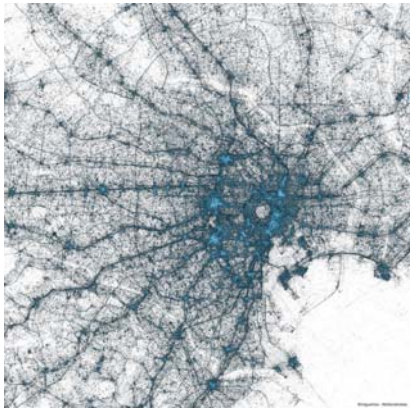
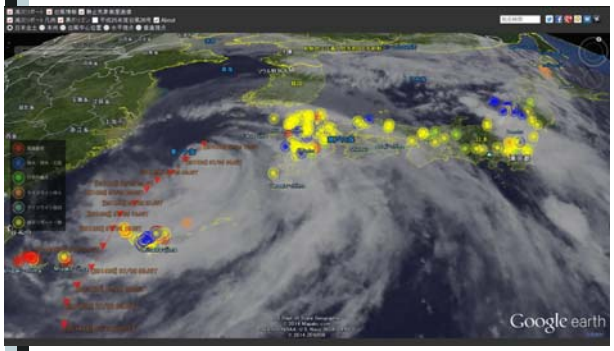
観光地理情報学特論Ⅰ 2017年度第12回

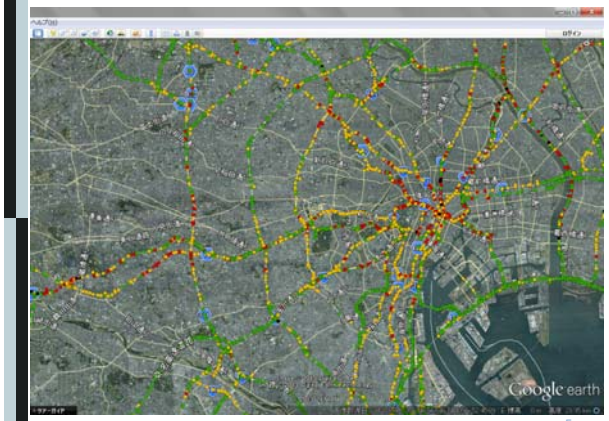
ビッグデータと観光の分析

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
TOKYO METROPOLITAN UNIVERSITY



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





http://safetymap.jp

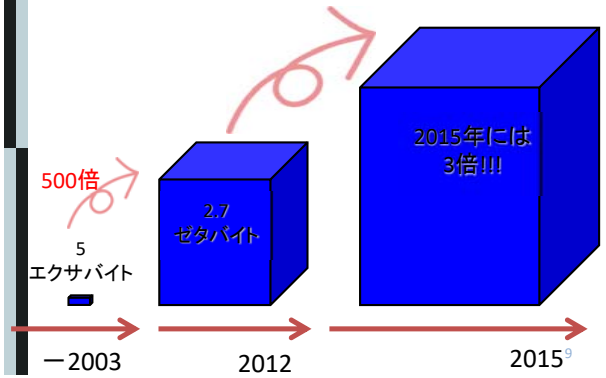
https://www.toyota.co.jp/jpn/auto/passable_route/map/tottori

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

ビッグデータの規模



観光とビッグデータ

旅行者個人が自ら巨大データベースにアクセスし、さらにSNSで情報を発信する時代

旅行前 → 旅行中 → 旅行後

興味	検索	予約	移動	整理共有
facebook	Google YAHOO! JAPAN	楽天トラベル JAL		facebook flickr

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
- 渡邉研（SD インダストリアルアート）
  - ウェザーリポートを用いた台風状況のリアルタイム可視化

4つの論点

1. いかにしてデータを取得するか？
2. いかにデータを精選するか？
3. いかに分析するか？
4. いかに結果を見せるか？

論点1：データの取得

データ取得の方法

- 買う
- もらう
- 狩る
  - APIの利用
  - スクレイピング

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、データが帰ってくる

- 例：http://maps.google.com/maps/api/geocode/json?address=南大沢駅

→プログラムでURLを自動生成してアクセスし、帰ってきたデータもプログラムで処理する

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

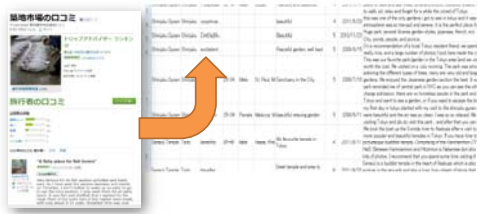
提供なし
- フォートラベル

例：Twitter APIからのデータ取得

- 二通りのAPI
  - 過去二週間分を検索
  - 流れてくるものを逐次取得
- 問題点
  - 位置情報が着いているものが少ない（0.2～0.6%）
  - 不完全な文章
  - ボットの存在
  - 中国人の投稿が少ない

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.



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

プログラムを書いて以下の処理を自動的に行う

1. 情報を入手したいページのURLを生成
2. 各ページのデータ（htmlコード）を入手
3. キーワードの前後に必ず出現する文字列を手がかりに各キーワードを抜き出す
4. 抜いたキーワードを表に書き込んでいく

実際には、誰か他の人の書いたプログラムをなんとなく書き換えるだけ

最近の野望

論点2：データの精選

なぜデータクレンジングが必要か？

- キーワード検索の限界
- ボットを除外したい
- 他国語のものを除外したい
- 観光客以外のものを除外したい

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

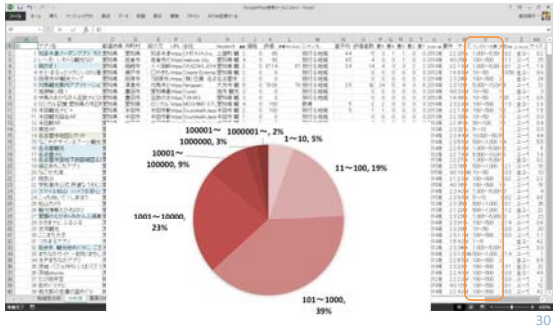
論点3：データの分析

集めた大量のデータをどうするの？

- 集計する
- 位置を含むデータの場合
  →地図化（点分布 or サーフェス）
- 自由記述を含むデータの場合
  →テキストマイニング

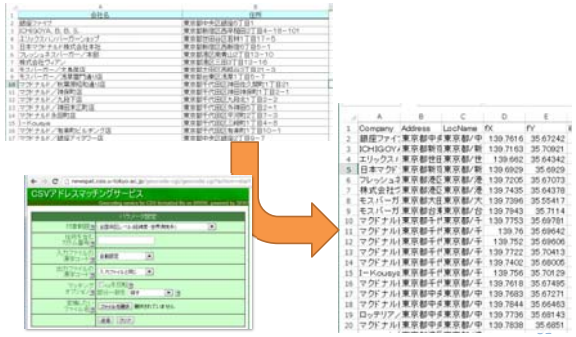
大量データの処理法

①集計



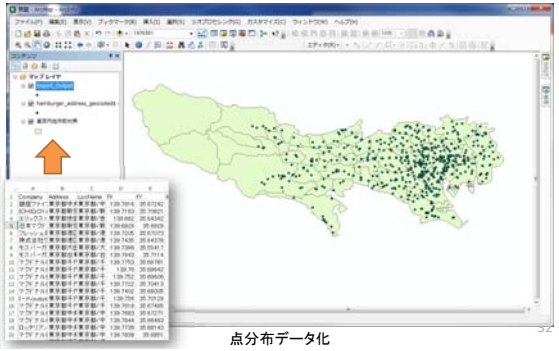
大量データの処理法

②地図化



大量データの処理法

②地図化

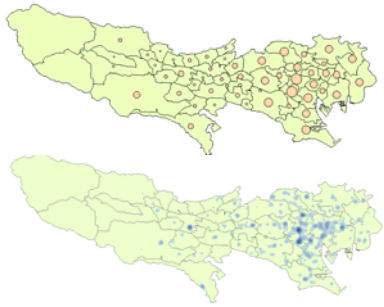


点分布データ化

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



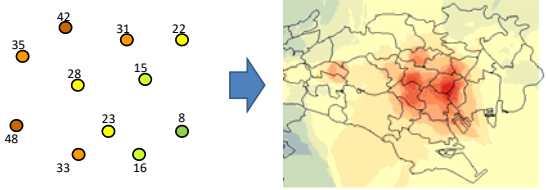
地区別集計

カーネル密度図

大量データの処理法

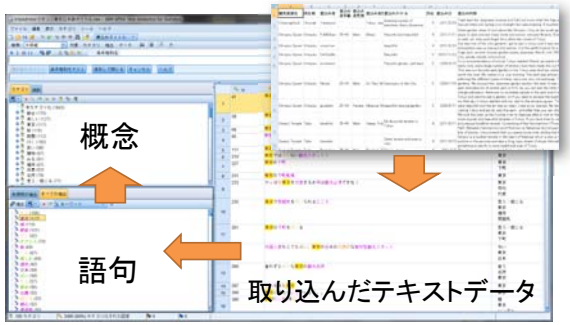
②地図化

位置＋数値の場合→サーフェス分析



大量データの処理法

③テキストマイニング



SPSS Text Analytics for Surveys

大量データの処理法

③テキストマイニング

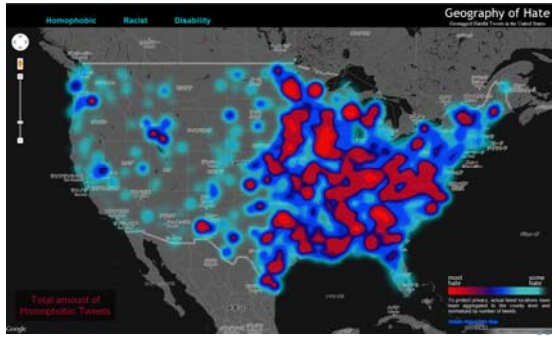
観光クチコミに使われる概念

good (275)	would recommend (52)	english (38)	fast (20)
place (160)	people (51)	photos (37)	displays (19)
tokyo (125)	right (51)	tour (37)	difficult (19)
large (107)	bad (48)	child (35)	friends (18)
shopping (105)	new (47)	busy (33)	long (18)
walk (74)	beautiful (46)	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 (58)	temple&shrine (40)	famous (22)	fish (15)
museum (55)	transportation (39)	hotel (21)	space (15)
like (52)		available (20)	

テキストマイニング × 多変量解析

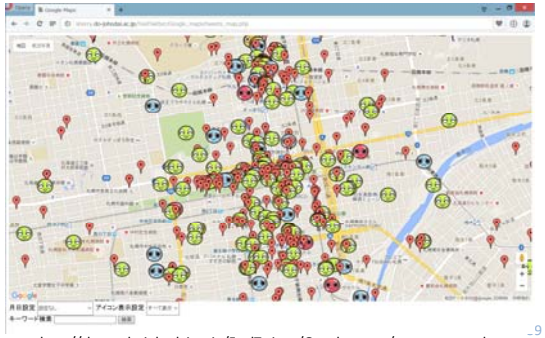
数量化Ⅰ類による評点との関連性分析

テキストマイニング × 地図化



テキストマイニング × 地図化

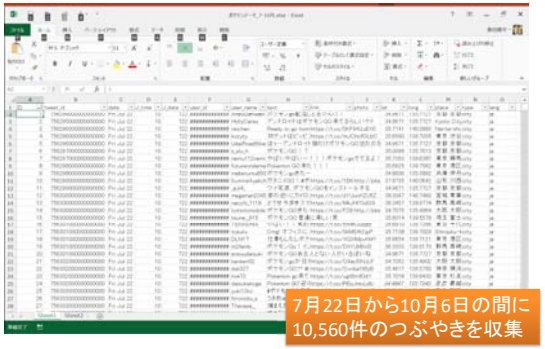
（斎藤 一，横川 祥司 2013）



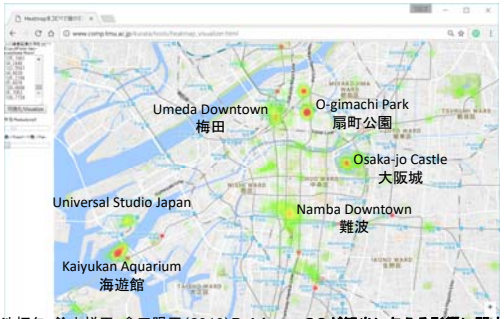
http://sherry.do-johodai.ac.jp/FeelTwitter/Google_maps/tweets_map.php

ビッグデータ分析の実例
ーポケモンGOマップー

ポケモンGO関連のツイートを収集



ポケモンGOマップ



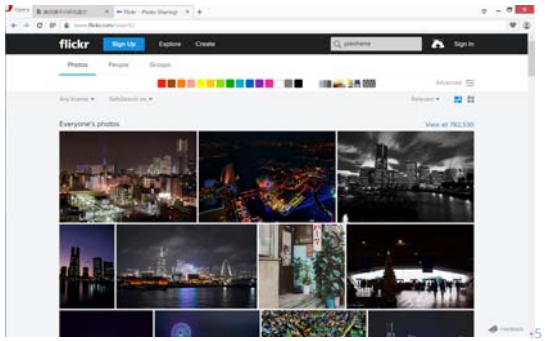
小池拓矢・鈴木祥平・倉田陽平（2016）**Pokémon GOが観光に与える影響に関する一考察ーTwitterの位置情報付きツイートを用いて**.地理学会2016年秋季学術大会

ビッグデータ分析の実例
ー観光ポテンシャルマップ研究ー

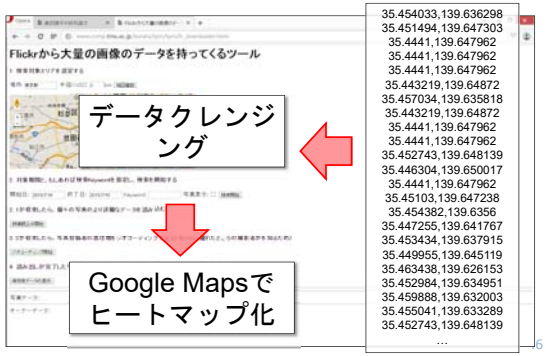
きっかけ：杉本君の研究(2010)

杉本興運(2010)：観光行動としての写真撮影に着目した景観資源の探索と「みどころ」の抽出　-都立石神井公園を事例として-. 地理情報システム学会講演論文集, 19

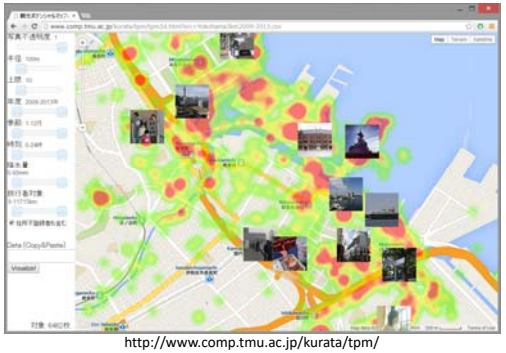
写真共有サイト「flickr」のデータを使うことを思いつく



flickrからのデータ抽出ツールを開発

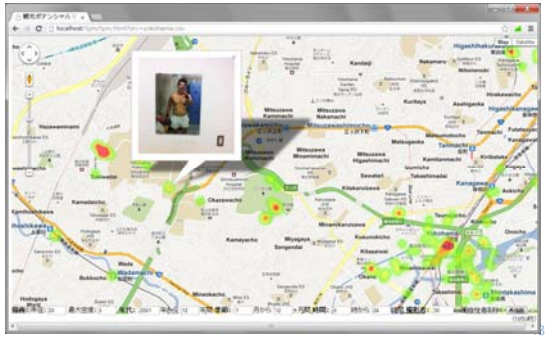


観光ポテンシャルマップ



http://www.comp.tmu.ac.jp/kurata/tpm/

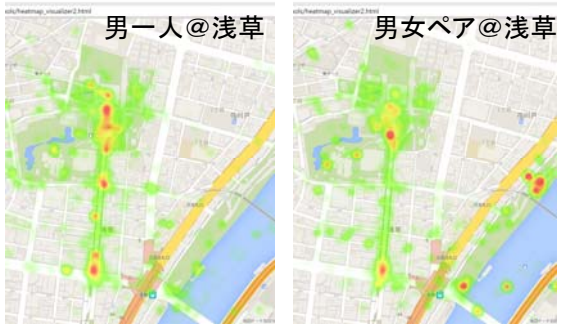
データクレンジングの必要性



観光ポテンシャルマップ×街路データ

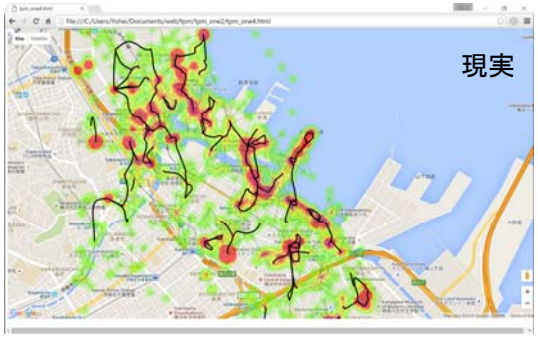
倉田陽平・相尚寿・真田風(2015) **写真共有サイト投稿データを利用した新たな観光マップの構築**. 観光科学研究, 8,

観光ポテンシャルマップ×顔認識



倉田陽平・鞠山彩実・石川博(2016) **Flickr画像を用いた観光空間内の写真撮影行動の来訪者類型別比較ー浅草を例としてー**. DEIM2016

観光ポテンシャルマップ×尾根線抽出



倉田陽平(2016) **大量写真データをもとにした観光地内の主要観光ルートの自動抽出の試みードロネー三角網の利用による再考ー**. 観光情報学会第13回全国大会

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