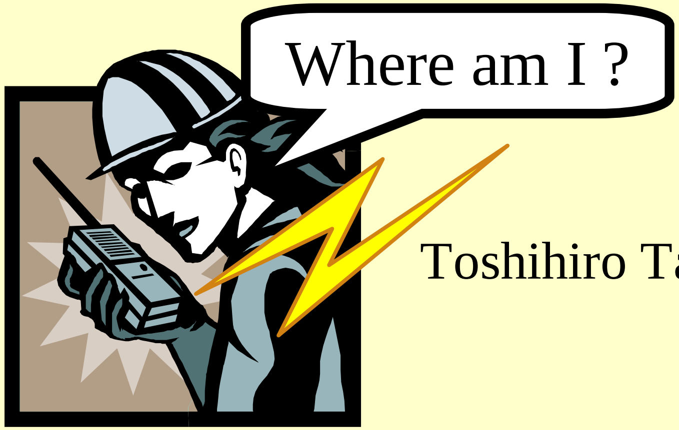
Tokyo Research Laboratory

Task 1:

A Semi-supervised Approach to Indoor Location Estimation



Hisashi Kashima, Shoko Suzuki, Shohei Hido, Yuta Tsuboi,
Toshihiro Takahashi, Tsuyoshi Ide, Rikiya Takahashi, and Akira Tajima

IBM Research, Tokyo Research Laboratory

Data Analytics Group

Presentation for task 1:
A Semi-supervised Approach to Indoor Location Estimation

- We formulated task1 as a semi-supervised learning [1] problem
- We employed the *label propagation* [2] as the semi-supervised learning method
  - A multi-class version of the label propagation method
  - Design of similarity measure using spatial information(=RSS values) and temporal information (=time stamps)

[1] X. Zhu. *Semi-supervised learning literature survey*. Technical Report, TR 1530, University of Wisconsin Madison, 2006.
[2] X. Zhu, Z. Ghahramani, and J. Lafferty. *Semi-supervised learning using gaussian fields and harmonic functions*. In *ICML*, 2003.

Task Review: Indoor location estimation by using machine learning

- Problem setting:
  - You want to know where you are in some building
  - In the building, there are several access points emitting radio signals
  - You have a client device by which you can know signal strength from each access point
- Difficulty: Triangulation is unsatisfactory because of high uncertainty in signals
- Solution: Apply machine learning techniques to estimate locations from received signal strengths

The task is formulated as a semi-supervised learning problem

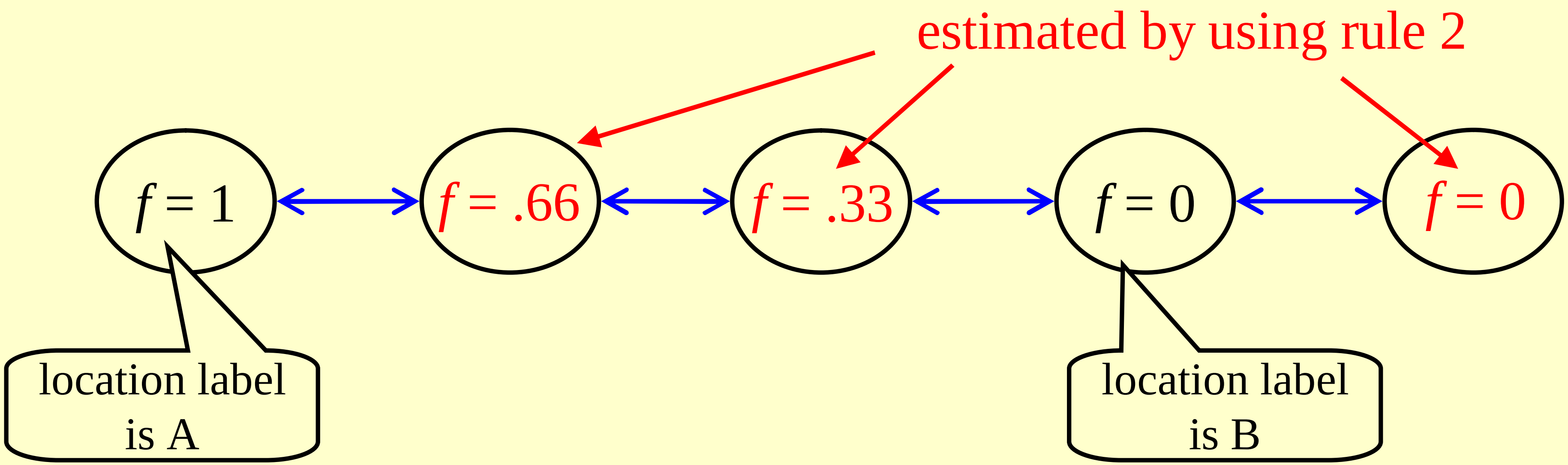
- Given: the i-th data is given as a tuple of $(\mathbf{x}^{(i)}, \mathrm{TID}^{(i)}, t^{(i)}, y^{(i)})$
  - spatial information: $\mathbf{x}^{(i)} \in \mathcal{X} = \mathbb{R}^{101}$ is the received signal strength (RSS) values
  - temporal information: $\mathrm{TID}^{(i)}$ (trace ID) and $t^{(i)}$ (time ID) indicate the time of the data observed
  - classs label: $y^{(i)} \in \mathcal{Y} = \{1, 2, ..., 247\}$ is a location label given only for a small fraction of the data
    - Semi-supervised learning problem
- Goal: predict $y^{(i)}$ for $\forall i \in$ UNLABELLED DATA whose location labels are " ? "(=not given)
  - Transduction problem

data ID (i)	trace ID (TID)	time ID (t)	location label ($y \in \{1, 2, ..., 247\}$)	RSS values ($\mathbf{x} \in [-100, 0]^{101}$) x_1	x_2	x_3	x_4	…	x_{101}
1		1	6	-58	-88	-100	-100	…	-80
2	1	2	?	-58	-95	-100	-100	…	-100
3		3	1	-65	-95	-100	-100	…	-75
4	2	1	23	-62	-83	-59		…	-93
5	3	1	?						-100
6		2	9						-100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Missing values are filled with -100 (the lowest RSS value)

We employed the *label propagation* as a semi-supervised learning method

- *Label propagation* tries to assign a location label to each observation with satisfying that
  1. labeled instances have the given labels, and
  2. similar instances have the similar class labels
- Example of two-class {A, B} case
  - f : the probability of the location label of the i-th instance being A
  - $(1 - f)$: the probability of the location label of the i-th instance being B
  - $ means "two observations are similar to each other"



We employed the *label propagation* as a semi-supervised learning method

- *Label propagation* tries to realize label assignments satisfying that
  1. labeled instances have the given labels, and
  2. similar instances have the similar class labels
- (Multi-class) label propagation is cast as an optimization problem

$$\textit{minimize}_{\mathbf{f}} \ \sum_{(i,j)} \mathrm{w}^{(i,j)} \sum_{y} \left(f^{(i)}(c) - f^{(j)}(c) \right)^2$$

where

- $f^{(i)}(c)$: the probability of the location label of the *i*-th instance being c
- $\mathrm{w}^{(i,j)}$: the similarity measure between the *i*-th and *j*-th examples

s.t. for each labeled instance i,

- $f^{(i)}(c) = 1$, if c is the true class label
- $f^{(i)}(c) = 0$, otherwise

- Prediction is made by $\mathrm{argmax}_c \ f^{(i)}(c)$ for each i
- Instead of a closed form solution requiring the inverse of a large matrix, we can use the following simple iterative update

$$f^{(i)}(c) \quad (\quad \left(\sum_j \mathrm{w}^{(i,j)} \sum_c f^{(j)}(c) \right) \Big/ \left(\sum_j \mathrm{w}^{(i,j)} \right)$$

Similarity measure $w^{(i,j)}$ is defined by RSS values and time stamps

- We have to define the *similarity measure* $w^{(i,j)}$ used in the label propagation
- Each instance is accompanied by two types of information
  1. spatial information: RSS values
  2. temporal information: a time stamp
- Two instances are similar if
  – their RSS values are similar, or
  – their time stamps are similar
- The similarity measure is defined by the maximum of two similarity measures

$$w^{(i,j)} = \max \{ w_X^{(i,j)},\ w_T^{(i,j)} \}$$

where

- $w_X^{(i,j)}$: similarity based on spatial information (=RSS values)
- $w_T^{(i,j)}$: similarity based on temporal information (=time stamps)

Robust similarity measure based on spatial information: $w_X^{(i,j)}$

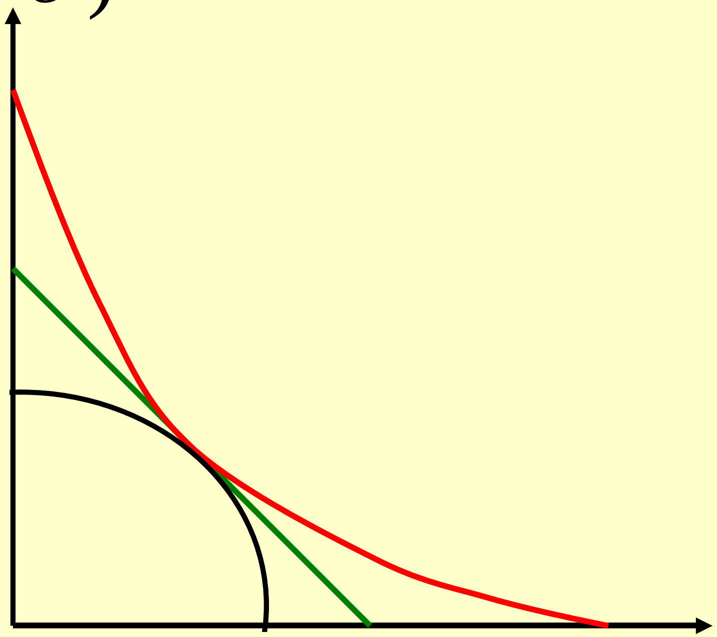
- Since RSS values are noisy,
  we need a similarity robust to noise caused by reflection, interference, and shielding
- RSS-based similarity $w_X^{(i,j)}$ is defined as

$$w_X^{(i,j)} = \exp\left(- \| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \|_p \; / \; \sigma \right)$$

where

- $\|\cdot\|_p$ is the p-norm (in submission, p = 0.5 (0.5-norm))
- σ is a constant scale parameter (in submission, σ = 0.5)

- We used p-norm with $p < 1$, which puts more importance on presence/absence of signals than the amount of change
  - Robust to drastic change of each RSS value
  - Sensitive to change of multiple RSS values



level curve of p-norm

Similarity measure based on temporal information: $w_T^{(i,j)}$

- Time-stamp-based similarity $w_T^{(i,j)}$ is defined as

$$w_T^{(i,j)} = \rho, \text{ if } i \text{ and } j \text{ are consecutive observations in a trace}$$
$$= 0, \text{ otherwise}$$

  – In submission, we used $\rho = 1$

- Probably, we could improve the similarity further ...
  – $\rho = 0.01$ performs better
  – Similarity function like that for RSS values

So, what was most important for performance improvement ? Design of similarity function is crucial

- Design of similarity function contributed most to improvement of prediction accuracy
  - Use of 0.5-norm in RSS similarity
  - Use of time-stamp-based similarity

- Nearest neighbour with 2-norm RSS similarity (baseline)

⇓ **+7% accuracy**

- Nearest neighbour with **0.5-**norm RSS similarity

⇓ + 1% accuracy

- **Label propagation** with 0.5-norm RSS similarity

⇓ **+ 5% accuracy**

- Label propagation with 0.5-norm RSS similarity and **time stamp similarity**

Conclusion and future work

- We applied a multi-class version of the label propagation to this task
- We designed a similarity measure using spatial information(=RSS values) and temporal information (=time stamps)
  - Metric design >> semi-supervised learning
- It is very difficult to beat the simple methods such as kNN

- Possible future work includes
  - Refinement of the time-based similarity
  - Out-of-sample prediction
    - In real situation, test data are not given in advance of test phase
    - Approximation or explicit learning of the mapping function

Thank you

WE ARE...
Hisashi Kashima, Shoko Suzuki, Shohei Hido, Yuta Tsuboi,
Toshihiro Takahashi, Tsuyoshi Ide, Rikiya Takahashi, and Akira Tajima
IBM Research, Tokyo Research Laboratory
Data Analytics Group