

李巧媚

2022:

1. Sentiment Word Aware Multimodal Refinement for Multimodal Sentiment Analysis with ASR Errors(ACL,CCF-A)

- Motivation:

实际应用程序只能访问自动语音识别 (ASR) 模型的文本输出, 由于模型容量的限制, 这可能会出现错误, 情感词 (文本情态中的关键情感元素) 被识别为其他词, 这使得文本的情感发生变化, 并直接损害了多模态情感模型的性能。为了解决这个问题, 本文提出了情感词感知多模态细化模型 (SWRM), 该模型可通过利用多模态情感线索来动态细化错误的情感词。

- Method:

首先使用情感词位置检测模块来获取情感词在文本中的最可能位置, 然后利用多模态情感词细化模块来动态细化情感词嵌入, 将细化的嵌入作为多模式特征融合模块的文本输入, 以预测情感标签。

- Experimental Result:

对包括 MOSI-Speechbrain, MOSI-IBM 和 MOSIiFlytek 在内的真实数据集进行了广泛的实验, 结果证明了模型的有效性, 该模型在三个数据集上超过了当前的最新模型, 且该方法可以轻松地适用于其他多模式特征融合模型。([Written by Qiaomei Li](#))

2022:

2. DARER: Dual-task Temporal Relational Recurrent Reasoning Network for Joint Dialog Sentiment Classification and Act Recognition(ACL,CCF-A)

- Motivation:

对于情感预测和行为预测双任务推理, 先前的工作只考虑了参数共享和语义级交互, 而标签信息没有集成到双任务交互中, 因此无法捕获两个任务之间的显式依赖关系。提出了一种新颖的双任务时态关系推理网络 (DARER), 它包括三个模块。对话理解模块通过 SATG 进行特定关系图转换 (RSGT), 以生成上下文、说话者和时间敏感的话语表示。

- Method:

随着对话的进行, 不同说话者的语义状态会发生变化, 这些语义状态转换会触发不同的情绪和行为。SATG 被设计为模拟扬声器语义状态转换, 这为两个任务提供了基本的指示语义。由于时间关系是双任务推理中的关键因素, DRTG 旨在整合任务内部和任务间的时间关系, 使双任务推理过程更加合理和有效。

- Experimental Result:

实验结果证明了该方法的优越性, 它不仅在性能上大大超过了以前的模型, 而且显著节省了计算资源。([Written by Qiaomei Li](#))

2022:

3. Seq2Path: Generating Sentiment Tuples as Paths of a Tree(ACL,CCF-A)

- Motivation:

最近利用生成的方式来解决方面级情感分析任务的方法虽然通过将输出转化情感元组的序列形式从而取得了一些很好的结果, 但是情感元组的顺序在文本中并不是显示存在的, 而且当前时刻下元组的生成也不应依赖于先前的元组。因此本文提出了一种新的生成方式:

Seq2Path, 该方式将情感元组的生成顺序转化为数的路径, 不仅可以有效的应对 1 对 n 的问题 (如一个方面实体对应多个意见词), 而且每个路径的生成都独立而不彼此依赖。

- **Method:**

将每一个元组视为一个独立的目标, 并计算其平均损失, 然后通过“树”的形似生成路径上的 token, 即利用束搜索并行且独立的生成路径, 最后通过一个二分类的判别 token “v” $\in \{\text{true}, \text{false}\}$ 判别器是否为有效路径, 并添加到路径的末尾。

- **Experimental Result:**

在训练阶段, 计算了 Seq2Seq 模型在路径上的平均损失。在推理阶段, 应用了带约束的束搜索 (beam search)。在此基础上, 通过引入附加的 token 自动选择有效地路径, 基本上在 ABSA 的各个子任务当中均取得了最好的结果。 ([Written by Qiaomei Li](#))

钟清山

2021

1. Empathetic Dialog Generation with Fine-Grained Intents (ACL)

- **Motivation:**

现有的方法并没有得到心理学文献的证实, 它们忽略了在人类对话中捕捉到的微妙互动, 其中听者往往表现出更中立的同情意图。听众更有可能回应另一个人的悲伤或愤怒的情绪, 而不是表达相似或相反的情绪

- **Method:**

提出了一种移情对话模型, 能够在更精确的水平上从对话数据中学习情感/意图交互, 并产生相应的移情反应。建模一组细粒度的移情反应意图, 以确保更精确地学习对话数据中显示的情感交互, 为了方便训练过程, 还从开放的子标题语料库中整理了一个大规模的对话框数据集。

- **Result:**

通过一个精心设计的众包实验, 作者评估并演示了作者的模型如何与基线相比产生更多的同理心对话。

(written by qingshan zhong) [BibTex](#)

2021

2. Exploring aspects of similarity between spoken personal narratives by disentangling them into narrative clause types

(EMNLP)

- **Motivation:**

在开发讲故事的机器或推断角色的特征方面, 人们已经做出了大量的努力。然而, 作者通常找不到比较叙事的模型。这项任务对机器来说非常具有挑战性, 因为它们和我们有时一样, 缺乏对相似性意味着什么的理解。

- **Method:**

首先引入了一个现实世界的个人口语叙事语料库, 其中包括来自 594 个视频文本中的 10,296 个叙事子句。其次, 要求非叙事专家在 Labov 的个人叙事社会语言学模型 (即行动、取向和评价条款类型) 下对这些子句进行注释, 并训练一个分类器。最后, 匹配故事, 并探索人们是否含蓄地依赖于拉博夫的框架来比较叙述。

- **Result:**

证明是故事匹配基础的方面也可以用于更广泛的任务，比如寻找事件之间的因果或时间关系，推断心理状态，或者在故事设置中建立行动和情绪。

(written by qingshan zhong) [BibTex](#)

2021

3. Similarity Analysis of Contextual Word Representation Models (IEEE)

• Motivation:

现存的模型适用性受到了语言注释的可用性的限制。此外，不同模型的比较是间接的，通过探测精度，使得不同模型的相似性和差异性难以评论。

• Method:

作者开发了互补的方法来分析上下文词表示的内部和内部相似性。虽然这种相似性分析不能告诉我们关于模型的绝对事实，但它允许在不订阅一种信息的情况下比较表示。作者考虑了几种基于不同信息定位/分布水平的相似性度量：从单个神经元的神经元水平成对比较到完整单词表征的表示水平比较。

• Result:

发现，不同的结构通常有相似的内部表征，但在单个神经元的水平上有所不同。较高的层比较低的层更局部化。比较精细模型和预训练模型，我们发现更高的层次更受其表征和注意力权重微调的影响，而变得更不局限化。

(written by qingshan zhong) [BibTex](#)

黄骏

2022

1、HMAI-BERT: Hierarchical Multimodal Aligment and Interaction Network-Enhanced BERT for Multimodal Sentiment Analysis (ICME CCF-B)

•Motivation:

多模态情感分析方法大多数以文本为主要信息，其他模态为补充信息，但是存在两个局限性：一是多模态数据固有的异构性，不同模态位于不同的特征空间，使得多模态融合困难。二是不同模态的时间序列数据采样率不一致导致的异步性。目前的方法没有能够缓解这两个问题。

•Method:

为了缓解以上问题，作者提出了 HMAI-BERT，一种分层的多模态对齐和交互网络增强的 BERT。在 HMAI-BERT 中，为了提高多模态交互的效率，在融合前引入了一个记忆网络来对齐不同的多模态表示。不同模态的表征在融合前通过共享记忆进行对齐，促进了模态间的相互作用。在对齐之后，提出了一种模态更新方法来解决异步问题，其中每个模态通过与其他模态的交互来加强。此外，通过引入一个融合模块来整合三种增强模态。

•Experimental Result:

作者在 CMU-MOSI 和 CMU-MOSEI 两个基准数据集上进行实验，其 F1 分数比 SOTA 高了 2 个点左右。说明作者提出的模型能够缓解多模态的异质性和异构性问题。([Written by Jun Huang](#))

2、NHFNET: A Non-Homogeneous Fusion Network for Multimodal Sentiment Analysis (ICME CCF-B)

•Motivation:

融合技术是多模态情感分析的关键技术，近年来基于注意的融合方法表现出较高的性能

和较强的鲁棒性，然而这些方法忽略了视觉和听觉在信息密度上的差异，即视觉和听觉具有低水平的信号特征，而文本具有高水平的语义特征。

•**Method:**

作者提出了一种非同质融合网络（NHENet）来实现多模态信息交互，为了克服模态信息密度的差异，鼓励公平的交互，首先设计了一个具有注意聚合的融合模块，使音频和视觉模态的低级特征增强为高级特征。该模块保证了两个模态在进行信息共享前对当前模态的信息进行组织和压缩，整合后的信息必须通过聚合模块进行传递，以实现互补的特征学习，此外该操作补偿了信息强度的差异，克服了跨模态平方复杂度。然后利用跨模态注意实现本文模态的信息强化和视听融合。最后通过一层全连接网络对原始特征和融合特征进行添加，得到情感预测。

•**Experimental Result:**

作者在基准数据集 CMU-MOSEI 上进行实验，其 F1 分数比基线高 1 左右。另外，作者在对齐和未对齐条件下进行了大量实验，证明了所提出模型的有效性。（[Written by Jun Huang](#)）
2021

3、Transformer-based Feature Reconstruction Network for Robust Multimodal Sentiment Analysis （ACM MM CCF-A）

•**Motivation:**

提高数据缺失的鲁棒性已经成为多模态情感分析的核心挑战之一，目前有基于翻译的方法和基于张量正则方法，但是这两种方法都无法处理非对齐序列中缺失的随机模态特征。

•**Method:**

在本文，作者提出了一种基于 Transformer 的特征重构网络(TFR-Net)，以提高模型对非齐模态序列中随机缺失的鲁棒性。首先，采用基于模态内和模态间的注意提取器学习模态序列中每个元素的鲁棒表示。然后，提出一个重构模块来生成缺失的模态特征。通过最小化生成序列与完整模态序列之间的重构损失，模型学会从不完整模态序列中提取语义。此外，利用后期融合策略对丰富的情感序列进行融合，并进行情感预测。

•**Experimental Result:**

作者在 MOSI 和 SIMS 两个基准数据集上进行实验，基线为 TFN、MuIT 和 MISA。其结果均比基线效果要好。另外通过消融实验，也进一步验证了模态融合的有效性。（[Written by Jun Huang](#)）

李外

1.Emotion Detection through Facial Expression using DeepLearning (IEEE,CCF-A)

- motivation:

通过非语言行为检测情绪是一项具有挑战性的任务。非语言行为包括面部表情、手势，如手部动作等。很多应用都是基于面部表情的，例如 HCI，警察局和医院的报警系统，医院里的病人的疼痛可以通过面部表情来识别。

- method:

利用人工神经网络通过面部表情来检测人的情绪。在各种研究中有 6 种预定义的面部表情。它包括快乐、惊讶、悲伤、厌恶、恐惧和愤怒。根据上面的面部表情，我们可以判断出这个人的行为和情绪。

- experimental result:

该模型的平均精度为 89.91%，与现有方法相当。(Written by WaiLi, <https://ieeexplore.ieee.org/document/9702451>)

2. Emotion Recognition in Simulated Social Interactions (IEEE,CCF-A)

- motivation:

社会背景在日常情感互动中扮演着重要的角色，而其他人脸通常在社会情境中提供语境线索。

- method:

利用合成面部表情来探讨社会情感推理机制对社会情绪识别的影响。参与者在屏幕中央的目标头像上判断羞耻-悲伤或愤怒-厌恶的混合面部表情，而背景头像则表达了一种情绪(厌恶、轻蔑和悲伤)或保持中立。

- experimental result:

当虚拟角色参与社交活动时，当情境脸表达厌恶时，目标表情混合的羞耻和悲伤被认为表达更多的羞耻;当情境脸表达悲伤时，目标表情混合的悲伤被认为表达更多的悲伤。

(Written by WaiLi, <https://ieeexplore.ieee.org/abstract/document/8319988>)

3. Expression training for complex emotions using facial expressions and head movements (IEEE,CCF-A)

- motivation:

面部表情和头部动作是人类用来向他人传达自己精神状态的重要模式。有趣的是，人们认为人类正确识别情绪的能力依赖于我们模仿他人呈现给我们的表情的能力。

- method:

提出了一个表情训练界面，鼓励和评估模仿面部表情和头部运动。该系统就用户的面部和头部动作在(a)传达复杂情绪和(b)精确模仿表情方面的成功情况向用户提供反馈。

- experimental result:

根据实验结果，该系统可以作为 ASC 的干预手段，也可以作为演员和其他面向客户的专业人员的培训工具。(Written by WaiLi, <https://ieeexplore.ieee.org/document/7344660>)

张泽浩

1、2022: It is AI' s Turn to Ask Humans a Question: Question-Answer Pair Generation for Children' s Story Books (ACL)

● Motivation:

现有的问题回答(QA)技术主要是为了回答人类提出的问题而创建的。但在教育应用中，为了帮助学生提高叙事理解能力，教师往往需要决定自己应该问什么问题。为了帮助教师在教学过程中的质量，我们希望能够自动生成能够测试学生理解能力各个维度的问答对。

● Method:

我们为这个教育场景设计了一个自动问答生成(QAG)系统:给定一本从幼儿园到八年级的故事书作为输入，我们的系统可以自动生成能够测试学生理解能力各个维度的问答对。我们提出的 QAG 模型架构使用一个新的专家注释的 FairytaleQA 数据集来演示，该数据集有 278 本儿童友好的故事书和 10580 对 QA。

● Experimental Result:

自动和人工评估表明，我们的模型优于最先进的 QAG 基线系统。在我们的 QAG 系统之上，我们还开始构建一个交互式的讲故事应用程序，以便未来在这个教育场景中部署。(Written by Zehao Zhang)

2、2022: Multimodal Dialogue Response Generation (ACL)

- Motivation:

现有的研究只关注于基于检索方法的多模态对话模型的探索，而忽略了生成方法。为了填补空白，我们首先提出了一个新的任务:在给定对话上下文的情况下，一个模型需要生成文本或图像作为响应。

- Method:

学习这样的 MDRG 模型通常需要包含文本和图像的多模态对话，这是很难获得的。由于实践中的挑战，我们考虑 MDRG 时自然假设只有有限的训练示例可用。在这种低资源设置下，我们设计了一个新的会话代理 Divter，以便从整个生成模型中分离依赖于多模态对话的参数。通过这种方法，可以分别从大量纯文本对话和文本对中学习模型的主要部分，然后只用少量的训练例就可以很好地拟合整个参数。

- Experimental Result:

大量的实验表明，我们的方法在自动和人工评价方面都取得了最先进的结果，并可以生成信息丰富的文本和高分辨率的图像响应。(Written by Zehao Zhang)

3、2022: Unsupervised multiple-choice question generation for out-of-domain Q&A fine-tuning (ACL)

- Motivation:

经过预先训练的模型在许多问答基准测试中表现出非常好的性能，特别是在同时对多个问答数据集进行微调时。

- Method:

我们提出了一种生成微调数据集的方法，这要归功于一种基于规则的算法，该算法从无注释的句子中生成问题和答案。

- Experimental Result:

我们表明，最先进的 UnifiedQA 模型可以从这样的系统中获得极大的好处，在物理、生物和化学的多项选择基准测试中，它从未接受过训练。我们进一步证明，通过选择最具挑战性的干扰因素(错误答案)，以及基于预先训练的 RoBERTa 模型的专用排名器，可以获得更好的表现。(Written by Zehao Zhang)

梁家振

1、Is Discourse Role Important for Emotion Recognition in Conversation? (NAACL)

- Motivation

对话是一系列话语，其中每个话语在表达特定情感时都扮演着特定的话语角色。本文提出了一种新的方法来利用话语的潜在话语角色信息来确定其在会话中所传达的情感。

- Method

我们使用变分自动编码器 (VAE) 的变体以无监督的方式对每个话语的上下文感知潜在话语角色进行建模。潜在话语角色表征进一步为话语表征提供了更准确的情感识别的显著线索。

- Experimental result

我们的实验表明，我们提出的方法在三个公开的会话情绪识别数据集上的表现优于最佳报告。这证明了话语的话语角色信息在情感识别任务中起着重要作用，这是前人没有研究过

的。[\(Written by Jiazhen Liang\)](#)

2 、 ConSERT: A Contrastive Framework for Self-Supervised Sentence Representation Transfer (ACL)

- Motivation

尽管基于 BERT 的预训练语言模型在许多下游任务上取得了很高的性能，但证明原生派生的句子表示是崩溃的，因此在语义文本相似性 (STS) 任务上产生了很差的性能。

- Method

在本文中，我们提出了 ConSERT，一种用于自我监督的内容表示转移的对比框架，它采用对比学习以无监督和有效的方式微调 BERT。通过使用未标记的文本，ConSERT 解决了 BERT 派生的句子表示的折叠问题，并使其更适用于下游任务。

- Experimental result

在 STS 数据集上的实验表明，ConSERT 比以前的最先进技术实现了 8% 的相对改进，甚至可以与受监督的 SBERT-NLI 相比。当进一步纳入 NLI 监督时，我们在 STS 任务上实现了新的艺术水平。此外，ConSERT 仅在 1000 个样本可用的情况下获得了可比较的结果，显示了其在数据稀缺情况下的鲁棒性。[\(Written by Jiazhen Liang\)](#)

3、 Contrast and Generation Make BART a Good Dialogue Emotion Recognizer (ACL)

- Motivation

在对话系统中，语义相似的话语在不同语境下可能具有不同的情感。因此，在对话情感识别中，利用说话者依赖性来建模长时间的语境情感关系起着至关重要的作用。同时，区分不同的情感类别是非常重要的，因为它们通常具有语义相似的情感。为此，我们采用有监督的对比学习，使不同的情绪相互排斥，以更好地识别相似的情绪。同时，我们利用辅助响应生成任务来增强模型处理上下文信息的能力，从而迫使模型在不同的上下文中识别具有相似语义的情绪。

- Method

为了实现这些目标，我们使用预训练的编码器-解码器模型 BART 作为主干模型，因为它非常适合理解和生成任务。

- Experimental result

在四个数据集上的实验表明，我们提出的模型在对话情绪识别方面获得了比最先进的模型更有利的结果。消融研究进一步证明了监督对比损失和生成损失的有效性。[\(Written by Jiazhen Liang\)](#)