

テンソル因子分解による欠損のあるセンシングデータの補間に関する研究

木下 和弥* 後藤 孝行** 高野 茂** 谷口 倫一郎**,**

(九州大学 *大学院システム情報科学府 **共進化社会システム創成拠点 ***大学院システム情報科学研究所)

1 はじめに

我々は人流データ分析に関する研究開発を進めており、キャンパス内に 14 台のポール型センサーノード (P-Sen) を設置し、歩行者の人流を収集、蓄積している [1]. この時、通信障害や障害物による遮蔽などの要因でデータに欠損が生じる. 人流を効率的に分析するためには、高精度な欠損データの補間が必要不可欠である.

本研究では、P-Sen が収集する人流データを、ある時間間隔における人量データに変換し、その欠損値を補間するためのテンソル因子分解 (Tensor Factorization, TF) [2] に基づく手法を提案する. 提案手法では、欠損のある人量データを、人流に関連するテンソルとして表現し、周囲のデータの関係を考慮したデータ補間を実現する.

2 提案手法

2.1 人量テンソル

各 P-Sen では、Laser Range Finder により約 0.1 秒間隔で前方の人物を検出し、その位置座標と、追跡を可能とする人物 ID を得ることができる. このように収集・蓄積された人流データから、単位時間ごとの人量を算出することにより、その場所の混雑度の把握や予測などのアプリケーションへの応用が考えられる. 提案手法では、人量データを、日付、時間、センサー番号をモードとする 3 階のテンソル $U \in \mathbb{R}^{I \times J \times K}$ で表現する. I, J, K は各モードの次元数であり、テンソル U の要素 U_{ijk} は、 i 日の j 時間帯における k 番の P-Sen での平均検出人数である. ここで平均検出人数とは、この j 時間帯における一回のセンシングで検出される人数の平均であり、その時間帯の混雑の指標となる.

2.2 テンソル因子分解による補間

欠損のある人量テンソル U に、TF を適用してデータを補間する手法を提案する. まず、テンソル U の欠損部に、初期値として日付モードに関する平均値を代入する. 次に、テンソル U に対して TF を適用すると、人量テンソルは次式のように近似される.

$$U_{ijk} \simeq \hat{U}_{ijk} = \sum_{i'=1}^{I'} \sum_{j'=1}^{J'} \sum_{k'=1}^{K'} Z_{i'j'k'} D_{ii'} H_{jj'} P_{kk'}.$$

ただし、 $Z \in \mathbb{R}^{I' \times J' \times K'}$ はコアテンソル、 $D \in \mathbb{R}^{I \times I'}$ 、 $H \in \mathbb{R}^{J \times J'}$ 、 $P \in \mathbb{R}^{K \times K'}$ は、それぞれのモードの相関情報をもつ因子行列である. 実際には、 U と \hat{U} との平均二乗誤差 (Root-Mean-Square Error, RMSE) を最小とする Z, D, H, P を求める. 提案手法では、 U の欠損部に \hat{U} で算出された補間値を代入し、この補間値が収束するまで U に TF を繰り返し適用する.

3 実験と考察

実験では、14 台の P-Sen で収集した 35 日間の人流をもとに構成した 3 階人量テンソル U を対象とし、人工的に加えた欠損部 (以下人工欠損部) に対する補間の精度を確かめる. それぞれのモードの次元数は、日付が 35、時間が 24、センサー番号が 14 となる.

テンソル U の全要素のうち一定の確率でランダムに欠損させた人量テンソルに対して提案手法 (ADTF) を適用し、人工欠損部の真値と補間値に関する RMSE を算出した. 20 回の試行による平均値を表 1 に示す. ここで、AD では提案手法の初期値をそのまま補間値とした. この結果から、欠損が多いデータの場合は、 $(I', J', K') = (6, 8, 4)$ のように各モードの次元数を適切に選択することで、補間の精度を上げることができると考えられる.

次に、各 P-Sen のデータが欠損した場合に、他の P-Sen のデータから欠損データを補間することができるかを調査した. その結果、精度よく補間できる P-Sen とそうでない P-Sen とに大別されることがわかった. これは、特定の P-Sen では、独立したデータを収集する傾向があったといえるが、上述のパラメータ選択の件と合わせて、本件に関しては、より深い考察が必要である.

4 まとめ

本研究では、テンソル因子分解による欠損のある人量データの補間手法を提案した. 評価実験では、適切な因子分解の次元数に関するパラメータを選択することにより、長期間の欠損データを効率よく補間することができることを示した. さらに、センサ間の相関を調べるために、各 P-Sen で計測される人量データが、他の P-Sen のデータから推定できるかを調査し、いくつかの P-Sen では、他の P-Sen との相関が小さく独立したデータを出力していることがわかった. 今後は、構成の異なる人量テンソルへの適用と、天気やシラバスなど他の関連データを考慮した人量データの補間手法について検討する.

謝辞

本研究は、独立行政法人科学技術振興機構の研究成果展開事業センター・オブ・イノベーション (COI) プログラムにより、助成を受けたものである.

参考文献

- [1] 高野茂, 後藤孝行, 辻徳生, 倉爪亮, 内田誠一, 森岡道雄, 谷口倫一郎, 村上和彰, 「地理情報システム上で活用できる人流・交通流ビッグデータ分析基盤の開発」, 全国共同利用研究発表大会「CSIS DAYS 2015」, 2015.
- [2] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4), 1324-1342, 2000.

表 1: U をランダムに欠損させた際の RMSE

ADTF			ランダム欠損率				
I'	J'	K'	5 %	10 %	30 %	50 %	70 %
1	1	1	1.00	1.01	1.02	1.03	1.03
6	8	4	0.53	0.53	0.55	0.61	0.77
10	15	6	0.40	0.40	0.49	0.66	0.90
15	21	9	0.30	0.35	0.56	0.79	1.02
19	28	11	0.25	0.36	0.65	0.86	1.06
AD			0.23	0.39	0.68	0.88	1.07