A Unified Model for Opinion Target Extraction and Target Sentiment Prediction[1]

Xin Li[1], Lidong Bing[2], Piji Li[3], Wai Lam[1]

[1]The Chinese University of Hong Kong

[2]Alibaba Damo Academy

[3]Tencent AI Lab

March 8, 2019

[1]Joint work with Tencent AI Lab

Outline

1. Aspect/Target-Based Sentiment Analysis
    - Task Introduction
    - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
    - Overview
    - Boundary Guided TBSA
    - Maintaining Sentiment Consistency
    - Auxiliary Target Word Detection
    - Training
3. Experiment
    - Settings
    - Results

Outline

1. Aspect/Target-Based Sentiment Analysis
   - Task Introduction
   - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
   - Overview
   - Boundary Guided TBSA
   - Maintaining Sentiment Consistency
   - Auxiliary Target Word Detection
   - Training
3. Experiment
   - Settings
   - Results

Task Introduction

Aspect/Target-Based Sentiment Analysis (ABSA or TBSA) usually involves two sub-tasks:

1. Opinion Target Extraction (OTE):
    - Extract the nouns or noun phrases where the users express their opinions from online reviews.
    - The extracted noun/noun phrase is called "Opinion Target" or "Aspect Term".
2. Target Sentiment Classification (TSC)
    - Predict sentiment over the pair of opinion target and sentence.
    - A typical text classification task with multi-source input (sentence + opinion target).

Example

- Great [**food**] but the [**service**] is dreadful.
    - **"food"** and **"service"** are the discussed opinion targets.
    - Given the target **"food"**, the sentiment polarity of the sentence is positive while if the input target is **"service"**, it becomes negative.

Outline

1. Aspect/Target-Based Sentiment Analysis
   - Task Introduction
   - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
   - Overview
   - Boundary Guided TBSA
   - Maintaining Sentiment Consistency
   - Auxiliary Target Word Detection
   - Training
3. Experiment
   - Settings
   - Results

Problem Formulation

- In this paper, we handle both of the sub-tasks, i.e., OTE and TSC in an end-to-end fashion.
  - We formulate the complete task as a sequence labeling task.
  - We utilize a specially-designed tag set to consolidate boundary and sentiment information.

Input	The	AMD	Turin	Processor	seems	to	always	perform	much	better	than	Intel	.
Joint	O	B	I	E	O	O	O	O	O	O	O	S	O
	O	POS	POS	POS	O	O	O	O	O	O	O	NEG	O
Unified (✔)	O	B-POS	I-POS	E-POS	O	O	O	O	O	O	O	S-NEG	O

Table: Example tagging scheme.

Outline

1. Aspect/Target-Based Sentiment Analysis
   - Task Introduction
   - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
   - Overview
   - Boundary Guided TBSA
   - Maintaining Sentiment Consistency
   - Auxiliary Target Word Detection
   - Training
3. Experiment
   - Settings
   - Results

- Motivation:
  (i) Most of the related works only tackle one of the sub-tasks alone, hindering their practical uses.
  (ii) The existing unified models do not consider the dependencies between the two sub-tasks.
  (iii) Sentiments within the same opinion target should be consistent.
- Our Contributions:
  (i) Propose a new unified model for the complete TBSA / ABSA task.
  (ii) Design a Boundary-Guided component to explicitly model the transition from boundary to sentiment.
  (iii) Propose a **SC** component to maintain the sentiment consistency in the RNN-based model.

Model Overview

- Key components:
  - Two LSTMs (**LSTM**$^{\mathcal{T}}$ and **LSTM**$^{\mathcal{S}}$) for the target boundary detection task (auxiliary) and the complete TBSA task (primary).
  - **BG** component for exploiting boundary information and **SC** component for maintaining sentiment consistency.
  - **OE** component for improving the quality of the boundary information.

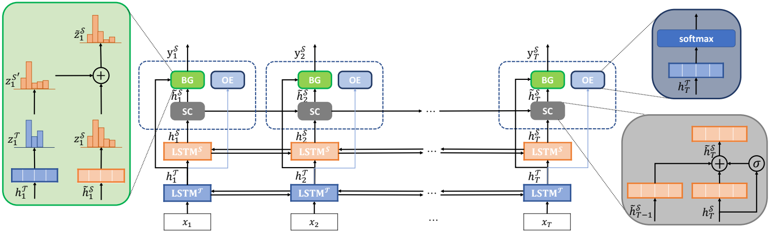

Figure: Architecture of the proposed framework.

Outline

1 Aspect/Target-Based Sentiment Analysis
- Task Introduction
- Problem Formulation

2 The Proposed Unified Solution for ABSA/TBSA
- Overview
- Boundary Guided TBSA
- Maintaining Sentiment Consistency
- Auxiliary Target Word Detection
- Training

3 Experiment
- Settings
- Results

Boundary Guided TBSA

Target boundary is useful for the target sentiment prediction, we design two schemes to incorporate boundary information:

1. Feature sharing for indirect boundary guidance.
2. Boundary transition for direct boundary guidance.

- Apart from the primary ABSA/TBSA task, we introduce an auxiliary target boundary detection task:
  - No external knowledge, human annotation or distant supervision is needed !
- Indirect guidance via feature sharing:
  - We stack $\textbf{LSTM}^{\mathcal{T}}$ and $\textbf{LSTM}^{\mathcal{S}}$ and regard the hidden representations from $\textbf{LSTM}^{\mathcal{T}}$ as the input of $\textbf{LSTM}^{\mathcal{S}}$ for the primary task.

$$\begin{aligned} h_t^{\mathcal{T}} &= [\overrightarrow{\text{LSTM}}^{\mathcal{T}}(x_t); \overleftarrow{\text{LSTM}}^{\mathcal{T}}(x_t)], \\ h_t^{\mathcal{S}} &= [\overrightarrow{\text{LSTM}}^{\mathcal{S}}(h_t^{\mathcal{T}}); \overleftarrow{\text{LSTM}}^{\mathcal{S}}(h_t^{\mathcal{T}})], t \in [1, T]. \end{aligned} \tag{1}$$

  - Boundary features are implicitly incorporated into the primary LSTM via feature sharing.

Direct Boundary Guidance

✶ Idea: make predictions based on both of the feature-based distributions and the boundary-based distributions over tag-set.

◆ Feature-based distribution is obtained based on the hidden representations from the primary LSTM:

$$z_t^{\mathcal{S}} = \mathbf{p}(y_t^{\mathcal{S}} | h_t^{\mathcal{S}}) = \mathrm{Softmax}(\mathbf{W}^{\mathcal{S}} h_t^{\mathcal{S}}) \tag{2}$$

◆ Boundary-based distributions is calculated by **BG** component.

Calculate Boundary-based Distributions

- First of all, we estimate the confidence α_t of the current boundary prediction $z_t^{\mathcal{T}}$ (GINI impurity):

$$\alpha_t = \epsilon c_t, c_t = (z_t^{\mathcal{T}})^\top z_t^{\mathcal{T}} \tag{3}$$

- Then, transiting boundary predictions to the boundary-based sentiment predictions:

$$z_t^{\mathcal{S}'} = (\mathbf{W}^{tr})^\top z_t^{\mathcal{T}} \tag{4}$$

- Since the boundary tags should be coherent with the sentiment tags, $\mathbf{W}^{tr}_{i,j}$ need to be carefully designed,

$$\mathbf{W}^{tr}_{i,j} = \begin{cases} \frac{1}{|\mathcal{B}_i|}, & \text{if } j \in \mathcal{B}_i \\ 0, & \text{Otherwise} \end{cases} \tag{5}$$

Example Transitions

	B-POS	B-NEG	B-NEU	I-POS	I-NEG	I-NEU	E-POS	E-NEG	E-NEU	S-POS	S-NEG	S-NEU	O
B	✔	✔	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Prediction

✶ Take both feature-based sentiment distribution and boundary-based sentiment distribution into consideration:

$$\tilde{z}_t^{\mathcal{S}} = \alpha_t z_t^{\mathcal{S}'} + (1 - \alpha_t) z_t^{\mathcal{S}}. \tag{6}$$

$$y_t^{\mathcal{S}} = \arg\max_y z_t^{\mathcal{S}} \tag{7}$$

Outline

1 Aspect/Target-Based Sentiment Analysis
- Task Introduction
- Problem Formulation

2 The Proposed Unified Solution for ABSA/TBSA
- Overview
- Boundary Guided TBSA
- Maintaining Sentiment Consistency
- Auxiliary Target Word Detection
- Training

3 Experiment
- Settings
- Results

Maintaining Sentiment Consistency

- RNN-based sequence labeling models do not impose constraint on the adjacent predictions:
  - e.g., `I-NEG` can appear right after the `B-POS`.
  - e.g., `O` may follow the prediction of `I-NEU`.
- To alleviate this issue, we introduce an additional gate on top of the output of **LSTM**$^{\mathcal{S}}$:

$$\begin{aligned} \tilde{h}_t^{\mathcal{S}} &= g_t \odot h_t^{\mathcal{S}} + (1 - g_t) \odot \tilde{h}_{t-1}^{\mathcal{S}} \\ g_t &= \sigma(\mathbf{W}^g h_t^{\mathcal{S}} + \mathbf{b}^g) \end{aligned} \tag{8}$$

Outline

1 Aspect/Target-Based Sentiment Analysis
- Task Introduction
- Problem Formulation

2 The Proposed Unified Solution for ABSA/TBSA
- Overview
- Boundary Guided TBSA
- Maintaining Sentiment Consistency
- Auxiliary Target Word Detection
- Training

3 Experiment
- Settings
- Results

Auxiliary Target Word Detection

- High-quality boundary information is essential:
  - The collocated opinion words can provide some clues for the target word discovery.
- We introduce a Target Word Detection task and design an **OE** component to improve the quality of boundary predictions:
  - At each step, **OE** predicts if the current word is a target word, i.e., solving a binary classification problem.

$$z_t^{\mathcal{O}} = \mathrm{Softmax}(\mathbf{W}^o h_t^{\mathcal{T}}) \tag{9}$$

  - This will help to locate target span more precisely.

Outline

1. Aspect/Target-Based Sentiment Analysis
   - Task Introduction
   - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
   - Overview
   - Boundary Guided TBSA
   - Maintaining Sentiment Consistency
   - Auxiliary Target Word Detection
   - Training
3. Experiment
   - Settings
   - Results

Training

- Token-level cross-entropy loss is employed as the loss function:

$$\mathcal{L}^{\mathcal{I}} = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}(y_t^{\mathcal{I},g}) \circ \log(z_t^{\mathcal{I}}), \mathcal{I} \in \{\mathcal{S}, \mathcal{T}, \mathcal{O}\} \tag{10}$$

- The following objective is minimized during training:

$$\mathcal{J}(\theta) = \mathcal{L}^{\mathcal{S}} + \mathcal{L}^{\mathcal{T}} + \mathcal{L}^{\mathcal{O}}. \tag{11}$$

Outline

1. Aspect/Target-Based Sentiment Analysis
   - Task Introduction
   - Problem Formulation
2. The Proposed Unified Solution for ABSA/TBSA
   - Overview
   - Boundary Guided TBSA
   - Maintaining Sentiment Consistency
   - Auxiliary Target Word Detection
   - Training
3. Experiment
   - Settings
   - Results

Settings

- Datasets:
  - $\mathbb{D}_{\mathrm{R}}$ (Restaurant Reviews), $\mathbb{D}_{\mathrm{L}}$ (Laptop Reviews): from SemEval ABSA challenges (2014-2016).
  - $\mathbb{D}_{\mathrm{T}}$: from Twitter posts.
- Evaluation metrics:
  - Precision, Recall and Macro-F1
  - Counting True-Positive is based on exact match of the target span and the sentiment.

Settings

- Compared methods:
  - **CRF**-{**pipeline**, **joint**, **unified**}: Conditional Random Fields (CRF) based sequence tagger.
  - **NN-CRF**-{**pipeline**, **joint**, **unified**}: Enhanced CRF models armed with word embeddings and neural network feature extractors.
  - **HAST-TNet**: HAST and TNet are the current state-of-the-art models on the tasks of OTE and TSC respectively. HAST-TNet is the pipline approach of these two models.
  - **LSTM-unified**: the standard LSTM model adopting the unified tagging scheme.
  - **LSTM-CRF-1**, **LSTM-CRF-2**: LSTM model with CRF decoding layer and no feature engineering is needed. Unified tagging scheme is adopted.
  - **LM-LSTM-CRF**: Language model enhanced LSTM-CRF model. Unified tagging scheme is adopted.

Outline

1 Aspect/Target-Based Sentiment Analysis
- Task Introduction
- Problem Formulation

2 The Proposed Unified Solution for ABSA/TBSA
- Overview
- Boundary Guided TBSA
- Maintaining Sentiment Consistency
- Auxiliary Target Word Detection
- Training

3 Experiment
- Settings
- Results

Main Results

	Model	$\mathbb{D}_L$			$\mathbb{D}_R$			$\mathbb{D}_T$		
		P	R	F1	P	R	F1	P	R	F1
Existing Baselines	CRF-joint	57.38	35.76	44.06	60.00	48.57	53.68	43.09	24.67	31.35
	CRF-unified	59.27	41.86	49.06	63.39	57.74	60.43	48.35	19.64	27.86
	NN-CRF-joint	55.64	34.48	45.49	61.56	50.00	55.18	44.62	35.84	39.67
	NN-CRF-unified	58.72	45.96	51.56	62.61	60.53	61.56	46.32	32.84	38.36
Pipeline Baselines	CRF-pipeline	59.69	47.54	52.93	52.28	51.01	51.64	42.97	25.21	31.73
	NN-CRF-pipeline	57.72	49.32	53.19	60.09	61.93	61.00	43.71	37.12	40.06
	HAST-TNet	56.42	54.20	55.29	62.18	73.49	67.36	46.30	49.13	47.66
Unified Baselines	LSTM-unified	57.91	46.21	51.40	62.80	63.49	63.14	51.45	37.62	43.41
	LSTM-CRF-1	58.61	50.47	54.24	66.10	66.30	66.20	51.67	44.08	47.52
	LSTM-CRF-2	58.66	51.26	54.71	61.56	67.26	64.29	53.74	42.21	47.26
	LM-LSTM-CRF	53.31	59.4	56.19	68.46	64.43	66.38	43.52	52.01	47.35
OURS	Base model	60.00	46.85	52.61	61.48	66.16	63.73	53.02	41.47	46.50
	Base model + **BG**	58.58	50.63	54.31	67.51	66.42	66.96	52.26	43.84	47.66
	Base model + **BG** + **SC**	58.95	53.00	55.81	63.95	69.65	66.68	53.12	43.60	47.79
	Base model + **BG** + **OE**	63.43	49.53	55.62	62.85	66.77	65.22	53.10	43.50	47.78
	Full model	61.27	54.89	**57.90**$^{\natural,\sharp}$	68.64	71.01	**69.80**$^{\natural,\sharp}$	53.08	43.56	**48.01**$^{\sharp}$

Table: Experimental Results. "Base model" refers to the stacked LSTMs. The markers $\natural$ and $\sharp$ refer to our full model significantly outperforms **HAST-TNet** and **LM-LSTM-CRF** respectively.

Case Analysis

Input	Base model		Base model + **BG**		**Full model**	
	Target	Complete	Target	Complete	Target	Complete
1. And the fact that it comes with an **[***i5 processor***]**$_{\texttt{POS}}$ definitely speeds things up	*i5 processor*	[*processor*]$_{\texttt{POS}}$ (✗)	*i5 processor*	[*i5 processor*]$_{\texttt{POS}}$	*i5 processor*	[*i5 processor*]$_{\texttt{POS}}$
2. There were small problems with **[***mac office***]**$_{\texttt{NEG}}$.	*mac office*	[*mac*]$_{\texttt{NEG}}$ (✗)	*mac office*	[*mac office*]$_{\texttt{NEG}}$	*mac office*	[*mac office*]$_{\texttt{NEG}}$
3. The **[***teas***]**$_{\texttt{POS}}$ are great and all the **[***sweets***]**$_{\texttt{POS}}$ are homemade	*teas, sweets*	[*teas*]$_{\texttt{POS}}$, [*sweets*]$_{\texttt{POS}}$	*teas, sweets, homemade* (✗)	[*teas*]$_{\texttt{POS}}$, [*sweets*]$_{\texttt{POS}}$, [*homemade*]$_{\texttt{POS}}$ (✗)	*teas, sweets*	[*teas*]$_{\texttt{POS}}$, [*sweets*]$_{\texttt{POS}}$
4. I love the **[***form factor***]**$_{\texttt{POS}}$	NONE	NONE	NONE	NONE	*form factor*	[*form factor*]$_{\texttt{POS}}$
5. I blame the **[***Mac OS***]**$_{\texttt{NEG}}$.	*Mac OS*	[*Mac*$_{\texttt{NEG}}$ *OS*$_{\texttt{NEU}}$] (✗)	*Mac OS*	[*Mac*$_{\texttt{NEG}}$ *OS*$_{\texttt{POS}}$] (✗)	*Mac OS*	[*Mac OS*]$_{\texttt{NEG}}$
6. Also, I personally wasn't a fan of the **[***portobello and asparagus mole***]**$_{\texttt{NEG}}$.	*portobello and asparagus mole*	[*portobello*$_{\texttt{NEG}}$ *and*$_{\texttt{NEG}}$ *asparagus*$_{\texttt{NEG}}$ *mole*$_{\texttt{NEU}}$] (✗)	*portobello and asparagus mole*	[*portobello*$_{\texttt{NEG}}$ *and*$_{\texttt{NEG}}$ *asparagus*$_{\texttt{NEU}}$ *mole*$_{\texttt{NEU}}$] (✗)	*portobello and asparagus mole*	[*portobello and asparagus mole*]$_{\texttt{NEG}}$

Table: Case analysis. The "Target" column contains the results from the auxiliary task of target boundary detection. The "Complete" column presents the output of the complete TBSA task, but note that we only show the sentiment part of the unified labels (i.e., POS, NEG, and NEU) and use brackets to indicate the boundary. The marker ✗ denotes the incorrect prediction.

Result Analysis

✶ Observations:

- In the first input and the second input, the "Base model" correctly predicts the target boundary but fails to produce the right sentiments.
- Only using the BG component may inherit the errors from the lower boundary detection task (see the third and the fourth case).
- Maintaining sentiment consistency within the same target mention, especially for whose with several words, is difficult for "Base model" and "Base model+BG" (see the last case).

✶ Analysis:

- Our **BG** component explicitly guiding the TBSA with boundary predictions can improve the coherence between the boundary tags and the sentiment tags.
- From the 3rd & 4th case, we can know that the performance of the auxiliary boundary detection task is critical to the TBSA task and employing the additional **OE** component is useful.
- With the help of **SC** component, the sentiment inconsistency in the 5th & 6th case can be alleviated.

Q&A

- Code available at `https://github.com/lixin4ever/E2E-TBSA`.
- If you still have problems after Q&A, feel free to send email to lixin4ever@gmail.com.