

# しましま/IBIS2017

[Top](#) / [しましま](#) / IBIS2017

## 第20回 情報論的学習理論ワークショップ (IBIS2017) <sup>±</sup>

このページはしましまが IBIS2017 に参加してとったメモです。私の主観や勘違いが含まれていたり、私が全く分かってなかったりしていますので、その点を注意してご覧ください。誤りがあれば、指摘してください。

- 第20回 情報論的学習理論ワークショップ (IBIS2017)
- 11月8日 (水) : ワークショップ 第1日
- 招待講演1: Supervised Learning without Discrimination
- 招待講演2: Advice to Young and New AI Scientists
- 11月9日 (木) : ワークショップ 第2日
- 企画セッション: 自然言語処理への機械学習の応用
  - 深層学習と自然言語処理
  - 音声対話アシスタントに関する最近の研究動向とYahoo! JAPAN研究所での取り組み
  - ニューラル機械翻訳の動向
- 企画セッション: 実社会への機械学習の応用
  - 自動車の自律化におけるロボティクスと学習
  - 実社会・実環境におけるロボットの機械学習
  - マルチモーダルカテゴリゼーション: 階層ベイズモデルに基づくロボットによる概念・言語獲得
- 11月10日 (金) : ワークショップ 第3日
- 企画セッション: 画像処理への機械学習の応用
  - 機械学習を用いた画像特徴量変換
  - 機械学習による視線推定とその実世界応用
  - ディープラーニングによる画像変換
- 招待講演: 学習理論よ何処へ

1

## 11月8日 (水) : ワークショップ 第1日 <sup>±</sup>

1

### 招待講演1: Supervised Learning without Discrimination <sup>±</sup>

Nathan Srebro (TTI-Chicago)

- 統計的予測が人種や性別で差別的にならないようにする
- たとえば、保護属性  $A$  が予測関数  $\hat{Y}(X)$  に含まれていなくても、 $A$  に依存する可能性
  - すくなくともテストのときは  $A$  を使って補正する必要
- demographic parity:  $A$  の各特徴ごとの割合を合わせる  $\rightarrow$  誤差を細かく制御できない
- equalized odds: 観測値が与えられたとき、予測値と保護属性が独立  $\rightarrow$  属性ごとに誤差が同じ
  - $\hat{Y}$  が実数のときは、ROC曲線を一致させるようにする

1

### 招待講演2: Advice to Young and New AI Scientists <sup>±</sup>

Edward Albert Feigenbaum

- If you see an area in which many people are working, then work somewhere else

知能とは

- 知覚 perception
  - deep learning: 数秒 (データから統計的の情報)
  - 知識を直接推論に使う
- 認識 cognition
  - 知識は高次の空間から部分空間を選ぶのに使う
  - deep thinking: 数年かかることも (シンボリックなデータからの情報), 論理・ベイズに説明
  - 狭い範囲の対象しか現状では扱えない (エキスパートシステム)
- Tim was mad at Joe, because he stole his lunch : because を so にすると, he/his の照応先が変わる  $\rightarrow$  世界についての知識がないと照応ができない
- 知識をどうやって得るのか?
  - いろいろ詳細な知識を集めても, edge caseがある.  $\rightarrow$  ML では edge case には対応できない
  - edge case は文化・社会的な手段で解決するもの
- 今後のAIシステム
  - 現在のAIシステムは, 完全なチェスの手を, 火事になっている部屋の中で指すもの

- 人と協調する augmented intelligence の方向へ
  - EURISKOシステム 1979: Trillion Credit Squadron (複雑な海戦ゲーム) で局所探索をするAIを使って人間が強くなった
- 説明のできるAI: 医療や法務関係では説明・ストーリーが必要に
- 50年代の「AI自体の発見」にあたるようなブレイクスルーをめざそう
  - Try to be paradigm-shifter, not an incremental scientist. and to not be afraid of failures. Failures are the learning step to success.

## 11月9日 (木) : ワークショップ 第2日 <sup>↑</sup>

### 企画セッション: [自然言語処理](#) への [機械学習](#) の応用 <sup>↑</sup>

#### 深層学習と[自然言語処理](#) <sup>↑</sup>

鶴岡 慶雅 (東京大学)

- 品詞予測 → bidirectional LSTM (BiLSTM): 入力→単語, 出力→品詞
  - CRF層を追加した性能向上
- チャンキング (shallow parsing): 意味のある節に分割 → 節の始めや終わりといったタグの予測問題と考えると品詞予測と同様の方法で解ける
- 固有表現認識: 文中の固有表現 (バイオなら遺伝子名など) を抽出, チャンキングと同様にできる
- 構文解析
  - 句構造解析: 主語・述語などの全体の構造を獲得
    - 木構造を深さ優先で系列の処理問題に変換
    - Shift-Reduce法: 何かの構造右側に増やしていくか, [部分木](#) を戻すかのタグを学習する → 縮むことのできる Stack LSTM
- 機械翻訳
  - 翻訳元 encoder RNN と翻訳後 decoder RNN をくっつける
  - アテンション: encoder と decoder を連結する部分で, 前後のどの部分を重視するか の重みも学習しておく
- 会話[モデル](#): 会話文の前の文とそれに対する応答を encoder-decoder で学習する
  - 電子メールの返信の自動生成
- 畳み込み[ニューラルネットワーク](#) (CNN): ネットの結合の局所化とパラメータの共有
- 画像をCNNで符号化し, 自然言語文を復号化 → 画像やビデオに対する説明文を作れる
- CNNによる文の分類: RNNの代わりにCNNを用いることも
- 質問応答
  - Dynamic Coattention net: 質問文・回答文を潜在空間に埋め込んで, その対応をencoder-decoderで学習 → Wikipedia の該当箇所を見つけることができるように
  - 推論を必要とする場合: 答えを導出するために必要な文を推定し, それらを潜在空間に埋め込んであとは同じ (?? memory network)
- pointer networks: 要約のための単語を選択する
  - magic the gathering のカードの能力をプログラムに変換するといったことも
  - Seq2QL: 自然言語をSQL文に変換 → SQL 文だけでなく, SQL文で実際に獲得できたデータも教師情報も使える
    - ある意味実世界への影響を解釈したともみなせるので, 文意の解釈ともみなせる

#### 音声対話アシスタントに関する最近の研究動向とYahoo! JAPAN研究所での取り組み <sup>↑</sup>

鍛冶 伸裕 (ヤフー株式会社)

- 音声対話: Yahoo! 音声アシスト, Siri, Google Home
- 対話システムはAIの古典的問題, Eliza (Weizenbaum 66), SHRDLU (Winograd 72)
  - 現在では[強化学習](#)に基づく理論的枠組みが既に確立されている → まだ課題はある?
- 実際に運用して顕在化する問題 (新ドメインの迅速な追加, タスクと雑談の切り分け, ユーザ満足度の自動化, システムエラーの自動検出)
- 従来システム: 対象が限定的 → 天気・探索・カレンダーなど複数タスクを扱うように
  - ドメイン適応: タスクに合わせて[モデル](#)を変形 → 実運用ではドメインの追加が起こる
    - ドメイン数を適切に抑えるのは: 入力を低次元に射影して分類器にかける問題 → ことなる仮想のドメインをアテンションの枠組みで組み合わせて使う
- タスク・雑談の区別
  - 宛先付きの Twitter を雑談発話, Webの検索ログを発話要求と見なしコーパスに使う
- 対話システムのユーザ評価
  - 本物のユーザを相手にした対話処理には正解がないので, 客観的評価が難しい
  - 利用者の行動パターンを使った予測 → 対話のアクションをラベル付けして, その[系列パターン](#)から満足・不満を分類
    - さらに, クリック, ユーザ発話, アクション発話などの追加特徴を用いる (コマンドの繰り返しなどは不満の現れ)
- 明示的なアノテーションをせずに, 対話システムを成長させる
  - ユーザの修正発話: 所望の結果が得られなかったときに, より詳しい発話をユーザが行う
  - エラーの発生源が特定できれば, 自動訂正ができるように
  - 発話の時間間隔などの補助情報も活用してエラーの発生源を特定できる

## ニューラル機械翻訳の動向 <sup>↑</sup>

中澤 敏明(科学技術振興機構(JST)) @Tzawa

- 2016年のGoogleの8言語の翻訳でニューラル翻訳は注目された
  - かなり向上した、英語を経由して翻訳しているようだ
- 以前のフレーズベース機械翻訳: 言語 **モデル** などを用いたもの、最もスコアの高くなる系列が翻訳文になる
- ニューラル翻訳: 入力文→出力分 の直接対応
- seq-2-seq: 入力・出力の長さが違うので二つのRNNを繋げる → 一つの低次元ベクトルに情報を詰め込んでるので限界がある
- アテンション: 入力系列中で次にどの部分を変換するかを、重み付けにより特定する
- Google翻訳では、入出力ともRNNをそれぞれ8層にしている
- ビームサーチ: 翻訳の途中の結果で良さそうなものを複数保持する
- ニューラル翻訳の長所: 入力を過不足なく網羅?
- ニューラル翻訳の短所: BLEUの **最適化** ではない、埋め込みはほとんどない単語の置換を引き起こす、アテンションはあまり信用できない、訳文の対応箇所は分からない
- 扱える語彙数が少ない
  - softmax の計算コストのため語彙数を制限する必要
  - 低頻度語をあとで辞書で復元、単語ではなく文字を使う(知らない単語のみ文字にする、サブワード: 高頻度の連続文字パターンを単語にする)
- 訳抜けと重複
  - アテンションが当たらないのが問題視されている
- 文の構造の学習
  - デコーダをRNN以外にするなどしてみる試み
- 多言語翻訳
  - 直接のend-to-endコーパスがなくても似た言語の **モデル** で翻訳するzero-shot翻訳
  - 言語のencoderやdecoderは言語依存だが、アテンションは共通
- **モデル** の軽量化
  - 蒸留など
- その他の **モデル**
  - RNNより計算コストの小さなCNNの利用
  - feed-forwardだけの **モデル**: self-attention機構というのがあり、代名詞の照応ができる
  - NMTの性能向上はかなり限界に達している
  - 単言語コーパス: encoderは言語間で共通で、デコードだけ言語依存
- 日本では対訳コーパスを集める

## 企画セッション: 実社会への **機械学習** の応用 <sup>↑</sup>

### 自動車の自律化におけるロボティクスと学習 <sup>↑</sup>

竹内 栄二郎(名古屋大学)

- 車輪付きロボットが目的のところに移動する知的システム
- 近年の自動運転研究開発の活発化: DARPA GrandChallenge, UrbanChallenge, Google: UrbanChallenge研究者が結集, Daimler: 2013年に100kmの市街地自動運転
- 名大の自動運転環境: 車3台, 模擬市街地, ガイドラインの制定
- 3次元地図情報: レンジセンサーから作成, 人手がかかる
- 位置情報確定: 3次元ライダーと3D地図のマッチング → 地図から信号なども認識できる
- 運転支援: 人がいたら減速, 運転者が意識を失うと安全なところで停車
- Autoware: 自動運転のオープンソースプラットフォーム
- 運転行動信号処理
  - 正しい状態だけではなく、異常状態から場合のデータも必要
- プランニングの拡張
  - 観測状態(見えているかどうか)をプランニングに導入 → 見通しのよくない道路に出るときは左右確認などの行動が獲得できる
- **最適化** 問題として解くと高計算コスト → 準最適解に
  - 深層学習で使う学習における **最適化** 問題として解く
- 物理・確率 **モデル** に基づく **最適化** 問題 → **機械学習** を使ったアプローチになる?

### 実社会・実環境におけるロボットの **機械学習** <sup>↑</sup>

高橋 城志(Preferred Networks, Inc.)

従来のロボット

- 認識(センサーと **モデル** のすりあわせ) → 計画(どう動くか) → 実行のループ
  - 行動した結果が予測と異なることは多くある

- 自由度: ざっくりと関節の数, 人間の肩は? 7自由度, 物体を自由に動かすには6自由度あればいい, それ以上は冗長自由度
- 順運動学: 関節角度と腕の長さで手先の位置は分かる ⇔ 逆運動学: 先端位置から関節状態, 解は一意にならない → 4自由度以上は解析解は難しい, 冗長自由度があると解は不定に
  - 数値解析的に解く → 目標と現状の差分から更新
- 現実についての内部**モデル**がある → 現実と乖離することがある
  - 速さ, 正確さ, コスト効率に優れるアプローチ
  - 工場など環境が安定していればいいが, オープンな環境では適応できない, 少量多品種, 複雑な**モデル**(お弁当を詰めるタスク)

#### 機械学習を用いたロボットの制御

- 「認識」の部分を**機械学習**で置き換える
- 掴む場所を経験により獲得する
  - 画像とレンジセンサーが入力 → 物体ごとに画像をセグメンテーションし, 掴む場所を出力
- 身体性人工知能: ハードとソフトが不可分
  - 環境の変動は邪魔なものではなく, それを利用して獲得
  - Brooks の subsumption architecture: 単純な機械構造

#### 機械学習を使った運動生成

- 欠点: 試行回数の多さ(ロボットが壊れる, 修理すると以前とは変わってしまう), シミュレーションとの差, 未知環境には弱い
- 試行回数の多さ
  - 模倣学習 (apprentice learning): 自己符号化器(画像特徴の系列を出力) → 時間窓の関節角度を出力 (time-delay 自己符号化器)
  - シミュレーション+実機: シミュレーションでは風やケーブルの摩擦は扱えない + 実機だけだと壊れる → 交互に行う
  - 教師あり + **強化学習**: 最初に人間の行動から**教師あり学習**して, そこから**強化学習**
  - 画像処理は照明などの変化に弱い → シミュレーションでいろいろな状況を試す
- 未知環境への対応
  - 異なる方策を保存しておいて, それから選択して再利用する
- マルチモーダル学習
  - 画像を見ながらの対話: 曖昧な指示に対して聞き返しを行う

### マルチモーダルカテゴリゼーション: **階層ベイズモデル**に基づくロボットによる概念・言語獲得 <sup>1</sup>

中村 友昭(電気通信大学)

- 概念・言語学習: 人のように言語を獲得するロボットの実現
  - **概念形成**, 語彙の獲得(単語辞書) 語意の獲得(記号接地) 文法の獲得
- ここでの概念の工学的定義: **クラスタリング**により形成されたカテゴリ
  - 触覚, 聴覚, 視覚入力に同時に依存するLDA**モデル**を利用
  - **生成モデル**一部の感覚情報があれば, 他の感覚がどうなるかを予測できる
- 語彙の獲得
  - 語彙の教師なし獲得: 音声認識できない, 単語の切れ目が分からない
  - 概念の形成と言語**モデル**の獲得を同時に行う
    - 概念, 単語とそれらの対応を学習する**モデル**
  - 100時間ほどロボットへの物体の提示と発話による教示から500程度の概念を獲得
- 文法の獲得
  - 教示者の行動についての概念と, 概念の系列から文法を獲得
  - 違うタイプの概念を同時に扱う mMLDA, 文法**モデル**はマルコフ系の**モデル**
- センサ情報に基づく概念獲得
  - 動作の分節も行う
- 人の**概念形成**の年齢による進捗と, この**モデル**の比較も行っている

## 11月10日 (金) : ワークショップ 第3日 <sup>1</sup>

### 企画セッション: 画像処理への**機械学習**の応用 <sup>1</sup>

#### **機械学習**を用いた画像特徴量変換 <sup>1</sup>

産業技術総合研究所 小林 匠

- 特徴量変換: データによって処理を変えない汎用性と, 特徴の識別力を向上させる弁別性
- 距離尺度:
  - ヒストグラムに基づく距離: EMD, Faster EMD, SiftDist (SIFTでのEMD) Diffusion距離(ヒストグラム上の拡散過程)
- 特徴量のテンソル構造を保持: X, Y, 特徴の3次テンソル, 各軸ごとに沿ったベクトルを三つ作って, テンソル構造を含んだ特徴ベクトルを使う

- Structural similarity Index Measure (SSIM): 画像の品質評価
  - ここでは 明るさ, コントラスト, 構造 パターンの三つの要素の重み付き線形和を考える
  - これら三つの要素を, テンソルから取り出したベクトルを作って構成.
- CNNのための特徴変換
  - cnnで得た特徴をヒストグラム用の変換を使うと性能が良くない
  - 学習を使った特徴変換
    - 加法的な**カーネル**(いくつかの**カーネル**の和で, その重みを学習する問題, ここでは**カーネル**にフーリエ変換の台を使う)
  - 複数のタスクに対応できるように, マルチタスクの共通部分を使う → 行列分解を使って共通因子を求める

## 機械学習による視線推定とその実世界応用 <sup>↑</sup>

大阪大学 菅野 裕介

- 視線推定・アイトラッキング: 人物がどこを見ているかを推定
  - 専用ハードを使わずに, ウェアラブルカメラなどの画像から推定できるように
- 入力と方向のアノテーションから**機械学習**で求める → 人物や照明が変わるとできない
- 統制環境下で, 50人分の3次元撮影顔画像と視点とのペアのデータを作成した → 専用ハードには及ばないが, そこそこいけた
- ラップトップに注視点を表示して, さらにいろいろな環境に対応できることをめざす
- さらに, 3D**モデル**をCGで作り, そこに眼球を埋め込む形で, 照明環境も制御できるようにした → CGベースのデータで最初に学習して, そこから実データで詳細に調整すると良かった
  - さらに深層ネットを用いて9度ぐらいの誤差を達成
- 目の画像だけでなく, 顔の方向も利用するとさらに精度を向上
- 課題
  - 現状の全体の画像から, 注目しそうなポイントの候補を予め推定する
    - パブリックディスプレイなどを用いて, 注目を集めそうな部分のデータを集める
  - アイコンタクト検出: ウェアラブルカメラを付けている人とのアイコンタクト
    - 物体の中央に注視しやすい傾向を利用した候補抽出, カメラに近いものへの注目という仮定を使う

## ディープラーニングによる画像変換 <sup>↑</sup>

早稲田大学 飯塚里志

- Fully CNN: 全ての層が畳み込み層の自己符号化器構造のネット → 画像のend-to-end学習ができる
  - バッチ**正規化**: 各層をミニバッチごとに正規化することで, 深いネットの学習を可能に
- CNNで学習した特徴を, 別の目的で使うようにもなった
- 白黒写真の着色
  - 入力: 輝度画像 → 出力: カラー情報 の対応
  - セグメンテーションと認識をしてから, 適切な色を選ぶ
    - シーンを認識しないと, 適切な判断はできない
  - 大域認識とローカル認識をするネットで認識してから, 色との対応を付けるネットを付加
  - 利用者が指示した色を付ける ← 注目テーマ
    - 入力: グレースケール+利用者アノテーション → 出力: 着色
- 画像補完
  - 従来法: パッチ→大域的な構造を学習できない, 平均: ぼやけた学習になる
  - GANによる補完: 画像を生成する生成器と, 生成画像と真の画像を識別する識別器, 生成器は識別器を突破するように, 識別器は識別精度を上げるように更新
  - 生成器は補完ネット, 識別器は補完部分用と大域用と二つ
- ラフスケッチの線画化
  - 入力: ラフスケッチ, 出力: 線画 → fully CNN
  - データの難しさ, イラストレータはラフ画から線画にするときに変える場合がある, イラストレータが書ける枚数は少ない
  - データ拡張: トーン, プレ, 汚れを足す
  - しかし, end-to-endのデータだけではどうしても限界
  - GAN: 生成器→線画化ネット, 識別器: ネット出力と実物の識別

## 招待講演: 学習理論よ何処へ <sup>↑</sup>

東京工業大学 渡辺澄夫

「人間力」の成功と挫折

- 次々とやってくる例題に先生が印しを付けて, その印しを予測するという基礎研究
- 基礎研究から実社会の間には「死の谷」があるといわれている
  - 死の谷は人間力で突破せよ! との上司の仰せ
  - アウトプットに学問はいらないのか?



- 歴史は後戻りしない: 線形分離しかできないとかいろいろあったが, そういう挫折を繰り返してきて現在にいたる
  - この先 → 世界の隅々まで広がる? 世の中に残る? 無に帰る?
- みかけの達成度は人間力で向上するが, 本当の目に見えない達成度はむしろある点からは下がる → この構造は学習理論のそのもの, 数理科学はみかけと真の達成度の差を測る
- 過剰な人間力が未来を閉塞させている
  - 海外ではAI研究は社会に出て行くと言われているが, 日本では悲観的 → 日本では数理科学の割合が少ないからではなからうか?

#### 学習理論

- 学習理論: 真の分布  $q(x)$  → データ:  $X_1 \dots X_n$  → [事後分布](#) → 予測分布
  - 真の分布と予測分布の誤差を  $G$  とする →  $G$  についての法則

$$E[G] = d / (2n) + o(1/n) \dots d \text{ はパラメータ数, } n \text{ はデータ数}$$

- [モデル1](#): 平均  $a$  分散  $s$  のガウス, [モデル2](#): 0平均と $b$ 平均のガウスを $a:1-a$ で混合
  - [モデル1](#)は正則[モデル](#)だが, [モデル2](#)は階層性があるので非正則[モデル](#)
- 正則[モデル](#)では[事後分布](#)を[正規分布](#)で近似しても良い推定だが, 非正則[モデル](#)ではよくない
- 事実をありのままに見ることは難しい
  - 深層学習のような複雑な[モデル](#)ではガウスでは無理と30年ぐらい前から知られていたが, 人間力で無視していた.
- 漸近理論ではデータが無限にあれば[事後分布](#)は[正規分布](#) ⇔ 一定のデータで[モデル](#)が複雑になると[正規分布](#)では表せなくなる
- [汎化誤差](#)の法則

$$E[G] = \lambda / 2 + o(1/n) \dots \lambda \text{ は実対数閾値, } n \text{ はデータ数}$$

- 双有理な世界: 実対数閾値は不変量
- 漸近正規性では解明できなかった世界が解明されてきた → 代数統計学の流れの一つになった

#### 学習をめぐる旅

- AICは正則[モデル](#)での[汎化誤差](#)を推定するもの
- 古の統計モデリング: 統計[モデル](#)は人間力で作るもので, その適切さを測る方法はないと想われていた
- $E[\text{汎化誤差}] = E[\text{学習誤差}] + d / n$ 
  - 真の分布が分からなくても[汎化誤差](#)を推測できるというブレイクスルー
- [次元の呪い](#): 予め作られた[モデル](#)は高次元空間にある複雑が大きな情報を学習できない
  - 深層学習など階層構造のある[モデル](#)は[次元の呪い](#)を克服できる (バロン, 1993) → でも[モデル](#)は正則ではない
- WAIC  $E[\text{汎化誤差}] = E[\text{学習誤差}] + E[\text{事後ゆらぎ}] \dots E[\text{事後ゆらぎ}] = V_w[\log p(X|w)]$
- 特異点論 → 代数幾何 → 超関数 → 経験過程 → 学習理論 という過程をたどってWAICを導出
  - $E[\text{散逸}] = E[\text{揺らぎ}]$  揺動散逸定理 ... なぜか物理と同じような結果が → これは必然

#### 学習理論よ何処へ

- データ科学と[機械学習](#)
  - データ科学: データ → 個別[モデル](#) → 最適設計 → 予測
  - [機械学習](#): 不偏[モデル](#) → 学習理論 → データ → 予測
- 双対構造: 関数  $f$  とデータ  $x$  を入れ換える構造
  - データ科学はデータ $x$  についての研究 ⇔ [機械学習](#)は関数 $f$ についての研究 → state of art はだいたい同じになる
  - より豊かな未来のために「異なり続ける努力」も必要
- 学習理論と人間
  - 宇宙人としてあつたとき, [ニューラルネット](#)などは作られているだろうか
- [機械学習](#)はしばしば人間の考えとは異なる方向に到達することがある
  - [機械学習](#)のさらなる研究は「理解する」とは何であるかについての新しい世界に到達するだろう
  - 人間力も学習理論も変わる必要: 人間は事故を起こしても責任をとれるからいい ⇔ 責任をとれるなら事故を起こしてもいいのか?
- 自然と情報
  - シヤノンは[確率分布](#)  $q(x)$  は  $-\log q(x)$  で符号化できる ⇔ フォン・ノイマンは[エントロピー](#)であると答えた → 自然と情報の一致は偶然か?
- 5年ほど前に, 情報まで含めた形で[エントロピー](#)保存則が成立する → 情報は物理的な存在
  - 学習理論とは, 物理学と無矛盾にならなければならない

Last-modified: 2017-11-11 (土) 19:00:49 (2d)

Site admin: [朱鷺の杜Wiki管理者](#)

PukiWiki 1.4.7 Copyright © 2001-2006 [PukiWiki Developers Team](#). License is [GPL](#).  
Based on "PukiWiki" 1.3 by [yu-ji](#). Powered by PHP 5.6.31. HTML convert time: 0.219 sec.