A Semi-supervised Approach to Indoor Location Estimation

Hisashi Kashima, Shoko Suzuki, Shohei Hido, Yuta Tsuboi,
Toshihiro Takahashi, Tsuyoshi Idé, Rikiya Takahashi, and Akira Tajima
IBM Research
Tokyo Research Laboratory
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502, Japan

Abstract

We describe the approach and the algorithm that we used for Task1 "Indoor Location Estimation" in ICDM Data Mining Contest 2007. We formulated the task as a transduction problem, and employed a semi-supervised learning approach based on a multi-class version of the label propagation method.

1 A formulation as a transduction problem

We formulate the indoor location estimation problem as a transductive multi-class classification problem.

Let the whole dataset consists of $N = 5,333$ instances, where the first $\ell = 505$ instances are the labeled data, and the rest of them are the unlabeled data. The i-th data is given as $(\mathbf{x}^{(i)}, y^{(i)})$, where $\mathbf{x}^{(i)} \in \mathbb{R}^{101}$ is the vector of the received signal strength (RSS) values from the WiFi Access Points (APs), and $y^{(i)} \in \{1, 2, \ldots, 247\}$ is the location label assigned to the RSS vector. As for the unobserved RSS values, we filled them by -100, since all RSS values are in the range of $[-100, 0]$ and unobserved RSS value implies that it was too weak to detect.

In addition to the RSS values, some portion of the instances are given trace IDs and observation times, which indicate the trace (i.e. experiment) each instance belongs to and the time when it was observed. We denote by $TID^{(i)}$ and $t^{(i)}$ the trace ID and the observation time of the i-th instance, respectively. For simplicity, we treat the observation times as just the orders of observation, i.e. integer values, although they are given as real values in the original data.

The objective of this task is to predict the location labels of the unlabeled data, $y^{(\ell+1)}, \ldots, y^{(N)}$, since they are not given. Note that since the inputs of the test data set are given in advance of the test phase, we can regard the problem as a transduction problem where test inputs are explicitly used.

2 A semi-supervised approach using label propagation

We employ a multi-class version of the *label propagation* method [2], which is one of the supervised learning approaches [1].

Let $f^{(i)}(c) \in [0, 1]$ indicate the probability with which the location label of the i-th instance is c. For the labeled data ($i \leq \ell$), the following holds

$$f^{(i)}(c) = \begin{cases} 1 & (\text{if } c = y^{(i)}) \\ 0 & (\text{otherwise}) \end{cases} . \tag{1}$$

The task is to predict $f^{(i)}(y^{(i)})$ for $i > \ell$ and $\forall c$, with which we obtain the prediction $\hat{c}^{(i)}$ for $i > \ell$ by

$$\hat{c}^{(i)} = \operatorname*{argmax}_{c} f^{(i)}(c). \tag{2}$$

In the label propagation framework, we try to minimize the discrepancies of the label distributions among neighbourhood instances, which is defined as

$$\sum_{(i,j)} w^{(i,j)} \sum_{c} \left(f^{(i)}(c) - f^{(j)}(c) \right)^2 ,$$

where $w^{(i,j)}$ is a constant called the *affinity* indicating the similarity between the i-th instance and the j-th instance, which we will define later. It is easy to see that the solution of the above optimization problem satisfies

$$f^{(i)}(c) = \frac{\sum_j w^{(i,j)} f^{(j)}(c)}{\sum_j w^{(i,j)}}, \tag{3}$$

for $\forall i > \ell$ and $\forall c$. Therefore, instead of solving the large optimization problem directly, we can iteratively apply (3) to make local updates of predictions until convergence.

The affinity $w^{(i,j)}$ is defined as the maximum of either $w_{\mathbf{x}}^{(i,j)}$ defined by RSS vectors and $w_t^{(i,j)}$ defined by trace

IDs and observation times, i.e.

$$w^{(i,j)} = \max\left\{w_{\mathbf{x}}^{(i,j)}, w_t^{(i,j)}\right\}. \tag{4}$$

Our definition of the affinity tries to imply that two instances are similar if their RSS vectors are similar or their observation times are consecutive.

For the affinity between two RSS vectors $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(i)}$, we used a heat-kernel like function

$$w_{\mathbf{x}}^{(i,j)} = \exp\left(-\frac{\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \|_q{}^q}{\sigma}\right), \tag{5}$$

where σ is a scale parameter, and we set $\sigma = 0.5$ in our submission. Also, $\| \cdot \|_q$ is the q-norm which is defined as

$$\| \mathbf{x} \|_q = \left(\sum_d |x_d|^q\right)^{\frac{1}{q}},$$

and we set $q = 0.5$ based on the observation that this choice performed well in our preliminary analysis using the nearest neighbour classifier for the labeled data.

The affinity between two pairs of a trace ID and an observation time is defined as

$$w_t^{(i,j)} = p \cdot \delta\left(TID^{(i)} = TID^{(i)}\right) \cdot \delta\left(|t^{(i)} - t^{(i)}| = 1\right), \tag{6}$$

where $p \in [0, 1]$ is a constant indicating the affinity of two consecutive observations, and we simply set $p = 1$ in the submission. Also, σ is a function that returns 1 if its argument is true, and 0 if otherwise.

3 The algorithm

Based on the discussion in the previous section, each step of the algorithm is summarized as follows.

1. Initialize $f^{(i)}$ by using (1) for the labeled instances ($i \leq \ell$).

2. Compute the affinities between all pairs of instances by using (4), (5) and (6).

3. Continue the follwing steps, 3a and 3b, until convergence.

   (a) Select $i > \ell$ uniformly at random.

   (b) Update $f^{(i)}(c)$ for $\forall c$ by using (3).

4. Output prediction for i in the test data by using (2).

References

[1] X. Zhu. Semi-supervised learning literature survey. Technical Report Computer Sciences, 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 *Proceedings of The 20th International Conference on Machine Learning*, 2003.