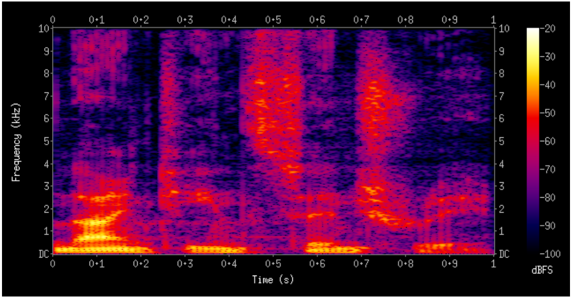
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Speech recognition

Постановка задачи

На вход поступает спектрограмма аудиозаписи с человеческой речью.



Спектрограмма — последовательность из векторов, т.е. матрица $n \times t$.
Нужно вывести результат распознавания в виде $\{a, ..., z, apostrophe, space\}^k$, где k заранее неизвестно.

BRNN

- Задачу можно решать с помощью bidirectional RNN.

BRNN

- Задачу можно решать с помощью bidirectional RNN.
- Функция активации *clipped ReLu*:
  $g(z) = \min(20, \max(0, z))$

BRNN

- Задачу можно решать с помощью bidirectional RNN.
- Функция активации *clipped ReLu*:
  $g(z) = \min(20, \max(0, z))$
- Регуляризация с помощью dropout 5-10%,
  накладывание шумов на аудиозаписи.

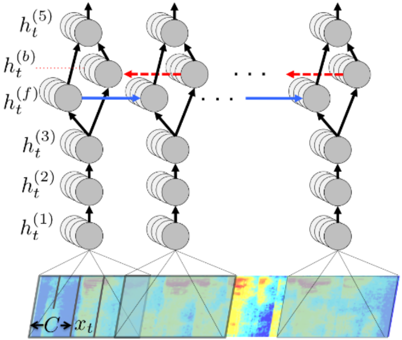
BRNN

- Задачу можно решать с помощью bidirectional RNN.
- Функция активации *clipped ReLu*:
  $g(z) = \min(20, \max(0, z))$
- Регуляризация с помощью dropout 5-10%, накладывание шумов на аудиозаписи.
- CTC-loss.

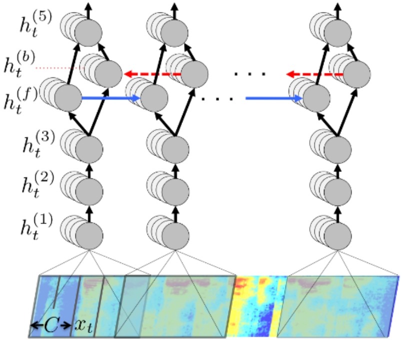
BRNN

- Задачу можно решать с помощью bidirectional RNN.
- Функция активации *clipped ReLu*:
  $g(z) = \min(20, \max(0, z))$
- Регуляризация с помощью dropout 5-10%, накладывание шумов на аудиозаписи.
- CTC-loss.
- Если у спектрограммы размер $n \times t$, то на выходе получаем t векторов с вероятностями для каждой метки.

BRNN

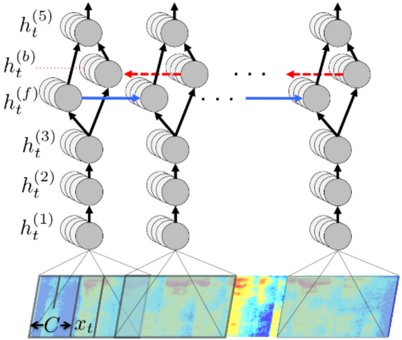


BRNN



- Что если спектрограмма состоит из нескольких тысяч векторов, а в аудиозаписи всего одно слово?

BRNN



- Что если спектрограмма состоит из нескольких тысяч векторов, а в аудиозаписи всего одно слово?
- Для получения конечного результата необходимо преобразовать выход модели.

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Пример: $ssssss \perp\perp\perp ppp \perp eeee \perp eeccchhhh$

1. $s \perp p \perp e \perp e \perp c \perp h \perp$

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Пример: $ssssss \perp\perp\perp ppp \perp eeee \perp eeccchhhh$

1. $s \perp p \perp e \perp e \perp c \perp h \perp$
2. $speech$

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Пример: $ssssss \perp\perp\perp ppp \perp eeee \perp eeccchhhh$

1. $s \perp p \perp e \perp e \perp c \perp h \perp$
2. $speech$

Порядок важен!

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Пример: $ssssss \perp\perp\perp ppp \perp eeee \perp eeccchhhh$

1. $s \perp p \perp e \perp e \perp c \perp h \perp$
2. $speech$

Порядок важен!

- Пусть $\mathcal{D}(w)$ преобразует слово w, выполняя шаги 2 и 3.

Преобразование выхода

Множество меток необходимо расширить: теперь это $\{a, ..., z, apostrophe, space, \perp\}$.

1. Каждому вектору сопоставим наиболее вероятный символ
2. Удаляем повторяющиеся идущие подряд символы
3. Удаляем $\perp$

Пример: $ssssss \perp\perp\perp ppp \perp eeee \perp eeccchhhh$

1. $s \perp p \perp e \perp e \perp c \perp h \perp$
2. $speech$

Порядок важен!

- Пусть $\mathcal{D}(w)$ преобразует слово w, выполняя шаги 2 и 3.
- Способ неплохой, но можно лучше.

Чем плох этот способ?

RNN output	Decoded Transcription
what is the weather like in bostin right now	what is the weather like in boston right now
prime miniter nerenr modi	prime minister narendra modi
arther n tickets for the game	are there any tickets for the game

Как решить эту проблему?

- Пусть S — спектрограмма размера $n \times t$.

Как решить эту проблему?

- Пусть S — спектрограмма размера $n \times t$.
- Пусть $\pi \in \{a, ..., z, apostrophe, space, \perp\}^k$ — путь длины k, $w \in \{a, ..., z, apostrophe, space\}^k$ — слово длины k.

Как решить эту проблему?

- Пусть S — спектрограмма размера $n \times t$.
- Пусть $\pi \in \{a, ..., z, apostrophe, space, \perp\}^k$ — путь длины k, $w \in \{a, ..., z, apostrophe, space\}^k$ — слово длины k.
- Пусть π — путь длины $k \leq t$: $P(\pi|S) = \prod_{i=1}^{k} y_{\pi_i}^{(i)}$, где $y^{(i)}$ — векторы вероятности.

Как решить эту проблему?

- Пусть S — спектрограмма размера $n \times t$.
- Пусть $\pi \in \{a, ..., z, apostrophe, space, \perp\}^k$ — путь длины k, $w \in \{a, ..., z, apostrophe, space\}^k$ — слово длины k.
- Пусть π — путь длины $k \le t$: $P(\pi|S) = \prod_{i=1}^{k} y_{\pi_i}^{(i)}$, где $y^{(i)}$ — векторы вероятности.
- Пусть w — слово длины $k \le t$

$$\mathbb{P}(w|S) \stackrel{def}{=} \sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=w}} P(\pi|S)$$

Как считать эту вероятность?

- Слово $w = w_1 w_2 ... w_k$.

Как считать эту вероятность?

- Слово $w = w_1 w_2 ... w_k$.
- Введем путь $l = \perp w_1 \perp w_2 \perp ... w_k \perp$.

Как считать эту вероятность?

- Слово $w = w_1 w_2 ... w_k$.
- Введем путь $l = \perp w_1 \perp w_2 \perp ... w_k \perp$.
- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

Как считать эту вероятность?

- Слово $w = w_1 w_2 ... w_k$.
- Введем путь $l = \perp w_1 \perp w_2 \perp ... w_k \perp$.
- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

$$a_i^{(j)} = \sum_{\substack{\pi \\ |\pi|=j \\ \mathcal{D}(\pi)=\mathcal{D}(w) \\ \pi_j = l_i}} P(\pi|S)$$

Как считать эту вероятность?

- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

Как считать эту вероятность?

- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.
- $a_1^{(1)} = y_\perp^{(1)}$

Как считать эту вероятность?

- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.
- $a_1^{(1)} = y_{\perp}^{(1)}$
- $a_2^{(1)} = y_{w_1}^{(1)}$

Как считать эту вероятность?

- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.
- $a_1^{(1)} = y_{\perp}^{(1)}$
- $a_2^{(1)} = y_{w_1}^{(1)}$
- $\forall i > 2 : a_i^{(1)} = 0$

Как считать эту вероятность?

- Динамическое программирование по l: $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.
- $a_1^{(1)} = y_\perp^{(1)}$
- $a_2^{(1)} = y_{w_1}^{(1)}$
- $\forall i > 2 : a_i^{(1)} = 0$
- $\mathbb{P}(w|S) = a_{2k}^{(t)} + a_{2k+1}^{(t)}$

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{\perp}^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- $\mathcal{D}(\pi_{1:j-1}) = \mathcal{D}(l_{1:i}) = \mathcal{D}(l_{1:i-1})$

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- $\mathcal{D}(\pi_{1:j-1}) = \mathcal{D}(l_{1:i}) = \mathcal{D}(l_{1:i-1})$
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда последний символ $\mathcal{D}(l_{1:j-1})$ равен $c \Rightarrow c = l_{i-1}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- $\mathcal{D}(\pi_{1:j-1}) = \mathcal{D}(l_{1:i}) = \mathcal{D}(l_{1:i-1})$
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда последний символ $\mathcal{D}(l_{1:j-1})$ равен $c \Rightarrow c = l_{i-1}$.
- Либо $\pi_{j-1} = \perp$, либо $\pi_{j-1} = l_{i-1}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- $\mathcal{D}(\pi_{1:j-1}) = \mathcal{D}(l_{1:i}) = \mathcal{D}(l_{1:i-1})$
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда последний символ $\mathcal{D}(l_{1:j-1})$ равен $c \Rightarrow c = l_{i-1}$.
- Либо $\pi_{j-1} = \perp$, либо $\pi_{j-1} = l_{i-1}$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = \perp$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_\perp^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- $\mathcal{D}(\pi_{1:j-1}) = \mathcal{D}(l_{1:i}) = \mathcal{D}(l_{1:i-1})$
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда последний символ $\mathcal{D}(l_{1:j-1})$ равен $c \Rightarrow c = l_{i-1}$.
- Либо $\pi_{j-1} = \perp$, либо $\pi_{j-1} = l_{i-1}$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = \perp$.
- $a_{i-1}^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_{i-1}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = l_{i-2} \neq \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = l_{i-2} \neq \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = l_{i-2} \neq \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$, но $l_i = l_{i-2} \Rightarrow c = l_i$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = l_{i-2} \neq \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$, но $l_i = l_{i-2} \Rightarrow c = l_i$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_i$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $l_i = l_{i-2} \neq \perp$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$, но $l_i = l_{i-2} \Rightarrow c = l_i$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_i$.
- $a_{i-1}^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = \perp$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_i$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

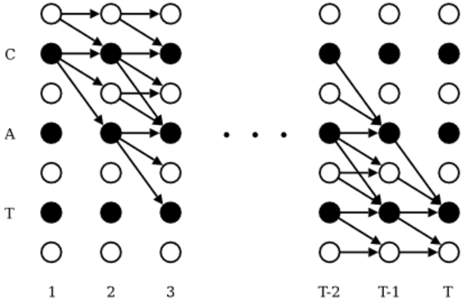
- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_i$.
- $a_{i-1}^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = \perp$.

Как считать эту вероятность?

$a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

- Если $\perp \neq l_i \neq l_{i-2}$, то $a_i^{(j)} = \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)}$.
- Пусть π учитывается в $a_i^{(j)}$.
- Пусть $\pi_{j-1} = c \neq \perp$. Тогда $c = l_i$, либо $c = l_{i-2}$.
- $a_i^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_i$.
- $a_{i-1}^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = \perp$.
- $a_{i-2}^{(j-1)}$ отвечает только за те пути π, что $\pi_{j-1} = l_{i-2}$

Как считать эту вероятность?



$$a_i^{(j)} = \begin{cases} \left(a_i^{(j-1)} + a_{i-1}^{(j-1)}\right) y_{l_i}^{(j)} & \text{если } l_i = \perp \text{ или } l_i = l_{i-2} \\ \left(a_i^{(j-1)} + a_{i-1}^{(j-1)} + a_{i-2}^{(j-1)}\right) y_{l_i}^{(j)} & \text{иначе} \end{cases}$$

n-gram

- Пусть есть большой корпус текста, c — n-грамма.
- $P_{lm}(c) = \frac{\mathrm{cnt}(c)}{\mathrm{cnt}(c_1 c_2 \ldots c_{n-1})}$

n-gram

- Пусть есть большой корпус текста, c — n-грамма.
- $P_{lm}(c) = \frac{\text{cnt}(c)}{\text{cnt}(c_1 c_2 ... c_{n-1})}$
- Вероятность слова w:
  $P_{lm}(w) = P_{lm}(w_1 ... w_n) P_{lm}(w_2 ... w_{n+1}) ... P_{lm}(w_{k-n+1} ... w_k)$

n-gram

- Пусть есть большой корпус текста, c — n-грамма.
- $P_{lm}(c) = \frac{\text{cnt}(c)}{\text{cnt}(c_1 c_2 ... c_{n-1})}$
- Вероятность слова w:
  $P_{lm}(w) = P_{lm}(w_1 ... w_n) P_{lm}(w_2 ... w_{n+1}) ... P_{lm}(w_{k-n+1} ... w_k)$

$$Q(w) = \mathbb{P}(w) + \alpha P_{lm}(w) + \beta \, \text{cnt}(w)$$

n-gram

- Пусть есть большой корпус текста, c — n-грамма.
- $P_{lm}(c) = \frac{\text{cnt}(c)}{\text{cnt}(c_1 c_2 ... c_{n-1})}$
- Вероятность слова w:
  $P_{lm}(w) = P_{lm}(w_1 ... w_n) P_{lm}(w_2 ... w_{n+1}) ... P_{lm}(w_{k-n+1} ... w_k)$

$$Q(w) = \mathbb{P}(w) + \alpha P_{lm}(w) + \beta \, \text{cnt}(w)$$

RNN output	Decoded Transcription
what is the weather like in bostin right now	what is the weather like in boston right now
prime miniter nerenr modi	prime minister narendra modi
arther n tickets for the game	are there any tickets for the game

CTC-loss

Пусть $D = (S_i, w_i)_{i=1}^{d}$ — данные, на которых считается ошибка.

$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

CTC-loss

Пусть $D = (S_i, w_i)_{i=1}^{d}$ — данные, на которых считается ошибка.

$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

- Как посчитать градиент?

CTC-loss

Пусть $D = (S_i, w_i)_{i=1}^{d}$ — данные, на которых считается ошибка.

$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

- Как посчитать градиент?
- $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.

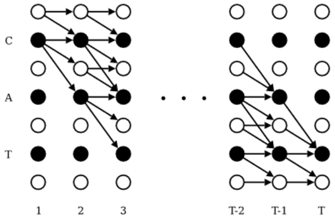
CTC-loss

Пусть $D = (S_i, w_i)_{i=1}^{d}$ — данные, на которых считается ошибка.

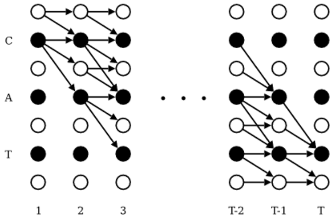
$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

- Как посчитать градиент?
- $a_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{1:i})$ и при этом $\pi_j = l_i$.
- $b_i^{(j)}$ — суммарная вероятность путей π длины j, таких, что $\mathcal{D}(\pi) = \mathcal{D}(l_{i:|l|})$ и при этом $\pi_1 = l_i$.

CTC-loss

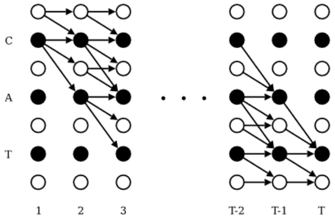


CTC-loss



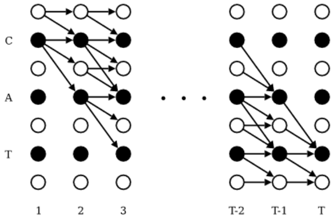
- Проставим $a_i^{(j)} = 0 \; \forall i < |l| - 2(t-j) - 1$ — верхний правый угол.

CTC-loss



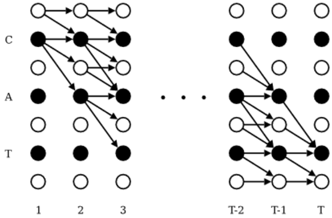
- Проставим $a_i^{(j)} = 0 \ \forall i < |l| - 2(t - j) - 1$ — верхний правый угол.
- Проставим $b_i^{(j)} = 0 \ \forall i > 2j$ — левый нижний угол.

CTC-loss



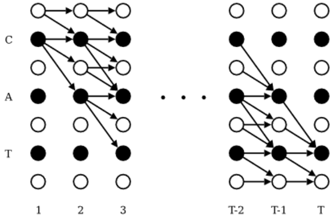
- Проставим $a_i^{(j)} = 0 \ \forall i < |l| - 2(t - j) - 1$ — верхний правый угол.
- Проставим $b_i^{(j)} = 0 \ \forall i > 2j$ — левый нижний угол.

CTC-loss



$$a_i^{(j)} b_i^{(j)} = y_{l_i}^{(j)} \cdot \sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=\mathcal{D}(l) \\ \pi_j = l_i}} P(\pi|S) = y_{l_i}^{(j)} \cdot \sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=\mathcal{D}(l) \\ \pi_j = l_i}} \prod_{u=1}^{t} y_{\pi_u}^{(u)}$$

CTC-loss



$$a_i^{(j)} b_i^{(j)} = y_{l_i}^{(j)} \cdot \sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=\mathcal{D}(l) \\ \pi_j = l_i}} P(\pi|S) = y_{l_i}^{(j)} \cdot \sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=\mathcal{D}(l) \\ \pi_j = l_i}} \prod_{u=1}^{t} y_{\pi_u}^{(u)}$$

$$\sum_{i=1}^{|l|} \frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} = \mathbb{P}(w|S)$$

CTC-loss

Нужно научиться брать производные такого вида: $\frac{\partial \mathcal{L}}{\partial y_i^{(j)}}$

CTC-loss

Нужно научиться брать производные такого вида: $\frac{\partial \mathcal{L}}{\partial y_i^{(j)}}$

$$\mathcal{L} = - \sum_{(S,w) \in D} \ln \left(\mathbb{P}(w|S) \right)$$

CTC-loss

Нужно научиться брать производные такого вида: $\frac{\partial \mathcal{L}}{\partial y_i^{(j)}}$

$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

То есть достаточно научиться брать производные такого вида: $\frac{\partial \mathbb{P}(w|S)}{\partial y_i^{(j)}}$

CTC-loss

Нужно научиться брать производные такого вида: $\frac{\partial \mathcal{L}}{\partial y_i^{(j)}}$

$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

То есть достаточно научиться брать производные такого вида: $\frac{\partial \mathbb{P}(w|S)}{\partial y_i^{(j)}}$

$$\sum_{i=1}^{|l|} \frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} = \mathbb{P}(w|S)$$

CTC-loss

Нужно научиться брать производные такого вида: $\frac{\partial \mathcal{L}}{\partial y_i^{(j)}}$

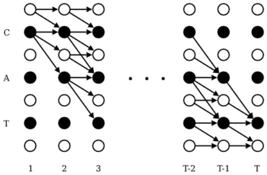
$$\mathcal{L} = -\sum_{(S,w)\in D} \ln\left(\mathbb{P}(w|S)\right)$$

То есть достаточно научиться брать производные такого вида: $\frac{\partial \mathbb{P}(w|S)}{\partial y_i^{(j)}}$

$$\sum_{i=1}^{|l|} \frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} = \mathbb{P}(w|S)$$

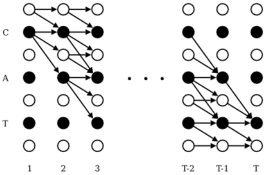
То есть достаточно научиться брать производные такого вида: $\frac{\partial}{\partial y_c^{(j)}} \left(\frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} \right)$

CTC-loss



Если $l_i \neq c$, то $\frac{\partial}{\partial y_c^{(j)}} \left(\frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} \right) = 0.$

CTC-loss



Если $l_i \neq c$, то $\frac{\partial}{\partial y_c^{(j)}} \left(\frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} \right) = 0.$

Иначе:

$$\frac{\partial}{\partial y_{l_i}^{(j)}} \left(\frac{a_i^{(j)} b_i^{(j)}}{y_{l_i}^{(j)}} \right) = \frac{\partial}{\partial y_{l_i}^{(j)}} \left(\sum_{\substack{\pi \\ |\pi|=t \\ \mathcal{D}(\pi)=\mathcal{D}(l) \\ \pi_j = l_i}} \prod_{u=1}^{t} y_{\pi_u}^{(u)} \right) = \frac{a_i^{(j)} b_i^{(j)}}{\left(y_{l_i}^{(j)}\right)^2}$$

Результаты

WER - word error rate

System	Clean (94)	Noisy (82)	Combined (176)
Apple Dictation	14.24	43.76	26.73
Bing Speech	11.73	36.12	22.05
Google API	6.64	30.47	16.72
wit.ai	7.94	35.06	19.41
Deep Speech	**6.56**	**19.06**	**11.85**

Список источников

- https://arxiv.org/abs/1412.5567
- https://www.cs.toronto.edu/~graves/icml_2006.pdf
- An Intuitive Explanation of Connectionist Temporal Classification