

経済行動に関する都市データを用いた統合分析システムの開発

株式会社竹中工務店 技術研究所 国本 陸斗、安藤 邦明、中村 壮志、畔上 泰彦、井上 僚
株式会社竹中工務店 情報エンジニアリング本部 高根 一晃、伊藤 翔
大阪公立大学大学院 情報学研究科 准教授 佐賀 亮介

1. はじめに

昨今の情報化社会における計算機能力の向上、クラウドプラットフォームの拡充及びスマートフォン等の普及¹⁾に伴い、都市における人間の移動・購買行動及びその動機発見のための場としての SNS 上における活動等、都市経済に関わる諸行動を記録した情報ソースが豊富に蓄積されるようになった。

このような状況の下、関連企業においては都市データの分析により意思決定を支援するための技術に関する研究開発が盛んに行われている²⁻⁴⁾。特に都市開発分野においては、計画策定のための意思決定が経済に重大な影響を及ぼすことから多角的な検討が求められる一方、現状普及している手法は個々のソースに閉じた分析に留まっており、異なるソースに跨る分析を効果的に行うための有効な手法は未だ確立されていない。

そこで我々は、異なるソースにおいて収集された人間の経済活動に関する各都市データを結合し、総括的に扱うことで得られるシナジーを生むための統合分析システム(以降、当システムと呼称)を新たに開発した。このシステムにおいては、後述する仮想人格(デジタルペルソナ、図 1)の考え方を新たに導入することでそれを実現している。

本論文では、当システムが対象とする課題の抽出・定義と、それを解決するための具体的なシステム構成および分析手法について提案する。また、当システムの有用性を検証するために行った大丸有エリアにおける実データを通じた分析実験について、その方法および結果を報告する。

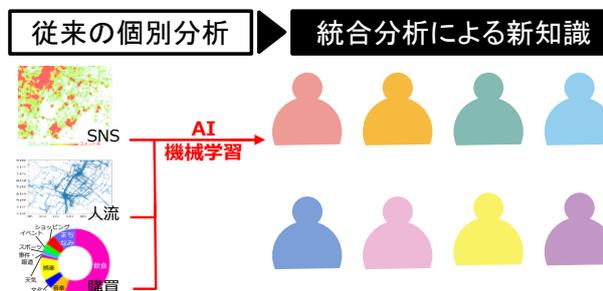


図1 デジタルペルソナの概要

2. システムアーキテクチャ

(ア) システム開発における課題及び機能目標

まず、当システムが持つべき機能目標を定義する。当システムを情報プラットフォームとして捉えたとき、小野ら⁵⁾がその機能要求条件として提唱する「収集・構築」「蓄積」「流通」「利用」の内、特に汎化が難しく各ドメインで最適設計されるべきものが「利用」機能であり、本システムはこの「利用」機能の提供において都市データの統合分析を可能とする必要がある。この達成にあたり、我々は下記の通り課題を見出し、その解決策として本システムが担うべき機能事項を決定した。

① 経済活動の主体としての仮想人格の生成

データを統合的に扱うことによるメリットを得るためには、複数ソースの情報を同時に扱うことが必要である。そこで我々は、経済活動の主体情報を仮想人格(デジタルペルソナ)として扱う手法を考案した(図 2)。

デジタルペルソナは、都市データがそもそも人間の経済活動が部分的に記録された表層情報であることに注目し、これらを後述の手法に従って結合することで、経済活動を行う主体(個々の人間)の特徴を補完的に求めることで生成される。これに基づき確率的に生成されるデジタルペルソナ集合に対し、適切なデータマイニング手法(アソシエーションルール、ベイジアンネットワーク等)を適用することにより効率的な知識発見が可能となる。例えば、あるエリア

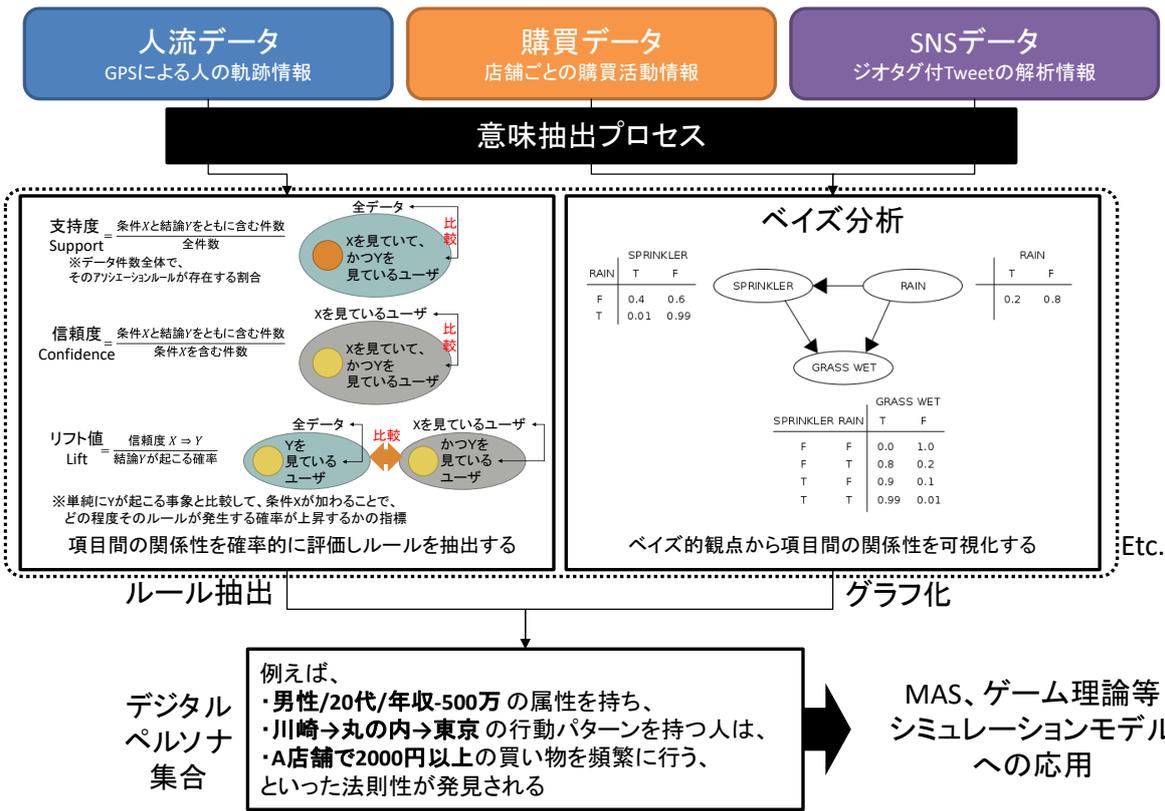


図2 デジタルペルソナと当システムを取り巻く外部要素との関係性

についてこれらを適用した時、そのエリアにおけるメジャーペルソナや、他のエリアに対し相対的に特色のあるユニークペルソナなど、エリアにおける経済活動主体の特徴を統計的に妥当な形で抽出することが可能である。

② データテーブルの結合手法

データ結合方法の概要を図3に示す。当システムが提案する手法はデータ融合のフレームワークに則る。抽象化されたデータ間における共通の属性項目を基に異種データを関連付けることで、疑似的に表層データを組合せることでレコードの属性項目を拡張することに成功した。

なお、結合ステップでは共通項目の値をある組合せで固定し、残りの属性の値の分布が、他の値の組合せで固定した場合の残りの属性の値の分布と比較して有意な差があるかを判断(多重比較検定等を用いた多項カテゴリーカルデータ群間の有意差検定)する必要がある。

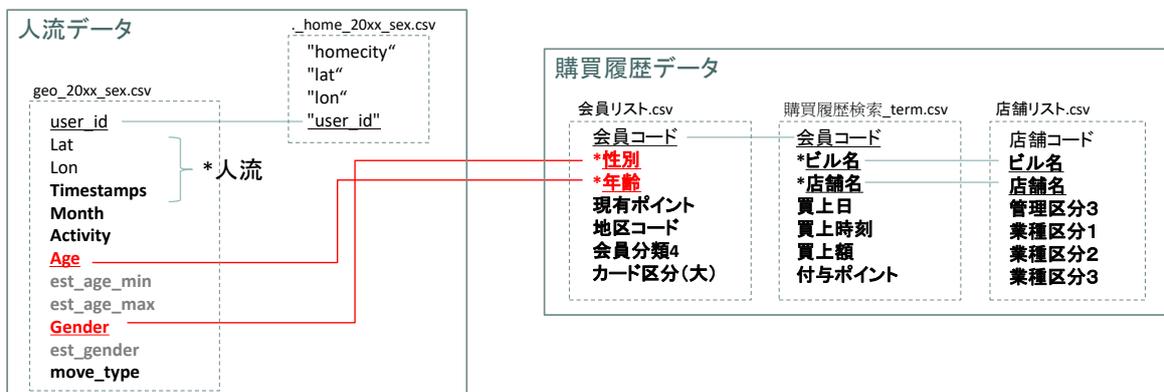


図3 データ間における共通属性の抽出とデータテーブル結合のイメージ

3. 大丸有エリアに関する実データを用いたシステム実験計画

(ア) 実験概要と目的

我々は、当システムの有用性を検証するためのフィールドとして、その経済活動の活発度を理由に大丸有エリアを選定した。当該地区の概要を図4に示す。

本実験では、大丸有エリアにおける実際の都市データに対し、2章で提案した分析フレームワークを適用することにより複数ソースを用いた統合的分析を試行した。また、その分析結果について考察することにより、当システムの有用性を検証することを目的とする。

区域面積	約120[ha]
就業人口	約280,000[人]
事業所数	約4,300 内、上場企業数:2,665 上記企業群による売上高の 全企業売上高に対する割合:8.85%
交通	13駅28路線 乗車人数:約117万[人/日]



図4 大丸有エリアにおける経済都市経済に関する主要情報とその概要⁶⁾

(イ) 実験条件

前述の目的を達成するため、本実験においては都市データとしての活用が盛んな「GPS(人流)」「ポイントカード使用履歴(購買)」「SNS(趣味・嗜好)」といった3つのソース^(*)から実際に得られたデータを用いた。4章で各データに対して意味抽出手法を設計・適用した結果を説明したのち、5章でそれらの結合処理を行い、最終的に得られたデジタルペルソナについて分析・考察を行った。

4. 実験に用いた各データの概要及び意味抽出手法とその適用結果

ここでは実際に使用した各データの構造・性質と、それらに対して本実験にて設計した各データ単体への意味抽出手法とその適用結果について説明する。なお、統合分析のサマリは5章にて説明する。

(ア) 人流データ(マルコフモデルとベイズ確率による移動パターン推定)

人流データとは、GPSで収集されたデバイス所持者による移動の時系列移動軌跡データである。当システムでは、軌跡データと併せて駅データベースの座標情報を用いることで、GPSの軌跡データを、駅をノードとする遷移ネットワークデータとして抽出した(図5)。このネットワークデータに対し、駅ごとに次点の移動対象となる駅への移動頻度を遷移確率としてマルコフ過程を構成し、そこにベイズ統計の手法に則った事前分布モデルを導入する。

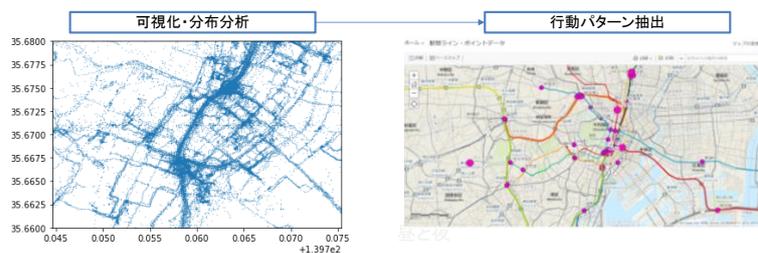


図5 駅ネットワークモデルによる人流データからの行動パターン抽出

上記手法の概要とその適用結果の例を図6に示す。この試行では「九段下駅」という入力に対し、その駅を始点あるいは通過点として

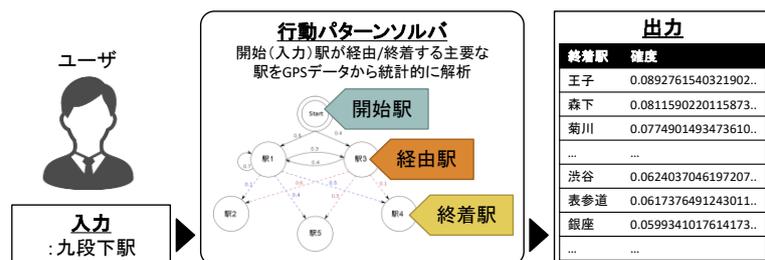


図6 行動パターンソルバによる解析のフロー概要と結果の例

持つ行動パターンペアとして確率が高い駅をその確度と併せ抽出した。出力結果を見ると、王子、森下、菊川等の駅が、渋谷・表参道・銀座等の、相対的に乗降数の多い駅群の上位にあること等から、この手法が統計的頻度によるバイアスをうまく除去できていることがわかる。

(イ) 購買データ(パターンマイニングによる購買パターン抽出)

購買データとは、取引物品の対象カテゴリ、日時、取引額、店舗等の購買行動に関するログである。なお、本論文では匿名加工されたポイントカード会員による購買履歴データを用いた。このような取引データは、頻出パターンマイニングの技術(図7)を用いてそのパターンを分類することができる。

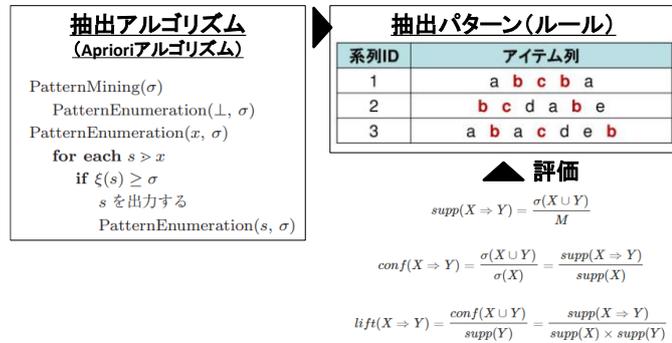


図7 パターンマイニングの概要

当システムによる分類結果を図8に示す。例えばID:1のパターンは美容サービス店の利用であり、大丸有エリアにおける一定数の従業員が終業後に利用する高額サービスの消費活動パターンとして存在することを示唆している。同様にID:2は昼休憩時のコンビニでの昼食購入、ID:3は終業後の夕食・会食、ID:4は終業後、或いは主(婦/夫)層による夕食の準備のための購買行為であるといったように、抽出された購買行動に関する属性値の組合せから、その活動内容を推定することが可能である。

ID	店舗	購入カテゴリ	取引価格	時間帯
1	※※※※※※※※※※	サービス	高(>5000)	夜(18:00-22:00)
2	※※※※※※※※※※	食品、飲料	低(<1000)	正午(11:00-13:00)
3	※※※※※※※※※※	外食	中(1000<Price<5000)	夜(18:00-22:00)
4	※※※※※※※※※※	食品、日用品	中(1000<Price<5000)	夕方(16:00-18:00)

図8 購買パターン抽出の例

(ウ) SNSデータ(人工知能(機械学習)によるSNSユーザ属性分類器の作成)

SNSデータとは、ユーザにより投稿される文章及びその日時、地理情報、画像・動画等のメディアデータを指す。

SNSデータからの意味抽出については、当社開発のSHM;ソーシャルヒートマップ®⁷⁾の出力を利用した。SHMは図9で示されるようにTwitterの投稿/アカウ

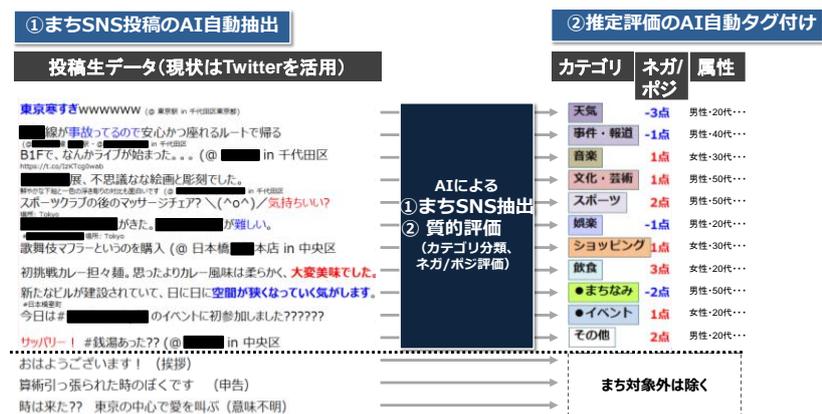


図9 SNS上での投稿データの構造とその例

ントにおけるポジネガ度や興味関心の対象を推定する機能を提供しており、推定された属性情報に共通属性となる項目(性別、年代等)が含まれているため、これを結合することが可能である。

5. システムによるデジタルペルソナ生成の試行と結果

4章では各単体データに対する個別分析を行ったが、5章ではそれらをデジタルペルソナの技術を用いて統合分析した結果について説明する。



図 10 座標情報としての人流データの内容とその可視化の例

当システムによるデジタルペルソナ生成の結果出力イメージを図 10 に示す。地図上に表示されているバブルは駅を基本単位とし、行動パターンの始/終点、あるいはハブとしての中心性に比例した半径サイズにより表現されている。ある単位エリア(本実験においては駅)を指定すると、当該の駅と、当該の駅と関連性の高い駅とが結ばれたエッジ(経路)が表示される。このエッジ、或いは単位エリアを選択することにより、その要素を特徴として持つ尤もらしいデジタルペルソナが提示される。

まず、単独エリアについてのデジタルペルソナ生成結果の一部(メジャー/ユニーク度合における上位

メジャーペルソナ							ユニークペルソナ						
性別	年代	配偶者	収入	購買カテゴリ	取引額	趣味嗜好	性別	年代	配偶者	収入	購買カテゴリ	取引額	趣味嗜好
女	40代	あり	>1000万	食物販	中(1000-3000)	事件・報道	男	30代	あり	>1000万	レストラン・カフェ	高(>5000)	事件・報道
男	40代	あり	>1000万	食物販	中(1000-3000)	事件・報道	男	40代	あり	>1000万	レストラン・カフェ	高(>5000)	事件・報道
女	30代	あり	700-1000万	食物販	中(1000-3000)	事件・報道	女	30代	あり	>1000万	宝飾品	高(>5000)	事件・報道
女	50代	あり	>1000万	食物販	中(1000-3000)	事件・報道	男	20代	あり	>1000万	宝飾品	高(>5000)	スポーツ
男	50代	あり	>1000万	食物販	中(1000-3000)	事件・報道	女	20代	なし	>1000万	宝飾品	高(>5000)	飲食
男	30代	あり	700-1000万	食物販	中(1000-3000)	事件・報道	女	50代	あり	>1000万	食物販	高(>5000)	事件・報道
女	20代	なし	700-1000万	食物販	中(1000-3000)	事件・報道	男	40代	あり	>1000万	衣料品	高(>5000)	事件・報道

図 11 品川駅エリアにおけるデジタルペルソナ分析

7つ)を図 11 に示す。なお、分析対象は「品川駅(港区在住)利用者で、大丸有エリアで勤務・買い物を行う人々」とし、対象となる属性項目については簡単のため一部を取捨選択した。メジャーペルソナとして抽出されたペルソナの属性値から、港区エリアの主体が、年代に関わらず有配偶率が高く(0.719)、かつ高所得者(年収において>700 万円レイヤが 0.885)であることがわかる。購買活動に関しては、購買データ全体の傾向と同様に食物販に関するものが上位を占めた。趣味嗜好に関しては、事件・報道である比率が他カテゴリに比べ有意に高く(11 カテゴリで 0.229)、上位ペルソナは当該の値をもつもので占められた。また、ユニークペルソナとして抽出されたペルソナの属性値から、港区エリア独特の人物層として年収が 1000 万円を超える高所得者層の割合が他エリアに対し有意に高く(他エリア平均 0.090 に対して 0.391)、20-30 代に条件を制約した際の有意性は更に高い(他エリア平均 0.001 に対し 0.095)。また、高額取引の割合に関しても同様に高い(他エリア平均 0.112 に対し 0.247)。

以上の分析結果から、品川駅(港区)利用者エリアにおいては若年層(20-30 代)を含む経済的エリート層による高額な購買活動が盛んである、という推定を行うことができる。

次に、複数エリアを対象に生成したデジタルペルソナの特徴について比較した分析例を図 12 に示す。この分析では、属性ごとの値の分布ではなく、属性値の組合せにより予めクラスタリングされたペルソナのクラスタに対し、各クラスタに属するペルソナの割合を比較したものである。今回は 8 つの購買カテゴリ

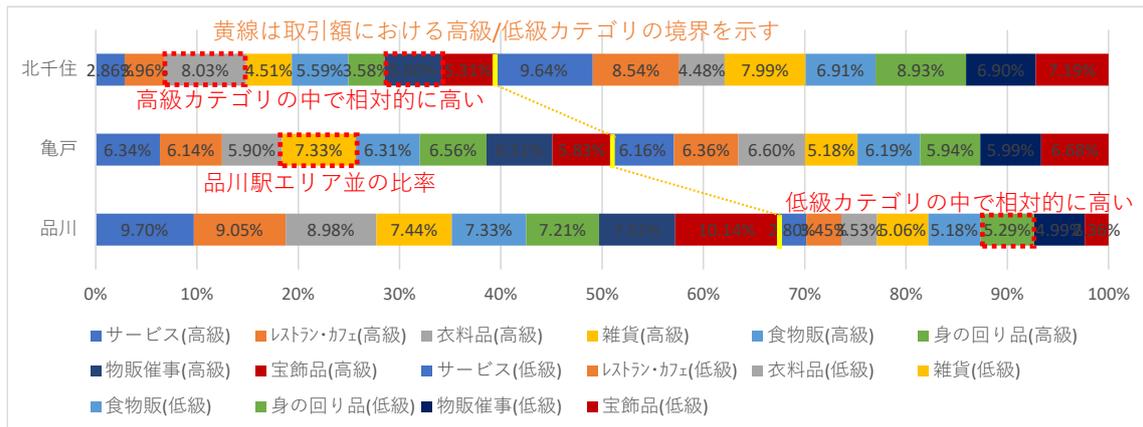


図 12 品川・亀戸・北千住駅間におけるデジタルペルソナ比較分析

を低級/高級の 2 クラスに分類した計 16 種のクラスタを学習した。対象エリアは、東京都内において平均世帯収入に顕著な差異が存在する⁸⁾品川駅(港区)、亀戸駅(江東区)、北千住駅(足立区)の 3 駅を選定した。解析結果を見てみると、実際にその統計上の特色が経済活動の特徴として表れていることが確認できる。ただし、北千住エリアにおいては相対的に高級衣料品及び物販催事クラスタの比率が高く、この地域では統計の事実的傾向に反して高級衣料品の催事が有効である可能性が示唆されている。亀戸エリアにおいては比較的特徴が穏やかであるが、高級雑貨が相対的にヒットする可能性が示唆されている。品川エリアにおいては全体的に高級志向が強いが、低級カテゴリの中でも身の回り品は比較的高いため、当該エリア内における市場開拓の可能性が示唆されている。

6. おわりに

本論文では、複数ソースによる都市データを統合的に分析するためデジタルペルソナのシステムを開発し、そのフレームワークによりエリアごとに経済行動の特徴を抽出することに成功した。今後、これを都市計画事業者へ展開することで、当システムがその策定における意思決定のための重要な支援ツールとして活用されていくことを期待する。また、外部システム連携や、ユーザフィードバックを始点とした改善 PDCA により、実務における有用性を高めつつより多くの課題を解決する機能を検討していき、業界のスタンダードツールとして広く認知・利用されるよう、当システムの研究開発を推進する予定である。

(※)人流データ、SNS データの取得に関しては、レイ・フロンティア(株)、(株)NTT データとの契約による。またポイントカードデータは三菱地所(株)の協力による。ここに謝意を示す。

【参考文献】

- 1) 総務省(2021), "デジタル・トランスフォーメーションによる経済へのインパクトに関する調査研究"
- 2) M.Shimosaka, K.Maeda, T.Tsukiji, K.Tsubouchi, "Forecasting urban dynamics with mobility logs by bilinear Poisson regression", ACM International Joint Conference on Pervasive and Ubiquitous Computing, p.535-546(2015)
- 3) G. Linden, B. Smith, J. York, "Amazon.com recommendations: item-to-item collaborative filtering", IEEE Internet Computing, Vol.7, I.1, p.76-80(2003)
- 4) M.Mathioudakis, N.Koudas, "Twittermonitor: trend detection over the twitter stream", ACM SIGMOD International Conference on Management of data, p.1155-1158(2010)
- 5) 小野欽司, 丸山勝巳, "情報プラットフォームの概念とその技術課題", NII Journal No. 3 (2001)
- 6) 一般社団法人大手町・丸の内・有楽町地区まちづくり協議会 (<https://www.tokyo-omy-council.jp/>)
- 7) 竹中工務店, "ソーシャルヒートマップ"(<https://www.takenaka.co.jp/solution/future/social/>)
- 8) 総務省, "令和 3 年度課税標準額段階別所得割額等に関する調"(2021)