P.O.Box 8718, Beijing 100080, China	Journal of Software, Feb. 2004,15(2):159-169
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2004 by The Editorial Department of Journal of Software

PRAM和LARPBS模型上的近似串匹配并行算法

钟 诚, 陈国良

钟 诚, 陈国良 (中国科学技术大学 计算机科学与技术系,安徽 合肥 230027)(国家高性能计算中心,安徽 合肥 230027)
作者简介: 钟诚(1964－),男,广西桂平人,博士,教授,主要研究领域为并行算法,网络信息安全,生物信息处理算法;陈国良(1938－),男,教授,博士生导师,中国科学院院士,主要研究领域为并行算法,并行计算.
联系人: 钟 诚 Phn: +86-771-3272685, E-mail: chzhong6517@sina.com, ciugxcn@163.com
Received 2002-12-05; Accepted 2003-09-26

Abstract
Approximate string matching technique has been widely applied to many fields such as network information retrieval, digital library, pattern recognition, text mining, IP routing searching, network intrusion detection, bioinformatics, and computing in musicology. The two concise parallel dynamic programming algorithms for approximate string matching with k-differences on CREW-PRAM (parallel random access machine with concurrent read and exclusive write) are presented by parallel computing the edition distance matrix D in the wave-front approach and by parallel computing the edition distance matrix D along the diagonal and horizontal directions, respectively. The first algorithm requires O(n) time and obtains a linear speedup by (m+1) processors. The second algorithm needs O(n/(+m) time in use of ((m+1) processors, where 1<(. The simulation experiment shows that both of the algorithms achieve a good speedup. Based on the divide-and-conquer strategy and split-merge approach, the two communication-efficient and scalable parallel algorithms for approximate string matching with k-mismatches on linear arrays with reconfigurable pipelined bus system (LARPBS) are presented by reconfiguring dynamically the optical bus system, applying neatly the optical message-passing technique, and computing in parallel the prefix sums. The first algorithm requires O(m) time by n processors and the second one has O(1) time in use of nm processors.

Zhong C, Chen GL. Parallel algorithms for approximate string matching on PRAM and LARPBS. *Journal of Software*, 2004,15(2):159~169.
http://www.jos.org.cn/1000-9825/15/159.htm

摘要
近似串匹配技术在网络信息搜索、数字图书馆、模式识别、文本挖掘、IP路由查找、网络入侵检测、生物信息学、音乐研究计算等领域具有广泛的应用.基于CREW-PRAM(parallel random access machine with concurrent read and exclusive write)模型,采用波前式并行推进的方法直接计算编辑距离矩阵D,设计了一个允许k-差别的近似串匹配动态规划并行算法,该算法使用(m+1)个处理器,时间复杂度为O(n),算法理论上达到线性加速;采取水平和斜向双并行计算编辑距离矩阵D的方法,设计了一个使用((m+1)个处理器和O(n/(+m)时间的、可伸缩的、允许k-差别的近似串匹配动态规划并行算法,.基于分治策略,通过灵活拆分总线和合并子总线动态重构光总线系统,并充分利用光总线的消息播送技术和并行计算前缀和的方法,实现了汉明距离的并行计算,设计了两个基于LARPBS(linear arrays with reconfigurable pipelined bus system)模型的通信高效、可扩放的允许k-误配的近似串匹配并行算法,其中一个算法使用n个处理器,时间为O(m);另一个为常数时间算法,使用mn个处理器.

基金项目：Supported by the National Grand Fundamental Research 973 Program of China under Grant No.G1998030403 (国家重点基础研究发展规划(973)); the National High-Tech Research and Development Plan of China under Grant No.2001AA111041 (国家高技术研究发展计划(863))

References:

[1] Navarro G. A guided tour to approximate string matching. ACM Computing Surveys, 2001,33(1):31~88.

[2] Navarro G, Baeza-Yates R. A hybrid indexing method for approximate string matching. Journal of Discrete Algorithms, 2000,1(1): 21~49.

[3] Lee H-C, Ercal F. RMESH algorithms for parallel string matching. In: Proc. of the 3rd Int'l. Symp. on Parallel Architectures, Algorithms, and Networks (I-SPAN'97). Los Alamitos: IEEE Computer Society Press, 1997. 223~226. http://ieeexplore.ieee.org/ xpl/tocresult.jsp?isNumber=13955&page=2.

[4] Jiang Y, Wright AH. O(k) parallel algorithms for approximate string matching. Journal of Neural Parallel and Scientific Computation, 1993,1:443~452.

[5] Amir A, Landau GM. Fast parallel and serial multidimensional approximate array matching. Theoretical Computer Science, 1991, 81 (1):97~115.

[6] Alan AB, Mei A. A residue number system on reconfigurable mesh with applications to prefix sums and approximate string matching. IEEE Trans. on Parallel and Distributed Systems, 2000,11(11):1186~1199.

[7] Park JH, George KM. Efficient parallel hardware algorithms for string matching. Microprocessors and Microsystems, 1999,23(3): 155~168.

[8] Lester B. The Art of Parallel Programming. Englewood Cliffs: Prentice Hall, 1993.

[9] Pan Y, Li Y, Li J, LI K, Zheng SQ. Efficient parallel algorithms for distance maps of 2-D binary images using an optical bus. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2002,32(2):228~236.

[10] Han Y, Pan Y, Shen H. Sublogarithmic deterministic selection on arrays with a reconfigurable optical bus. IEEE Trans. on Computers, 2002,51(6):702~706.