第 57 回土木計画学研究発表会・講演集

大規模かつ動的な旅行時間データの効率的な圧縮

原田 日郎[1]・Genaro Peque Jr.[2]・浦田 淳司[3]・井料 隆雅[4]

[1]学生会員　神戸大学大学院工学研究科市民工学専攻（〒657-8501 神戸市灘区六甲台町1-1）
E-mail: 171t139t@stu.kobe-u.ac.jp

[2]非会員　神戸大学大学院工学研究科市民工学専攻（〒657-8501 神戸市灘区六甲台町1-1）
E-mail: gpequejr@panda.kobe-u.ac.jp

[3]正会員　神戸大学大学院工学研究科市民工学専攻（〒657-8501 神戸市灘区六甲台町1-1）
E-mail: urata@person.kobe-u.ac.jp

[4]正会員　神戸大学大学院工学研究科市民工学専攻（〒657-8501 神戸市灘区六甲台町1-1）
E-mail: iryo@kobe-u.ac.jp

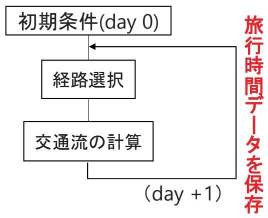
計算技術の発達により，Day to dayダイナミクスを用いた交通流シミュレーションが近年発展しつつある．これらのシミュレーションにおいて，各ドライバーは一つ前の反復計算での旅行時間を元に経路選択をするため，旅行時間を次の反復計算まで保存する必要がある．しかし，ネットワーク規模が大きくなると膨大なデータの保存が必要になるため，データ容量が限られる場合は圧縮して保存する必要がある．本論文では，シミュレーションの結果精度を維持するため，一定の精度で効率的に圧縮する手法を提案する．データ圧縮にあたり，区分線形曲線近似アルゴリズムを用い，圧縮後のデータから任意時刻における旅行時間を得るために二分探索及び線形補完を使用する．また，ノイズが多く含まれるデータに関しては，平滑化の後に圧縮することを提案する．

Key Words : piecewise linear approximation, data reduction, travel time smoothing, traffic simulation

1. はじめに

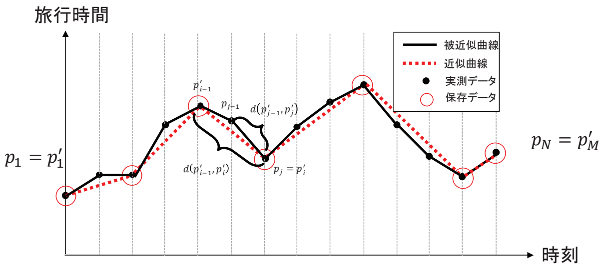
近年の計算技術の発達で，MATSIM [1]や TRANSIMS [2]のような Day to day ダイナミクスを用いた動的かつ大規模な交通ミクロシミュレーションが発展しつつあり，発展的な交通管理システムの構築がされつつある．これらのシミュレーションは，多くの反復計算により各行動者が効用最大化を行って，行動や経路を決定し，最終的には定常状態に到達する．実際のシミュレーションの経路選択では **図–1** に示すように，各ドライバーは一つ前の反復計算のリンク旅行時間の結果を元に選択を行う。そのため，各リンクの旅行時間を次の反復計算まで保存する必要がある．しかしながら，リンク旅行時間は当然リンク単位で生成され、ネットワークが大規模で計算対象の時間軸が長いほど，保存するデータ量は膨大になる．その際，コンピュータのデータ容量が限られる場合はデータ量を圧縮して保存する必要がある．

近年，マルチコア CPU による並列コンピューティングが浸透しつつあるが，経路探索においては既存研究 [3]が示すように，ネットワーク情報を分割せずに CPU ごとに持つほうが望ましい．しかしながら，ムーアの法則が曲がり角を迎え，主記憶メモリの容量増大の頭打ちになりつつあり，なるべくネットワーク情報のデータ量を小さくする必要がある．また，データを効率的に保存・取り出し可能とすることができれば，経路探索計算自体の高速化にも寄与するだろう．



図–1　交通流シミュレーションの計算概要

本研究では，リンク旅行時間データの効率的な圧縮手法を提案し，数値計算によりアルゴリズムの効率を明らかにする．対象とするリンク旅行時間は，時間依存のある動的データであり，この連続的に測定された旅行時間データを時刻で離散化して保存し、データ量を圧縮する．データ量を圧縮する際には，経路探索の精度を保つため



図–2　区分線形曲線近似法の概要

に，一定の圧縮精度が必要である．

精度を保証した曲線近似のデータ圧縮に関する既往研究として，Ramer[4], Pavlidis and Horowitz[5], Cox et al.[6]等がある．動的なリンク旅行時間データは，時間軸上では離散的な平面曲線として表現可能であるため，こうした手法を援用する．また，交通シミュレーションでは多くの繰り返し計算を行うため，近似の際の計算コストがなるべく小さいことが望ましい．

本研究では，データ圧縮にあたり，反復近似法 [4]を応用した区分線形曲線近似法 [7]を用いる．区分線形曲線近似法は，曲線上の点を数か所選定し，それらを繋ぐことで近似曲線を生成する手法である．提案するデータ圧縮アルゴリズムの特徴は，外生パラメータとして閾値近似誤差を与えることで結果を得ることができる．精度誤差を直接指定している点が強みである．また，経路探索に必要となる任意時刻のリンク旅行時間データは，圧縮後の点列から，線形補間と二分探索を用いて取得できる．

2. 計算手法

(1) 区分線形曲線近似法

A) 区分線形曲線近似法の概要

本論文ではデータ圧縮に区分線形曲線近似法 [7]を適用する．区分線形曲線近似法を以下にて説明する．**図–2** のように，N個の点から成る平面曲線（以下，被近似曲線）を，M個の点から成る区分線形曲線（以下，近似曲線）を用いて近似する問題を考える．被近似曲線を構成する点の集合をPとするとPは式(1)で表される．

$$P = \{p_j\}_{j \in \{1,2,\ldots,N\}} \tag{1}$$

近似曲線を構成する点の集合をP'とするとP'は式(2)で表される．

$$P' = \{p'_i\}_{i \in \{1,2,\ldots \mathrm{M}\}} \tag{2}$$

この時，式(3)の関係が成り立つ．

$$P' \subset P, \mathrm{N} > \mathrm{M} \tag{3}$$

区分線形曲線近似法において最適な近似を行うことは，Pから最適なP'を選択することに等しい．したがって以後は，最適なP'を選択する手法について説明する．

ここで，ξはjとiの対応を表す関数であり，以後ξを点選択関数と呼ぶ．$p_j = p'_i$（点p_jと点p'_iが同じ点）の時，jとiの関係はξを用いて式(4)のように表される．

$$j = \xi(i) \tag{4}$$

点集合Pの初期要素であるp_1，最終要素であるp_Nは必ず，P'の集合に含まれることに留意する．

$$p_1 = p'_1, p_N = p'_M \tag{5}$$

被近似曲線の総延長と近似曲線の総延長の差($E(P')$)を最小にするP'の選択が最適な選択と定義する [7]．すなわち式(6)が成り立つとき，点集合P^*が最適な近似曲線を構成する．

$$E(\mathrm{P}^*) = \min E(P') \tag{6}$$

$E(P')$の計算手法について説明する．$d(p_k, p_l)$はp_k, p_l間のユークリッド距離とし，$E(P')$は式(7)で表される．

$$E(P') = \sum_{j=2}^{N} d(p_{j-1}, p_j) - \sum_{i=2}^{M} d(p'_{i-1}, p'_i)] \tag{7}$$

式(7)を式(6)に代入すると，次となる．

$$\begin{aligned} E(P^*) = \min[&\sum_{j=2}^{N} d(p_{j-1}, p_j) \\ &- \sum_{i=2}^{M} d(p'_{i-1}, p'_i)] \end{aligned} \tag{8}$$

簡略化のため，次と置く．

$$L = \sum_{j=2}^{N} d(p_{j-1}, p_j) \tag{9}$$

$$L'(P') = \sum_{i=2}^{M} d(p'_{i-1}, p'_i) \tag{10}$$

ここで式(9)は被近似曲線の総延長であるため定数であるが，式(10)は近似曲線の総延長であるため，P'の関数である．式(9), (10), (8)から式(11)が導出される．

$$\begin{aligned} E(P^*) &= \min_{P'}(L - L'(P')) \\ &= L - \max_{P'} L'(P') \end{aligned} \tag{11}$$

式(11)から，最適な近似を求める問題は，近似曲線の総延長$L'(P')$の最大化問題と等価であることがわかる．

B) 計算アルゴリズム

前項の最適化問題を解くためのアルゴリズム [7]について説明する．上記の近似曲線の総延長$L'(P')$の最大化問題の最大値は明らかに$P' = P$となる時であり， Mが多いほどL'は大きくなる．つまり，分割数(M)が大きくなるほど近似の精度は向上，すなわち$E(P^*)$は小さくなる．計算アルゴリズムでは，近似曲線の分割数を反復計算で増加させ，徐々に$E(P^*)$の値を小さくする．これを$E(P^*)$の値が一定値以下になるまで続ける．反復計算における分割過程を**図–3** に示す．効率的に分割するため，最も誤差が大きい区間（式(12)の区間k_{max}）を分割する．なおg_kは区間kにおける被近似曲線，近似曲線の総延長の差である．

$$k_{max} = arg\,max\,g_k \qquad (k=1,2,\ldots,M\text{-}1) \qquad (12)$$

そして，近似誤差εが設定された閾値近似誤差ε_0を下回れば計算を完了し，近似曲線を得る．ここでの近似誤差εは式(13)のように被近似曲線の総延長に対する$E(P^*)$の割合であることを注意する．

$$\varepsilon = \frac{E(P^*)}{L} \times 100 \qquad (13)$$

区分線形曲線近似法はユークリッド距離の反復計算を必要とするが，ユークリッド距離の反復計算コストを節減するために，Memoization[8]を用いて計算する．

実際の過程を**図-3** を用いて説明する．1 回目の過程では$k = 1$の区間が分割される． $p_1 - p_{14}$間において，近似曲線の総延長L'が最も大きくなるp_{10}が新しく分割される点として選択される．この部分が前項における最適化問題に該当する．2 回目の過程では$k = 1$の区間が分割される．$p_1 - p_{10}$間において，近似曲線の総延長L'が最も大きくなるp_7が新しく分割される点として選択される．この作業を，$\varepsilon < \varepsilon_0$になるまで繰り返す．

(2) 圧縮データからの旅行時間の取得

区分線形曲線近似により，旅行時間データは**図-4** の黒点のような離散的な点保存されている．ドライバーが時刻t'における旅行時間を必要とするとき，まず，時刻t'の直前時刻t_1, 直後時刻t_2を二分探索で求める．t_1，t_2が求まると，**図-4** のようにt_1，t_2を結ぶ一次関数$f(t)$を求め，$f(t')$を時刻t'における旅行時間として得る．

(3) ノイズデータに対するガウス平滑化

リンク旅行時間データをセンサーで取得した場合は，激しいノイズを含むことが想定される．そのデータに直接，区分線形曲線近似法を適用すると，ノイズに対する近似曲線が生成されてしまい，よい近似が得られない．このため，区分線形曲線近似を行う前に，データの平滑化を行う必要がある．平滑化にはガウス平滑化[9]を適用することができる．ガウス平滑後のリンク旅行時間データを$f(t)$は式(14)で表される．

$$f(t) = \sum_{j=t-\frac{w-1}{2}}^{t+\frac{w-1}{2}} \tilde{f}(t)\phi_{j-t} \qquad (14)$$

$$\left[\phi_{-\left(\frac{w-1}{2}\right)}, \ldots, \phi_{-1}, \phi_0, \phi_1, \ldots, \phi_{\left(\frac{w-1}{2}\right)}\right] \qquad (15)$$

なお，$\tilde{f}(t)$をノイズを含む観測されたリンク旅行時間データを示し， ϕ_{j-t}は正規分布$\mathrm{N}(0, \sigma^2)$から与えられる一次元ベクトル(式(15))の要素である．**図-5** の通り，ϕ_tは$\mathrm{x} = \mathrm{t}$となる生起確率を示す．また，一次元ベクトル(式(15))は y 軸対称であり，Memoization[8]により，実装の際の計算速度が向上する．

3. 計算結果

上述したアルゴリズムを実装し，区分線形曲線近似法を用いてテストデータを実際に圧縮し，その精度と効率

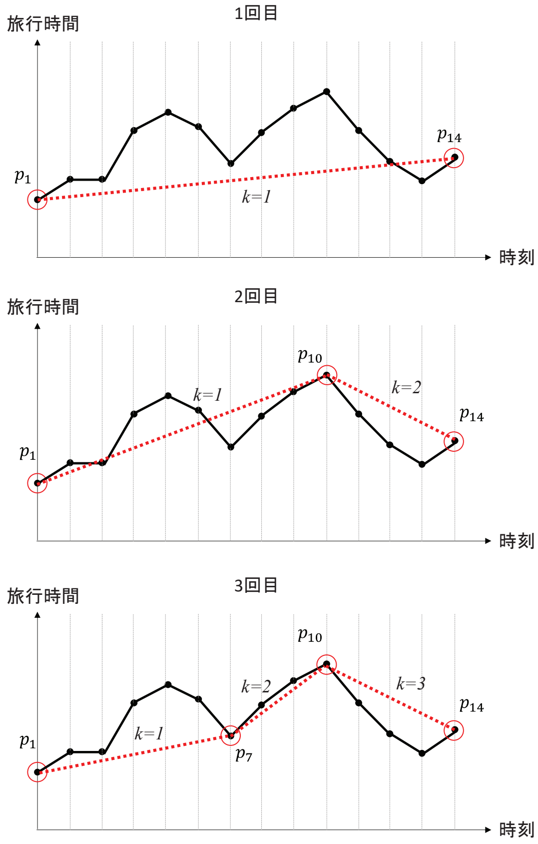


図-3 計算アルゴリズムの計算例

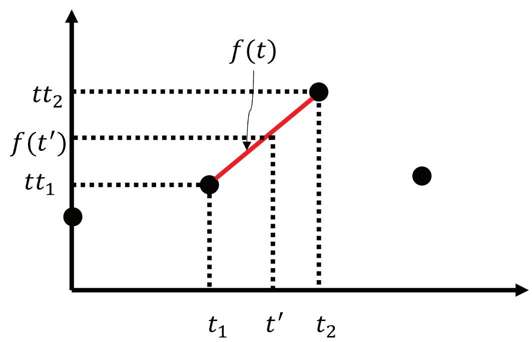
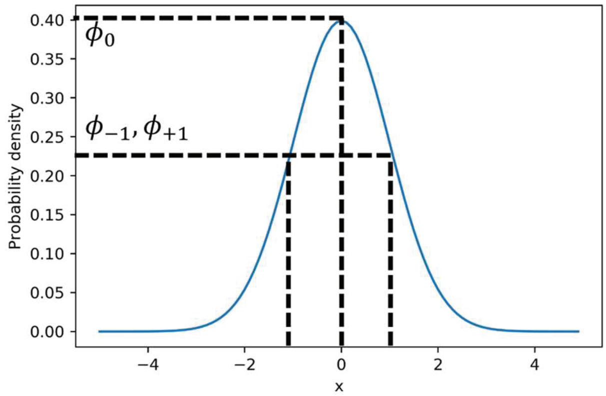
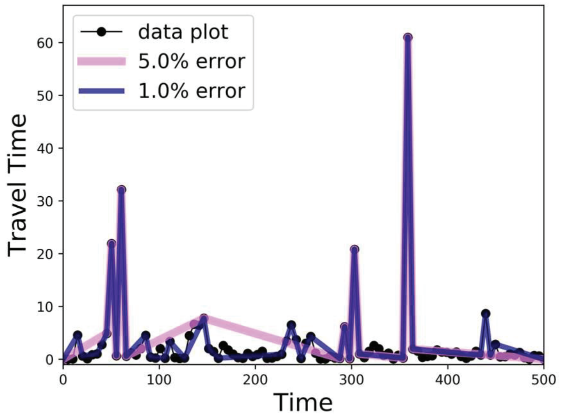


図-4 線形補間

性を検証する．まず，リンク旅行時間テストデータの作成方法を説明する．テストデータは時刻 5 単位ごとに旅行時間を対数正規分布を用いて生成したデータであり，データ数は 100 個である．テストデータは**図-6** の黒プロットで示す．**図-6** は横軸が時刻，縦軸がリンク旅行時間である．閾値近似誤差ε_0は 1%と 5%で設定し，圧縮アルゴリズムを適用した．**図-6** のグラフにおいて，濃い紫の実線はε_0を 1%で近似したときの近似曲線，薄い紫の実線はε_0を 5%で近似したときの近似曲線を表す．当然ではあるが，ε_0を 1%で近似した濃い紫の実線の方が



図−5　正規分布によるガウス平滑化



図−6　圧縮近似の計算結果

表−1　圧縮データ量の計算結果

閾値近似誤差	圧縮前データ数	圧縮後データ数
1%	100	34
5%	100	16

より黒点に沿っている（精度が高い）ことが確認できる．表−1 はデータ圧縮量の結果を示す． ε_0=1%の近似では66%のデータ量を圧縮，ε_0=5%の近似では 84%のデータ量を圧縮できた．圧縮後のデータ数は近似曲線を構成する点の個数(M) である．また，近似曲線の分割繰り返し計算は M-2 回行われている．したがって，閾値近似誤差 ε_0 を大きく設定すればデータ容量が縮減されるだけでなく，近似計算に要する計算コストの削減にも寄与する．本アルゴリズムは，インプットとなるリンク旅行時間データの質によっても圧縮後データ数は変化する．同じ閾値近似誤差 ε_0 の設定でも，変化が緩やかなデータであるほど大幅な圧縮が望めるだろう．

4.　結論

本研究ではリンク旅行時間データの圧縮から，圧縮されたデータから任意時刻における旅行時間を取得するまでの手法を提案した．旅行時間データの圧縮は区分線形曲線近似法を用い，任意時刻における旅行時間の探索は，二分探索・線形補間を用いた．旅行時間データ圧縮では，設定パラメータを閾値近似誤差を直接指定することができる．これにより，一定の精度を保証した圧縮を実現する．数値計算では，1%の閾値誤差を設定した場合にデータ量の66%を圧縮できることが確認できた．また，閾値誤差を5%に設定した場合はさらに84%のデータを圧縮でき，かつ圧縮にかかる計算時間も短縮できる．

本アルゴリズムの課題として，近似を行う前の段階で，全時間帯の旅行時間データが必要なことが挙げられるが，の Day to day のシミュレーション計算においては解決可能である．今後は，実際のシミュレータにアルゴリズムを適用し，メモリ，計算時間の効率性及び近似精度の影響を明らかにする必要がある．

参考文献

1) Horni, A., Nagel, K. and Axhausen, K.W. (2016). The Multi-Agent Transport Simulation MATSim. London: Ubiquity Press.
2) Smith, L., Beckman, R and Baggerly, K. (1995). TRANSIMS: Transportation analysis and simulation system. 10.2172/88648.
3) 福田 和輝，井料 隆雅，Wasuwat Petprakob，Lalith Wijerathne，浦田 淳司，Genaro Peque Jr. (2017) 大規模ネットワークにおける並列計算機を用いた動的利用者均衡配分の高速計算，第 55 回土木計画学研究発表会，CD-ROM.
4) Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. Comput. Graphics Image Process. 1, 244-256.
5) Pavlidis, T. and Horowitz, S.L. (1974). Segmenting of plane curve. IEEE Trans. Comput. C-23, 860-870.
6) Cox, M., Harris, P. and Kenward, P. (2001). Fixed- and free-knot univariate least-squares data approximation by polynomial splines. Algorithms For Approximation IV. Proceeding of the 2001 Inter. Symposium
7) Sato, Y. (1992). Piecewise linear approximation of plane curves by perimeter optimization.
8) Michie, Donald. (1968) Memo Functions and Machine Learning, Nature, No. 218, pp. 19-22.
9) Shapiro, L. G. and Stockman, G. C. (2001) Computer Vision, page 137, 150. Prentice Hall.

(2018. 4.27 受付)

EFFICIENT COMPRESSION OF LARGE-SCALE AND DYNAMICS TRAVEL TIME DATA

Hiro HARADA, Genaro Peque Jr., Junji URATA, Takamasa IRYO

High-resolution large-scale traffic simulations generate a large number of time-dependent traffic data for each link at each iteration that gradually changes over time. These are used to calculate important traffic information which in turn are used by drivers in the network to make travel decisions. Although high resolution traffic data can have high accuracy, these cannot be stored easily in a computer's memory because its capacity is usually limited. Moreover, if the size of the traffic data needs to be reduced, a certain level of threshold needs to be retained in order to maintain the accuracy of the simulation result. Thus, we are motivated in reducing data points in the time-dependent link traffic data without significantly compromising its level of accuracy with the least amount of computational resource.

In this paper, a piecewise linear approximation algorithm for planar curves is used to reduce the size of noisy or noiseless travel time data produced by a simulation. The algorithm fits lines on certain areas of the travel time data and discards data points based on some collinearity criteria. The algorithm then terminates once an exogenous error threshold is reached. Travel time value at any point in time can easily be retrieved using binary search and linear interpolation between two reduced data points it belongs to. Due to the gradual changes in travel time value over time, results show a significant reduction of data points while retaining a good approximation of the data.