

一种基于 P2P 的短视频分享系统

陈 卓^{1,2}, 李 彦¹

(1. 重庆理工大学计算机科学与工程学院, 重庆 400054; 2. 电子科技大学通信抗干扰技术国家重点实验室, 成都 611731)

摘 要: 现有在线短视频分享策略通常采用 C/S 架构, 给视频服务器带来较大的带宽压力。为此, 提出一种采用点对点方式的在线短视频分享系统 IShare, 该系统结合用户点播偏好和视频文件之间的社会网络特性实现视频分享。IShare 主要包括基于点播兴趣的节点分簇和视频数据源节点的查找 2 个核心技术。实验结果表明, IShare 具备较好的视频数据源节点查找能力, 可降低视频服务器带宽资源消耗。

关键词: 点对点; 查找算法; 社会网络; 重叠网; 视频点播; 兴趣簇

P2P-based Short Video Sharing System

CHEN Zhuo^{1,2}, LI Yan¹

(1. College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China; 2. National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China)

【Abstract】 Online short video sharing is reshaping the way people watching videos, so this paper presents sharing policies mainly adopts the Client/Server architecture, which can result huge bandwidth pressure at streaming server. A P2P based online short video sharing System IShare is proposed, which combines viewing interests of users and social network characteristic of short videos, which can share short videos between users. IShare mainly includes interest based peers clustering and streaming source peers searching. Experimental results show IShare has a high efficient streaming source peers searching ability and can greatly reduce the bandwidth consumption of streaming server.

【Key words】 Peer-to-Peer(P2P); searching algorithm; social network; overlay network; Video-on-Demand(VoD); interest cluster

DOI: 10.3969/j.issn.1000-3428.2012.03.090

1 概述

近年来, 在线网络视频分享成为新的 Internet 重要应用。大量在线视频分享网站如 YouTube、优酷、土豆网每天吸引上百万的点播量^[1]。大多数短视频分享网站采用的是客户机/服务器(C/S)的架构实现, 以这种架构带来的问题是视频服务器的带宽消耗巨大。基于点对点(Peer-to-Peer, P2P)的视频类应用越来越多^[2-4], 这类系统中的用户在从其他节点获取视频流的同时也贡献自己的上行带宽, 提供缓存的视频流给其他节点。但由于在线视频分享自身的特点使得直接应用已有的 P2P 视频播放技术存在诸多挑战。主要的技术难点在于: (1)视频分享网站中的视频节目通常较小, 在同一时刻观看同一个视频节目的节点通常不多, 这使得节点之间共享视频流的难度增大; (2)如何在较短时间内高效地查找视频源节点也成为必须研究的问题。为此, 本文提出一种采用点对点方式的在线短视频分享系统 IShare。

2 基于 P2P 的在线短视频分享系统 IShare

IShare 的主要设计思想是结合节点的点播兴趣偏好和视频文件之间所具有的社会网络特性, 并建立混合结构重叠网, 该重叠网结构如图 1 所示。该图中的 P1 为一个新访问 IShare 视频分享网站的节点, P1 点播了视频 V1。由于 P2 和 P3 都曾经点播了 V1, 因此 P2 和 P3 可以向 P1 提供 V1。这时 P1、P2、P3 同属于基于社会特性的重叠网, P2 和 P3 具有较大的概率缓存了 P1 即将点播的另一个视频节目。另外, 根据对用户点播兴趣偏好, IShare 把节点组织成多个兴趣簇, 每个簇中的节点对同一类视频有同样的点播偏好, P1 根据和自己点播兴趣的相似程度和多个节点建立了邻居关系, 如 P1

和簇 1 中的多个节点 P4、P5、P6 建立了邻居关系, 这里 P4、P5、P6 不一定缓存了 V1, 但由于和 P1 的点播兴趣最相似, 因此 P1 观看完 V1 后 P4、P5、P6 具有很大的概率缓存了 P1 即将点播的视频。

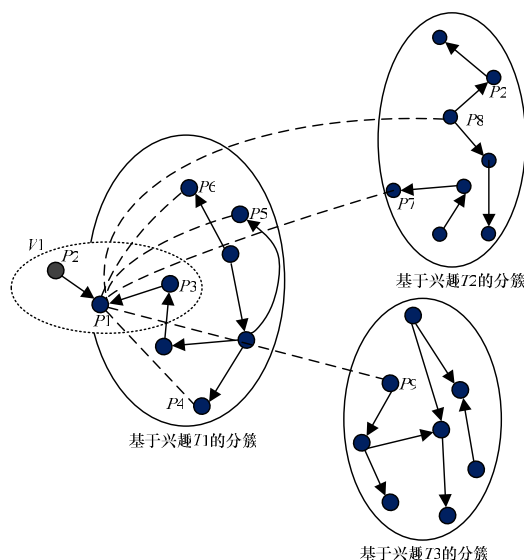


图 1 IShare 的重叠网结构

基金项目: 国家科技基金资助重大项目(2008ZX03004); 重庆市教委科技基金资助项目(KJ110831)

作者简介: 陈 卓(1980—), 男, 讲师、博士研究生, 主研方向: 分布式网络, 网络通信协议; 李 彦, 副教授

收稿日期: 2011-07-27 **E-mail:** uestccz@gmail.com

2.1 混合重叠网络结构的建立

IShare 的重叠网络结构把用户的点播偏好和视频文件之间形成的社会特性进行结合, 主要目的是为了点播节点在选择点播新的视频文件时, 能够以较小的开销找到更多的视频数据源节点。这里首先介绍在一段时间内用户的点播兴趣计算方法。

设视频节目共有 n 种分类, 分别为 T_1, T_2, \dots, T_n 。另外假设点播节点 i 对应于一个兴趣向量 $V_i = (w_{i,T_1}, w_{i,T_2}, \dots, w_{i,T_n})$, 该向量中的元素 w_{i,T_j} 对应于节点 i 对 T_j 类视频节目的偏好, 而 $w_{i,T_j} = \frac{Num_{i,T_j}}{Num_i}$, $Num_i = \sum_{j=1}^n Num_{i,T_j}$ 。其中, Num_{i,T_j} 表示点播节点 i 曾经点播过 T_j 类视频节目的次数, Num_i 表示节点 i 曾经点播 IShare 网站的视频总次数。节点 i 在访问 IShare 时, 首先计算得到自己的兴趣向量并把向量信息发送给 IShare。IShare 的 Tracker 服务器会计算出节点 i 最感兴趣的视频分类 T_{fav} , 即 T_{fav} 满足:

$$w_{i,T_{fav}} = \max(\frac{Num_{i,T_j}}{Num_i}), 1 \leq j \leq n$$

并且向节点 i 提供分类 T_{fav} 上的部分在线节点的信息作为节点 i 的候选邻居节点。在这些候选节点中, 节点 i 会进一步通过比较点播相似度选择其中和自己点播相似度最高的节点建立邻居关系。在 IShare 中, 通过计算 i 和 j 的兴趣向量的欧氏距离得到 i 和 j 的点播相似度。如: $V_i = (w_{i,T_1}, w_{i,T_2}, \dots, w_{i,T_n})$, $V_j = (w_{j,T_1}, w_{j,T_2}, \dots, w_{j,T_n})$, 则 i 和 j 的点播相似度 $S_{i,j}$ 计算方法为:

$$S_{i,j} = \frac{Num_i}{Num_j} \sqrt{(w_{i,T_1} - w_{j,T_1})^2 + (w_{i,T_2} - w_{j,T_2})^2 + \dots + (w_{i,T_n} - w_{j,T_n})^2} \quad (1)$$

$S_{i,j}$ 越小表示 i 和 j 的点播相似程度越高。 i 通过计算和候选邻居节点的点播相似度, 和相似度最大的 m 个节点建立起实际的邻居关系, 称节点在 T_{fav} 分类中建立起的邻居关系称为强关联邻居关系。节点 i 可能还点播过其他几种类型的视频。因此, IShare 中的每个点播用户还会和自己曾经点播过的视频类建立弱关联, 具体的方式是通过 Tracker 服务器, 根据 V_i , 提供少量属于其他类型的节点。

另外, 文献[4]通过 Tracking 实验说明在线视频分享网站的视频文件之间具有社会网络特点。NetTube 基于这个特点实现了一个具有社会特点的重叠网络。IShare 实现的具有社会网络特性的重叠网络类似于 NetTube。因此, IShare 的重叠网络结构是一个混合结构, 更好地结合了点播兴趣和社会网络属性。

2.2 视频数据源节点查找算法

2.2.1 算法的基本描述

IShare 的资源查找算法的伪码描述如下。

输入

v : 被 N_i 选择用来观察下一节点的短视频

$T(v)$: 短视频 v 的类型

StrongNbri: 强邻居列表 N_i

WeakNbri: 弱邻居列表 N_i

SocialNbri: 基于社会关系的邻居列表 N_i

输出

能提供 v 给 N_i 的流资源节点 SP_v

if $T(v) = T_{fav}(N_i)$ then

$SP_v = \text{SearchSP}(\text{StrongNbri}, v)$;

$SP_v = SP_v + \text{SearchSP}(\text{SocialNbri}, v)$;

endif

if $T(v) \neq T_{fav}(N_i)$ then

if can find WeakNbri which $T_{fav}(\text{WeakNbri}) = T(v)$ then

$SP_v = \text{SearchSP}(\text{WeakNbri})$;

endif

$SP_v = SP_v + \text{SearchSP}(\text{SocialNbri}, v)$;

endif

if $SP_v = \Phi$ then

$SP_v = \text{SendReqtoServer}(v)$;

endif

2.2.2 查找跳数 TTL 的确定

查找策略需要重点考虑的一个问题是如何确定查找跳数 $TTL(hop)$ 。文献[4]通过大量实验数据得出在线视频分享中视频的点播次数和流行度排名基本满足 Zipf 分布。假设节点单播的视频 v 的流行度排名为 k , 而 IShare 中所有视频文件数量为 M , Zipf 分布是指特性参数为 S 。则根据 Zipf 分布的性质, 节点点播流行度排名为 k 的视频 v 的概率可表示为:

$$f(k; S, M) = \frac{1/k^S}{\sum_{n=1}^M (1/n^S)} = \frac{1}{k^S H_{M,S}} \quad (2)$$

其中, $H_{M,S}$ 为谐波参数, $H_{M,S} = \sum_{k=1}^M \frac{1}{k^S}$; 设 $S=1$ 时, $H_{M,1} = \sum_{k=1}^M \frac{1}{k} = \ln M + \gamma$; γ 是 Euler-Mascheroni 系数, 值为 0.577 215 664 9。另外, 设在线的点播节点数为 N , 那么在线节点中点播并缓存视频 v 的节点数 N_k 可粗略估计为:

$$N_k = N \cdot f(k; S, M) = \frac{N}{k(\ln M + \gamma)} \quad (3)$$

因此, 节点 i 要查找到视频 v 需要访问的节点数量为:

$$N_v = \left\lceil \frac{N}{N_k} \right\rceil = \lceil k(\ln M + \gamma) \rceil \quad (4)$$

另外, 还需要考虑的是类型为 $T(v)$ 的簇的紧密程度, 这里采用文献[5]中的方法, 对一个属于类型为 $T(v)$ 兴趣簇的节点 j 其局部簇系数可表示为:

$$CC_j = \frac{|E(\Gamma_j)|}{C_{Nb_j}^2}$$

其中, Γ_j 为节点 j 的邻居集合; $|\Gamma_j|$ 为节点 j 的邻居的总数; $|E(\Gamma_j)|$ 表示 Γ_j 集合中实际的连接总数; $C_{Nb_j}^2$ 表示 Γ_j 集合中可能产生的连接总数; CC_j 越大表示簇的紧密程度越大。再假设 IShare 中一个节点的强邻居节点个数为 x , 那么在 TTL 跳内可以访问到的 $T(v)$ 类型簇的节点个数为 $\sum_{i=1}^{TTL} (x')^i$, 其中, $x' = x \cdot (1 - CC_j)$ 。所以有:

$$\sum_{i=1}^{TTL} (x')^i \geq N_v \quad (5)$$

$$TTL \geq \log_x \left(\frac{(x'-1)N_v}{x'} + 1 \right) \quad (6)$$

IShare 通过式(3)~式(6)的计算得到一个流行度排名为 k 的视频 v 的查找跳数 TTL 。

3 实验仿真与分析

本文采用 OverSim^[6] 系统对 IShare 进行仿真, 比较前后 2 次点播的视频节目无强关联性时的查找效率, 如图 2 所示。

通过图 2 的对比可以看到 IShare 的查找效率显著优于 NetTube, 在这种情况下 IShare 的平均查找效率基本达到了 60%, 而部分情况下 NetTube 却基本无法直接通过 P2P 的方式找到源节点而必须借助于服务器的帮助, 而这显然又增加了服务器的压力。

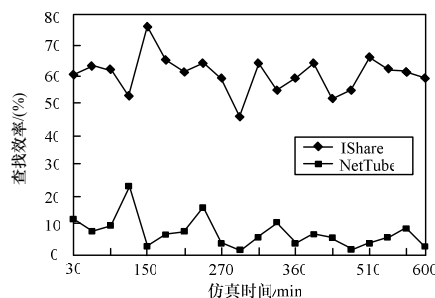


图2 视频关联度较小时查找视频源节点的效率

实验对服务器的带宽消耗也进行了比较,对比了 IShare、NetTube 和 C/S 架构时视频服务器的带宽消耗。从图 3 可以看到, C/S 架构的带宽消耗最大,在同样数量的用户点播请求下接近 4 Gb/s。IShare 比 NetTube 有了更好的改善,能够节约 75% 的带宽消耗,这是通过进一步增加节点间的数据共享实现的。

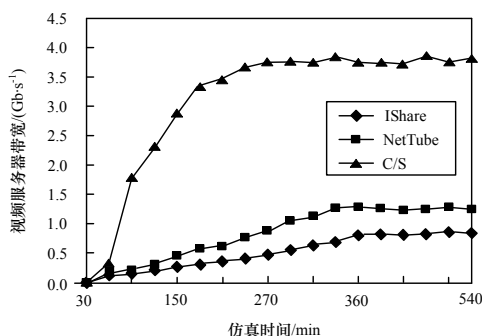


图3 视频服务器的带宽对比

4 结束语

本文提出了一种基于 P2P 的在线视频分享策略,通过实验仿真表明了 IShare 比现有系统的性能优势,下一步将继续通过现实系统的部署完善和验证 IShare。

参考文献

- [1] Gill P, Arlitt M, Li Zongpeng, et al. YouTube Traffic Characterization: A View from the Edge[C]//Proc. of the 7th ACM Conference on Internet Measurement. San Diego, USA: ACM Press, 2007.
- [2] 苏少伟, 王劲林, 尤佳莉. 一种带宽自适应 P2P 视频点播数据调度策略[J]. 计算机工程, 2011, 37(1): 13-15.
- [3] Yan Huang, Fu T Z J, Chiu Dah-Ming. Challenges, Design and Analysis of a Large-scale P2P-VoD System[C]//Proc. of ACM Conference on Data Communication. Seattle, USA: ACM Press, 2008.
- [4] Xu Cheng, Liu Jiangchuan. Net Tube: Exploring Social Networks for Peer-to-Peer Short Video Sharing[C]//Proc. of IEEE INFOCOM'09. [S. l.]: IEEE Press, 2009: 1152-1160.
- [5] Hui K Y K, Lui J C S, Yau D K Y. Small-world Overlay P2P Networks: Construction and Handling Dynamic Flash Crowd[J]. Computer Networks, 2006, 50(15): 2727-2746.
- [6] Baumgart I, Heep B, Krause S, et al. OverSim: A Scalable and Flexible Overlay Framework for Simulation and Real Network Applications[C]//Proc. of the 9th International Conference on Peer-to-Peer Computing. Berlin, Germany: [s. n.], 2009: 87-88.

编辑 顾逸斐

(上接第 272 页)

从图 4 可以看出,在信号数量不同的情况下,OKCPA 算法的迭代次数仍然明显少于 SAKM 算法。由此可知,无论是信号方差变化还是信号数量变化的情况下,OKCPA 算法都比 SAKM 算法有更少的迭代次数。

4 结束语

本文提出一种新的变参数在线核聚类算法,并将其应用于雷达辐射源信号分选中。该算法基于支持向量机的思想,采用核映射技术将数据映射到高维线性空间,因此,可以用线性方式处理任何复杂分布形状的数据,并利用随机梯度下降法更新类的边界函数,使雷达辐射源信号可以实现在线地分选,同时更新步长和惩罚项参数可以随着信号的到来而动态地调整,加快聚类分选的速度。仿真实验结果证明,该算法无论在分选准确率还是在处理速度上,都达到了较好的水平。

参考文献

- [1] 郭杰, 陈军文. 一种处理未知雷达信号的聚类分选方法[J]. 系统工程与电子技术, 2006, 28(6): 853-856.
- [2] 张万军, 樊甫华, 谭营. 聚类方法在雷达信号分选中的应用[J]. 雷达科学与技术, 2004, 2(4): 219-223.
- [3] Guo Qiang, Zhang Xingzhou, Li Zheng. SVC & K-Means and Type-entropy Based DeInterleaving/Recognition System of Radar

Pulses[C]//Proc. of International Conference on Information Acquisition. Weihai, China: IEEE Press, 2006: 742-747.

- [4] 陈彬, 骆鲁秦, 赵贵喜. 基于核模糊聚类的雷达信号分选算法[J]. 舰船电子对抗, 2009, 32(4): 76-79.
- [5] 普运伟, 朱明, 金炜东, 等. 核聚类算法最佳聚类数的自适应确定方法[J]. 计算机工程, 2007, 33(4): 11-13.
- [6] Amadou-Boubacar H, Lecoecue S. SAKM: A Kernel-based Algorithm for Online Clustering[J]. Neural Networks, 2008, 21(9): 1287-1301.
- [7] Kivinen J, Smola A, Williamson R. Online Learning with Kernels[J]. IEEE Trans. on Signal Processing, 2004, 52(8): 2165-2176.
- [8] Scholkopf B, Platt J, Taylor J S, et al. Estimating the Support of a High-dimensional Distribution[J]. Neural Computation, 2001, 13(7): 1443-1471.
- [9] Vapnik V. An Overview of Statistical Learning Theory[J]. IEEE Trans. on Neural Networks, 1999, 10(5): 988-999.
- [10] Vishwanathan S V N, Schraudolph N N, Smola A J. Step Size Adaptation in Reproducing Kernel Hilbert Space[J]. Journal of Machine Learning Research, 2006, 7: 1107-1133.

编辑 陆燕菲