

アドバンス・トップエスイー 最先端ソフトウェアゼミ成果発表

AI・機械学習 グループ

2019年 3月 22日

太田 裕一

北野 健太

早川 芳昭

目次

1. 問題意識・目標

2. 取組内容

① 代表的な研究の調査・整理と概要説明

(1) 大域的、(2) 局所的、(3) 説明可能なモデル、

② クレジットカードデフォルト予測に代表手法を適用

(1) BATree、(2) LIME、(3) interpretable decision sets

③ 適用結果を踏まえた考察、評価

3. まとめ・所感

1. 問題意識・目標

前期は「機械学習」をテーマに活動した。後期は…

■現在のAIは説明が苦手

- 得意なこと
 - 高い精度での予測・認識
- 苦手なこと
 - 予測・認識の判断根拠の説明

判断根拠が説明できない

→AIへの不信／導入の阻害

「機械学習モデルの判断根拠の説明」大阪大学 産業科学研究所 原聡 引用

<https://www.slideshare.net/SatoshiHara3/ss-126157179?ref=https://stair.connpass.com/event/109983/presentation/>

2. 取組内容

以下を明確にするため、①～③の取り組みを実施した。

- ・機械学習の解釈性・説明性とは何か？
- ・どのような技術があるのか？
- ・機械学習の開発・運用・保守のどのプロセスで利用できるのか？
- ・誰向けの説明として使えるのか？

①代表的な研究の調査・整理と概要説明

②クレジットカードデフォルト予測に代表手法を適用

③ 適用結果を踏まえた考察、評価

2. 取組内容①代表的な研究の調査・整理

解釈性・説明性に関する近年の代表的な研究は以下の通り

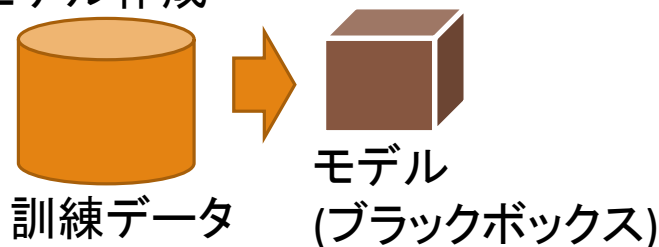
分類	概要	手法例
1.大域的な説明	複雑なブラックボックスモデルを可読性の高い解釈可能なモデルで表現することで説明とする方法。	Born Again Tree
2.局所的な説明	特定の入力に対するブラックボックスモデルの予測の根拠を提示することで説明とする方法。	LIME、S-LIME、Anchor
3.説明可能なモデルの設計	そもそも最初から可読性の高い解釈可能なモデルを作ってしまう方法。	Interpretable Decision Sets
4.深層学習モデルの説明	深層学習モデル、特に画像認識モデルの説明法。アプローチとしては2の局所的な説明に該当。	SmoothGrad

2. 取組内容①手法紹介(1) Born Again Tree

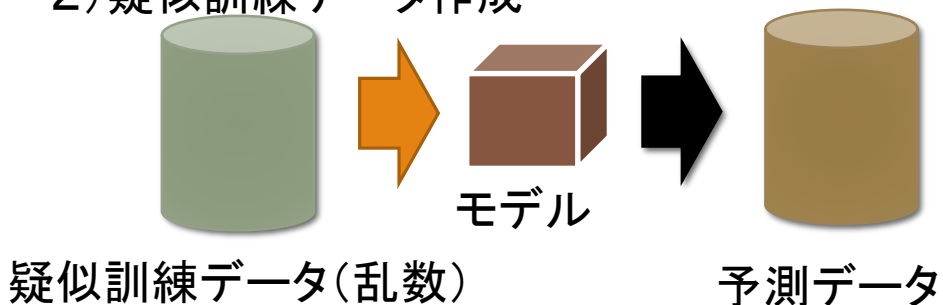
• 取り組み

- Leo Breimanの論文「Born Again Tree」では、アンサンブル学習、ニューラルネットワーク等、一度作成したモデルを用いて追加の教師データを大量に生成する
- この追加データを使って決定木を学習する。決定木のモデルのため、予測の背景を人間が読み取ることができる。

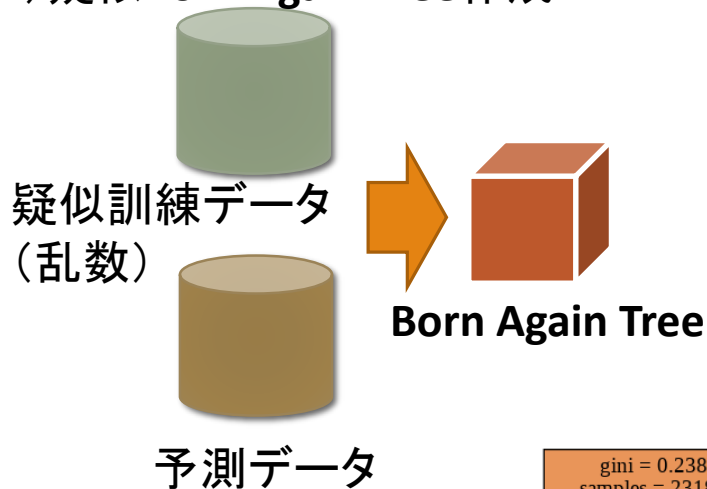
1) モデル作成



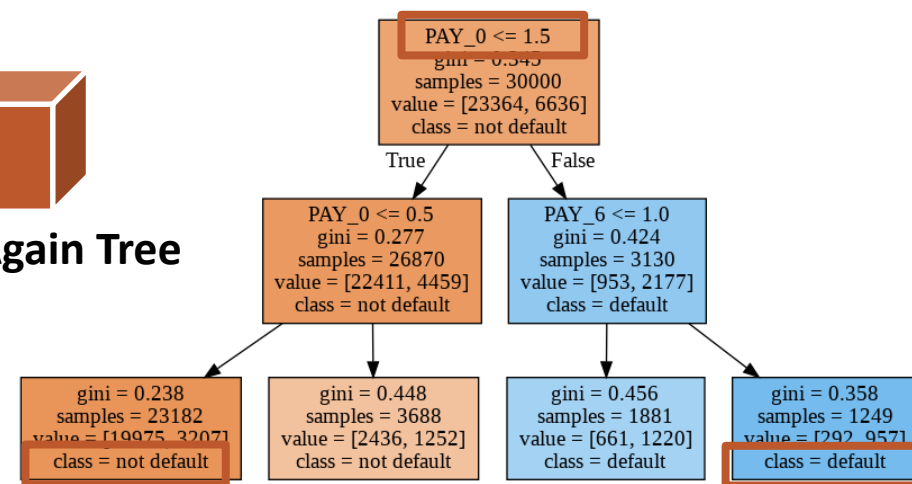
2) 疑似訓練データ作成



3) 疑似Born Again Tree作成



予測
訓練



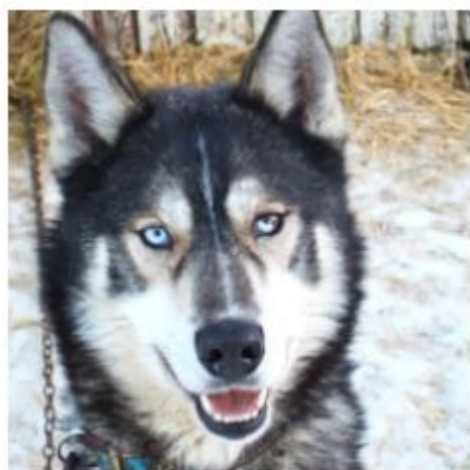
クレジットカード利用者における、ある月のデフォルト予測モデルの可視化例

2. 取組内容①手法紹介(2)LIME

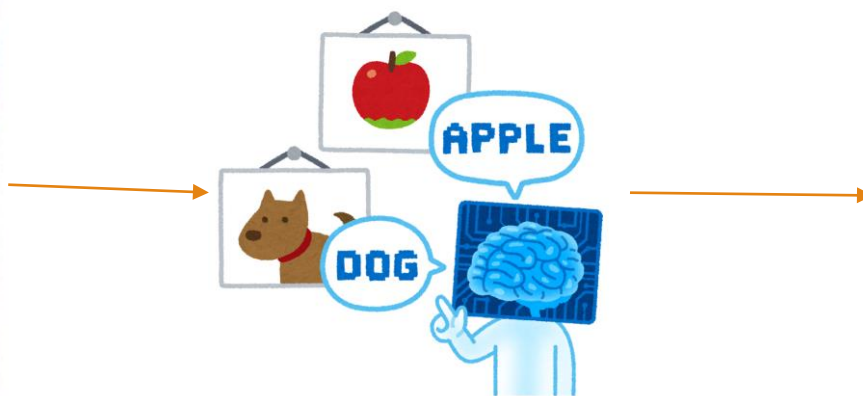
2.局所的な説明

KDD'16論文。解釈性研究の代表例として扱われることが多い。線形モデル/ルールモデルを用いた局所的な説明を生成する方法を提案。任意のモデルについて簡単に局所的な説明を生成できる点が優れている。

ある分類器では、ハスキーをオオカミとして分類してしまう。それはなぜか？



ハスキー

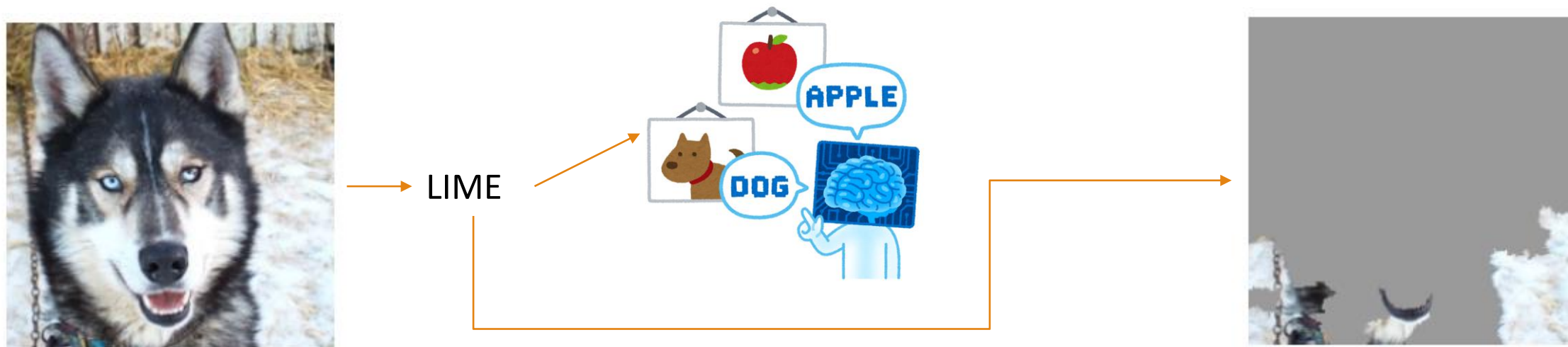


オオカミ

(参考) <https://publicdomainq.net/wolf-animal-0015458/>

2. 取組内容① (2) LIME

LIMEは、特定のクラスに対して正の重みを有するスーパーピクセルのハイライトする。
結果の画像から雪を特徴として、分類している可能性が見受けられる。



ハスキー

○推測

学習データのオオカミ画像の背景が雪景色なものが多いか？

○対応

背景によらないデータを収集し、再度学習させて、精度向上を狙う。

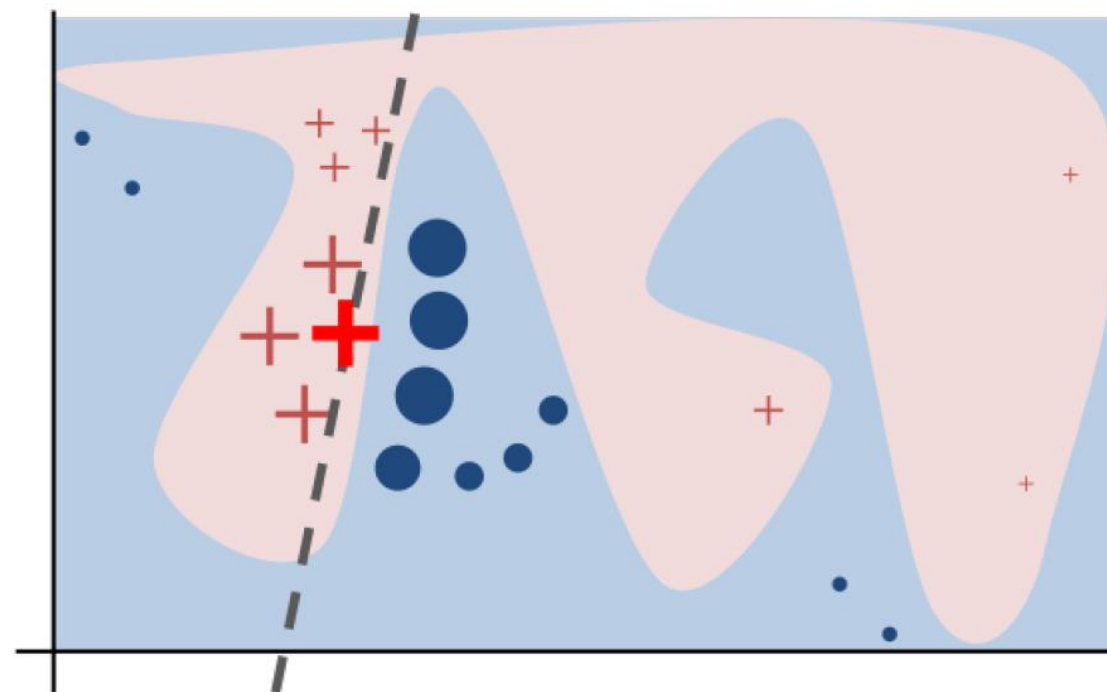
2. 取組内容① (2) LIME

局所的なデータに注目することで、線形な分類を行う。

ブラックボックスモデルの複雑な関数 F (LIMEには未知)。
データをピンクと青に分類する。
太赤十字は説明したいインスタンス。

LIMEは、データをサンプリングして、 f を使用して予測結果を取得する。

予測結果から、破線で示される説明可能な線形モデルを作成し、データの説明に使用する。



2. 取組内容①手法紹介(3) interpretable decision sets

提案手法

```
If Respiratory-Illness=Yes and Smoker=Yes and Age  $\geq$  50 then Lung Cancer
If Risk-LungCancer=Yes and Blood-Pressure  $\geq$  0.3 then Lung Cancer
If Risk-Depression=Yes and Past-Depression=Yes then Depression
If BMI  $\geq$  0.3 and Insurance=None and Blood-Pressure  $\geq$  0.2 then Depression
If Smoker=Yes and BMI  $\geq$  0.2 and Age  $\geq$  60 then Diabetes
If Risk-Diabetes=Yes and BMI  $\geq$  0.4 and Prob-Infections  $\geq$  0.2 then Diabetes
If Doctor-Visits  $\geq$  0.4 and Childhood-Obesity=Yes then Diabetes
```

一般的なデシジョンリスト

```
If Respiratory-Illness=Yes and Smoker=Yes and Age  $\geq$  50 then Lung Cancer
Else if Risk-Depression=Yes then Depression
Else if BMI  $\geq$  0.2 and Age  $\geq$  60 then Diabetes
Else if Headaches=Yes and Dizziness=Yes, then Depression
Else if Doctor-Visits  $\geq$  0.3 then Diabetes
Else if Disposition-Tiredness=Yes then Depression
Else Diabetes
```

- ・解釈可能なデシジョン・セット(interpretable decision sets)を提案
- ・解釈可能モデルの代表格であるデシジョンリストには欠点がある(決定木も同様)
- ・左右は同じものであるが、提案手法は、各ルールが階層ではなく、すべて独立したルールなのでわかりやすい
- ・一般的なデシジョンリストは階層が深くなるにつれて解釈性が下がっている
- ・BMI \geq 0.2、Age \geq 60の場合、Diabetes(糖尿病)は単純で解釈しやすいように見えるが、前の2つのルール条件が偽で、BMI \geq 0.2、年齢 \geq 60の場合は、第3規則の正しい解釈が実際に行われる(左だと、呼吸器疾患と抑うつリスクが、No、であることが一目瞭然)

2. 取組内容② クレジットカードデフォルト予測

- ・「default of credit card clients Data Set (*)」のデータを使ってクレジットカードのデフォルト予測を行なった
(ここでのデフォルト=翌月払えなかった)

- ・構築したモデルに対して、先述の手法を適用し、“精度”と“解釈性”を評価するもの

30,000レコード
23属性項目(詳細は右記)

データ属性

分類	概要
ID	ID
LIMIT_BAL	利用可能上限額
SEX	性別
EDUCATION	学歴
MARRIAGE	結婚状況
AGE	年齢
PAY_0 ~ PAY_6	支払ステータス(月別)
BILL_AMT1 ~ BILL_AMT6	請求合計額(月別)
PAY_AMT1 ~ PAY_AMT6	支払合計額(月別)
default.payment.next.month	デフォルト(1:yes, 0:no)

2. 取組内容②(1) Born Again Tree

適用

クレジットカードのデフォルト予測において、ホールドアウト法により元データから訓練データ(24000件)テストデータ(6000件)を生成。

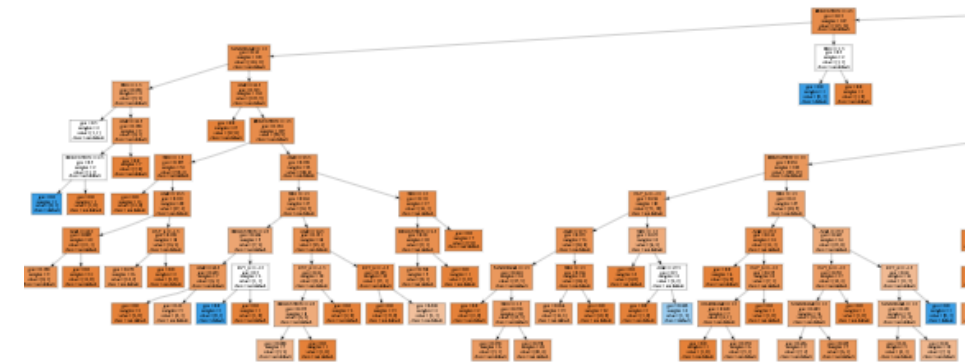
- モデルは訓練データから2モデル(アンサンブル学習と決定木)を作成。
- その後、Born Again Treeの手法を用いて、アンサンブル学習の結果から、疑似学習データ(100000件)を作成し、Born Again Tree(決定木)を生成する。
- 最後に各予測モデルにおいて、誤り率を計測

考察

- テストデータにおける、誤り率は、アンサンブル学習の方が低く予想通りの結果となった。
- Born Again Tree(決定木)の誤り率は、アンサンブル学習に近い値を出しており、予測モデルを近似しているといえる。これにより、実際の機械学習システムでも、精度面でBorn Again Treeを採用可能と言え、可読性の高いモデルが利用できる。

利用モデル	テストデータの誤り率
アンサンブル学習	0.1712
決定木	0.2103
Born Again Tree	0.1735

クレジットカード利用者におけるある月のデフォルト予測モデルのモデルと誤り率の比較



Born Again Treeモデルの可視化例(部分抜粋)

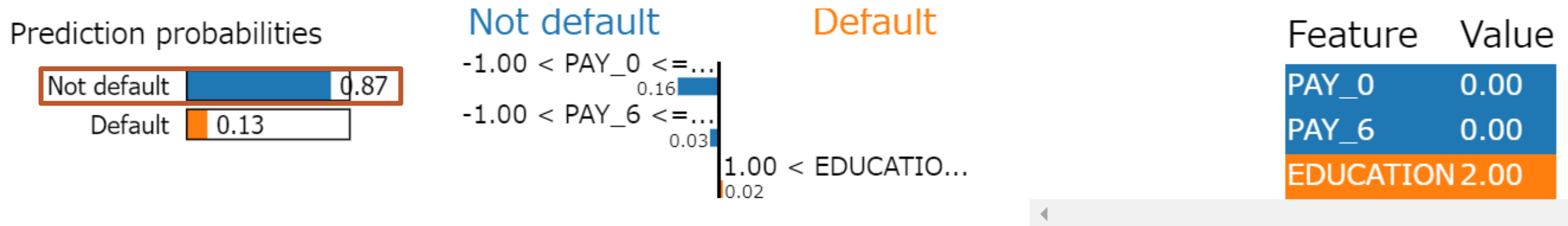
2. 取組内容②(2) LIME

- 適用

クレジットカードのデフォルト予測において、アンサンブル学習後、任意のレコード(対象: No11)に対してLIMEを適用。デフォルトしない確率が約87%、その理由としてPAY_0/PAY_6の値が0を根拠としていると可視化された。

- 考察

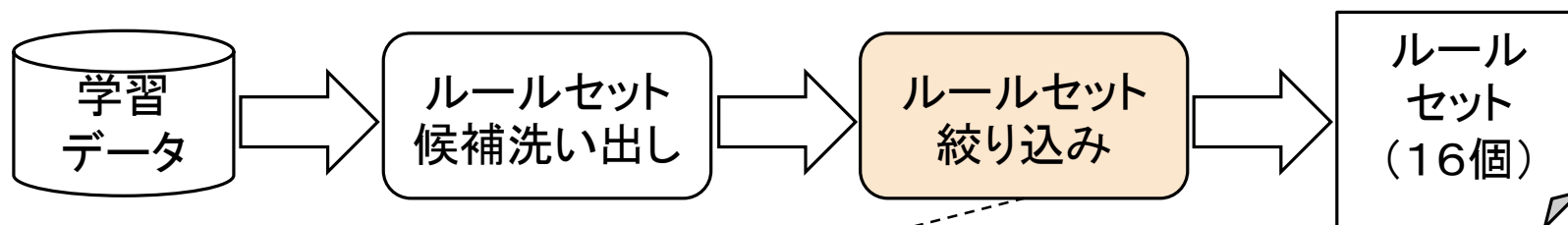
- 予測がどのパラメータに基づいているか、近似であっても説明があると、予測の説得力が高まる。



LIMEによる予測の可視化例

2. 取組内容②(3) interpretable decision sets

- ・組み合わせが膨大になるにつれ、ルール生成処理に時間がかかる
- ・説明変数 : 'SEX','EDUCATION','MARRIAGE','PAY_0','PAY_6'
- ・SEX{0,1}、EDUCATION{1~6}、MARRIAGE{1,2,3}、...



<評価関数:7つの関数を定義(解釈性・正確性の観点)>

- ・ルールセット自体のサイズを考慮
- ・1ルールの述語の数 (多すぎると理解が困難に)
- ・「Overlap」:ルールの重なり(重複がない方が良い)
- ・・・など

$$\arg \max_{\mathcal{R} \subseteq \mathcal{S} \times \mathcal{C}} \sum_{i=1}^7 \lambda_i f_i(\mathcal{R})$$

The Solution Set is:

{2, 3, 34, 36, 37, 8,
40, 42, 11, 13, 45, 15,
46, 56, 57, 31}

(参考)説明変数

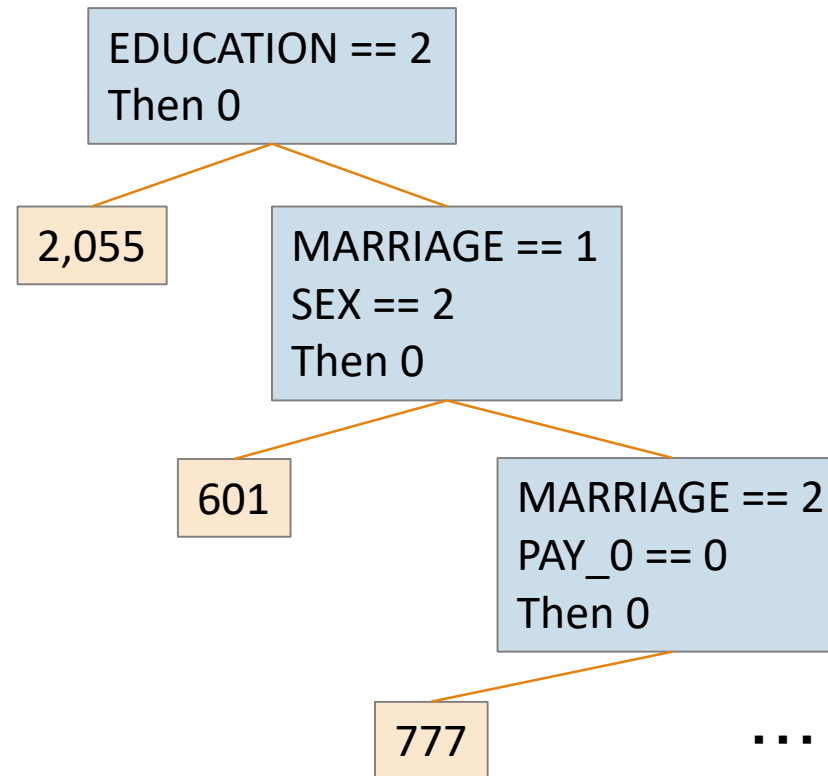
SEX	Gender (1=male, 2=female)
EDUCATION	(1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
MARRIAGE	Marital status (1=married, 2=single, 3=others)
PAY_0	Repayment status in September, 2005 (-2=no consumption, -1=pay duly, 0=the use of revolving credit, 1=payment delay for one month, 2=payment delay for two months, ... 8=payment delay for eight months, 9=payment delay for nine months and above)
PAY_6	Repayment status in April, 2005 (scale same as above)

2. 取組内容②(3) interpretable decision sets

- interpretable decision sets を適用、16のルールが生成
- 矛盾する3つのルールを排除したもので評価

<ルール>

If EDUCATION == 2, then 0
If EDUCATION == 2, then 1
If MARRIAGE == 1 and SEX == 2, then 0
If MARRIAGE == 2 and PAY_0 == 0, then 0
If MARRIAGE == 2 and PAY_0 == 0, then 1
If PAY_0 == 0, then 0
If SEX == 1 and MARRIAGE == 2, then 0
If SEX == 2 and MARRIAGE == 2, then 0
If PAY_6 == 0, then 1
If SEX == 1, then 1
If PAY_0 == 0 and PAY_6 == 0, then 1
If SEX == 2, then 1
If SEX == 2 and PAY_0 == 0, then 0
If SEX == 2 and PAY_0 == 0 and PAY_6 == 0, then 0
If SEX == 2 and PAY_0 == 0 and PAY_6 == 0, then 1
If MARRIAGE == 1 and PAY_0 == 0, then 1



<結果>

全レコード : 6000
defaultした人数 : 1266

カバーした数 : 5664
カバーできてない数 : 336
カバー率 : 0.944

2. ③ 適用結果を踏まえた評価、考察

○: 他手法対比優れる、△: 劣る

観点	(1) BATree	(2) LIME	(3) interpretable decision sets (IDS)
対応可能なデータ (画像/表形式)	△ 表形式のみ	○ 画像/表形式両方に対応	△ 表形式のみ
実装コスト	△ ライブラリがなく、一から実装	○ ライブラリが整備、実装容易	△ ライブラリがなく、一から実装
既存モデル への適用	○ 既存モデルに適用可能	○ 既存モデルに適用可能	△ 新たにモデル構築が必要
考察	木が深くなる傾向があり、読み解くのが手間	根拠をわかりやすく提示してくれる	ルール生成にかかる処理時間が長い

3. まとめ・所感

- ・機械学習の解釈性・説明性とは何か？

- ・解釈性: 内部状況が解釈できるか(論理的)
- ・説明性: 人が理解、納得できるか(心理的要素・実績)

- ・どのような技術があるのか？

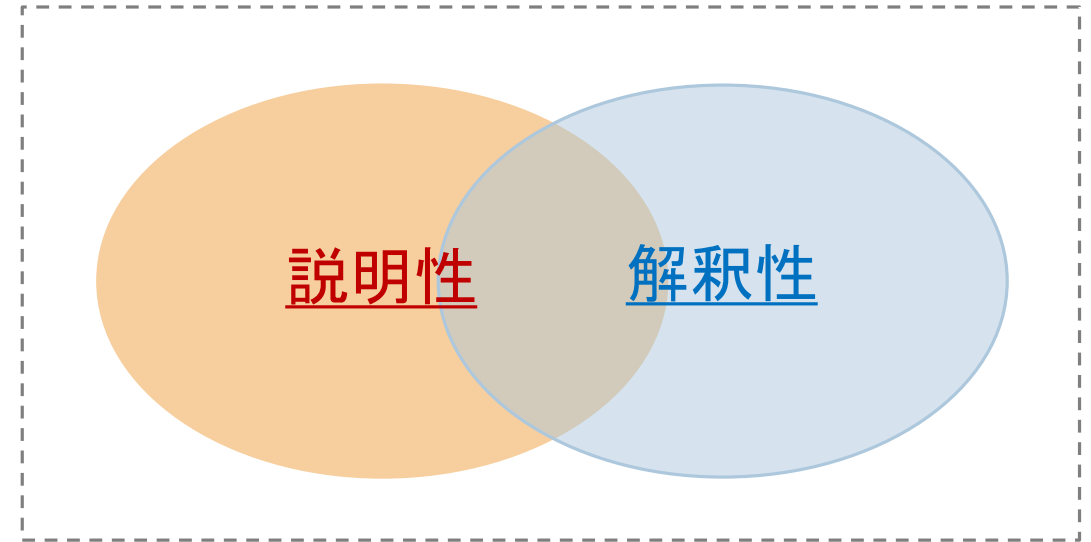
- ・大きく4分類、研究開発が進展

- ・機械学習の開発・運用・保守のどのプロセスで
使用できるのか？

- ・デバッグ、チューニングへ利用可能
- ・すでに運用しているモデルの解釈

- ・誰向けの説明として使えるのか？

- ・今回試行、考察した範囲ではエンジニア向け



【今後について】

- ・システム利用ユーザ、最終顧客に
対する解釈性・説明性
- ・研究開発の動向注視