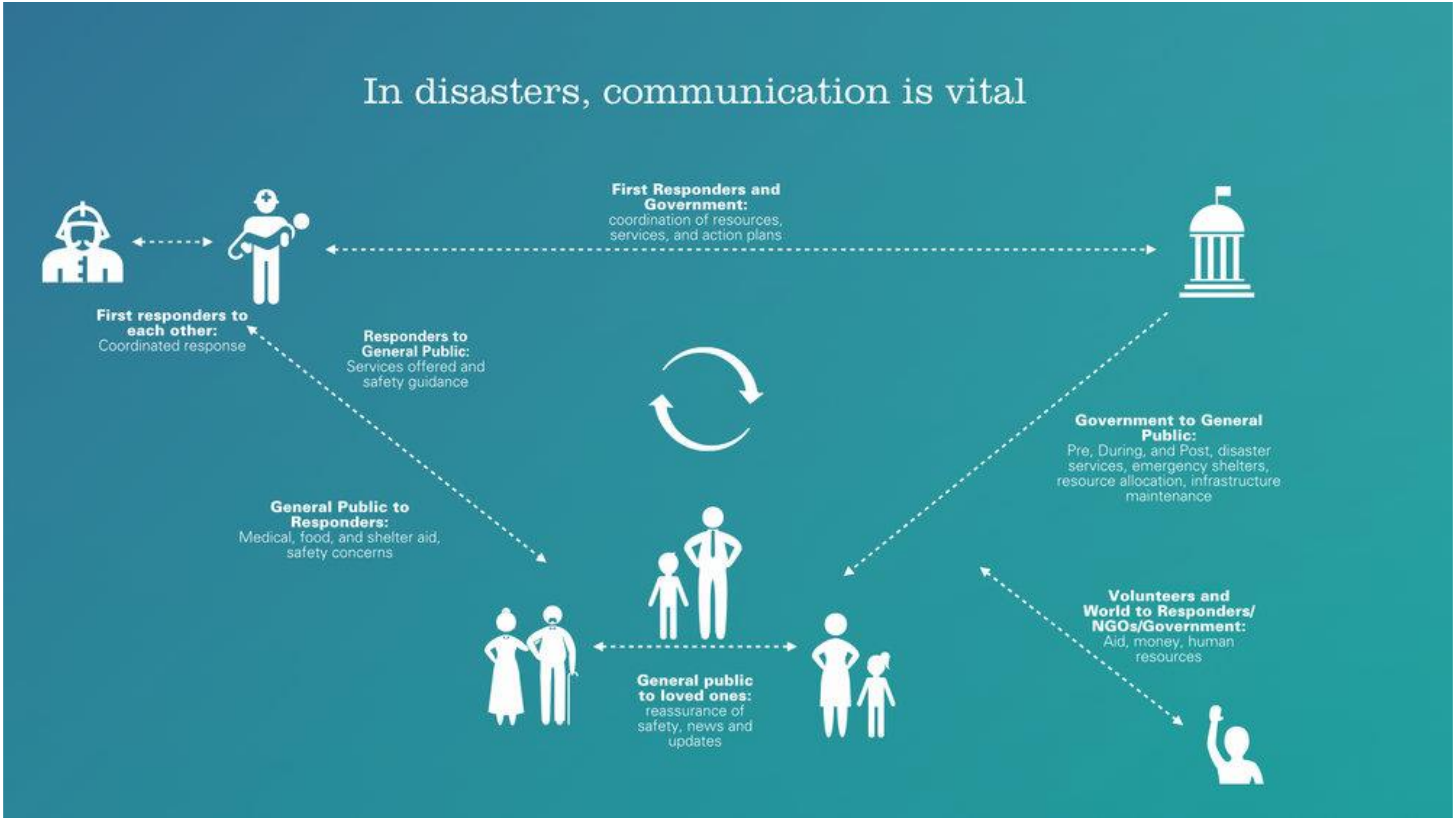
UNIVERSITAS GADJAH MADA

What can NLP do in a disaster?

Tutorial: ICOICT 2020

Yunita Sari, Ph.D

Department of Computer Science and Electronics
Universitas Gadjah Mada, Yogyakarta



Source: http://arianefund.com/cohere


FIDELITE STORE
PRODUITS ALIMENTAIRES
RUE DU CENTRE

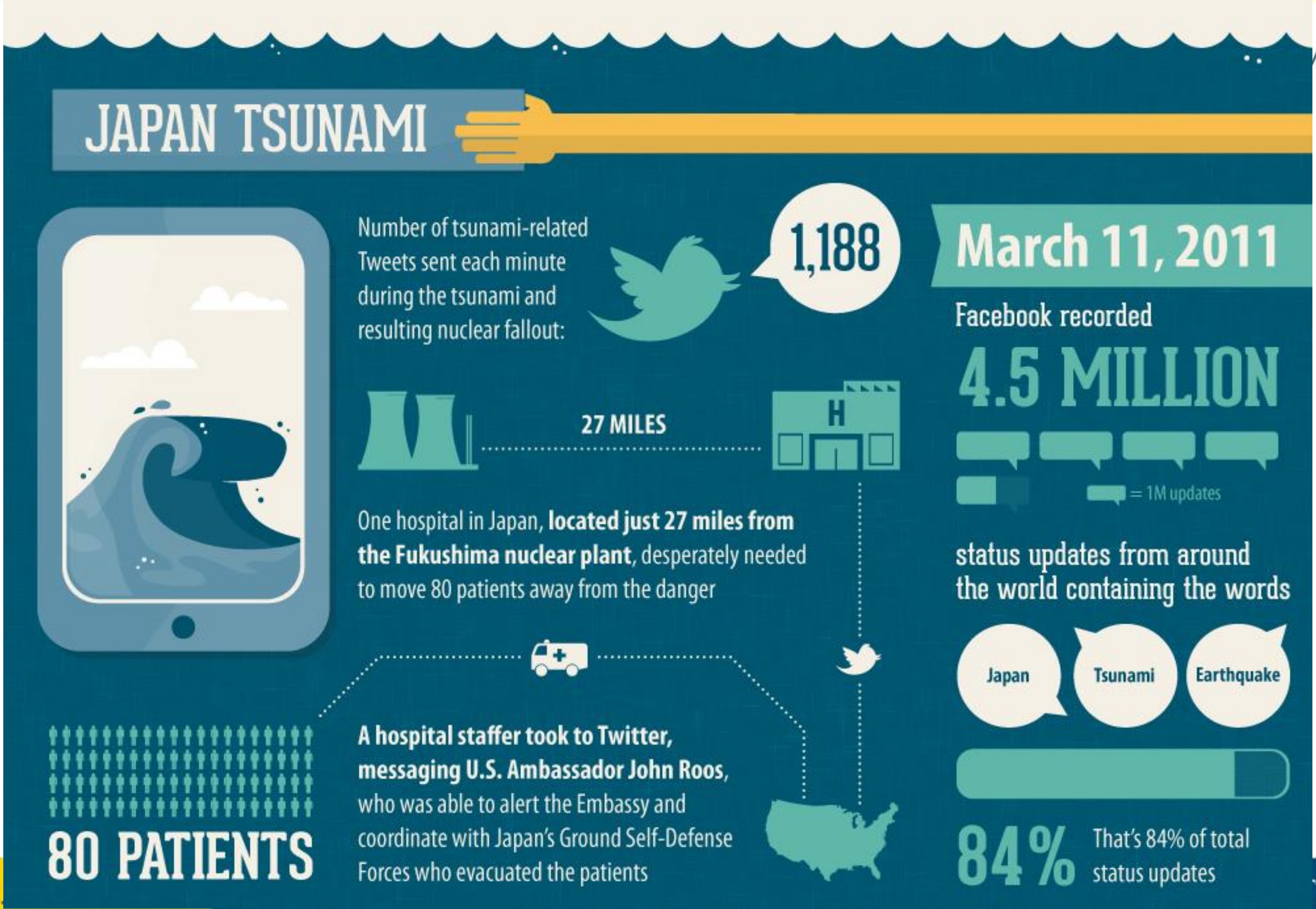
SOCIAL MEDIA
THE NEW FACE of DISASTER RESPONSE

An international study conducted by researchers at the Universidad of Madrid, National Information Communications Technology Australia and the University of California in San Diego have found that social network activity during and in the hours following a natural disaster can quickly reveal the extent of damage that took place in a particular area and time period. (https://www.firestorm.co)

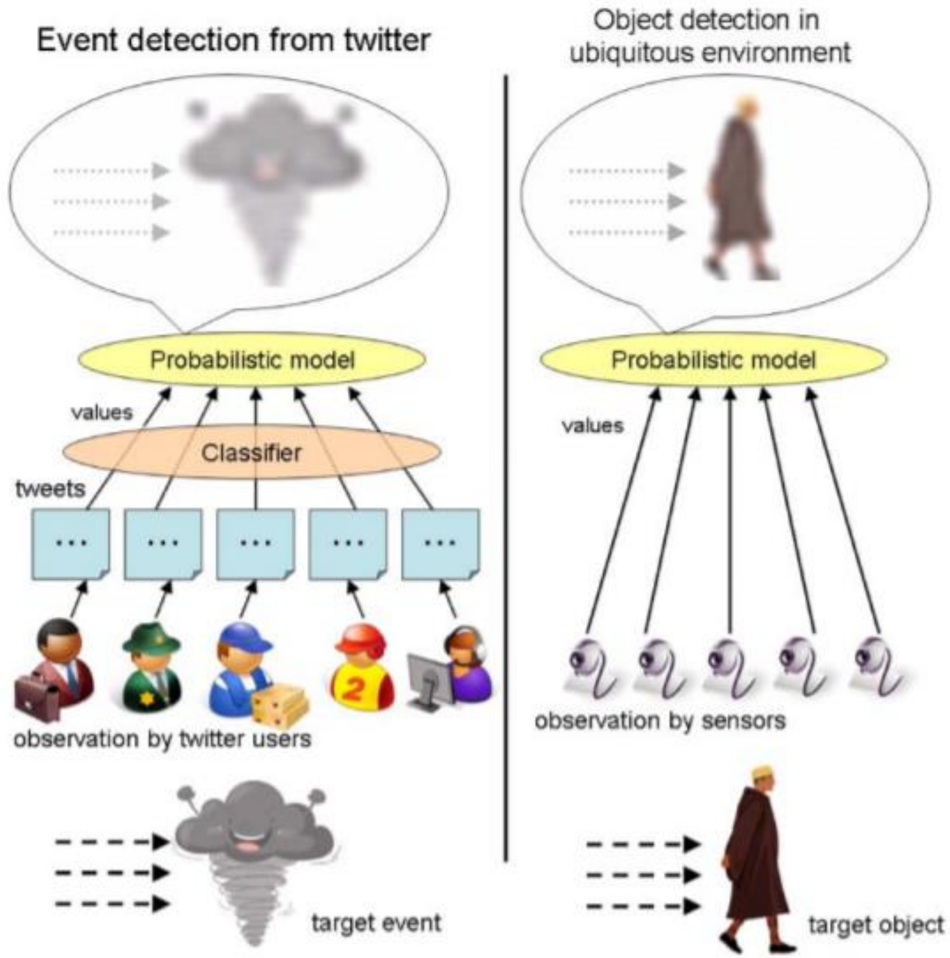
University of San Francisco MPA Program created an infographics that explains how social media is creating major changes in disaster recovery and response.

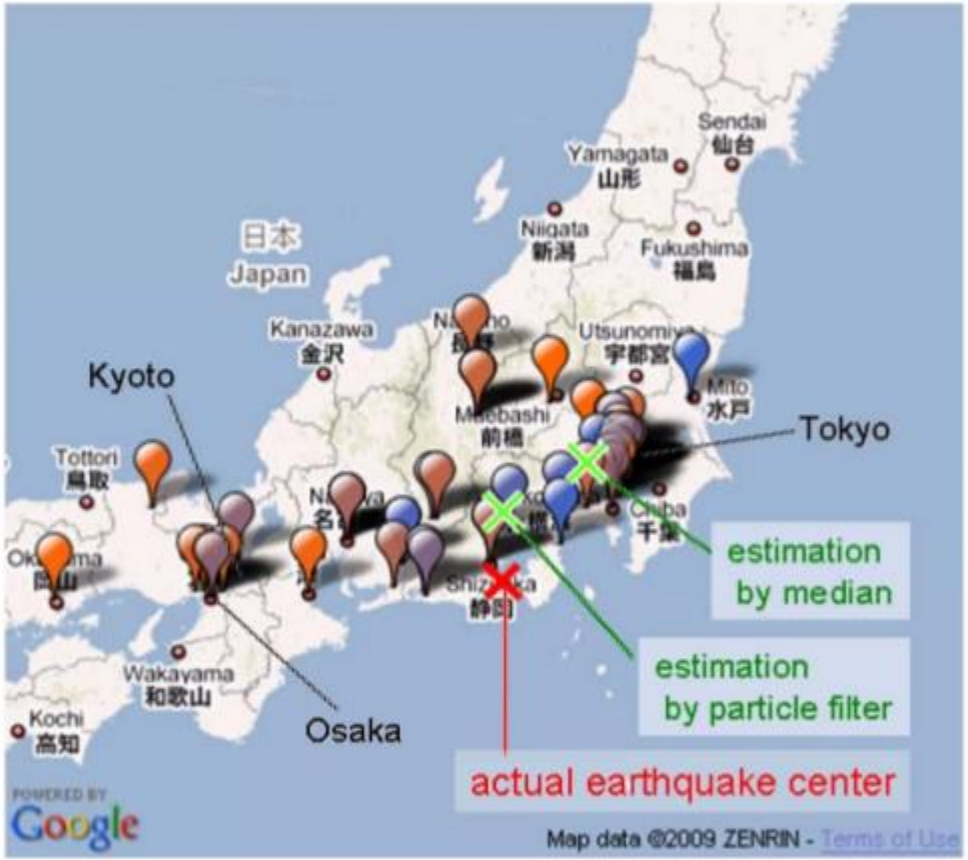
Communication Is the Key to Disaster Management


TORNADO SEASON
1,665
Number of tornadoes that ravaged the U.S., making 2011 the deadliest tornado year in the U.S. ever
A Facebook page dedicated to tornado recovery in Joplin, MO attracted
123,000
members within days of a devastating tornado
The page mobilized volunteers & assisted in the search for survivors
facebook
1,100
An employee of a hospital in Joplin used Facebook to successfully locate 1,100 missing hospital workers
Tuscaloosa, AL, created Tuscaloosa Forward, a social media website that let residents share ideas for rebuilding
300 IDEAS
shared by over
4,000 VISITORS
= 200 visitors
A school system in Tuscaloosa posted a request for volunteers to help with school cleanup efforts on social networks
80
volunteers
arrived in under
30
minutes


JAPAN TSUNAMI
Number of tsunami-related Tweets sent each minute during the tsunami and resulting nuclear fallout:
1,188
27 MILES
H
One hospital in Japan, **located just 27 miles from the Fukushima nuclear plant**, desperately needed to move 80 patients away from the danger
80 PATIENTS
A hospital staffer took to Twitter, messaging U.S. Ambassador John Roos, who was able to alert the Embassy and coordinate with Japan's Ground Self-Defense Forces who evacuated the patients
March 11, 2011
Facebook recorded
4.5 MILLION
= 1M updates
status updates from around the world containing the words
Japan
Tsunami
Earthquake
84%
That's 84% of total status updates

Tweet as a Sensory Value (Sakaki et al., 2013)





Tweet as social sensor

Earthquake detector

Earle et al. (2012); Robinson et al. (2013)

Firsthand reports	Not firsthand reports
Woah! Earthquake	Magnitude 3.8 earthquake shakes Wellington: Wellingtonians were shaken awake by a magnitude 3.8 earthquake early…http://t.co/rT4UvjzH
Earthquake!! 2 small 3-second-each tremors just now!!	i thought there was an earthquake or some sort of world ending experience but then i realised my brother was running around upstairs.. woops
That was a goodun. #eqnz	Large earthquake struck Vanuatu. Imagine the thoughts running through their heads when the earth started to shake
oooh, big wobble, heard that coming way off #eqnz	Can't believe it's been 2 years today since Christchurch had its major earthquake #KiaKaha

Intensity prediction based on impact descriptions

Massive earthquake. House covered in glass. Bookshelf on floor. Lights fallen out. Still shaking → classified as 'VI Strong' in the MMI scale

Influenza Epidemic Detection

Aramaki et al. (2011);

Headache? You might have flu. **[Suspicions]** → (-)

A-bad- influenza-is-going-around-in-our-lab. → (+)

Hay Fever map generation

Takahashi et al. (2011)

1. NE Recognition

people · people · location

TANAKA Taro and Hanako who lived in Kesennuma City can’t be reached. Does anybody know where they are?

2. Safety Information Classification

Label	Definition	Example
I	Him/Herself is alive	I’m XXX in YYY City. I’m all right.
L	Alive	Mr./Ms. XXX in YYY City is at ZZZ Shelter. He/She is alive.
P	Passed away	—
M	Missing	The safety of Mr./Ms. XXX living in YYY City is unknown.
H	Help Request	Mr./Ms. XXX is left in YYY and needs help! My relatives/parents/... staying in South XXX City are missing.
S	Information request	Refugees at XXX School are provided enough daily supplies? (Safety information of unspecific individual, region, etc.)
O	Not safety information	You can post safety information on this site!
R	External link	Survivor list of XXX City: http://...
U	Unknown	(Non-Japanese or nonsense postings)

(Neubig and Matsubayashi, 2010)-- Safety Information Mining

Main Steps

UNIVERSITAS GADJAH MADA

Data Collection

Collecting/scrapping data from the internet

Data Filtering

Filtering related data/tweet to non-related one. Not every tweet with disaster keywords is related to disaster

Annotation

Manual annotation to create training data, usually utilize crowdsourcing annotation tools/employ volunteers

Training

Training the machine learning model, mostly supervised learning model

Prediction

Using the trained model to predict new data

From Annotation to Corpus Development

CRISIS NLP
Natural Language Processing

Resources for Research on Crisis Informatics Topics
https://crisisnlp.qcri.org/

CrisisLex.org

a repository of crisis-related social media data and tools.
https://crisislex.org/

COVID-19 NLP Resources
https://www.nlpcovid19workshop.org/resources

Emotera
Emotion-annotated Tweets for Disaster Risk Assessment Corpus
http://tinyurl.com/emoteracorpus

FloDusTA
Saudi Tweets Dataset for Flood, Dust Storm, and Traffic Accident Events
https://github.com/BatoolHamawi/FloDusTA

Challenges

Supervised learning → needs a lot of training data

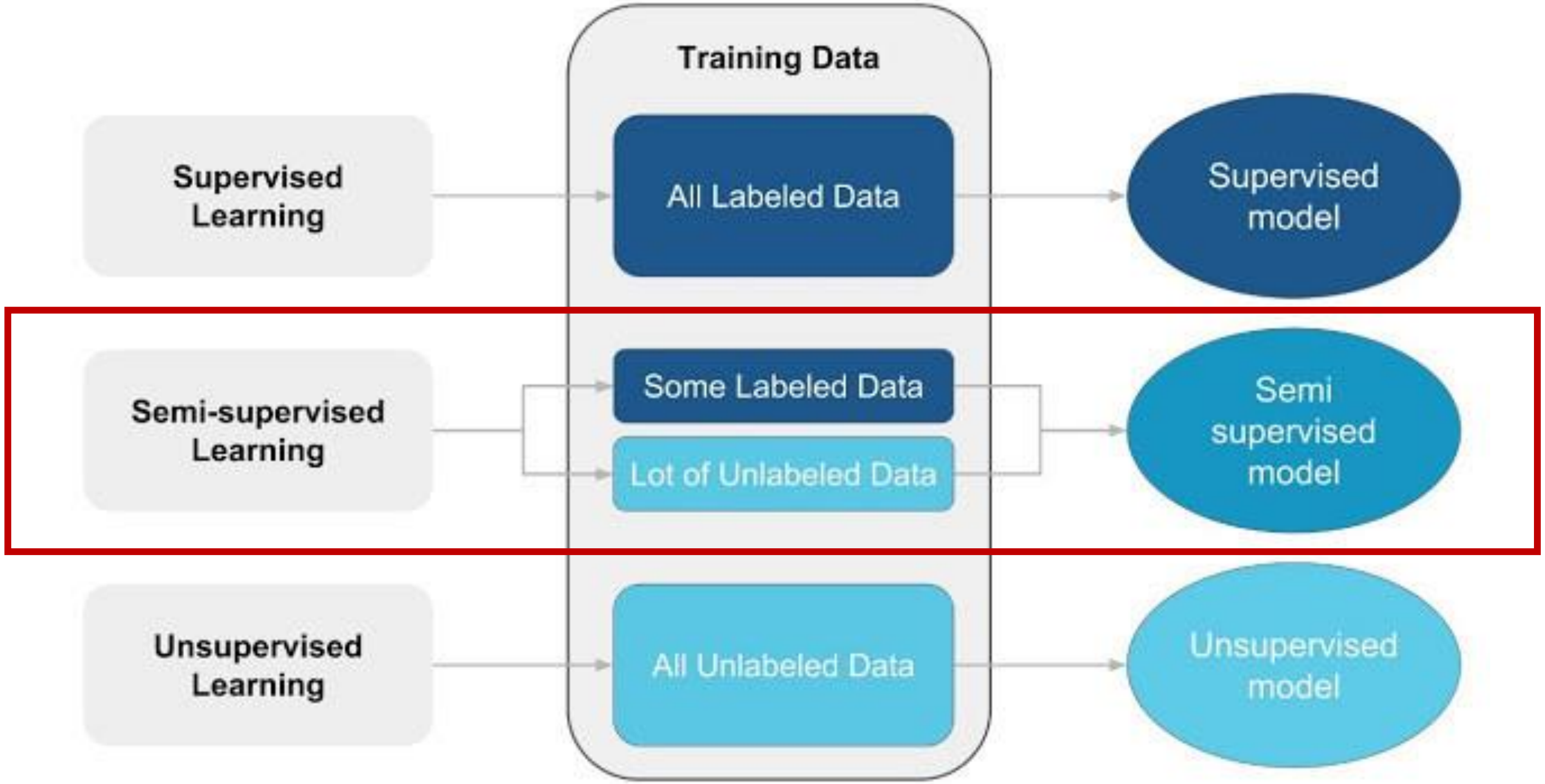
Problems in creating training data

- Annotation may need a lot of efforts, especially in the time of disaster
- Annotation is expensive
- Labeling may need experts
- Labeling may require specific tools (i.e. crowdsourcing platform)
- Time consuming and boring

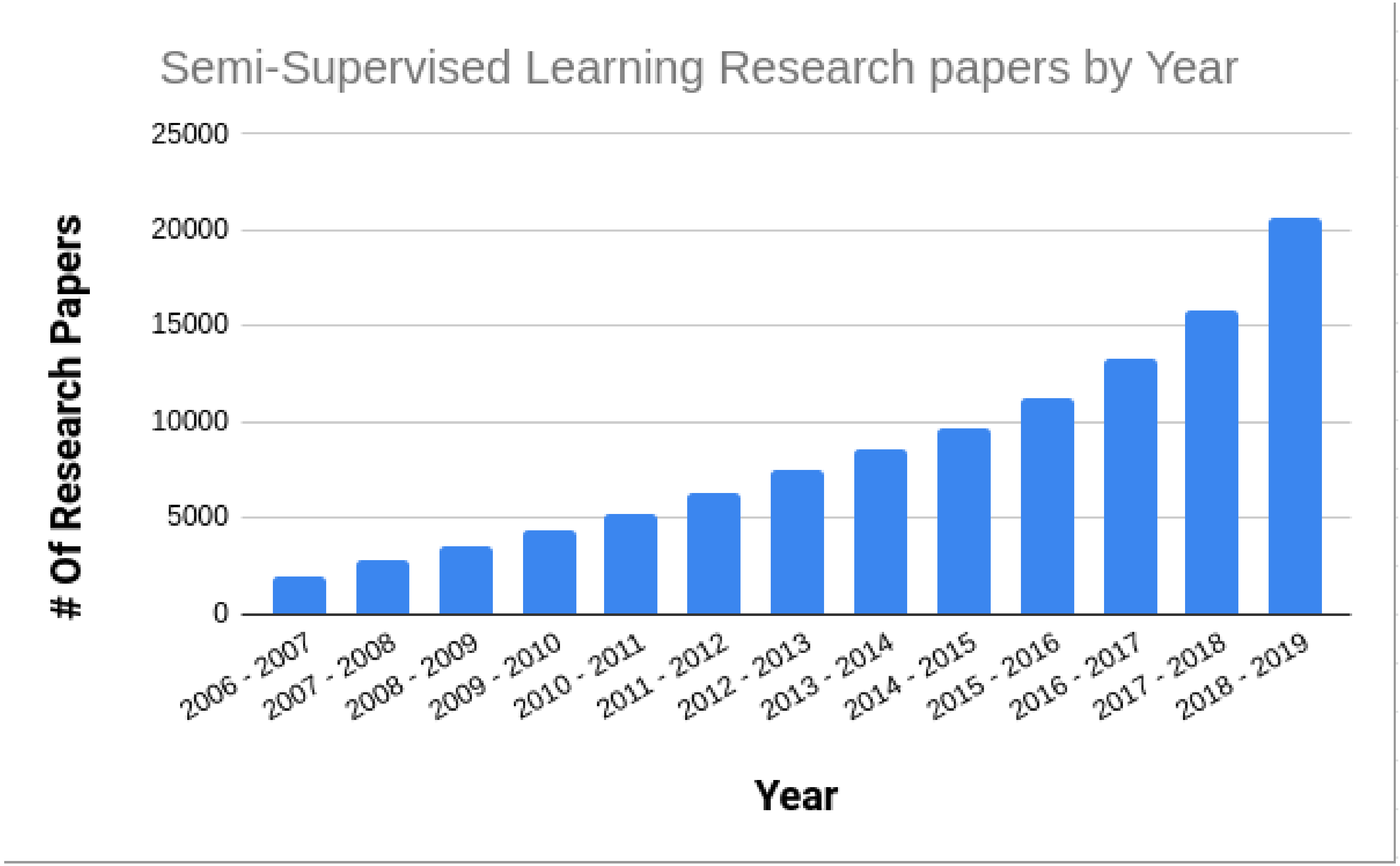
Fact: unlabeled data >> labeled data

Can we make use of them?

Leveraging unlabeled data



Pic: https://business.blogthinkbig.com/semi-supervised-learning-the-great-unknown/



Source: https://towardsdatascience.com/pseudo-labeling-to-deal-with-small-datasets-what-why-how-fd6f903213af

Do we learn from both labeled and unlabeled data?

Learning exists long before machine learning.
Do humans perform semi-supervised learning? Yes it seems!

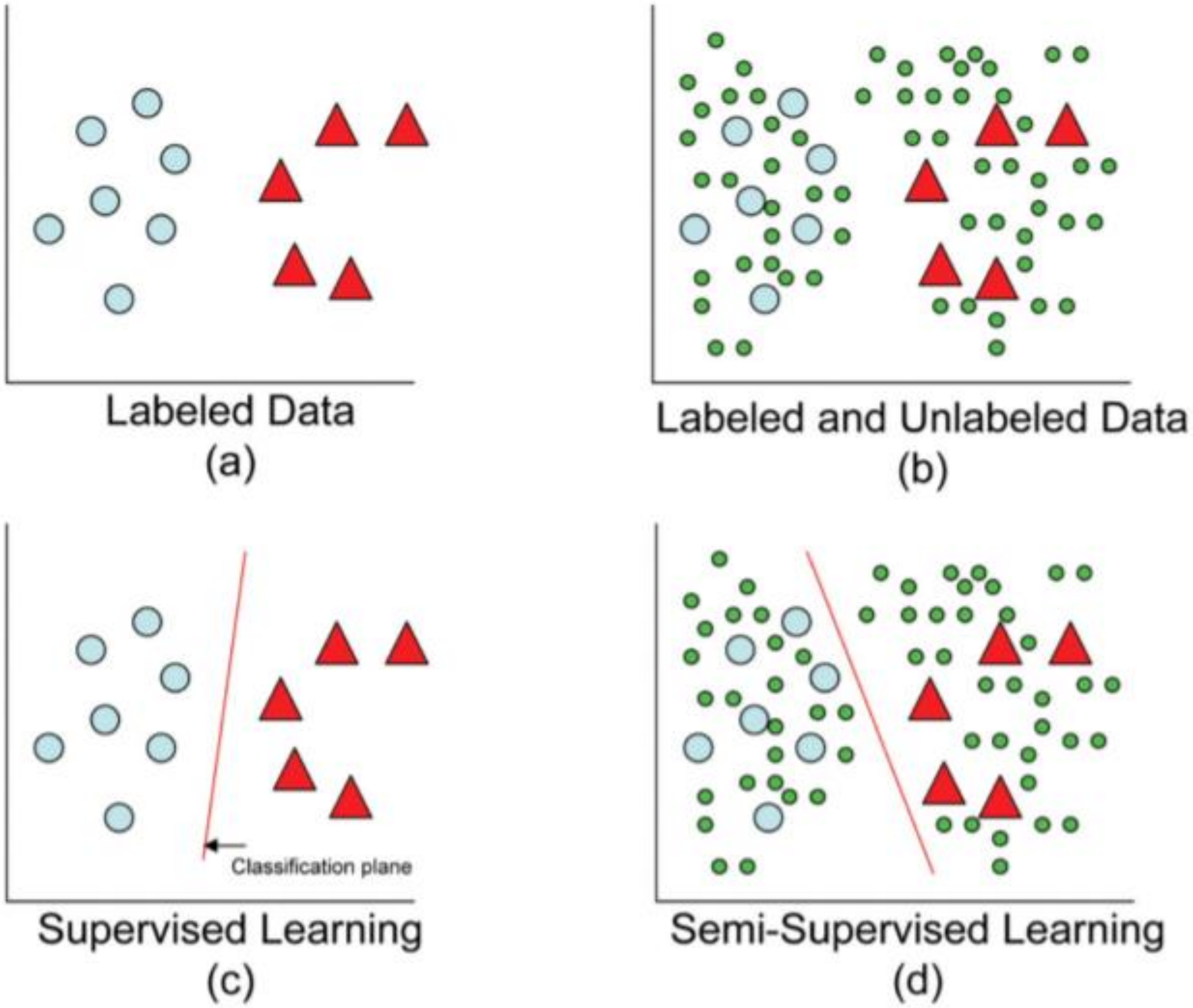
Cognitive Science

Computational model of how humans learn form labeled and unlabeled data

- Concept learning in children $x = animal,\ y = concept\ (e.g. cat)$
- Children learn their surrounding (ex: parents point out to furry white animal and say “cat”)
- Children also observer animal by themselves

How can unlabeled data be helpful?

Use accessible data to improve decision boundaries and better classify unlabeled data



Will be discussed:

- Self-training
- Pseudo labeling
- Multi-view Co-training

Notation

- Instance x label y
- Learner $f: X \rightarrow Y$
- Labeled data $(X_l, Y_l) = \{(x_{1:l}, y_{1:l})\}$
- Unlabeled data $X_u = \{x_{l+1:l+u}\}$, available during training, usually

$l \ll u$. Let $n = l + u$

- Test data $\{(x_{n+1\ldots}, y_{n+1\ldots})\}$, not available during training

Self Training

UNIVERSITAS GADJAH MADA

First proposed by Yarowsky (1995) as an approach to word sense disambiguation in text documents, predicting the meaning of words based on their context.

Input: labeled data $\{(x_i, y_i)\}_{i=1}^{l}$ and unlabeled data $\{x_j\}_{j=l+1}^{l+u}$

1. Initially, let $L = \{(x_i, y_i)\}_{i=1}^{l}$ and $U = \{x_j\}_{j=l+1}^{l+u}$
2. Repeat
   a. Train f from L using supervised learning
   b. Apply f to the unlabeled instances in U;
   c. Remove a subset S from U; add $\{(x, f(x)) | x \epsilon S\}$ to L

Xioajin Zhu (2007)

Assumptions

One’s own high confidence predictions are correct.

Variations of Self-Training

Add a few most confident $(x, f(x))$ to labeled data
Add all $(x, f(x))$ to labeled data
Add all $(x, f(x))$ to labeled data, weigh each by confidence

Notes

Self-training is a wrapper method, the choice of f is left completely open
Good for many real-world tasks like NLP
But mistake by f can reinforce itself

Xioajin Zhu (2007)

Pseudo-Labeling (Lee, 2013)

→ Self-training applied to Deep Neural Network

Consider multi-layer neural networks with M-layers of hidden units :

$$h_i^k = s^k \left(\sum_{j=1}^{d^k} W_{ij}^k h_j^{k-1} + b_i^k \right), \quad k = 1, ..., M+1$$

Where s^k is a non-linear activation function of the k −th layer such as sigmoid, $f_i = h_i^{M+1}$ are output units used for predicting target class and $x_j = h_j^0$ are input values.

Sigmoid activation function $s(x) = \frac{1}{1+e^{-x}}$

The whole network can be trained by minimizing supervised loss function

$$\sum_{i=1}^{C} L(y_i, f_i(x)),$$

where C is the number of labels, y_i's is the 1-of-K code of the label,
f_i is the network output for i'th label, x is input vector.
If we use sigmoid output unit, we can choose Cross Entropy as a loss function:

$$L(y_i, f_i) = -y_i \log f_i - (1 - y_i) \log(1 - f_i)$$

Pseudo-Label

Pseudo-Labels are target classes for unlabeled data as if they were true labels. Lee (2013) picked up the class which has maximum predicted probability for each unlabeled sample.

$$y'_i = \begin{cases} 1 & \text{if } i = \text{argmax}_{i'} f_{i'}(x) \\ 0 & \text{otherwise} \end{cases}$$

The pre-trained network is trained in a supervised fashion with labeled and unlabeled data simultaneously. The overall loss function is

$$L = \frac{1}{n} \sum_{m=1}^{n} \sum_{i=1}^{C} L(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^{C} L(y_i'^m, f_i'^m),$$

Loss function for labeled data

$$L = \boxed{\frac{1}{n}\sum_{m=1}^{n}\sum_{i=1}^{C} L(y_i^m, f_i^m)} + \alpha(t)\boxed{\frac{1}{n'}\sum_{m=1}^{n'}\sum_{i=1}^{C} L(y_i'^m, f_i'^m),}$$

Loss function for unlabeled data

a coefficient balancing loss of labeled and unlabeled data

Note: The total number of labeled and unlabeled data is quite different and the training balance between them is quite important for the network performance.

$$\alpha(t) = \begin{cases} 0 & t < T_1 \\ \frac{t-T_1}{T_2-T_1}\alpha_f & T_1 \le t < T_2 \\ \alpha_f & T_2 \le t \end{cases}$$

$\alpha\ (t)$ is slowly increased, is expected to help the optimization process to avoid poor local minima (Grandvalet et al., 2006) so that the pseudo-labels of unlabeled data are similar to true labels as much as possible.

with $f = 3, T_1 = 100, T_2 = 600$ without pretraining,
$T_1 = 200, T_2 = 800$ with DAE.

Multi-view Co-training (Blum and Mitchell, 1998)

- An extension of self-training to multiple supervised classifiers.
- In co-training, two or more supervised classifiers are iteratively trained on the labelled data, adding their most confident predictions to the labelled data set of the other supervised classifiers in each iteration.
- For co-training to succeed, it is important that the base learners are not too strongly correlated in their predictions.

Each instance is represented by two sets of features $x = [x(1); x(2)]$

- $x(1)$ = 1st feature
- $x(2)$ = 2nd feature
- This is a natural feature split (or multiple views)

Co-training idea:

- Train an image classifier and a text classifier
- The two classifiers teach each other

Xioajin Zhu (2007)

Assumptions:

- feature split $x = [x(1); x(2)]$ exists
- $x(1)$ or $x(2)$ alone is sufficient to train a good classifier
- $x(1)$ and $x(2)$ are conditionally independent given the class

Co-training algorithm

1. Train two classifiers: $f^{(1)}$from $\left(X_l^{(1)}, Y_l\right)$,$f^{(2)}$from $\left(X_l^{(2)}, Y_l\right)$
2. Classify X_u with$f^{(1)}$and$f^{(2)}$separately.
3. Add$f^{(1)}$'s k-most-confident $\left(x, f^{(1)}(x)\right)$ to $f^{(2)}$'s labeled data.
4. Add $f^{(1)}$'s k-most-confident $\left(x, f^{(2)}(x)\right)$ to $f^{(1)}$'s labeled data.
5. Repeat.

Xioajin Zhu (2007)

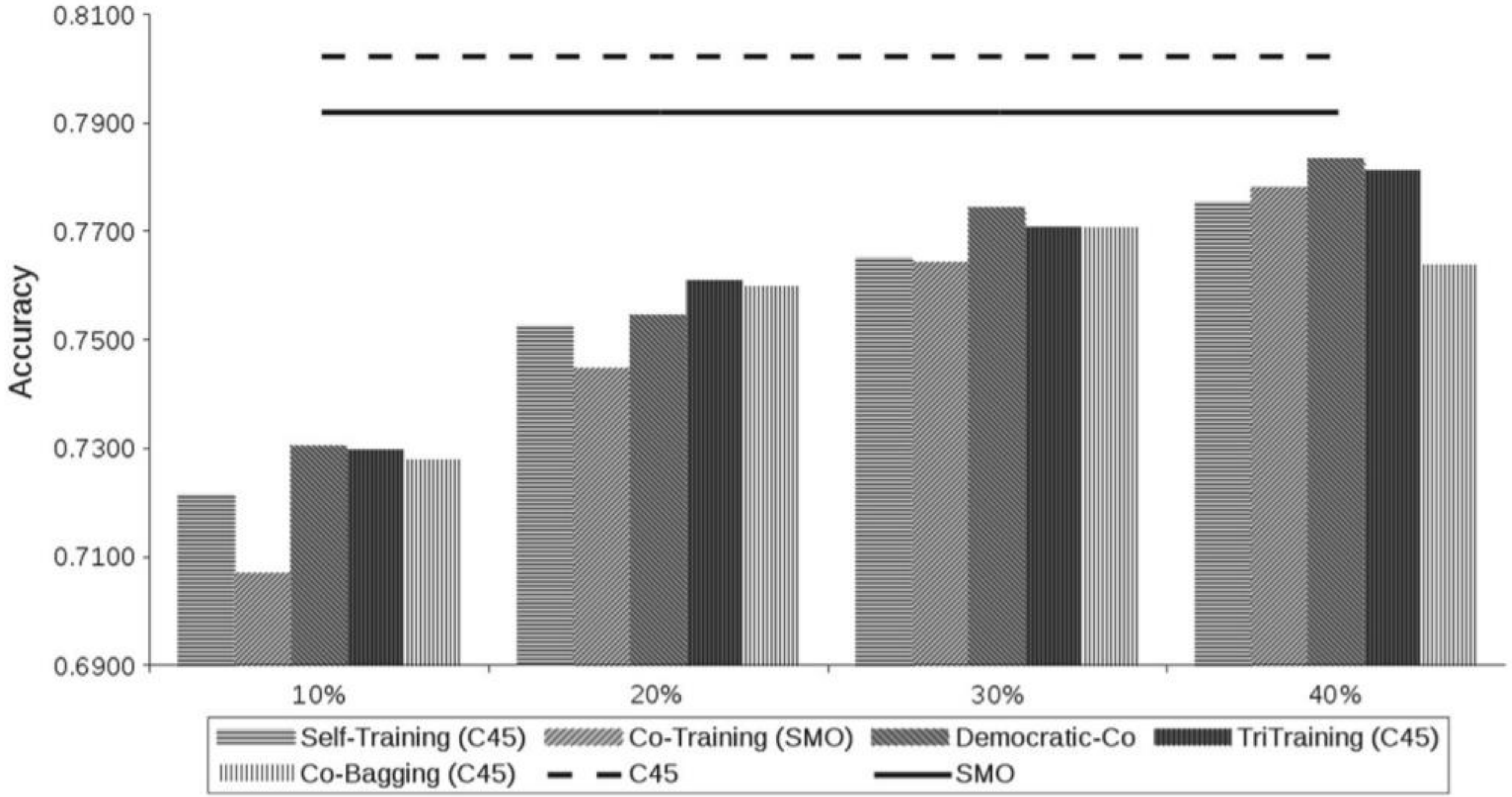
Pros

- Simple wrapper method. Applies to almost all existing classifiers
- Less sensitive to mistakes than self-training

Cons

- Natural feature splits may not exist
- Models using BOTH features should do better

Xioajin Zhu (2007)



Differences in average test results between outstanding models and C4.5 and SMO (I. Triguero et al, 2015)

Yan Lecun’s cake analogy

“Most of human and animal learning is unsupervised learning.

If intelligence is a cake, the bulk of the cake is unsupervised learning, the icing on the cake is supervised learning, and the cherry on the cake is reinforcement learning (RL).”

References

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting of the association for computational linguistics, association for computational linguistics* (pp. 189–196).

Lee, D. H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Proceedings of the 30th ICML workshop on challenges in representation learning* (Vol. 3, p. 2).

Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th annual conference on computational learning theory* (pp. 92–100). ACM.

Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using twitter. In EMNLP, pages 1568–1576.

Graham Neubig, Yuichiroh Matsubayashi, Masato Hagiwara, and Koji Murakami. 2011. Safety information mining — What can NLP do in a disaster—. In Proceedings of 5th International Joint Conference on Natural Language Processing, pages 965–973.

References

- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2013. Tweet analysis for real-time event detection and earthquake reporting system development. IEEE Transactions on Knowledge and Data Engineering, 25(4):919–931
- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In WWW, pages 851–860.
- Tetsuro Takahashi, Shuya Abe, and Nobuyuki Igata. 2011. Can twitter be an alternative of real-world sensors? In HCI (3), pages 240–249.
- Triguero, I., García, S.,&Herrera, F. (2015). Self-labeled techniques for semi-supervised learning: Taxonomy,software and empirical study. *Knowledge and Information Systems*, *42*(2), 245–284.
- Yves Grandvalet and Yoshua Bengio, Entropy Regularization, In: Semi-Supervised Learning, pages 151-168, MIT Press, 2006.
- Xioajin Zhu (2007), Semi-Supervised Learning Tutorial, ICML 2007
- J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” Machine Learning, vol. 109, no. 2, pp. 373–440, feb 2020