

RFIDによる画像識別モデルパラメータの自動学習

Learning Model Parameters of Object Image Classification Using RFID

白坂 洋平*¹ 柴田 潤一*¹ 金崎 弘文*² 矢入 健久*² 町田 和雄*²
 Yohei Shirasaka Junichi Shibata Hirofumi Kanazaki Takehisa Yairi Kazuo Machida

*¹東京大学大学院 工学系研究科 航空宇宙工学専攻
 Department of Aeronautics and Astronautics, School of Engineering, University of Tokyo

*²東京大学先端科学技術研究センター
 Research Center for Advanced Science and Technology

For position estimation of multiple objects, a construction of algorithms for integrating information from various types of sensors is one of the most significant technologies. Among these sensors, vision sensors are indispensable for position estimation. But we need large amount of training data for constructing classifiers. This paper proposes a framework for reducing the cost and time of building the visual object classifiers. In this method, the RFID device is used as a supervisor to semiautomatically construct the training data set for each objects. We cannot use the auxiliary information of RFID directly as the perfect training data because of the observation errors. In this research, we construct a semi-supervised object image learning method using EM algorithm and verify the effectiveness of our method.

1. はじめに

異種のセンサ情報を統合して、物体位置を推定する技術は、屋内環境での物品管理や、移動ロボットが活動する上での基盤技術といえる。活用されるセンサ群の中でも、画像センサと識別センサは有用なセンサとして注目されている。画像センサは物体の高度な位置計測精度を示すが、物体検出、識別性能を向上させるには事前調整が必要となる。一方 RFID Tag Reader に代表される識別センサは、極めて正確な物体識別性能を示すが、詳細な位置計測が困難であるなどの問題点もある。

このように異種なセンサ間では、物体識別性能、位置計測精度において、その優位性に差がある。各センサが相対的に優位なセンサ情報を、他センサと共有して観測モデル学習に活用すれば、事前に構築されたモデルを実環境に対しロバストに調整できると考えられる。また、appearance based な画像識別学習を行う場合、事前に大量の教師データを必要とするが、RFID 情報を付加すれば、自動的に教師データを形成することができる。本研究では、センサ情報の質的価値に対し相補的な性質をもつ画像センサと識別センサ (RFID) に関して、過去の両センサの観測情報を統合して観測モデルパラメータを環境適応的に自動調整する枠組みを提案する。

識別センサを用いた画像センサの物体認識モデルのチューニングは、物体に付加された RFID の識別データを用いた画像物体識別学習と解釈できる。RFID による補助データは読み落としなどの観測ノイズ、画像センサによる観測では未検出や誤識別などの観測ノイズが生じることが予想される。また、RFID が付加されていない物体が画像中に混入する状況や、同時に複数物体の観測を得る状況などを考慮すると、画像データと RFID 情報を正確に対応付けることは困難であり、完全な教師データとして扱うことはできないといえる。本研究では、画像センサの観測データと、RFID の不完全な ID 補助データからの識別学習問題を、不完全な教師付学習の枠組みで取り扱う。こうした意味で、本問題は半教師付き学習 [Nigam 98, Chapelle 06]

と関連性が高い。

2. 問題定義

2.1 環境及び物体特徴に関する前提条件

RFID Tag が貼り付けられた物体 (登録物体) 及び貼り付けられていない物体 (未登録物体) の、物体数をそれぞれ n, m とする。位置推定の対象となる空間的領域を \mathcal{D} とし、領域 \mathcal{D} を、画像センサ及び識別センサ (RFID) の観測対象領域が包括しているものとする。物体 i の形状及び色などの潜在的特徴を表すパラメータを O_i とし、時刻 t における照明・背景などの環境条件を表すパラメータを E_t とする。

2.2 観測情報

時刻 t における RFID Tag Reader の観測値を D_t とし、 $D_t = (d_{t,1}, \dots, d_{t,n})^T$ なるベクトルで表現する。ここで、 $d_{t,i} = 1$ は、物体 i が RFID Tag Reader によって読み取られたことを意味し、 $d_{t,i} = 0$ は、読み取られなかったことを意味することとする。時刻 t における画像センサの観測情報は、添え字 j を用いて $Y_{t,j}$ とする。各々の $Y_{t,j}$ は、画像センサによって検出された物体を指すと思われる画像特徴量を表し、それぞれの観測データを、添え字 $j = 1, \dots, m_t$ を用いて区別する。尚、 m_t は時刻 t において検出された、物体とおぼしき画像特徴量の総数である。また、 $Y_t = (Y_{t,1}, \dots, Y_{t,m_t})$ と定義する。

2.3 隠れ変数 (欠損情報)

上では、問題の前提条件となる潜在パラメータ n, m, O_i, E_t と、観測情報 $D_t, Y_{t,j}$ を定義した。本研究では他に、直接観測することはできないが、モデル構造の定式化に有用な欠損データを定義する。 L_t は、時刻 t において対象領域 \mathcal{D} 内に入っている物体リストを表すベクトル $L_t = (l_{t,1}, \dots, l_{t,n})^T$ とする。ここで、 $l_{t,i} = 1$ は、物体 i が領域 \mathcal{D} 内に入っていることを意味し、 $l_{t,i} = 0$ は、 \mathcal{D} 外にあることを意味することとする。また、画像センサによる複数の物体を対象とする推定問題の場合は、観測によって得られたデータと、ターゲット物体の ID との対応付け (データアソシエーション) が自明ではない。本論文では、データアソシエーションを記述するため、変数 $r_{t,j}$ を

導入する。ここで、 $r_{t,j} = i$ は時刻 t に画像センサが観測したデータ $Y_{t,j}$ が、物体 i に対応していることを意味する。

2.4 観測モデルパラメータ

本論文では、学習すべき観測モデルパラメータとして、RFID Tag Reader の検出率 δ_i と、画像センサによる物体の識別パラメータ θ_i を導入する。 δ_i は、物体 i が対象領域 D 内に存在するときに、RFID Tag Reader が物体 i の RFID Tag を検出できる確率を表す。また θ_i は、画像センサが物体 i を観測したときに、物体の特徴量を表現するパラメータであり、全物体に関するパラメータ $\theta_1, \dots, \theta_{n+m}$ を用いることによって、画像センサから得られる特徴量から、その物体 ID を識別させることを前提としている。

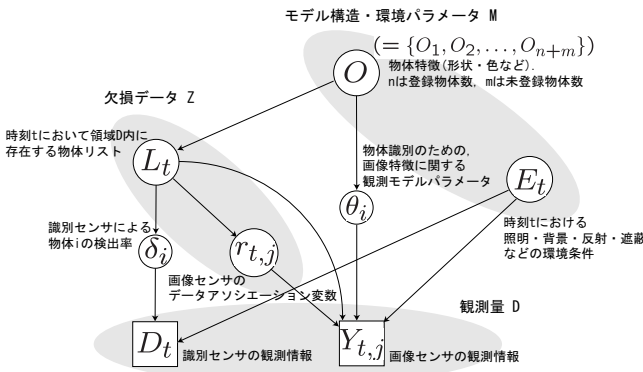


図 1: 相互学習問題におけるデータ間のモデル構造

2.5 目的

本研究では、前提となる図 1 のモデル構造において、欠損データ $r_{t,j}, L_t$ が存在する条件下で、観測情報 $Y_{1:t}$ 及び $D_{1:t}$ から、画像センサ及び識別センサの観測モデルパラメータ θ_i, δ_i を学習することを目的とする。図 2 に問題設定の概略図を示す。

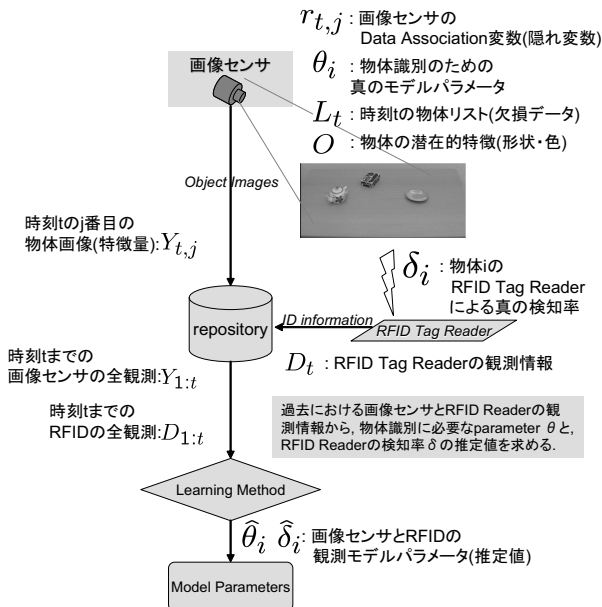


図 2: 問題設定の概略

3. 学習手法

3.1 手法の選択

本研究で扱う問題は、直接観測できない隠れ変数 $L_{1:T}, r_{1:T}$ が存在するという前提で、観測情報 $D_{1:T}, Y_{1:T}$ を最もうまく説明できるパラメータ θ, δ を見つける、最適化問題と解釈できる。EM アルゴリズムは、潜在変数をもつ統計モデルの、モデルパラメータの最尤推定値を効率的に求めるための一般的数値解法である [McLachlan 96]。本研究では、EM アルゴリズムを用いた推定アルゴリズムを定式化する。

3.2 EM アルゴリズムによるパラメータ推定法

問題を隠れ変数 $Z = (L_{1:T}, r_{1:T})$ 、観測データ $D = (D_{1:T}, Y_{1:T})$ 、求めるべきモデルパラメータ $\Psi = (\theta, \delta)$ として手法を定式化する。まず、完全データの対数尤度の期待値 Q は、

$$Q(\Psi|\Psi^{(\tau)}) = E[\log p(D, Z; \Psi) | D; \Psi^{(\tau)}] \\ = \sum_{t=1}^T \sum_{j_t=1}^{m_t} \sum_{r_{t,j_t}=1}^{n+m} \sum_{L_t} \left[P(r_{t,j_t}, L_t | D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)}) \times \log p_c(D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)}) \right] \quad (1)$$

となり、これを最大化する $\theta^{(\tau)}, \delta^{(\tau)}$ を求める。ここで、 $p_c(\cdot)$ とは、完全データの確率モデルを表す。以下、 $p_c(D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)})$ を完全データの尤度関数、 $P(r_{t,j_t}, L_t | D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)})$ を隠れ変数の事後分布と呼び、これらをモデル化する。

3.3 完全データの尤度関数

確率分布 $p_c(D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)})$ をモデル化するにあたり、まず観測情報の独立性を仮定して、

$$p_c(D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)}) = p_c(Y_{1:T}; \theta^{(\tau)}) p_c(D_{1:T}; \delta^{(\tau)}) \quad (2)$$

とする。

$p_c(Y_{1:T}; \theta^{(\tau)})$ は、欠損データ r_{t,j_t}, L_t が与えられた際の、画像識別に関する尤度関数であると解釈できる。 $Y_{1:T}$ は極めて高次元な特徴データであるため、本研究では Kernel 主成分分析によって Y_{t,j_t} を、より低次元なデータ f_{t,j_t} へ変換し、このデータ f_{t,j_t} が、次式のような混合正規分布モデルに従うとする。

$$p(f; \theta) = \sum_{i=1}^{n+m} \alpha_i \mathcal{N}(f; \mu_i, \sigma_i), \alpha_i > 0, \sum_{i=1}^{n+m} \alpha_i = 1 \quad (3)$$

ここで、 $\mathcal{N}(\cdot; \mu, \sigma)$ は平均ベクトル μ 、共分散行列 σ からなる正規分布を表す。

$p_c(D_t; \delta^{(\tau)})$ は、欠損データ r_{t,j_t}, L_t が与えられた際の、RFID Tag の検知率に関する尤度関数である。これを、

$$p_c(D_t; \delta^{(\tau)}) = \prod_{i=1}^{n+m} q_i \quad (4)$$

where

$$q_i = \begin{cases} \delta_i & (D_{t,i} = 1, l_{t,i} = 1) \\ 1 - \delta_i & (D_{t,i} = 0, l_{t,i} = 1) \\ 0 & (D_{t,i} = 1, l_{t,i} = 0) \\ 1 & (D_{t,i} = 0, l_{t,i} = 0). \end{cases} \quad (5)$$

と、モデル化する。

3.4 隠れ変数の事後分布

$P(r_{t,j_t}, L_t | D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)})$ は、観測モデルパラメータ $\theta^{(\tau)}, \delta^{(\tau)}$ と、観測データ $D_{1:T}, Y_{1:T}$ が与えられた条件における、欠損データの確率分布を表す。これを、

$$\begin{aligned} & P(r_{t,j_t}, L_t | D_{1:T}, Y_{1:T}; \theta^{(\tau)}, \delta^{(\tau)}) \\ & \propto P(D_{1:T} | r_{t,j_t}; \theta^{(\tau)}, \delta^{(\tau)}) p(Y_{1:T} | r_{t,j_t}; \theta^{(\tau)}, \delta^{(\tau)}) \\ & \quad \times P(D_{1:T} | L_t; \theta^{(\tau)}, \delta^{(\tau)}) p(Y_{1:T} | L_t; \theta^{(\tau)}, \delta^{(\tau)}) \\ & \quad \times P(r_{t,j_t}; \theta^{(\tau)}, \delta^{(\tau)}) P(L_t; \theta^{(\tau)}, \delta^{(\tau)}) \end{aligned} \quad (6)$$

と展開する。

$P(D_{1:T} | r_{t,j_t}; \theta^{(\tau)}, \delta^{(\tau)})$ は、画像センサ観測情報 Y_{t,j_t} のデータアソシエーション変数 $r_{t,j_t} = i$ が与えられているときの確率モデルである。本研究では、

$$P(D_{1:T} | r_{t,j_t} = i; \theta^{(\tau)}, \delta^{(\tau)}) \propto \beta_i \quad (7)$$

where

$$\beta_i = \begin{cases} \delta_i^{(\tau)} & (D_{t,i} = 1) \\ 1 - \delta_i^{(\tau)} & (D_{t,i} = 0) \end{cases} \quad (8)$$

とモデル化する。

$p(Y_{1:T} | r_{t,j_t}; \theta^{(\tau)}, \delta^{(\tau)})$ は、3.3 でのモデル化に基づき、

$$p(Y_{1:T} | r_{t,j_t} = i; \theta^{(\tau)}, \delta^{(\tau)}) = \alpha_i^{(\tau)} \mathcal{N}(f_{t,j_t}; \mu_i^{(\tau)}, \sigma_i^{(\tau)}) \quad (9)$$

とモデル化する。

$P(D_{1:T} | L_t; \theta^{(\tau)}, \delta^{(\tau)})$ に関しては、 $L_t = (l_{t,1}, \dots, l_{t,n})^T$ が与えられたときの、RFID Tag Reader の観測値を表す確率モデルであると解釈できるので、以下のようにモデル化することができる。

$$P(D_{1:T} | L_t; \theta^{(\tau)}, \delta^{(\tau)}) = \prod_{i=1}^{n+m} q_i^{(\tau)} \quad (10)$$

尚 $q_i^{(\tau)}$ は、5 式に基づく。

3.5 最尤推定量の導出

以上のモデル化から完全データの対数尤度の期待値 $Q(\theta | \theta^{(\tau)})$ は、

$$\begin{aligned} Q(\Psi | \Psi^{(\tau)}) & \propto \sum_{t=1}^T \sum_{j_t=1}^{m_t} \sum_{r_{t,j_t}=1}^{n+m} \sum_{L_t} \left[\beta_i^{(\tau)} \times \alpha_i^{(\tau)} \mathcal{N}(f_{t,j_t}; \mu_i^{(\tau)}, \Sigma_i^{(\tau)}) \right. \\ & \quad \left. \times \prod_{i=1}^{n+m} q_i^{(\tau)} \times \log(\alpha_i \mathcal{N}(f_{t,j_t}; \mu_i, \Sigma_i) \times \prod_{i=1}^{n+m} q_i) \right] \quad (11) \end{aligned}$$

とまとめることができる。

$\frac{\partial}{\partial \mu_i} Q(\Psi | \Psi^{(\tau)}) = 0, \frac{\partial}{\partial \Sigma_i} Q(\Psi | \Psi^{(\tau)}) = 0$ を解くことにより、 μ_i, Σ_i の最尤推定量 $\mu_i^{(\tau+1)}, \Sigma_i^{(\tau+1)}$ は、

$$\mu_i^{(\tau+1)} = \frac{1}{N_i^{(\tau)}} \sum_{t=1}^T \sum_{j_t=1}^{m_t} \sum_{r_{t,j_t}=1}^{n+m} \left[\beta_i^{(\tau)} \times \alpha_i^{(\tau)} \mathcal{N}(f_{t,j_t}; \mu_i^{(\tau)}, \Sigma_i^{(\tau)}) \times f_{t,j_t} \right] \quad (12)$$

$$\begin{aligned} \Sigma_i^{(\tau+1)} & = \frac{1}{N_i^{(\tau)}} \sum_{t=1}^T \sum_{j_t=1}^{m_t} \sum_{r_{t,j_t}=1}^{n+m} \left[\beta_i^{(\tau)} \times \alpha_i^{(\tau)} \mathcal{N}(f_{t,j_t}; \mu_i^{(\tau)}, \Sigma_i^{(\tau)}) \right. \\ & \quad \left. \times (f_{t,j_t} - \mu_i^{(\tau+1)})(f_{t,j_t} - \mu_i^{(\tau+1)})^T \right] \end{aligned} \quad (13)$$

と計算される。ここで、

$$N_i^{(\tau)} = \sum_{t=1}^T \sum_{j_t=1}^{m_t} \sum_{r_{t,j_t}=1}^{n+m} \left[\beta_i^{(\tau)} \times \alpha_i^{(\tau)} \mathcal{N}(f_{t,j_t}; \mu_i^{(\tau)}, \Sigma_i^{(\tau)}) \right] \quad (14)$$

である。また、 $\alpha_i^{(\tau+1)}$ に関しては、制約条件 $\sum_i \alpha_i = 1$ のもとで、 Q の α_i に関する最大化より、 $\alpha_i^{(\tau+1)} = N_i^{(\tau)} / \sum_i N_i^{(\tau)}$ を得る。

次に、RFID 検知率の最尤推定量を導出する。 $\mu_i^{(\tau+1)}$ と同様に、 $\frac{\partial}{\partial \delta_i} Q(\Psi | \Psi^{(\tau)}) = 0$ を解くことにより求めた δ_i の最尤推定量は、解析的に解くことが困難である。そこで本研究では、より直感的に導出された検知率を近似解として以下のように決定する。まず、画像識別モデル θ_i によって各観測情報 $Y_{t,j}$ の最有力 ID を求める。次に、RFID による観測情報 D_t 及び $Y_{t,1}, \dots, Y_{t,m_t}$ の最有力ラベルから、時刻 t のリスト変数の推定値 \hat{L}_t を求める。そして最後に、 $\delta_i = \sum_t D_{t,i} / \sum_t \hat{L}_{t,i}$ として検知率の推定値を決定する。

4. 実験結果

4.1 実験システム

観測対象領域 \mathcal{D} は、家庭用テーブル上とした。CCD カメラは、テーブル上を斜め上から見下ろす位置に設置し、RFID Tag Reader は、テーブル下に上向きに設置した。対象物体には、登録物体 (RFID 付き)10 種類と、未登録物体 (RFID なし)5 種類を選んだ。対象物体及びその ID を図 4 に示す。図 3 に CCD Camera によって取得した画像を示す。



図 3: CCD Camera で取得した画像 (左: [照明 1], 中央: [照明 2], 右: [照明 3]).

4.2 取得データ

3 種類の照明条件を用いて実験データを取得し、各照明条件の下で、カメラ、RFID リーダそれぞれ 130 回の観測を行った。各回の観測においては、物体のリスト L_t はランダムに決定して配置した。

各照明条件のもとで、各観測で同時に配置する物体の数は、1 物体から 3 物体に設定した。また、各照明条件における観測回数は、1 物体を 30 回、2 物体を 50 回、3 物体を 50 回とした。

本研究では、物体の存在しない背景のみの画像と、生の観測データとして得られた画像との差分を取り、一定値以上の差分が見られる領域を物体部分画像として抽出した。また、物体部分画像は 30×30 の pixel を持つ RGB 画像データにリサイズした。また、各 pixel は、その RGB 成分に対して、輝度 0-255 の値をもつものとした。したがって、物体特徴量 Y_{t,j_t} は、 $30 \times 30 \times 3 = 2700$ 次元のベクトル量となる。

表 1: 過去の学習データに対する識別成功率 ([照明 1,2,3] の統合データ), 及び RFID 検知率の真値・推定値

ID	1	2	3	4	5	6	7	8	9	10	平均
識別成功率	0.712	0.938	1.000	1.000	0.966	0.964	0.980	0.983	0.939	0.964	0.945
真の検知率	0.833	0.872	0.750	0.965	0.917	0.862	0.460	0.918	0.922	0.947	0.845
推定検知率	0.395	0.591	0.640	0.632	0.545	0.862	0.170	0.949	0.839	0.771	0.639

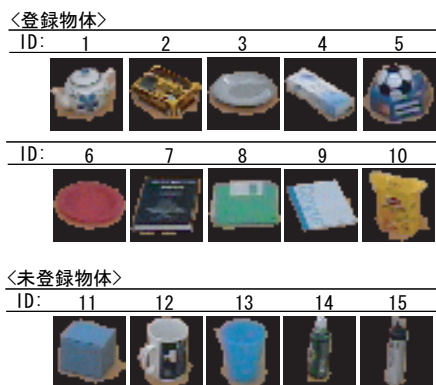


図 4: 観測対象物体及び ID

4.3 学習結果

EM アルゴリズムによって学習した, 画像識別モデルパラメータが, どの程度の識別性能を示すか検証した. その際, 登録物体数 $n = 10$, 未登録物体数 $m = 6$ とした. 表 1 は, 過去に取得した画像センサによる観測データ $Y_{t,j}$ に対し, 識別器の出力と真の ID を照合して, どの程度の識別成功率であったかを, 各 (真の)ID に対して計算した結果である. 表 1 から, 平均して 90%以上の識別成功率を示していることが分かる.

次に, EM アルゴリズムによって学習した画像識別パラメータが, 未登録物体画像に対してどのような識別性能を示すか検証した. 結果を表 2 に示す. 表 2 から, 登録物体に対する識別

表 2: 未登録物体検出結果

未登録物体判定の成功率	456
未登録物体の全データ数	206
未登録物体判定の成功率	0.452

成功率が平均 90%以上と高いのに対し, 未登録物体に対する検出判定成功率が著しく低いことが分かる.

検知率を学習した結果を, 真の検知率と比較して示す. 真の検知率は, 4.2 節で取得したデータから計算した. RFID Tag Reader の観測情報 D_t と, RFID Tag Reader の観測対象領域 \mathcal{D} 内においた真の物体リスト L_t との比を取ることによって計算している. 表 1 から, 各推定検知率は, 真値よりも低くなっていることがわかる. この原因としては, 3.5 で述べた δ の近似計算の際に, 各観測情報 $Y_{t,j}$ の最有力 ID の推定が, 登録物体に偏る傾向があるために, 物体リストの推定時に, 登録物体を多くカウントしていることがあげられる.

5. 考察・結論

RFID 情報という不完全な教師データを用いて識別学習を行っても, 良好な識別性能を示すことができる. 副産物的な問題として, RFID 検出モデルが未知な状態でも, 近似的ではあるが, これを同時に推定しながら識別学習を行うことができる. 本手法は, 登録物体に対する識別性能に対しては, 未登録物体や誤検出画像の影響に対しても一定のロバスト性を示すことができた. しかし, 未登録物体に対する識別性能, いわゆる outlier detection に対しては, 登録物体に対する識別性能ほどの性能を示すことができなかった.

参考文献

- [McLachlan 96] McLachlan, G. J., Krishnan, T.: The EM Algorithm and Extensions (Wiley Series in Probability and Statistics), pp.1-44(1996).
- [Nigam 98] Nigam, K., McCallum, A., Thrun, S. and Mitchell, T.: Learning to Classify Text from Labeled and Unlabeled Documents, in *Proceedings of the Fifteenth National Conference on Artificial Intelligence (1998)*, pp.792-799(1998).
- [Chapelle 06] Chapelle, O., Scholkopf, B. and Zien, A.: Semi-Supervised Learning, pp. 1-8 (2006).