CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

Clustering Geo-Markets using Self-Organizing Maps: Application to a Business Venture in Natural Disaster Planning and Recovery

Salman Kimiagari
Benoit Montreuil

October 2013

CIRRELT-2013-62

Document de travail également publié par la Faculté des sciences de l'administration de l'Université Laval, sous le numéro FSA-2013-012.

Clustering Geo-Markets using Self-Organizing Maps: Application to a Business Venture in Natural Disaster Planning and Recevory†

Salman Kimiagari[1,2,*], Benoit Montreuil[1,3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Management Department, 2325, de la Terrasse, Université Laval, Québec, Canada G1V 0A6

[3] Operations and Decision Systems Department, 2325, de la Terrasse, Université Laval, Québec, Canada G1V 0A6

Abstract. This paper focuses on the clustering of geo-markets within an overall target territory to help ventures, businesses or organizations plan their growth. It proposes using self-organizing maps for such purposes due to their combined multi-criteria, spatial and visualization capabilities. It exposes the methodology for generating geo-market clusters based on self-organizing maps (SOM), notably exploiting spatial information databases. It illustrates and analyzes the approach through an application to a business venture targeting natural disaster planning and recovery related geo-markets.

Keywords. Geo-market clustering approach, self-organizing map, natural disaster planning and recovery.

Acknowledgements. This research has been supported by the Canada Research Chair in Enterprise Engineering.

† This article is submitted to the *Academy of Marketing Science*.

---

* Corresponding author: Salman.Kimiagari@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2013

1. Introduction

In today's fast-paced global economy, ventures, businesses and organizations with large-scale and vast-scope strategic intents increasingly have to be holistically designed in their early stages in order to capture their intended markets. Targeting and clustering markets become crucial for helping further design and plan the activities, resources, networks, financial flows and expected value creation of the intended business (Hwang , Thill 2007;Kim, Ahn,2008). This paper focuses on geographical market (geo-market) clustering in such contexts. Geo marketing consolidates and processes data based on geographic patterns of customers. Building a database from demographic and geographic information and integrating the information obtained in marketing mixes in order to design a sound marketing plan are amongst The geo marketing process can be used in any aspect of the marketing mix : the product, price, promotion, or place. In this study, we utilize the geo-targeting concept to exploit market selection in the market strategy. Geo-markets can refer to territories such as countries, regions or cities as pertinent. The key deliverable of the geo-market clustering decision process is a set of geo-market clusters organized in terms of similarity related to a set of key clustering variables pertinent for the specific business context. The clusters can be visualized and assessed by the decision makers.

The paper proposes using self-organizing maps (SOM) (Liu et al, .2012) for geo-market clustering due to their combined multi-criteria, spatial and visualization capabilities. It introduces the methodology for generating geo-market clusters based on self-organizing maps by exploiting spatial information databases.

The proposed approach is illustrated and analyzed through an application to natural disaster planning and recovery, using as a case study a business venture proposed by entrepreneurs, here called Global Relief Supply (GRS). The intended mission of GRS is to improve the readiness and reaction of worldwide client cities by helping them provide timely supply of effective tools, goods, drugs and food when facing a disaster. Its service offer is to complement local governments and relief agencies in preparing cities so as to economically maximize fast response and minimize the effects of natural disasters (Guha-Sapir and Santos, 2012). GRS aims to fulfill the needs of a huge potential market worldwide, illustrated in Figure 1 by the fast growth of natural disaster frequency and in Figure 2 by the growth in the number of people affected by such disasters. The intent of the GRS venture entrepreneurs is to serve a significant percentage of all cities worldwide within the next fifteen years. The paper examines the clustering process of GRS venture's geo-markets within the U.S.A. territory, a key subset of their intended worldwide terrain, aimed to sustain the entrepreneurs in planning their market deployment.

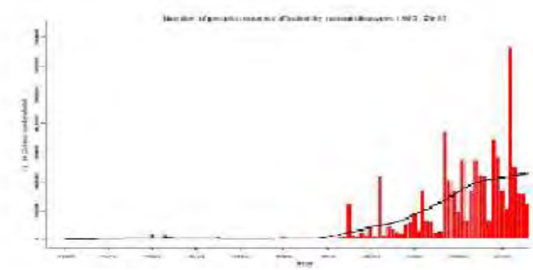

Fig1: Natural disasters (1900-2010) Fig2: Number of people affected by natural disasters (1900-2010) - CRED database

This paper is structured as follows. Section two provides a brief literature review on market segmentation and quantitative approaches for market clustering. This literature review justifies the appropriateness of using self-organizing maps for geo-market clustering purposes. Section three describes the generic SOM clustering method. Section four presents the proposed geo-market clustering methodology, illustrating and analyzing its application to the GRS case. Finally, section five provides conclusive remarks and avenues for further research.

2. Literature review

Market clustering originated as a result of market segmentation that was first defined by Smith (1956) to provide a conceptual view of heterogeneous markets (Liu et al., 2012). The strategic intent underlying market segmentation is aiming toward developing target marketing programs and products customized to each segment, as well as sustaining market segment selection decisions and market deployment plans (Kotler, 2009). Market characteristics targeted for segmentation purposes can be demographics or socioeconomic factors, geographic locations or product related behavioral characteristics such as purchase or consumption behaviors (Hanafizadeh and Mirzazadeh, 2011). Clustering approaches aim to classify market subsets and situate them into specific differentiated groups. These approaches aid to identify the consumers in each subset, and put them into homogenous groups, in order to facilitate marketing plan provision (Hanafizadeh and Mirzazadeh, 2011). Numerous methods have been proposed for market clustering, as highlighted in Table 1. For each method, Table 1 provides the advantages and limitations reported in the literature, with the pertinent sources of reference.

Table1: Market clustering methods

Methods	Advantage(S)	Limiataions(S)
K-means clustering(Cheung,2003; Kuo et al.,2004)	1.Robust (Rujasiri and Chomtee ,2009) 2.Efficient in terms of executive time(Augen,2004) 3.High accuracy (Guha-Sapir and Santos ,2012)	1.May fall in local optima(Liu et al.,2012) 2.Depends on the initial cluster centers(Augen ,2004; Rujasiri.and Chomtee ,2009; Liu et al.,2012) 3.Lack of visualisation (Hanafizadeh, Mirzazadeh 2011) 4.Need to have predefined cluster numbers (Augen ,2004; Rujasiri and Chomtee 2009;Hanafizadeh, Mirzazadeh 2011)
SOM(Hanafizadeh, Mirzazadeh 2011)	1.Reducing the amount of data 2.Projecting the data non-linearly onto a lower-dimensional subspace(Augen ,2004) 3. Visualisation benefit (Hanafizadeh, Mirzazadeh 2011) 4.Detecting non-linier correlation between variables (Venugopal, and Baets,1994) 5.Robust(Venugopal, and Baets,1994) 6.Not significantly affected by missing data(Venugopal, and Baets,1994) 7.Does not require any prior assumption about the underlying distribution of the data(Venugopal, and Baets,1994) 8. High accuracy (Guha-Sapir and Santos ,2012)	1.Setting initial weights(Kim , Ahn,2008) 2.Setting stop condition(Kim , Ahn,2008)
Extended SOM (Kiang , Hu, Fisher,2007)	1.Same as SOM 2.Non overlapping clustering(Hanafizadeh, Mirzazadeh 2011) 3.Independent from sample size(Hanafizadeh, Mirzazadeh 2011)	1.Same as SOM 2.Lack of visualisation (Hanafizadeh, Mirzazadeh 2011)
Fuzzy clustering (Hwang , Thill ,2007).	1. Not significantly affected by missing data 2. Does not require any prior assumption about the underlying distribution of the data 3.Appropriate when the sample size is large(Guha-Sapir and Santos ,2012)	1. Lack of visualisation (Hanafizadeh, Mirzazadeh 2011) 2.Overalpping clustering (Hanafizadeh, Mirzazadeh 2011) 3.Hard to decide the number of clusters 4.Chossing the initial cluster centroids
SOM and GA K-means(Kritboonyalai and Avatchanakorn ,2003; Kuo et al,2006)	1.Same as SOM 2.Non – overlapping clustering (Augen,2004;Hanafizadeh, Mirzazadeh 2011)	1.Same as SOM 2.Lack of visualisation (Hanafizadeh, Mirzazadeh 2011)
GA K-means(Kim,Ahn,2008)	1.Using GA to identify initial seed (Augen,2004; Hanafizadeh, Mirzazadeh 2011) 2. Non – overlapping clustering(Augen,2004; Hanafizadeh, Mirzazadeh 2011)	1.weak visualization (Hanafizadeh, Mirzazadeh 2011)
Cluster wise regression (Liu et al.,2012)	1.Non – overlapping clustering	1. Lack of visualisation 2.Not opmtimized with in segment homogeneity(Liu et al.,2012)
Automatic interaction detection (Vavrik, & Mazanec ,1990)	1.Non – overlapping clustering(Augen,2004) 2. Appropriate when the sample size is small(Liu et al.,2012)	1.Not opmtimized with in segment homogeneity(Liu et al.,2012) 2. Lack of visualisation
Simulated annealing(Kuo et al,. 2004)	1.Avoid local optima	1.Show in reaching the optimal solution(Jarboui et al., 2007) 2. Lack of visualisation
Hierarchical clustering (Guha-Sapir and Santos ,2012)	1.Non – overlapping clustering 2.Appropriate when the sample size is small(Liu et al.,2012)	1.Lack of visualisation(Liu et al.,2012) 2.Low accuracy(Liu et al.,2012)

As revealed in Table 1 reveals, the self-organizing map (SOM) method is the most appropriate method for geo-market clustering, due to the combination of its capability of taking into consideration the location of the markets and its conceptual and geographical visualization benefits. In general, the SOM method has the ability to display multi-dimensional space in two-dimension space (Fish, Ruby, 2009). Moreover, it can demonstrate the correlation between segmentation variables and their effects on each other. Through cluster analysis based on SOM, an analyst can group data in d-dimensional space in order to maximize the similarity within the clusters and minimize the difference between two different clusters. The SOM method has been widely used in marketing, visualization and market studies (Oja,Kaski, Kohonen,2002; Pöllä, Honkela,and Kohonen,2007). Some of these studies include application of SOM in direct marketing (Curry et al., 2003), visualization of marketing data (Lisboa and Patel, 2004) and application in data mining planning (Hwang, Thill, 2007). SOM has been selected in this paper for purposes of geo-market clustering. The following section presents in more depth the SOM based clustering method.

3. Self-organizing map based clustering

A Self-Organizing Map consists of a grid of artificial neurons. Each neuron has a multi-dimensional weight vector. Each dimension corresponds to an informational variable. The neurons are trained to adjust their weight using a method well documented in the literature (Kritboonyalai. and Avatchanakorn, 2003; Kiang, Hu, Fisher, 2007; Hanafizadeh and Mirzazadeh, 2011), using the multi-variable data points as input. The final Self-Organizing Map is generally displayed using a U-matrix and variable matrices such as illustrated in Figure 3.

In the SOM matrices, colors are used to display the distance between neighboring neurons in terms of their weight vectors. Colors range from dark blue to hot red as the distance increases. In the U-matrix, the coloring is based on the overall multidimensional distance, while in a variable matrix it is based on the one-dimensional distance related to the selected variable. The variable matrices of Figure 3 have distinctive coloring patterns, except the matrices related to the strongly correlated GDP (Gross Domestic Product) and Population variables that are almost identical.

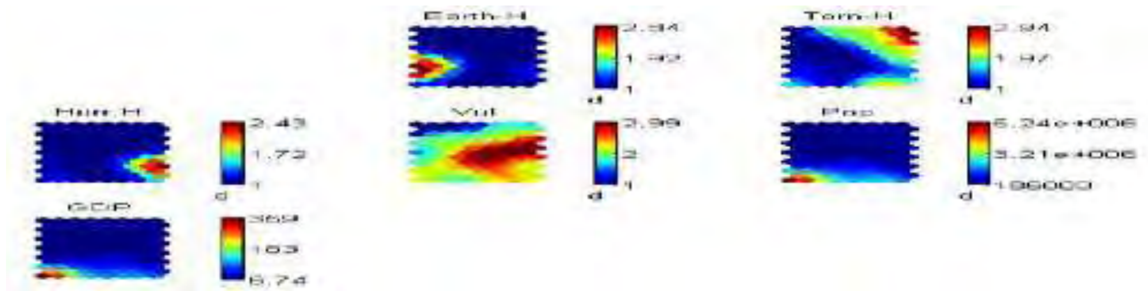

Fig3: Variable Matrix Views of a Self-Organizing Map(Earth-H, Torn-H ,Hurr-H, Vul, Pop, GDP stand for Earthquake hazard, Tornado hazard, Hurricane hazard, Vulnerability, Population and Gross Domestic Product respectively)

In the U-matrix of Figure 3, there are several dark blue zones corresponding to sets of neurons near to each other, surrounded by lighter and hotter color zones acting as high-distance separations (Pires, Lobo, Bação, 2007). Conceptually, dark blue zones correspond to neurons that are prone to be clustered together based on the multidimensional information they were trained with. Clustering neurons in self-organized maps is based on algorithms that exploit this phenomenon. Figure 4 displays a set of clusters obtained when requesting that five clusters be *generated.* Note that some neurons end up in no clusters due to their high distance from their neighbours.

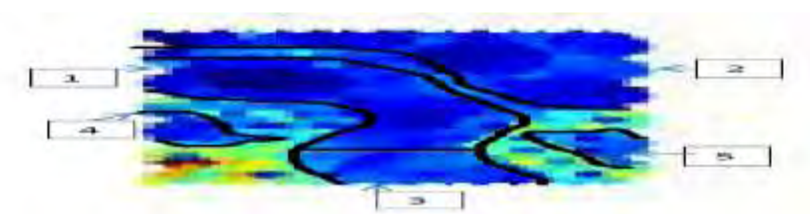

Fig4: Set of five Neuronal Clusters in a Self-Organized Map, displayed on a U-matrix

The original training data points, and new ones as pertinent, can be located on the SOM matrices by being assigned to their minimal-multidimensional-distance neuron on the map. This allows transposing the neuronal clusters into data point clusters.

4. Self-organizing map based geo-market clustering

The proposed method for geo-market clustering based on self-organizing maps is synthesized in Table 2. It exploits the generic SOM clustering method described in section 3, where the geo-markets correspond to the data points, described using a set of segregating variables.

Hereafter, the key steps of the methodology are described using the Global Relief Supply venture case for illustration purposes.

Table 2: Proposed geo-market clustering methodology

Phase 1: Strategic intent definition
1. Define the strategic business intent;
Phase 2: Creating segregating data sets and variables
2. Define the set of geo-markets; **3. Explore the set of databases pertinent to the geo-markets and the strategic business intent;** **4. Define a set of geo-market segregating variables computable from the available databases;** **5. Add to the set some variables defining the location of each geo-market;** **6. Assess the value of each variable for each geo-market;**
Phase 3: Generating a set of geo-market clusters based on self-organizing map
7. Define a neuronal grid structurally representative of the overall territory; **8. Train the self-organizing map using the geo-markets as input points;** **9. Analyze the U-matrix and its implications on market segmentation as to decide on the number of clusters;** **10. Generate neuronal clusters in the Self-Organizing Map;** **11. Transpose the neuronal clusters into geo-market clusters;**
Phase 4: Result interpretation & evaluation
12. Analyze the resulting set of geo-market clusters, saving it if deemed satisfying; **13. Evaluation of results based on cluster validation metrics** **14. Return to one of the first nine steps above as pertinent, unless stopping because the current set of satisfying cluster sets is deemed sufficient for supporting geo-market clustering decision-making.**

4.1 Phase 1: Defining the strategic business intent

The mission of GRS is to improve the readiness and reaction of worldwide client cities by enabling fast and reliable supply availability of effective tools and supplies before, during and after a natural disaster. To this end, the entrepreneurs at the core of the GRS are planning to achieve this mission via an appropriate business model. The entrepreneurs are developing a market deployment plan for this global venture. At the culmination of this 15-year plan, GRS intends to be serving cities covering most of the disaster-prone countries. As an important phase in market deployment plan, segmentation helps GRS identify market demands and design better strategies for servicing. Considering the different variables which affect target market selection, GRS needs a clustering approach to project the data to lower dimensions. This approach should provide robust results with high accuracy and benefit from visualization and efficiency in term of execution time. Moreover, the utilized technique needs to be independent from sample size, the distribution of the data and missing data and pre-defined clusters number. Hence, SOM is a convenient approach for this purpose.

4.2 Phase 2: Creating segregating data sets and variables

The market clustering is intended to be data driven. So a key activity is the creation of a reference data set to sustain the study, and the selection of a set of segregating variables to be used for the clustering analysis. The overall process and the calculation method as well as data set are detailed in the following subsections.

4.2.1. Define the set of geo-markets

The first step in data set preparation is selecting the target geo markets. For this study, we focused on metropolitan cities in the USA. The GRS venture intends to target cities with over than 100,000 inhabitants. This has led to consider 365 US metropolitan cities as constituting the set of geo markets for this study. Each one was geocoded with a specific longitude and latitude.

4.2.2. Explore the set of relevant databases

Database exploration is essential step in market clustering. This step can be performed based on the case under study and market selection factors such as socioeconomic factors or demand side factors. For this study we used different databases. These databases include U.S. Census database (2007), Seismic-Hazard Maps for the Conterminous United States (2008), Severe Weather Database Files[1] and National hurricane center data bases[2].

4.2.3. Define a set of geo-market segregating variables computable from the available databases

For finding the market selection factors for GRS, we need to review the literature related to natural disasters. According to Skidmore and Toya (2002), variables that determine deaths from natural disasters are population, land area and disaster type. In this study, we considered population, GDP, hazard and vulnerability as the effective variables in market segmentation. Population and GDP are important for estimation of contract value. Moreover, these factors influence the vulnerability of target markets. The hazard and vulnerability are the parameters for calculation of target market risk. The populations and GDP were extracted from the U.S. 2009 Census database. For hazard occurrence calculation, we used Seismic-Hazard Maps for the Conterminous United States, 2008, Severe Weather Database Files and National hurricane center data bases. For vulnerability calculation, we used Borden's (2007) study. In order to prepare the data for hazard and vulnerability for our geo market data set, we used procedures which are described in detail in 4.2.5.

4.2.4. Add to the set some variables defining the location of each geo-market

Jones and Pearce (1999) explain the importance of geography in today's marketing activities according to four aspects: demand, supply, logistic chains and the nature of some economic activities. For example, geography affects demand as perceived through revenue per household,

[1] http://www.spc.noaa.gov/

[2] http://www.nhc.noaa.gov/

budgetary spending structures and local lifestyles (Cliquet, 2006). In this study we exploited longitude and latitude of each target market their geo-positioning variables.

4.2.5. Assess the value of each variable for each geo-market

Based on the major natural disasters which have occurred in US states, we considered three types of natural disasters: earthquake, tornado and hurricane. According to Zhang et al. (2006), "Hazard represents an extreme natural event that adversely affects human life, property or activity and to the extent of causing a disaster with a certain degree of probability and severity". Borden et al.,(2007) defined a natural disaster hazard as the potential threat from an environmental process, such as a hurricane, tornado, or earthquake. For each city in the geo-market set, we extracted the information of these three types of hazards as follows.

For earthquakes, we used Seismic-Hazard Maps for the Conterminous United States, 2008. These summarize the available quantitative information about seismic ground motion hazard for the conterminous United States from geological and geophysical sources (Petersen et al., 2011).

For tornadoes, we used Severe Weather Database Files. The database provides files for tornados, hail, and damaging wind data as compiled in Storm Data. We utilized data from 1980 to 2010. These tables include the tornado's name, the date and the time of incidence, the starting latitude, starting longitude, ending latitude and ending longitude, the wind speed, and hail size or the intensity of tornado based on the Fujita Tornado Scale. We applied an algorithm in order to calculate the hazard occurrence. We searched the database to find cases occurring in the same latitude and longitude as cities in our reference dataset, while considering a radius effect for each incident proportional to its intensity. Hence for each city, we considered several types of incidents over 30 years. We took into account the frequency of each type and then calculated the sum of frequencies to estimate the tornado hazard occurrence.

We applied the same approach to determinate hurricane hazard. For this hazard type, we used Database Files from the U.S. National hurricane center and Saffir-Simpson Hurricane Wind Scale.

Figure 5 shows a map of natural disaster history in the U.S.A. According to Zhang et al. (2006), "Vulnerability denotes the degree of resistance of the asset & population against hazard and it decides the loss degree caused by hazard". Borden et al. (2007) considered vulnerability as the susceptibility to harm from the risk posed by hazard events at a particular location and the potential for social disruption. The authors studied the vulnerability of the 132 urban areas in U.S using three vulnerability indices: social, built environment, and hazard impact. We used this study in order to calculate the vulnerability of cities to natural disasters. We utilized social and built environment vulnerability indices of this study and calculated the results for our dataset through interpolation. The vulnerability to disaster (V_d) is calculated from the sum of the social vulnerability$(V_S)$and the built environmental vulnerability(V_b) $(\text{Borden et al.}, 2007)$:

$$V_d = V_S + V_b \tag{1}$$

Figure5: Natural disaster history on map in USA[3]

[3] http://earthquake.usgs.gov/, http://www.nhc.noaa.gov/,http://www.spc.noaa.gov/

4.3 Phase 3: Generating a set of geo-market clusters based on self-organizing map

The application of SOM in our case study involves defining a grid of neurons corresponding to the set of geo markets. For this case study, a network was developed and trained in the MATLAB 7 environment as described in the following subsections.

4.3.1 Define a neuronal grid

The first step of applying SOM is creating an appropriate neural grid for training and analysis. In this study, a 20x20 hexagonal SOM was trained based on the SOM Toolbox 2.0 developed by Vesanto (2000).

4.3.2 Train the self-organizing map

The neurons were trained to adjust their weight based on training algorithms. The SOM Toolbox 2.0 provides two training algorithms including sequential and batch algorithms detailed in Vesanto (2000). We experimented with both algorithms and, considering the quantization error, we utilized batch algorithm for our case study.

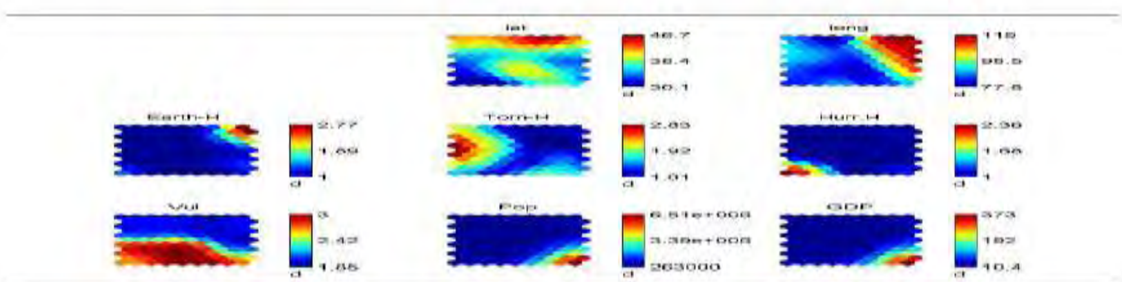

Fig 6: Variables Planes of self-organizing map

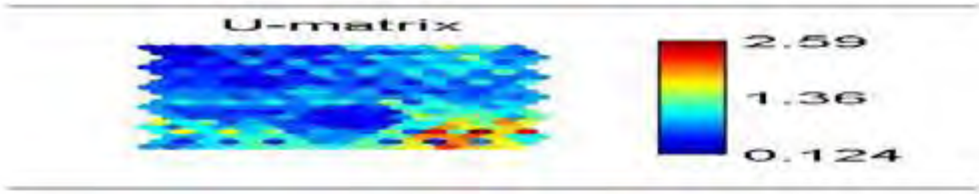

Fig 7: U-matrix of SOM for U.S natural disaster market of GRS venture

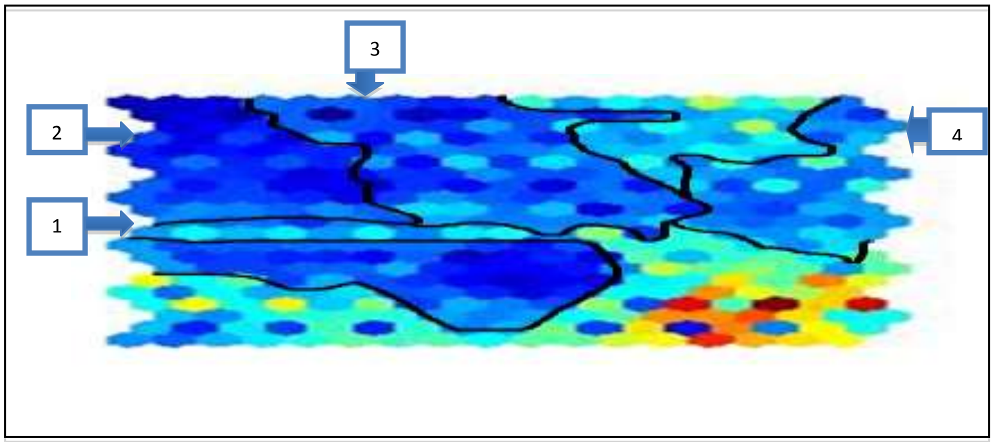

Fig 8: U-matrix clusters of SOM for U.S natural disaster market

4.3.3 Analyze the U-matrix and component planes

Figure 6 presents the variable planes using a 20x20 hexagonal SOM for GRS. High values are associated with dark red colors and low values with dark blue colors. Figure 7 presents the U-matrix resulting from the SOM process. Then Figure 8 overlays on this U-matrix the proposed geo-market clusters, as described further in the next sub-section.

4.3.4 Generate neuronal clusters in the Self-Organizing Map

As shown in Figure 8, four specific clusters have been identified approximately from the U-matrix. This estimation is based on the color spectrums. For clustering, we need to find the border points. These are located at the margin of densely distributed data, and they represent a subset of the population that possibly belongs to two or more classes (Xia et al., 2006). Each cluster of the map links to a segment in the market. There are different regions of dark blue that correspond to low values in the U-*matrix,* and therefore to clusters in the data. These regions are disjointed by lighter colours, which correspond to separations between the clusters

When examining the clusters identified in Figure 8, it can be recognized that each cluster has special characteristics based on variables are used for clustering. Cluster 1 groups cities with earthquake hazards. Cluster 2 groups cities with medium tornado hazards. In section 3 are grouped cities with both tornado and hurricane hazard. These cities are more vulnerable and more populated. Section 4 groups cities with a high risk of hurricane.

4.3.5 Transpose the neuronal clusters into geo-market clusters

In order to evaluate which of the clustering results are best, ranging from three to six clusters, we used two different indices. The basis for this evaluation is measuring high similarity within a cluster and low similarity between clusters. We utilized the Davies–Bouldin index and the Dunn index. The Davies–Bouldin(1979) index can be calculated as follows:

$$DB = \frac{1}{n}\sum_{i=j}^{n} \max_{i \neq j}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \qquad (2)$$

Where n is the number of clusters, σ_x is the average distance of all elements in cluster x to cluster center c_x, and $d(c_i, c_j)$ is the distance between center of clusters $c_i$and c_j . The results with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) are preferred. Hence the smallest Davies–Bouldin index indicates the best result.The Dunn index (Dunn, 1974) is calculated as follows:

$$D = \min_{1 \leq i \leq n}\left\{\min_{1 \leq j \leq n, i \neq j}\left\{\frac{d(i,j)}{max_{1\leq k\leq n} d`(k)}\right\}\right\} \qquad (3)$$

Where $d(i,j)$ represents the distance between clusters i and j, and $d`(k)$measures the intra-cluster distance of cluster k. Since internal criteria look for clusters with high intra-cluster similarity and low inter-cluster similarity, the results with a high Dunn index are more desirable. The results are shown in table 3. The best result based on these two indices correspond to four clusters with a minimum Davies–Bouldin index value and a high value for the Dunn index.

Table 3: Cluster evaluation

Number of clusters	3	4	5	6
Davies–Bouldin Index	14,8	5,4	6,1	11,8
Dunn Index	0,0032	0,0130	0,0062	0,0035

The next step is to transpose the neural clusters into a clustered geo-market set, resulting in the four clusters depicted on the USA map as shown in Figure 9. Cluster 1 corresponds to cities depicted in pink on the USA map. The earthquake-prone cities of cluster 1 are located in the Pacific and Mountain West of the USA, except Wyoming, Colorado and New Mexico that are in cluster 2. Note that Alaskan cities have been included in cluster 1. In the western central zone of the USA, cluster 2 indeed includes green-colored cities with medium tornado hazard level, ranging from North Dakota in the north down to New Mexico, Texas and Louisiana in the south.

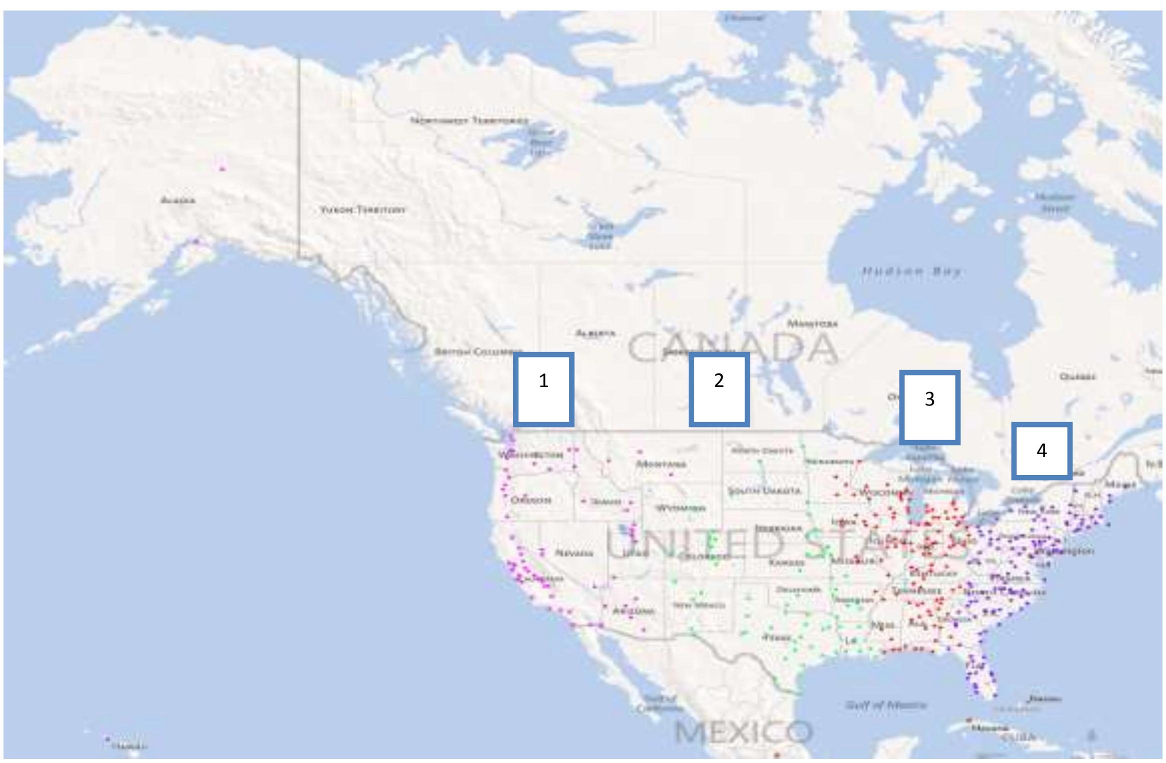

Fig 9: Four clusters on US map

The red-colored tornado+ hurricane-prone cities of cluster 3 are located in the eastern central zone of the USA, ranging from Minnesota, Wisconsin and Michigan in the North to Mississippi, Alabama and the northwestern tip of Florida. Section 4 includes the purple-colored cities with high hurricane vulnerability, located in the eastern zone of the USA.

The information, as visualized in Figure 9, provides managers with an overview of the market clusters that can be interpreted in a way such as illustrated above, facilitating the target market selection procedure.

4.4 Phase 4: Result interpretation & Evaluation

In the result interpretation and evaluation step of methodology, we need to analyze the results of the geo-market clustering and save them if deemed satisfying. If the results are inappropriate, we return to previous steps (e.g. change variables being considered) in order to obtain satisfying results. Focusing on validating the four-cluster solution, the U-matrix and component planes obtained in Figure 6 and 7 and the SOM clustering depicted in Figure 8 are deemed to be satisfying as their transposition into the geo-market clusters on the USA map of Figure 9 relate nicely with the history maps and USGS data, making it easy to justify to managers. In general, this step can be performed by experts who are familiar with the business context and the strategic intent of the business venture. For example removing the location information from input variables can change the results completely (See the figure 10). Moreover, the U-matrix and component planes will be different if GDP and population are removed (See figure 11, 12, 13 and 14). In these cases, quantization error and topographic error for SOM structure are increased and deemed unacceptable. Hence in general, with these procedures, satisfying results are for example achieved by iteratively adding or removing variables and interpreting the results, until these are deemed adequate.

Fig 10: Four clusters on US map without location variables

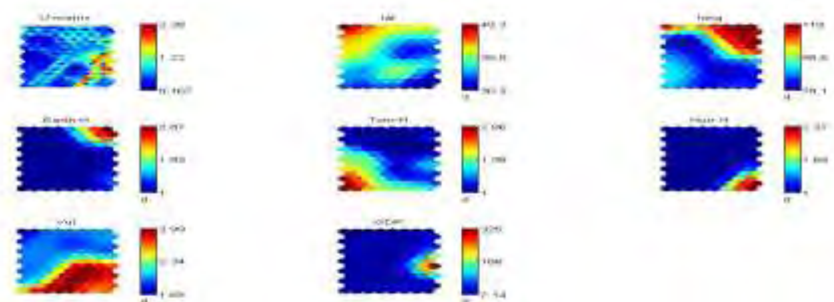

Fig 11: U-matrix and component planes without population variable

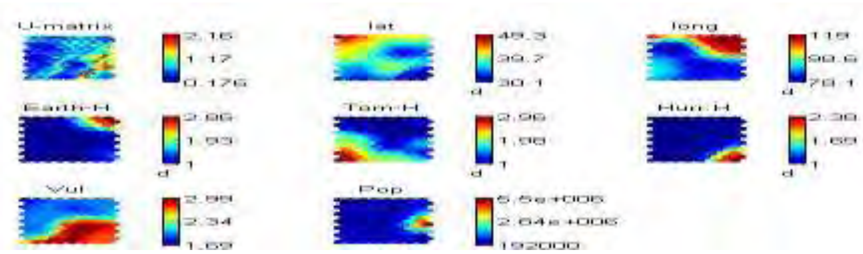

Fig 12: U-matrix and component planes without GDP variable

Fig 13: Four clusters on US map without population

Fig 14: Four clusters on US map without GDP

Beyond the cluster-by-cluster interpretation spelled out in the previous section, the following insights can be gained from the overall results:

- ✓ In the U.S.A., the cities with a high earthquake hazard do not have high tornado or hurricane hazards, as can be reviewed using Natural disaster history maps and USGS data. This justifies creating an earthquake-prone cluster of geo-markets.
- ✓ The number of cities affected by a medium-to-high tornado hazard is higher than the number of cities with earthquake and hurricane hazards (Figure 6), as can be verified by cross referencing tornado report data between 1950 and 2004 (http://www.spc.noaa.gov/)
- ✓ The cities with high and medium hurricane hazards are more populated than the cities with high tornado hazard. These cities are shown in Green, yellow and light blue in corresponding variable Planes map(see Figure 6)

- ✓ Population and GDP have the same color pattern in Figure 6 and it shows that there is a positive correlation between these two variables.
- ✓ The cities with high GDP and population are more vulnerable. These cities are shown in red and dark areas in Component Planes map of vulnerability, GDP and population (see Figure 6).
- ✓ The cities with high tornado and hurricane hazards are also in the areas with high vulnerability. These matters can be proven by social and build vulnerability data from Borden's (2007) study as well as the tornado and hurricane history maps.

4. Conclusion

Market segmentation, clustering and visualization techniques simultaneously fulfill the needs of marketing strategies particularly in the early stages of the business venture creation. These techniques can be considered as effective tools in developing market deployment plans and mapping the buyer's propensities in global business activities. In this study, we applied SOM for clustering and visualization, using as an illustrative case the GRS venture project that intends to improve the readiness and reaction of worldwide client cities in natural disaster cases. The essential variables of this market have been identified and a data set was prepared to train SOM. U-Matrix and variable maps have been provided and analyzed. The investigation of the adaptation of the proposed methodology for born global cases provides an exciting avenue for further research, notably dealing with international regions where data availability is sparser. Moreover, there is a rich research avenue in investigating how to better incorporate the information related to the supply chain and logistic networks into the geo-market clustering approach. Finally, there is strong potential for research on the joint application of this geo-market clustering approach with other quantitative approaches such as optimization techniques in order to develop time-phased market deployment plans guiding business and ventures through their growth.

Acknowledgements

This research has been supported by the Canada Research Chair in Enterprise Engineering.

References:

1. Augen, J. (2004). Bioinformatics in the Post-Genomic Era: Genome, Transcriptome, Proteome, and Information-Based Medicine by Addison-Wesley Professional.Edition: 1st ISBN-10: 0-321-17386-4.
2. Borden, K.A., Schmidtlein, M.C., Emrich, C.T, Piegorsch, W.W., & Cutter.S.L. (2007). Vulnerability of U.S. Cities to Environmental Hazards Journal of Homeland Security and Emergency Management. 4(2), ISSN (Online) 1547-7355, DOI: 10.2202/1547-7355.1279, June.
3. Cliquet, G. (2006). Spatial Marketing. Geomarketing: Methods and Strategies in Spatial Marketing, 11-25. ISTE Ltd, ISBN-13: 978-1-905209-07-1
4. Curry, B., Davies, F., Evans, M., Moutinho, L., & Phillips, P. A. (2003). The Kohonen self-organising map as an alternative to cluster analysis: an application to direct marketing. International Journal of Market Research, 45(2), 191-211.
5. Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2), 224-227.
6. Dunn, J. C. (1973). A Fuzzy Relative of the ISO DATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3 (3): 32–57.
7. Fish. K, & Ruby, P (2009). An artificial intelligence foreign market screening method for small, businesses. International Journal of Entrepreneurship 13, 65-81.
8. Guha-Sapir.D. & Santos .I (2012), the Economic Impacts of Natural Disasters, OUP USA 978-0-19-984193-6. Oxford University Press.
9. Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method–ADSL market of a telecommunication company. Expert Systems with Applications, 38(1), 198-205.
10. http://earthquake.usgs.gov/
11. http://www.nhc.noaa.gov/
12. http://www.spc.noaa.gov/
13. Hwang, S., & Thill, J.C. (2007). Using fuzzy clustering methods for delineating urban housing submarkets. In Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems (p. 14). ACM. , November.

14. Jarboui, B., Cheikh, M., Siarry, P., & Rebai, A. (2007). Combinatorial particle swarm optimization (CPSO) for partitional clustering problem. Applied Mathematics and Computation, 192(2), 337-345.
15. Jones, K., & Pearce, M. (1999). The Geography of markets: Spatial analysis for retailers. Ivey Business Journal, 63(3), 66-70.
16. Kiang, M. Y., Hu, M. Y., & Fisher, D. M. (2007). The effect of sample size on the extended self-organizing map network—A market segmentation application. Computational Statistics & Data Analysis, 51(12), 5940-5948.
17. Kim K-j, & Ahn.H. (2008). A recommender system using GA K-means clustering in an online shopping market. Expert Systems with Applications, 34(2), 1200–1209.
18. Kotler, P. (2009). Marketing management. Pearson Education India.
19. Kritboonyalai.W, and Avatchanakorn.V. (2003). Data mining for multi-level marketing planning in e-commerce. In H. Leung, editor, Proceedings of the Seventh IASTED International Conference on Artificial Intelligence and Soft Computing, pages 307–312. ACTA Press, Anaheim, CA, USA.
20. Kuo R. J., Chang K., Chien S. Y. (2004). Integration of self-organizing feature maps and genetic-algorithm-based clustering method for market segmentation. Journal of Organizational Computing and Electronic Commerce, 14(1), 43–60.
21. Kuo R.J., An Y.L., Wang H.S., Chung W.J. (2006). Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation. Expert Systems with Applications, 30, 313–324.
22. Lisboa, P.J.G & Patel, S. (2004). Cluster-based visualisation of marketing data. In Intelligent Data Engineering and Automated Learning Ideal, Proceedings, Lecture Notes in Computer Science, pages 552–558. Springer Berlin Heidelberg.
23. Liu, Y, Kiang, M., Brusco, M. (2012). A unified framework for market segmentation and its applications ,Expert Systems with Applications, 39, (11), 1 September, 10292–10302.
24. Oja, M., Kaski, S., & Kohonen, T. (2002). Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum Helsinki University of Technology, Neural Networks Research Centre Neural Computing Surveys 3, 1-156, http: //www.icsi.berkeley.edu/ jagota/NCS.
25. Petersen, M.D., Frankel, A.D., Harmsen, S.C., Mueller, C.S., Haller, K.M., Wheeler, R.L, Wesson R.L., Yuehua, Zeng, Oliver S. Boyd, David M. Perkins, Nicolas Luco, Edward H. Field, Wills, C.J., & Kenneth S. (2011). Rukstales Seismic-Hazard Maps for the Conterminous United States, 2008 Prepared in cooperation with the California Geological Survey.

26. Pires, F.J, Lobo, V, & Bação, F, (2007). Insights on the interpretation of SOM and U-Matrices with an example clustering based in oceanographic data 10th AGILE International Conference on Geographic Information Science Aalborg University, Denmark
27. Pöllä, M., Honkela, T., & Kohonen, T. (2007). Bibliography of self-organizing map (SOM) papers: 2002-2005 addendum. Neural Computing Surveys.
28. Rujasiri, P., & Chomtee, B. (2009). Comparison of Clustering Techniques for Cluster Analysis. Kasetsart J. (Nat. Sci.), 43, 378-388.
29. Skidmore, M., & Toya, H. (2002). Do natural disasters promote long run growth? Economic Inquiry, 40(4), 664-687.
30. Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. The Journal of Marketing, 21(1), 3-8.
31. Vavrik, U. & J. Mazanec (1990). A Priori and Posteriori Market Segmentation: Tailoring Automatic Interaction Detection and Cluster Analysis for Tourism Marketing. Caliers du Tourisnie, Serie C, No.62 Aix-en-Provence: Centre des Hautes Etudes Touristiques
32. Venugopal, V., & Baets, W. (1994). Neural networks and statistical techniques in marketing research: A conceptual comparison. Marketing Intelligence and Planning, 12(7), 30–38.
33. Vesanto, J , Himberg, J., Alhoniemi, E. , & Parhankangas, J., (2000). SOM Toolbox for Matlab 5. SOM toolbox team, Helsinki University of technology report A57 April.
34. Xia, C., Hsu, W., Lee, M. L., & Ooi, B. C. (2006). BORDER: efficient computation of boundary points. Knowledge and Data Engineering, IEEE Transactions on, 18(3), 289-303.
35. Cheung, Y. M. (2003). K*-Means: A new generalized k means clustering algorithm. Pattern Recognition Letters, 24(15), 2883-2893.
36. Zhang, J., Okada, N., & Tatano, H. (2006). Integrated natural disaster risk management: comprehensive and integrated model and Chinese strategy choice. Ziran Zaihai Xuebao/ Journal of Natural Disasters, 15(1), 29-37.