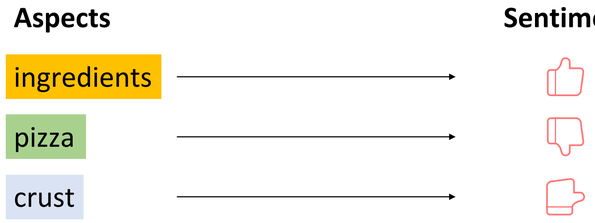
Syntax-Aware Aspect-Level Sentiment Classification with Proximity-Weighted Convolution Network

Chen Zhang[1], Qiuchi Li[2], Dawei Song[1*]

[1]Beijing Institute of Technology & Zhejiang Lab, [2]University of Padua. [*]Corresponding author

BACKGROUND

"They use fancy ingredients but even fancy ingredients don't make for good pizza unless someone knows how to get the crust right."



MOTIVATION

Limitations of the State of the Art

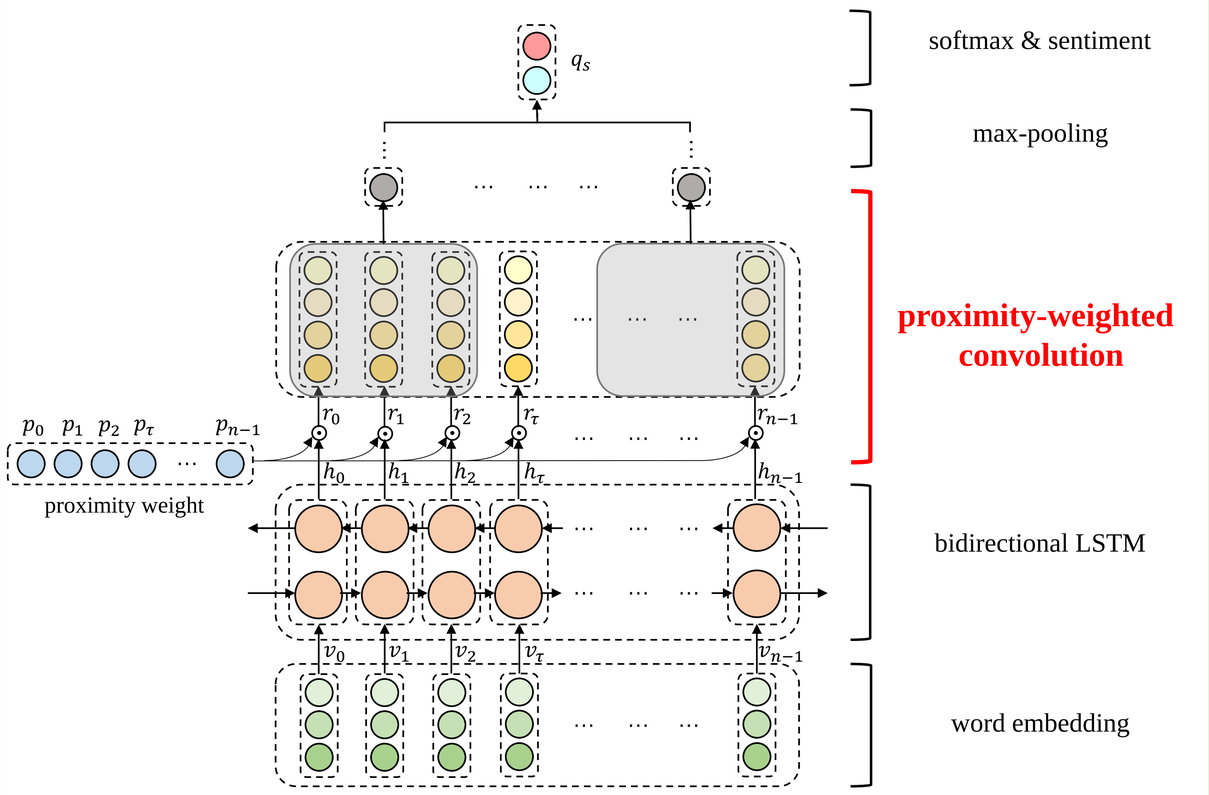
- **Syntax** has been generally neglected in aspect-level sentiment classification.
- The sentiment polarity of an aspect needs to be determined by key **phrases** instead of single **words**.

Research Questions

- How to capture syntactic information and n-gram level features in relation to aspects in a unified framework?
- Can syntactic information help improve aspect-level sentiment classification?
- Will n-gram level features work better than word level ones?

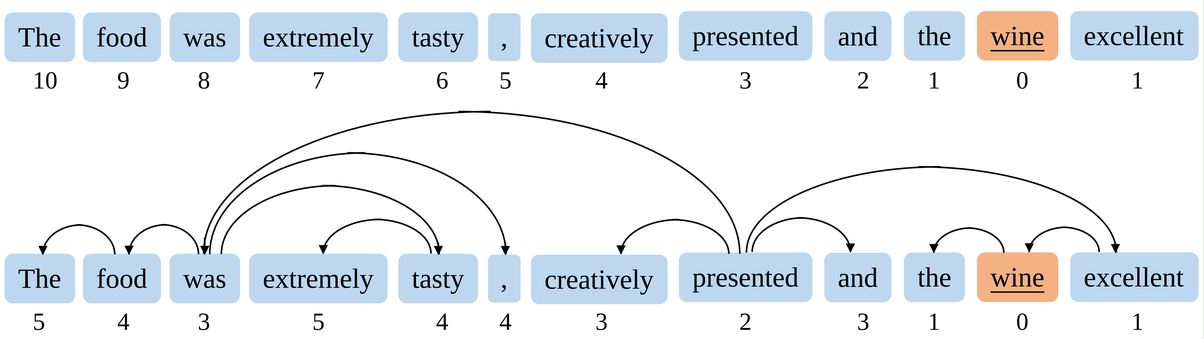
MODEL

Proximity-**W**eighted **C**onvolution **N**etwork (**PWCN**)



Proximity weight -- Capturing Syntactic Information

- Position / dependency distance



- Position / dependency proximity weight

$$p_i = 1 - \frac{d_i}{n}$$

p_i -- proximity weight, d_i -- position / dependency distance, n -- sentence length.

Accordingly, we have two variants of **PWCN**: **PWCN-Pos** and **PWCN-Dep**

Proximity-weighted convolution -- Capturing n-gram level features

- Proximity weight assigning

$$r_i = h_i \cdot p_i$$

r_i -- proximity-weighted representation, h_i -- hidden states output by bidirectional LSTM.

- Convolution

EXPERIMEMTS

Datasets

- Experiments on two benchmarking datasets from **SemEval 2014**.
- The datasets consist of reviews and comments from two categories: **laptop** and **restaurant**, respectively.

Baselines

- **LSTM** only uses the last hidden state vector to predict sentiment polarity.
- **RAM** considers hidden state vectors of context as external memory and applies Gated Recurrent Unit (GRU) structure to multi-hop attention. The top-most representation is used for predicting polarity.
- **IAN** models attention between aspect and its context interactively with two LSTMs.
- **TNet-LF** leverages Context-Preserving Transformation to preserve and strengthen the informative part of context. It also benefits from a multi-layer architecture.

Variants of PWCN-Pos

- **Att-PWCN-Pos:** the proximity weight is **multiplied by the normalized attention weight**, to integrate semantic relatedness and syntax relationship.
- **Point-PWCN-Pos**: n-gram level convolution is set to **1-gram level**, which degrades the convolution process to point-wise (word level) feed-forward network.

Main results

Average accuracy and macro-F1 score over 3 runs with random initialization. The best results are in bold. The marker † refers to p-value < 0.05 when comparing with IAN, while the marker ‡ refers to p-value < 0.05 when comparing with TNet-LF. The relative increase over the LSTM baseline is given in bracket.

Model	Laptop		Restaurant	
	Acc	Macro-F1	Acc	Macro-F1
LSTM	69.63	63.51	77.99	66.91
RAM	72.81(+4.57%)	68.59(+8.00%)	79.89(+2.44%)	69.49(+3.86%)
IAN	71.63(+2.87%)	65.94(+3.83%)	78.59(+0.77%)	68.41(+2.24%)
TNet-LF	75.16(+7.94%)	71.10(+11.95%)	80.20(+2.83%)	70.78(+5.78%)
Att-PWCN-Pos	72.92(+4.72%)	68.14(+7.29%)	80.15(+2.77%)	70.17(+4.87%)
Point-PWCN-Pos	74.45(+6.92%)	69.47(+9.38%)	80.00(+2.58%)	69.93(+4.51%)
PWCN-Pos	75.23†(+8.17%)	70.71†‡(+11.34%)	**81.12**†‡(4.01%)	71.81†(+7.32%)
PWCN-Dep	**76.12**†‡(+9.32%)	**72.12**†(+13.56%)	80.96†(+3.81%)	**72.21**†(+7.92%)

Impact of Syntax

Visualization of a case with respect to *food*

Method	Visualization	Prediction
Att.	great food but the service was dreadful !	Negative
Pos.	great food but the service was dreadful !	Positive
Dep.	great food but the service was dreadful !	Positive

CONCLUSIONS AND FUTURE WORK

Conclusions

- We have proposed a framework that leverages n-gram information and syntactic dependency between aspect and contextual terms into an applicable model.
- Experimental results have demonstrated the effectiveness of our proposed models and suggested that syntactic dependency is more beneficial to aspect-level sentiment classification than semantic relatedness.
- N-gram level features are more significant than word level ones in aspect-level sentiment classification.

Future work

- In-depth analysis of the difference between PWCN models and attention-based models to achieve a deep understanding of where the syntactical dependencies overwhelm semantic relatedness.

ACKNOWLEDGEMENTS

- This work is supported by The National Key Research and Development Program of China (grant No. 2018YFC0831700), Natural Science Foundation of China (grant No. U1636203), and Major Project Program of Zhejiang Lab (grant No. 2019DH0ZX01).
- Co-author Chen Zhang has received Travel Grants from ACM SIGIR.

dwsong@bit.edu.cn