

TVORBA DISTRIBUČNEJ SIETE S POUŽITÍM ZHLUKOVACÍCH METÓD

CREATION OF DISTRIBUTION NETWORK WITH A USE OF AGGREGATION METHODS

Peter Matis¹

Anotace: Príspevok skúma možnosti použitia zhlukovacích metód pri tvorbe a optimalizácii distribučnej siete v intravilánoch miest alebo na rozľahlom území. V príspevku sú navrhnuté nové metódy merania kvality zhlukovania. Niekoľko známych a dve nové zhlukovacie metódy boli implementované a porovnávané použitím týchto metód merania kvality zhlukovania. V úlohe majú požiadavky odberateľov často ne-deterministický charakter. V príspevku sú požiadavky aproximované pomocou fuzzy množiny.

Kľúčová slova: tvorba siete, zhlukovanie, fuzzy logika

Summary: This paper presents possibilities of use aggregation methods for creation and optimization of distribution network in the cities or wide spread area. New methods of measurements for the quality of aggregation methods are presented. Aggregation methods are then compared using these measurements. Several known and two new aggregation methods have been implemented and compared using these quality measurements. In many real cases customer's demand is not deterministic. In the paper demand has been approximated by fuzzy set.

Keywords: network creation, aggregation, fuzzy logic

1. ÚVOD

Distribúcia tovaru alebo iných komodít predstavuje významnú časť dopravných tokov v obciach, mestách a v celej krajine. Distribúcia je obvykle realizovaná z niekoľkých centrálnych bodov k odberateľom prostredníctvom jednej alebo viacerých vrstiev distribučných centier (medzi-skladov) [5], [15], [16]. V každej vrstve môže byť počet distribučných centier pomerne veľký. V posledných rokoch je možné pozorovať snahu o zníženie počtu vrstiev a absolútneho počtu distribučných centier, za účelom zvýšenia efektívnosti a kvality procesov prebiehajúcich pri distribúcii. Skúmané distribučné systémy sú rozľahlé a obsahujú veľké množstvo odberateľov. Napríklad Slovenská Pošta a.s. obsluhuje celé územie SR a jej distribučná sieť pozostáva zo 4 hlavných distribučných centier, 42 regionálnych distribučných centier a 1265 lokálnych distribučných centier, z ktorých je služba poskytovaná koncovým odberateľom. Počet odberateľov v takýchto systémoch je obrovský a navrhnuť distribučný systém tak, aby bol efektívny a spĺňal požiadavky na kvalitu a rýchlosť distribúcie je zložitý proces. Obvykle sa distribučná sieť tvorí tak, že sa najskôr určia polohy hlavných distribučných centier, následne sa určia polohy centier

¹ Ing. Peter Matis, PhD., Žilinská univerzita v Žiline, Fakulta riadenia a informatiky, Katedra dopravných sietí, Univerzitná 8215/1 010 26 Žilina, Tel.: +421-41-5134225, E-mail: Peter.Matis@fri.uniza.sk

v nasledujúcej vrstve atď. až po poslednú uvažovanú distribučnú vrstvu. Na záver sa určia pre koncové distribučné centrá distribučné rajóny pre jednotlivé vozidlá, ktoré obsluhujú odberateľov. Aj takto separované pod-problémy sú veľmi náročné na riešenie a zhlukovanie zákazníkov alebo medzi-skladov do zhlukov môže výrazne pomôcť pri ich riešení.

2. ZHLUKOVANIE

Zhlukovanie predstavuje proces spájania niekoľkých odberateľov na základe určitých vlastností, obvykle polohy, do zhluku. Na zhluk sa následne pozeráme ako na samostatnú jednotku a rozmer úlohy sa tým redukuje. V príspevku sú skúmané možnosti zhlukovania odberateľov reprezentovaných v grafe G vrcholmi V . V literatúre sa okrem zhlukovania vrcholov vyskytuje aj zhlukovanie hrán alebo plôch [9].

Existuje veľké množstvo heuristik, ktoré môžu byť použité pri zhlukovaní. Ja som vytvoril alebo implementoval nasledujúce heuristiky:

- Mriežkové zhlukovanie používa mriežku určitého tvaru (štvorcovú, trojuholníkovú, šesťhrannú, atď.). Heuristika začína s veľkým počtom buniek v mriežke. Ak to dovoľujú kapacitné alebo iné obmedzenia, zákazníci z dvoch buniek sú spojení do zhluku a väčšia bunka je vytvorená. Proces je opakovaný pokiaľ kapacitné alebo iné obmedzenia nedovoľujú spojiť žiadne dve susediace bunky.
- Stieracia heuristika začína z určitého bodu v priestore a nechá rásť určitý geometrický tvar, obvykle kruh, okolo tohto bodu. Expanzia je zastavená, keď sú prekročené kapacitné alebo iné obmedzenia. Následne je nový bod mimo tohto útvaru vytvorený a algoritmus pokračuje s expanziou geometrického tvaru okolo tohto bodu. Algoritmus je ukončený, keď sú všetci odberatelia zahrnutí do zhlukov. Prekrývajúce sa časti geometrických útvarov sú priradené prvému zhluku.
- Vzdialenostné zhlukovanie [6] je špeciálnym prípadom stieracej heuristiky, keď sú Euklidovské vzdialenosti nahradené vzdialenosťami po dopravnej infraštruktúre, ktorá spája odberateľov a distribučné centrá.
- Líniové zhlukovanie [6] je špeciálnym prípadom stieracej heuristiky. Odberatelia sú vložení do jedného zhluku, ak sú umiestnení na jednej ulici (línii). Výsledné zhluky majú špecifický pozdĺžny tvar.
- Segmentové zhlukovanie [6], [10] je špeciálnym prípadom líniového zhlukovania. Odberatelia sú vložení do jedného zhluku, ak sú umiestnení na jednom segmente ulice (od križovatky po križovatku).
- Logické zhlukovanie [6] využíva určitý typ informácie, napr. adresu, na tvorbu zhluku.
- Zhlukovanie pomocou K -centier [1] je dobre známou heuristikou, ktorá vyžaduje aby počet zhlukov k bol dopredu známy.
- Fuzzy K -center zhlukovanie [2] je fuzzy verziou K -center heuristiky. Bola použitá lineárna fuzzy funkcia príslušnosti v tvare lichobežníka.
- Pažravá heuristika [2] je ďalšou verziou K -center heuristiky. Cieľom je priradiť odberateľa k zhluku na základe najkratšej vzdialenosti po infraštruktúre bez ohľadu na kapacitné a iné

obmedzenia. V ďalšej fáze sú odberatelia medzi zhlukmi vymieňaní, tak aby boli kapacitné a iné podmienky splnené.

- Zhlukovanie na základe histogramu [3] vytvorí histogram výskytu počtu (frekvencie) odberateľov v dvojrozmernom priestore. Na histograme sú hľadané vrcholy. Podľa dopredu určeného počtu zhlukov k je vybrané k vrcholov histogramu. V miestach týchto vrcholov sú umiestnené centrá budúcich zhlukov. Ďalšia fáza zhlukovania môže prebiehať podobne ako v pažravej heuristike.

3. KVALITATÍVNE KRITÉRIA TVARU ZHLUKU

V mnohých publikáciách sa pri tvorbe distribučnej siete zohľadňujú zväčša celkové náklady, alebo návratnosť investícií [16]. Menej sa v publikáciách zohľadňujú veličiny, ktoré sú ťažko merateľné. Medzi také patrí pri zhlukovaní aj tvar výsledného zhluku. Tvar vytvoreného zhluku je pre mnohé praktické aplikácie veľmi dôležitý [2], [11], [12]. Je potrebné zabrániť zhlukom ktorých tvar je komplikovaný, alebo sa prekrývajú, alebo sú tvorené niekoľkými samostatnými plochami. Tieto neprirodené zhluky sa niekedy nazývajú aj *gerrymanders* [4].

Z pohľadu kvalitatívnych parametrov zhluku je dôležité rozloženie odberateľov v zhluku. Rovnomerné rozloženie odberateľov pomáha zvýšiť efektívnosť distribúcie. Kvantitatívne meranie kvality agregácie je subjektívny proces. Pre rozmiestnenie odberateľov v zhluku bola použitá miera umiestnenia odberateľov voči gravitačnému ťažisku zhluku [5]:

$$DGRB_j = 2 * \left(1 - \frac{|\hat{O}_j|}{|O_j|} \right) - 1 \quad (1)$$

Kde $|\hat{O}_j|$ je počet odberateľov, ktorí patria do zhluku j a vzdialenosťou je im bližšie gravitačné ťažisko iného zhluku. $|O_j|$ je počet všetkých odberateľov v zhluku j . V ideálnych prípadoch je množina \hat{O}_j prázdna a výraz dosahuje hodnotu 1. V prípadoch, keď množina \hat{O}_j obsahuje veľký počet odberateľov, by mohol výraz (1) nadobúdať záporné hodnoty. Pre tieto prípady a hodnoty blízke 0 sa hodnota výrazu stanoví na 0,1. Vzdialenosť je meraná ako Euklidovská vzdialenosť.

Ďalšou použitou mierou je normalizovaná vzdialenosť odberateľov od medoidu zhluku. Medoid je reprezentovaný odberateľom, ktorý je najbližšie ku gravitačnému ťažisku zhluku.

$$DMED_j = 1 - \frac{\sum_{i=1}^{N_j} dis(cu_{ji}, m_j)}{ND} \quad (2)$$

$$ND = \frac{\max(dis(cu_l, cu_k))}{K}; \quad \forall l, \forall k \quad (3)$$

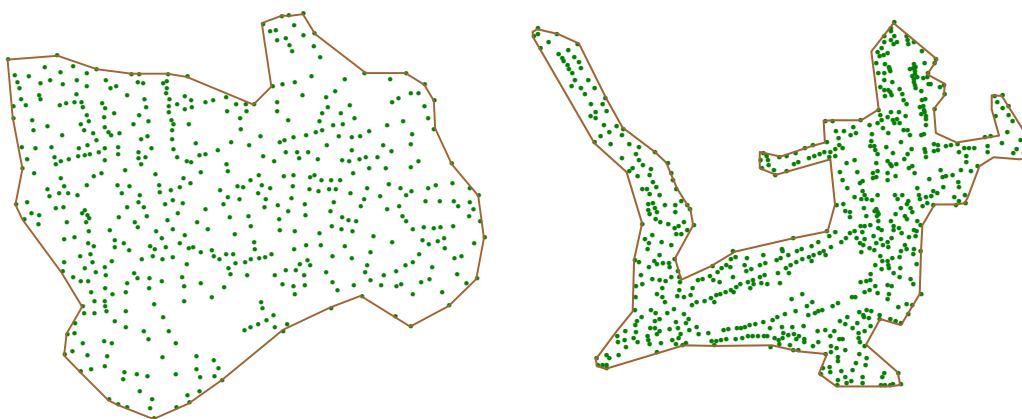
Kde $dis(cu_{ji}, m_j)$ je cestná vzdialenosť medzi odberateľom i a medoidom zhluku j . N_j je počet odberateľov v agregáte j . ND je odhadovaná najväčšia cestná vzdialenosť medzi odberateľmi zhluku. K je predpokladaný počet agregátov pre obec alebo sídelný celok. $\max(dis(cu_l, cu_k))$

predstavuje maximálnu cestnú vzdialenosť pre odberateľov celého sídelného celku. Veličinu normalizujeme do intervalu $\langle 0,1; 1 \rangle$.

Poslednou mierou je kompaktnosť zhuku:

$$COMP_j = \sqrt{\frac{Area_j}{AreaC_j}} \quad (4)$$

Kde $Area_j$ je plocha najmenšieho nekonvexného obalu okolo odberateľov, ktorí patria do zhuku j . $AreaC_j$ je plocha kruhu, ktorý má rovnakú dĺžku obvodu ako je dĺžka obvodu tohto nekonvexného obalu. V ideálnom prípade, keď sú zákazníci sústredení do kruhového útvaru dosahuje hodnota kompaktnosti 1. Pre prípady keď rozloženie odberateľov tvorí tvarovo komplikované útvary sa môže hodnota kompaktnosti blížiť k číslu 0. Z praktických dôvodov limitujem kompaktnosť intervalom $\langle 0,1; 1 \rangle$. Obrázok 1 ukazuje výsledky zhukovania s dobrou a malou hodnotou kompaktnosti.



Zdroj: Autor

Obr. 1 - Agregáty s kompaktnosťou 0.79 a 0.35

Bol uskutočnený experiment s niekoľkými heuristikami pre tvorbu agregátov a boli sledované priemerné hodnoty vyššie spomenutých veličín. Tabuľka 1 obsahuje výsledky experimentu. Experiment bol vykonaný nad 52 sídelnými celkami s celkovým počtom odberateľov 75 238.

Tab. 1 - Kvalitatívne vlastnosti vybraných zhukovacích heuristik

Heuristika	Počet agregátov	Priemerná <i>DGRB</i>	Priemerná <i>DMED</i>	Priemerná <i>COMP</i>
Štvorcová mriežka	339	0.485	0.535	0.613
Vzdialenostná	321	0.513	0.564	0.587
K-center	312	0.561	0.575	0.631
Fuzzy K-center	302	0.581	0.583	0.617
Pažravá	309	0.492	0.562	0.643
Histogram	320	0.512	0.512	0.532

Zdroj: Autor

4. ODHAD PARAMETROV ZHLUKU

Pri riešení úloh tvorby distribučnej siete pomocou zhlukovania je potrebné odhadnúť parametre, ktoré sú nutné a sledované pri riešení neagregovaného prípadu. Medzi také najčastejšie patrí odhad času obsluhy, odhad dĺžky trasy pri obsluhu odberateľov, odhad požadovaného množstva. Dobrú vzorku algoritmov na odhad dĺžky trasy pre jeden zhuk poskytuje napríklad [8]. Experimentálne som zistil, že odhady jednotlivých autorov sú vhodné pre prípady v extravilánoch obcí. Pre intravilány obcí som navrhol vlastné odhady dĺžky trasy pre obsluhu zhuku. Nové odhady boli porovnané s odhadmi publikovanými v [13].

$$TSP(D) = 0.765\sqrt{NA} \quad (5)$$

$$TSP(D) = [0.83 - 0.0011(N + 1) + \frac{1.11S}{N + 1}]\sqrt{NA} \quad (6)$$

Kde $TSP(D)$ je odhadovaná dĺžka trasy pre obsluhu odberateľov zhuku. A je plocha zhuku. S je pomer medzi dlhšou a kratšou stranou opísaného obdĺžnika okolo zhuku, kde pomer je vypočítaný tak, aby to bolo číslo väčšie alebo rovné 1. Autori pri výpočtoch vždy používajú Euklidovské alebo Manhatenské vzdialenosti.

Navrhol som dve nové formulácie odhadu dĺžky trasy pre obsluhu odberateľov zhuku:

$$TSP(D) = \max_{i,j \in zhuku} (c_{ij}) / v \quad (7)$$

Kde c_{ij} je vzdialenosť medzi odberateľmi i a j po infraštruktúre. v je priemerná rýchlosť pohybu vozidla.

$$TSP(D) = 1.445 * SD \quad (8)$$

Kde SD je celková dĺžka všetkých uličných úsekov obsiahnutých v skúmanom zhuku. Vzťah (8) bol vytvorený regresiou s využitím vzorky 78 agregátov a ich porovnaním s výsledkami najlepších riešení neagregovaných prípadov. Tabuľka 2 porovnáva hodnoty odhadov dĺžky trasy potrebnej na obsluhu odberateľov v agregáte. Experiment bol vykonaný celkovo na 350 agregátoch, ktoré boli vytvorené vyššie spomenutými heuristikami na zhlukovanie.

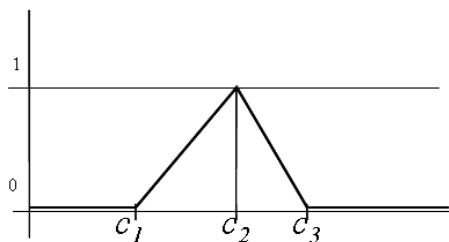
Tab. 2 - Porovnanie odhadu dĺžky trasy na obsluhu agregátu

Odhad	Pomer	Odchýlka
(5)	1.99	0.616
(6)	0.56	0.371
(7)	0.54	0.327
(8)	1.03	0.129

Zdroj: Autor

Požiadavky odberateľov majú kolísavý charakter a sú dni, keď niektorí odberatelia nemusia žiadať obsluhu. Pri zostavovaní agregátov som aproximoval požiadavky odberateľov fuzzy funkciou príslušnosti v tvare trojuholníka [7], [14]. Fuzzy funkcia príslušnosti v tvare trojuholníka podľa obrázku 2 je konvexná a nazýva sa aj trojuholníkové fuzzy číslo [7] a je ju možné popísať nasledujúcim zápisom $\mu_c : \mathbf{c} = \langle c_1, c_2, c_3 \rangle$. Funkcia príslušnosti μ_c popisuje mieru nášho presvedčenia, že hodnota konkrétnej požiadavky odberateľa patrí k \mathbf{c} . Základňou trojuholníkového

čísla \mathbf{c} rozumieme tú časť fuzzy množiny, kde hodnoty funkcie príslušnosti μ_c spĺňajú požiadavku $\mu_c > 0$. Pre trojuholníkové číslo \mathbf{c} je dĺžka základne definovaná výrazom $z_c = c_3 - c_1$.



Zdroj: Ramík, J., Vlach, M. [7]

Obr. 2 - Trojuholníkové fuzzy číslo

Pri zhlukovaní odberateľov je potrebné jednotlivé požiadavky odberateľov sčítať. V prípade dvoch trojuholníkových fuzzy čísel platí nasledovné tvrdenie:

$$\mathbf{a} + \mathbf{b} = \langle a_1, a_2, a_3 \rangle + \langle b_1, b_2, b_3 \rangle = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle \quad (9)$$

Dôkaz tvrdenia (9) je možné nájsť v [7]. Využitím tvrdenia (9) je možné odhadnúť celkové požiadavky odberateľov (hmotnostné, objemové, časové) v zhluku. Je možné skúmať vplyv jednotlivých zhlukovacích heuristik na dĺžku základne výsledného trojuholníkového čísla a rozptyl tejto veličiny. Bol vykonaný experiment, kde trojuholníkovým číslom bol ohodnotený čas obsluhy odberateľa. Trojuholníkové číslo bolo vygenerované pre každého odberateľa náhodne v pevne stanovenom rozsahu a zostalo rovnaké pre daného odberateľa pre všetky vykonané experimenty. Tabuľka 3 sumarizuje výsledky experimentu.

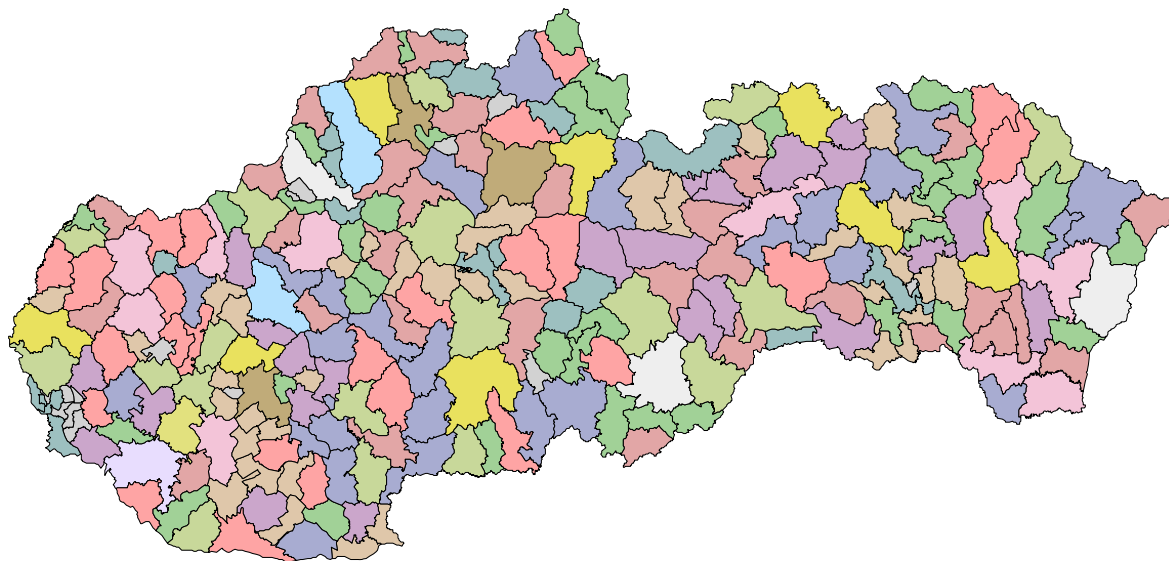
Tab. 3 - Hodnota priemernej dĺžky základne fuzzy čísla reprezentujúceho celkovú dobu obsluhy zhluku a jej rozptylu pre vybrané zhlukovacie heuristiky

Heuristika	Priemer dĺžky základne [min]	Rozptyl dĺžky základne [min]
Štvorcová mriežka	67.2	45.4
Vzdialenostná	71.3	43.2
K-center	65.2	38.4
Fuzzy K-center	66.6	37.8
Pažravá	61.3	41.2
Histogram	70.5	44.7

Zdroj: Autor

Z experimentu vyplýva, že najvyvázenejšie heuristiky sú fuzzy K-center a K-center, ktoré vykazujú najmenší rozptyl základne výsledného trojuholníkového čísla. Naopak pomerne veľký rozptyl je viditeľný pri heuristike s využitím štvorcovej mriežky.

Obrázok 3 dokumentuje použitie pažravej heuristiky pre určenie spádových oblastí distribučných centier pre územie SR. Pre výpočet boli stanovené obmedzenia na počet sídelných celkov v spádovej oblasti (max. 40) a počet obyvateľov (max. 80 000).



Zdroj: Autor

Obr. 3 - Hranice lokálnych distribučných centier získané pažravou heuristikou



Zdroj: Autor

Obr. 4 - Zhluky odberateľov v Žiline a okolí vytvorené Fuzzy K-center heuristikou.

Obrázok 4 dokumentuje použitie zhlukovacej heuristiky Fuzzy K – center na určenie koncových distribučných rajónov pre mesto Žilina a širšie okolie. Pre výpočet bolo stanovené obmedzenie na počet odberateľov v jednom zhluku (max. 300).

5. ZÁVER

V článku boli spomenuté možnosti použitia zhlukovacích metód pre tvorbu distribučnej siete na makro aj mikro úrovni. Zhlukovanie je vhodnou technikou redukcie zložitosti úlohy pri veľkých dopravných a distribučných problémoch. Niekoľko mier pre meranie kvality bolo použitých a jednotlivé heuristiky boli porovnané. Pre experimenty boli použité reálne údaje. Pre dosiahnutie výsledkov v reálnom čase pri úlohách reálnych rozmerov je potrebné odhadnúť dĺžku trasy potrebnej na obsluhu zhluku a ďalšie kvantitatívne parametre zhluku. Boli navrhnuté nové metódy odhadu týchto dĺžok a porovnané s existujúcimi odhadmi. Nové odhady vo všeobecnosti dosahujú lepšie výsledky pre úlohy v intravilánoch obcí. Bola preskúmaná citlivosť riešenia pri zmene vstupných parametrov s využitím fuzzy logiky.

Problém návrhu distribučnej siete je zložitý a nebolo cieľom tohto článku popísať celý proces návrhu. Existuje veľké množstvo parametrov a obmedzení, ktoré sa v praktických úlohách vyskytujú. Tieto môžu ovplyvniť možnosti použitia metód zhlukovania na tvorbu reálnej distribučnej siete. Predmetná úloha bude cieľom ďalšieho výskumu.

POUŽITÁ LITERATURA

- [1] MACKAY, D. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003, ISBN 0-521-64298-1
- [2] MATIS, P. Decision support system for solving the street routing problem, *Transport 3*, 2008, ISSN 1648-4142, s. 230 – 236.
- [3] RICHARDS, J.A., JIA, X. *Remote Sensing digital Image Analysis, an introduction*, Springer, 2005, ISBN 978-3-540-25128-6
- [4] ALTMAN, M. *Districting principles and democratic representation*, PhD. thesis, California Institute of Technology, 1998, Dostupné z <http://etd.caltech.edu/etd/available/etd-05192004-142452/unrestricted/dissertation_full.pdf>
- [5] SLIVONĚ, M. Řešení problému lokace hubů pomocí genetického algoritmu, *Perner's Contacts 4*, 2008, ISSN 1801-674X, s. 96 – 108.
- [6] MATIS, P. Management of street routing problems using decisions support system, *Communications 3*, 2006, ISSN 1335-4205, s. 5 – 8
- [7] RAMÍK, J., VLACH, M. Concept of generalized concavity based on triangular norms, *Journal of Statistics and Management Systems 1*, 2003, ISSN 0972-0510, s. 86 – 106
- [8] FIGLIOZZI, M.A. Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer demands and Routing Constraints, *Transportation Research Record: Journal of the Transportation Board*, No 2089, 2008, s. 1 – 8.
- [9] MUYLDERMANS, L., CATTRYSSE, D., OUDHEUSDEN, D.V. District design for arc-routing applications, *Journal of the Operational Research Society 11*, 2003, s. 1209 – 1221.
- [10] OPPEN, J., LOKKETANGEN, A., Arc routing in a node routing environment, *Computers & Operations Research 4*, 2006, s. 1033–1055.
- [11] POOT, A., KANT, G., WAGELMANS, A. A Saving based method for real-life vehicle routing problems, *Journal of the Operational Research Society*, 2002, s. 57-6.
- [12] TANG, H., MILLER-HOOKS, E. An interactive heuristic for a practical vehicle routing problem with solution shape constraints, *85 annual meeting of the Transportation research Board*, 2006, 20 s.

- [13] KWON, O., GOLDEN, B., WASIL, E. Estimating the length of the optimal TSP tour - an empirical-study using regression and neural networks, *Computers and Operations Research* 22(10), 1995, s. 1039-1046
- [14] GÁBRIŠOVÁ, L. Kapacitne obmedzená lokační úloha v podmímkach neistoty, *Perner's Contacts* 5, 2008, ISSN 1801-674X, s. 98 – 106.
- [15] ŠIROKÝ, J., SLIVONĚ, M., CEMPÍREK V. Centra nákladní dopravy a jejich optimalizace na vybrané dopravní síti, *Perner's Contacts* 2, 2008, ISSN 1801-674X, s. 81 – 94.
- [16] FIDLER, T. Lokace a alokace z hlediska návratnosti, *Perner's Contacts* 3, 2009, ISSN 1801-674X, s. 83 – 88.