

【ICDE 2014勉強会】

## Session 25: Uncertain and Probabilistic Data

担当：石川佳治(名大)

Some figures are copied from ICDE 2014 proceedings.

# User-Driven Refinement of Imprecise Queries

- ▶ Bahar Qarabaqi, Mirek Riedewald (Northeastern U., USA)

- ▶ 目的: 曖昧な検索条件を対話的に洗練

- ▶ 特徴: 確率的なアプローチ

- ▶ 例: 鳥の観察画像の検索

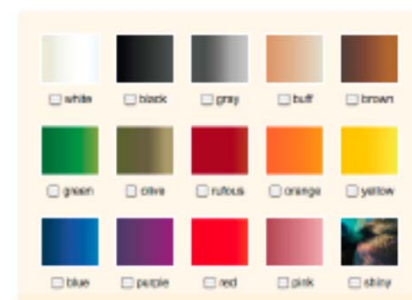

- ▶ 複数の属性: 羽の色, 大きさ, . . .

- ▶ ユーザは特定の画像を検索したいのではなく, 意図する鳥の種別 (例: ツグミ) で絞りこみたい

- ▶ ユーザは属性に対し**信念確率**を付与可能

- ▶ 例: 意図している鳥の多くが赤い羽根を持つとき, hasRedColorWing属性に (yes = 80%, no = 20%) という確率分布を指定

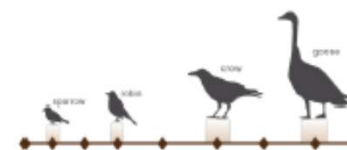
What color is the wing?  
Select 1-3 colors.



I'm sure Pretty Sure Not Sure

Select the approximate size of the bird.

Click on a point on the scale below



I'm sure Pretty Sure Not Sure

# インタフェースのイメージ

- ▶ 属性  $X_2 = \text{Size}$  と属性  $X_4 = \text{MainColor}$  に分布を指定した後の状況
- ▶ 類似度によるランキング(中)
- ▶ 未指定の属性で結果を改善しそうなものの順位
- ▶ 既指定の属性の感度 (sensitivity)

感度: 確率分布の値をいじるとどのくらい敏感に反応するか

Attribute	Quality Improvement	Rank	Species	Image	Specified Attribute	Sensitivity Score
$X_3$ :ShapeGroup	83.27	1	$C_{245}$ :Eastern Bluebird		$X_2$ :Size	206.88
$X_{14}$ :BillLength	81.12	2	$C_{211}$ :Blue Jay		$X_4$ :MainColor	18.01
$X_{11}$ :WingColor	74.98	3	$C_{223}$ :Barn Swallow			
$X_1$ :Time	69.47	4	$C_{212}$ :Western Scrub-Jay			
$X_7$ :Location	65.24	5	$C_{233}$ :Red-breasted Nuthatch			
$X_5$ :BreastColor	64.02	6	$C_{242}$ :Blue-grey Gnatcatcher			
$X_6$ :BreastPattern	63.18	7	$C_{221}$ :Tree Swallow			
$X_9$ :BackColor	57.79	8	$C_{331}$ :Indigo Bunting			
$X_{18}$ :LegColor	49.35	9	$C_{246}$ :Western Bluebird			
$X_8$ :BellyPattern	48.81	10	$C_{210}$ :Steller's Jay			

# 技術的な部分

---

- ▶ 感度分析 (Sensitivity Analysis)
  - ▶ 取りうる確率分布の値の変化で、**ランクが最大限どの程度変化するか**を評価: この論文で頑張っている部分 (面倒)
  - ▶ ランクの違いはMinkowski距離で評価
  - ▶ 確率に関する性質をうまく使って効率的に判定
- ▶ 追加条件のランキング (画面の左側)
  - ▶ **改善度の期待値** (expected improvement) を評価
    - ▶ 注目する画像 (エンティティ) を高く, 他をより低く評価するのがよい
  - ▶ エントロピーの概念を用いた定式化
  - ▶ 実直な評価手法では対話性に難あり
    - ▶ 決定木のアンサンブルによる近似的な評価