

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Кафедра КІТС

Магістерська кваліфікаційної робота

на тему: **Нейромережевий персональний асистент користувача
для контролю свого часу та автоматизації повсякденних справ**

Виконав магістрант групи КІТм-21-2
Ткачук Олександр Костянтинович

Науковий керівник:
доктор технічних наук, професор
Корабльов Микола Михайлович

Харків - 2022

Актуальність



2

Мета кваліфікаційної роботи

► **Метою магістерської роботи** розробка нейромережевого персонального асистента користувача для контролю свого часу та автоматизації повсякденних справ.

Вимогами до системи:

- розпізнавання мовлення ;
- синтез мовлення;
- слідування за календарем користувача та його попередження про важливі справи ;
- надання користувачу змоги шукати потрібні речі на сучасних торгових платформах.

У роботі вирішуються наступні завдання:

- Дослідження існуючих персональних асистентів з голосовим управлінням.
- Дослідження методів і алгоритмів нейронних мереж для розпізнавання та синтезу мовлення.
- Дослідження функцій для автоматизації людини у повсякденному житті.
- Розробка нейромережевої моделі для розпізнавання мовлення, а також нейромережевого асистента на основі даної моделі.
- Проведення експериментальних досліджень розробленої нейромережевої моделі для розпізнавання та синтезу мовлення.

3

Аналіз існуючих персональних асистентів

На світовому ринку лідерами серед персональних асистентів є:

- ▶ Siri (системний асистент продукції Apple)
- ▶ Amazon Alexa (периферійний асистент від Amazon)
- ▶ Google Now (системний асистент пристроїв Google)
- ▶ Cortana (персональний асистент ОС Windows 10/11)
- ▶ BlackBerry Assistant (асистент пристроїв BlackBerry)
- ▶ Bixby (мобільний асистент пристроїв Samsung)

4

Переваги та недоліки існуючих персональних асистентів (1)

1. Siri (системний асистент продукції Apple).

Переваги - добре розуміє природне мовлення та обізнаний на новинах, погоді, спорті, кіно, маршрутах, а також вміє взаємодіяти з деякими елементами «розумного будинку».

Основні недоліки полягають у повільній роботі, низькому рівні «людяності», у недоступності для багатьох програм та сервісів.

2. Amazon Alexa (периферійний асистент від Amazon) - є частиною розумного будинку.

Переваги - відкриває жалюзі та двері гаража, включає світло та телевізор, налаштовує термостат та систему кондиціонування, Читає ранкові новини, програвє улюблені пісні, ставить будильники, диктує аудіокниги, викликає таксі.

Основні недоліки полягають у наступному: система намагається перекласти всі відповіді з метою придбання товару; використання обмежене будинком.

5

Переваги та недоліки існуючих персональних асистентів (2)

3. GoogleNow (системний асистент пристроїв Google).
Переваги - розуміє розмовну мову, підтримує найпростіші діалоги. може дзвонити, бронювати квитки та місця у ресторанах.

Основні недоліки: надто ініціативна, показує повідомлення, на які ви не підписувались; вкрай низький рівень «людяності», не може підтримати розмову.

4. Cortana (персональний асистент ОС Windows 10/11).

Переваги - можете вводити унікальні функції, асистент їх запам'ятає та виконуватиме автоматично; бізнес-помічник.

Основні недоліки полягають у недоопрацьованій реалізації унікальних функцій.

5. Vixby (мобільний асистент пристроїв Samsung) - орієнтований на роботу всередині телефону: зміна налаштувань, відправлення текстів та створення нотаток.

Недоліки: обмежена функціональність.

6

Аналіз розпізнавання мовлення для голосового управління

1. Артикуляційний аналіз - полягає у дискретизації рухів усіх мовних артикуляторів.

2. Голосова ідентифікація особи - використовує акустичні характеристики вхідного аудіо потоку для ідентифікації (розпізнавання) особи, щорозмовляє.

Розрізняють два типи голосового визначення особи:

- а) ідентифікація особи;
- б) верифікація особи.

Вимірювання характеристик частотного спектру є ключовим елементом ідентифікації і верифікації особи.

7

Синтез мовлення для голосового управління

Складається з декількох етапів:

1. Трансформація мови та натуральних природніх (звуків) шумів у аналоговий сигнал.
2. Дискретизація аналогового сигналу для отримання цифрового аналога мови.
3. Акустичний аналіз - цифровий аудіо сигнал розділяється на короткі ділянки (фрейми) з тривалістю 20-25 мілісекунд.
4. Аналіз вимовляння або модель вимови для побудови диктор-незалежного розпізнавання
5. Побудова мовної моделі - статистична модель, яка враховує положення слів у реченні або частоту послідовності слів у реченні.
6. Імплементация моделі.

8

Аналіз функцій для автоматизації людини у повсякденному житті

Розроблена реалізація персонального асистента, окрім лише представлення вербального методу введення команд до пристрою, також повинна пропонувати наступні функції для користувача:

- ▶ Знаходження товарів та послуг на різних платформах.
- ▶ Ведення та організація персонального календаря користувача.

9

Інтелектуальні системи розпізнавання мовлення для отримання інформації персональним асистентом

Системи за принципом роботи поділяють на наступні підкатегорії:

- ▶ Інтелектуальні системи розпізнавання злитого мовлення.
- ▶ Інтелектуальні системи голосового управління роботехнічними комплексами.
- ▶ Інтелектуальні системи дикторонезалежних систем розпізнавання.

10

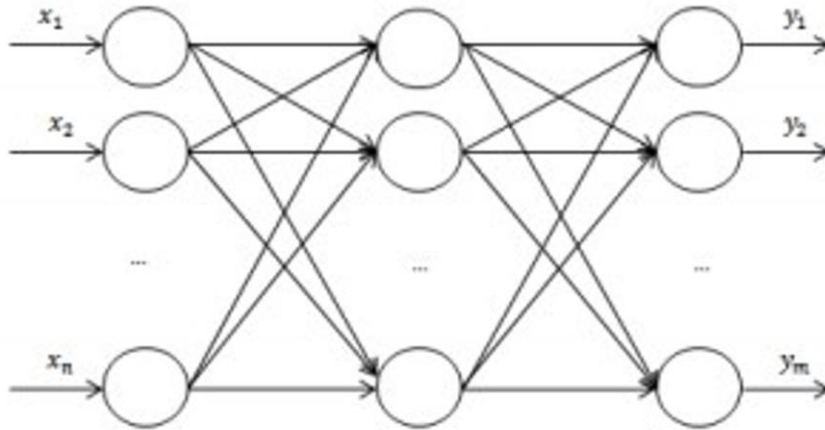
Вибір моделі нейронної мережі для розпізнавання та синтезу мовлення

Основними моделями нейронних мереж виступають:

- ▶ Багатошаровий персептрон – є найбільш поширеною схемою побудови нейронних мереж для вирішення задач обробки та розпізнавання аудіо інформації
- ▶ Згорткова нейронна мережа – це тип багатошарової прямокутної нейронної мережі, яка має у своїй структурі один або декілька згорткових рівнів, і яка може мати повне або часткове з'єднання з нейронами сусідніх шарів.
- ▶ Рекурсивна нейронна мережа – відрізняються від прямокутних багатошарових мереж тим, що вони мають зворотні зв'язки, які формують циклічне коло обробки інформації.

11

Побудова нейромережевої моделі для розпізнавання мовлення

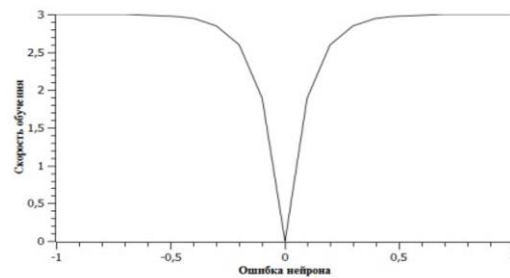


Кількість входів та виходів мережі залежить від кількості вимовлених слів.

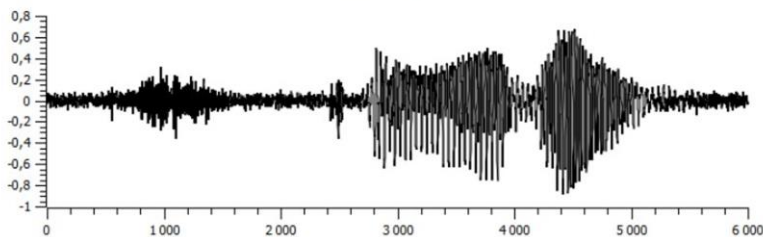
12

Дослідження нейромережевої моделі для розпізнавання мовлення

Графік швидкості навчання окремого нейрона



Тимчасова діаграма слова



13

Експериментальні дослідження

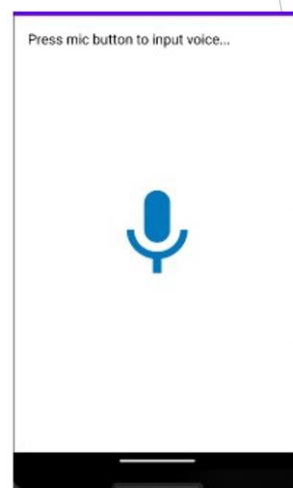
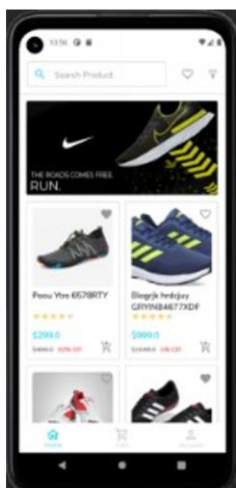
В основі розробленого персонального асистента лежить взаємодія із сервісом Google “Speech-To-Text API”, що використовує технологію трансформуючих рекурентних нейронних мереж, яка принципово відрізняється від інших варіантів нейронних мереж. Структурно схема розпізнавання голосу Google “Text-To-Speech API” також принципово відрізняється від аналогічних сервісів. Класична сучасна модель SARPM передбачає поступове оброблення вхідної аудіо інформації усіма компонентами системи.

Google Speech-to-Text має три основних методи розпізнавання, а саме: синхронне розпізнавання, асинхронне розпізнавання та потокове розпізнавання.

- ▶ Метод синхронного розпізнавання призначено для розпізнавання голосу у коротких аудіо фрагментах з тривалістю до 15 хвилини. Результат розпізнавання надається лише після обробки всього поточного аудіо фрагмента.
- ▶ Асинхронне розпізнавання-цей режим призначений для тривалих аудіо фрагментів до 480 хвилин, і дозволяє додаткові запити для отримання текстових фрагментів, що вже отримані в результаті розпізнавання.
- ▶ Режим потокового розпізнавання призначений для розпізнавання в режимі реального часу, що дозволяє отримувати результати розпізнавання під час вимови користувача.

14

Скріншоти роботи програми



15

Висновки

- ▶ Метою магістерської кваліфікаційної роботи є розробка нейромережевого персонального асистента користувача для контролю часу та організації розпорядку дня і справ користувача та автоматизації його повсякденних справ. Програмне забезпечення працює на платформі Android, разом з цим структура і побудова програмного забезпечення передбачає масштабування для роботи на платформах Windows і Linux.
- ▶ Нейронна мережа, яка використовується у роботі в повному обсязі задовольняє технічним вимогам програмного асистента, розробка якого є метою роботи. Так, тестові випробування показали, що швидкість розпізнавання дозволяє в режимі реального часу виводити на екран результати розпізнавання.
- ▶ Основний висновок стосується використання нейронних мереж при побудові прикладних програмних засобів. В ході експериментальних досліджень було перевірено, що використання нейронних мереж для розпізнавання і синтезу природної людської мови має наймовірний потенціал і дозволить багатократно масштабувати персонального асистента, що розроблений в частині реалізованих команд користувача і підтримку для різних мов спілкування.

16

Дякую за увагу!

17

