

# アーケード商店街のデータサイエンス

## アーケード商店街の地理情報分析とクチコミ件数予測モデル構築

榎原千爽子 高橋信雄 神沼英里

アーケード商店街は、屋根付きの歩道空間を提供するため持続可能な都市構築目標に沿い、存続が望まれる。しかし通常の商店街よりも、アーケードの維持管理コストがかかり負担が大きい。維持管理費用の捻出には、アーケード商店街の活性化が課題となる。本研究では、日本全国のアーケード商店街の地理情報分析と、機械学習によるクチコミ件数予測モデルの構築を通して、データサイエンスの観点からアーケード商店街の活性化のヒントを探究した。Moran 空間自己相関検定でアーケード商店街数の都道府県行政区別の地理的集積状況を調査した所、大阪府・兵庫県・奈良県がホットスポットで、中部地方から東北地方にかけての地域がコールドスポットであった。また、じゃらん net 観光情報サイトからクチコミ件数と投稿者の属性情報を収集して、投稿者属性を入力とするクチコミ件数の機械学習予測モデルを構築した。Random Forest 回帰アルゴリズムを基に、クチコミ件数予測モデルの属性重要度を調査した。結果として、「60代」と「カップル・夫婦」の属性を持つ訪問者数の増加がアーケード商店街のクチコミ件数増加に寄与する可能性が示唆された。

キーワード：アーケード商店街・Moran 空間自己相関検定・クチコミ件数予測・機械学習

### 1. はじめに

“アーケード商店街”は、地域の商店が密集する商店街の道路上にアーケードが架けられているものを指す。アーケードは屋根付きの歩道空間を提供するため、雨避けの機能を持ち、天候に左右されずに訪問者は商店街を散策する事ができる。商店街の道路を覆うアーケードの維持・管理には、費用が嵩む<sup>1)</sup>が、持続可能な開発目標(SDGs)<sup>2)</sup>の達成が求められている今日、アーケード商店街の存続は、住み続けられるまちづくりの開発目標(SDGs11番)に貢献すると考えられる。アーケードの維持管理コストの捻出には、アーケード商店街の活性化が重要になる。

本研究では、日本全国のアーケード商店街のデータを収集して、データサイエンス的なアプローチにより、アーケード商店街の活性化のヒントを探る。具体的には、まずアーケード商店街の地理位置情報<sup>3)</sup> 4)を収集して、アーケ

ード商店街の日本における地理的分布を調査する。次に、観光のクチコミサイトからアーケード商店街毎のクチコミ件数を収集して、機械学習によるクチコミ件数予測モデルを構築する。特に、予測モデルの重要属性を分析することで、アーケード商店街のクチコミ件数増加に寄与するデータ属性について検討する。

### 2. 研究の方法

#### 2.1 アーケード商店街の地理位置分析

##### (1) 商店街の緯度経度データの収集

本研究ではアーケード商店街名が記載された2つウェブサイト「アーケード商店街写真館」<sup>5)</sup>とWikipediaの「日本のアーケード商店街」<sup>6)</sup>に2021年10月7日にアクセスして、アーケード商店街の住所情報を収集した。次に、Yahoo!ジオコード API<sup>7)</sup>を用いて、調査した住所情報か

ら、アーケード商店街毎の緯度・経度の情報を取得した。この緯度・経度の情報を入力して、Google マイマップ 8) を用いてアーケード商店街の地点の可視化を行なった。Google マイマップは、オンライン地図作成アプリで、ユーザーが緯度・経度の地理位置情報を指定すると地図上にマーカを作成する機能を持つ。

### (2) コロプレス図の作成

コロプレス図(階級区分図、Choropleth Map)は、地理的な区切りで統計情報等に合わせて色調を変化させて色塗りをした地図を指す。本研究では、地理的な区切りは47都道府県とした。各県に存在するアーケード商店街の数を、商店街の住所を基に集計を行い、コロプレス図を作成した。コロプレス図の作成は、Python プログラミングにより GeoPandas 9) ライブラリを用いて行った。日本の都道府県の行政区域データは、GADM データベース 10) から ESRI Shapefile 形式 11) ファイルをダウンロードした。

### (3) Moran 空間自己相関検定

地理的分布を可視化するコロプレス図だけでは分からない、地域集積性の地理位置を検出する指標として LISA(Local Indicator of Spatial Association) 12) がある。LISA として Local Moran 統計量が提案され、GIS 分野や空間疫学分野で使われている。本研究では、Python ライブラリ PyGeoDa 13) を用いて、Local Moran 統計量に基づいた空間自己相関検定を実施した。

## 2.2 観光クチコミ件数予測モデルの構築

### (1) 観光クチコミデータの収集

観光地クチコミ情報サイトの1つである“じゃらん net 観光ガイド” 14) を用いて、アーケード商店街のクチコミ投稿数と投稿者の属性情報を収集した。“じゃらん観光ガイド”のキーワード検索機能を利用して、“観光スポット情報”サイトにアーケード商店街名をキーワードとして入力し検索を実施した。検索にヒットしたアーケード商店街からクチコミ件数、投稿者の女性割合、投稿者の年代(10

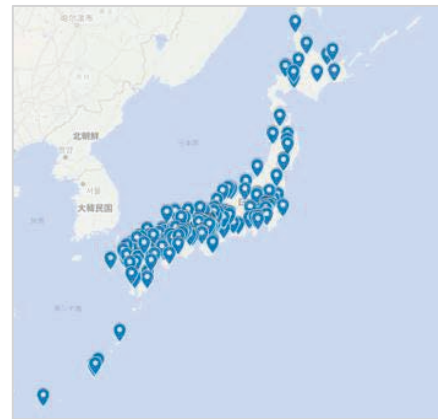


図 1. 日本全国 603 件のアーケード商店街の地理位置情報

代～80 代)、訪問時の投稿者のグループ属性(友達同士、家族、カップル・夫婦、一人、その他)の14の属性情報を収集した。10件よりクチコミ件数が少ないアーケード商店街はデータを採用しなかった。また属性情報の収集数は、アーケード商店街毎に、最新のクチコミから50件までとした。10～50件より少ないクチコミ件数の場合も含めて、全ての入力変数を0～1に正規化した。

### (2) 機械学習を用いたクチコミ件数予測モデルの構築

クチコミ件数の予測モデル構築は、Python の自動機械学習 (AutoML) ライブラリの“PyCaret ver. 2.3.5” 15) を用いて回帰木のアルゴリズムを比較した。データは訓練とテストの割合を8:2に分割した。プログラミング実行環境は、Web ブラウザ経由でクラウド上の計算機を利用できる Google Colab 16) を用いた。

### (3) クチコミ件数予測への各属性の重要度評価

属性情報毎のクチコミ件数予測への重要度を、PyCaret の Feature Importance 機能を用いて計算した。属性間の相関係数は、GeoPandas ライブラリ 9) で Pearson の積率相関係数 17) を計算した。

## 3. 研究の結果

### 3.1 日本のアーケード商店街の地理情報分析

アーケード商店街の住所情報は、ウェブサイト2件に掲載されているリストを2021年10月7日にアクセスして、

表 1. 都道府県別のアーケード商店街数

行政区	商店街数	行政区	商店街数	行政区	商店街数
大阪府	173	大分県	9	佐賀県	3
兵庫県	71	愛知県	8	富山県	3
福岡県	51	滋賀県	8	秋田県	2
東京都	37	静岡県	7	千葉県	2
京都府	18	和歌山県	7	福井県	2
三重県	16	高知県	6	茨城県	2
神奈川県	15	熊本県	6	埼玉県	2
愛媛県	14	徳島県	6	山梨県	2
北海道	14	群馬県	5	青森県	1
香川県	14	宮城県	5	福島県	1
奈良県	13	鹿児島県	4	石川県	1
岡山県	13	宮崎県	4	長野県	1
長崎県	12	鳥取県	4	島根県	1
山口県	11	岐阜県	3	栃木県	1
沖縄県	10	岩手県	3	山形県	0
広島県	9	新潟県	3		

手作業で収集した。1 件目は個人が運営しているウェブサイト「アーケード商店街写真館」4) で、2 件目は Wikipedia 日本語版の「日本のアーケード商店街」5)である。両サイトの重複を除外した所、日本全国のアーケード商店街の総数は 603 件だった。住所情報からジオコーディングで緯度経度の地理情報に変換し、Google マイマップで地図上のマーカーを作成した(図 1)。図 1 より、マーカーで重なっている箇所は視認が難しいが、アーケード商店街の地理的分布は、北海道から沖縄まで幅広い事が判る。

次に、都道府県別の偏りを見る為に、アーケード商店街数の行政区別データセットを作成した(表 1)。大阪府が最も多く 173 個のアーケード商店街が存在しており、多い順では兵庫県、福岡県と続く。一方少ない場合には、山形県はアーケード商店街を持たなかった。青森県、福島県、石川県、長野県、島根県、栃木県のアーケード商店街数は 1 個だった。図 2 に、アーケード商店街数のコロプレス図を示す(沖縄県を除く)。右のカラーバーは濃淡色による商店街数を表し、最も濃い黒色で 173 件の商店街数となる。

更に、アーケード商店街の地理的集積性を分析する為に、Moran 空間自己相関検定を実施した(図 3)。図 3 で大阪府・兵庫県・京都府・奈良県が High-High 領域(赤色)で有意になった(P=0.004-0.024)。High-High 領域は、一般的にホットスポットと呼ばれており、アーケード商店街数が



図 2. 都道府県別アーケード商店街数のコロプレス図 (商店街数が濃淡色情報で色分けされている)

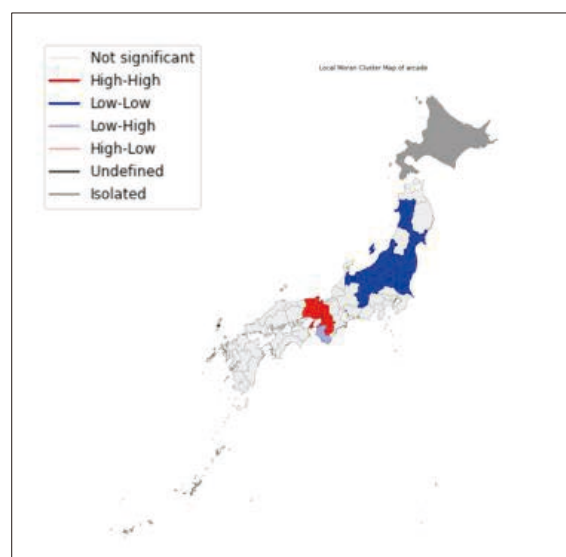


図 3. Moran 空間自己相関検定のクラスターマップ(High-High =ホットスポット、Low-Low=コールドスポット)

多く、近隣県も数が多い地域を表す。一方、Low-High 領域(グレー色)はアーケード商店街数が少ないが近隣県は数が多い地域で、和歌山県のみが有意だった(P=0.01)。Low-Low 領域(青色)はコールドスポットと呼ばれ、自身も近隣県もアーケード商店街数が少ない地域を表す。有意だった地域は、中部地方から東北地方に亘る 9 行政区域 (秋田県・宮城県・福島県・新潟県・富山県・長野県・群馬県・栃木県・茨城県) だった(P=0.002-0.039)。北海道と沖縄県は隣接県が無いクラスタ(Isolated)に分類された。

### 3.2 アーケード商店街クチコミ件数の機械学習予測

#### (1) クチコミ投稿件数と投稿者属性情報の抽出

本研究では、アーケード商店街について、観光地のクチコミサイトである“じゃらん net 観光ガイド” 14)の投稿情報をもとに分析を行った。“じゃらん net 観光ガイド”の検索を利用し、“観光スポット情報”に、603箇所 of アーケード商店街名をキーワードで入力した。検索にヒットして、投稿者のクチコミ投稿件数が 10 以上だったアーケード商店街に限定してデータを収集したところ、27 箇所(約 4%)しかなかった。27 箇所のアーケード商店街のクチコミ投稿件数(以後、クチコミ件数)は、11 から 2,834 と幅広かった。投稿者によるアーケード商店街毎の 5 段階評価値もサイトに掲載されていたが、多くが 4 前後で値に差がない。またクチコミ件数が多いほど賑わいのあるアーケード商店街と考えられる。この為、本研究では、アーケード商店街の活性度を示唆する定量指標として、訪問者の 5 段階評価値ではなくクチコミ件数を採用した。

(2) クチコミ件数の機械学習予測モデルの構築

じゃらん net 観光ガイドのクチコミ欄には、投稿者の属性情報が付与されている。アーケード商店街のクチコミ件数を投稿者属性情報から予測する機械学習モデルを構築して、クチコミ件数に寄与する属性情報を調べる事とした。投稿者属性情報として、①性別(男性、女性) ②年代(10代~80代) ③訪問時のグループ属性(友達同士、家族、カップル・夫婦、一人、その他)をじゃらん net サイトから収集した。属性情報データから、女性割合・8種類の年代・5種類のグループ属性の 14 種類の正規化特徴量を作成した。

クチコミ件数予測モデルの構築には、Python のローコード AutoML ライブラリである PyCaret 15)を用いて、3 件の回帰木機械学習アルゴリズム「Decision Tree Regressor 18)」「Random Forest Regressor 19)」「Extremely Randomized Trees Regressor 20)」を比較した。5 分割交差検証を行い、平均 2 乗誤差 (Mean Square Error, MSE) の評価尺度を用いてアルゴリズムを比較した所、Random Forest 回帰アルゴリズム 19)が最も MSE

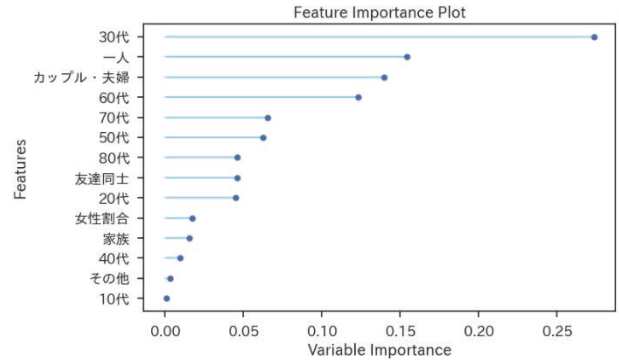


図 4. 投稿者属性のクチコミ件数予測への重要度ランキング

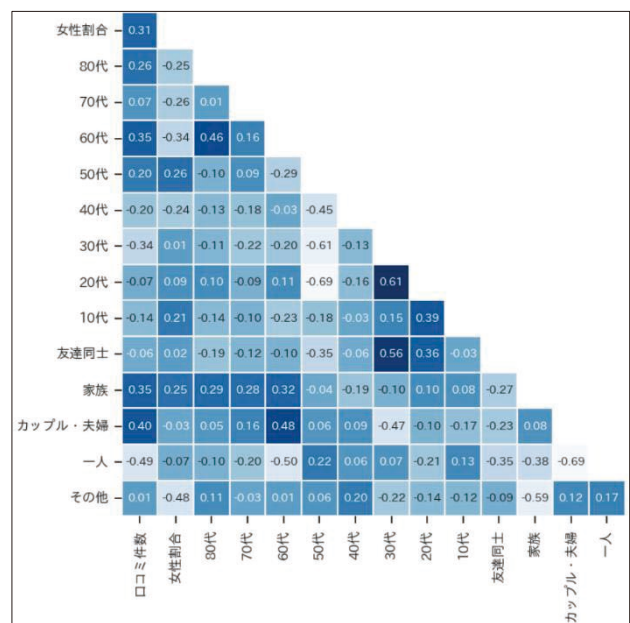


図 5. アーケード商店街のクチコミ投稿者の属性間相関係数

の値が小さかった。

(3) クチコミ件数予測に寄与する投稿者属性分析

Random Forest 回帰アルゴリズムでクチコミ件数予測モデルを構築した後で、PyCaret の機能を用いて、属性重要度を調査した。図 4 はアーケード商店街のクチコミ件数に各投稿者属性情報 (Feature) がどのくらい影響を与えているか重要度 (Feature Importance) をランキングしたものである。横軸の値が大きい程、入力変数としてモデル出力であるクチコミ変数への寄与度が高い。一方、図 5 に入力属性・出力属性を合わせた投稿者属性間のピアソン積率相関係数 17)を示す。図 4 の属性重要度が 1 位・2 位だっ



た「30代」と「一人」の属性は、「クチコミ件数」とは負の相関関係にある事が判る。相関係数の値は、それぞれ「30代」属性が $-0.34$  ( $P<0.10$ ) , 「一人」属性が $-0.49$  ( $P<0.01$ )だった。重要度が3位・4位の「カップル・夫婦」と「60代」の属性は、「クチコミ件数」と正の相関関係にあり、相関係数は各々、 $0.40$  ( $P<0.05$ )と $0.35$  ( $P<0.10$ )だった。この結果から、アーケード商店街への訪問者が「60代」や「カップル・夫婦」で占められている場合はクチコミ件数が多く、逆に「30代」や「一人」の訪問者で占められる場合は、クチコミ件数が少ない可能性が示唆された。

#### 4. まとめ

本研究では、データサイエンスの観点でアーケード商店街の活性化のヒントを探究する為に、日本全国のアーケード商店街の地理情報分析と、機械学習によるクチコミ件数予測モデルの構築を行った。またクチコミ件数予測モデルの属性情報分析から、アーケード商店街への訪問者が「60代」や「カップル・夫婦」が多い場合は、クチコミ件数も増える可能性が示唆された。クチコミ投稿者の属性間相関係数を見れば、他にも特徴を確認する事ができる。例えば、図5からは「友達同士」と「30代」・「20代」には関連性が伺える。アーケード商店街ではなく、子育て世代が通うショッピングモールで本研究と同様の属性情報分析をすれば、アーケード商店街とは異なる訪問者属性でクチコミ件数との関連がみられるかもしれない。また地理情報分析で603箇所のアーケード商店街を分析したが、クチコミ属性分析では27箇所のデータしか得られなかった。今後、アーケード商店街のクチコミのデータ数が増えれば、より精度が高いクチコミ件数予測モデルが構築できるだろう。

#### 謝辞

本稿をまとめるにあたり名古屋市立大学の栗原康行教授に、厚く御礼申し上げます。

#### 参考文献

- 1) 築根広明、北原理雄、商店街における全蓋式アーケードのコストと利点、(2010)、日本建築学会技術報告集、16巻32号、pp. 335-338.
- 2) Cf, O. : Transforming our world: the 2030 Agenda for Sustainable Development. (2015), United Nations: New York, NY, USA.
- 3) 国土交通省 : GIS ホームページ、  
<https://nlftp.mlit.go.jp/index.html>
- 4) 国土交通省国土地理院 : GIS・国土の情報、  
<https://www.gsi.go.jp/gis.html>
- 5) アーケード商店街写真館: 全国のアーケード商店街一覧、  
<https://syoutengai-net.jp/list.html>
- 6) Wikipedia : 日本のアーケード商店街、  
<https://ja.wikipedia.org/wiki/日本のアーケード商店街>
- 7) Yahoo!ジオコード API、  
<https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/geocoder.html> .
- 8) Google MyMap :  
<https://www.google.co.jp/intl/ja/maps/about/mymaps/>
- 9) Jordahl, K. : GeoPandas: Python tools for geographic data. (2014) <https://github.com/geopandas/geopandas>.
- 10) Global Administrative Areas : GADM database of Global Administrative Areas, version 2.0. (2012), <https://www.gadm.org>.
- 11) ESRI : Esri Shapefile Technical Description, (1998), An ESRI White Paper.
- 12) Anselin, L.: Local Indicators of Spatial Association –LISA, (1995) *Geographical Analysis*, 27:93-115.
- 13) Anselin, L. et al. : GeoDa, from the Desktop to Ecosystem for Exploring Spatial Data, (2021), *Geographical Analysis (Early Release)*.

- 14) Recruit Co., Ltd.: ジャらん net 観光ガイド,  
<https://www.jalan.net/kankou/>.
- 15) Moez, A. : PyCaret : An open source, low-code  
machine learning library in Python, (2020)  
<https://www.pycaret.org>.
- 16) Bisong E. : Google Colaboratory, In:Building  
Machine Learning and Deep Learning Models on  
Google Cloud Platform, (2019), Springer, pp.59-64.
- 17) Pearson, K. : Notes on regression and inheritance in  
the case of two parents, (1895), Proc. of the Royal  
Society of London., 58, pp.240-242.
- 18) Breiman, L. et al: Classification and Regression  
Trees, (1984), Wadsworth, Belmont.
- 19) Liaw, A. and Wiener, M.: Classification and  
Regression by randomForest, (2002), R News 2(3),  
pp.18-22.
- 20) Geurts, P., Ernst D., and Wehenkel, L. Extremely  
Randomized Trees, (2006), Machine Learning, 63,  
pp.3-42.