

五十嵐 健夫

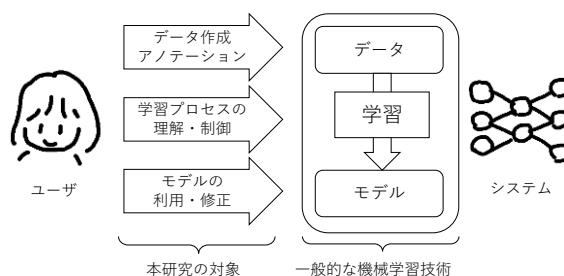
東京大学大学院情報理工学系研究科  
教授

## データ駆動型知的情報システムの理解・制御のためのインタラクション

### § 1. 研究成果の概要

本プロジェクトでは、機械学習におけるヒューマンファクタに関する研究に取り組んでいる。2019年度における主な成果は下記の3点である。

1 点目は、統計モデリングを調べるためのインタラクティブな散布図を備えた統合開発環境である。統計モデリングプログラムの開発には、プログラムの動作を観察および検証するためのサンプルデータが必要である。通常、このようなサンプルデータは、既存のデータセット



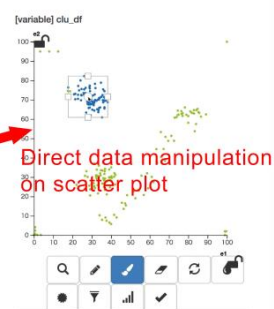
から取得されるか、コマンドを使用して合成される。場合によっては、任意のデータセットを直接設計したり、インタラクティブに変更したりしたい場合がある。しかし、現在の開発環境では、そういったことは困難である。本研究では、統計モデリングアルゴリズムの動作を効率的に理解するための手法として、コードエディターとインタラクティブな散布図エディターを組み合わせることを提案している。ユーザーは、散布図エディターでデータセットをインタラクティブに作成および変更し、システムはエディターでコードを継続的に実行し、データを入力として受け取り、結果をエディターに表示する。これにより、既存あるいは開発中のアルゴリズムの動作の理解を助けることができる。本研究成果は、ヒューマンコンピュータインタラクション分野のトップカンファレンスである ACM CHI 2020 に採

### Live programming system

#### Continuous execution

```
1 params = {'quantile': .3,  
2 'eps': .3,  
3 'damping': .9,  
4 'preference': -200,  
5 'n_neighbors': 10,  
6 'n_clusters': 3}  
7  
8 clu_df = pd.read_csv('./mysite/uploads/min_points.csv')  
9  
10  
11 #bind(clu_df)  
12  
13 clu_df['two_means'] = label_dict['two_means']  
14 #bind(clu_df, {"color": "two_means", "title": "two_means.cluste"}  
15
```

Data in variables is changed.



択された。

2点目は、三つ組みの比較データからの分類器学習手法の開発である。三つ組み比較では、ユーザが3つのデータを提示され、最初の1つが2つ目と3つ目のどちらに近いかを返答する。システムは、ユーザからの回答をもとに分類器を学習する。3つ組比較は、直接のラベル付けよりもユーザへの負荷が少ないものと期待される。本研究では、この問題について、はじめて3つ組比較のみから分類器を学習できることを示した。提案された手法は、ニューラルネットワークを含む任意の代理損失関数と任意のモデルに簡単に適用できるという利点がある。さらに、理論的に提案された経験的リスク最小化法の限界推定誤差を確立した。実験により、提案手法がさまざまなベースライン方法よりも優れていることを示した。

3点目は、画像ラベリング作業の効率化手法の開発である。手動で画像にラベルを付ける(画像に適切な「カテゴリ」を選択する)作業は、特に多数のカテゴリからラベルを選択する場合、非常に面倒で時間がかかる。この研究では、ワーカーがカテゴリグループ内の画像をサブカテゴリグループに再帰的に分類し、一度に1つのレベルで作業するラベリングタスクを限定する、階層割り当てを提案している。ユーザ実験により、ラベリングタスクの階層割り当てと通常の(非階層)割り当てとを比較し、階層割り当てによりラベル付けタスクを完了するために必要な合計時間が短くできることを示した。さらに、ラベリングプロセスの学習効果が、階層割り当てでより強く表れることを確認した。

#### 【代表的な原著論文】

1. "Kentaro Asai, Tsukasa Fukusato, and Takeo Igarashi. ""Integrated Development Environment with Interactive Scatter Plot for Examining Statistical Modeling."" In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2020).
2. Zhenghang Cui, Nontawat Charoenphakdee, Issei Sato and Masashi Sugiyama. Classification from Triplet Comparison Data. Neural Computation, Volume 32 | Issue 3 | March 2020 p.659-681.
3. Chia-Ming Chang, Siddharth Deepak Mishra and Takeo Igarashi. "A Hierarchical Task Assignment for Manual Image Labeling." The IEEE Symposium on Visual Languages & Human-Centric Computing (VL/HCC 2019), Memphis, Tennessee, US, 14-18 October 2019

## § 2. 研究実施体制

### (1) 研究代表者グループ

- ① 研究代表者:五十嵐 健夫 (東京大学大学院情報理工学系研究科創造情報学専攻 教授)
- ② 研究項目
  - ・研究プロジェクト全体のとりまとめ。
  - ・データ駆動型知的情報システムの理解
  - ・制御のためのインタラクション手法の研究開発
  - ・カメラや各種のセンサーからの入力を理解してユーザ支援を行うシステムの開発
  - ・メディアコンテンツの生成を補助する知的情報処理システムの開発

### (2) 共同研究者グループ(1)

- ① 主たる共同研究者:佐藤 一誠 (東京大学大学院新領域創成科学研究科複雑理工学専攻 講師)
- ② 研究項目
  - ・人の活動と先端インタラクション・インターフェースに基づく機械学習基盤技術の開発、および建築デザインへの応用

### (3) 共同研究者グループ(2)

- ① 主たる共同研究者:楽 詠瀬 (青山学院理工学部 准教授)
- ② 研究項目
  - ・クラウドソーシングと機械学習を活用した創造的データ生成のための基盤技術の開発と、メディアコンテンツの生成を補助する知的情報処理システムへの応用

### (4) 共同研究者グループ(3)

- ① 主たる共同研究者:金 太一 (東京大学医学部脳神経外科 助教)
- ② 研究項目
  - ・機械学習を用いた脳神経外科領域の診断・治療支援技術の開発