the data challenge in music information retrieval

FSRM 2023

alexander lerch

Georgia Tech | Center for Music Technology
College of Design

about
about me

- **education**
  - Electrical Engineering (Technical University Berlin)
  - Tonmeister (music production, University of Arts Berlin)

- **professional**
  - Associate Professor at the School of Music, Georgia Institute of Technology
  - 2000-2013: CEO at zplane.development

- **background**
  - audio algorithm design (20+ years)
  - commercial music software development (10+ years)
  - entrepreneurship (10+ years)

www.linkedin.com/in/lerch

about
research directions

- field: music information retrieval, audio content analysis
  - audio classification
    - ▶ genre, instrument, auto-tagging, . . .
  - music transcription
    - ▶ pitch, chord, performance data, . . .
  - music processing
    - ▶ separation, . . .
  - sound and music generation
    - ▶ controllability

- technical areas of interest
  - representation learning
  - machine learning with insufficient data
  - evaluation of generative systems

www.alexanderlerch.ch

about
research directions

Georgia Tech | **Center for Music Technology**
College of Design

- field: music information retrieval, audio content analysis
  - audio classification
    - ▶ genre, instrument, auto-tagging, ...
  - music transcription
    - ▶ pitch, chord, performance data, ...
  - music processing
    - ▶ separation, ...
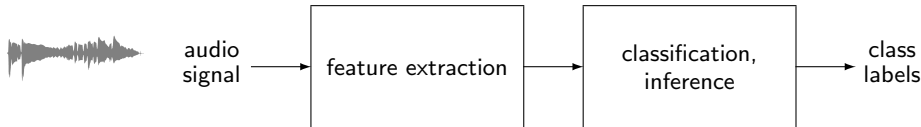 - sound and music generation
    - ▶ controllability

- technical areas of interest
  - representation learning
  - machine learning with insufficient data
  - evaluation of generative systems

www.alexanderlerch.com

about
○○

intro
●○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

introduction
audio classification — traditional

Georgia Tech | Center for Music Technology
College of Design

audio
signal → feature extraction → classification,
inference → class
labels

feature representation

- compact and non-redundant
- task-relevant
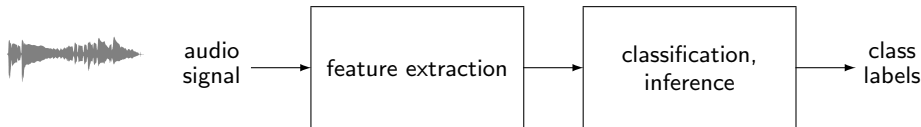- easy to analyze
- e.g., MFCCs etc.

classification

- map or convert feature to comprehensible domain
- e.g., Support Vector Machines etc.

---

[1] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," *Journal of the Audio Engineering Society (JAES)*, vol. 52, no. 7/8, pp. 724–739, 2004. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2016/10/Burred-and-Lerch-2004-Hierarchical-Automatic-Audio-Signal-Classification.pdf.

about
○○

intro
●○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

introduction
audio classification — traditional

Georgia | Center for Music
Tech | Technology
College of Design

feature representation

- compact and non-redundant
- task-relevant
- easy to analyze
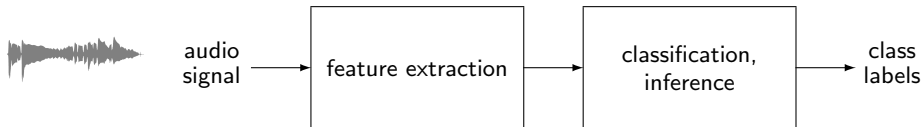- e.g., MFCCs etc.

classification

- map or convert feature to comprehensible domain
- e.g., Support Vector Machines etc.

---

[1] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," *Journal of the Audio Engineering Society (JAES)*, vol. 52, no. 7/8, pp. 724–739, 2004. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2016/10/Burred-and-Lerch-2004-Hierarchical-Automatic-Audio-Signal-Classification.pdf.

about
○○

intro
●○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

introduction
audio classification — traditional

feature representation

- compact and non-redundant
- task-relevant
- easy to analyze
- e.g., MFCCs etc.

classification

- map or convert feature to comprehensible domain
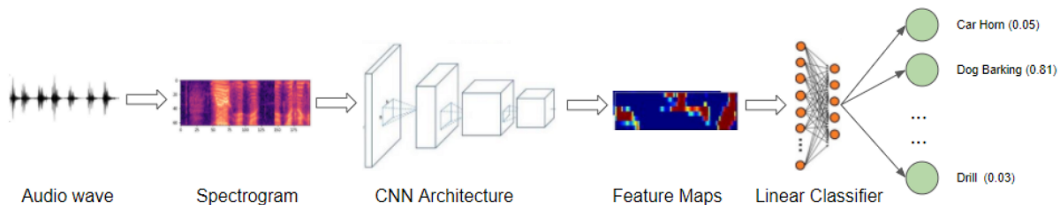- e.g., Support Vector Machines etc.

---

[1] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," *Journal of the Audio Engineering Society (JAES)*, vol. 52, no. 7/8, pp. 724–739, 2004. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2016/10/Burred-and-Lerch-2004-Hierarchical-Automatic-Audio-Signal-Classification.pdf.

about
○○

intro
○●

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

introduction
neural network based approaches

Georgia | Center for Music
Tech | Technology
College of Design

- no custom-designed features anymore
- learn features from basic inputs (like spectrograms)

Audio wave   Spectrogram   CNN Architecture   Feature Maps   Linear Classifier

Car Horn (0.05)
Dog Barking (0.81)
...
...
Drill (0.03)

Fig.: towardsdatascience.com

- less required expert-knowledge, more complex systems
- less expert-tweaking, more rigorous experimental requirement
- much **higher data requirements**

about
○○

intro
○●

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

introduction
neural network based approaches

Georgia Tech | Center for Music Technology
College of Design

- no custom-designed features anymore
- learn features from basic inputs (like spectrograms)

Audio wave → Spectrogram → CNN Architecture → Feature Maps → Linear Classifier → Car Horn (0.05), Dog Barking (0.81), ..., ..., Drill (0.03)
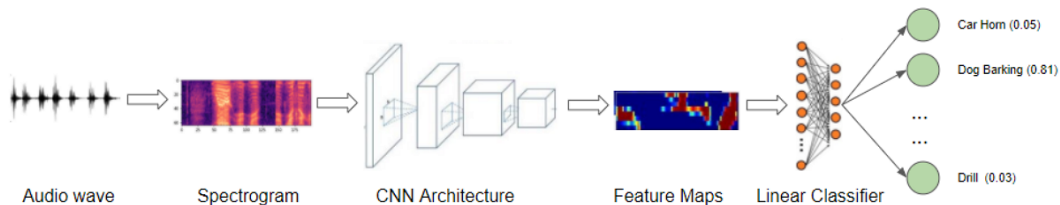
Fig.: towardsdatascience.com

- less required expert-knowledge, more complex systems
- less expert-tweaking, more rigorous experimental requirement
- much **higher data requirements**

data
importance of data

machine learning: generic algorithm mapping an input to an output

- mapping function is learned from patterns and characteristics **from data**
- ⇒ model **success largely depends on training data**

■ **general challenges** concerning data

- subjectivity
- noisiness
- imbalance & bias
- diversity & representativeness
- amount

data
importance of data

machine learning: generic algorithm mapping an input to an output

- mapping function is learned from patterns and characteristics **from data**
- ⇒ model **success largely depends on training data**

■ **general challenges** concerning data

- subjectivity
- noisiness
- imbalance & bias
- diversity & representativeness
- amount

https://imgs.xkcd.com/comics/machine_learning.png

about
oo

intro
oo

data
●oo

overview
o

reprogramming
ooooo

east
oooo

data revisited
o

conclusion
o

thanks
o

data
importance of data

machine learning: generic algorithm mapping
an input to an output

- mapping function is learned from patterns and
  characteristics **from data**
- ⇒ model **success largely depends on training
  data**

■ **general challenges** concerning data

- subjectivity
- noisiness
- imbalance & bias
- diversity & representativeness
- amount

data
importance of data

Georgia | Center for Music
Tech | Technology
College of Design

machine learning: generic algorithm mapping
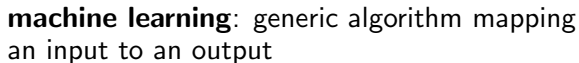an input to an output

- mapping function is learned from patterns and
  characteristics **from data**
- ⇒ model **success largely depends on training
  data**

■ **general challenges** concerning data

- subjectivity
- noisiness
- imbalance & bias
- diversity & representativeness
- **amount**

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG
PILE OF LINEAR ALGEBRA, THEN COLLECT
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL
THEY START LOOKING RIGHT.

https://imgs.xkcd.com/comics/machine_learning.png

about
○○

intro
○○

data
○●○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

data
insufficient data

Georgia | Center for Music
Tech | Technology
College of Design

insufficient data in music

data
insufficient data

Georgia Tech | Center for Music Technology
College of Design

insufficient data in music

- **music data** itself is not scarce (although there might be copyright issues...)

- **consumer annotations** are more difficult to collect, but there are some large collections

about   intro   **data**   overview   reprogramming   east   data revisited   conclusion   thanks
00   00   0●0   0   00000   0000   0   0   0

data
insufficient data

Georgia | Center for Music
Tech | Technology
          College of Design

insufficient data in music

- **music data** itself is not scarce (although there might be copyright issues…)

- **consumer annotations** are more difficult to collect, but there are some large collections

- **detailed musical annotations** are hard to come by, because
  - time consuming & tedious annotation process
  - experts needed for annotations

about
○○

intro
○○

data
○○●

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

data
previous work on insufficient data

Georgia Tech | Center for Music Technology
College of Design

■ literature proposes many ways of **dealing with insufficient data**

- data synthesis
- data augmentation[2]
- transfer learning
- semi- and self-supervised approaches
- ...

---

[2] Y. Qin and A. Lerch, "Tuning Frequency Dependency in Music Classification," en, in *Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Brighton, UK: Institute of Electrical and Electronics Engineers (IEEE), 2019, pp. 401–405. DOI: 10.1109/ICASSP.2019.8683340. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2019/04/Qin-and-Lerch-2019-Tuning-Frequency-Dependency-in-Music-Classificatio.pdf.

data
previous work on insufficient data

- literature proposes many ways of **dealing with insufficient data**
  - data synthesis
  - data augmentation
  - transfer learning[2]
  - semi- and self-supervised approaches
  - ...

---

[2] S. Gururani, M. Sharma, and A. Lerch, "An Attention Mechanism for Music Instrument Recognition," in *ISMIR*, Delft, 2019.

about
○○

intro
○○

data
○○●

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
○

data
previous work on insufficient data

Georgia **Center for Music**
Tech **Technology**
College of Design

- literature proposes many ways of **dealing with insufficient data**
  - data synthesis
  - data augmentation
  - transfer learning
  - semi- and self-supervised approaches[23]
  - . . .

---

[2]C.-W. Wu and A. Lerch, "Automatic drum transcription using the student-teacher learning paradigm with unlabeled music data," in *ISMIR*, Suzhou, 2017.

[3]S. Gururani and A. Lerch, "Semi-Supervised Audio Classification with Partially Labeled Data," in *Proceedings of the IEEE International Symposium on Multimedia (ISM)*, online: Institute of Electrical and Electronics Engineers (IEEE), 2021. [Online]. Available: https://arxiv.org/abs/2111.12761.

about   intro   data   overview   reprogramming   east   data revisited   conclusion   thanks
  oo   oo   ooo   ●   ooooo   oooo   o   o   o

overview
overview

Georgia | Center for Music
Tech | Technology
          College of Design

1 **self-supervised representation learning**
- utilize pre-trained features to improve classification

2 **reprogramming**
- utilize pre-trained model to improve classification

3 **data challenge revisited**

reprogramming
introduction

- **observation**
  - pre-trained deep models can be very powerful if trained with sufficient data, even for different tasks

- **idea**
  - re-using pre-trained models for a new task **without** re-training

- **goals**
  - keep number of training parameters minimal
  - utilize unmodified network trained on different task

reprogramming
introduction

Georgia | Center for Music
Tech | Technology
College of Design

- **observation**
  - pre-trained deep models can be very powerful if trained with sufficient data, even for different tasks

- **idea**
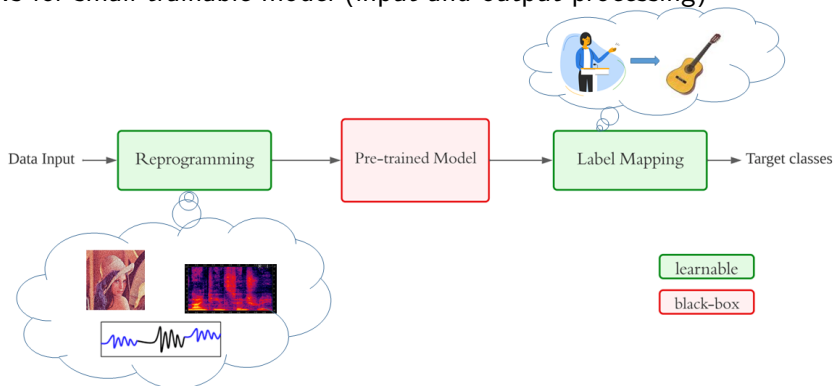 - re-using pre-trained models for a new task **without** re-training

- **goals**
  - keep number of training parameters minimal
  - utilize unmodified network trained on different task

reprogramming
overview

- inspired by
  - transfer learning
  - adversarial learning
- allows for small trainable model (input and output processing)

reprogramming
experimental setup: data

- OpenMic:
  - 20 classes of musical instruments
  - 10 s audio snippets (20000)

about
∘∘

intro
∘∘

data
∘∘∘

overview
∘

reprogramming
∘∘∘●∘

east
∘∘∘∘

data revisited
∘

conclusion
∘

thanks
∘

reprogramming
experimental setup: baselines

Georgia ‖ Center for Music
Tech ‖ Technology
College of Design

- Baseline AST:
  - good performance on audio event classification[4]

- ablation study:
  - CNN only
  - U-Net only
  - CNN + AST + FC
  - U-Net + AST + FC

---

[4]Y. Gong, Y.-A. Chung, and J. Glass, "AST: Audio Spectrogram Transformer," in *Proceedings of Interspeech*, arXiv: 2104.01778, Brno, Czechia, Jul. 2021. [Online]. Available: http://arxiv.org/abs/2104.01778 (visited on 04/17/2022).

about
oo

intro
oo

data
ooo

overview
o

reprogramming
ooooo●

east
oooo

data revisited
o

conclusion
o

thanks
o

reprogramming
results: classification metrics

method	F1 (macro)	train. param. (M)
AST + simple output mapping	62.03	0.001
CNN	60.77	0.017
U-Net	62.73	0.017
CNN + AST + FC	78.08	0.017
U-Net + AST + FC	**81.60**	0.018

MIC over OpenMIC comparison
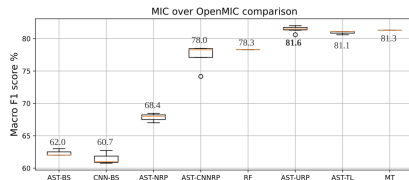
- a powerful model trained on a different task cannot easily be used directly
- proper input and output processing can significantly improve performance
- *re-programming can beat the state-of-the-art* with a fraction of trainable parameters (at least factor 10)

[5]H.-H. Chen and A. Lerch, "Music Instrument Classification Reprogrammed," in *Proceedings of the International Conference on Multimedia Modeling (MMM)*, Bergen, Norway, 2023. [Online]. Available: https://arxiv.org/abs/2211.08379.

embeddings as teachers
introduction

- **question**:
  - how can we provide extra training information without additional data labels

- **idea**:
  - use proven pre-trained embeddings (e.g., VGGish, OpenL3, ...)

- **goals**:
  - *impart knowledge* of pre-trained deep models
  - *improve model generalization* by utilizing pre-trained embeddings
  - *reduce model complexity*

- **general approach**:
  - combine transfer learning and knowledge distillation ideas

embeddings as teachers
introduction

Georgia | Center for Music
Tech | Technology
College of Design

- **question**:
  - how can we provide extra training information without additional data labels

- **idea**:
 - use proven pre-trained embeddings (e.g., VGGish, OpenL3, . . .)

- **goals**:
  - *impart knowledge* of pre-trained deep models
  - *improve model generalization* by utilizing pre-trained embeddings
  - *reduce model complexity*

- **general approach**:
  - combine transfer learning and knowledge distillation ideas

embeddings as teachers
introduction

- **question**:
  - how can we provide extra training information without additional data labels

- **idea**:
  - use proven pre-trained embeddings (e.g., VGGish, OpenL3, . . .)

- **goals**:
  - *impart knowledge* of pre-trained deep models
  - *improve model generalization* by utilizing pre-trained embeddings
  - *reduce model complexity*

- **general approach**:
  - combine transfer learning and knowledge distillation ideas

embeddings as teachers
method overview

Georgia Tech | Center for Music Technology
College of Design
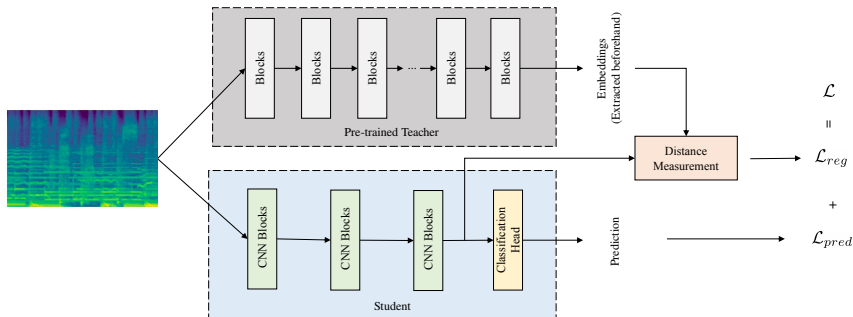
- **transfer learning**
  - use embeddings from a different task for the target task
- **knowledge distillation**
  - use a teacher to train a less complex student on the same task
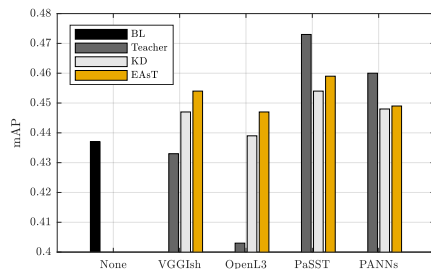
embeddings as teachers
experimental setup

- task: auto-tagging
  - MagnaTagATune (MTAT) dataset:
    - ▶ 50 music tags
    - ▶ 30 s audio snippets ($\approx$ 21000)

- systems:
  - baseline: student without teacher
  - teacher: embedding plus logistic regression
    - ▶ VGGish
    - ▶ OpenL3
    - ▶ PaSST
    - ▶ PANNs
  - KD: student trained with soft targets from teacher
  - EAsT: student regularized with teacher embeddings

embeddings as teachers
results

Georgia Tech | Center for Music Technology
College of Design

- student model consistently outperforms baseline

- student model consistently outperforms knowledge distillation

- student model outperforms teacher for "old" embeddings

- modern embeddings are powerful but complex

---

[6] Y. Ding and A. Lerch, "Audio Embeddings as Teachers for Music Classification," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, 2023. DOI: 10.48550/arXiv.2306.17424. [Online]. Available: http://arxiv.org/abs/2306.17424 (visited on 07/03/2023).

about
○○

intro
○○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
●

conclusion
○

thanks
○

data challenge revisited
insufficiency vs. representativeness

Georgia Tech | Center for Music Technology
College of Design

moderate improvements can be made to deal with insufficient data, but

data challenge revisited
insufficiency vs. representativeness

moderate improvements can be made to deal with insufficient data, but

is the amount of data really the main issue

❓

- a closer look at music datasets for popular tasks
  - chord detection: Beatles dataset
    - ▶ audio data
    - ▶ chord vocabulary
  - genre classification: several
    - ▶ genre labeling
  - source separation: MUSDB
    - ▶ choice of classes
- homogeneity questions representativeness of data
- category restrictions question generalization to task
- problem for methodology: data might impact architectural choices
- problem for training: system cannot learn what is hasn't seen or what seems

conclusion
data challenge

Georgia Tech | Center for Music Technology
College of Design

- we presented **2 recent approaches**
  - a novel *self-supervised regularization loss*
  - *reprogramming* for audio classification

- all approaches perform **at or above the state-of-the-art** with different trade-offs between
  - *training complexity*
  - *inference complexity*
  - *classification accuracy*
- remaining challenges might be bigger

about
○○

intro
○○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
●

thank you!

Georgia **Center for Music**
Tech **Technology**
College of Design

links

alexander lerch: www.linkedin.com/in/lerch

mail: alexander.lerch@gatech.edu

book: www.AudioContentAnalysis.org

zplane.development: www.zplane.de

music informatics group: musicinformatics.gatech.edu

✉

github.com/alexanderlerch

thank you!

Georgia Tech | Center for Music Technology
College of Design

links

alexander lerch: www.linkedin.com/in/lerch

mail: alexander.lerch@gatech.edu

book: www.AudioContentAnalysis.org

zplane.development: www.zplane.de

music informatics group: musicinformatics.gatech.edu

Second Edition
An Introduction to
Audio
Content Analysis
Music Information Retrieval
Tasks & Applications
Alexander Lerch
IEEE PRESS   WILEY

github.com/alexanderlerch

about ○○

intro ○○

data ○○○

overview ○

reprogramming ○○○○○

east ○○○○

data revisited ○

conclusion ○

thanks ●

thank you!

Georgia | Center for Music
Tech | Technology
College of Design

links

alexander lerch: www.linkedin.com/in/lerch

mail: alexander.lerch@gatech.edu

book: www.AudioContentAnalysis.org

zplane.development: www.zplane.de

music informatics group: musicinformatics.gatech.edu

Second Edition

An Introduction to
**Audio
Content Analysis**
Music Information Retrieval
Tasks & Applications

Alexander Lerch

IEEE PRESS   WILEY

github.com/alexanderlerch

about
○○

intro
○○

data
○○○

overview
○

reprogramming
○○○○○

east
○○○○

data revisited
○

conclusion
○

thanks
●

thank you!

Georgia
Tech

Center for Music
Technology

College of Design

links

alexander lerch: www.linkedin.com/in/lerch

mail: alexander.lerch@gatech.edu

book: www.AudioContentAnalysis.org

zplane.development: www.zplane.de

music informatics group: musicinformatics.gatech.edu

Second Edition

An Introduction to
**Audio
Content Analysis**
Music Information Retrieval
Tasks & Applications

Alexander Lerch

IEEE PRESS   WILEY

✉

github.com/alexanderlerch