FUZZY CLUSTERING: APPLICATION ON ORGANIZATIONAL METAPHORS IN BRAZILIAN COMPANIES

Angel Cobo
Rocío Rocha
Universidad de Cantabria, Spain
Adolfo Alberto Vanti
Gustavo Schneider
Universidade do Vale do Rio dos Sinos, Brasil

_____

ABSTRACT

Different theories of organization and management are based on implicit images or metaphors. Nevertheless, a quantitative approach is needed to minimize human subjectivity or bias on metaphors studies. Hence, this paper analyzed the presence of metaphors and clustered them using fuzzy data mining techniques in a sample of 61 Brazilian companies that operate in the state of Rio Grande do Sul. For this purpose the results of a questionnaire answered by 198 employees of companies in the sample were analyzed by R free software. The results show that it is difficult to find a clear image in most organizations. In most cases characteristics of different images or metaphors are observed, so soft computing techniques are particularly appropriate for this type of analysis. However, according to these results, it is noted that the most present image in the organizations studied is that of "organisms" and the least present image is that of a "political system" and of an "instrument of domination".

Keywords: organizational images, soft clustering, data mining.

_____

1. INTRODUCTION

Images are metaphors that permit an interpretation of what is happening to the organizational culture (Morgan, 1986). It allows to better understand the behavior of employees in order to adhere or not to organizational compliance, as stated in the study performed by Spears and Barki (2010) which analyzed their awareness on information security. In Bulgurcu, Cavusoglu and Benbasat (2010) employees` behavior on compliance matters were also analyzed, in this case involving rationality-based beliefs and information security awareness.

The study of organizational images is widely used in management courses because they provide a very detailed view from organizational studies. However, when it is narrowed to a sector or regional study, this kind of approach becomes limited. To overcome this limitation this study presents an approach using Data Mining and Soft Clustering techniques to understand what can happen in an organizational culture environment through images in a large number of companies.

The developed application reached 61 enterprises but using this method it can be extended to a significant number of companies, coming to a complete study applied to a whole country. The sections in this paper are divided into a theoretical basis, methodological aspects, a case study and finally discussions on the results.

2. IMAGES OF ORGANIZATIONS

Morgan (1986) believes that one can better understand organizations by recognizing metaphors that prompt one to view organizations through a certain angle. Metaphors play a paradoxical role: they are vital to understand and highlight certain aspects of organizations, while they restrict understanding by back grounding or ignoring others. Morgan illustrates his ideas by exploring eight archetypical metaphors of organization: Machines, Organisms, Brains, Cultures, Political Systems, Psychic Prisons, Flux and Transformation, Instruments of Domination. In this work and based on (Knorst, Vanti, Andrade, & Johann, 2011), brains and culture metaphors are considered as a single image, so 7 metaphors or images are analyzed.

Mechanistic (M): Organizations that impose rigid routines and patterns, hierarchically distributed. Dealings are impersonal and control of the organization is bureaucratic. Because it is very predictable, it is no longer regarded as ideal, even in stable and authoritarian institutions. This style also presents difficulties for innovation.

Psychic Prisons (PP): Inflexibility is a characteristic of this image, becoming a prisoner of past events, allied to fundamental attitudes by their idealizers. Some of their traps are false assumptions, rules without questioning and fanaticism around the charisma of the leader.

Political Systems (PS): This view is not often in the interest of the group and often favors authoritarian executives. This includes companies with participatory management that is encompassed in political systems because although there is a certain distribution of power, the central objective will be executed by both subordinates and the owners of the capital.

Instruments of Domination (ID): In organizations viewed as instruments of domination, the employees and managers need to completely dedicate themselves to the company. They feel insecure about their employment and experience a lot stress on the job.

Organisms (O): The fundamental principal of organisms is that it is based on the employees' intellectual capital. Motivation is a substantial factor. Because of constant innovation and deadlines, employees tend to obey a biological clock because there are targets to reach and constantly innovations to develop.

Brain/Cybernetic (C): Intellectual capital is highly valued and is constantly being stimulated to improve. Decision-making needs to be done "through formal or temporary processes, producing policies and plans that offer a point of reference or a structure for information processing" (Johann, 2008, p. 33). The definition of cybernetic is given due to the fact that information technology is permanently present, which ensures better conditions in the review of political norms and procedures, in addition to learning how to absorb changes in the environment.

Flux and Transformation (FT): Organizations that best mirror flux and transformation are those that modify and evolve to conform to change and evolution in the environment. Their survival depends on their internal and external environments

These images represent the employees` behavior as Morgan (1986) stated. Analysis can be performed properly with this approach considering organizational case studies but has some limitations in sector and regional studies using the same criteria. To balance it this paper proposes the application of Data Mining (DM) and Soft Clustering techniques that are presented below. Organizational identity can also reside in metaphorical images internalized in the members of the organization (Taber, 2007). Each employee`s perception about the image projected by the organization can even be quite different, so that the use of a fuzzy approach is particularly appropriate.

3. DATA MINING AND SOFT CLUSTERING

Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data (Han & Kamber, 2006). This area has attracted a great deal of attention in the information industry and in society as a whole in recent years and data mining techniques have been applied to a wide variety of areas. Data mining techniques have been used for trying to predict behavioral patterns, generate forecasts, identify trends or changes thereto, as well as to discover relationships between information pieces in order to optimize decision making. Thus, there is no doubt as to the practical application in those processes where a large amount of data must be handled. This explains, therefore, why this area of knowledge has drawn the attention of different sectors of the information industry in recent years.

Cluster analysis or clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields (Kaufman & Rousseeuw, 2008). Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible (Witten & Frank, 2005). The potential of clustering algorithms is to reveal the underlying structures in data and it can be exploited in a wide variety of applications, including classification, image processing

and pattern recognition, modeling and identification. In particular, data mining techniques can be used to identify categories or behavioral patterns in organizations.

Many clustering algorithms have been introduced in the literature (Pedrycz, 2005). A widespread accepted classification scheme subdivides these techniques into two main groups: hard (crisp) or soft (fuzzy) clustering. In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster, however in fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels that indicate the strength of the association between that data element and a particular cluster. Due to the fuzzy nature of many practical problems, a number of fuzzy clustering methods have been developed following the general fuzzy set theory strategies outlined by (Zadeh, 1965). Fuzzy set theory deals with the representation of classes whose boundaries are not well defined. The key idea is to associate a membership function that takes values in the interval [0,1], with 0 corresponding to non membership in the class and 1 corresponding to full membership. Thus, membership is a notion intrinsically gradual instead of abrupt as in conventional Boolean logic.

The concept of fuzzy partition is essential for cluster analysis and identification techniques that are based on fuzzy clustering. The most known method of fuzzy clustering is the Fuzzy c-Means method (FCM), initially proposed by Dunn (1973) and generalized by Bezdek (1981) and other authors; in Kruse, Hoppner, Klawonn and Runkler (1999) an overview is presented. The FCM is based on an optimization problem which objective function is defined as:

$$\sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - c_j \right\|^2$$

where $\{x_1, x_2, ..., x_n\}$ is the input sample set, that is, the objects that have to be clustered, c is the number of clusters, $\{c_1, c_2, ..., c_n\}$ the centroids of the clusters, which can be defined by a given matrix or randomly chosen, and u_{ij} is the degree of membership of x_i in the cluster j. Finally, the parameter m is a real number greater than 1 that is a weighting factor called fuzzifier. Normally the Euclidean distance is used, but any norm $\|*\|$ expressing the dissimilarity between any measured data and the center can be used. One of the drawbacks of FCM is the requirement for the number of clusters, c, to be specified before the algorithm is applied. In the literature, methods for selecting the number of clusters for the algorithm can be found (Pham, Dimov, & Nguyen, 2005).

Fuzzy partitioning is carried out through an iterative minimization of the objective function under the following fuzzy constraints:

$$0 \leq u_{ij} \leq 1 \quad \forall i, j$$

$$\sum_{j=1}^{c} u_{ij} = 1 \quad \forall i$$

In the approach proposed by Bezdek (1981) in each iteration membership levels u_{ij} and centroid positions c_j are updated applying the technique of Lagrange multipliers. The algorithm stops when a maximum number of iterations is reached, or when the algorithm is unable to reduce the current value of the objective function.

Given the fact that different organizational images can often be linked to an organization, in this work a soft clustering approach is considered more appropriate. Using the FCM technique, each organization is allowed to belong to many clusters with different degrees of membership and therefore they have multiple images or metaphors linked. In the paper the results of the analysis are presented.

Any data mining process is composed of the following basic phases or stages: data compilation; data processing (in which it is cleaned, transformed and reduced); application of data mining (determining the model to use, carrying out statistical analysis, and graphically visualizing data to obtain a first approximation); and finally, interpretation and evaluation of results obtained. Following the previous stages, in the next sections we will show the practical application of data mining techniques to identify behavioral features in a sample of Brazilian companies.

4. DATA COMPILATION AND PROCESSING: INSTRUMENT FOR THE IDENTIFICATION OF ORGANIZATIONAL IMAGES IN A SAMPLE OF BRAZILIAN COMPANIES

For the identification of images, an instrument developed by Johann (2004) was used. This instrument is a questionnaire with 35 questions on organizational aspects that are grouped into 7 blocks; each block is associated with one of the images considered. In order to identify characteristics of the images in an organization, a set of employees can make a quantitative assessment on each of the 35 questions of the questionnaire. The evaluation uses a discrete scale with values between 1 and 4, according to the following criteria: 4 if there is a strong presence, 3 if there is a reasonable presence, 2 if there is little impact and 1 if there is virtually no presence.

The Appendix shows the 35 questions selected and Table 1 shows the relationship of each question with one of the 7 images defined by Morgan.

Image	**M**	**O**	**SP**	**B/C**	**ID**	**FT**	**PP**
	1	2	3	4	5	6	7
	14	13	12	11	10	9	8
Question	15	16	17	18	19	20	21
	28	27	26	25	24	23	22
	29	30	31	32	33	34	35

Table 1. Relationships between questions and organizational images.

With the answers to 35 questions, 7 numerical values can be generated with the sum of the scores for each of the 5 questions related to each of the 7 images. These 7 values can be taken into account in determining the most relevant image in the company, according to the opinions of the employee interviewed. An example of the tabulation of answers to the questionnaire is shown in Table 2. The sums of the scores associated with each of the images are shown in the last row, for example, in the case

presented in Table 2, the most visible organizational images are those of the "political system (SP)", but images M, C and ID also obtain high scores.

Company name:	COMPANY_1												
Sector of activity:	FINANCIAL												
Main activity:	SERVICES												
Size of company:	LARGE												
Location:	ESTEIO												
Position of respondent:	SERVICE COORDINATOR												
Date:	20/03/2011												
Test tabulation													
M=Machine O=Organism SP=Politic System C=Brain and Culture													
ID=Instruments of Domination FT=Flux and Transformation													
PP=Psychic Prisons													
M		O		SP		B/C		ID		FT		PP	
01:	4	02:	3	03:	3	04:	2	05:	2	06:	3	07:	2
14:	3	13:	3	12:	3	11:	4	10:	4	09:	4	08:	4
15:	4	16:	2	17:	3	18:	4	19:	4	20:	2	21:	3
28:	2	27:	2	26:	4	25:	2	24:	2	23:	3	22:	3
29:	3	30:	3	31:	4	32:	4	33:	4	34:	3	35:	3
Total	16		13		17		16		16		15		15

Table 2. Example of answers to the questionnaire.

To analyze the organizational images with greater presence in the state of Rio Grande do Sul (Brazil), a sample of 61 companies from various sectors and sizes was selected. In each company a group of up to 4 employees were interviewed, resulting in a total of 198 responses to the questionnaire (mean of 3.25 responses per company). All data were pre-processed for analysis with data mining techniques.

5. APPLICATION OF DATA MINING: FUZZY CLUSTERS IDENTIFICATION

Clustering algorithms were applied to try to identify groups of companies responding, according to their employees, to similar images. We used a free software environment for statistical computing and graphics; this software is R and can be downloaded from the following site http://www.r-project.org/. This software implements a great variety of clustering algorithms; the Fuzzy C-Means (FCM) algorithm, implemented in package 'e1071', was selected.

The *cmeans* command needs several parameters to run:

- The data matrix where columns correspond to variables and rows to observations. In our case 7 variables were considered with the average values corresponding to the sum of the scores of the 5 questions from each of the 7 blocks given by each employee in the company. The data matrix has 61 rows (companies).
- Number of clusters or initial values for cluster centers. In our case we decided to give 7 initial cluster centers. The center of cluster *i* was initially defined as:

$$c_i = (c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}, c_{i6}, c_{i7}) \ where \ c_{ij} = \begin{cases} 5 & if \ i \neq j \\ 20 & if \ i = j \end{cases}$$

Note that 5 is the minimum value and 20 the maximum in a block of 5 questions.

- Maximum number of iterations, the value 500 was used.
- Distance measure to use; we used the "euclidean" distance.
- Degree of fuzzification (*m*). We used *m=2*.

The algorithm needed a total number of 218 iterations to converge, and the final error was 3,8954. After the execution of the 218 iterations, the cluster centers were updated as shown in Table 3. Bold identifies the highest values in each centroid, that is, the images obtained higher scores in each group.

Table 4 shows a ranking of the images with the greatest presence in each group. As can be seen, image O is clearly the most relevant in most groups. Another image with a high presence in the groups is FT. In relation to the images of smaller presence in the sample, they are those corresponding to PP and SP.

	Image M	*Image O*	*Image SP*	*Image B/C*	*Image ID*	*Image FT*	*Image PP*	*More relevant images in the cluster*
1	**14.06969**	**14.14340**	10.75342	12.25663	12.81599	**13.73587**	10.68910	O, M, FT
2	**13.43621**	**14.91055**	11.87856	13.15049	12.30009	**13.90021**	10.71262	O, FT, M
3	13.87489	13.30586	**14.32899**	12.72550	**14.97622**	**14.11905**	13.77686	ID, SP, FT
4	14.89808	**16.73978**	12.87971	**15.28379**	14.23304	**16.49389**	12.38194	O, FT, C
5	14.21077	**14.71124**	13.08771	13.61664	**14.26382**	**14.52736**	12.50887	O, FT, ID
6	**15.03585**	**15.98669**	12.03754	14.72027	13.31240	**14.95138**	11.05963	O, M, FT
7	14.31467	**14.60668**	13.31190	13.63852	**14.39336**	**14.56143**	12.62568	O, FT, ID

Table 3. Cluster centers and more relevant images after the execution of the FCM algorithm.

As a result of the algorithm we obtained a matrix with the degrees of membership of each company for each of the 7 groups identified. Table 5 shows this matrix; the membership levels can also be shown using a density plot, as in Figure 1. A graphical representation of relationships between variables in the clusters is also shown in Figure 2. Figure 1 shows the companies in the vertical axis and membership levels on the horizontal axis. The darker shades in that graph correspond to higher membership values. As shown, in some companies there is a clear association with one of the groups, but in most cases the association with a single group is not as clear. The same conclusion can be reached by observing Figure 2, which shows the pairwise relationships between the variables used for performing the process of clustering.

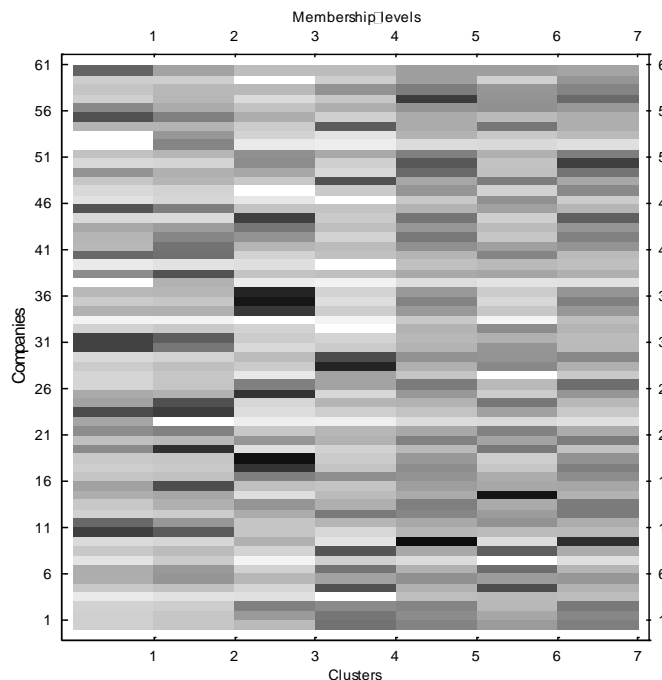| Cluster	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7
1	O	M	FT	ID	B/C	SP	PP
2	O	FT	M	C	ID	SP	PP
3	ID	SP	FT	M	PP	O	C
4	O	FT	C	M	ID	SP	PP
5	O	FT	ID	M	C	SP	PP
6	O	M	FT	C	ID	SP	PP
7	O	FT	ID	M	C	SP	PP

Table 4. Ranking of the images

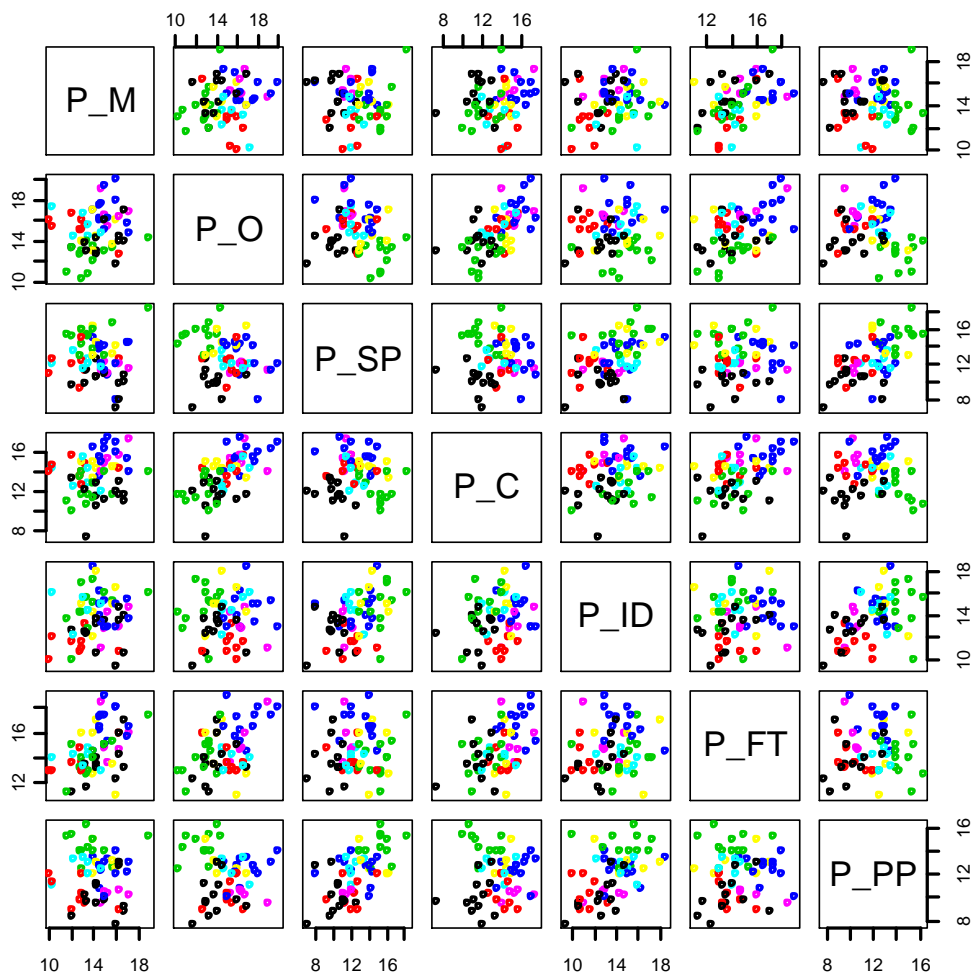*Figure 1. Density plot of membership levels in the fuzzy clustering process.*

Figure 2. Graphic clustering.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
1	0.07066084	0.08658671	0.10323928	0.21063065	0.18665701	0.15041535	0.19181016
2	0.07312988	0.08607538	0.15161303	0.20526897	0.17179575	0.13193341	0.18018358
3	0.06859525	0.07444648	0.19060567	0.17475744	0.18399386	0.10722599	0.20037532
4	0.03169135	0.04403128	0.04741090	0.57385733	0.09821968	0.10476523	0.10002422
5	0.08308866	0.09396503	0.06310459	0.26138938	0.11674858	0.26730794	0.11439583
6	0.12172084	0.15924327	0.10787176	0.13952651	0.16655198	0.14846257	0.15662307
7	0.12279663	0.15318111	0.06888526	0.20865810	0.11651387	0.21916242	0.11080261
8	0.04220773	0.07790881	0.01870347	0.07060410	0.05522701	0.68612636	0.04922252
9	0.08219092	0.10195635	0.06719794	0.25448306	0.13011700	0.23951137	0.12454337
10	0.05546326	0.06238707	0.11696803	0.04217665	0.36042338	0.05174491	0.31083671
11	0.28728923	0.24772864	0.08701666	0.05989234	0.11032077	0.10274618	0.10500617
12	0.22593052	0.15658323	0.08749638	0.11131703	0.13027708	0.16301567	0.12538010

13	0.06920147	0.07903842	0.12510591	0.19929293	0.18164698	0.14795461	0.19775967
14	0.08285978	0.11859962	0.15931363	0.12035732	0.19264877	0.12761317	0.19860771
15	0.12177958	0.13361607	0.04982584	0.10184959	0.12635401	0.35147288	0.11510203
16	0.14045734	0.26530835	0.09467698	0.08950505	0.14555268	0.13023758	0.13426202
17	0.08681304	0.09328667	0.19122849	0.16892750	0.16149141	0.12522848	0.17302441
18	0.07497735	0.08073145	0.30462401	0.09105763	0.17141417	0.08550092	0.19169448
19	0.08341590	0.08578571	0.35908959	0.07812405	0.15503223	0.07478851	0.16376401
20	0.17479429	0.30830196	0.04631832	0.07003819	0.10548488	0.20155637	0.09350599
21	0.09634673	0.10469832	0.15798910	0.11641307	0.18966021	0.13881451	0.19607807
22	0.17231070	0.18867734	0.08467363	0.10917158	0.13054105	0.18768695	0.12693876
23	0.13553735	0.65585995	0.02651960	0.02495840	0.05331346	0.05677979	0.04703145
24	0.26976889	0.29072870	0.05170888	0.07291884	0.09093024	0.14135547	0.08258898
25	0.13949925	0.26258352	0.06216253	0.11056620	0.11528520	0.20376653	0.10613677
26	0.12697591	0.11798789	0.30413710	0.06087029	0.15423258	0.07650053	0.15929570
27	0.06342657	0.08722344	0.19070266	0.13850639	0.19781697	0.10569781	0.21662616
28	0.06385745	0.08590340	0.03551134	0.14001793	0.08742439	0.50523607	0.08204943
29	0.07901129	0.09561119	0.07833999	0.32820361	0.12081321	0.17904222	0.11897848
30	0.05739890	0.06748329	0.10023415	0.26933224	0.16538393	0.16067014	0.17949735
31	0.28326301	0.20365793	0.05764492	0.07854032	0.11426179	0.15872703	0.10390500
32	0.28461438	0.24248080	0.07755410	0.06658642	0.10871574	0.11847521	0.10157336
33	0.06903735	0.08165823	0.06779278	0.38233559	0.11484726	0.17214823	0.11218056
34	0.01807095	0.01896151	0.74341633	0.01876586	0.08306922	0.01836714	0.09934898
35	0.11663941	0.11931074	0.29643997	0.07474561	0.15255682	0.08365965	0.15664780
36	0.07817857	0.08536278	0.34888223	0.04989400	0.18663021	0.06024328	0.19080894
37	0.10714989	0.11048183	0.32739166	0.06609067	0.15243308	0.07717875	0.15927413
38	0.69337887	0.11684761	0.02909171	0.02083147	0.05103369	0.04323923	0.04557743
39	0.17273715	0.25856266	0.08966250	0.09588023	0.12877008	0.13496245	0.11942493
40	0.03241751	0.04420640	0.04902702	0.57711622	0.09526600	0.10528039	0.09668645
41	0.22437489	0.21413734	0.06785018	0.09530541	0.11097166	0.18305758	0.10430293
42	0.10765563	0.21007287	0.11086404	0.10154859	0.16660240	0.14315675	0.16009971
43	0.11133460	0.18090745	0.16155856	0.06774603	0.20270129	0.08630506	0.18944702
44	0.13048799	0.14819301	0.20418735	0.09798384	0.15786594	0.10286958	0.15841229
45	0.05764661	0.05624164	0.28681837	0.08039807	0.20725059	0.07240928	0.23923545
46	0.25750525	0.19386599	0.09092025	0.08955001	0.11618149	0.14053572	0.11144130
47	0.04572654	0.06161214	0.03807164	0.52587451	0.08298042	0.16708356	0.07865120
48	0.05939674	0.06714226	0.38750100	0.08041471	0.15888866	0.06972908	0.17692754
49	0.08586828	0.11065564	0.08367132	0.26059800	0.13202895	0.19531702	0.13186079
50	0.16231335	0.11824507	0.13174813	0.05896488	0.22471806	0.09544003	0.20857048

51	0.06117608	0.06424937	0.17273495	0.07234683	0.25039101	0.09209677	0.28700499
52	0.09123197	0.11071931	0.16669303	0.12941725	0.18977795	0.11894709	0.19321339
53	0.59796974	0.18279490	0.03202861	0.02766130	0.05408170	0.05685057	0.04861317
54	0.43332432	0.15658670	0.06662395	0.04240830	0.11423622	0.08300395	0.10381654
55	0.11348891	0.11319047	0.07704351	0.24432758	0.12596022	0.20477560	0.12121372
56	0.26191311	0.19090609	0.12260103	0.07201902	0.12645872	0.10397421	0.12212780
57	0.17894663	0.12674166	0.09227940	0.12114093	0.16112533	0.16691343	0.15285262
58	0.07228857	0.11036817	0.05615977	0.08543059	0.29143982	0.16136069	0.22295239
59	0.08730678	0.10696052	0.10468709	0.16331970	0.19488071	0.15817972	0.18466549
60	0.07500784	0.07422026	0.39973367	0.07578448	0.14605454	0.07041335	0.15878586
61	0.23432493	0.13894778	0.09988333	0.10115031	0.14476995	0.14358131	0.13734238

Table 5. Membership levels in the fuzzy clustering process.

6. EVALUATION OF RESULTS

As shown in Table 3, after executing the algorithm the centroids of each group are not clearly related to a single image. Instead, each group has values assigned to each feature (image) which are very different from those initially chosen. Although in most cases the image initially linked to the centroid is among those most present in the final centroid, in one case, that corresponding to group 7, the initial image (PP) does not have a strong presence in the final centroid. In fact, its value is the lowest value obtained in the centroid of the group. This seems to confirm that this image does not have a strong presence in the sample analyzed. In addition, these final centroids show that most of the companies seem to fit a mixed image, with a combination of characteristics from different images or organizational metaphors.

The membership levels allow us to analyze the presence of organizational images for the companies in the sample. For example, the FCM algorithm has assigned the following levels of group membership to company 1:

$$u_1 = (0.0707, 0.0866, 0.1032, 0.2106, 0.1867, 0.1504, 0.1918)$$

As can be seen, in this case no single cluster can be clearly linked to this company. Instead, there are four clusters with similar degrees of membership and quite different from the rest; specifically clusters 4, 5, 6 and 7. According to the centroids of these groups (see Table 3), the most relevant organizational images in these clusters are O and FT.

Specifically, if an image k is considered, the membership levels (u_{ij}) and the final centroids (c_{ij}) could be used to obtain a quantitative assessment *eval(k,p)* of the presence of the image k in the company p, using the following expression:

$$eval(k, p) = \sum_{j=1}^{7} u_{pj} \, c_{jk}$$

In the case of company 1, the maximum values for *eval(k,1)* are reached in *k=2* (associated to image O) and *k=6* (associated to image FT), with values 15.14 and 14.86, respectively.

In some cases, the FCM algorithm is able to allocate a cluster for a particular company more clearly. For example, for company 34, the membership levels obtained by the algorithm are:

$$u_{34} = (0.0181, 0.0190, 0.7434, 0.0188, 0.0831, 0.0184, 0.0993)$$

In this case, we can see a clear link between the company and group 3. Figure 3 shows the difference with the previous case. If the evaluation of function *eval()* is carried out, the most relevant images in company 34 are ID and FT. These results confirm that this company seems to also have characteristics common in different images.
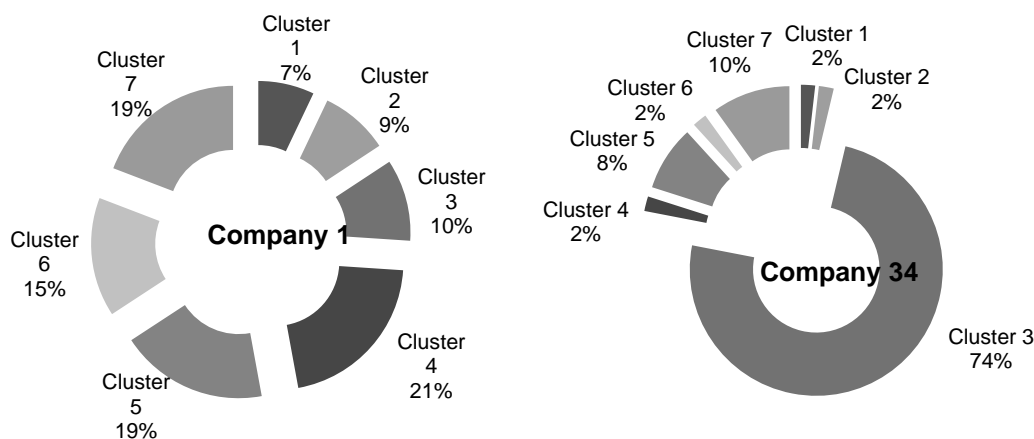

Figure 3. Image levels for companies 1 and 34 in the sample.

The R *cmeans* command also generates the closest hard clustering solution. This information is also useful for identifying significant groups. Table 6 shows the number of companies in the 7 hard clusters after the execution of the algorithm.

Cluster	Number of companies
1	12
2	8
3	12
4	12
5	6
6	5
7	6

Table 6. Cluster sizes in the closest hard clustering.

As Table 6 shows, three groups have more frequency than the rest; they are clusters 1, 3 and 4. Cluster 6 is the one with the least number of companies. The next section will discuss the results reached and possible lines for future work.

7. DISCUSSIONS

In conclusion, this paper analyzed the potential of data mining techniques to extract knowledge about organizational aspects in a sample of Brazilian companies. In addition, soft computing has shown itself to be a very suitable tool for identifying organizational patterns, where the difference between some patterns and others is not so clear.

The results of applying the techniques of soft clustering confirm the difficulty associating a single image or metaphor to a company, as features of the other images are also present. However, according to these results, it is noted that the most visible image in the organizations studied is that of "organisms". This image is the most relevant in 6 out of the 7 groups identified. This metaphor means seeing the businesses as behaving in similar ways to our own biological mechanisms; successful businesses are often adaptable and open to change and the structures and procedures are less rigid. Central to this metaphor is the theory of open systems which are "open" to their environment and have to achieve appropriate relationships with their environment in order to survive.

Also, it is remarkable the fact that the "flux and transformation" image appears in all groups with a high value. So, characteristics as constant change, dynamic equilibrium, flow, self-organization, systemic wisdom, attractors, chaos, complexity, butterfly effect, emergent properties, dialectics, and paradox are also present in most of companies.

All groups obtained seem to have a very similar structure, with the most similar and relevant images, but one of them (cluster 3) is clearly different from the rest, with a high degree of association with images that are less present in the other groups, such as "political system" and "instrument of domination" images.

In summary, this work confirms the difficulty linking a company with a single image, but it has allowed seeing images that have a greater presence in companies operating in Rio Grande do Sul.

With respect to obtaining organizational patterns, it is necessary to point out that the valuations must be carried out in the context of the specific experience analyzed. Thus, it is important to remember that the data analyzed correspond to a small sample of companies. The sample includes companies from various sectors and sizes, making it difficult to draw conclusions that can be generalized. It is necessary to extend the study with a larger sample size. It would also be interesting to carry out sector analysis to try to identify organizational features which are typical of companies in certain sectors, as well as geographically comparative studies.

In each company a group of up to 4 employees were interviewed; in some cases significant differences in the perception of different employees were observed. It would therefore be interesting to try to analyze these differences in perception, depending on the type of company and the employee profile.

From the point of view of applying soft clustering techniques, another line of research that opens from this work is the application of other soft clustering algorithms, in particular the use of algorithms that do not require the previous definition of the number of groups to be created.

In any case, the study has served to demonstrate the usefulness of the methodology proposed and to draw some conclusions about organizational images that seem to have a presence in Brazilian companies.

REFERENCES

BULGURCU, B.; CAVUSOGLU, H. BENBASAT, I. Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*. Vol.34, no.3, pp.523-548, set, 2010.

BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms**. Plenum Press, New York: 1981.

DUNN, J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, **Journal of Cybernetics** 3, pages 32-57. 1973.

HAN, J.; KAMBER, M. **Data Mining. Concepts and Techniques** (2nd Edition). Morgan Kaufmann Publishers: 2006.

JOHANN, S. L. **Gestão da cultura corporativa – como as organizações de alto desempenho gerenciam a sua cultura organizacional**. São Paulo: Editora Saraiva. 2004.

_____. Gestão da cultura organizacional. Working in paper - **SIGA - Sistema de Informação e Gestão,** Rio de Janeiro: Acadêmica, da Fundação Getúlio Vargas (FGV), 2008.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding Groups in Data: An Introduction to Cluster Analysis**. John Wiley & Sons, Inc., Hoboken, NJ, USA: 2008.

KNORST, A; VANTI, A.A.; ANDRADE, R.;JOHANN, S. Aligning information security with the image of the organization and prioritization based on fuzzy logic or the industrial automation sector. **Journal of Information Systems and Technology Management.** Vol. 8, No. 3, p. 571-596, 2011.

KRUSE, R.; HOPPNER, F.; KLAWONN, F.; RUNKLER, T. **Fuzzy Cluster Analysis**. John Wiley and Sons: 1999.

MORGAN, G. **Images of Organization** (1st ed.). SAGE Publications, Newbury Park, CA. 1986.

PEDRYCZ, W. **Knowledge-Based Clustering**. John Wiley & Sons, Inc: 2005.

PHAM, D. T,; DIMOV, S. S.; NGUYEN, C. D. Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: **Journal of Mechanical Engineering Science** 219(1):103-119. 2005.

SPEARS, J.; BARKI, H. User participation in information systems security risk management. *MIS Quarterly*. Vol.34, no.3, pp.503-522, sept. 2010.

TABER, T.D. Using metaphors to teach organization theory. **Journal of Management Education**, vol. 31(4), pag. 541-545. 2007.

WITTEN, I.; FRANK, E. Data **Mining. Practical Machine Learning Tools and Techniques**. Morgan Kaufmann Publishers. 2005.

ZADEH, L. Fuzzy Sets. **Information and Control**, 8, pp. 338–352. 1965.

Appendix: Questionnaire for the identification of organizational images

1) Procedures, operations and processes are standardized.

2) Changes in the organization are normally a reaction to changes that already occurred in the macro business environment.

3) Administrators frequently talk about authority, power and superior-subordinate relationships.

4) Flexible and creative action.

5) Working in inadequate circumstances and conditions is considered a proof of loyalty to the organization.

6) The organization sees itself as a part of a larger system where there is an interdependence that involves the community, suppliers and the competition.

7) People and groups tend to display infantile behavior.

8) Past achievements are constantly cited as references and as examples on how to deal with present situations and how to face future adversities.

9) The organization evolves in harmony and balance with its macro environment.

10) People act under constant stress and pressure.

11) There is constant questioning and redirection of actions.

12) Power serves to provide discipline and achieve order in conflicts of interest.

13) The organization considers the motivations and needs of people.

14) There are rigid patterns and uniformity in people's behavior.

15) The company has and utilizes a great number of rules, norms and regulations about operational aspects of the business.

17) The delegation of power to operational levels tends to be very restricted.

18) Negative feedback is encouraged to correct the organizational direction.

19) The organization expects complete devotion and dedication from its employees.

20) The company benefits more from external events (environmental, etc.) than from strict planning.

21) There are many taboos and prejudices in the organization.

22) The relationships between superiors and subordinates tend to contain elements of love and hate.

23) Long term achievements will be achieved in partnership with the forces acting with the macro-environment and not against it.

24) To dismiss people and streamline activities are part of the game.

25) Most people think about and influence on the destiny of the company.

26) Interpersonal gossip consumes energy and diverts attention from productivity.

27) Organizational objectives and people's needs can be met simultaneously.

28) The organization is a realm of bureaucracy.

29) The organization is expected to operate in a routine, efficient, reliable and predictable manner.

30) Employees are seen as valuable resources who can offer rich and varied contributions to the organizations activities, provided that the organization attends to their needs and motivations.

31) Rumors and gossip are frequent.

32) The organization tends to offer quick answers to changes in their macro-environment.

33) The organization values executives who appear framed and faithful to the mode of being of the company

34) In strategic decision making the company normally abandons the simple view and prefers to take into account the complexity of the situation.

35) People are dedicated to the organization because they feel they belong to something greater, which transcends their existence and individual limitations.