

情報流のデータマイニング

— 通話からみた日本の構造と通話地図 —

Data Mining of Information Flow

— Phone Traffic Space and Phone Traffic Map of Japan

古藤 浩

KOTOH Hiroshi

長谷川文雄

HASEGAWA Fumio

Abstract: In this paper, we discuss tendencies of telephone call frequencies in Japan. In our analysis, we developed a new data-visualization method using backward-calculated distances from gravity model. Tendencies of phone traffic data seem to follow the gravity model as a whole, but if we analyze details, the tendencies are very complicated. The reason might be influences of big cities or regional geographical features. We do not develop a model that explains the general phone traffic structure, but we develop a new method for visualizing the regional structure of phone-traffic. We use the gravity model as a basis, and using the concept of backward-calculated distance. Finally, the multi-dimensional scaling method is applied for visualizing the data. The concept of backward-calculated distance was developed by Plane(1984), in analyzing the migration space. Our method is unique because we apply the backward-calculated distance into multi-dimensional scaling method. We calculated distances from non-constrained gravity model, and it enables us to analyze regional structure in many aspects simply. We analyze Japanese phone traffic data which is between 567 MAs'.

はじめに

データマイニングとは大量のデータを想定し、マーケティングなどに活用するために、その傾向をつかみ、分析するための一連のプロセス・システムを意味する。それは大まかには、データベースの構築・データの洗浄・データの整理／分析と解釈、といった手順を踏むと言える。データの整理／分析では、その「視覚化」といった技術も重要視される。最終的には判断は人間がする以上、如何にしてデータ分析結果を浮き彫りにするかということとはデータマイニングでは重要な問題である。

本研究では、通話データとその視覚化に注目して、その方法と結果を説明する。それは、地域間のつながりの傾向や通話の構造を分析する方法の提示と、それによる日本の通話に関する構造分析となる。

地域のつながりの強さは、人や物の動き、通話の回数、時間などから知ることができよう。人口移動や通勤などの面からの研究が進んでいる一方（石川, 1994; Plane, 1984; Yano, 1993）、情報通信、通話トラフィックに関する研究も、近年の情報化の進展に応じて、研究が活発になっている（川手他, 1994; 山崎他, 1993; Rietveld et al., 1990; Murnion et al., 1998; Shen, 2000; Katz et al., 1998）。これらの研究は通信インフラの整備への指針を示す上で重要である。

通話の活発さと距離の関係は、発着点の距離が長くなるにつれて急減することが知られている。このような傾向の説明には指数減衰型のいわゆる重力モデルが基本として使われることが多い。しかし、実際には重力モデル

だけでは十分説明できない。通話トラフィックは全体としては重力モデルに従う傾向でも、予測値と真の値を比べるとかなりの誤差が見られ、またそれはモデルを多少改良しても説明力の大幅な向上を得ることは非常に難しい(高嶋, 1997)。

その原因は地域個々の要因のためと考えられる。それは三種類の要因に分けて考察することができる。第一はいくつかの地域グループの存在であり、グループ間とグループ内では同じ距離でも通話の回数が異なる傾向がある(Mackay, 1958; Rietvelt and Janssen, 1990; Noronha et al., 1992)。それは例えばMackay, J.R.(1958)によって国内通話と国際通話のずれとして説明されている。Mackayの研究はカナダから北米への通話の分析であるが、三つのグループに分けられることを示している。それはカナダケベック地域、その他のカナダ、アメリカである。ここからは国境と言語によるグループ化が推定できる。国や言語だけでなく、様々な理由について考察がなされている(Nijkamp et al., 1990; Noronha et al., 1992)。

第二は地域個々の要因である。首都であること、地域の核都市であることは、人口等によるウェイトを勘案しても、通話では活発なことが確認できる。

第三はその他の要因で、第一、第二の要素を勘案しても残る分である。例えば交通網の影響、歴史的な地域の関係等の理由が考えられ、それもかなりの重要性を持つ。

これら三つの要因を勘案し、本研究ではそれを説明するのではなく、そのばらつき具合を、視覚化し、構造の理解を支援することを目的とする。具体的には逆算距離の概念と多次元尺度構成法を応用し、通話の地域構造を仮想地図で示す。構造の理解を目的とした研究では、主にクラスター分析を利用したものをいくつか挙げる事ができる(長谷川他, 1985; 中村・長谷川, 1991; 山崎他, 1993)。本研究は、人口とのマッチングによって規模の影響を除いた上での通話の構造を見ること、逆算距離の導入によって多次元尺度構成法の利用を導き、構造を二次元的に視覚化したことの点で、これまでにない知見を得ることを可能とした。なお、通話データに多次元尺度構成法を活用することは長谷川他(1985)でもおこなっているが、それは「負の結合度におけるMDS」という特別な地域を浮き彫りにするための特殊な方法の議論であり、本研究とは目的が異なる。また本研究での結果は日

本全国562MAによる通話地図を示すので、規模の大きい応用結果を示す。

なお、逆算距離の概念はPlane(1984)によって導入され、二重制約形の空間相互作用モデルによって米国の人口移動の地域構造を視覚化している。また、Noronha et.al(1992)も同様の概念を提唱している。本研究での方法は無制約形の空間相互作用モデルを使うので、Planeの方法に比較して、理論的な面では課題を残すが、簡便に活用できるという利点を持つ。

以下では、2章で理論的なモデルや視覚化の考え方について説明し、3章ではデータの概要を説明した上で、分析・視覚化した結果を考察する。なお、データの単位は回数・時間の2種類、また契約形式は住宅用・事務用の2種類の計4種類の枠組みで分析した。

1. 理論

(1) 本研究での重力モデル

本研究では単位料金通話地域をMA(i.e. Message Area)と呼び、それを分析の単位地域とする。同一MA内では最低料金(本研究の分析データの時点では10円)で一定時間の通話ができる。日本は567のMAに分割されている。MAは通話の交流上、概ね一体と考えられる地域(NTT電話サービス契約約款より)であり、大概の場合、複数の市町村の集合体となっている。規模は2千人ほど(南大東)人から823万人(東京23区)までと様々である。

まず、いくつかの記号を定義する。

c_{ij} : MA iからMA jへの発信回数

p_i : MA iの人口

d_{ij} : MA iとMA jの間の物理的な距離

本研究では通話時間データも扱うが、ここでは発信回数を例にとりてモデルを説明する。以下の式は通話回数を通話時間に置き換えても、そのまま利用可能である。次に基準発信回数 y_{ij} を対応する関係二つのMAの人口の積で割った値

$$y_{ij} = c_{ij}/p_i p_j \quad (1)$$

と定義する。その考え方は、個人個人の組み合わせの数が通話発生の可能性と比例するという考えによる。異なるMA間では、大まかにはこの値が通話の回数を推定す

るための基準となろう。なお、MA内々での発信回数を考えると、MAの人口に比例する傾向があり、この仮定は十分とはいえない。これは内々データの特異性に起因すると考えられるが、本研究でのモデルの計算には内々データを用いないこととした。

本研究での重力モデルは式(2)で与える。

$$y_{ij} = \alpha d_{ij}^{-\beta} \quad (2)$$

ここで、 α は重力モデルでの距離1(km)での推定値、 β は重力減衰係数(>0)である。式(2)は無制約型の空間相互作用モデルともいえる。

このモデルはその論理的な意味や、係数の次元等に関連していくつかの問題が指摘されている(Hynes, 1975; 石川, 1988)。しかし、後述するが通話データの大まかな傾向の説明はこのモデルで十分と思われたこと、単純な式なので応用が容易であり、簡明に逆算距離が定義できるといった利点を持つことに注目し、本研究ではメカニズムとしての重力モデルという意味よりは、全体的な経験的傾向として式(2)を採用し、そこからのずれによって地域構造を分析することにした。

なお、モデルとしては二重制約形の空間相互作用モデルや対数線型モデル(矢野, 1991; より詳しくはYano, 1993)、ロジットモデルの方が理論的には進んだ面を持っているよう。しかし、二重制約型のモデルや対数線型モデルの場合は、扱いが複雑で、得たパラメータの解釈も難解になるという問題がある。また距離や規模以外のパラメータを含むモデルを作る場合はロジットモデルが非常に有効であるが、本研究の場合、データの特異性の理由でそのようなパラメータを加えることが困難という背景がある。

一方、山崎他(1993)で使っているような、規模の効果に関する係数を導入することも検討したが、データに対する説明力が期待したほどの効果を上げなかったので、本研究では最も古典的なモデル(式(2))を使うことにした。

係数の推定では、式(2)の対数をとると、

$$\log y_{ij} = \log \alpha - \beta \log d_{ij} \quad (3)$$

という直線関係となることを用いて係数 α 、 β を推定する。非線型の式(2)のままの係数推定も可能だが、その場合、遠距離の通話がほとんど無視されることになる。対数をとることで、遠距離間の傾向も生かすことができるので、全体的な構造を知る上では、この方法は有効で

ある。

(2) 逆算距離による通話空間の視覚化

ここでは、逆算距離を用い通話地図を作成する方法を述べる。まず発信回数から対応する距離を逆算して求められる「通話距離」を定義する。それは重力モデルの適用によって得る係数を利用し、基準通話回数を用いて、MA間の仮想的な距離を逆算することで定義される。重力モデルの係数が求められていて、基準発信量 y_{ij} がわかれば、MA間の通話距離 \tilde{d}_{ij} は式(4)のように得る。

$$\tilde{d}_{ij} = (y_{ij} / \alpha)^{-1/\beta} \quad (4)$$

次に通話距離を用い、多次元尺度構成法を応用して、通話距離からMA間の位置関係を総合的に図化する。具体的には清水(1992)による最小二乗法を用いた多次元尺度構成法を用いる。

仮想地図上でのMA i とMA j の距離を D_{ij} 、MAの数を n とおいたときに、最も単純に考えられる通話地図の作成法は、

$$\min \sum_{i=1}^n \sum_{j=1}^n (D_{ij} - \tilde{d}_{ij})^2 \quad (5)$$

という最小化問題によってMAの位置を決定する方法である。この方法は古藤・長谷川(2002)で採用しているが、通話量が少ないペアの比較では、ほんの少し通話量が異なるだけで大きく距離が相違するという問題がある。また、遠距離でのわずかのずれ(例えば1000kmに対する10kmのずれ)を近距離でのわずかのずれ(例えば20kmに対する10kmのずれ)と同じように評価することになるので、通話の量としては圧倒的な比率を占める近距離の通話が相対的にあまりに軽んじられることになると考えた。

この不都合を解消するため、本研究では距離に対数をとって推定に活用することにした。それは、次の目的関数

$$\min \sum_{i=1}^n \sum_{j=1}^n (\log D_{ij} - \log \tilde{d}_{ij})^2 \quad (6)$$

によって通話地図上でのMAの位置を決定する。

実際の計算では、通話の頻度が0回(時間)のデータは距離が無限遠になるので、計算から除かなくてはならない。その他にも年度によって不安定なデータを計算から除外したりする(後述)ので、データをリスト構造で管理することにし、データ数(利用するMAペアの数)

を m, h 番目のデータの発信地MAを c_h 、受信地MAを r_h 、通話地図上でMAiの座標を (X, Y) とおくならば、通話地図を作成するときの目的関数 S は、

$$\min S = \sum_{h=1}^m \left\{ \log \sqrt{(X_{c_h} - X_{r_h})^2 + (Y_{c_h} - Y_{r_h})^2} + \log(y_h / \alpha) / \beta \right\}^2 \quad (7)$$

と整理できる。式(7)の括弧内の第一項は仮想地図上での距離、第二項は通話距離を示す。この式から得る通話地図は、単位が距離(km)となるので、物理的な地図との比較が容易という利点を持つ。

通話地図の精度の評価は、通話地図上の距離と通話距離のそれぞれの対数値間の相関係数が考えられる。しかし、相関係数では通話地図の精度をどのようにとらえたらよいかははっきりしない。そこで、目的関数値から導かれる標準誤差を利用することとした。それは、 $\sqrt{S/m}$ と与えられる。推定結果での残差が正規分布に従うならば、MA間の位置のずれは、 $\pm \sqrt{S/m}$ の範囲に約3分の2のデータが収まっていることを意味する。式(7)は常用対数を用いているので、10の $\sqrt{S/m}$ 乗がそのずれの大きさ(対数上での値なので、距離の比となる)を示す。もちろん、相関係数と標準誤差のいずれも、特定のMAについて計算して評価することが可能である。

2. 日本の通話地図

(1) データについて

ここでは日本の通話の構造を概説する。利用可能データは1990年と1991年の発信回数・通話時間のMA間通話データで、さらに住宅用契約、事務用契約(電話の所有者は企業となる)に分けて分析する。

10年以上前のデータであるが、この時点の通話トラフィックの構造を把握することは重要である。なぜならば、近年、携帯電話が急速に普及したこと等の理由によって、通信の全体構造を把握できる通話データの入手は困難となったからである。また携帯電話に関するデータが入手できたとしても、それは基地局単位であり、固定式電話と対応させることは非常に難しい。高度成長期に比べれば、近年の人口の増減や産業構造の変化が小さい点も考えると、この時点は通話の全体構造が把握できる最

表1 MAリスト(東北地方の部分)とその人口

MAリスト番号	MA名	1990年人口
93	仙台	1205622
94	石巻	250420
95	古川	223144
96	気仙沼	109553
97	白石	58312
98	岩沼	86660
99	大河原	137831
100	築館	98656
101	迫	78360
102	秋田	394788
103	能代	109635
104	横手	119495
105	大館	90098
106	本荘	127327
107	男鹿	46009
108	湯沢	83443
109	大曲	98487
110	鹿角	50442
111	鷹巣	50406
112	角館	57348
113	山形	371852
114	米沢	180795
115	鶴岡	159103
116	酒田	169260
117	新庄	102214
118	寒河江	94058
119	村山	108541
120	長井	72567
121	福島	405549
122	会津若松	207056
123	郡山	314642
124	いわき	355812
125	白河	110216
126	原町	133211
127	須賀川	92201
128	喜多方	64501
129	二本松	104295
130	田島	22962
131	会津山口	14995
132	柳津	27271
133	石川	94379
134	三春	79394
135	磐城富岡	77574

近年代とも考えられるのではないだろうか。もちろん現状を考える場合には情報化の進展によるデータ通信の増大の事実を忘れてはならない。一方で、IP電話等の普及により、現在の通話トラフィックの全容をつかむことはいよいよ不可能と考えられることも付記する。

日本全国は567のMAに分けられるが、MAの境界は、(各MAに対応して構成した)市町村の集合体の境界と概ね一致している。しかし、完全に一致している訳で

表2 総発信回数・通話時間（年間）

	発信回数(100万回単位)				通話時間(100万時間単位)			
	住宅用	事務用	計	一人あたり	住宅用	事務用	計	一人あたり
1990	26241	46371	72613	587.4 (回/年)	1900	1453	3353	27.1 (時間/年)
1991	27638	45632	73270	592.7 (回/年)	2031	1413	3445	27.9 (時間/年)

表3 一通話当たりの時間（分）

	住宅用	事務用
1990	4.34	1.88
1991	4.41	1.86

表4 内々通話の比率

	90年住宅用	91年住宅用	90年事務用	91年事務用
発信回数	74.8%	74.3%	65.7%	66.5%
通話時間	62.7%	62.4%	60.1%	61.3%

はなく、一部のMAでは数百人程度の誤差がある。またMAの境界と市町村界が著しく異なる場合は、その境界で接するMAを統合した。その結果、本研究での分析MA数は562となった。例として東北地方の部分に属するMAのリスト順の名前とその1990年国勢調査人口を表1に示す。この範囲だけで見てもMAの人口で120万人超の仙台から1万5千人弱の会津山口まで、様々であることがわかる。ちなみに最大のMAは東京23区で823万人、最小は硫黄島（鹿児島県）の500人強である。なお、通話回数データは100回単位（90年のデータの一部は10回単位）・通話時間は10時間単位で1年分の合計が集計されている。

(2) データの傾向

まず、総発信回数・通話時間を見る。日本人口（1990国勢調査：124,611,167人）を用いて、一人あたりも含め算出した結果が表2である。発信回数の63%が事務用電話からだが、通話時間では事務用からは42%にすぎない。このことから、住宅用電話での1通話あたりの通話時間が比較的長いことがわかり、それは表3のように示される。住宅用の電話での一通話あたりの時間は事務用の通話に比べ、2.3倍ほど長い。一方、発信の総量に対するMA内々通話の比率は高く、表4のように、発信回数では住宅用の約4分の3が、それ以外でも6割以上が内々通話によって占められる。

1990年と1991年でデータの集計値には若干の違いが見られるが、2時点では時系列的な変化傾向を捉えるのは困難である。本研究では、この2時点データをデータの信頼性・安定性の評価のために使い、吟味の結果からMA間の通話がいずれかの年で1000回以上、100時間以上ならば有効データとしてとりあげることにした。そして、2時点の平均値をもって代表値とした。各カテゴリでのデータ

数、全体に対する比率を表5に示す。これによってデータの総数は40%未満となったが、すべてのMAでの発信回数の90%MA以上（内々は除いて）は採用したデータに含まれている。ただし、離島のMAでは除外されるデータが比較的多かった。

ここからの議論では1990年国勢調査人口によって通話データとマッチングさせ、一人あたり・人口ペアあたりの通話を分析していく。まず、一人あたりの年間発信回数は図1のように表され、MAによってかなりの相違が現れた。すなわち、住宅用電話では一人あたり年間発信回数が最小値81回（愛知県田原MA）、中央値198回、最大値451回（沖縄県南大東MA）というばらつきがあること、事務用ではさらにばらつきが大きく、最小値101回（奈良県大和榛原MA）、中央値276回、最大値919回（東京MA）という傾向がわかった。表1から事務用電話の方が発信回数が多いことがわかっているが、分布としても事務用電話は裾が広く、MAによつての相違がより大きいことがわかる。通話時間でも同様に図2に示すと、通話回数の場合と同じように、住宅用で最小値3.6時間（愛知県田原MA）、中央値8.3時間、最大値30.6時間（神奈川県相模原MA）、事務用で最小値5.5時間（奈良県大和榛原MA）、中央値12.8時間、最大値30.2時間（東京MA）と様々なMAがあることを知ることができた。表1から通話時間では住宅用電話の方が事務用電話より長いことがわかっているが分布の大まかな傾向はそれほど

表5 採用したデータ数

	事務用	住宅用
通話回数	121489	113377
	38.50%	35.90%
通話時間	94120	120853
	29.80%	38.30%

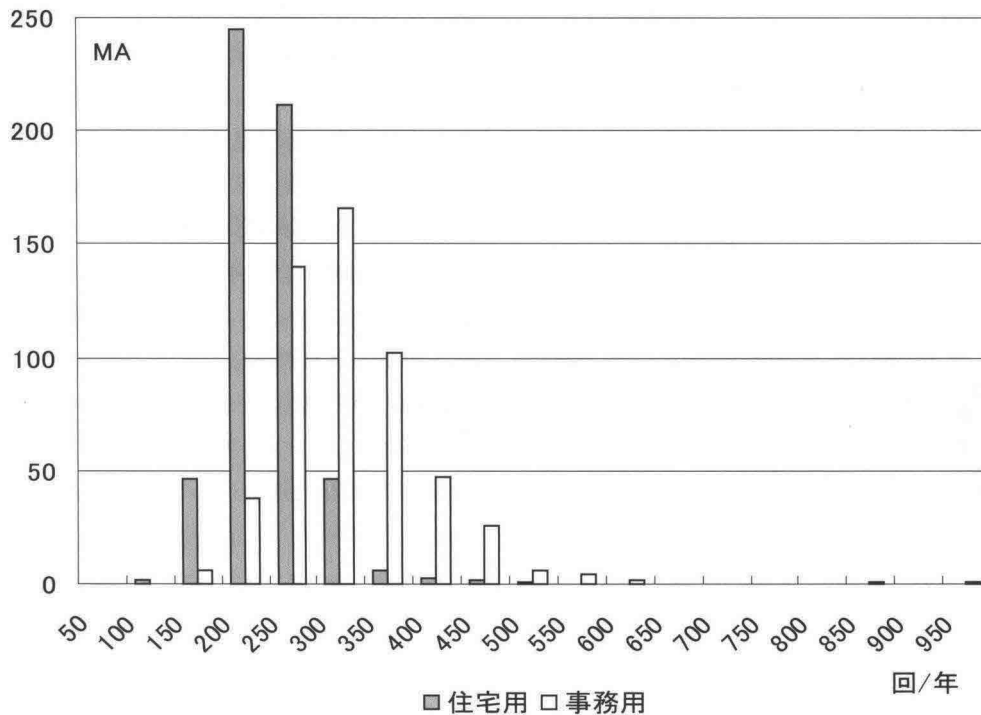


図1 一人あたり年間発信回数 (MA単位での通話回数のヒストグラム)

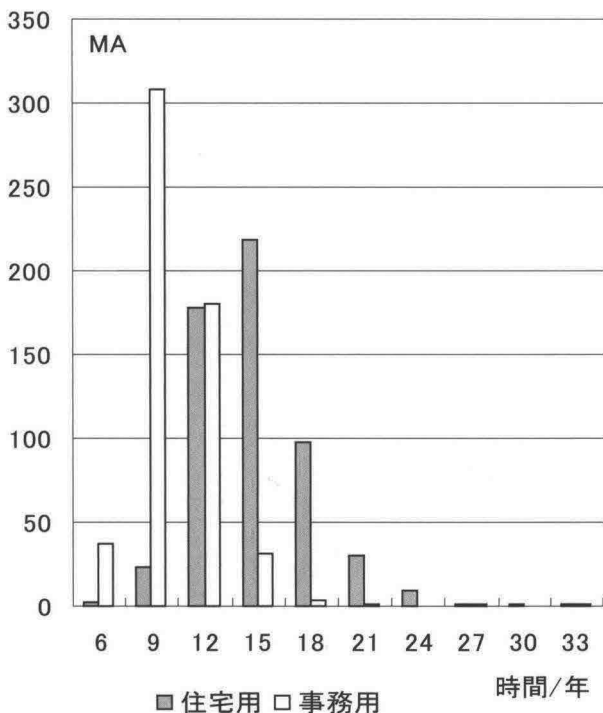


図2 一人あたり年間通話時間 (MA単位での通話時間のヒストグラム)

違いは見られないようである。なお、本稿では示さないが、就業者従業地人口を事務用電話に対応させれば、もう少しばらつきが小さくなるかもしれない。特に東京MAが事務用電話の最大値となる通話ような通話の活発さを示しているのは、就業者従業地人口が多いためと解釈できよう。

MAの位置と住宅用電話で、一人あたり年間発信回数を地図に表せば図3になる。ここでは年間発信回数を[平均] ± [標準偏差] との大小で4つに区分し、発信回数が多いほど円を大きく表した。また円の色の濃度は、内々通話を除いて、発信に対して受信がどのくらいの比率かを4段階に表し、発信に対する受信回数が多いほど色を濃く示した。

図3でMA平均の年間発信回数は200.87回、標準偏差は43.2回であり、[受信/発信]の値は平均1.053 (MAの規模が様々なので1.0にはならない)、標準偏差0.337であった。この図によって、いわゆる太平洋ベルト地帯の一人あたりの発信が多いことと、当該地域の発信に対する受信も多いことがわかる。産業の発達している地域

一人当たり年間通話回数

- :244 ~
- :200~244
- :157~200
- :~157

受信/発信

- :1.39~
- :1.05~1.39
- :0.72~1.05
- :~0.72

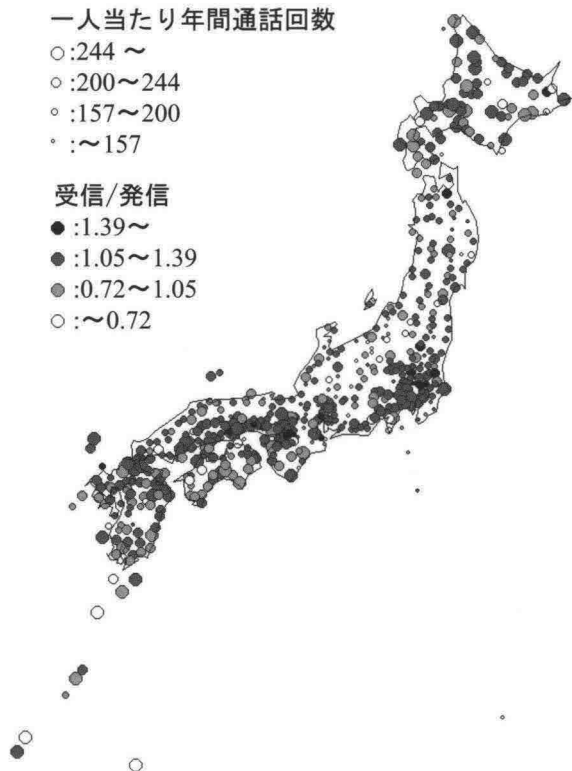


図3 MA別一人当たり発信回数と受信率（住宅用電話）

一人当たり年間通話回数

- :366 ~
- :287~366
- :209~287
- :~209

受信/発信

- :1.28~
- :1.00~1.28
- :0.71~1.00
- :~0.71

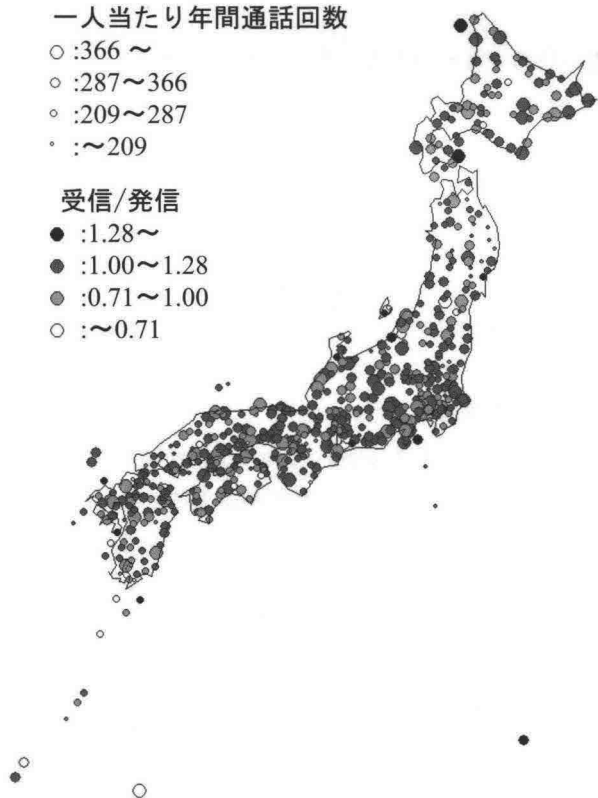


図4 MA別一人当たり発信回数と受信率（事務用電話）

なので、その反映という仮説も立てうるが、同様の傾向は北海道などでも見られる。またMAそのものの密度が太平洋ベルト地帯で高いため、この地域が本当に通話が活発で、かつ受信が発信に対し相対的に多いのか、または単にMAがたくさんあるために線の部分も含めて作画されている範囲が大きいことのように見えるのか、判断は難しい。このような図から視覚的に十分傾向をつかむことは容易ではない。

一方、同様の図を事務用電話について作成したのが図4である。図4ではMA平均の年間発信回数が287.57回、標準偏差は78.8回であり、[受信/発信]の値は平均0.996、標準偏差0.286であった。事務用電話での地域間格差は、図3の住宅用電話に比べると小さめである。一人あたり年間通話回数の少ないMAも大きなMAも遍在しているため、小さいMAのマークが隠されることが多く、図3より地域差が小さい印象を得る図となっている。

(3) 重力モデルに適用した結果

この節では、通話回数データを重力モデルに適用した結果を示す。本分析でのMA数は562であるが、一方「都道府県」という地域単位にも注目する。

各MAには、通話料金を決めるための代表点がNTTによって決められている。それは概ね人口重心に当たる位置にあり、本研究でもその代表点の座標を使ってMA間の距離を与える。なお、隣接MA間の距離など近距離での距離の誤差の問題を緩和するため、各MAの面積に等しい円を想定し、その場合の領域間距離の近似式（栗田・腰塚、1988）によって与えた。

まず、式(2)で与えた重力モデルの適用結果を表6に示す。なお、2.1節で述べた理由でMA内々の通話はここでの計算からはずした。それによって表6でのデータ数は表5よりも562だけ少なくなっている。重力係数 β の

表6 重力モデルの適用結果

		重力 β	切片 α	データ数	決定係数
回数	事務用	1.40	10E -3.39	120927	0.58
	住宅用	1.20	10E -3.95	112815	0.53
時間	事務用	1.31	10E -4.86	93558	0.58
	住宅用	1.13	10E -5.07	120291	0.49

※切片は指数表記。例えば10E-2は10のマイナス2乗を意味する。

絶対値は事務用電話の方が大きく、距離による減衰が大きいことが分かる。個人（住宅用）での遠距離通話の頻度の減衰に比べ、企業（事務用）での遠距離通話の必要はそれほど減らないということなのだろう。決定係数は0.5前後と十分大きいとはいえない。

事務用電話の回数に関して距離と通話回数の関係を示したのが図5である。ここでは、両対数平面で示した。縦軸の単位は式(1)に従い、通話回数を人口の積で割ったものの対数値である（底は10）。データ数が多く、全ての表示は無理なので、乱数によってデータの1%を抽出し表示した（表示データ数=1217）。また、県内通話と県間通話を分けて表示した。同一距離に対する通話回数は県内で明らかに多いことがわかる。同様の図を住宅用電話や通話の時間に関して図6（通話回数・住宅用；表示データ数=1131）、図7（通話時間・事務用；表示データ数=944）、図8（通話時間・住宅用；表示データ数=1213）に示す。通話時間で住宅用電話での傾向では多少データが広く散布する傾向が見られるが、どの場合でも概ね重力モデルに従う傾向にあることは理解できる。

このような、べき型の重力モデルを当てはめた結果の考察は、これまでもいくつか示されている（Mackay, 1958; Rietvelt and Janssen, 1990; Noronha et.al, 1992）。そこでは、言語や国境という障壁によって通話

の頻度が大きく低下することが示されている。一方、日本国内の通話データについては都道府県単位のデータによって分析がなされていたが、MA単位ではあまりされていなかった。言語や国境に比べ、移動の障壁が小さい都道府県という単位でも、地域のグループが明確に見られることは、一つの知見といえよう。なお、以降の分析で述べるが、一つの地域グループとはならない都道府県も存在する。それをどのように調べていくかは次節の主題の一つである。

ところで、通話時間を通話回数で割ることによって、一通話あたりの平均時間を得ることができる。図9は通話時間・通話回数の両方のデータが有効なMAのペアについて、一通話あたりの平均時間と距離の関係を調べた結果である。ここから遠距離間の通話ほど平均通話時間が長くなること、住宅用の電話の方が通話時間が長いことがわかる。なお、線でつないだ点列と離れた距離が短い位置の点はMA内々通話を示す。図9の傾向は直接会えないための通話による代替効果の現れと考えられる。著者らは通話の目的は大きく三つに分類できると考えている。それは、直接会うことの「代替」、会った後の「補完」、日常会うことによる「相乗」である。「代替」が「補完」「相乗」に比べ長い時間を要することは明らかであり、それが図9の結果に表れていると考えられよう。

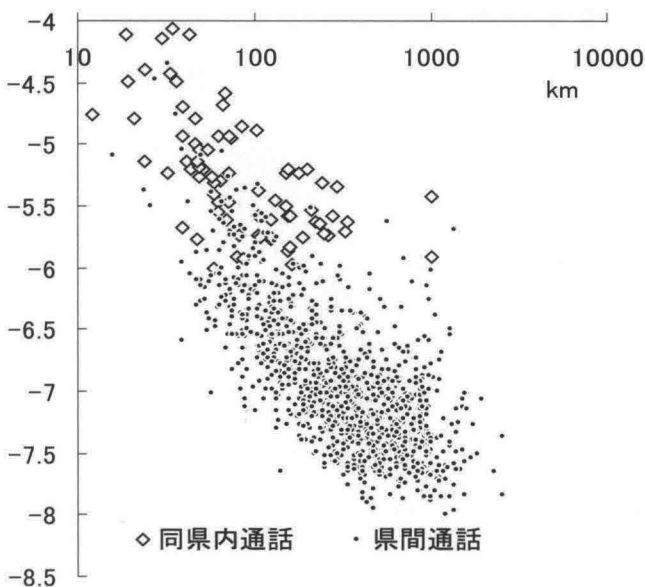


図5. 通話回数と距離の関係（事務用電話）

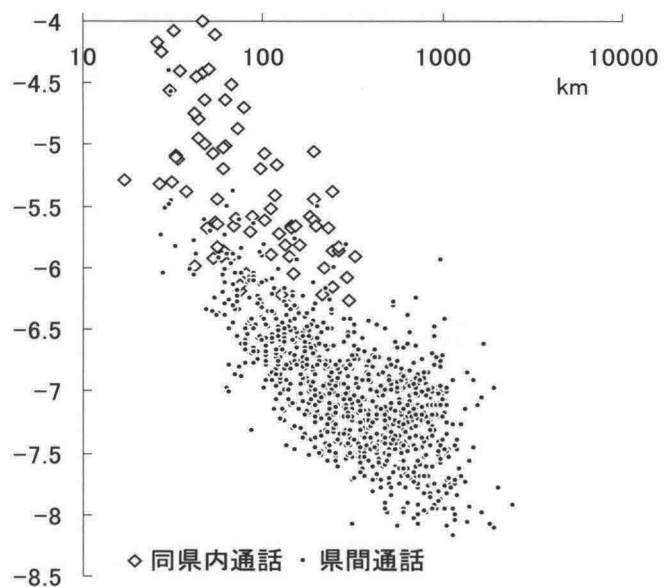


図6. 通話回数と距離の関係（住宅用電話）

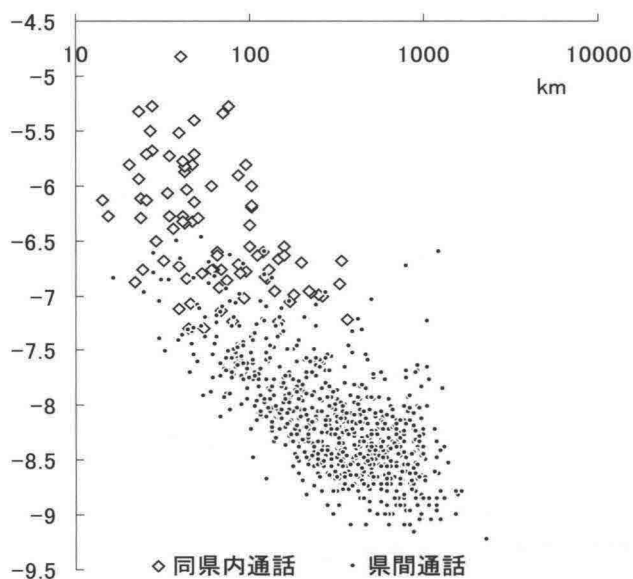


図7 通話時間と距離の関係（事務用電話）

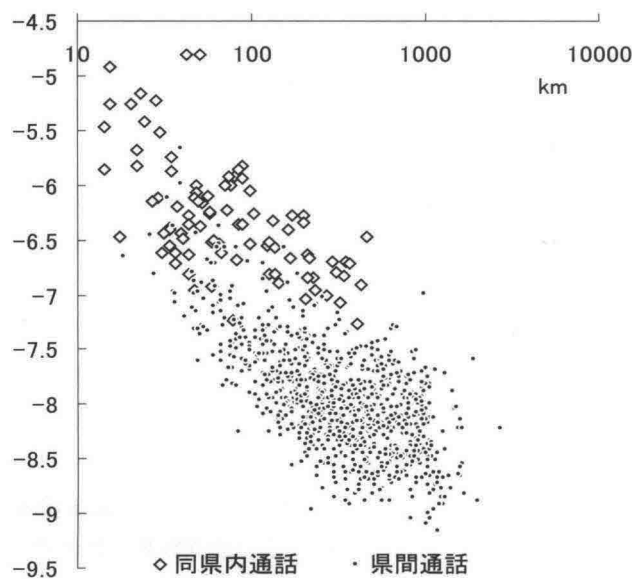


図8 通話時間と距離の関係（住宅用電話）

より直接的に通話の傾向を見ることを試みよう。例として住宅用電話の通話時間を521のMA間のマトリックス形式で表示したのが図10である。行が発信MAを、部分的に凡例をつけた列が受信MAを示す。ここでは、最も基準通話時間が長いペアを水色〔主に対角線上〕で表し（10E-5）、基準通話時間が短くなるにつれて徐々に白（10E-7）に色を変えている。更に基準通話時間が短くなるにつれて赤（10E-8）、暗赤（10E-10）となっている〔対角線上以外は明るいほど通話が活発であることを示している〕。これらの色は値に対応して連続的に変化させてある。また黒は通話時間が短いためデータを採用しなかったMAペアを示す。

対角線上はMA内々の通話を示すので、必然的に通話時間は長く、水色となっている。対角線上に沿って地域名を説明すれば、MAの番号は都道府県コードに従うので、左上から北海道、青森県、岩手県、……と続き、一番右下は沖縄県となる。例えば一番のMAは札幌MAなので、一番上の行は、札幌からの通話時間、大まかに、右へ行くほど西方の地域への通話を示す。対角線上に様々な大きさの正方形の連なりを感じることができるが、多くはこれは都道府県としての通話のまとまりの反映である。さらに、北海道の支庁でのまとまりや、東北、四国、九州といったより広い地域としてのまとまりも見ることができる。また南関東地方の全国への通話が活発

な様子も見ることができる。

細かく見ていくと、例えば、愛知県は静岡県より岐阜県とのつながりが強いことや、四国で全国との通話が比較的活発な線があり、それは高松MAであることがわかる（90年での通信販売の販売額第一位の業者は高松に立地している影響と推測される）。

このように、通話マトリックスはすべてのデータを示

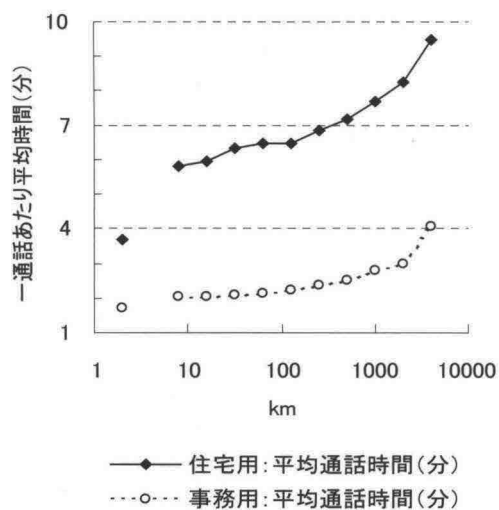


図9 一通話あたりの平均時間

しているので、詳細に調べればたくさんの知見を得ることができ、分析に手間がかかることや、より大きな日本全体の構造を知ることは困難なので、活用には問題も大きい。

(4) 多次元尺度構成法による視覚化

考察をさらに進め、逆算距離の活用によって、県のまとまり具合、その空間的位置関係、さらにはより大きな地方というくくりの影響がどのように通話に見ることができるかを吟味する。

通話地図は式(7)、最小二乗法多次元尺度構成法によって作成した。なお、重力モデルの当てはめと同様に内々

データをはずして視覚化の計算もおこなった。計算は近似計算であり、準ニュートン法を用いて通話地図での座標を求めた。このとき、目的関数S (式(7)) の座標位置による偏微分値が必要となるが、例えば X_{c_h} での偏微分の数式は、

$$Z_h = \log \sqrt{(X_{c_h} - X_{r_h})^2 + (Y_{c_h} - Y_{r_h})^2} + \log(y_h / \alpha) / \beta$$

とにおいて、

$$\frac{\partial S}{\partial X_{c_h}} = -\frac{2}{\ln 10} \sum_{h=1}^m Z_h \frac{X_{c_h} - X_{r_h}}{(X_{c_h} - X_{r_h})^2 + (Y_{c_h} - Y_{r_h})^2}$$

(8)

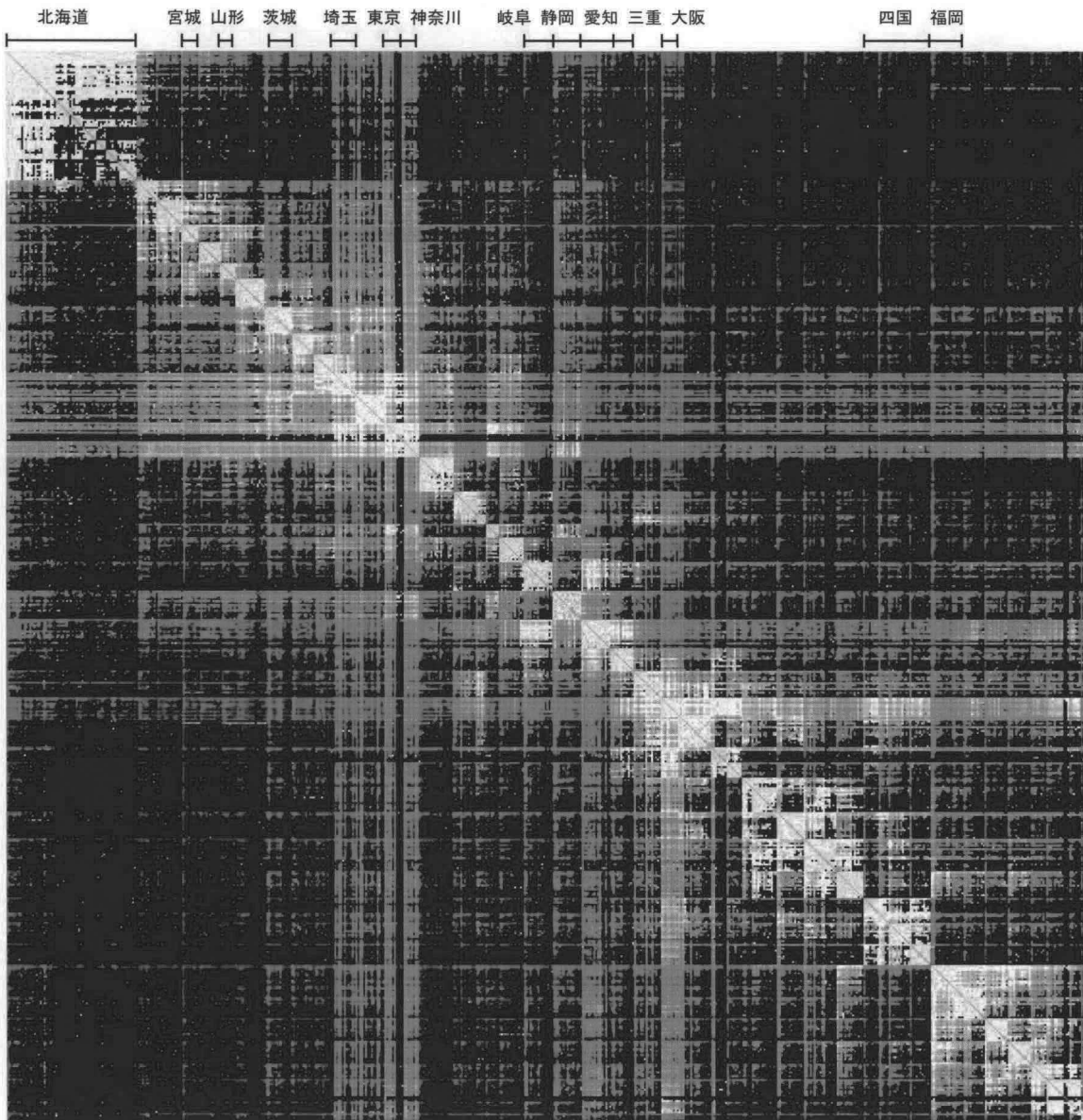


図10 通話時間マトリックス (住宅用電話)

となる。なお、 X_m, Y_{ch}, Y_m どれについても偏微分したときの式の形は(8)式と対称になる。

準ニュートン法による計算で、最適性は保証されないため、初期値のセットとして、日本の物理的な位置と9種類の乱数で与えたものの合計10種類を準備し、それぞれの初期値からの計算の結果で、目的関数値が最小となったものを採用した。なお、係数 α, β は表6の値を用いた。

比較の便宜のため図11に物理的な地図を載せた。凡例・縮尺は図11～図15まで共通となるが、各都道府県は県庁所在地を中心とした星形グラフで示した。色は地方の区別である。通話地図(図12～図13)での色は図

11を凡例とされたい。また、記号が円で囲まれているのは政令指定都市を含むMAである。さらに、図12～図15に関して吟味に取り上げるMAには矢印でその名前を示した。

まず、事務用電話での通話回数の通話地図が図12である。この地図上でのMA間距離と通話距離の相関係数は0.866、標準誤差は0.242(距離の比に直せば1.748倍)であった。地図上では、他のMAとの残差が平均で0.301以上のMAを白抜きマークで示した。これは $10^{0.301}$ がほぼ2なので、他のMAとの位置関係の誤差(距離の比)が相乗平均で2倍以上あるMAであることを意味する。

結果として日本は逆S字型に変形された。それは、北

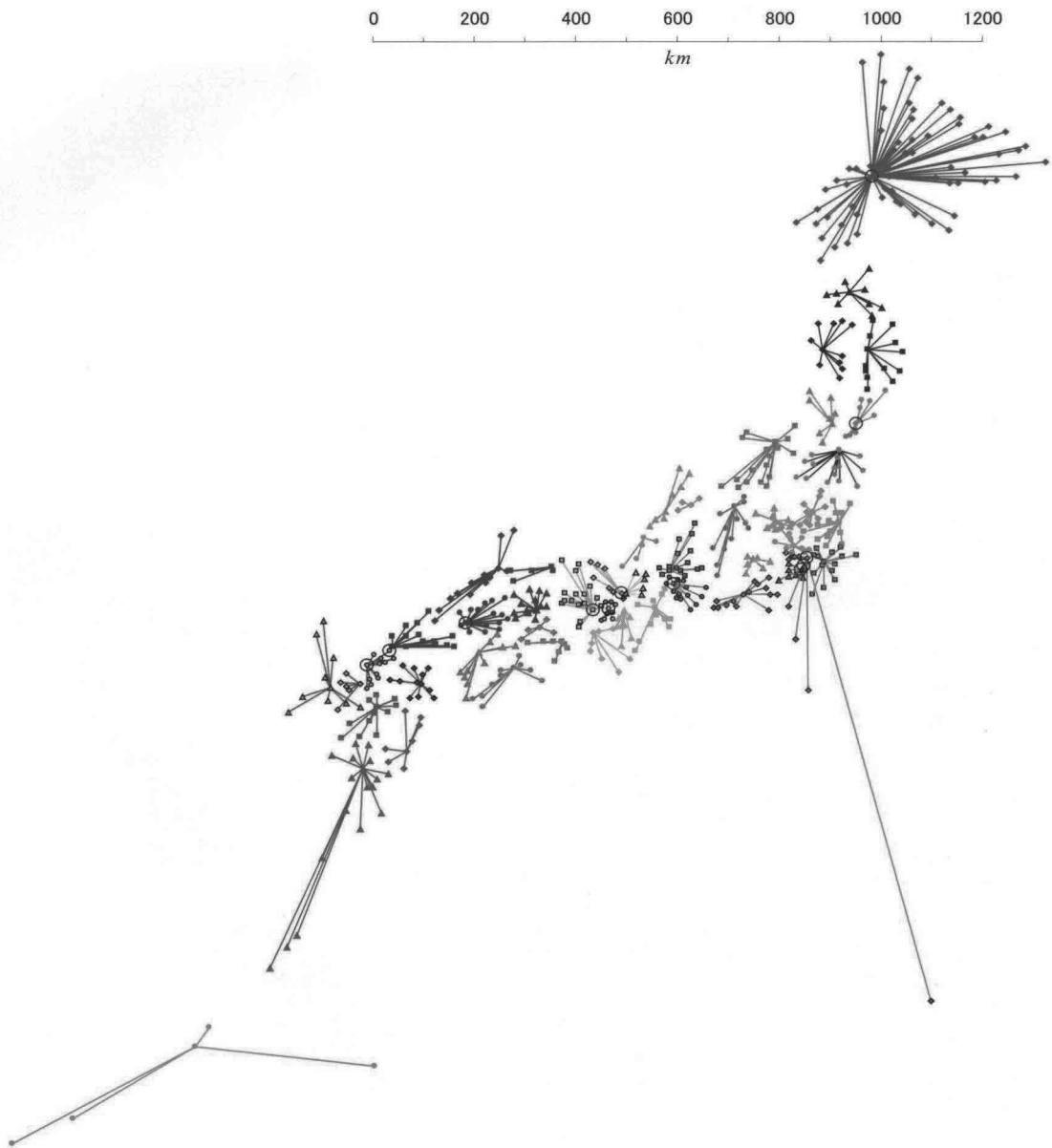


図11 物理的な地図

海道や九州・沖縄が多少、日本中央部に近く位置づけられたと言える。これは長距離通話は料金が距離と無関係であるため、相対的に通話の回数が大きくなっているためかもしれない。なお、これらの通話地図は回転や鏡像変換をしても等価なものである。

通話地図に共通した特徴は、多くの各都道府県が真の位置関係よりもはるかに集まって配置されたことであ

る。これは、重力モデルで見たときに都道府県内の通話が対数重力モデルでの推定値よりもはるかに通話が活発であること（図5～図8）の反映といえる。

図12から個別に読みとれることを列記しよう。

- (i)多くの都道府県では属するMAが狭い範囲にまとまった。
- (ii)いくつかの政令指定都市は、その属する県の中央よ

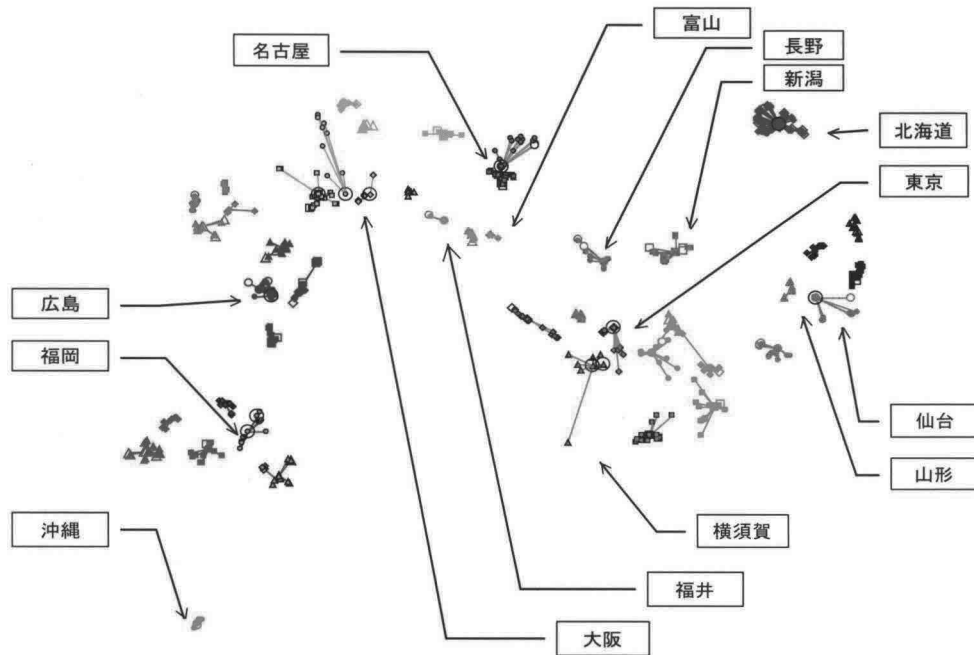


図12 事務用電話、発信回数に関する通話地図（縮尺は図11に同じ）

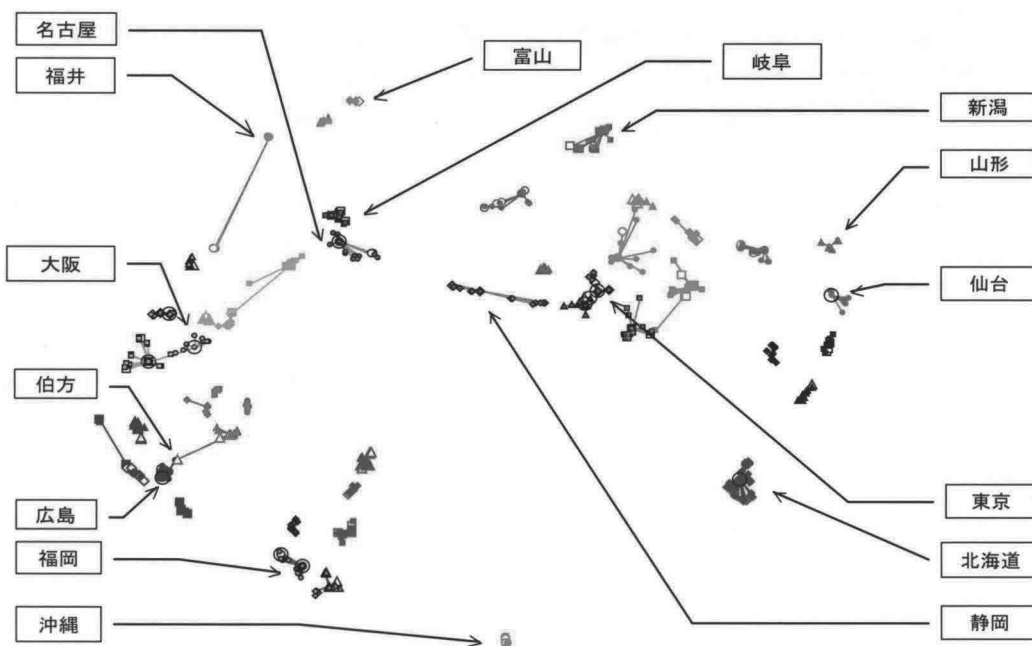


図13 住宅用電話、発信回数に関する通話地図（縮尺は図11に同じ）

りも、属する地域の中央に位置する傾向となる (ex. 仙台MA→東北地方の中央、広島MA→中国地方の中央、福岡MA→九州の中央)。

(iii) 東京MAの位置は関東の中央というよりは、むしろ東日本の中央に近くなった。

(iv) 大阪府のMAは大阪市のみ大阪府の他のMAとは離れ、また近畿地方の端に位置した。概ね近畿・中国・

四国の3地方の中央に位置した。

(v) 物理的な位置関係では接しているにもかかわらず、非常に離れて位置した県がある (ex.新潟と山形、長野と岐阜)。物理的な交通の便が悪いので、その反映と考えられる。

(vi) 横須賀MAは神奈川県以外のMAと位置が離れた。横須賀には大きな米軍基地があるので米軍基地間 (横

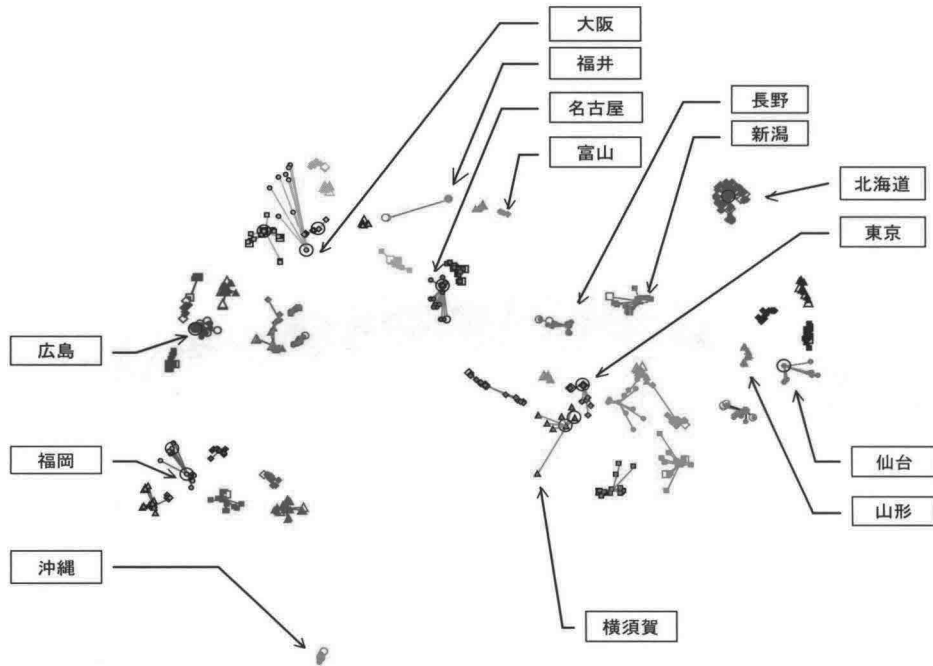


図14 事務用電話、通話時間に関する通話地図 (縮尺は図11に同じ)

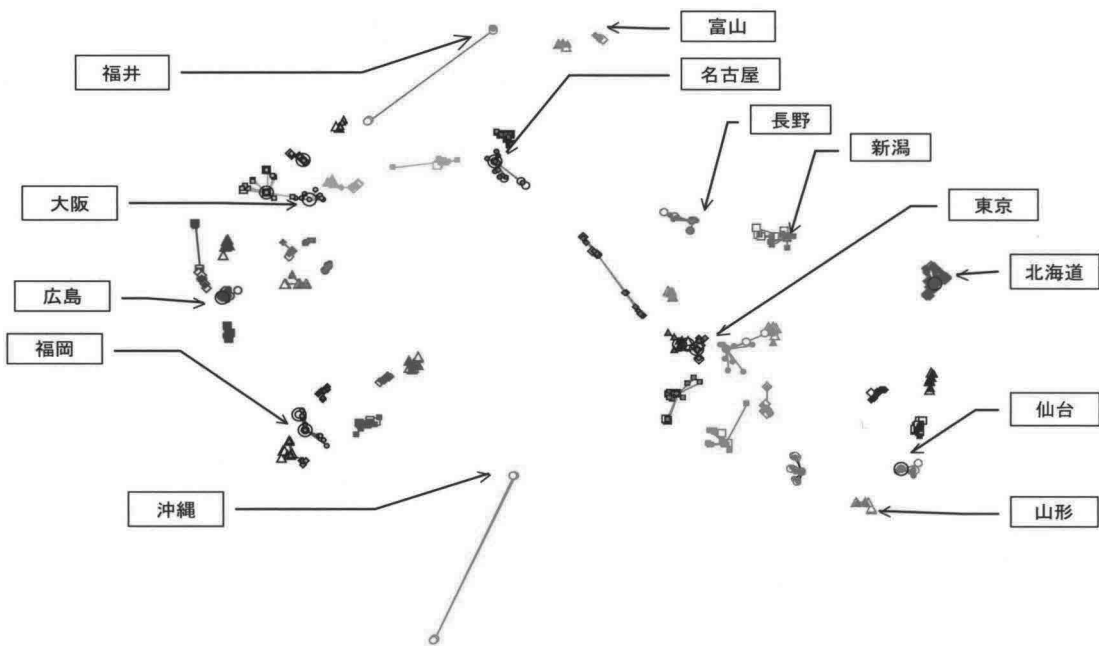


図15 住宅用電話、通話時間に関する通話地図 (縮尺は図11に同じ)

須賀一沖縄間)の通話が多いことを意味しているのかもしれない。または同じ県であっても、他の神奈川県内とは通話が少ないことの反映らしい。

図12に海岸線を描いてみると中部地方、新潟県―静岡県と富山県―愛知県の間で交差がおきる。すなわちこの地図は日本の中央でねじれている。この通話地図を元に、ねじれた部分を裏返してそれを初期値に計算しても図12に比べ目的関数値が劣る結果だった。これは、新潟県と富山県、愛知県と静岡県のつながりが、ほかの隣県との関係に比べると薄いこと、西日本側で四国や南九州地域よりも中国地方や北九州地域の方が関東地方との関係が強いことの反映と読み取られた。このようなねじれは他の通話地図では見られなかったもので、例外的なものかもしれない。

住宅用電話の回数についての通話地図が図13である。住宅用では英文字のCのような形になった。この地図上でのMA間距離と通話距離の相関係数は0.867、標準誤差は0.251(距離の比に直せば1.781倍)で図12の場合とほぼ同じである。ここで見られる傾向を列記する

- (i)図12同様、多くの都道府県では属するMAが狭い範囲にまとまった。
- (ii)政令指定都市は、その属する県の端に位置した。住宅用電話の場合、地方中心の位置としての効果はそれほど現れない。
- (iii)東京MAは関東の中央(寄り)に位置した。
- (iv)大阪市MAも近畿地方の中央寄りに位置した。
- (v)新潟と山形、長野と岐阜など、物理的な位置関係では接しているにもかかわらず、非常に離れて位置した県は、事務用電話での場合より、より極端に位置が離れた。
- (vi)静岡県、福井県などは県内MAが集まらず、大きく散開して位置した。例えば静岡県は東京方面と中京地域方面の綱引きの結果のように見えるし、福井県では越前地方と若狭地方(但し、誤差は大きい)の相違と見える。また愛媛県の伯方MAは広島県とのつながりが強いらしい。

図14は事務用電話の通話時間についての地図である。傾向は事務用電話の発信回数の通話地図と似ていて、英文字のSのような形になった。但し、図12と異なりねじ

れはおきず、ねじった配置を初期値にして再計算しても、図14の結果以上の目的関数値とはならない。この地図上でのMA間距離と通話距離の相関係数は0.867、標準誤差は0.242(距離の比に直せば1.745倍)である。図12との比較に視点を置いてここで見られる傾向を列記する

- (i)図12同様、多くの都道府県では属するMAが狭い範囲にまとまり、政令指定都市の多くは、その属する県の中央よりも、属する地域の中央に位置する傾向となった。
- (ii)図12との相違として、中国地方と四国地方の位置関係、合わせて九州地方の南北の向きが逆になった。ただし、四国と中国地方の位置関係を逆にしても大きくは相関係数等の値は大きくは変化しなかった。

最後に、住宅用電話の通話時間についての地図が図15である。個々の傾向は住宅用電話の発信回数の通話地図と似たが、全体の形は英文字のSのような形となり、住宅用の場合とは異なる形となった。この地図上でのMA間距離と通話距離の相関係数は0.867、標準誤差は0.275(距離の比に直せば1.882倍)である。図13同様、多くの都道府県では属するMAが狭い範囲にまとまり、また政令指定都市の多くは、その属する県の端に位置した。住宅用電話の共通性らしいが、ここでも地方中心の位置としての効果はそれほど現れなかった。

住宅用の電話の全体としての傾向は、事務用に比べ、より狭い範囲での通話が活発と読み取れる。ビジネスの広がり個人生活の範囲の違いと理解できるだろう。

図13のデータ(発信回数&住宅用)の通話地図では、目的関数値が0.29%だけ劣る結果で、逆S字型の結果も得られた。また図12のデータ(発信回数&事務用)の場合でも目的関数値が0.27%劣る結果でC字型が得られた。図14・図15でも同様である。通話地図を求める計算は数十種類の初期値・データに対して行ったが、様々なカテゴリ(回数、時間、事務用、住宅用、合算など)での最適な結果と比べ、目的関数値が1%未満劣る程度の結果は全て(逆)S字型か、C字型となった。元々日本国土の形態はわずかに逆S字型であるが、通話の構造では、それがより強くS字型になる、またはC字型になるとわかった。なお、C字型になる視覚化結果は(都道府県単位の)人口移動に関する研究で矢野(1991)でも挙げ

られている。これらは、中央から離れた地域が物理的な距離よりも中央に近く位置付くことを意味する。それは実際に会いに行くことを、代わりに電話で済ますことの効果、すなわち電話によるコミュニケーションの“代替効果”の反映という可能性を考えているが、人口移動でも同じような結果が出ているので、より精細な研究の結果を待つ必要がある。

一方、県としてのまとまりが強くては、日頃のつきあいが、その確認や予約などで電話をよくすることになる結果、すなわち電話によるコミュニケーションの“相乗効果”、“補完効果”と呼ぶことができると考える。まとめれば、通話地図で見る日本国の形は、S字型かC字型かということはあまり重要ではなく、「遠くがより近くに位置し、各都道府県レベルのつながりは一般的には強い」ということが全体的な傾向と結論づけられる。

主として、事業用の電話で政令指定都市など大都市MAが属する地方の中心に位置したのは、大都市の通信活動に関するある種の強さを示している。それはモデルに規模の効果の係数；例えば人口に対して γ 乗するような係数を与えれば、それが1未満であることを意味している。実際、本研究の重力モデルに係数を追加し、一律に人口に γ 乗してデータを説明することを考えた場合、住宅用の発信回数の場合で $\gamma = 0.74$ を得ることができる。

3. まとめ

以上によって逆算距離の応用による、通話の仮想地図を作成する方法、その結果の吟味についての議論をおこなう。通話は単純な距離との関係では十分説明できない。本研究で方法を示した仮想地図は、その複雑な関係を吟味するのに有効な方法と考える。もちろん、通話地図は十分精度が高いとは言えず相当の誤差を持つてはいるが、各地域の特質を精査しながら通話の全体構造を考察する道具としては効果的ではないかと考えている。

本研究による日本の通話構造の分析から、都道府県を中心とする地域グループの存在や大都市の通話に関する強い吸引力の傾向（特に事務用電話で顕在）などが明確に視覚化された。

一方、本研究では古典的なモデルで分析説明したが、

制約形のモデルや、規模の効果の係数の導入など、様々な可能性が考えられるが、それは今後の課題としたい。

参考文献

- 石川義孝(1988)『空間相互作用モデル—その系譜と体系—』、知人書房。
- 石川義孝(1994)『人口移動の計量地理学』、古今書院。
- 伊藤 悟(1997)：時間地図のデイリーリズム—都市の時空間分析にかかわる1つの試み—。地理情報システム学会講演論文集, 6, 275-280
- 茨木俊秀、福島雅夫(1991)『Fortran77最適化プログラミング』、岩波書店。
- 川手弘明、大山達雄(1994)：区域間通話量流動現象の空間相互作用モデル分析。Journal of the Operations Research Society of Japan. 37(2), 114-131。
- 栗田 治・腰塚武志(1988)：領域間平均距離の近似理論とその応用、都市計画論文集、No23, pp.43-48。
- 古藤 浩、長谷川文雄(1998)：重力モデルを利用した地域間交流の表現—通話データを例にとって。日本地理情報システム学会講演論文集、253-258。
- 古藤 浩、長谷川文雄(2002)：通話トラフィックデータによる地域構造の分析—重力モデルによる逆算距離を活用した視覚化を通して—。都市計画論文集、37, 61-66。
- 清水英範(1992)：時間地図の作成手法と応用可能性。土木計画学研究論文集、10, 15-29。
- 高嶋裕一(1994)：MA間電話トラフィックデータを用いた通話圏の識別、InfoComREVIEW、1, 77-89。
- 高嶋裕一(1997)：都市間通話需要モデルにおける距離減衰パラメータの分析、InfoComREVIEW、4, 80-86。
- 中村雄一、長谷川文雄(1992)：単位料金区域から見た通話トラフィックの基本特性に関する研究、平成3年度情報通信学会年報、106-119。
- 長谷川文雄、中村雄一、出石宏彦(1985)：ODデータを基にした圏域同定に関する一手法、計測自動制御学会論文集、21(5), 469-474。
- 山崎 健他(1993)：日本における通話構造の定量的分析—通話トラフィックデータに依拠して—。地域学研究（日本地域学会）、23(1), 109-130。
- 矢野桂司(1991)：空間相互作用モデルの精緻化に関する研究。都市計画論文集、26, 517-522。
- Haynes,R.M.(1975): Dimensional analysis: Some applications in human geography. Geographical Analysis 7,51-68.
- Katz,J.E. and P.Aspden(1998): Theories, data, and potential impacts of mobile communications a longitudinal analysis of US National Surveys. Technological Forecasting and Social Change. 57(1-2),133-156.
- Mackay,J.R.(1958): The interactance hypothesis and boundaries in Canada: A preliminary study. Canadian Geographer 11, 1-8.
- Murnion,S. and R.G.Healey(1998): Modeling distance decay effects in Web server information flows. Geographical

Analysis 30(4),285-303.

Noronha,V.T. and M.F.Goodchild(1992): Modeling Interregional Interaction - Implications for Defining Functional Regions. Annals of the Association of American Geographers, 82,86-102.

Nijkamp, Peter, Piet Rietveld, and Ilan Salomon(1990): Barriers in Spatial Interactions and Communications: A Conceptual Exploration. Annals of Regional Science, 24,237-252.

Plane,D.A.(1984): Migration Space: Doubly constrained gravity model mapping of relative interstate separation. Annals of the Association of American Geographers, 74,244-256.

Shen,Q.(2000): An approach to representing the spatial structure of the information society. Urban Geography, 21(6), 543-560.

Rietveld, Piet and Leon Janssen(1990): Telephone Calls and Communication Barriers: The Case of the Netherlands. Annals of Regional Science, 24,307-18.

Yano, K(1993): Integration of spatial interaction models: towards a general theory of spatial interaction. Geographical Reports of Tokyo Metropolitan University, Department of Geography, Tokyo Metropolitan University, 28,33-78.

執筆者

古藤 浩 デザイン工学部 情報デザイン学科
KOTOH Hiroshi Faculty of Design / Department of Informatique Design
助教授
Associate Professor

長谷川文雄 デザイン工学部 情報デザイン学科
HASEGAWA Fumio Faculty of Design / Department of Informatique Design
教授
Professor