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

Powering Comparative Classification with Sentiment Analysis via Domain Adaptive Knowledge Transfer

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

Paper ID: 1731

Comparative Preference Classification

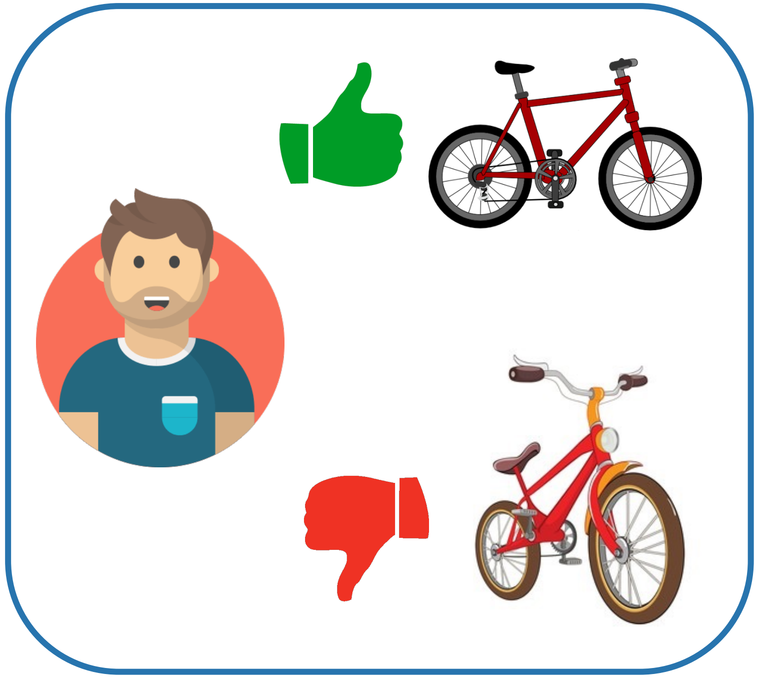
WHAT IS CPC?

- Whether a preference comparison exists between two entities in a sentence?
- For example:
    - Sentence: ***Python*** *is better suited for data analysis than* ***MATLAB*** *due to the many available deep learning libraries.*
    - Entities:
        - Entity A: Python; Entity B: MATLAB
    - Prefer “python” to “MATLAB”?
        - “**Better**”, “Worse”, “None”

Why useful?

USE CASES

- Identity a comparison:
  - In a piece of shopping review comparing two items
  - In a news article or a wiki page for relation extraction
  - On social media posts
- With the comparisons:
  - Build product graph for better recommender system
  - Understand users' preferences towards items
  - Extract comparative facts
  - And more



Existing works and Challenges

- Existing works:
  - Some model CPC as a sentence classification task without highlighting the two entities.
  - ED-GAT [MMWL'20]:
    - Purely dependency parsing-based, semantics deprived.
- Dataset:
  - CompSent-19 [PBFHB'19]
  - 7.2K sentences in total
- For a better solution:
  - Semantics
  - More training data as training knowledge

Intuition of Design

SEMANTICS

- Add a semantics module to understand the entire sentences

INCORPORATING MORE KNOWLEDGE

- Aspect-based Sentiment Analysis (ABSA)
  - Goal: identifying the fine-grained opinion polarity towards a specific aspect associated with a given target.
  - E.g.: “I liked the *service* and the *staff*, but not the *food”.*
    - Aspects: service, staff, food
    - Sentiments: positive, positive, negative
- How about incorporating ABSA to CPC?
  - The preferred entity usually receives a positive sentiment while its rival gets a relatively negative one

Intuition of Design

SEMANTICS

- Add a semantics module to understand the entire sentences

INCORPORATING MORE KNOWLEDGE

- Aspect-based Sentiment Analysis (ABSA)
- How about incorporating ABSA to CPC?
- How to incorporate?
    - Incorporate a trained sentiment analyzer
    - Incorporate the architecture only with untrained parameters and jointly optimize them with the CPC task

Intuition of Design

SEMANTICS

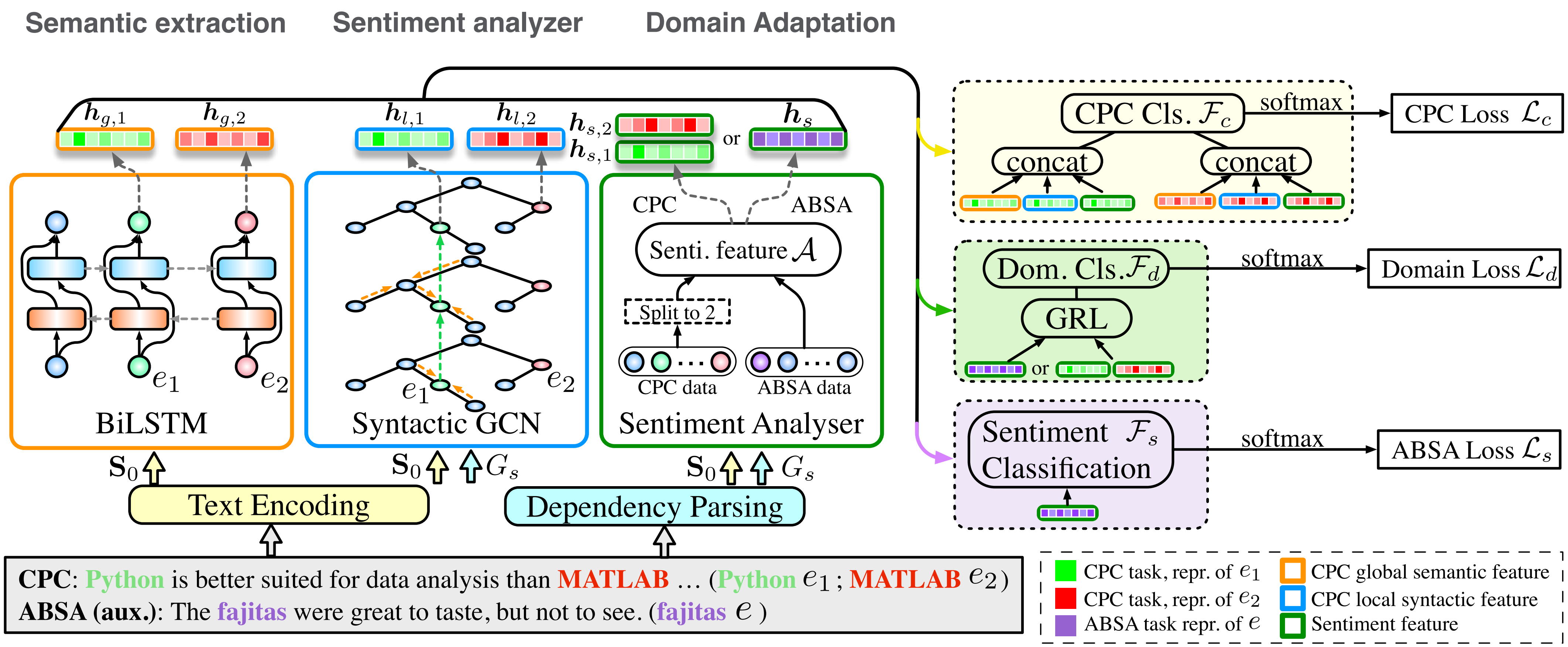
- Add a semantics module to understand the entire sentences

INCORPORATING MORE KNOWLEDGE

- Aspect-based Sentiment Analysis (ABSA)
- How about incorporating ABSA to CPC?
- How to incorporate?
  - Incorporate a trained sentiment analyzer → Domain shift
  - Incorporate the architecture only with untrained parameters and jointly optimize them with the CPC task → "Closeable" domain shift

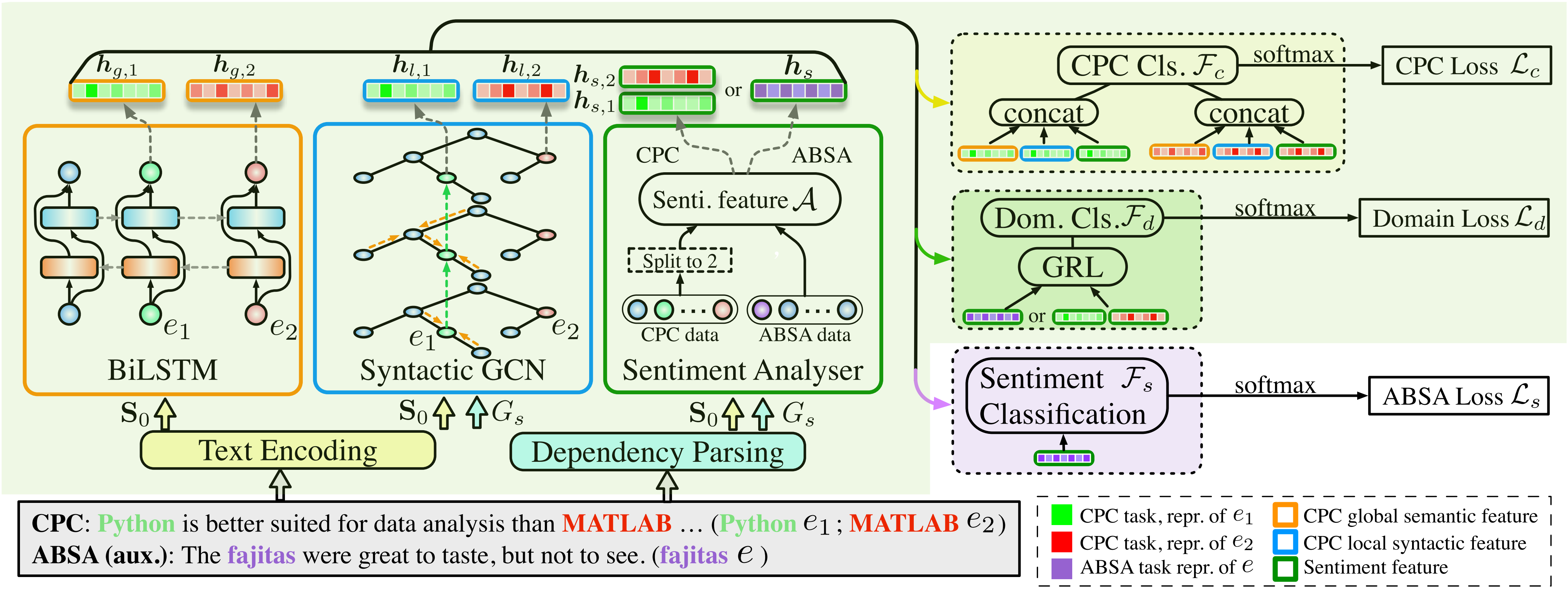
SAECON – Overall Arch

SENTIMENT ANALYSIS ENHANCED COMPARATIVE CLASSIFICATION NETWORK



SAECON – Training

- Different route for CPC and ABSA
- Forward pass for CPC input (in the green shade)



SAECON – Training

CPC TASK

1. Global Semantic Context

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

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

SAECON – Training

CPC TASK

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

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

2. Local Syntactic Context (Syntactic GCN [MT'17])

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

Direction and label of (u,v)

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

Aggregation function
(e.g. sum/mean/etc)

SAECON – Training

CPC TASK

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

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

2. Local Syntactic Context (Syntactic GCN [MT'17])

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

Direction and label of (u,v)

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

Aggregation function
(e.g. sum/mean/etc)

3. Sentiment Analyzing representation for each entity

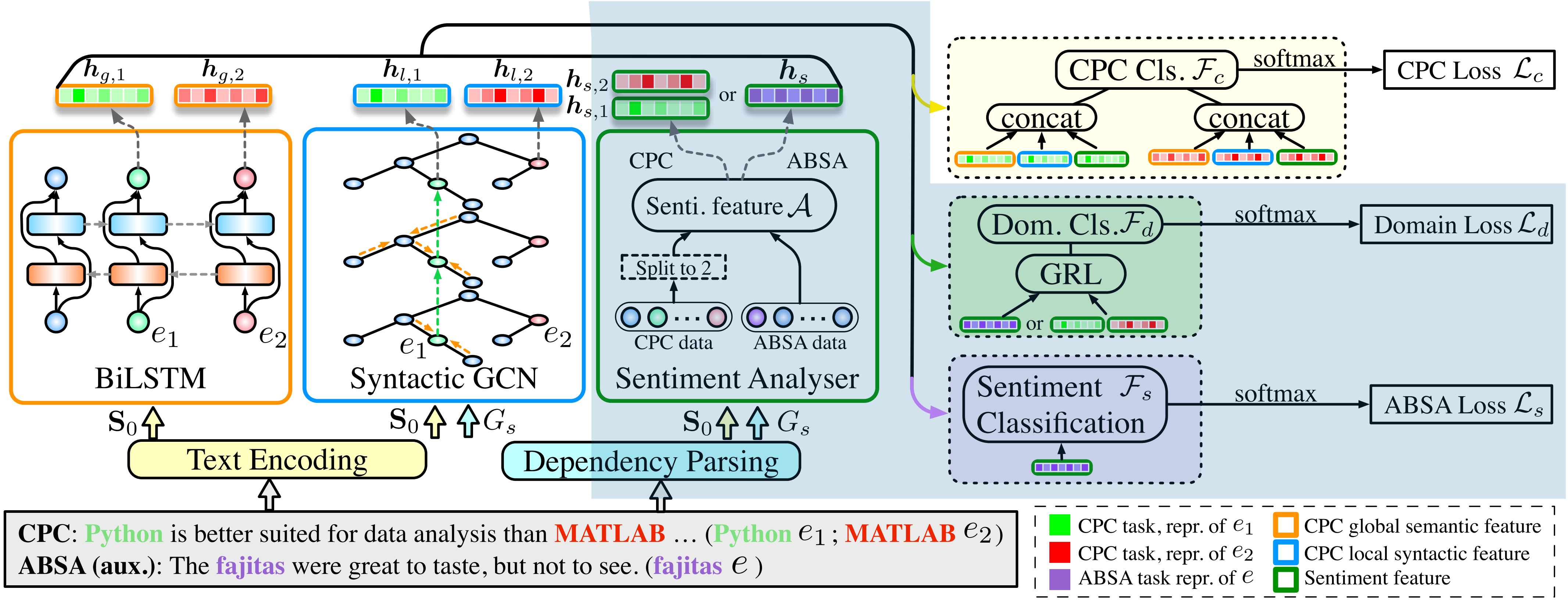
$$\mathcal{A}(\mathbf{S}_0, G_s, E) = \begin{cases} \boldsymbol{h}_{s,1}, \boldsymbol{h}_{s,2} & \text{if } s \in D_c, \\ \boldsymbol{h}_s & \text{if } s \in D_s. \end{cases}$$

For CPC task/input, with CPC labels

For ABSA task/input, with ABSA labels

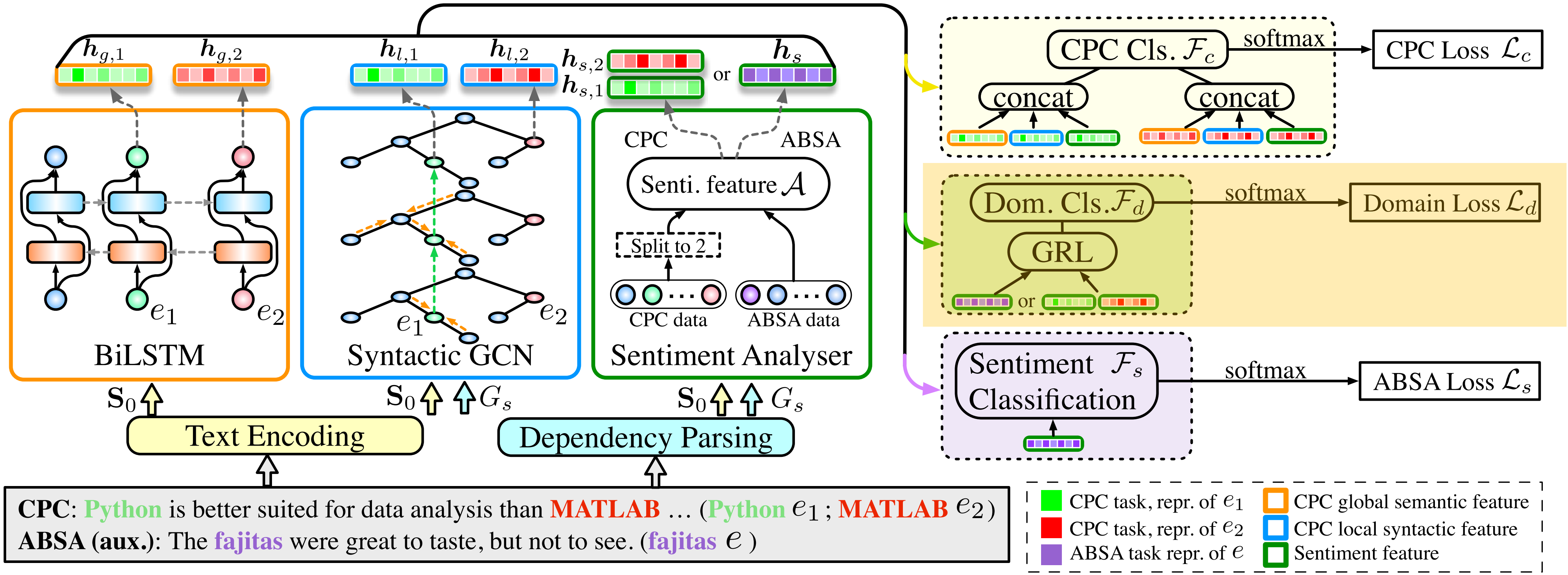
SAECON – Training

- Different route for CPC and ABSA
- Forward pass for ABSA input (in the blue shade)



SAECON – Training

- Different route for CPC and ABSA
- For both types of input, we train the domain classification layer



SAECON – Training

SENTIMENT ANALYSIS

Sentiment analyzer output different representations for different task

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

For CPC task/input, with CPC labels

For ABSA task/input, with ABSA labels

DOMAIN SHIFT VIA GRADIENT REVERSAL LAYER (GRL)

$$\frac{\partial \mathrm{GRL}_\alpha}{\partial \boldsymbol{x}} = -\alpha \mathbf{I}.$$

OBJECTIVE AND OPTIMIZATION

$$\begin{aligned} \hat{y}_c &= \delta(\mathcal{F}_c([\mathcal{F}(\boldsymbol{h}_{e_1}); \mathcal{F}(\boldsymbol{h}_{e_2})])) && \text{(CPC only)}, \\ \hat{y}_s &= \delta(\mathcal{F}_s(\boldsymbol{h}_s)) && \text{(ABSA only)}, \\ \hat{y}_d &= \delta(\mathcal{F}_d(\mathrm{GRL}(\mathcal{A}(\mathbf{S}_0, G_s, E)))) && \text{(Both tasks)}, \end{aligned}$$

Binary/Multiclass classifications

Experiments – Setup

COMPSENT-19

1. Statistics
2. Imbalanced Data
    1. Flipping labels
    2. Upsampling
    3. Weighted loss
3. Evaluation Metric
    1. F1 score of each label (B, W, N)
    2. Micro-averaging F1

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

Experiments – Performance

COMPARING WITH BASELINES & ABLATION STUDY

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

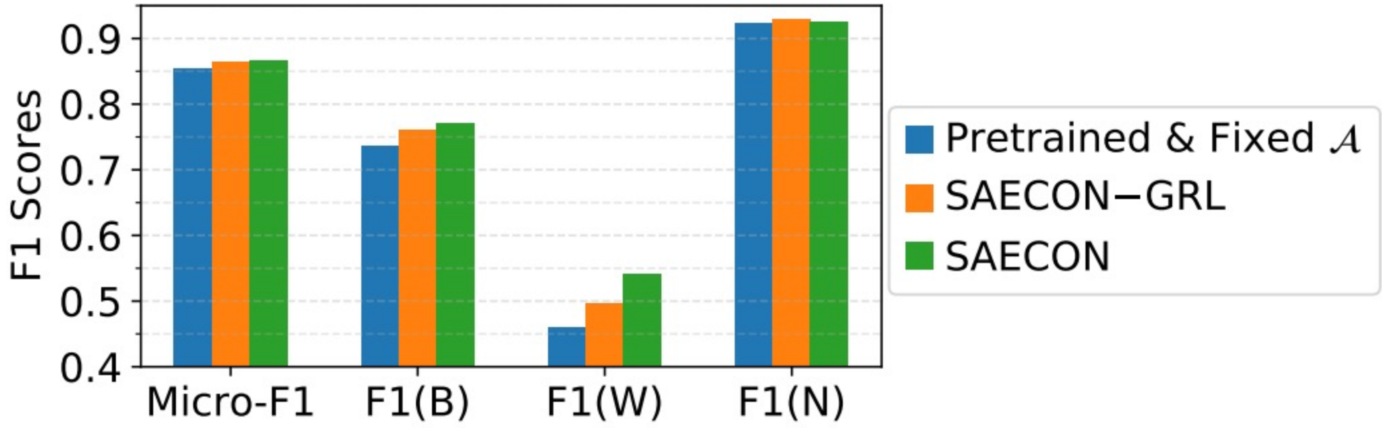
Experiments – Analyses

DATA IMBALANCE AND ALTERNATIVE TRAINING

Data Imbalance

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**

Alternative Training



Experiments – Case Study

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

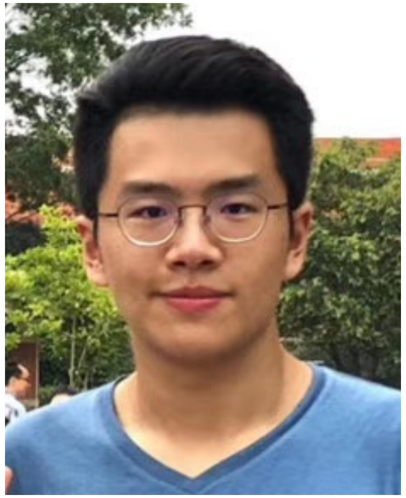
Thanks!

Code on GitHub: https://github.com/zyli93/SAECON

Paper ID: 1731

Looking forward to seeing you in the **POSTER** session!

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



UCLA Samueli Computer Science