

## 無限混合 SVM の関係モデル化

## On Modeling Relational Infinite SVM

小山田 昌史 \*<sup>1</sup> 中台 慎二 \*<sup>1</sup>

Masafumi Oyamada Shinji Nakadai

\*<sup>1</sup>日本電気株式会社 情報・ナレッジ研究所

Knowledge Discovery Research Laboratories, NEC Corporation

We present *Relational Infinite-SVM (R-iSVM)*, a Bayesian nonparametric mixture-of-experts model that leverages Infinite Relational-Model (IRM) as gating functions and SVM as discrimination functions. R-iSVM captures hidden relationships of entities from entity relationship data, constructs clusters of entities, and learns a localized SVM model in each cluster. In doing so, R-iSVM successfully increases the prediction performance of SVM, and furthermore, constructs better clusters compared to IRM thanks to supervisors of SVM.

## 1. 概要

本稿は, iSVM (Infinite-SVM) [7] を IRM (Infinite Relational Model) [5] にもとづき関係モデル化した R-iSVM (Relational Infinite-SVM) を提案する. iSVM は DPM (Dirichlet Process Mixture) により得られたクラスごとにローカリティの高い SVM モデルを学習することで予測性能を高める, 高速な多クラス分類器である. 提案する R-iSVM は iSVM を関係モデルへと拡張し, 顧客と商品の購買関係のような関係データを用いてクラスタリングをおこなうことで, 予測性能を向上させる.

R-iSVM は生成モデルとして扱うことのできない SVM をモデル中に含むため, ベイズ推論でよく用いられる MCMC や変分ベイズ法を直接適用することができない. そこで, 本稿は [7] にならい, IRM の変分ベイズ推論と SVM モデルの学習がどちらも関数の最小化として扱えることを利用し, R-iSVM の学習問題をこれら関数の同時最小化として定義する. そして, その問題の解法として変分ベイズ法と二次計画法の組み合わせによる学習アルゴリズムを示す.

本稿は提案する R-iSVM に対して実データを用いた実験をおこない, R-iSVM が欠損属性値の予測タスクにおいて, 関係データを用いた従来の属性生成と比べて高い予測精度を実現することを確認する.

## 2. 背景

データの属性値の予測は一般的な機械学習タスクである. 例えば, 図 1 の Users テーブルのような顧客データが与えられたとき, 属性 occupation の予測は, 説明変数を age, income 属性, 目的変数を occupation 属性とする多クラス分類問題となる. 実際の業務においては単一のテーブルだけでなく, 商品データや, 顧客と商品を関連付ける関係データ (例: 顧客の商品購買ログ) が利用可能である場合も多く, その場合は図 2 に示すような属性生成 (Feature engineering) によって顧客データに購買ログにもとづく属性を追加し, 説明変数を増やすことで, 予測精度の向上を図ることができる.

図 2 に示すような従来の属性生成では, 購買ログという関係データから得た商品データを顧客ごとに集計して利用するため,

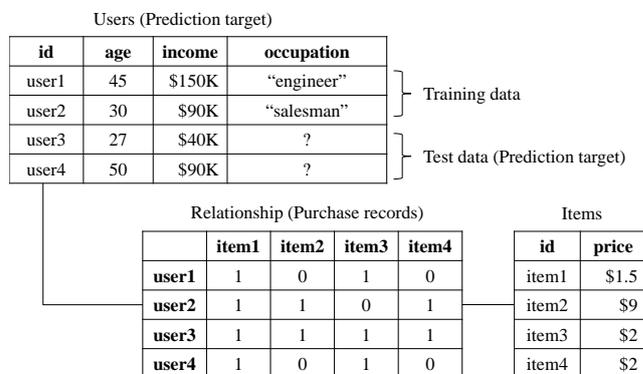


図 1 顧客データ, 商品データ, そして顧客による商品の購買の有無をあらわす関係データ.

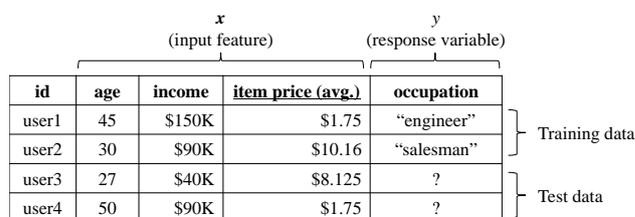


図 2 関係データによる属性生成の例: 各顧客が購入した商品情報を集約し, 顧客の属性に追加する. この例では購入した商品の平均価格を顧客の新たな属性としている.

各顧客がどの商品を購入したかの関係情報が失われ, 情報の粒度が粗くなる. また, 顧客の購入した全ての商品データの情報を説明変数とする属性生成の戦略も考えられるが, 情報の粒度が細かく, よい説明変数になりにくい. これに対し, 提案する R-iSVM (Relational Infinite-SVM) は, 関係データを直接的に用いて顧客ないし商品から背後にあるクラスタを抽出し, 適度な粒度の情報としたうえで, 予測性能の向上に役立てる.

## 3. 準備

はじめに, 提案する R-iSVM モデルのベースとする iSVM と IRM について述べる.

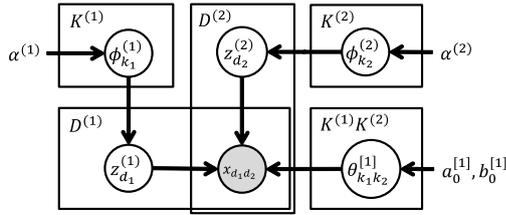


図3 IRMのグラフィカルモデル.

### 3.1 iSVM (Infinite-SVM)

iSVM (Infinite-SVM) [7] は Zhu らによって提案された多クラス分類器である. iSVM は DPM (Dirichlet Process Mixture) と SVM を組み合わせた確率的潜在変数モデルであり,

1. SVM の教師情報によるクラスタリングの質の向上
2. クラスタ毎にローカリティの高い SVM モデルを学習することによる分類性能の向上

の2つを同時に達成する. これまでも分類性能を向上するためにデータをクラスタリングしてからクラスタ毎に分類器を学習する試み [2] はあったが, このアプローチでは2. の分類器の性能向上しか実現できないのに対し, iSVM ではクラスタリングと SVM が相互に作用しあうことで, 1. のクラスタリングの質の向上と, 2. の分類性能の向上を同時に達成している.

iSVM は SVM の DPM (Dirichlet Process Mixture) [1] による混合モデルであり, DPM から得られたクラスタ毎にマージン最大化をおこない, クラスタ毎に異なる SVM のモデル  $\eta_k$  を学習する\*1. 各データ  $x_i$  がクラスタ  $k$  において SVM モデル  $\eta_k$  の学習 (マージン最大化) に与える影響の大きさは, データ  $x_i$  のクラスタ  $k$  への所属確率  $q(z_i = k)$  により定まる. 各クラスタにおいてはデータのローカリティが高まること期待されるため, クラスタ毎に作成した SVM モデルの予測性能はデータ全体に対して作成された SVM モデルと比べ高くなること期待される. また, 全体として線形分離の不可能なデータがクラスタ内では線形分離可能であることもあり, この場合は iSVM により線形 SVM を組み合わせることで, 計算コストの高い非線形カーネルと同等の分離性能を低コストに得ることができる [7].

### 3.2 IRM (Infinite Relational Model)

IRM (Infinite Relational Model) [5, 6] は関係クラスタリングを実現する潜在変数モデルのひとつである. 関係クラスタリングは, 複数のデータ (例: 顧客  $d_1$ , 商品  $d_2$ ) と, それらのデータ間の関係をあらわす関係データ (例:  $x_{d_1d_2} = 1$  ならば, 顧客  $d_1$  は商品  $d_2$  を買ったことがある) が与えられたとき, 両データをそれぞれクラスタリングするタスクである. IRM は図3に示すような生成モデルを持ち, 潜在変数 ( $z_{d_1}^{(1)}, z_{d_2}^{(2)}$ ) の値を推定することでデータのクラスタリングをおこない, またクラスタの組み合わせの上に定義されたモデルのパラメータ  $\theta_{k_1k_2}^{[1]}$  を推定することで, クラスタ間の関係の強さを推定する. なお, IRM はノンパラメトリックベイズモデルであり, 各データの混合数は無限であるが, 本稿は図3に示すように SBP (Stick-Breaking Process) によるモデル化をしており, 混合数は SBP の打ち切り数  $K^{(1)}, K^{(2)}$  となる.

\*1 DPM はノンパラメトリックベイズモデルであり, 混合数は理論的に無限である. すなわち, iSVM はモデルとして無限の SVM モデルを持つ. これが, iSVM が無限混合 SVM と呼ばれるゆえんである.

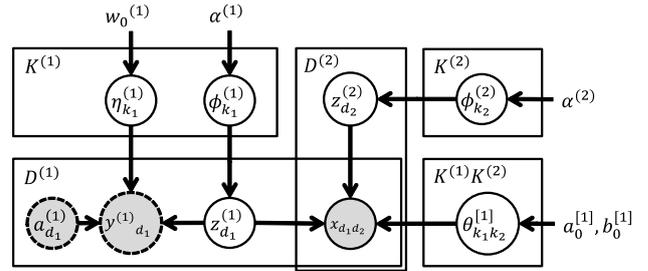


図4 提案モデル R-iSVM のグラフィカルモデル.

## 4. R-iSVM (Relational Infinite-SVM)

R-iSVM (Relational Infinite-SVM) は iSVM [7] を IRM [5] の考え方にもとづき関係モデル化する.

### 4.1 R-iSVM のモデルと生成過程

R-iSVM のグラフィカルモデルを図4に示す. ここでは  $D^{(1)}$  個の顧客と  $D^{(2)}$  個の商品を関係モデルにより接続している. 顧客クラスタに関しては, 未知の顧客属性  $y_{d_1}^{(1)}$  を既知の顧客属性  $a_{d_1}^{(1)}$  から予測するため, SVM による予測モデル  $\eta_{k_1}^{(1)}$  を関連付ける. なお, 破線のノードは SVM の予測結果/入力であり確率変数ではないが, SVM が  $y_{d_1}^{(1)}$  の予測時に  $\eta_{k_1}^{(1)}$  と  $z_{d_1}^{(1)}$  を参照することから, 分かりやすさのためにグラフィカルモデル中に記載している.

図4のモデルの生成過程を以下に記す. 下線を引いた1bが通常のIRMと異なる部分であり, IRMによって生成された各顧客クラスタ毎に, SVM の分離平面が生成される. R-iSVM の学習では, この逆問題として各クラスタで SVM の分離平面を推定することになる (詳しくは後述する).

1. 各顧客クラスタ  $k_1 = 1, \dots, K^{(1)}$  について,
  - (a) SBP によりクラスタ  $k_1$  の重み  $\phi_{k_1}^{(1)}$  を生成

$$v_{k_1}^{(1)} | \alpha^{(1)} \sim \text{Beta}(1, \alpha^{(1)})$$

$$\phi_{k_1}^{(1)} = v_{k_1}^{(1)} \prod_{k'_1=1}^{k_1-1} (1 - v_{k'_1}^{(1)})$$

- (b) 顧客の属性推定モデル  $\eta_{k_1}^{(1)} | w_0^{(1)} \sim \text{N}(0, w_0^{(1)})$  を生成

2. 各商品クラスタ  $k_2 = 1, \dots, K^{(2)}$  について,
  - (a) SBP によりクラスタ  $k_2$  の重み  $\phi_{k_2}^{(2)}$  を生成

$$v_{k_2}^{(2)} | \alpha^{(2)} \sim \text{Beta}(1, \alpha^{(2)})$$

$$\phi_{k_2}^{(2)} = v_{k_2}^{(2)} \prod_{k'_2=1}^{k_2-1} (1 - v_{k'_2}^{(2)})$$

3. 顧客クラスタと商品クラスタの組み合わせ  $k_1 = 1, \dots, K^{(1)}$  と  $k_2 = 1, \dots, K^{(2)}$  について, クラスタ関連度  $\theta_{k_1k_2}^{[1]} \sim \text{Beta}(a_0^{[1]}, b_0^{[1]})$  を生成
4. 各顧客  $d_1 = 1, \dots, D^{(1)}$  について, 顧客  $d_1$  のクラスタ割り当て

$$z_{d_1}^{(1)} | \phi^{(1)} \sim \text{Multinomial}(\phi^{(1)})$$

を生成

5. 各商品  $d_2 = 1, \dots, D^{(2)}$  について, 商品  $d_2$  のクラスタ割り当て

$$z_{d_2}^{(2)} | \phi^{(2)} \sim \text{Multinomial}(\phi^{(2)})$$

を生成

6. 各顧客  $d_1 = 1, \dots, D^{(1)}$  と各商品  $d_2 = 1, \dots, D^{(2)}$  について, 顧客  $d_1$  による商品  $d_2$  の購買情報

$$x_{d_1 d_2}^{[1]} | z^{(1)}, z^{(2)}, \theta^{[1]} \sim \text{Bernoulli}(\theta_{z_{d_1}^{(1)} z_{d_2}^{(2)}}^{[1]})$$

を生成

以上のように, R-iSVM は IRM と iSVM を組み合わせた生成過程を持ち, 関係情報を利用してデータをクラスタリングし, クラスタ毎に SVM のモデルを学習する. これにより, 実用上の効果として

- iSVM が (顧客の年齢や身長など) データ自体に静的に与えられた性質でデータをクラスタリングするのにに対し, R-iSVM は (顧客の実際の購買情報など) 動的な情報を利用してデータをクラスタリングしてクラスタ毎に予測モデルを構築するため, 現状をより反映した予測モデルの構築が期待できる.
- R-iSVM は予測モデルの予測性能が向上するように IRM のクラスタリングへ補正を加えるため, 生成モデル化が難しく従来はクラスタリングに用いることのできなかつた属性 (パラメトリックに確率分布をあてはめにくい属性) を予測モデルに含めることで, 間接的にクラスタリングに用いることができる.

が得られる.

## 4.2 R-iSVM の学習

R-iSVM は生成モデルとして扱うことのできない SVM をモデル中に含むため, ベイズ推論でよく用いられる MCMC や変分ベイズ法を直接適用することができない. そこで, 本稿は [7] にならない, IRM の変分ベイズ推論と SVM モデルの学習がどちらも関数の最小化として扱えることを利用し, R-iSVM の変分ベイズ法による学習問題を, これら関数の最小化を同時におこなう次の問題として定義する.

**Problem 1** (R-iSVM 変分ベイズ学習問題).

$$\begin{aligned} \min_{q \in \mathcal{Q}, \xi} & -C_2^{(1)} F_{IRM} + KL(q(z^{(1)}, \eta^{(1)}) || p(z^{(1)}, \eta^{(1)})) + C_1^{(1)} \sum_{d_1}^{D^{(1)}} \xi_{d_1}^{(1)} \\ \text{s.t.} & l_{d_1}^{\Delta}(y_1) - \mathbb{E}_{q(z^{(1)}, \eta^{(1)})} [\eta_{z_{d_1}^{(1)}}^{(1)}]^{\top} f_{d_1}^{\Delta}(y_1) \leq \xi_{d_1}^{(1)}, \xi_{d_1}^{(1)} \geq 0, \forall d_1, y_1 \end{aligned}$$

第一項は IRM の負の変分下限であり, この最小化は IRM の変分ベイズ推論 [3, 6] に対応する \*2. 第二項, 第三項, そして制約項は多クラス構造 SVM [4] の学習問題に対応する. すなわち, 問題 1 は IRM の変分ベイズ推論と多クラス構造 SVM の 2 つ同時に学習する問題である. これらは変分事後分布  $q$  を通じて互いに依存しあうため, 交互の最適化が必要となる. なお,  $C_2^{(1)} (> 0)$  は SVM が IRM の関係クラスタリングに与える効果を調節するパラメータであり,  $q \in \mathcal{Q}$  は変分事後分布の平均場近似をあらわす.

\*2 変分ベイズ法は変分下限  $F_{IRM}$  の最大化であるが, SVM の学習問題が最小化問題であるため, ここでは負の変分下限  $-F_{IRM}$  の最小化として扱い, SVM の学習問題と結合している.

本稿の以降では, R-iSVM 変分ベイズ学習問題 1 に対する解法として, 変分ベイズ法と二次計画法の組み合わせによる最適化アルゴリズムを示す.

### 4.2.1 変分事後分布の更新

R-iSVM 学習問題 1 についてラグランジュ関数を定め双対問題を解くと, IRM の変分ベイズ推論 [3, 6] と同様に, IRM の変分事後分布の更新式が得られる. IRM と R-iSVM の変分事後分布の更新式はほぼ同一であるため, IRM と異なる  $q(z_{d_1}^{(1)} = k_1)$  の計算についてのみ述べる. R-iSVM における顧客  $d_1$  のクラスタ  $k_1$  所属確率  $q(z_{d_1}^{(1)} = k_1)$  の更新式は

$$\begin{aligned} \log q(z_{d_1}^{(1)} = k_1) &= \log \hat{\phi}_{d_1, k_1}^{(1)} \\ &\propto \mathbb{E}_{q(v_{k_1}^{(1)})} [\log p(z_{d_1}^{(1)} = k_1 | v_{k_1}^{(1)})] \\ &+ \rho \sum_{d_2}^{D^{(2)}} \sum_{k_2}^{K^{(2)}} \hat{\phi}_{d_2, k_2}^{(2)} \mathbb{E}_{q(\theta_{k_1 k_2}^{[1]})} [\log p(x_{d_1 d_2}^{[1]} | \theta_{k_1 k_2}^{[1]})] \\ &+ (1 - \rho) \left\{ \sum_{y_1}^{Y^{(1)}} \omega_{y_1, d_1}^{(1)} \mathbb{E}_{q(\eta_{k_1}^{(1)})} [\eta_{k_1}^{(1)}]^{\top} f_{d_1}^{\Delta}(y_1) \right\} \quad (1) \end{aligned}$$

となる. 第一項と ( $\rho$  を除いた) 第二項については通常の IRM と同一である. 注目すべきは第三項であり, これはデータ  $d_1$  に対応するデータをクラスタ  $k_1$  の SVM 予測モデル  $\eta_{k_1}$  で分類した際のスコアをあらわし, 分類がうまくいくほど高い値となる. すなわち, R-iSVM はデータのクラスタ割り当てを決める際, 生成モデルから算出されたクラスタ割り当ての確率値を, そのデータの分類がよりうまくゆくように補正する. これにより, 単純に IRM でクラスタリングをした後に SVM モデルをつくる場合と比べ, 高い予測性能が期待される. また, SVM の教師情報がクラスタリングに影響をあたえることで, 関係クラスタリングの質が向上することも期待される.

### 4.2.2 ラグランジュ乗数の算出

式 1 に示した変分事後分布  $q(z_{d_1}^{(1)} = k_1)$  の更新式では, 確率の補正項にラグランジュ乗数  $\omega_{y_1, d_1}^{(1)}$  が含まれる. R-iSVM はこの値を得るため, SVM の双対問題に対応する以下の二次計画問題を変分ベイズの各イテレーション毎に解く.

**Problem 2** (Relational-iSVM 二次計画問題 (Dual)).

$$\begin{aligned} \max_{\omega^{(1)}} & -\frac{1}{2} \sum_{k_1}^{K^{(1)}} \hat{\mu}_{k_1}^{(1)\top} \hat{\mu}_{k_1}^{(1)} + \sum_{y_1}^{Y^{(1)}} \sum_{d_1}^{D^{(1)}} \omega_{y_1, d_1}^{(1)} l_{d_1}^{\Delta}(y_1) \\ \text{s.t.} & 0 \leq \sum_{y_1}^{Y^{(1)}} \omega_{y_1, d_1}^{(1)} \leq C_1^{(1)}, \forall d_1 \quad (2) \end{aligned}$$

なお,  $\hat{\mu}_{k_1}^{(1)}$  はクラスタ  $k_1$  における SVM 分離平面の変分事後分布  $q(\eta_{k_1}^{(1)})$  の平均パラメータであり,

$$\hat{\mu}_{k_1}^{(1)} = w_0^{(1)} + \sum_{d_1}^{D^{(1)}} \hat{\phi}_{d_1, k_1}^{(1)} \sum_{y_1}^{Y^{(1)}} \omega_{y_1, d_1}^{(1)} f_{d_1}^{\Delta}(y_1) \quad (3)$$

である.

### 4.2.3 R-iSVM の学習アルゴリズム

以上から, R-iSVM の学習アルゴリズムはアルゴリズム 1 に示すような変分ベイズ法と二次計画法の組み合わせによる最適化となる.

データセット (タスクの予測対象)	Y	正解率 (F 値)		
		Linear R-iSVM	IRM + S-SVM	S-SVM
Synthetic Data	2	<b>0.99 (0.99)</b>	0.96 (0.95)	0.58 (0.53)
MovieLens 100K (ユーザの “gender” 属性)	2	<b>0.63 (0.33)</b>	0.62 (0.31)	0.58 (0.27)
MovieLens 100K (ユーザが “Horror” 映画を高く評価するか)	2	<b>0.78 (0.61)</b>	0.78 (0.59)	0.74 (0.59)
MovieLens 100K (ユーザの “occupation” 属性)	21	0.13 (0.16)	<b>0.16 (0.18)</b>	0.09 (0.11)

表 1 各データセット, タスクにおける実験結果. 数値は正解率であり, カッコ内は F 値. Y は分類タスクのクラス数.

---

#### Algorithm 1: Relational-iSVM の学習アルゴリズム

---

```

while not converged do
  begin: (QP-Step) Solve SVM's dual QP:
    Solve QP 2 to obtain lagrangeans.
    for  $k_1 \in K^{(1)}$  do
      Update  $q(\eta_{k_1}^{(1)})$  using the obtained
      lagrangeans.
  begin (VB E-Step) Update latent variables:
    for  $d_1 \in D^{(1)}$  do
      for  $k_1 \in K^{(1)}$  do
        Update  $q(z_{d_1}^{(1)} = k_1)$  according to 1.
    for  $d_2 \in D^{(2)}$  do
      for  $k_2 \in K^{(2)}$  do
        Update  $q(z_{d_2}^{(1)} = k_2)$ .
  begin (VB M-Step) Update parameters:
    for  $k_1 \in K^{(1)}$  do
      Update  $q(v_{k_1}^{(1)})$ .
    for  $k_2 \in K^{(2)}$  do
      Update  $q(v_{k_2}^{(2)})$ .
    for  $(k_1, k_2) \in K^{(1)} \times K^{(2)}$  do
      Update  $q(\theta_{k_1 k_2}^{[1]})$ .

```

---

## 5. 実験

R-iSVM を実データに適用し, その予測性能を従来の方式と比較する. 比較対象の方式は, 多クラス構造 SVM [4] (S-SVM) と, IRM でデータをクラスタリングした上でクラス毎に S-SVM のモデルを構築するもの (IRM + S-SVM) とした. 各方式のパラメータは同じパラメータ空間でのグリッドサーチにより決定した.

### 5.1 人工データ

はじめに, 人工的に生成したデータに対する結果について述べる. 4.1 節で述べた R-iSVM の生成過程に従い, 顧客データ, 商品データ, そして顧客による商品の購買データを生成した. データ数は顧客数  $D_1 = 1000$ , 商品数  $D_2 = 300$  とした. 顧客の既知属性は, 顧客のクラス毎に生成された分離平面  $\eta_k$  を平均とする多変量正規分布から生成し, 未知属性は既知属性が分離平面のどちらに位置するかで正か負に設定した.

人工データに対する実験結果を表 1 の Synthetic Data に示す. R-iSVM が正解率, F 値共に最も良い. また, 生成した人工データの性質上, 真のクラスが推定できていれば正解率と F 値はどちらも 1 となることから, R-iSVM が IRM と比べ真のクラスに近い推定結果となったことが推測できる.

## 5.2 MovieLens

次に, 著名な関係データセットである MovieLens 100K のデータにおける, ユーザの属性情報予測の結果を述べる. 本実験では関係データによる属性生成として, ユーザの属性にユーザが過去に高く評価した映画のジャンルを 0/1 ベクトル化したものを追加し用いる. データ数は, ユーザ数  $D_1 = 943$ , 映画数  $D_2 = 1692$  である.

ユーザの性別, ホラー映画に対する評価, 職業を予測した結果を表 1 に示す. 性別とホラー映画の評価に関しては, R-iSVM が最も良い結果を示している. 他方, 職業については IRM+S-SVM が最も良い. R-iSVM が IRM+S-SVM よりも悪い結果となった理由は, 職業予測がクラス数 21 の多クラス問題であり, 各クラスで各クラスについて十分な数の訓練データが得られなかったことと推察される. 今後, データ数を増やして実験をおこなう予定である.

## 6. 結論

本稿は, iSVM (Infinite-SVM) [7] を IRM (Infinite Relational Model) [5] にもとづき関係モデル化した R-iSVM (Relational Infinite-SVM) を提案した. R-iSVM は関係データを用いることで SVM の予測性能を向上させる. 実験の結果, 未知属性の予測タスクにおいて, R-iSVM が従来の関係データを用いた属性生成と比べて高い性能を実現することが確認できた.

## 参考文献

- [1] C. E. Antoniak. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *Ann. Statist.*, 2(6):1152–1174, nov 1974.
- [2] R. Collobert, S. Bengio, Y. Bengio. A Parallel Mixture of SVMs for Very Large Scale Problems. *Neural Computation*, 14(5):1105–1114, 2002.
- [3] K. Ishiguro, I. Sato, N. Ueda. Collapsed Variational Bayes Inference of Infinite Relational Model. *arXiv*, 1:arXiv:1409.4757 [cs.LG], 2014.
- [4] T. Joachims, T. Finley, C. N. J. Yu. Cutting-Plane Training of Structural SVMs. *Machine Learning*, 77(1):27–59, 2009.
- [5] C. Kemp *et al.* Learning Systems of Concepts with an Infinite Relational Model. In *AAAI*, pages 381–388, 2006.
- [6] Z. Xu. *Statistical Relational Learning with Nonparametric Bayesian Models*. PhD thesis, 2007.
- [7] J. Zhu *et al.* Infinite SVM: a Dirichlet Process Mixture of Large-margin Kernel Machines. In *ICML*, pages 617–624, 2011.