Aspect Based Sentiment Analysis of Reviews

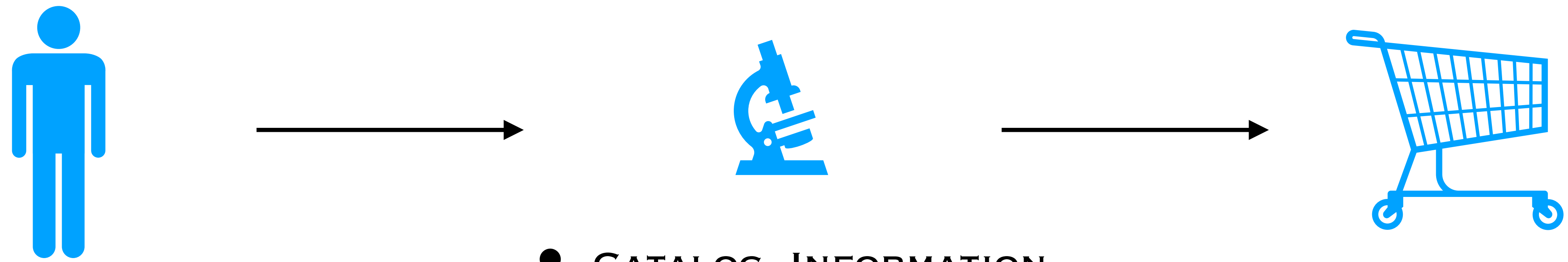
Shreyas Shetty M

23 January, 2020
DL for NLP
IIT Patna

Outline

- Aspect Based Sentiment Analysis
- (Quick) Background Refresher
- SemEval 2014 Challenge
- Approaches to model ABSA

Aspect Based Sentiment Analysis



Customer

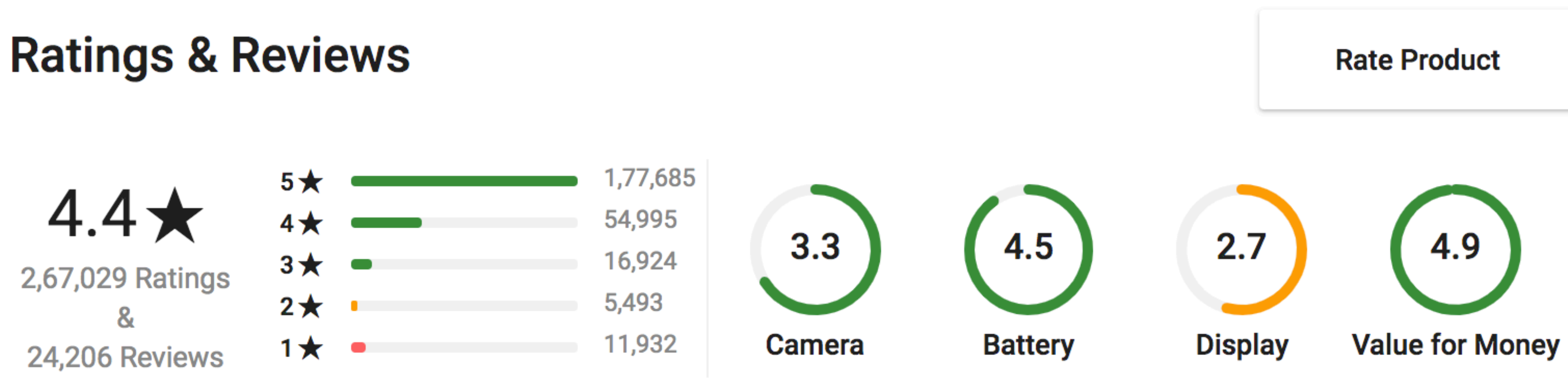
- Catalog Information
- Product Comparison
- Product Reviews

Purchase

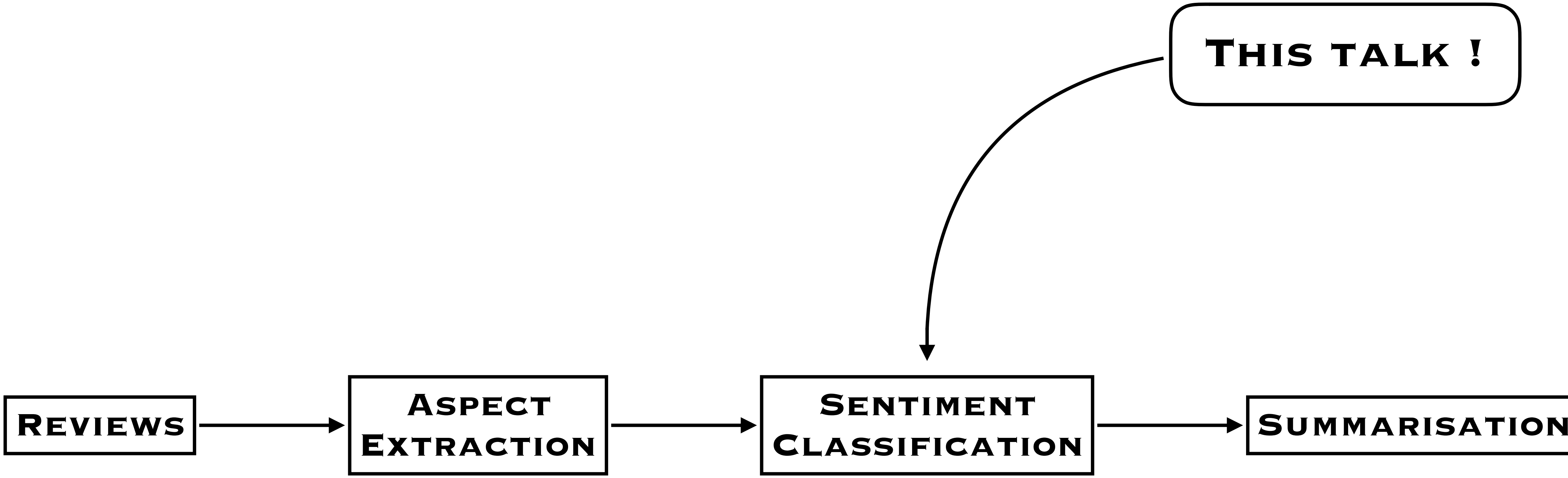
Product Reviews

Crucial for purchase decisions
BUT
Can be potentially too many in number

Aspect (Feature) Based Review Summarisation



Aspect Level Summary For a Mobile Phone



The camera is great but the battery is terrible.

- camera
- battery

- camera (positive)
- battery (negative)

Aspect Based Sentiment Analysis (ABSA)

Predict sentiment corresponding to aspect(s) in a review

*SemEval 2014 Task 4 : Subtask 2

ABSA challenge for reviews from Restaurant and Laptop domain
Classify aspects to Positive, Negative and Neutral Sentiments

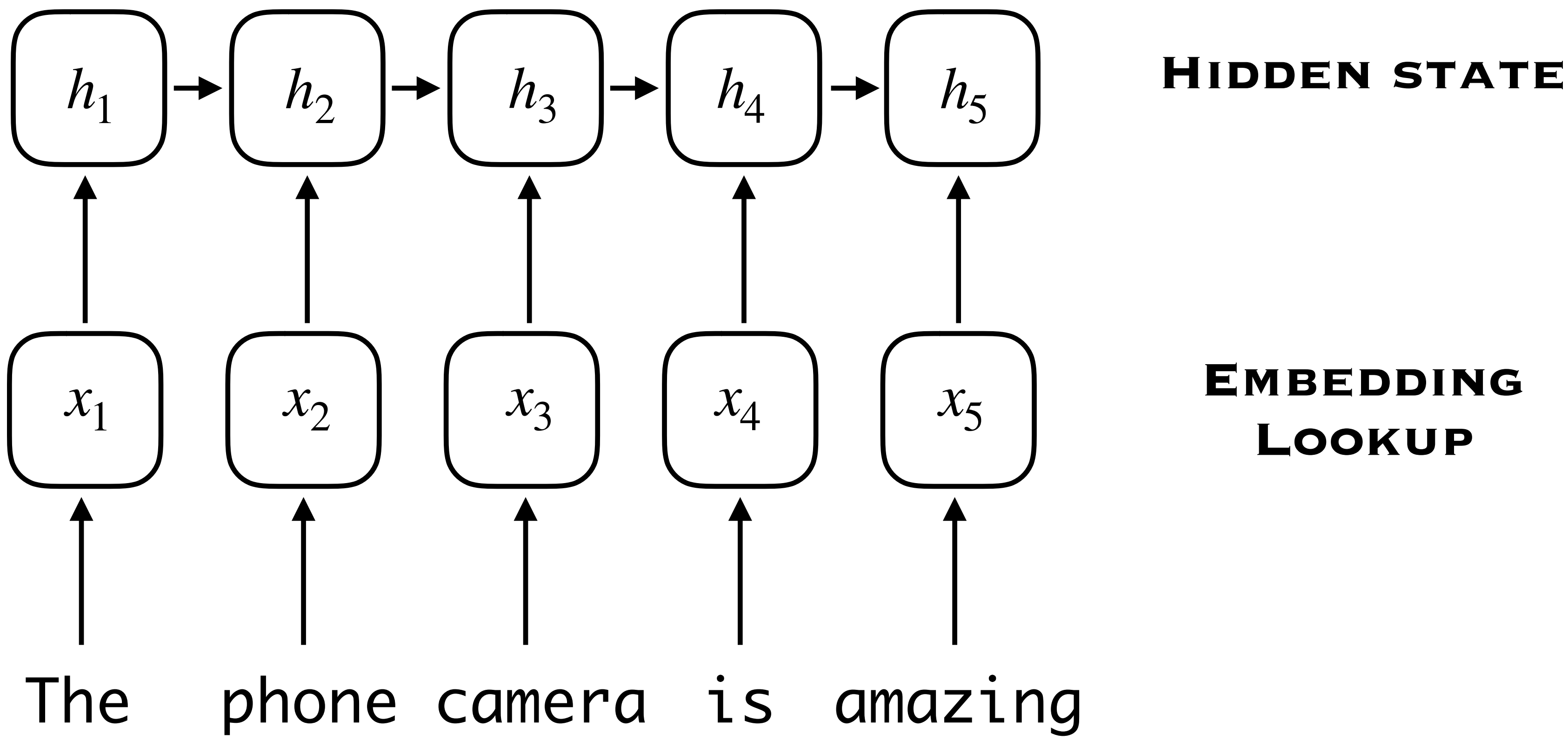
* alt.qcri.org/semeval2014/task4

(Quick) Background Refresher

Important Models

- Recurrent Neural Networks
- Memory Networks
- Transformers (BERT)

Recurrent Neural Nets



Capture Sequentiality

$$h_t = f(W_{hh}h_{t-1} + W_{hx}x_t + b)$$

Popular Variants

- LSTM
- GRU

Representation

- Last state h_n
- "Attention" weighted

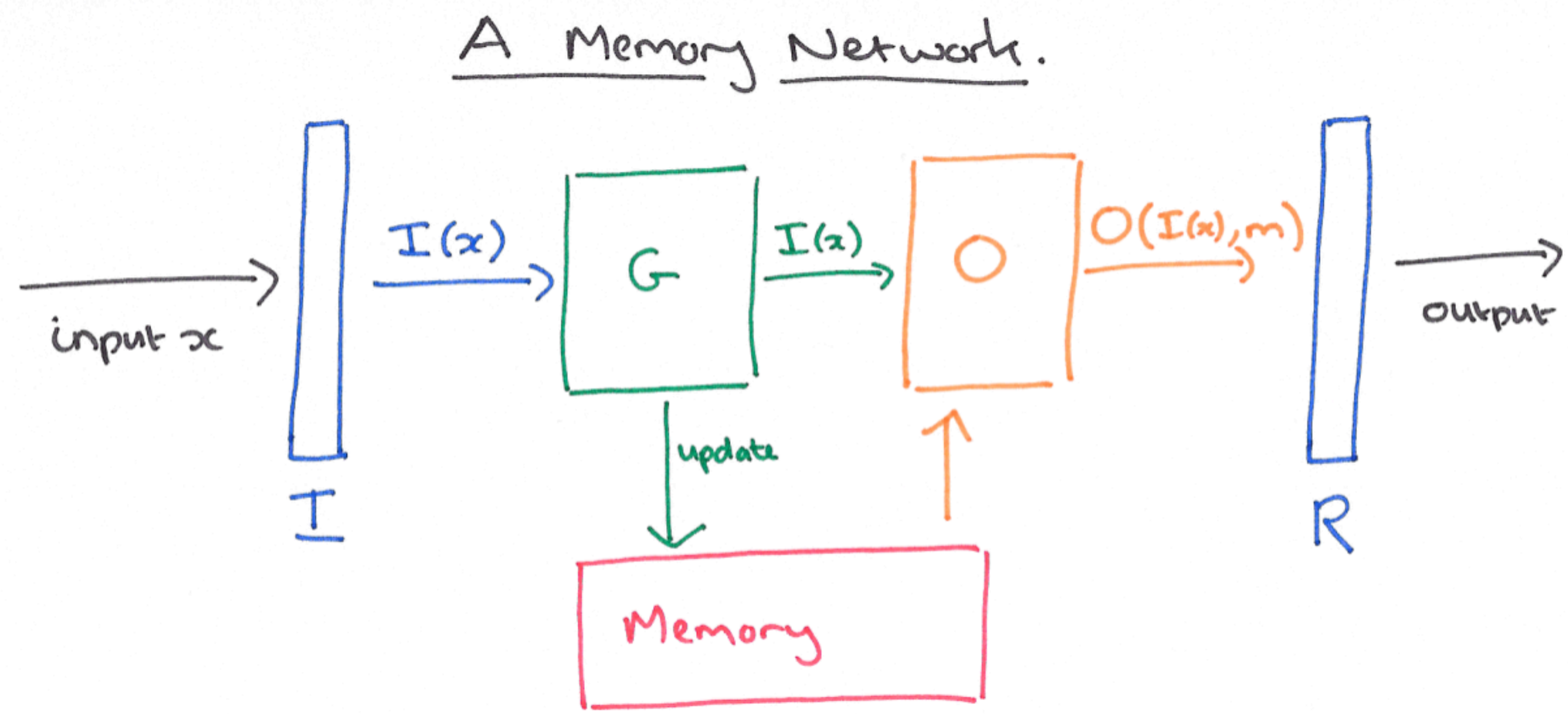
$$c_t = \sum_{i=1}^{n} \alpha_i h_i$$

- Appropriate concat (Bi-LSTM)

Memory Networks

CORE IDEA: Memory which can be

- Read from
- Written into
- Jointly Learned



*Components of a memory network

Components

- Input (**I**)
- Generalization (**G**)
- Output (**O**)
- Response (**R**)

Memory Networks,
Weston et al., ICLR 2015

End-to-End Memory Networks,
Sukhbaatar et al., NIPS 2015

* borrowed from Adrian Colyer's blog
blog.acolyer.org/2016/03/10/memory-networks/

BERT..

Using pre-trained language representations

- Feature-based
- Fine-tuning

BERT - Bidirectional **E**ncoder **R**epresentations from **T**ransformers

Pre-training

- Task #1: Masked Language Model (**MLM**)
- Task #2: Next Sentence Prediction (**NSP**)

Fine-Tuning

- Plug in task specific inputs and fine tune parameters end-to-end

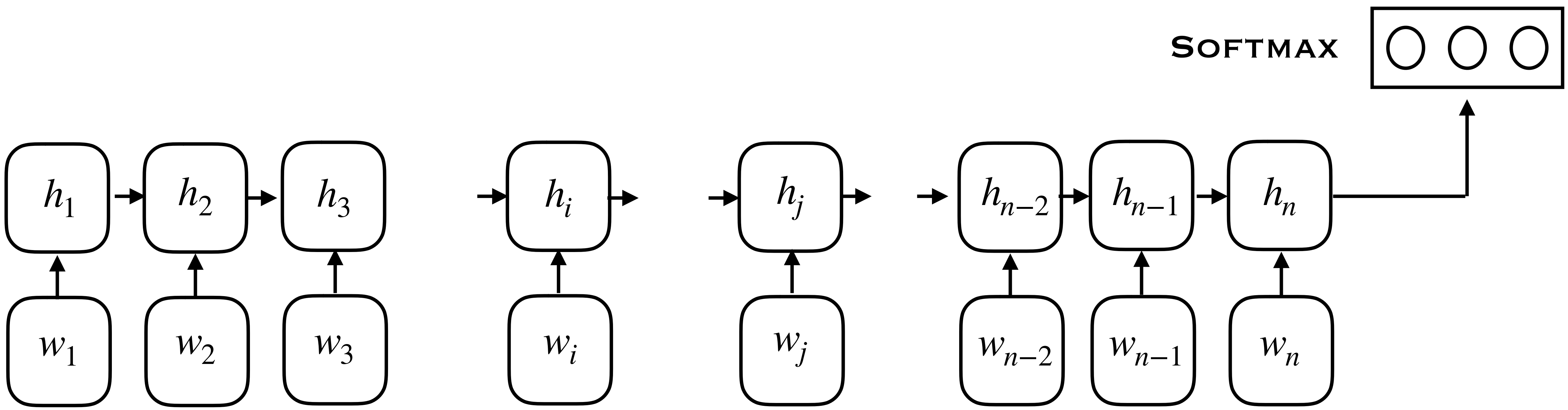
Resulting embeddings are contextual and can be adapted to new downstream tasks

BERT: Pre-training of deep bidirectional transformers for language understanding,
Devlin et Al., NAACL 2019

Approaches to model ABSA

LSTM Based Models

LSTM model



$w_i, \ldots, w_j$

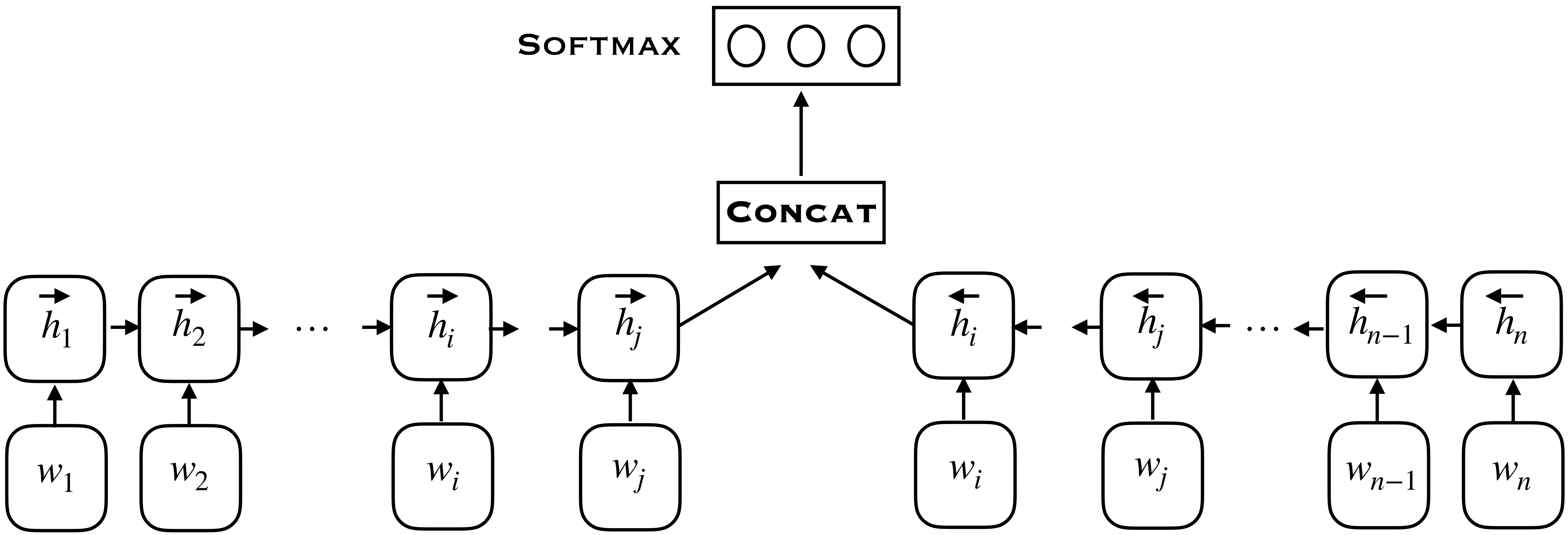
Aspect Phrase

Review Sentence

$w_1, w_2, \ldots, w_{i-1}, w_i, \ldots, w_j, w_{j+1}, \ldots, w_{n-1}, w_n$

Effective LSTMs for Target-Dependent Sentiment Classification,
Tang et al., COLING 2016

Target Dependent TD-LSTM model



$w_i, \ldots, w_j$ $w_i, \ldots, w_j$

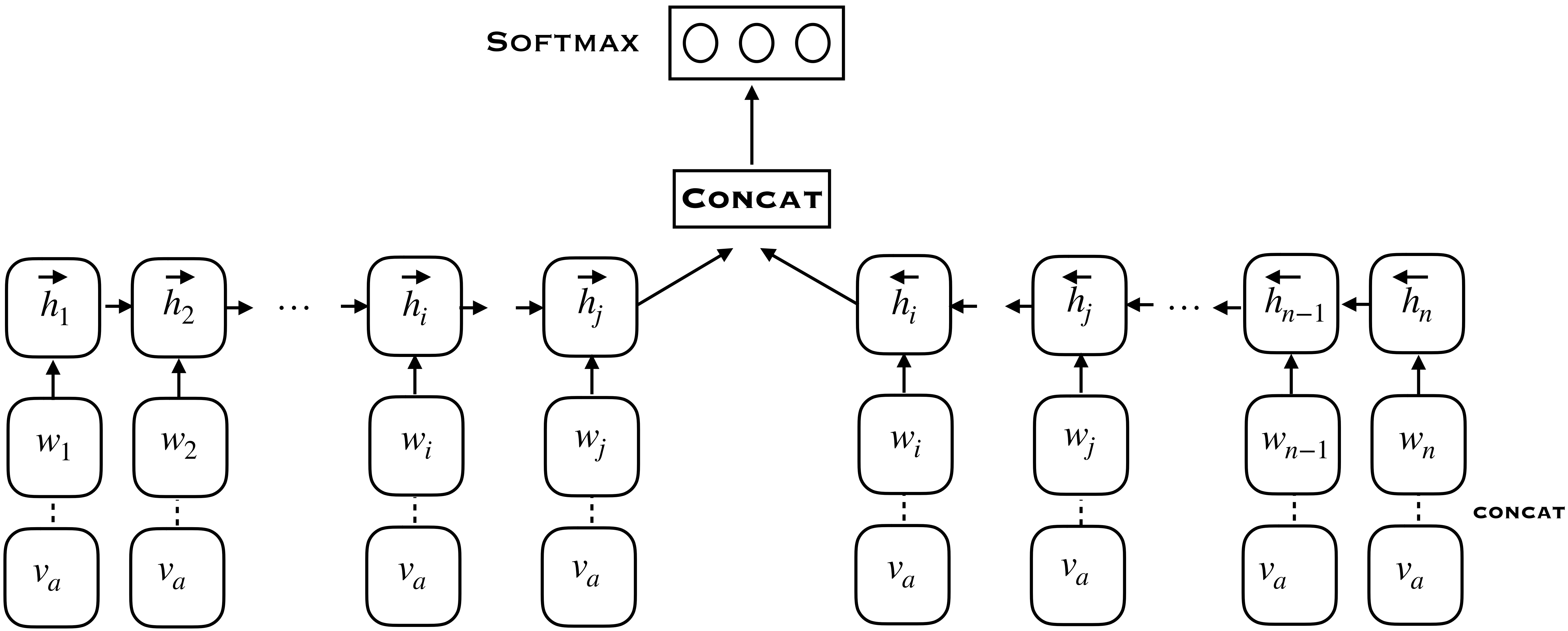
Aspect Phrase **Aspect Phrase**

Review Sentence

$w_1, w_2, \ldots, w_{i-1}, w_i, \ldots, w_j, w_{j+1}, \ldots, w_{n-1}, w_n$

Effective LSTMs for Target-Dependent Sentiment Classification,
Tang et Al., COLING 2016

Target Connection TC-LSTM model



Review Sentence

$w_1, w_2, \ldots, w_{i-1}, w_i, \ldots, w_j, w_{j+1}, \ldots, w_{n-1}, w_n$

v_a **Aspect Representation***

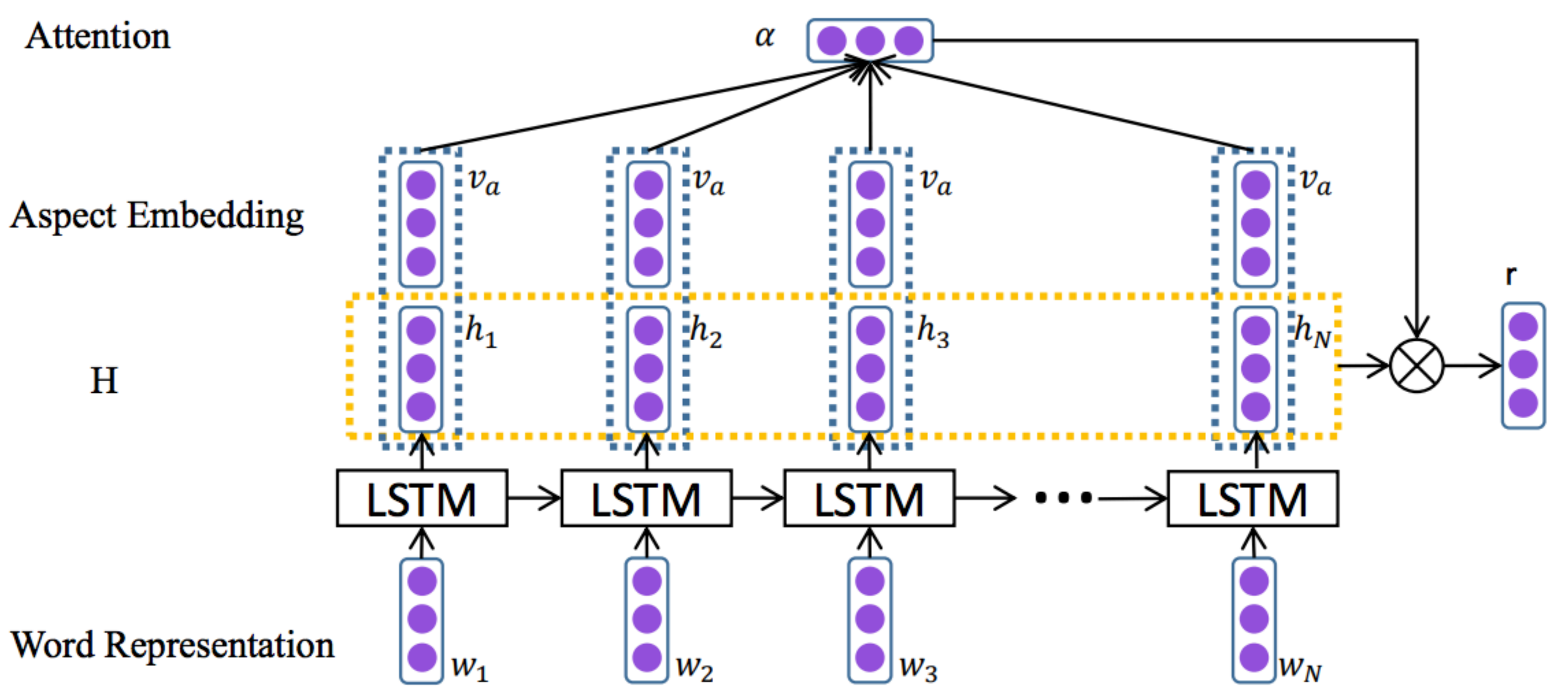
***Averaged for phrases**

Effective LSTMs for Target-Dependent Sentiment Classification,
Tang et Al., COLING 2016

Input sentence - Sequence of tokens

	LSTM	TD-LSTM	TC-LSTM
Model Aspect (Target)	✗	✓	✓
Context around aspect	✗	✓	✓
Aspect info. at each token	✗	✗	✓

AT-LSTM



Attention Based LSTM

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

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

$$r = H\alpha^T$$

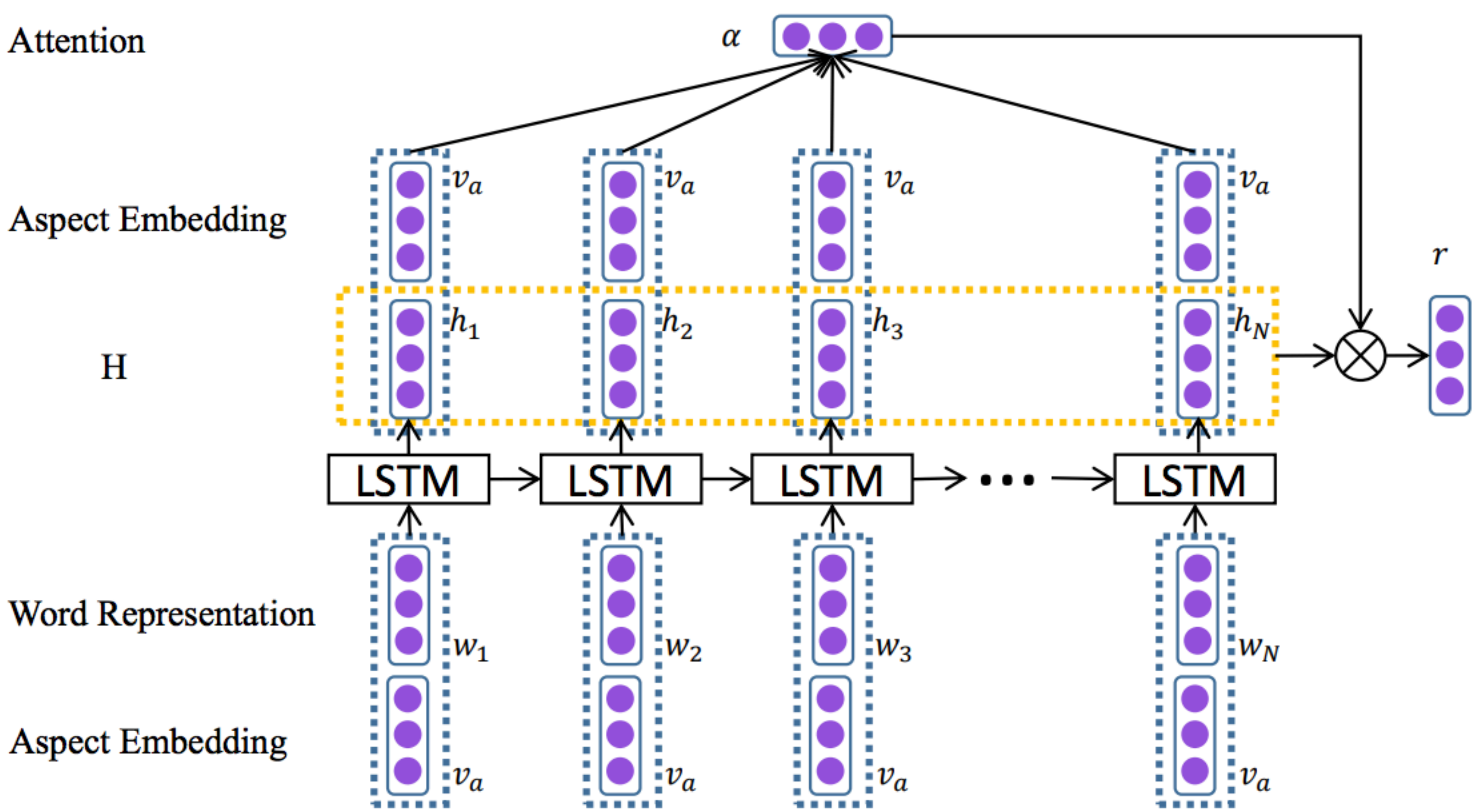
Attention Computation

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

Final representation

Attention Based LSTM for Aspect-Level Sentiment Classification,
Wang et al., EMNLP 2016

ATAE-LSTM



Attention Based LSTM with Aspect Embedding

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

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

$$r = H\alpha^T$$

Attention Computation

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

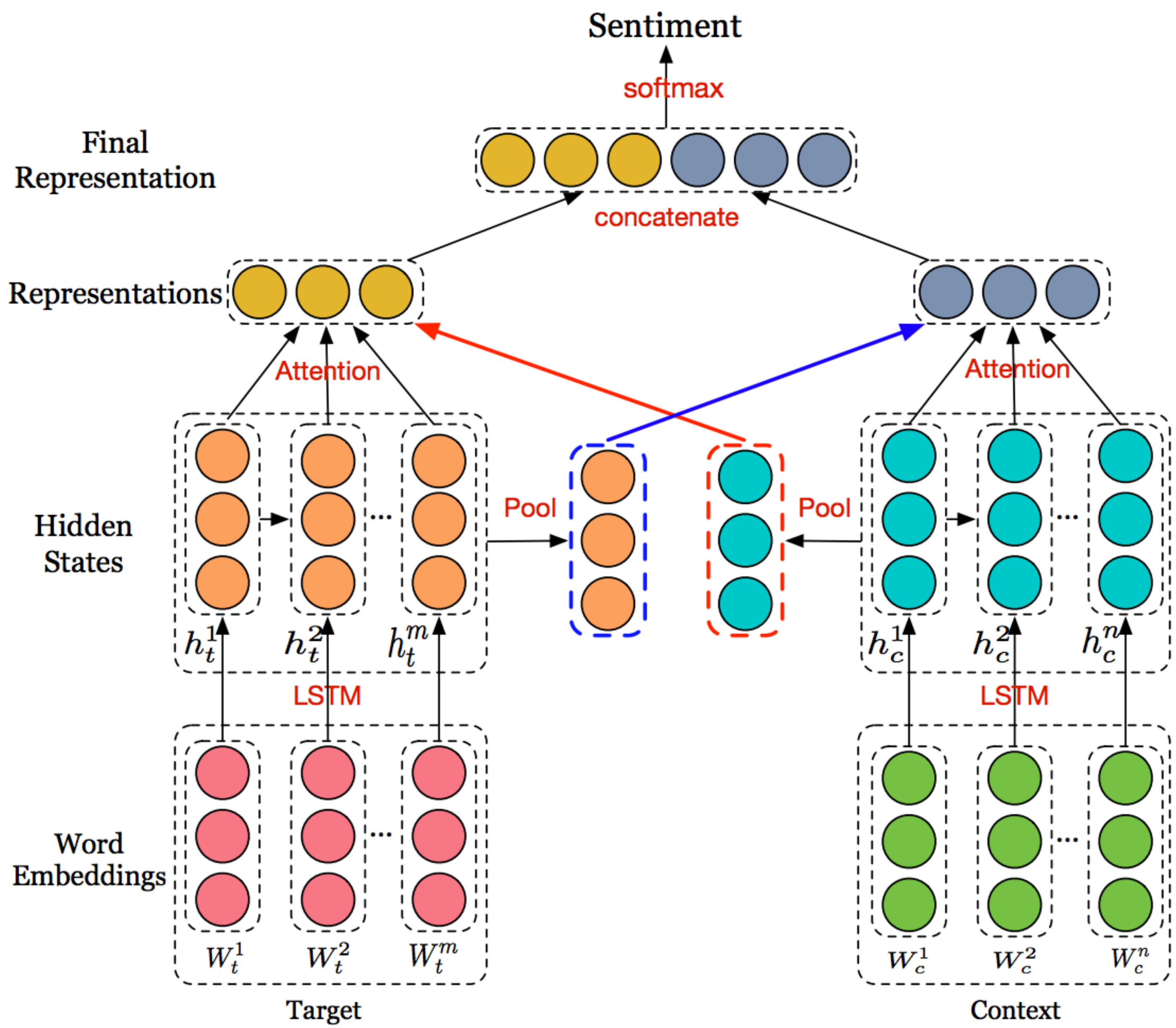
Final representation

Attention Based LSTM for Aspect-Level Sentiment Classification,
Wang et al., EMNLP 2016

Idea: Capture important information in response to a given aspect

Use appropriate attention weighting scheme

	AT-LSTM	ATAE-LSTM
Aspect Representation used to derive LSTM states	✗	✓
Aspect representation used to derive attention weights	✓	✓



Interactive Attention Networks

$$\alpha_i = \frac{exp(\gamma(h_c^i, t_{avg}))}{\sum_{j=1}^{n} exp(\gamma(h_c^j, t_{avg}))}$$

Attention score computation

$$\gamma(h_c^i, t_{avg}) = tanh(h_c^i \cdot W_a \cdot t_{avg}{}^T + b_a)$$

Similarity score

$$c_r = \sum_{i=1}^{n} \alpha_i h_c^i$$

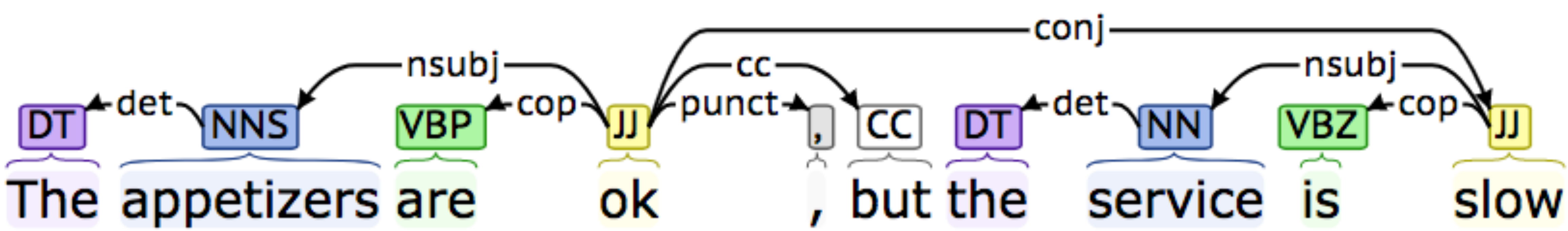
$$t_r = \sum_{i=1}^{m} \beta_i h_t^i$$

Context and target representation

Interactive Attention Networks for Aspect-Level Sentiment Classification, Ma et al., IJCAI 2017

IDEA: Capture both targets and contexts and model interaction them

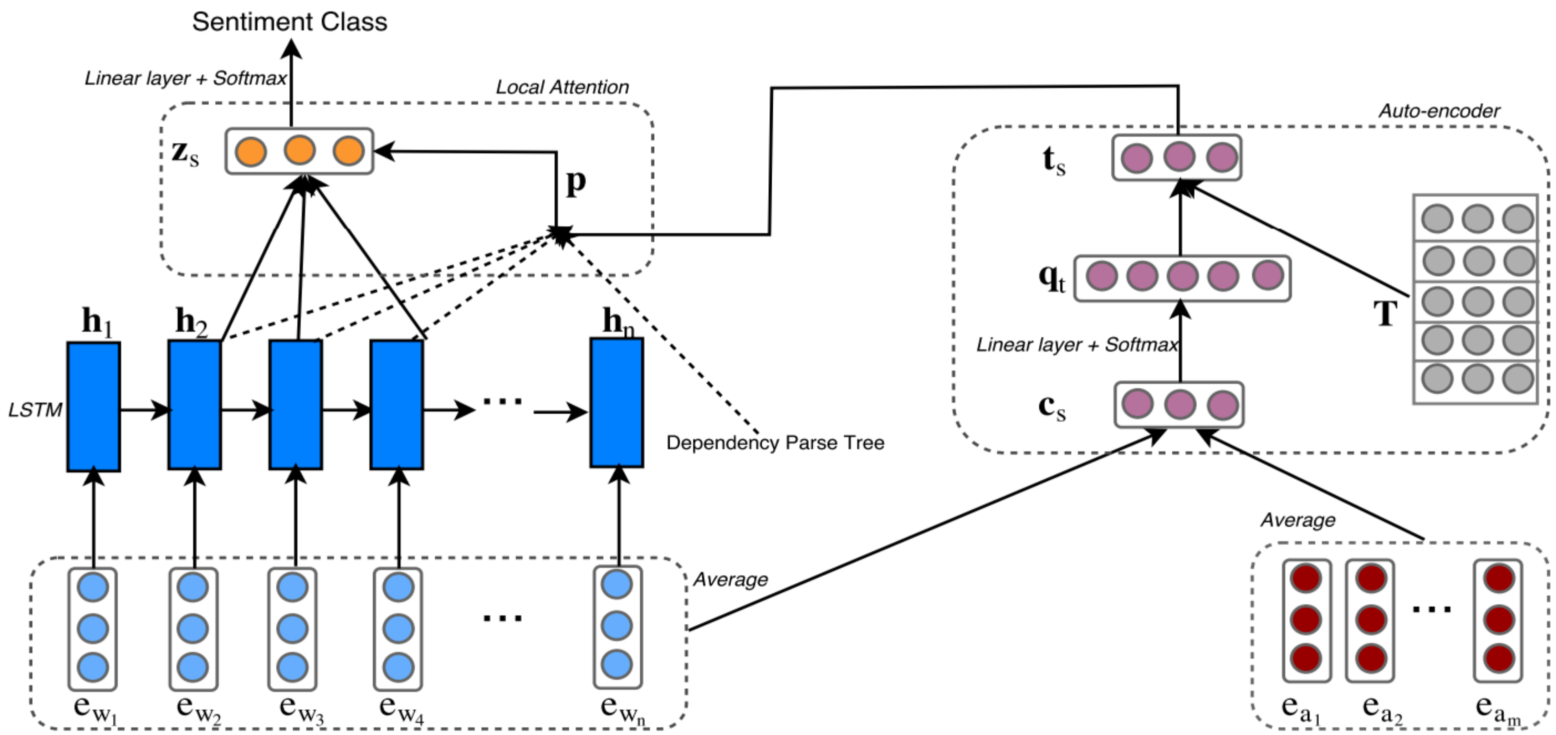
Target	Represent target sequence using an LSTM
Context	Represent context sequence using an LSTM
Interaction	Use tanh non-linearity to capture similarity between token (context) representation and average context (token) representation



Dependency parse visualisation of a sample review sentence

$$\mathbf{t}_s = \mathbf{T}^\top \cdot \mathbf{q}_t$$
$$\mathbf{q}_t = \mathit{softmax}(\mathbf{W}_t \cdot \mathbf{c}_s + \mathbf{b}_t)$$
$$\mathbf{c}_s = \mathit{Average}\left(\frac{1}{m}\sum_{i=1}^{m} \mathbf{e}_{a_i}, \frac{1}{n}\sum_{j=1}^{n} \mathbf{e}_{w_j}\right)$$

Attention Equations



$$f_{score}(\mathbf{h}_i, \mathbf{t}_s) = tanh(\mathbf{h}_i^T \cdot \mathbf{W}_a \cdot \mathbf{t}_s)$$

Scoring function

$$p_i = \frac{d_i}{\sum_j d_j}$$
$$d_i = \begin{cases} \frac{1}{2^{(l_i-1)}} \cdot exp(f_{score}(\mathbf{h}_i, \mathbf{t}_s))), & \text{if } l_i \in [1, ws] \\ 0, & \text{otherwise} \end{cases}$$

Incorporating syntactic information

$$\mathbf{z}_s = \sum_{i=1}^{n} p_i \mathbf{h}_i$$

Input representation

Effective Attention Modelling

Effective Attention Modelling for Aspect-Level Sentiment Classification,
He et al., CoLING 2018

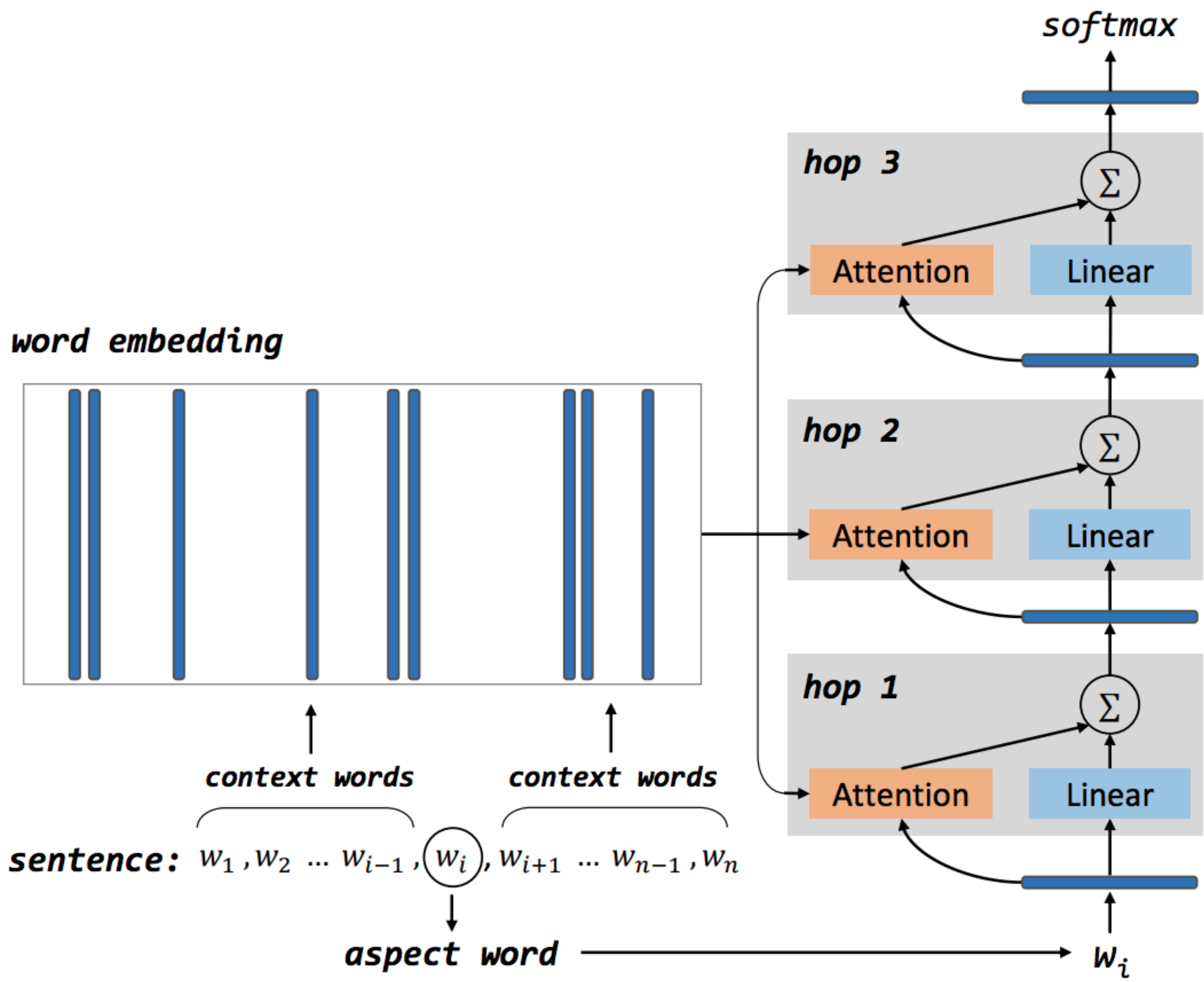
IDEA: Target represented as a weighted summation of meta-aspect embeddings

meta-aspect - can be thought of as a grouping of actual aspects mentioned

Syntactic information is important to determine target polarity

Words that are near the target, or have a modifier relation to target should get higher attention weight

Memory Network Based Models



Deep Memory Network

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

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

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

Content Attention

$$m_i = e_i \odot v_i$$

$$v_i = 1 - l_i / n$$

Location Attention

Aspect-Level Sentiment Classification with Deep Memory Network,
Tang et al., EMNLP 2016

IDEA: Iteratively refine memory using context and location weighted memory cells

Content Attention

Compute importance of each word in the input with respect to the aspect

Location Attention

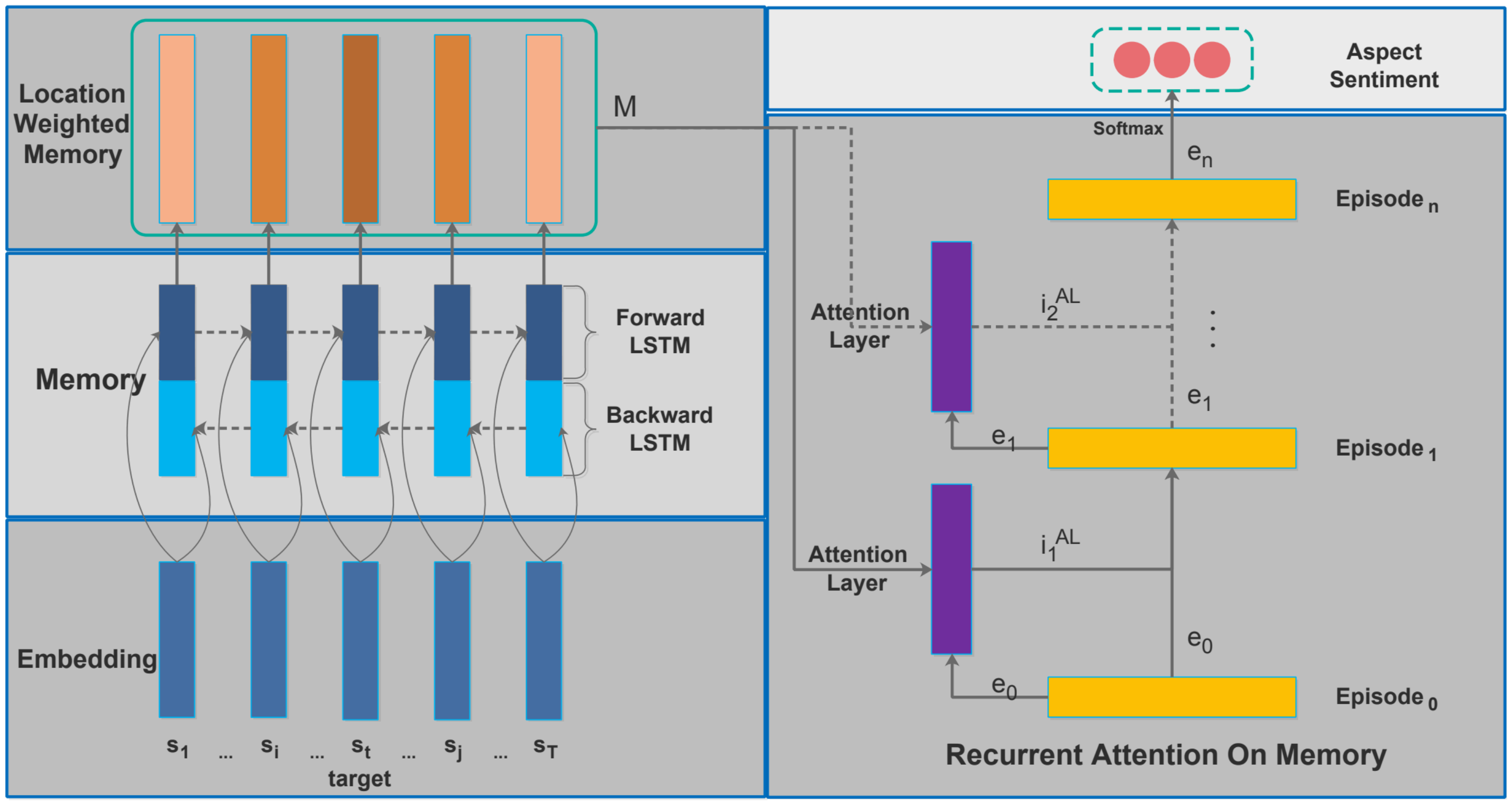
Model the importance of the words in the input based on its "distance" from the aspect

Multiple hops

Derive progressively abstract representations of input

$$w_t = 1 - \frac{|t - \tau|}{t_{max}}$$

Position Weighting



Recurrent Attention on Memory

$$g_j^t = W_t^{AL}(m_j, e_{t-1}[, v_\tau]) + b_t^{AL}$$

Attention Computation

$$\alpha_j^t = \frac{\exp(g_j^t)}{\sum_k \exp(g_k^t)}$$

Attention Weights

$$i_t^{AL} = \sum_{j=1}^{T} \alpha_j^t m_j$$

Context Vector

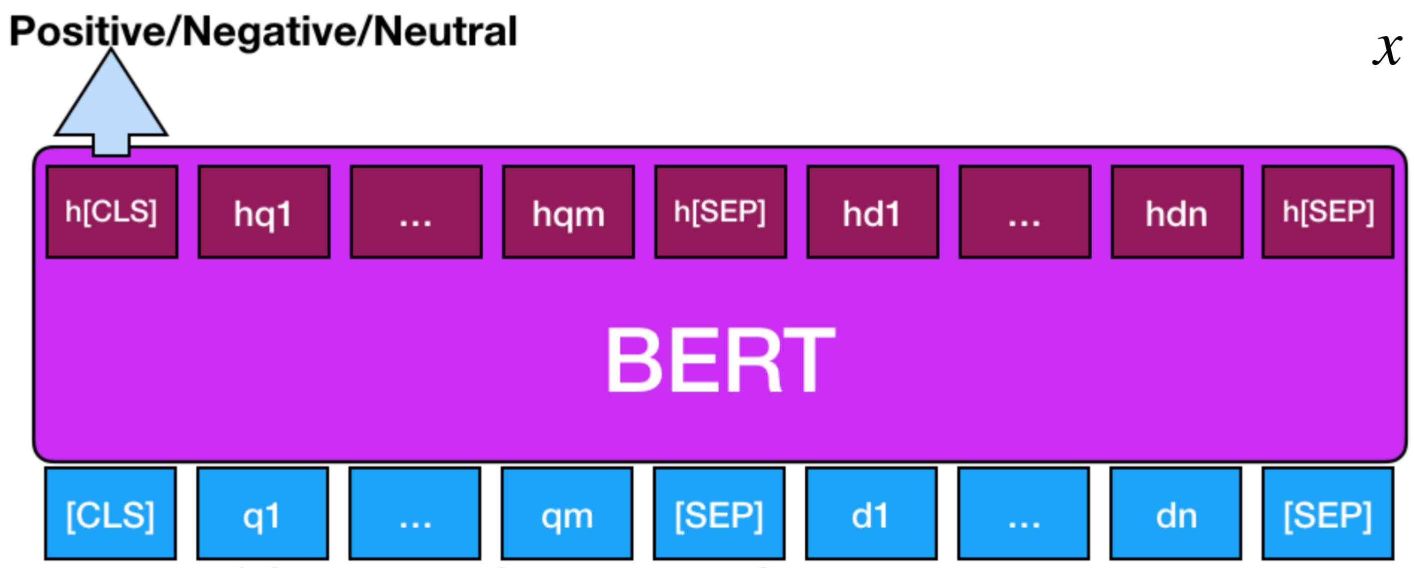
Recurrent Attention on memory for Aspect Sentiment Classification, CHEN et al., EMNLP 2017

IDEA: Refine location-weighted memory iteratively to derive input representation

Memory: BiLSTM states weighted depending on the distance from target word

Use a GRU to (recurrently) refine attention weighted memory

Transformer Based Models



$$x = ([CLS], q_1, \ldots, q_m, [SEP], d_1, \ldots, d_n, [SEP])$$

$$h = BERT(x)$$

$$l = softmax(W \,.\, h_{[CLS]} + b)$$

Aspect Phrase $q_1, \ldots, q_m$

Review Sentence $d_1, \ldots, d_n$

BERT Post-training

BERT Post-training for review reading comprehension and Aspect Based Sentiment Analysis,
XU et al., NAACL 2019

IDEA: Fine tuned contextual representations help with sentiment classification

Fine tuning: Domain Knowledge (review data) and Review Reading Comprehension (RRC*)

- Tune BERT model to learn better representation of words appearing in reviews
- Aspect Sentiment Classification is very similar to RRC predicting sentiment equivalent to answering the question about the polarity of the aspect

* RRC - Task and dataset introduced in the paper

Leaderboard

Model	Laptop (Acc)	Restaurant (Acc)
TD-LSTM	68.13	75.63
ATAE-LSTM	68.7	77.2
IAN	72.1	78.6
LSTM+SynATT+TarRep	71.94	80.63
MemNet	72.21	80.95
RAM	74.49	80.23
BERT-PT	78.07	84.95

References

- Memory Networks, Weston et Al., ICLR 2015
- End-to-end Memory Networks, Sukhbaatar et Al., NIPS 2015
- BERT: Pre-training of deep bidirectional transformers for language understanding, Devlin et Al., NAACL 2019
- Effective LSTMs for Target-Dependent Sentiment Classification, Tang et Al., COLING 2016
- Attention Based LSTM for Aspect-Level Sentiment Classification, Wang et al., EMNLP 2016
- Interactive Attention Networks for Aspect-Level Sentiment Classification, Ma et al., IJCAI 2017
- Effective Attention Modelling for Aspect-Level Sentiment Classification, He et al., CoLING 2018
- Aspect-Level Sentiment Classification with Deep Memory Network, Tang et al., EMNLP 2016
- Recurrent Attention on memory for Aspect Sentiment Classification, CHEN et al., EMNLP 2017
- BERT Post-training for review reading comprehension and Aspect Based Sentiment Analysis, XU et al., NAACL 2019
- PapersWITHCode: https://paperswithcode.com/sota/aspect-based-sentiment-analysis-on-semeval