Appl Intell (2016) 44:377–398
DOI 10.1007/s10489-015-0705-7

Fuzzy geographically weighted clustering using artificial bee colony: An efficient geo-demographic analysis algorithm and applications to the analysis of crime behavior in population

Arie Wahyu Wijayanto[1] · Ayu Purwarianti[1] · Le Hoang Son[2]

Published online: 9 August 2015

Abstract Geo-demographic analysis is an essential part of a geographical information system (GIS) for predicting people's behavior based on statistical models and their residential location. Fuzzy Geographically Weighted Clustering (FGWC) serves as one of the most efficient algorithms in geo-demographic analysis. Despite being an effective algorithm, FGWC is sensitive to initialize when the random selection of cluster centers makes the iterative process falling into the local optimal solution easily. Artificial Bee Colony (ABC), one of the most popular meta-heuristic algorithms, can be regarded as the tool to achieve global optimization solutions. This research aims to propose a novel geo-demographic analysis algorithm that integrates FGWC to the optimization scheme of ABC for improving geo-demographic clustering accuracy. Experimental results on various datasets show that the clustering quality of the proposed algorithm called FGWC-ABC is better than those of other relevant methods. The proposed algorithm is also applied to a decision-making application for analyzing crime behavior problem in the population using the US communities and crime dataset. It provides fuzzy rules to determine the violent crime rate in terms of linguistic labels from socioeconomic variables. These results are significant to make predictions of further US violent crime rate and to facilitate appropriate decisions on prevention such the situations in the future.

Keywords Artificial bee colony · Clustering quality · Decision support process · Fuzzy geographically weighted clustering · Geo-demographic analysis

Lists of abbreviation

Terms	**Explanation**
GIS	Geographical Information Systems
GDA	Geo-Demographic Analysis
FCM	Fuzzy C-Means
NE	Neighborhood Effects
FGWC	Fuzzy Geographically Weighted Clustering
ABC	Artificial Bee Colony
FGWC-ABC	Fuzzy Geographically Weighted Clustering based on Artificial Bee Colony
UNO	United Nation Organization
PC	Partition Coefficient
CE	Classification Entropy
SC	Partition Index
S	Separation Index
XB	Xie and Beni's Index
IFV	A spatial cluster validity index

1 Introduction

Geo-demographic analysis (GDA) is a successful implementation of geographical information systems (GIS), when used in private sector, that enables marketers to predict

✉ Le Hoang Son
sonlh@vnu.edu.vn; chinhson2002@gmail.com

[1] School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung, Indonesia

[2] VNU University of Science, Vietnam National University, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

consumer behavior based on statistical models and their residential location [9]. Geo-demographic analysis is the study of the attributes of population demographics based on geographical position, using spatially explicit analytical approaches [23]. GDA is usually performed to investigate the underlying rules from geographical data [26]. In doing the analysis of data, it is clear that common business processes exist across all statistical collections, indicating that there is an opportunity for common system components, techniques and tools to be used [4].

Data mining, as a popular computing technique, is widely used in the process of discovering patterns in data, which is similar to the objective of GDA [38]. For a country with a huge population and household census data, data mining is an ideal approach for analyzing this information [6]. Clustering, a kind of data mining methods is the process to create segmentation from the entire data set into relatively homogeneous subgroups or clusters, in which the similarity of the instances within the cluster is maximized and the similarity to instances outside the cluster is minimized [11, 19, 38]. This method is typically used to break down statistical information to get valuable data from substantial scale records [10].

In GDA, the fuzzy clustering method that typically used is Fuzzy C-Means (FCM) [3], one of a well known data mining algorithm for clustering task. FCM is an improvement method of k-means algorithm which permits a data item to belong to some clusters with a defined fuzzy membership grade [39]. *Fuzzy Geographically Weighted Clustering* (FGWC), a variant of FCM, serves as one of the most efficient algorithms in GDA. It belongs to the class of iterative algorithms aiming to enhance the cluster memberships and centers inspired by the spatial interaction model until a pre-defined stopping condition holds [20, 26, 31]

Numerous current literatures introduced that FCM is efficient, and has excellent performance in handling a large volume of data [14, 39]. The standard FCM must be initiated with a given number of clusters and selects random cluster centers based on it [19]. However, the initialization phase of FCM has typical issues regarding its random selection of the cluster centers which can easily deliver the local optimal solution [14]. This weakness of FCM does not ensure to give the global optimal solution [19]. All of those weaknesses of FCM are also found in FGWC.

There is a popular collection of rules for algorithmic advancement to contrasting optimization issues with only necessary little changes that known as a meta-heuristic [21]. Meta-heuristic algorithms are usually used as the global optimization tools. *Artificial Bee Colony* (ABC) [17], one of meta-heuristic algorithms, is effective when used on optimization of fuzzy clustering [16]. ABC is an optimization algorithm which simulates the insightful searching conduct of a honey bee swarm [17]. It is exceptionally straightforward and very flexible when contrasted to the other current swarm based algorithms [17].

Some authors introduced several metaheuristic approach to optimize FCM clustering such as reported by Karaboga & Ozturk [16], Niu & Huang [22], and Runkler & Katz [24]. Specifically, Runkler & Katz presented new methods for optimizing the reformulated objective functions of FCM model by particle swarm optimization (PSO) [24]. Niu & Huang [22] proposed an enhanced PSO algorithm to overcome the issue of premature convergence of FCM by improving Runkler & Katz [24] model. Karaboga & Ozturk [16] proposed a robust optimization technique of clustering using the ABC algorithm, which is experimentally proven as the better approach for classification purpose than PSO and other popular techniques.

As a summary, the differences of FGWC-ABC with those other previous clustering methods in those literatures are two folds: Firstly, the FGWC-ABC, as derivative of FGWC, is specially designed for the geo-demographic analysis problem which requires the modification of geographical spatial effects to the methods itself and not covered by classical fuzzy clustering such as Fuzzy C-Means; secondly, it is focussed on overcoming the limitation of FGWC in initialization phase by performing ABC to select cluster centers or membership matrix under constraint of objective function minimization.

On the other hand, recent other literatures of FGWC improvement such as in [20, 26, 27, 31] did not focus on overcoming the local optimization issue of FGWC methods. Performing an improvement of this FGWC limitation is more complicated than previous works in FCM, regarding the effect of spatial interaction among clustering elements and its iterative process of geographical weighting that should be considered as the characteristics of FGWC [20].

This research aims to **propose a novel geo-demographic analysis algorithm** that integrates FGWC to the optimization scheme of ABC for **improving geo-demographic clustering accuracy**. The basic idea is utilizing ABC to find the appropriate cluster center and cluster partition through iterative process in a set of geographically weighted clusters as its fitness function. This proposed method is called the *Fuzzy Geographically Weighted Clustering based on Artificial Bee Colony* (FGWC-ABC) algorithm. The performance of the proposed algorithm is evaluated and compared with the relevant clustering algorithms such as FCM, Neighborhood Effects (NE) and FGWC through several benchmark datasets namely the socio-economic demographic from United Nations [36], a benchmark UCI Machine Learning Repository called the Wisconsin Breast Cancer Data [12] and a real dataset of demographic and economic variables from Indonesia Population Census 2010 province level [5].

An important role of the proposed algorithm is to support effectively for the decision-making process. Specifically, the FGWC-ABC algorithm is **applied to a decision-making application for analyzing crime behavior problem in population** using the communities and crime dataset, which consists of socio-economic variables from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR [1]. The decision support process provides i) fuzzy rules to determine the violent crime rate in terms of linguistic labels from socio-economic variables; and ii) a distribution map of the US violent crime rate showing the states, which have very high violent crime rate among all. These results are significant to make predictions of further US violent crime rate and to facilitate appropriate decisions on prevention such the situations in the future.

The remainder of the paper is organized as follows: Section 2 introduces some related references of this research including geo-demographic analysis, the fuzzy geographically weighted clustering and the artificial bee colony optimization. Section 3 presents the proposed method of FGWC-ABC algorithm. Performance evaluation and experimental results on the various geo-demographic datasets are presented in details in Section 4. Section 5 elaborates steps to analyze the crime behavior problem. Finally, some conclusions are made in Section 6.

2 Related works

In this section, a brief overview of the literature that is more relevant to this work is given as a theoretical basis for the proposed method. We firstly describe the geo-demographic analysis principles and the relevant works in Section 2.1. Since the proposed FGWC-ABC used the ideas of FGWC and ABC techniques in geo-demographic analysis, basic concepts of FGWC and ABC are described in Sections 2.2 and 2.3, respectively.

2.1 The GDA principles and relevant works

Geo-demographic Analysis plays a critical role in many current business processes not only in the private sector but also in public issues [9]. It is expressed in numerous current literatures that GDA is generally characterized as the investigation of spatially referenced geo-demographic information, which investigates the individuals based on their residential status [10, 20]. In order to make the geo-demographic data more meaningful and manageable, some clustering methods are utilized in GDA to classify those data into several clusters [31]. There are two primary hidden presumptions in GDA: Firstly, the demographic characteristics of some people that live in the same region are more similar than those who live in different areas. Secondly, we can characterize the profile of two regions regarding their population data that measured using some demographic indicators. Based on these two presumptions, clustering process is performed to group geo-demographic data into meaningful clusters that capture existing regularities, or relevant geo-demographic profiles, therefore making the data more manageable for the further analysis [26, 31]. Fuzzy clustering methods are often used in GDA because they assign a membership value for each area instead of assigning a geographical area to a single group, so that the issues of ecological fallacy can be solved [31]. Geo-demographic have been adopted with apparent success, and delineating the social and demographical profile of small areas [10]. Literature shows that there are some relevant works concerning the applications and algorithms for GDA such as in [8, 27–30, 32–35]. Among all existing relevant works, *Fuzzy Geographically Weighted Clustering* (FGWC) [20] is considered one of the most efficient algorithms for the GDA problem.

2.2 Fuzzy geographically weighted clustering (FGWC)

Fuzzy Geographically Weighted Clustering supports the ability to apply populace and geographical separation impacts for analyzing a geo-demographic cluster in order to improve the standard Fuzzy C-Means algorithm [20]. The influence of one area upon another is considered by FGWC as the product of the populations of the areas. The divisor implements a distance decay effect through the weighting factor. In every cycle of fuzzy clustering, the adjusted cluster membership for the FGWC algorithm is calculated using formulas in (1) [20].

$$u'_i = \alpha \times u_i + \beta \times \frac{1}{A} \times \sum_{j=1}^{n} w_{ij} \times u_j, \tag{1}$$

where u'_i is the new cluster membership of area i and u_i is the old cluster membership of area i. w_{ij} is the weight measuring the amount of interaction between a pair of areas. The weight is decided by distance between centers of the areas or the length of common boundary between them, or both. The A parameter is determined to ensure that the average of weighted membership values is still in the range of zero and one. The α and β are respectively weights to old membership and the mean of membership values of surrounding areas and are defined as in formulas (2–3).

$$\alpha + \beta = 1, \tag{2}$$

The parameters α and β are the scaling variables that influence the ratio of the original membership and the weighted (calculated) membership, respectively [20]. They are also to represent distinctive geo-demographic effect concept, which are the relative significance of demographic characteristic and the spatial interaction respectively [9, 20]. If it is assumed that spatial interaction has the same impact as demographic features of people's behavior, then $\alpha = \beta = \frac{1}{2}$ [9, 20].

$$w_{ij} = \frac{(m_i \times m_j)^b}{d_{ij}^a}, \tag{3}$$

The w_{ij} is the weighting function showing the influence of area i^{th} to area j^{th} defined through formula (3). In this formula, m_i and m_j are the population of areas i^{th} and j^{th}, respectively. d_{ij} is the distance between areas i^{th} and j^{th}. The a and b are user definable parameters, which are determined by considering how critical the role of population and distance in spatial autocorrelation between areas [9, 20]. Value $a = 1$ and $b = 1$ is given if both variables spatial autocorrelation is assumed to have the same degree of interest [9, 20].

FGWC proposed an integration of classical fuzzy clustering concept and geographical sense where the cluster memberships and characteristics evolve throughout the process of fuzzy clustering [20]. FGWC also incorporated geography into geo-demographic analysis so the clusters are sensitive to neighborhood effects and have an influence on the cluster center values to create "geographically aware" clusters. This method proposed an enhancement of the previous research- Neighborhood Effect (NE) by Feng and Flowerdew [9], which gave ex post facto adjustment of the cluster memberships after original fuzzy clustering. Feng and Flowerdew incorporated neighborhood effects after the process of fuzzy clustering. Figure 1 introduces a clear conceptual overview of this strategy and its modification to classical fuzzy clustering method. FGWC characterized the cluster partitions as well as alters its membership matrix by performing iterative geographical adjustments amid clustering iteration.

2.3 Artificial bee colony optimization

Meta-heuristic algorithms, which are frequently nature-inspired, are often efficient in practice in solving difficult optimization problems [41]. Heuristic means 'to find' or 'to discover by trial and error', and meta means 'beyond' or 'higher level' [40]. Artificial Bee Colony optimization is one of the most popular meta-heuristic algorithms [17], which have successfully implemented in many research issues and fields [15]. The most essential element of optimization is the principal algorithms utilized to search optimal solutions to a given problem under different constraints [13, 40, 41]. The colony is designed into three kinds of artificial bees: employed bees, onlooker bees and scout bees. Heaps of nectar are delivered by the employed bees from the nourishment source to the colony hive. They also impart the data about nourishment source in the dancing area. The data contains location of food sources and their certain likelihood. The onlooker bees hold up in the dances territory for

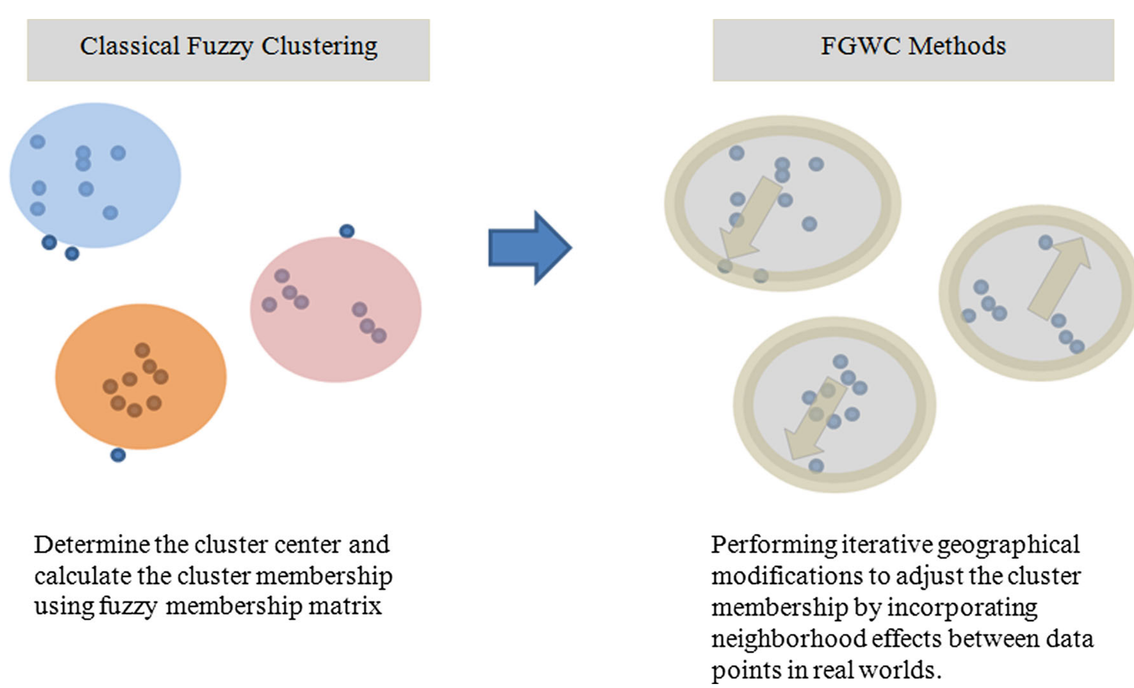

Fig. 1 Conceptual overview of Fuzzy Geographically Weighted Clustering [20]

settling on a choice on the determination of a nourishment source relying upon the likelihood conveyed by employed bees [37]. The calculation of likelihood is focused around the measures of the colony food source. The other sort of bees is the scout bee that does arbitrary looks for new food sources. The employed bee of a relinquished sustenance source turns into a scout and when it discovers another food source it gets to be utilized once more. As it were, each one hunt cycle of the ABC method contains three steps. In the beginning stage, the employed bees are conveyed into their food sources and the measures of nectar are assessed. In the wake of imparting this data about the nectar, the onlooker bees select the food source areas and assess the measure of nectar in the sustenance sources. The scout bees are then picked and conveyed to discover new nourishment sources [37].

3 The proposed method

In this section, we present the FGWC-ABC algorithm in detail. As mentioned beforehand, the FGWC algorithm has limitation in the initialization stage. The cluster centers are created randomly so that they could make the iterative process falling into the local optimal solution easily, which could affect the quality of the resulting cluster. We utilized the Artificial Bee Colony optimization algorithm (ABC) [15] to determine the cluster centers automatically in the initialization stage of FGWC. The preliminary idea also stated in [37]. The objective function which will be minimized is expressed in formula (4).

$$J_{FGWC}(U, V; X) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\|v_i - x_k\|^2 \rightarrow \text{min}, \tag{4}$$

where $U$is a membership matrix, $V$is a cluster center matrix, $X$is a data matrix, $m$is a weighting exponent which determines the fuzziness of the clusters, $u_{ik}$is an element of membership matrix, v_i is a cluster center and x_k is a data point. The cluster centers can be determined in formula (5).

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}}, \tag{5}$$

We also define the membership matrix of fuzzy cluster before geographical modification in formula (6). It employs the cluster center and the data point to calculate an appropriate element of membership matrix.

$$u_{ik} = \frac{1}{\sum_{j=1}^{C}\left(\frac{\|v_i - x_k\|}{\|v_j - x_k\|}\right)^{\frac{2}{m-1}}}. \tag{6}$$

After being calculated, we modify the membership matrix using formulas in (1–3). Inspired by popular method to optimize objective function in classical fuzzy clustering as reported by Runkler and Katz [24], we can reformulate the subsequent computation of cluster centers in formula (5) into a new objective function which only contains cluster center as a single independent variable, as described in formula (7).

$$J_{FGWC}(V; X) = \sum_{i=1}^{c}\sum_{k=1}^{n} \frac{\|v_i - x_k\|^2}{\left(\sum_{j=1}^{C}\left(\frac{\|v_i - x_k\|}{\|v_j - x_k\|}\right)^{\frac{2}{m-1}}\right)^{m}} \rightarrow \text{min}, \tag{7}$$

where V is a cluster center matrix, X is a data matrix, v_i is a cluster center, m is a weighting exponent which determines the fuzziness of the clusters and x_k is a data point.

Using similar technique, the reformulation of the membership matrix computation in formulas (6) and (1) can provide another new objective functions which employ membership matrix as a single independent variable, as provided in formula (8).

$$J_{FGWC}(U; X) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left\|\frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}} - x_k\right\|^2 \rightarrow \text{min}, \tag{8}$$

where U is a membership matrix, X is a data matrix, u_{ik} is an element of membership matrix, m is a weighting exponent which determines the fuzziness of the clusters, and x_k is a data point.

Based on the result reported in [22, 24], it is necessary to distinguish the treatment for different dataset due to its dimension. For a dataset containing n number of records and d number of variables (dimensions), there are at least two different treatments. If the number of data records is greater than the number of dimension ($n > d$), encode of the objective function using cluster centers is simple. The other case is treated analogously. Encoded by cluster centers is simpler to compute and could better handle data sets that the value of artificial bee should meet the constraint of FGWC objective function [22, 24]. Thus, in the proposed method we distinguish the treatment of these conditions of data. The termination criteria of FGWC-ABC are reaching the maximum iteration or finding the best solution that is equal to the global minimum.

We also illustrate it in the pseudo-code of the FGWC-ABC algorithm below. The detailed flowchart of the proposed method is shown in Fig. 2.

Input: Geo-demographic data, clustering parameters
Output: Final cluster centers

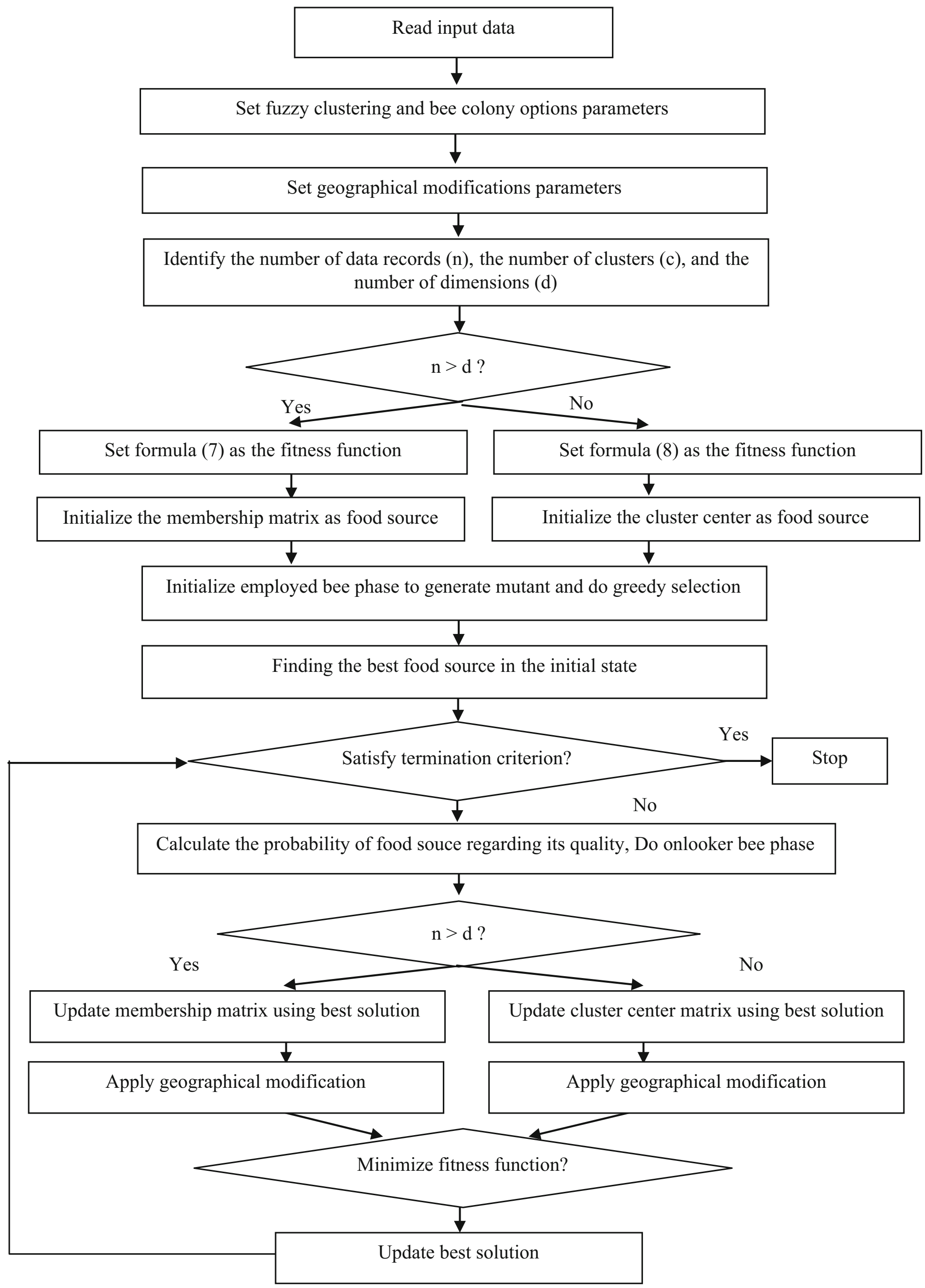

Fig. 2 Flowchart of improved FGWC using ABC algorithm

FGWC-ABC

Step 1: After performing the data reading process, set the number of clusters, threshold $\varepsilon > 0$ and other parameters such as the fuzziness m. Some parameters of ABC such as the number of colony size, the number of food sources, and the number of cycles for foraging (the stopping criteria) are also defined. Some geographical parameters such as α, β, a, b also set up.

Step 2: Identify number of data records (n), number of clusters (c), and number of dimensions (d) to determine the objective function that must be evaluated.

Step 3: Perform an iterative process to check whether the termination criteria, which are reaching the maximum iteration or finding the best solution that is equal to global minimum, are satisfied. If the number of data records is greater than the number of dimension ($n > d$), then use formula (7) as the objective function and initialize the membership matrix as food source, otherwise use formula (8) as the objective function and initialize the cluster center as food source. Those objective functions subject to the set of cluster centers $V = \{v_1, \ldots, v_n\} \subset \Re^d$ and the membership matrix $U \in M_{FGWC}$, where

$$M_{FGWC} = \left\{ U \in [0,1]^{cxn} \left| \sum_{i=1}^{c} u_{ik} = 1, k = 1, \ldots, n, \right. \right.$$
$$\left. \sum_{i=1}^{c} u_{ik} > 0, i = 1, \ldots, c \right\} \tag{9}$$

Thus, the FGWC search space contains the continuous elements of cluster centers V and membership matrix U.

Step 4: Initialize the employed bee phase to generate mutant and do greedy selection, then find best food source in this initial state.

Step 5: Calculate the probability of the food source. A food source is chosen with the probability which is proportional to its quality using the following formula:

$$prob_i = \frac{fitness_i}{\sum_{x=1}^{S} fitness_x} \tag{10}$$

where $prob_i$ is the probability of certain food source, $fitness_i$ is the value of fitness function of the solution i and S is the number of food sources [15]. Then, perform the onlooker bee phase to generate new best solution.

Step 6: Update the membership matrix or the cluster center using best solution. If the number of data records is greater than the number of dimension ($n > d$), use formula (6) to calculate the membership values and compute the objective function in formula (8). Otherwise, use formula (5) to calculate the cluster centers. This process will compute the objective function in formula (7). The distance used herein is the Euclidean function. Perform geographic modifications through formulas (1–3) to involve the neighborhood effect.

Step 7: Check whether the termination criteria, which are reaching the maximum iteration or finding the best solution that is equal to the global minimum, hold. If yes, stop and print the results. Otherwise, back to Step 4.

4 Results and discussions

Firstly, we describe the experimental environments such as,

- *Experimental tools*: The proposed method is implemented in C language on the basis of the source code of ABC implementation by Karaboga & Basturk [15] and executed under the environment of Intel Core i5-3210M CPU @2.50GHz, 4GB RAM and Windows 7, 64bit operating system. It was tested against FCM [3], Neighborhood Effect (NE) [9] and FGWC [20].
- *Parameters setting*: $\varepsilon = 10^{-2}$; $m = 3$; $a = b = 1$; $\alpha = 0.7$; $\beta = 0.3$.
- *Experimental dataset*:
  - A real dataset of socio-economic demographic variables from United Nations Organization (UNO) [36]. These data are constructed from a set of questionnaires dispatched annually by the United Nations Statistics Division to over 230 national statistical offices. It has been collected from national statistical authorities since 1948. The variables contain information on ethnicity and language, household characteristics, housing, population size and composition, marriage and divorce on an annual basis, economic activity, births, deaths, educational attainment, etc.;
  - A real dataset used from public UCI Machine Learning Repository, namely the Wisconsin Breast Cancer Data [12]. This breast cancer databases was obtained from the university of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. It consists of 699 numbers of instances and 10 numbers of attributes plus the class attribute. It collect information about the cancer case such as clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, etc in numerical ordinal value from 1–10. The class attribute has two values, which determine benign or malignant. There are also 16 records that contain a single missing attribute value;
  - A real dataset of demographic and economic variables from Indonesia Population Census 2010 [4] province level. The variables contain 110 characteristics on ethnicity and language, household characteristics, housing, population size and composition, marriage and divorce, economic activity, births, deaths, educational attainment, etc.

- *Cluster validity measurement*: Partition Coefficient (PC), Classification Entropy (CE), Partition Index (SC), Separation Index (S), Xie and Beni's Index (XB), and IFV index. Those measurements are usually used to measure the performance of clustering algorithms such as in [2, 7, 10, 18] and are used as observed variables in this research. According to this research framework, the accuracy of fuzzy geo-demographic clustering problem is aimed to be improved.

  – The PC index measures the amount of overlapping between clusters and for c clusters as in formula (11) where u_{ik} is the membership of data point k$^{\text{th}}$ to cluster i$^{\text{th}}$. From the formula, it can be concluded that the value of PC index range in $[1/c, 1]$. PC is the-higher-the-better.

$$PC = \frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^2. \tag{11}$$

  – The CE index measures the fuzziness of the cluster partition and is defined as in formula (12). CE is the-smaller-the-better.

$$CE = -\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}\log_{\alpha}(u_{ik}). \tag{12}$$

  – The SC index is the ratio of the sum of compactness and separation of the clusters and is defined as in formula (13). SC is the-smaller-the-better.

$$SC = \sum_{i=1}^{c} \frac{\sum_{j=1}^{n} u_{ij}^m \|v_i - x_j\|^2}{N_i \sum_{k=1}^{c} \|v_i - v_k\|^2}. \tag{13}$$

  – The S index uses a minimum-distance separation for partition validity in formula (14). S is the-smaller-the-better.

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2 \left\|v_i - x_j\right\|^2}{N \min_{i,k} \|v_i - v_k\|^2}. \tag{14}$$

  – XB aims to quantify the ratio of the total variation within clusters and the separation of clusters in formula (15). XB is the-smaller-the-better.

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \|v_i - x_j\|^2}{N \min_{i,j} \|v_i - x_j\|^2}. \tag{15}$$

  – IFV is usually used as a validity function of fuzzy clustering for spatial data, because its robustness and stability. When IFV $\rightarrow$ max, the value of IFV is said to yield the most

Table 1 Comparison of PC, CE, and SC Indices for Case 1

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.949	0.854	0.931	0.949	0.090	0.257	0.120	0.091	1.6E-03	3.8E-03	2.7E-02	1.6E-03
3	0.914	0.756	0.866	0.887	0.160	0.456	0.243	0.202	1.5E-04	3.2E-04	1.7E-02	2.2E-04
4	0.884	0.687	0.846	0.868	0.222	0.606	0.292	0.246	4.2E-05	6.9E-05	7.7E-05	4.2E-05
5	0.859	0.658	0.843	0.859	0.273	0.696	0.308	0.272	1.8E-05	2.4E-05	1.9E-05	1.8E-05
6	0.852	0.623	0.793	0.830	0.298	0.800	0.411	0.338	9.4E-06	1.3E-05	1.0E-05	9.2E-06
7	0.803	0.560	0.793	0.811	0.392	0.939	0.418	0.380	5.5E-06	7.8E-06	5.4E-06	5.6E-06
8	0.800	0.546	0.777	0.794	0.408	0.996	0.464	0.423	3.7E-06	4.9E-06	3.6E-06	3.6E-06
9	0.852	0.516	0.854	0.865	0.315	1.096	0.316	0.290	1.8E-06	3.9E-06	1.9E-06	7.0E-07
10	0.807	0.502	0.780	0.797	0.404	1.144	0.473	0.429	1.1E-06	1.6E-06	1.7E-06	8.7E-07

Table 2 Comparison of S, XB, and IFV Indices for Case 1

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	2.3E+04	3.8E+04	5.9E+04	2.2E+04	9.2E+08	2.4E+09	5.9E+10	1.1E+07	7.511	3.014	4.344	6.778
3	7.1E+03	1.2E+04	3.6E+04	9.1E+03	1.1E+08	1.3E+09	8.9E+09	1.0E+09	16.276	7.562	5.976	19.661
4	3.4E+03	5.4E+03	5.9E+03	3.6E+03	3.8E+08	1.4E+09	5.9E+09	1.0E+08	23.975	10.984	33.271	26.718
5	2.1E+03	3.1E+03	2.5E+03	2.1E+03	2.1E+09	0.0E+00	0.0E+00	5.3E+08	35.028	14.333	38.960	34.723
6	1.5E+03	2.2E+03	1.9E+03	1.6E+03	0.0E+00	0.0E+00	0.0E+00	0.0E+00	44.758	16.940	48.328	43.943
7	1.2E+03	1.8E+03	1.3E+03	1.2E+03	0.0E+00	0.0E+00	0.0E+00	0.0E+00	52.885	19.214	54.879	50.325
8	9.4E+02	1.2E+03	1.0E+03	9.9E+02	9.4E+08	1.2E+09	0.0E+00	0.0E+00	58.531	19.396	58.626	61.790
9	5.2E+02	1.2E+03	5.4E+02	2.3E+02	0.0E+00	0.0E+00	0.0E+00	0.0E+00	37.602	21.715	41.274	36.256
10	4.9E+02	7.4E+02	6.1E+02	3.5E+02	0.0E+00	0.0E+00	0.0E+00	0.0E+00	53.415	18.233	64.929	57.785

Table 3 Comparison of PCAES Index,Number of Iterations, and Running Time for Case 1

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	9.536	9.078	6.570	9.681	20	18	15	10	0.078	0.015	0.062	1.029
3	20.081	17.572	9.644	14.450	61	51	38	10	0.39	0.046	0.249	2.044
4	26.156	29.436	15.580	24.262	66	97	38	10	0.483	0.124	0.327	3.931
5	28.918	34.218	24.931	28.974	70	182	67	10	0.67	0.327	0.67	5.21
6	30.045	35.234	23.441	28.558	63	60	108	10	0.702	0.124	1.263	8.174
7	27.830	32.829	26.888	28.300	307	57	38	10	3.978	0.156	0.53	10.935
8	30.812	47.292	26.100	27.269	76	97	38	10	1.294	0.327	0.67	12.838
9	140.331	45.102	50.815	165.982	45	58	30	10	0.748	0.219	0.608	15.303
10	45.786	43.448	42.325	150.167	53	47	139	10	0.982	0.218	2.667	21.185

Table 4 Comparison of PC, CE, and SC Indices for Case 2

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.845	0.690	0.834	0.839	0.259	0.481	0.278	0.269	1.1E-02	2.1E-02	2.2E-02	1.2E-02
3	0.732	0.519	0.704	0.729	0.483	0.834	0.529	0.492	1.7E-03	4.8E-03	2.1E-03	1.9E-03
4	0.561	0.366	0.551	0.598	0.785	1.148	0.811	0.755	5.5E-04	1.2E-03	7.0E-04	6.6E-04
5	0.526	0.322	0.507	0.524	0.899	1.329	0.945	0.924	1.9E-04	5.1E-04	2.4E-04	2.1E-04
6	0.504	0.250	0.481	0.482	0.990	1.545	1.039	1.043	7.6E-05	2.4E-04	8.7E-05	8.8E-05
7	0.484	0.229	0.401	0.464	1.075	1.668	1.232	1.124	4.1E-05	1.3E-04	4.8E-05	4.3E-05
8	0.399	0.214	0.387	0.446	1.262	1.777	1.301	1.204	2.2E-05	8.0E-05	2.5E-05	2.5E-05
9	0.389	0.182	0.344	0.379	1.323	1.915	1.445	1.357	1.4E-05	4.7E-05	1.6E-05	1.5E-05
10	0.347	0.189	0.329	0.370	1.460	1.973	1.508	1.414	9.8E-06	4.1E-05	1.0E-05	1.0E-05

Table 5 Comparison of S, XB, and IFV Indices for Case 2

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.0E+03	5.470	2.040	3.672	4.441
3	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	8.115	2.029	8.515	7.013
4	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	10.625	2.374	10.920	9.349
5	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	10.973	2.059	11.527	9.948
6	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	12.544	1.960	10.916	12.530
7	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	12.387	1.694	12.416	12.546
8	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	11.932	1.498	12.279	12.105
9	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	11.464	1.420	11.474	11.676
10	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	10.891	1.211	10.798	11.234

Table 6 Comparison of PCAES Index,Number of Iterations, and Running Time for Case 2

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	1.978	2.119	1.812	2.095	7	13	7	5	0.078	0.031	0.093	2.169
3	6.734	4.669	6.257	6.907	13	57	12	5	0.234	0.187	0.249	4.056
4	5.406	4.743	5.882	8.917	12	48	10	5	0.28	0.218	0.249	7.753
5	11.624	7.878	13.356	14.954	59	37	13	5	1.716	0.218	0.452	11.827
6	23.475	7.443	23.417	18.563	65	79	46	5	2.574	0.577	1.684	15.975
7	35.150	10.665	18.525	28.268	57	129	30	5	2.277	1.138	1.326	32.994
8	30.122	14.204	26.200	37.460	37	47	66	5	1.747	0.499	3.182	50.965
9	36.640	13.968	23.990	32.460	69	75	49	5	3.588	0.951	2.62	37.815
10	33.417	21.414	28.422	37.865	166	67	53	5	9.594	0.982	3.588	49.92

Table 7 Comparison of PC, CE, and SC Indices for Case 3

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.603	0.500	0.577	0.577	0.583	0.693	0.612	0.612	2.2E-01	0.0E+00	6.3E-01	6.2E-01
3	0.536	0.333	0.519	0.519	0.790	1.099	0.810	0.810	2.5E-03	0.0E+00	2.5E-03	2.5E-03
4	0.433	0.250	0.392	0.421	1.036	1.386	1.129	1.053	6.6E-04	1.9E+06	4.2E-03	6.6E-04
5	0.420	0.200	0.350	0.386	1.158	1.609	1.282	1.218	2.8E-04	2.0E+05	9.5E-04	7.6E-04
6	0.394	0.167	0.349	0.363	1.281	1.792	1.365	1.321	2.2E-04	6.4E+04	2.0E-04	1.2E-04
7	0.380	0.143	0.334	0.363	1.369	1.946	1.488	1.394	7.7E-05	2.3E+04	4.7E-04	7.2E-05
8	0.386	0.125	0.347	0.357	1.405	2.079	1.508	1.471	3.4E-05	7.1E+03	7.1E-05	4.4E-05
9	0.397	0.111	0.369	0.387	1.444	2.197	1.508	1.447	3.3E-05	3.7E+03	3.4E-05	2.3E-05
10	0.420	0.100	0.383	0.393	1.424	2.303	1.534	1.489	1.3E-05	1.8E+03	3.5E-05	1.7E-05

Table 8 Comparison of S, XB, and IFV Indices for Case 3

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	1.4E+01	0.0E+00	1.5E+01	1.5E+01	2.4E+01	1.5E+01	1.5E+01	1.5E+01	0.503	0.000	0.191	0.191
3	6.0E+00	0.0E+00	5.8E+00	5.8E+00	6.4E+00	6.7E+00	5.8E+00	5.8E+00	5.994	0.000	5.710	5.710
4	3.4E+00	0.0E+00	3.7E+00	3.3E+00	8.2E+01	3.8E+00	6.1E+01	8.9E+01	9.852	0.000	3.311	9.725
5	2.1E+00	0.0E+00	0.0E+00	2.6E+00	1.0E+03	2.4E+00	3.1E+05	2.9E+05	12.734	0.000	4.475	5.006
6	1.6E+00	0.0E+00	1.5E+00	0.0E+00	9.5E+02	1.7E+00	6.5E+02	1.3E+04	13.293	0.000	10.755	21.692
7	1.3E+00	0.0E+00	1.4E+00	0.0E+00	8.0E+04	1.2E+00	8.3E+03	6.9E+04	25.904	0.000	4.426	25.298
8	8.8E-01	0.0E+00	9.2E-01	8.4E-01	0.0E+00	9.4E-01	2.3E+05	8.4E+05	18.782	0.000	18.158	27.815
9	7.1E-01	0.0E+00	6.7E-01	6.1E-01	0.0E+00	7.4E-01	0.0E+00	0.0E+00	36.397	0.000	34.385	31.307
10	4.9E-01	0.0E+00	5.9E-01	4.9E-01	0.0E+00	6.0E-01	0.0E+00	0.0E+00	42.159	0.000	24.694	35.596

Table 9 Comparison of PCAES Index,Number of Iterations, and Running Time for Case 3

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.029	0.000	0.283	0.283	21	36	17	10	0.043	0.028	0.035	0.587
3	6.406	0.000	6.021	6.021	71	35	158	10	0.195	0.042	0.47	1.364
4	6.159	0.001	1.917	5.951	26	22	48	10	0.095	0.035	0.202	1.494
5	6.215	0.001	4.962	5.995	78	25	31	10	0.361	0.048	0.145	2.25
6	4.577	0.002	3.066	3.717	20	32	46	10	0.11	0.076	0.27	3.079
7	3.607	0.003	2.995	3.184	21	30	64	10	0.136	0.085	0.434	3.965
8	4.262	0.005	2.893	3.168	34	32	92	10	0.252	0.1	0.739	4.867
9	2.867	0.006	9.919	3.255	12	28	15	10	0.109	0.093	0.124	6.286
10	3.383	0.007	15.430	9.108	13	34	15	10	0.118	0.14	0.146	7.421

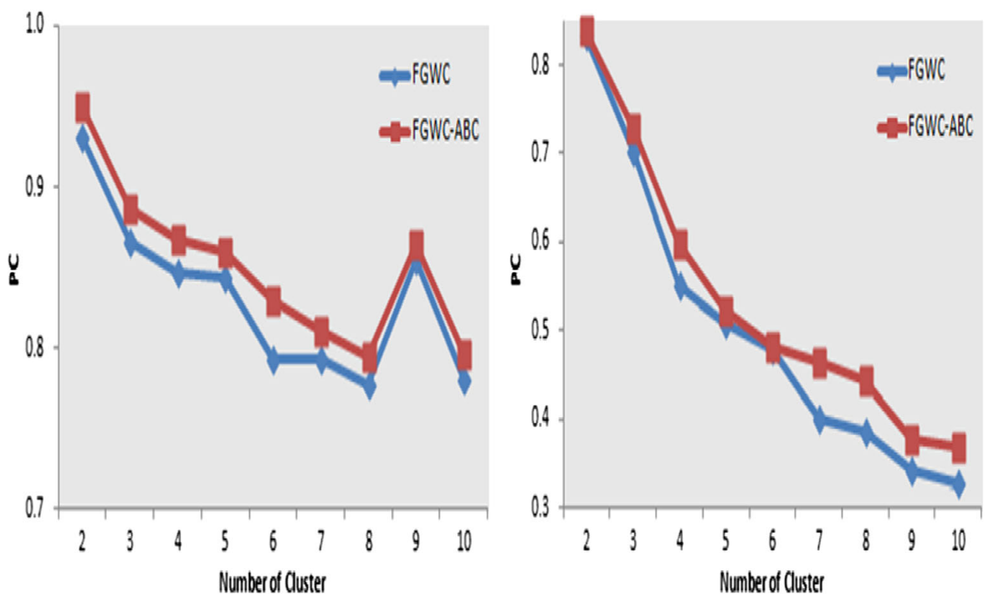

Fig. 3 PC index of FGWC and FGWC-ABC in Case 1 (*left*) and Case 2 (*right*)

optimal of the dataset. It is defined as in formulas (16–18).

$$IFV = (1/C)\sum_{j=1}^{C}\left\{(1/N)\sum_{k=1}^{N}u_{kj}^{2}\left[\log_2 C - (1/N)\sum_{k=1}^{N}\log_2 u_{kj}\right]^2\right\} \times (SD_{\max}/\overline{\sigma_D}), \quad (16)$$

$$SD_{\max} = \max_{k\neq j}\|V_k - V_j\|^2, \quad (17)$$

$$\overline{\sigma_D} = (1/C)\sum_{j=1}^{C}\left((1/N)\sum_{k=1}^{N}\|X_k - V_j\|^2\right). \quad (18)$$

- *Experimental objective*: to evaluate the clustering qualities and the computational time of all algorithms.

The experimental results under these environments are illustrated from Tables 1, 2, 3, 4, 5, 6, 7, 8 to 9. From these results, it can be concluded that the proposed FGWC-ABC is successfully improved the original FGWC in terms of validity index measurement for most cases. For some cases, FGWC-ABC is not better than the other method such as FCM and NE, but the difference is not so far. As the main objective of this work is to improve the quality of standard FGWC that is previously known as better than classical FCM in terms of incorporating the geographical feature and the neighborhood effect. Thus, the performance is still acceptable.

Figure 3 gives an example of the comparison of PC index evaluation of FGWC and the proposed method FGWC-ABC. The greater value of the PC index denotes the better quality of resulting cluster. Using UNSD Socio-economic 2011 data in case 1 simulation, the PC validity index measurement of FGWC-ABC is always higher than the standard FGWC, which indicated superiority of FGWC-ABC clustering quality against the standard FGWC. The experimental result of the Breast Cancer dataset in case 2 also reflects the similar condition since FGWC-ABC provides a better PC index than FGWC in various number of clusters.

Case 1. UNSD Socio-economic 2011 Data set

Case 2. The Breast Cancer Data set

Case 3. The Indonesia Population Census 2010 Data set

The evaluation of standard deviation under the environment stated in the parameter settings are illustrated from Tables 10, 11, 12, 13, 14, 15, 16, 17 to 18. Using 10

Table 10 Comparison of Standard Deviation of PC, CE, and SC Indices for Case 1

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	2.34.E-16	8.23.E-07	9.67.E-03	5.92.E-03	1.46.E-17	8.43.E-07	1.53.E-02	9.35.E-03	9.80.E-09	3.75.E-07	1.30.E-02	7.96.E-03
3	4.83.E-07	2.71.E-05	1.18.E-02	6.99.E-07	7.38.E-07	3.39.E-05	1.95.E-02	8.23.E-07	1.20.E-09	9.57.E-08	5.26.E-03	3.39.E-09
4	9.19.E-07	6.20.E-06	1.39.E-02	7.79.E-03	1.26.E-06	5.54.E-06	2.66.E-02	9.71.E-03	1.47.E-10	1.27.E-07	1.77.E-05	2.53.E-07
5	1.17.E-16	1.51.E-06	7.35.E-03	6.74.E-03	5.16.E-07	3.01.E-06	1.57.E-02	1.49.E-02	3.86.E-11	2.46.E-11	6.11.E-07	5.79.E-07
6	9.61.E-06	1.06.E-05	2.52.E-02	1.27.E-02	1.46.E-05	1.27.E-05	4.83.E-02	2.26.E-02	8.97.E-11	3.69.E-10	4.56.E-07	1.67.E-07
7	1.10.E-02	8.62.E-03	1.15.E-02	1.27.E-02	1.84.E-02	9.84.E-03	2.93.E-02	3.09.E-02	1.32.E-07	1.99.E-07	4.26.E-07	3.29.E-07
8	1.31.E-02	9.19.E-03	9.83.E-03	4.72.E-03	2.56.E-02	2.37.E-02	2.31.E-02	1.07.E-02	6.69.E-07	9.13.E-07	6.77.E-08	8.89.E-07
9	8.39.E-04	7.42.E-03	8.74.E-03	4.31.E-03	2.02.E-03	2.20.E-02	2.04.E-02	1.04.E-02	1.82.E-07	8.35.E-07	6.71.E-07	5.74.E-07
10	1.13.E-02	1.20.E-02	9.47.E-03	6.40.E-03	2.26.E-02	3.16.E-02	2.32.E-02	1.43.E-02	4.01.E-07	2.11.E-06	3.58.E-07	4.54.E-07

Table 11 Comparison of Standard Deviation of S, XB, and IFV Indices for Case 1

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	8.43.E-03	1.49.E+00	1.88.E+04	1.15.E+04	3.56.E+02	9.33.E+04	3.04.E+10	1.86.E+10	5.61.E-06	2.08.E-05	1.26.E+00	7.70.E-01
3	2.90.E-02	2.82.E+00	8.51.E+03	8.88.E-03	4.61.E+02	3.14.E+05	2.49.E+09	9.43.E+02	2.76.E-05	9.69.E-04	4.44.E+00	6.07.E-05
4	9.92.E-03	1.97.E+00	1.20.E+03	1.16.E+02	1.12.E+03	4.94.E+05	2.88.E+09	1.36.E+09	1.25.E-04	2.00.E-03	4.18.E+00	2.26.E+00
5	5.70.E-03	1.47.E-02	1.73.E+02	1.47.E+02	5.70.E+03	0.00.E+00	2.57.E+08	2.25.E+08	2.53.E-04	7.40.E-05	3.05.E+00	1.79.E+00
6	3.52.E-02	1.30.E-01	1.89.E+02	4.11.E+01	0.00.E+00	0.00.E+00	6.37.E+08	7.80.E+08	3.65.E-03	8.78.E-04	2.35.E+00	2.08.E-01
7	2.93.E+01	7.21.E+01	1.38.E+02	1.27.E+02	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	1.58.E+00	7.04.E-01	1.81.E+00	2.29.E+00
8	1.62.E+02	1.86.E+02	4.14.E+01	2.03.E+02	4.53.E+08	5.37.E+08	0.00.E+00	2.64.E+08	4.58.E+00	2.09.E+00	2.91.E+00	5.52.E+00
9	6.13.E+01	1.54.E+02	1.36.E+02	1.53.E+02	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	2.97.E+00	1.96.E+00	4.79.E+00	3.00.E+00
10	1.16.E+02	2.81.E+02	9.70.E+01	1.12.E+02	2.34.E+08	0.00.E+00	0.00.E+00	0.00.E+00	8.04.E+00	2.08.E-01	5.17.E+00	5.52.E+00

number of runs for each number of clusters and for each of the three public datasets, we measured the standard deviation of the data as well as the average of data that are presented in the previous Tables 1–9. From the evaluation, it can be concluded that the proposed FGWC-ABC not only successfully improved the original FGWC in terms of validity index measurement for most cases, but also provided the optimum range of results. For some cases, FGWC-ABC is not better than the other methods such as FCM and NE, but the differences are not far. As the main objective of this work is to improve the quality of standard FGWC that is previously known as better than classical FCM in terms of incorporating the geographical feature and the neighborhood effect. Because of the average result of FGWC-ABC in previous Table 1–9 is optimum, thus the range of performance results is still acceptable.

STANDARD DEVIATION MEASUREMENT
Case 1. UNSD Socio-economic 2011 Data set
Case 2. The Breast Cancer Data set
Case 3. The Indonesia Population Census 2010 Data set

5 An application of FGWC-ABC for analyzing crime behavior in population

- *The problem and the dataset*:

This section presents a decision-making application based on FGWC-ABC for analyzing crime behavior problem in population of the city using the communities and crime dataset, which consists of socio-economic variables from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR [1]. The dataset consists of many parameters involve the community, such as the percent of the population considered urban, and the median family income, and involving law enforcement, such as per capita number of police officers, and percent of officers assigned to drug units. All numeric data was normalized into the decimal range [0.00, 1.00] using an unsupervised, equal-interval binning method. Attributes retain their distribution and skew; hence for example the population attribute has a mean value of 0.06 because most communities are small. The attributes which have missing values are then being ignored. The government typically has many programs in controlling and reducing crime rates. Many stakeholders also have similar interest in crime control and analysis, e.g. police agency, investor, multinational enterprise, etc. In this section we define and analyze the crime rate in the certain population using demographic and socio-economic variable analysis. The per capita violent crimes variable, which is defined for crime rate, was usually calculated using

Table 12 Comparison of Standard Deviation of PCAES Index,Number of Iterations, and Running Time for Case 1

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	1.59.E-05	2.65.E-04	1.61.E+00	9.84.E-01	1.0801	0.7888	1.9465	0.0000	0.0059	0.0006	0.0116	0.0525
3	1.44.E-04	4.24.E-03	2.57.E+00	1.40.E-04	2.3781	4.9486	8.4123	0.0000	0.0157	0.0058	0.0572	0.1806
4	1.68.E-04	1.64.E-02	4.72.E+00	9.37.E-01	2.4060	15.2840	10.8853	0.0000	0.0286	0.0210	0.0906	0.2082
5	1.22.E-04	1.52.E-04	1.97.E+00	1.70.E+00	4.6774	5.8271	12.8776	0.0000	0.0412	0.0140	0.1219	0.3540
6	1.55.E-03	1.18.E-03	3.00.E+00	8.67.E-01	9.6517	11.8771	19.3509	0.0000	0.1086	0.0288	0.2330	0.4048
7	9.49.E-01	1.79.E+00	2.14.E+00	1.97.E+00	89.5845	23.3866	7.6920	0.0000	1.1314	0.0636	0.0962	0.5219
8	3.78.E+01	6.77.E+00	2.28.E+00	4.94.E+01	45.2209	10.7373	14.4380	0.0000	0.6542	0.0364	0.2563	1.3783
9	3.39.E+00	8.25.E-01	9.17.E+00	3.71.E+01	9.7434	14.9179	2.7183	0.0000	0.1727	0.0603	0.0591	1.2376
10	5.22.E+01	5.76.E+00	4.44.E+01	5.34.E+01	13.1217	58.5401	58.3910	0.0000	0.2533	0.2605	1.2649	2.6209

population and the sum of crime variables considered violent crimes in the United States: murder, rape, robbery, and assault. Numerous variables in the dataset were included so that FGWC-ABC can select or learn weights for attributes and then be further evaluated. **The aim of research is to perform the identification of crime cases in United States**.

- *A procedure to determine fuzzy rules*:

There are some alternatives to define fuzzy rules from a certain data set [8]. Berenji & Khedkar proposed the learning and tuning fuzzy logic controllers through reinforcements, which also provide a simple procedure of decision selection processes [40]. As presented in Fig. 4, antecedent attributes are determined from data input to create fuzzy rules. Some rules can be directly used for the decision or action making. The other rules can be selected to determine the consequent attributes [40].

Suppose that we have the following rules:

Rule 1: IF A is X_1 and B is Y_1 THEN C is Z_1

Rule 2: IF A is X_2 and B is Y_2 THEN C is Z_2

Rule 3: IF A is X_3 and B is Y_3 THEN C is Z_3

Each rule has antecedent attributes which describe some preconditions, in those examples are X (X_1, X_2, and X_3) and Y (Y_1, Y_2, and Y_3). It also has a consequent attribute that provides the decisions as output [40].

Table 13 Comparison of Standard Deviation of PC, CE, and SC Indices for Case 2

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	5.52.E-05	1.50.E-05	3.12.E-03	3.02.E-03	1.19.E-04	1.88.E-05	5.10.E-03	4.94.E-03	5.53.E-05	1.52.E-06	5.28.E-03	5.23.E-03
3	5.26.E-02	5.27.E-07	1.36.E-02	1.07.E-02	7.32.E-02	3.16.E-07	2.02.E-02	1.60.E-02	4.07.E-04	5.42.E-07	1.28.E-04	9.73.E-05
4	5.49.E-02	2.73.E-05	2.03.E-02	1.53.E-02	6.84.E-02	3.42.E-05	2.44.E-02	1.84.E-02	9.35.E-06	1.60.E-06	2.56.E-05	2.04.E-05
5	2.52.E-04	4.90.E-05	8.53.E-03	6.95.E-03	5.11.E-04	6.26.E-05	1.05.E-02	8.46.E-03	4.79.E-06	4.28.E-07	1.05.E-05	1.15.E-05
6	1.63.E-03	1.80.E-02	2.83.E-02	1.87.E-03	3.13.E-03	2.73.E-02	4.90.E-02	4.13.E-03	2.72.E-06	1.98.E-05	1.20.E-05	3.43.E-06
7	2.65.E-02	9.65.E-03	3.04.E-02	6.40.E-03	4.43.E-02	3.88.E-02	4.94.E-02	2.84.E-02	1.70.E-05	3.85.E-05	2.46.E-05	1.40.E-05
8	1.09.E-04	9.27.E-03	3.34.E-02	2.75.E-02	3.09.E-04	2.07.E-02	6.12.E-02	4.53.E-02	9.95.E-08	1.04.E-06	2.05.E-06	1.23.E-07
9	1.27.E-02	6.01.E-03	3.59.E-02	2.48.E-02	3.33.E-02	1.15.E-02	6.91.E-02	3.87.E-02	3.05.E-07	2.54.E-06	7.88.E-07	6.64.E-07
10	1.16.E-02	6.50.E-03	1.47.E-02	2.47.E-02	2.45.E-02	1.90.E-02	3.52.E-02	4.32.E-02	3.59.E-07	3.60.E-08	1.12.E-07	6.14.E-07

Table 14 Comparison of Standard Deviation of S, XB, and IFV Indices for Case 2

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	5.27.E+02	5.27.E+02	1.33.E-02	2.32.E-04	4.05.E-01	3.95.E-01
3	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	4.51.E-02	2.47.E-03	7.66.E-01	6.19.E-01
4	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	7.48.E-01	6.41.E-03	6.47.E-01	4.91.E-01
5	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	2.34.E-01	1.38.E-02	1.10.E+00	8.85.E-01
6	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	1.62.E-01	7.85.E-02	6.80.E-01	7.01.E-01
7	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	4.40.E-01	8.63.E-02	5.08.E-01	7.12.E-01
8	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	2.94.E-02	3.63.E-02	7.72.E-02	7.25.E-02
9	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	6.16.E-02	1.80.E-02	1.06.E-01	8.89.E-02
10	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	0.00.E+00	3.46.E-01	3.45.E-02	1.16.E-01	7.85.E-02

In this case, we evaluate the prediction of crime rate measured in total number of violent crimes per 100K population variable using socio-economic and demographic variables from the pre-defined dataset. We compute the rule weight by the following characteristics of fuzzy rules [8, 40] :

- Associativity: F(x, F(y,z)) = F(F(x,y),z).
- Commutativity: F(x,y) = F(y,x).
- Monotonicity: F(x,y) <F(z,w) if x<z and y<w.
- Identity: F(x,1) = x.

As presented in Fig. 5, we then create a procedure to determine fuzzy rules in this case, as follows:

1. Select a number of antecedents attributes for the rules. The attributes are selected manually regarding its importance for specific task in determining class attribute. The selection of antecedent attributes require the predefined knowledge about the objective of the decision making process. Thus, expert judgement is required to clarify the pertinence of the selection.
2. Perform fuzzy geo-demographic clustering using FGWC-ABC on the dataset and obtain the corresponding fuzzy membership matrix.
3. Perform other fuzzy geo-demographic clustering using FGWC-ABC to the selected antecedent attributes and obtain the corresponding fuzzy membership matrices.
4. Perform other fuzzy geo-demographic clustering using FGWC-ABC to the consequent attribute.
5. Utilize the obtained membership matrices to evaluate the rule weight. If the rule weight is equal or greater than the given threshold, the rule can be considered as the valid rule.

- *Initial experimental results*:

  In this part, we describe the initial simulation of the procedure using the communities and crime dataset. All running processes are performed using the FGWC-ABC parameters such as $\varepsilon = 10^{-3}$; $m = 3$; $a = b = 1$; $\alpha = 0.7$; $\beta = 0.3$. Six antecedent attributes are chosen for the rules as follows.

  - PCI: per capita income;
  - Poverty: percentage of people under the poverty level;
  - Unemployed: percentage of people 16 and over, in the labor force, and unemployed;
  - Divorce: percentage of population who are divorced;
  - Homeless: number of homeless people counted in the street;

Table 15 Comparison of Standard Deviation of PCAES Index,Number of Iterations, and Running Time for Case 2

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	3.20.E-03	2.30.E-04	1.45.E-01	1.46.E-01	0.6992	0.9487	0.9487	0.0000	0.0112	0.0066	0.0165	0.2076
3	2.33.E+00	4.27.E-04	3.04.E-01	2.53.E-01	2.2010	7.5873	1.3166	0.0000	0.0422	0.0261	0.0395	0.0839
4	4.28.E+00	1.85.E-03	1.29.E+00	1.01.E+00	14.9224	9.5621	2.2608	0.0000	0.3483	0.0464	0.0636	0.5893
5	5.01.E-01	6.24.E-03	6.87.E-01	7.47.E-01	5.1381	2.5144	5.1478	0.0000	0.1713	0.0180	0.1668	1.0027
6	1.86.E+00	1.79.E+00	2.62.E+00	3.80.E+00	15.4546	5.4732	5.9217	0.0000	0.5304	0.0485	0.2226	2.2389
7	6.16.E+00	1.47.E+00	5.42.E+00	2.72.E+00	11.0960	9.8319	7.3341	0.0000	0.5317	0.0830	0.3846	4.3484
8	3.84.E-01	1.63.E+00	6.85.E+00	6.15.E+00	12.4316	9.0486	6.1183	0.0000	0.7220	0.0967	0.2927	1.3292
9	4.00.E+00	1.23.E+00	7.74.E+00	5.73.E+00	9.2039	11.9238	5.4416	0.0000	0.5055	0.1503	0.3153	2.5741
10	3.07.E+00	1.87.E+00	3.83.E+00	6.18.E+00	12.6474	12.1856	3.6148	0.0000	0.7526	0.1816	0.2189	2.4055

- WorkingMom: percentage of moms of kids 6 and under in labor force.

Among other available attributes, those antecedent attributes were selected regarding some literatures that proposed appropriate factors influencing crime rate, such as Snook et al [42], and Levit [43]. We classify each antecendent attribute into 5 clusters: “Very High”, “High”, “Medium”, “Low”, and “Very Low” respectively. The class attribute (violent crime per population) is also classified into 5 clusters as above.

To amplify the utilization of the proposed method in generating fuzzy rules, we provide 2 examples of calculation case in different states. These examples are performed following the procedure that previously stated. Using the same procedure, we can duplicate to generate fuzzy set rules to the other states. Here are the example of rule calculation:

1. *Case of District of Columbia (DC)*:
   - From the fuzzy membership matrix of Per Capita Income we have vector [0.040731 0.020157 0.108814 0.049846 0.780453]. The membership degree 0.780453 is chosen to proposition Per Capita Income of DC is High;
   - The vector of membership of Poverty is [0.019558 0.038629 0.012589 0.013019 0.916205], while the

Table 16 Comparison of Standard Deviation of PC, CE, and SC Indices for Case 3

C	PC Index				CE Index				SC Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	2.81.E-06	2.02.E-04	4.94.E-05	4.94.E-05	4.00.E-06	2.02.E-04	5.54.E-05	5.54.E-05	3.55.E-05	6.53.E+04	1.16.E-03	1.16.E-03
3	1.42.E-06	1.26.E-04	2.73.E-06	2.63.E-06	7.85.E-06	1.88.E-04	2.80.E-06	2.67.E-06	9.22.E-08	1.22.E+04	1.59.E-07	1.44.E-07
4	6.95.E-03	9.77.E-05	1.34.E-02	1.46.E-02	8.48.E-03	1.95.E-04	3.36.E-02	3.69.E-02	1.99.E-05	2.63.E+02	1.58.E-03	1.73.E-03
5	2.28.E-02	9.88.E-05	1.93.E-02	8.88.E-03	5.33.E-02	2.47.E-04	4.44.E-02	2.43.E-02	6.73.E-04	1.20.E+02	1.02.E-03	4.81.E-04
6	8.37.E-03	5.51.E-05	1.03.E-02	7.52.E-03	1.76.E-02	1.65.E-04	3.07.E-02	2.48.E-02	3.47.E-05	4.47.E+00	3.63.E-04	3.45.E-04
7	1.07.E-02	9.39.E-05	7.88.E-03	6.70.E-03	2.88.E-02	3.29.E-04	2.12.E-02	1.81.E-02	1.26.E-05	3.76.E+01	3.79.E-05	3.81.E-05
8	9.55.E-03	9.68.E-05	6.83.E-03	7.21.E-03	2.77.E-02	3.88.E-04	2.44.E-02	2.36.E-02	7.55.E-06	5.91.E+00	7.26.E-05	2.97.E-05
9	4.20.E-03	7.96.E-05	6.44.E-03	2.38.E-03	1.40.E-02	3.58.E-04	2.46.E-02	1.01.E-02	3.53.E-06	2.15.E+00	4.93.E-06	3.30.E-06
10	6.36.E-03	6.05.E-05	6.27.E-03	2.96.E-03	1.99.E-02	3.02.E-04	2.55.E-02	1.21.E-02	2.50.E-06	1.67.E+00	7.34.E-06	7.39.E-06

Table 17 Comparison of Standard Deviation of S, XB, and IFV Indices for Case 3

C	S Index				XB Index				IFV Index			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	6.78.E-04	0.00.E+00	9.12.E-04	9.12.E-04	1.12.E-03	2.88.E-04	9.12.E-04	9.12.E-04	5.39.E-05	1.42.E-05	3.33.E-04	3.33.E-04
3	3.60.E-04	0.00.E+00	2.97.E-04	2.99.E-04	3.84.E-04	1.32.E-04	2.97.E-04	2.99.E-04	3.42.E-04	4.49.E-05	3.44.E-04	3.33.E-04
4	2.06.E-02	0.00.E+00	2.15.E-01	2.43.E-01	1.47.E+01	1.07.E-04	1.17.E+02	1.17.E+02	1.68.E-01	1.16.E-04	2.70.E+00	3.03.E+00
5	1.83.E-01	0.00.E+00	8.13.E-01	1.99.E-01	9.48.E+04	6.87.E-05	1.26.E+05	1.49.E+05	6.70.E+00	2.21.E-04	3.41.E+00	3.34.E+00
6	3.70.E-02	0.00.E+00	1.56.E-01	1.66.E-01	2.94.E+03	6.63.E-05	4.42.E+03	4.42.E+03	1.43.E+00	3.12.E-04	5.58.E+00	5.52.E+00
7	4.14.E-02	0.00.E+00	7.53.E-02	7.83.E-02	5.05.E+04	4.70.E-05	1.21.E+05	1.27.E+05	5.32.E+00	2.85.E-04	6.02.E+00	5.52.E+00
8	3.71.E-02	0.00.E+00	1.04.E-01	1.16.E-01	3.54.E+05	5.62.E-05	4.64.E+05	4.68.E+05	5.66.E+00	5.16.E-04	9.18.E+00	7.97.E+00
9	3.71.E-02	0.00.E+00	4.23.E-02	4.01.E-02	3.39.E+05	3.41.E-05	0.00.E+00	0.00.E+00	2.46.E+00	6.99.E-04	1.21.E+00	6.85.E-01
10	3.53.E-02	0.00.E+00	6.32.E-02	6.12.E-02	0.00.E+00	4.06.E-05	0.00.E+00	0.00.E+00	3.07.E+00	3.12.E-04	7.76.E+00	6.86.E+00

value 0.916205 is chosen to proposition Poverty of DC is High;

- The attribute of Unemployed gives vector membership [0.000398 0.000778 0.000213 0.000107 0.998505], which leads the value 0.998505 to be chosen to proposition Unemployed of DC is High;
- Membership matrix of Divorce for District of Columbia is [0.002571 0.016808 0.048019 0.926014 0.006589]. The maximum value 0.926014 is then chosen to proposition Divorce of DC is Very High;
- Similar to previous attributes, we then obtain vector membership of Homeless, which is [0.174139 0.034012 0.233458 0.378699 0.179691]. It is clear that from value 0.378699 we can propose Homeless of DC is Medium;
- The last antecedent attribute, Working Mom vector membership for DC is [0.213729 0.027792 0.008482 0.704056 0.045941]. This leads to propose Working Mom of DC is High from value 0.704056;
- On the other hand, after classify the class attribute of data set, we get vector membership [0.012723 0.947883 0.003291 0.031064 0.00504], which leads the value 0.947883 to be chosen to proposition Violent Crime Rate of DC is Very High.

From those steps, we get the following fuzzy rule:

Rule 1: IF Per Capita Income is High AND Poverty is High AND Unemployed is High AND Divorce is Very High AND Homeless is Medium AND Working Mom is High THEN Violent Crime Rate is Very High.

1. *Case of State of Delaware*:

   - From the fuzzy membership matrix of Per Capita Income we have vector [0.007084 0.00032 0.97017 0.011833 0.010593]. The membership degree 0.97017 is chosen to proposition Per Capita Income of DC is Medium;
   - The vector of membership of Poverty is [0.102203 0.416416 0.049904 0.012208 0.419268], while the value 0.419268 is chosen to proposition Poverty of DC is High;
   - The attribute of Unemployed gives vector membership [0.427019 0.068328 0.49181 0.001583 0.01126], which leads the value 0.49181 to be chosen to proposition Unemployed of DC is Very Low;
   - Membership matrix of Divorce for District of Columbia is [0.000842 0.981126 0.0094 0.002328 0.006304]. The maximum value 0.981126 is then

Table 18 Comparison of Standard Deviation of PCAES Index,Number of Iterations, and Running Time for Case 3

C	PCAES Index				Number of Iterations				Running Time (s)			
	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC	FCM	NE	FGWC	FGWC-ABC
2	2.95.E-04	8.75.E-04	5.93.E-04	5.93.E-04	3.1552	4.3512	1.7127	0.0000	0.0056	0.0072	0.0066	0.0106
3	8.94.E-04	2.55.E-03	6.56.E-04	6.36.E-04	12.7122	2.5033	20.8062	0.0000	0.0344	0.0028	0.0585	0.0791
4	2.93.E-02	5.83.E-03	1.85.E+00	2.00.E+00	8.7057	3.0984	5.2377	0.0000	0.0288	0.0072	0.0197	0.0312
5	1.84.E+00	4.59.E-03	1.83.E+00	1.67.E+00	18.0604	1.4181	14.9458	0.0000	0.0801	0.0077	0.0676	0.0456
6	4.39.E-01	5.90.E-03	9.92.E-01	7.95.E-01	7.9868	1.6633	9.4287	0.0000	0.0442	0.0089	0.0546	0.0710
7	8.14.E-01	1.63.E-02	3.14.E-01	2.69.E-01	10.4222	0.9487	7.6920	0.0000	0.0697	0.0067	0.0513	0.0957
8	5.84.E-01	1.88.E-02	1.52.E+00	1.59.E+00	11.5667	1.0328	12.8794	0.0000	0.2464	0.0190	0.1011	0.4355
9	3.65.E-01	2.26.E-02	1.58.E+00	3.89.E-01	3.0840	0.9661	1.6865	0.0000	0.0244	0.0064	0.0130	0.1344
10	2.71.E-01	1.74.E-02	2.60.E+00	9.38.E-01	6.5354	0.5270	9.1948	0.0000	0.0619	0.0080	0.0863	0.0687

chosen to proposition Divorce of DC is Very Medium;

- Similar to previous attributes, we then obtain vector membership of Homeless, which is [0.000173 0.000064 0.075503 0.008516 0.915743]. It is clear that from value 0.915743 we can propose Homeless of DC is Very Low;
- The last antecedent attribute, Working Mom vector membership for DC is [0.091994 0.018468 0.005325 0.852533 0.03168]. This leads to propose Working Mom of DC is High from value 0.852533;
- On the other hand, after classify the class attribute of data set, we get vector membership [0.332505 0.023732 0.083359 0.074192 0.486213], which leads the value 0.486213 to be chosen to proposition Violent Crime Rate of DC is Low.

From those steps, we get another fuzzy rule:

Rule 2: IF Per Capita Income is Medium AND Poverty is High AND Unemployed is Very Low AND Divorce is Medium AND Homeless is Very Low AND Working Mom is High THEN Violent Crime Rate is Low.

3. From two case above, Table 19 clearly presents these samples of fuzzy rules for determining the violent crime rate.

We have also implement a web-based GIS application mapping to present the distribution of the violent crime rate. From this figure, decision makers could observe some states

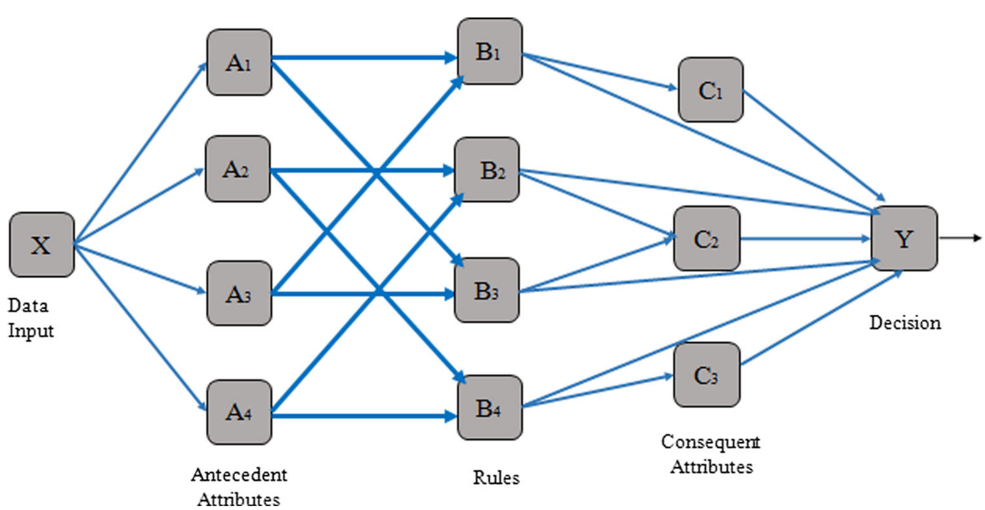

Fig. 4 Simple diagram of decision selection process [44]

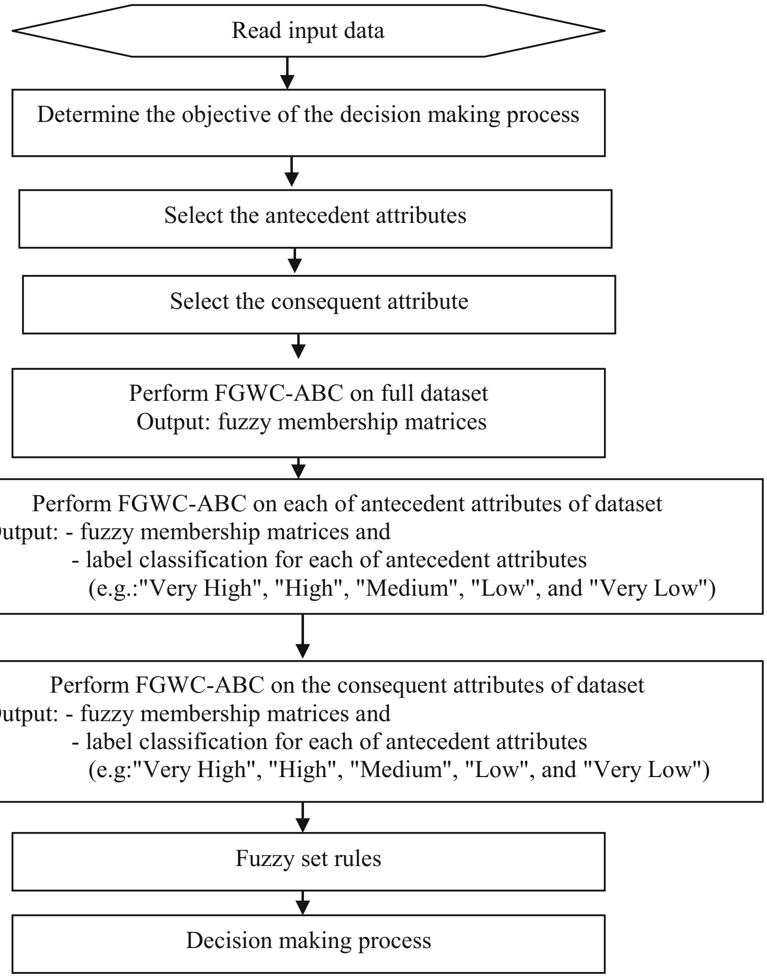

Fig. 5 Procedure to build fuzzy rules using FGWC-ABC

having very high violent crime rate such as Oregon, New Jersey, and Connecticut.

Figure 6 presented the distribution of violent crime rate that resulted after summing up different records into their respective states using the fuzzy weighted aggregation operator. The labels represent for the cluster of respective antecedent and consequent attributes that are calculated by FGWC-ABC. After generating fuzzy rules for all available states and obtaining the complete rules as exemplified in Table 19, we performed geo-demographic mapping of the violent crime rate attribute as the goal of this decision making process.

Table 19 Sample of Fuzzy Rules for Determining The Violent Crime Rate

State	Per Capita Income	Poverty	Unemployed	Divorce	Homeless	Working Mom	Violent Crime Rate
Delaware	High	High	High	Very High	Medium	High	Very High
District of Columbia	Medium	High	Very Low	Medium	Very Low	High	Low

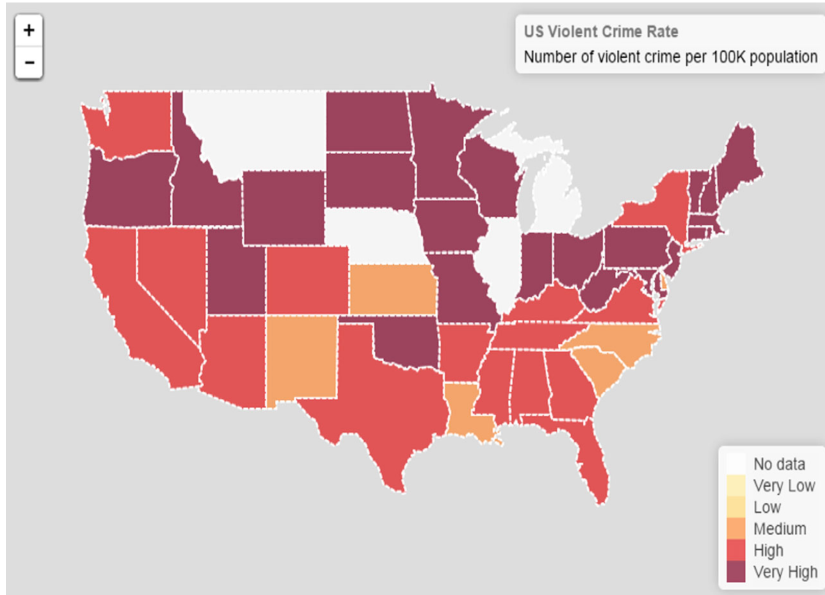

Fig. 6 Mapping geo-demographic cluster in US based on violent crime rate

6 Conclusions

This paper aimed to propose the design for improvement of the limitations in fuzzy geo-demographic clustering algorithm by giving an integration of Artificial Bee Colony (ABC) algorithm based optimization and Fuzzy Geographically Weighted Clustering (FGWC) algorithm to reach a better geo-demographic clustering accuracy. The new algorithm used the ABC algorithm to select the cluster centers (centroids) or the membership matrix automatically in the initialization phase of FGWC clustering. Different objective functions were proposed to distinguish the treatment for different datasets. The proposed design was implemented as a contribution to the fuzzy geo-demographic clustering field.

The simulation results on various geo-demographic datasets showed that the clustering quality of the proposed algorithm so-called FGWC-ABC is better than those of the relevant works such as FCM, Neighborhood Effects (NE) and FGWC. Based on various validity indices for fuzzy clustering, the proposed FGWC-ABC was verified as the robust and efficient method. FGWC-ABC was also applied to a decision-making application for analyzing crime behavior problems in population using the communities and crime dataset, which consists of socio-economic variables from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. The aim of application was to perform the identification of crime cases in United States. A novel procedure based on FGWC-ABC was proposed to generate fuzzy rules of violent crime rate in US. A web-based GIS application mapping to present the distribution of the violent crime rate using the fuzzy weighted aggregation operator was designed. The fuzzy rules and distribution maps would help decision makers manage the violent crime rate more efficiently.

Future works include the comparison using other meta-heuristic optimization and the use of context information within FGWC-ABC. It is also possible to elaborate the use of FGWC-ABC in spatial interaction model application. Other considerable challenges are the implementation of this algorithm in real world geo-demographic applications. This method still leaves a weakness, i.e., the computational time is longer than those of other methods reported previously. Further investigation is in progress to shorten this time.

Acknowledgments The authors are greatly indebted to the editor-in-chief, Prof. Moonis Ali; anonymous reviewers for their comments and their valuable suggestions that improved the quality and clarity of paper. The authors wish to give special thanks to Mr. Nguyen Tho Thong, VNU for suggestion on experimental simulation of this work. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2014.01.

References

1. Bache K, Lichman M (2013) UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine. Available at: http://archive.ics.uci.edu/ml. [Accessed: 01-Sep-2014]
2. Balasko B, Abonyi J, Feil B (2005) Fuzzy Clustering and Data Analysis Toolbox: For Use with Matlab, Veszprem, Hungary

3. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms, p 256. Kluwer Academic Publishers, Norwell
4. BPS (2011) BPS Analysis Document, version 3.1, vol 1, Badan Pusat Statistik, Indonesia
5. BPS (2014) Statistics Indonesia - 2010 Population Census. Available at: sp2010.bps.go.id. [Accessed: 01-Sep-2014]
6. Chang C, Shyue SW (2009) A study on the application of data mining to disadvantaged social classes in Taiwan's population census. Expert Syst Appl 36(1):510–518
7. Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ (2006) Fuzzy c-means clustering with spatial information for image segmentation. Comput Med Imaging Graph 30(1):9–15
8. Cuong BC, Son LH, Chau HTM (2010) Some context fuzzy clustering methods for classification problems. In: Proceedings of the 2010 ACM Symposium on Information and Communication Technology, pp 34–40
9. Feng ZQ, Flowerdew R (1998) Fuzzy geodemographics: a contribution from fuzzy clustering methods. In: Carver S. (ed) Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK (GISRUK). Taylor & Francis Group, London, pp 112–120
10. Grekousis G, Thomas H (2012) Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods. Appl Geogr 34:125–136
11. Han J, Kamber M (2006) Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann
12. Hettich S, Bay SD (1999) The UCI KDD Archive, University of California, Department of Information and Computer Science, Irvine. Available at: http://kdd.ics.uci.edu. [Accessed: 01-Sep-2014]
13. Horng M (2011) Multilevel thresholding selection based on the artificial bee colony algorithm for image segmentation. Expert Syst Appl 38(11):13785–13791
14. Izakian H, Abraham A (2011) Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert Syst Appl 38(3):1835–1838
15. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8(1):687–697
16. Karaboga D, Ozturk C (2010) Fuzzy clustering with artificial bee colony algorithm. Sci Res Essays 5(14):1899–1902
17. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization, TechnicalReport-TR06, Kayseri/Türkiye
18. Lu Y, Ma T, Yin C, Xie X, Tian W, Zhong S (2013) Implementation of the Fuzzy C-Means Clustering Algorithm in Meteorological Data. Int J Database Theory Appl 6(6):1–18
19. Maimon O, Rokach L (2010) Data Mining and Knowledge Discovery Handbook, 2nd Edn. Springer, US
20. Mason GA, Jacobson RD (2007) Fuzzy Geographically Weighted Clustering. In: Proceedings of the 9th International Conference on Geocomputation, pp 1–7
21. Mullen RJ, Monekosso D, Barman S, Remagnino P (2009) A review of ant algorithms. Expert Syst Appl 36(6):9608–9617
22. Niu Q, Huang X (2011) An improved fuzzy C-means clustering algorithm based on PSO. J Softw 6(5):873–879
23. Páez A, Trépanier M, Morency C (2011) Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. Transp Res Part A Policy Pract 45(7):640–652
24. Runkler TA, Katz C (2006) Fuzzy Clustering by Particle Swarm Optimization. IEEE International Conference on Fuzzy Systems 3(2006):601–608
25. Simon D (2013) Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence. Wiley
26. Son LH, Cuong BC, Lanzi PL, Thong NT (2012) A Novel Intuitionistic Fuzzy Clustering Method for Geo-Demographic Analysis. Expert Syst Appl 39(10):9848–9859
27. Son LH, Cuong BC, Long HV (2013) Spatial interaction – modification model and applications to geo-demographic analysis. Knowl-Based Syst 49:152–170
28. Son LH (2015) DPFCM: A Novel Distributed Picture Fuzzy Clustering Method on Picture Fuzzy Sets. Expert Syst Appl 42(1):51–66
29. Son LH (2014) Enhancing Clustering Quality of Geo-Demographic Analysis Using Context Fuzzy Clustering Type-2 and Particle Swarm Optimization. Appl Soft Comput 22:566–584
30. Son LH (2014) HU-FCF: A Hybrid User-Based Fuzzy Collaborative Filtering Method in Recommender Systems. Expert Syst Appl 41(15):6861–6870
31. Son LH, Lanzi PL, Cuong BC, Hung HA (2012) Data Mining in GIS: A Novel Context-Based Fuzzy Geographically Weighted Clustering Algorithm. International Journal of Machine Learning and Computing 2(3):235–238
32. Son LH, Linh ND, Long HV (2014) A lossless DEM compression for fast retrieval method using fuzzy clustering and MANFIS neural network. Eng Appl Artif Intell 29:33–42
33. Son LH (2014) Optimizing Municipal Solid Waste Collection Using Chaotic Particle Swarm Optimization in GIS Based Environments: A Case Study at Danang City, Vietnam. Expert Syst Appl 41(18):8062–8074
34. Son LH, Thong NT (2015) Intuitionistic Fuzzy Recommender Systems: An Effective Tool for Medical Diagnosis. Knowl-Based Syst 74:133–150
35. Thong PH, Son LH (2014) A new approach to multi-variables fuzzy forecasting using picture fuzzy clustering and picture fuzzy rules interpolation method. In: Proceeding of 6th International Conference on Knowledge and Systems Engineering, pp 679–690
36. UNSD (2011) UNSD Statistical Databases: Demographic Yearbook, 16
37. Wijayanto AW, Purwarianti A (2014) Improvement Design of Fuzzy Geodemographic Clustering using Artificial Bee Colony Optimization. 3rd International Conference on Information Technology for Cyber and IT Service Management 1(2):D13–D16
38. Witten IH, Frank E, Hall MA (2011) Data Mining, Practical Machine Learning Tools and Techniques, 3rd edn. Elsevier, Netherland
39. Wu J, Xiong H, Liu C, Chen J (2012) A Generalization of Distance Functions for Fuzzy c-Means Clustering With Centroids of Arithmetic Means. IEEE Trans Fuzzy Syst 20(3):557–571
40. Yang X, Press L (2010) Nature-Inspired Metaheuristic Algorithms. Luniver Press, 2nd Edn
41. Yang X (2012) Nature-Inspired Mateheuristic Algorithms: Success and New Challenges. J Comput Eng Inf Technol 1(01)
42. Levitt SD (2004) Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not. The Journal of Economic Perspectives 18(1):163–190
43. Snook B, Cullen RM, Mokros A, Harbort S (2005) Serial Murderers' Spatial Decisions: Factors that Influence Crime Location Choice. J Investig Psychol Offender Profiling 164(2): 147–164
44. Berenji HR, Khedkar P (1992) Learning and Tuning Fuzzy Logic Controllers Through Reinforcements. IEEE Transactions on Neural Networks 3(5):724–740

Arie Wahyu Wijayanto is currently a Master student at School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia. His research interests are Fuzzy Clustering, Data Mining, Machine Learning and Service Oriented Architecture. So far, he has published some papers in national and international conferences. Email: ariewahyu@students.itb.ac.id

Dr. Ayu Purwarianti is the Chair of Informatics/Computer Science Undergraduate Program, School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia. She obtained Dr.Eng. degree on Informatics from Toyohashi University of Technology, Japan and until now she has been Assistant Professor of Institut Teknologi Bandung, Indonesia. Her research interests are Natural language processing, Speech Processing, Intelligent Tutoring System and Knowledge Management System. She is the member of IEEE/ACM and has published many papers in national and international conferences and journals. Email: ayu@stei.itb.ac.id

Dr. Le Hoang Son obtained the Bachelor, Master and PhD degrees on Mathematics – Informatics at VNU University of Science, Vietnam National University (VNU). He has been working as a researcher and lecturer at the Center for High Performance Computing, VNU University of Science, Vietnam National University since 2007. His major field includes Soft Computing, Fuzzy Clustering, Recommender Systems, Geographic Information Systems (GIS) and Particle Swarm Optimization. He is a member of IACSIT and also an associate editor of the International Journal of Engineering and Technology (IJET).

Dr. Son served as a reviewer for various international journals and conferences such as PACIS 2010, ICMET 2011, ICCTD 2011, KSE 2013, BAFI 2014, NICS 2014 & 2015, ACIIDS 2015, ICNSC15, GIS-2015, International Journal of Computer and Electrical Engineering, Imaging Science Journal, International Journal of Intelligent Systems Technologies and Applications, IEEE Transactions on Fuzzy Systems, Expert Systems with Applications, International Journal of Electrical Power and Energy Systems, Neural Computing and Applications, International Journal of Fuzzy System Applications, Intelligent Data Analysis, Computer Methods and Programs in Biomedicine, World Journal of Modeling and Simulation, Knowledge-Based Systems.

Dr. Son has got 43 publications in prestigious journals and conferences including 15 SCI/SCIE papers and undertaken more than 20 major joint international and national research projects. He has published 2 books on mobile and GIS applications. So far, he has awarded the "2014 VNU Research Award for Young Scientists". Email: sonlh@vnu.edu.vn / chinhson2002@gmail.com