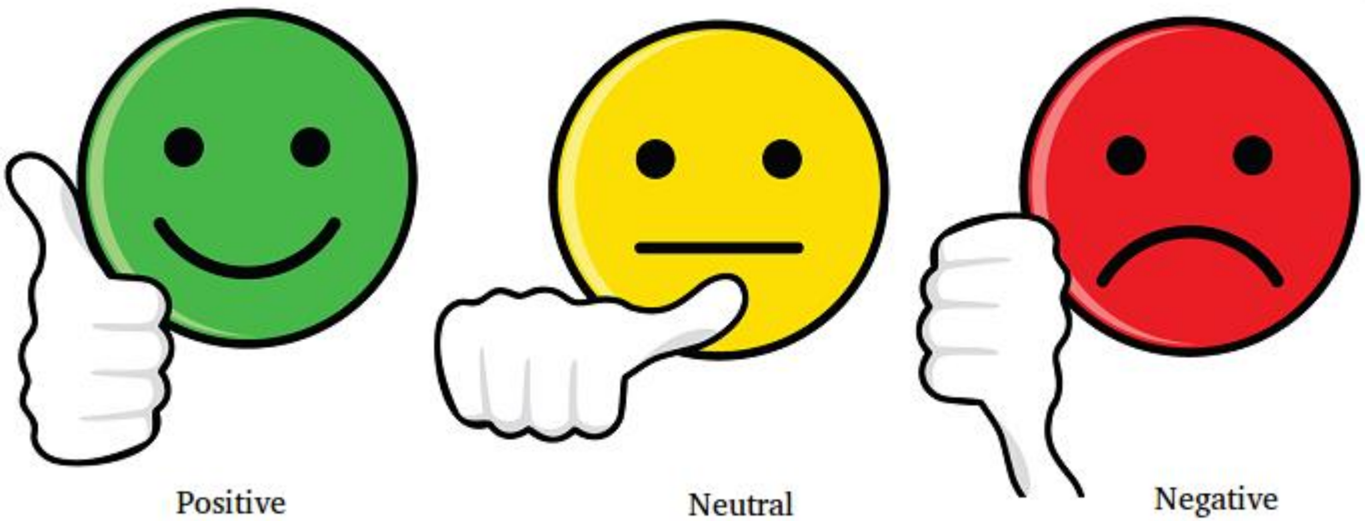
Aspect-based Sentiment Classification with Graph Convolutional Networks

Hyunjae Kim

Data Mining & Information Systems Lab.
Department of Computer Science and Engineering,
College of Informatics, Korea University

Sentiment Classification



Aspect-Based Sentiment Classification

Great food but the service was dreadful

Aspects : Food, Service
Sentiments :
Food → good, Service → bad

Limitations of Previous Models (1/2)

- Attention-based models

Its size is ideal and the weight is acceptable.

Good food bad service

Limitations of Previous Models (1/2)

- CNN-based models

The staff a bit more friendly.

The staff should be a bit more friendly.

Limitations of Previous Models (2/2)

The staff should be a bit more friendly

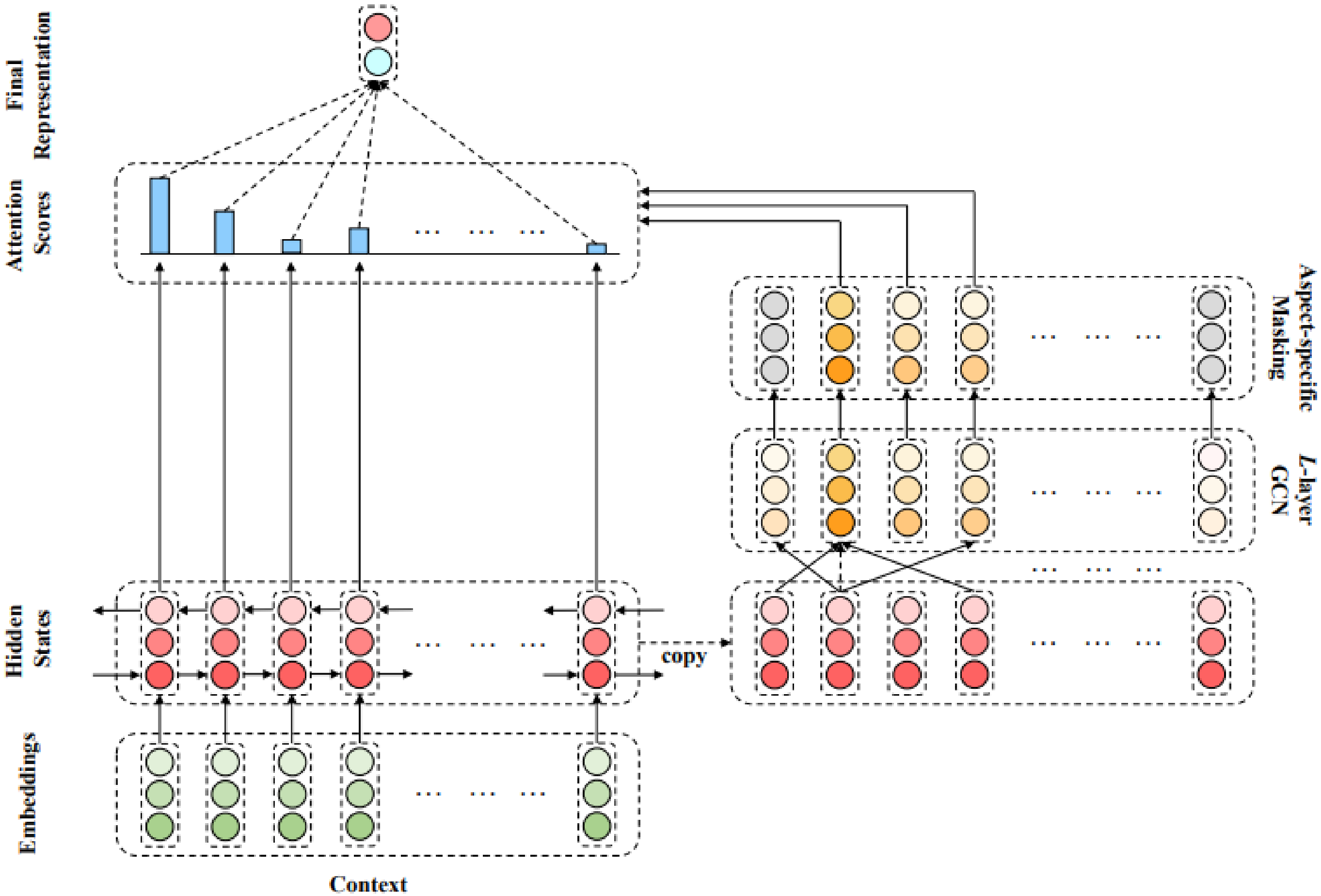
CNNs - the sentiment of an aspect is usually determined by key phrases instead of individual words.

Solution

Utilizing GCN!

- **Two papers** accepted at EMNLP 2019
  - Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks
  - Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks

Approach 1 (1/3)



Approach 1 (2/3)

$$\tilde{\mathbf{h}}_i^l = \sum_{j=1}^{n} \mathbf{A}_{ij} \mathbf{W}^l \mathbf{g}_j^{l-1} \tag{2}$$

$$\mathbf{h}_i^l = \mathrm{ReLU}(\tilde{\mathbf{h}}_i^l / (d_i + 1) + \mathbf{b}^l) \tag{3}$$

$$\mathbf{g}_i^l = \mathcal{F}(\mathbf{h}_i^l) \tag{4}$$

$$q_i = \begin{cases} 1 - \frac{\tau+1-i}{n} & 1 \le i < \tau + 1 \\ 0 & \tau + 1 \le i \le \tau + m \\ 1 - \frac{i-\tau-m}{n} & \tau + m < i \le n \end{cases} \tag{5}$$

- position weights

$$\mathcal{F}(\mathbf{h}_i^l) = q_i \mathbf{h}_i^l \tag{6}$$

Approach 1 (3/3)

$$\beta_t = \sum_{i=1}^{n} \mathbf{h}_t^{c\top} \mathbf{h}_i^{L} = \sum_{i=\tau+1}^{\tau+m} \mathbf{h}_t^{c\top} \mathbf{h}_i^{L} \tag{8}$$

- Aspect-specific Masking

$$\alpha_t = \frac{\exp(\beta_t)}{\sum_{i=1}^{n} \exp(\beta_i)} \tag{9}$$

- Aspect-specific Attention

$$\mathbf{r} = \sum_{t=1}^{n} \alpha_t \mathbf{h}_t^{c} \tag{10}$$

$$\mathbf{p} = \mathrm{softmax}(\mathbf{W}_p \mathbf{r} + \mathbf{b}_p) \tag{11}$$

Approach 2

$$h_{l+1}^{i} = \Big\Vert_{k=1}^{K} \sigma\Big(\sum_{j \in n[i]} \alpha_{lk}^{ij} W_{lk} h_l^j\Big) \tag{1}$$

$$\alpha_{lk}^{ij} = \frac{exp(f(a_{lk}^T[W_{lk}h_l^i || W_{lk}h_l^j]))}{\sum_{u \in n[i]} exp(f(a_{lk}^T[W_{lk}h_l^i || W_{lk}h_l^u]))} \tag{2}$$

$$H_{l+1} = GAT(H_l, A; \Theta_l) \tag{3}$$

$$H_{l+1}, C_{l+1} = LSTM(GAT(H_l, A; \Theta_l), (H_l, C_l))$$
$$H_0, C_0 = LSTM(XW_p + [b_p]_N, (0, 0))$$

Datasets

Dataset		# Pos.	# Neu.	# Neg.
Twitter	Train	1561	3127	1560
	Test	173	346	173
Lap14	Train	994	464	870
	Test	341	169	128
Rest14	Train	2164	637	807
	Test	728	196	196
Rest15	Train	912	36	256
	Test	326	34	182
Rest16	Train	1240	69	439
	Test	469	30	117

Results

Model	TWITTER		LAP14		REST14		REST15		REST16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
SVM	63.40♯	63.30♯	70.49♮	N/A	80.16♮	N/A	N/A	N/A	N/A	N/A
LSTM	69.56	67.70	69.28	63.09	78.13	67.47	77.37	55.17	86.80	63.88
MemNet	71.48	69.90	70.64	65.17	79.61	69.64	77.31	58.28	85.44	65.99
AOA	72.30	70.20	72.62	67.52	79.97	70.42	78.17	57.02	87.50	66.21
IAN	**72.50**	**70.81**	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
TNet-LF	**72.98**	**71.43**	**74.61**	**70.14**	80.42	71.03	78.47	59.47	**89.07**	**70.43**
ASCNN	71.05	69.45	72.62	66.72	**81.73**	**73.10**	78.47	58.90	87.39	64.56
ASGCN-DT	71.53	69.68	74.14†	69.24†	**80.86**‡	**72.19**‡	**79.34**†‡	**60.78**†‡	88.69†	66.64†
ASGCN-DG	72.15†	70.40†	**75.55**†‡	**71.05**†‡	80.77‡	72.02‡	**79.89**†‡	**61.89**†‡	**88.99**†	**67.48**†

a possible reason, supect, conjecture, ...

Ablation Study

Model	TWITTER		LAP14		REST14		REST15		REST16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
BiLSTM+Attn	71.24	69.55	72.83	67.82	79.85	70.03	78.97	58.18	87.28	68.18
ASGCN-DG	72.15	70.40	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
ASGCN-DG w/o pos.	72.69	70.59	73.93	69.63	81.22	72.94	79.58	61.55	88.04	66.63
ASGCN-DG w/o mask	72.64	70.63	72.05	66.56	79.02	68.29	77.80	57.51	86.36	61.41
ASGCN-DG w/o GCN	71.92	70.63	73.51	68.83	79.40	69.43	79.40	61.18	87.55	66.19

- Removal of pos → performance increases on Twitter and Rest14.
- The GCN does not work well as expected on the datasets not sensitive to syntax information.

Case Study

Model	Aspect	Attention visualization	Prediction	Label
MemNet	food	great food but the service was dreadful !	negative✗	positive
	staff	The staff should be a bit more friendly .	positive✗	negative
	Windows 8	Did not enjoy the new Windows 8 and touchscreen functions .	positive✗	negative
IAN	food	great food but the service was dreadful !	positive✓	positive
	staff	The staff should be a bit more friendly .	positive✗	negative
	Windows 8	Did not enjoy the new Windows 8 and touchscreen functions .	neutral✗	negative
ASCNN	food	great food but the service was dreadful !	positive✓	positive
	staff	The staff should be a bit more friendly .	neutral✗	negative
	Windows 8	Did not enjoy the new Windows 8 and touchscreen functions .	negative✓	negative
ASGCN-DG	food	great food but the service was dreadful !	positive✓	positive
	staff	The staff should be a bit more friendly .	negative✓	negative
	Windows 8	Did not enjoy the new Windows 8 and touchscreen functions .	negative✓	negative

0/3

1/3

2/3

3/3

- long-range
- multi-word

GGeut