JASEM ISSN 1119-8362

Full-text Available Online at
www.ajol.info and
www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage. Dec, 2011
Vol. 15 (4) 635 - 641

Using the Hopfield– Fuzzy C Means Algorithm for Clustering of People based on Food Insecurity and Obesity in the Northwest of Iran

*HELDA TUTUNCHI; ALIREZA OSTADRAHIMI; SOLTANALI MAHBOOB; REZA MAHDAVI; SIROUS J.TABRIZI

School of Public Health and Nutrition, Nutritional Science Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

ABSTRACT: It is commonly recognized that food insecurity and obesity is the major sections of academic researches, but the activities' rate of food insecurity and surveillance is always Low. Based on food insecurity and obesity' characteristics, in this paper attempt are made to present a model based on Hopfield– fuzzy C Means clustering algorithm. Firstly, it is capable of identifying the reasons behind the emergence of the present status. Secondly, the suggested model must represent the clustering of the people based on food insecurity and surveillance in different levels. Finally, it tests the validity of the suggested model with comparing by other models (Hopfield–K-Means, K-Means, and fuzzy C Means). @JASEM

Keywords: Learning, Clustering, Algorithm, Hopfield.

*Corresponding Author: Helda Tutunch, **Email:** HeldaTutunchi@yahoo.com

The concept of household food insecurity includes problems with the quantity of available food, uncertainty about food supply, and experience of hunger in life (Alaimo et. all, 1998). Food insecurity is frequent in both developed and developing countries, affecting from 5% to 25% of the general population in different research reports (Blumberg et. all, 1999). It has considerable health impacts on the physical, social, and psychological status of individuals in communities suffering from food insecurity. It may also affect the quality of life of households (Frongillo et. all, 1997). Various techniques and methods have been used to measure food insecurity in many countries (Gulliford et. all, 2004). The aim of this study was to document the epidemiologic features of food insecurity in the northwest region of Iran, and to evaluate the sensitivity and specificity of a short-form (six items) questionnaire for screening for food insecurity in the region.

Obesity has been described by the World Health Organization as an 'escalating epidemic' as great as that of smoking1. Obesity has now become a global epidemic and its prevalence continues to increase in both developed and developing countries3. The total prevalence of overweight and obesity is estimated at between 50 and 65% of the general population in some developed countries. An increasing trend in overweight and obesity prevalence has also been reported from other developing countries in recent years. Hereditary, environmental, metabolic and behavioral factors may all have a role in the development and progression of obesity. Obesity is associated with common causes of morbidity and mortality in the population, including diabetes9, coronary heart disease, hypertension and dyslipidemia, some types of cancer1 and mental health problems.

Since, goal of improving of the Obesity is decreasing of distance between the activated nodes, to improve productivity it would be necessary to identify the present status at first and then the causes and the solutions are describe based on suitable algorithm. Thus, it needs to use of a Learning Automata-based Clustering Algorithm. There are several algorithms for clustering of WSN (such as the hierarchical clustering, K-Means, C-Means, Hopfield, SOM models). In this paper, we present the Hopfield–fuzzy C Means algorithm in order to overcome these limitations. This approach eliminates the randomness of the initial solution provided by fuzzy C-Means based algorithms and it moves closer to the global optimum.

The study is set seven major sections; the second part presents the related works. The third part describes the proposed algorithm based on the Hopfield– fuzzy C Means algorithm. The forth part is expressed case study, and in the next sections, it will be discussed analysis and presentation of research findings and suggestions for future research results.

Related works: Hunger and food security have been identified as national priorities that, in principle, should have particular relevance for nutrition education.1 For instance, the U.S. Department of Health and Human Services has adopted the goal of increasing the prevalence of food security among U.S. households as one of the health objectives for the nation for the year 2010.2 Maxwell et al. (2011) Measuring food insecurity: Can an indicator based on localized coping behaviors be used to compare across contexts? The Coping Strategies Index (CSI) was developed as a context-specific indicator of food insecurity that counts up and weights coping behaviors at the household level. It has proven useful to operational humanitarian agencies and researchers in measuring localized food insecurity, but to date has not been useful to compare the relative severity of different crises and has therefore has not been particularly useful for geographic targeting or resource allocation. This paper analyzes data from 14 surveys in crisis affected or chronically vulnerable countries in Sub-Saharan Africa that incorporated the context-specific CSI. The paper identifies a sub-set of individual coping behaviors common to all surveys, whose severity is regarded as broadly similar by households across these studies. Data from these studies were re-analyzed using a reduced index constructed from only these behaviors. Correlations of this new index with other known food security indicators are similar to those of the complete, context-specific CSI. This suggests the possibility that an indicator based on these common behaviors could be used to compare the types of food security crises analyzed here across different contexts – particularly in Sub-Saharan Africa – to improve geographic targeting and resource allocation, according to the severity of crises. This new, more comparative indicator can be generated with no loss to the context-specific nature of the original CSI, which has proven useful for assessment and monitoring purposes. There are few universally valid indicators of food security that are applicable in crisis situations. Nutritional status, if properly measured, is widely accepted as comparable across different contexts. But while nutritional status can be one indicator of food security status, it may equally reflect elements of health status, care practices, water quality, and other determinants of nutrition (Young and Jaspars, 2006). Some analysts suggest that measuring actual food consumption at the household level by a 24-h recall should be the ''gold standard'' by which other food security indicators are measured (Hoddinott and Yohannes, 2002; Weismann et al., 2006). But while 24-h recall data accurately reflects current consumption status, it does not capture other elements of the complex notion of food security. And the methodology is far too time-consuming to be useful in the applications discussed above – early warning, assessment, targeting or monitoring – all of which are very time-sensitive.

Patterns of behavioral responses in relation to a food shortage have been documented previously by several researchers (Davies, 1996). Watts (1983) presented a sequence of options based on their reversibility and commitment of domestic resources. Modest dietary adjustments (such as eating less-preferred foods or reducing portion size), for example, are highly reversible strategies that do not jeopardize household assets. More extreme behaviors, such as sales of productive assets to purchase food, hold more long-term consequences for the household. As a food security situation worsens, households are more likely to employ strategies that are less reversible, and therefore represent a more severe form of coping and greater food insecurity (Corbett, 1988; Devereux, 1993).

The Coping Strategies Index is similar in many respects to other measures of food security but distinct in that it queries household behaviors directly, and factors in the severity of different behaviors. Given that no one ''gold standard'' indicator has emerged, particularly for use in humanitarian emergencies, different measures of food security are needed for triangulation or complementary analysis. Attempts at developing and refining indicators of food access have revealed a number of critical considerations. First, food security is a ''managed process'' with predictable patterns – people can foresee a food access problem before it arises and thus begin to alter behavior long before an actual crisis hits a household (Christiaensen and Boisvert, 2000). Second, with respect to coping strategies, it must be noted that some strategies do not necessarily reflect the same severity of food insecurity, nor are they equally acceptable to vulnerable households in different cultures (Coates et al., 2006a,b). To develop more broadly applicable measures of food security, adequate attention must be given to developing methods of translating or adapting measures from one context to another (Swindale and Bilinsky, 2006; Coates et al., 2006a,b; Webb et al., 2006; Maxwell et al., 1999). And third, although some progress has been made, the search for more broadly applicable measures of food security continues.

There is evidence that food insecurity, particularly transitory food insecurity, has been getting worse in Malawi. In 2001–2003 Malawi suffered a food crisis. This was manifested in a six-fold increase in food prices, which left around 3.5 million people food insecure. The crisis was the combined result of climatic shocks, mis-management of the country's strategic grain reserve, poor crop estimates and a chaotic delayed response in terms of maize imports (Stevens et al., 2002; World Development Movement, 2002).

Tsai and Lin (2011) review Fuzzy C-means based clustering for linearly and nonlinearly separable data. In this paper, they present a new distance metric that incorporates the distance variation in a cluster to regularize the distance between a data point and the cluster centroid. It is then applied to the conventional fuzzy C-means (FCM) clustering in data space and the kernel fuzzy C-means (KFCM) clustering in a high-dimensional feature space. Experiments on two-dimensional artificial datasets, real datasets from public data libraries and color image segmentation have shown that the proposed FCM and KFCM with the new distance metric generally have better performance on non-spherically distributed data with uneven density for linear and nonlinear separation. The experiments on the 2D artificial datasets have shown that the kernel-based clustering methods can well partition nonlinearly distributed data, but only up to quadratic functions. It is worthy of further investigation for extending the kernel-based clustering to a higher polynomial function. A good computational strategy is also required in the future to use the kernel-based clustering for datasets with a huge number of data points, especially for color segmentation in a very large image.

Mingoti and Lima (2006) Compare SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. In this paper, they present a comparison among some nonhierarchical and hierarchical clustering algorithms including SOM (Self-Organization Map) neural network and Fuzzy c-means methods. Data were simulated considering correlated and uncorrelated variables, non overlapping and overlapping clusters with and without outliers. A total of 2530 data sets were simulated. SOM neural network did not perform well in almost all cases being very affected by the number of variables and clusters. The traditional hierarchical clustering and K-means methods presented similar performance. The results showed that Fuzzy c-means had a very good performance in all cases being very stable even in the presence of outliers and overlapping. All other clustering algorithms were very affected by the amount of overlapping and outliers.

Loópez et al. (2011) purpose Hopfield–K-Means clustering algorithm as a proposal for the segmentation of electricity customers. Customer classification aims at providing electric utilities with a volume of information to enable them to establish different types of tariffs. Several methods have been used to segment electricity customers, including, among others, the hierarchical clustering, Modified Follow the Leader and K-Means methods. These, however, entail problems with the pre-allocation of the number of clusters (Follow the Leader), randomness of the solution (K-Means) and improvement of the solution obtained (hierarchical algorithm). Another segmentation method used is Hopfield's autonomous recurrent neural network, although the solution obtained only guarantees that it is a local minimum. In this paper, they present the Hopfield–K Means algorithm in order to overcome these limitations. This approach eliminates the randomness of the initial solution provided by K-Means based algorithms and it moves closer to the global optimum. The proposed algorithm is also compared against other customer segmentation and characterization techniques, on the basis of relative validation indexes. Finally, the results obtained by this algorithm with a set of 230 electricity customers (residential, industrial and administrative) are presented.

According the Mingoti and Lima (2006), they say that Fuzzy c-means had a very good performance in all cases being very stable even in the presence of outliers and overlapping (rather than SOM, HC, K MEAN, …), and so according to Loópez et al. (2011), they say that the Hopfield–K Means algorithm has good performance rather than other algorithms. Therefore, in this paper, I claim that the Hopfield–fuzzy C Means algorithm has better performance than the Hopfield–K Means algorithm.
Therefore, I present the Hopfield– fuzzy C Means algorithm in order to overcome these limitations. This approach eliminates the randomness of the initial solution provided by fuzzy C-Means based algorithms and it moves closer to the global optimum.

The Hopfield–Fuzzy C Means Algorithm: The algorithm that is presented in this paper has been developed to classify in an optimal manner a set X comprising a large number Q of Mean Load Curves (MLC) of electric energy customers previously characterized and standardized. The load curves are represented as follows:

$$X(i,j) = x_j^i \quad \forall i = 1,\ldots,Q; \quad j = 1,\ldots,\nu$$

$$X = \left\{x^{(i)}, \quad i = 1,\ldots,Q\right\}$$

$$C^{(X)} = \frac{1}{Q}\sum_{x\in X} x^{(i)}$$

Where $x^{(i)}$ represents a characterized and standardized node i and x^i_j represents term j of node i. The value of v will depend on the characterization of node i. Finally, $C^{(X)}$ is the centroid of load curve set X. The objective of this segmentation algorithm is to form a certain number of clusters X(k) ⊂X, k=1, . . ., K with a great deal of dissimilarity among the different classes and very little dissimilarity inside each class. H-ANN–K is a two-stage hybrid algorithm. In the first stage, an initial segmentation is obtained using the H-ANN algorithm. On the basis of the initial segmentation, in the second stage, the K-Means algorithm is applied to refine the solution obtained by H-ANN until the final segmentation is achieved. It is worth noticing that, as apposed to classical K-Means algorithms, the random formation of the initial cluster centroids is replaced with an initial assignment applying the H-ANN algorithm. The algorithm is robust since it always achieves the same solution. Distance is used as the measure of dissimilarity:

$$d(x^{(i)},x^{(l)}) = \left\{\sum_{j=1}^{\nu}\left(x_j^i - x_j^l\right)^p\right\}^{1/p} \quad ; \quad \forall (x^{(i)},x^{(l)})\in X$$

-H-ANN algorithm: first stage: In general terms, a H-ANN is a complete graph G= (V, A) whose vertices (V) represent N neurons connected to each other by means of edges or arcs (A). A synaptic weight w_{ij}, which is a real number that represents the strength of the link between the neurons i, j, is associated to each axis or connection. Synaptic weights (SW) are the distances, and are represented as:
$SW(i, j) = w_{i,j}\ \forall i, j = 1, \ldots, N$
This matrix is symmetric ($w_{ij} = w_{ji}$) and has zero diagonal (w_{ii} =0). In general terms, the basic computational element for a H-ANN network is a bipolar processing unit. Its mathematical function is defined in the set $\{-1,1\}^N$, and can be described by means of the following expression:

$$f(x_1,x_2,\ldots,x_N) = \begin{cases} 1 & \text{if} \quad x_1w_1+\ldots+x_Nw_N \geq \theta \\ -1 & \text{if} \quad x_2w_2+\ldots+x_Nw_N < \theta \end{cases}$$

where x_N represents the state of the network's N neurons. Furthermore, in the case of multi-valued recurrent networks, state x_i of each N neuron will be characterized by its output s_i which, in a general formulation, can take any value from a set that we will call M. Said set can take values in R, or in a non-numerical (qualitative) set. Since the network evolves over time, the state of the ith neuron will be characterized by the value of its output at that instant $s_i(t)$. The vector $S(t) = [s_1(t), s_2(t), \ldots, s_N(t)]$, which describes the state of the network's neurons at the given instant t, will be called the state vector of the network at instant t. An energy function (E) is associated to this state vector and defined as:

$$E(t) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,j} f(s_i, s_j) + \sum_{i=1}^{N}\theta_i s_i$$

Where function f is an application of M×M→R and measures the analogy or similarity between the outputs of ith and jth neurons. The network's dynamic evolution follows an asynchronous sequential procedure. The H-ANN begins with an initial state vector S(0). The states of all the neurons for instants 1,2, . . ., t, t + 1 are calculated according to the following computation rule:

$$s_i(t+1) = \begin{cases} 1 & \text{if } \sum_{j=1}^{N} w_{ij}s_j(t) \geq \theta \\ -1 & \text{if } \sum_{j=1}^{N} w_{ij}s_j(t) < \theta \end{cases}$$

For each instant 1, 2, . . ., t, t + 1, there is a state vector S(1), S(2), . . ., S(t), S(t + 1); the union of which will represent the network's state space. The network evolves in such a way that at every instant the value of the energy function associated to its state vector decreases as much as possible. The final objective is to minimize the energy function E for the simulation interval considered, thus guaranteeing a local minimum. The final solution is the state vector for the minimum energy function value. For the purpose of segmentation, a multistate H-ANN network must be constructed taking into account the following considerations:

(1) Synaptic weights w_{ii} are calculated with distance p. Cases p=1 and p = 2 (Euclidean distance) have been considered.
(2) The output (si) of each neuron is a value of the set M= {1, 2, . . . K}, K being the number of clusters.
(3) The thresholds θ_i of all network neurons are considered null.
(4) The similarity function used is that of identity: f(x, y) = (x≡y), which can be expressed as:

$$f(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

(5) The energy function for an instant t will be determined by the following expression:

$$E(t) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,j}(s_i(t) \equiv s_j(t))$$

where (si ≡sj) will be equal to 1, if the states of ith and jth neurons coincide, and 0 if they do not.

(6) The solution obtained is the state vector S(t) that corresponds to the minimum energy function (E) of the simulation interval.

(7) Classes K, associated with the state vector S(t), are determined. The subset $X^{(k)}{}_0 \subset X$, k = 1, . . . , K contains the MLC whose neurons have the same output (si ≡sj).
(8) The centroids $C^{(k)}{}_0$ of the classes formed ($X^{(k)}{}_0$) are calculated, and will be used later in the K-Means algorithm for the final segmentation.

Fuzzy C-Means: second stage: FCM partitions the dataset X into C clusters by minimizing the errors in terms of the weighted distance of each data point x_i to all centroids of the C clusters. The algorithm's steps are the following:
(1) The data set is randomly partitioned into C classes, calculating the initial centroids. In our case said initial _centroids are obtained by the H-ANN algorithm, $C0 = (C^{(1)}{}_0, C^{(2)}{}_0, \ldots, C^{(k)}{}_0, \ldots, C^{(K)}{}_0)$.

(2) For each element of set X, the distance (p = 1 and p = 2) of each element $x^{(i)} \in X^{(k)}{}_0$ to its own centroid, and to the centroids of the other classes, is calculated. If any element is not currently in the class whose centre is the closest, change the sample class and recalculate the set of centroids. The new set of centroids will be $C1 = (C^{(1)}{}_1, C^{(2)}{}_1, \ldots, C^{(k)}{}_1, \ldots, C^{(K)}{}_1)$. After it, we should be obtaining followings:

E-step:

$$w_{ic} = 1 \Big/ \sum_{j=1}^{C}\left(\frac{d_{ic}^2}{d_{ij}^2}\right)^{1/(p-1)} \quad \text{for } i = 1, 2, \ldots, N \text{ and } c = 1, 2, \ldots, C$$

where

$$d_{ic}^2 = \|\mathbf{x}_i - \mathbf{v}_c\|^2$$

M-step:

$$\mathbf{v}_c = \frac{\sum_{j=1}^{N} w_{jc}^p \cdot \mathbf{x}_j}{\sum_{j=1}^{N} w_{jc}^p} \quad \text{for } c = 1, 2, \ldots, C$$

Repeat step (2) until the algorithm reaches convergence, in other words, until there are no variations of class centroids in two consecutive iterations. In this case a subset $X^{(k)}{}_f \subset X$, k = 1, · · ·,K and associated centroids

$C_f = (C^{(1)}{}_f, C^{(2)}{}_f, \ldots, C^{(k)}{}_f, \ldots, C^{(K)}{}_f)$are obtained.

This is the final solution obtained by the H-ANN–fuzzy C Mean algorithm. This cross-sectional study was conducted on selected randomly in the Asadabadi area of the northwest of Iran.

Case study: The case study used consisted of information regarding nodes periodically report 300 subjects (132 male and 168 female). All simulations have been implemented using Matlab 2010. The H-ANN–K (HK) segmentation method was compared with Hopfield's recurrent neural network (H), K-Means (K) [2,9], SOM–K-Means (SOMK), the Modified Follow the Leader (F) and hierarchical clustering with two criteria of linkage, mean (DM) and Ward's (DW). The number of classes was considered in a range from 5 to 30.

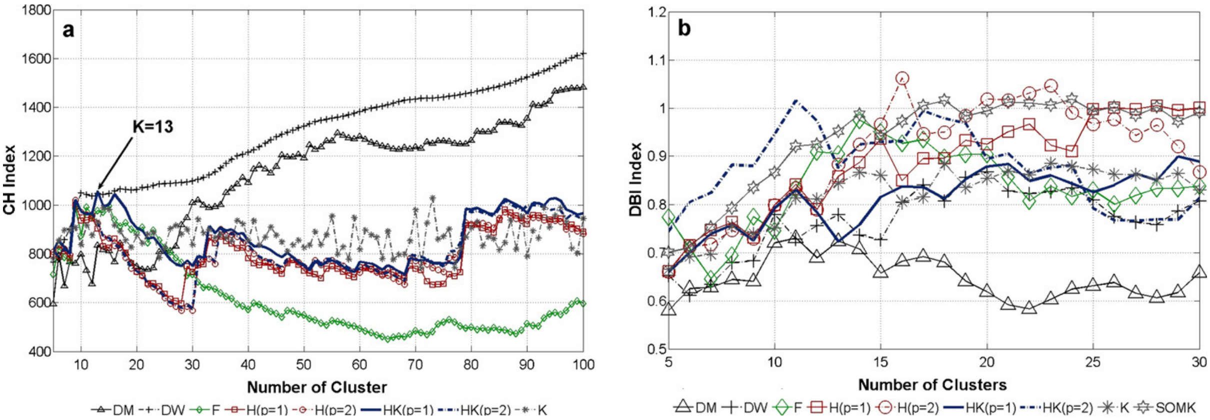

Fig. 1. (a) MIA index from 5 to 30 clusters. (b) Relationship of variability from 5 to 30 clusters.

The efficiency of the algorithm proposed in this paper was compared with the other segmentation methods, using relative validation indexes.

Conclusion: We have defined the Learning Automata-based Clustering Algorithm based on Hopfield– Fuzzy C Means clustering algorithm. The proposed clustering algorithm, in an iterative process tries to find a policy that determines a cluster-head set with the minimum cardinality. This approach eliminates the randomness of the initial solution provided by Fuzzy C-Means based algorithms and it moves closer to the global optimum.

Acknowledgement: This work was supported by a grant from Tabriz University of Medical Sciences.

REFERENCES

Alaimo K., Briefel R.R., Frongillo E.A. Jr., Olson C.M., (1998). Food insufficiency exists in the United States: Results from the third National Health and Nutrition Examination Survey (NHANES III), Am J Public Health,419–26.

Asadabadi, (2003). Population and Health Survey, Tabriz Med J.,suppl 59:1s–42s, 115s–120s.

Bickel G., Nord M., Price C., Hamilton W., Cook J., (2000). Measuring food security in the United States. Guide to measuring food security, Washington, DC: US Department of Agriculture, Food and Nutrition Service.

Blumberg S.J., Bialostosky K., Hamilton W.L., Briefel R.R., (1999). The effectiveness of a short form of the Household Food Security Scale, Am J Public Health, 12131–4.

Campbell C.C., (1991). Food insecurity: A nutritional outcome or a predictor variable?, J Nutr, 121:408–15.

Carlson S.J., Andrews M.S., Bickel G.W. (1999). Measuring food insecurity and hunger in the United States: Development of a national benchmark measure and prevalence estimates, J Nutr, 129(2S suppl):510S–6S.

Coates J., Frongillo E.A., Rogers B.L., Webb P., Wilde P.E., Houser R., (2006). Commonalities in the experience of household food insecurity across cultures: What are measures missing?, J Nutr, 136:1438S–48S.

Frongillo E.A., Olson C.M., Rauschenbach B.A., Kendall A., (1997). Nutrition consequences of food insecurity in a rural New York State County, Institute for Research on Poverty Discussion Paper No. 1120- 97, Madison, WI, USA: University of Wisconsin.

Gulliford M.C., Mahabir D., Rocke B., (2004). Reliability and validity of a short form household food security scale in a Caribbean community, BMC Public Health,4:22.

Gulliford M.C., Mahabir D., Rocke B., (2003). Food insecurity, food choices, and body mass index in adults: Nutrition transition in Trinidad and Tobago, Int J Epidemiol, 32:508–16.

Gulliford M.C., Nunes C., Rocke B., (2006). The 18 Household Food Security Survey items provide valid food security classifications for adults and children in the Caribbean, BMC Public Health, 6:26.

Hamilton W.L., Cook J.T., Thompson W.W., Buron L.F., Frongillo E.A. Jr., Olson C.M., Wehler C.A., (1997). Household food security in the United States in 1995: Summary report of the Food Security Measurement Project. Alexandria, Va, USA: US Department of Agriculture, Food and Consumer Service.

Kaiser L.L., Townsend M.S., Melgar-Quinonez H.R., Fujii M.L., Crawford P.B., (2004). Choice of instrument influences relations between food insecurity and obesity in Latino women, Am J Clin Nutr, 80:1372–8.

Kendall A., Olson C.M., Frongillo E.A. Jr., (1996). Relationship of hunger and food insecurity to food availability and consumption, J Am Diet Assoc, 96:1019–24.

Laquatra I., (2004). Nutrition for weight management In: Mahan K, Escot-Stump S, eds. Krause's food, nutrition and diet therapy, 11th ed. Philadelphia, Pa, USA: Saunders,565.

Olson C.M., (1999). Nutrition and health outcomes associated with food insecurity and hunger. J Nutr, 129(2S suppl):521S–4S.

Qasemi H., (1998). Food security programme in Islamic Republic of Iran. Tehran: Institute of Food and Nutrition Research, (in Persian).

Radimer K.L., Olson C.M., Campell C.C., (1990). Development of indicators to assess hunger. J Nutr, 120(suppl 11):1544–8.

Sarlio-Lahteenkorva S., Lahelma E., (2001). Food insecurity is associated with past and present economic disadvantage and body mass index. J Nutr, 131:2880–4.

Swindale A., Bilinsky P., (2006). Development of a universally applicable household food insecurity measurement tool: Process, current status, and outstanding issues, J Nutr, 136:1449S–52S.

Townsend M.S., Peerson J., Love B., Achterberg C., Murphy S.P., (2001). Food insecurity is positively related to overweight in women. J Nutr, 131:1738–45.

Webb P., Coates J., Frongillo E.A., Rogers B.L., Swindale A., Bilinsky A., (2006). Measuring household food insecurity: Why it's so important and yet so difficult to do, J Nutr 2006, 136:1404S–8S.

Zarafati Shoa N., (2003). Validity of Radimer-Cornell questionnaire in Tehran. MSc Thesis, Tehran Shahid Beheshti University of Medical Sciences, (in Persian).