

ДОСЛІДЖЕННЯ ТА ЗАСТОСУВАННЯ ГРАФІВ ЗНАТЬ ТА НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДЛЯ ОБРОБКИ ПРИРОДНОМОВНИХ ТЕКСТІВ

Рябова Н. В., Громак О. В.

Харківський національний університет радіоелектроніки, Харків, Україна

Використання глибинного навчання нейронних мереж для вирішення задач обробки природної мови (Natural Language Processing, NLP) та сучасних бібліотек для створення та опрацювання нейронних моделей за допомогою мови Python і потужних платформ машинного навчання TensorFlow та PyTorch за останні роки продемонструвало вражаючі результати.

Нейромережевий підхід до NLP має переваги в першу чергу завдяки використанню векторного представлення об'єктів дослідження, наприклад слів (word2vec). При цьому зберігається семантичний зв'язок між словами завдяки їх вкладенню (embeddings) у векторний простір ознак.

З іншого боку стрімко розвивається підхід до NLP із застосуванням графів знань (Knowledge Graphs, KG). KG можна розглядати як специфічну семантичну мережу, центральним вузлом якої є об'єкт інтересу, пов'язаний різними типами відношень з іншими вузлами, які представляють конкретні об'єкти, поняття, інформаційні ресурси, або дані про них, а ребра представляють семантичні відносини між вузлами.

Таким чином, графи знань пропонують широко відомий формат для представлення інформації у формі, що обробляється комп'ютером. Така модель фокусується на зберіганні всіх даних як на рівні схеми, так і на рівні окремих сутностей.

Метою даної роботи є дослідження та використання гібридних моделей на підставі графів знань та нейронних мереж для вирішення задач NLP. Вирішується актуальна задача видобування іменованих сутностей з текстів (Named Entity Extraction, NEE). NEE - це завдання, яке включає розпізнавання згадування іменованого об'єкта в тексті (Named Entity Recognition, NER), усунення неоднозначності його можливих посилань (NED) і зв'язування названої сутності з об'єктом у базі знань (NEL). NER є напрямком досліджень, що швидко розвивається і є одною із найпопулярніших завдань NLP. Задача NER полягає у виділенні та класифікації за певними категоріями іменованих сутностей у тексті, в якій ми намагаємось отримати контекстуальне значення слів, використовуючи вкладення слів.

Список літератури

1. Gomez-Perez Jose Manuel, Denaux R., Garcia-Silva A. A Practical Guide to Hybrid Natural Language Processing. Combining Neural Models and Knowledge Graphs for NLP. Springer. 2020. 281 p.