

# A Longitudinal Analysis of Labour Market Data with SOM

**Patrick Rousset**  
*CEREQ, France*

**Jean-Francois Giret**  
*CEREQ, France*

## INTRODUCTION

The aim of this paper is to present a typology of career paths in France drawn up with the Kohonen algorithm and its extension to a clustering method of life history analysis based on the use of Self Organizing Maps (SOMs). Several methods have previously been presented for transforming qualitative into quantitative information so as to be able to apply clustering algorithms such as SOMs based on the Euclidean distance. Our approach consists in performing quantitative encoding on labor market situation proximities across time. Using SOMs, the preservation of the topology also makes it possible to check whether this new method of encoding preserves the particularities of the life history according to our economic approach to careers. Lastly, this quantitative encoding preprocessing, which can be easily applied to analysis methods of life history, completes the set of methods extending the use of SOM to qualitative data.

## BACKGROUND

Several methods are generally used to study the dynamic aspects of careers. The first method, which estimates some reduced-form transition models, has been extensively used in labor microeconometrics, using event-history models for continuous-time data or discrete-time panel data with Markov processes. Those of the second kind, which include the method presented here, are sequence analysis methods dealing with complex information about individual labor market histories, such as the various states undergone, the duration of the spells, multiple transitions between the states, etc.. The idea was to empirically generate a statistical typology of sequences by performing cluster analysis (Lebart, 2006). This method thus makes it pos-

sible to define “cluster paths” constituting endogenous variables and explained in terms of individual characteristics such as gender, educational level or parental socio-economic status. The optimal matching method, which has been widely used in social science since the pioneering paper by Abott (Abbott & Hrycak, 1990), is an attractive solution for analysing longitudinal data of this kind. The basic idea underlying this method is to take a pair of sequences and calculate the cost of transforming them into each other by performing a series of elementary operations (insertion, deletion and substitution). However, this method has been heavily criticized because it may be difficult to determine the values of these elementary operations. Here we adopt another strategy. First, in order to classify sequences into groups, we have defined a measure of the distance between each trajectory, which is coherent with our data and with some well-known theoretical hypotheses in the field of labor economics. We then use Self Organizing Maps (the Kohonen algorithm) for classification and purposes.

Self Organizing Maps (see Kohonen, 2001, Fort, 2006) are known to be a powerful clustering and projection method. Since this method accounts efficiently for changes occurring with time, SOMs yield accurate predictions (see for example Cottrell, Girard & Rousset, 1998, Dablemont, Simon, Lendasse, Ruttiens, Blayo & Verleysen, 2003, Souza, Barreto & Mota, 2005). Life histories can be considered as a qualitative record of information, while SOMs are based on Euclidean distance. Many attempts have been made to transform qualitative variables into quantitative ones: using for example the Burt description (see the KACM presentation in Cottrell & Letremy, 1995) or using the multidimensional scaling (Miret, Garcia-Lagos, Joya, Arazoza & Sandoval, 2005). In our approach, the quantitative recoding focuses on the proximity between items considering particularities of the data (a life his-

tory) according to our economic approach. When the preprocessing of recoding is performed, SelfOrganizing Maps is a useful clustering tool, first considering its pre-mentioned clustering and projection qualities and also because of its ability to make the efficiency of our new encode emerge.

## CLUSTERING LIFE HISTORY WITH SOM

### An Example of a Life History

#### Career Paths

Labor economists have generally assumed that the beginning of a career results from a matching process (Jovanovich, 1979). Employers and job seekers lack information about each other: employers need to know how productive their potential employee is and job applicants want to know whether the characteristics of the job correspond to their expectations. Job turnover and temporary employment contracts can therefore be viewed as the consequences of this trial-by-error process. However, individuals' first employment situations may also act as a signal of employability to the labor market. For example, a long spell of unemployment during the first years in a person's career may be interpreted by potential employers as sign of low work efficiency; whereas working at a temp agency may be regarded as a sign of motivation and adaptability. This is consistent with the following path dependency hypothesis: the influence of past job experience on the subsequent career depends on the "cost" associated with the change of occupational situation. However, empirical studies have shown that employers mainly recruit on the basis of recent work experience (Allaire, Cahuzac & Tahar, 2000). The effects of less recent employment situations on a person's career therefore decrease over time.

#### Data

The data used in this study were based on the "Generation 98" survey carried out by Céreq: 22 000 young people who had left initial training in 1998 at all levels and in all training specializations were interviewed in spring 2001 and 2003 and autumn 2005. This sample

was representative of the 750 000 young people leaving the education system for the first time that year in France. This survey provided useful information about the young people's characteristics (their family's socio-economic status, age, highest grade completed, highest grade attended, discipline, any jobs taken during their studies, work placement) and the month-by-month work history from 1998 to 2005. We therefore have a complete and detailed record of the labor market status of the respondents during the 88-month period from July 1998 to November 2005. Employment spells were coded as follows, depending on the nature of the labor contract: 1 = permanent labor contract, 2 = fixed term contract, 3 = apprenticeship contract, 4 = public temporary labor contract, 5 = interim/templing). Other unemployed situations were coded as follows: 6 = unemployment, 7 = inactivity, 8 = military service, 9 = at school.

### Preprocessing Phase: Life History Encoding

The encoding of the trajectories involved a two-step preprocessing phase : defining a distance between states including time dynamics and the resulting quantitative encoding of trajectories. These two steps refer to the specificity of the data set structures of life history samples: the variables items (the states) are some qualitative information while the variables order records some quantitative information (the timing and the duration of events).

#### The Distance Between Situations

Working with pairs (state, time), called situations, allows to include the time dynamics in the proximities between occupational states. The proximity between two situations is measured on the basis of their common future, in line with our Economic approach. A situation is assumed as a potential for its own future, depending on its influence on this future. The similarity between two situations is deduced from comparisons between their referring potential. The *potential future*  $P^S$  of a situation  $S$  among  $n$  monthly periods and  $p$  states is defined as the  $p \times n$  dimensional vector given in (1). Its components  $P_s^{S'}$  are the product of terms  $\varphi$  and  $\beta$ .  $\varphi$  measures the flow between situation  $S$  and any situation  $S'$  as the empirical probability of reaching  $S'$  starting

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/longitudinal-analysis-labour-market-data/10369](http://www.igi-global.com/chapter/longitudinal-analysis-labour-market-data/10369)

## Related Content

---

### An Intuitionistic Fuzzy Multi-Criteria Decision-Making Method Based on an Exponential-Related Function

Daniel Aikhuele and Faiz Turan (2017). *International Journal of Fuzzy System Applications* (pp. 33-46). [www.irma-international.org/article/an-intuitionistic-fuzzy-multi-criteria-decision-making-method-based-on-an-exponential-related-function/189128](http://www.irma-international.org/article/an-intuitionistic-fuzzy-multi-criteria-decision-making-method-based-on-an-exponential-related-function/189128)

### A Novel Long Short-Term Memory Method for Model for Stock Price Prediction

K. S. Archana, B. Sivakumar, B. Ebenezer Abishek, Shaik Ghouhar Taj, V. Kavitha Reddy and A. Vijayalakshmi (2023). *Handbook of Research on Advancements in AI and IoT Convergence Technologies* (pp. 73-87). [www.irma-international.org/chapter/a-novel-long-short-term-memory-method-for-model-for-stock-price-prediction/330060](http://www.irma-international.org/chapter/a-novel-long-short-term-memory-method-for-model-for-stock-price-prediction/330060)

### Using Event B to Specify Context Awareness for Service Discovery in Pervasive Environments

Karima Belgharbi and Mahmoud Boufaïda (2017). *International Journal of Ambient Computing and Intelligence* (pp. 1-22). [www.irma-international.org/article/using-event-b-to-specify-context-awareness-for-service-discovery-in-pervasive-environments/176711](http://www.irma-international.org/article/using-event-b-to-specify-context-awareness-for-service-discovery-in-pervasive-environments/176711)

### Development and Evaluation of Intelligent Agent- Based Teaching Assistant in E-Learning Portals

Saeed Rouhani and Seyed Vahid Mirhosseini (2018). *Intelligent Systems: Concepts, Methodologies, Tools, and Applications* (pp. 798-809). [www.irma-international.org/chapter/development-and-evaluation-of-intelligent-agent--based-teaching-assistant-in-e-learning-portals/205809](http://www.irma-international.org/chapter/development-and-evaluation-of-intelligent-agent--based-teaching-assistant-in-e-learning-portals/205809)

### Understanding Phatic Aspects of Narrative when Designing Assistive and Augmentative Communication Interfaces

Benjamin Slotznick (2014). *International Journal of Ambient Computing and Intelligence* (pp. 75-94). [www.irma-international.org/article/understanding-phatic-aspects-of-narrative-when-designing-assistive-and-augmentative-communication-interfaces/147384](http://www.irma-international.org/article/understanding-phatic-aspects-of-narrative-when-designing-assistive-and-augmentative-communication-interfaces/147384)