

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І.І.МЕЧНИКОВА

(повне найменування вищого навчального закладу)

Факультет математики, фізики та інформаційних технологій

(повне найменування інституту, назва факультету (відділення))

Кафедра комп'ютерних систем та технологій

(повна назва кафедри (предметної, циклової комісії))

Дипломна робота

На здобуття ступеня вищої освіти «магістр»

(освітньо-кваліфікаційний рівень)

на тему: Інформаційна система розпізнання мови / Speech recognition information system

Виконав: студент денної форми навчання

напряму підготовки 123 – Комп'ютерна інженерія

(шифр і назва напряму підготовки, спеціальності)

Козачков Владислав Георгійович

(прізвище, ім'я, по-батькові)

Керівник д.т.н., проф. Гунченко Ю. О.

(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рецензент к. ф.-м. н., доц. Петрушина Т. І.

(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рекомендовано до захисту:

Протокол засідання кафедри

№ від « » 2020 р.

Завідувач кафедри

Захищено на засіданні ЕК №

протокол № від « » 2020 р.

Оцінка / /

(за національною шкалою, шкалою ECTS, бали)

Голова ЕК

Ю.О. Гунченко

(підпис)

(прізвище, ініціали)

Н.Ф. Казакова

(підпис)

(прізвище, ініціали)

Одеса - 2020

АНОТАЦІЯ

Мета дипломної роботи – підвищення якості розпізнання мови у системи з відкритим вихідним кодом.

В роботі зроблено огляд і аналіз систем і методів розпізнання, обрана система для модифікації, обрано метод поліпшення і складено алгоритм його застосування, здійснена модифікація системи і проведено порівняльне тестування декількох варіантів системи.

Базовою системою виступає система CMU Sphinx, яка для акустичного моделювання застосовує метод прихованих Марковських моделей. В процес навчання моделі було додано уточнення параметрів навчання за допомогою нейронної мережі.

Базова система написана на мові C/C++, на ньому ж реалізована модифікація системи. Робота над системою здійснювалася в середовищі розробки Microsoft Visual Studio 2013. Для підготовки і форматування тестових даних з корпусу TED-LIUM під стандарти CMU Sphinx були написані скрипти на мові Python з використанням середовища розробки Python IDLE.

Тестування показало невелике поліпшення характеристик розпізнавання для модифікованої системи, що частково компенсується за рахунок значного підвищення часу навчання системи.

Отримані і навчені варіанти систем можна застосовувати тільки для науково-дослідницьких цілей через ліцензійні обмеження, накладені на корпус TED-LIUM.

ABSTRACT

The purpose of the thesis is to improve the quality of speech recognition in an open source system.

In this work, a review and analysis of systems and recognition methods is made, a system for modification is selected, an improvement method is selected and an algorithm for its application is drawn up, the system is modified and comparative testing of some variants of the system is carried out.

The basic system is the CMU Sphinx system, which uses the method of hidden Markov models for acoustic modeling. In the process of training the model, the refinement of the training parameters using a neural network was added.

The base system is written in the C/C++ language, and the system modification is implemented in it as well. The work on the system was carried out in the Microsoft Visual Studio 2013 development environment. To prepare and format test data from the TED-LIUM corpus for the CMU Sphinx standards, scripts were written in Python, the development environment used was Python IDLE.

Testing showed a slight improvement in recognition performance for the modified system, which is partially offset by a significant increase in the system learning time.

The received and trained versions of the systems can only be used for research purposes due to the licensing restrictions imposed on the TED-LIUM corpus.

АННОТАЦИЯ

Цель дипломной работы – повышение качества распознавания речи у системы с открытым исходным кодом.

В работе сделан обзор и анализ систем и методов распознавания, выбрана система для модификации, выбран метод улучшения и составлен алгоритм его применения, осуществлена модификация системы и проведено сравнительное тестирование нескольких вариантов системы.

Базовой системой выступает система CMU Sphinx, которая для акустического моделирования применяет метод скрытых Марковских моделей. В процесс обучения модели было добавлено уточнение параметров обучения при помощи нейронной сети.

Базовая система написана на языке C/C++, на нем же реализована модификация системы. Работа над системой осуществлялась в среде разработки Microsoft Visual Studio 2013. Для подготовки и форматирования тестовых данных из корпуса TED-LIUM под стандарты CMU Sphinx были написаны скрипты на языке Python, используемая среда разработки – Python IDLE.

Тестирование показало небольшое улучшение характеристик распознавания для модифицированной системы, что частично компенсируется за счет значительного повышения времени обучения системы.

Полученные и обученные варианты систем можно применять только для научно-исследовательских целей из-за лицензионных ограничений, наложенных на корпус TED-LIUM.

ЗМІСТ

	стр.
ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ І ТЕРМІНІВ	8
ВСТУП	9
1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Класифікація систем розпізнання мови	11
1.2 Огляд методів розпізнання мови	13
1.3 Огляд систем розпізнання мови.....	16
1.4 Порівняння відкритих систем розпізнання мови.....	22
2 МЕТОДИ ТА АЛГОРИТМИ АКУСТИЧНОГО МОДЕЛЮВАННЯ.....	26
2.1 Приховані Марковські моделі	26
2.1.1 Алгоритм Баума-Велша.....	30
2.1.2 Алгоритм Вітербі	33
2.2 Нейронні мережі.....	34
2.3 Гібридні моделі ПММ/НМ.....	40
3 ПРОЕКТУВАННЯ СИСТЕМИ.....	45
3.1 Система Sphinx	45
3.1.1 Обчислення ознак.....	46
3.1.2 Перевірка вхідних даних та параметрів.....	46
3.1.3 Тренування контекстно-незалежних моделей.....	47
3.1.4 Тренування незв'язаних контекстно-залежних моделей	48
3.1.5 Побудова дерев.....	50
3.1.6 Скорочення дерев.....	50
3.1.7 Тренування зв'язаних контекстно-залежних моделей	51
3.1.8 Тестування системи	52

3.2 Модифікація.....	53
4 ВИПРОБУВАННЯ СИСТЕМИ.....	55
4.1 Підготовка тестових даних.....	55
4.2 Тестування систем.....	55
ВИСНОВОК.....	59
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60
Скрипти Python для підготовки навчальних та тестових даних	65
Слайди презентації доповіді	70
Копія публікації на ICT-2019.....	82

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ І ТЕРМІНІВ

MFCC – Mel Frequency Cepstral Coefficients

НМ – нейронна мережа

ПММ – прихована Марковська модель

WER – Word Error Rate

WRR – Word Recognition Rate

SF – Speed Factor

БВ – Баум-Велш

ВТ – Вітербі

БП – багатошаровий персептрон

PER – Phone Error Rate

ВСТУП

Застосування комп'ютерних систем у сучасному світі є надзвичайно широким і вони присутні практично у всіх сферах людської діяльності. Комп'ютерні системи призначені для того, щоб знизити втрати часу та сил при опрацюванні різних даних та інформації, незалежно від їх типу та області, до якої вони відносяться. Проте, взаємодія людини з комп'ютером та з приладами поки що не є ідеальною або досить ефективною. Зазвичай для взаємодії з системами потрібні специфічні засоби введення, а також певні навички роботи з ними.

Ситуація значно полегшується, коли при зверненні до систем або їх компонентів люди, що працюють з цими системами, можуть використовувати природну мову. Мовне спілкування – інструмент, який використовується людством протягом багатьох тисяч років, і для людей воно давно є природним та зручним. Застосування мовного спілкування для доступу до штучних систем дозволяє значно знизити час, що втрачається на взаємодію, а також зменшити складність систем за рахунок заміни більшості пристроїв введення на мікрофони з включенням у програмне забезпечення підсистем автоматичного розпізнання мови.

Перші системи для розпізнання мови з'явилися ще у 50-х роках ХХ ст., тоді вони були здатні виконувати зовсім невеликі задачі, такі як розпізнання цифр, вимовлених людиною [1]. Комерційні програми для розпізнання мови з'явилися наприкінці 80-х – початку 90-х років ХХ ст.. З часом надійність та швидкість програм для розпізнавання поступово покращуються, але збільшується і їх складність та обсяг ресурсів, яких вони потребують. Програми розпізнання мови стали створювати і для мобільних пристроїв, коли їх обчислювальні потужності досягли достатнього рівня.

На даний момент існує значна кількість програм та утиліт, що можуть виконувати або надавати функції для розпізнання мови. Вони можуть бути готовими прикладними програмами або движками, що вбудовуються у інші

системи, вони можуть бути з відкритим або закритим вихідним кодом, і всі вони мають різні властивості та різні показники ефективності при виконанні цільової задачі.

Об'єктом дослідження є системи розпізнання природної мови.

Предметом дослідження є методи та процеси акустичного моделювання у розпізнанні мови.

Метою роботи є підвищення якості системи розпізнання мови з відкритим вихідним кодом.

Під підвищенням якості будемо розуміти поліпшення таких показників, як точність та швидкість процесу розпізнання.

Для досягнення поставленої мети у роботі необхідно вирішити наступні взаємопов'язані задачі:

- 1) огляд та аналіз предметної області, зокрема відкритих систем розпізнання мови;
- 2) огляд та аналіз методів та алгоритмів акустичного моделювання;
- 3) модифікація існуючої системи;
- 4) підготовка навчальних та тестових даних;
- 5) тестування та порівняння декількох варіантів систем.

ВИСНОВОК

В результаті проведеного у роботі огляду та аналізу предметної області обрано базову систему для подальшої роботи та критерії для оцінки майбутніх результатів.

В результаті проведеного у роботі огляду та аналізу методів моделювання обрано метод та спосіб для його реалізації.

Визначено та реалізовано алгоритм та структуру нової системи, виконано модифікацію обраної раніше системи.

Навчальні та тестові дані перевірено, переведено у потрібний формат. Сформовано навчальний масив даних значного обсягу, сформовано тестові масиви даних.

Проведено порівняльне тестування систем, проаналізовано результати тестування. Новий варіант системи має кращі показники точності розпізнання, ніж базовий варіант, і майже ідентичні показники швидкості розпізнання, проте потребує значно більше часу для навчання системи.

В майбутньому дослідження можна продовжити декількома способами. Можна спробувати далі покращити якість роботи систем, можна пробувати експериментувати з архітектурою модифікованої системи для подальшого покращення точності або швидкості навчання, можна задіяти навчальні та тестові масиви даних значно більшого обсягу.

Можливості практичного застосування наявних систем обмежені, бо хоча система Sphinx розповсюджується під ліцензією типа BSD, яка дозволяє комерційне поширення, але використаний корпус TED-LIUM має обмеження Attribution-NonCommercial-NoDerivs. Системи та додатки, побудовані на TED-LIUM, можна застосовувати лише у науково-дослідницьких цілях.

Автором опубліковано тези, відповідні даній роботі, на конференції «Інформаційні системи та технології ICT-2019» [32] Копію публікації представлено у додатку В.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Распознавание речи – Национальная библиотека им. Н.Э. Баумана [Электронный ресурс] – Режим доступа: https://ru.bmstu.wiki/%D0%A0%D0%B0%D1%81%D0%BF%D0%BE%D0%B7%D0%BD%D0%B0%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D1%80%D0%B5%D1%87%D0%B8.
2. Александр Прохоров. Системы автоматического распознавания речи | КомпьютерПресс [Электронный ресурс] – Режим доступа: <https://compress.ru/article.aspx?id=11331>.
3. Воробьева С.А. Методы распознавания речи // Молодой ученый. — 2016. — №26. — С. 136-141. [Электронный ресурс] – Режим доступа: <https://moluch.ru/archive/130/36213/>.
4. Беленко М.В., Балакшин П.В. Сравнительный анализ систем распознавания речи с открытым кодом // Международный научно-исследовательский журнал. – 2017. - №4(58) – С. 13-18. [Электронный ресурс] – Режим доступа: <https://research-journal.org/technical/sravnitelnyj-analiz-sistem-raspoznavaniya-rechi-s-otkryтым-kodom/>.
5. А. К. Алимуратов, П. П. Чураков. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи // Измерение. Мониторинг. Управление. Контроль. - 2015. -№ 2 (12). - С. 27-35 [Электронный ресурс] – Режим доступа: <https://cyberleninka.ru/article/n/obzor-i-klassifikatsiya-metodov-obrabotki-rechevyh-signalov-v-sistemah-raspoznavaniya-rechi>.
6. Алборова Ж. В. Алгоритм и методы распознавания речи // Молодежный научно-технический вестник. – 2016. – [Электронный ресурс] – Режим доступа: <https://docplayer.ru/42471925-Algorithm-i-metody-raspoznavaniya-rechi.html>.

7. Алюнов Д.Ю., Сергеев Е.С., Пигачев П.В., Мытников А.Н. Реализация алгоритма обработки и распознавания речи // Современные наукоемкие технологии. – 2016. – № 3-2. – С. 225-230 – [Электронный ресурс] – Режим доступа: <http://www.top-technologies.ru/ru/article/view?id=35724>.
8. Маковкин К.А. Гибридные модели: скрытые Марковские модели и нейронные сети, их применение в системах распознавания речи // Модели, методы, алгоритмы и архитектуры систем распознавания речи – 2006.
9. Локтионов Н.П. Обзор систем распознавания речи // Молодежный научный форум: Технические и математические науки: электр. сб. ст. по мат. XLIII междунар. студ. науч.-практ. конф. № 3(43). URL: [https://nauchforum.ru/archive/MNF_tech/3\(43\).pdf](https://nauchforum.ru/archive/MNF_tech/3(43).pdf).
10. Darren Allan. Best voice recognition software of 2019 | TechRadar [Электронный ресурс] – Режим доступа: <https://www.techradar.com/best/best-voice-recognition-software>.
11. J. Simpson. 5 Best Speech-to-Text APIs | Nordic APIs | [Электронный ресурс] – Режим доступа: <https://nordicapis.com/5-best-speech-to-text-apis/>.
12. Voice Recognition on Apple devices [Электронный ресурс] – Режим доступа: https://www.learningapps.co.uk/moodle/xertetoolkits/play.php?template_id=1315.
13. Amazon Transcribe – Автоматическое распознавание речи – AWS [Электронный ресурс] – Режим доступа: <https://aws.amazon.com/ru/transcribe/>.
14. Basic concepts of speech recognition – CMUSphinx Open Source Speech Recognition [Электронный ресурс] – Режим доступа: <https://cmusphinx.github.io/wiki/tutorialconcepts/>.

15. About Simon | Simon [Электронный ресурс] – Режим доступа: <https://simon.kde.org/>.
16. Даниил Бойко. Системы распознавания речи с открытым исходным кодом | LVEE [Электронный ресурс] – Режим доступа: <https://lvee.org/en/abstracts/273>.
17. Д. В. Бобкин, К. Ю. Жигалов. Исследование надежности распознавания речи системой Google Voice Search // Cloud of Science – 2015 [Электронный ресурс] – Режим доступа: <https://cyberleninka.ru/article/n/issledovanie-nadezhnosti-raspoznavaniya-rechi-sistemoy-google-voice-search>.
18. Cloud Speech-to-Text – Speech Recognition | Cloud Speech-to-Text | Google Cloud. [Электронный ресурс] – Режим доступа: <https://cloud.google.com/speech-to-text/>.
19. Распознавание речи от Яндекса. Под капотом у Yandex.SpeechKit / Блог компании Яндекс / Хабр (2013). [Электронный ресурс] – Режим доступа: <https://habr.com/ru/company/yandex/blog/198556/>.
20. The Baum–Welch algorithm for estimating a Hidden Markov Model. [Электронный ресурс] – Режим доступа: <https://web.archive.org/web/20070929091507/http://www.ph.biu.ac.il/faculty/kanter/BW.pdf>
21. Jeff A. Bilmes. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models – 1998. [Электронный ресурс] – Режим доступа: <https://imaging.mrc-cbu.cam.ac.uk/methods/BayesianStuff?action=AttachFile&do=get&target=bilmes-em-algorithm.pdf>
22. L. J. Rodriguez, I. Torres. Comparative Study of the Baum-Welch and Viterbi Training Algorithms // Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, 2652/2003, 847 (2003).

- 23.Armen Allahverdyan, Aram Galstyan. Comparative Analysis of Viterbi Training and Maximum Likelihood Estimation for HMMs. [Электронный ресурс] – Режим доступа: https://www.isi.edu/~galstyan/papers/hmm_parameter_nips.pdf
- 24.Обучение Витерби или алгоритм Баума-Уэлча для оценки вероятностей перехода и эмиссии? - CodeRoad. [Электронный ресурс] – Режим доступа: <https://coderoad.ru/13360892/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5-%D0%92%D0%B8%D1%82%D0%B5%D1%80%D0%B1%D0%B8-%D0%B8%D0%BB%D0%B8-%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC-%D0%91%D0%B0%D1%83%D0%BC%D0%B0-%D0%A3%D1%8D%D0%BB%D1%87%D0%B0-%D0%B4%D0%BB%D1%8F-%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B8-%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9-%D0%BF%D0%B5%D1%80%D0%B5%D1%85%D0%BE%D0%B4%D0%B0-%D0%B8>
- 25.G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition (2012)
- 26.CMUSphinx Documentation – CMUSphinx Open Source Speech Recognition [Электронный ресурс] – Режим доступа: <https://cmusphinx.github.io/wiki/>
- 27.C++ Language Reference [Электронный ресурс] – Режим доступа: <https://msdn.microsoft.com/en-us/library/3bstk3k5.aspx>
- 28.ActiveState Platform [Электронный ресурс] – Режим доступа: <https://platform.activestate.com/>

29. TED-LIUM Release 1 - Laboratoire d'Informatique de l'Université du Mans [Електронний ресурс] – Режим доступу: <https://lium.univ-lemans.fr/en/ted-lium1/>
30. Download Release 1 – CORPUS TED LIUM [Електронний ресурс] – Режим доступу: <https://projets-lium.univ-lemans.fr/ted-lium/release1/>
31. Python Tutorials – Real Python [Електронний ресурс] – Режим доступу: <https://realpython.com/>
32. Владислав Козачков, Олексій Маслєєв, Юрій Гунченко. Використання нейронних мереж у системах розпізнання мови // Інформаційні системи та технології ICT-2019. С. 270–272