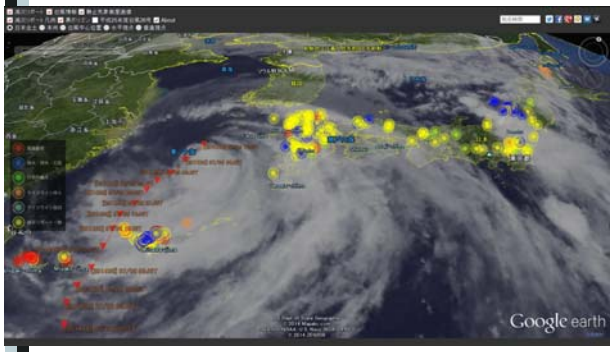
観光地理情報学 2017年度第13回

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



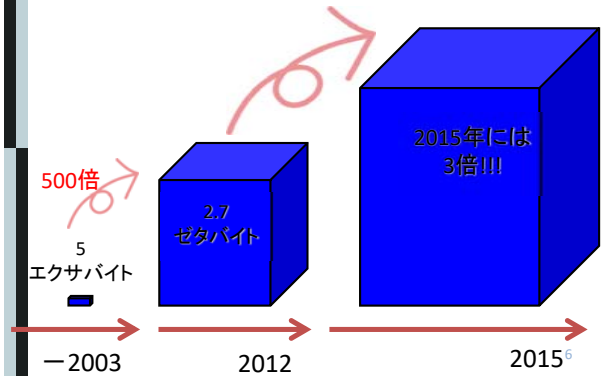
http://safetymap.jp

ビッグデータとは？

- SNS書き込み
- 移動（カーナビ・IC乗車券）
- 検索履歴
- 電子商取引
- 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



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
- 狩りをする（ネットから取得する）
  - –APIの利用 合法
  - –スクレイピング グレー

APIを利用したデータ取得とは

特定のアドレスにパラメータをつけてアクセスすると、人間が見る用のwebページの代わりに、データそのものが帰ってくる

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

→上のようなアドレスを機械的に生成してアクセスし、帰ってきたデータを機械的に記録していく

観光関連サイトのAPI提供状況

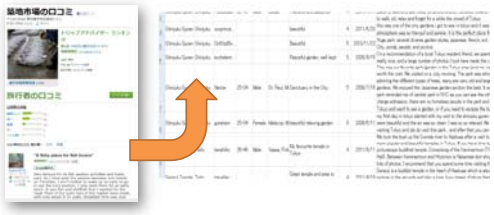
無償
- Flickr
- twitter
- Facebook
- Instagram
- 楽天トラベル
- ぐるなび
- Expedia

有償（企業のみ）
- TripAdvisor

提供なし
- フォートラベル

スクレイピング

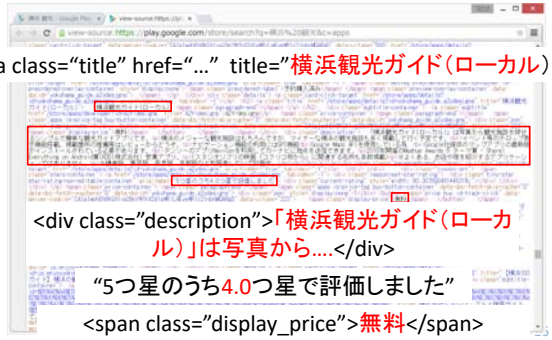
「人間が閲覧する用のwebページ」を巡回し、テキスト解析によって必要とするデータを機械的に集めてくること。「クローリング」ともいう.



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

Excel上でプログラム（VBA）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アドレスを生成
2. 各ページの「表示変換前のソース」（htmlドキュメント）を入手
3. 必要とするデータの前後に必ず出現する文字列を手がかりに、欲しいデータを抜き出す
4. 抜いたデータを順次、表に書き込んでいく

難しそうに聞こえるかも知れませんが、実際には他人のプログラムを改造するだけ

最近の野望

論点2：データの精選

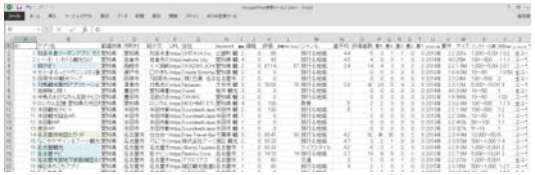
なぜデータクレンジングが必要か？

- キーワード検索の限界
- ボットが参加している
- 他国語が混じっている
- 観光以外のデータが混じっている

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

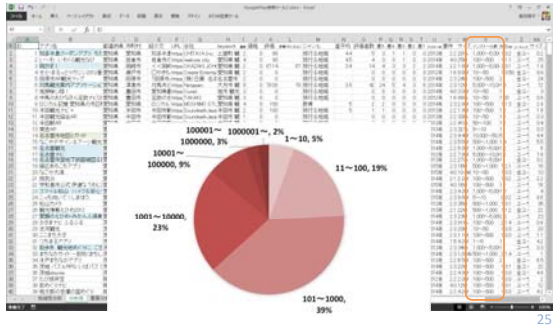


論点3：データの分析

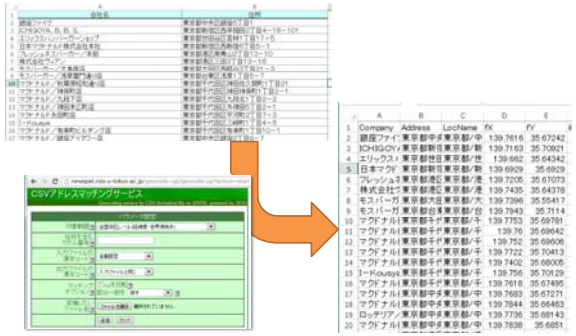
集めた大量のデータをどうするの？

- 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

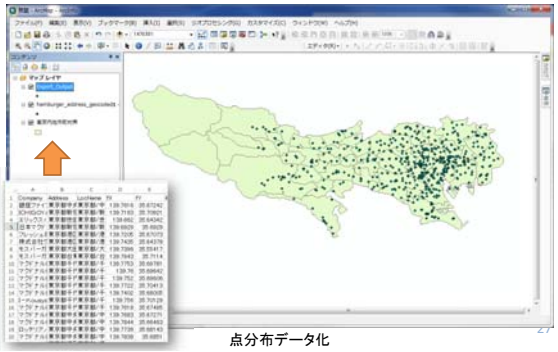
大量データの処理法
①集計



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

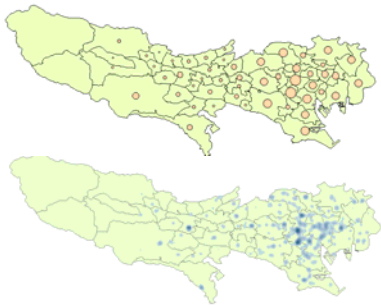


点分布データ化

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



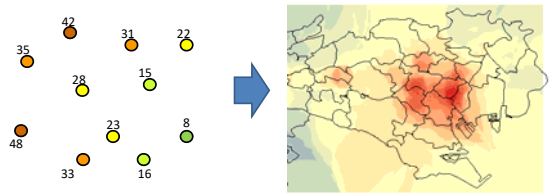
地区別集計

カーネル密度図

大量データの処理法

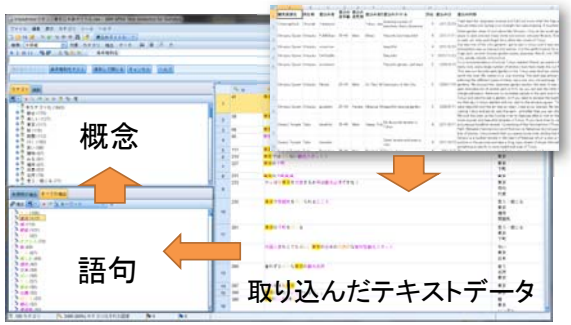
②地図化

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

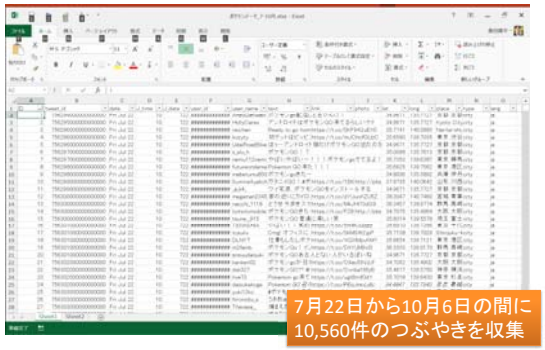
③テキストマイニング

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)	

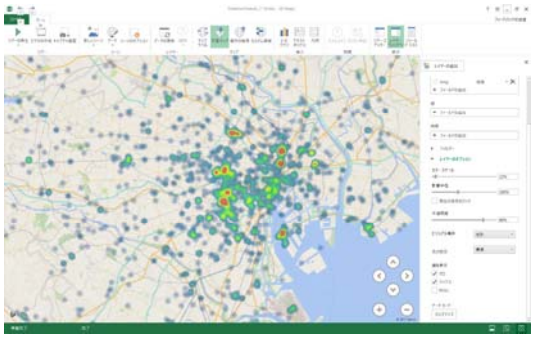
ビッグデータ分析の実例Ⅰ
ーポケモンGOマップー

ポケモンGO関連のツイートを収集

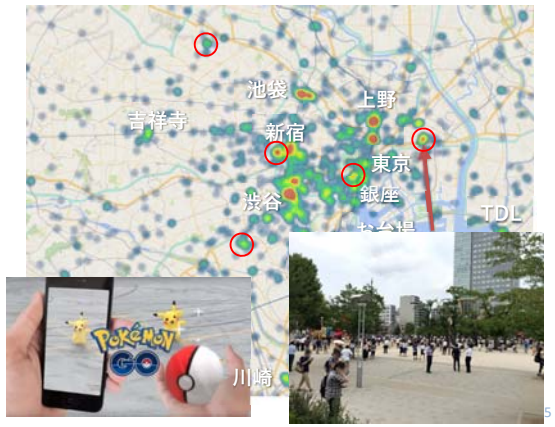


7月22日から10月6日の間に
10,560件のつぶやきを収集

ポケモンGOツイートの地図化

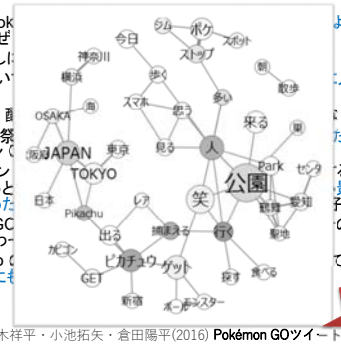


鈴木祥平・小池拓矢・倉田陽平(2016) Pokémon GOツイートから
読みとる人の動きと気づき. 観光情報学会第14回研究発表会



ポケモンGOツイート中の
頻出語句とその共起関係

- 今話題の pok… よりもよい景色を get だぜ
- ポケモン探し…
- お家から歩い… 入りスポット発見
- ポケモン go …
- 江戸崎祇園祭… たかもしれない。ポケモン …
- 結局ポケモン… る日々。(笑) だけどちょっと… 景色や変わってしまった… 子屋
- Pokemon GO… の存在も知らんかったわ…
- ポケモン Go … て遊ぶだけではなく勉強に…



世界最速の学術研究

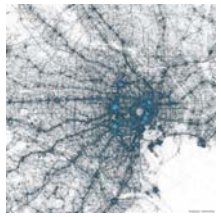
鈴木祥平・小池拓矢・倉田陽平(2016) Pokémon GOツイートから
読みとる人の動きと気づき. 観光情報学会第14回研究発表会

ビッグデータ分析の実例Ⅱ
ー観光ポテンシャルマップ研究ー

観光関連の分析を考えたときのツイッターの問題点

- 日常性が高い
- 位置情報つきが少ない(0.2~0.6%)
- 意味不明な文章が多い
- 中国人の投稿が少ない

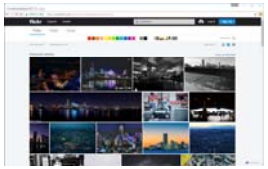
南大沢なう



我々は「Flickr」に注目

http://www.flickr.com

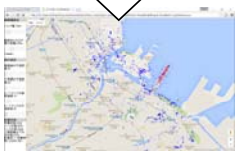
我々のアイデア
(2010年頃)



プログラムによる自動データ取得

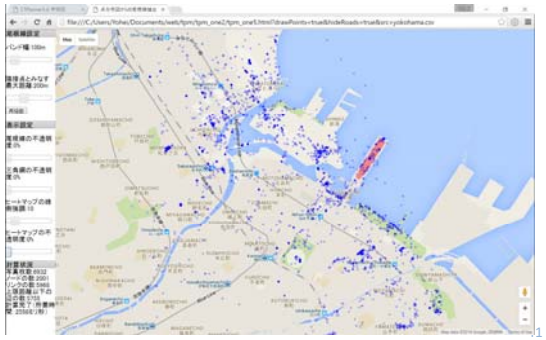
緯度経度　撮影者の居住地

地図に描く



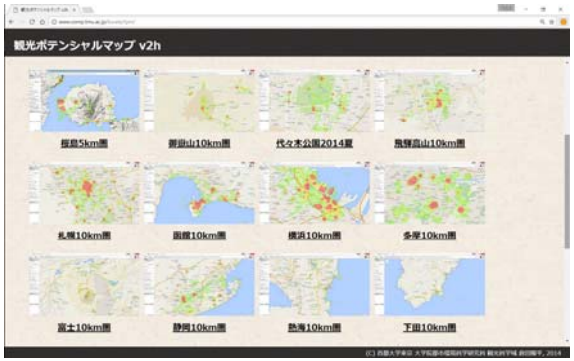
密度図化

写真投稿箇所(横浜)



写真投稿密度(横浜)

命名「観光ポテンシャルマップ」



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

応用:猫ポテンシャルマップ

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

猫ポテンシャルマップの作り方

Flickrの投稿写真を利用した観光研究は、他にもいっぱいある

- 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）
- 名所マップの自動作成（Chen *et al.* 2009）
- 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）

flickr

観光研究に使えるのはTwitterやFlickrだけでしょうか？

観光が生み出す大量のデータ

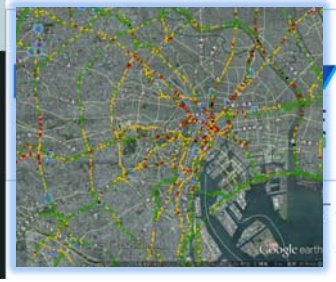
旅行前 → 旅行中 → 旅行後

興味	検索	予約	移動	共有
facebook YouTube	Google YAHOO! JAPAN NAVITIME	ANA 楽天トラベル		facebook twitter flickr

どの段階でも、人が何かをするたびにデータが生じる
そのデータをうまく使えば、何か生み出せるかも

観光が生み出す大量のデータ

旅行前 → 旅行中 → 旅行後



移動	共有
	facebook twitter flickr

るたびにデータが生じる
何か生み出せるかも

観光が生み出

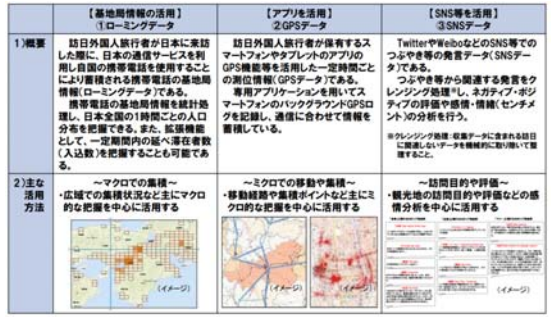
旅行前

興味	検索
facebook YouTube	Google YAHOO! JAPAN NAVITIME

flickr

どの段階でも、人が何かをするたびにデータが生じる
そのデータをうまく使えば、何か生み出せるかも

国（観光庁）も…



	【基地局情報の活用】①ローミングデータ	【アプリを活用】②GPSデータ	【SNS等を活用】③SNSデータ
1）概要	訪日外国人旅行者が日本に来訪した際に、日本の通信サービスを利用し自国の携帯電話を使用することにより蓄積される携帯電話の基地局情報（ローミングデータ）である。 携帯電話の基地局情報を統計処理し、日本全国の1時間ごとの人口分布を把握できる。また、拡張機能として、一定期間内の延べ滞在者数（入込数）を把握することも可能である。	訪日外国人旅行者が保有するスマートフォンやタブレットのアプリのGPS機能等を活用した一定時間ごとの測位情報（GPSデータ）である。 専用アプリケーションを用いてスマートフォンのバックグラウンドGPSログを記録し、通信に合わせて情報を蓄積している。	Twitterやweiboなどの SNS等でのつぶやき等の発言データ（SNSデータ）である。 つぶやき等から関連する発言をクレンジング処理※1し、ネガティブ・ポジティブの評価や感情・情緒（センチメント）の分析を行う。 ※1クレンジング処理：収集データに含まれる目的に関連しないデータを機械的に取り除いて管理すること。
2）主な活用方法	～マクロでの集積～ ・広域での集積状況など主にマクロ的な把握を中心に活用する	～ミクロでの移動や集積～ ・移動経路や集積ポイントなど主にミクロ的な把握を中心に活用する	～訪問目的や評価～ ・観光地の訪問目的や評価などの感情分析を中心に活用する

ＩＣＴを活用した訪日外国人観光動態調査 報告書（平成２７年度）

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

期末試験

- 8/4（金）13:00より11-209教室で
- 配点90点
  - 正誤10問×1点
  - 三拓20問×2点
  - 記述5問×6点
- 残り20点は課題提出状況