Consumer Data Research Centre

An ESRC Data Investment

Deriving the Spatial Extent of Retail Centres

Michalis Pavlis, Les Dolega and Alex Singleton
University of Liverpool

Impacts

- Explores new methods of defining the spatial extents of retail centres using a novel data source.
- Provides a means of updating retail boundaries on a regular basis, therefore providing more accurate representations of the dynamic structure of centres.
- Outcomes can inform existing boundary data that are widely used by academics, local authorities and private organisations.

Project Background

Town centres form the core of many urban areas and can be viewed as complex economic systems that are constantly evolving. The spatial extents (or boundaries) and composition of these centres typically expand and contract over time, with changes being linked to planning, rising property values, levels of accessibility and other forces such as economic shocks and the rise of Internet sales.

The classification of shopping destinations and their spatial extents is essential if we want to gain a better understanding of the relationship between retail space and changing consumer behaviours. It has long been recognised that the structure of shopping destinations are complex, with retail centre size, function, competition and agglomeration factors influencing their attraction and market potential. However, research has typically been constrained by availability of data and current models are unable to represent their constantly evolving structures.

The availability of more accurate and comprehensive spatial data regarding retail unit locations in Great Britain (GB) has improved significantly in recent years, providing scope for exploring new methods of defining their spatial extents. In particular, definitions are required that are consistent, updatable and cover a national extent. In collaboration with the Local Data Company (LDC), the objective of this research was to develop a method to address these challenges. The work evaluates a number of clustering methods and compares outputs to existing retail centre boundary data. It then presents a novel methodology that hopes to provide better spatial coverage and represent the dynamic characteristics of retail centres.

Data and Methods

National retail centre location data were provided by the LDC, describing over 500,000 retail destinations across GB. These were collected via a large pool of local surveying teams in 2015 and contained detailed information about the current occupiers, type of retailers, vacant outlets and their locations.

Additional data regarding retail areas were obtained for comparative purposes. Firstly, local authority boundary data were acquired, that were produced as part of the National Planning Policy Framework's (2012) retail centre "health checks" in the UK. Secondly, boundaries for the top 339 retail centres were acquired from the company Geolytix, who provide frequently updated information regarding centres across in the UK. These data were created using multiple variables (including the locations of retail units) and although only represented a subset of the total retail boundaries, provided an additional sample of retail areas suitable for comparison.

A key task for this research was to identify a suitable spatial clustering technique that would enable automated delineation of realistic retail agglomeration boundaries. Five of the most frequently used clustering methods were compared across 8 representative locations in Great Britain were selected based on their retail location density, size and morphology. The DBSCAN clustering algorithm was chosen as it provided the most accurate representation of retail areas relative to formal definitions. However, in order to address a well-known issue that DBSCAN does not perform well in areas of varying densities, the method was adapted so that it could be iteratively applied within smaller more homogeneous sites (using a k-NN sparse graph representation of the retail locations).

Retail clusters were then created from the LDC data using the modified DBSCAN algorithm and compared to the Geolytix boundaries. Given that the creation of polygon boundaries would result in an additional source of error, it was decided to compare the Geolytix boundaries against the clustered retail unit locations only.

Key Findings

Implementing this alternative dataset and methodology produced results that were in most areas similar to the spatial extents produced by Geolytix and the local authority

Deriving the Spatial Extent of Retail Centres

Michalis Pavlis, Les Dolega and Alex Singleton
University of Liverpool

data. On average (based on the median value) almost 90% of the clustered points were within the Geolytix boundaries.

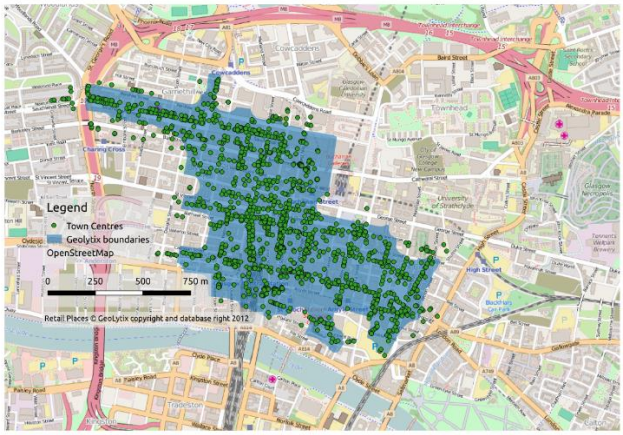

Figure 1. Comparison of results and Geolytix boundaries - Glasgow

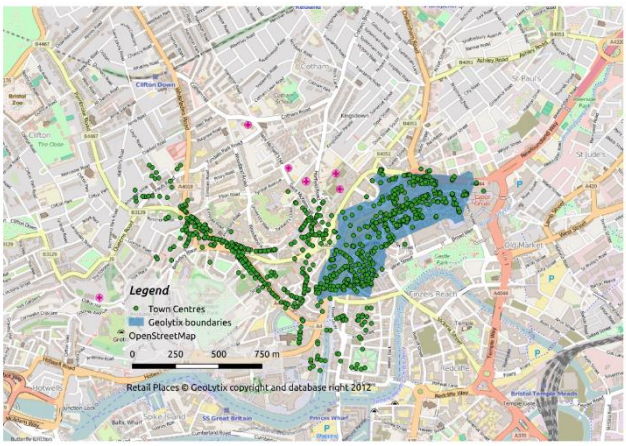

Figure 2. Comparison of results and Geolytix boundaries - Bristol

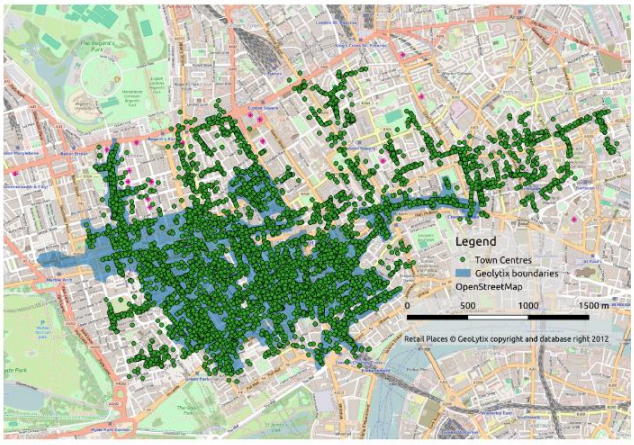

Figure 3. Comparison of results and Geolytix boundaries - Central London

Example clusters and comparisons to the Geolytix data are shown for Glasgow (Figure 1), Bristol (Figure 2) and Central London (Figure 3). Glasgow demonstrates an example where the two datasets mostly overlap, yet the spatial extent of the clusters produced was on average larger. This is likely a result of the pre-processing of Geolytix boundaries to be constrained by road networks.

In contrast, Bristol and London demonstrate examples where the two datasets have significant differences. Concerning Bristol, Geolytix currently split the city centre into several smaller clusters of which only Broadmead was available as open data. However, the novel cluster was inclusive of the surrounding high streets such as Stokes Croft, south of Ashley Road, Christmas Steps and Old Market. It could be argued, based on local knowledge and acknowledgement by Bristol City Council in 2008, that this expanded definition may now be more appropriate.

Geolytix also split Central London into smaller clusters, 7 of which were available for the area that was identified by the modified DBSCAN method as a single cluster. Differences could be due to Geolytix's adoption of additional variables into their clustering method, which, particularly for London, would result in identifying clusters based on different retail activities rather than just retail density.

Given that existing boundary data are widely used by academics, local authorities and private organizations across the country, it is hoped that these results will provide valuable insights and tools for future analysis. In addition, the full paper demonstrates a thorough comparison of clustering techniques, showing that a novel modified version of DBSCAN may be the most appropriate for this context.

Future Directions

The developed methodology aims to be open source (via github) and data links will be added post review. This hopes to provide a straightforward means of updating the retail boundaries on a regular basis and therefore provide more accurate representations of the dynamic structure of centres. In addition, the methodology could be applicable to alternative datasets and across different international locations.