ABSTRACT

Maximum flows in distribution networks

by
Ju-Chun Lin

The maximum flow (max-flow) problem is a fundamental network optimization problem which computes for the largest possible amount of flow sent through the network from a source node to a sink node. This problem appears in many applications and has been investigated extensively over the recent four decades. Traditional max-flow problem may require some modification in its constraints to deal with real-world applications. Fang and Qi propose a new max-flow model, named as manufacturing network flow model, to describe a network with special distillation nodes or combination nodes. Based on this new model, Ting proposes a multi-labeling method to solve the max-flow problem in a distribution network which contains both ordinary and distillation nodes. The approach identifies an augmenting subgraph connecting both source and sink nodes which can be further decomposed into several components where flows in each component can be expressed by a single variable and solved by a system of homogeneous linear equations. The method requires manual detection for components and thus is not trivial to implement. In this paper, we present a preprocessing procedure that reduces the size of a distribution network in polynomial time and then give detailed procedures for implementing the multi-labeling method. We also implement a random network generator that generates random distribution networks guaranteed to be acyclic, connected and in a compact form for our computational testing. Modification on the preflow-push algorithm and the max-flow min-cut theorem is investigated and discussed to provide more theoretical insights as well.

Key words : Maximum flow; Distribution network

Abstract

Reverse logistics has attracted a lot of research attention recently, especially when the disposition of electrical appliances increases. Different recycled appliances may consume different levels of resources and may require very different recycling processes. Conventional reverse logistics optimization model usually considers only the optimal facility locations and assumes the facility can conduct all different recycling processes. This paper investigates a mixed integer programming model that solves for both optimal facility locations and facility planning for a general reverse logistics network. In particular, in a planning stage for a reverse logistics company, our model considers both location and facility type —a facility that can only process a specific category of recycled products, or a facility that can process more categories of recycled products. Since our mathematical programming model is more complex and NP-hard, we propose several heuristic solution methods based on previous research and conduct computational experiments on several scenarios using data collected from Taiwan's reverse logistics market. The computational tests show that our proposed solution methods can give a high-quality solution in a promising time much shorter than the optimization software CPLEX.

Keywords: Reverse logistics; Facility location; Heuristics; Mixed Integer programming

ABSTRACT

Uncapacitated minimum cost flow problems in a distribution network

by
Yu-Hui Yang

In this thesis, we consider three special minimum cost flow problems in a distribution network, a kind of manufacturing network recently introduced by Fang and Qi. The network differs from a traditional network model because a new kind of nodes, called D-nodes, are incorporated to describe a distilling operation that decomposes one raw-material to several products with fixed ratios. Besides, all the arcs in our models have no upper bounds. We treat these uncapacitated minimum cost flow problems as specialized shortest path problems with side constraints in a distribution network. We give a preprocessing that compacts a distribution network to an equivalent network of smaller size, derive their mathematical formulations and develop efficient solution methods.

Keywords : Manufacturing network, distribution network, minimum cost flow problem, uncapacitated, shortest path

Abstract

The minimum cost flow problem seeks an optimal flow assignment over a network satisfying the node flow balance constraints and arc flow bounds constraints. These constraints are too simplified to model some real cases. To model the distillation or decomposition of products in some manufacturing processes, a minimum distribution cost problem (MDCP) on a specialized manufacturing network flow model has been investigated. In an MDCP, a specialized node called D-node is used to model a distillation process which only connects with a single incoming arc and several outgoing arcs. The flows entering a D-node have to be distributed according to a pre-specified ratio associated with each of its outgoing arcs. Such a proportional relationship between the arc flows associated with each D-node complicates the problem and makes the MDCP harder than conventional min-cost network flow problem. A network simplex algorithm for uncapacitated MDCP has been outlined in literature, but its detailed graphical procedures such as initial basic feasible solution computation, dual variables updates, and flow pivoting operations have not yet been given. In this thesis, we resolve these issues by upper bound techniques as well as graphical operations which decompose each pivoting graph into several components for calculating both the arc flows and the dual variables. Other issues regarding efficient ways to obtain an initial primal basic feasible solution to start with our algorithm and mathematical insights for solving the MDCP on distribution networks will also be investigated and discussed.

keywords: network optimization, minimum distribution cost problem, minimum cost flow, network simplex algorithm

ABSTRACT

The network flow problem is a specialized Linear Programming problem (LP) due to its special constraint structure. An LP solution method may have a more efficient implementation when applied for solving the network flow problem. Recently, a new primal-dual algorithm called the least-squares primal-dual (LSPD) method has been proposed to solve LPs with good performance since it guarantees nondegenerate pivoting in each iteration. In each prmial-dual iteration, the LSPD algorithm solves a nonnegative least squares (NNLS) subproblem to obtain an improving direction for its dual variables.

In this thesis, we develop techniques related with the LSPD method for solving network flow problems. Issues regarding efficient ways to obtain an initial dual feasible solution and techniques to deal with capacitated network flow problems will be investigated. We also propose a new least-squares dual-primal (LSDP) algorithm which differs from the LSPD algorithm in exchanging the roles of the primal and dual formulations. When solving for the max-flow problem, the NNLS subproblem in our LSDP algorithm can be treated as an algorithm to calculate the current on an electrical network with diodes, where the unit-flow cost associated with an arc can be treated the electrical potential. Thus the Kirchhoff's law can be applied to solve the NNLS subproblem. Similar techniques can be applied to solve the MCF problem.

keywords: network optimization, minimum cost flow problem, primal-dual algorithm, nonnegative least-squares, degeneracy, maximum flow problem

ABSTRACT

Due to the completion of Human Genome Project, we have known more about the structure and sequence, but not yet the function of human DNA. The function of DNA may be a key for human disease. To analyze the function of DNA, researchers have to obtain each haplotype, the genetic constitution of an individual chromosome, of an individual for analysis. Nevertheless, considering the significant efforts required in collecting haplotypes, usually the descriptions of one conflated pair of haplotypes called genotypes are collected.

Since the genotype data contains insufficient information to identify the combination of DNA sequence in each copy of a chromosome, one has to solve the population haplotype inference (PHI) problem which infers haplotype data from genotype data for a population. Previous researchers use mathematical programming methods and heuristic algorithms to solve the population haplotype inference problem. This thesis surveys these methods and conducts computational experiments on the efficiency and effectiveness for these methods in solving a population haplotype inference problem based on pure parsimony criterion (HIPP) which seeks the minimum number of distinct haplotypes to infer a given genotype matrix.

We propose two heuristic algorithms to solve the HIPP problem with promising performance. The first heuristic algorithm exploits the compatible relations among genotypes to solve a reduced integer linear programming problem in a smaller solution space. The second heuristic algorithm selects popular haplotypes that can resolve more genotypes in a greedy fashion. Extensive computational experiments have been

conducted for several PHI solution methods on both the biological and simulated data. The results show that our proposed algorithms are efficient and effective, especially for solving cases with larger recombination rates. Finally, we give a divde-and-concur technique to solve large-scale HIPP problems. We also improve a parsimonious tree growing heuristic to obtain all the multiple optimal solutions for an HIPP problem.

Key words : Haplotype inference, Genotype, Pure parsimony, Heuristic algorithm, Integer programming

Abstract

In order to carry out the clinical test system of the biologicaltechnology industry plan, the Executive Yuan has been actively helpinghospitals establish clinical trial management. Site ManagementOrganization (SMO) has replaced the old management model that onedoctor handles one plan at a time. SMO can effectively help schedulenurses to serve more patients and reduce the time to search forpatients. The original management model before SMO usually requireslonger time to search for patients because the sources of the patientsare scattered around places. On the other hand, SMO has more complexrules to assign nurses to the patients. Our research focuses ontechniques to search for and also serve all patients with less humanresource. According to our studies, SMO is used in the third stage ofthe clinical trial process which is also the stage of the largestcost. In order to improve the operations of SMO, one needs toefficiently decide the schedules of treatments for each patient andeach served nurse so that the cost for human resources is reduced.This optimization problem can be considered as a Nurse RosteringProblem (NRP), which is often solved by integer programming.In this thesis, we propose a mathematical model based on integerprogramming that leads to better schedules and reduces the cost ofhuman resource. In particular, several objectives including theminimization of the number of served nurses, and balancing the dailyworkload are formulated in our IP models and solve by CPLEX. However,the solution calculated by CPLEX seems to be inconvenient comparedwith the original schedule in the sense that the solution tends topartition the daily schedule for each nurse. Moreover, CPLEX consumesa lot of computational time. To construct a more

convenient schedulein a shorter time such that the original objectives are achieved as-much as possible, we propose several greedy heuristics and build arostering information system to help SMO managers design a goodschedule that not only reduces the human resource costs but alsoassigns workload more reasonably.

Keywords: clinical trial, SMO, nurse rostering problem, integerprogramming, heursitics

Abstract

In this thesis, we first propose a random project scheduling network generator by integrating additional time window or time schedule constraints with a popular project network generator (RanGen). We then give two polynomial heuristic algorithms (Greedy I and Greedy II) to solve the integer programming problems, where Greedy I exploits the special structure of the models so that it can suggest a feasible start time for each task in a topological order to reduce the objective values, while Greedy II further takes the objective weights associated with variables into consideration and improves the quality of the solutions obtained by Greedy I. To further get solutions of better qualities, we propose a genetic algorithm and conduct computational experiments to evaluate the effectiveness (optimality gap) and efficiency (running time) of our proposed algorithms (Greedy I, Greedy II, and GA) and a popular optimization solver, CPLEX. The results show that our greedy heuristics are very efficient but not as effective, GA can be both efficient and effective, while CPLEX usually consumes much more computational time and is especially inefficient for problems of larger scale.

keywords: project scheduling, stable baseline scheduling, uncertainty, genetic algorithm

Abstract

Pattern discovery is a common task in data mining. Given the transaction datasets of multi periods, we are concerned with a temporal data mining problem that detects any pattern of interested changes that have been consistent from some period to the last period. Discovering such changes from the transaction database of multi periods will help the managers to detect the tendency of customer needs so that potential customers may be identified. To the best of our knowledge, previous studies in change mining only focus on datasets of two datasets, although the tendency of changes are more meaningful for datasets of multi periods in real-world applications. Conventional data mining techniques that seek frequent patterns could be modified for mining changes from datasets of multi periods, but such approaches would require many pairwise comparisons between datasets of consecutive periods and thus not so efficient.

In this thesis, we propose an algorithm called MCP for mining changes from multi-period datasets. MCP is based on a novel data structure modified from the popular frequent-pattern tree (FP-tree), and seeks the target patterns in a very efficient way. In particular, starting from the last two periods, our algorithm first constructs a candidate-pattern forest (CP-forest) to store those patterns of qualified changes, and then iteratively updates the CP-forest using the dataset of each period. The CP-forest is carefully designed such that useless information will not be stored and qualified patterns can be easily identified by tree traversals.

Computational experiments have been conducted to compare MCP and another algorithm called modiFP which is modified from the popular FP-growth algorithm for

mining the changes of patterns from multi-period datasets. Several parameters have be used to evaluate the performance of MCP and modiFP, and the results show that MCP is much more efficient than modiFP, especially when the number of periods increases or when the datasets of consecutive periods share more similarities.

keywords: pattern discovery, change mining, candidate-pattern forest, temporal data mining, algorithm, data structure

ABSTRACT

Optimal Paths Based on Time and Fares in Transit Networks

Michael Wang

In a public transportation oriented metropolis such as Tokyo or Taipei, transit system is a convenient means for personal trip. Historically, most transit passengers rely on printed schedules for trip planning. However, the transit network nowadays has become too complicated to navigate manually. With the advance of technology, various navigation applications have been developed for guiding private vehicles, but few are designed for public transportation. Given an origin, destination, and intended departure time, this study proposes two timetable-based algorithms to search for optimal itineraries so that the total travel time for an individual is minimized in an transit network. Itineraries showing the suggested routes with walking access and egress, bus stops, and Mass Rapid Transit (MRT) information will be generated, considering the time to wait for, to transfer between, and to stay in transit vehicles, as well as the time to walk between transit stations. In addition, this study proposes an innovative fare-based algorithm to search for the cheapest itineraries with non-linear fare structure, and an alternative fare model is specifically developed for Taipei transit system. Optimizing trip planning according to time or fare meets the common practices of passengers using transit system in a metropolitan area.

Keywords: shortest path, transit network, multimodal transportation, trip planning, timetable-based algorithm, non-linear fare structure.

Abstract

Among all possible DNA sequence variations, Single Nucleotide Polymorphism（SNP）is the most common genetic variation. Sequence of closely linked SNPs composes the haplotype. Changes in the sequence can have significant influence on disease occurrences and the phenotype of human traits. The changes can be applied for identifying diseases and other medical research. At present, there are voluminous data of discovered SNPs. To make SNP database more cost-effective, many researchers propose tagging SNPs, a minimal subset of SNPs called tagSNP, to capture the full information of the original SNP sequence（haplotype）.

Various SNP tagging methods have been proposed in the literature, based on different purposes of applications. This paper focuses on the most commonly cited definition of tagSNP and proposes methods to select them. In particular, we first seek the smallest SNP subset to identify all Haplotype patterns. The tagSNP Selection Problem can be modeled as a 0-1 binary integer programming problem with multiple optimal solutions. Previously, scholars focused on improving the efficiency of their solution methods, and ignored the differences among multiple optimal solutions. To this end, we proposes a Heuristic algorithm based on graph theory that first solves for all the multiple optimal solutions and then recommends a more informative set of optimal solution based on the concept of Linkage Disequilibrium（LD）. Our method calculates the relevancy between a selected tagSNP and the other SNPs, which later serves as an indicator for assessing the amount of information it carries. Among all the multiple optimal solutions, we recommend the one with the largest sum of LD values. In addition, we also propose a bi-objective integer programming model which tries to minimize the number of selected tagSNPs while maximize the sum of their LD values.

To deal with large-scale tagSNP selection problems, we propose a heuristic called

LRH, based on the theory of Lagrangian Relaxation. In particular, a modified subgradient method is proposed to update the Lagrangian multiplier which in turn approaches to the optimal solution step by step. In LRH, we incorporate the concept of Greedy Algorithm which selects some good SNP column, gradually reduces the problem size, and thus greatly improves the efficiency and quality. The computational results show that LRH has good efficiency and effectiveness, since it can quickly converges to a better solution. Moreover, we also give a two-stage solution method（MIX） which first uses LRH to select good candidate SNP columns and then solve a reduced tagSNP selection problem of much smaller size based on the selected candidate SNPs. The computational results show that the proposed two-stage approach converges to a better solution in a much shorter time.

Finally, in a suggested future research topic, we consider the case where the selected tagSNPs is to be included in a biochip, while the capacity for the biochip is sufficiently large. In this case, we no longer have to select the SNPs of minimum size. Instead, we focus on selecting those tagSNPs that can be used to differentiate haplotypes as many as possible so that the selected SNPs are more robust or reliable.

Kayowrd：tagSNP 、Linkage Disequilibrium、Haplotype、Algorithm、Graph、Lagrangian Relaxation

Abstract

In a manufacturing network, there often exist operations that produce the same products or semi-products of different qualities. Such a phenomenon can be described by a distillation process using a specialized D-node introduced by Fang and Qi (2003) where the flow enters a D-node has to be distributed according to a pre-specified vector of ratios associated with its outgoing arcs. The existence of D-node complicates the manufacturing process since it creates the relation of flow dependency. In particular, to send flow successfully from a source node to a sink node, an augmenting subgraph containing many paths instead of an augmenting simple path has to be constructed, since the flow passing through a D-node has to be distributed to all of its outgoing arcs. As a result, when the capacity associated with each arc obeys a multi-state probability distribution, calculating the system reliability of shipping a given amount of flows for a multi-state capacitated manufacturing network containing D-nodes becomes a difficult task. This paper focuses on techniques to derive the system reliability for such a multi-state capacitated manufacturing network containing D-nodes. We will first introduce polynomial-time preprocessing techniques that simplify the problem structure, and then propose algorithms for calculating the network reliability. In addition, we also investigate the problem of constructing a manufacturing network with multi-state arc capacity that satisfies the reliability lower bound with minimum total cost, and study the problem of stochastic shortest path using our proposed solution methods as well.

keywords: Network reliability, Multi-state arc capacity, Manufacturing network, Distillation process.

Abstract

In a metropolitan area where the conventional four-wheel vehicles may have difficulties in shipping small consignments efficiently due to the limited parking spaces and congested traffic, the motorcycles become perfect means for such a task. Two major network configurations are used in the industry of the metropolitan motorcycle courier services: the hub-and-spoke network (HB) and the point-to-point network (PP). Conventional motorcycle courier services usually employ PP as their network structure, yet the largest motorcycle courier service in Taipei metropolitan area uses HB to achieve higher quality of service and thus dominates the market of shipping small consignments. This paper, with both network strategies employed, analyzes the efficiency of service and trend of profits for both the monopoly and competition models, respectively. In the monopoly model, we derive the total amount, on-time amount, and delayed amount of consignments as the performance indices representing the service efficiency for each network strategy. Then, we try to exploit the theories of games and incorporate the penalty of delayed consignments into the profit function to analyze the profits for two motorcycle courier services in a Duopoly market.

By exploring the three different network strategy combinations (HS,HS), (PP,PP)and (PP,PP)in a simultaneous-move Cournot game where both players serve the same service zones, we observe the following three facts: First, the HS strategy will always have more profits when the economies of scale is sufficiently significant so that it can conduct more consignments with less cost (and thus is considered to have more service advantages); Second, the PP strategy may beat the HS strategy in the cases where the HS strategy does not have sufficient service advantages, and then the equilibrium would

exist; Third, when both the HS and PP strategies have the same service efficiency, then the PP strategy becomes the only profitable choice. In the cases when both players are free to serve different amount of service zones and their network strategies as well, generally the PP strategy is competitive to, or even sometimes beats the HS strategy. Nevertheless, if the HS strategy has sufficient service advantages, it becomes the profitable choice for both players. Besides, we also propose a two types of services model to obtain the optimal decision of allotting capability while the proportions of market demand of two types services are known.

Key Words: Hub-and-Spoke, Point-to-Point, Monopoly, Network Strategy, Game Theory, Duopoly, Cournot, Stackelberg