IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 December 2023, accepted 21 January 2024, date of publication 26 January 2024, date of current version 5 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359053

RESEARCH ARTICLE

GCZRec: Generative Collaborative Zero-Shot Framework for Cold Start News Recommendation

SYED ZAIN UL HASSAN [1], MUHAMMAD RAFI [1], AND JAROSLAV FRNDA [2,3], (Senior Member, IEEE)

[1]Department of Computer Science, School of Computing, National University of Computer and Emerging Sciences, Islamabad 44000, Pakistan
[2]Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communications, University of Žilina, 01026 Žilina, Slovakia
[3]Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB—Technical University of Ostrava, 70800 Ostrava, Czech Republic

Corresponding author: Jaroslav Frnda (jaroslav.frnda@uniza.sk)

This work was supported by the European Union within the REFRESH Project—Research Excellence for Region Sustainability and High-Tech Industries of the European Just Transition Fund under Grant CZ.10.03.01/00/22 003/0000048.

ABSTRACT The aim of personalized news recommendation is to suggest news stories to the users that are most interesting for them. To improve the user experience, it is important that these news items are not only relevant to the user but also get recommended to them as soon as they are available. The inability of traditional collaborative filtering approach to recommend such cold start items has led to techniques that incorporate latent features of items in order to make cold start recommendations such as content based filtering and deep neural network-based approaches. However, these existing techniques do not make use of any collaborative information between users and items as well as latent features at the same time and thus fail to provide any serendipity which is an important aspect of any recommender system. Moreover, these underlying collaborative signals between users and items are crucial to improving the overall quality of recommender systems and can also be utilized to make cold start recommendations. In this paper, we propose the Generative Collaborative Zero-Shot Recommender System framework (GCZRec) which makes use of both the latent user and item features as well as the underlying collaborative information to generate both warm start and cold start recommendations. We evaluate our framework for news recommendation task given cold start and warm start cases for both users and news items. We also discuss that our model can be plugged in and used as preprocessing to improve the performance of an existing recommender system.

INDEX TERMS News recommendation, cold start problem, zero-shot learning, recommender system.

I. INTRODUCTION

The improvement in media technology and online services have resulted in an overload of information especially with online news articles as the people realize the need to be well-informed at all times [1], [2]. Recommender systems can therefore improve the user experience by suggesting news articles that are most recent, relevant and contain value for her. These systems can help the users find information that is interesting and personalized. But compared to recommending movies and products, news article recommendations often entail some additional challenges such as the latest news articles being posted frequently and lacking any historical interactions that can be used for recommending these news items [3]. This severe case of cold start problem is a challenge in news recommendations. Moreover, from the user point of view, these news stories need to be recent but highly personalized, while from the item perspective, it should be recommended to the users based strictly on its relevance to those particular users.

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

The conventional collaborative recommendation algorithms rely on historical interaction data of users and items to find hidden patterns based on similarity [4], [5]. The performance of these algorithms decreases when the data contains missing user interaction entries for the items. This lack of data is mostly seen in the case of news articles which are often posted without any prior interaction information. This leads to a severe case of cold start problem.

Other techniques such as Matrix Factorization [6] and content-based filtering [7] also suffer from cold start user problem [8], [9], [10]. In case of Matrix Factorization, it can additionally suffer from both over-fitting and under-fitting given the available historical data. Another problem that both of these techniques face is the assumption that features are always independent. This condition is difficult to hold true in most real-world scenarios where not only the features but items also have relative dependence on features and themselves.

The cold start problem in recommender systems can be remodeled as a classical zero-shot learning task which comes from the computer vision domain [11], [12]. In zero-shot classification, the set of classes in the training data and set of classes in the samples to be classified can be disjoint. Similarly, in cold start item recommendations, the aim is to predict whether an item should be recommended to a particular user without any available historical interactions for that item. In cold start user case, items are to be recommended to a particular user for which there are no existing historical information [13]. Following this intuition, the features of news items and users can be used to deduce the behavioral context of cold start items and users in recommendation scheme just like a class label can be predicted for an unseen data sample using the generalization from known samples in zero-shot classification. Some existing studies [14], [15] have used this relation to propose recommendation models for cold start items.

But these techniques do not take into account serendipity, which is an important aspect of a recommender system [16], [17], [18], [19]. This lack of diversity stems from the inability of these models to make use of the latent collaborative information between users as well as items. These neighborhood signals are therefore important to make fine-grained recommendations that are not only relevant to the active users but also provide diversity in choices for them.

We observe that by directly synthesizing the interactions based on feature representations can eliminate the need for any external click predictor model and can also provide an effective method to not only allow item-to-user interactions prediction but also projection of user-to-item interactions. This synthesis of interactions can also allow us an efficient method to rank the predicted interactions. This can be achieved by incorporating a generative network with conditional information to learn the latent collaborative information between users and items. This allows us to use these hidden patterns in the available historical data to directly synthesize the interactions for cold start users and items. In the same way that an unseen class label is used for prediction by leveraging the features of the novel sample, the conditional input of the generator network can also be learned from the available item and user feature representations.

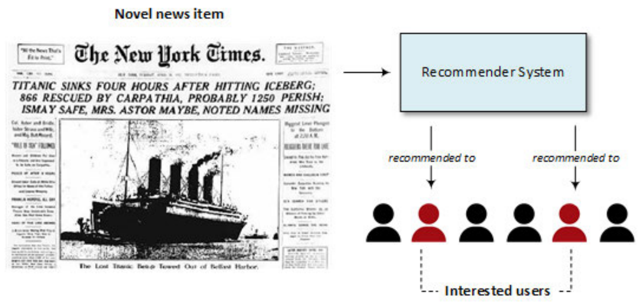


FIGURE 1. Illustration of the cold start news recommendation problem.

Based on the previous discussion, we propose a novel recommender system framework, GCZRec, to synthesize both cold start and warm start interactions for users and news items. Our technique utilizes the hidden feature information of users and items to perform cold start recommendations as zero-shot predictions. The proposed model is capable of learning collaborative signals between users and among items to generate interactions thus allowing diverse recommendations. The framework also allows the ranking of these recommendations. At its core, GCZRec framework consists of two separate classifiers for zero-shot labelling of cold start news and cold start user. These predicted classes are used as input to conditional Wasserstein GAN (cWGAN) for generating interactions. During training, two separate generator networks are independently trained such that each training sample of the first network represent a news item with interaction. This generator network is trained on samples each one of which is an interaction vector containing both interactions of users for news items. The experiments were conducted on two publicly available news recommendation datasets Microsoft News Dataset (MIND) [20] and Addressa [21] in order to provide the proof of concept for our research. Furthermore, our framework allows this problem to be formulated as an extreme multi-label classification task where the class labels are news items to be recommended.

The main contributions of this research are as follows:

- We propose a novel GCZRec framework capable of using latent collaborative information to make both cold start and warm start recommendations of news items in generative manner and allowing the recommended items to be ranked.
- We present a formulation of cold start recommendation as zero-shot learning problem and utilize hidden features of both users and items in order to make recommendations.
- Our framework can also be used for typical extreme multi-label classification task and provides an efficient

approach for predicting the subset of labels from a large space given a new instance.

II. PRELIMINARIES

The goal of a recommender system is to present the users with an ordered set of items which are ranked based on the preference and relevance of these items for each particular user. This section defines the relevant concepts pertaining to the overall recommendation problem and provides the necessary basis for further discussion on these topics in the subsequent sections.

Definition 1: Given set of users U and items I, the U x I interaction matrix $\Re$ represents the historical choices of users and items $r(u \in U, i \in I)$. A cold start user problem occur when $r(u_{new}, i)$ is undefined for a novel user u_{new} and all values of items i in I. Whereas, a cold start item problem occurs when $r(u, i_{new})$ is undefined for a novel item i_{new} and all values of u in U.

The cold start problem in recommender systems is comparable to zero-shot classification problem in computer vision.

Definition 2: In zero-shot learning, the classification model generalizes feature information from seen classes to an unseen class in order to predict it. Mathematically, given a set of instances X and set of labels Y where Y contains both seen and unseen classes, and feature space Z, the objective of zero-shot learning is to learn the mapping f from input state X to semantic space Z:

$$f : X \rightarrow Z \tag{1}$$

And also learn the mapping g from semantic space Z to label space Y:

$$g : Z \rightarrow Y \tag{2}$$

Since, in a recommendation problem, there are typically a large number of users and items involved. The selection of a small subset of relevant items for the user from a large space of available items is analogous to predicting class labels in an extreme multi-label classification problem.

Definition 3: In extreme multi-label classification, the objective is to predict a subset of most relevant labels from a high-dimensional label space containing a vast number of potential labels, given an input instance. Mathematically, given an input space χ and a high-dimensional label space $|\mathcal{L}|$. The objective in extreme multi-label classification is to train a model that can find a set l containing relevant labels for a novel instance x given $l \subseteq |\mathcal{L}|$.

We now introduce serendipity which can generally be seen as the measure of diversity in recommendations produced by an algorithm and is an important characteristic for improving the overall user experience.

Definition 4: In the context of recommender systems, serendipity refers to the ability of an algorithm to recommend unexpected and diverse items to the users to expand their taste into neighboring interest areas.

III. RELATED WORK

Over the years, numerous techniques have been proposed to deal with recommendation problem with Collaborative Filtering [4], Content-based Filtering [6], [22] and Matrix Factorization [7] among the prominent approaches. However, the problem of news recommendation presents an additional challenge that the item must be linked to a target set of readers soon upon entry into the system.

In this section, we first review the news recommendation problem and the techniques that were employed for this specific task and then we shift our attention to generative adversarial approaches for recommendations that are present in literature.

A. NEWS RECOMMENDATION

The earliest news recommendations were focused on similarity and classical machine learning algorithms. In [22], similarity between user model and news articles are exploited to generate personalized recommendations. For finding relevant news items, [23] proposed the idea of using semantics of the news articles. SF-IDF in combination with different semantic similarity measures were used to find relevant news items where the only semantic context they incorporated was based on synonyms. The approach of using SF-IDF was further extended by [24] in their work which used an updated SF-IDF measure for finding semantic similarity while taking into account the relationship between synonym sets. In a graph-based approach, [25] discussed the use of knowledge graphs by connecting named entities, events and places present in the news articles.

The idea of employing collaborative filtering along with content-based approach to make news recommendation was also explored in research. One such example was NewsDude [26] which recommended news by sequentially employing three modules. A content-based recommender, followed by classical collaborative component and a Naïve Bayes classifier. In [27], a hybrid algorithm was presented that combined content-based recommender system with collaborative filtering to recommend sports news articles. The inability of collaborative version to handle cold start items was dissimulated by the content-based component. In another such work [28] proposed the technique for fusion of collaborative filtering and content-based modelling to generate news recommendations. The content-based module was used to construct user profile while user groups similar to the active user were found in much the same way as in a collaborative approach. Then a fusion model with user's current and potential interests was developed to recommend news by finding similarity between the fusion model and content of the news articles.

In a different approach for finding personalized news articles, [29], [30], [31] used deep neural networks as their recommendation model. In [29], a news encoder and user encoder were trained such that the news encoder used attention mechanism to find topic information from news articles through classification. The user encoder was constructed with the help of users click behavior on news articles. The news encoder was constructed in much the same way by [30]. However, they argued that capturing both long-term and short-term interests of the users is necessary for recommending highly personalized news items. The long-term representations were captured by the embeddings of user IDs while the short-term representation of the users was guided by their browsed articles using a GRU network.

The idea of different users who click on the same article with attention on different aspects was discussed by [31] in their paper. They used convolutional neural network (CNN) to learn news item representation from its title. The attention mechanism was used at news-level and word-level in the news model since a particular news may have different importance or relevance for different users.

B. GENERATIVE METHODS

Among the first to use GAN for recommendation problem were IRGAN [32] and GraphGAN [33]. These methods explored the potential of GAN for recommender systems but suffered from the well-known ''label confusion'' problem; that is the model learning to label an item with positive and negative labels at the same time resulting in performance degradation of the model. As an application of minimax optimization inherently present in GANs, [32] proposed item recommendation as a generalized information retrieval task with an objective function of matching top-k relevant documents to the user.

In their paper, [33] proposed a model that set an objective of generating the connectivity distribution for a given vertex. In the recommendation application, the connectivity distribution between a given vertex and all relevant items was discussed. It was discussed by [34] in their paper that treating missing user-item as negative rating can deteriorate the recommendation performance since the negative ratings could just be due to the user unaware of the item. They used GANs to generate pairwise recommendation for each user and item with positive-unlabeled sampling. The idea of using conditional variant of GAN for recommendation was presented by [35] in their research. Their GAN was conditioned on fashion item as a class, given which another complementary item was generated as a recommendation.

A GAN-based approach to handle the problem of data imbalance in recommender systems was proposed by [36]. They made use of conditional Wasserstein GAN to generate missing data for minority class to perform recommendations. Their work used PacGAN in the discriminator architecture with an aim to alleviate the performance of missing data and improve the performance of recommendation models. In another Wasserstein GAN based framework, [37] proposed GAZRec model to generate synthetic feature representations for both cold start news and user. To find the probability of click behavior, their framework adopted a separate click predictor module given a single user and news item. The model did not use the behavioral representations to train the generator for learning distribution of interactions directly for items across different users. Due to the click prediction objective of their work, the recommendation task is reduced to binary classification and could not be extended to allow for multi-label formulation of the problem.

In an earlier work on generative recommendations, [38] proposed autoencoders are generators for collaborative recommendations in CAAE model and to extract latent factors from user-item interactions, however, their framework did not utilize the separate feature space of users and items to make recommendation in case there was a cold start user or product.

IV. GCZREC FRAMEWORK

The architecture of our proposed GCZRec framework consists of dual generator networks, implemented as conditional Wasserstein GAN. The generator for news-to-user interaction is trained on mapping a given news item to a distribution representing users' interest score for the item. Whereas the companion generator for user-to-news interaction is trained to generate a distribution of interaction scores of all news items for a given user. Another important component of the GCZRec model are two independent classifiers for news and users. These pre-trained classifiers are used to perform zero shot prediction of a cold start news or user in order to provide the generator networks their conditional input for synthesizing the interactions.

The proposed framework utilizes generative capabilities of the traditional GAN architecture to synthesize interactions. The individual classifiers are trained to use semantic space and classify both seen and unseen news item and user in order to provide our generators a conditional input. This design of our model also opens the door for a novel way of performing zero-shot extreme multi-label classification efficiently. In the subsequent subsections, each component of our model is discussed in detail. In Fig. 2, the overall architecture of GCZRec is illustrated.

A. NOTATIONAL CONVENTIONS

In the remainder of this paper, the general notation used for a news is N and for active user it is U. We also denote warm news, cold news, warm user and cold user by $\boldsymbol{w_n}$, $\boldsymbol{c_n}$, $\boldsymbol{w_u}$, $\boldsymbol{c_u}$ respectively. The four possible cases to be considered are thus represented as $\boldsymbol{w_n w_u}$, $\boldsymbol{w_n c_u}$, $\boldsymbol{c_n w_u}$, $\boldsymbol{c_n c_u}$. These cases are represented in the model with the help of a 2-bit vector which serves as the item-user state gate g_s and can determine the synthesizer to be used for generating interest vector. We refer to the generator responsible for synthesizing interactions for each user given a particular news item as $\boldsymbol{gen_N}$ and its companion generator which is responsible for generating interactions for each news given a user as $\boldsymbol{gen_U}$. Apart from these generators, the zero-shot classifiers for novel news item and user will be called news label predictor $\boldsymbol{P_N}$ and user label predictor $\boldsymbol{P_U}$. These classifiers are jointly referred to as zero-shot predictors. For encoding the identifiers of warm start news item and warm start users and map them to a unique numeric identifier, the encoders employed are referred to as $\boldsymbol{E_N}$ and $\boldsymbol{E_U}$ respectively.

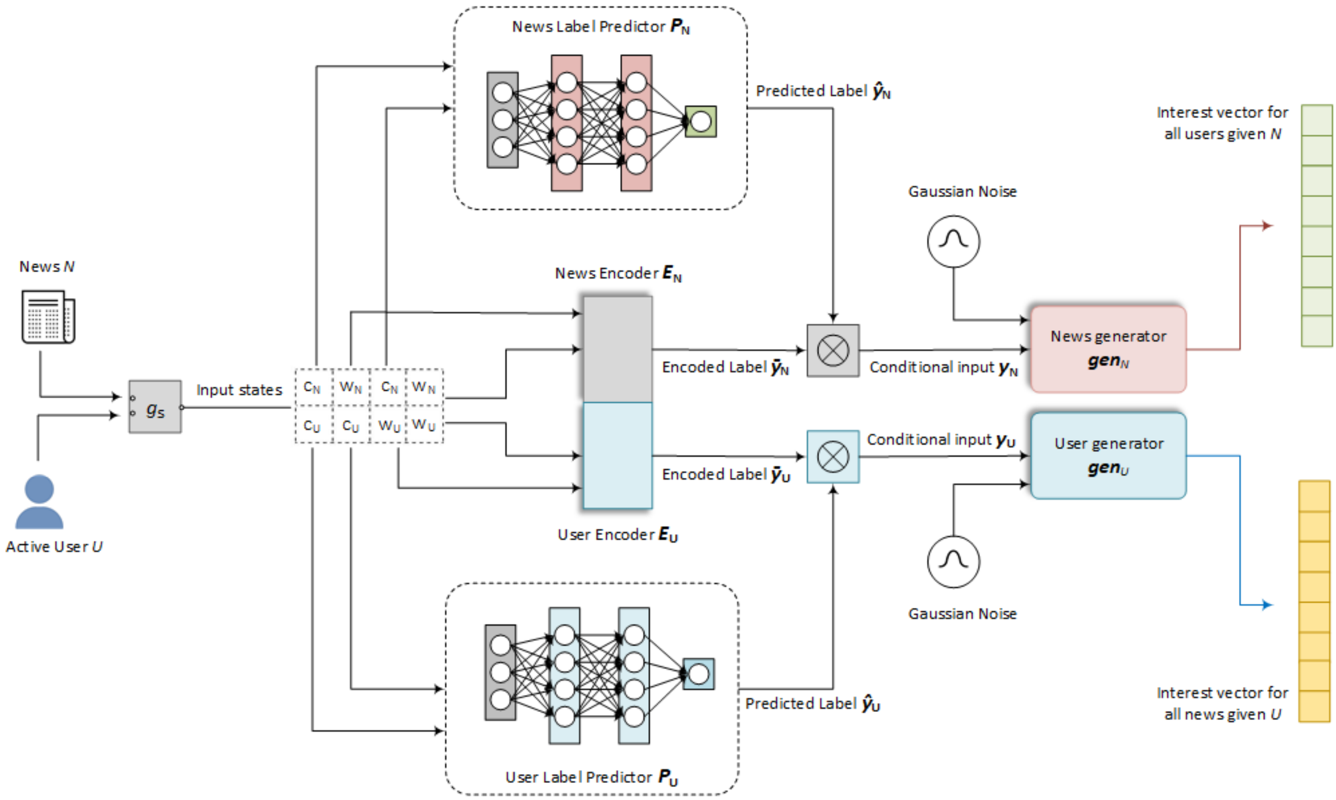


FIGURE 2. Architecture of proposed GCZRec framework.

B. NEWS INTEREST SCORE GENERATOR GEN_N

The generator $\boldsymbol{gen_N}$ in GCZRec framework is responsible for handling the states $\boldsymbol{w_n w_u}$, $\boldsymbol{c_n w_u}$ ***and*** $\boldsymbol{c_n c_u}$. These input states are determined prior to it by state gate g_s. This network synthesizes interaction score for each of the users given the news item label y_N as its conditional input. The relevancy of an active news item for a user can be determined from the corresponding value generated by the network where this value is essentially an interest score. The overall output of $\boldsymbol{gen_N}$ is a vector of interest scores predicted to be given by each user in the system to the active news item. Each position in this vector represents a unique user and the value is a score that shows preference of that particular user for the active item. With the value closer to *+1* meaning that the user would like this news article whereas any score closer to *-1* implying the user's dislike for the item. In state $\boldsymbol{w_n w_u}$, the $\boldsymbol{gen_N}$ takes encoded news label as conditional input from news encoder $\boldsymbol{E_N}$ to generate the interaction scores as its output distribution. For both states $\boldsymbol{c_n w_u}$ and $\boldsymbol{c_n c_u}$ the generator $\boldsymbol{gen_N}$ uses label provided by $\boldsymbol{P_N}$ for synthesizing interaction scores. Due to its stability and the inherent sparsity in the historical interactions present in our data, we used a conditional gan that uses Wasserstein loss called conditional Wasserstein GAN (cWGAN) with the critic network during training to implement the $\boldsymbol{gen_N}$ model.

The general objective function of cWGAN is given as:

$$min_G max_D V(D, G) = \mathbb{E}_{c,x \sim true}[D(x_{true}, c)] - \mathbb{E}_{c,z}[D(G(z, c)), c] \tag{3}$$

In the context of news recommendation, the generator gen_N aims to minimize this combined objective function, while the critic aims to maximize it. This leads to a minimax game where the generator tries to produce realistic synthetic samples, and the critic tries to effectively distinguish between real and synthetic samples. Where $\mathbb{E}_{c,x\ true}[D(x_{true},c)]$ represents expectation over real data whereas $\mathbb{E}_{c,z}[D(G(z,c)),c]$ is the expectation over values generated by synthesis. In terms of generating interest scores of users given news item as conditional input.

The objective function of $\boldsymbol{gen_N}$ can therefore be stated as:

$$minL_{gen_N} = -\mathbb{E}_{y_N, z \sim P_{g(x)}}[D(G(z, y_N)), y_N] \tag{4}$$

Formally, this generator takes random noise z from a guassian distribution $g(x)$ as latent input and given a news class label y_N, it generates a vector of synthetic interest scores $G(\cdot)$ for all users and aims to minimize the distance between fake and ground truth interactions between user-news pairs. The critic evaluates how well the generated scores match real user interest scores given the corresponding news item and produces $D(\cdot)$ which is the output when it evaluates

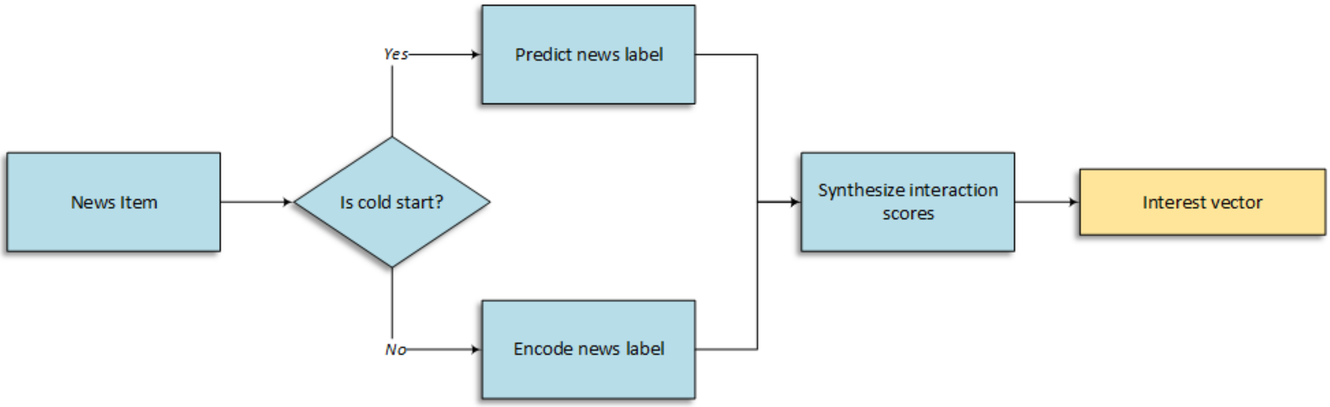


FIGURE 3. Flow diagram of news recommendation in GCZRec framework.

the sample $G(\cdot)$ generated by $\boldsymbol{gen_N}$. As part of adversarial training, the critic network aims to discern the synthetic interaction distribution from the real one that is produced by the generator. The activation used in the dense and output layers of this network are LeakyReLU and tanh respectively. The critic network uses linear activations instead of sigmoid in the output layer and its output is the approximation of Wasserstein distance hence assigning lower values to fake interactions. In the dense layers of this model, LeakyReLU activations are used. During training, the weights of the critic are clamped to a small range and this network is updated five times compared to a single update of the generator in order to improve the generation quality.

C. USER INTEREST SCORE GENERATOR GEN_U

The input states determine the use of $\boldsymbol{gen_U}$ for synthesizing interest scores for active news. These states are handled by g_s. The $\boldsymbol{gen_U}$ takes the user class label y_U of the active user as its conditional input and generates fake interest scores for each news item in the system. The possible states managed by the $\boldsymbol{gen_U}$ are $\boldsymbol{w_n w_u}$, $\boldsymbol{w_n c_u}$ and $\boldsymbol{c_n c_u}$. The output of $\boldsymbol{gen_U}$ is a vector of interest scores showing preferences given by this user to each one of the news items. Each position of this interaction vector representing a unique news and the value at that index indicating a score in range *-1* to *+1* to show if that particular item can be interesting for the active user. For state $\boldsymbol{w_n w_u}$, the conditional variable for this model is provided by the user encoder $\boldsymbol{E_U}$ as $\bar{y}_U$. For cases $\boldsymbol{w_n c_u}$ and $\boldsymbol{c_n c_u}$ the predicted class label $\hat{y}_U$ from the zero-shot predictor P_U is used. Similar to $\boldsymbol{gen_U}$ the training of this network is done in an adversarial manner by employing a cWGAN and a critic that uses Wasserstein loss. The activation in the dense layers of both generator and critic are LeakyReLU while the output layer of the critic uses linear activation and tanh activation is used for the output layer of the synthesizer.

The objective function of $\boldsymbol{gen_U}$ can be stated as:

$$minL_{gen_U} = -\mathbb{E}_{y_U, Z \sim P_{g(x)}}[D(G(z, y_U)), y_U] \tag{5}$$

The model $\boldsymbol{gen_U}$ takes latent vector z from the gaussian distribution as input along with user class label y_U to generate the distribution $G(\cdot)$ of synthetic interest scores for all news items with respect to the active user. The critic outputs its evaluation $D(\cdot)$ of the generated interaction scores produced and $\boldsymbol{gen_U}$ tries to minimize the loss between real and fake distribution of interaction scores.

D. WARM START ENCODERS

For warm start news, the class label $\bar{y}_N$ to be served to the interest score generator $\boldsymbol{gen_N}$ is encoded by mapping the raw identifier of the active news item to a unique numeric id. This encoded id is then used by $\boldsymbol{E_N}$ to collect the corresponding label of the news from historical data. In the same way, warm start user id is encoded by $\boldsymbol{E_U}$ to a unique numeric id in order to extract the available class $\bar{y}_U$ of this active warm start user in order to provide conditional input to $\boldsymbol{gen_U}$ network.

E. LATENT FEATURE REPRESENTATION

In the GCZRec approach, we represent each news item N as a latent feature vector, denoted by δ. This representation is obtained by feature extraction process θ using pre-trained embedding to extract informative features from the textual content of the news item. The feature representation yielded is $\delta = \theta(N)$.

Since the MIND and Addressa datasets do not contain any explicit user entity features, we transformed each user U into latent feature representation λ with the help of her historical interactions with the news categories. All the news the user interacted with previously are treated as positive samples and use to extract the hidden features. These features are constructed as a process ρ which converts the list of categories and subcategories of each interacted news into one-hot encoding. Hence $\lambda = \rho(U)$ becomes the user profile of active user.

F. ZERO-SHOT CLASSIFIERS

The cold start problem for both news items and users is treated as zero-shot classification task in the GCZRec approach. For

TABLE 1. Statistics of adressa and MIND datasets.

	Adressa	MIND
# of users	56733	50000
# of news	11207	51282
# of impressions	201498	156965
# of categories	24	17
avg. # of words per title	51	66
avg. # of words per body	235	205

these zero-shot predictions, we employ two classifiers that use the latent feature representation to predict a class label for item and user. As a result, this allows the predictors to leverage hidden collaborative signals between items and also users for predicting labels in terms of similarity in the latent feature space.

1) NEWS LABEL PREDICTOR P_N

Given the latent news feature representation δ, we classify a novel item into one of K predefined categories, denoted by $y_{N_1}, y_{N_2}, y_{N_3}, \ldots, y_{N_K}$ where K is the total number of news categories in the domain. We implement the news label predictor as a 1D convolutional neural network with softmax activation in the output layer for prediction. The classifier calculates the probability $P(y_{N_i}|\delta)$ for the given news item belonging to class y_{N_i} as stated in equation 6.

$$P(y_{N_i}|\delta) = \frac{e^{w_i} \cdot \delta}{\sum_{j=1}^{k} e^{w_j} \cdot \delta} \tag{6}$$

The assigned news category $\hat{y}_N$ is expressed as:

$$\hat{y}_N = arg\,\max_i P(y_{N_i}|\delta) \tag{7}$$

2) USER LABEL PREDICTOR P_U

This classifier is used to predict the label for a user based on its feature representation λ. Similar to news label predictor, the architecture of this model is a 1D convolutional network with softmax function for finding the probability $P(y_{U_i}|\lambda)$ of the user falling into one of the $y_{U_1}, y_{U_2}, y_{U_3}, \ldots, y_{U_M}$ categories. The posterior probability for finding the user label and label assignment is shown as follows:

$$P(y_{U_i}|\lambda) = \frac{e^{w_i} \cdot \lambda}{\sum_{j=1}^{k} e^{w_j} \cdot \lambda} \tag{8}$$

$$\hat{y}_U = arg\,\max_i P(y_{U_i}|\lambda) \tag{9}$$

V. EXPERIMENTS

A. DATASET DETAILS

For the experiments, we used the publicly available MIND [20] and Adressa [21] news recommendation datasets. The key statistics for both of these datasets are provided in Table 1. The datasets contains click behavior of users for news items. The data include information like impressions, news categories, subcategories, abstract and textual content.

B. DATASET PREPROCESSING

From the users' behavioral data provided including their impressions log and news click history, we first sampled 70% data for training our model and left the 30% for post-training evaluation. For each user, the news item for which they have positive interactions were found by extracting the news id among their news click history and also from the impressions where the user had a "1" as a click behavior for a particular news. We encoded all the positive interactions between user and news as the value "1" during training data construction. The negative interactions between users and news were found when the user did not click the news and hence has a 0 for that particular news id in the impressions log. We encoded this negative interaction as "-1" in the training data. For all the news not present in either a user's historical interactions or impressions log, it was assumed that the user was never show the news item and did not interact with it. These interactions are encoded as "0" for training. Moreover, for indexing purpose, each news id and user id is mapped to a unique numeric news id and numeric user id respectively. Based on their numeric indices, the final training set for $\boldsymbol{gen_N}$ was constructed by using the numeric id as index for a unique instance (row) and each numeric user id as a feature value (column). In a similar manner, the final training data for $\boldsymbol{gen_U}$ was constructed by using numeric user id as index for the instance (row) and each numeric news id as index for feature value (column).

C. IMPLEMENTATION

For constructing latent feature representations, we used hierarchical clustering to assign contextual labels to each news item based on its rich textual features. The number of clusters selected based on silhouette score and discernment was 32. For user labels, the hyperparameter value for number of classes was set to 18 classes. The embedding size for news and user is set to 300 to allow for baseline comparison. For news, pre-trained Word2Vec embedding are used whereas for users we used count vectorization to perform the behavior encoding. The same architecture for $\boldsymbol{gen_N}$ and $\boldsymbol{gen_U}$ is used with a dropout rate of 0.5, learning rate of 0.0002, LeakyReLU as activation in the dense layers, tanh as non-linearity for the generator output layer. Adam is used as optimizer with hyperparameters $\beta 1$ =0.9 and $\beta 2$ =0.999. As part of the cWGAN, the critic is trained with clipped weights. Both the zero-shot predictors are trained as multi-class classifiers with conv1d hidden layers, batch norm regularization, dropout rate of 0.5, learning rate of 0.0005 and softmax activation.

D. BASELINE MODELS

In terms of the recommendation objective, the GCZRec framework is compared with the existing recommender models to validate the performance of the proposed approach. The models are listed as:

- **GAZRec-NPA** [37]: A three-tower generative zero-shot framework to generate generalized behavior

TABLE 2. Comparative results of GCZRec on MIND and adressa datasets in exclusively cold start case.

MIND Dataset					
Method	**AUC**	**ndcg@1**	**ndcg@5**	**ndcg@10**	**MAP**
NAML	0.6140	0.2967	0.5451	0.6580	0.2960
GAZRec-NPA	0.6856	0.3880	0.6290	0.7003	0.3737
GNUD	*nil*	*nil*	*nil*	*nil*	*nil*
GCZRec	0.6970	0.3912	0.6598	0.7435	0.3864
Adressa Dataset					
Method	**AUC**	**ndcg@1**	**ndcg@5**	**ndcg@10**	**MAP**
NAML	0.5384	0.2333	0.4666	0.6086	0.2217
GAZRec-NPA	0.6275	0.4539	0.5339	0.6909	0.2850
GNUD	*nil*	*nil*	*nil*	*nil*	*nil*
GCZRec	0.6103	0.5009	0.6116	0.7281	0.3217

TABLE 3. Comparative results of GCZRec on MIND and adressa datasets in mixed cold start and warm start case.

MIND Dataset					
Method	**AUC**	**ndcg@1**	**ndcg@5**	**ndcg@10**	**MAP**
NAML	0.6800	0.3557	0.5820	0.6847	0.3581
GAZRec-NPA	0.8098	0.5236	0.7379	0.7769	0.5711
GNUD	*nil*	*nil*	*nil*	*nil*	*nil*
GCZRec	0.8287	0.5497	0.7502	0.8105	0.5992
Adressa Dataset					
Method	**AUC**	**ndcg@1**	**ndcg@5**	**ndcg@10**	**MAP**
NAML	0.5835	0.2939	0.5201	0.6377	0.2628
GAZRec-NPA	0.6509	0.5557	0.5703	0.6879	0.3255
GNUD	0.8401	*nil*	*nil*	*nil*	*nil*
GCZRec	0.6320	0.6004	0.6277	0.6870	0.3549

representations of users and items for recommendation and then use these representation for cold start and warm start predictions using a neural click predictor.

- **GNUD** [39]: The user and news interactions are treated as high-order graph in order to exploit latent preference factors of the user to perform recommendation.
- **NAML** [40]: A neural news recommendation approach with attentive multi-view learning in which user representation is learned using their browsed history and other information as well as news attributes such as title and category are used for item representation.

E. EVALUATION METRICS

To evaluate the performance of the proposed GCZRec framework against the baseline, four evaluation measures are used as performance indicators. These metrics are Area Under Curve (AUC), normalized Discounted Cumulative Gain (nDCG@k) and Mean Average Precision (MAP).

The AUC can be measured in terms of true positive rate (TPR) and false positive rate (FPR) as:

$$AUC \approx \sum_{i=1}^{n} \frac{1}{2}(TPR_i + TPR_{i-1})(FPR_i - FPR_{i-1}) \tag{10}$$

The nDCG@k is a measure of ranking quality in the list of recommended items with IDCG as the ideal DCG and position k:

$$nDCG@k = \frac{DCG@k}{IDCG@k} \tag{11}$$

To measure the performance of recommender system using average precision given *top-k* recommendations over multiple values of k we use MAP which is defined as:

$$MAP = \frac{\sum_{k=1}^{K} Average\ Precision@k}{K} \tag{12}$$

F. TEST ENVIRONMENT

For model training and performance evaluation we divided the test data into two distinct sets. From the total test data we selected 50% cold start items for evaluating the model in an exclusively cold start setting. The remaining cold start items along with the warm start data was used to generate recommendations for mixed cold-warm news items.

The threshold value for recommendation of a given item is fixed to 0.5 and the values used for hyperparameter k are 1, 5 and 10.

VI. RESULTS AND DISCUSSION

In this section, the effectiveness of the proposed approach is evaluated and the results indicating the performance on benchmark datasets are reported. These results are summarized in Table 2 for cold start case and in Table 3 for mixed case of both cold start and warm start items.

A break down of model performance into different aspects is needed in order to effectively discuss the outcomes of GCZRec framework. These performance aspects are presented in the following subsections.

A. CLASSIFICATION PERFORMANCE

The performance of generator networks in the GCZRec for scores generation is done in the context of number of correct interest score generation for a given news item. With the help of threshold value, each individual interest score produced in the interest vector can itself be treated as a binary class prediction. The combined performance of these positive and negative scores generation are represented by the AUC values as presented in Table 2 and Table 3. The results show significant improvement in cold start case for MIND but slightly under-performed on Adressa against the baseline for mixed cold-warm start case. This may be due to the label encoding scheme used for the Adressa categories. It can be further investigated whether category condensation in the dataset affected the prediction accuracy.

B. PRECISION-RECALL TRADE-OFF

The GCZRec model offers significant improvement over the existing approaches and the positional relevance of recommended news items are taken into account by the synthetic interaction generators. For both MIND and Adressa dataset, the MAP score given by the proposed model shows an improvement in both cold start and mixed warm-cold start cases. But compared to purely cold start items, the improvements in mixed case recommendations were much more significant. The precision-recall curve for k=1, 5 and 10 for MIND and Adressa datasets in both cases is illustrated in Fig. 4 and Fig. 5.

C. RANKING QUALITY

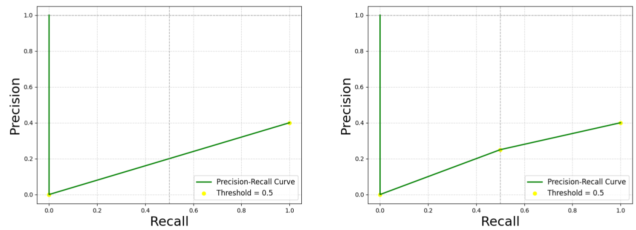
In terms of the ranking quality of news items in both cold start and warm start cases, the proposed GCZRec framework clearly outperforms baseline models with the highest average improvement of +0.1113 is observed when top-5 items are considered as shown by ndcg@k values for k=1, 5 and 10. The overall value of relative ranking in the proposed approach can be attributed to the $\boldsymbol{gen_N}$ and $\boldsymbol{gen_N}$ learning the underlying interest distribution from the data to produce synthetic interest scores. These scores in their raw form are used as is to provide the ranking of relevant items that are recommended.

D. SERENDIPITY

For the inherently challenging and subjective aspect of evaluating the proposed system in terms of expanding the interest of users into neighbouring news categories, we model the results of GCZRec as a collaborative recommendation outcome. This is done in an implicit manner as the output generated by $\boldsymbol{gen_N}$ and $\boldsymbol{gen_U}$ use the interaction between similar user and news. We measure the diversity of recommendations produced using GCZRec framework by finding the percentage of new high interest news item found over 5, 10, 25 and 50 generations given the same user as input to $\boldsymbol{gen_U}$. A summary of this is presented in Table 4. It can also be argued that the diversity is measurable for generations produced by the $\boldsymbol{gen_N}$ in the same manner.

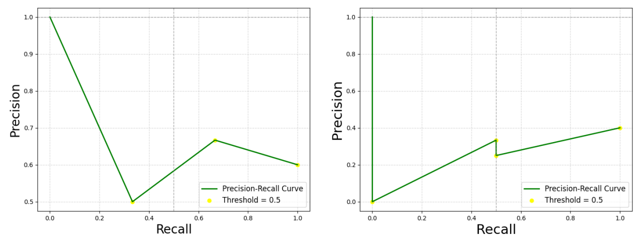
TABLE 4. Percentage of novel news.

Generations	5	10	25	50
% of novel news	0.04	0.06	0.10	0.15



(a) MIND dataset (b) Adressa dataset

FIGURE 4. Precision-recall curve for cold start case.



(a) MIND dataset (b) Adressa dataset

FIGURE 5. Precision-recall curve for mixed cold-warm start case.

Based on the comparative results, it can be stated that the proposed GCZRec framework provides more accurate and relevant ranked recommendation of cold start and warm start news items to users also incorporates diversity by leveraging latent collaborative information present in feature space of users and items.

VII. CONCLUSION

In this paper, we presented the GCZRec framework for cold start news recommendation. We formulated the problem of cold start recommendation as zero-shot classification task and proposed that these recommendations can be diverse and have serendipity if user and item information are implicitly used during training. Unlike existing models, the GCZRec approach allows the interest scores to be generated directly for a given news or user in both warm start and cold start cases. Two separate wCGAN networks are trained on interaction between users and news items in order to allow collaborative signals to be implicitly used for producing synthetic interactions at testing time. For any unseen user or news item, the model makes use of zero-shot predictors implemented as 1D-CNN classifiers. Results on two benchmark datasets indicate that our proposed approach offers significant improvement in the accuracy and ranking of news items for cold start recommendation and also sets a standard for incorporating serendipity by implicitly using collaborative information with a generative recommender system in zero-shot manner.

A. LIMITATIONS AND FUTURE WORK

The current limitation of our model include its inability to consider the correlation between news items and temporal relation between news clicks. Both of these aspects, if incorporated, can be important in further improving the recommendation quality of the GCZRec model.

In future work, we also aim to improve our framework to allow cross-domain recommendation problems to be handled. For this, existing knowledge distillation models can be used to allow learned knowledge from source domain to be transferred to a model set to recommend items that are present in target domain.

REFERENCES

[1] A. Bermes, "Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during COVID-19," *J. Retailing Consum. Services*, vol. 61, Jul. 2021, Art. no. 102555.

[2] S. Feng, J. Meng, and J. Zhang, "News recommendation systems in the era of information overload," *J. Web Eng.*, vol. 20, no. 2, pp. 459–470, Mar. 2021.

[3] M. Zihayat, A. Ayanso, and X. Zhao, "A utility-based news recommendation system," *Decis. Support Syst.*, vol. 117, pp. 14–27, Feb. 2019.

[4] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992.

[5] Y. Koren, S. Rendle, and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2021, pp. 91–142.

[6] P. De Handschutter, N. Gillis, and X. Siebert, "A survey on deep matrix factorizations," *Comput. Sci. Rev.*, vol. 42, Nov. 2021, Art. no. 100423.

[7] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*. Berlin, Germany: Springer, 2007, pp. 325–341.

[8] C. N. Sunilkumar, "A review of movie recommendation system: Limitations, survey and challenges," *ELCVIA Electron. Lett. Comput. Vis. Image Anal.*, vol. 19, no. 3, pp. 18–37, Sep. 2020.

[9] P. Lops, D. Jannach, C. Musto, T. Bogers, and M. Koolen, "Trends in content-based recommendation: Preface to the special issue on recommender systems based on rich item descriptions," *User Model. User-Adapted Interact.*, vol. 29, no. 2, pp. 239–249, Apr. 2019.

[10] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, "Recommender systems challenges and solutions survey," in *Proc. Int. Conf. Innov. Trends Comput. Eng. (ITCE)*, Feb. 2019, pp. 149–155.

[11] J. Li, M. Jing, K. Lu, L. Zhu, Y. Yang, and Z. Huang, "From zero-shot learning to cold-start recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 4189–4196.

[12] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–37, Mar. 2019.

[13] S. Yin and X. Luo, "A survey of learning-based methods for cold-start, social recommendation, and data sparsity in e-commerce recommendation systems," in *Proc. 16th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nov. 2021, pp. 276–283.

[14] T. Wu, E. K.-I. Chio, H.-T. Cheng, Y. Du, S. Rendle, D. Kuzmin, R. Agarwal, L. Zhang, J. Anderson, S. Singh, T. Chandra, E. H. Chi, W. Li, A. Kumar, X. Ma, A. Soares, N. Jindal, and P. Cao, "Zero-shot heterogeneous transfer learning from recommender systems to cold-start search retrieval," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2821–2828.

[15] H. Ding, Y. Ma, A. Deoras, Y. Wang, and H. Wang, "Zero-shot recommender systems," 2021, *arXiv:2105.08318*.

[16] R. J. Ziarani and R. Ravanmehr, "Serendipity in recommender systems: A systematic literature review," *J. Comput. Sci. Technol.*, vol. 36, no. 2, pp. 375–396, Apr. 2021.

[17] D. Kotkov, J. Veijalainen, and S. Wang, "How does serendipity affect diversity in recommender systems? A serendipity-oriented greedy algorithm," *Computing*, vol. 102, no. 2, pp. 393–411, Feb. 2020.

[18] S. Inoue and M. Tokumaru, "Serendipity recommender system for academic disciplines," in *Proc. Joint 11th Int. Conf. Soft Comput. Intell. Syst. 21st Int. Symp. Adv. Intell. Syst. (SCIS-ISIS)*, Dec. 2020, pp. 1–4.

[19] T. Dorjmaa and T. Shin, "Evaluating the quality of recommendation system by using serendipity measure," vol. 25, no. 4, pp. 89–103, 2019.

[20] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, and W. Wu, "MIND: A large-scale dataset for news recommendation," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3597–3606.

[21] J. A. Gulla, L. Zhang, P. Liu, Ö. Özgöbek, and X. Su, "The adressa dataset for news recommendation," in *Proc. Int. Conf. Web Intell.*, Aug. 2017, pp. 1042–1048.

[22] M. Kompan and M. Bieliková, "Content-based news recommendation," in *Proc. 11th Int. Conf. e-commerce Web Technol. (EC-Web)* (Lecture Notes in Business Information Processing). Berlin, Germany: Springer, 2010, pp. 61–72.

[23] M. Capelle, F. Frasincar, M. Moerland, and F. Hogenboom, "Semantics-based news recommendation," in *Proc. 2nd Int. Conf. Web Intell., Mining Semantics*, Jun. 2012, pp. 1–9.

[24] M. Moerland, F. Hogenboom, M. Capelle, and F. Frasincar, "Semantics-based news recommendation with SF-IDF+," in *Proc. 3rd Int. Conf. Web Intell., Mining Semantics*, Jun. 2013, pp. 1–8.

[25] K. Joseph and H. Jiang, "Content based news recommendation via shortest entity distance over knowledge graphs," in *Proc. Companion World Wide Web Conf.*, May 2019, pp. 690–699.

[26] D. Billsus and M. J. Pazzani, "User modeling for adaptive news access," *User Model. User-Adapted Interact.*, vol. 10, pp. 147–180, Jun. 2000.

[27] P. Lenhart and D. Herzog, "Combining content-based and collaborative filtering for personalized sports news recommendations," in *Proc. CBRecSys@ RecSys*, 2016, pp. 3–10.

[28] W. Yang, R. Tang, and L. Lu, "News recommendation method by fusion of content-based recommendation and collaborative filtering," *J. Comput. Appl.*, vol. 36, no. 2, p. 414, 2016.

[29] C. Wu, F. Wu, M. An, Y. Huang, and X. Xie, "Neural news recommendation with topic-aware news representation," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1154–1159.

[30] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, and X. Xie, "Neural news recommendation with long- and short-term user representations," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 336–345.

[31] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, "NPA: Neural news recommendation with personalized attention," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2576–2584.

[32] J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang, "IRGAN: A minimax game for unifying generative and discriminative information retrieval models," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 515–524.

[33] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, "GraphGAN: Graph representation learning with generative adversarial nets," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–8.

[34] Y. Zhou, J. Xu, J. Wu, Z. T. Nasrabadi, E. Korpeoglu, K. Achan, and J. He, "GAN-based recommendation with positive-unlabeled sampling," 2020, *arXiv:2012.06901*.

[35] S. Kumar and M. D. Gupta, "c^+GAN: Complementary fashion item recommendation," 2019, *arXiv:1906.05596*.

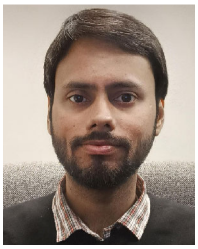
[36] W. Shafqat and Y.-C. Byun, "A hybrid GAN-based approach to solve imbalanced data problem in recommendation systems," *IEEE Access*, vol. 10, pp. 11036–11047, 2022.

[37] M. A. Alshehri and X. Zhang, "Generative adversarial zero-shot learning for cold-start news recommendation," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 26–36.

[38] D.-K. Chae, J. A. Shin, and S.-W. Kim, "Collaborative adversarial autoencoders: An effective collaborative filtering model under the GAN framework," *IEEE Access*, vol. 7, pp. 37650–37663, 2019.

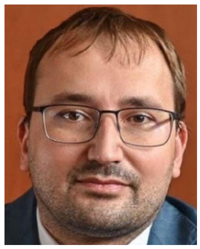
[39] L. Hu, S. Xu, C. Li, C. Yang, C. Shi, N. Duan, X. Xie, and M. Zhou, "Graph neural news recommendation with unsupervised preference disentanglement," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 4255–4264.

[40] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, "Neural news recommendation with attentive multi-view learning," 2019, *arXiv:1907.05576*.



SYED ZAIN UL HASSAN was born in Karachi, Pakistan. He received the M.C.S. degree from the University of Karachi, in 2014, and the M.S. degree in computer science from the National University of Computer and Emerging Sciences, in 2018, where he is currently pursuing the Ph.D. degree in computer science. He has more than eight years of teaching experience and was a Developer at a software service provider for almost two years. His research interests include machine learning, generative AI, recommendation systems, and large language model.

MUHAMMAD RAFI was born in Karachi, Pakistan. He received the B.S. and M.S. degrees in computer science from the FAST-Institute of Computer Science, University of Karachi, Pakistan, in 1996 and 2000, respectively, and the Ph.D. degree in computer science in 2017. He has more than ten years of experience in software development and also a Consultant for the local software industry. His current research interests include algorithm development, machine learning, information retrieval, text/data mining, time series analysis, and natural language processing. He has received several travel grant awards for presenting his work at the top conferences. He has served as a Judge and a Technical Quality Review Team at many versions for IEEEXtreme Programming Competitions. He has served as a reviewer for various international journals of high impact.



JAROSLAV FRNDA (Senior Member, IEEE) was born in Slovakia, in 1989. He received the M.Sc. and Ph.D. degrees from the Department of Telecommunications, VSB—Technical University of Ostrava, Czech Republic, in 2013 and 2018, respectively. He is currently an Assistant Professor with the University of Žilina, Slovakia. He has authored or coauthored more than 85 journal articles in Web of Science. His research interests include the quality of multimedia services in IP networks, data analysis, and machine learning algorithms. In 2022, he was the Finalist of the Category Outstanding Scientist in Slovakia under the age of 35-ESET Science Award, in 2022.