arXiv:2301.02521v1 [cs.CL] 6 Jan 2023

SAIDS：一种结合方言与讽刺信息的情感分析新方法

AbdelrahmanKaseb 和**MonaFarouk**开罗大学计算机工程系，埃及吉萨 {abdelrahman.kaseb,mona_farouk}@eng.cu.edu.eg

摘要

情感分析已成为各类社交网络的重要组成部分，因为它能帮助决策者更全面地了解用户在生活各个方面的观点。尽管情感分析十分重要，但它仍面临诸多问题，例如讽刺文本所表达的情感，这也是情感分析的一大挑战。本文通过提出一种新型系统SAIDS，解决这一挑战。该系统能够预测阿拉伯语推文的情感、讽刺和方言。SAIDS 将其对讽刺和方言的预测作为已知信息，用于预测情感。系统采用 MARBERT 作为语言模型生成句子嵌入，再将其输入讽刺模型和方言模型，最后把三个模型的输出拼接后输入情感分析模型。本文对多种系统设计方案进行了实验并报告结果。SAIDS 应用于 ArSarcasm-v2 数据集后，在情感分析任务上的表现优于当前最先进的模型。通过联合训练所有任务，SAIDS 在情感分析、讽刺检测和方言识别任务上分别取得了 75.98 FPN、59.09 F1 分数和 71.13 F1 分数。该系统设计可用于提升任何依赖其他任务的任务的性能。

1 引言

情感分析（SA）是自然语言处理（NLP）领域的主要任务之一。它用于观点挖掘，帮助决策者制定决策。情感分析的研究开展得相对较早，例如，Pang 等人(2002) 对电影评论中的情感进行了分析，并将其划分为正面和负面。此后，情感分析成为自然语言处理领域最重要的主题之一，尤其是随着网站和社交媒体平台上的评论数量不断增加。从那时起，英语情感分析领域已开展了大量研究，而阿拉伯语相关研究则相对较少。自从Abbasi 等人(2008) 开始研究阿拉伯语情感分析后，许多研究者也陆续开展了相关研究。如今，已有一些知名的阿拉伯语情感分析模型，例如（Alayba 等人, 2018年; Abdulla 等人, 2013; Abu Farha 和 Magdy,2021年;Elshakankery 和 Farouk,2019年）。当然，阿拉伯语研究面临诸多挑战，其中最棘手的问题之一是阿拉伯语复杂的形态学（Kaseb 和 Farouk, 2016年; Abdul-Mageed, 2019年）。另一项挑战是阿拉伯语方言种类繁多（Abdul-Mageed, 2019年）。此外，所有语言的情感分析都面临一个众所周知的挑战，即讽刺，因为使用讽刺的人会用词语表达与其本意相反的意思。例如，“在这种天气里出门，我真是求之不得！”这表达的是正面情感还是负面情感？由于存在讽刺，我们无法准确判断其情感。

一些相关研究结合情感分析处理英语讽刺检测（OpreaandMagdy，2020；AbercrombieandHovy，2016年；Barbieri 等人，2014）。另一方面，关于阿拉伯语情感与讽刺的研究寥寥无几。已有两项关于讽刺检测的共享任务（Ghanemetal.，2019年），但针对讽刺和情感两者的共享任务仅有一项Abu Farha 等人(2021年)，但每个子任务相互独立，这意味着参赛团队可以为每项任务提交不同的模型。一些参与者在情感和讽刺任务中使用了相同的模型（El Mahdaouy 等人，2021年）。

本文没有将情感训练与讽刺训练分开，而是提出了一种新的模型架构，采用多任务训练，同时训练这两项任务。所提出的架构还有其他改进：首先，它也使用方言进行训练。其次，最初预测出的讽刺和方言会用于情感预测。换言之，情感模型会利用讽刺模型和方言模型的输出。本文的贡献如下：

• 设计一种新颖的模型架构，用于处理依赖于另一项任务的复杂任务，例如依赖于讽刺检测的情感分析。

• 研究新架构的设计方案，并找出最佳方案。

• 在 ArSarcam-v2 数据集上训练模型，并在情感分析任务中取得了 75.98 FPN 的最先进结果。

本文结构安排如下：第2节介绍情感分析、讽刺检测和方言识别的相关工作。第3 节介绍本文所用的数据集并展示数据统计信息。第4节介绍 SAIDS 模型及所有设计方案。第5 节展示实验结果，最后第6 节总结本文。

2 相关工作

SAIDS 涉及三项任务：情感分析、讽刺检测和方言识别。本节将讨论各项任务的现有方法。

2.1 情感分析

阿拉伯语情感分析始于 Abbasi 等人(2008) 的研究。此后，许多研究人员对其进行了发展。起初，研究重点主要放在现代标准阿拉伯语（MSA）上，但随着时间推移，研究人员开始关注阿拉伯方言（Mourad 和 Darwish, 2013; Kaseb 和 Farouk,2021年）。

关于数据集，根据 Alyafeai 等人（2021年）的研究，情感分析数据集超过五十个，其中包括 Elshakankery 等人（ 2021 年）、 Kaseb 和 Farouk（2019）、Kiritchenko 等人（2016）、Rosenthal 等人（2017）以及 Elmadany 等人（2018年）的数据集。由于数据集数量庞大，阿拉伯语情感分析也有大量的系统方法（Abu Farha 和 Magdy，2019；Alayba 等人，2018；El-Beltagy 等人，2017）。根据 Abu Farha 和 Magdy（2021）的比较研究，采用词嵌入的深度学习模型优于传统机器学习模型，而基于 Transformer 的模型又优于前两者。阿拉伯语基于 Transformer 的模型数量相当可观，例如 AraBERT（Antoun 等人，2020）和 MAR-BERT（Abdul-Mageed 等人，2021），大多数阿拉伯语情感分析论文都会使用它们。

2.2 讽刺检测

与阿拉伯语情感分析不同，阿拉伯语讽刺检测迄今尚未受到足够关注。只有少数研究探讨了这一问题，而且阿拉伯语讽刺数据集仍明显不足，例如 Karoui 等人(2017); Abu Farha 等人(2022年). Abbes 等人 (2020)收集了一个讽刺性推文数据集，他们使用主题标签（例如 # sarcasm）来收集数据集。随后，他们构建了多个传统机器学习模型，包括支持向量机、朴素贝叶斯和逻辑回归，最佳 F1 分数为 0.73。

此后，Ghanem 等人(2019年) 在一场关于阿拉伯语讽刺检测的研讨会上组织了一项共享任务。他们通过收集不同主题的推文，并利用主题标签设定类别，构建了数据集。此外，他们还抽取了部分数据集并进行人工标注，作为额外步骤。在这项共享任务中，共有十八支团队参与讽刺检测。Khalifa 和 Hussein (2019年) 是排名第一的团队，取得了 0.85 的 F1 分数。

随后，Abu Farha 等人 (2021年) 基于 ArSarcasm-v2 数据集设置了两项任务：情感分析和讽刺检测。共有 27 支团队参加了研讨会，排名前列的团队在讽刺检测和情感分析任务中分别取得了 62.25 的 F1 分数和 74.80 的 FPN。

2.3 方言识别

阿拉伯语方言识别是一项自然语言处理任务，旨在识别书面文本所使用的方言。它分为三个层级：第一个层级是识别现代标准阿拉伯语、古典阿拉伯语（CA）和阿拉伯方言（McWhorter，2004）。第二个层级是根据五种主要阿拉伯语方言识别方言：埃及方言（EGY）、黎凡特方言（LEV）、NOR、海湾方言（Gulf）和现代标准阿拉伯语（MSA）（El-Haj，2020；Khalifa 等人，2016年；Sadat 等人，2014；Al-Sabbagh 和 Girju，2012年；Egan，2010）。第三个层级是识别国家级方言（Abdul-Mageed 等，2020）。

关于数据集，已有二十多个标注了方言的阿拉伯语数据集。其中最受欢迎的数据集之一是 MADAR（Bouamor 等人，2018年），该数据集以城市级粒度标注了 25 个阿拉伯城市的数据。Abdul-Mageed 等人（2020）组织了一项方言检测共享任务，并发布了三项不同的共享任务。在 2020 年的任务中，共有 60 个团队参与，最佳

结果分别为国家级和城市级方言的 26.78 和 6.39 F1 分数。

3 数据集

ArSarcasm-v2（Abu Farha 等人，2021年) 是本文使用的主要数据集，于 WANLP 2021共享任务中发布，包含讽刺和情感分析两项任务。该数据集约有 1.5 万条推文，分为 1.2 万条训练集和 3000 条测试集；本文使用了 WANLP 2021 发布的同一测试集。每条推文都标注了情感（正面（POS）、中性（NEU）和负面（NEG））、讽刺（真、假）以及方言（现代标准阿拉伯语、埃及（埃及方言）、黎凡特方言（黎凡特方言）、马格里布（NOR）和海湾方言）。该数据集的作者通过众包平台对数据进行标注。该数据集最初由两个数据集合并而成：第一个是 ArSarcasm（Abu Farha 和 Magdy，2020），第二个是 DAICT（Abbes 等人，2020），Abu Farha 等人(2021年) 将这两个数据集合并在一起。

3.1 数据集统计

在本小节中，我们介绍一些促使我们开展 SAIDS 研究的数据集统计信息。ArSarcasm-v2 数据集包含 15,548 条推文，其中 3,000 条用于测试，其余用于训练。表 1 展示了训练集中所有任务标签的样本数量。正如我们所见，大多数数据分别被标记为现代标准阿拉伯语和非讽刺的。

Task	标签	数量
情感	正面	2,180
	中性	5,747
	负面	4,621
讽刺	讽刺的	2,168
	非讽刺的	10,380
方言	MSA	8,562
	EGY	2,675
	Gulf	644
	LEV	624
	NOR	43
总计		12,548

表 1：训练集中情感、讽刺和方言标签的数量

情感标签与讽刺、方言各自之间的关系可见表2。对于情感／讽刺部分，我们可以看到，大约90%的讽刺性推文被标注为负面情感，而大约50%的非讽刺推文被标注为中性情感。另一方面，对于情感／方言部分，我们可以看到，大约50%的现代标准阿拉伯语推文被标注为中性情感，而大约50%的埃及方言推文被标注为负面情感。从该表可以得出结论，关于讽刺和方言的信息将有助于情感分析任务。

	POS	NEU	NEG
非讽刺的	2,122	5,576	2,682
讽刺的	58	171	1,939
MSA	1,405	4 ,486	2 ,671
EGY	506	793	1,376
Gulf	121	259	264
LEV	142	197	285
NOR	6	12	25

表2：训练集上情感标签与讽刺和方言标签的交叉统计

表3 展示了各方言中讽刺性推文的百分比。由于 NOR 推文数量有限，其百分比并不可靠，因此可以看出，埃及人的推文最具讽刺性。这印证了表2 中的事实：大多数 EGY 推文都是负面的，而且大多数讽刺性推文都是负面推文。

方言	讽刺比例
MSA	10.83 %
EGY	34.77 %
Gulf	24.38 %
LEV	22.12 %
NOR	34.88 %

表3：训练集中各方言的讽刺性推文百分比

4 所提系统

本节详细介绍所提系统。SAIDS 同时学习情感分析、讽刺检测和方言识别（多任务训练）。此外，它将讽刺检测和方言识别的输出作为额外输入，传递给情感分析模型，这一过程称为“知情决策”。SAIDS 利用

讽刺类别和方言类别的相关信息来判断情感类别，这两者也都是系统本身的输出。SAIDS 的核心思想基于对数据集统计的分析，如第3节所示：大多数讽刺性推文被归类为负面推文，而大多数现代标准阿拉伯语推文被归类为中性推文。

4.1 系统架构

图1 展示了 SAIDS 架构。该架构由四个主要模块组成。第一个模块是 MARBERTv2 (Abdul-Mageed 等，2021年)，它是一种基于Transformer的模型，输入为推文，输出为一个长度为 768 的向量，即句子嵌入。第二个模块是“讽刺模型”，它是用于识别讽刺的二分类器，输入为句子嵌入，输出两个值，分别对应讽刺性推文和非讽刺推文。第三个模块是“方言模型”，它与“讽刺模型”相同，但输出五个类别（埃及方言、黎凡特方言、NOR、海湾方言和现代标准阿拉伯语）。第四个模块是“情感模型”，它是一个用于识别情感的分类器，输入为句子嵌入、讽刺模型输出和方言模型输出的拼接。

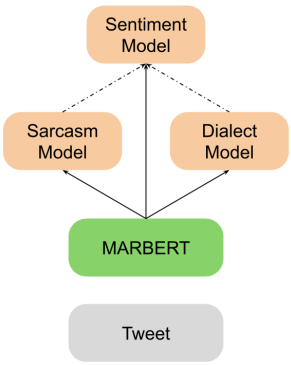

图1：SAIDS架构

情感和方言使用的损失函数是交叉熵。当然，由于讽刺是二元的，因此我们对其使用二元交叉熵。

4.2 训练设置

本小节介绍了为获得最佳模型性能而采用的多种设置。实验采用了多种架构设置和训练策略。

模块架构 我们为“情感模型”“讽刺模型”和“方言模型”测试了多种架构。作为概念验证针对这一想法，我们首先在每个任务模型中构建了一个简单的随机森林模型（随机森林版本）。在实际场景中，我们使用了多层神经网络（MNN）模型。第一种也是最简单的模型只有一个输出层，没有隐藏层。第二种模型包含一层或两层隐藏层，之后是输出层。第三种模型包含一层或两层隐藏层，模块的输出就是隐藏层的输出；这意味着“情感模型”的输入不是“讽刺模型”的输出层，而是其最后一个隐藏层。第四种设置是将最后一个隐藏层与输出层拼接起来，然后传递给“情感模型”。

应传递的信息 SAIDS架构图 1 显示，“情感模型”的输入是“讽刺模型”和“方言模型”的输出，但我们在这一部分尝试了多种设置：仅利用讽刺、仅利用方言，以及同时利用讽刺和方言进行情感分析。

受限反向传播 我们限制了图 1中虚线所示路径上的反向传播，以确保“讽刺模型”和“方言模型”能够正确学习各自的主要目标。当模型未能正确预测情感时，其损失会沿实线直接传播至MARBERTv2模型，而不会沿虚线传播。此外，我们还评估了不限制反向传播的SAIDS，即损失在所有路径上传播的情况，以及部分限制反向传播的情况。只有当“讽刺模型”包含隐藏层时，才能采用部分限制。此时，我们限制损失通过讽刺模型的输出层进行反向传播，但允许其通过隐藏层传播。

激活函数 实验使用 Softmax 作为所有模块输出的激活函数。不过，为了进行比较，我们也进行了不使用 Softmax 处理模块输出的训练，这意味着这些值不在 0 到 1 之间。

按任务训练 由于我们使用同一个模型同时训练三个任务，因此我们尝试先对第一层模型“讽刺模型”和“方言模型”训练若干轮次，然后再对完整系统进行多轮次联合训练。这个想法背后的动机是，只要第一层模型正常工作，情感分析也就会相应地正常工作。我们尝试了多种训练顺序，例如在第一层模型和完整系统之间交替训练，等等。

其他训练参数在我们的实验中，

我们构建了 SAIDS，并使用 HuggingFace 的 Transformers 库（Wolf 等人，2020）提供的 MARBERTv2 模型。除学习率较低时训练了二十个训练轮次外，大多数实验均训练了五个训练轮次。对于学习率，我们使用了从 $1e^{-4}$ 到 $1e^{-6}$ 的范围。序列被截断为最长 128 个词元。所有模型均使用 Adam（Kingma 和 Ba，2015）作为优化器。

5 结果

本节将讨论 SAIDS 取得的结果。为便于比较，我们为该系统构建了基线模型。为了初步评估这一构想本身，我们构建了一个随机森林模型基线，并将其与 SAIDS 的随机森林版本进行比较。针对实际场景的基线包括：基线一（B1），与 HuggingFace 的 BERTModelForSequenceClassification 类完全相同（Wolf 等人，2020），该模型将 MARBERTv2 句子嵌入传递给输出层进行分类；基线二（B2），在分类层之前使用两个隐藏层，其隐藏层大小与“情感模型”的隐藏层大小相同；以及基线三（B3），使用更大的隐藏层，以匹配 SAIDS 模型的训练参数总数。

在评估中，我们采用了数据集原文所述的指标（Abu Farha 等人，2021年）。对于情感分析，指标为负面和正面类别的 F1 分数平均值（FPN）。对于讽刺检测，指标仅为反讽类的 F1 分数（FSar）。对于方言识别，我们采用所有方言 F1 分数的加权平均值（WFS）。

5.1 不同训练设置的结果

本小节展示了不同训练设置的结果，并介绍为所提出的模型选定的最佳设置。在本小节的每个部分中，其他设置均保持不变，以确保比较公平。

模块架构作为我们系统的概念验证，将随机森林（RF）基线模型与信息增强随机森林（IRF）进行比较，后者是 SAIDS 的随机森林版本。表4 显示，IRF 的表现优于 RF，正面指标提高了 3 个百分点，这归功于所提出的架构。从新输入“讽刺模型的输出”和“方言模型的输出”中获得的信息分别为 5 和 4 个百分点。这意味着情感分析决策中约有 10% 来自新添加的信息。

模型	FPN
随机森林	59.36
信息增强随机森林 62.34	

表 4：概念验证在验证集上的性能比较

针对模块的 MNN 架构，我们训练了多个不同隐藏层数量的模型。在每次实验中，所有模块的隐藏层数量都相同。表5 显示，使用零个隐藏层时效果最佳。因此，SAIDS 未采用隐藏层配置。

模型	FPN
0 个隐藏层	75.23
1 个隐藏层	74.90
2 个隐藏层	74.89

表 5：验证集上模块隐藏层数量的性能比较

应传递的信息 此外，我们还进行了实验，以找出分析情感时最适合使用的特征。表6 显示，同时使用方言和讽刺优于只使用其中一项，当然也优于两者都不使用的情况，后者是基线。通过快速观察发现，方言对情感分析的帮助大于讽刺；谈到现代标准阿拉伯语推文时，这一点显而易见，因为其中大多数在情感标注中都被标为中性。因此，SAIDS 使用了讽刺和方言信息。

模型	FPN
未提供信息（B1）	72.40
提供讽刺信息	73.67
提供方言信息	74.41
提供讽刺和方言信息	75.23

表 6：验证集上应传递的信息的性能比较

受限反向传播 我们还进行了实验，以找出反向传播的最佳作用路径。“完全限制”是指损失不会通过“讽刺模型”和“方言模型”传播；“部分限制”是指损失会通过部分层传播

，而“无限制”是指损失会通过所有层传播。在进行这些实验时，模型由两个隐藏层组成。表7显示，“ 部分限制 ”比其他方式取得了更好的结果，但在 SAIDS 中，我们采用了无隐藏层的设置，因此没有使用该方式，而是使用了“完全限制”反向传播。

模型	FPN
完全限制	74.23
部分限制	74.89
无限制	72.31

表 7：验证集上限制反向传播的性能比较

激活函数 为了进行比较，实验中移除了模型输出层中的 Softmax 层。表8比较了这两种设置。结果表明，正如预期，使用 Softmax 比不使用它更好，因为它能够量化属于讽刺或某种方言的概率。因此，在 SAIDS 中，每个模块都使用了 Softmax。

模型	FPN
使用 Softmax	75.23
不使用 Softmax	72.15

表 8：验证集上不同激活函数设置的性能比较

逐任务训练 我们还尝试同时训练三个任务（所有任务），并采用多种训练序列设置。第一种是先对讽刺和方言训练一轮，其余训练轮次用于完整系统（序列 1）。第二种是在奇数轮次训练讽刺和方言，在偶数轮次训练完整系统（序列 2）。第三种是先对讽刺和方言训练两轮，其余轮次仅训练情感（序列 3）。表9显示，序列 1 的表现优于其他序列，因此我们将其用于最终模型训练。

模型	FPN
所有任务	74.35
序列 1	75.23
序列 2	73.49
序列 3	73.01

表9：验证集上不同模型训练序列的性能比较

所用设置概要 SAIDS 使用了讽刺和方言模型的信息，这两个模型都只有一个分类层，没有隐藏层；情感损失不会通过讽刺和方言模型反向传播，并且每个模型的输出都使用 Softmax 激活函数。所采用的训练序列是先对讽刺和方言训练一轮，其余轮次用于完整系统。

5.2 与文献的结果比较

SAIDS经过训练，并与我们构建的基线模型以及当前最先进的模型进行了比较。表10显示，SAIDS在情感分析任务上的表现优于现有的当前最先进模型。SAIDS的主要任务是情感分析，讽刺检测和方言识别则被视为次要输出。尽管SAIDS的FSar分数相当高，但在当前最先进的模型中排名第三。另一方面，大多数取得当前最先进结果的研究都为每项任务使用不同的模型，而在所提出的架构中，一个模型便可同时完成两项任务。该模型还会输出方言，在加权F1分数指标上达到71.13%，但文献尚未报告方言方面的表现，因此表中未列出。

模型	FPN	FSar
基线1	71.60	58.41
基线2	72.53	58.61
基线3	73.11	58.62
El Mahdaouy 等人（2021年）	74.80	60.00
Song 等人 （2021年）	73.92	**61.27**
阿卜杜勒-萨拉姆(2021年)	73.21	56.62
Wadhawan (2021年)	72.55	58.72
SAIDS	**75.98**	59.09

表10：先进模型与 SAIDS 在测试集上的性能比较

6 结论

情感分析是一种重要的系统，广泛应用于决策制定，但也有一些不足，例如难以处理讽刺性句子。在这项工作中，我们提出了 SAIDS，这是一种旨在解决这一问题的新型模型架构。SAIDS 能够在考虑句子的讽刺和方言信息的同时，提升情感分析效果。为此，我们使用经过情感、讽刺和方言标注的 ArSarcasm-v2 数据集进行训练。

SAIDS 的主要目标是预测推文的情感。它经过训练，能够预测方言和讽刺，然后利用这些信息预测推文的情感。这意味着模型在预测情感时，会参考其对讽刺和方言的预测结果。SAIDS 在 ArSarcasm-v2 数据集上的情感预测表现达到了最先进水平；对于负面和正面情感，其平均 F1 分数为 75.98%。在讽刺检测方面，SAIDS 对反讽类取得了 59.09% 的 F1 分数；在方言识别方面，它对所有方言取得了 71.13% 的加权 F1 分数。我们认为，这种模型架构可作为应对情感分析各种挑战的起点。它不仅适用于情感分析，还是一种通用架构，可用于任何任务预测依赖于其他任务的情境。其背后的思路直观明了：针对两项任务进行训练，并将独立任务的输出提供给模型，使其了解相关依赖任务。

参考文献

Ahmed Abbasi、Hsinchun Chen 和 Arab Salem。2008年。多语言情感分析：网络论坛观点分类的特征选择。*ACM Trans. Inf.Syst.*，26(3)。

Ines Abbes、Wajdi Zaghouani、Omaima El-Hardlo 和 Faten Ashour。2020。DAICT：从 Twitter 提取的阿拉伯方言反讽语料库。载于第 *12* 届语言资源与评测会议论文集，第 6265–6271 页，法国马赛。欧洲语言资源协会。

Reem Abdel-Salam。2021年。WANLP 2021 共享任务：使用多头 LSTM-CNN-GRU 和 MaRBERT 对阿拉伯语推文中的反讽和情感进行检测。载于第六届阿拉伯语自然语言处理研讨会论文集，第 306–311 页，乌克兰基辅（线上）。计算语言学协会。

Muhammad Abdul-Mageed。2019年。词汇空间中的阿拉伯语主观性与情感建模。信息处理与管理，56(2)：291–307。阿拉伯语自然语言处理进展（ANLP）及其应用。

Muhammad Abdul-Mageed、AbdelRahim Elmadany 和 El Moatez Billah Nagoudi。2021年。ARBERT 与 MARBERT：用于阿拉伯语的深度双向变换器。载于第*59*届计算语言学协会年会暨第*11*届自然语言处理国际联合会议论文集（第*1*卷：长篇论文），第7088–7105页，线上。计算语言学协会。

Muhammad Abdul-Mageed、Chiyu Zhang、Houda Bouamor 和 Nizar Habash。2020年。NADI 2020：首届细粒度阿拉伯语方言识别共享任务。载于第五届阿拉伯语自然语言处理研讨会论文集，第97–110页，西班牙巴塞罗那（线上）。计算语言学协会。

Nawaf A. Abdulla、Nizar A. Ahmed、Mohammed A. Shehab 和 Mahmoud Al-Ayyoub。2013。阿拉伯语情感分析：基于词典和基于语料库的方法。载于*2013*年*IEEE*约旦应用电气工程与计算技术会议（*AEECT*），第1–6页。

Gavin Abercrombie 和 Dirk Hovy。2016年。将讽刺检测置于语境中：类别不平衡和人工标注对 Twitter 对话监督式机器分类的影响。载于*ACL 2016*学生研究研讨会论文集，第107–113页，德国柏林。计算语言学协会。

Ibrahim Abu Farha 和 Walid Magdy。2019年。Mazajak：一种线上阿拉伯语情感分析器。收录于第四届阿拉伯语自然语言处理研讨会论文集，第192–198页，意大利佛罗伦萨。计算语言学协会。

Ibrahim Abu Farha 和 Walid Magdy。2020。从阿拉伯语情感分析到讽刺检测：Ar-Sarcasm数据集。收录于第四届阿拉伯语开放源代码语料库与处理工具研讨会论文集（含冒犯性语言检测共享任务），第32–39页，法国马赛。欧洲语言资源协会。

Ibrahim Abu Farha 和 Walid Magdy。2021年。阿拉伯语情感分析有效方法的比较研究。信息处理与管理，58(2):102438。

Ibrahim Abu Farha、Silviu Vlad Oprea、Steven Wilson 和 Walid Magdy。2022年。SemEval-2022第6项任务：iSarcasmEval，英语和阿拉伯语中的意图反讽检测。收录于第*16*届国际语义评测研讨会（*SemEval-2022*）论文集，第802–814页，美国西雅图。计算语言学协会。

Ibrahim Abu Farha、Wajdi Zaghouani 和 Walid Magdy。2021年。关于阿拉伯语讽刺与情感检测的 WANLP2021 共享任务概述。载于第六届阿拉伯语自然语言处理研讨会论文集，第 296–305 页，乌克兰基辅（线上）。计算语言学协会。

Rania Al-Sabbagh 和 Roxana Girju。2012年。YADAC：又一个阿拉伯方言语料库。载于第八届国际语言资源与评估会议论文集（*LREC'12*），第

2882–2889页，土耳其伊斯坦布尔。欧洲语言资源协会（ELRA）。

Abdulaziz M. Alayba、Vasile Palade、Matthew Eng- l and 和 Rahat Iqbal。2018年。一种用于阿拉伯语情感分析的卷积神经网络与长短期记忆网络组合模型。收录于机器学习与知识提取，第179–191页，尚。SpringerInternationalPublishing。

Zaid Alyafeai、Maraim Masoud、Mustafa Ghaleb 和 Maged Saeed AlShaibani。2021年。Masader：阿拉伯语文本和语音数据资源的元数据来源。计算研究知识库，abs/2110.06744。

Wissam Antoun、Fady Baly 和 Hazem Hajj。2020。AraBERT：用于阿拉伯语理解的基于Transformer的模型。收录于第四届开放源码阿拉伯语语料库与处理工具研讨会论文集，并设有一项冒犯性语言检测共享任务，第9–15页，马赛，法国。欧洲语言资源协会。

Francesco Barbieri、Horacio Saggion 和 Francesco Ronzano。2014。Twitter讽刺建模：一种新颖的方法。载于第五届主观性、情感与社交媒体分析计算方法研讨会论文集，第50–58页，美国马里兰州巴尔的摩。计算语言学协会。

Houda Bouamor、Nizar Habash、Mohammad Salameh、Wajdi Zaghouani、Owen Rambow、Dana Abdulrahim、Ossama Obeid、Salam Khalifa、Fadhl Eryani、Alexander Erdmann 和 Kemal Oflazer。2018年。MADAR阿拉伯语方言语料库与词典。载于第十一届国际语言资源与评估会议（*LREC 2018*）论文集，日本宫崎。欧洲语言资源协会（ELRA）。

Kathleen Egan。2010。跨语言阿拉伯语博客警报（COLABA）。载于美洲机器翻译协会第九届会议论文集：政府机器翻译用户计划，美国科罗拉多州丹佛。美洲机器翻译协会。

Samhaa R. El-Beltagy、Mona El Kalamawy 和 Abu Bakr Soliman。2017年。 NileTMRG 在 SemEval-2017 任务 4 中的工作：阿拉伯语情感分析。载于 第 *11* 届语义评测国际研讨会论文集（*SemEval-2017*），第 790–795 页，加拿大温哥华。计算语言学协会。

Mahmoud El-Haj。2020年。 Habibi——多方言多国别阿拉伯语歌词语料库。载于 第 *12* 届语言资源与评测会议论文集，第 1318–1326 页，法国马赛。欧洲语言资源协会。

Abdelkader El Mahdaouy、Abdellah El Mekki、Ka- bil Essefar、Nabil El Mamoun、Ismail Berrada 和 AhmedKhoumsi。2021年。用于阿拉伯语讽刺检测和情感分析的深度多任务模型

语言。载于第六届阿拉伯自然语言处理研讨会论文集，第334–339页，基辅，乌克兰（线上）。计算语言学协会。AbdelRahim A. Elmadany、Hamdy Mubarak 和 Walid Magdy。2018年。阿拉伯语推文语音行为与情感语料库。载于第十一届国际语言资源与评估会议论文集（*LREC 2018*）。欧洲语言资源协会（ELRA）。第三届开放源码阿拉伯语语料库与处理工具研讨会（OSACT3）；会议日期：2018年5月8日。

Kariman Elshakankery 和 Mona Farouk。2019年。Hilatsa：一种用于阿拉伯语推文情感分析的混合增量学习方法。《埃及信息学杂志》，20(3)：163–171。

Kariman Elshakankery、Magda Fayek 和 Mona Farouk。2021年。 Lastd：一个经人工标注和测试的大型阿拉伯语情感推文数据集。载于 *2021* 年第五届信息系统与数据挖掘国际会议，ICISDM 2021，第 62–66 页，美国纽约州纽约市。计算机协会。

Bilal Ghanem、Jihen Karoui、Farah Bena mara、Véronique Moriceau 和 Paolo Rosso。2019年。IDAT 在 FIRE2019：阿拉伯语推文反讽检测专题概述。见第十一届信息检索评估论坛论文集，FIRE '19，第 10–13 页，美国纽约。计算机协会。

Jihen Karoui、Farah Banamara Zitoune 和 Véronique Moriceau。2017。Soukhria：面向社交媒体阿拉伯语反讽检测系统。《计算机科学 *Procedia*》，117：161–168。阿拉伯语计算语言学。

Gehad S.Kaseb 和 Mona Farouk。2016年。阿拉伯语情感分析方法：分析性综述。《国际科学与工程研究期刊》，7(10)。

Gehad S. Kaseb 和 Mona Farouk。2019年。扩展版 atsd：阿拉伯语推文情感数据集。《工程与应用科学杂志》，14。

Gehad S. Kaseb 和 Mona Farouk。2021年。一种基于增强型支持向量机的阿拉伯语不同方言推文情感分类方法。《国际电子与计算机科学进展杂志》，8。

Muhammad Khalifa 和 Noura Hussein。2019年。用于阿拉伯语推文反讽检测的集成学习。载于*FIRE 2019* 工作札记——信息检索评估论坛，印度加尔各答，*2019*年*12*月*12*日至*15*日，第2517卷*CEUR* 研讨会论文集，第433–438页。CEUR-WS.org。

Salam Khalifa、Nizar Habash、Dana Abdulrahim 和 Sara Hassan。2016年。大型海湾阿拉伯语语料库。载于第十届国际语言资源与评估会议论文集（*LREC'16*），第4282–4289页，斯洛文尼亚波尔托罗日。欧洲语言资源协会（ELRA）。

Diederik P. Kingma 和 Jimmy Ba。2015年。Adam：一种随机优化方法。载于第三届国际学习表征会议（*ICLR 2015*），美国加利福尼亚州圣迭戈，*2015*年*5*月*7*日至*9*日，会议论文集。

Svetlana Kiritchenko、Saif Mohammad 和 Mohammad Salameh。2016年。SemEval-2016任务7：确定英语和阿拉伯语短语的情感强度。载于第十届语义评测国际研讨会论文集（*SemEval-2016*），第42–51页，美国加利福尼亚州圣迭戈。计算语言学协会。

John H. McWhorter。2004年。《口语阿拉伯语句法评论：摩洛哥、埃及、叙利亚和科威特方言的比较研究》。载于*Language*（第*80*卷），第338–339页。计算语言学协会。

Ahmed Mourad 和 Kareem Darwish。2013。现代标准阿拉伯语和阿拉伯语微博的主观性与情感分析。载于第*4*届主观性、情感与社交媒体分析计算方法研讨会论文集，第55–64页，佐治亚州亚特兰大。计算语言学协会。

Silviu Oprea 和 Walid Magdy。2020。iSarcasm：一个有意讽刺数据集。载于计算语言学协会第*58*届年会论文集，第1279–1289页，线上。计算语言学协会。

Bo Pang、Lillian Lee 和 Shivakumar Vaithyanathan。2002。点赞？使用机器学习技术进行情感分类。载于*2002*年自然语言处理实证方法会议论文集（*EMNLP 2002*），第79–86页。计算语言学协会。

Sara Rosenthal、Noura Farra 和 Preslav Nakov。2017。SemEval-2017任务4：Twitter上的情感分析。收录于 第十一届语义评测国际研讨会（*SemEval-2017*）论文集，第 502–518 页，加拿大温哥华。计算语言学协会。

Fatiha Sadat、Farnazeh Kazemi 和 Atefeh Farzindar。2014。社交媒体中阿拉伯语方言的自动识别。收录于 首届社交媒体检索与分析国际研讨会论文集，SoMeRA'14，第 35–40 页，美国纽约州纽约市。计算机协会。

Bingyan Song、Chunguang Pan、Shengguang Wang 和 Zhipeng Luo。2021年。DeepBlueAI 在 WANLP-EACL2021 任务 2 中：一种基于深度集成的方法用于阿拉伯语中的讽刺和情感检测。载于第六届阿拉伯语自然语言处理研讨会论文集，第 390–394 页，乌克兰基辅（线上）。计算语言学协会。

Anshul Wadhawan。2021年。基于 AraBERT 和 farasa 分词的方法，用于阿拉伯语推文中的讽刺与情感检测。收录于 第六届阿拉伯语自然语言处理研讨会论文集，第 395–400 页，乌克兰基辅（线上）。计算语言学协会。

Thomas Wolf、Lysandre Debut、Victor Sanh、J ulien Chaumond、Clement Delangue、Anthony Moi、Pier- ric Cistac、Tim Rault、Remi Louf、Morgan Funtow- icz、Joe Davison、Sam Shleifer、Patrick von Platen、Clara Ma、Yacine Jernite、Julien Plu、Canwen Xu、Teven Le Scao、Sylvain Gugger、Mariama Drame、Quentin Lhoest 和 Alexander Rush。2020 年。Transformers：自然语言处理领域的最先进技术。载于*2020* 年自然语言处理实证方法会议论文集：系统演示，第 38–45 页，线上。计算语言学协会。