Summary of Aspect-based Sentiment Analysis

Bin Liang

Harbin Institute of Technology, Shenzhen

2019-06-14

Overview

- Background & Task
- Proposed Methods
- Other Flash

TDSA vs ABSA vs TABSA

I bought a new camera. The ***picture quality*** is **amazing** but the ***battery life*** is too **short**.

Result: ***picture quality*** → ***Positive*** ***battery life*** → ***Negative***

The ***sushi*** seemed **pretty fresh** and was **proportioned**. (disordered)

Result: ***sushi, FOOD#QUALITY*** → ***Positive***

sushi, FOOD#STYLE_OPTIONS → ***Positive (negative)***

TDSA vs ABSA vs TABSA

Fulham (location1) is your *best bet for secure* although *expensive* and **Brent** (location2) is too *far*.

Result: **(location1, SAFETY, Positive)**
(location1, PRICE, Negative)
(location2, TRANSIT-LOCATION, Negative)

Brent is your *best bet for secure* although *expensive* and **Fulham** is too *far*.

TDSA vs ABSA vs TABSA

Model	paper	Acc.	conference
TDLSTM	Effective LSTMs for Target-Dependent Sentiment Classification	75.63	COLING-2016
ATAE-LSTM	Attention-based LSTM for Aspect-level Sentiment Classification	77.2	EMNLP-2016
MemNet	Aspect Level Sentiment Classification with Deep Memory Network	80.95	EMNLP-2016
GCAE	Aspect Based Sentiment Analysis with Gated Convolutional Networks	77.28	ACL-2018
HP-LSTM	A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis	85.3	EMNLP-2016
SenticLSTM	Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM	76.47	AAAI-2018

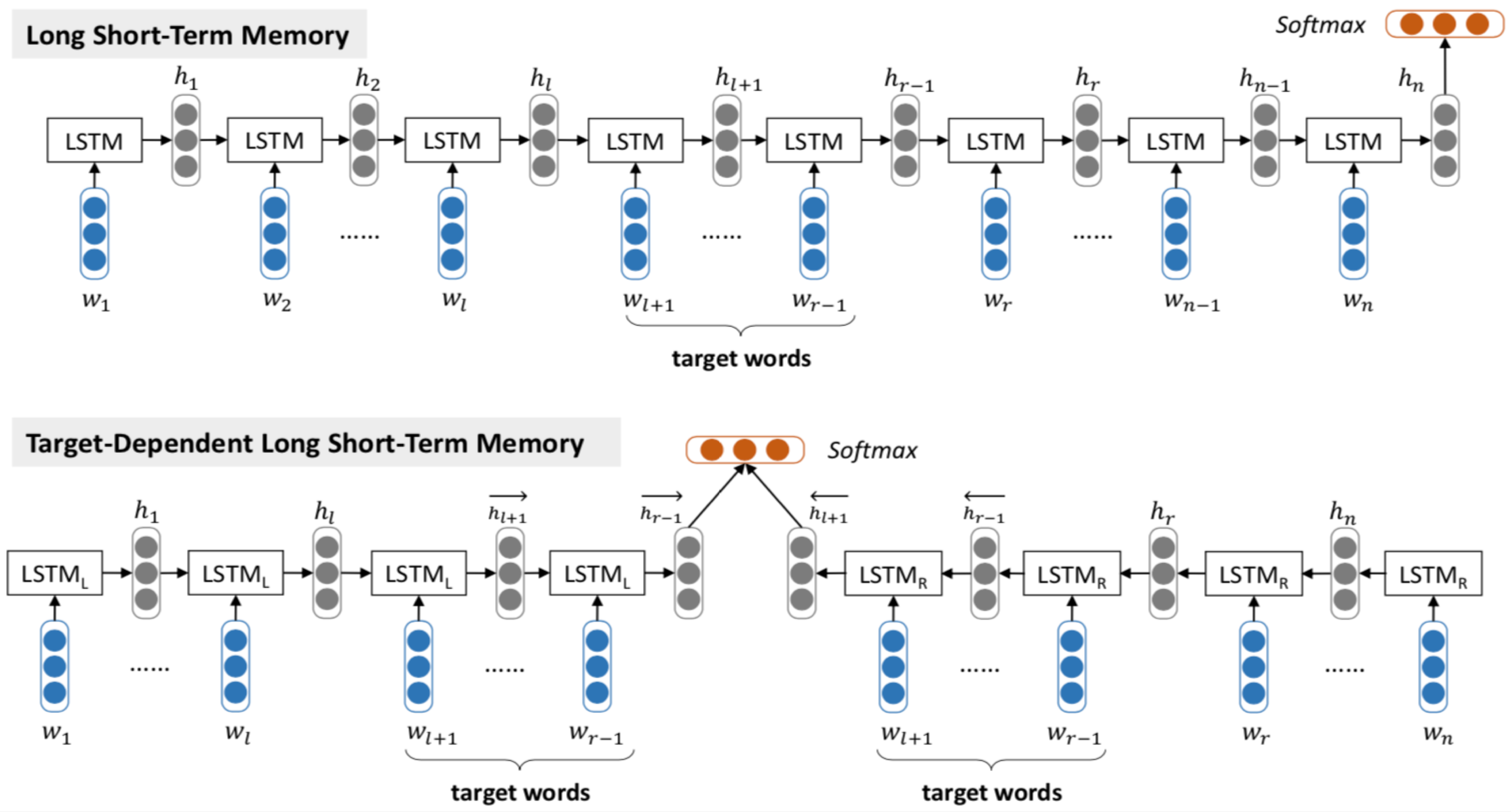
https://github.com/Derll/Sentiment-analysis-statistics

TDLSTM: Motivation

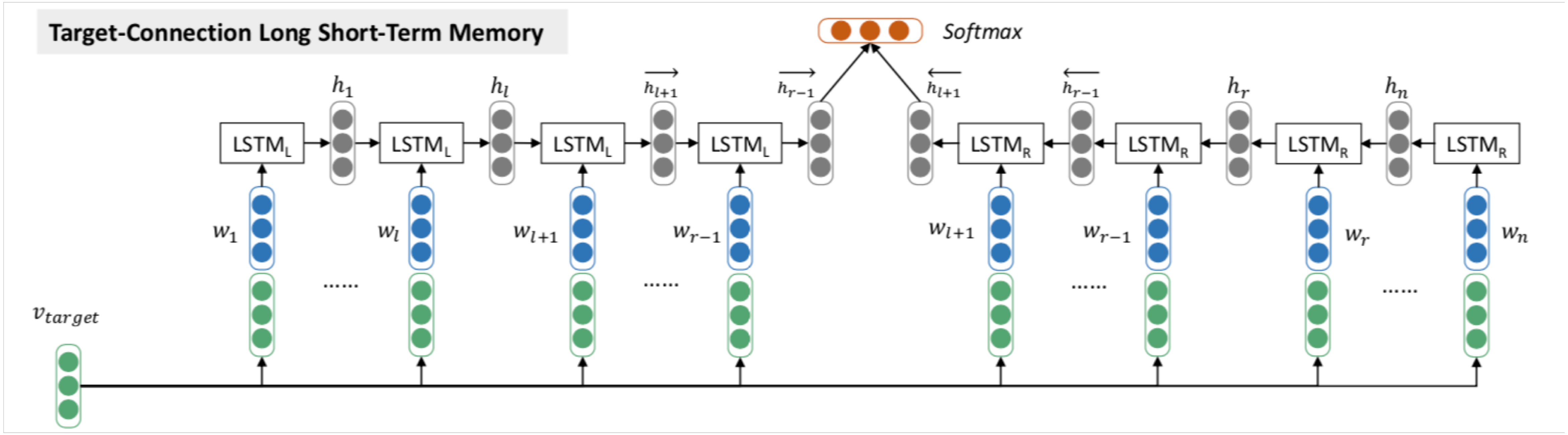
- How to effectively model the semantic relatedness of a target word with its context words in a sentence
- A person asked to do this task will naturally “look at” parts of relevant context words which are helpful to determine the sentiment polarity of a sentence towards the target

Effective LSTMs for Target-Dependent Sentiment Classification, COLING 2016

TDLSTM: Model



TDLSTM: TCLSTM



Context plus target string

TDLSTM: Result

Method	Accuracy	Macro-F1
SVM-indep	0.627	0.602
SVM-dep	0.634	0.633
Recursive NN	0.630	0.628
AdaRNN-w/oE	0.649	0.644
AdaRNN-w/E	0.658	0.655
AdaRNN-comb	0.663	0.659
Target-dep	0.697	0.680
Target-dep^{+}	0.711	**0.699**
LSTM	0.665	0.647
TD-LSTM	0.708	0.690
TC-LSTM	**0.715**	0.695

TDLSTM: Inspiration

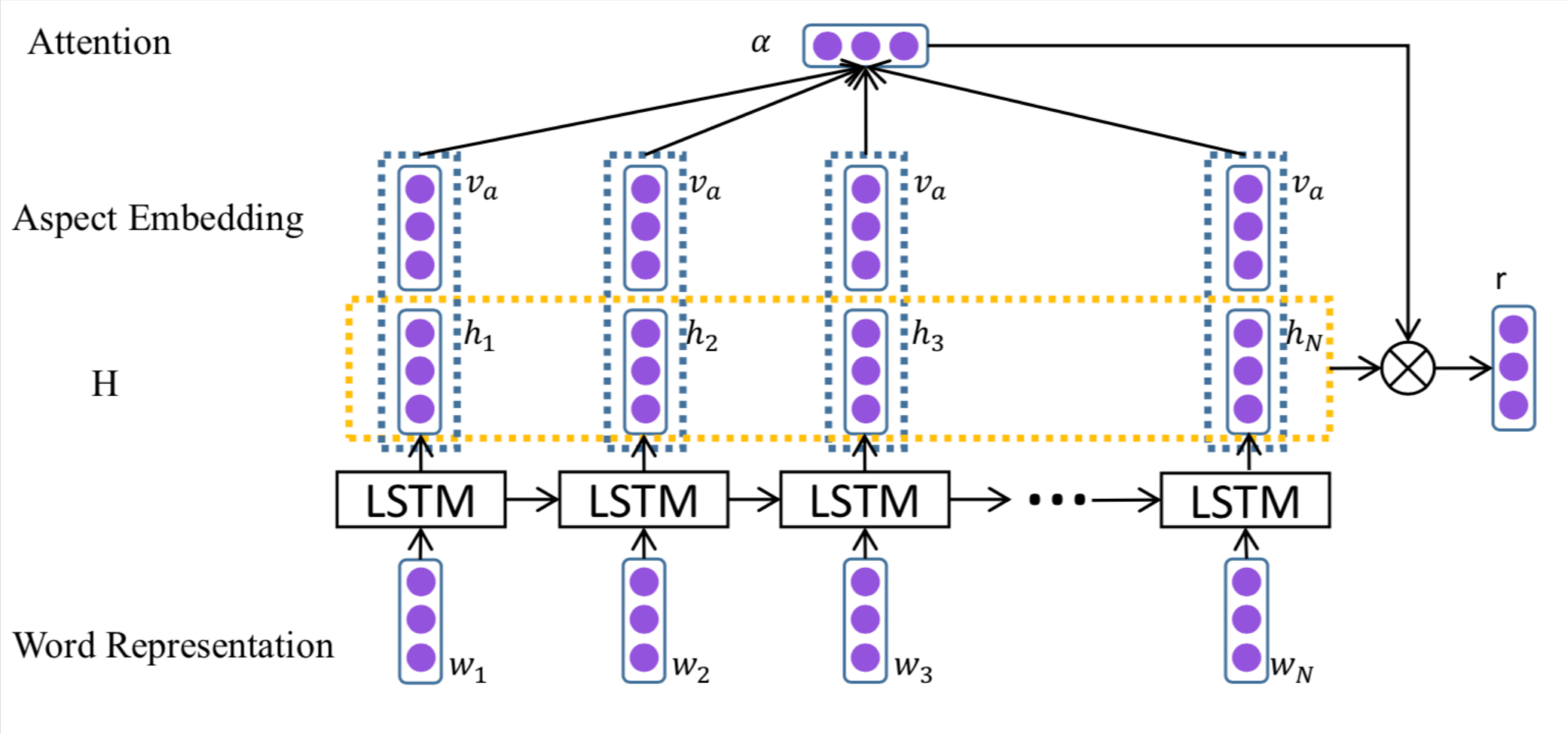
- Different outputs can be achieved from different inputs
- The interrelation of different targets in the same sentence

ATAE-LSTM: Motivation

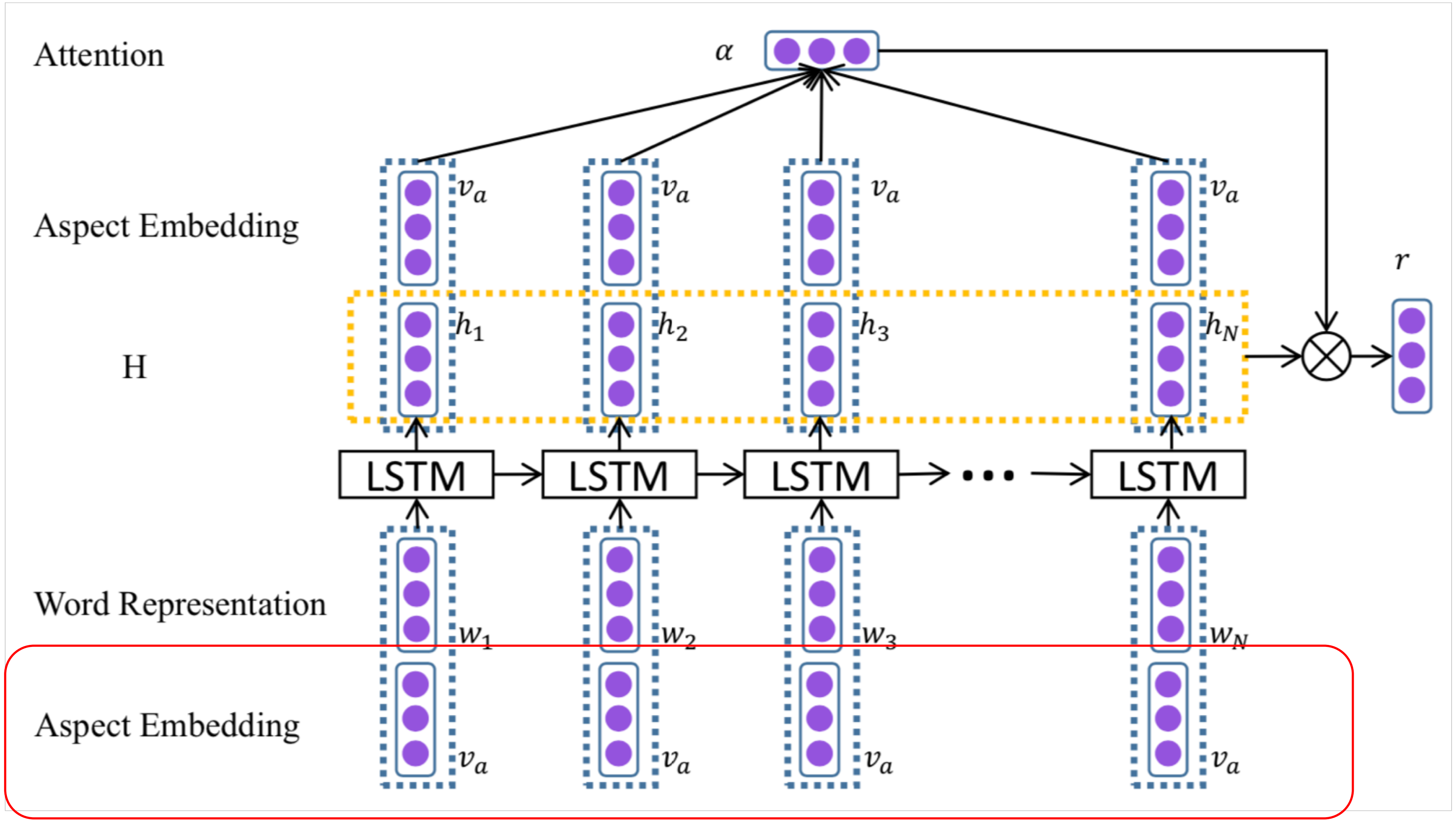
- Previous models can only take into consideration the target but not aspect information which is proved to be crucial for aspect-level classification
- The attention-based models are able to attend different parts of a sentence when different aspects are concerned
- It is the first time to propose aspect embedding

Attention-based LSTM for Aspect-level Sentiment Classification, EMNLP 2016

ATAE-LSTM: Models



ATAE-LSTM: Models



ATAE-LSTM: Models

$$M = tanh(\begin{bmatrix} W_h H \\ W_v v_a \otimes e_N \end{bmatrix})$$

$$\alpha = softmax(w^T M)$$

$$r = H\alpha^T$$

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

The attention mechanism allows the model to capture the most important part of a sentence when different aspects are considered.

ATAE-LSTM: Result

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**

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

The gap between the results of the three-way classification and binary classification is still evident.

ATAE-LSTM: Inspiration

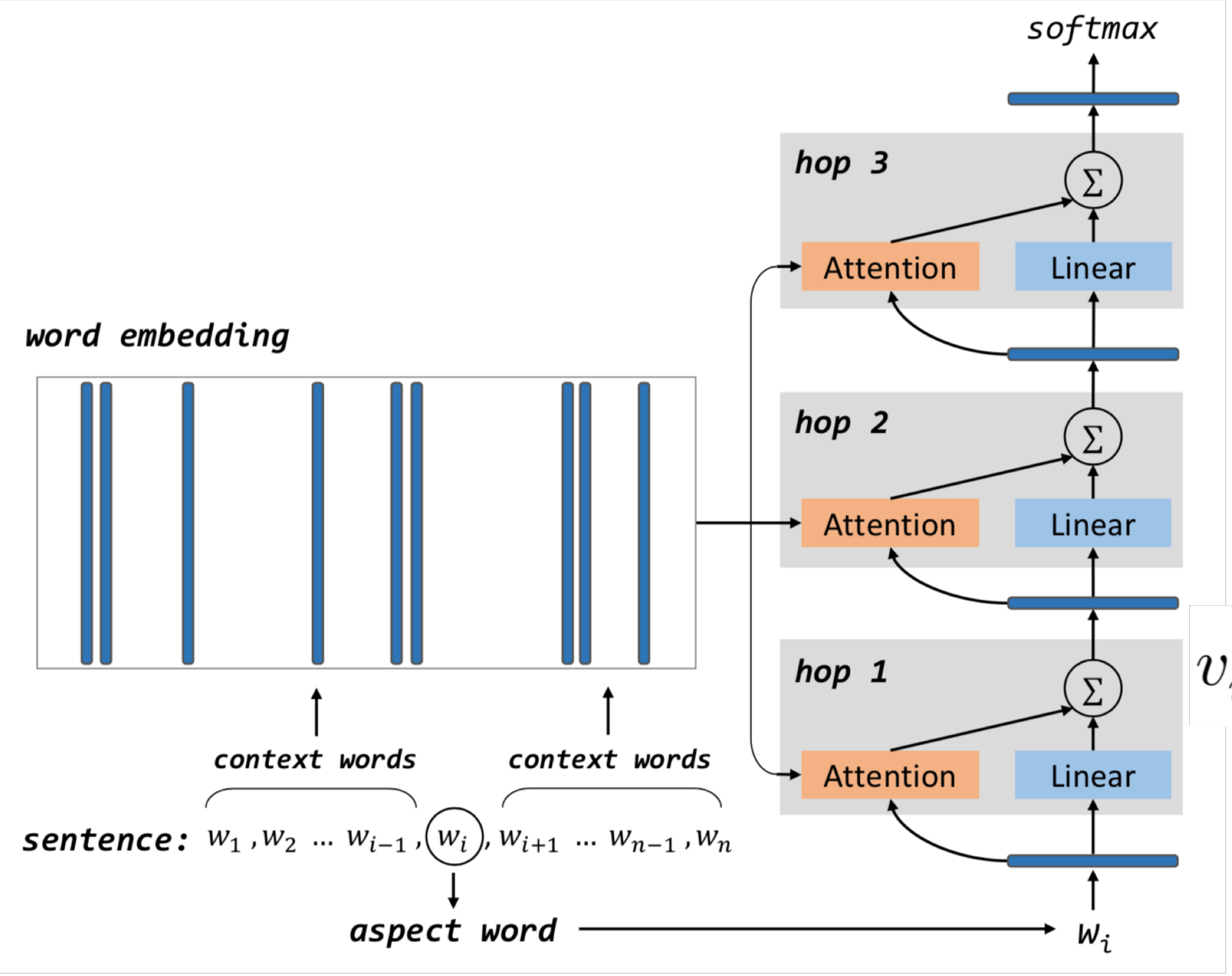
- Attending different parts of a sentence when different aspects or targets are concerned is important for ABSA.
- This paper uses the whole sentence to construct the attention mechanism. It does not fully consider the situation that there are multiple targets and multiple aspects in the sentence.
- How to obtain the correlation between target (aspect term) and aspect?
- How to distinguish different relations of targets and aspects?
- How to improve the performance of three-way (multi) classification?

MemNet: Motivation

- Standard LSTM works in a sequential way and manipulates each context word with the same operation, so that it can not explicitly reveal the importance of each context word.
- A desirable solution should be capable of explicitly capturing the importance of context words and using that information to build up features for the sentence after given an aspect word.
- Data-driven, computationally. The first attention-based memory work.

Aspect Level Sentiment Classification with Deep Memory Network, EMNLP 2016

MemNet: Models



$$vec = \sum_{i=1}^{k} \alpha_i m_i$$

$$g_i = tanh(W_{att}[m_i; v_{aspect}] + b_{att})$$

$$\alpha_i = \frac{exp(g_i)}{\sum_{j=1}^{k} exp(g_j)}$$

$$m_i = e_i \odot v_i$$

$$m_i = e_i + v_i$$

$$v_i^k = (1 - l_i/n) - (k/d)(1 - 2 \times l_i/n)$$

$$v_i = 1 - l_i/n$$

MemNet: Result

	Laptop	Restaurant
Majority	53.45	65.00
Feature+SVM	**72.10**	**80.89**
LSTM	66.45	74.28
TDLSTM	68.13	75.63
TDLSTM+ATT	66.24	74.31
ContextAVG	61.22	71.33
MemNet (1)	67.66	76.10
MemNet (2)	71.14	78.61
MemNet (3)	71.74	79.06
MemNet (4)	72.21	79.87
MemNet (5)	71.89	80.14
MemNet (6)	72.21	80.05
MemNet (7)	**72.37**	80.32
MemNet (8)	72.05	80.14
MemNet (9)	72.21	**80.95**

Method	Time cost
LSTM	417
TDLSTM	490
TDLSTM + ATT	520
MemNet (1)	3
MemNet (2)	7
MemNet (3)	9
MemNet (4)	15
MemNet (5)	20
MemNet (6)	24
MemNet (7)	26
MemNet (8)	27
MemNet (9)	29

MemNet: Inspiration

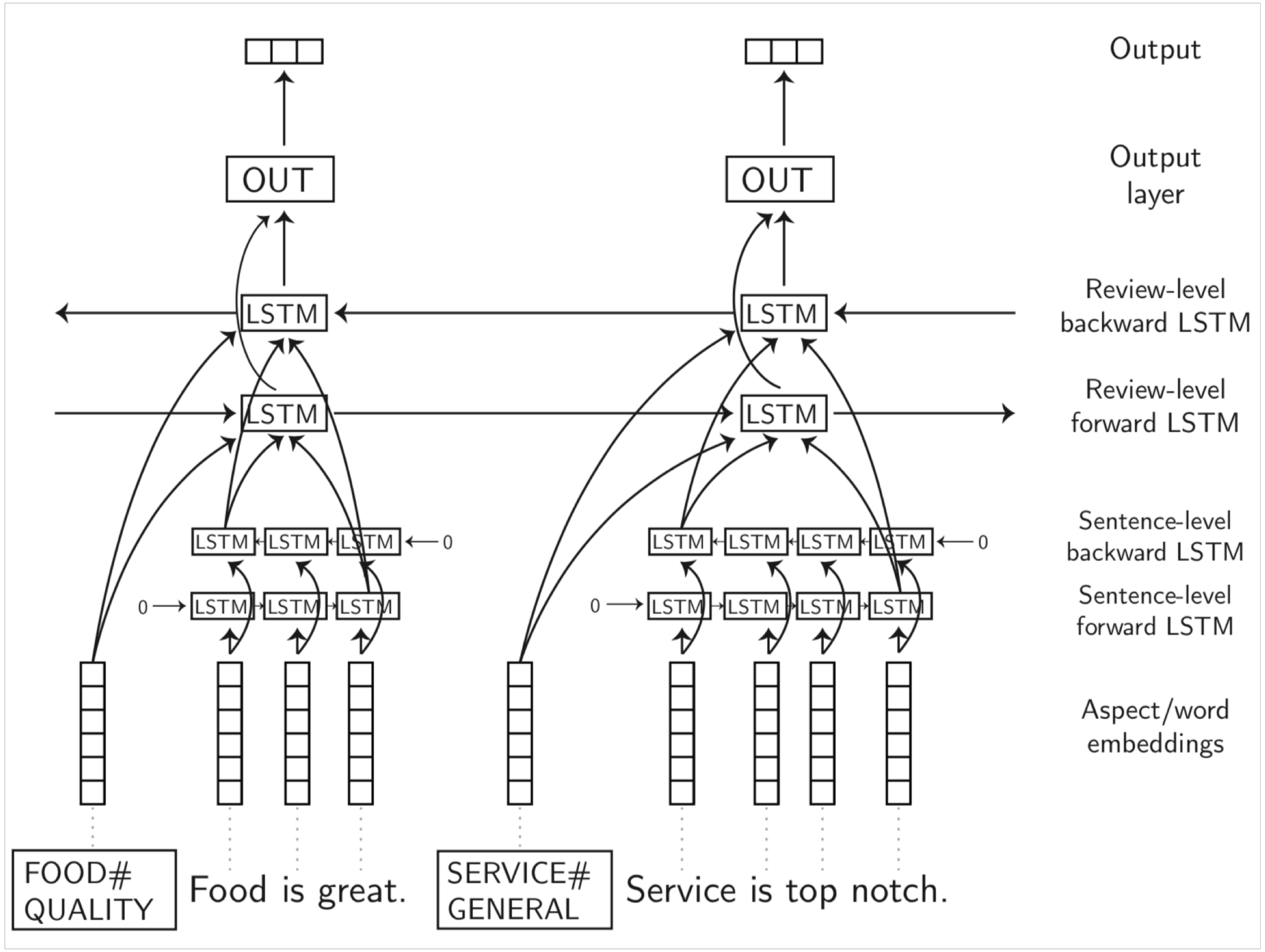
- Useful methods for new tasks. Additionally, some task-specific innovations are added.
- How to make full use of location information for the task of ABSA, TABSA, et al.?
- The improvement of computational performance is also important. (ACL 2019).

HP-LSTM: Motivation

- Sentences in reviews, however, are usually classified independently, even though they form part of a review's argumentative structure.
- Sentences in a review build and elaborate upon each other; knowledge of the review structure and sentential context should thus inform the classification of each sentence.
- To leverage both intra- and inter-sentence relations is important in review-level ABSA.

A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis, EMNLP 2016

HP-LSTM: Model



There is no formula, only simple concatenation!!!

HP-LSTM: Result

Language	Domain	Best	XRCE	IIT-TUDA	CNN	LSTM	H-LSTM	HP-LSTM
English	Restaurants	**88.1**	**88.1**	86.7	82.1	81.4	83.0	85.3
Spanish	Restaurants	**83.6**	-	**83.6**	79.6	75.7	79.5	81.8
French	Restaurants	**78.8**	**78.8**	72.2	73.2	69.8	73.6	75.4
Russian	Restaurants	77.9	-	73.6	75.1	73.9	**78.1**	77.4
Dutch	Restaurants	77.8	-	77.0	75.0	73.6	82.2	**84.8**
Turkish	Restaurants	**84.3**	-	**84.3**	74.2	73.6	76.7	79.2
Arabic	Hotels	82.7	-	81.7	82.7	80.5	82.8	**82.9**
English	Laptops	**82.8**	-	**82.8**	78.4	76.0	77.4	80.1
Dutch	Phones	83.3	-	82.6	83.3	81.8	81.3	**83.6**
Chinese	Cameras	**80.5**	-	-	78.2	77.6	78.6	78.8
Chinese	Phones	73.3	-	-	72.4	70.3	**74.1**	73.3

HP-LSTM: Inspiration

- Attention mechanism.
- How to make full use of relations between different sentences in the same review (document) in the task of review-level (document-level) ABSA

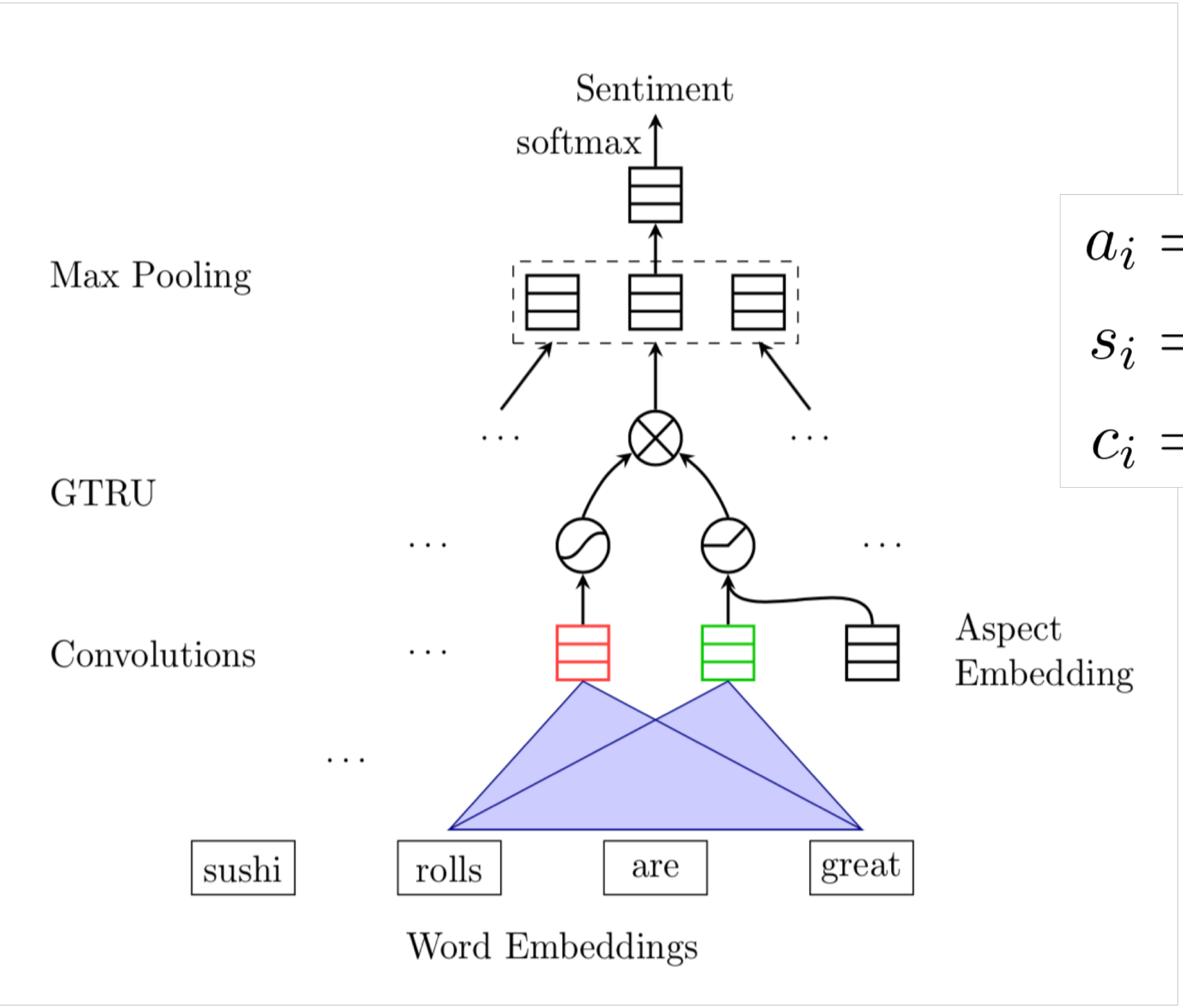
A bad service. ………………. My assessment of food is the same as that of service.

GCAE: Motivation

- LSTM and attention mechanism to predict the sentiment of the concerned targets, which are often complicated, need more training time and cannot take advantage of highly-parallelized sources.
- Models based on convolutions and gating mechanisms, which has much less training time than LSTM based networks, but with better accuracy.
- The first gated CNN model for ABSA

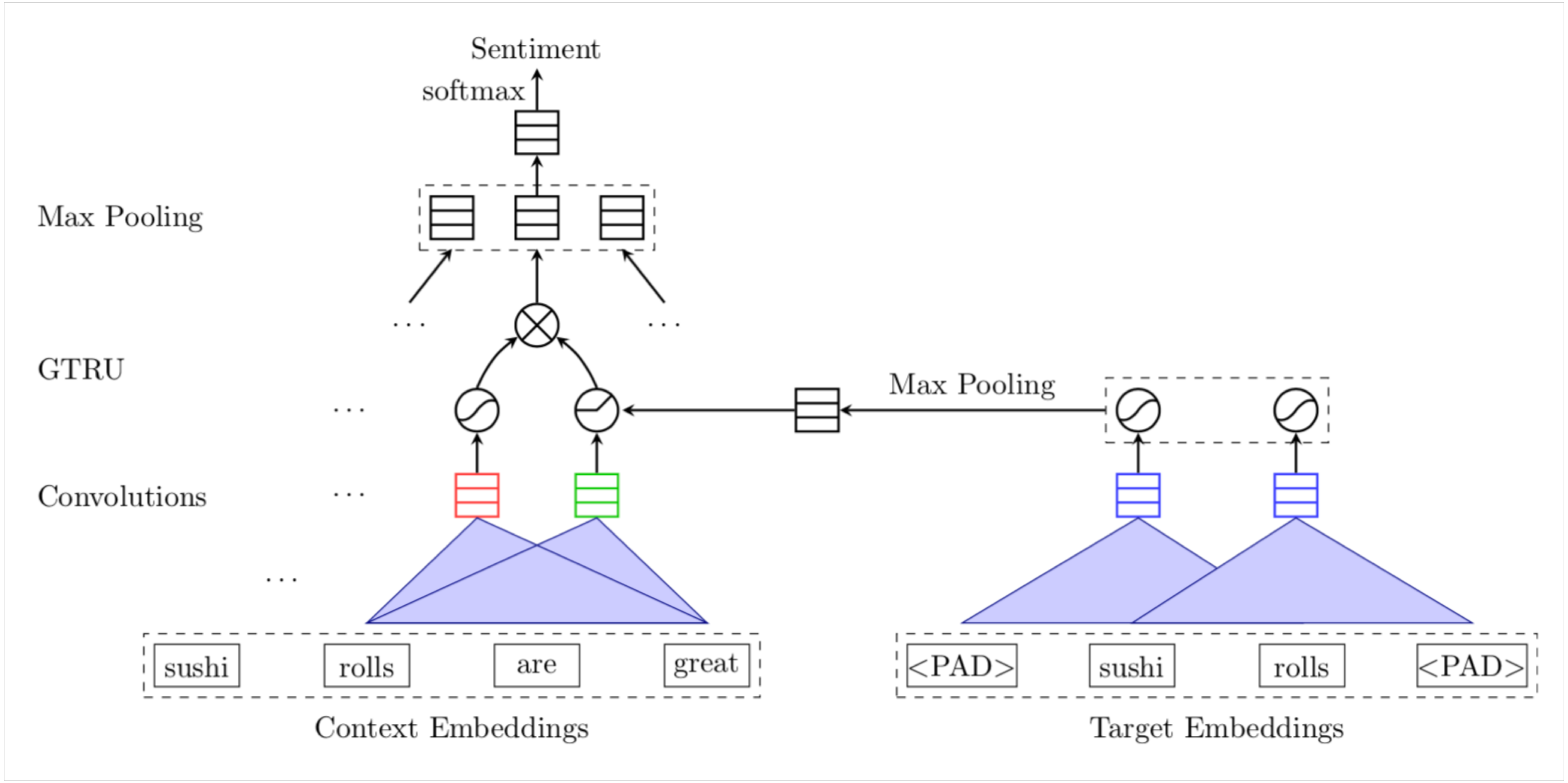
Aspect Based Sentiment Analysis with Gated Convolutional Networks, ACL 2018

GCAE: Model



$$
\begin{aligned}
a_i &= \mathrm{relu}(\mathbf{X}_{i:i+k} * \mathbf{W}_a + \mathbf{V}_a \boldsymbol{v}_a + b_a) \\
s_i &= \tanh(\mathbf{X}_{i:i+k} * \mathbf{W}_s + b_s) \\
c_i &= s_i \times a_i \quad ,
\end{aligned}
$$

GCAE: Model


Sentiment
softmax
Max Pooling
GTRU
Convolutions
Max Pooling
sushi
rolls
are
great
Context Embeddings
<PAD>
sushi
rolls
<PAD>
Target Embeddings

GCAE: Result

Models	Restaurant-Large		Restaurant 2014	
	Test	Hard Test	Test	Hard Test
SVM*	-	-	75.32	-
SVM + lexicons*	-	-	**82.93**	-
ATAE-LSTM	83.91±0.49	66.32±2.28	78.29±0.68	45.62±0.90
CNN	84.28±0.15	50.43±0.38	79.47±0.32	44.94±0.01
GCN	84.48±0.06	50.08±0.31	**79.67±0.35**	44.49±1.52
GCAE	**85.92±0.27**	**70.75±1.19**	79.35±0.34	**50.55±1.83**

Models	Restaurant		Laptop	
	Test	Hard Test	Test	Hard Test
SVM*	77.13	-	63.61	-
SVM + lexicons*	**80.16**	-	**70.49**	-
TD-LSTM	73.44±1.17	56.48±2.46	62.23±0.92	46.11±1.89
ATAE-LSTM	73.74±3.01	50.98±2.27	64.38±4.52	40.39±1.30
IAN	76.34±0.27	55.16±1.97	68.49±0.57	44.51±0.48
RAM	76.97±0.64	55.85±1.60	68.48±0.85	45.37±2.03
GCAE	**77.28±0.32**	**56.73±0.56**	**69.14±0.32**	**47.06±2.45**

Model	ATSA
ATAE	25.28
IAN	82.87
RAM	64.16
TD-LSTM	19.39
GCAE	3.33

GCAE : Inspiration

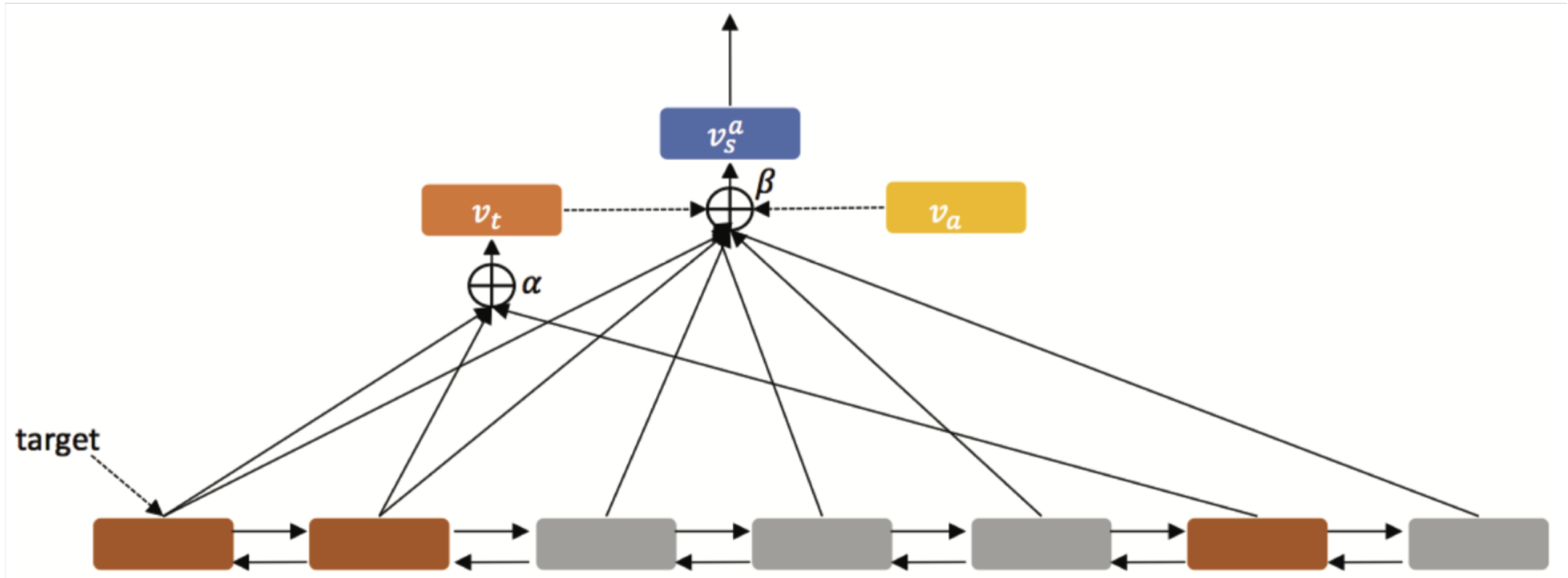
- Using GCNN in ABSA task is an interesting thing.
- There is only one loss function in the model. Can such learning achieve expectations?

SenticLSTM: Motivation

- Existing research assumes all instances are of equal importance and simply computes an average vector over such instances.
- Hierarchical attention exploited by existing methods only implicitly models the process of inferring the sentiment-bearing words related to the given target and aspect as black-box.
- Existing research falls short in effectively incorporating into the deep neural network external knowledge.

Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM, AAAI 2018

SenticLSTM: Model



$$v_t = H'\alpha = \sum_j \alpha_j h_{t_j}$$

$$\alpha = softmax(W_a^{(2)} tanh(W_a^{(1)} H'))$$

$$v_{s,t}^a = H\beta = \sum_i \beta_i h_i$$

$$\beta_a = softmax(v_a^T tanh(W_m(H' \odot v_t)))$$

SenticLSTM: Model

$$
\begin{aligned}
f_i &= \sigma(W_f[x_i, h_{i-1}, \mu_i] + b_f) \\
I_i &= \sigma(W_I[x_i, h_{i-1}, \mu_i] + b_I) \\
\widetilde{C}_i &= tanh(W_C[x_i, h_{i-1}] + b_C) \\
C_i &= f_i * C_{i-1} + I_i * \widetilde{C}_i \\
o_i &= \sigma(W_o[x_i, h_{i-1}, \mu_i] + b_o) \\
o_i^c &= \sigma(W_{co}[x_i, h_{i-1}, \mu_i] + b_{co}) \\
h_i &= o_i * tanh(C_i) + o_i^c * tanh(W_c \mu_i)
\end{aligned}
$$

SenticLSTM: Result

	Aspect Categorization						Sentiment	
	Strict Acc. (%)		Macro F1 (%)		Micro F1 (%)		Sentiment Acc. (%)	
	dev	test	dev	test	dev	test	dev	test
TDLSTM	50.27	50.83	59.03	58.17	55.72	55.78	82.60	81.82
LSTM + TA	54.17	52.02	62.90	61.07	60.56	59.02	83.80	84.29
LSTM + TA + SA	68.83	66.42	79.36	76.69	79.14	76.64	86.00	86.75
LSTM + TA + DMN SA	60.66	60.14	68.89	70.19	67.28	68.37	84.80	83.36
LSTM + TA + SA + KB Feat	**69.38**	64.76	**80.00**	76.33	**79.79**	76.08	87.00	88.70
LSTM + TA + SA + KBA	68.08	65.12	78.68	76.40	78.73	76.46	87.40	87.98
Recall LSTM + TA + SA	68.64	64.66	78.44	75.61	78.53	75.91	86.80	86.85
Sentic LSTM + TA + SA	69.20	**67.43**	78.84	**78.18**	79.09	**77.66**	**88.80**	**89.32**

SenticLSTM : Inspiration

- How to use external knowledge better?
- Since target itself can construct attention information, can the relation between aspect and target construct attention information?

SenticLSTM : Inspiration

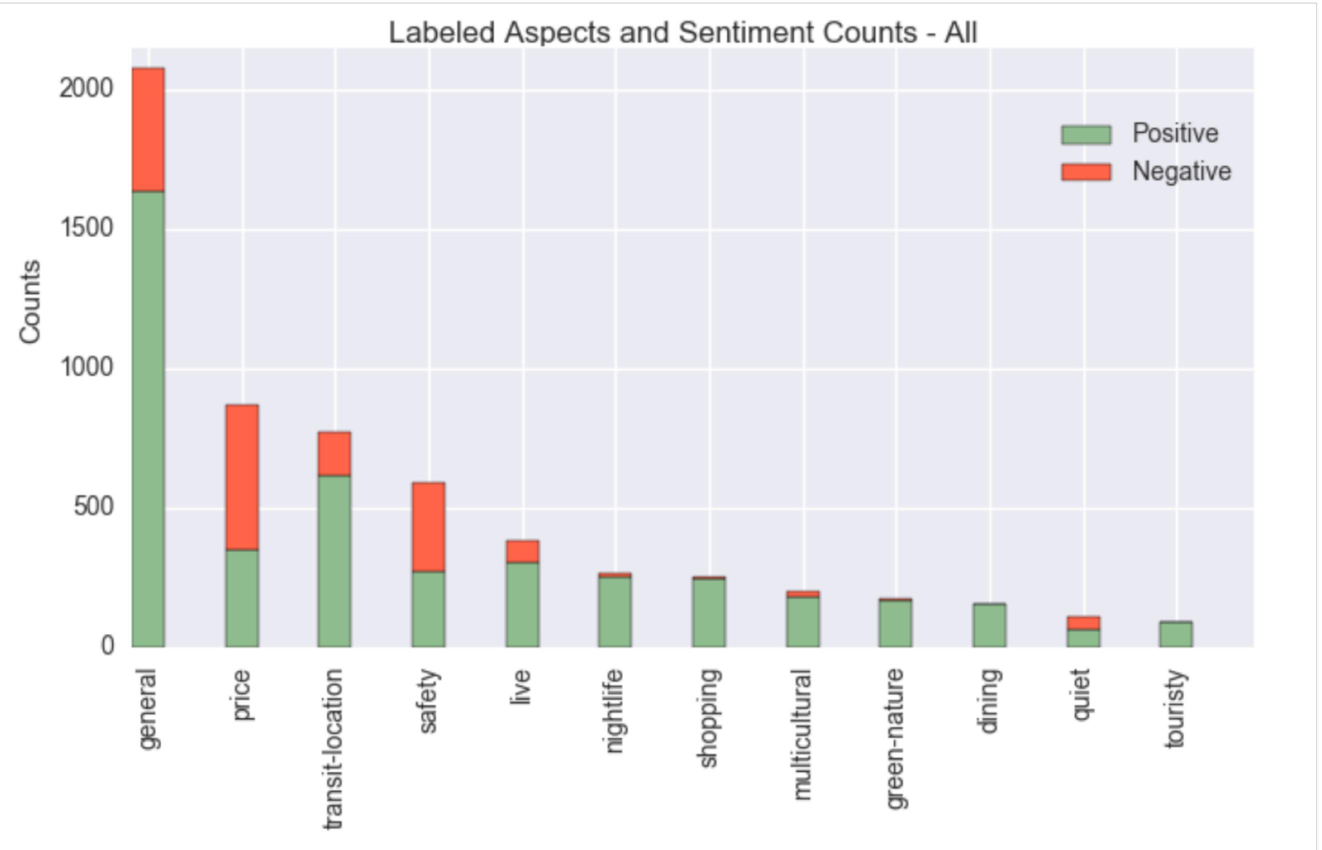
lord_bacon	0	0	0.124	0	0	0	0	0	0	0	0	0	0	0	0.615	0	0
circuitry	0.157	0	0	0	0	0.169	0.285	0	0	0	0	0.165	0	0	0.344	0	0
congressional_gold_medal	0	0	0	0	0	0.198	0	0	0.37	0	0	0	0	0	0	0	0
final_buzzer	0.177	0	0	0.134	0	0	0.194	0	0	0.081	0.047	0.098	0	0.082	0.142	0	0
use_match	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hastily	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.534	0	0
comically	0	0	0	0.091	0	0	0	0	0	0	0	0	0	0	0.501	0	0
abductor_muscle	0	0	0	0	0	0.076	0.132	0	0	0	0	0	0	0	0	0	0.169
regularize	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.618	0	0
fibrillose	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Other Flash

Methods	D1 Acc.	D1 Macro-F1	D2 Acc.	D2 Macro-F1	D3 Acc.	D3 Macro-F1	D4 Acc.	D4 Macro-F1
Feature-based SVM	80.16	NA	70.49	NA	NA	NA	NA	NA
LSTM	75.23	64.21	66.79	64.02	75.28	54.10	81.94	58.11
LSTM+ATT	76.83	66.48	68.07	65.27	77.38	60.52	82.73	59.12
TDLSTM	75.37	64.51	68.25	65.96	76.39	58.70	82.16	54.21
TDLSTM+ATT	75.66	65.23	67.82	64.37	77.10	59.46	83.11	57.53
ATAE-LSTM	78.60	67.02	68.88	65.93	78.48	62.84	83.77	61.71
MM	76.87	66.40	68.91	63.95	77.89	59.52	83.04	57.91
RAM	78.48	68.54	72.08	68.43	79.98	60.57	83.88	62.14
Ours: LSTM+ATT+TarRep	78.95	68.67	70.69	66.59	80.05	**68.73**	84.24	**68.62**
Ours: LSTM+SynATT	80.45	71.26	**72.57**	69.13	80.28	65.46	83.39	66.83
Ours: LSTM+SynATT+TarRep	**80.63***	**71.32***	71.94	**69.23**	**81.67***	66.05*	**84.61***	67.45*

Effective Attention Modeling for Aspect-Level Sentiment Classification, COLING 2018

Other Flash



SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods, AAAI 2016

Q & A