Aspect Term Extraction with History Attention and Selective Transformation[1]

Xin Li[1], Lidong Bing[2], Piji Li[1], Wai Lam[1], Zhimou Yang[3]
Presenter: Lin Ma[2]

[1]The Chinese University of Hong Kong

[2]Tencent AI Lab

[3]Northeastern University

IJCAI 2018

[1]Joint work with Tencent AI Lab

Outline

Outline

1. Aspect Term Extraction
   - What is "Aspect Term"?
   - Problem Formulation
2. The Proposed Model
   - Motivation
   - Our Model
3. Comparative Study
   - Baselines
   - Main Results
   - Effectiveness of "History Attention" and "Selective Transformation"

What is “Aspect Term”?

- **Definition**: Explicitly mentioned entities / product attributes in the review sentences where the users express their opinions.
  - Also called “Aspect Phrase” or “Opinion Target” in the existing works [4].

Examples

- *Its size is ideal and the weight is acceptable.*
- *The pizza is overpriced and soggy.*

Outline

Problem Formulation

- Aspect Term Extraction is to automatically extract the aspect term from user reviews.
- As a natural information extraction problem, it can be formulated as a sequence labeling problem or a token-level classification problem.

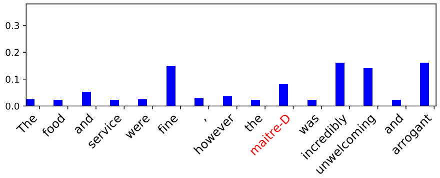
Examples

	I	love	the	operating	system	and	the	preloaded	software
O-T	O	O	O	T	T	O	O	T	T
B-I-O	O	O	O	B	I	O	O	B	I

Outline

Motivation

1. We still adopt the aspect-opinion joint modeling strategy [3, 5, 11, 12] in our model.
   - The existence of opinion (aspect) term can provide indicative clues for finding the collocated / correlated aspect (opinion) term.
2. Local attention and global soft attention have some limitations.
   - Local attention [3] can **NOT** capture the long term dependency between the aspect term and the opinion words.
     - Example: We ordered the special, grilled branzino, that was so infused with bone, it was difficult to eat .
   - Global soft attention [12] may introduce some **irrelevant information**.



Motivation

3. The previous predictions can help the current prediction to reduce the error space.
   - If the previous prediction is "O", then current prediction cannot be "I".
   - Some previously predicted commonly-used aspect terms can guide the model to find the co-occurred infrequent aspect terms.
     - Example: Apple is unmatched in product quality, aesthetics, craftmanship, and customer service.
     - If we know "product quality" is an aspect, then "aesthetics" and "craftmanship" which belong to the same co-ordinate structure with "product quality" are very likely to be aspect terms.

Outline

Model Overview

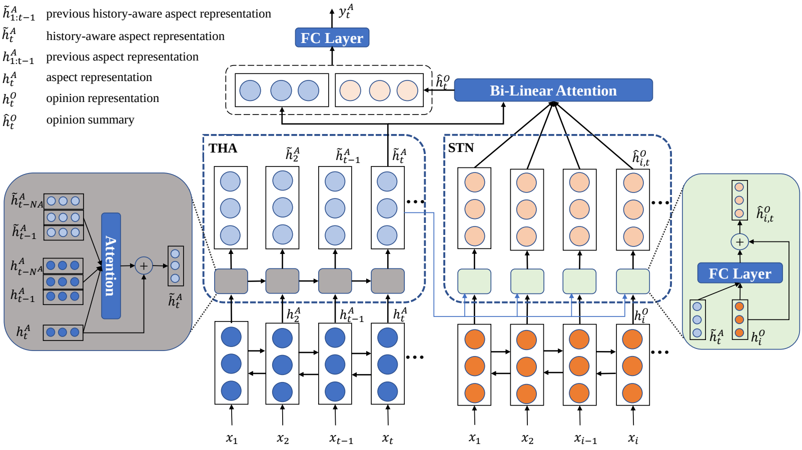


Figure: The proposed architecture for Aspect Term Extraction

Core components of the proposed model

- Long Short-Term Memory Networks (LSTMs)
  - Learning word-level representations.
- Truncated History Attention (THA) component
  - Explicitly modeling the aspect-aspect relation based on self-attention.
- Selective Transformation Networks (STN)
  - Making use of global opinion information without introducing too much noise.

Truncated History Attention (THA)

The primary goal of THA is to explicitly model the relation between the previous predictions and the current prediction.

- Adding more constraints on the current prediction.
  - E.g., if the previous hidden vector h_{t-1} was predicted as tag "O" then the current tag cannot be "I".
- Providing more information for the current predictions based on the collocated aspects.
  - Example: Apple is unmatched in product quality, aesthetics, craftmanship, and customer service.
  - Given the current input "aesthetics", modeling the relation between it and "product quality" implicitly captures the co-ordinate structure.

Truncated History Attention (THA)

Solutions provided By THA:

1. Calculate the association scores between the previous representations (h_i^A & $\tilde{h}_i^A$) and the current representation h_t^A (self-attention):

$$a_i^t = \mathbf{v}^\top \tanh(\mathbf{W}_1 h_i^A + \mathbf{W}_2 h_t^A + \mathbf{W}_3 \tilde{h}_i^A),$$

$$s_i^t = \mathrm{Softmax}(a_i^t).$$

2. Incorporate the aspect history $\hat{h}_t^A$ into the aspect representation h_t^A:

$$\hat{h}_t^A = \sum_{i=t-N^A}^{t-1} s_i^t \times \tilde{h}_i^A.$$

$$\tilde{h}_t^A = h_t^A + \mathrm{ReLU}(\hat{h}_t^A),$$

Selective Transformation Networks (STN)

This component tries to make use of the global information without introducing too much noises.

- Global soft attention [10]:
    1. Computing association scores between aspect and opinion representations
    2. Aggregating the opinion features based on association scores
- Local attention [3]:
    - Assume the aspect is close to its opinion modifier.
    - Only paying attention to a few surrounding words (i.e., opinion representations)

Selective Transformation Networks (STN)

- Our STN:
  - Capture long-term aspect-opinion dependency: make use of the global opinion information.
  - Reduce noises: add more constraint on the opinion representation h_i^O with current aspect representation $\tilde{h}_t^A$.
  - Refine opinion representations h_i^O: introduce a residual block [2] to combine the original and the transformed opinion representations.
  - The produced opinion features $\hat{h}_t^O$ is aspect-dependent or time-dependent.

$$\hat{h}_{i,t}^O = h_i^O + \mathrm{ReLU}(\mathbf{W}^A \tilde{h}_t^A + \mathbf{W}^O h_i^O),$$

$$w_{i,t} = \mathrm{Softmax}(\tanh(\tilde{h}_t^A \mathbf{W}_{bi} \hat{h}_{i,t}^O + \mathbf{b}_{bi})),$$

$$\hat{h}_t^O = \sum_{i=1}^{T} w_{i,t} \times \hat{h}_{i,t}^O.$$

Outline

Baselines

- CRF and Semi-CRF [7]
- SemEval ABSA winning systems [1, 9, 6, 8]
- LSTMs
- WDEmb [13]
- Memory Interaction Networks (MIN) [3]
- Recursive Neural Conditional Random Fields (RNCRF) [11]
- Coupled Multi-Layer Attention (CMLA) [12]

Outline

Main Results

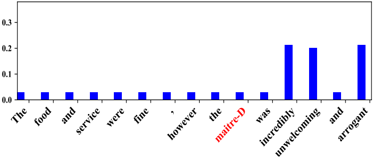
Models	D_1 (LAPTOP14)	D_2 (REST14)	D_3 (REST15)	D_4 (REST16)
CRF-1	72.77	79.72	62.67	66.96
CRF-2	74.01	82.33	67.54	69.56
Semi-CRF	68.75	79.60	62.69	66.35
LSTM	75.71	82.01	68.26	70.35
IHS_RD (D_1 **winner**)	74.55	79.62	-	-
DLIREC (D_2 **winner**)	73.78	84.01	-	-
EliXa (D_3 **winner**)	-	-	70.04	-
NLANGP (D_4 **winner**)	-	-	67.12	72.34
WDEmb (**IJCAI 2016**)	75.16	84.97	69.73	-
MIN (**EMNLP 2017**)	77.58	-	-	73.44
RNCRF (**EMNLP 2016**)	78.42	84.93	67.74$^\natural$	69.72*
CMLA (**AAAI 2017**)	77.80	85.29	70.73	72.77*
OURS w/o **THA**	77.64	84.30	70.89	72.62
OURS w/o **STN**	77.45	83.88	70.09	72.18
OURS w/o **THA** & **STN**	76.95	83.48	69.77	71.87
OURS	**79.52**	**85.61**	**71.46**	**73.61**

Table: Experimental results (F_1 score, %).

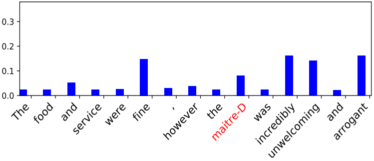
Outline

Effectiveness of "History Attention" and "Selective Transformation"

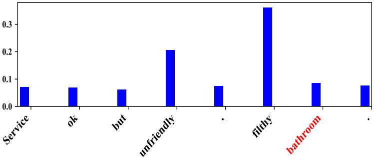
The generated attention scores of our model and our model w/o STN:



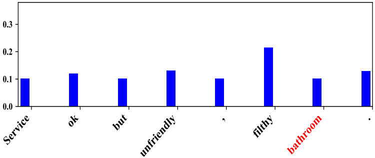
(a) OURS



(b) OURS w/o STN.



(c) OURS



(d) OURS w/o STN.

Effectiveness of "History Attention" and "Selective Transformation"

We also compare the output of our model and its variants:

Input sentences	Output of LSTM	Output of OURS w/o THA & STN	Output of OURS
1. *the device speaks about it self*	*device*	**NONE**	**NONE**
2. *Great survice !*	**NONE**	*survice*	*survice*
3. *Apple is unmatched in product quality, aesthetics, craftmanship, and custormer service*	*quality, aesthetics, custormer service*	*quality, customer service*	*product quality, aesthetics, craftmanship, custormer service*
4. *I am pleased with the fast log on, speedy WiFi connection and the long battery life*	*WiFi connection, battery life*	*log, WiFi connection, battery life*	*log on, WiFi connection, battery life*
5. *Also, I personally wasn't a fan of the portobello and asparagus mole*	*asparagus mole*	*asparagus mole*	*portobello and asparagus mole*

Table: The gold standard aspect terms are underlined and in red.

Summary

- In this paper, we design a convolution-based framework for Aspect Term Extraction, which achieves state-of-the-art results on four SemEval ABSA datasets.
- The proposed THA component explicitly models the aspect-aspect relation for more accurate extraction.
- The proposed STN component makes full use of the opinion information without introducing too much noises.

References:

[1] M. Chernyshevich. Ihs r&d belarus: Cross-domain extraction of product features using crf. In *Proc. of SemEval*, pages 309–313, 2014.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. of CVPR*, pages 770–778, 2016.

[3] X. Li and W. Lam. Deep multi-task learning for aspect term extraction with memory interaction. In *Proc. of EMNLP*, pages 2886–2892, 2017.

[4] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[5] G. Qiu, B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, 2011.

[6] I. n. San Vicente, X. Saralegi, and R. Agerri. Elixa: A modular and flexible absa platform. In *Proc. of SemEval*, pages 748–752, 2015.

[7] S. Sarawagi, W. W. Cohen, et al. Semi-markov conditional random fields for information extraction. In *Proc. of NIPS*, pages 1185–1192, 2004.

[8] Z. Toh and J. Su. Nlangp at semeval-2016 task 5: Improving aspect based sentiment analysis using neural network features. In *Proc. of SemEval*, pages 282–288, 2016.

[9] Z. Toh and W. Wang. Dlirec: Aspect term extraction and term polarity classification system. In *Proc. of SemEval*, pages 235–240, 2014.

[10] W. Wang, S. J. Pan, and D. Dahlmeier. Multi-task coupled attentions for category-specific aspect and opinion terms co-extraction. *arXiv preprint arXiv:1702.01776*, 2017b.

[11] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proc. of EMNLP*, pages 616–626, 2016.

[12] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proc. of AAAI*, pages 3316–3322, 2017.

[13] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang, and M. Zhou. Unsupervised word and dependency path embeddings for aspect term extraction. In *Proc. of IJCAI*, pages 2979–2985, 2016.