Detection, Classification and Visualization of Place-triggered Geotagged Tweets

Shinya Hiruta [1]
Takuro Yonezawa [1]
Marko Jurmu [1,2]
Hideyuki Tokuda [1]

[1] Keio University, [2] University of Oulu

hide tokuda lab.

Background: Real World Event Detection with Location-Based Social Networks

- **Real world event**

  Structured as a collection of descriptive attributes

  ‣ e.g. Place, Time, Content, ...
    - "**Baseball game** will be held at **PNC park** from **6:00 PM**"

  However, attributes are often dynamic
    - e.g. Baseball game that gets postponed because of rain
    - e.g. A traffic accident occurring on a way and causing traffic congestion

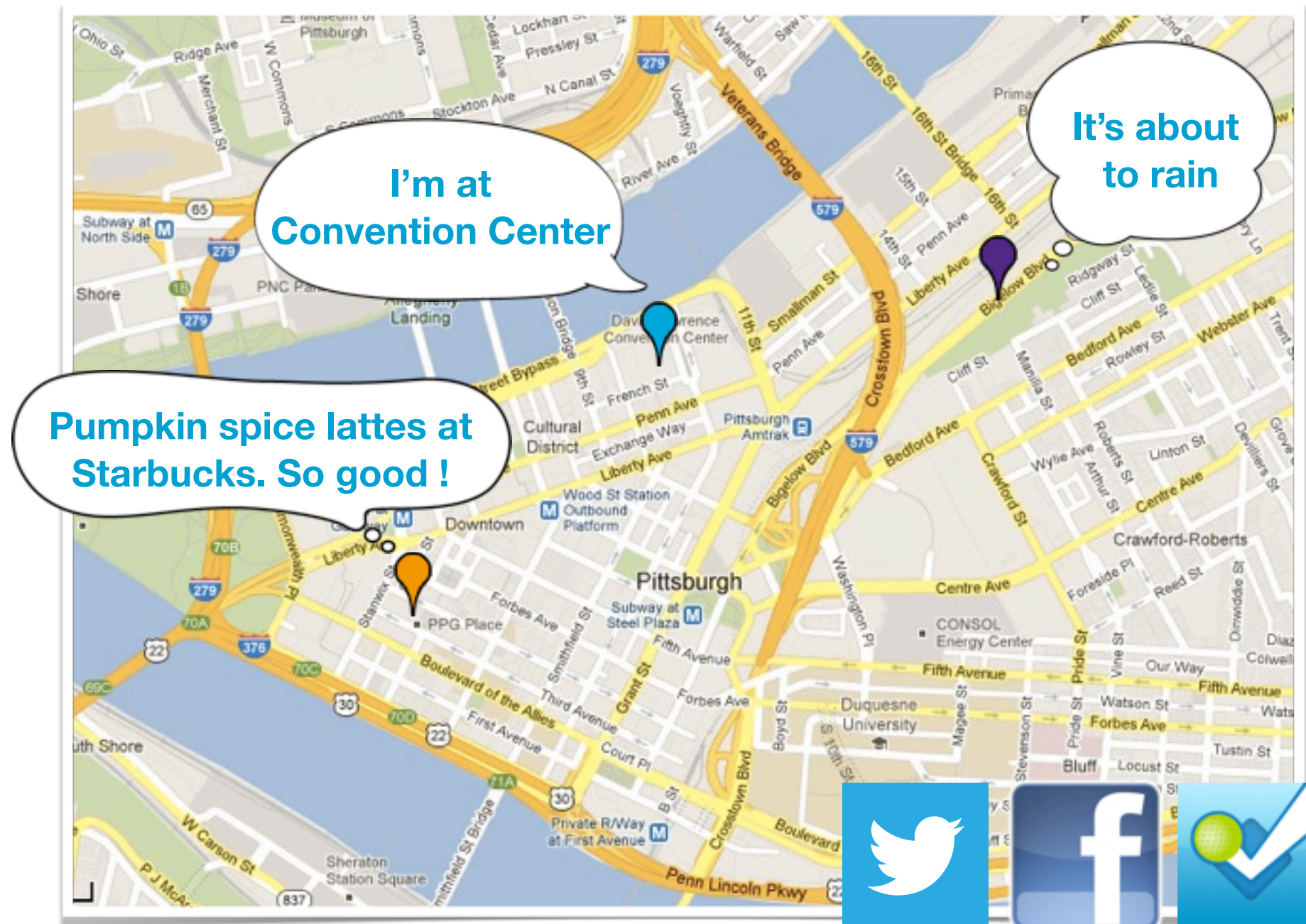 LBSN are suitable for extraction of dynamic information

hide tokuda lab.

Motivation:
Geotagged tweets are not always useful for real world event detection!
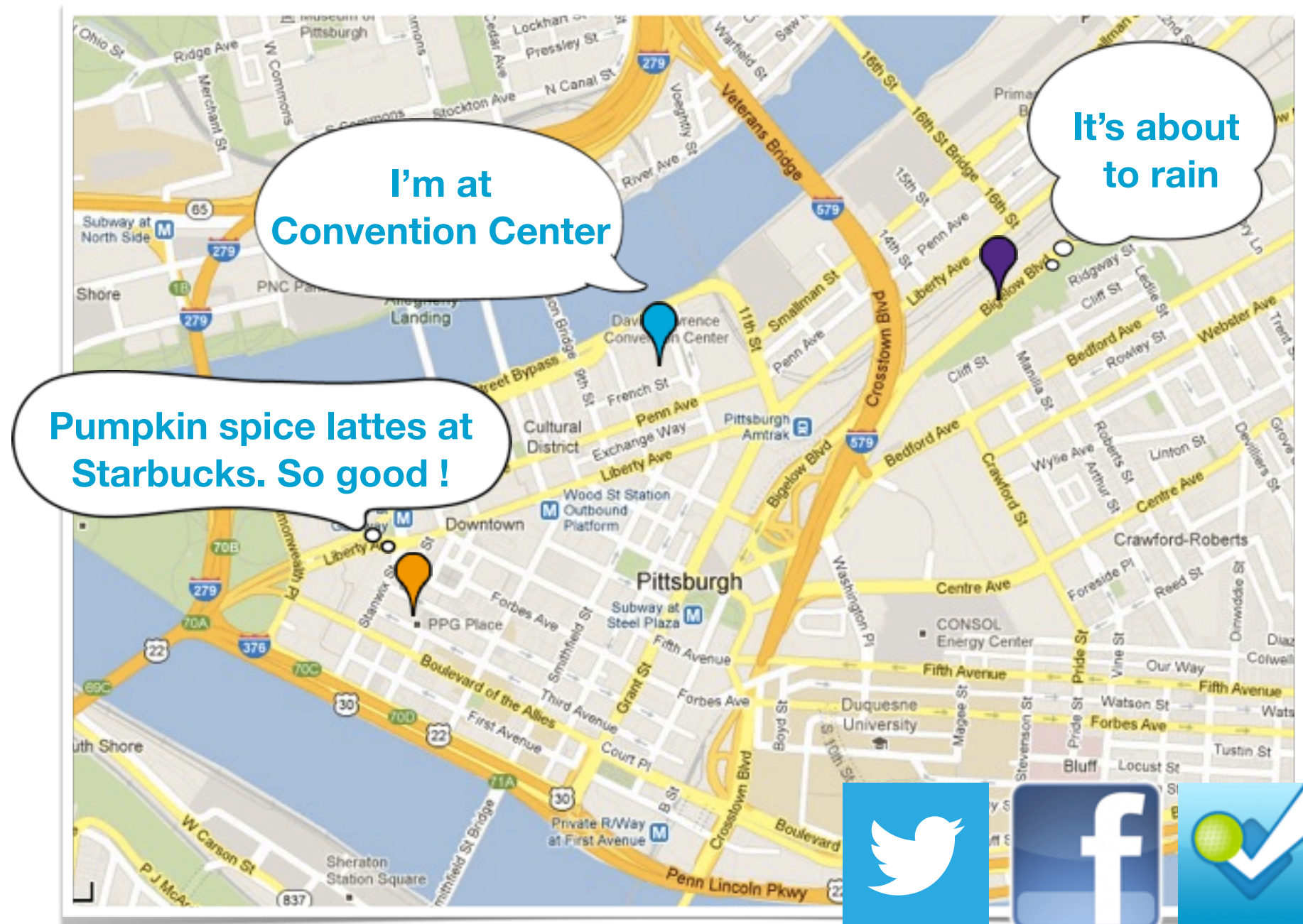
hide tokuda lab.

Motivation:
Geotagged tweets are not always useful for real world event detection!

Motivation:
Geotagged tweets are not always useful for real world event detection!

Motivation:
Geotagged tweets are not always useful for real world event detection!

Motivation:
Geotagged tweets are not always useful for real world event detection!
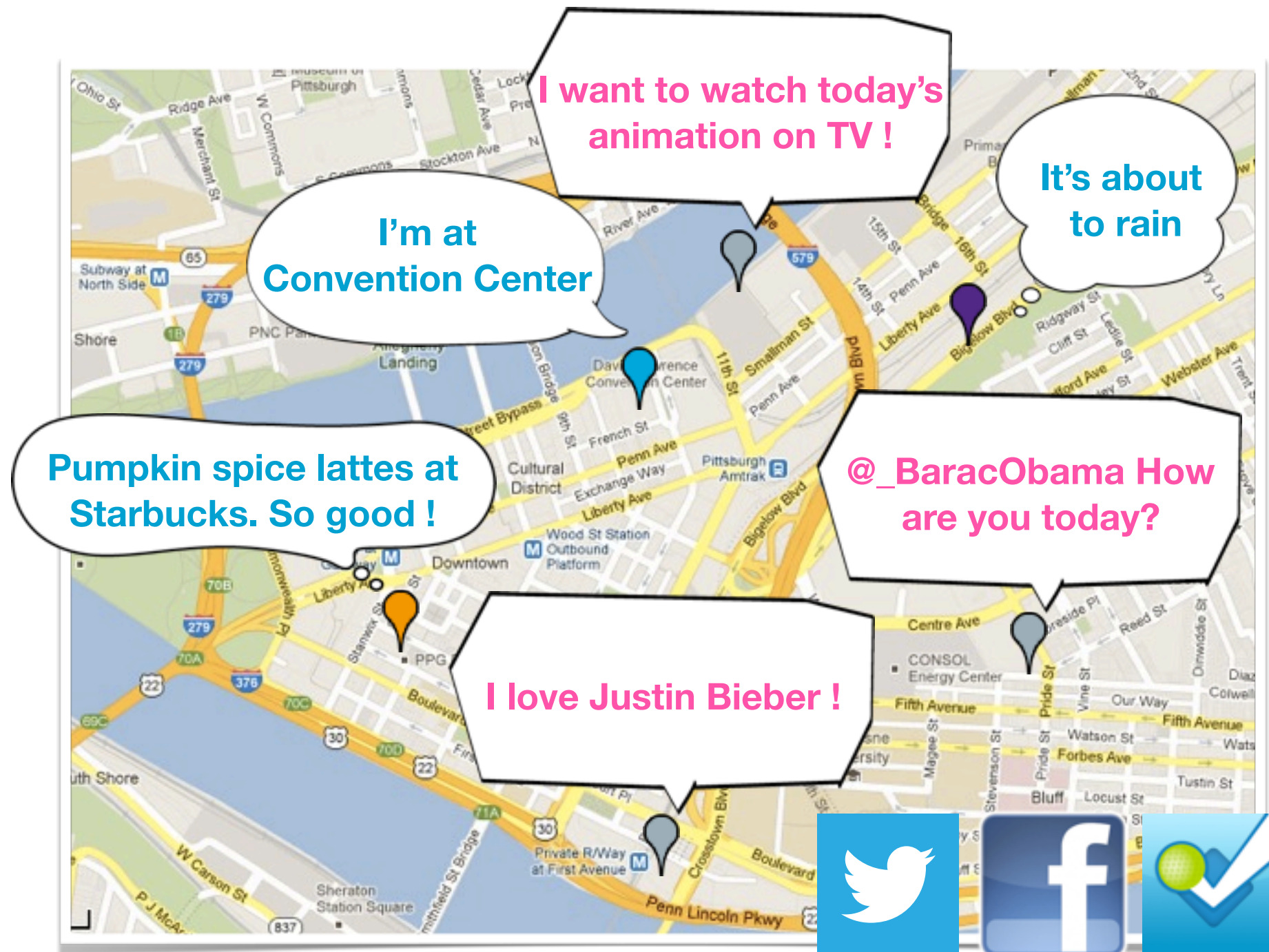
Place-triggered Geotagged Tweets

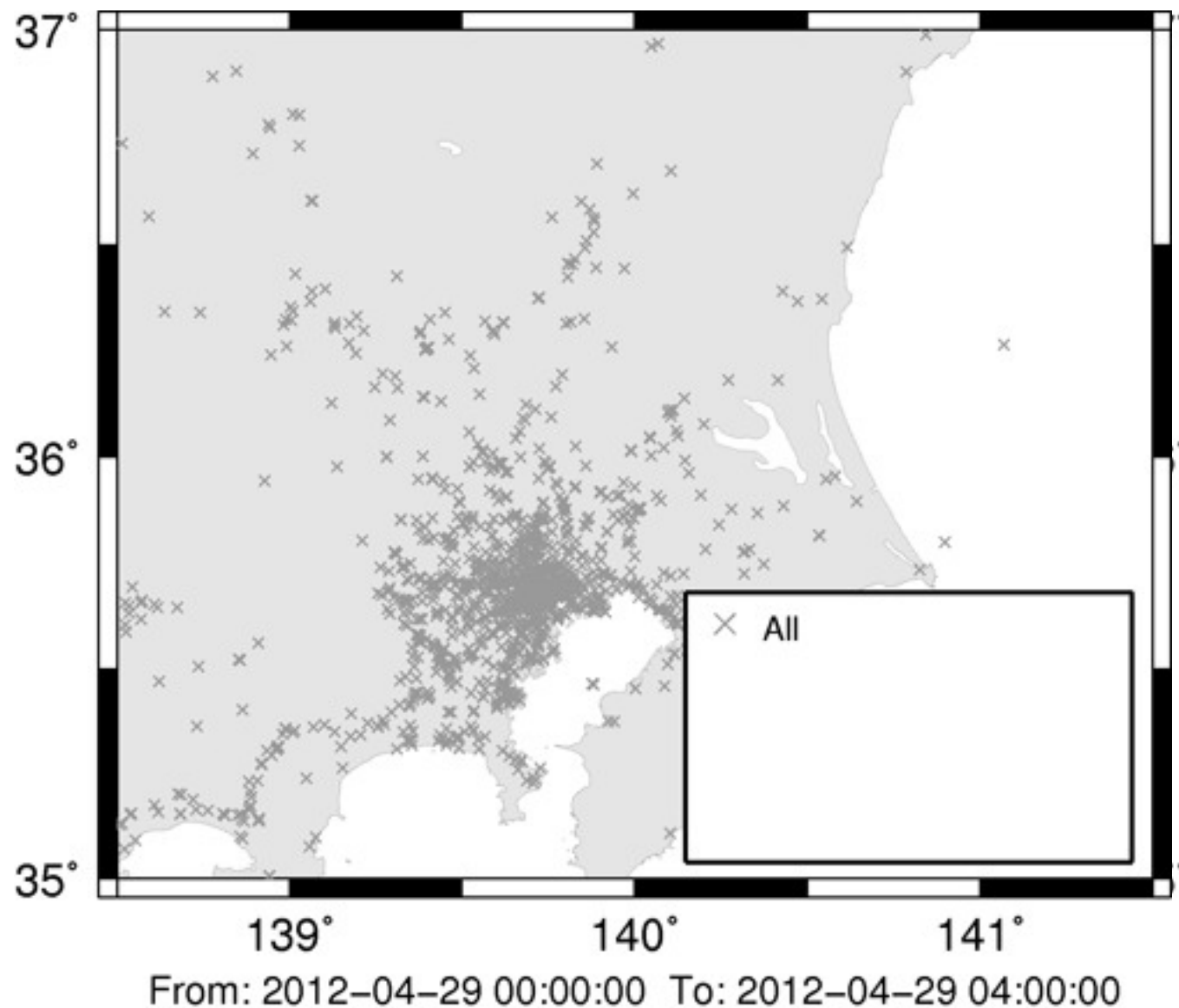- **Definition**

  Tweets that have both:
  - ‣ Geotag metadata
  - ‣ Content relevant to the associated location

- **Research Goal**
  - ‣ Detection
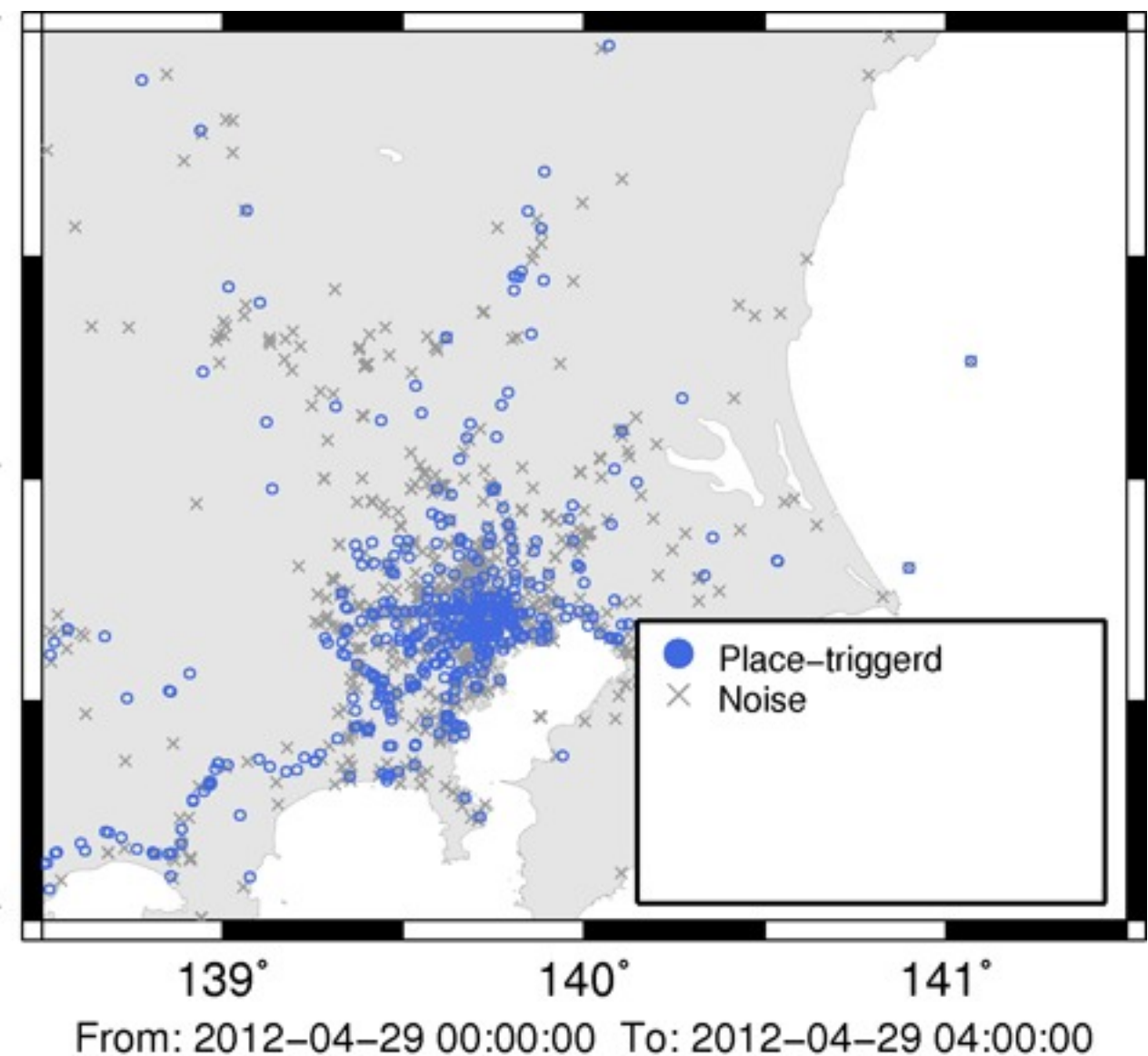 - ‣ Classification
  - ‣ Application

Detecting Place-triggered Geotagged Tweets

Without our system

With our system

hide tokuda lab.

Detecting Place-triggered Geotagged Tweets

Without our system

With our system

hide tokuda lab.

Detecting Place-triggered Geotagged Tweets

Without our system

With our system

hide tokuda lab.

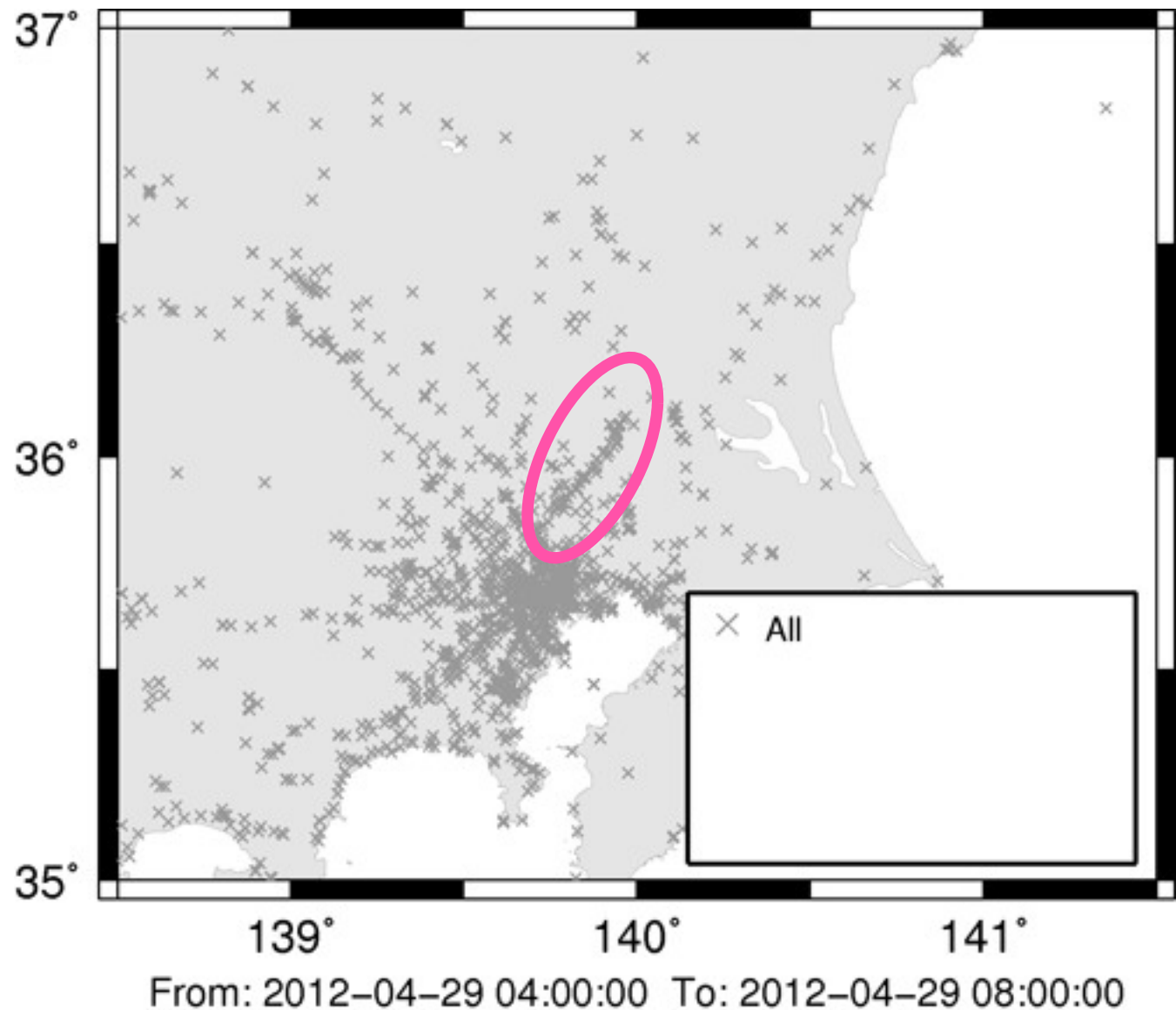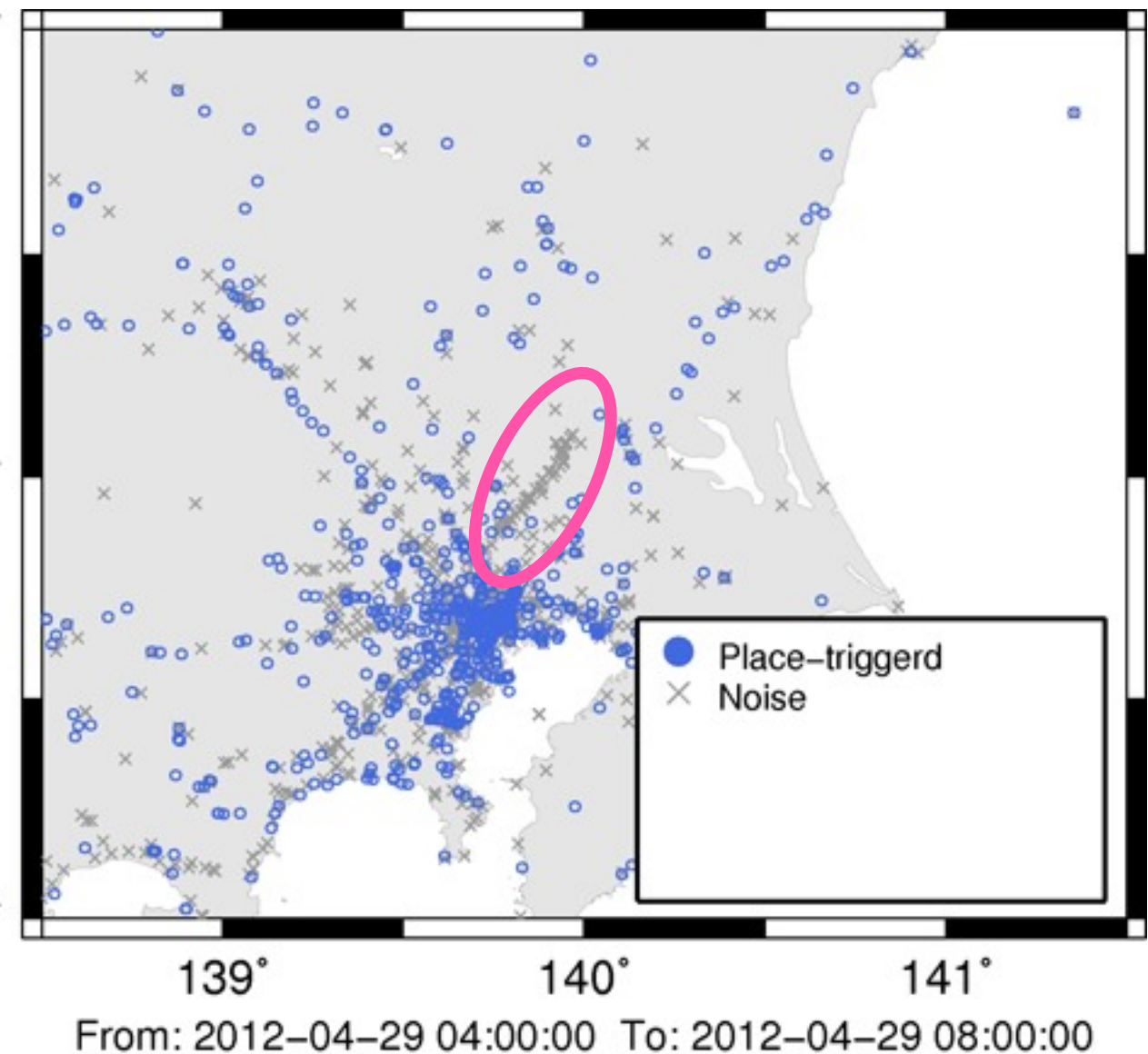# Detecting Place-triggered Geotagged Tweets
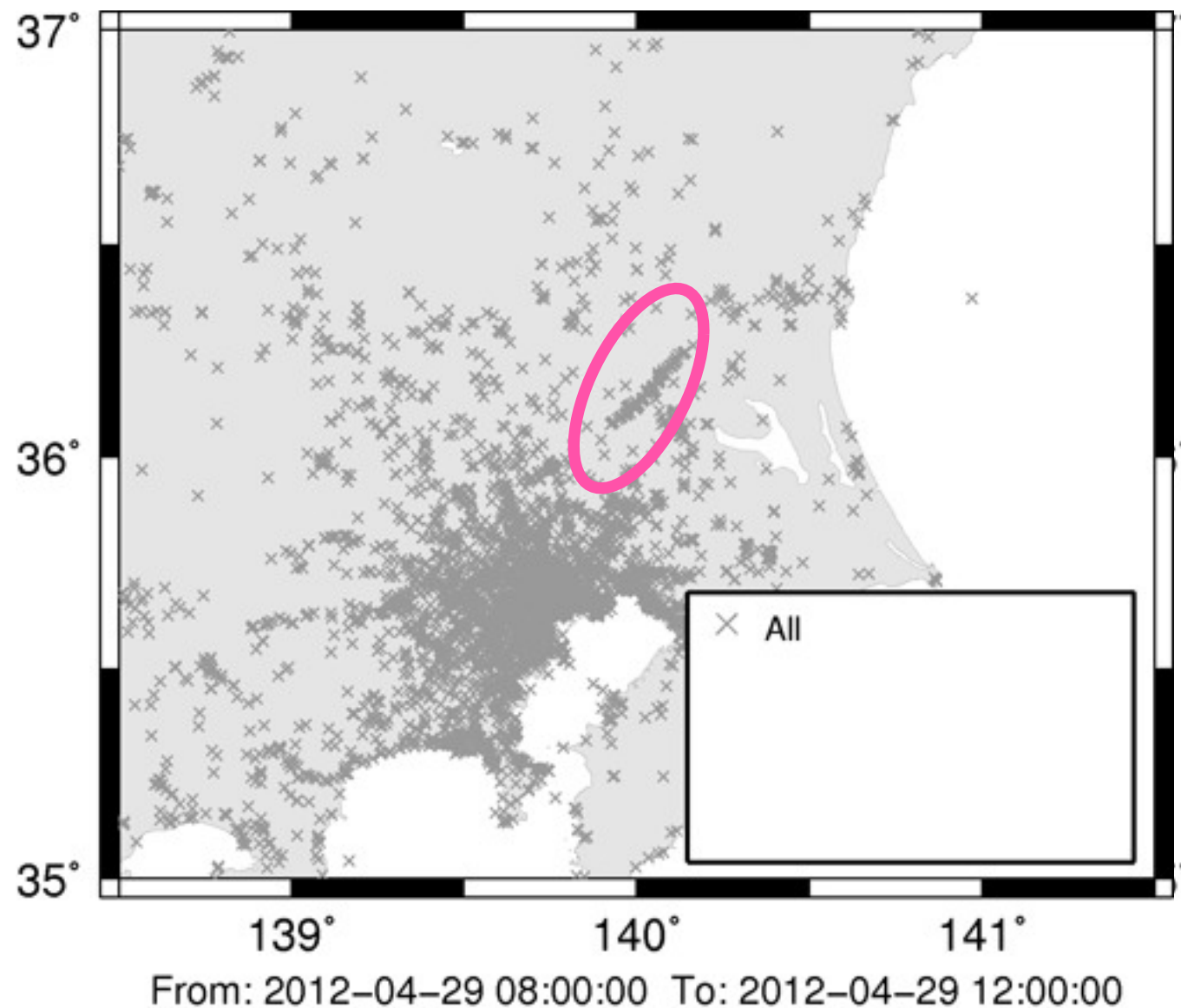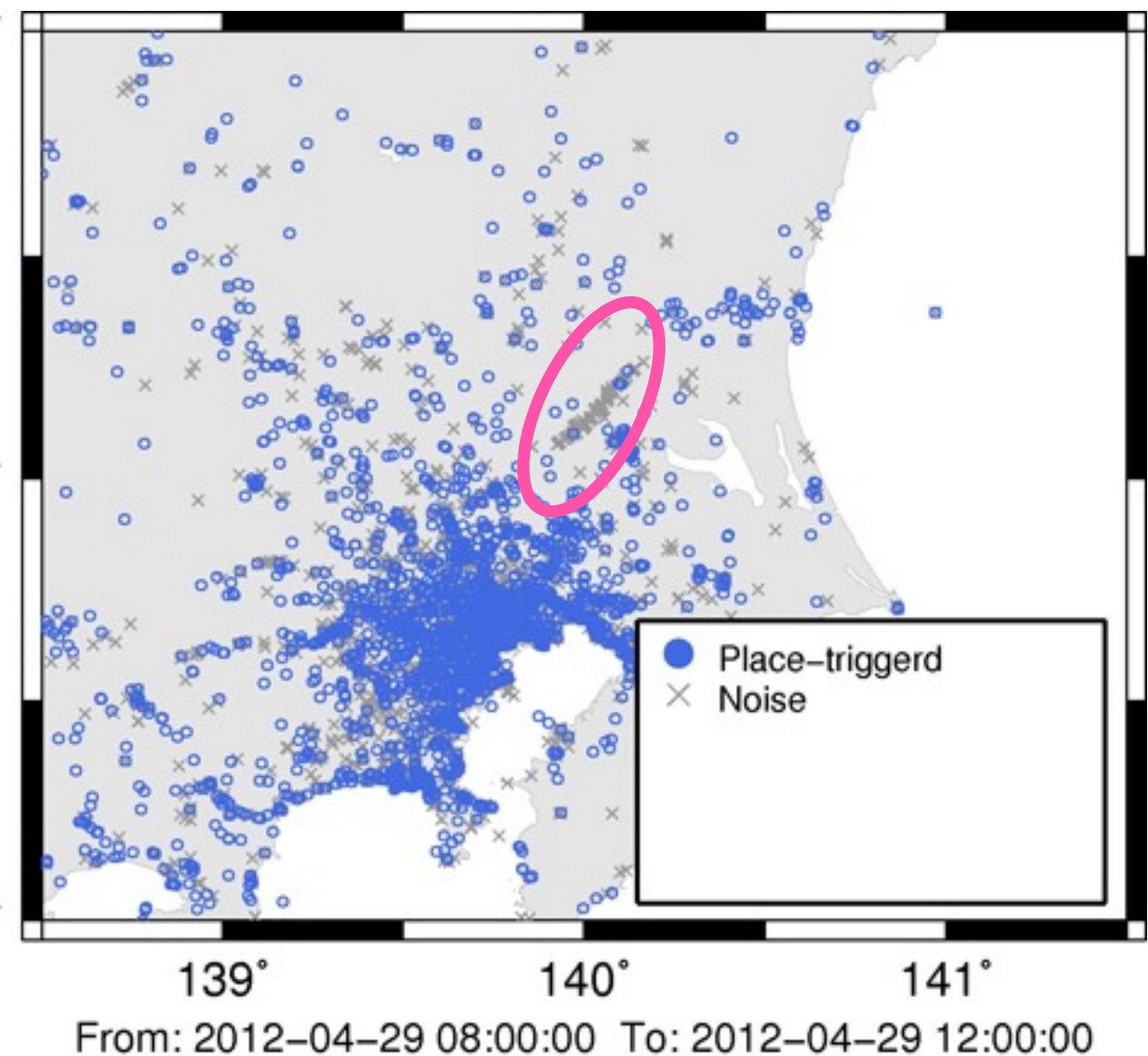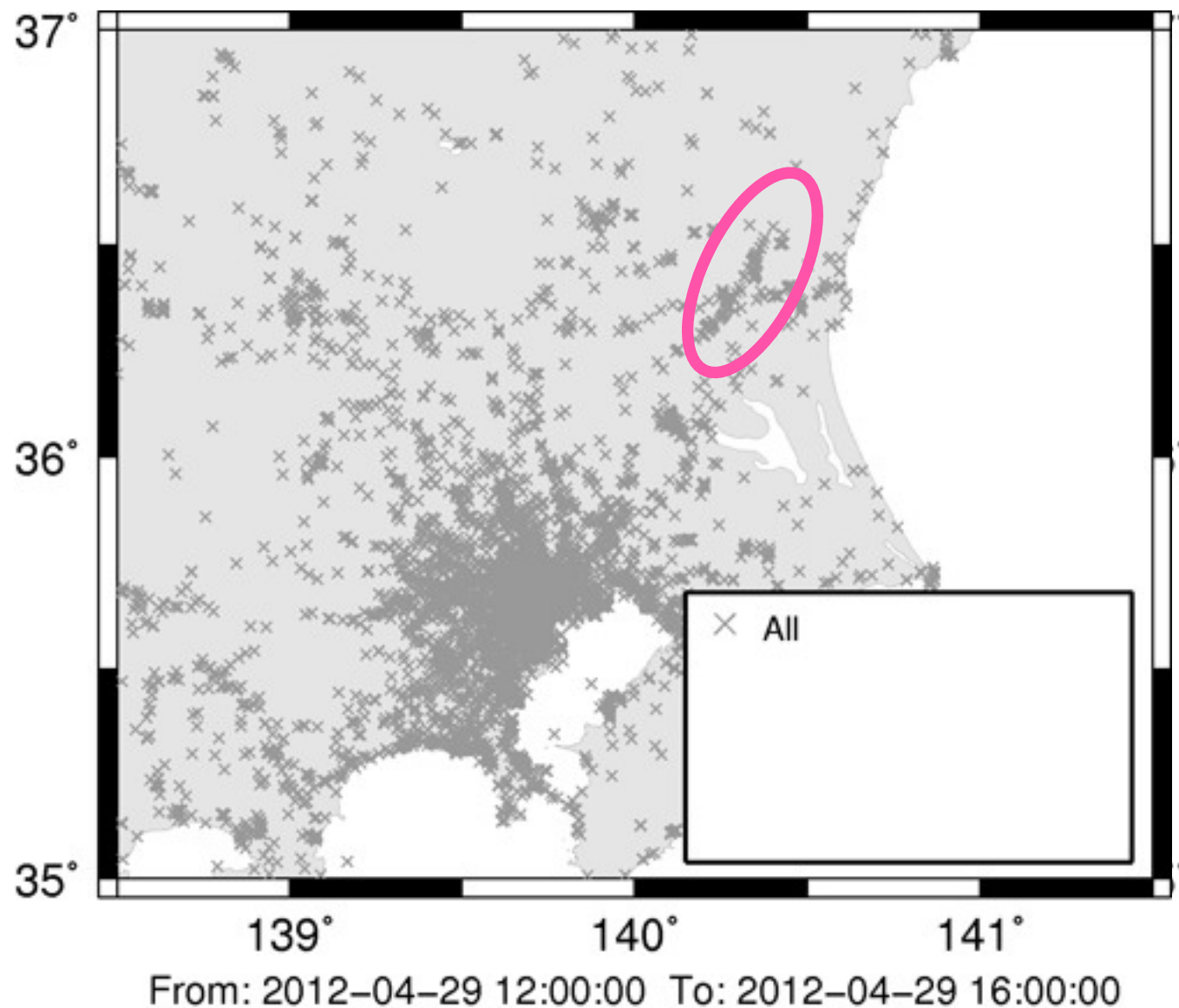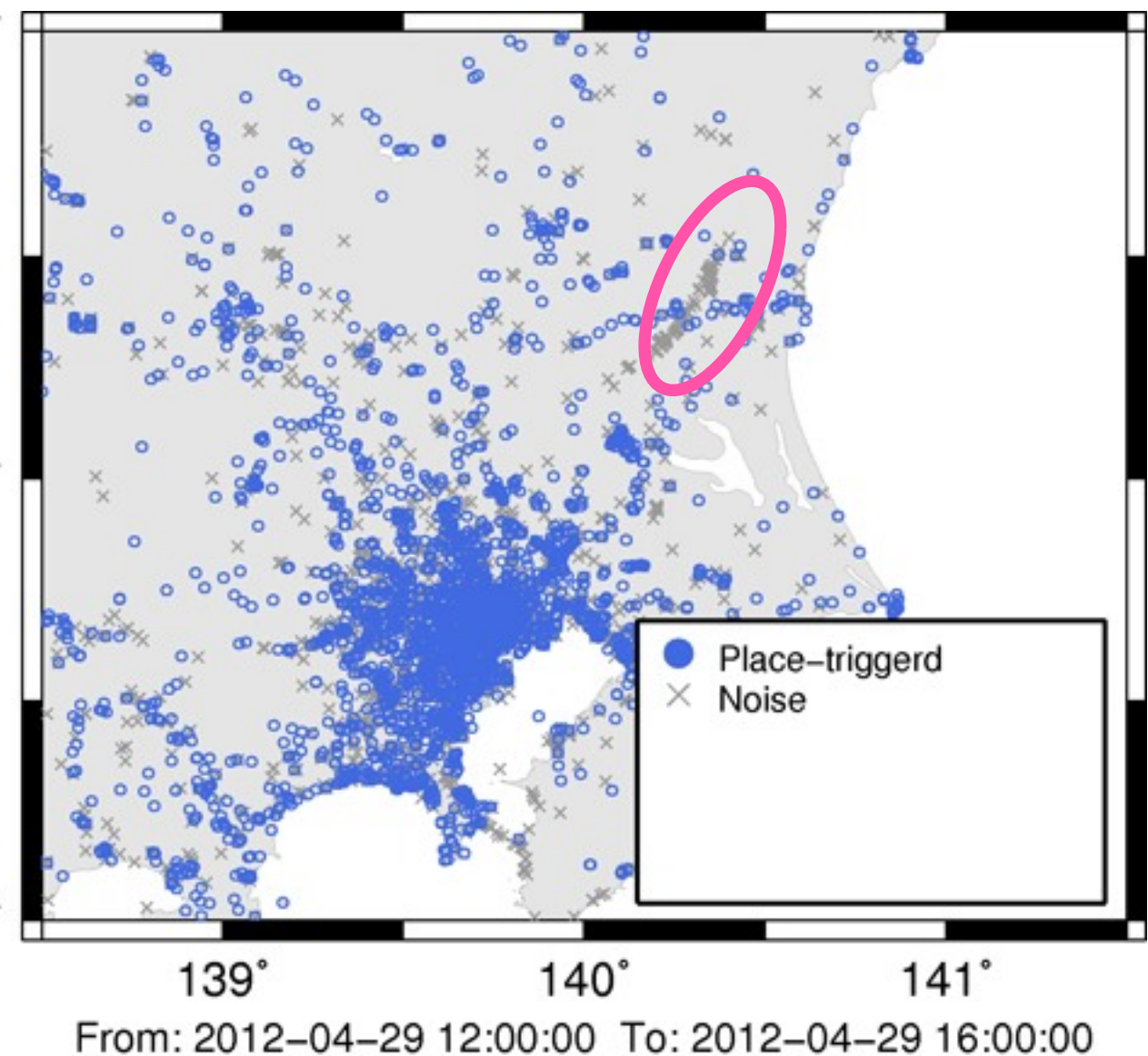
Without our system

With our system

Detecting Place-triggered Geotagged Tweets
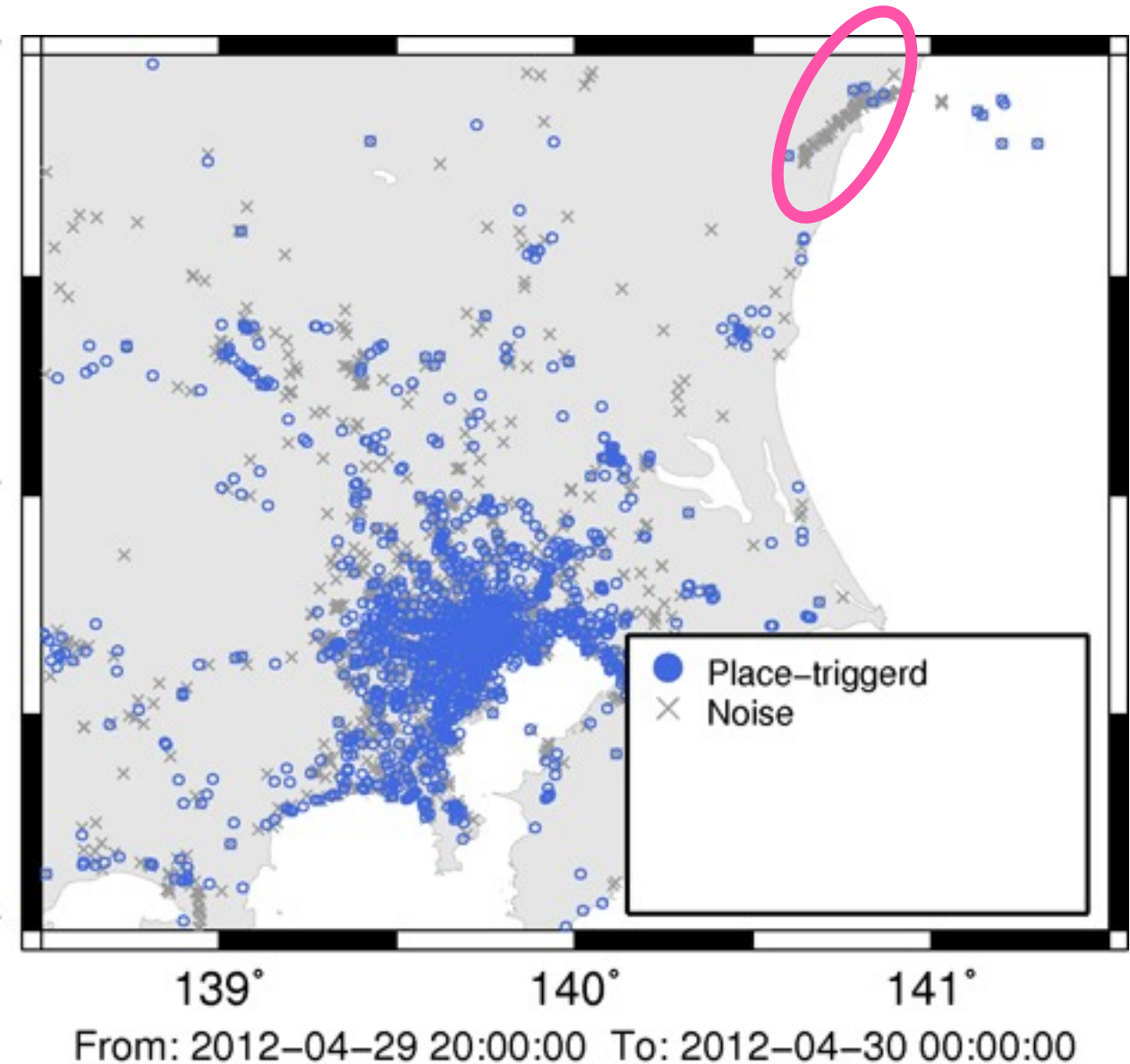
Without our system

With our system

hide tokuda lab.

Detecting Place-triggered Geotagged Tweets

Without our system

With our system

hide tokuda lab.

Detecting Place-triggered Geotagged Tweets

Without our system

With our system

Noise

hide tokuda lab.
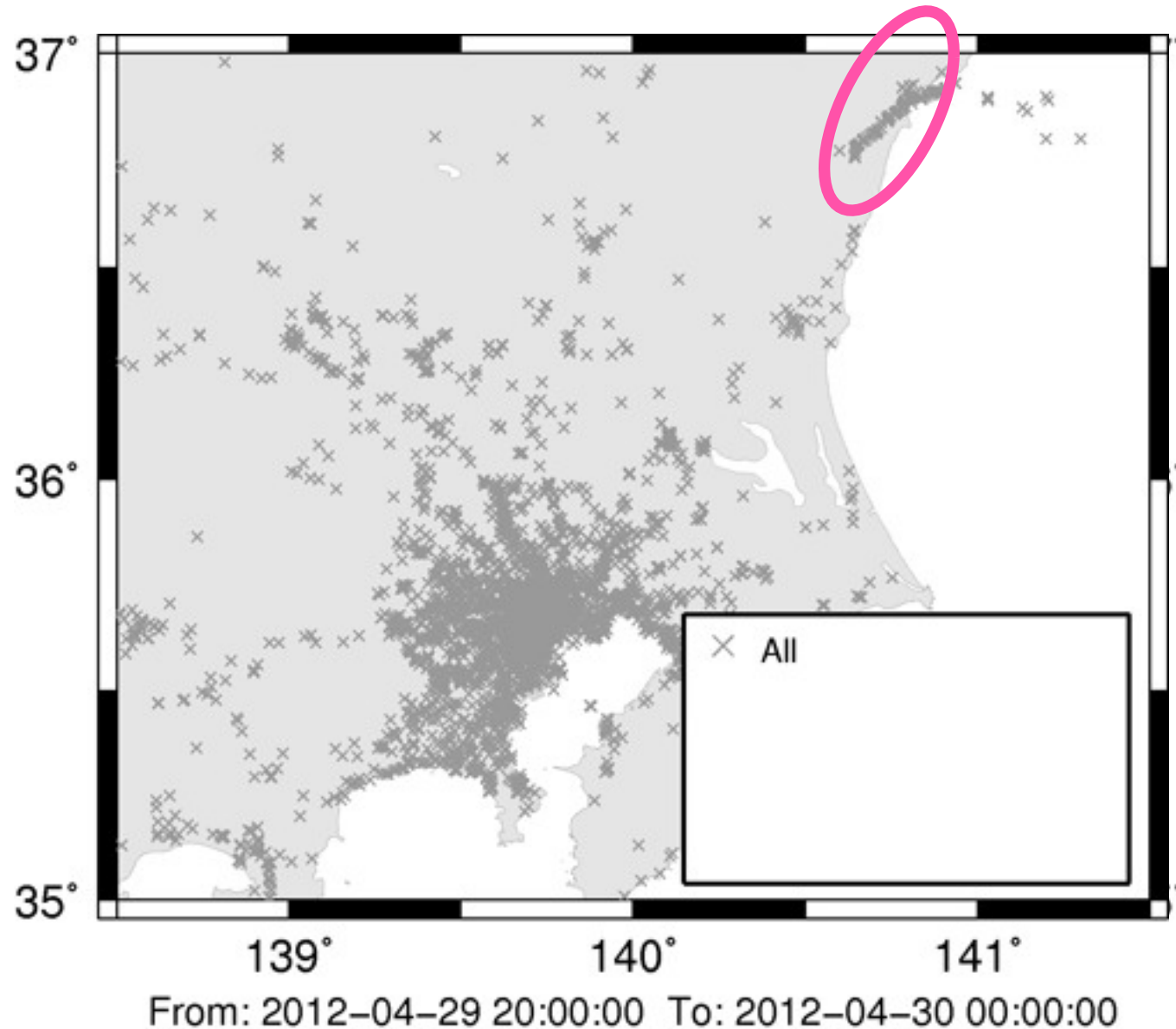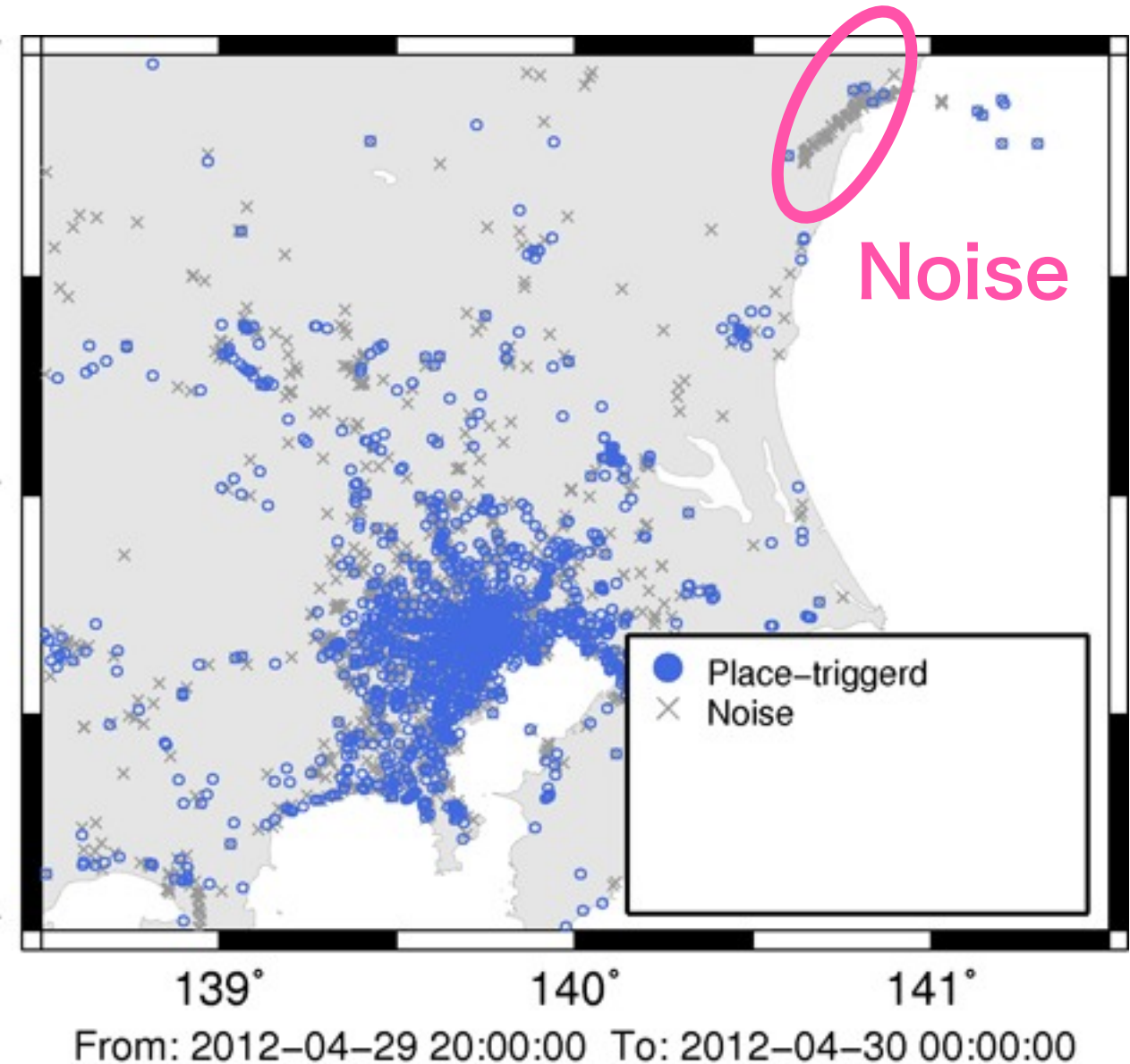
Related Work

- **Earthquake shakes twitter users: Real-time event detection by social sensors.**

  T. Sakaki, M. Okazaki, and Y. Matsuo.

  ‣ In Proceedings of the 19th International Conference on World Wide Web, pages 851–860, 2010.

- **Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection.**

  R. Lee and K. Sumiya.

  ‣ In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pages 1–10, 2010.

Comparison with Related Work
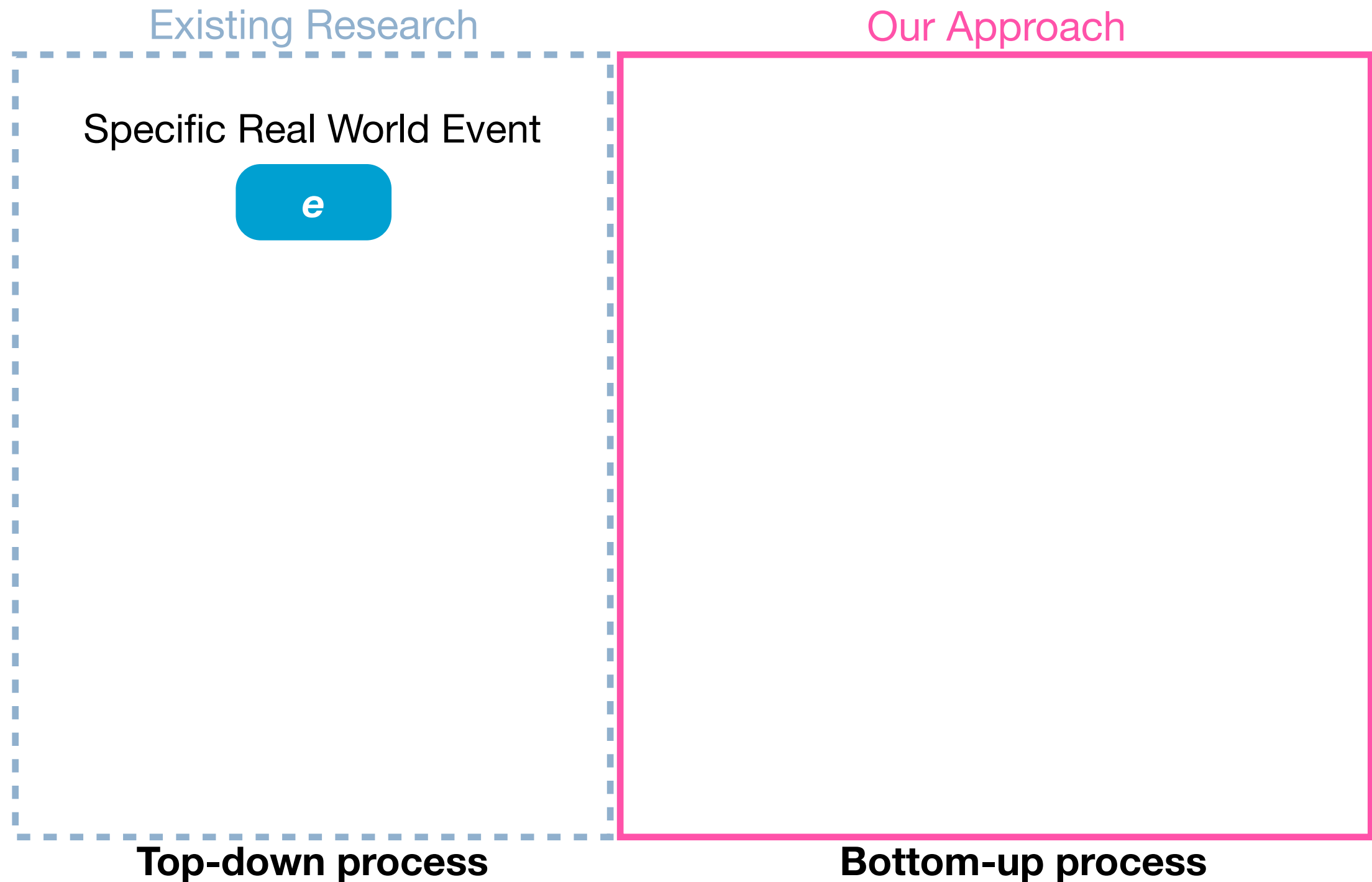
Specific Real World Event

e

Top-down process

Bottom-up process

hide tokuda lab.

12

Comparison with Related Work
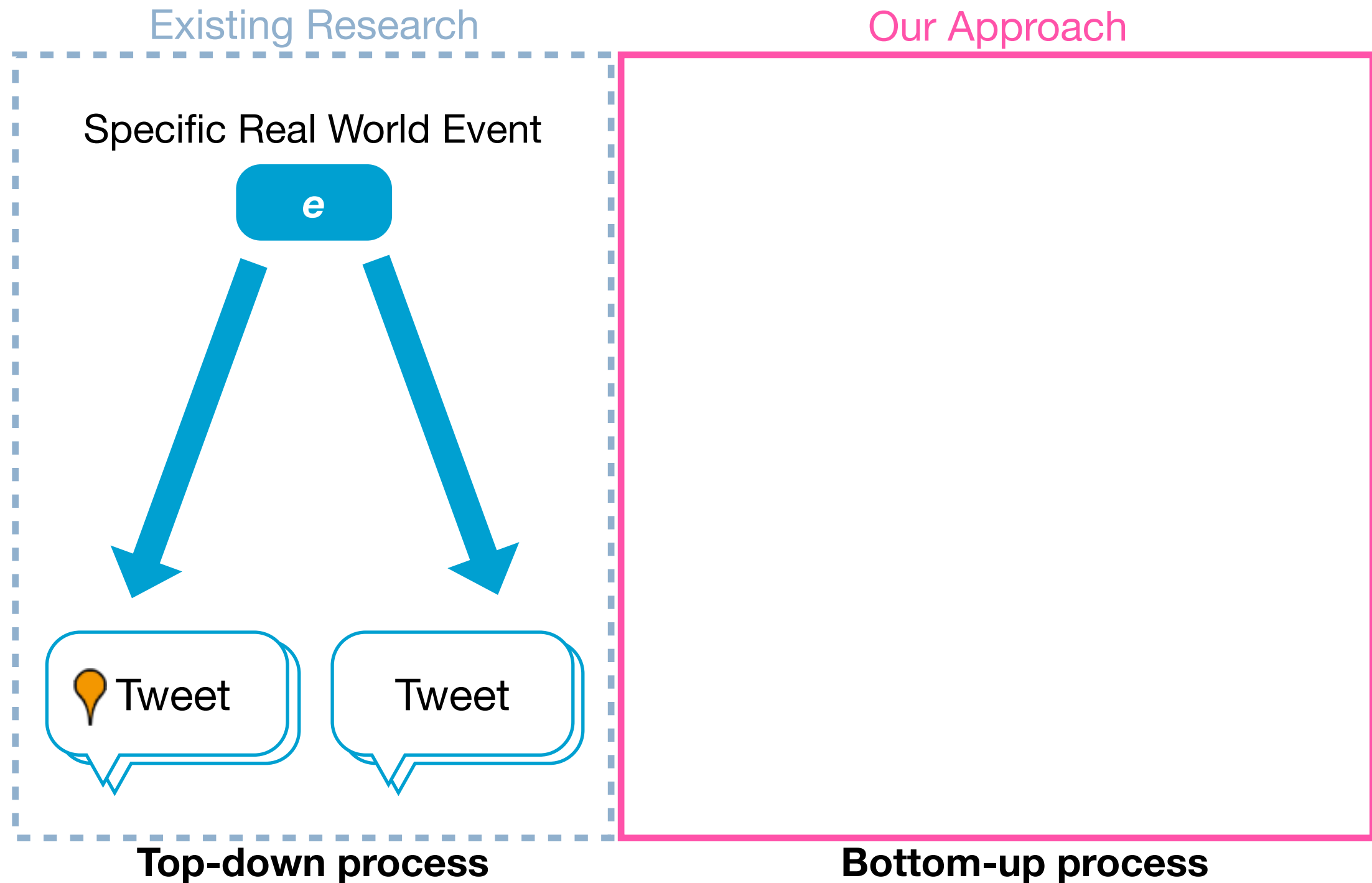
Comparison with Related Work
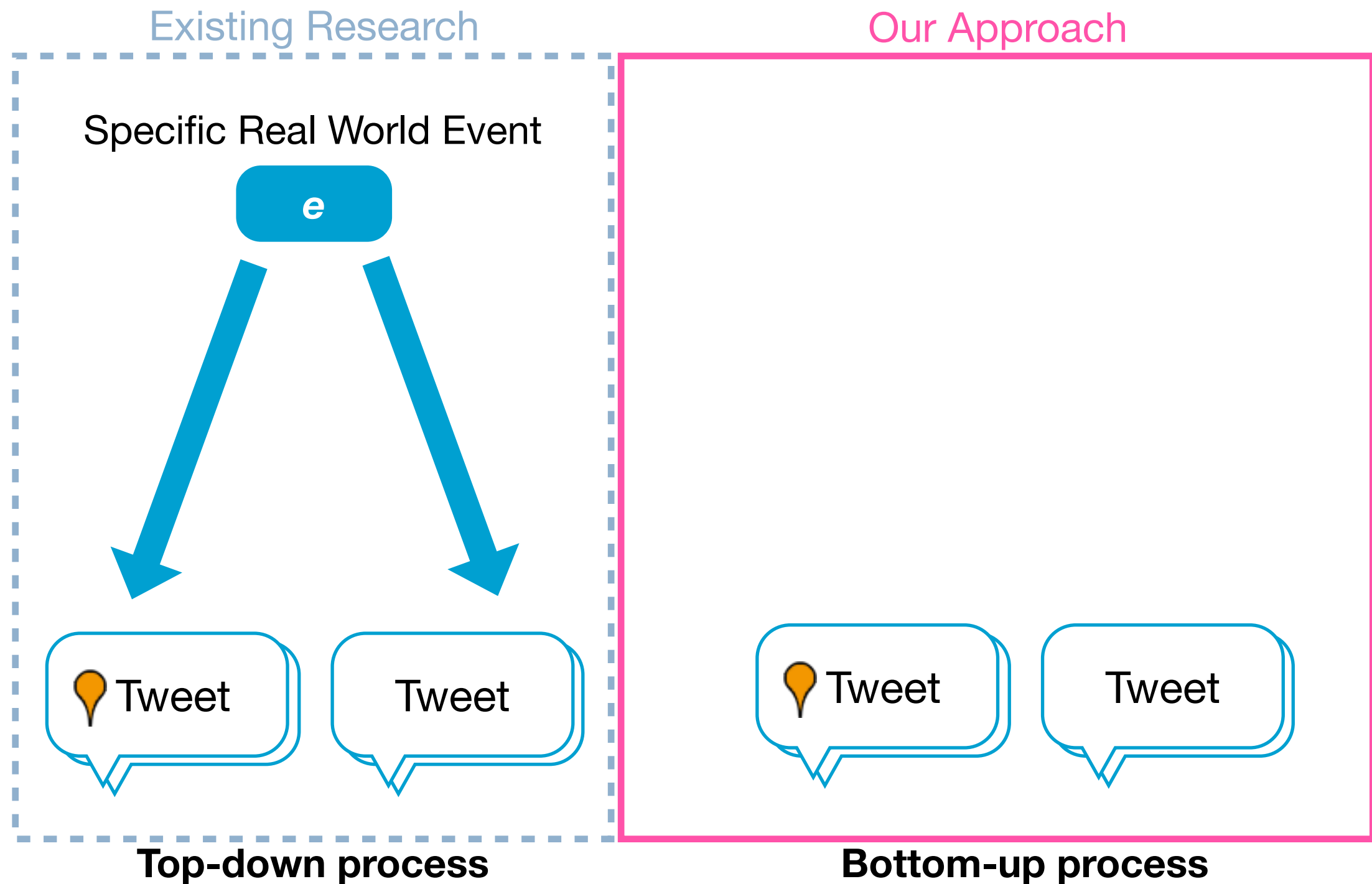
Comparison with Related Work
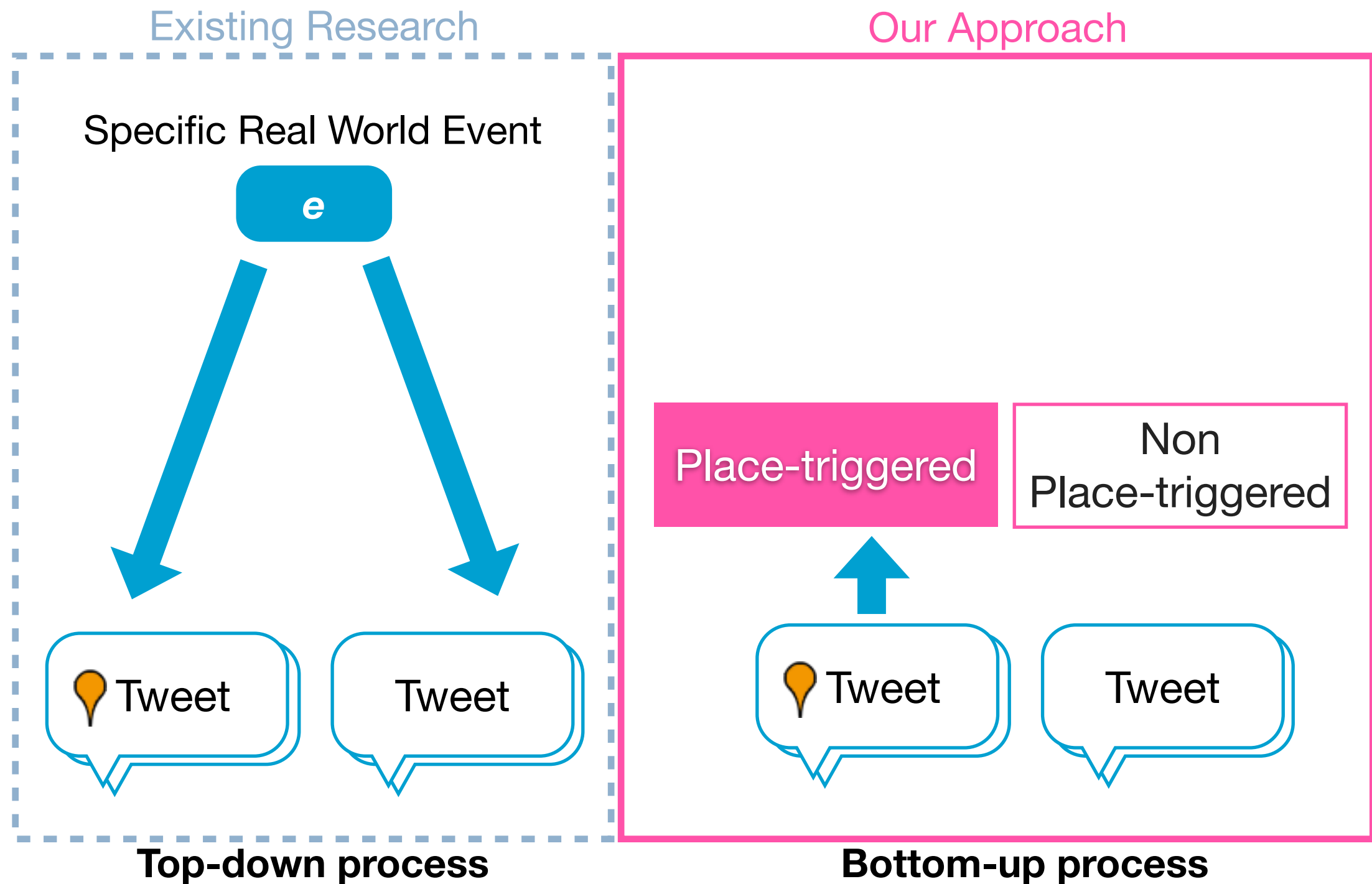
Comparison with Related Work
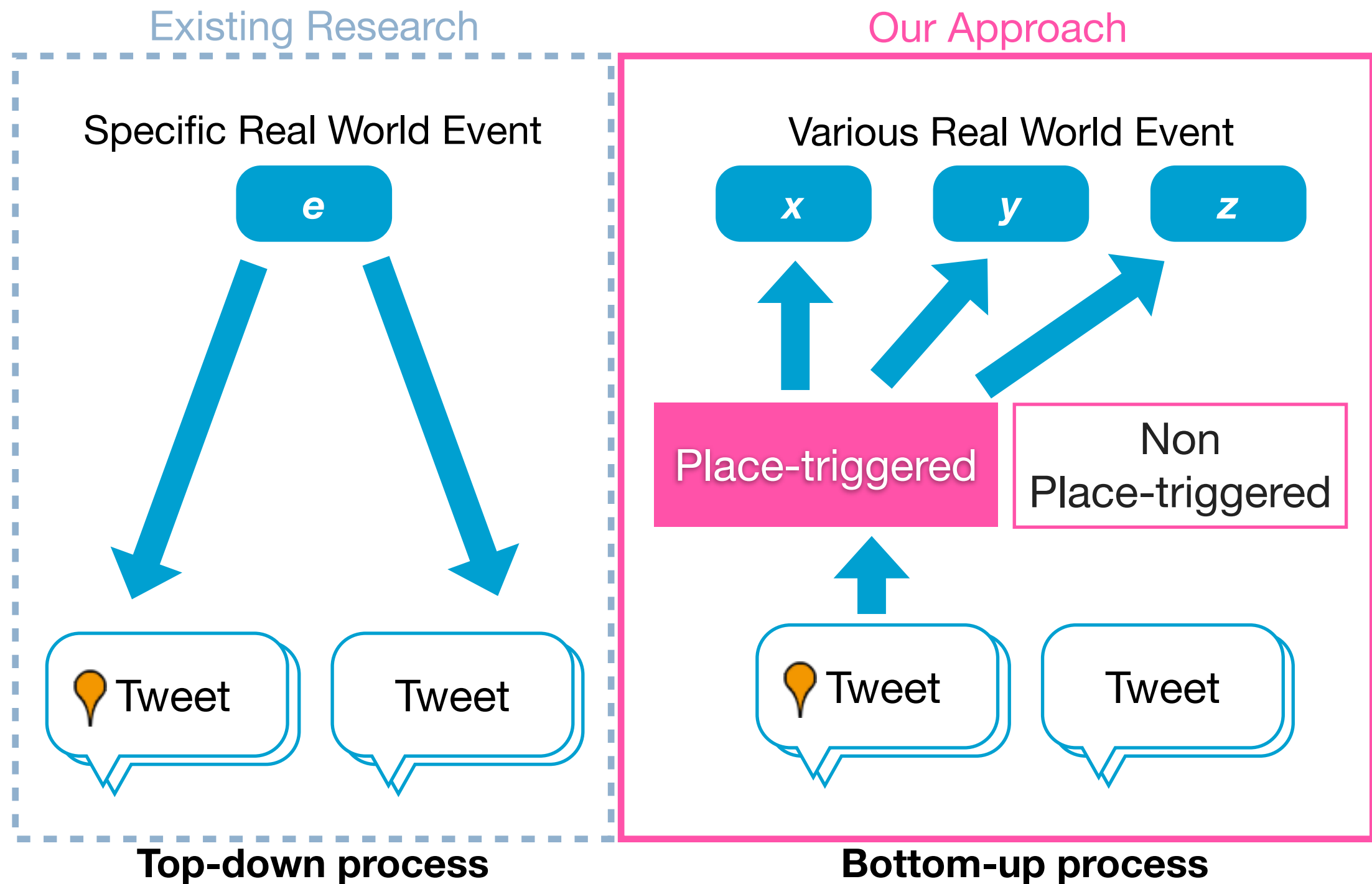
Preliminary Survey

- **Geotagged tweets in Twitter around Japan**

  Period: From 2011-11-21 to 2011-12-31

  Number of sample: 2,000

  Classified these tweets to certain types based on their content

Accident
Earthquake
Other Events
Traffic
Shopping
Moving
Back at Home
Weather
Food

Noise

Report of Whereabouts
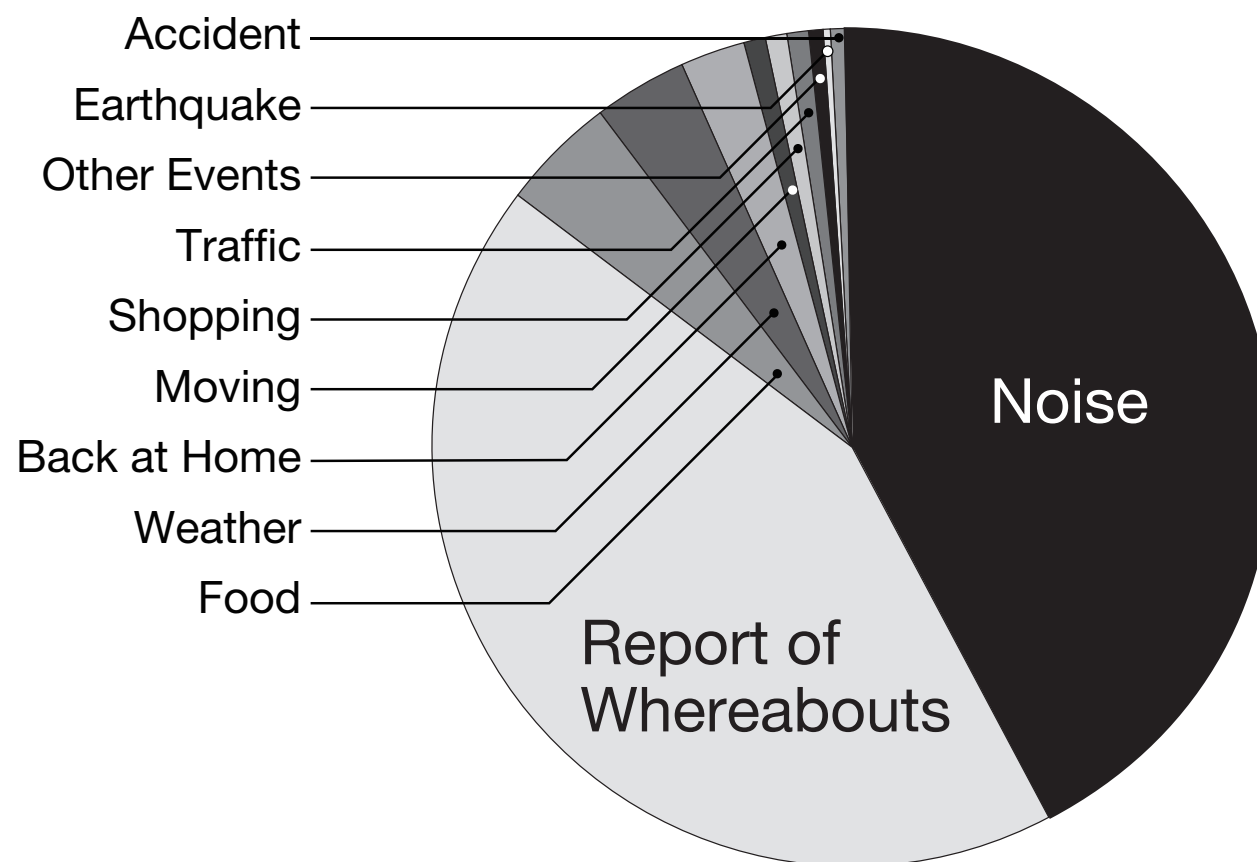
Most of the tweets (42.5%) were classified as noise

Classification of the Place-triggered Geotagged Tweets

- **Classified to Five types:**

Report of whereabouts

‣ A tweet that user refers to his/her current location

Food

‣ A tweet where user shares information regarding current food or drink

Weather

‣ A tweet about weather of the location

Back at home

‣ A tweet where user reports the fact that he/she is back at home

Earthquake

‣ A tweet in which user reports the feeling of the earthquake

hide tokuda lab.

Approach

- **How do we detect Place-triggered Geotagged Tweets?**

  We started with straightforward approach

  ‣ Report of whereabouts

    – Detecting checkin activity
      (Foursquare, Loctouch, Imakoko-now)

  ‣ Food, Weather, Back at home and Earthquake

    – Naive keyword matching method with dictionary

      – We assume that people tend to classify tweets mainly by distinctive keywords
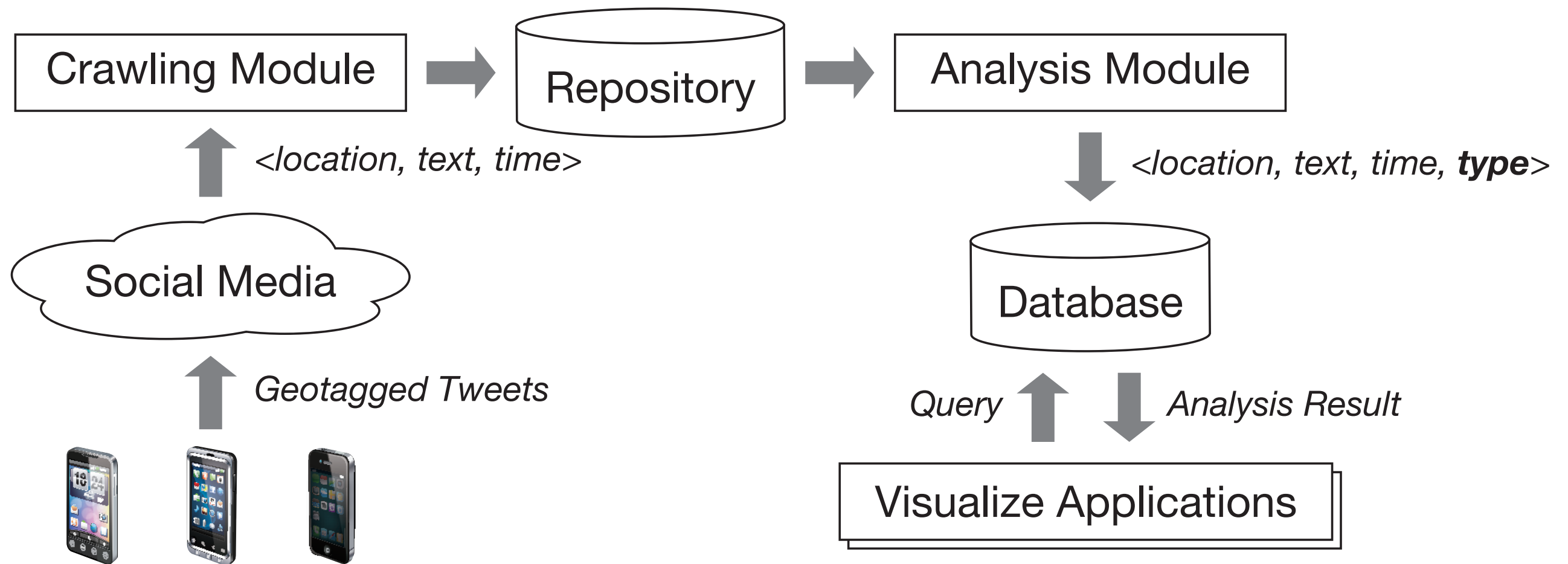
hide tokuda lab.

Design and Implementation

Crawling Module → Repository → Analysis Module

<location, text, time>

*<location, text, time, **type**>*

Social Media

Database

Geotagged Tweets

Query ↑ ↓ *Analysis Result*

Visualize Applications

Interactive Visualization of Place-triggered Geotagged Tweets

Filtering by date / time
Animation view

Each tweet is colord by type

Filtering by types of tweet

Plotting area

Interactive Visualization of Place-triggered Geotagged Tweets

Filtering by date / time
Animation view

Each tweet is colord by type

Filtering by types of tweet

Plotting area

Demo...

hide tokuda lab.
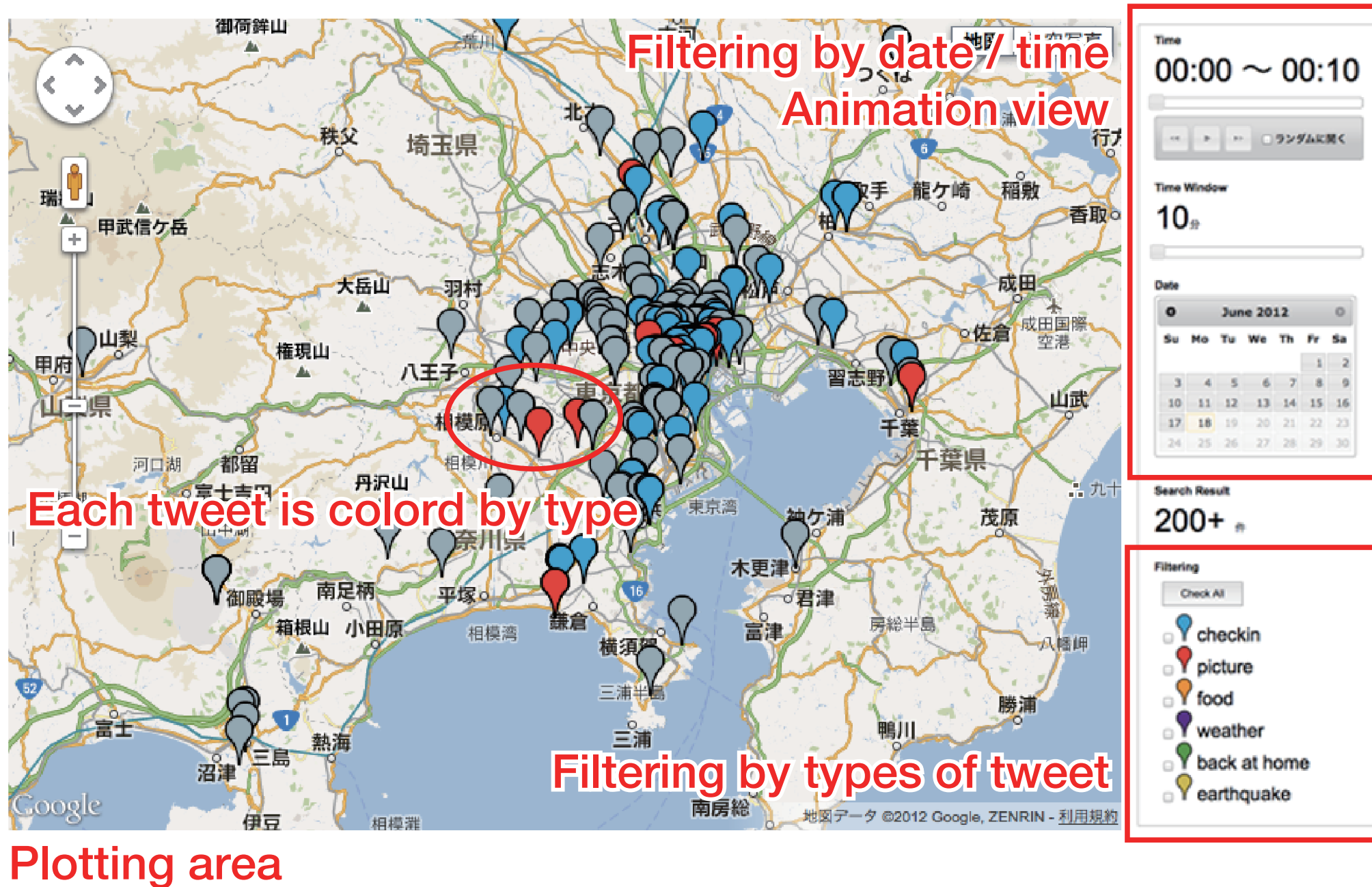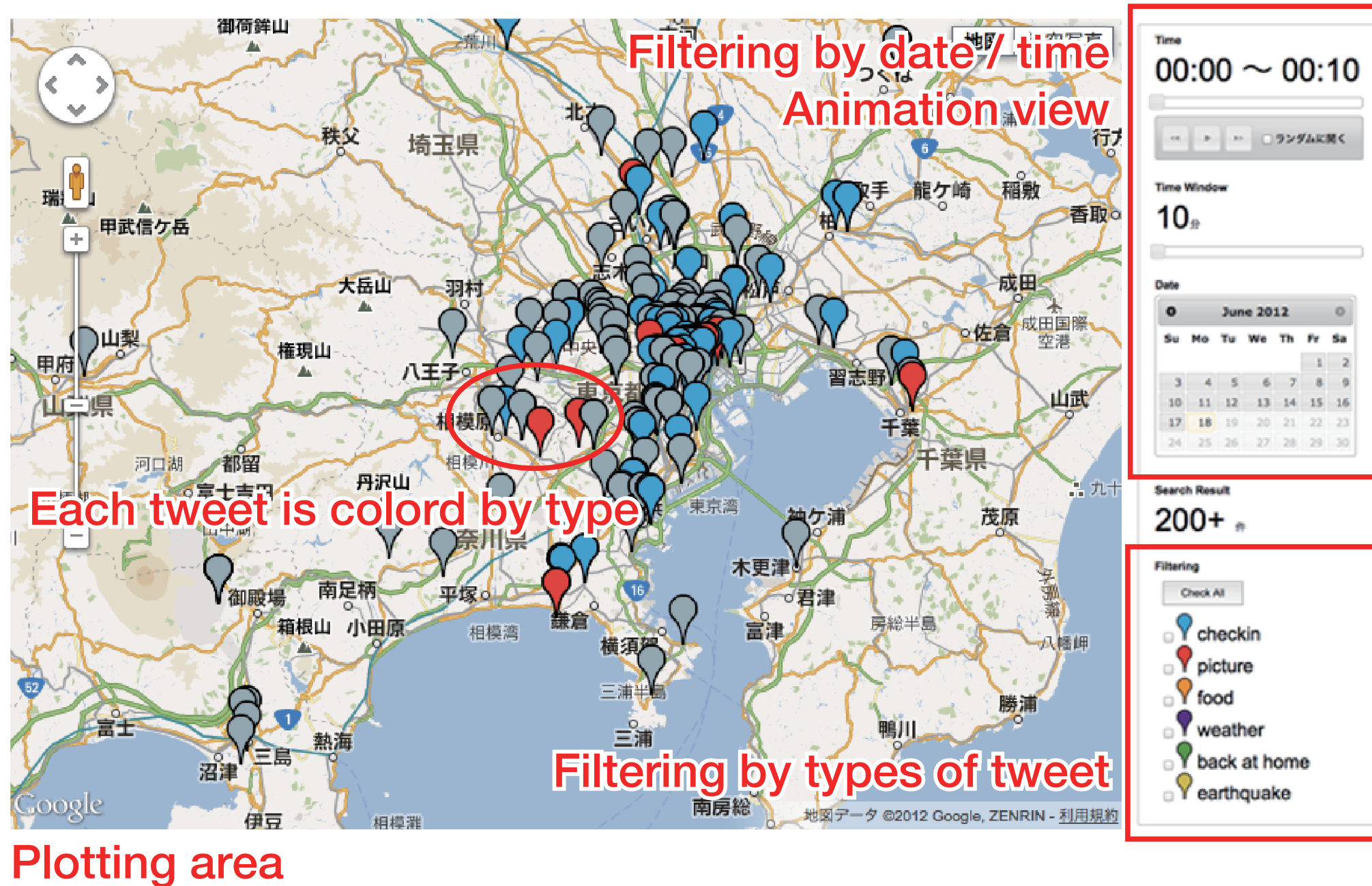
Evaluation

- **Methodology**
  - ‣ Creating Ground-truth
    - – Asked 18 third party people to classify tweets
      - – 12 men in their 20s
      - – 2 men in their 30s
      - – 5 women in their 20s
  - ‣ Dataset
    - – Geotagged tweets nearby Japan
    - – Period: From 2012-01-01 to 2012-03-31
    - – Total amount: 4,524,257
  - ‣ Each participants reviewed 500 tweets which were randomly sampled from the dataset

Evaluation Result

Type of Tweets	Precision	Recall	F-measure
Report of whereabouts	93.18%	77.16%	84.42%
Food	53.6%	17.8%	26.7%
Weather	57%	21%	30%
Back at Home	54%	23%	32%
Earthquake	76%	66%	71%

Table 1. Classification result by the system

	Positive	Negative
TRUE	40.09%	False Negative 15.84%
FALSE	False Positive 2.18%	41.89%

Table 2. Accuracy rate of detecting place-triggered geotagged tweets

Future Work

- **Expanding the classification**
  - ‣ Expand to other countries
  - ‣ More complete categories

- **Improving detection accuracy**
  - ‣ Linguistic analysis, slang

- **Discovering real events**
  - ‣ Automatic event detection
  - ‣ Temporal-spacial analysis should be investigated

Conclusion

- **We defined Place-triggered Geotagged Tweets**
  - ‣ Tweets containing both geotag and content-based relation to your location

- **We classified the place-triggered geotagged tweets as 5 types**
  - ‣ Report of whereabouts, Food, Weather, Back at home and Earthquake

- **We conducted evaluation study**
  - ‣ Showed that the system can detect place-triggered geotagged tweets with an overall accuracy of 82%

- **Contact us**

  hiru@ht.sfc.keio.ac.jp

  http://www.ht.sfc.keio.ac.jp/cpsf/

  Thank you for listening!

hide tokuda lab.

Thank you for listening!

hide tokuda lab.

Twitter Statistics (2011)

- **Tweets per second (TPS)**

  6,939 tweets/sec (Max)

  ‣ 2011.1.1 0:00:04, JST

- **Tweets per day**

  140,000,000 tweets/day (Average)

- **Language**

  ‣ 1. English - 61%

  ‣ 2. Portuguese - 11%

  ‣ 3. Japanese - 6%

  ‣ 4. Spanish - 4%

- **Geotagged**

  0.6% (Jun. 22, 2010)

http://www.marketinggum.com/twitter-statistics-2011-updated-stats/
http://www.thomascrampton.com/twitter/asia-twitter-stats/

hide tokuda lab.

Filtering Module

- **Classifier**

Report of whereabouts

‣ Checkin activity (Foursquare, Loctouch, Imakoko-now)

Food

‣ 86 words

  – breakfast, eat, dinner, ...

Weather

‣ 131 words

  – sunny, cloudy, rainy, cold, ...

Back at home

‣ 5 words

  – back at home, ...

Earthquake

‣ 5 words

  – earthquake, shaking, shook, ...

hide tokuda lab.

Applications

- **For end-user**
  - ‣ Dynamic recommendation service
    - – Restaurant, Entertainment, Road, ...

- **For company**
  - ‣ Traffic accident/congestion detection
  - ‣ Efficient advertisement

- **For city planner**
  - ‣ Provide people moving pattern

hide tokuda lab.

Outline

- **Real world event detection**
  - ‣ From social networking services

- **Place-triggered Geotagged Tweets**
  - ‣ <u>New concept</u> of classifying tweets
  - ‣ Preliminary survey
  - ‣ Approach to detect Place-triggered Geotagged Tweets

- **Prototype system**
  - ‣ Design and implementation

- **Evaluation**
  - ‣ Using ground truth created by 18 third party people
  - ‣ Future work

hide tokuda lab.

Point of View

- **System requirements**

  System which *extract*, *classify* and *provide real-time* **dynamic attributes** of the event

- **Key attribute**

  Location

  ‣ Location is the most common denominator for a wide variety of events

  ‣ In many cases, it's the single most important one

- **Data sources**

  Social networking services are suitable

  ‣ Twitter