音声・音楽・映像・テキストデータの判別予測方式に関する研究

モデリング研究系　知能情報モデリンググループ

教授　　松井　知子

1　はじめに

本年度は、機能と帰納プロジェクトのサブプロジェクト「マルチモーダルデータからの不変情報の発見とその方法論の研究」において PrefixSpan based subsequence boosting (pboost) [Nowozin(2007)]やグラフカーネルを利用した映像分類に関する研究[Vert(2009), Matsui(2009)]を、奈良先端大の学生研究指導を通じてカーネルマシンを利用したタスク外発話検出[藤田(2009a. 2009b, 2010)]やトピックス分類[Torres(2009a, 2009b, 2010)]に関する研究を、総研大の学生研究指導を通じて話者認識のためのカーネル法[Yamada(2009, 2010a, 2010b)]、カーネルマシンを利用したピアノ演奏の音高推定[今村(2009)]および Web ユーザビリティの低いページの検出[山田(2009)]に関する研究を、情報研との共同研究において音声コーパスの可視化の研究[菊池(2009), Yamakawa(2009)]を行った。さらに融合シーズ探索プロジェクト「複雑な人間・社会データから「構造」を発見するための方法論に関する研究」において構造学習の研究を、国立国語研究所の「コーパス日本語学の創成」プロジェクトに参加してコーパスを利用した音声・対話研究の可能性を幅広く探る研究を開始した。以下、pboost を利用した映像分類の研究を取り上げて紹介する。

2　pboostを利用した映像分類

現在、インタネット上を含め、いろいろな映像データが大量に利用できるようになっている。それらの映像データを有効に活用するためには高度な映像検索技術が不可欠である。本研究ではその要素技術として、ある映像ショット（一連の画像系列）に特定の概念が含まれているかどうかを判定する、映像ショット分類の方法について検討した。

一般に映像ショット分類では、計算コストを考慮して、そのショットを代表する画像（キーフレーム）を一枚（もしくは数枚）選択することにより、画像系列分類を画像分類の問題に簡略化して扱うことが多い。これまで、Support Vector Machine (SVM)や Neural Network (NN)に基づくいろいろな画像分類の方法が報告されている。筆者らも昨年度は、画像の部分的な空間構造をグラフカーネルによってモデル化する方法について検討を行った[Vert(2009)]。本年度はもともとの画像系列分類の問題を簡略化せずに、可変長の映像ショットの時間構造を pboost によってモデル化することを試みた。

pboost は系列パターンの列挙問題のためのアルゴリズム PrefixSpan を拡張したものである。PrefixSpan が頻度の高い系列パターンを抽出するのに対して、pboost では識別的な系列パターンを抽出する。TRECVID データ[Matsui(2009)]を用いた小規模の分類実験（分類対象は 6 つの概念）において、pboost により、各ショットから「飛行機」などの動的な概念について、識別に有効な部分画像の系列パターンを抽出することができ、従来の SVM に基づく方法と比べて同程度の性能が得られることがわかった。なお、pboost については pboost toolbox として公開されているソフトウェアを基本的に利用した(http://www.kyb.mpg.de/bs/people/nowozin/pboost/)。本実験で用いたデータでは正例の数が負例の数よりも極端に少ない（正例の数は負例の数の 0.3～4%程度）。pboost のパラメータを頑健に推定するために、正例と負例の数に応じて損失関数に重み付けをするなどの改良を行った。

今後は、pboost で抽出した識別的な有効な部分画像の系列パターンを可視化するなどして分析を行い、分類方法の高精度化をはかっていきたい。

参考文献

Øystein Birkenes, Tomoko Matsui, Kunio Tanabe, Sabato Marco Siniscalchi, Tor Andr´e Myrvoll, and Magne Hallstein Johnsen (2010). Penalized Logistic Regression with HMM Log-Likelihood Regression for Speech Recognition, *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING* (to appear).

Makoto Yamada, Masashi Sugiyama, and Tomoko Matsui (2010a). Semi-supervised Speaker Identification under Covariate Shift, *Signal Processing* (to appear).

今村武史, 松井知子 (2010). SVM による頑健なピアノ演奏の音高推定, *電子情報通信学会大会予稿集*.

山田俊哉, 中道上, 松井知子 (2009). パターン認識手法による Web ユーザビリティの低いページの検出, *日本計算機統計学会第 23 回シンポジウム予稿集*.

Makoto Yamada, Masashi Sugiyama, and Tomoko Matsui (2009). Covariate shift adaptation for semisupervised speaker identification, *Proc. 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Makoto Yamada, Masashi Sugiyama, Gordon Wichern, and Tomoko Matsui (2010b). Acceleration of Sequence Kernel Computation for Real-time Speaker Identification, *Proc. 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Jean-Philippe Vert, Tomoko Matsui, Shin'ichi Satoh, and Yuji Uchiyama (2009). HIGH-LEVEL FEATURE EXTRACTION USING SVM WITH WALK-BASED GRAPH KERNEL, *Proc. 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Tomoko Matsui, Shin'ichi Satoh, and Yuji Uchiyama (2009). ISM TRECVID2009 High-level Feature Extraction, *Proc. TRECVID 2009 Workshop*.

Rafael TORRES, Shota TAKEUCHI, Hiromichi KAWANAMI, Tomoko MATSUI, Hiroshi SARUWATARI, and Kiyohiro SHIKANO (2009a). Inquiry Classification in a Speech-Oriented Guidance System Using Discriminative Learning, *情報処理学会研究報告*.

藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏 (2009a). 音声情報案内システムにおける SVM を用いたタスク外発話の検出, *情報処理学会研究報告*.

トーレス・ラファエル, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏 (2009b). Comparison of Discriminative Learning-Based Inquiry Classification Methods for a Speech-Oriented Guidance System, *日本音響学会 2009 年秋季研究発表会予稿集*.

藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏 (2009b). SVM を用いたタスク外発話検出における特徴量の組み合わせに関する検討, *日本音響学会 2009 年秋季研究発表会予稿集*.

トーレス・ラファエル, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏 (2010). PrefixSpan Boosting-Based Inquiry Classification for a Speech-Oriented Guidance System, *日本音響学会 2010 年春季研究発表会予稿集*.

藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏 (2010). Bag-of-Words を用いた SVM による無効発話の棄却, *日本音響学会 2010 年春季研究発表会予稿集*.

菊池英明, 沈睿, 山川仁子, 松井知子, 板橋秀一 (2009). 音声言語コーパスの類似性可視化システムの構築, *日本音響学会 2009 年秋季研究発表会予稿集*.

Kimiko Yamakawa, Hideaki Kikiuchi, Tomoko Matsui, and Shuichi Itahashi (2009). Utilization of Acoustical Feature in Visualization of Multiple Speech Corpora, *Proc. Oriental COCOSDA 2009*.

Sebastian Nowozin, Gökhan BakIr, and Koji Tsuda (2007). Discriminative Subsequence Mining for Action Classification, *ICCV 2007*.