Deep Neural Network for Speech Emotion Recognition
—A Study of Deep Learning—
Zhuowei Han
Institut für Signalverarbeitung und Systemtheorie
Universität Stuttgart
05.06.2014

Deep Neural Network for Speech Emotion Recognition

—A Study of Deep Learning—

Zhuowei Han

Institut für Signalverarbeitung
und Systemtheorie

Universität Stuttgart

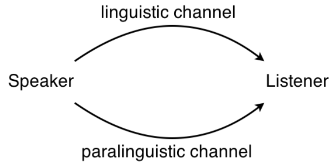
05.06.2014

Motivation

Why speech emotion recognition

- Most current work focuses on speech processing based on linguistic information, e.g.: Skype Translator
- More natural human-machine interaction requires paralinguistic information such as age, gender, emotion.
- Speech Recognition / Speeker Identification / Emotion

Recognition



Deep Network Applications

- Handwriting Digit Recognition
- Image Recognition

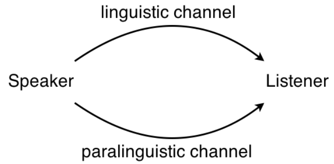
Motivation
Why speech emotion recognition
Most current work focuses on speech processing based on linguistic information, e.g.: Skype Translator
More natural human-machine interaction requires paralinguistic information such as age, gender, emotion.
Speech Recognition / Speaker Identification / Emotion
Recognition
Deep Network Applications
Handwriting Digit Recognition
Image Recognition

Motivation

Why speech emotion recognition

- Most current work focuses on speech processing based on linguistic information, e.g.: Skype Translator
- More natural human-machine interaction requires paralinguistic information such as age, gender, emotion.
- Speech Recognition / Speaker Identification / Emotion

Recognition



Deep Network Applications

- Handwriting Digit Recognition
- Image Recognition

Table of Contents

Foundations
Mel Frequency Cepstral Features
Emotion Recognition Approaches
Conditional Restricted Boltzmann Machine
Concept
Deep Neural Networks
Concept
Problems and Solutions
Long Short Term Memory
Recurrent Neural Network

Table of Contents

Foundations
- Mel Frequency Cepstral Features
- Emotion Recognition Approaches

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks
- Concept
- Problems and Solutions

Long Short Term Memory
- Recurrent Neural Network

Table of Contents
Foundations
Mel Frequency Cepstral Features
Emotion Recognition Approaches
Conditional Restricted Boltzmann Machine
Deep Neural Networks
Long Short Term Memory

Table of Contents

Foundations
- Mel Frequency Cepstral Features
- Emotion Recognition Approaches

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks
- Concept
- Problems and Solutions

Long Short Term Memory
- Recurrent Neural Network

2015-03-25

Deep Learning
└ Foundations
  └ Mel Frequency Cepstral Features
    └ Mel Frequency Cepstral Features

Mel Frequency Cepstral Features

- short-term power spectrum
- mel-scale approximate human perception
- widely-used in speech recognition tasks

Transformation between Mel and Hertz scale

$$f_{mel} = 1125 \ \ln\left(1 + f_{Hz}/700\right) \quad (1)$$

$$f_{Hz} = 700 \left(\exp(f_{mel}/1125) - 1\right) \quad (2)$$

Mel Frequency Cepstral Features
short-term power spectrum
mel-scale approximate human perception
widely-used in speech recognition tasks
Transformation between Mel and Hertz scale
$f_{mel} = 1125 \ \ln\left(1 + f_{Hz}/700\right)$ (1)
$f_{Hz} = 700 \left(\exp(f_{mel}/1125) - 1\right)$ (2)

Emotion Recognition Approaches
Traditional Approaches
- pre-selected features
- supervised training
- low-level features not appropriate for claasification
- shallow structure of classifiers
Deep Learning Approaches
- learning representations from high-dim data
- extracting appropriate features without hand-crafting
- low-level features are used to build high-level features as network gets deeper
-

Emotion Recognition Approaches

Traditional Approaches

- pre-selected features
- supervised training
- low-level features not appropriate for claasification
- shallow structure of classifiers

Deep Learning Approaches

- learning representations from high-dim data
- extracting appropriate features without hand-crafting
- low-level features are used to build high-level features as network gets deeper
-

Table of Contents
Foundations
Conditional Restricted Boltzmann Machine
Concept
Deep Neural Networks
Long Short Term Memory

Table of Contents

Foundations
- Mel Frequency Cepstral Features
- Emotion Recognition Approaches

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks
- Concept
- Problems and Solutions

Long Short Term Memory
- Recurrent Neural Network

Table of Contents

Foundations

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks

Long Short Term Memory

Table of Contents
Foundations
Conditional Restricted Boltzmann Machine
Deep Neural Networks
Concept
Problems and Solutions
Long Short Term Memory

Table of Contents

Foundations
- Mel Frequency Cepstral Features
- Emotion Recognition Approaches

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks
- Concept
- Problems and Solutions

Long Short Term Memory
- Recurrent Neural Network

Table of Contents

Foundations
Conditional Restricted Boltzmann Machine
Deep Neural Networks
Concept
Problems and Solutions
Long Short Term Memory

Computing net-activation

$$
\begin{aligned}
\underline{z}_k^{(l+1)} &= \mathbf{W}^{(l)}\underline{a}_k^{(l)} + \underline{b}^{(l)} \\
\underline{a}_k^{(l+1)} &= \underline{\mathit{\Phi}}\left(\underline{z}_k^{(l+1)}\right) \\
\underline{\hat{y}}_k &= \underline{a}_k^{(ol)}
\end{aligned}
$$

- Arbitrary non-linear mapping from $\underline{x}_k$ to $\underline{\hat{y}}_k$ possible
- Relation $N \Leftrightarrow$ Complexity
- Deep Architectures $(l\uparrow)$ more efficient than shallow ones $(l\downarrow, N_l\uparrow)$

Determining the parameters

Training objective

$$J(\mathbf{W}, \underline{b}) = \sum_{\forall k} \frac{1}{2} ||\underline{y}_k - \hat{\underline{y}}_k||^2 + \frac{\lambda}{2} \sum_{\forall l} ||\mathbf{W}^{(l)}||_F^2 \quad (3)$$

$$\mathbf{W}, \underline{b} = \arg\min_{\mathbf{W}, \underline{b}} J(\mathbf{W}, \underline{b}) \quad (4)$$

Numerical minimization

- Gradient calculation with Backpropagation
- Stochastic gradient descent
- **L**imited memory **B**royden-**F**letcher-**G**oldfarb-**S**hanno (L-BFGS)

Determining the parameters
Training objective
$$J(\mathbf{W}, \underline{b}) = \sum_{\forall k} \frac{1}{2} ||\underline{y}_k - \hat{\underline{y}}_k||^2 + \frac{\lambda}{2} \sum_{\forall l} ||\mathbf{W}^{(l)}||_F^2 \quad (3)$$
$$\mathbf{W}, \underline{b} = \arg\min_{\mathbf{W}, \underline{b}} J(\mathbf{W}, \underline{b}) \quad (4)$$
Numerical minimization
- Gradient calculation with Backpropagation
- Stochastic gradient descent
- Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS)

Problems
- Optimization problem non-convex
  $\Rightarrow$ getting stuck in poor local minima
- Diffusion of gradients
- Large p small n problem $\Rightarrow$ overfitting

Problems

- Optimization problem non-convex
  $\Rightarrow$ getting stuck in poor local minima
- Diffusion of gradients
- Large p small n problem $\Rightarrow$ overfitting

Solutions
- Layerwise Pre-training
- Layerwise Pre-training

Solutions

- Layerwise Pre-training
- Layerwise Pre-training

Table of Contents

Foundations
Conditional Restricted Boltzmann Machine
Deep Neural Networks
Long Short Term Memory
Recurrent Neural Network

Table of Contents

Foundations
- Mel Frequency Cepstral Features
- Emotion Recognition Approaches

Conditional Restricted Boltzmann Machine
- Concept

Deep Neural Networks
- Concept
- Problems and Solutions

Long Short Term Memory
- Recurrent Neural Network

2015-03-25

Deep Learning
└Long Short Term Memory
 └Recurrent Neural Network
  └Recurrent Neural Network

Recurrent Neural Network
Concepts of RNN
- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with backpropagation through time (BPTT)

Recurrent Neural Network

Concepts of RNN

- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with backpropagation through time (BPTT)

Recurrent Neural Network
Concepts of RNN
Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.
$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

tell me a story.

Recurrent Neural Network

Concepts of RNN

- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with **b**ack**p**ropagation **t**hrough **t**ime (BPTT)

Recurrent Neural Network
Concepts of RNN
Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.
Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.

Recurrent Neural Network

Concepts of RNN

- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with **b**ack**p**ropagation **t**hrough **t**ime (BPTT)

Recurrent Neural Network
Concepts of RNN
Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.
$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$
$y_t = W_{hy}h_t + b_y$
Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
Potentially to model arbitary dynamic system.
Trained with backpropagation through time (BPTT)

Recurrent Neural Network

Concepts of RNN

- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with **b**ack**p**ropagation **t**hrough **t**ime (BPTT)

2015-03-25

Deep Learning
└─Long Short Term Memory
  └─Recurrent Neural Network
    └─Recurrent Neural Network

Recurrent Neural Network

Concepts of RNN

- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

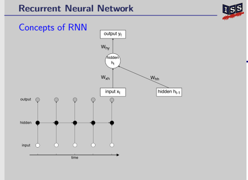
- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with **back**propagation **t**hrough **t**ime (BPTT)

Concepts of RNN

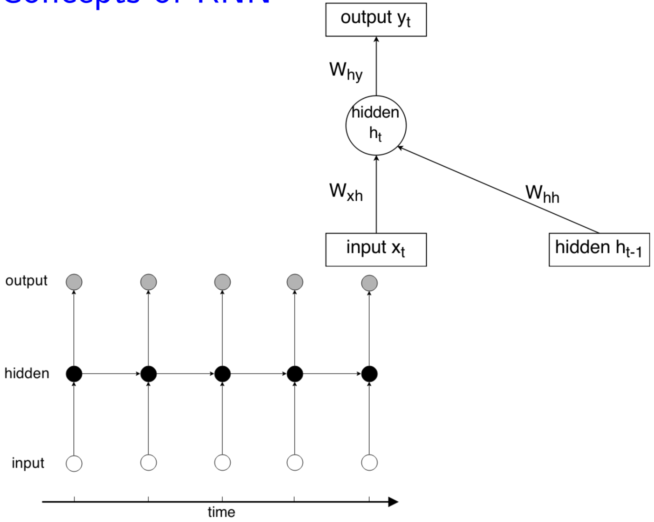
- Same Structure as MLP but differs from feed-forward network, enabling nonlinear mapping.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

- Feedback connection between previous hidden units and current hidden units, enabling memory past hidden state.
- Potentially to model arbitary dynamic system.
- Trained with **back**propagation **t**hrough **t**ime (BPTT)



Concepts of RNN



Long short term memory

Long short term memory