Profiling and Labour Market Accessibility for the Graduates in Economics at Naples University

Simona Balbi, Maria Gabriella Grassia[1]

Department of Mathematics and Statistics, University of Naples "Federico II", Italy

Summary. In this paper, after defining a pseudo-panel of groups observed at subsequent times, we propose a strategy for the construction of a set of association rules related to different survey occasions. First, we measure the similarity between systems built at different times for understanding the stability of the phenomenon. We apply a procedure developed for symbolic data analysis for this purpose. The procedure consists of two phases: the definition of the pseudo-panel and that of a system of rules referred to the semantic marking technique. Then, the agreement between the systems is measured. We applied such a strategy for studying the labour market accessibility for graduate in Economics, the University of Naples "Federico II", and the market evolution during an eight-year time span.
Keywords: Semantic marking technique; Pseudo-panel; Association rule; Symbolic objects.

1. The pseudo-panel definition

The Faculty of Economics at the University of Naples "Federico II" has been carrying out for over twenty years recurrent sample surveys on its graduates in order to evaluate their labour market accessibility. We will examine the last three surveys for evaluating the evolution of the phenomenon and apply a coherent policy.

The questionnaire and the survey methods for these surveys are constant in time. Therefore, it is possible to examine the evolution of the graduates' be-

[1] This paper is the result of the joint effort of the two authors. S. Balbi was responsible for the final editing of Sections 1, 6 and 7, whereas M.G. Grassia was responsible for the other Sections.

haviours and destiny by constructing a *pseudo-panel* formed by *cohorts*, that is to say, sets of individuals identified according to characteristics that do not vary in time according to the studied phenomenon.

The analyses will be carried out on higher order units, formed by aggregating the elementary units, which are present at each survey occasion. For instance, it is possible to study if the selection devices used for graduate women who have obtained the maximum final score have changed in a given period.

The literature on symbolic objects has produced statistical methods for the analysis of complex structures. The complexity relates both to the characteristics of the units and the membership relationship linking each elementary unit to its own object (Section 2).

In the following, we put forward a strategy that, taking an advantage from the tools developed within the analysis of symbolic objects, makes a pseudo-panel approach feasible in the described context. At each survey occasion, the proposed strategy defines the constitutive elements of a pseudo-panel in terms of association between descriptors.

We propose a data-driven strategy suitable to set up a pseudo-panel according to the data association structure. The rules (called *implication rules* or *logical rules "if-then"*) will be referred to the survey waves and associated to measures of the rule authenticity.

The comparison between the rules may give a measure of the structural stability of the phenomenon. The rules refer to a symbolic data analysis frame, the *symbolic marking*, and the comparison between rules will be carried out by means of a similarity measure between symbolic objects.

2. The symbolic objects

A symbolic object, *s*, is defined as a triplet:

$$s = (a, R, d)$$

where: $d=(d_1 , ..., d_j, ..., d_p)$ is a set of values on p descriptors, $Y=(Y_1,...,Y_j,...,Y_p)$, of the object,

a is a *recognition* function ,

$R=(R_1, ..., R_j, ..., R_p)$ is the type of relation applied for the comparison between the description provided at a conceptual level, in *intention*, from d, and an observation.

The descriptors of a symbolic object can be on a nominal, continuous or discrete scale and can have several categories for each object. The Boolean function a has categories *true* and *false* and identifies those elements which belong to the d description set and that are the *extension* of the s object, $ext(s)$.

The construction of conceptual models described in terms of symbolic objects (Bock & Diday, 2000) may be based on:

1) the expert opinion,
2) the knowledge acquired from repeated surveys.

We will merge the two approaches by using the implied longitudinal nature of the data. In order to construct a *pseudo-panel* we have to identify the structural characteristics that allow a partition of the time-related samples. Then, we will interpret the partition through the associated symbolic object.

Let us consider a set of units $E = \{1, 2, \ldots, Q\}$ to which a questionnaire was administered T times. This set can be partitioned in t subsets E_t, with $t = 1,\ldots, T$, composed by the units who participated to wave t. If, for example, a graduate participated in several waves, he or she is considered each time as a different individual. So, $E_1 \cap E_2 \cap \ldots \cap E_t = \varnothing$ and $\cup E_t = E$, for $t = 1, \ldots, T$.

Let us assume that the E elements are described by the same P variables $Y = \{Y_1, \ldots, Y_j, \ldots, Y_p\}$ and that each Y_j variable has m_j response categories. Continuous variables are made discrete with the same scale at all times. If the elementary units of the E set are groups of individuals that possess common characteristics, the data structure will be a symbolic matrix whose generic element is the marginal frequency distribution of the Y_j modal variable.

The s_k symbolic object is then defined as:

$$s_k = \bigwedge_{j=1}^{P} \left[Y_j = \left\{ y_{jm}, p_{jm} \right\}_{m=1,2,\ldots,m_j} \right], \tag{1}$$

where p_{jm} is the relative frequency of y_{jm}, m-th category of Y_j.

In defining a symbolic object, it can be useful to consider the object implications, assuming relations that can be expressed as logical rules (*if-then*, see: Agrawal *et al*., 1993). The symbolic object s_k is defined as:

$$s_k = \underbrace{\bigwedge_{a=1}^{A} \left[Y_a = \left\{ y_{am}, p_{am} \right\}_{m=1,2\ldots m_a} \right]}_{(A)} \Rightarrow \underbrace{\bigwedge_{c=1}^{C} \left[Y_c = \left\{ y_{cm}, p_{cm} \right\}_{m=1,2\ldots m_c} \right]}_{(C)} \tag{2}$$

$$\text{with } A, C \subseteq Y \quad \text{and} \quad A \cap C = \varnothing,$$

where A is the set of the antecedent categories (whenever possible, independent and exclusive variables) and C is the set of consequent categories.

With reference to T times, it is possible to define s_{kt} in each E_t. The use of complex structures allows us to measure the similarity between the objects collected in different times, as well as to value the stability of the structures. The comparison is done among the implication logical rules: for instance, two generic objects s_{kt} and $s_{k't'}$ with the same expression for A and a different expression for C, imply a change of the individual behaviour going from time t to time t'.

3. Selection of descriptors that define the objects implication

For the selection of the descriptors apt to define the intension of the symbolic object, having chosen the *consequence* variable, we use the semantic marking technique (Gettler-Summa, 1998; Grassia e Muratore, 2001), considering all the remaining variables that are part of the set of *antecedent* characteristics within the *implication* of a logical rule.

The semantic marking is a non-binary segmentation technique aimed at pointing out the characteristics of a class *K* (it may be a natural partition, or derived from a cluster analysis), considering the conjunctions and disjunctions logical links among the attributes that describe the units. The procedure determines some *marking cores*, that is to say, groups of individuals that are identical according to a set of "traits":

$$mc_g : [Y_1 = y_{1m}] \wedge ..... \wedge [Y_r = y_{rm'}] \quad \text{with} \quad r \leq P . \tag{3}$$

The union of the *G marking cores* mc_g (expressed in terms of logical AND), based on the OR disjunction operator, forms the description of the K class:

$$K : mc_1 \vee mc_2 \vee ..... \vee mc_g \vee ..... \vee mc_G \ . \tag{4}$$

By using the semantic marking, we build abstractions based on the two criteria of: homogeneity of the *K* elements and difference with the *NOT-K* elements. The parameters of the algorithm are the indexes:

1. *Rec* = Card [$ext_K(mc_g)$] ,
2. *Deb* = Card [$ext_{\bar{K}}(mc_g)$] ,

where mc_g is a generic marking that is a subset of *K*, characterized by the same categories of one or more descriptors. The *Rec* index is the percentage of elementary units belonging to *K* that satisfy the conditions defined by the mc_g marking. The *Deb* index is the percentage of elementary units that satisfy the marking conditions, but do not belong to *K*.

The semantic marking is a procedure for constructing symbolic objects, because its output is a symbolic matrix of smaller dimensions than the input matrix, with the same variables expressed in modal form, that is to say, with the respective frequency or probability distributions.

4. The comparison among objects

We compare two symbolic objects in times *t* and *t'* with a dissimilarity measure (Bock & Diday, 2000; Bruzzese & Davino, 2002) based on the Minkowski L_1 distance:

$$d(s_{k_t}, s_{k't'}) = \sum_{j=1}^{P} \frac{1}{P} d(Y_j(s_{k_t}), Y_j(s_{k't'})) \tag{5}$$

where $d(Y_j(s_{kt}), Y_j(s_{k't'}))$ compares the frequency distributions so that:

$$d(Y_j(s_{kt}), Y_j(s_{k't'})) = \frac{1}{m_j} \sum_{j=1}^{m_j} \frac{\left| p_{y_j}(s_{k_t}) - p_{y_j}(s_{k't'}) \right|}{\max \left| p_{y_j}(s_{k_t}) - p_{y_j}(s_{k't'}) \right|}. \tag{6}$$

The dissimilarity index varies between 0 if the two objects have the same frequency distribution for each variable, and 1 if they are completely different, that is to say, if the dissimilarity is maximum for all variables.

If we introduce the logical relationships (2), the dissimilarity between two symbolic objects can be the average of the dissimilarities among the frequency distributions of the variables in antecedent (A) and consequent (C) categories:

$$d(s_{kt}, s_{k't'}) = \frac{1}{2}\left(d(A_{kt}, A_{k't'}) + d(C_{kt}, C_{k't'})\right), \tag{7}$$

where $d(A_{kt}, A_{kt'})$ e $d(C_{kt}, C_{kt'})$ are the dissimilarities among the objects s_{kt} and $s_{kt'}$ obtained with formulas (5) and (6), considering in the first case only the variables in antecedent and in the second only the variables in consequence.

5. A synthetic index for rule evaluation

Let us consider a single modal consequent variable. The extension of this variable to other consequent variables is immediate, through the construction of composite variables. By using the semantic marking, which characterizes a natural partition obtained from the categories of the consequent variable, we can have objects that, compared in T times, can assume the same or different consequence.

If the Y_c consequent variable has m_c response categories (y_{c1},…y_{cm},..y_{cm_c}), the dissimilarity between times t and t' will be the mean obtained for each category by averaging the means of the dissimilarities among the antecedences of the objects at times t and t' that have an identical consequence and averaging the similarities among the antecedences of the objects at times t and t' having different consequences:

$$d(t,t') = \sum_{y_{cm}=1}^{m_c} \frac{1}{2m_c} \left\{ \frac{1}{ng} \sum_{k=1}^{n} \sum_{k'=1}^{g} d(A_{s_{k_{y_{cm}t}}}, A_{s_{k'_{y_{cm}t'}}}) + \frac{1}{nh} \sum_{k=1}^{n} \sum_{k''=1}^{h} (1 - d(A_{s_{k_{y_{cm}t}}}, A_{s_{k''_{\bar{y}_{cm}t'}}})) \right\}, \tag{8}$$

where n is the number of symbolic objects at time t having y_{cm} category for the answer variable (consequent), g is the number of symbolic objects at time t'

having the same modality m_j for the response variable, h is the number of symbolic objects at time t' having any other category for the variable Y_c.

6. Labour market accessibility for Economics graduates

We applied the strategy set forth in the previous chapters in order to study the labour market accessibility for students who graduated in Economics at the University of Naples "Federico II".

The data at hand were collected in three repeated surveys performed in 1997, 2000 and 2002. The questionnaire was structured into a set of common questions and specific modules with each one relating to a survey occasion. The comparisons concerned the common parts of the questionnaires (Table 1).

We analysed 1030 units: 385 from the 1997 survey, 397 from 2000 and 248 from 2002 (Table 2).

The *consequence* variable is the employment status with three categories: *unemployed, not permanently employed, permanently employed.* The aim of our analysis is to study the evolution of the labour market accessibility from 1997 to 2002. There are 29 *antecedent* variables in the construction of logical rules. By using the semantic marking, 43 marking cores have been pointed out for the different periods (13 in 1997, 18 in 2000, and 12 in 2002). A measure

Table 1. Common variables of the questionnaire

1	Gender	16	Number of job interviews
2	Residence during studies	17	Job conditions
3	Diploma degree	18	Time spent searching for job
4	Age	19	Job position
5	Type of secondary school	20	Type of job
6	Secondary school degree	21	Company's economic sector
7	Type of studies	22	Channels used for job finding
8	University attendance	23	Work site
9	Subject of diploma	24	Second job
10	Years spent getting degree	25	Job sector condition
11	English knowledge	26	University satisfaction
12	Informatics knowledge	27	Job satisfaction
13	PhD	28	University education
14	Job qualification	29	Job mobility
15	Worked during studies	30	Research channels

Table 2. Dataset structure

	ID	*Gender*	*Study Residence*	*Diploma degree*	*Age*	*…*	*Job satis-faction*	*University education*	*Research channels*
1997	*I1*	1	3	4	2	….	1	4	4
	I21	2	2	2	3	….	2	2	2
	….	….	….	….	….	….	….	….	….
	I254	2	1	1	2	….	1	1	1
2000	*I1*	1	2	3	1	….	2	3	3
	I2	1	2	1	1	….	2	1	1
	….	….	….	….	….	….	….	….	….
	I332	1	1	2	2	….	1	2	2
2002	*I1*	2	3	1	2	….	2	1	1
	I2	2	1	1	2	….	3	1	1
	….	….	….	….	….	….	….	….	….
	I220	1	2	1	4	….	1	1	1

of importance related to the previously described *Rec* and *Deb* indexes has been associated to each *marking core*.

For example, let us consider the first marking for the *not permanently employed* category during the years 1997, 2000 and 2002 (Tables 3, 4 and 5). By using the semantic marking, the individuals who answered not to have a permanent job constitute pseudo-panels represented by 7 symbolic objects in 1997 and in 2000 and by 5 in 2002. All the non-marked individuals form a residual symbolic object for each year.

Within the wide group of interviewees who declared not to have a permanent job, in 1997 there was a subgroup of individuals who had found a job thanks to family ties. The persons of this subgroup were *self-employed*, *did not work* during their university studies and *were unsatisfied* of the university education received.

So, the logical rules were expressed as:

IF Channels Used = *family ties* **And** Job Position = *self-employed* **And** University Satisfaction = *no* **And** Worked during studies = *no* **And** Company's economic sector = *n.a.* (*not applicable*)
THEN Job Position= *not permanently employed*.

This rule was good for the 13% of the *not permanently employed* persons and was bad for the 2% of the other persons.

Table 3. First marking of the not permanently employed category in 1997

Category: not permanently employed category year 1997 (119)					
	Weight	**Percentage**	**T-Value**		
Marking 1	**18**	**1,75**	**7,443**	**Category**	**Variable**
			8,118	*Family ties*	Channels used
REC	16	13,45	7,933	**And** *Self-employed*	Job position
REC Correct	16	13,45	7,299	**And** *No*	University satisfaction
REC Cumulated	16	13,45	3,623	**And** *No*	Worked during studies
DEB	2	11,11	2,048	**And** *not applicable*	Company's economic sector

For the year 2000 survey, we have the following rule:

IF University Satisfaction = *partial* **And** Company's economic sector = *services for firms* **And** Channels Used = *family ties* **And** Job Position = *self-employed person* **And** Type of Diploma = *technical diploma* **THEN** Job Position= *not permanently employed.*

This rule was good for the 12% of the *not permanently employed* persons and was bad for the 1% of the other graduates.

Table 4. The first marking of the not permanently employed category in year 2000

Category: not permanently employed category year 2000 (181)					
	Weight	**Percentage**	**T-Value**		
Marking 1	**24**	**2,33**	**7,411**	**Category**	**Variable**
			9,248	*Partial*	University satisfaction
REC	21	11,60	7,552	**And** *Services for firms*	Company's economic sector
REC Correct	21	11,60	7,247	**And** *Family ties*	Channels used
REC Cumulated	21	11,60	4,324	**And** *Self-employed*	Job position
DEB	3	12,50	1,598	**And** *Technical Diploma*	Type of diploma

For the year 2002 survey, we have the following rule:

IF University satisfaction = *missing* **And** Job position = *self-employed* **And** Job satisfaction = *No*
THEN Job Position= *not permanently employed.*

This rule was valid for 50% of the *not permanently employed* and, in any case, it was false.

Table 5. The first marking of the not permanently employed category in year 2002

Category: not permanently employed category year 2000 (181)					
	Weight	**Percentage**	**T-Value**		
Making 1	**76**	**7,38**	**18,015**	**Category**	**Variable**
REC	76	50,33			
REC Correct	76	50,33	8,018	*Missing*	University satisfaction
REC Cumulated	76	50,33	7,473	**And** *Self-employed*	Job position
DEB	0	0,00	6,578	**And** *No*	Job satisfaction

We highlight the radical modifications occurred in the last period for the necessary qualifications requested to enrol at the Register of Graduates in Economics and Commerce, a regulated profession that represents one of the most important employment opportunities for graduates in Economics in Naples.

Therefore, this event can be measured, by applying the proposed strategy, considering the characteristics in the antecedent part of the rules referred to different periods.

The aim of the proposed strategy is an evaluation of all the modifications registered in the structure of the rules that represent the phenomenon in each period. The three previous marking cores represent only one of the 7 observed subsets (5 for 2002). The global evaluation must then consider all the rules that have the same consequence and all the situations when the antecedences have produced different consequences in different times. We have also considered, while computing this index, all the variables pointed out on the studied object.

By using the marking cores, we have built the matrix of the symbolic objects in modal form, composed by 43 rows and 30 columns (Table 6).

In Table 7 we show, for each compared pair of years, the value of the first term of (8), $\sum^{m_c} \frac{1}{n\,g\,m_c} \sum_{k=1}^{n} \sum_{k'=1}^{g} d(A_{s_{k y_{cm} t}}, A_{s_{k' y_{cm} t'}})$, that is to say the dissimilarity of the rules that have identical consequence.

The similarity among the rules in each time *t* and the rules of the other years having different consequences, that is to say the value of the second term $\sum_{y_{cm}=1}^{m_c} \frac{1}{n h m_c} \sum_{k=1}^{n} \sum_{k''=1}^{h} (1 - d(A_{s_{k y_{cm} t}}, A_{s_{k'' \bar{y}_{cm} t'}})$ of (8), is shown in Table 8.

Table 6. Symbolic matrix of data

	Gender		Residence			Diploma Degree				....	To move				
	Male	Female	Naples	Province of Naples	Other Province	- 95	96 - 105	106 - 110	110 with Lode	... ...	Yes	Yes only in Italy	Yes only in my region	No	NR
1997 unemployed 1	35,00	65,00	45,00	35,00	20,00	0,00	100,00	0,00	0,00	... ...	40,00	25,00	25,00	5,00	5,00
1997 unemployed 2	52,38	47,62	52,38	23,81	23,81	23,81	33,33	23,81	23,81	... ...	42,86	33,33	14,29	0,00	9,52
1997 unemployed 3	46,15	53,85	38,46	46,15	15,38	0,00	100,00	0,00	0,00	... ...	38,46	7,69	23,08	30,77	0,00
1997 unemployed 4	53,66	46,34	50,00	26,83	23,17	21,95	37,80	21,95	20,73	... ...	45,12	30,49	10,98	9,76	3,66
1997 unemployed without marking	100,00	0,00	17,65	41,18	41,18	5,88	70,59	5,88	0,00	... ...	23,53	17,65	35,29	23,53	0,00
1997 not permanently 1	38,89	61,11	27,78	38,89	33,33	16,67	72,22	16,67	5,56	... ...	33,33	11,11	5,56	11,11	38,89
1997 not permanently 2	46,15	53,85	38,46	23,08	38,46	15,38	46,15	15,38	23,08	... ...	38,46	23,08	0,00	0,00	38,46
1997 not permanently 3	42,86	57,14	7,14	35,71	57,14	21,43	71,43	21,43	0,00	... ...	35,71	7,14	14,29	7,14	35,71
1997 not permanently 4	0,00	100,00	27,27	18,18	54,55	0,00	100,00	0,00	0,00	... ...	27,27	0,00	9,09	9,09	54,55
1997 not permanently 5	12,50	87,50	37,50	37,50	25,00	25,00	50,00	25,00	0,00	... ...	12,50	25,00	0,00	12,50	50,00
1997 not permanently 6	40,00	60,00	100,00	0,00	0,00	0,00	60,00	0,00	20,00	... ...	20,00	20,00	20,00	0,00	40,00
1997 not permanently without marking	53,49	46,51	34,88	37,21	27,91	16,28	32,56	16,28	18,60	... ...	25,58	0,00	4,65	4,65	65,12
1997 permanently without marking	55,56	44,44	38,89	38,89	22,22	5,56	27,78	5,56	27,78	... ...	22,22	44,44	16,67	5,56	11,11
2000 unemployed 1	66,67	33,33	16,67	16,67	66,67	0,00	100,00	0,00	0,00	... ...	50,00	0,00	0,00	0,00	50,00
2000 unemployed 2	50,00	50,00	16,67	16,67	66,67	0,00	100,00	0,00	0,00	... ...	0,00	0,00	33,33	0,00	66,67
2000 unemployed 3	57,14	42,86	0,00	100,00	0,00	14,29	42,86	14,29	14,29	... ...	28,57	14,29	28,57	0,00	28,57
2000 unemployed 4	60,00	40,00	40,00	60,00	0,00	0,00	100,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 unemployed 5	25,00	75,00	100,00	0,00	0,00	0,00	100,00	0,00	0,00	... ...	25,00	0,00	25,00	0,00	50,00
2000 unemployed without marking	43,94	56,06	46,97	30,30	22,73	22,73	31,82	22,73	22,73	... ...	13,64	21,21	7,58	9,09	48,48
2000 not permanently 1	70,83	29,17	33,33	45,83	20,83	16,67	45,83	16,67	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 2	50,00	50,00	37,50	37,50	25,00	18,75	43,75	18,75	18,75	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 3	25,00	75,00	37,50	50,00	12,50	12,50	43,75	12,50	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 4	60,00	40,00	60,00	30,00	10,00	50,00	30,00	50,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 5	33,33	66,67	66,67	16,67	16,67	16,67	50,00	16,67	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 6	0,00	100,00	50,00	37,50	12,50	12,50	25,00	12,50	25,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently without marking	55,66	44,34	48,11	28,30	23,58	23,58	38,68	23,58	17,92	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 1	0,00	100,00	25,00	62,50	12,50	25,00	25,00	25,00	25,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 2	37,50	62,50	25,00	75,00	0,00	12,50	50,00	12,50	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 3	40,00	60,00	60,00	40,00	0,00	20,00	80,00	20,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 4	100,00	0,00	25,00	75,00	0,00	0,00	75,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently without marking	44,90	55,10	48,98	38,78	12,24	10,20	22,45	10,20	34,69	... ...	0,00	0,00	0,00	2,04	97,96
2002 unemployed 1	75,00	25,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	... ...	25,00	50,00	0,00	0,00	25,00
2002 unemployed 1	33,33	66,67	0,00	100,00	0,00	0,00	0,00	0,00	0,00	... ...	0,00	100,00	0,00	0,00	0,00
2002 unemployed without marking	29,46	70,54	53,49	31,01	15,50	18,60	37,98	18,60	20,93	... ...	3,88	6,20	3,10	0,00	86,82
2002 not permanently 1	47,37	52,63	36,84	48,68	14,47	9,21	52,63	9,21	7,89	... ...	38,16	31,58	15,79	11,84	2,63
2002 not permanently 2	50,00	50,00	50,00	41,67	8,33	8,33	25,00	8,33	50,00	... ...	75,00	25,00	0,00	0,00	0,00
2002 not permanently 3	12,50	87,50	37,50	25,00	37,50	25,00	37,50	25,00	12,50	... ...	0,00	0,00	12,50	0,00	87,50
2002 not permanently 4	42,86	57,14	85,71	0,00	14,29	28,57	28,57	28,57	42,86	... ...	0,00	0,00	0,00	0,00	100,00
2002 not permanently without marking	30,51	69,49	43,22	33,90	22,88	15,25	40,68	15,25	16,10	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 1	14,29	85,71	57,14	28,57	14,29	57,14	0,00	57,14	28,57	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 1	0,00	100,00	42,86	57,14	0,00	14,29	28,57	14,29	28,57	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 3	60,00	40,00	60,00	40,00	0,00	0,00	100,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently without marking	48,28	51,72	41,38	41,38	17,24	10,34	46,55	10,34	12,07	... ...	0,00	0,00	0,00	0,00	100,00

Table 7. Dissimilarity of the rules that have identical consequence

Unemployed 1997-2000	0,19
Unemployed 1997-2002	0,23
Unemployed 2000-2002	0,27
Not permanently employed 1997-2000	0,25
Not permanently employed 1997-2002	0,29
Not permanently employed 2000-2002	0,24
Employed 1997-2000	0,40
Employed 1997-2002	0,39
Employed 2000-2002	0,21

Table 8. Similarity among the rules at time *t*

Unemployed 1997 and all other rules 2000	0,59
Unemployed 1997 and all other rules 2002	0,62
Unemployed 2000 and all other rules 2002	0,63
Not permanently employed 1997 and all other rules 2000	0,66
Not permanently employed 1997 and all other rules 2002	0,65
Not permanently employed 2000 and all other rules 2002	0,71
Employed 1997 and all other rules 2000	0,74
Employed 1997 and all other rules 2002	0,74
Employed 2000 and all other rules 2002	0,68

Table 9. Dissimilarity among the rules system in the three surveys

1997-2000	*1997-2002*	*2000-2002*
0,47	0,49	0,26

By applying formula (8), we obtain the dissimilarity among the rules system in the three surveys shown in Table 9.

So, it is possible to see how the first survey, related to a period of unemployment for Neapolitan graduates, shows anomalous accessibility rules to the labour market, while “a getting back to normality” is expressed by the poor value of the dissimilarity index relative to the comparison between the years 2000 and 2002.

7. Conclusions and future developments

We proposed a strategy, the symbolic marking, as a tool for the analysis of the evolution of a phenomenon in a given time span. We have shown it is applicable to the examined dataset and may be applied in other cases.

The proposed comparison measure may be further enriched if, for its calculation, we introduce a weighting system related to the different strength of the applied logical rules. Other future developments may derive from the possibility of simultaneous treatment of different types of variables (multinomial, modal, continuous, interval variables) without operating any previous transformations.

References

AGRAWAL R., IMIELINSKI T., SWAMI A. (1993) Mining Associations between Sets of Items in Massive Databases, *Proceedings of the ACM-SIGMOD 1993 Int'l Conference on Management of Data*, Washington D.C.: 207-216.

BOCK H., DIDAY E. (2000) *Analysis of Symbolic Data*, Springer Verlag, Heidelberg.

BOCCI L., RIZZI A. (2000) Misure di prossimità nell'analisi dei dati simbolici. In: *Atti della XL Riunione Scientifica della Società Italiana di Statistica, Sessioni Plenarie e specializzate*, Firenze, 91-102.

BRUZZESE D., DAVINO C. (2002) Post Analysis of Association Rules in a Symbolic Framework. In: *Atti della XLI Riunione Scientifica della SIS*, Milano, 63-66.

GRASSIA M.G., MURATORE M.G. (2001) I controlli nei questionari CATI dell'indagine ISTAT sulla sicurezza dei cittadini. In: FABBRIS L. (ed) *CAPTOR 2000. Qualità della didattica e sistemi computer-assisted*, CLEUP, Padova: 309-326.

GETTLER-SUMMA M. (1998) *MGS in SODAS: Marking and Generalization by Symbolic Objects in the Symbolic Official Data Analysis Software*, Cahier9935, Université Dauphine LISE CEREMADE, Paris.