Learning Sentiment Memories for Sentiment Modification without Parallel Data

Yi Zhang, Jingjing Xu, Pengcheng Yang, Xu Sun
MOE Key Laboratory of Computational Linguistics, Peking University
School of Electronics Engineering and Computer Science, Peking University

INTRODUCTION

Task: sentiment modification

- Reversing the sentiment of the input to the opposite
- Preserving the sentiment-independent content

Challenge:

- Aligned sentences with the same content but different sentiments are usually unavailable.

MOTIVATION

Staff → *friendly* / *rude*
Food → *delicious* / *terrible*

We propose a novel method that automatically extracts appropriate sentiment information from learned sentiment memories according to specific context.

MODEL

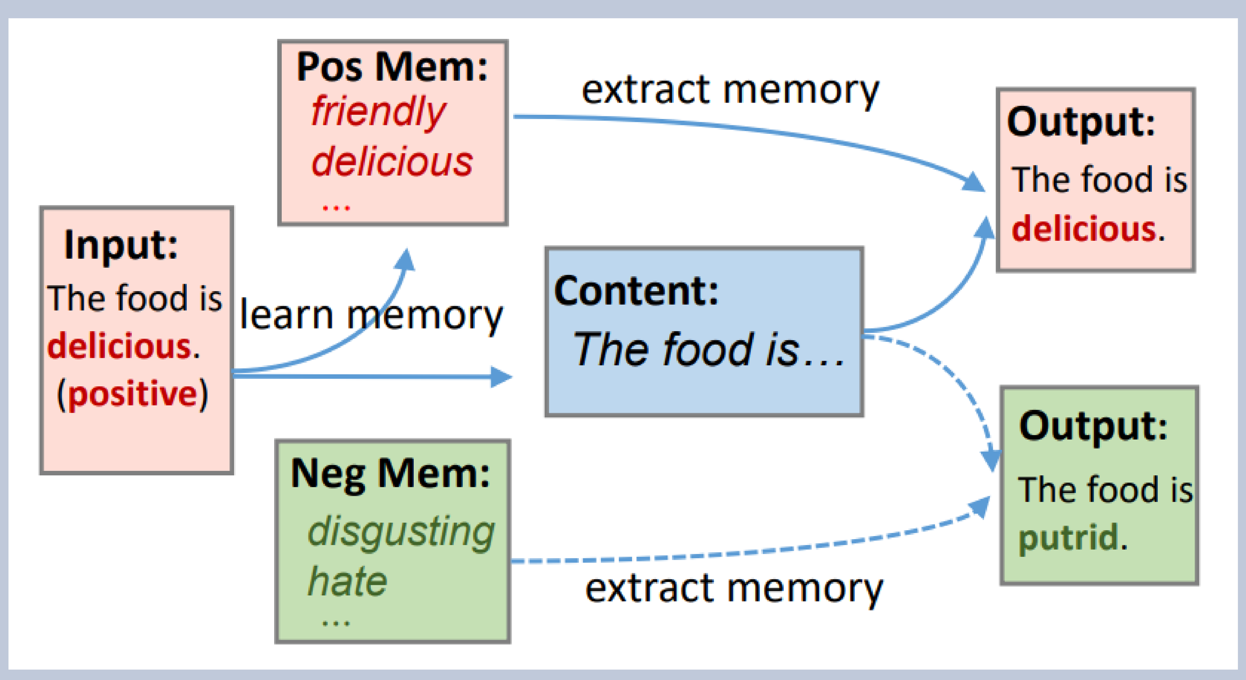

Figure1: Illustration of the proposed model with a positive input. Solid and dashed lines indicate the training process and the testing process, respectively. The process with a negative input is in a similar way.

- **Emotional Words Detection Model**

We first find the emotional words that have the most discriminative power for sentiment polarity. This work is done by training a sentiment classifier with a simple self-attention mechanism.

It really is n't worth coming here .
Been here once and loved going here.

- **Sentiment-Memory Based Auto-Encoder**

Update memory

We first sum the embeddings of the emotional words to get a vector representation of sentiment. Then we add the weighted vector to the columns of the sentiment matrix. The weights are computed by an attention mechanism and reflect the relevance of sentiment vector and the columns of matrix.

Extract memory

We first sum the embedding of the context words to get a vector representation of context. Then we compute the attention weights over the columns of sentiment memory matrix. Finally we sum these weighted columns as the extracted memory.

EVALUATION

Input: Very helpful and informative staff!
CAE: Worst service ever
MAE: Very nice here and poor!
Proposed: Very rude and careless staff !
Input: I will never go here again.
CAE: I love this place here!
MAE: I had say this place here.
Proposed: I will never go anywhere else.
Input: The worst and would never recommend anyone to use them.
CAE: The best place I 've been to go here!
MAE: The first experience is so happy and nice.
Proposed: The best and would definitely recommend anyone to use them.

Table 1: Examples generated by the proposed method and baselines

Model	Sentiment	Content	Fluency
CAE	6.55	4.46	5.98
MAE	6.64	4.43	5.36
SMAE	6.57	5.98	6.69

Table 2: Results of human evaluation

Model	ACC	BLEU
CAE	71.96	2.77
MAE	74.59	5.45
SMAE	76.64	24.00

Table 3: Results of automatic evaluation

The staff here is very rude.
It really is n't worth coming here .
Very pleased with this business.
Been here once and loved going here.

Table 4: The effectiveness of the memory module with examples. The brown words are absent in the input but generated with the help of sentiment memories.

Models	Acc	Bleu
SMAE	76.64	24.00
SMAE(w/o memories)	14.08	26.09

Table 5: Ablation test of memory module

CONCLCUSIONS

- We propose a model that learns sentiment memories without parallel data and then automatically extract sentiment information to adapt different contexts when decoding.
- Experimental results show that our method substantially improves content preservation.

MISC

- ARXIV: arxiv.org/pdf/1808.07311.pdf
- CODE: github.com/lancopku/SMAE