Earthquake Shakes Twitter Users: Real-time event Detection by Social Sensors

Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo
The University of Tokyo

Presented by
Taymuraz Abaev

Plan

- Introduction: Twitter
- Event Detection
- Model
- Experiment and Evaluation
- Earthquake Reporting System
- Conclusion

Twitter

- Popular microblogging service
- 50 million active users
- Real-time nature

Event Detection

Event - an arbitrary classification of a space/time region.

Target events properties:

- Large scale
- Influence people's daily life
- Have both spatial and temporal regions

Semantic Analysis on Tweet

- Search by keyword
- Tweets classification
  - Statistical features
  - Keyword features
  - Word context features

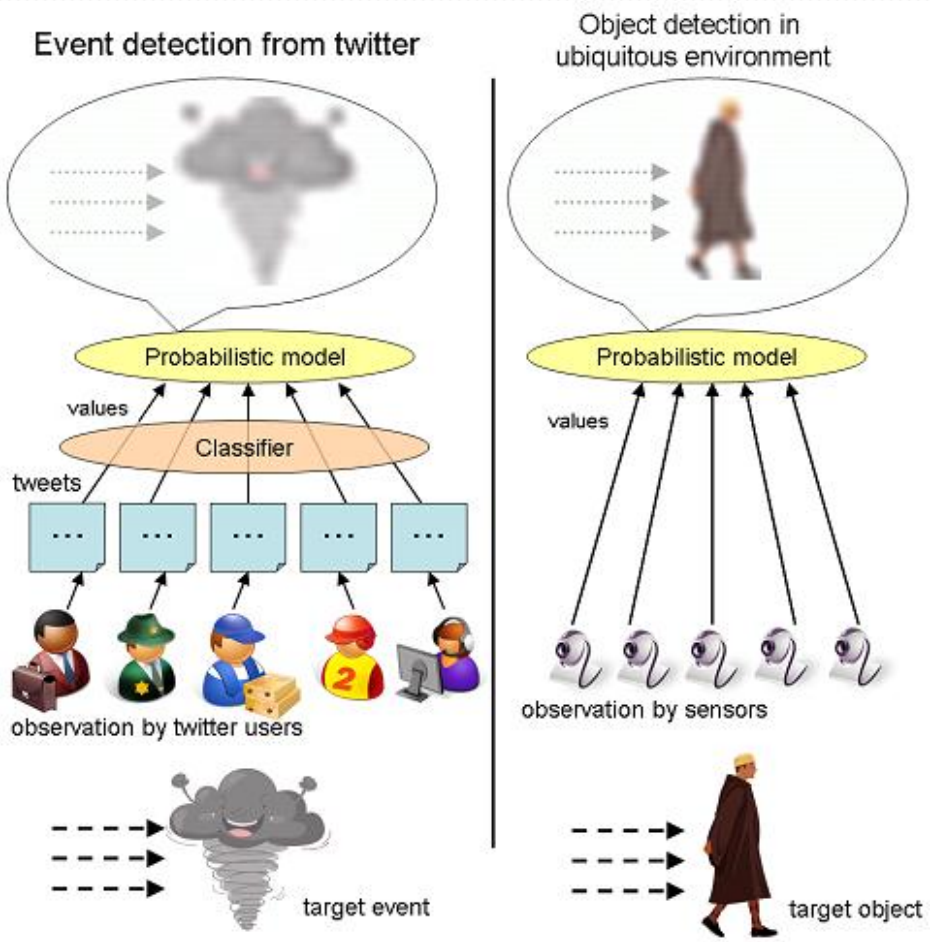
Tweet as a Sensory Value

User = sensor

Tweet = sensor reading

Sensor characteristics:

- Activated / inoperable / operating incorrectly
- Time and location



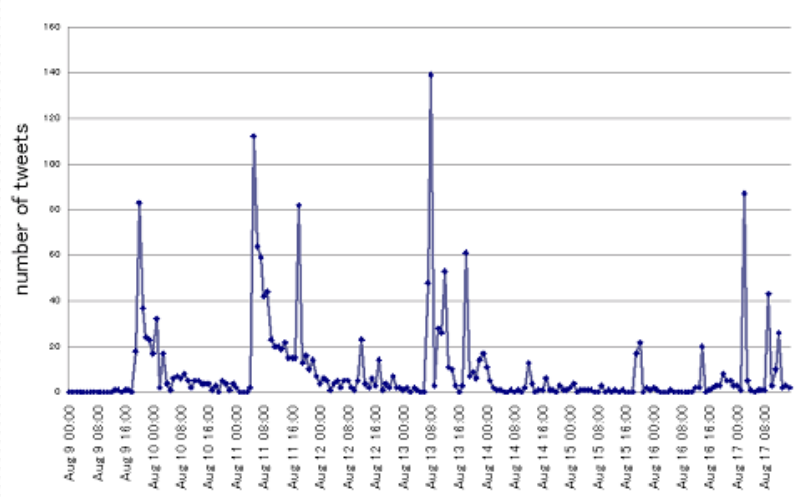
Temporal Model

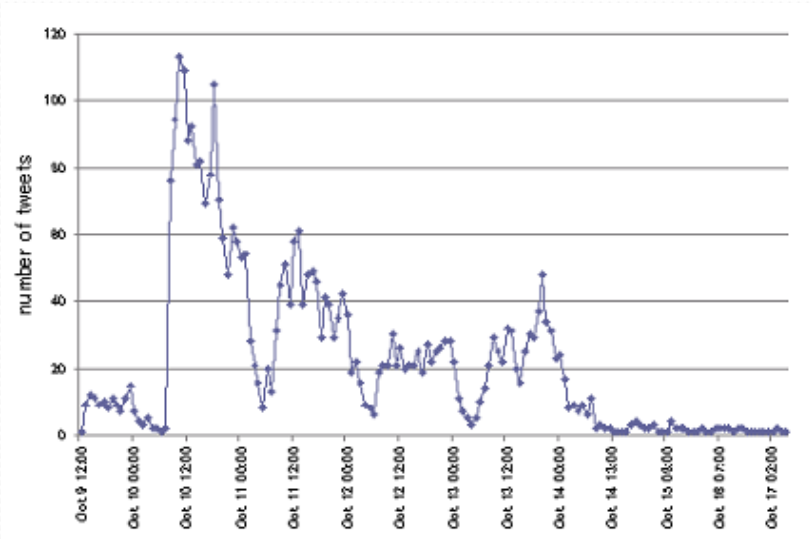
$$p_{occur}(t) = 1 - p_f^{n_0(1-e^{-\lambda(t+1)})/(1-e^{-\lambda})}$$

p_f = 0.35 - false-positive ratio

n_0 - number of sensors at time 0

λ = 0.34





Spatial Model

$$\{x_t, t \epsilon N\}$$
$$x_t = f_t(x_{t-1}, v_{t-1})$$
$$v_{t-1} - i.i.d. process\ noise\ sequence$$

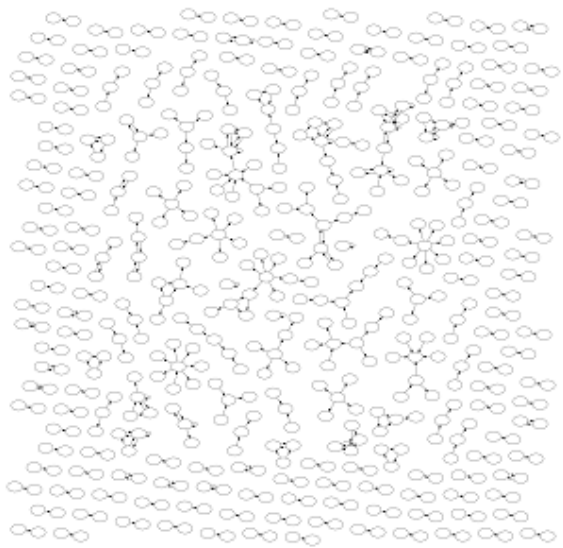
- Kalman Filters

$$p(x_t|z_t) = N(x_t; m_t, P_t) - Gaussian\ density$$
$$N(x; m, P) - Gaussian\ density$$
$$m - mean, P - covariance, z_t - data\ at\ moment\ t$$

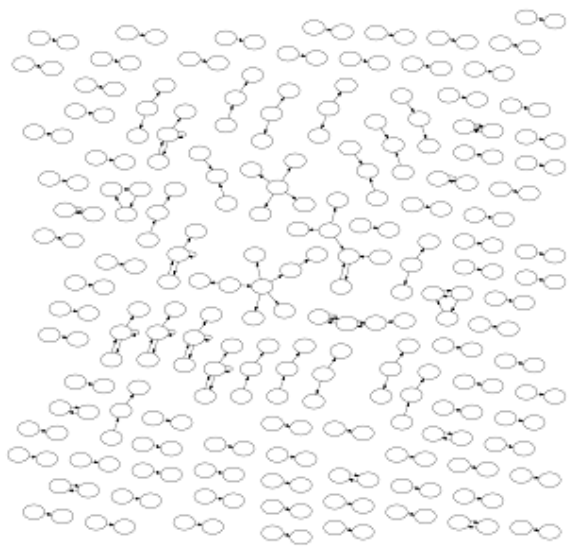
- Particle filters

$$Bel(x_t) = \{x_t^i, w_t^i\}, i = 1..n$$
$$x_t^i - location\ hypothesis\ at\ t, w_t^i - weight$$

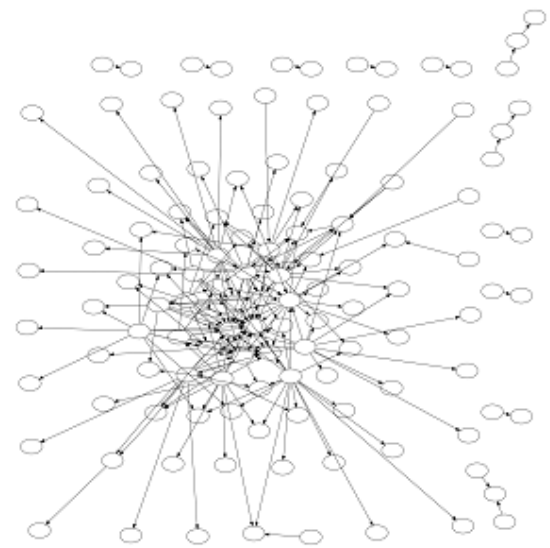
Information Diffusion related to a Real-time Event



Earthquake information diffusion network



Typhoon information diffusion network



A new Nintendo game information diffusion network

Evaluation by Semantic Analysis

Table 1: Performance of classification.

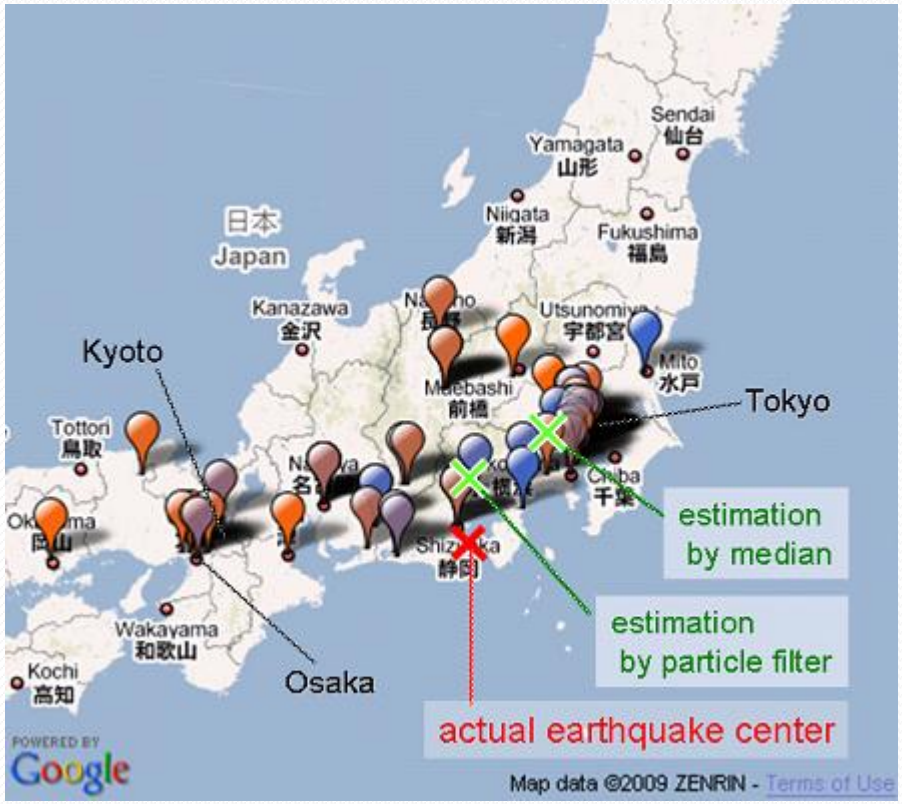
(i) *earthquake* query:

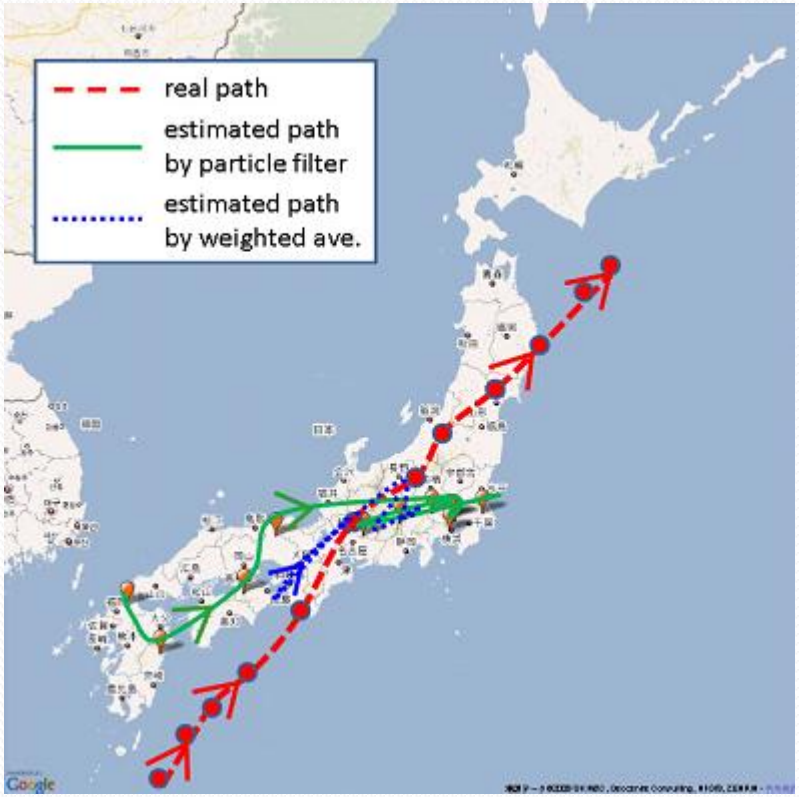
Features	Recall	Precision	F-value
A	87.50%	63.64%	73.69%
B	87.50%	38.89%	53.85%
C	50.00%	66.67%	57.14%
All	87.50 %	63.64%	73.69%

(ii) *shaking* query:

Features	Recall	Precision	F-value
A	66.67%	68.57%	67.61%
B	86.11%	57.41%	68.89%
C	52.78%	86.36%	68.20%
All	80.56 %	65.91%	72.50%

Evaluation of Spatial Estimation(1)





Evaluation of Spatial Estimation(2)

Table 2: Location estimation accuracy of earthquakes from tweets. For each method, we show the difference of the estimated latitude and the longitude to the actual ones, and the Euclid distance of them. Smaller distance means better performance.

Date	Actual center		Median (baseline)			Weighted ave. (baseline)			Kalman filters			**Particle filters**		
	lat.	long.	lat.	long.	dist.	lat.	long.	dist.	lat.	long.	dist.	lat.	long.	dist.
Aug. 10 01:00	33.10	138.50	3.40	-0.80	3.49	2.70	-0.10	2.70	2.67	-0.50	2.72	2.60	0.50	**2.65**
Aug. 11 05:00	34.80	138.50	0.90	-0.90	1.27	0.70	-0.30	0.76	0.60	-0.20	**0.63**	0.30	-0.90	0.95
Aug. 13 07:50	33.00	140.80	1.30	-9.60	9.69	2.30	-2.30	**3.25**	1.63	-3.75	4.09	2.70	-2.70	3.82
Aug. 17 20:40	33.70	130.20	4.60	6.00	7.56	0.90	3.20	3.32	1.63	4.35	4.65	0.10	-0.80	**0.81**
Aug. 18 22:17	23.30	123.50	7.80	9.90	12.60	8.70	10.90	13.95	8.32	10.13	13.11	5.60	8.10	**9.85**
Aug. 21 08.51	35.70	140.00	0.50	-4.40	4.43	0.10	-1.00	1.00	0.00	-0.60	**0.60**	-0.80	0.48	0.93
Aug. 24 13:30	37.50	138.60	-0.40	0.00	**0.40**	-0.50	0.40	0.64	-0.50	0.30	0.58	2.40	0.70	2.50
Aug. 24 14:40	41.10	140.30	-1.90	1.10	2.20	-1.30	0.50	**1.39**	-1.50	0.50	1.58	3.10	2.00	3.69
Aug. 25 02:22	42.10	142.80	-2.90	-3.90	4.86	-6.10	-3.80	7.19	-5.20	-3.70	6.38	-1.80	-1.90	**2.62**
Aug. 25 20:19	35.40	140.40	1.60	-1.80	2.41	2.20	-0.70	2.31	0.70	-1.60	1.75	1.40	0.10	**1.40**
Aug. 31 00:46	37.20	141.50	-0.40	-3.60	3.62	-1.10	-2.30	2.55	-1.30	-2.20	2.56	-0.30	-0.30	**0.42**
Aug. 31 21:11	33.40	130.90	-4.50	-3.60	5.76	0.50	2.10	2.16	0.70	1.90	2.02	-0.20	-1.70	**1.71**
Sep. 3 22:26	31.10	130.30	6.20	-0.10	6.20	4.00	5.00	6.40	4.90	7.20	8.71	2.40	2.10	**3.19**
Sep. 4 11:30	35.80	140.10	3.10	-1.70	3.54	0.20	-0.90	**0.92**	0.00	-1.00	1.00	0.80	1.40	1.61
Sep. 05 10:59	37.00	140.20	-2.70	-8.30	8.73	-1.40	-3.10	**3.40**	-1.30	-3.30	3.55	-2.10	-5.80	6.17
Sep. 08 01:24	42.20	143.00	-3.60	-8.90	9.60	-2.50	-3.90	4.63	-4.50	-6.00	7.50	1.30	-3.60	**3.83**
Sep. 10 18:29	43.20	146.20	-5.90	-10.20	11.78	-4.90	-7.10	8.63	-4.50	-7.20	8.49	-0.90	-7.00	**7.06**
Sep. 16 21:38	33.40	130.90	1.10	-0.20	**1.12**	0.90	2.10	2.28	0.50	1.40	1.49	-0.20	-2.50	2.51
Sep. 22 20:40	47.60	141.70	-11.10	-7.50	13.40	-10.80	-3.10	11.24	-11.30	-3.80	11.92	-7.80	-3.00	**8.36**
Oct. 1 19:43	36.40	140.70	0.70	-3.80	3.86	-0.60	-1.80	1.90	-0.30	-1.50	1.53	-0.70	0.30	**0.76**
Oct. 5 09:35	42.40	141.60	-3.70	-3.10	4.83	-2.70	-2.00	3.36	-2.60	-1.60	3.05	1.10	-1.70	**2.02**
Oct. 6 07:49	35.90	137.60	0.50	1.20	1.30	-0.20	0.80	0.82	-0.10	0.90	0.91	0.30	0.50	**0.58**
Oct. 10 17:43	41.80	142.20	-3.50	-5.40	6.44	-1.40	-2.10	**2.52**	-2.20	-2.60	3.41	2.40	-1.30	2.73
Oct. 12 16:10	35.90	137.60	2.80	0.50	2.84	0.80	1.20	**1.44**	0.80	1.60	1.79	3.60	1.40	3.86
Oct. 12 18:42	37.40	139.70	-2.00	-4.40	4.83	-1.50	-0.90	1.75	-1.70	-1.40	2.20	-1.00	-0.60	**1.17**
Average distance					5.47			3.62			3.85			**3.01**

Table 3: Trajectory estimation accuracy of typhoon Melor from tweets.

Date	Location		Median (baseline)			Weighted ave. (baseline)			Kalman filters			**Particle filters**		
	lat.	long.	lat.	long.	dist.	lat.	long.	dist.	lat.	long.	dist.	lat.	long.	dist.
Oct. 7 12:00	29.00	131.80	-1.90	-1.90	**2.69**	-5.20	-3.60	6.32	-3.90	-1.10	4.05	-4.70	1.10	4.83
Oct. 7 15:00	29.90	132.50	-3.70	-2.60	4.52	-3.80	-2.40	4.49	3.20	3.10	4.46	-2.70	0.90	**2.85**
Oct. 7 18:00	30.80	133.20	-4.10	-1.90	4.52	-4.40	-3.50	5.62	-6.40	5.40	8.37	-3.20	-0.70	**3.28**
Oct. 7 21:00	31.60	134.30	-3.90	-3.50	5.24	-3.60	-3.30	4.88	-10.90	-1.60	11.02	-3.70	-0.50	**3.73**
Oct. 8 0:00	32.90	135.60	-2.30	-0.10	**2.30**	-2.30	-0.90	2.47	-12.60	-20.40	23.98	-2.90	-3.50	4.55
Oct. 8 6:00	35.10	137.20	1.60	3.00	3.40	0.80	1.70	**1.88**	4.20	16.00	16.54	-0.60	-2.50	2.57
Oct. 8 9:00	36.10	138.80	-0.60	3.60	3.65	0.00	0.50	**0.50**	0.50	2.60	2.65	0.70	-0.80	1.06
Oct. 8 12:00	37.10	139.70	1.70	3.90	4.25	1.50	1.20	1.92	2.10	1.60	2.64	1.40	0.10	**1.40**
Oct. 8 15:00	38.00	140.90	2.30	3.20	3.94	2.40	2.20	**3.26**	1.70	7.60	7.79	2.40	2.70	3.61
Oct. 8 18:00	39.00	142.30	3.20	7.30	7.97	3.50	5.10	**6.19**	2.10	-18.80	18.92	3.70	5.10	6.30
Oct. 8 21:00	40.00	143.60	4.30	3.90	5.81	4.00	5.30	6.64	1.60	4.50	4.78	4.20	3.10	**5.22**
Average distance					4.39			4.02			9.56			**3.58**

Earthquake Reporting System

Table 4: Facts about earthquake detection.

Date	Magnitude	Location	Time	E-mail sent time	#tweets within 10 min	Announce of JMA
Aug. 18	4.5	Tochigi	6:58:55	7:00:30	35	07:08
Aug. 18	3.1	Suruga-wan	19:22:48	19:23:14	17	19:28
Aug. 21	4.1	Chiba	8:51:16	8:51:35	52	8:56
Aug. 25	4.3	Uraga-oki	2:22:49	2:23:21	23	02:27
Aug. 25	3.5	Fukushima	22:21:16	22:22:29	13	22:26
Aug. 27	3.9	Wakayama	17:47:30	17:48:11	16	17:53
Aug. 27	2.8	Suruga-wan	20:26:23	20:26:45	14	20:31
Aug. 31	4.5	Fukushima	00:45:54	00:46:24	32	00:51
Sep. 2	3.3	Suruga-wan	13:04:45	13:05:04	18	13:10
Sep. 2	3.6	Bungo-suido	17:37:53	17:38:27	3	17:43

Table 5: Earthquake detection performance for two months from August 2009.

JMA intensity scale	2 or more	3 or more	4 or more
Num. of earthquakes	78	25	3
Detected	70(89.7%)	24 (96.0%)	3 (100.0%)
Promptly detected[14]	53 (67.9%)	20 (80.0%)	3 (100.0%)

Conclusion