Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction

Xin Li   Wai Lam
Presenter: Piji Li

Department of Systems Engineering & Engineering Management
The Chinese University of Hong Kong

Sept. 2017

Aspect Term Extraction (ATE)

Goal: Detect the <u>aspect phrases</u> (also called "aspect terms") where the users express their opinions from the online reviews.

Potential aspect terms

- Mention of product / entity
- Mention of product attributes

Examples

- *Its size is ideal and the weight is acceptable.*
- *The pizza is overpriced and soggy.*

Aspect Term Extraction (ATE)

- The supporting task for Aspect-Based Sentiment Analysis (ABSA)
- Can be formulated as a sequence labeling problem or token-level classification problem.

Examples

Its	size	is	ideal	and	the	weight	is	acceptable
O	T	O	O	O	O	T	O	O

Aspect Term Extraction (ATE)

- A natural information extraction task.
- Some specific characteristics in the context of sentiment analysis can facilitate the ATE task.

Examples

- *Its size is ideal and the weight is acceptable.*
- *The pizza is overpriced and soggy.*

Aspect Terms and Opinion Words always co-occur.

Preliminary Works

1. Models just focusing on aspect term modeling.
   - **CRF, Semi-CRF**
   - **IHS_RD** (Chernyshevich et al., SemEval 2014, Winning system on laptop dataset)
   - **NLANGP** (Toh and Su, SemEval 2016, Winning system on restaurant dataset)
   - **LSTM** (Liu et al., EMNLP 2015)
   - **WDEmb** (Yin et al., IJCAI 2016)

Preliminary Works

2. Models capturing the aspect-opinion relation.
   - **Double Propagation** (Qiu et al., Computational Linguistics 2011)
   - **WTM** (Liu et al., EMNLP 2012)
   - **CR_WP** (Liu et al., ACL 2014)
   - **RNCRF** (Wang et al., EMNLP 2016)
   - **CMLA** (Wang et al., AAAI 2017)

Our Approach

Motivations:

1. Opinion words can provide <u>indicative clues</u> for finding aspects.
2. <u>Non-sentimental sentences</u> cannot have aspect terms and knowing this kind of features will help us to find aspect more accurately.

Ideas:

1. Modeling the aspect-opinion relation through the <u>memory interactions</u> between aspect term extraction and opinion word extraction.
2. Capturing <u>sentimental features</u> for the whole sentence and incorporating them into the aspect predictions.

Our Approach

We propose a LSTM-based multi-task learning framework named **M**emory **I**nteraction **N**etworks (**MIN**) for tackling aspect extraction.

Overview

- **MIN** involves three tasks: **aspect term extraction** (ATE), **opinion word extraction** and **sentimental sentence classification**.

- Two coupled LSTMs with extended memories are proposed to extract aspect terms and opinion words.

- Neural memory operations are defined over the extended memory to perform memory interactions.

- Sentimental representation of the whole sentence generated from another generic LSTM is incorporated into the aspect prediction.

Our Approach

Architecture: A-LSTM, O-LSTM, S-LSTM are responsible for aspect term extraction (ATE), opinion word detection and sentimental sentence classification respectively. (Note: outputs of O-LSTM and S-LSTM are ignored)
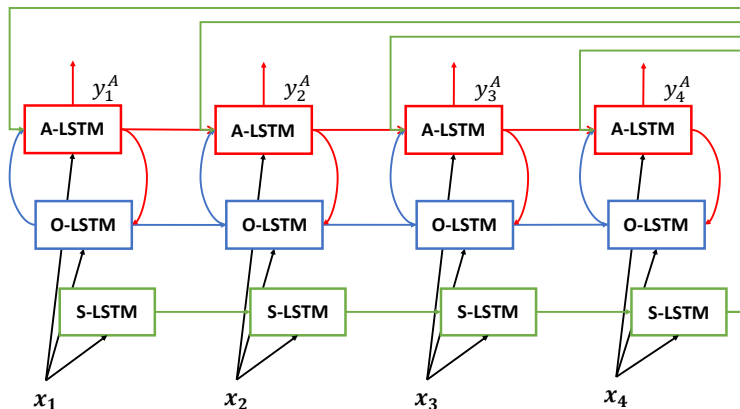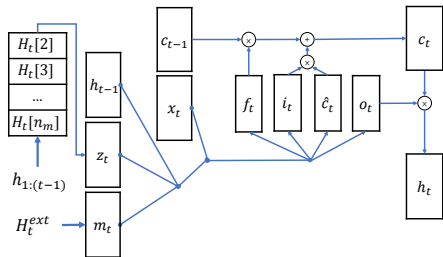
Our Approach

Our **MIN** consists of two components:

- Memory interactions over the task level memories (A-LSTM & O-LSTM).
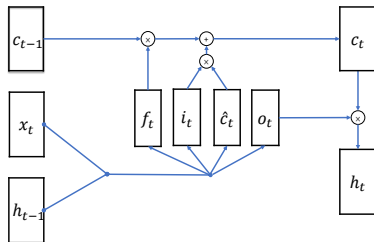- Sentimental feature learning for review sentences (S-LSTM).

1. We propose LSTMs with extended memories.

LSTM in our **MIN**

Vanilla LSTM (Jozefowicz et al., ICML 2015)

Memory Interaction

2. We design several kinds of memory operations

1. **READ**: select n_m pieces of aspect (opinion) hidden states from the past memories and build $H_t^A(H_t^O)$

2. **DIGEST**: distill an aspect (opinion)-specific summary $m_t^A(m_t^O)$

3. **INTERACT**: perform interactions between A-LSTM and O-LSTM.

Sentimental Feature Learning

We use another LSTM (namely, S-LSTM) to discriminate sentimental and non-sentimental sentences and use the learned features to facilitate the aspect prediction.

> **Remark**
>
> Sentimental sentences are the sentences containing gold standard aspects
>
> - *The pizza is overpriced and soggy.* ($\sqrt{}$)
>
> Non-sentimental sentences are the sentences where the users do not express any opinion.
>
> - *And I've been to many NYC delis.* ($\times$)

Experiment

We evaluate our MIN on two benchmark datasets from Sem-Eval challenges.

	#TRAIN/#TEST Sentences	#TRAIN/#TEST Aspects
D_1 (laptop)	3045/800	2358/654
D_2 (restaurant)	2000/676	1743/622

Table: Statistics

Experiment

Baselines

1. **CRF**, **Semi-CRF**: CRF-based models with basic feature templates.
2. **LSTM**: LSTM with word embeddings.
3. **IHS_RD**, **NLANGP**, **DLIREC**, **AUEB**: Top-ranked systems in SemEval competitions.
4. **WDEmb**, **RNCRF**: Current SOTA.

Experiment

Comparison

	D_1	D_2
CRF	74.01%	69.56%
Semi-CRF	68.75%	66.35%
LSTM	75.25%	71.26%
IHS_RD	74.55%	-
DLIREC	73.78%	-
NLANGP	-	72.34%
AUEB	-	70.44%
WDEmb	75.16%	-
RNCRF	77.26%	69.74%
Our MIN	**77.58%**	**73.44%**

Table: Experiment results

Experiment

Architecture ablation

	D_1	D_2
MIN w/o bi-directionality	75.59%	71.87%
MIN w/o S-LSTM	76.04%	72.55%
MIN w/o O-LSTM	75.97%	71.80%
MIN	**77.58%**	**73.44%**

Table: Ablation experiment results

Summary

In this work, we propose Memory Interaction Networks (MIN), a multi-task learning framework, to detect aspect terms from the online user reviews.

- Aspect-Opinion relation is modeled through the memory interaction between ATE task and opinion word extraction.
- A novel LSTM unit with extended memories is developed for memory interactions.
- To further improve the aspect extraction, MIN incorporates a sentimental feature learning component.

Thanks!