

УДК 004.8

Нікуліна А.С.¹, Леощенко С.Д.²

¹ студ. гр. КНТ-114м НУ «Запорізька політехніка»

² д-р філос., доц. НУ «Запорізька політехніка»

ТОПОЛОГІЧНІ ОСОБЛИВОСТІ СУЧАСНИХ АРХІТЕКТУР ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ АНАЛІЗУ МОВЛЕННЯ

Структурні особливості рекурентних нейронних мереж (RNN) роблять їх особливо ефективними для завдань аналізу мови завдяки їх здатності обробляти послідовності та залежності з часом. Розглянемо конкретні кейси таких структурних особливостей RNN, що сприяють поліпшенню результатів аналізу мовлення.

Управління тимчасовими залежностями. Послідовна обробка даних – на відміну від мереж прямого зв'язку, RNN мають у своїй архітектурі цикли, які

дозволяють їм зберігати інформацію з попередніх часових кроків [1]. Це важливо для аналізу мови, оскільки мовні сигнали за своєю суттю є послідовними; кожен звук або фонема залежать від попередніх. Пам'ять про минулі стани – RNN зберігають інформацію про попередні вхідні сигнали, що дозволяє їм обробляти та пов'язувати звуки з часом, що є важливою функцією для розуміння контексту мови. Ця можливість розширює можливості таких завдань, як розпізнавання фонем, сегментація слів та ідентифікація мовця, зберігаючи контекст у звуковому сигналі [2].

Обробка послідовностей змінної довжини. Адаптація до різної довжини вхідних даних – мовленнєві сегменти, такі як слова, фрази або пропозиції, розрізняються по довжині [1]. RNN, природно, враховують ці відмінності без необхідності фіксованих розмірів вхідних даних, що робить їх ефективними при аналізі безперервної або неоднорідної мови. Обробка в режимі реального часу – оскільки RNN обробляють введення по одному часовому кроку за раз, вони можуть обробляти потокові дані та забезпечувати розпізнавання мови в режимі реального часу, що важливо в таких додатках, як голосові помічники та транскрипція в режимі реального часу [1].

Довготривала короткочасна пам'ять (LSTM) і закриті рекурентні блоки (GRU). Вирішення проблем із зникаючим градієнтом – на практиці стандартні RNN стикаються з довгостроковими залежностями через зникаючі градієнти, які можуть призвести до того, що мережа забуде попередні частини послідовності. Архітектури LSTM та GRU, дві просунуті форми RNN, вирішують цю проблему, впроваджуючи механізми управління, які вибірково оновлюють, забувають або зберігають інформацію протягом більш тривалих періодів часу [2].

Уловлювання контекстуальних нюансів у мовленні – LSTM та GRU покращують виконання мовленнєвих завдань, допомагаючи моделі запам'ятовувати важливі контекстуальні сигнали в довгих послідовностях, що особливо важливо для розуміння довгих фраз, контекстуального тону або навіть настроїв у мовленні [2].

Двонаправлена обробка. Прямий та зворотний контекст – двонаправлені RNN обробляють дані як у прямому, так і в зворотному напрямку, що дозволяє їм використовувати як минулий, так і майбутній контекст у мовленнєвій послідовності. Це корисно в таких завданнях, як розпізнавання мови та обробка природної мови, де розуміння майбутнього контексту (наприклад, цілого речення) може прояснити неоднозначні звуки чи слова. Покращене розпізнавання фонем – при розпізнаванні мови деякі звуки можуть бути фонетично неоднозначними без урахування навколишнього контексту. Двонаправлені RNN допомагають усунути неоднозначність цих звуків, отримуючи доступ як до попередньої, так і до подальшої фонетичної інформації, що дозволяє більш точно розпізнавати фонему та слова [1].

Здатність справлятися з контекстуальними варіаціями. Чутливість до акустичних коливань – RNN за своєю суттю чутливі до послідовних даних, що робить їх ефективними при розпізнаванні акустичних змін з часом. Така чутливість забезпечує більш високу продуктивність при виявленні унікальних рис мовця, емоційного тону або варіацій висоти тону, які мають вирішальне значення для таких завдань, як ідентифікація мовця, розпізнавання емоцій і аналіз інтонації [1], [2]. Стійкість до перешкод – оскільки RNN можуть використовувати минулу інформацію для прогнозування або згладжування поточного сигналу, вони більш стійкі до шуму, що є поширеною проблемою при обробці мовних даних у реальному світі. Така надійність забезпечує більш високу продуктивність при аналізі мови в неідеальних умовах запису [2].

Структурні особливості RNN, зокрема їх механізми пам'яті, двонаправлена обробка та спеціалізовані модулі (такі як LSTM та GRU), дозволяють їм фіксувати та підтримувати залежності в послідовних мовних даних. Ці функції мають вирішальне значення для ефективного аналізу, Транскрибування та інтерпретації розмовної мови, що робить RNN добре придатним для вирішення складних проблем аналізу мовлення.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. What is RNN (Recurrent Neural Network) [Electronic resource]. – Access mode: https://aws.amazon.com/what-is/recurrent-neural-network/?nc1=h_ls
2. Understanding Recurrent Neural Networks (RNN) — NLP [Electronic resource]. – Access mode: <https://medium.com/@praveenraj.gowd/understanding-recurrent-neural-networks-rnn-nlp-c2f4cae03a4f>