

ВИКОРИСТАННЯ СУЧАСНИХ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИРІШЕННЯ ПРОБЛЕМ ГЕНОМІКИ

Запорізька державна інженерна академія, кафедра ПЗАС

Наразі геноміка вважається однією з передових галузей біології та медицини. Предметом дослідження геноміки є організація та функціонування геномів живих організмів. Методи геноміки спрямовані на розшифровку нових закономірностей біологічних систем і процесів. Геноміка людини є основою молекулярної медицини і має найважливіше значення для розробки методів діагностики, лікування і профілактики спадкових і неспадкових захворювань.

Завдання геноміки - застосувати отримані знання для поліпшення якості життя людини. Геноміка дозволяє виразити сутність організму - його потенційні можливості, видові (і навіть індивідуальні) відмінності від інших організмів, передбачити реакцію на зовнішні впливи, знаючи послідовність нуклеотидів в кожному з генів і число генів.

Мета геноміки - отримання інформації про всіх потенційних властивості клітини, які не реалізуються на даний момент, наприклад, "мовчазні гени". Мінімальні геноми мікроорганізмів деяких видів складаються з декількох сотень генів. Геном людини наближається до 25000 тисяч генів. Розміри окремих генів варіюють приблизно від однієї тисячі пар нуклеотидів і вище. Таким чином, кількість пар нуклеотидів, складових індивідуального геному, вимірюється як мінімум сотнями тисяч, звичайно ж багатьма мільйонами пар нуклеотидів. Отже, для повного знання геному організму треба визначити послідовність кількох мільйонів пар нуклеотидів (А - Т — аденін - тіамін, G - C — гуанін-цитозін).

Дослідження в Геноміці почалися з проекту "Геном Людини", міжнародної дослідницької роботи яка визначала ДНК послідовність усього еталонного геному людини. Цей проект розпочався в 1990 році та завершився у квітні 2003 року. Така важлива ДНК послідовність геному людини була визначена, всі початкові цілі проекту були виконані, а деякі також були перевищені. Всі ці досяги потребували розвитку та покращення сучасних програмних технологій та активної програми для забезпечення підтримки розробки технологій.

Геноміка є дуже пов'язаною з терміном "точна медицина" (персоналізована медицина). Точна медицина — це підхід, який охоплює генетику, оточуючу середу та поведінку людини, що досліджується, та розробляє індивідуальний метод лікування для окремого індивіду; на відміну від традиційного підходу який має рекомендаційний характер для широких груп населення.

Наразі існує два бар'єра, які заважають реалізації точної медицини: високі кошти та технологічні обмеження. Для того, щоб мати можливість впорядкувати та оброблювати великі обсяги даних пацієнта, проаналізувати їх, а також знизити трати коштів, вчені застосовують методи машинного навчання.

Метою роботи є дослідження можливості використання методів штучного інтелекту для вирішення проблем геноміки, проаналізувати існуючі рішення та запропонувати нові потенціально вагомні підгалузі геноміки, задачі яких можуть бути поліпшені, використовуючі інтелектуальні та евристичні алгоритми.

На сьогоднішній день є дві найпоширеніших категорії застосування машинного навчання в геноміці:

- Секвенсування геному (точна медицина): Вчені використовують машинне навчання для вивчення та ідентифікації шаблонів з великим об'ємом генетичних наборів даних. Ці шаблони потім транслюються на комп'ютер в комп'ютерні мільйони, які можуть допомогти передбачати можливість розвитку окремих

захворювань у індивіду або допомогати інформувати про потенціальні засоби терапії;

- Business-to-Customer (прямий підхід для користувача) геноміка: ця категорія охоплює компанії, що пропонують секвенсування геному як функція для індивідуальних клієнтів. Компанії найчастіш використовують машинне навчання для досягнення більшої глибини в інтерпритації генетичної інформації, такої, наприклад, як вплив генів індивіду на вагу цієї людини.

В грудні 2017 року командою Google Brain Team було представлено платформу для впровадження експериментів в області геноміки. Інструменти платформи надають можливість будувати справжній біологічний пайплайн. Деякі інструменти, використовуючи штучний інтелект надають можливість побудувати більш чітке представлення геному людини з наданих даних секвенсування. Глибока модель навчання представлена у деяких інструментах DeepVariant, не має спеціалізованих знань про геноміку чи секвенсування, але у 2017 році вона виграла найвищу нагороду з точністю передбачування одиничних мутацій у престижній команді True Challenge precisionFDA, перевершуючи найсучасніші методи. З того часу рівень помилок було зменшено більш ніж на 50%. Окрім цього DeepVariant також, як багато інших розробників у цій галузі (більшість розробляють пропріетарне програмне забезпечення) використовують конволюційні нейронні мережі для обробки та аналізу геному. Наразі ця платформа активно використовується серед вчених як невеликих лабораторій, так й серед дослідників крупних фармацевтичних компаній.

Таким чином можна побачити, що використання методів штучного інтелекту в геноміці є новою, неширокодослідженою областю, однак яка вже довела свою ефективність та надає багато перспектив для пришвидшення швидкості, покращення якості научних досліджень в сфері геноміки.

Висновки:

- Проведено дослідження можливостей використання методів штучного інтелекту для вирішення проблем геноміки.
- Проведено дослідження існуючих технологічних та обчислювальних проблем в сфері геноміки.
- Були розглянуті та вивчені основні теоретичні відомості по існуючим засобам аналізу та обробки даних в геноміці;
- Проведено аналіз існуючих впроваджені системи штучного інтелекту;
- Виявлено основні проблеми геноміки, які вже вирішуються засобами штучного інтелекту
- Виявлено потенціальні напрями у геноміці, які можуть бути покращенні за допомогою машинного навчання.
- Доведено актуальність розвитку та реалізації теми роботи в подальшому.

Література

1. Using Artificial Intelligence to Revolutionise Precision Medicine // URL: <http://www.frontlinegenomics.com/news/16724/using-artificial-intelligence-revolutionise-precision-medicine/>
2. DeepVariant: Highly Accurate Genomes With Deep Neural Networks // URL: <https://research.googleblog.com/2017/12/deepvariant-highly-accurate-genomes.html>
3. Северин С.Е. Біохімія і медицина - нові підходи та досягнення / С.Е. Северин. - М.: Російський лікар, 1998. - 94 с.