

МЕТОДИ РОЗПІЗНАВАННЯ РУКОПИСНОГО ВВОДУ ЗА ДОПОМОГОЮ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ

Шишков Олег Миколайович, КА-41м
к.т.н., доцент, Тимощук Оксана Леонідівна

- **Мета:** розробити систему для розпізнавання рукописного вводу
- **Об'єкт:** задача розпізнавання рукописного вводу.
- **Предмет:** розпізнавання рукописного вводу за допомогою рекурентних нейронних мереж з нейромережевою часовою класифікацією

Типи рукописних даних

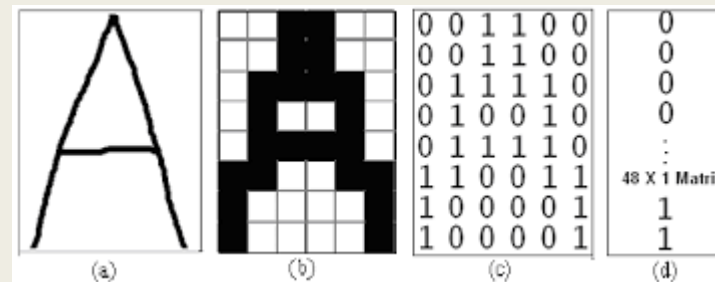
Online

- Послідовність точок
- Більше інформації
- Розпізнавання неперервного рукописного вводу



Offline

- Піксельне зображення
- Висока якість лише для друкованого тексту



Проблема розмітки даних

При використанні класичних методів машинного навчання необхідна додаткова поточкова розмітка даних. Ручна або автоматична розмітка робить розпізнавання менш точним та дорогим

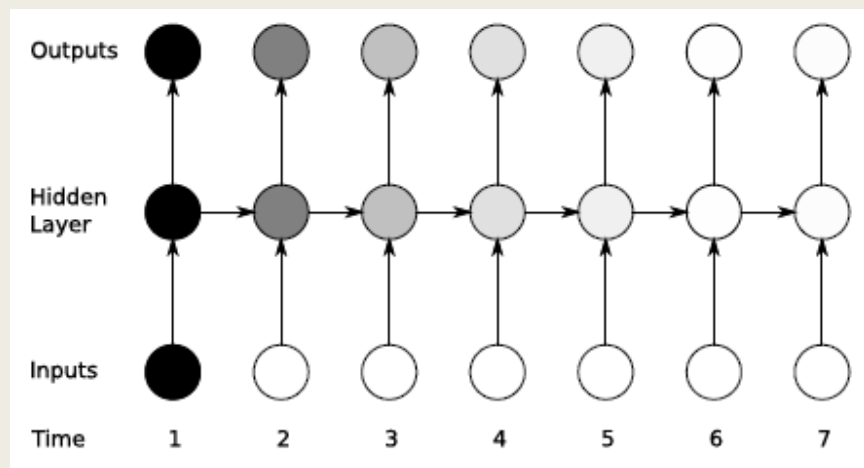
x	y	l
0.1	0.4	a
0.2	0.5	a
0.3	0.6	a
0.4	0.7	b
0.5	0.8	b
0.6	0.9	b

При використанні СТС відсутня необхідність додаткової розмітки, необхідний лише вихідний текст, який писав респондент

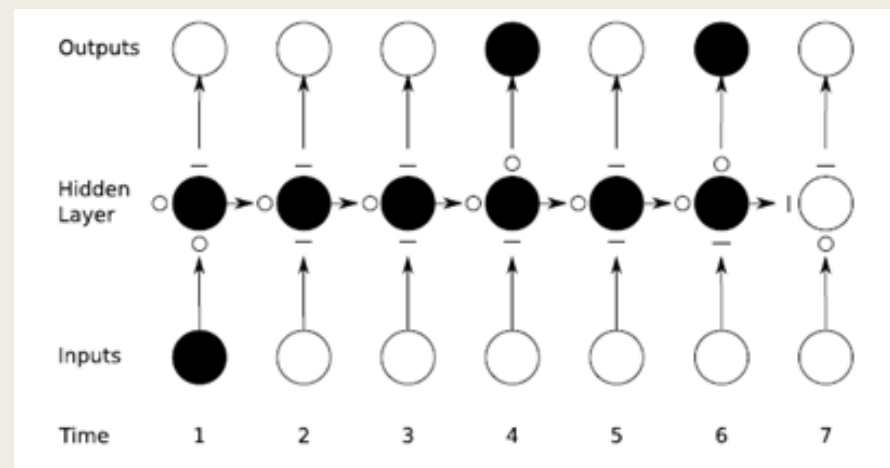
x	y	l
0.1	0.4	ab
0.2	0.5	
0.3	0.6	
0.4	0.7	
0.5	0.8	
0.6	0.9	

Архітектура РНН

Long short-term memory



Ілюстрація проблеми розмиття градієнту [1].
В звичайних РНН вхідні значення з часом
втрачають вплив



LSTM може зберігати необхідну інформацію [1].

[1]. A. Graves. Supervised Sequence Labelling with Recurrent Neural Networks. Textbook, Studies in Computational Intelligence, Springer, 2012

Алгоритм нейромережевої часової класифікації Connectionist Temporal Classification (CTC)

$$A = \{a, b, \dots, z\}$$

$$A' = A \cup \{blank\}$$

\mathbf{x} – вхідна послідовність

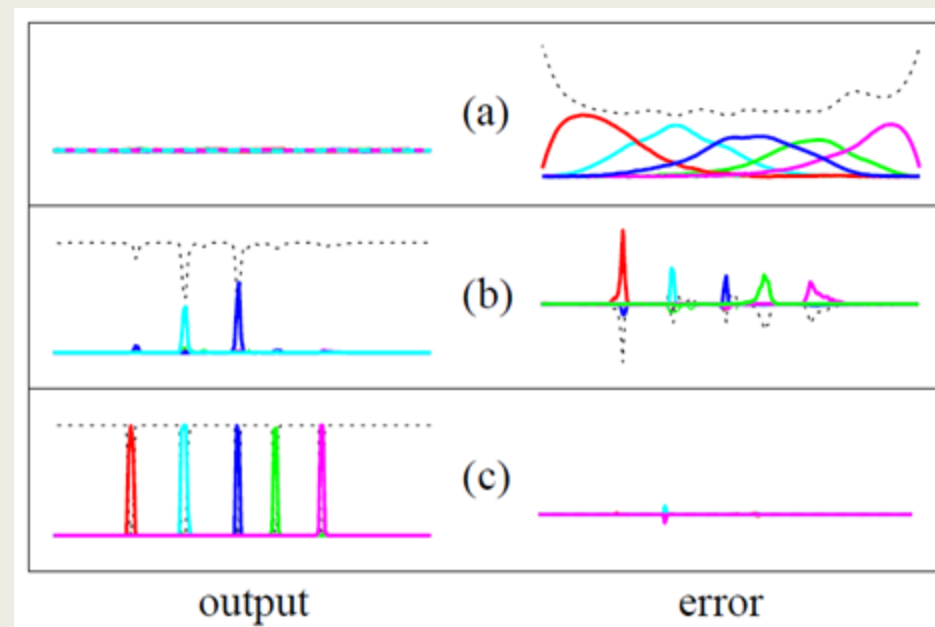
\mathbf{y} – виходи нейронної мережі

$\mathbf{l} \in A^*$ - правильне слово

$$\pi \in A'^T, T = |\mathbf{x}|$$

$$\mathcal{F}: A'^T \rightarrow A^{*\leq T} (- -aa - a - bb) = aab$$

$$p(\mathbf{l}|\mathbf{x}) = \sum_{\pi \in \mathcal{F}^{-1}(\mathbf{l})} p(\pi|\mathbf{x})$$



Ілюстрація процесу навчання нейронної мережі з CTC [1].

[1]. A. Graves. Supervised Sequence Labelling with Recurrent Neural Networks. Textbook, Studies in Computational Intelligence, Springer, 2012

Алгоритм прямого-звортного ходу

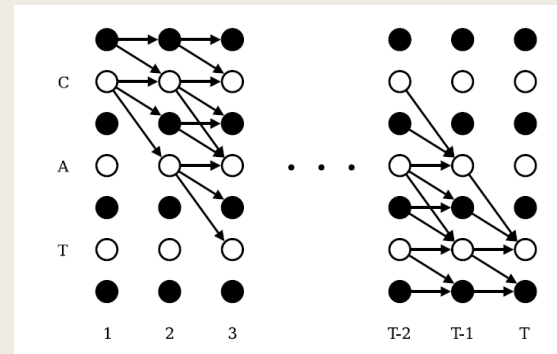
$$\alpha(t, u) = \sum_{\pi \in V(t, u)} \prod_{i=1}^t y_{\pi_i}^i$$

$$V(t, u) = \{\pi \in A'^t : \mathcal{F}(\pi) = \mathbf{l}_{1:\frac{u}{2}}, \pi_t = l'_u\}$$

$$p(\mathbf{l}|\mathbf{x}) = \alpha(T, U') + \alpha(T, U' - 1)$$

- $\alpha(1, 1) = y_b^1$
- $\alpha(1, 2) = y_{l_1}^1$
- $\alpha(1, u) = 0, \forall u > 2$

$$\alpha(t, u) = y_{l'_u}^t \sum_{i=f(u)}^u \alpha(t-1, i)$$



Алгоритм прямого-обратного хода

$$\beta(t, u) = \sum_{\pi \in W(t, u)} \prod_{i=1}^{T-t} y_{\pi_i}^{t+i}$$

$$\beta(t, u) = \sum_{i=u}^{g(u)} \beta(t+1, i) y_{l'_i}^{t+1}$$

$$\frac{\partial p(\mathbf{z}|\mathbf{x})}{\partial y_k^t} = -\frac{1}{p(\mathbf{z}|\mathbf{x}) y_k^t} \sum_{u \in B(\mathbf{z}, k)} \alpha(t, u) \beta(t, u)$$

Архітектура мережі

- Вхідний шар з 3 входами
- 2 BLSTM шари по 200 вузлів (100 forward + 100 backward)
- Вихідний шар Softmax на 27 виходів
- Функція помилки CTC

330627 вагових коефіцієнтів

- Декодування найімовірнішої траєкторії

$$\pi = \arg \max_{\pi^*} p(\pi^* | \mathbf{x})$$
$$l^* \approx \mathcal{F}(\pi^*)$$

Модернізація процесу пакетного навчання

Довгі послідовності заважають на початку навчання:

- Швидкість обробки одного пакету залежить від найдовної послідовності
- Більша імовірність перенавчитись або застрягти в локальному мінімумі

Варіант вирішення проблеми:

- Відсортувати все послідовності по довжині
- Розбити на декілька груп
- Навчати послідовно на кожній із груп до досягнення прийнятною помилки

Набори даних рукописного вводу

Unipen

14544 речень

75529 окремих слів

122628 літер

IAM On-Line

Handwriting
Database

1700 рукописних
фрагментів по 50 слів

86272 рукописних слів

11059 слів у словнику

221 респондентів

ICDAR 2003

Informal Competition for the
Recognition of On-line Words

13119 рукописних слів

884 слів у словнику

72 респондента

Результати роботи

- Алгоритм CTC адаптовано під паралельні обчислення на GPU
- Імплементовано алгоритм CTC на технології CUDA
- Досліджено та покращено процес пакетного навчання рекурентних нейронних мереж
- Натреновано та протестовано нейронну мережу для розпізнавання окремих слів

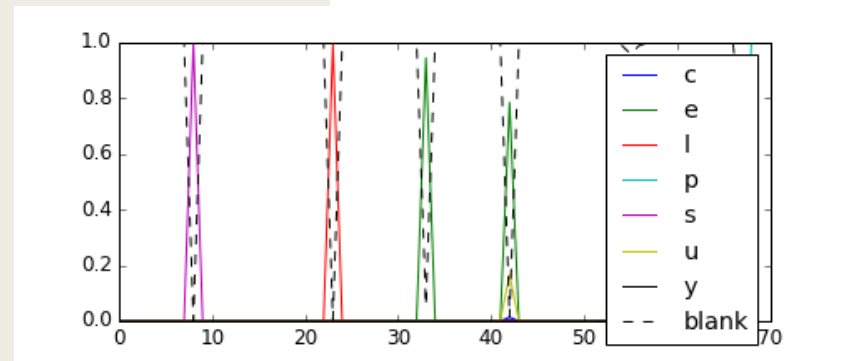
Результаты работы

Тестова вибірка: 6145 окремих рукописних слів

Результаты работы нейронной сети:

середня помилка на символ $= \frac{\text{кількість помилок}}{\text{загальна довжина слів}} = 0.082$

- 76% правильно розпізнаних слів
- 3.4% слів з однією помилкою
- 14% слів з двома помилками



Можливі шляхи подальших досліджень

- Використання іншої архітектури рекурентних нейронних мереж з пам'яттю. Наприклад, Gate Recurrent Unit (GRU)
- Застосування складніших алгоритмів декодування, що використовують інформацію про модель мови
- Розширення алфавіту додатковими символами, цифрами. Розпізнавання кириличного вводу

Дякую за увагу!