

ДОСЛІДЖЕННЯ ТА ІНТЕГРАЦІЯ АЛГОРИТМУ SVM В ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ ТЕХНОЛОГІЯХ ДЛЯ ВИБОРУ ПРОФЕСІЇ ЛІКАРЯ

Олексій Ізмалков

ORCID: <https://orcid.org/0009-0005-3732-7474>

Дніпровський національний університет імені Олеся Гончара, Дніпро

Вступ

Стрімкий розвиток інформаційно-вимірювальних технологій (ІВТ) проклав шлях до значних покращень у різних галузях, включаючи медичну сферу. Однією з ключових сфер, де ІМТ можуть мати значний вплив, є допомога людям у виборі медичної професії. У цій статті розглядається роль алгоритмів машин опорних векторів (SVM) у цьому контексті, досліджується їхній потенціал для покращення процесу прийняття рішень майбутніми медичними працівниками.

МЕТА ТА ЗАДАЧІ

Визначити можливості використання інтегрованого рішення алгоритму SVM та його аналіз та обробку.

Розуміння алгоритму SVM.

Машина опорних векторів (Support Vector Machine, SVM) – це алгоритм машинного навчання, який широко використовується для задач класифікації та регресії. Основний принцип SVM полягає в тому, щоб знайти гіперплощину, яка найкраще розділяє набір даних на класи. Це досягається шляхом максимізації відстані між точками даних різних класів, що гарантує, що класифікатор має найкращу можливу здатність до узагальнення на невидимих даних. У контексті вибору медичної кар'єри SVM можна використовувати для аналізу різних факторів, таких як особисті інтереси, академічна успішність, набори навичок та інші відповідні атрибути, щоб класифікувати і передбачити найбільш підходящу медичну професію для людини. Ефективний збір та попередня обробка даних є важливим етапом для використання алгоритму SVM у прогнозуванні медичних професій для студентів. Академічні записи, психологічні оцінки, інтереси та здібності є основними джерелами даних. Для обробки відсутніх значень використовуються методи імпутації та прогностичного моделювання. Нормалізація ознак використовується для забезпечення однакового впливу ознак на модель. Кодування категоріальних змінних перетворює категоріальні дані у числовий формат. Інженерія функцій

дозволяє створювати нові функції для покращення продуктивності моделі. Збір і обробка даних створюють підґрунтя для розробки точної та надійної моделі, яка може забезпечити персоналізовану проф-орієнтацію для майбутніх медичних працівників.

МАТЕРІАЛИ ТА МЕТОДИ

Застосування SVM у виборі медичної кар'єри.

Збір та попередня обробка даних.

– Перший крок передбачає збір великої кількості даних про студентів, включаючи академічні записи, психологічні оцінки, інтереси та здібності. Ці дані слугують вхідними даними для SVM-моделі.

– Попередня обробка даних для обробки пропущених значень, нормалізації ознак і кодування категорійних змінних забезпечує оптимальну роботу алгоритму SVM.

Збір даних.

а) Академічні записи.

– Типи даних: Сюди входять оцінки, результати тестів, курсові роботи, записи про відвідування та інші показники академічної успішності.

– Джерела: Шкільні записи, бази даних стандартизованих тестів та дані, надані учнями самостійно.

– Мета: Академічні записи дають змогу оцінити базу знань студента, його здатність до навчання та стабільність успішності, що є критично важливими факторами для визначення придатності до різних медичних професій.

б) Психологічні оцінки.

– Типи даних: Особистісні тести (наприклад, індикатор типу Майерс-Бріггс), тести когнітивних здібностей та оцінка емоційного інтелекту.

– Джерела: Психометричні тестові компанії, шкільні консультанти та самостійне онлайн-тестування.

– Мета: Психологічне тестування допомагає визначити такі риси, як емпатія, стресостійкість, здатність приймати рішення та навички міжособистісного спілкування, які мають вирішальне значення для успіху в різних медичних ролях.

с) Інтереси та здібності.

– Типи даних: Інвентаризація інтересів, позакласні заходи, хобі, волонтерська робота та стажування.

– Джерела: Опитування учнів, записи про позакласну діяльність та особисті інтерв'ю.

– Мета: Ці дані дають уявлення про те, якими галузями медицини

захоплюються студенти і де вони досягають найбільших успіхів, допомагаючи узгодити вибір кар'єри з особистими інтересами та сильними сторонами.

Попередня обробка даних.

а) Обробка відсутніх значень.

– Ідентифікація: Визначення відсутніх значень у наборі даних, які можуть виникнути через неповний збір даних або помилки у звітності.

Технології.

– Імпутація: Заміна відсутніх значень середнім значенням, медіаною для неперервних змінних або найчастішим значенням для категоріальних змінних.

– Прогностичне моделювання: Використовуйте алгоритми, такі як k -найближчих сусідів (k -NN) або регресійні моделі, для прогнозування та заповнення відсутніх значень на основі інших наявних даних.

– Видалення: У випадках, коли дані відсутні повністю випадково і становлять невеликий відсоток набору даних, відповідні записи можна видалити.

б) Нормалізація ознак.

– Мета: Забезпечити, щоб всі ознаки однаково впливали на продуктивність моделі, запобігаючи непропорційному впливу ознак з великими числовими діапазонами на результати.

Методика.

– Мінімальне-максимальне масштабування: Зміна масштабу значень ознаки до фіксованого діапазону, зазвичай від 0 до 1.

– Стандартизація: Перетворення значень ознак до нульового середнього значення та одиничної дисперсії, що особливо корисно, коли ознаки мають гаусівський розподіл.

– Надійне масштабування: Масштабування ознак відповідно до статистичних даних, стійких до викидів, наприклад, з використанням медіани та інтерквартильного діапазону.

с) Кодування категоріальних змінних.

– Мета: Перетворити категоріальні дані в числовий формат, який можна обробити за допомогою алгоритму SVM.

Методика.

– Одноголосне кодування: Створює двійкові стовпчики для кожної категорії, де наявність категорії позначається 1, а її відсутність - 0. Це ефективно для номінальних категоріальних змінних, які не мають порядкового зв'язку.

– Кодування міток: Присвоїти кожній категорії унікальне ціле число. Цей метод більше підходить для порядкових категоріальних

змінних, де порядок категорій має значення.

– Двійкове кодування: Перетворення категорій у двійкові числа, а потім розбиття цифр на окремі стовпчики, що допомагає зменшити розмірність порівняно з однозначним кодуванням.

d) Інженерія функцій.

– Створення нових функцій: Розробка нових функцій, які можуть підвищити продуктивність моделі, об'єднавши існуючі або витягнувши більше інформації.

– Доменна експертиза: Використання знання з освітніх та психологічних досліджень для створення функцій, які відображають відповідні закономірності та взаємозв'язки.

Ефективний збір і попередня обробка даних мають вирішальне значення для використання алгоритму SVM у прогнозуванні відповідних медичних професій для студентів. Ретельно збираючи широкий спектр релевантних даних і застосовуючи надійні методи попередньої обробки, ми можемо гарантувати, що вхідні дані для SVM моделі є вичерпними і чистими. Цей фундаментальний крок створює основу для розробки точної, надійної та справедливої класифікаційної моделі, яка може забезпечити персоналізовану профорієнтацію для майбутніх медичних працівників.

Вибір функцій.

Виявлення найбільш релевантних ознак, які впливають на вибір кар'єри, має вирішальне значення. Такі методи, як аналіз головних компонент (PCA) та рекурсивне виключення ознак (RFE), можуть бути використані для зменшення розмірності та покращення продуктивності моделі.

Моделна підготовка.

– Потім оброблені дані розбиваються на навчальну та тестову вибірки. Алгоритм SVM тренується на навчальному наборі, навчаючись виявляти закономірності та взаємозв'язки в даних.

– Можна експериментувати з різними функціями ядра (лінійними, поліноміальними, радіальними базисними функціями), щоб визначити, яка з них найкраще підходить для даних.

Оцінка моделі.

– Навчена SVM-модель оцінюється за допомогою тестового набору для вимірювання її точності, достовірності, запам'ятовування та оцінки F1.

Методи перехресної перевірки можуть додатково забезпечити надійність моделі.

Прогнозування та інтерпретація.

– Після валідації SVM-модель може прогнозувати найбільш підходящу медичну професію для нових точок даних (студентів).

Інтерпретованість моделі може бути покращена за допомогою таких методів, як SHAP (SHapley Additive exPlanations) для пояснення прогнозів.

Переваги використання SVM у виборі медичної кар'єри.

1. Персоналізоване керівництво:

– Модель SVM забезпечує персоналізовану кар'єрну орієнтацію, засновану на комплексному аналізі індивідуальних характеристик, що призводить до більш усвідомленого та відповідного вибору професії.

2. Ефективність і точність.

– Здатність SVM обробляти багатовимірні дані та його надійність у завданнях класифікації забезпечують високу точність у прогнозуванні найкращої медичної професії.

Прийняття об'єктивних рішень:

– Підхід алгоритму, заснований на даних, виключає упередженість і суб'єктивізм, пропонуючи об'єктивні рекомендації, засновані на емпіричних даних.

3. Масштабованість.

– SVM-моделі можна легко масштабувати та оновлювати за допомогою нових даних, що дозволяє постійно вдосконалюватись та адаптуватись до мінливих тенденцій у медичній галузі.

Виклики та майбутні напрямки.

Хоча SVM має численні переваги, необхідно вирішити певні проблеми:

1. Якість та доступність даних.

– Забезпечення якісних і всебічних даних має вирішальне значення для успіху моделі. Необхідно докласти зусиль для збору різноманітних і репрезентативних даних.

2. Інтерпретованість моделі.

– Покращення інтерпретованості SVM-моделей має важливе значення для завоювання довіри з боку користувачів (студентів та консультантів з питань кар'єри). Необхідні методи, які дають уявлення про процес прийняття рішень в моделі.

3. Інтеграція з існуючими системами.

– Безперешкодна інтеграція систем профорієнтації на основі SVM з існуючими освітніми та консультаційними системами є життєво важливою для практичної реалізації.

Майбутні дослідження можуть вивчати гібридні моделі, що поєднують SVM з іншими алгоритмами машинного навчання для подальшого підвищення точності та надійності прогнозування. Включення механізмів зворотного зв'язку для вдосконалення моделі на основі досвіду користувачів може підвищити її ефективність.

ВИСНОВОК

Алгоритм машина опорних векторів (SVM) можна використовувати для класифікації людей за різними медичними професіями на основі їхніх психофізичних особливостей. Він працює шляхом створення гіперплощини, яка розділяє дані на різні класи.

Застосування алгоритмів SVM в інформаційно-вимірjuвальних технологіях має значні перспективи для покращення обґрунтованості вибору медичної професії. Використовуючи сильні сторони SVM в обробці складних і багатовимірних даних, ми можемо за-пропонувати персоналізовану, точну і об'єктивну профорієнтацію майбутнім медикам. Оскільки ми продовжуємо вдосконалювати ці моделі і вирішувати пов'язані з ними проблеми, потенціал SVM в цій галузі буде тільки зростати, що в кінцевому підсумку сприятиме більш усвідомленому і повноцінному вибору кар'єри в медичній сфері.

ПОСИЛАННЯ

1. Watts, A. G. (2002). The role of information and communication technologies in integrated career information and guidance systems: A policy perspective. *International Journal for Educational and Vocational Guidance*, 2(3), 139-155.
2. VidyaShreeram, N., & Muthukumaravel, A. (2021, June). Student career prediction using machine learning approaches. In *First International Conference on Computing, Communication and Control System* (p. 444).
3. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.