

安全知データの3方向モデル表現

Three-way data modeling and its application for safety problem

○¹辻 光宏, ¹堀之内公亮, ¹泉本優輝,
○¹Mitsuhiro Tsuji, ¹Kosuke Horinouchi, ¹Yuki Izumoto,
 ²下川 敏雄, ¹田中成典
²Toshio Shimokawa, ¹Shigenori Tanaka

¹ 関西大学

¹Kansai University

² 山梨大学

²University of Yamanashi

Abstract: We investigated some applications of three-way data modeling for safety problem. The target safety problem has three spaces as follows; secure space, geographical space and time space.

At first, the spatio-temporal GIS approach was adopted. At the same time, we made development of the dynamic graphics of matrix type presentation.

To the next, the validity of integration of MDS and Cluster Analysis was observed as follows; (i) The result of MDS was verified by Cluster Analysis, (ii) The result of Cluster Analysis was visualized on the map from MDS. We adopted INDSCAL model and INDCLUS model. Then we created some kinds of maps on the same plane which would show the 0latent structure.

We developed the matrix-type presentation which is the effective application to execute the dynamic graphics, where we can make zoom-up, zoom-down and moving by the simple way like the smart phone.

1 はじめに

安全知の問題, 特にここでは飲酒と医療に関連する情報に関して, 3方向から検討した. 3方向とは, 飲酒や医療に関する情報と地域空間と時空間とする. 飲酒や医療の情報 [9] それぞれに注目すると, 酒の種類や医療施設に関するそれぞれの情報を地域空間にマッピングした地図を時空間移動することで, 地域ごとに酒の種類別飲酒量や医療施設情報に関する時間的推移を把握することができる. 時空間上に展開された地図情報をマトリックス状に表現し, ダイナミックに拡大/縮小/移動できる視覚化ツールも開発してきた.

また, 多次元尺度構成法 [2] とクラスター解析 [3][10] を連携させて解析することにより, 多種多様な分析が可能となるので, いくつかの例を紹介する.

例えば, 特定な期間でのいろんな種類の酒の飲酒に関する情報を集約することに注目すると, 多次元尺度法 (INDSCAL モデル [1]) によって縮約し, 共通の2次元または3次元空間に酒の種類に基づく都道府県の認知地図を布置することができる. この布置に関係する酒の種類については, 重み値を布置することで都道府県布置の軸と関連付けることができる.

INDSCAL モデルと同じデータをクラスター解析 (INDCLUS モデル [1]) することによって, 都道府県を分類 (非階層重複クラスタリング) することができる. このクラスター解析結果を上記の都道府県認知地図に追記することによって適切な分類結果を表現でき, 多次元尺度構成法による縮約の歪みを検出することができる.

これらの結果を, 飲酒に関する共通空間として, 医療施設に関するデータを追加してこの空間にマッピングすることによって, 飲酒という概念に基づく医療施設の位置づけを導出することができる. すなわち, 飲酒に関する都道府県の共通空間を初期布置として, 医療施設データに基づく都道府県の INDSCAL モデルの収束計算をすぐに停止することで可能となる [7]. 重みとして導出される医療施設情報を布置することができる [8]. 医療施設データに関するクラスタリング結果も同様に追記することができる.

多次元尺度構成法とクラスター解析を連携させて解析して空間に布置した散布図1つずつをマトリックス状に表現し, ダイナミックに拡大/縮小/移動できる視覚化ツールを用いることで比較検討が容易になる.

2 時空間 GIS データのマトリックス表現

時間の推移とともに変化する GIS データを、動的なマトリックス表現を行うことで、多面的かつ多様な視点で評価できる方法を採用し、そのサービスを開発してきた [5] .

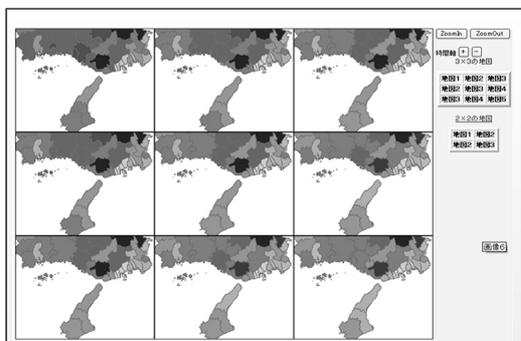


図 1: 時空間 GIS のマトリックス表現

サービスを開始すると、ウィンドウ画面に 3 × 3 の地図のマトリックス提示画面 (図 1) が表示される。ズーム機能による複数の地図の拡大と縮小、ドラッグ機能による複数の時間や地図位置の移動を実行できる。2 × 2 の地図のマトリックス提示画面に切り替えることも可能である。

3 データ解析

3 方向モデルのデータに対する多次元尺度法 (indscal) とクラスター解析 (indclus) について説明する。

3.1 3 方向モデル

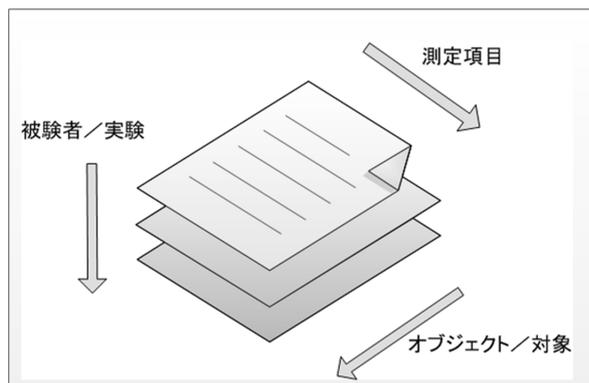


図 2: 3 方向モデルのテンプレート

本研究での解析での 3 方向モデル (図 2) は、個人差 (individual) に配慮した 2 相 3 方向データとして取り扱う。個人差を、飲酒および医療データ空間 (測定項目) とし、時間空間 (被験者 / 実験) で集計した都道府県の地域 (オブジェクト / 対象) の間の距離モデル

を分析する。

3.2 個人差多次元尺度構成法 INDSCAL モデル

INDSCAL モデルは、共通飲酒空間の中での酒類 k に対する都道府県 i と都道府県 j との距離データ行列

$$D \equiv \{d_{ij,k}\} \quad (1)$$

を、入力データ行列 $\Delta \equiv \{\delta_{ij,k}\}$ から推定する。再現するという表現もとられる。

共通の空間上での距離データは、対象の都道府県の布置と次元に対する重みを用いて、次式で表される。

$$d_{ij,k} = \sqrt{\sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2} \quad (2)$$

ここで、 w_{kr} は酒類 k に対する次元 r における判断や認知の重要度を表す重みであり、次元 r に対して伸縮することができるので、多種多様な個人差を表現することができる。また、 x_{ir} と x_{jr} は R 次元の共通飲酒空間での都道府県 i と都道府県 j の布置である。

多次元尺度構成法の計算は一意的なものではなく、変換による歪みをできる限り少なくする非線形最適化問題を解くことになる。最適化計算は、布置の初期値に伴う局所解の存在という問題を伴っている。その際には、適合値を示す VAF 値が妥当であることと、重みが負値でないこと (絶対値が小さな負値は 0.0 とみなす) とが最適値の基準となるように、試行錯誤的に初期値を調整しながら再計算を繰り返す必要がある。

本研究では、Arabie ら [1] のプログラムを用いることにより、初期値を何度も乱数発生させて入力とした最適化計算を何度も繰り返し実行し、有意な適合値 VAF を得ることに努めた。

3.3 個人差クラスタリング INDCLUS モデル

飲酒を個人差とする INDCLUS モデルは、酒類 k に関する都道府県 i と都道府県 j の間の類似度を次式で表現する。

$$s_{ij,k} \cong \sum_{r=1}^R w_{kr} p_{ir} p_{jr} + c_k \quad (3)$$

ここで、 w_{kr} はクラスター r に対する酒類 r の重み、 p_{ir} は都道府県 i がクラスター r に属しているか ($p_{ir} = 1$) 属していないか ($p_{ir} = 0$) を示し、 c_k は酒類 k の追加定数である。INDCLUS は、複雑なデータ構造に対応できるように、対象が複数のクラスターに属することを許す重複クラスタリングであり、クラスター融合に規則性を要求しない非階層の特性をもつ。

最適化計算は、INDSCALと同様に類似度の初期値に伴う局所解の存在という問題を伴っている。その際には、適合値を示す VAF 値が妥当であることと、重みが負値でないことが最適値の基準となる。本研究では、Arabie ら [1] のプログラムを用いて、初期値を乱数発生させて入力とした最適化計算を試行錯誤的に何度も繰り返し実行し、有意な適合値 VAF を得ることに努めた。

非階層の特性のため、クラスター個数それぞれに別の結果を得ることになる。また、重複クラスタリングのため、ある都道府県は複数のクラスターに属する。

3.4 INDCLUS 結果の図示方法の選択

クラスタリング結果を図示する代表的な方法として有名なものは、階層型クラスタリングの場合のデンドログラム（樹状図）である。INDCLUS は非階層型クラスタリングであるので、低次元データ表現の方法として個人差多次元尺度構成法 INDSCAL を採用する。

都道府県に共通の場としての低次元データ表現を目指している。そのため、飲酒を共通の場に布置し、さまざまな都道府県を対象としても共通の場はあまり変えないように個人差として取り扱うように工夫した。

本研究では、まず最初に酒類それぞれを個人差として都道府県の共通布置を求め、その共通布置を踏襲して医療施設に関しても同じ都道府県の共通布置を導出するように工夫した。

4 安全知データへの応用

結果の一部を紹介する。

4.1 時空間 GIS

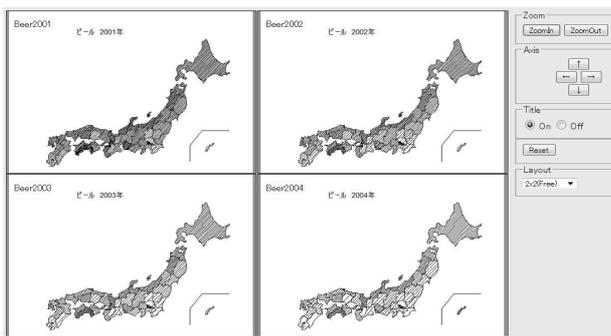


図 3: 1万人あたりのビール消費本数の時空間 GIS のマトリックス表現

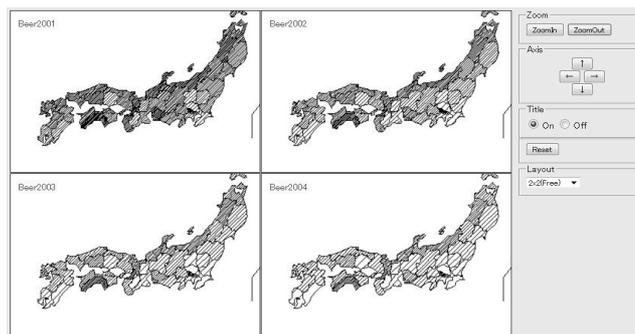


図 4: 時空間 GIS のマトリックス表現 (ズームアップ)

朝日新聞社「民力」から、1万人当たりのさまざまな酒類（ビール、ウィスキー、果実酒、清酒、しょうちゅうなど）の販売量を対象にして、時空間 GIS データのマトリックス表現により視覚的に表現し（図 3、図 4）、地域特性を検証した。

4.2 個人差多次元尺度構成法 INDSCAL モデル

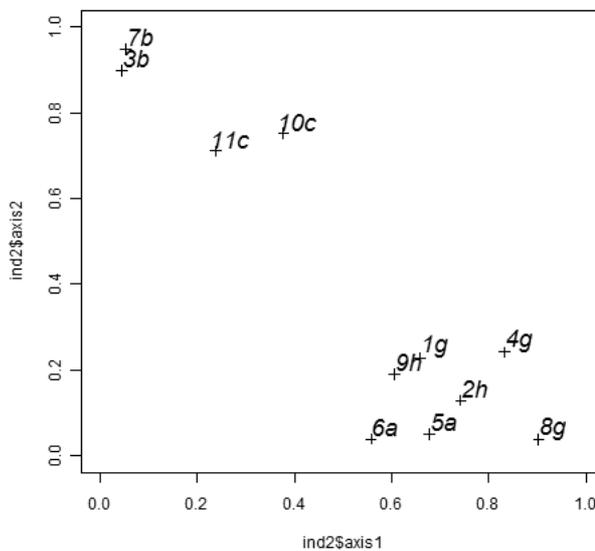


図 5: INDSCAL 2次元布置の例

4.3 個人差クラスタリング INDCLUS モデル

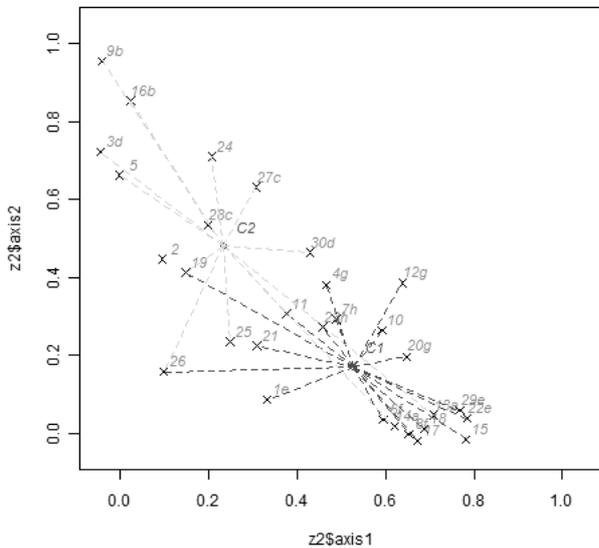


図 6: INDCLUS 結果の重み空間布置の例

5 謝辞

本研究は、文部科学省私立大学戦略的研究基盤形成支援事業「セキュアライフ創出のための安全知循環ネットワークに関する研究(研究代表者:堀雅洋(関西大学))」によって行われた。ここに御礼申し上げます。

参考文献

- [1] P. Arabie, J. D. Carroll, and W. S. DeSarbo : Three-way Scaling and Clustering, Sage, Newbury Park, 1987. 岡太彬訓, 今泉忠 共訳: 3元データの分析, 共立出版, 1990.
- [2] I. Borg, and P. Groenen: Modern Multidimensional Scaling, Springer, 2005
- [3] B. Everitt, et al: Cluster Analysis (5th ed.). Wiley, 2011
- [4] 岸原万梨子, 辻光宏: 感情価を用いた音楽作品分類への試み. 情報処理学会 MUS-67,33-36, 2006.
- [5] 辻光宏, 名取良太, 紺社聖司, 木村真理子, 田中成典: 地域活動に関する時空間 GIS 情報サービスの開発 - 投票行動と道路計画と経済状況の PHE 3次元モデルの試行 - . 情報処理学会 第83回人文科学とコンピュータ研究会発表会, Vol.2009-CH-83, pp. 205-212, 2009.

- [6] M. Tsuji, T. Shimokawa, and A. Okada: Three-way Scaling and Clustering Approach to Musical Structural Analysis. In Locarek - Junge, H. and Weihs, C. (eds.), Classification as a Tool for Research, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag Berlin Heidelberg, pp. 767-77, 2010
- [7] 辻光宏: 個人差モデル SINDSCAL と INDCLUS の音楽作品感情価分析への適用, 日本音楽知覚認知学会 2010 年秋季研究発表会, 2010
- [8] 辻光宏, 小宮山達也: 異分野の音楽に対する感情価共通布置でのクラスタリング視覚化に関する考察, 日本音楽知覚認知学会 2011 年春季研究発表会, 2011
- [9] 朝日新聞出版 編: 民力 2010. 朝日新聞出版, 2010.
- [10] R. Xu, and C. Wunsch : CLUSTERING. IEEE Press, 2009