

МЕТОДИ КЛАСТЕРНОГО АНАЛІЗУ У СПРИЙНЯТТІ СТУДЕНТАМИ НАВЧАЛЬНОГО МАТЕРІАЛУ ДИСТАНЦІЙНОГО КУРСУ

Вступ. В загальному випадку, більшість систем дистанційного навчання (СДН) спрямовані на подачу навчального матеріалу (НМ) студентам дистанційного курсу навчання (ДКН) за заздалегідь визначеним принципом без урахування індивідуального підходу. Отже, актуальним є провести кластерний аналіз сприйняття НМ студентами ДКН, які відрізняються системами індивідуального підходу.

Постановка задачі. Постановка задачі: поставимо задачу дослідити особливості сприйняття НМ ДКН за допомогою існуючих методів кластерного аналізу.

Постановка експерименту. Керівниками експерименту були розроблені два курси дистанційного навчання для тієї ж дисципліни. Для обох курсів використовувалися ті самі вихідні дані, тобто, обсяг, суть і якість навчального матеріалу ідентичні. Відмінність становить подача матеріалу. Подачу навчального матеріалу у першому випадку можна віднести до так званого, традиційного підходу. Другий варіант курсу розроблений у запропонованій адаптивній системі дистанційного навчання (АСДН) з використанням принципів непрямого оцінювання (НО), тобто, з використанням елементів індивідуального підходу до сприйняття НМ студентами ДКН. Для контролю були розроблені тести.

Непряме оцінювання побудовано на шкалі, яка складається з дійсних значень на противагу дискретній шкалі. Це дозволить оцінювати студентів з більш високою точністю. Даний експеримент базується на припущенні, що збільшиться кластер встигаючих студентів (відмінників і тих хто вчиться на «добре») запропонованої АСДН у порівнянні з результатами тестування студентів, що навчалися та здавали тести в інших СДН.

Збір даних. На етапі відбору вибірки для кластеризації було відібрано 100 студентів. Усі відібрані студенти згідно з навчальним навантаженням вивчали дисципліну і пройшли тестування на виявлення залишкових знань. Половина, тобто, 50 студентів повторили навчальний матеріал і здавали тести в запропонованій АСДН. Інша половина – у так званій, СДН з традиційним підходом до подачі навчального матеріалу [1].

На першому етапі масивом реальних даних є оцінки, зібрані в середовищі запропонованої АСДН за допомогою спеціально розробленого програмного забезпечення клієнт-серверного типу – Deloss. Зовнішній вигляд Deloss та сформований у його середовищі список студентів можна побачити на рисунку 1.

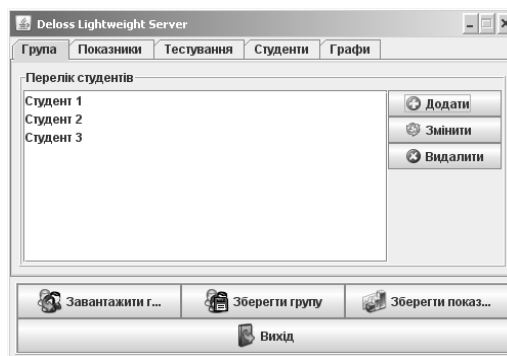


Рис. 1. Формування списків студентів у середовищі Deloss

Перед проходженням курсу студенту пропонується простий еталонний текст. Після прочитання тексту дається кілька нетривіальних запитань, відповіді на які передавали б

його сутність [2, с. 3]. Зовнішній вигляд еталонного тексту у вікні браузера наведено на рисунку 3.

Після прочитання еталонного тексту передбачено кнопку «Текст прочитано» (рисунок 2), яку студент натискає тоді, коли прочитав матеріал.

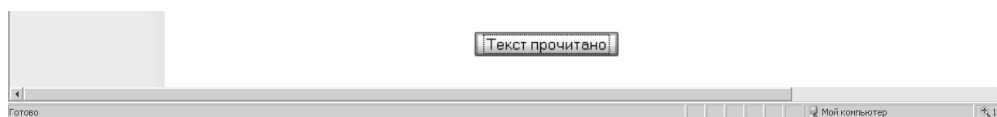


Рис. 2. Зовнішній вигляд кнопки «Текст прочитано»

Фіксується час, який він витратив на сприйняття даної порції навчального матеріалу,

а також, обсяг тексту $V_{\text{ет тексту}}$.

Зовнішній вигляд еталонного тесту у вікні браузера наведено на рисунку 3.

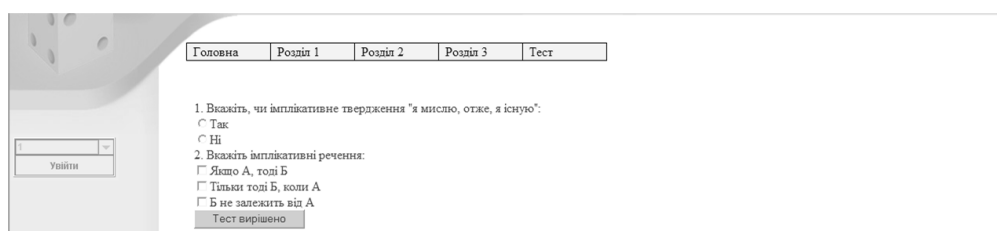


Рис. 3. Зовнішній вигляд еталонного тесту

Результати здачі $t_{\text{ет тексту}}$ (сек.), $t_{\text{тесту}}$ (сек.) та $V_{\text{ет тексту}}$ фіксуються у розробленому додатковому клієнт-серверному програмному забезпеченні «Deloss». ПЗ «Deloss» також

обчислює значення $t_{\text{тесту}_n}$, $t_{\text{тесту}_n}$ та $V_{\text{тесту}_n}$ і оцінює важкість сприйняття студентом НМ ДКН. Важкість сприйняття тексту – це оцінка, яку в академічному навчанні викладач робить інтуїтивно [3].

Таблиця із зібраними даними містить непрямі оцінки, що виставляються запропонованою АСДН та враховує значення прямої оцінки.

На другому етапі експерименту масивом реальних даних є оцінки, зібрані в середовищі СДН з традиційною подачею навчального матеріалу.

Отже, необхідні для експерименту статистичні дані зібрано. Оскільки зібраних даних для аналізу класичними статистичними методами замало, використовуємо кластерний аналіз.

Кластерний аналіз. Як науковий напрямок кластерний аналіз почав розвиватись ще в середині 60-х років і з тих пір активно розвивається, являючись однією із гілок найбільш інтенсивного росту статистичної науки. Перевагою даного методу є те, що він може працювати навіть тоді, коли даних мало і не виконуються вимоги класичних методів статистичного аналізу. Основною метою кластерного аналізу є виділення в вихідних багатомірних даних таких однорідних підмножин, щоб об'єкти всередині груп були схожі, а об'єкти з різних груп – не схожі [4, с.5].

Методи кластерного аналізу можна розділити на дві групи:

- ієрархічні;
- неієрархічні.

Кожна з цих груп містить багато підходів та алгоритмів. Використовуючи різні методи кластерного аналізу, аналітик може отримати різні результати для тих самих даних. Це вважається нормальним явищем.

Незважаючи на відмінності в цілях, типах даних і застосованих методах, усі дослідження, що використовують кластерний аналіз, використовують наступні п'ять основних кроків [4, с.5]:

1) відбір вибірки для кластеризації;

У нашому випадку вибірку складають 100 осіб, що проходили тестування. Причому, 50 осіб проходили тестування в запропонованій АСДН, 50 осіб – в традиційній СДН.

2) визначення множини ознак, за якими оцінюватимуться об'єкти у вибірці;

Для традиційного підходу множиною ознак є пряма оцінка, отримана за тестування та час проходження тесту; а для запропонованої АСДН множинами ознак будуть: пряма оцінка – *ПО*, непряма оцінка – *НО*, число повернень до якогось вузла у курсі – $N_{нов}$, шлях

студента у ДК – W , швидкість прочитання еталонного тексту $t_{етексту}^{ет}$, швидкість складання

еталонного тесту $t_{тесту}^{ет}$, швидкість прочитання n -ного тексту $t_{тексту_n}$, швидкість

складання n -ного тесту $t_{тесту_n}$.

3) обчислення значень тієї або іншої міри подібності (близькості);

При аналізі існуючих мір подібності було обрано евклідову відстань, оскільки вона найбільш чітко виражає кореляційну залежність.

4) використання методу кластерного аналізу для створення груп схожих об'єктів;

З метою більш зрозумілого та наглядного представлення результатів аналізу отриманих даних було обрано один з ієрархічних методів кластерного аналізу.

5) перевірка достовірності результатів кластерного рішення.

Перевірку достовірності результатів аналізу являє собою спростування або підтвердження висунутої гіпотези.

Розглянемо більш детально ієрархічні методи, а саме агломеративний метод кластерного аналізу. Прикладом агломеративної кластеризації є дерево ієрархічної структури або дендрограма (з грецької мови dendron – дерево), яка відображує процес агломерації, злиття окремих спостережень в єдиний кінцевий кластер. Дендрограма показує ступінь близькості окремих об'єктів кластера, а також наглядно демонструє в графічному вигляді послідовність їх об'єднання.

Алгоритм створення дендрограми:

1) на початковому етапі алгоритму кожен об'єкт вважається окремим кластером;

2) визначається матриця відстаней між об'єктами (найбільш поширений спосіб визначення відстаней – евклідова відстань);

3) об'єднуються два ближчих об'єкти, які створюють новий кластер;

4) визначаються відстані від цього кластера до всіх інших об'єктів, і розмірність матриці відстаней скорочується на 1.

5) процедура повторюється доки усі об'єкти не об'єднаються в один кластер.

Кластерний аналіз дозволяє обраховувати різні типи відстаней (евклідова відстань, манхетенська, Чебишева і т.д.). Евклідова відстань – це найбільш загальний тип відстаней, вона являється геометричною відстанню в багатомірному просторі і обраховується наступним чином (1) [6]:

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}, \quad (1)$$

де x, y – об'єкти тестування в запропонованій АСДН.

Отже, за допомогою математичного пакету «Статистика», проаналізуємо отримані дані та побудуємо дендрограми розподілу студентів на кластери по результатам тестування засобами запропонованої АСДН та засобами традиційних СДН.

Аналіз отриманих даних. Кластерний аналіз проводився за допомогою математичного пакету «STATISTICA v.7» на експериментальних даних, що були зібрані в результаті тестування студентів.

На початковому етапі аналізу кожен студент-об'єкт вважається окремим кластером. Для традиційного підходу при подачі НМ ознаками об'єкту є пряма оцінка та час проходження тесту. При використанні запропонованої АСДН існує шість ознак об'єкта, а

саме: $PO, HO, N_{нов}, W, t_{\text{тексту}}, t_{\text{тесту}}$.

Оскільки ознаки об'єктів виміряні в різних одиницях виміру, то необхідно провести нормування даних за формулою:

$$n = \frac{x_{ij}}{x_{j \max}}$$

де x_{ij} – це значення самої ознаки, а $x_{j \max}$ – максимальне значення цієї ж ознаки.

Тоді n_{ij} – це нормоване значення ознаки об'єкта, де i – кількість ознак об'єкта, j – множина об'єктів.

Обрахунок міри близькості між об'єктами виконані на основі евклідової метрики, і в результаті чого були отримані матриці відстаней (рисунки 4 та 5).

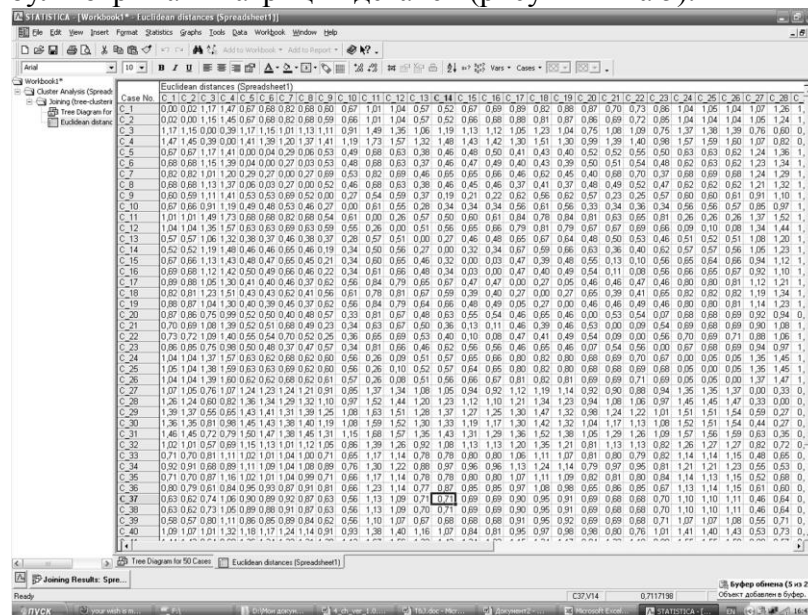


Рис. 4. Приклад визначення матриці евклідових відстаней за результатами тестування засобами запропонованої АСДН

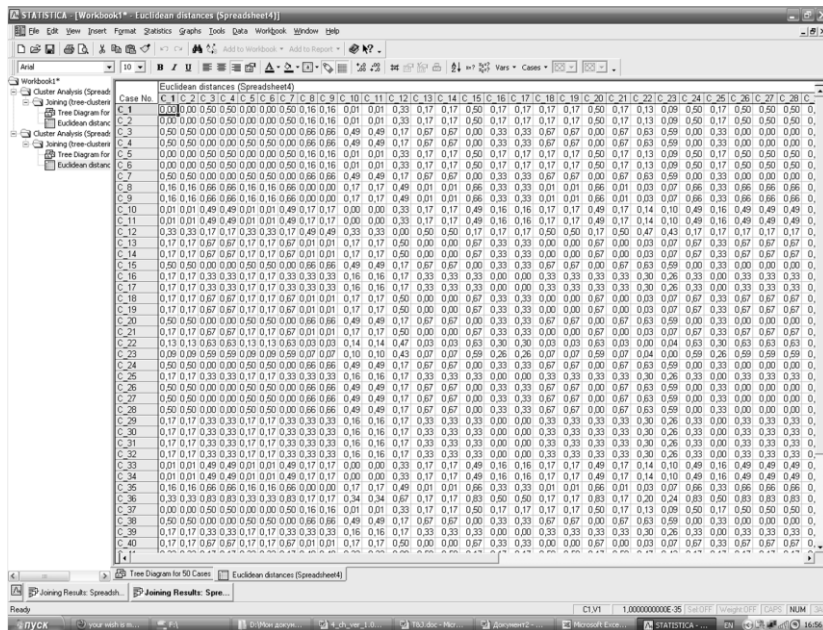


Рис. 5. Наочний приклад визначення матриці евклідових відстаней за результатами тестування засобами традиційної СДН

За допомогою матриць відстаней будемо дендрограми, по кожному із видів тестувань: засобами АСДН, що розробляється (рисунок 6) та засобами традиційних СДН (рис. 7).

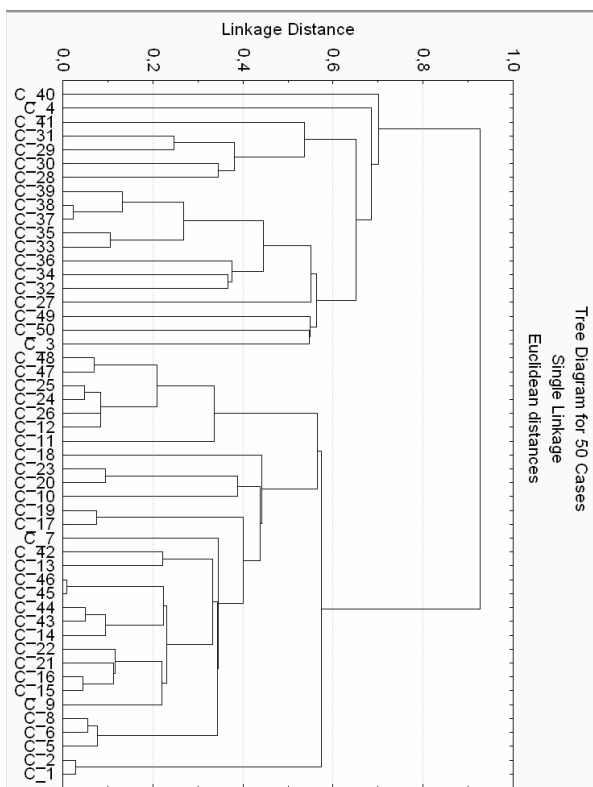


Рис. 6. Дендрограма розподілу студентів на кластери за результатами тестування засобами запропонованої АСДН

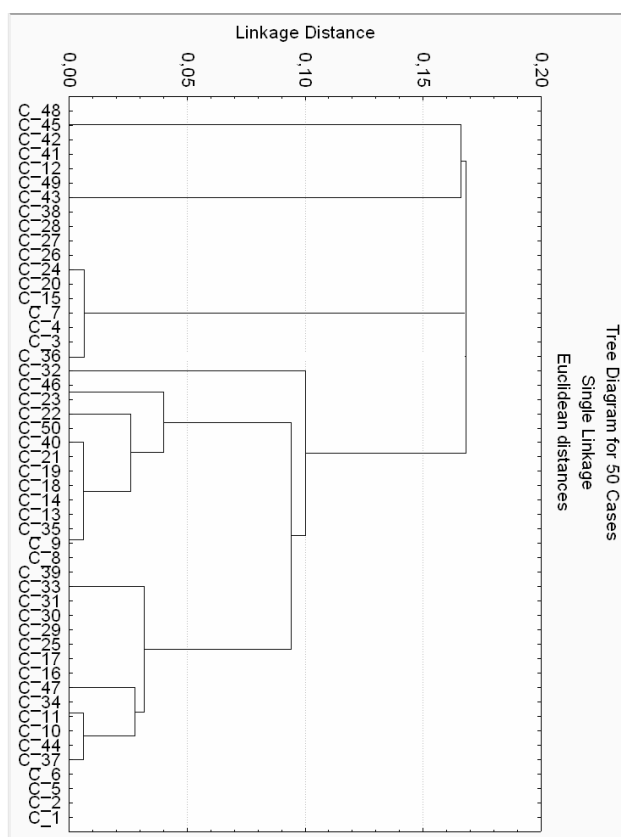


Рис. 7. Дендрограма розподілу студентів на кластери за результатами тестування засобами традиційних СДН

У результаті проведення експерименту, на дендрограмах можна побачити утворені кластери студентів за рівнем успішності здачі тесту.

Зокрема, на рисунку 6 можна виділити кластер студентів – 12 осіб, що здали тестування на «відмінно»: C_3, C_49-C_50, C_27, C_32-C_39; та кластер студентів – 38 осіб, що вчаться на «добре» (на «добре» 14 осіб – C_28-C_31, C_40-C_41, C_4, C_11-C_12, C_24-C_26, C_47-C_48 та на «задовільно» 24 особи – C_1-C_2, C_5-C_10, C_12-C_23, C_43-C_46). Ключовим результатом є те, що зник кластер невстигаючих студентів.

Розглянувши дендрограму з тестування засобами традиційних СДН (рис. 7), можна зробити висновок, що кластер студентів – 11 осіб, які не встигають, є присутнім (C_26-C_29, C_41-C_43, C_45, C_38, C_12, C_49), а кластер відмінників (у даному випадку кластер складається з сімох осіб – C_24, C_20, C_15, C_7, C_3, C_4, C_36) є вдвічі меншим за чисельністю у порівнянні з результатами, що представлені на рисунку 6. Кластер хорошистів складає 32 особи (на «добре» склали тест 14 осіб – C_8-C_9, C_13-C_14, C_35, C_18-C_19, C_21-C_23, C_46, C_32 та 18 студентів що отримали оцінку «задовільно» – C_1-C_2, C_5-C_6, C_37, C_44, C_10-C_11, C_34, C_47, C_16-C_17, C_25, C_29-C_31, C_33, C_39).

Отже гіпотеза підтвердилась, так як на дендрограмі за результатами тестування засобами запропонованої АСДН кластер студентів, що склали тестування на «відмінно» змінився з 14% при традиційному підході до 28% засобами АСДН, а також, в АСДН та зник кластер невстигаючих студентів.

Висновки. У результаті проведеного експерименту було виявлено, що подача навчального матеріалу в запропонованій адаптивній системі дистанційного навчання є ефективнішою, оскільки за результатами тестування засобами запропонованої АСДН кластер успішних студентів збільшився та зник кластер невстигаючих студентів у порівнянні з традиційним підходом.

Література:

1. Кондратенко Н.Р., Трояновська Т.І., Слободянюк Ю.Я. Дослідження особливостей сприйняття студентами дистанційного курсу навчання на базі кластерного аналізу // Збірник праць другої міжнародної конференції «Нові інформаційні технології в освіті для всіх: стан та перспективи розвитку», Київ, 2009, с. – 249-254.
2. Гороховський О.І., Трояновська Т.І. Автоматизування роботи викладача в дистанційній формі навчання // Международная научно-техническая конференция «ИОН-2006». – Материалы международной НТК – С. 142-145.
3. Гороховський О.І., Трояновська Т.І., Кисюк Д.В. Автоматизація роботи викладача дистанційної форми навчання за допомогою непрямих оцінок // «Наукові дослідження – теорія та експеримент 2007», Полтава, 14-16 травня, 2007. Полтава, «Інтерграфіка» 2007. С. – 127-131.
4. Гороховський О.І., Трояновська Т.І. Аспекти створення, моделювання та практики автоматизованих систем дистанційного навчання // «Автоматика-2006» XII МК, ВІННИЦЯ, 25-28 вересня 2006р., Вінниця. В.:ВНТУ, 2006. С. – 453
5. Кондратенко Н.Р., Трояновська Т.І., Слободянюк Ю.Я. Дослідження особливостей сприйняття студентами дистанційного курсу навчання на базі кластерного аналізу. // Вісник Хмельницького національного університету, 2009, №4. Хмельницький, ХНУ. С. 249-253.
6. Дюран Н. Кластерный анализ / Дюран Н. , Оделл П. – М.: Статистика, 1977. –128 с.

Дана стаття присвячена кластерному аналізу даних, отриманих у результаті дослідження сприйняття студентами навчального матеріалу дистанційного курсу навчання.

Ключові слова: *еталонний текст, еталонний тест, непрямі оцінки, адаптивна система дистанційного навчання, дистанційний курс, навчальний матеріал, кластерний аналіз.*

Данная статья посвящена кластерному анализу данных, полученных в результате изучения восприятия студентами подачи учебного материала в дистанционном курсе обучения.

Ключевые слова: *эталонный текст, эталонный тест, непрямые оценки, адаптивная система дистанционного обучения, дистанционный курс, учебный материал, кластерный анализ.*

This article is devoted to cluster analysis of data received during learning by the students the distant course educational material

Keywords: *etalon text, etalon test, non-straight marks, adaptive system of distant learning, distant course, learning material, cluster analysis.*