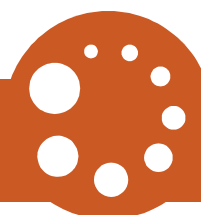




ML_SoSi: Parcours individuels dans le système de sécurité sociale

Résumé

EXPERIMENTAL STATISTICS



Neuchâtel, 2019

Éditeur: Office fédéral de la statistique (OFS)
Contenu: Luzius von Gunten, Nora Meister,
Philippe Meyer, Thomas Ruch
Domaine: 00 Bases statistiques des généralités
Langue du texte original: Allemand
Traduction: Services linguistiques de l'OFS

Concept de mise en page: Section DIAM
Téléchargement: www.statistique.ch
Copyright: OFS, Neuchâtel 2019
La reproduction est autorisée, sauf à des fins
commerciales, si la source est mentionnée



Projet pilote ML_SoSi: Résumé

Le 30 juin 2019

Le chômage peut être vécu de manière très différente d'une personne à l'autre. Les parcours des chômeurs peuvent ainsi être marqués par l'octroi (répété) de prestations sociales du système de sécurité sociale (assurance-chômage, assurance-invalidité, aide sociale), par des périodes d'activité rémunérée ou par le retrait de la vie active. Dans ce projet, les données sur les parcours individuels ont été analysées à l'aide de méthodes statistiques inductives, dans le but d'identifier des parcours types. L'analyse a été réalisée à l'aide du fichier de données SHIVALV+IK, qui contient des informations mensuelles sur les bénéficiaires de prestations de l'aide sociale (SH), de l'assurance-invalidité (IV) et de l'assurance-chômage (ALV), ainsi que sur l'activité professionnelle (IK). L'univers de base est constitué d'une cohorte de personnes âgées de 18 à 64 ans ayant perçu des indemnités journalières de l'assurance-chômage pour la première fois en 2010. L'analyse prend en compte les informations liées aux prestations sociales perçues et à l'activité professionnelle des 48 mois suivants.

En utilisant une méthode de *sequence clustering*, les parcours similaires ont été groupés sur la base d'un algorithme de manière à identifier des parcours type qui n'étaient pas connus au préalable (*unsupervised machine learning*). Ce processus inductif a permis d'identifier 10 parcours type (*clusters*). Chacun de ces groupes a fait l'objet d'une analyse approfondie et d'une interprétation du contenu à l'aide d'indicateurs longitudinaux et de représentations graphiques («State Distribution Plots», «Sankey Plots»). L'analyse fait ressortir plusieurs groupes de personnes qui réintègrent le marché du travail après une phase de chômage (ill. 1, *clusters* 3, 7 et 6). Ces groupes se différencient les uns des autres par la durée du chômage. L'analyse met aussi en évidence un groupe de personnes ayant réalisé un gain intermédiaire dont la durée moyenne des indemnités de chômage perçues se monte à 22 mois au total (ill. 1; cluster 9). D'autres groupes se démarquent par une tendance claire à bénéficier durablement de rentes AI ou de prestations de l'aide sociale (ill. 1; clusters 2 et 8, 10). Dans l'un de ces groupes, le revenu tiré d'une activité professionnelle est complété par des prestations de l'aide sociale (cluster 10). La durée moyenne des rentes AI et des prestations de l'aide sociale perçues dépasse 30 mois dans les trois groupes. Un autre groupe est composé des personnes qui sont sorties durablement des systèmes considérés pendant la période étudiée (ill. 1; cluster 1). Plusieurs raisons peuvent expliquer ce résultat : les personnes ont entamé une formation, se sont retirées du marché du travail ou ont quitté la Suisse. Il y a ensuite deux plus petits groupes réunissant des personnes ayant des parcours hétérogènes (ill. 1, clusters 5 et 4).

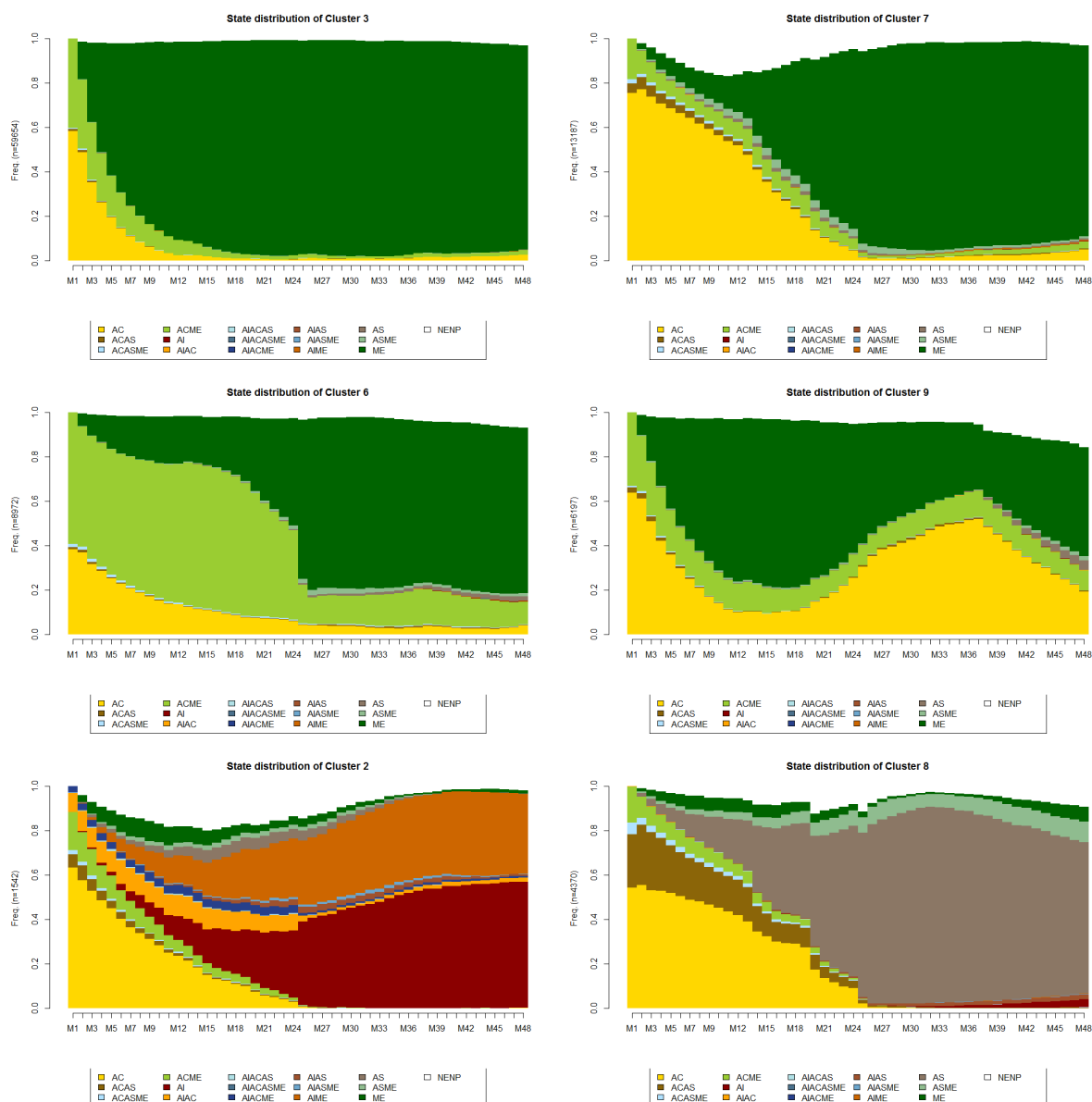
Ce projet montre ainsi que le regroupement de séquences est une méthode prometteuse pour produire des résultats valides et pertinents sur le plan analytique. Il permet de réduire considérablement la complexité des données et élargit ainsi les possibilités d'analyse grâce à l'identification de parcours types qu'il n'est pas toujours possible d'identifier de prime abord. Des analyses de sensibilité seront réalisées lors des prochaines étapes du projet sur différentes cohortes et pour chacune des étapes du traitement des données. On examinera enfin comment implémenter un modèle de *cluster* stable dans le temps sur de nouvelles cohortes. Dans une deuxième étape, à l'aide de modèles prévisionnels, une estimation de la probabilité pour chaque membre de la cohorte d'être attribué à l'un ou l'autre des dix groupes de parcours va être réalisée (*supervised machine learning*). Les modèles sont développés (entraînés) sur les cohortes pour lesquelles l'appartenance de chaque personne à un groupe (*cluster*) est connue. Le but est de pouvoir les utiliser ensuite pour des cohortes, dont la première indemnité de chômage perçue est plus récente. De cette manière, et en fonction de la robustesse du modèle, cela doit permettre de prédire d'emblée comment se développeront les parcours des individus par la suite.

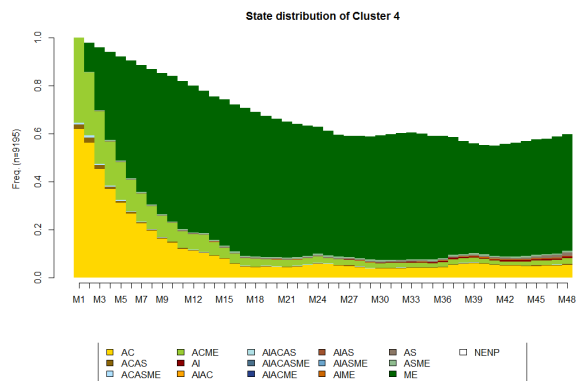
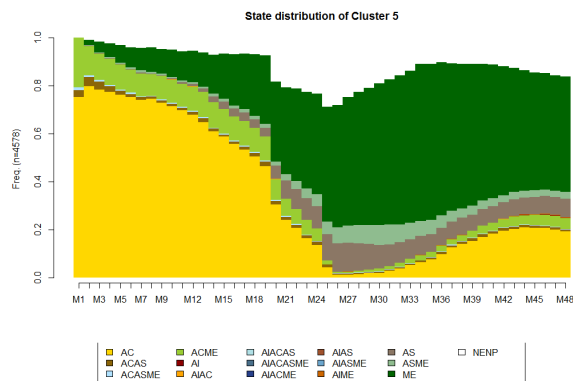
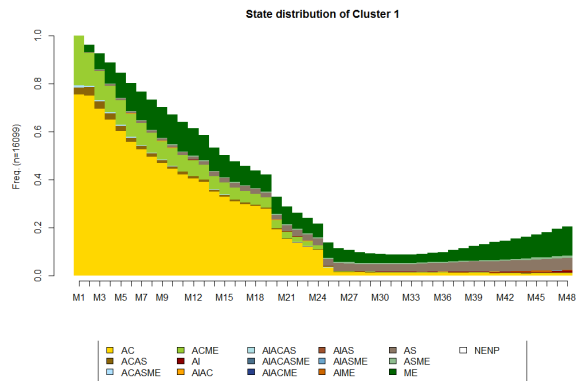
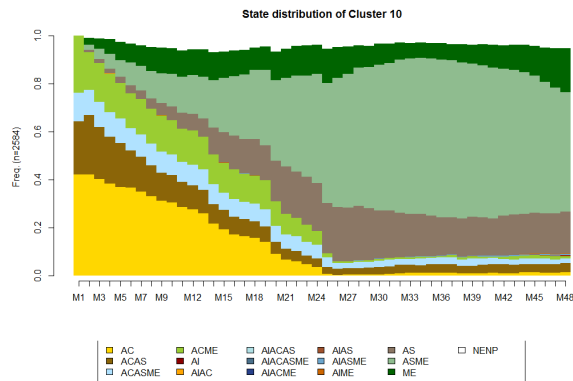


Méthodes d'analyse complémentaires utilisées:

- *kmeans* et *optimal matching* (agnes avec matrice de distance d'appariement optimal tirée de *seqdist*) pour «Sequence Clustering»
- «State Distribution Plots» (package *TraMineR*) et «Sankey Plots» pour la visualisation.

Illustration1: *state distribution plots* pour dix groupes de parcours, cohorte 2010





AC	=	Assurance-chômage (ALV)
ACME	=	Assurance-chômage (ALV) + activité professionnelle
AIACAS	=	Assurance-invalidité (IV) + assurance-chômage (ALV) + aide sociale (SH)
AIAS	=	Assurance-invalidité (IV) + aide sociale (SH)
AS	=	Aide sociale (SH)
ACAS	=	Assurance-chômage (IV) + aide sociale (SH)
AI	=	Assurance-invalidité (IV)
AIACASME	=	Assurance-invalidité (IV) + assurance-chômage (ALV) + aide sociale (SH) + activité professionnelle
AIASME	=	Assurance-invalidité (IV) + aide sociale (SH) + activité professionnelle
ASME	=	Aide sociale (SH) + activité professionnelle
ACASME	=	Assurance-chômage (ALV) + aide sociale (SH) + activité professionnelle
AIAC	=	Assurance-invalidité (IV) + assurance-chômage (ALV)
AIACME	=	Assurance-invalidité (IV) + assurance-chômage (ALV) + activité professionnelle
AIME	=	Assurance-invalidité (IV) + activité professionnelle
ME	=	Activité professionnelle
NENP	=	Pas d'activité professionnelle + pas de prestations sociales

Source: SHIVALV-IK 2010-2015, propres calculs