

## Neural Headline Generation

### Our challenge in news-aggregator in Yahoo! JAPAN

Condensed headline is used to be shown in the Yahoo! JAPAN website as a hyperlink to a news article instead of the original headline and to allow its visitors to swiftly determine if they should read the article. Editing assistant system which shows candidates of condensed headline to editors would be useful.

### Condensed headline generation with encoder-decoder model

Recent automatic summarization systems depend heavily on machine learning which based on a lot of training examples. However, since the condensed headlines are manually written by human workers in addition to their original headlines, making them is costly, and therefore it is not easy to prepare them in large quantities.

Original Headline	Condensed Headline
<天皇陛下>退位後に「赤坂御用地」に転居 宮内庁検討 His Majesty the Emperor to move to Akasaka Palace after abdication - Imperial Household Agency plans	陛下 退位後は赤坂に転居案 His Majesty to move to Akasaka after abdication planned

## Proposed Method

### Step 1. Train baseline model

A baseline model is learned from pairs of headlines and their corresponding condensed headlines.

### Step 2. Generate pseudo headlines

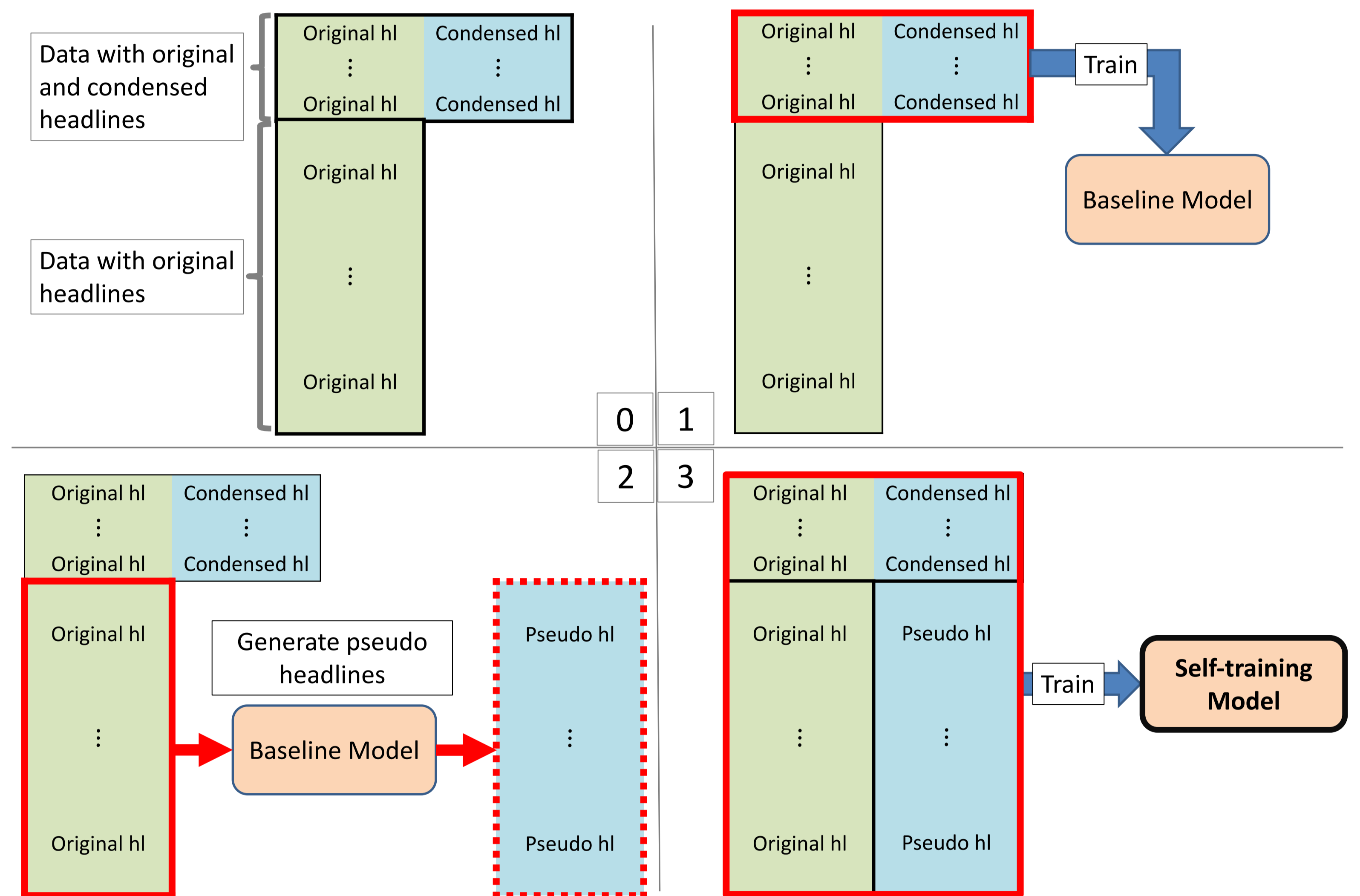
We run the baseline model to generate pseudo condensed headlines from the headlines without condensed ones.

### Step 3. Train proposed model

Our proposed model is learned from both original and pseudo condensed headlines.



Our contents partner provides articles (right) to us. Yahoo! JAPAN news editor team adds information such as related links to it, writes condensed headline which is 13 and half or less characters (13 multibyte characters and one singlebyte character) and displays it to the top page of Yahoo! JAPAN.



## Experiments

### ◆ Implementation and settings

- Generate condensed headline from original one
- Baseline method employs basic encoder-decoder model with an attention mechanism (with LSTM, character based). Developed with OpenNMT-Toolkit.

### ◆ Data

- Original and condensed headlines: about 100K
- Split into: Training 60%, Validation 20%, Test 20%
- Original headlines: 2M
- Compare number of unlabeled training examples from 200K to 2M
- ※ Gathered by web crawling

### ◆ Compared method

- Baseline method
- Pre-training method [Dai and Le, NIPS 2015]
- Proposed method (with 600K unlabeled headlines)

Original Headline	Gold Standard	Baseline (ROUGE-1)	Proposed (ROUGE-1)
ソフトバンクがロボット事業に参入、日本のロボット産業どうなる? SoftBank enters robot industry - what will become of robot industry in Japan?	日本のロボット産業 未来は? Robot industry in Japan - future?	日本ロボットロボットに参入 (0.462) Japan robot enters robot	ソフト B ロボット事業に参入 (0.385) Soft B enters robot business
<ボクシング>亀田ジムが新会長での再出発を検討 Boxing: Kameda gym plans to restart with new president	亀田ジム 新会長で再出発検討 Kameda gym plans to restart with new president	亀田ジム新会長で再出発検討 (1.000) Kameda gym plans to restart with new president	亀田ジム新会長で再出発検討 (1.000) Kameda gym plans to restart with new president
ドワンゴの入社試験有料化、その是非をどう考えるべきか Dwanggo charges for employment exam - how should we consider: good or bad?	入社試験「有料化」の是非は “Fee-charging” employment exam - good or bad?	ドワンゴ有料化その是非どう (0.462) Dwanggo charges for - how good or bad?	ドワンゴの入社試験有料化へ (0.615) Dwanggo charges for employment exam

### ◆ Results

**Proposed method significantly overcame baseline method. (p<0.05)**

ROUGE-1 result

	ROUGE-1
Baseline	0.571
Pre-Training(w/o attention weights)	0.520
Pre-Training(all weights)	0.503
Proposed (600K)	<b>0.574</b>

### ◆ Discussion

- Known problem in self-training  
Too much unlabeled training examples generated from an initial model overwhelms the labeled examples (right figure). This implies that if we could filter out the inferior training examples and choose only good training examples, we could improve the accuracy.
- Consideration of news description  
Human editors carefully read the body of articles as well as their headlines in order to make their condensed headlines, and therefore information written in the body should also be considered to generate better condensed headlines.

