

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки
Кафедра програмної інженерії

ЗВІТ

Індивідуального завдання (ІНДЗ)
з дисципліни «Інтелектуальний аналіз даних»
на тему «Порівняння RNN та LSTM для обробки природної мови»

Виконав

студент групи ІПЗм-24-2

Голодніков Дмитро

Перевірів

ст. викл. Онищенко К.Г.

Харків 2024

1 МЕТА РОБОТИ

Провести порівняльний аналіз рекурентних нейронних мереж (RNN) та мереж з довгою короткостроковою пам'яттю (LSTM) для задач обробки послідовностей. Дослідити проблему зникаючого градієнта та переваги архітектури LSTM.

2 ХІД ВИКОНАННЯ РОБОТИ

2.1 Архітектура RNN

Рекурентна нейронна мережа (RNN) обробляє послідовності, передаючи прихований стан з кожного часового кроку на наступний. Формула оновлення: $h(t) = \tanh(W \cdot h(t-1) + U \cdot x(t) + b)$. RNN страждає від проблеми зникаючого градієнта при довгих послідовностях.

2.2 Архітектура LSTM

LSTM (Long Short-Term Memory) вирішує проблему зникаючого градієнта через механізм воріт: forget gate, input gate, output gate. Додатковий cell state дозволяє зберігати інформацію на довгих часових проміжках.

2.3 Експеримент 1: Стандартні послідовності

Порівняння на задачі класифікації послідовностей середньої довжини (50 кроків). Обидві архітектури показують схожі результати на простих задачах.

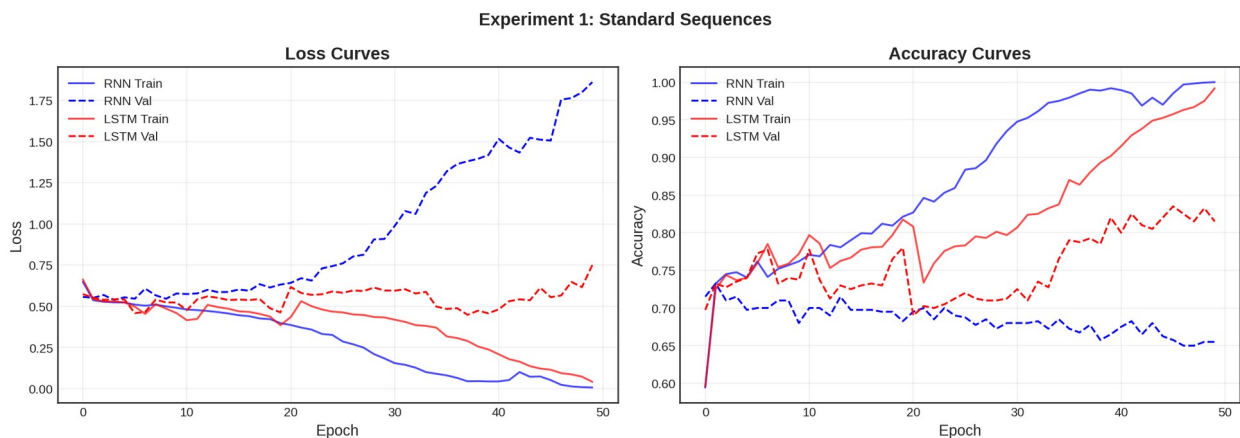


Рис. 2.1 - Криві навчання на стандартних послідовностях

2.4 Експеримент 2: Довгострокові залежності

Тестування здатності моделей запам'ятовувати інформацію з початку послідовності. LSTM значно перевершує RNN на цій задачі.

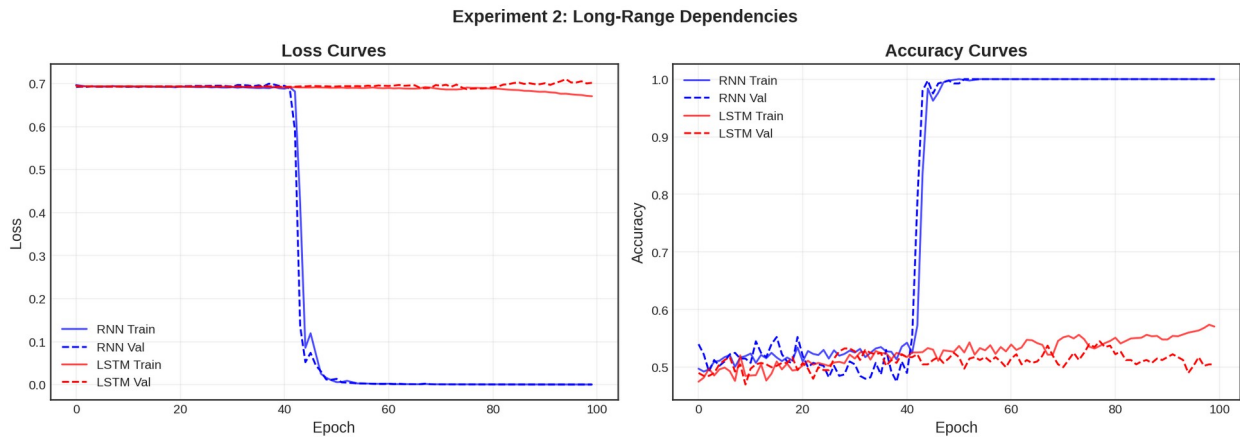


Рис. 2.2 - Криві навчання на довгих залежностях

2.5 Експеримент 3: Вплив довжини послідовності

Дослідження деградації точності при збільшенні довжини послідовності. RNN швидко втрачає здатність навчатися на довгих послідовностях.

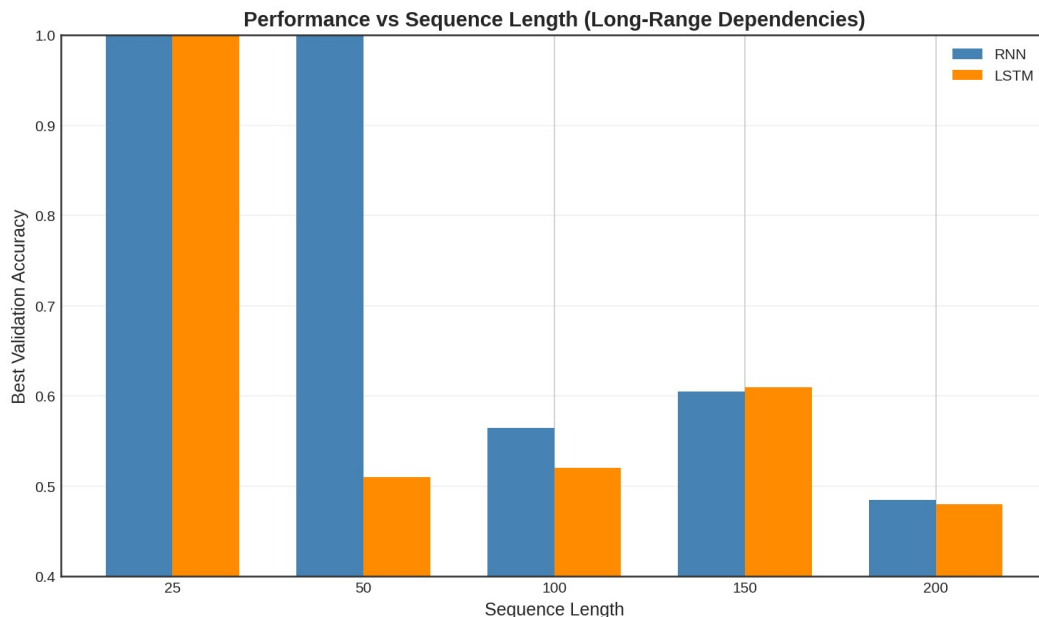


Рис. 2.3 - Залежність точності від довжини послідовності

3 ВИСНОВКИ

Проведено експериментальне порівняння архітектур RNN та LSTM. На стандартних послідовностях обидві моделі показують схожу точність. LSTM значно перевершує RNN на задачах з довгостроковими залежностями. При збільшенні довжини послідовності точність RNN швидко падає, тоді як LSTM зберігає здатність навчатися. Cell state в LSTM діє як "шосе" для градієнтів, дозволяючи інформації зберігатися на багатьох часових кроках. Рекомендується використовувати LSTM для задач NLP з довгими текстами.

ПОСИЛАННЯ

Код проєкту доступний у репозиторії GitHub: <https://github.com/na-naina/data-analysis-khnure>