

# クラウドソーシングによる 食事画像認識モデルの自動構築

電気通信大学 情報工学科

大澤翔吾 柳井啓司

# 研究背景・目的

- 動機

- 食事画像認識モデルの自動構築
- 学習データの作成に人手が必要  
→ クラウドソーシングの利用

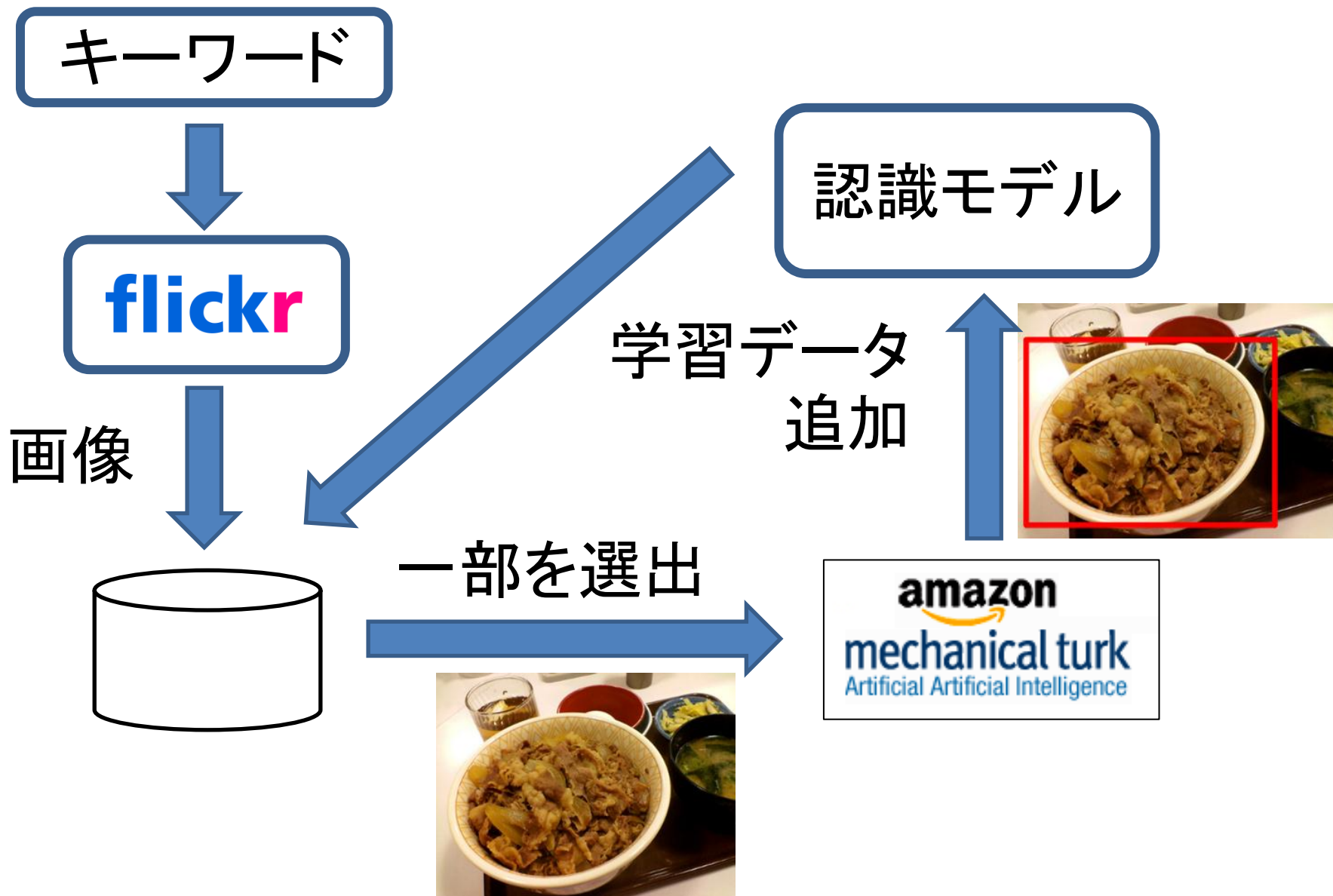
- 問題

- クラウドソーシング費用

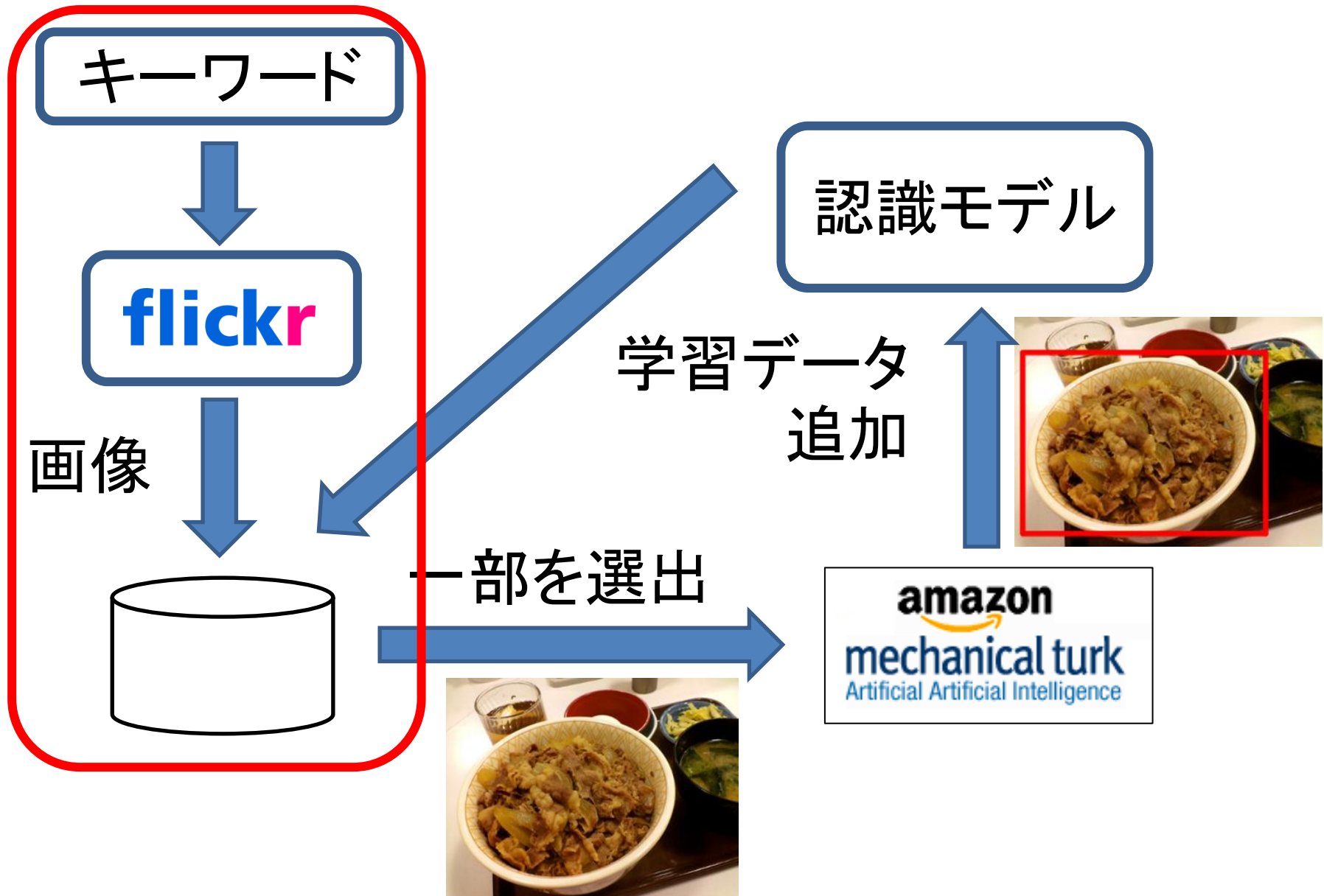
- 検討

- 費用を抑えて高精度なモデルを自動構築する手法

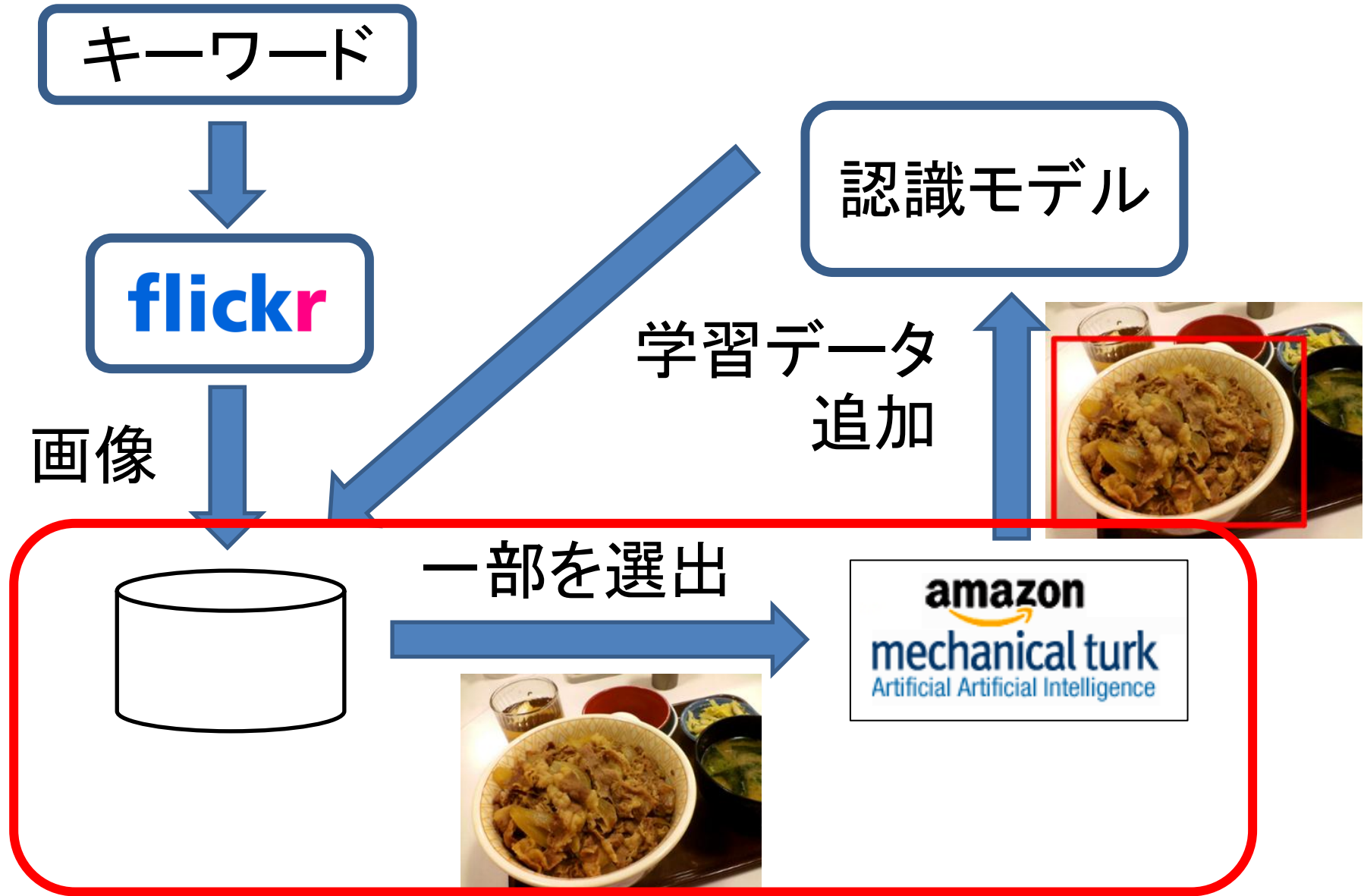
# 画像認識モデルの自動構築



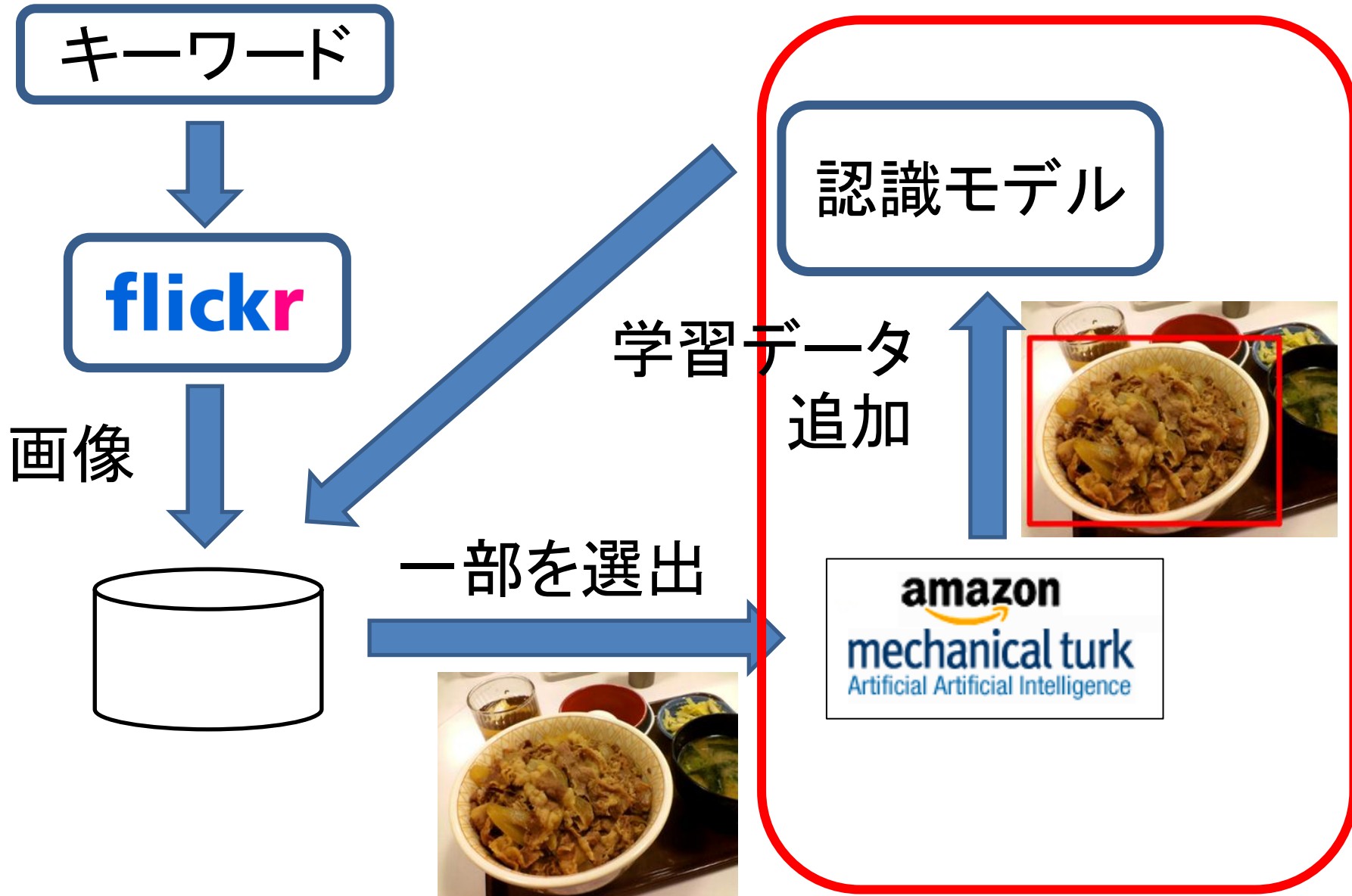
# 画像認識モデルの自動構築



# 画像認識モデルの自動構築



# 画像認識モデルの自動構築



# 画像認識モデルの自動構築



# 画像認識モデル

- Deformable Part Model\* を利用
- 線形SVMベース



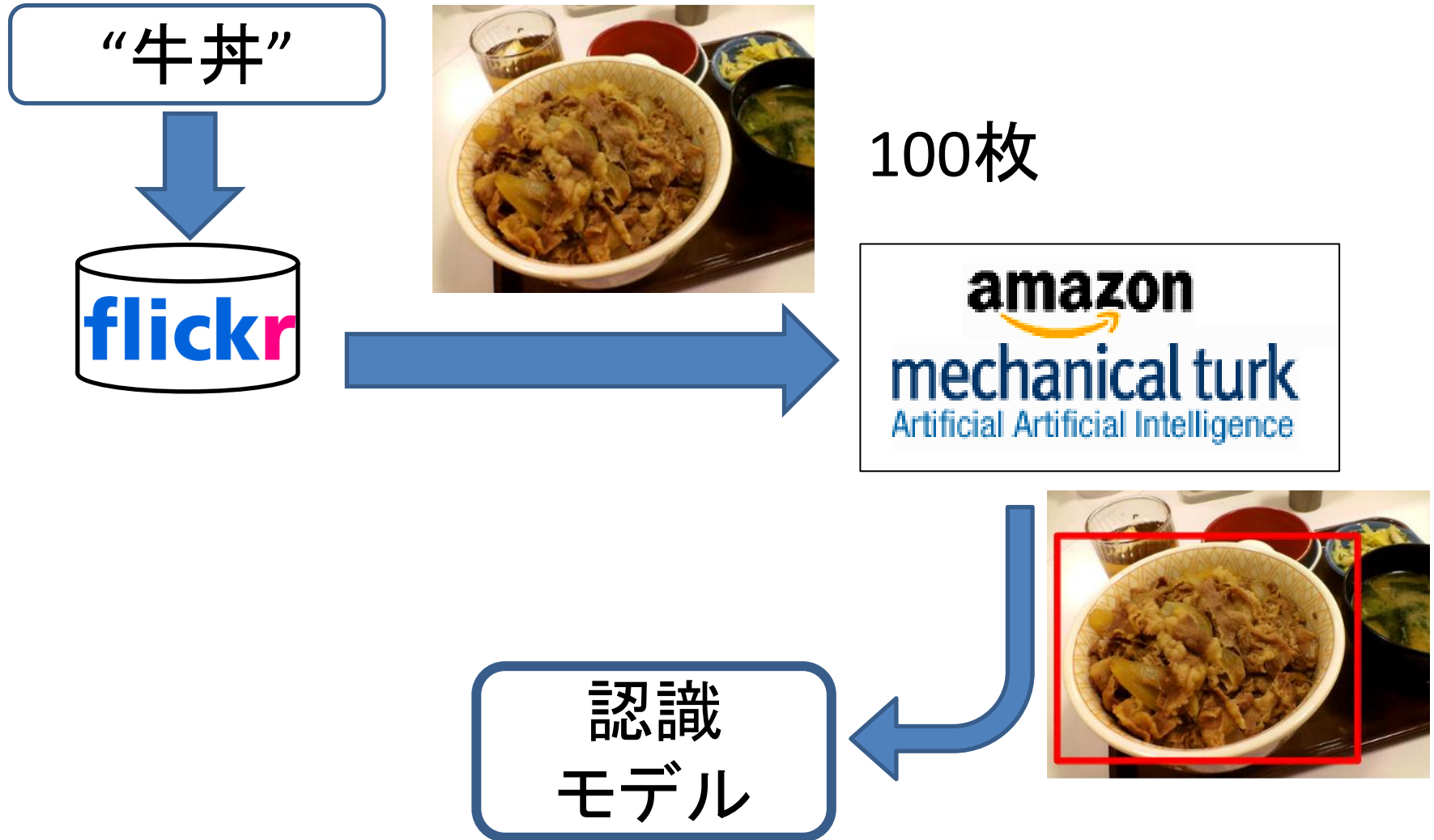
ポジティブ画像

ネガティブ画像

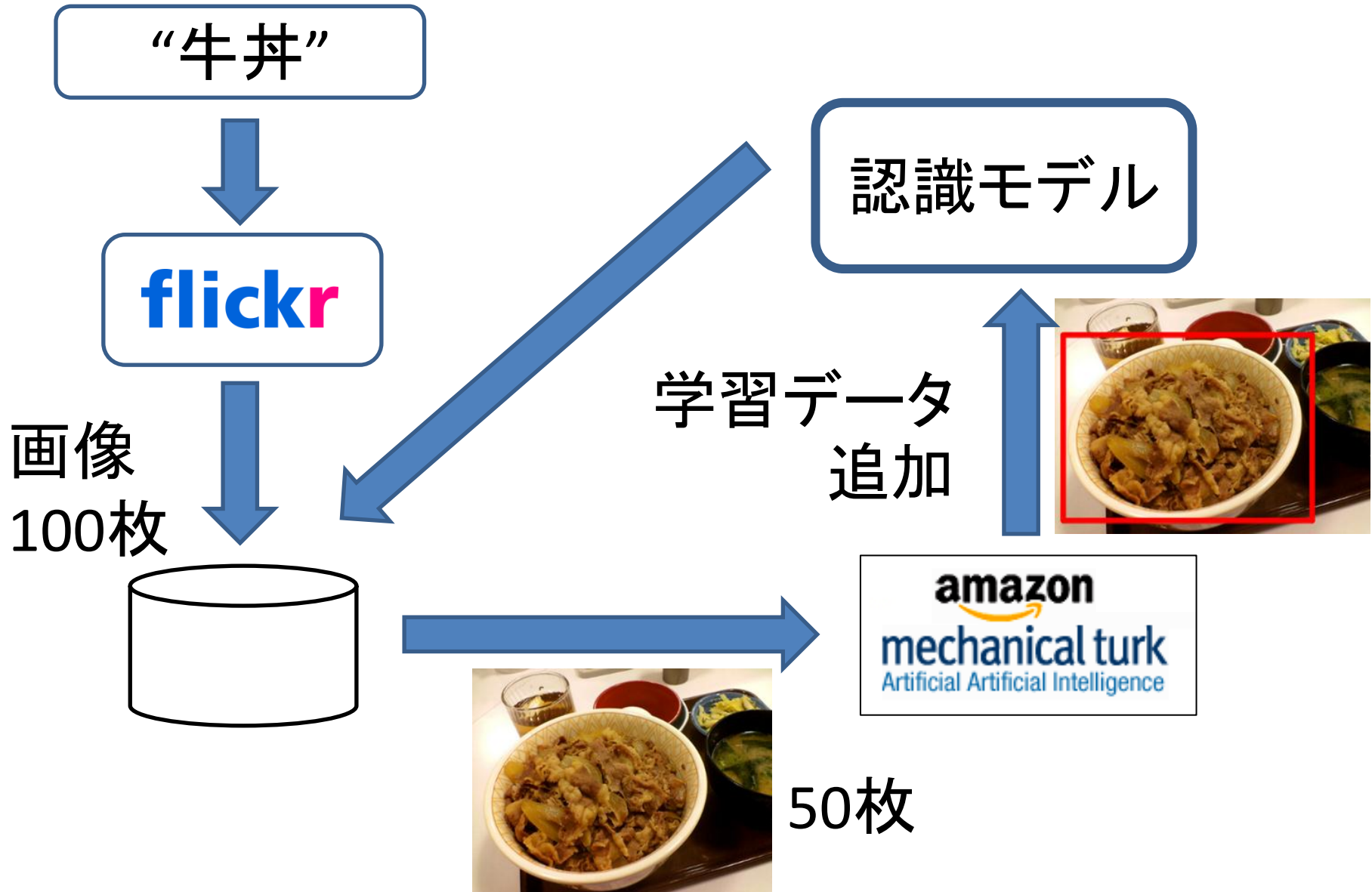
(\* ) P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, Object detection with discriminatively trained part-based models, PAMI 2010



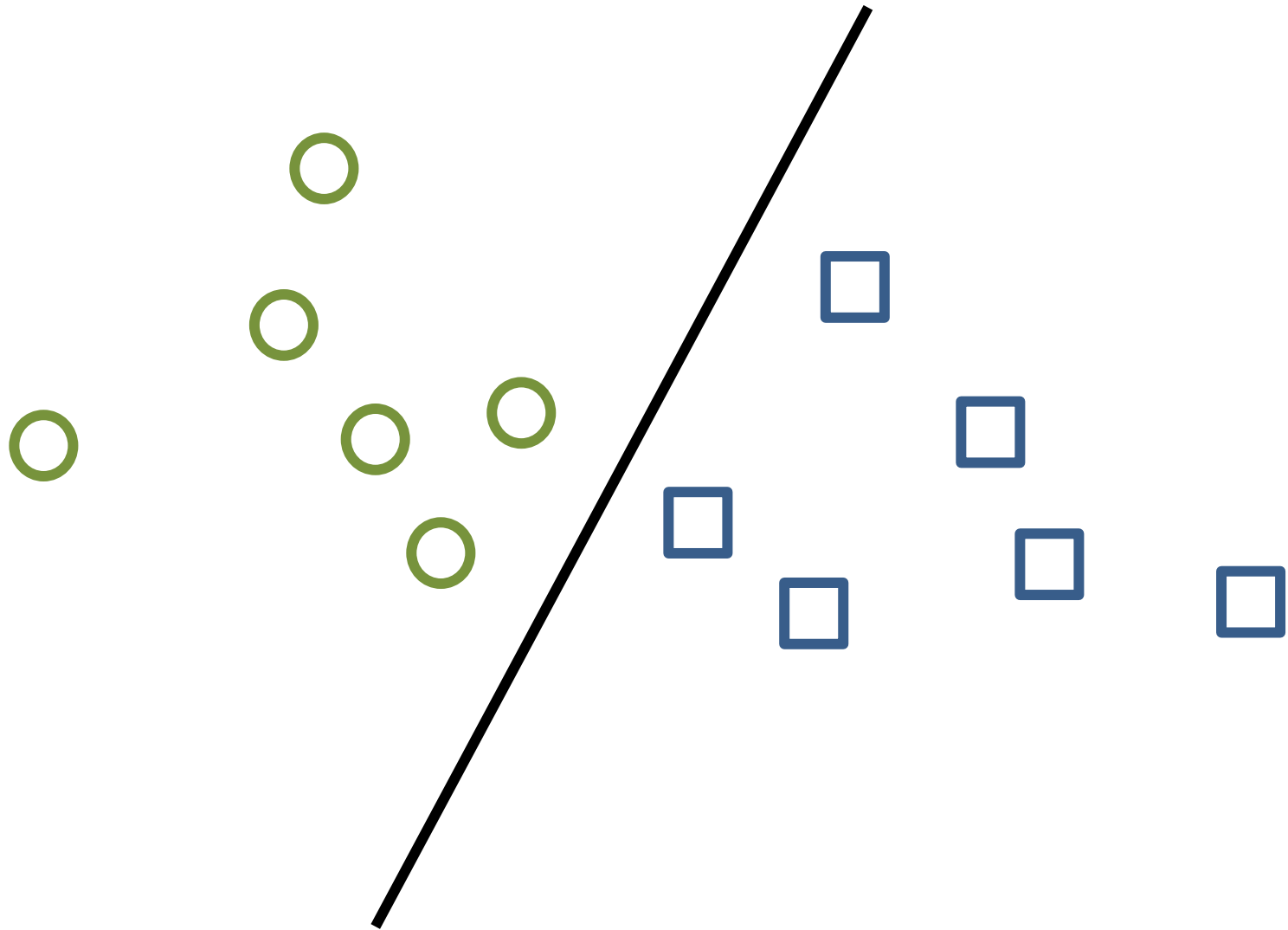
# 提案手法（認識モデル初期学習）



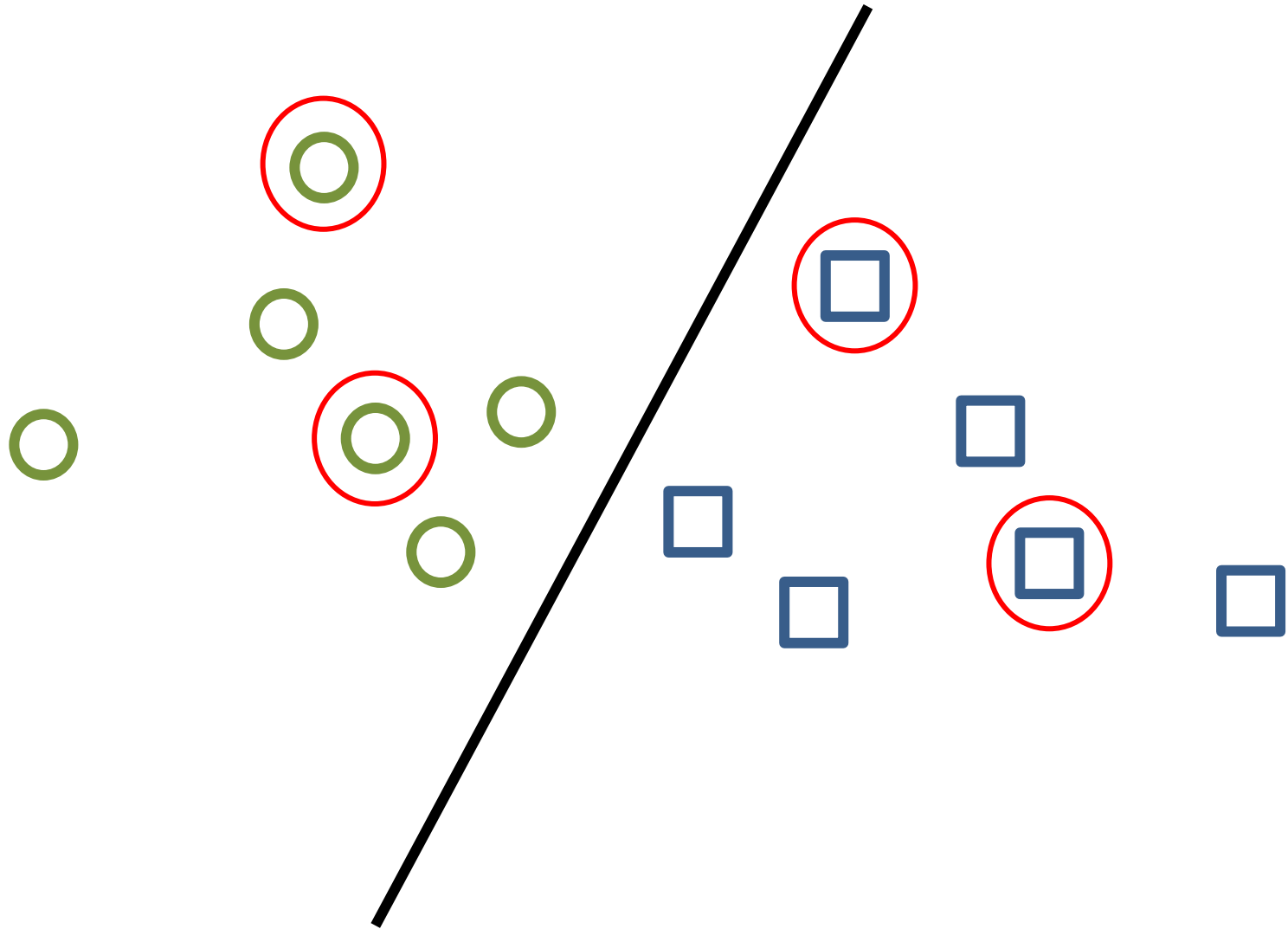
# 提案手法（ループ学習）



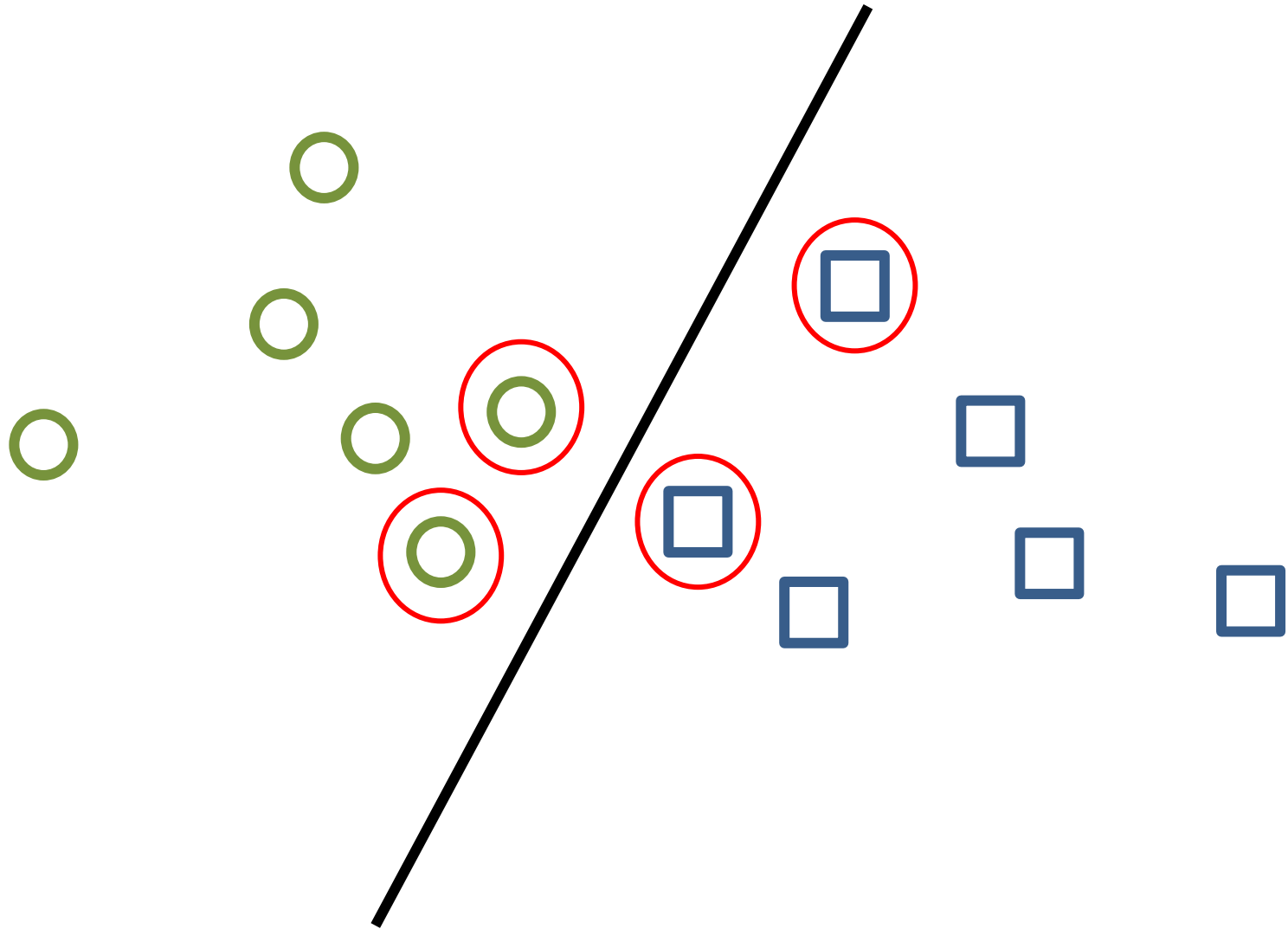
# 画像選出戦略 (Random)



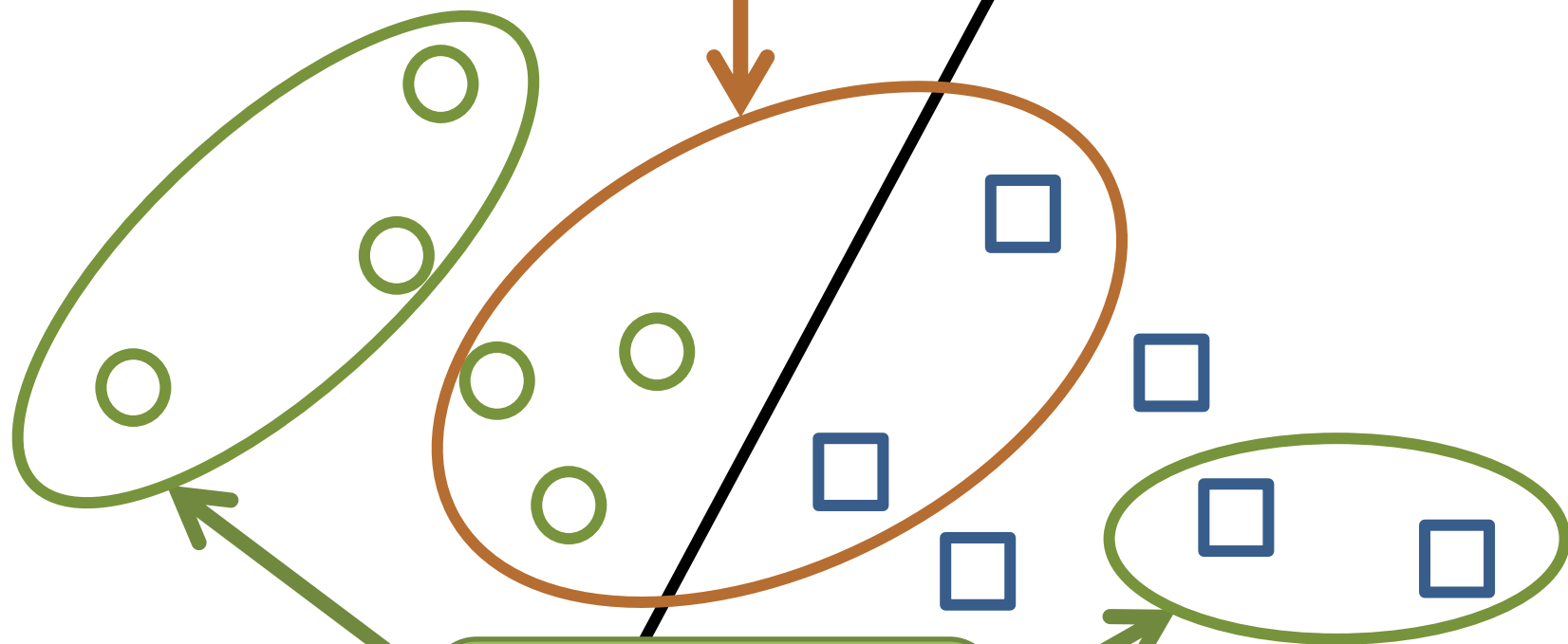
# 画像選出戦略 (Random)



# 画像選出戦略 (Near)

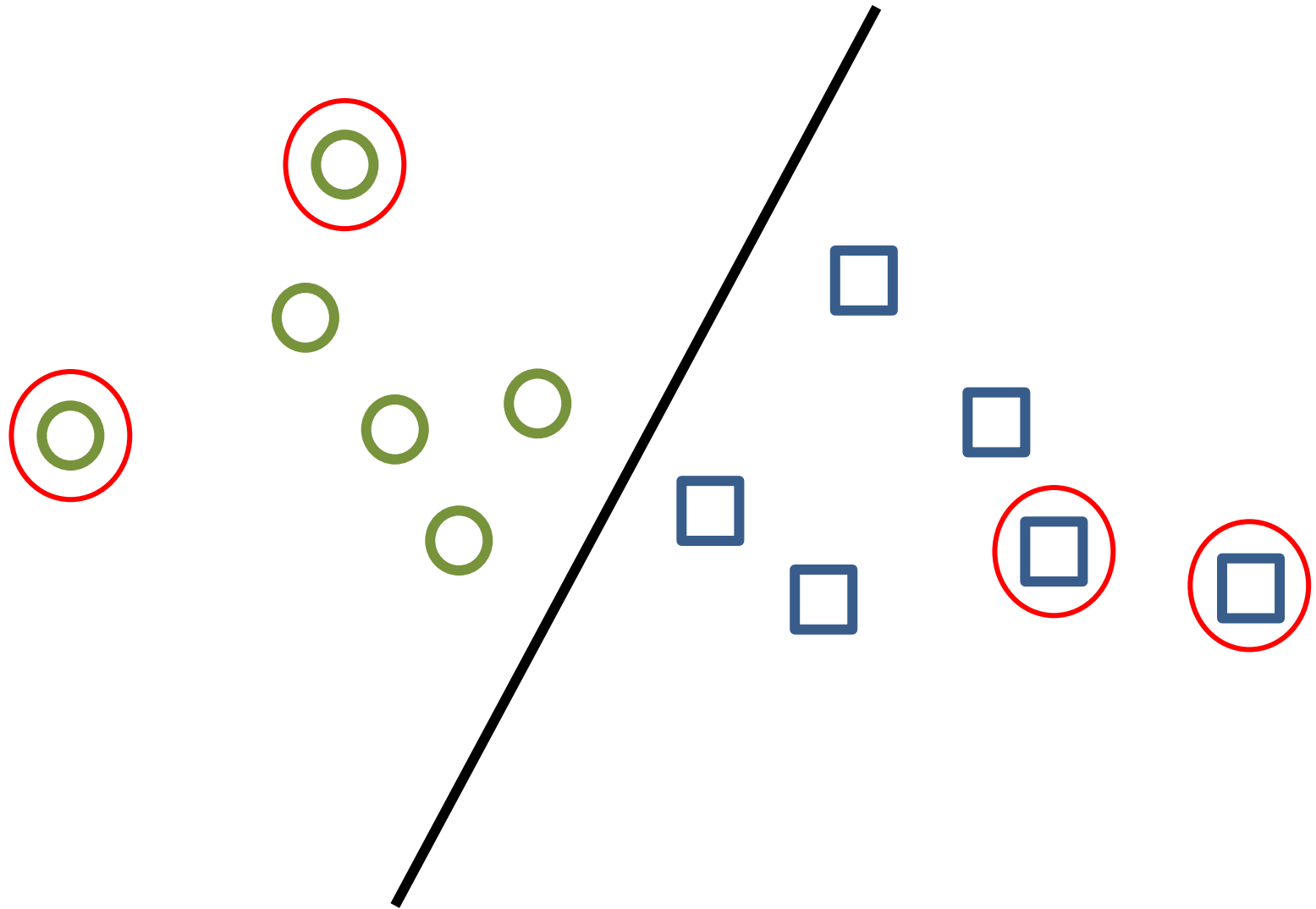


画像 (信頼できない Near)



信頼できる

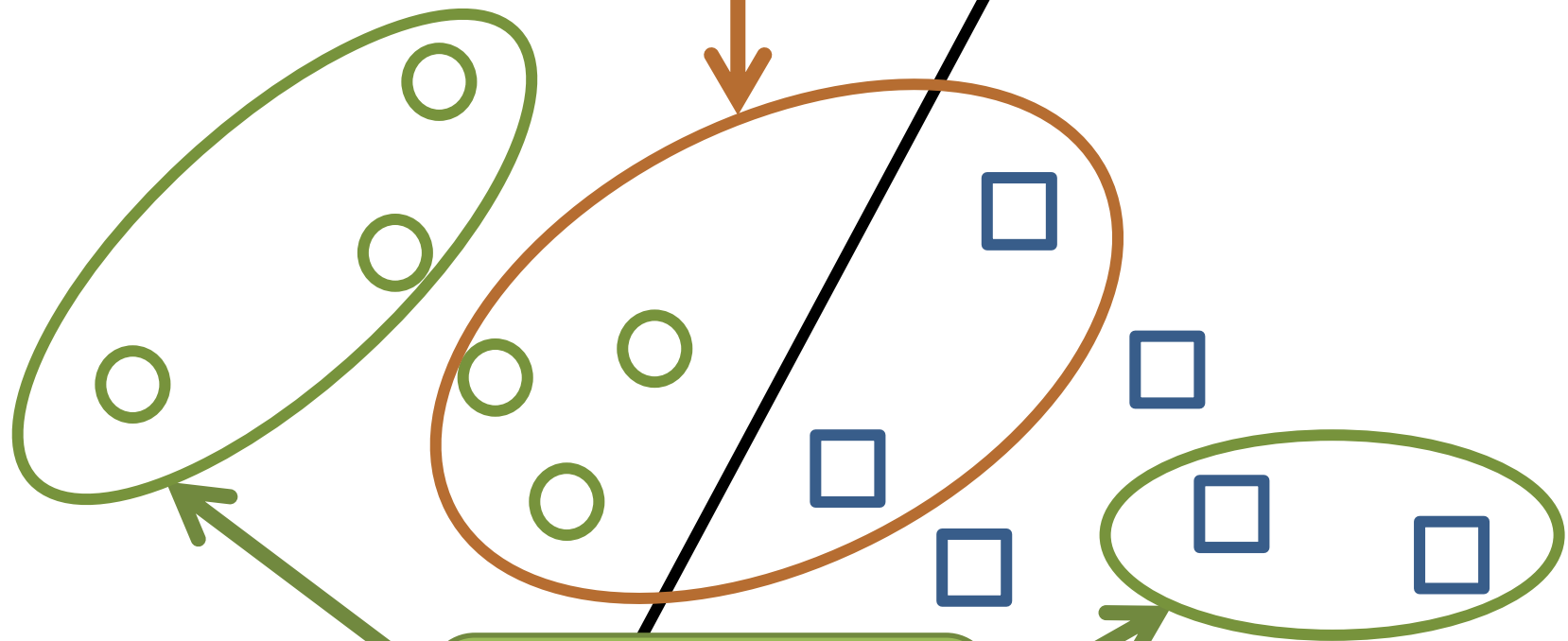
# 画像選出戦略 (Near)



画像

変化小

(Near)



変化大



# 実験

- 5カテゴリ(日本食)でループ学習

- 牛丼, 肉じゃが, お好み焼き, ラーメン, たい焼き



- Amazon Mechanical Turkを利用

- 各戦略で比較

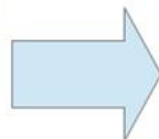
- モデルの認識精度 (F値)

- ポジティブ学習画像枚数

# 作業内容



牛丼が写っている



Bounding-Boxを描画



牛丼が写っていない



「牛丼なし」として  
チェック

# 作業ページの例


amazonmechanical turk Artificial Intelligence Your Account HITs Qualifications 354,788 HITs available now MultimediaMiningLab

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing that pay at least \$ 0.00 for which you are qualified require Master Qualification

Timer: 00:00:00 of 10 minutes Want to work on this HIT? **Accept HIT** Total Earned: \$2.59 Total HITs Submitted: 76

Draw bounding box around a Japanese food in images  
Requester: MultimediaMiningLab Reward: \$0.05 per HIT HITs Available: 1 Duration: 10 minutes  
Qualifications Required: HIT approval rate (%) is greater than 95



The image shows a screenshot of an Amazon Mechanical Turk HIT page. The task is to draw a bounding box around a Japanese food item in a set of six images. The top-left image, showing two bowls of food, is marked with a large red 'X' and is not selected. The other five images, showing various bowls of Japanese food like beef bowls, chicken bowls, and rice bowls, have red bounding boxes drawn around them, indicating they are the correct answers for the HIT.

# クラウドソーシングの利用

	1作業	1カテゴリ	総数
作業数		50	250
画像枚数	10	500	2500
報酬	\$0.05	\$2.5	\$12.5

$\$12.5 \times 5人 = \$62.5$  (全部依頼)

$\$62.5 \times 0.6 = \$37.5$  (提案手法)

# 認識モデルの評価方法

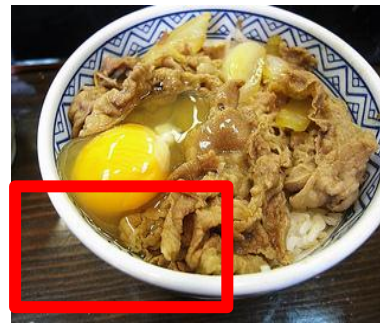
- テストデータを用意
  - 各カテゴリ50枚, 計250枚
  - 正解データを付与 (boxの位置 or 無し)
- 正解データと50%オーバーラップ



正解



○



×



×

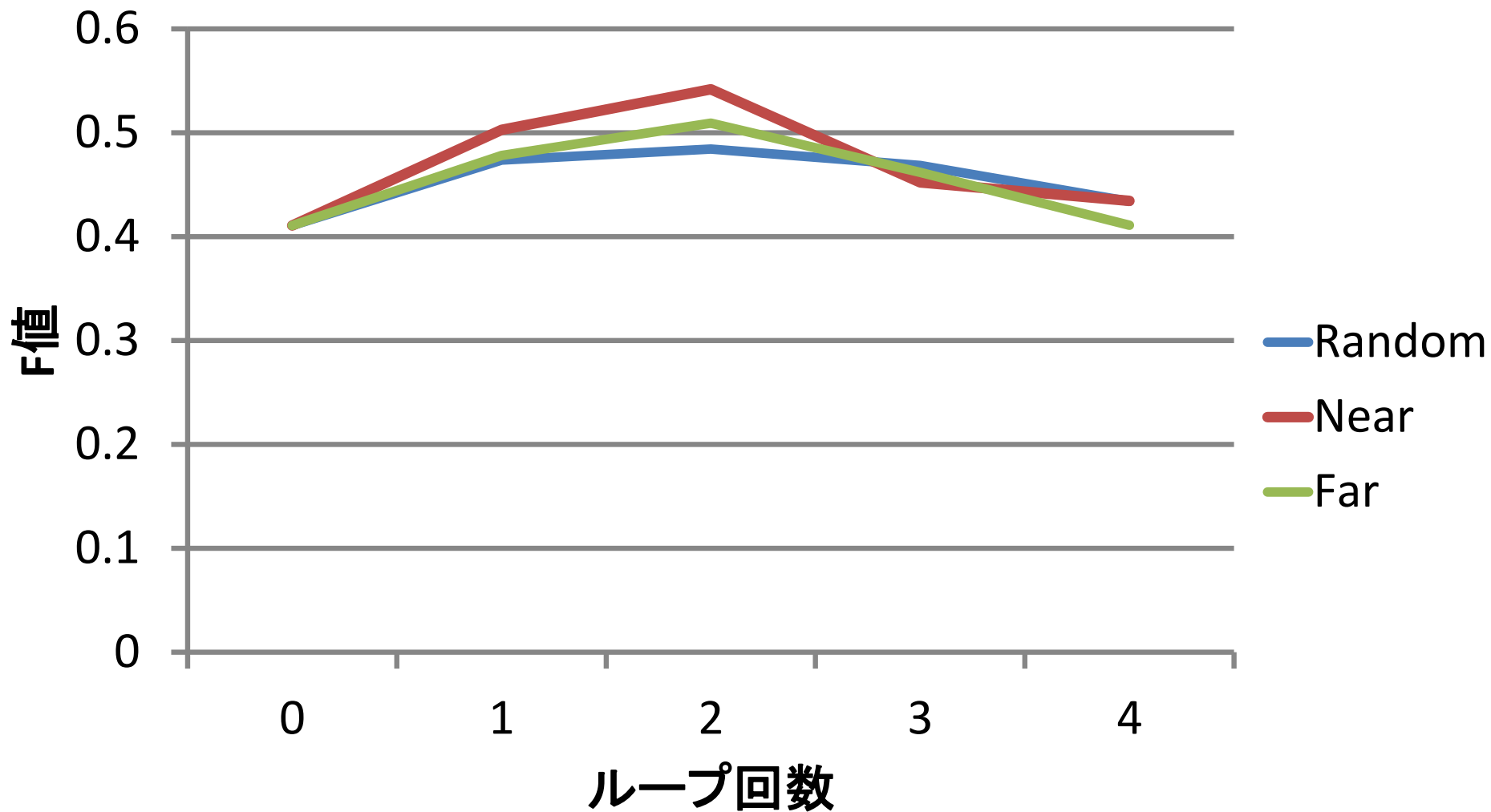
# 認識モデルの評価方法

適合率 = 認識されたうち正解した数

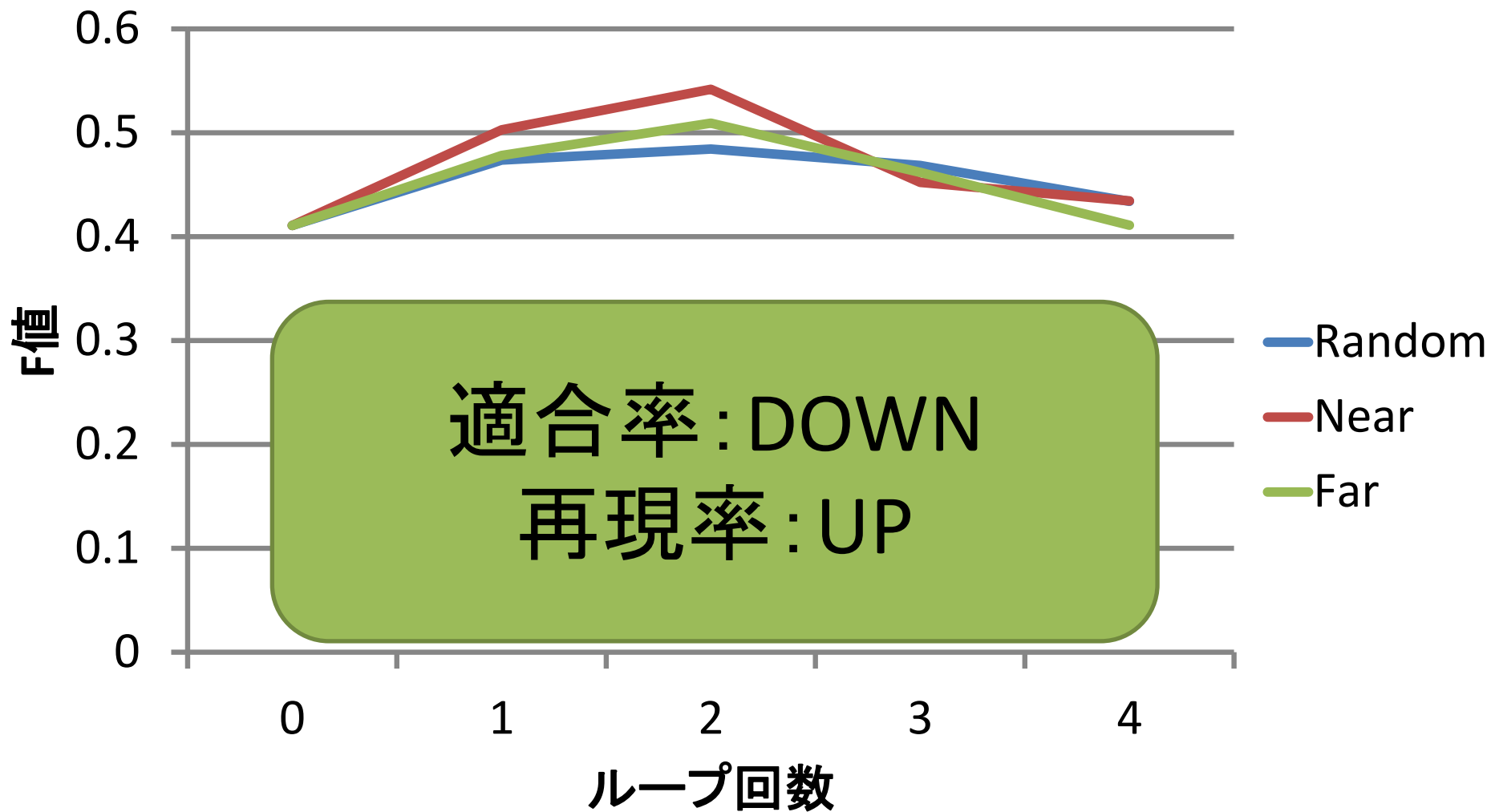
再現率 = 全正解画像中認識した数

$$F\text{値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

# モデルの認識精度 (F値)

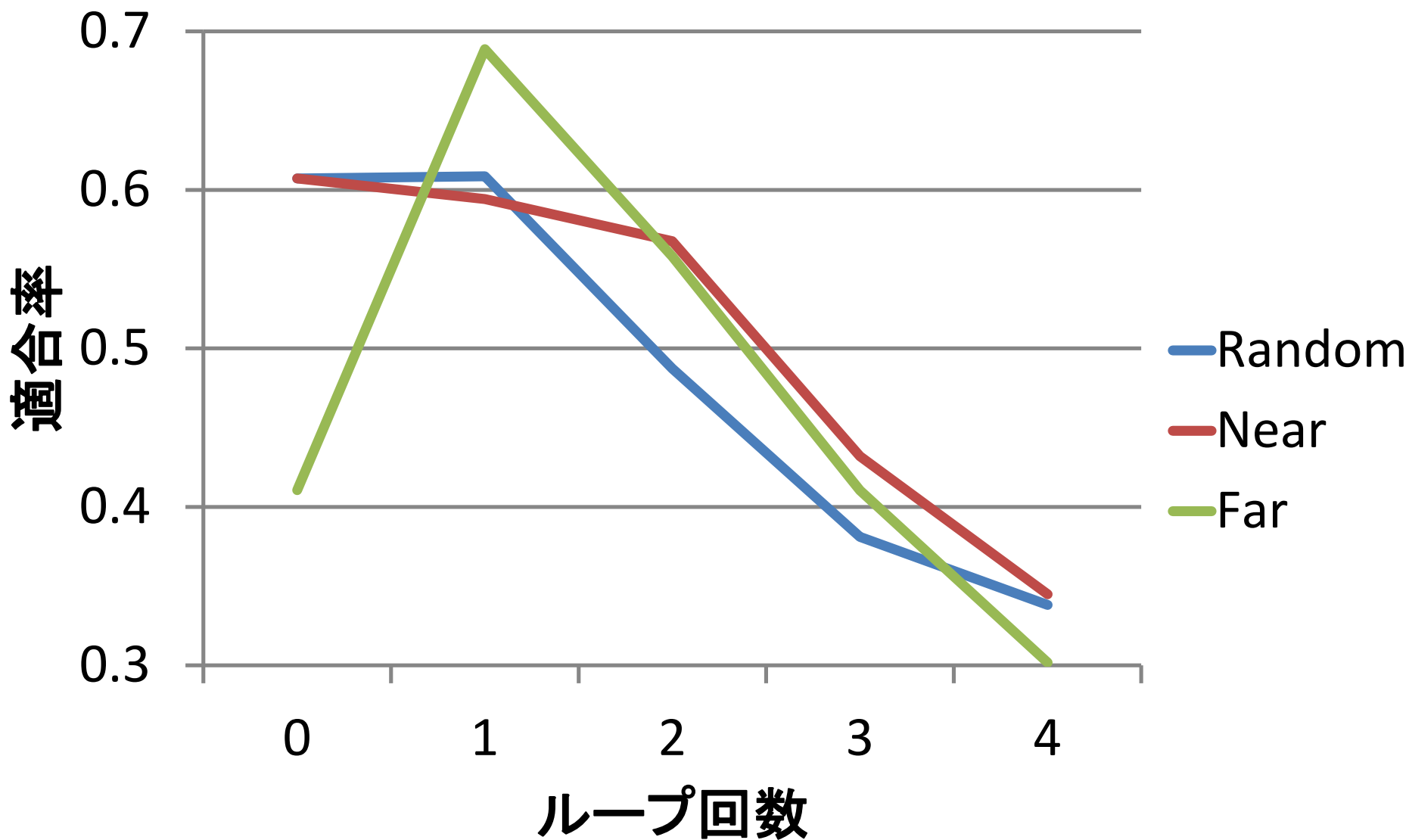


# モデルの認識精度 (F値)

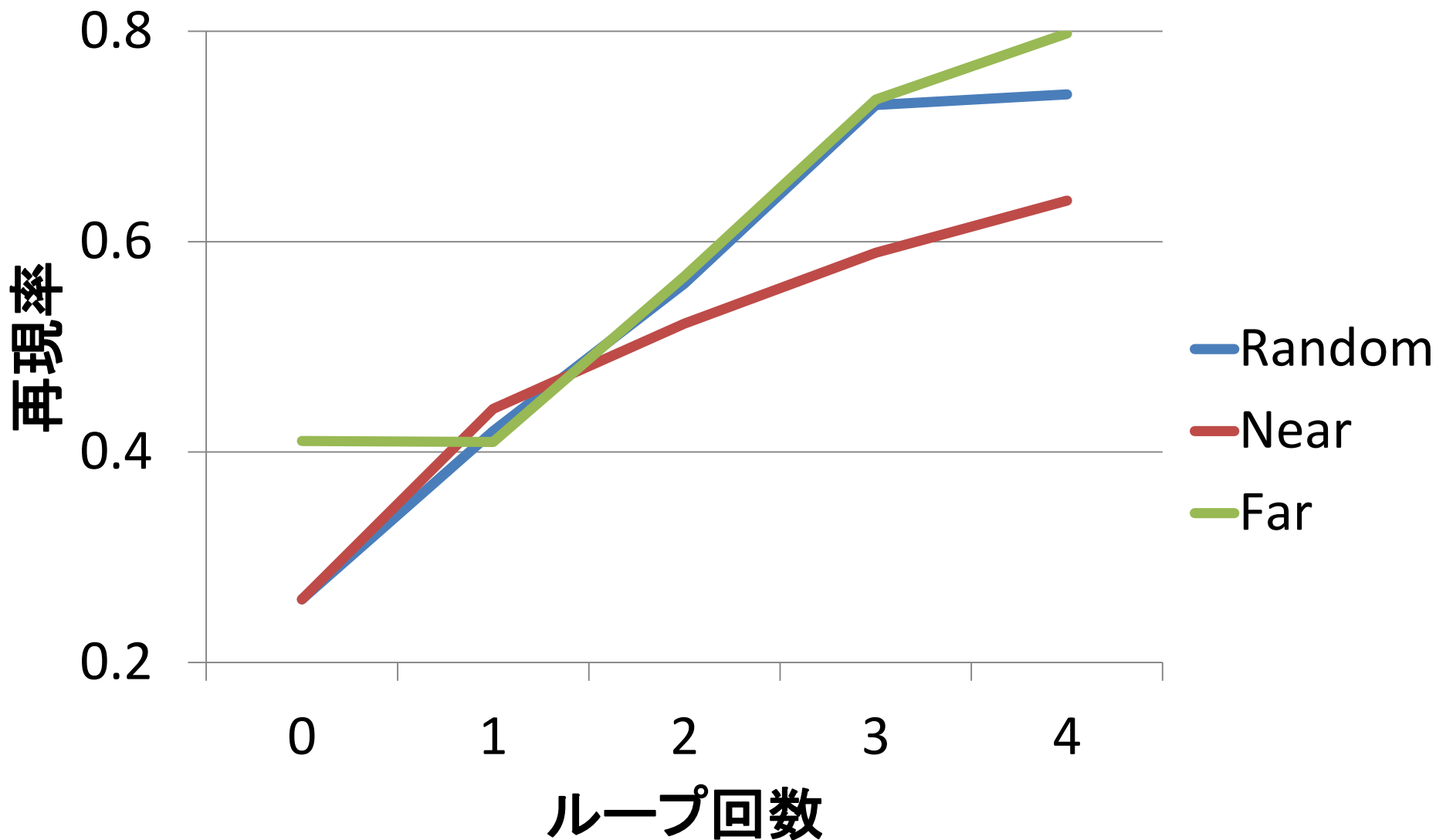




# モデルの認識精度（適合率）



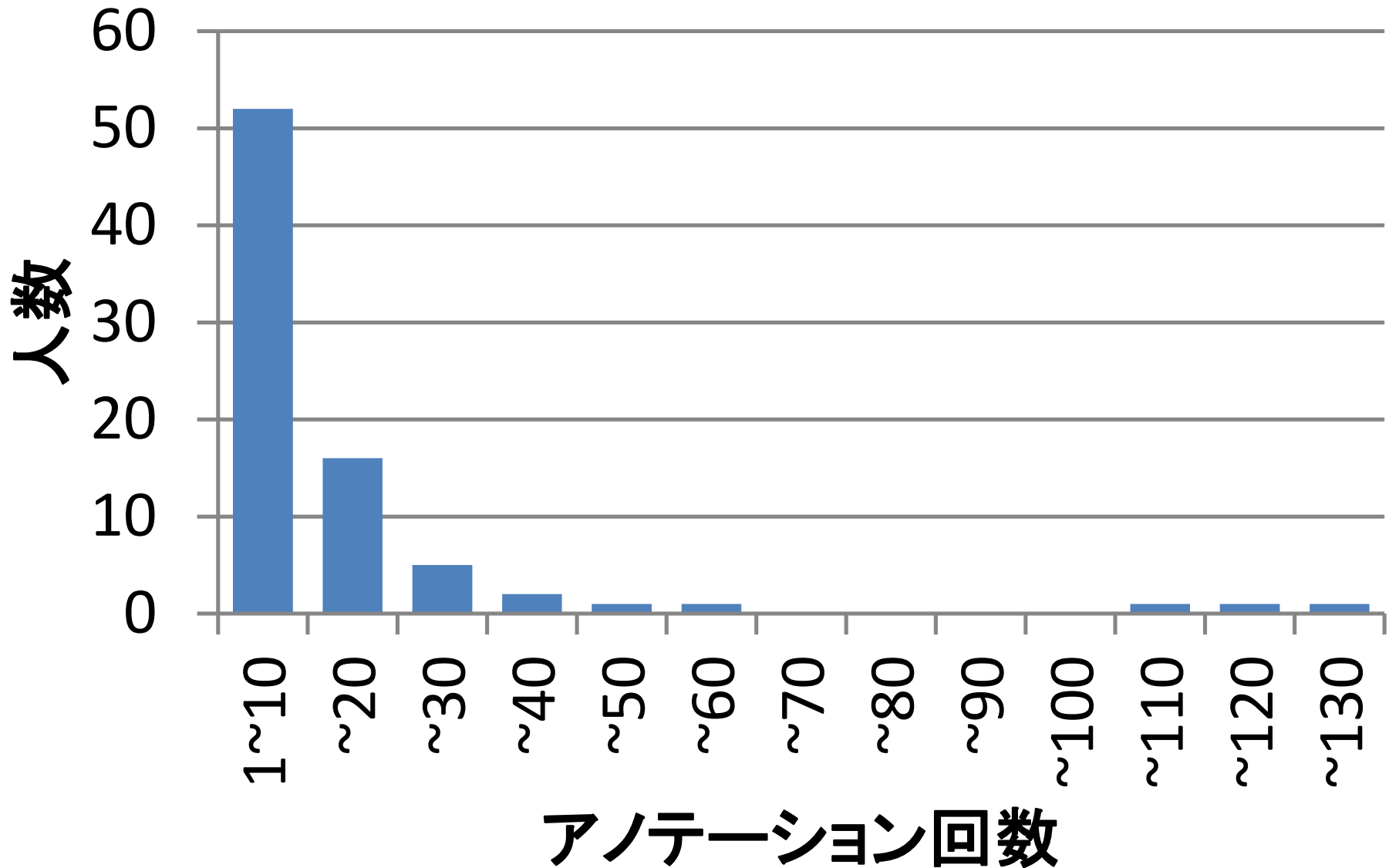
# モデルの認識精度（再現率）



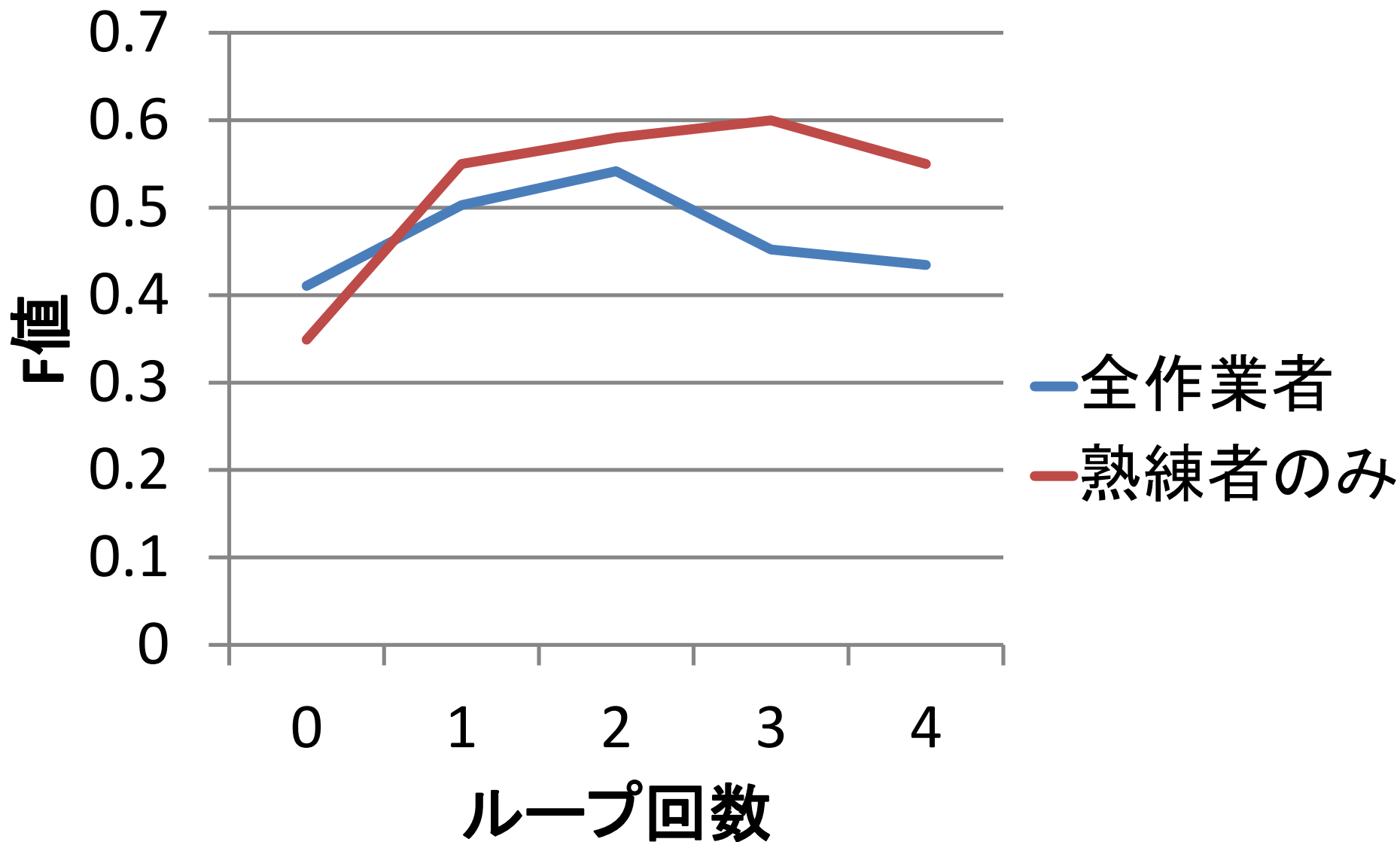
# ノイジーなアノテーション



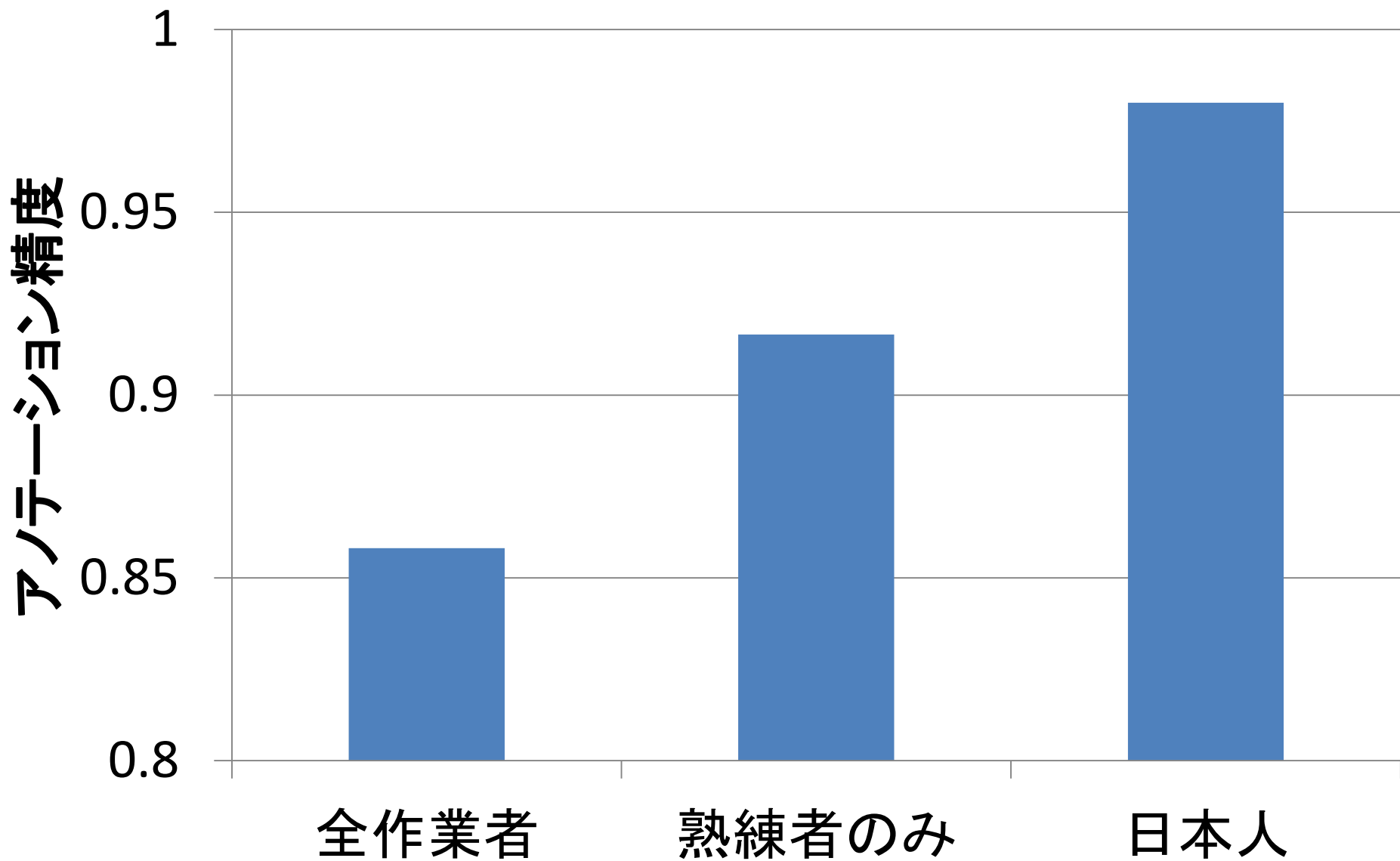
# アノテーション回数の分布



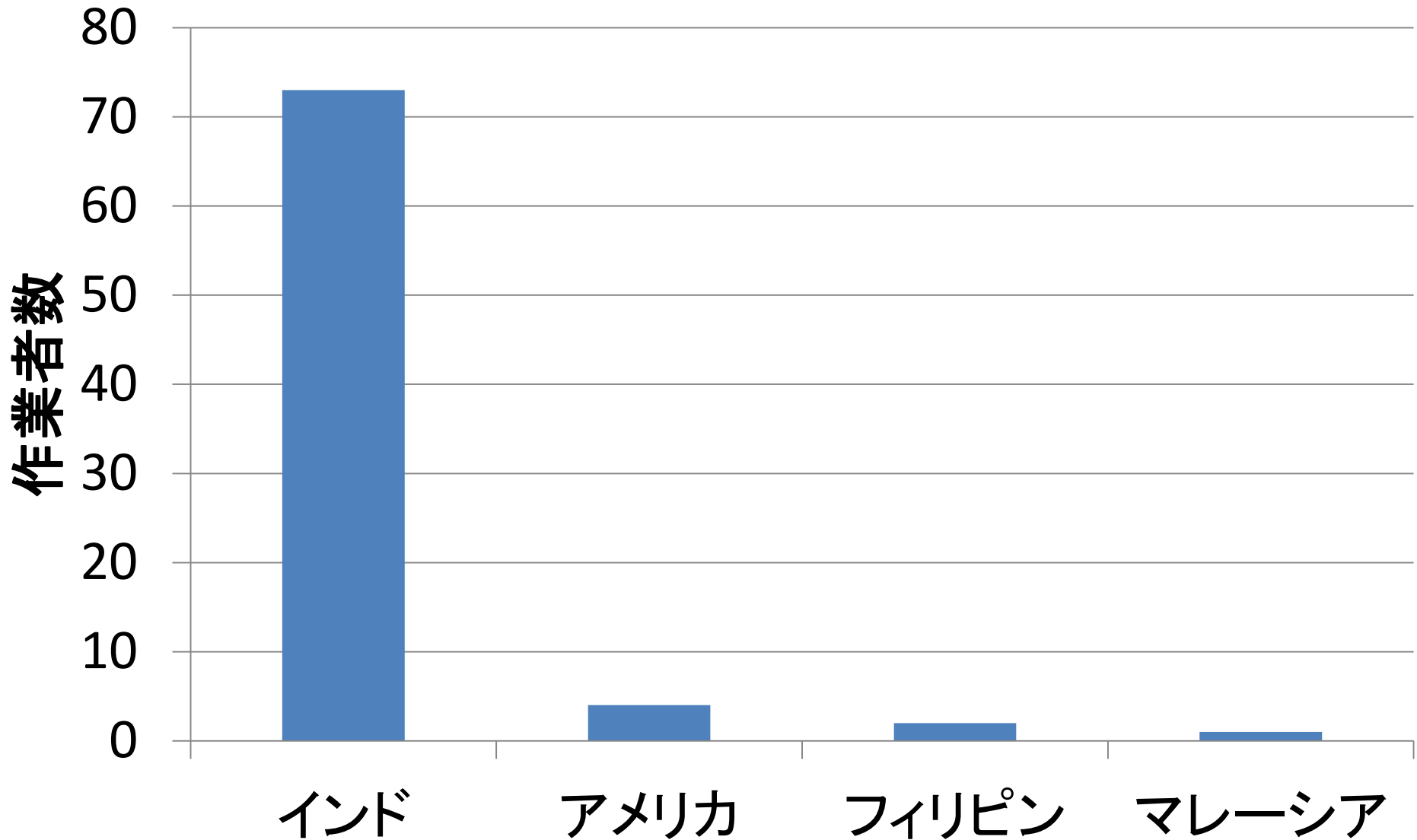
# 熟練者(10回以上)のみで学習



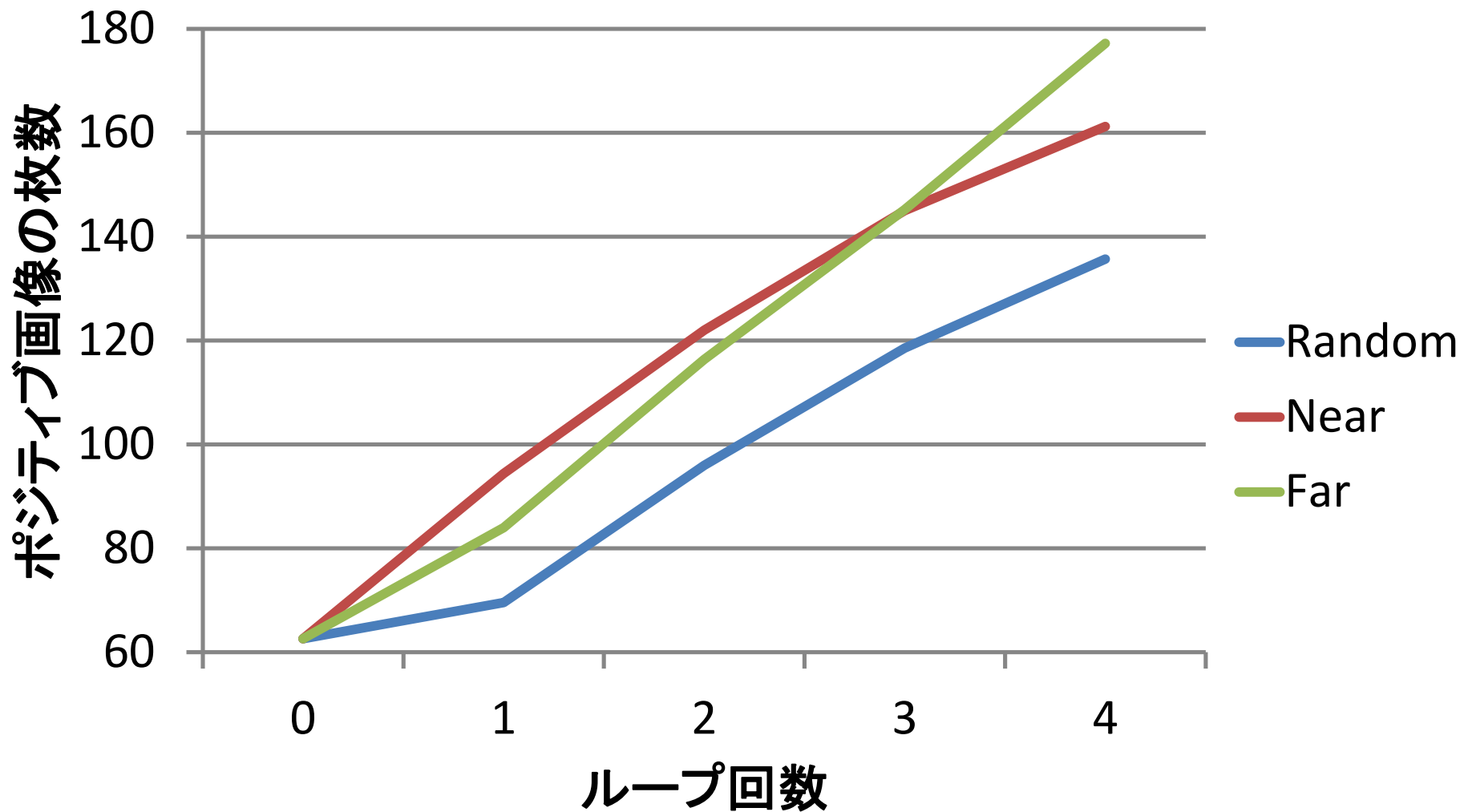
# アノテーション精度



# 出身国の分布

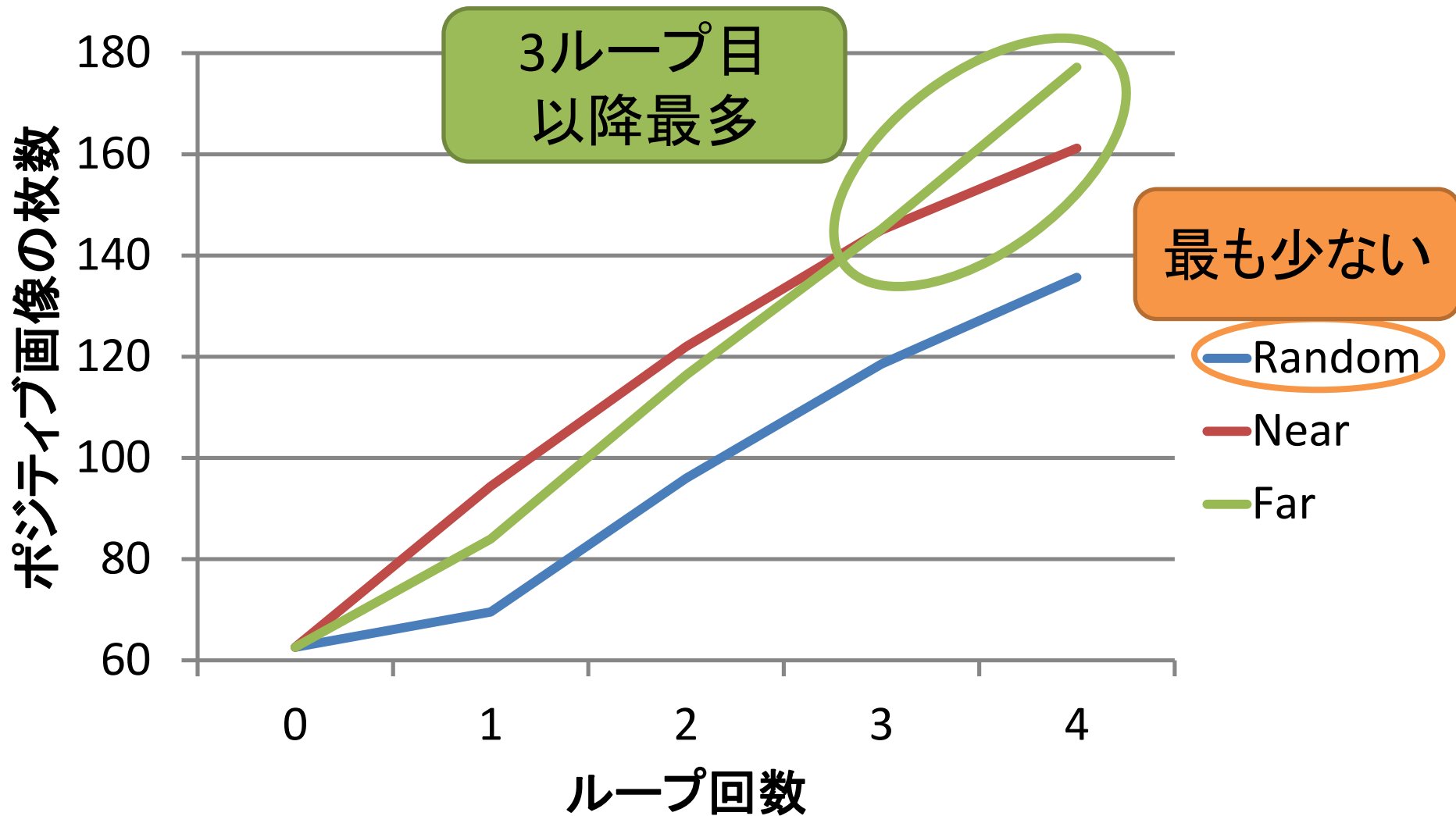


# ポジティブ学習画像の枚数

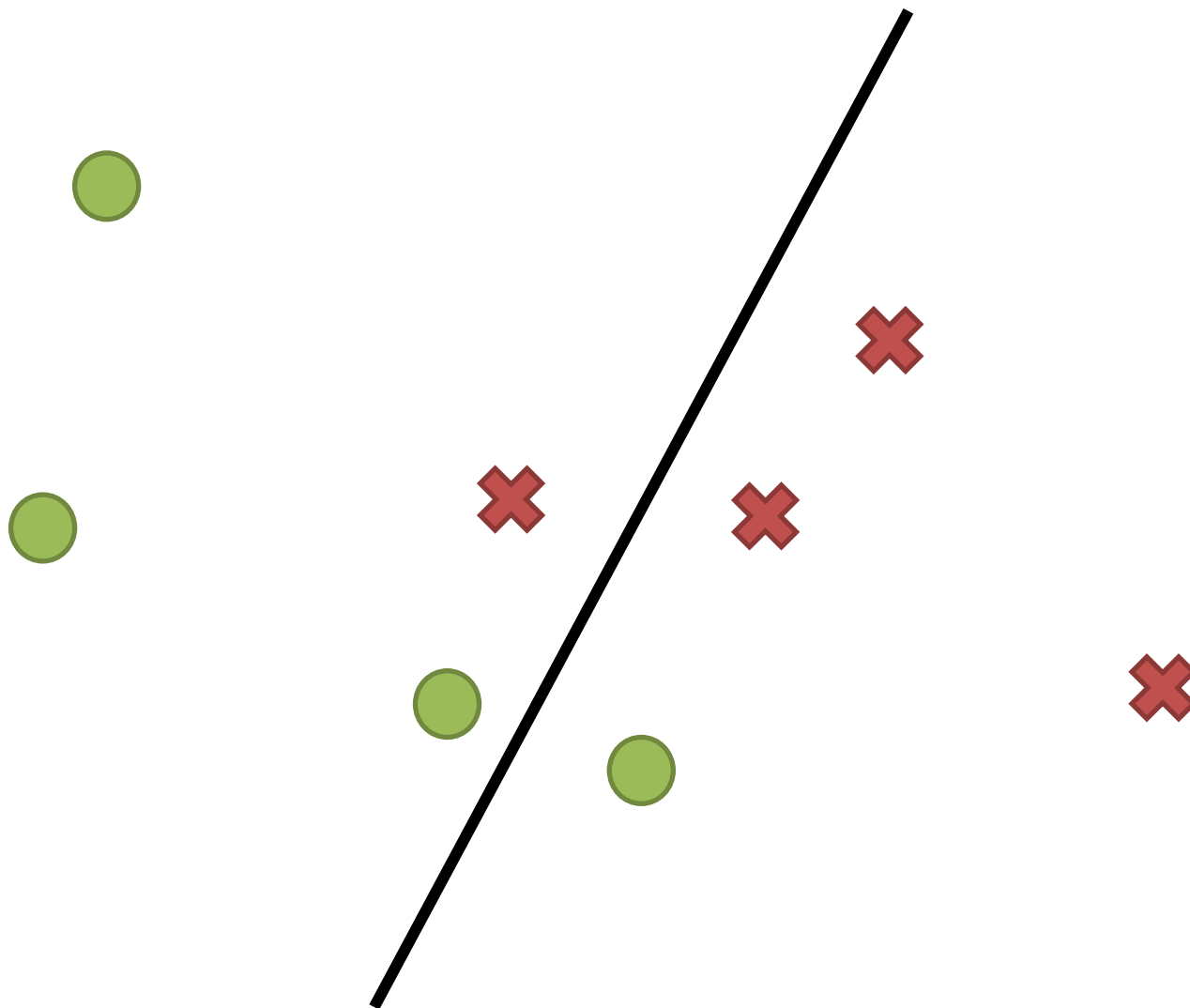




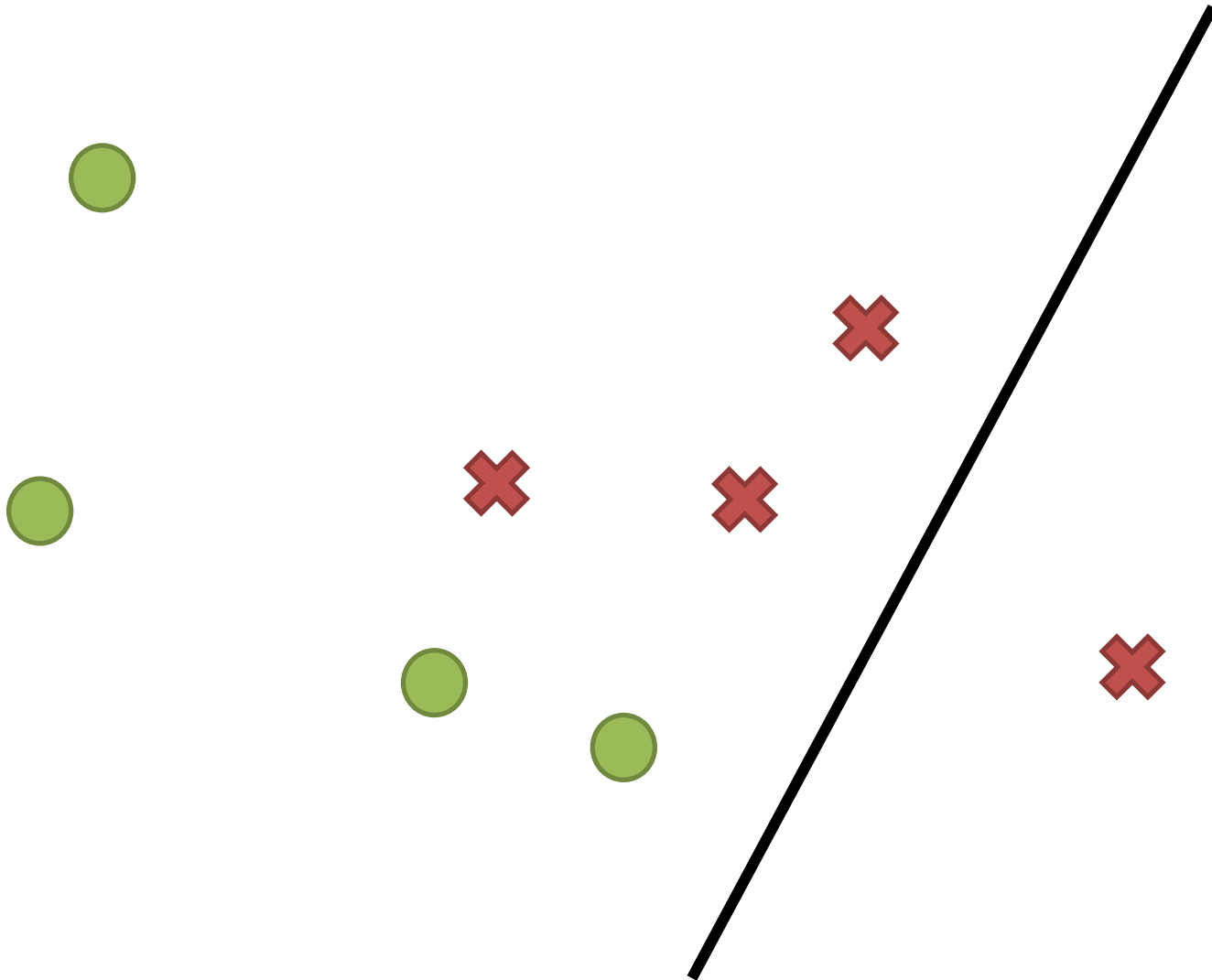
# ポジティブ学習画像の枚数



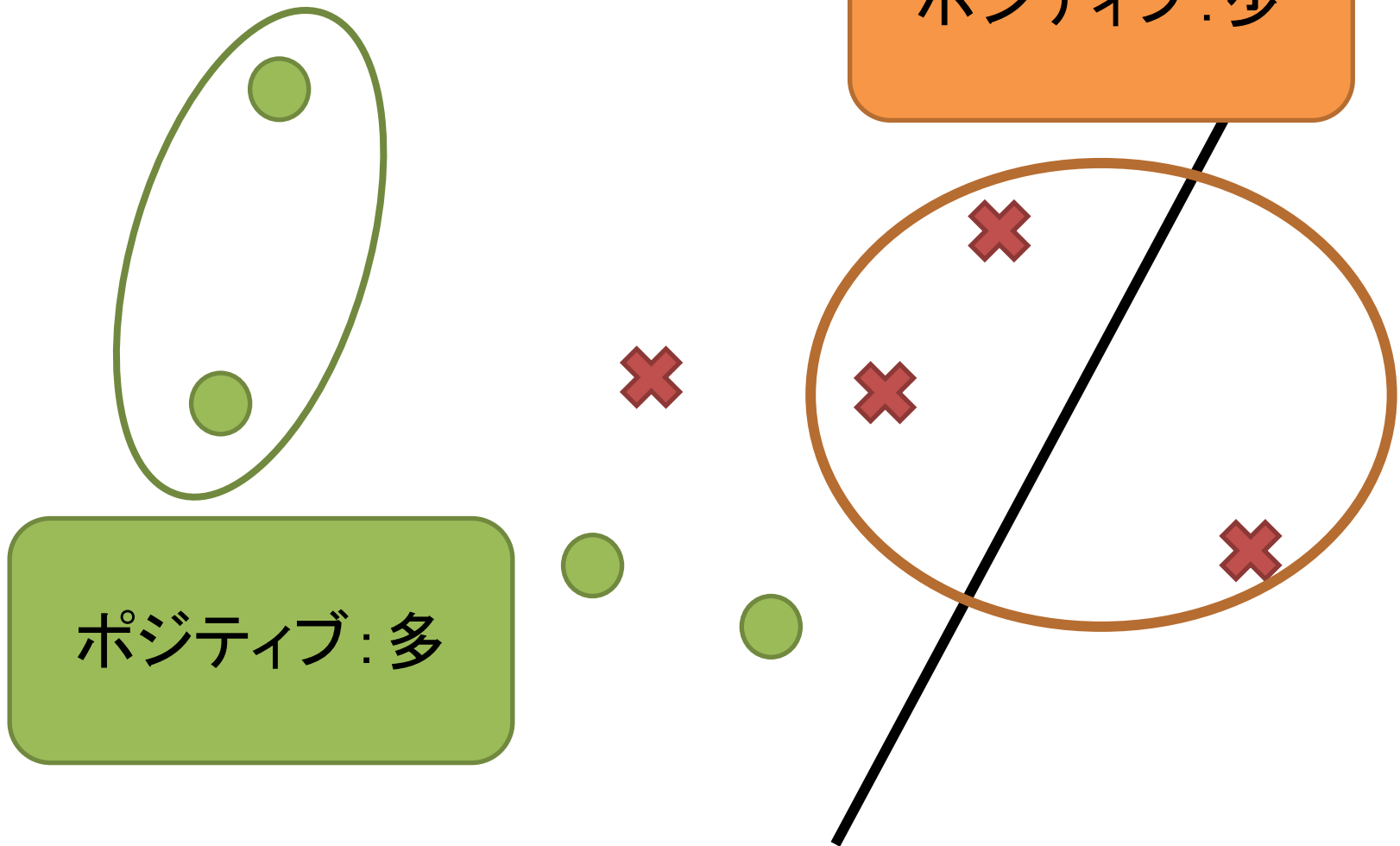
# 初期状態の認識モデル



# ループが進むと



# ループが進むと



# まとめ

- ループ学習の画像選出戦略を考案
  - Random, Near, Far
- 認識モデルのF値
  - どの戦略も横ばい
  - アノテーション精度に問題
- ポジティブ学習画像の枚数
  - ループを重ねるならFarが良い

# まとめ

- 今後の課題
  - アノテーション精度の向上
  - さらなる費用の削減