

# 記号列化した人流データからの特徴抽出と可視化

宮城 優里<sup>†</sup> 大西 正輝<sup>‡</sup> 渡辺 知恵美<sup>§</sup> 伊藤 貴之<sup>†</sup> 高塚 正浩<sup>¶</sup>

<sup>†</sup>お茶の水女子大学 人間文化創成科学研究科 〒112-8610 東京都文京区大塚 2-1-1

<sup>‡</sup>産業技術総合研究所 人工知能研究センター 〒305-8560 茨城県つくば市梅園 1-1-1

<sup>§</sup>筑波大学 システム情報系 情報工学域 〒305-8573 茨城県つくば市天王台 1-1-1

<sup>¶</sup>シドニー大学 情報工学部 オーストラリア連邦ニューサウスウェールズ州 2006

## 1. はじめに

監視カメラで記録される映像から歩行者の行動パターンを見つけることにより、交通、防災、マーケティングなどの分野で有益な情報を得ることができる。しかし、蓄積された映像は膨大となるため、人流の全体像把握や重要な現象の発見を効率よく進めるのは困難な場合が多い。計算機を用いた人流解析の手法は開発途上にあり、特に大規模な人流情報の圧縮とマイニングについて検討する必要がある。

我々は人流情報の圧縮・マイニング・可視化に関する手法を提案している[1]。この手法ではまず UniversalSAX を適用し、人流データを記号列化して圧縮する。これらの記号列から自然言語処理手法によって特徴的な情報を抽出して可視化する。本報告では主に特徴抽出および可視化に関する進展を報告する。

## 2. 関連研究

人流や自動車などの交通流のパターンを可視化した研究としては Wang らによる研究[2]がある。数ヶ月間にわたり取得した車両の動きのデータをグラフ構造によって分析し、地図上へ描写することで交通量の推移や車両の速度を可視化している。GPS を用いて取得したデータをそのまま活用し、圧縮を行っていない点において改良の余地がある。矢田ら[3]は RFID によって買い物客の歩行動線を記録し、文字列化した上で行動パターン分析を行った。この研究では、売り場ごとの滞在時間が分析されており、さらにユーザの関心に沿って提示内容を変更できる可視化機能が実装されていない点で本手法と異なる。

## 3. 処理手順の流れ

本手法では、人流データを記号列に変換した後に特徴抽出を行い、その結果を可視化する。

### 3.1 人流データの取得と記号列化

図 1 は、モーションキャプチャデバイス Xtion を用いて取得した人流データを記号列化する処理の流れである。2次元時系列データの人流データを記号列化し、最終的にランレングス符号化する(詳細は[1]を参照)。この処理には「データ量の削減」と「特徴抽出をしやすい形式

への変換」という 2 つの意味を持つ。ランレングス符号の中で、文字の種類は歩行者の通過地点を表し、連続出現回数はそれぞれの場所での滞在時間の長さを表す。

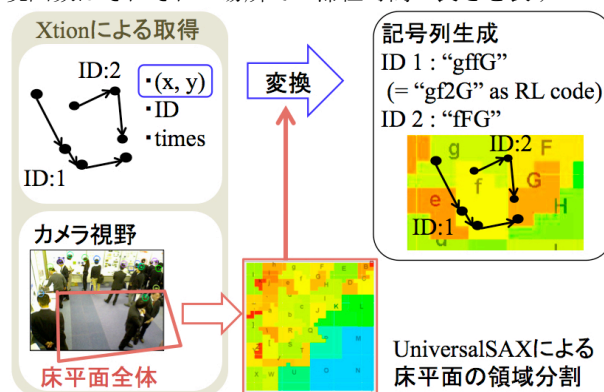


図 1: 人流データの記号列化の流れ

### 3.2 歩行経路の特徴抽出とそれらの比較

3.1 節で生成した記号列の特徴を調べることで、人流の性質を把握する。使用する記号列の形式は[1]で論じたとおりである。ここで抽出したい特徴は主に、歩行者の通過地点と滞在時間である。続いて、歩行経路を要約/検索するために、各歩行経路の特徴抽出結果を比較する。現時点で比較の処理として行っているのは、通過地点に焦点を当てての歩行経路間の非類似度の計算である。

非類似度の定義には重み付きレーベンシュタイン距離を利用し、記号列  $s_1$  を  $s_2$  に変換するために必要となる操作の合計コストの高さを非類似度とする。可能な操作は 1 文字の挿入、削除、変換の 3 種類である。1 回の操作におけるコストは、操作の種類だけでなく操作対象の記号に応じて、その都度決定する。具体的には、各記号に対応する地点間の距離からコストを算出し、非類似度に反映させる。上記の操作を繰り返し、 $s_1$  を  $s_2$  に変換し終えた時の合計コストを、存在するすべての変換手順について比較する。これらの合計コストの中での最小値を  $s_1$  と  $s_2$  の非類似度とする。

以上によって算出された非類似度の値に反映される歩行経路の特徴は、歩行範囲の広さや通過地点、向きの違いである。事前に各操作のコストを調整してから計算し、特定の条件に基づいた検索や分類を行うこともできる。例えば挿入と削除のコストを高く設定することで、記号列の長さの違い(=歩行範囲の広さの違い)に基づく分類が容易になる。

### 3.3 人流データの可視化

最後に、記号列化した人流データを、その特徴を強調

Feature Extraction and Visualization Using Symbolized People Flow Data

<sup>†</sup> Yuri Miyagi and Takayuki Itoh, Ochanomizu University

<sup>‡</sup> Masaki Onishi, National Institute of Advanced Industrial Science and Technology

<sup>§</sup> Chiemi Watanabe, University of Tsukuba

<sup>¶</sup> Masahiro Takatsuka, The University of Sydney

しながら可視化する。3.1節で作成した領域分割結果を参照し、各領域の中心座標を1つのノードとし、各ノード間を線で結ぶことで歩行経路を表す。大まかには目的の異なる以下の2通りの可視化を実現する。

1. 各領域についての概要表示
2. 選択した地点周辺の詳細表示

操作手順としては、1.の機能を用いて詳細を調べる場所を特定し、続いて特定された場所について2.の機能を使用する、といった手順を想定する。

1.では、ノードとエッジのサイズや色を変えることで各地点での通過人数や平均滞在時間を表現し、滞留が発生している地点などを可視化する。ノードの円の半径やエッジの太さは人数を表し、赤から黄色まで変化するノードの色は平均滞在時間を示す。

2.ではノードを1つ選択し、その場所に関するより細かな情報を可視化する。例えば、選択地点を通過した歩行者の行動を、動線アニメーションで表現する。各歩行者の具体的な歩行経路の把握や、通過時間に基づいた動線の色分けにより交通量の遷移を確認することができる。さらに、多くの歩行者が選択地点の通過前後に立ち寄りやすい場所の表示も可能である。

さらにまた、歩行経路を表す記号列を入力とする経路検索も行うことができ、与えた記号列との非類似度が閾値以下の歩行経路のみの描写も可能である。

#### 4. 展示会における人流分析の一例

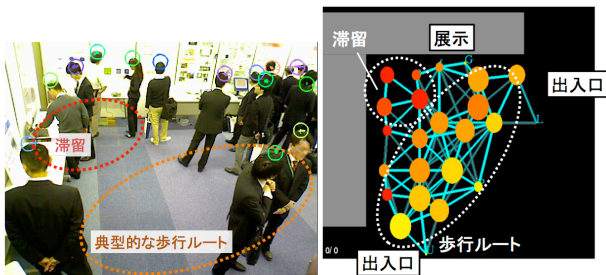


図2: (左) Xtionの視野 (右) 滞留表示結果

図2は、ある展示会場で8時間にわたって取得した、延べ人数5,531人分の人流データを可視化した実行例である。部屋の左端から上端にかけて展示物があり、出入口は下端と右上の2カ所がある(図2左)。

まず滞留発生地点を把握するため、各地点の通過人数と平均滞在時間を表示し歩行者の多かった場所を調べた(図2右)。展示物前の付近は、平均滞在時間が長いことを示す赤い円が多くなっている。特に左上隅の円は半径が大きく、この付近で足を止めた人が最も多かったことが読み取れる。一方下部から右上にかけての領域には、短

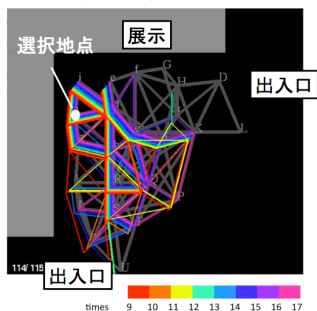


図3: 展示物付近を通過した歩行者の動線

時間で通過されやすかったことを意味する黄色やオレンジの円が目立ち、多くの人々がスムーズに通過した歩行ルートにあたるのがわかる。

図3は、展示物付近のノードを1つ選択し、その場所を通過した歩行者の具体的な歩行経路と、交通量の時間変化を観察した結果である。9:00-16:00の8時間を8色に色分けし、各歩行者の歩行経路を動線として描写した。展示物付近は赤から紫までの8色すべてが確認でき、どの時間帯でも人が絶えなかったことが読み取れる。特に14:00-15:00の歩行を示す青い線が多く描写され、この時間帯が最も混雑していたことが推測できた。

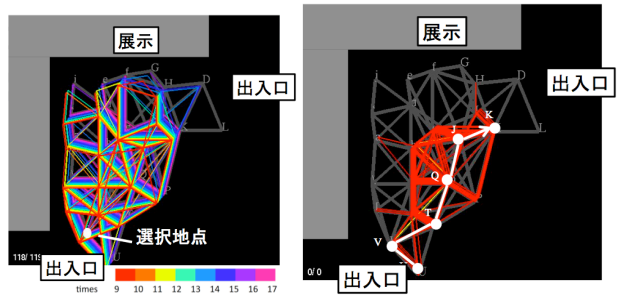


図4: (左) 下の出入口を通過した歩行者の動線 (右) 経路検索結果(白:入力 赤:検索結果)

続いて、展示物から離れた場所のうち、スムーズに歩行者が移動していた場所について分析した。下端の入口を通過した歩行者の歩行ルートを可視化すると(図4左)、展示物付近より典型的な歩行ルート上(図2右)の方が、多くの動線が描写された。具体的に、展示物の方に行かず2つの出入口を結ぶ“UVTQJK”というルートを検索すると、類似した歩行パターンが実在し、展示物を素通りした人が一定数いたことが確認された(図4右)。

このように、記号列化した人流データからいくつかの行動パターンを抽出することができた。データの記号列化により、時刻や歩行経路に関する情報の粒度は粗くなっている。一方で、記号化前のデータがおよそ669MBだったのに対し、可視化に使用したデータは62KBの記号列データと53KBの領域情報のみだった。このことから、人流の特徴を残しながらデータを軽量化することに成功し、さらに短時間で人流の特徴を把握できる手法を開発できたと考える。

#### 5. まとめと今後の課題

本報告では、記号列化した人流データから通過地点や滞在時間などの特徴を抽出し、それらの結果を概要表示と詳細表示の2段階で可視化する手法を提案した。今後の課題として(1)歩行経路要約結果の可視化(2)経路検索機能拡張(3)ユーザによる領域形状指定などが挙げられる。

#### 参考文献

- [1] 宮城, 大西, 渡辺, 伊藤, “記号化による人流データの圧縮と可視化”, 第18回画像の認識・理解シンポジウム, SS5-9, 2015.
- [2] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu, “Visual Exploration of Sparse Traffic Trajectory Data,” IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1813–1822, 2014.
- [3] 矢田, “スーパーマーケットにおける顧客動線分析と文字列解析”, 統計数理, vol. 56, no. 2, pp. 199–213, 2008.