



(12) 发明专利申请

(10) 申请公布号 CN 122596889 A

(43) 申请公布日 2026. 08. 18

(21) 申请号 202610875275.2

G06F 16/907 (2019.01)

(22) 申请日 2026.06.17

G06F 16/2457 (2019.01)

(71) 申请人 北京印刷学院

地址 102600 北京市大兴区兴华大街二段1号

申请人 北京正乙科技有限公司

(72) 发明人 刘玉琴 田宗霖

(74) 专利代理机构 北京中和戎智知识产权代理有限公司 11995

专利代理师 刘晓玲

(51) Int. Cl.

G06Q 10/1053 (2023.01)

G06Q 10/0639 (2023.01)

G06Q 50/20 (2012.01)

G06F 16/901 (2019.01)

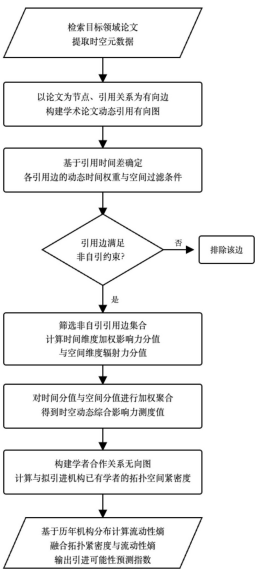
权利要求书4页 说明书17页 附图3页

(54) 发明名称

基于动态引用图谱的学术人才评价及引进预测方法及系统

(57) 摘要

本发明涉及学术人才评价技术领域,尤其涉及一种基于动态引用图谱的学术人才评价及引进预测方法。通过检索论文提取时空元数据,构建动态引用有向图并确定动态时间权重和空间过滤条件,筛选非自引引用边计算时间加权影响力分值与空间辐射力分值并加权聚合为时空动态综合影响力测度值,构建合作无向图确定拓扑空间紧密度与流动性指标并融合为引进可能性预测指数,实现了学术人才评价的时空动态综合准确性与引进预测可靠性的提升。



1. 基于动态引用图谱的学术人才评价及引进预测方法,其特征在于,包括:

从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值;

构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数。

2. 根据权利要求1所述的方法,其特征在于,从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据,包括:

按照预设领域范围与检索条件从学术文献数据库中检索并获取论文集合,提取每篇论文的数字对象唯一标识符、发表年份和月份构成的时间戳以及全部作者的姓名列表,形成携带完整元字段的原始论文记录数据集;

对所述作者列表中的每位作者,通过关联数据库分别获取其对应论文的所属机构全称和所属国家全称,以作者姓名为关联建表键生成包含文献标识符、时间戳、作者姓名、机构全称和国家全称的扁平化元数据条目;

将所述扁平化元数据条目按文献标识符进行分组归并,对同一文献标识符下因作者在不同机构兼职或机构名称书写变体导致的重复作者条目执行消歧处理,合并相同作者在该文献关联下的机构及国家归属记录,消除冗余条目并保留时间戳最早的归属信息;

将归并后的元数据条目按作者姓名进行跨论文的二次聚合,以作者为唯一标识构建覆盖各篇论文的发表时间阶段及相应机构和国家归属标注的完整时空元数据集。

3. 根据权利要求1所述的方法,其特征在于,以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件,包括:

以学术论文为节点并以被引论文指向施引论文的方向为有向边,构建学术论文动态引用有向图,为所述有向图中的每个节点赋予该节点对应论文的发表时间戳属性及作者的机构名称和国家名称归属标签;

遍历所述动态引用有向图的所有有向边,提取每条有向边所连接的被引论文节点与施引论文节点的发表时间戳,计算被引论文发表时间至当前观测时间之间的绝对时间差值,通过单调非增的时间衰减映射函数将所述绝对时间差值转换为位于预设区间内的动态时间权重;

针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与

被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记。

4.根据权利要求3所述的方法,其特征在于,针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记,包括:

对于所述动态引用有向图中的每条有向边,逐一采集施引论文节点上全部作者所对应的机构名称集合和国家名称集合,以及被引论文节点上全部作者所对应的机构名称集合和国家名称集合,形成施引方归属信息与被引方归属信息;

对所述施引方机构名称集合与所述被引方机构名称集合执行交集运算,在交集为空集时判定该有向边为跨机构引用边并赋予第一级跨机构空间过滤标记值,在交集为非空集时进一步比较施引方机构名称集合是否完全包含于被引方机构名称集合以识别机构层面的部分跨界情形并赋予对应的部分跨机构标记值;

对所述施引方国家名称集合与所述被引方国家名称集合执行交集运算,在交集为空集时判定该有向边为跨国引用边并赋予第二级跨国空间过滤标记值;

综合所述第一级跨机构标记值、部分跨机构标记值和第二级跨国标记值,为每条有向边组合输出包含跨机构标记位和跨国标记位的多级空间过滤标记向量。

5.根据权利要求1所述的方法,其特征在于,根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值,包括:

在所述动态引用有向图中筛选目标学者作为被引论文作者之一且施引论文作者集合与被引论文作者集合的交集为空集的有向边,构成目标学者的非自引引用边集合,记录所述非自引引用边集合中每条有向边携带的动态时间权重值和空间过滤标记值;

对所述非自引引用边集合中各边的动态时间权重值执行累加求和运算得到权重值总和,将目标学者的每条非自引引用边对应的动态时间权重值与所述权重值总和相除后累加,输出目标学者在时间维度上的加权影响力分值;

在机构维度和国家维度上分别对所述非自引引用边集合施加空间过滤标记值筛选,计算被标记为跨机构引用的有向边数占集合总边数的第一比例值作为机构维度辐射力分值,计算被标记为跨国引用的有向边数占集合总边数的第二比例值作为国家级维度辐射力分值;

为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值。

6.根据权利要求5所述的方法,其特征在于,为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值,包括:

接收所述时间维度加权影响力分值、机构维度辐射力分值及国家级维度辐射力分值,将三个分值分别通过极值归一化方法映射至零到一的闭区间内,消除各分值在量纲单位和数值量级范围上的差异性;

获取表征当前评价任务对时效性、跨机构传播力和跨国传播力三项评价维度侧重程度的一组权重配置参数,所述权重配置参数包含分别对应所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值的三个权重系数,且所述三个权重系数的累加结果为一;

将归一化后的时间维度加权影响力分值乘以对应的时间权重系数得到时间贡献分量,将归一化后的机构维度辐射力分值乘以对应的机构权重系数得到机构辐射分量,将归一化后的国家级维度辐射力分值乘以对应的国家权重系数得到国家辐射分量;

对所述时间贡献分量、机构辐射分量和国家辐射分量执行累加求和运算,将所得的累加结果输出为目标学者的时空动态综合影响力测度值。

7.根据权利要求1所述的方法,其特征在于,构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数,包括:

以学者为节点并以两学者在同一篇论文中存在合著关系为无向边,构建覆盖全部学者的合作关系无向图,在所述无向图中定位目标学者对应的节点以及拟引进机构当前已聘任的全部学者所对应的节点集合;

以目标学者节点为起点对所述无向图执行广度优先遍历,分别计算目标学者节点与所述节点集合中各个节点之间的最短路径长度,对所述节点集合中所有最短路径长度分别取倒数后计算算术平均值,将所述算术平均值作为目标学者相对于拟引进机构的拓扑空间紧密度值;

统计目标学者历年发表的论文在各所属机构上的分布频次,将所述分布频次代入以自然常数为底的信息熵函数计算得到机构分布熵值,以所述机构分布熵值作为量化目标学者机构变动倾向的流动性指标;

将所述拓扑空间紧密度值和所述机构分布熵值分别进行量纲归一化处理,为归一化后的紧密度值和流动性指标分配第一融合权重和第二融合权重,采用加权求和方式将两者耦合为引进可能性预测指数。

8.基于动态引用图谱的学术人才评价及引进预测系统,用于实现如权利要求1-7中任一项所述的方法,其特征在于,包括:

元数据单元,用于从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

构图单元,用于以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

影响力单元,用于根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动

态综合影响力测度值；

预测单元,用于构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数。

9.一种电子设备,其特征在于,包括:

处理器;

用于存储处理器可执行指令的存储器;

其中,所述处理器被配置为调用所述存储器存储的指令,以执行权利要求1至7中任意一项所述的方法。

10.一种计算机可读存储介质,其上存储有计算机程序指令,其特征在于,所述计算机程序指令被处理器执行时实现权利要求1至7中任意一项所述的方法。

基于动态引用图谱的学术人才评价及引进预测方法及系统

技术领域

[0001] 本发明涉及学术人才评价技术领域,尤其涉及基于动态引用图谱的学术人才评价及引进预测方法及系统。

背景技术

[0002] 在学术人才评价与引进预测领域,现有技术通常依赖于传统的文献计量指标体系进行评估。常见的做法包括统计论文发表数量、总被引频次、H指数及其衍生指标,以及基于期刊影响因子或学科归一化引用影响因子的加权计算。这些方法多采用静态累计或平均统计模式,以学者发表的所有论文为对象,计算其在全部时间跨度内的总量或均值,据此评判其学术水平和影响力。部分改进方案尝试引入引用半衰期或即时指数来考虑引用时间因素,但通常仅对引用总量进行简单的时间窗口截断或线性衰减处理,未建立引用关系随时间动态变化的精细权重模型。此外,现有评价模型大多在单一机构或学科内部进行横向比较,缺少对学者在跨机构、跨国家维度上引用辐射范围与强度的系统性分析。

[0003] 上述传统方法存在明显缺陷。静态或半静态的引用统计无法反映学者影响力在时间维度上的动态演化特征,早期发表的高被引论文与近期发表的引用关系在影响力贡献上被等同对待,导致评价结果滞后于学术表现的实际变化。更关键的是,现有评价体系普遍缺失对学者在机构与国家层面上的空间辐射能力的衡量,即施引论文所属机构或国家之间的地理关联未被纳入考量,难以揭示学者学术影响力在不同地域空间的扩散模式及其潜在的合作网络优势。在人才引进预测方面,常规做法极度依赖专家经验对候选人的履历和学术产出进行主观评判,缺乏对学者合作关系网络紧密度及其历史机构流动模式的量化分析,引进决策往往基于静态的简历解读而非对学者未来融入新环境的可预测性指标,导致评估结果片面且存在不确定性。

发明内容

[0004] 本发明实施例提供基于动态引用图谱的学术人才评价及引进预测方法及系统,能够解决现有技术中的问题。

[0005] 本发明实施例的第一方面,提供基于动态引用图谱的学术人才评价及引进预测方法,包括:

[0006] 从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

[0007] 以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

[0008] 根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力

测度值；

[0009] 构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数。

[0010] 从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据,包括:

[0011] 按照预设领域范围与检索条件从学术文献数据库中检索并获取论文集合,提取每篇论文的数字对象唯一标识符、发表年份和月份构成的时间戳以及全部作者的姓名列表,形成携带完整元字段的原始论文记录数据集;

[0012] 对所述作者列表中的每位作者,通过关联数据库分别获取其对应论文的所属机构全称和所属国家全称,以作者姓名为关联建表键生成包含文献标识符、时间戳、作者姓名、机构全称和国家全称的扁平化元数据条目;

[0013] 将所述扁平化元数据条目按文献标识符进行分组归并,对同一文献标识符下因作者在不同机构兼职或机构名称书写变体导致的重复作者条目执行消歧处理,合并相同作者在该文献关联下的机构及国家归属记录,消除冗余条目并保留时间戳最早的归属信息;

[0014] 将归并后的元数据条目按作者姓名进行跨论文的二次聚合,以作者为唯一标识构建覆盖各篇论文的发表时间阶段及相应机构和国家归属标注的完整时空元数据集。

[0015] 以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件,包括:

[0016] 以学术论文为节点并以被引论文指向施引论文的方向为有向边,构建学术论文动态引用有向图,为所述有向图中的每个节点赋予该节点对应论文的发表时间戳属性及作者的机构名称和国家名称归属标签;

[0017] 遍历所述动态引用有向图的所有有向边,提取每条有向边所连接的被引论文节点与施引论文节点的发表时间戳,计算被引论文发表时间至当前观测时间之间的绝对时间差值,通过单调非增的时间衰减映射函数将所述绝对时间差值转换为位于预设区间内的动态时间权重;

[0018] 针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记。

[0019] 针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记,包括:

[0020] 对于所述动态引用有向图中的每条有向边,逐一采集施引论文节点上全部作者所对应的机构名称集合和国家名称集合,以及被引论文节点上全部作者所对应的机构名称集

合和国家名称集合,形成施引方归属信息与被引方归属信息;

[0021] 对所述施引方机构名称集合与所述被引方机构名称集合执行交集运算,在交集为空集时判定该有向边为跨机构引用边并赋予第一级跨机构空间过滤标记值,在交集为非空集时进一步比较施引方机构名称集合是否完全包含于被引方机构名称集合以识别机构层面的部分跨界情形并赋予对应的部分跨机构标记值;

[0022] 对所述施引方国家名称集合与所述被引方国家名称集合执行交集运算,在交集为空集时判定该有向边为跨国引用边并赋予第二级跨国空间过滤标记值;

[0023] 综合所述第一级跨机构标记值、部分跨机构标记值和第二级跨国标记值,为每条有向边组合输出包含跨机构标记位和跨国标记位的多级空间过滤标记向量。

[0024] 根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值,包括:

[0025] 在所述动态引用有向图中筛选目标学者作为被引论文作者之一且施引论文作者集合与被引论文作者集合的交集为空集的有向边,构成目标学者的非自引引用边集合,记录所述非自引引用边集合中每条有向边携带的动态时间权重值和空间过滤标记值;

[0026] 对所述非自引引用边集合中各边的动态时间权重值执行累加求和运算得到权重值总和,将目标学者的每条非自引引用边对应的动态时间权重值与所述权重值总和相除后累加,输出目标学者在时间维度上的加权影响力分值;

[0027] 在机构维度和国家维度上分别对所述非自引引用边集合施加空间过滤标记值筛选,计算被标记为跨机构引用的有向边数占集合总边数的第一比例值作为机构维度辐射力分值,计算被标记为跨国引用的有向边数占集合总边数的第二比例值作为国家级维度辐射力分值;

[0028] 为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值。

[0029] 为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值,包括:

[0030] 接收所述时间维度加权影响力分值、机构维度辐射力分值及国家级维度辐射力分值,将三个分值分别通过极值归一化方法映射至零到一的闭区间内,消除各分值在量纲单位和数值量级范围上的差异性;

[0031] 获取表征当前评价任务对时效性、跨机构传播力和跨国传播力三项评价维度侧重程度的一组权重配置参数,所述权重配置参数包含分别对应所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值的三个权重系数,且所述三个权重系数的累加结果为一;

[0032] 将归一化后的时间维度加权影响力分值乘以对应的时间权重系数得到时间贡献分量,将归一化后的机构维度辐射力分值乘以对应的机构权重系数得到机构辐射分量,将归一化后的国家级维度辐射力分值乘以对应的国家权重系数得到国家辐射分量;

[0033] 对所述时间贡献分量、机构辐射分量和国家辐射分量执行累加求和运算,将所得的累加结果输出为目标学者的时空动态综合影响力测度值。

[0034] 构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数,包括:

[0035] 以学者为节点并以两学者在同一篇论文中存在合著关系为无向边,构建覆盖全部学者的合作关系无向图,在所述无向图中定位目标学者对应的节点以及拟引进机构当前已聘任的全部学者所对应的节点集合;

[0036] 以目标学者节点为起点对所述无向图执行广度优先遍历,分别计算目标学者节点与所述节点集合中各个节点之间的最短路径长度,对所述节点集合中所有最短路径长度分别取倒数后计算算术平均值,将所述算术平均值作为目标学者相对于拟引进机构的拓扑空间紧密度值;

[0037] 统计目标学者历年发表的论文在各所属机构上的分布频次,将所述分布频次代入以自然常数为底的信息熵函数计算得到机构分布熵值,以所述机构分布熵值作为量化目标学者机构变动倾向的流动性指标;

[0038] 将所述拓扑空间紧密度值和所述机构分布熵值分别进行量纲归一化处理,为归一化后的紧密度值和流动性指标分配第一融合权重和第二融合权重,采用加权求和方式将两者耦合为引进可能性预测指数。

[0039] 本发明实施例的第二方面,提供一种基于动态引用图谱的学术人才评价及引进预测系统,包括:

[0040] 元数据单元,用于从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

[0041] 构图单元,用于以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

[0042] 影响力单元,用于根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值;

[0043] 预测单元,用于构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数。

[0044] 本发明实施例的第三方面,提供一种电子设备,包括:

[0045] 处理器;

[0046] 用于存储处理器可执行指令的存储器;

[0047] 其中,所述处理器被配置为调用所述存储器存储的指令,以执行前述所述的方法。

[0048] 本发明实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机程序指令,所述计算机程序指令被处理器执行时实现前述所述的方法。

[0049] 基于动态引用图谱的时间权重修正机制,使各引用边按与当前观测时间的的时间差自适应衰减,从而精准凸显学者近期学术产出在引文网络中的实际贡献,避免早期高被引论文对评价结果的过度影响。空间过滤条件从机构或国家维度对引用边进行筛选,有效排除同机构或同国家内部引用的同质化效应,使空间辐射力分值客观反映学者跨区域学术影响的广度与深度。

[0050] 时空动态综合影响力测度值通过时间分值与空间分值的加权聚合,同时兼顾学术成果的持续影响能力与跨地域传播能力,形成对学者学术地位的全面量化。该指标既能识别在时间纵向上保持活跃且影响力稳定的学者,又能甄别在空间横向上具有广泛合作网络与辐射效应的学者,为人才评价提供多维度、高分辨率的判别依据。

[0051] 基于学者合作关系无向图的最短路径长度,可将目标学者与拟引进机构现有学者之间的合作亲疏程度量化为拓扑空间紧密度,直接体现双方在学术社群中的衔接潜力。流动性指标通过学者历年所属机构的分布熵值衡量其不同科研环境间的迁移活跃度,两者融合形成的引进可能性预测指数,能够科学预判目标学者对特定机构的嵌入难度与适应风险,显著提升学术人才引进决策的准确率与效率。

附图说明

[0052] 图1为基于动态引用图谱的学术人才评价及引进预测方法的流程示意图;

[0053] 图2为跨时空学术影响力权重计算示意图;

[0054] 图3为人才引进决策矩阵示意图。

具体实施方式

[0055] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0056] 下面以具体地实施例对本发明的技术方案进行详细说明。下面这几个具体的实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例不再赘述。

[0057] 参考图1至图3,本申请基于动态引用图谱的学术人才评价及引进预测方法包括:

[0058] 从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

[0059] 以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

[0060] 根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值;

[0061] 构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵

值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数。

[0062] 从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据,包括:

[0063] 按照预设领域范围与检索条件从学术文献数据库中检索并获取论文集合,提取每篇论文的数字对象唯一标识符、发表年份和月份构成的时间戳以及全部作者的姓名列表,形成携带完整元字段的原始论文记录数据集;

[0064] 对所述作者列表中的每位作者,通过关联数据库分别获取其对应论文的所属机构全称和所属国家全称,以作者姓名为关联建表键生成包含文献标识符、时间戳、作者姓名、机构全称和国家全称的扁平化元数据条目;

[0065] 将所述扁平化元数据条目按文献标识符进行分组归并,对同一文献标识符下因作者在不同机构兼职或机构名称书写变体导致的重复作者条目执行消歧处理,合并相同作者在该文献关联下的机构及国家归属记录,消除冗余条目并保留时间戳最早的归属信息;

[0066] 将归并后的元数据条目按作者姓名进行跨论文的二次聚合,以作者为唯一标识构建覆盖各论文的发表时间阶段及相应机构和国家归属标注的完整时空元数据集。

[0067] 按照预设领域范围与检索条件,从学术文献数据库(如Web of Science、Scopus或PubMed等)中执行检索操作,获取目标领域内的论文集合。预设领域范围可通过学科分类代码、关键词组合或期刊来源列表等方式加以限定,检索条件可包含时间窗口约束、语言类型过滤以及文献类型筛选(如仅保留期刊论文和会议论文,排除评论、勘误等非研究性文献)。对检索命中的每篇论文,提取其数字对象唯一标识符(DOI)作为全局唯一的文献主键,同时提取发表年份与发表月份构成时间戳,并提取全部作者的姓名列表。上述字段共同形成携带完整元字段的原始论文记录数据集,为后续作者归属关联提供结构化基础。

[0068] 对原始论文记录数据集中每篇论文的作者列表,以作者姓名为关联键,通过关联数据库(如ORCID、机构权威库或领域专属作者数据库)分别查询该作者在该篇论文发表时所对应的机构全称和国家全称。需要注意的是,同一作者在不同论文中可能隶属于不同机构,因此机构与国家的归属信息须与具体文献标识符进行绑定,而非仅与作者姓名绑定。完成关联查询后,将文献标识符、时间戳、作者姓名、机构全称和国家全称五个字段拼合为一条扁平化元数据条目,使得每位作者在每篇论文下均对应独立的一条记录。这种扁平化结构便于后续的分组归并与消歧处理,同时也保留了作者在不同时期机构归属动态变化的完整历史信息。

[0069] 将所有扁平化元数据条目按文献标识符进行分组归并,针对同一文献标识符下出现的重复作者条目执行消歧处理。重复条目的产生主要源于两类情形:其一,同一作者在该论文发表时同时持有多家机构的兼职身份,导致该作者在同一文献下存在多条机构归属记录;其二,同一机构存在名称书写变体(如缩写形式、不同语言版本或历史沿革导致的旧称),使得同一机构被识别为多条不同的机构条目。对上述两类重复情形,分别采取不同的处理策略:对于兼职导致的多机构条目,将同一作者在该文献关联下的所有机构全称和国家全称合并为集合形式保留,不进行丢弃;对于名称变体导致的冗余条目,通过字符串规范化(如统一大小写、去除标点、扩展缩写)和机构权威词表匹配,将变体名称归并至标准全称,消除冗余条目。在时间戳处理上,若同一文献标识符下因数据来源差异存在多个时间戳记录,则保留时间戳最早的归属信息,以确保时间维度上的一致性与可比性。

[0070] 归并与消歧完成后,对元数据条目集合按作者姓名执行跨论文的二次聚合操作,以作者为唯一标识,将该作者参与的所有论文的发表时间戳、对应机构全称集合和国家全称集合整合为一条完整的作者级时空元数据记录。对于作者姓名本身存在拼写变体或同名歧义的情形,可结合ORCID标识符、邮箱域名或共同作者网络等辅助信息进行作者消歧,确保同一自然人的所有论文记录被归并至同一标识符下,不同自然人的记录不被混淆合并。经过二次聚合后,每位作者对应一条完整时空元数据集合,其中包含该作者在不同论文发表阶段的时间节点以及相应的机构和国家归属标注,完整刻画了该作者在学术生涯中的时空分布轨迹。

[0071] 该完整时空元数据集合在结构上可表示为以作者为行索引、以各篇论文的时间戳和机构/国家归属为列字段的宽表形式,也可表示为以(作者标识符,文献标识符)为复合主键的长表形式,具体存储格式视后续计算需求而定。长表形式便于基于时间窗口的动态切片操作,宽表形式便于作者级别的横向比较分析。无论采用何种存储形式,均需保证时间戳的精度统一(建议统一至年月级别),机构全称和国家全称的标准化一致,以及作者标识符的全局唯一性,从而为后续动态引用有向图的构建、时空动态综合影响力测度值的计算以及引进可能性预测指数的融合计算提供高质量、结构完整的基础数据支撑。

[0072] 在实际工程实现中,检索与元数据提取阶段可采用批量API调用方式,通过分页参数控制每次请求的论文数量,避免单次请求数据量过大导致超时或内存溢出。关联数据库的查询可采用缓存机制,对已查询过的作者姓名与文献标识符组合的结果进行本地缓存,减少重复查询开销。消歧处理中的机构名称规范化可预先构建机构别名词表,将常见缩写、历史旧称和多语言变体统一映射至标准全称,提升归并准确率。跨论文二次聚合阶段可采用分布式计算框架对大规模数据集进行并行处理,以作者标识符为分区键将数据分发至不同计算节点,各节点独立完成本分区内的聚合操作,最终汇总输出完整时空元数据集合,满足大规模学术数据处理的性能要求。

[0073] 以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件,包括:

[0074] 以学术论文为节点并以被引论文指向施引论文的方向为有向边,构建学术论文动态引用有向图,为所述有向图中的每个节点赋予该节点对应论文的发表时间戳属性及作者的机构名称和国家名称归属标签;

[0075] 遍历所述动态引用有向图的所有有向边,提取每条有向边所连接的被引论文节点与施引论文节点的发表时间戳,计算被引论文发表时间至当前观测时间之间的绝对时间差值,通过单调非增的时间衰减映射函数将所述绝对时间差值转换为位于预设区间内的动态时间权重;

[0076] 针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记。

[0077] 构建学术论文动态引用有向图时,以每篇学术论文作为图中的一个独立节点,以

引用关系为连边方向依据,规定被引论文指向施引论文的方向为有向边的正方向。这一方向选择的逻辑在于:被引论文是知识来源,施引论文是知识接受方,从来源指向接受方可以直观反映知识流动的路径。在完成节点与边的基本拓扑结构构建之后,为每个节点赋予多维属性标签。具体地,每个论文节点携带该论文的发表时间戳属性,用于后续时间差计算;同时携带该论文全部作者的机构名称归属标签集合与国家名称归属标签集合,这两类标签均以集合形式存储,以应对一篇论文存在多位来自不同机构或不同国家的共同作者的情形。时间戳采用统一的日历日期精度,机构名称和国家名称在入库前经过规范化处理,以消除同一机构在不同文献中因缩写、译名差异导致的实体不一致问题。

[0078] 在为有向图中的所有有向边计算动态时间权重时,遍历图中的每一条有向边,提取该边所连接的被引论文节点的发表时间戳与当前观测时间,计算二者之间的绝对时间差值。设被引论文的发表时间戳对应的时间点为 t_p ,当前观测时间为 t_0 ,则绝对时间差值 Δt 满足 $\Delta t = t_0 - t_p$,其中 $\Delta t \geq 0$ 。之所以采用被引论文的发表时间而非施引论文的发表时间,是因为被引论文代表了被评价知识单元的产出时间,其距离当前时间越远,在当前学术生态中的时效性相对越低,以此为基础计算的衰减权重更能反映该引用关系对当前时期学术影响力的贡献程度。

[0079] 将绝对时间差值 Δt 输入单调非增的时间衰减映射函数,得到对应的动态时间权重 w_t 。所谓单调非增,是指随着 Δt 的增大, w_t 不增大,即时间越久远的引用边所获得的权重越小或不变。时间衰减映射函数的输出值被约束在预设区间内,通常将该区间设置为 $(0,1]$,使得最新产生的引用关系对应权重趋近于1,而年代极为久远的引用关系对应权重趋近于某一正下界而非归零,从而保留历史引用的基础贡献,避免因完全衰减导致信息丢失。时间衰减映射函数的具体形式可选取指数衰减函数,例如 $w_t = e^{-\lambda \Delta t}$,其中 $\lambda > 0$ 为衰减速率超参数,控制权重随时间下降的快慢;也可选取基于对数变换的分段线性映射,视具体应用场景中学科知识更新速度的差异进行调整。对于知识迭代较快的工程技术类领域,可设置较大的 λ 值以加速衰减;对于知识积累周期较长的基础理论类领域,可设置较小的 λ 值以减缓衰减。每条有向边在完成时间权重计算后,将 w_t 作为该边的权重属性存储于图结构中,供后续影响力计算阶段直接调用。

[0080] 完成动态时间权重的赋值之后,针对每条有向边执行空间过滤条件的生成步骤。对于图中的任意一条有向边,分别读取其施引论文节点所携带的作者机构名称集合(记为 A_{inst})与作者国家名称集合(记为 A_{ctry}),以及被引论文节点所携带的作者机构名称集合(记为 B_{inst})与作者国家名称集合(记为 B_{ctry})。在机构维度上,计算 A_{inst} 与 B_{inst} 的交集 $A_{inst} \cap B_{inst}$;若该交集为空集,则说明施引方与被引方不存在共同所属机构,即该引用关系属于跨机构引用情形,为该边生成跨机构引用标记,标记值置为真;若交集非空,则说明施引方与被引方至少存在一个共同机构,属于同机构内部引用,标记值置为假。

[0081] 在国家维度上,类似地计算 A_{ctry} 与 B_{ctry} 的交集 $A_{ctry} \cap B_{ctry}$;若该交集为空集,则说明施引方与被引方不存在共同所属国家,该引用关系属于跨国引用情形,为该边生成跨国引用标记,标记值置为真;若交集非空,则说明双方至少共享一个国家归属,属于同国引用,标记值置为假。上述双维度比较独立执行,互不干扰,一条有向边可以同时携带跨机构引用标记为真且跨国引用标记为真(即既跨机构又跨国)、跨机构引用标记为真而跨国

引用标记为假(跨机构但同国)、跨机构引用标记为假而跨国引用标记为真(同机构但跨国,此情形在实践中较为少见但理论上可能存在于跨国联合机构场景)、以及双标记均为假(同机构同国)等四种组合情形。

[0082] 这种双维度独立标记的设计使得后续在计算空间辐射力分值时,可以灵活地根据分析目标选择不同的过滤粒度。若分析目标是衡量目标学者的国际影响力,则筛选跨国引用标记为真的引用边集合;若分析目标是衡量目标学者突破所在机构圈层的影响力,则筛选跨机构引用标记为真的引用边集合;若需要同时评估两个维度的辐射范围,则可分别基于两类标记独立筛选后再进行加权融合。空间过滤标记与动态时间权重共同作为有向边的属性被持久化存储,构成完整的动态引用有向图的边属性体系,为后续学者影响力计算和引进预测提供结构化的图数据基础。

[0083] 在实际工程实现中,考虑到学术文献数据库中论文数量庞大、引用关系稠密,遍历所有有向边并为每条边执行集合交集运算的计算开销较大。为提升效率,可在图构建阶段对机构名称集合和国家名称集合进行哈希索引预处理,将集合交集判断转化为哈希集合的查找操作,将单次交集判断的时间复杂度从线性降低至近似常数级别。此外,对于同一机构或同一国家的作者群体之间的大量内部引用边,可通过批量标记的方式减少重复计算,进一步提升整体图构建的处理速度,保证在大规模文献数据场景下动态引用有向图能够在合理时间内完成构建与属性赋值。

[0084] 针对每条有向边,分别提取其施引论文节点的作者机构名称集合和国家名称集合以及被引论文节点的作者机构名称集合和国家名称集合,在机构维度上比较施引方机构集合与被引方机构集合二者的交集是否为空,在国家维度上比较施引方国家集合与被引方国家集合二者的交集是否为空,根据双维度比较结果生成对应跨机构引用情形和跨国引用情形的空间过滤标记,包括:

[0085] 对于所述动态引用有向图中的每条有向边,逐一采集施引论文节点上全部作者所对应的机构名称集合和国家名称集合,以及被引论文节点上全部作者所对应的机构名称集合和国家名称集合,形成施引方归属信息与被引方归属信息;

[0086] 对所述施引方机构名称集合与所述被引方机构名称集合执行交集运算,在交集为空集时判定该有向边为跨机构引用边并赋予第一级跨机构空间过滤标记值,在交集为非空集时进一步比较施引方机构名称集合是否完全包含于被引方机构名称集合以识别机构层面的部分跨界情形并赋予对应的部分跨机构标记值;

[0087] 对所述施引方国家名称集合与所述被引方国家名称集合执行交集运算,在交集为空集时判定该有向边为跨国引用边并赋予第二级跨国空间过滤标记值;

[0088] 综合所述第一级跨机构标记值、部分跨机构标记值和第二级跨国标记值,为每条有向边组合输出包含跨机构标记位和跨国标记位的多级空间过滤标记向量。

[0089] 在完成动态引用有向图的构建之后,需要对图中每条有向边逐一执行空间归属信息的采集与比对操作,以生成反映引用关系跨界程度的空间过滤标记。针对图中的任意一条有向边,其起点为施引论文节点,终点为被引论文节点。从施引论文节点出发,遍历该论文的全部署名作者,依次读取每位作者在发表该论文时所归属的机构名称与所在国家名称,将所有作者的机构名称汇总为施引方机构名称集合,将所有作者的国家名称汇总为施引方国家名称集合。对被引论文节点执行同样的遍历操作,形成被引方机构名称集合与被

引方国家名称集合。上述四个集合共同构成该有向边的双维度归属信息,是后续空间过滤标记生成的基础输入。

[0090] 在机构维度上,对施引方机构名称集合与被引方机构名称集合执行集合交集运算。若交集结果为空集,则表明施引论文的所有作者所属机构与被引论文的所有作者所属机构之间不存在任何重叠,该引用行为跨越了机构边界,判定该有向边为跨机构引用边,并为其赋予第一级跨机构空间过滤标记值,记为标记位 f_{inst} 取值为1。这一标记直接反映了知识在不同机构之间的流动事实,在后续计算空间辐射力分值时具有重要的筛选意义。

[0091] 当施引方机构集合与被引方机构集合的交集为非空集时,说明施引方与被引方之间存在至少一个共同机构,但这并不意味着双方机构完全一致。此时进一步判断施引方机构名称集合是否完全包含于被引方机构名称集合,即检验施引方机构集合是否为被引方机构集合的真子集。若施引方机构集合中存在至少一个机构名称不属于被引方机构集合,则认定该有向边存在机构层面的部分跨界情形,为其赋予部分跨机构标记值,记为标记位 f_{inst} 取值为0.5,以区别于完全跨机构(标记值为1)和完全同机构(标记值为0)两种极端情形。引入部分跨机构标记的原因在于,多机构合著论文的引用关系往往同时涵盖机构内引用与机构间引用两种成分,单纯以二值标记无法精确刻画这类复合情形,而连续化的标记值能够更细腻地表达跨界程度的差异,从而使后续空间辐射力的计算更具区分度。

[0092] 在国家维度上,对施引方国家名称集合与被引方国家名称集合执行相同逻辑的交集运算。若交集结果为空集,则判定该有向边为跨国引用边,为其赋予第二级跨国空间过滤标记值,记为标记位 f_{ctry} 取值为1;若交集非空,则记标记位 f_{ctry} 取值为0,表示该引用行为在国家维度上属于境内引用。国家维度的过滤标记独立于机构维度的标记进行赋值,二者并行生成,互不干扰。这样的设计保证了在后续分析中既可以单独考察机构层面的跨界引用模式,也可以单独考察国家层面的跨界引用模式,或者将两个维度联合使用以支持更精细的空间辐射力分析。

[0093] 在完成机构维度与国家维度的标记赋值之后,将两个维度的标记值组合输出,形成每条有向边的多级空间过滤标记向量。该标记向量记为 $v_{spatial}$,其形式为 (f_{inst}, f_{ctry}) ,其中 f_{inst} 取值范围为 $\{0, 0.5, 1\}$, f_{ctry} 取值范围为 $\{0, 1\}$ 。整个向量的维度固定为2,每个分量对应一个空间维度的过滤标记位,两个分量的组合共可形成6种不同的标记模式。例如,当 $f_{inst} = 1$ 且 $f_{ctry} = 1$ 时,表示该引用边同时满足跨机构和跨国两个条件,属于最高级别的跨界引用;当 $f_{inst} = 0$ 且 $f_{ctry} = 0$ 时,表示该引用边在机构和国家两个维度上均属于内部引用,对空间辐射力的贡献最低。

[0094] 在实际处理过程中,机构名称集合与国家名称集合的构建需要对原始数据进行规范化预处理。由于同一机构在不同文献记录中可能存在名称缩写、全称、别称等多种表述形式,需要在数据入库阶段对机构名称执行统一的标准化映射,确保同一机构在所有论文记录中以唯一的规范名称出现。同理,国家名称也需要统一映射至标准国家代码或规范全称,以避免因名称不一致导致交集计算结果出现偏差。经过规范化处理后,集合交集运算的结果才能真实反映机构或国家层面的归属重叠情况。

[0095] 多级空间过滤标记向量的生成是面向整个动态引用有向图中所有有向边批量执行的。在具体实现上,可以按照有向边的列表顺序依次处理每条边,对每条边完成四个集合

的采集、两次交集运算以及标记赋值操作,最终将所有边的标记向量存储为与边列表一一对应的向量序列,供后续计算目标学者空间辐射力分值时按需检索与过滤。这种批量化的处理方式能够有效支持大规模学术文献数据库场景下的高效计算,保证空间过滤标记生成过程的可扩展性与一致性。

[0096] 根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值,包括:

[0097] 在所述动态引用有向图中筛选目标学者作为被引论文作者之一且施引论文作者集合与被引论文作者集合的交集为空集的有向边,构成目标学者的非自引引用边集合,记录所述非自引引用边集合中每条有向边携带的动态时间权重值和空间过滤标记值;

[0098] 对所述非自引引用边集合中各边的动态时间权重值执行累加求和运算得到权重值总和,将目标学者的每条非自引引用边对应的动态时间权重值与所述权重值总和相除后累加,输出目标学者在时间维度上的加权影响力分值;

[0099] 在机构维度和国家维度上分别对所述非自引引用边集合施加空间过滤标记值筛选,计算被标记为跨机构引用的有向边数占集合总边数的第一比例值作为机构维度辐射力分值,计算被标记为跨国引用的有向边数占集合总边数的第二比例值作为国家级维度辐射力分值;

[0100] 为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值。

[0101] 在完成动态引用有向图的构建、各边动态时间权重赋值及空间过滤标记向量生成之后,针对目标学者进行影响力量化计算的前提是从全图中精确筛选出有效的非自引引用边集合。具体而言,遍历整个动态引用有向图中的所有有向边,保留满足以下两个条件的边:其一,目标学者属于被引论文的作者集合;其二,施引论文的作者集合与被引论文的作者集合的交集为空集,即两篇论文之间不存在任何共同作者。满足上述两个条件的有向边共同构成目标学者的非自引引用边集合,记为 $E_{scholar}$ 。对集合 $E_{scholar}$ 中的每条有向边,同步记录其已携带的动态时间权重值 w_t 以及空间过滤标记向量 $v_{spatial}$ 中的各分量值,供后续各维度分值计算直接调用。

[0102] 引入非自引约束的必要性在于,若允许作者引用自己参与撰写的论文,则所统计的引用次数会虚增,无法真实反映该学者在学术共同体中的外部影响力。通过交集为空集这一严格判断,可以有效排除自引噪声,使后续分值更具客观性。在实际数据处理中,作者姓名存在多种形式变体(如缩写、译名差异等),建议在构建作者集合时先对作者标识符进行规范化消歧处理,以保证交集判断的准确性。

[0103] 完成非自引引用边集合 $E_{scholar}$ 的筛选后,计算目标学者在时间维度上的加权影响力分值 S_t 。设集合 $E_{scholar}$ 共包含 n 条有向边,第 k 条有向边对应的动态时间权重值为 $w_t^{(k)}$,则所有边的权重值总和为 $w_{sum} = \sum_{k=1}^n w_t^{(k)}$ 。在此基础上,将每条边的权重值除以权重值总和后累加,得到归一化加权影响力分值:

$$[0104] \quad S_t = \sum_{k=1}^n \frac{w_t^{(k)}}{w_{sum}}$$

[0105] 从计算结果来看, S_t 的数值实际上等于对各引用边权重进行归一化后的加权求和,其物理含义为:近期被引用的边对分值贡献更大,而时间久远的引用边因权重衰减而贡献相对较小。这种设计使得 S_t 不仅反映了目标学者被引次数的绝对量,还体现了其学术影响力在时间维度上的活跃程度——若目标学者近年来持续被引用,则 S_t 相对较高;若引用主要集中于早期,则 S_t 相对偏低。

[0106] 在机构维度辐射力分值的计算中,从集合 $E_{scholar}$ 中筛选出空间过滤标记向量中机构维度分量 f_{inst} 取值为 0.5 或 1 的有向边,即被标记为部分跨机构引用或完全跨机构引用的边,统计其数量记为 n_{inst} 。将 n_{inst} 除以集合 $E_{scholar}$ 的总边数 n ,得到机构维度辐射力分值:

$$[0107] \quad S_{inst} = \frac{n_{inst}}{n}$$

[0108] S_{inst} 反映了目标学者的研究成果被来自不同机构的学者所引用的比例,数值越高说明其学术影响力在机构层面的扩散范围越广,不局限于本机构内部的引用。值得注意的是,此处将 f_{inst} 取值为 0.5 (部分跨机构) 和 1 (完全跨机构) 的边均纳入计数,是因为只要施引论文中存在来自其他机构的作者,就意味着该引用具有一定的跨机构传播效应。

[0109] 在国家维度辐射力分值的计算中,从集合 $E_{scholar}$ 中筛选出空间过滤标记向量中国家维度分量 f_{ctry} 取值为 1 的有向边,即被标记为跨国引用的边,统计其数量记为 n_{ctry} 。将 n_{ctry} 除以集合总边数 n ,得到国家维度辐射力分值:

$$[0110] \quad S_{ctry} = \frac{n_{ctry}}{n}$$

[0111] S_{ctry} 衡量了目标学者的研究成果在国际范围内被引用的比例,是评价其国际学术影响力的重要指标。跨国引用比例高的学者,其研究成果往往具有更广泛的国际认可度,在学术人才引进评价中具有重要参考价值。

[0112] 完成三个维度的分值计算后,为时间维度加权影响力分值 S_t 、机构维度辐射力分值 S_{inst} 和国家维度辐射力分值 S_{ctry} 分别配置对应的加权系数 α_t 、 α_{inst} 和 α_{ctry} ,要求三个加权系数满足归一化约束,即 $\alpha_t + \alpha_{inst} + \alpha_{ctry} = 1$,且各系数均为非负实数。将三个分值分别与对应加权系数相乘后求和,得到时空动态综合影响力测度值 $I_{comprehensive}$:

$$[0113] \quad I_{comprehensive} = \alpha_t \cdot S_t + \alpha_{inst} \cdot S_{inst} + \alpha_{ctry} \cdot S_{ctry}$$

[0114] 加权系数的取值可根据评价场景的侧重点进行灵活调整。例如,在评价国际化人才时,可适当提高 α_{ctry} 的权重;在评价跨学科领域合作活跃度时,可适当提高 α_{inst} 的权重;在评价近期学术活跃度时,可适当提高 α_t 的权重。在缺乏先验偏好的默认场景下,可将三个系数设置为等权重,即 $\alpha_t = \alpha_{inst} = \alpha_{ctry} = 1/3$ 。

[0115] $I_{comprehensive}$ 的取值范围为 [0,1],数值越接近 1 表示目标学者在时间活跃度、跨机构影响力和国际影响力三个维度上的综合表现越优秀。该测度值将作为后续学者引进可能性预测模型的重要输入特征之一,与基于合作关系图的拓扑空间紧密度和流动性指标共同构成完整的学术人才评价与引进预测体系。在实际应用中,可针对同一目标领域内的多位候选学者分别计算 $I_{comprehensive}$,并按照该值进行排序,为拟引进机构提供量化的学术影响力参考依据。

[0116] 为所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值分别配置对应的加权系数,将三个分值分别与对应加权系数相乘后求和,输出时空动态综合影响力测度值,包括:

[0117] 接收所述时间维度加权影响力分值、机构维度辐射力分值及国家级维度辐射力分值,将三个分值分别通过极值归一化方法映射至零到一的闭区间内,消除各分值在量纲单位和数值量级范围上的差异性;

[0118] 获取表征当前评价任务对时效性、跨机构传播力和跨国传播力三项评价维度侧重程度的一组权重配置参数,所述权重配置参数包含分别对应所述时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值的三个权重系数,且所述三个权重系数的累加结果为一;

[0119] 将归一化后的时间维度加权影响力分值乘以对应的时间权重系数得到时间贡献分量,将归一化后的机构维度辐射力分值乘以对应的机构权重系数得到机构辐射分量,将归一化后的国家级维度辐射力分值乘以对应的国家权重系数得到国家辐射分量;

[0120] 对所述时间贡献分量、机构辐射分量和国家辐射分量执行累加求和运算,将所得的累加结果输出为目标学者的时空动态综合影响力测度值。

[0121] 在完成时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值的独立计算后,三个分值往往因计算方式不同而在数值量级和量纲单位上存在显著差异。若直接对三个原始分值施加加权求和,会导致数值量级较大的分值对最终结果产生不成比例的支配作用,从而掩盖其他维度的真实贡献。因此,在执行加权聚合之前,需要对三个分值分别进行极值归一化处理,将其统一映射至零到一的闭区间内。

[0122] 极值归一化的具体做法是:针对同一批次评价任务中所有被评价学者在同一维度上的分值集合,分别取该集合中的最大值与最小值,将每位学者的原始分值减去最小值后除以最大值与最小值之差,从而得到归一化后的分值。以时间维度加权影响力分值为例,设当前批次中所有学者的时间维度加权影响力分值中最大值为 $S_{t,max}$,最小值为 $S_{t,min}$,则目标学者归一化后的时间维度加权影响力分值 $\hat{S}_t$ 为 $\hat{S}_t = \frac{S_t - S_{t,min}}{S_{t,max} - S_{t,min}}$ 。当批次中所有学者的某一维度分值完全相同、即最大值等于最小值时,为避免分母为零的情形,将该维度所有学者的归一化分值统一赋值为零,表明该维度在当前批次中不具备区分能力。机构维度辐射力分值和国家级维度辐射力分值的归一化处理方式与上述过程完全一致,分别在各自维度的分值集合内独立完成归一化,归一化后的机构维度辐射力分值记为 $\hat{S}_{inst}$,归一化后的国家级维度辐射力分值记为 $\hat{S}_{ctry}$ 。

[0123] 完成归一化后,需要获取一组权重配置参数,用于表征当前评价任务对时效性、跨机构传播力和跨国传播力三项评价维度的侧重程度。这组权重配置参数包含三个权重系数,分别对应时间维度加权影响力分值、机构维度辐射力分值和国家级维度辐射力分值,记为 β_t 、 β_{inst} 和 β_{ctry} 。三个权重系数必须满足非负约束,且三者之和严格等于一,即 $\beta_t + \beta_{inst} + \beta_{ctry} = 1$,以确保最终综合影响力测度值的输出范围仍落在零到一的闭区间内,便于后续跨学者横向比较。

[0124] 权重配置参数的具体取值由评价任务的应用场景和评价目标决定。在以引进国际顶尖学者为导向的评价场景中,跨国传播力维度对于判断学者的国际影响力具有更高的参

考价值,此时可适当提高 β_{ctry} 的取值;在以推动国内机构间学术交流为目标的评价场景中,跨机构传播力维度更能反映学者在国内学术生态中的辐射能力,此时可适当提高 β_{inst} 的取值;在以评价学者近期学术活跃度和成果时效性为核心的场景中,时间维度加权影响力分值的权重 β_t 应相应提高。权重配置参数可由评价机构的专家组根据实际需求手动设定,也可通过历史评价数据的统计分析自动推荐一组默认参数值。

[0125] 获得归一化分值和权重配置参数后,分别计算三个维度的加权贡献分量。将归一化后的时间维度加权影响力分值 $\hat{S}_t$ 乘以对应的时间权重系数 β_t ,得到时间贡献分量 $C_t = \beta_t \cdot \hat{S}_t$;将归一化后的机构维度辐射力分值 $\hat{S}_{inst}$ 乘以对应的机构权重系数 β_{inst} ,得到机构辐射分量 $C_{inst} = \beta_{inst} \cdot \hat{S}_{inst}$;将归一化后的国家级维度辐射力分值 $\hat{S}_{ctry}$ 乘以对应的国家权重系数 β_{ctry} ,得到国家辐射分量 $C_{ctry} = \beta_{ctry} \cdot \hat{S}_{ctry}$ 。三个分量分别独立计算,互不干扰,每个分量的取值范围均在零到对应权重系数之间。

[0126] 对时间贡献分量、机构辐射分量和国家辐射分量执行累加求和运算,将所得累加结果作为目标学者的时空动态综合影响力测度值输出,即 $I_{comprehensive} = C_t + C_{inst} + C_{ctry}$ 。由于三个归一化分值均在零到一之间,且三个权重系数之和为一,因此 $I_{comprehensive}$ 的取值范围同样落在零到一的闭区间内。 $I_{comprehensive}$ 越接近一,表明目标学者在时效性、跨机构传播力和跨国传播力三个维度上的综合表现越突出; $I_{comprehensive}$ 越接近零,则表明目标学者在上述三个维度上的综合影响力相对较弱。

[0127] 在实际评价流程中, $I_{comprehensive}$ 可作为学者排序和筛选的核心量化依据。评价机构可根据预设的阈值对候选学者进行分级,将 $I_{comprehensive}$ 超过特定阈值的学者纳入重点考察名单,进而结合后续构建的引进可能性预测指数,对候选学者的引进优先级作出综合研判。此外,由于权重配置参数 β_t 、 β_{inst} 和 β_{ctry} 可根据评价任务动态调整,同一批学者在不同评价场景下可能呈现出不同的排序结果,这种灵活性使得该方法能够适应多样化的学术人才评价需求,而无需重新采集或计算底层分值数据,只需在聚合阶段调整权重配置即可完成评价策略的切换。

[0128] 构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性预测指数,包括:

[0129] 以学者为节点并以两学者在同一篇论文中存在合著关系为无向边,构建覆盖全部学者的合作关系无向图,在所述无向图中定位目标学者对应的节点以及拟引进机构当前已聘任的全部学者所对应的节点集合;

[0130] 以目标学者节点为起点对所述无向图执行广度优先遍历,分别计算目标学者节点与所述节点集合中各个节点之间的最短路径长度,对所述节点集合中所有最短路径长度分别取倒数后计算算术平均值,将所述算术平均值作为目标学者相对于拟引进机构的拓扑空间紧密度值;

[0131] 统计目标学者历年发表的论文在各所属机构上的分布频次,将所述分布频次代入以自然常数为底的信息熵函数计算得到机构分布熵值,以所述机构分布熵值作为量化目标学者机构变动倾向的流动性指标;

[0132] 将所述拓扑空间紧密度值和所述机构分布熵值分别进行量纲归一化处理,为归一

化后的紧密度值和流动性指标分配第一融合权重和第二融合权重,采用加权求和方式将两者耦合为引进可能性预测指数。

[0133] 构建学者合作关系无向图时,以学术文献数据库中检索到的全部学者作为节点集合,若任意两位学者在同一篇论文中存在合著关系,则在两者对应节点之间添加一条无向边,从而形成覆盖该领域全量学者的合作关系无向图。合著关系的判断依据为论文作者列表中同时出现两位学者的署名,无论其排名先后。该无向图的规模通常较大,节点数量与领域内活跃学者总数相当,边的数量则取决于合著论文的密集程度。在构建完成后,在图中定位目标学者对应的节点,同时检索拟引进机构当前已聘任的全部学者,将其在无向图中对应的节点汇聚为一个节点集合,记为 V_{inst} ,该集合的规模反映了拟引进机构现有学术团队的体量。

[0134] 以目标学者节点为起点,对合作关系无向图执行广度优先遍历(BFS)。广度优先遍历从起始节点出发,按照与起始节点的跳数由近及远逐层扩展,能够高效地计算起始节点到图中所有可达节点的最短路径长度。对于 V_{inst} 中的每一个节点 v_j ,记目标学者节点到 v_j 的最短路径长度为 d_j ,即两者之间在合作关系无向图中经过的最少边数。若目标学者节点与 v_j 之间不可达(即两者处于不同的连通分量中),则将 d_j 设为一个足够大的惩罚值,使其倒数趋近于零,从而在后续计算中自然地体现出极低的关联程度。

[0135] 对节点集合 V_{inst} 中所有节点的最短路径长度分别取倒数,再计算这些倒数的算术平均值,将该算术平均值作为目标学者相对于拟引进机构的拓扑空间紧密度值,记为 C_{topo} 。具体地,设 V_{inst} 中共有 M 个节点,则拓扑空间紧密度值的计算方式为:

$$[0136] \quad C_{topo} = \frac{1}{M} \sum_{j=1}^M \frac{1}{d_j}$$

[0137] 其中 d_j 为目标学者节点到 V_{inst} 中第 j 个节点的最短路径长度, M 为 V_{inst} 中节点的总数量。 C_{topo} 的取值范围为 $[0,1]$,当目标学者与拟引进机构所有已聘学者均直接合著(即 $d_j = 1$)时, C_{topo} 取得最大值 1;当目标学者与所有已聘学者均不可达时, C_{topo} 趋近于 0。该指标从合作网络的拓扑结构出发,客观地量化了目标学者与拟引进机构学术团队之间的学术社交距离,路径越短说明学术圈层交叠越紧密,被引进后融入团队的可能性也越高。

[0138] 统计目标学者历年发表的全部论文,提取每篇论文所标注的作者所属机构信息,以论文为统计单元,汇总目标学者在各所属机构上的出现频次,形成一个机构频次分布。设目标学者历史上共归属于 R 个不同机构,第 r 个机构对应的论文频次为 c_r ,则目标学者历年论文在各机构上的分布概率 p_r 为:

$$[0139] \quad p_r = \frac{c_r}{\sum_{r=1}^R c_r}$$

[0140] 将上述分布概率代入以自然常数 e 为底的信息熵函数,计算得到机构分布熵值,记为 H_{mob} ,其计算方式为:

$$[0141] \quad H_{mob} = -\sum_{r=1}^R p_r \ln p_r$$

[0142] H_{mob} 的取值随机构分布的均匀程度而变化。若目标学者长期固定在单一机构发表论文,则 p_r 集中于某一个机构,熵值趋近于 0,表明流动倾向极低;若目标学者历年频繁更换所属机构,各机构出现频次趋于均等,则熵值较高,表明其具有较强的机构变动倾

向。以 H_{mob} 作为量化目标学者机构变动倾向的流动性指标,能够有效区分长期稳定型学者与频繁流动型学者,为引进可能性的预测提供重要的行为特征依据。

[0143] 在融合拓扑空间紧密度值与流动性指标之前,需要分别对两者进行量纲归一化处理,以消除两个指标在数值尺度上的差异。对拓扑空间紧密度值 C_{topo} 在当前批次所有候选学者中执行最小-最大归一化,记归一化后的紧密度值为 $\hat{C}_{topo}$,归一化公式为:

$$[0144] \quad \hat{C}_{topo} = \frac{C_{topo} - C_{topo,min}}{C_{topo,max} - C_{topo,min}}$$

[0145] 其中 $C_{topo,max}$ 和 $C_{topo,min}$ 分别为当前批次所有候选学者拓扑空间紧密度值中的最大值和最小值。对机构分布熵值 H_{mob} 采用相同的最小-最大归一化方式,记归一化后的流动性指标为 $\hat{H}_{mob}$,归一化公式为:

$$[0146] \quad \hat{H}_{mob} = \frac{H_{mob} - H_{mob,min}}{H_{mob,max} - H_{mob,min}}$$

[0147] 其中 $H_{mob,max}$ 和 $H_{mob,min}$ 分别为当前批次所有候选学者机构分布熵值中的最大值和最小值。归一化处理后,两个指标均被压缩至 $[0,1]$ 区间,具备直接进行加权融合的数值基础。

[0148] 为归一化后的拓扑空间紧密度值 $\hat{C}_{topo}$ 分配第一融合权重 γ_1 ,为归一化后的流动性指标 $\hat{H}_{mob}$ 分配第二融合权重 γ_2 ,采用加权求和方式将两者耦合为引进可能性预测指数,记为 $P_{recruit}$,计算方式为:

$$[0149] \quad P_{recruit} = \gamma_1 \cdot \hat{C}_{topo} + \gamma_2 \cdot \hat{H}_{mob}$$

[0150] 其中 $\gamma_1 + \gamma_2 = 1$,且 $\gamma_1, \gamma_2 \geq 0$ 。两个融合权重的具体取值可根据实际引进场景的侧重点灵活调整:若拟引进机构更看重候选学者与现有团队的学术协同潜力,则可适当提高 γ_1 ;若更看重候选学者主动寻求机构变动的意愿,则可适当提高 γ_2 。在实际应用中,也可通过对历史引进案例的回归分析来数据驱动地确定 γ_1 和 γ_2 的最优取值。 $P_{recruit}$ 的值越高,表明目标学者与拟引进机构的学术关联越紧密且本身具有较强的流动意愿,被成功引进的可能性越大,可作为机构人才引进决策的定量参考依据。

[0151] 本发明实施例的第二方面,提供一种基于动态引用图谱的学术人才评价及引进预测系统,包括:

[0152] 元数据单元,用于从学术文献数据库中检索目标领域论文,提取论文的发表时间、作者集合、作者所属机构及国家集合作为时空元数据;

[0153] 构图单元,用于以论文为节点、以引用关系为有向边构建学术论文动态引用有向图,基于引用时间与当前观测时间的时间差确定各引用边的动态时间权重,并根据施引论文与被引论文在机构或国家维度上的归属关系确定各引用边的空间过滤条件;

[0154] 影响力单元,用于根据所述动态时间权重和空间过滤条件,针对目标学者筛选其作为作者且满足非自引约束的引用边集合,计算其在时间维度上的加权影响力分值和在机构或国家维度上的空间辐射力分值,并对所述时间分值与空间分值进行加权聚合,得到时空动态综合影响力测度值;

[0155] 预测单元,用于构建学者合作关系无向图,根据目标学者与拟引进机构的已有学者在所述合作无向图中的最短路径长度确定拓扑空间紧密度,并根据目标学者在历年所属机构上的分布熵值确定流动性指标,融合所述拓扑空间紧密度和流动性指标为引进可能性

预测指数。

[0156] 本发明实施例的第三方面,提供一种电子设备,包括:

[0157] 处理器;

[0158] 用于存储处理器可执行指令的存储器;

[0159] 其中,所述处理器被配置为调用所述存储器存储的指令,以执行前述所述的方法。

[0160] 本发明实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机程序指令,所述计算机程序指令被处理器执行时实现前述所述的方法。

[0161] 本发明可以是方法、装置、系统和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质,其上载有用于执行本发明的各个方面的计算机可读程序指令。

[0162] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

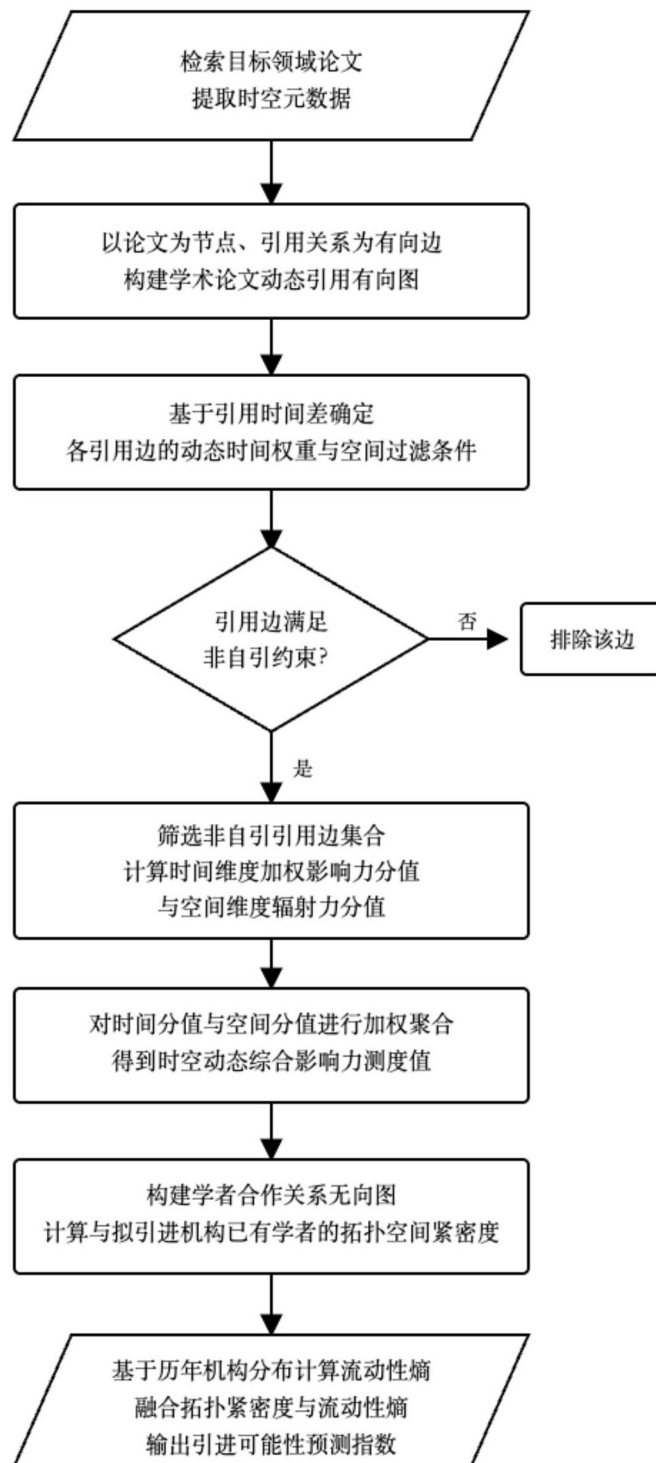


图1

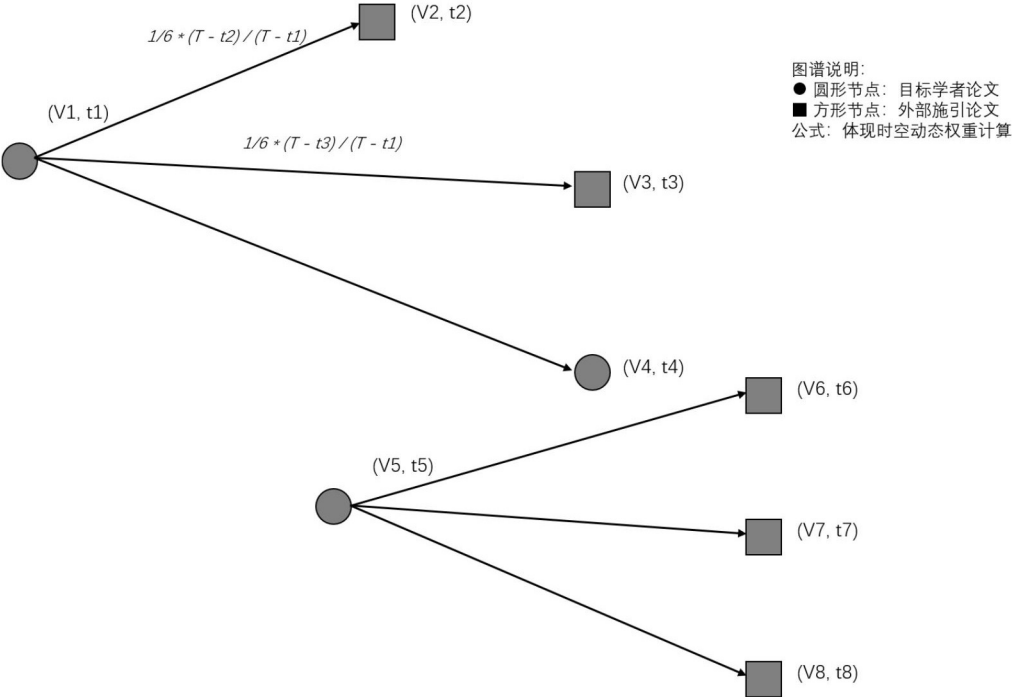


图2

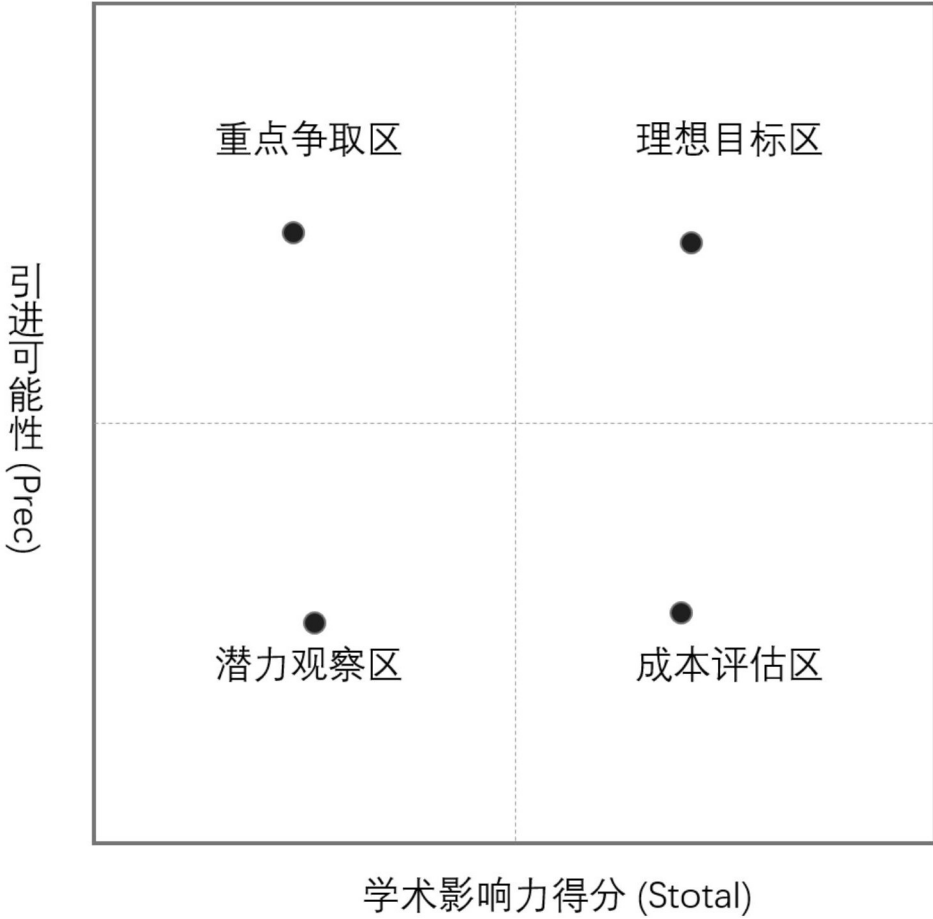


图3