ЧАСОВА ТРАНСФОРМАЦІЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Юрій Рашкевич, Роман Ткаченко, Зореслава Шпак

Державний університет "Львівська політехніка"
290646, м. Львів-13, вул. Ст. Бандери, 12, тел.398-793
електронна пошта: rashkev@polynet.lviv.ua

Наведено результати використання штучних нейронних мереж для задач перетворення часового масштабу мовних сигналів. Описана структура мережі та представлені результати експериментів прогнозування зміни тривалості мовних одиниць при сповільненні темпу відтворення мовної інформації.

1. ВСТУП.

Починаючи із кінця 80-х років, штучні нейронні мережі (ШНМ) знаходять широке застосування для розв'язування багатьох типів задач оброблення мовних сигналів, включаючи задачі розпізнавання слів та фраз, верифікації дикторів, виділення ключових слів тощо. Особливо перспективним є застосування ШНМ для оброблення сигналів із змінними в часі параметрами, оскільки завдяки особливій прогностичній здатності ШНМ часто з дивовижною точністю передбачають значення сигналу, чи його параметрів.

Дуже важливою для задач регулювання темпу мови, є властивість ШНМ акумулювати та використовувати в процесі роботи інформацію про кореляційні залежності між сусідніми сегментами сигналу, тобто відслідковувати взаємовплив тривалостей сусідніх сегментів, що необхідно для збереження збалансованої темпоральної структури слова в цілому не тільки на вході моделі (що є властивим для регресійних моделей), але й на виході.

2. МОДЕЛЬ ЕКСПЕРИМЕНТУ.

Метою дослідження є встановлення здатності ШНМ відслідковувати закономірності у зміні тривалостей звуків різних класів при сповільненні темпу мови з урахуванням кореляційного впливу як тривалості попереднього звуку, так і звуку наступного. Така постановка задачі у випадку прискорення розглянута в [1], де наведені результати прогнозування тривалостей стаціонарних ділянок мовного сигналу 5-шаровою ШНМ, в якій функції перетворення входів у виходи відрізняються від сигмоїдних і зображуються сплайном Ерміта. Мовний сигнал подавався у вигляді тривалостей біжучої і двох сусідніх (попередньої і наступної) стаціонарних ділянок. Незважаючи на те, що середня похибка дещо перевищувала 20 %, отримані результати підтвердили можливість і перспективність використання ШНМ в такого типу задачах.

В наших експериментах сигнал на вході ШНМ подавався у вигляді послідовності векторів:

$$(l_{i-1}, k_{i-1}, l_i, k_i, l_{i+1}, k_{i+1}),$$

де символом l позначені тривалості біжучого, попереднього та наступного звуків, а символом k - відповідні ознаки класифікації. На основі цих даних ШНМ прогнозувала тривалість біжучого звуку в сповільненому темпі.

Мовний сигнал сегментувався на окремі класи звуків згідно із запропонованим в [2] алго-ритмом сегментації та маркірування. Виділялися 5 класів звуків - наголошені голосні, ненаголо-шені голосні, вокалізовані приголосні, невокалі-зовані приголосні, вибухові звуки, а також міжслівні паузи.

В експериментах використана гетерогенна ШНМ з проективно-латеральними синаптичними зв'язками, яка відноситься до класу мереж прямого поширення Feed Forward. Мережа забезпечує відтворення складних поверхонь на навчальній множині даних як завгодно точно, однак оптимальні прогностичні властивості моделі встановлюються відповідним вибором параметрів ШНМ на основі зовнішнього критерію якості. Загалом процес вибору по суті відповідає концепції методу групового врахування аргументів О.Г.Івахненка та здійснюється в автоматичному режимі. Процес навчання такої ШНМ є неітераційним, здійснюється за час, що не перевищує декількох секунд, а принцип балансу точності відтворення забезпечує адекватне відображення закономірностей з одночасним вилученням чисто випадкових факторів.

3.РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ.

Випробовування проводилися на мовних текстах, висловлених одним диктором у різних темпах: швидкому розмовному та сповільненому виразному. Загальний коефіцієнт зміни темпу мовлення складав 2,1. Обидва записи було просегментовано на ділянки, що відповідали звукам мови та паузам. Для кожної ділянки встановлювався її клас у відповідності із типом звуку.

Проведено дві серії експериментів: прогнозування зміни тривалостей звуків без вказування їх типів та передбачення з урахуванням класів звуків.

У першій серії при навчанні мережі на вхід ШНМ подавались трійки значень, які задавали тривалості трьох послідовних звуків у швидкому темпі, а на вихід - відповідні їм тривалості звуків при повільній вимові. Навчальна множина складалась із 100 наборів. Встановлено оптимальні для задачі такого класу параметри ШНМ: число асоціативних нейронів - 5, коефіцієнт нелінійності гіперповерхні процесу - 0,38. Результати передбачення для 25 наборів значень тривалостей звуків наступного за навчальним фрагменту мовного тексту наведені на рис. 1. Темні стовбчики відображають еталонні значення, я ясні - значення прогнозу. Середнє значення відхилення при передбаченні тривалостей звуків у випадку неврахування їх типів складало 31 мс, що відповідає 19,1% від усередненої тривалості звуків у повільному темпі. Проте, оскільки для різних звуків прогнозовані значення відрізнялись від еталонних як у сторону збільшення, так і в сторону зменшення, то загальне відхилення у передбаченні цілого фрагменту склало 7,2%.

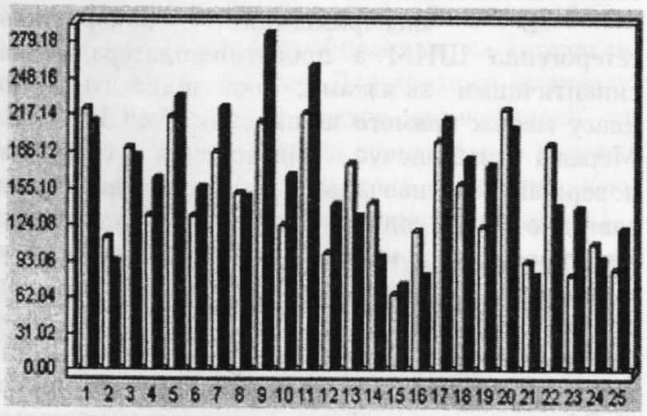

Рис.1. Прогонозування тривалостей звукових елементів без урахування ознак класифікації

На рис. 2 зображено гістограму результатів передбачення для тих самих навчальних та експериментальних наборів, але з вказанням класів звуків (на вході задавалися тривалість і клас кожного із трійки звуків). Аналіз результатів підтвердив, що введення класів звуків забезпечило вищу точність у передбаченні тривалостей як окремих звуків (середнє відхилення складало 21 мс або 12,9% від середньої тривалості звуку в еталонному тексті), так і для цілого фрагменту, використаного в експерименті (сумарна прогнозована тривалість відрізнялася від еталонної тільки на 2,4%). Тобто, з високою точністю був витриманий загальний коефіцієнт сповільнення темпу.

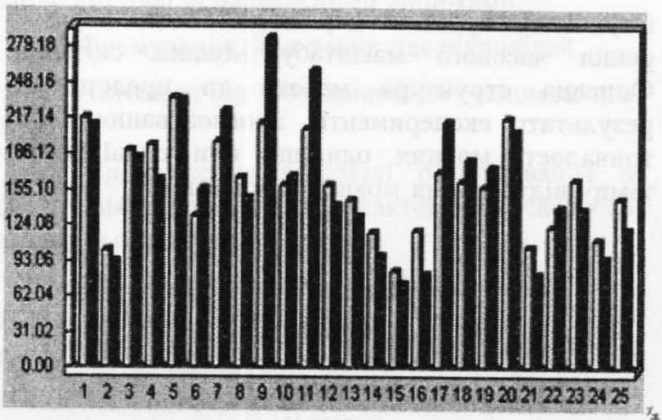

Рис.2. Прогнозування тривалостей звукових елементів з урахуванням класів звуків

При зменшенні навчальної вибірки до 50 наборів результати прогнозування різко погіршились - середнє відхилення у тривалостях звуків складало 23%, а для цілого есперимнтального фрагменту - 11,4%.

4. ВИСНОВКИ.

Отримані результати свідчать, що введення додаткової ознаки - класу звуку дозволяє суттєво підвищити прогностичні властивості нейронної мережі. Подальше покращення результатів може бути досягнуте шляхом введення в структуру ШНМ нелінійних синаптичних зв'язків. Передбачається також використання можливостей ШНМ для проведення класифікації звуків.

ЛІТЕРАТУРА.

1. Rashkevych Yu. Non-linear time-scale modification of speech by Neural Networks//Proc. of the Summer School on Neural Network Application to Signal Processing. - Czestochowa (Poland). - 1997. - P. 393-395.
2. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. - Львів: Академічний експрес, 1997. - 140 с.