ECML-PKDD 2020

Hierarchical Interaction Networks with Rethinking Mechanism for Document-level Sentiment Analysis

Lingwei Wei[*1,3], **Dou Hu**[*2], Wei Zhou[1], Xuehai Tang[1], Xiaodan Zhang[1], Xin Wang[1], Jizhong Han[1], and Songlin Hu[1,3]

1. *Institute of Information Engineering, Chinese Academy of Sciences*
2. *National Computer System Engineering Research Institute of China*
3. *School of Cyber Security, University of Chinese Academy of Sciences*

[*] *These authors contributed equally to the work.*

Background

- **Document-level Sentiment Analysis** (DSA) aims to understand user attitudes and identify sentiment polarity expressed at document-level.
- One of the most active research areas in Natural Language Processing (NLP)
- Applications: intent identification, recommender systems and misinformation detection
- **Challenge:**
  - ✓ ***vague semantic links***
  - ✓ ***complicate sentiment information***
- Recent works leveraging text summarization have achieved promising results but ignore the inherent interactions between the summary and document

Motivation

subject

Summary	Quality is a reflection of **Customer Service**.
Document	... They just sent a new camera and it showed up without any warning or communication about the bad one. Minimal **Customer Service**. That's OK...The 1st camera was promising and worked so well for about two weeks; ...

- The subject Customer Service in the summary can help better predict the key sentiment of the document, such as bad and minimal
- The document can supplement ambiguously semantic features in the summary, such as the details of Customer Service.
- They are complementary.

The auxiliary of the summary is significant for subject mining and semantic understanding in DSA.

Approach

- This paper studies how to effectively generate a discriminative representation with explicit subject patterns and sentiment contexts for Document-level Sentiment Analysis.
- A **Hierarchical Interaction Networks** (HIN) is proposed to explore bi-directional interactions between the summary and document at multiple granularities and learn subject-oriented document representations for sentiment classification.
- Furthermore, we design **a Sentiment-based Rethinking mechanism** (SR) by refining the HIN with sentiment label information to learn a more sentiment-aware document representation.

Experiments

- Datasets
  - ✓ **Toys & Games, Sports & Outdoors**: are parts of Stanford Network Analysis Project.
  - ✓ **Online News Datasets (News)** is from Emotional Analysis of Internet News task.
- Results

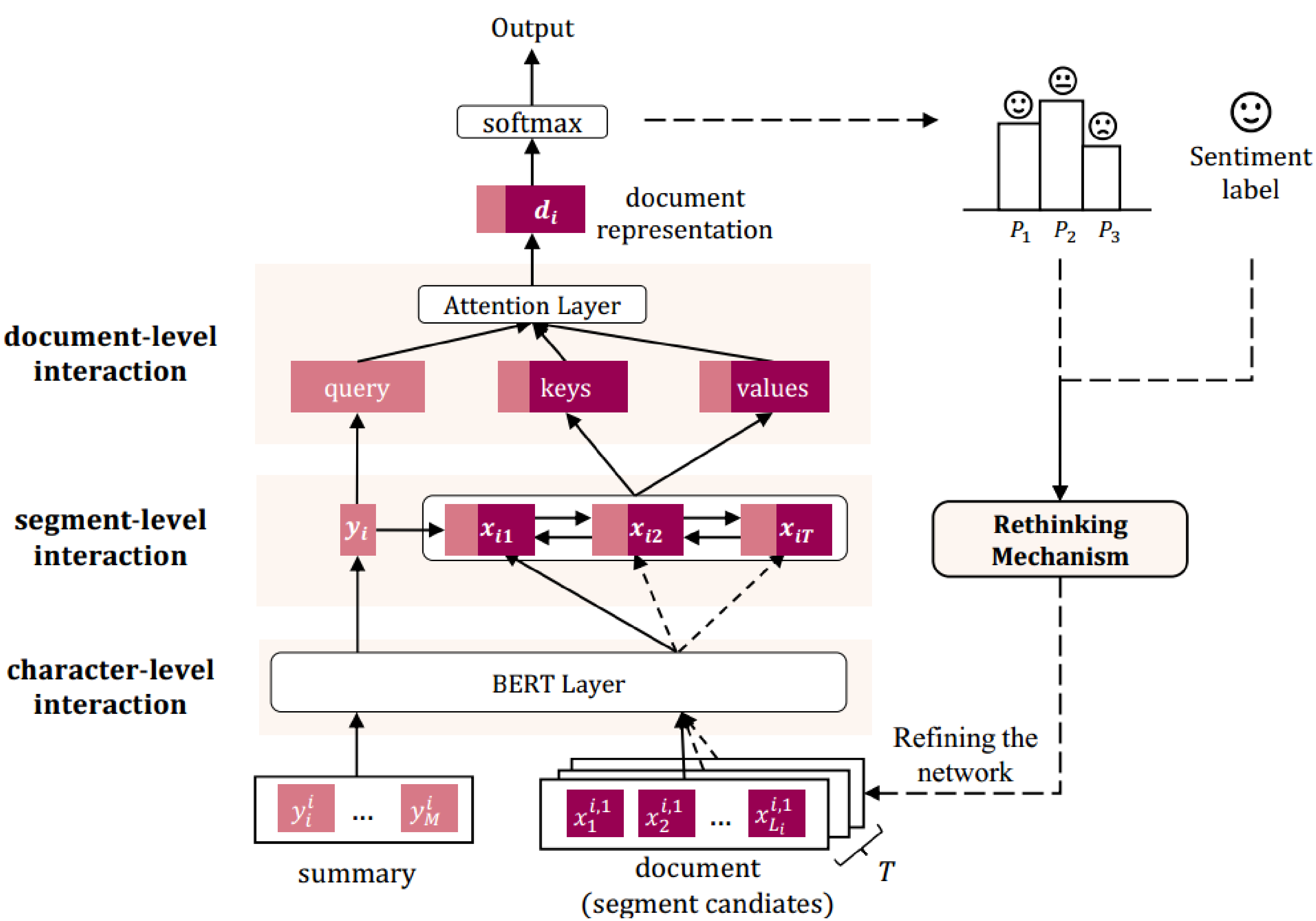
Models	Toys & Games	Sports & Outdoors	News
TextCNN [5]	70.5	72.0	77.5
Bi-LSTM [6]	70.7	71.9	76.5
BERT [7]	75.5	74.2	85.7
BERT(head+tail) [8]	75.9	74.0	86.1
HAN [9]	69.1	72.3	78.6
CAHAN [11]	70.8	73.0	79.9
HSSC [14]	71.9*	73.2*	-
SAHSSC [15]	72.5*	73.6*	-
HIN	**77.5**	**76.7**	**89.0**
HIN-SR	**78.1**	**77.2**	**89.3**

- ✓ The results consistently demonstrate the effectiveness of our proposed models.
- ✓ HIN-SR outperforms various state-of-the-art methods.
- ✓ The proposed model can tackle long documents with the vague semantic links and abundant sentiments effectively.

Outline

- Overview
- Approach
- Experiments
- Conclusion

HIN-SR



Output
softmax
document representation
Attention Layer
document-level interaction
query
keys
values
segment-level interaction
Rethinking Mechanism
character-level interaction
BERT Layer
Refining the network
summary
document
(segment candiates)
T
Sentiment label

Hierarchical Interactions Networks

Character-level Interaction Encoding

$\{[CLS], y^i, [SEP], x^{i,j}, [SEP]\}$ $\xRightarrow{}$ BERT

- ✓ The summary vector $\boldsymbol{y}_i^c$
- ✓ the candidates' vectors $\boldsymbol{x}_{ij}$

Segment-level Interaction Encoding

- ✓ Bi-GRU
- ✓ The candidates' vector at segment-level

$$\overrightarrow{\boldsymbol{h}_{ij}} = \overrightarrow{\mathrm{GRU}}([\boldsymbol{y}_i^c; \boldsymbol{x}_{ij}]), j \in [1, T],$$
$$\overleftarrow{\boldsymbol{h}_{ij}} = \overleftarrow{\mathrm{GRU}}([\boldsymbol{y}_i^c; \boldsymbol{x}_{ij}]), j \in [T, 1].$$
$$\boldsymbol{h}_{ij} = [\overrightarrow{\boldsymbol{h}_{ij}}; \overleftarrow{\boldsymbol{h}_{ij}}]$$

- ✓ The summary vector at segment-level

$$\boldsymbol{y}_i^s = \frac{1}{T} \sum_{j=1}^{T} tanh(\boldsymbol{W}_s \boldsymbol{h}_{ij} + \boldsymbol{b}_s)$$

Hierarchical Interactions Networks

■ Document-level Interaction Encoding

✓ The final document representation

$$\boldsymbol{u}_{ij} = tanh(\boldsymbol{W}_d \boldsymbol{h}_{ij} + \boldsymbol{b}_d),$$

$$\alpha_{ij} = \frac{exp(\boldsymbol{u}_{ij} \boldsymbol{y}_i^s)}{\sum_{j=1}^{T} exp(\boldsymbol{u}_{ij} \boldsymbol{y}_i^s)},$$

$$\boldsymbol{d}_i = \sum_{j=1}^{T} \alpha_{ij} \boldsymbol{h}_{ij}.$$

■ Decoding and Training

✓ The probability distribution of the sentiment label

$$P(l|x,y) = softmax(\boldsymbol{W}_c \boldsymbol{d} + \boldsymbol{b}_c).$$

✓ Loss function

$$L = -\sum_{l=1}^{K} \hat{p}(l,x,y) \log P(l|x,y),$$

Sentiment-based Rethinking Mechanism

- We introduce the **gold sentiment label information** as the high-level features to guide the previous layers based on the current posterior probabilities of these confusable categories.
  - ✓ The state of i-th document $\boldsymbol{s}_i = \boldsymbol{d}_i$,
  - ✓ A feedback layer $\hat{P}(l|x,y) = softmax(\boldsymbol{W}_r \boldsymbol{s}_i + \boldsymbol{b}_r)$,
  - ✓ In the t-th episode, the reward of the sample is defined:

$$r^{(t)} = \lambda r^{(t-1)} + (1-\lambda) log \hat{P}(\hat{l}|x,y),$$

  - ✓ Reweight the lower-level features to enable it to selectively emphasize some discriminative sentiment features, and suppress the feature causing confusion in the classification.
  - ✓ Loss function

$$L' = -\sum_{l=1}^{K} r^{(t)} \hat{p}(l,x,y) log \hat{P}(l|x,y),$$

Outline

- Overview
- Approach
- Experiments
- Conclusion

Experiment Setup

Datasets

- ✓ **Toys & Games** and **Sports & Outdoors**: are parts of Stanford Network Analysis Project (SNAP). These datasets consist of reviews from Amazon spanning May 1996-July 2014.
- ✓ Online News Datasets **(News)** is from *Emotional Analysis of Internet News task*. The dataset is collected from websites including news websites, WeChat, blogs, Baidu Tieba, etc.

Dataset	Total size	Classes	# Summary	# Document
Toys & Games	167,597	5	4.4	99.9
Sports & Outdoors	296,337	5	4.2	87.2
News	14,696	3	23.8	1216.1

Evaluation Metrics

- ✓ Accuracy and F1

Experiment Setup

☐ Compared Methods

- ✓ **TextCNN**: used CNN to learn document representations.
- ✓ **Bi-LSTM** was the baseline that directly took the whole document as a single sequence using a bidirectional LSTM network for DSA.
- ✓ **BERT** was a pre-trained model with deep bidirectional transformer.
- ✓ **BERT(head+tail)** fine-tuned BERT with different truncated methods.
- ✓ **HAN** classified documents via hierarchical attention networks.
- ✓ **CAHAN** extended the HAN by making sentence encoder context-aware.
- ✓ **HSSC** applied a hierarchical end-to-end model for joint learning of text summarization and sentiment classification.
- ✓ **SAHSSC** jointly established text summarization and sentiment classification via self-attention.

Results

- Our proposed models perform the best among all baselines on three datasets.

Models	Toys & Games	Sports & Outdoors	News
TextCNN [5]	70.5	72.0	77.5
Bi-LSTM [6]	70.7	71.9	76.5
BERT [7]	75.5	74.2	85.7
BERT(head+tail) [8]	75.9	74.0	86.1
HAN [9]	69.1	72.3	78.6
CAHAN [11]	70.8	73.0	79.9
HSSC [14]	71.9*	73.2*	-
SAHSSC [15]	72.5*	73.6*	-
HIN	**77.5**	**76.7**	**89.0**
HIN-SR	**78.1**	**77.2**	**89.3**

- **First block**
  - ✓ HIN and HIN-SR outperform traditional methods
  - ✓ HAN and CAHAN with hierarchical attention achieve better results than TextCNN and Bi-LSTM.
  - ✓ BERT and BERT(head+tail) achieve mediocre results.
- **Second block**
  - ✓ HIN and HIN-SR outperform text-summarization methods HSSC and SAHSSC.
- **Third block**
  - ✓ HIN-SR outperforms HIN among three datasets.

Ablation Study of HIN

Comparisons

	Summary	Character-level interaction	Segment-level interaction	Document-level Interaction
HIN	✓	✓	✓	✓
w/o Doc.	✓	✓	✓	×
w/o Doc. and Seg.	✓	✓	×	×
w/o Interactions	✓	×	×	×
w/o Summary	×	×	×	×

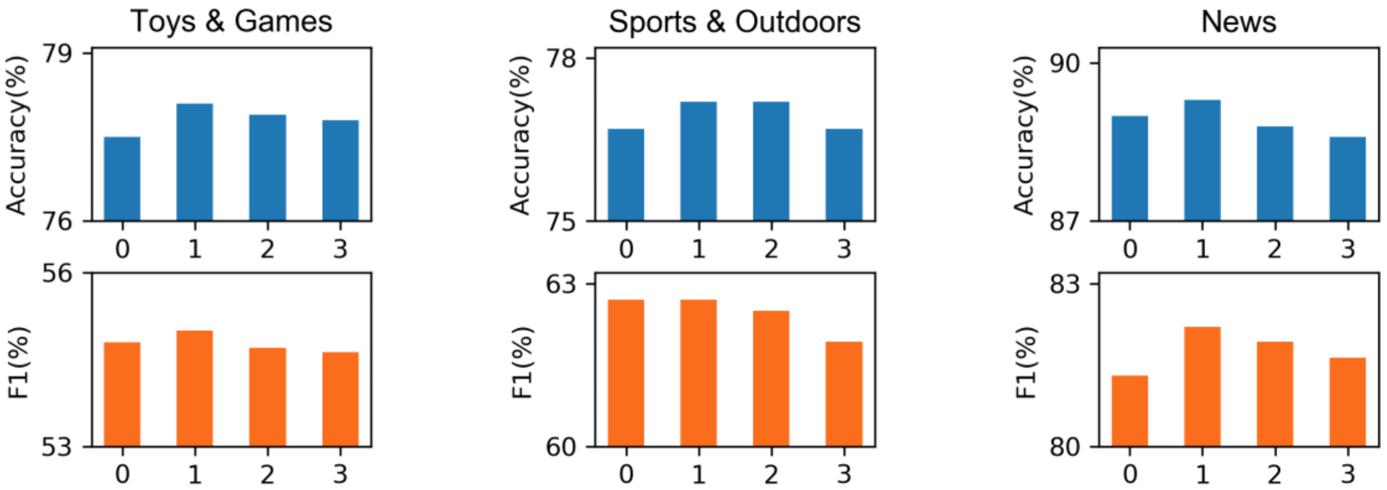
Result

Ablation Settings	Toys & Games		Sports & Outdoors		News	
	Acc	F1	Acc	F1	Acc	F1
HIN	77.5	**54.8**	**76.7**	**62.7**	**89.0**	**81.3**
- w/o Doc.	76.9	53.0	76.2	62.3	88.9	81.1
- w/o Doc. and Seg.	**77.8**	40.0	75.9	62.2	87.9	80.2
- w/o Interactions	76.0	50.6	75.3	61.7	87.8	79.6
- w/o Summary	75.5	49.3	74.2	61.0	85.7	78.1

- ✓ The full model yields significant improvements on F1.
- ✓ The module can aggregate multiple segments with rich subject information.
- ✓ Bi-GRU was able to capture more contextual information.

Analysis of Training Episodes

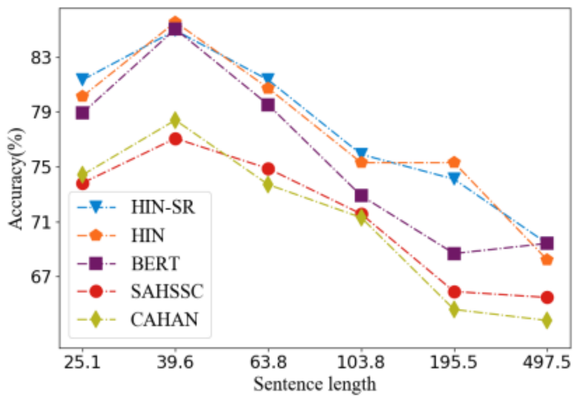
- The number of episodes is an important parameter in SR module.



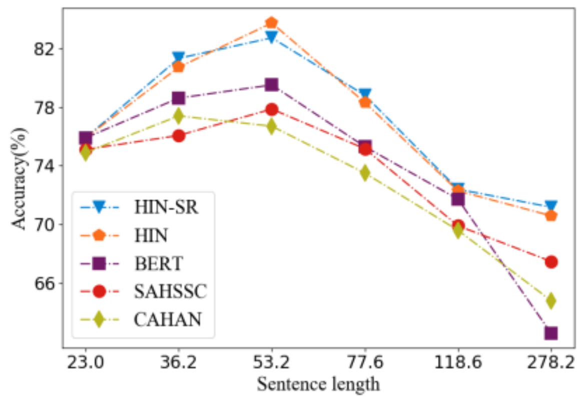
- Result
  - ✓ As the number of training episodes increases, the performance grows significantly and then decline slightly on both accuracy and F1.
  - ✓ The best performance is when the number of training episodes is set to 1.
  - ✓ Three datasets show different meliorations of performance when training episodes is set to 1 than 0.

Analysis of Document Length

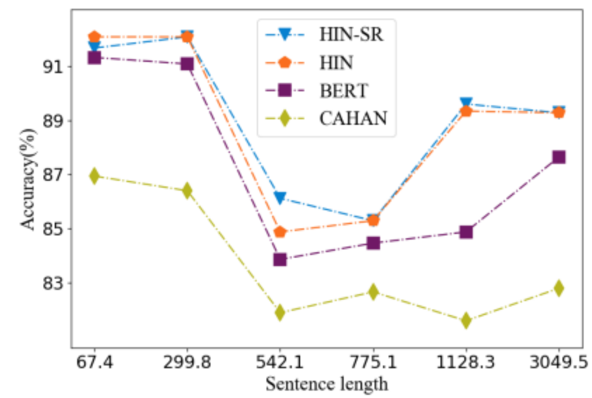
- The testing dataset is divided into six parts according to quantiles of document length.



(a) Toys & Games dataset.



(b) Sports & Outdoors dataset.



(c) News dataset.

- Results
  - ✓ As document length increases, the performance of all models clearly decreases after a certain range.
  - ✓ HIN achieves a considerable improvement than other baselines on three datasets over almost any range of document length.
  - ✓ HIN-SR is not sensitive to document length and is adaptive in both short and long documents.

Case Study

- An example from Toys & Games dataset with a negative sentiment label.

Summary	Rating		Reviews
simulator?	1	1	“a simulater replicates what the real thing is ever used a rc controller? well throttle is on the left pitch yaw, elevator.. on right.. right? well to tally reversed on this ..
		2	so I’m trying build ..instincts on the sticks .. well learn onthis when you got 200 or more on a model, crash your cash its backwards!!! i have flown real airplanes have 83 hrs solo.
		3	this would kill you in the real world.. nowsims are awesome this will waste your model... because... the inputs on this.. mirror revers from spektrum.. futuba.. nice game though...lol forget this #$% ok get a real one.. ok?"

Each block represents one segment candidate of the document. Blue color denotes the word weight after character- and segment-level interaction encoding. Red color indicates candidate weight in the document-level interaction encoding module.

- ✓ HIN addresses the challenge of vague semantic links and complicate sentiment information
- ✓ Subject information is affirmative for capturing accurate sentiment information.
- ✓ Our proposed model can explicitly explore multi-granularities interactions between the summary and document to learn a subject-oriented document representation

Conclusion

- In this work, we have investigated the task of document-level sentiment analysis.
- We design a hierarchical interaction networks (HIN) model to learn a subject-oriented document representation for sentiment classification by exploiting bidirectional interactions between the user-generated summary and document.
- A sentiment-based rethinking mechanism (SR) refines the weights of document features to learn a more sentiment-aware document representation and alleviate the negative impact of noisy data.
- Experimental Results on three widely public datasets have demonstrated that HIN-SR outperforms significantly and tackles long documents with the vague semantic links and abundant sentiments effectively.

Thank you !

Q & A

hudou18@mails.ucas.edu.cn

zhouwei@iie.ac.cn

code