Звіт про виконану роботу

Протягом січня 2020 року

Студента 3 курсу бакалаврату

Групи «комп'ютерна статистика»

Горбунова Даніела Денисовича

За період проведення зимових канікул, мені вдалося підготуватися до навчальної практики з інформаційних технологій. Процес підготовки включає в себе не тільки прочитання певної літератури, але й спроби закріпити вивчений матеріал на базі вирішення практичних завдань.

Безпосередньо по головній темі навчальної практики, а саме «Security Text Detection», виконано наступне:

1. Ознайомився з історією та головними завданнями, що відносяться до обробки природної мови (NLP). Побачив переваги у застосуванні алгоритмів машинного навчання замість прийомів строгої систематизації мовних правил у задачах NLP.

* <https://arxiv.org/pdf/1807.10854.pdf>
* <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32>
* <https://en.wikipedia.org/wiki/Natural_language_processing>
* <https://www.wonderflow.co/blog/natural-language-processing-examples>

1. Розглянута загальна концепція створення математичних моделей на основі текстових даних та певної задачі, яка пов’язана с методами обробки природної мови.

* <https://www.youtube.com/watch?v=rURRYI66E54>
* <https://towardsdatascience.com/what-the-heck-is-word-embedding-b30f67f01c81>
* <https://towardsdatascience.com/nlp-text-preprocessing-a-practical-guide-and-template-d80874676e79>

1. Детально розібрана теоретична частина про векторне представлення слів, з використанням одного з трьох найбільш розповсюджених інструментів для цього: word2vec, GloVe, fasttext.

* <https://arxiv.org/pdf/1301.3781.pdf> (word2vec)
* <https://towardsdatascience.com/an-implementation-guide-to-word2vec-using-numpy-and-google-sheets-13445eebd281> (word2vec)
* <https://nlp.stanford.edu/pubs/glove.pdf> (GloVe)
* <https://arxiv.org/pdf/1411.5595.pdf> (GloVe, word2vec – порівняльна характеристика)
* <https://fasttext.cc/docs/en/supervised-tutorial.html> (fasttext)
* <https://arxiv.org/pdf/1607.04606.pdf> (fasttext)

1. Повільні, але достатньо значні кроки в сторону вивчення теорії машинного навчання:

* <https://developers.google.com/machine-learning/crash-course/>

1. Освоєння програмної оболонки для інтерактивних обчислень мовою python – Jupyter Notebook. Базові поняття, принцип роботи не пройшли повз моєї уваги. Так звані перші кроки можна побачити у репозиторії, який було створено на початку семестру.

* <https://github.com/MilkyCousin/python-probability-statistics/tree/law_of_large_numbers> (невелика праця на тему «Закон великих чисел»)
* <https://jupyter.org/> (загальна інформація)

1. Google colab – середовище для публікації блокнотів Jupyter Notebook та розробки програм.

* <https://colab.research.google.com/>

1. Програмний пакет для аналізу даних в python:

* <https://pandas.pydata.org/pandas-docs/stable/getting_started/index.html>

Окрім необхідного матеріалу з обраної теми, виконані наступні дії:

1. Перші кроки у світ математичної статистики. Проаналізований матеріал на тему «Перевірка статистичних гіпотез».

* Теория вероятностей и математическая статистика / И. И. Гихман, А. В. Скороход, М. И. Ядренко. – 2-е изд., перераб. и доп. – К.: Выща шк. Головное изд-во, 1988. – 439 с.: ил. (стор.360, 421)
* Теория вероятностей. Сборник задач. Дороговцев А. Я., Сильвестров Д. С., Скороход А. В., Ядренко М. И. Издательское объединение «Вища школа», 1976, 384 с. (на украинском языке)
* <http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf> (стор. 297)

1. Покращив володіння стандартним методом головних компонент, що відноситься до математичної статистики.

* <https://arxiv.org/pdf/1404.1100.pdf>

1. Згадав деякі елементи теорії лінійної алгебри. Матричні перетворення, нормальна форма, власні числа та вектори, інші деталі.

* <http://www.mechmat.univ.kiev.ua/wp-content/uploads/2018/03/IM-Gel-fand-Lektsii-po-linejnoj-algebre-1.pdf> (стор.110-240)
* <http://www.mechmat.univ.kiev.ua/wp-content/uploads/2018/03/LinAlgAlgorithms.pdf> (стор.23-38)