

一种基于拥塞发现的强化学习 P2P 网络视频点播预取策略

沈项军¹ 姚 银¹ 查正军²

(江苏大学计算机科学与通信工程学院 镇江 212013)¹

(中国科学院合肥智能机械研究所 合肥 230031)²

摘 要 在 P2P(Peer to Peer)网络视频点播中设计合理的视频内容预取策略,可以有效地解决视频观看中出现的缓冲等待时间过长、服务器负载过大等问题。而现有的 P2P 网络视频预取方法往往考虑了视频内容的发现而忽略了节点本身的状态监控,在网络拥塞发生时存在视频播放观看不佳的情况。提出了一种基于拥塞发现的强化学习 P2P 网络视频点播预取策略,通过监测节点的拥塞状况以及带宽等参数,采用 Q-learning 学习算法综合评估网络节点,引导视频预取节点的选择,减少了对网络拥塞节点视频内容的预取。实验结果表明,该方法能提高视频播放流畅性,避免用户在节点发生拥塞时等待时间过长,提高了视频播放效率。

关键词 P2P 网络,视频预取,拥塞发现,强化学习

中图法分类号 TP393

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2015.7.055

Video Prefetching Strategy Based on Congestion Finding with Reinforcement Learning in P2P VOD Networks

SHEN Xiang-jun¹ YAO Yin¹ ZHA Zheng-jun²

(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)¹

(Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China)²

Abstract A proper video prefetching strategy in P2P (peer to peer) VOD (Video on Demand) networks can effectively solve such circumstances as waiting time for buffering is too long and server is overloaded. While the existing video prefetching methods in P2P networks just consider video content finding but ignore monitoring nodes status, which makes poor video play when the network congestions occur. This paper presented a video prefetching strategy based on congestion finding with reinforcement learning in P2P VOD network. Through monitoring nodes status of congestion, bandwidth and other parameters, the Q-learning learning algorithm was used to evaluate network nodes, which can guide nodes selection and reduce the video prefetching from congested nodes. Simulation results show that the proposed method can make video playback smooth. It can also avoid too long waiting time when congestions happen and improve the efficiency of video playback.

Keywords P2P network, Video prefetching, Congestion finding, Reinforcement learning

1 引言

近年来,随着互联网、多媒体技术以及社交网络的飞速发展,网络视频的在线点播日益成为人们关注的焦点,出现了诸如 youtube、youku 等由用户上传的在线视频播放网络。据统计,2010 年 5 月 youtube 上观看的视频数据数量达到了 146 亿次^[1]。2017 年,互联网上的视频共享所用流量将占整个互联网流量的 69%^[2]。为了解决互联网上如此海量的视频传输与播放难题,出现了诸如 CDN(Content Delivery Network)网络与 P2P 网络等方面的研究^[3,4]。P2P 采用与 CDN 完全不同的内容传递方式,通过对等方式进行视频内容传输,实现了不依赖服务器而在众多用户节点上快速交换视频内容。

P2P 网络可以让更多用户参与并提供视频内容的共享和分发,降低网络建设成本,因此引起了学术界和工业界的广泛研究。

由于 P2P 网络的用户对等性,视频播放中存在用户搜索服务延迟以及系统拥塞等问题,因此利用视频内容预取策略预测用户未来视频播放的数据需求,利用空闲资源提前获取部分数据,可以有效提高视频播放效率^[5]。当前,P2P 网络中视频内容预取研究主要分为以下 3 类:1)基于合作内容预取的研究,即通过分析不同预取策略的性能,提高了查询的命中率并有效地降低了开销。如 Thampi 等^[5]提出通过提高一个视频段中稀少资源的可用性,在提高了命中率的同时降低了资源查找的带宽等开销;Jeffrey 等^[6]提出了 P2P 合作缓存和

到稿日期:2014-08-28 返修日期:2014-12-24 本文受国家自然科学基金项目(61005017)资助。

沈项军(1977—),男,博士,副教授,主要研究领域为模式识别、分布式多媒体计算,E-mail: xjshen@ujs.edu.cn;姚 银(1991—),女,硕士生,主要研究领域为 P2P 流媒体,E-mail: 634629768@qq.com;查正军 男,博士,研究员,博士生导师,主要研究领域为多媒体计算、计算机视觉,E-mail: zhazj@iim.ac.cn.

OGC WMS 数据预取合作的方式,运用金字塔方案实现平滑启发式的查询和预取统计功能;Toufik 等^[7]提出 SVC(Scalable Video Coding)的使用,利用合作预取以及 SVC 最小化视频点播的流媒体寻求延迟的同时提供持续的寻道操作服务。2)基于用户行为特征的研究,即通过分析用户观看视频时的行为,实现了基于数据块流行度感知的预取策略。如 Guo 等^[8]提出一种新的自适应数据搜索结构 U-VOD,该机制根据用户的自主访问行为自适应地组成行为相似团体,降低了用户 VCR 操作引起的跳转时延,并减轻了对服务器的访问压力;Yu 等^[9]分析和模仿了大规模视频点播环境中的用户行为和相关访问方式,发现视频流行度的改变受到新内容的引进以及外部因素的影响,其优势在于使得视频点播系统能够更好地根据资源分配情况来决定预取行为;Riad 等^[10]提出在 P2P 视频点播系统中利用策略和双缓冲机制来缓冲媒体数据,通过分析用户行为,预取播放次数高的视频,最终减少了网络上的流量使用以及服务器负载,并加快了读取视频块到缓冲区的过程。3)基于节点管理和资源定位的研究,此类方法通过分析 P2P 流媒体体系结构及调度策略,实现了快速准确定位所需视频资源。如 He 等^[11]提出的 Guided Seeks,基于有效的数据统计,用户能够更有效地寻求高点击率和有趣的片段。王庆波等^[12]提出了 Barnet——一个基于特征信息定位的 P2P 网络模型,其目标是为广域网络构建一个高性能、负载均衡的海量信息资源服务平台;并讨论了在 Barnet 中使用基于特征信息定位技术对具体信息资源进行查找、定位的具体策略。张玉洁等^[13]提出了基于用户需求的 P2P 系统的查找定位行为,其目标是快速、准确地对资源定位,每个用户节点通过对系统资源位置的掌握和处理,最终减小了查找的深度和广度,提高了查找效率和命中率。

综上,现有的 P2P 视频点播预取考虑了视频内容的发现而忽略了对节点本身的监控,在网络发生拥塞时会存在视频播放不佳的情况,进而影响视频点播中用户的播放体验。为此本文提出了一种基于拥塞发现的强化学习 P2P 网络视频点播预取策略,通过监测节点的状态,特别是节点的拥塞状况以及带宽等参数,并采用 Q-learning 学习算法综合评估网络节点,减少了对网络拥塞节点视频内容的预取。实验表明,本文方法能够有效地控制节点的负载,避免网络视频播放时发生拥塞,减少用户观看视频的等待时间,提高视频播放的流畅性。

2 基于拥塞发现的强化学习视频预取策略

2.1 强化学习模型

强化学习是指在环境中智能体 Agent 通过“试错”的方法与环境进行交互^[14],从中学习最优策略,使自己的效用最大化。强化学习的理论构架如图 1 所示。

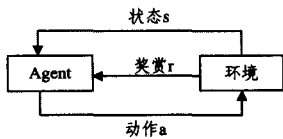


图 1 强化学习模型

当前强化学习算法有很多,如时间差分法(Time difference)TD(λ)^[15]、Q-learning^[16]以及决策评价学习(actor-critic learning)^[17]。而 Q-learning 作为一种无模型、无监督的在线强化学习算法,是基于马尔可夫决策过程(Markov Decision Process, MDP)的递进式动态规划算法^[18]。马尔可夫决策过

程可定义为四元组 $\langle S, A, T, R \rangle$,其中 S 是环境的离散状态有限集, A 是 Agent 行为的可选集, T 为状态转移模型, R 为奖赏函数。其形式定义如下: $T: S \times A \times S \rightarrow [0, 1]$, $T(S_i, a_i, S_{i+1})$ 表示当环境状态为 S_i , Agent 采取行动 a_i 时,到达状态 S_{i+1} 的概率; $R: S \times A \times S \rightarrow R$, $R(S_i, a_i, S_{i+1})$ 表示 Agent 在状态 S_i 采取行动 a_i 到达 S_{i+1} 的奖赏值。

2.2 基于拥塞发现的强化学习的预取策略

为了充分利用 P2P 网络视频点播的优势,避免视频播放时发生拥塞,本研究通过成组资源管理方法,将具有相同资源的节点连接在一起。当用户发出请求时,智能体监测收集邻居节点的拥塞值和奖赏值,通过 Q-learning 算法综合评估网络节点,选择综合指标最佳的预取节点,从而减轻过载节点上查询消息过多的负担,更好地完成视频传输与播放。

本文采用一种在现有强化学习中考虑拥塞因素的预取模型。为了引导视频查询尽量在轻载节点中进行,需要评估其拥塞状况,因此本文利用一个消息在一个节点的等待时间来确定节点的拥塞程度(CL)^[19],通过计算节点的拥塞程度 CL 来判断节点的负载状态。节点 P_i 在时刻 t 的拥塞程度计算如下:

$$CL_i(t) = \frac{1 + Q_i(t)}{C_i} \quad (1)$$

其中, C_i 表示节点 P_i 在单位时间可以处理的消息数,当一个节点正在处理消息时标记为‘busy’,转发到这个节点的消息都放在该节点的输入缓存队列中; $Q_i(t)$ 表示在时刻 t 节点 P_i 的输入缓存队列中的消息数。定义两个阈值 B_{thred} 、 U_{thred} 用于确定节点状态,当 $CL \geq U_{thred}$ 时节点为过载;当 $CL \leq B_{thred}$ 时节点为轻载,否则节点为正常情况下的拥塞情况。

因此,加入节点拥塞发现、带宽参数等节点状态发现的 Q-learning 学习公式如下:

$$Q(s_{t+1}, a_{t+1}) = \partial \times [R_t + \gamma \times \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] + \beta \times CL_i \quad (2)$$

其中,参数 ∂ 是学习速度参数, γ 是折扣因子, β 是拥塞因子。 R_t 为奖赏和的最大值,其计算公式如下:

$$R_t = \sum_{k=0}^{\infty} \phi^k r_{t+k+1} \quad (3)$$

其中, ϕ 为折扣率, $0 \leq \phi \leq 1$,决定下一个状态的奖赏值; r_t 为 s_t 执行 a_t 后被观察到的奖赏,计算公式为:

$$r_t = \frac{A_i \times [\epsilon B_u + (1 - \epsilon) B_d]}{n_i} \quad (4)$$

其中, A_i 为节点 P_i 的吸引能力, B_u 为当前节点的上行带宽, B_d 为当前节点的下行带宽, ϵ 为比例参数, n_i 为节点 P_i 上的资源数。节点 P_i 的吸引能力计算公式为:

$$A_i = \sum_{h=1}^{h=k} \frac{N(i, h)}{h^\delta} \times C_i \quad (5)$$

其中, C_i 是节点 P_i 的处理能力, $N(i, h)$ 是与节点 P_i 相距 h 跳的节点个数, δ 是控制因数, h^δ 用来控制不同距离节点对节点吸引能力影响的权重比。

图 2 为网络节点模拟图,当前节点为 P ,每个节点都维护一个邻居表(如表 1 所列),用于保存所有邻居节点的奖赏值和拥塞值,智能体根据表格选择下一个访问节点。节点的奖赏值的计算方法见式(4),拥塞值的计算方法见式(1)。如果想要观看的视频关键字段已存在于当前节点,则查找成功,直接读取;否则节点 P 发出搜索消息,该消息在网络中以随机走(random walk)的方式在其邻居节点搜索,用存活时间(TTL)限制搜索路径的长度。每到达一个节点, TTL 减 1,若

该消息经过的节点 P_k 与发起节点 P 有相同的资源,就将 P_k 加入到候选节点集中。若邻居节点中不存在目标资源节点,则采用 random walk 的方式转发查询消息给邻居节点,该消息一直转发,直到该消息的存活时间到零为止。通过对候选节点集中节点的奖赏值 r_i 和拥塞值 CL 的计算,得出最优 Q 值,其对应节点即为最佳资源预取节点。

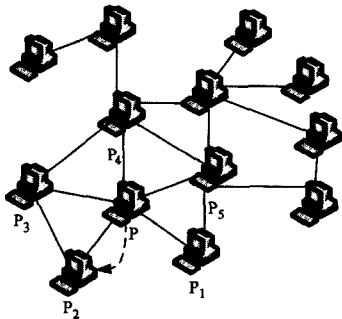


图2 网络节点模拟图

表1 邻居表

Peer	P ₁	P ₂	P ₃	P ₄	P ₅
奖赏	1.4	0.8	0.8	1.0	1.1
CL	1	0.01	1	0.1	0.5
资源数	1	2	1	4	4

假设节点 P 的邻居节点存在目标视频资源,例如是关于资源‘a’的,节点 P_1 到 P_5 上的资源分别为[‘e’],[‘a’,’d’],[‘a’],[‘e’,’c’,’d’,’f’],[‘a’,’c’,’f’,’g’],通过 random walk 方式搜索到的具有此资源的节点为{ P_2, P_3, P_5 },根据式(2)计算得出节点 P_2 为最佳视频内容预取节点,则选择节点 P_2 为目标节点。如果节点 P 的邻居节点不存在目标视频资源,则转发查询消息给邻居节点,进行一次新的查询。

上述方法的算法流程如图3所示。

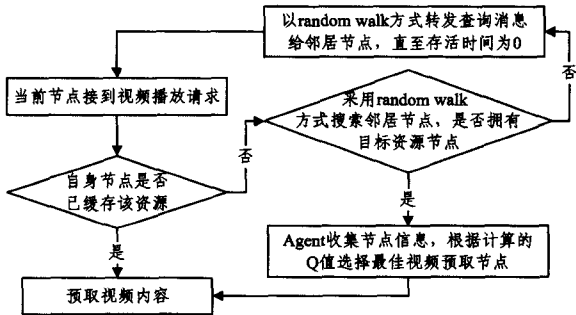


图3 网络资源节点查找流程

3 模拟实验

3.1 实验参数

本实验在 Windows 平台下利用 python2.6 进行程序的编写及运行,根据 Gnutella 协议设计网络,该网络中包括 5000 个节点,流媒体传输比率为 256kbps^[20]。视频流的长度为 30min,视频被分为 90 个视频段,每个视频段时长 20s。网络中设定每个节点上行带宽分布在 500kbps 和 1400kbps 之间,下行带宽分布在 800kbps 和 1700kbps 之间。由于网络中不同因素对视频播放的影响比率不同,实验中设置学习速度参数 θ 为 0.3,拥塞因素 β 为 0.5。折扣因子越大,上一次状态选择因素对此次选择的影响越大,实验中设置 γ 为 0.3。计算奖赏值的比例参数 ϵ 为 0.5。

网络中的拥塞主要是由过多查询的视频播放引起的,本

文假设所有节点所能服务的请求播放数是相同的,记为 V_{pi} 。若 $V_{pi}=50$,则最多能服务 50 个节点请求播放。假设在本系统中每个节点每秒会发起 100 个查询消息,用选择率来表示每秒发起查询会有多少个状态选择,每个状态选择称作一个播放(play),假设选择率为 0.01,那么长度为 30min 的视频流可能有 1800 个 play。采用 random walk 进行资源搜索,每个查询产生 5 个 walker,转发查询消息给邻居节点的 TTL 设置为 8。实验中设置节点状态的两个阈值 B_{thred} 和 U_{thred} 分别为 0.05 和 0.1。计算吸引能力的两个参数 k_c 和 δ 分别为 2 和 1。完成播放时,用户直接重连或离开系统。

3.2 实验过程

本实验主要从两个方面研究基于拥塞发现的强化学习 P2P 网络视频点播预取策略,首先分析了网络查询拥塞下的视频播放性能,如节点请求时的拥塞变化及不同拥塞情况下平均搜索路径的变化,并对比了考虑拥塞与否情况下的最佳行为率^[14],以此来体现播放性能;其次是预取机制下的性能分析,分析了强化学习模型优化下的平均查询时间、成功预取资源所需的 hops 数以及对比分析强化学习预取与随机预取情况下的预取命中率。

(1) 查询拥塞下的性能分析

图4反映了基于拥塞发现的强化学习模型下节点请求时的网络拥塞变化。实验中对发起视频资源请求的节点的拥塞求平均值得到图4的结果,从开始到第 15min 的时间段内,由于大量节点同时请求播放同一节点资源,导致节点拥塞值逐渐增大。而第 15min 到第 65min 在强化学习方法的优化下,节点的拥塞值会逐渐降低,最终达到稳定状态。以上说明,强化学习模型能及时发现、均衡拥塞,使拥塞值逐渐减小且达到稳定。

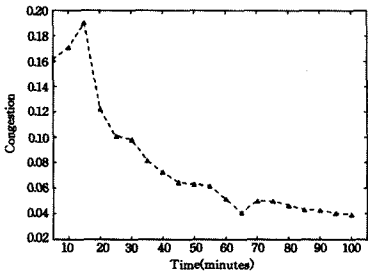


图4 强化学习模型下的拥塞变化

图5反映了不同网络节点消息处理负载下平均搜索路径长度的变化。从 3 种不同负载状况下($V_{pi}=10, V_{pi}=50, V_{pi}=100$)的对比可以看出,3 种网络负载状态下,在前 20min,由于网络处于学习状态,平均搜索路径长度不稳定,下降趋势不明显。在第 20min 到第 80min,网络经过多次基于强化学习的视频预取后,不同负载下所需的平均搜索路径迅速减少,且最终达到稳定状态。以上说明了不同负载下,本文方法都能很好地提高视频内容预取的性能。

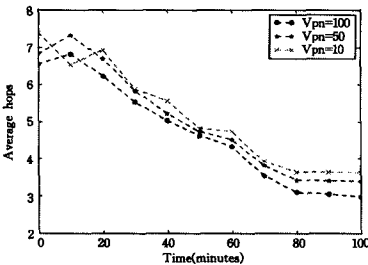


图5 不同负载下平均搜索路径长度的变化

图6反映了考虑拥塞因素和未考虑拥塞因素情况下的最佳行为率的对比结果。最佳行为率 $O_i(a)$ 是指基于强化学习方法的预取策略指导节点最优状态选择率,其计算方法为:

$$O_i(a) = \frac{r_1 + r_2 + \dots + r_{k_a}}{k_a} \quad (6)$$

r_{k_a} 表示视频播放每次行为 action 对应的奖赏值, k_a 为总的次数。

实验中以每 100 个 play 作为一个单位观察行为率的变化,从图6中可以看出,随着播放数目的增加,最佳行为率逐渐上升。考虑拥塞因素的行为率明显高于未考虑拥塞因素的行为率,说明了通过监测节点拥塞因素,视频播放预取效率得到明显提高。

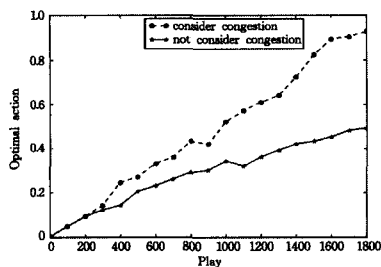


图6 最佳行为率

(2) 预取机制下的性能分析

图7反映了预取条件下视频的平均查询时间的变化趋势,实验中对 5000 个用户节点请求时的查询延迟时间进行了分析,可以看出随着强化学习模型的优化,经过预取处理后的平均查询时间比开始时减少了很多,而平均查询延迟时间越少,说明用户播放视频时等待时间越少,可见基于强化学习的预取策略提高了视频的播放流畅性。

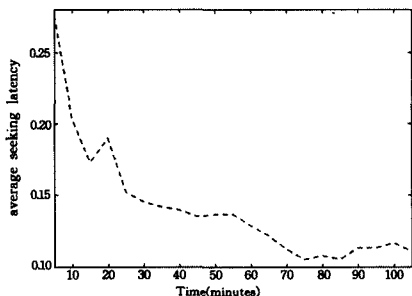


图7 平均查询延迟时间的变化

图8反映了成功预取资源所需的平均跳数(hops),为对每 1000 个成功预取的 hops 求均值得到的结果。随着时间的增加,预取需要的平均 hops 逐渐减少。从图8中可以看出,强化学习优化后成功预取资源需要的平均跳数比开始时减少了很多,而查询需要的平均跳数越少,说明资源搜索的效率越高,可见基于拥塞发现的强化学习 P2P 网络视频内容预取策略大大优化了网络的搜索性能。

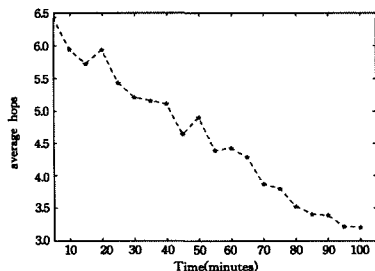


图8 成功预取资源所需的平均 hops 数

图9反映了随机预取和基于强化学习的预取下的命中率的变化。为了了解预取的效率,用命中率^[21]来比较网络的预取性能。将基于拥塞发现的强化学习预取策略与随机预取策略进行对比来更好地观察实验效果。假设在网络环境中,每个预取节点根据下载带宽预取所需要的下一个状态,然后顺序地预取状态集合中所需节点资源。从图9中可以看出,基于强化学习预取的命中率明显高于随机预取下的命中率,当用户群超过 1500 时,网络中的资源充足,因此命中率趋于稳定,说明了基于强化学习下的预取能够充分利用网络中的资源,预取的命中率有了显著的提高。

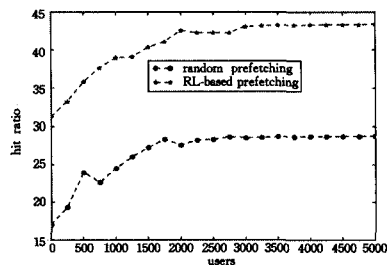


图9 随机预取和基于强化学习预取下命中率的对比

结束语 本文提出了一种基于拥塞发现的强化学习 P2P 网络视频内容预取策略,通过监测节点的状态,特别是节点的拥塞状况以及带宽、节点吸引力等参数,并采用 Q-learning 学习算法来综合评估网络节点,减少了对网络拥塞节点视频内容的预取。实验证明,该方法通过考虑拥塞等因素,能够有效地在强化学习基础上控制并解决节点的负载,避免网络查询过程中发生拥塞,最终提高视频播放的流畅性,减少用户观看视频的等待时间,从而提高了整个视频播放的预取性能。

参考文献

- [1] Rao L. comScore: YouTube Reaches All-Time High of 14.6 Billion Videos Viewed In May [OL]. <http://techcrunch.com/2010/06/24/comscore-youtube-reaches-all-time-high-of-14-6-billion-videos-viewed-in-may/>, 2010
- [2] Cisco I. Cisco Visual Networking Index: Forecast and Methodology[S]. Cisco System, 2008
- [3] Passarella A. A survey on content-centric technologies for the current Internet CDN and P2P solutions[J]. Computer Communications, 2012, 35(1): 1-32
- [4] Thampi S M, Sekaran C K. Survey of search and replication schemes in unstructured P2P networks[J]. Network Protocols and Algorithms, 2010, 2(1): 93-131
- [5] Thampi S M, Sekaran C K. Q-learning Based Collaborative Load Balancing Using Distributed Search for Unstructured P2P Networks[C]// 33rd IEEE Conference on Local Computer Networks, 2008: 797-802
- [6] Bergamini J A, Haungs M. Enabling P2P Cooperation WMS proxy Caching and Prefetching in an Educational Environment [C]// 10th AGILE International Conference on Geographic Information Science. 2007: 1-14
- [7] Ahmed T, Abbasi U, Medjah S. Efficient IPTV Services Delivery using SVC Adaptation and Cooperative Prefetching [J]. IEEE Comsoc Mmtc E-Letter, 2011, 6(1): 49-50
- [8] Guo Hong-fang, Lin Yu-song, Wang Zong-min. User-behavior based self-adaptive indexing for VoD service[J]. Computer Applications Research, 2012, 29(4): 86

- classification based on I2C distance and label dependency[J]. Computer Science, 2014, 41(1): 88-90
- [3] Trohidis K, Tsoumakas G, Kalliris G, et al. Multi-label classification of music into emotions[C]//Proceeding of 9th International Conference on Music Information Retrieval (ISMIR). Philadelphia, PA, USA, 2008: 325-330
- [4] Elisseeff A, Weston J. A kernel method for multi-labelled classification[C]//Advances in Neural Information Processing Systems 14. Cambridge, MA: MIT Press, 2002: 681-687
- [5] Tsoumakas G, Katakis I. Multi-Label Classification: An Overview[J]. International Journal of Data Warehousing and Mining, 2007, 3(3): 1-13
- [6] Zhang Min-ling, Zhou Zhi-hua. A Review on Multi-Label Learning Algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837
- [7] Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356
- [8] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General Systems, 1990, 17: 191-208
- [9] Hu Qing-hua, Zhang Lei, An Shuang, et al. On robust fuzzy rough set models[J]. IEEE Transactions on Fuzzy systems, 2012, 20(4): 636-651
- [10] McCallum A. Multi-label text classification with a mixture model trained by EM[C]//Proc of Working Notes of the AAAI'99 Workshop on Text Learning. Menlo Park, CA: AAAI Press, 1999
- [11] Ueda N, Saito K. Parametric mixture models for multi-label text [C]//Advances in Neural Information Processing Systems 15. Cambridge, MA: MIT Press, 2003: 721-728
- [12] Zhang Min-ling, Zhou Zhi-hua. Multi-label neural networks with applications to functional genomics and text categorization [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338-1351
- [13] Zhang Min-ling, Zhou Zhi-hua. ML-kNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048
- [14] Comit   F D, Gilleron R, Tommasi M. Learning multi-label alternating decision tree from texts and data[C]//Lecture Notes in Computer Science 2734. Berlin: Springer, 2003: 35-49
- [15] Yeung D S, Chen D G, Tsang E C C, et al. On the Generalization of Fuzzy Rough Sets[J]. IEEE Transactions on Fuzzy Systems, 2005, 13(3): 343-361
- [16] 郑伟, 王朝坤, 刘璋, 等. 一种基于随机游走模型的多标记分类算法[J]. 计算机学报, 2010, 33(8): 1418-1426
Zheng Wei, Wang Chao-kun, Liu Zhang, et al. A multi-label classification algorithm based on random walk model[J]. Chinese Journal of Computers, 2010, 33(8): 1418-1426
- [17] 广凯, 潘金贵. 一种基于向量夹角的 k 近邻多标记文本分类算法[J]. 计算机科学, 2008, 35(4): 205-207
Guang Kai, Pan Jin-gui. An kNN algorithm based on vector angle for multi-label text categorization [J]. Computer Science, 2008, 35(4): 205-207
- [18] Tsoumakas G, Katakis I, Vlahavas I. Mining Multi-label Data [M]//Data Mining and Knowledge Discovery Handbook. Berlin: Springer, 2010: 667-685

(上接第 261 页)

- [9] Yu Hong-liang, Zheng Dong-dong, Zhao B Y, et al. Understanding User Behavior in Large-scale Video Streaming Services [J]. Computer Science Department, 2006, 40(4): 18-21
- [10] Riad A M, Elmogy M, Schehab A I. A Framework for Cloud P2P VoD System based on User's Behavior Analysis[J]. International Journal of Computer Applications, 2013, 76(6): 20-26
- [11] He Yi-feng, Shen Guo-bin, Xiong Yong-qiang. Optimal Prefetching Scheme in P2P VoD Applications With Guided Seeks[J]. IEEE Transaction on Multimedia, 2009, 11(2): 138-151
- [12] 王庆波, 代亚非, 田敬, 等. 基于特征信息定位的 P2P 网络模型: Barnet[J]. 软件学报, 2003, 14(8): 1481-1488
Wang Qing-bo, Dai Ya-fei, Tian Jing, et al. An Infrastructure for Attribute Addressable P2P Network: Barnet[J]. Journal of Software, 2003, 14(8): 1481-1488
- [13] 张玉洁, 何明, 孟祥武. 基于用户需求的内容分发点对点网络系统研究[J]. 软件学报, 2014, 25(1): 98-117
Zhang Yu-jie, He Ming, Meng Xiang-wu. Research on CDN-P2P system over user requirements[J]. Journal of Software, 2014, 25(1): 98-117
- [14] Sutton R S, Barto A G. Reinforcement Learning—An Introduction[M]. MIT Press, 1998
- [15] Tanner B, Sutton R S. Temporal-Difference Networks with Eligibility Traces[C]//Proceedings of the 22nd International Conference on Machine Learning, 2005: 1313-1320
- [16] Murphy S A. A Generalization Error for Q-Learning[J]. Journal of Machine Learning Research, 2005, 6: 1073-1097
- [17] Grondman I, Busoniu L, Lopes G A D, et al. A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients[J]. IEEE Transactions on Systems, 2012, 42(6): 1291-1307
- [18] Mahadevan S, Maggioni M. Proto-value Functions: A Laplacian Framework for Learning Representation and Control in Markov Decision Processes[J]. Journal of Machine Learning Research, 2007, 8(10): 2169-2231
- [19] Kwong K W, Tsang D H K. A Congestion-Aware Search Protocol for Unstructured Peer-to-Peer Networks[C]//Parallel and Distributed Processing and Applications; Second International Symposium (ISPA). HongKong, China, 2005: 319-329
- [20] Huang Cheng, Li Jin, Ross K W. Can Internet Video-on-Demand be Profitable? [J]. ACM SIGCOMM Computer Communication Review, 2007, 37(4): 133-144
- [21] Abbasi U, Ahmed T. COOCHING: Cooperative Prefetching Strategy for P2P Video-on-Demand System[C]//Wired-Wireless Multimedia Networks and Services Management; IEEE International Conference on Management of Multimedia and Mobile Networks and Service (MIMNS 2009). 2009: 195-200