

ПРОГРАМНИЙ МОДУЛЬ ІНТЕЛЕКТУАЛЬНОГО ЛІНГВІСТИЧНОГО ПОШУКУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ ХЕМІНГА

Вінницький національний технічний університет

Анотація

В даній роботі описується програма для інтелектуального лінгвістичного пошуку, яка може розпізнавати слова тексту, набрані із помилками. Для реалізації цієї програми використовується штучна нейронна мережа Хемінга. Зроблено тестування програмного модуля та проаналізовано доцільність використання даного підходу для розробки програми інтелектуального лінгвістичного пошуку.

Ключові слова: нейронна мережа Хемінга, штучні нейронні мережі, інтелектуальний лінгвістичний пошук, розпізнавання тексту.

Abstract

This paper describes the software for intelligent linguistic search that can recognize words of text typed with errors. To implement this program uses artificial Hamming neural network. Made software module testing and analysis of the feasibility of using this approach to develop intelligent applications linguistic search.

Keywords: neural network Hemet, artificial neural networks, intelligent linguistic search, text recognition.

На сучасному етапі розвитку інформаційних технологій і безперервним збільшенням обсягів інформації, доступної в глобальній мережі Інтернет, все більшої актуальності набувають питання ефективного пошуку і доступу до даних. Як показує практика, досить складно знайти потрібну інформацію використовуючи тільки пошук на точний збіг ключових слів, оскільки слова у запиті можуть бути з помилками або мати різні граматичні форми одного і того ж слова. Тому для більш точного лінгвістичного пошуку потрібна перевірка правильності написання слів запиту. Один із способів підвищення точності результатів лінгвістичного пошуку - застосування мережі Хемінга[1] для інтелектуального пошуку.

Найчастіше стандартний пошук з використанням ключових слів не дає бажаного результату, у зв'язку з тим, що такий підхід не враховує мовного та смислового взаємозв'язку між словами запиту. Як показує практика, досить складно знайти потрібну інформацію використовуючи тільки пошук на точний збіг ключових слів. Слова можуть зустрічатися в документах абсолютно в різних формах, що відрізняються від вихідних форм, зазначених у пошуковому запиті.

При введенні інформації, наприклад при скануванні, з паперових носіїв в базу даних надходить неминуче спотворення інформації, помилки і друкарські помилки різного ступеня тяжкості, що ускладнює подальший пошук і робить його практично неможливим. Таким чином, гостро постає проблема пошуку документів, які неточно відповідають пошуковому шаблону. Пошукова система повинна доповнювати запит, що вводиться користувачем, асоціативно зв'язаними і близькими за значенням словами.

Один із способів підвищення точності результатів пошуку - інтелектуальний пошук за подібністю з доповненням запиту синонімами ключових слів. Звичайні алгоритми пошуку виконують перевірку на точний збіг рядків. Традиційні алгоритми порівняння рядків дають тільки дві відповіді - рівні дані рядки чи ні. Алгоритми інтелектуального пошуку дають більш розширену відповідь - наскільки відрізняються рядки один від одного[2].

Однак, алгоритми інтелектуального пошуку за подібністю поки що практично не реалізовані в сучасних пошукових системах. Інтелектуальний пошук за подібністю може бути реалізований на базі різних алгоритмів і функцій пошуку. Для нечіткого пошуку можуть бути використані широко відомі

алгоритми та коди Хемінга [3]. Коди Хемінга давно і успішно застосовуються при кодуванні і декодуванні інформації, дозволяючи успішно відновити загублену при передачі інформацію.

Нейронна мережа Хемінга (Hamming) реалізує класифікатор, що базується на найменшій похибці для векторів двійкових входів, де похибка визначається відстанню Хемінга. Відстань Хеммінга визначається як число біт, що відрізняються між двома відповідними входними векторами фіксованої довжини. Один входний вектор є незашумленим прикладом образу, інший - зіпсованим чином. Вектор виходів навчальної множини є вектором класів, до яких належать образи. У режимі навчання входні вектори розподіляються за категоріями, для яких відстань між зразковими входними векторами і поточним входним вектором є мінімальним.

В подальшому для покращення процесу лінгвістичного пошуку пропонується використовувати імпульсні нейронні мережі [4]. Це покращить швидкість та результативність пошуку. Крім того, максимум переваг від імпульсних нейронних мереж можна досягти при їх апаратній реалізації [5,6]. Також вони є основою для побудови операційного блоку нейрокомп'ютерів [7], які незабаром придуть на допомогу цифровим комп'ютерам.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Руденко О.В. Штучні нейронні мережі: Навчальний посібник / О.В.Руденко, Є.В.Бодяньський. - Харків : ТОВ «Компанія СМІТ», 2006. — 404 с. - ISBN 966-8630-73-Х.
2. Хайкин Саймон Нейронные сети: полный курс, 2-е изд. / Пер. с англ. под ред. КуССуль Н. Н. – М. : Издательский дом «Вильямс», 2006. – 1104 с. – ISBN 5-8459-0890-6.
3. Бардаченко В.Ф. Перспективи застосування імпульсних нейронних мереж з таймерним представленням інформації для розпізнавання динамічних образів / Бардаченко В.Ф., Колесницький О.К., Василецький С.А. // УСІМ.-2003-№6.- С. 73-82.
4. Бардаченко В. Ф. Таймерні нейронні елементи та структури. Монографія / В. Ф. Бардаченко, О. К. Колесницький, С. А. Василецький. – Вінниця : УНІВЕРСУМ-Вінниця, 2005, 126 с. – ISBN 966-641-109-1.
5. Колесницький О. К. Аналітичний огляд апаратних реалізацій спайкових нейронних мереж / О. К. Колесницький // Математичні машини і системи. – 2015. – №1, С.3-19. ISSN 1028-9763 [Електронний ресурс]. Режим доступу http://www.immsp.kiev.ua/publications/articles/2015/2015_1/01_2015_Kolesnytsky.pdf
6. О. К. Kolesnytskyj Optoelectronic Implementation of Pulsed Neurons and Neural Networks Using Bispin-Devices / О. К. Kolesnytskyj, I. V. Bokotsey, S. S. Yaremchuk // Optical Memory & Neural Networks (Information Optics), 2010, Vol.19, №2, pp.154-165.
7. Колесницький О. К. Принципи побудови архітектури спайкових нейрокомп'ютерів / О. К. Колесницький // Вісник Вінницького політехнічного інституту. – Вінниця: УНІВЕРСУМ-Вінниця. – 2014. – №4 (115), С.70-78. [Електронний ресурс]. Режим доступу - <http://visnyk.vntu.edu.ua/article/view/3697/5416>

Павло Миколайович Семенюк – студент групи ІКН-15м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: semenukpasha93@gmail.com.

Олег Костянтинівич Колесницький – кандидат техн. наук, доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: okk_vin@ukr.net.

Pavlo M. Semeniuk – Department of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: semenukpasha93@gmail.com.

Oleh K. Kolesnytskyj – Cand. Sc. (Eng), Assistant Professor, Assistant Professor of Computer Science Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: okk_vin@ukr.net.