FUSION 2008 Tutorial Proposal: Fundamentals of the Class-Specific Method

Dr. Paul M. Baggenstoss
Naval Undersea Warfare Center
Newport RI, 02841
401-832-8240 (TEL)
p.m.baggenstoss@ieee.org

January 15, 2008

Abstract

The class-specific method (CSM) is an approach for constructing classifiers with class-dependent features that blends signal processing with classifcation theory. CSM is based on the mathematical identity called the probability density function (PDF) projection theorem that extends classical Bayesian theory. In contrast to conventional classifiers, feature extraction is an integral part of the theory. Because CSM does not need a common feature space, the dimensionality curse can be avoided. The tutorial covers fundamentals including generative and discriminative classiifers, the PDF projection theorem, class-specific modules and the chain rule. Many intuitive examples are providded. Some advanced examples are also covered.

1 Overview

The class-specific method (CSM) is emerging as an important new generative method in signal modelling and classification. Numerous papers (see bibliography below), a tutorial article[1], and a conference tutorial [2] have appeared on CSM.

2 Background

While the ultimate goal of all classifiers is to distinguish the various classes, the difference between generative and discriminative classifiers is the approach taken. While discriminative classifiers achieve the goal directly through construction of decision boundaries in a common feature space, generative classifiers achieve the goal indirectly by statistically modelling each class. Each method has a good argument: proponents of discriminative classifiers argue that it is better to estimate decision boundaries directly while proponents of generative classifiers argue that generative classifier are embodiments of the optimal Bayesian classifier and by fully modelling each class, there is a better chance of rejecting new unseen data types. The best argument, however, is that the best classifiers contain elements of both generative *and* discriminative classifiers.

Recently, most attention has been paid to discriminative methods. The treatment of generative methods is inadequate, not just from the point of view of attention, but from the point of view that most treatments of generative classifiers are in the context of a common feature space. Within the constraints of a common feature space, discriminative classifier are often better. However, with the introduction of the PDF projection theorem (PPT) in 2000 [4], [5], [23], the situation has changed for generative classifiers. The constraint of living within a common feature space has been removed and each data class can be modelled using a dedicated signal processing chain.

The class-specific method (CSM) blends signal processing with classification theory using the theoretical foundation of the PDF projection theorem [5]. At times alone, but mainly when combined with discriminative methods, CSM can produce substantial performance improvements. The so-called curse of dimensionality plagues conventional generative and discriminative classification methods because when restricted to a common feature space, it is necessary to seek a compromise between conflicting goals. If the feature dimension is too low, critical information is lost. If the feature dimension is too high, estimation of PDFs and decision boundaries are plagued by the dimensionality issues.

In CSM, because a given feature set does not need to represent all classes, the features are required only to distinguish a given class from a reference hypothesis that can be chosen separately for each class. As new classes are added, no changes are required for existing models. While many methods exist for class-dependent feature extraction, [6],[7], [8], none except CSM are based on a rigid general theory that is directly tied to the optimal Bayesian classifier. Thus, they are subject to approximations and restrictions, whereas CSM is a generalization of classical theory. Because CSM is a generative approach (focused on modeling each class) and not a discriminative method (focused on discriminating classes), it can operate hand in hand with discriminative methods to produce performance better that either method alone.

Because CSM is signal-modelling intensive, it facilitates intuitive monitoring of performance at every stage through signal synthesis and statistical PDF model synthesis. The promise of improved performance comes at a price. CSM requires careful attention to signal and PDF modelling and careful balancing of numerical and statistical errors. Thus, there is a substantial learning curve for first-users.

3 Tutorial

3.1 Brief Description

This tutorial is designed to to inform the audience about the fundamentals of CSM. The examples ranging from time-series and spectral modeling to speech signal modeling are provided to give the audience a good feel for what kinds of problems would benefit from CSM how to apply it when it is. It is designed to provide information at enough detail to be of interest to researchers in the field of signal processing, pattern recognition, speech recognition, and related fields. The introduction and basic examples are at a high-enough level to make the talk accessible to audience members with a general interest in finding out what is CSM.

3.2 Tutorial Outline

1. **Background and Theory.**
   (a) Discriminative and Generative classifiers.
   (b) Conventional classifier and the Curse of Dimensionality.
   (c) PDF projection theorem.
   (d) Fixed and floating reference hypothesis.
   (e) The chain rule.
   (f) The class-sepific module.
2. **Practical Design issues.**
   (a) Choosing reference hypotheses.
   (b) Choosing features.
   (c) Signal re-synthesis from features.
   (d) Signal re-synthesis from PDF.
   (e) Hybrid Discriminative-CSM classifier.
3. **Class-Specific Model Examples.**
   (a) AR, MA, ARMA modeling.
   (b) Spectral Modelling.
   (c) Order statistics.
4. **Working classification examples.**
   (a) Synthetic signals.
   (b) Speech signals.
5. **Advanced Concepts.**
   (a) Class-specific iterated subspace.
   (b) Multiresolution HMM.

3.3 Target Audience

Audience members with a general understanding of signal processing and classification theory will benefit from most of the tutorial. To fully absorb the material, a familiarity with advanced statistical methods such as hidden Markov models (HMM), Gaussian mixtures, and a basic understanding of decision theory and probability density functions (PDFs) is required.

3.4 Interest to FUSION 2008 Audience

Because CSM is a blend of signal processing and classification, it is of interest to a wide group within FUSION 2008.

4 About The Author

The author has worked in sonar signal processing since 1979 at Raytheon corporation in Portsmouth, RI, USA, and has been with the Naval Undersea Warfare Center (NUWC) in Newport, RI, USA since 1996. At NUWC, Dr. Baggenstoss has worked exclusively in statistical signal classification and modeling. He has published a wide range of conference papers [4], [9], [10], [3], [11], [12], [13], [14], [15], and journal articles [16], [17], [18], [19], [20], [21], including landmark articles in classification theory [22], [23], [5], [24]. Related articles published by Steve Kay [25], [26] are also notable.

Dr. Baggenstoss has worked as an adjunct professor at the university of connecticut, Storrs, where he taught detection theory and digital signal processing. In February 2000, he began a 1-year joint research effort with the Pattern Recognition Chair, University of Erlangen, Erlangen, Germany. The author lives in Newport, RI, USA, speaks fluent German and has working knowledge of French and Spanish.

Dr. Baggenstoss published a tutorial article on CSM [1] and has delivered a conference tutorial on CSM in 2003 [2]: `http://www.oceanicengineering.org/main.cfm?id=147&r1=5.00&r2=3.05&level=2`

4.1 Contacting the Author

Dr. Paul M Baggenstoss
PO Box 2651

Newport, RI, 02840, USA
Tel: 001-401-832-8240
Fax: 001-401-832-7453
Email: p.m.baggenstoss@ieee.org
Email: pbaggens@verizon.net

References

[1] P. M. Baggenstoss, "The class-specific classifier: Avoiding the curse of dimensionality (tutorial)," *IEEE Aerospace and Electronic Systems Magazine, special Tutorial addendum*, vol. 19, pp. 37–52, January 2002.

[2] P. M. Baggenstoss, "Target classification architectures: The class-specific method," in *OCEANS 2003 Turorial T14*, 2003.

[3] P. M. Baggenstoss and T. E. Luginbuhl, "An E-M algorithm for joint model estimation.," in *Proceedings of the 1999 International conference on Acoustics, Speech, and Signal Processing (ASSP)*, pp. 1825–1828, 1999.

[4] P. M. Baggenstoss, "A theoretically optimum approach to classification using class-specific features.," *Proceedings of ICPR, Barcelona*, 2000.

[5] P. M. Baggenstoss, "The PDF projection theorem and the class-specific method," *IEEE Trans Signal Processing*, pp. 672–685, March 2003.

[6] H. Watanabe and S. Katagiri, "Hmm speech recognizer based on discriminative metric design," in *Proc. 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP97)*, vol. 4, 1997.

[7] M. K. Omar and M. Hasegawa-Johnson, "Strong-sense class-dependent features for statistical recognition," in *Proc. IEEE Workshop on Statistical Signal Processing*, vol. 4, pp. 490–493, 2003.

[8] A. Bailey, *Class-dependent features and multicategory classification*. PhD thesis, University of Southampton, 2001.

[9] P. M. Baggenstoss, "The chain-rule processor: Optimal classification through signal processing," *Proceedings of ICPR, Quebec*, August 2002.

[10] P. M. Baggenstoss, "A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces.," *Proceedings of ICASSP, Istanbul*, 2000.

[11] P. M. Baggenstoss, "Class-specific features in classification.," in *IASTED International Conference on Signal and Image Processing*, 1998.

[12] P. M. Baggenstoss, "Structural learning for classification of high-dimensional data," *Proceedings of the 1977 International Conference on Intelligent Systems and Semiotics (ISAS 97)*, pp. 124–129, 1997.

[13] P. M. Baggenstoss, "Improved echo classification in shallow water," in *OCEANS 96 MTS/IEEE Proceedings*, pp. 1197–1203, IEEE, 1996.

[14] P. M. Baggenstoss, "Structural learning for classification of high dimensional data," in *Proceedings of the 1997 International Conference on Intelligent Systems and Semiotics*, pp. 124–129, National Institute of Standards and Technology, 1997.

[15] P. M. Baggenstoss, "Class-specific feature sets in classification," in *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC), 1998*, pp. 413–416, National Institute of Standards and Technology, 1998.

[16] P. M. Baggenstoss and S. M. Kay, "An adaptive detector for deterministic signals in noise of unknown spectra using the Rao test," *IEEE Trans. SP*, vol. 40, no. 6, pp. 1460–1468, 1992.

[17] P. M. Baggenstoss and S. M. Kay, "Detection of broadband planewaves in the presence of Gaussian noise of unknown covariance: Asymptotically optimum tests using 2-d autoregressive noise model," *IEEE Trans. SP*, vol. 43, pp. 950–966, April 1995.

[18] P. M. Baggenstoss, "Improved active classification incorporating spatial energy distribution of reverberation and environmental adaptation.," *U.S. Navy Journal of Underwater Acoustics (Accepted for publication)*, 1997.

[19] P. M. Baggenstoss, "On detecting linear frequency modulated waveforms in frequency and time dispersive channels: Alternatives to segmented replica correlation.," *IEEE Trans of Oceanic Engineering*, vol. 19, pp. 591–598, October 1994.

[20] P. M. Baggenstoss, "Image distortion analysis using polynomial series expansion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[21] P. M. Baggenstoss, "Adaptive pulselength correction (APLECORR): A strategy for waveform optimization in ultrawideband active sonar.," *IEEE Trans of Oceanic Engineering*, vol. 23, pp. 1–11, January 1998.

[22] P. M. Baggenstoss, "Class-specific features in classification.," *IEEE Trans Signal Processing*, pp. 3428–3432, December 1999.

[23] P. M. Baggenstoss, "A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces.," *IEEE Trans. Speech and Audio*, pp. 411–416, May 2001.

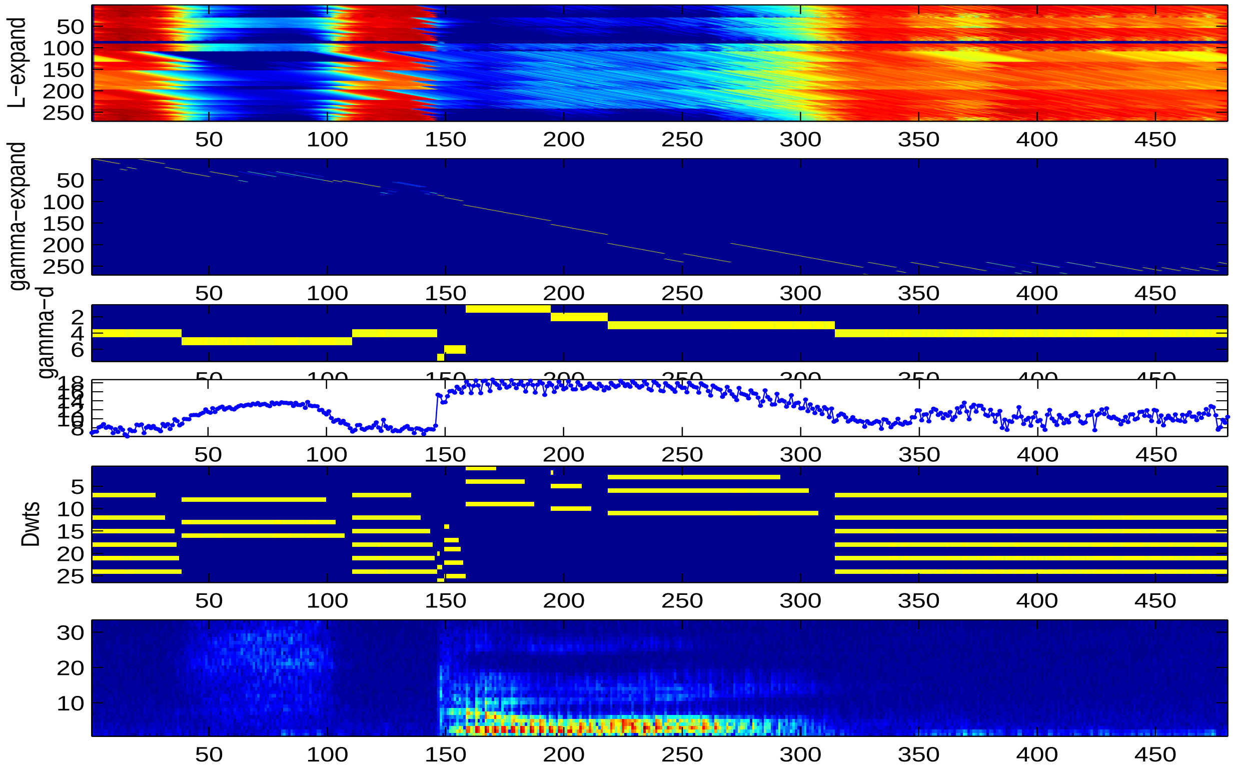

Figure 1: This figure shows how CSM can be applied to speech analysis. Pictured is the word "stool" analyzed by the multi-resolution hidden Markov model (MR-HMM). The MR-HMM is a means of processing data at multiple resulutions, yet combining the likelihood functions from various processing lengths into a log-likelihood function for the full raw data record. The MR-HMM computes the likelihood function of the raw data averaged over all possible paths through the state trellis. Each state is constrained by a minimum number of consecutive visits to the state, thereby controlling the processing length of each state. Use of multiple processing lengths is important in speech applications where long vowel sounds ("oo" in "stool") exist together with short sounds (the "t" in "stool"). The CSM's PDF projection theorem is the theoretical key to the MR-HMM.

[24] S. M. Kay, A. H. Nuttall, and P. M. Baggenstoss, "Multidimensional probability density function approximation for detection, classification and model order selection," *IEEE Trans. Signal Processing*, pp. 2240–2252, Oct 2001.

[25] S. Kay, "Sufficiency, classification, and the class-specific feature theorem," *IEEE Trans. Information Theory*, vol. 46, pp. 1654–1658, July 2000.

[26] S. Kay, "Conditional model order estimation," *IEEE Trans. Signal Proc.*, vol. 49, pp. 1910–1917, Sept 2001.