, Acquisition and Applications.
arXiv:2002.00388 [cs], February 2020.

[6] Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok.
SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings.
In *AAAI*, pages 1795–1802, 2018.

Reference II

[7] Yukun Ma, Haiyun Peng, and Erik Cambria.
Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm.
In *AAAI*, pages 5876–5883, 2018.

[8] Duyu Tang, Bing Qin, and Ting Liu.
Aspect Level Sentiment Classification with Deep Memory Network.
In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, 2016.

[9] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu.
Effective LSTMs for target-dependent sentiment classification.
In *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, pages 3298–3307, 2016.

[10] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao.
Attentional encoder network for targeted sentiment classification.
arXiv preprint arXiv:1902.09314, 2019.

[11] Pinlong Zhaoa, Linlin Houb, and Ou Wua.
Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification.
arXiv preprint arXiv:1906.04501, 2019.

[12] Zheng Li, Ying Wei, Yu Zhang, Xiang Zhang, Xin Li, and Qiang Yang.
Exploiting Coarse-to-Fine Task Transfer for Aspect-level Sentiment Classification.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4253—-4260, 2019.

Reference III

[13] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier.
Exploiting document knowledge for aspect-level sentiment classification.
In *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, volume 2, pages 579–585, 2018.

[14] Ruining He and Julian McAuley.
Ups and Downs.
In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, pages 507–517, New York, New York, USA, 2016. ACM Press.

[15] Sepp Hochreiter and Jürgen Schmidhuber.
Long short-term memory.
Neural computation, 9(8):1735–1780, 1997.

[16] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu.
BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis.
In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335. Association for Computational Linguistics, 2019.

[17] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao.
Targeted sentiment classification with attentional encoder network.
Lecture Notes in Computer Science, page 93–103, 2019.

Reference IV

[18] Jiaxin Pei, Aixin Sun, and Chenliang Li.
Targeted sentiment analysis: A data-driven categorization, 2019.

[19] Akbar Karimi, Leonardo Rossi, Andrea Prati, and Katharina Full.
Adversarial training for aspect-based sentiment analysis with bert, 2020.

[20] Takeru Miyato, Andrew M. Dai, and Ian Goodfellow.
Adversarial training methods for semi-supervised text classification, 2016.

[21] Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu.
Bert post-training for review reading comprehension and aspect-based sentiment analysis, 2019.

[22] Z. Gao, A. Feng, X. Song, and X. Wu.
Target-dependent sentiment classification with bert.
IEEE Access, 7:154290–154299, 2019.

[23] Margaret Mitchell, Jacqui Aguilar, Theresa Wilson, and Benjamin Van Durme.
Open domain targeted sentiment.
In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1643–1654, 2013.

[24] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski.
Matching the blanks: Distributional similarity for relation learning, 2019.

Reference V

[25] Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl.
Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification, 2019.