

STATISTICKÁ ANALÝZA MAYEROVY METODY A MODIFIKOVANÉ FERNANDEZ DE LA VEGA-LUECKEROVY METODY PRO ŘEŠENÍ VÍCEOKRUHOVÉHO OKRUŽNÍHO DOPRAVNÍHO PROBLÉMU

STATISTICAL ANALYSIS OF THE MAYER METHOD AND THE MODIFIED FERNANDEZ DE LA VEGA-LUECKER METHOD FOR SOLVING THE MULTIPLE-TOURS TRAVELING SALESMAN PROBLEM

Petr Kučera

Anotace:

Víceokruhový okružní dopravní problém je úloha nalézt optimální rozvoz určitého materiálu z centrálního místa do několika dalších míst po několika okružních trasách (několika vozidly). Pro tuto úlohu není znám efektivní algoritmus, který by našel její teoretické optimum, a tak jsou používány aproximační metody. Ty se zaměřují především na minimalizaci celkové ujeté vzdálenosti (či nákladů na vlastní přepravu). Členové autorské katedry navrhli před několika lety postup více zohledňující počet použitých vozidel. V tomto příspěvku je provedena statistická analýza výsledků dvou verzí tohoto přístupu a dvou verzí „klasické“ Mayerovy metody na deseti náhodně generovaných testovacích příkladech s cílem dát uživateli návod, kterou metodu vybrat pro konkrétní úlohu podle jejích specifických vlastností.

Klíčová slova:

Víceokruhový okružní dopravní problém, aproximační metoda, regresní analýza, korelační analýza, analýza hlavních komponent.

Abstract:

The multiple-tours traveling salesman problem is the task to find the optimal distribution of some supplies from the central place to some other places using several cyclic routes (several vehicles). There is no efficient algorithm for finding its theoretical optimum and so heuristics (approximation methods) are used for its solution. They are concentrated especially on minimization of the total distance (or costs for the proper transportation). Members of the author's department have proposed an approach having a greater regard for the number of the vehicles. In this contribution the statistical analysis of results of two versions of this approach and other two versions of the "classical" Mayer method at ten randomly generated instances is implemented to give instructions for user what method of these to choose according to specific properties of the task to solve.

Key words:

Multiple-tours traveling salesman problem, approximation methods (heuristics), regression analysis, correlation analysis, principle components analysis.

ÚVOD

Víceokruhový okružní dopravní problém (VODP) je úloha optimalizovat způsob rozvozu určitého materiálu z jednoho centrálního místa do dalších míst (případně svozu opačným směrem). U každého z těchto míst je zadána kapacita, tj. množství, které je tam

třeba dopravit. Rovněž vozidlo má omezenou kapacitu, přičemž ta nestačí pro všechna místa zároveň, a tak je třeba použít k rozvozu více vozidel, případně jedno vozidlo musí absolvovat více jízd. Rozvoz se má uskutečnit okružním způsobem, tj. vozidlo vždy vyjede z centrálního místa, obslouží několik dalších míst a do centrálního se opět vrátí. Za předpokladu přímé úměrnosti nákladů ujeté vzdálenosti jde o to, aby celková délka tras rozvozu byla minimální.

Úloha patří mezi tzv. NP-úplné problémy, pro které neexistuje efektivní algoritmus, který by našel přesné teoretické optimum. Navíc se vlastně jedná o úlohu vícekriteriální optimalizace: kromě již zmíněné délky přepravní trasy je také účelné používat k přepravě co nejmenší počet vozidel.

Pro takové úlohy bývá vždy k dispozici celá řada aproximačních metod, které dávají přibližné, ale pro praktické účely dostatečně kvalitní řešení. Pro VODP se často používá Mayerova metoda, která je zaměřena především na optimalizaci přepravní vzdálenosti. Pro minimalizaci počtu okruhů (vozidel) byla navržena v [2], [3], [4] modifikace Fernandez de la Vega-Luekerovy (FVL) metody z [1] určené pro řešení tzv. „bin packing“ problému, tj. úlohy naskládat určité předměty se zadanými vahami (či objemem apod.) do minimálního počtu nádob dané kapacity. Metoda je založena na tom, že nejprve jsou do nádob rozmístěny „těžké“ předměty a teprve potom „lehké“, přičemž na tom, jak je určena váhová hranice mezi těmito dvěma skupinami, závisí garantovaná přesnost výpočtu. Při aplikaci na VODP je zřejmé, že „nádobami“ budou vozidla a „vahami předmětů“ kapacity míst.

CÍL A METODIKA

V [3] jsou dvě verze Mayerovy metody a dvě verze FVL metody (tj. celkem čtyři výpočetní postupy) testovány na deseti náhodně generovaných příkladech (v [3] lze najít také přesný popis jednotlivých metod a způsobu generování, což v tomto příspěvku pro jeho omezený disponibilní rozsah není uvedeno). Každá z metod byla u jednotlivých příkladů různě úspěšná. Proto je cílem tohoto příspěvku dát uživateli návod, která z nich by měla pravděpodobně dát pro konkrétní úlohu podle jejích vlastností, které si může sám snadno vyhodnotit, nejlepší výsledek. Z vlastností a výsledků úloh z [3] byl tedy sestaven statistický soubor dat, který byl podroben regresní a korelační analýze a analýze hlavních komponent.

Výsledky čtyř testovaných metod určily prvních osm znaků statistického souboru. První čtyři charakterizují délky vypočtených tras a jejich hodnoty udávají, o kolik se tyto liší od aritmetického průměru pro daný příklad (statistickou jednotku). Tento způsob byl zvolen proto, aby se jejich rozdělení pravděpodobnosti co nejvíce blížilo normálnímu. Vzhledem k diskrétnímu charakteru dat bylo ovšem jen stěží uvěřitelné, že by bylo možno nezamítnout hypotézu o normalitě rozdělení na jakékoliv rozumné hladině významnosti nejen u těchto, ale i u všech ostatních níže popsanych znaků, což test normality rozdělení potvrdil. Další čtyři znaky byly počty okruhů (použitých vozidel) v řešeních získaných jednotlivými metodami.

Z vlastností, podle kterých může úlohy hodnotit uživatel, bylo jako další statistické znaky vybráno následujících pět:

- (i) počet míst s velkými kapacitami,
- (ii) poměr sumy velkých kapacit ku sumě malých kapacit,
- (iii) počet míst určujících konvexní obal obsluhované oblasti,
- (iv) poměr mezi největší a nejmenší vzdáleností mezi některým z míst určujících konvexní obal a centrálním místem,
- (v) počet míst s velkými kapacitami mezi místy určujícími konvexní obal.

Statistický soubor tedy má celkem 13 znaků a 10 jednotek a je uveden v tabulce 1.

VÝSLEDKY A DISKUSE

Regresní analýza

Abychom určili, jak jistá vlastnost úlohy ovlivňuje úspěšnost dané metody, musíme především určit, zda s rostoucí hodnotou statistického znaku, který ji popisuje, hodnota statistického znaku charakterizujícího výsledek roste nebo klesá. Vzhledem k tomu, že jak délka trasy, tak počet vozidel jsou minimalizační kritéria, v prvně jmenovaném případě si metoda vede úspěšně při nízkých hodnotách statistického znaku vlastnosti, v opačném případě při vysokých. Ke zjištění tohoto plně stačí lineární regrese. Poznamenejme, že z prováděných analýz je to jediná, která nepředpokládá normalitu rozdělení pravděpodobnosti.

V tabulce 2 jsou parametry regresních nadrovin (včetně absolutního členu, který nemá pro účely tohoto rozboru význam) pro jednotlivé metody a kritéria (tj. délka přepravy a počet vozidel). Důležitá jsou zejména jejich znaménka, absolutní hodnoty jsou podstatné až při porovnávání různých metod pro úlohy se stejnou vlastností. V tabulce 3 jsou tzv. p-hodnoty (p-values), které udávají, s jakou pravděpodobností leží parametr v intervalu symetrickém podle své střední hodnoty (z tabulky 2), jehož jednou z hraničních hodnot je nula. Pravděpodobnost, že ve skutečnosti má koeficient opačné znaménko, je tedy polovina p-hodnoty. Zhruba lze říci, že čím je p-hodnota nižší, tím vypočtená regresní nadrovina, a tak i lineární regrese vůbec, lépe charakterizuje realitu. Případy p-hodnot nižších než 0,05 (tj. kdy má parametr ve skutečnosti opačné znaménko s pravděpodobností menší než 2,5%) jsou v tabulkách 2 a 3 vyznačeny tučně.

Výsledky regresní analýzy jsou ještě přehledně uvedeny v tabulce 4. V každé buňce je uvedeno pro danou vlastnost úlohy a metodu před lomítkem hodnocení z hlediska celkové délky řešení a za lomítkem z hlediska počtu vozidel. Šipky nahoru udávají, že metoda je vhodná pro vysoké hodnoty vlastnosti, dolů pro nízké, počet šipek vlastní hodnocení (tři šipky – „výborně“ až jedna – „dobře“). Tj. např. FVL metoda – verze 1 dosahuje velmi dobrých délek trasy pro vysoké hodnoty vlastnosti (v) a dobrých počtů okruhů pro vysoké hodnoty vlastnosti (iii), ostatní vlastnosti její úspěšnost neovlivňují, ovšem metodu má smysl použít i tehdy, kdy si dobře nevede žádná jiná metoda, tj. pro nízké hodnoty vlastností (i) a (iv).

Pří- klad	Mayer v.1 d	Mayer v.2 d	FVL v.1 d	FVL v.2 d	Mayer v.1 o	Mayer v.2 o	FVL v.1 o	FVL v.2 o	(i)	(ii)	(iii)	(iv)	(v)
1	2,0%	-0,7%	-0,7%	-0,7%	5	4	4	4	8	6,6	7	1,6845	5
2	-18,2%	-8,6%	14,5%	12,3%	5	5	4	4	4	5,3333	8	1,7145	2
3	-6,2%	0,8%	-7,5%	12,9%	5	5	5	4	6	6	7	1,7636	5
4	-5,9%	-4,9%	-0,8%	11,6%	5	5	4	4	5	5,6	8	2,1014	4
5	-4,7%	12,4%	-6,1%	-1,6%	5	5	4	4	7	5	6	1,2643	5
6	-1,4%	-11,4%	-1,6%	14,4%	5	4	4	4	6	5,6	7	1,3040	2
7	3,8%	8,7%	-8,7%	-3,8%	5	5	4	5	7	4,6	8	1,3245	6
8	-0,7%	-3,9%	-18,8%	23,5%	6	5	5	5	5	6	8	1,5214	4
9	-9,3%	5,4%	-0,1%	4,0%	5	5	5	4	7	4,1667	6	1,2417	3
10	-1,4%	-0,2%	4,0%	-2,4%	5	4	4	4	6	6	9	1,6578	5

Tabulka 1: Statistický soubor dat – vlastnosti příkladů VODP a jejich řešení získaných různými metodami

Parametr	Mayer v.1 d	Mayer v.2 d	FVL v.1 d	FVL v.2 d	Mayer v.1 o	Mayer v.2 o	FVL v.1 o	FVL v.2 o
abs.člen	1,1729	2,0057	1,1776	2,3373	6,6476	10,8674	7,8736	5,2564
(i)	-0,0276	-0,0627	0,0003	-0,1323	-0,2605	-0,4712	-0,2807	-0,1992
(ii)	0,0396	-0,0331	-0,0410	0,0705	0,2581	-0,3801	0,1005	0,0090
(iii)	0,0178	-0,0390	0,0322	-0,0562	-0,0587	-0,3864	-0,3199	0,1375
(iv)	-0,2352	-0,1744	0,0660	-0,2216	-0,8977	0,2706	-0,3993	-1,1038
(v)	0,0147	0,0501	-0,0566	0,0007	0,1116	0,3004	0,1405	0,1979

Tabulka 2: Parametry regresní analýzy

Parametr	Mayer v.1 d	Mayer v.2 d	FVL v.1 d	FVL v.2 d	Mayer v.1 o	Mayer v.2 o	FVL v.1 o	FVL v.2 o
abs.člen	0,0398	0,0009	0,0640	0,0148	0,0296	0,0014	0,0978	0,0995
(i)	0,4900	0,0399	0,9942	0,0668	0,2350	0,0214	0,4555	0,4328
(ii)	0,3861	0,2230	0,4432	0,3005	0,2847	0,0576	0,8051	0,9735
(iii)	0,6426	0,1290	0,4866	0,3381	0,7634	0,0369	0,3898	0,5713
(iv)	0,1316	0,0716	0,6781	0,2892	0,2331	0,5720	0,7492	0,2319
(v)	0,5925	0,0261	0,1318	0,9848	0,4379	0,0282	0,5843	0,2812

Tabulka 3: p-hodnoty regresní analýzy

Vlastnost	Mayer v.1	Mayer v.2	FVL v.1	FVL v.2
(i)	-/↑	↑↑↑/↑↑↑	-/-	↑↑/-
(ii)	↓/↓	↑/↑↑	-/-	↓/-
(iii)	-/-	↑↑↑↑↑	-/↑	↑/-
(iv)	↑↑/↑	↑↑/-	-/-	↑/↑
(v)	-/-	↓↓↓/↓↓↓	↑↑/-	-/↓

Tabulka 4: Hodnocení podle regresní analýzy

Korelační analýza

Korelační koeficienty udávají míru těsnosti lineární závislosti mezi dvěma náhodnými veličinami. Korelační analýza předpokládá normalitu jejich rozdělení v tom smyslu, že při nesplnění tohoto požadavku z nekorelovanosti dvou statistických znaků ještě nevyplývá jejich nezávislost. V našem případě má tedy smysl si všimnout v korelační matici v tabulce 5 vysokých absolutních hodnot (kolem 0,65 – 0,70), které indikují významně těsnou lineární závislost, zatímco nízké hodnoty žádný praktický význam nemají, neboť náhodné veličiny nemají normální rozdělení a navíc závislost mezi nimi nemusí být zrovna lineární.

V „prvním kvadrantu“ korelační matice lze vyhledat dvojice metod, které se vzájemně doplňují v tom smyslu, že pokud jedna dá špatné řešení, druhá dá dobré, a naopak. Pokud uživatel zkusí pro libovolnou úlohu obě, má velkou šanci, že pomocí jedné z nich získá dobré řešení. Tyto dvojice jsou určeny zápornými korelačními koeficienty s vysokou absolutní hodnotou. Vzhledem k omezeným možnostem rozsahu příspěvku ponecháváme jejich vyhledání v tabulce 5 na čtenáři.

„Druhý a třetí kvadrant“ korelační matice by měl alespoň v některých případech potvrdit výsledky regresní analýzy o vhodnosti jednotlivých metod pro úlohy s určitými vlastnostmi. Ke zcela opačnému výsledku se ale došlo v případě Mayerovy metody – verze 2 a vypočtené celkové délky řešení pro vlastnost (i). Podle korelační analýzy by tato metoda měla nacházet krátká řešení pro úlohy s malým počtem velkých kapacit. Pro opačný případ odhalila korelační analýza naděje Mayerovy metody – verze 1 (opět co se týče délky řešení),

kteře regresní analýze zůstaly skryty. Ohledně délky tras pro úlohy s nízkým počtem velkých kapacit mezi místy určujícími konvexní obal (vlastnost (v)) se potvrdily kvality Mayerovy metody – verze 2 a navíc se ukázala jako vhodná i Mayerova metoda – verze 1, pro opačný případ je zde doporučení, i když méně jednoznačné, obou verzí FVL metody. V ostatních případech se buďto potvrdily výsledky regresní analýzy nebo korelační analýza nepřinesla žádné významné výsledky.

Absence vysokých absolutních hodnot ve „čtvrtém kvadrantu“ svědčí o tom, že korelační analýza nenalezla mezi vlastnostmi vybranými pro charakteristiku úloh takovou dvojici, kde by vlastnosti navzájem jedna na druhé závisely.

Analýza hlavních komponent

V předchozím textu byla posuzována úspěšnost metod pro řešení VODP u konkrétních úloh v závislosti na celkem pěti jejich vlastnostech. Daná metoda ovšem může být podle některé z nich pro určitou úlohu vhodná, zatímco podle jiné nikoliv. Za účelem zjednodušení situace pro uživatele byla provedena poněkud netradičním způsobem analýza hlavních komponent. Nešlo totiž vyjádření jednotlivých proměnných pomocí menšího počtu nově zavedených, ale naopak o definici dvou nově agregovaných proměnných (vlastností) pomocí původních.

Konkrétní postup bude předveden v rámci prezentace na konferenci. Jeho výsledkem bylo vyjádření dvou nově definovaných vlastností y_1 a y_2 pomocí vlastností (i) a (ii), které se ukázaly nejvíce významné a jejichž hodnoty udávají proměnné x_1 a x_2 :

$$y_1 = -0,10x_1 + 1,22x_2, y_2 = 1,96x_1 + 1,66x_2.$$

Za jednu z nových agregovaných vlastností lze tedy považovat téměř výlučně vlastnost (ii) (koeficient $-0,10$ je v porovnání s koeficientem $1,22$ zanedbatelný), druhá bude potom „součet“ obou vlastností (i) a (ii) se slabě vyšší vahou u první z nich.

Výsledky lineární regrese jsou v tabulkách 6 a 7, zhodnocení analogické tabulce 8 necht' si čtenář provede sám.

	May. v.1 d	May. v.2 d	FVL v.1 d	FVL v.2 d	May. v.1 o	May. v.2 o	FVL v.1 o	FVL v.2 o	(i)	(ii)	(iii)	(iv)	(v)
M.v.1 d	1,00	0,27	-0,64	-0,30	0,19	-0,43	-0,13	0,48	0,60	0,27	0,13	-0,22	0,64
M.v.2 d	0,27	1,00	-0,33	-0,69	-0,17	0,36	0,09	0,19	0,64	-0,48	-0,40	-0,44	0,70
F.v.1 d	-0,64	-0,33	1,00	-0,25	-0,65	-0,25	-0,49	-0,68	-0,25	-0,07	0,14	0,25	-0,47
F.v.2 d	-0,30	-0,69	-0,25	1,00	0,63	0,24	0,48	0,16	-0,70	0,27	0,10	0,27	-0,56
M.v.1 o	0,19	-0,17	-0,65	0,63	1,00	0,22	0,51	0,67	-0,32	0,24	0,22	-0,05	-0,03
M.v.2 o	-0,43	0,36	-0,25	0,24	0,22	1,00	0,43	0,33	-0,33	-0,54	-0,19	0,02	0,05
F.v.1 o	-0,13	0,09	-0,49	0,48	0,51	0,43	1,00	0,22	-0,06	-0,10	-0,29	-0,12	-0,05
F.v.2 o	0,48	0,19	-0,68	0,16	0,67	0,33	0,22	1,00	-0,04	-0,14	0,33	-0,26	0,35
(i)	0,60	0,64	-0,25	-0,70	-0,32	-0,33	-0,06	-0,04	1,00	-0,10	-0,52	-0,47	0,54
(ii)	0,27	-0,48	-0,07	0,27	0,24	-0,54	-0,10	-0,14	-0,10	1,00	0,38	0,58	0,14
(iii)	0,13	-0,40	0,14	0,10	0,22	-0,19	-0,29	0,33	-0,52	0,38	1,00	0,51	0,13
(iv)	-0,22	-0,44	0,25	0,27	-0,05	0,02	-0,12	-0,26	-0,47	0,58	0,51	1,00	0,05
(v)	0,64	0,70	-0,47	-0,56	-0,03	0,05	-0,05	0,35	0,54	0,14	0,13	0,05	1,00

Tabulka 5: Korelační matice

Parametr	Mayer v.1 d	Mayer v.2 d	FVL v.1 d	FVL v.2 d	Mayer v.1 o	Mayer v.2 o	FVL v.1 o	FVL v.2 o
abs.člen	-0,4050	0,0073	0,1476	0,2500	5,0790	7,7462	4,8352	4,7741
y_1	0,0002	-0,0571	0,0037	0,0532	0,1219	-0,1916	-0,0335	-0,0492
y_2	0,0171	0,0159	-0,0093	-0,0238	-0,1340	-0,0894	-0,0157	-0,0131

Tabulka 6: Parametry regresní analýzy

Parametr	Mayer v.1 d	Mayer v.2 d	FVL v.1 d	FVL v.2 d	Mayer v.1 o	Mayer v.2 o	FVL v.1 o	FVL v.2 o
abs.člen	0,0445	0,9677	0,6428	0,3055	0,0019	0,0006	0,0269	0,0157
y_1	0,9918	0,0276	0,9209	0,0868	0,3598	0,2539	0,8746	0,7903
y_2	0,0458	0,0696	0,4988	0,0432	0,4742	0,1526	0,8375	0,8444

Tabulka 7: p-hodnoty regresní analýzy

ZÁVĚR

Ve výsledcích těchto analýz může uživatel nalézt cenné poznatky pro řešení svých konkrétních úloh. Je ale třeba mít na paměti, že pro úlohy poněkud odlišného typu (jiný počet míst, charakteristika jejich rozmístění, komunikační síť atd.) mohou jednotlivé metody dosahovat jiných výsledků než zde. Příspěvek je ale zároveň návodem pro uživatele, jak si sám může v takovém případě provést podobnou analýzu pro svůj typ úloh.

LITERATURA

- [1] Fernandez de la Vega, W., Lueker, G.S.: Bin Packing Solved within $1+\varepsilon$ in Linear time, *Combinatorica* 1, str. 349-355, 1981
- [2] Kučera P., Dömeová L.: Řešení víceokružového okružního dopravního problému se zřetelem na minimalizaci okruhů. Sborník: Kvantitativne metody v ekonomii a podnikání. Bratislava, FHI EU Bratislava, s. 74-77, 2002
- [3] Kučera P., Dömeová L.: Solution of the multipletours Travelling Salesman Problem with Vehicle Capacity preference. Proceedings of the 21th International Conference Mathematical Methods in Economics 2003, Praha, s.171-176, 2003
- [4] Kučera P., Domeová L.: Využití metod pro řešení „bin packing“ problému pro okružní dopravní problém. Sborník: SYSTE 02, System Engineering, Plzeň, Evida, s.135-144, 2002

kontaktní adresa

RNDr. Petr Kučera
Katedra operační a systémové analýzy
Fakulta provozně ekonomická
Česká zemědělská univerzita v Praze
165 21 Praha 6 – Suchbát
tel.: 224382351
email: kucera@pef.czu.cz