

卒業論文 2019 年度 (令和元年)

マハラノビス距離を用いた
エリア推定手法の提案

慶應義塾大学 環境情報学部
山田 航太郎

マハラノビス距離を用いた エリア推定手法の提案

現在、欧米で普及が進んでいた Nomadic Device や PND(Portable Navigation Device) などの可搬型の端末によるナビゲーションは、スマートフォンやタブレットといった IT モバイル機器との融合が進み、利用者の位置と移動を扱う情報環境は急速に発展が進んでいる。GPS に代表される GNSS は、自動車のナビゲーションシステムによる普及を経て、現在はモバイル機器の代表的な位置測位システムとなっている。こうした位置測位システムの普及展開を経て、google maps や Yahoo!乗り換え案内などの、利用者のスムーズな移動をサポートするサービスがインターネット上で展開されている。

位置情報は大きく分けて、GNSS などから得られる緯度・経度・高度として示される座標情報と、ビーコンや WiFi 基地局などから得られる電波情報と専用のデータベースを照合して得られる所在情報がある。現在のモバイル機器では座標情報を用いて現在地を把握し、それを元にナビゲーションなどを行う。しかし、入り組んだ施設や細街路などでは頻繁に位置がずれて利用者が混乱するなどの問題がある。一方、所在情報についてはコンテキストに基づいたサービスを構築することが可能である一方、対象とする空間に位置測位基盤を敷設する必要があり、導入コストが高い問題がある。

そこで本研究では、モバイル機器に広く普及した GNSS による座標情報から、マハラノビス距離という統計学上の距離を利用することによって所在情報を推定する手法を提案する。都市部の駅や大型施設など、空間へのアクセスに限りがあり、かつ内部に空間的な広がりがある環境においては、座標情報による点よりも所在情報を活用することで、利用者のデマンドに即した情報提供が可能となる。同時に、施設の管理者にとっては空間内に存在するデバイスの機器から利用者の人数や滞留状況などが推定可能となり、施設の利用状況の把握や運用性の向上が可能となる。

提案の有効性を示すため、本研究では、大学キャンパスの最寄駅から都心部にかけての鉄道、およびその駅構内をターゲットとした所在情報提供システムを構築し、実際の鉄道路線で取得した実座標情報から、所在情報への変換に際しての精度や品質、処理コストなどを検証し、本研究が提案するアルゴリズムの妥当性を評価した。提案手法を活用することで、博物館のような通路と展示の間に明確に区切りがないところであっても、所在情報を活用した情報サービスが可能となる。また、駅や高速道路などの限られたアクセス経路でしか入れないようなエリアについて、カーナビゲーションや乗り換え案内などの精度や品質の向上が期待される。

キーワード

1. 位置情報, 2. 鉄道, 3. 博物館, 4. ナビゲーション

慶應義塾大学 環境情報学部
山田 航太郎

目次

第1章	序論	1
1.1	本研究の背景	1
1.2	本研究の目的	2
1.3	本論文の構成	3
第2章	関連する技術や先行研究	4
第3章	提案手法	5
3.1	本提案手法の位置付け	5
3.2	用語	5
3.3	目的と仮説	5
3.3.1	本提案手法の目的	5
3.3.2	本研究の仮説	6
3.4	本提案手法の設計	6
3.4.1	マハラノビス距離	6
3.4.2	路線の推定	7
3.4.3	駅・駅間の推定	7
3.4.4	屋内空間の推定	7
3.5	フロー図	8
第4章	サービスの実装	10
4.1	設計	10
4.1.1	路線について	10
4.1.2	路線と駅	10
4.1.3	ユーザーについて	11
4.2	路線・駅の推定	12
4.2.1	路線の推定手法	12
4.2.2	駅の推定手法	13
4.3	実装環境	14
第5章	評価	15
5.1	仮定	15
5.1.1	実験環境について	15
5.1.2	条件設定	15

5.2	評価実験	15
5.2.1	評価実験 2 : 駐車区画が保有する駐車枠数の推定	15
第 6 章	結論	16
6.1	本研究のまとめ	16
6.2	将来的な展望	16

図 目 次

1.1	ヤフー乗り換えの現在地指定 ¹	1
1.2	日本科学未来館に設置されている Bluetooth ビーコン ²	2
3.1	マハラノビス距離のイメージ	7
3.2	本提案手法のフローチャート	8
4.1	データ構造	12
4.2	路線推定範囲 ³	13

表 目 次

4.1	路線を表す <i>rosen</i> クラス	10
4.2	駅を表す <i>eki</i> クラス	11
4.3	リクエストを送る人物を表す <i>User</i> クラス	11
4.4	シミュレータの実行環境	14

第1章 序論

1.1 本研究の背景

複雑な鉄道網を持つ日本では最適な経路の選択が難しいため、多くの乗り換え案内サービスが存在する。このようなサービスでは現在地を利用して駅名を入力せずに経路検索を行うことができる。これにより不慣れな都市で最寄りの駅が分からなからな場合や急いでる場合でも目的の駅さえ入力すれば検索が可能になる。



図 1.1: ヤフー乗り換えの現在地指定¹

しかし、現状ホームや線路上という判定はできていない。そのため、駅まで歩く経路を加味した上で経路検索が実行される。このためホーム上で目の前の電車に乗車するかの判断には使いづらい。また今乗っている電車の特定も不可能なため乗車している電車の目的地への到着時刻の検索といった用途には使用できない。

¹ヤフー乗り換えのスクリーンショット [?]



図 1.2: 日本科学未来館に設置されている Bluetooth ビーコン²

また、博物館や大型の商業施設において利用客への案内や人流可視化などのニーズが存在する。これを実現するために Bluetooth ビーコンや wifi などの電波情報より独自のデータベースを参照し所在情報を提供するサービスが存在する。これによって日本科学未来館では音声ガイドのサービスを提供している。

しかし、設置・維持管理するコストが設置者にかかるほか独自のデータベースが必要となりコストが高いことが問題に挙げられる。また全世界に展開するとなると膨大なサイズのデータ量になるためローカル向きの技術であると考えられる。

1.2 本研究の目的

既存の位置情報の考え方は現在地というものは緯度・経度・高度で表されているものであった。これは空や海や草原など壁がなく自由に移動が行える場面では有効であった。しかし都市では建物があり道があり空間が隔てられている。さらに電車やバスなどの決まったルールに従って移動を行う乗り物も存在する。このような空間の場合各空間において入り口と出口のような移動のルールが存在しているため所在情報を推定することがナビゲーションの精度に直結する。

本研究ではマハラノビス距離という統計学上の距離を利用することによって既存の座標情報から所在情報を推定する手法を提案する。これにより都市で使いやすい形の位置情報の提供ができよりよいナビゲーションや新しいサービスが提供できるようになる。

1.3 本論文の構成

本論文の構成を以下に示す．第??章では本研究に関連する技術や先行研究を述べる．第3章では本提案手法の目的と仮説，具体的なアプローチに関して述べる．第4章には仮説を評価するためのシステムの実装について記述する．第5章では評価実験の概要と結果，考察について述べる．第6章では本研究のまとめを記述する．

第2章 関連する技術や先行研究

第3章 提案手法

3.1 本提案手法の位置付け

本章では、緯度・経度・高度情報から所在情報を推定するために地図上の区画を検証可能な数理モデル化する手法について述べる。

本研究では、スマートフォンを利用し既存の GNSS などの情報を利用し座標情報を収集する事を想定している。これにより特別なデバイスや新たな部品を使用せずにスマートフォン用アプリなどで利用が可能である。また、施設者側もあらたな設備などの新たなコストがかからない利点がある。

また本提案では事前に計算を行うことが可能であり、これにより非常に少ない計算コストとデータ量によってシステムが構築できるという利点がある。

3.2 用語

本論文内で用いる用語について、以下に定義する。

- 座標情報: x, y, z 軸の直交座標上の情報を示す。位置情報の場合は緯度・経度・高さ。
- 所在情報: 施設名や展示名といったような空間名の情報。
- 路線: 鉄道が通る線路に名称をつけたもの。今回の場合国土数値情報 鉄道データ [1] の名称に準拠
- 駅: 列車が停止し人が乗降をする場所。駅ビルや駅コンコースは含まない。
- 展示エリア: 博物館などで利用客が展示を見る空間のこと。

3.3 目的と仮説

3.3.1 本提案手法の目的

本提案手法では既存の所在情報提供手法の諸問題に以下のように取り組む。

新たな装置・設備・組織が不要

Bluetooth ビーコンや wifi の電波情報を利用した所在提供の場合施設側に設置・維持管理コストがかかる。また電波情報と所在情報を紐付けるデータベースが必要となり、全世界で展開するとなると膨大なデータベースになるほか識別子の割り当てが必要である。このためには IP アドレスでいう ICANN のような組織が必要になる。しかし本提案手法の場合既存の GNSS を利用した座標情報を利用するため、新たな装置・設備・組織が不要である。

少ないデータ量・計算量によって構築可能

各区画に最小で 3 点のデータによって構築できる。また大量のデータを区画に指定した場合でも必要とするデータは区画内の点の平均、標準偏差、2*2 の行列でありこれを事前に計算することによってデータ量・計算量ともに小さくすることが可能である。

文脈に沿った所在情報の提供が可能

3.3.2 本研究の仮説

本研究では、座標情報を所在情報に変換する手法を提案する。マハラノビス距離という統計上の距離を利用することによって少ない計算量でどこの区画に属しているかを推定できると予想している。また少ないデータ量でも適切に区画に線引きできると考えられる。

3.4 本提案手法の設計

3.4.1 マハラノビス距離

マハラノビス距離の定義と用途

マハラノビス距離とはプラサンタ・チャンドラ・マハラノビス氏が考案した距離であり、データ群の相関や分散を考慮した上でどのくらいデータがデータ群の平均から離れているかを示す。マハラノビス距離は以下のような数式で定義されている。

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (3.1)$$

$$\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T = \text{平均ベクトル} \quad (3.2)$$

$$\Sigma = \text{共分散行列} \quad (3.3)$$

$$x = (x_1, x_2, x_3, \dots, x_p)^T = \text{多変量ベクトル} \quad (3.4)$$

これを 2 次元上でしめすと以下の図 3.1 のようになる。図 3.1 にある青色の十字がデータの平均値であり、青色の楕円がこのデータ群におけるマハラノビス距離の等距離である。赤色のユークリッド距離における等距離とは異なる。

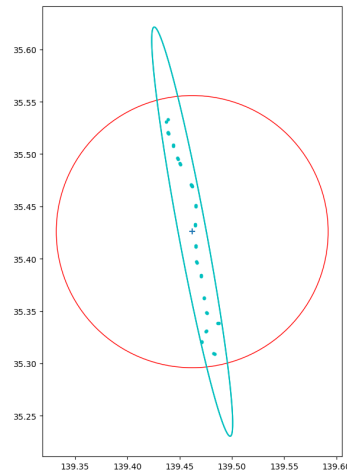


図 3.1: マハラノビス距離のイメージ

この距離は統計学で多数使われ異常値の検出手法として主に用いられている。本提案の場合、推定したいエリアの内側では取りえない値（異常値）ではないことを確認した場合そこに居ると推定する。

マハラノビス距離は多次元で定義されている距離であり、今回は緯度・経度の2次元でデータを与えた。GNSSでは高度のデータも取得できるが誤差が大きいため使用しなかった。

3.4.2 路線の推定

国土数値情報 鉄道データ [1] を利用し路線の駅の情報を取得し、各路線のデータ群を作成する。GNSSによって取得した座標データの各路線ごとのマハラノビス距離を計算し閾値を下回った路線を候補とする。

3.4.3 駅・駅間の推定

3.4.2項で推定された路線を基に、国土数値情報 鉄道データ [1] を用いその路線に含まれている駅ごとのデータ群を作成する。これを前述の座標データを用い各駅ごとのマハラノビス距離を計算し閾値を下回った駅にしていると推定する。このとき駅が属する路線にしていると推定する。

3.4.4 屋内空間の推定

区画に設定したい空間の角の座標データをあらかじめ取得し、このデータを基にデータ群を作成する。これを屋内位置測位技術によって取得した座標データを用いマハラ

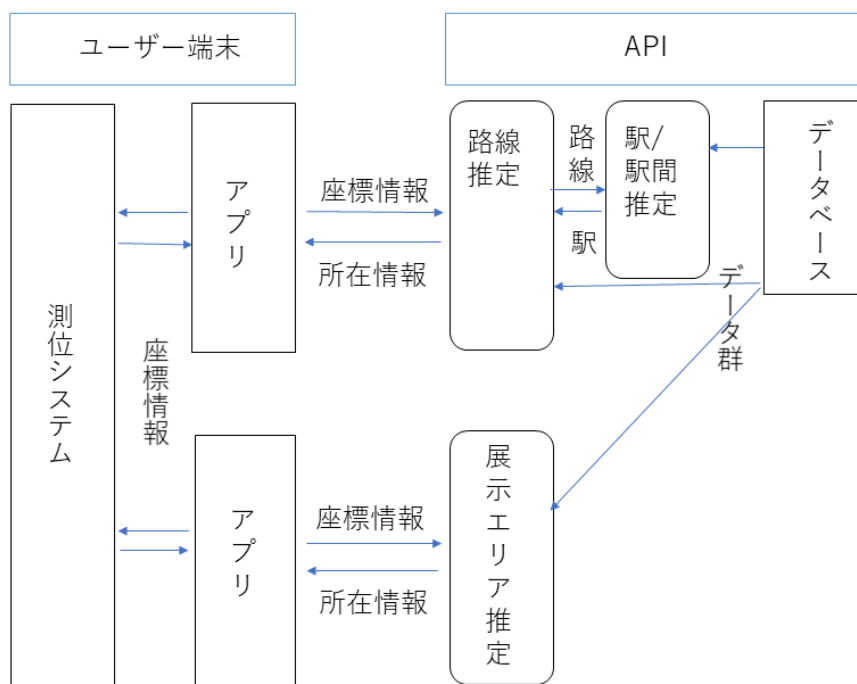


図 3.2: 本提案手法のフローチャート

ノビス距離を計算し閾値を下回った空間にいると推定する。

3.5 フロー図

図 3.2 に本提案手法のフローを表す。以下に具体的な動作を列挙する。

1. 位置情報の取得

アプリケーションがリクエストを出し端末の GNSS などを用いて座標情報を取得する。

2. 路線推定

座標情報を 3.4.2 項にて述べた手法で解析を行う。このとき必要なデータ群をデータベースより取得する。その結果を推定された路線の候補を座標情報とともに駅・駅間推定 3.4.3 により解析する。

3. データベース

各推定に使用するデータ群を提供する。データ群は処理済みであり、マハラノビス距離で必要なデータ群の平均・標準偏差・共分散行列の状態である。なおかつツリー型のデータ構造を持ち路線が駅の親のような形である。

4. API

推定された所在情報は、様々なアプリケーションが使用することを想定する。図 3.2 ではアプリとありスマートフォンや web での使用であるが、カーナビなどの他のサービスでも使用したい。基本的には座標情報を含めた上でリクエストを送り所在情報が帰ってくることを想定しているが、セキュリティの問題やオフラインデバイスでの利用を考慮し、リクエストが送られて場合必要なデータ群を送信するなども考えられる。

第4章 サービスの実装

本提案手法の妥当性を検討する評価実験を行うために、実際にマハラノビス距離を利用して所在情報の特定を行うサービスを実装した。本章ではサービスの構成と、実験データの収集について述べる。

4.1 設計

本サービスはマハラノビス距離を利用しエリアを数理モデル化することによりスマートフォンなどで取得した座標情報を所在情報に変換するサービスである。

4.1.1 路線について

本サービスでは路線を *rosen* クラスとして実装する。表 4.1 に主要なプロパティとメソッドを列挙する。

class rosen	
プロパティ名	概要
rosenData	路線に含まれるすべての座標データ
rosenAvg	路線に含まれるすべてのデータの平均ベクトル
rosenStd	路線に含まれるすべてのデータの標準偏差ベクトル
rosenInvR	路線に含まれるすべてのデータの共分散行列
メソッド名	概要
getData	国土数値情報 鉄道データ [1] から取得した json を parse し rosenData に格納する
beforemaha	データが未処理だった場合平均・標準偏差・共分散行列を計算し json 形式で保存する

表 4.1: 路線を表す *rosen* クラス

4.1.2 路線と駅

路線データは駅データを内包しており、駅データの座標を集め路線データとしている。このため 4.1.1 で路線データを取得する時に路線ごとに駅データをまとめる。

駅

駅は *eki* クラスとして実装される．表 4.2 に *eki* クラスの主要なプロパティとメソッドを列挙する．

class Eki	
プロパティ名	概要
ekiData	路線内の各駅のデータを格納する配列
ekiAvg	路線内の各駅のデータの平均を格納する配列
ekiStd	路線内の各駅のデータの標準偏差を格納する配列
ekiInvR	路線内の各駅のデータの共分散行列を格納する配列
ekiName	路線内の各駅の名前を格納する配列
メソッド名	概要
getData	駅のデータを路線ごとにまとめられた json から読み格納する
beforemaha	データが未処理だった場合平均・標準偏差・共分散行列を計算し json 形式で保存する

表 4.2: 駅を表す *eki* クラス

4.1.3 ユーザーについて

本実装ではリクエストを行う人物を *user* クラスとして実装する．表 4.3 に *user* クラスの主要なプロパティとメソッドを列挙する

class User	
プロパティ名	概要
positions	リクエストに存在する座標情報を格納する配列
ansRosen	マハラノビス距離を利用し推定した路線の候補を格納する配列
ansEki	マハラノビス距離を利用し推定した駅の候補を格納する配列
メソッド名	概要
getRosen	マハラノビス距離を利用し路線を推定する
getEki	マハラノビス距離を利用し駅を推定する

表 4.3: リクエストを送る人物を表す *User* クラス

データ構造

本項では本実装においてどのようなデータ構造をしているかを示す．

1. 鉄道ディレクトリ

今回の場合ルートに当たる階層である

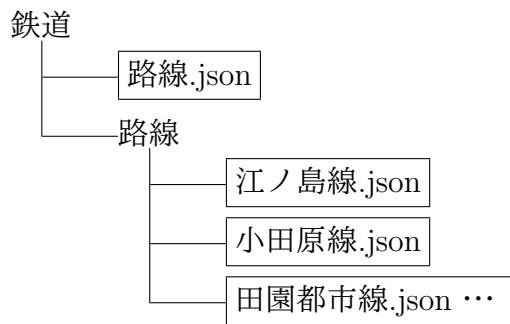


図 4.1: データ構造

2. rosen.json
各路線の平均値ベクトル・標準偏差ベクトル・共分散行列が格納されている
3. 路線ディレクトリ
このなかに各駅の情報が入った json が格納されている
4. 江ノ島線.json
各駅の平均値ベクトル・標準偏差ベクトル・共分散行列が格納されている

4.2 路線・駅の推定

本実装では第 3 章で述べた提案手法を用いて、路線と駅の推定を行う。

4.2.1 路線の推定手法

マハラノビス距離を用いて路線の推定を行う。路線が絞られることにより内包している駅についても候補が絞られるために計算量を少なくすることができる。

用いるデータ

国土数値情報 鉄道データ [1] より各路線のセクションデータおよび駅データの緯度・経度を抽出。これを事前に処理しデータの平均及び標準偏差・分散行列を格納した路線.json を使用する。

用いる実装

GNSS などから座標情報を取得し路線ごとに 3.4.1 で述べたマハラノビス距離を計算する。このとき分布が正規分布に従うと考えそのデータ群が取る確率が 99% である 2.58 よりも小さい場合その地点にいと推定する。これをすべての路線に対して行う。この範囲をプロットすると下図 4.2 のようになる。青が小田原線、緑が田園都市線、水

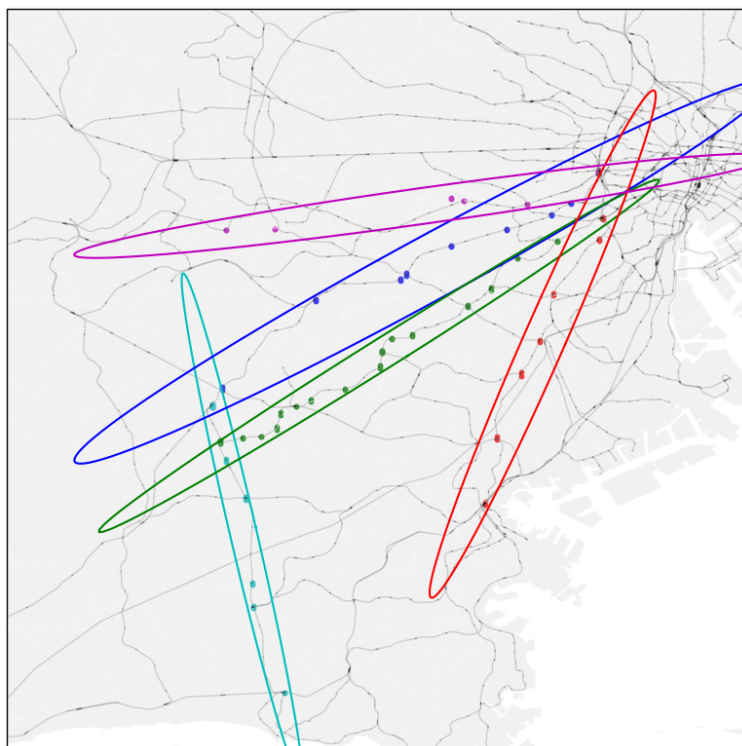


図 4.2: 路線推定範囲¹

色が江ノ島線、赤が東横線、紫が京王線である。この円の中に居る場合その路線を候補とする。範囲が被った場合は相関係数が小さい方を優先する。

4.2.2 駅の推定手法

4.2.1 と同様にマハラノビス距離を利用し一定数値を下回った場合その地点にいると推定する。

用いる実装

路線の候補から駅のデータ群が含まれている json を読み出す。これにより必要な計算リソースを削減することができる。駅のデータは駅に幅がない。そのためあらかじめデータの平均に影響を与えないように分散した 2 点を追加し計算を行っている。駅推定の流れは下に示す。

1. データ群を取得する

候補の路線名を利用し json ファイルを読み、そこからデータの平均及び標準偏差・分散行列を取得する。

¹国土数値情報 鉄道データを加工 [1]

2. マハラノビス距離を算出する
入力された座標情報とデータ群を比較し閾値を下回った場合駅にいると推定する。
3. 路線の確定
駅にいることが確認できた場合その駅が属している路線にいると推定する。。

4.3 実装環境

本説では本シミュレータの実行環境について記述する。
本シミュレータ及び推定システムはPython3²とそのライブラリ群を用いて実装した。
図 4.3 は第章で記述する評価実験を実行した環境を表している。

実行環境	
ソフトウェア	バージョン
Python	3.7.4(anaconda3-5.3.1)
NumPy ⁵	1.15.4
pandas ⁷	0.23.4
SciPy ⁸	0.19.1
matplotlib ⁹	3.0.1
json ¹⁰	2.0.9
ハードウェア	チップセット
FUJITSU LIFEBOOK WS1/W Intel(R) Core(TM) CPU i5-6200U @ 2.40GHz	

表 4.4: シミュレータの実行環境

²2018 年 1 月現在, Python 言語の現行仕様バージョンは Python3 と呼ばれる. <https://www.python.org/>

⁵数学的な計算や行列式に用いるライブラリ. <http://www.numpy.org/>

⁷データを高速に扱うライブラリ. <https://github.com/pandas-dev/pandas>

⁸数値計算用ライブラリ. <https://www.scipy.org/>

⁹図表のプロットを行う. <https://matplotlib.org/>

¹⁰JSON を利用するライブラリ. <https://github.com/dvarrazzo/py-setproctitle>

第5章 評価

第4章にて述べたシミュレータを用いて，本提案手法の妥当性を検証するための評価実験を行う．本章では評価実験の概要と，それにより得られた結果及び考察について述べる．

5.1 仮定

本節に本評価実験の仮定条件となる事項を記述する．

5.1.1 実験環境について

実験環境についてしらべる

各路線の特徴

5.1.2 条件設定

本評価実験の際に設定する条件について記述する．

5.2 評価実験

評価項目

説明変数と目的変数

結果

考察

5.2.1 評価実験2：駐車区画が保有する駐車枠数の推定

評価項目

結果

考察

第6章 結論

本章では本論文のまとめと今後解決すべき課題を示す.

6.1 本研究のまとめ

6.2 将来的な展望

謝辭

参考文献

- [1] 国土交通省. 国土数値情報 鉄道データ, 2018. http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02-v2_3.html 最終確認日 2019 年 12 月 22 日.