UDC 519.254+001.2
DOI: 10.20535/SRIT.2308-8893.2025.4.08

ALGORITHMS FOR ASSIGNMENT OF EXTERNAL REVIEWERS FOR PHD-THESIS DEFENSE

SERHIY SHTOVBA, MYKOLA PETRYCHKO

Abstract. We propose an approach to assigning external reviewers. In the proposed approach, only the semantic similarity between applications and reviewers is taken into account; the similarity indices are assessed, and the necessary number of reviewers is assigned to ensure the maximum suitability level of the reviewers with the application, according to some criteria. We also perform a comparative analysis of various optimization algorithms using the criterion of "assignment quality–optimization time". Experiments on the dataset showed that a reasonable balance between the "assignment quality" and "optimization time" criteria for the assignment of external reviewers can be achieved using a greedy algorithm without elitism or brute-force search on a truncated set of candidates. An application of the proposed algorithms improves the average quality of PhD committees by 13–34% across the entire dataset, depending on the algorithm used.

Keywords: external reviewers, reviewer assignment problem, categorization, optimization, brute force algorithm, greedy algorithm, assignment in isolation, PhD-thesis, Dimensions, ANZSRC 2020, research group.

INTRODUCTION

External reviewers are persons from outside an institution who are invited to provide an independent evaluation or assessment of a particular project, document, research paper, or system. They are often selected for their expertise in a relevant field and are expected to offer objective, unbiased feedback. In academia, external reviewers are used in the peer-reviewing to evaluate the quality, relevance, and originality of academic papers before publication. They may also be used for reviewing PhD-thesis.

In Ukraine, a PhD thesis is defended in front of a committee. A PhD-committee consists of 5 scientists with expertise in the thesis subject. The chairman and 1 or 2 reviewers are from the PhD-student's institution, and 2 or 3 external reviewers are invited from other institutions. The members of the PhD-committee are assigned manually, which has several disadvantages. First of all, there are corruption risks when the committee is formed exclusively from friendly persons who a priori give only favorable reviews regardless of the results of the thesis. Second, a lot of time is spent on manual search and analysis of candidates for the committee. Third, the combining competence of the committee may not fully correspond to the thesis topic due to the fact that some of the good candidates were missed during the manual search. Therefore, there is an interest in automating the assignment of reviewers to eliminate the specified risks of the human factor influence.

The general task of assigning the reviewers consists of three stages [1]: 1) forming of a pool of potential reviewers and subsequently choosing a method of data representation for reviewers and applications; 2) assessing the similarities between the application and the reviewers; 3) assignment of applications to reviewers to maximize combined similarity across all the subjects with some constraints. Typical constraints include balancing reviewer workloads, taking into account their preferences, and preventing conflicts of interest. In this work, it is assumed that the pool of potential reviewers is available.

Automatic assignment of reviewers assumes that some initial information about reviewers and applications is available. A structured set of such information is called a reviewer profile and an application profile. The following information about reviewer's publications is used usually to build a reviewer's profile: title, abstract, keywords, full text, list of references, and list of citations [2]. Abstract, full text, keywords and title are most often used to create an application profile [2].

Applications' profiles and reviewers' profiles are built using various natural language processing methods based on bag of words [2; 3; 4], hidden semantic analysis [5; 6], topic modeling [7; 8], static language models with deep learning [9; 10; 11] and contextual models with deep learning [12]. Approaches to solving the problem of automatic assignment of reviewers in most cases require a fairly large amount of initial information about the reviewers' publications, their interaction with other scientists, and similar information about the authors of applications. Analyzing this information is costly and will not be expedient if thousands of candidates are to be analyzed in detail for each team of reviewers.

Our paper is dedicated to the assignment of external reviewers for PhD thesis defense. A candidate list of available internal reviewers is usually too short; hence it makes no sense to optimize it. We focus on the task of express assignment of external reviewers, where a long initial list of candidates is to be reduced drastically. The subsequent short list can be analyzed manually, or a fine assignment procedure can be activated, which is resource-intensive and requires a much larger volume of initial information than is required for express assignment. During express assignment, only the semantic similarity between applications and reviewers is taken into account, which provide the maximal level of collective competence of the committee. In this paper we perform comparative analysis of various optimization algorithms by using the criteria of *"assignment quality – optimization time"* in order to better understand the tradeoffs when choosing "assignment quality" over "optimization time" or vice versa.

DATA REPRESENTATION

At the first stage of assigning the reviewers, it is necessary to choose the source data for decision-making, as well as the method of its representation in vector form. In the case of an application, a list of its keywords is used, and in the case of a reviewer, a list of keywords obtained from available data is used. In general, this list of keywords can be from the candidate's recent publications, from his CV or from a profile from some register of scientists. In the second case, keywords or research interests are formed by the candidate at his own discretion, that is, they are presented in an arbitrary form without reference to any rubric or classifier.

The source data is usually processed using statistical models, topic models and embedding models. Some of them analyze the frequency of occurrence of words in the text, others form representation vectors based on the co-occurrence of words. Usually, the resulting vector representations are difficult to interpret. In addition, obtaining such representations requires a large amount of data. We suggest using the approach from [13], according to which a set of keywords is categorized as a vector in the space of research groups from the Australian and New Zealand Standard Research Classification — ANZSRC 2020. ANZSRC 2020 includes 171 research groups from 22 divisions. Therefore, the final representation of the application and reviewer profiles looks like a distribution over the 171 research groups from ANZSRC 2020.

In order to carry out a categorization, it is necessary to have a corpus of marked articles that are assigned to one or more research groups, and a machine learning model that, based on keywords, assigns the analyzed profile to certain research groups. We use the information resources of the Dimensions, in which more than 100M publications are already categorized according to ANZSRC 2020. For a search query in the form of a keyword, Dimensions produces an output that indicates how many publications with that keyword are assigned to each of the research group. This procedure is shown schematically in Fig. 1. It also shows that in the collection of marked documents an article can be categorized into several research groups, for example, *Article 1* is assigned to *Research Group 1* and *Research Group 2*. Based on this output, the distribution of a keyword's occurrence in the context of various research groups can be built. For example, for the keyword from Fig. 1 distribution looks like this: *Research Group 1* — 3 appearances, *Research Group 2* — 2 appearances, *Research Group 3* — 2 appearances, and *Research Group K* — 1 appearance. On the basis of this distribution, the keyword *"some keyword"* is further categorized within the framework of the research classification system. To categorize a set of keywords, the algorithm from [13] is applied, which is based on the resources and services of Dimensions. This algorithm takes into account both the occurrence of isolated keywords from a profile, as well as the co-occurrence of keyword pairs. The algorithm allows to filter the information noise caused by both stop words and rare keywords that have low reliability of the conclusions.

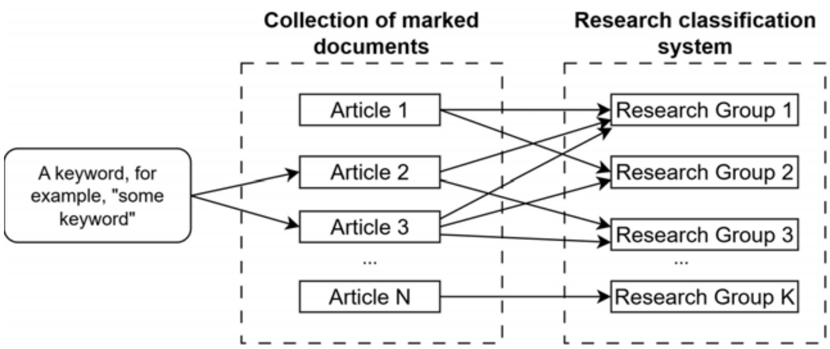

Fig. 1. Keyword categorization schema

The categorization algorithm consists of 3 stages. For a set of two keywords the procedure of categorization is schematically shown on Fig. 2. In the first stage the set E of search queries is created using the initial keywords and their pairwise combinations. At the second stage the membership degrees of queries to research groups are computed. For this the overall distribution of the number of publications over research groups using Dimensions API is found. Then the same is done for each search query with subsequent stop-words detection and noise filtering. Having done this, the relative frequencies of search queries based on the overall distribution is found and the noise reduction using cumulative contribution of research groups is done. On the third stage all the queries distributions are averaged that produces one-dimensional vector. We further perform truncation to at most *RG_max* research groups with non-zero membership degree. A reviewer by the proposed algorithm can be categorized to at most *T_max* research groups, and the smallest membership degree is restricted to be at least *RG_min_degree*. The truncation is done in the last step of the third stage by removing research groups with low membership degree.

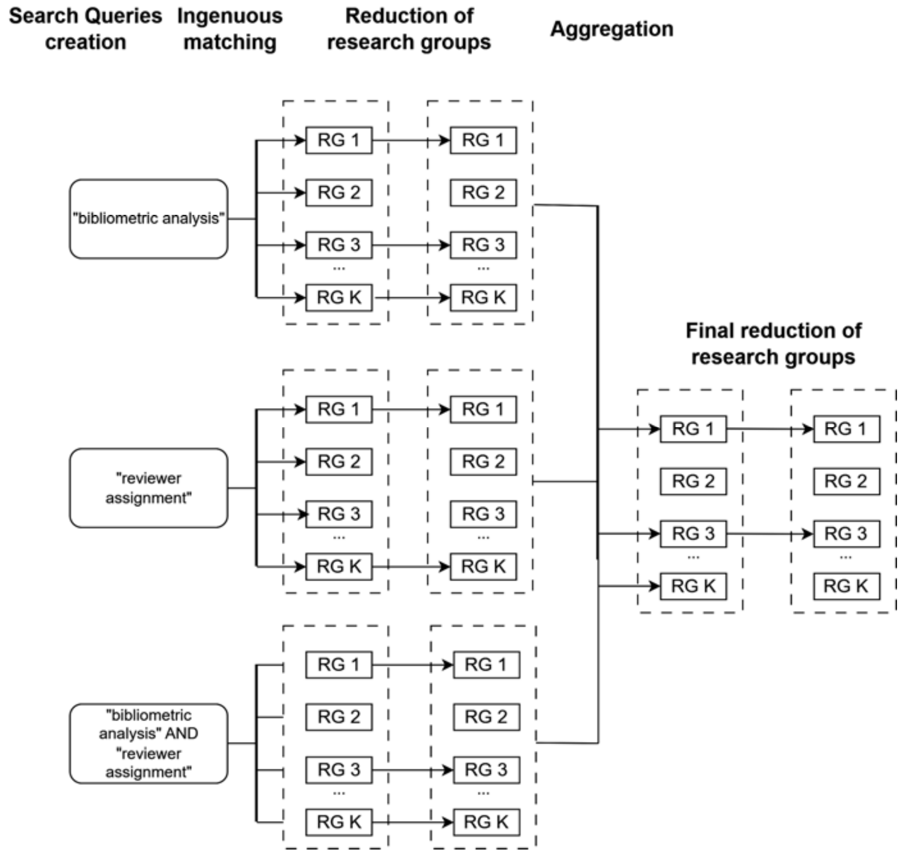

Fig. 2. Keywords detailed categorization schema

The MATLAB-style pseudocode of the categorization algorithm is as follows:

```
%STAGE #1 – creating the set E of search queries from the key-
% words w
E=w
for i=1:length(w)-1
     for j=i:length(w)
       E={E; [w(i) 'AND' w(j)] }
     end
end
%STAGE #2 – compute membership degrees to research groups by
% each query
```

```
< Find the total number of publications in each research groups
  N=[N(1), N(2), …, N(m)], m=171 >
Counter=0  % the counter of successful query responses
for i=1:length(E)
     < Find Q — the total number of publications in Dimensions,
       that contain E{i} >
     If Q>Threshold_StopWord continue   % ignoring the stop-
     words
     end
     If Q<Threshold_Noise continue;   % ignoring the rare key-
     words
     end
     < Find t(1), t(2), …, t(m) — the number of publications in
     each
       research group for query E{i} >
     %Ignoring the research group with a tiny number of publica
     % tions:
     index=find(t<Threshold_topic)
     t(index)=0
     if max(t)==0 continue
     end
     r=t./N   %frequency of E{i}'s occurrence in research groups
     %Normalizing the frequency distributions:
     Gamma=r./sum(r)
     < Choosing the most popular research groups that have cumu-
     lative
       contribution in Gamma >= Tail. ID-numbers of the remain-
     ing research groups
       are put in vector Rejected >
     %Ignoring the research groups with contribution lower than
     % Tail:
     Gamma(Rejected)=0
     Gamma=Gamma./sum(Gamma)   %normalizing again
     Counter=Counter+1
     Mu(Counter)=Gamma
end
If Counter==0 return ('Unsuccessful')
end
%STAGE #3 — compute membership degrees using all queries
Mu_mean=mean(Mu)   % averaging all successful queries
%Computing the current number of the selected research groups:
Current_N_RGs=sum(Mu_mean>0)
[Mu, RG_ID, Current_N_RGs]=Top_RG(Mu_mean, Source_RG_ID, RG_max)
% Top_RG — forms RG_ID as a selection of RG_max research groups
% with
% highest membership degree from Source_RG_ID. RG_ID is descend
% ing order
% list of research groups according to their membership degrees
% Mu.
% Vector Mu is normalized in [0; 1].
%Finish truncation based on kinship of research groups:
while (true)
      if (Current_N_RGs<=Tmax AND Mu(end)>RG_min_degree) break
      end
```

```
        if (Current_N_RGs<=1) break
        end
        < Drop the minor groups and redistribute its contribution
        to
          others based on their kinship >
        for target=1:Current_N_RGs-1
            akin_factor=Jaccard(RG_ID(target),
        RG_ID(Current_N_RGs))
            Mu(target)=Mu(target)+Mu(Current_N_RGs)*akin_factor
        end
[Mu, RG_ID, Current_N_RGs]=Top_RG(Mu, RG_ID, Current_N_RGs-1);
end
Return(Mu, RG_ID)
```

At the last stage of the algorithm when dropping a minor research group its contribution is redistributed to other research groups based on their kinship. The additional value is proportional to the kinship level between the target research group and the research group being removed. The kinship level is assessed using Jaccard index, where the size of the intersection is the number of publications categorized to belong to both research groups, and the size of the union is the number of publications categorized to either of research groups [14]. We formed the matrix of Jaccard indices for research groups using Dimensions API for the data period of 2019–2023. The intuition behind this step lays in the fact that we want to increase the influence of the subset of research groups that are more akin than others.

For example, a researcher is categorized tentatively to research groups *4410 Sociology*, *4611 Machine Learning*, *3508 Tourism*, and *3504 Commercial Services* as follows: $\left(\frac{0.4}{\mathit{4410}}, \frac{0.25}{\mathit{4611}}, \frac{0.2}{\mathit{3508}}, \frac{0.15}{\mathit{3504}}\right)$. Let us drop the minor research group *3504*. For this, we first compute Jaccard indices between *4609* and other research groups using the method from [14]. For the data of 2019–2023 they are:

$J(\mathit{4410}, \mathit{3504}) = 0.044$;

$J(\mathit{4611}, \mathit{3504}) = 0$;

$J(\mathit{3508}, \mathit{3504}) = 0.478$.

By taking into account the kinships, the contribution of the research group *4609* is redistributed in the following way:

$$\left(\frac{0.4 + 0.044 \cdot 0.15}{\mathit{4410}}, \frac{0.25 + 0 \cdot 0.15}{\mathit{4611}}, \frac{0.2 + 0.478 \cdot 0.15}{\mathit{3508}}\right).$$

As a result, we get: $\left(\frac{0.466}{\mathit{4410}}, \frac{0.25}{\mathit{4611}}, \frac{0.271}{\mathit{3508}}\right)$. After norming: $\left(\frac{0.472}{\mathit{4410}}, \frac{0.275}{\mathit{3508}}, \frac{0.253}{\mathit{4611}}\right)$. As a result, research group *3508 Tourism* has been strongly reinforced. This research group is closely related to *3504 Commercial Services*, which has been eliminated. If we simply discard the minor research group, then after normalization we get $\left(\frac{0.47}{\mathit{4410}}, \frac{0.29}{\mathit{4611}}, \frac{0.24}{\mathit{3508}}\right)$. In this case, there was no additional reinforcement of the *3508* research group.

Let's present a step-by-step example of how the proposed algorithm works. For this, *Susan Dumais* is considered as a potential reviewer. The reviewer's information is taken from her Google Scholar profile that contains a set of research interests. Those interests may be interpreted as a set of initial keywords. For this reviewer the keywords are: *"Information Retrieval"*, *"Human-Computer Interaction"*. Interests often complement each other thus making the research topics more focused. To take this into account, additional keywords are synthesized as pairs of initial interests. Interests in a pair are combined by a logical operation AND as follows: *"Information Retrieval"* AND *"Human-Computer Interaction"*. Fig. 3 shows the initial distribution of membership degrees to research groups for the research interests of *Susan Dumais*. For each of the reviewer's interest and conjunction of her interests the distribution to research groups from Dimensions is found. Then the research groups with cumulative contribution less than *Tail* is dropped to reduce the noise (Fig. 4). *Tail* is set to be 0.93. The next step is to average over all interests' distribution (Fig. 5) and further restrict the max number of non-zero membership degrees to be at most *RG_max*. *RG_max* is set to be 12. The noise reduction steps and the restriction on the max number of non-zero membership degrees are based on the assumption that researchers usually are proficient only in a few research fields at once. In the end in case of $T_{\max} = 4$ *Susan Dumais* is represented by the following research groups:

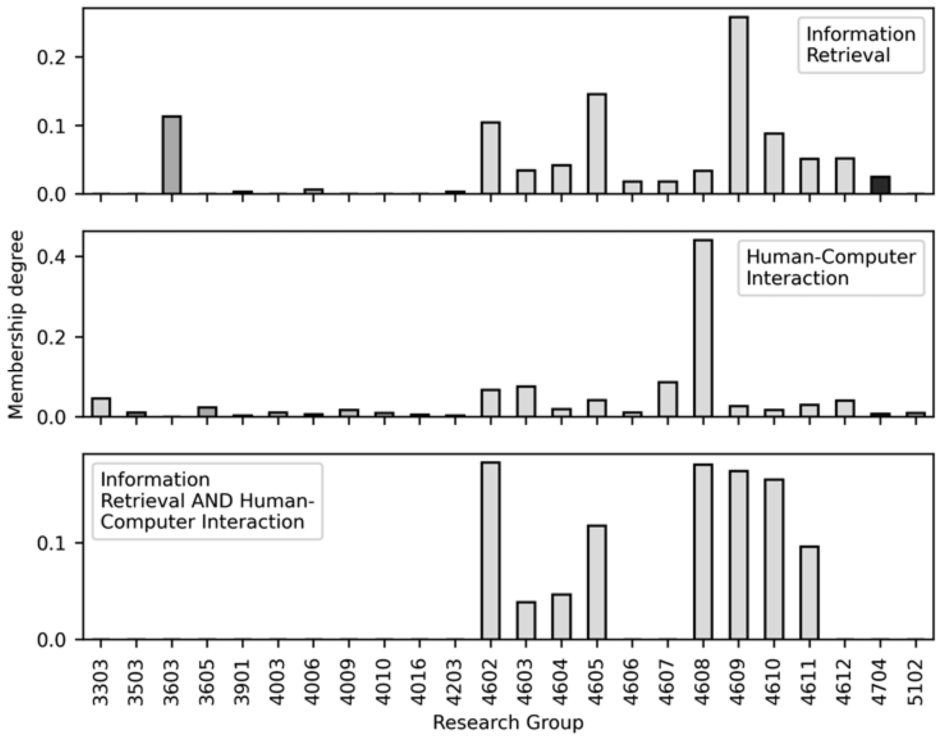

Fig. 3. The initial interests' distributions for *Susan Dumais*

4608 Human-Centred Computing	with degree 0.35;
4609 Information Systems	with degree 0.25;
4602 Artificial Intelligence	with degree 0.21;
4605 Data Management and Data Science	with degree 0.19.

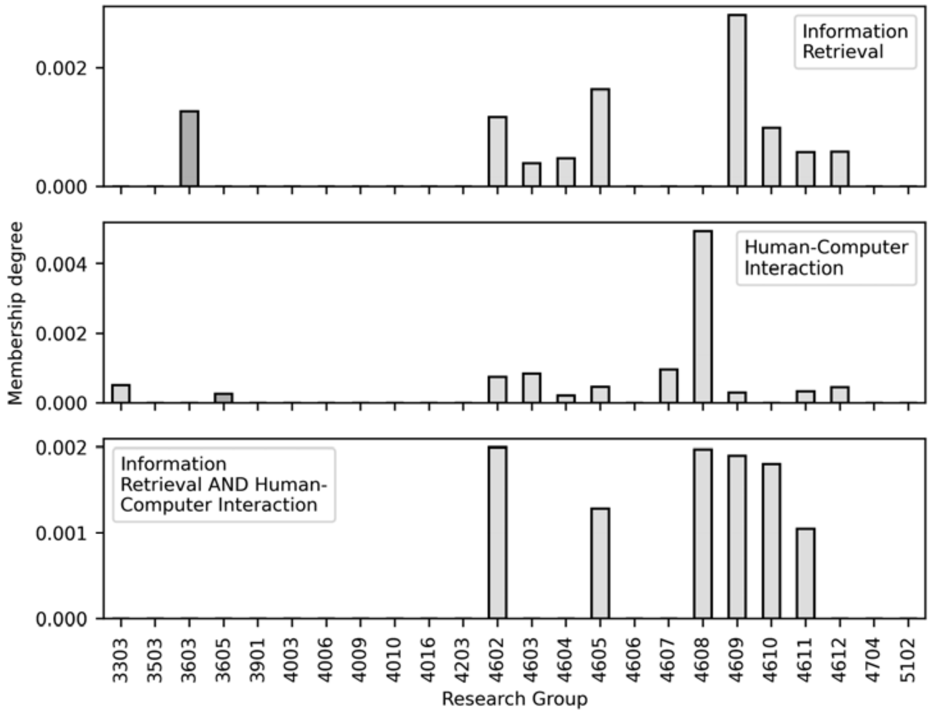

Fig. 4. Interests' distributions after filtering by *Tail*

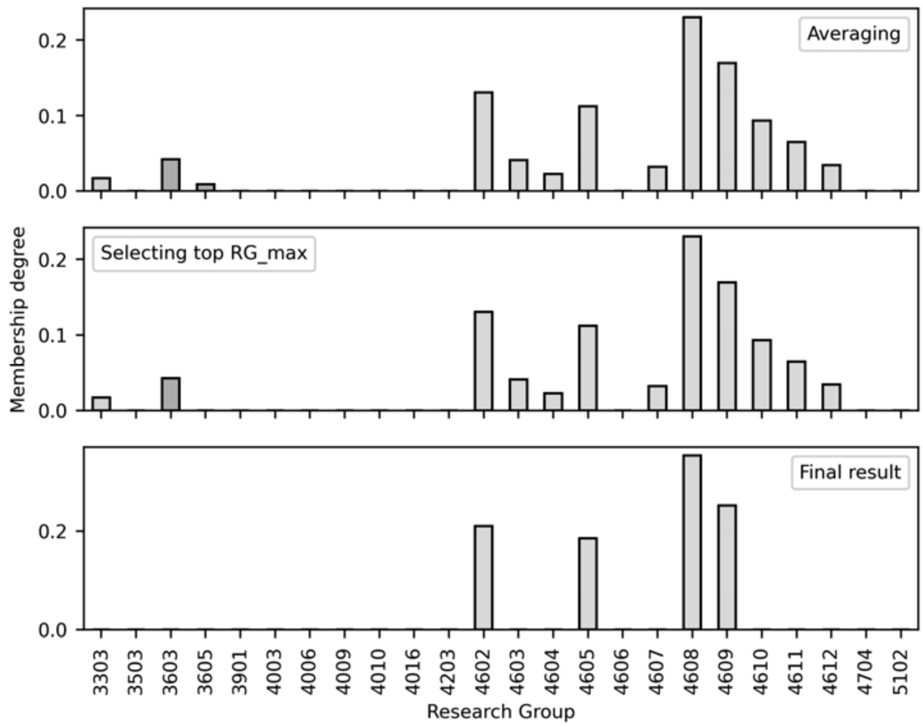

Fig. 5. Reviewer's distribution after averaging over all insterests' distributions and final result

As the result of categorization, an application profile, defined as a set of keywords $A_w = \{w_1, w_2, ..., w_n\}$, is transformed into a profile defined as a categorical distribution over research groups $A_t = \{\mu_{t_1}(A), \mu_{t_2}(A), ..., \mu_{t_m}(A)\}$, where $\mu_{t_i}(A) \in [0;1]$ denotes membership degree of application A to research group t_i, $i = \overline{1, m}$. Similarly, a reviewer's profile, defined as a set of keywords or research interests $R_w = \{w_1, w_2, ..., w_n\}$, is transformed into a profile defined as a categorical distribution over research groups $R_t = \{\mu_{t_1}(R), \mu_{t_2}(R), ..., \mu_{t_m}(R)\}$.

SIMILARITY ASSESMENT

To match reviewers and applications, a similarity metric between 2 categorical distributions, the reviewer keywords' research groups distribution and the application keywords' research groups distribution, has to be defined. For this, the metric from [15] is used. The metric calculates the similarity of two objects X and Y with the following categorical distributions $(\mu_1(X), \mu_2(X), ..., \mu_m(X))$ and $(\mu_1(Y), \mu_2(Y), ..., \mu_m(Y))$, where m denotes the number of categories, that are research groups in our case, $\mu_i(X)$ denotes membership degree of object X to i-th category, $\mu_i(Y)$ denotes membership degree of object Y to i-th category, $i = \overline{1, m}$. Distributions are normalized and satisfy the following conditions:

$$\mu_i(X) \in [0;1], \qquad \mu_i(Y) \in [0;1], \qquad i = \overline{1, m};$$
$$\sum_{i=\overline{1, m}} \mu_i(X) = 1; \qquad \sum_{i=\overline{1, m}} \mu_i(Y) = 1.$$

The categorical distributions of objects X and Y look like two fuzzy sets on universal sets of all categories. Therefore, to calculate the similarity of objects X and Y, it is proposed to use an intersection of the corresponding fuzzy sets. This is reflected in the metric [15], according to which the similarity of objects X and Y is defined as follows:

$$Fit(X, Y) = \sum_{i=\overline{1, m}} \min(\mu_i(X), \mu_i(Y)) + \Delta F(X, Y), \tag{1}$$

where $\sum_{i=\overline{1, m}} \min(\mu_i(X), \mu_i(Y))$ is an addend that evaluates the direct similarity of objects X and Y; $\Delta F(X, Y)$ is an addend that evaluates the similarity of objects X and Y through akin categories (akin research groups in our case). Across the all research groups, kinship is conveniently represented by a binary fuzzy relationship in the form of an $m \times m$ matrix. Each element of the matrix corresponds to the kinship level of two corresponding research groups. An identification of this kinship matrix is easily performed by the method [14], which uses the Jaccard index on data from Dimensions.

TASK STATEMENT OF ASSIGNMENT OPTIMIZATION

Consider the task of assigning a team of reviewers, who are collectively the best suited for reviewing an application. For this task, 2 cases are possible: forming a team from scratch and supplementing the team with new members.

Given: an application profile $A_t = \{\mu_{t_1}(A), \mu_{t_2}(A), ..., \mu_{t_m}(A)\}$ and profiles of k-th potential reviewers $R_{tj} = \{\mu_{t_1}(R_j), \mu_{t_2}(R_j), ..., \mu_{t_m}(R_j)\}$, $j = \overline{1, k}$ in the space of m research groups. The entire set of reviewers is denoted as $\mathbf{R} = \{R_1, R_2, .., R_k\}$.

Find out: subset of reviewers $S \subset \mathbf{R}$ with the highest overall suitability level to all the topics of the application:

$$Fit(A, Agg(S)) \to \max ,$$

where $Agg(S)$ denotes aggregation function of categorical distributions of the assigned reviewers set.

Aggregation of categorical distributions by reviewer profiles R_{tj}, $j = \overline{1, k}$ in the space of research groups from ANZSRC 2020 is implemented using the third stage of the above described categorization algorithm.

The number of reviewers for an application is denoted by $c = |S|$. This quantity is constant; usually it is from 2 to 5 people. The level of suitability between the application and the team of reviewers is calculated by formula (1).

REVIEWER ASSIGNMENT ALGORITHMS

The task of assigning reviewers from a mathematical point of view is to find a subset of fixed cardinality. To solve such problems in practice, mostly approximate algorithms are used. Among the set of possible algorithms, it is necessary to choose the one that provides a balance between assignment quality and efforts for solution finding. The following algorithms are proposed to be used.

Brute force. The best solution can be found by trivial brute force. For application A, among all possible teams of size c from the reviewers set **R,** a team with the maximum level of suitability has to be found. The complexity of brute force grows exponentially. The number of operations is proportional to the binomial coefficient: $\frac{n!}{(n-c)!c!}$. So even for medium-sized problems, it is unrealistic to walk through all possible options and adhere to some time constraints. Moreover, the number of options depends very much on the c.

Brute force on a truncated set of candidates. In practice, candidates with a low level of similarity are unlikely to be assigned as reviewers. Therefore, the rational step would be to ignore potential reviewers with very low similarity. By rejecting candidates with low similarity to the application, for example, at the level of 0.1 or 0.2, the search time can be significantly reduced. The number of operations is still proportional to the binomial coefficient but on a much smaller set of reviewers: $n \cdot p(r > truncation\_level)$, where $p(r > truncation\_level)$ is the probability that a reviewer r will have at least $truncation\_level$ similarity level with the application. The more we thin out the initial list of candidates, the shorter the duration of optimization will be, but the risks of deviating far from the optimum increase.

Pure greedy algorithm. The reviewers are assigned iteratively to ensure at each step the maximum suitability of the current fragment of the team to the application. The algorithm is performed in c iterations. At each iteration, one new member is added to the team of reviewers, who at this iteration maximizes the level of combined suitability of the current composition with the application. In the first iteration, we find the candidate with the highest similarity to the application. In the second iteration, we choose the candidate who, together with the already selected member of the team, has the highest suitability level to the application. The number of operations with this approach is significantly reduced and is proportional to n^c, but the solution may turn out to be suboptimal.

Greedy algorithm with elitism. The candidate with the highest value of suitability to the application is added first. At the same time, the level of combined suitability of updated reviewer team to the application is not taken into account. Other reviewers are assigned according to the pure greedy algorithm, that is, candidates are assigned who, in the current iteration, maximize the team's suitability level to the application. The greedy algorithm with elitism significantly shortens the duration of the optimization but still is proportional to n^{c-1}.

Assignment in isolation. The easiest way to assign reviewers is to choose those who are the most similar to the application. The combined suitability of the team is not taken into account. It is assumed that the stronger each of the candidates corresponds to the application, the better the team will be. Roughly speaking, the combined suitability level of the team is considered to be the sum of the similarity levels of each member. Algorithmically, assignment in isolation is implemented by sorting the candidates in descending order of similarity to the application and selecting the first c candidates. The number of operations is proportional to $n \cdot c$ in the best case. This is a very fast algorithm, but with a small chance of getting to the optimum.

DATASET FOR ASSIGNING EXTERNAL REVIEWERS

For experiments on the assignment of external reviewers, a dataset of PhD-thesis was collected [16]. For this, the information system of Ukrainian National Agency for Higher Education Quality Assurance was used. The collected theses belong to various research fields (Fig. 6) with the predominance of *Information Technologies*.

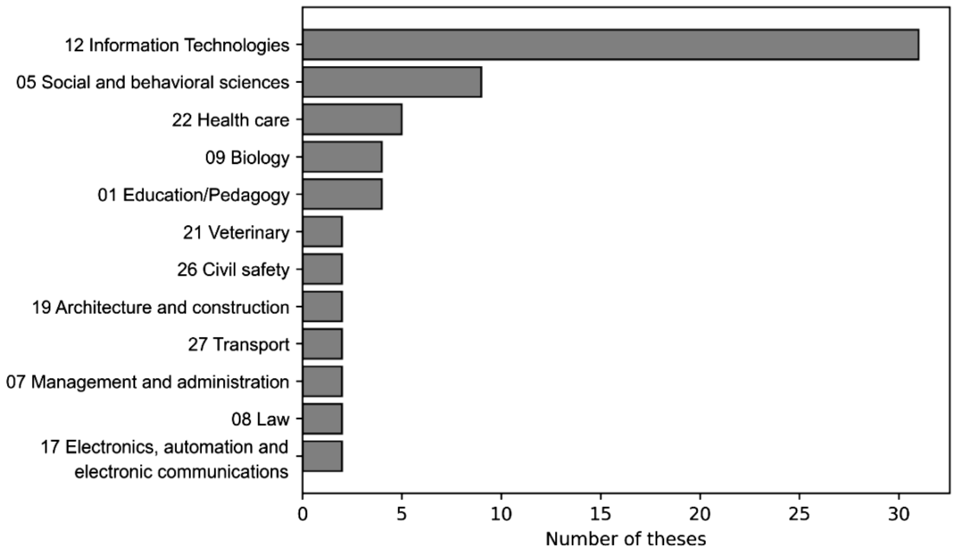

Fig. 6. PhD-theses distribution over research fields

EXPERIMENTS ON ASSIGNING EXTERNAL REVIEWERS

Experiments on external reviewers' assignment are conducted on the formed dataset of theses. At first, a thesis's keywords are categorized according to the keyword categorization algorithm within the research groups from ANZSRC 2020. Next, in a similar way, the keywords of the articles of the committees' members are categorized. Pairs of keywords are combined into additional queries only within one article. For each committee, the external reviewers are removed and new ones are assigned from other committees to maximize combined suitability. After removing the external reviewers, we get a set of fragments of committees, containing the chairman and two or one internal reviewers. The task is to find external reviewers whose addition to the fragments of committee ensures their maximum of combined suitability level to the topic of the theses.

The results of the reviewers' assignment are compared with the version of the committee, which is formed by the institution. The effect is estimated by an average level of change in the suitability level of committees:

$$E(F^{new}, F^{current}) = \frac{\sum\limits_{i=\overline{1,N}} (F_i^{new} - F_i^{current})}{\sum\limits_{i=\overline{1,N}} F_i^{current}} \cdot 100\%,$$

where N denotes number of theses; F_i^{new} denotes suitability level of the committee for i-th thesis after optimization, $i = \overline{1, N}$; $F_i^{current}$ denotes suitability level of the committee for i-th thesis before optimization, $i = \overline{1, N}$.

Fig. 7 presents the results of optimization using various assignment algorithms. Most of the committees from institutions have the suitability level

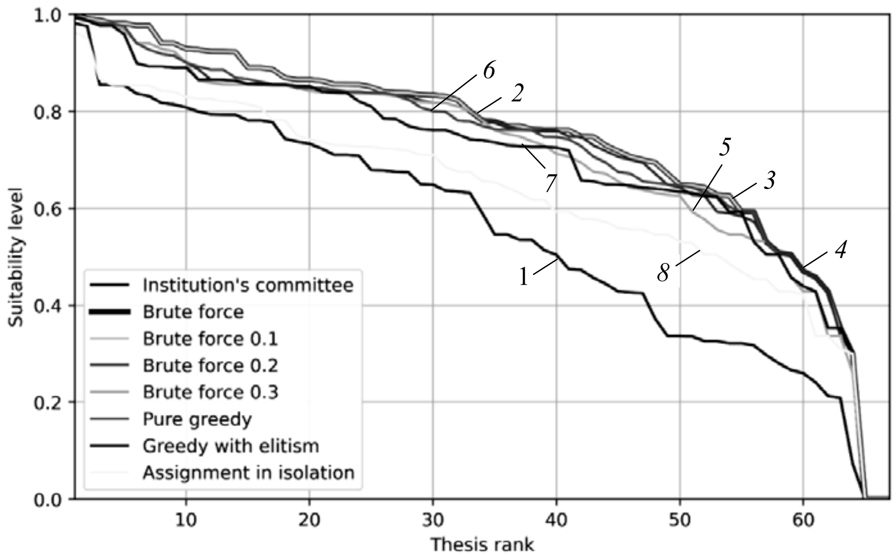

Fig. 7. Distribution of committees' suitability level depending on the algorithm used

above 0.2. The interquartile range is approximately equal to [0.4; 0.8]. With brute force there is a significant improvement in the suitability levels for the majority of committees. Some committees are not improved or the improvement level is low. This is due primarily to the fact that the distribution of theses by fields in the dataset is uneven and the dataset has a relatively small size. In almost all cases, committees from institutions have a lower suitability level to thesis than found by any assignment algorithm. By manually creating committees with limited opportunities for choosing committee's members, we get an average level of suitability to the thesis. On the other hand, with the automatic assignment of committee's members and a sufficiently large pool of candidates, we get a significant improvement of the committees only by changing external reviewers.

Fig. 8 compares suitability levels of committees' found by brute force with the committees found by other algorithms including brute force on a truncated set of candidates. Brute force on a truncated set of candidates with similarity threshold 0.1 performs almost identically as regular brute force, but the optimization time is reduced (Fig. 9). Brute force on a truncated set of candidates with similarity thresholds 0.2 and 0.3 performs very similar to the regular brute force, but there are a few suboptimal committees in both cases. Committees found by pure greedy algorithm are also suboptimal. Its performance is very close to the brute force 0.2 and is somewhat better than the brute force 0.3, but the time of optimization is significantly better (Fig. 9). Greedy algorithm with elitism performs slightly worse than pure greedy algorithm, there are slightly more suboptimal committees, but it is close to the brute force 0.3 with the optimization time reduced (Fig. 9). Under the assignment in isolation, most of the committees are suboptimal but it is the fastest among the algorithms (Fig. 9). This is due to the fact that the high similarity of a candidate with a thesis does not mean that the team formed by assignment in isolation covers the entire research groups' distribution of the thesis.

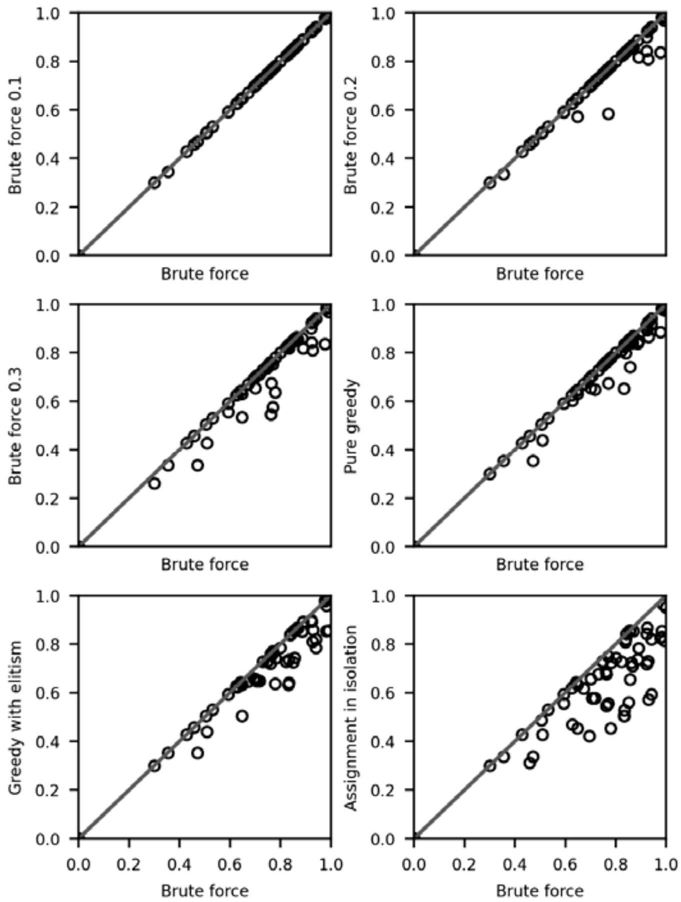

Fig. 8. Comparison of committees found by brute force with the committees found by faster algorithms

Fig. 9 compares the results of committees' assignments according to various optimization algorithms. Optimizing the truncated set of candidates with the similarity threshold of 0.3 is clearly unsuccessful. All others form a Pareto set. Therefore, when choosing an algorithm, it is necessary to take into account priorities, what is needed — a quick result or a high-quality one. From Fig. 9, it can be seen that the level of change due to the skip from pure greedy algorithm to brute force algorithms grows slowly. But the optimization time increases significantly. Therefore, the pure greedy assignment algorithm can be considered the most balanced. An alternative to it can be the brute force on truncated set of candidates with the similarity threshold in the vicinity of 0.25. These conclusions are based on experiments on a small dataset. With real databases of large volume, the optimization time by brute force algorithms can increase drastically.

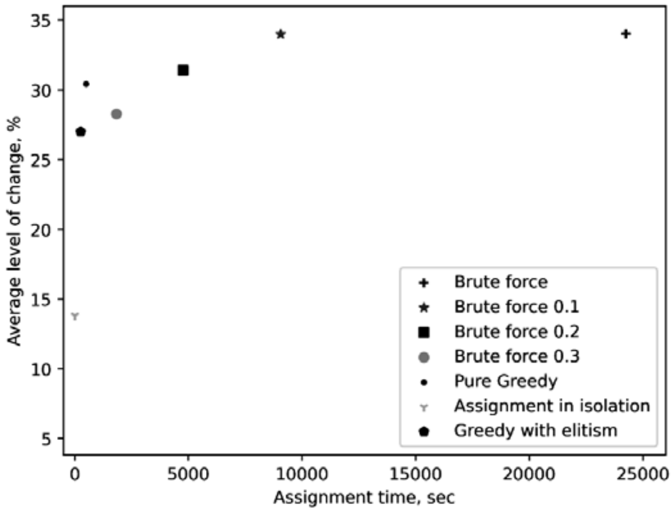

Fig. 9. Comparison of assignment algorithms according to the "duration — quality" criteria

AN EXAMPLE OF ASSIGNING A COMMITTEE

Let's consider an example of assigning a committee for the following thesis: "Models and methods of data processing of the system of remote monitoring of the condition of patients with diabetes". The thesis identifier in National Agency for Higher Education Quality Assurance is 4756.

The thesis's keywords are: edge devices; IoT; diagnostics; diseases; intelligent data analysis; information technologies; medical information systems; modeling; monitoring; data processing; patient; forecasting; software component model; system design; diabetes. After categorizing these keywords, we get the following result:

4605 Data Management and Data Science — 0.382;
4606 Distributed Computing and Systems Software — 0.255;
4609 Information Systems — 0.205;
4203 Health Services and Systems — 0.158.

The thesis is represented by the following vector:

$$A_t = \left(\frac{0.382}{4605}, \frac{0.255}{4606}, \frac{0.205}{4609}, \frac{0.158}{4203} \right)$$

In National Agency for Higher Education Quality Assurance, the research topics of each committee member are represented by the keywords of 3 or 4 of his/her papers. To categorize them, the principle of a bag of keywords is applied. Categorization of a member takes place as follows: 1) for each set of keywords of one paper, their paired combinations is created; 2) the received sets of keywords of different papers are combined into into one bag; 3) categorize the received set of keywords according to the algorithm [13]. The result of the committee categorization is as follows.

Research groups of the chairman are:

4609 Information Systems — 0.381;
4203 Health Services and Systems — 0.225;
4606 Distributed Computing and Systems Software — 0.214;
4601 Applied Computing — 0.180.

Suitability level of the chairman is:

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.381}{4609},\frac{0.225}{4203},\frac{0.214}{4606},\frac{0.180}{4601}\right)\right)=0.577\,.$$

Research groups of the first inner reviewer are:

4606 Distributed Computing and Systems Software — 0.337;
4605 Data Management and Data Science — 0.256;
4003 Biomedical Engineering — 0.244;
3208 Medical Physiology — 0.162.

Suitability level of the first inner reviewer is:

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.337}{4606},\frac{0.256}{4605},\frac{0.244}{4003},\frac{0.162}{3208}\right)\right)=0.564\,.$$

Research groups of the second inner reviewer are:

4606 Distributed Computing and Systems Software — 0.426;
4605 Data Management and Data Science — 0.299;
4003 Biomedical Engineering — 0.138;
4604 Cybersecurity and Privacy — 0.135.

Suitability level of the second inner reviewer is:

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.426}{4606},\frac{0.299}{4605},\frac{0.138}{4003},\frac{0.135}{4604}\right)\right)=0.521\,.$$

Research groups of the first external reviewer are:

3201 Cardiovascular Medicine and Haematology — 0.387;
3203 Dentistry — 0.215;
4605 Data Management and Data Science — 0.205;
4602 Artificial Intelligence — 0.192.

Suitability level of the first external reviewer is:

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.387}{3201},\frac{0.215}{3203},\frac{0.205}{4605},\frac{0.192}{4602}\right)\right)=0.239\,.$$

Research groups of the second external reviewer are:

4602 Artificial Intelligence — 0.435;
4611 Machine Learning — 0.357;
4605 Data Management and Data Science — 0.208.

Suitability level of the second external reviewer is:

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.435}{4602},\frac{0.357}{4611},\frac{0.208}{4605}\right)\right)=0.227\,.$$

The result of the committee aggregation is as follows:

$$Agg\begin{pmatrix}\left(\frac{0.381}{4609},\frac{0.225}{4203},\frac{0.214}{4606},\frac{0.180}{4601}\right)\\ \left(\frac{0.337}{4606},\frac{0.256}{4605},\frac{0.244}{4003},\frac{0.162}{3208}\right)\\ \left(\frac{0.426}{4606},\frac{0.299}{4605},\frac{0.138}{4003},\frac{0.135}{4604}\right)\\ \left(\frac{0.387}{3201},\frac{0.215}{3203},\frac{0.205}{4605},\frac{0.192}{4602}\right)\\ \left(\frac{0.435}{4602},\frac{0.357}{4611},\frac{0.208}{4605}\right)\end{pmatrix}=\left(\frac{0.389}{4606},\frac{0.374}{4605},\frac{0.236}{4602}\right).$$

The combined suitability level of the committee to the thesis is $Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.389}{4606},\frac{0.374}{4605},\frac{0.236}{4602}\right)\right)=0.631$. This is a relatively good suitability level, which is mainly due to the strong overlap in two of the four research groups.

Let's try to choose the best external reviewers to increase the combined suitability level. The members of all other committees of the dataset are used as candidates. As the result of brute force, the two new external reviewers are found. Their profiles are as follows: $\left(\frac{0.274}{3210},\frac{0.261}{4203},\frac{0.251}{4202},\frac{0.214}{3205}\right)$ with suitability level 0.158, and $\left(\frac{0.555}{4605},\frac{0.306}{4611},\frac{0.139}{4609}\right)$ with suitability level 0.542. After aggregating all members of the new committee we get the following categorization:

$$Agg\begin{pmatrix}\left(\frac{0.281}{4611},\frac{0.269}{4605},\frac{0.228}{4602},\frac{0.222}{4608}\right)\\ \left(\frac{0.556}{4612},\frac{0.302}{4602},\frac{0.142}{4007}\right)\\ \left(\frac{0.457}{4611},\frac{0.196}{4603},\frac{0.183}{4605},\frac{0.163}{4609}\right)\\ \left(\frac{0.274}{3210},\frac{0.261}{4203},\frac{0.251}{4202},\frac{0.214}{3205}\right)\\ \left(\frac{0.555}{4605},\frac{0.306}{4611},\frac{0.139}{4609}\right)\end{pmatrix}=\left(\frac{0.361}{4605},\frac{0.335}{4606},\frac{0.156}{4609},\frac{0.148}{4203}\right).$$

The combined suitability level of the new committee to the thesis is

$$Fit\left(\left(\frac{0.382}{4605},\frac{0.255}{4606},\frac{0.205}{4609},\frac{0.158}{4203}\right),\left(\frac{0.361}{4605},\frac{0.335}{4606},\frac{0.156}{4609},\frac{0.148}{4203}\right)\right)=0.923.$$

Comparing with the initial committee, a significant improvement in the level of suitability is observed, the new committee has the same research groups as the thesis. The improvement is about 46%.

From the given example, it can be seen that although the individual similarity of an individual member of a committee may be mediocre, the overall suitability level of the committee may turn out to be high. This is due to the fact that the new external reviewers cover the so-called minor part of the thesis topic, which is outside the field of expertise of other committee members. This is clearly visible on Fig. 10 where the difference between the distributions of thesis, institution's committee and proposed committee is shown. The thesis and proposed committee intersect in all their research groups. The institution's committee lacks the research groups *4609 Information Systems* and *4203 Health Services and Systems,* which makes it less similar to the thesis's research field.

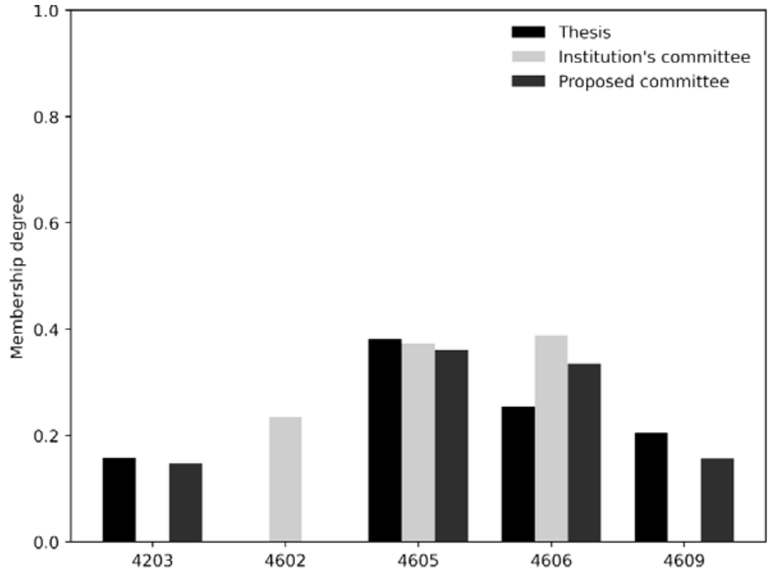

Fig. 10. Comparison of initial committee and proposed committee

CONCLUSIONS

The paper proposes an express method of assigning the external reviewers for PhD defense committee. On the first stage of assignment, the application and potential reviewers are categorized by presenting their profiles as vectors in the space of research groups from ANZSRC 2020. At the second stage, the suitability levels of potential reviewers to the application topic are calculated, taking into account the kinship of research groups. At the third stage, a team of reviewers is assigned, which corresponds to the topic of the application to the maximum possible extent. To implement the third stage, the various optimization algorithms are proposed: brute force, brute force on a truncated set of candidates, greedy algorithm without elitism and with elitism, and on assignment in isolation. Experiments on the dataset of 67 PhD theses showed that the best balance in terms of assignment quality criteria and team searching duration provides greedy algorithm without elitism and brute force on a truncated set of candidates. As a result of the optimization, it was possible to improve the combined quality of committees by an average of 13–34% over all the dataset, depending on the type of algorithm used. Optimizing the truncated set of candidates with the similarity threshold of

0.3 is clearly unsuccessful. All others form a Pareto set. Therefore, when choosing an algorithm, it is necessary to take into account priorities, what is needed — a quick result or a high-quality one.

The proposed method can be used to improve the efficiency of managing the processes of assigning reviewer teams in various fields, for example, for evaluation of grant applications. The method can also be used for auditing to quickly check the correctness of the assigned committees with subsequent thorough resource-intensive examination of suspicious cases.

Further research may include: studying whether using Large Language Models is a better choice for modeling the keywords representation than the proposed method; using the proposed method of express assignment in more time-consuming and iterative procedures for assigning a team of reviewers, when it is necessary to take into account not only the relevance of the topic of the application, but also the absence of a conflict of interests, the balance of the load on the reviewers, and other possible limitations. It is advisable to take into account not only the relevance of the subject of the reviewers and the application, but also the qualification level of the experts during the assignment.

Acknoledgment. The authors are grateful to Digital Science & Research Solutions Inc. for the provision of access to Dimensions as part of the DIM-371 project.

REFERENCES

1. F. Wang, N. Shi, B. Chen, "A comprehensive survey of the reviewer assignment problem," *International Journal of Information Technology and Decision Making*, 9(4), pp. 645–668, 2010. doi: https://doi.org/10.1142/S0219622010003993
2. M. Aksoy, S. Yanik, M.F. Amasyali, "Reviewer assignment problem: A systematic review of the literature," *Journal of Artificial Intelligence Research*, vol. 76, 2023. doi: https://doi.org/ 10.1613/JAIR.1.14318
3. S. Tan, Z. Duan, S. Zhao, J. Chen, Y. Zhang, "Improved reviewer assignment based on both word and semantic features," *Information Retrieval Journal*, 24(3), pp. 175–204, 2021. doi: https://doi.org/10.1007/s10791-021-09390-8
4. D. Yarowsky, R. Florian, "Taking the load off the conference chairs: Towards a digital paper-routing assistant," *Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, EMNLP 1999*, pp. 220–230.
5. M. Karimzadehgan, C.X. Zhai, G. Belford, "Multi-aspect expertise matching for review assignment," *Proceedings of International Conference on Information and Knowledge Management*, pp. 1113–1122, 2008. doi: https://doi.org/10.1145/ 1458082.1458230
6. M. Mirzaei, J. Sander, E. Stroulia, "Multi-aspect review-team assignment using latent research areas," *Information Processing and Management*, 56(3), pp. 858–878, 2019. doi: https://doi.org/10.1016/j.ipm.2019.01.007
7. E. Ekinci, S.I. Omurca, "NET-LDA: A novel topic modeling method based on semantic document similarity," *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(4), pp. 2244–2260, 2020. doi: https://doi.org/10.3906/ELK-1912-62
8. O. Anjum, H. Gong, S. Bhat, J. Xiong, W.M. Hwu, "Pare: A paper-reviewer matching approach using a common topic space," *EMNLP-IJCNLP 2019 – 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 518–528. doi: https://doi.org/ 10.18653/v1/d19-1049
9. C. Sun, K.T.J. Ng, P. Henville, R. Marchant, "Hierarchical word mover distance for collaboration recommender system," *Communications in Computer and Information Science*, vol. 996, pp. 289–302. Springer Verlag, 2019. doi: https://doi.org/10.1007/978-981-13-6661-1_23
10. X. Kong, H. Jiang, Z. Yang, Z. Xu, F. Xia, A. Tolba, "Exploiting publication contents and collaboration networks for collaborator recommendation," *PLOS One*, 11(2): e0148492, 2016. doi: https://doi.org/10.1371/journal.pone.0148492
11. B. Bhaisare, R. Bharati, "Advancing Peer Review Integrity: Automated Reviewer Assignment Techniques with a Focus on Deep Learning Applications," in *A.K. Bairwa, V. Tiwari, S.K. Vishwakarma, M. Tuba, T. Ganokratanaa, (eds) Computation of Artificial Intelligence and*

Machine Learning. ICCAIML 2024. Communications in Computer and Information Science, vol 2184. Springer, Cham, 2024. doi: https://doi.org/10.1007/978-3-031-71481-8_25

12. Y. Zhao, J. Tang, Z. Du, "EFCNN: A restricted convolutional neural network for expert finding," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11440 LNAI, pp. 96–107. Springer Verlag, 2019. doi: https://doi.org/10.1007/978-3-030-16145-3_8
13. S. Shtovba, M. Petrychko, "Topic modeling of researchers based on their interests from Google Scholar," *System Research and Information Technologies*, no. 2, pp. 113–129, 2021. doi: https://doi.org/10.20535/SRIT.2308-8893.2021.2.09
14. S. Shtovba, M. Petrychko, "Jaccard index-based assessing the similarity of research fields in dimensions," *CEUR Workshop Proceedings*, vol. 2533, pp. 117–128, 2019.
15. S. Shtovba, M. Petrychko, O. Shtovba, "Similarity metric of categorical distributions for topic modeling problems with akin categories," *CEUR Workshop Proceedings*, vol. 3392 "The Sixth International Workshop on Computer Modeling and Intelligent Systems", pp. 76–85, 2023. doi: https://doi.org/10.32782/cmis/3392-7
16. M. Petrychko, S. Shtovba, "Dataset for PhD theses reviewers' assignments," *ResearchGate*, 2024. doi: http://dx.doi.org/10.13140/RG.2.2.23147.35362

Received 25.11.2024

INFORMATION ON THE ARTICLE

Serhiy D. Shtovba, ORCID: 0000-0003-1302-4899, Vasyl' Stus Donetsk National University, Vinnytsia National Technical Universiry, Ukraine, e-mail: s.shtovba@donnu.edu.ua

Mykola V. Petrychko, ORCID: 0000-0001-6836-7843, Vinnytsia National Technical Universiry, Ukraine, e-mail: mpetrychko@vntu.edu.ua

АЛГОРИТМИ ПРИЗНАЧЕННЯ ЗОВНІШНІХ РЕЦЕНЗЕНТІВ ДЛЯ ЗАХИСТУ PHD-ДИСЕРТАЦІЙ / С.Д. Штовба, М.В. Петричко

Анотація. Запропоновано підхід до призначення зовнішніх рецензентів. У ньому враховується лише семантична схожість між заявками та рецензентами, оцінюються індекси схожості та призначається необхідна кількість таких рецензентів, за яких забезпечується максимальний рівень відповідності рецензентів заявці за деякими критеріями. Виконано порівняльний аналіз різних алгоритмів оптимізації за критерієм «якість призначення – тривалість оптимізації». Експерименти на тестовому датасеті показали, що прийнятний баланс за критеріями «якість призначення» та «тривалість оптимізації» для призначення зовнішніх рецензентів забезпечує жадібний алгоритм без елітизму та за повного перебору на проріджений множині кандидатів. Застосування запропонованих алгоритмів покращує якість роботи докторських рад в середньому на 13–34% за усього набору даних, залежно від типу використовуваного алгоритму.

Ключові слова: зовнішні рецензенти, задача призначення рецензентів, категоризація, оптимізація, повний перебір, жадібний алгоритм, ізольоване призначення, PhD-дисертація, Dimensions, ANZSRC 2020, галузь досліджень.