

Ми і є ШІ: Беремо технології під контроль

Ми і є ШІ № 2: Навчаючись на даних

Обкладинка:

Утілений ШІ-робот стає дедалі більшим на екрані. Унизу — сила силенна «даних», на яких ШІ тримає долоню і з яких черпає силу

мови використання

Усі ілюстрації в цьому коміксі доступні за ліцензією CC BY-NC-ND 4.0. Будь ласка, перейдіть на сторінку ліцензії, щоб дізнатися більше про те, як можете використовувати ці роботи.

Не соромтеся використовувати панелі/групи панелей у презентаціях/статтях, якщо

1. належно цитуєте їх;
2. не вносите змін в окремі панелі.

Цитувати за:

Джулія Стоянович та Фала Аріф Хан. «Навчаючись на даних?». Ми і є ШІ Комікси, том 2 (2021)

https://dataresponsibly.github.io/we-are-ai/comics/vol2_en.pdf

Сторінка 1

Говорячи про штучний інтелект (ШІ), ми зазвичай розрізняємо системи «класичного ШІ» — ті, які працюють за заздалегідь визначеними правилами, як-от Roomba — розумний пилосмок, круглий, металевий робот, що обертається навколо, і ті, які «вивчають» правила з даних, — наприклад, штучний інтелект, що грає в шахи.

На екрані чотири різні шахові фігури (зліва направо): пішак, король, ферзь і кінь або безпілотний автомобіль.

Уявіть жінку в безпілотному автомобілі: вона на водійському сидінні, але відкинулася назад, закинувши руки за голову й поставивши ноги на приладову панель. Автомобіль перебуває в режимі автономного керування, кермо підсвічується, а на світлодіодному екрані на приладовій панелі з'являється робот-асистент.

Ця відмінність не строга.

Насправді більшість сучасних систем штучного інтелекту поєднують у собі правила, створені вручну, з певною формою навчання.

Розгляньмо приклади таких правил і подивімося, як машини вивчають їх на основі даних!

Сторінка 2

Уявіть, що маєте спроектувати розумну систему освітлення, яка автономно вмикає та вимикає світло у вашій оселі.

З'являється ШІ-джин, що зачаровує лампочку, містично контролюючи її світіння та згасання.

Як ваш ШІ має вирішувати, коли так діяти?

Ми можемо почати проектувати таку систему, спираючись на власний повсякденний «інтелект».

Назвімо це Правилем 1. Воно просте:

Вмикай світло, коли надворі темно, Уявіть собі простий карикатурний будиночок із дверима та одним вікном. Надворі темно: місяця немає, небо фіолетове, а всередині будинку світиться — у вікні видніється жовте світло, що розливається кімнатою.

В іншому разі тримайте вимикач вимкненим.

У будинку з дверима тепер темно — у вікні видно темно-фіолетовий відтінок, що осідає над будинком, тоді як надворі сонце зійшло, а небо яскраво-синє.

Цей алгоритм дуже простий, він має лише один крок — приймає зовнішні умови (чи темно надворі) як вхідні дані, і, використовуючи Правило 1, прогнозує прийнятний результат — чи вмикати світло.

Просте символічне зображення правила 1: сонце і місяць ліворуч, стрілка посередині, яскрава (жовта) лампочка і/або тьмяна (сіра) лампочка праворуч

Це правило просте, але чи добре воно працює?

Чи правильно воно відображає те, як вручну керуєте світлом у вашому домі?

Виконаймо експеримент, щоб з'ясувати це!

Сторінка 3

Ми збиратимемо дані про те, чи темно надворі й чи ввімкнене світло в будинку.

Для кожного спостереження перевіримо, чи збігається прогноз за нашим алгоритмом із тим, що спостерігали.

Сторінка поділена на чотири панелі: a, b, c і d.

Панель a: Пізній вечір, жінка лежить на дивані й між іншим перемикає канали. Надворі темно, але кімната добре освітлена.

Панель б: Дві жінки за обіднім столом підіймають склянки на честь свята: перед ними — тарілки з вишуканими частуваннями. Надворі темно, але стіл добре освітлений підвісною лампою.

Панель в: Жінка сидить за столом і читає, тримаючи ранкову каву в одній руці, а iPad — у другій. Позаду вимкнена лампа. З іншого боку позаду відчинене вікно, і приємне світанкове сонце освітлює кімнату.

Панель d: Загорнувшись у ковдру, жінка міцно спить у своєму ліжку. У кімнаті темно, світло вимкнене, тому бачимо її у відображенні фіолетових і синіх відтінків.

У спостереженнях а і b надворі темно, але ми дивимося телевизор і вечеряємо, тому залишили світло.

І це справді те, що передбачало Правило 1!

Це означає, що а і b підтверджують гіпотезу, закодовану у Правилі 1.

Спостереження с теж підтверджує гіпотезу:

надворі любо і ясно, тож ми не вимикали світло.

Тепер погляньмо на заувагу d:

надворі темно, ми вкладаємося спати, світло вимкнули.

Однак, згідно з правилом 1, алгоритм передбачив, що світло буде «увімкнене». Правило 1 тут не працює!

Сторінка 4

Отже, що ми дізналися про Правило 1 з нашого експерименту — Наше правило змогло правильно передбачити результат 3 рази з 4.

Темрява — вимкнено

Світло — увімкнено

Прогнозування

Ця таблиця підсумовує спостереження з експерименту:

Для кожного спостереження є три стовпчики: «Темно» [місяць/сонце], «Світло» [жовта лампочка/сіра лампочка] і «прогноз» [жовта лампочка/сіра лампочка]. Таблиця має такий вигляд: А — темно, світло увімкнене, світло увімкнене; В — темно, світло увімкнене, світло

увімкнене; C — світло, світло вимкнене, світло вимкнене; D — темно, світло вимкнене, світло увімкнене.

Точність передбачення цього правила — 75 %.

Наш ШІ-джин хизується перед камерою, але його біцепс «заповнений» кольором лише на 3 частини з 4. На зігнутому біцепсі написано «75 %».

Підправмо тут ще трохи термінології

Зазвичай вхідні дані правила — у нашому прикладі п'ятьма — мають назву вхідна ознака, або просто ознака.

Коли ззовні темно, ми зображаємо це місяцем, а коли ясно — малюємо сонце.

Вихідні дані — стан наших світильників (увімкнено/вимкнено) — зазвичай називають результатом, або міткою.

Різні можливі результати — вимкнути світло чи залишити його ввімкненим, як у нашому прикладі, — це класи.

Якщо світло ввімкнене, лампочка має жовтий колір, а якщо вимкнене — сірий.

Правило, з яким працюємо, називають «класифікатор» — воно надає спостереженню мітку класу.

Класифікатор, який ми створили за допомогою Правила 1, може зробити один із двох виборів — увімкнути або вимкнути світло.

Ми називаємо такі класифікатори «бінарними».

Сторінка 5

Повернімося до спостереження d, яке хибно класифікувало правило 1:

Нагадаємо приклад d: жінка міцно спить у своєму ліжку, загорнувшись у ковдру. У кімнаті темно, світло вимкнене, тому бачимо її у віддзеркалених відтінках фіолетового та синього.

Правило прийняло б на вхід — «надворі темно», а на вихід повернуло б «світло увімкнене», тим самим схибивши: увімкнуло б світло й розбудило б жінку.

Яка ціна цієї помилки?

Жінка раптово прокидається від сліпучого світла — зазомбовані очі й скуйовджене волосся відображають розлючений вираз її обличчя й розкинуті руки.

Як ви, напевно, здогадалися, розумне світло розбудило вас посеред ночі!

Як можемо зарадити?

Якщо порівняємо спостереження a , b та d , то побачимо, що вони мають однакове значення вхідної ознаки (на вулиці темно), але різні вихідні.

Темно — вимкнено

Світло — увімкнено

Прогнозування

Це наштовхує на думку, що потрібні додаткові ознаки, щоб розрізняти такі ситуації.

Тепер рішення, яке ви, як розробник цієї системи, маєте прийняти, стосується того, які ще вхідні дані можуть бути корисні

Занурена в роздуми жінка схрестила руки й підперла кулаком підборіддя — над нею з'являється мисленнєва бульбашка із зображенням будильника, чашки та блюдця з чаєм, а також термометра.

Які ознаки, на вашу думку, допоможуть передбачити результат?

Температура на вулиці? Ні!

Ціна чаю в Китаї? Ні!

Гаразд, а як щодо того, чи час спати?

Почнімо вдосконалювати наше правило.

Сторінка 6

Ось наше вдосконалене правило — назвімо його Правило 2.

«Вмикай світло, якщо надворі темно, але ще не час спати.

В іншому разі не вмикай світла».

Цього разу карикатурне правило має два вхідні параметри: символи сонця або місяця, які вказують на те, світло чи темно надворі, і будильник, який вказує на те, час спати чи ні. Посередині є стрілка, а вихідні дані праворуч такі самі — жовта лампочка (світло увімкнене) та сіра лампочка (світло вимкнене)

Повторімо експеримент, зібравши ще кілька спостережень та оцінивши наше нове правило.

На сторінці чотири різні знімки: e, f, g і h

Кадр e: Жінка сидить, тримаючи горнятко ранкової кави в одній руці, а iPad — у другій. Позаду вимкнена лампа. З іншого боку позаду відчинене вікно, тож лагідне досвітнє проміння освітлює кімнату.

Кадр f: Жінка лежить на дивані, закинувши ноги на підлокітник. Вона говорить по телефону, жваво жестикулюючи. У вікні видніється темне небо, а кімната добре освітлена світлом згори.

Кадр g: Жінка сидить за столом перед ноутбуком. Вона пильно вдивляється в екран, тримаючи в одній руці чашку чаю/кави. Настільна лампа стоїть поряд, але вимкнена, адже у вікні позаду видно яскраво-блакитне небо й денного світла достатньо.

Кадр h: Жінка, згорнувшись калачиком, спить на дивані: вона лежить, поклавши ноги на підлокітник, і, здається, задрімала перед ноутбуком, який прямісінько перед нею. Настільна лампа ввімкнена й розливає світло по кімнаті. З'являється ШІ-джин і, зависаючи між світлом та заснулою жінкою, за допомогою магії вимикає лампу.

Вуаля! Здається, усе працює ідеально!

Нам не потрібне ввімкнене світло, хіба що за вікном темно й ми не спимо, а щось робимо!

І подивіться, цього разу він вимикає світло, якщо засинаємо на дивані (після того, як нам час спати) — так дбайливо!

Сторінка 7

Правило 2 продемонструвало бездоганну точність — наш алгоритм здатен правильно передбачити результат у 100 % випадків!

Темно ззовні

Час спати

Світло ввімкнене

Передбачення

З'явилася нова таблиця, яка підсумовує спостереження експерименту:

Для кожного спостереження є 4 стовпчики: «Темрява» [місяць/сонце], «Час спати», «Світло» [жовта лампочка/сіра лампочка] і «Прогноз» [жовта лампочка/сіра лампочка]. Таблиця така: Е — світло, не час спати, світло вимкнене, світло вимкнене; F — темно, не час спати, світло увімкнене, світло увімкнене; G — світло, не час спати, світло вимкнене, світло вимкнене; H — темно, час спати, світло вимкнене, світло вимкнене.

Наш ШІ-джин знову хизується перед камерою, цього разу його біцепс повністю «залитий» кольором. Поперек зігнутого біцепса написано «100 %».

Ми використали власні інтуїцію й досвід, щоб розробити належне правило, а також зібрали дані, прагнучи перевірити, як воно спрацьовує.

і вносили потрібні корективи, аж доки досягли результату, який нам сподобався!

Насправді ми так задоволені нашим смартсвітлом — тим, як це зручно і як допомагає нам заощаджувати енергію

Тож хочемо запропонувати іншим теж покористуватися.

Ми відрекомендували нашого ШІ-джина іншим людям! ШІ-джин тисне руку жінці. Наш головний герой обіймає і ШІ-джина, і жінку, знайомлячи їх.

Подивимось, чи спрацює це для нашої подруги Анни в її офісі.

Однак перед тим, як візьмемося переробляти електрику в усій офісній будівлі Анни, зберімо деякі типові приклади використання, щоб зрозуміти, чи буде наше правило ефективне.

Ви знаєте правила: гайда експериментувати!

Сторінка 8

Пригадайте правило 2 — «Увімкніть світло, якщо надворі темно, але ще не час спати. В іншому разі не вмикайте світла»

Сторінка поділена на 4 кадри: I, J, K і L

Панель I: Розмаїта група людей сидить за столом і дивиться презентацію на ноутбучі. На столі — інші гаджети й кілька чашок чаю/кави. Люди працюють у відкритому, сучасному робочому просторі, з великим вікном позаду, що доповнює світло всередині офісу.

Панель J: Збільшене зображення офісної будівлі ззовні — темно, тому бачимо два поверхи офісів у фіолетово-рожевих відтінках. Офіси порожні, на кожному поверсі стоять кілька порожніх стільців і столів.

Панель K: Вид на офіс зсередини — здається, що зараз полудень, адже небо ясне, а офіс освітлений природним світлом. Офіс порожній, тому бачимо лише великі й довгі офісні столи для нарад і багацько порожніх стільців, але не людей!

Панель L: Жінка сидить над своїм завданням і несамовито друкує на ноуті. Настільна лампа не ввімкнена, як і інші лампи в офісі. Жінка виснажена, схоже, вона працює понаднормово, адже вікно за її спиною відкриває темне нічне небо.

В офісі ми тримаємо світло ввімкненим упродовж дня, коли люди працюють.

Однак Правило 2 некоректно спрацює, адже надворі світло.

Переважно вечорами, коли всі розходяться по домівках, світло вимикається.

Однак Правило 2 некоректно ввімкне його, щойно на вулиці потемніє!

На вихідних офіс пустує, тому світло вимкнене протягом усього дня!

У цьому разі Правило 2 працює правильно: вимикає світло, адже надворі ясно.

Час від часу хтось затримується в офісі допізна, щоб встигнути завершити роботу до критично важливого речення.

У цьому сценарії нам потрібно увімкнути світло, але Правило 2 знову хибне: воно вимкне світло, адже вже пізно (після того, як зазвичай лягаємо спати).

Сторінка 9

О, ні! Це означає, що ми отримали лише 1 правильний прогноз з 4!

Якщо припустити, що світло ввімкнене приблизно половину часу, то було б краще, якби ми просто підкинули монетку, щоб вирішити, чи вмикати світло!

Чому точність Правила 2 така низька?

Це пов'язано з тим, що набір даних, на якому тестували класифікатор, стосується інакшої ситуації, ніж набір даних, на якому його навчали.

Ми «натренували» класифікатор — визначили відповідні правила — на основі наших споживацьких звичок удома Знімок жінки з-за плеча в оселі: жінка сидить із чашкою кави, схрестивши ноги на письмовому столі перед собою. Яскраве небо освітлює кімнату, тому настільна лампа вимкнена.

Але Анна «потестувала» його у своєму офісі.

Світлина Анни, яка працює на ноутбуці в офісі з увімкненою настільною лампою, попри пізній час і темряву за вікном.

Пора для сну чи темрява за вікном уже не суттєві!

Важливо, чи відкритий офіс!

Сторінка 10

Яка тут ціна помилки?

Ціна хибно позитивного — невиправдане ввімкнення світла, коли воно має бути вимкнене, — змарнована електроенергія і кругленький рахунок компанії за неї.

Анна тримає в руках довжелезний рахунок за електроенергію для компанії й вигукує, дивлячись на суму. Жінка схопилася рукою за голову й здається засмученою.

Ціна хибно негативного — вимкнення світла, коли воно потрібне, — у перериванні людей посеред роботи, а отже, у їхній нижчій продуктивності.

Знімок поточної зустрічі в офісі: світло раптово вимкнули, і ми можемо бачити лише фіолетово-сині відтінки людей у кімнаті. Вони вигукують щось через раптове вимкнення світла, а на їхніх обличчях відверта зневага, бо обговоренню завадили.

Що нам робити? Знову повернутися на початок?

Подумайте, які додаткові ознаки потрібні, зберіть ці дані

і виконайте експеримент, щоб перевірити, чи працює наше правило насправді?

Але це так нудно!

Головна героїня занурена в міркування: вона тримається рукою за обличчя й задумливо дивиться вгору. У віддзеркаленні на її райдужній оболонці бачимо голівки роботів, коли вона розмірковує про використання ШІ.

І найголовніше — нам буде дуже складно далі формулювати ці правила, які стають щораз складнішими

залежно від місця, де використовують смартсвітло, та їхніх різних вимог!

Рішення таке: використайте алгоритм машинного навчання, щоб визначити статистичні закономірності в наших даних — і автоматично “навчимо” правил

Однак так само, як ми розробляли експерименти, щоб перевірити, чи застосовні вигадані правила на практиці, варто розробити експерименти, щоб перевірити, чи працюють автоматично “навчені” правила в разі впровадження.

Бо правило повністю залежить від історичних даних — якщо ці дані репрезентативні для майбутнього використання системи смартосвітлення,

тоді правило працюватиме добре.

Однак якщо фактичний випадок використання відрізняється — правило призведе до багатьох помилок Уявіть, що ШІ-джин тримає апельсин, але думає, що це яблуко (як показано в його мисленнєвій бульбашці), адже він бачив тільки яблука.

Наприклад, увімкнути світло в порожніх офісних будівлях, ШІ-джин з’явився в порожньому офісі. Там немає людей, але ШІ-джин за допомогою магії вмикає всі лампочки в офісі або пробуджує малюка серед денного сну.

ШІ-джин перемістився в будинок і магічно керує всім світлом: він боязко дивиться на камеру, а сліпуче вичаклуване світло змушує сонного малюка підхопитися і нестримно репетувати.

Незалежно від того, правила записані людьми чи вивчені на основі історичних даних машинами, ми повинні обов’язково використовувати науковий метод — формулювання та перевіряння спростовної гіпотези, яку можна фальсифікувати.

Як казав філософ науки Карл Поппер:

«Теорія або ідея не має бути науковою, якщо її не можна довести як хибну».

Карикатурний Карл Поппер підкидає монету в повітрі.

Сторінка 12

Науковий метод починається зі спостереження.

Наша героїня вбрана як Шерлок Голмс: вона тримає лупу й дивиться прямо в камеру через неї. Це спотворює її око й робить його дуже великим для глядача.

За цим іде формулювання спростовної гіпотези, яку можна сфальсифікувати — яку можна довести як хибну.

Головна героїня сидючи споглядає щось — обидві руки схрещені перед губами, а вираз обличчя суворий.

Далі потрібно розробити експеримент, щоб перевірити, чи можна спростувати гіпотезу.

Головна героїня тримає дві мензурки над великою каструлею й у передчутті несамовито виливає їхній вміст у вариво.

І якщо так, ми уточнюємо гіпотезу, розробляємо новий експеримент, «прополіскуємо» його й повторюємо.

Головна героїня переживає момент «Еврики»: одна її рука на вустах, а друга вказує вгору.

Сторінка 13

Ключове питання: «Чи працює класифікатор?» формалізоване як «Чи передбачення класифікатора точніші, ніж випадкове вгадування?»

Це найнижча планка для точності — ми не хочемо зрештою створити дурненький ШІ, який підкидатиме монети.

ШІ-джин чаклує над лампочкою й контролює її потужність. На одному знімку за ним зображена людина, яка працює в темряві й потребує світла, а на іншому — спляча жінка й порожній офіс, і в обох випадках світло потрібно вимкнути.

Коли маємо достовірну інформацію (спостереження за тим, чи справді має бути світло) — можемо перевірити.

А якщо гіпотеза буде спростована, то геть її.

Ми розробляємо нову процедуру генерування даних і перенавчаємо модель, щоб вона вивчила нове правило.

Важливо, що навіть з усією додатковою складністю ми однаково навряд чи побачимо ідеальну точність.

Одна з причин цього — невизначеність у світі.

Фото жвавого офісу: люди працюють на різних робочих місцях у великому відкритому просторі. Навколо кілька знаків питання, що вказують на сумніви, чи потрібне їм світло, чи воліють вони працювати при тьмяному світлі та/або чи скоро вони підуть, а чи гаруватимуть допізна.

А ще те, що правила іноді порушують: люди можуть виходити на роботу у вихідні або затримуватися на роботі довше, щоб вкластися в реченець.

Отже, у реальних життєвих ситуаціях можемо сподіватися лише на те, що класифікатор працюватиме більшість часу.

Але інколи хтось матиме встати й увімкнути або вимкнути світло.

Головна героїня підходить до вимикача й збирається на нього натиснути. Вона повертається до камери/глядача й навмисно підморгує.