Artificial Intelligence Research and Development
I. Sanz et al. (Eds.)

doi:10.3233/FAIA230659

Expert Finding for Citizen Science

Bjoern KOMANDER [a,c,1], Jesus CERQUIDES [a], Jaume PIERA [b], Jeffrey CHAN [c] and Azadeh ALAVI [c]

[a] *IIIA-CSIC, Campus UAB, 08193 Cerdanyola, Spain*
[b] *EMBIMOS, ICM-CSIC, Barcelona, Spain*
[c] *School of Computing Technologies, RMIT University*

Abstract. Citizen Science brings together scientists and public participants to collaborate on a wide range of applications and fields. With this approach, Citizen Science advances scientific research and communication while accounting for various stakeholders. Across many Citizen Science projects, digital discussion platforms play an essential role for self-governance and self-organisation. In order to increase the quality of the discussions held on these platforms, we propose a model that recommends users to new discussions in which they are likely to contribute meaningful content. Our model learns relevant user representations based on the quality of past interactions between users and discussion threads, as well as the text content of questions, using a ranking loss function, an approximation of the NDCG metric, and matrix factorization. We demonstrate that our approach is able to predict potential experts on unseen discussion threads and outperforms several baselines. Compared to state-of-the-art expert finding techniques, the architecture of our model is significantly less complex, while focusing on a mostly overlooked ranking loss function.

Keywords. Expert Finding, Citizen Science, Learning-To-Rank, Online Discussion, Natural Language Processing

1. Introduction

Citizen Science (CS) [14] is scientific research conducted with participation from the public (who are sometimes referred to as amateur/nonprofessional scientists). An extension of this, Extreme Citizen Science (ExCiteS) [3], is a situated, bottom-up practice that takes into account local needs, practices and culture and works with broad networks of people to design and build new devices and knowledge creation processes to advance science and potentially help the community. In the quest for ExCiteS, digital tools which allow for discussion and deliberation among citizens are essential. Rich deliberation will enable proper self-governance for the CS project and will potentially increase the impact of the project on future policy[12].

Nowadays, the state-of-the-art in digital debate tools in CS projects range from non-existent for many small projects, to discourse[2] forums for some of the largest projects, such as iNaturalist[13][3] or Sensor.Community [16].

[1]Corresponding Author: Bjoern Komander, bkomander@iiia.csic.es
[2]https://discourse.org/
[3]https://www.inaturalist.org/

These discussion forums provide a tool to organise discussions and questions, where users can post a question or start a discussion in a *topic* across several *categories*. Other users can contribute to discussions or answer questions by posting within topics and yet other users can signal the usefulness of posts by awarding likes. Such discussion platforms play a vital role in the inclusive participation of all users in the Citizen Science process.

However, with increasing usage and content, these platforms become more and more cluttered, which can lead to several inefficiencies. As the amount of content continues to increase, potential answerers - i.e., participants who can contribute relevant answers - must filter through a larger volume of material to locate the post that aligns with their expertise. At the same time, question askers will have to wait for a longer time to receive the best (right) answer for their question. Tackling this problem by directing a new question to the right users and as quickly as possible is a prominent research area, called *expert finding* or *question routing*. While expert finding is a common research setting for online platforms like Stackoverflow and Yahoo! Answers, it has not yet been applied to Citizen Science discussion platforms. Particularly in these platforms, we can expect a very heterogenous distribution of knowledge across users with varying levels of expertise. An efficient allocation of users knowledge to the right content can therefore enhance the overall interaction with the platform and the inclusive participation of all users.

In this paper, we propose a model to enhance user engagement in Citizen Science discussion fora, by identifying potential experts for newly asked questions. Our approach combines multiple state-of-the-art concepts from expert finding to provide a ranking of potential experts for a new question. We utilize the question content, represented by Sentence-Bert embeddings [10], and past user-question interaction data to estimate the expertise of users and their likelihood of contributing valuable insights. Our model learns user embeddings by factorizing the interaction matrix using a ranking loss function. Unlike other expert finding methods, our model is computationally simple, yet outperforms several baselines. Moreover, our model can recommend experts for questions asked by unseen users, enabling it to generalize to new situations.

The remainder of this paper is structured as follows. Section 2 discusses related papers from the field of expert finding. Section 3 presents our concise problem and discusses the methodology of our approach. Section 4 introduces the data we obtained from several Citizen Science discussion platforms and section 5 presents our results. We conclude this work with a short discussion in section 6.

2. Related Work

To the best of our knowledge, there is only a small number of papers that address expert finding in the Citizen Science context. However these papers propose methods to find potential experts for the specific task of the Citizen Science platform, like the identification of insects in Spipoll [11]. We deviate from this problem as we seek to find experts for the respective Citizen Science discussion platform. This is more related to the task of expert finding/ question routing in community question answering, a sub-field of recommender systems. On community question answering websites such as StackOverflow and Yahoo! Answers, expert finding aims to identify user(s) with the adequate knowledge to answer a new question. The proposed methods in this field differ by the considered signals, i.e., the way they model questions and their content, as well as the modeling of users, the interactions between different users and the interaction between potential experts and questions. Moreover, these works propose a range of methods to make use of the different signals, from purely ranking and no learning algorithms such as PageRank to deep learning applications such as Convolutional Neural Networks. We categorize the different works into approaches that uses graphs to reflect the expertise of users, approaches primarily focused on the ranking objective, and deep learning approaches.

For the class of proposed methods which represent the interaction signals between questions and users, and among users in the form of graphs, Zhang et al. [17] note that these graphs differ from other online community graph structures, such as social networks like Twitter, as they are not based on social relationships but rather on the shared interests of the users. For instance Zhang et al. [17] and Wang et al. [15] construct a graph by connecting users with a directed edge from the author of a post to the answerer. Others [4,7] connect users if they answered the same question.

To make appropriate use of the graph structure, Wang et al. [15] apply an adapted version of PageRank, where they model a users expertise authority by the links pointing towards the user. Then, they combine the expert authority with a measure of expertise relevance by using text similarity based on TF-IDF embeddings. For each expert, they create an embedding for all past answered questions and then compare this to the new question through a distance measure. While they find their approach to be effective, they do not provide comparisons to other baselines or other discussion forums.

Similarly, Ghasemi et al. [4] use text similarity between a new question and all answers of existing questions in combination with user similarity between question asker and potential experts. They use Node2Vec to build user embeddings for all users in their graph and propose to use Word2Vec or TF-IDF for text embeddings. Then, they use a weighted sum to combine these two measures to create a ranking across potential experts for a new questions. They find that including the graph embeddings approach significantly increases the model's performance across all metrics. However, potential improvements of graph embeddings depend on the properties of the considered graph, as they find larger improvements for graphs with more links and nodes. Additionally, this approach requires sufficient information about all users for the user embeddings, and fails if an unseen user asks a new question.

With a learning-to-rank (LTR) approach, Ji et al. [5] apply RankingSVM, in a pairwise ranking approach, on user specific features, such as their activity and sum of best answers, and question specific features, such as text content. This approach is outperformed by Cheng et al. [2], who make use of LambdaMart, a popular Learning-To-Rank algorithm. They model expert finding as a multi-objective ranking problem, where they optimize answer quality and answer possibility of a potential expert at the same time. Their results suggest a very strong increase in performance over previous models.

Recently the expert finding task received increased attention in researchers that employ deep learning models: Li et al. [6] use a metapath guided network embedding to jointly embed information about the question raiser, question answerer and the question content. They then apply a Convolutional Neural Network ranking model to predict the expertise across users for a new question. Again, as their model relies on data of the question raiser, it does not work for unseen or new question askers.

Liu et al. [8] aimed to improve the quality of text representations by retraining BERT embeddings with a label augmented masked language modelling task. To achieve this, they concatenated a user's past answers to a new question and used the binary label "expert" or "no expert" as the special CLS token.. They then fine tune the model on a binary expert / no expert classification task and report the best performance compared to a wide range of other models. For instance their model outperforms the approach by Li et al. [6] and Ghasemi et al. [4]. However, it is important to note that the pre-training stage of the model is computationally expensive. Additionally, the use of a binary classification in both pretraining and finetuning might limit the models ability to learn more meaningful text representations. When only considering one expert per question, the potentially helpful information of other answers is ignored. In general, it remains an open question if a ranking over more potential experts outperforms a binary classification for expert finding.

The existing literature in expert finding lacks a clear consensus on the optimal approach for representing questions, users, and their interactions. There is no evidence that graph based approaches outperform models that focus more on the text content and vice versa. Developing meaningful user representations that reflect their expertise based on their past interactions with questions is a major challenge in the expert finding literature. While previous studies have relied on graph-based methods to generate user representations, we propose a novel model that directly learns user embeddings by taking into account the quality of past interactions and the text representation of past questions. Our approach offers a simple and flexible way of learning meaningful user representations that help to recommend experts for unseen questions. Importantly, in comparison to graph-based user embeddings, our method does not fail to predict for unseen users. Furthermore, we integrate a ranking objective directly into the learning of user embeddings. This allows us to consider multiple potential experts for each question while learning user embeddings, through which we can better capture the complexity of the expert finding task.

3. Methodology

3.1. Problem Statement

Consider a (potentially non-finite) set of possible questions $\mathscr{Q}$, and a probability distribution P over $\mathscr{Q}$. Furthermore, consider a set of users $U = \{u_1, u_2, \ldots, u_n\}$, and a set of already answered questions $Q = \{q_1, q_2, \ldots, q_m\}$ which is a sample of P of size m. Let X_{ij} denote the matrix that reflects the expertise of user i for existing question j. For a new question q_{new}, we want to identify the user(s) with the adequate knowledge to answer q_{new}. We therefore define a function *f(text)* that, given the text features of q_{new} and the observed X_{ij}, creates a ranking across the set of users U, while maximizing the expected quality of the produced ranking:

$$\arg\max_{f} \mathop{\mathbb{E}}_{q\sim P} \kappa(f(q)), \tag{1}$$

where κ is a function to assert the quality of the produced ranking.

For the purposes of this study, we use the term "question" to refer to all types of initial posts in online discussion fora, including statements of opinion or requests for features. While for Community Question Answering fora (CQA), initial post are typically phrased as questions, in the Citizen Science discussion platforms, they may take on various forms. In this work we refer to all types of initial posts as questions, as we want to find the user with the proper knowledge to contribute to a post/discussion, regardless of the type. Moreover there might be more than one user with the appropriate knowledge for a given question. While many existing approaches view expert finding as a binary classification problem due to the prevalence of accepted answers on CQA platforms, our focus is on ranking potential experts based on their likelihood of providing a high-quality answer to a new question. By adopting this approach, we can identify multiple users who are well-suited to the task of expert finding, which is a critical step in encouraging greater participation among platform users.

In this paper, we define the knowledge adequacy $\tilde{X}_{ij}$ of user i for question j as

$$\tilde{X}_{ij} = \begin{cases} 0 & \text{if user did not engage} \\ \frac{\sum_{p\in R_{ij}} \text{likes(p)}+1}{\sum_{p\in R_j q(p)=j} \text{likes}(p)+1} & \text{otherwise,} \end{cases} \tag{2}$$

where R_{ij} is a set containing the responses to question j by user i and $R_j = \bigcup_{i\neq a(j)} R_{ij}$ is a set containing every response p provided to question j except those coming from the author. To model the difference of users who posted to a question but did not receive any likes from users that did not engage at all, we add one like for each user composing a response. We then normalize the adequacy to arrive at expertise matrix X_{ij} of user i for existing question j.

$$X_{ij} = \frac{\tilde{X}_{ij}}{\max_{i\neq a(j)} \tilde{X}_{ij}}. \tag{3}$$

Furthermore, in this paper we evaluate the quality of a produced ranking κ with the Normalized Discounted Cumulative Gain (NDCG), which can be defined as

$$NDCG(X_j, \hat{X}_j) = \frac{DCG(X_j, \hat{X}_j)}{DCG(X_j, X_j)}, \tag{4}$$

where DCG is defined as

$$DCG(X_j, \hat{X}_j) = \sum_{i=1}^{n} \frac{2^{X_{ij}}-1}{\log_2(1+\pi_i)}. \tag{5}$$

To make the DCG measure comparable across different questions, each DCG measure is normalized by its ideal DCG, to arrive at NDCG. This measure reflects the quality of a predicted ranking by considering both the discounted rank of the prediction, π_i, as well as the relevance scores, in our case the adequacy scores, in the original ranking as $2^{X_{ij}}$.

3.2. Proposed Model

In this paper we propose a function that produces a ranking across potential experts in the form of

$$f(q,W) = \text{rank}(W \times H_q), \tag{6}$$

where W denotes a user embedding matrix and H_q denotes the encoded text representation of question q. As we want to avoid the shortcomings of methods that require exhaustive information about the question asker, and therefore fail for unseen users, we concentrate on information available for all users.

To model the question content H, we use Sentence-BERT, a computationally efficient adaptation of BERT that extracts semantically meaningful document representations [10]. The resulting question representation matrix has dimensions of *questions x 768*. We append an additional column of 1's to the question representation matrix, which serves as a constant term or bias for the subsequent learning process.

The main challenge is to develop a representation of the potential experts that reflects the relationship between users and questions, so that we can adequately predict their expertise for a new question. While previous studies have often modeled users using node embeddings in graph-based approaches, we propose factorizing the adequacy matrix X using a user embedding W and the encoded text representation H:

$$X = WH^T. \tag{7}$$

Therefore we can learn a user representation $\hat{W}$, which minimizes a loss $\mathscr{L}$ between X and the resulting $\hat{X}$. As our objective in predicting experts is to maximize the quality of the produced ranking in $\hat{X}$ denoted by κ, we can rewrite the objective in estimating $\hat{W}$ as

$$\hat{W} = \min \sum_{q}^{Q_t} \kappa(\hat{W}H^T, X). \tag{8}$$

Essentially we want to estimate $\hat{W}$ that maximizes the ranking quality κ of $\hat{X}$ with respect to X across all training questions $q \in Q_t$.

3.3. Loss Function

Previously, we defined κ as the NDCG. However since this metric, like other ranking metrics, is not differentiable, it cannot be used for direct optimization. As a result, there have been several approaches proposed in recent years to overcome this problem in listwise Learning-To-Rank (LTR) algorithms. Notably, algorithms like LambdaMART and LambdaRank, both based on Ranknet, work with gradient boosting or gradient optimization to optimize listwise ranking and related metrics, like the NDCG. For instance, LambdaMart uses a decision tree ensemble and gradient boosting to optimize pairwise differences in relevance scores between pairs of items. Other approaches aim to approximate the ranking metrics by a smooth and differentiable function. Recently, an older approach proposed by Qin et al.[9] gained more attention from researchers in the field, including [1]. Qin et al.[9] propose an approximation of NDCG, that can be applied to other ranking metrics, where the rank π in the DCG metric is approximated by

$$\hat{\pi}_i = 1 + \sum_{k \neq i} \frac{1}{1 + \exp^{-\alpha(t-s)}}. \tag{9}$$

This approximation is now differentiable and thus can be optimized using (stochastic) gradient descent. We therefore define our loss function as

$$\mathscr{L}(X, \hat{W}H^T) = \min \sum_{q}^{Q_t} \hat{\kappa}(X, \hat{W}H^T), \tag{10}$$

where $\hat{\kappa}$ is the approximation of the NDCG. At the question level, the loss is defined as

$$\mathscr{L}(X_j, \hat{X}_j) = -\frac{1}{DCG(X_j, X_j)} \sum_{i=1}^{n} \frac{2^{X_{ij}} - 1}{\log_2(1 + \hat{\pi}_i)}. \tag{11}$$

3.4. Evaluation Metrics

The proposed approach is evaluated using several metrics including the NDCG at various cutoffs, namely for the top 5, 10, and 20. We also evaluate the performance using Mean Average Precision at k (MAP@k), where we compute the mean the average precision at k for all questions, as the intersection of ranking predictions and actual ranking of users at the cutoff k. Moreover, we use the Mean Recipropal Rank (MRR) as

$$MRR = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\pi_i}, \tag{12}$$

where π_i denotes the predicted rank of the actual top ranked experts, across questions $j \in \{j_1, j_2, \ldots, J\}$

3.5. Baselines

We compare the model's performance to several baselines. The *CONSTANT* baselines predict the same ranking for all questions, based on the sum of training adequacy scores. Additionally we include an adaptation of the baseline *score*, where for a new question q_{new}, we weight the adequacy score for each training question by a similarity function $s()$ between the new the question and the respective text question

$$f(q) = \text{rank}(\sum_{j=k}^{J} s(q_{new}, q_i) \times X_{.i}). \tag{13}$$

We implement two different similarity functions, *EUC*

$$s_1(q_i, q_j) = e^{-d(H_{.i}, H_{.j})}, \tag{14}$$

and *COS*, where s_2 is defined as the cosine similarity between q_j and q_i.

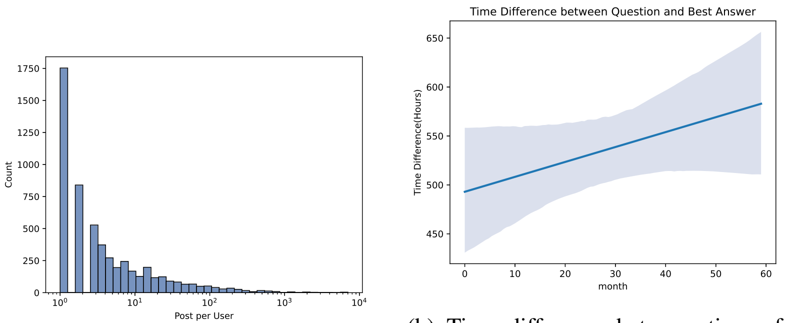
Name	Language	Active Users	Total Posts	Total Questions	Categories
iNaturalist	English	5568	161829	11886	9

Table 1. Characteristic of iNaturalist Discussion Forum.

4. Data

We crawled several citizen science discussion fora, based on Discourse. We choose the largest forum, iNaturalist for training our model and evaluating our approach. Here users can post and discuss across several topics, such as *nature talk* and **feature requests**. An overview of the basic characteristics of iNaturalist is presented in table 1.

Similar to other online community discussion / question answering platforms, the iNaturalist discourse platforms consist of a small number of very active users as visible in figure 1a.



(a) Distribution of posts per user.

(b) Time difference between time of question and best answer in hours.

Figure 1.

The number of posts made by users can vary significantly, suggesting a difference in expertise levels. As shown in Figure 1b, we also observe a noticeable increase in the average wait time for obtaining a *best answer*, which is defined as the answer receiving the most likes, and it is noteworthy that this average time is already relatively high. To address this inefficiency, our model aims to route new questions to users who are most likely to provide meaningful answers, thus optimizing the process and reducing wait times. This approach could be implemented in the iNaturalist forum and other CS-fora by notifying users who our model identifies as top experts for a new question.

4.1. Experiments

After removing all questions with no answers, as well as any potential expert, where the sum of adequacy scores across all questions is below 1, we arrive at 11652 questions and 1625 potential experts. We split the questions into 80% training 10% validation and 10% testing using 5 folded cross validation.

We train the proposed model using stochastic gradient descent. Specifically we train each split for 150 epochs, with a batch size of 64 and a learning rate of 0.5. Our experiments suggest that training the model for more than 150 epochs will result in overfitting.

Model	MRR	MAP@1	MAP@3	MAP@5	MAP@10	MAP@20	NDCG	NDCG@5	NDCG@10	NDCG@20
MF	**0.25***	**0.129***	**0.225****	**0.227**	**0.184****	**0.132****	**0.5****	**0.325****	**0.368****	**0.407****
	(0.013)	(0.011)	(0.004)	(0.003)	(0.002)	(0.005)	(0.008)	(0.007)	(0.007)	0.007
COS	0.234	0.117	0.211	0.224	0.177	0.129	0.474	0.296	0.336	0.375
	(0.01)	(0.009)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.006)	(0.005)	(0.004)
EUC	0.241	0.124	0.214	0.219	0.179	0.129	0.486	0.307	0.35	0.39
	(0.01)	(0.009)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.007)	(0.004)	(0.005)
CONSTANT	0.226	0.105	0.2	0.224	0.177	0.128	0.468	0.291	0.331	0.369
	(0.009)	(0.005)	(0.017)	(0.002)	(0.003)	(0.002)	(0.004)	(0.007)	(0.006)	(0.005)

Table 2. Performance of several Models across Ranking Metrics with Standard Deviation in parenthesis. Best model for each score highlighted in bold. Stars indicate significance level of improvement over baselines at $p < 0.05$. * for baseline *CONSTANT*, ** for baseline *CONSTANT* & best baseline

5. Results

To assess the performance of our models we evaluate them on the previously discussed ranking metrics, presented in table 2. For each model we report the mean and standard deviation across the cross validation splits. We refer to our proposed model, the matrix factorization model as *MF*. Furthermore for the proposed baselines based on the distance functions, *EUC* denotes the baseline utilising the euclidean distance and *COS* the cosine similarity.

As it can be observed from the above table, our proposed model outperforms all the baselines across all metrics. For example, on average the most relevant expert is predicted at fourth position. However it should be noted that the other baselines already work reasonably well without any learning. All the baselines of comparison heavily rely on the sum of the adequacy score, which results in favoring the users that post more. While post frequency could be useful in indicating overall knowledgeability, we have found that incorporating question title embeddings into the *EUC* and *COS* models yields significant improvements over the *BASE* model. By further enhancing the signals from question titles, our approach has the potential to learn more meaningful user representations.

5.1. Quality of Predictions

We can further asses the quality of predictions across several models and several structural properties of iNaturalist. As figure 2 illustrates, the prediction quality for test questions differs across iNaturalist categories. For example all models perform relatively better for the categories *Bug Reports* and *Feature Requests*, categories where we would expect a well defined distribution of knowledge across users. We assume that these categories are more similar to community question answering fora such as Stackoverflow, where there are unique best answers to questions. However for the categories *Nature Talk*, *Curators* and *Tutorials* all models perform relatively bad. The knowledge for questions within these topics might be more evenly distributed, as these questions are usually more open. Interestingly, our model *MF* significantly outperforms the baselines in the categories *Nature Talk, News and Updates* and *Curators*.

6. Discussion

With this paper we present a model that can predict users with the adequate knowledge to contribute to new questions or posts in Citizen Science discussion platforms. We use

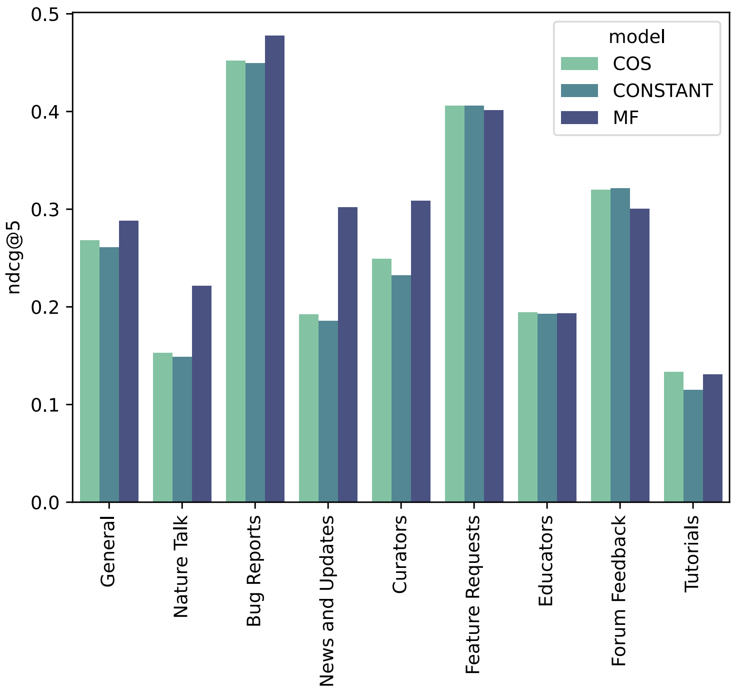


Figure 2. Differences in NDCG@5 across several categories and models

state-of-the-art techniques from the Expert Finding literature, including a novel Ranking Loss function and Sentence-BERT embedding with the goal to increase the usability of these discussion platform. Our proposed model outperforms several baselines while being significantly less computationally complex than state-of-the-art models. Moreover our model performs well on unseen question askers and is able to predict several potential experts for new questions. We believe that expert finding for online discussion can be a helpful tool for improving the quality and utility of these discussions. Our proposed model is adaptable and can be readily implemented on various online discussion platforms. However, our model can be improved by creating more meaningful question content representations and including more structural signals from the discussion fora.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Nº 101034328, HumanE-AI-Net project N. 952026, and from the Horizon Europe programme project GUARDEN 101060693.

References

[1] Sebastian Bruch, Masrour Zoghi, Michael Bendersky, and Marc Najork. Revisiting Approximate Metric Optimization in the Age of Deep Neural Networks. In *Proceedings of the 42nd International ACM SIGIR*

Conference on Research and Development in Information Retrieval, SIGIR'19, pages 1241–1244, New York, NY, USA, July 2019. Association for Computing Machinery.

[2] Xiang Cheng, Shuguang Zhu, Sen Su, and Gang Chen. A Multi-Objective Optimization Approach for Question Routing in Community Question Answering Services. *IEEE Transactions on Knowledge and Data Engineering*, 29(9):1779–1792, September 2017. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[3] Rafael Morais Chiaravalloti, Artemis Skarlatidou, Simon Hoyte, Marcos Moreu Badia, Muki Haklay, and Jerome Lewis. Extreme citizen science: Lessons learned from initiatives around the globe. *Conservation Science and Practice*, 4(2):e577, 2022.

[4] Negin Ghasemi, Ramin Fatourechi, and Saeedeh Momtazi. User Embedding for Expert Finding in Community Question Answering. *ACM Transactions on Knowledge Discovery from Data*, 15(4):70:1–70:16, March 2021.

[5] Zongcheng Ji and Bin Wang. Learning to rank for question routing in community question answering. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, CIKM '13, pages 2363–2368, New York, NY, USA, October 2013. Association for Computing Machinery.

[6] Zeyu Li, Jyun-Yu Jiang, Yizhou Sun, and Wei Wang. Personalized Question Routing via Heterogeneous Network Embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):192–199, July 2019. Number: 01.

[7] Chen Liu, Yuchen Hao, Wei Shan, and Zhihong Dai. Identifying Experts in Community Question Answering Website Based on Graph Convolutional Neural Network. *IEEE Access*, 8:137799–137811, 2020. Conference Name: IEEE Access.

[8] Hongtao Liu, Zhepeng Lv, Qing Yang, Dongliang Xu, and Qiyao Peng. ExpertBert: Pretraining Expert Finding. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pages 4244–4248, New York, NY, USA, October 2022. Association for Computing Machinery.

[9] Tao Qin, Tie-Yan Liu, and Hang Li. A general approximation framework for direct optimization of information retrieval measures. *Information Retrieval*, 13(4):375–397, August 2010.

[10] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, August 2019. arXiv:1908.10084 [cs].

[11] Zakaria Saoud and Colin Fontaine. Expert finding in Citizen Science platform via weighted PageRank algorithm. October 2018.

[12] Lea A. Shanley, Alison Parker, Sven Schade, and Aletta Bonn. Policy Perspectives on Citizen Science and Crowdsourcing. *Citizen Science: Theory and Practice*, 4(1):30, December 2019. Number: 1 Publisher: Ubiquity Press.

[13] Shem Unger, Mark Rollins, Allison Tietz, and Hailey Dumais. iNaturalist as an engaging tool for identifying organisms in outdoor activities. *Journal of Biological Education*, 55(5):537–547, 2021. Publisher: Taylor & Francis.

[14] Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht. *The science of citizen science*. Springer Nature, 2021.

[15] G. Alan Wang, Jian Jiao, Alan S. Abrahams, Weiguo Fan, and Zhongju Zhang. ExpertRank: A topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems*, 54(3):1442–1451, February 2013.

[16] Michał Wróblewski, Joanna Suchomska, and Katarzyna Tamborska. Citizens or Consumers? Air Quality Sensor Users and Their Involvement in Sensor.Community. Results from Qualitative Case Study. *Sustainability*, 13(20):11406, January 2021. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[17] Jun Zhang, Mark S. Ackerman, and Lada Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 221–230, New York, NY, USA, May 2007. Association for Computing Machinery.