Powering Comparative Classification with Sentiment Analysis via Domain Adaptive Knowledge Transfer

EMNLP 2021
Online & in the Dominican Republic
Conference on Empirical Methods in Natural Language Processing

Zeyu Li, Yilong Qin, Zihan Liu, Wei Wang
Department of Computer Science, University of California, Los Angeles
UCLA Samueli Computer Science

1. Introduction

- Comparative Preference Classification (**CPC**) explores whether a preference comparison exists between **two entities** in a sentence.
- Example: ***Python*** *is better suited for data analysis than* ***MATLAB*** *due to the many available deep learning libraries.* We would like to know if there is a preference of Python over MATLAB.
- Existing work: (1) Some model CPC as a sentence classification task without highlighting the two entities; (2) The state-of-the-art ED-GAT (Ma et al., 2020) only considers syntactic information (dependency parsing) and ignores the critical semantic relations and the sentiments to the compared entities.
- Another challenge: dataset is small (CompSent-19 with 7.2k sentences).
- We propose Sentiment Analysis Enhanced COmparative Network **(SAECON)** which improves CPC accuracy with a ***sentiment analyzer*** that learns sentiments to individual entities via ***domain adaptive knowledge transfer***.

2. Architecture Intuition

- Semantics: Add a semantics module to understand the entire sentences.
- Incorporating more knowledge:
  - Aspect Based Sentiment Analysis (ABSA) tries to identify the fine-grained opinion polarity towards a specific aspect associated with a given target.
  - Use ABSA to identify the sentiment to each entity. The preferred entity usually gets a positive sentiment. Its rival gets a relatively negative one.
- How to?
  - Directly incorporating a trained sentiment analyzer would cause a performance degradation due to **domain shift** between the CPC & ABSA.
  - Incorporating the architecture only with untrained parameters and jointly optimize them with the CPC task closes domain shift via GRL.

3. SAECON

- CPC task:
  - Global Semantic Context (orange)

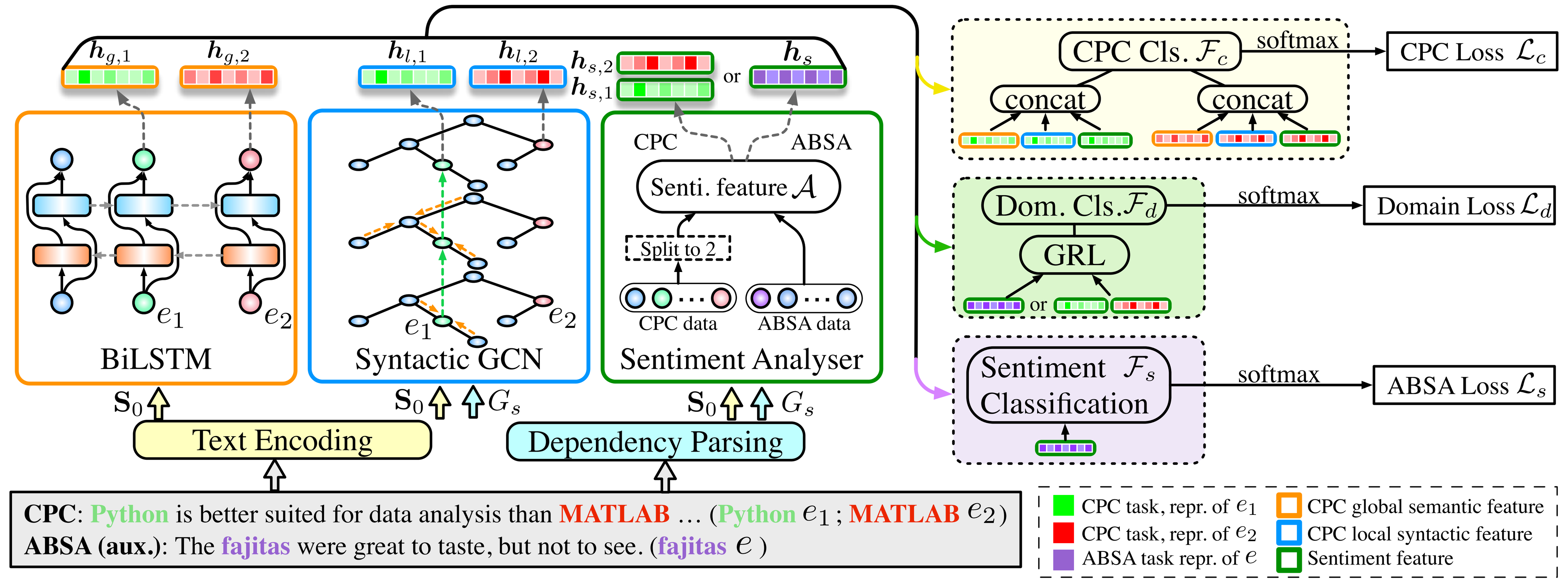
$$\overrightarrow{\boldsymbol{h}_{g,i}}, \overleftarrow{\boldsymbol{h}_{g,i}} = \mathrm{BiLSTM}(\mathbf{S}_0)[e_i.\mathrm{index}], \quad i = 1, 2$$

$$\boldsymbol{h}_{g,i} = \frac{1}{2}\left(\overrightarrow{\boldsymbol{h}_{g,i}} + \overleftarrow{\boldsymbol{h}_{g,i}}\right), \boldsymbol{h}_{g,i} \in \mathbb{R}^{d_g}.$$

  - Local Syntactic Context (Syntactic GCN, blue)

$$g_{uv}^{(j)} = \sigma\left(\boldsymbol{h}_u^{(j)} \cdot \boldsymbol{\beta}_{d_{uv}}^{(j)} + \gamma_{l_{uv}}^{(j)}\right), \quad g_{uv}^{(j)} \in \mathbb{R},$$

$$\boldsymbol{h}_v^{(j+1)} = \rho\left(\sum_{u \in \mathcal{N}(v)} g_{uv}^{(j)} \left(\mathbf{W}_{d_{uv}}^{(j)} \boldsymbol{h}_u^{(j)} + \boldsymbol{b}_{l_{uv}}^{(j)}\right)\right)$$



3. SAECON (cont'd)

- Sentiment Analyzing representation for each entity (green)

$$\mathcal{A}(\mathbf{S}_0, G_s, E) = \begin{cases} \boldsymbol{h}_{s,1}, \boldsymbol{h}_{s,2} & \text{if } s \in D_c, \text{ For CPC task/input, with CPC labels} \\ \boldsymbol{h}_s & \text{if } s \in D_s. \text{ For ABSA task/input, with ABSA labels} \end{cases}$$

- Domain shift via Gradient Reversal Layer (GRL) $\frac{\partial \mathrm{GRL}_\alpha}{\partial \boldsymbol{x}} = -\alpha \mathbf{I}.$
- Objective and Optimization

$$\hat{y}_c = \delta(\mathcal{F}_c([\mathcal{F}(\boldsymbol{h}_{e_1}); \mathcal{F}(\boldsymbol{h}_{e_2})])) \qquad \text{(CPC only)},$$

$$\hat{y}_s = \delta(\mathcal{F}_s(\boldsymbol{h}_s)) \qquad \text{(ABSA only)},$$

$$\hat{y}_d = \delta(\mathcal{F}_d(\mathrm{GRL}(\mathcal{A}(\mathbf{S}_0, G_s, E)))) \qquad \text{(Both tasks)},$$

4. CompSent-19 Dataset

- Statistics (see right)
- Imbalanced Data
  - Flipping labels
  - Upsampling
  - Weighted loss

Dataset	Better	Worse	None	Total
Train	872 (19%)	379 (8%)	3,355 (73%)	4,606
Development	219 (19%)	95 (8%)	839 (73%)	1,153
Test	273 (19%)	119 (8%)	1,048 (73%)	1,440
Total	1,346 (19%)	593 (8%)	5,242 (73%)	7,199
Flipping labels	1,251 (21%)	1,251 (21%)	3,355 (58%)	5,857
Upsampling	3,355 (33%)	3,355 (33%)	3,355 (33%)	10,065

5. Experiments

- Dataset: CompSent-19 (see above)
- ABSA dataset: SemEval 14 to 16
- Evaluation Metric:
  - F1 score of each label (Better, Worse, None)
  - Micro-averaging F1 score

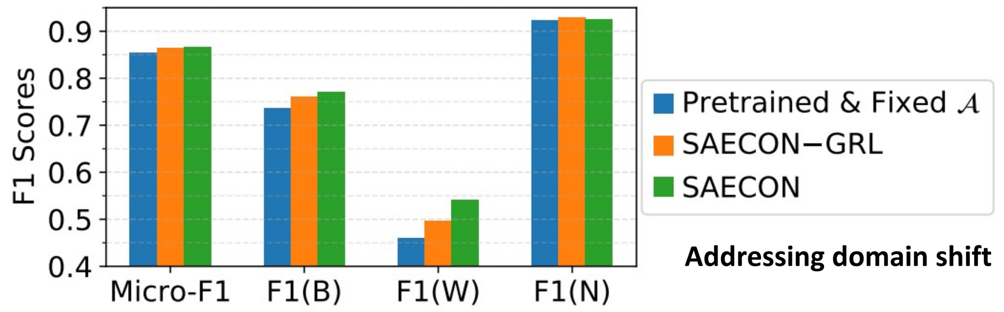
5. Experiments (cont'd)

- Experimental results demonstrate an increase in CPC task performance.
- Our ablation study suggests the effectiveness of our modules.
- Our techniques in addressing domain shift improve all F1 scores.
- Weighted loss proves to be the most effective technique in addressing data imbalance.

Model	Micro.	F1(B)	F1(W)	F1(N)
Majority	68.95	0.0	0.0	81.62
SE-Lin	79.31	62.71	37.61	88.42
SE-XGB	85.00	75.00	43.00	92.00
SVM-Tree	68.12	53.35	13.90	78.13
BERT-CLS	83.12	69.62	50.37	89.84
AvgWE-G	76.32	48.28	20.12	86.34
AvgWE-B	77.64	53.94	26.88	87.47
ED-GAT-G	82.73	70.23	43.30	89.84
ED-GAT-B	85.42	71.65	47.29	92.34
SAECON-G	83.78	71.06	45.90	91.05
SAECON-B	**86.74**	**77.10**	**54.08**	**92.64**

Variants	Micro.	F1(B)	F1(W)	F1(N)
SAECON	**86.74**	**77.10**	**54.08**	92.64
−BiLSTM	85.21	72.94	43.86	92.63
−SGCN	86.53	76.22	51.38	92.24
−GRL	86.53	76.16	49.77	**92.93**
−($\mathcal{A}$+GRL)	85.97	74.82	52.44	92.45

Methods	Micro.	F1(B)	F1(W)	F1(N)
Weighted loss (WL)	**86.74**	**77.10**	**54.08**	92.64
Original (OR)	85.97	73.80	46.15	92.90
Flipping labels (FL)	84.93	73.07	42.45	91.99
Upsampling (UP)	85.83	73.11	46.36	**92.95**



Addressing domain shift

Directed	Gating	Micro.	F1(B)	F1(W)	F1(N)
✓	✓	**86.74**	**77.10**	**54.08**	92.64
✗	✓	86.18	75.72	49.78	92.40
✓	✗	85.35	74.03	43.27	92.34
✗	✗	85.35	73.39	35.78	**93.04**

CPC sentences with sentiment predictions by $\mathcal{A}$	Label	Δ
S1: This is all done via the gigabit [**Ethernet**:POS] interface, rather than the much slower [**USB**:NEG] interface.	Better	+2
S2: Also, [**Bash**:NEG] may not be the best language to do arithmetic heavy operations in something like [**Python**:NEU] might be a better choice.	Worse	−1
S3: It shows how [**JavaScript**:POS] and [**PHP**:POS] can be used in tandem to make a user's experience faster and more pleasant.	None	0
S4: He broke his hand against [**Georgia Tech**:NEU] and made it worse playing against [**Virginia Tech**:NEU].	None	0

Supplementary CPC sentences with sentiment predictions by $\mathcal{A}$	Label	Δ
S1: [**Ruby**:NEU] wasn't designed to "exemplify best practices", it was to be a better [**Perl**:NEG].	Better	+1
S2: And from my experience the ticks are much worse in [**Mid Missouri**:NEG] than they are in [**South Georgia**:POS] which is much warmer year round.	Worse	−2
S3: As an industry rule, [**hockey**:NEG] and [**basketball**:NEG] sell comparatively poorly everywhere.	None	0
S4: [**Milk**:NEG], [**juice**:NEG] and soda make it ten times worse.	None	0

***Open Source:** https://github.com/zyli93/SAECON