Document Modeling with Gated Recurrent Neural Network for Sentiment Classification

Proceedings of the 2015 conference on empirical methods in natural language processing

Duyu Tang, Bing Qin, Ting Liu

Harbin Institue of Technology, Harbin, China

2015

Introduction

- Infant attempt combines neural networks to sentimental analysis
- Document level sentiment classification - rating 1-5 or 1-10
- bottom-up fashion algorithm
  - Learns sentence representation w/ CNN or LSTM
  - Sementics of sentences and their relatios are adaptively encoded with GRU
- Achieve state-of-the-arts results for IMDB and Yelp datasets.

Overview

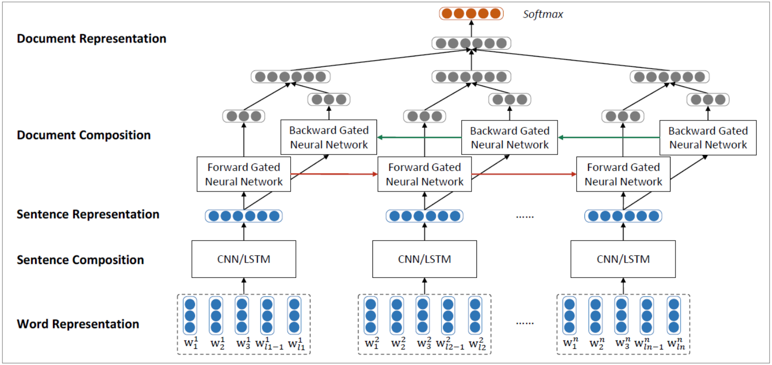

Figure: Fig. 1 of thesis

Sentence Composition

- Sentence embedding from word embedding
  - Word embedding: word2vec
- CNN and LSTM method

CNN Sentence Composition

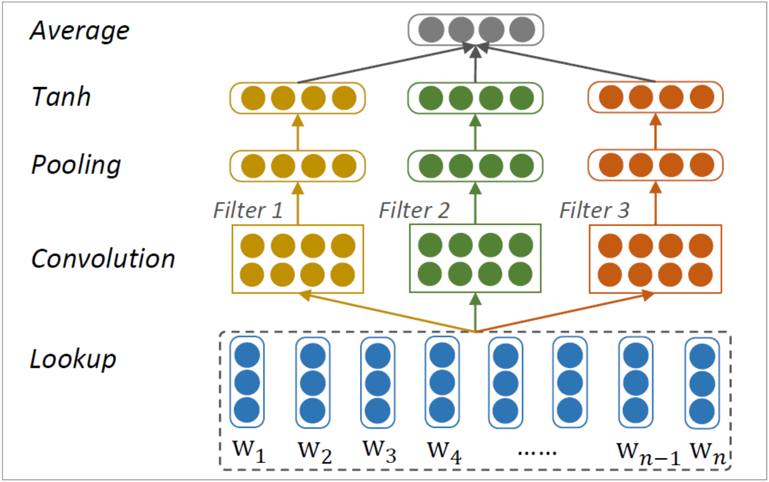

Figure: Fig. 2 of thesis. Filter: *n*-gram

Document composition

- Simple averaging sentence vector fails to capture complex linguistic relations(e.g. “cause” and “contrast”)
- Standard RNN suffers from gradient vanishing / exploding
- GatedRNN, Average, Bidirectional

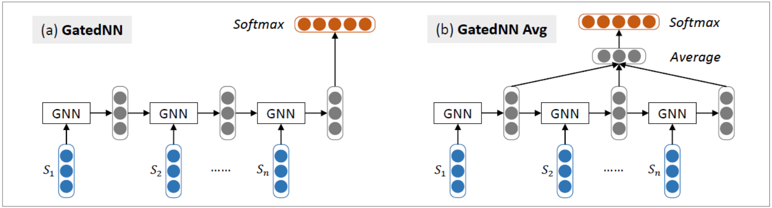

Figure: Fig. 3 of thesis

Datasets

Corpus	#docs	#s/d	#w/d	$\|V\|$	#class	Class Distribution
Yelp 2013	335,018	8.90	151.6	211,245	5	.09/.09/.14/.33/.36
Yelp 2014	1,125,457	9.22	156.9	476,191	5	.10/.09/.15/.30/.36
Yelp 2015	1,569,264	8.97	151.9	612,636	5	.10/.09/.14/.30/.37
IMDB	348,415	14.02	325.6	115,831	10	.07/.04/.05/.05/.08/.11/.15/.17/.12/.18

Figure: Fig. 4 of thesis

Comparison to Other Methods

	Yelp 2013		Yelp 2014		Yelp 2015		IMDB	
	Accuracy	MSE	Accuracy	MSE	Accuracy	MSE	Accuracy	MSE
Majority	0.356	3.06	0.361	3.28	0.369	3.30	0.179	17.46
SVM + Unigrams	0.589	0.79	0.600	0.78	0.611	0.75	0.399	4.23
SVM + Bigrams	0.576	0.75	0.616	0.65	0.624	0.63	0.409	3.74
SVM + TextFeatures	0.598	0.68	0.618	0.63	0.624	0.60	0.405	3.56
SVM + AverageSG	0.543	1.11	0.557	1.08	0.568	1.04	0.319	5.57
SVM + SSWE	0.535	1.12	0.543	1.13	0.554	1.11	0.262	9.16
JMARS	N/A	–	N/A	–	N/A	–	N/A	4.97
Paragraph Vector	0.577	0.86	0.592	0.70	0.605	0.61	0.341	4.69
Convolutional NN	0.597	0.76	0.610	0.68	0.615	0.68	0.376	3.30
Conv-GRNN	0.637	0.56	0.655	0.51	0.660	0.50	0.425	**2.71**
LSTM-GRNN	**0.651**	**0.50**	**0.671**	**0.48**	**0.676**	**0.49**	**0.453**	3.00

Figure: Fig. 5 of thesis

Model Analysis

	Yelp 2013		Yelp 2014		Yelp 2015		IMDB	
	Accuracy	MSE	Accuracy	MSE	Accuracy	MSE	Accuracy	MSE
Average	0.598	0.65	0.605	0.75	0.614	0.67	0.366	3.91
Recurrent	0.377	1.37	0.306	1.75	0.383	1.67	0.176	12.29
Recurrent Avg	0.582	0.69	0.591	0.70	0.597	0.74	0.344	3.71
Bi Recurrent Avg	0.587	0.73	0.597	0.73	0.577	0.82	0.372	3.32
GatedNN	0.636	0.58	0.656	0.52	0.651	0.51	**0.430**	2.95
GatedNN Avg	0.635	0.57	**0.659**	0.52	0.657	0.56	0.416	2.78
Bi GatedNN Avg	**0.637**	**0.56**	0.655	**0.51**	**0.660**	**0.50**	0.425	**2.71**

Figure: Fig. 6 of thesis

Attention-based LSTM for Aspect-level Sentiment Classification

Proceedings of the 2016 conference on empirical methods in natural language processing

Yequan Wang, Minlie Huang, and Li Zhao* and Xiaoyan Zhu

State Key Laboratory on Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, China *Microsoft Research Asia

2016

Introduction

- Aspect-level Sentimental classification:
  One review has various (sometimes opposite) views for different aspects.

  ex) *"The appetizers are ok, but the service is slow"*

  Suggest aspect embedding vector for the first time

- Attention mechanism:
  The model concentrate on different parts of a sentence when different aspects are taken as input.

- LSTM-based sentence analysis

- Experiments

LSTM

- N : length of sentence. $h_i \in \mathbb{R}^d$: hidden, w: word embedding.

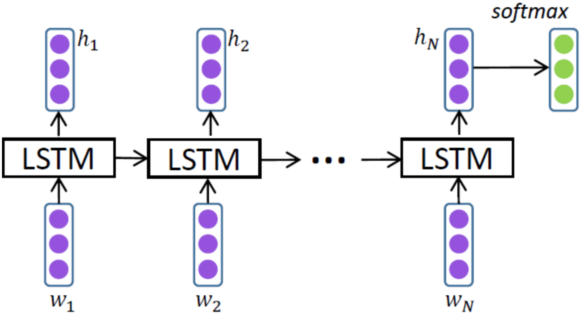

Figure: Fig 1 of thesis

Attension-based LSTM

- $v_{a_i} \in \mathbb{R}^{d_a}$: embedding of aspect i, d_a: dim. of aspect embedding.
  $A = (v'_{a_1}, \cdots, v'_{a_{|A|}}) \in \mathbb{R}^{d_a \times |A|}$: matrix of all aspect embedding.
  $H \in \mathbb{R}^{d \times N} = [h'_1, \cdots, h'_N]$
  $e_N \in \mathbb{R}^N = (1, 1, \cdots, 1)'$
- Attension mechanism: from h_1 to h_N, where should we focus on? (α)

$$M = tanh(\begin{bmatrix} W_h H \\ W_v v_a \otimes e_N \end{bmatrix})$$
$$\alpha = softmax(w^\top M)$$
$$r = H\alpha^T$$

- Final sentence representation: linear combination of r and h_N (not r only)

$$h* = tanh(W_p r + W_x h_N)$$

- output probability

$$y = softmax((W_s h * + b_s)$$

AT-LSTM

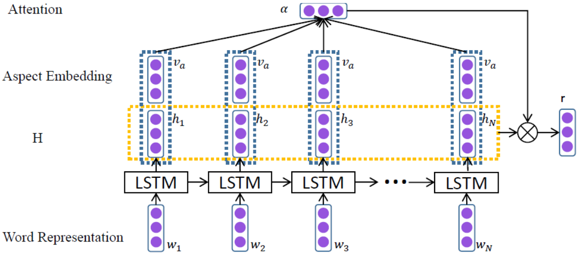

Figure: Fig 2 of thesis. How to achieve r in AT-LSTM alg.

Attention-based LSTM with Aspect Embedding

- In AT-LSTM, v_a only is only used for computing attention weight α.
- By adding v_a as input of LSTM, h can have information of aspects.

ATAE-LSTM

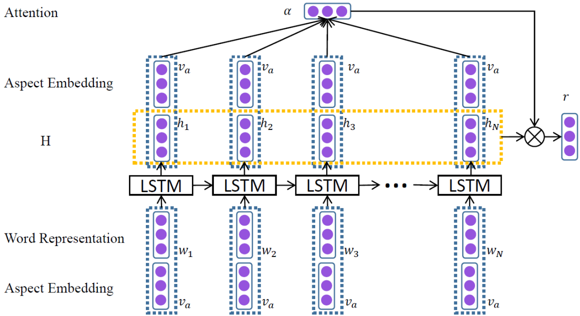

Figure: Fig 3 of thesis. How to achieve r in ATAE-LSTM alg.

Dataset

- All aspects term is fixed

Asp.	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Fo.	867	302	209	69	90	31
Pr.	179	51	115	28	10	1
Se.	324	101	218	63	20	3
Am.	263	76	98	21	23	8
An.	546	127	199	41	357	51
Total	2179	657	839	222	500	94

Figure: Table 1 of thesis

Aspect-level classification

Models	Three-way	Pos./Neg.
LSTM	82.0	88.3
TD-LSTM	82.6	89.1
TC-LSTM	81.9	89.2
AE-LSTM	82.5	88.9
AT-LSTM	83.1	89.6
ATAE-LSTM	**84.0**	**89.9**

Figure: Table 2 of thesis

Aspect-term-level classification

Models	Three-way	Pos./Neg.
LSTM	74.3	-
TD-LSTM	75.6	-
AE-LSTM	76.6	89.6
ATAE-LSTM	**77.2**	**90.9**

Figure: Table 3 of thesis. Aspect term polarity classification about **restaurants**.

Models	Three-way	Pos./Neg.
LSTM	66.5	-
TD-LSTM	68.1	-
AE-LSTM	**68.9**	87.4
ATAE-LSTM	68.7	**87.6**

Figure: Table 4 of thesis. Aspect term polarity classification about **laptops**.

Qualitative analysis

- Visualize α
- Attend proper words for aspects

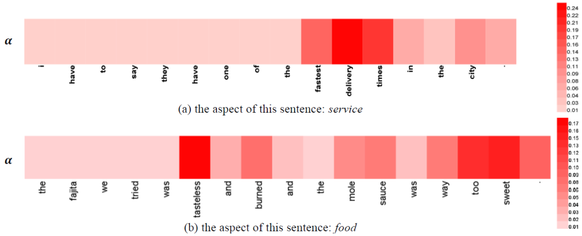

Figure: Fig 4 of thesis.

Case Study

- Sentence with different aspects
- Keypoints are distributed and interpret the word *'not'* correctly
- Long and complicated sentences

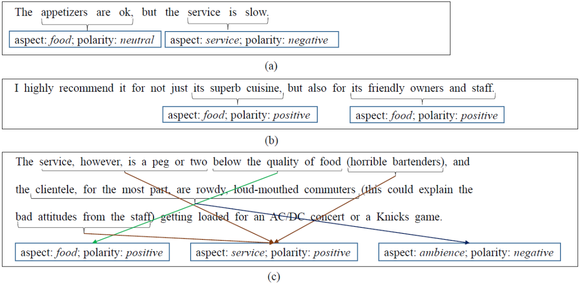

Figure: Fig 5 of thesis.