

2020年IEEE国际数据挖掘会议(ICDM)

自下而上和自上而下：用心理语言学和语言模型特征预测人格

Yash Mehta*

盖茨比计算神经科学组 伦敦大学学院

Samin Fatehi*

计算机工程学院
伊朗科学和技术大学

Amirmohammad Kazameini

计算机工程学院
伊朗科学和技术大学

Clemens Stachl

斯坦福大学通信系

Erik Cambria

南洋理工大学计算机科学与工程学院

绍雷-埃特马迪

计算机工程学院
伊朗科学和技术大学

摘要

最先进的文本数据个性预测大多依赖于自下而上的、自动化的特征生成，作为深度学习过程的一部分。更多的传统模型依赖于手工制作的、基于理论的文本特征分类。我们提出了一个新的基于深度学习的模型，它将传统的心理语言学特征与语言模型嵌入结合起来，从Essays数据集的Big-Five和Kaggle数据集的MBTI中预测人格。通过这种方法，我们实现了最先进的模型性能。此外，我们使用可解释的机器学习来可视化和量化各种语言特征对各自人格预测模型的影响。最后，我们讨论了这项工作对计算建模和心理科学的潜力。¹

索引词—语言模型，自动人格预言，心理语言学特征，NLP

2374-8486/20/\$31.00 ©2020 IEEE doi:
10.1109/icdm50108.2020.00146

I. 简介

人格特征通常被称为相对稳定的思想、情感和行为模式，与一系列重要的生活结果和选择相关[1]，[2]。具体来说，人格特质与个人（如幸福感、精神病态）、人际（如关系满意度）和社会机构结果（如职业选择、工作成功；[3]，[4]）都有反复的关系。因此，人们对开发能够利用人类行为和偏好的在线数据（即数字足迹）来自动预测个人的人格特征水平的模型越来越感兴趣，以用于推荐系统[5]，[6]，产品和服务的个性化[7]，[8]工作筛选[9]，社会网络分析[10]，和情感分析[11]。

A. 人格理论和评估

随着时间的推移，人们提出了许多分类法和模型来全面、系统地描述人类的人格[12]。五因素模型（Big Five）在心理科学中被广泛接受[13]，包括五种广泛的人格维度（经验开放性、自觉性、合意性、外向性和神经质或正向键，情绪稳定性）。

*这些作者贡献相同

¹代码: <https://github.com/yashmehta/personality-prediction>

为了估算出一个人在这些维度上的得分，人们会使用标准化的自我报告问卷（例如NEO-PI-3, [14]）。虽然基于五因素特质模型的人格评估在人格科学中最为常用，但迈尔斯-布里格斯类型指标是另一个在应用环境中广泛使用的问卷[15]。与大五人格特质分类法（将人格概念化为潜在的特质分数）不同，MBTI以16种类型的形式描述人格，这些类型是由四个维度的二元分配组合而成的：内向与外向、感觉与直觉、思维与感觉、判断与感知[15]。迈尔斯-布里格斯类型指标（MBTI）由于在方法上存在诸多缺陷而受到严厉批评[16]。但是，它仍然是世界上最广泛使用的人格问卷之一[17]。

B. 伦理方面的考虑

在传统的人格评估中，自我报告的问卷被用来估计人们的潜在特质水平。然而，最近在自动预测人格方面的发展表明，数字足迹和行为数据可以用来自动推断人们自我报告的人格特质水平，并具有一定的准确性，而且无需明确同意[18]-[20]。计算性人格评估很有吸引力，因为它有希望消除填写调查问卷的需要。虽然这些模型的性能还不够高，无法根据人们的特质进行精确的区分，但平均而言，预测仍然是“正确的”，可以用于个性化服务和产品以及数字大众说服[6]。在这方面，计算性人格特质评估也引起了对个人隐私和知情同意概念的严重关切[21]。

C. 性格和语言使用

自弗洛伊德早期以来，语言使用的个体差异一直被认为是心理现象的反映[22]。在过去的十年中，许多经验性研究将人们的语言使用与他们的自我报告联系起来。

人格特质得分[20], [23]-[25]。例如, 在外向性方面得分较高的人被发现使用更多的积极情绪词(例如, 伟大、惊人、快乐), 而那些神经质方面得分较高的人被发现更频繁地使用第一人称单数(例如, 我、我的、我)[26]。最初, 这些发现导致了心理语言学词汇分类的发展(例如, 语言学调查和词汇计数)。

- LIWC[26], [27]), 以便对心理学中的语言数据进行系统分析。

D. 理论依据

基于语言的人格评估的最新和最前途的发展之一是使用转移学习技术。在大量的无标签数据上使用无监督学习对语言模型进行预训练, 以获得对语言的基本结构的理解。这些语言模型已经被用来在许多著名的NLP基准中获得最先进的结果, 包括GLUE[28]和SQuAD[29]。在本文中, 我们利用这些语言模型的力量, 进行广泛的实证实验, 并在著名的Essays[26]和Kaggle²人格数据集上取得了最先进的结果。我们还通过分析个别心理语言学特征对预测特定人格特征的影响, 研究了传统手工设计的心理语言学特征的贡献。额外的资源、语法和数据可在我们的开放科学资源库中获得, 以保证透明度和完全可重复性: <https://osf.io/rg5tf/>。

II. 相关的工作

大量的研究都致力于从文本模式中进行自动匹配的人格预测。早期关于作者个性预测的工作集中在从文本中提取基于词库、句法、写作风格和主题的特征, 然后使用相关指标(如皮尔逊相关)来观察这些特征中哪些与个性特征高度相关[30]。使用LIWC的实证结果证明了其在各种实验环境中检测意义的能力, 包括显示注意力集中、情绪化、社会关系、思维方式和个体差异。Mairesse等人[31]开发了一个用于个性预测的文档级特征集, 由84个特征组成。这些文本特征随后被送入传统的机器学习分类器, 如逻辑回归、支持向量机(SVM) [32]、Naïve Bayes等, 以获得最终的人格预测。

最近的工作依赖于深度学习的进展, 并利用预训练的单词嵌入, 如Word2Vec[33]和Glove[34]来建立性能更好的人格预言模型。研究发现, 将常识性知识与心理语言学特征相结合, 使准确率有了明显的提高[35]。这个方向上的另一项工作是Majumder等人提出的著名的一维CNN n-grams模型[36], 该模型取得了最先进的人格预测性能, 直到被

²<https://www.kaggle.com/datasnaek/mbti-type>

Kazemeini等人[37]的基于语言模型的集合方法(BB-SVM)。最近, Mehta等人[38]回顾了基于深度学习的自动人格的最新进展, 重点是有效的多模式人格预测。

III. 方法

A. 数据集

我们在分析中使用了以下公开的人格数据集。

1) 作文。著名的意识流数据集由2468篇学生写的作文组成, 并标注了大五人格特征的二元标签, 这些特征是通过标准化的自我报告问卷发现的[26]。

2) Kaggle

MBTI: 这个数据是通过PersonalityCafe论坛收集的, 因此提供了一个在非正式在线社交环境下互动的多样化选择。这个数据集包含了8675条记录, 其中包括个人在网站上发布的最后50件事情以及他们的MBTI二元性格类型。

表一

文件概念的senticnet值与论文数据集的五大人格特征之间的点-比邻关系水平。*p <.05, **p<.001, 双尾。相关系数是在完整的数据集上计算的。

	O	C	E	A	N
愉悦	0.041*	0.066*	0.032	0.025	-0.075**
注意	0.113**	-0.026	0.013	-0.007	-0.017
敏感度	-0.011	-0.052*	-0.064*	-0.034	-0.022
适应性	-0.045*	0.112**	0.052*	0.081**	-0.020
极性	0.000	0.081**	0.037	0.056*	-0.058*

B. 特征提取

从文本数据中, 我们提取两种不同类型的特征, 即心理语言学特征(以前发现的与个性相关的一组固定特征)和语言模型嵌入。

1) 心理语言学特征。我们从上述文本数据集(P=123)中提取了文献得出的心理语言学特征。此外, 我们还从文本中检索出元信息(称为"可读性"), 并研究这些特征与个性的相关程度。

- Mairesse[31]。一共有84个特征, 由LIWC、医学研究委员会[39]、发音和语气类型的特征组成。这些都是在传统的基于机器学习的人格预测模型中广泛使用的"手工设计"特征。
- SenticNet [40]: 一个由超过100,000个常识性概念组成的词库, 其中标注了愉悦度、注意力、敏感性、适应性和极性等学习值。我们创建了自己的高效的解析器, 以提取最长长度概念的这些数值。这个子特征的最终值是所有从文件中提取的概念的平均值。这些SenticNet特征与五大人格特征之间的相关性见表一。

表二

モンク レーン	論文						Kaggle MBTI				
											P/JA平
	OCE						I/E	N/S	T/F	均	
多数派基线	51.5	50.8	51.7	53.1	50.0	51.4	77.0	85.3	54.1	60.4	69.2
Majumder等人的CNN模型[36]	61.1	56.7	58.1	56.7	57.3	58.0	-	-	-	-	-
神学院[37] [43]	62.1	57.8	59.3	56.5	59.4	59.0	79.0	86.0	74.2	65.4	76.1
心理语言学+MLP	60.4	57.3	56.9	57.0	59.8	58.3	77.6	86.3	72.0	61.9	74.5
BERT-base + SVM	63.2	56.2	57.8	57.4	58.8	58.7	77.0	86.2	73.7	60.5	74.4
BERT-base + MLP	64.6	59.2	60.0	58.8	60.5	60.6	78.3	86.4	74.4	64.4	75.9
所有特征（基础）+MLP	61.1	57.4	57.9	58.6	60.5	59.1	78.4	86.6	75.9	64.4	76.3
BERT-大型+MLP	63.4	58.9	59.2	58.3	58.9	59.7	78.8	86.3	76.1	67.2	77.1

- **NRC情绪词典**[41]。由超过14000个英语单词组成的词库，注释了8种情绪的数值：愤怒、期待、厌恶、恐惧、喜悦、悲伤、惊讶和信任。这个子特征的最终值是一个8维向量，它是文件中存在的所有情绪化词汇的平均值。
- **VAD词典**[42]。一个由超过20,000个英语单词组成的词库，标注了它们的价值、唤醒和支配分数。如上所述，文档的VAD值是所有组成词的平均值。
- **可读性**。一些基于文本的简单表面特征而计算出来的可读性衡量标准。这些衡量标准基本上是基于单词、音节和句子的数量进行的线性回归。

2) **语言模型特征**。我们用多种不同的语言模型（BERT [44]、Albert[45]和Roberta[46]）进行实验，但我们发现它们的性能相似，因此在本文中只报告了BERT-base和BERT-large的结果。我们进行了广泛的实验，以达到语言模型的最佳配置。我们调整的其他配置因素包括语言层使用的模型嵌入（因为研究表明[47]，语言模型的不同层在一个句子中编码不同的语言信息），选择标记嵌入（CLS与平均值），文本预处理的方法，以及从文本的哪个部分选择512个标记（例如，前512个、后512个、前256个和后256个）。

C. 实验配置

由于基于权重初始化和数据顺序的模型性能存在差异，我们报告了外部重采样循环的10倍交叉验证性能的汇总，10个种子的平均值（图1）。在我们的finetuning设置中，我们试验了逻辑回归、SVM和具有50个隐藏单元和 "relu"非线性的多层感知器（MLP）。使用的优化器是Adam[48]，具有二元交叉熵损失函数。我们在表二中报告了表现最好的模型的结果。在finetuning时，我们还试验了更大的MLP结构，但是，它没有带来明显的性能提升。

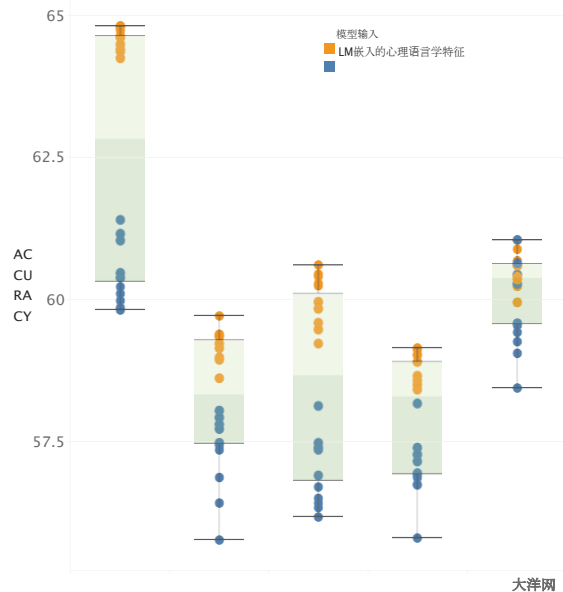


图1.图中说明了Essays数据集上不同初始化种子的性能差异。

IV. 结果

A. 用语言数据预测性格

我们在Essays和Kaggle数据集上取得了最先进的结果。从表二中可以看出，正如预期的那样，基于语言模型的方法在个性预测方面远远超过了传统的封闭词汇。然而，我们发现，使用更大的语言模型并不总是能带来更高的性能。在不同的模型初始化种子的运行中，预测的准确性也有很大差异（图1）。所使用的语言模型的规格配置（如第II-B2节所述）也会导致模型性能的高差异。更多的结果，请参考该项目的开放科学资源库。

B. 心理语言学特征的重要性

可解释的机器学习可以帮助发现算法的偏差和发现无效的模型（例如，使用错误信息进行预测的模型[49]）。在此，我们

每个词的音节是最重要的。此外，图2表明表达焦虑和抑制倾向的词对预测神经很重要。

总之，我们的可解释性机器学习分析的结果与过去人格心理学的关联性结论部分一致，并强调了语言使用对个体差异的表达能力[53]。然而，这些结果也强调，简单的线性关联分析只能很好地描述文本中的语言特征和人格特征之间的关系。

我们对基于语言的人格预测（自下而上和自上而下的特征）的综合和解释方法可能有助于缩小人格科学的计算方法和理论驱动方法之间的差距。

V. 讨论

在本文中，我们提出了一个基于深度学习的新型模型，用于基于语言的人格特质预测。在这个模型中，我们使用传统的心理语言学特征和语言模型嵌入作为特征。此外，我们还分析了个别心理语言学特征对人格特征最终预测的贡献。我们的结果表明，语言建模特征一直在击败传统的心理语言学特征。总的来说，BERT-base + MLP模型在预测大五人格特征方面占优势，BERT-large + MLP在预测MBTI维度方面占优势。在Essays数据集上，我们模型的预测性能比目前最先进的模型高出1.6%，在Kaggle数据集上高出1%。此外，我们的可解释性机器学习分析结果与过去的心理学研究[20], [24], [25]部分一致。

A. 限制与展望

虽然我们的结果显示了对其他使用语言数据的深度学习模型的改进，但也有一些限制影响了本研究。在心理学的人格特征评估中，人格是以连续的分数的衡量的，然而，现有的基准数据集大多只以人工分化的形式提供人格特征的分数的。未来的研究应该以使用提供人格特征连续分数的数据集为目标。如同语言建模中常见的那样，我们在完整的数据集上测试了不同的模型设置（如标记嵌入），以确定性能评估的最佳模型设置。这种方法可能会导致对模型性能的高估。因此，未来的研究应该使用嵌套交叉验证方法评估不同的模型设置[54]。

Stachl等人[49]进一步谈论了研究人员在建立、解释和验证用于人格评估的机器学习模型时面临的主要挑战。另一个很大的缺点是，自我评估与观察到的，以及在线与现场的个性标记之间存在差异。此外，尽管我们的发现与先前的证据相吻合，但结果可能会因分析的社会文化而有所不同。

群体。Lewis[55]进一步探讨了个人行为的多样性。最后，未来的工作将研究SenticNet 6[56]和新的沙漏模型[57]的应用是否能提高人格预测的准确性。

参考文献

- [1] D.J.Ozer and V.Benet-Martínez, "个性和后果性结果的预测", 《心理学年度评论》, 第57卷, 第1期, 第401-421页, 2006年。
- [2] B.W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg, "The Power of Personality: 人格特征、社会经济地位和认知能力对预测重要生活结果的比较效力". *心理科学的观点: 心理科学协会的期刊*, 第2卷, 第4期, 4, pp.313-45, 2007年12月。
- [3] C.J. Soto, "人格特质和生活结果之间的联系的可复制性如何? The Life Outcomes of Personality Replication Project," *Psychological Science*, vol. 30, no.5, pp. 711-727, may 2019.[在线]. Available: <http://journals.sagepub.com/doi/10.1177/0956797619831612>
- [4] Ziegler, D. Bensch, U. Maaß, V. Schult, M. Vogel, and M. Bühner, "大五面作为工作培训绩效的预测因素. The role of specific job demands," *Learning and Individual Differences*, vol. 29, pp.
- [5] H.Yin, Y. Wang, Q. Li, W. Xu, Y. Yu, and T. Zhang, "A network-enhanced prediction method for automobile purchase classification using deep learning," in *PACIS*, 2018, p. 111.
- [6] C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion," *Proceedings of the national academy of sciences*, vol. 114, no.48, pp. 12 714-12 719, 2017.
- [7] H.Schneider, K. Schauer, C. Stachl, and A. Butz, "Your data, your vis:Personalizing personal data visualizations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10515 LNCS. Springer, Cham, sep 2017, pp.374-392.[在线]. Available: http://link.springer.com/10.1007/978-3-319-67687-6_25
- [8] S.T. Völkel, R. Schödel, D. Buschek, C. Stachl, Q. Au, B. Bischl, M.B. "uhner, and H. Hussmann, "在人机交互中利用个性特征进行个性化的机会和挑战: 实现人机交互和心理学共同视角", 在 *个性化的人机交互*。德国Oldenbourg。De Gruyter, 2019.
- [9] C.C. Liem, M. Langer, A. Demetriou, A. M. Hiemstra, A. S. Wicaksana, M.P. Born, and C. J. König, "心理学遇到机器学习. Interdisciplinary perspectives on algorithmic job candidate screening," in *Explainable and Interpretable Models in Computer Vision and Machine Learning*.Springer, 2018, 第197-253页。
- [10] J.Maria Balmaceda, S. Schiaffino, and D. Godoy, "个性特征如何影响在线社交网络中用户之间的交流?" *在线信息评论*, 第38卷, 第1期, 第136-153页, 2014年。
- [11] E.Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74-80, 2017.
- [12] R.J. Gregory, "心理测试的历史", 在 *心理测试*。历史、原则和应用。第7版。Pearson, 2013, ch. 2, pp.
- [13] O.P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative Big Five Trait taxonomy," *Handbook of personality: 理论与研究*, 第114-158页, 2008年。
- [14] R.R. McCrae, P. T. Costa, T. A. Martin, P. T. Costa Jr, and T. A. Martin, "The NEO-PI-3: A More Readable Revised NEO Personality Inventory," *Journal of Personality Assessment*, Vol. 84, no.3, pp.261-270, 2005.
- [15] I.布里格斯-迈尔斯, "类型介绍。了解你的迈尔斯-布里格斯类型指标结果的指南(由LK KIRBY和KD MYERS修订)。
- [16] D.J. Pittenger, "关于Myers-Briggs类型指标的警告性评论". *咨询心理学杂志. 实践与研究*, 第57卷, 第3期, 3, p. 210, 2005.
- [17] C.Luo, "mbti理论在销售人员招聘中的应用", 在 *第19届工业工程和工程管理国际会议上*。Springer, 2013, pp.703-709.

- [18] C. Stachl, Q. Au, R. Schoedel, S. D. Gosling, G. M. Harari, D. Buschek, S. Theres, V. Olkel, T. Schuwerk, M. Oldemeier, T. Ullmann, H. Hussmann, B. Bischl, and M. Bühner, "Predicting personality from patterns of behavior collected with smartphones," *Proceedings of the National Academy of Sciences*, jul 2020.[在线]. Available: www.pnas.org/cgi/doi/10.1073/pnas.1920484117
- [19] W. Youyou, M. Kosinski, and D. Stillwell, "基于计算机的人格判断比人类的判断更准确", *美国国家科学院院刊*, 第112卷, 第. 4, pp. 1036-40, jan 2015.
- [20] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman, "通过社交媒体语言自动评估人格", *人格与社会心理学杂志*, 第108卷, 第6期, 第934页, 2015年.
- [21] S. C. Matz, R. E. Appel, and M. Kosinski, "Privacy in the age of psychological targeting," *Current Opinion in Psychology*, vol. 31, pp. 116-121, feb 2020.
- [22] S. 弗洛伊德, "关于失语症: 一项批判性研究". 1953.
- [23] M. E. Ireland 和 J. W. Pennebaker, "写作中的语言风格匹配。作文、通信和诗歌中的同步性", *人格与社会心理学杂志*, 第99卷, 第3期, 3, p. 549, 2010.
- [24] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "社交媒体语言中的个性、性别和年龄: 开放词汇法。", *PloS one*, vol. 8, no. 9, p. e73791, jan 2013.
- [25] T. Yarkoni, "100,000字的个性。对博客中的人格和用词的大规模分析," *《人格研究》杂志*, 第44卷, 第3期, 3, pp. 363-373, jun 2010.
- [26] J. W. Pennebaker 和 L. A. King, "语言风格。语言使用是一种个体差异", *人格与社会心理学杂志*, 第77卷, 第6期, 第1296页, 1999年。
- [27] J. Pennebaker, M. Francis, and R. Booth, "Linguistic inquiry and word count (liwc): A computerized program," *Mahwah, NJ: Erlbaum*, 2001.
- [28] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Glue: A multi-task benchmark and analysis platform for natural language understanding," *arXiv preprint arXiv:1804.07461*, 2018.
- [29] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000多个用于机器理解文本的问题," *arXiv预印本 arXiv:1606.05250*, 2016.
- [30] E. C. Fieller, H. O. Hartley, and E. S. Pearson, "Test for rank correlation coefficients. i," *Biometrika*, vol. 44, no. 3/4, 第470-481页, 1957年。
- [31] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457- 500, 2007.
- [32] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18-28, 1998.
- [33] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [34] J. Pennington, R. Socher, and C. Manning, "Glove: 全局向量的单词表示," 在 *2014年自然语言处理经验方法会议 (EMNLP)* 上, 2014, 第1532-1543页。
- [35] S. Poria, A. Gelbukh, B. Agarwal, E. Cambria, and N. Howard, "Common sense knowledge based personality recognition from text," in *Mexican International Conference on Artificial Intelligence*. Springer, 2013, pp. 484-496.
- [36] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74-79, 2017.
- [37] A. Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, "Personality trait detection using bagged svm over bert word embedding ensembles," in *Proceedings of the ACL 2020 workshop on Widening NLP*. 计算语言学协会, 2020年。
- [38] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artificial Intelligence Review*, vol. 53, pp. 2313-2339, 2020.
- [39] M. Coltheart, "The mrc psycholinguistic database," *The Quarterly Journal of Experimental Psychology Section A*, Vol. 33, no. 4, pp. 497-505, 1981.
- [40] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *AAAI*, 2018, pp. 1795-1802.
- [41] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," vol. 29, no. 3, pp. 436-465, 2013.
- [42] S. M. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words," in *Proceedings of The Annual Conference of Association for Computational Linguistics (ACL)*, Melbourne, Australia, 2018.
- [43] M. H. Amirhosseini 和 H. Kazemian, "基于 myers-briggs 类型指标 @ 的人格类型预测的机器学习方法," *多模态技术和交互*, 第4卷, 第1期, 第9页, 2020年。
- [44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [45] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.
- [46] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
- [47] B. van Aken, B. Winter, A. Löser, and F. A. Gers, "How does bert answer questions? a layer-wise analysis of transformer representations," in *Proceedings of 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1823-1832.
- [48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv预印本 arXiv:1412.6980*, 2014.
- [49] C. Stachl, F. Pargent, S. Hilbert, G. M. Harari, R. Schoedel, S. Vaid, S. D. Gosling, and M. Bühner, "机器学习时代的人格研究和评估", *《欧洲人格杂志》*, 第per.2257页, 2020年5月。[在线]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/per.2257>
- [50] S. M. Lundberg 和 S.-I. Lee, "解释模型预测的统一方法", 载于 *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. 李, "解释模型预测的统一方法", 载于 *《神经信息处理系统进展》第30期*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765-4774.[在线]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- [51] C. G. DeYoung, "开放性/智力。A dimension of personality reflecting cognitive exploration," in *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences*, ser. M. Mikulincer, P. R. Shaver, M. L. Cooper, and R. J. Larsen, Eds. Washington, DC, US: 美国心理学会, 2015年, 第369-399页。
- [52] W. G. Graziano 和 R. M. Tobin, "Agreeableness," in *Handbook of individual differences in social behavior*, M. R. L. R. Hoyle, Ed. New York, NY, US: Guilford Press, 2009, pp. 46-61.
- [53] R. L. Boyd 和 J. W. Pennebaker, "基于语言的人格: 数字世界中人格的新方法", 第63-68页, 2017年12月。
- [54] B. Bischl, O. Mersmann, H. Trautmann, and C. Weihs, "Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation," *Evolutionary Computation*, vol. 20, no. 2, pp. 249-275, 2012.
- [55] R. 刘易斯, 《当文化碰撞》。尼古拉斯-布雷利出版社, 波士顿, 马萨诸塞州, 2010年。
- [56] E. Cambria, Y. Li, F. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis," in *CIKM*, 2020.
- [57] Y. Susanto, A. Livingstone, B. C. Ng, and E. Cambria, "The hourglass model revisited," *IEEE Intelligent Systems*, vol. 35, no. 5, 2020.

