

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

на дисертаційну роботу *Мельникової Наталії Іванівни* на тему:

«Моделі та методи підтримки персоналізованих рішень

у медичних системах», подану на здобуття наукового ступеня

доктора технічних наук за спеціальністю

05.13.23 - системи та засоби штучного інтелекту

Актуальність теми дослідження. Проблема персоналізації рішень спостерігається в бізнесі, навчанні, консалтингових та маркетингових дослідженнях, а особливо - у медицині. Персоналізація актуалізується за рахунок важливості прийняття рішень респондентами при вирішенні задач, з урахуванням індивідуальних особливостей досліджуваних об'єктів. В системі охорони здоров'я усі процеси дослідження та визначення стану пацієнта вимагають оцінки якості його персоналізованих даних, а також джерел їхнього надходження.

Прийняття персоніфікованих медичних рішень для вибору лікування та способу діагностики захворювань можуть бути перетворені в класифікаційні або передбачувані задачі машинного навчання, де оптимальним рішенням для людини є правило прийняття рішень, яке дає кращий майбутній клінічний результат або покращує точність діагнозу. Проте при аналізі складних медичних даних виникають певні ускладнення. З одного боку, це неповнота або відсутність даних, а з іншого боку, це нові типи даних і більші масштаби даних вимагають інновацій, що поєднують статистичне моделювання, доменні знання та інформаційні технології.

Визначення необхідних індивідуальних характеристик для розв'язання задачі персоналізації залежить від ключових факторів ідентифікації об'єкта. Для формалізованого подання досліджуваного об'єкта в медицині беруть до уваги основні параметри його загального стану з визначеними його характеристиками. Аналізуючи стан хворого під час лікування, експерти визнають, що важливим показником одужання є позитивна динаміка зміни основних показників загального стану у часі.

Існуючі прикладні системи, орієнтовані на забезпечення персоналізації рішень, в рамках виду робіт при опрацюванні індивідуальних медичних даних пацієнтів,

характеризуються процесом прогнозування, яке спрямоване лише для одного пацієнта, а не для групи пацієнтів з однаковими параметрами, та які у часі змінюватися під впливом різних факторів, що значно знижує якість прийнятих персоналізованих медичних рішень.

Тому при опрацюванні таких даних, що у свою чергу характеризуються наявністю викидів, шумів, надлишковості та пропусків даних, не відповідають очікуваній поведінці стану пацієнта та не несуть інформаційного змісту для аналізу вибірки, виникає потреба пошуку комплексного підходу щодо введення евристик, припущення про незалежність параметрів, пошуку асоціативних залежностей з малою підтримкою, пошуку оптимального рішення, підвищення точності прийняття лікарських рішень.

Дисертаційна робота Мельникової Н.І. присвячена вирішенню актуальної науково-прикладної проблеми розроблення нових та удосконалення існуючих методів щодо підвищення точності процесу обробки великих та малих мультимодальних медичних даних для адаптації персоналізованих медичних рішень до пацієнта.

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації, їхня достовірність. Наукові положення, висновки та рекомендації, сформульовані в дисертації, повною мірою обґрунтовані, оскільки вони логічно випливають із результатів, отриманих за допомогою чітких математичних викладок з коректним використанням теоретичних основ реляційної алгебри, методів прийняття рішень та машинного навчання. Логічна послідовність етапів проведених досліджень також зумовлює коректність викладених положень.

Достовірність одержаних результатів базується на обґрунтованості більшості зроблених припущень, перевірці застосування моделі відображення стану пацієнта в багатовимірному просторі умов та розроблених методів, алгоритмів щодо опрацювання, аналізу персоналізованих даних пацієнта в задачах класифікації, кластеризації та прогнозування, а також їхню узгодженість з результатами інших дослідників та успішній апробації результатів дослідження на науково-технічних конференціях.

Наукова новизна дисертаційної роботи полягає в наступному.

Вперше:

– розроблено модель відображення стану пацієнта в багатовимірному просторі умов, яка за рахунок додаткових вимірів та параметрів забезпечила

підвищення точності прогнозування;

- розроблено метод двоетапної обробки малих наборів даних на основі ієрархічного предиктора, який за рахунок кластеризації та прогнозування забезпечує підвищення точності опрацювання нових вхідних наборів даних;

- розроблено нову схему стекування моделей машинного навчання, яка за рахунок врахування основних ознак та повторного навчання забезпечує підвищення точності прогнозування даних;

- розроблено метод зменшення розмірності вхідних даних, який за рахунок класифікаторів, асоціативних правил та узагальненого рангу ознак на основі індексу Жакара забезпечує підвищення точності вибору пріоритетних ознак на великих наборах даних.

Удосконалено:

- метод заповнення пропусків даних, який за рахунок використання асоціативних та продукційних правил забезпечує підвищення стійкості моделі до помилок даних;

- метод пошуку зміни стану пацієнта, який за рахунок використання простору умов та аналізу характеристик стану забезпечує підвищення точності підбору схеми лікування;

- метод консолідації мультимодальних даних, який за рахунок попереднього визначення структури та семантики даних забезпечує агрегування даних з різною структурою.

Одержала подальший розвиток теорія підтримки прийняття персоналізованих рішень, яка ґрунтується на розробленні моделі відображення стану пацієнта в багатовимірному просторі умов та розроблених методах двоетапної обробки малих наборів даних, групування моделей машинного навчання, зменшення розмірності вхідних даних, заповнення пропусків даних, пошуку зміни стану пацієнта та консолідації мультимодальних даних, що забезпечило адаптацію прийняття персоналізованих медичних рішень.

Значення одержаних результатів для науки і практики. З допомогою розроблених методів та засобів:

- розроблено модель простору станів пацієнта, що подано у вигляді

гіперкуба, як відображення функціонального відношення загального стану пацієнта;

– розроблено паралельний та розподілений режим опрацювання мультимодальних персоналізованих медичних даних, що є удосконаленням методу упорядкованого пошуку та надає чіткості та направленості у пошуку рішень стосовно вибору цільових схем лікування, що дає змогу зменшити ймовірність появи похибки при виборі схеми лікування;

– введено нову абстракцію керування даними – простору умов для прогнозування цільових змінних, що дало змогу підвищити точність прогнозування цільових показників в підмножині простору умов, що забезпечило індивідуальний підхід до моніторингу стану пацієнта на основі тривалого спостереження та контролю лікаря;

– розроблено стекінгову модель, для забезпечення точності прогнозування даних та паралелізації процесу обробки даних, що на відміну від існуючих моделей забезпечує універсальність пошуку рішень щодо малих та великих наборів персоналізованих даних;

– розроблено два алгоритми ймовірнісної продукційної залежності для заповнення відсутніх медичних персоналізованих даних, що забезпечують збереження характеристик стійкості до помилок у даних.

Практичну цінність роботи підтверджує впровадження її результатів на підприємстві та у місцевих клініках, а також використання її результатів при виконанні досліджень у міжнародних, національних проектах та держбюджетних темах Національного університету «Львівська політехніка».

Рекомендації щодо використання результатів дисертації. Запропонований підхід, методи, моделі та програмні засоби можуть бути використані розробниками інформаційних інтелектуальних системах при створенні експертних, консалтингових системах пошуку рішень щодо опрацювання мультимодальних персоналізованих даних досліджуваного об'єкта у медичній, соціальній та військовій областях.

Повнота викладу в опублікованих працях. Основні результати дисертаційного дослідження опубліковано в 56 наукових працях: 21 праця опублікована у міжнародних та фахових наукових виданнях, з яких 5 статей опубліковано в журналах, індексованих у наукометричних базах, з квантилями Q1-Q2,

14 статей у фахових іноземних виданнях та у наукових фахових виданнях України, 2 монографії та 35 публікацій у матеріалах конференцій, 23 з яких індексуються у наукометричній базі Scopus.

Аналіз наукових праць та тез доповідей на міжнародних конференціях показав, що в них в достатньому обсязі опубліковано всі основні результати дисертаційного дослідження, апробація проведена на належному рівні.

Оцінка змісту дисертаційної роботи, її завершеність. Дисертаційна робота є завершеною науковою працею. Вона складається із вступу, шести розділів, висновків, переліку літературних джерел із 302 найменування та додатку. Загальний обсяг дисертації складає 293 сторінки, з них 239 сторінок основного тексту.

У вступі обґрунтовується актуальність теми, формулюється мета та основні задачі досліджень, подається короткий зміст роботи.

У першому розділі введено поняття персоналізації медичних даних та персоналізованої медицини, що характеризуються процесом пошуку рішень лікаря для розв'язання проблеми призначення або корекції лікування пацієнта.

Проаналізовано існуючі практичні рішення, що відображають основні підходи до використання медичних даних та урахування персоналізації. За результатами аналізу існуючих систем штучного інтелекту у сфері медицини для вирішення проблем діагностики захворювань, дослідження геному, розробки ліків, медичної візуалізації та ін. виокремлено проблеми, що ще залишаються не розкритими, а саме: консолідація даних про хворого, його індивідуальних особливостей, враховуючи міжнародні стандарти лікування та існуючу фармацевтичну продукцію; визначення персоналізованого лікування та прогнозування результатів його застосування.

Проведено порівняльний аналіз класичних методів визначено обмеження, які характерні при опрацюванні медичних персоналізованих даних. Визначені обмеження існуючих методів опрацювання та аналізу великих та малих наборів мультимодальних медичних наборів даних дали змогу сформулювати науково-прикладну проблему.

У другому розділі введено модель стану пацієнта, яка спрощено відображає структуру параметрів пацієнта і зв'язків між ними, визначено продукційні правила, які формулюють рішення щодо перегляду та зміни тактики лікування, а саме стратегічних

рішень щодо зміни стану, зміни медикаментів, зміни дозування, перелік лабораторних досліджень, фізіотерапевтичних заходів та ін.

Результатами другого розділу є модель простору станів пацієнта, яку представлено у вигляді гіперкуба, як відображення функціонального відношення загального стану пацієнта; метод пошуку шаблонів, який базується на модифікації методу асоціативних правил, що дозволяє зменшити трудомісткість і використовувати паралельний та розподілений режим для розрахунку; метод консолідації мультимодальних даних за рахунок попереднього визначення структур даних та узгодження семантики, а також проведено аналіз проблеми опрацювання мультимодальних персоналізованих медичних даних та попередньої обробки даних та аналізу якості моделі даних.

У третьому розділі проведено аналіз опрацювання медичних персоналізованих даних та досліджено процес прогнозування, та визначені етапи препроцесингу отриманих даних, здійснено класифікацію завдань видобутку даних, що орієнтовані на передбачення даних та формулювання рішень, визначені інструменти для очистки, попередньої обробки, аналізу даних та виявлення взаємозалежностей між ними, що мають великий вплив на процеси опрацювання даних в системах діагностування та прогнозування станів пацієнта.

Результатом розділу є розроблення методу двоетапної розробки на основі ієрархічного предиктора, що забезпечує підвищення точності процесу узагальнення результатів на малих наборах даних, за рахунок ієрархічної класифікації та поєднання різних моделей машинного навчання, методу групування моделей машинного навчання, як стекінгову модель, для забезпечення точності прогнозування даних, що на відміну від існуючих моделей забезпечує універсальність пошуку рішень щодо малих та великих наборів персоналізованих даних.

У четвертому розділі введено модель вибору функцій гібридного ансамблю, що містить кілька селекторів за допомогою агрегування результатів, які будуть використовуватися на етапі попередньої обробки.

Проведено порівняння точності прогнозування для стандартних моделей зменшення розмірності під час наступних кроків, для цього використано аналіз головних компонентів із вісьмома компонентами

Розроблено ансамбль моделей машинного навчання для вибору пріоритетних ознак на великих наборах даних, який складається з класифікаторів, асоціативних правил та узагальненого рангу ознак на основі індексу Жакара, що дозволяє уникнути кореляції ознак та збільшує узагальнення моделі. Розроблено модель тришарового стекування ансамблю методів, що дає можливість об'єднати асоціативну класифікацію зі слабкими класифікаторами в ансамбль для узагальнення результатів; описані етапи укладання.

Результатом розділу є розроблення методу зменшення розмірності вхідних даних на основі ансамблю моделей слабких предикторів для вибору найважливіших ознак, що дає змогу уникнути кореляції результатів слабких предикторів і збільшує узагальнення моделі.

У п'ятому розділі визначені типи аномалій та шляхи їхнього пошуку, що дає можливість диференціації підходів щодо аналізу персоналізованих медичних даних та визначені особливості їхнього застосування і основні проблеми, що виникають при обробці медичних персоналізованих даних, за рахунок чого постає проблема розробки ефективної стратегії аналізу даних, яка може бути застосована до розосереджених баз даних різних доменів.

Результатом розділу є розроблення та апробація методу заповнення відсутніх даних, який ґрунтується на пошуку подібності даних, що можуть впливати на доменні значення та функціональні залежності між ними і забезпечує їхнє включення у навчальні дані, що дає змогу забезпечити стійкість до помилок даних, паралелізації обчислень та аналізу різнотипних даних.

У шостому розділі спроектовано архітектуру системи підтримки прийняття медичних рішень щодо прогнозування станів пацієнта на підставі опрацювання та аналізу персоналізованих медичних даних.

Подано результати імплементації рішень у системі підтримки прийняття рішень для лікування хворих на ПІД з порушеннями антитілоутвореннями, результати апробації програмного модуля щодо прогнозування динаміки поширення COVID-19 з урахуванням процесу класифікації пацієнтів відповідно до стану, проведення оцінки помилок одержаних рішень та результати імплементації системи для виявлення помилок, що дає змогу експериментувати із різними параметрами для моделі, знову і знову на одних і тих же самих даних.

У додатку наведені акти впровадження результатів дисертаційної роботи.

Загалом дисертаційна робота написана грамотно, технічною мовою. Матеріали досліджень подано логічно, послідовно та доказово. Текст реферату відповідає змісту дисертаційної роботи.

Відповідність дисертації паспорту спеціальності. Дисертація повністю відповідає паспорту спеціальності 05.13.23. - системи та засоби штучного інтелекту, а саме за такими розділами:

- розроблення теоретичних засад створення та застосування систем штучного інтелекту різноманітного призначення;
- моделювання інтелектуальної діяльності людини та його застосування в системах штучного інтелекту;
- розроблення принципів, методів й архітектурних розв'язань побудови баз знань і технологія їх експертування (експертні та багатоагентні системи).

Помічені недоліки та зауваження до дисертації можна зазначити такі.

1. Запропонований метод заповнення пропусків не зможе працювати на неструктурованих даних.

2. Підрозділ 4.2 «Розробка методу зменшення розмірності вхідних даних» в основному побудовано на експериментах. Варто було приділити більшу увагу його теоретичному обґрунтуванню.

3. Підрозділ 1.2.3 «Аналіз методів штучного інтелекту для прикладних задач персоналізації» мав би бути краще структурований, оскільки включає аналіз табличних даних, часових рядів, зображень та ін.

4. Рис 2.1 нечитабельний через малий розмір шрифту тексту.

5. Метод консолідації мультимодальний даних мав би бути поданий з урахування описаного у розділі 2 простору умов.

6. Як результат оцінювання точності ієрархічного предиктора доцільно було б порахувати ймовірність успіху ансамблю (формула (3.10)).

7. Подання матеріалу дисертаційної роботи є незвичним, оскільки практична імплементація кожного методу наводиться в тому ж розділі, що і теорія, а не в окремому розділі експериментів та аналізу результатів.

8. Текст дисертації містить пунктуаційні, поліграфічні та стилістичні помилки.

9. Формула (3.15) подає частковий випадок зважування результатів ансамблю та фактично дублює формулу (3.14).

Зазначені недоліки не знижують значущість і якість одержаних результатів.

Незважаючи на зауваження, робота справляє добре враження рівнем аналізу об'єкта досліджень та різноманітністю використаного математичного апарату.

Висновок. В цілому дисертаційна робота Мельникової Наталії Іванівни «Моделі та методи підтримки персоналізованих рішень у медичних системах» є завершеною науковою роботою, в якій отримані нові науково обґрунтовані результати, що в сукупності вирішують поставлену науково-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізація результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

За актуальністю обраної теми, обсягом та рівнем виконаних досліджень, повнотою охоплення наукової проблеми, новизною і ступінню обґрунтованості отриманих результатів, практичних висновків та рекомендацій робота задовольняє вимогам, що пред'являються до докторських дисертацій.

Таким чином, дисертаційна робота відповідає вимогам п. 7 та 9 Порядку присудження та позбавлення наукового ступеня доктора наук, затвердженого постановою Кабінету Міністрів України, а її автор, Мельникова Наталія Іванівна, заслуговує на присудження їй наукового ступеня доктора технічних наук за спеціальністю 05.13.23 - системи та засоби штучного інтелекту.

Офіційний опонент,

завідувач кафедри програмних засобів,

Національного університету «Запорізька політехніка»,

доктор технічних наук, професор

Сергій СУББОТІН

Підпис проф. С.О. Субботіна засвідчую.



Курбача ІВ