

Self-Organizing Maps of Nutrition, Lifestyle and Health Situation in the World

Yasir Mehmood¹, Mudassar Abbas², Xi Chen¹, and Timo Honkela¹

¹ Department of Information and Computer Science

² Department of Biomedical Engineering and Computational Science
Aalto University School of Science
ymehmood@cc.hut.fi, {xi.chen,timo.honkela}@tkk.fi,
mudassar.sabir@aalto.fi

Abstract. In this article, we present an analysis of the impact of nutrition and lifestyle on health at a global level. We have used Self-organizing Maps (SOM) algorithm as the analysis technique. SOM enables us to visualize the relative position of each country against a set of the variables related to nutrition, lifestyle and health. The positioning of the countries follows the basic understanding of their status with respect to their socioeconomic conditions. We have also studied the relationships between the variables supported by the SOM visualization. This analysis presents many obvious correlations but also some surprising findings that are worth further analyses.

1 Introduction

The overall relationship between unhealthy diet and deteriorating health is obvious and generally well understood. Establishing this relationship from data may lead us to further identify the severity of the relationship concerning different specific aspects, and to take appropriate corrective actions. Moreover, these types of analyses can be used to create social awareness regarding the strong impact of certain nutrition on the health of individuals and thus promoting overall health and wellbeing in the society. The technical report series 916 of World Health Organization explains that the rapid industrialization, in the previous decade, has affected the health and nutrition especially in the developing countries, which has resulted in “inappropriate dietary patterns, decreased physical activities and increased tobacco use, and a corresponding increase in diet-related chronic diseases, especially among poor people” [1]. The report also states that existing scientific evidence has helped in identifying the role of diet in controlling various diseases. However, the evidence is contradicting at times [1].

A possible way to study correlations of different elements of nutrition and lifestyle with the diseases could be to focus on the eating and drinking trends in different parts of the world. The patterns of nutrition intake vary to a great extent in different regions of the world. As this is also true in the case of the prevalence of various diseases. The connections between the nutrition intake and health can be further examined (see e.g. [2]). Volkert [2] studies the nutrition and

lifestyle of elderly people in Europe and finds out that these vary widely even within Europe. Volkert continues to state that the elderly in the south consume more vegetables, grains, fruit, lean meat and olive oil whereas relatively more milk products are consumed in the northern European countries. These kind of findings are particularly interesting because a deeper analysis of world-wide eating trends and prevalence of diseases may enable us to identify the regions where some improvements in terms of changing eating habits are essential to promote wellbeing. In our research, we have carried out a similar investigation where we can clearly identify large differences in nutrition intake at the global level. Thus, we can group different regions of the world depending on nutrition intake profiles of each region. We have also explored the links between the prevalence of certain diseases in various countries and citizens' diet in those countries. The dataset, that we have used for our analysis, contains more than one hundred nutrition, lifestyle and health indicators in 86 countries. The dataset has been obtained from Canibais e Reis [3], with FAO (Food and Agriculture Organization) [4], WHO (World Health Organization) [5] and British Heart Foundation [6]. as the main sources of the data.

In order to inspect the aforementioned aspects of our research, sophisticated analysis techniques are required since the dimensionality of data is large owing to a lot of health, lifestyle and nutrition related features. For this reason, we have used Self-Organizing Maps (SOM) algorithm, a well-known data analysis and visualization technique, to mine interesting correlations. SOM is a suitable means to create an ordered representation of high-dimensional data. The method reduces the complexity of the data and reveals meaningful relationships by mapping the data into an intuitive two-dimensional space. This also helps in understanding dependencies between variables in the data. In our analysis, we have used SOM for a countrywide grouping of the data or data subsets. For instance, SOM enables us to visually perceive the groups of countries based on the spread of different diseases or the intake of certain nutrition elements. SOM also helps in visualizing the relationships between different food items and diseases.

Section 2 provides a brief introduction of SOM and the dataset. Section 3 presents the results of experimentations on the data and finally Section 4 concludes the paper.

2 Method and Data

This section first describes in brevity different features of the dataset and then sheds some light on the technical and mathematical details of SOM.

2.1 Nutrition, Lifestyle and Health Database

The dataset under investigation is comprised of statistics that can be divided into three categories namely health, diet and lifestyle. The first category contains information such as obesity prevalence, incidence of tuberculosis, mortality rates and related variables in different countries. The dietary information includes the

consumption of proteins, sugar and milk products, and various other components of nutrition (see a subset shown in Fig. 1). The lifestyle category provides information related to the drinking and smoking habits etc. This categorization has helped us, as shown in section 3, to group different countries based on the similarity of food consumption and the spread of diseases.

Nutrition	Argentina	China	Ethiopia	Finland	...	USA
Protein (g/day)	94	82	54	102	...	114
Fats (g/day)	100	90	20	127	...	156
Carbohydrates (g/day)	477.5	450.5	366	399.75	...	426
Animal Products (kcal/day)	823	644	96	1164	...	1045
Animal Fats (kcal/day)	72	46	13	131	...	116
Bovine Meat (kcal/day)	342	27	29	90	...	115
Butter, Ghee (kcal/day)	28	1	5	78	...	40
Cheese (kcal/day)	90	1	0	164	...	149
Eggs (kcal/day)	24	74	1	32	...	55
Fats, Animals, Raw (kcal/day)	42	44	7	17	...	75
Fish, Seafood (kcal/day)	10	35	0	59	...	28
Freshwater Fish (kcal/day)	1	20	0	14	...	5
Honey (kcal/day)	1	1	4	4	...	4
Meat (kcal/day)	475	440	44	497	...	451
Milk - Excluding Butter (kcal/day)	222	30	33	438	...	390
Milk, Whole (kcal/day)	127	28	27	218	...	199
Mutton & Goat Meat (kcal/day)	8	15	6	2	...	3
Offals, Edible (kcal/day)	17	10	3	4	...	3
Pelagic Fish (kcal/day)	1	0	0	34	...	7
Pigmeat (kcal/day)	34	343	0	348	...	132
Poultry Meat (kcal/day)	83	51	2	53	...	197
Vegetal Products (kcal/day)	2135	2296	1761	1978	...	2708
Alcoholic Beverages (kcal/day)	67	150	12	183	...	102
...

Fig. 1. A selection of nutrition-related data

2.2 Self-Organizing Maps

Self-organizing Maps (SOM) is an unsupervised learning algorithm, often referred to as an artificial neural network, that models input patterns in a dataset as an ordered collection of model vectors [7]. The self-organizing map is famous as an effective visualization methods, with which high-dimensional input data can be reduced to 2-dimensional data space that can be interpreted and perceived more easily. SOM has been used in many real world problems ranging from analysis of complex data sets to monitoring of large industrial processes [8,9].

In Self-organizing Maps, the output layer of neurons is usually arranged in a 2-dimensional lattice structure. Each input vector in the input space is connected to each neuron in the output space through a synaptic weight connection which is initialized arbitrarily and learned or updated during the learning process.

The result of the learning process is a particular topological arrangement of the neurons in the output layer. The learning mechanism of SOM consists of three main processes, namely competitive process, co-operative process and adaptive process. These three processes combined can be called the learning algorithm of SOM [7].

The experiments in this work have been conducted using the SOM Toolbox Matlab package [10].

3 Analysis and Results

Our analysis on the data is twofold. In the first phase, we visualize the distribution of countries on the 2-dimensional space of SOM, based on the consumption of different nutrients. This explains how countries group with each other regarding the intake of nutrition. The second phase of the analysis shows relationships between the nutrition variables as well as between the nutrition variables and diseases.

3.1 Nutrition Analysis

For the world-wide nutrition analysis, 50 nutrition variables were selected as an input for the SOM algorithm. The groups shown on the SOM map (see Fig. 2) are based on the commonality of consumption of these nutrients in all 86 countries.

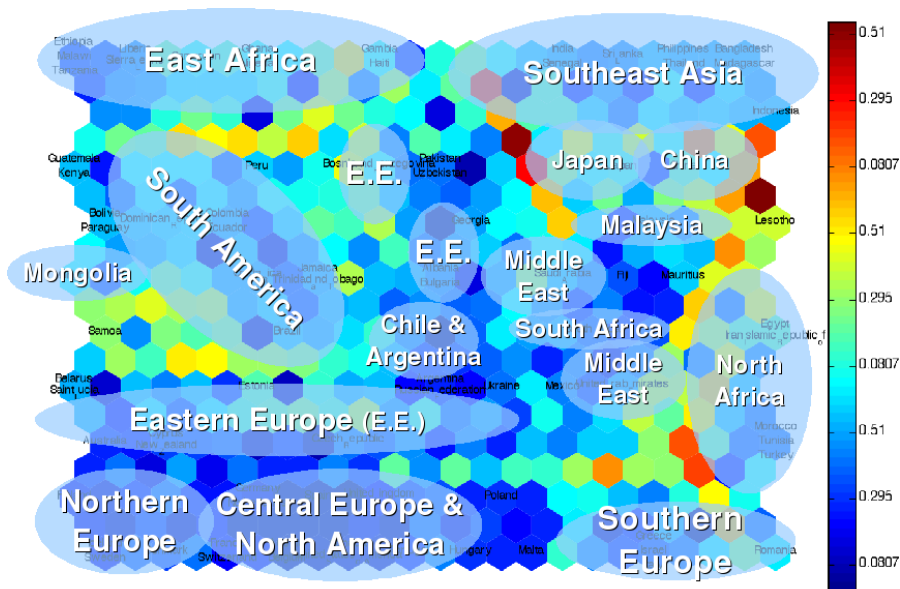


Fig. 2. A map of nutrition variables with an illustration of division of larger geographical groups

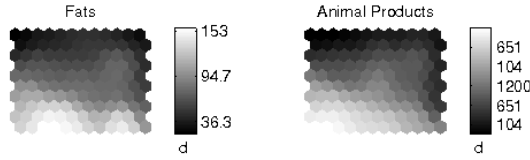


Fig. 3. Component distributions of fats and animal products

In Fig. 2, the structure of clusters on the map is illuminated by the visualization of distances in the original high-dimensional space. The shades of red color indicate large distances in the original space whereas shades of blue refers to relatively shorter distances. The visualization is interesting since the groups of countries on the SOM map appear to have a close resemblance with the region-wise grouping of countries on the globe. This possibly explains the fact that people living in different regions of the world have common preferences for certain foods when compared with others. We have annotated the SOM map by making virtual group boundaries. However, this demarcation is not fully correct since some countries fall in the group of those countries that are not their geographical neighbors. Interesting exceptions include Egypt, Iran and Turkey appearing in the group of North African countries. Moreover, Israel appears in the group of South European countries, and Jamaica in South America. These kind of exceptions are because of the resemblance in eating trends.

We have also performed similar analyses by combining various disease, lifestyle and nutrition variables. The maps of disease and lifestyle variables (disease map is shown in section 3.3) show various similar groups. For example, groups of various South Asian and European countries can be identified (see section 3.3).

3.2 Correlation Analysis

In this section, we present the correlation analysis of variables or components of data in the form of different component maps. The component maps are shown

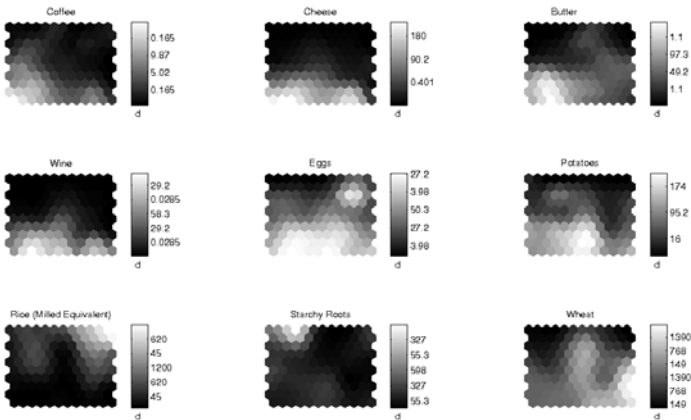


Fig. 4. Component distributions of common food products

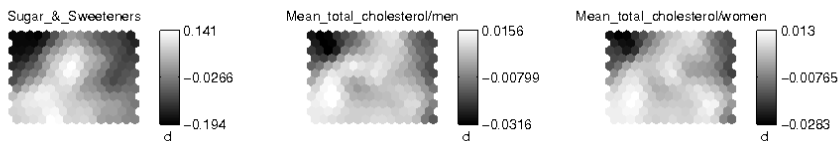


Fig. 5. Relationship between the intake of sugar and sweeteners (kcal/day) and the mean total cholesterol value (mg/dl) in men and women

on the gray scale where the dark shades represent low values and lighter shades represent high values of variables.

The correlation between fats and animal products (see Fig. 3) is significant since it shows strong correlations on both high and low values. This explains the fact that high consumption of fats in European countries is strongly correlated with high consumption of animal products. Similarly, low consumption of fats in some East African and South Asian countries is strongly correlated with low consumption of animal products.

Another component map of nine commonly used food constituents is shown in Fig 4. The map clearly shows strong correlations on relatively low values/ consumption of coffee, cheese, butter, wine, eggs, potatoes, starch and wheat products in various parts of the world.

Disease-Food correlation. Another important aspect of our research was to explore the connection between food and health. For this reason, we studied the

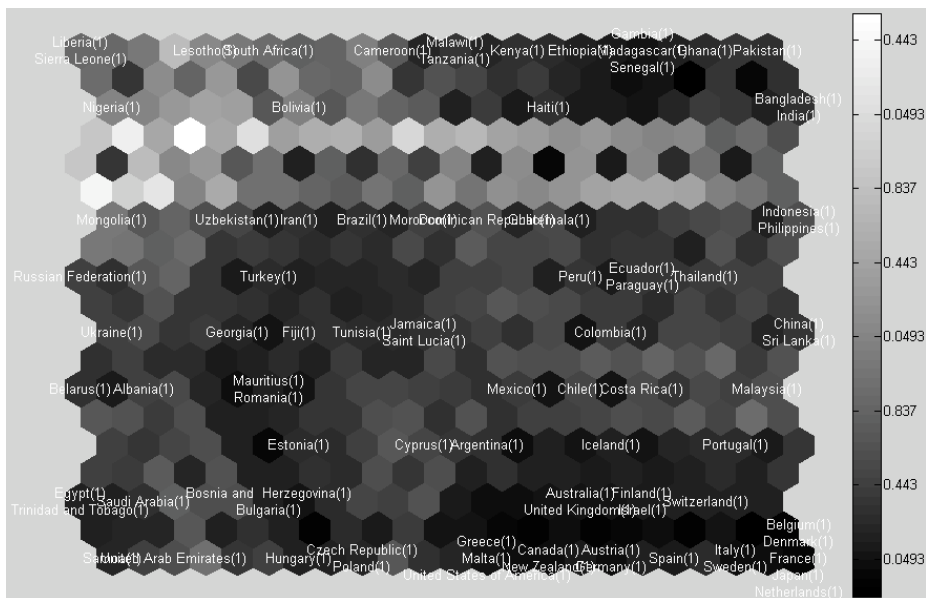


Fig. 6. SOM analysis of disease variables

relationships between some disease-related components and food constituents. An interesting finding, shown in figure 5, indicated a clear correlation between high consumption of sugar or sweeteners and a prevalence of cholesterol in men and women. This result appears to be an important finding from the health point of view (see also [11,12]).

3.3 Disease Analysis

In addition to the nutrition-based analysis, we have also analyzed the prevalence of different diseases in all 86 countries. The disease dataset includes prevalence of heart diseases in men and women, obesity in both genders, cholesterol and many others. The analysis leads us to conclude that countries, like in nutrition-based analysis in section 3.1, tend to fall in the same cluster in which their geographical neighbors reside. The SOM map of disease dataset (see Fig. 6) clearly shows some South Asian countries in the top right corner whereas much of the Europe is clustered in the bottom right.

4 Conclusion

The results obtained using SOM analysis provide a good understanding of the data by not only showing the underlying correlations within different food components but also between food components and diseases.

The study also shows the significance of machine learning techniques in order to infer useful information from different eating and drinking trends in a population. Moreover, deeper analyses performed on richer datasets may bring forth information that can be used for the societal and individual wellbeing.

Acknowledgments: We are grateful to EIT-ICT labs for funding Well-being Innovation Camp, which has greatly helped in conceiving the idea of this paper. We are also thankful to the organizers of the Camp as well as the participants.

References

1. WHO. Technical report series 916 (2003), <http://whqlibdoc.who.int/trs/>
2. Volkert, D.: Nutrition and lifestyle of the elderly in europe. Journal of Public Health 13, 56–61 (2005)
3. Canibais e Reis. Integrated nutrition, lifestyle and health database: epidemiological information for an improved understanding of diseases of civilization (2009), <http://www.canibaisereis.com/2009/03/21/nutrition-and-health-database/>
4. FAO: Statistical yearbook 2005-2006 - consumption, <http://www.fao.org/economic/ess/en/>
5. WHO. Global health atlas, <http://apps.who.int/globalatlas/dataQuery/>
6. British heart foundation statistics website, <http://www.heartstats.org/atozindex.asp?id=8>
7. Kohonen, T.: Self-organizing maps. Springer Series in Information Sciences (2001)

8. Díaz, I., Domínguez, M., Cuadrado, A.A., Fuertes, J.J.: A new approach to exploratory analysis of system dynamics using SOM. Applications to industrial processes. *Expert Systems with Applications* 34(4), 2953–2965 (2008)
9. Abonyi, J., Nemeth, S., Vincze, C., Arva, P.: Process analysis and product quality estimation by self-organizing maps with an application to polyethylene production. *Computers in Industry* 52(3), 221–234 (2003)
10. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J.: Self-organizing map in matlab: the som toolbox. In: *Proceedings of the Matlab DSP Conference 1999*, Espoo, Finland, November 1999, pp. 35–40 (1999)
11. Welsh, J.A., Sharma, A., Abramson, J.L., Vaccarino, V., Gillespie, C., Vos, M.B.: Caloric sweetener consumption and dyslipidemia among us adults. *Journal of American Medical Association* 303, 1490–1497 (2010)
12. Welsh, J.A., Sharma, A., Cunningham, S.A., Vos, M.B.: Consumption of added sugars and indicators of cardiovascular disease risk among us adolescents. *Circulation* 123, 249–257 (2011)