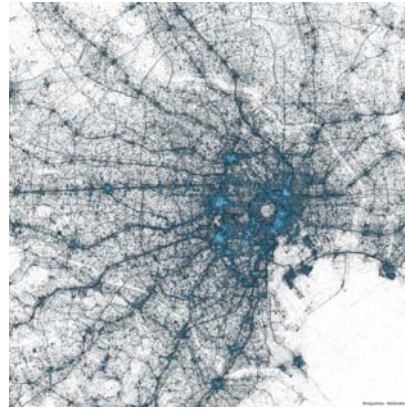


# ビッグデータと観光の分析

首都大学東京大学院 都市環境科学研究科 観光科学域

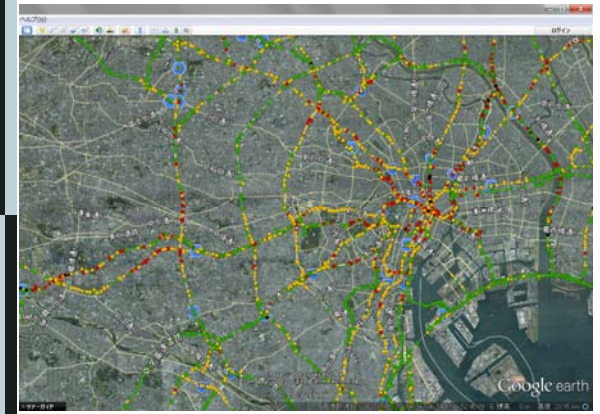
倉田 陽平

ykurata@tmu.ac.jp



<https://blog.twitter.com/2013/geography-tweets-3>

2



3



<http://safetymap.jp>

4

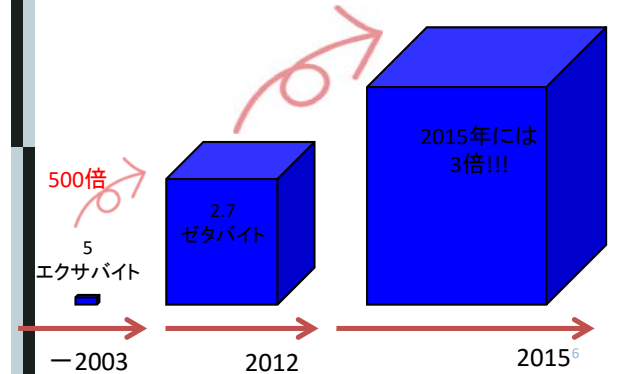
## ビッグデータとは？

- SNS書き込み
- 検索履歴
- 商取引
- 移動(カーナビ・IC乗車券)
- 人工衛星画像
- 防犯カメラ
- ヘルスケア

デジタル社会の中で日々創られる超膨大なデータ

5

## ビッグデータの規模



## ビッグデータを用いた観光研究の例

- 倉田研
  - Suica利用履歴を用いた訪日外国人の行動分析
  - Flickrデータを用いた観光地の「見所」の分布の地図化
  - ニコニコ動画データを用いた観光関連コンテンツの共創状況の分析
- 清水研
  - Honda車走行履歴データを用いた事故危険箇所の分析
  - ナビタイムカーナビアプリのデータを用いた晴天・雨天時の立ち寄り行動
- 沼田研
  - 衛星画像を用いた熱帯雨林の降雨量推定
  - 衛星画像を用いた桜の開花状況推定
- 石川研 (SD 情報通信)
  - Flickrデータを用いた有名撮影対象の撮影箇所・方向の分析
  - twitterデータを用いた各国訪日外国人のつぶやく街ランキング
  - twitterデータを用いた桜やアジサイの開花時期の分析
- 渡邊研 (SD インダストリアルアート)
  - ウェザーレポートを用いた台風状況のリアルタイム可視化

7

## 4つのステップ

1. 取得
2. 精選
3. 分析
4. 可視化



## Step 1. 大量データの取得

- 買う
- もらう
- 狩る
  - APIの利用
  - スクレイピング (クローリング)



## APIを利用したデータ取得とは

特定のURLにパラメータをつけてリクエストすると、データだけが帰ってくる

例: <https://maps.googleapis.com/maps/api/geocode/json?address=南大沢駅>



→このようなURLをプログラムで自動生成してリクエストし、帰ってきたデータもプログラムで処理する

10

## 観光関連サイトのAPI提供状況

無償

- Flickr
- twitter
- Facebook
- Instagram
- 楽天トラベル
- ぐるなび
- Expedia

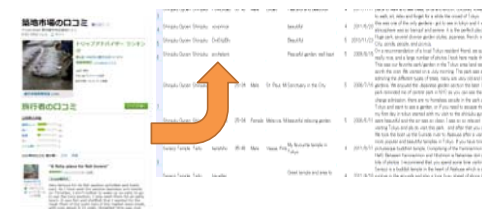
有償(企業のみ)

- TripAdvisor
- 提供やめた
- フォートラベル



## スクレイピング

テキスト解析によって大量のウェブページから必要とするデータを自動的に集めてくること。「クローリング」ともいう。



12

## 例: Google Playのスクレイピング



## スクレイピングの原理



## スクレイピングの原理



## スクレイピングのやり方

プログラム(あるいはエクセルのマクロ)を書いて、以下の処理を自動的に行う

1. 情報を入力したいページのURLを生成
2. 各ページのソース(html)を自動入手(※スピードが速すぎるとアクセス遮断されるリスクあり)
3. ほしいデータの前後に必ず出現する文字列を手がかりに各データを抜き出す
4. 抜いたデータを記録していく

難しく聞こえるが、実際には、誰か他人の書いたプログラムを改造するだけ

16

## Step 2. データの精選

収集した膨大なデータの中から、自分の分析に必要なデータだけを抜き出すステップ  
クレンジングとも言う

- ゴミが入り混む理由:
- キーワード検索の限界
  - ボットが参加している
  - 他言語のデータ
  - 観光客以外のデータ



## クレンジングの実例: Google Playのご当地観光アプリのケース



## クレンジングの実例: Google Playのご当地観光アプリのケース

- スクレイピング直後: 4915
- 外国語除去後: 1340
- ご当地観光アプリを選別後: 764

## Step 3. データの分析

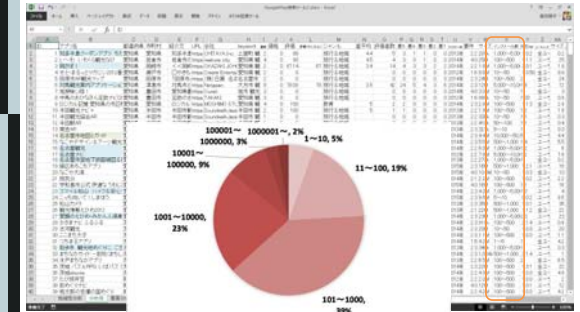
次のいずれかに限られる

- 数値データ → 統計分析(集計)
- 位置を含むデータ → 空間解析
- 自由記述データ → テキストマイニング



大量データの処理法

### ①集計

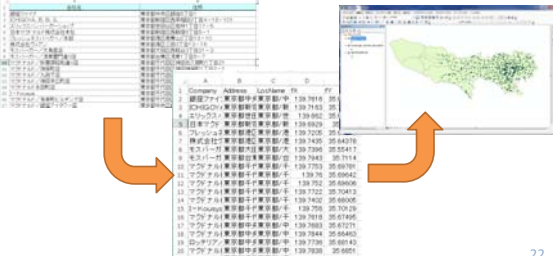


21

大量データの処理法

### ②地図化

- 住所など位置情報があったら、何はともあれジオコーディングして地図にプロット



22

大量データの処理法

### ②地図化

- 位置のみの場合 → 点分布分析



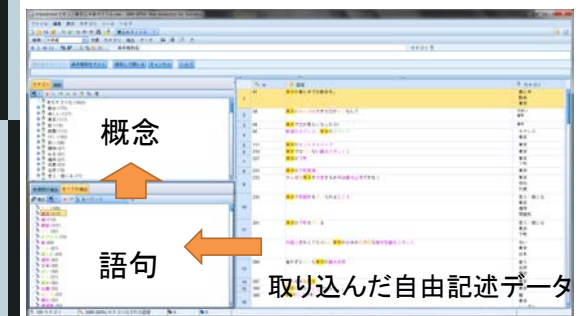
- 位置 + 数値 → サーフェス分析



23

大量データの処理法

### ③テキストマイニング



SPSS Text Analytics for Surveys

24

大量データの処理法

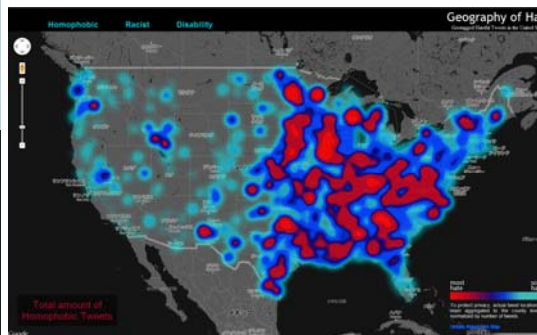
### ③テキストマイニング

観光クチコミに使われる概念

good (275)	would recommend (52)	english (36)	fast (20)
place (160)	people (51)	photos (37)	displays (19)
tokyo (125)	night (51)	tour (37)	difficult (19)
large (107)	bad (46)	child (35)	friends (18)
shopping (105)	new (47)	busy (33)	long (18)
walk (74)	beautiful (45)	expensive (33)	arrive (18)
park&garden (73)	would be good (46)	stop (31)	family (17)
food (73)	interesting (45)	old (30)	night (17)
visit (71)	japanese (42)	building (29)	easy (17)
enjoyable (63)	relaxing (42)	open (26)	entrance (15)
cheap/free (61)	city (42)	tourists (25)	real (15)
worth (60)	attractions (41)	stay (24)	world (15)
small (59)	nearby (40)	price (23)	disappointing (15)
japan (59)	experience (40)	problem (23)	closed (15)
view (53)	temple&shrine (40)	famous (22)	fish (15)
museum (55)	transportation (39)	hotel (21)	space (15)
like (52)	available (20)		

25

## テキストマイニングと地図化の組み合わせコンボ



## ビッグデータ分析の実例 —観光ポテンシャルマップ研究—



27

## flickrからのデータ抽出ツールを開発

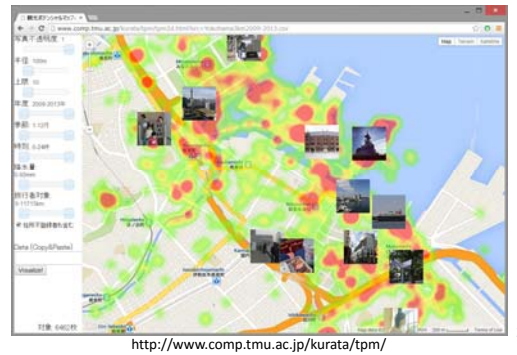
Flickrから大量の画像のデータを持ってくるツール

データクレンジング

Google Mapsでヒートマップ化

35.454033, 139.636298  
35.451494, 139.647303  
35.4441, 139.647962  
35.4441, 139.647962  
35.4441, 139.647962  
35.443219, 139.64872  
35.457034, 139.635818  
35.443219, 139.64872  
35.4441, 139.647962  
35.4441, 139.647962  
35.452743, 139.648139  
35.446304, 139.650017  
35.4441, 139.647962  
35.45103, 139.647238  
35.454382, 139.6356  
35.447255, 139.641767  
35.453434, 139.637915  
35.449955, 139.645119  
35.463438, 139.626153  
35.452984, 139.634951  
35.459888, 139.632003  
35.455041, 139.633289  
35.452743, 139.648139

## 観光ポテンシャルマップ



## 観光ポテンシャルマップ ×キーワード検索



真田風・倉田陽平・相尚寿 (2015) 写真共有サイトに投稿された写真群を活用したテーマ別観光マップの作成. 情報処理学会第77回全国大会

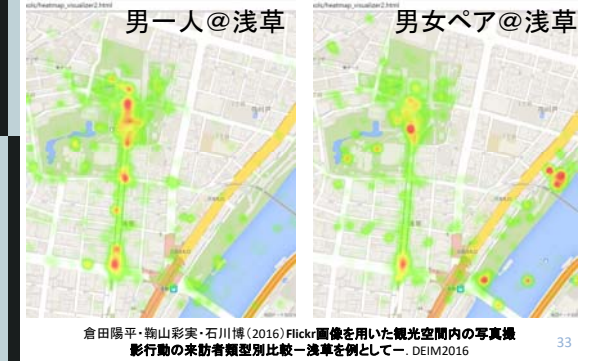
## 観光ポテンシャルマップ×気象データ



## 観光ポテンシャルマップ×街路データ



## 旅行者類型別ポテンシャルマップ



## Flickrデータを利用した他の研究例

- 観光客の分布推定 (日高・磯田 2010)
- 名所マップの自動作成 (Chen *et al.* 2009)
- 特定ランドマークに対する人気撮影スポットの抽出 (Shirai *et al.* 2012)
- 旅行者の観光地間の移動軌跡の推定 (Giradin, *et al.* 2008)
- 旅行者の観光地内の移動軌跡の推定 (Kisilevich, *et al.* 2010; Lu, *et al.* 2010)
- 移動軌跡をもとにした旅程推薦ツールの開発 (De Choudhury, *et al.* 2010; 奥山・柳井 2011)



## 今日のまとめ

- 現在、データは既に膨大にある
- それを研究で有効利用することも可能
- 位置情報のついたデータは地図化、文章のついたデータはテキストマイニングが常套手段
- データの大きさに恐れず、さらにほかのデータとかけあわせると、面白いことが見えることも多い

## 今日のキーワード

- スクレイピング (クローリング)
- データクレンジング
- ヒートマップ化
- テキストマイニング
- Twitter
- Flickr
- 観光ポテンシャルマップ

