

社会連携・産学協創推進室シンポジウム
AI応用ワークショップ

ジオタグ付きTweetデータを用いた観光行動分析

東京大学大学院工学系研究科
鳥海不二夫

計算社会科学

- 人間・社会と計算の組み合わせが可能になったことで実現した新しい学際科学
 - 現実世界を映し出すビッグデータ
 - 観察・実験の場としてのウェブ
 - 複雑な社会を扱う数理と情報技術

社会心理学・社会学
社会情報学・政治学

社会
科学

計算社
会科学

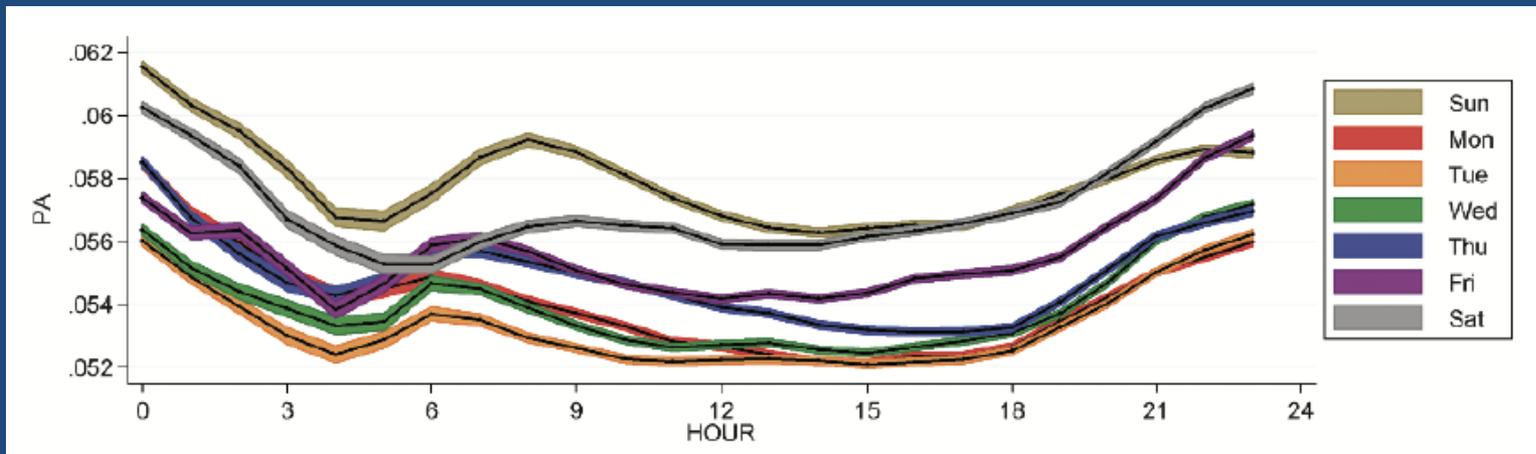
情報
技術

数理
手法

計算機科学
ビッグデータ
人工知能
機械学習
自然言語処理

統計数理
経済物理学, 社会物理学
ネットワーク科学
ゲーム理論・ABS

Twitterにおける気分の時間変化



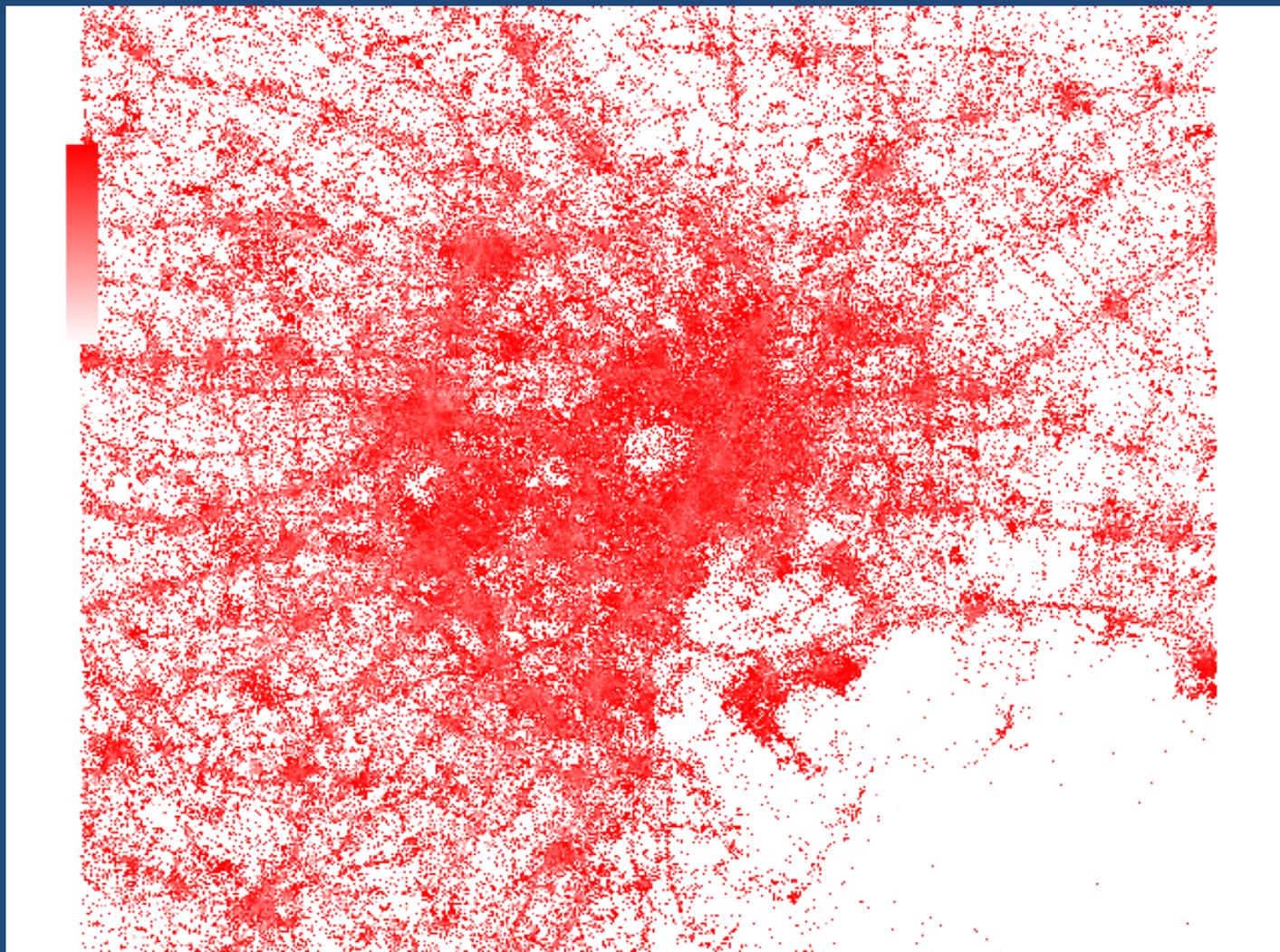
Golder and Macy. Science, 2011

- Twitterを分析することで人間の「気分」を計測

ジオタグ付きTweetデータ

- データ
 - 毎日、日本で30万件以上
 - リアルタイムで収集可能
 - 精緻な位置情報(緯度経度)
- 使い方
 - 普段の何気ないつぶやき(家・職場)
 - 特別なイベント(旅行など)

都内のツイート



国際旅行客増加と地域経済

- 全世界で国際旅行客数の増加

1990年	2000年	2010年	2014年
4.4億人	6.7億人	9.5億人	11.3億人

全世界の国際旅行客数:国連世界観光機関調べ

- 観光振興の議論は

「いかに外国人観光客を呼び寄せるか」から、
「いかに外国人観光客を国内の各地域に結びつけるか」へ

観光とツイート

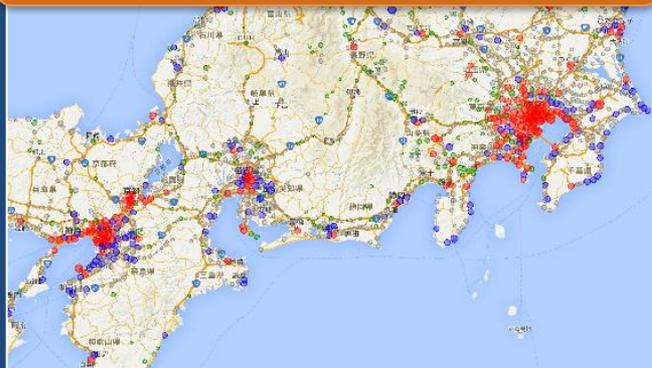
- 観光地でのツイート
 - 多くの人が観光地で感想をツイート
- 感想から観光地の特性を推定
 - 外国人観光客の期待に見合う観光資源を提供

研究目的

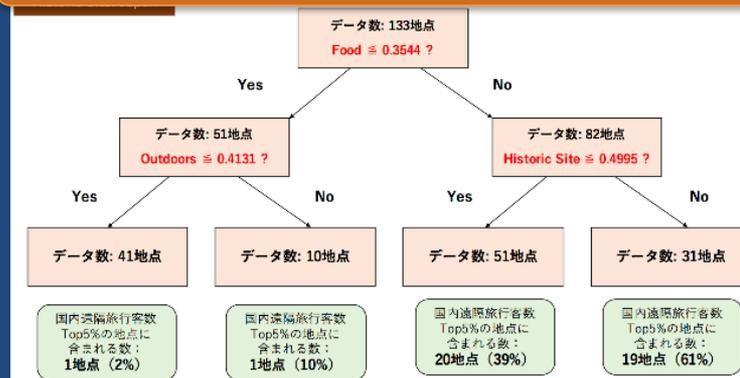


ジオタグ付き
Tweet

観光地情報の抽出



観光客数の決定要因分析

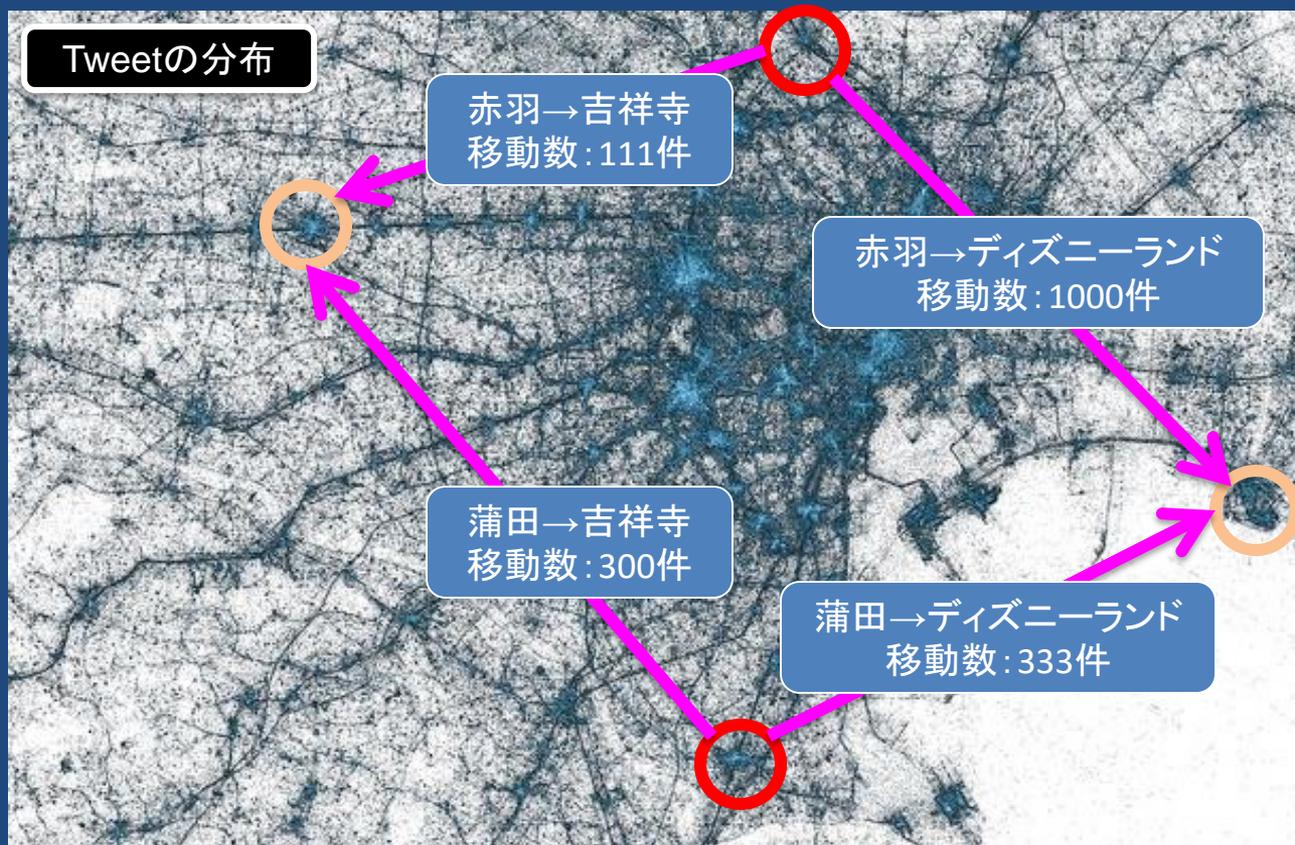


【社会応用】国内の観光資源に外国人観光客を呼び寄せするための改善策を提示

地域分類



Twitterからの地域間移動抽出



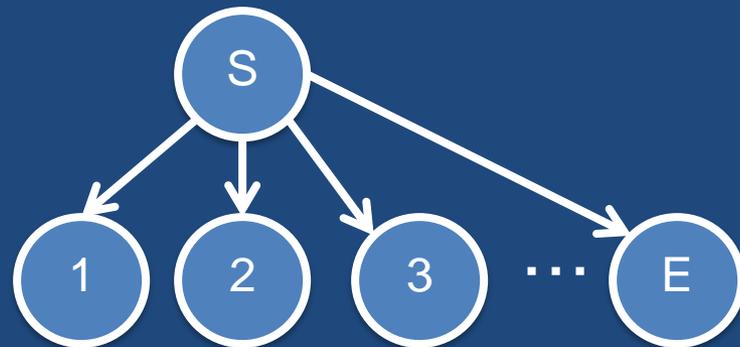
- 1.各ユーザの居住地と目的地(旅行・外出先)の特定
- 2.全ユーザの居住地・目的地をHotspotへ集約

目的地の競合を考慮した 目的地の魅力推定

$$\text{地域 } S \rightarrow E \text{ の移動数} \propto \frac{(\text{地域 } S \text{ の総出発数}) \cdot (\text{地域 } E \text{ の魅力})}{(S \cdot E \text{ 間の距離})^\alpha}$$



$$\frac{\text{地域 } S \rightarrow E \text{ の移動数}}{\text{地域 } S \text{ の総出発数}} = \frac{\text{地域 } E \text{ の魅力}}{(S \cdot E \text{ 間の距離})^\alpha} / \sum_X \frac{\text{地域 } X \text{ の魅力}}{(S \cdot X \text{ 間の距離})^\alpha}$$



魅力の上位地域

順位	地域	順位	地域
1	那覇空港	11	宜野湾
2	秋葉原	12	石垣島
3	博多	13	お台場
4	国際通り・首里城	14	ディズニーランド
5	新宿	15	名古屋
6	渋谷	16	札幌
7	石垣島	17	品川
8	梅田	18	成田空港
9	なんば	19	ユニバーサル・スタジオ・ ジャパン
10	北谷	20	川平湾

2014年8月のデータ

テキスト情報による固有性判定

固有性:高

富士山周辺

【高頻度出現単語】

- 富士山
- 御来光
- ~合目

他では登場しない単語

固有性:低

イオン周辺

【高頻度出現単語】

- イオン
- 駐車場
- 買い物

ありがちな単語

その場所の登場頻度の高い単語が他の場所で登場数が少ない

固有性が高い

結果

【魅力】

- 35位 : 宮崎イオン
- 65位 : 高知イオン
- 138位 : 倉敷イオン
- 142位 : 伊勢神宮
- 329位 : 厳島神社



【固有性】

- 25位 : 厳島神社
- 135位 : 伊勢神宮

- 545位 : 倉敷イオン
- 1042位 : 高知イオン
- 1284位 : 宮崎イオン

地域分類

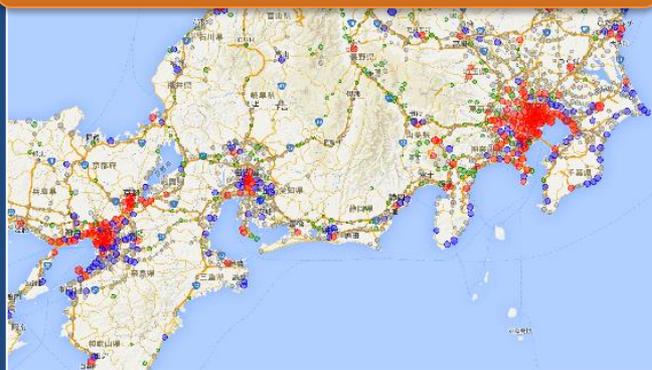


研究目的

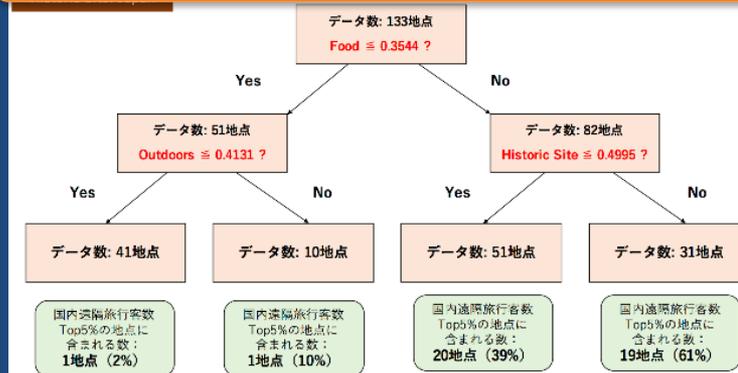


ジオタグ付き
Tweet

観光地情報の抽出



観光客数の決定要因分析



外国人観光客数・国内観光客数の の違いの分析

外国人・国内観光客数の分布の違いを把握

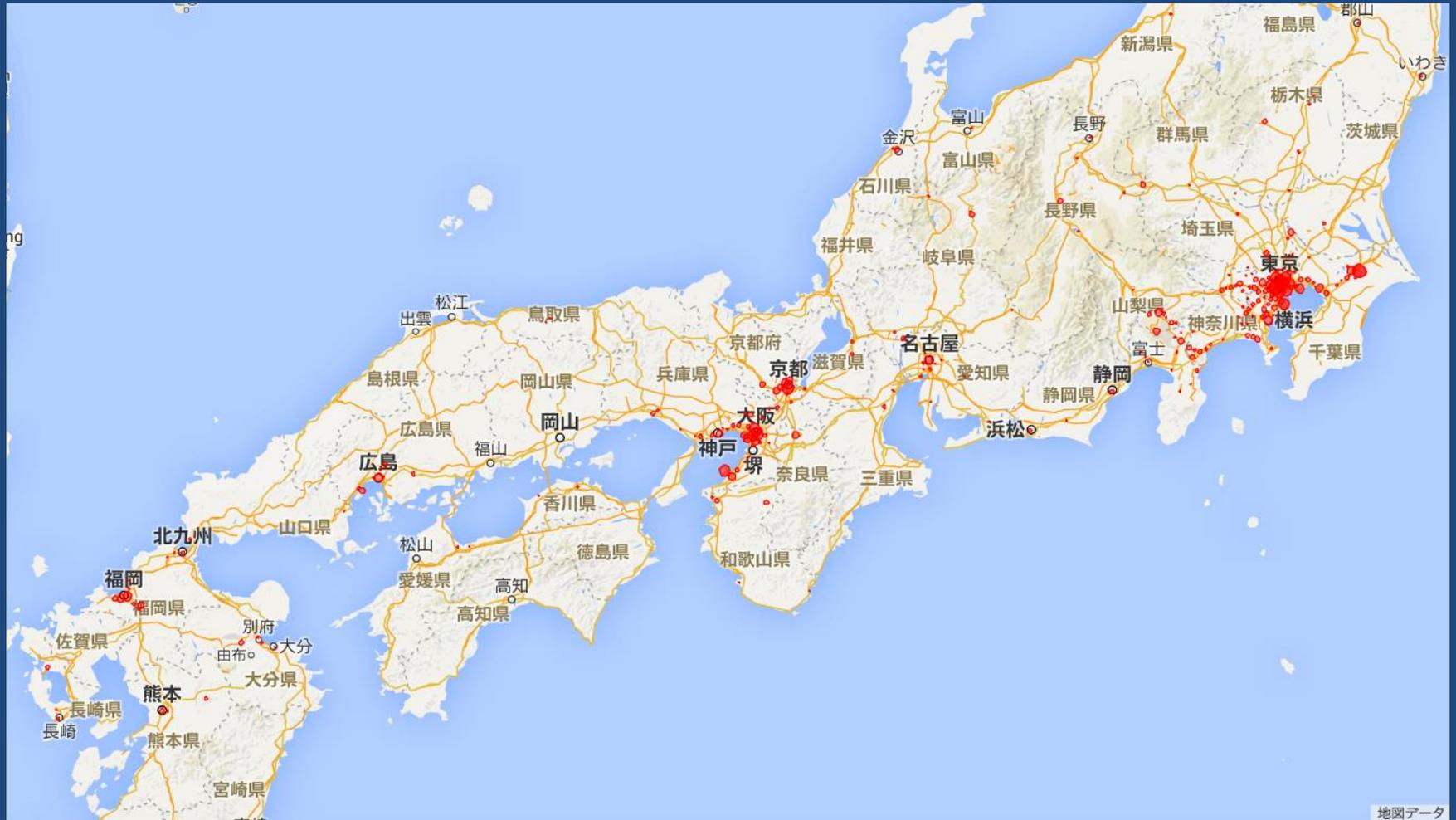


各観光地周辺の特徴の把握



観光客数の決定要因の分析

外国人観光客訪問地の分布



Foursquareを用いた特徴の把握

Foursquare

- ユーザが自身の行った場所(Venueと呼ぶ)を評価するソーシャルメディア
- 各Venueにはカテゴリ(ホテル・レストランなど)

特徴付与: 観光地の周囲の各カテゴリの充実度

Foursquare

ホテル

史跡

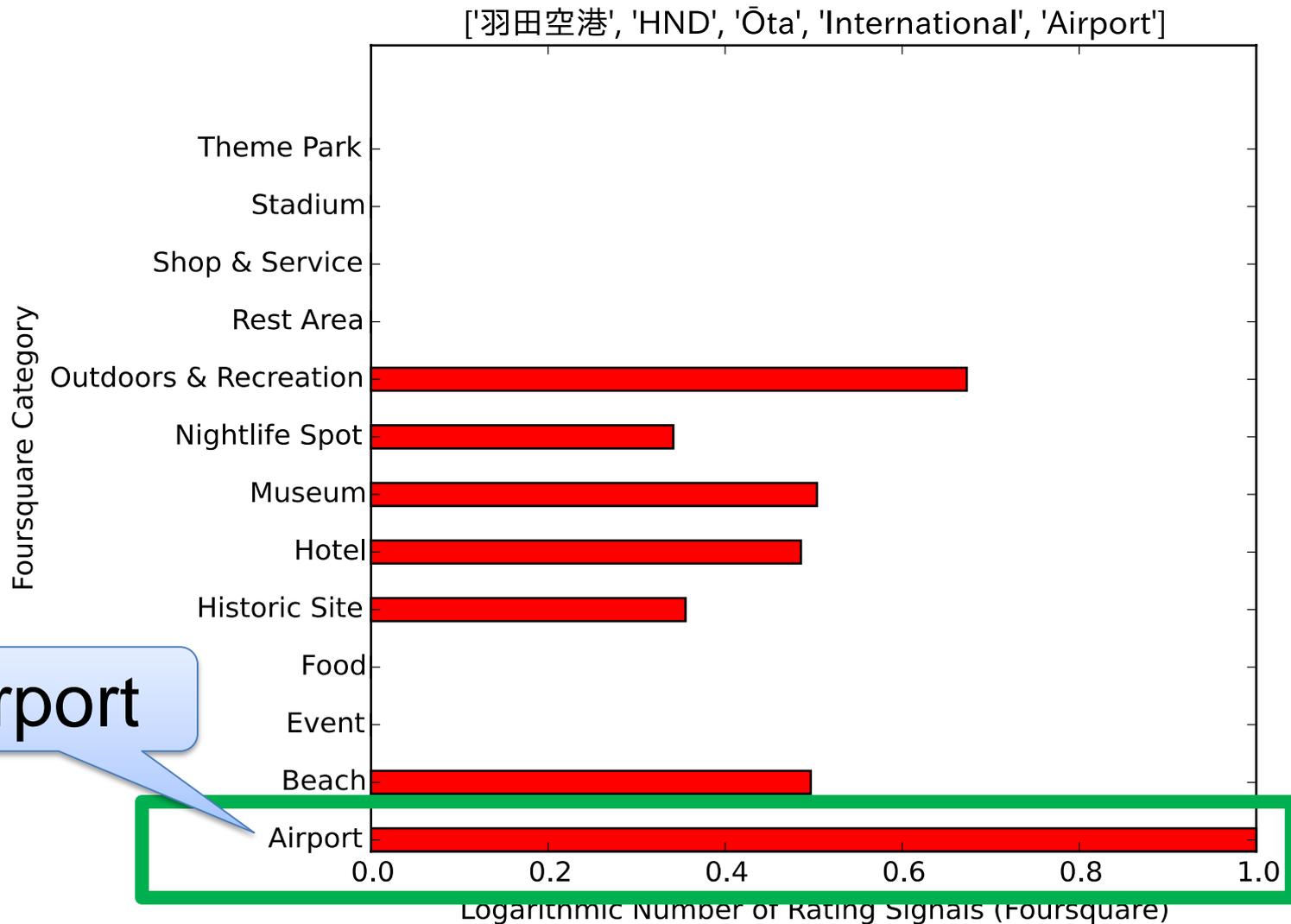
レストラン

商店

Foursquareのカテゴリー一覧

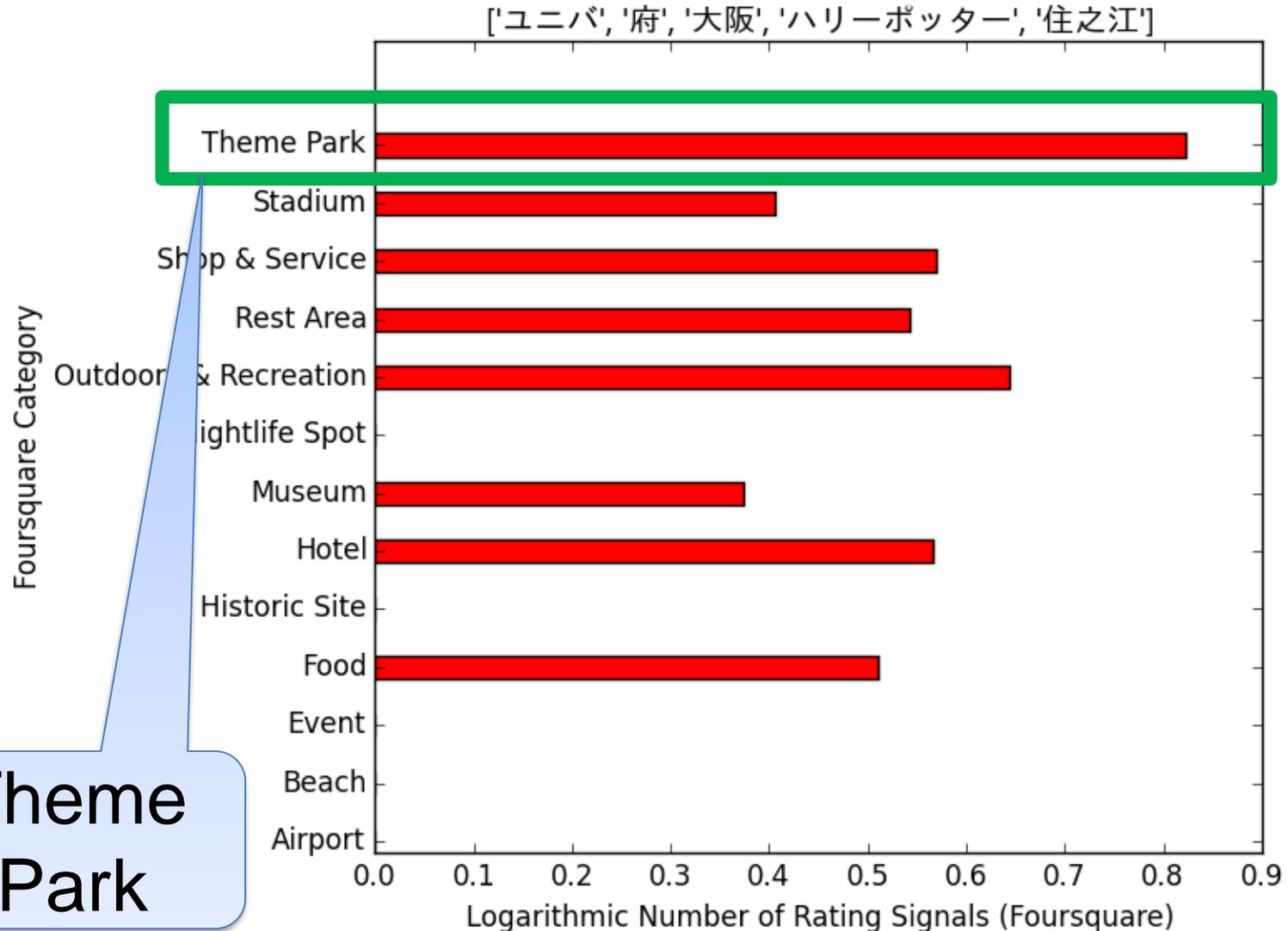
カテゴリー名	内容
Airport	空港
Beach	海岸, ビーチ
Event	イベント
Food	レストラン
Historic Site	史跡
Hotel	ホテル
Museum	博物館, 美術館
Nightlife Spot	ナイトスポット(バー, クラブなど)
Outdoors & Recreation	公園, アウトドアスポーツなど
Rest Area	サービスエリア
Shop & Service	商店全般
Stadium	競技場
Theme Park	テーマパーク, 遊園地

特徴付与の結果(例):羽田空港



Airport

特徴付与の結果(例): USJ

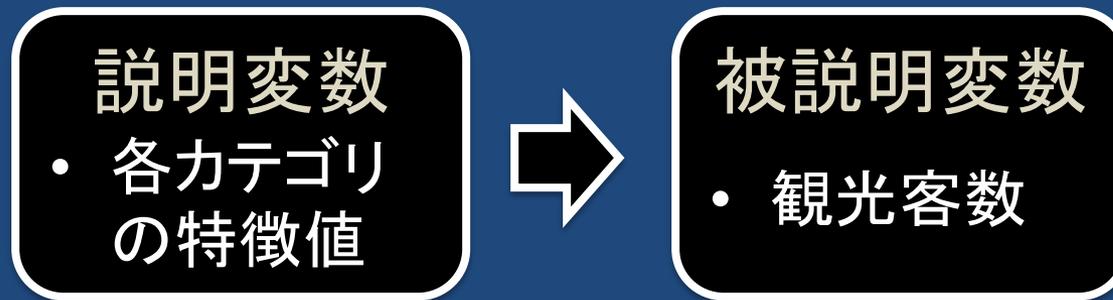


Theme
Park

決定木分析

- 決定木

- 説明変数と被説明変数の関係を木構造で表現



- 対象

- 史跡, 遊園地の特徴が上位5%の地点
- 外国人観光客, 国内観光客

決定木(史跡): 外国人観光客

データ数: 133地点
ナイトスポット ≤ 0.4273 ?

外国人観光客数
Top5%の地点数

Yes

No

データ数: 93地点
ホテル ≤ 0.5072 ?

データ数: 40地点
レストラン ≤ 0.5392 ?

Yes

No

Yes

No

データ数:
89地点

データ数:
4地点

データ数:
3地点

データ数:
37地点

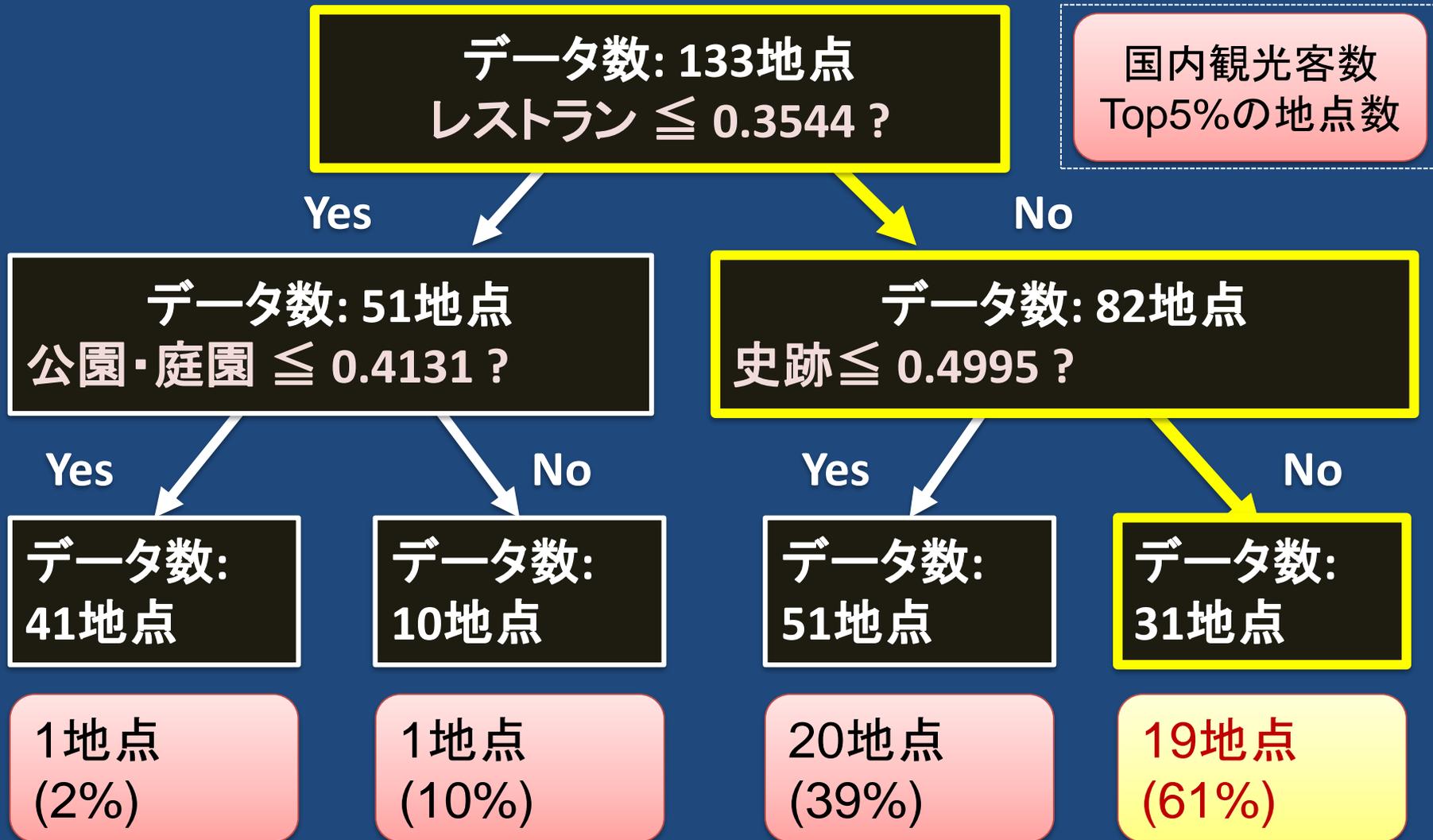
7地点
(8%)

3地点
(75%)

0地点
(0%)

30地点
(81%)

決定木(史跡): 国内観光客



日本の観光資源の比較(史跡)

外国人が訪れる場所

- 東京
- 京都
- 金沢
- 函館
- 広島

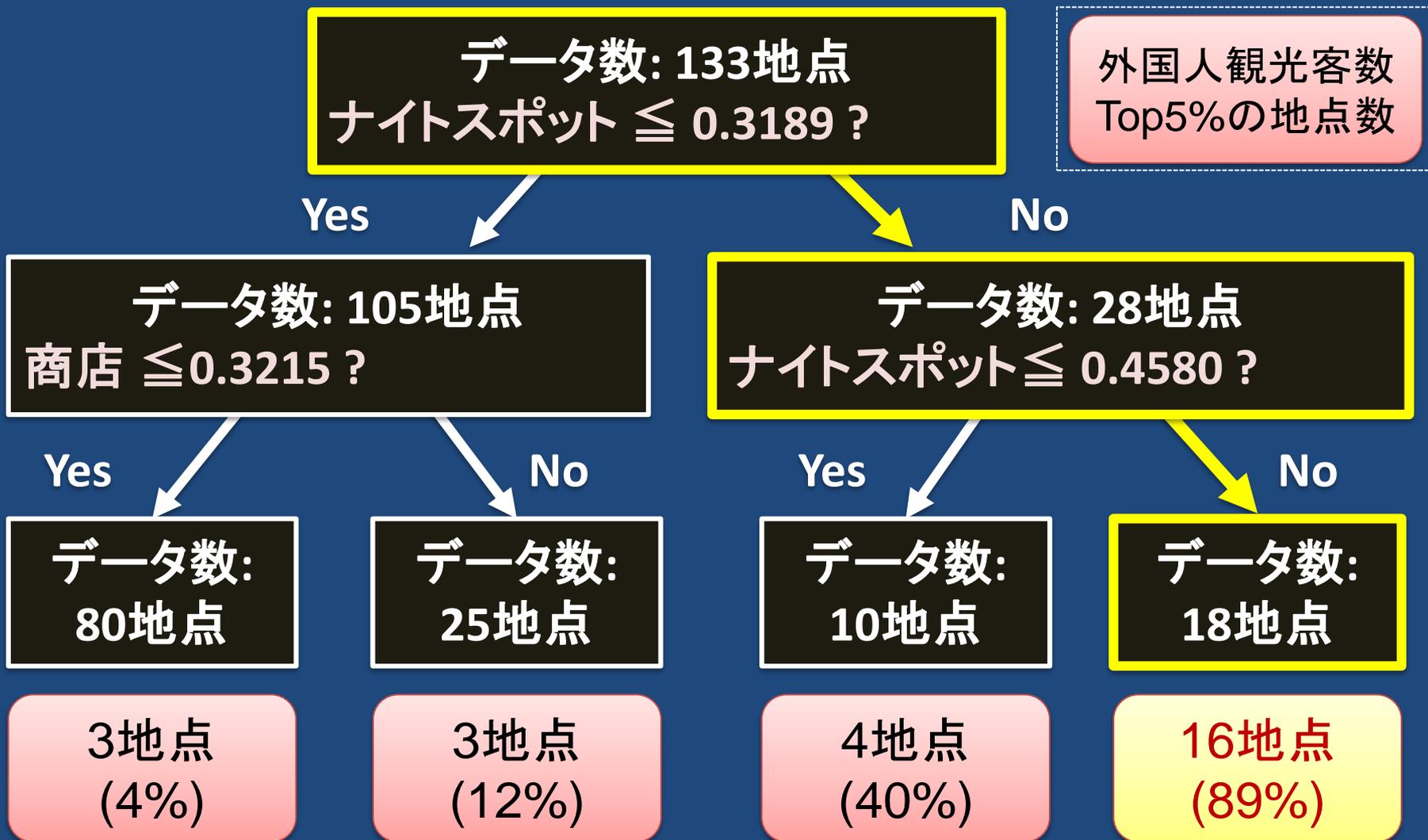
ナイトスポット,
ホテルともに充実

外国人が訪れない場所

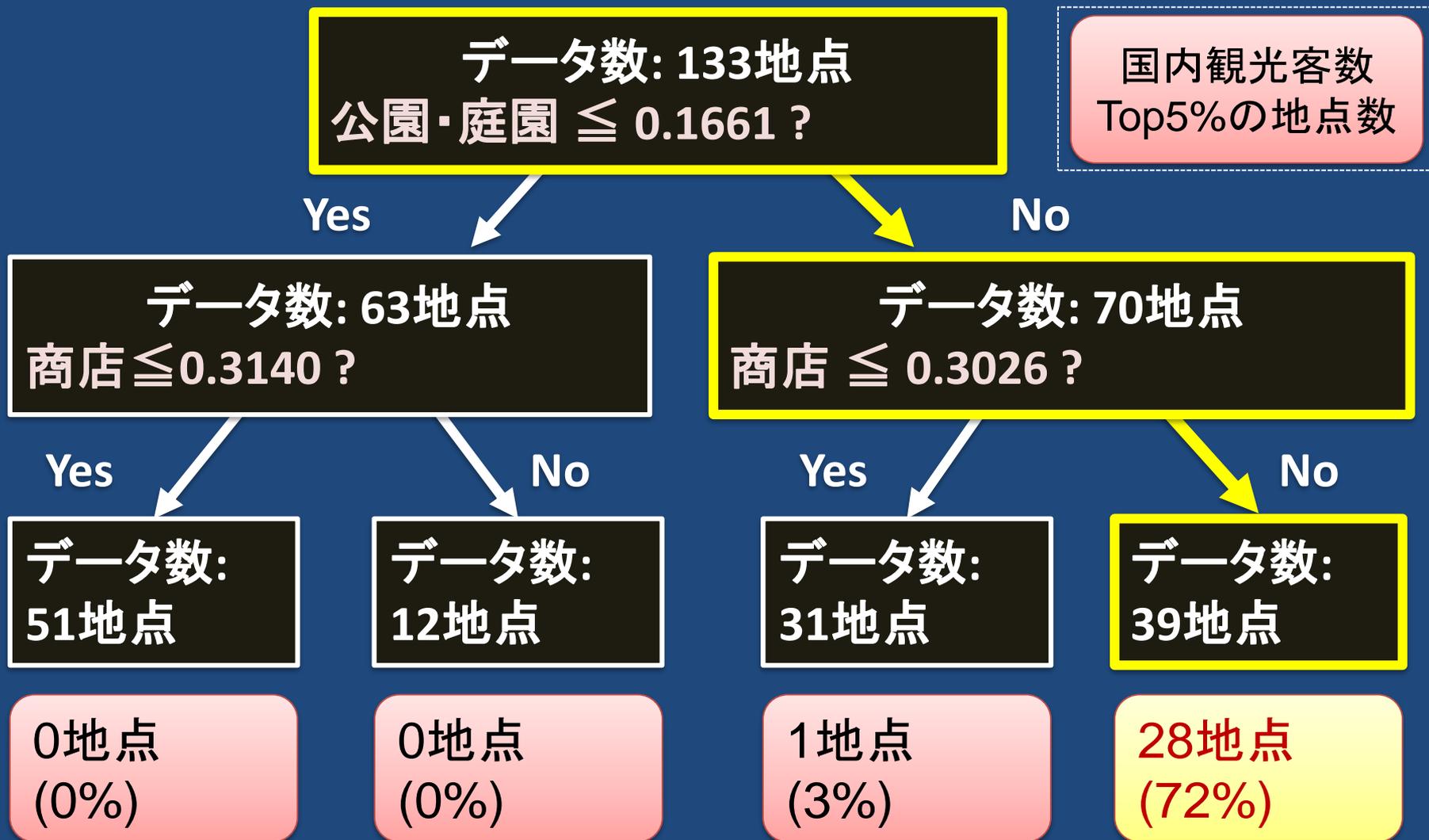
- 白川郷
- 松江
- 伊勢
- 法隆寺
- 姫路城

ナイトスポット,
ホテルが少ない

決定木(遊園地): 外国人観光客

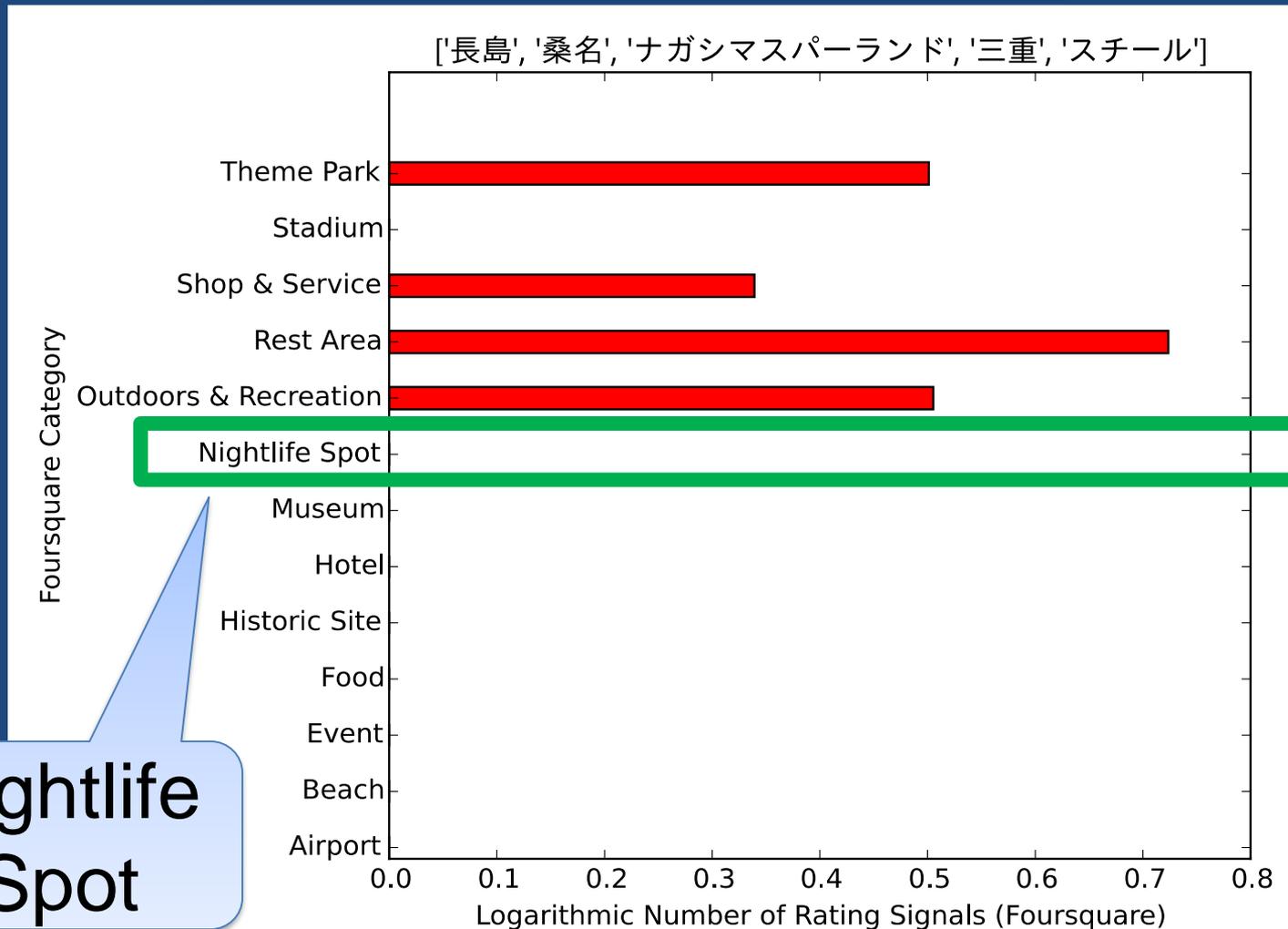


決定木(遊園地): 国内観光客



外国人が訪れない遊園地

ナガシマスパーランドの特徴



Nightlife
Spot

Twitterからわかる観光戦略

- 外国人観光客
 - 史跡に行くためには
 - ナイトスポットとレストランが必要
 - 遊園地に行くためには
 - ナイトスポットが必要
- ナイトスポットの充実が外国人観光客増加につながる
 - 夜まで遊ぶ
 - 日本人には効果なし
 - 旅館でまったりを好む？

結論

- 行動データが社会を可視化する
 - 見えなかったものが見えてくる
 - 予想の裏付け
 - 予想外の発見
 - 情報技術が支援する人間・社会
 - デジタルに刻まれた大量の行動痕跡



- 社会の理解が変わる
 - より詳細に・より一般性を持って
 - 本当の社会を理解しサービスに生かす