

# マルチタスク学習と誘因両立性

## Analysis of inseparability of data generation from multi-task learning

大岩 秀和\*1  
Hidekazu Oiwa

中川 裕志\*2  
Hiroshi Nakagawa

東京大学 情報理工学系研究科\*1

The University of Tokyo, Graduate School of Information Science and Technology

東京大学 情報基盤センター\*2

The University of Tokyo, Information Technology Center

Multitask learning algorithms have been well used for achieving personalization due to their highly sophisticated model variations and theoretical foundations. However, users' incentives have crucial effects for the generation of data and the performance of algorithms. This paper investigates the relationship between these algorithms and users' incentives.

### 1. はじめに

複数の予測タスクが存在する際に、全タスクのための予測モデルを同時に学習する事をマルチタスク学習と呼ぶ。特に学習に十分なデータ量が存在しないタスクが多い場合には、個々のタスクで独立に予測モデルを構築するよりもタスク間で情報共有することで高い予測精度を実現できる事が示されてきた [Evgeniou 04, Argyriou 08, Kumar 12, Crammer 12]. 理論面でも、多くのアルゴリズムで未知データへの予測精度を示す汎化誤差の上限保証が与えられている [Ben-David 10, Blanchard 11, Crammer 12, Kakade 12]. このように、マルチタスク学習はその有用性から理論面および応用事例が豊富であり非常に様々な側面から研究が進められてきた。

ここで、メールフィルタリング・商品推薦・広告最適化などの多くのユーザーを抱えるサービスに機械学習アルゴリズムを適用する場合を考える。これらは多くの場合ユーザー毎に選好が異なるため、一ユーザーを一タスクとみなしたマルチタスク学習によるユーザー選好を考慮した予測モデル構築手法が近年盛んに研究されている [Aberdeen 10]. この定式化により、購入履歴やユーザープロフィール等の情報を学習データとして用いた予測モデル構築およびユーザー適応が可能となる。

本稿ではこれらのユーザーフィードバック情報を学習データとして用いたマルチタスク学習を分析する。はじめに、この問題設定では多くの既存アルゴリズムで汎化誤差等の理論保証が崩れる事を示す。既存手法の理論保証は、データがあるデータ分布に従い独立同分布で生成されるという重要な仮定が置かれている。これはマルチタスク学習の重要な仮定として考えられてきた\*1。一方で、ユーザーフィードバック環境下で各ユーザーは利己的にデータを供給すると考えられる。したがって、ユーザーが自身の効用にあった予測モデルを構築できるならば、ユーザーの真の効用関数に従わない偽のデータ生成を促しうる。既存アルゴリズムの多くはこのようなユーザーのインセンティブ構造を考慮していないため、偽データの生成を防ぐ機構を持たない。偽データ生成は一部のユーザーにとって利益と

なってもユーザー全体の効用を減少させる可能性があるため偽データ生成のインセンティブをユーザーに与えない機構を持つアルゴリズムの開発が求められる。このような性質を誘因両立性と呼び、誘因両立性を持つアルゴリズムの設計がこの問題設定では重要となる。

次に、マルチタスク学習アルゴリズムを誘因両立性の観点から複数のモデルに分類し、各モデルの理論的性質を検証する。誘因両立性を欠いたモデルやアルゴリズムは前述のとおり偽データの影響で独立同分布の仮定が崩れるため、既存の各種理論保証が適用不可能となる。ユーザーフィードバック環境下ではアルゴリズムがデータ生成に影響を与えうるため、データ生成に悪影響をおよぼすアルゴリズムは適用困難になる。各種モデルにおける誘因両立性と統計的学習理論の関係をまとめたのが本稿の貢献であり、加えて誘因両立性を満たすマルチタスク学習アルゴリズムの例として CRD-SHAMO アルゴリズムを開発した。結果の概要を表 1 に示している。

### 1.1 関連研究

#### 1.1.1 機械学習と誘因両立性

機械学習アルゴリズムと誘因両立性の関連性を分析した代表的な既存研究として [Meir 12] と [Dekel 10] が挙げられる。[Meir 12] は効用関数が異なる複数のエージェントからデータが生成される環境下で単一の予測器を学習する分類問題に対し、誘因両立性とアルゴリズムの関係を分析している。本研究は複数の予測器を持つ予測モデルへの拡張とみなせる。[Dekel 10] では、回帰学習アルゴリズムと誘因両立性との関係性を理論面から解析している。誘因両立性を満たすアルゴリズムの期待損失の上限値は損失関数によって異なる事を示し、数種の特別な損失関数について理論的最適なアルゴリズムを導出している。

各ユーザー効用の金銭による定量化とユーザー・サービス間の金銭のやり取りが可能であれば、VCG メカニズム\*2に代表される金銭ベースのメカニズムが利用可能になる。顕示原理 [Nisan 07] により、多くの問題設定でユーザーのインセンティブを金銭で制御し誘因両立性を満たすことが可能な事が知られている [Dekel 10, Meir 12]. しかし、金銭ベースの手法は金銭譲受を満たすシステムが実現不可能あるいは計算量的な問題から多くの場合適用が困難である。

連絡先: 大岩秀和, 東京大学情報理工学系研究科, 東京都文京区本郷 7-3-1 総合図書館 4F, 03-5841-2729, 03-5841-2738, oiwa@r.dl.itc.u-tokyo.ac.jp

\*1 共変量シフトなどの例外も存在するが、本研究では扱わない。

\*2 VCG メカニズムの概要は [Nisan 07] に詳しい。

## 2. 問題設定

本章では、インセンティブを導入したマルチタスク学習の問題設定を定義する。マルチタスク学習では、各タスクのデータから全タスクの予測モデルを同時に学習するのが目的となる。データの入力空間は  $r$  次元の空間  $\mathcal{X} \subset \mathbb{R}^r$ 、出力空間は一次元空間  $\mathcal{Y} \subset \mathbb{R}$  で定義され、これらは全タスクで共通と仮定する。本稿では二値分類問題を対象とし、出力空間は  $\mathcal{Y} = \{-1, 1\}$  で定義される。

入力データは入力空間上の確率分布  $D$  に従って生成され、 $T$  個の各タスク  $t \in \mathcal{T}$  に  $n$  個のラベル無しデータ  $U = \{(x_i)\}_{i=1}^n$  が分配される。  $U$  は全タスクで共通と仮定する\*3。生成されたラベル無しデータに各タスクのエージェントがラベル付けを行い、教師データ  $\mathcal{L}_t = \{(x_i, y_{t,i})\}_{i=1}^n$  を生成する。学習アルゴリズム  $A : \{\mathcal{L}_t\}_{t \in \mathcal{T}} \rightarrow \{\hat{h}_t\}_{t \in \mathcal{T}}$  は上の手順で生成された教師データを受け取り、入力データから出力ラベルを返す関数の部分集合である仮説集合  $\mathcal{H}$  から各タスクの予測モデル  $\hat{h}_t(\cdot) : \mathcal{X} \rightarrow \mathcal{Y}$  を出力する。エージェントはそれぞれ固有の識別関数  $h_t(\cdot) : \mathcal{X} \rightarrow \mathcal{Y}$  を持ち、全タスクの識別関数を適切に表現できる予測モデルの生成が理想となる。各タスクの目的関数は以下の予測誤差最小化問題として記述できる。

$$R_t(\hat{h}_t) = \mathbb{E}_{x \sim D} [h_t(x) \neq \hat{h}_t(x)]. \quad (1)$$

全タスクの目的関数をあわせたサービスとしての目的関数は以下の式で記述される。本稿では簡単のため全エージェントの重要度は同一と仮定するが、不均一な場合へ拡張可能でありこれからの議論も基本的に同じである。

$$R(\{\hat{h}_t\}_{t \in \mathcal{T}}) = \frac{1}{T} \sum_{t \in \mathcal{T}} \mathbb{E}_{x \sim D} [h_t(x) \neq \hat{h}_t(x)]. \quad (2)$$

アルゴリズムの目的関数は、入力データに関する予測誤差の期待値として定義される。ここで、アルゴリズムに乱択要素が含まれる場合はデータに関する期待値と同じく乱択部分に関しても期待値をとる。

$$R(A) = \mathbb{E}_{U \sim (D)^n} [R(A(\{\mathcal{L}_t\}_{t \in \mathcal{T}}))]. \quad (3)$$

通常のマルチタスク学習の問題設定では、教師データ  $\{\mathcal{L}_t\}_{t \in \mathcal{T}}$  は各エージェントの識別関数にしたがって生成されると仮定している。一方でエージェントのインセンティブを導入した場合、各エージェントはタスク全体としての最適化 (2) ではなく、自身のタスクに関する予測誤差 (1) が最小となるようにデータを生成する。真の関数に従わないデータの生成によって自身の予測精度が改善するならば、各エージェントは偽のデータを生成する。これらの偽データ生成は教師データの性質を変化させた時に導出される予測モデルがどのように変化するか依存するため、アルゴリズムの設計が誘因両立性に強い影響を与える。偽データ生成を許すアルゴリズムは、タスク全体の予測精度に影響を及ぼす事に加え、独立同分布仮定が崩れるために既存の理論保証が適用出来なくなる。

本稿の目的は、エージェントが利己的に行動する環境下において汎化誤差最小化の面で各種マルチタスク学習アルゴリズムの有益性の分析を行うことである。上記のように、この分析の上で誘因両立性は非常に重要な要因となる。

\*3 入力データが共通である仮定を置かない場合には、十分に効率的な最適性を持つアルゴリズムを導出不可能であることが先行研究により証明されている [Meir 10].

## 3. マルチタスク学習アルゴリズム

マルチタスク学習の手法として非常に様々なモデルが考案されてきた。単純な学習方法として、全タスクをまとめて一タスクと見立て一つの予測モデルを学習する単一予測モデルアプローチと、各タスクに予測モデルを一つ割り当てそれぞれ独立に最適化する単純独立予測モデルアプローチが存在する。これらのアプローチはタスク間の関連性を無視するなど非常にナイーブであるが、本稿の分析の出発点として最初に検討する。有効的なマルチタスク学習アルゴリズムの多くはこれらの中間的手法として考えられ、本稿では二種類のモデルに分類する。

第一の方法は各タスクに対して線形の予測モデルを一つ割り当て、タスク間の情報交換を促進する罰則項を加えた最適化問題を解くアプローチである。罰則項の種類には様々なバリエーションが存在し [Evgeniou 04, Cavallanti 10]、汎化誤差上限を導出する解析結果も得られているため [Kakade 12]、非常に有用なアプローチとして知られている。

第二の方法は少数の予測モデルをタスク間で共有するアプローチである。[Crammer 12] は、少数データしか持たない大量のタスクから効率的に予測モデルを学習する方法として、複数のタスクで同じ予測器を共有するマルチタスク学習を実現する SHAMO アルゴリズムを提案した。このアルゴリズムは予測器の割り当てと学習の同時最適化を実現し、理論面からもモデルの VC 次元を導出している。

## 4. マルチタスク学習と誘因両立性

本稿では 2 章で定義された問題設定について四種のマルチタスク学習のアプローチの分析を行う。ここで、学習アルゴリズムは分布  $D$  そのものは確認出来ず分布から生成されたサンプルの情報のみから学習を行う。そのため生成サンプルが非常に偏る可能性があり、エージェントが一人のみの問題設定であっても誘因両立性を必ず満たすアルゴリズムの非存在性が既存研究により証明されている [Meir 12]。ただし、汎化誤差上限を持つアルゴリズムは非常に高い確率で誘因両立性を満たすアルゴリズムによっても示すことが可能なため、本稿では  $\epsilon$ -truthful と呼ばれる概念を用いて分析を行う。  $\epsilon$ -truthful なエージェントとは、真のラベル付けに対して自身のタスクに関する予測誤差が  $\epsilon$  以上改善しない限り偽データを生成しない行動原理を持つエージェントのことを指す。偽データを能動的に作成するには一定のコストがかかる場合が多く、そのコストを  $\epsilon$  で制御していると考えられる。ページ数の制約により、以下で導入される定理の証明は本稿からは省く。

### 4.1 単一予測モデル

はじめに、全タスクを一つのタスクと見立て単一の予測モデルで学習を行うアプローチの分析を行う。マルチタスク学習問題を単一タスク教師あり学習問題へ簡略化しているため、先行研究の分析結果をそのまま適用出来る [Meir 12]。先行研究で提案された CRD メカニズム (Algorithm 1) は、以下の定理により期待損失上限が導出できることが示されている。

**定理 1** [Meir 12] 全てのエージェントが  $\epsilon$ -truthful であり、仮説空間  $\mathcal{H}$  は有限次元であると仮定する。この時、いかなる  $\epsilon$  についても  $n > 64 \frac{V_C}{\epsilon^2} \log(256 \frac{V_C T}{\epsilon^3})$  個のデータを生成して CRD メカニズムを適用すれば以下の式が成立する。

$$R(A) \leq \left(3 - \frac{2}{T}\right) \inf_{\hat{h} \in \mathcal{H}} R(\hat{h}) + \epsilon. \quad (4)$$

ここで  $V_C$  は仮説空間の VC 次元である。

モデル	誘因両立性	近似レート	必要サンプル数	最適値の定義
単一予測モデル	✓	$3 - \frac{2}{T}$	$64 \frac{V_C}{\epsilon^2} \log(256 \frac{V_C \cdot T}{\epsilon^3})$	$\inf_{\hat{h} \in \mathcal{H}} \sum_{t \in \mathcal{T}} R_t(\hat{h})$
単純独立予測モデル	✓	1	$64 \frac{V_C \cdot T}{\epsilon^2} \log(256 \frac{V_C}{\epsilon^3})$	$\sum_{t \in \mathcal{T}} \inf_{\hat{h} \in \mathcal{H}} R_t(\hat{h})$
罰則付き独立予測モデル	—	—	—	—
共有予測モデル	✓	$3 - \frac{2}{T_k}$	$\sum_{k \in \mathcal{K}} 64 \frac{V_C}{\epsilon^2} \log(256 \frac{V_C \cdot T_k}{\epsilon^3})$	$\sum_{k \in \mathcal{K}} \inf_{\hat{h} \in \mathcal{H}} \frac{1}{T_k} \sum_{t \in \mathcal{T}_k} R_t(\hat{h})$

表 1: 分析結果の概要

**Algorithm 1** CRD メカニズム

データ分布  $D$  より  $n$  個のデータ  $U$  を独立同分布でサンプル  
**for** 各タスク  $t \in \mathcal{T}$  について **do**  
 エージェントが  $U$  のラベル付けを行い,  $L_t$  を生成  
**end for**  
 タスク  $t' \in \mathcal{T}$  を一つランダムに選択  
 $L_{t'}$  に関する経験リスク最小化により仮説  $\hat{h}$  を導出

CRD メカニズムで導出される予測モデルは, 仮説空間内の最適仮説で評価される期待損失に対して  $3 - \frac{2}{T}$  の近似レートで上限が抑えられる.

**4.2 単純独立予測モデル**

次に, 各タスクに一つの予測モデルを割り当てそれぞれ独立に最適化するアプローチを分析する. このアプローチでは CRD メカニズムを各タスク独立に適用する事が可能であり, 結果としてリスク最小化問題をタスクごとに解くアルゴリズムとなる. 分析としても CRD メカニズムの性質を適用でき, 以下の定理が成立する.

**定理 2** エージェント  $t \in \mathcal{T}$  が  $\epsilon$ -truthful であり, 仮説空間  $\mathcal{H}$  は有限次元であると仮定する. この時, いかなる  $\epsilon$  についても  $n > 64 \frac{V_C}{\epsilon^2} \log(256 \frac{V_C}{\epsilon^3})$  個のデータをサンプルして CRD メカニズムを適用すれば以下の式が成立する.

$$R_t(A) \leq \inf_{\hat{h}_t \in \mathcal{H}} R_t(\hat{h}_t) + \epsilon. \quad (5)$$

先の単一予測モデルの分析結果と比較すると, 以下の点で性質が異なる.

- 各エージェント個別の最適仮説に対する理論保証を与えているため, 全エージェント共通の最適仮説よりも小さな期待損失への近似レート上限を与えている.
- 一方で,  $\epsilon$  誤差を達成するために必要なデータ数はタスク数  $T$  に線形オーダーとなる. したがって, タスク数に比べデータ数が十分に多い場合以外には予測モデルの上限保証は有益なものとならない.

**4.3 罰則付き独立予測モデル**

次に罰則付き独立予測モデルを分析する. ここで罰則付き独立予測モデルとは, 以下の形式で最適化問題を記述できるアルゴリズムの事を本稿では指す.

$$W = \operatorname{argmin}_W \left( \sum_{i,t} \ell_{i,t}(w_t) + \Omega(W) \right). \quad (6)$$

ここで  $W$  は重み行列で  $w_t$  はその  $t$  行ベクトル,  $\ell_{i,t}(\cdot)$  は  $t$  番目のタスクの  $i$  番目のデータに関する損失関数,  $\Omega(\cdot)$  は重み行列に課される正則化項である. 重み行列の各行ベクトル  $w_t$  が各タスクの予測モデル  $h_t(\cdot)$  (線形識別器) に対応しており, 重

**Algorithm 2** SHAMO アルゴリズム

各タスク  $t \in \mathcal{T}$  を予測モデル集合  $\mathcal{K}$  のいずれかにランダムに割り当てる  
**while** 所定の反復回数に達しない **and** モデル割り当てに変更がある **do**  
 各予測モデルに割り当てられたタスクのデータを用いて  $\{\hat{h}_k\}_{k \in \mathcal{K}}$  を学習  
 各タスクを, そのデータに関する予測誤差最小の予測モデルへ割り当て直す  
**end while**  
 各予測モデルに割り当てられたタスクのデータを用いて最終的な  $\{\hat{h}_k\}_{k \in \mathcal{K}}$  を学習

み行列の学習が予測モデル  $\{\hat{h}_t\}_{t \in \mathcal{T}}$  の学習に対応している. 正則化項には重みベクトル間のユークリッド距離 [Evgeniou 04] や重み行列の特異値 [Cavallanti 10] などが使用される.

残念ながら, これらの罰則付き独立予測モデルのアルゴリズムは誘因両立性を満たさない問題設定が存在する.

**定理 3** 重みベクトル間のユークリッド距離 [Evgeniou 04] や重み行列の特異値 [Cavallanti 10] を罰則項に用いたアルゴリズムは, 無視できない高い確率で誘因両立性を満たさない実験設定が存在する.

Proof Sketch: この証明では, [Evgeniou 04] のみを対象とする. 入力空間は一次元, エージェントは二人のケースを想定する. 各エージェントは,  $h_1(x) = 1(x > 1), -1(x \leq 1)$  と  $h_2(x) = 1(x \geq -1), -1(x < -1)$  という形の識別関数を持つ. この時, 各エージェントの最適な線形識別器は  $\langle w_1, x \rangle = x - 1$  と  $\langle w_2, x \rangle = x + 1$  になる. 正則化項は各予測モデルを互いに近づけるように働いたため, エージェント 1 は真のデータを生成すると  $x < 1$  のデータも  $y = -1$  とラベル付けされる予測モデルが生成されてしまう. エージェント 1 は  $x = 1$  に近接したデータは  $x > 1$  であっても  $y = -1$  とラベル付けする偽データを生成し, エージェント 2 の予測モデルから距離的に離し, 正則化項を含めた最適化問題の出力として  $\hat{h}_1(x) = x - 1$  となるようデータを改変するインセンティブを持つ. このため, アルゴリズムは誘因両立性を満たさない. [Cavallanti 10] の場合も上と同様のケースを検証することで, 誘因両立性を満たさないことの証明が可能である. ■

この性質より, 過去に示された理論保証 [Kakade 12] もユーザーフィードバック環境下では成立しないため, これらの罰則付き独立予測モデルは適用が困難であるとかんがえられる. その他一般の独立予測モデルが誘因両立性を満たす条件は現在 Open Problem である.

**4.4 共有予測モデル**

最後に共有予測モデルを誘因両立性の面から分析する. 共有予測モデルでは,  $K$  個の予測モデル集合  $\{\hat{h}_k\}_{k \in \mathcal{K}}$  をタスク間で共有するモデルである. 各タスクは予測モデル集合のうち

**Algorithm 3** CRD-SHAMO アルゴリズム

---

データ分布  $D$  より  $n$  個のデータ  $U$  を独立同分布でサンプル  
**for** 各タスク  $t \in \mathcal{T}$  について **do**  
 エージェントが  $U$  のラベル付けを行い,  $L_t$  を生成  
**end for**  
 各タスク  $t \in \mathcal{T}$  を予測モデル集合  $K$  のいずれかにランダムに割り当てる  
**while** 所定の反復回数に達しない **do**  
 各予測モデル  $k \in K$  から一つタスク  $t \in \mathcal{T}$  をランダムに選択. 選ばれたタスクのデータから  $\hat{h}_k$  を学習  
 各タスクを, そのデータに関する予測誤差最小の予測モデルへ割り当て直す  
**end while**  
 各予測モデル  $k \in K$  から一つタスク  $t \in \mathcal{T}$  をランダムに選択. 選ばれたタスクのデータから最終的な  $\hat{h}_k$  を学習

---

一つの子モデルに割り当てられる。予測モデル数  $K$  がタスク数  $T$  より少ない時予測モデルが複数のタスクで共有されるため、共有予測モデルと呼ばれる。共有予測モデルの代表例として SHAMO アルゴリズム [Crammer 12] (Algorithm 2) が提案されている。ここで、本稿で SHAMO アルゴリズムとして扱うアルゴリズムは、既存手法から各タスクの訓練・テストデータ分割処理を除いたものになっている。既存のモデルでは各反復で学習手法として経験リスク最小化を用いている。

経験リスク最小化ベースの SHAMO アルゴリズムは、今回の問題設定に適さない事が次の定理によって示される。

**定理 4** 経験リスク最小化ベースの SHAMO アルゴリズムも誘因両立性を満たさない。

この定理は複数エージェントを含む経験リスク最小化が誘因両立性を満たさない [Meir 12] ことから証明可能である。

そこで本稿では、各反復時に CRD メカニズムにもとづいて予測モデルの学習を行う CRD-SHAMO アルゴリズムを提案する。CRD-SHAMO アルゴリズムの概要を Algorithm 3 に示す。CRD アルゴリズムは乱択アルゴリズムのため、CRD-SHAMO アルゴリズムではタスクの割り当て変更が生じない場合も反復を停止しない。

CRD-SHAMO アルゴリズムは誘因両立性を満たし、さらに以下の定理を満たす。

**定理 5** 全てのエージェントが  $\epsilon$ -truthful であり、仮説空間  $\mathcal{H}$  は有限次元であると仮定する。この時、予測モデル  $k$  に最終的に割り当てられたタスク集合を  $\mathcal{T}_k$ 、タスクの数を  $T_k$  とおくと、いかなる  $\epsilon$  についても  $n > 64 \frac{V_C}{\epsilon^2} \log(256 \frac{V_C \cdot T_k}{\epsilon^3})$  個のデータをサンプルして CRD-SHAMO メカニズムを適用すれば以下の式が成立する。

$$\frac{1}{T_k} \sum_{t \in \mathcal{T}_k} R_t(A) \leq \left(3 - \frac{2}{T_k}\right) \inf_{\hat{h} \in \mathcal{H}} \frac{1}{T_k} \sum_{t \in \mathcal{T}_k} R_t(\hat{h}) + \epsilon. \quad (7)$$

他のモデルとの性質の比較は以下になる。

- 最適仮説は予測モデル毎に独立であるため、単一予測モデルにより良い最適値に対する上限を与えている。また、ほぼ同じ最適仮説を持つ性質が似たタスク同士を同じ予測モデルに割り当てる事ができれば、最適値は単純独立予測モデルとほぼ同じとなると考えられる。

- $\epsilon$  誤差を達成するのに必要なサンプル数は予測モデル数に依存し、予測モデル数をタスク数に対して対数オーダー以下に設定すれば対数オーダーで抑えられる。このため、単純独立予測モデルの線形オーダーから改善される。少数データしか持たない大量のタスクを同時に処理するユーザーフィードバック環境下において、予測モデルの数を調整して必要サンプル数を調整できる CRD-SHAMO アルゴリズムは非常に有効な手法であると考えられる。
- 単純独立予測モデルと比較すると、必要サンプル数のみでなく予測モデルのメモリ使用量も大幅に削減可能となる。

## 5. 今後の課題

本稿ではマルチタスク学習と誘因両立性の関係性について分析を行った。今後の課題として、1) 罰則付き独立予測モデルにおける誘因両立性を持つアルゴリズムの有無の検証、2) 誘因両立性を満たすアルゴリズムの予測誤差上限値の解析、3) 誘因両立性を満たす最適アルゴリズムの導出、4) 各アルゴリズムの性能を比較する実験、を進めることで、この分野のより包括的な分析と検証が求められる。

## 参考文献

- [Aberdeen 10] Aberdeen, D., Pacovsky, O., and Slater, A.: The learning behind gmail priority inbox, in *LCCC: NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds* (2010)
- [Argyriou 08] Argyriou, A., Evgeniou, T., and Pontil, M.: Convex Multi-task Feature Learning, *Machine Learning*, Vol. 73, No. 3, pp. 243–272 (2008)
- [Ben-David 10] Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. W.: A Theory of Learning from Different Domains, *Machine Learning*, Vol. 79, No. 1-2, pp. 151–175 (2010)
- [Blanchard 11] Blanchard, G., Lee, G., and Scott, C.: Generalizing from Several Related Classification Tasks to a New Unlabeled Sample, in *NIPS*, pp. 2178–2186 (2011)
- [Cavallanti 10] Cavallanti, G., Cesa-Bianchi, N., and Gentile, C.: Linear Algorithms for Online Multitask Classification, *Journal of Machine Learning Research*, Vol. 11, pp. 2901–2934 (2010)
- [Crammer 12] Crammer, K. and Mansour, Y.: Learning Multiple Tasks using Shared Hypotheses, in *NIPS*, pp. 1484–1492 (2012)
- [Dekel 10] Dekel, O., Fischer, F. A., and Procaccia, A. D.: Incentive compatible regression learning, *Journal of Computer and System Sciences*, Vol. 76, No. 8, pp. 759–777 (2010)
- [Evgeniou 04] Evgeniou, T. and Pontil, M.: Regularized Multi-task Learning, in *KDD*, pp. 109–117 (2004)
- [Kakade 12] Kakade, S. M., Shalev-Shwartz, S., and Tewari, A.: Regularization Techniques for Learning with Matrices, *Journal of Machine Learning Research*, Vol. 13, pp. 1865–1890 (2012)
- [Kumar 12] Kumar, A. and III, H. D.: Learning Task Grouping and Overlap in Multi-task Learning, in *ICML* (2012)
- [Meir 10] Meir, R., Procaccia, A. D., and Rosenschein, J. S.: On the Limits of Dictatorial Classification, in *AAMAS*, pp. 609–616 (2010)
- [Meir 12] Meir, R., Procaccia, A. D., and Rosenschein, J. S.: Algorithms for Strategyproof Classification, *Journal of Artificial Intelligence*, Vol. 186, pp. 123–156 (2012)
- [Nisan 07] Nisan, N.: Introduction to mechanism design (for computer scientists), *Algorithmic game theory*, pp. 209–242 (2007)