评估TCFD报告：零样本分析在气候相关财务披露中的新应用

Alix Auzepy, [1] **David Lenz,** [1] **Elena T¨onjes,** [1] **Christoph Funk** [2]1吉森尤斯图斯·李比希大学经济与工商管理学院，Licher Strasse 64，35394 吉森，德国[2]吉森尤斯图斯·李比希大学国际发展与环境研究中心（ZEU），森肯贝格街 3 号，35390 吉森，德国
Alix.Auzepy@wi.jlug.de，David.Lenz@wi.jlug.de，Elena.T¨onjes@wi.jlug.de，Christoph.Funks@wi.jlug.de

arXiv:2302.00326v1 [cs.CY] 1 Feb 2023

摘要

我们考察了由正式支持气候相关财务披露工作组（TCFD）建议的银行发布的大量报告中的气候相关披露。为此，我们介绍了零样本文本分类的一种新应用。通过开发一套细粒度TCFD标签，我们表明，零样本分析是一种有用的工具，无需进一步训练模型即可对气候相关披露进行分类。总体而言，我们的研究结果表明，TCFD建议发布后，企业气候相关披露迅速增长。然而，不同建议披露主题的报告程度存在明显差异，这表明部分建议尚未得到充分落实。我们的研究结果为气候相关披露框架的设计提供了重要结论。

1 引言

金融稳定委员会气候相关财务披露工作组（TCFD）于2017年发布的建议，被英国政府称为“最有效的框架之一，可帮助公司分析、理解并最终披露气候相关财务信息”（商业、能源与产业战略部，2021年）。

气候相关财务信息披露工作组建议迄今已获全球4,000 多家公司正式认可。这是一套自愿性披露指南，旨在向投资者及其他重要企业利益相关者提供一致的气候相关信息（Friederich等人，2021）。与其他报告框架（如碳信息披露项目、全球报告倡议组织）相比，其一个重点是披露将气候相关风险纳入风险管理流程、控制结构及业务运营关键方面的信息（Beyene、Ongena 和 Delis，2022）。总体而言，这些建议围绕四大披露类别（治理、战略、风险管理、指标与目标）及其下属的 11 项建议披露内容展开。有关气候相关财务信息披露工作组建议的概述，请参见附录中的图 3。

由于可靠的气候相关风险敞口信息对于作出明智的投资决策和适当定价风险至关重要，越来越多的投资者向公司施压，要求其发布包含全面气候相关财务披露的报告（Krueger、Sautner和Starks，2020年）。此外，包括英国和瑞士在内的多个国家已采取措施，要求其辖区内的大型公司强制进行TCFD报告。从公司的角度看，披露气候相关信息往往表明其意识到气候相关问题并已做好准备；相反，缺乏披露则可能表明公司并未处理此类问题（Sullivan和Gouldson，2012年；Bingler 等人，2022）。Jung、Herbohn和Clarkson（2018年）以及Subramaniam等人（2015年）的研究印证了上述观点：当企业同时披露与气候变化相关的风险信息时，它们更有可能将这些风险纳入整体风险管理。

在此背景下，气候相关披露方面的研究十分稀少，令人惊讶。更重要的是，先前研究对于气候相关财务信息披露工作组指南能否提供高质量且重大的信息，得出了截然不同的结果，原因包括漂绿，以及缺乏标准化、定量数据和透明度。例如，Bin- gler 等人（2022）基于2015年至2020年间818家支持TCFD的企业样本，研究了与四大核心披露支柱相关的气候相关披露。作

版权所有 © 2022，美国人工智能促进协会（www.aaai.org）。保留所有权利。

者得出的结论是，气候相关报告与选择性披露有关，这表明企业披露气候相关信息时，主要侧重于重要性最低的TCFD披露类别，即治理和风险管理。Ding、Liu和Chang（2022）基于TCFD分析了碳排放如何影响自愿气候相关披露。研究结果表明，碳排放水平较高的企业会披露更多气候相关信息。具体而言，他们发现，碳排放会推动战略、风险管理以及指标与目标支柱在类别层级上的披露，但不会推动治理支柱的披露。

迄今为止，这些研究基本未对建议披露事项进行细致分析。具体而言，现有文献关注的是气候相关财务信息披露工作组类别层级上的信息数量，而非其内容（Mehra、Louka 和 Zhang，2022年）。然而，深入分析至关重要，因为披露内容及其重要性会因相关公司的所属行业而异。尤其是，气候相关财务信息披露工作组建议主要面向金融机构。气候变化会影响不同类型资产的信用风险，并带来资产搁浅的风险，因此，金融机构面临气候变化的高风险敞口也可能加剧金融不稳定风险（Beyene 等人，2021）。因此，气候相关财务信息披露工作组建议特别强调披露金融业与碳相关资产的集中度（金融稳定委员会，2015年）。此外，气候相关财务信息披露工作组还发布了针对金融业的补充指南，并鼓励银行提供用于评估气候相关风险对其贷款及其他金融活动影响的指标。

本文尝试从两个重要方面弥补这一空白。首先，我们通过分析2010年至2021年间由支持TCFD的银行发布的3.335份报告样本，为了解有关TCFD建议的披露现状提供了新的见解。我们通过检索所有公开声明支持TCFD、且被列为TCFD网站官方支持者的银行，来确定TCFD支持者。基于前文所述原因，我们明确将研究重点放在银行上。总体而言，我们通过引入零样本文本分类这一新方法，对大量报告中的文本信息进行系统化、自动化提取，从而为自然语言处理在气候相关财务信息披露领域的应用研究作出贡献。此外，我们还设计了一组细粒度标签，使我们能够利用四大核心支柱之外有关TCFD建议的信息。

零样本分类会为每个提取出的文本序列返回标签概率。因此，可以将其解读为相应序列与标签相匹配的概率，换言之，即该文本序列涉及标签所指主题的概率。我们假设，较高的标签概率可作为披露质量的代理指标，因为较高的标签概率表明文本序列的语义很可能与相关标签的语义相匹配。因此，在零样本文本分类的语境下，当文本序列准确、明确地阐述标签所表达的主题时，其对应的标签概率会较高。在一定程度上，较高的标签概率也表明了某个特定主题的披露程度，因为如果文本序列详细讨论了相关主题，其标注对应的概率就会较高。

与其他语言模型相比，零样本方法的一项关键优势是，即使没有接受过相关训练，也能使用标签对句子进行分类。自动识别和分类气候相关内容的算法有一个弱点：这类模型需要大量由人工标注的句子组成的训练集。尤其是，人工标注句子不仅耗时，也容易出错。因此，出于质量和一致性的考虑，需要由训练有素的专业“标注人员”进行标注，这也会增加标注过程的成本。此外，模型分类方案中纳入的类别（或标签类别）越多，为确保每个类别都有数量合理的示例，所需的标注数据就越多；在某些情境下，这可能成为限制因素。由于我们的方法不需要任何带标签的训练数据，因此也不受类别数量的限制，使我们能够对气候相关财务信息披露工作组建议的披露内容进行更细致的分析。此外，气候相关财务信息披露工作组建议非常适合零样本分析，因为它们为我们提供了一个现成的框架和语义（Ding、Liu 和 Chang，2022年）。总之，我们设计了14个细粒度标签，采用相近的语义来捕捉气候相关财务信息披露工作组建议中对银行最核心的方面。

本文得出以下几方面的研究发现。首先，我们考察类别层级的披露水平。具体而言，我们分析样本中的文本序列涉及四个核心TCFD类别之一的平均概率。我们发现，样本期（2010年至2021年）内，与一般标签（即未明确提及气候的治理、战略、风险管理以及指标与目标）相关的平均概率保持稳定；与此同时，同一时期，类别层级上所有气候相关标签（即气候相关治理、气候相关战略、气候相关风险管理以及气候相关指标与目标）的概率均有所上升。具体来说，我们报告称，与“气候相关战略”和“气候相关指标与目标”标签相关的披露增长尤为显著，到2021年，其平均概率分别达到22%和20%。乍看之下，这或许表明，支持TCFD的银行更有可能关注气候相关风险对其业务的战略意义，或制定具有相应目标的气候相关业务战略。

然而，更准确地评估TCFD建议披露的质量及公司相应的落实情况，一种方法或许是不仅考察宽泛的TCFD类别，还考察公司在每个宽泛类别下对建议披露事项的报告情况。据我们所知，迄今为止，关于TCFD的现有文献仅研究了类别层级的报告情况。由于公司报告中的文本序列往往同时对应多个类别，此类分析的结果可能会导致不同的结论。因此，我们在第二步中分析与细粒度标签相关的平均概率，这些标签涵盖了底层的建议披露事项。我们的结果表明，各项披露之间存在显著差异，包括同一TCFD类别内部的差异。在内容最全面、且包含多项针对银行的具体建议披露事项的战略领域，我们发现，与向碳密集型行业（如化石燃料行业）提供融资和投资活动相关的披露，其标签概率较低。同样，支持TCFD的银行似乎较少在披露中明确说明气候相关情景模型的使用情况。

在指标与目标方面，我们发现，与碳足迹和减排目标相关的标签相比，将气候相关绩效指标纳入薪酬政策与较低的标签概率相关。在治理方面，支持TCFD的银行在报告董事会监督气候相关问题的职责以及管理层评估和管理气候相关问题的作用时，披露程度似乎相近。不过，董事会监督的最高值更高，表明部分报告对董事会监督气候相关问题的作用披露质量更高。

本文其余部分安排如下。首先介绍数据，随后介绍研究方法和模型性能评估。结果部分分为两部分。第一部分呈现类别层级的零样本分类结果。第二部分分析细粒度标签的结果，这些标签涵盖了气候相关财务信息披露工作组建议披露的内容。最后一节对结果进行总结和讨论。

2 数据

我们将零样本分类应用于2010年至2021年间人工收集的3,335份报告样本。由于样本中支持TCFD的银行网站首页差异很大，因此无法采用完全自动化的抓取方法。首先，我们根据TCFD网站上提供的行业类别（“银行”“中央银行”和“资本市场”）进行筛选，从TCFD网站获取所有支持TCFD的银行名称。剔除无法识别的银行以及无法在线获取报告的银行后，我们最终得到188家支持TCFD的银行。表1列出了银行层面的数据。

接下来，我们根据样本中银行总部所在地区及其总资产规模对这些银行进行分类。在分析中，总资产超过5,000亿美元的银行被视为“大型”银行，总资产介于500亿至5,000亿美元之间的银行被视为“中型”银行，而总资产低于500亿美元的银行则被视为“小型”银行。有趣的是，我们的样本中近一半是来自亚太地区的银行。欧洲银行占三分之一，北美银行约占样本的10%。样本中的大多数银行是中型银行，总资产介于500亿至5,000亿美元之间。

表1：支持TCFD的银行的规模和地区

地区	大型	中型	小型	∑
亚太地区	15	51	24	90
欧洲	23	26	17	66
拉丁美洲	0	4	3	7
中东和非洲	0	3	2	5
北美	9	8	3	20
∑	47	92	49	188

我们采用 Bingler 等人（2022）的方法，收集2010年至2021年期间可获取的银行报告，以获取气候相关财务信息披露工作组于2017年6月发布建议前后的文本数据。报告按以下类别分类：年度报告、CDP报告、企业治理报告、综合报告、薪酬报告、可持续发展报告和TCFD报告。我们并非仅依赖TCFD报告，因为大多数TCFD支持者不会发布独立TCFD报告来披露气候相关信息，而是倾向于将关键信息纳入年度报告和可持续发展报告。这与TCFD建议一致；建议强调，气候相关披露应纳入“主流（即公开的）年度财务文件”（TCFD 2017）。我们的样本中，大多数报告为年度报告和可持续发展报告。

将报告解析为适合进行零样本分类的原始文本格式后，我们最终得到共3,335份银行报告的样本，如表2所示。由于大规模提取气候披露信息需要投入大量人工工作，零样本方法使我们能够考察相较于以往与TCFD相关的研究规模更大的报告样本。例如，Ding、Liu和Chang（2022）对来自TCFD签署方的140份报告样本应用计算机文本分析；Demaria和Rigot（2020）则分析了2015年至2018年法国CAC 40企业的样本。

表2：样本构成

报告类别	报告数量	平均页数	平均句子数
年度报告	1869	207.98	2711.21
CDP报告	75	63.43	699.79
公司治理报告	148	69.44	1014.25
综合报告	183	163.98	2354.95
薪酬报告	83	36.88	494.42
可持续发展报告	896	81.01	1158.54
TCFD报告	81	36.68	544.37
	$\sum = 3335$	$\bar{x} = 94.20$	$\bar{x} = 1282.50$

PDF解析

我们分析中使用的所有报告均为PDF 格式。将PDF 文档中包含的文本信息转换为适合进一步进行自然语言处理分析的格式，并不像分析存储在CSV或TXT文件中的文本信息那样简单。本文采用版面解析模型，从PDF 文档中检测并提取实际文本。在我们的分析中，我们纳入报告中的实际文本，并有意省略表格和图表中的文本，这不仅提高了数据质量，也节省了计算时间。

我们的解析模型基于 Shen 等人（2021）提出的视觉布局（VILA）组。VILA 能够将文本数据转换为词元组（文本行或文本块），并为这些词元分配版面标签。VILA 有多个变体，称为 H-VILA（视觉布局引导的层次模型）和 I-VILA（注入视觉布局指示器）。经过多次试验，我们选择了基于 layoutLM 模型（Xu 等人，2020年）作为基础模型、在 grotoap2 上训练的 H-VILA 文本块变体，因为它在信息提取和词元化方面表现最佳。输出由提取出的文本词元组及其对应的版面标签组成。根据训练集的不同，版面标签可以是图表、正文内容、摘要和标题。在我们的分析中，我们保留标记为正文内容和摘要的部分。

零样本分类

一项应用广泛且重要的自然语言处理任务是文本分类（Belinkov 和 Glass 2019年）。文本分类通过根据各个文本序列的主题为其分配类别（即所谓的“标签”），来组织和分析海量文本数据。这些文本序列可能由句子、段落或整页内容组成。通常使用简单的神经网络模型，也会使用带有分类末层的更复杂语言模型来执行文本分类。为此，基础模型（例如 BERT 或 BART）会在大量文本数据上进行预训练，以学习语义（Mehra、Louka 和 Zhang 2022年）。随后，这些模型可用于多种自然语言处理任务，并在微调阶段针对特定任务进行训练。为此，可以采用多种预训练语言模型的不同组合（基础模型）和特定任务末层可以结合使用。

微调基础模型的一种方法是使用带标签的训练数据对其进行训练。因此，模型的准确性往往也取决于训练数据的规模和质量。构建句子分类训练集存在若干缺点：首先，人工标注大量文本非常耗时，也需要投入大量人力。其次，当需要更改类别或添加新标签时，必须重复标注过程。第三，即使是人类，有时也难以为句子分配正确的标签，因为某些句子可能会被作不同解读。另一个缺点是，负责标注数据集的人往往存在偏见，因此难以获得具有代表性的训练集（Beltagy、Cohan 和 Lo 2019年）。最后，由于训练数据不能与实际分析中使用的数据相同，寻找足够且合适的训练数据可能是一项挑战。例如，在分析 TCFD 报告时，用于训练模型的已标注报告之后不能再用于分析。此外，由于银行发布的 TCFD 报告数量有限，而且公司进行的 TCFD 报告披露也并不完善，因此高质量训练数据的获取自然受到限制。

鉴于上述缺点，本文采用 Davison（2020）提出的零样本文本分类模型。该模型无需进一步训练，即可根据输入文本序列的语义和标签对文本序列进行分类。我们模型架构的简化示意图见附录中的图4。作为零样本分类的基础模型，我们使用 BART。该模型在约 160 GB 的英文维基百科和 BookCorpus 数据集文本上进行预训练，以“理解”文本的语义（Mehra、Louka 和 Zhang，2022年）。由于 TCFD建议并未包含特定的金融用语，而是采用一般性语义，我们认为这样的基础模型非常适合用于我们的分析。此外，由于样本中的报告也使用英语，我们还受益于该模型曾在规模庞大的英语语料库上进行预训练这一点。与 BERT 等其他模型相比，BART模型（Lewis 等人（2019））不仅采用带有双向编码器（BERT）的序列到序列转换架构，还使用从左到右的自回归解码器（GPT 模型），因此兼具两者的特点。结合零样本文本分类后，BART 作为基础模型展现出优异的性能（Davison，2020年）。

我们用于零样本分类的具体自然语言处理任务是多自然语言推理（MNLI）（Davison（2020））。零样本文本分类会将文本中的句子（词语序列）和标签本身嵌入同一潜在空间。在这样的潜在空间中，可以计算句子与标签之间的距离。标签越接近句子，该标签与句子相匹配的概率就越高。在零样本文本分类的语境中，标签因而被用来衡量文本序列是否涉及一个或多个标注主题。

曾用于衡量一段文本序列涉及一个或多个标注主题的概率。

虽然零样本文本分类会为每个标签生成从0到100%的概率，但 Bingler 等人（2022）使用的 ClimateBERT 模型不会将概率作为输出，而是输出二元结果：其中一个标签被视为真，其他所有标签均视为假。例如，某个与治理相关的段落，其分类输出可能是：标签治理为1，而其他所有标签均为0，即战略为0、风险管理为0，等等。与我们的方法不同，作者通过汇总所有标签的结果，并将其与段落数量进行对比，来衡量企业报告中与气候相关财务信息披露工作组相关内容的比例。

相比之下，我们的文本分类模型将文本序列视为前提，将标签视为假设。模型会对二者进行分词，并利用底层语言模型分别对句子和标签进行嵌入。该模型将二者都输入预训练 MNLI 层。MNLI 末端层是一个简单的全连接神经网络，其输出是一个 logit 分数向量，对应三种结果：“中性”“矛盾”和“蕴含”（Lewis 等人（2019）；Davison（2020））。因此，模型会针对前提检验假设，结果可以分类为蕴含、矛盾或中性。我们舍弃“中性”的分数，并对“矛盾”和“蕴含”的分数应用 softmax 函数，以便将其解释为0到100%范围内的概率。在我们的分析中，所显示的分数仅对应“蕴含”。因此，这些分数可以解释为相应序列与标签相匹配的概率，换言之，即该标签为真的概率。下一步，可以使用微调后的末端层进行零样本分类，而无需进一步训练。

最后，由于TCFD建议彼此关联，而且一个句子可能同时符合多个建议披露事项，因此我们采用多标签方法。因此，我们不会像单标签方法那样，强制零样本文本分类返回总和为1的概率。相反，我们选择一种方法，使模型能够为每个标签分别分配0到1之间的概率（多标签方法），以考虑一个句子可能符合多个标签这一事实。因此，每个句子的所有标签结果之和不必为1。

3 标签评估

零样本文本分类无需使用预先标注的数据进行专门训练。因此，我们无法像通常那样验证模型：将数据集划分为训练集和测试集，再使用测试集验证模型。尽管如此，我们仍可以进行一系列评估测试，以评估零样本文本分类模型的文本序列识别和分类性能。

我们通过手动收集符合气候相关财务信息披露工作组建议的句子，创建了一个新的测试数据集。该数据集包含来自气候相关财务信息披露工作组良好实践手册（CDSB 2019年、2021年）的句子，其中收录了由气候相关财务信息披露工作组选出的最佳实践披露示例。此外，我们还从非银行业公司的气候相关财务信息披露工作组报告中提取句子。最后，我们也使用了 Webersinke等人（2021年）提供的训练资料库中的句子。考虑到一个文本句子可能对应多个标签，我们采用多标签方法进行零样本分类。因此，每个句子对应的所有标签的结果之和不等于一。

我们采用表6所列的细粒度标签。此外，我们还在分类任务中加入“无”标签。该标签用于识别非气候相关文本，即不符合任何细粒度标签的文本序列。它还可以确保零样本方法不会随机分配标签，而是确实根据文本序列的语义进行分配。为了检验模型在“无”标签方面的表现，我们使用上述训练资料库中标注为“无”的句子，以及其他由我们标注为“无”的句子。

使用我们细粒度标签进行零样本文本分类的结果见附录中的图 5。人工标注的句子显示在 x 轴上，而基于气候相关标签的零样本分析结果显示在 y 轴上。结果表示零样本模型返回的平均概率，即模型认为列中的句子涉及行中标签所对应主题的概率。颜色越深，模型判定的可能性越高。在理想的零样本文本分类中，对角线上的单元格颜色应最深。

这些结果表明，我们的模型表现令人满意，能够为相关标签赋予较高的概率。同时，我们也发现，在文本序列识别方面，标签 ST.1.1（气候相关转型风险）、ST.1.3（气候相关问题的重大财务影响）、ST.1.7（银行战略的韧性）、RM.1.1（识别、评估和管理气候相关风险，并将其纳入整体风险管理的流程）以及 RM 1.2（气候相关风险与财务风险之间的关系）的表现不如其他标签。例如，对于标签 RM.1.1 和 RM1.2，不仅我们标注为风险管理文本序列的内容被赋予了这些标签，还有若干句子组也被赋予了这些标签，且概率达到或超过 60%。

这些结果可能有多种原因。首先，一个潜在原因可能是标签的定义过于宽泛，且所涉及的主题过于抽象，不适合进行零样本文本分类。例如，“韧性”或“风险管理流程”等概念可以有许多不同的描述方式，因此可能会部分反映在许多文本序列中。

其次，气候相关财务信息披露工作组建议涉及的主题往往彼此密切相关，因此文本序列通常可以同时符合多个标签。例如，唯一一个自身句子未获得最高概率的标签是 MT.1.1（碳足迹、直接和间接温室气体排放）；零样本分类为 MT.1.3（减排和碳中和目标）赋予了更高概率。由于这些主题密切相关，两组句子都获得较高数值并不令人意外。不过，为了尽可能贴合气候相关财务信息披露工作组建议的语义，我们有意决定不更改这些标签。

“无”标签在所有句子中的概率都很低。这表明模型并非偶然正确地进行标签分类，而是真正识别并纳入了标签的语义。此外，对于几乎所有标签，零样本文本分类为标记为“无”的句子判定的概率都相对较低。标签“气候相关转型风险”（ST.1.1）在“无”标签句子中的概率略高，为 37%。这可能与该标签涵盖多种不同类型的风险有关，例如政治和法律风险、技术风险、市场风险和声誉风险，这些都属于“转型风险”类别。对于一些描述此类风险的句子，零样本分类可能无法直接识别其与气候变化的联系。此外，我们还注意到，RM.1.1 和 RM.1.2 等较为抽象的标签在“无”标签句子中的概率也较高。

除图形分析外，我们还通过考察总体F1分数以及各标签对应的F1分数来评估模型性能，如表3所示。我们基于此前图5（见附录）中使用的测试数据评估模型，重点关注细粒度标签。总体而言，我们的模型微平均F1分数为0.60，宏平均F1分数为0.57。考虑到我们有14个类别，这一结果令人满意。我们还发现各类别的性能差异很大，其中，识别财务问题的重大财务影响（ST.1.3）最为困难（F1分数为0.34），而将气候相关绩效指标纳入薪酬政策（MT.1.2）则最为容易（F1分数为0.78）。这可以解释为，财务重大影响是一个更宽泛的概念，而识别相应句子的任务即使对人类而言也可能很困难。我们还观察到，治理标签的精确率相对较高。相比之下，与转型风险（ST.1.1）、气候相关风险与财务风险之间的关系（RM.1.2）以及减排目标（MT.1.3）有关的标签，精确率相对较低，说明假阳性数量较多。此外，我们的标签召回率总体较高，但识别财务问题的重大财务影响（ST.1.3）以及碳密集型行业的融资与投资（ST.1.5）这两项除外。

总的来说，我们的零样本文本分类似乎并非纯粹随机地分配概率。TCFD建议似乎彼此交织，这也支持了我们采用多标签方法。此外，我们发现，当标签基于界定清晰且定义精确的概念时，零样本分类能够取得更好的结果。

4 结果

按 TCFD 大类划分的气候相关披露

TCFD建议的主要目标是为企业提供指导，帮助其披露一致且对关键利益相关者有决策价值的信息（Bingler 等人，2022）。由于气候相关披露有望减少报告企业与利益相关者之间的信息不对称，并展现公司对气候相关议题的一定认知（Krueger、Sautner和Starks，2020；Jung、Herbohn和Clarkson，2018；Ding、Liu 和 Chang，2022），我们考察类别层级的披露水平。具体而言，我们检验企业报告与四个主要TCFD支柱之一相关的概率。

表4列出了与一般标签“治理”“战略”“风险管理”和“指标与目标”相关的概率，以及分别与我们的标签“气候相关治理”（GO.1）、“气候相关战略”（ST.1）、“气候相关风险管理”（RM.1）和“气候相关指标与目标”（MT.1）相关的概率。我们有意纳入这两类标签，作为比较一般议题披露与更具体的气候相关议题披露的第一步。

根据这些结果，可以得出几项观察：首先，一般议题的平均标签概率高于具体的气候相关议题。例如，在整个样本期内，报告“治理”的概率高于报告“气候相关治理”（GO.1）的概率。其他主要类别也呈现相同结果。这一发现似乎合乎情理，因为我们从不同类型的报告中提取了文本序列，其中包括企业治理报告和年度报告；这些报告讨论了各种治理相关议题，并非只关注气候相关治理议题。因此，零样本文本分类似乎能够很好地区分气候相关与非气候相关的文本数据。

其次，我们发现，在2010年至2021年的样本期内，一般标签（未明确提及气候）的平均概率保持稳定，而2017年后所有气候相关标签的概率均有所上升。我们发现，与其他标签相比，一般标签“战略”的平均概率最高

表3：基于F1分数的性能比较

标签	GO.1.1	GO.1.2	ST.1.1	ST.1.2	ST.1.3	ST.1.4	ST.1.5	ST.1.6	ST.1.7	RM.1.1	RM.1.2	MT.1.1	MT.1.2	MT.1.3
召回率	0.53	0.74	0.66	0.61	0.28	0.52	0.31	0.57	0.51	0.73	0.59	0.58	0.68	0.93
精确率	0.97	0.72	0.29	0.79	0.46	0.54	1.00	0.77	0.50	0.61	0.29	0.90	0.91	0.24
F1分数	0.69	0.73	0.40	0.69	0.34	0.53	0.48	0.65	0.51	0.66	0.39	0.70	0.78	0.39

整体表现分　　数为：微平均F1分数：0.6029，宏平均F1分数：0.5668，加权　F1分数：0.6281。

。该标签的概率随时间保持在40%（或0.39%），而“气候相关战略”标签的概率则呈现出尤为显著的增长，从2010年的12%升至2022年的22%。这表明，在我们的样本中，涉及“气候相关战略”的文本序列在2010年报告中的概率仅为12%，到2021年则升至22%。

此外，我们发现，在样本期内，与其他标签相比，“指标与目标”标签的平均概率位居第二，平均概率介于31%至34%之间。观察“气候相关指标与目标”的平均概率后，我们发现其增长相对温和，从2010年的12%升至2021年的20%；但在2021年，其水平仍高于“气候相关治理”和“气候相关风险管理”。

为了进一步考察类别层级的气候相关披露，我们分析了2017年气候相关财务信息披露工作组建议发布前后的报告趋势。图1展示了样本期内全部四个TCFD类别的这些趋势。具体而言，蓝线表示一般标签的标签概率，橙线则表示明确提及“气候相关”的类别标签的概率，这些标签在表6中缩写为GO.1、ST.1、RM.1和MT.1。我们发现，所有气候相关标签的概率均有所上升，这一变化在2017年前后尤为明显。

我们发现，所有气候相关标签的概率均有所上升，尤其是2017年之后更为明显。对于“气候相关风险管理”（RM.1）和“气候相关指标与目标”（MT.1），我们报告的增幅分别为60%和54%。对于“气候相关战略”（ST.1）标签，我们观察到2017年至2021年间标签概率上升了57%。对于“气候相关治理”（GO.1）标签，我们报告的2017年至2021年间标签概率增幅为46%。总的来说，这些结果表明，在TCFD建议发布后，气候相关问题的报告有所增加；尤其是气候相关风险管理报告的增幅最大，这与Ding、Liu和Chang（2022）的研究结果一致。

表4：每个财年的类别层级标签概率均值

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
治理	0.31	0.31	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32
GO.1	0.11	0.11	0.12	0.11	0.12	0.12	0.12	0.13	0.14	0.15	0.17	0.19
战略	0.40	0.40	0.40	0.40	0.39	0.40	0.39	0.40	0.39	0.40	0.40	0.40
ST.1	0.12	0.12	0.12	0.12	0.12	0.13	0.13	0.14	0.15	0.17	0.20	0.22
风险管　理	0.23	0.23	0.24	0.23	0.24	0.23	0.23	0.23	0.23	0.23	0.24	0.24
RM.1	0.09	0.08	0.09	0.09	0.09	0.09	0.09	0.10	0.11	0.12	0.15	0.16
指标与目标	0.31	0.31	0.31	0.31	0.31	0.32	0.31	0.32	0.32	0.32	0.33	0.34
MT.1	0.12	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.14	0.16	0.18	0.20

该表呈现了基于全样本3,355份报告计算的通用标签和类别层级气候相关标签的标签概率均值（范围为0至1）。

按细粒度TCFD标签划分的气候相关披露

要更准确地评估气候相关披露的质量，一种方法或许是不仅考察各个TCFD宽泛类别的报告数量，还更仔细地审视企业对各类别所建议披露内容的报告。因此，在本分析中，我们不考察总体标签GO.1、ST.1、RM.1和MT.1，而是关注涉及与TCFD建议披露内容相关的具体主题的细分类标签。如前所述，TCFD为金融业提供了额外指引，涵盖银行、保险公司、资产管理者和资产所有者（TCFD 2017）。针对银行的额外指引尤其侧重于战略、风险管理以及指标与目标方面的披露。

我们假设，较高的标签概率可以作为披露质量的替代指标，因为较高的标签概率表明，文本序列的语义很可能与相关标签的语义相匹配。因此，在零样本文本分类的情境下，当文本序列精准、明确地涉及标签所表达的主题时，其对应的标签概率较高。在某种程度上，较高的标签概率也能反映某个特定主题的披露程度，因为如果文本序列详细讨论了该主题，标注结果对应的概率就会更高。

零样本文本分类的结果以汇总平均标签概率表示，如图2所示。作为比较，表5列出了2010年至2021年各年的平均标签概率。结果表明，在每个TCFD类别内，披露程度都存在相当大的差异。在治理领域，支持TCFD的银行似乎报告

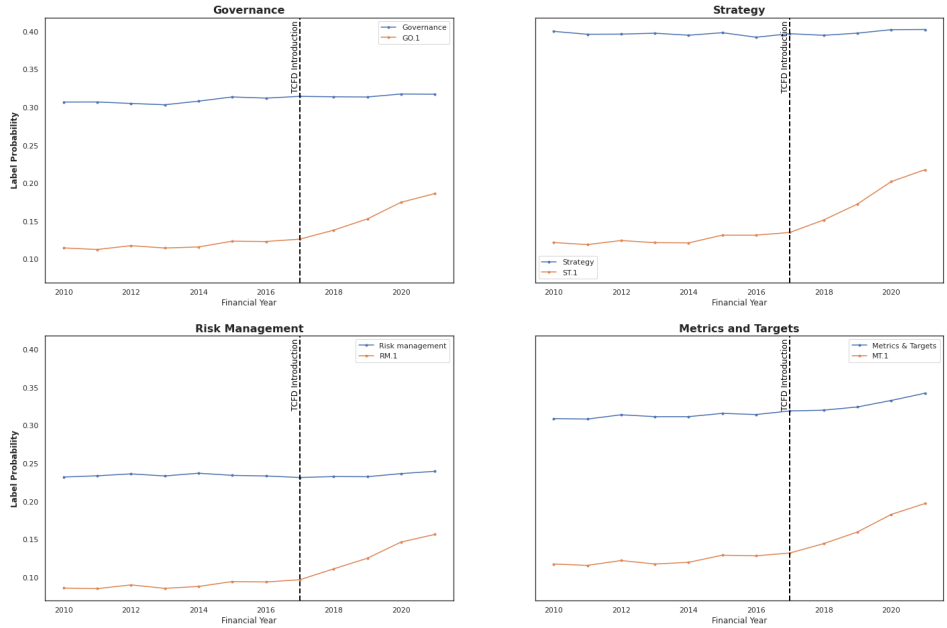

图1：按广义TCFD类别划分的气候相关披露

了相近水平的信息，分别涉及董事会监督气候相关问题的职责（GO.1.1），以及管理层在评估和管理气候相关问题方面的作用（GO.1.2）。两个标签的总体平均概率均较低，为9%；GO.1.1的中位数约为7%，GO.1.2则约为8%。不过，GO.1.1的最大值更高，这表明银行在报告董事会监督气候相关问题的作用方面，报告质量较高。这可能是因为我们的样本涵盖了各种规模的银行，从大型银行到小型银行。尽管所有银行很可能都已将气候相关职责分配给管理层级的职位或委员会，但规模较小的银行可能尚未在董事会层面广泛监督气候相关风险，因此也可能较少报告这一事项的相关信息。

战略领域最为全面，其中包含多项针对银行的具体建议披露事项，这也解释了为什么与其他支柱相比，该领域包含更多标签。转向与战略相关的标签后，我们发现，关于气候相关问题的重大财务影响（ST.1.3）的报告，其总体平均标签概率相对较高，达到35%。然而，正如前文所强调的，零样本模型在该标签上的表现较差（F1分数：0.34），因此这一结果可能被高估。相比之下，对于碳密集型行业（如化石燃料行业）的融资与投资相关披露（ST.1.5），报告质量似乎较低，因为该标签的总体平均概率仅为7%。这意味着，在从报告完整样本中提取并由零样本模型分类的所有文本序列中，某些文本序列符合ST.1.5标签语义的概率仅为7%。表5也对此进行了说明，其中

我们可以看到，ST.1.5的年度平均标签概率在2021年之前一直处于相对较低水平，即使在气候相关财务信息披露工作组建议发布之后也是如此。同样，支持气候相关财务信息披露工作组的银行似乎也较少披露与气候相关物理风险（ST.1.2）以及气候相关情景模型的使用（ST.1.6）有关的信息，因为这两个标签的总体平均概率分别仅为17%和14%。

这些结果可能有多方面的解释。例如，我们样本中的报告可能仅在有限程度上涉及情景分析或气候相关物理风险等问题，因为一些支持TCFD的银行可能尚不具备开展此类分析或识别此类风险所需的工具和专业知识。这与Bingler 等人（2022）关于气候情景工具的研究结果一致：现有的16种气候情景工具中，只有10种能够评估气候相关物理风险。对于碳密集型行业相关披露（ST.1.5）的标签概率较低这一现象，我们观察到，银行往往很少报告这一主题的相关信息。在某些情况下，被赋予该标签且概率较高的文本序列，涉及的是停止为化石燃料资产提供融资和投资的正式承诺，而非对这类投资进行具体量化。相比之下，银行似乎会披露更多有关其对碳相关行业信贷风险敞口的信息。这并不奇怪，因为该项目的定义更宽泛，还包括运输业和公用事业行业。

近期研究表明，自《巴黎协定》签署以来，国际银行对化石燃料企业的融资并未减少，而且国际银行仍在继续提供

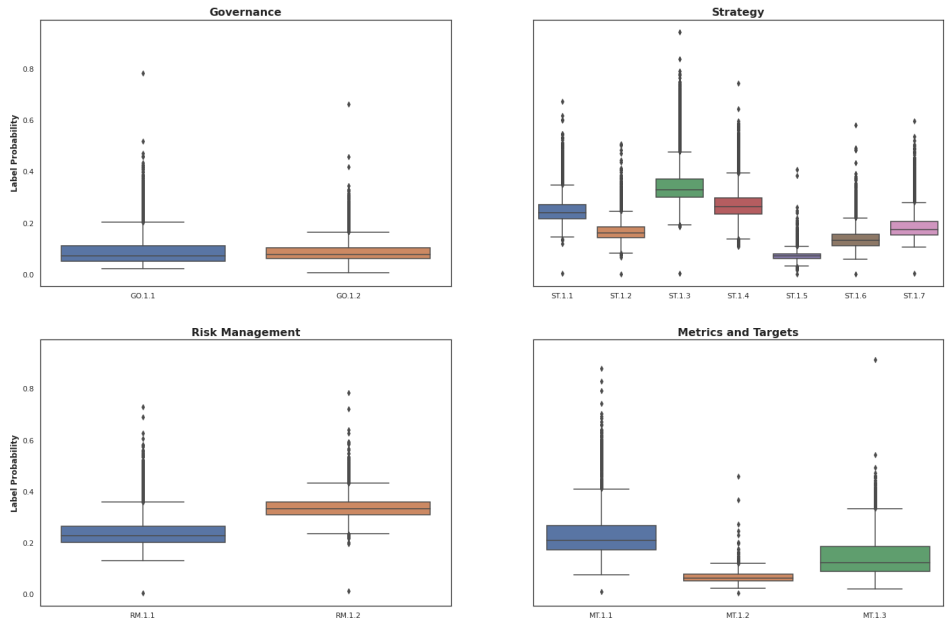

图2：按细粒度标签划分的气候相关披露

融资，无论化石燃料企业是否面临搁浅资产风险（Beyene 等人，2021）。具体而言，Beyene 等人（2021）的分析显示，2017年6月以来加入气候相关财务信息披露工作组的多家银行，在2007—2018年间都是最频繁的化石燃料贷款机构之一。作者的分析尤其指出，摩根大通（自2017年12月起成为气候相关财务信息披露工作组成员）、法国巴黎银行（自2017年6月起成为气候相关财务信息披露工作组成员）和富国银行（自2019年11月起成为气候相关财务信息披露工作组成员）在2007—2018年间都是最频繁的化石燃料贷款机构之一。因此，气候相关财务信息披露工作组成员银行也可能进行某种形式的选择性披露，正如 Bingler 等人（2022）所提出的那样，出于声誉顾虑而省略信息。与此同时，有些银行可能完全不报告某些事项，例如其对化石燃料的风险敞口，因为 TCFD建议措辞十分宽泛，且未提供衡量这种风险敞口的具体方法（Beyene、Ongena 和 Delis，2022）。

在风险管理领域，我们发现，银行报告中涉及识别、评估和管理气候相关风险并将其纳入整体风险管理流程（RM.1.1）的内容，平均少于涉及气候相关风险与信用风险、市场风险、流动性风险和操作风险等财务风险之间关系（RM.1.2）的内容。RM.1.1 的中位数也低于 RM.1.2（分别为 23% 和 33%）。不过，零样本文本分类给出的概率可能略高于实际报告情况，因为零样本模型在 RM.1.2 上的表现较差（F1分数：0.39）。如前所示，有几组句子被判定为符合 RM.1.1 和 RM.1.2 的概率较高，这可能表明标签及相应的建议披露事项定义得过于宽泛。尤其值得注意的是，其他支柱中的若干建议披露事项（例如，管理层在评估和管理气候相关议题方面的作用，GO.1.2）与风险管理问题有关。

在指标与目标方面，我们发现，将气候相关绩效指标纳入薪酬政策（MT.1.2），与其相关的标签概率低于碳足迹相关指标（MT.1.1）和减排目标（MT.1.3）。其总体平均标签概率仅为7%，而MT.1.1为24%，MT.1.3为15%。这一结果符合我们的预期，因为与更具象征性地承诺实现碳中和目标相比，金融机构不太可能将薪酬政策与气候相关绩效指标挂钩；即便它们之后可能无法兑现这些承诺，参见例如Gibson等人（2019）。

表5：按财年划分的细粒度标签的标签概率均值

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
GO.1.1	0,07	0,07	0,08	0,07	0,07	0,08	0,08	0,08	0,09	0,11	0,13	0,14
GO.1.2	0,08	0,07	0,08	0,07	0,08	0,08	0,08	0,08	0,09	0,10	0,11	0,12
ST.1.1	0,24	0,24	0,24	0,23	0,24	0,24	0,24	0,24	0,25	0,26	0,28	0,28
ST.1.2	0,16	0,16	0,16	0,16	0,16	0,16	0,16	0,16	0,17	0,18	0,19	0,19
ST.1.3	0,34	0,34	0,34	0,33	0,33	0,34	0,33	0,33	0,34	0,36	0,40	0,40
ST.1.4	0,25	0,25	0,26	0,25	0,26	0,26	0,27	0,26	0,27	0,28	0,31	0,32
ST.1.5	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,08	0,08
ST.1.6	0,13	0,13	0,13	0,12	0,13	0,13	0,13	0,13	0,14	0,15	0,16	0,16
ST.1.7	0,18	0,17	0,18	0,17	0,17	0,18	0,18	0,18	0,19	0,20	0,22	0,22
RM.1.1	0,22	0,22	0,23	0,22	0,23	0,23	0,23	0,23	0,24	0,25	0,27	0,28
RM.1.2	0,34	0,34	0,34	0,33	0,33	0,33	0,33	0,32	0,33	0,34	0,35	0,35
MT.1.1	0,21	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23	0,25	0,28	0,29
MT.1.2	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07	0,07
MT.1.3	0,12	0,12	0,13	0,12	0,13	0,13	0,14	0,14	0,15	0,16	0,18	0,19

该表展示了基于全部3,355份报告计算的细粒度TCFD标签的标签概率（范围为0至1）的描述性统计。

5 结论

本文采用零样本文本分类这一新颖的计算机化文本分析方法，考察支持气候相关财务信息披露工作组（TCFD）的银行所作的气候相关披露。我们为关于自愿性气候相关企业报告的不断增长的文献作出贡献（例如，Friederich等人（2021）；Bingler 等人（2022）；Ding、Liu和Chang（2022））。将TCFD建议和额外的金融业指导意见转化为细粒度标签后，我们发现气候相关报告存在很大差异，不仅体现在宽泛的TCFD类别上，也体现在基础的建议披露主题上。具体而言，我们发现，银行较少报告气候相关物理风险、对碳密集型行业（如化石燃料行业）的融资和投资、利用气候相关情景模型分析气候相关风险的影响，以及将气候相关绩效指标纳入薪酬政策等主题。这些结果表明，我们样本中的支持TCFD的银行尚未以同等程度落实所有建议。

本研究存在一些局限性，值得审慎考量，但也可能凸显进一步研究的潜力。首先，尽管我们总体上发现，发布TCFD建议后，气候相关报告有所增加，但在本研究的背景下，我们无法判断银行是否确实更加积极地处理了这些问题。换言之，披露更多信息并不一定意味着采取了更多内部措施来应对相关问题。其次，银行也可能因为尚未采取内部措施、因而没有可披露的内容，而未报告某些具体问题。第三，一些银行也可能出于声誉顾虑，故意隐瞒信息并进行选择性披露（即漂绿）。因此，企业也可能试图通过掌控叙事，并在TCFD框架内选择性披露有助于使其行动显得正当的信息，来改善公众形象（Ding、Liu和Chang，2022）。

尽管存在这些局限性，我们的研究也具有重要的实践意义。具体而言，我们的研究凸显了在气候相关披露框架方面提出精确且具体建议的必要性。由于缺乏一致的方法以及对建议披露主题的具体定义，报告范围似乎存在较大差异。就银行业而言，尽管TCFD建议朝着正确的方向迈出了一步，但由于缺乏具体指导，很难准确推导出银行对化石燃料行业的风险敞口，并最终根据所披露的碳相关指标确定银行对搁浅资产的风险敞口。

致谢

Christoph Funk感谢德国学术交流中心（DAAD）利用联邦经济合作部（BMZ）、SDGnexus网络（拨款编号：57526248）项目“exceed——发展合作领域的高校卓越计划”提供的资金支持。Alix Auzepy感谢为SATISFY项目提供的资金支持，该项目属于德国联邦教育与研究部（BMBF）的气候保护与金融（KlimFi）研究基金。

参考文献

Belinkov, Y.; and Glass,J. 2019年. 神经语言处理中的分析方法：综述. 计算语言学协会会刊, 7: 49–72.

Beltagy, I.; Cohan, A.; and Lo, K. 2019年. SciBERT：面向科学文本的预训练上下文化嵌入.*CoRR*, abs/1903.10676. 最后访问于2022年2月28日.

Beyene, W.; De Greiff, K.; Delis, M. D.; and Ongena, S. R. G. 2021. Too-Big-To-Strand：向低碳经济转型过程中的债券对银行融资的替代。*CEPR*讨论论文第*DP16692*号. Beyene, W.; Ongena, S.; and Delis, M. D. 2022. 银行化石燃料风险敞口披露。《欧洲经济——银行、监管与实体部门》, 4(2): 89–105. Bingler, J. A.; Kraus, M.; Leippold, M.; and Webersinke, N. 2022. 空谈与挑拣：ClimateBert对企业气候风险披露有何见解。《金融研究快报》, 47: 102776. CDSB. 2019. *TCFD*良好实践手册。 气候披露标准委员会。CDSB. 2021. *TCFD*良好实践手册（第*2*版）。气候披露标准委员会。Davison, J. 2020. 现代自然语言处理中的零样本学习。最后访问于2022年2月28日。Demaria, S.; and Rigot, S. 2020. 企业环境报告：法国企业是否遵守气候相关财务信息披露工作组的建议? 《商业战略与环境》, 30(1): 721–738. 商业、能源与产业战略部。2021. 咨询回应：上市公司、大型私营公司及有限责任合伙企业的强制性气候相关财务披露。Ding, D.; Liu, B.; and Chang, M. 2022. 碳排放与符合TCFD要求的气候相关信息披露。《商业伦理学杂志》. Friederich, D.; Kaack, L. H.; Luccioni, A.; and Steffen, B. 2021. 自动识别企业年度报告中的气候风险披露。金融稳定委员会。2015. 气候相关风险披露工作组提案。金融稳定委员会。

Gibson, R.; Glossner, S.; Krueger, P.; Matos, P.; and Steffen, T. 2019. 全球负责任的机构投资。*SSRN*电子期刊. Jung, J.; Herbohn, K.; and Clarkson, P. 2018. 碳风险、碳风险意识与债务成本。《商业伦理学杂志》, (150): 1151–1171. Krueger, P.; Sautner, Z.; and Starks, L. T. 2020. 气候风险对机构投资者的重要性。《金融研究评论》, 33(3): 1067–1111. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mo-hamed, A.; Levy, O.; Stoyanov, V.; and Zettlemoyer, L. 2019. BART：面向自然语言生成、翻译与理解的序列到序列去噪预训练。*CoRR*, abs/1910.13461. 最后访问于2022年2月28日。Mehra, S.; Louka, R.; and Zhang, Y. 2022. ESGBERT：用于公司环境、社会和治理实践相关分类任务的语言模型。Shen, Z.; Lo, K.; Wang, L. L.; Kuehl, B.; Weld, D. S.; and Downey, D. 2021. VILA：利用视觉布局组改进科学类PDF中的结构化内容提取。最后访问于2022年2月28日。Subramaniam, N.; Wahyuni, D.; Cooper, B. J.; Leung, P.; and Wines, G. 2015. 将碳风险与机遇纳入企业风险管理系统：来自澳大利亚企业的证据。《清洁生产杂志》, 96: 407–417. Sullivan, R.; and Gouldson, A. 2012. 自愿碳报告是否满足投资者的需求？《清洁生产杂志》, 36: 60–67. 气候相关财务信息披露工作组。2017. 最终报告：气候相关财务信息披露工作组的建议。气候相关财务信息披露工作组。Webersinke, N.; Kraus, M.; Bingler, J. A.; and Leippold, M. 2021. ClimateBert：一种面向气候相关文本的预训练语言模型。*CoRR*, abs/2110.12010. Xu, Y.; Li, M.; Cui, L.; Huang, S.; Wei, F.; and Zhou, M. 2020. Layoutlm：面向文档图像理解的文本与布局预训练。载于第*26*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集, 1192–1200.

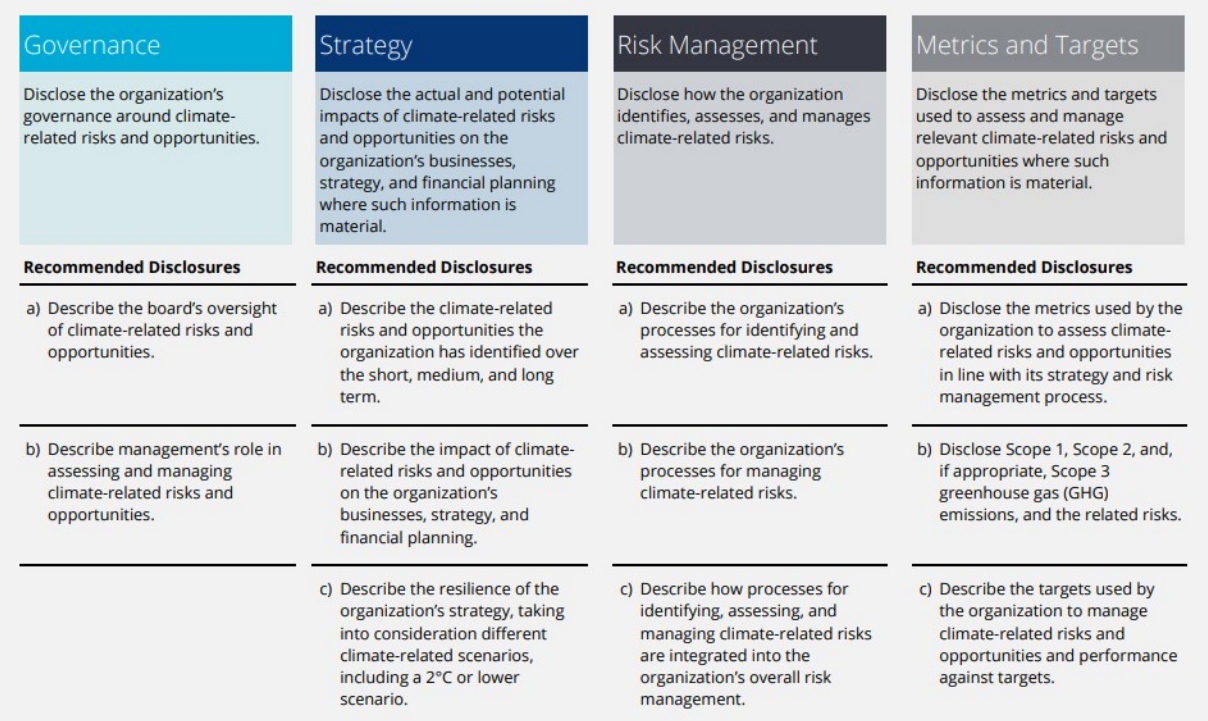

图3：TCFD披露类别及其下属建议披露事项。（TCFD 2017）

TCFD类别	标签名称	标签说明
治理	GO.1.	气候相关治理
	GO.1.1	董事会负责监督气候相关问题
	GO.1.2	高级管理层在气候相关问题评估和管理方面的战略角色
战略	ST.1.	气候相关问题 气候相关战略
	ST.1.1	气候相关转型风险，例如政策、法律、技术、市场及气候变化所引发的声誉风险
	ST.1.2	气候相关物理风险，例如急性天气事件以及天气模式的长期变化 天气模式
	ST.1.3	气候相关议题的重大财务影响
	ST.1.4	对碳相关行业的信贷风险敞口
	ST.1.5	针对碳密集型行业（如化石燃料行业）的融资与投资
	ST.1.6	使用气候相关情景模型分析气候相关风险的影响
	ST.1.7	不同气候相关情景下银行战略的韧性
风险管理	RM.1.	气候相关风险管理
	RM.1.1	识别、评估和管理气候相关风险并将其纳入其中的流程 纳入整体风险管理
	RM.1.2	气候相关风险与财务风险之间的关系，例如信用风险、市场风险、流动性风险和操作风险
指标与目标	MT.1.	气候相关指标与目标
	MT.1.1	碳足迹、直接和间接温室气体排放
	MT.1.2	将气候相关绩效指标纳入薪酬政策
	MT.1.3	减排和碳中和目标

表6：细粒度TCFD标签概览

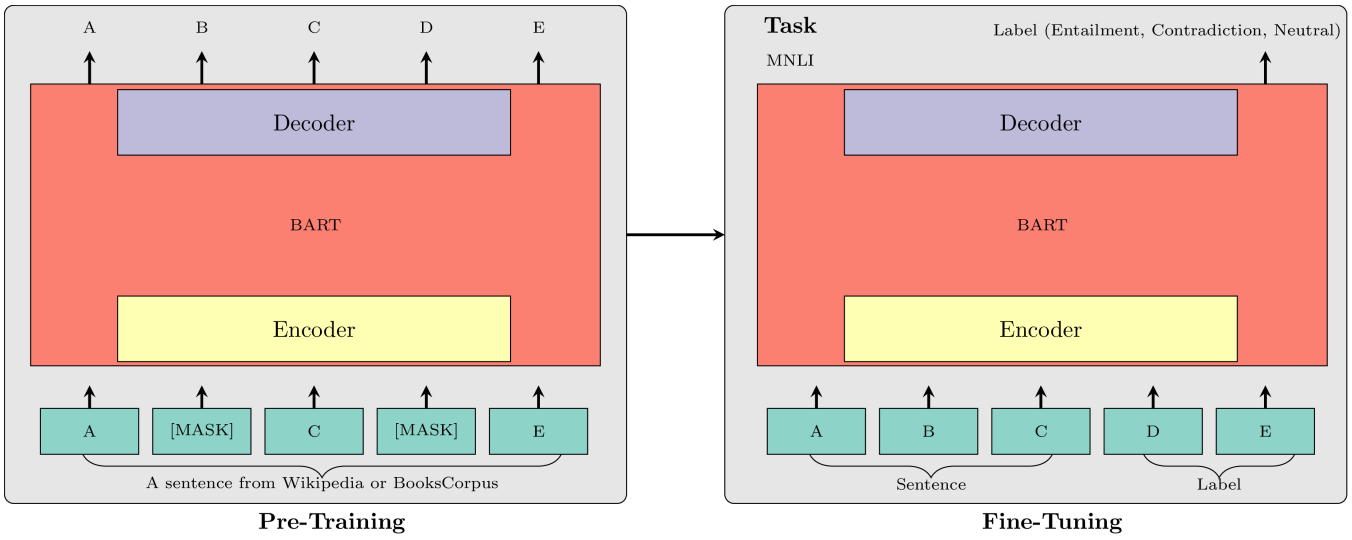

图4：左侧展示了在所有英文维基百科文章和 BooksCorpus 数据集上预训练的 BART 模型。通过遮蔽句子中的部分内容（[MASK]），模型经过训练以学习语义并预测缺失部分。此过程在预训练数据集中的所有句子上重复进行。右侧展示了在多项自然语言推理任务上微调的模型，其输出蕴含、矛盾和中性的概率。根据 Davison（2020）改编。

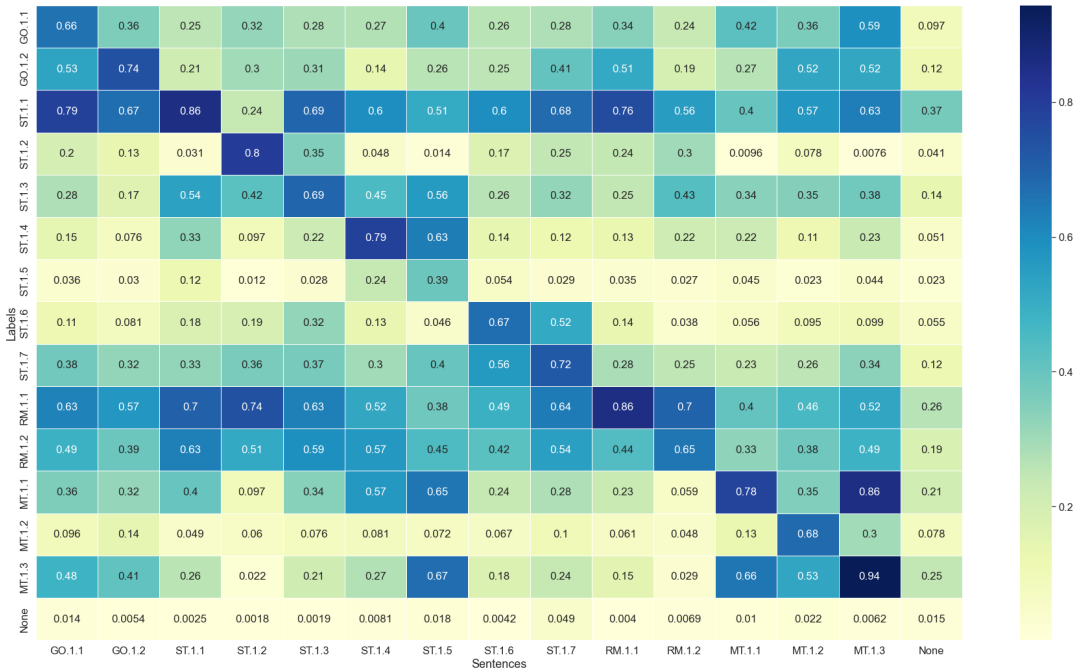

图5：该矩阵呈现了对我们的测试数据集进行零样本文本分类所得的结果。细粒度TCFD标签基于建议披露事项及金融业补充指南。