

РОЗДІЛ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ»

УДК 004.934

DOI 10.31319/2519-2884.32.2018.174

ЯЛОВА К.М., к.т.н., доцент
ЯШИНА К.В., к.т.н., доцент
ВАСИЛЬЄВА А.В., магістр

Дніпровський державний технічний університет, м. Кам'янське

ДОСЛІДЖЕННЯ СИСТЕМ ТА МЕТОДІВ РОЗПІЗНАВАННЯ МОВНОГО СИГНАЛУ

Вступ. Мова – основний вид передачі думок, ідей та почуттів у людському середовищі. Можливість контролю оточення голосом підштовхнув науковців та експертів передових корпорацій до створення таких програмних додатків, як Cortana від Microsoft, Siri від Apple та інших. Вирішення задачі розпізнавання голосу дозволив перейти від командного та WIMP (Window, Image, Menu, Pointer) до SILK (Speech, Image, Language, Knowledge) інтерфейсу програмних систем. На відміну від взаємодії користувача з комп'ютерною системою за допомогою клавіатури, миші, джойстика та дисплею мовний інтерфейс має наступні переваги:

- для спілкування з комп'ютером людині немає необхідності мати спеціальні навички або уміння в галузі інформаційних технологій;
- мова знижує психологічну та фізичну відстань між людиною і комп'ютером та може бути пов'язана з ним через системи комунікацій, наприклад, телефон;
- мовний інтерфейс надає оперативність і мобільність спілкування, звільнення рук і розвантаження зорового каналу при отриманні інформації [1].

Постановка задачі. Метою даної роботи є дослідження та проведення аналітичного порівняння систем автоматичного розпізнавання мови (САРМ), методів та алгоритмів, що використовуються при розпізнаванні мовного сигналу.

Результати роботи. Під розпізнаванням мови розуміють процес трансформації мовного сигналу в цифрову інформацію (наприклад, текстові дані) [2]. Для цього існує безліч САРМ – систем, що перетворюють вхідний мовний сигнал в розпізнане повідомлення [3]. При цьому повідомлення може бути представлено як у формі тексту цього повідомлення, так і одразу перетворено в зручну для його подальшої обробки форму з метою формування відповідної реакції програмної системи. САРМ класифікуються за такими ознаками, як:

1) розмір словника (обмежений набір слів або великий словник) – чим більший розмір словника, з яким працює система розпізнавання мови, тим більше помилок при розпізнаванні слів. Словник, що складається тільки з цифр, може бути розпізнаний практично безпомилково, тоді як вірогідність помилок при розпізнаванні словника в сто тисяч слів може досягати 45%. Потрібно також враховувати унікальність слів в словнику. Якщо слова дуже схожі, то похибка розпізнавання збільшується;

2) залежність від диктора (дикторозалежні або дикторонезалежні) – дикторозалежна система призначена для роботи тільки з людиною, яка навчала цю систему, в той час як дикторонезалежна система призначена для роботи з будь-яким диктором. На поточному етапі розвитку САРМ вірогідність виникнення помилок в дикторонезалежній системі в 3-5 разів більша, ніж у дикторозалежних;

3) тип мови (злитна, роздільна). Роздільна мова – це мова, в якій слова відокремлюються одне від одного проміжком тиші. Злитна мова – це природно вимовлений

текст. Розпізнавання злитного мовлення складніше, тому що у вимовлених слів немає чітких меж. САРМ, що працюють з ізольованими словами, досягли високо рівня точності розпізнавання – 95-99%, в той час як задача розпізнавання злитної мови в достатній мірі не вирішена [4]. У системах для розпізнавання злитного мовлення від ІВМ та Microsoft частота помилок становить 5.5-5.9%;

4) призначення (системи диктування, командні системи) – визначає необхідний рівень абстракції, на якому буде відбуватися розпізнавання мови. Системи голосового набору мобільного телефону, де здійснюється розпізнавання за шаблоном, називаються командними. На відміну від них, система диктування вимагає розпізнавання на базі виділення лексичних елементів. При інтерпретації виголошеної фрази вона буде покладатися не тільки на те, що було виголошено в поточний момент, але і на те, як це співвідноситься з тим, що було вимовлено до цього. Також в таку систему повинен бути вбудований набір граматичних правил. Чим суворіші ці правила, тим простіше реалізувати систему розпізнавання, але набір слів, які вона зможе розпізнати, буде меншим;

5) алгоритм, що використовується. Після того, як мовний сигнал розбивається на певні частини, відбувається імовірнісна оцінка належності цих частин до того чи іншого елемента словника, що здійснюється за допомогою одного з алгоритмів розпізнавання;

6) по типу структурної одиниці (фрази, слова, фонем, діфони, алофони) – САРМ, які використовують цілі слова або фрази, називаються САРМ за шаблоном. Вони як правило дикторозалежні, і їх реалізація є простішою, ніж створення САРМ, які розпізнають мовлення на базі виділення лексичних елементів. У таких системах структурними одиницями мови є лексичні елементи;

7) по принципу виділення структурних одиниць. Найпоширеніший підхід виділення структурних одиниць заснований на перетворенні Фур'є, яке переводить вихідний сигнал з амплітудно-часового простору в частотний. Однак аналіз Фур'є має цілу низку недоліків, в результаті яких відбувається втрата інформації стосовно часових характеристик оброблюваних сигналів. У зв'язку з цим для завдання виділення структурних одиниць мови виправдано використання вейвлет-аналізу. Вейвлет – це математична функція, яка дозволяє аналізувати частотні компоненти даних. В загальному випадку, аналіз сигналів проводиться в площині вейвлет-коефіцієнтів – масштаб-час-рівень (Scale-Time-Amplitude). Отримані вейвлет-спектри відрізняються від спектрів Фур'є тим, що дають чітку прив'язку властивостей сигналу до часу. Крім вейвлет і Фур'є-

аналізу в САРМ використовується кепстральний аналіз, але створення таких систем є трудомістким і вимагає високої кваліфікації розробника.

Спрощену структурну схему роботи САРМ наведено на рис.1.

Метою аналізу мовного сигналу є виділення в складі отриманого сигналу компонентів, які є основними для розпізнавання отриманого повідомлення. До таких компонентів належать параметри, що описують мову, аналогічні тим, які формуються в процесі синтезу мови. Набір зазначених параметрів залежить від обраного методу розпізнавання.

Модель розпізнавання мови і прийняття рішення – це блок, в рамках якого здійснюється формування розпізнаного повідомлення на основі аналізу послідовності параметрів, отриманих після аналізу мовного сигналу. Наприклад, якщо вико-



Рисунок 1 – Структурна схема автоматичного розпізнавання мови

ристовується формантна модель опису мови, то на основі отриманих в першому блоці частот формант будується послідовність розпізнаних фонем, що складають вхідне повідомлення. При цьому здійснюється прийняття рішення про те, чи розпізнано вхідне повідомлення правильно. При прийнятті рішення можливі наступні варіанти: повідомлення розпізнано правильно (підтвердженням цього є текст, що відповідає нормам природної мови) або повідомлення не розпізнається, або розпізнано неправильно (таке рішення приймається в разі наявності в розпізаному повідомленні явних помилок, які важко виправити автоматично, або взагалі повна нісенітниця).

У процесі розпізнавання мови найскладніше полягає в здійсненні процедури порівняння вхідного та еталонного елемента, задача ускладнюється ще й тим, що вхідні сигнали характеризуються протяжністю в часі. На теперішній час існують багато методів та алгоритмів розпізнавання мови, найпоширенішими з яких є: лінійні моделі, нейронно-мережеві методи, приховані марковські моделі, метод динамічного трансформування часу тощо.

Лінійні моделі. Першими з'явилися лінійні моделі розпізнавання мови. В них передбачається, що для порівняння мовного сигналу з еталоном досить простого масштабування в часі. Метод лінійної екстраполяції передбачає представлення мовного сигналу $s(n)$ в якості комбінації попередніх відліків сигналу, а математично модель сигналу представляється у вигляді:

$$s(n) = - \sum_{i=1}^{N_{LP}} \alpha_{LP}(i) s(n-i) + e(n),$$

де N_{LP} – кількість коефіцієнтів моделі або порядок передбачення; α_{LP} – коефіцієнти лінійного передбачення; $e(n)$ – функція помилки моделі (різниця між передбаченим значенням і реально зміненим значенням). Помилка передбачення визначається у вигляді різниці між вихідним та передбаченим відліками сигналу. Основна задача методу лінійного передбачення зводиться до визначення набору коефіцієнтів передбачення, які б забезпечили мінімізацію помилки передбачення.

Оскільки в мові мають місце нелінійні спотворення часу, тобто лінійна модель передбачала порівняння реалізації з еталоном за лінійним законом, тоді як зміни в реалізації піддаються нелінійних спотворень.

Нейронні мережі. Одним з найбільш ефективних методів розпізнавання мови є метод з використанням нейронних мереж. Нейрон представляє собою комірку мережі і може знаходитись у збудженому або загальмованому стані та має зв'язки з іншими нейронами мережі – синапси (однонаправлені вхідні зв'язки), аксони – вихідні зв'язки нейрона, по яким сигнали (збудження або гальмування) надходять до синапсів наступних нейронів.

Кожний односпрямований зв'язок характеризується вагою w_i . Позитивні та негативні значення w_i відповідають збудженому або загальмованому стану синапсів. Сума всіх входів визначає поточний стан нейрона, що визначається за формулою:

$$S = \sum_{i=1}^n x_i w_i.$$

Результат підсумовування передається до функції активації, яка може бути представлена нарізно, але найпоширенішою є логістична функція, яку можна розрахувати за формулою:

$$F(S) = \frac{1}{1 + e^{-\alpha S}}.$$

При використанні нейронних мереж для розпізнавання мовних сигналів необхідно побудувати відповідну призначену для цієї задачі мережу та підібрати вагові коефіцієнти синапсів для мінімізації помилок.

Динамічне трансформування часу (ДТЧ). (ДТЧ – Dynamic Time Warp). Вимова одного й того ж слова, зазвичай, має різну тривалість. Навіть якщо слово було вимовлено з однаковою тривалістю, тривалість вимови окремих частин слова може бути різною. Тому, щоб отримати оцінку розходження між двома мовними сигналами у вигляді вектора, має бути виконано вирівнювання за часом, яке реалізується за допомогою ДТЧ. ДТЧ – метод еластичного порівняння вектора спостереження зі збереженим шаблоном. Вектор спостережень та шаблон лежать на відповідних осях сітки. Для кожної комірки сітки розраховується різниця між відповідними фрагментами вектора спостережень та шаблону.

Метод ДТЧ працює з фрагментами мовного сигналу, тобто аналіз ознак складається з обробки вектора ознак в регулярних інтервалах. Вектор ознак може мати велику кількість фрагментів, тому є потреба у засобах розрахунку локальної оцінки відстані між точками сигналу x та шаблону y у n -вимірному просторі за формулою:

$$(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

де x_i, y_i – елементи, що порівнюються, i – номер фрагмента.

Таким чином, вхідний сигнал порівнюється зі всіма шаблонами. Результатом порівняння буде шаблон, для якого знайдено мінімальне розходження між вхідним сигналом і шаблоном, що є сумою локальних відстаней між фрагментами сигналу і шаблону.

Оскільки для визначення основи послідовності в динамічному програмуванні оптимальним є використання методу зворотного програмування, необхідно використовувати певний динамічний стек. Подібно будь-якому динамічному алгоритму програмування ДТЧ має поліноміальну складність. У випадку великих послідовностей виникають незручності із збереження великих числових матриць та виконання великої кількості розрахунків відхилень. Існує поліпшена версія алгоритму, FastDWT, яка вирішує дві вищевказані проблеми. Рішення полягає в розбитті матриці станів на 2, 4, 8, 16 і т.д. менших за розміром матриць за допомогою повторюваного процесу розбиття послідовності введення на дві частини. Таким чином, розрахунки відхилення здійснюються тільки на цих невеликих матрицях і шляхах деформації, розрахованих для невеликих матриць.

У роботі [5] авторів Запрягаєва С.А. та Коновалова А.Ю. представлено результати розробки програмного додатку для розпізнавання командних слів. Якість розпізнавання методом ДТЧ в даній роботі становить 94%. В роботі [6] авторів Є.С.Малькової і О.А.Шабаліної описано методи розпізнавання мови в задачі автоматизованого виявлення дефектів вимови. Автори статті змогли створити простий класифікатор із 80% коректністю класифікації при застосуванні методу ДТЧ.

Приховані марковські моделі. В основі прихованої марковської моделі (ПММ) лежить кінцевий автомат, що складається з N прихованих станів. Перехід між станами в кожний дискретний момент часу t не є детермінованим, а відбувається відповідно до ймовірнісного закону і описується матрицею переходів. Знаходження моделі в стані i відповідає певній стаціонарності сигналу на обмеженому інтервалі часу. При здійсненні чергового переходу в новий стан i в момент часу t здійснюється генерація вихідного параметричного вектора x_t у відповідності до багатомірної функції розподілу ймовірностей $f_j(x)$. Результатом роботи ПММ є послідовність векторів спостереження довжиною T . Перевагою ПММ є можливість обробки послідовностей і сигналів різної

довжини, що ускладнено при роботі зі штучними нейронними мережами. Функція щільності ймовірностей $f_j(x)$ для стану j описується:

$$f_j(x) = \sum_{i=1}^M w_i p_i(x),$$

де M – кількість компонентів; w_i – вага компонента; $p_i(x)$ – нормальний розподіл для D -вимірного випадку.

Функція $p_i(x)$ описується наступним виразом:

$$p_i(x) = \frac{1}{2\pi^{\frac{D}{2}} |\sigma_i|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2(x - \mu_i)^T \sigma_i^{-1} (x - \mu_i)} \right\},$$

де D – розмірність вектору; μ_i – вектор математичного очікування; σ_i – матриця ко-варіації.

Робота з ПММ проводиться в 2 етапи:

- 1) визначення параметрів моделі (алгоритм Баума-Велча);
- 2) визначення ймовірності того, що послідовність векторів, за якими ведеться спостереження, була згенерована даною моделлю (алгоритм Вітербі).

Для якісного навчання ПММ необхідна велика кількість зразків сигналу. Також необхідно дотримуватись умов лінійної незалежності навчальних зразків, у разі невиконання чого має місце виродження матриці коваріації, що може призвести до повної непрацездатності моделі. Однією з найуспішніших САРМ, що використовує ПММ є дикторонезалежна система розпізнавання неперервної мови Sphinx.

Висновки. Швидкий розвиток інформаційних технологій та комп'ютерної техніки ставить перед науковцями та винахідниками задачу оптимізації машинно-людинного інтерфейсу та зменшення проблем комунікації між людиною та комп'ютером. Створення мовних інтерфейсів може знайти застосування в системах різного призначення: голосове управління для людей з обмеженими можливостями, автовідповідачі, оброблення в автоматичному режимі сотні тисяч дзвінків на добу (наприклад, в системі продажу авіаквитків) тощо.

Одним із можливих способів скорочення фізичної та психологічної відстані між користувачем та комп'ютерною технікою є застосування мовного інтерфейсу для керування комп'ютерними пристроями та системами на відстані, базуючись на мовні сигнали користувача. В даній роботі подано класифікацію САРМ за різними властивостями та признаками, перелічені представники САРМ та описані якісні показники розпізнавання мовного сигналу. На основі аналітичного огляду представлено основні методи обробки мовних сигналів, які використовуються в САРМ, а саме: лінійне передбачення мовного сигналу, нейронні мережі, приховані марковські моделі та метод динамічного трансформування часу. Поданий опис дозволяє оцінити можливості існуючих методів обробки мовних сигналів і визначити перспективність застосування їх математичних апаратів у задачах обробки мовних сигналів у САРМ.

ЛІТЕРАТУРА

1. Алимуратов А.К. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи / А.К.Алимуратов, П.П.Чураков // Измерение. Мониторинг. Управление. Контроль. – 2015. – №2(12). – С.27-35.
2. Федосин С.А. Классификация систем распознавания речи [Электронный ресурс] / Федосин С.А., Еремин А.Ю. – Режим доступа: <http://fetmag.mrsu.ru/2010-2/pdf/SpeechRecognition.pdf>.

3. Аграновский А.В. Теоретические аспекты алгоритмов и классификации речевых сигналов / А.В.Аграновский, Д.А.Леднов. – М.: Радио и связь, 2004. – 164с.
4. Титов Ю.Н. Современные технологии распознавания речи / Ю.Н.Титов // Вестник ТГУ. – 2006. – Т.11. – Вип.4. –С.571-574.
5. Запрягаев С.А. Распознавание речевых сигналов / С.А.Запрягаев, А.Ю.Коновалов // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2009. – №2. – С.37-46.
6. Малькова Е.С. Методы распознавания речи в задаче автоматизированного выявления дефектов произношения / Е.С.Малькова, О.А.Шабалина // Известия Волгоградского государственного технического университета. – 2015. – №2. – С.65-71.

Надійшла до редколегії 29.01.2018.

УДК 004.42

DOI 10.31319/2519-2884.32.2018.175

ДЕМЧЕНКО Ю.Ю., студент
БАБЕНКО М.В., к.т.н., доцент

Дніпровський державний технічний університет, м. Кам'янське

ВИКОРИСТАННЯ КОЛІРНОЇ МОДЕЛІ RGB ТА МЕТОДУ LSB ПРИ СТЕГАНОГРАФІЧНОМУ ЗАХИСТІ ІНФОРМАЦІЇ У ФАЙЛАХ ФОРМАТУ OFFICE OPEN XML

Вступ. Розглядаючи способи захисту інформації, доречно звернути увагу на стеганографічні методи. Сам термін «стеганографія» означає приховане повідомлення, яке повністю виключає можливість дізнатися про його існування третій особі. В якості сучасного прикладу можна привести випадок роздрукування контрактів з малопомітними викривленнями обрисів певних символів тексту на ЕОМ. Таким чином вносились дані про умови складання контракту, які необхідно було зашифрувати.

Комп'ютерна стеганографія ґрунтується на двох основних принципах [1]. По-перше, файли з оцифрованими зображеннями, а також аудіо- та відеофайли можна певною мірою змінити без втрати їх функціональності. По-друге, можливості людини розрізнати незначні зміни звуку або кольору досить обмежені. Стеганографічні методи дають можливість замінити несуттєві частки даних потрібною інформацією. Це означає, що сімейне фото може містити інформацію комерційного характеру, а файл з улюбленою мелодією – секретне повідомлення.

Проте найчастіше стеганографія застосовується для створення цифрових водяних знаків, які на відміну від звичайних можна виявити, лише використовуючи необхідне програмне забезпечення. Цифрові водяні знаки записуються у вигляді псевдовипадкових послідовностей сигналів шуму, які згенеровані на базі секретних ключів. Такі знаки забезпечують автентичність або недоторканість документа, дають можливість ідентифікувати власника або автора, перевірити права користувача або дистриб'ютора навіть в тому випадку, коли файл був спотворений або оброблений.

Щодо впровадження засобів програмно-технічного захисту в ІС, виділяють два головних способи:

1 – вбудований захист – механізми захисту розподілені за іншими компонентами системи або реалізуються у вигляді окремих складових ІС;

2 – додатковий захист – засоби захисту являють собою доповнення до основних