

# 統計的機械学習による音声/画像/SNS 情報解析に関する研究

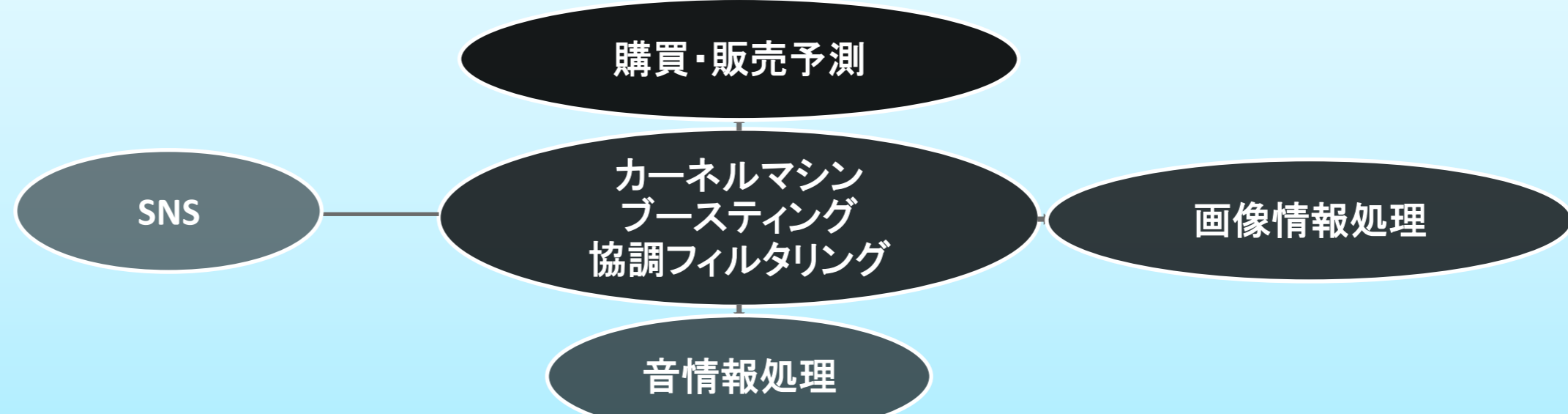
松井 知子 モデリング研究系 教授

## 【概要】

本研究室では統計的学習機械を用いて、音声/音楽/画像/SNSなどを処理する方法について研究しています。具体的にはカーネルマシン、ブースティング、協調フィルタリングの手法を用いて、

1. 音声・話者認識
2. 音楽ジャンル分類
3. 画像識別
4. SNS解析
5. トピック分類
6. WEBユーザビリティ評価 など

の研究課題に取り組んでいます。



本研究室では統計的機械学習とその応用研究に興味のある学生さんを募集しています！

## 【統計的機械学習】

- 統計科学を用いて、
  - データから、内在する数学的な構造を発見する。
  - その数学的な構造に基づいて、予測や判別などの情報処理を行う。
- 帰納的アプローチ  
v.s.  
自然科学でよく見られる演繹的アプローチ
  - 仮説をたて、推論し、実験的または理論的に検証する。
- カーネルマシン
  - 自動的な特徴(モデル)選択機構を含む。
  - 非線形の扱いに優れている。
  - サポートベクターマシン(SVM)、罰金付ロジスティック回帰マシン
- いろいろな確率モデルによる方法
  - 混合ガウス分布モデル
  - 隠れマルコフモデル
- 協調フィルタリング など

## 【スパースコーディングを用いた自己教師学習による音楽ジャンル分類】

### いろいろな音楽データベース

複数のデータベースを活用するには？

### スパースコーディング

$$X_{(d,n)} \approx B_{(d,k)} \times A_{(k,n)}$$

入力データ  $X_{(d,n)}$  は ベース(辞書)  $B_{(d,k)}$  と アクティベーション(スパース)  $A_{(k,n)}$  の積として表される。

$$x_i \approx \sum_j a_j^i b_j$$

### 利用データ

データベース	GTZAN	ISMIR (2004)
ジャンル	Classical, Country, Disco, Hip-Hop, Jazz, Rock, Pop, Blues, Reggae, Metal	Classical, Electronic, Jazz-Blues, Metal-Punk, Rock-Pop, World
サイズ	各ジャンルにつき 100 30秒クリップ	各ジャンルごとに異なる長さ
役割	ラベルなしデータ	ラベルありデータ
分析単位	5秒	5秒

### 半教師あり学習

ラベルありデータ → 学習システム → モデル

### 転移学習

ラベルありデータ (ソースドメイン) → 学習システム → 知識 → 学習システム → モデル (目標ドメイン)

### 自己教師学習

ラベルありデータ → 高次特徴量 → 教師あり学習 → モデル

ラベルなしデータ → 教師なし学習 → 高次特徴量

### 定式化

- ラベルありデータ:  $C$ 個のクラス  $X^l = \{(x_i^l, y_i)\}_{i=1}^n$   $x_i^l \in R^d$ ,  $y_i \in \{1, \dots, C\}$
- ラベルなしデータ  $X^u = \{x_i^u\}_{i=1}^m$   $x_i^u \in R^d$
- ベース(辞書)学習 (Feature-sign search [Lee et al. 2006])

$$\min_{a,b} \sum_i \|x_i^u - \sum_j a_j^i b_j\|_2^2 + \beta \sum_i \|a^i\|_1 \quad \text{s.t. } \|b_j\|_2 \leq 1$$

再構成誤差      スパース度のペナルティ

### 実験結果

#### Performance with bases learned from GT-50

### 自己教師学習のアプローチ:

- ラベルありデータが少ない時に有効
- 辞書サイズが大きくなるほど効果も大きい
- ラベルなしデータのサイズの影響は少ない