

Twitter食事画像からの詳細カテゴリ発見

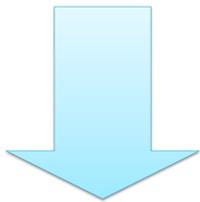
電気通信大学大学院 情報理工学研究所 情報学専攻 伊藤 祥文 柳井 啓司

背景と目的

Twitterでの画像付き投稿が増加

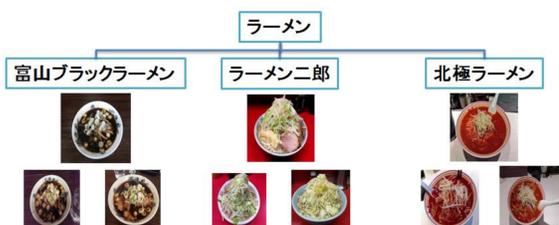
流行のものなど、新しいデータが使える

画像データとテキスト情報が利用可能



大量のツイート内容を分析し、ある単語に対する詳細カテゴリを発見

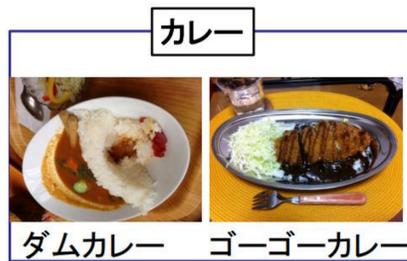
発見されたカテゴリを用いて、投稿画像からデータセットを作成



各カテゴリごとの固有の特徴を発見し、画像認識可能な下位カテゴリを発見



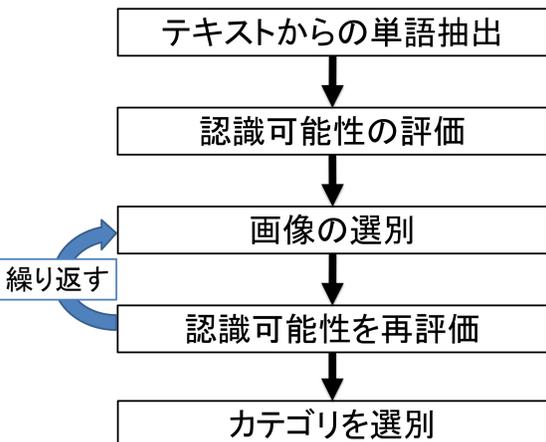
「ラーメン」から得られた詳細カテゴリとその画像の例



「カレー」から得られた詳細カテゴリとその画像の例

「ラーメン」と「カレー」の2つのカテゴリについて、データセットを作成した

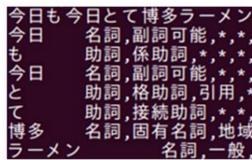
手法概要



手法詳細

1. 単語抽出

- Twitter過去ログを用いる
- 食事画像つきツイート約5年分
- MeCabを用いて単語抽出



今日
今日
博多ラーメン

- 対象の単語を含むものを用いる

2. 認識可能性の評価

- Deep Learning
- Caffeを用いてマルチクラス分類
- Digitsを用いて学習

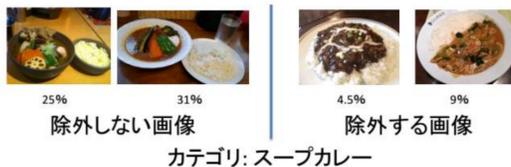


図 4 認識可能性の評価を行った例

➢ GoogLeNetを用いてfine tuning

3. 画像の選別

- ノイズ画像を除去
- 画像ごとの確率値で判定
- 確率値の低い画像は除外



- ノイズ画像の選別方法

- ①各カテゴリから一定数除外
 - ・確率値の低い順に画像を除外
 - 各カテゴリの画像数に変化しない
 - ×十分にノイズを除去できない
 - ×非ノイズ画像を除去する可能性
- ②一定の確率値以下の画像を除外
 - ・特定の確率値以下の画像を除外
 - 高確率でノイズ画像が除外できる
 - ×カテゴリ間の画像数に差が発生

4. カテゴリの選別

- ノイズとなるカテゴリを除去
- カテゴリの全画像の確率値で判定
- 全画像の確率値の平均を用いる
- 確率値の低いカテゴリは除外
- ノイズカテゴリの選別方法
- 平均が一定以下のカテゴリを除外
- 画像数に関係なく選別可能
- 確率値の低いノイズ画像によって結果が変わる可能性

実験・結果

ラーメンの詳細カテゴリ

味噌ラーメン	塩ラーメン	醤油ラーメン	台湾ラーメン	家系ラーメン
豚骨ラーメン	とんこつラーメン	徳島ラーメン	博多ラーメン	カレーラーメン
トマトラーメン	尾道ラーメン	カップラーメン	佐野ラーメン	チキンラーメン
喜多方ラーメン	みそラーメン	朝ラーメン	ネギラーメン	和歌山ラーメン
インスタントラーメン	辛味噌ラーメン	熊本ラーメン	野郎ラーメン	長浜ラーメン
ラーメンテロ	しょうゆラーメン	特製ラーメン	ラーメン最高	二郎系ラーメン
ラーメン二郎	北極ラーメン	札幌ラーメン	旭川ラーメン	激辛ラーメン
八王子ラーメン	野菜ラーメン	ラーメンセット	ラーメン犬	スタミナラーメン
最近ラーメン	ラーメン画像	牛骨ラーメン	味玉ラーメン	彩華ラーメン
大砲ラーメン	ラーメン人生	味噌カレー牛乳ラーメン	横浜家系ラーメン	8番ラーメン

データセットの評価

ラーメンの分類率

	Top-1	Top-5	データ数
ノイズ除去なし	40.2%	61.9%	12500
ノイズ除去1回	49.6%	69.9%	10000
ノイズ除去2回	63.6%	84.9%	8018

得られた画像



カレーの詳細カテゴリ

カレーうどん	スープカレー	カレーライス	チキンカレー	キーマカレー
グリーンカレー	インドカレー	朝カレー	ドライカレー	ビーフカレー
野菜カレー	カレーパン	バターチキンカレー	夏野菜カレー	タイカレー
ゴーゴーカレー	ポークカレー	トマトカレー	チーズカレー	オムカレー
シーフードカレー	牛すじカレー	カレーランチ	海軍カレー	焼きカレー
カレードリア	レッドカレー	マトンカレー	インドアンカレー	牛タンカレー
カレー鍋	インディアンカレー	金沢カレー	スライムカレー	ほうれん草カレー
牛スジカレー	カレー曜日	金曜カレー	黒カレー	手作りカレー
カレーセット	カレーそば	豆カレー	ダムカレー	ナイスカレー
カレー皿	カレースープ	唐揚げカレー	カツカレー	エビカレー

データセットの評価

カレーの分類率

	Top-1	Top-5	データ数
ノイズ除去なし	34.3%	59.3%	12500
ノイズ除去1回	41.4%	67.2%	10000
ノイズ除去2回	54.8%	82.7%	7839

得られた画像



まとめ・今後の課題

- ◆ Twitterの内容を用いて詳細カテゴリを発見
- ◆ 発見したカテゴリをもとに詳細分類データセットを作成
- ◆ 対象の単語を含まないカテゴリのデータセット作成
- ◆ 学習用の画像が必要