# Deep Mask Memory Network with Semantic Dependency and Context Moment for Aspect Level Sentiment Classification

Authors: Peiqin Lin, Meng Yang, Jianhuang Lai

Presenter: Peiqin Lin

# Content

# Task Description

Aspect Level Sentiment Classification:
Identifying the sentiment polarity (i.e., negative, neutral, or positive) of each given aspect term in a sentence

Example1: “The price is very cheap, but the service is very poor”
Result: price (positive), service (negative)
Example2: “From the speed to the multi touch gestures this operating system beats Windows easily.”
Result: speed (positive), multi touch gestures (positive), operating system (positive), Windows(negative)

# Background

Existed work

- Approaches with memory network have achieved promising results (Chen et al., 2017)
- Inter-aspect information have been considered to solve the problem (Majumder et al., 2018; Hazarika et al., 2018)

Crucial issues:

- Semantic information between context and aspect
- The relation between aspects in the same sentence are not well exploited.

# Contribution of DMMN-SDCM

1. Integrating semantic parsing information instead of location information into deep memory network to guide the attention mechanism;
2. Integrating semantic parsing information into inter-aspect modeling for utilizing the information provided by the nearby aspects in a better way;
3. Designing an auxiliary task to learn the sentiment distribution of the entire sentence, which can provide a background for the sentiment analysis of the desired aspect;

# Overview

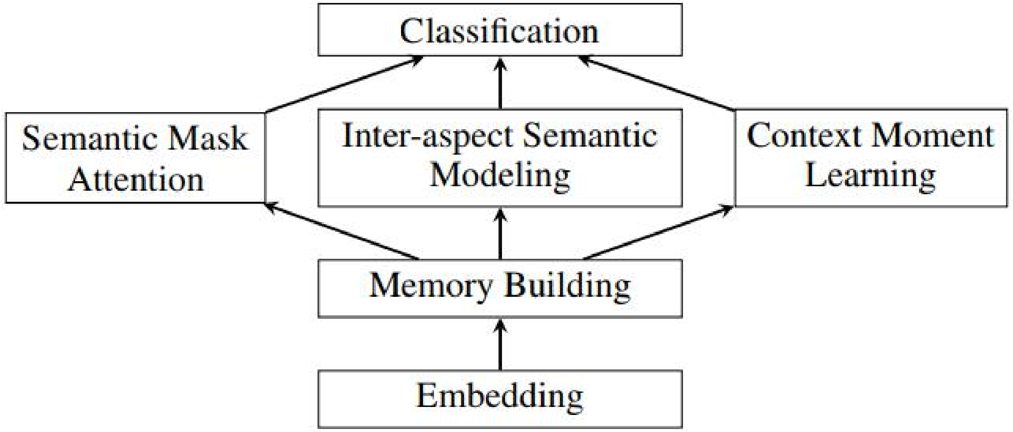


Figure 1: The overall architecture of DMMN-SDCM.

Six modules:

- embedding
- memory building
- semantic mask attention inter-
- aspect semantic modeling
- context-moment learning
- classification

# Semantic mask attention and aspect semantic modeling

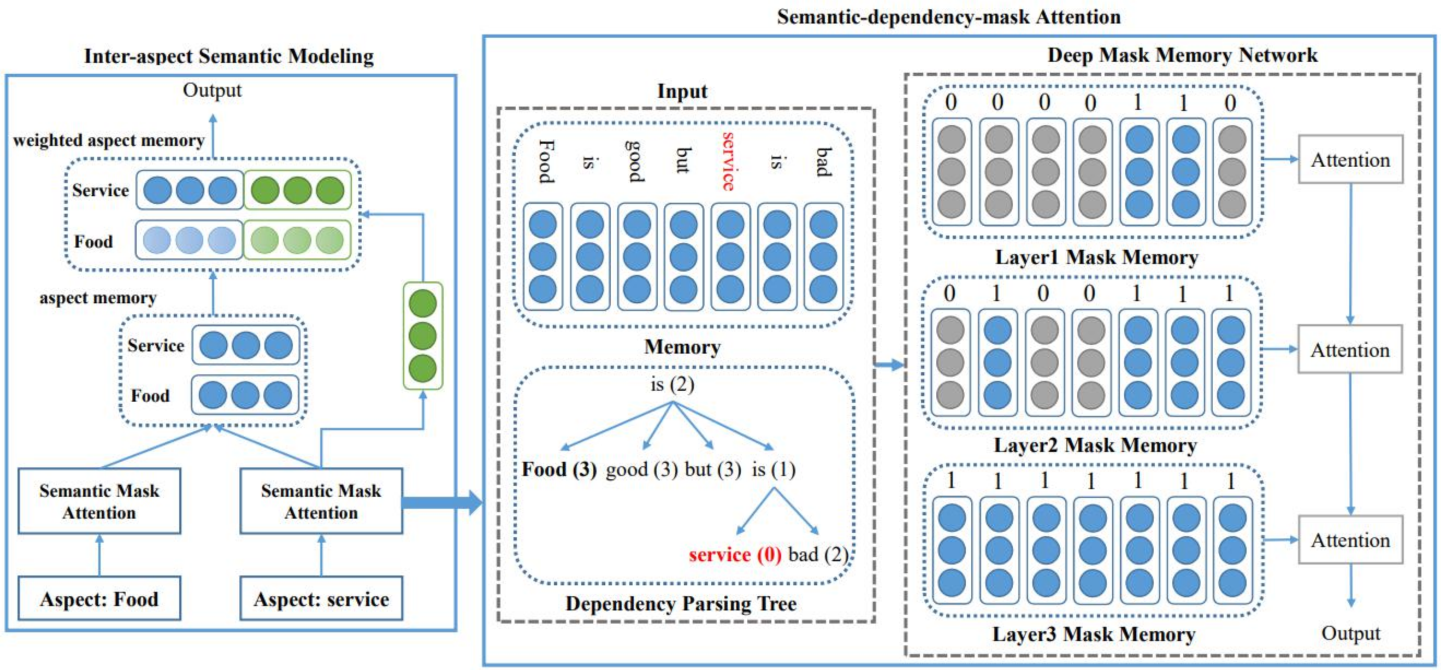


Figure 2: Architecture of inter-aspect semantic modeling and semantic-dependency-mask attention. The left part of the semantic mask attention module illustrates the input of the deep mask memory network, namely the memory $\{\boldsymbol{m}_1^*, \ldots, \boldsymbol{m}_i^*, \ldots, \boldsymbol{m}_n^*\}$ and the dependency parsing tree whose nodes are in the form of $word(dist(w_i, w_t))$, where $dist(w_i, w_t)$ is the distance from the aspect $w_t$ to the context word $w_i$ in the dependency tree. The right part shows how the deep mask memory network works. Attention mechanism is only applied on the selected memory slices remaining colored in the figure.

# Context moment learning

Context moment:

Let –1, 0 and 1 denote negative, neutral and positive, and the moment is defined as follows:

$$\mu_i = E((x-\mu)^i)$$

Specifically:

- $\mu_1$ is mean, which portrays the overall sentiment of the sentence
- $\mu_2$ is variance, which portrays the relationship between aspects of the sentence

# Context moment learning

| Sentences | Distribution | Mean | Var |
|---|---|---|---|
| **Food** and **service** are good. | [1, 1] | 1 | 0 |
| **Food** and **service** are bad. | [-1, -1] | 0 | 0 |
| **Food** are good, **service** are bad. | [1, -1] | 0.5 | 1 |

Table 1: Context moment calculation examples with two aspects. The normalized value of mean and variance, namely the first and second moment are given in the table.

- The greater the mean, the more positive the overall sentiment of the sentence.
- The closer the variance is to 1, the closer the relation between aspects is to comparison, while the closer the variance is to 0, the closer the relation is to coordination.

Conclusion: context moment can describe the overall sentiment of the sentence and the relation between aspects.

# Context moment learning

An auxiliary task which learns the sentiment distribution:

- Inputs: original memory of sentence
- Network: an attention layer and two FC layers
- Outputs: the estimates of context moments

The input of last layer will be taken as the feature output of the module, and used in the classification module.

# Experimental settings

- Word vector: glove.840B.300d.vec, oov words are initialized by U(–0.25, 0.25)
- Optimizer: Adam with learning rate 0.01 and weight decay 0.0001
- Dimension of model layers: 50
- Evaluation metrics are Accuracy and Macro-F1.

| Dataset | Positive | | Negative | | Neutral | |
|---|---|---|---|---|---|---|
| | train | test | train | test | train | test |
| Laptop | 994 | 341 | 870 | 128 | 464 | 169 |
| Restaurant | 2164 | 728 | 807 | 196 | 637 | 196 |

Table 2: Details of the experimental datasets.

# Comparisons

Baseline methods: TD-LSTM [Tang et al., 2015], ATAE-LSTM [Wang et al., 2016], MemNet [Tang et al., 2016], IAN [Ma et al., 2017], BILSTM-ATT-G [Liu and Zhang, 2017], RAM [Chen et al., 2017], Conv-Memnet [Fan et al., 2018] and TNet [Li et al., 2018]

| Models | Laptop | | Restaurant | |
|---|---|---|---|---|
| | ACC | Macro-F1 | ACC | Macro-F1 |
| TD-LSTM | 71.83 | 68.43 | 78.00 | 66.73 |
| ATAE-LSTM | 68.65 | 62.45 | 77.23 | 64.95 |
| MemNet | 70.33 | 64.09 | 78.16 | 65.83 |
| IAN | 72.10 | 67.48 | 77.95 | 67.90 |
| BILSTM-ATT-G | 74.37 | 69.90 | 80.38 | 70.78 |
| RAM | 75.01 | 70.51 | 79.79 | 68.86 |
| Conv-Memnet | 76.37 | 72.10 | 78.26 | 68.38 |
| TNet | 76.54 | 71.75 | 80.79 | 71.27 |
| DMMN-SDCM | **77.59** | **73.61** | **81.16** | **71.50** |

# Extensive experiments

| Contained Modules | ACC | Macro-F1 |
|---|---|---|
| $M_{sd}$ | 75.08 | 71.68 |
| $M_{sd} + M_{im}$ | 76.80 | 72.82 |
| $M_{sd} + M_{cm}$ | 76.18 | 71.91 |
| $M_{sd} + M_{im} + M_{cm}$ | **77.59** | **73.61** |

Table 4: Effects of each module on the Laptop dataset. $M_{sd}$, $M_{im}$ and $M_{cm}$ represents the mask attention module, the inter-aspect modeling module and the context-moment module, respectively.

Conclusion: with the three creative modules, our model performs best.

# Extensive experiments

| Models | ACC | Macro-F1 |
|---|---|---|
| Without Inter-aspect Modeling | 76.18 | 71.91 |
| BiLSTM | 75.08 | 70.40 |
| BiLSTM + Semantic Weighting | 75.24 | 70.83 |
| Attention | 75.55 | 71.73 |
| Attention + Semantic Weighting | **77.59** | **73.61** |

Table 5: Effects of inter-aspect modeling on the Laptop dataset.

Conclusion: introducing the semantic information can improve the performance.

# Conclusions

In this paper, we design a deep mask memory network with semantic dependency and context moment (DMMN-SDCM) which integrates semantic parsing information and context moment learning into deep memory network for the first time. Experiments show that our model performance is state-of-the-art.

# THANKS

Deep Mask Memory Network with Semantic Dependency and Context Moment for Aspect Level Sentiment Classification


- Peiqin Lin, Meng Yang, Jianhuang Lai