

画像教師データ判別作業における改善手法の調査

千葉工業大学

情報科学部 情報ネットワーク学科

原研究室

1532096 高橋 宏如

背景

- オープンソースで提供されている機械学習ライブラリが進化
 - 機械学習モデルの構築が容易に
 - 活用例：画像判別
- 教師あり学習ではモデルの構築とは別に教師データが必要
 - 正解のデータが大量に必要
 - webからの自動収集

問題点と目的

問題点

- 自動収集した画像には必要ない画像が含まれている
- 不要な画像を取り除く判別作業が必要 ←**手動**



画像の判別順によって作業に変化があるのではないかな

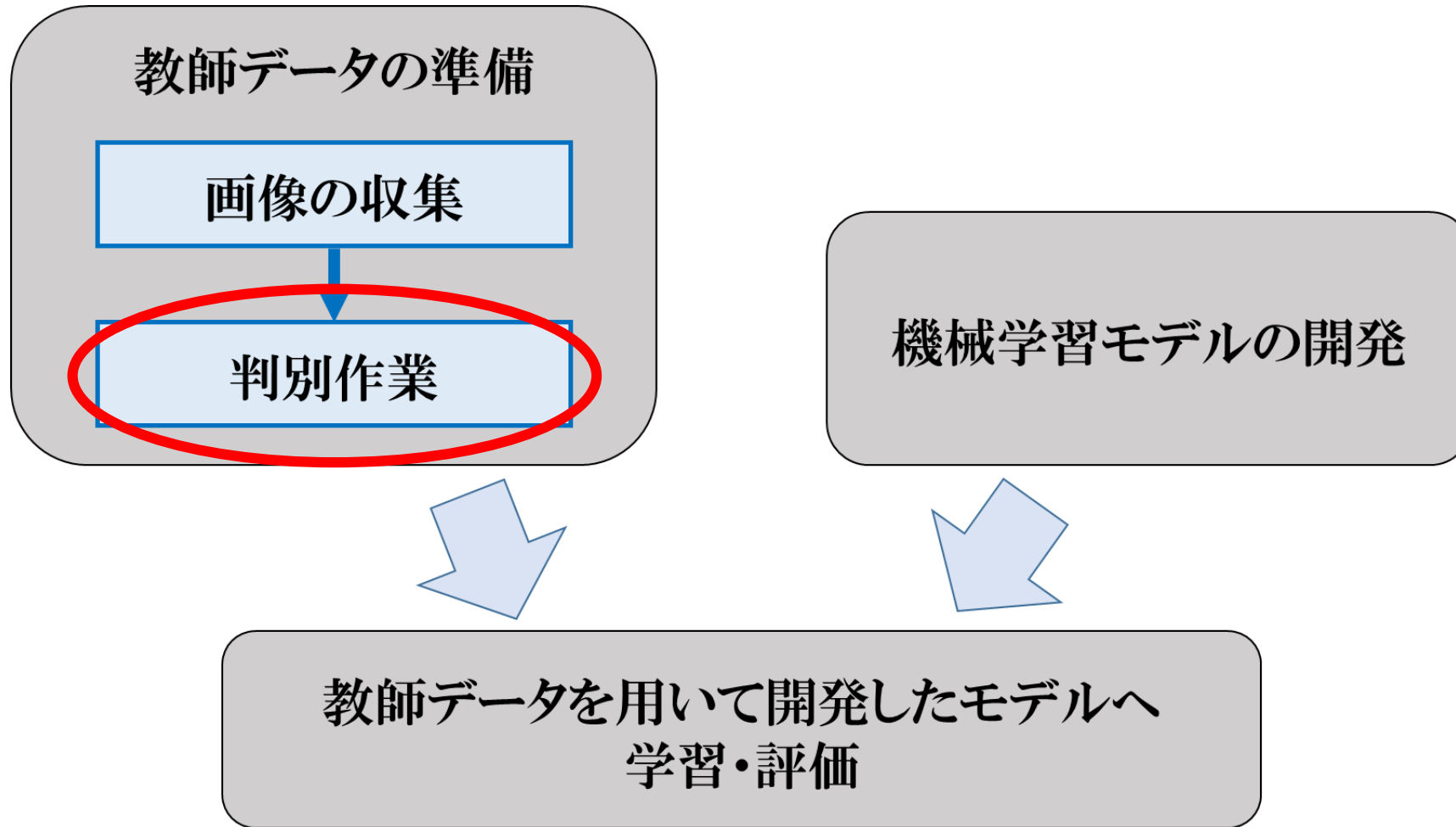
目的

- 画像の判別順によって起こる変化の調査
- 判別作業の改善



図1 判別作業の例

教師あり学習の手順



実験方法

1.画像の収集



2.画像の判別作業

・・・作業時間の調査



3.データセットの作成



4.モデルへの学習・テスト

・・・教師データの質を調査

実験方法

1.画像の収集

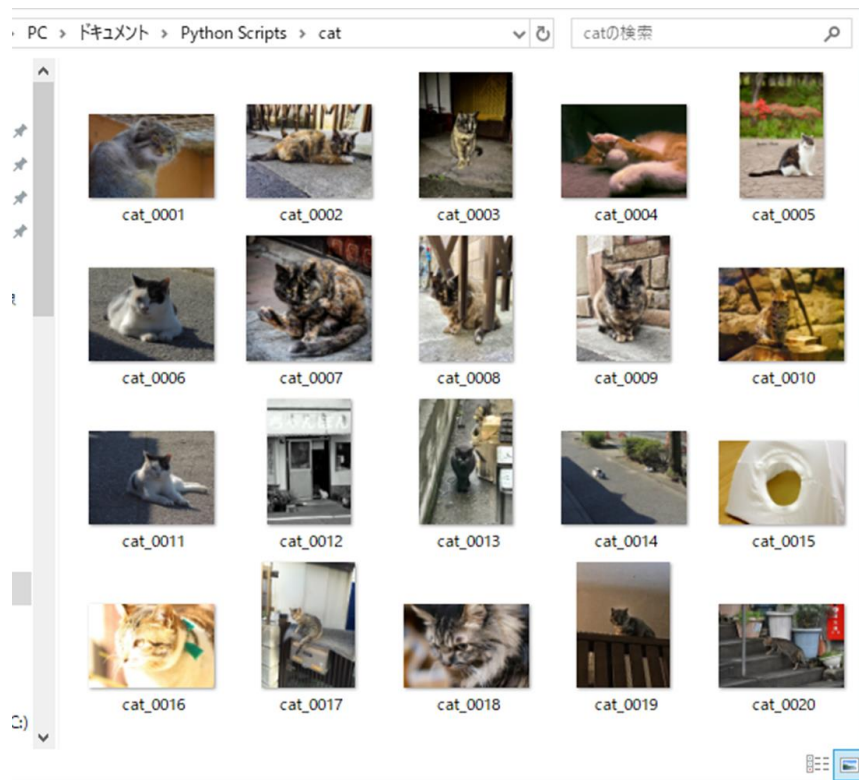


図2 画像の取得例

検索語
保存先フォルダ名
取得枚数を指定

動物 { イヌ
ペンギン
ウサギ

花 { コスモス
ハイビスカス
ヒマワリ

野菜・果物 { リンゴ
ナス
イチゴ

実験方法

2.画像の判別作業

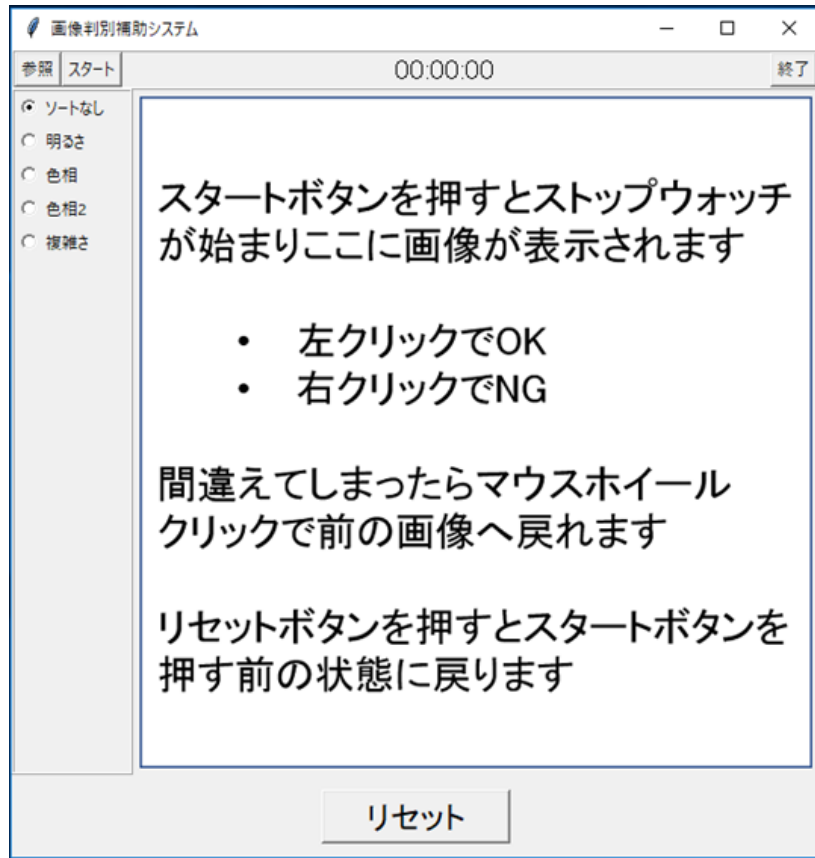
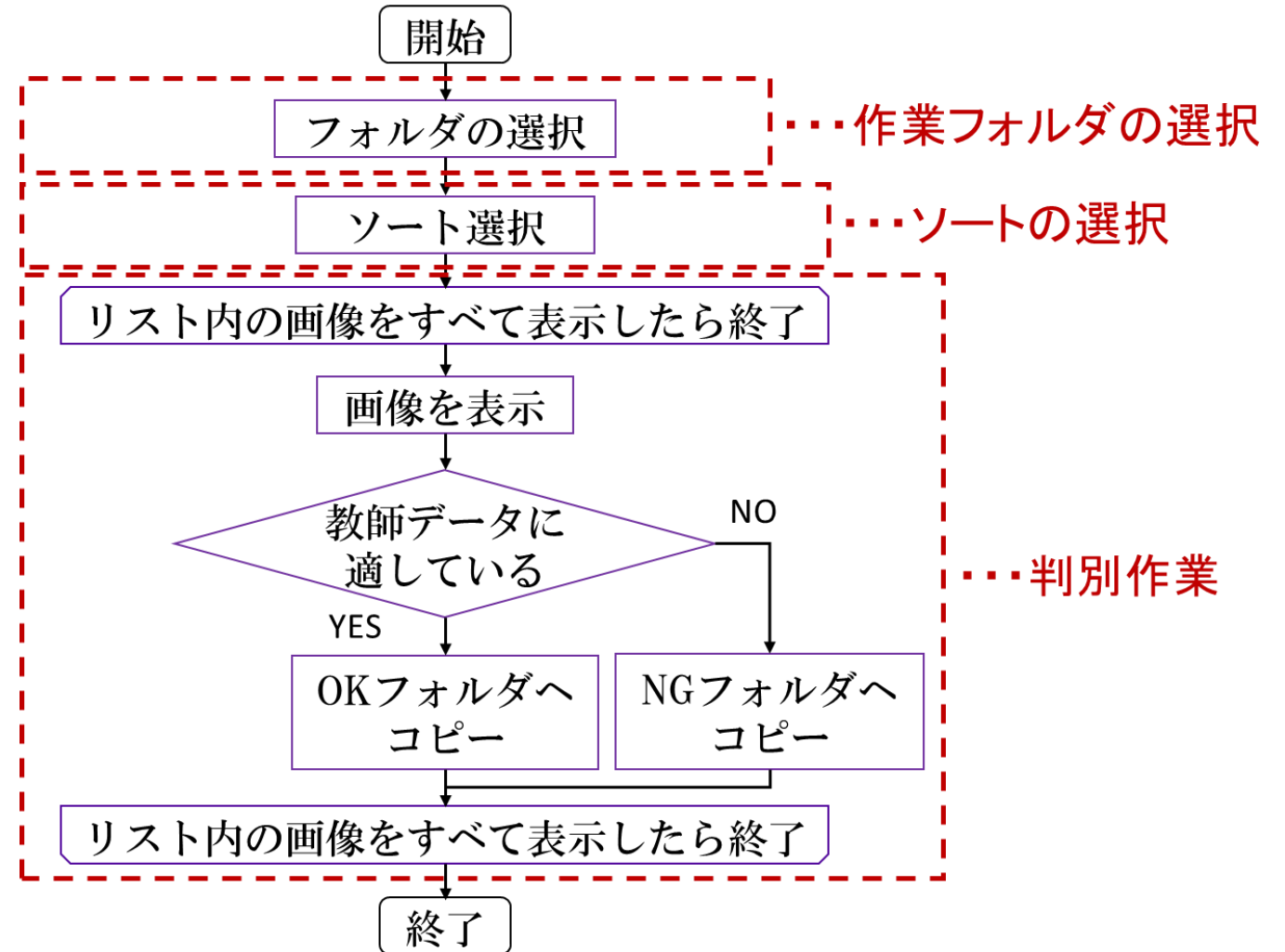


図3 実際の作業画面



実験方法

判別作業で用いるソート

－ 明るさソート

- ・ 画像全体の明るさの平均

－ 色相ソート

- ・ 画像中で最も使用されている色相

－ 色相ソート2

- ・ 色相ソート後画像を半分に分割し交互に表示

－ 複雑性ソート

- ・ 画像全体におけるエッジ(全方向エッジ)の割合

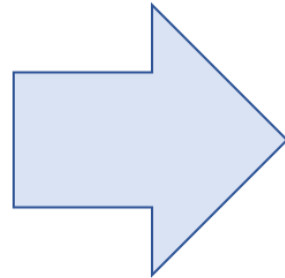
実験方法

3. データセットの作成 同カテゴリ・同ソート

イヌ_ok1

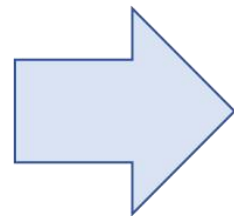
ペンギン_ok1

ウサギ_ok1



学習用データ

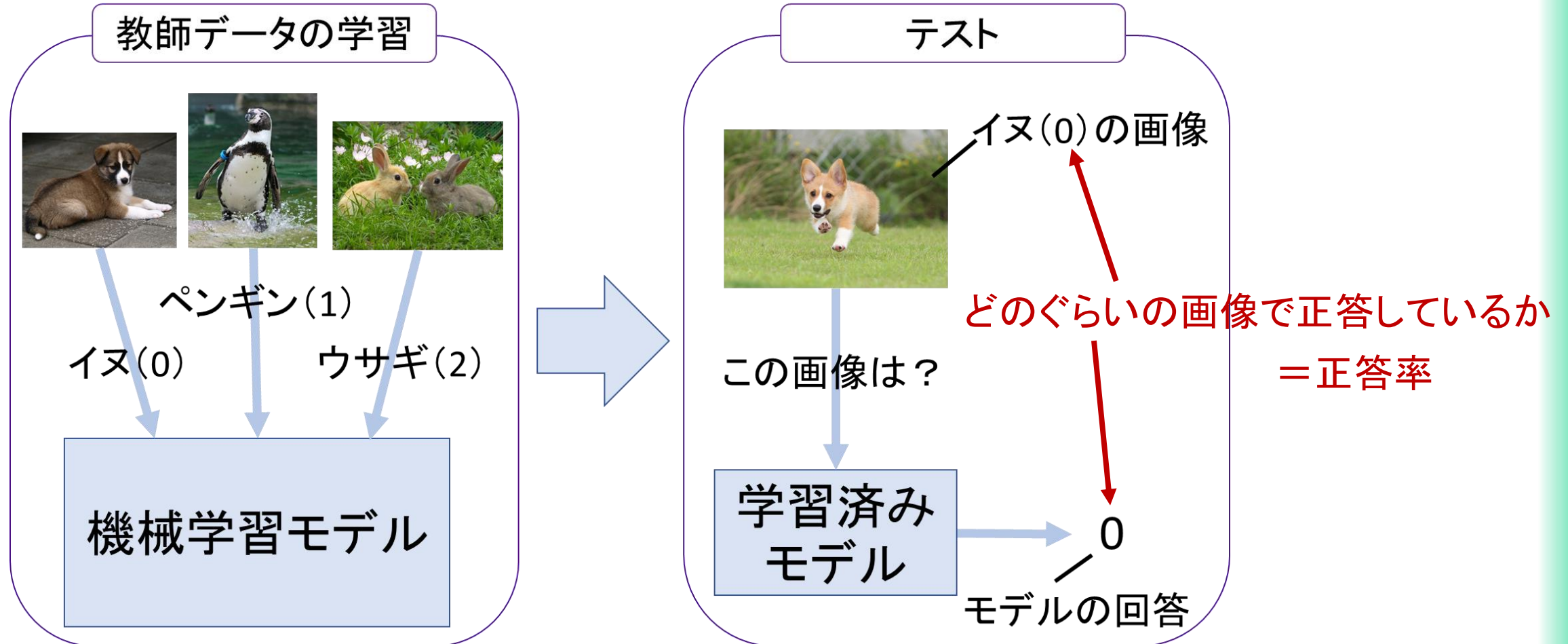
テスト用データ



[[画像のピクセルデータ], 対応するラベル]

実験方法

4.モデルへの学習・テスト



実験結果

表1 ソートごとの作業時間

画像\ソート	ソートなし	明るさ	色相	色相2	複雑さ
イヌ	7:04	6:54	6:52	7:02	6:49
ペンギン	7:10	6:42	6:45	7:14	6:51
ウサギ	7:05	6:51	6:31	6:59	7:04
コスモス	6:33	6:25	6:50	7:10	6:35
ハイビスカス	6:36	6:13	6:11	6:29	6:32
ヒマワリ	6:55	6:30	6:47	7:34	6:39
リンゴ	6:32	6:35	6:23	6:46	6:25
ナス	6:36	6:14	6:23	6:20	6:28
イチゴ	6:52	6:50	6:51	7:05	6:46

実験結果

表2 ソートごとの正答率(教師データの質)

画像\ソート	ソートなし	明るさ	色相	色相2	複雑さ
動物	48.4%	50.1%	48.0%	51.5%	46.6%
花	70.3%	71.1%	73.7%	72.7%	72.8%
果物・野菜	64.4%	65.4%	63.7%	61.6%	61.4%

考察

- 似ている画像を連続して判別する場合に作業時間が短くなる
- 教師データの質について明確な結果が得られなかった原因
→テストデータの一貫性がなかった

まとめ

目的

- 画像の判別順によって起こる変化の調査
- 判別作業の改善

結果

- 似ている画像を連続して判別する場合に作業時間が短縮
- 教師データの質においては特徴を掴めなかった

まとめ

課題

- 原因を改善し教師データの質がどのように変化するか再度調べる
- 作業時間の短縮と教師データの質を両立させた判別作業の手法を見つける

