

УДК 004.738.5

DOI 10.31891/2307-5732-2020-285-3-13

С. П. ЗЕМБИЦЬКИЙ, Н. В. ГРИПИНСЬКА

Хмельницький національний університет

ПРОЕКТУВАННЯ АРХІТЕКТУРИ ХМАРНОГО ОБЧИСЛЮВАЛЬНОГО СЕРЕДОВИЩА ДЛЯ ОБРОБКИ АНАМНЕЗУ ПАЦІЄНТІВ ІЗ ЗАХВОРЮВАННЯМ ДИХАЛЬНОЇ СИСТЕМИ НА ОСНОВІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

У статті представлено проектування архітектури хмарного середовища для обробки анамнезу захворювання дихальної системи та подальшого прогнозування діагнозу. Під час проведення наукової роботи було опрацьовано попередні публікації та наукові дослідження в галузі сімейної медицини. Проаналізувавши попередні дослідження було сформовано основне невирішене питання, яке полягає в розробці архітектури обробки епізоду звернення пацієнта з використанням методів машинного навчання та з подальшим розгортанням на серверах - хостерах з використанням хмарної обчислювальної технології. Також було сформовано основні цілі проведення даного наукового дослідження. Розглянуто основні характеристики для побудови архітектури та обрано їх оптимальні значення. Обґрунтовано доцільність використання технології передачі даних AJAX. Також було обґрунтовано доцільність використання баз знань формату csv та структурованого об'єкту JSON. Розглянуто алгоритми класифікації з машинним навчанням та обрано оптимальний алгоритм класифікації. Результатом дослідження є розроблена архітектура, в вигляді блок – схеми, функціонування хмарного обчислювального середовища.

Ключові слова: алгоритми класифікації, машинне навчання, обробка анамнезу

S. ZEMBITSKYI, N. HRYPYNska

Khmelnytskyi National University

CREATING OF ARCHITECTURE OF CLOUD COMPUTING ENVIRONMENT FOR ANALYZING ANAMNESIS OF PATIENTS WITH RESPIRATORY SYSTEM DISEASE ON BASE MACHINE LEARNING

Algorithms of machine learning are widely used in small galuzy, did not omit medical practice. The machine of modernization in the medical field has known the storage in different directions, such as the processing of biomedical images, the analysis of computer tomography, the analysis of electrocardiograms, the development of licenses, the pathology, the detection. Algorithms of machine learning are also used for the administration of the plant, as it is not necessary for the practical practice, but for the medical galusa. For example, private clientele vikoristovuyut systems and piece neural framing for the establishment and processing of routing of patients in the clinic, keruvannya chergami in the clinic, the establishment of interactive knowledge bases for detailed information to the medical staff. But it's a pity in the family of medicine, for the present moment, it is not possible to see in the implementation of algorithms of machine technology. In the out-of-town medical practice of testing systems of piece neural fences, to optimize the working process of the family doctor, to change the vitrate for an hour at the time of the patient's reception, and to set up the backward pattern of anamnesis of the patient's anamnesis. Therefore, the main task of this scientific advancement is to develop the architecture of a gloomy numerical center for reviewing the anamnesis of patients with mental illness on the basis of machine learning algorithms, so that we can give the values of the robotic field of recognition.

This work present process of developing architecture cloud processing environment for analyzing anamnesis of patients with respiratory system disease. During doing this scientific work was processed previous publications and scientific experiments in family medicine direction. Having analyzed previous experiments was created main problem this scientific work. Main problem is creating of architecture processing episode patient treatment with using machine learning and future integration on servers – hosters with cloud processing technologies. Also was created main targets this scientific work. Reviewed main features for creating of architecture and was selected them optimal correct values. Justified expediency using technologie of transmit data such as AJAX. Also was justified expediency using knowledge database that using csv format and using structured object such as JSON. Reviewed machine learning algorithms of classifications and was selected optimal algorithm. Result this experiment is created architecture in view block – schema, function cloud processing environment.

Keywords: algorithm classification, anamnesis processing, machine learning

Постановка задачі. Алгоритми машинного навчання широко використовуються в різних галузях, не минуло це і медичну практику. Машинне навчання в медичній галузі знайшло застосування в різних напрямках, таких як обробка біомедичних зображень, аналіз комп'ютерної томографії, аналіз електрокардіограми, розробка ліків, радіологія, виявлення патологій. Також алгоритми машинного навчання застосовуються для виконання завдань, які не стосуються лікувальної практики, але відносяться до медичної галузі. Наприклад, приватні клініки використовують системи штучних нейронних мереж для створення та обробки маршрутизації пацієнтів в клініці, керування чергами в клініці, створення інтерактивних баз знань для вдосконалення досвіду медичного персоналу. Але на жаль в сімейній медицині на даний момент не відбувається впровадження алгоритмів машинного навчання. В загальній медичній практиці використання систем штучних нейронних мереж створить оптимізацію робочого процесу сімейного лікаря, зменшить витрату часу на прийом пацієнта та обробку його анамнезу захворювання, встановить загальний шаблон алгоритму та порядок дій проведення прийому пацієнта та обробки анамнезу. Тому основною задачею даного наукового дослідження являється розробка архітектури хмарного обчислювального середовища для обробки анамнезу пацієнтів з захворюванням дихальної системи на основі алгоритмів машинного навчання, що надасть значні полегшення в роботі сімейного лікаря.

Аналіз досліджень та публікацій. На даний момент наукова спільнота проводить дослідження та розробки в сфері діагностики респіраторних захворювань за допомогою алгоритмів машинного навчання. Одне із таких досліджень викладене в дисертації Поревої Ганни Сергіївни «Методи аналізу звуків легень

для оцінки стану дихальної системи людини». Дана дисертаційна робота присвячена актуальній темі обробки та аналізу звуків легень людини з метою отримання діагностично цінних параметрів для використання їх як в якості самостійних критеріїв оцінювання стану дихальної системи людини і як вхідних аргументів класифікаторів для автоматизації прийняття рішень щодо певних захворювань [1]. Проводиться розробка нових структур баз знань та модифікації алгоритмів машинного навчання [2]. Також проводиться розробка в області аналізу біомедичних зображень, а саме аналіз флюорографічних знімків легень та виявлення аномалій в знімках.

Виділення невирішених частин. Проаналізувавши існуючі наукові дослідження та публікації, можна зробити висновок, що основною невирішеною проблемою є відсутність реалізації алгоритмів машинного навчання для аналізу анамнезу пацієнта, розгорнутих в хмарному обчислювальному середовищі. Саме використання систем машинного навчання в хмарних технологіях надасть можливість кінцевому користувачу використовувати менш ресурсоємні пристрої за рахунок проведення всіх обчислень на стороні сервера та відправки опрацьованого пулу даних, клієнту. Також середовище хмарного обчислення допоможе в проектуванні кросплатформового застосунку, що збільшить кількість потенційних користувачів, за рахунок виконання додатка на web платформі, та зменшить витрати ресурсів на розробку кінцевого додатка.

Формулювання цілей. В ході проведення даного наукового дослідження було сформульовано наступні цілі:

- ознайомитися з типами хостингових серверів та визначити оптимальний тип та конфігурацію сервера;
- визначити тип web застосунку;
- визначити метод «спілкування» клієнта з сервером;
- визначити тип баз даних для збереження карт навчання та іншої інформації, яка забезпечує функціонування середовища;
- визначити алгоритм машинного навчання для обробки та прогнозування або класифікації діагнозу відповідно до відібраного анамнезу;
- розробити архітектуру функціонування хмарного обчислювального середовища.

Виклад основного матеріалу. Для реалізації хмарної архітектури першим основним кроком було визначення типу та конфігурації сервера хостингу. Існують наступні типи web хостингів:

- віртуальний хостинг (Shared hosting) розширений тип хостингу, який задовольняє більшість користувачів та власників невеликих сайтів. Плюсами даного типу сервера є простота в використанні та наявність універсальної панелі керування. Недоліками є відсутність доступу до сервера по SSH протоколу для встановлення розширень та налаштування сервера, невеликий об'єм виділеного місця, кількісне обмеження створення таблиць та баз даних, присутність «сусідів»;
- віртуальний виділений сервер (Virtual dedicated server) виділена частина сервера, яка містить для кожного клієнта окрему віртуальну машину. Використовується для середніх проєктів. Плюсами даного типу сервера є можливість встановлення потрібної операційної системи та програмного забезпечення для функціонування сервісу, конфігурування обчислювальних ресурсів та наявність доступу до сервера по SSH протоколу. Основним недоліком є присутність «сусідів»;
- виділений сервер (Dedicated server) окремий обчислювальний пристрій з окремим виділеним каналом зв'язку. Використовується для великих проєктів. Наявний повний контроль над програмною частиною сервера з повним доступом до всіх налаштувань сервера. Апаратну частину адмініструє хостинг-провайдер;
- колокейшн (Colocation) надання фізичного місця та виділеного каналу зв'язку в дата центрі хостинг – провайдера;
- хмарний сервер (Cloud hosting) віртуальний виділений сервер з гнучкою системою розподілення обчислювального ресурсу для виконання задач.

Таким чином для реалізації архітектури хмарного обчислення з використанням алгоритмів машинного навчання оптимальним варіантом сервера є віртуальний виділений сервер, виділений сервер або хмарний сервер.

Тип web застосунку – SPA (Single Page Application) односторінковий web додаток. Порівняно з звичайним багатосторінковим web додатком SPA оптимізований для зменшення використання інтернет трафіку за рахунок разового завантаження та кешування файлів ресурсів на пристрій кінцевого користувача, таким чином відбувається обмін лише інформацією для обробки між клієнтом та сервером в вигляді пулу даних. Також використовуючи дану технологію можна використовувати методику шаблонізації, що значно зменшить об'єм кінцевого web додатку.

Спілкування з сервером відбувається за методикою AJAX, що дозволить кінцевому користувачу працювати без примусового перезавантаження web сторінки при відправці та отриманні пулу даних. Також використання даної методики роботи з сервером зменшує використання інтернет трафіку та зменшує навантаження на сервер.

База знань організована у вигляді структурованих векторів клініки респіраторних захворювань та анамнезу пацієнтів. Об'єкт бази знань представлений в вигляді файлу з типом розширення “.csv”. Порівняно з звичайною базою даних MySQL організація структури бази знань в файлі з типом розширення “.csv” дає змогу системі алгоритму машинного навчання з мінімальною затримкою на вибірку інформації отримати повний зліпок бази знань для чергового перенавчання, побудови нової моделі навчання та створення

класифікації або прогнозу відповідно до нової моделі. Також використовуючи об'єкт файлу з типом розширення ".csv" для організації бази знань з легкістю можна вносити зміни в структуру самої бази знань не руйнуючи вже існуючу базу знань. Плани лікування та шифри ICPC2, МКХ – 10 потребують організації складної структури даних, тому для реалізації даної потреби було обрано формат JSON. Використовуючи даний формат можна сформувати складну модель збереження та представлення даних та інтерпретувати дані в об'єкти, що допоможе швидко та легко отримати доступ до об'єкту даних та отримати значення обраного об'єкту.

Для реалізації задачі класифікації було обрано наступні алгоритми машинного навчання: алгоритм наївного Баєса (далі АНБ), метод опорних векторів, метод К-ближніх сусідів. Алгоритм наївного Баєса – це ймовірнісний алгоритм класифікації в основі якого лежить теорема Баєса про наївні припущення незалежності змінних. Цей алгоритм є мультикласифікатором і здатний опрацьовувати великі об'єми даних. Основною вимогою даного алгоритму являється деталізована база знань, яка будується на основі статистичних вибірок, в деяких випадках ця вимога вважається недоліком через неможливість відбору великої статистичної вибірки. На відміну від алгоритму наївного Баєса метод опорних векторів є бінарним класифікатором. Задача класифікації методом опорних векторів полягає в відшукуванні гіперплощини в n -вимірному просторі для розділення об'єктів навчальної вибірки представлених в вигляді точок на два класи та побудові апроксимації вхідного вектору для відшукування приналежності до першого або другого класу. Через потребу мультикласифікації метод опорних векторів як самостійний окремий алгоритм використовувати неможливо тому для створення мультикласифікатора на основі даного методу використовується модель – надбудова «Один проти всіх» або «Один проти одного». Для методу опорних векторів наявність великої бази знань не обов'язкова, достатньо лише загальні характеристики для проведення класифікації. Метод К-ближніх сусідів – це метод класифікації, який використовує властивості евклідових відстаней для визначення приналежності до певного класу. Цей метод є бінарним і для багатокласової класифікації потрібно також використовувати модель надбудову «Один проти всіх» або «Один проти одного», що призводить до збільшення витрати обчислювального ресурсу та збільшення часу на обробку. Таким чином для проведення класифікації було обрано алгоритм наївного Баєса тому, що даний алгоритм є мультикласифікатором і відсутня необхідність реалізації моделі – надбудови. Також алгоритм використовує мало обчислювальних ресурсів для проведення класифікації та швидко відпрацьовує на великих об'ємах бази знань. Деталізація бази знань буде використовуватися для більш чіткої класифікації.

Проаналізувавши описані вимоги розроблено архітектуру функціонування хмарного обчислювального середовища для обробки анамнезу пацієнтів з захворюванням дихальної системи на основі алгоритму машинного навчання (рис. 1).

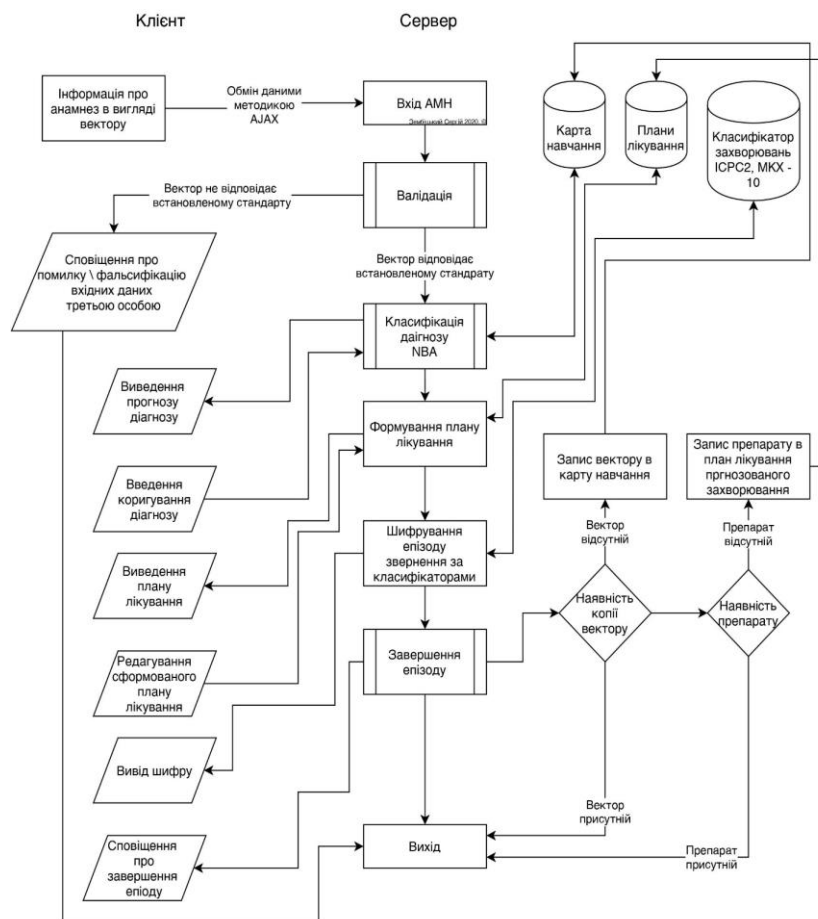


Рис. 1. Блок-схема архітектури хмарного обчислювального середовища

Середовище функціонує за наступним алгоритмом: після введення анамнезу пацієнта на web сторінці на вхід подається методом AJAX вектор анамнезу пацієнта. Вектор проходить валідацію, за невідповідності вектора до встановленої структури користувачу виводиться на екран сповіщення про помилку в системі або підробку даних з боку третіх осіб і рекомендується перезавантажити сторінку, якщо вектор задовольняє всі умови та відповідає встановленій структурі він передається на вхід обчислювального ядра алгоритму наївного Баєса. На етапі класифікації ядро АНБ отримує актуальний зліпок бази знань з карти навчання, проводить класифікацію вхідного вектору, передає методом AJAX користувачу інтерпретований вихідний вектор, який містить в собі назву діагнозу відповідно до введенного анамнезу, та виводить його на екран. Задля збереження людського фактора в системі розроблена можливість змінювати діагноз якщо на думку користувача (сімейного лікаря) він не відповідає тому, що встановив АНБ. На етапі формування плану лікування пацієнта на вхід методом AJAX подається вектор діагнозу, ядро відповідальне за формування плану лікування звертається до JSON структурованого об'єкту, який містить плани лікування та формує план лікування згідно вказаних умов. Сформований план лікування передається користувачу методом AJAX, пулом даних та виводиться на екран. За потреби користувач може редагувати план лікування. Далі процес виконання передається в модуль шифрування епізоду звернення пацієнта. На даному етапі ядро модулю шифрування епізоду звертається до JSON структурованого об'єкту, який містить шифри ICPC2 та МКХ – 10 та за вхідним вектором кінцевого діагнозу, переданим методом AJAX, проводить шифрування епізоду за системами ICPC2 та МКХ – 10. Модуль шифрування епізоду звернення повертає значення шифру методом AJAX та виводить на екран користувача. На завершальному етапі проводиться підготовка до вивантаження пулу даних до баз знань. Спершу проходить перевірка на наявність точної копії в базі знань введенного користувачем вектору анамнезу. За відсутності точної копії вектору анамнезу модуль завершення епізоду проводить перевірку на відповідність структури вектору та вносить новий вектор в базу знань векторів. Якщо точна копія вектору наявна в базі знань модуль завершення епізоду переходить до аналізу плану лікування. Аналізатор плану лікування проводить пошук в структурованому об'єкті JSON бази знань препарати наявні в плані лікування. За відсутності препарати вносяться в базу знань, за наявності – ігноруються. Після внесення пулу інформації до баз знань клієнту виводиться повідомлення про успішність завершення епізоду.

Висновки. На основі поданого матеріалу можна зробити висновок, що реалізований, за даною архітектурою, кінцевий програмний продукт допоможе сімейному лікарю оптимізувати робочий процес, зменшити витрату часу на прийом пацієнта та обробку його анамнезу захворювання, встановити загальний шаблон алгоритму та порядок дій проведення та обробки прийому пацієнта. Дана архітектура має можливість масштабування та розгортання на різних типах хмарних обчислювальних середовищ. Завдяки використанню алгоритму наївного Баєса відповідь від сервера на запит користувач отримує швидко, також даний алгоритм дає можливість обробляти великі навчальні вибірки та швидко проводити класифікацію діагнозу. Основною позитивною властивістю являється модульність архітектури. Зі сторони розробки модульність виражається в можливості легкої та швидкої інтеграції нового розробленого модулю, а зі сторони користувача модульність виражається в можливості використання незалежно один від одного модулів встановлення діагнозу, розробки плану лікування та шифрування звернення пацієнта.

Література

1. Порєва Г.С. Методи аналізу звуків легень для оцінки стану дихальної системи людини / Порєва Г.С. / Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», 2020. – 2 с.
2. Шахгельдян К.И., Гельцер Б.И., Курпатов И.А. Методы машинного обучения для дифференциальной диагностики болезней органов дыхания / Шахгельдян К.И., Гельцер Б.И., Курпатов И.А. / Владивостокский государственный университет экономики и сервиса, 2019. – С. 10–13.

References

1. Porieva H.S. Metody analizy zvukiv lehen dlia otsinky stanu dykhalnoi systemy liudyny / Porieva H.S. / Natsionalnyi tekhnichnyi universytet Ukrainy «Kyivskiy politekhnichnyi instytut imeni Ihoria Sikorskoho», 2020. – 2 s.
2. Shahgeldyan K.I., Gelcer B.I., Kurpatov I.A. Metody mashinnogo obucheniya dlya differencialnoy diagnostiki boleznej organov dyhaniya / Shahgeldyan K.I., Gelcer B.I., Kurpatov I.A. / Vladivostokskij gosudarstvennyj universitet ekonomiki i servisa, 2019. – S. 10–13.

Надійшла / Paper received: 27.04.2020

Надрукована / Paper Printed : 05.06.2020