

ВІДГУК

офіційного опонента на дисертаційну роботу
Шаптали Романа Віталійовича
на тему «Класифікація документів на основі векторних представлень
словників при обробці природної мови у малоресурсному середовищі»,
представлену на здобуття ступеня доктора філософії
в галузі знань Інформаційні технології
за спеціальністю 122 – Комп'ютерні науки

Актуальність теми дисертації.

Актуальність теми дисертації полягає у важливості обробки природної мови як однієї з найбільш прикладних областей застосування комп'ютерних наук та штучного інтелекту. Це стосується накопичення знань та методів для автоматичної обробки текстової інформації, яка генерується під час людської соціальної активності. Важливим завданням є надання комп'ютерам здатності «розуміти» зміст документів з урахуванням контексту та мовних нюансів.

Проблема поглиблюється при обробці природної мови у малоресурсних середовищах. Малоресурсність середовищ відзначається обмеженою доступністю даних, які комп'ютерні системи використовують для навчання моделей мови в конкретних прикладних областях.

Дослідження з обробки природної мови часто спрямовані на розвиток методів за наявності великої кількості даних, зокрема при роботі з англійською мовою. Малоресурсні мови, наприклад українська, значно менш популярні у дослідженнях, а отже кількість даних для ефективного навчання моделей обробки природних мов та роботи з ними недостатня. Тому розвиток методів обробки природної мови у малоресурсних середовищах – актуальна та важлива задача.

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1. Запропоновано метод класифікації документів на основі векторних представлень словників при обробці природної мови, який через поєднання векторних представлень документів та векторних представлень слів зі словника синонімів покращує F1-міру якості класифікації документів у малоресурсному середовищі.

2. Запропоновано векторну модель слів зі словника синонімів, яка будується на основі графового представлення словника за допомогою методів кодування вузлів графу.

3. Модифіковано методи злиття векторних представлень слів, а саме конкатенації та зваженої суми, через додатковий крок пошуку відповідника слову з документа у словнику синонімів на основі критеріїв міжрядкової відстані.

Достовірність результатів дисертації забезпечена використанням математичних методів штучного інтелекту та інтелектуального аналізу даних. Результати експериментів, які проводились на основі набору даних петицій до Київської міської ради, були перевірені на статистичну значущість за допомогою Хі-квадрат тесту за методом МакНемара. Цей підхід дозволив підтвердити високу достовірність та надійність наукових висновків дослідження.

У ході дисертаційного дослідження було проаналізовано актуальну літературу, яка стосується тематики обробки природної мови та методів штучного інтелекту. Систематичний огляд наукових джерел допоміг поглибити розуміння сучасних тенденцій та підходів у даній галузі. Були виявлені перспективні напрямки досліджень, а також розкрито питання, які є необхідними для подальшого розвитку методів обробки природної мови в малоресурсних умовах.

Одним із ключових елементів дисертації є запропонований метод класифікації документів, призначений для обробки природної мови у малоресурсному середовищі. Цей метод суттєво відрізняється від існуючих підходів, які базуються на доповненні даних з використанням словників. У даному дослідженні здійснюється поєднання векторних представлень

документів з векторними представленнями елементів лінгвістичних словників. Це дозволяє забезпечити більш точну та ефективну класифікацію документів у малоресурсних умовах. Завдяки такому підходу до обробки природної мови, результати дослідження мають значний потенціал для подальших практичних застосувань та розвитку області.

Отже, в дисертаційній роботі поставлене наукове завдання розробки методів обробки природної мови на основі векторних представлень словників у малоресурсному середовищі виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.

За своїм змістом дисертаційна робота здобувача Шаптали Р.В. повністю відповідає Стандарту вищої освіти зі спеціальності 122 – Комп'ютерні науки та напрямкам досліджень відповідно до освітньої програми Комп'ютерні науки.

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям Комп'ютерні науки.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Шаптали Романа Віталійовича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

Мова та стиль викладення результатів.

Дисертаційна робота написана українською мовою.

Робота характеризується високою послідовністю та логічною структурою викладення, що дозволяє читачеві з легкістю розуміти розвиток дослідження та логічні зв'язки між розділами. Представлення інформації чітке та зрозуміле, що

сприяє засвоєнню основних понять та методів, які використовуються в дослідженні.

Дисертація демонструє високий професіоналізм та володіння загальноприйнятою термінологією у даній науковій області. Автор вміло використовує терміни та поняття, що визначені в наукових джерелах, це дозволяє дисертації бути актуальною та зрозумілою для наукової спільноти, а також забезпечує легке порівняння з існуючими підходами та методиками.

Також варто відзначити стиль мовлення, який характеризується виразністю та чіткістю. Автор використовує влучні приклади та ілюстрації, що допомагають усвідомити основні ідеї та результати дослідження. Структура роботи включає чітко сформульовані мету та завдання, а також змістовні розділи, які підкреслюють актуальність та наукову цінність дослідження.

Дисертація складається з вступу, трьох розділів, висновків, списку літератури та додатків. Загальний обсяг дисертації 151 сторінка.

У вступі викладаються мета та основні завдання дослідження, а також обґрунтовується його актуальність. Описуються проблематичні аспекти, що виникають у зв'язку з існуючими підходами, та наголошується на науковій і практичній новизні досягнутих результатів. Здобувач вказує на власний внесок у дослідження, подає інформацію про апробацію матеріалів дисертації та перелік публікацій, пов'язаних із темою дослідження.

У першому розділі дисертації викладено аналіз сучасних методів та проведених досліджень в галузі обробки природних мов у малоресурсних середовищах. В основі даного аналізу лежить систематичний опис різноманітних методів обробки природних мов, які знаходять застосування в умовах обмеженого обсягу даних. Для цього була створена класифікація методів обробки природних мов у малоресурсних середовищах, що дозволило виділити основні групи підходів та способів розв'язання завдань в контексті малоресурсного середовища. Аналіз кожного класу підходів виявив обмеження та припущення, що його характеризують, а також область застосування, для якої він є оптимальним.

У другому розділі дисертації висвітлено методику побудови векторних представлень словників, а також методи злиття таких векторних представлень. Значна увага приділена детальному аналізу підходів, що лежать в основі створення векторних представлень графів. Зокрема, розглянуті методи на основі факторизації графу, глибокого навчання та випадкових блукань. У даному розділі здобувач запропонував власну модифікацію методу злиття векторних представлень слів, додавши етап пошуку відповідностей між словами на основі міжрядкових відстаней.

Третій розділ дисертації присвячено детальному опису експериментальних результатів, отриманих за допомогою запропонованого автором методу. З метою перевірки якості даного методу, та його порівняння із загальновідомими підходами, використано множину даних петицій до Київської міської ради з розміткою за темами звернень. У розділі проведено ретельний аналіз набору даних, його поділ на вибірки для тестування та тренування. Було зіставлено результати експериментів з різними архітектурами та компонентами, а також перевірена статистична значущість отриманих результатів.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

Оприлюднення результатів дисертаційної роботи.

Наукові результати дисертації висвітлені у 7 наукових публікаціях здобувача, серед яких: 5 статей у наукових фахових виданнях України, 4 з яких включені на дату опублікування до переліку наукових фахових видань України за спеціальністю 122 Комп'ютерні науки, та 1 стаття у періодичному науковому виданні, проіндексованому у наукометричній базі даних Scopus.

Також результати дисертації були апробовані на 1 науковій фаховій конференції.

Усі публікації здобувача демонструють високий науковий рівень, і в них ретельно висвітлюються основні наукові результати досліджень. Здобувач зробив значний особистий внесок у публікації, особливо у розкритті

експериментальних аспектів роботи. Важливо відзначити, що в усіх публікаціях збережено принципи академічної доброчесності, і в жодній з них не було порушено науково-етичних норм.

Таким чином, наукові результати описані в дисертаційній роботі повністю висвітлені у наукових публікаціях здобувача.

Недоліки та зауваження до дисертаційної роботи.

1. У дисертації наведена класифікація методів обробки природної мови у малоресурсному середовищі, але деякі класи методів описані поверхнево. Наприклад, для метанавчання не зазначено специфіки застосування саме у малоресурсних середовищах.

2. У роботі зазначено, що для вирішення задачі класифікації було проведено підбір гіперпараметрів багат шарового перцептронну, але аналіз впливу кожного гіперпараметра на результат експериментів не наводиться. Це допомогло б зрозуміти оптимальну область гіперпараметрів для задачі.

3. У дослідженні використовується метод усереднення векторних представлень слів, але у літературі для покращення результатів цього процесу часто застосовується механізм уваги, коли мережа самостійно навчається приділяти більшу вагу конкретним частинам тексту. Варто було провести експерименти із застосуванням механізму уваги.

4. Набір даних, використаний в експериментальній частині дослідження – незбалансований, що додатково ускладнює задачу. Було б корисно спочатку дослідити метод у контрольованому середовищі зі збалансованими даними.

5. Попри опис мотивації вибору зваженої F1-міри як ключової метрики дослідження, не наведено формули її обчислення. Це дозволило б покращити сприйняття матеріалу науковою спільнотою.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.


Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Шаптали Романа Віталійовича на тему «Класифікація документів на основі векторних представлень словників при обробці природної мови у малоресурсному середовищі» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для інформаційних технологій. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п. 6–9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Шаптала Роман Віталійович заслуговує на присудження ступеня доктора філософії в галузі знань Інформаційні технології за спеціальністю 122 – Комп'ютерні науки.

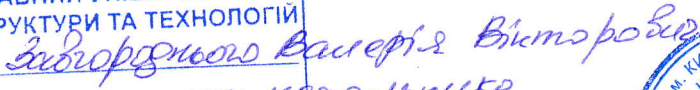
Офіційний опонент:

завідувач кафедри
інформаційних технологій
Державного університету
інфраструктури та технологій,
д.т.н., професор

/  /

Валерій ЗАВГОРОДНИЙ

М.П. «14» серпня 2023 року

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФРАСТРУКТУРИ ТА ТЕХНОЛОГІЙ
Підпис: 
Засвідчує: 