P ONTIFÍCIA U NIVERSIDADE C ATÓLICA
DO RIO DE JANEIRO

Marco Serpa Molinaro

Improved Approximations for the k-Hotlink Assignment Problem and for Binary Searching in Trees

MsC Thesis

Thesis presented to the Post–graduate Program in Computer Science of the Computer Science Department, PUC–Rio as partial fulfillment of the requirements for the degree of Master in Computer Science

Adviser: Prof. Eduardo Sany Laber

Rio de Janeiro
April 2008

Marco Serpa Molinaro

Improved Approximations for the k-Hotlink Assignment Problem and for Binary Searching in Trees

Thesis presented to the Post–graduate Program in Computer Science of the Computer Science Department, PUC–Rio as partial fulfillment of the requirements for the degree of Master in Computer Science. Approved by the following commission:

Prof. Eduardo Sany Laber
Adviser
Department of Computer Science — PUC–Rio

Prof. Marcus Poggi de Aragão
Departamento de Informática — PUC-RIO

Prof. Valmir Carneiro Barbosa
COPPE — UFRJ

Prof. Jayme Luiz Szwarcfiter
COPPE — UFRJ

Prof. José Eugenio Leal
Head of the Science and Engineering Center — PUC–Rio

Rio de Janeiro — April 28, 2008

Marco Serpa Molinaro

Marco Molinaro graduated from Pontifícia Universidade Católica do Rio de Janeiro in Computer Engineering. During his undergraduate years, he participated in multiple research projects funded by CNPq. For the duration of his Masters, he was supported by an academic scholarship and performed theoretical research focusing on approximation algorithms.

Ficha Catalográfica

Acknowledgments

Abstract

Molinaro, Marco Serpa; Laber, Eduardo Sany. **Improved Approximations for the k-Hotlink Assignment Problem and for Binary Searching in Trees**. Rio de Janeiro, 2008. 89p. MsC Thesis — Department of Computer Science, Pontifícia Universidade Católica do Rio de Janeiro.

Here we present a study on two optimization problems in trees: the k-Hotlink Assignment Problem and the problem of Binary Searching in Trees. As a result, we obtain improved approximation algorithms for both problem. The k-Hotlink Assignment Problem can be defined as follows. Let $G = (V, E)$ be a directed acyclic graph representing a web site, where nodes correspond to pages and arcs to hyperlinks. In this context, hotlinks are defined as shortcuts (new arcs) added to web pages of G in order to reduce the time spent by users to reach their desired information. Here we consider the problem where G is a rooted directed tree and the goal is minimizing the expected time spent by users by assigning at most k hotlinks to each node. For the most studied version of this problem where at most one hotlink can be assigned from each node, we prove the existence of an FPTAS, improving upon the constant factor algorithm recently obtained in [Jacobs, WADS 2007]. In addition, we develop the first constant factor approximation algorithm for the most general version where k hotlinks can be assigned from each node.

In the second part of this work, we consider the problem of computing efficient strategies for searching in trees. As a generalization of the classical binary search for ordered lists, suppose one wishes to find a (unknown) specific node of a tree by asking queries to its arcs, where each query indicates the endpoint closer to the desired node. Given the likelihood of each node being the one searched, the objective is to compute a search strategy that minimizes the expected number of queries. Practical applications of this problem include file system synchronization and software testing. Here we present a linear time algorithm which is the first constant factor approximation for this problem. This represents a significant improvement over previous $O(\log n)$-approximation.

Keywords

Approximation Algorithms.   Hotlinks.   Search.   Trees.

Resumo

Molinaro, Marco Serpa; Laber, Eduardo Sany. **Improved Approximations for the k-Hotlink Assignment Problem and for Binary Searching in Trees**. Rio de Janeiro, 2008. 89p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese apresenta um estudo sobre dois problemas de otimização em árvores: o *k-Hotlink Assignment Problem* e o problema de busca binária em árvores. Como resultado, obtemos algoritmos aproximativos melhorados para ambos os problemas.

O *k-Hotlink Assignment Problem* pode ser definido da seguinte maneira. Seja $G = (V, E)$ um dígrafo acíclico representando um web site, onde nós correspondem a páginas e arcos correspondem a hyperlinks. Nesse contexto, hotlinks são definidos como atalhos (novos arcos) adicionados a páginas em G de modo a encurtar o tempo gasto por usuários para alcançar as informações desejadas. Aqui consideramos o problema onde G é uma árvore enraizada e o objetivo é minimizar o tempo médio gasto por usuários adicionando no máximo k hotlinks a cada nó. Para a versão mais estudada desse problema onde no máximo um hotlink pode ser atribuído a cada nó, nós provamos a existência de um FPTAS, uma melhora sobre a aproximação de fator constante obtida recentemente em [Jacobs, WADS 2007]. Além disso, desenvolvemos a primeira aproximação de fator constante para a versão mais geral onde k hotlniks podem ser atribuídos a cada nó.

Na segunda parte deste trabalho, consideramos o problema de computar eficientemente estratégias para buscar em árvores. Como uma generalização da clássica busca binária em listas ordenadas, suponha que deseja-se encontrar um nó específico (e desconhecido) de uma árvore através de consultas a seus arcos, onde cada consulta indica qual extremidade do arco está mais próxima ao nó desejado. Dada a probabilidade de cada nó ser aquele procurado, o objetivo é encontrar uma estratégia que minimize o número esperado de consultas. Aplicações práticas desse problema incluem sincronização de sistemas de arquivo e criação de testes de software. Apresentamos um algoritmo que executa em tempo linear e que provê a primeira aproximação de fator constante para esse problema. Esse resultado representa uma significante melhora sobre a $O(\log n)$-aproximação conhecida anteriormente.

Palavras–chave

Algoritmos Aproximativos. Hotlinks. Busca. Árvores.

Contents

Contents