Neurocomputing 147 (2015) 185–196

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Analysis of professional trajectories using disconnected self-organizing maps

Etienne Côme [a,*], Marie Cottrell [b], Patrice Gaubert [c]

[a] Université Paris-Est, IFSTTAR, GRETTIA, F-93166 Noisy-Le-Grand, France
[b] Laboratoire SAMM, Université Paris 1 Panthéon-Sorbonne, 90 rue de Tolbiac, F-75634 Paris Cedex 13, France
[c] ERUDITE, Université Paris Est Créteil, 61, avenue du Général De Gaulle, 94010 Créteil, France

ARTICLE INFO

Article history:
Received 11 April 2013
Received in revised form
30 November 2013
Accepted 17 December 2013
Available online 11 June 2014

Keywords:
Kohonen algorithm
Planar graphs
Labor market
Markov chains

ABSTRACT

In this paper we address an important economic question. Is there, as mainstream economic theory asserts it, a homogeneous labor market with mechanisms which govern supply and demand for work, producing an equilibrium with its remarkable properties? Using the Panel Study of Income Dynamics (PSID) collected on the period 1984–2003, we study the situations of American workers with respect to employment. The data include all heads of household (men or women) as well as the partners who are on the labor market, working or not. They are extracted from the complete survey and we compute a few relevant features which characterize the worker's situations.

To perform this analysis, we suggest using a Self-Organizing Map (SOM, Kohonen algorithm) with specific structure based on planar graphs, with disconnected components (called D-SOM), especially interesting for clustering. We compare the results to those obtained with a classical SOM grid and a star-shaped map (called SOS). Each component of D-SOM takes the form of a string and corresponds to an organized cluster.

From this clustering, we study the trajectories of the individuals among the classes by using the transition probability matrices for each period and the corresponding stationary distributions.

As a matter of fact, we find clear evidence of heterogeneous parts, each one with high homogeneity, representing situations well identified in terms of activity and wage levels and in degree of stability in the workplace. These results and their interpretation in economic terms contribute to the debate about flexibility which is commonly seen as a way to obtain a better level of equilibrium on the labor market.

1. Introduction

The aim of this study is to identify and to analyze the succession of situations occupied by workers on a modern labor market (1984–2003), it is an extended version of [1].

Basically the dominant economic theory (neo-classical sense) is based on the concept of the labor market where supplies (individuals) and demands (firms) meet. An equilibrium price (salary) makes the adequacy of supply and demand [2]. This theory defines mechanisms explaining labor supply by the wage level and predicting the stability of the relation between a firm and a worker and its evolution over time (a career). Unfortunately, these mechanisms are not observed in most real situations.

This is the pure neo-classical model. To get closer to the real economy, many developments have been made in the representation of the behavior of economic agents, in particular with the theory of job search taking into account different types of imperfections (incomplete information, the presence of institutions, regulation of relations between firms and employees, etc.), see for example chapter 39 in [3]. But the basic hypothesis is unchanged: a single market whose functioning is flawed with the constraints and inefficiencies caused by the actual conditions. The result is closer to the real economy, but the deeper understanding of the phenomena that affect this sector of the economy in the last 30 years and their dynamics in the changing conditions of this period cannot be properly identified.

Instead of completing the pure neo-classical market model by a set of constraints more or less complex, the idea of this work is to show that the market is not homogeneous, it is the assembly of parts whose main characteristics are very different: that is the meaning of the assertion that the labor market is heterogeneous as mentioned in [4] and [5]. To find evidence of this heterogeneity, we construct a classification of the labor market observed over 20 years: this should identify the specific characteristics of each component and, secondly, permit to observe the situation of employees, over time, in these specific markets.

Our contribution consists in the identification and characterization of each class essentially using the variables used for classification

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2013.12.058
0925-2312/

(as well as some qualitative variables, subject to a satisfactory quality of this information.) A result to be expected is the dynamic view of trajectories of employees between these classes which can be obtained by observing the transition matrices between classes.

Let us identify the diversity of situations in terms of activity. A "situation" is defined by a combination of quantitative variables as shown in these two examples:

- full time job for the whole year, high wages, seniority in the same job;
- precarious conditions, wages lower than the average, part-time jobs, short-term contracts, on-call jobs, holding of a second job.

On the basis of individual characteristics, we construct a classification of situations observed every 2 years on a specific labor market, the US labor market over two consecutive periods of nine (1984–1992) and eleven (1993–2003) years respectively. So for each individual, we can observe the sequence of the classes he belongs to. That is what we call *professional trajectories*.

We need to study the temporal changes to answer some important questions linked to the evolution of the macroeconomic environment. Recall that in 1992, the end of Reaganomics and the beginning of Clinton period lead to a global reduction of unemployment during the rest of the decade. And this change leads to ask several questions. For instance what are the real changes at the individual level after 1992? More generally, what is the impact of a reduction of unemployment associated with "new forms of employment" . And also are there different conclusions if we observe the specific situation of women on the labor market?

This article follows another paper [6] but contains necessary material (and possibly redundant) to be self-contained. It is organized as follows: Section 2 presents the data and the notations used throughout the paper. The methodology and the global architecture of the proposed procedure are described in Section 3. Section 4 discusses how to choose the more efficient topology for the map. In Sections 5 and 6, the classes are analyzed from an economic point of view. Finally Section 7 presents the transitions from one class to another, according to the period and gender. Section 8 is devoted to a discussion of recent articles and to a conclusion which summarizes the main results.

2. The data: first period (1984, 86, 88, 90, 92) and second period (93, 95, 97, 99, 2001, 03)

We use the PSID (Panel Study of Income Dynamics),[1] dividing the observations in two periods in order to meet two objectives: on one hand to observe a number of workers large enough to obtain statistical indicators representative of the whole population and on the other hand to keep only individuals present all along each period to identify trajectories.

We create a sample for each period (1984–1992, 1993–2003). By looking at descriptive statistics for the quantitative variables for each period, we can assume that both periods have the same rough characteristics. So we can make the classifications with all the observations together.

In the PSID, we select households for which the head (man or woman) is present in the household every year of the period and we do it separately for each period. The administrative rule is that if there is a male in the household, he is the head, if not the head is a woman. Fortunately quite the same variables concerning the activity on the labor market are available for the wife/partner of the head, if there is one. Retrieving this information, we constitute a set of individuals (3965 in period 1, and 3607 in period 2) observed every 2 years in each period, with a proportion of women close to that observed in the whole population.

An observation consists of a couple (year, individual). Each one is described by 8 quantitative variables and 2 qualitative variables. See Table 1 for the list of variables and their meaning.

The pre-processing consists of removing observations with clearly inconsistent values such as a number of weeks per year greater than 52. After this filtering 41,467 observations constitute our working database. Observed current wages are converted in real dollars using the Price Index of PIB in 1992 (first period) or 2003 (second period). Eventually, the 8 quantitative variables were centered and reduced to standardize the order of magnitude. We can compute the correlation matrix of these variables, displayed in Table 2.

We observe that variables *Number of worked hours per week* (nbhtrav), *Number of worked weeks* (nbstrav), *Hourly wages in dollars* (salhor) and *Seniority in current work in months* (anctrav) are strongly positively correlated, and that they are opposite to *Number of unemployed weeks* (nbschom) and *Number of weeks out of the labor market* (nbsret). The variables related to *extrajobs* are not correlated with the others.

3. SOM, disconnected self-organizing maps (D-SOM), self organizing star (SOS)

3.1. The Kohonen algorithm (SOM)

In its classical presentation [7,8], the SOM algorithm is an iterative algorithm, which iterates the two following steps over training patterns $\mathbf{x}_j$ for computing the set of code-vectors $\mathbf{m}_i, i \in \{1, \ldots, K\}$ which define the map:

- *Competitive step,* this step aims at finding the best matching unit (BMU) for sample $\mathbf{x}_j$:

$$c = \arg \min_{i \in \{1,\ldots,K\}} ||\mathbf{x}_j - \mathbf{m}_i||. \quad (1)$$

Table 1
Variable name, description and type, for PSID dataset.

Name	Description	Min-Max	Type
nbhtrav	Number of worked hours per week	0–112	Quant
nbstrav	Number of worked weeks	0–52	Quant
nbschom	Number of unemployed weeks	0–52	Quant
nbsret	Number of weeks out of the labor market	0–52	Quant
salhor	Hourly wages in real dollars	0–83.85	Quant
nbex	Number of extra jobs	0–5	Quant
hortex	Number of hours worked in extra jobs	0–1664	Quant
anctrav	Seniority in current job in months	0–780	Quant
Gender	Gender	2 modalities	Qual
Age	Age group (<30, 30–45, >45)	3 modalities	Qual

Table 2
Correlation matrix of the quantitative variables.

Variables	*nbhtrav*	*nbstrav*	*nbschom*	*nbsret*	*salhor*	*nbex*	*hortex*	*anctrav*
nbhtrav	1	0.72	−0.04	−0.14	0.36	0.05	0.01	0.23
nbstrav	0.72	1	−0.23	−0.30	0.38	0.06	0.01	0.30
nbschom	−0.04	−0.23	1	0.02	−0.09	−0.01	−0.01	−0.11
nbsret	−0.14	−0.30	0.02	1	−0.10	−0.04	−0.04	−0.12
salhor	0.36	0.38	−0.09	−0.10	1	0.07	0.05	0.31
nbex	0.05	0.06	−0.01	−0.04	0.07	1	0.72	0.00
hortex	0.01	0.01	−0.01	−0.04	0.05	0.72	1	−0.01
anctrav	0.23	0.30	−0.11	−0.12	0.31	0.00	−0.01	1

[1] Available online at http://psidonline.isr.umich.edu/.