International Journal of Population Data Science

Journal Website: www.ijpds.org

2021 Census England And Wales: Developing Record Speed Linkage Methods to Produce Outputs in a Year

Shipsey, R^{1*}, Tomlin, A¹, White, Z¹, Plachta, J¹, Gammon, S¹, and Dygas, P¹

¹Centre for Public Health, Queen's University Belfast, Belfast, UK

Introduction

2021 will herald the next census in England and Wales. The Office for National Statistics (ONS) have a goal of publishing outputs within one year, 4 months earlier than in 2011. Since we produce estimates rather than counts, the linkage of the 2021 Census to the Census Coverage Survey which comprises 710,000 person and 370,000 household records, has to be carried out in record time (eight weeks) whilst maintaining incredibly high accuracy (less than 0.1% false positives and 0.25% false negatives).

Objectives and Approach

Our approach is to utilise the ONS Distributed Access Platform to write automated matching algorithms that are both efficient and accurate. These methods use parallelisation to speed things up, active machine learning to iteratively improve our parameters, and associative matching to squeeze every last match out automatically without impairing the accuracy.

As in 2011, we will be using clerical matchers to resolve cases that cannot be matched automatically. Speeding up the clerical matching process is imperative. We have therefore developed a pre-search algorithm that takes the hard work out of clerical matching by replacing clerical searching (here's a record can you find a match?) with clerical resolution (here are two or more records, do they match?).

Results

As a result of our improvements we estimate that we have increased our automatic matching rates from 70% to 91% for person matching, and from 60% to 95% for household matching, without loss of accuracy. However, the biggest gains in terms of speed are delivered by our pre-search algorithm which, at the current iteration, is limiting false negatives to 0.13% according to the 2011 gold standard.

Conclusion

We estimate that overall our improvements will mean that in 2021 we will need less than half the clerical resource that was required in 2011 and will meet our eight-week deadline.

*Corresponding Author:

Email Address: rachel.shipsey@ons.gov.uk (R Shipsey)