

В. А. ЛАБЧУК

аспірант кафедри програмних засобів і технологій
Херсонський національний технічний університет
ORCID: 0009-0006-9252-761X

SMALL DATA ЯК ПЕРСОНАЛІЗОВАНЕ РІШЕННЯ ДЛЯ ФОРМУВАННЯ СТРАТЕГІЇ КОРЕКЦІЇ ВАГИ ЛЮДИНИ

У статті розглянуто сутність роботи з даними, які на перший погляд здаються універсальними, однак такими не є. Навіть попри можливість зведення розв'язуваних задач до декількох типів, їх класифікації та алгоритмізації, й успішного застосування *big data*, на практиці універсальні рішення будь-яких задач не завжди є доцільними та придатними до масового тиражування, навіть якщо здається що проблема виглядає однаково для всіх людей чи групи. Спричиняти таку поведінку можуть взаємозалежності між атрибутами та групами атрибутів даних, а також наявність контексту, що може не враховуватися у випадку використання великих даних. Тож у статті розглянуто доцільність застосування *small data* на прикладі персональної боротьби з ожирінням та вплив різних атрибутів, досліджених емпірично, на загальний результат. Завдання корекції ваги – це саме той приклад, де загальні рішення можуть призводити не лише до різного ефекту у схожих людей, але навіть до різного результату в 1-ї людини в різні часові проміжки життя.

Для застосування концепції *small data* для такої прикладної задачі, доцільне застосування методів багатокритеріального аналізу, зокрема TOPSIS для вибору найбільш близького до ідеального рішення при ранжуванні за багатьма критеріями одночасно, різних методів аналізу ієрархій для порівняння альтернатив (в даному випадку різних стратегій корекції ваги для однієї людини) за кількома критеріями. Доцільні й методи кореляційного аналізу з метою виявлення зв'язків між різними атрибутами (наприклад, видом тренування, його здатністю до витрат калорій, рівнем та бажанням споживання їжі після нього, а також мотивацією займатися саме цим тренуванням конкретній людині), та регресійного аналізу, а також інтелектуальний аналіз даних.

Науковою новизною є дослідження ситуації на прикладі реальних емпіричних даних та взаємозв'язку між ними.

Ключові слова: *small data*, обробка даних, персоналізована система, медичні дані, складна динамічна система, розмір та розмірність вибірки, контекст даних, правдивість даних.

V. A. LABCHUK

Postgraduate Student at the Department of Software Tools and Technologies
Kherson National Technical University
ORCID: 0009-0006-9252-761X

SMALL DATA AS A PERSONALIZED SOLUTION FOR FORMATION OF THE STRATEGY OF THE PERSON'S WEIGHT CORRECTION

The article shows the handling with so-called universal data and problems during this, but which isn't universal in real due to some hidden patterns between some attributes. Even when algorithmization and classification of tasks for only several big branches are possible and Big data exists – even now it's not the best idea to always use universal methods for dealing with all problems. Such behavior can look strange a little but can be caused by coupling between some of attributes in the multiattributes samples of a data. And even more – context exists as well but big data doesn't always very good to deal with it. So the aim of the article is to find some problem which shows this principle brightly and one of them is task of normalizing the human's weight. This problem is insidious due to causing not only the different effect during the same actions in quite similar people, but even between the different version of one people in different time periods.

For using the conception of small data for so applied tasks, it will be useful to use some methods like multicriteria analysis, f.e. TOPSIS to deal with the most close to ideal solution while choosing between many criterias simultaneously. Different methods for ierarchy analysis as well so we can compare the alternatives (it's about different strategies of weight's correction for the same person). Also the methods of correlation analysis will be useful due to allowing to establish connection between different attributes from the multiattributes data samples (f.e. between type of sport activity, amount of possible kkal for burn, amount of desire to eat after that, and motivation to exercise with the help of this sport for concrete person in general).

Scientific innovation here is in researchement of situation with the help of real empiric data from real world and finding out the impact on the person.

Key words: *small data*, data processing, personalized system, medical data, complex dynamic system, size and dimension of the sample, data context, data veracity.



Постановка проблеми

Існує проблема боротьби з проблемами, які потребують високоперсоналізованих рішень, навіть якщо здаються такими, які можуть обійтися без таких. Одним зі способів забезпечення цього є парадигма small data та розгляд об'єкту і предмету дослідження як складної динамічної системи. Іншою проблемою є виявлення залежності результату від конкретних емпіричних даних на практиці.

Аналіз останніх досліджень і публікацій

Дослідженням проблеми вирішення завдань за допомогою small data, які на перший погляд здаються загальними, однак такими не є, і відповідно потребують персоналізованих рішень, займалися науковці такі як Eric B. Hekler, P. Klasnja, G. Chevance, N. M. Golaszewski, D. Lewis, I. Sim. Шаховська Н.Б. займалася дослідженням питань якості даних і класифікацією невизначеностей, боротьбою з кожним видом з них. Святний В. А. досліджував складні динамічні системи, прикладом якої може бути персоналізована система слідування за вагою пацієнта, а Шерстюк В.Г. досліджував вплив розміру та розмірності вибірок на можливість аналізу даних.

Формулювання мети дослідження

Метою дослідження є аналіз показників та їх взаємозалежності між собою з метою формування оптимальної динамічної стратегії корекції ваги людини на кожному етапі, зокрема на етапі значного відхилення від комфортної маси, незначного відхилення, в умовах значних коливань ваги або ж на етапі повторної корекції маси тіла. Для цього дослідження має проаналізувати основні фактори впливу на вагу, можливість різної поведінки в разі схожих дій але для різних людей, а також можливість різної поведінки навіть для однієї людини однак в різні періоди часу.

Викладення основного матеріалу дослідження

Проблеми в будь-якій сфері характеризуються або здатністю бути вирішеними загальними універсальними методами, або ж здатністю потребувати комплексних і специфічних саме для конкретної ситуації рішень. Для першого випадку існує автоматизація, другий же випадок складніший. Більше того, проблема може здаватися вирішуваною загальними методами і засобами на перший погляд, однак такою не бути. Зазвичай це ті явища, з якими часто стикаються в житті, однак які мають таку велику кількість атрибутів, що встановити чітку залежність результату від зміни того чи іншого параметру не вдається або ж дуже важко. Одним з таких явищ є питання нестандартної ваги людей, зокрема ожиріння. На перший погляд, це загальна проблема, з якою стикаються мільйони людей. І рецепт вирішення проблеми теж на початку здається простий і універсальний – потрібно лише щоб витрата кілокалорій була більшою, ніж споживання. Але атрибутів настільки багато, що загальні рішення можуть не підходити [1]. А до атрибутів можна віднести:

1. Ступінь зайвої ваги – надмірна чи дорівнює «пару зайвих кілограм». При надмірній масі є з чого зменшувати і можливі зменшення по багато кг/міс, однак навіть проста фізична активність для людини є важкою (витрачається багато ккал в розрахунку на 1 зусилля). Крім того при надмірній вазі є заборона на багато видів фізичної активності та їх інтенсивність, зокрема біг, стрибки, катання на велосипеді (схили та високоінтенсивне). В другому ж випадку людина витримує інтенсивні та тривалі навантаження, менше втомлюється, однак не має значної зайвої ваги для швидких видимих покращень.

2. Наявність супутніх хвороб – захворювання, як діабет, заважають або ж і унеможливають процес зменшення ваги. А проблеми з суглобами накладають значні обмеження на фізичну активність. Судинні ж проблеми вимагають регуляції фізичної активності, але малоінтенсивної (пульс не більше 150).

3. Вік – обмеження на гнучкість, швидкість, динамічність і тривалість.

4. Стиль харчування – середня добова норма споживання калорій, а також (не)улюблені продукти (уникнення може відбуватися як з причин несмаку, так і цінових причин або наявності продукту в регіоні, що збільшує розмірність вибірки). Також можливі певні фізичні особливості, наприклад, головний біль при відсутності виробів з борошна в раціоні, що унеможливить взагалі подальшу фізичну активність, а з іншого боку вимагає споживання більшої кількості калорій, а отже і необхідності в більш інтенсивних або ж довгих тренуваннях.

5. Добова витрата калорій та фізична активність – залежить від виду фізичної активності, тривалості та тренуваності.

6. Самодисциплінованість і емоційна стабільність – людина може не вміти поводитися достатньо ефективно, якщо не бачить результату. До того ж, якщо підібране для людини тренування є якраз її нелюбленим (наприклад, не любить бігати), то процес швидко припиниться, людина почне робити все через силу, що спричинить до швидкої втрати інтересу до досягнення результату.

Насправді атрибутів може бути значно більше і ризиком використання такого багатоатрибутного набору даних є прокляття розмірності, яке найяскравіше проявляється за умов, коли розмірність (кількість атрибутів) перевищує розмір [2]. Однак, так як це рішення глибокоперсоналізоване і щоразу застосовуватиметься для однієї конкретної людини, то ця проблема менш актуальна. Пояснити це можна тим, що множинні тільки атрибути, тоді як розмір становить 1, що не призводить до кількості комбінацій, яка дорівнює добутку розміру та розмірності. Кількість комбінацій лише становить величині розмірностей в даному випадку. Таким чином, хоч вплив певного параметру на загальний ефект не завжди інтуїтивно очевидний, особливо враховуючи можливість взаємовпливу

різних параметрів між собою, однак справжнє прокляття розмірності яскраво б проявлялося, якщо б аналіз відбувався зразу на декількох пацієнтах. Тоді б, навіть у схожих пацієнтів можлива дещо різна реакція на однакові зовнішні чинники і їх взаємодію між собою, що сильно б ускладнювало аналіз даних і розуміння меж впливу окремих параметрів.

Для матеріалів статті взяті експериментальні дані, які неодноразово вимірювалися автором, втім у інших людей навіть співвідношення між витратою при різних активностях відрізнятиметься. Так, наприклад, було досліджено, що найвищу витрату калорій вдається досягти саме при високоінтенсивному динамічному катанні на велосипеді у місцевостях з чітко вираженим рельєфом (наявністю схилів, спусків), а не наприклад бігу, на рівні аж до 800 ккал/год. Ще продуктивнішим видом спалювання можуть виступати стрибки, але без можливості зробити тривалість тренування достатньо довгою і велика витрата на одиницю часу компенсуватиметься малою протяжністю тренування і постійними перервами (тоді як велосипед часто дозволяє безперервну їзду). Для порівняння, під час бігу вдається спалювати не більше ніж 600 ккал/год, до того ж досягти відповідної тривалості, а також тренування на рельєфі – тяжче, ніж у випадку з велосипедом, який дозволяє зробити високоінтенсивне тренування 2-годинним, забезпечивши спалювання на рівні аж 1400 ккал (800 за першу годину та 600 за другу) лише за одне тренування. В той же час існують і інші фізичні активності, які можуть бути висококалорійними і навіть приємними, але можуть мати свої власні обмеження, наприклад недоступність протягом більшої частини року, зокрема ковзани. З іншого боку велосипед є достатньо гнучким інструментом і за необхідності дозволяє досягти навантажень менших ніж при бігу, і навіть при ходьбі, за умови спокійної їзди та відсутності рельєфу, що може бути рекомендованим для людей з обмеженнями на інтенсивність фізичних активностей, і частково це пояснюється можливістю велосипеда їхати накатом (при ходьбі все одно потрібно здійснювати рух ногами на кожен крок).

Оскільки йшлося про можливу взаємозалежність атрибутів між собою, то важливо щоб тренування не лише призводило до значного спалювання калорій, але й не призводило до несамовитого бажання споживати багато їжі після нього, що автоматично нівелювало б кількість витраченого обсягу. І реакція після тренування може сильно залежати від внутрішніх особливостей організму, а можливими з них є: втома, сильне виснаження, сильне бажання споживати їжу, обезводнення, відсутність реакції (в разі короткого, легкого тренування або хорошої адаптованості). До того ж для спалювання саме жирів, необхідна робота на певному пульсі (120–135, середньому), рівень якого втім може сильно різнитися навіть при абсолютно однаковій активності для різних людей, а спалювання саме жирів починається не відразу, а орієнтовно після 40 хвилин, тому короткі, навіть інтенсивні тренування, не завжди підходять для цієї цілі (а для людей з занадто надмірною вагою навіть протипоказані). Але навіть для 1-єї людини і в межах ітерації процес корекції можна розглядати як складну динамічну систему через можливість змін [3–5]. Одна ітерація означає процес схуднення за один підхід, тоді як різні ітерації – повторну потребу в схудненні в іншому віці. Зміна стратегії, навіть в межах ітерації, пояснюється тим, що коли людина має занадто надмірну вагу, то їй не можна займатися багатьма видами фізичної активності, а з іншого боку втрата калорій навіть при легкому русі і умовні –4кг за місяць – не важко (за наявності невеликого обсягу зайвої ваги це занадто інтенсивно). Але з часом організм пристосовується, активність стає легшою, її динамічність дозволяється збільшувати, однак і витрати стають меншими, тому рекомендаційна система потребує коригування. Крім того, типовим при схудненні є можливі відкоти системи до попереднього стану (так звані зриви) і їх ступінь та частота проявів залежить від конкретної людини. Так наприклад, при виснаженні від процесу і зниженні мотивації, можливе повернення до попередньої ваги і навіть ще більшої. Якщо ж розглядати процес зниження ваги навіть в контекстах різних ітерацій, то незважаючи на те, що це наче залишається одна і та ж людина, система все одно носить характер динамічності. І це можна довести через поняття віку. Якщо людина успішно пройшла процес корекції ваги за 7 років тому і зараз розраховує на те ж саме після повернення до попереднього стану, то така стратегія може не спрацювати, адже вік впливає на діапазон можливих фізичних активностей, споживання їжі, настрій і багато факторів, які можуть відрізнятися від попереднього стану людини декілька років тому.

Існує й ризик дезінформації, яка в умовах *small data* – шлях до кардинально неправильних висновків, адже *small data* більш чутливе до дезінформації в силу своїх розмірів (рис. 1). Це проявлятиметься коли пацієнт соромиться надавати показники, змінюючи їх зазвичай на певну реалістичну величину. Можлива і брехня про рівень мотивації на поточний момент та інші показники, що спричинить до отримання некоректних даних. Для боротьби з нею доцільно застосовувати ймовірнісні способи взаємодії з інформацією у *small data*, які крім даних включають і ймовірність появи їх. Так є можливість аналізувати поточні показники і прогрес людини в реальному часі і виявляти місця, де особа може прибріхувати за допомогою визначення ймовірності правдивості деяких даних, наданих користувачем. Для таких цілей підходять Баєсівські мережі довіри, що містять серію вершин та ребер і саме їх використання доцільне в контексті малих даних [6, 7]. Крім дезінформації є і неточності, зокрема округлення (наприклад, 70 кг замість 70.9). Авторка [8, 9] пропонує досліджувати інформацію за неточностями, чітко класифікуючи і поділяючи інформацію за видами неточності, оскільки методи боротьби з цим можуть бути різні.

Дезінформацію ж згідно [10] варто розрізняти за метою здійснення. Так інформація може бути *misinformation* (дезінформація, але без цілі зашкодити), *disinformation* (з метою введення в оману) або ж *malinformation* (правдива

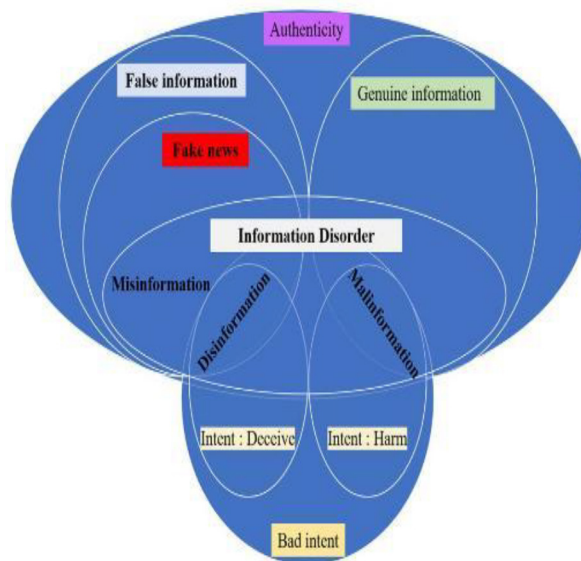


Рис. 1. Види дезінформації та її вплив на прийняття рішень

інформація, яка однак використана в контексті, який має на меті зашкодити, зокрема дані вирвані з контексту. В контексті пацієнта – «я схуд на 4 кг». Це можуть бути правдиві дані, але якщо подані як велике досягнення, але вага 160 кг, то 4 кг невідчутно, і це вже malinformation).

Висновки

У роботі досліджено прийняття рішень в умовах коли таке рішення здатне дати різний ефект для 2-ох схожих об'єктів, і навіть для одного в різні моменти його життєвого циклу. Науковою новизною роботи може вважатися дослідження взаємовпливу атрибутів на базі реальних емпіричних даних, отриманих із повсякденної життєвої практики індивіда (реальні поведінкові патерни, харчові звички, рівень фізичної активності, біометричні показники тощо), у медичній сфері, зокрема для моделювання персональної траєкторії корекції маси тіла та оцінки значущості кожного чинника в динаміці змін.

Список використаної літератури

1. Eric B. Hekler, Predrag Klasnja, Guillaume Chevanche, Natalie M. Golaszewski, Dana Lewis, Ida Sim. Why we need a small data paradigm. *BMC Medicine*. DOI: <https://doi.org/10.1186/s12916-019-1366-x> (дата звернення: 11.12.2025).
2. Я. Повод, В. Шерстюк. Дослідження методів зменшення навчальної вибірки даних. URL: https://scholar.google.com/citations?view_op=view_citation&hl=uk&user=-I16K2EAAAAAJ&cstart=20&pagesize=80&citation_for_view=-I16K2EAAAAAJ:nb7KW1ujOQ8C (дата звернення 11.12.2025).
3. В. А. Святний. Паралельне моделювання складних динамічних систем. URL: <https://pmap.donntu.edu.ua/sites/upload/articles/art137.pdf> (дата звернення 14.12.2025).
4. Hiroki Sayama. Introduction to the modeling and analysis of complex systems. С. 3–11, 27–35. URL: <https://knightscholar.geneseo.edu/cgi/viewcontent.cgi?article=1013&context=oer-ost> (дата звернення 14.12.2025).
5. M. San Miguelm, J. H. Johnson, J. Kertesz, K. Kaski, A. Diaz-Guilera, R. S. MacKay, V. Loreto, P. Erdi, D. Helbing. Challenges in complex systems science. DOI: <https://doi.org/10.1140/epjst/e2012-01694-y> (дата звернення 14.12.2025).
6. Lindsay C. Todman, Alex Bush, Amelia S.C. Hood. Small data for big insights in ecology. DOI: <https://doi.org/10.1016/j.tree.2023.01.015> (дата звернення 13.12.2025).
7. Огнева О. Є. Інформаційна система підтримки прийняття рішень для моделювання врожайності гороху овочевого. URL: <https://scholar.archive.org/work/vpcfxfxrukzjbs5lyv53vqgk6zp4/access/wayback/http://journals.kntu.net.ua/index.php/pit/article/download/625/784> (дата звернення 13.12.2025).
8. Шаховська Н. Б. Оцінювання якості консолідованих даних. URL: <http://dspace.nbu.gov.ua/bitstream/handle/123456789/113605/05-Shahovska.pdf?sequence=1> (дата звернення 12.12.2025).
9. Шаховська Н. Б. Дослідження якості консолідованих даних у просторах даних. URL: http://www.immsp.kiev.ua/publications/articles/2012/2012_1/01_2012_Shahovska.pdf (дата звернення 12.12.2025).
10. Esma Aimeur, Sabrine Amri, Gilles Brassard. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and mining*. DOI: <https://doi.org/10.1007/s13278-023-01028-5> (дата звернення: 24.12.2025).

References

1. Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC Medicine*, 17(1), 133. <https://doi.org/10.1186/s12916-019-1366-x>
2. Povod, Y., Sherstjuk, V. (2020). Study of data sample reduction methods. *Problems of Information Technologies*, 27(1), 98–107. https://scholar.google.com/citations?view_op=view_citation&hl=uk&user=-I16K2EAAAAAJ&cstart=20&pagesize=80&citation_for_view=-I16K2EAAAAAJ:nb7KW1ujOQ8C
3. Svjatnyj, V. A. (2006). Parallel modelling of complex dynamic systems. <https://pmap.donntu.edu.ua/sites/upload/articles/art137.pdf>
4. Sayama, Hiroki. (2015). Introduction to the modeling and analysis of complex systems. pp. 3–11, 27–35. Open SUNY Textbooks. <https://knightscholar.geneseo.edu/cgi/viewcontent.cgi?article=1013&context=oer-ost>
5. Miguel, M. S., Johnson, J. H., Kertesz, J., Kaski, K., Díaz-Guilera, A., MacKay, R. S., Loreto, V., Érdi, P., & Helbing, D. (2012). Challenges in complex systems science. *The European Physical Journal Special Topics*, 214(1), 245–271. <https://doi.org/10.1140/epjst/e2012-01694-y>
6. Todman, L. C., Bush, A., & Hood, A. S. (2023). ‘Small Data’ for big insights in ecology. *Trends in Ecology & Evolution*, 38(7), 615–622. <https://doi.org/10.1016/j.tree.2023.01.015>
7. Ohnieva, O. Y. (2019). Decision support information system for modeling green pea yield. *Problems of information technologies*, 26(1). <https://scholar.archive.org/work/vpcfxfukzjbs5lyv53vqgk6zp4/access/wayback/http://journals.kntu.net.ua/index.php/pit/article/download/625/784>
8. Shakhovska, N. B. (2014). Consolidated data quality assessment. <http://dspace.nbu.gov.ua/bitstream/handle/123456789/113605/05-Shahovska.pdf?sequence=1>
9. Shakhovska, N. B. (2012). Consolidated data quality researchment at data spaces. *Mathematical machines and systems*, 1(1). http://www.immsp.kiev.ua/publications/articles/2012/2012_1/01_2012_Shakhovska.pdf
10. Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30. <https://doi.org/10.1007/s13278-023-01028-5>

Дата першого надходження статті до видання: 06.01.2026

Дата прийняття статті до друку після рецензування: 09.02.2026

Дата публікації (оприлюднення) статті: 30.04.2026