arXiv:2305.06047v1 [cs.IR] 10 May 2023

在 BIP! 服务中试点主题感知的研究影响评估功能

Serafeim Chatzopoulos[1 [0000−0003−1714−5225]]、Kleanthis Vichos[1 [0000−0002−8955−9489]]、Ilias Kanellos[1 [0000−0003−2146−3795]] 和 Thanasis Vergoulis[1 [0000−0003−0555−4128]]

IMSI, Athena RC,雅典, 希腊{schatz, kvichos, ilias.kanellos, vergoulis}@athenarc.gr

摘要。各种研究活动都依赖基于引文的影响力指标。然而，这些指标通常是在全局范围内计算的，因而难以在研究评估和知识发现等应用中得到正确解读。在本研究中，我们倡导使用主题感知型分类影响力指标，以缓解上述问题。此外，我们扩展了 BIP! 服务，使其支持这些指标，并展示了它们在真实研究活动中的优势。

1 引言

近年来，基于引文的影响力指标（如引文计数）被广泛用于促进各种与科研相关的活动。首先，科学文献搜索引擎（例如 Semantic Scholar[1]，BIP Finder[2]）使用这些指标对关键词搜索结果进行排序，帮助研究人员确定阅读优先级。此外，这些指标还被用于促进科研评估活动 [4], ，并成为监测科学产出的基础（例如 [2]）。大多数此类指标都基于网络分析算法，而这些算法依赖引文数据和出版物元数据（例如出版年份、作者列表等）。

然而，影响力指标也带来了困扰整个科研群体的各种问题。例如，科学文献搜索引擎所采用的指标数量有限，只能反映科学影响力的狭窄视角 [5]。具体而言，大多数搜索引擎仅支持引文计数，而引文计数存在一些已知问题（例如对近期文章存在偏见、容易受到过度自引攻击的影响）。此外，在科研评估中，评估者往往过度依赖影响力指标，而不深入考察研究人员的简历和出版物。将指标作为“评估捷径”已被认定为一种有问题的方法 [4] ，并且往往会导致不公平的科研评估。

但一个更重要的问题是，在上述应用中，用户可以比较来自不同领域的文章，这可能会导致误解。学术搜索引擎通常会返回

[1] Semantic Scholar：https://www.semanticscholar.org/[2] BIP! Finder：https://bip.imsi.athenarc.gr/

来自不同主题的结果，因为相同的关键词可能与多个领域相关。类似地，学术简历通常包含来自多个领域的出版物，因此直接比较影响力指标很可能会导致误判。

由于实际上不可能解决所有与影响力指标相关的问题，因此使用时始终应保持谨慎，并且只能将其作为其他（定性）证据的补充。不过，影响力指标在各种应用中仍可发挥辅助作用，因此，缓解其中一些问题仍然很有价值。基于此，我们将 BIP! 数据库 [7] 提供的多视角影响力指标改编为一组主题感知型分类指标。为将原始指标的数值转换为分类指标，我们将其转化为百分位排名类别（与 [1] 中所述的方法类似）。我们认为这很有用，因为分类指标更容易解读。最后，我们通过扩展 BIP! 服务 [3]，将这些指标纳入其中，展示它们在实际应用中的优势。

2 实施

2.1 主题感知型分类影响指标

如前所述，为缓解第 1 节提到的问题，我们主张采用 BIP! 数据库 [7] 提供的一种影响力指标变体。该数据库已采用多视角方法，提供多种指标，以反映出版物科学影响力的不同方面。本研究聚焦于以下指标：[4]

– 热度。 它根据底层引文网络，反映文章当前的影响力／受关注程度（即“热度”）。– 影响力。 它根据底层引文网络，反映文章在整个研究社群中的总体／长期影响力。– 引文数量。 文章收到的引文数量（也反映总体／长期影响力）。– 初始势头。 它根据底层引文网络，反映文章发表后初期的势头。

BIP! 数据库基于整个引文网络计算分数。根据指标值，可以按照每篇论文所属的百分位数[5] 为其指定一个全局分类值。因此，可以为每个初始指标定义五个分类影响指标。我们依次将这些分类指标称为“受欢迎度类别”“影响力类别”“引文数量类别”和“影响力类别”。

[3] BIP! 服务：https://bip.imsi.athenarc.gr/ [4] 有关这些指标的更多详情（例如计算算法）请参见：https://bip.imsi.athenarc.gr/site/indicators[5] 百分位数不易受到异常值的影响，即使底层数据严重偏斜，也能轻松计算。

我们进一步为每篇文章标注其相关主题（详情见第 2.2 节），然后计算上述分类指标的主题专属版本。这样，除了全局影响类别外，我们还会为每篇文章计算其所有相关主题下的主题专属影响类别。具体而言，对于每个主题，我们按以下方式计算所有指标的百分位数：首先，按照给定影响指标的降序排列与该主题相关的文章；然后，根据分数分布为每篇出版物指定一个百分位数，并为该文章分配相应类别。所有分类指标采用以下影响类别：前 0.01%、前 0.1%、前 1%、前 10%、平均（其余 90%）。

2.2 数据收集、处理与发布

为计算主题感知的分类影响指标，我们从 BIP! 数据库获取相应的影响指标分数。本研究使用的是第 8 版，其中包含近 1.34 亿篇文章的指标。[6] 随后，我们将这些文章与 OpenAlex [3] 中的（L2）主题相关联（共 284 个）。对于每篇出版物，我们仅保留置信度分数最高的三个主题，且分数必须大于 0。3。经过这一处理，BIP! 数据库中超过 75% 的文章至少关联了一个主题。接着，我们为每篇出版物计算主题特定影响力类别。对于给定主题，每篇出版物都会被分配一个影响类别，取自集合 {C1、C2、C3、C4、C5}；其中 C1 对应前 0.01% 类别，C5 对应平均影响力。我们将这些指标整合到 BIP! 数据库的数据集中，并在 Zenodo 上公开提供。

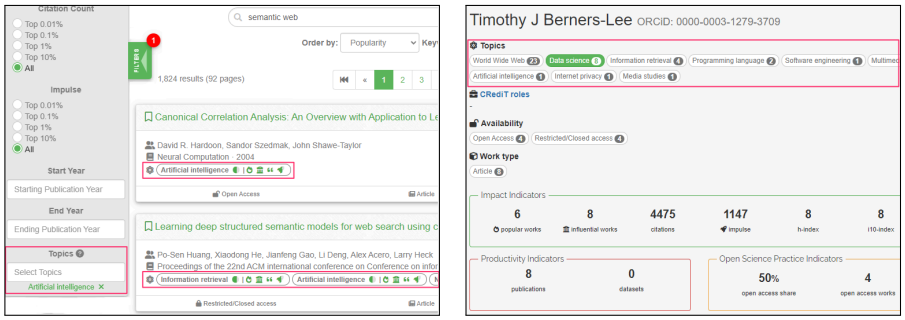
2.3 BIP!Services 扩展

为展示前述指标在实践中的实用性，我们聚焦于两个使用案例：科学知识发现和研究影响监测。对于前者，我们相应地扩展了 BIP! Finder [5] 学术搜索引擎，通过修改用户界面来（a）显示每个结果的主题及其依据各主题划分的影响类别，以及（b）支持基于主题的筛选。为实现影响类别的可视化，我们采用了一种紧凑的可视化方式，为每个类别的图标设置特定的颜色编码。图 1a 展示了“semantic web”查询的结果列表和筛选器。对于后者，我们扩展了 BIP! Scholar [6] 服务，该服务提供总结研究人员职业生涯的研究人员个人资料页面。具体而言，我们添加了一个主题筛选项，使研究人员能够查看其在所选主题上的影响（图 1b）。

3 个演示场景

在会议上，听众将有机会体验 BIP! 服务，并了解主题特定影响力指标在各种用例中带来的益处。我们还将演示以下场景。

[6] https://doi.org/10.5281/zenodo.4386934



（a）科学知识发现。 （b）监测研究者影响力。

图 1：BIP! 服务中的主题感知型出版物影响力指标。

科学知识发现。 一位听众在 BIP! Finder 中搜索关键词“语义网”，并使用主题筛选器确定自己只对与“人工智能”主题相关的文章感兴趣（图 1a）。每条结果都会列出相关主题，而影响力图标则会针对每个主题告知用户该结果所属的主题特定影响力类别。**监测研究者影响力。** 同一位听众使用 BIP! Scholar 查看蒂姆·伯纳斯-李的个人资料。他是一位知名的网络技术领域研究者（图 1b）。用户点击个人资料上方的各个主题分面（“数据科学”），即可查看伯纳斯-李在相应主题中的影响力（例如，他有多少篇热门作品）。

致谢

本研究由欧盟“地平线欧洲”计划项目 SciLake（GA：101058573）和 GraspOS（GA：101095129），以及欧盟 H2020 计划项目 OpenAIRE-Nexus（GA：101017452）共同资助。

参考文献

1. Bornmann, L., Leydesdorff, L., Mutz, R.：在文献计量数据分析中使用百分位数和百分位排名类别：机遇与局限。《信息计量学杂志》**7**(1), 158–165 (2013) 2. Papastefanatos, G., Papadopoulou, E., Meimaris, M., Lempesis, A., Martziou, S., Manghi, P., Manola, N.：开放科学观测站：监测欧洲的开放科学。载于：ADBiS/TPDL/EDA研讨会。第1260卷，第341–346页（2020）

3. Priem, J., Piwowar, H., Orr, R.：OpenAlex：一个完全开放的学术作品、作者、出版场所、机构和概念索引（2022），https://arxiv.org/abs/2205. 01833 4. Strinzel, M., Brown, J., Kaltenbrunner, W., de Rijcke, S., Hill, M.：改善学术简历、实现更公平科研评估的十种方法。《人文与社会科学传播》8(1), 1–4（2021）5. Vergoulis, T., Chatzopoulos, S., Kanellos, I., Deligiannis, P., Tryfonopoulos, C., Dalamagas, T.：BIP! Finder：利用基于影响力的排名促进科学文献检索。载于：CIKM。第2937–2940页（2019）6. Vergoulis, T., Chatzopoulos, S., Vichos, K., Kanellos, I., Mannocci, A., Manola, N., Manghi, P.：BIP! Scholar：促进公平研究人员评估的服务。载于：JCDL。第1–5页（2022）7. Vergoulis, T., Kanellos, I., Atzori, C., Mannocci, A., Chatzopoulos, S., Bruzzo, S.L., Manola, N., Manghi, P.：BIP! 数据库：科学出版物影响力指标数据集。载于：2021年万维网大会附属论文集。第456–460页（2021）