

基于 DeepWalk 的在线教育平台智能学习资源推荐研究^①

王玲艳^{1,2} 李婵³ 张梦琪^{1,2}

¹ (福州大学信息管理研究所, 福建福州, 350108)

² (福州大学图书馆, 福建福州, 350108)

³ (福建农林大学公共管理学院, 福建福州, 350108)

摘要 [目的] 本文将蕴含特征表示能力的深度随机游走方法引入在线学习推荐领域, 挖掘学习者偏好特征从而进行智能学习资源推荐。[方法] 该方法利用学习者评分数据构造的二部图生成随机游走顶点序列, 并利用SkipGram语言模型获取图顶点的低维嵌入向量表示。然后结合协同过滤的思想, 使用学习者向量表示计算学习者间的相似度, 从而进行在线课程个性化推荐。[结果] 利用中国大学MOOC在线教育平台真实数据集将本文方法与基线方法进行对比, 实验表明本文提出的方法效果更优。[局限] 本文仅挖掘学习者对课程的用户偏好, 未来可考虑结合信任机制, 引入信任网络, 建模用户偏好模型, 获得更加准确的学习者特征表示。[结论] 深度游走模型可以分析学习者的行为偏好, 有效提升在线学习资源个性化推荐效果。

关键词 DeepWalk; SkipGram语言模型; 学习资源推荐

Research on Intelligent Learning Resource Recommendation for Online Education Platform Based on DeepWalk

Wang Lingyan^{1,2} Li Chan³ Zhang Mengqi^{1,2}

¹(Institute of Information Management, Fuzhou University, Fuzhou, Fujian, 350108, China)

²(Fuzhou University Library, Fuzhou, Fujian, 350108, China)

³(School of Public Management, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350108, China)

Abstract [Objective] In this paper, the DeepWalk method with feature representation ability is introduced

^①本文系福建省中青年教师科研项目(信息化专项)基金项目编号: JAT211018 和 JAT211012 的研究成果之一。

into the field of online learning recommendation, so as to mine learners' personal preferences and make intelligent recommendations. [**Methods**] According to the interactive data between learners and online courses, a bipartite graph model was constructed. Random walks are carried out on bipartite graphs to generate vertex sequences, and the neighbor structure model of graph vertices is captured by using SkipGram language model to learn the low-dimensional embedding vector representation of graph vertices. Calculate the similarity between learners' vertices, and use collaborative filtering idea to make personalized recommendation of online courses. [**Results**] The real data set of MOOC online education platform of China University was used to compare the proposed method with the baseline method, and the experiment showed that the proposed method was more effective. [**Limitations**] This paper only explores learners' preferences for courses. In the future, we can consider introducing trust network to model users' preferences in combination with trust mechanism to obtain a more accurate representation of learners' characteristics. [**Conclusions**] The DeepWalk model can analyze learners' behavior preferences and effectively improve the personalized recommendation effect of online learning resources.

Keywords DeepWalk; SkipGram Language Model; Learning Resource Recommendation

1 引言

教育信息化的推进使在线教育平台蓬勃发展, 最具代表性的有爱课程、中国大学 MOOC、Coursera 等^[1]。这些在线教育平台提供的在线学习方式是一种重要的学习交互方式。然而, 各式各样的课程资源给学习者带来了信息过载问题, 尤其突出的是平台采用类似的教学模式, 学习者要耗费大量时间和精力才能找到符合自己认知能力、学习风格和兴趣爱好的课程。所以, 深入了解学习者的行为偏好、学习效果和能力等, 并提高个性化学习课程的推荐效果, 已经成为构建在线学习课程系统的挑战。

图推荐模型能够捕捉学习者和在线课程的特征以及两者间的拓扑关系, 对图结构数据有很好的建模效果。其中, 应用于图推荐领域的图推荐模型一般可分为两大类: 图游走模型和图卷积模型。随机游走方法就是图游走模型在推荐系统的重要研究, 但是, 传统的图游走模型难以将蕴含的复杂拓扑关系建模图结构中,

深度随机游走方法为解决图结构中蕴含的复杂拓扑关系提供了很好的解决思路, 也逐渐成为学者们的一个重要研究方向。

推荐系统是指导用户发现满足自身需求信息的工具, 能够有效解决信息过载问题。目前, 常用的推荐技术包括基于协同过滤的、基于知识的、基于模型的推荐方法等。这些推荐方法已在多个领域得到广泛应用, 例如: 淘宝商品推荐、今日头条新闻推荐、亚马逊图书推荐、网易云音乐推荐、Netflix 电影推荐和 Coursera 课程推荐等。推荐系统凭借用户属性特征和在线浏览数据对其进行建模, 主动挖掘潜在用户及其感兴趣的产品和项目, 有效地解决信息过载问题, 从而提高用户满意度。

2 相关研究

2.1 在线学习资源研究现状

在线学习资源领域使用比较多的推荐方法是协同过滤推荐方法, 它已被较多学者作为重要研究方向进行研究。刘芳等^[2]对融入静

态和动态特征改进协同过滤方法,在此基础上改进学习者模型提出一种在线学习资源推荐方法。赵继春等^[3]基于知识图谱技术,提出一种结合标签生成的领域知识图谱推荐方法。Tahir S 等^[4]为课程设计提供了一个基于机器学习和协同过滤的情境感知的推荐系统,可以方便地搜索和访问多种语义相关的视点。Mbipom B 等^[5]利用结构化的教材集合作为识别重要领域概念的指南,创建一种系统过滤方法,可以自动构建学习领域的背景知识。Ibrahim T S 等^[6]提出了基于雾计算技术的协同过滤推荐系统,雾计算技术可以丰富基于遥感的 E-Learning 技术,它可以成功地用于提升在线学习课程推荐环境的性能。但是,这些研究都侧重于使用协同过滤方法进行个性化推荐,这会造成数据稀疏性问题。因为在大多数情况下,由于数据量不足,协同过滤方法很难准确地找到相关的学习者和在线课程之间的关联性。

2.2 深度随机游走研究现状

深度随机游走方法(DeepWalk)是一种基于随机游走的节点嵌入技术,用于将图结构中的节点转换为低维向量表示。即 DeepWalk 可以将节点的高维结构嵌入到低维空间中,并保留节点之间的结构关系^[7]。该技术可以应用于各种图形数据分析任务,如:社交网络分析、推荐系统、自然语言处理等^[8]。

例如, Yang 等^[9]提出了一种新的社交和序列感知下的一个 POI 推荐模型。利用异构网络上改进的 DeepWalk 来学习更好的 POI 表示,在一定程度上解决冷启动问题和可解释性问题,进行更好的推荐。Qu 等^[10]提出了一种基于深度图的神经网络框架,用

于社交媒体平台上的好友推荐。通过使用 DeepWalk 方法获得了用户特征的全面表示,进而在考虑用户关系和个人属性的情况下生成更合理的推荐列表。Cai 等^[11]由于链接预测在社交网络分析、生物信息学和个性化推荐中的广泛应用,把深度随机游走方法引入链接预测。将 DeepWalk 得到的距离指数与传统相似性指数相结合,提出深度亲和性指数提升推荐效果。刘靖凯^[12]基于 DeepWalk 方法学习用户的嵌入向量,并基于学习到的向量表示计算用户相似度,进行准确的个性化项目推荐。冯曦等^[13]使用深度游走模型改进标签传播社区发现,并利用深度随机游走方式解决传统标签传播精准率低下的问题。然而,目前很少有研究在在线学习资源推荐领域应用深度随机游走方法,以提升在线学习资源推荐性能。

综上所述,本文面向在线教育平台提出了一种基于 DeepWalk 的智能学习资源推荐方法,以解决传统方法面临的数据稀疏问题,并进一步提升在线学习资源推荐性能。

3 基于深度随机游走的学习资源推荐模型

本文提出的基于深度随机游走的学习资源推荐模型的框架,如图 1 所示。首先利用中国大学 MOOC 在线学习课程的真实数据集构建图游走模型,得到用于训练的顶点序列;其次,使用 SkipGram 语义模型学习顶点序列的深层语义信息,并得到每个图顶点的低维嵌入向量表示;最后,根据学习者的向量表示计算学习者间的余弦相似度,并结合协同过滤的思想进行个性化推荐。

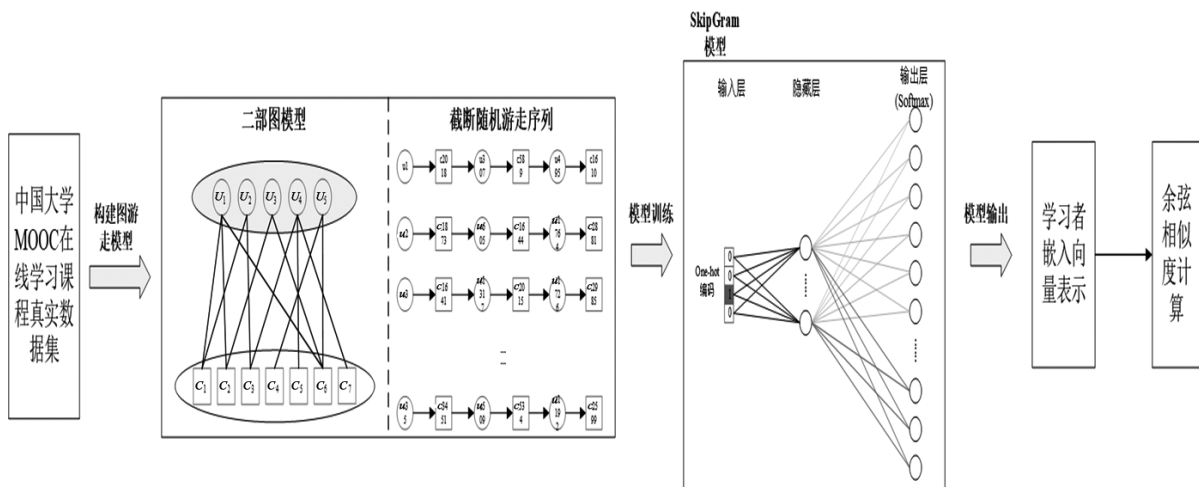


图1 基于深度随机游走的学习资源推荐模型的架构图

3.1 基于二部图模型的截断随机游走

主要介绍基于学习者-在线课程二部图的顶点序列生成过程。首先,使用 $U=\{u_1, u_2, \dots, u_n\}$ 表示学习者集合,其中, u_i 代表第 i 个学习

者;使用 $C=\{c_1, c_2, \dots, c_m\}$ 表示在线课程集合,其中, c_j 代表第 j 门在线课程。其次,利用评分数据构造学习者-在线课程二部图模型 $G=(V, E)$, 具体如图2所示。

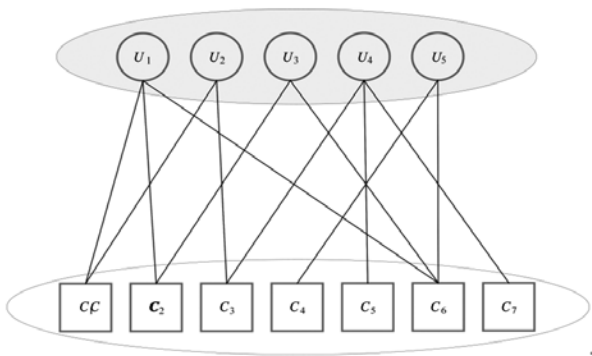


图2 学习者-在线课程二部图

其中, $V=U \cup C$, 由学习者顶点集合 U 和在线课程顶点集合 C 组成。对于图中每条边 $(u_i, c_j) \in E$, 表示学习者 u_i 对在线课程 c_j 有过评分行为。此外, 将学习者 u_i 对在线课程 c_j 的评分数值作为边 (u_i, c_j) 的权重。

接着, 对构建的二部图模型进行截断随机游走, 进而探索顶点和边之间的关联性, 以捕捉学习者-在线课程二部图模型中的结构信息。具体而言, 从二部图中的任意一个顶点开始, 然后从此顶点的邻居列表中随机选择一个

相邻顶点进行移动, 直到顶点序列达到给定截断长度为止, 具体如公式(1)所示:

$$W_{vi} = [W_{vi}^1, W_{vi}^2, \dots, W_{vi}^K] \quad \text{公式(1)}$$

其中, 该顶点序列以 v_i 作为源顶点, 截断随机游走长度设置为 K 。例如, 在图2中的二部图模型上, 选择顶点 U_1 作为初始顶点, 则一个长度固定为5的游走序列可以为 $U_1 \rightarrow C_1 \rightarrow U_2 \rightarrow C_3 \rightarrow U_4 \rightarrow C_5$ 。其中, 初始顶点 U_1 分别与顶点 C_1 、 C_2 和 C_6 存在连接, 则顶点 U_1 随机选取下一相邻顶点的概率由加

权平均法进行计算, 具体如下式所示:

$$P(v_j | v_i) = \begin{cases} \frac{M_{ij}}{\sum_{j \in N_+(v_i)} M_{ij}}, & j \in N_+(v_i) \\ 0, & v_i \text{ 与 } v_j \text{ 不存在连接} \end{cases} \quad \text{公式 (2)}$$

其中, $N_+(v_i)$ 代表顶点 v_i 的所有边的顶点集合, M_{ij} 代表顶点 v_i 与顶点 v_j 之间相连边的权重。

3.2 SkipGram 语言模型

主要介绍利用 SkipGram 语言模型学习图顶点的低维嵌入向量表示。由于在随机游走顶点序列中, 顶点的出现频率与自然语言处理中单词的出现频率有相似的幂律分布特性。这种分布特性是指只有少数顶点在序列中频繁出现, 而大部分顶点只会偶尔出现^[7]。因此, 可以类比自然语言的建模方法对图结构进行建

模, 将随机游走序列视为一个虚拟的文本语料库, 其中二部图的顶点被视作单词。基于此, 可以应用自然语言处理方法的 SkipGram 模型来解决图结构中的顶点嵌入问题。

SkipGram 模型作为一种基于神经网络的词向量模型, 主要用于学习词语的分布式表示。它的基本思想是通过预测一个词语的周围词语来学习该词语的向量表示。在深度随机游走方法中, SkipGram 模型的目标是对于给定一个顶点 v_i , 使得它的邻居顶点在随机游走中出现的概率最大化, 如公式 (3) 所示:

$$\min -\log \Pr(\{v_{i-w}, \dots, v_{i-1}, v_{i+1}, \dots, v_{i+w}\} | v_i) \quad \text{公式 (3)}$$

其中, w 表示目标顶点 v_i 的上下文窗口大小。

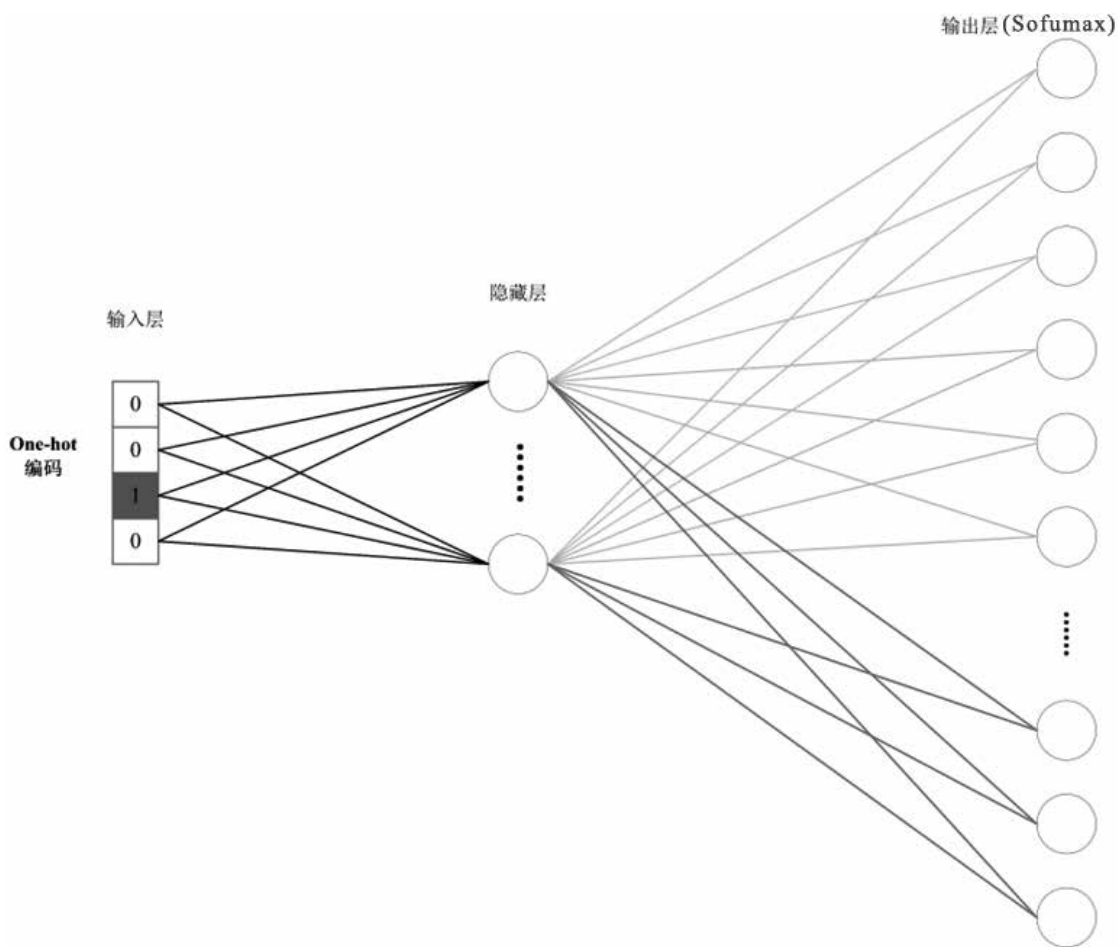


图 3 Skip-Gram 结构图

如图 3 所示, SkipGram 模型的神经网络结构包括一个输入层、一个隐藏层和一个输出层。即, SkipGram 模型的输入层将一个顶点表示为一个 $1 \times |V|$ 的 one-hot 向量 \mathbf{x} , 其中 $|V|$ 表示顶点个数。然后, 输入层通过查找表 (Lookup Table) 将这个 one-hot 向量 \mathbf{u} 转换为一个 d 维的嵌入向量。

其次, 隐藏层包含所有顶点的嵌入向量, 它们被组合成一个大小为 $|V| \times d$ 的向量矩阵 $W \in R^{|V| \times d}$ 。在这个矩阵中, 每一行代表一个顶点的嵌入向量, 每一列代表顶点嵌入向量的一个维度。通过将输入顶点的嵌入向量乘以该矩阵, 则得到了隐藏层的输出:

$$\mathbf{h} = W^T \mathbf{u} \quad \text{公式 (4)}$$

其中, \mathbf{h} 是一个大小为 $|V|$ 的隐藏向量, 其包含了所有顶点与输入顶点之间的相似度得分。通过选择相似度得分最高的顶点, SkipGram 模型可以学习到顶点的潜在信息。

最后, 输出层主要是一个 Softmax 分类器, 用于预测在给定顶点的情况下, 每个上下文顶点出现的概率。具体地, 输出层将从隐藏层中接收到的隐藏向量 \mathbf{h} 作为输入, 并通过 Softmax 回归函数计算每个上下文顶点的条件概率, 具体如下所示:

$$p(w_i | w_j) = \frac{e^{u_i^T h}}{\sum_{j=1}^{|V|} e^{u_j^T h}} \quad \text{公式 (5)}$$

其中, u_i 是顶点 v_i 的嵌入向量。

在训练过程中, SkipGram 模型使用损失函数公式 (6) 来计算预测分布与真实分布之间的差异。接着, 通过使用批量随机梯度下降方法来更新权重, 并在每个批次上运行反向传播。当迭代结束, 模型可以学习到每个顶点的向量表示。

$$Loss = -\frac{1}{K} \sum_{t=1}^K \sum_{-w \leq j \leq w, j \neq 0} \log \Pr(v_{t+j} | v_t) \quad \text{公式 (6)}$$

3.3 基于顶点嵌入向量的在线课程推荐

由于学习者嵌入向量能够体现其对在线课程交互行为的相关信息, 因此具有类似偏好的两个学习者在向量空间中距离较近。本文根据这一特点, 结合协同过滤的思想, 利用学习者的嵌入向量表示计算相应的余弦相似度从而进行个性化推荐。首先, 假设学习者 A 的嵌入向量为 $A = (A_1, A_2, \dots, A_d)$, 学习者 B 的嵌入向量为 $B = (B_1, B_2, \dots, B_d)$, 则学习者 A 和 B 之间的余弦相似度计算公式如下所示:

$$S_{A,B} = \cos\theta = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^d (A_i \times B_i)}{\sqrt{\sum_{i=1}^d A_i^2} \times \sqrt{\sum_{i=1}^d B_i^2}} \quad \text{公式 (7)}$$

即余弦相似度为嵌入向量 A 和 B 的点积除以两个嵌入向量各自的 L_2 范数的乘积。因此, 当两个顶点的余弦值越大时, 表示这两个顶点的相似度越高。

然后, 利用基于用户的协同过滤思想为学习者进行资源推荐。首先, 使用计算目标学习者和其他学习者之间的相似度, 接着寻找与目标学习者兴趣偏好最相似的学习者集合; 然后, 将兴趣偏好相似的学习者观看的但目标学习者没有观看过的课程资源, 按照目标学习者对其偏好程度进行推荐。其中, 目标学习者 u 对在线课程 c 的偏好度计算公式如下:

$$p(u, c) = \sum_{v \in S(u, K) \cap N(c)} S_{uv} r_{vc} \quad \text{公式 (8)}$$

其中, $S(u, K)$ 代表和目标学习者 u 最相似的前 K 个学习者的集合, $N(c)$ 代表观看过在线课程 c 的学习者的集合, s_{uv} 表示学习者 u 和学习者 v 的相似度大小, r_{vc} 表示学习者 v 对在线课程 c 的评分数值。

4 实验结果及分析

4.1 实验环境与数据准备

本文研究基于 Python3.6 的实验环境,

SkipGram 模型基于 Tensorflow2.0.0 框架构建, 相关数据预处理使用 Numpy、Pandas 模块。接着, 本文从在线教育平台中国大学 MOOC 收集的学习者与在线课程的相关信息构成数据集。其中, 主要使用数据集中的 “UserID” “CourseID” “Rating” 三个字段的相关数据, 然后, 将实验数据集随机划分成训练集和测试集, 其中划分比例是 8 : 2。

此外, 本文使用精准率 (Precision)、召回率 (Recall) 和覆盖率 (Coverage) 三个评测指标, 以评估本文提出的推荐方法的具体性能。

精准率表示推荐列表中学习者喜欢的在线课程占列表总数的比例, 其计算公式如下:

$$Precision@K = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} R(u)} \quad \text{公式 (9)}$$

召回率表示推荐列表中学习者喜欢的在线

课程占测试集上学习者所有喜欢在线课程总数的比例, 其计算公式如下:

$$Recall@K = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} T(u)} \quad \text{公式 (10)}$$

覆盖率表示推荐的在线课程占总课程数的比例, 其计算公式如下:

$$Coverage@K = \frac{|\bigcup_{u \in U} R(u)|}{|C|} \quad \text{公式 (11)}$$

其中, K 表示在推荐场景中选取推荐列表中排名前 K 个项目, $R(u)$ 表示算法为学习者 u 计算得出的在线课程推荐列表, $T(u)$ 表示测试集中学习者 u 的在线课程观看列表, $|C|$ 表示在线课程的总数目。

4.2 参数设置与基线方法

本小节主要介绍本文提出的基于 DeepWalk 的智能学习资源推荐方法的具体参数设置, 如表 1 所示。

表 1 基于 DeepWalk 的智能学习资源推荐方法的参数设置

参数名	描述	参数值
t	游走路径长度	90
γ	游走路径数量	80
d	嵌入维度	200
w	窗口大小	30
$iter$	迭代次数	20

其中, t 表示进行随机游走的路径长度; γ 表示以每个顶点作为源顶点进行游走的路径数量; d 代表二部图中每个顶点的嵌入向量的维度大小; w 代表 SkipGram 语言模型中所使用的窗口大小; $iter$ 表示模型训练的迭代总次数。

此外, 本文选取两种经典的推荐方法作为基准方法进行对比实验, 从而验证本文所提方法的有效性, 具体描述如下:

ItemCF: 基于物品的协同过滤方法。主要思想是根据用户过去的行为数据, 计算不同

物品之间的相似度, 从而为用户推荐相似的物品。

UserCF: 基于用户的协同过滤方法。主要思想是根据用户过去的行为数据, 寻找与目标用户兴趣相似的其他用户集合, 从而进行相关推荐。

4.3 实验结果及分析讨论

本文将基于 DeepWalk 的智能学习资源推荐方法 (DRW)、ItemCF 和 UserCF 方法在 MOOC 评分数据集上进行多次实验, 然后取

评测指标 Precision@10、Recall@10 和 Coverage@10 的平均值作为最终的实验结果，具体如表 2

表 2 本文方法与基线方法在 TOP-10 下的实验结果

方法	准确率	召回率	覆盖率
ItemCF	0.3015	0.0901	0.1565
UserCF	0.3059	0.0915	0.2035
DRW	0.3201	0.0957	0.2152

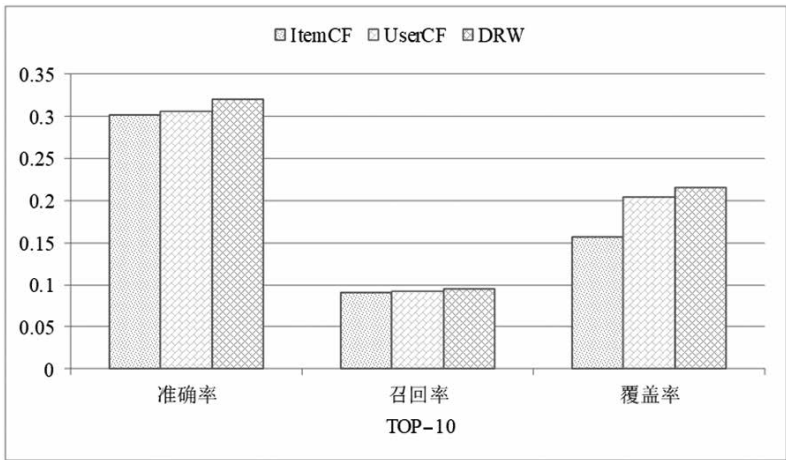


图 4 本文方法与基线方法在 TOP-10 下的实验结果

由表 2 以及图 4 可知，本文提出的推荐方法在准确率、召回率和覆盖率三个评测指标上均优于基线方法。具体而言，本文方法在准确率上比 ItemCF 和 UserCF 方法分别提高 5.8% 和 4.4%；在召回率上比 ItemCF 和 UserCF 方法分别提高 5.9% 和 4.3%；在覆盖率上比 ItemCF 和 UserCF 方法分别提高 27.3% 和 5.4%。这是因为本文提出的基于 DeepWalk 的推荐方法通过深度学习模型学习图顶点的邻居结构，从而更准确地捕捉顶点之间的相似性以进行在线课程的个性化推荐。

综上所述，在 TOP-10 场景下，本文提出的推荐方法在三个评测指标上均表现较为优秀，从而证明了基于 DeepWalk 的在线教育平台智能学习资源推荐方法的有效性。

5 结语

本文使用中国大学 MOOC 在线学习课程的真实数据集构建图游走模型并在图上进行截断随机游走，得到用于训练的顶点序列；其次，利用 SkipGram 语义模型学习顶点序列的深层语义信息，并获得每个图顶点的低维嵌入向量表示；然后，结合协同过滤的思想，使用学习者的向量表示计算余弦相似度，从而进行个性化推荐。最终，基于在线教育平台中国大学 MOOC 数据集，将本文所提方法与基线方法进行对比实验。实验结果表明，本文方法在准确率、召回率和覆盖率三个评测指标上均有较优的表现，从而验证了所提出方法的有效性。

在未来研究中，将考虑信任机制在模型构造中引入信任网络，以细粒度的方式建模

用户偏好, 进而得到性能更加精准的用户特征表示。

参考文献

- [1] 董永峰, 王雅琮, 董瑶, 等. 在线学习资源推荐综述 [J/OL]. 计算机应用: 1-10 [2023-02-25]. <http://kns.cnki.net/kcms/detail/51.1307.tp.20221107.1453.004.html>.
- [2] 刘芳, 田枫, 李欣, 等. 融入学习者模型在线学习资源协同过滤推荐方法 [J]. 智能系统学报, 2021, 16(6): 1117-1125.
- [3] 赵继春, 孙素芬, 郭建鑫, 等. 农业在线学习资源知识图谱构建与推荐技术研究 [J]. 计算机应用与软件, 2022, 39(8): 69-75.
- [4] TAHIR S, HAFEEZ Y, ABBAS M A, et al. Smart Learning Objects Retrieval for E-Learning with Contextual Recommendation based on Collaborative Filtering [J]. Education and Information Technologies, 2022: 1-38.
- [5] MBIPOM B, CRAW S, MASSIE S. Improving E-Learning Recommendation by Using Background Knowledge [J]. Expert Systems, 2021, 38(7): e12265.
- [6] IBRAHIM T S, SALEH A I, ELGAML N, et al. A Fog Based Recommendation System for Promoting the Performance of E-Learning Environments [J]. Computers & Electrical Engineering, 2020, 87: 106791.
- [7] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online Learning of Social Representations [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 701-710.
- [8] 崔冬, 王明, 李刚, 等. 基于多级深度特征与随机游走的显著性检测 [J]. 华南理工大学学报 (自然科学版), 2020, 48(8): 49-55.
- [9] YANG K, ZHU J. Next Poi Recommendation via Graph Embedding Representation from H-DeepWalk on Hybrid Network [J]. IEEE Access, 2019, 7: 171105-171113.
- [10] QU Z, LI B, WANG X, et al. An Efficient Recommendation Framework on Social Media Platforms Based on Deep Learning [C] // 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2018: 599-602.
- [11] CAI L, WANG J, HE T, et al. A Novel Link Prediction Algorithm Based on DeepWalk and Clustering Method [C] // Journal of Physics: Conference Series. IOP Publishing, 2018, 1069(1): 012131.
- [12] 刘靖凯. 基于深度随机游走的协同过滤推荐算法 [J]. 科学技术创新, 2021(6): 93-94.
- [13] 冯曦, 朱福喜, 刘世超. 基于深度游走模型的标签传播社区发现算法 [J]. 计算机工程, 2018, 44(3): 220-225.
- (通讯作者: 李婵, 联系电话: 18046046096, E-mail: shandongweifanglch@163.com)

基于智能合约的知识产权共享保护体系构建研究——以在线社区为例^①

潘星宇¹ 蔡均益²

¹ (福州大学图书馆, 福建福州, 350000)

² (漳州城市职业学院图书馆, 福建漳州, 363000)

摘 要 [目的] 本文目的在于结合智能合约与精益管理理论, 分析区块链智能合约技术对在线社区知识产权共享保护的优化作用, 并提出相应发展建议。[方法] 基于网络调查法及文献调研法, 从当前在线社区知识产权共享保护存在的问题入手, 深挖其症结所在, 有针对性地提出了基于区块链智能合约的在线社区知识产权共享保护精益管理体系。[结果] 提出了基于区块链智能合约技术的在线社区知识产权共享保护链(OCPC), 并构建了在线社区用户知识产权共享保护精益管理体系, 具体分析了其中主要功能模块的运行逻辑。[局限] 如果能够将区块链、智能合约等技术应用到在线社区知识产权共享保护中, 将会大大提高效率、降低成本, 但同时, 由于技术上的缺陷, 可能使得智能合同的实施会产生不合法的行为。[结论] 本研究旨在为解决当前在线社区存在的知识产权侵权问题提供解决方案, 助力在线社区健康发展。

关键词 区块链; 智能合约; 在线社区; 数字作品; 知识产权保护

Research on the Construction of Intellectual Property Sharing and Protection System Based on Smart Contracts – Taking Online Communities as an Example

Pan Xingyu¹ Cai Junyi²

¹(Fuzhou University Library, Fuzhou, Fujian, 350000, China)

²(Zhangzhou City Vocational College Library, Zhangzhou, Fujian, 362000, China)

Abstract [Purpose] The purpose of this paper is to analyze the optimization effect of blockchain smart

①本文系“2022年度福建省高校教育信息化项目”(课题编号:FJGX22004)研究成果之一。

contract technology on online community intellectual property sharing protection, and further propose corresponding development suggestions. [**Method**] Based on the methods of network survey and literature research, this paper starts with the problems existing in online community intellectual property sharing protection, and digs deep into the root causes. A lean management system for online community knowledge sharing and intellectual property protection based on blockchain smart contracts is proposed accordingly. [**Result**] The Online Community Intellectual Property Sharing Protection Chain (OCPC) based on blockchain smart contract technology is proposed, and a lean management system for protecting the intellectual property sharing of online community users is constructed, with the running logic of the main functional modules analyzed in detail. [**Limitation**] The application of technologies such as blockchain and smart contracts to online community intellectual property sharing protection can greatly improve efficiency and reduce costs, but at the same time, illegal behaviors may arise due to technical deficiencies in implementing smart contracts. [**Conclusion**] This study aims to provide solutions for solving the problem of intellectual property infringement in current online communities and to promote the healthy development of online communities.

Keywords Blockchain; Smart Contracts; Online Communities; Digital Works; Intellectual Property Protection

1 引言

据统计,截至 2021 年 12 月,我国网民规模已达 10.32 亿^[1],与此同时,随着移动互联网的高速发展,人们的信息交流模式日益改变,传统的知识共享模式有所转变,交流过程延伸至网络空间,在此背景下,基于用户生成内容的在线社区迅速发展。根据国家版权局发布的《中国网络版权产业发展报告(2021)》显示,近年来我国网络版权产业市场规模增长迅猛(图 1),2020 年我国网络版权产业市场规模首次

突破一万亿元^[2]。在线社区作为内容产出社区,不可避免地存在知识产权确权、侵权及维权方面的问题,降低用户内容产出的意愿,同时也制约了在线社区的发展。如何确定用户产出内容的归属权、如何保证用户知识产权不受侵犯等问题成为在线社区发展亟待解决的问题。

传统的知识产权共享保护依赖第三方中介机构的信用背书,通常受到严格的限制,其保护流程往往十分复杂且信息不透明,使得需求链的两端难以实现有效对接。区块链中的智能合约技术具备自动化、透明公开、智能化等特

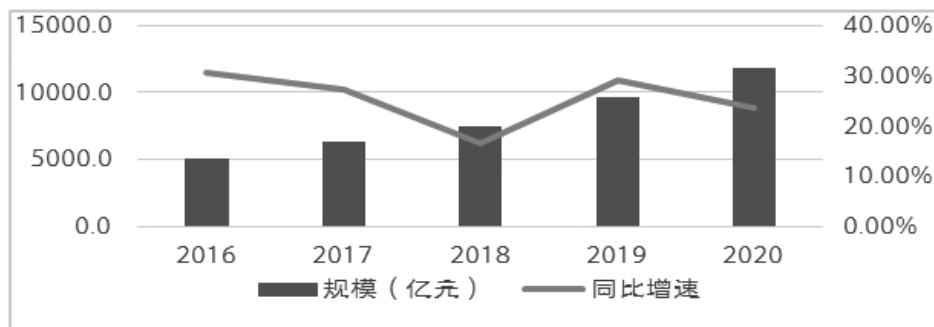


图 1 2016—2020 年中国网络版权产业市场规模

点,打破知识产权共享保护的传统限制,重构知识产权共享保护体系,为解决数字知识产权确权难、侵权问题等提供了解决方案,助力在线社区健康发展。

2 相关研究综述

2.1 在线社区的内涵

在线社区在国内外许多文献中又被称为

“Online community” “Cyberspace community” “Computer-mediated community” “Virtual community” “网上社区” “虚拟社区” “网络社区” “电子社区”等。在概念发展过程中,诸多学者从不同视角给出了在线社区的定义,表1梳理了代表性研究。

表1 在线社区定义梳理

研究视角	学者	定义	期刊名称
社会学视角	Plant ^[3] (2004)	在线社区是一种群体或组织,其中的个体因有共同的兴趣领域而通过电子媒介,暂时或永久地汇聚在一起互相交流	Technology in Society
	徐小龙等 ^[4] (2007)	在线社区是指在互联网上,人们通过网络进行沟通,以满足某种需要而产生的群体关系总和	外国经济与管理
管理学视角	Preece ^[5] (2001)	在线社区是人们为了学习、寻求陪伴或获得支持而聚集为一个群体的虚拟社交空间	Behaviour & Information Technology
	Gupta等 ^[6] (2004)	在线社区是指为一群具有共同目的的陌生人,以网络空间为载体,提供互动和沟通,建立关系,分享知识,享受乐趣等活动	Proceedings of the Tenth America's Conference on Information System
情报学视角	Lee ^[7] (2002)	在线空间是指在技术的支持下,以参与者的沟通和互动为中心建立社会关系的网络空间	Hawaii International Conference on System Sciences
信任视角	Ridings ^[8] (2002)	在线社区是有着相同利益与经历的一群人,通过某些机制在网络上长时间、频繁地沟通互动而形成的一种群体	Journal of Strategic Information Systems

虽然不同学者对在线社区给出的定义不尽相同,但相关定义的共同特征是都包含群体、虚拟空间、沟通和知识等要素。本研究将在线社区的种种概念加以融合,将其主要内涵定义为:用户可以自由分享知识,获取知识的内容导向虚拟社交空间。

2.2 面向知识产权的区块链智能合约研究

目前,国内外学者已经在知识产权管理领域对智能合约的应用进行了诸多探索。Sung讨论了智能合约在版权认证和保存方面的应用,并指出其相关技术的发展将会对国家现行

版权保护制度产生颠覆性影响^[9]。Yu等认为,区块链可以为知识产权注册管理带来历史性变革,在源头上筑起有效遏制知识产权违法行为的金盾^[10]。Martin以Monegraph为例,讨论了区块链智能合约技术在数字艺术品中的应用^[11]。针对互联网上存在的数字版权侵权问题,Xu等学者提出了一种基于区块链技术的网络媒体数字版权管理方案,管理版权、交易和用户行为^[12]。国内学者赵丰描述了数字知识资产的应用管理,并分析了国内外多个企业的相关应用^[13]。在此基础上,郑阳等人讨论

了区块链技术在数字版权方面的优势和问题, 并指出需要以发展眼光、包容心态和辩证思维审视其发展和应用^[14]。此外, 宁梦月等学者建立了基于智能合约的多角色、多链数字版权保护方案, 可实现产权登记、转让和维权^[15]。就新闻版权管理而言, 林爱珺等学者从版权登记、版权交易和版权维护三个角度分析了区块链智能合约在版权管理和保护机制上的独特优势^[16]。总之, 智能合约在知识产权管理方面的研究趋向于探索更广泛的应用, 需要更多结合实际工程情境的研究以发掘其应用深度。

2.3 研究述评

通过文献调研发现, 国内将对智能合约技术应用于知识产权的保护已有一定数量的研究, 并取得了一定的研究成果。现有文献主要分为探讨智能合约技术与知识产权共享保护结合的可行性、前景及存在的困境三方面。将智能合约技术运用至知识产权的保护方面, 能够解决当前知识产权共享保护存在的诸多问题,

值得深入研究。然而, 鲜有研究将智能合约技术应用到在线社区知识产权共享保护中, 在线社区作为内容高产出社区, 所面临的知识产权共享保护问题十分严峻, 因此考虑将智能合约技术嵌入在线社区的知识产权共享保护中具有一定的现实意义。再次, 现有研究中, 大部分学者还只是将研究停留在理论层面, 如何具体实现在线社区与智能合约知识产权共享保护相结合, 包括文字识别技术、合约设计、主要功能实现等问题都有待进一步解决。

3 在线社区知识产权共享存在问题分析

在线社区蓬勃发展, 成了许多优秀数字作品的摇篮, 与此同时, 在线社区知识产权共享保护也衍生出诸多问题, 包括在线社区用户知识产权归属、知识产权取证难、知识产权交易、知识产权损害难以评估等问题, 具体如图 2 所示。

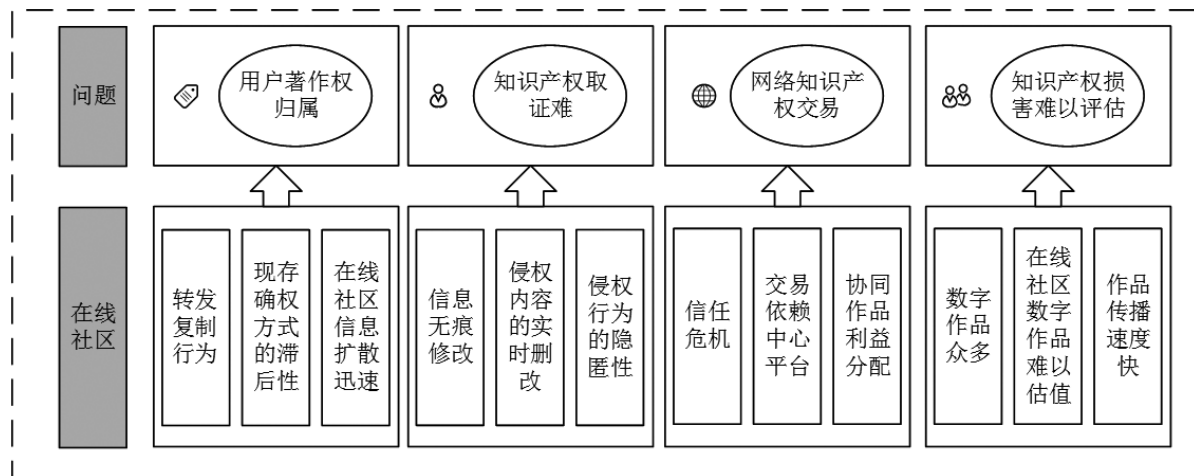


图 2 在线社区知识产权共享保护问题概览

3.1 在线社区用户数字知识产权归属问题

在线社区知识产权是指知识产权人对受知识产权法保护的作品在网络环境下所享有的知识产权权利, 其含义主要包括两个层面: 一是

指作品被上传至互联网时知识产权人对作品享有的“信息网络传播权”; 二是指作品上传至互联网后知识产权人对该数字化作品所享有的衍生权利, 如复制权、署名权等^[17]。随着互

联网的蓬勃发展,数字作品的传播变得十分容易,然而随着大量转发、复制行为,数字作品的原作者往往难以确定,在线社区用户知识产权的归属问题成为一大难题。我国现行的数字版权登记,一般是由国家版权局和省级版权中心完成,其流程包括申请受理、审查决定、颁证送达,历时较长且费用昂贵,与互联网在线社区上迅速扩散的信息相比,这种中心化确权方式存在着严重的滞后性。

3.2 在线社区用户数字知识产权取证难问题

互联网技术为信息的无痕修改提供了极大的便利,电子化存储使得产权“洗白”、作品删改等侵权行为频发,且在线社区数字知识产权侵权行为存在很强的隐匿性,跨社区的侵权行为难以取证。此外,在线社区内容的实时修改功能使得一旦侵权行为发生,侵权人可以立刻删除侵权内容,导致侵权行为难以认定。2016年,新浪“知乎大神”在未经授权的情况下,擅自复制、传播知乎上的问题及回复,发表含侵权内容的微博数达一万余条,此后他被知乎告上了法庭,指控其侵犯了用户的信息网络传播权和署名权。该案由于取证难度大、跨平台等问题,历时两个月才完成取证和准备工作^[18]。

3.3 在线社区用户数字知识产权交易问题

当前,类似抖音、微博等内容生产的在线社区,其数字知识产权交易往往依赖于中心化平台。例如抖音原创作品的收益来自作品的点击量,然而这种依赖中心平台的交易方式往往易引发“信任危机”。此外,对于协同作品,也存在着如何确定多个作者的创作贡献与收益分配的比例问题,从而限制了在线社区的知识产权交易效率。

3.4 在线社区用户数字知识产权损害难以评估问题

数字知识产权是一种基于知识价值的评估

机制,它突出了知识的未来收益所带来的价值,因此要求利用已有的条件对其未来的市场价值进行估算,以便进行知识产权侵权时的损害赔偿^[19]。由于在线社区数字作品众多,且大部分的网络作品都没有经过市场的检验,又缺乏线下实体产品,因此在线社区中的数字作品缺乏可信赖的价值评估机制,难以量化侵权时的赔偿金额。此外,由于在线社区基于互联网技术,其传播速度极快,因此数字知识产权侵权行为一旦发生,其给知识产权人造成的损失是难以估算的。倘若侵权赔偿过低,难以约束侵权人的侵权行为,就会造成侵权行为难以制止的恶性循环。

4 在线社区知识产权共享保护精益管理体系构建

4.1 精益管理理论概述

精益管理(Lean Management)是以消除浪费、提高价值流和提高质量为核心的管理理论,通过不断地改善过程、降低成本、提高质量和快速响应客户需求来提高企业竞争力^[20]。将在线社区知识产权共享保护与精益管理理论相结合,能够提高企业的竞争力和可持续发展能力。具体而言,其适配性体现在以下几个方面:

(1) 优化数字作品确权流程。基于精益管理理论,可以对知识产权共享保护系统的各个流程进行优化和改进,发现当前数字作品确权流程中的瓶颈和浪费,进一步减少不必要的环节和步骤,提高效率和效益。

(2) 强化反馈机制。精益管理理论强调反馈和学习,帮助企业不断改进和提高。在线社区知识产权共享保护系统以区块链技术为基础,可以通过反馈机制获取用户的反馈和建议,不断改进和提高系统的性能和功能。

(3) 持续改善。精益管理的核心是持续改善,与在线社区知识产权共享保护体系的特点相契合。在智能合约技术的指导下,通过自动化的改善活动来不断优化系统的性能和功能,提高系统的可靠性、安全性和用户体验,从而更好地满足用户和市场的需求。

4.2 基于区块链的在线社区知识产权共享保护体系技术框架

本研究将区块链技术、智能合约技术、信息加密技术及精益管理理论运用于在线社区知识产权共享保护中,构建基于区块链智能合约技术的在线社区知识产权共享保护链(Online community copyright protection chain based

on blockchain smart contract technology, OCPC)。区块链技术的去中心化、不可篡改性等特点有助于在线社区知识产权确权,同时解决了数字知识产权取证难问题;“智能合约”保证了合约的履行不受主观个人因素的影响,有利于利益分配及作品共享交易;信息加密技术是利用密码学算法对个人隐私、商业秘密等进行加密,以确保数据的传输和存储的安全性。本研究通过分析在线社区用户数字知识产权存在的侵权问题,结合在线社区中各个主体的交互过程,将上述技术嵌入在线社区中,降低知识产权保护成本,减少浪费。在线社区用户数字知识产权侵权问题技术破解框架如图3所示。

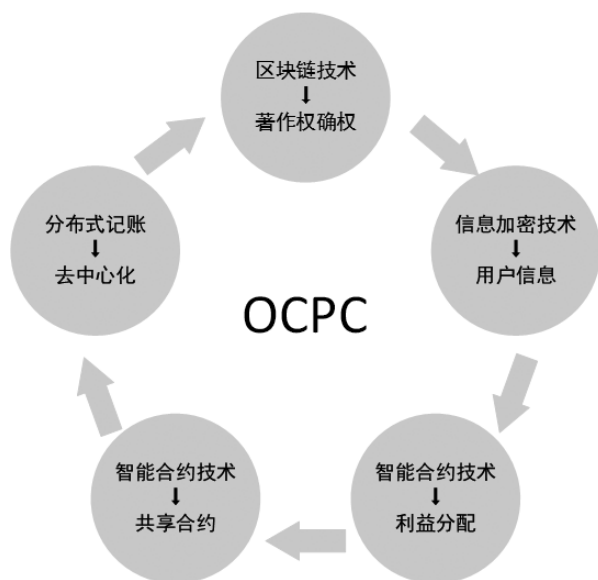


图3 在线社区用户数字知识产权侵权问题技术破解框架

4.3 基于区块链的在线社区知识产权共享保护体系具体搭建方案

本研究设计的基于区块链的在线社区知识产权共享保护体系由六个层级构成,由下往上依次为:区块链服务平台层、业务支撑层、业务及应用层、用户层及服务界面。其中区块链服务平台层将区块链功能模块与区块链底层技

术进行封装,核心功能模块包括数字作品确权模块、作品共享交易模块、著作财产权保护模块及版权信息检索模块,区块链底层技术主要包括共识算法、对等网络、智能合约、加密算法等,通过上述技术,保证了区块链溯源系统的安全性^[20];业务支撑层主要包括用户权限控制、数字作品存储及在线社区数据管理,该

层负责为整个平台提供功能方面的技术支持；业务应用层主要是对底层的服务进行具体的应用，并调用一系列的底层 API 来实现业务应用；用户层涵盖了 OCPC 涉及的所有用户，包括数字知识产权人、浏览用户、审查机构、公信力机构及其他相关企业；服务界面旨在为使用

者提供一个友好的操作界面，并将基础底层技术透明化，减少使用者的使用成本。在线社区用户数字知识产权共享保护体系框架由图 4 所示，下面将对框架中涉及的功能模块展开详细介绍。

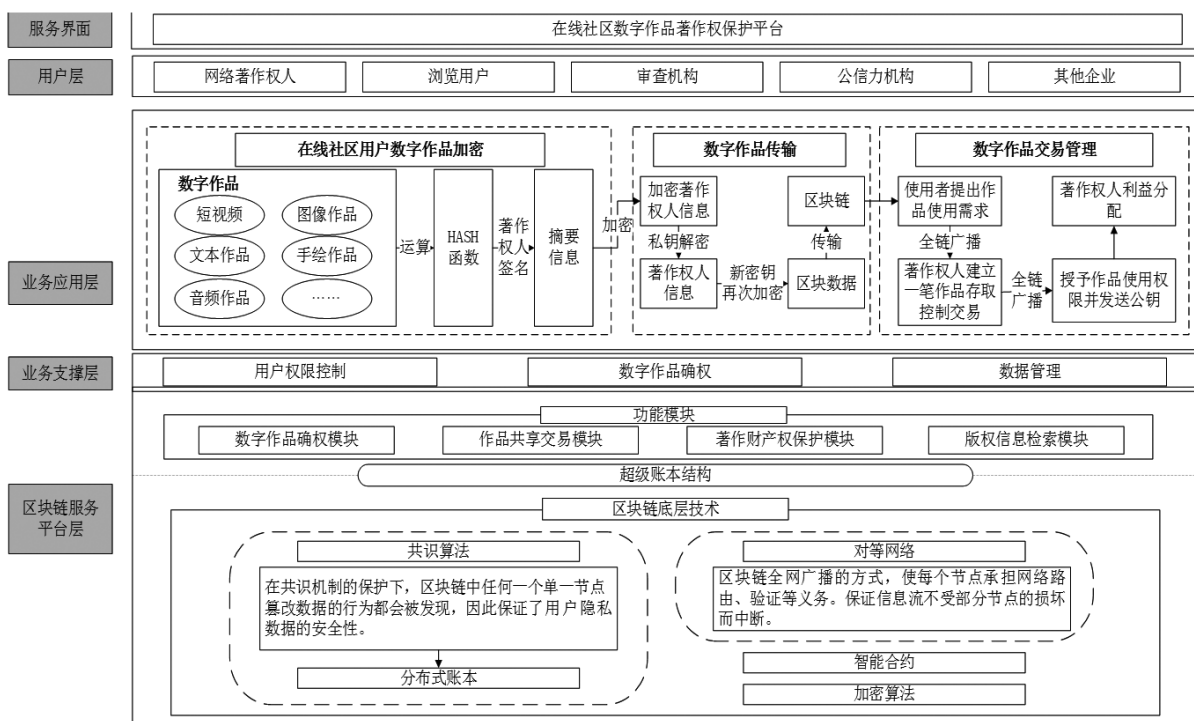


图 4 在线社区知识产权共享保护精益管理体系框架

（1）数字作品确权模块

数字作品确权模块的流程图见图 5，该模块主要涵盖的功能包括：数字作品上传、数字作品原创性检测、星际文件系统存储、数字作品上链及生成确权码。区块链技术应用于数字作品确权中，为解决目前我国数字知识产权共享保护所面临的诸如成本高、易遭破坏、确权难等问题提供了可行的途径：区块链技术中的

时间戳与哈希值可以载明任何数字作品的权利信息，加盖在数据块上不可更改的时间戳及唯一哈希值为数字作品的完成时间、知识产权归属等提供了有力证据；此外，据相关研究，相对于传统版权注册，使用区块链进行版权注册所需的比特币成本仅 0.4 美金左右，大大降低了确权成本^[21]。

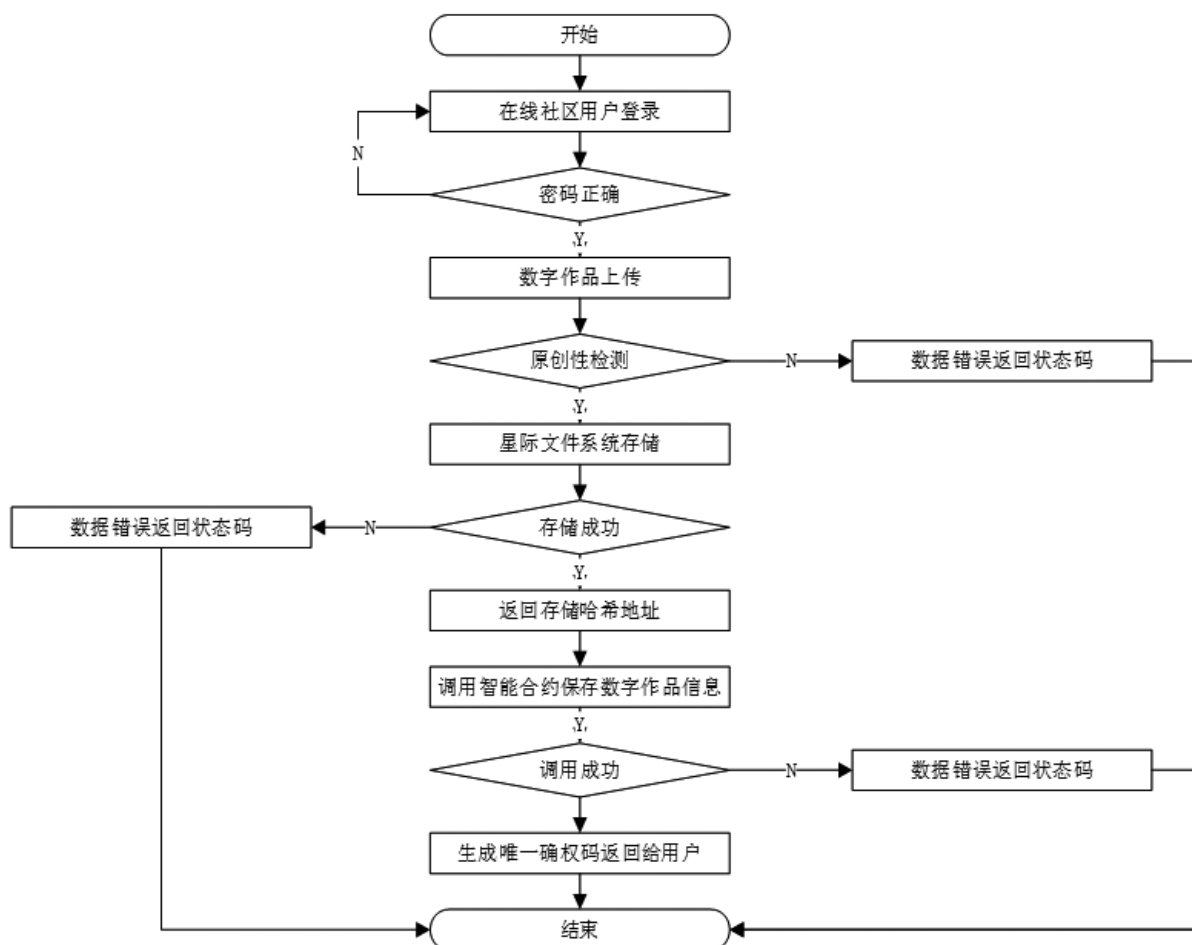


图5 数字作品确权模块流程图

(2) 作品共享交易模块

图6显示的是数字作品共享交易流程,在数字作品交易的过程中,引入了智能合约技术,建立基于智能合约的数字作品共享交易模式,并通过状态触发实现了智能合约的自动执行。该模式下,任何交易都可追溯,基于OCPC上的公开、透明的交易信息,例如数字产品每次下载、授权、改编、转让等的记录,既为数字作品共享交易提供了保证,同时也解决了网络侵权案件中损害难以计算的问题。

在该模块中,在线社区数字作品知识产权人与交易人将先通过在线社区线上签订数字作品共享交易协议,订约数字作品共享的边界、期限、要求、报酬等。此后,根据协议内容和

触发条件,系统自动编码生成区块链智能合约,由P2P网络对该合约进行扩散,OCPC上各个节点将对协议内容进行验证,将其存储在数据库中并生效。合约生效后,将会对合约状态进行定期的检验,触发条件未满足时该合约不会自动执行;满足触发条件时,该合约将被传输至OCPC上的各个节点进行验证,确认后合约中的协议将自动执行。当数字作品被在线社区知识产权人提交给交易人并确认接收后,触发智能合约执行条件,将按照双方协议支付报酬。若双方任一方违反了智能合约,系统将自动执行惩罚机制。该智能合约的整个过程均由系统完成,任何人都无法篡改智能合约内容,这保证了数字作品共享交易的安全性。

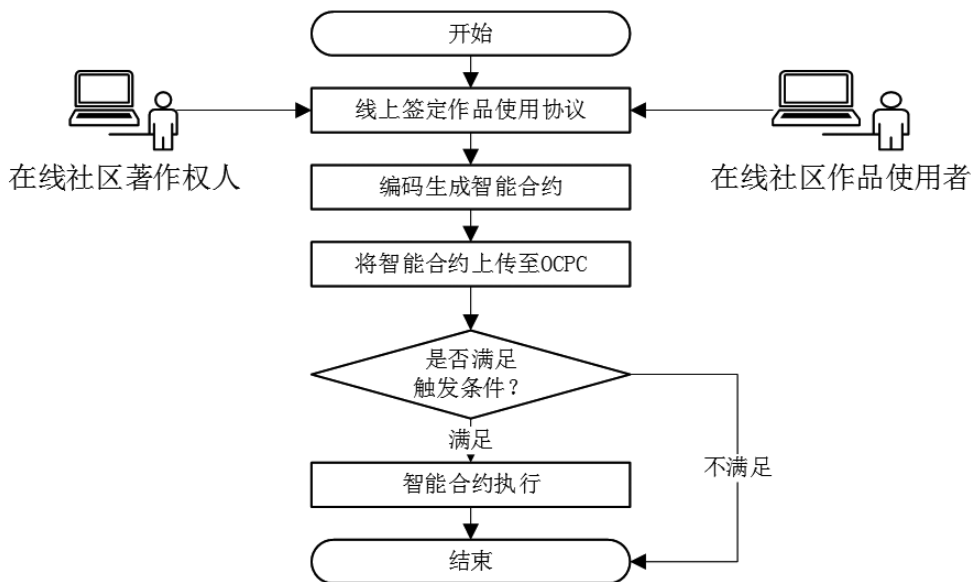


图6 数字作品共享交易流程图

通过区块链智能合约技术进行在线社区数字作品共享交易，有利于打破在线社区中心平台而导致的“信任危机”，优化数字知识产权共享保护运作生态。“去中心化”的区块链有助于建立“点对点”的互联网知识产权交易模式，减少交易过程中的中间环节，进而借助区块链的分布式技术消除信息壁垒，有效打破知识产权人信任危机^[22]。此外，区块链安全地保存了所有的知识产权交易数据，通过“时间戳技术”清晰记录所有数字作品的流转记录、使用记录，从源头上解决数字知识产权纠纷。再次，区块链中的非对称加密技术、PoW 共识机制等强大算力可以有效抵挡外部攻击，确保了 OCPC 上的数据安全性。最后，区块链技术的存证溯源机制和分布式记账通过为用户进行多维度记账的方式，建立起用户个人数字身份，有效解决了以往数字知识产权纠纷中存在的取证难问题^[23]。

（3）著作财产权保护模块

利用区块链智能合约技术，有助于解决传统知识产权交易中数据分散的问题，保证著作财产权。传统的数字作品共享交易缺少一个统一的平台，使得知识产权人的作品所有权可能以分散的方式出售，导致著作财产权的实现表现出各自独立的交易合同，甚至知识产权人不清楚自己的著作财产权内容等^[24]。基于智能合约的著作财产权保护平台可以有效改善这种情况，区块链将分散的数据集中到链上，让知识产权人、需求人等可以更集中地进行数字作品的交易，从而更好地保护权利人的著作财产权。在线社区著作财产权的保护主要是通过智能合约来进行的，智能合约所具有的自动交付和强制执行的特点使其不需要中间方，如法院、仲裁机构对交易人进行督促，保证了在线社区知识产权人的资金安全，具体如图7所示。

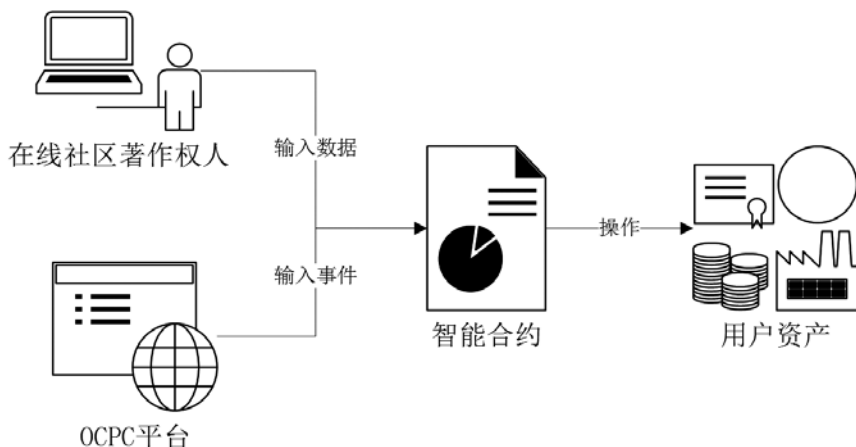


图7 在线社区用户资产保护示意图

（4）版权信息检索模块

对于在线社区用户来说，数字作品的知识产权查询是一个必不可少的环节。在线社区作为内容生产的平台，很多数字作品知识产权人可能会主动公开自己的数字作品、开放转发功能，以此来扩大该数字产品的影响力，因此数字作品需求者在使用数字作品之前，应当先对该作品的版权信息进行查询。该模块提供数字作品的版权信息查询，包括数字作品的标题、确权码、确权时间、知识产权人、使用范围等。

5 结语

本文主要研究了基于智能合约的知识产权共享保护精益管理体系构建。以在线社区为例，通过分析传统知识产权保护体系存在的问题，如知识产权归属难以确定、知识产权取证难、知识产权交易、知识产权损害难以评估等，结合精益管理理论和智能合约技术，提出了一种精益管理体系构建方案，以实现在线社区知识产权共享保护的高效和可靠。尽管将区块链、智能合约等技术应用到在线社区知识产权共享保护中，将会大大提高效率、降低成本，但同时，由于技术上的缺陷，可能使得智能合

同的实施产生不合法的行为。因此，在推广并应用区块链技术的同时，需要采取适当的管理和监管措施，以确保实现数字知识产权共享保护的同时维护公共利益。

参考文献

- [1] 中国互联网络信息中心. 第49次中国互联网络发展状况统计报告 [EB/OL]. (2022-02-25) [2022-06-17]. <https://www.cnnic.cn/n4/2022/0401/c88-1131.html>.
- [2] 国家版权局. 2020年中国网络版权产业发展报告 [EB/OL]. (2021-05-17) [2022-07-17]. <http://www.199it.com/archives/1269995.html>.
- [3] PLANT R. Online Communities [J]. Technology in Society, 2004, 26 (1): 51-65.
- [4] 徐小龙, 王方华. 虚拟社区研究前沿探析 [J]. 外国经济与管理, 2007 (9): 10-16.
- [5] JENNY P. Sociability and Usability in Online Communities: Determining and Measuring Success [J]. Behaviour & Information Technology, 2001,

- 20(5):347-356.
- [6] GUPTA S, KIM H W. Virtual Community: Concepts, Implications, and Future Research Directions [J]. Proceedings of the Tenth America's Conference on Information System. 2004.
- [7] LEE F S L, VOGEL D, LIMAYEM M. Virtual Community Informatics: What We Know and What We Need to Know [C]. Hawaii International Conference on System Sciences. IEEE Xplore, 2002: 2863-2872.
- [8] RIDINGS C M, GEFEN D, ARINZE B. Some Antecedents and Effects of Trust in Virtual Communities [J]. Journal of Strategic Information Systems, 2002, 11(3-4): 271-295.
- [9] SUNG H C. Prospects and Challenges Posed by Blockchain Technology on the Copyright Legal System [J]. Queen Mary Journal of Intellectual Property, 2019, 9(4): 430-451.
- [10] YU L, TSAI W T, LI G, et al. Smart-Contract Execution with Concurrent Block Building [C] //2017 IEEE Symposium on Service-Oriented System Engineering (SOSE). IEEE, 2017: 160-167.
- [11] ZEILINGER M. Digital Art as "Monetised Graphics": Enforcing Intellectual Property on the Blockchain [J]. Philosophy & Technology, 2018, 31(1): 15-41.
- [12] XU R, ZHANG L, ZHAO H, et al. Design of Network Media' s Digital Rights Management Scheme Based on Blockchain Technology [C] //2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS). IEEE, 2017: 128-133.
- [13] 赵丰, 周围. 基于区块链技术保护数字版权问题探析 [J]. 科技与法律, 2017(1): 59-70.
- [14] 郑阳, 杜荣. 区块链技术在数字知识资产管理中的应用 [J]. 出版科学, 2018, 26(3): 97.
- [15] 宁梦月, 刘东苏. 基于区块链技术的数字知识产权保护方案研究 [J]. 情报理论与实践, 2020, 43(7): 144-150.
- [16] 林爱珺, 章梦天. 网络内容生态治理的多元主体责任规制 [J]. 新闻爱好者, 2021(4): 14-16.
- [17] 胡勇. 基于联盟区块链的网络著作权保护研究 [J]. 广西政法管理干部学院学报, 2021, 36(4): 8-13.
- [18] 金强, 李智鸣. 网络社区版权保护的现状、问题及对策——以知乎社区为例 [J]. 中国编辑, 2021(8): 37-42.
- [19] 华劼. 区块链技术与智能合约在知识产权确权 and 交易中的运用及其法律规制 [J]. 知识产权, 2018(2): 13-19.
- [20] 束飞. 运用精益管理理论推行成本精益管理的思考 [J]. 中国经贸, 2016 (18): 150.
- [21] 王群, 李馥娟, 王振力, 等. 区块链原理及关键技术 [J]. 计算机科学与探索, 2020, 14(10): 1621-1643.
- [22] 王秋红. 基于区块链技术重构网络著作权保护制度 [J]. 四川民族学院学报, 2020, 29(3): 66-73.

[23] 俞锋, 谷凯月. 网络版权保护体系变革:来自区块链技术的支持与想象 [J]. 中国出版, 2021(2): 66-69.

[24] 王清, 陈潇婷. 区块链技术在数字著作

权保护中的运用与法律规制 [J]. 湖北大学学报(哲学社会科学版), 2019, 46(3): 150-157.