Прийняття рішення при розпізнаванні мови на основі нейронних мереж

О.М. КАРПОВ, Г.В. ЗІРНЄЄВА

Дніпропетровський національний університет

В статті представлена архітектура нейронної мережі для прийняття рішення при розпізнаванні мови, алгоритм сегментації мовного сигналу, алгоритм навчання та проаналізована ефективність побудованої нейронної мережі.

В статье представлена архитектура нейронной сети для принятия решения при распознавании речи, алгоритм сегментации речевого сигнала, алгоритм обучения и проанализирована эффективность построенной нейронной сети.

In article the architecture of a neural network for making-decision is presented at recognition of speech, algorithm of segmentation of a speech signal, algorithm of training and efficiency of the constructed neural network is analyzed.

Постановка проблеми. Проблема розпізнавання мови є дуже актуальною в наші дні і мало вирішеною. Особливо актуальною є проблема побудови системи штучного інтелекту за допомогою нейронних мереж. Однією з важливих наукових задач, є створення нейронної мережі, такої архітектури, щоб забезпечити найбільш високий відсоток розпізнавання мови.

Аналіз останніх досліджень і публікацій. Уперше теорія нейронних мереж як новий напрямок була позначена в роботі Маккаллока і Питтса. Великий внесок у розпізнаванні мовного сигналу за допомогою нейронних мереж внесли наступні учені: П.Е. Овчинников, Ю.А.Семин - навчання персептрона без сегментації слів з навчальної вибірки в задачі розпізнавання звуків, В. Л. Арлазаров - використання острівного нейромережевого аналізу мовного сигналу в кореляції з виділенням стійких ознак і застосуванні фонологічних і інших "інженерних" знань, О.И.Федяев розробив нейромережевий метод аналізу фонетичної структури мовного сигналу. А також розпізнаванням мовних сигналів за допомогою нейронних мереж займалися наступні учені Д.Разумихин, О.М. Карпов, П.А. Лалетин та ін.

Проблема розпізнавання мови як одне зі складових штучного інтелекту давно залучала дослідників, і на сьогоднішній день хоч і досягнуті визначені успіхи, вона залишається відкритої. Не цілком вирішеними задачами в даній області, є задача сегментації мовного сигналу і безпосередньо розпізнавання мовного сигналу. У розпізнаванні мови за допомогою нейронних мереж запропоновані архітектури не дають високого відсотка розпізнавання вхідних сигналів.

Постановка задачі. Побудувати нейронну мережу для задачі ухвалення рішення при розпізнаванні мови. Побудувати таку архітектуру зв'язків синоптичних ваг, та такий спосіб їх з'єднання, який забезпечував найкраще розпізнавання мовного сигналу. Створити програмну реалізацію нейронної мережі, алгоритму навчання і безпосередньо розпізнавання мови. Використовувати мережну архітектуру для тимчасової обробки. Проаналізувати ефективність створеної нейронної мережі. Дані про навчений нейронної мережі зберігати у форматі xml.

Основний матеріал. Для реалізації системи прийняття рішень при розпізнаванні мови на основі нейронної мережі необхідними етапами є обробка вхідного сигналу.

На першому етапі сигнал, що надійшов, нормалізуємо, відтинаємо латентні періоди, переводимо в спектральне представлення за допомогою швидкого перетворення Фур'є, далі перетворимо в спектрально-смугове представлення і забираємо шуми за допомогою фільтрації ФНЧ.

На другому етапі проводимо розбивку вхідного сигналу на фонеми за допомогою методу верифікації по сукупності параметрів [1]. Метод верифікації по сукупності параметрів полягає в наступному:

Припустимо, кожна еталонна послідовність Y_k вміщує $[y_{k0}(\omega),...,y_{ki}(\omega),...,y_{km}(\omega)]$ елементів, об'єд-наних в μ_k груп послідовності, кожна група в еталонній послідовності містить γ_j фонем, $j=1\div\mu_k$, вхідна послідовність X складається з $x_0(\omega),...,x_l(\omega),...,x_t(\omega)$ елементів, об'єднаних в θ груп послідовності, кожна група вхідної послідовності містить σ_p фонем, $p=1\div\theta$. Задача сегментації визначити межі між групами чи фонемами. Сегментація мови верифікацією на приналежність елементів $[y_{k0}(\omega),...,y_{ki}(\omega),...,y_{km}(\omega)]$; $[x_0(\omega),...,x_l(\omega),...,x_t(\omega)]$ до деякого класу с близькими параметрами відбувається по правилу

$$d=\text{extrem}\ \{x_u\}\#\{x_t\}. \qquad (1)$$

Межа визначається як

$$d_1=\max\ \{x_u\}\#\{x_{u+\Delta}\}, \qquad (2)$$

$$u=v+\delta;\ v=0,\ \Delta,\ 2\Delta\ ,...,r-2\Delta\ ;\ \delta=1\div\Delta$$

при умові

$$T_{seg}>0{,}2\text{сек}, \qquad (3)$$

де Δ– шаг верифікації(в даній роботі $\Delta=3;2;2$); # – операція зіставлення (бітового в хемінговом або десяткового в Евклідовому просторах).

$$d_{1u}=\sum_{v=0}^{r-2\Delta}\sum_{u=v}^{v+\Delta}\sum_{j=1}^{n}\left[x_u(\omega_j)-x_{u+\Delta}(\omega_j)\right]^2$$

В евклідовому просторі

$$d_{2u}=\sum_{v=0}^{r-2\Delta}\sum_{u=v}^{v+\Delta}\sum_{\delta=0}^{\Delta-1}\sum_{j=1}^{n}\left[x_u(\omega_j)-x_{u+\Delta+\delta}(\omega_j)\right]^2$$

Если $y_{ki}(\omega)$ и $x_l(\omega)$ – битовые последовательности по ω, то хеммінгово расстояние

$$d^h_{1u}=\sum_{v=0}^{r-2\Delta}\sum_{u=v}^{v+\Delta}\sum_{j=1}^{n}unc\left[x_u(\omega_j)\wedge x_{u+\Delta}(\omega_j)\right]^2,$$

$$d^h_{2u}=\sum_{v=0}^{r-2\Delta}\sum_{u=v}^{v+\Delta}\sum_{\delta=0}^{\Delta-1}\sum_{j=1}^{n}unc\left[x_u(\omega_j)\wedge x_{u+\Delta+\delta}(\omega_j)\right]^2,$$

де unc() – функция несовпадений бит операции “исключающее ИЛИ – ^”, а $x_u(\omega_j)$ – это $S(\Omega_r, t_l)$.

Верифікацію і, відповідно, сегментацію можна здійснювати на всій довжині $T_{фр}$ слова чи фрази. На рис.1. представлений результат сегментації методом верифікації по сукупності параметрів.

Далі необхідно спроектувати архітектуру нейронної мережі. Задача застосування нейронних мереж для ухвалення рішення при розпізнаванні слів мови має свої особливості в порівнянні з іншими методами останнього кроку, а саме, Марківськими процесами, динамічного програмування, принципу максимуму Понтрягіна, локальних екстремумів, для яких вибір розпізнаного образа зважується як найкраще наближення пред'явленої й еталонної реалізацій. Звичайно в цьому випадку чи обчислюється мінімальна відстань чи максимальна подоба.

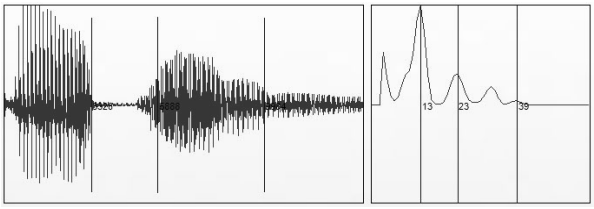

Рис. 1. Сегментація методом верифікації по сукупності параметрів. Вхідний сигнал - слово «один».

Для нейронних мереж у принципі адаптація при навчанні теж реалізується як найкраще наближення деякого опису, що відповідає одиничному стану деякого нейрона, призначеному вчителем для даного образа.

Однак у випадку мовного сигналу $s(t)$, що є функцією часу, параметричний опис також розподілений у часі. Нехай, для сигналу $s(t)$ знайдене спектрально-часове представлення $S(\omega_k, T_l)$, де ω_k-частота k-ї гармоніки перетворення Фур'є, T_l - l-й інтервал аналізу при $T_l = 12{,}5$ *Мсек*, $k = \overline{1,128}$, $l = \overline{1,L}$, $L = T_0 / T_l$, T_0 - тривалість мовного висловлення.

В [2] розглядається деяка просторово-тимчасова модель нейрона, заснована на фокусованому нейронному фільтрі і FIR-фільтрі (фільтр із кінцевою імпульсною характеристикою).

Вихід $v_j(n)$ має вид

$$v_j(n) = \sum_{i=1}^{m_0} \sum_{l=0}^{p} w_{ji}(l) x_i(n-l) + b_j ,$$

де w_{ji} – вага l-го вторинного синапса, що належить i-му первинному синапсу; $x_i(n)$ – вхідний сигнал, застосовуваний до i-му первинного синапсу в момент часу n; b_j – зсув, застосовуваний до нейрона, $v_j(n)$ – індуковане локальне поле чи нейрона аргумент функції активації φ.

Вихідна функція нейрона має вид $y_j(n) = \varphi\left(v_j(n)\right) = 1/(1+\exp(-a v_j(n)))$ для сигмоідальної функції активації. З погляду розпізнавання мовних сигналів необхідно аналізувати функцію $S(\omega_k, T_l)$ розподілену в часі. В часі функція представлена деякою сукупністю сегментів (фонем), кожен сегмент – сукупністю відліків – спектральних зрізів. Кожен спектральний зріз – вектор спектральних параметрів (128 чи 110 коефіцієнтів) чи спектрально-смугових (9 коефіцієнтів).

У даній роботі будується нейронна мережа для спектрально-полосових параметрів. Схема підключення синапсів може бути прямої й інверсної. Будемо вважати схему з'єднання нейронів у FIR-фільтрі в роботі [2] прямою. Інверсна схема припускає, що кожен спектральний зріз є варіантом образа і нейрон повинний навчаться на сукупність спектральних зрізів, що на інтервалі сегмента утворять образ. Нейронна мережа, таким чином, повинна бути багатошарової: Перший шар настроювання на спектральний зріз з вихідним значенням $v_{1j}(n)$, де n– номер спектрального зрізу, j номер сегмента (фонеми) у слові; другий шар – розподілена в часі n послідовність $v_{1j}(n)$ на сегменті j. Вихід другого шару $-v_{2j}(j*n, j*n-p_j)$, де $(j*n, j*n-p_j)$ відповідає перебору значень $v_{1j}(n)$ від $(j*n-p_j)$ до $(j*n)$ для кожного сегмента $j = \overline{1,m}$, m– кількість сегментів.

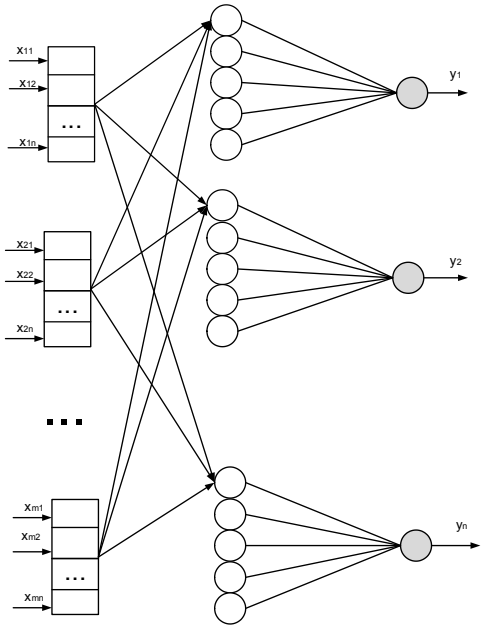

Рис. 2. Структура нейронної мережі

На рис. 2. представлена архітектура нейронної мережі, де n– кількість фонем в слові, m=9 – кількість смуг.

Для навчання нейронної мережі необхідно вибрати алгоритм навчання. Існує велике число алгоритмів навчання, орієнтованих на рішення різних задач. Серед них виділяється алгоритм зворотного поширення помилки, що є одним з найбільш успішних сучасних алгоритмів. Його основна ідея полягає в тім, що зміна ваг синапсів відбувається з урахуванням локального градієнта функції помилки. Різниця між реальними і правильними відповідями нейронної мережі, обумовленими на вихідному шарі, поширюється в зворотному напрямку – назустріч потоку сигналів. У підсумку кожен нейрон здатний визначити внесок кожної своєї ваги в сумарну помилку мережі.

Найпростіше правило навчання відповідає методу найшвидшого спуска, тобто зміни синаптичних ваг пропорційно їх внеску в загальну помилку.

Навчання нейронної мережі здійснюється послідовно за допомогою алгоритму зворотного поширення [3]. Суть алгоритму зворотного поширення полягає в наступному:

Сигнал похибки вихідного нейрона j на ітерации n (відповідає n – прикладу навчання) визначається співвідношенням:

$$e_j(n) = d_j(n) - y_j(n) \quad (4)$$

Поточне значення загальної енергії помилки обчислюється по формулі:

$$E(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (5)$$

де множина C включає всі нейрони вихідного шару мережі.

Енергія среднеквадратической помилки:

$$E_{av}(n) = \frac{1}{N} \sum_{n=1}^{N} E(n) \quad (6)$$

Метою процесу навчання є настроювання вільних параметрів мережі з метою мінімізації величини $E_{av}(n)$.

Корекція $\Delta w_{ij}(n)$, застосовувана до ваг $w_{ij}(n)$, визначається відповідно до дельта-правила:

$$\Delta w_{ij}(n) = -\eta \frac{\partial E(n)}{\partial w_{ij}(n)} \quad (7)$$

Після того як мінімізували величину $E_{av}(n)$, вводимо ідентифікатор слова і записуємо ваги мережі у файл еталона.

Збереження даних по навченої нейронної мережі організуємо у форматі xml, що дозволяє створювати універсальний файл еталона, що може бути використаний у будь-якій системі по розпізнаванню мови. У такий спосіб файл еталона зі значеннями ваг нейронної мережі буде стерпним, не залежати від платформи, на якій розробляється програмне забезпечення, не залежати від мови розробки і також використання файлу в такому форматі полегшує реконфігурацію програмного забезпечення. Приклад створеного файлу еталона представлений на рис.3.

```
<networks>
<word>
  <id>ноль</id>
  <phonems>3</phonems>
  <layer>
   <line>
    <item>1,12</item>
    …
    <item>0,49</item>
   </line>
   ….
   <line>
   …
   </line>
  </layer>
  <layer>
  ….
  </layer>
 </word>
…
<word>
 <word>
  <id>девять</id>
  <phonems>5</phonems>
  <layer>
  …
  </layer>
  <layer>
  …
  </layer>
 </word>
</networks>
```

Рис. 3. Структура файла эталона

Провівши навчання нейронної мережі на вибірці достатнього розміру, можна проводити розпізнавання.

Для розпізнавання вхідний сигнал обробляється по тім же принципі, що і при навчанні. На вхід нейронної мережі надходить спектрально-смугове представлення сигналу, з файлу еталона послідовно зчитуються вагові коефіцієнти для нейронної мережі й у залежності від близькості до 1 вхідних значень приймається рішення про результат розпізнавання. На рис.4. представлено результат роботи програми по прийняттю рішень на основі нейронної мережі. У таблиці відображені відстані до кожного з еталонів, структура нейронной мережі для слова "шість" і результат розпізнавання .

Рис. 4. Результат розпізнавання слова «шесть»

Висновок

Використання нейронної мережі запропонованої архітектури, дозволяє поліпшити якість розпізнавання вхідних сигналів, завдяки вирішенню внутрішньої задачі апроксимації. А використання формату xml для збереження даних про нейронної мережі, дозволяє використовувати навчену нейрону мережу в будь-якому іншо-

му програмному забезпеченні та полегшує його реконфігурацію.

ЛІТЕРАТУРА

1. Карпов О.Н., Габович А.Г., Марченко Б.Г., Хорошко В.А., Щербак Л.Н. Компьютерные технологии распознавания речевых сигналов. Монографія. –К.: ООО ”Полиграф-Консалтинг”, 2005 .–138 с. : ил. Парал. тит. англ.
2. Саймон Хайкин. Нейронные сети: полный курс, 2-е изд.: Пер. С англ. – М. : ООО «И.Д. Вильямс», 2006. – 1104 с.
3. Дж. Ф. Люгер Искусственный интеллект: Стратегии и методы решения сложных проблем, 4-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 804 с. : ил. Парал. тит. англ.