Mach Learn
DOI 10.1007/s10994-013-5421-0

Tracking people over time in 19th century Canada for longitudinal analysis

Luiza Antonie · Kris Inwood · Daniel J. Lizotte · J. Andrew Ross

Received: 21 November 2012 / Accepted: 26 September 2013

Abstract Linking multiple databases to create longitudinal data is an important research problem with multiple applications. Longitudinal data allows analysts to perform studies that would be unfeasible otherwise. We have linked historical census databases to create longitudinal data that allow tracking people over time. These longitudinal data have already been used by social scientists and historians to investigate historical trends and to address questions about society, history and economy, and this comparative, systematic research would not be possible without the linked data. The goal of the linking is to identify the same person in multiple census collections. Data imprecision in historical census data and the lack of unique personal identifiers make this task a challenging one. In this paper we design and employ a record linkage system that incorporates a supervised learning module for classifying pairs of records as matches and non-matches. We show that our system performs large scale linkage producing high quality links and generating sufficient longitudinal data to allow meaningful social science studies. We demonstrate the impact of the longitudinal data through a study of the economic changes in 19th century Canada.

Keywords Record linkage · Classification · Historical census

Editors: Kiri Wagstaff and Cynthia Rudin.

L. Antonie (✉)
Historical Data Research Unit, University of Guelph, Guelph, Canada
e-mail: lantonie@uoguelph.ca

K. Inwood
Department of Economics and Finance, University of Guelph, Guelph, Canada
e-mail: kinwood@uoguelph.ca

D.J. Lizotte
David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada
e-mail: dlizotte@uwaterloo.ca

J. Andrew Ross
Department of History, University of Guelph, Guelph, Canada
e-mail: jaross@uoguelph.com

Published online: 01 November 2013