

Fig.4 Performance of OCSM with varying  $K$

图 4 不同  $K$  值对 OCSM 的影响

通过图 4 的实验结果,本文得出如下结论.

- (1) 随着  $K$  值的不断增加,数据项样本聚类生成的子类类别数呈下降趋势;
- (2) 在所有请求截止期和数据请求分布组合下,AAT 随  $K$  值的增大先保持稳定,随后呈现上升趋势,最后逐渐趋于稳定.这是因为随着  $K$  值的增大,样本的聚类精度变差,系统广播调度效果受到聚类精度的影响,从而导致平均等待时间变长;
- (3) 当  $K$  值处于 10~30 区间时,请求失效率呈现轻微下降趋势;当  $K$  值处于 35~45 区间时,请求失效率迅速增大,然后随着  $K$  值的增大,失效率基本保持在较高的服务水平状态.

通过上面的分析,本文得出最优  $K$  值以及对应信道划分  $N$  值见表 6.

**Table 6** Optimal  $K$  and number of sub-channels

**表 6** 最优  $K$  值与对应信道划分  $N$  值

截止期分布	数据请求分布	最佳 $K$ 值	信道个数 $N$
固定分布	正态分布	22	6
固定分布	均匀分布	17	4
固定分布	Zipf 分布	20	6
均匀分布	正态分布	11	4
均匀分布	均匀分布	18	6
均匀分布	Zipf 分布	25	7
指数分布	正态分布	24	5
指数分布	均匀分布	20	6
指数分布	Zipf 分布	17	6

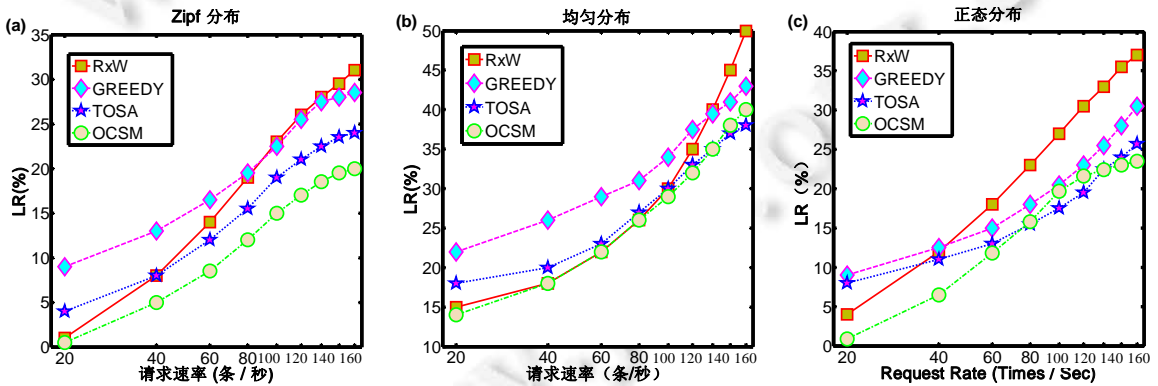
**4.3 不同数据请求分布下OCSM有效性验证**

本组实验对比在数据请求服从不同分布(Zipf分布、均匀分布、正态分布)下,OCSM,R×W,GREEDY 和 TOSA 算法的效率,评价指标采用 LR.理论上数据请求越集中,广播效率越高.因为请求越集中,广播单个数据项所满足

的用户请求数越多.数据请求在 Zipf 分布下的概率密度为  $P_i = \frac{(1/i)^\theta}{\sum_{i=1}^{20000} (1/i)^\theta}$ ,其中, $P_i$ 为数据项  $d_i$ 被访问的概率,设定

参数 $\theta=1$ .数据请求均匀分布设定平均值为 10 000,数据正态分布为标准正态分布,即 $\mu=0,\sigma^2=1$ .对于不考虑数据项大小的算法 R×W,GREEDY,数据项大小设定为 1 个单位,其他对比算法设定数据项大小服从均匀分布,平均大小为 1 个单位.请求截止期服从均匀分布,平均大小为 60 秒.数据请求服从 Zipf 分布、均匀分布、正态分布时对应的  $K$  值分别设定为 25,18,11.

图 5(a)~图 5(c)分别展示了数据请求服从 Zipf 分布、均匀分布、正态分布时,请求速率变化对各对比算法的影响.



**Fig.5** LR under different data request distributions

**图 5** 数据请求服从不同分布下各算法的失效率

从图中可以得出如下结论.

- (1) 4 种对比算法的失效率都随着请求速率的增大而升高.其中,两种静态调度算法 GREEDY 和 TOSA 在请求速率较低时失效率高于 R×W 和 OCSM.这是因为静态调度算法所广播的数据不是由当前用户请求决定,用户必须等待系统广播.当请求速率较高时,用户所请求的数据覆盖了大部分静态广播内容,此时,静态多信道广播算法优势明显;
- (2) 当请求服从 Zipf 分布情况下,请求热点相对于其余两种分布更集中,请求数据特征明显,4 种算法的表现都比较好.其中,OCSM 表现最好,这也证实了在数据特征明显的情况下,OCSM 中的聚类效果能够

较准确地找到适合广播的信道划分方式;

- (3) 当数据请求服从正态分布、请求速率大于 60 条/s 时, GREEDY 和 TOSA 失效率低于  $R \times W$  算法. TOSA 失效率一直优于 GREEDY 算法, 这得益于 TOSA 两层优化模式. 同其余 3 种算法相比, OCSM 算法表现优于  $R \times W$  和 GREEDY 算法, 在请求速率较低时优于 TOSA; 而随着请求速率的增大, 具有与 TOSA 算法表现相近. 这是因为当请求服从正态分布时, 请求数据具有较大的偏向性, 这种情况更能发挥 OCSM 自适应划分信道的优势, 从而采用更适合的信道大小广播数据;
- (4) 当数据请求服从均匀分布时, 可以看出, 所有算法的失效率普遍偏高. 当请求服从均匀分布时, 请求数据相对分散, 即热点数据不集中, 这种情况对数据广播调度十分不利. 同时也可以发现: 静态广播算法由于受请求分布影响较小, 随着请求速率增大, 失效率增长速度低于按需调度算法. OCSM 在请求速率较低时同 TOSA,  $R \times W$  算法表现相近, 而当请求速率较高时, 其表现劣于 TOSA 算法. 这是因为热点数据分散, 聚类算法不能发挥作用;
- (5) 通过对比 3 种情况下各算法的表现可以发现, 请求热点数据相对集中有利于数据广播调度. 因为热点数据集中, 广播一个热点数据所服务的用户量增多, 从而提高系统调度效率. 同时也可以发现, 静态调度算法在请求速率较低时表现不如按需调度算法, 但其受速率变化影响较小, 在请求速率较高时仍能保持较低的失效率. 在这 3 种请求分布下, OCSM 算法综合表现较好, 尤其是在请求数据特征明显的 Zipf 分布时, 请求失效率相对于 TOSA 低接近 5%; 在其余两种情况下, 也同 TOSA 算法保持相近的失效率.

#### 4.4 不同截止期分布下 OCSM 有效性验证

截止期是影响调度算法效率的重要因素之一, 截止期越短, 满足该请求的可能性越小, 请求失效率越高. OCSM 算法采用降低系统总即将失效请求的思路, 充分考虑请求截止期这一因素. 本组实验对比在截止期服从不同分布时, 算法 OCSM,  $R \times W$ , GREEDY, TOSA 的 LR 变化情况. 对比算法中: 不考虑数据项大小的调度算法, 数据项大小固定为 1 个单位; 考虑数据项大小的调度算法, 数据项大小服从均匀分布, 平均大小设定为 1 个单位, 并设定数据请求分布服从 Zipf 分布, 请求速率为 100 条/秒. 截止期服从均匀分布、指数分布、固定分布时  $K$  值分别设定为 18, 17, 17.

图 6(a)~图 6(c) 分别展示了截止期服从均匀分布、指数分布和固定分布时, 截止期大小变化对各对比调度算法的影响.

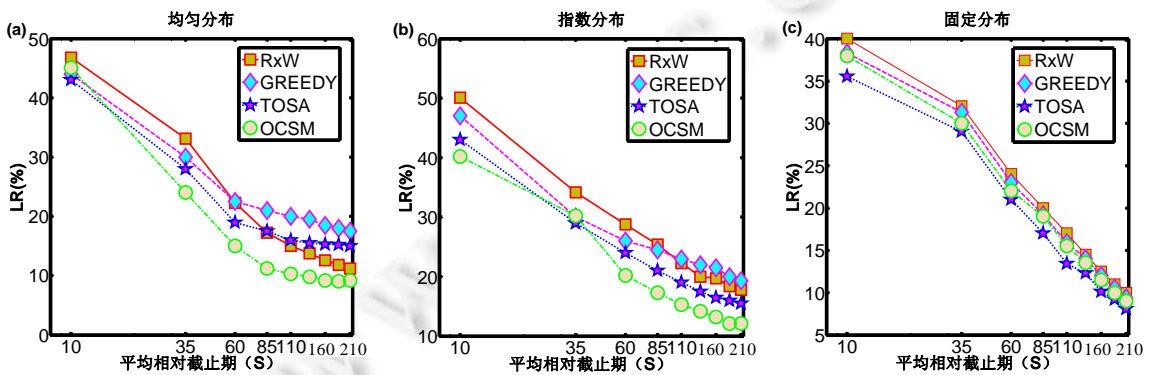


Fig.6 LR under different deadline distributions

图 6 截止期服从不同分布情况下各对比算法失效率

从图中可以得出如下结论.

- (1) 不同分布情况下, 各对比算法请求失效率随着截止期增大的变化趋势基本相似, 即: 随着截止期不断增大, 各算法的请求失效率降低. 当截止期处于区间 10~85 时, 请求失效率下降迅速; 当截止期处于 85

到 210 区间时,所有算法的失效率逐渐趋于平缓.在请求截止期较短的情况下,4 种算法的表现相近,失效率都在 40% 以上;而随着截止期的延长,两种静态调度算法 GREEDY 和 TOSA 的失效率下降速度低于按需调度算法  $R \times W$  和 OCSM.这是因为静态调度算法广播数据项是由系统根据广播之前得到的数据热度来决定的,其中有一部分数据并不会被广播,因此即使截止期继续延长,静态调度算法也会保持较高的失效率.OCSM 将截止期因素考虑进去,能够更加准确地衡量一个数据项的紧急度,因此能够在截止期服从不同分布的情况下都取得比  $R \times W$  算法更优的失效率,后者仅考虑了数据请求数以及最长等待时间;

- (2) 当截止期服从均匀分布时, $R \times W$  和 OCSM 在区间 10~60 内,LR 下降速度明显快于 GREEDY 和 TOSA.其中:OCSM 表现最好,一直处于请求失效率最低的状态;而  $R \times W$  算法在截止期大于 85 以后表现仅次于 OCSM.两个静态调度算法 GREEDY,TOSA 则一直保持较高的失效率,其中,TOSA 算法略优于 GREEDY;
- (3) 当截止期服从指数分布时,请求截止出现的概率随着截止期长度的增大而减小,故截止期分布相对集中于较小的一端,4 种对比算法在这种情况下都表现最差.而由于截止期分布较集中,特征明显,有利于 OCSM 中聚类的准确性,OCSM 在这种情况下仍然能够取得很好的表现,请求失效率明显低于其余 3 种算法;
- (4) 当截止期服从固定分布时,OCSM 算法表现跟  $R \times W$  算法相近,因为所有请求截止期都相同,这使得考虑了截止期因素的 OCSM 算法相对于不考虑截止期因素的  $R \times W$  算法的优势不突出.然而,得益于 OCSM 算法的自适应多信道广播,其表现还是稍微优于  $R \times W$  算法.因为截止期固定,使得在截止期较小的 10~60 区间内,各算法的失效率基本呈梯度下降趋势.这说明在截止期较小时,截止期是影响失效率的主要因素;但随着截止期的增大,各算法失效率降低速度逐渐放缓.在这种情况下,TOSA 算法的失效率略低于 OCSM 算法,说明固定截止期更适合静态调度算法.

综上,通过在截止期服从均匀分布、指数分布、固定分布时,4 种算法请求失效率的对比,验证了 OCSM 算法的具有较好的有效性.当截止期服从均匀分布以及指数分布时,OCSM 算法比  $R \times W$ ,GREEDY 和 TOSA 算法表现更优;当截止期服从固定分布时,OCSM 算法表现接近 TOSA.

#### 4.5 数据项大小变化OCSM有效性验证

数据项大小是影响广播调度的另一个重要因素,数据项越大,单位时间广播的数据越少,所能满足的请求越少,请求失效率越高.OCSM 算法综合运用数据项大小决定广播信道大小的思想,自适应寻求更适合的信道大小.本组实验对比了 OCSM 同  $R \times W$ ,GREEDY,TOSA 算法在数据项大小服从固定分布下的算法效率.设定数据请求服从 Zipf 分布,请求速率为 100 条/秒,请求截止期服从均匀分布,截止期平均大小为 60 秒,K 值设定为 18(如图 7 所示).

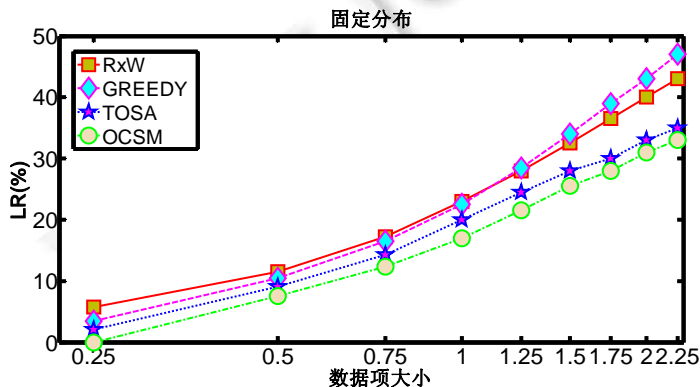


Fig.7 LR under a fixed distribution of data item size  
图 7 数据项大小服从固定分布大小变化对调度的影响

从图 7 得出如下结论.

- (1) 由于  $R \times W$ , GREEDY 算法没有考虑数据项大小的因素,其表现较差,LR 随着数据项的递增,几乎呈线性递增的趋势.其中, GREEDY 算法受数据项大小的影响最大,随着数据项的增大而广播周期变长,从而导致大量截止期较短的请求得不到满足而失效;
- (2) TOSA 和 OCSM 算法考虑到数据项单位大小的广播价值,优先广播单位价值高的数据,从而随着数据项大小的增加,仍能取得较低的失效率,其中, OCSM 算法表现一直优于 TOSA.

## 5 结论与展望

为了克服现有固定信道数据广播调度的不足,进一步提高实时按需数据广播的调度效率,本文提出了一种基于实时按需数据广播调度的信道划分与分配方法 OCSM. OCSM 首先通过  $R \times W/SL$  算法确定数据项权重值;然后利用 WASC 算法根据数据项权重值以及数据项大小对数据进行均衡聚类;最后利用 CSA 算法划分信道,组织数据进行广播调度.为了验证 OCSM 的性能,本文模拟现实广播环境,进行了大量的对比实验,并得出以下结论.

- (1) 相对于单信道按需数据广播调度算法  $R \times W$ , OCSM 提出的自适应信道划分策略能够充分利用多信道并行广播的特性,更好地降低失效率;
- (2) WASC 算法在请求数据倾斜程度较大时表现更好,因为此时的数据特征明显,使得 WASC 算法聚类更准确;
- (3) 相对于多信道数据广播调度算法 GREEDY 和 TOSA, OCSM 具有更强的自适应性,在不同的情况下都有较好的表现.

本文在如下几个方面需要进一步展开研究:(1) 在自适应信道划分与分配的基础上,研究索引策略,进一步缩短用户请求的调谐时间;(2) 考虑用户请求包含多个数据项的情况,研究相应的信道划分算法.

**致谢** 我们真诚地向对本文的工作给予支持和建议的审稿人、主编、编辑、同行、同学和老师表示感谢.

## References:

- [1] Wang Q, Wu SJ, Li MS. Software defect prediction. Ruan Jian Xue Bao/Journal of Software, 2008,19(7):1565–1580 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/1565.htm>
- [2] Li LJ, Liu HF, Yang ZY, Huang XY. Broadcasting methods in vehicular ad hoc networks. Ruan Jian Xue Bao/Journal of Software, 2010,21(7):1620–1634 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3845.htm> [doi: 10.3724/SP.J.1001.2010.03845]
- [3] Xuan P, Sen S, Gonzalez O, Fernandez J. Broadcast on demand: Efficient and timely dissemination of data in mobile environments. Real-Time Technology and Applications Symp., 1997,4(5):38–48.
- [4] Jung H, Chung Y, Liu L. Processing generalized  $k$ -nearest neighbor queries on a wireless broadcast stream. Information Sciences, 2012,188(4):64–79.
- [5] Yu C, Yao D, Li X, Zhang Y, Yang L, Xiong N, Jin H. Location-Aware private service discovery in pervasive computing environment. Information Sciences, 2012,230(5):78–93.
- [6] Simunic T, Boyd S, Glynn P. Managing power consumption in networks on chips. IEEE Trans. on Very Large Scale Integration Systems, 2004,12(1):96–107.
- [7] Sun W, Qin Y, Wu J, *et al.* Air indexing for on-demand XML data broadcast. IEEE Trans. on Parallel and Distributed Systems, 2014,25(6):1371–1381.
- [8] Zhao RQ, Liu ZJ, Wen AJ. An efficient energy-saving broadcast mechanism for wireless sensor networks. Chinese Journal of Electronics, 2009,37(11):2457–2462 (in Chinese with English abstract).

- [9] Xuan P, Sen S, Gonzalez O, Fernandez J, Ramamritham K. Broadcast on demand: Efficient and timely dissemination of data in mobile environments. In: Proc. of the 3rd IEEE Real-Time Technology and Applications Symp. Montreal: IEEE Real-Time Technology and Applications Symp., Vol.38. 1997. 38–48.
- [10] Lei M, Vrbsky SV, Xiao Y. Scheduling on-demand data broadcast in mixed-type request environments. *Computer Networks*, 2010, 54(5):811–825.
- [11] Kalyanasundaram B, Velauthapillai M. On-Demand broadcasting under deadline. In: Proc. of the 11th Annual European Symp. on Algorithms, Vol.2832. Berlin, Heidelberg: Springer-Verlag, 2003. 313–324.
- [12] Ng J, Lee V, Hui C. Client-Side caching strategies and on-Demand broadcast algorithms for real-time information dispatch system. *IEEE Trans. on Broadcasting*, 2008,54(1):24–35.
- [13] Lu Z, Wu W, Li W W, *et al.* Efficient scheduling algorithms for on-demand wireless data broadcast. In: Proc. of the IEEE Int'l Conf. on Computer Communications. IEEE Infocom, 2016. 1–9.
- [14] Dewri R, Whitley D, Ray I, *et al.* Optimizing real-time ordered-data broadcasts in pervasive environments using evolution strategy. In: Proc. of the Parallel Problem Solving From Nature (PPSN X). LNCS, DBLP, 2008. 991–1000.
- [15] Liu CM, Su TC. Broadcasting on-demand data with time constraints using multiple channels in wireless broadcast environments. *Information Sciences*, 2013,242(2):76–91.
- [16] Aksoy D, Franklin M.  $R \times W$ : A scheduling approach for largescale on-demand data broadcast. *IEEE/ACM Trans. on Networking*, 1999,7(6):846–860.
- [17] Wu X, Lee VCS. Wireless real-time on-demand data broadcast scheduling with dual deadlines. *Journal of Parallel and Distributed Computing*, 2005,65(6):714–728.
- [18] Hu W, Fan C, Luo J, *et al.* An on-demand data broadcasting scheduling algorithm based on dynamic index strategy. *Wireless Communications and Mobile Computing*, 2015,15(5):947–965.
- [19] Hu W, Xia C, Du B, *et al.* An on-demand data broadcasting scheduling considering the data item size. *Wireless Networks*, 2015, 21(1):35–56.
- [20] Saxena N, Pinotiti M. On-Line balanced  $k$ -channel data allocation with hybrid schedule per channel. In: Proc. of the 6th Int'l Conf. on Mobile Data Management. ACM Press, 2005. 239–246.
- [21] Yee WG, Navathe SB, Omiecinski E, Jermaine C. Efficient data allocation over multiple channels at broadcast servers. *IEEE Trans. on Computers*, 2002,52(10):1231–1236.
- [22] Waluyo A, Srinivasan B, Taniar D, Rahayu W. Incorporating global index with data placement scheme for multi channels mobile broadcast environment. In: Proc. of the Embedded and Ubiquitous Computing (EUC 2005). Berlin Heidelberg: Springer-Verlag, 2005. 755–764.
- [23] Lee W, Hu Q, Lee D. A study on channel allocation for data dissemination in mobile computing environments. *Mobile Networks and Applications*, 1999,4(2):117–129.
- [24] Gao X, Yang Y, Chen G, Lu X, Zhong HF. Global optimization for multi-channel wireless data broadcast with ah-tree indexing scheme. *IEEE Trans. on Computers*, 2015,65(7):2104–2117.
- [25] Lim JH, Naito K, Yun JH, Gerla M. Revisiting overlapped channels: Efficient broadcast in multi-channel wireless networks. In: Proc. of the 2015 IEEE Conf. on Computer Communications (INFOCOM). IEEE, 2015. 1984–1992.
- [26] Nawaz Ali GGM, Lee VCS, Chan E, Li M, Liu K, Lv JS, Chen J. Admission control-based multichannel data broadcasting for real-time multi-item queries. *IEEE Trans. on Broadcasting*, 2014,60(4):589–605.
- [27] Hu C, Chen M. Adaptive balanced hybrid data delivery for multi-channel data broadcast. In: Proc. of the IEEE Int'l Conf. on Communications 2002. IEEE, 2002. 960–964.
- [28] Zheng B, Wu X, Jin X, *et al.* TOSA: A near-optimal scheduling algorithm for multi-channel data broadcast. In: Proc. of the 6th Int'l Conf. on Mobile Data Management. ACM Press, 2005. 20–37.
- [29] Hu W, Qiu Z, Wang H, *et al.* A real-time scheduling algorithm for on-demand wireless XML data broadcasting. *Journal of Network and Computer Applications*, 2016,68:151–163.
- [30] He P, Shen H, Tian H. On-Demand data broadcast with deadlines for avoiding conflicts in wireless networks. *Journal of Systems and Software*, 2015,103(C):118–127.



- [31] Cacciatore S, Luchinat C, Leonardo T. Knowledge discovery by accuracy maximization. In: Proc. of the National Academy of Sciences of the United States of America. 2014.
- [32] World Cup 98 Web site access logs. 1998. <http://ita.ee.lbl.gov/html/contrib/WorldCup.html>
- [33] Lu F, Du N, Wen CL. A fuzzy-evidential  $k$  nearest neighbor classification algorithm. Chinese Journal of Electronics, 2012,40(12): 2390–2395 (in Chinese with English abstract).

#### 附中参考文献:

- [1] 王青,伍书剑,李明树.软件缺陷预测技术.软件学报,2008,19(7):1565–1580. <http://www.jos.org.cn/1000-9825/19/1565.htm>
- [2] 李丽君,刘鸿飞,杨祖元,黄席樾.车用自组网信息广播.软件学报,2010,21(7):1620–1634. <http://www.jos.org.cn/1000-9825/3845.htm> [doi: 10.3724/SP.J.1001.2010.03845]
- [8] 赵瑞琴,刘增基,文爱军.一种适用于无线传感器网络的高效节能广播机制.电子学报,2009,37(11):2457–2462.
- [33] 吕锋,杜妮,文成林.一种模糊-证据  $k$ NN 分类方法.电子学报,2012,40(12):2390–2395.



胡文斌(1976—),男,湖北武汉人,博士,副教授,博士生导师,主要研究领域为人工智能与智能仿真优化,社会网络仿真与优化,大数据分析 with 优化理论,智能计算与智能决策,智能交通系统.



王欢(1989—),男,博士生,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



邱振宇(1992—),男,硕士生,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



严丽平(1980—),女,博士生,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



聂聪(1993—),男,硕士生,主要研究领域为智能仿真与优化.



杜博(1983—),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为人工智能,图像分析.