SEQ$^3$: Differentiable Sequence-to-Sequence-to-Sequence Autoencoder for Unsupervised Abstractive Sentence Compression

Christos Baziotis, Ion Androutsopoulos, Ioannis Konstas, Alexandros Potamianos

NAACL-HLT 2019, Minneapolis, USA
Introduction

Machine Translation

the big black cat ...

η μεγάλη μαύρη γάτα...

Dialogue

A: What do you want to do tonight?

B: Let's go for a movie!

Text to Code

sort a list of numbers

for i in range(len(A)):
    min_idx = i
    for j in range(i+1, len(A)):
        if A[min_idx] > A[j]:
            min_idx = j

Sentence Compression

Text to Tree

the big black cat ...

SEQ³ Autoencoder
**Introduction**

**Machine Translation**
- The big black cat ...
- η μεγάλη μαύρη γάτα...

**Text to Tree**
- The big black cat ...

**Dialogue**
- A: What do you want to do tonight?
- B: Let's go for a movie!

**Text to Code**
- Sort a list of numbers
  ```python
  for i in range(len(A)):
      min_idx = i
      for j in range(i+1, len(A)):
          if A[min_idx] > A[j]:
              min_idx = j
  ```

**Sentence Compression**

**SEQ³: Sequence-to-Sequence-to-Sequence Autoencoder**

**Input Sentence**
- ── ── ── ── ──
- ── ── ── ── ──

**Compression**
- ── ── ── ── ──

**Reconstruction**
- ── ── ── ── ──
Unsupervised Models for Language

Vanilla Autoencoders

\[ x_1, x_2, \ldots, x_N \rightarrow \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N \]
Unsupervised Models for Language

Vanilla Autoencoders

$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \rightarrow \mathbf{\hat{x}}_1, \mathbf{\hat{x}}_2, \ldots, \mathbf{\hat{x}}_N$

Discrete Latent Variable Autoencoders

$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \rightarrow \mathbf{\hat{x}}_1, \mathbf{\hat{x}}_2, \ldots, \mathbf{\hat{x}}_N$

+ Model the **discreteness** of language
  - Sampling is **not differentiable**
  - REINFORCE: sample **inefficient** and **unstable**
 Contributions

<table>
<thead>
<tr>
<th>Model</th>
<th>Supervision</th>
<th>Abstractive</th>
<th>Differentiable</th>
<th>Latent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miao &amp; Blunsom (2016)</td>
<td>semi</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Wang &amp; Lee (2018)</td>
<td>weak</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fevry &amp; Phang (2018)</td>
<td>none</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>SEQ³</td>
<td>none</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**SEQ³ Features** (+ contributions)

- Fully **unsupervised** and **abstractive**
- Fully **differentiable** (continuous approximations)
- **Topic**-grounded compressions
  - **Human-readable** compressions via **LM prior**
  - **User-defined** flexible compression ratio

**SOTA in unsupervised sentence compression**
Reconstruction loss: distill input into the latent sequence
LM Prior loss: human-readable compressions
Topic loss: similar topic as input
Length constraints: user-defined shorter

Compressor
Encoder Decoder
\(e_1^s\) \(e_2^s\) \(...\) \(e_N^s\)
\(x_1\) \(x_2\) \(x_N\)
Overview

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter

Compressor

Encoder

Decoder

\[ e \]

\[ \text{BOS} \]

\[ x_1 \]

\[ x_2 \]

\[ x_N \]

\[ e_1^s \]

\[ e_2^s \]

\[ \ldots \]

\[ e_N^s \]
Overview

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter length

\[ e_{\text{BOS}} \]

\[ x_1 \rightarrow x_2 \rightarrow x_N \]

Compressor
Encoder
Decoder

\[ e_{\text{BOS}} \]

\[ e_1 \rightarrow e_2 \rightarrow \ldots \rightarrow e_N \]
**Overview**

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter length

---

Compressor

Encoder

Decoder

\[ e^c_1 \]

\[ y_1 \]

\[ x_1 \]

\[ x_2 \]

\[ x_N \]

\[ e^s_1 \]

\[ e^s_2 \]

\[ \ldots \]

\[ e^s_N \]

\[ e_{BOS} \]
**Overview**

- **Reconstruction loss:** distill input into the latent sequence
- **LM Prior loss:** human-readable compressions
- **Topic loss:** similar topic as input
- **Length constraints:** user-defined shorter

---

**SEQ^3 Autoencoder**

Baziotis et al.
**SEQ³ Overview**

Reconstruction loss:
- distill input into the latent sequence

LM Prior loss:
- human-readable compressions

Topic loss:
- similar topic as input

Length constraints: user-defined shorter

\[ e_{BOS} \]

\[ x_1 \]

\[ x_2 \]

\[ x_N \]

\[ e_1^s \]

\[ e_2^s \]

\[ \ldots \]

\[ e_N^s \]

\[ e_{BOS} \]

\[ e_1^c \]

\[ e_2^c \]

\[ \ldots \]

\[ e_N^c \]

\[ y_1 \]

Compressor

Encoder

Decoder

Baziotis et al.
Overview

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined, shorter

SEQ$^3$ Autoencoder
Overview

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter

\[ e_{BOS} x_1 x_2 \]

\[ e_{1} \]

\[ e_{2} \]

\[ e_{N} \]

\[ y_{1} \]

\[ y_{2} \]

Compressor

Encoder

Decoder

\[ x_{1} \]

\[ x_{2} \]

\[ x_{N} \]
**SEQ$^3$ Overview**

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter

![Diagram](image)
**Overview**

SEQ³ Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter

\[
x_1 \rightarrow \epsilon_{BOS} \rightarrow x_2 \rightarrow \ldots \rightarrow x_N \rightarrow e_1^c \rightarrow \epsilon_{M-1} \rightarrow \ldots \rightarrow \epsilon_M \rightarrow y_1 \rightarrow \ldots \rightarrow y_M \rightarrow e_{EOS} \rightarrow e_1^s \rightarrow \ldots \rightarrow e_N^s \rightarrow \epsilon_1^s \rightarrow \epsilon_2^s \rightarrow \ldots \rightarrow \epsilon_M^s
\]

Baziotis et al.
SEQ³ Overview

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined shorter

Compressor

Reconstructor

Encoder

Decoder

Encoder

Decoder

Baziotis et al.
**Overview**

Reconstruction loss: distill input into the latent sequence

LM Prior loss: human-readable compressions

Topic loss: similar topic as input

Length constraints: user-defined, shorter

\[ e_{BOS} \]
\[ e_1 \]
\[ e_{M-1} \]

\[ y_1 \]
\[ y_2 \]
\[ y_M \]

Compressor

Reconstructor

Encoder
Decoder

Encoder
Decoder

Encoder
Decoder

Encoder
Decoder

\[ x_1 \]
\[ x_2 \]
\[ x_N \]

\[ \hat{x}_1 \]
\[ \hat{x}_2 \]
\[ \hat{x}_N \]
**Overview**

- **Reconstruction loss**: distill input into the latent sequence

### Reconstruction Loss

Minimize input reconstruction error:

\[ L_R(x, \hat{x}) = -\sum_{i=1}^{N} \log p_R(\hat{x}_i = x_i) \]
- **Reconstruction** loss: distill input into the latent sequence
- **LM Prior** loss: human-readable compressions
**Overview**

- **Reconstruction loss**: distill input into the latent sequence
- **LM Prior loss**: human-readable compressions

---

**LM Prior Loss**

Minimize $D_{KL}$ between Compressor and LM:

$$L_P = \frac{1}{M} \sum_{t=1}^{M} D_{KL}(p_C(y_t|y_{<t}, x) \parallel p_{LM}(y_t|y_{<t}, x))$$

---

**SEQ³ Autoencoder**

Baziotis et al.
**Overview**

- **Reconstruction** loss: **distill** input into the latent sequence
- **LM Prior** loss: **human-readable** compressions

**LM Prior Loss**

Minimize $D_{KL}$ between Compressor and LM:

$$L_P = \frac{1}{M} \sum_{t=1}^{M} D_{KL}(p_C(y_t|y_{<t}, x) \parallel p_{LM}(y_t|y_{<t}))$$
**Overview**

- **Reconstruction loss**: distill input into the latent sequence
- **LM Prior loss**: human-readable compressions
- **Topic loss**: similar topic as input

---

**Topic Loss**

\[ \mathbf{v}^x: \text{IDF-weighted average of } e_i^s \]
**Overview**

- **Reconstruction loss**: distill input into the latent sequence
- **LM Prior loss**: human-readable compressions
- **Topic loss**: similar topic as input

---

**Topic Loss**

\[ \mathbf{v}^x: \text{IDF-weighted average of } e^s_i \]

\[ \mathbf{v}^y: \text{average of } e^c_i \]
**Overview**

- **Reconstruction** loss: **distill** input into the latent sequence
- **LM Prior** loss: **human-readable** compressions
- **Topic** loss: similar **topic** as input

**Topic Loss**

\[ v^x: \text{IDF-weighted average of } e^s_i \]
\[ v^y: \text{average of } e^c_i \]
\[ L_T = 1 - \cos(v^x, v^y) \]
**SEQ³ Overview**

- **Reconstruction** loss: *distill* input into the latent sequence
- **LM Prior** loss: *human-readable* compressions
- **Topic** loss: similar *topic* as input
- **Length** constraints: user-defined *shorter* length

---

**Length Constraints**

1. **Length-aware** decoder initialization
2. **Countdown** inputs
3. Explicit *length penalty*

---

Diagram showing the flow of input compression through the **Compressor**, **Encoder**, **Decoder**, and **Reconstructor**.
Differentiable Sampling

**Straight-Through + Gumbel-softmax**

(Bengio et al., 2013, Maddison et al., 2017; Jang et al., 2017)

**Forward-pass:** **Discrete** embedding

\[
\text{argmax}\left(\frac{a_i + \xi_i}{\tau}\right)
\]

**Backward-pass:** **Mixture** of embeddings

\[
\text{softmax}\left(\frac{a_i + \xi_i}{\tau}\right)
\]
Differentiable Sampling

**Straight-Through + Gumbel-softmax**
(Bengio et al., 2013; Maddison et al., 2017; Jang et al., 2017)

**Forward-pass:** **Discrete** embedding

\[
\text{argmax}(\frac{(a_i + \xi_i)/\tau}{e})
\]

**Backward-pass:** **Mixture** of embeddings

\[
\text{softmax}(\frac{(a_i + \xi_i)/\tau}{\tilde{e}})
\]

\[
\nabla_\theta e \approx \nabla_\theta \tilde{e}
\]
Differentiable Sampling

Straight-Through + Gumbel-softmax
(Bengio et al., 2013, Maddison et al., 2017; Jang et al., 2017)

Forward-pass: **Discrete** embedding

\[
\text{argmax}\left(\frac{(a_i + \xi_i)}{\tau}\right)
\]

Backward-pass: **Mixture** of embeddings

\[
\text{softmax}\left(\frac{(a_i + \xi_i)}{\tau}\right)
\]
Experimental Setup

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigaword (English)</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>(source sentences)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUC-2003</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>DUC-2004</td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>

Training

- Train LM (LM prior) → Train $\text{SEQ}^3$
- **Never** exposed to target sentences (compressions)
- Vocabulary: 15K most frequent words in source sentences

Metrics

- Average F1 of ROUGE-1, ROUGE-2, ROUGE-L
Results on Gigaword

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Model</th>
<th>R-1</th>
<th>R-2</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised</td>
<td>Lead-8 (Rush et al., 2015)</td>
<td>21.86</td>
<td>7.66</td>
<td>20.45</td>
</tr>
<tr>
<td></td>
<td>Pretrained Generator (Wang &amp; Lee, 2018)</td>
<td>21.26</td>
<td>5.60</td>
<td>18.89</td>
</tr>
<tr>
<td></td>
<td>SEQ³</td>
<td>25.39</td>
<td>8.21</td>
<td>22.68</td>
</tr>
</tbody>
</table>

**Table**: Results on (English) Gigaword for sentence compression.
### Results on Gigaword

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Model</th>
<th>R-1</th>
<th>R-2</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised</td>
<td>LEAD-8 (Rush et al., 2015)</td>
<td>21.86</td>
<td>7.66</td>
<td>20.45</td>
</tr>
<tr>
<td></td>
<td>Pretrained Generator (Wang &amp; Lee, 2018)</td>
<td>21.26</td>
<td>5.60</td>
<td>18.89</td>
</tr>
<tr>
<td></td>
<td>$\text{SEQ}^3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>25.39</strong></td>
<td><strong>8.21</strong></td>
<td><strong>22.68</strong></td>
</tr>
<tr>
<td>Supervised</td>
<td>ABS (Rush et al., 2015)</td>
<td>29.55</td>
<td>11.32</td>
<td>26.42</td>
</tr>
<tr>
<td></td>
<td>SEASS (Zhou et al., 2017)</td>
<td>36.15</td>
<td>17.54</td>
<td>33.63</td>
</tr>
<tr>
<td></td>
<td>words-lvt5k-1sent (Nallapati et al., 2016)</td>
<td>36.40</td>
<td>17.70</td>
<td>33.71</td>
</tr>
</tbody>
</table>

Table: Results on (English) Gigaword for sentence compression.
## Ablation

<table>
<thead>
<tr>
<th>Model</th>
<th>R-1</th>
<th>R-2</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ$^3$ (Full)</td>
<td>25.39</td>
<td>8.21</td>
<td>22.68</td>
</tr>
<tr>
<td>SEQ$^3$ w/o LM</td>
<td>24.48 (-0.91)</td>
<td>6.68 (-1.53)</td>
<td>21.79 (-0.89)</td>
</tr>
<tr>
<td>SEQ$^3$ w/o TOPIC</td>
<td>3.89</td>
<td>0.10</td>
<td>3.75</td>
</tr>
</tbody>
</table>

**Table:** Ablation results on Gigaword.

Both topic and LM losses work in **synergy**

- **LM** prior loss: **how** words should be included
- **Topic** loss: **what** words to include
the central election commission (cec) on monday decided that
taiwan will hold another election of national assembly members
on may #.

GOLD  national <unk> election scheduled for may

SEQ³  the central election commission (cec) announced elections __

Dave Bassett resigned as manager of struggling English prem-

GOLD  forest manager bassett quits

SEQ³  Dave Bassett resigned as manager of struggling English premier

league side UNK forest on knocked round press
Conclusions and Future Work

Conclusions

- Fully **differentiable** seq2seq2seq (SEQ$^3$) autoencoder
- SOTA in unsupervised abstractive sentence compression
- **Topic** loss is essential for convergence
- **LM prior** improves **readability**

Next Step: unsupervised machine translation
Conclusions and Future Work

Conclusions

- Fully **differentiable** seq2seq2seq ($\text{SEQ}^3$) autoencoder
- SOTA in unsupervised abstractive sentence compression
- **Topic** loss is essential for convergence
- **LM prior** improves **readability**

Next Step: unsupervised machine translation

---

**Machine Translation**

- the big black cat ...
- η μεγάλη μαύρη γάτα...

**Dialogue**

- A: What do you want to do tonight?
- B: Let’s go for a movie!

**Text to Code**

- sort a list of numbers
  - for i in range(len(A)): 
    - min_idx = i 
    - for j in range(i+1, len(A)): 
        - min_idx = j 

**Sentence Compression**

- Baziotis et al.
Questions?

Source code

👉 https://github.com/cbaziotis/seq3

Contact me

✉️ christos.baziotis@gmail.com

🐦 @cbaziotis
Bonus Slides
**Soft-argmax:** Weighted sum of embeddings from peaked softmax (Goyal et al., 2017)

- **logits** $a_i$
- **softmax** ($a_i / \tau$)
- **Embeddings** $\bar{e}$
**Soft-argmax**: Weighted sum of embeddings from peaked softmax (Goyal et al., 2017)

\[
\text{logits} \rightarrow \text{Soft-argmax} \approx \text{Gumbel-softmax}
\]

**Gumbel-Softmax**

**Gumbel-max trick:**

\[
y \sim \text{softmax}(a_i) \\
= \text{argmax}(a_i + \xi_i), \quad \xi_i \sim \text{Gumbel}
\]

**Gumbel-softmax relaxation:**

\[
\hat{y} = \text{softmax}(a_i + \xi_i), \quad \xi_i \sim \text{Gumbel}
\]
**Soft-argmax**: Weighted sum of embeddings from peaked softmax (Goyal et al., 2017)

- **Softmax** function formula: \( \text{softmax}(\frac{a_i}{\tau}) \)
- Logits: \( a_i \)

**Gumbel-softmax**: Differentiable approximation to sampling (Maddison et al., 2017; Jang et al., 2017)

- Gumbel distribution: \( \xi_i \sim \text{Gumbel} \)
- Gumbel noise addition: \( \frac{a_i + \xi_i}{\tau} \)
**Soft-argmax:** Weighted sum of embeddings from peaked softmax 
(Goyal et al., 2017)

\[
\text{logits } a_i
\]

\[
\text{softmax}\left(\frac{a_i}{\tau}\right)
\]

\[
\bar{e}
\]

**Gumbel-softmax:** Differentiable approximation to sampling 
(Maddison et al., 2017; Jang et al., 2017)

**Straight-Through:** forward-pass: one-hot, backward-pass: soft 
(Bengio et al., 2013)

\[
\text{logits } a_i
\]

\[
\xi_i \sim \text{Gumbel}
\]

\[
\text{argmax}\left(\frac{a_i + \xi_i}{\tau}\right)
\]

\[
e
\]
Out of Vocabulary (OOV) Words

We copy OOV words using the approach of Fevry and Phang (2018). Simpler alternative to pointer networks (See et al., 2017).

1. We use a set of **special OOV tokens**: $OOV_1, OOV_2, \ldots, OOV_N$.
2. We replace the $i$th unknown word in the input with the $OOV_i$ token.
3. If all the OOV tokens are used, we use the generic UNK token.
4. In inference, we replace the special tokens with the original words.

### OOV Handling Example

<table>
<thead>
<tr>
<th>RAW</th>
<th>“John arrived in Rome yesterday. While in Rome, John had fun.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>INPUT</td>
<td>“$OOV_1$ arrived in $OOV_2$ yesterday. While in $OOV_2$, $OOV_1$ had fun.”</td>
</tr>
<tr>
<td>OOVs</td>
<td>John, Rome</td>
</tr>
</tbody>
</table>
Temperature $\tau$ does not affect the forward pass, but it affects gradients.

1. Jang et al. (2017) anneal $\tau \rightarrow 0$.
2. Gulcehre et al. (2017) learn $\tau$:

$$\tau(h^c_t) = \frac{1}{\log(1 + \exp(w^\top \tau h^c_t)) + 1}$$

3. Havrylov & Titov (2017) tune bound $\tau_0$:

$$\tau(h^c_t) = \frac{1}{\log(1 + \exp(w^\top \tau h^c_t)) + \tau_0}$$

In our experiments the learned temperature lead to instability. We fix $\tau = 0.5$ following (Gu et al., 2018).
Implementation Details

Hyper-Parameters
- Encoders: 2-layer bidirectional LSTM with size 300
- Decoders: 2-layer unidirectional LSTM with size 300
- Embedding: initialize with 100d GloVe (Pennington et al., 2014)

Parameter Sharing
- **Tied encoders** of the compressor and reconstructor.
- **Shared embedding** layer for all encoders and decoders.
- **Tied embedding-output** layers of both decoders.
1 **Sample** target length \( M \).
Length Control

1. **Sample** target length $M$.
2. Decoder’s state **length-aware initialization**.
1 **Sample** target length $M$.
2 Decoder’s state **length-aware initialization**.
3 **Countdown** input.

Length Control

1. Sample target length $M$.
2. Decoder’s state **length-aware initialization**.
3. Countdown input.
1 **Sample** target length $M$.
2 Decoder’s state **length-aware initialization**.
3 **Countdown** input.
4 Explicit length **penalty**.

---

**Diagram**:
- **Input sequence**: $x_1, x_2, ..., x_N$.
- **Sample target length**: $M$.
- **Decoder state**: $y_1, y_{M-1}, y_M, y_{M+1}, y_{M+2}$.
- **Countdown input**: $<\text{EOS}>$.
- **Explicit length penalty**: $f(\cdot)$.
### Results on DUC Shared Tasks

<table>
<thead>
<tr>
<th>Model</th>
<th>R-1</th>
<th>R-2</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topiary</strong> (Zajic et al., 2007)</td>
<td>25.12</td>
<td>6.46</td>
<td>20.12</td>
</tr>
<tr>
<td>(Woodsend et al., 2010)</td>
<td>22.00</td>
<td>6.00</td>
<td>17.00</td>
</tr>
<tr>
<td><strong>ABS</strong> (Rush et al., 2015)</td>
<td>28.18</td>
<td>8.49</td>
<td>23.81</td>
</tr>
<tr>
<td><strong>PREFIX</strong></td>
<td>20.91</td>
<td>5.52</td>
<td>18.20</td>
</tr>
<tr>
<td><strong>SEQ³</strong> (Full)</td>
<td>22.13</td>
<td>6.18</td>
<td>19.3</td>
</tr>
</tbody>
</table>

**Table:** Results on the DUC-2004

<table>
<thead>
<tr>
<th>Model</th>
<th>R-1</th>
<th>R-2</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS (Rush et al., 2015)</td>
<td>28.48</td>
<td>8.91</td>
<td>23.97</td>
</tr>
<tr>
<td><strong>PREFIX</strong></td>
<td>21.3</td>
<td>6.38</td>
<td>18.82</td>
</tr>
<tr>
<td><strong>SEQ³</strong> (Full)</td>
<td>20.90</td>
<td>6.08</td>
<td>18.55</td>
</tr>
</tbody>
</table>

**Table:** Results on the DUC-2003
the american sailors who thwarted somali pirates flew home to the u.s. on wednesday but without their captain, who was still aboard a navy destroyer after being rescued from the hijackers.

us sailors who thwarted pirate hijackers fly home

the american sailors who foiled somali pirates flew home after crew hijacked.