

「近代東京ジオコーディングシステム」の活用 —1922年発行職業別電話名簿を用いて—

石川 和樹（首都大学東京大学院 都市環境科学研究科）

中山 大地（首都大学東京 都市環境科学研究科）

筆者らは、明治・大正期の住所を現在の位置座標へ変換する「近代東京ジオコーディングシステム」を構築し、無償で公開している。本稿ではこのシステムの概要について述べたあと、国立国会図書館のデジタルアーカイブから1922年発行の職業別電話名簿を用いて当時の魚商店の分布を地図化するとともに、それを用いた分析事例について述べる。

Application of the Modern Tokyo Geocoding System

- A Case of Using Classified Telephone Directory Published in 1922 -

Kazuki Ishikawa (Graduate School of Department of Geography, Tokyo Metropolitan University)

Daichi Nakayama (Department of Geography, Tokyo Metropolitan University)

We developed the Modern Tokyo Geocoding System which converts old address of Tokyo city in Meiji and Taisho era into coordinates (latitude/longitude) and opened this system on the internet freely. In this paper, we explain the outline of this system, and make a fish merchant map from the classified telephone directory published in 1922 on digital archive of national diet library, and analyze their locations.

1. はじめに

ジオコーディングとは、あらゆる地物や住所に対して位置座標を付与し、空間データとして扱えるようにする手法であり、主に住所を座標値に変換して地図表示する際に利用される。現代の住所を効率的に位置座標に変換するシステムとして、東京大学空間情報科学研究センター（CSIS）が提供する「CSV アドレスマッチングサービス」[1]が挙げられる。このシステムは現代の住所を位置座標に一括変換することが可能であるが、明治期や大正期の住所を位置座標に変換することはできない。そのため、歴史的な住所を地図化するには、旧版地形図や古地図などを参照しながら必要な住所の位置座標を特定する必要がある。しかし、大量の住所データを扱う場合は作業量が膨大になるため、この作業を効率的に行うシステムの構築が求められていた。

そこで筆者らは、歴史的な住所を位置座標へ一括変換するシステムを構築し、「近代東京ジオコーディングシステム」という名称で一般に公開した（URL: <http://www.bokutachi.org/geocoding>）（図1）。本システムはユーザー登録のみで誰もが利用することが可能である。本システムを利用することで、簡便な操作で大量の歴史的な住所に対して位置座標を付与し、地理情報システム等を用いて地図化することが可能である。また、地図

化の結果からさまざまな分析を行うことで歴史的な事象の把握ができる。そこで本稿では、本システムで利用している住所データベースやアルゴリズムについての概要を示し、その後、デジタルアーカイブにおいて公開されている1922年発行の史資料「職業別電話名簿」を用いて行った地図化事例や分析事例を提示することによりシステムの有用性を示すことを目的とする。

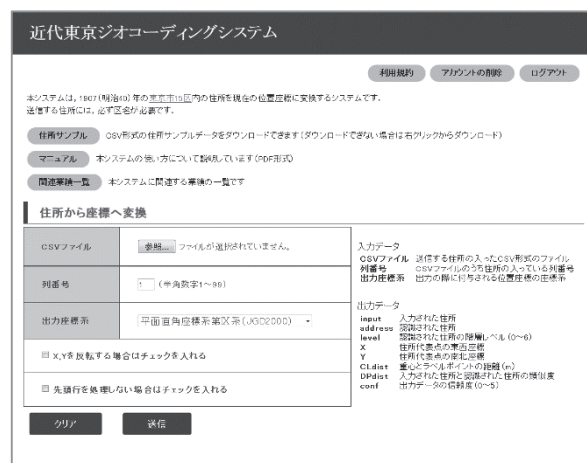


図1 「近代東京ジオコーディングシステム」のメイン画面

2. システムの概要

2. 1. 住所データベース

本システムは、歴史的な住所とその代表点座標などからなる住所データベースを利用することでジオコーディングを行っている。住所データベースの作成には、1907（明治 40）年に発行された「東京郵便局 東京市十五区番地界入地図 明治四十年調査」を用いた。この地図は各区 1 枚の計 15 枚枚なり、図 2 のように地図には当時の番地や番地界が記載されている。地理情報システムを用いて地図画像に座標を付与した後、番地界等に沿ってポリゴンデータを作成し、属性情報として住所や代表点座標値などを付与することで東京市 15 区内すべての住所データベースを作成した。このため、本システムで扱うことのできる住所の範囲は東京市 15 区内（図 3）となる。

2. 2. 住所検索アルゴリズム

本システムは歴史的な住所を扱うことから、住所の表記ゆれに可能な限り対応し、より正しい住所を認識する住所検索アルゴリズムが必要である。そのため、本システムでは、漢字データベース[2]を利用してシステムに入力された住所内の旧字体を新字体に変換し、その住所文字列を新字体で表記されている住所データベース上の住所において検索を行い、該当するものを取得する。検索は入力された住所の最初の文字から 1 文字

ずつ延ばしながら検索を行う。検索途中で検索結果が無くなった場合は、その 1 文字を任意の 1 文字に変換する。そしてその後ろの文字から再度検索を行い、再び検索結果がなくなるか入力住所を最後まで延ばして検索するまで継続する。これにより、旧字体以外の誤字などの表記ゆれについてもある程度対応することが可能となる。

住所のうち、丁目や番地、号の表記には算用数字や漢数字、また区切り文字には「一」や「ノ」が使用される場合があるなど、表現方法が多様である。そこで丁目以降の住所については数字に関連する文字を半角算用数字に、それ以外の文字列を「#」へ変換し、統一の形式へ変換することで検索を行う。たとえば、「一ノ二四」は「1#24」と変換される。この文字列を「#」で区切り、丁目、番地、号それぞれの列で検索を行い最終的な住所を取得する。

2. 3. システムの利用方法

本システムの入力フォーム画面と入出力データの例は図 4 の通りである。利用者は入力フォームにて住所の入力された CSV 形式のファイルを指定し、変換したい住所の入っている列番号を入力してシステムに送信する。そのほかに、出力する座標系（図 4 では平面直角座標系第 IX 系（JGD2000）を選択）の指定や出力座標値（X, Y）の反転、先頭行処理の有無を設定することが可能である。

システムの出力は表 1 のフォーマットの通りである。入力された住所（表 1 の input）とシステムによって認識された住所（表 1 の address）、その代表点座標値（表 1 の X, Y）のほかに、認識された住所の階層レベル（表 1 の level）や出力データの信頼度（表 1 の conf）などが伴って出力される。住所の階層レベルは表 2 の通りになっており、値が大きいほどより詳細な住所であることを示す。また、出力データの信頼度は値が大きいほど正しい住所を認識している可能性が高いことを示している。信頼度は入力された住所と認識された住所の類似性を示す値（表 1 の DPdist）に閾値を設定して求められており、位置的な精度とは無関係である。システムの利用者は地図化する住所の対象を住所の階層レベルや出力データの信頼度を用いて取捨選択することができる。



図 2 「東京郵便局 東京市十五区番地界入地図 明治四十年調査」の一部

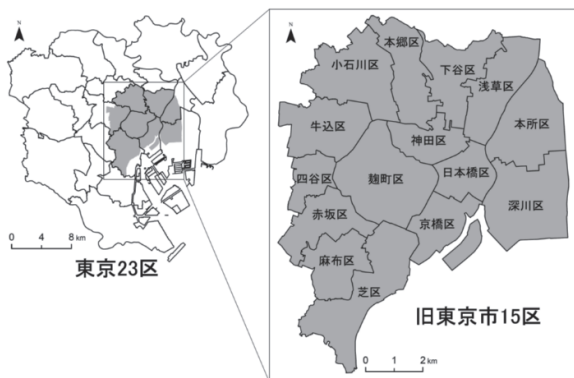


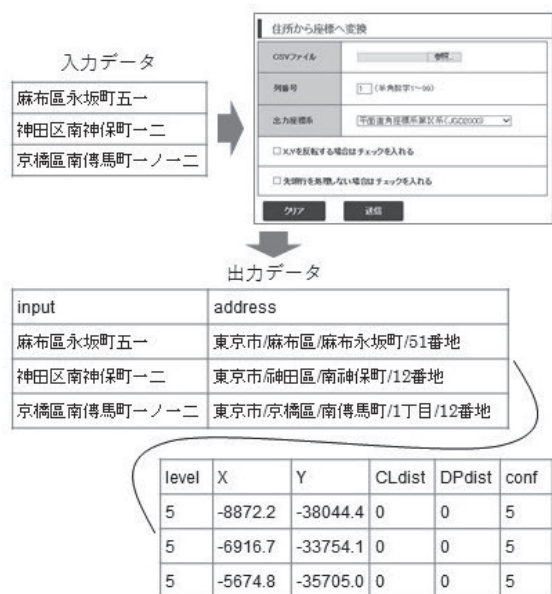
図 3 システムで利用可能な住所の対象範囲

3. 地図化事例

3. 1. 使用データと地図化の流れ

国立国会図書館のデジタルアーカイブである「国立国会図書館近代デジタルコレクション」において公開されている史資料に記載されている住所に対し、本システムを用いて住所を位置座標に変換し地図化した事例について述べる。

史資料には、1922 年発行の職業別電話名簿[3]を用いた。この資料は 1922（大正 11）年に発行された職業別の電話帳であり、電話番号をはじめ氏名や店名、住所などが記載されている。職業の



中から東京市内の魚商の住所（合計 637 件）を手作業にて入力し、CSV 形式のリストを作成した。作成したファイルをシステムに送信し、返ってきた位置座標を、地理情報システムを用いてポイントデータにした。その際、信頼度 4 以上の住所を地図化の対象とした（578 件）。表 3 はシステムに送信した魚商店の住所のうち、各信頼度の件数をまとめたものである。信頼度 5 となった住所は 546 件あるが、これらは入力された住所と認識された住所が完全に一致していることを示しているため、その全てにおいて正しい結果を返している。信頼度 4 となった住所は 32 件ある。信頼度 4 とは信頼度 4 となった住所データのうち、おおよそ 90% が正しく認識できる指標である。そのため、地図化の対象となる住所全てにおいて正しい結果を返しているとは限らず、場合によっては目視による変換結果の精査が必要である。このようにして作成した魚商店のポイントデータを、住所データベース作成の際に独自に作成した町丁目ごとのポリゴンデータを用いて地図化した。

3. 2. 結果と考察

図 5 は 1922（大正 11）年の東京市 15 区における魚商店の分布を示したものである。東京市の全域にわたって分布していることがわかるが、局所的に分布の集中している地域として、かつて日本橋に存在した魚市場（日本橋魚河岸）が挙げられる。日本橋魚河岸は江戸開府の頃に摂津国佃村・大和田村の有力者たちがこの付近に魚店を開いたのが始まりとされ[4]、関東大震災によって壊滅した後に築地市場に移転するまでの約 300 年の間利用されていた。また、隅田川河口東岸の魚商店の集積している地域は「深川漁師町」とい

表 1 出力データのフォーマット

ヘッダー名	説明
input	入力された住所
address	認識された住所
level	住所の階層レベル (0~6)
X	住所代表点のX座標
Y	住所代表点のY座標
CLdist	代表点と重心の距離 (m)
DPdist	入力された住所と認識された住所の類似性を示す値
conf	出力データの信頼度 (0~5)

表 2 住所の階層レベル

住所の階層レベル	値
検索結果なし	0
市レベル	1
区レベル	2
町名レベル	3
丁目レベル	4
番地レベル	5
号レベル	6

表 3 各信頼度の住所件数

信頼度	住所件数
5	546
4	32
3	22
2	15
1	20
0	2
合計	637

う総称で呼ばれた地域で、漁業が盛んであっただけでなく、魚市場も存在していた。

以上の事例のように、近代東京ジオコーディングシステムを用いることで容易に歴史的な住所を空間データ化することができ、本事例において示した魚商店のみならず、当時のさまざまな住所の分布を地図化することが可能となる。

3. 3. 地図化対象外の住所データ

本事例でシステムへ送信した魚商店の住所データは 637 件であるが、そのうち信頼度が 4 以上で認識結果が良好となった住所は 578 件であった。一方で、信頼度 3 以下であったため地図化の対象外とした住所が 59 件あった。本システムで住所の認識がうまくいかない要因には大きく分けて以下の 3 点が考えられる。①検索アルゴリズムで対応不可能な文字等が入力住所に存在する場合、②住所データベース上の住所表記自体に誤りがある場合、③入力された住所の時期が住所デ

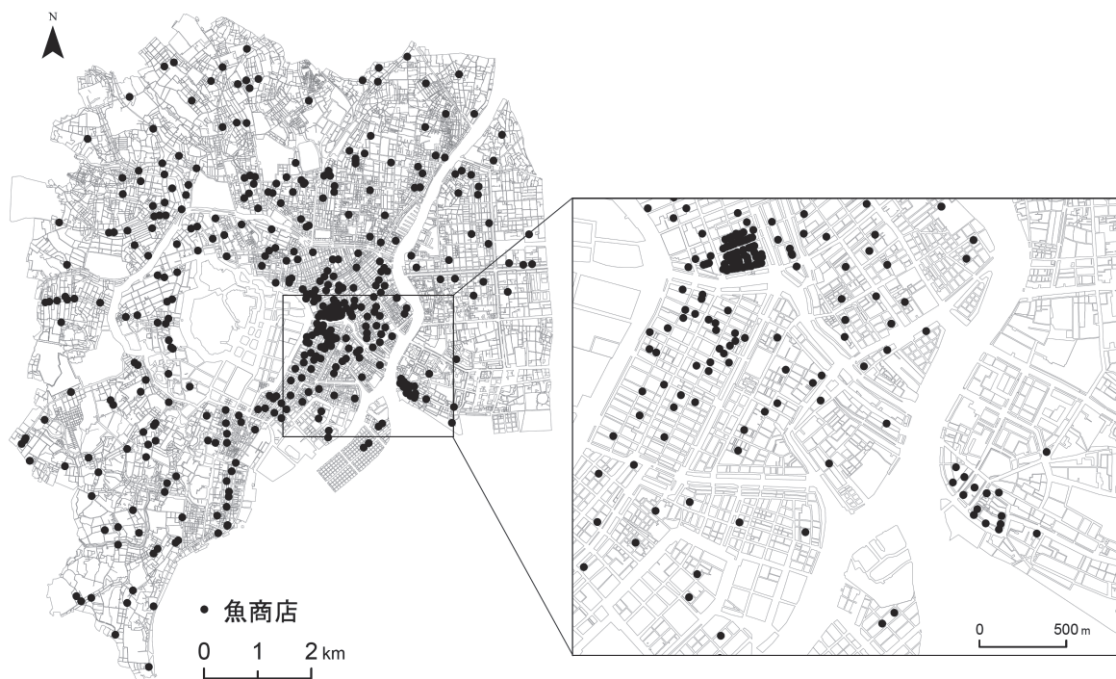


図5 1922年東京市15区における魚商店の分布

データベースの時期と異なるため、住所データベース上に住所が存在しない場合、である。

以下の分析では、入力住所の階層レベルと認識住所の階層レベルを比較して、認識がうまくいかない住所が生じる原因を考察する。入力された住所の階層レベルと認識された住所の階層レベルを表4にまとめた。対象は信頼度が3以下で認識がうまくいかなかったとされる59件である。

この中で最も多いパターンは番地レベル(レベル5)の住所を入力し町名レベル(レベル3)の住所が結果として返ってきたものである(29件)。これは、入力された住所の町名レベルまではデータベース内に存在しているが、丁目や番地のレベルが当該町名に存在していないことを示している。たとえば、ある時点では町名の下に直接番地がある住所スタイルの地域だったものが、のちに区画整理などで町名の下に丁目加わった場合などが挙げられる。これは前述した住所の認識がうまくいかない要因のうち、入力された住所とデータベース内の住所の時期的な違いによる③のパターンが大部分と考えられる。

二番目に多いパターンは、入力された住所と認識された住所の階層レベルの両者が番地レベル(レベル5)であるケースである。これらは入力された住所が住所データベース上に存在しないが、システムが類似した住所を結果として返しているケースである。このケースの原因は①と②のパターンに相当すると考えられるが、どちらの原因に帰着するかを明確にすることは困難である。なぜなら、①と②はともに入力住所とデータベース内の住所の齟齬から生じるものであるが、齟齬

表4 地図化対象外データの内訳

入力された住所の階層レベル	認識された住所の階層レベル				
	6	5	4	3	0
6(号)	0	0	0	5	0
5(番地)	0	22	0	29	2
4(丁目)	0	0	0	0	0
3(町名)	0	0	0	1	0

単位：件

の原因が入力された住所の誤りに起因するのか、データベース内の住所の誤りに起因するのか、それとも両者の誤りに起因するかを特定できないためである。

たとえば入力データに誤りがあることを証明するためには、ユーザーが正しいと考えている住所を知る必要がある。しかし、ユーザーが正しいと考えている住所が何であるかをシステムが知ることは不可能であり、入力された住所はユーザーが入力した正しい現存した住所であるという前提でシステムは検索を行わなければならない。このため、正しく認識された住所とは入力された住所とデータベース内の住所を比較した結果からしか決定できない。

表1にあるDPdistは入力された住所とデータベース内の住所をDPマッチングにより得点化したものであり、両者の一致度が高いほどDPdistが小さくなる。信頼度を示すconfはDPdistに基づいて算出しており、confが5の場合は入力され

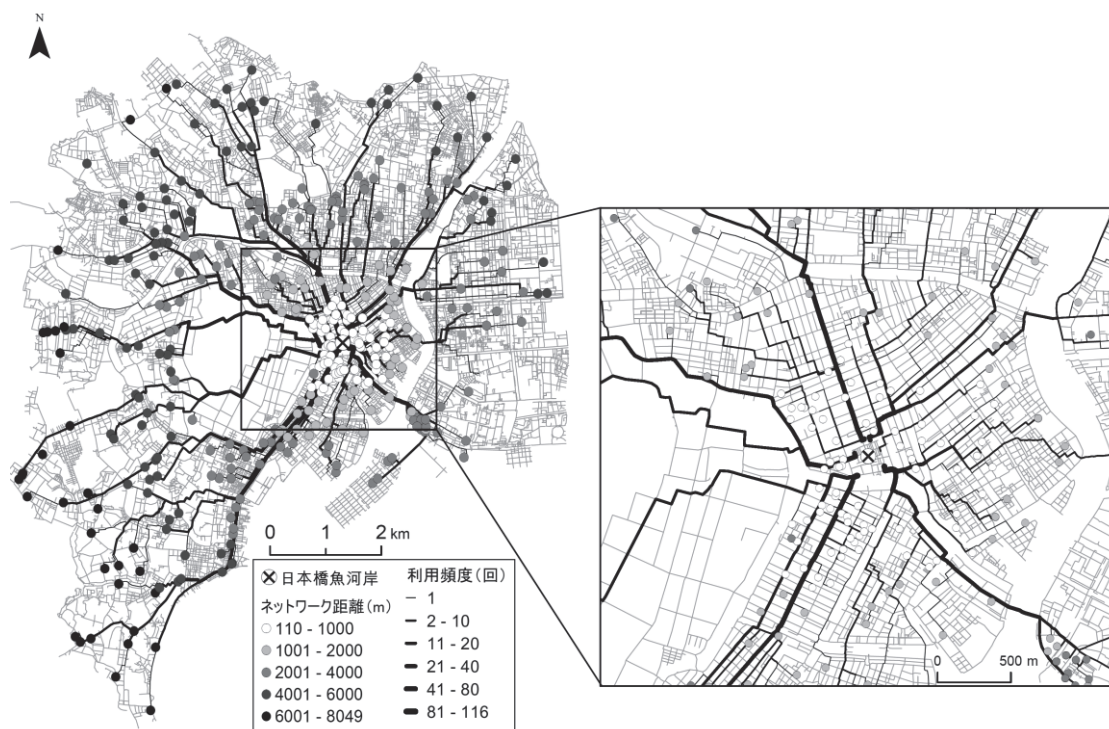


図6 日本橋魚河岸から各魚商店までのネットワーク距離と利用頻度

た住所とデータベース内の住所が完全一致することを示す. *conf* が4以下のものは入力された住所とデータベース内の住所が部分一致していることを示し, *conf* が小さくなるほど部分一致している文字数が少なくなる. ただし, *conf* が4の住所は結果としてその90%が正しく認識されている. また, たとえ *conf* が3以下だとしても, 中には正しい住所が返されている場合がある. これはたとえば入力された住所の一部に当て字などが使われている場合であり, このような例は目視により人間が判断する必要がある.

4. 分析事例

4. 1. 分析手法

本システムを用いることで, 歴史的な住所の地図化のみならず, その空間分布に関する分析を行うことが可能になる. これにより, 当時の空間的・社会的な構造や, 歴史的な事象を把握することができる. 本章では, 前章において作成した魚商店の分布図を利用して行った分析事例について述べる.

日本橋魚河岸に存在した魚市場は近海から魚介類が集められ, 市内の多くの小売店や料理店は, この魚市場へ出向いて, 仕入れを行っていた[5]. 本論では, 日本橋魚河岸から各魚商店への最短経路による道路ネットワーク距離を計算し, 市場から小売店へ魚介類が流通する際にそれらの輸送に利用される道路を推定した. 分析に使用した道

路ネットワークは, 「東京郵便局 東京市十五区番地界入地図 明治四十年調査」を用いて筆者らが独自に作成したものである. 魚商店分布と多少時期が異なるが, 分析にはほとんど影響しないと判断したためそのまま用いた. また, 前述した深川の魚市場は考えないものとした.

4. 2. 結果と考察

図6は日本橋魚河岸から各魚商店へのネットワーク距離と利用頻度を示す. 利用頻度は, 日本橋魚河岸から各魚商店へのネットワークを計算した際に, その道路が利用された頻度を表している. 魚商店は東京市全域にわたって分布しているため, 日本橋魚河岸から放射状に最短経路の道路ネットワークが形成されているのがわかる. また, 最長距離はおおよそ8kmであることがわかった. 先行研究[6]において, 1936(昭和11)年の例であるが, 当時の物資輸送ではその距離に応じて利用する運搬手段が異なることが示されており, おおよそ市場から4里(1里はおおよそ4km)以内が徒歩圏(荷車), 6里以内が畜力圏(牛・馬車)であるとされている. そのため, 東京市内であれば徒歩による輸送が主であったことがわかる.

また, 利用頻度についてみると, 日本橋魚河岸周辺に利用頻度の多い道路がいくつかあり, 特に南に向かって利用頻度が多くなっている道路は, 現在の中央通りである. この通りは当時から東京市のメインストリートとして機能してお

り、市の中心的地域として周辺には魚商店が多く立地していた。また、旧東海道の一部でもあることから周辺に魚商店が多くみられ、物資の輸送の際には多く利用されていたことが推測される。

現中央通りにおける荷車の交通量をみるために、当時の交通量調査データを使用して荷車の交通量を示す。データには、東京市交通調査統計表大正十四年六月三日調査[7]を用いた。この調査は1925年6月3日に行われた大規模な交通量調査で、東京市内とその周辺291地点で観測が行われた。調査項目には手荷車をはじめ乗用自動車や自転車など13項目にも及んでいる。ただし、本調査は1923年の関東大震災後に実施されたものであり、震災によって日本橋魚河岸が壊滅的な被害を受けたことから発災数か月後には築地に魚市場が移転していることに留意する必要がある。震災以前に同様の交通調査はないことから、ここでは、本調査によるデータを用いる。

日本橋魚河岸周辺の中央通り沿いにおいて交通量の調査が行われた地点を図7に示し、北から順に番号を振った。各地点における手荷車の交通量は図8の通りである。どの地点においても6時台から7時台の早朝の時間帯に交通量が多くなっており、それ以降は急激に交通量が減少している。このことから、現中央通りでは、魚市場や青果物市場からの物資の輸送が頻繁であったことがわかる。

5. まとめと今後の課題

本稿では、「近代東京ジオコーディングシステム」の概要を示し、その後デジタルアーカイブの史資料を活用して歴史的住所の地図化を行ったほか、実際に分析を行うことでシステムの有用性を確認した。本システムを利用することで、容易に当時の地物の分布を再現できるようになるほか、本システムが、歴史学や地理学における新たなツールのひとつとなることが期待される。

システムの発展のためには、住所データベースの拡張が欠かせない。現在のシステムでは、おおよそ1890年代から1920年代の住所において高い変換精度をもつことを確認している。しかし1930年代以降は町名や番地の整理が多くなされ、その結果システムでの変換精度が落ちている。そのため、異なる時期の住所データ整備は不可欠で、これにより時期の異なる住所が入力された場合でも検索精度を高いまま維持することができる。また、異なる時期の住所データベースが整備されることで時系列の分析にも対応することが可能となるだろう。

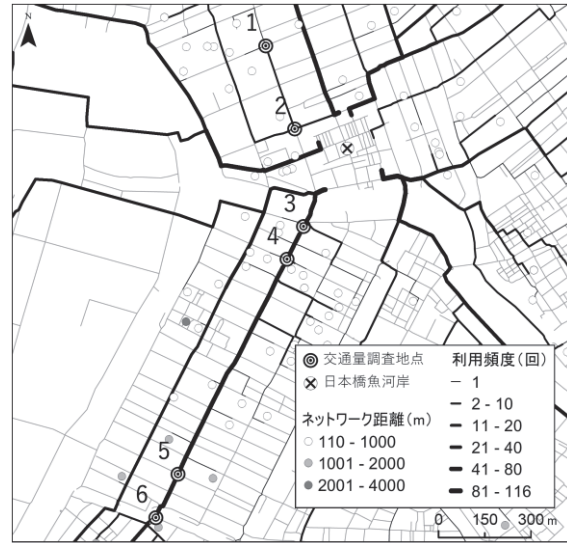


図7 中央通りにおける交通量調査地点

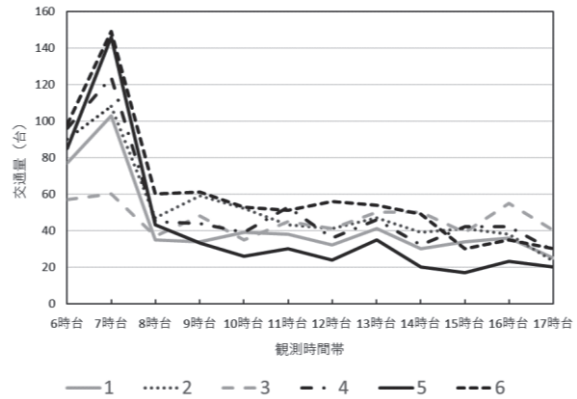


図8 中央通りにおける荷車交通量の時間変化

参考文献

- 1) 相良 毅, 有川正俊, 坂内正夫: 分散位置参照サービス, 情報処理学会論文誌, Vol. 42, No. 12, pp. 2928-2940 (2001).
- 2) 漢字データベースプロジェクト: 漢字データベース, 入手先 <http://kanji-database.sourceforge.net/> (参照 2016-10-30).
- 3) 日本商工通信社: 職業別電話名簿, 日本商工通信社 (1922).
- 4) 北原 進: 中央区の歴史, 名著出版 (1979).
- 5) 原田東風: 諸種営業実地商業案内, 大学館 (1905).
- 6) 永森忠正: 市場距離と運輸方式との相関—千住青物市場調査 第一報—, 地理, Vol. 2, No. 3, pp. 333-348 (1938).
- 7) 東京市統計課: 東京市 交通調査統計表 大正十四年六月三日調査, 東京市役所 (1926).