

統計的機械学習によるメディア データ解析に関する研究

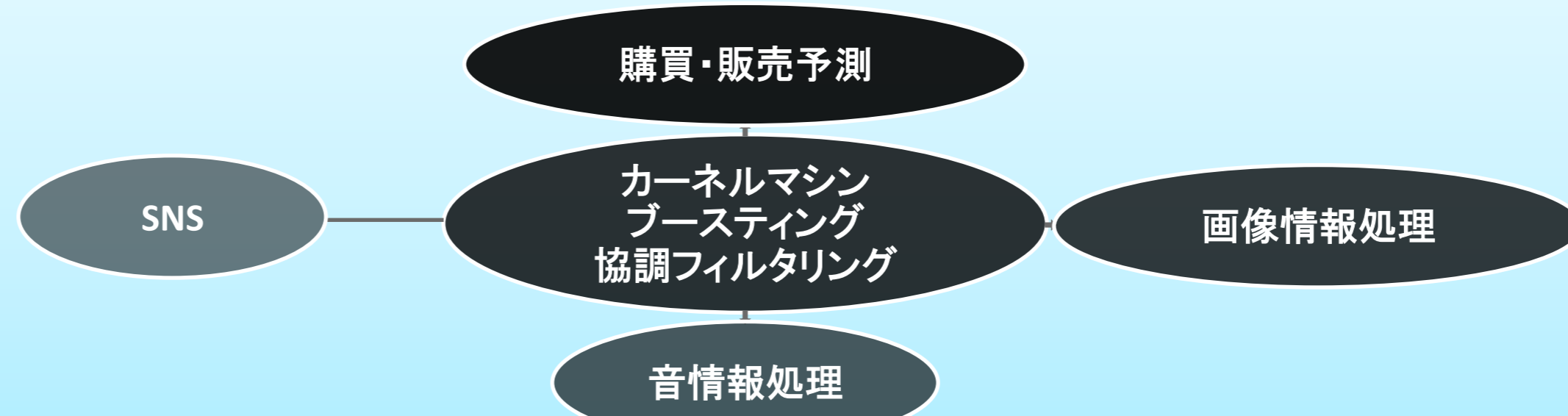
松井 知子 モデリング研究系 教授

【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
- ガウス過程状態空間モデル など

【ガウス過程状態空間モデルによる音楽感情認識】

音楽情報検索への応用

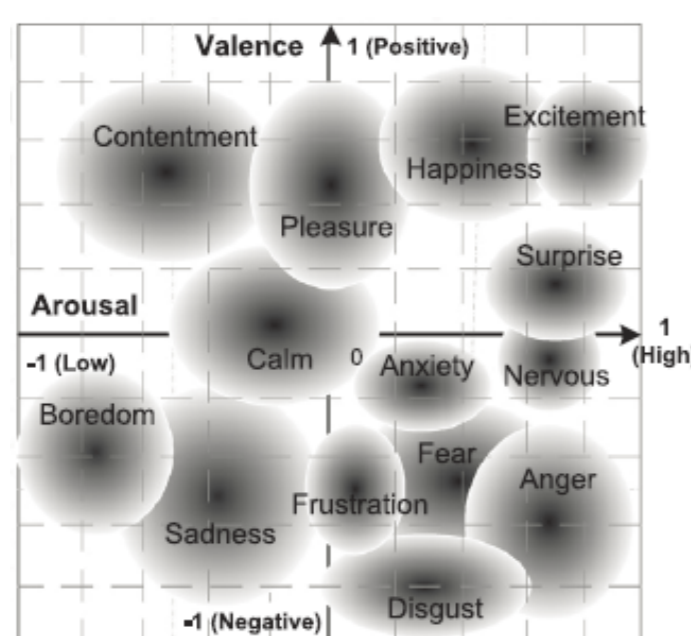
- なぜ必要か?
 - A lot of online music – fast and efficient access is demanded.
 - Intelligent music services – dynamic / personalized playlist generation, recommendations, etc.
- 主なタスク
 - Musical genre / mood recognition.
 - Artist / album recognition.
 - Music transcription.
 - Music structure analysis,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



MediaEval2014: Subtask focus

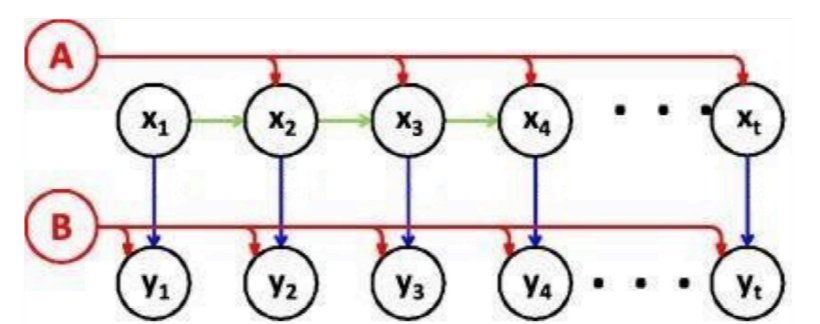
- Subtask 1: Feature development
 - Static affect prediction
 - New features
 - Signal Processing challenge
- Subtask 2: Dynamic estimation
 - Dynamic affect prediction
 - New modeling approaches
 - Machine learning challenge

Affect trajectory

- Dynamic emotion recognition is estimation of the **affect trajectory in time!**
- Apply *time series analysis* tools:
- Trajectory estimation is a **time series filtering/smoothing** task.
- Well suited are the **State-Space Models (SSM)**.

State-Space Models

- Also known as **dynamic (temporal) systems**:



$$\begin{cases} x_t = f(x_{t-1}, A) + v_t, & v_t \sim \mathcal{N}(0, R) \\ y_t = g(x_t, B) + v_t, & v_t \sim \mathcal{N}(0, Q) \end{cases}$$

Gaussian Filtering

- It can be shown that: (Deisenroth, 2011)

$$p(x_t | y_{1:t}) = \mathcal{N}(x_t | \mu_t^x, \Sigma_t^x), \quad \text{where}$$

$$\begin{aligned} \mu_t^x &= \mu_t^{x, pred} + \Sigma_t^{x,y} (\Sigma_t^y)^{-1} (y_t - \mu_t^y) \\ \Sigma_t^x &= \Sigma_t^{x, pred} - \Sigma_t^{x,y} (\Sigma_t^y)^{-1} (\Sigma_t^{x,y})^T \end{aligned}$$

$$p(x_t | y_{1:t-1}) = \mathcal{N}(x_t | \mu_t^{x, pred}, \Sigma_t^{x, pred})$$

Kalman Filter (KF)

- Linear state-space model:

$$\begin{cases} x_t = Fx_{t-1} + v_t \\ y_t = Gx_t + v_t \end{cases}$$

- Advantages:
 - Analytic approximation to $p(x_t | y_{1:t})$
 - Fast
- Disadvantages:
 - Linearity assumption

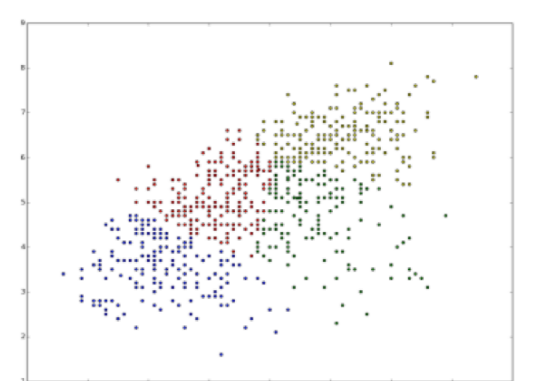
Gaussian Process SSM (GP-SSM)

- Gaussian Processes based state-space model:

$$\begin{cases} x_t = f(x_{t-1}) + v_t, & f(x) \sim \mathcal{GP}(0, K_f) \\ y_t = g(x_t) + v_t, & g(x) \sim \mathcal{GP}(0, K_g) \end{cases}$$
- Advantages:
 - Non-linear, Non-parametric
 - Flexible.
- Disadvantages:
 - No standard algorithms for training and inference.
 - Analytic moment matching approximation (Deisenroth, 2012)
 - Computationally expensive.

Experiments

- Feature extraction.
 - Marsyas tool.
 - mfcc – MelFreq Cepstral Coefficients
 - spfe – zero-cross, spectral flux, centroid, rolloff.
 - **baseline** – Features used in the official baseline system.
- Independent* state and observation model learning.
 - Multivariate linear regression for the KF.
 - GP regression learning for the GP-SSM.
- Development data.
 - Training set - 600 clips.
 - Validation set – 144 clips.
- Training set clustering.
 - Four clusters based on clips static A-V vectors.
 - Separate SSM trained for each cluster.
 - Maximum likelihood based model selection.



Results on Test Data

Feature	Kalman filter		GP-SSM filter	
	R	RMSE	R	RMSE
AROUSAL – MULTIPLE MODELS				
mfcc	0.1914	0.0852	0.044	0.1089
spfe	0.0526	0.1986	0.1015	0.2066
baseline	-0.1520	0.3824	0.0301	0.2073
VALENCE – MULTIPLE MODELS				
mfcc	-0.065	0.1590	-0.017	0.1096
spfe	-0.075	0.2679	-0.023	0.1920
baseline	-0.099	0.2325	-0.049	0.2267

Conclusions

- KF vs. GP-SSM
 - Similar performance under similar conditions.
 - GP-SSM a bit better for Valence estimation.
- Feature types.
 - Baseline features seem to perform well.
 - No definite winner.

共同研究者: K. Markov氏 (会津大)、G. Peters氏 (UCL)