清華大學
Tsinghua University

A Unified Model for Opinion Target Extraction and Target Sentiment Prediction

报告人：史可为、吴冠陞

2021-12-23

金融＋情感分析

社群平台

weibo

主观性文本

金融领域平台

基金

雪球

相关研究

金融词库：中国财经媒体领域的正负面词库

中文情感分析分类模型

一、研究背景

二、研究方法

三、实验

四、结论

研究背景

基于目标的情感分析（Target-based sentiment analysis, TBSA）：1. 检测意见目标， 2. 并预测意见目标的情绪极性

一般方法：分为两个子任务

- 目标提取
- 检测文本中的意见目标

- 情绪分类
- 预测给定意见目标的情感极性

传统工作：大部分只针对其中的一个子任务解决。

清華大學 Tsinghua University

已有的集成解决方案：

- 两个子任务的模型联合训练
  利用一组目标边界标签（例如：B，I，E，S和O）
  和一组情感标签（例如POS，NEG，NEU）
  进行联合训练

Unified Tagging Scheme：
两个子任务的边界，利用一组专门设计的标签（“统一标记方案”）

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 1: Tagging schemes used in the integrated approaches. “Joint” and “Unified” refers to joint and unified approaches respectively.

问题 提出一种更完整的模型来解决这一问题

A Unified Model for Opinion Target Extraction and Target Sentiment Prediction

unified tagging scheme

本文提出一种端到端的方案，通过一个应用”统一标记方案”的统一模型来完成TBSA。

模型包括两个LSTM：

上层预测统一标签进行初步TBSA

下层通过辅助目标边界预测来指导上层网络提升性能

清華大學
Tsinghua University

任务定义

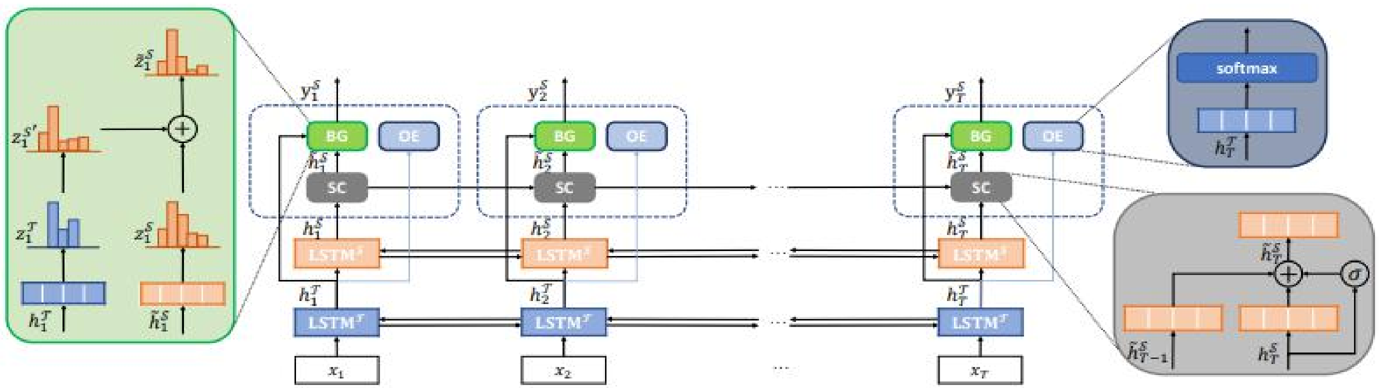
本文提出了一个 Target-Based Sentiment Analysis（**TBSA**）任务作为一个序列标注任务，给定一个序列采用如下标签进行标记：

其中B,I,E,S分别代表一个 aspect 的开始、一个 aspect 的中间部分、一个 aspect 的结尾、和单个 aspect；O 代表该字符不在一个 aspect 之内；POS、NEG、NEU 分别代表情感倾向为正向、负向、中性。如下图所示：

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

模型图

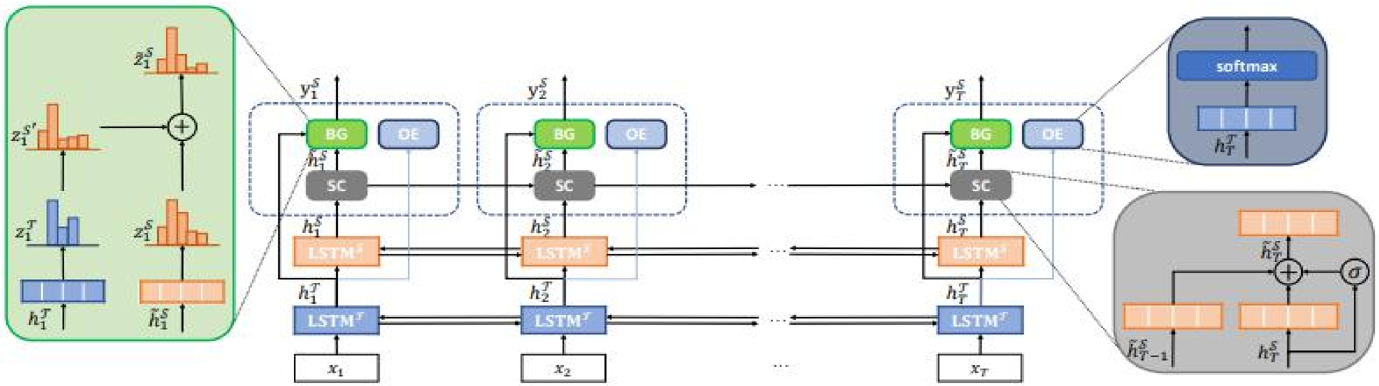
针对给定的输入序列 $\mathrm{X} = (x_1, \dots, x_T)$，我们的目标是输出 $\mathrm{Y}^S = (y_1^S, \dots, y_T^S)$，其中 $y_i^S \in \mathcal{Y}^S$ 。



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

低层LSTM: 主要负责目标边界检测，即预测给定的 token 标签
高层LSTM: 主要负责预测标签序列，并以低层LSTM生成的隐含表示作为指导信息
Sentiment Consistency (SC): 情感一致性组件，结合前文特征到当前预测中，保持多词目标的情感依赖
Boundary Guidance (BG): 边界引导组件，从辅助任务提取边界信息指导统一标签预测
Opinion-Enhanced (OE): 观点增强目标词检测组件，判定当前词是否为目标词

Target Boundary Guided TBSA



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

- 低层LSTM主要负责预测目标边界，即预测给定的 token 标签 $\mathcal{Y}^T \in \{B, I, E, S, O\}$
- 高层LSTM预测标签序列，即标签POS/NEG/NEU
- 低层LSTM生成的隐含表示作为高层的指导信息。

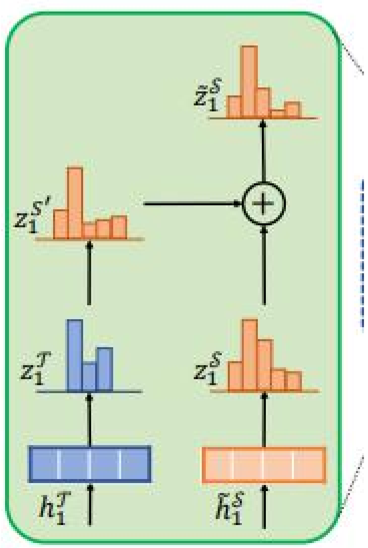
$$h_t^{\mathcal{T}} = [\overrightarrow{\mathrm{LSTM}}^{\mathcal{T}}(x_t); \overleftarrow{\mathrm{LSTM}}^{\mathcal{T}}(x_t)],$$

$$h_t^{\mathcal{S}} = [\overrightarrow{\mathrm{LSTM}}^{\mathcal{S}}(h_t^{\mathcal{T}}); \overleftarrow{\mathrm{LSTM}}^{\mathcal{S}}(h_t^{\mathcal{T}})], \text{ t}\in[1,\mathrm{T}]$$

清華大学 Tsinghua University

Target Boundary Guided TBSA

- 边界标签和统一标签的概率分数：



$$z_t^{\mathcal{T}} = \mathbf{p}\left(y_t^{\mathcal{T}} | x_t\right) = \mathrm{Softmax}\left(\mathbf{W}^{\mathcal{T}} h_t^{\mathcal{T}}\right)$$

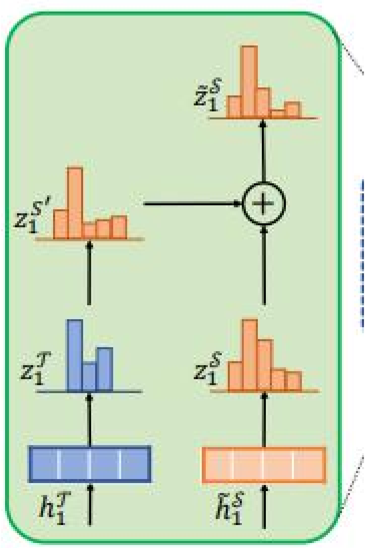
$$z_t^{\mathcal{S}} = \mathbf{p}\left(y_t^{\mathcal{S}} | h_t^{\mathcal{T}}\right) = \mathrm{Softmax}\left(\mathbf{W}^{\mathcal{S}} h_t^{\mathcal{S}}\right)$$

$z_t^{\mathcal{T}} \in \mathbb{R}^{|\mathcal{Y}^{\mathcal{T}}|}$ 代表着该 token 作为 boundary tag 的概率

$z_t^{\mathcal{S}} \in \mathbb{R}^{|\mathcal{Y}^{\mathcal{S}}|}$ 在前者的基础上计算其附加情感倾向的概率

Target Boundary Guided TBSA

- Boundary guided transition matrix



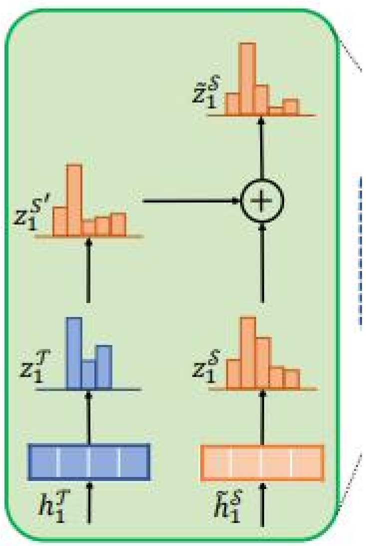
$$\mathbf{W}^{tr} \in \mathbb{R}^{|\mathcal{Y}^T| \times |\mathcal{Y}^S|}$$

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

这里的 $\mathcal{B}_i$ 是在给定 boundary tag i 的情况下，unified tag 的合理输出，比如 B 对应着 B-POS、B-NEG、B-NEU 。最后 transition-based sentiment score $z_t^{S'} \in \mathbb{R}^{|\mathcal{Y}^S|}$ 计算方式如下：

$$z_t^{S'} = \left(\mathbf{W}^{tr}\right)^{\top} z_t^T$$

Target Boundary Guided TBSA



如果 z_t^T 差不多满足均匀分布的情况下，那么 $z_t^{S'}$ 也会差不多满足均匀分布

- Proportion scor $\alpha_t \in \mathbb{R}$:

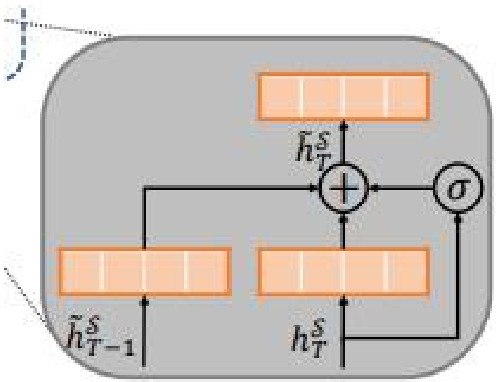
$$c_t = \left(z_t^T\right)^\top z_t^T$$
$$\alpha_t = \epsilon c_t$$

- Final scores

$$\tilde{z}_t^S = \alpha_t z_t^{S'} + (1 - \alpha_t)\, z_t^S$$

Maintaining Sentiment Consistency

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

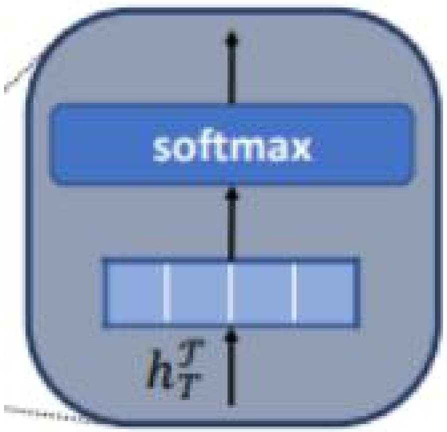


- 从当前和之前步的特征中预测当前统一标签（使用门机制）：

$$\tilde{h}_t^S = g_t \odot h_t^S + (1 - g_t) \odot \tilde{h}_{t-1}^S$$
$$g_t = \sigma\left(\mathbf{W}^g h_t^S + \mathbf{b}^g\right)$$

- 之前的特征也会被考虑到当前预测中，避免同一目标中出现不同的情感。

Auxiliary Target Word Detection



该部分的主要作用就是，确保情感词和aspect 同时出现，采用如下计算方式：

$$z_t^O = \mathrm{Softmax}\left(\mathbf{W}^o h_t^T\right)$$
$$y_t^O = \arg\max_y z_t^O$$

实验部分

数据集

来源：

1. SemEval ABSA Challenges
2. 推特数据库

DL：笔记型电脑产品评价

DR：2014 – 2016餐厅数据

DT：2013年收集的推特推文数据

Dataset		Train	Dev	Test	Total
$\mathbb{D}_L$	# POS	883	104	339	1326
	# NEG	754	106	130	990
	# NEU	404	46	165	615
$\mathbb{D}_R$	# POS	2337	270	1524	4131
	# NEG	942	93	500	1535
	# NEU	614	50	263	927
$\mathbb{D}_T$	# POS		-		692
	# NEG		-		263
	# NEU		-		2244

Table 2: Statistics of the datasets.

清華大學
Tsinghua University

对比模型

CRF- Conditional Random Fields 条件随机场 (Mitcheel 2013)

- CRF–pipeline：
- CRF–joint ： joint tagging scheme
- CRF–unified ： unified tagging scheme

NN–CRF 增强的CRF 模型，利用NN进行特征提取 （Zhang 2015）

- NN– CRF–pipeline：
- NN– CRF–joint ： joint tagging scheme
- NN– CRF–unified ： unified tagging scheme

HAST–TNet（Li等人）

目标检测＋情感分类 两个模型的 Pipeline

对比模型

LSTM–unified

标准 LSTM模型：unified tagging scheme

LSTM– CRF

LSTM+CRF decoding 层

LM–LSTM–CRF

Language mode 增强LSTM–CRF模型

实验结果

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

评估指标：standard precision（P）、recall（R）和F1 Score

	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**$^{\sharp}$

消融研究

base model中辅助LSTM预测的边界信息对性能有所提升，加入BG后，由于边界限制，提升更加明显，SC和OE如果只是单独加入，则另外的组件信息无法提供指导，有时性能反而下降。

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**$^{\sharp}$

清華大學
Tsinghua University

例子

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*]$_{\text{POS}}$ definitely speeds things up	*i5 processor*	[*processor*]$_{\text{POS}}$ (✗)	*i5 processor*	[*i5 processor*]$_{\text{POS}}$	*i5 processor*	[*i5 processor*]$_{\text{POS}}$
2. There were small problems with [*mac office*]$_{\text{NEG}}$.	*mac office*	[*mac*]$_{\text{NEG}}$ (✗)	*mac office*	[*mac office*]$_{\text{NEG}}$	*mac office*	[*mac office*]$_{\text{NEG}}$
3. The [*teas*]$_{\text{POS}}$ are great and all the [*sweets*]$_{\text{POS}}$ are homemade	*teas, sweets*	[*teas*]$_{\text{POS}}$, [*sweets*]$_{\text{POS}}$	*teas, sweets, homemade* (✗)	[*teas*]$_{\text{POS}}$, [*sweets*]$_{\text{POS}}$, [*homemade*]$_{\text{POS}}$ (✗)	*teas, sweets*	[*teas*]$_{\text{POS}}$, [*sweets*]$_{\text{POS}}$
4. I love the [*form factor*]$_{\text{POS}}$	NONE	NONE	NONE	NONE	*form factor*	[*form factor*]$_{\text{POS}}$
5. I blame the [*Mac OS*]$_{\text{NEG}}$.	*Mac OS*	[*Mac*$_{\text{NEG}}$ *OS*$_{\text{NEU}}$] (✗)	*Mac OS*	[*Mac*$_{\text{NEG}}$ *OS*$_{\text{POS}}$] (✗)	*Mac OS*	[*Mac OS*]$_{\text{NEG}}$
6. Also, I personally wasn't a fan of the [*portobello and asparagus mole*]$_{\text{NEG}}$.	*portobello and asparagus mole*	[*portobello*$_{\text{NEG}}$ *and*$_{\text{NEG}}$ *asparagus*$_{\text{NEG}}$ *mole*$_{\text{NEU}}$] (✗)	*portobello and asparagus mole*	[*portobello*$_{\text{NEG}}$ *and*$_{\text{NEG}}$ *asparagus*$_{\text{NEU}}$ *mole*$_{\text{NEU}}$] (✗)	*portobello and asparagus mole*	[*portobello and asparagus mole*]$_{\text{NEG}}$

base可以预测目标边界，但目标情感错误，两个LSTM没有准确预测边界信息，BG可以准确预测边界信息，但可能会存在边界检测错误，OE可以解决这个问题，SC针对一些较长的目标词也可准确识别。

总结

本文的模型中，三个组件起到了很好的辅助作用，只加入BG时由于增加了边界信息，准确率会上升，但可能会从低层LSTM中获取错误的边界信息，检测边界错误，OE与SC的加入会改善这个问题，但单独加入时有时性能反而会下降，其中原因可能是OE会提供准确的目标词与非目标词给SC进行情感依赖处理，而SC会控制同一目标词中相同的情感信息，后续可以验证SC与OE的相互关系。