Themen für Bachelorarbeiten

aus dem Bereich der

Kombinatorischen Optimierung

Dr. Eranda Dragoti-Cela
Institut für Diskrete Mathematik, TU Graz

März 2017

1. Online Exploration in Graphen: das Problem des kanadischen Tourenbetreibers (the Canadian tour operator problem (CTOP))

   Der Input des CTOP besteht aus einem ungerichteten kanten- und knotengewichteten Graphen $G = (V, E)$ mit Kantengewichten $w\colon E \to \mathbb{R}_+$, Knotengewichten $\pi\colon V \to \mathbb{R}^+$, und einem Startknoten v_0. Der Tourenbetrieber muss eine in v_o startende und in v_o endende geschlossene Tour T mit Kantenmenge $E(T)$, Knotenmenge $V(T)$, und minimalen Kosten $c(T)$ bestimmen, wobei die Kosten der Tour als $c(T) := \sum_{e\in E(T)} w(e) + \sum_{i\in V\setminus V(T)} \pi(i)$ gebeben sind. Die Zielfunktion ist also als Summe der Kantengewichte aller von der Tour durchlaufenen Kanten plus die Summe der Knotengewichte der von der Tour nicht besuchten Knoten gegeben. Die Gewichte können folgendermaßen interpretiert werden. Die Kantengewichte sind proportional zu den Kosten für die Zurücklegung der Strecke zwischen den Endknoten der Kante (sehenswürdige Orte). Es sollten idealerweise alle Orte (Knoten) besucht werden, und wenn ein Ort nicht besucht wird, dann wird eine Strafe, die dem jeweiligen Knotengewicht entspricht, fällig.

   Manche der Kanten in E könnten durch Lawinen „gesperrt“ sein, und der Tourenbetreiber weiß erst bei der Ankunft an einem Endknoten einer solchen Kante, dass sie gesperrt ist. Der Rest der geplanten Route muss dann ggf. geändert werden. Somit ist das CTOP ein sogenanntes *Online Problem*. Die Variante des Problems, die entsteht wenn alle gesperrten Kanten im Voraus bekannt sind, nennt man das *Offline CTOP*; das entspricht dem *prize-collecting travelling salesman problem*.

   Ein deterministischer online Algorithmus ALG für ein Online Problem heißt c-kompetitiv falls für jede Instanz I des Online Problems die von ALG ermittelte Lösung einen Zielfunktionswert $ALG(I)$ von höchstens $c \cdot OPT(I)$ liefert, wobei $OPT(I)$ den optimalen Zielfunktionswert des dazugehörigen Offline Problems bezeichnet. Die kleinste Zahl c, für die ALG c- kompetitiv ist, heißt Kompetitivität-Quotient von ALG. Im allgemeinen ist bei Online Problemen das Ziel, Algorithmen mit kleinem Kompetitivität-Quotienten zu entwicklen [3, 6]. Es ist oft schwierig den Kompetitivität-Quotienten eines bestimmten Algorithmus zu ermitteln, selbst die Ermittlung von oberen und unteren Schranken dafür ist i.A. ein anspruchsvolles Ziel. Derartige Untersuchungen sind Subjekt der Kompetitivitätsanalyse

   Für das CTOP in beliebigen Graphen ist der Kompetitivität-Quotient unbeschränkt. Falls G ein Pfad ist, dann ist der goldene Schnitt $(1 + \sqrt{5})/2$ eine untere Schranke für den Kompetitivität-Quotienten eines jeden deterministischen Algorithmus, siehe Büttner und Krumke [7]. In [7] wird auch ein Algorithmus beschrieben, der die oben genannte untere Schranke erreicht und somit ein „optimaler“ Algorithmus aus der Sicht der Kompetitivitätsanalyse ist.

In einer Bachelorarbeit zu diesem Thema sollen die Grunddefinitionen und Konzepte aus der Kompetitivitätsanalyse von Online Algorithmen eingeführt und insbesondere das CTOP in Pfaden näher betrachtet werden. Der in [7] beschriebene Algorithmus soll anhand von Beispielen veranschaulicht werden.

2. Das Rucksackproblem mit Setups: dynamische Optimierung und Approximationsalgorithmen

   Der Input des klassischen binären Rucksackproblem (KP) besteht aus einee Menge von n Gegenständen $\{1, 2, \dots, n\}$, $n \in \mathbb{N}$, mit nichtnegativen Gewichten w_i und nichtnegativen Nutzenkoeffizienten p_i, $1 \leq i \leq n$, sowie ein Gewichtslimit $b \in \mathbb{R}_+$. Das Ziel ist es eine Teilmenge $M \subseteq \{1, 2, \dots, n\}$ von Gegenständen zu ermitteln, sodass ihr Gesamtgewicht das Limit b nicht überschreitet, $\sum_{i \in M} w_i \leq b$, und der Gesamtnutzen $\sum_{i \in M} p_i$ maximiert wird. KP ist eines der bestuntersuchten Probleme in der kombinatorischen Optimierung, ein schwieriges Problem, das jedoch durch dynamische Optimierung mit einer von b polynomiell abhängenden Zeitkomplexität gelöst werden kann. Weiters ist das Problem gut approximierbar, d.h. es besitz ein sogenanntes *fully polynomial time approximation scheme (FPTAS)*. Eine unfassende Literaturquelle über KP und verwandte Probleme ist zB. [11].

   Das Rucksackproblem mit Setup-Kosten (KPS) ist eine Verallgemeinerung des KP in dem die Gegenstände in Familien eingeteilt sind. Im Allgemeinen bilden die Familien jedoch keine Partition der Menge der Ggenstände. Wenn ein Gegenstand aus einer bestimmten Familie in den Rucksack eingepackt wird so muss die Familie „freigeschaltet“ werden, dafür fallen familienabhängige Setup-Kosten an. Weiters beansprucht jede Familie bei Freischaltung auch ein Teil des gesamten Gewichtslimits. Der Gewichts- bzw. Nutzenkoeffizient eines jeden Gegenstands hängt nun auch davon ab, aus welcher Familie dieser Gegenstand selektiert wird. Das Ziel ist, wie bei KP, die Zusammensetzung eines Rucksacks, der das Gewichtslimit respektiert (inkl. der Gewichte der freigeschalteten Familien) und den Gesamtnutzen abzüglich der Setup-Kosten maximiert. KPS wurde in [5] eingeführt. Einige neue Ergebnisse, darunter eine neuer effizienterer dynamischer Optimierungsansatz, wurden in [13] präsentiert. Darüber hinaus wurden in [13] auch Approxmiationsergebnisse für einige relavante Spezialfälle von KPS, sowie ein Nicht-Approximierbarkeit-Resultat für das allgemeine KPS präsentiert.

   In einer Bachelorarbeit zu diesem Thema kann zunächst das KP und einige dazugehörige grundlegende Ergebnisse eingeführt werden. Weiters soll das KPS eingeführt und der neuer dynamischer Optimierungsansatz präsentiert werden. Je nach Präferenz des Studierenden kann dieser Algorithmus implementiert und getestet, oder die Approximierbarkeit des Problems behandelt werden.

3. Spezialfälle des quadratischen Kürzeste Wegeproblems

   Der Input des quadratischen Kürzeste Wegeproblems (*quadratic shortest path problem (QSPP)*) besteht aus einem gerichteten Graphen $G = (V, E)$, einem Start- und einem Endknoten $s \in V$ bzw. $t \in V$, Kantenlängen $c_e \in \mathbb{R}_+$, fär alle Kanten $e \in E$, sowie quadratischen Kosten $q_{ef} \in \mathbb{R}_+$, für alle $e, f \in E$. Die quadratischen Kosten bilden eine durch die Kanten des Graphen indexierte symmetrische nicht-negative $|E| \times |E|$ matrix $Q = (q_{ef})$. Es wird ein Pfad P von s nach t gesucht, der die Summe der linearen Kosten $c(P)$ und der quadratischen Kosten $q(P)$ minimiert. Die linearen und die quadratischen Kosten von P sind als $c(P) := \sum_{e \in E(P)} c_e$ bzw. $q(P) := \sum_{e,f \in E(P)} c_{qf}$ gegeben, wobei $E(P)$ die Menge der Kanten des Pfades P bezeichnet. Klarerweise ist das QSPP eine Verallgemeinerung des kürzesten Wegeproblems (*shortest path problem (SPP)*): wenn Q die Nullmatrix ist, dann entspricht das QSPP dem SPP mit Input G, s, t und c.

In [14] wurde gezeigt, dass das QSPP ein schwieriges Problem ist: für das QSPP gibt es keinen polynomiellen Algorithmus mit konstanter Approximationsgüte, vorausgesetzt $P \neq NP$. Es gibt jedoch einige Varianten des Problems, die polynomiell lösbar sind, etwa das sogenannte adjazente quadratische kürzeste Wegeproblem (*adjacent quadratic shortetst paths problem (AQSPP)*) auf seriell-parallele oder azyklische Graphen [10]. Beim adjazenten quadratischen Kürzeste Wegeproblem tragen nur die adjazenten Kanten eines Pfades etwas bei den quadratischen Kosten des Pfades P bei, d.h. die quadratischen Kosten $q(P)$ von P sind folgendermaßen gegeben:

$$q(P) = \sum_{\substack{e,f \in E(P) \\ e,\, f \text{ teilen einen Knoten}}} q_{ef}\,.$$

Es gibt auch einige polynomial lösbare Spezialfälle von QSPP insbesondere das *linearisierbare QSPP* [10]. Eine Instanz des QSPP mit Input G, c, s, t und Q heißt *linearisiebar*, wenn es Kantenlängen c'_e, $e \in E(G)$ gibt, sodass für jeden s-t-Pfad P in G die Gleichung $c(P)+q(P) = c'(P)$ mit $c'(P) := \sum_{e\in E(P)} c'_e$ gilt. In [10] werden grundlegende Eigenschaften des linearisierbaren QSPP untersucht, notwendige und hinreichende Bedingungen für die Linearisierbarkeit gegeben, sowie einige linearisierbare Spezialfälle des QSPP, etwa das QSPP in Gridgraphen, charakterisiert. Weiters wird in [10] ein effizienter Algorithmus zur Erkennung von linearisierbaren QSPP in Gridgraphen dargestellt.

Eine Bachelorarbeit zu diesem Thema sollte sich mit dem QSPP und dem AQSSP im Allgemeinen und insbesondere mit linearisierbaren Spezialfällen von QSPP befassen. Dazu kann der oben genannter Algorithmus zur Erkennung vom linearisierbaren QSPP in Gridgraphen implementiert und getestet werden.

4. Spannende Bäume mit minimalen Kommunikationskosten: Modellierung und exakte Algorithmen

   Das abstandoptimale Spannbaumproblem (*optimum distance spanning tree (ODST)*) ist ein gut untersuchtes Problem im Netzwerkdesign, siehe zB. [8]. Der Input des ODST ist ein ungerichteter kantengewichteter Graph $G = (V, E)$ mit Kantengewichten $c_e \in \mathbb{R}_+$, für $e \in E$. Das Ziel ist es ein Spannbaum T zu finden, der die Summe der Längen der eindeutigen u-v-Pfade in T über alle $u, v \in V$ minimiert. Die Knoten des Graphen modellieren Einrichtungen, die mit einander entlang Pfaden kommunizieren sollen, die Kanten entsprechen direkten Verbindungen zwischen den jewiligen Endknoten, und die Kantengewichte entsprechen den für die Verwendung der entsprechenden direkten Verbindung anfallenden Kosten. Somit entspricht das gesuchte Spannbaum einem Netzwerk mit minimalen Gesamtkommunikationskosten. (In diesem vereinfachten Netzwerkdesignmodell werden sowohl die Errichtungskosten für die direkten Verbindungen als auch der Kommunikationbedarf für je zwei Einrichtungen vernachlässigt bzw. als konstante Größe über alle Eninrichtungen bzw. Einrichtungspaare betrachtet.) ODST ist eine schwieriges Problem und viele dafür entwickelte exakte Lösungsverfahren beruhen auf Formulierungen des Problems als gemischt-ganzahliges lineares Programm (*mixed integer linear program (MILP)*). Die MILP Formulierungen werden mit Hilfe von enumerativen Verfahren, wie *branch-and-bound* und *branch-and-cut*, gelöst. In der Literatur gibt es unterschiedliche MILP Formulierungen für das Problem, deren linearen Relaxationen unterschiedlich gute untere Schranken für den optimalen Zielfunktionswert des Problems liefern und in unterschiedlichen branch-and-bound Verfahren verwendet werden. Oft werden die dazugehörigen linearen Relaxationen mit Hilfe von *column generation* Ansätzen gelöst, siehe z.B. [8].

Eine Bachelorarbeit zu diesem Thema sollte einige LP-Formulierungen des ODST beschreiben und darauf basierend die Berechnung von unteren Schranken diskutieren. Weiters sollte ein exaktes branch-and-bound Verfahren zur Lösung des Problems beschrieben und and kleinen Instanzen veranschaulicht werden. Als Grundreferenz dazu könnte [8] dienen.

5. Robuste Routenplanung im urbanen öffentlichen Verkehr

   Ein typisches urbanes öffentliches Verkehrnetz umfasst mehrere Transportmittel, wie zB. Busse und Straßenbahnen, ist dicht, und weist eine hohe Frequenz auf. Eine Verkehrslinie ist eine Folge von Haltestellen, die mit einem bestimmten Verkehrsmittel in einer fixen Reihenfolge periodisch bedient werden. Das Verkehrsnetz kann mit Hilfe eines gerichteten Graphen modelliert werden in dem die Knoten den Haltestellen unterschiedlicher Transportlinien entsprechen. Eine Kante (s, t) entspricht einer direkten Verbindung von s nach t, d.h. es gibt eine Verkehrslinie, die die Haltestelle t unmittelbar nach s besucht. Dabei kann es zu Mehrfachkanten kommen, etwa wenn s und t zwei unmittelbar nach einander folgende Haltestellen von mehr als eine Linie sind. Eine Route von Haltestelle a zu Haltestelle b ist eine Folge von Haltestellen $a = s_o, s_1, \ldots, s_k = b$, für ein $k \in \mathbb{N}$, sodass für jedes $i \in \{0, 1, \ldots, k-1\}$ eine Linie L_i gibt die s_i und s_{i+1} in dieser Reihenfolge bedient. Für jede Linie jedes Transportmittels gibt es einen periodischen (etwa täglichen) Fahrplan. Ein Routenplanungsproblem (*routing problem (RP)*) besteht darin, für eine gegebene Starthaltestelle a, eine Endhaltestelle b und eine Ankunftszeit t_b die schnellste Route von a nach b zu berechnen, unter Berücksichtigung der Fahrpläne aller Linien [12].

   Bei der robusten Routenplanung sollen auch die eventuellen Verspätungen im Verkehr berücksichtigt werden; diese könnten eine im Voraus berrechnete Route unzulässig machen. In [4] wird die robuste Routenplanung angesichts von aufgezeichneten historisch realisierten Verspätungen analysiert. Die Autoren führen als mathematisches Modell das robuste Routenplanungsproblem (*robust routing problem (RRP)*) ein. Das Ziel besteht darin, für ein gegebenes Tripel (a, b, t_b) einer Starthaltestelle a, einer Endhaltestelle b und einer Ankunftszeit t_b, eine (oder mehrere) *robuste Routen* zu ermitteln. Eine Route heißt *robust*, wenn sie für einen bestimmten Anteil der historisch realisierten Fahrzeiten zulässig gewesen wäre. Die Autoren führen auch ein sogennantes *Zuverlässigkeitsmaß* (*measure of reliability*) einer Route ein und präsentieren einen iterativen Konstruktions- und Verbesserungsalgorithmus zur Lösung des RRP.

   Eine Bachelorarbeit zu diesem Thema sollte sich mit dem RRP Modell aus [4] befassen, die nicht-trivialen Modellierungsaspekte behandeln und Algorithmen zur Berechnung von robusten Routen und dem dazugehörigen Zuverlässigkeitsmaß beschreiben. Die Funktionsweise des Algorithmus soll an einem kleinen Testbeispiel veranschaulicht werden.

Weitere Themen

6. Kombinatorische Optimierungsprobleme mit Budgetrestriktionen (combinatorial optimization with budget constraints)

   Viele klassische polynomial lösbare kombinatorische Optimierungsprobleme wie das minimale Spanbaumproblem, das maximale Matchingproblem, das kurzeste Wegeproblem, können als Spezialfälle des folgenden generischen Problems betrachtet werden. Sei $\mathcal{F}$ eine Menge von zulässigen Lösungen und $w\colon \mathcal{F} \to \mathbb{R}$ eine Abbildung, die jeder zulassigen Lösung $S \in \mathcal{F}$ ein Gewicht $w(S)$ zuordnet. Gesucht wird eine zulässige Lösung S^* mit maximalem Gewicht, i.e. $w(S^*) = \max\{w(S)\colon S \in \mathcal{F}\}$. Betrachten wir nun eine Modifikation des Problems in der den zulässigen Lösungen neben den Gewichten $w(S)$ auch Kosten $c(S)$, $S \in \mathcal{F}$, zugeordnet werden. Weiters gäbe es ein Budget B, das nicht überschritten werden darf, d.h. zulässig sind nur jene Lösungen S aus $\mathcal{F}$, für die $c(S) \le B$ gilt. Das Optimierungsproblem lautet nun „Bestimme $S^* \in \mathcal{F}$, sodass $w(S^*) = \max\{w(S)\colon S \in \mathcal{F}, c(S) \le B\}$“. In diesem Fall spricht man von einem Optimierungsproblem mit einer Budgetrestriktion. Wenn es mehrere Kostenfuktions $c_i\colon \mathcal{S} \to \mathbb{R}$ und mehrere Budgetgrenzen B_i, $1 \le i \le k$, $k \in \mathbb{N}$, gibt, so spricht man von einem Optimierungsproblem mit k Budgetrestriktionen; gesucht wird nun ein $S^* \in \mathcal{F}$, sodass $w(S^*) = \max\{w(S)\colon S \in \mathcal{F}, c_i(S) \le B_i, 1 \le i \le k\}$.

Alle oben genannten Probleme sind bereits NP-schwer für $k = 1$, es gibt jedoch polynomielle Approximationsschemas für alle drei Probleme in diesem Fall. Das polynomielle Approximationsschema für das maximale Matchingproblem mit einer Budgetrestriktion beruht auf die Lagrange Relaxation des Problems bei der die Budgetrestriktion relaxiert wird. Zwei Lösungen dieser Lagrange Relaxation werden geschickt miteinander kombiniert um eine approximative Lösung des ursprünglichen Problems zu erhalten. Als Nebenergebnis wird dabei auch ein altes kombinatorisches Puzzle gelöst, siehe [2].

Für $k \ge 2$ ist die Situation generell eine andere; das kürzeste Wege Problem und das Spannbaumproblem sind nicht approximierbar für $k \ge 2$, siehe [9]. Für das maximale matching Problem hingegen gibt es im Fall $k = 2$ ein polynomielles Approximationsschema [9].

Im Rahmen einer Bachelorarbeit könnten die PTAS für das maximale Matchingproblem mit einer Budgetrestriktion bzw. zwei Budgetrestriktionen behandelt werden.

7. Das Frequenzzuordnungsproblem (the frequency assignment problem) (FZP)

   Das Frequenzzuordnungsproblem in seinen vielen Varianten entspricht praktischen Fragestellungen in Telekommunikationsnetzwerken. Das grundlegende Problem kann in ein allgemeines Framework folgendermaßen beschrieben. Ein Telekommunikationsnetzwerk wird als Graph modelliert, dessen Knoten Kommunikationszellen (Antennen, Basisstationen udgl.) entsprechen. Die Kanten entsprechen den Verbindungen (Kanälen) zwischen den Kommunikationszellen. Jedem Knoten muss eine gewisse Anzahl von Frequenzen zugeordnet werden, damit der Knoten die ihm zugeordneten Verbindungen realisieren kann. In manchen Anwendungen gibt es für jeden Knoten eine im Voraus festgelegte Menge von Frequenzen, die diesem Knoten zugeordnet werden können. Abhängig von der Topologie des Graphen und den vorgenommenen Frequenzzuordnungen können zwischen bestimmten Verbindungspaaren kommunikationstörende Interferenzen eintreten. Die potentiellen Interferenzen werden oft anhand eines sogenannten Interferenzgraphen modelliert. In manchen Anwendungen muss das Eintreten von Interferenzen gänzlich ausgeschlossen werden, in Anderen wiederum wird das Eintreten von Interferenzen anhand von Straftermen in der Zielfunktion penalisiert. Das Ziel ist es eine Zuordnung von Frequenzen zu den Kommunikationszellen zu finden, die alle Anforderungen erfüllt und eine bestimmte Zielfunktion optimiert. Eine häufig verwendete Zielfunktion ist die Minimierung der Gesamtanzahl der verwendeteten Frequenzen. Eine andere in der Praxis relevante Zielfunktion ist die Minimierung der Differenz zwischen der höchsten und der niedrigsten Frequenz, die demselben Knoten zugeordnet werden.

   In der Regel sind die Varianten des FZP NP-schwer und es besteht ein starker Zusammenhang mit Färbungsproblemen in Graphen. Einige Forschungsaspekte im Bereich der Frequenzzuordnungsprobleme sind: gemischt-ganzzahlige Formulierungen, die eine approximative Lösung des Problems ermöglichen, strukturelle Untersuchungen des dazugehörigen Polytops, die Ermittlung von unteren (oberen) Schranken, die Entwicklung von heuristischen Verfahren, die die Ermittlung von guten Lösungen für praktische Probleme ermöglichen.

Als allgemeine Orientierungshilfe und als Quelle von weiteren adequaten und modellspezifischen Referenzen kann die Arbeit von Aardal et al. [1] dienen.

Literatur

[1] K.I. Aardal, S.P.M. Van Hoesel, A.M.C.A. Koster, C. Mannino und A. Sassano, Models and solutions techniques for frequency assignment problems, 4OR, *Operations Research Quarterly* **1(4)**, 261-317, 2003.

[2] André Berger, Vincenzo Bonifaci, Fabrizio Grandoni, Guido Schäfer, Budgeted matching and budgeted matroid intersection via the gasoline puzzle, *Mathematical Programming, Ser. A* **128**, 2011, 355–372.

[3] A. Borodin und E. El-Yaniv, Online Computation and Competitive Analysis, Cambridge University Press, New York, 2005.

[4] K. Böhmova, M. Mihalák, T. Pröger, R. Srámek und P. Widmayer, Robust routing in urban public transportation: how to find reliable journeys based on past observations, 2013, `http://drops.dagstuhl.de/opus/volltexte/2013/4242`.

[5] E.D. Chajakis und M. Guignard, Exact algorithms for the setup knapscak problem, INFOR **32**, 1994, 124–142.

[6] A. Fiat und G.J. Woeginger, Competitive analysis of algorithms, in *Online algorithms: the state of the art*, A. Fiat und G.J. Woeginger, Hrsg., *Lecture Notes in Computer Science* **1442**, Springer, Berlin, 1998, S. 1–12.

[7] S. Büttner und S. Krumke, The Canadian tour operator problem on paths: tight bounds and resource augmentation, *Journal of Combinatorial Optimization* **38**, 2016, 842–854.

[8] M. Fischetti, G. Lanca und P. Serafini, Exact algorithms for minimum routing cost trees, *Networks* **39(3)**, 2002, 161–173.

[9] Fabrizio Grandoni, Rico Zenklusen, Optimization with More than One Budget http://arxiv.org/abs/1002.2147

[10] H. Hu und R. Sotirov, Special cases of the quadratic shortest paths problem, 2016, `https://arxiv.org/abs/1611.07682`

[11] H. Kellerer, U. Pferschy und D. Pisinger, Knapsack problems, Springer, 2004.

[12] M. Müller-Hannemann und M. Schnee, Efficient timetable information in the presence of delays, in R.K. Ahuja, R.H. Möhring ind C.D. Zaroliagis, Hrsg., Robust and Online Large Scale Optimization, *Lecture Notes in Computer Science* **5868**, Springer, 2009, S. 249–272.

[13] U. Pferschy und R. Scatamacchia, Improved dynamic programming and approximation results for the knapsack problem with setups, Optimization online, 2016, `www.optimization-online.org/DB_FILE/2016/07/5539.pdf`.

[14] B. Rostami, A. Chassein, M. Hopf, D. Frei, C. Buchheim, F. Malucelli, M. Goerigk, The quadratic shortest path problem: complexity, approximability, and solution methods, Optimization online, 2016, `http://www.optimization-online.org/DB_HTML/2016/02/5341.html`.