

Порівняльний аналіз алгоритмів формування разових рад із захисту PhD дисертацій

¹ Вінницький національний технічний університет;

² Донецький національний університет імені Василя Стуса

В Україні PhD-дисертації захищають у разових радах, які формують вручну. Це може займати багато часу і при цьому експертиза підібраних членів може виявитись сумнівною, не кажучи вже про корупційні ризики. Тому автоматизація процесу формування рад є актуальною. В даній роботі проводиться порівняльний аналіз алгоритмів призначення рецензентів за критеріями якості підбору і витратами ресурсів на пошук членів разової ради. Тематика дисертації та тематика статей потенційних членів ради представляються категоріальними розподілами над науковими спеціальностями з системи класифікації наук ANZSRC-2020. Як критерій схожості використовується удосконалена метрика Чекановського, що додатково враховує спорідненість різних наукових спеціальностей за допомогою індексу Жаккара. Порівняння ефективності алгоритмів призначення рад здійснюється експериментально на датасеті з 67 PhD-дисертацій. Для формування датасету використано інформаційну систему НАЗЯВО. Експерименти проводились для таких алгоритмів оптимізації: повний перебір; повний перебір на прорідженій множині кандидатів (враховуються лише кандидати з певним мінімальним рівнем схожості); жадібний алгоритм; жадібний алгоритм з елітизмом; ізольований підбір (найбільш схожі кандидати вважаються оптимальними). Завдяки оптимізації склад разових рад вдалося покращити в середньому на 13-34% залежно від задіяного алгоритму. Таке покращення відбулося за невеликого пулу кандидатів. Збільшення кількості кандидатів призведе до більшого ефекту від застосування запропонованих алгоритмів підбору опонентів. В результаті експериментів встановлено, що добрий баланс між якістю підбору та ресурсозатратністю досягається жадібним алгоритмом без елітизму та повним перебором на прорідженій множині кандидатів. При цьому за будь-яким алгоритмом, новостворені разові ради не гірше, а як правило кращі, ніж разові ради, які запропонував заклад.

Ключові слова: задача підбору рецензентів, алгоритми оптимізації, категоріальний розподіл, ANZSRC-2020, PhD-дисертація, разова рада.

Вступ

В Україні PhD-дисертації захищають у разових радах, які складаються з 5 фахівців. Голова та 1-2 рецензенти представляють заклад, де утворена рада, а 2-3 опоненти запрошуються з інших установ. Склад ради затверджує вчена рада закладу. Ручне формування разових рад має кілька недоліків. По-перше, це корупційні ризики, коли раду складають із дружніх осіб, які завжди надають схвальні відгуки. Успішний захист вигідний як тим, хто обирає членів ради, так і закладу, що її затверджує. За даними НАЗЯВО, вже відбулося 6 433 успішні захисти, жодного випадку відмови не було. По-друге, пошук кандидатів у раду займає багато часу. По-третє, склад ради може слабо відповідати тематиці дисертації через людський фактор при ручному підборі. Тому автоматизація процесу формування рад є актуальною.

Формування разових рад є типовою задачею призначення (підбору) рецензента (Reviewer assignment problem). Задача призначення рецензента складається з трьох етапів [1]: 1) вибір методу представлення даних про рецензентів та заявки; 2) підрахунок схожості між заявкою та рецензентами; 3) розподіл заявок для максимізації загальної схожості з урахуванням обмежень, таких як баланс навантаження, вподобання рецензентів та відсутність конфлікту інтересів.

Автоматичний підбір рецензентів передбачає доступність деякої початкової інформації про рецензентів та заявки. В цій роботі як початкова інформація використовуються ключові слова дисертації та ключові слова статей членів разових рад. Тематика дисертації та тематики досліджень членів ради формалізуються за підходом з [2], в якому вони представляються як категоріальний розподіл над науковими спеціальностями з системи

класифікації наук ANZSRC-2020. Критерієм схожості обрано удосконалену метрику Чекановського [3], яка додатково враховує спорідненість наукових спеціальностей за допомогою індексу Жаккара [4].

Результати дослідження

Для експериментів із підбору рецензентів використовується датасет дисертацій [5]. Для його формування використано інформаційну систему НАЗЯВО (NAQA.Svr). Зібрано інформацію про 67 дисертацій, які належать до різних спеціальностей, переважною з галузі знань «Інформаційні технології». Експерименти проводились для таких алгоритмів оптимізації: повний перебір; повний перебір на прорідженій множині кандидатів (враховуються лише кандидати з певним мінімальним рівнем схожості); жадібний алгоритм; жадібний алгоритм з елітизмом; ізольований підбір (найбільш схожі кандидати вважаються оптимальними). Суть експериментів полягала в наступному: для кожної разової ради підбрано нових опонентів з членів інших разових рад з датасету. На рис. 1 показано результати оптимізації для різних алгоритмів підбору. У більшості випадків разові ради, сформовані закладом, мають нижчу відповідність тематиці дисертацій в порівнянні з тими, що сформовано за автоматичними методами підбору. Ручне формування рад призводить до середнього рівня відповідності через неможливість враховувати всіх потенційних кандидатів. Натомість автоматичний підбір навіть за невеликого пулу кандидатів дозволяє значно покращити відповідність рад лише завдяки оптимальному добору опонентів.

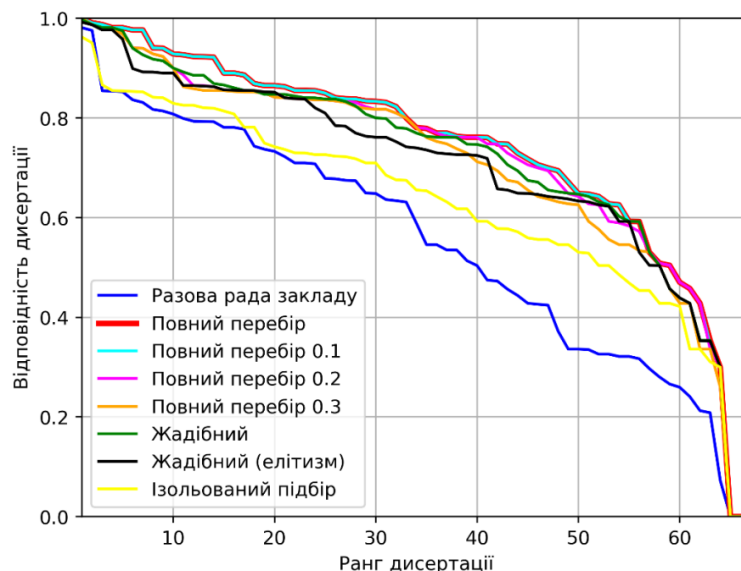


Рис. 1. Порівняння результатів підбору за різних алгоритмів оптимізації

На рис. 2 порівнюється робота різних алгоритмів оптимізації за критеріями «тривалість – якість». Якість у даному контексті визначається як відносний рівень покращення разових рад щодо разових рад від закладу. Перебір на прорідженій множині кандидатів за порогу у 0.3 однозначно невдалий. Усі інші утворюють множину Парето. Тому під час вибору алгоритму необхідно врахувати пріоритети – потрібен швидкий результат чи якісний. З рис. 2 видно, що рівень покращення внаслідок переходу з жадібного алгоритму на алгоритми повного перебору зростає повільно. Але тривалість оптимізації зростає суттєво. Найбільш збалансованим можна вважати жадібний алгоритм підбору без елітизму.

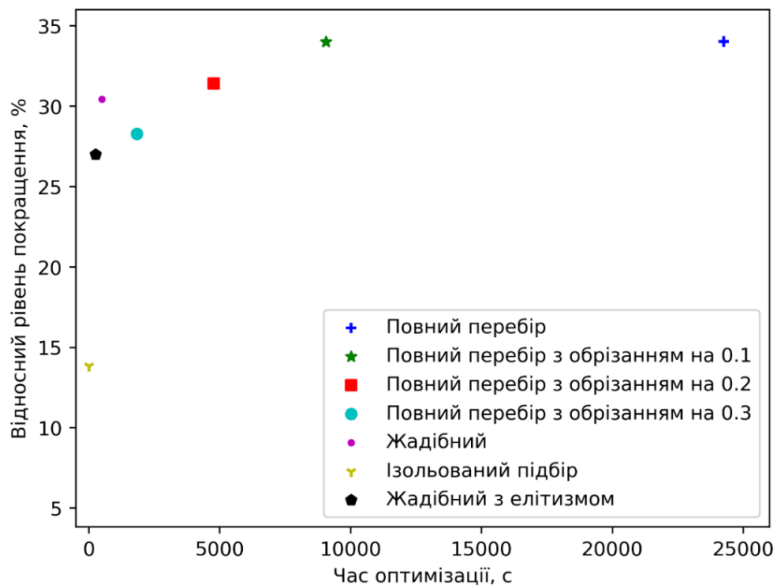


Рис. 2. Порівняння роботи різних алгоритмів оптимізації за критеріями «тривалість – якість»

Висновки

Порівняння роботи алгоритмів оптимізації на датасеті з 67 PhD-дисертацій показало, що добрий баланс між якістю підбору та ресурсозатратністю досягається жадібним алгоритмом без елітизму та повним перебором на прорідженій множині кандидатів. Для досягнення бажаних результатів можна також обирати між різними алгоритмами оптимізації в залежності від того який результат потрібний – швидкий чи якісний. Завдяки оптимізації склад разових рад вдалося покращити в середньому на 13-34% залежно від задіяного алгоритму. Таке покращення відбулося за невеликого пулу кандидатів. Збільшення кількості кандидатів призведе до більшого ефекту від застосування запропонованих алгоритмів підбору опонентів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Петричко, М. В., & Штовба, С. Д. (2024). Автоматизація підбору наукових рецензентів: огляд задач і методів. Вісник Вінницького політехнічного інституту, (1), 56-64. <https://doi.org/10.31649/1997-9266-2024-172-1-56-64>
- [2] Штовба С.Д., Петричко М.В. (2021). Тематичне моделювання науковців на основі їх інтересів у Google Scholar, Системні дослідження та інформаційні технології, №2, вересень, с. 113-129. <https://doi.org/10.20535/SRIT.2308-8893.2021.2.09>.
- [3] Shtovba, S., Petrychko, M., & Shtovba, O. (2023). Similarity Metric of Categorical Distributions for Topic Modeling Problems with Akin Categories. In CEUR Workshop Proceedings (Vol. 3392, pp. 76–85). CEUR-WS. <https://doi.org/10.32782/cm/3392-7>.
- [4] Shtovba, S., & Petrychko, M. (2019). Jaccard index-based assessing the similarity of research fields in dimensions. In CEUR Workshop Proceedings (Vol. 2533, pp. 117–128). CEUR-WS.
- [5] Petrychko, M., & Shtovba, S. (2024). Dataset for PhD theses reviewers assignments. ResearchGate. <http://dx.doi.org/10.13140/RG.2.2.23147.35362>

Петричко Микола Володимирович — доктор філософії, асистент кафедри комп'ютерних систем управління, e-mail: m.petrychko@vntu.edu.ua;

Вінницький національний технічний університет;

Штовба Сергій Дмитрович — д.т.н., професор, професор кафедри інформаційних технологій, e-mail: s.shtovba@donnu.edu.ua.

Донецький національний університет імені Василя Стуса.

M. V. Petrychko¹
S. D. Shtovba²

Comparative Analysis of Reviewer Assignment Algorithms for PhD-Theses Committees

¹Vinnitsia National Technical University

²Vasyl Stus' Donetsk National University

In Ukraine, PhD theses are defended in one-time committees, which are formed manually. This can take a lot of time, and at the same time, the expertise of the selected members can be questionable, not to mention the corruption risks. Therefore, the automation of the process of assigning committees is important. In this work, a comparative analysis of algorithms for assigning reviewers based on assignment quality criteria and resource costs for finding members of the one-time committee is carried out. The subject of a thesis and the subject of the articles of potential committee members are represented by categorical distributions over research specialties from the ANZSRC-2020 research classification system. As a similarity criterion, improved Chekanovsky metric is used, which additionally takes into account the kinship of different research specialties using Jaccard index. The comparison of the efficiency of the algorithms for the assignment of committees is carried out experimentally on a dataset of 67 PhD theses. The NAQA information system was used to create the dataset. Experiments were conducted for the following optimization algorithms: brute force; brute force on a truncated set of candidates (only candidates with a certain minimum level of similarity are considered); greedy algorithm; greedy algorithm with elitism; isolated assignment (the most similar candidates are considered optimal). Thanks to optimization, it was possible to improve the composition of one-time committees by an average of 13-34%, depending on the algorithm used. This improvement occurred with a small candidate pool. An increase in the number of candidates will lead to a greater effect of applying the proposed algorithms for assigning opponents. As a result of the experiments, it was established that a good balance between the quality of assignment and resource consumption is achieved by a greedy algorithm without elitism and a brute force on a truncated set of candidates. At the same time, according to any algorithm, the newly created one-time committees are not worse, but usually better, than the one-time committees offered by the institution.

Keywords: reviewer assignment problem, optimization algorithms, categorical distribution, ANZSRC-2020, PhD-this, one-time committee.

Petrychko Mykola Volodymyrovych — PhD, assistant on Computer Control Systems Department, e-mail: m.petrychko@vntu.edu.ua;

Vinnitsia National Technical University;

Shtovba Serhiy Dmytrovych — Doctor of Science, Full Professor, Professor on Information Technology Department, e-mail: s.shtovba@donnu.edu.ua.

Vasyl Stus' Donetsk National University.