Emotion Analysis in Texts

Rui Xia
Nanjing University of Science & Technology

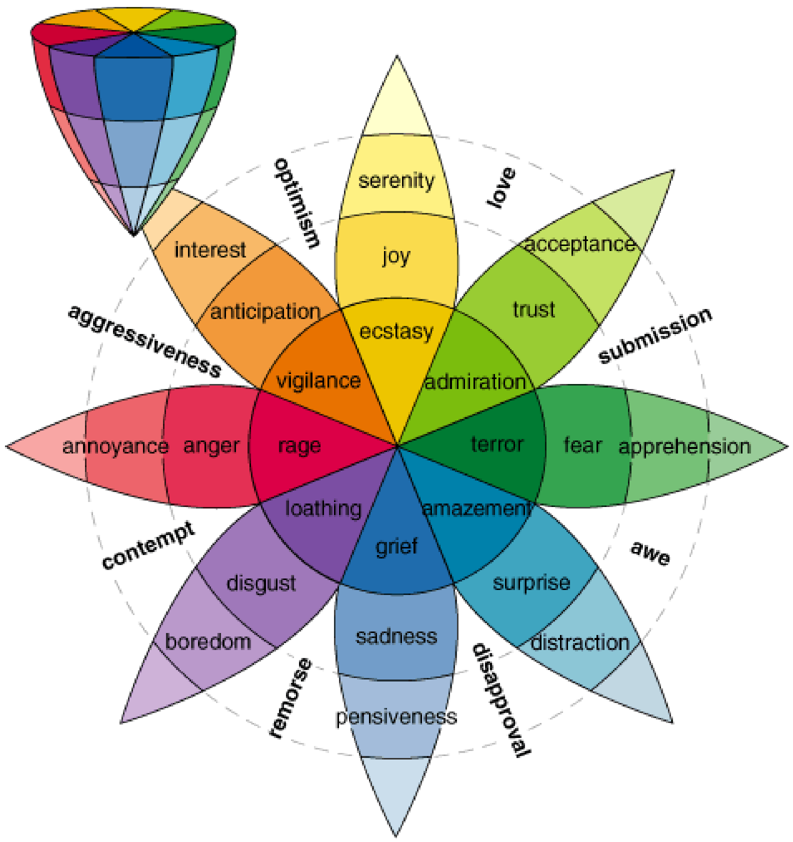
ICDM SENTIRE Workshop
2019.11.8

Text Emotion Analysis Tasks

Two main tasks of emotion analysis in texts

- **Emotion Classification**
  It is an extension to sentiment classification which aims at predicting the people's emotional attitudes in texts (such as joy, anger, sadness, etc.), from the perspective of human psychology.

- **Emotion Cause Extraction (ECE)**
  It is a fine-grained task of emotion analysis with the goal to discover the potential causes that lead to certain emotion expressions in texts.



Plutchik's wheel of emotions (1980)

Emotion Cause Extraction (ECE)

- The ECE task was firstly proposed in [Lee et al., 2010] and was defined as a word-level sequence labeling problem.
- To solve the shortcomings of describing emotion cause at word/phrase level, [Gui et al., 2016a] released a new corpus and re-formalized the ECE task as a clause-level classification problem.
- This framework was followed by most of the recent studies in this field [Gui et al., 2016a; Li et al., 2018; Xu et al., 2019; Yu et al., 2019; Ding et al., 2019; Xia et al., 2019].

```
6,happiness,1,欣喜,-5,no,2013 年 6 月
6,happiness,2,欣喜,-4,no,在 深圳 打拼 10 年 的 吴树粱 终于 拿到 大红 的 深圳市 户口
6,happiness,3,欣喜,-3,no,儿子 吴同 也 随之 迁入 深圳
6,happiness,4,欣喜,-2,no,但 妻子 丁维清 却 必须 等候 吴树粱 入户 满 两年 才能 随迁
6,happiness,5,欣喜,-1,no,半年 后
6,happiness,6,欣喜,0,yes,当初 获得 入户 指标 的 那份 欣喜
6,happiness,7,欣喜,1,no,因为 老 吴 患上 肺癌 晚期 的 噩耗 而 荡然无存
6,happiness,8,欣喜,2,no,取而代之 的
6,happiness,9,欣喜,3,no,是 他 对 自己 生存期 的 忧虑
6,happiness,10,欣喜,4,no,医生 的 判决 是 36 个 月
6,happiness,11,欣喜,5,no,这 意味着
6,happiness,12,欣喜,6,no,老 吴 可能 等 不到 妻子 随 迁入 户 深圳
```

Previous Studies

- **Rule-based methods**
  - RB: rule based method [Lee et al., 2010];
  - CB: common-sense based method [Russo et al., 2011];
- **Traditional machine learning methods**
  - RB+CB+SVM: SVM classifier trained on features including rules [Lee et al., 2010] and Chinese Emotion Cognition Lexicon [Xu et al., 2017];
  - Ngrams+SVM: SVM classifier that uses the unigram, bigram and trigram features. It was a baseline system in [Gui et al., 2017];
  - Multi-Kernel: multi-kernel based structure modeling method [Gui et al., 2016a];
- **Deep learning methods**
  - Word2vec+SVM: SVM classifier using word embeddings of Word2vec as features;
  - Memnet: convolutional multiple-slot deep memory network [Gui et al., 2017];
  - CANN: co-attention neural network with emotional context awareness [Li et al., 2018].

Defined as a Clause Classification Problem

Document

Yesterday morning, a policeman visited the old man with the lost money and told him that the thief was caught. The old man was very **happy (emotion annotation)**. Accompanied by the policeman, he deposited the money in the bank.

Clauses in a document	emotion cause?
Clause 1: Yesterday morning,	no
Clause 2: a policeman visited the old man with the lost money,	yes
Clause 3: and told him that the thief was caught.	yes
Clause 4: The old man was very **happy (emotion annotation)**.	no
Clause 5: Accompanied by the policeman,	no
Clause 6: he deposited the money in the bank.	no

ECE = A set of clause-level binary classification problem?

Part 1. PAE-DGL

Zixiang Ding, Huihui He, Mengran Zhang, and Rui Xia*. From Independent Prediction to Reordered Prediction: Integrating Relative Position and Global Label Information to Emotion Cause Identification. AAAI 2019.

Motivation

clauses in a document	relative position	emotion cause
Clause 1: Yesterday morning,	−3	no
Clause 2: a policeman visited the old man with the lost money,	−2	yes
Clause 3: and told him that the thief was caught.	−1	yes
Clause 4: The old man was very **happy** (emotion annotation).	0	no
Clause 5: Accompanied by the policeman,	+1	no
Clause 6: he deposited the money in the bank.	+2	no

We find that, in addition to the content information, there are two other kinds of information that are also very important for this task:

- **relative position**, which means the position of current clause relative to the emotion clause;
- **global label**, which means the predicted labels of the other clauses in the whole document.

Observations

The percentage of clauses at different positions being an emotion cause

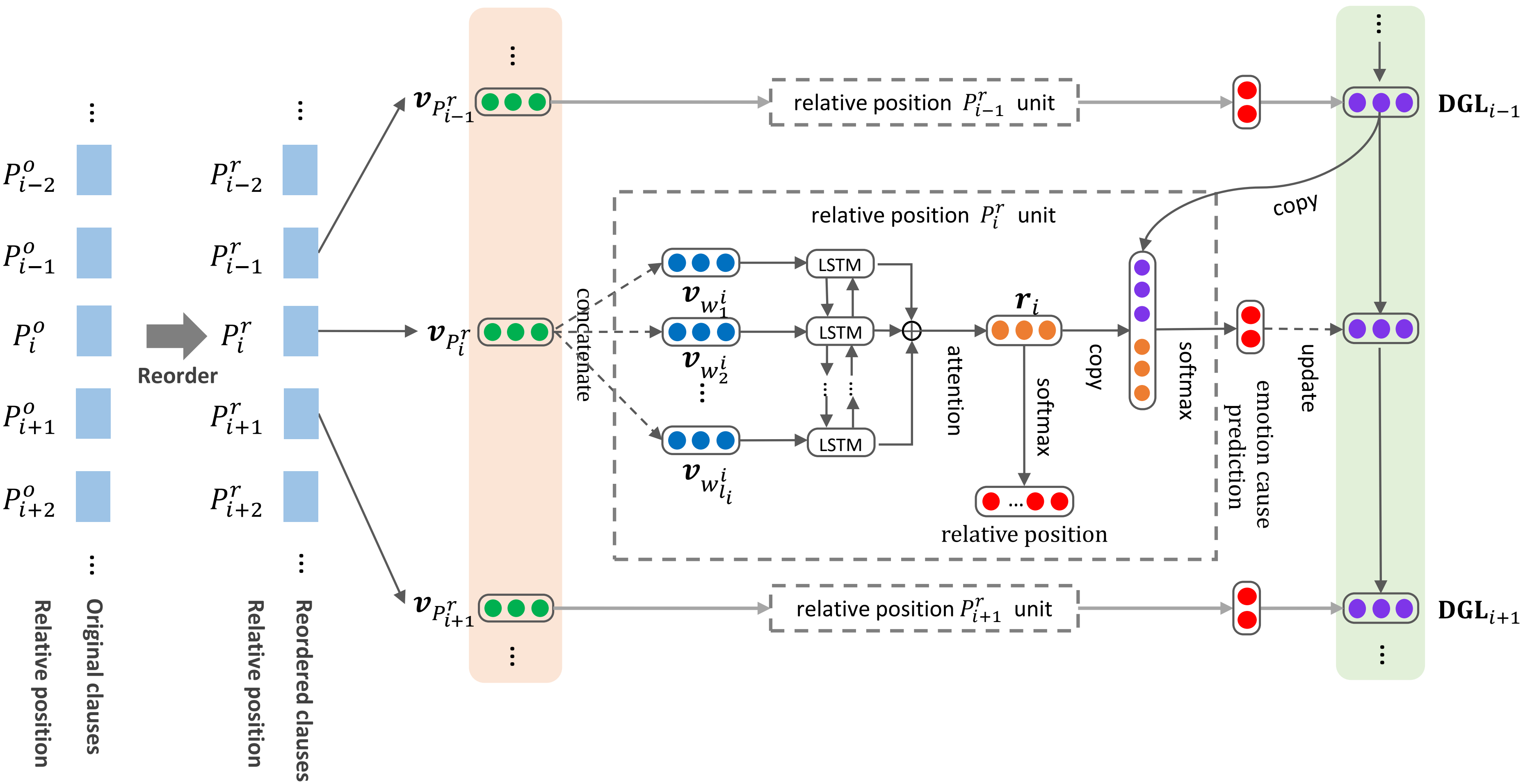
Relative Position	Number	Percentage
-3	37	1.71%
-2	167	7.71%
-1	1,180	54.45%
0	511	23.58%
+1	162	7.47%
+2	48	2.22%
+3	11	0.51%
Others	42	1.94%

The proportion of documents with different number of emotion causes

	Number	Percentage
Document with one cause	2046	97.20%
Document with two causes	56	2.66%
Document with three causes	3	0.14%
All	2105	100%

- clauses that are closer to the emotion expression are more likely to be an emotion cause;
- The clause order ranked by relative position: $0;\ -1;\ +1;\ -2;\ +2;\ -3;\ +3, \ldots$
- If the previous clause in this order has been predicted as a cause (with a high probability), the probability of subsequent clause being a cause should be reduced; otherwise it should be increased.

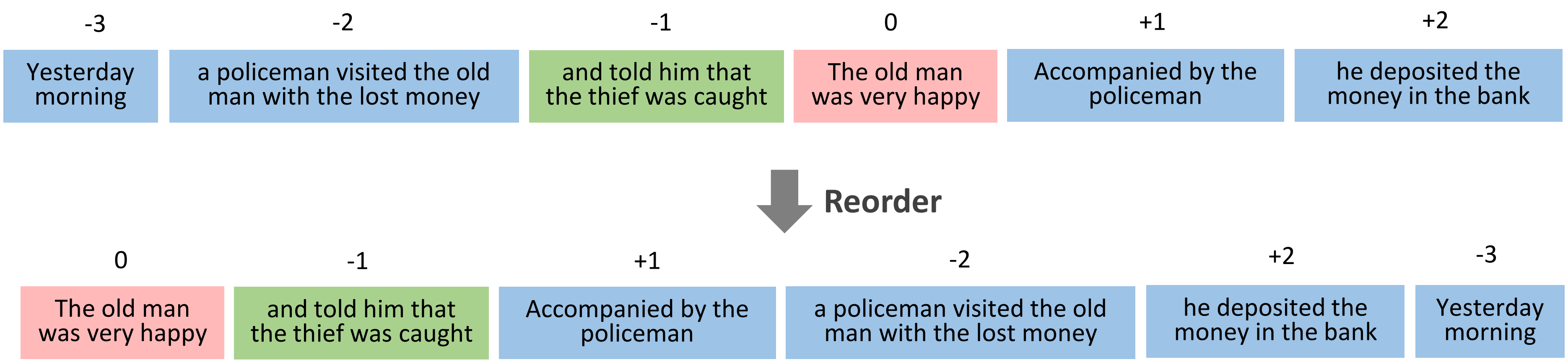
Overall Architecture



From Independent Prediction to Reordered Prediction

- Clause Reordering

We sort the clauses according to the absolute value of relative position in an order (i.e., $0, -1, +1, -2, +2, -3, +3 \ldots$).



before reordering, The order of clauses is $P^o = [-3, -2, -1, 0, +1, +2]$

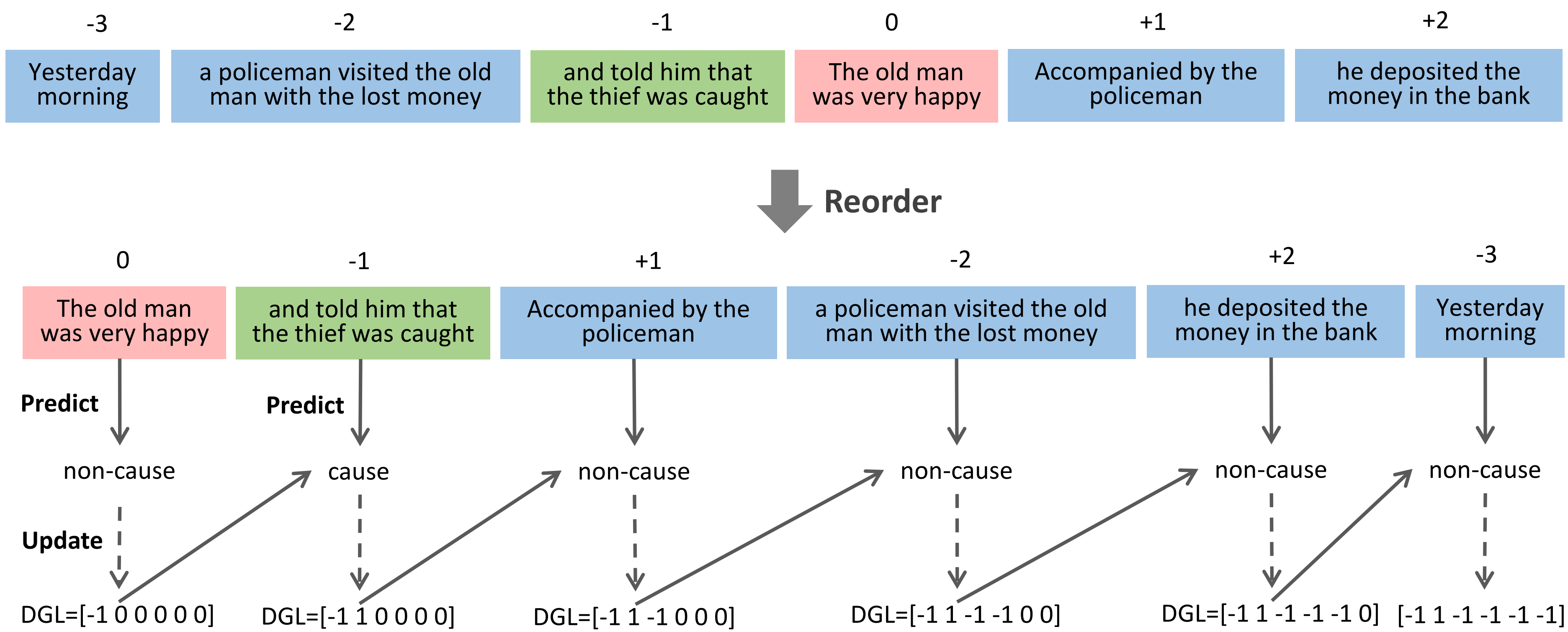
after reordering, The order of clauses is $P^r = [0, -1, +1, -2, +2, -3]$

Dynamic Global Labels (DGL)

- Dynamic Global Label

We use a vector $DGL \in \mathbb{R}^q$ to represent the predictions of all clauses in a document and concatenate DGL_{i-1} and r_i as feature for emotion cause classification.

$$\hat{y}_i = \mathrm{softmax}(W_c[r_i \oplus DGL_{i-1}] + b_c)$$



Experiments

- **Dataset:** Chinese emotion cause corpus (Gui et al. 2016a)
- **Metrics:** Precision, Recall, F1

$$P = \frac{\sum correct\_causes}{\sum proposed\_causes} \qquad R = \frac{\sum correct\_causes}{\sum annotated\_causes} \qquad F = \frac{2 \times P \times R}{P + R}$$

	P	R	F
RB	0.6747	0.4287	0.5243
CB	0.2672	**0.7130**	0.3887
RB+CB	0.5435	0.5307	0.5370
RB+CB+ML	0.5921	0.5307	0.5597
SVM	0.4200	0.4375	0.4285
Word2vec	0.4301	0.4233	0.4136
CNN	0.6215	0.5944	0.6076
Multi-Kernel	0.6588	0.6927	0.6752
Memnet	0.5922	0.6354	0.6131
ConvMS-Memnet	0.7076	0.6838	0.6955
PAE-DGL	**0.7619**	0.6908	**0.7242**

✓ Our method achieves significant improvements in Precision, without reducing Recall.

✓ Our method outperforms the state-of-the-art method by 2.87% in F1 measure.

Experiments

- Ablation Study

	P	R	F
Bi-LSTM	0.5445	0.1663	0.2529
PAE	0.6897	0.6794	0.6836
PAE-DGL	**0.7619**	**0.6908**	**0.7242**

- Different prediction orders

	P	R	F
DGL-P^o	0.6997	0.6561	0.6764
DGL-P^r	**0.7619**	**0.6908**	**0.7242**

- Different ways of modeling positions

	P	R	F
PL	0.7018	0.6496	0.6743
PEC	**0.7081**	0.5867	0.6405
PAE	0.6897	**0.6794**	**0.6836**

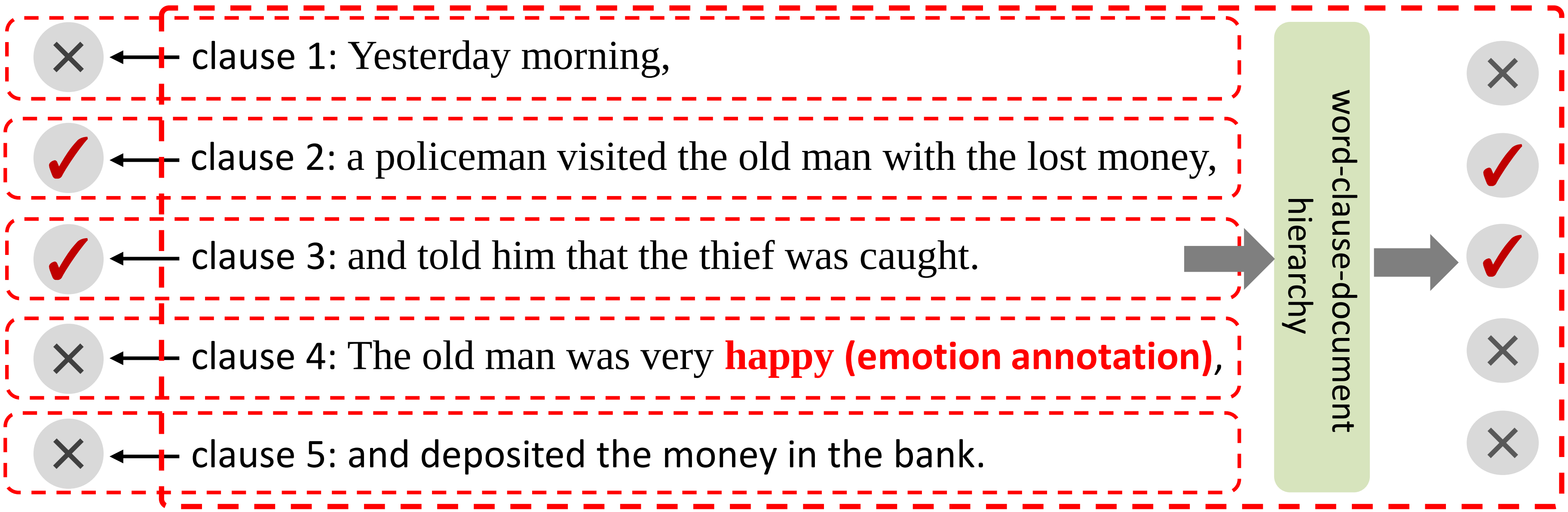
- Upper-bound test of using DGL

	P	R	F
DGL	**0.7412**	0.6866	0.7129
DGL-Upper-Bound	0.7402	**0.7880**	**0.7633**

Part 2. RTHN

Rui Xia, Mengran Zhang, and Zixiang Ding. RTHN: A RNN-Transformer Hierarchical Network for Emotion Cause Extraction. IJCAI 2019.

Motivation



Shortcoming of PAE-DGL: Although PAE-DGL converted the task to a reordered clause classification, its performance depends on the clause order and only the predictions of previous clauses rather than subsequent clauses can be incorporated.

Motivation of this work: A RNN-Transformer Hierarchical Network (RTHN) is proposed to model the relations between multiple clauses in a document and classify them synchronously in a joint framework.

RNN-Transformer Hierarchy

- **Word-level Encoder Based on BiLSTMs**

$$h_{i,t} = \mathrm{BiLSTM}(w_{i,t})$$

- **Clause-level Encoder Based on Transformer**
  - Multi-head Self-attention

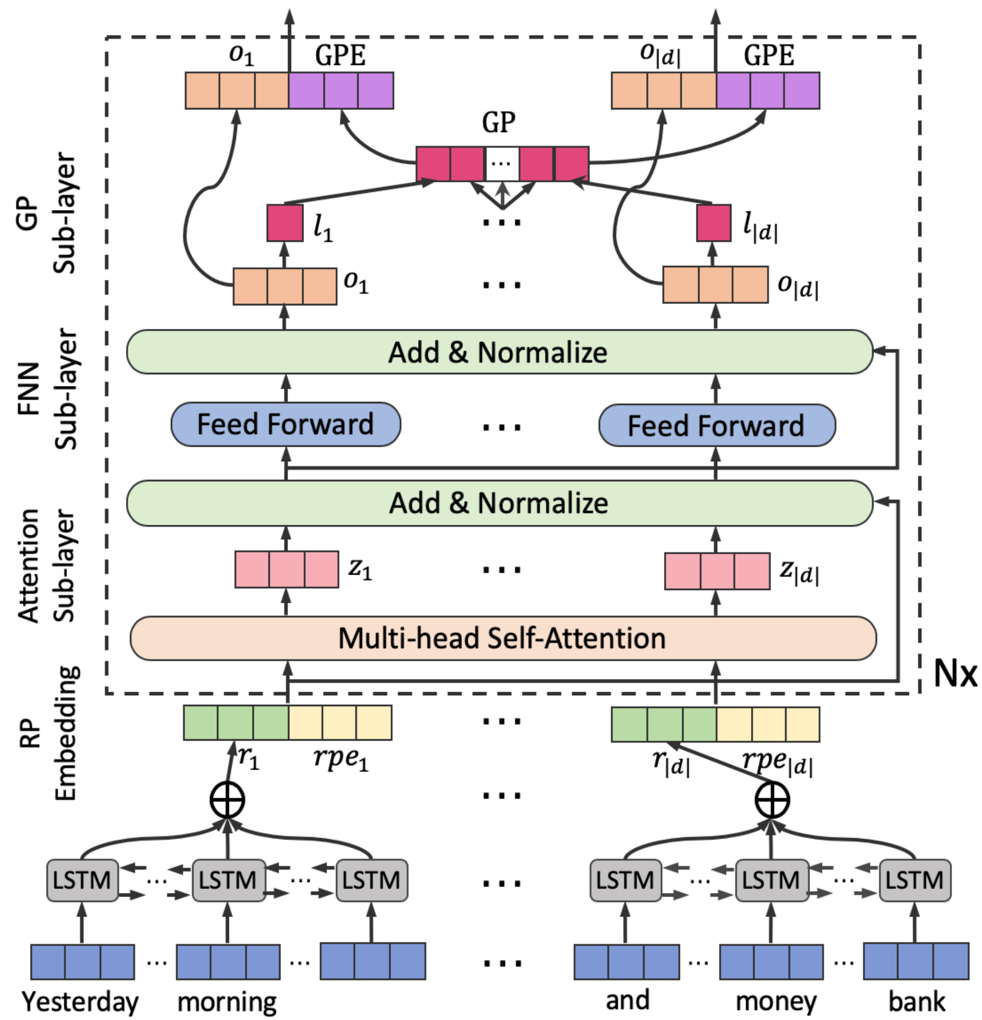
$$\beta_{i,j} = \frac{\exp(q_i \cdot k_i)}{\sum_{j'} \exp(q_i \cdot k_{j'})}$$

  - Feed-Forward Network

$$e_i = \mathrm{ReLU}(z_i W_1 + b_1) W_2 + b_2$$

  - Residual connection and normalization

$$o_i = \mathrm{Normalize}(e_i + x_i)$$



Encoding Relative Position and Global Prediction

- **Relative Position Embedding (RPE)**

$$x_i = r_i \oplus rpe_i$$

- **Global Prediction Embedding (GPE)**
  - The prediction label

$$l_i \leftarrow \mathrm{softmax}(W o_i + b)$$

  - The global prediction(GP) vector

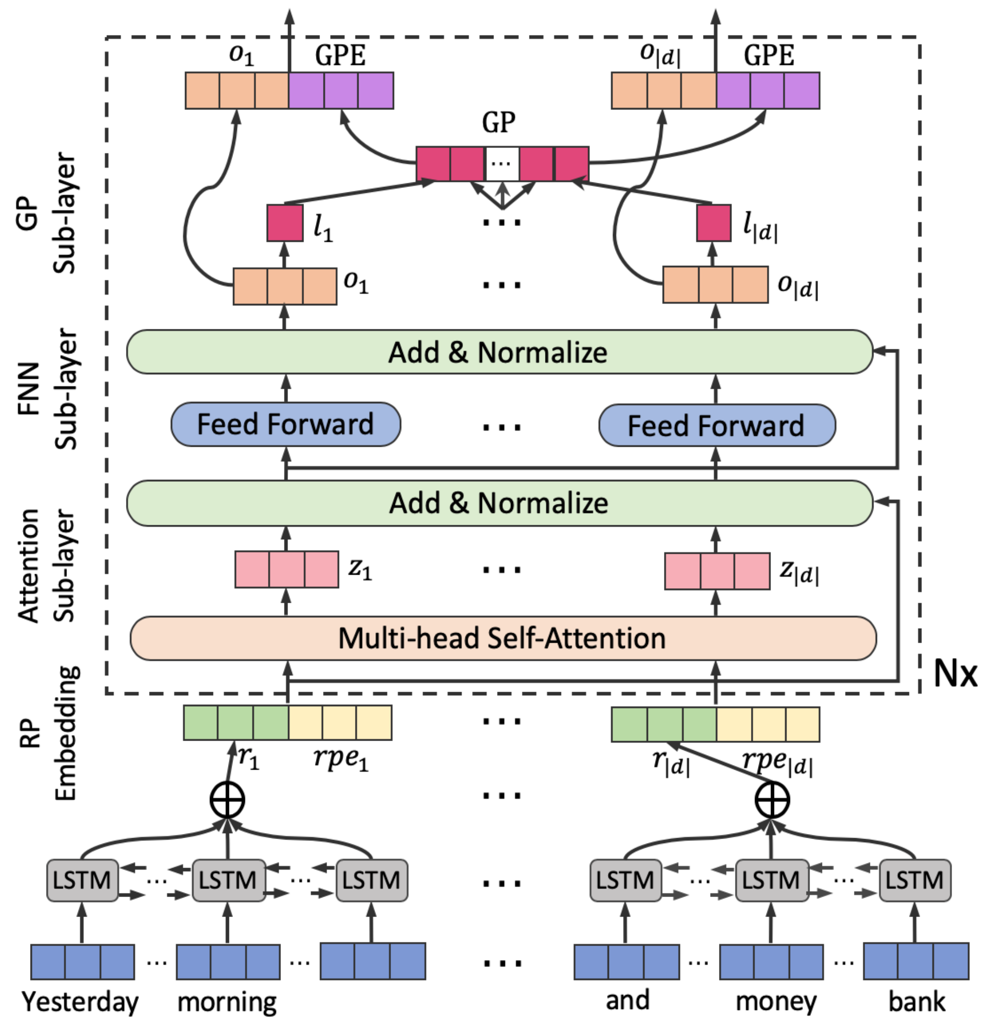
$$GP = [\ldots, l_{i_{-2}}, l_{i_{-1}}, l_{i_0}, l_{i_{+1}}, l_{i_{+2}}, \ldots]$$

  - The Global Prediction Embedding

$$GPE = \mathrm{Tanh}(W_{gpe} GP + b_{gpe})$$

  - Next layer's input

$$x_i^{(l+1)} = o_i^{(l)} \oplus \frac{1}{l} \sum_l GPE^{(l)}$$



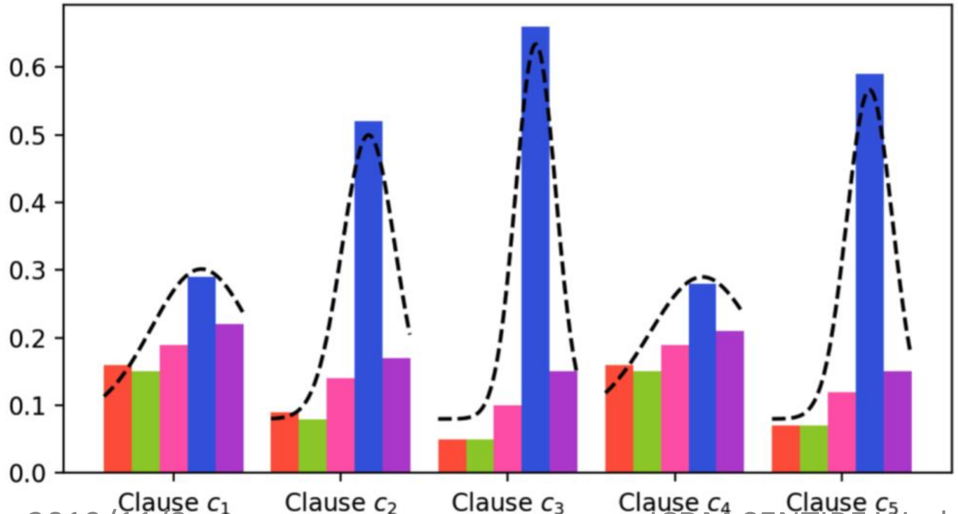
Experiments

	P	R	F1
RB [Lee *et al.*, 2010]	0.6747	0.4287	0.5243
CB [Russo *et al.*, 2011]	0.2672	0.7130	0.3887
RB+CB	0.5435	0.5307	0.5370
RB+CB+SVM	0.5921	0.5307	0.5597
Ngrams+SVM	0.4200	0.4375	0.4285
Word2vec+SVM	0.4301	0.4233	0.4136
Multi-Kernel [Gui *et al.*, 2016a]	0.6588	0.6927	0.6752
CNN [Kim, 2014]	0.6215	0.5944	0.6076
Memnet [Gui *et al.*, 2017]	0.7076	0.6838	0.6955
CANN [Li *et al.*, 2018]	0.7721	0.6891	0.7266
PAE-DGL [Ding *et al.*, 2019]	0.7619	0.6908	0.7242
HCS [Yu *et al.*, 2019]	0.7388	0.7154	0.7269
RTHN (layer 1)	0.7696	0.7333	0.7501
RTHN (layer 2)	0.7644	0.7566	0.7601
RTHN (layer 3)	0.7697	0.7662	**0.7677**
RTHN (layer 4)	0.7604	0.7699	0.7646
RTHN (layer 5)	0.7592	0.7684	0.7634

- Our RTHN model achieves a much higher Recall score than PAE-DGL, without reducing the Precision score (the improvement is more than 7%) ;
- We improve the F1 score of the state-of-the-art from 72.69% to 76.77%.

Experiments

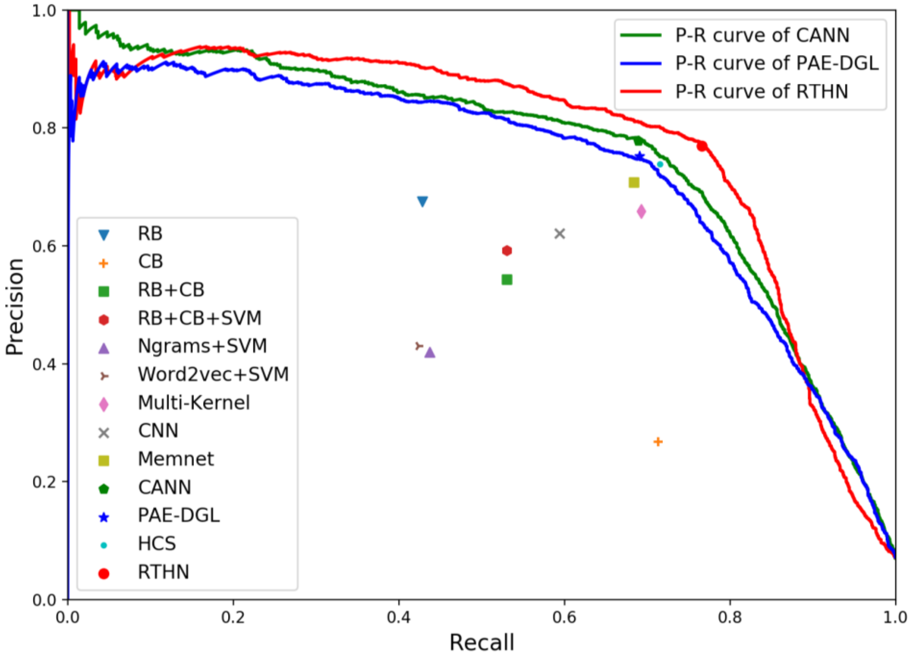
clauses in a document	relative position	emotion cause
Clause 1: Yesterday morning,	-3	no
Clause 2: a policeman visited the old man with the lost money,	-2	yes
Clause 3: and told him that the thief was caught.	-1	yes
Clause 4: The old man was very **happy (emotion annotation)**,	0	no
Clause 5: and deposited the money in the bank.	$+1$	no



- The clause with higher probability being an emotion tends to have more concentrated distribution.
- For clause c_3, the weight at the emotion expression clause is the largest and gradually becomes smaller towards both sides.

Experiments

- The Precision-Recall Curve



- Effects of encoding relative position and global prediction

	P	R	F1
RTHN-No-GPE	0.7369	0.7276	0.7314
RTHN-No-RPE	0.4588	0.3804	0.4145
RTHN-APE	0.5800	0.5618	0.5694
RTHN	0.7697	0.7662	**0.7677**

- Different hierarchy combination

	P	R	F1	Training Time (s)
RRHN	0.7831	0.7273	0.7534	732
TTHN	0.7123	0.6798	0.6952	281
RTHN	0.7697	0.7662	**0.7677**	360

Part 3. ECPE

Rui Xia and Zixiang Ding. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. ACL 2019. (**Outstanding Paper Award**)

Motivation of Emotion-Cause Pair Extraction (ECPE)

Document

Yesterday morning, a policeman visited the old man with the lost money and told him that the thief was caught. The old man was very happy. But he still feels worried as he doesn't know how to keep so much money.

Emotion Cause Extraction (ECE)

happy → a policeman visited the old man with the lost money

happy → and told him that the thief was caught

worried → as he doesn't know how to keep so much money

Emotion-Cause Pair Extraction (ECPE)

(The old man was very happy, a policeman visited the old man with the lost money)

(The old man was very happy, and told him that the thief was caught)

(But he still feels worried, as he doesn't know how to keep so much money)

Shortcomings of the ECE task

- The emotion must be annotated in advance;
- Only supports one emotion in one document;
- Ignores the fact that emotions and causes are mutually indicative.

Advantages of our ECPE task

- Does not need the emotion annotation on test documents;
- Supports multiple emotions and causes in one document;
- Emotions and causes will be learned mutually.

Method

- The Definition of ECPE
  - Input: a document consisting of multiple clauses $d = [c_1, c_2, \cdots, c_{|d|}]$
  - Output: A set of emotion-cause pairs $P = \{\cdots, (c^e, c^c), \cdots\}$

- A Two-step Framework
  - Step 1: Individual Emotion and Cause Extraction
  - Step 2: Emotion-Cause Pairing and Filtering

c1:Yesterday morning,
c2: a policeman visited the old man with the lost money
c3: and told him that the thief was caught
c4: The old man was very happy
c5: But he still feels worried
c6: as he doesn't know how to keep so much money

Step 1

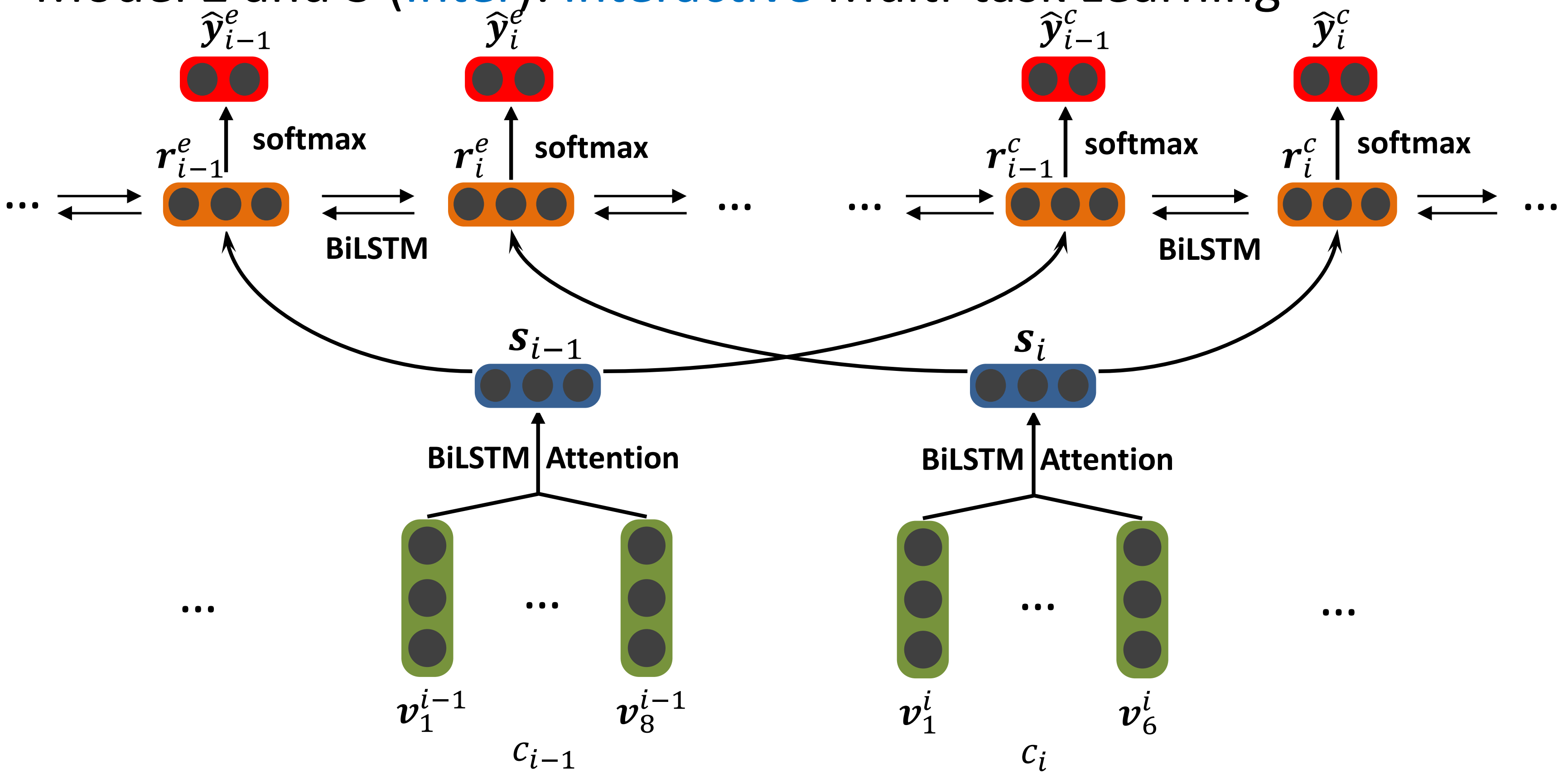
Emotion set: {c4, c5}
Cause set: {c2, c3, c6}

Step 2

Emotion-Cause Pairs: {(c4, c2), (c4, c3), (c5, c6)}

Step 1: Individual Emotion and Cause Extraction

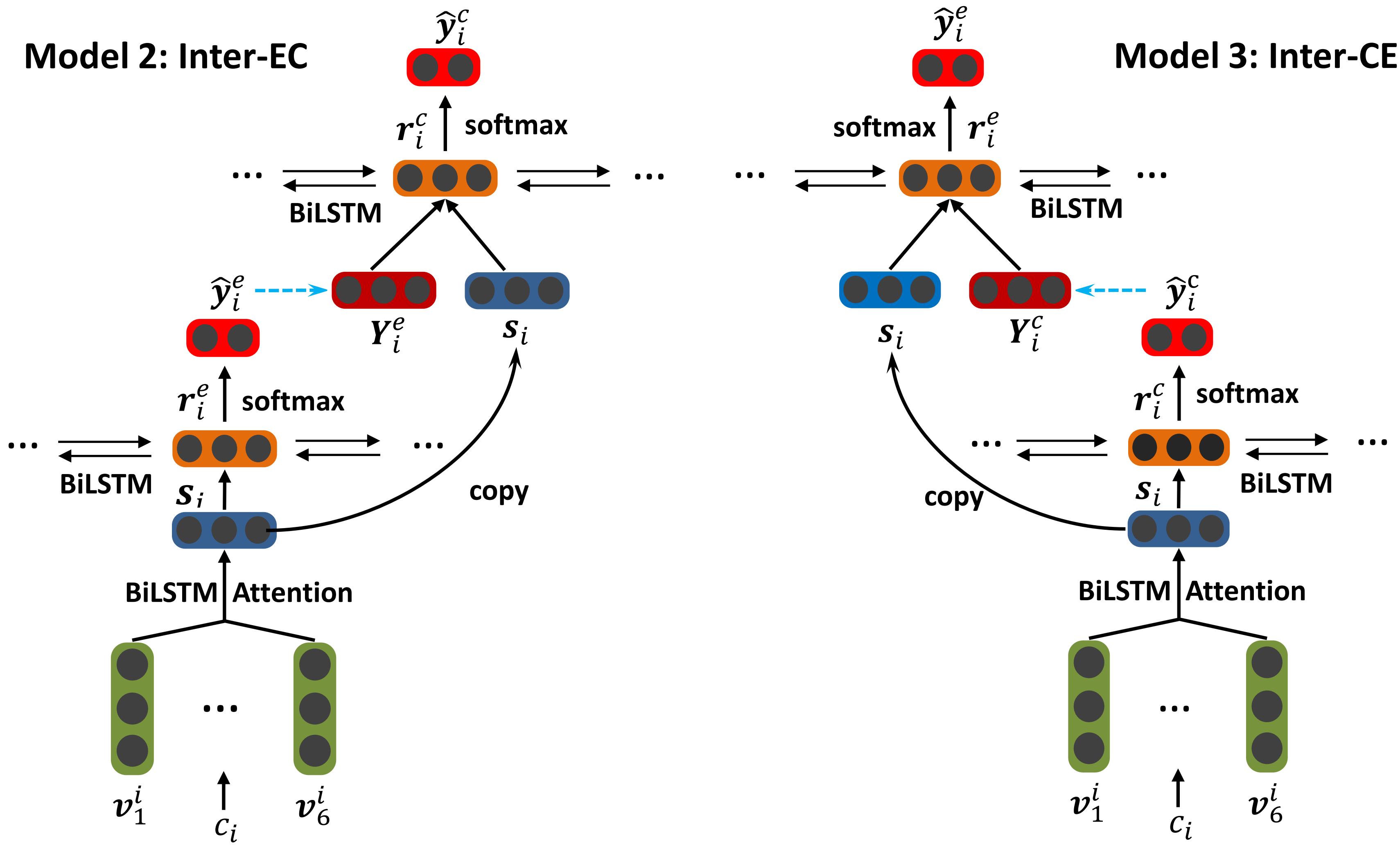
- Model 1 (Indep): Independent Multi-task Learning
- Model 2 and 3 (Inter): Interactive Multi-task Learning



… and told him that the thief was caught The old man was very happy …

Step 1: Individual Emotion and Cause Extraction

- Model 2 and 3: Interactive Multi-task Learning (Inter)



Step 2: Emotion-Cause Pairing and Filtering

(c4, c2)	(The old man was very happy, a policeman visited the old man with the lost money)	Valid
(c4, c3)	(The old man was very happy, and told him that the thief was caught)	Valid
(c4, c6)	(The old man was very happy, as he doesn’t know how to keep so much money)	Invalid
(c5, c2)	(But he still feels worried, a policeman visited the old man with the lost money)	Invalid
(c5, c3)	(But he still feels worried, and told him that the thief was caught)	Invalid
(c5, c6)	(But he still feels worried, as he doesn’t know how to keep so much money)	Valid

All possible Emotion-Cause Pairs: {(c4, c2), (c4, c3), (c4, c6), (c5, c2), (c5, c3), (c5, c6)}

Step 2 - Emotion-Cause Filtering

Valid Emotion-Cause Pairs: {(c4, c2), (c4, c3), ~~(c4, c6), (c5, c2), (c5, c3),~~ (c5, c6)}

Step 2: Emotion-Cause Pairing and Filtering

- Pairing: We first apply a Cartesian product to the emotion set and the cause set, and obtain a set of all possible emotion-cause pairs $P_{all} = \{\cdots, (\mathrm{c}_i^{\mathrm{e}}, \mathrm{c}_j^{\mathrm{c}}), \cdots\}$;
- Meta-learning: We then represent each pair in P_{all} by a feature vector composed of three kinds of features $x_{(\mathrm{c}_i^{\mathrm{e}}, \mathrm{c}_j^{\mathrm{c}})} = [s_i^e, s_j^c, v^d]$. In such a manner we construct a meta-learning training set;
- Filtering: A Logistic regression model is finally trained to detect for each emotion-cause pair $(\mathrm{c}_i^{\mathrm{e}}, \mathrm{c}_j^{\mathrm{c}})$ extracted from the test document, if it is valid or not: $\hat{y}_{(\mathrm{c}_i^{\mathrm{e}}, \mathrm{c}_j^{\mathrm{c}})} \leftarrow \delta(\theta \cdot x_{(c_i^e, c_j^c)})$.

Dataset

Benchmark ECE dataset [Gui et al., 2016]

```
6,happiness,1,欣喜,-5,no,2013 年 6 月
6,happiness,2,欣喜,-4,no,在 深圳 打拼 10 年 的 吴树梁 终于 拿到 大红 的 深圳市 户口
6,happiness,3,欣喜,-3,no,儿子 吴同 也 随之 迁入 深圳
6,happiness,4,欣喜,-2,no,但 妻子 丁维清 却 必须 等候 吴树梁 入户 满 两年 才能 随迁
6,happiness,5,欣喜,-1,no,半年 后
6,happiness,6,欣喜,0,yes,当初 获得 入户 指标 的 那份 欣喜
6,happiness,7,欣喜,1,no,因为 老 吴 患上 肺癌 晚期 的 噩耗 而 荡然无存
6,happiness,8,欣喜,2,no,取而代之 的
6,happiness,9,欣喜,3,no,是 他 对 自己 生存期 的 忧虑
6,happiness,10,欣喜,4,no,医生 的 判决 是 36 个 月
6,happiness,11,欣喜,5,no,这 意味着
6,happiness,12,欣喜,6,no,老 吴 可能 等 不到 妻子 随 迁入 户 深圳
```

```
7,sadness,1,忧虑,-8,no,2013 年 6 月
7,sadness,2,忧虑,-7,no,在 深圳 打拼 10 年 的 吴树梁 终于 拿到 大红 的 深圳市 户口
7,sadness,3,忧虑,-6,no,儿子 吴同 也 随之 迁入 深圳
7,sadness,4,忧虑,-5,no,但 妻子 丁维清 却 必须 等候 吴树梁 入户 满 两年 才能 随迁
7,sadness,5,忧虑,-4,no,半年 后
7,sadness,6,忧虑,-3,no,当初 获得 入户 指标 的 那份 欣喜
7,sadness,7,忧虑,-2,no,因为 老 吴 患上 肺癌 晚期 的 噩耗 而 荡然无存
7,sadness,8,忧虑,-1,no,取而代之 的
7,sadness,9,忧虑,0,no,是 他 对 自己 生存期 的 忧虑
7,sadness,10,忧虑,1,no,医生 的 判决 是 36 个 月
7,sadness,11,忧虑,2,no,这 意味着
7,sadness,12,忧虑,3,yes,老 吴 可能 等 不到 妻子 随 迁入 户 深圳
```

The ECPE dataset

```
6 12
 (6,6), (9,12)
1,null,null,2013 年 6 月
2,null,null,在 深圳 打拼 10 年 的 吴树梁 终于 拿到 大红 的 深圳市 户口
3,null,null,儿子 吴同 也 随之 迁入 深圳
4,null,null,但 妻子 丁维清 却 必须 等候 吴树梁 入户 满 两年 才能 随迁
5,null,null,半年 后
6,happiness,欣喜,当初 获得 入户 指标 的 那份 欣喜
7,null,null,因为 老 吴 患上 肺癌 晚期 的 噩耗 而 荡然无存
8,null,null,取而代之 的
9,sadness,忧虑,是 他 对 自己 生存期 的 忧虑
10,null,null,医生 的 判决 是 36 个 月
11,null,null,这 意味着
12,null,null,老 吴 可能 等 不到 妻子 随 迁入 户 深圳
```

Available at
https://github.com/NUSTM/ECPE/tree/master/data_combine

The proportion of documents with different number of emotion-cause pairs

	Number	Percentage
Documents with one emotion-cause pair	1746	89.77%
Documents with two emotion-cause pairs	177	9.10%
Documents with more than two emotion-cause pairs	22	1.13%
All	1945	100%

Experiments

- Evaluation metrics

$$P = \frac{\sum correct\_pairs}{\sum proposed\_pairs} \qquad R = \frac{\sum correct\_pairs}{\sum annotated\_pairs} \qquad F1 = \frac{2 \times P \times R}{P + R}$$

- Overall performance

	emotion extraction			cause extraction			emotion-cause pair extraction		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
Indep	0.8375	0.8071	0.8210	0.6902	0.5673	0.6205	0.6832	0.5082	0.5818
Inter-CE	**0.8494**	**0.8122**	**0.8300**	0.6809	0.5634	0.6151	**0.6902**	0.5135	0.5901
Inter-EC	0.8364	0.8107	0.8230	**0.7041**	**0.6083**	**0.6507**	0.6721	**0.5705**	**0.6128**

- ✓ Compared with Indep, Inter-EC gets significant improvements on cause extraction. It shows that the predicted emotion is of great help for cause extraction.
- ✓ Similary, Inter-CE gets significant improvements on emotion extraction, due to the help of cause prediction.
- ✓ In emotion-cause pair-extraction, two Inter methods perform significantly better than Indep and Inter-EC is the best.

Experiments

- Upper-Bound of Emotion and Cause Interaction

	emotion extraction			cause extraction			emotion-cause pair extraction		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
Inter-CE-Bound	#0.9144	#0.8894	#0.9016	#1.0000	#1.0000	#1.0000	#0.8682	#0.8806	#0.8742
Inter-EC-Bound	#1.0000	#1.0000	#1.0000	#0.7842	#0.7116	#0.7452	#0.7610	#0.7084	#0.7328

✓ The results on three tasks all prove that emotion and cause are strongly indicative to each other.

- Effect of Emotion-Cause Pair Filtering

	without emotion-cause pair filtering			with emotion-cause pair filtering			
	P	R	$F1$	P	R	$F1$	$keep\_rate$
Indep	0.5894	0.5114	0.5451	0.6832	0.5082	0.5818	0.8507
Inter-CE	0.5883	0.5192	0.5500	**0.6902**	0.5135	0.5901	0.8412
Inter-EC	**0.6019**	**0.5775**	**0.5842**	0.6721	**0.5705**	**0.6128**	0.8889
Inter-CE-Bound	#0.8116	#0.8880	#0.8477	#0.8682	#0.8806	#0.8742	0.9271
Inter-EC-Bound	#0.6941	#0.7118	#0.7018	#0.7610	#0.7084	#0.7328	0.9088

✓ It can be seen that After applying filtering, the performance is significantly improved across different methods, due to the removal of invalid emotion-cause pairs.

Experiments

- Evaluation on the ECE subtask

		P	R	$F1$
use the emotion annotations on the test data	**RB**	0.6747	0.4287	0.5243
	CB	0.2672	0.7130	0.3887
	RB+CB+ML	0.5921	0.5307	0.5597
	Multi-Kernel	0.6588	0.6927	0.6752
	Memnet	0.5922	0.6354	0.6134
	ConvMS-Memnet	0.7076	0.6838	0.6955
	CANN	07721	0.6891	0.7266
do not use the emotion annotations on the test data	**CANN-E**	0.4826	0.3160	0.3797
	Inter-EC	0.7041	0.6083	0.6507

- ✓ Although our approach does not use emotion annotations in test documents, it still achieves comparable results with most of the traditional ECE methods.
- ✓ We furthermore implemented a simplification of CANN which removes the dependency of emotion annotation (CANN-E). It can be seen that Inter-EC has significant advantage.

Source Code and Future Work

- Source Code
  - PAE-DGL: https://github.com/NUSTM/PAEDGL
  - RTHN: https://github.com/NUSTM/RTHN
  - ECPE: https://github.com/NUSTM/ECPE

- Future Work
  - End-to-end joint model for ECPE
  - Evaluation on an English Emotion Cause Corpus
  - Multimodal Emotion Identification and Cause Extraction

Thanks for your attention!