

Результатом роботи є проект інформаційно-аналітичної системи для обліку та прогнозування прибутковості інтернет-магазину комп'ютерної техніки, що включає БД, програмні засоби обробки даних і програмний модуль аналізу та прогнозування динаміки зміни продажів інтернет-магазину.

Реалізація запропонованих засобів дозволить підвищити оперативність і якість прийняття рішень з управління інтернет-магазином.

Список використаних джерел

- [1] Інформаційний інтернет-портал All Retail // Україна в 2020 р стала лідером по зростанню e-commerce в Східній Європі — Euromonitor International - 2020. URL: - <https://allretail.ua/news/67700-ukrajina-v-2020-r-stala-liderom-po-zrostannyu-e-commerce-v-shidniy-yevropi-euromonitor-international> (дата звернення 20.04.2021).
- [2] Abraham B., Ledolter J. Statistical methods for forecasting. – New York: Wiley, 2003. – 421 p.

УДК 004.62.77

СИСТЕМА ДЛЯ АНАЛІЗУ ВІДВІДУВАННЯ САЙТІВ

Абросімова Т. В., Пронін С.В.

Харківський національний автомобільно-дорожній університет, Харків

Розвиток Інтернету в глобальну інформаційну інфраструктуру дає звичайним користувачам можливість для поучення різноманітної інформації з багатьох джерел. Це ставить перед постачальниками контенту необхідність в постійному покращенні послуг. В результаті чого встає проблема визначення профілю користувача.

Розуміння профілю користувача важливо для отримання цінних даних, які можуть підвищити ефективність послуг, підвищити відвідуваність сайту і дати розуміння того які дані здійснюють позитивні, а які негативні наслідки.

Для ефективного вирішення цих завдань може допомогти такий напрямок в методології аналізу даних - Web Mining. Web Mining розвивається на

перетині таких дисциплін як виявлення знань в базах даних, ефективний пошук інформації, штучний інтелект, машинне навчання і обробка природних мов[1-2].

Завдання Web Mining [1-2]:

- пошук інформації;
- аналіз структури сегмента мережі;
- виявлення знань з веб-ресурсів;
- персоналізація інформації;
- пошук шаблонів в поведінці користувачів.

Рішення нашої задачі пов'язано з розв'язанням таких завдань Web Mining як персоналізації веб-простору та пошук шаблонів в поведінці користувачів.

Персоналізації веб-простору - завдання по створенню веб-систем, які адаптують свої можливості (навігація, контент, банери і інші рекламні пропозиції) під користувача на підставі зібраної та проаналізованої інформації про переваги користувача.

Метою пошуку шаблонів в поведінці користувачів є пошук закономірностей в шаблонах взаємодії користувача з веб-ресурсом з метою прогнозування його подальших дій. Аналізовані дії користувачів можуть включати не тільки переходи по посиланнях, але і відправку форм, прокрутку сторінок, додавання в обрані сторінки і т.д. Знайдені шаблони використовуються в подальшому для оптимізації структури сайту, вивчення цільової аудиторії та для прямого маркетингу.

Для аналізу даних при пошуку шаблонів користувача поведінки найчастіше використовуються наступні методики: кластеризація, асоціації, аналіз послідовностей.

Задачу аналізу відвідування сайту можливо вирішити застосовуючи методику кластеризації. Використання кластеризації дасть можливість визначити групу користувачів та їх можливі переваги. Серед методів кластеризації для рішення задачі була обрана мережа Кохонена.

Процес Web Mining підрозділяється на наступні етапи [2]:

Етапи Web Mining [2]

1. вхідний етап (input stage) - отримання "сирих" даних з джерел (логи серверів, тексти електронних документів);
2. етап попередньої обробки (preprocessing stage) - дані подаються у формі, необхідної для успішної побудови тієї чи іншої моделі;
3. етап моделювання (pattern discovery stage);
4. етап аналізу моделі (pattern analysis stage) - інтерпретація отриманих результатів.

Самоорганізовані нейронні мережі

Властивість самоорганізації є одним з найбільш привабливих властивостей нейронних мереж. Таким властивістю володіють самоорганізовані нейронні мережі, описанні фінським вченим Т. Кохоненом. Нейрони самоорганізованої мережі можуть бути навчені виявлення груп (кластерів) векторів входу, що володіють деякими загальними властивостями. При вивченні самоорганізованих нейронних мереж, або мереж Кохонена, істотно розрізняти мережі з неврегульованими нейронами, які часто називають шарами Кохонена, і мережі з упорядкуванням нейронів, які часто називають картами Кохонена. Останні відображають структуру даних таким чином, що близьким кластерам даних на карті відповідають близько розташовані нейрони [3].

Приклад на рис. 1 демонструє нейронну мережу з єдиним шаром, завдання якої полягає в тому, щоб правильно згрупувати (кластерізувати) вектори входу.

Правило навчання шару Кохонена:

1 етап: Ініціалізація параметрів мережі. Ваги мережі можна формувати до випадковими значеннями, так і випадково вибраними значеннями навчальних образів. Початкове значення параметра активності:

$$c_0 = 1/N,$$

де N - число нейронів в мережі: Зсув пов'язано з параметром активності співвідношенням:

$$b = e/c_0.$$

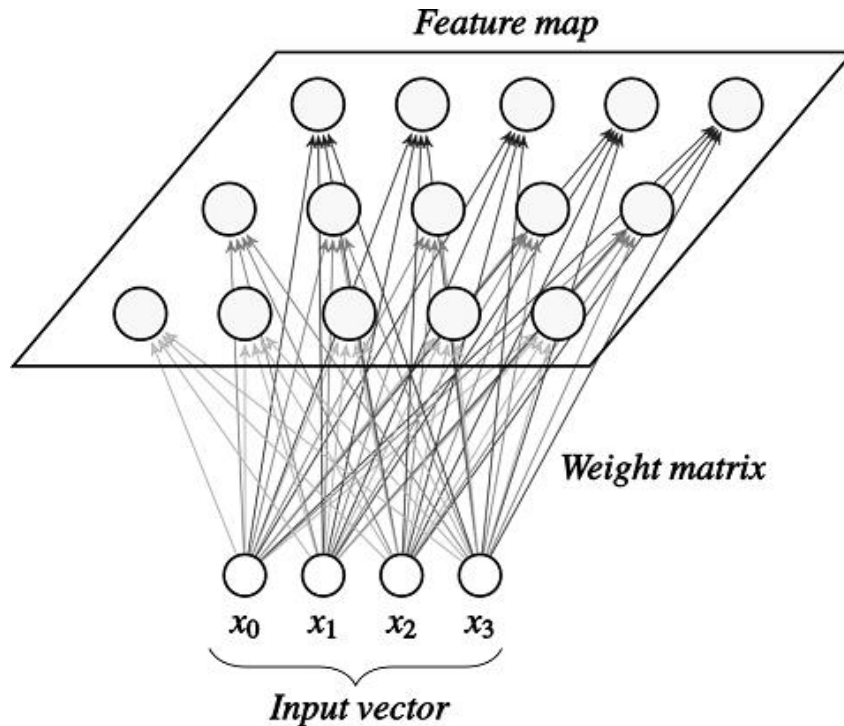


Рисунок 1 - Архітектура шару Кохонена

Завдання початкової швидкості навчання α_0 . Швидкість залежить від поточної ітерації:

$$\alpha(q) = \alpha_0/q,$$

де q номер поточної ітерації.

2 етап: Навчання мережі.

1. Вибір випадкового способу з навчальної послідовності. Пошук нейрона-переможця. Переможцем є той нейрон, у якого сума від'ємного евклидова відстані між ним і способом мислення й зміщення приймає найбільше значення.

$$i^* = \arg(\max(-\text{dist}(n_i, p_j) + b_i)),$$

де i^* - індекс переміг нейрона, p_j - вхідний образ на j -й ітерації, $\text{dist}(n_i, p_j)$ - евклидова відстань між n_i нейроном і p_j чином.

2. Для переміг нейрона коригуємо ваги:

$$W(q) = W(q-1) + \alpha(q) * (p(q) - W(q-1)),$$

де q - номер поточної ітерації, α - швидкість навчання на даній ітерації. 3.

Для всіх нейронів встановлюємо нове значення параметра активності:

$$c_i = c_{i-1} + lr * (a_i - c_{i-1}),$$

де $lr = 0.001$ параметр швидкості настройки, a_i відповідає виходу даного нейрона - 1, якщо це нейрон-переможець, 0 - в іншому випадку. Відповідно до новим значенням параметра активності, перераховуємо зміщення для кожного нейрона:

$$b = e / c_i$$

Повторюємо ітерації 1-3 стільки разів, скільки задано епох, або поки не буде досягнута точність, тобто вектори ваг будуть змінюватися менше будь-якого заданої точності.

Після навчання мережі потрібно провести її тестування. Для цього на вхід мережі подається тестовий приклад, в результаті тільки один нейрон з максимальним значенням ($\max(-\text{dist}(n_i, p_j) + b_i)$) на виході видасть 1.

Для активації «мертвих» нейронів використовується збільшення значення їх зміщення b . За рахунок цього вони вводяться в конкуруючий шар. Для них значення параметра активності зменшується, так як доданок $lr * (a_i - c_{i-1})$ завжди буде менше 0, а для переміг нейрона навпаки, $lr * (a_i - c_{i-1})$ більше 0, отже значення активності збільшується і зменшується зміщення b . Застосування випадкової вибірки навчальних образів необхідно для того, щоб переможцями ставали різні нейрони, інакше ймовірність перемоги одного і того ж нейрона дуже велика.

Список використаних джерел

- [1] Web Mining: основные понятия. [Он-лайн]. Доступно: <https://basegroup.ru/community/articles/basic-conceptions>.
- [2] Web Mining - добыча знаний из World Wide Web. [Он-лайн]. Доступно: <http://www.interface.ru/home.asp?artId=25898>.
- [3] Web Mining: анализ использования веб-ресурсов, обработка веб-лога [Он-лайн]. Доступно: <https://basegroup.ru/community/articles/web-usage-mining-part1>.
- [4] Карта самоорганизации (Self-organizing map) на TensorFlow. [Он-лайн]. Доступно: <https://habr.com/ru/post/334810/>.