British Journal of Applied Science & Technology
18(4): 1-10, 2016; Article no.BJAST.30676
ISSN: 2231-0843, NLM ID: 101664541

SCIENCEDOMAIN *international*
www.sciencedomain.org

A New Efficient Machine Learning Algorithm to Solve Facility Location Selection Problem of Geoinformatics

Ali Sever[1*]

[1]*Department of Computer Information Systems, Pfeiffer University, Charlotte, NC 28109, USA.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/BJAST/2016/30676
Editor(s):
(1) Xu Jianhua, Department of Geography, East China Normal University, China.
(2) Harry E. Ruda, Stan Meek Chair Professor in Nanotechnology, University of Toronto, Director, Centre for Advanced Nanotechnology, University of Toronto, Canada.
Reviewers:
(1) Hassan A. Karimi, University of Pittsburgh, USA.
(2) Helmi Zulhaidi Mohd Shafri, Universiti Putra Malaysia (UPM), Malaysia.
(3) V. J. Rehna, Visvesvaraya Technological University, India.
(4) Jiang Minlan, Zhejiang Normal University, China.
(5) Oyeranmi Adigun, Yaba College of Technology, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/17503

Original Research Article

Received 25th November 2016
Accepted 5th January 2017
Published 12th January 2017

ABSTRACT

The problem of selecting store locations has received increased attention in the literature during the past decade, and varieties of models have been promoted to select those sites. In this paper, we address the problem of finding the optimal deployment of site locations in a certain geographic area with a given wide range of factors affecting decision making. This problem is complex and should be tackled as a multiple-objective problem. The combination of several criteria in the selection of store location must be considered. It should be noted that facility location problems require knowledge of a key parameter, "aggregated degree of importance" (ADI) indicator. This is where the discrete inverse problems can help. In this article, we discuss existing models on this problem and sketch how inverse problems (IP) can be formulated to yield a smooth ADI indicator surface. The latter is very useful both in the accurate locating of facilities as well as in computing sensitivities. A computational analysis on a non-spatial and spatial data set describing the algorithm in the site selection problem illustrates the effectiveness of the approach. Application of the facility location model is demonstrated using an example of a drug store's selection problem in a given area.

**Corresponding author: E-mail: ali.sever@pfeiffer.edu;*

Keywords: Multi-objective optimization; facility location; inverse problems; geoinformatics.

1. INTRODUCTION

Geographic facility location selection is commonly defined as the process of: *"determining the probable spatial location of a store within the context of the locations of, and the spatial and non-spatial relationships between, the various existing store sites"* ([1,2 and 3]).

Facility location selection decision is a critical element in decision making with regard to the success of management for a wide range of organizations. The goal of this process is to determine the most appropriate location to provide good customer service, to attain a competitive advantage, to improve the distribution network, and to create new businesses and markets. Facility location selection is vitally important with regard to the success of companies. Location of the facility is one of the most important variable factors affecting the profitability of a business. Due to its vital importance, the decisions of facility location selection have to be made.

1.1 Related Work

The selection of location problem has been an important research topic during the last several years. Some studies classify facility location problems as a single objective supply chain design problem ([4,5 and 6]). Bhatnagar and Sohal in their empirical study show that it is a multi-objective problem influenced by a wide range of decision making factors such as business services, cost, infrastructure, government, labor, customers, suppliers, and competitors. Nwogugu [7] discussed approaches for retrieval of input data for a new facility location model created specifically for a retail store location. In [8,9,10,11 and 12], Literature Review on Selection Criteria of Store Location is studied extensively. Therefore, the facility location problem should be tackled as a multi-objective problem and a wide range of factors should be included in any successful design.

In [13], it was indicated that the facility location problem is, even in medium-sized problems, a computationally difficult problem, and solving it directly often would not produce any meaningful results. Many existing models use continuous-attributed data, not semantic data ([14]). Similar problem in manufacturing systems with stochastic demand is studied in [15]. Here, we developed an alternative approach to solve the facility location problem described below.

1.2 Our Contribution

The inverse problem approach to machine learning provides for a bottom-up approach where our goal will be to express a generalization based on a known conclusion-that is, the existing store location data points and an ADI indicator surface- through the application of inverse resolution and inverse implication. Inverse problems, like most of the inverse problems encountered in science and engineering may be reformulated as an optimization problem. Therefore, many available techniques of solving the optimization problems are available as methods of solving the IPs.

A reasonable and realistic approach is to think of the aggregated degree of importance (ADI) indicator as a surface, $\sigma = \sigma(x)$, rather than a constant. Here we assume that one cluster (set of locations) Ω contains existing locations while another cluster Ω_k contains locations, which are potential.

There are many typical selection criteria and factors for the selection criteria for the selection of facility location. We will recap a literature review on a selection criteria of store Location and implementation of data gathering process in Table 1 ([16,17,18,19 and 20]).

Table 1. The classification of factors relevant to facility location

Factor	Data type	Source type
Economic Factors	Non-spatial	Commercial
Saturation Level	Spatial	Public
Business services	Non-spatial	Commercial, field work
Competitive Condition	Spatial	Commercial, field work, public
Store Characteristics	Non-spatial	Public, commercial
Customer/market	Spatial	Public, commercial
Suppliers/resources	Spatial	Commercial, field work

In our study, the main criteria have been identified in compliance with the most suitable facility location definition objective. The most convenient main criteria has been defined with regards to economic factors, saturation level, competitive conditions, population structure, physical features for the stores, and the location of preferred site for the most suitable facility location selection.

To be able to produce facility location selections, we use and modify the model discussed in [13], using a manageable subset of a dataset, assuming that a relationship between a facility location selection and some of its various attributes exists. As discussed in [13], we will treat this problem as an inverse problem of finding the $\sigma = \sigma(x_1) > 0$ in

$$-\text{div}\,(\sigma\nabla u) = \delta\,(-x_2) \text{ in } \Re^2 \quad (1)$$

where δ is the Dirac delta function, and σ is a bounded and measurable reconstruction of volatility of the ADI (Aggregated Degree of Importance) indicator function. We assume Ω (includes existing facility locations) is a bounded domain in $\Re^2$ and $\sigma = 1 + \sigma_1$ with supp$\sigma_1 \subset \Omega$. The solution $u(x_1, x_2)$ is given for x_1, $x_2 \in \Omega_k$ (potential new locations), which is a bounded domain in $\Re^2$, whose closure does not intersect. Notice that this imposed constraint incorporates prior knowledge about the problem. We refer to [13] for the detailed formulation of Eq. (1). In Eq. (1), u is the ADI indicator function (i.e., what we are trying to determine in Ω - potential locations with a degree of importance and what is known in Ω_k – existing locations. A value of σ is required before Eq. (1) can be of practical value. One should notice that σ is not directly observable but can be determined in two different ways. The first is to use a 1-dimensional regression problem. But this does not yield the actual facility location for any known ADI indicators for existing locations. The second approach is to calculate implied volatility to yield a value for σ. Implied volatility is determined by solving an inverse problem involving similar facility locations on the same underlying complex ADD indicator infrastructure. Implied volatility is thus that the value ofσ, when substituted in to Equation (1), yields known facility location. Hence, for each facility location, there is a corresponding implied volatility, and one such value can be used to determine the "close" approximation of a new facility location on the same underlying complex ADD indicator infrastructure.

However, in this model, there are problems with implied volatility use as explained in [13]. These kinds of inverse problems are usually "ill-posed." The choice of σ can become very important for the sensitivity calculation and so an arbitrary choice may not lead to a desired results. Therefore we will treat volatility as a surface rather than constant. Thus, we may think of an inverse problem of facility location selection problem as determination of the distribution of expected store locations implied by known store locations.

There are many challenges to solve these kinds of problems due to their nonlinearity, for most computational methods, we assume that a linear problem is a good approximation to a nonlinear one. Here we modified the solution that was suggested in [13] to split the inverse problem into a simpler, ill-posed problem with an integral equation of a Riesz-type kernel and a well-posed problem. This will allow us to isolate and better control the propagation of errors due to the ill-posedness.

In order to compute the aggregated degree of importance indicator for existing facilities, a weighted sum of individual facility goodness indicators is computed. Importance of each indicator can be determined according to the results of empirical studies on practical importance of different facility location criteria. The model presented below uses just five facility location criteria although other criteria mentioned in [8] can be incorporated, if necessary. The main difference of the proposed model from the traditional discrete facility location models is that ADI indicator is not specified by discrete data points and these data points are not assigned to the specific location. Instead of that the ADI indicator approximates the ADI indicator around the potential facility, and facilities are located at sites with high ADI indicator density.

The rest of the paper is organized as follows: the next Section describes our model and justifies its use. In Section III, we formulate the proposed linearized unsupervised learning algorithm and present main experimental results and discussions. Finally, we conclude the paper with some remarks and future work in Section IV.

2. A LINEARIZED APPROACH

In this section, we modify the linearization of the inverse problem in Eq. (1) using perturbation methods and reduce our original non-linear inverse problem to the linear integral Eq. (1) with the Riesz-type kernel,

$$w(x) = (1/2\pi)^2 \int_{\Omega} \sigma_1 (y) \, |x-y|^{-2} dy \quad x \in \Omega^* \tag{2}$$

For detailed theoretical derivation of the perturbation Eq. (2) we refer to [13].

New linearized inverse problem can be stated as follows: Find an ADI indicator function $\sigma_1 \in L_\infty(\Omega)$, given the function $F(x_k) = w(x_k)$, $x_k \in \Omega_k$, and using the following integral equation of first kind

$$A\sigma_1 (x) = F(x) \quad x \in \Omega_k \tag{3}$$

where

$A\sigma_1 (x) = \int_{\Omega} k(x,y)\sigma_1 (y)\, dy$ and $k(x,y) = (1/2\pi)^2 |x - y|^{-2}$

and A is considered as an operator from $L_\infty(\Omega)$ into $L_\infty(\Omega_k)$.

For the integral Eq. (3), the uniqueness result and a logarithmic type of stability estimate with a Riesz-type kernel and non-intersecting domains Ω and Ω_k is provided in [13].

We, first, make clear connection between regularization theory for inverse problems and machine learning to introduce a new machine learning regularization algorithm. Then we apply the proposed algorithm together multi-objective decision support system to an actual problem of selecting store locations.

This problem is complex and has been tackled as a multiple-objective problem. We noted that facility location problems require knowledge of a key parameter, "aggregated degree of importance" indicator and can be stated as machine learning regularization of an inverse problem. Then we discussed existing models on this problem and sketched how discrete inverse problems could be formulated to yield a smooth ADI indicator surface with an optimizer which is needed for unsupervised learning.

Inverse problems (IP), like most of the inverse problems encountered in science and engineering may be reformulated as an optimization problem. Therefore, many available techniques of solving the optimization problems are available as methods of solving the IPs. Here we use the Tikhonov regularization method to solve the ill-posed problem in Eq. (3). In this method, instead of Eq. (3), we solve the following regularized equation:

$$(A^*A + \alpha \mathbf{I})\sigma_1 (\alpha) = A^*F \tag{4}$$

where α is a regularizer parameter and $\mathbf{I}$ is the identity operator.

A known theory of regularization guarantees existence of the solution σ_1 (α) and its convergence to the solution f when $\alpha \to 0$, provided σ_1 exists and the uniqueness for the original equation is known [13]. Now one has to prove that Eq. (4) has a solution, which is given below.

Theorem 1: Let S(y) be the set of all solutions to

$$\min_{x \in X_0} \; || Ax - y ||_Y \tag{5}$$

where x is solution to Eq. (4), X_0 is a compact subset of X, and X= Ω and Y = Ω_k in Eq. (5). Then S(y) is continuous with respect to y.

Proof: Suppose we pick up any sequence $y_k \in Y$ such that $y_k \to y_0$ for all $y_0 \in Y$ as $k \to \infty$ and $\varepsilon > 0$ such that $d(S(y_k), S(y)) > \varepsilon$. That is, there exists $x_k \in S(y_k)$ such that

$$||x_k - x||_X > \varepsilon \tag{6}$$

for any $x \in S(y)$.

Now let $x_0 \in X_0$. Since X_0 is compact so the lower semi-continous functional $\Phi(x;y) = ||Ax - y||_Y^2$ has a minimum point x^* on X. The set of all min points is closed deu to semicontinuity of Φ. We denote it by S(y). Then since $x_k \in X_0$ and X_0 is compact, by extracting a subsequence, we can assume that $x_k \to x' \in X_0$. Therefore we have

$$\Phi(x_k;y_k) \le \Phi(x;y_k) \text{ for any } x \in X_0 \tag{7}$$

Since x_k's are minimum points in X_0. Also since Φ is lower semi continuous with respect to x_k, we can pass to limit

$$\Phi(x';y) \le \Phi(x;y) \text{ for all } x \in X_0 \tag{8}$$

and therefore $x' \in S(y)$. On the other hand $||x' - x||_X \ge \varepsilon$ for any $x \in S(y)$ and this contradiction proves that S(y) is continuous with respect to y. Since X_0 is a compact subset of X there exists a solution to (5).

Let Ω and Ω_k be the domain in $\Re^2$. For computational reasons, we will regard the integral operator

$$A\sigma_1 (x) = \int_{\Omega} \sigma_1 (y) / |x-y|^2 \, dy \qquad (9)$$

as defined from $L_2(\Omega)$ into $L_2(\Omega_k)$. By using the definition of an adjoint operator in L_2, we have $A^*: L_2(\Omega_k) \rightarrow L_2(\Omega)$ defined by

$$A^*F(y) = \int_{\Omega^k} F(x)/|x-y|^2 \qquad y \in \Omega \qquad (10)$$

and $A^*A\sigma_1$ (y) becomes

$$A^*A\sigma_1 (y) = \int_{\Omega_k} \int_{\Omega} \sigma_1 (y') \, |x-y|^2 \, |x-y'|^2 \, dy' dx \qquad (11)$$

where $x \in \Omega_k$, and y, y' $\in \Omega$. By discretizing Eq. (11), we have

$$A^*A\sigma_1 (y_i) = \sum_{k,j=1}^{n} w_j \, w_k \; \sigma_1 (y_j^{'}) / |x_k - y_i|^2 \, |x_k - y_j^{'}|^2 \qquad y_i, y_j^{'} \in \Omega \qquad (12)$$

$$\cong \sum_{j=1}^{n} \Theta_{i,j} \sigma_1 (y_j^{'})$$

where

$$\Theta_{i,j} \cong \sum_{k,j=1}^{n} w_j \, w_k \; / |x_k - y_i|^2 \, |x_k - y_j^{'}|^2 \qquad (13)$$

and w_j, w_k are the weight functions. By discretizing Eq. (10), we have

$$A^*F(y_i) \cong \sum_{k=1}^{n} w_k F(x_k) / |x_k - y_i|^2 \qquad (14)$$

From Eqs. (12), (13), and (14) we have the discretized matrix equation in the form of

$$(\alpha \mathbf{I} + \Theta) \, \sigma_1 (y) = A^*F \qquad (15)$$

for some regularization parameter, $\alpha > 0$. Notice that the linearization method led us to a Fredholm integral equation of first kind with Riesz kernel. Since this equation represents an ill-posed problem, we needed some kind regularization method to overcome this difficulty. We used the Tikhonov regularization method and, by discretizing the regularized equation (15), ended up with the system of linear equations. With this enhancement, now the problem is reduced to solving systems of linear equations. This is done conventionally, using MATLAB routines. Typical run times are in seconds for several machines and operating systems.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Our particular model is designed to investigate the volatility of ADI indicators that can be used to predict accurately what facility locations will be selected based on several criteria and determine the aggregated degree of importance of the potential facility locations. The model will then populate a Google map with the existing stores of data given as input. These data are marked on the map with colored dots. Potential facility locations and their aggregated degree of importance, on the other hand, are marked with thumbtack icons. Next, the model calculates the circular area containing the area of interest. This is displayed in a large red circle. Then, it calculates the smallest circular area with the highest ADI indicator. This is displayed in a medium-sized blue circle. Using the proposed model, estimation is made as to the most probable area where the best location resides. This is displayed in a small green circle.

Experimental studies are conducted to measure computational performance of the proposed model using distributed real data sources and to analyze decision-making results obtained using the proposed model-solving algorithm.

The sample store location problem considered in this paper deals with locating drug stores. There are a number of existing facility location sites and the total number of facilities to be open is limited. It is aimed to locate drug stores at sites having the largest number of customers and the smallest number of competitors in its proximity, reasonable traffic density, and having acceptable real estate costs.

The necessary input data for modeling are ADI indicators and addresses of existing facility location sites. The number of potential sites is varied from 5 to 10 - what corresponds to small to medium-sized facility location problems. The value of the coverage radius is 15 miles. The model is solved using: (1) the algorithm described in Section 2; and (2) the Geocoding service, a geographical information system is used to represent the modeling results by using spatial data. The importance of each ADI indicator in the objective function is set according to the results.

Using the existing facility location and ADI indicator data, the facility location problem is solved using a DIP model-solving algorithm. For the given dimensions of the facility location problem, the model solving time is negligible, and it is within 1 min for 10 existing and 5 potential sites. However, the model solving time can be very long for an increased number of existing and potential facility location problem. Therefore, the time limit is imposed for the model and the optimal solution occasionally is not found within the specified time limit. For the drug store experiment, the model-solving is performed for ten existing stores and generated five candidate facility locations.

The number of selected units usually is only a fraction of existing locations that is affected by impact of competition and real estate cost, limiting incentives to open maximum number of facilities. For the upper level values of r, there must also be a limited number of candidate locations.

The important feature of spatial data processing is an ability to represent data graphically using various cartographical tools. Fig. 1 shows a sample facility location results for ten existing locations and five potential locations within a 15-mile radius. Technically, the graphical representation is created: (1) constructing the skeleton of a web page with JavaScript and (2) loading it into the HTML browser. Map-type can be street, satellite, or hybrid. Displaying the Google map in an HTML widget requires Common Graphics (CG). The CG html-widget always works on MS Windows, and will work on Linux only if you have installed the Mozilla GTK widget as needed.

Table 1 shows the ADI indicators for existing drug stores derived empirically. Ei shows existing location and Pj shows potential locations at location i and j respectively, $1 \le i \le 10$; $1 \le j \le 5$. Notice that existing locations are ranked from best to worst based on five criteria with the aid of expert analysts form the investigated store.

Our test calculations were made on Table 2 and the domain was divided up into the 100-element mesh and the regularization parameter $\alpha = 10^{-7}$ in (21) for the reconstructed potential store location distribution. We did not do a systematic study to determine an optimal strategy for choosing the regularization parameter and the number of grid points. We have to mention that our aim is to investigate the specific characteristics of this method as a technique for reconstruction. The approach used here is to calculate implied volatility to yield a value for σ. The reconstructed potential store locations with importance level using the model is shown in different forms in Figs. 1 and 2. Notice that potential locations are ranked from best to worst based on five criteria with the aid of the suggested model for the investigated store. The Fig. 1 shows that there is strong preference for locations with high population density. As the number of existing locations in question increases, the accuracy of the prediction should also increase, thus weeding out extremes.

Existing and Potential Facility Locations with ADI indicators are displayed in Figs. 3 and 4.

Implied volatility is determined by solving an inverse problem involving similar aggregated degree of importance indicators on the same category of stores. Implied volatility is thus the value of σ that, when substituted into Eq. (1), yields a known aggregated degree of importance indicators of existing stores. Hence, for each aggregated degree of importance indicators, there is a corresponding implied volatility, and one such value can be used to determine the "close" approximation of a new facility location on the same underlying.

Eq. (21) involves a main parameter that must be adjusted for greatest efficiency: the regularization parameter and the number of stores. The numerical experiments described above have shown that the reconstructed surface is smooth and close to the true volatility surface in Fig. 5. The reconstruction was usually a fair representation of the shape of σ.

Table 2. ADI indicators for existing drug stores

Main criteria / Existing stores	**E1**	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**	**E10**
Economic factors	0.21	0.17	0.05	0.09	0.1	0.04	0.03	0.12	0.14	0.05
Saturation level	0.11	0.12	0.08	0.1	0.08	0.07	0.09	0.11	0.13	0.11
Competitive condition	0.17	0.03	0.11	0.12	0.09	0.05	0.06	0.14	0.12	0.01
Population structure	0.14	0.12	0.09	0.07	0.07	0.11	0.12	0.09	0.05	0.14
Store characteristics	0.19	0.11	0.08	0.06	0.09	0.07	0.05	0.11	0.12	0.03

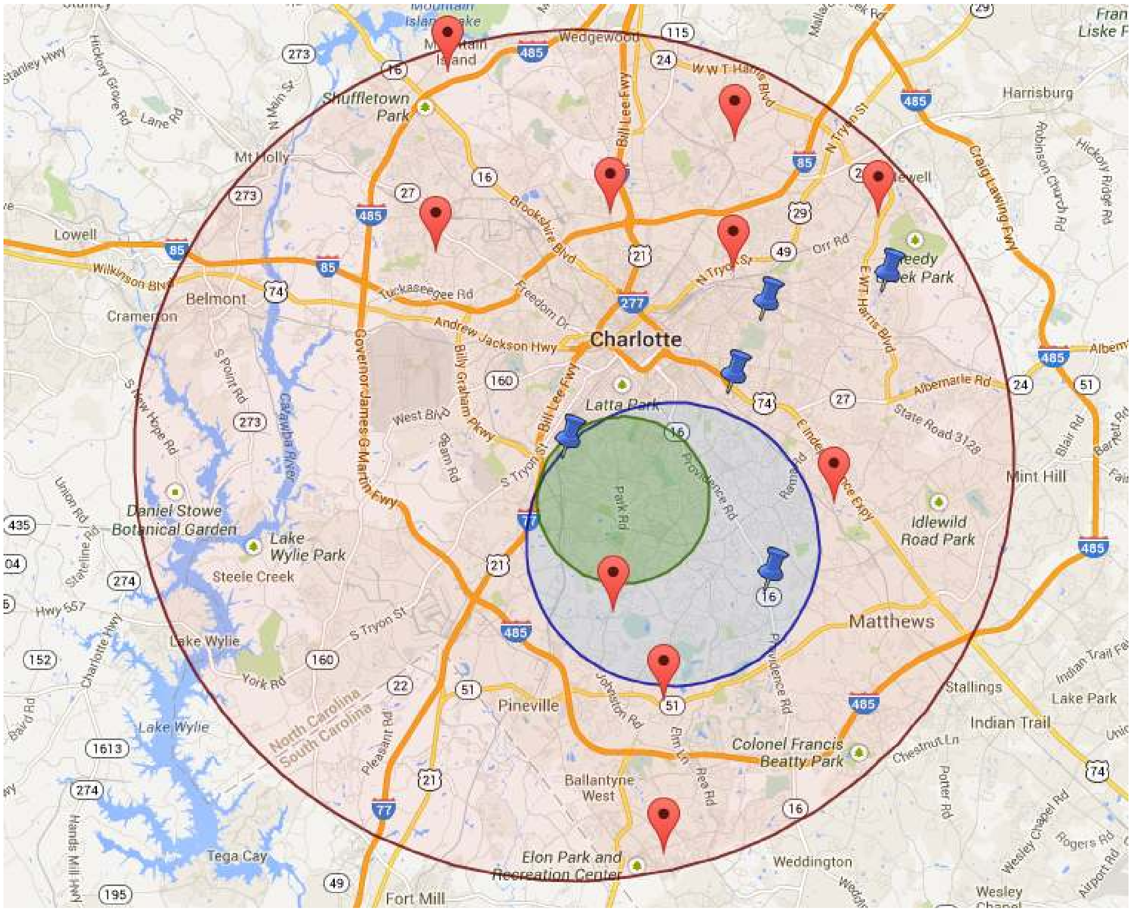

Fig. 1. Reconstructed Potential Store Locations – Marked with thumbtack icons

Fig. 2. Reconstructed Potential Store Locations – Potential Location #1 - #5 (Lat – latitude; Lng – longitude)

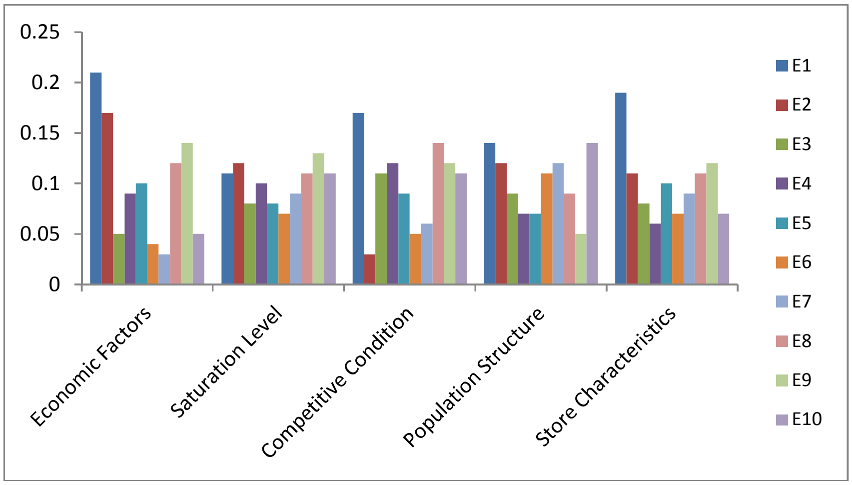

Fig. 3. Existing facility locations with ADI indicators

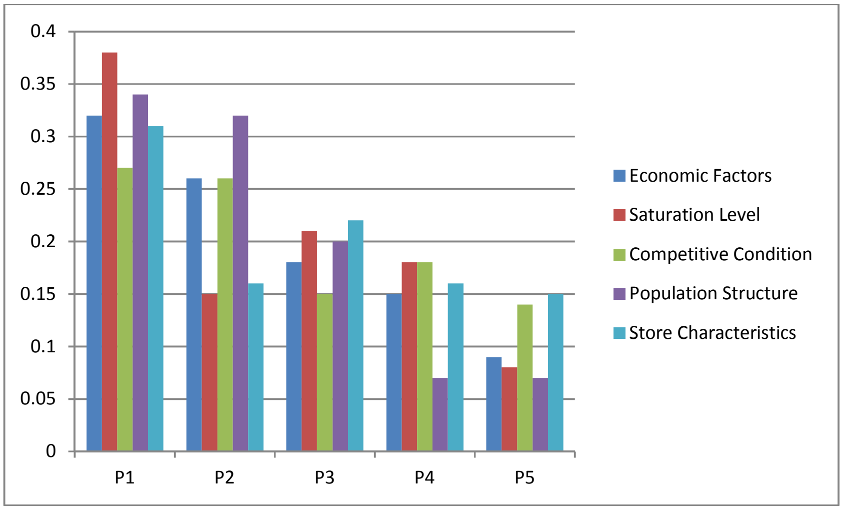

Fig. 4. Potential facility locations with ADI indicators

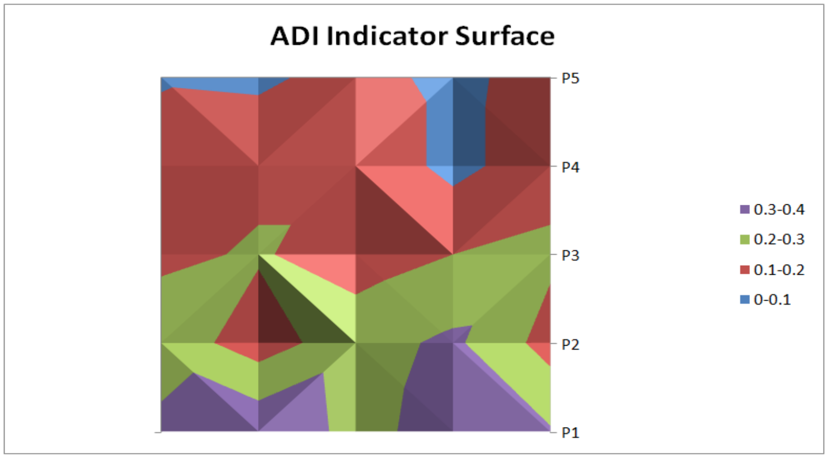

Fig. 5. Reconstructed surface of ADI indicators

4. CONCLUDING REMARKS AND FUTURE WORK

The paper elaborates a facility location model based on utilization of distributed spatial data sources. It develops and tests a computational approach for potential store location data. This approach provides an end-to-end solution starting with data gathering until representation of the results. The proposed data retrieval architecture allows structuring the data retrieval process. However, it still requires involvement of a data retrieval expert and a decision maker cannot be completely relieved from data retrieval concerns. Given the special character of facility location, the graphical representation provides the decision maker with efficient depiction of modeling results.

The implementation of the proposed algorithm shows that the method is reasonably accurate for the reconstruction of volatility of the store locations, which is useful both in the accurate user predictions as well as in the computing sensitivity. Much better results can be achieved if the model used to identify degree of importance appropriately describes dependencies in data.

The algorithm provides satisfactory results compared to the optimal results though there are multiple ways to improve the algorithm. But, this work is by no means exhaustive of the method discussed here. One could do much more detailed work to improve data retrieval process consists of multiple interrelated steps since there is a large number of standards and technologies used in distributed data processing and the effectiveness of the data retrieval process automation in general.

The investigation shows that there are a multiple challenges to be considered for successful utilization of distributed spatial data. Cartographical resources enable representation of modeling results through a combination of data representation layers from multiple sources involving manual operations.

Even though the experimental results show the advantage of the proposed algorithm for multi-objective decision support system, the proposed facility location model can be expanded to include additional criteria subject suggested in literature.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Arrondo GA, Redondo LJ, Fernandez J, Ortigosa MP. Parallelization of a non-linear multi-objective optimization algorithm: Application to a location problem. Applied Mathematics and Computation. 2015;255: 11-124.
2. Carillo M, Jorge MJ. A multi-objective DEA approach to ranking alternatives. Expert Systems with applications. 2016;50:130-139.
3. Erbiyik H, Ozcan S, Karaboga K. Retail store location selection problem with multiple analytical hierarchy process. Procedia - Social and Behavioral Sciences. 2012;58:1405–1414.
4. Julka N, Baines T, Tjahjono B, Lendermann P, Vitanov V. A review of multi-factor capacity expansion models for manufacturing plants: Searching for a holistic decision aid. International Journal of Production Economics. 2007;106(2): 607–621.
5. Melo MT, Nickel S, Saldanha-da-Gama F. Facility location and supply chain management – A review. European Journal of Operational Research. 2009;19: 401–412.
6. Xiaoyuan Zhang, Jianzhong Zhou, Changqin Wang. Multi-class support vector machine optimized by inter-cluster distance and self-adaptive deferential evolution. Applied Mathematics and Computation. 2012;218(9):4973-4987.
7. Nwogugu M. Site selection in the US retailing industry. Applied Mathematics and Computation. 2006;182:1725–1734.
8. Randhawa SU, West MT. An integrated approach to facility location problems. Computers Ind. Eng. 1995;29(1-4):261-265.
9. Seyed Abbas Sadatsakkak, Mohammad H. Ahmadi, Roham Bayat, Seyed Mohsen Pourkiaei, Michel Feidt. Optimization density power and thermal efficiency of an endoreversible Braysson cycle by using non-dominated sorting genetic algorithm. Energy Conversion and Management. 2015;93:31-39.
10. Turhan G, Akalin M, Zehir C. Literature review on selection criteria of store location based on performance measure. 9th International Strategic Management Conference, Published by Elsevier Ltd.; 2013.

11. Yang CL, Chuang SP, Huang RH, Tai CC. Location selection based on AHP/ANP approach. International Conference on Industrial Engineering and Engineering Management. 2008;1148-1153.
12. Newman MEJ. Networks: An introduction. Oxford: Oxford University Press; 2010.
13. Sever A. A problem of data mining in E-Commerce. Applied Mathematics and Computation. 2011;217(24):9966–9980.
14. Grabis J, et al. Use of distributed data sources in facility location. Computers and Industrial Engineering. 2012;63:855-863.
15. Ingerman DV. Discrete and continuous Dirichlet-to-Neumann maps in the layered case. SIMA. 2000;31(6):121-123.
16. Kampars J, Grabis J. An approach to parallelization of remote data integration tasks. Scientific Proceedings of RTU: Computer Science. 2011;49:24–30.
17. Li Z, Yang X, Gu H, Zhang H. Kernel-improved SVM for Semanteme data. Appl. Math. Comput. 2013;219(17):8876–8880.
18. Steve W, Jonathan R. Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. Geoforum. 2012;43:1076–1087.
19. Cai L, Zhu Y. The challenges of data quality and data quality assessment in the big data era. Data Science Journal. 2015;14:2.

    DOI: http://doi.org/10.5334/dsj-2015-002
20. Tavakkoli MR, et al. Design of a facilty layout problem in cellular manufacturing systems with stochastic demands. Applied Mathematics and Computation. 2012;184: 721–728.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/17503