Optimallösungen ausreizen - vom Partial Set Covering zum Multiple Coverage Partial Set Covering als Anwendung im Schienenverkehr

Oliver Annen, Günter Törner, Duisburg-Essen

1 Einführung

In den OR News Nr. 11/März 2001 berichteten wir über ein Optimierungsproblem aus dem Schienenverkehr, das im Jahre 2000 Gegenstand eines Projekts mit der DB Cargo war. Von Seiten der DB Cargo bestand seinerzeit Interesse, Güterwaggons, die im gesamten Güterzug-Streckennetz der DB an Zugfahrten teilnehmen, durch Messstellen zu erfassen, wobei Messstellen auf Gleisen installiert werden und aufgrund bekannter Zuginformationen jeden passierenden Waggon identifizieren können. Auf betriebsspezifische Details soll hier nicht eingegangen werden. Die Optimierungsaufgabe bestand darin, einen geforderten Anteil von **x** % aller Güterwaggons durch Messstellen zu erfassen bei gleichzeitiger Minimierung der Anzahl der verwendeten Messstellen. Es versteht sich, dass nur solche Güterwaggons berücksichtigt werden können, die innerhalb des betrachteten Optimierungszeitraums an Zugfahrten teilnahmen. Das Problem wurde schließlich als *Partial Set Covering Problem* (PSCP) modelliert. Das PSCP (vgl. [1,2,3,4,5,6]) stellt eine Verallgemeinerung des bekannten *Set Covering Problem* (SCP) dar und lässt sich wie folgt als BIP (**B**inary **I**nteger **P**rogram) formulieren:

$$\min \sum_{e\in E} c_e \cdot x_e$$

$$y_k + \sum_{e\in E_k} x_e \geq 1 \text{ für alle } k \in M$$

$$\sum_{k\in M} y_k \leq m - \alpha$$

$$x_e, y_k \in \{0,1\}.$$

Im Falle des Projekts mit der DB wurde $c_e = 1$ für alle $e \in E$ gesetzt. Hierbei ist E die Menge der Standorte und $M = \{1,..,m\}$ die Menge der Waggons. Die Route eines Waggons $k \in M$ wird dabei durch die Menge $E_k \subseteq E$ beschrieben. Der Parameter $\alpha \in \{1,..,m\}$ gibt die Anzahl der zu erfassenden Waggons an.

In diesem Beitrag berichten wir kurz über eine Erweiterung der Aufgabenstellung, die durch die nachgeschaltete Analyse motiviert wurde und dabei – wo immer möglich - eine *mehrfache* Erfassung von einzelnen Waggons durch Messstellen vorsieht. Der Vorteil einer mehrfachen Erfassung von Waggons ist offensichtlich, da die erzielte Redundanz zu einer Verbesserung der gesamten Datenbasis führt.

Die hier vorgestellten Ansätze sind Teil einer im Institut für Mathematik der Universität Duisburg-Essen angenommenen Dissertation des ersten Autors.

2 Das Partial Set Covering Problem mit Multiple Coverage

Wir betrachten die erwähnte Erweiterung unseres Ausgangsproblems als allgemeines Standortproblem mit einer Menge M von Bedarfsknoten und einer Menge E von möglichen Standorten. Ferner seien alle Kosten für Standorte positiv, d.h. es gilt $c_e \geq 0$. Das PSCP soll nun unter dem Aspekt des *Multiple Coverage* erweitert werden. Unter dem Begriff Multiple Coverage werden im Bereich der Standortplanung Optimierungsziele zusammengefasst, die eine Versorgung bzw. Überdeckung einzelner Bedarfsknoten durch mehrere Standorte anstreben (vgl. [7,8,9]).

Die Aufgabenstellung des PSCP wird dahingehend erweitert, dass jeder Bedarfsknoten möglichst häufig von gewählten Standorten überdeckt werden soll. *Zusätzliche Überdeckungen* (eine zusätzliche Überdeckung eines Bedarfsknotens durch einen (gewählten) Standort liegt vor, wenn der Bedarfsknoten bereits von einem anderen (gewählten) Standort überdeckt wird) eines Bedarfsknotens, die nicht durch die Standortwahl realisiert werden, werden dabei durch Strafkosten in Form von Gewichten in der Zielfunktion berücksichtigt. Diese können für einzelne Überdeckungen unterschiedlich sein, falls eine Priorität von zusätzlichen Überdeckungen vorliegt. Die Anzahl der zusätzlichen Überdeckungen, die pro Bedarfsknoten zu berücksichtigen sind, sei hier zunächst auf $n \geq 0$ zusätzliche Überdeckungen begrenzt. Abhängig von n kann für ein Bedarfsknoten $k \in M$ der Fall $|E_k| < n+1$ eintreten, d.h. eine Realisierung von n zusätzlichen Überdeckungen ist nicht möglich. In solchen Fällen betrachte man für $k \in M$ maximal $r_k = \max\{0, \min\{|E_k|-1, n\}\}$ zusätzliche Überdeckungen.

Der Einfachheit halber bezeichnen wir das geschilderte Problem als *Multiple Coverage Partial Set Covering Problem* (MCPSCP). Die Aufgabenstellung des MCPSCP lässt sich mathematisch wie folgt als BIP formulieren:

$$\min \sum_{e \in E} c_e \cdot x_e + \sum_{k \in M} \sum_{s=1}^{r_k} \omega_s \cdot z_k^s$$

$$y_k + \sum_{e \in E_k} x_e + \sum_{s=1}^{r_k} z_k^s \geq r_k + 1 \quad \text{für alle } k \in M$$

$$y_k \leq z_k^1 \quad \text{für alle } k \in M$$

$$z_k^{s-1} \leq z_k^s \quad \text{für alle } k \in M,\ s = 2,..,r_k$$

$$\sum_{k \in M} y_k \leq m - \alpha$$

$$x_e, y_k, z_k^s \in \{0,1\}.$$

Wir verzichten auf eine vollständige Beschreibung des Modells und gehen lediglich auf einige wichtige Details ein. Die Zielfunktion

$$\sum_{e \in E} c_e \cdot x_e + \sum_{k \in M} \sum_{s=1}^{r_k} \omega_s \cdot z_k^s$$

berücksichtigt die Kosten für Standorte sowie Strafkosten für nicht-realisierte zusätzliche Überdeckungen einzelner Bedarfsknoten. Das Gewicht $\omega_s \geq 0$ beschreibt die Strafkosten für die s-te zusätzliche Überdeckung eines Bedarfsknotens, die nicht realisiert wurde.

3 Ergebnisse

Zur Lösung des MCPSCP wurde ein *Simulated Annealing Algorithmus* entwickelt und implementiert. Als Anwendungsbeispiel präsentieren wir hier die Ergebnisse des Algorithmus auf realitätsnahen Daten, die uns aus dem Projekt mit der DB Cargo zur Verfügung standen. Auf eine nähere Beschreibung des Algorithmus sowie auf eine detaillierte Beschreibung der Datenbasis und der Gewichtewahl $\omega_1,..,\omega_n$ soll an dieser Stelle allerdings verzichtet werden. Ferner betrachten wir hier der Einfachheit halber nur den Spezialfall, dass alle zusätzlichen Erfassungen (= zus. Überdeckungen) von Waggons die gleiche Priorität haben, d.h. es gilt $\omega_1 = ... = \omega_n$.

Aus Tabelle 1 lassen sich die wichtigsten Angaben zu den verwendeten Probleminstanzen entnehmen. Für jede Instanz ist die Dauer des Optimierungszeitraums, die Anzahl der zu berücksichtigenden Waggons sowie die Anzahl möglicher Standorte für Messstellen angegeben. Die Spalte p gibt den prozentualen Anteil der zu erfassenden Waggons an, d.h. es gilt $\alpha := \lceil m \cdot p \cdot 0.01 \rceil$. Der Name einer Instanz ist in der Spalte *Instanz* angegeben.

Instanz	Zeitraum in Tagen	p	Anzahl m Waggons	Anzahl Standorte
Bahn_1	1	100	11246	407
Bahn_2	1	95	11246	407
Bahn_3	1	90	11246	407
Bahn_4	1	100	12841	431
Bahn_5	1	100	55036	712
Bahn_6	3	100	102844	804
Bahn_7	3	90	102844	804
Bahn_8	14	100	134086	826
Bahn_9	14	90	134086	826
Bahn_10	14	90	165219	881

Tabelle 1: Reale Instanzen der DB

Anhand der Instanzen aus Tabelle 1 sollen nun die Lösungen des MCPSCP mit den Lösungen des PSCP unter dem Aspekt des Multiple Coverage verglichen werden. Wir betrachten hierfür die Tabelle 2.

Instanz	Lösung MCPSCP	Opt. Lösung PSCP	Diff. Erf.	Proz. Diff. Erf.
Bahn_1	90	90	2550	35,54 %
Bahn_2	63	63	1579	29,88 %
Bahn_3	51	51	1678	44,33 %
Bahn_4	101	101	3111	27,16 %
Bahn_5	272	272	3680	3,46 %
Bahn_6	283	283	8173	2,87 %
Bahn_7	70	70	11173	8,17 %
Bahn_8	285	285	5727	1,3 %
Bahn_9	55	55	13048	6,43 %
Bahn_10	44	44	26582	10,02 %

Tabelle 2: MCPSCP vs. PSCP

Aus Tabelle 2 lassen sich zunächst für jede der obigen Instanzen die berechnete Anzahl von Messstellen für das MCPSCP (Spalte *Lösung MCPSCP*) sowie für das PSCP (Spalte *Opt. Lösung PSCP*) entnehmen. Die Lösungen für das PSCP wurden hierbei mit Hilfe von CPLEX 7.0 berechnet. In allen Fällen handelt es sich um Optimallösungen.

Vergleicht man die Lösungen der beiden Modelle miteinander, so zeigt sich für jede Instanz, dass die Lösung des MCPSCP stets eine Optimallösung des PSCP darstellt. Da die Lösungen sich allerdings voneinander unterscheiden (dies ergab ein detaillierter Vergleich der Lösungen hinsichtlich der Standortwahl), wurden die Lösungen des PSCP und des MCPSCP zusätzlich unter dem Aspekt des Multiple Coverage miteinander verglichen. Als Vergleichswert verwendeten wir die Anzahl der zusätzlichen Erfassungen von Waggons, die durch die jeweilige Lösung bzw. Positionierung von Messstellen realisiert wird. Die Spalten *Diff. Erf.* und *Proz. Diff. Erf.* aus Tabelle 2 vergleichen die Lösungen der beiden Modelle unter diesem Aspekt. Die Spalte *Diff. Erf.* gibt an, um wie viel zusätzliche Erfassungen sich die Lösung des MCPSCP von der Lösung des PSCP unterscheidet. Die prozentuale Verbesserung ist in der Spalte *Proz. Diff. Erf.* angegeben. Die Tabelle 2 zeigt, dass aus dem MCPSCP Lösungen resultieren, die für das PSCP optimal sind und bezüglich der Anzahl von zusätzlichen Erfassungen deutlich bessere Ergebnisse liefern als das PSCP. Bei kurzen Zeiträumen von einem Tag kann die Anzahl der zusätzlichen Erfassungen um bis zu 44,33 % verbessert werden. Bei längeren Zeiträumen von bis zu zwei Wochen sind Verbesserungen von bis zu 10,02 % möglich. Im Falle der Instanz Bahn_10 entspricht dies einer Verbesserung um 26582 zusätzliche Erfassungen.

Um die Qualität des SA-Algorithmus beurteilen zu können, bestimmten wir mit CPLEX 7.0 untere Schranken für die Probleme. In allen Fällen berechnete der Algorithmus Näherungslösungen mit einem *Gap* unter 1%. Die Berechnungsdauer lag im Durchschnitt bei ca. 60 Sekunden. CPLEX 7.0 erzielte erwartungsgemäß schlechtere Lösungen als der SA-Algorithmus. Dies gilt insbesondere für „größere" Instanzen mit einem Optimierungszeitraum von zwei Wochen (z.B. Instanz Bahn_10). Hier berechnete CPLEX 7.0 auch nach einer Rechenzeit von 1-2 Wochen lediglich Lösungen mit einem *Gap* von ca. 4%.

Zusammenfassend konnte anhand eines realen Standortproblems aus dem Schienenverkehr gezeigt werden, dass das PSCP mehrer Optimallösungen besitzen kann, die sich unter dem Aspekt des Multiple Coverage erheblich unterscheiden. Mit einer Formulierung des Standortproblems als MCPSCP war schließlich eine Berechnung solcher Optimallösungen des PSCP möglich, die im Vergleich zu einer anderen Optimallösung des PSCP deutlich bessere Ergebnisse hinsichtlich einer mehrfachen Erfassung von Güterwaggons erzielten.

Literaturverzeichnis

[1] M. J. Kearns (Ed.), *The computational complexity of machine learning,* MIT Press, Cambridge, 1990

[2] P. Slavik, *Improved performance of the greedy algorithm for partial cover,* Information Processing Letters **64** (1997), 251-254

[3] L. Burroughs, *Approximation algorithms for covering problems*, Masters`s thesis, University of Calgary (1998)

[4] M.S. Daskin und S.H. Owen, *Two new location covering problems: The partial p-center problem and the partial set covering problem,* Geographical Analysis **31** (1999), 217-235

[5] R. Bar-Yehuda, *Using homogeneous weights for approximating the partial cover problem,* Information Processing Letters **39** (2001), 137-144

[6] R. Gandhi, S. Khuller und A. Srinivasan, *Approximation algorithms for partial covering problems,* In: International Colloquium on Automata, Languages and Programming (ICALP) 2001, Track A (2001)

[7] M.S. Daskin und E.H. Stern, A hierarchical objective set covering model for emergency medical service vehicle deployment, Transportation Science **15** (1981), 137-152

[8] K. Hogan und C. ReVelle, *Backup coverage concepts in the location of emergency sciences,* Management Science **32** (1986), 1434-1444

[9] M.S. Daskin, K. Hogan und C. ReVelle, *Integration of multiple, excess, backup, and expected covering models,* Environment and Planning B **15** (1988), 15-35