

РОЗРОБКА СИСТЕМИ МАШИННОГО ПЕРЕКЛАДУ НА ОСНОВІ НЕЙРОМЕРЕЖЕВИХ ТЕХНОГОЛІЙ З ВИКОРИСТАННЯМ ВЕКТОРА МЕТРИК ЯКОСТІ

UDC 004.415.5

M. Baranskyi

(Ternopil Ivan Puluj National Technical University, Ukraine)

DEVELOPMENT OF MACHINE TRANSLATION SYSTEM BASED ON TECHNOLOGY OF NEURAL NETWORKS WITH USING A QUALITY METRIC VECTOR

Доповідь присвячена програмній реалізації системи машинного перекладу на основі нейромережових технологій з використанням вектора метрик якості. Проведено аналіз предметної області. Після проведення перед проектного дослідження предметної області, поставлено завдання на розробку системи, сформульовано вимоги до ресурсу.

Професійних перекладачів вчать передавати зміст своїми словами, не прив'язуючись до структури речень оригінального тексту. Адекватний переклад повинен прагнути від дослівній і пофразовій передачі до сенсових трансформацій. Правильна «конвертація» граматичних конструкцій однієї мови засобами іншої це межа для статистичного машинного перекладу. Це обмеження не скасовується навіть на базі споріднених мов і створює більше проблем в інтерпретації, але і менше граматичної схожості мов в парі.

Як і статистичний переклад, нейронний переклад вимагає для навчання паралельні корпуси, що дозволяють порівняти автоматичний переклад з еталонним «людським», тільки в процесі навчання оперує не окремими фразами і словосполученнями, а цілими реченнями [1]. Основна проблема в тому, що для тренування такої системи потрібно істотно більше обчислювальних потужностей.

Наступний етап полягав у виборі архітектур для розробки системи машинного навчання та тренування нейронних мереж кожної з архітектур. Наступним кроком є визначення точності перекладу за допомогою метрики METEOR – метрика для оцінювання якості машинного перекладу. Метрика базується на використанні n-gram та орієнтована на використання статистичної та точної оцінки вихідного тексту[2].

Далі було створено систему машинного перекладу, використовуючи мову програмування Python. Для реалізації архітектури було обрано бібліотеку управління базами даних Keras та дистрибутив розробки Anaconda, яка включає у собі дану бібліотеку. У якості даних для навчання будуть використовуватися двісті тисяч перекладів речень з англійської на німецьку.

Розробка архітектури системи машинного перекладу є одним з ключових моментів його створення. Практичне значення отриманих результатів дослідження полягає у тому, що створена комп'ютерна система допомагає визначити котра з обраних архітектур є найбільш точною для використання при розробці різних систем машинного перекладу.

Література

1. Alex Galea, Luis Capelo «Applied Deep Learning with Python. Use scikit-learn, TensorFlow, and Keras to create intelligent system and machine learning solutions» 2018 p. С. 214–215, С. 216–217.
2. Steven Bird, Ewan Klein, Edward Loper “Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit”, с. 292–304.