Audio Signal Classification Using Linear Predictive Coding and Random Forests

Lăcrimioara GRAMA, Corneliu RUSU

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group
Outline

- Research aim
- Acoustic Wildlife Intruder Detection System
- Wildlife Database
- Linear Predictive Coding
- Random Forests
- Stratified 10-fold cross validation
- Results
- Conclusion
Research Aim

- Audio signal classification system based on Linear Predictive Coding and Random Forests
  - Acoustic wildlife intruder detection system (WIDS)
- Sound classification has been the focus of intensive research and several approaches have been proposed in different domains
  - Medical applications: hearing aids and remote monitoring
  - Identification of the musical instruments from an audio recording
  - Environmental sound classification
  - Classification of the kitchen sounds
  - Vehicle identification
Why this research?

- The number of events that imply
  - Illegal logging, hunting,
  - Trespassing of natural reservations, parks, forests

increased so much in the past decade

⇒ On a high demand became the design of WIDS

- To detect in time unwanted activities within the protected areas + help the authorities to take an action
Why this research?

- Over 25 environmental agencies and organizations worldwide are being proactive in tracking illegal logging and hunting.

- About 25 million birds are killed illegally in the Mediterranean every year [BirdLife International 2017]

- Romania: in 2015 the authorities registered 34,870 cases of illegal logging, which means 96 cases/day [Greenpeace 2015]
  - Regarding the gravity of the deeds, of all cases of illegal logging recorded in 2015, 32% of them were classified as criminal offences, while 68% were contraventions.
Acoustic Wildlife Intruder Detection System

Intruder models → Intruder detection → Alarm → Response to alarm → Data processing

Database selection → Feature extraction LPC-10
Classification Stratified 10-fold cross validation + Random Forests → Output performance

SpeD 2017 | Audio Signal Classification Using Linear Predictive Coding and Random Forests
Wildlife Database

**Birds dataset** – 654 audio files originated from 70 different species of birds (Internet)

**Chainsaws dataset** – 356 audio files originated from 18 different types of chainsaws (SPG)

**Gunshots dataset** – 120 audio files originated from 40 different types of guns (Internet)

**Human voice dataset** – 207 speech sounds originated from 50 different former students from the TUCN

**Tractors dataset** – 260 audio files originated from 17 different types of tractors (SPG)

- 16 kHz, 16-bit
- None of the audio signals are studio recordings ⇒ they are subject to some additive noise from surroundings
Linear Predictive Coding Coefficients

Framing

Pre-emphasis

Windowing

H(z) = 1 - 0.97z^{-1}

25 ms frames (60% overlap)

Hamming window

LPCs

Levinson-Durbin

Features vector $F_k = \begin{bmatrix} \sigma_k^2 & a_{k,1} & a_{k,2} & \ldots & a_{k,10} \end{bmatrix}$

Features matrix $F_{Nx11} = \begin{bmatrix} \sigma_1^2 & a_{1,1} & \ldots & a_{1,10} \\ \sigma_2^2 & a_{2,1} & \ldots & a_{2,10} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_N^2 & a_{N,1} & \ldots & a_{N,10} \end{bmatrix}$

- $\sigma_k^2$ – prediction error variance
- $a_{k,i}$ – last 10 LPC coefficients

- $N = 1597$ – number of audio files
Why Random Forests?

- Acoustic WIDS – look for suspicious sound signals
  - Attack/unauthorized access to the natural environment
  - At an abstract level – WIDS purpose – to classify the input correctly as non-intruders or intruders
- Tradition systems can detect known intruders but cannot identify unknown ones
  - Nowadays machine learning techniques are attempting to be apply to this area of cybersecurity
- Many industries use machine learning techniques to better automate
  - Security screening
  - Border entry
  - College applicant selection
- Almost all kind of stuffs can be tackled with machine learning in order to take good decisions

Why Random Forests?

- SpeD 2017 | Audio Signal Classification Using Linear Predictive Coding and Random Forests

9/19
Why Random Forests?

- IBM – machine learning techniques
  - Applied to historical alert data
    - Can significantly improve classification accuracy
    - Can decrease research time for analysts
    - Can supplement analysts with additional data and insights to make better judgments
  - Very effective
    - In the elimination of white noise
    - Classification of benign data with a high degree of accuracy
      - For our framework, the benign data are the non-intruders
  - Machine learning and security are old friends
    ⇒ We should use for classification Random Forests
Random Forests

**RF of trees hypothesis**

**Bootstrap sets**

**OOB data**

**Predictions**

**Final prediction**

11 features in the dataset + target

Each node is split by choosing a feature out of a random sample of \( x = \sqrt{11} \) features (3)

The splitting feature gives the best information gain

Each tree is grown to the largest extent possible

**Full dataset**

1 2 3 ... N

draw N with replacement

Bootstrap 1

1 2 3 ... N

Bootstrap 2

1 2 3 ... N

Bootstrap t

Bootstrap 1                           Bootstrap 2                                                    Bootstrap t

Final prediction

Cᵢ
Stratified 10-fold cross validation

- Stratification
  - Is important for classification problems involving imbalanced datasets
  - Preserves classes distributions during training and testing
  - Reduces the estimate’s variance
Results

- 49 classifiers
  - Open source software issued under the GNU General Public License
  - A collection of machine learning algorithms for data mining tasks
  - Tools for data pre-processing, classification, regression, clustering, association rules, and even visualization
- 10 times stratified 10-fold cross validation
  - 27 classifiers out of 49 average CCR >90%
  - Random Forests

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Average CCR [%] (St.Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>94.88 (1.73)</td>
</tr>
<tr>
<td>Logistic</td>
<td>92.77 (1.57)</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>93.35 (1.72)</td>
</tr>
<tr>
<td>SVM (linear kernel)</td>
<td>97.64 (1.14)</td>
</tr>
<tr>
<td>SVM (radial basis kernel)</td>
<td>98.90 (0.81)</td>
</tr>
<tr>
<td>lazy.IBk</td>
<td>98.52 (0.98)</td>
</tr>
<tr>
<td>lazy.IBkLG</td>
<td>98.52 (0.98)</td>
</tr>
<tr>
<td>lazy.KStar</td>
<td>98.70 (1.04)</td>
</tr>
<tr>
<td>Logit Boost</td>
<td>92.60 (1.95)</td>
</tr>
<tr>
<td>CHIRP</td>
<td>92.68 (1.92)</td>
</tr>
<tr>
<td>JRip</td>
<td>92.75 (2.10)</td>
</tr>
<tr>
<td>PART</td>
<td>94.84 (1.64)</td>
</tr>
<tr>
<td>J48</td>
<td>94.70 (1.94)</td>
</tr>
<tr>
<td>Logistic Model Tree</td>
<td>96.96 (1.61)</td>
</tr>
<tr>
<td><strong>Random Forest</strong></td>
<td><strong>98.95 (0.91)</strong></td>
</tr>
<tr>
<td>Random Tree</td>
<td>95.97 (1.58)</td>
</tr>
<tr>
<td>REP Tree</td>
<td>92.13 (2.37)</td>
</tr>
</tbody>
</table>
Results – Random Forests

- 100 times stratified 10-fold cross validation
- Test phase
- Averaged CCR of each run
  - Minimum: 98.183%
    (frequency of apparition 1)
  - Maximum: 99.249%
    (frequency of apparition 10)
  - Mean value: 98.879%;
    Std.Dev.: 0.246
Results – Random Forests

- OOB error is evaluated by computing the error rate for each class and then averaging over all classes (the misclassification probability)
- Averaged OOB of each run
  - Minimum: 0.01113
  - Maximum: 0.01378
  - Mean value: 0.01239; Std.Dev.: 0.00053

Histogram of the out-of-bag error

⇒ good model for classification
Results – Random Forests

- **Precision vs recall curve** – insensitive to classes distribution
- **One-vs-all approach**
  - I.e., the dotted red line labeled ‘Bird’ means that the positive class is the class of birds, while the negative class consists of chainsaws, gunshots, human voices and tractors
  - All five possible variations are illustrated
### Results – Random Forests

#### Confusion matrix

<table>
<thead>
<tr>
<th>Output class</th>
<th>B</th>
<th>C</th>
<th>G</th>
<th>H</th>
<th>T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>652</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>99.69%</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>351</td>
<td></td>
<td></td>
<td>4</td>
<td>98.60%</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td>118</td>
<td></td>
<td></td>
<td>98.33%</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td>207</td>
<td></td>
<td>100.00%</td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>257</td>
<td>98.85%</td>
</tr>
</tbody>
</table>

#### Prevalence (BR)

- **B**: 40.83%
- **C**: 21.98%
- **G**: 7.39%
- **H**: 98.87%
- **T**: 98.87%

#### Probability of detection (TPR)

- **B**: 0.31%
- **C**: 1.40%
- **G**: 1.67%
- **H**: 0.00%
- **T**: 0.30%

#### Miss rate (FNR)

- **B**: 0.00%
- **C**: 0.40%
- **G**: 0.14%
- **H**: 0.00%
- **T**: 0.30%

#### Precision (PPV)

- **B**: 100.00%
- **C**: 96.72%
- **G**: 100.00%
- **H**: 98.47%
- **T**: 99.25%

#### False discovery rate

- **B**: 0.00%
- **C**: 1.13%
- **G**: 3.28%
- **H**: 1.53%
- **T**: 0.75%

#### Table: FAR and FOR

<table>
<thead>
<tr>
<th></th>
<th>FAR</th>
<th>FOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.00%</td>
<td>0.21%</td>
</tr>
<tr>
<td>C</td>
<td>0.32%</td>
<td>0.40%</td>
</tr>
<tr>
<td>G</td>
<td>0.27%</td>
<td>0.14%</td>
</tr>
<tr>
<td>H</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>T</td>
<td>0.30%</td>
<td>0.22%</td>
</tr>
</tbody>
</table>

- 12 audio signals out of 1597 are misclassified.
Conclusion

• A model for audio signal classification: LPC + RF
• The signals under classification belong to the class of sounds from WID applications
• The step by step model building was illustrated
• Evaluation of the proposed classification system: 100 x stratified 10-fold CV
• Multiclass classification – average CCR: 99.25%
  • There is no probability of false alarms: birds + human voices
  • For the other three classes the probability is low (~0.3%)
  • The false omission rate is also low: ~0.2% for birds and tractors, a little bit higher for chainsaws (0.4%), lower for gunshots (0.14%) and zero for human voices

⇒ Proposed audio classification system can be used as a good detection system, i.e. for WID problems
Audio Signal Classification Using Linear Predictive Coding and Random Forests

Lăcrimioara GRAMA, Corneliu RUSU

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

The 9th Conference on Speech Technology and Human-Computer Dialog – July 6, 2017