Predicting Reputation Score of Users in Stack-overflow with Alternate Data

Sahil Yerawar[1], Sagar Jinde[1], P. K. Srijith[1], Maunendra Sankar Desarkar[1], K. M. Annervaz[2] and Shubhashis Sengupta[2]
[1]*Indian Institute of Technology Hyderabad, India*
[2]*Accenture Technology Labs, India*

Keywords: Recommendation System, Cold-start Recommendation, Knowledge Distillation, Transfer Learning.

Abstract: The community question and answering (CQA) sites such as Stack Overflow are used by many users around the world to obtain answers to technical questions. Here, the reliability of a user is determined using metrics such as reputation score. It is important for the CQA sites to assess the reputation score of the new users joining the site. Accurate estimation of reputation scores of these *cold start* users can help in tasks like question routing, expert recommendation and ranking etc. However, lack of activity information makes it quite difficult to assess the reputation score for new users. We propose an approach which makes use of alternate data associated with the users to predict the reputation score of the new users. We show that the alternate data obtained using users' personal websites could improve the reputation score performance. We develop deep learning models based on feature distillation, such as the student-teacher models, to improve the reputation score prediction of new users from the alternate data. We demonstrate the effectiveness of the proposed approaches on the publicly available stack overflow data and publicly available alternate data.

1 INTRODUCTION

A large majority of internet users refer to community question answering (CQA) sites to gather information, knowledge and general consensus views for their own purposes. Examples of such CQA sites include Reddit, Stack Exchange, Quora etc. The success of these community sites are mostly attributed to efficient moderation of user questions and answers, as well as the presence of a large number of users who consistently provide high quality content. The reliability of these consistent users is quantified by using a points based system (Cavusoglu et al., 2015), where the user is encouraged to interact with the community by providing comments, posts, questions and answers to earn their points which is also a measure of trust and expertise. In Stack Overflow, the points earned by the user are often called the reputation score (Overflow, 2022b). The reputation score of the users can be utilized by the system to find expert users to answer questions, rank the answers by users, and in providing some privileges. However, when a new user joins the community QA website, due to lack of any registered activity, they will be assigned a very low reputation score and consequently their expertise can not be tapped completely by the system.

In this work, we aim to predict the reputation score new users could have achieved, so that their expertise can be fully utilized by the system to get expert answers to the questions, and provide a better user experience (Slag et al., 2015). We address this by considering alternate data provided by users such as their publicly available website information. In addition to the activity based features like views, upvotes and downvotes, stack overflow often contains user's profile information and publicly available website URLs such as personal websites and GitHub accounts. We believe that by leveraging the content of the personal website, we can determine technical expertise, work and education backgrounds of the user. Consequently, this can be used to predict an approximate reputation score, which can be used by the system to provide a better user experience. This would help in providing a fair estimate of the users' expertise in the form of reputation score, which can help in recommending new users with a certain level of proficiency to different questions in Stack Overflow. The proposed approach addresses the cold start problem of assigning reputa-

Yerawar, S., Jinde, S., Srijith, P., Desarkar, M., Annervaz, K. and Sengupta, S.
Predicting Reputation Score of Users in Stack-overflow with Alternate Data.
DOI: 10.5220/0011591900003335
In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) - Volume 1: KDIR*, pages 355-362
ISBN: 978-989-758-614-9; ISSN: 2184-3228

tion scores for new users, which would help the downstream Recommender Systems for CQA websites.

To predict the reputation score of new users, we develop deep learning models which could predict the reputation score from the alternate data. Another major contribution of our work is to develop effective deep learning models which learns not only from the alternate data but also from the activity information by treating them as privileged information and performing feature distillation using a teacher-student learning technique (Hinton et al., 2015; Lopez-Paz et al., 2016; Xu et al., 2019). We propose an auto-encoder based teacher and student model which allows transfer of knowledge from a teacher network trained on activity information and alternate data, to a student network trained on alternate data alone. We propose several ways to transfer the knowledge between teacher and student for effective learning. We demonstrate first the usefulness of the alternate data in predicting the reputation score. Following this, we demonstrate the effectiveness of the proposed approaches in predicting the reputation score of the new users.

The contributions of this paper are three-fold:

- We propose to use alternate data such as personal website information to improve reputation score prediction in stack overflow and use them in particular to predict reputation score for new users.
- We develop an auto-encoder based teacher and student model and feature distillation techniques to predict the reputation score of users from alternate data.
- We conducted experiments on the publicly available stack overflow data and publicly available alternate data, demonstrating usefulness of the alternate data and proposed methodologies for reputation score prediction.

2 RELATED WORK

2.1 Cold Start in Recommendation

Collaborative Filtering (CF) in Recommendation Systems have given good performances in recommending items to users who have past user-item interaction history. Traditionally, Collaborative Filtering involves matrix factorization to obtain low level embeddings of user and item vectors. However, most CF algorithms fail to work in the case of discovering interactions between both new users and new items (Lam et al., 2008; Park and Chu, 2009).

To alleviate the drawbacks of CF-based algorithms in the context of cold start problem, many researchers also utilize content information of these users to formulate content based methods. These methods try to learn an appropriate new user/item representation based on content information and learn a mapping to convert these representations into user-item interaction vector space. However, these methods fail to model the complex nature of interaction between embedding space and the user content embedding space, due to which a new user embedding cannot associate with the embedding space. More recent methods to address cold start problem involve a hybrid approach of using CF and content based methods (Sujithra Alias Kanmani et al., 2021; **?**), using meta-learning (Zheng et al., 2021; Chen et al., 2021; Wang et al., 2021a; Du et al., 2022), using pre-training (Hao et al., 2021) etc. However, most of these approaches aim to produce a good ranking of recommended items for the end users, and do not focus on finding the *true value* of the cold-start item or the *true activity-based user profile* of the cold-start user.

2.2 Stack Overflow

Stack Overflow is one of the largest community question-answer (CQA) websites, playing a crucial role in the daily activities of numerous researchers, developers and scholars all over the world. In Stack Overflow, the point based system is known as reputation score which can be earned by asking good questions and providing answers on the Stack Overflow website. Subsequently, there have been numerous studies on Stack Overflow reputation score and how it behaves for different users. However, to the best of our knowledge, there is no work which aims to predict the reputation score of new users based on the alternate data such as the website information.

Movshovitz-Attias et al. (Movshovitz-Attias et al., 2013) studied participation patterns of expert and non-expert users present in StackOverflow reputation system. Their study concluded that influential users have a good contribution during the first month of activity and that expert users contribute drastically more as soon as they join the site. They also found that high reputation users were the primary sources of high-quality answers. The study in Anderson et al. (Anderson et al., 2012) looked at the way the community determines the answers, including how votes are cast. The authors found that the answerer's reputation correlates with the speed of answers and that the arrival time of an answer impacts its chances of being chosen as the best one. In Slag et al. (Slag et al., 2015), the authors found that no response to a question or unsatisfactory answer is the main issue a new user faces. Due to this reason, the new users may get demotivated and stop using the site. Their analysis showed that half of them have just one contribution on the site, and 60% of users have a reputation score of one. These points towards the need to identify expert users early, get satisfactory answers, provide more privileges, and encourage them to participate in the QA site.

In Morrison et al. (Morrison and Murphy-Hill, 2013), the authors found a positive correlation between reputation score and age. They also examined users' familiarity with various skills and technologies using tags. MacLeod et al. (MacLeod, 2014) performed an exploratory analysis of stack exchange data. They found that the reputation score of a user is positively correlated to the number of tags. Users with high reputation scores contribute to a diverse number of topics. Recent approaches to user reputation score prediction (Woldemariam, 2020; Banati and Seema, 2021) made use of the textual content of the user posts and its syntactic and semantic information in determining the reputation score of the users. Unlike all these previous works, our aim is to develop an approach to predict the reputation score of *cold start users* in the question answering sites.

3 BACKGROUND AND PROBLEM DEFINITION

3.1 Features from the Dataset

In this work, we want to utilize stack overflow dataset for identifying the user reputation scores. We try to develop models that can predict the user's reputation scores based on the alternate data. This can help the QA systems to make use of the expertise of new users in providing relevant answers to questions posted by users. We divide the features available in the stack overflow dataset (Overflow, 2022a) as follows:

- **Activity Information:** The features reflect the activity of the Stack Overflow user in the system. It includes the *number of views*, *number of upvotes received*, *number of downvotes received* and the *number of years* the user has been active. We calculate the *number of years* by subtracting the *user creation date* from *last accessed date* present in the dataset. We call these features the *privileged features* which the QA systems use to compute the reputation score. The new users joining the system lack these features, while only old users will have these features.
- **About Me Text:** Stack Overflow users can write about themselves and their interest in a separate section named "About Me". This textual data could be utilized as our alternate data which provides an indication about the user's capabilities. All the data present in "About Me" section is publicly available.
- **Website Text:** In Stack Overflow, users can also share their personal websites containing more information about their skill sets and achievements. This textual data can act as another source of alternate data which could be of great help in reputation score prediction. All the data obtained from personal websites are publicly available.
- **Website Category:** There are various kinds of websites provided by the users. We make a broad classification of these websites into three categories: Social media websites, Academic pages and Personal blogs. We consider this information also as part of the alternate data.

3.2 Problem Definition

We consider the problem predicting the reputation score of a user $y \in \mathcal{R}$, given their alternate data based features $X \in \mathcal{R}^H$ and privileged activity features $X^* \in \mathcal{R}^{H'}$. Assume we are given a training data set $\mathcal{D} = \{X_i, X_i^*, y_i\}_{i=1}^N$ which consists of users for whom both the privileged data and alternate data are available (experienced users), and test data $\hat{\mathcal{D}} = \{\hat{X}_i, \hat{y}_i\}_{i=1}^{\hat{N}}$ which consists of users for whom only alternate data are available, i.e. new users. We aim to learn a regression model $f : X \to y$ to predict the reputation score of users using alternate data, but also make use of the availability of the privileged data X^* during training.

3.3 Learning using Privileged Information

Standard machine learning and deep learning techniques assume the same set of features to be available with all the users. However, it is different in our case where we have a set of users for whom we have activity (privileged) data while the rest does not have. The recent works along the direction of learning from privileged information (LUPI) (Pechyony and Vapnik, 2010; Vapnik and Vashist, 2009; Vapnik and Izmailov, 2015) have shown that making use of the additional privileged features associated with some subset of examples could improve the generalization performance of the learning algorithms. Therefore,

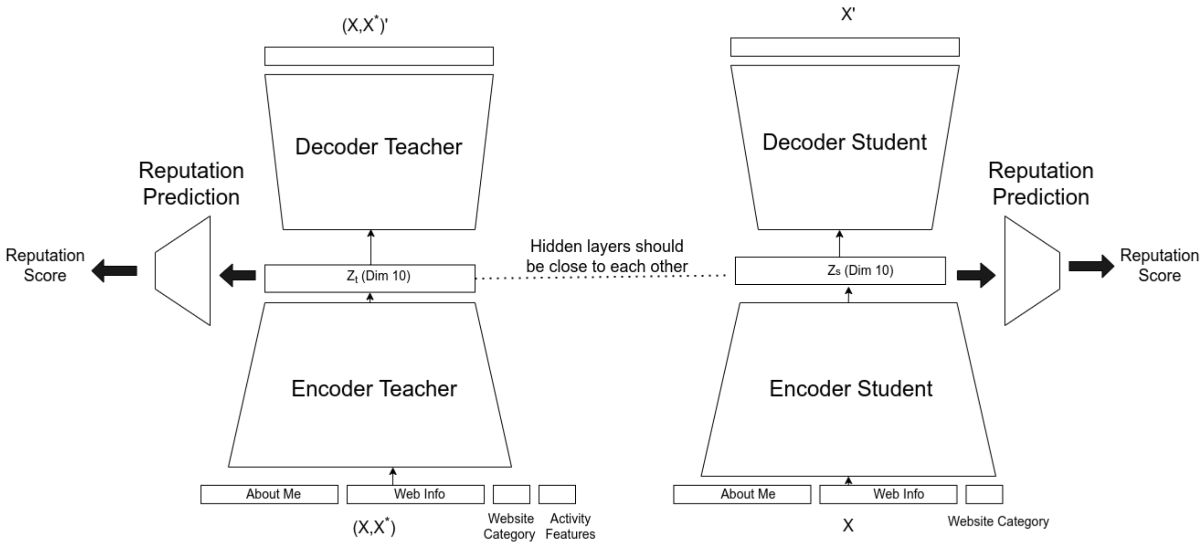

Figure 1: Model Architecture of Teacher and Student model with direct constraints on the latent representations.

we develop our models for reputation score prediction using the framework of learning using privileged information (Hinton et al., 2015; Lopez-Paz et al., 2016) and more specifically the feature distillation approach (Xu et al., 2019; Wang et al., 2021b).

A more general framework of knowledge distillation has been used for several problems which involve transfer of knowledge from one model to another. In particular, knowledge distillation (Hinton et al., 2015; Mirzadeh et al., 2020; Zhou et al., 2018) consider two models: teacher $f_t()$ with its weights W_t and student $f_s()$ with model parameters W_s. The teacher model is more complex than the student model. The goal of this technique is to improve the training of the student model with the help of predictions done by the teacher model. Knowledge distillation is usually employed in cases where the teacher model is very complex in architecture and we want to emulate its prediction distribution within the more simpler student model. For the standard model distillation, the loss is governed by the following objective function:

$$
\begin{aligned}
& min_{W_s} \sum_{i=1}^{N} L_p(y_i, f_s(X_i;W_s) \\
& + \lambda * L_d(f_t(X_i;W_t), f_s(X_i;W_s)
\end{aligned}
\tag{1}
$$

The loss component $L_p()$ is the standard loss (e.g root mean square error for regression) while L_d helps in distilling the knowledge learned by the teacher and aids in better training of W_s. For instance, through L_d, the student model tries to match the prediction $f_t(X;W_t)$ generated by the teacher model. Here λ is a regularization constant that controls the weight associated with each loss term.

We consider a different setup of knowledge distillation known as feature distillation (Vapnik and Vashist, 2009; Xu et al., 2019) where the feature set being visible to the student and teacher is different. It assumes the teacher has access to both X and X^* while the student has access to only X. In feature distillation, the aim of the student is to learn to predict only using feature X but with the help of the additional knowledge available with the teacher model. The learning of the student is done through feature distillation where the following loss is used.

$$
\begin{aligned}
& min_{W_s} \sum_{i=1}^{N} L_p(y_i, f_s(X_i;W_s) \\
& + \lambda * (L_d(f_t(X_i, X_i^*;W_t), f_s(X_i;W_s))
\end{aligned}
\tag{2}
$$

Here, $L_p()$ is the standard loss (root mean square error) and $L_d()$ is the distillation loss. The teacher is trained using the standard loss alone but using both the features X and $X*$. Through feature distillation, we aim the student to not only predict the output from X but also learn representations similar to the teacher through the distillation loss term. The teacher is expected to have better representation learning capability as it learns from both the features and can assist the student model to learn better representations through the distillation process.

4 MODELS FOR PRIVILEGED LEARNING OF REPUTATION SCORES

We provide a detailed description of the teacher-student model and the distillation techniques used to transfer knowledge from teacher to the student in predicting the reputation score of users.

We use three models to aid the learning process - the teacher model $f_t : (X, X^*) \rightarrow y$ which has access to both privileged and common features, the baseline model $f_b : X \rightarrow y$ which has access to only common features and no information provided by the teacher, and student model $f_s : X \rightarrow y$ which can access only common features and is guided by the teacher. All the models are based on auto-encoders, which learns a latent representation of the input using an encoder and then tries to reproduce the input through an decoder. The auto-encoder allows one to learn a good latent representation of the features and we intend to transfer knowledge through these latent representations. The latent representation is used to predict the reputation score using a neural network. In all these models, the reputation score prediction from the latent representation is done using a two layer neural network. The teacher will be able to learn to predict reputation scores from both the alternate data and privileged data while the student has access to only alternate data, but we intend to improve student learning using the teacher.

We transfer the knowledge from teacher to student by constraining the latent space of the teacher and student to be similar. This is achieved by putting an RMSE loss directly on the latent representations of the student z_s and teacher z_t. We follow two variations of training this student-teacher model: 1) training student and teacher simultaneously (co-train), 2) training the teacher model first and then the student model with a constraint on the latent representations (pre-train). Figure 1 represents the overall architecture of the proposed model.

4.1 Loss Function

In the first case (co-train), we use joint loss functions over both teacher and student models along with the RMSE loss over the latent space ($L_{constr}()$) which ensures similarity.

$$\begin{aligned} min_{W_t, W_s} \sum_{i=1}^{N} (L_p(y_i, f_t(X_i, X_i^*; W_t)) \\ + \lambda_1 L_{rec}((X_i, X_i^*), (X_i', X_i^{*\prime})) + L_p(y_i, f_s(X_i; W_s)) \\ + \lambda_2 L_{rec}(X_i, X_i') + \lambda_3 L_{constr}(z_{ti}, z_{si})) \end{aligned} \tag{3}$$

where $(X', X^{*\prime})$ and X' are reconstructed input by the decoder of teacher and student networks respectively. Here, $L_{rec}()$ is the auto-encoder reconstruction loss (RMSE) between the output of the decoder and input of the encoder. The hyper-parameters $\lambda_1, \lambda_2, \lambda_3$ control the degree of regularization through the reconstruction loss and strength of transfer.

In the second case (pre-train), we train the teacher model first using the following loss

$$\begin{aligned} min_{W_t} \sum_{i=1}^{N} L_p(y_i, f_t(X_i, X_i^*; W_t)) \\ + \lambda L_{rec}((X_i, X_i^*), (X_i', X_i^{*\prime})) \end{aligned} \tag{4}$$

Then, we train the student model with the following loss which includes the constraint loss component.

$$\begin{aligned} min_{W_s} \sum_{i=1}^{N} L_p(y_i, f_s(X_i; W_s)) \\ + \lambda_1 L_{rec}(X_i, X_i') + \lambda_2 L_{constr}(z_{si}, z_{ti}) \end{aligned} \tag{5}$$

4.2 Experimental Details

We perform the experiments on a subset of the Stack Overflow dataset. From 13 million users, we select 65957 users who have valid *about me*, *website* data and have their reputation scores to be between 2 and 10000, both inclusive. We exclude users with reputation score value of 1 as every new user is assigned that score at the time of creation and signifies no user activity for a lot of users. We use LDA (Blei et al., 2003) and GLoVE (Pennington et al., 2014) to convert about me and textual data into embedding vectors. As presented in section 3.1, we also use the website category feature, which is a vector of size 3. We split our dataset into train split and test split in about 80:20 ratio and use the results as reported on the test data. We use the root mean square error (RMSE) metric to evaluate the reputation score prediction capability of the model. The teacher model will have access to both the privileged and alternate data during evaluation while the student will have access to only alternate data. The experiments use a 5 fold cross validation technique on the train split and the hyper-parameters are selected using grid search. We also consider Bayesian optimization (Shahriari et al., 2016) to get the optimal hyper-parameters for the co-train model. The batch size for all the experiments is 64 and the learning rate is 0.001 which is used in Adam Learning Optimizer (Kingma and Ba, 2014). In Figure 1, while using LDA we consider input vectors of size 20 for both about me and web data features (and 100 for Glove), a 3 dimensional web category and a 4 dimensional vector of activity features. The teacher encoder will have 4 fully connected layers of size 30, 20, 15 and 10. We provide the activity feature vector in the second layer of the encoder. The teacher decoder consists of fully connected layers of size 15, 24, 30 and 43. The reputation prediction module consists of two fully connected layers of size 5 and 1.

Table 1: Prediction errors (RMSE) of the models on the stack overflow data.

Model	Training	RMSE (LDA)	RMSE (GLoVE)
Feed Forward Network	only activity features	943.6	
Baseline Auto-encoder	only alternate data	2181.4	2266.5
Teacher Student	Pre-train and Grid search	608.5 1960.0	756.5 1968.8
Teacher Student	Co-train and Grid search	626.6 1937.6	757.5 2173.5
Teacher Student	Co-train and Bayesian optimization	589.7 1949.3	649.1 2130.2

The student component of the model is very similar to that of the teacher, except that the second layer of the encoder and third last layer of decoder is of size 20, since we are not passing any activity features directly to the student. For reference, we have developed a simple autoencoder network which uses only activity features to predict the reputation score. The layers in the encoder are fully connected layers of size 4,3,3,2 and that of decoder is 3,3,4,4. The reputation prediction module consists of two fully connected layers of size 2 and 1.

The hyper-parameters selected are as follows:

- For the Teacher model, the learning rate is 0.01, batch size is 128, number of epochs is 100 and the λ used in reconstruction loss is 0.95.
- For the Teacher assisted student model, the learning rate is 0.001, batch size is 128, number of epochs is 100 and the λ used in reconstruction loss is 0.8.

4.3 Results

Table 1 shows the results of the various models used for reputation prediction. Note that all reported values correspond to the RMSE errors between reputation prediction and actual score. When comparing the results of the feed forward model with the teacher base model, it is clear that a combination of activity and alternate features gives better results than considering only activity features. We can also see that for all the proposed student models, the RMSE is less than the *baseline auto-encoder* without any distillation. This shows the effectiveness of our distillation approach and the student auto-encoder. Although feature distillation reduces RMSE, the gap between teacher and student models still needs some improvement.

Comparing the two representations of the text, we observe that the LDA based representation gives a better performance than the Glove representation. This is possibly due to two reasons. The dimensionality of GLoVE embeddings is higher than the LDA based embeddings. This may distract the model from the activity based features which are lesser in number. The student and the distilled model learn better for LDA as it has a lesser number of dimensions. Secondly, LDA tries to identify the topic distribution in the text holistically by considering the entire text together. On the other hand, the GLoVE based method aggregates the information from individual words to get the final vector. Thus, LDA is more useful in capturing the expertise of the user and consequently predict the reputation score.

Table 2 shows a case study involving two users from the dataset. Both the users have a reasonable amount of alternate data available on their website. Our student model could effectively learn from the alternate data and has assigned reasonably good scores to both of them, higher scores to one with higher ground truth reputation score and vice-versa. We see many more such examples from our experiments which support the applicability of the approach.

5 CONCLUSIONS

We have developed an auto-encoder based regression model and utilized privileged distillation to transfer the knowledge of teachers, to assist the training of the student model. We consider two different training procedures for the proposed model. The proposed approaches were found to be effective in predicting reputations score of users in the stack overflow data. We found the alternate data to provide additional information which improved the reputation score prediction performance. Further, we found the student model learnt using feature distillation to also improve the performance compared to the baseline model. As a future work, we would like to further reduce the performance gap between student and the teacher in predicting the reputation score.

Table 2: Actual and predicted scores for two users from the test split of dataset using student model with LDA input vectors. The examples are tken from public data dump of the Stack Exchange forum.

User name	**About_Me**	**Web_Data**	**Prediction**	**Target**
******	https://www.*****.com	How To Watch For Changes Tuesday March 30 2021 5 minute read I needed to customize the reviews UI on a product detail page for a Shopify site whenever there were no reviews. However, I ran into a series of unexpected problems ...	1227.6	3114
*******	Computer ***** Student at Georgia Tech	A 60 GHz phased arrayfor $10 A 60 GHz phased array for $10 In 2018, I gave an talk that was basically Phased Arrays 101...	628.7	1223

REFERENCES

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2012). Discovering value from community activity on focused question answering sites: A case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 850–858, New York, NY, USA. Association for Computing Machinery.

Banati, H. and Seema (2021). Proficiency assessment of experts in online social network q/a communities. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 1–5.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. volume 3. JMLR.org.

Cavusoglu, H., Li, Z., and Huang, K.-W. (2015). Can gamification motivate voluntary contributions? the case of stackoverflow q&a community. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, CSCW'15 Companion, page 171–174, New York, NY, USA. Association for Computing Machinery.

Chen, Z., Wang, D., and Yin, S. (2021). Improving cold-start recommendation via multi-prior meta-learning. In *European Conference on Information Retrieval*, pages 249–256. Springer.

Du, Y., Zhu, X., Chen, L., Fang, Z., and Gao, Y. (2022). Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Hao, B., Zhang, J., Yin, H., Li, C., and Chen, H. (2021). Pre-training graph neural networks for cold-start users and items representation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 265–273.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Lam, X. N., Vu, T., Le, T. D., and Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 208–211.

Lopez-Paz, D., Schölkopf, B., Bottou, L., and Vapnik, V. (2016). Unifying distillation and privileged information. In *International Conference on Learning Representations (ICLR)*.

MacLeod, L. (2014). Reputation on stack exchange: Tag, you're it! In *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, pages 670–674.

Mirzadeh, S. I., Farajtabar, M., Li, A., Levine, N., Matsukawa, A., and Ghasemzadeh, H. (2020). Improved knowledge distillation via teacher assistant. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):5191–5198.

Morrison, P. and Murphy-Hill, E. (2013). Is programming knowledge related to age? an exploration of stack overflow. In *2013 10th Working Conference on Mining Software Repositories (MSR)*, pages 69–72.

Movshovitz-Attias, D., Movshovitz-Attias, Y., Steenkiste, P., and Faloutsos, C. (2013). Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pages 886–893.

Overflow, S. (2022a). Stack Overflow Datadump.

Overflow, S. (2022b). What is reputation? How do I earn (and lose) it? - Help Center - Stack Overflow.

Park, S.-T. and Chu, W. (2009). Pairwise preference regression for cold-start recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 21–28.

Pechyony, D. and Vapnik, V. N. (2010). On the theory of learning with privileged information. In *NIPS 2010*.

Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global vectors for word representation. In *Proceed-*

ings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N. (2016). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

Slag, R., de Waard, M., and Bacchelli, A. (2015). One-day flies on stackoverflow - why the vast majority of stackoverflow users only posts once. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 458–461.

Sujithra Alias Kanmani, R., Surendiran, B., and Ibrahim, S. (2021). Recency augmented hybrid collaborative movie recommendation system. *International Journal of Information Technology*, 13(5):1829–1836.

Vapnik, V. and Izmailov, R. (2015). Learning using privileged information: similarity control and knowledge transfer. *J. Mach. Learn. Res.*, 16:2023–2049.

Vapnik, V. and Vashist, A. (2009). 2009 special issue: A new learning paradigm: Learning using privileged information. *Neural Netw.*, 22(5–6):544–557.

Wang, J., Ding, K., and Caverlee, J. (2021a). Sequential recommendation for cold-start users with meta transitional learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1783–1787.

Wang, S., Zhang, K., Wu, L., Ma, H., Hong, R., and Wang, M. (2021b). Privileged graph distillation for cold start recommendation. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Woldemariam, Y. (2020). Assessing users' reputation from syntactic and semantic information in community question answering. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5383–5391, Marseille, France. European Language Resources Association.

Xu, C., Li, Q., Ge, J., Gao, J., Yang, X., Pei, C., Sun, F., Wu, J., Sun, H., and Ou, W. (2019). Privileged features distillation at taobao recommendations.

Zheng, Y., Liu, S., Li, Z., and Wu, S. (2021). Cold-start sequential recommendation via meta learner. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4706–4713.

Zhou, G., Fan, Y., Cui, R., Bian, W., Zhu, X., and Gai, K. (2018). Rocket launching: A universal and efficient framework for training well-performing light net. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).