

- державного університету. Серія Економічні науки. Вип. 34. С. 157 – 163. DOI: 10.32999/ksu2307-8030/2019-34-34
4. Соловйов А. І. (2014) Ефективне управління агровиробництвом на базі технологій точного землеробства. *Вісник ХНАУ. Серія : Економічні науки.* № 6. С. 169 – 176. URL: http://nbuv.gov.ua/UJRN/Vkhnau_ekon_2014_6_26.
5. McBratney A., Whelan B., Ancev T. (2005) Future Directions of Precision Agriculture. *Precision Agriculture.* No 6. P. 7 – 23.

УДК 004.622

Смиш О.Р. аспірант спеціальності 122 Комп'ютерні науки
Науковий керівник: Жежерун О.П., кандидат ф.-м. н., доцент кафедри
мультимедійних систем факультету інформатики
(Національний університет «Києво-Могилянська академія», м. Київ, Україна)

ВИПРАВЛЕННЯ ПОМИЛОК ЛЕМАТИЗАЦІЇ ТЕКСТУ ЗА ДОПОМОГОЮ СЛОВНИКА

Працюючи з обробкою природної мови важливим аспектом є точність отримуваних даних. У власному дисертаційному дослідженні, присвяченому створенню системи для розв'язування задач з геометрії, які записані природною українською мовою [1], головною і першочерговою частиною застосунку є модуль обробки тексту задач. Цей модуль поетапно виконує попереднє опрацювання сирого тексту, далі застосовує UDPipe [2] аналізатор для токенизації, тегування, лематизації та аналізу залежностей, потім проводить уніфікацію словосполук. Хоча для української мови UD 2.10 демонструє точність лематизації понад 97% [3], помилки трапляються, – інколи аналізатор може залишити початкове слово без змін або ж некоректно змінити це слово.

Оскільки лематизація тексту є суттєвим етапом обробки мови, варто підвищити точність результатів додатковими методами. Одним з таких підходів є використання сформованого словника слів зі словоформами, який можна використати для перевірки утворених після аналізатора слів. Великий електронний словник української мови (ВЕСУМ), що містить більш як 415 тисяч лем, призначений саме для використання в обробці української мови [4]. ВЕСУМ має таку структуру: з нового рядка записана лема слова, через пробіл записаний ключ цієї лемі (тег, що розрізняє різні слова), якщо для цього слова існують словоформи, то всі форми записуються з нового рядка через два пробіли та через один пробіл від слова записано ключ.

Очевидно, що використовуючи ВЕСУМ, як *.txt файл, час виконання алгоритму значно зростає. Для збільшення швидкодії вирішено оптимізувати словник: ключі слів прибрати, а структуру файлу замінити на *.json формат (де ключем будуть лемі слів, а словоформи – значеннями ключа). Таким чином швидкість проходження по словнику зростає. Алгоритм такий: взявши лематизоване після UDPipe слово, яке треба перевірити, проходимося по лемах у файлі, якщо є повний збіг зі словом, беремо наступне слово; якщо збігу немає, то здійснюємо пошук по всіх словоформах, якщо збіг є, то замінюємо слово на ключ, якщо ж збігу немає, повертаємо помилку, припускаючи, що такого слова не існує. Підсумовуючи, варто зазначити, що запропонований механізм підвищує кінцеву точність результату лематизації. Запропонований алгоритм може бути застосованим для роботи з обробкою природної мови, як для етапу налагодження моделі, так і в готових застосунках.

Список використаних джерел:

1. Жежерун, О., Смиш, О. (2020). Автоматизація розв'язування задач з планіметрії,

записаних природною українською мовою. Проблеми програмування, (4), 71-80. DOI: <https://doi.org/10.15407/pp2020.04.071>

2. Milan Straka (2018): UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In: Proceedings of CoNLL 2018: The SIGNLL Conference on Computational Natural Language Learning, pp. 197-207, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-948087-72-8

3. Zeman, Daniel; et al., 2022, Universal Dependencies 2.10, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-4758>.

4. Рисін А., Старко В. Великий електронний словник української мови (ВЕСУМ). Вебверсія 6.0.1. 2005-2022. URL: <https://r2u.org.ua/vesum/>

УДК 004.93

Хабарлак К.С. аспірант кафедри Системного аналізу та управління

(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна)

КОНФІГУРАЦІЯ ПІСЛЯ НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ

Сегментація зображень використовується для вирішення багатьох практичних задач, таких як аналіз медичних знімків, фотографій навколишнього середовища тощо. Найбільшу якість вирішення даної задачі демонструють згорткові нейронні мережі. Проблема є обчислювально складною. Великий час виконання нейронних мереж ускладнює їх впровадження на мобільних пристроях [1], особливо враховуючи той факт, що архітектуру нейронної мережі (і, як наслідок, час передбачення) зазвичай неможливо змінити без повторного навчання нейронної мережі.

Кілька останніх робіт [2; 3] пропонують модифікації процедури навчання нейронної мережі так, щоб результуюча мережа стала адаптивною. Зокрема, в [2] вперше було запропоновано підхід навчання та побудови адаптивної після навчання мережі (РТА).

Підхід РТА в задачі класифікації надає кілька переваг: 1) після завершення навчання можна змінити архітектуру мережі. В залежності від конфігурації час виконання становить від 80% до 107% часу роботи нединамічної нейронної мережі; 2) загальний час навчання зменшився на 14%; 3) точність прогнозування зросла з 96,74% для підходу без РТА до 97,85% для підходу РТА.

Проте запропонований підхід не застосовувався до завдання семантичної сегментації. У цій роботі побудовано нейронну мережу U-Net+РТА, яка здатна вирішити проблему сегментації зображення, а її конфігурацію можна змінювати після завершення навчання нейронної мережі. Отже, метою є розробка адаптивної після навчання мережі для вирішення задачі сегментації.

Для досягнення поставленої мети в роботі сформовані і вирішені такі завдання:

1. Розроблено власну реалізацію алгоритму нейро-мережевої сегментації зображень U-Net.

2. Розроблено модифікацію підходу адаптивних після навчання нейронних мереж (РТА), що є застосовним для мережі U-Net та може ефективно вирішувати задачу сегментації зображень.

3. Експериментально підтверджено ефективність розробленого підходу із врахуванням як якості вирішення задачі, так і часу, необхідного для передбачення маски сегментації.

Основою роботи є модель U-Net, яка показала хороші результати в задачі