

Ткачук Андрій Віталійович

студент

*Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»*

ВИКОРИСТАННЯ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ В ОБРОБЦІ ПРИРОДНОЇ МОВИ

***Анотація.** У роботі обґрунтовано складність обробки текстів, написаних природними мовами, розглянуто структуру рекурентних нейронних мереж та їх переваги при застосуванні у моделях обробки текстових даних. Досліджено моделі для машинного перекладу і генерування/предикції тексту.*

***Ключові слова:** рекурентні нейронні мережі, обробка природної мови, інтелектуальний аналіз, штучний інтелект.*

Вступ. Обробка природної мови є галуззю, що поєднує у собі штучний інтелект і математичну лінгвістику і знаходить своє застосування в інтелектуальному аналізі даних, які представлені у формі мови чи тексту.

Інтелектуальним аналізом тексту є його обробка із використанням засобів штучного інтелекту, тому машинний переклад та створення нового тексту, що відбувається за допомогою рекурентних нейронних мереж, може розглядатися із точки зору інтелектуального аналізу.

Основний текст. Основною проблемою, із якою стикаються вчені при розробці моделей та засобів для обробки природної мови, є її неоднозначність, яка проявляється у сенсі і відмінках слів, синтаксисі речень, відношенні слів між собою у реченнях. Саме розуміння сенсу

речення або уривку тексту і його детерміноване представлення у системі для можливості його подальшого опрацювання є задачею обробки природномовного тексту. Проблема полягає у складності встановлення конкретного відображення дійсної семантико-синтаксичної структури речення у його внутрішнє логічне уявлення, яке автоматично генерується системою [1]. Тому при роботі із природномовним текстом така проблема найбільш гостро постає при машинному перекладі та створенні нового тексту, оскільки важливим є не сам дослівний переклад, а збереження сенсу і створення тексту, який матиме смислове навантаження.

Раніше інтелектуальний аналіз тексту полягав у послідовному проведенні морфологічного, синтаксичного та семантичного аналізів. Такий підхід вимагає великої кількості ресурсів та попередніх знань для отримання коректних результатів, а їх якість прямо пропорційно залежить від кількості внесених у систему правил мови, якою написаний текст. Такі процедури втрачають точність, коли стикаються із явищами полісемії та омонімії [2]. Тому сьогодні існують лінгвістичні програми, які здатні виявляти неоднозначності такого роду у дуже вузьких сферах.

Рекурентні нейронні мережі є чудовим засобом для обробки природних мов. Це пояснюється тим, що їх структура чітко відображає структуру речення. У рекурентній нейронній мережі значення виходу на кожному кроці залежить від значення виходу на попередньому кроці. Іншими словами – такі мережі мають пам'ять. Відповідно значення речення залежить від порядку слів. Сьогодні існують шари із вдосконаленою внутрішньою структурою, що дозволяють запам'ятовувати чи забувати дані із певних кроків (шари Long Short-Term Memory - LSTM). Такий тип шарів використовується найчастіше, оскільки дозволяє досягти найкращих результатів. Також використовуються різноманітні функції активації. Основною вимогою є недопускання виникнення проблеми зникання градієнту (оскільки при навчанні використовуються градієнти

методи, наприклад метод зворотного поширення).

Якщо розгорнути структуру РНН у часі і співставити її кожен крок із словом у реченні, то отримаємо однозначну відповідність: кожен попередній крок має вплив на наступний, або від попереднього слова залежить наступне. Така структура зображена на рис. 1.

Таку модель (саму рекурентну нейронну мережу) можна використовувати для створення тексту або передбачення наступного слова для частини тексту. Такі задачі вирішуються однією імплементацією мережі, що описує ще одну перевагу використання РНН.

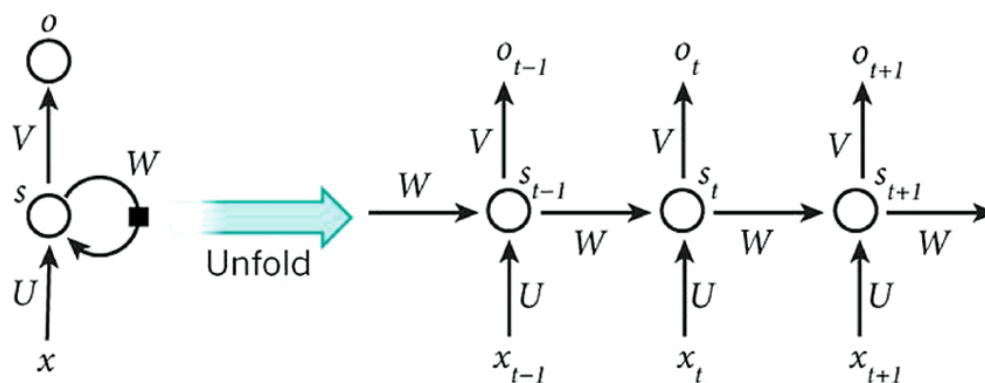


Рис. 1. Розгортання РНН [3]

Для того, аби описана вище модель змогла генерувати текст, її потрібно навчити на спеціально створеному наборі даних. Зрозуміло, що для того, аби написати наступне слово, необхідно володіти інформацією про те, які слова були раніше. Відповідно до цього твердження будується навчання мережі і відбувається за наступним принципом: на вхід мережі на поточному кроці для навчання подається перша частина речення, а за вихід приймається це ж речення, але зсунуте на одне слово вліво. Таким чином, мережа вивчить, що для вхідної послідовності наступним словом буде те, яке з'явилося при зсуванні речення (останнє слово у реченні, що приймалось за «ідеальний» вихід) [3].

При здійсненні машинного перекладу за допомогою нейронних мереж виникає ще одна проблема – речення мають різну довжину у

вихідній мові та у цільовій. Така проблема вирішується за допомогою введення додаткового проміжного стану даних, а обробка цих даних у такому випадку відбувається вже двома рекурентними нейронними мережами різного розміру.

На рис. 2 показано структуру моделі Sequence2Sequence, запропоновану співробітниками Google.

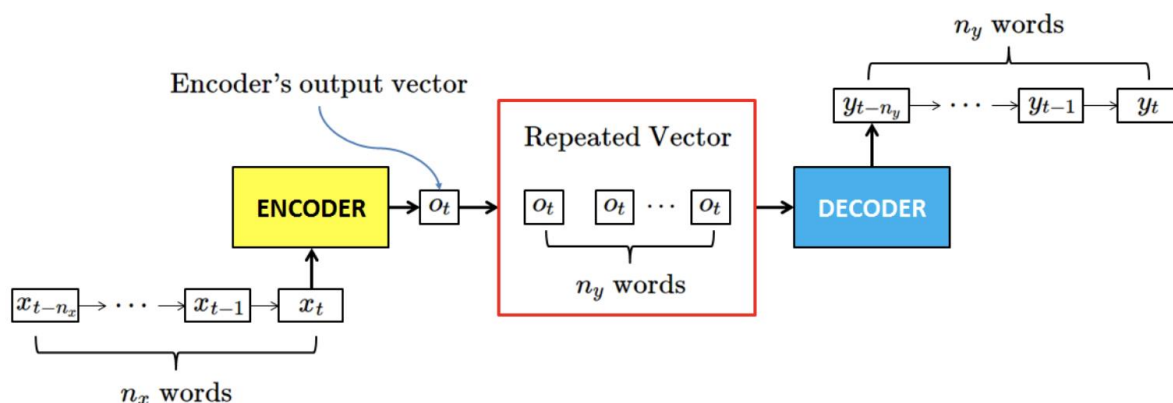


Рис. 2. Структура моделі Sequence2Sequence [4]

Для того, аби здійснювати переклад із використанням такої моделі, не потрібно мати жодних додаткових відомостей про структуру мови чи її правила. Достатньо навчити модель на підготовленому наборі даних, який складається із пар речення – переклад. Це дозволяє здійснювати переклад із однієї мови на іншу напряму (без використання перекладу на проміжну мову), отримувати переклад у певному контексті (витримується контекст навчальних даних) [4].

Очевидно, що комп'ютер не здатен розуміти значення слів, тому для можливості роботи із словами здійснюється їх перетворення у зрозумілу для машин форму – числа. При навчанні моделей із навчальних даних будуються словники, у якому кожному слову присвоюється номер. Оскільки слів дуже багато, то на практиці часто розмір словника зменшують, тому в нього входять лише слова, які зустрічаються найчастіше у навчальному наборі. Розмір речення (послідовності), яке

може бути опрацьовано за один раз, теж є обмеженим. Це пов'язано із розміром самої нейронної мережі, із збільшенням якого значно збільшується кількість ресурсів, які необхідні для забезпечення її нормального функціонування. Саме через це у лінгвістичних програмах (перекладачах) існує обмеження на уривок тексту, який може бути опрацьовано за один раз.

Висновки. Із проведеного дослідження можна зробити висновок, що рекурентні нейронні мережі чудово підходять для обробки текстових даних, написаних природними мовами. Вони дозволяють зменшити кількість процедур та початкових даних, які необхідні для проведення інтелектуального аналізу. Використання методів та засобів штучного інтелекту, а також їх поєднання для обробки текстів, написаних природними мовами, дозволяє отримати якісно нові кращі результати.

Література

1. Система обработки текстов на естественном языке / авт. Анисимов А. В. Марченко А. А. // «Штучний інтелект». - Київ : НАН України, 2002 р. 4. - С. 157.
2. Інтелектуальна обробка текстів / авт. Тарануха В.Ю.. - Київ : Електронна публікація, 2014. - Т. 1 - С. 80.
3. Recurrent Neural Networks Tutorial. – Режим доступу: <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>
4. Creating a Language Translation Model Using Sequence-To-Sequence Learning Approach. – Режим доступу: <https://chunml.github.io/ChunML.github.io/project/Sequence-To-Sequence/>