

International Journal of Population Data Science

Journal Website: www.ijpds.org



Detecting 'Lifestates' From Digital Footprint Time Series Data

Sam Smith¹, Gavin Smith¹, and John Harvey¹

¹University of Nottingham, Nottingham, United Kingdom

Introduction & Background

As digital footprints data continue to grow in complexity and volume, understanding and summarising large, high-dimensional time series is becoming increasingly important for analysing behavioural patterns. In various domains, mass data collection is routine and crucial to detecting breakpoints (significant statistical changes) and underlying patterns. Examples of such data include: transactional records; internet history records; social media data. And yet, despite this explosion of data, it remains challenging to uncover commonalities across real-world time series that are both high-dimensional and noisy.

Objectives & Approach

We introduce a new method for deciphering complex digital footprints data named ALI (Automatic Lifestate Identification), a parameter-free algorithm which aims to create interpretable summaries of the data. ALI efficiently identifies breakpoints, and clusters resulting segments, referred to as lifestates, across multiple time series. ALI uses an Expectation Maximisation (EM) algorithm that iteratively searches for breakpoints and lifestates, meaning that each is informed by the other and improved on each iteration. The result is an efficient and parsimonious characterisation of the time series; that is, a collection of lifestates describing the dataset and its patterns.

Relevance to Digital Footprints

ALI has the potential to improve how digital footprints are analysed across many different domains.

Results

This work demonstrates the practical utility of the ALI algorithm by applying it to two real-world datasets: a transactional dataset from a UK-based pharmaceutical company, and a dataset from a Sri Lankan telecommunications company. The two datasets represent the purchase history for over 18,000 new mothers, and the mobile application usage of a 'super-app' which has over 1 million downloads. This work

demonstrates how digital footprints data can be used to understand shared experiences across multiple domains. Understanding cross-individual patterns can be useful for providing tailored services to people, particularly for understanding when individuals transition from one stage of their life to another.

Conclusions & Implications

The experiments show that ALI is successful in identifying transitions into parenthood more accurately and reliably when compared to previous methods. Furthermore, ALI finds shared states across the whole set of mothers in the dataset, allowing for comparisons between the retail behaviours of different mothers before, during, and after pregnancy. The results of this work show ALI's capability to extract meaningful lifestates in challenging, real-world scenarios.

Keywords

Time Series, Segmentation, Clustering, Change Point Detection, Digital Footprints



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.