

ラーメン vs カレー: 2年分のログデータと 高速食事画像認識エンジンを用いた Twitter分析とデータセット自動構築

電気通信大学 総合情報学専攻

河野 憲之, 柳井 啓司

はじめに

- Twitterの流行
 - 気軽にコメントや画像を投稿可能
 - 食事を撮影して投稿する例も多々
- データセットの大規模化
 - 数百種類や数千種類など大規模化
 - より実用的に

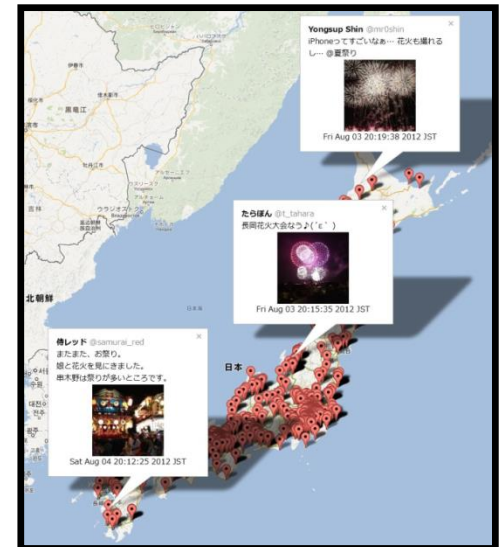
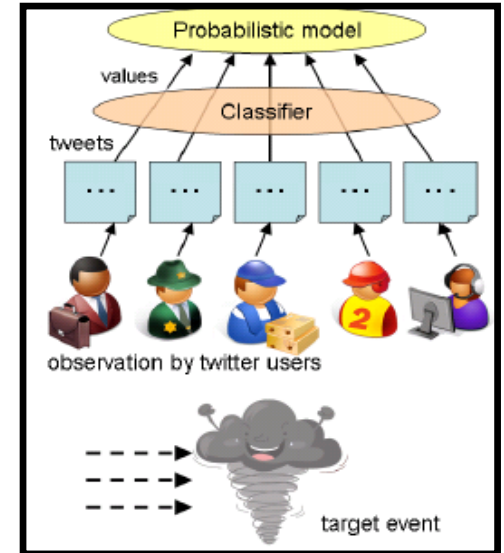
目的

- Twitter 食事画像分析
 - 投稿される食事画像の地域性の発見
 - ラーメン vs カレー
 - 既知カテゴリの画像収集
- データセットの自動構築
 - 食事画像データセットを対象
 - クラウドソーシングを用いる
 - 未知カテゴリの画像収集

高速食事画像認識エンジンを用いた
Twitter食事画像分析
～既知カテゴリの画像収集～

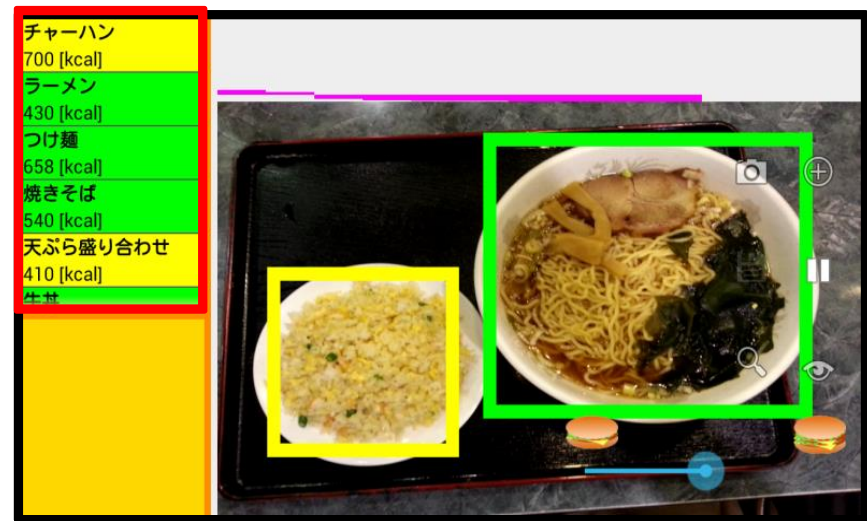
関連研究

- ツイートテキストの分析(榊ら WWW '10)
 - SVM+確率モデル
 - 地震の検出
- ツイート画像の利用(金子ら ICME WS'13)
 - グリッドごとにイベント
 - クラスタリング



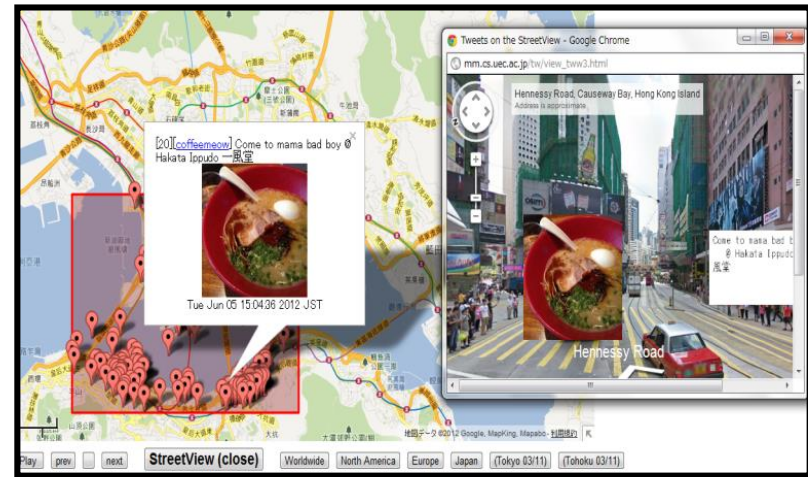
既発表システム (MIRU '13)

- 高速食事画像認識システム **FoodCam** (河野ら MIRU '13)
 - スマートフォン上で認識
 - HOG-FV + Color-FV, SVM
- 本発表ではPC上に構築
 - 100種類5位以内80%以上の分類率
 - 認識時間0.024秒



既発表システム (ICMR '12)

- 実時間ツイート画像地図への対応付け& 収集システム (柳井 ICMR '12)
 - Twitterストリームの監視
 - クラスタリング



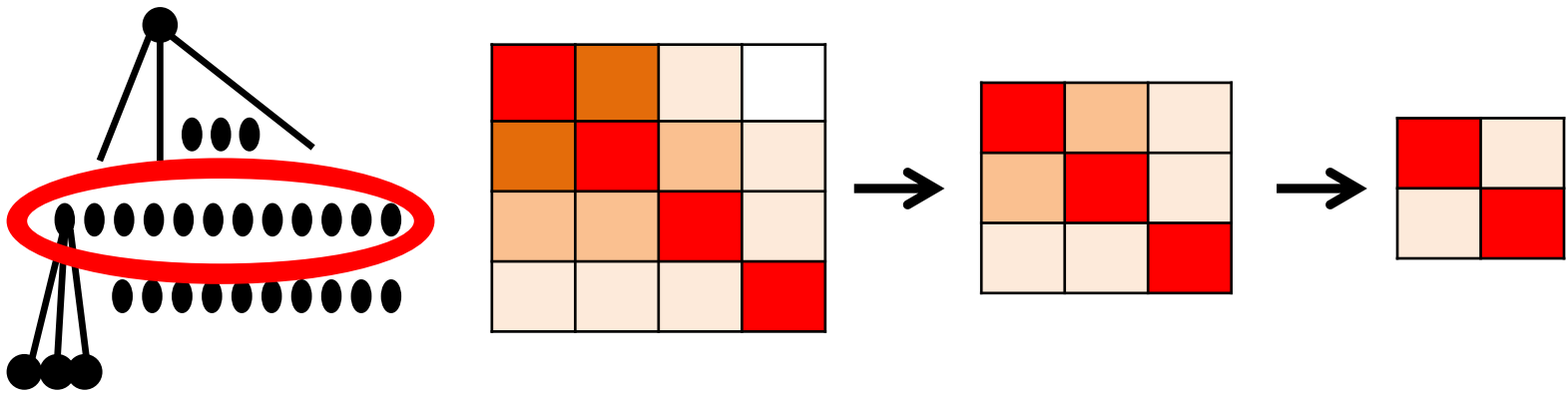
- 本発表では蓄えられた2年4か月分のツイートを分析

本発表

- 2つの既発表システムを組み合わせる
 - 実時間ツイート食事画像地図への対応付け&収集システム
 - 食事画像判別エンジンの導入
- Twitter食事画像の分析
 - 識別によるアプローチ
- 食事画像収集
 - 低ノイズ
 - より実用的な画像?

食事画像判別エンジン

- 既存の100種類食事画像のグループ化
 - 1-vs-rest SVMs の混同行列から混合の多いカテゴリ同士を結合、再帰的に実行
 - 13の食事グループの作成



食事画像判別エンジン

- 各グループでSVMの学習
 - 正例: グループに属する食事領域
 - 負例: 事前に手動で収集
 - レストラン、出店、飲食店、夕食会
 - 最終的な評価値はSVMの最大スコア
 - 最大反応

食事画像判別エンジン

作成されたグループ例

- 麺系(ラーメン、つけ麺、かけうどん、天ぷらうどん etc)
- 黄色系(オムレツ、オムライス、コーンスープ、茶わん蒸し etc)
- 汁物系(味噌汁、豚汁、けんちん汁、中華スープ etc)
- 茶色系(餃子、焼きそば、たこ焼き、お好み焼き etc)
- 揚げ物系(焼き鳥、唐揚げ、コロッケ、豚カツ etc)
- サラダ系(グリーンサラダ、ほうれん草炒め、野菜の天ぷら etc)
- パン系(サンドウィッチ、ブドウパン、ロールパン、クロワッサン etc)
- 海鮮系(刺身、寿司、海鮮丼)
- ご飯系(チキンライス、ピラフ、チャーハン、ご飯)
- 魚系(鮭の塩焼、干物、さんまの塩焼、魚の照り焼き etc)
- 煮物系(筑前煮、がんもどきの煮物、肉じゃが、おでん etc)
- ??(スパゲッティ、スパゲッティミートソース、エビチリ、シチュー etc)
- ??(野菜炒め、ゴーヤチャンプルー、チンジャオロース、きんぴらごぼう etc)

ツイート食事画像分析の流れ

1. ツイートテキストからキーワード検索、
画像が添付されたツイートを対象
2. 食事画像判別エンジンにかける
 - ノイズの除去
3. 食事画像認識エンジンにかける
 - 上位5位以内に入っていれば取得
 - 対象の食事画像の取得
4. 結果の分析

分析項目

- 100種類の食事のそれぞれの出現枚数
- 食事画像集合の適合率 (300枚無作為抽出)
 - キーワード検索のみ
 - 食事画像判別エンジン
 - 100種類食事画像認識エンジン
 - 上記2エンジンを組み合わせ
- 位置情報付きツイートの地理的な分布の比較
 - ラーメン vs カレー

食事画像取得結果

- 対象の2年4カ月の位置情報付きツイートデータ
 - 約10億件(サンプリング)
- 100種類のキーワード検索 (日本語)
 - 約170万枚
- 食事画像判別エンジン+
100種類食事画像認識エンジンで上位5位以内
 - 約47万枚

100種類の出現頻度ランキング上位10位

| 出現頻度の順位 | 食事名 | 出現頻度 |
|---------|-------|-------|
| 1 | ラーメン | 80021 |
| 2 | カレー | 59264 |
| 3 | 寿司 | 25898 |
| 4 | つけ麺 | 22158 |
| 5 | オムライス | 17520 |
| 6 | ピザ | 16921 |
| 7 | 餃子 | 16014 |
| 8 | お好み焼き | 15234 |
| 9 | ご飯 | 14264 |
| 10 | 刺身 | 13927 |

収集された画像例

- オムライスが6位、ケチャップで絵が描かれたものが多数



- よく食べるハンバーガー(30位)や、牛丼(27位)はチェーン店が多く、必ずしも上位ではない
- つけ麺も意外と多い

食事画像集合の適合率

| 品目 | キーワードのみ | 食事検出 |
|---------|----------------|----------------|
| ラーメン | 275652 (72.0%) | 200173 (92.7%) |
| チャーシュー麺 | 861 (94.3%) | 811 (99.0%) |
| カレー | 224685 (75.0%) | 163047 (95.0%) |
| カツカレー | 10443 (92.7%) | 9073 (98.0%) |
| 寿司 | 86509 (69.0%) | 43536 (86.0%) |
| つけ麺 | 33165 (88.7%) | 24896 (96.3%) |
| オムライス | 34125 (90.0%) | 28887 (96.3%) |

| 品目 | 100 種類認識 | 検出+認識 |
|---------|---------------|---------------|
| ラーメン | 84189 (95.0%) | 80021 (99.7%) |
| チャーシュー麺 | 558 (99.7%) | 555 (99.7%) |
| カレー | 62824 (97.0%) | 59264 (99.3%) |
| カツカレー | 6544 (98.7%) | 6339 (99.3%) |
| 寿司 | 48019 (72.3%) | 25898 (92.7%) |
| つけ麺 | 28846 (93.7%) | 22158 (99.0%) |
| オムライス | 18370 (98.0%) | 17520 (99.0%) |

キーワードのみの例(ラーメン)



食事検出の例



食事認識の例



食事検出+認識後の結果例



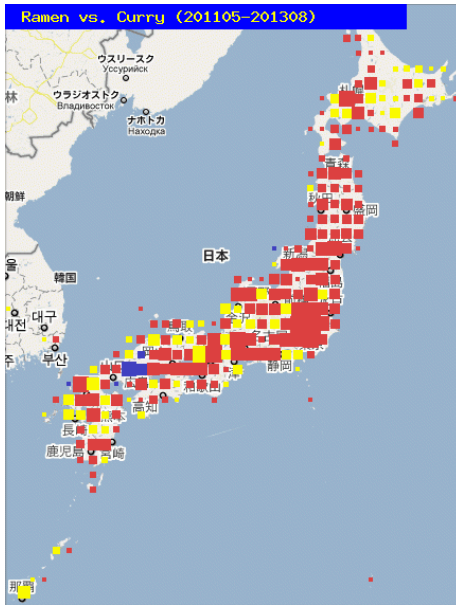
キーワードのみの例(カレー)



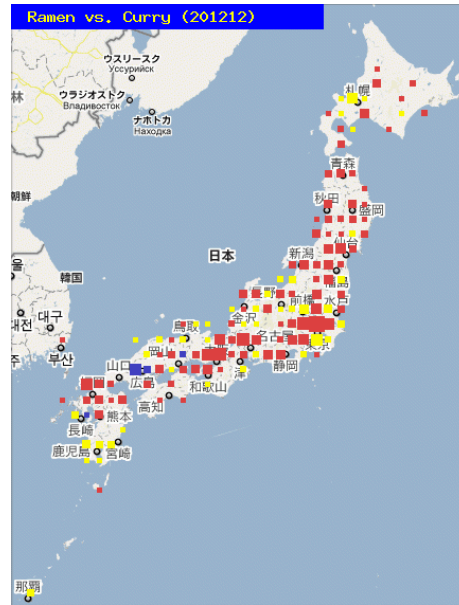
検出+認識後の結果例



ラーメンとカレーの地理的分布

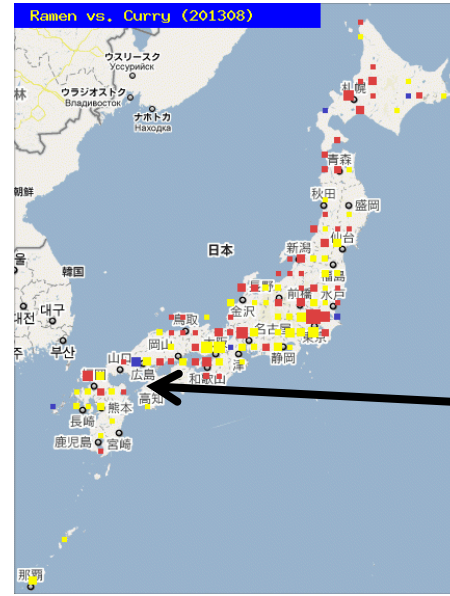


合計



12月 (冬季)

ラーメンが多い



8月 (夏季)

カレーが多い

広島周辺では
お好み焼き

● ラーメン, ● カレー, ● お好み焼き

画像収集2

- 同時にTwitter stream APIから「ラーメン」で取得
 - APIの上限の緩和
 - 1カ月でラーメン画像**2万枚**
 - 4年後には**100万枚**
 - 「ラーメン」FGVC用データセットの構築
- Twitter streamを監視し、実時間処理

まとめ

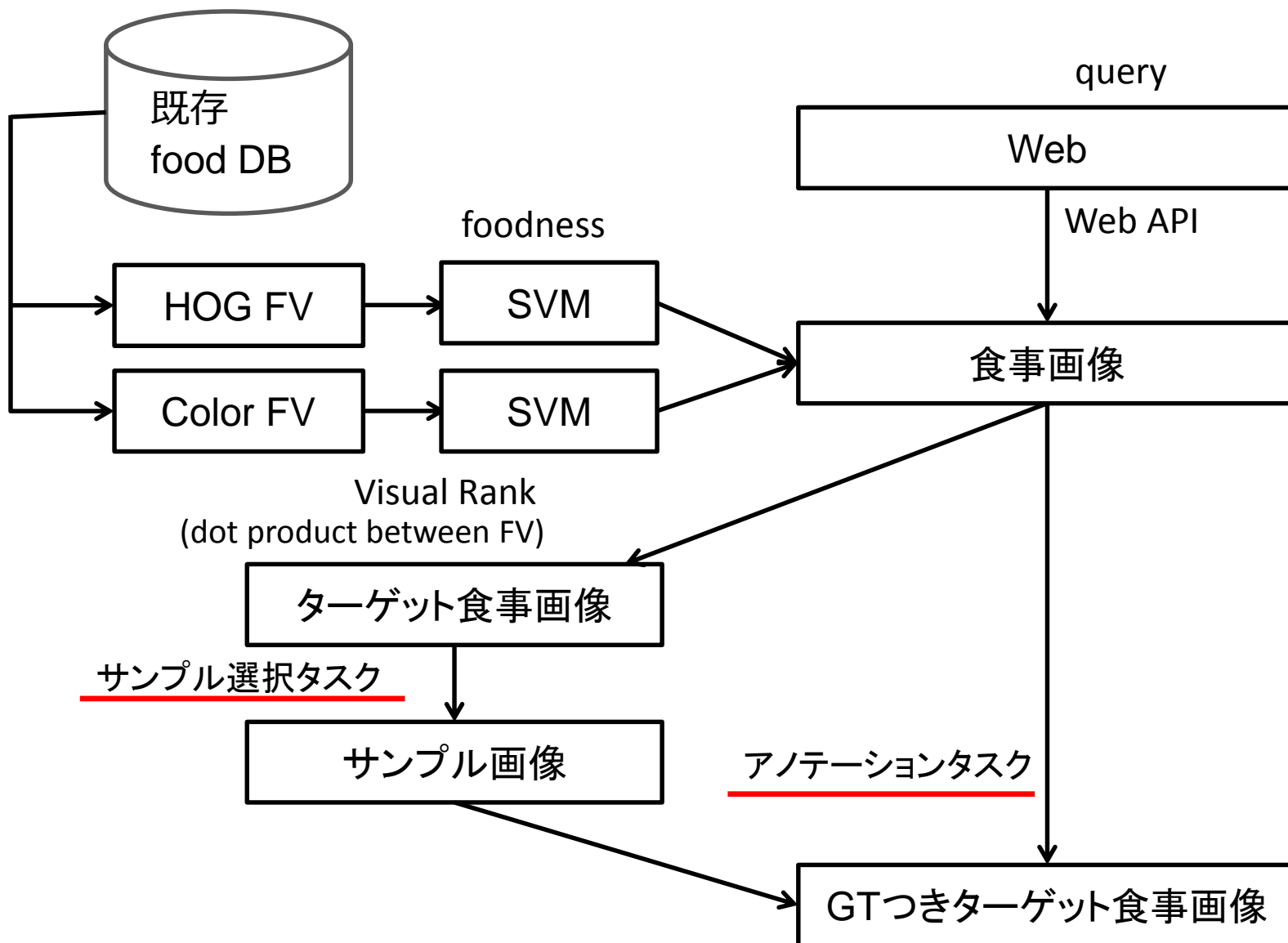
- Twitterのログデータに対して、食事画像の分析
 - 食事画像判別エンジン、100種類食事画像認識エンジンにより、高精度で食事画像を収集可能
- 高速画像認識エンジンの応用例として、投稿されたツイートを実時間で食事画像判定し、つぶやく

クラウドソーシングを用いた
データセット自動構築
～未知カテゴリの画像収集～

データセット自動構築の流れ

1. Webから食事画像を収集
 - ノイズのフィルタリング
2. クラウドソーシングでサンプル画像を取得
3. クラウドソーシングでBB付き食事画像を取得

システムの流れ



データセット自動構築の流れ

1. Webから食事画像を収集
 - ノイズのフィルタリング
2. クラウドソーシングでサンプル画像を取得
3. クラウドソーシングでBB付き食事画像を取得

Webから食事画像を収集

- 問題点
 - ノイズが多い
 - 全てをAMTに用いるとコストが高い
- 解決法
 - 食事画像判別エンジンによるフィルタリング
 - 未知のカテゴリに適用

食事画像判別前

- 沖縄そば



食事画像判別後

- 沖縄そば



ターゲットと判定に貢献したカテゴリの一部

ターゲット

食事グループの一部



データセット自動構築の流れ

1. Webから食事画像を収集
 - ノイズのフィルタリング
2. クラウドソーシングでサンプル画像を取得
3. クラウドソーシングでBB付き食事画像を取得

クラウドソーシングでサンプル画像を取得

- 問題点

- ターカーは対象の料理について知らない
- ノイズを除去しても、少なからず含まれる

- 解決法

- バウンディングボックス付と作業前にサンプル画像を収集
- ターカーに30枚画像を投げ、最も対象の料理らしいサンプルを選択してもらう

画像選択

- 収集した画像から、ターカーに投げる30枚を決定
 - SVMのスコア順だと対象以外も含まれる恐れ
 - Visual Rank
 - 上位300枚を順位付け (上位100枚に重み)
- Visual Rank
 - 類似度---Fisher Vectorの内積
- Fisher Vector
 - GMMは作成済みを使用 (未知カテゴリの情報はない)

サンプル選択タスク

- タスクの特徴

- 30枚を与えて、最大7枚、
一般的なサンプルを選択してもらう
- リンクを与えるので勉強してもらう
- \$0.06、4人
 - 少し高め、外部リンクで勉強させる、また重要であるため
- 質の向上
 - チェックをしないとヒットを提出できない
 - 設置したリンクをたどらないとヒットを提出できない

Visual Rank適用前

ターゲット以外の食事画像も含まれる



Visual Rank適用後



大域特徴との比較

Color Gist



タスク例

チゲのサンプル画像選択タスク



收集結果例



サンプル選択結果

- 上位5枚が対象の食事である—98%
 - アップで撮影された食事画像も若干数
- ワーカーからの評価 (BB付与とタスクで質問)

| 有用であった | まずまず | 有用でなかった |
|--------|------|---------|
| 763 | 103 | 22 |

実際の影響については未調査

データセット自動構築の流れ

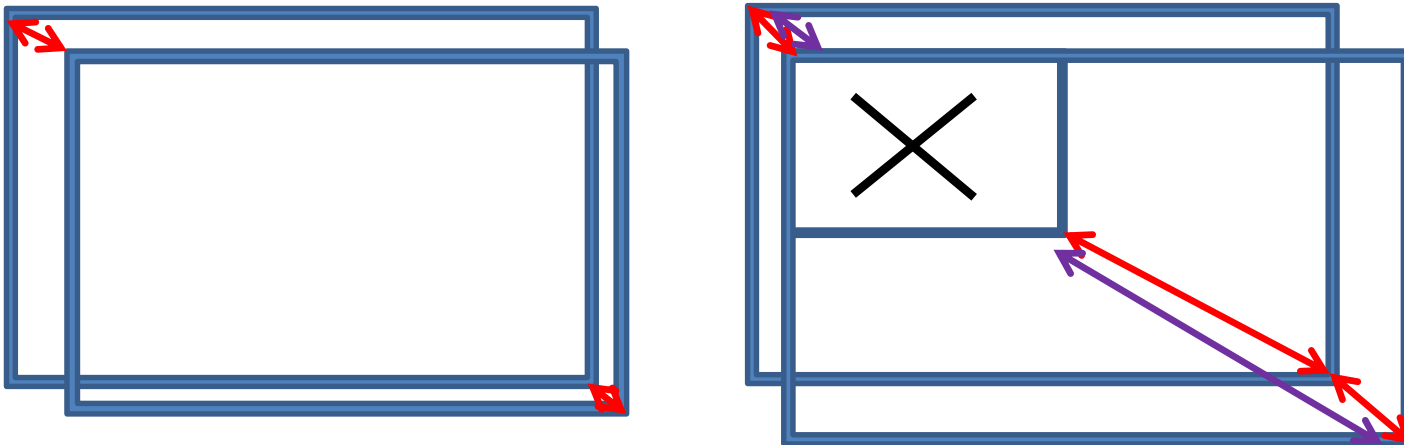
1. Webから食事画像を収集
2. クラウドソーシングでサンプル画像を取得
3. クラウドソーシングでBB付き食事画像を取得

クラウドソーシングでBB付き食事画像を取得

- タスクの目的
 - 人手による対象でない食事画像の除去
 - バウンディングボックスの付与
- タスクの特徴
 - \$0.05、4人
 - 質の向上
 - 現在の状況の提示
 - 入力時にノイズのチェック

結果の結合

- 複数のワーカーによる結果の結合
 - 複数のバウンディングボックスが付近x%内に存在 (始点、終点)
 - そのバウンディングボックスの平均
 - データベースに加える



タスク例

取得したサンプル画像



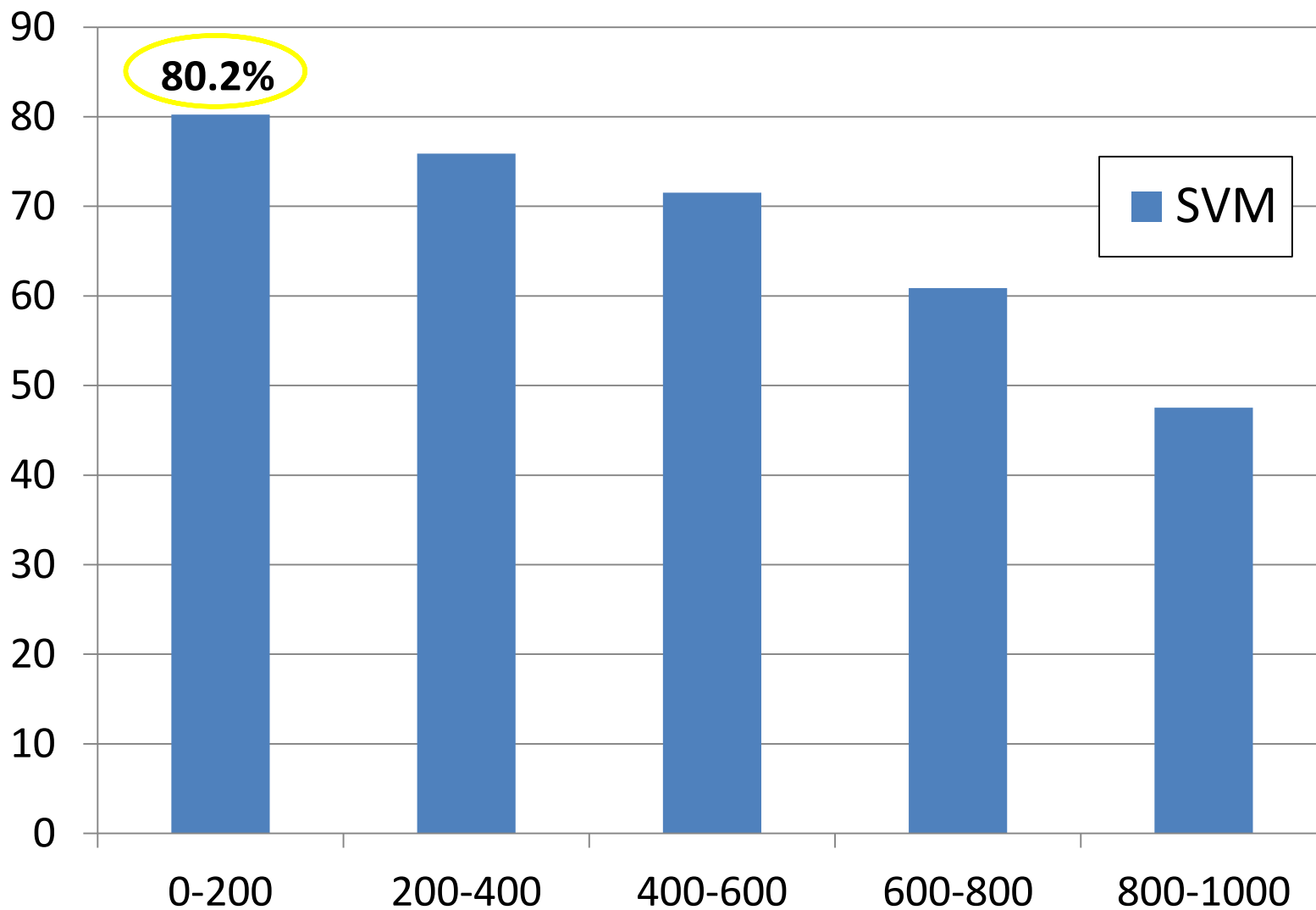
バウンディングボックス付与タスク



実験設定

- 選別方法
 - 食事画像判別エンジン (SVM)
- 収集ターゲット
 - 15種類の未知の食事
 - 各1000枚以上の画像
- 評価
 - 対象の食事の適合率

実験結果



実験設定

- 選別方法
 - クラウドソーシング (AMT)
- 収集ターゲット
 - 20種類の未知の食事
- 評価
 - 対象の食事の適合率

実験結果

| | 適合率 |
|---------------|----------------|
| ランダム | 48.6% |
| 食事画像判別器 (15種) | 80.2% (上位200枚) |
| AMT | 96.9% (平均158枚) |

AMTでの結果

- 一般的な食事画像以外の画像も多々



- ワーカーの国別人口

| india | usa | Canada | French | Philippine | Asian |
|-------|-----|--------|--------|------------|-------|
| 1000 | 271 | 19 | 2 | 1 | 1 |

まとめと課題

- クラウドソーシングを用い、食事画像データセットの自動拡張に取り組んだ
- 未知の食事に対して、既存の食事画像を用いることは有効
- ターゲットの食事の知識を導入する
- 人手で付与したバウンディングボックスとの比較を行う

収集ターゲット食事画像

- 沖縄そば、マンゴープリン、杏仁豆腐、冷麺、ティラミス、串カツ、ベーグル、ちゃんぽん、釜飯、ナチョス、パエリア、シーザーサラダ、ワッフル、アップルパイ、ショートケーキ、ホットケーキ、ミネストローネ、custard tart、oyster omelette、stinky tofu

認識実験設定

- データセット1 (自動)
 - 今回作成したデータセット
- データセット2 (手動)
 - 既存食事画像データセットから20種類選択
 - カテゴリはそれぞれ異なる
- 特徴量
 - HOG FV (D:64, K:64, SP:0)

実験結果

