

**Ю. І. ДОРОФЄЄВ**, доц., каф. САіУ, НТУ «ХП»;  
**К. С. БУЛАВІНА**, студентка, каф. САіУ, НТУ «ХП»

## АНАЛІЗ СПОЖИВЧОГО РИНКУ ЗАСОБІВ МОБІЛЬНОГО ЗВ'ЯЗКУ ЗА ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Метою дослідження є сегментація споживачів та виявлення взаємозв'язків між економіко-демографічними характеристиками респондентів та їх поведінкою на ринку. В даній роботі виконується анкетування 230 респондентів, які є споживачами засобів мобільного зв'язку. Результати анкетування проаналізовані за допомогою штучної нейронної мережі Кохонена, що самоорганізується. Отримані результати представлені візуально за допомогою метода паралельних координат.

Целью исследования является сегментация потребителей и выявление взаимосвязей между экономико-демографическими характеристиками респондентов и их поведением на рынке. В данной работе выполняется анкетирование 230 респондентов, которые являются потребителями средств мобильной связи. Результаты анкетирования проанализированы с помощью самоорганизующейся искусственной нейронной сети Кохонена. Полученные результаты представлены визуально с помощью метода параллельных координат.

The primary focus of the research concentrates on consumers and the relationship between interconnection of economic and demographic activities of the respondents and their behavior in the marketplace. This questionnaire is designed for 230 respondents who are consumers of mobile communication. The results of the questionnaire have been analyzed by using the self-organizing artificial neural network of Kohonen as the instrument. The data can be seen by referring to parallel coordinates as the visual instrument.

**Введення.** Проведення маркетингових досліджень – це складний багатоступінчастий процес, що вимагає глибокого знання об'єкта вивчення, від точності та своєчасності результатів якого багато в чому залежить успішне функціонування всього підприємства.

Ринок складається зі споживачів, що мають схожі потреби, однак сукупність покупців ніколи не буває однорідною. Споживачів занадто багато, а їхні бажання й потреби часом діаметрально протилежні. Вони відрізняються вибором товару, якому віддають перевагу, сумою грошей, яку готові заплатити за товар, і джерелами одержання інформації. Отже, метою маркетингових досліджень є сегментація споживчого ринку, щоб вибрати один або декілька сегментів для освоєння.

Ринковий сегмент – це група споживачів, що мають особливі, істотно важливі для розробки стратегії маркетингу, характеристики. На більшості ринків необхідність розробки пропозицій стосовно до конкретних сегментів очевидна, оскільки один різновид товару не здатний одночасно задовольнити попит всіх споживачів [1]. Маркетингові дослідження дозволяють прийняти рішення про випуск нового продукту та його позиціонування на ринку, про фінансування нового проекту або спрямованість рекламних акцій.

У теперішній час для здійснення аналізу великих масивів даних все частіше використовують концепції та методи штучного інтелекту, зокрема штучні нейронні мережі. Для складних недостатньо формалізованих задач застосування нейронних мереж дозволяє швидко отримати результати, які можуть значно перевершувати результати, що отримані за допомогою традиційних методів розв'язання інтелектуальних задач [2].

**Постановка задачі.** На основі опитування респондентів провести аналіз споживчого ринку засобів мобільного зв'язку за допомогою штучних нейронних мереж, тобто встановити взаємозв'язки між економіко-демографічними показниками споживачів та їх вибором і поведінкою на ринку засобів мобільного зв'язку. Перевірити придатність отриманих результатів для використання у маркетинговій діяльності.

**Отримання вхідних даних.** Для збору даних була сформована анкета для опитування респондентів та проведено опитування у місті Харкові.

В анкетуванні взяло участь 230 респондентів. Серед опитуваних були як чоловіки, так і жінки, у віці від 18 до 55 років, з різним рівнем доходу. Опитування проводилося як серед перехожих на вулиці й у метро, так і за допомогою соціальних мереж.

У перелік характеристик, які необхідно повідомити респондентові про себе, входять наступні:

- вік;
- стать;
- рід занять;
- середній місячний дохід на одного члена родини.

В процесі обробки анкет ці дані були закодовані шляхом присвоєння порядкового номеру відповіді та представлені у вигляді масиву вхідних даних  $X$  розміром  $4 \times 230$ .

Для графічного відображення результатів анкетування використано секторну діаграму, яку приведено на рис. 1.

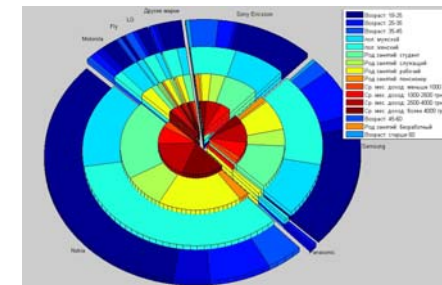


Рис. 1 – Відображення результатів анкетування

Кожен сектор діаграми відповідає одній з торгівельних марок засобів мобільного зв'язку, якими користуються респонденти на цей час. Довжина

дуги сектора вираховується з процентного відношення респондентів, які належать до цього сектора, до всіх опитуваних респондентів.

Рівень сектора описує економіко-демографічний показник опитаних респондентів. Кожен рівень має свої кольори, розбиті на сектора, які вказують на процентне відношення категорій в економіко-демографічному показнику до всієї множини цього показника. Відношення вираховується згідно обраного сектора, тобто марки телефону.

**Метод кластерного аналізу.** Кластерний аналіз є одним з найбільш відомих методів математичної статистики, що використовуються для сегментування споживачів. Він є одним з методів розвідувального аналізу даних, створених для виявлення можливих угруповань, які присутні у сукупності даних. Основним критерієм для об'єднання даних у групи є відстань: об'єкти, розташовані "близько" один до одного, повинні попадати в той самий кластер, тоді як "досить далекі" об'єкти повинні бути в різних кластерних групах. В ідеальному випадку всі об'єкти усередині кластера повинні бути досить однорідними, але значно відрізнятися від об'єктів з інших кластерів. Результатом успішно проведеної кластеризації є виділення ряду сегментів усередині загального набору даних.

Основною метою при проведенні кластерного аналізу є виділення природно виникаючих груп досліджуваних об'єктів на основі їхньої схожості. Але при практичному застосуванні конкретних методів з'являються відмінності, які пов'язані з різними способами обчислення відстані між двома спостереженнями (або кластерами) та з правилами, які використовуються для формування кластерів.

Методи кластерного аналізу можна розділити на дві групи: ієрархічні та неієрархічні. Ієрархічні алгоритми кластерного аналізу можуть бути двох типів – агломеративні та дивізімні.

Ієрархічні агломеративні методи характеризуються послідовним об'єднанням елементів і відповідним зменшенням числа кластерів. На початку роботи алгоритму всі об'єкти є окремими кластерами. На першому кроці найбільш схожі об'єкти об'єднуються в кластер. На наступних кроках об'єднання продовжується до тих пір, поки всі об'єкти не будуть складати один кластер.

Ієрархічні дивізімні методи є логічною протилежністю агломеративних методів. На початку роботи алгоритму всі об'єкти належать одному кластеру, який на наступних кроках ділиться на менші кластери, в результаті утворюється послідовність розщеплених груп.

Неієрархічні методи не вимагають, щоб об'єкти, що потрапили в кластер, залишалися в цьому ж кластері протягом усього подальшого процесу кластеризації. Одним з найпоширеніших методів неієрархічної кластеризації є алгоритм  $k$ -середніх. При його використанні дослідник повинен попередньо задати необхідне число кластерів  $k$ , і робота алгоритму приведе до виявлення в даних саме заданого числа кластерів.

Результати кластеризації даних можуть бути зображені у вигляді дендрограми. Дендрограма описує близькість окремих точок та кластерів один до одного та графічно зображує послідовність об'єднання (роз'єднання) кластерів (рис. 2).

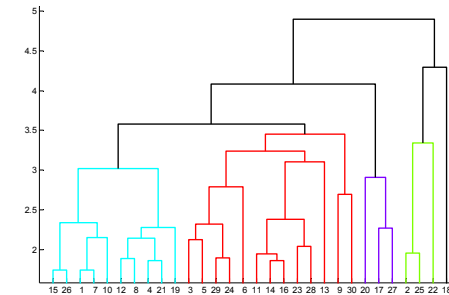


Рис. 2 – Зображення результатів кластерного аналізу у вигляді дендрограми

На рисунку можна виділити п'ять груп взаємозв'язаних показників, що представляються єдиною точкою п'ятивимірного простору: стать, вік, рід занять, середній місячний дохід та марка мобільного телефону, якою користується респондент на даний момент. Причому не можливо виявити внутрішнього взаємозв'язку між характеристиками об'єкта та виявити їх вплив на вибір марки телефону, якою зараз користується респондент. Тому для подальшого аналізу обрано штучні нейронні мережі Кохонена, що самоорганізуються.

**Експериментальний аналіз ринку засобів мобільного зв'язку за допомогою штучних нейронних мереж.** Штучна нейронна мережа Кохонена являє собою набір впорядкованих нейронів, навчання яких здійснюється «без вчителя» шляхом самоорганізації.

Основною метою шару Кохонена є співвіднесення векторів-сигналів, що подаються на нейронну мережу, з вихідним результатом кластеризації, здійснюючи таким чином розв'язання задачі кластеризації і розпізнавання образів. Це досягається за допомогою такого підстроювання ваг нейронів, при якому близькі вхідні вектори активують один і той же нейрон.

На рис. 3 зображена загальна архітектура мережі Кохонена.

Вектор входів  $X$ , розмірністю  $N \times 1$ , містить в собі параметри кластеризації. Матриця  $W$  – вагова матриця розмірністю  $K \times N$ ,  $i$ -й рядок якої містить ваги  $i$ -го нейрона, а вектор  $W_0$  розмірністю  $K \times 1$  – вектор зміщень нейронів. Вони представляють собою параметри мережі Кохонена, що налаштовуються. Блок  $S$  описує обчислення відстані між векторами ваг нейронів та поданим на вхід вектором  $X$ . Відстань між нейронами обчислюється згідно з заданою функцією: евклідова, принцип квадратних

ансамблів, принцип ланцюга, відстань Манхетена.  $F$  – активаційна функція нейронів.  $Y$  – вектор виходів мережі, який представляє результат кластеризації.

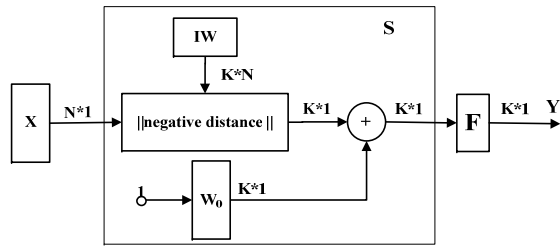


Рис. 3 – Загальна архітектура мережі Кохонена

Алгоритм навчання мережі Кохонена передбачає самонавчання по правилу “переможець забирає все”. Ваговим коефіцієнтам мережі перед початком навчання треба присвоїти початкові значення. Загально прийнятою практикою при роботі з нейронними мережами є присвоєння вагам невеликих випадкових значень. При навчанні шару Кохонена випадково обрані вагові вектори варто нормалізувати. Остаточні значення вагових векторів після навчання збігаються з нормованими вхідними векторами. Тому нормалізація перед початком навчання наближає вагові вектори до їх остаточних значень, скорочуючи, таким чином, процес навчання [3].

Після навчання така нейронна мережа візуально відображає багатомірні вхідні дані на площині нейронів. Таке зображення даних дає можливість побачити наявність або відсутність взаємозв'язків у вхідних даних. Перед мережею, що навчається «без учителя» ставиться мета виявити скриті закономірності, що присутні в наборі вхідних даних.

Для формування нейронної мережі Кохонена необхідно визначити кількість кластерів, на які розбиваються вхідні дані. В якості даних, що підлягають кластеризації, виступають характеристики респондента: вік, стать, рід занять, середній місячний дохід на одного члена родини, та відповідь на чергове запитання про пристрасті та переваги респондента.

Після формування архітектури нейронної мережі необхідно виконати ініціалізацію її параметрів та здійснити процес навчання. Розташування кожного нейрона у просторі ознак характеризується значеннями нейронних ваг. За допомогою нейронної мережі, навчання якої завершено, здійснюється кластеризація всіх векторів з вибірки. При цьому кожний вектор порівнюється з вектором ваг нейрона, який є центром кластера, що являє собою деяке узагальнення властивостей, характерних певної області багатомірних даних.

Для візуалізації результатів використовується метод паралельних координат, згідно з яким будується декілька паралельних горизонтальних осей, які перетинають одну вертикальну вісь координат. Для кожного кластера на осях відмічаються значення ознак, які є для нього типовими, потім отримані точки з'єднуються. Отриманий набір ліній вказує на зв'язок між ознаками.

Результати експериментів показали, що торгівельною маркою Nokia користуються в основному чоловіки, службовці з середнім місячним доходом в 2500–4000 грн у віці 18–35 років; торгівельною маркою LG в основному користуються робітники з середнім місячним доходом в 1000–2500 грн; торговій марці Motorola віддають перевагу жінки віком 25–35 років. Дані результати приведені на рис. 4.

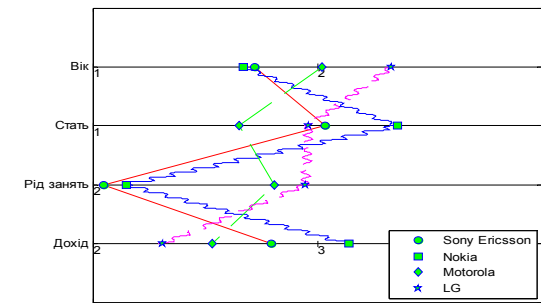


Рис. 4 – Результат візуалізації результатів методом паралельних координат

**Висновки.** В роботі запропоновано методику автоматизації процесу аналізу великих масивів вихідних даних за допомогою штучних нейронних мереж Кохонена, що самоорганізуються. Запропоновану методику застосовано для розв'язання задачі кластеризації та послідуочної сегментації користувачів мобільного зв'язку. Описаний підхід може бути застосований при прийнятті маркетингових рішень, наприклад, про випуск нового продукту та його позиціонування на ринку, про фінансування нового проекту або спрямованість рекламних акцій.

**Список літератури:** 1. Котлер Ф. Основы маркетинга / Ф. Котлер. – М. : «Прогресс», 1991. – 796 с. 2. Данько Т. П. Системы искусственного интеллекта в разработке корпоративных маркетинговых стратегий / Т. П. Данько, М. А. Ходимчук. – М. : «Финпресс», 2001. – 255 с. 3. Гвидо Д. Анализ финансовых данных с помощью самоорганизующихся карт / Д. Гвидо, Т. Кохонен. – М. : «Альпина Бизнес Букс», 2001. – 317 с.

Надійшла до редколегії 20.05.2011