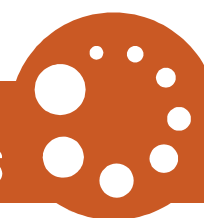




ML_SoSi: Individual trajectories in the social security system

Summary

EXPERIMENTAL STATISTICS



Neuchâtel, 2019

Published by: Federal Statistical Office (FSO)
Contents: Luzius von Gunten, Nora Meister,
Philippe Meyer, Thomas Ruch
Topic: 00 Statistical Basis and Overviews
Original text: German
Translation: FSO language services

Layout concept: Section DIAM
Downloads: www.statistics.ch
Copyright: FSO, Neuchâtel 2019
Reproduction with mention of source authorised
(except for commercial purposes)



Pilotprojekt ML_SoSi: Summary

30 June 2019

Unemployment trajectories can vary widely for those affected. The resulting individual trajectories can be shaped by (repeated) drawing of welfare benefits from the social security system (unemployment, invalidity insurance, and social assistance), through employment or by withdrawing from working life. This project analyses information on individual trajectories using inductive statistical methods to identify typical patterns. It uses the SHIVALV+IK dataset, which contains monthly information on the drawing of individual benefits from social assistance (SH), invalidity insurance (IV) and unemployment insurance (ALV), and also on employment (IK).

The total population is a cohort of persons aged between 18 and 64 who first drew unemployment benefits (ALV) in 2010. All information on the receipt of social benefits and on employment for the subsequent 48 months was incorporated in the analysis.

Using sequence clustering, similar trajectories were grouped together based on an algorithm and typical patterns could be identified without them being known in advance (unsupervised machine learning). This inductive method revealed 10 clusters of typical patterns. Each of these clusters was then analysed and interpreted in detail using trajectory indicators and graphical representations (state distribution plots and Sankey plots).

Among several clusters, people are re-integrated into the labour market after a period of unemployment (Fig. 1: clusters 3, 6 and 7). The difference between these clusters lies in the duration of the unemployment. One cluster is also clearly differentiated with income from work while registered unemployed in which the average duration of receipt of ALV unemployment benefits is 22 months in total (Fig. 1: cluster 9). There are also clusters that show clear tendencies towards permanent drawing of either IV pensions or social assistance benefits (Fig. 1: clusters 2, 8 and 10). In one of these (cluster 10), income from employment is supplemented by social assistance benefits. The average duration of receipt of invalidity pensions/social assistance benefits is over 30 months in all three clusters. Another cluster comprises people who permanently disappear from the systems being studied during the observation period (Fig. 1: cluster 1). This can be for different reasons: either they took part in a course, embarked on training, withdrew from the labour market or left the country. In addition, two other clusters emerge which combine heterogeneous trajectories (Fig. 1: clusters 4 and 5).

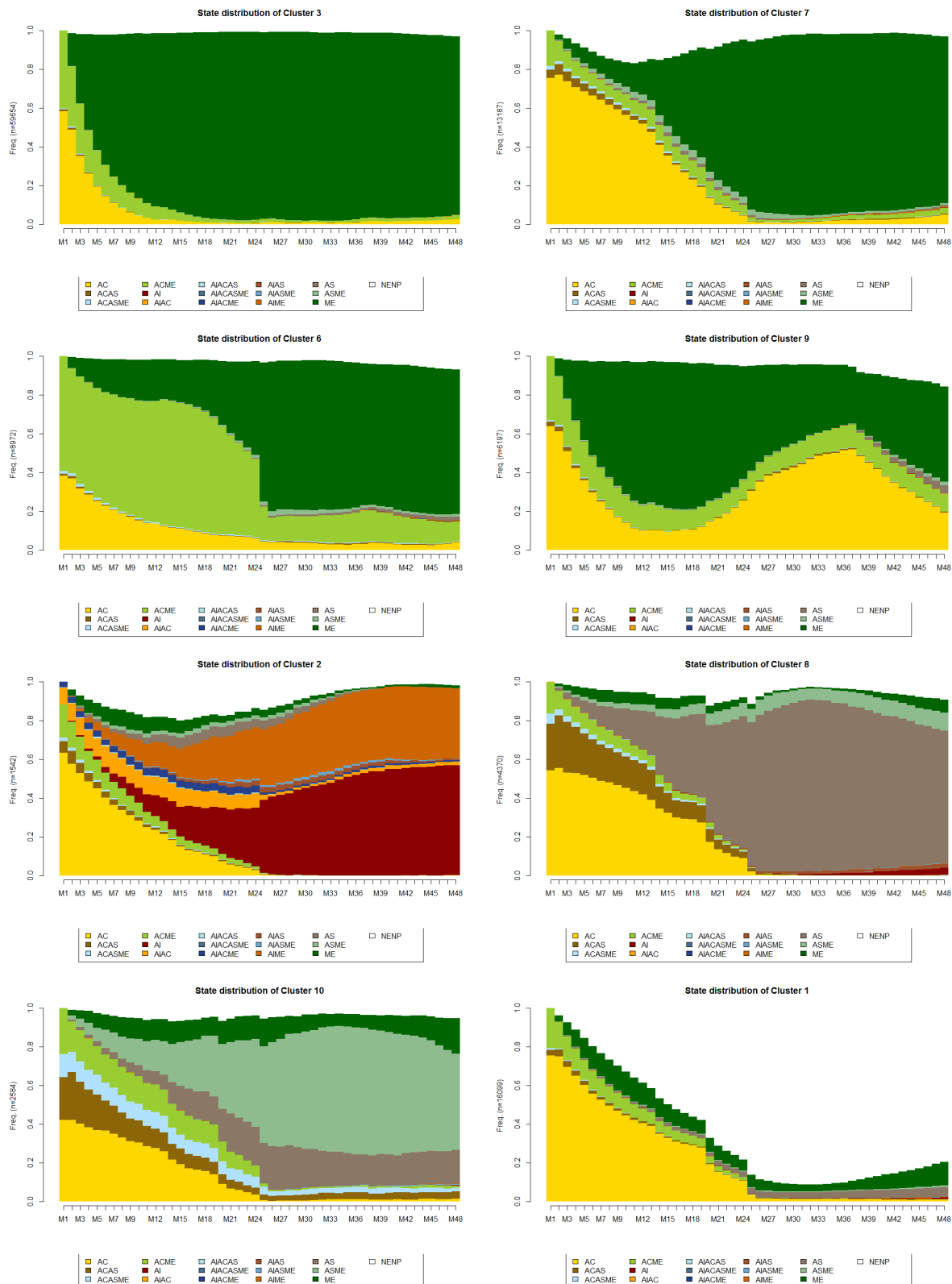
This project has shown that sequence clustering is a promising method with which to produce substantively valid and analytically relevant results. It significantly reduces the complexity of the trajectory data and therefore broadens the analysis options by detecting patterns that are not anticipated. In the next phase of the project, sensitivity analyses will be introduced for different cohorts as well as for each individual data processing steps. The project will then also examine how a temporally stable cluster model can be transferred to new cohorts. In a second phase, prediction models are to be used to assess the probability of each person in the cohort to belong to each of the ten trajectory clusters (supervised machine learning). The models are developed (trained) on cohorts where as the dependent variable the cluster affiliation of each person is known. These models can then be applied to cohorts whose first receipt of unemployment benefits is more recent. In this way and depending on the robustness of the models, the prospective trajectories can be identified at an early stage.

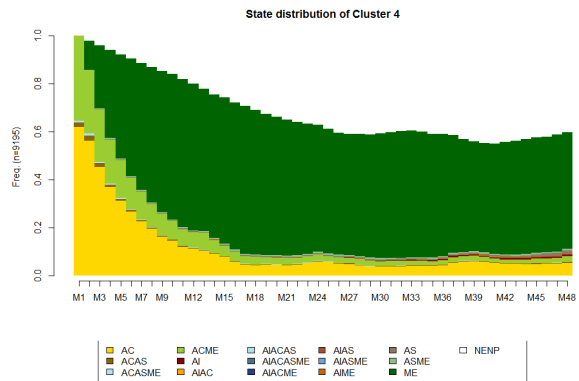
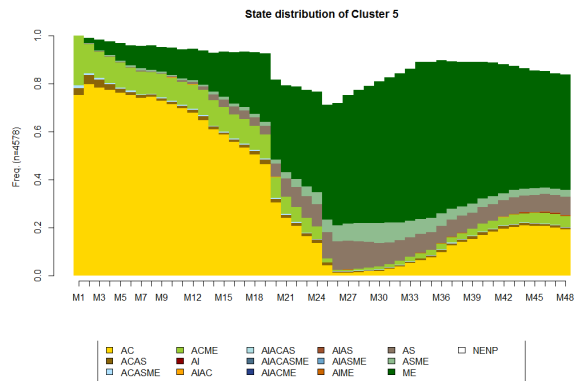
Complementary analytical methods used:

- kmeans and optimal matching (agnes with optimal matching distance matrix resulting from seqdist) for sequence clustering
- State distribution plots (TraMineR package) and Sankey plots for visualisation.



Figure 1: state distribution plots for ten trajectory clusters, 2010 cohort





AC	=	Unemployment insurance (ALV)
ACME	=	Unemployment insurance (ALV) + employment
AIACAS	=	Invalidity insurance (IV) + unemployment insurance (ALV) + social assistance (SH)
AIAS	=	Invalidity insurance (IV) + social assistance (SH)
AS	=	Social assistance (SH)
ACAS	=	Unemployment insurance (ALV) + social assistance (SH)
AI	=	Invalidity insurance (IV)
AIACASME	=	Invalidity insurance (IV) + unemployment insurance (ALV) + social assistance (SH) + employment
AIASME	=	Invalidity insurance (IV) + social assistance (SH) + employment
ASME	=	Social assistance (SH) + employment
ACASME	=	Unemployment insurance (ALV) + social assistance (SH) + employment
AIAC	=	Invalidity insurance (IV) + unemployment insurance (ALV)
AIACME	=	Invalidity insurance (IV) + unemployment insurance (ALV) + employment
AIME	=	Invalidity insurance (IV) + employment
ME	=	Employment
NENP	=	No employment + no welfare benefit

Source: SHIVALV-IK 2010-2015, own calculations