Sound Recognition

Jonas Papmeier

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik, Arbeitsbereich TAMS
Project Intelligent Robitics

22. April 2015

Gliederung

General Introduction

Goal, applications

Goal

extracting information, attach "semantic" meaning to audio signal enabeling reasoning and decision-making [Sch13, p. 3f]

applications

- activity recognition
- monitoring/surveilance
- human-robot interaction

General Introduction

historical origin, related fields

related fields

- historical origion: speech recognition
- Computational auditory scene analysis
- Sound detection
- Sound source seperation

Motivation

Why is (general) sound useful despite other sensor modalities?

state indicators may:

- not be visual detectable
- be outside of field of view

example kitchen environment
state indicators: alarms, bells, buzzers, water boiling [KMSW05]

Definition of Problem: Sound Recognition

Classification of Sound Events

given a sound event
determine to which class within the set of classes the sound event belongs
Classification Problem

General steps

- Feature Extraction/Reduction/Generation
- oftenl Mel-frequency cepstral coefficients (MFCC)
- learning phase
- utilizes machine learning techniques from pattern recognition (HMM, SVM)
- use learned model to classify incomming sound events
- (evaluate performance on test data)

Simplified classification

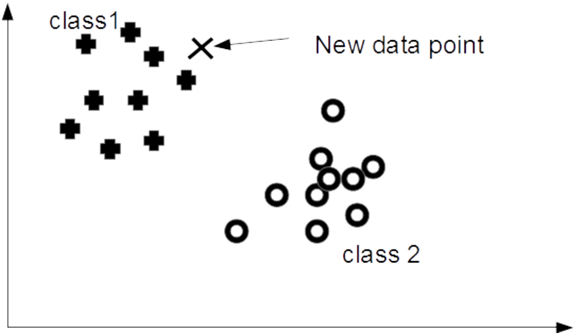

Abbildung: simple classification problem: crosses class 1, circles class 2, given new data point

Example Approach

Paper: Temporal ICA for Classification of Acoustic Events in a Kitchen Environment

Title: Temporal ICA for Classification of Acoustic Events in a Kitchen Environment
Authors: Florian Kraft, Robert Malkin, Thomas Schaaf, Alex Waibel
published: Interspeech Conference 2005
Scenario: Robot assists in kitchen, uses sound cues to identify states

Example Approach

Used Classes/ Sounds to be identified

class	# training ex. (dur. in sec)	# test ex. (dur. in sec)	total # ex. (dur. in sec)
boiling	221 (662)	98 (319)	319 (981)
bread_cutter	25 (40)	11 (27)	36 (67)
cutting_vegetables	134 (89)	58 (41)	192 (130)
door	114 (101)	50 (44)	164 (144)
door_bell	50 (110)	22 (55)	72 (164)
egg_timer_ring	11 (34)	6 (17)	17 (51)
footsteps	240 (140)	104 (66)	344 (206)
lighter	84 (42)	37 (20)	121 (61)
match	141 (131)	62 (59)	203 (189)
microwave_beep	110 (30)	49 (17)	159 (47)
others	858 (1130)	369 (547)	1277 (1677)
oven_switch	472 (133)	208 (60)	680 (194)
oven_timer	12 (16)	6 (8)	18 (24)
overboiling	186 (129)	81 (70)	267 (199)
pan_stove	584 (308)	256 (132)	840 (439)
pan_sizzling	107 (343)	46 (146)	153 (489)
telephone	134 (920)	63 (393)	197 (1313)
speech	125 (82)	55 (38)	180 (120)
stove_error	18 (12)	8 (5)	26 (17)
toaster	119 (92)	53 (46)	172 (138)
water	421 (1129)	184 (464)	605 (1593)
total	4166 (5670)	1826 (2573)	5992 (8243)

Abbildung: Classes and their sample counts; [KMSW05, Table 1]

Features

Feature Extraction and Selection/Reduction/Generation

- Raw audio has ”too much” information
- Feature give compact form
- ideally contain just information relevant for discrimination of the classes
- extractet over short frames
- all features represented as feature vector

Features

Approach in the Paper

- Independent Component Analysis (ICA)
- incorporates time-dependencies
- uses linear transformation of basic features over multiple frames
- represents a change over time of a spectral view

Features

Feature Vector to calculate ICA over 1 and 7 frames

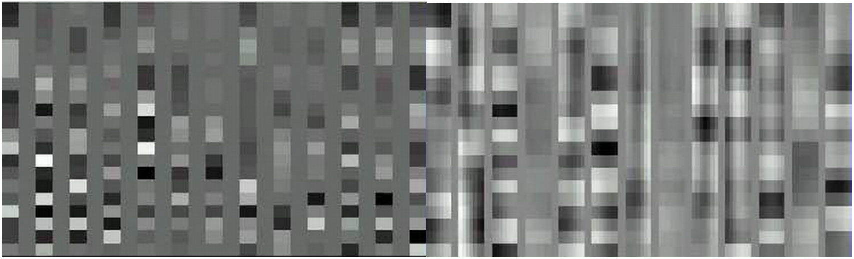

Abbildung: left: ICA over 1 Frame, vertical Mel bins, horizontal time [KMSW05, Figure 1]; right: ICA over 7 frames [KMSW05, Figure 2]

HMM

Hidden Markov Model as an example for a statistical classifier

- dynamic approach -> able to model time series
- based on statistical theory
- most frequently used [Sch13, p. 118]

HMM

The Model

- finite-state automaton
- transitions occur with probability
- states are hidden (current state probability depends only on previous state)
- but produce "observations" with probability (depends only on current state)

HMM

Simple biased coin example

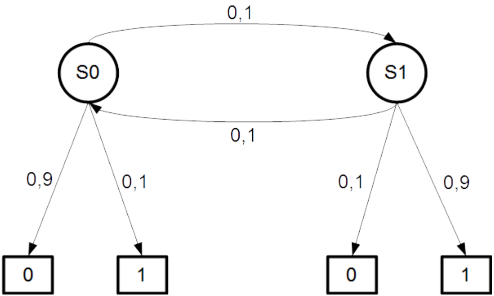

Abbildung: Simple HMM for biased coin; circle represents state, rectangle represents observation, arrows represent conditional probabilities

Example for generated sequence:
000010001111111111000001110111111

HMM

Mapping to Sound Events

- hidden states are some form of physical process
- observation is the current feature vector
- step in series of feature vectors is a time step -> transition can occur
- each class is modelled by own HMM

HMM

How to learn?

- Goal: create HMM for each class
- Training data as series of feature vectors given with classlabel
- Problem: How to find transition and observation probability?
- -> create HMM which will produce the series of feature vectors with high probability
- (Baum-Welch Forward-Backward Algorithm)

HMM

How to classifie/recognize?

- Goal: Find correct class for new series of feature vectors from sensor input
- Choose HMM with highest probability to create this observation
- Problem: How to find probability to create observation given class (HMM)
- summation over all possible state sequences

Evaluation

Error/precision rates

a)

System	2-state	3-state	4-state
BASE	20.2%	14.4%	16.1%
ICA1	12.2%	11.5%	11.5%
ICA3	11.9%	10.8%	11.7%
ICA5	11.8%	11.5%	11.1%
ICA7	**10.7%**	**9.4%**	**10.5%**
ICA9	11.8%	10.4%	11.3%

b)

System	2-state	3-state	4-state
BASE	79.3%	83.5%	82.2%
ICA1	81.2%	82.5%	82.2%
ICA3	81.8%	82.1%	82.6%
ICA5	83.1%	81.7%	84.4%
ICA7	**84.0%**	**85.5%**	**84.8%**
ICA9	83.5%	84.7%	85.7%

Abbildung: a) Errors for different HMM's and Features [KMSW05, Table 4]; b) Precision for different HMM's and Features [KMSW05, Table 5]

Integration of temporal Features lead to a better performance

Typical Problems

- Features are important, typically used MFCC may not be suitable (for general audio)
- closed set of classes and training data was used
- labeled training data is needed, training can be expensive (human resources and computational resources)
- Number of HMM states could be adressed and optimized
- open question: how to use the created knowledge?

Dufaux, Alain ; Besacier, Laurent ; Ansorge, Michael ; Pellandini, Fausto:
Automatic sound detection and recognition for noisy environment.
In: *Proc. of the X European Signal Processing Conference* Citeseer, 2000

Geiger, Jürgen T. ; Lakhal, Mohamed A. ; Schuller, Björn ; Rigoll, Gerhard:
Learning New Acoustic Events in an HMM-Based System Using MAP Adaptation.
In: *INTERSPEECH*, ISCA, 2011, S. 293–296

Kraft, Florian ; Malkin, Robert ; Schaaf, Thomas ; Waibel, Alex:
Temporal ICA for classification of acoustic events i a kitchen environment.
In: *INTERSPEECH*, ISCA, 2005, S. 2689–2692

Ntalampiras, Stavros ; Potamitis, Ilyas ; Fakotakis, Nikos:
Exploiting Temporal Feature Integration for Generalized Sound Recognition.
In: *EURASIP J. Adv. Sig. Proc.* 2009 (2009)

Schuller, Bjrn W.:
Intelligent Audio Analysis.
Springer Publishing Company, Incorporated, 2013. –
ISBN 3642368050, 9783642368059