

PETER MATIS, Ph.D.
E-mail: Peter.Matis@fria.utc.sk
University of Zilina,
Faculty of Management Science and Informatics
Department of Transportation Networks
Univerzitná 8215/1, 010 26 Zilina, Slovak Republic

Distribution Logistics
Preliminary Communication
Accepted: Nov. 27, 2009
Approved: Oct. 28, 2010

CREATION OF NEW DISTRIBUTION NETWORK – DISTRIBUTION OF MAIL IN SLOVAK REPUBLIC

ABSTRACT

In recent years the providers of logistic services are under big pressure to lower their expenses and improve their quality. One way of accomplishing this task is the centralization of logistic activities. The paper presents possible methods for restructuring the distribution network on the case of Slovak Post p.l.c. By reducing the number of centres for distribution and by creating larger centres one can expect improvement in the process efficiency. There is no simple method to find the optimal distribution network. One possible way to tackle this problem is to use aggregation methods to group customers on the micro and macro levels. Several aggregation methods, including two new ones are presented and compared for the centres' and delivery districts. New measurements for the quality of aggregation are created and tested on real data using all the discussed aggregation methods.

KEY WORDS

network creation, aggregation, quantitative methods, fuzzy logic, distribution districts

1. INTRODUCTION

In recent years we have seen many centralization activities in postal delivery companies to improve the performance of their processes [11], [12], [13]. The Slovak Post p.l.c. has been providing universal services for Slovakia. The concept of universal service was explained by the EU Commission in its communication of 11 September 1996 on Services of General Interest in Europe (COM(96) 443 final). The aim of universal service is to ensure the provision of high-quality services at affordable prices to all. In this paper we want to address the need for cutting down the expenses in the process of mail distribution keeping the level of quality guaranteed by the universal service.

The distribution of mail by the Slovak Post is a hierarchical process. It is distributed by large volumes from 4 Main Distribution Centres (MDC) to 42 Regional Distribution Centres (RDC). From RDC the mail is distrib-

uted using regularly routes by cars every day or twice a day to 1,265 Distribution Centres (DC). From DC the mail is distributed usually using public transport and on foot by mail carriers to Distribution Districts (DD). In recent years the Slovak Post p.l.c. has been restructuring the distribution network. Part of that process is reducing the number of DCs and creating larger centres to improve the process efficiency. The Slovak Post is automating the process of mail sorting to the district level thus leaving more time available for delivery. This also decreases the number of wrong mail deliveries and increases the quality of service. It also opens a gap for the decrease in the number of mail carriers or the ability to offer this free time of mail carriers for new services.

On the other hand, the process of restructuring requires some large investments in the automation of mail sorting, because currently this process is automated up to the DC level. We evaluate the expenses for the distribution of mail to DC and then to DD. Other expenses are related to physical distribution of mail carriers. Currently, they use public transport or walk to get to their districts. New scenario plans the use of cars for moving the mail carriers to and from districts. The mail carrier is driven by car to the first point of the service area, serves the area on foot, and then returns to the depot by car.

In this paper we will deal only with the costs that are related to the number of DCs created, number of districts created and their relation to DC. The paper also focuses on expenses that are related to distribution in DD. Even such part of the system is complex and it is impossible to find an optimal solution in a reasonable computing time. These problems have to be separated and first a good set of DCs and assignment of each village or city district to these DCs found. In the second step a new DD with account of available time for the delivery has to be created. Both problems are complex and difficult to solve within reasonable time. Good aggregation techniques can size down the complexity of them to solvable sizes.

A solution is proposed for a regular season. A solution is provided for the peaks that appear in the distribution during the year. These can be solved separately or could be part of the agreement between employer and the workers.

2. AGGREGATION

Aggregation is the process of combining several customers into one large pool. By aggregating customers, the problem size decreases and sometimes there is an improvement in the quality of delivery and reduction in costs. In postal deliveries there is a node aggregation. Each aggregate can be treated as a node in the new aggregation process. From that perspective there is an aggregation at the micro-level and at the macro-level. In this paper both of these perspectives will be used. The customer can be included in several aggregates or always in one aggregate. For our problem the second option is used.

There are several techniques at the micro and macro levels that can be used for aggregation. We have created or used several of them:

- Grid aggregation uses a grid structure (e.g. triangle, square, and hexagon). It starts with the structure with many cells. Customers from two cells are aggregated and a larger cell is created. The process can be repeated and it stops when no neighbour cells can be aggregated. This aggregation requires starting cells to be small enough so that the customers in one cell fit the capacity and other restrictions.
- Sweep aggregation starts from some point in space and expands the shape (e.g. circle) until the restriction limits are reached. Then a new sweeping point is created and the algorithm continues with the next aggregate. The overlapping parts of shapes are assigned to the first one.
- Distance aggregation [5] is a special case of sweep aggregation when a circle is used and the distances are measured on the road network that connects the customers and the centres.
- Line aggregation [5] is a special case of sweep aggregation when customers are aggregated along the network infrastructure (roads) and create aggregates with line shape.
- Segment aggregation [5] is a special case of line aggregation when the aggregation stops also when line segments end (usually in the crossing of roads).
- Logical aggregation [5] uses logical information (e.g. address or block) to connect the customers to an aggregate.
- K-means aggregation [1] is a well-known aggregation method that requires knowing the number of aggregates to create - k .
- Fuzzy K-means aggregation [2] is fuzzy (soft) version of K-means aggregation. The triangle fuzzy

membership function, based on the road distances, is used for the measurement that a customer is a member of a particular aggregate.

- Greedy aggregation [2] is another version of K-means aggregation with the goal of minimum distances regardless of other conditions and then parts of the aggregates can be exchanged to meet the capacity and other conditions.
- Histogram peak aggregation [3] creates a histogram of frequency of customers in space and then finds the peaks in this histogram. Usually, the number of peaks is known and it is the number of aggregates. Peaks are ordered by frequency and tops, up to the limit, are used. Centres of aggregates are placed in these peaks.

For both problems the shape of the aggregate is important. We want to avoid the shapes that are complicated or cross one another or are split to several separated polygons. These non-natural shapes are called also gerrymanders [4]. To measure the shape parameters there are several measurements available:

- The centre of aggregate is usually measured as a gravity centre of all customers in the aggregate. Sometimes it is the weighted gravity point, where each customer can have different weights depending on some parameters (e.g. the average time a mail carrier spends delivering to this customer).

$$c_x = \frac{\sum_{i=1}^N w_i cu_{x_i}}{N}; \quad c_y = \frac{\sum_{i=1}^N w_i cu_{y_i}}{N} \quad (1)$$

where c_x , c_y are computed coordinates of the centre, N is the number of customers in the aggregate, w_i is the weight of customer i , cu_{x_i} , cu_{y_i} are coordinates of customer i .

- Medoid is the customer that has sum of all distances to all customers in the aggregate minimal. It is similar to centre, but the point has to be placed in the coordinates of customer and also for the distance measurements other kinds of measurements could be used (e.g. distance over the transportation network).
- The radius is usually measured as the average distance of boundary customers to the gravity centre.

$$r = \frac{\sum_i \sqrt{(cu_{x_i} - c_x)^2 + (cu_{y_i} - c_y)^2}}{N} \quad (2)$$

- The single link is the shortest distance between two aggregates.

$$s_l(A_j, A_m) = \min(\text{dis}(cu_{ji}, cu_{mk})); \quad \forall cu_{ji} \in A_j; \forall cu_{mk} \in A_m; j \neq m \quad (3)$$

where A_j , A_m are aggregates j and m , operator dis is distance measurement operator, cu_{ji} , cu_{mk} are customer i from aggregate j and customer k from aggregate m .

- The complete link is the farthest distance between two aggregates. Usually it is calculated similarly as a single link, but the maximum is used.
- The average distance between two aggregates is important for merging of two aggregates into one. It can be calculated as a distance between their centres, medoids or as follows:

$$\text{dis}(A_j, A_m) = \sum_{i=1}^{N_j} \sum_{k=1}^{N_m} \frac{\text{dis}(cu_{ji}, cu_{mk})}{N_j N_m} \quad (4)$$

And for dis operator the Euclidean distance, network distance, Manhattan distance, etc. can be used. N_j, N_m is the number of customers in cluster j and m .

It is important to measure the distribution of customers in the aggregates and their shapes. The human inspection on the output may be the most intuitive validation method as it compares the result with the decision-maker's intention in a natural way. However, human inspection lacks the scalability to large and complicated problems. In addition, manual inspection is not always desirable and not feasible in real-life applications. Therefore, quantitative assessment of aggregation quality is of great importance for these applications. The quantitative measure of the aggregation quality is subjective. The following measurements have been used:

- The distribution of customers in aggregate j is: [5]

$$DGRB_j = 2 * \left(1 - \frac{|\hat{O}_j|}{|O_j|}\right) - 1 \quad (5)$$

where $|\hat{O}_j|$ is the number of customers that belong to aggregate j and are closer to the centre of another aggregate, and $|O_j|$ is the number of all customers in aggregate j . For good solutions this number equals 1. For poor solutions this number is -1. For practical use the resulting values are limited to interval $\langle 0.1, 1 \rangle$. The Euclidean distance has been used for the measurement of distance between the customer and the centre of the aggregate.

- The normalized distance of customers in aggregate j from the medoid is:

$$DMED_j = 1 - \frac{\sum_{i=1}^{N_j} \text{dis}(cu_{ji}, m_j)}{ND} \quad (6)$$

$$ND = \frac{\max(\text{dis}(cu_l, cu_k))}{K}; \quad \forall l, \forall k \quad (7)$$

where m_j is a medoid of aggregate j , as distance dis the distance along the given network of routes has been used, ND is calculated as the expected maximum network distance among customers in aggregate, K is the number of aggregates to be created, $\max(\text{dis}(cu_l, cu_k))$ represents maximum network distance in the studied area because it is calculated for all l and k . Good aggregates give the normalized distance to medoid close to 1. Poor aggregates can give a value close to 0. For practical

use we limit the resulting values to the interval $\langle 0.1, 1 \rangle$.

- The compactness is a measure of the aggregate's shape and the definition published in [8] has been used:

$$COMP_j = \sqrt{\frac{Area_j}{AreaC_j}} \quad (8)$$

where $Area_j$ is the area of smallest non-convex hull around the customers in aggregate j and $AreaC_j$ is the area enclosed by a circle that has the same perimeter as the non-convex hull around the customers. Good aggregates have values close to 1. For practical use the resulting values are limited to interval $\langle 0.1, 1 \rangle$. There are many non-convex hulls for a set of points and finding the smallest one is the difficult task of NP. This is estimated by creating the non-convex hull using shape onion-peeling algorithm [9].

2. LOCATIONS OF DC AND NEW DISTRICTS - EFFECT ON TRANSPORTATION

Location of DC is important in the delivery of services and goods on the large-scale territory with a high-density transport network and uniform or non-uniform settlement. The problem of centres location is complex and consistent if the location is studied from the different views of different groups of people that are influenced by these locations.

The criteria used for the location of DCs and their districts represent the economical and sociological parameters that have major impact on the distribution process. The study covered the entire Slovak Republic. The main goal has been to slim down the operations and increase the efficiency of the existing network of DCs with the maintenance of the quality for services provided by the Slovak Post. The following criteria have been used:

- Expenses used for transportation of postal matters. To lower these expenses one has to place the DC in the area where most of the customers are located. Decreasing the number of DCs lowers these expenses as well. Good district allocation can lower incurring costs from the distribution of postal matters from DC to DD.
- The shape of districts for DC can greatly influence not only the economy of transportation but all the processes in logistics. The best shapes are those that have good compactness and distribution of customers in the district.
- The accessibility of services provided by the DC for potential customers and speed of delivery are key variables for the maintenance of quality for the universal service. Long distances between DC and DD create usually problems in regular deliveries.

These criteria have been evaluated and the triangle fuzzy membership function [6] for each criterion was

created. These membership functions were combined to one norm using the triangular norm (t-norm) introduced by Dubois & Prade (1980) [10]. This t-norm with parameter λ is:

$$\mu_{ABC\lambda} = \frac{\mu_A * \mu_B * \mu_C}{\max(\mu_A, \mu_B, \mu_C, \lambda)} \quad (9)$$

where $\mu_{ABC\lambda}$ is the resulting membership function, μ_A , μ_B , μ_C are membership functions for each evaluation criterion, λ is the parameter in range $\langle 0, 1 \rangle$. We have used $\lambda = 0.5$. The resulting membership function gives values in $\langle 0, 1 \rangle$.

From the perspective of DC the customers are villages and city districts. These can be represented as centres of the settled area. Several parameters for these centres were assigned: the number of inhabitants, the number of companies with higher volume of mail, the average volume of deliveries in the previous year, the current number of DDs. All these parameters were weighted and a linear combination form was used for the total measure of importance for each village or city district. These inhabited units cannot be divided between two DCs.

Our experimental sample was from all cities villages in Slovakia. The data were collected manually from source maps ZM 1:10,000, purchased from the Geodetic and Cartographic Institute of the Slovak Republic. Some of the data were collected using GPS receivers. These data are not publicly available at the moment.

Several aggregation methods mentioned in the previous section were used to compare the results from the fuzzy membership function and the measure of aggregation quality. The number of new DCs was set at 260 (currently there are 1,416 DCs). Then, these aggregations were run with different starting conditions ten times. In total 2,600 clusters for each heuristics were received and Table 1 presents the resulting average values of the measurements of quality for five aggregation heuristics.

In Table 1 the average $\mu_{ABC\lambda}$ represents an average value of the total membership function for all of the cases in the experimental sample. Higher value of $\mu_{ABC\lambda}$ means a better solution.

In our experiments K-means and fuzzy K-means heuristics produced the clusters that met the best combined criterion $\mu_{ABC\lambda}$ from the economic perspective and visual attractiveness of the DC districts. The

square grid heuristics and both K-means generated compact shapes. Greedy heuristics is strong in the creation of aggregates where customers are close to medoids of the aggregates. Figure 1 shows one output of DC districts for Greedy heuristics.

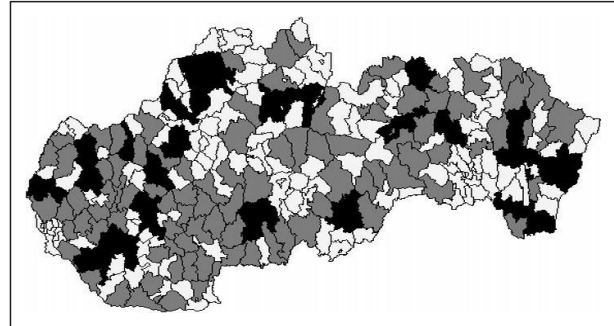


Figure 1 - 260 DC districts from the Greedy heuristics

3. CREATION OF DD

Created districts for DC have on the average 9,500 real customers, where service personnel (mail carriers) have to deliver the post items. Each customer is represented as a point on the map and in reality it is one address where the delivery has to be made. In such large agglomerations, service can be performed using mixed transportation modes. The service personnel are driven by car to the first point of the service area, serve the area on foot, and then return to the depot by car.

One key problem in solving very large distribution problems is to aggregate customers into natural clusters. Each cluster is then treated as one DD and is served by one mail carrier. The size of the cluster is determined by the estimated length of the mail carrier's route, the travel speed of the mail carrier, the delivery time for customers in the cluster, and the total time available for serving the cluster.

By aggregating customers the problem size is decreased. On the other hand, one must assume that error is introduced, as customers may be aggregated that do not belong to the same cluster in the optimal solution. In reality, optimality is not the goal, but getting a good solution in a short period of time is crucial. If aggregation can help in getting such a good, but not necessarily optimal solution quickly, this will be of interest.

Table 1 - Quality of different aggregation heuristics for the DC district creation

Heuristics	Average $\mu_{ABC\lambda}$	Average DGRB	Average DMED	Average COMP
Square Grid	0.147	0.623	0.215	0.658
Distance	0.254	0.535	0.198	0.584
K-means	0.321	0.643	0.186	0.636
Fuzzy K-means	0.342	0.617	0.205	0.628
Greedy	0.286	0.487	0.248	0.515

In literature there are some good examples of how other authors estimate the length of the Traveling Salesman Problem (TSP) [7]. The topology of the network for these cases is specific and the following estimation of the length for one mail carrier’s path was used:

$$TSP(D) = 1.445 * SD \tag{10}$$

where *SD* is the total distance of all the street segments that are served by one mail carrier. This formula was created from a simple regression. The experiment for the estimation of *TSP(D)* was done using sample data with 78 service clusters. On each cluster several heuristics described by Matis [2] were used to estimate the total length of the mail carrier’s path. The shortest calculated distance was used. *Figure 2* summarizes the results of the experiment. The horizontal axis represents the total distance of all the street segments *SD*. The vertical axis represents the length of the shortest found TSP route.

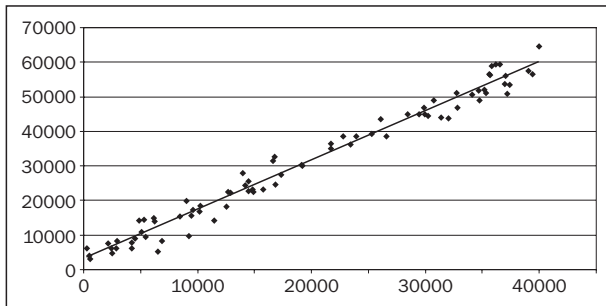


Figure 2 - Estimation of TSP(D)

To create the DD we have to estimate the service time (time that the mail carrier has available for the delivery) and compare it with the available service time. The estimated time should not exceed the available service time.

$$ESWT_j = \left(\sum_{i=1}^{N_j} CST_{ji} + TSP(D)_j \right) \leq SWT \tag{11}$$

where *ESWT_j* is estimated service time for DD *j*, *SWT* is available service time of the mail carrier, *CST_{ji}* is the average service time for customer *i* that belongs to aggregate *j*.

The calculation of *DMED_j* was modified to:

$$DMED_j = 1 - \frac{\sum_{i=1}^{N_j} \text{dis}(cu_{ji}, m_j)}{NDVI} \tag{12}$$

$$NDVI = \frac{\max(\text{dis}(cu_i, cu_k))}{KVI}; \quad cu_i \in VI; \quad cu_k \in VI \tag{13}$$

where *NDVI* is calculated as the expected maximum network distance among customers in the aggregate, *KVI* is the number of aggregates to be created for the village or city districts, $\max(\text{dis}(cu_i, cu_k))$ represents maximum network distance for the studied village or city district, *VI* is a set of all customers in the studied village or city district. The values from updated DMED formula for DD are usually higher than the values from their calculation for DC districts, because the density of customers in a village is higher than the density of village across the entire state. We normalized it to the range of <0.1, 1>.

Fuzzy K-means aggregation was used for DC districts creation. As shown in the previous section, this heuristic gives the best combined economic value and reasonable other measurements of quality. For each DC several aggregation heuristics have been applied.

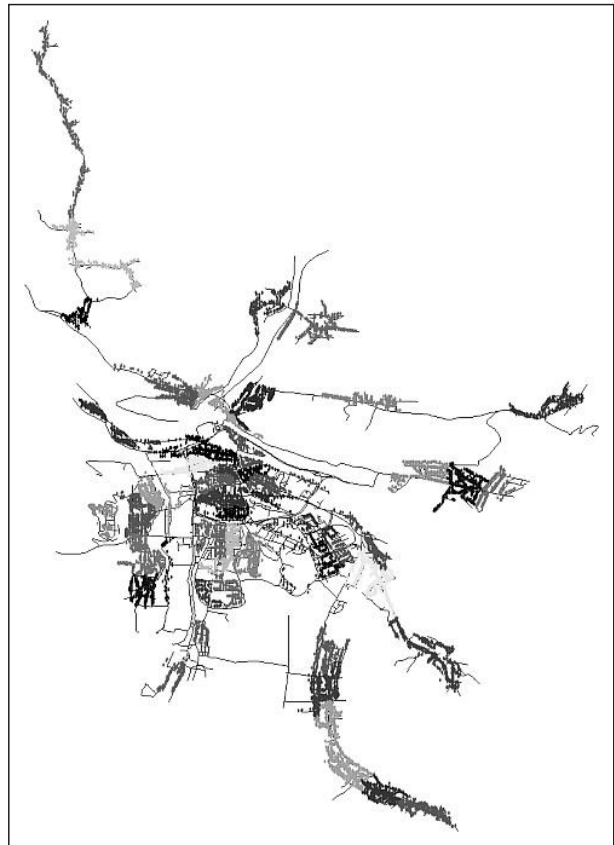


Figure 3 - Distribution districts created in the wider Zilina vicinity using Fuzzy K-means heuristics

Table 2 - Quality of different aggregation heuristics for the DD creation

Heuristics	Average TSP(D) [km]	Number of DDs	Average DGRB	Average DMED	Average COMP
Square Grid	7.212	5,745	0.467	0.542	0.572
Distance	6.954	5,593	0.492	0.573	0.553
K-means	6.853	5,186	0.554	0.589	0.594
Fuzzy K-means	6.537	4,987	0.576	0.603	0.592
Greedy	6.487	5,254	0.503	0.592	0.582

Table 2 provides the important results. The best and the worst heuristics vary more than 15% in the resulting number of DDs for the whole Slovakia.

4. DISCUSSION

The aggregation of customers is a valuable technique for solving large transportation and distribution problems. Several measurements for the quality of aggregation have been proposed and tested in the practical cases. Good measurements are compactness of the created shape, the average distance to medoid, the distribution of customers in the aggregate. Several known and two own heuristics for the aggregation in two specific cases were tested. In the first case we aggregated the villages and city districts to large districts named DCs. All heuristics have been used for three different numbers of DCs and one result with 260 DCs was presented.

For the second case customers have been aggregated from each city and village in each created DC separately and a new DD was created. The results of Fuzzy K-means were by 4% better than the real number of DDs currently used for delivery in Slovakia.

5. CONCLUSION

Aggregation can improve the efficiency of the delivery network. In this paper some of the possible methods for aggregation with the use of real data have been tested. For practical use of the method it is necessary to evaluate the real transportation expenses in the current system after the aggregation. It is also important to understand the functional dependency of micro level aggregation on the previous used macro level aggregation.

Ing. PETER MATIS, Ph.D.

E-mail: Peter.Matis@fria.utc.sk

Žilinská univerzita, Fakulta riadenia a informatiky

Katedra dopravných sietí

Univerzitná 8215/1, 010 26 Zilina, Slovak Republic

TVORBA NOVEJ DISTRIBUČNEJ SIETE - DISTRIBÚCIE POŠTY V SLOVENSKEJ REPUBLIKE

Poskytovatelia logistických služieb sú v posledných rokoch pod veľkým tlakom na znižovanie nákladov a zlepšovanie kvality služieb. Jedným zo spôsobov na vyriešenie tejto úlohy je centralizácia logistických aktivít. Článok ukazuje možné spôsoby reštrukturalizácie distribučnej siete na prípade Slovenskej pošty a.s. Znižovaním počtu distribučných centier a ich zväčšovaním môžeme očakávať zefektívnenie procesov vykonávaných v týchto centrách. Neexistuje jednoduchá metodika na nájdenie optimálnej distribučnej siete. Jednou z možných metód riešenia tejto úlohy je použitie zhlukovacích metód na zhlukovanie zákazníkov na mikro a makro úrovni. Niekoľko agregáčnych metód včítane dvoch

nových je prezentovaných a porovnaných na úrovni tvorby distribučných centier a ich rajónov a doručovacích rajónov. Nové metódy merania kvality zhlukovania sú prezentované a testované na reálnych údajoch s použitím všetkých diskutovaných zhlukovacích metód.

KLÚČOVÉ SLOVÁ

tvorba siete, zhlukovanie, kvantitatívne metódy, fuzzy logika, doručovacie rajóny

LITERATURE

- [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, Vol. 23, No. 3, 2008, ISSN 1648-4142, pp. 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, http://etd.caltech.edu/etd/available/etd-05192004-142452/unrestricted/dissertation_full.pdf
- [5] Matis, P.: *Management of street routing problems using decisions support system*, Communications, Vol. 7, No. 3, 2006, ISSN 1335-4205, pp. 5 – 8
- [6] Ramik, J., Vlach, M.: *Concept of generalized concavity based on triangular norms*, Journal of Statistics and Management Systems, Vol. 5, No. 1, 2003, ISSN 0972-0510, pp. 86 – 106
- [7] 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, pp. 1 – 8.
- [8] Smith, M., Goodchild, M.F., Longley, P.A.: *A comprehensive guide to principles, techniques and software tools*, Matador, 2007, ISBN 978-1905886-69-9.
- [9] Fadili, M.J., Melkemi, M., ElMoataz, A.: *Non-convex onion-peeling using a shape hull algorithm*, Pattern recognition letters, Vol. 25, No. 14, 2004, ISSN 0167-8655, pp. 1577 – 1585.
- [10] Dubois, D., Prade, H.: *Fuzzy sets and systems - Theory and applications*, Academic press, 1980, ISBN 0-12-222750-6.
- [11] Tabak, P., Kljak, T.: *Traffic flows model of postal items based on input-output demands in the public postal network: case study of Croatia*, Promet - Traffic&Transportation, Vol. 21, No. 6, 2009, ISSN 0353-5320, pp. 407 – 414.
- [12] Trupac, I.: *More competitiveness for Slovenia and its companies through the Slovenian transport logistics cluster*, Promet - Traffic&Transportation, Vol. 20, No. 1, 2008, ISSN 0353-5320, pp. 19 – 30.
- [13] Safran, M., Babic, D., Tomasic, D.: *Defining the optimization criteria for the functioning of logistics and distribution centres*, Promet - Traffic&Transportation, Vol. 20, No. 1, 2008, ISSN 0353-5320, pp. 59 – 65.