

International Journal of Population Data Science

Journal Website: www.ijpds.org



Describing multidimensional life course sequences capturing a child's context using vector embeddings

Maximilian Reichert¹

¹Erasmus Universitz, Rotterdam, Netherlands

The early years of childhood are among the most formative of a person's life. In this study I ask: what typical composition of contextual resources characterises the early years of a child's life? I set out to describe the contextual resources of a Dutch cohort of children born in 2013 over the course of the first 12 years using tools from natural language processing. I construct a wide variety of indicators covering several contextual life course domains, such as parental income, wealth, residential mobility, and family configurations using Dutch register data. I apply a Long-Short-Term-Memory (LSTM) recurrent neural network, to encode these multi-domain sequences into two sets

of vector embeddings for each child: First, one global vector embedding representing the entire person-sequence. Second, eleven yearly embeddings representing one person-year from 1 to 12 each. I cluster the life course sequences using the global vector embeddings and then describe the clusters. Additionally, I use the yearly embeddings to measure life course instability in contextual resources by tracing relative movement through the vector space over time. I then correlate these clusters and instability measures with parental sociodemographic circumstances measured before birth and/or children's outcomes in adolescence.

