МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМ. ІГОРЯ СІКОРСЬКОГО»
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Дипломна робота на тему:

Використання методу довгої короткочасної пам'яті в задачах розпізнавання активності людини

Виконав:
Валентин Мельничук, КА-41
Науковий керівник:
к.ф.-м.н., доц. Яковлева А.П.

1. Вступ

Об'єкт дослідження

Алгоритм глибокого навчання - **довга короткочасна пам'ять (ДКЧП)** (англ. long short-term memory LSTM) для класифікації активності людини.

Предмет дослідження

Застосування ДКЧП у трьох задачах класифікації:

- класифікація дій людини за відеозапису (біг, оплески, махання тощо)
- класифікація дій людини послідовностями даних гіроскопа та акселерометра (ходіння, підйом/спуск по сходах, тощо)
- розпізнавання читання за траекторією погляду людини

Мета дослідження

- перевірка ефективності алгоритму ДКЧП у задачах класифікації активності.
- дослідження його модифікацій та поєднання з іншими методами глибокого навчання.
- формування переваг та недоліків алгоритму.

2. Постановка задачі

Модель класифікації послідовностей описується наступним чином:

Вхідна послідоність $\{x^{(t)} \in \mathbb{R}^n, t \in (1, 2 \dots \tau)\}$ -> Клас з деякої фіксованої множини $\hat{y} \in (1, 2 \dots K)$

Необхідно побудувати модель, яка буде моделювати залежність згідно до заданого **критерію якості** на основі заданої вибірки:

$$D = \Big\{(x_i, y_i); x_i = \{x_i^{(t)} \in \mathbb{R}^n, t \in (1, 2 \dots \tau_i)\},$$

$$y_i \in (1, 2 \dots K), i \in (1, \dots, m)\Big\}$$

3.1. Набір даних KTH

Набір даних KTH був наданий Schuldt та ін. у 2004 році і є найпоширенішим набором даних активностей людини.

Вхідні данні:

- 600 відеозаписів різної довжини (100 для кожного классу)
- Відеозапис - це послідовність монохромних кадрів (матриць) розміром 120*160

Вихідні данні:

- 6 типів активності: 'boxing', 'handclapping', 'handwaving', 'jogging', 'running', 'walking'

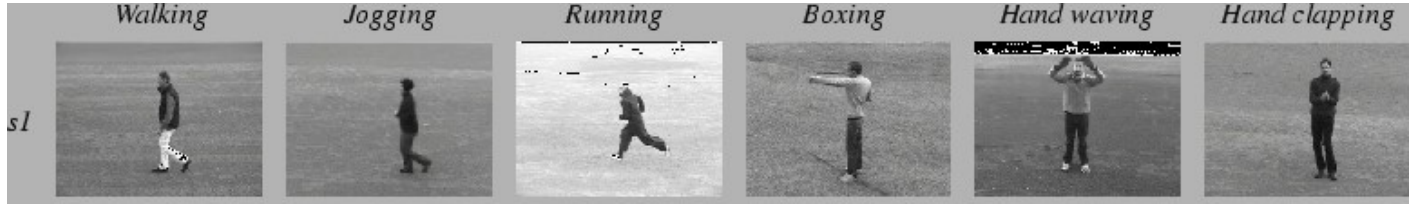


Рис.1 6 класів активностей

3.2. Набір даних HAR

База даних HAR була побудована з записів 30 учасників дослідження, які виконували 6 щоденних активностей тримаючи на талії смартфон із вбудованим акселерометром та гіроскопом.

Вхідні данні:

- 7352 часових рядів довжиною 128 часових кроків
- Кожний часовий ряд - послідовність 9-вимірних векторів, які містять інформацію про прискорення від акселерометра та кутову швидкість від гіроскопа

Вихідні данні:

- 6 типів активності: 'walking', 'walking upstairs', 'walking downstairs', 'sitting', 'standing', 'laying'

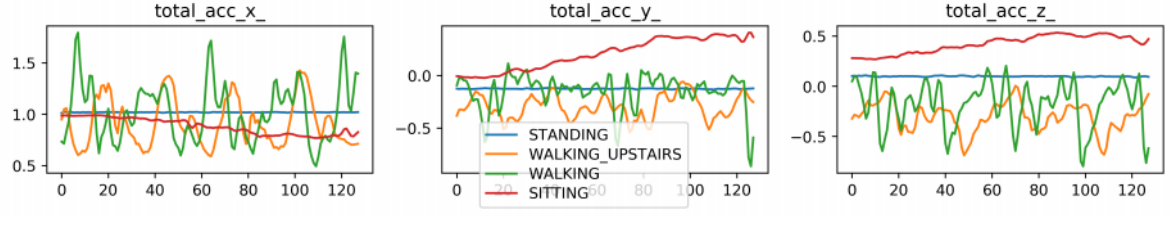


Рис.2 Прискорення у трьох вимірах для 4 класів активностей

3.3. Набір даних GAZEPOINT

Використовується набір даних, що генерується трекером погляду Gazepoint. У записі даних брало участь 15 учасників, кожен з яких мав прочитати два тексти, виконати задачу пошуку картинки та пошуку у тексті та продивитися відеозапис.

Вхідні данні:

- 143 послідовності, на кожному з якої записана траекторія переміщення погляду людини з частотою 60 Hz.
- кожна послідвність має такі ознаки: координати погляду на площині екрану та довжина фіксації.

Вихідні данні:

- 2 типи активності: 'reading', 'non-reading'

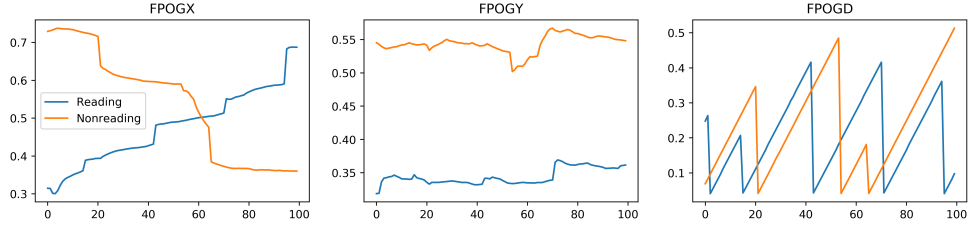


Рис.3 Траекторія погляду та довжина фіксації для читання та не читання

4. Критерії якості підбору моделі

Критеріями якості $J(D, \theta)$ моделі з параметрами θ було обрано наступні метрики:

- середнє значення загальної функціїї втрат L на валідаційній підвибірці:

$$J(D, \theta) = \frac{1}{m_{valid}} \sum_{i=1}^{m_{valid}} L(x_i, y_i)$$

- точність класифікації на валідаційній підвибірці:

$$J(D, \theta) = \frac{1}{m_{valid}} \sum_{i=1}^{m_{valid}} \mathbb{I}_{y_i = \tilde{y}_i}$$

- матриця невідповідностей (англ. confusion matrix) на валідаційній підвибірці:

$$J(D, \theta) = || \sum_{i=1}^{m_{valid}} \mathbb{I}_{y_i = k, \tilde{y}_j = l} ||_{k,l \in (1,..,K)}$$

5. Огляд існуючих підходів до розв'язку задачі

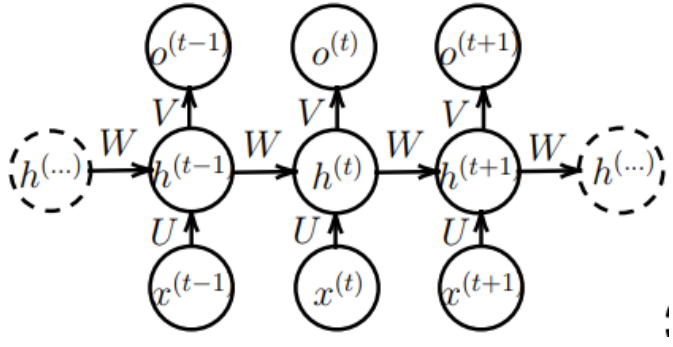
- Класичні підходи, присвячені розпізнаванню активності, вимагають **інженерію ознак** (англ. feature engineering), оскільки традиційні **нейронні мережі прямого поширення** не можуть працювати з послідовностями різної довжини та не володіють властивістю **пам'яті**.

- Вирішити ці недоліки допомагають рекурентні нейронні мережі (РНМ), які містять **зворотні зв'язки** і дозволяють зберігати інформацію про усю послідовність вхідних даних. Основна проблема РНМ - **проблема затухання градієнту** і звичайна РНМ не враховує довготривалі залежності.

- ДКЧП частково вирішує цю проблему, оскільки вона рекурсивно передає стан без застосування активації чи множення на матрицю вагів, за допомогою вентилів.

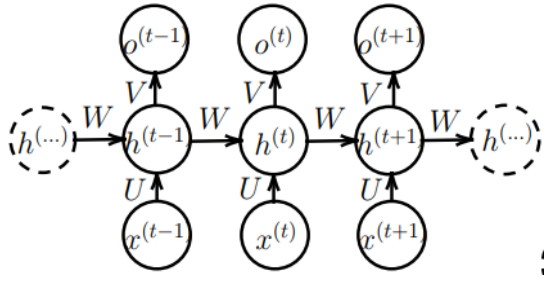
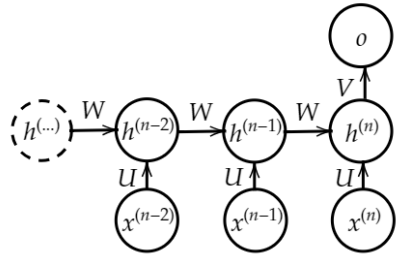
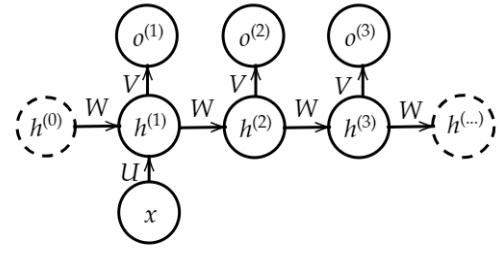
6.1. Рекурентні нейронні мережі

Рекурентна нейронна мережа - нейронна мережа, яка реалізує ідею **обміну параметрів**, і призначена для роботи з послідовностями значень довільної довжини $x^{(1)}, \ldots, x^{(\tau)}$. Фактично - це нелінійна авторегресійна екзогенна модель (англ. nonlinear autoregressive exogenous model, NARX).

Обчислювальний граф звичайної РНМ у розгорнутому вигляді	**Рівняння моделі**
	$a^{(t)} = b + Wh^{(t-1)} + Ux^{(t)}$ $h^{(t)} = tanh(a^{(t)})$ $o^{(t)} = c + Vh^{(t)}$

6.2. Види рекурентних нейронних мереж

Існують три основних архітектури РНМ, в залежності від задачі:

Many-to-many	Many-to-one	One-to-many
		

6.3. Рекурентні нейронні мережі в задачах класифікації

Найчастіше використовується архітектура **many-to-one**, інколи **many-to-many** (неперевна класифікація).

Прогнозом моделі **many-to-one** будемо вважати:

$$\hat{y} = softmax(o); softmax(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}, i = 1, \dots, K$$

Функцією втрат L для пари (x, y) вхідної послідовності та вихідного класу буде кросс-ентропія (негативна логарифмована функція правдоподібності):

$$L(\{x^{(1)}, \dots, x^{(\tau)}\}, y\}) = -\sum_{i} (y)_i * ln((\hat{y})_i)$$

де $\hat{y}$ - прогнозоване значення виходу, y - справжнє значення виходу в унітарному кодуванні.

6.4. Рекурентні нейронні мережі в задачах класифікації

Для знаходження параметрів моделі необхідно мінімізувати загальну функцію втрат:

$$J = \frac{1}{m_{train}} \sum_{i=1}^{m_{train}} L(x_i, y_i)$$

Зв'язки РНМ між $\hat{y}$ та $\{x^{(1)}, \ldots, x^{(\tau)}\}$ описуються диференційованими функціями, можна використати градієнтні методи.

У роботі застосовано **оптимізатор Адама**, який є модифікованим методом стохастичного градієнтного спуску.

7.1. Довга коротчкочасна пам'ять

Довга короткочасна пам'ять (ДКЧП, англ. long short-term memory, LSTM) — це архітектура рекурентних нейронних мереж, запропонована 1997 року Зеппом Хохрайтером та Юргеном Шмідгубером.

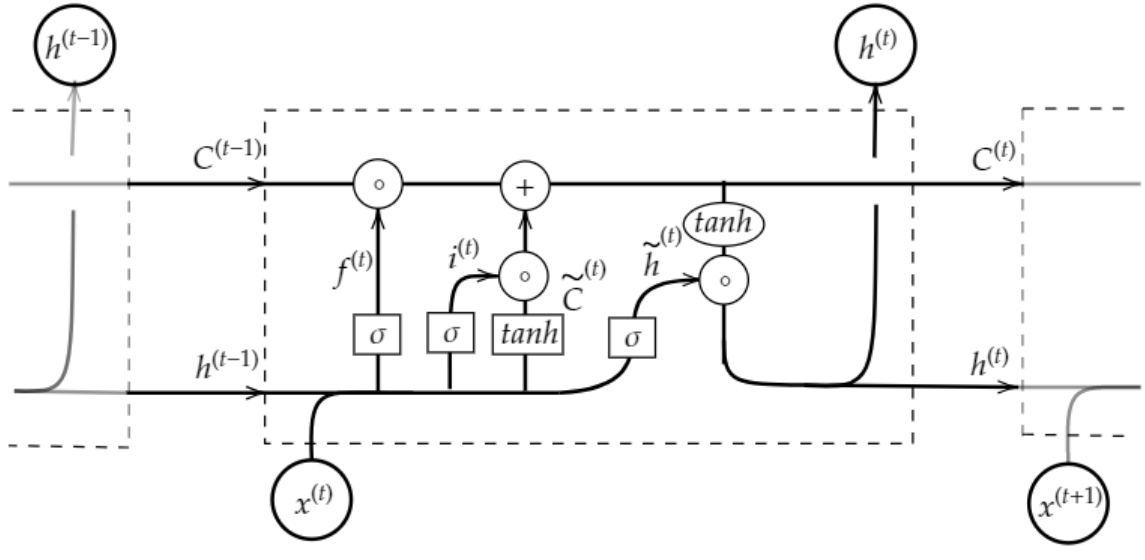
ДКЧП розроблені спеціально для того, щоб уникнути **проблеми затухання градієнту**, яка властива звичайним РНМ.

ДКЧП описується наступними рівняннями:

$$
\begin{aligned}
f^{(t)} &= \sigma(W_f[h^{(t-1)}, x^{(t)}] + b_f) \\
i^{(t)} &= \sigma(W_i[h^{(t-1)}, x^{(t)}] + b_i) \\
C^{(t)} &= f^{(t)} \circ C^{(t-1)} + i^{(t)} \circ tanh(W_C[h^{(t-1)}, x^{(t)}] + b_C) \\
\tilde{h}^{(t)} &= \sigma(W_o[h^{(t-1)}, x^{(t)}] + b_o) \\
h^{(t)} &= \tilde{h}^{(t)} \circ tanh(C^{(t)})
\end{aligned}
$$

7.2. Обчислювальний граф ДКЧП

ДКЧП містить чотири рекурентних шари та три вентилі: вхідний, забувальний та вихідний.

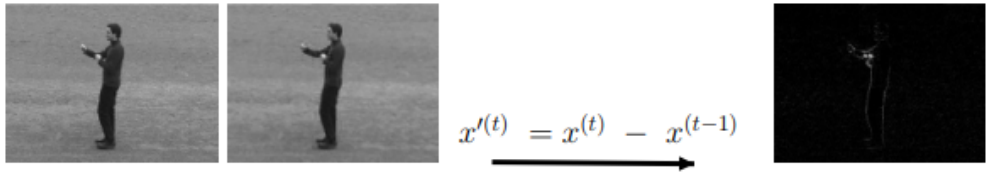


де квадратами позначено шари мережі, ○ - добуток Адамара

8.1. КТН - результати

Попередня обробка даних:

- Диференціювання $x'^{(t)} = x^{(t)} - x^{(t-1)}$, щоб враховувати лише рухомі частнини кадрів і нормалізувати данні
- Нарізка на підпослідовності довжиною 25 кадрів.



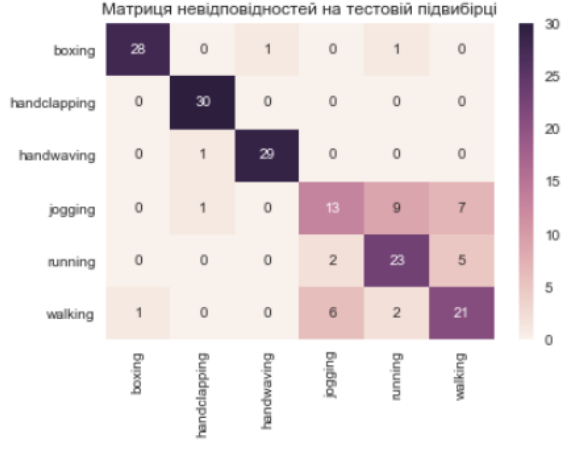
Було побудовано дві моделі:

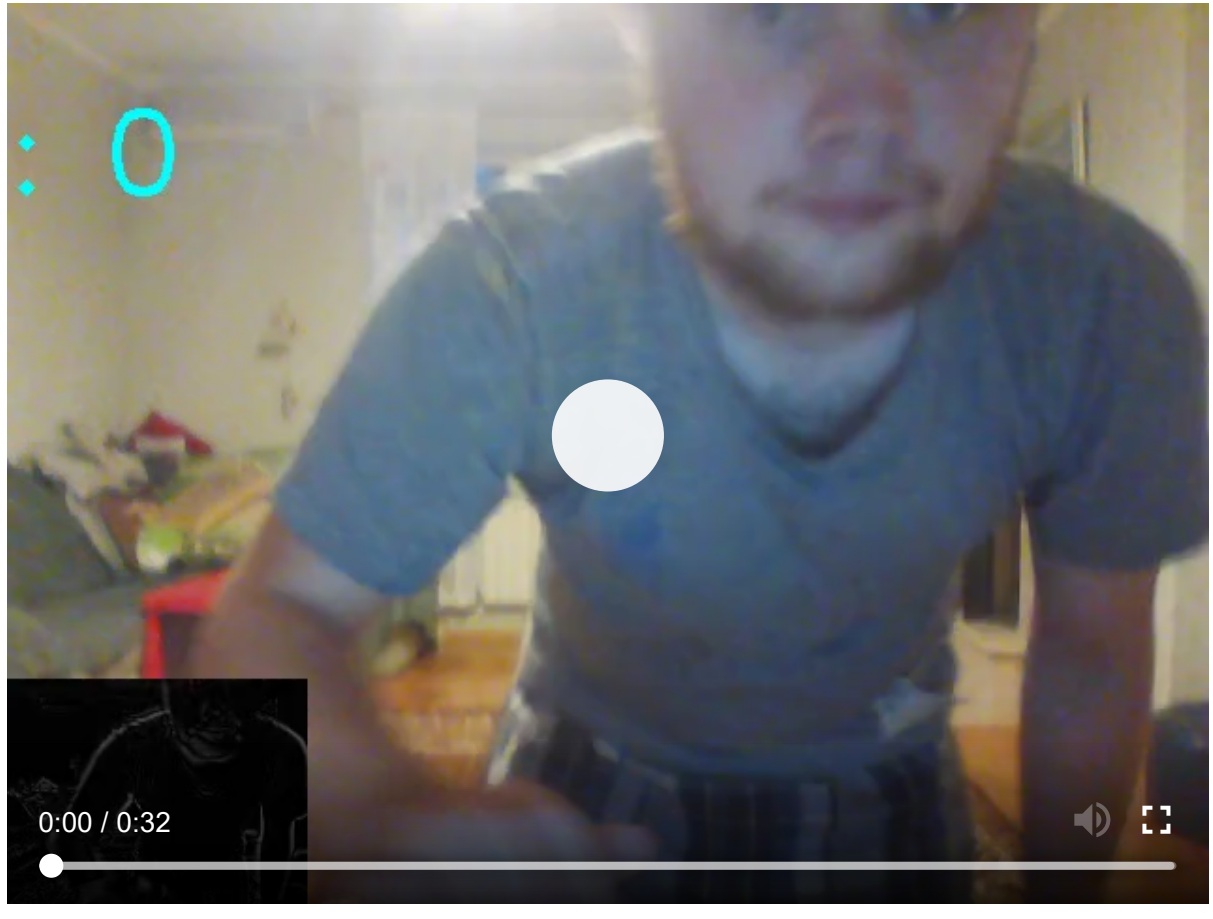
- CNN-LSTM - 3-вимірна згоркова мережа + 2-шарова ДКЧП
- CONV-LSTM - згорткова ДКЧП (3 шари)

Результати

Назва моделі	Accuracy	Loss	Кількість параметрів
CNN-LSTM	80.1%	0.5708	10,238
CONV-LSTM	50.0%	0.958	5,174

Матриця несумісностей найкращої моделі




0
0:00 / 0:32

8.2. HAR - результати

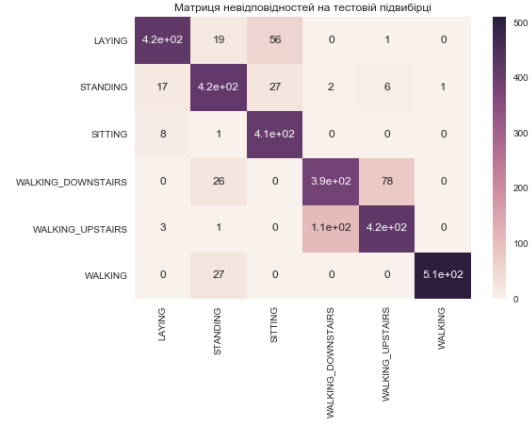
Попередня обробка відсутня.

Було побудовано три моделі:

- LSTM - звичайна ДКЧП
- LSTM-L_2 - ДКЧП з L_2 регуляризацією
- LSTM-dropout - ДКЧП з випадковим відключенням зв'язків під час навчання

Результати	**Матриця несумісностей найкращої моделі**

Назва моделі	**Accuracy**	**Loss**	**Кількість параметрів**
LSTM	90.0%	0.3911	13,894
LSTM-L_2	92.3%	0.2234	13,894
LSTM-dropout	94.65%	0.0879	13,894



8.3. GAZEPOINT - результати

Попередня обробка даних:

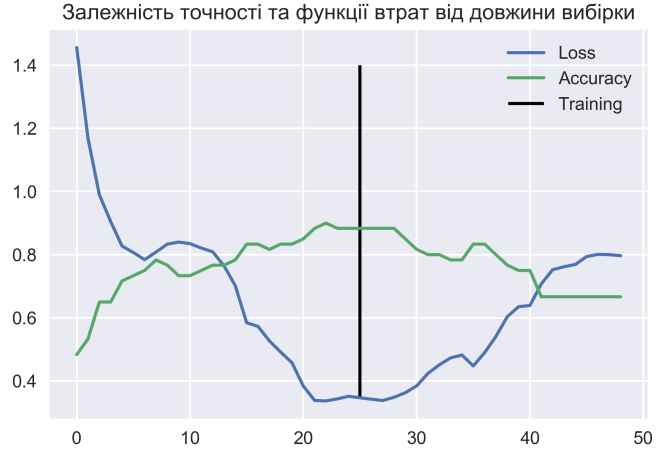
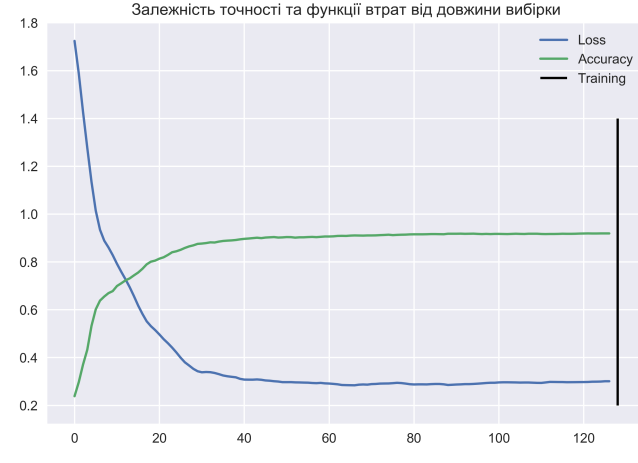
- Диференціювання траекторії погляду $x'^{(t)} = x^{(t)} - x^{(t-1)}$.

Було побудовано дві моделі:

- LSTM - звичайна ДКЧП з 3-вимірної вхідна послідовністю, 100 часових кроків
- LSTM-reshape(6) - ДКЧП з 6-вимірної вхідної послідовностю, 50 часових кроків
- LSTM-reshape(12) - ДКЧП з 12-вимірної вхідної послідовностю, 25 часових кроків
- CNN-5 - п'ятишарова згорткова мережа (для порівняння)

Назва моделі	**Accuracy**	**Loss**	**Кількість параметрів**
LSTM	63.46%	0.6572	12,961
LSTM-reshape(6)	84.88%	0.3363	13,345
LSTM-reshape(12)	86.50%	0.3086	14,113
CNN-5	95.3%	0.1192	4,536

9. Залежність якості моделі від довжини вхідної послідовності

KTH	HAR
	

10. Подальші ідеї для дослідження

- Навчання на послідовностях різної довжини (треба мати апріорне знання про тривалість залежності).
- Порівняння різноманітних архітектур ДКЧП (вічкова, з об'єднаними вентилями входу та забування).

11.1. Висновки - переваги

У роботі проведено дослідження щодо можливості та ефективності застосування ДКЧП для задач класифікації активності людини.

Серед переваг ДКЧП можна зазначити такі:

- Відсутність попередньої обробки даних
- Масштабованість та універсальність
- Можливість працювати з різними довжинами вхідних послідовностей

11.2. Висновки - недоліки

Крім перерахованих у роботі переваг методу, було виявлено певні його недоліки:

- ДКЧП дійсно не вимагає інженерії ознак, але в деяких випадках точніше будувати модель на вхідних даних з попередньою обробкою
- ДКЧП не є універсальним алгоритмом для роботи з послідовностями, вона наприклад не змогла добре класифікувати 3-вимірні послідовності GAZEPOINT.
- Інколи може бути складно підібрати оптимальну архітектуру моделі
- Навчання даних моделей може займати певний час

Дякую за увагу!