

統計的機械学習による音声/音楽/画像/WEBユーザビリティ 情報解析に関する研究

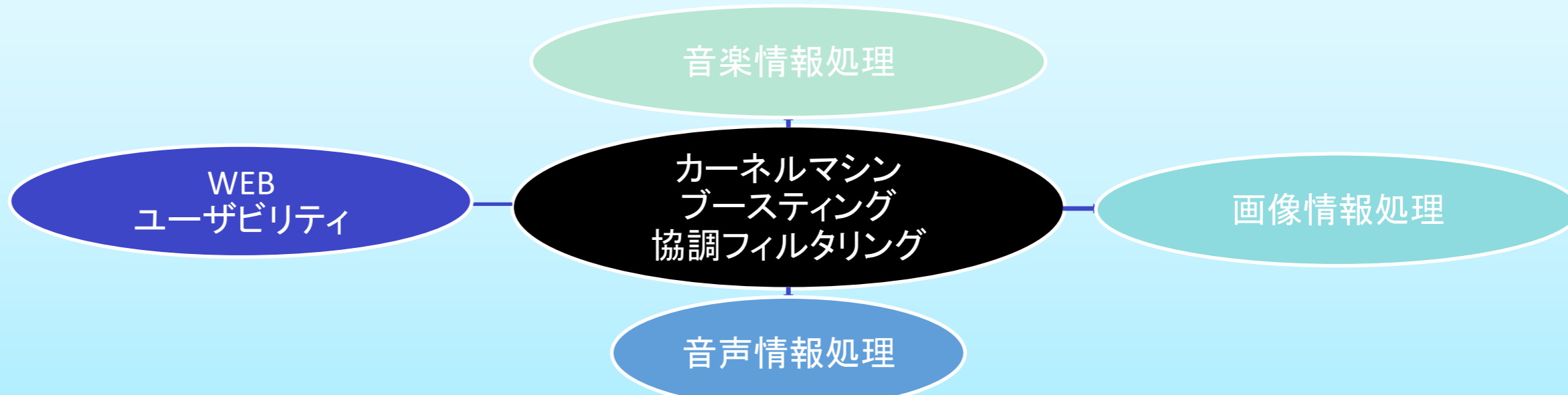
松井 知子 モデリング研究系 教授

【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/WEBユーザビリティデータ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. 画像識別
3. テキスト分類
4. ピアノ演奏の自動採譜
5. 音楽ジャンル分類
6. WEBユーザビリティ評価値の補完

などの研究課題に取り組んでいます。



【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

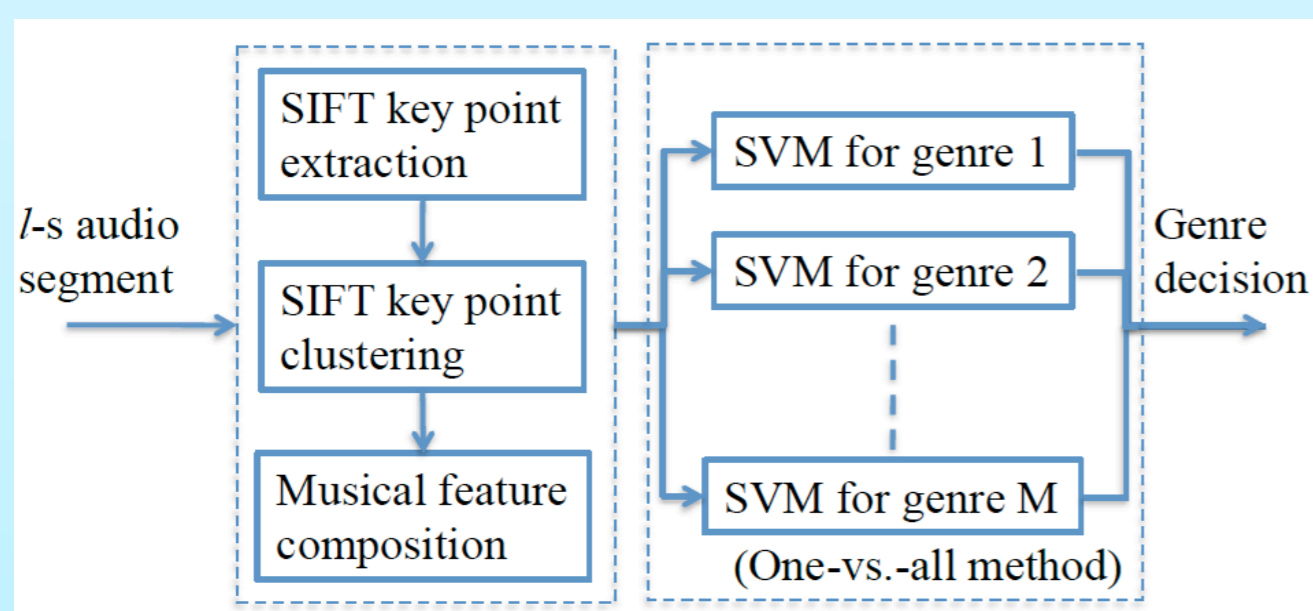
【SVMによる音楽ジャンル分類】

方針:

- 音楽セグメント(5秒程度の長さ)について、テンポ非依存、周波数非依存にスペクトル上の動的な特徴を捉える。
- SIFT (scale-invariant feature transform)を用いて、
 - ✓ テンポ非依存性を特徴的な点(クラスタID)の連鎖情報として表すことで実現する。
 - ✓ 周波数非依存性を局所的な勾配情報として表すことで実現する。
- 高次元の音楽特徴量(例えばクラスタ数(10³)の連鎖数(3)乗=10⁹次元)を扱うためにSVMを利用する。

手続き:

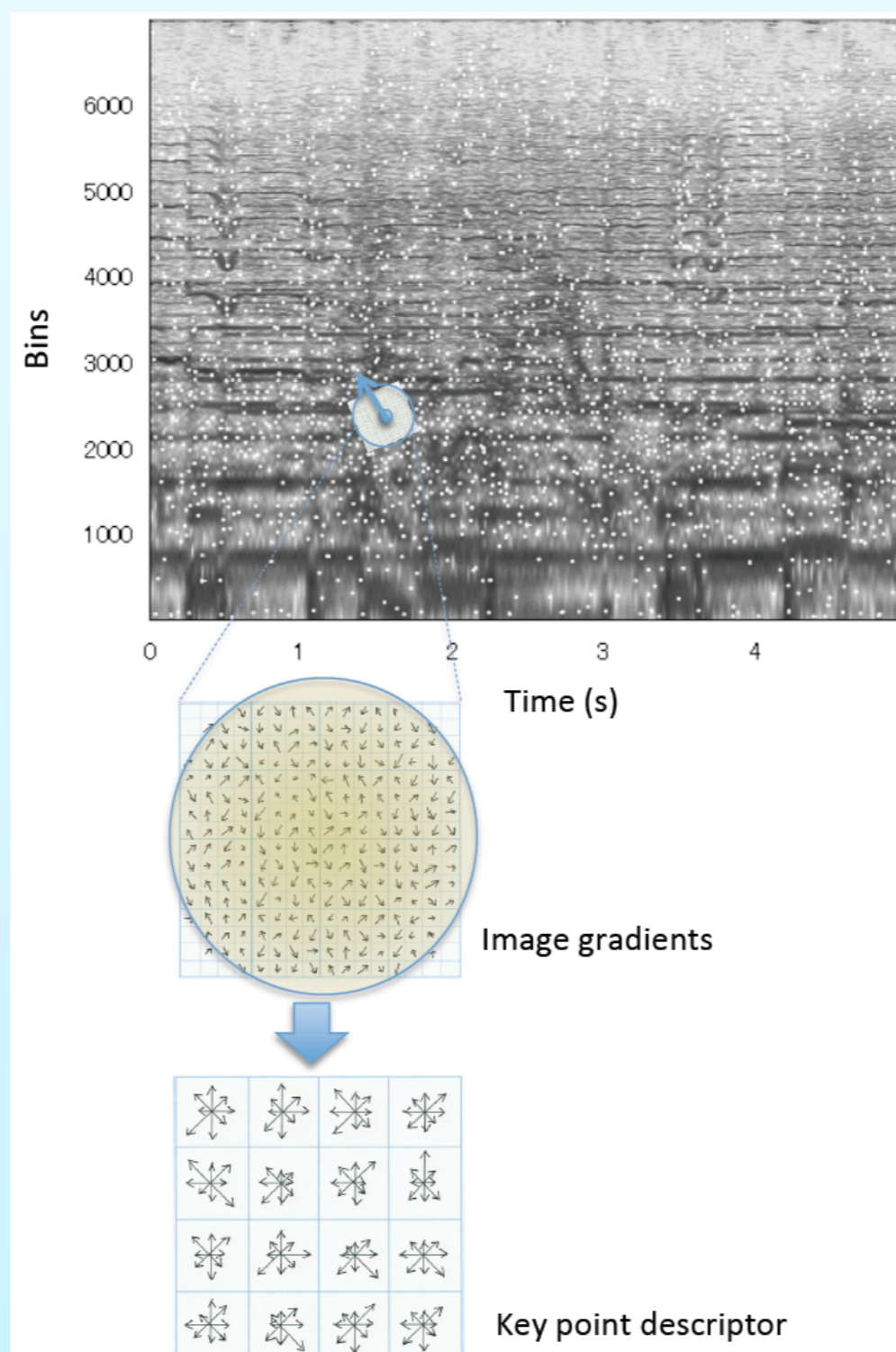
- 各音楽セグメントをスペクトログラムの画像で表現
- SIFT特徴点(4x4x8=128次元で表現)の抽出[Lowe, 2004]
- SIFT特徴点のクラスタリング(クラスタ数は10³程度)
- クラスタIDで表されたSIFT特徴点の一連鎖、二連鎖、三連鎖の頻度情報による音楽特徴量の構成
- SVMの適用



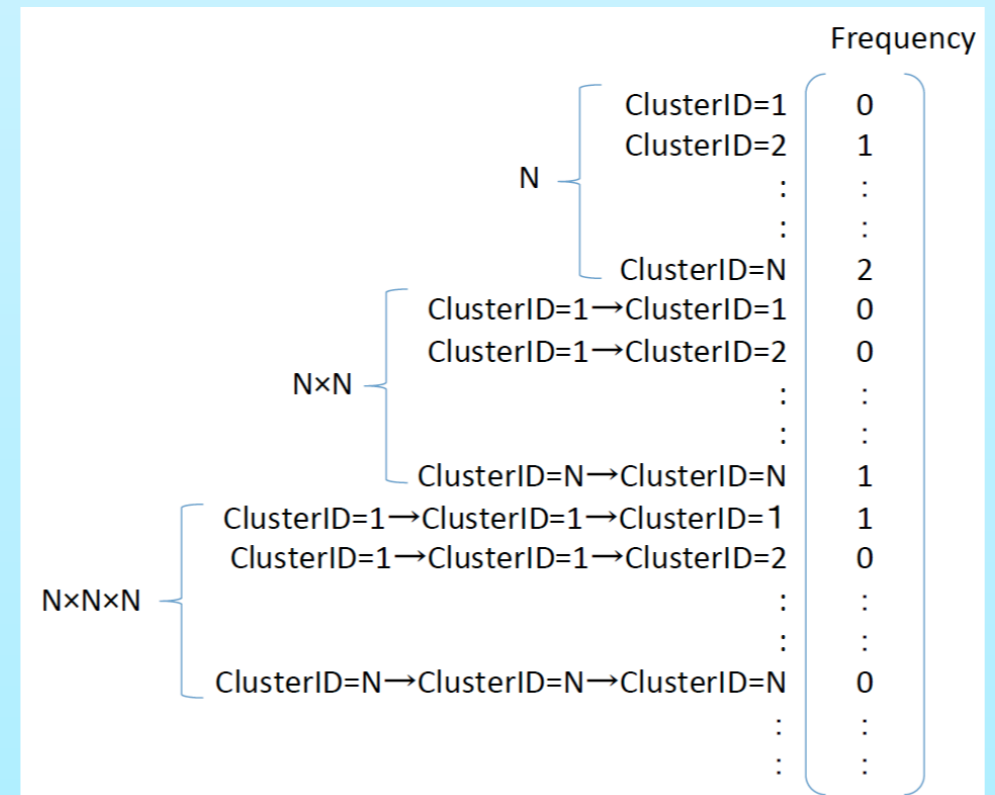
音楽ジャンル分類の手続き

性能:

- ISMIR2004の音楽ジャンル分類タスク用データセットを用いて評価したところ、従来の短時間(32ミリ秒)の音楽セグメントから抽出したMFCC特徴量を用いたGMMによる分類法を利用した場合と比べて、ジャンルの種類に寄らず高精度で分類できた。
- テンポを元の音楽データの1/4, 1/3, 2倍にしたり、音の高さを半オクターブ低くしたり、高くしたデータを用いて評価したところ、従来法と比べて、比較的安定して音楽ジャンル分類できることを確認した。



音楽“rock_pop”セグメントから抽出したSIFT特徴点と局所的な勾配情報を表すSIFT特徴点のイメージ



クラスタIDで表したSIFT特徴量の一連鎖、二連鎖、三連鎖の頻度情報で構成される音楽特徴量

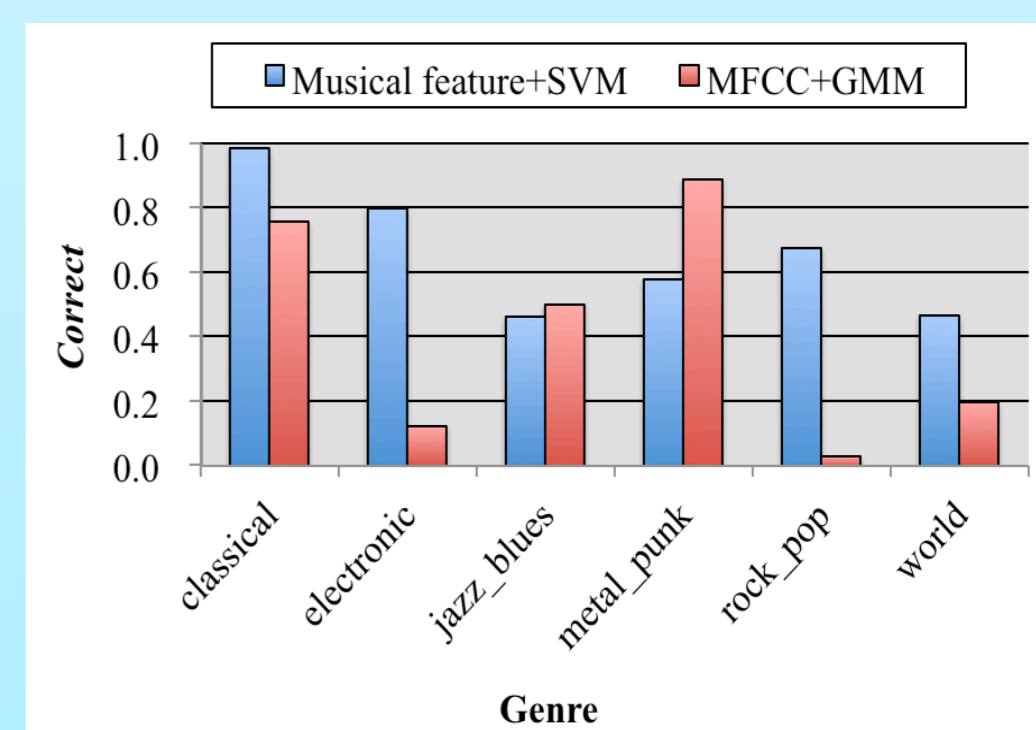
Genre	No. of samples	No. of images	No. of SIFT key points
classical	[201.0]	320	3200
electronic	[327.9]	115	1150
jazz_blues	[230.0]	26	260
metal_punk	[251.4]	45	450
rock_pop	[226.1]	101	1010
world	[351.6]	122	1220
Total/Average	Average [253.8]	Total 729	Total 7290
		Average 2161.8	

ISMIR2004の音楽ジャンル分類タスク用データセットに含まれるサンプル数、利用したセグメント数、平均のSIFT特徴点数

No. of clusters	100	200	300	500	1000
L1	0.711	0.781	0.779	0.778	0.775
L2	0.763	0.796	0.801	0.812	0.823
L3	0.785	0.796	0.804	0.811	0.827

SIFT特徴点を表すためのクラスタ数を変えた時の音楽ジャンル分類の正解度

- L1: 一連鎖の頻度情報のみで音楽特徴量を構成した場合
- L2: 一連鎖、二連鎖の頻度情報を利用した場合
- L3: 一連鎖、二連鎖、三連鎖の頻度情報を利用した場合



従来のMFCC+GMMの方法との音楽ジャンル別の正解度

【グラフカーネルによる画像識別】

方針:

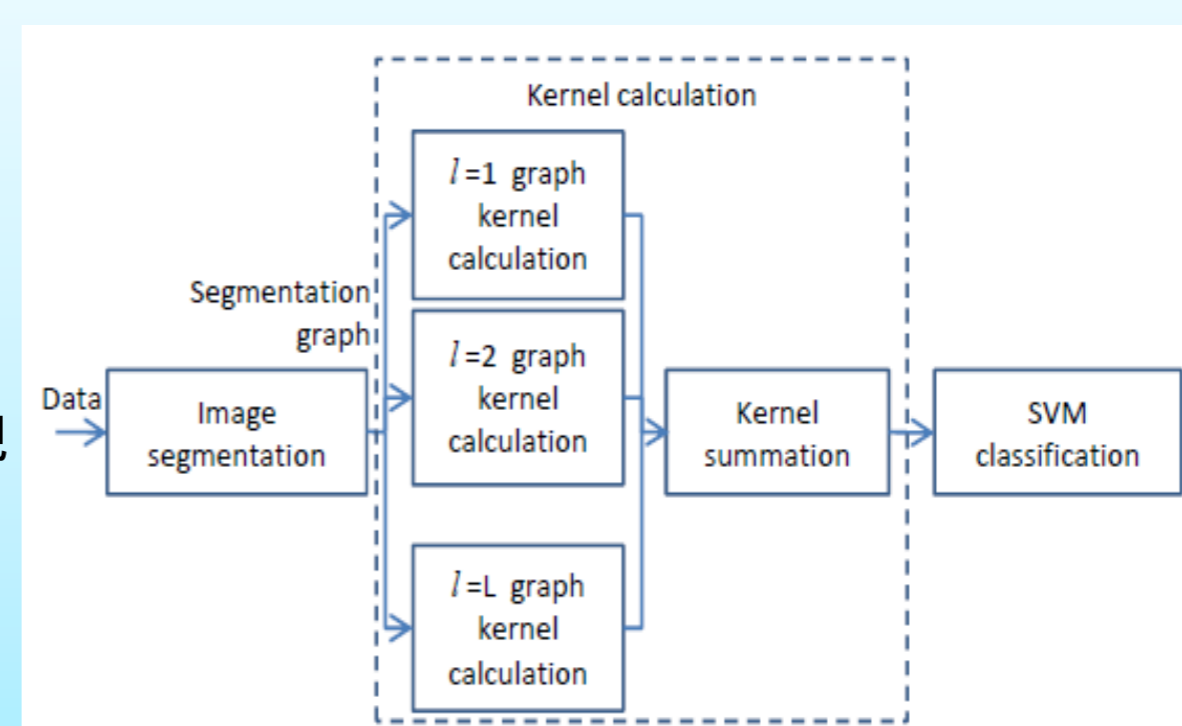
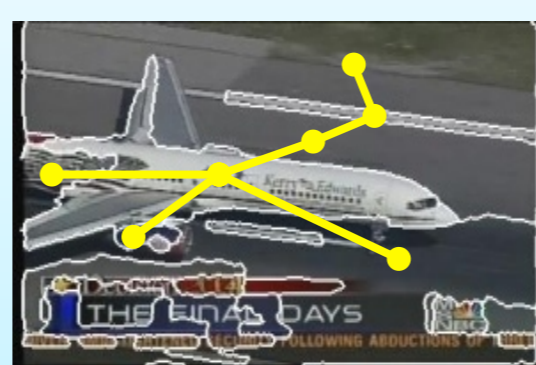
- 画像の部分的な幾何情報(グラフで表現)を利用して性能向上をはかる。
- グラフカーネル+SVMを利用する。

手続き:

- 画像を自動セグメント化 [Deng et al., 2001]
- セグメントを頂点、隣接するセグメント間を辺として扱い、画像中の各セグメントの隣接関係をグラフとして表現
- グラフカーネル[Mahe et al., 2005]を拡張し、グラフ表現した画像間の類似度をはかる => SVMへ

性能:

- MediaMillベンチマークデータセットでは、従来法と比べて58%の性能向上



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!