

# Інтелектуальний аналіз для оцінювання науковців з нейротематики

Володимир Красиленко, Діана Нікітович, Ольга Касіяненко

Кафедра журналістики та інформаційних технологій, Вінницький інститут університету “Україна”, УКРАЇНА, м. Вінниця, вул. Першотравнева, 8  
E-mail: krasilenko@mail.ru; Nikitovitch.Diana@yandex.ru

*The article presents the results of cluster analysis of scientometric indicators in Google Scholar for the world's best scientists in Neural Nets.*

Ключові слова – інтегральне рейтингове оцінювання науковців, наукометрія, інтелектуальні технології, кластеризація, регресійна модель, нейропакети.

## Вступ

Рейтинги науковців визначаються на основі бібліометричних показників наукометричних баз Scopus, Google Scholar та інші [1]. Проте в рейтингових таблицях місця ранжуються на основі ієрархічного методу та окремих показників. Ми ж пропонуємо здійснювати багатоаспектне кластерне групування, визначити значущість показників і лише потім визначити інтегральну оцінку та рейтинг [2]. Для такого групування та інтегрального оцінювання ми використовуємо інтелектуальні технології [2]. **Метою роботи** є демонстрація ефективності використання нейропакетів для інтелектуального кластерного аналізу та рейтингового оцінювання діяльності науковців.

## Виклад матеріалу та результатів дослідження

Для модельних експериментів нами була створена вибірка статистичних даних з наукометричної бази даних Google Scholar стосовно показників 400 кращих науковців з нейротематики [1]. Незалежними змінними ми обрали показники: кількість публікацій, загальна кількість цитувань, кількість цитувань Since 2010, h-індекс, h-індекс Since 2010, i10-index, i10-index Since 2010, кількість років публікації, середня кількість статей за рік, середня кількість цитувань на статтю, кількість публікацій за 5 років, середня кількість статей за рік за 5 років. Нижче ми наводимо лише основні результати дослідження, а більш детальний опис процесів та отриманих результатів буде зроблено в доповіді. Оптимальним є поділ вибірки на 11 кластерів. Результати кластеризації 400 науковців картою Кохонена нейропакету Excel Neural Package дозволяють зробити наступні висновки. Серед 11 кластерів виділяються три кластери, які відповідають явним лідерам. Наприклад, найкращі показники мають 4 науковці, усереднені показники яких такі: кількість публікацій - 644, кількість цитувань нараховує декілька тисяч,

індекс Гірша становить - 92,4, мають значний (29) стаж та публікували щорічно приблизно 25 робіт, нарощуючи темпи за останні 5 років. До 2-го кластеру, який характеризується нижчими показниками і оцінений нами загальною оцінкою в 10 балів, потрапило 18 науковців, але у порівнянні з кластером A1 h-індекс знизився для цієї групи до 46,7, кількість публікацій зменшилась до 378, кількість цитувань знизилась у 4,3 рази. Аналогічно можна детально проаналізувати центроїдні показники кожної групи науковців та пояснити відмінності і причини визначення результуючої оцінки. Для визначення впливовості змінних ми використовували програми NeuroPro та Winnet 3.0. Крім того, значущість змінних визначалась для різних модифікованих нейромоделей шляхом зміни їх структури та параметрів. Експерименти показали, що видалення найменш впливових 10-ої, 11-ої змінних, спрощення моделей та бінаризація зв'язків приводять до зміни порядку ранжування показників. Зміни середньої та максимальної похибки для всіх таких експериментів були незначними. Результати визначення впливовості змінних різними підпрограмами показали розбіжності. Побудова на основі тієї ж вибірки моделі множинної лінійної регресії (МЛР) показала, що між незалежними змінними моделі існує високий рівень кореляції, тобто мультиколінеарність. Найбільшу кореляцію з вихідною оцінкою має 4-ий фактор, трохи меншу - 5-ий, потім - 6-ий, далі -1-ий, 7-ий та інші. Отримані похибки оцінок МЛР: максимальне відхилення значення оцінки регресійної моделі від реального значення склало 2,905, середня похибка склала 19,51%, а дисперсія оцінок склала 0,3141. Експерименти показали, що у всіх виконаних експериментах програма виділяє 2-ий фактор, як найбільш значимий. Другим по значимості є 3-ій фактор, оскільки в трьох випадках із чотирьох експериментів він зайняв два рази перше і один раз друге місце.

## Висновок

Результати моделювання показали, що нейропакети є ефективним інструментом і дає менші похибки оцінювання (середня похибка не перевищувала 8 %) в порівнянні з регресійними моделями.

## Література

1. Google Scholar [Електронний ресурс]. - Режим доступу: [http://scholar.google.com/citations?view\\_op=search\\_authors&hl=en&mauthors=1&label=neural\\_networks](http://scholar.google.com/citations?view_op=search_authors&hl=en&mauthors=1&label=neural_networks)
2. Красиленко В.Г. Інтелектуальний кластерний аналіз наукометричних показників наукової діяльності кращих науковців планети з нейротематика за даними Google Scholar / В.Г. Красиленко, Д.В. Нікітович // Актуальні питання документознавства та інформаційної діяльності: теорії та інновації: Матеріали І МПК. – Одеса, ОНПУ, 2015. - С. 275-280.