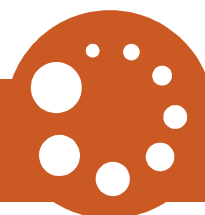




Verläufe im System der Sozialen Sicherheit

Zusammenfassung

EXPERIMENTAL STATISTICS



Neuchâtel, 2019

Herausgeber: Bundesamt für Statistik (BFS)
Inhalt: Luzius von Gunten, Nora Meister,
Philippe Meyer, Thomas Ruch
Themenbereich: 00 Statistische Grundlagen
Originaltext: Deutsch
Übersetzung: Sprachdienste BFS

Layoutkonzept: Sektion DIAM
Download: www.statistik.ch
Copyright: BFS, Neuchâtel 2019
Wiedergabe unter Angabe der Quelle
für nichtkommerzielle Nutzung gestattet



Pilotprojekt ML_SoSi: Zusammenfassung

30. Juni 2019

Arbeitslosigkeit kann für die betroffenen Personen sehr unterschiedlich verlaufen. Die so entstehenden Verlaufsbiographien sind geprägt durch (wiederholte) Sozialleistungsbezüge aus dem System der Sozialen Sicherheit (Arbeitslosen-, Invalidenversicherung, Sozialhilfe), Erwerbsarbeit oder auch Rückzug aus dem Erwerbsleben. Im vorliegenden Projekt sind Angaben zu den individuellen Verläufen unter Anwendung induktiver statistischer Methoden analysiert worden, um typische Verlaufsmuster zu identifizieren. Dazu wurde der Datensatz SHIVALV+IK verwendet. Dieser enthält monatsgenaue Informationen zu individuellen Sozialleistungsbezügen aus der Sozialhilfe (SH), der Invalidenversicherung (IV) und der Arbeitslosenversicherung (ALV), sowie zur Erwerbstätigkeit (IK).

Als Grundgesamtheit diente eine Kohorte von Personen zwischen 18 und 64 Jahren, die im Jahr 2010 erstmals Taggelder der Arbeitslosenversicherung (ALV) bezogen hatten. Alle Informationen zum Bezug von Sozialleistungen sowie zur Erwerbstätigkeit während den folgenden 48 Monaten sind in die Analyse miteinbezogen worden.

Mit Sequence Clustering werden ähnliche Verläufe auf der Basis eines Algorithmus gruppiert und so typische Verlaufsmuster identifiziert, ohne diese im Vorherein zu kennen (unsupervised machine learning). Mittels diesem induktiven Verfahren haben sich 10 Cluster typischer Verlaufsmuster herausgebildet. Jedes dieser Cluster wurde anschliessend mit Verlaufsindikatoren und grafischen Darstellungen («State Distribution Plots», «Sankey Plots») vertieft analysiert und inhaltlich interpretiert. Es zeichnen sich mehrere Cluster ab, in denen sich die Personen nach einer Phase der Arbeitslosigkeit wieder in den Arbeitsmarkt integrieren (Abb. 1; Cluster 3, 7 und 6). Untereinander unterscheiden sich diese Cluster nach der Dauer der Arbeitslosigkeit. Deutlich wird auch die Ausdifferenzierung eines Clusters mit Zwischenverdienst in dem die durchschnittliche Dauer des Bezugs von ALV gesamthaft 22 Monate beträgt (Abb. 1; Cluster 9). Daneben entstehen Cluster, die klare Tendenzen entweder zum dauerhaften Bezug von IV-Renten oder Leistungen der Sozialhilfe zeigen (Abb. 1; Cluster 2 und 8, 10). In einem davon (Cluster 10) wird das Einkommen aus Erwerbstätigkeit durch Leistungen der Sozialhilfe ergänzt. Die durchschnittlichen Dauern von IV-Renten respektive des Bezugs von Leistungen der Sozialhilfe betragen in allen drei Clustern über 30 Monate. Ein anderes Cluster vereint diejenigen Personen, die während des Beobachtungszeitraumes dauerhaft aus den untersuchten Systemen hinausfallen (Abb. 1; Cluster 1). Unter anderem durch Teilnahme an einer Bildungsmassnahme wegen Aufnahme einer Ausbildung, Rückzug aus dem Arbeitsmarkt oder Ausreise. Daneben entstehen zwei kleinere Cluster, die heterogene Verläufe vereinen (Abb. 1; Cluster 5 und 4).

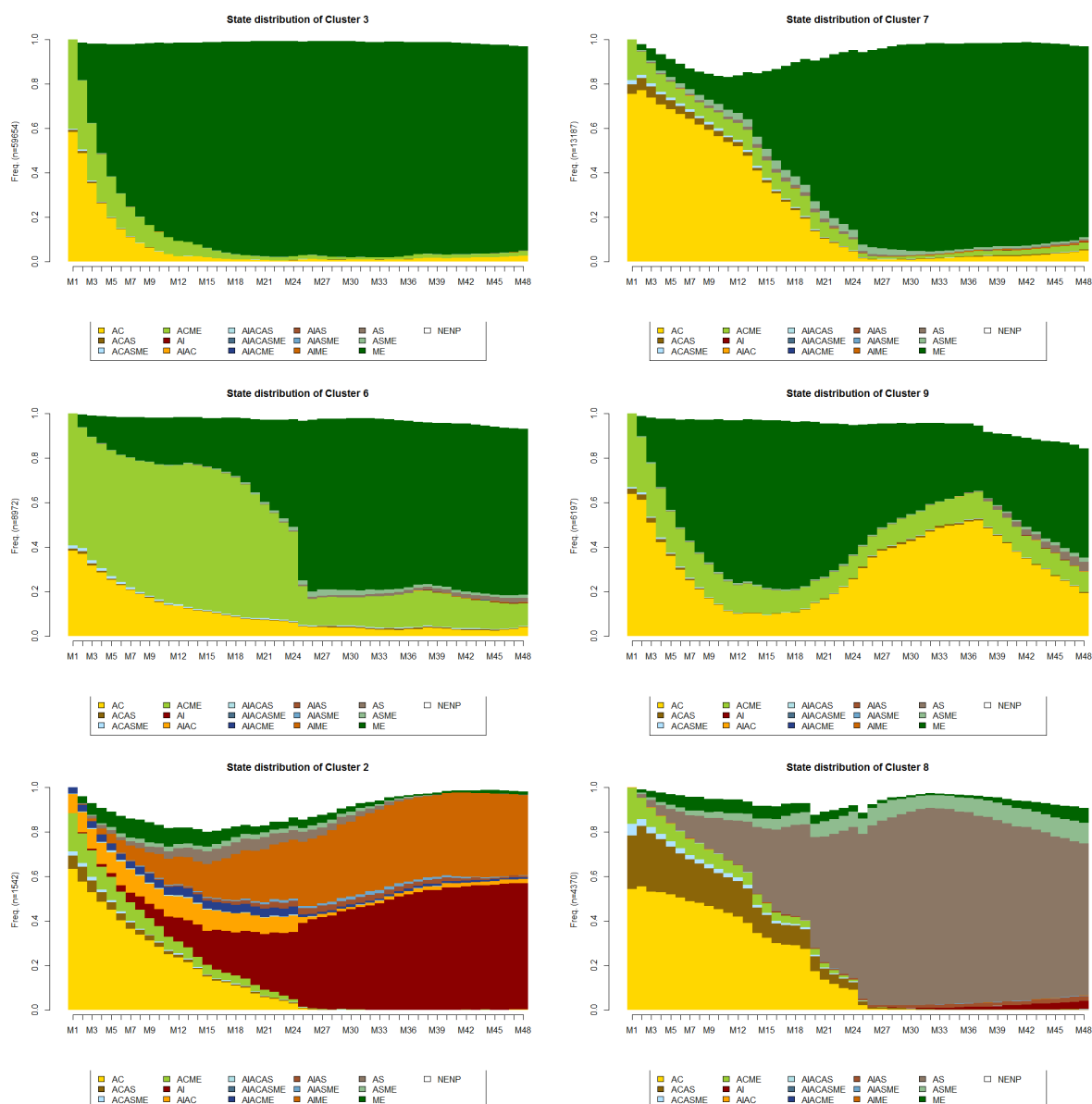
Das Projekt hat somit gezeigt, dass das «Sequence Clustering» ein vielversprechendes Verfahren ist, um inhaltlich valide und analytisch relevante Resultate zu erzeugen. Es erlaubt eine deutliche Verringerung der Komplexität der Verlaufsdaten und erweitert damit die Analysemöglichkeiten durch die Erkennung von Mustern, die deduktiv nicht antizipiert werden. Im weiteren Verlauf des Projektes werden Sensitivitätsanalysen für unterschiedliche Kohorten sowie einzelner Datenaufbereitungsschritte durchgeführt. Anschliessend wird geprüft, wie ein zeitlich stabiles Clustermodell auf neue Kohorten übertragen werden kann. In einem zweiten Projektteil soll mithilfe von Prädiktionsmodellen für jede Person in der Kohorte die Wahrscheinlichkeit für die Zugehörigkeit zu jedem der zehn Verlaufscluster geschätzt werden (supervised machine learning). Die Modelle werden auf denjenigen Kohorten entwickelt (trainiert), bei welchen als abhängige Variable die Clusterzugehörigkeit für jede Person bekannt ist. Diese Modelle können anschliessend auf Kohorten angewendet werden, deren Erstbezug von ALV weniger weit in der Vergangenheit liegt. Auf diese Art können – je nach Robustheit der Modelle – die voraussichtlichen Verläufe frühzeitig erkannt werden.

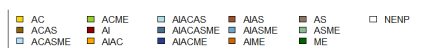
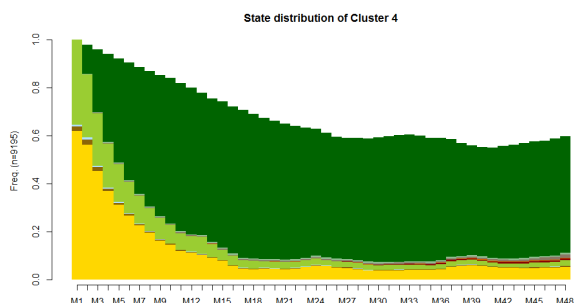
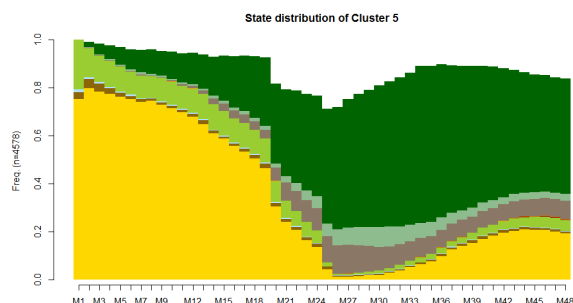
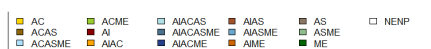
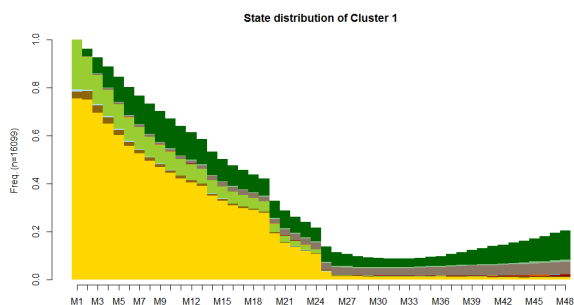
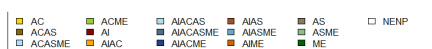
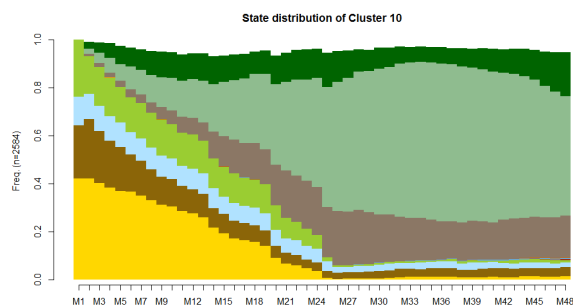


Verwendete komplementäre Analysemethoden:

- kmeans und optimal matching (agnes mit optimal matching Distanzmatrix aus seqdist resultierend) für «Sequence Clustering»
- «State Distribution Plots» (package TraMineR) und «Sankey Plots» für die Visualisierung.

Abbildung 1: state distribution plots für zehn Verlaufskluster, Kohorte 2010





AC	=	Arbeitslosenversicherung (ALV)
ACME	=	Arbeitslosenversicherung (ALV) + Erwerbstätigkeit
AIACAS	=	Invalidenversicherung (IV) + Arbeitslosenversicherung (ALV) + Sozialhilfe (SH)
AIAS	=	Invalidenversicherung (IV) + Sozialhilfe (SH)
AS	=	Sozialhilfe (SH)
ACAS	=	Arbeitslosenversicherung (ALV) + Sozialhilfe (SH)
AI	=	Invalidenversicherung (IV)
AIACASME	=	Invalidenversicherung (IV) + Arbeitslosenversicherung (ALV) + Sozialhilfe (SH) + Erwerbstätigkeit
AIASME	=	Invalidenversicherung (IV) + Sozialhilfe (SH) + Erwerbstätigkeit
ASME	=	Sozialhilfe (SH) + Erwerbstätigkeit
ACASME	=	Arbeitslosenversicherung (ALV) + Sozialhilfe (SH) + Erwerbstätigkeit
AIAC	=	Invalidenversicherung (IV) + Arbeitslosenversicherung (ALV)
AIACME	=	Invalidenversicherung (IV) + Arbeitslosenversicherung (ALV) + Erwerbstätigkeit
AIME	=	Invalidenversicherung (IV) + Erwerbstätigkeit
ME	=	Erwerbstätigkeit
NENP	=	keine Erwerbstätigkeit + keine Sozialleistung

Quelle: SHIVALV-IK 2010-2015, eigene Berechnungen