

ВИБІР БАЗОВОЇ АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУВАННЯ ЧАСУ ВИКОНАННЯ ЗАДАЧ НА ОБЧИСЛЮВАЛЬНОМУ ВУЗЛІ

Водяний Р.В.

Тернопільський національний економічний університет, магістрант

Першочерговою задачею проектування нейронної мережі для прогнозування часу виконання задач на обчислювальному вузлі є вибір архітектури, яка здатна швидко навчатися і потребує малу кількість системних ресурсів.

Існують різноманітні архітектури нейронних мереж, які використовуються для вирішення тих або інших складних інженерних задач. Найбільш поширені – багатошарові перцептрони (Multilayer Perceptron – MLP) [1], які характеризуються прямим розповсюдженням вхідного сигналу від шару до шару і складаються з множини вхідних нейронних елементів, одного або декількох прихованих шарів нейронних елементів і вихідного шару. Однією з головних переваг таких мереж є можливість вирішувати задачі, що складно формалізуються, або задачі, для яких алгоритмічне рішення невідоме, але для яких можливо скласти репрезентативний набір прикладів з відомими рішеннями. MLP при навчанні, за рахунок своєї внутрішньої будови, виявляє закономірності і взаємозв'язки між вхідними і вихідними образами, узагальнюючи отриманий на навчальній вибірці досвід.

В процесі вибору архітектури нейронної мережі для прогнозування часу виконання задач на обчислювальному вузлі необхідно врахувати, що до такої системи ставиться ряд жорстких вимог, однією з яких є функціонування в режимі реального часу. В результаті цього, необхідно мінімізувати часові витрати, пов'язані з навчанням нейронної мережі.

Розглянемо характеристики MLP [1, 2]. Багатошарові перцептрони навчаються за допомогою алгоритму зворотного розповсюдження помилки (back-propagation algorithm) [1, 2] і успішно застосовуються для вирішення багатьох складних задач класифікації, розпізнавання та ін. У [2] зазначено, що на здатність нейронної мережі до коректного узагальнення впливають розмір навчальної вибірки і архітектура нейронної мережі.

Для коректного навчання нейронній мережі досить, щоб розмір навчальної вибірки [2]

$$L \approx O(W / \varepsilon),$$

де $O()$ – порядок величини, тобто, наприклад, для помилки в 5% кількість прикладів навчання повинна в 5 разів перевершувати кількість вільних параметрів мережі W ;

W – загальна кількість параметрів, що налаштовуються (вагових коефіцієнтів і порогових значень);

ε - допустима помилка класифікації.

При цьому

$$W = m(n + k + 1) + k,$$

де n – кількість вхідних нейронів; m – кількість нейронів прихованого шару; k – кількість нейронів вихідного шару.

Для вирішення задачі прогнозування часу виконання задач на обчислювальному вузлі спочатку виділяється список параметрів, від яких залежить час виконання задачі. Це такі параметри, як значення вхідних параметрів, об'єми інформації, яка оброблятиметься, завантаженість мережі передачі даних до обчислювальних вузлів, завантаження підсистем обчислювальних вузлів і т. д. Далі проводиться нормалізація вхідних даних. Для наявних даних про виконання різних задач на обчислювальному вузлі проведено моделювання прогнозування часу виконання задачі в середовищі MatLab.

Таким чином, в даній роботі для вирішення задачі прогнозування часу виконання задач на обчислювальному вузлі запропоновано вибрати нейронну мережу MLP. Отримані результати моделювання підтверджують доцільність використання штучних нейронних мереж для прогнозування часу виконання задач на обчислювальному вузлі.

Список використаних джерел

1. Головкин В.А. Нейронные сети: обучение, организация, применение / В.А. Головкин // Нейрокомпьютеры и их применение : учеб. пособие. – М., 2001. – 256 с.
2. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2006. – 1104 с.