統計的機械学習によるマルチメディアデータ解析に関する研究

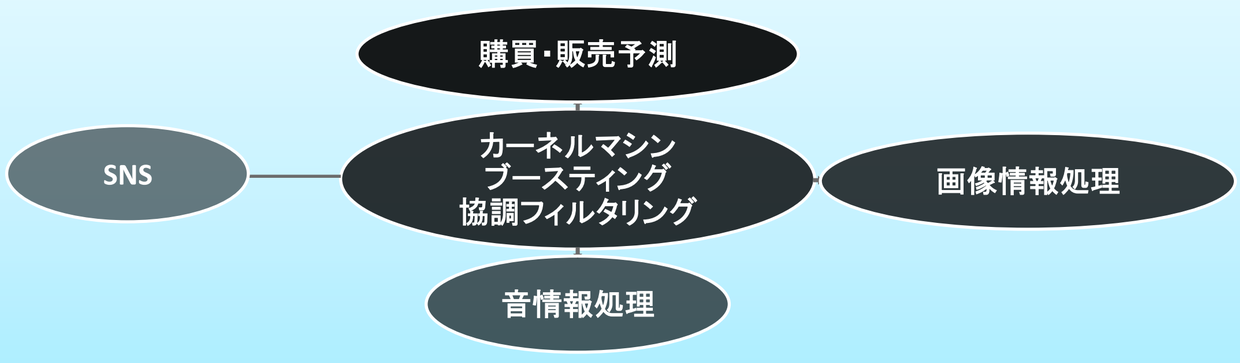
松井　知子　　モデリング研究系　教授

【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. 音楽情報処理
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

【ガウス過程による音楽情報処理】

ガウス過程

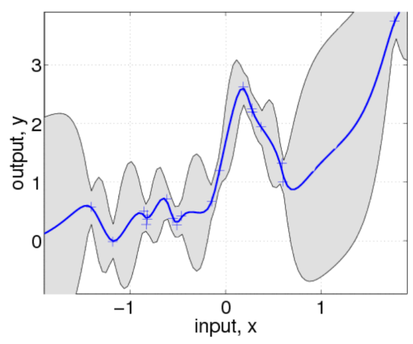
- GP回帰:

$$\mathbf{f} = \left[f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)\right] \quad n = 1, \ldots, \infty \quad \mathbf{f} \mid \mathbf{X} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$$

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')),$$

$$m(\mathbf{x}) = E[f(\mathbf{x})]$$

$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$



音楽情報検索への応用

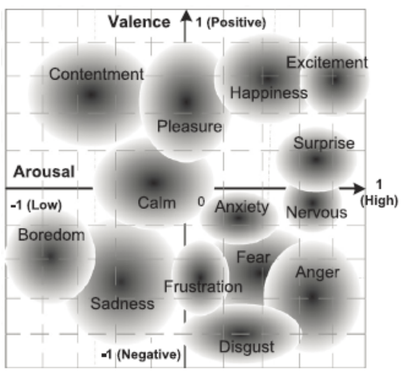
- なぜ必要か?
  - A lot of online music – fast and efficient access is demanded.
  - Intelligent music services – dynamic / personalized playlist generation, recommendations, etc.
- 主なタスク
  - **Musical genre / mood recognition.**
  - Artist / album recognition.
  - Music transcription.
  - Music structure analysis, etc.

音楽ムード推定

- 人間の感情
  - Happiness, fear, sadness, anger, etc.
  - High degree of subjectivity.
  - Varying levels of intensity – very happy, little nervous, somewhat bored, etc.
- 音楽ムード
  - Corresponds to human emotions – Happy, fearful, sad, angry, etc.
  - Even more subjective than human emotion.
  - Difficult to formalize.

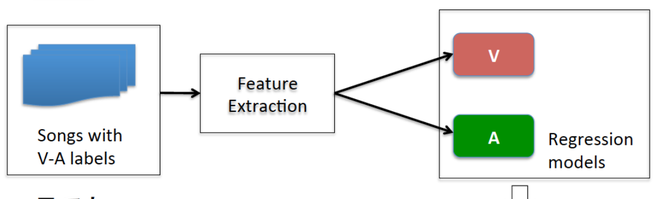
人間の感情モデル

- Valence(感情の positive と negative の度合い)-Arousal(感情の興奮度合い) による感情空間

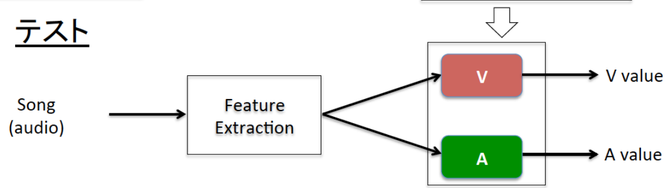


音楽ムード推定システム

- 学習



- テスト

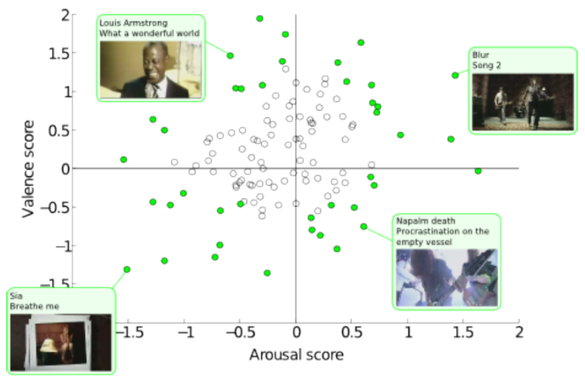


- V, Aごとにガウス過程による回帰モデル化

$$V = f_V(\mathbf{x}) + \varepsilon_V$$

$$A = f_A(\mathbf{x}) + \varepsilon_A$$

- 評価



- Two scoring cases:
  - One V-A value for the whole song – coefficient of determination (**R squared**)

$$R^2 = \frac{\mathrm{cov}(X,Y)^2}{\mathrm{var}(X)\mathrm{var}(X)}$$

obtained separately for V and A.

  - Multiple V-A pairs per song (for variable mood) – **Kendall's tau** coefficient.

実験:音楽ムード推定

- データベース
  - EmotionInMusic2013.
    - Part of MediaEval2013 benchmarking initiative for multimedia evaluation. (www.multimediaeval.org/mediaeval2013)
  - Data description
    - Western music.
    - 700 song clips (45 seconds).
    - A-V labels obtained from multiple annotators (mean and std).

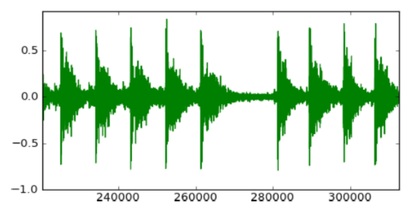
実験:特徴抽出

- One feature vector per clip consisting of statistics of frame level:
  - 13 MFCCs.
  - 24 sub-bands spectral crest factor coefficients:

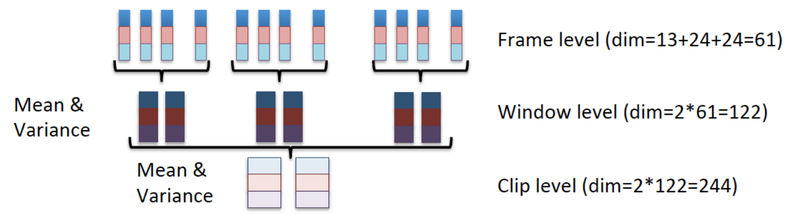
$$C = \max(s(k)) / \mathrm{mean}(s(k))$$

  - 24 sub-band spectral flatness coefficients:

$$F = \sqrt[K]{\prod_{k=0}^{K-1} s(k)} / \mathrm{mean}(s(k))$$



Frame length	23.2 ms
Frame shift	23.2 ms
Window length	43 frames
Window shift	1 frame



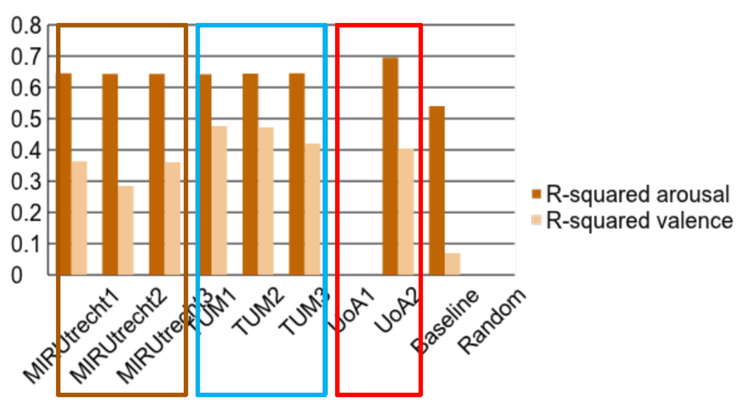
実験:結果

- Results, 7-fold CV (in terms of R^2).

Kernel	SVR		GPR	
	Valence	Arousal	Valence	Arousal
Linear	0.611	0.328	0.616	0.333
RBF / SE	0.659	0.401	0.658	0.383
Rational	-	-	0.661	0.430
Matern 3	-	-	0.667	0.401

  - GPR – Gaussian likelihood, Constant mean, Exact inference (GPML).

- MediaEval2013ベンチマークワークショップ
  - 歌唱データのV-Aスコア推定タスク(1000曲の歌唱データ)
  - ガウス過程(UoA) vs. Recurrent NN(TUM) vs. 重回帰分析(MIRUtrecht)



共同研究者:K. Markov氏(会津大)、G. Peters氏(UCL)