VendorLink：一种用于识别和关联暗网市场中的供应商迁移者与潜在别名的自然语言处理方法

Vageesh Saxena
法律与技术实验室 马斯特里赫特大学
v.saxena@maastrichtuniversity.nl

Nils Rethmeier语音与语言技术实验室 DFKI, 柏林
nils.rethmeier@dfki.de

Gijs Van Dijck
法律与技术实验室 马斯特里赫特大学
gijs.vandijck@maastrichtuniversity.nl

Gerasimos Spanakis
法律与技术实验室 马斯特里赫特大学
jerry.spanakis@maastrichtuniversity.nl

arXiv:2305.02763v1 [cs.CY] 4 May 2023

摘要

暗网的匿名性使供应商能够通过使用多个供应商别名，或频繁地在市场之间迁移，从而避免被发现。因此，非法市场及其在暗网上的关联关系难以被揭示。为了识别非法市场与其供应商之间的关系，我们提出了VendorLink，这是一种基于自然语言处理的方法，通过分析写作模式，验证、识别并关联七个公开暗网市场上的文本广告（广告）中的唯一供应商账户。与现有研究不同，VendorLink 利用监督式预训练的优势，执行封闭集卖家验证、开放集卖家识别和低资源市场适应任务。通过 VendorLink，我们在（i）Alphabay-Dreams-Silk 数据集中发现了 15 个迁移者和 71 个潜在别名；在（ii）Valhalla-Berlusconi 数据集中发现了 17 个迁移者和 3 个潜在别名；在（iii）Traderoute-Agora数据集中发现了 75 个迁移者和 10 个潜在别名。总的来说，我们的方法能够帮助执法机构通过验证和识别现有及新兴低资源（LR）暗网市场上的迁移供应商及其潜在别名，从而作出更明智的决策。[1]

1 引言

传统搜索引擎收录的表层网络网站仅占整个互联网的 4%（Georgiev，2021年）。其余部分中，90% 为深网（未被收录），6% 为暗网，暗网使用先进的增强匿名性协议（Georgiev，2021年）。前者用于需要匿名性的合法目的，而后者也被用于金融欺诈等非法活动（ENISA，2018年）、儿童剥削（BruggenandBlokland，2021年），以及非法武器交易（Weimann，2016年；Persi Paoli etal.，2017年）、违禁药物和化学品（Kruithof et al.，2016年）。

鉴于暗网的范围、规模和匿名性，执法机构很难发现非法市场之间的联系（Vogt，2017年）。虽然手动检测此类联系是一个耗时且资源密集的过程，但近年来在线爬虫（Fu 等人，2010年；Hayes 等人，2018年）和监控系统（Schäfer 等人，2019年；Godawatte 等人，2019年）的成功，使研究人员和执法机构能够分析（Easttom，2018年；Faizan 和 Khan，2019年；Goodison 等人，2019年；Davies，2020年）并自动识别（Al Nabki 等人，2017年；Ghosh 等人，2017年；Ubbink 等人，2019年；He 等人，2019年）暗网内容。本研究提出了一种供应商验证和识别方法，通过关联供应商、减轻人工工作负担并生成基于相似性的分析，帮助执法机构做出更好的决策。与现有的暗网研究文献（He 等人，2015年；Ekambaranathan，2018年；Tai 等人，2019年；Kumar 等人，2020年；Manolache 等人，2022年）相比，VendorLink（如图 1 所示）着重于以下对暗网市场中供应商验证和识别问题的贡献：

(i) 封闭集供应商验证任务：由于资源有限，执法机构会根据暗网卖家的交易规模和性质，优先开展调查。因此，暗网卖家常常将业务分散到多个市场，以免被发现。同样，有些卖家会在市场被查封后迁往其他市场，重新开展业务（Booij 等人，2021年）。我们将这些

[1]我们的代码实现公开发布于 https://github.com/maastrichtlawtech/VendorLink.git

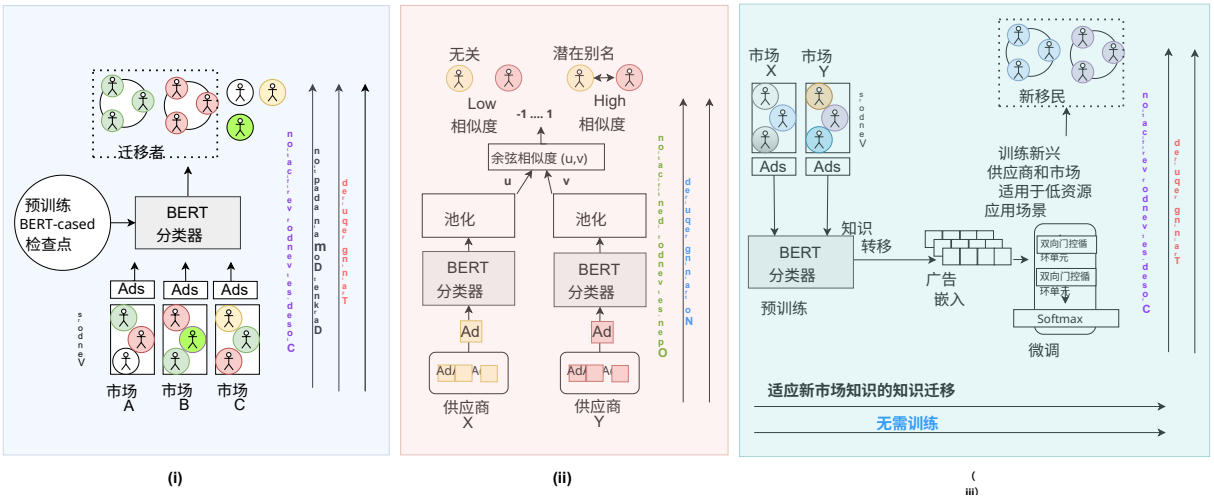


图 1：**（i）封闭集供应商验证任务：**在封闭集环境中使用 BERT-cased 分类器进行分类，以验证现有市场中独特的供应商迁移者的监督式预训练任务，**（ii）开放集卖家识别任务：**在开放集环境中利用已建立的 BERT-cased 分类器所生成风格表征的文本相似度任务，用于验证已知供应商并识别潜在别名，**（iii）低资源市场适应任务：**在封闭集环境中进行知识迁移，以适应新的市场知识，并验证低资源（LR）新兴市场中的迁移者。

迁移供应商简称为迁移者。遗憾的是，这种迁移使执法机构无法准确估算卖家的经营规模。为了协助执法机构，我们在封闭集环境中进行多类分类，以开展监督式预训练（Zhou 等人，2021a），分析文本广告中不同的写作风格，并将迁移卖家归入三个暗网市场中各自对应的卖家账户。此外，研究人员发现，暗网网站与表层网络网站在语言结构方面存在显著差异（Choshen 等人，2019年；Jin 等人，2022年）。由于大多数上下文模型都基于表层网络数据进行训练，监督式预训练步骤使我们的模型能够适应暗网市场领域知识。

(ii) 开放集卖家识别任务：暗网卖家通常会创建别名并组成团体，在多个市场上销售产品，从而在不被执法机构发现的情况下拓展业务。此外，鉴于暗网的广泛性和匿名性，手动关联这些资料并不可行。暗网上每天都会涌现数百个新市场和卖家。尽管现有文献已证明供应商验证任务取得了出色的表现，但在真实到封闭世界的场景中，任何训练好的分类器在推理时都会因遇到来自新兴市场的未知供应商而失效。因此，本研究使用预训练分类器的风格表征来计算文本广告之间的余弦相似度，以验证现有供应商并在开放集环境中识别潜在别名和未知供应商（Zhou 等人，2021a）。

（iii）低资源市场适应任务：尽管研究已证明，在暗网供应商验证任务中取得了令人瞩目的表现（Kumar 等人，2020年；Manolache 等人，2022年），高昂的计算和存储需求给执法机构带来了重大挑战。此外，暗网市场和供应商数量呈指数级增长，每年都有新内容涌现，因此亟需能够从已知数据库中验证现有供应商，同时适应来自新兴供应商和市场的新市场知识的系统。毕竟，并非所有执法机构都有资源从头开始训练计算成本高昂的模型。因此，本实验考察我们的分类器在低资源设置中利用迁移学习（Ruder 等人，2019年）适应新市场知识，并对新兴（即将出现的）供应商和市场执行封闭集供应商验证的能力。最后，我们评估知识迁移对所训练低资源模型的影响，并与零样本（Srivastava 等人，2018年）和基于 Transformer 的基线进行比较。

2 相关研究

供应商验证 - 一项监督式作者归属（作者归属）任务：研究人员此前

已采用各种自然语言处理（Ekambaranathan, 2018年；Tai 等人，2019年；Manolache 等人，2022年）和计算机视觉（Wang 等人，2018年；He 等人，2015年）技术来识别和关联不同暗网市场中的供应商。例如，在其研究中，Zhang 等人（2019年）提出了 uStyle-uID，利用写作风格和摄影风格来识别毒品交易市场中的供应商。类似地，Kumar 等人（2020年）提出利用多视图学习范式和领域特定知识，通过风格计量表示和位置表示来提升跨领域性能。

暗网广告由产品标题和描述、供应商名称、产品价格，以及偶尔出现的一些元数据和图片组成。虽然这些信息大多包含在广告描述中，但手动提取这些特征需要投入大量标注工作。因此，我们着重研究一种端到端方法，该方法只需广告的标题和描述，即可分析写作模式，用于卖家验证与识别。此外，由于我们对暗网广告的文本序列进行多类分类，因此认为我们的方法类似于自然语言处理中的作者归属分析任务。

随着自然语言处理的发展，作者归属领域已涌现大量研究，并证明了基于 TF-IDF 的聚类和分类技术（Agarwal 等人，2019年；Izzet Bozkurt 等人，2007年）、卷积神经网络（Rhodes，2015年；Shrestha 等人，2017年）、循环神经网络（Zhao 等人，2018年；Jafariakinabad 等人，2019年；Gupta 等人，2019年）以及 Transformer 架构（Fabien 等人，2020年；Ordoñez 等人，2020年；Uchendu 等人，2020年a）的成功。此外，研究人员还发现，暗网与表层网络网站的语言结构存在显著差异（Choshen 等人，2019年；Jin 等人，2022年）。因此，探索作者归属任务在暗网语言中的应用至关重要。

卖家识别；一项文本相似度任务：文本相似度技术对作者归属领域的研究人员而言并不新鲜（Sapkota 等人, 2013; Castro Castro 等人, 2015年; Rexha 等人, 2018年; Boenninghoff 等人, 2019年）。然而，随着 Transformer 近期取得成功（Reimers 和 Gurevych, 2019a; Yang 等人, 2019年b; Jiang 等人, 2022年），研究人员如今正探索将语义表征应用于释义检测（Timmer 等人, 2021年; Olney, 2021年; Ko 和 Choi, 2020年）、文本摘要（Miller, 2019年; Cai 等人, 2022年）、语义解析（Geet al., 2019; Ferraroand Suominen, 2020年）、问答（Yang 等人, 2019a; Vold and Conrad, 2021; Louis and Spanakis, 2021）以及作者归属（Fabien 等人, 2020年; Li 等人, 2020年; Custódio and Paraboni, 2021; Uchendu 等人, 2020年b）。

近期关于风格表征的研究（Hay 等人，2020年；Zhu 和 Jurgens，2021年）揭示了在著者验证任务中值得探索的一条有前景的途径。在他们的研究中，Wegmann 等人（2022年）发现，这些表征之所以成功，源于它们能够通过捕捉虚假的内容相关性来表征风格。此外，作者建议在对比设置中采用内容控制，以更好地表征风格，使其与内容的关联更加独立。在本研究中，我们采用类似的方法，将暗网卖家的广告输入到一个针对封闭集供应商验证任务预训练的基于Transformer的分类器中，从而提取风格表征。接下来，我们利用这些表征计算不同供应商广告之间的文本相似度（余弦相似度）。尽管我们认可对风格表征进行内容控制的潜力，本研究仍侧重于在暗网市场上建立一个基线。尽管如此，我们计划在未来的实验中尝试内容控制。

知识适应；一项迁移学习任务：在其研究中，Ruder (2019年) 提出了迁移学习，旨在从源域设置中提取知识，并将其迁移到目标域设置中。此后，许多研究人员探索了迁移学习在跨领域和主题作者归属任务中的成功应用（Sapkota 等人，2014；Barlas和Stamatatos，2021年）。与（Devlin等人，2019年；Horne等人，2020年）中的实验类似，本研究提出利用知识迁移，从新兴的暗网供应商和市场中获取并适应新的市场知识。该迁移方法采用预训练的风格表征，训练一个计算效率较高的双向门控循环单元（BiGRU）分类器，以执行封闭集供应商验证任务。

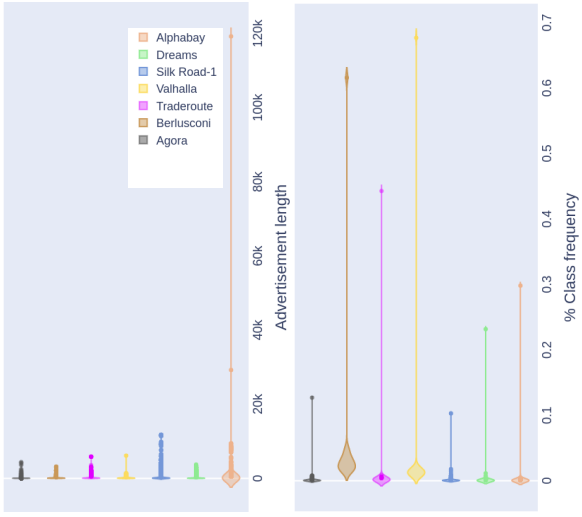
3 数据集

许多研究人员曾使用从活跃暗网市场抓取的数据开展类似实验。然而，由于执法部门现已查封并关闭这些市场，我们既无法复现研究结果，也无法获取其数据。因此——

Task	数据集	Ads.	供应商
基线 / 监督式 预训练	AlphaBay	100,429	1,457
	Dreams	93,586	1,422
	丝绸之路-1	78,681	1,392
	AlphaBay-Dreams-Silk	**272,696**	**4,271**
低资源 监督式 市场适应	Valhalla	2,175	110
	Berlusconi	1,437	84
	Valhalla-Berlusconi	**3,612**	**194**
高资源 监督式 市场适应	Traderoute	19,952	612
	Agora	109,644	3,187
	Traderoute-Agora	**129,586**	**3,799**

表1：每个市场的唯一广告和供应商账户数量。

，为了确保研究可复现并支持未来研究，我们使用来自以下非匿名市场的公开数据集开展分析：Alphabay（Van Wegberg 等人，2018年；Baravalle 和 Lee，2018年；CMU，2017—2018a）、Dreams、Traderoute、Valhalla 和 Berlusconi（Carr 等人，2019年；CMU，2017—2018b）、Agora（Branwen 等人，2015年）以及丝绸之路（Christin，2013；CMU，2012—13）。[2]



Dataset	Min	Median	Max
Alphabay	3	97	119.229k
Dreams	2	67	3,713
Silk Road-1	3	61	11.742k
Berlusconi	5	95	3,161
Traderoute	3	100	5,835
Valhalla	3	88	6,119
Agora	3	40	4,367

(a)

Dataset	Min	Median	Max
Alphabay	199.678u	348.480u	0.2981
Dreams	213.678u	384.623u	0.2316
Silk Road-1	167.515u	385.285u	0.1029
Berlusconi	0.0160	0.0271	0.6158
Traderoute	0.0001	0.0002	0.4427
Valhalla	0.0009	0.0124	0.6767
Agora	182.333u	392.017u	0.1270

(b)

图2：（a）每条广告的词元长度分布（b）每个供应商的广告数量。

预处理： 图 2（a）展示了我们数据集中的所有输入广告的词元数量分布。在小提琴图中，概率分布在中位数附近达到最大值，而表格 2(a)显示，我们所选数据集的中位数介于 40 和 100 之间。因此，为了公平比较其他基线分类器和基于 Transformer 的模型，我们将广告截断为前 512 个词元。另一方面，图 2(b)展示了我们数据集中每个卖家账户的广告数量存在类别不平衡。可以看出，有些市场比其他市场更不平衡。因此，与早期研究侧重于训练好的模型在准确率和微平均 F1 上的表现不同，我们还使用宏平均 F1 评估训练好的模型，该指标对所有类别赋予相同权重。

表 1 展示了每个市场中的唯一广告（输入序列）和卖家账户数量。[3] 首先，我们使用“[SEP]”标记合并广告的标题和描述，以形成输入序列。然后，我们删除数据集中每个供应商的所有重复广告。大多数广告使用英语，少数供应商会使用多种语言。我们认为，数据中的噪声大致反映了各个供应商独特的写作风格。例如，我们发现供应商“CaliforniaDreams420”将药品称为“medi...”，“SAPIOWAX”使用多个“-”表示换行，而“QualityKing”的广告只使用大写字母。因此，任何清理和处理都只会适得其反。不过，由于我们将供应商账户视为分类任务的金标准标签，因此我们将所有供应商名称转换为小写，以尽可能减少数据集中的供应商数量。换句话说，我们假设供应商“agentq”和“AgentQ”是同一实体。该表展示了我们如何划分数据集，用于监督式预训练、低资源和高资源微调步骤。最后，我们将所有广告少于 20 条的供应商归入新的类别标签“其他”，使分类器能够在零样本设置下运行。

虽然我们不对广告数量少于 20 条的供应商进行分类，但通过开放集卖家识别任务，我们能够捕捉这些供应商广告中的相似性。话虽如此，20 这个数字并非随意设定，而是通过实验确定的。我们还在相同设置下进行了实验，分别移除广告数量少于 5、10、20、50 和 100 条的供应商。结果表明，我们的

[2]托管方IMPACT Cyber Trust门户

[3]在本研究中，市场数据指来自单个暗网市场的广告和供应商账户。另一方面，数据集指来自两个或更多市场的合并数据。

模型需要每个供应商至少提供 20 条广告，才能实现最优分类。

4 实验

在开展实验之前，我们先进行合理性检查，通过采用基于 textdistance 的传统文体计量方法（orsinium，2022年）（参见附录 A.1.1），考察暗网广告之间的相似度，以评估是否有必要使用机器学习算法。我们的分析表明，这些传统方法无法识别广告内容差异较大的供应商，说明有必要采用更复杂的特征提取技术。此外，这些方法也有助于我们从后续分析中剔除重复广告。

4.1 封闭集供应商验证任务

架构基线： 为了验证存在于多个市场中的供应商迁移者，我们首先训练多个分类器，分析暗网广告中不同的写作风格，并针对各种基于机器学习和神经网络的算法建立基准。鉴于我们可用的资源，在合并后的 AlphaBay、Dreams 和丝绸之路数据集上训练模型，计算成本高昂且耗时。因此，我们首先通过训练以下模型建立架构基线：(i) 基于 TF-IDF 的统计模型（多项式朴素贝叶斯、逻辑回归器、随机森林、支持向量机和多层感知机网络）；(ii) 使用 fastText 嵌入的双向门控循环单元（Gupta 等人，2019年）、基于字符 n-gram 的卷积神经网络（Shrestha 等人，2017年）；(iii) 预训练的 BERT-base-cased（Devlin等人，2019年）、RoBERTa-base（Liu 等人，2019年）和 DistilBERT-base-cased（Sanh 等人，2019年）序列分类器，以从 Dreams 市场的 93,586 条广告中识别出 1,422 个独立的供应商账户。

方法基线： 我们进一步建立方法基线，以研究不同训练方法对由 AlphaBay、Dreams 和丝绸之路 1 组成的数据集的影响。该数据集包含 272,696 条广告和 3,896 个独立供应商。首先，我们训练 BERT-base-cased 和不区分大小写的分类器，以研究广告中的大写和小写模式对模型性能的影响。其次，我们研究从一个在暗网广告上针对语言任务进行训练的 BERT-cased 模型中进行知识迁移，是否能够提升分类性能。在本研究中，我们将训练得到的语言模型称为 *DarkBERT-LM*，将分类器称为 *DarkBERT*分类器。在另一项研究中，Houlsby 等人（2019年）提出，与其更新预训练模型的权重，不如拼接适配器层并对其进行更新，同时保持预训练模型冻结，这样效率要高得多。因此，我们最终训练了一个带有适配器层的 BERT-cased 分类器（即 *Adapter-BERT*），并计算其性能。

4.2 开放集供应商识别任务

在他们的研究中，（Kornblith 等人，2019年；Phang 等人，2021年）提出了中心化核对齐（*CKA*）作为一种相似度度量，用于可靠地计算以不同初始化训练的网络中表征之间的对应关系。在本研究中，我们计算了训练所得分类器的表征层与一个可用的预训练检查点（未使用暗网数据训练）之间的 CKA 相似度。最后，我们考察相似度最低的层，即训练过程中变化最大且 CKA 相似度较低的层，以从暗网市场的广告中提取具有语义意义的风格表征。[4]

与Reimers 和 Gurevych(2019b) 类似，我们通过计算两个供应商广告中提取的风格表征之间的余弦相似度，来计算这两个供应商之间的相似度。随后，将其中一个供应商指定为父级卖家，并对数据集中的所有其他供应商重复这一过程。然而，余弦距离表示一个线性空间，其中所有维度的权重均相同。因此，Xiao(2018年) 建议关注秩，而非表示两个供应商之间相似度的绝对值。此外，暗网卖家会在多个类别中宣传其产品。对于在多个类别中销售产品的两个供应商 A 和 B，其广告之间的余弦相似度通常会偏低。因此，我们不比较相似贸易类别下的广告（这需要投入标注工作，且不利于本研究），而是提出归一化相似度（sim_{norm}）作为衡量两个供应商广告之间余弦相似度（sim）的指标，并结合其广告中的自相似度（sim_{self}），计算公式如下：

$$sim_{norm} = 2 * \frac{sim(A,B)}{sim_{self}(A,A) + sim_{self}(B,B)}$$

4.3 低资源市场适应任务

为验证来自新兴市场的供应商迁移者，我们在 LR 数据集上开展实验，

[4]附录中的算法-?? ??展示了用于计算我们训练的分类器与可用预训练检查点各层之间 CKA 相似度的伪代码。

即 Valhalla-Berlusconi，该数据集包含 3,612 条广告和 194 个供应商。首先，我们从 LR 数据集中所有广告的预训练分类器的“[CLS]”词元中提取风格表征（第 4.1 节）。接着，参照（Devlin等人，2019年），我们通过使用提取的风格表征初始化双层双向 GRU 分类器，将预训练分类器的知识迁移至该分类器。随后，对双向 GRU 分类器进行微调，使其适应新的市场知识，并验证 LR 数据集中的迁移者。我们将此模型称为迁移 *BiGRU* 模型。知识迁移有助于我们现有的分类器随着新兴供应商和暗网市场数据不断演进。在评估过程中，我们将迁移 BiGRU 的性能与 BERT-base-cased 和双层双向门控循环单元（使用 fastText 嵌入）分类器（又称端到端基线）进行比较；后两者均在 LR 数据集上从头开始训练。最后，对于封闭集供应商验证任务，我们还评估了架构分类器和方法分类器（又称零样本基线）的零样本性能，并将其与迁移 BiGRU 进行比较。

5 结果

5.1 开放集卖家验证任务

架构基线：表2 展示了我们的架构基线在 Dreams 市场上的评估表现。在所有统计模型中，我们发现，使用二元组 TF-IDF 特征的多层感知机（MLP）表现最佳。尽管字符级 CNN 和使用 fastText 嵌入的双向门控循环单元等传统神经网络的表现优于统计模型，但我们注意到，基于 Transformer 的架构在我们的数据集上取得了显著的性能提升。出乎意料的是，RoBERTa-base 模型的表现不及 BERT-base-cased 架构。尽管我们提出利用写作风格来识别不同的供应商，但暗网市场在设计时刻意加入了随机噪声，以阻挠任何自动化系统。此外，由于 RoBERTa 分词器采用“字节级 BPE”，我们认为训练数据不足以让模型学习到这些特征。因此，我们将 Dreams 市场上训练的 BERT-cased 分类器确立为架构基线中的基准分类器。

方法基线：表3 展示了我们方法基线的性能

Data	模型	准确率	微平均 F1	宏平均 F1
Dreams 市场	统计模型			
	多项式朴素贝叶斯	0.0183	0.0144	0.0059
	随机森林	0.0102	0.1093	0.0449
	逻辑回归回归性	0.0045	0.0090	0.0037
	SVM	0.2480	0.3974	0.3703
	神经网络			
	MLP	0.6614	0.6603	0.6594
	字符级卷积神经网络	0.7266	0.7256	0.7248
	BiGRU-Fasttext	0.7374	0.7415	0.7360
	Transformer网络			
	BERT-cased	**0.8978**	**0.8978**	**0.9002**
	区分大小写的DistilBERT	0.8886	0.8885	0.8889
	RoBERTa-base	0.8776	0.8797	0.8736

表 2：Dreams 市场上架构基线的性能。

Data	模型 准确率		微平均 F1	宏平均 F1
AlphaBay-Dreams-Silk 数据集	BERT-uncased	0.8947	0.8939	0.8768
	BERT-cased	**0.9046**	**0.9066**	**0.9013**
	DarkBERT-分类器	0.9000	0.9090	0.9073
	Adapter BERT	0.8398	0.8330	0.8188

表3：方法基线在合并的 Alphabay-Dreams-Silk 数据集上的表现。

在合并后的Alphabay-Dreams-Silk Road-1 测试数据集上进行评估。我们的第一个实验考察写作风格，即大小写模式，对分类任务的影响。如结果所示，BERT-cased 分类器的表现明显优于不区分大小写的分类器（在3,896个类别标签上约高出3%）。我们认为，性能提升源于训练时加入了大写和小写模式。接下来，我们尝试对用于语言任务的DarkBERT-LM在广告上进行持续预训练[5]，使测试困惑度达到 2.07。与BERT-cased 分类器相比，我们观察到微调后的DarkBERT分类器性能略有提升。然而，我们认为如此微小的提升并不值得投入全部训练。此外，DarkBERT-LM的低性能反映了暗网卖家在广告中使用的行话难以预测且嘈杂。我们还推测，使用大规模数据集对模型进行进一步预训练，有助于提升基线模型的性能。最后，Adapter BERT的表现也不及原始区分大小写的BERT分类器。因此，我们将基于封闭集供应商验证任务训练的 BERT-cased架构确立为

[5]针对语言任务进行 BERT预训练需要耗费大量资源。遗憾的是，我们没有足够的资源继续预训练直至收敛，因此只对模型进行了 20 轮训练。

Alphabay-Dreams-Silk Road 暗网数据集的基准分类器。

5.2 开放集卖家识别任务

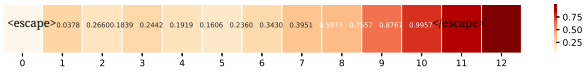


图 3：在使用 Alphabay-Dreams-Silk 数据集训练前后，对比 BERT- cased 方法分类器各层之间的 CKA 距离。

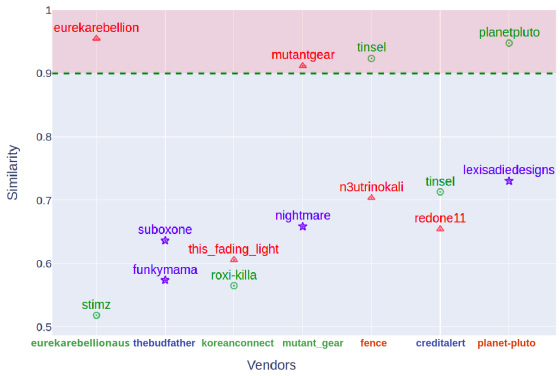


图 4：来自AlphaBay、Dreams 和丝绸之路-1 市场中，父级卖家（横轴）与其潜在别名（纵轴上的散点）之间的散点

图3 显示了我们的 BERT-cased 分类器最后四层表征之间较高的 CKA 距离，即较低的 CKA 相似度。因此，从最后四层的加权和中提取信息，能够为我们在 Alphabay-Dreams-Silk 数据集中的广告提供最有意义的风格表征。如第4.2节所述，我们使用这些风格表征计算卖家广告之间的余弦相似度。图4 在横轴上展示了随机选取的一些母供应商及其两个潜在别名（散点），纵轴则显示了其广告的相似度分数。 [6] 我们的分析表明，“eurekare- bellionaus”和“eurekarebellion”、“mutant_gear”和“mutantgear”、“fence”和“tinsel”，以及“planet- pluto”和“planetpluto”的广告相似度都非常高，因此可能来自同一供应商。相似度越高，两个供应商属于同一实体的可能性就越大。 [7] 为了更清

[6]我们使用Plotly生成散点图，这使我们能够无限缩放查看任何卖家。不过，为了让图表更清晰、更易读，我们只展示选定的卖家及其两个潜在别名。[7]如前所述，暗网卖家经常创建别名来躲避执法机构（LEA）的追查。然而，由于每年都有大量卖家出现在暗网市场上，执法部门很难手动关联这些

	父级卖家	别名／仿冒者	相似度
High（潜在别名）	houseofdank	houseofdank2.0	0.9844
	incorporated	incorporatedv2	0.9769
	castro6969	castro69696	0.9541
	thewizard	thewizzardnl	0.9480
	europills	europills2	0.9467
Low (潜在的模仿者)	topgear	topgear69	0.0367
	dutchpirates	dutchpiratesshop	-0.1015
	whitey	whiteyford	-0.1410
	g3cko	壁虎	-0.2292
	aussieimportpills	aussieimportpillsv2	-0.2560

表4：父级卖家与其潜在别名/仿冒者之间的归一化相似度，按降序排列。

楚地展示，这些供应商在散点图中以红框标出。

通常，供应商别名会使用与卖家名称相似的账号名，以便获得辨识度并垄断其业务。虽然大多数名称相似的账号都可以通过基于字符串的匹配技术检测出来，例如字符串_分组器(Chrisvan denBerg，2021年），但我们的实验揭示了仿冒者的存在：他们的写作风格迥异，广告相似度也很低。例如，我们的实验发现，在 AlphaBay-Dreams-Silk 数据集中，名称相似的供应商别名配对中，广告相似度分数达到 0.7 或以上的仅约占 24%。表4 展示了 10 个此类父级卖家与其可能的别名或仿冒者之间的广告相似度。最后，我们认为我们的实验还可以帮助执法部门发现供应商名称完全不相关的潜在供应商别名配对，例如“fence”和“tinsel”（见图4），但它们的广告却高度相似。

5.3 低资源市场适配任务

为设定零样本基线，我们首先使用成熟的 BERT-cased 架构分类器和方法分类器，在 LR 数据集 Valhalla-Berlusconi 上进行零样本供应商验证。由于新兴的 LR 数据集中出现了新的供应商，我们将所有这些新供应商账户都归入“其他”类别。然而，由于宏平均 F1 是对所有类别标签的 F1 值进行不加权算术平均后计算得出的，而先前

别名指向某个父级卖家。由于缺乏真实情况数据，我们难以评估这些别名是否存在于我们的数据集中。因此，如果没有定性案例研究，我们无法对基于相似性的分析之准确率作出有把握的评价。我们鼓励执法机构不要仅仅依赖这些相似性，而应将其作为开展人工调查的起点。此外，我们强烈不建议执法机构将这些分析作为调查或起诉的证据。本研究的唯一目的是帮助执法机构理解线上暗网市场数据。

已存在的供应商并未出现在新兴 LR 市场中，因此宏平均 F1 的结果并不可靠。因此，我们重点关注零样本基线在微平均 F1 上的表现。尽管未使用 LR 数据进行训练，这些基线仍表现出良好的性能，微平均 F1 分别为 0.7702 和 0.7388。从架构基线到方法基线，宏平均 F1 性能有所下降，这是因为供应商账户数量从 Dreams 市场中的 1,442 个增加到了 Alphabay-Dreams-丝绸之路数据集中的 3,896 个。

模型	层	微平均 F1	宏平均 F1
零样本基线			
架构	-	0.7702	0.2927
方法论	- 0.7388		0.2401
端到端基线			
BERT-cased	-	**0.8987**	**0.8148**
BiGRU-Fasttext	-	0.7797	0.6957
迁移基线			
迁移- 双向门控循环单元	嵌入	0.7653	0.6408
	Last	0.8590	0.7809
	倒数第二层	0.8951	0.7884
	全部12层加权求和	0.8928	0.7837
	加权和最后一层 4 0.8946		**0.8132**

表 5：零样本、端到端和迁移基线在 Valhalla-Berlusconi 数据集上的性能。

GPU	模型	可训练参数	训练时间（小时：分钟）
特斯拉-V100（32 GB）	BERT-cased	110M	0:54
	BiGRU-Fasttext	13M	0:12
	迁移 BiGRU	24M	0:32
GE-MX110（2 GB）	迁移 BiGRU	24M	2:40

表 6：在 LR、Valhalla-Berlusconi 数据集上训练的分类器的计算详情。

此外，我们还从头开始训练了一个 BERT-cased 分类器和一个使用 fastText 嵌入的双向门控循环单元分类器，以适应新兴 LR 数据集带来的新市场知识和供应商。如表 5所示，与零样本基线相比，端到端基线在微平均 F1 和宏平均 F1 得分上均显著提升。最后，遵循（Devlin等人，2019年），我们通过提取 BERT-cased 方法分类器多个层的风格表征，并用其初始化分类层之前的双向门控循环单元，来进行知识迁移。表 5 表明，当使用最后四层的加权和进行初始化时，迁移双向门控循环单元分类器从知识迁移中获益最大，在新兴 LR 数据集上的表现可与端到端 BERT-cased 分类器相媲美。因此，我们将基于封闭集供应商验证任务训练的迁移双向门控循环单元架构确定为 LR、Valhalla-Berlusconi 数据集的基准分类器。

商验证任务作为 LR、Valhalla-Berlusconi 数据集的基准分类器。[8]

最后，表6 通过比较在 LR 数据集上训练的分类器的可训练参数数量和训练时间，展示了训练好的模型在计算方面的表现。可以看出，与 BERT-cased 相比，我们的迁移 BiGRU 分类器更加碳高效（参见附录9），可训练参数减少了 78%，训练时间约为其一半。此外，我们还展示了迁移 BiGRU 在低端显卡 GeForce-MX110 上的训练可行性，该显卡配备 2 GB GPU 内存。因此，我们的低计算量迁移 BiGRU 分类器能够帮助执法部门以极小的性能损失，将我们的方法推广至新兴市场。

6 误差分析

供应商	Pred	文本 A	文本 B
房屋 ofdank	TP	** 1 磅的 酸味 [毒品1] (温室) **	** 1 盎司的 温室 [毒品1] 温室种植 **
房屋 ofdank2.0	TP	** 1 盎司的 [毒品2] 温室 种植 **	** 1 磅的 [毒品2] 温室 种植 **
appleinc	FP	10 x €50 欧元 假冒品 纸币（非常好） 质量)	5 张 100 美元钞票 假冒品 药板 高质量 钞票
加拿大药房	FP	[美国到美国] [毒品3] 80毫克，仅售19.99 每粒仅需美元	[美国到美国] 30药片 [药物3] 100毫克 19.99 美元 Ultra m

表7：BERT-cased 分类器（在 Alphabay-Dreams-Silk Road 数据集上训练）对真正例（TP）和假正例（FP）预测的定性分析。

为了更好地了解训练好的模型的优势和不足，我们对表7中 BERT-cased 分类器（在 Alphabay-Dreams-Silk Road 数据集上训练）的预测结果进行了定性分析。请注意，受篇幅和可见性限制，我们仅展示这些广告的标题。如前两个示例所示，我们训练好的分类器能够识别广告中的许多模式，例如“**”、“[毒品1]”、“[毒品1]”以及“greenhouse gown,.”。

[8]我们还测试了基线模型在新兴高资源（HR）数据集 Traderoute-Agora 上的性能。附录表 10 显示，与端到端 BERT-cased 分类器相比，迁移 BiGRU 模型的表现较差。换言之，在新兴高资源市场上应用知识迁移并不能带来最佳性能。更多详情请参阅附录中的第A.1.3 节。

前两个示例还展示了供应商“houseof-dank”和“houseofdank2.0”的广告有多么相似。这一点也体现在这两家供应商广告的高相似度上（参见4）。最后，下表中的后两个示例展示了假正例的情况。如图所示，网络会混淆“COUNTERFEIT”（假冒）、“质量”、“超级美钞”、来源地和目的地等词汇，以及 [毒品3]，和产品价格。

此外，我们还分析了已训练的 BERT-cased 分类器失效、但迁移 GRU 分类器在知识迁移后成功的案例。表8 展示了供应商广告中的写作风格在 Alphabay-Dreams-Silk Road 与 Valhalla-Berlusconi 数据集之间发生剧烈变化的情况。因此，在零样本设置下，我们的 BERT-cased 分类器无法验证来自 Valhalla-Berlusconi 数据集的供应商。最后，在应用知识迁移并对迁移 BiGRU 模型进行微调后，该模型能够迅速适应这些供应商广告中的新写作风格。

供应商	Alphabay-Dreams- 丝绸之路	Valhalla- Berlusconi
cannacorner	[毒品1] 35克 —|梅赛德斯|	7克 [毒品1] —|兰博|
medicalznl	5 克 哥伦比亚 [DRUGX] 哥伦比亚 [DRUGX] 93% + 免运费	2.5 克 - 90+% 纯度高，未经掺杂
颜色	信用卡可以成为 无需安全码	最后一次更新信用卡 在此文件中。

表 8：迁移 BiGRU 分类器（在 Valhalla-Berlusconi 数据集上训练）对真正例（TP）和假正例（FP）预测结果的定性分析。

7 讨论与未来工作

我们在附录部分讨论补充实验和训练设置的详情A.1 和A.2。

未来，我们计划通过研究内容控制对比学习方法（A.3 和 A.4 附录部分指出的假设和局限性，Wegmann等人，2022年），在现有和新兴的暗网数据集上开展卖家验证与识别。

8 结论

本研究提出了一种基于自然语言处理的供应商验证与识别方法 VendorLink，供执法部门验证、识别并关联供应商迁移者以及现有及新兴暗网市场上的潜在别名。在本研究中，我们首先进行监督式预训练，以适应暗网市场知识，并建立 BERT-cased 分类器，用于在封闭集环境中验证市场之间现有的供应商迁移者。随后，我们从训练后的 BERT-cased 分类器中提取风格表征，以计算开放集环境中供应商广告的文本相似度，并将供应商与其潜在别名关联起来。最后，我们将训练后的 BERT-cased 分类器中的知识迁移应用于低计算资源的双向门控循环单元分类器，以适应新的市场知识，并在新兴 LR市场上执行封闭集供应商验证。通过实验，我们发现：(i) 在 Alphabay-Dreams-Silk 数据集中，有 15 个迁移者和 71 个潜在别名；(ii) 在 Valhalla-Berlusconi 数据集中，有 17 个迁移者和 3 个潜在别名；(iii) 在 Traderoute-Agora数据集中，有 75 个迁移者和 10 个潜在别名；供应商广告与其潜在别名广告之间的余弦相似度均达到 0.8 及以上。

9 广泛影响

本节讨论强制性数据收集规程、伦理考量、潜在风险，以及法律、社会和环境影响。

数据收集规程： 网络爬取相关的伦理问题不适用于我们的研究，因为所用的线上暗网数据是通过与IMPACT Cyber Trust门户（ICC）签署协议备忘录（MoA）后获取的。因此，这些数据可自由获取、依法收集，并被分发用于大规模网络安全分析，使研究人员能够推进网络风险研发和决策支持领域的技术水平。

法律影响： 本研究旨在为执法机构梳理暗网市场上大量可获取的线上数据，并赋予其结构和意义。我们无法预测本研究是否会对执法机构的工作流程产生影响，但本研究旨在识别非法商品供应商之间可能存在的关联，并为执法机构的内部工作流程提供更广泛的信息基础。请注意，我们绝不声称能够提供对任何罪犯提起刑事起诉所需的证据。

伦理与隐私考量： 我们承认，在研究中使用供应商名称可能会被滥用并被识别为一

隐私顾虑。然而，在我们所在机构进行数据隐私影响评估（数据隐私影响评估）后，委员会认定，本研究中使用的供应商名称均为化名，不对应任何个人身份。此外，研究表明，暗网供应商和交易市场的存续时间从几个月到两年不等。（Booij 等人，2021年；Broadhurst 等人，2021年；UNDOC，2020年）。由于我们的数据集中市场广告的时间跨度为 2011年-2018年，任何供应商使用相同用户名仍然存在的可能性都非常低。最后，根据第 6 条“处理的合法性”，通用数据保护条例条款指出，只要任务是出于公共利益而开展，处理个人数据就是合法的。鉴于暗网上非法活动的性质，尽管存在所有潜在风险，我们仍认为本研究有望惠及执法机构并挽救生命。话虽如此，在分析中使用供应商名称有助于提高读者间的透明度和可复现性，我们仍鼓励这些供应商在有任何顾虑时与我们联系。在这种情况下，我们将承担全部责任，立即采取行动并从研究中删除其信息。

社会影响与潜在风险：在他们的研究中，Juola(2020年）阐述了作者身份研究和社交媒体分析在基于目标的推荐系统，以及员工、政治、医疗、性别、人口统计和种族画像方面的阴暗面。尽管我们的方法可能遭到滥用，但鉴于暗网与表层网络网站的语言存在极大差异（Choshen 等人，2019年），我们认为任何人利用我们的研究成果的可能性都很低。尽管如此，我们承认，在犯罪市场之外，通过匹配用户在各个公共平台上的活动，可能会侵犯隐私。例如，心怀恶意的第三方和组织可能会利用我们的研究成果，在公共社交媒体平台上揭穿个人身份。因此，我们提醒读者，在网络安全场景内外运用我们的研究成果开发作者身份技术时，应意识到其中的伦理两面性。

环境影响： 考虑到并非所有执法机构都拥有训练计算开销巨大的架构所需的资源，本研究探讨利用知识迁移来训练计算资源需求较低的模型。因此，我们的迁移 BiGRU 分类器的碳效率为 0.07 $kgCO_2$ eq/kWh 和 2.25 $kgCO_2$eq/kWh，而 BERT-cased 分类器的碳效率为 0.12 $kgCO_2$ eq/kWh 和 4.21 $kgCO_2$eq/kWh，分别对应 Vallhalla-Berlusconi 和 Traderoute-Agoradatasets。这些估算使用 Tesla V100-SXM2-32GB（TDP 为 300W）并借助 MachineLearn-ingImpact calculator进行，相关方法见（Lacoste 等人，2019年）。换言之，本研究表明，将知识迁移应用于从现有市场到新兴市场的迁移，可以帮助执法部门训练计算资源需求较低、性能出色、训练时间更短且碳足迹更小的模型。

10 致谢

本研究得到荷兰教育、文化与科学部数字法律研究部门计划以及 Cora4NLP 项目的支持。该项目由德国联邦教育与研究部（BMBF）资助，资助代码为 01IW20010。最后，本研究得以开展，离不开马斯特里赫特大学托管的数据科学研究基础设施（DSRI）。

参考文献

Lucky Agarwal、Kartik Thakral、Gaurav Bhatt 和 Ankush Mittal。2019年。使用 tf-idf 加权词嵌入进行作者聚类。载于 第 *11* 届信息检索评估论坛论文集，FIRE '19，第 24–29 页，纽约，纽约州，美国。计算机协会。

Mhd Wesam Al Nabki、Eduardo Fidalgo、Enrique Ale- gre 和 Ivan de Paz。2017年。基于网页文本内容对 Tor 网络中的非法活动进行分类。载于欧洲计算语言学协会第*15*届会议论文集：第*1*卷，长篇论文，第35–43页，西班牙瓦伦西亚。计算语言学协会。

安德烈斯·巴拉瓦莱和李欣。2018年。暗网市场：揭开 *AlphaBay* 的神秘面纱：第*19*届国际会议，阿联酋迪拜，*2018*年*11*月*12*日至*15*日，论文集，第二部分，第502–514页。

Georgios Barlas 和 Efstathios Stamatatos。2021年。一种用于跨领域作者身份归属的迁移学习方法。 *Evol. Syst.*，12(3):625–643。

Benedikt Boenninghoff、Steffen Hessler、Dorothea Kolossa 和 Robert M. Nickel。2019年。通过基于注意力的相似度学习进行社交媒体中的可解释作者身份验证。

Tim M. Booij、Thijmen Verburgh、Federico Falconieri 和 Rolf S. van Wegberg。2021 年。发家致富还是继续努力：暗网市场卖家职业生涯轨迹。载于 *2021*年*IEEE* 欧洲安全与隐私研讨会研讨会（*EuroS PW*），第 202–212 页。

Gwern Branwen、NicolasChristin、David Décary-Hétu、Rasmus Munksgaard Andersen、StExo、El Presi- dente、Anonymous、Daryl Lau、Delyan Kratunov So- hhlz、Vince Cakic、Van Buskirk、Whom、Michael McKenna 和 Sigi Goode。2015年。暗网市场档案，2011—2015年。https://www.gwern.net/DNM-archives。访问日期：DATE。

Roderic Broadhurst、MatthewBall、Chuxuan Jiang、Joy Wang 和 Harry Trivedi。2021年。暗网市场查封对阿片类药物供应情况的影响。

Madeleine Bruggen 和 Arjan Blokland。2021年。暗网上的儿童性剥削社区：它们的组织化程度如何？，第259–280页。

XiaoyanCai、SenLiu、Libin Yang、YanLu、Jintao Zhao、Dinggang Shen 和 Tianming Liu。2022年。Covid-sum：一种面向 COVID-19 科学论文的语言学增强型、基于 SciBERT 的摘要生成模型。《生物医学信息学杂志》，127：103999。

Theo Carr、Jun Zhuang、Dwight Sablan、Emma LaRue、Yubao Wu、Mohammad Al Hasan 和 George Mohler。2019年。走入梦境：探索梦境市场。载于*2019*年 *IEEE Big Data* 国际会议（*Big Data*），第 1432–1441 页。Daniel Castro Castro、Yaritza Adame Arcia、María Pelaez Brioso 和 Rafael Muñoz Guillena。2015年。作者身份验证、平均相似度分析。载于国际自然语言处理最新进展会议论文集，第 84–90 页，保加利亚希萨尔。INCOMA Ltd.，保加利亚舒门。Leshem Choshen、Dan Eldad、Daniel Hershcovich、Elior Sulem 和 Omri Abend。2019年。暗网上合法与非法活动的语言。载于计算语言学协会第 *57* 届年会论文集，第 4271–4279 页，意大利佛罗伦萨。计算语言学协会。Chris van den Berg。2021年。字符串_分组器。 [线上；访问日期：2022-09-01]。Nicolas Christin。2013年。丝绸之路之旅：对大型匿名在线市场的测量分析。载于第 *22* 届万维网国际会议论文集，WWW '13，第 213–224 页，美国纽约州纽约。计算机协会。卡内基梅隆大学 CMU。2012—2013年。丝绸之路之旅：未匿名化数据集。卡内基梅隆大学 CMU。2017—2018a。Alphabay 市场：未匿名化数据集，2017—2018年。卡内基梅隆大学 CMU。2017—2018b。梦境、Traderoute、Berlusconi 和 Valhalla 市场，2017—2018年：未匿名化数据集。

José Eleandro Custódio 和 Ivandré Paraboni。2021年。数字文本的堆叠式作者归属判定。《*Expert Systems with Applications*》，176:114866。

Gemma Davies。2020年。透视暗网执法：英国调查权力分析。《*The Journal of Criminal Law*》，84(5):407–426。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 KristinaToutanova。2019年。 BERT：用于语言理解的深度双向变换器预训练。

C Easttom。2018年。 在暗网上开展调查。 《信息战期刊》，17(4):26–37。

Anirudh Ekambaranathan。2018年。 利用文体计量学追踪暗网论坛中的网络罪犯。

ENISA。2018年。 数字空间中的金融欺诈。

Maël Fabien、Esau Villatoro-Tello、Petr Motlicek 和 Shantipriya Parida。2020年。BertAA：用于作者归属判定的 BERT 微调。载于第 *17* 届国际自然语言处理会议（*ICON*）论文集，第 127–137 页，印度理工学院巴特那分校，印度巴特那。印度自然语言处理协会（NLPAI）。Mohd Faizan 和 Raees Ahmad Khan。2019年。探索和分析暗网：一种新的炼金术。Gabriela Ferraro 和 Hanna Suominen。2020年。基于变换器的语义解析。载于澳大拉西亚语言技术协会第 *18* 届年度研讨会论文集，第 121–126 页，线上研讨会。澳大拉西亚语言技术协会。Tianjun Fu、Ahmed Abbasi 和 Hsinchun Chen。2010年。面向暗网论坛的定向爬虫。美国信息科学与技术学会杂志，61(6)：1213–1231。Donglai Ge、Junhui Li 和 Muhua Zhu。2019年。面向 NLPCC-2019 共享任务 2 的基于变换器的语义解析器。载于《自然语言处理与中文计算》，第 772–781 页，Cham。Springer International Publishing。Deyan Georgiev。2021年。2021年互联网中有多少属于暗网？：令人担忧的暗网统计数据。Shalini Ghosh、Ariyam Das、Phil Porras、Vinod Yegneswaran 和 Ashish Gehani。2017年。用于分析暗网生态系统的洋葱网站自动分类。载于第 *23* 届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集，KDD '17，第 1793–1802 页，美国纽约州纽约。计算机协会。

K. Godawatte、M. Raza、M. Murtaza 和 A. Saeed。2019年。暗网及其营销与监控。载于 *2019* 年第 *20* 届并行与分布式计算、应用与技术国际会议（*PDCAT*），第 483–485 页。

Sean E. Goodison、Dulani Woods、Jeremy D. Barnum、Adam R. Kemerer 和 Brian A. Jackson。2019年。识别执法部门在涉及暗网证据的刑事调查中所需的支持。兰德公司，加利福尼亚州圣莫尼卡。

Shriya TP Gupta、Jajati Keshari Sahoo 和 Rajen- dra Kumar Roul。2019年。使用循环神经网络进行作者身份识别。载于*2019*年第*3*届国际信息系统与数据挖掘会议论文集，ICISDM 2019，第133–137页，美国纽约州纽约。计算机协会。

Julien Hay、Bich-Lien Doan、Fabrice Popineau 和 Ouassim Ait Elhara。2020年。写作风格的表征学习。载于第六届嘈杂用户生成文本研讨会论文集（*W-NUT 2020*），第232–243页，线上。计算语言学协会。

Darren R Hayes、Francesco Cappa 和 James Cardon。2018年。《更有效地开展暗网市场调查的框架》。《信息》，9(8):186。

何恺明、张祥雨、任少卿和孙剑。2015年。用于图像识别的深度残差学习。*CoRR*，abs/1512.03385。

Siyu He、Yongzhong He 和 Mingzhe Li。2019年。暗网上非法活动的分类。载于*2019*年第二届国际信息科学与系统会议论文集，ICISS 2019，第73–78页，美国纽约州纽约市。计算机协会。Leo Horne、Matthias Matti、Pouya Pourjafar 和 Zuowen Wang。2020年。GRUBERT：一种基于门控循环单元的方法，用于融合 BERT 隐藏层以进行推特情感分析。载于亚太计算语言学协会分会第一届会议暨自然语言处理第十届国际联合会议论文集：学生研究研讨会，第130–138页，中国苏州。计算语言学协会。

Neil Houlsby、Andrei Giurgiu、Stanislaw Jastrzebski、Bruna Morrone、Quentin de Laroussilhe、Andrea Gesmundo、Mona Attariyan 和 Sylvain Gelly。2019年。面向自然语言处理的参数高效迁移学习。Fereshteh Jafariakinabad、Sansiri Tarnpradab 和 Kien A. Hua。2019年。用于作者归属判定的句法循环神经网络。Ting Jiang、Shaohan Huang、Zihan Zhang、Deqing Wang、Fuzhen Zhuang、Furu Wei、Haizhen Huang、Liangjie Zhang 和 Qi Zhang。2022年。PromptBERT：利用提示改进 BERT 句子嵌入。*arXiv* 预印本 *arXiv:2201.04337*。

Youngjin Jin、Eugene Jang、Yongjae Lee、Seungwon Shin 和 Jin-Woo Chung。2022年。重新认识暗网语言。

Patrick Juola。2020年。作者身份研究与社交媒体分析的阴暗面。《通用计算机科学杂志》，26:156–170。

Bowon Ko 和 Ho-Jin Choi。2020年。采用多任务学习的双向释义转换器。载于*2020*年 *IEEE*大数据与智能计算国际会议 *(BigComp)*，第217–220页。

Narine Kokhlikyan、Vivek Miglani、Miguel Martin、Edward Wang、Bilal Alsallakh、Jonathan Reynolds、Alexander Melnikov、Natalia Kliushkina、Carlos Araya、Siqi Yan 和 Orion Reblitz-Richardson。2020年。Captum：一个统一通用的 PyTorch 模型可解释性库。

Simon Kornblith、Mohammad Norouzi、Honglak Lee 和 GeoffreyHinton。2019年。重新审视神经网络表征的相似度。

Kristy Kruithof、Judith Aldridge、David Décary Hétu、Megan Sim、Elma Dujso 和 Stijn Hoorens。2016年。暗网在非法毒品贸易中的作用。兰德公司，加利福尼亚州圣莫尼卡。

Ramnath Kumar、Shweta Yadav、Raminta Daniulaityte、Francois Lamy、Krishnaprasad Thirunarayan、Usha Lokala 和 Amit Sheth。2020年。Edarkfind：用于 Sybil 账户检测的无监督多视图学习。载于 万维网大会 *2020* 论文集，WWW '20，第1955–1965页，纽约，纽约州，美国。计算机协会。Alexandre Lacoste、Alexandra Luccioni、Victor Schmidt 和 Thomas Dandres。2019年。量化机器学习的碳排放。*CoRR*，abs/1910.09700。

Bohan Li、Hao Zhou、Junxian He、Mingxuan Wang、Yiming Yang 和 Lei Li。2020年。关于预训练语言模型生成的句子嵌入。

Yinhan Liu、Myle Ott、Naman Goyal、Jingfei Du、Mandar Joshi、Danqi Chen、Omer Levy、Mike Lewis、Luke Zettle moyer 和 Veselin Stoyanov。2019年。RoBERTa：一种稳健优化的 BERT 预训练方法。

Antoine Louis 和 Gerasimos Spanakis。2021年。法语法律条文检索数据集。Andrei Manolache、Florin Brad、Antonio Barbalau、Radu Tudor Ionescu 和 Marius Popescu。2022年。Veridark：暗网上用于作者身份验证的大规模基准。Derek Miller。2019年。利用 BERT 对讲座进行抽取式文本摘要。Andrew M. Olney。2021年。学术文本改写：解剖学反向翻译研究

以及利用 Transformer 处理生理学。载于 人工智能教育：第*22*届国际会议，*AIED 2021*，乌得勒支，荷兰，*2021*年*6*月*14*日至*18*日，论文集，第二部分，第279–284页，柏林，海德堡。施普林格出版社。Juanita Ordoñez、Rafael Rivera Soto 和 Barry Y. Chen。2020年。Longformer 能否适用于作者身份验证？CLEF 2020 PAN 任务笔记本。载于*CLEF*。orsinium。2022年。textdistance。 [线上；访问日期：2022年-09-01]。Lin Pan、Chung-Wei Hang、Avirup Sil 和 Saloni Potdar。2021年。通过对比式对抗训练改进文本分类。

Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Te- jani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。2019年。PyTorch：一种命令式风格的高性能深度学习库。载于《神经信息处理系统进展》第*32*卷，第8024–8035页。Curran Associates, Inc.。

F. Pedregosa、G. Varoquaux、A. Gramfort、V. Michel、B. Thirion、O. Grisel、M. Blondel、P. Prettenhofer、R. Weiss、V. Dubourg、J. Vanderplas、A. Passos、D.Cournapeau、M.Brucher、M.Perrot 和 E.Duch- esnay。2011年。Scikit-learn：Python 中的机器学习。《机器学习研究期刊》，12:2825–2830。

Giacomo Persi Paoli、Judith Aldridge、Nathan Ryan 和 Richard Warnes。2017年。幕后：暗网上枪支、爆炸物和弹药的非法交易。兰德公司，加利福尼亚州圣莫尼卡。

Jason Phang、Haokun Liu 和 Samuel R. Bowman。2021年。微调后的Transformer在不同层展现出相似表征的簇。*CoRR*，abs/2109.08406。

Charles Pierse。2021年。TransformersInterpret。Nils Reimers 和 Iryna Gurevych。2019a。Sentence-BERT：使用孪生 BERT 网络的句子嵌入。收录于*2019*年自然语言处理实证方法会议暨第九届国际自然语言处理联合会议（*EMNLP-IJCNLP*）论文集，第3982–3992页，中国香港。计算语言学协会。Nils Reimers 和 Iryna Gurevych。2019b。Sentence-BERT：使用孪生 BERT 网络的句子嵌入。

Andi Rexha、Mark Kröll、Hermann Ziak 和 Roman Kern。2018年。高内容相似度文档的作者身份识别。*Scientometrics*，115(1):223–237。Dylan Rhodes。2015年。使用 CNN 进行作者归属判定。Sebastian Ruder。2019年。面向自然语言处理的神经迁移学习。博士论文，NUI Galway。Sebastian Ruder、Matthew E. Peters、Swabha Swayamdipta 和 Thomas Wolf。2019年。自然语言处理中的迁移学习。收录于计算语言学协会北美分会*2019*年会议论文集：教程，第15–18页，美国明尼苏达州明尼阿波利斯。计算语言学协会。Victor Sanh、Lysandre Debut、Julien Chaumond 和 Thomas Wolf。2019年。DistilBERT：BERT 的精简版本，更小、更快、更便宜、更轻量。*CoRR*，abs/1910.01108。

Upendra Sapkota、Thamar Solorio、Manuel Montes、Steven Bethard 和 Paolo Rosso。2014年。跨主题作者归属判定：主题外数据会有帮助吗？载于*COLING 2014*论文集：第*25*届国际计算语言学会议技术论文，第1228–1237页，爱尔兰都柏林。都柏林城市大学和计算语言学协会。

Upendra Sapkota、Thamar Solorio、Manuel Montes-y Gómez 和 Paolo Rosso。2013年。正交相似关系在作者身份预测中的应用。载于计算语言学与智能文本处理，第463–475页，柏林，海德堡。Springer Berlin Heidelberg。

Matthias Schäfer、Markus Fuchs、Martin Strohmeier、Markus Engel、Marc Liechti 和 Vincent Lenders。2019年。Blackwidow：监控暗网以获取网络安全信息。载于第 *11* 届国际网络冲突会议（*CyCon*），第 900 卷，第 1–21 页。

Prasha Shrestha、Sebastian Sierra、Fabio González、Manuel Montes、Paolo Rosso 和 Thamar Solorio。2017年。用于短文本作者归属判定的卷积神经网络。载于欧洲计算语言学协会第*15*届会议论文集：第*2*卷，短篇论文，第669–674页，西班牙瓦伦西亚。计算语言学协会。Shashank Srivastava、Igor Labutov 和 Tom Mitchell。2018年。从自然语言量化中零样本学习分类器。载于计算语言学协会第*56*届年会论文集（第*1*卷：长篇论文），第306–316页，澳大利亚墨尔本。计算语言学协会。

Xiao Hui Tai、Kyle Soska 和 Nicolas Christin。2019年。暗网市场供应商账户的对抗性匹配。载于第*25*届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，KDD '19，第1871–1880页，美国纽约州纽约市。计算机协会。

Roelien C. Timmer、David Liebowitz、Surya Nepal 和 Salil S. Kanhere。2021年。预训练 Transformer 能否用于检测复杂的敏感句子？——孟山都案例研究。载于*2021*年第三届 *IEEE* 智能系统与应用领域信任、隐私与安全国际会议（*TPS-ISA*），第 90–97 页。

Jeroen Ubbink、Luca Allodia 博士、Alexander Serebrenik 博士和 Decebal Mocanu 博士。2019年。暗网非法武器市场的特征分析。

Adaku Uchendu、Thai Le、Kai Shu 和 Dongwon Lee。2020年a。神经文本生成的作者归属判定。载于*2020*年自然语言处理实证方法会议论文集（*EMNLP*），第 8384–8395 页，线上。计算语言学协会。

Adaku Uchendu、Thai Le、Kai Shu 和 Dongwon Lee。2020年b。神经文本生成的作者归属判定。载于*2020*年自然语言处理实证方法会议论文集（*EMNLP*），第 8384–8395 页，线上。计算语言学协会。

UNDOC。2020年。聚焦暗网上的人口贩运 4——联合国毒品和犯罪问题办公室。

Guido Van Rossum 和 Fred L Drake Jr。1995年。*Python*参考手册。阿姆斯特丹数学与计算机科学研究所，阿姆斯特丹。

Rolf Van Wegberg、Samaneh Tajalizadehkhoob、Kyle Soska、Ugur Akyazi、Carlos Gañán、Bram Klievink、Nicolas Christin 和 Michel Van Eeten。2018年。诱饵与捕获？通过在线匿名市场衡量网络犯罪的商品化。载于第 *27* 届 *USENIX* 安全研讨会论文集，SEC'18，第 1009–1026 页，美国。USENIX协会。

Sophia Dastagir Vogt。2017年。数字地下世界：在现代社会打击暗网犯罪。

Andrew Vold 和 Jack G. Conrad。2021年。使用 *Transformer* 提升法律问题的答案检索，第 245–249 页。纽约州纽约市：计算机协会。

Cindy Wang 和 Michele Banko。2021年。 实用的基于 Transformer 的多语言文本分类。载于《*2021*年北美计算语言学协会年会论文集：人类语言技术：工业论文》，第 121–129 页，线上。计算语言学协会。

Xiangwen Wang、Peng Peng、Chun Wang 和 Gang Wang。2018年。 你就是你的照片：检测暗网市场中供应商的多个身份。载于 《*2018*年亚洲计算机与通信安全会议论文集》，ASIACCS '18，第 431–442 页，纽约，纽约州，美国。计算机协会。

Anna Wegmann、Marijn Schraagen 和 Dong Nguyen。2022年。作者相同，还是主题相同？迈向与内容无关的风格表示。载于 第七届自然语言处理表示学习研讨会论文集，第 249–268 页，爱尔兰都柏林。计算语言学协会。

Gabriel Weimann。2016年。恐怖分子向暗网迁移。《恐怖主义视角》，10(3)：40–44。

Thomas Wolf、Lysandre Debut、Victor Sanh、J ulien Chaumond、Clement Delangue、Anthony Moi、Pierric Cistac、Tim Rault、Rémi Louf、Morgan Funtowicz、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 M. Rush。2020年。Transformer：最先进的自然语言处理技术。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。

Han Xiao. 2018年。 bert-as-service。

WeiYang, YuqingXie, Aileen Lin,Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin. 2019a。使用进行端到端开放域问答。载于 *2019*年北方。计算语言学协会。

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019b。 XLNet：用于语言理解的广义自回归预训练。 *CoRR*，abs/1906.08237。

Xiang Zhang, Junbo Zhao, and Yann LeCun. 2016年。字符级卷积网络文本分类。

Yiming Zhang, Yujie Fan, Wei Song, Shifu Hou, Yan- fang Ye, Xin Li, Liang Zhao, Chuan Shi, Jiabin Wang, and Qi Xiong. 2019年。 你的风格，你的身份：利用写作和摄影风格，通过属性异构信息网络识别暗网市场中的毒品贩运者。载于万维网大会，WWW '19，第3448–3454页，纽约，纽约州，美国。计算机协会。

Chen Zhao、Wei Song、Xianjun Liu、Lizhen Liu 和 Xinlei Zhao。2018年。基于循环神经网络的文章片段作者归属研究。载于*2018*年 *IEEE* 第九届软件工程与服务科学国际会议（*ICSESS*），第 156–159 页。

Da-Wei Zhou、Han-Jia Ye 和 De-Chuan Zhan。2021a。用于开放集识别的占位符学习。

Wenxuan Zhou、FangyuLiu 和 Muhao Chen。2021b。针对预训练 Transformer 的对比式分布外检测。载于*2021*年自然语言处理实证方法会议论文集，第 1100–1111 页，线上及多米尼加共和国蓬塔卡纳。计算语言学协会。

Jian Zhu 和 David Jurgens。2021年。独特却非任意：通过在线语域中的个人语体学习，揭示鲜明而一致的个人风格。载于*2021*年自然语言处理实证方法会议论文集，第 279–297 页，线上及多米尼加共和国蓬塔卡纳。计算语言学协会。

˙Izzet Bozkurt、O. Baghoglu 和 Erkan Uyar。2007年。作者归属判定。*2007*年第 *22* 届国际计算机与信息科学研讨会，第 1–5 页。

A 附录

A.1 补充实验

A.1.1 SanityCheck：文体计量分析方法

作为合理性检查，我们通过研究传统文体计量方法能否识别暗网广告中的写作模式，来探究是否需要机器学习算法。由于语言由字符、词元和句子级元素构成，我们使用 textdistance 中的 Damerau-Levenshtein 距离、Jaccard指数和 Ratcliff-Obershelp 模式识别技术，计算广告之间的字符串、词元和基于序列的相似度。我们将两则卖家广告之间的相似度定义为上述三项指标的平均值。对于拥有多则广告的供应商（例如供应商 A），我们计算其所有广告两两之间的相似度，并取平均值。类似地，对于出现在多个市场的供应商 B，我们取市场 X 中的所有广告，逐一计算它们与市场 Y 中广告的相似度。最后，我们计算供应商 B 在所有市场中广告之间相似度的平均值。

图 5 展示了传统文体计量方法在箱线图上的表现。该图呈现了 AlphaBay-AlphaBay、Dreams-Dreams、Silk Road-Silk Road 市场内部，以及 AlphaBay-Dreams、Dreams-Silk Road 和 AlphaBay-Silk Road 市场之间广告的平均相似度分布及其偏度。可以看出，大多数广告的平均相似度低于 0.20。尽管存在相似度较高的离群值，但只有一位供应商“cyanspore”的相似度分数在 AlphaBay-Dreams 和 Dreams-Silk 数据集中达到 1.0。由于该供应商发布的广告完全相同，我们在后续所有分析中将其移除。

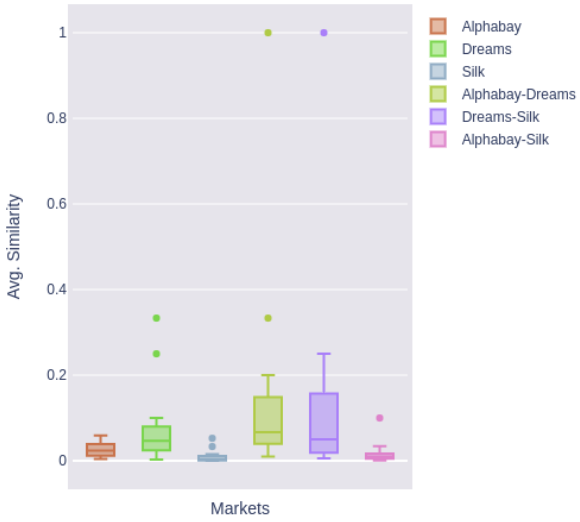


图 5：传统文体计量技术在暗网数据集内部及不同暗网数据集之间供应商广告平均相似度方面的表现。

数据集内部和跨数据集的相似度分数较低，表明传统文体计量框架的能力有限，也说明需要能够在更高层次抽象特征的数学模型。较低的分数也起到了合理性检查的作用，表明暗网上的供应商在不同市场之间以及同一市场内的广告中使用了不同的词汇和风格，这说明需要更深入的特征抽象技术。

A.1.2 供应商验证任务：广告频率和贸易类别对分类器性能的影响

供应商	广告频次	类别	F1 分数
谷歌眼	349	63	0.9340
依替唑仑	186	59	0.9462
有牛奶吗	842	48	0.9893
rinran	437	47	0.9816
钟表机芯	135	41	0.9925
citizen5	35	1	1.0000
corktech	35	1	0.9714
sabinas	26	1	0.9615
mrsupermario	24	1	0.9615
emperium	22	1	1.000

表9：供应商广告频率和贸易类别下的 F1 分数。

Alphabay-Dreams-Silk Road 数据集包含 272,696 条独立广告、3,896 个供应商，以及 322 个不同类别。表9 展示了性能——

我们建立的 BERT-cased 分类器在五个跨最多不同类别销售商品的供应商，以及五个仅在一个类别中销售商品的供应商上的表现。可以看出，分类器的性能基本不受贸易类别数量和广告频率的影响。性能始终稳定，这表明尽管交易涉及不同类别，暗网卖家往往会以相似的方式为其产品刊登广告，使我们的分类器能够区分不同供应商各自独特的写作风格。

A.1.3 应用知识迁移：适配以验证来自高资源（HR）新兴市场的供应商

模型	层	微平均 F1	宏平均 F1
零样本基线			
架构	-	0.7305	0.2173
方法论	-	0.6498	0.1563
端到端基线			
BERT-cased	-	**0.8750**	**0.8700**
BiGRU-Fasttext	-	0.6577	0.6539
迁移基线			
迁移- 双向门控循环单元	嵌入	0.6707	0.6698
	Last	0.7061	0.7153
	倒数第二层	0.6992	0.6911
	全部12层加权求和	0.6698	0.6703
	最后4层加权求和	**0.8065**	**0.8177**

表 10：零样本、端到端和迁移基线在 Traderoute-Agora数据集上的性能。

GPU	模型	可训练参数	训练时间（小时：分钟）
特斯拉-V100 (32 GB)	BERT-cased	112M	32:30
	BiGRU-Fasttext	31M	2:25
	迁移 BiGRU	42M	17:23

表11：在 Traderoute-Agora数据集上训练的分类器的计算详情在 Traderoute-Agora数据集上。

在本研究中，我们展示了我们的方法如何利用计算效率高的网络（迁移 BiGRU），适应并验证从新兴 LR 市场迁移而来的供应商。与第5.3节中呈现的结果类似，表格10和11展示了迁移 BiGRU 分类器在高风险新兴 Traderoute-Agora 数据集上的性能和计算细节。如表所示，尽管可训练参数更少、训练时间更短，我们的迁移 BiGRU 性能仍逊于端到端区分大小写的 BERT 基线。因此，我们并不声称知识迁移方法能够扩展至高风险暗网市场中的新兴供应商。

A.1.4 种子运行

由于资源有限，我们只分析了不同初始化对模型在既有基准测试中性能的影响。如表12所示，标准差、方差和平均性能表明，初始化对模型性能的影响约为 1%。本研究报告的所有性能结果，均基于使用种子值 1111 进行的实验。

Seed 值	BERT-cased Alpha-Dreams-Silk 数据集	BERT-cased Valhalla-Berlusconi 数据集	transfer- 双向门控循环单元
40	0.8969	0.8039	0.7798
100	0.8824	0.8278	0.8005
500	0.8813	0.7837	17:23
1100	0.8861	0.8089	0.8019
1111	**0.9013**	**0.8290**	**0.8132**
Var.	6.46 × 10^{-5}	0.0002	0.0002
Std.	0.0080	0.0167	0.0143
Avg.	**0.8896**	**0.8106**	**0.8035**

表 12：不同初始化对宏平均 F1 性能的影响。

A.1.5 模型解释

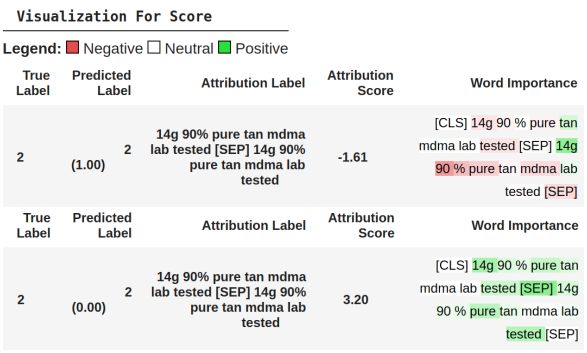


图 6：不同可解释性框架中的模型解释不一致性。

我们还在 BERT 大小写敏感方法分类器上开展了各种基于词归因的可解释性实验，以理解我们的模型决策。图6 展示了由供应商“pck-abml”发布的同一则广告的词归因，这些归因通过 captum (Kokhlikyan 等人，2020年）和 transformers-interpret（Pierse，2021年）框架生成。如图所示，尽管广告内容相同，不同的可解释性框架会生成不同的词归因，导致我们的解释不一致。

另一方面，图7展示了供应商“uridol”的相似广告中基于 Captum 的词归因。可以看出，尽管

广告相似，并且使用相同的框架生成解释，我们得到的词归因却各不相同，导致解释前后不一致。我们认为，通过 [CLS] 标记而非整则广告来计算词归因，可能是造成这些不一致的原因之一。虽然我们尚不清楚解释出现差异背后的原因，但计划在未来对此展开调查。

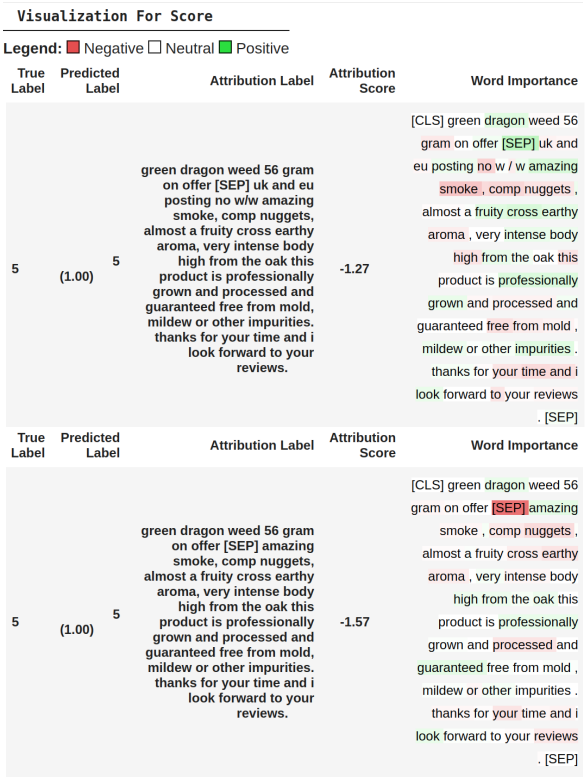


图 7：同一供应商发布的相似广告，其模型解释存在不一致。

A.2 基础设施与日程安排

数据： 我们使用标准的 0.75:0.05:0.20 划分比例，分别将数据划分为训练集、验证集和测试数据集。

训练： 我们使用一块配备 32 GB 内存的 Tesla V100 GPU 训练和评估神经网络。统计分类器的训练与评估则在一台配备 Intel Xeon Processor E5-2698 v4 处理器和 512 GB 随机存取存储器的服务器上进行。最后，我们使用一块配备 2 GB 内存的 GeForce-MX110 显卡，训练用于低资源设置的蒸馏迁移 BiGRU 模型。

我们使用 Adam 优化器，参数为 $\beta 1 = 0.9$, $\beta 2 = 0.999$，L2 权重衰减为 0.01，学习率为 0.001，并在前 500 个训练步骤中进行预热，随后采用线性衰减。

架构与超参数 [9]**：** 我们使用一元语法和二元语法特征以及平衡的类别权重训练所有统计模型。我们针对支持向量机测试线性核和径向基函数（RBF）核，针对随机森林测试 n_个估计器（100 和 1000）、max_深度（5、10 和 20），并使用包含 100 层、每层 100 个神经元的多层感知机。最后，我们采用 5 折嵌套交叉验证方法，在测试数据集上评估统计模型。

我们的卷积神经网络架构处理从暗网广告中提取的字符 n 元语法序列。然后，我们将提取的嵌入依次输入六个带最大池化的卷积层和三个全连接层。受（Zhang 等人，2016年）启发，我们将输入长度设为 1,014，将每个含 768 个神经元的全连接层的丢弃率设为 0.5，前两个卷积层的核大小设为 7，其余层设为 3。最后，我们将滤波器大小设为 32，并以批次大小为 32 训练模型，直到收敛。

循环神经网络架构包含一个双向 GRU 模型，其中有两层，并配有两个全连接层和 fastText 嵌入。我们首先通过 PyTorch 函数对可变长度的输入序列进行打包和填充，然后将其传递给嵌入层。从 Bi-GRU 层生成文本表示后，我们最终将输出传递至 Softmax 层，并对其进行分类。经过一些实验，我们将隐藏单元数量设为 768，丢弃率设为 0.65，批次大小设为 32，并训练模型直至收敛。

最后，我们训练了多个 Transformer 模型（BERT-base-cased、BERT-base-uncased、RoBERTa-base 和 DistilBERT-base-cased），在顶部添加序列分类头，批次大小为 32 [10]，架构基线训练 40 轮（受计算资源限制），方法基线则训练至收敛。我们还训练了一个 BERT-base-uncased 模型，在语言任务上训练 20 轮。所有基于 Transformer 的架构均初始化

[9]所有模型均使用 Python 实现（VanRossum and Drake Jr，1995年），并使用 Sklearn （Pedregosa et al.，2011年）、PyTorch（Paszke et al.，2019年）和 Hugging-face （Wolfet al.，2020年）框架。[10]在不出现内存问题的情况下，我们的资源所允许的最大批次大小。

自预训练模型检查点。

代码仓库：我们鼓励读者通过VendorLink访问并复现我们的代码。

计算详情：表13 和14展示了架构和方法基线中所有训练模型的可训练参数数量及执行时间。

模型（在 Dreams 数据上训练）	可训练参数	训练时间（小时）
多项式朴素贝叶斯	-	53:56
随机森林	-	68:27
逻辑回归	-	79:42
SVM	-	81:08
MLP	-	94:18
字符级卷积神经网络	16M	0:54
GRU-FastText	39M	1:12
BERT	110M	25:14
RoBERTa	125M	23:40
DistilBERT	68M	17:57

表 13：架构基线的可训练参数数量和训练时间。

模型（在 AlphaBay-Dreams -丝绸之路数据集）	可训练参数	训练时间（小时）
BERT-uncased	111M	67:02
BERT-cased	112M	66:58
DarkBERT-LM	108M	156:14
DarkBERT 分类器	112M	49:39
Adapter BERT	4M	51:00

表14：方法基线的可训练参数数量和训练时间。

评估指标：我们使用 Scikit-learn 的分类报告，根据准确率、微平均 F1 和宏平均 F1（通常称为宏平均 F1 和微平均 F1）评估训练好的分类器。我们认为，宏平均 F1 会分别计算每个类别的分数，然后取平均值（对多数类和少数类一视同仁）。鉴于我们的数据集中存在类别不平衡，我们尤其重视训练好的模型在宏平均 F1 指标上的表现。此外，我们还使用损失和困惑度评估 BERT-base 语言模型。最后，我们使用中心化核对齐（CKA）评估并计算方法学基线表示在微调前后之间的对应关系。

微调前后我们的方法学基线表征之间的对应关系。

A.3 假设

本研究在预处理步骤中将卖家名称转换为小写，并假设供应商账户“agentq”和“AgentQ”属于同一实体。然而，实际上，这些实体可能指代两个不同的供应商。此外，我们在多类分类设置下训练分类器，并假设广告仅对应一个单独的供应商账户。然而，我们的实验发现，暗网市场中存在仿冒者。在现实中，多个供应商始终有可能使用相似的卖家名称共存；因此，任何监督式方法都只能产生偏斜的结果。未来，我们计划研究对比学习方法（Pan 等人，2021年；Zhou 等人，2021b；Wegmann等人，2022年），以避免这些假设。

A.4 局限性

架构局限性：本研究建立了一个 BERT-base-cased 分类器，用于验证现有及新兴暗网市场中的迁移供应商。尽管我们承认，使用更大的 BERT 模型并结合滑动窗口可能会提升分类性能，但鉴于我们可用的资源，我们决定不采用该方案。此外，如前文所述，本研究使用的大多数广告为英文，只有少数例外，即供应商使用多种语言。因此，将基于多语言 Transformer 的模型应用于分类任务（Wang 和 Banko，2021年）可以提升我们方法的性能。

无监督和高风险设置：如附录部分A.3所述，我们方法的核心在于金标准标签的可用性。VendorLink 利用监督式预训练步骤来进行知识迁移和文本相似度任务。因此，在缺少这些真实标签或处于无监督设置时，我们的方法会受到重大限制。此外，如A.1.3所述，我们的方法无法很好地扩展到 HR新兴数据集中的供应商迁移者验证。未来，我们计划将 VendorLink 应用于对比学习方法，以学习通用表示并克服这一问题。

多样化广告：在半监督式任务中，我们计算其似然度

通过计算两个供应商的广告之间的相似度，判断两个供应商账户是否属于同一实体。由于本研究的创新之一在于端到端训练方向，我们避免使用人工设计的标签来应用内容控制，以生成与内容无关的风格表示。然而，如第4.2节所述，毒品类别的广告可能与武器类别的广告非常不同。因此，未来我们计划训练另一个分类器，在执行供应商验证任务之前，将暗网广告分类为不同的贸易类别。

XAI 的局限性：可解释人工智能（XAI）对于增进最终用户的信任与理解至关重要。从执法机构的角度来看，缺少这一功能可能会被视为疏忽和不可靠。我们承认，目前的方法尚不具备 XAI 功能；未来，我们计划在实验A.1.5 的基础上，建立一种可靠的方法，以理解并解释模型的决策。