УДК 004.02

Новічков Є. М., магістрант,
Чижмотря О. В., ст. викладач
Державний університет «Житомирська політехніка»

ВИКОРИСТАННЯ BI-LSTM МОДЕЛІ З МЕХАНІЗМОМ УВАГИ ДЛЯ АВТОМАТИЧНОГО ТЕГУВАННЯ ТЕКСТУ

Для побудови системи автоматичного тегування тексту, нам потрібно побудувати рекурентну нейронну мережу, головна відмінність якої – наявність «пам'яті», що формується через подачу виходу попереднього кроку часу як входу у наступний крок.

Звичайні рекурентні нейронні мережі мають проблему зникнення градієнту, що заважає їм вивчати довготривалі зв'язки між словами. Натомість, Long short-term memory (LSTM), одна із різновидів рекурентних мереж, має структуру комірок пам'яті [1]. Ця структура підходить для завдань, де результат залежить від аналізу семантики всього тексту. Архітектура такої моделі починається з шару вкладень, що подає слова як числові вектори. У цьому векторному просторі слова зі схожими значеннями розташовані ближче одне до одного. Ми можемо скористатися вже навченою моделлю вкладень, даючи змогу моделі узагальнювати синоніми навіть якщо даних для тренування небагато.

Ядром такої архітектури є LSTM комірка, що керує потоком інформації за допомогою трьох механізмів: вхідний вентиль, вентиль забуття та вихідний вентиль. Ці вентилі використовують сигмоїдні функції активації для видачі значень від нуля до одиниці [1]. Вентиль забуття вирішує, яка інформація з попереднього стану комірки більше не є актуальною та має бути відкинута. Вхідний вентиль визначає, яка нова інформація є достатньо значущою для оновлення стану комірки, а вихідний вентиль контролює, яка інформація передається до наступного прихованого стану.

Проте лінійна обробка тексту «зліва направо» не є найкращою для завдань класифікації, де весь документ доступний одночасно. У складних текстах процес визначення значення слова часто залежить від контексту навколо нього. Тому кращою архітектурою для даної задачі є двонапрямлена LSTM (Bi-LSTM). Вона передбачає тренування двох окремих шарів LSTM: прямого та зворотного LSTM, які обробляють послідовність від початку до кінця й від кінця до початку відповідно. На кожному конкретному часовому кроці приховані стани цих двох шарів об'єднуються. Це гарантує, що кінцеве представлення слова містить повний контекст до та після нього [2].

Незважаючи на ефективність Bi-LSTM, її стандартна архітектура має обмеження, відоме як інформаційне вузьке місце (information bottleneck). Мережа стискає семантичний вміст усього документа в кінцевий вектор прихованих станів, що призводить до втрати деталей. Щоб вирішити цю проблему, можна додати механізм уваги. Цей шар дозволяє моделі брати до уваги усі попередні приховані стани, а не лише кінцевий. Механізм працює шляхом обчислення оцінки для кожного часового кроку, які потім нормалізуються для отримання вагових коефіцієнтів уваги [1]. Це дозволяє моделі присвоювати високу значущість найбільш релевантним словам, практично ігноруючи стоп-слова. Це усуває проблему вузького місця та спрощує доступ до

Оскільки тегування тексту є завданням класифікації з кількома мітками, де документ може належати до кількох не взаємовиключних категорій, стандартна softmax функція активації є недоречною. Замість неї краще використати сигмоїдну функцію активації. Вона обчислює кінцеві значення незалежно для кожного класу від нуля до одиниці. Потім застосовується поріг ймовірності для визначення приналежності класу до документа. Щоб зменшити ризик перенавчання алгоритм повинен використовувати шари dropout для випадкового вимкнення нейронів під час навчання, запобігаючи коадаптації, та batch нормалізацію для стабілізації розподілу вхідних даних. Процес навчання повинен керуватися оптимізатором adam для ефективної конвергенції, використовуючи бінарну перехресну ентропію як функцію втрат для належного покарання за помилки.

Використання описаного підходу забезпечує створення ефективної системи автоматичного тегування тексту, що спирається не лише на частотність термінів, а й на їхнє контекстуальне значення.

Список використаних джерел:

1. Liriam Enamoto, Andre R.A.S. Santos, Ricardo Maia, Li Weigang and Geraldo P. Rocha Filho. Multi-label legal text classification with BiLSTM and attention. International Journal of Computer Applications in Technology. 2022. Vol. 68, No. 4. P. 369-378. URL: https://www.researchgate.net/publication/363244291_Multi-label_legal_text_classification_with_BiLSTM_and_attention (дата звернення: 19.11.2025).

2. Understanding Bidirectional LSTM for Sequential Data Processing. URL: https://medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befc (дата звернення: 19.11.2025).