Automatic Summarization
(and other stuff)

Berkeley

NLP

Taylor Berg-Kirkpatrick
CS 288 UC Berkeley
Summarization
Summarization
Multidocument Summarization
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail.
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.
Maximum Marginal Relevance
Maximum Marginal Relevance

\[ S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9 \]
Maximum Marginal Relevance

S_9

S_2

S_1 \quad S_4

S_7 \quad S_3

S_5

S_8

S_6
Maximum Marginal Relevance

S_1: She traveled to France on Friday.
Maximum Marginal Relevance

$S_1$: She traveled to France on Friday.
Maximum Marginal Relevance

S1: She traveled to France on Friday.
S4: On Friday, She took a trip to France.
Maximum Marginal Relevance

S_9

S_2

S_5

S_1 \rightarrow S_4

S_7

S_3

S_8

S_6

S_1: She traveled to France on Friday.
S_4: On Friday, She took a trip to France.
Maximum Marginal Relevance

$S_1$: She traveled to France on Friday.

$S_4$: On Friday, She took a trip to France.

$S_8$: She plans to stay for two weeks.
Maximum Marginal Relevance

S1: She traveled to France on Friday.
S4: On Friday, She took a trip to France.
S8: She plans to stay for two weeks.
Maximum Marginal Relevance

S_9

S_2

S_1 S_4

S_7 S_3 S_8

S_5

S_6
Maximum Marginal Relevance

S1  S4  S8

S2

S9  S5

S7  S3  S6
Maximum Marginal Relevance

S_1  S_4

S_2

S_3

S_7

S_8

S_9

S_5

S_6
Maximum Marginal Relevance
Maximum Marginal Relevance

\[ \text{argmax}_{i \in D \setminus S} \left[ \lambda \cdot \text{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \text{sim}(S_i, S_j) \right] \]

[Carbonell and Goldstein, 1998]
Maximum Marginal Relevance

$$\arg \max_{i \in D \setminus S} \left[ \lambda \cdot \text{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \text{sim}(S_i, S_j) \right]$$

[Carbonell and Goldstein, 1998]
Maximum Marginal Relevance

\[
\text{argmax}_{i \in D \setminus S} \left[ \lambda \cdot \text{sim}(S_i, D) - (1 - \lambda) \cdot \max_j (\text{sim}(S_i, S_j)) \right]
\]

[Carbonell and Goldstein, 1998]
Max Coverage

She stopped in France. In France she remained.
Max Coverage

She stopped in France. In France she remained.
Max Coverage

She stopped in France. In France she remained.
She stopped in France. In France she remained.
Max Coverage

A summary is a set of cuts: $S$

She stopped in France. In France she remained.
Max Coverage

She stopped in France. In France she remained.
She stopped in France. In France she remained.
Max Coverage

She stopped in France.

In France she remained.

(she, stopped) ✓
(stopped, in) ✓
(in, france) ✓
(france, she)

(she, remained)
Max Coverage

She stopped in France.  In France she remained.

(she, stopped)  (in, france) ✓  (she, remained) ✓
(stopped, in)  (france, she) ✓
Max Coverage

She stopped in France.
In France she remained.

Set of bigrams covered by summary: \( B(s) \)
Max Coverage

\[
\max_{s} \text{coverage}(s)
\]
Max Coverage

\[
\max_{s} \text{coverage}(s)
\]

\[
s.t. \quad \text{length}(s) \leq L_{\text{max}}
\]
Max Coverage

\[
\max_s \sum_{b \in B(s)} \text{value}(b)
\]

s.t. \( \text{length}(s) \leq L_{max} \)
Max Coverage

\[
\max_{s} \sum_{b \in B(s)} \text{value}(b)
\]

s.t. \ length(s) \leq L_{\text{max}}

\[
\text{value}(b) = \text{freq}(b)
\]
Max Coverage

$$\max_s \sum_{b \in B(s)} \text{value}(b)$$

s.t.  \(\text{length}(s) \leq L_{max}\)

$$\text{value}(b) = \text{freq}(b)$$

[Gillick and Favre 2008]
\[
\max_s \sum_{b \in B(s)} \text{value}(b)
\]

s.t. \(\text{length}(s) \leq L_{\text{max}}\)

[Gillick and Favre 2008]
ILP for Decoding

$$\max_s \sum_{b \in B(s)} \text{value}(b)$$

s.t. $$\text{length}(s) \leq L_{\text{max}}$$
ILP for Decoding

\[
\max_{s,z} \sum_b z_b \cdot \text{value}(b)
\]

s.t. \( \text{length}(s) \leq L_{\text{max}} \)
ILP for Decoding

\[
\max_{s, z} \sum_{b} z_b \cdot \text{value}(b)
\]

s.t. \( \text{length}(s) \leq L_{\text{max}} \)

bigrams in \( s \) are covered
ILP for Decoding

$$\max_{s, z} \sum_b z_b \cdot \text{value}(b)$$

s.t. \( \text{length}(s) \leq L_{max} \)

bigrams in \( s \) are covered

only bigrams in \( s \) are covered
ILP for Decoding

\[
\begin{align*}
\max_{s,z} \quad & \sum_{b} z_b \cdot \text{value}(b) \\
\text{s.t.} \quad & \sum_{n} l_n s_n \leq L_{\text{max}} \\
& \forall n,b \quad s_n Q_{nb} \leq z_b \\
& \forall b \quad \sum_{n} s_n Q_{nb} \geq z_b
\end{align*}
\]
Linear Model for Extraction

$$\max_s \sum_{b \in B(s)} \text{value}(b)$$

s.t. $\text{length}(s) \leq L_{max}$

$$\text{value}(b) = \text{freq}(b)$$
Linear Model for Extraction

$$\max_s \sum_{b \in B(s)} \text{value}(b)$$

s.t. \( \text{length}(s) \leq L_{\text{max}} \)

Parameterize using features:

$$\text{value}(b) = w^\top f(b)$$
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.

Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail.

The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.

[Martins and Smith 2009] [Woodsend and Lapata 2010]
Extraction and Compression

She stopped in France. In France she remained.
She stopped in France. In France she remained.
She stopped in France. In France she remained.
She stopped in France. In France she remained.
She stopped in France. In France she remained.
Joint Extractive / Compressive Model

\[ S \]

She stopped in France. In France she remained.

- (she, stopped) ✓
- (in, france) ✓
- (stopped, in) ✓
- (france, she) ✓
- (she, remained) ✓
She stopped in France. In France she remained.

\[ B(s) \]
She stopped in France. In France she remained.

\( B(s) \)
Joint Extractive / Compressive Model

\[
\max_s \left[ \sum_{b \in B(s)} \text{value}(b) \right]
\]
Joint Extractive / Compressive Model

\[
\max_s \left[ \sum_{b \in B(s)} \text{value}(b) + \sum_{c \in s} \text{value}(c) \right]
\]
Joint Extractive / Compressive Model

\[
\max_s \left[ \sum_{b \in B(s)} \text{value}(b) + \sum_{c \in s} \text{value}(c) \right]
\]

Parameterize using features:

\[
\text{value}(b) = w^\top f(b)
\]

\[
\text{value}(c) = w^\top f(c)
\]
Learning

Linear prediction:

\[ \text{score}(s) = w^\top f(s) \]
Learning

Linear prediction:

\[ \text{score}(s) = w^\top f(s) \]

Feature function factors:

\[ f(s) = \sum_{b \in B(s)} f(b) + \sum_{c \in s} f(c) \]
Features
Features

Bigram Features $f(b)$
Features

Bigram Features $f(b)$

Cut Features $f(c)$
Features

Bigram Features $f(b)$  
Cut Features $f(c)$

COUNT: Bucketed document counts
STOP: Stop word indicators
POSITION: First document position indicators
CONJ: All two- and three-way conjunctions of above
BIAS: Always one
## Features

**Bigram Features**  $f(b)$  

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUNT:</td>
<td>Bucketed document counts</td>
</tr>
<tr>
<td>STOP:</td>
<td>Stop word indicators</td>
</tr>
<tr>
<td>POSITION:</td>
<td>First document position indicators</td>
</tr>
<tr>
<td>CONJ:</td>
<td>All two- and three-way conjunctions of above</td>
</tr>
<tr>
<td>BIAS:</td>
<td>Always one</td>
</tr>
</tbody>
</table>

**Cut Features**  $f(c)$  

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COORD:</td>
<td>Coordinated phrase, four versions: NP, VP, S, SBAR</td>
</tr>
<tr>
<td>S-ADJUNCT:</td>
<td>Adjunct to matrix verb, four versions: CC, PP, ADVP, SBAR</td>
</tr>
<tr>
<td>REL-C:</td>
<td>Relative clause indicator</td>
</tr>
<tr>
<td>ATTR-C:</td>
<td>Attribution clause indicator</td>
</tr>
<tr>
<td>ATTR-PP:</td>
<td>PP attribution indicator</td>
</tr>
<tr>
<td>TEMP-PP:</td>
<td>Temporal PP indicator</td>
</tr>
<tr>
<td>TEMP-NP</td>
<td>Temporal NP indicator</td>
</tr>
<tr>
<td>BIAS:</td>
<td>Always one</td>
</tr>
</tbody>
</table>
Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.
Obtaining Training Data

TAC 2009: 44 Document Collections

Intermediate Extraction
Obtaining Training Data

TAC 2009: 44 Document Collections

Maximize Gold Recall

Intermediate Extraction
Obtaining Training Data

TAC 2009: 44 Document Collections

Intermediate Extraction

Maximize Gold Recall

Mechanical Turk
Learning
Learning

Structured SVM Training:
Structured SVM Training:

$$\min_w \| w \|_2^2$$
Learning

Structured SVM Training:

$$\min_{w} \| w \|_2^2$$

s.t. for all possible guess summaries:

score of gold exceeds score of guess by loss
Learning

Structured SVM Training:

$$\min_w \|w\|^2_2$$

s.t. for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S)$$
Learning

Structured SVM Training:

\[
\min_w \|w\|_2^2
\]

s.t. for all possible guess summaries:

\[
w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi
\]
Structured SVM Training:

$$\min_w \left[ \|w\|^2_2 + \xi \right]$$

s.t. for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$
Structured SVM Training:

$$\min_w \left[ \|w\|_2^2 + \xi \right]$$

s.t. for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$

End-to-end summarization quality
Structured SVM Training:

$$\min_{w} \left[ \|w\|_2^2 + \xi \right]$$

s.t. for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$

Exponentially many constraints!
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Learning

[Tsochantaridis et al. 2004]
Results

Rouge-2
Results

Rouge-2

Last Doc Baseline
Results

Rouge-2

Pyramid

☐ Last Doc Baseline
Results

Rouge-2

Pyramid

Linguistic Quality

Last Doc Baseline
Results

Linguistic Quality

Rouge-2

Pyramid

Last Doc Baseline
Extractive Baseline

5.9
10.1
23.5
35
6.2
7.2
Results

Rouge-2

<table>
<thead>
<tr>
<th>Linguistic Quality</th>
<th>Last Doc Baseline</th>
<th>Extractive Baseline</th>
<th>Learned Extractive</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>7.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pyramid

<table>
<thead>
<tr>
<th>Linguistic Quality</th>
<th>Last Doc Baseline</th>
<th>Extractive Baseline</th>
<th>Learned Extractive</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Rouge-2

Pyramid

Linguistic Quality

- Last Doc Baseline
- Extractive Baseline
- Learned Extractive
- Learned E + C
Coarse-to-fine Decoding
Coarse-to-fine Decoding
Coarse-to-fine Decoding

Intermediate Extraction
Coarse-to-fine Decoding
Coarse-to-fine Decoding

Intermediate Extraction
Coarse-to-fine Decoding

Intermediate Extraction
Coarse-to-fine Decoding
Coarse-to-fine Decoding
Coarse-to-fine Decoding

Intermediate Extraction

Exact E + C

Approximate E + C
Coarse-to-fine Decoding

Avg Objective

Avg Recall

Seconds per instance

Intermediate extraction size

Full Extractive
Coarse-to-fine Decoding

- Avg Objective
- Avg Recall
- Seconds per instance

- Full Extractive
- Full Exact E+C
- Approx E+C
Coarse-to-fine Decoding

Avg Objective

Avg Recall

Seconds per instance

Intermediate extraction size

- Full Extractive
- Full Exact E+C
- Approx E+C
- LP Relaxation
Piano Music Transcription
Piano Music Transcription
Piano Music Transcription
Piano Music Transcription

note

time

[Waveform diagram]
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds
Piano Sounds

freq

time
Piano Sounds
Piano Sounds
Spectral Shape

freq

time
Spectral Shape

![Spectral Shape Diagram](image)
Spectral Shape
Spectral Shape
Temporal Shape

freq

time
Temporal Shape
Temporal Shape
Temporal Shape
Temporal Shape

freq

time
Temporal Shape

(freq)
Temporal Shape

(freq, time)
Temporal Shape

![Diagram showing spectral analysis and time series data]
Temporal Shape
Temporal Shape

freq

time
Temporal Shape
Temporal Shape
Temporal Shape
Polyphony
Polyphony
Polyphony
Polyphony
Polyphony
Polyphony
Polyphony
Generative Model
Generative Model
Generative Model

Note events

velocity

time

note

time
Generative Model

Note events

$M_n$
Generative Model

$M_{n}$

Note events

velocity

time
Generative Model

\[ M_{rl} \]

Note events

\[ \text{time} \]
Generative Model

Note events

Note:
- PLAY
- REST
- duration
- velocity

\[ \mu_n \rightarrow M_n \]
Generative Model

Note events
time

Activation
time

PLAY
REST
duration velocity

$\mu_n \rightarrow M_n$
Generative Model

\[ \mu_n \rightarrow M_n \rightarrow A_n \]

Note events

Activation

time

duration
velocity

PLAY
REST

\( \mu_n \)
Generative Model

Note events

Activation

PLAY
REST
duration velocity

µn → Mn

αn → An

time
Generative Model

Note events

[Diagram showing note events with time and duration]

Activation

[Diagram showing activation with time and frequency]

Component spectrogram

[Diagram showing component spectrogram with time and frequency]
Generative Model

Note events

Activation

Component spectrogram

$\mu_n \rightarrow M_n \rightarrow A_n \rightarrow S_n$
Generative Model
Generative Model

- Note events
- Activation
- Component spectrogram
Generative Model

Note events

Activation

Component spectrogram

Spectrogram
Generative Model

Latent variables

Note events

Activation

Component spectrogram

Spectrogram

PLAY
REST
duration velocity

μ_n

M_n

α_n

A_n

σ_n

S_n

X

freq

time

freq

time

freq

time
Generative Model

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram
Note Event Model
Note Event Model

$M_n$
Note Event Model

$M_n$
Note Event Model

$M_n$

Event type
Note Event Model

\[ M_n \]

Event type

PLAY

\[ E_1 \]
Note Event Model

Event type

PLAY $E_1$ REST $E_2$
Note Event Model

Event type

PLAY $E_1$ \rightarrow REST $E_2$ \rightarrow PLAY $E_3$
Note Event Model

Event type

PLAY $E_1$ → REST $E_2$ → PLAY $E_3$
Note Event Model

Event type

PLAY

REST

PLAY

$E_1$ --> $E_2$ --> $E_3$
Note Event Model

Event type

$M_n$

PLAY

REST

$D_1$

Duration

$E_1$  $E_2$  $E_3$

PLAY  REST  PLAY
Note Event Model

Event type

PLAY

REST

PLAY

Event Model

$M_n$

Duration

$D_1$

$D_2$
Note Event Model

Event type

PLAY

REST

PLAY
Note Event Model

Event type

PLAY $E_1$ → REST $E_2$ → PLAY $E_3$

Duration

$D_1$ $D_2$ $D_3$
Note Event Model

Event type

PLAY

REST

duration

Velocity

$M_n$

$E_1$ $E_2$ $E_3$

$D_1$ $D_2$ $D_3$

$V_1$

$D_1$ $D_2$ $D_3$
Note Event Model

Event type

Duration

Velocity

Mn

E1

E2

E3

D1

D2

D3

V1

V2
Note Event Model

Event type

PLAY

REST

duration

velocity

Event Model

$M_n$

$E_1$ → $E_2$ → $E_3$

Duration

$D_1$ → $D_2$ → $D_3$

Velocity

$V_1$ → $V_2$ → $V_3$
Note Event Model

Event type

PLAY

REST

PLAY

Duration

Velocity

$E_1$ → $E_2$ → $E_3$

$D_1$ → $D_2$ → $D_3$

$V_1$ → $V_2$ → $V_3$

$E_n$ $M_n$ $\mu_n$

duration

velocity
Activation Model
Activation Model
Activation Model
Activation Model
Activation Model
Activation Model
Activation Model

Activation

$A_n$
Activation Model

$D_1$, $D_2$, $D_3$

$V_1$, $V_2$, $V_3$

$A_n$ Temporal shape
Activation Model

$V_1$, $V_2$, $V_3$, $D_1$, $D_2$, $D_3$, $\alpha_n$, $A_n$

Temporal shape
Activation Model

$D_1$, $D_2$, $D_3$

$V_1$, $V_2$, $V_3$

$A_n$, $\alpha_n$

Temporal shape

Copy temporal shape
Activation Model

$D_1 \rightarrow V_1$

$D_2 \rightarrow V_2$

$D_3 \rightarrow V_3$

$\alpha_n$ Temporal shape

truncate to duration

Activation

$A_n$
Activation Model

\[ V_1, V_2, V_3 \]

\[ D_1, D_2, D_3 \]

\[ A_n \]

Temporal shape

truncates to duration
Activation Model

$D_1$ $V_1$ $D_2$ $V_2$ $D_3$ $V_3$

$A_n$ Temporal shape

scale to velocity

Activation
Activation Model

\[ A_n \]

\[ V_1 \]
\[ V_2 \]
\[ V_3 \]

\[ D_1 \]
\[ D_2 \]
\[ D_3 \]

\[ \alpha_n \] Temporal shape

scale to velocity
Activation Model

D_1 \rightarrow V_1 \rightarrow \text{Activation} \rightarrow A_n
D_2 \rightarrow V_2
D_3 \rightarrow V_3

\alpha_n \text{ Temporal shape

Add Gaussian noise}
Activation Model

\[ \begin{align*}
V_1 & \quad D_1 \\
V_2 & \quad D_2 \\
V_3 & \quad D_3
\end{align*} \]

\[ \alpha_n \quad \text{Temporal shape} \]

\[ \text{add Gaussian noise} \]

Activation

\[ A_n \]
Component Spectrogram Model

Activation

$A_n$
Component Spectrogram Model

Activation

$A_n$

Spectral shape

$\sigma_n$
Component Spectrogram Model

Activation

$A_n$

Spectral shape

$\sigma_n$
Component Spectrogram Model

Activation

\[ A_n \]

Spectral shape

\[ \sigma_n \]

Poisson noise
Component Spectrogram Model

Activation

$A_n$

Spectral shape

$\sigma_n$

Component spectrogram

$S_n$

Poisson noise
Total Spectrogram Model
Total Spectrogram Model

\[ \mathbf{X} = \sum_{i=1}^{N} \sigma_i \cdot \mathbf{S}_i \]
Learning and Inference

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram
Learning and Inference

Parameters
- PLAY
- REST
- duration
- velocity

Latent variables
- Note events
- Activation
- Component spectrogram
- Spectrogram

Parameters:
- $\mu_n$
- $\alpha_n$
- $\sigma_n$

Latent variables:
- $M_n$
- $A_n$
- $S_n$
- $X$

$\mu_n$, $\alpha_n$, and $\sigma_n$ are functions of time.

$X$ is a function of frequency and time.
Learning and Inference

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram

PLAY
REST
durationvelocity

time

\[ \mu_n \]

\[ M_n \]

\[ \alpha_n \]

\[ A_n \]

\[ \sigma_n \]

\[ S_n \]

\[ X \]

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram
Learning and Inference

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram

Parameters

Latent variables

Note events

Activation

Component spectrogram

Spectrogram
Learning and Inference
Learning and Inference

Note events update:

\[ M | A, \alpha, \mu \]
Learning and Inference

Note events update:

\[ M | A, \alpha, \mu \]

Semi-Markov dynamic program
Learning and Inference

Note events update:

$M | A, \alpha, \mu$

Semi-Markov dynamic program

Temporal shapes update:

$\alpha | A, M$
Learning and Inference

Note events update:

\[ M | A, \alpha, \mu \]

Semi-Markov dynamic program

Temporal shapes update:

\[ \alpha | A, M \]

Closed form update
Learning and Inference

Note events update:

\[ M | A, \alpha, \mu \]

Semi-Markov dynamic program

Temporal shapes update:

\[ \alpha | A, M \]

Closed form update

Activations update:

\[ A | M, X, \alpha, \sigma \]
Learning and Inference

Note events update:
\[ M \mid A, \alpha, \mu \]
Semi-Markov dynamic program

Temporal shapes update:
\[ \alpha \mid A, M \]
Closed form update

Activations update:
\[ A \mid M, X, \alpha, \sigma \]
Exponentiated gradient ascent
Learning and Inference

Note events update:
\[ M | A, \alpha, \mu \]
Semi-Markov dynamic program

Temporal shapes update:
\[ \alpha | A, M \]
Closed form update

Activations update:
\[ A | M, X, \alpha, \sigma \]
Exponentiated gradient ascent

Spectral shapes update:
\[ \sigma | A, X \]
Learning and Inference

Note events update: $M|A, \alpha, \mu$

Semi-Markov dynamic program

Temporal shapes update: $\alpha|A, M$

Closed form update

Activations update: $A|M, X, \alpha, \sigma$

Exponentiated gradient ascent

Spectral shapes update: $\sigma|A, X$

Exponentiated gradient ascent
Evaluation

Onset F1

note

time
Evaluation

Onset F1

note

time
Results

MAPS Corpus

Onset F1

80
70
60
50

[Valentin et al. 2010]
Results

MAPS Corpus

Onset F1

80
70
60
50

58.3

O’Hanlon
2014

[Valentin et al. 2010]
Results

MAPS Corpus

Onset F1

58.3

68.6

O’Hanlon 2014

Benetos 2014

[Valentin et al. 2010]
Results

MAPS Corpus

Onset F1

50 60 70 80

58.3 68.6 69.0

O’Hanlon 2014  Benetos 2014  Vincent 2013

[Valentin et al. 2010]
MAPS Corpus

Onset F1

- O’Hanlon 2014: 58.3
- Benetos 2014: 68.6
- Vincent 2013: 69.0
- Our System*: 82.1

[Valentin et al. 2010]
[Berg-Kirkpatrick et al. 2014]
Transcription
Transcription

Reference
Transcription

Reference

Predicted
Resynthesized Examples
Resynthesized Examples

Grieg input
Resynthesized Examples

Grieg input
Grieg resynth piano
Resynthesized Examples

Grieg input
Grieg resynth piano
Grieg resynth guitar