

## 相互依存モデルによるマルチラベル分類

## Multi-label Classification by Interdependence Model

吉村 阜亮      馬場 雪乃      鹿島 久嗣  
Kosuke YOSHIMURA      Yukino BABA      Hisashi KASHIMA

## 京都大学大学院情報学研究科知能情報学専攻

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

Multi-label classification is a supervised classification problem where one or multiple labels are assigned to each instance. Applications of multi-label classification range from recommender systems and classification to drug activation prediction. In this paper, we propose a novel multi-label classification model based on the idea of the interdependence model along with three prediction methods. Our experimental results show the proposed method makes effective predictions especially when some of the labels for each instance are available.

## 1. はじめに

機械学習における研究対象の一つにマルチラベル分類がある。マルチラベル分類とは、一つの対象に対して複数のラベルが割り当てられる分類問題である。この他の分類問題として、対象物を二つの集合に分ける二値分類や、二つ以上の集合に分けるマルチクラス分類があり、これらはマルチラベル分類と区別してシングルラベル分類と呼ばれる。マルチラベル分類は、シングルラベル分類の一般化であることが知られている。マルチラベル分類には様々な応用先があり、具体的には、感情分類、テキスト分類、薬物活性予測、そして、レコメンドなどが挙げられる。

マルチラベル分類がシングルラベル分類と大きく異なる点は、予測の際にラベル間の共起関係を推定に生かすことで精度改善を見込める点である。そのため、ラベル間の関係を予測にうまく利用することができるモデルが必要となる。そこで、本研究では、多変量回帰モデルとして岩瀬らが提案した相互依存モデル [岩瀬 14] をマルチラベル分類に拡張する。相互依存モデルでは、各ラベルの予測に他のラベルを入力として用いるような予測モデルをラベルの種類数分用意する。そのため、予測時に非線形連立方程式を解く必要がある。本研究では、マルチラベル分類に適した予測法として、3種類の予測方法を提案する。

相互依存モデルと新たに提案したラベル予測手法の組み合わせに対して、既存手法との精度比較実験を行い提案手法が精度改善を達成することを示す。加えて、予測対象となるラベルの一部が事前にわかっている状況を想定した精度比較実験も行い、提案手法を用いることでラベルの一部が事前にわかっている場合に、より高い予測精度を達成することも示す。

## 2. 問題設定

訓練データとして、 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  が与えられているとする。 $x_i$  はインスタンス  $i$  の特徴ベクトルを表す。 $y_i$  はインスタンス  $i$  に対する各候補ラベルの有無を表すベクトルであり、 $y_i$  の各要素はインスタンス  $i$  が対応するラベルを持つ場合には 1 を、持たない場合には 0 をとる。つまり、候補ラベル集合を  $\mathcal{Y}$  とすると、各  $y_i$  は、 $y_i \in \{0, 1\}^{|\mathcal{Y}|}$  を満たすことになる。本研究では、訓練データを用いて学習したモデルを用い

て、新たな入力  $x$  に対応するラベル  $y = (y^{(1)}, y^{(2)}, \dots, y^{(|\mathcal{Y}|)})$  を予測することが目標となる。加えて、一部のラベルについてのみ、ラベルの有無がわかっている状況を仮定する。この状況を限定的にシミュレートするため、事前に各インスタンスごとに任意のラベル  $y^{(i)}$  の値が一つだけ与えられている時に、残りのラベルを予測することも目標とする。

## 3. 関連研究

マルチラベル分類におけるラベルの予測手法は大きく次の二つのアプローチに分けられる。一つ目がマルチラベル分類をそのまま取り扱えるように分類器を拡張する方法、二つ目がマルチラベル分類をシングルラベル分類に変形させてシングルラベル分類器を用いて予測する方法である。前者の例としては、決定木のアルゴリズムである C4.5 アルゴリズムを、マルチラベル分類に適用できるように拡張した手法 [Quinlan 93] などがある。一方、後者のアプローチに関しては更に Binary Relevance 法 (BR 法) [Tsoumakas 07a] と Label Powersets 法 (LP 法) の二つの手法に分けられる。BR 法は各ラベルごとに別々に分類器を学習し、ラベルを予測する手法である。つまり、 $f$ : 入力ドメイン  $x \rightarrow \{0, 1\}$  となる分類器  $f$  をラベル数 ( $|\mathcal{Y}|$  個) 分学習し、その各々でラベル予測を行う。BR 法は計算量が小さいという利点がある一方で、ラベル間の共起関係を利用した予測ができないという欠点をもつ。LP 法は候補ラベル集合の冪集合、つまり、考えられる全ての組み合わせの中から一つの組み合わせを選び出すように分類器を学習し、ラベルを予測する方法である。つまり、 $f$ : 入力ドメイン  $x \rightarrow \{0, 1\}^{|\mathcal{Y}|}$  となる分類器  $f$  を一つ学習し、ラベル予測を行う。LP 法はラベル間の共起関係を利用した予測ができるという利点がある一方で、計算量がラベル数に対して指数関数的に増大してしまうという欠点をもつ。以上から、BR 法と LP 法の間には、“計算量” と “ラベル間の共起関係を利用することによる精度向上” の二つの間のトレードオフ関係があることがわかる。

このトレードオフ関係を改善するための手法として RANdom  $k$  labELsets (RA $k$ EL) [Tsoumakas 07b] がある。RA $k$ EL はある種のアンサンブル手法であり、BR 法と LP 法の一般化であることが知られている。RA $k$ EL は、以下の四ステップを経てラベル予測を行う。(1) 候補ラベル集合からサイズが  $k$  の部分集合 (小タスク) を作る。(2) その中から  $M$  個の小タスクをランダムに選択する。(3) 選択した小タスクごとに LP 法を適

用する。(4)LP法で得られる各小タスクごとの予測結果を全て統合し、ラベルを予測をする。なお、RAkELでは、部分集合のサイズ  $k$  により考慮する共起ラベルの個数を設定し、小タスクのサンプリング数  $M$  により計算量を調整する。

また、先述したBR法の計算量が小さいという利点を残したまま、ラベル間の共起関係を利用する手法として、Meta Stacking法 (MS法)[Godbole 04a] や Classifier Chains法 (CC法)[Read 09] などがある。MS法は、まず、特徴ベクトルのみを用いてBR法で全てのラベルを予測し、次に特徴ベクトルに加えて予測により得られたラベル情報を用いて、もう一度各ラベルをBR法で予測するという二段階の手法である。これにより、BR法に比べて計算量は約2倍になってしまうが、ラベル間の共起関係を利用したラベル予測が可能となる。CC法は、各ラベルを順にBR法で予測していく際に、それまでの予測で得られたラベル情報も用いて予測する方法である。これにより、計算量はBR法とほとんど同じままで、ラベル間の共起関係を利用したラベル予測が可能となる。

## 4. マルチラベル分類のための相互依存モデル

本研究では、岩瀬らにより提案された多変量回帰モデルとしての相互依存モデル [岩瀬 14] をマルチラベル分類に拡張する。以下では、マルチラベル分類の文脈における相互依存モデルについて説明する。

### 4.1 相互依存モデルの学習

候補ラベル数が  $|\mathcal{Y}|$  個であるマルチラベル分類を考える。各ラベルに番号を割り当て、 $j$  番目のラベルの値を  $y^{(j)}$  とする。この時、 $y^{(j)}$  に対応する予測モデル  $f_j$  を以下のように定義する。

$$y^{(j)} = f_j(x, \mathbf{y}^{(-j)}; \mathbf{w}_j) \quad (1)$$

ただし、 $\mathbf{y}^{(-j)}$  は、 $\mathbf{y}$  の  $j$  次元目を除いたベクトルであり、 $\mathbf{y}^{(-j)} = (y^{(1)}, \dots, y^{(j-1)}, y^{(j+1)}, \dots, y^{(|\mathcal{Y}|)})^T$  と書ける。また、 $\mathbf{w}_j$  は  $j$  番目のラベルの値を推定するためのパラメタベクトルである。以上から、 $f_j$  は  $j$  番目のラベルの値を、それ以外のラベルの値と特徴ベクトルから予測するモデルということになる。

特徴ベクトル  $x$  が与えられた時の  $\mathbf{y}$  の予測値は、式 (1) を連立した以下の方程式

$$\begin{aligned} y^{(1)} &= f_1(x, \mathbf{y}^{(-1)}; \mathbf{w}_1) \\ y^{(2)} &= f_2(x, \mathbf{y}^{(-2)}; \mathbf{w}_2) \\ &\vdots \\ y^{(|\mathcal{Y}|)} &= f_{|\mathcal{Y}|}(x, \mathbf{y}^{(-|\mathcal{Y}|)}; \mathbf{w}_{|\mathcal{Y}|}) \end{aligned}$$

の解として得られる。

以降、簡単のため、上記の連立方程式をまとめて、

$$\mathbf{y} = \mathbf{f}(x, \mathbf{y}; \mathbf{W}) \quad (2)$$

と表すことにする。ただし、 $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{|\mathcal{Y}|})$  は、各ラベルを予測するモデルのパラメタベクトルをまとめた行列である。

モデルの予測値は連立方程式 (2) の解として得られるため、学習時にこのモデルを訓練データに適合させることを考えると、以下の  $\mathbf{W}$  についての最適化問題

$$\text{minimize} \quad \sum_i L(\mathbf{y}_i, \hat{\mathbf{y}}_i) + \gamma \|\mathbf{W}\|_{\text{F}}^2 \quad (3)$$

$$\text{subject to} \quad \hat{\mathbf{y}}_i = \mathbf{f}(x, \hat{\mathbf{y}}_i; \mathbf{W}) \quad (4)$$

を解くことになる。ただし、 $\hat{\mathbf{y}}_i$  はモデルによる予測であり、この値が制約 (4) を充しながら、目的関数 (3) の第一項目で教師データの値  $\mathbf{y}_i$  を再現するようなモデルパラメタ  $\mathbf{W}$  を求める。なお、 $L(\cdot, \cdot)$  は損失関数を表し、 $\gamma > 0$  は正則化パラメタを表す。また、目的関数内の  $\|\cdot\|_{\text{F}}$  はフロベニウスノルムを表す。

制約 (4) は複雑なため、この制約付き最適化問題を直接解くことは煩雑なものになる。そのため、制約 (4) の右辺の  $\hat{\mathbf{y}}_i$  を教師データの値  $\mathbf{y}_i$  で置き換えることで、この制約を取り除く。これにより、次の制約なしの最適化問題

$$\text{minimize} \quad \sum_i L(\mathbf{y}_i, \mathbf{f}(x, \mathbf{y}_i; \mathbf{W})) + \gamma \|\mathbf{W}\|_{\text{F}}^2 \quad (5)$$

を得る。 $w_1, w_2, \dots, w_{|\mathcal{Y}|}$  を相互に結びつける制約が取り除かれたことで、各モデルごとに独立に最適化できる。結果として、各ラベルごとに別々の  $w_j$  についての最適化問題

$$\text{minimize} \quad \sum_i L(y_i^{(j)}, \mathbf{f}(x, \mathbf{y}_i^{(-j)}; \mathbf{w}_j)) + \gamma \|\mathbf{w}_j\|_2^2 \quad (6)$$

を解けば良いことになる。

### 4.2 相互依存モデルの予測

岩瀬らは多変量回帰モデルとしての相互依存モデルを用いて各変数を予測するために、MATLAB の `fsolve` 関数を用いている。しかし、対象とする問題を多変量回帰からマルチラベル分類に変更するに伴い、我々はマルチラベル分類に適した新たな予測手法を三つ提案する。

#### 4.2.1 不動点反復法による連立方程式の求解

不動点反復法 (Fixed Point Iteration Method) は、非線形連立方程式の解法として知られており、以下の更新規則

$$\mathbf{y}_{(n+1)} = \mathbf{f}(x, \mathbf{y}_{(n)}; \mathbf{W}) \quad (7)$$

を用いて繰り返し値を更新することで非線形連立方程式を解く手法である。先述した通り、予測値  $\hat{\mathbf{y}}$  は非線形連立方程式 (2) の解として得られるため、この不動点反復法を利用することができる。最終的には、一定回数  $\mathbf{y}_{(n)}$  の更新を繰り返すか、収束した時点で閾値を用いて予測ラベルを決定する。ただし、十分に大きな  $n$  を取ったとしても、 $\mathbf{y}_{(n)}$  が収束する保証はない。以降、簡単のためこの手法を FPI と表すことにする。

#### 4.2.2 Gibbs サンプリングを用いたラベル分布の近似

相互依存モデルでは、各ラベルの予測モデルが確率モデルの場合には、各ラベルの条件付確率が与えられる形になるため、これをそのまま Gibbs サンプリングに用いることでラベル集合上の事後分布を近似できる。そのため、相互依存モデルを用いてラベル予測をする際に Gibbs サンプリングを用いることができる。一つのラベルをサンプリングする回数を  $N$  とすると、サンプリングを始めてから  $n = N * 0.01$  までのサンプルは burn-in 期間とみなしてカウントしないこととしている。

#### 4.2.3 全探索

全探索 (Exhaustive Search) を用いてクロスエントロピーの最適化を行う。以下の  $\mathbf{y}$  についての最適化問題

$$\text{minimize} \quad - \sum_j \{y^{(j)} \ln f_j + (1 - y^{(j)}) \ln (1 - f_j)\} \quad (8)$$

$$\text{subject to} \quad y^{(j)} \in \{0, 1\}, \quad j = 1, \dots, |\mathcal{Y}| \quad (9)$$

を全探索により解くことでラベル予測を行う。ただし、 $f_j = f_j(x, \mathbf{y}^{(-j)}; \mathbf{w}_j)$  と省略した。以降、簡単のためこの手法を ES と表すことにする。

## 5. 実験

この章では、各種実験設定と実験の結果について述べる。まず、比較手法と評価指標に関して説明して、次にデータセットの詳細、事前情報を与えた場合の実験の実施方法について述べる。最後に、実験の結果を示す。

### 5.1 比較手法

比較手法として Binary Relevance 法 (BR 法)、Label Powersets 法 (LP 法)、Meta Stacking 法 (MS 法)、Classifier Chains 法 (CC 法) の 4 つの手法を用いる。なお、提案手法も含め内部のモデル (分類器) には全てロジスティック回帰を用いる。閾値はいずれも  $\theta = 0.5$  とし、ロジスティック回帰のパラメタ  $C$  は  $\{0.01, 0.1, 1, 10, 100\}$  の中から最適なものを用いる。各手法の精度は、10 分割の交差検定を用いて求める。

### 5.2 評価指標

評価指標として、以下で定義される Accuracy と macro averaged F-measure を用いる [Read 09]。Accuracy は例ベースの評価指標であるためラベル間の相関も考慮した評価ができる評価指標であり、部分的なラベルの正しさを測ることができる。Accuracy は次式で定義される。

$$\text{Accuracy} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|S_i \cap Y_i|}{|S_i \cup Y_i|} \quad (10)$$

ただし、 $D$  はデータ集合を表し、 $|D|$  はインスタンス数を表す。また、 $S_i$  は  $i$  番目のインスタンスの真のラベル集合を表し、 $Y_i$  は  $i$  番目のインスタンスの予測ラベル集合を表す。一方で、macro averaged F-measure はラベルベースの評価指標であり、よりサンプルが少ないクラスの影響を強く受ける手法であるため、出現頻度の低いラベルを正しく予測できているかをうまく測ることができる。macro averaged F-measure は次式で定義される。

$$F1_{\text{macro}} = \frac{1}{|Y|} \sum_{j=1}^{|Y|} \frac{2p_j r_j}{p_j + r_j} \quad (11)$$

ただし、 $p_j$ 、 $r_j$  はそれぞれ  $j$  番目のラベルについての precision と recall を表す。この時、分母が 0 となる場合には値を 0 としている。

### 5.3 データセット

Scene データセット [Boutell 04]、Emotions データセット [Trohidis 08]、Yeast データセット [Elisseeff 01] の 3 つのデータセットを用いて実験を行った。それぞれ、Scene データセットは、インスタンス数 2407 個、特徴量 294 個、候補ラベル 6 個、Emotions データセットはインスタンス数 593 個、特徴量 72 個、候補ラベル 6 個、Yeast データセットはインスタンス数 2417 個、特徴量 103 個、候補ラベル 14 個である。

### 5.4 事前知識を与えた状況での実験

各インスタンスの特徴量のみを与えてラベル集合を推定する実験に加えて、一部のラベルの有無がわかっている状況を仮定した実験を行う。この実験では、各インスタンスごとにランダムに一つラベルを選び、その値 (0/1) を既知とする。一部の既知となっているラベル情報も利用して他のラベルを推定した時の各手法の精度を比較する。

この実験では、既知のラベル情報を利用するために既存手法を含めて、ある一定の拡張を施す必要がある。以下では、各手法の拡張について述べる。まず、既存の 4 手法の拡張は以下

に示す通りである。BR 法の拡張として、一度 BR 法でラベルを予測し、予測ラベルと既知のラベル情報との食い違いがある場合には既知の情報で予測ラベルを上書きする。LP 法の拡張として、既知のラベルの値を固定することでラベル集合の候補を絞り、その中から最も確率の高いラベル集合を予測ラベルとする。しかし、この LP 法の拡張は本質的な問題として、訓練データには存在しないラベルの組み合わせがテストデータの ground truth の中に存在する時に複数のラベルの既知情報で候補となる集合を絞ると、予測ラベル集合として出力できる選択肢がなくなってしまう可能性がある。CC 法の拡張として、各ラベルの有無を予測することにラベルの既知情報がないかを確認し、もしある場合には予測ラベルを既知の情報で上書きする。MS 法の拡張として、一段階目のラベル予測の際に、既知情報があるラベルに関しては全て既知の情報で上書きして二段階目の予測を行う。

次に、提案手法の拡張は以下に示す通りである。相互依存モデル+FPI の拡張として、各回の更新後に既知情報と食い違いがある予測ラベルを既知情報で上書きする。相互依存モデル+Gibbs サンプリングの拡張として、各回のサンプリングで得られるラベルの有無 (0/1) をラベルの既知情報で毎回上書きして、次のラベルの付与確率も上書きした予測ラベルを用いて計算する。相互依存モデル+ES の拡張として、既知のラベルの値を固定することで探索範囲を狭める。

### 5.5 結果

一般的なマルチラベル分類の文脈での精度比較実験の結果と一部のラベルが既知の状況における精度比較実験の結果を示す。

#### 5.5.1 精度比較実験の結果

表 1 に実験結果を示す。この表において IDD とは相互依存モデル (InterDepenDence Model) を表す。表 1 において値が太字となっている手法は、その実験設定において最も精度が高いことを表す。また、値に下線を引いている手法は、提案手法の中で最も精度が良い手法を表す。

Emotions データセットで評価指標として Accuracy を用いた場合には、提案手法の一つである相互依存モデル+ES が最も良い精度を示している。その他のデータセットや評価指標の組み合わせでは、一部をのぞいて LP 法が最も良い精度を示している。しかし、Scene データセットと Emotions データセットにおいては、提案手法である相互依存モデル+ES は LP 法に匹敵する精度はあることがわかる。提案手法だけに注目すると、相互依存モデル+ES がほとんどの場合で最も良い精度を示している。

#### 5.5.2 事前知識を与えた状況での精度比較実験の結果

表 2 に、一部のラベルの有無が事前情報として与えられている場合の実験の結果を示す。Emotions データセットで評価指標として F-measure を用いた場合には、提案手法の一つである相互依存モデル+ES が最も良い精度を示している。その他のデータセットや評価指標の組み合わせでは、LP 法が最も良い精度を示している。しかし、Emotions データセットにおいては、評価指標として Accuracy を用いた場合も提案手法である相互依存モデル+ES は LP 法に匹敵している。提案手法だけに注目すると、相互依存モデル+ES がほとんどの場合で最も良い精度を示しているが、Yeast データセットにおいては相互依存モデル+Gibbs サンプリングが有効に働いていることも見て取れる。

表 1: 実験結果

Metric	Dataset	BR	LP	MS	CC	IDD + FPI	IDD + Gibbs	IDD + ES
Accuracy	Scene	.593(±.029)	<b>.734</b> (±.020)	.645(±.024)	.693(±.022)	.604(±.028)	.640(±.027)	<u>.722</u> (±.020)
	Emotions	.524(±.042)	.548(±.049)	.543(±.036)	.543(±.033)	.548(±.033)	.506(±.036)	<b>.550</b> (±.042)
	Yeast	.504(±.018)	<b>.530</b> (±.015)	.505(±.017)	.493(±.015)	.382(±.008)	.495(±.019)	<u>.500</u> (±.022)
F-measure	Scene	.692(±.014)	<b>.746</b> (±.021)	.712(±.012)	.716(±.016)	.697(±.020)	.708(±.022)	<u>.742</u> (±.018)
	Emotions	.641(±.044)	.639(±.051)	<b>.646</b> (±.041)	.629(±.042)	<u>.642</u> (±.036)	.618(±.044)	.640(±.042)
	Yeast	.383(±.020)	<b>.432</b> (±.025)	.394(±.016)	.417(±.014)	.231(±.004)	<u>.406</u> (±.017)	.346(±.018)

表 2: 一つのインスタンスあたり一個のラベルの有無が既知の状況における実験の結果

Metric	Dataset	BR	LP	MS	CC	IDD + FPI	IDD + Gibbs	IDD + ES
Accuracy	Scene	.656(±.036)	<b>.795</b> (±.022)	.679(±.027)	.751(±.022)	.686(±.029)	.737(±.016)	<u>.769</u> (±.022)
	Emotions	.583(±.054)	<b>.631</b> (±.035)	.563(±.055)	.607(±.054)	<u>.630</u> (±.053)	.602(±.052)	.629(±.052)
	Yeast	.529(±.018)	<b>.574</b> (±.021)	.515(±.019)	.528(±.019)	.436(±.013)	.550(±.020)	<u>.550</u> (±.024)
F-measure	Scene	.748(±.025)	<b>.811</b> (±.024)	.743(±.025)	.774(±.026)	.772(±.019)	.787(±.016)	<u>.789</u> (±.018)
	Emotions	.675(±.050)	.699(±.046)	.659(±.050)	.684(±.048)	.711(±.046)	.706(±.048)	<b>.714</b> (±.054)
	Yeast	.435(±.023)	<b>.508</b> (±.023)	.393(±.024)	.475(±.027)	.356(±.021)	<u>.477</u> (±.021)	.453(±.015)

## 6. おわりに

マルチラベル分類のための新たな予測モデルとして、多変量回帰の文脈で用いられていた相互依存モデルをマルチラベル分類に拡張した手法を提案した。対象とする問題設定を多変量回帰からマルチラベル分類へ変更するに伴い、相互依存モデルを用いた予測手法として FPI、Gibbs サンプリングに基づく予測手法、ES の 3 つも提案した。

提案手法と既存手法について、二つの実験を行い精度比較を行った。一つ目の実験が、一般的なマルチラベル分類の文脈での実験であり、Emotions データセットにおいて提案手法のうちの相互依存モデル+ES が最も高い Accuracy 値を示した。また、Scene データセットと Emotions データセットでは、LP 法に匹敵する精度を示した。二つ目の実験が、一部のラベルが既知である状況における実験であり、Emotions データセットにおいて提案手法のうちの相互依存モデル+ES が最も高い F-measure 値を示した。また、Emotions データセットでは、相互依存モデル+ES は Accuracy 値でも LP 法に匹敵する精度を示した。

一部のラベルが既知である状況に対応するための今回の LP 法の拡張には、5.4 節で述べた通り本質的な問題を含んでいる。このため、一部のラベルが既知である状況においては、提案手法を用いる意味は十分にあると考えられる。今後の課題としては、各インスタンスごとに既知となっているラベルの数を増やした場合に提案手法の優位性があるか否かを検証する必要がある。

## 参考文献

- [Boutell 04] Boutell, M. R., Luo, J., Shen, X., and Brown, C. M.: Learning multi-label scene classification (2004)
- [Elisseeff 01] Elisseeff, A. and Weston, J.: A Kernel Method for Multi-Labelled Classification, in *In Advances in Neural Information Processing Systems 14*, pp. 681–687, MIT Press (2001)
- [Godbole 04a] Godbole, S. and Sarawagi, S.: Discriminative Methods for Multi-labeled Classification, *Advances*

*in Knowledge Discovery and Data Mining*, pp. 22–30 (2004)

- [Godbole 04b] Godbole, S. and Sarawagi, S.: Discriminative Methods for Multi-labeled Classification, *Advances in Knowledge Discovery and Data Mining*, pp. 22–30 (2004)

- [Quinlan 93] Quinlan, J. R.: *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)

- [Read 09] Read, J., Pfahringer, B., Holmes, G., and Frank, E.: Classifier Chains for Multi-label Classification, in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, ECML PKDD '09, pp. 254–269, Berlin, Heidelberg (2009), Springer-Verlag

- [Schapire 99] Schapire, R. E. and Singer, Y.: Improved Boosting Algorithms Using Confidence-rated Predictions, *Mach. Learn.*, Vol. 37, No. 3, pp. 297–336 (1999)

- [Trohidis 08] Trohidis, K., Tsoumakas, G., Kalliris, G., and Vlahavas, I. P.: Multi-Label Classification of Music into Emotions., in Bello, J. P., Chew, E., and Turnbull, D. eds., *ISMIR*, pp. 325–330 (2008)

- [Tsoumakas 07a] Tsoumakas, G. and Katakis, I.: Multi-label classification: An overview, *Int J Data Warehousing and Mining*, Vol. 2007, pp. 1–13 (2007)

- [Tsoumakas 07b] Tsoumakas, G. and Vlahavas, I.: Random k-Labelsets: An Ensemble Method for Multilabel Classification, in *Proceedings of the 18th European Conference on Machine Learning*, ECML '07, pp. 406–417, Berlin, Heidelberg (2007), Springer-Verlag

- [岩瀬 14] 岩瀬 智亮, 世古 敦人, 鹿島 久嗣: マルチタスク学習を用いた複数物性値の同時予測 (統計数理・機械学習・データマイニング・一般), 電子情報通信学会技術研究報告. IBISML, 情報論的学習理論と機械学習, Vol. 113, No. 476, pp. 9–13 (2014)