УДК 004.8(075) + 004.93(075)

А.В. Коростиленко, О.Б. Крамаренко

Інститут підготовки слідчих кадрів для СБ України
у складі Національної юридичної академії України ім. Ярослава Мудрого, Харків

МЕТОДОЛОГІЧНИЙ ПІДХІД ЩОДО РОЗПІЗНАВАННЯ МОВЛЕННЯ З ВИКОРИСТАННЯМ МОЖЛИВОСТЕЙ ПІДСИСТЕМИ РОЗУМІННЯ ВИСЛОВЛЮВАНЬ

Розглянута пропозиція використання можливостей підсистеми розуміння висловлювань, які побудовані на базі стохастичних нейроподібних мереж з ансамблевою організацією (А-мереж) і апарату активних семантичних мереж (М-мереж) з метою підвищення якості розпізнавання природної мови.

підсистема розуміння висловлювань, стохастичні нейроподібні мережі

Вступ

Постановка проблеми. Стан розвитку інформаційних ресурсів і засобів їх активізації багато в чому визначає потенційні можливості забезпечення національних інтересів країни.

Існує велике коло інформаційних та інформаційно-аналітичних задач, які при формуванні інформаційного ресурсу вирішуються, а саме: пошук потрібної інформації, її класифікація, аналіз змісту текстових та мовних джерел (у тому числі й різномовних) на сумісність, суперечливість і новизну, формування аналітичних оглядів, довідок тощо.

Рівень автоматизації обробки інформації визначається можливостями інформаційно-пошукових систем, систем автоматичного індексування й реферування текстових джерел та систем машинного перекладу, які на сьогодні перебувають в експлуатації. Віддаючи належне високому рівню теоретичних і практичних досягнень у сфері розроблення таких систем, треба зазначити, що опубліковані праці із цих питань недостатньо орієнтовані на автоматизацію обробки знань, які містяться в вербальних джерелах інформації (радіо, телебачення і т.д). Розвиток технологій обробки вербальної інформації повинен йти через розвиток методів автоматизації вилучення знань із таких джерел та їх обробки.

Системи розпізнавання мовлення характеризуються багатьма параметрами. Одним з основних параметрів є помилка розпізнавання слів (ПРС). ПРС являє собою відношення кількості нерозпізнаних слів до загальної кількості вимовлених слів.

У табл. 1 наведені порівняльні характеристики деяких дикторо-незалежних систем автоматичного розпізнавання мовлення (АРМ) [1]. З таблиці слідує, що помилка розпізнавання слів значно збільшується, якщо на стиль мовлення не накладаються обмеження або погіршуються умови введення акустичних сигналів. Складність рішення завдання розпізнавання мовлення пояснюється великою мінливістю акустичних сигналів. Ця мінливість залежить від декількох причин. По-перше, різна реалізація фонем − основних одиниць звукового строю мови. В українській мові нараховують біля чотирьох десятків фонем. Мінливість реалізації фонем викликана впливом сусідніх звуків у потоці мовлення. Загальне число алофонів української мови перевищує 9000 [2]. По-друге, положення і характеристики акустичних приймачів. По-третє, зміни параметрів мовлення диктора, які обумовлені різним емоційним станом диктора, темпом мовлення.

Таблиця 1

Порівняльні характеристики деяких дикторо-незалежних систем автоматичного розпізнавання мовлення

Система АРМ	Розмір словника, слів	Стиль мовлення	Рівень шумів	Якість каналу	ПРС, %
Інформаційна система повітряних судів	2000	Спонтанний	Низький	Широкосмуговий	2,1
Північноамериканська система ділових новин	60000	Детермінований	Низький	Широкосмуговий	6,6
Система новин радіо й телебачення	60000	Комбінований	Різний	Різне	27,1
Система розпізнавання телефонних повідомлень	23000	Спонтанний	Низький	Телефонна лінія	35,1

Формулювання мети статті. Таким чином, системи АРМ поки погано працюють в умовах введення мовних сигналів за допомогою телефонних ліній при наявності високого рівня навколишніх акустичних шумів, коли не обмежується розмір словника, і мова є злитою. Метою цієї статті є обґрун-

тування пропозиції використання можливостей підсистеми розуміння висловлювань для підвищення якості розпізнавання природної мови.

Виклад основного матеріалу

На рис. 1 зображені основні компоненти системи розпізнавання мовлення. Оцифрований мовний сигнал надходить на вхід блоку попередньої обробки, де здійснюється виділення ознак, необхідних для розпізнавання звуків. Розпізнавання звуків часто здійснюють за допомогою моделей штучних нейронних мереж (ШНМ). На етапі розпізнавання звуків у багатьох системах АРМ застосовуються алгоритми, основані на векторному квантуванні. Якщо на сегменті мовного сигналу обчислюється n ознак, то вектор ознак можна представити точкою в n-мірному просторі. Конкретний звук може бути заданий зазначенням еталонного вектору. У процесі розпізнавання вектор ознак мовного сигналу відображається в деякий номер еталона.

Пошук послідовності слів виконується в сучасних системах АРМ за допомогою акустичної, лексичної і мовної моделей [3].

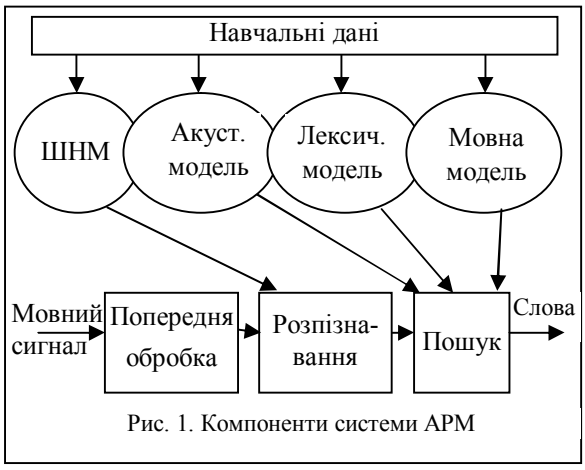

Рис. 1. Компоненти системи АРМ

Розпізнавання мови може розглядатися як задача відновлення послідовності слів W по відомому акустичному сигналу, представленому номерами векторів еталонів X.

Часто використовують статистичний підхід до пошуку рішень в умовах невизначеності, основаному на правилі Байеса. Завдання полягає в тому, щоб по сукупності номерів векторів еталонів X знайти найбільш імовірну послідовність слів W*, що забезпечує максимум апостеріорної ймовірності P(W|X).

Моделі, які використовуються в системах АРМ, не дозволяють безпосередньо обчислювати апостеріорну ймовірність P(W|X). Однак вони забезпечують обчислення ймовірності P(X|W), що являє собою апріорну ймовірність формування послідовності X при проголошенні послідовності слів W. Апостеріорна ймовірність P(W|X) може бути визначена із правила Байеса:

$$P(W|X) = \frac{P(W)P(X|W)}{P(X)} = \frac{P(W)P(X|W)}{\sum_{W_j} P(X|W_j)P(W_j)}, \quad (1)$$

де P(W) − апріорна ймовірність послідовності слів W; P(X) − апріорна ймовірність послідовності X (номера векторів еталонів).

У процесі розпізнавання P(X) не змінює своїх значень і може не прийматися до уваги при ухваленні рішення відповідно до (1). Отже, послідовність слів W*, яка максимізує P(W|X), також максимізує добуток P(W)P(X|W). Безумовна ймовірність послідовності слів P(W) обчислюється на основі моделі мови, що дозволяє відбирати в ході пошуку найбільш імовірні послідовності слів W. Умовна ймовірність P(X|W) обчислюється на основі акустичної моделі, що забезпечує відбір найбільш ймовірних варіантів звукової реалізації послідовності слів W, представленої номерами векторів еталонів X. Таким чином, формально завдання АРМ може бути сформульована в наступному виді:

$$W^* = \arg\max_W P(X|W_j)P(W_j). \quad (2)$$

Параметри мовної й акустичної моделі оцінюються за експериментальним даними з певною погрішністю. Ця погрішність може бути різною для зазначених вище моделей. Внесок кожної з моделей у результат пошуку послідовності W* регулюється за допомогою вагового коефіцієнта β (β ≤ 1) [3]:

$$W^* = \arg\max_W \{\log P(X|W) + \beta P(W)\}. \quad (3)$$

Вираз (3) може розглядатися як критерій ухвалення рішення в умовах двох експертних оцінок.

Мовна модель (ММ) забезпечує формування послідовностей слів, які потенційно можуть бути розпізнані системою АРМ. Зокрема, вона дозволяє оцінити ймовірність появи слова w_i в оточенні слів з послідовності W. ММ інтегрує в собі лінгвістичні знання, знання про предметну область і іншу інформацію з метою скорочення простору пошуку.

Так як ймовірність появи слова залежить від раніше вимовлених слів, то ймовірність послідовності слів $w_1w_2w_3\ldots w_M$ можна представити у вигляді:

$$P(w_1 \ldots w_M) = \prod_{i=1}^{M} P(w_i | w_1 \ldots w_{i-1}). \quad (4)$$

Вираз (4) дозволяє ввести поняття N-грамної моделі, відповідно до якої ймовірність появи деякого слова залежить тільки від появи N-1 попередніх слів:

$$\hat{P}(w_i | W_1^{i-1}) = P(w_i | W_{i-N+1}^{i-1}). \quad (5)$$

Визначення умовних імовірностей (5) при великих значеннях N вимагає значних обчислювальних ресурсів. Тому на практиці розглядають імовірність появи слова w_i при обмеженій довжині попереднього контексту. Найбільше часто використову-

ється біграмна модель (N = 2). У цьому випадку умовна ймовірність (5) апроксимується ймовірністю $P(w_i|w_{i-1})$. Іншими словами, імовірність появи слова w_i обумовлюється тільки попереднім словом w_{i-1} у послідовності W. Тоді спільна ймовірність появи послідовності слів W визначиться з формули:

$$P(w_1 ... w_M) = \prod_{i=1}^{M} P(w_i | w_{i-1}).$$

Істотною перевагою біграмної моделі є простота оцінювання її параметрів. Оцінка $P(w_i|w_{i-1})$ утворюється шляхом підрахунку частоти появи в навчальних даних слова w_i, якщо попереднім було слово w_{i-1}. Якщо навчальні дані не повні, то припустимим послідовностям слів можуть бути призначені нульові ймовірності.

Крім біграмної моделі, часто застосовується триграмна мовна модель, заснована на використанні умовних імовірностей виду $\hat{P}(w_i|w_{i-1},w_{i-2})$. Ця модель дає більше можливостей. Оцінка умовних імовірностей триграмних моделей виконується за допомогою формули:

$$\hat{P}(w_i | w_{i-1}, w_{i-2}) = \frac{C(i-2, i-1, i)}{C(i-2, i-1)}. \qquad (6)$$

де C(i–2, i–1, i) і C(i–2, i–1) − кількість випадків, коли спостерігалися послідовності слів (w_{i-2}, w_{i-1}, w_i) і (w_{i-2}, w_{i-1}) відповідно.

Застосування триграмних моделей вимагає значних обсягів даних для коректного оцінювання відповідних імовірностей. Для словника розміром V існує V^3 можливих триграм. Однак реальні навчальні дані можуть містити не всі триграми. З урахуванням формули (6) триграмам, які не зустрілися в навчальному наборі даних, будуть призначені нульові оцінки ймовірностей появи. Щоб уникнути цього, застосовують інтерполяційну оцінку ймовірності, що комбінує триграмную, біграмну й уніграмну оцінки:

$$P = \lambda_3 \hat{P}(w_i | w_{i-1}, w_{i-2}) + \lambda_2 \hat{P}(w_i | w_{i-1}) + \lambda_1 \hat{P}(w_i) + \lambda_0 \frac{1}{V}, \qquad (7)$$

де λ_3, λ_2, λ_1, λ_0 − коефіцієнти інтерполяції.

Біграмні й триграмні моделі повністю ігнорують лінгвістичну структуру речень. Зокрема, вони не дозволяють урахувати "далекі" синтаксичні відносини, що існують між словами речення.

З метою усунення цього недоліку у дійсний час застосовується наступний підхід [3]. Так як для рішення задачі розпізнавання мови, необхідно обчислювати ймовірність P(W|X), тобто по заданому мовному сигналу, представленому послідовністю X, відновлювати послідовність слів W, що відповідає цьому сигналу, то будується композиційна модель, що поєднує наступні моделі:

− модель мови, що представляється біграмами або триграмами та характеризується імовірностями $P(w_i | w)$ і $P(w_i | w_{i-1}, w_{i-2})$;

− модель вимови слова (лексична модель), що звичайно представляється у вигляді ланцюга Маркова й характеризується імовірністю P(фонеми|слово);

− модель фонеми, з використанням апарату прихованих марківських моделей, яка дозволяє обчислити ймовірність P(X|фонема), X − послідовність номерів векторів еталонів. Це вимагає квантування безперервних значень векторів-ознак мовного сигналу. Даний процес реалізується за допомогою векторного квантувача.

У роботі [3] показано, що використовуючи композиційну модель, можна приступити до пошуку послідовності слів W*, яка максимізує імовірність P(W|X). Однак, складність композиційної моделі зростає в міру збільшення розміру словника. Для великих словників розміри пошукових просторів виявляються величезними. Тому безпосередній пошук, оснований на перегляді всіх варіантів послідовностей слів, неефективний. Більше того, при великих розмірах словника, що характерно для розпізнавання вербальних джерел, він практично не здійсненний.

Скористаємося тим фактом, що вихід підсистеми розпізнавання мовлення служить входом підсистеми розуміння тексту (ПРТ). В [4] показано, що навіть без урахування різноманітних логік тексту (часової, просторової, каузальної й т.п.), які здатні породити інформацію, яка явно відсутня в тексті, необхідно проводити морфологічний, синтаксичний і семантичний аналіз тексту. На виході лінгвістичного процесора, що здійснює указані види аналізу, утворюється внутрішнє представлення тексту, з якими може працювати блок виводу. Використовуючи спеціальні процедури, цей блок формує відповіді. Інакше кажучи, уже розуміння на цьому рівні вимагає від ПРТ певних засобів представлення даних і виводу на цих даних.

Найбільш часто використовуваний шлях побудови систем знань полягає у тому, що класу об'єктів відразу ж приписується ім'я, що саме по собі не несе ніяких ознак об'єктів, але представляє цілий їхній клас у всіх подальших процедурах обробки інформації. Такий підхід породжує принаймні дві категорії проблем: у різноманітних випадках виявляються істотними самі різні ознаки об'єктів, а застигла класифікація, відбита в іменах класів, приводить до негнучкості всієї системи; імена класів нічого не говорять про близькість цих класів друг до друга, про наявність у них загальних властивостей або розходжень. Щораз, коли така інформація потрібна, доводиться спеціально звертатися до записів властивостей класів, що в складній системі знань породжує громіздкі пошукові процедури.

Перебороти зазначені недоліки в побудові баз знань удалося застосовуючи стохастичні нейроподібні мережі з ансамблевою організацією (А-мережі) і апарат активних семантичних мереж (М-мережі) [5]. Ці мережі були розроблені для побудови систем знаннь, у яких ім'я класу не буде відірване від ознак цього класу, і на всіх рівнях обробки інформації воно буде брати участь нарівні й одночасно з набором ознак. Основна ідея полягає у тому, що й ім'я класу, і кожний з ознак об'єкта, і взагалі будь-які функціональні елементи такої системи знань являють собою підмножини нейронів у певних нейронних полях. Ці підмножини оформляються у вигляді нейронних ансамблів, які багаті внутріансамблевими збуджуючими зв'язками й тому виступаючих при роботі нейроподібних мереж як єдине ціле. При переході на верхні рівні ієрархії такі ансамблі об'єднуються між собою, що дозволяє злити в єдине ціле нейрони-представники ім'я класу й нейрони-представники різних ознак. Сформований у такий спосіб нейронний ансамбль виступає надалі як інформаційний елемент, що несе у своїй структурі не тільки ім'я, а й окремі ознаки об'єктів. Такий елемент може сам посилати своїх представників в елементи, що формуються, більш високих рівнів.

У роботі [5] показано, що М-мережа є окремим випадком продукційних систем (ПС). М-мережа, розглянута як ПС, характеризується також тим, що база даних і база знань у ній сполучені. Дані в М-мережі представлені сукупністю збуджень її вузлів, а знання − сукупністю зв'язків між вузлами, а також функціональними характеристиками вузлів і зв'язків, зокрема вагами останніх. Для організації роботи М-мережі, яка полягає в послідовному виконанні перерахувань активності, не потрібно, таким чином, яких-небудь спеціальних засобів ведення бази даних. Що стосується інтерпретатора ПС, то в М-мережі він представлений програмами, що реалізують перерахування активностей вузлів. Ці програми досить прості і їхня структура не залежить від змісту знань (семантики вузлів) і особливостей предметної області.

Важливим елементом А и М-мереж є система посилення-гальмування (СПГ). СПГ являє собою специфічну систему, що у кожний момент часу забезпечує домінування одного інформаційного дискрету над всіма іншими. У кожний момент часу вона вибирає найбільш збуджений вузол, додатково підвищує його збудження й зменшує збудження інших вузлів (пригальмовує їх). СПГ може виділити не один, а групу близьких по величині збудження вузлів. Алгоритми СПГ такі, що збудження виділених нею вузлів поступово зменшується в часі. При цьому пропорційно розгальмовуються інші вузли. Збудження від вузлів, виділених СПГ, поширюється по мережі, викликаючи перерозподіл активності. У результаті СПГ переключається на інші вузли, і процес повторюється. Функції СПГ у роботі мережі можуть бути зіставлені з роллю уваги в процесі мислення.

Висновки

Таким чином, рівень збудження тих або інших інформаційних дискретів у будь-який момент часу залежить від попередньої історії процесу, що дозволяє побудувати процедуру уточнення коефіцієнтів інтерполяції у виразі (7). Так, якщо максимального збудження досяг вузол бази знань відповідному слову w_i, а залишкове збудження спостерігається у вузлах відповідним словам w_{i-2} і w_{i-1}, то має сенс більше довіряти триграмній оцінці: $\lambda_3 = 1$; $\lambda_2,\lambda_1 << 1$. Відповідно, якщо залишкове збудження спостерігається у вузлі відповідному слову w_{i-1}, а збудження вузла відповідному слову w_{i-2} відсутнє, то $\lambda_2 = 1$; $\lambda_3,\lambda_1 << 1$.

Запропонована процедура дозволить урахувати лінгвістичну структуру речень, "далекі" синтаксичні відносини, що існують між словами речення й підвищити якість розпізнавання спонтанного мовлення. З обліком того, що А и М-мережі належать до класу нейронних мереж, у яких функції вирішувача сполучені зі знаннями й даними, ухвалення рішення про значення коефіцієнтів потенційно буде здійснено у реальному масштабі часу.

Подальші дослідження доцільно проводити в наступних напрямках: визначення оптимальної швидкості згасання активності вузлів з урахуванням необхідності відстеження передісторії оброблюваної одиниці інформації на даний момент часу; визначення оптимальних рівнів коефіцієнтів інтерполяції в тих чи інших умовах.

Список літератури

1. Stolcke A. Linguistic Knowledge and Empirical Methods in Speech Recognition/ A. Stolcke // AI MAGAZINE. − 1997. − V. 18 − № 4 − P. 25-31.

2. Искусственный интеллект [В 3-х кн.]. − Кн.1. Системы общения и экспертные системы: Справочник / Под. ред. Э.В. Попова. − М.: Радио и связь, 1990. − 464 с.

3. Бондарев В.Н., Аде Ф.Г. Искусственный интеллект: Учеб. пособие. − Севастополь: СевНТУ, 2002. − 615 с.

4. Искуственный интеллект. − В 3-х кн. Кн. 2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. − М.: Радио и связь, 1990. − 304 с.

5.. Нейрокомпьютеры и интеллектуальные роботы / Н.М. Амосов, Т.Н. Байдык, А.Д. Гольцев, А.М. Касаткин и др. − К.: Наук. думка, 1994. − 272 с.

Надійшла до редколегії 30.05.2007

Рецензент: канд. техн. наук В.Ф. Столбов, Інститут підготовки слідчих кадрів для СБ України у складі НЮАУ ім. Ярослава Мудрого, Харків.