

# 招行论文调研

关于招行任务的一些思考

- 1、优秀话术抽取初步定在句子级别
- 2、感觉任务上更加倾向ECE，情感原因抽取（数据标注情感）

满意度分析一些思考：

满意度分析和对话情感分析有些不同，满意度分析主要受客户最后的几句话的影响（可以结合单个句子的情感识别和对话情感识别）

## 一、数据集调研

目前获取到的数据可用性比较差

shenyang_wanquan_20214256_8	36770	40060	在哪？你要是刚好我在打那动物呢？
shenyang_wanquan_20214256_9	40180	41860	是不是是一个流水机？
shenyang_wanquan_20214256_10	42520	46300	什么是工七位落马的。

主要问题：1、没有情感标注

2、存在很多噪音

需要找一些公开数据集进行先行实验

表 9-1 对话情绪识别常用数据集

Dataset	Dialogues	Utterances	Classes
IEMOCAP	151	7,433	6
AVEC 2012	95	5,798	-
DailyDialog	12,218	103,607	7
EmotionLines	2,000	29,245	7
EmoContext	38,421	115,263	4
MELD	1,433	13,708	7

IEMOCAP。南加州大学的SAIL实验室收集，由人扮演的双人对话，12小时的多模态视听数据。10个专业演员（5男5女），一共分为5个Session，每个Session分配1男1女。对话分为两部分，一部分是固定的剧本，另一部分是给定主题情景下的自由发挥。151段对话，共7433句。标注了6类情绪：Neutral, Happiness, Sadness, Anger, Frustrated, Excited，非中性情绪占比77%。IEMOCAP是对话情绪识别中最常用的数据集，质量较高，优点是有多模态信息，缺点是数据规模较小。

数据集链接：<https://sail.usc.edu/iemocap/>

SEMAINE。SEMAINE数据库收集的多模态对话数据，由四个固定形象的机器人與人进行对话，曾用于AVEC2012挑战赛。AVEC2012使用的数据有95段对话，共5798句。标注了4个情感维度：Valence (愉悦度), Arousal (激活度), Expectancy (预期), Power (力量)。Valence表示情感积极的程度，Arousal表示兴奋的程度，Expectancy表示与预期相符的程度，Power表示情感影响力。其中Valence、Arousal和Expectancy为[-1, 1]范围内的连续值，Power为大于等于0的连续值。SEMAINE是对话情绪识别中常用的数据集之一，缺点是数据规模较小。

数据集链接：<https://semaine-db.eu/>

DailyDialog。高质量多轮对话数据集，纯文本，噪声小，对话反映不同主题的日常生活，无固定说话人。数据集除了7类情绪标注，还有10类主题标注以及4类对话行为标注。12218段对话，共103607句。标注了7类情绪：Neutral, Happiness, Surprise, Sadness, Anger, Disgust, Fear，非中性情绪占比16.8%。DailyDialog在对话情绪识别中应用较少，优点是数据规模较大，缺点是中性情绪占比过高。

数据集链接：<http://yanran.li/dailydialog>

EmotionLines。来源于老友记（多人对话）和私人Facebook聊天记录（双人对话），纯文本，有固定说话人。曾用于SocialNLP 2018 EmotionX Challenge。内容上两部分独立，各1000段对话，共29245句。标注了7类情绪：Neutral, Happiness, Surprise, Sadness, Anger, Disgust, Fear，非中性情绪占比44.5%。EmotionLines在对话情绪识别中应用较少，通常使用多模态的MELD数据集。

数据集链接：<http://doraemon.iis.sinica.edu.tw/emotionlines/index.html>

EmoContext。纯文本双人对话，每段对话三句，仅最后一句有情感标签。用于SemEval-2019 Task 3。38421段对话，共115263句。标注了4类情绪：Happiness, Sadness, Anger, Other，非中性情绪占比42.8%。EmoContext在对话情绪识别中应用较少，优点是数据规模较大，缺点是对话长度过短和仅标注最后一句。

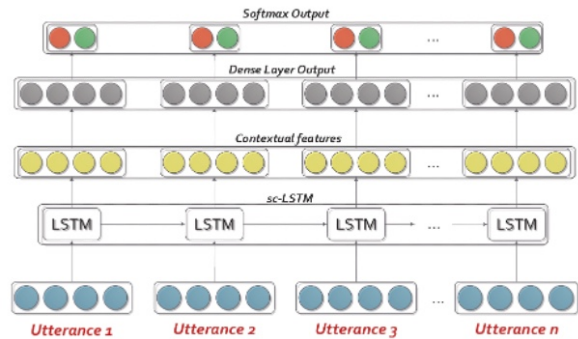
数据集链接：<https://www.humanizing-ai.com/emocontext.html>

MELD。来源于老友记，多人对话形式，是EmotionLines老友记部分的多模态扩充（文本+视频）。1433段对话，共13708句。标注了7类情绪：Neutral, Happiness, Surprise, Sadness, Anger, Disgust, Fear和3类情感：Positive, Negative, Neutral，非中性情绪占比53%。MELD是对话情绪识别中常用的数据集之一，优点是数据集质量较高并且有多模态信息，缺点是数据集中的对话涉及到的剧情背景太多，情绪识别难度很大。

数据集链接：<https://affective-meld.github.io/>

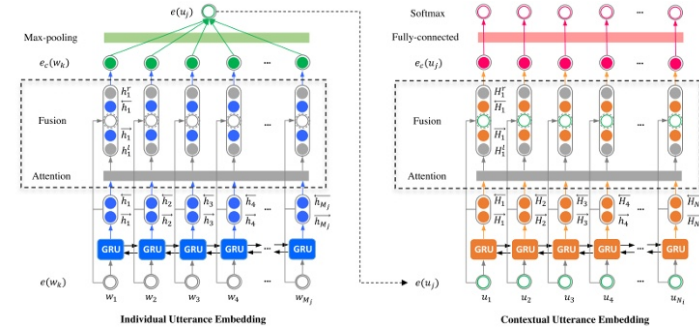
## 二、对话情感识别方法调研

### 1、bc-LSTM



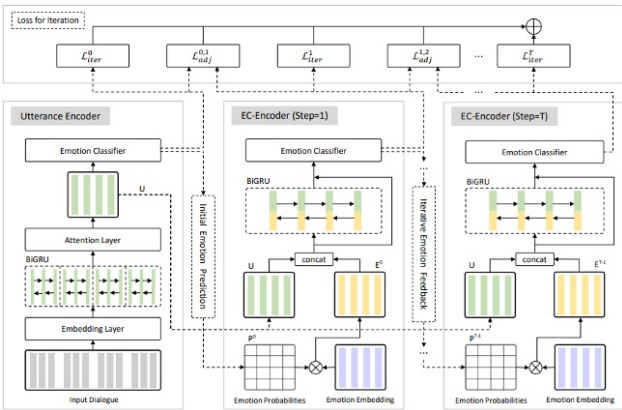
bc-LSTM模型，基于LSTM捕捉上下文信息，其模型架构如图9-4所示。将对话中的话语输入双向LSTM单元，捕捉话语上下文的信息，再经过全连接层得到话语表示，进行情感分类

### 2、HiGRU



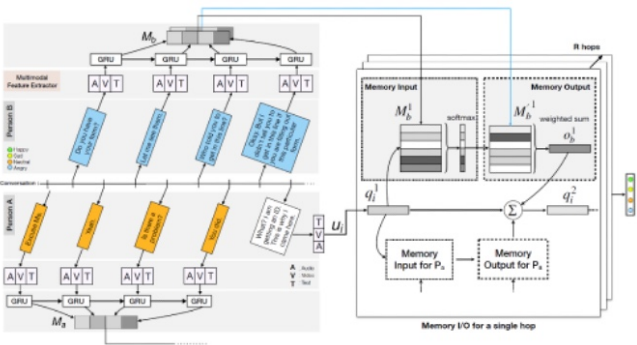
分层的GRU模型HiGRU，下层的GRU用于建模词级别的输入，上层的GRU用于捕捉对话级别的上下文信息。并且采用了单独的特征融合和自注意力特征融合两种方案，以便于更加充分地利用词级别、话语级别的输入和长距离的上下文信息

3、迭代情绪交互模型



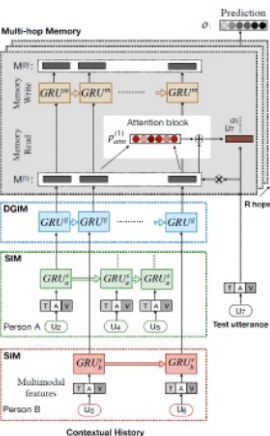
考虑到对话中上下文的话语情绪是相互影响的。现有工作通常都建模对话上下文的内容，以此来隐式地建模话语的情绪交互，但这种做法常被语言中的复杂表达所干扰，导致情绪交互变得不可靠。Lu等人注意到话语的情绪标签可以提供显式且精确的情绪交互，但情绪标签在测试阶段是不可获得作为输入的。为了解决这个问题，其提出了一个迭代情绪交互模型。该模型使用迭代预测的情绪标签代替真实情绪标签，在迭代过程中不断更正预测并反馈输入，实现逐步增强的显式情绪交互。实验结果表明，使用迭代预测标签有效地保留了显式建模的性能优势，并在迭代过程中可以实现有效的预测修正。最终，该方法在两个公开数据集上取得了最好的结果。

4、CMN



CMN模型，对于当前要识别的话语，通过GRU分别建模每个说话者的历史话语，作为记忆单元。然后通过注意力机制将每个说话者的记忆与当前话语的表示进行融合，得到的结果用于话语分类，从而模拟说话者个人的状态以及不同说话者的状态对当前话语的影响。

5、ICON

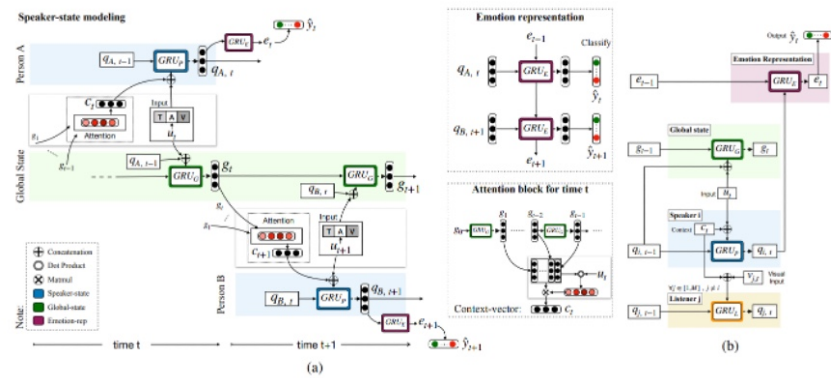


CMN对于不同的说话者采用了独立的记忆单元，在此基础上，Hazarika 等人[5]又提出了ICON模型，使用交互式的记忆单元。

对于当前要识别的话语，ICON通过SIM(Self-Influence Module)模块分别对每个说话者的历史话语进行建模，接着通过DGIM(Dynamic Global Influence Module)模块对说话者之间的影响进行建模得到全局的状态，存入记忆单

元，然后使用Attention机制得到记忆单元与当前话语表示的融合结果，用于话语分类。  
CMN、ICON等模型虽然对不同的说话人信息进行了建模，但是对于最终要识别的话语，并未区分该话语是哪个说话人

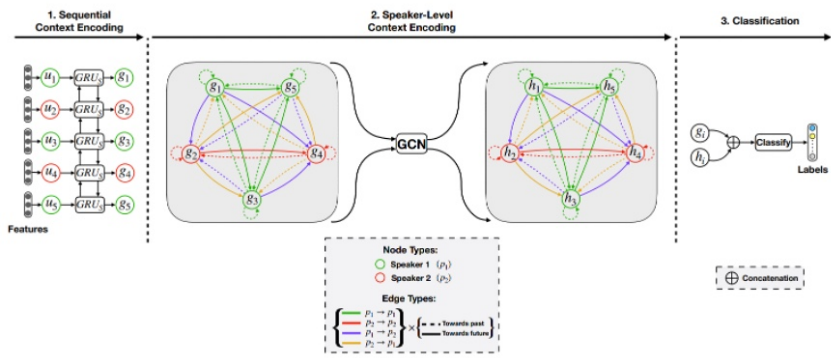
6、DialogueRNN: An Attentive RNN for Emotion Detection in Conversations



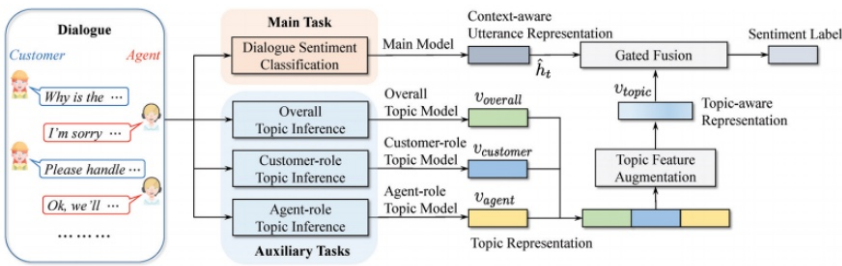
DialogueRNN，认为对话中的话语情感取决于三个因素：说话人信息，先前话语的语境和情感信息，并分别使用GRU对说话者状态、全局状态和情感状态进行捕捉。对于当前时刻的话语，全局状态由前一时刻的全局状态，当前话语的表示，当前话语说话者前一时刻的状态进行更新，说话者状态由当前话语说话者前一时刻的状态，当前话语的表示，之前时刻的全局状态进行更新，情感状态由说话者当前时刻的状态以及上一时刻的情感状态更新，之后用当前时刻的情感状态进行当前话语的分类

7、DialogueGCN

GNN（图神经网络）  
DialogueGCN，先用GRU建模上下文信息得到话语表示，然后以话语作为图的节点，说话者与话语顺序的不同区分边的类型，建立有向图，然后不断更新图得到话语表示，进行情感分类



8、TML



主题感知的多任务学习框架TML，并聚焦在客服对话领域的情绪识别任务。TML通过增加整体、顾客以及客服三个方面的主题推断任务，学习主题增强的话语表示，进而提高对话情绪识别任务的性能。

三、情感原因对抽取（ECPE）



# 1、Emotion-Cause Pair Extraction:A New Task to Emotion Analysis in Texts

EEEI 2019

ECE和ECPE的emotion和cause都是句子级别的

情感原因抽取（ECE）

子句级二分类问题。目的是检测文档中的每个子句，是否是情感的原因

情感原因抽取（ECE）存在的问题（2010提出）：

- 1) 在ECE中，在提取原因之前必须对情感进行注释，这大大限制了其在现实场景中的应用；
- 2) 这种先诠释情感然后提取原因的方法忽略了它们是相互指示的事实

提出情感原因对抽取（ECPE）

提出两步设计方法，首先通过多任务学习进行个体情感提取和原因提取，然后进行情感-原因配对和过滤。

详细版：

Step1:两种多任务学习网络将情绪-原因对提取任务转化为两个独立子任务（情绪提取和原因提取），目的是提取一组情绪从句（E）和一组原因从句（C）

Step2:执行情感原因配对和过滤。我们将这两个集合的所有元素组合成一对（使用笛卡尔积），最后训练一个过滤器来消除不包含因果关系的对。

优势：

不需要先提供情感注释

## 模型详细介绍

Step1 :

Independent Multi-task Learning 和 Interactive Multi-task Learning

后者是在前者的基础上进一步捕捉情感与原因之间相关性的增强版。

(1) Independent Multi-task Learning

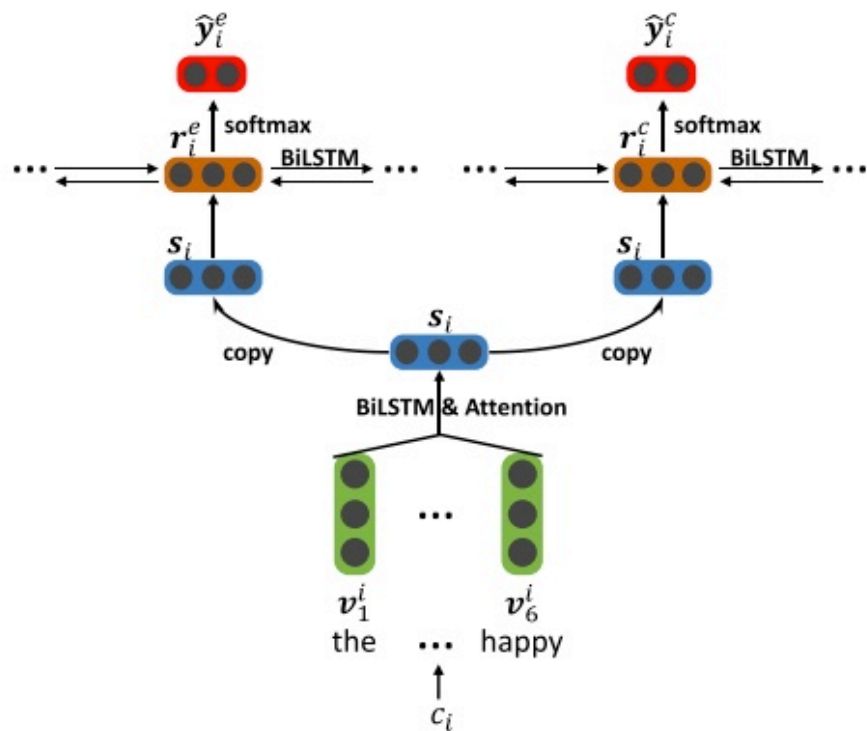


Figure 2: The Model for Independent Multi-task Learning (Indep).

## (2) Interactive Multi-task Learning

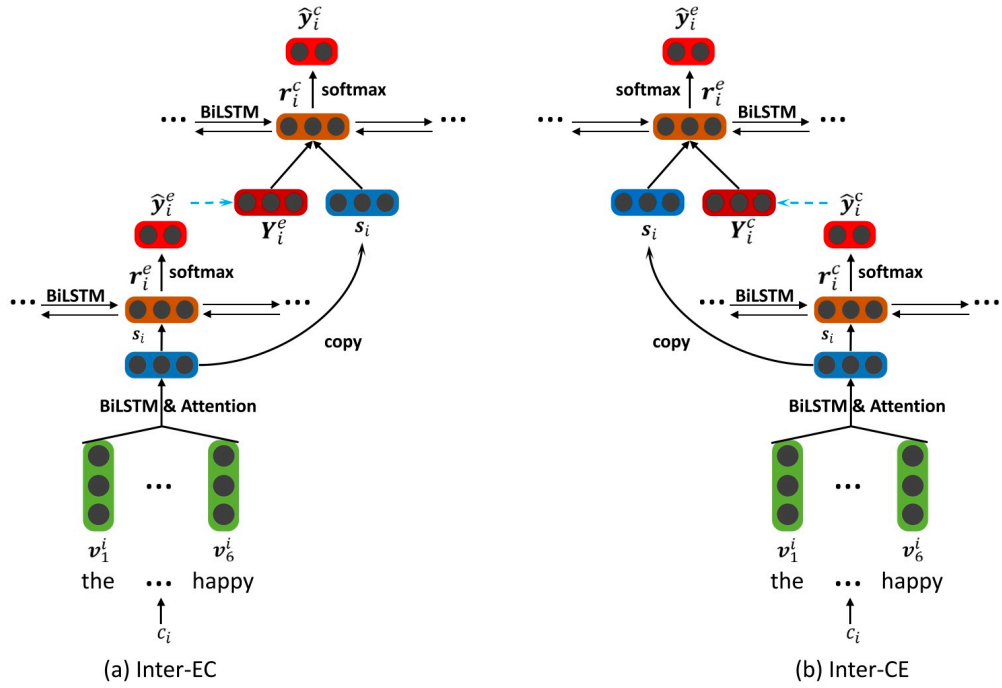


Figure 3: Two Models for Interactive Multi-task Learning: (a) Inter-EC, which uses emotion extraction to improve cause extraction (b) Inter-CE, which uses cause extraction to enhance emotion extraction.

## 实验设置

从ECE数据集（一个情感对应一个或多个原因）上merge得到的数据集

	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%

Table 1: The proportion of documents with different number of emotion-cause pairs in the merged dataset.

1、结果：

Indep (emotion extraction and cause extraction are independently modeled by two Bi-LSTMs)

Inter-CE (the predictions of cause extraction are used to improve emotion extraction.)

Inter-EC (the predictions of emotion extraction are used to enhance cause extraction.)

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

Table 2: Experimental results of all proposed models and variants using precision, recall, and F1-measure as metrics on the ECPE task as well as the two sub-tasks.

2、Upper-Bound of Emotion and Cause Inter- action

使用标签（探究上限，发现对于Inter-EC-Bound与Inter-EC性能有差距，是因为emotion extraction 中存在很多误差）

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

Table 3: Results of upperbound experiments for Inter-CE and Inter-EC.

### 3、Effect of Emotion-Cause Pair Filtering

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

Table 4: Experimental results of all proposed models and variants using precision, recall, and F1-measure as metrics on the ECPE task with or without the pair filter.

### 4、Evaluation on the ECE task

	<i>P</i>	<i>R</i>	<i>F1</i>
<b>RB</b>	0.6747	0.4287	0.5243
<b>CB</b>	0.2672	0.7130	0.3887
<b>RB+CB+ML</b>	0.5921	0.5307	0.5597
<b>Multi-Kernel</b>	0.6588	0.6927	0.6752
<b>Memnet</b>	0.5922	0.6354	0.6134
<b>ConvMS-Memnet</b>	0.7076	0.6838	0.6955
<b>CANN</b>	0.7721	0.6891	0.7266
<b>CANN-E</b>	0.4826	0.3160	0.3797
<b>Inter-EC</b>	0.7041	0.6083	0.6507

Table 5: Experimental results of some existing ECE approaches and our model on the ECE task.

CANN-E 移除测试数据中情感注释的依赖关系

## 2、A Question Answering Approach to Emotion Cause Extraction

EMNLP 2017 引用量: 75

主要思路: 把情感原因任务转换为阅读理解任务 (Yes/No)

模型CNN

优势: extract both word level sequence features and lexical features

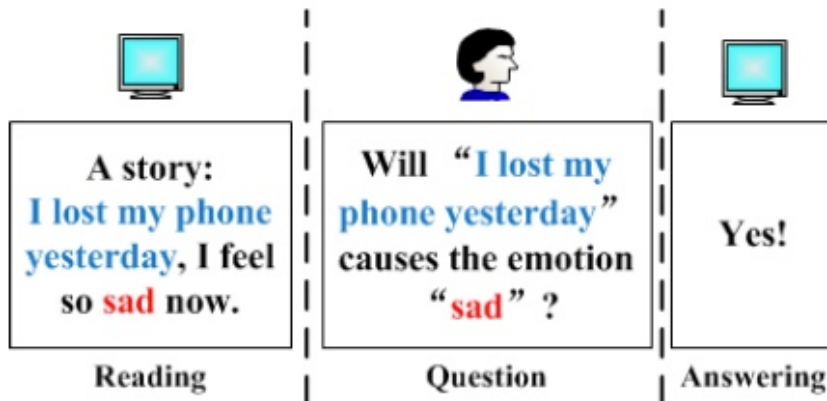


Figure 1: An example of emotion cause extraction based on the QA framework.

## 模型详细介绍

模型首先评估text中每个单词的重要性（注意力机制）

根据注意力结果，把text map into a low dimensional vector space.

这个vector用于生成答案

把text当作阅读理解的文本，把emotion（sad）转换为问题（这个句子causes the emotion “sad”），建模query和text之间的深层关系，

之前的工作，大多依赖于信息提取中的常用方法，如基于规则的模板匹配、序列标注和基于分类的方法。它们大多使用语言规则或词汇特征，但没有考虑语义信息，忽略了情感词与情感原因之间的关系

## 3、Emotion-Cause Pair Extraction as Sequence Labeling Based on A Novel Tagging Scheme

EMNLP 2020

The task of emotion-cause pair extraction deals with finding all emotions and the corresponding causes in unannotated emotion texts.

出发点：之前的工作都用笛卡尔积做，计算量比较大

序列标注，基于新的标记方案（将linked components之间的不一致性编码到tags，从而使情绪和相应的原因能够同时提取）

线性时间复杂度，端到端模型