Construction of Autosimilar Electoral Units using Self-Organizing Maps

Alberto García Aguilar[1], José Carlos Méndez de la Torre[1], and Leopoldo Trueba Vázquez[1]

Universidad Autónoma de Zacatecas, Unidad Académica de Matemáticas,
Paseo a la Bufa esquina con Calzada Solidaridad,
Zacatecas, Zac., México, C.P. 98068

Abstract. In this paper we present a method of constructing autosimilar electoral units, using the Self-Organizing Maps (SOM) for detecting those electoral sections that present similar results to the complete universe. A particular case is presented that show the benefits of the application, in the capital of Zacatecas state; this method can be applied to electoral surveys and preliminar results of a given election.

1 Introduction

The obtention of satisfactory results in surveys for measure the vote intention and in the preliminary results programs is not easy. In first place, the electoral universes generally present a big cuantity of electoral units for recolecting the votes and in adition, some universes present high geographic and population dispersion. Another of the principal problems is derived of the fact that the electoral units do not have, by one side, an homogeneus behavior in the intention of the vote emission that allows using the method of standar stratified sampling.

Taking into acount various elements that are presented in the electoral units leads us to a problem with the design of the methodology that is going to be used in the selection of the units like representative units. It is usual to use the random selection acording the number of electors, however, this present the problem that sampling frame is based in the cuantity of electors and not in the historical results of the election. The statistic orthodoxy implies that the classical methods are good when sampling frame is good, and this good frame is precisely the electoral behavior and not the cuantity of people that make up the electoral units of the universe. For example, a universe with high electoral dispersion requires of selecting a big number of elements by each electoral unit and this take us to a bigger size and cost of the sample.

The principal target has to be selecting those electoral units that present a similar behavior to the real historical development of the trends of results of the vote orientation and the citizen's opinions. This is what we define like autosimilar electoral units, this denomination has its origin in the mathematical concept of autosimilarity, used in fractals and that is the property that guarantee the reproduction of the geometric structure in a different scale.

Advances in Artificial Intelligence: Algorithms and Applications
Research in Computing Science 40, 2008, pp. 39-48

An electoral unit is generally composed by one or some boxes that are the basic structure in the elections, so then one of the first steps for studing the electoral results should be analyse a set of the electoral units, but, Wich units has to be studied? It is not desirable to study all of them, a procceding for clostering the units should be done to make the study easier. In this clasification, those units that are closer or that are similar to the behaviour of the whole electoral universe are very important.

Extrapolating the mathematical concept of autosimilarity to the electoral results implies a change of the scale that can be made using the percentages obtained by the different political parties in the unit and in the total of all the units. Then, the problem is reduced to find the units or the sections that have very similar historical percentages of elections to the complete universe under study.

2 The Principle of the SOM

Acording to [1], in the documentation included of the software "SOM_PAK" and that is reproduced in this section: There exist many version of the SOM [2]. The basic philosophy, however, is very simple and already effective as such.

The SOM here defines a mapping from the input data space $\mathbb{R}^n$ onto a regular two-dimensional array of nodes. With every node i, a parametric reference vector, $m_i \in \mathbb{R}^n$ is associated. An input vector $x \in \mathbb{R}^n$ is compared with the m_i and the best match is defined as *response*: the input is thus mapped onto this location. The array and the location of the response (image of input) on it are supposed to be presented as a graphic display.

One might say that the SOM is a *nonlinear projection* of the probability density function of the high-dimensional input data onto the two-dimensional display. Let $x \in \mathbb{R}^n$ be an input data vector. It may be compared with all the m_i in any metric; in practical applications, the smallest of the Euclidean distances $||x - m_i||$ is usually made to define the best-matching node, signified by the subscript c:

$$||x - m_c|| = \min_i \{||x - m_i||\} \text{ or } c = arg\min_i \{||x - m_i||\} \tag{1}$$

Thus x is mapped onto the node c relative to the parameter values m_i. An *optimal* mapping would be one that maps the probability density function $p(x)$ in the most *faithful* fashion, trying to preserve at least the local structures of $p(x)$. (You might think of $p(x)$ as a flower that is pressed!). In the practical applications for wich such maps are intended, it may be usually self-evident from daily routines how a particular input data set ought to be interpreted. By inputting a number of typical, manually analized data set and looking where the best matches on the map according to Eq. (1) lie, the map or at least a subset of its nodes can be labeled to delineate a *coordinate system* or at least a set of characteristic reference points on it according to their manual interpretation. Since this mapping is assumed to be continuous along some hypotetical *elastic*

surface, it may be self-evident how the unknown data are interpreted by means of interpolation and extrapolation with respect to these calibrated points.

3 Autosimilar Units

The SOM is a good tool that, adapted, may allow us clasify the Electoral Units (EU) in groups and subgroups characterized by the electoral results. Their application in this case, is to group the electoral results of the EU, in some groups and subgropus that have similar characteristics. In our case is of vital importance the group of EU that give us an homogeneous group or something similar with the results of all the universe, this is finding the autosimilar EU.

To make a detailed analysis and knowing if the electoral behavior of one or some EU can be similar to the whole universe is necessary to define how the data will be treated. The results of one election is normally presented in absolutes votes and in tabular way (see Tab. 1), in wich the EU are the rows and the political parties (PP) are the columns, filling the array with the obtained votes by each PP in every EU, plus that, we add columns for the null votes and the total of electors or electoral roll.

Table 1. Typical representation of absolute electoral results.

	$\mathbf{PP}_1$	$\mathbf{PP}_2$	...	$\mathbf{PP}_n$	**Effective votation**	**Electoral Roll**
$\mathbf{UE}_1$	$V_{1,1}$	$V_{2,1}$	...	$V_{n,1}$	$T_1 = \sum V_i, 1$	P_1
$\mathbf{UE}_2$	$V_{1,2}$	$V_{2,2}$	...	$V_{n,2}$	$T_1 = \sum V_i, 2$	P_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{UE}_m$	$V_{1,m}$	$V_{2,m}$	...	$V_{n,m}$	$T_1 = \sum V_i, m$	P_m
Total	$V_1 = \sum V_{1,j}$	$V_2 = \sum V_{2,j}$	...	$V_n = \sum V_{n,j}$	$T = \sum T_i$	$P = \sum P_i$

In first place, we need to prepare the electoral data in a way such that some comparations can be made, this implies using an adecuated scale for all the EU. The adecuated scale is transforming the absolut data to a relative scale for each EU, this is, to take the percentages obtained for each PP in each EU

$$v_{i,j} = \frac{V_{i,j}}{T_j} \qquad (2)$$
$$v_i = \frac{V_i}{T}$$

This give us the advantage that all the EU sums one, regardless of the sum of the votes and the cuantity of the electoral roll (see Tab. 2).

Table 2. Typical representation of relative electoral results.

	PP_1	PP_2	...	PP_n	Null	Effective	Citizen Part.
UE_1	$V_{1,1}$	$V_{2,1}$	...	$V_{n,1}$	$V_{a,1}$	1	p_1
UE_2	$V_{1,2}$	$V_{2,2}$	...	$V_{n,2}$	$V_{a,2}$	1	p_1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
UE_m	$V_{1,m}$	$V_{2,m}$	...	$V_{n,m}$	$V_{a,m}$	1	p_m
Total	$V_1 = \sum V_{1,j}$	$V_2 = \sum V_{2,j}$	...	$V_n = \sum V_{n,j}$	V_a	1	p

In this data array, the rows represent to the EU, the columns to the PP y the cross-values are the percentages obtained for the different PP that have participated in the election process. After that, it is added the total percentage as an additional row in the data array and it is considered like another unit.

4 Zacatecas Case

Before going inside in a detailed analisys of the methodology used for finding the autosimilar sections, let us see the electoral stage: in the city of Zacatecas, capital town of Zacatecas, the City Hall is renewed every 3 years, aditionally, for being soberan part of the Zacatecas state, the citizenship participates also in the election of the local legislature representation at the same time. The state goverment is renewed every 6 years, and this election takes place in a concurrent way when both periods match. In the other hand, the election of the republic's president and senators its made every 6 years and the election of federal members of parlament every 3 years, this two last elections are, in the same way, concurrent but independent of the local ones. Of the foregoing, we have that in Zacatecas 20 elections have been done in 9 times, and there are 111 electoral units in the town of Zacatecas.

With the general view of the behavior of the electoral preference for the political parties, we can ask, How the behavior at the electoral section level is?, or even more, Does one or some electoral sections that aproximate the historical behavior of all the political parties exist? If it is true, it would be better to study that or those autosimilar sections that could be used like good aproximations of posible future elections results.

In this work, we present a method for finding the autosimilar electoral sections and after that the best of them will be used for having a good approximation to the results of the 2006 federal elections.

5 Results

In this section are presented the results of aplying the SOM to the array formed using the information of electoral sections and the total (rows) with the votes obtained by the PP from the 1995 to 2000 (columns) in the Zacatecas town [3].

Different scenarios are proposed: autosimilar sections of each election from 1995 to 2000, autosimilar elections for some combinations of elections and for all the federal and local elections, finally it is compared the use of the autosimilar section with the best approximation of the results in the federal elections of 2006.

5.1 Different Scenarios

Aplying the SOM to a local electoral process that is composed by the election of the principal political parties in the election of Mayor and local members of parlament in 1995 we get that the section 1840 is the best approximation to the total percentage of the town.

Also, there exist other sections that can be used like autosimilar: 1788, 1790, 1821 and 1827. This can be observed in the central region of the Fig. 1.

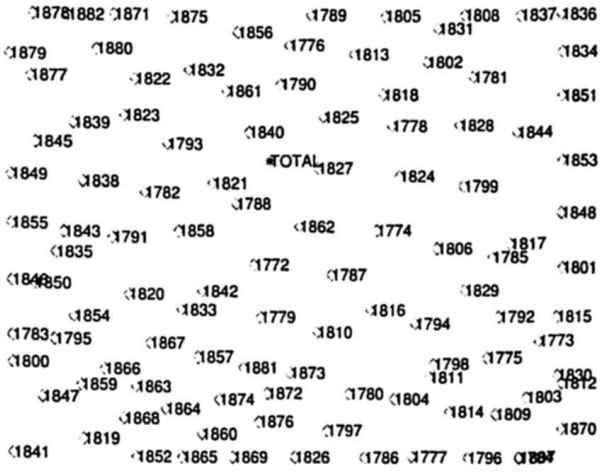

Fig. 1. SOM's result for the Mayor election and local members of parlament in 1995.

We obtain the Fig. 2 applying the SOM to the correspondent data of the election for federal members of parlament in 1997. Now, we find 2 sections with a percentage behavior very close to the whole town behavior. These are the sections 1810 and 1835. Considering those distances we can find other sections that could be proposed like autosimlar: 1804, 1820, 1838, and 1840.

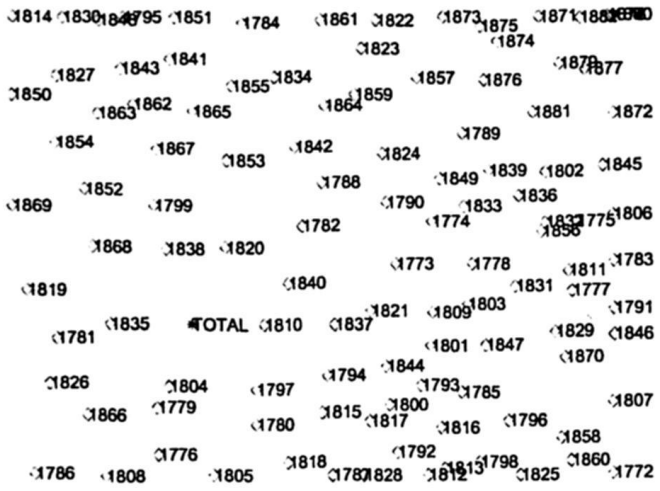

Fig. 2. SOM's result for the federal members of parlament election in 1997.

Now, let us take the 3 variables of 1998, the Mayor election, local members of parlament and governor. The result can be observed in the Fig. 3, where we can check that the best porcentual approximation section is the 1864 section, and other good approximations are: 1795, 1805 and 1844.

To conclude this part, we show the result of the SOM for the Republic's President elections and federal members of parlament in the 2000 year. In the Fig. 4 it can be analyzed that in this ocassion we can not find a section that is the best percentage approximation, those that can be used like that, are: 1780, 1822, 1854, 1863 and 1867.

This is a good point for comment that the analysis of some scenarios[4] that we take, have already been done using the cluster technic [5, 6]. The remarkable fact here, is that comparing this results with the results of the cluster analysis, similar conclusions are obtained [7]. We get the same results in the autosimilar sections, but when we look to the *other good approximations*, we realized that they are not the same, but there are coincidences.

In the Tab. 3, we have a summary that contains the election year, the best electoral section for each election and in the third column, the other good autosimilar sections.

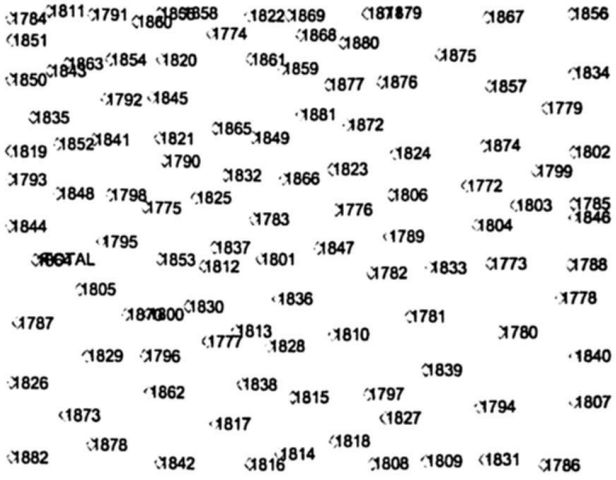

Fig. 3. SOM's result for the Mayor election, local members of parlament and governor in 1998.

Table 3. Summary of autosimilar electoral sections.

Year	Best section	Other approximations
1995	1840	1788, 1790, 1821, 1825, 1827
1997	1810, 1840	1804, 1820, 1835, 1838
1998	1864	1795, 1805, 1844
2000	1780, 1854	1822, 1863, 1867

5.2 Historic Autosimilar Section

Applying the SOM technic for the 111 sections and the town behavior, only considering the votation for the principal political parties in the whole local and federal elections since 1995 to 2001, was obtained as the best autosimilar electoral section, the section 1793, because it is the closest to the historic town behavior. This is shown in the Fig. 5.

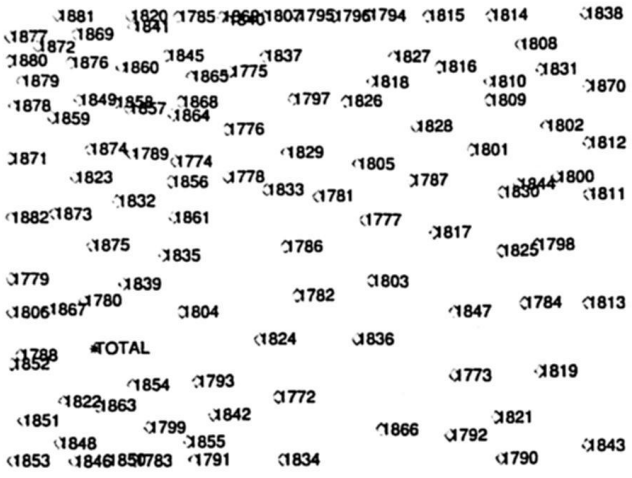

Fig. 4. SOM's result for the Republic's President election and federales members of parlament in 2000.

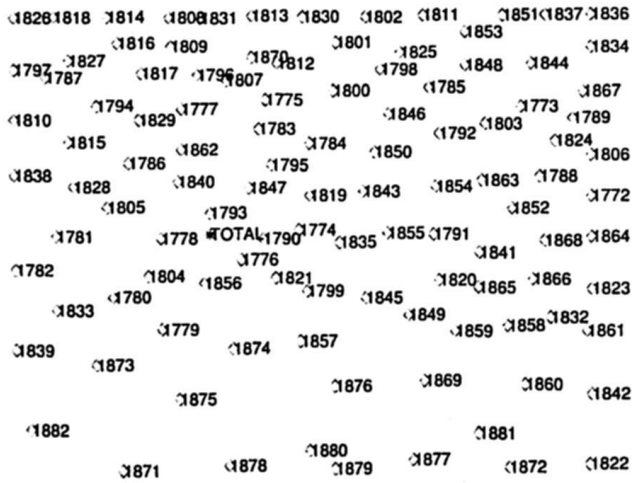

Fig. 5. SOM's result for the historic behavior considering all the elections since 1995 to 2001.

The Fig. 6 presents the historical behavior of the section 1793 as the best historically, compared with the town's historic total percentage in the different elections. In the notation, the first letter corresponds to the type of the election: M for mayor, D for members of parlament, G for governor and R for the republic's president; the las two numbers indicate the year of the election and in the middle we have the name of the political party: A for PAN, I for PRI, D for PRD, T for the PT, A*C for Alianza por el Cambio and A*M for Alianza por México.

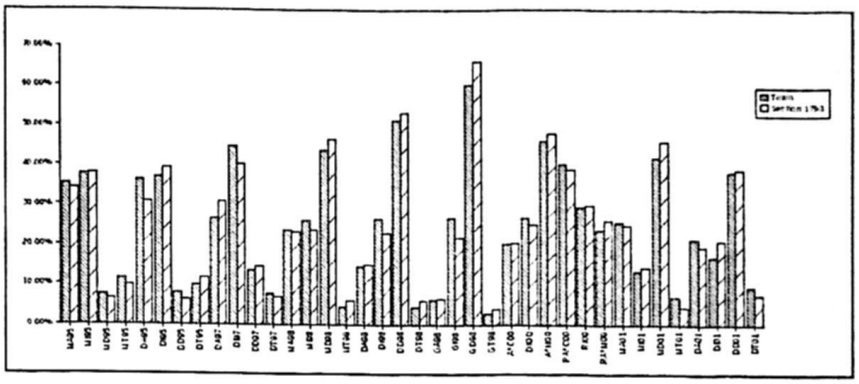

Fig. 6. Historical behavior of the section 1793 respect the town official results in the elections of 1995 to 2001.

5.3 Using Autosimilar Sections to Forecast Results

Now, we introduce the use of autosimilar sections like the best approximations for electoral processes results, it is used the best of the sections, the section 1793, taked since 1995 to 2001. Our goal is to compare this section with the town percentages obtained by the political parties that participate in the electoral process of the 2006 for President of the Republic. The summay of the analysis is shown in Tab. 4.

There, in Tab. 4 we can see that the maximum error is less than 2.5%, the half of what is commonly used in classical sampling for this kind of issues. So, we have that the autosimilar sections obtained with the SOM technic can be used to forecast results of future elections.

Table 4. Comparation of the section 1793 percentages with the results of Zacatecas town for the electoral process in the 2006 for Republic's President.

Political Party	Section 1793	Town Behavior	Error
PAN	32.7 %	35.1 %	2.4%
Alianza por México	20.2 %	17.4 %	2.8%
Coalición por el Bien de Todos	38.0%	37.2 %	0.9%
Nueva Alianza	1.1%	1.3%	0.2%
Alternativa	1.9%	4.3%	2.4%
Not registrated	3.8%	2.7%	1.1%
Nulls	2.3%	2.1%	0.1%

6 Conclusions

Through the SOM application, it was possible to find electoral sections that have an historical behavior similar to the development of local and federal electoral processes in the town of Zacatecas since 1995 to 2001.

This technique for finding autosimilar sections applying SOM, can be used not only in the Zacatecas town but can also be applied for finding similar sections in Local and Federal Electoral Districts at the State level and for the whole country. Then, the goal is to find a group of electoral sections that have a similar historical behavior to the universe of electoral sections.

The importance of having a group of autosimilar sections is to have an instrument that can be used for improve work about the surveys of voting intentions, because you can generate control surveys, and with those control surveys you can verify if you get the same results of a survey that has been designed under an statistical methodology.

The autosimilar sections can also be applied as an instrument in itself for making surveys of voting intentions, that allows to have estimators trend with lower or equal approximation errors to the classical sampling.

References

1. Kohonen, T.; Hynninen, J.; Kangas, J.; Laaksonen, J., *SOM_PAK The Self Organizing Map Program Package.* SOM Programming Team of the Helsinki University of Technology Laboratory of Computer and Information Science. (1995)
2. Kohonen, T.: Self-Organizing Maps. Springer-Verlag, Berlin (2001)
3. *Instituto Electoral de Estado de Zacatecas.* http://www.ieez.org.mx/principal.htm
4. Velasco Olvera, Hilda. ***Análisis de la votación municipal histórica en Zacatecas.*** Universidad Autnoma de Zacatecas (2002).
5. Dallas E. Johnson., ***Métodos multivariados aplicados al análisis de datos***, International Thomson Editores (1998).
6. Hair Jr. Joseph F., Anderson Rolph E., Tatham Ronald, Black William C. ***Análisis Multivariable*** 5 Ed., Prentice Hall ed. Madrid (1999).
7. Garcia A., Trueba L., Mercado G., ***Construcción de Unidades Electorales Autosimilares.*** Revista IEEM Apuntes Electorales, Año VII, Número 33, pp. 11-42 (2008).