Personality-Driven Social Multimedia Content Recommendation
Qi Yang  
yangqi@itmo.ru  
ITMO University  
Saint Petersburg, Russia  
Alfred Huang  
alfred@somin.ai  
Multimedia Research Lab, SoMin.ai  
Singapore

Sergey Nikolenko  
nikolenko@logic.pdmi.ras.ru  
Steklov Institute of Mathematics at St. Petersburg  
Saint Petersburg, Russia

Aleksandr Farseev  
farseev@itmo.ru  
ITMO University  
Saint Petersburg, Russia

ABSTRACT
Social media marketing plays a vital role in promoting brand and product values to wide audiences. In order to boost their advertising revenues, global media buying platforms such as Facebook Ads constantly reduce the reach of branded organic posts, pushing brands to spend more on paid media ads. In order to run organic and paid social media marketing efficiently, it is necessary to understand the audience, tailoring the content to fit their interests and online behaviours, which is impossible to do manually at a large scale. At the same time, various personality type categorization schemes such as the Myers-Briggs Personality Type indicator make it possible to reveal the dependencies between personality traits and user content preferences on a wider scale by categorizing audience behaviours in a unified and structured manner. Still, McKinsey-style manual categorization is a very labour-intensive task that is probably impractical in a real-world scenario, so automated incorporation of audience behaviour and personality mining into industrial applications is necessary. This problem is yet to be studied in depth by the research community, while the level of impact of different personality traits on content recommendation accuracy has not been widely utilised and comprehensively evaluated so far. Even worse, there is no dataset available for the research community to serve as a benchmark and drive further research in this direction. The present study is one of the first attempts to bridge this important industrial gap, contributing not just a novel personality-driven content recommendation approach and dataset, but also facilitating a real-world ready solution which is scalable and sufficiently accurate to be applied in real-world settings. Specifically, in this work we investigate the impact of human personality traits on the content recommendation model by applying a novel personality-driven multi-view content recommender system called Personality Content Marketing Recommender Engine, or PersiC. Our experimental results and real-world case study demonstrate not just PersiC’s ability to perform efficient human personality-driven multi-view content recommendation, but also allow for actionable digital ad strategy recommendations, which when deployed are able to improve digital advertising efficiency by over 420% as compared to the original human-guided approach.

CCS CONCEPTS
• Information systems → Learning to rank; Multimedia and multimodal retrieval; Computational advertising.

KEYWORDS
User Profiling, Multimedia Retrieval, Machine Learning, Recommender System, Deep Learning

ACM Reference Format:
Qi Yang, Sergey Nikolenko, Alfred Huang, and Aleksandr Farseev. 2022. Personality-Driven Social Multimedia Content Recommendation. In Proceedings of the 30th ACM International Conference on Multimedia (MM ‘22), October 10–14, 2022, Lisbon, Portugal. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3548769

1 INTRODUCTION
Over the past decade, social networks have become an integral part of our lives. Hundreds of millions, if not billions, of people use social networks on a daily basis, reading their feeds, communicating with friends, watching videos... and watching ads. The growing use of social media has led to a corresponding exponential growth in social media advertising, with ads on Facebook and other social media playing an increasingly important role in product promotion and customer human making. At the same time, another important recent trend to note is a decisive decrease of the so-called [50] audience reach, which is the number of people who can see one’s post on social media with no paid digital advertising (also known as Facebook posts boosting) involved. It looks like the users are becoming increasingly harder to “trick” into engaging with the digital content and spreading the word about the ads they see organically, while Facebook itself reduces the organic ad impressions artificially leading brands to scale on digital via paid ads.

Following this global trend, brands and marketing agencies promote products on social media via paid advertising campaigns on ad distribution engines such as Meta Ads (formerly Facebook Ads). A typical campaign consists of multiple ads grouped together, and even in the simplest case an ad can contain both text and image data. Moreover, modern advertisers and agencies, like Somin.ai, are often constrained by particular ad targeting capabilities aimed at
which breaks down human personality into 16 types, one for every composition of four binary components (traits): Extroversion vs. Introversion, Sensing vs. Intuition, Thinking vs. Feeling, and Judging vs. Perceiving. MBTI categorization is widely accepted in psychology and used for personality profiling in, e.g., career planning and market research. However, until now there has not been much research devoted to the incorporation of personality traits into industry-applicable actionable content recommendations, and we are making one of the first large-scale attempts to solve the problem at a new level of adaptability and performance.

Encouraged by the industrial needs and research gap described above, we pose the following research questions:

RQ1 is it possible to predict personality traits from data available to recommender systems in a social network setting?

RQ2 can we improve content recommendations by incorporating personality traits predicted from social media data?

RQ3 what is the difference in content consumption preferences between users exhibiting different personality traits and if such insights could add value to digital advertising routines?

To answer our research questions, in this study we make one of the first attempts to perform personality-driven content recommendation, that is sufficiently accurate to be applied in real-world settings. Particularly, we investigate the impact of human personality on the content recommendation model performance. Our experimental results and real-world case study demonstrate not just PersiC’s ability to perform efficient human personality-driven multi-view content recommendation, but also allow for gaining actionable digital ad strategy insights, which are able to drive 420% digital advertising performance boost as compared to traditional human-based approaches. The paper is organized as follows: in Section 2 we give an overview of previous research on personality traits and recommender systems, Section 3 introduces the datasets used to learn a personality trait model and give recommendations, Section 4 presents our approach, the PersiC framework for recommendation with personality trait features, Section 5 presents a comprehensive experimental evaluation, and Section 6 concludes the paper.

2 RELATED WORK

This work presents a new recommender system with personality features so the nearest prior art is related to recommendation systems. Classical collaborative filtering, i.e., completion of the user-item feedback matrix, is mostly based on matrix factorization (MF) [4, 39, 52]; MF-based approaches still present a competitive baseline, and many variations and applications have been developed over the years [60], but there are several viable alternatives even for the standard matrix completion problem. In particular, we note factorization machines [58] that combine SVMs and matrix factorization and are able to train on implicit feedback. Most popular modern approaches utilize user and item embeddings, vectors in a space that should represent the similarity between users and items that would be useful for recommendations, and the mapping to the embedding space does not have to be linear as in classical MF. In particular, autoencoder-based methods replace direct learning of the user feature matrix with learning a function that maps user feedback to user embeddings; this field began with shallow autoencoders [61, 76] and then moved to variational autoencoders that allow training deeper and more complex models [22, 36, 44, 47, 51, 63]. Another approach to collaborative filtering is based on graph convolutional

1Special thanks to Somin.ai for providing the environment of experiment.
networks, including NFCF [74] and LightGCN [31]; training such networks is computationally intensive but they demonstrate impressive performance, and recent approaches such as GF-CF [62] and UltraGCN [49] improved both performance and computational efficiency. As collaborative filtering baselines, in this work we use basic MF, Neural Collaborative Filtering (NeuCF) that uses deep neural networks to learn user and item embeddings [32], and Bilateral Variational Autoencoder for Collaborative Filtering (BiVAE CF) that learns embeddings via variational autoencoders [70].

We propose a model that takes into account personality traits, so we need recommender systems able to incorporate external information. Such recommender systems usually employ a content analysis module that extracts item representations from unstructured information such as text and images and a user profiling module that extracts user representations from features such as demographics (or, in our case, personality traits). Most modern approaches are actually hybrid recommender systems [9, 19] able to leverage both content/demographics and the user-item matrix; additional information can include external knowledge bases [11], content-related features such as tags or keywords [64], dynamically chosen expert users whose opinions should be trusted more [42], and so on. In embedding-based approaches, additional user and item features are used to inform either the neural network producing these embeddings or the neural network that performs recommendations based on them; such approaches include the widely used DeepCoNN [83], YouTube recommendations [10], embeddings based on topic models [72], and more. As the primary baseline for this work, we selected the Personalized Content Discovery (PCD) model [29] because it was tailored to a similar problem of content discovery for brands but also note several other works that extend recommender systems with extra features and extra data modalities [8, 20, 28, 34, 45, 46, 68, 69, 73, 77, 81].

Finally, we note prior art in the research on personality traits. There have been several research directions that model human personality traits with different statistical approaches. First, the Big Five, or NEO-PI model proposed in [13] was based on the assumption that human personality is reflected in their written language, so statistical analysis of the latter can inform us of the former. Inspired by this idea, the LIWC word categorization scheme [54] later provided a numerical connection between personality traits and written language utilization patterns; relations between the Big Five and MBTI personality types are well known [27].

Automatic personality profiling started in the early 2000s, with cross-disciplinary research using machine learning techniques for automatic human personality inference based on data collected from questionnaires and personality tests [1, 48]. Early studies, however, were conducted on relatively small datasets and did not make use of the huge data provided by social networks. This problem was acknowledged and partially mitigated in the MyPersonality project [40] that was able to provide the first large-scale personality-labeled dataset that includes user-generated data from Facebook. This dataset soon became the basis for larger-scale studies in social media personality profiling research [30, 41, 66]. These studies made a big leap in the field, but most of them were still focused on a single data source such as, e.g., Facebook, or a single data modality, e.g., text. In particular, the Linguistic Inquiry and Word Count (LIWC) works are mostly focused on text processing and predicting personality by using personality-labeled word categories [35, 65], while the works [2, 67] instead utilized pretrained GloVe embeddings [56] and were the first to report the results of machine learning-driven unimodal personality inference. Image-based personality recognition methods can utilize, e.g., correlating specific image features with personality traits [37, 57], localization of attention on images a user liked [84], joint learning of personality types and emotions from facial images [82], and more. Finally, there are user behavior-based personality extraction methods based, e.g., on Facebook likes [66].

Modern social network data, however, is multi-source, multi-view, and multimodal, combining text, images, videos, and other data sources for a single person. There exist several studies that have approached user profiling from a multimodal data perspective. For example, Farseev et al. [21] proposed a multimodal ensemble model for the demographic profiling problem from multimodal data, a work that was later extended to leverage sensor data and multi-source multi-task learning for wellness profiling Farseev and Chua [17, 18]. In Buraya et al. [7], relationship status between social network users was predicted by applying classical machine learning techniques on early-fused data from Twitter, Instagram, Facebook, and Foursquare, achieving a significant 17% increase in performance compared to unimodal learning. Going further, Tsai et al. [71] proposed a factorization method to model the intra-modal and inter-modal relationships within multimodal data inputs, which proved to be important for the incorporation of multimodal data into user profiling, while Buraya et al. [6] instead leveraged the temporal component of the multimodal data, being the first to apply deep learning methods for multi-view personality profiling. While multimodal data has already been tackled in these works, all of them still lacked multi-source cross-social network data processing [16], which limits their applicability in the majority of real-world scenarios. Therefore, in this work, we base our personality profiling on the PERS framework [79] that is able to learn from multi-view data for personality profiling by efficiently leveraging highly varied data from diverse social multimedia sources.

Recommender systems have previously used personality traits in several different ways [12]. Most of them use automatic personality profiling since questionnaire data can hardly be assumed to be available in a real-world setting. Classical approaches use personality features to define a similarity score (proximity function) between users and use it (perhaps together with standard CF proximity) for recommendations [3, 26, 75] or add personality features to matrix factorization models in a way similar to SVD++ [15, 24, 25]. Approaches based on deep learning have only recently begun to incorporate personality features, and so far these approaches have not used standard personality types but rather inferred their own personality feature vectors [23, 33, 43]. In this work, we propose a hybrid approach that uses predicted MBTI personality type to inform a deep-learning-based recommender system.

3 DATASET

Our main contribution in this work is to infer a user’s personality with multi-modal data from their timeline and recommend suitable content based on the user’s content preferences guided by their personality traits. Therefore, to benchmark our model on real-world data, we need a large-scale dataset of user interaction with historical
social media content of various brands; it would also be preferable to cover several different industries in order to minimize industry bias. There exist several datasets which consist of posts from social media networks and personality traits. For example, the myPersonality dataset [40] had been widely used in research but is not available anymore, and we know of no other large-scale efforts that would be suitable for our task. Thus, we need to build our own dataset; in this section, we outline our data gathering and preprocessing methodology.

3.1 Data acquisition

We choose Twitter as our main data source since it is one of the most open social media outlets, known to concentrate on the users’ self-expression rather than their identity and capture more about public personality intended for the broader public because of its high engagement rate. Inspired by [29], we composed a list of official Twitter accounts of various brands and collected all of their historical posts. We then collected a list of users who liked the posts and the corresponding users’ posts and liked posts in their timeline. As for image content of timeline data and brand posts, we selected image previews for video posts and chose the first image for multiple-image posts. Finally, we filtered out brands with less than 100 posts, kept the most recent 100 posts for every brand, and filtered out users who had only one interaction with a brand’s post. The resulting dataset statistics are shown in Table 1.

3.2 Data representation

For all models utilized in this work, we extracted both textual and visual features from the ads as follows.

For **textual features**: 
- for each of Brand posts, we extracted the tf-idf features for every post to form the document-term matrix and then applied latent semantic analysis (LSA) [14] to reduce the textual feature dimension to 100;
- for each user, we concatenated their timeline data into the corresponding user-specific “documents” and then extracted the tf-idf features and the sentiment feature by LIWC lexicon;

For **visual features**, we have chosen to represent visual data in terms of emotional and sentiment-related concepts. To do that, we used a pretrained visual concept detector model SentiBank [5] to extract visual features; the model outputs a distribution of 2089 visual sentiment concepts such as BeautifulNight, HappyFace, or ClassicDesign. For each brand post, we extracted the visual sentiment concept distribution, and to represent a user’s visual preferences we extracted the concept distribution of every image in the user’s timeline and averaged across the concepts, getting a distribution of user preferences with respect to the concepts.

4 PERSIC FRAMEWORK

This section is dedicated to presenting our **Personality-Driven Content Recommendation** (PersiC) framework for the problem of content recommendation for users. We first present the problem setting, then describe the framework itself, and finally report details of the optimization method we used to train the model.

4.1 Problem setting

We denote users by \( \mathcal{U} = \{ u_1, u_2, \ldots, u_i, \ldots, u_n \} \) and brand posts by \( \mathcal{P} = \{ p_1, p_2, \ldots, p_i, \ldots, p_m \} \). In this notation, the model’s goal is to learn a scoring function \( f \) to recommend content for users such that for a post \( p_x \) that a user \( u \) likes and a post \( p_y \) that user \( u \) did not interact with we would have \( f(u, p_x) > f(u, p_y) \). In this work, given a group of brand posts, we aim to learn a ranking model \( f \) to rank the content that the user has not interacted with to indicate which posts have a higher chance that the user will like them according to the user’s content preferences.

4.2 Proposed Method

The structure of the PersiC framework is illustrated in Figure 2. It has two primary components: post representation and user representation learning.

4.2.1 User Representation Learning. The first component is designed to map the users into a common latent space. The most common approach to learn user representations is through the one-hot representation of features [78, 80] or a fixed size latent vector [29]. However, in our case personality traits do not provide enough information to distinguish users: with the default approach users who share very similar personality traits would have very high similarity and we would fail to learn fine-grained user representations and fail to recommend relevant content. For this reason, PersiC learns a fine-grained user representation by utilizing the rich multi-modal data from the user’s timeline in a social media platform, using personality traits as additional features to inform this representation.
Following the data representation method described in Section 3.2, we represent the timeline data of a user $u_i$ as a collection of features:

$$d_i = (T_{pi}, I_{pi}, T_{li}, I_{li}, L_{pi}, L_{li}),$$

where $T$ denotes text features, $I$ represents the average value of the image concepts distribution, and $L$ denotes the LIWC feature [55]. The subscript $p$ represents features extracted from the user’s timeline and $l$ represents features extracted from the user’s historical favorite posts.

To obtain a representation of personality traits, we used the PERS model proposed in [79]. This model predicts user personality traits from multi-modal inputs (collected user posts), producing four classifiers for each of the four MBTI trait pairs (E-I, S-N, T-F, J-P). We use the features from the penultimate layer of the PERS model, with 3 features for each classifier, so in total we obtain a vector $\text{pers}_i \in \mathbb{R}^{12}$. Using this output, we represent the $i$th user with

$$u_i = \text{Concat}(\psi(d_i), \text{pers}_i),$$

where

$$\psi(d_i) = \text{Linear}(\text{Concat}(d_i))$$

is a linear perceptron with ReLU activation that fuses the previously extracted features $d_i$; as a result, we get a user latent representation as $u_i \in \mathbb{R}^{512}$.

4.2.2 Post Representation. The second component of PersiC is obtaining post representations. Pure collaborative filtering methods would represent posts as one-hot encoded items and would not utilize any information from the post; our approach follows the latent representation paradigm (see Section 2 for an overview): PersiC learns from the post’s multi-modal data and projects the post in a common latent space with the users, where recommendations will be made. To achieve this goal, similar to the user representation, we denote the $j$th collection of extracted textual and visual features from a brand’s post by $e_j = (T_{bj}, I_{bj})$, where $T$ denotes text features, $I$ represents the average value of image concepts distribution, and subscript $b$ denotes features extracted from brand posts. We then add one fully connected layer $\gamma$ with ReLU activations to fuse the textual and image features, and obtain the final representation as $p_j = \gamma(\text{Concat}(e_j))$.

4.3 Loss function and optimization

Our dataset $D := \{(u, p)\}$ consists of user-post pairs that show a post liked by a given user. Inspired by [29], we utilize pairwise learning, sampling a fixed number of negative samples for each pair. Based on the assumption that relevant content, i.e., a liked post $p_{pos}$ should have a higher score than irrelevant content $p_{neg}$ for a user $u$, we adopted the Bayesian personalized ranking (BPR) loss function [59] for our model. Thus, for learning we use $(u, p_{pos}, p_{neg})$ triples; to construct them, we take a pair $(u, p_{pos}) \in D$ and uniformly sample 20 negative posts $p_{neg}$ (posts that user $u$ did not like) for each positive pair.

As a result, we learn $f : U \times P \rightarrow \mathbb{R}$ by minimizing

$$L(D) = \sum_{u, p_{pos}, p_{neg} \in D} \ell(u, p_{pos}, p_{neg}),$$

where triples $(u, p_{pos}, p_{neg})$ are sampled as above, and

$$\ell(u, p_{pos}, p_{neg}) = \log \sigma(\hat{y}(u, p_{pos}) - \hat{y}(u, p_{neg})) - \lambda_0\| \Theta \|^2,$$

where $\hat{y}(u, p_{pos})$ and $\hat{y}(u, p_{neg})$ are scores predicted by our model for user $u$ and items $p_{pos}$ and $p_{neg}$ respectively, and $\Theta$ represents the model parameters.

We trained the model with Adam optimizer for 30 epochs, using mini-batches of 64 user-post pairs shuffled at the beginning of every epoch. We also utilized dropout with rate 0.3.

5 EXPERIMENTAL EVALUATION

In this section, we report on our experimental study conducted to evaluate the performance of PersiC. We begin by describing the experimental setup, then provide the results of various tests to illustrate the efficiency of our proposed model, including comparisons with baselines and an ablation study to illustrate the impact of personality traits on performance. Finally, we present qualitative case studies on different personality traits.

5.1 Quality metrics

In order to evaluate the impact of human personality traits on content recommendation, we have used the following standard metrics:

- **Area Under Curve (AUC):** AUC computes the area under the ROC (Receiver Operating Characteristic) curve for classification problems; larger AUC is better;
- **Normalized Discounted Cumulative Gain (nDCG):** nDCG measures the quality of ranking models based on evaluating a ranked list of $n$ top results predicted by the model as follows:

$$\text{nDCG}_n = \frac{\sum_{i=1}^{n} \frac{r_i}{\log_2(i+1)}}{\text{IDCG}_n}, \quad \text{IDCG}_n = \frac{\sum_{i=1}^{n} \frac{r_i}{\log_2(i+1)}}{\text{DCG}_n},$$

where $r_i$ is the ground truth relevance of the $i$th item on the list, and IDCG is the ideal DCG, i.e., DCG that would be obtained if the results were ranked according to their actual relevance, so nDCG is a number between 0 and 1; higher nDCG is better;
- **F1-measure:** $F_1$-measure is defined as the harmonic average of precision and recall, $F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$; higher precision and recall are better, so higher $F_1$-measure is also better.

In this work, we truncated the nDCG and $F_1$-measure ranked lists of results at 10 and 50 respectively.

5.2 Baselines

To the best of our knowledge, there are no preexisting models developed specifically for personality-driven content recommendation.

| Item | Brands | Brands Posts | Interaction | Users | User Posts | User Images | Sparsity |
|------|--------|--------------|-------------|-------|------------|-------------|----------|
| Quantity | 48     | 4800         | 330545      | 41901 | 6547342    | 1407775     | 99.835%  |
based on brand posts on social media platforms. Thus, we compare our approach with several recommender system baselines (see Section 2 for a general survey).

5.2.1 Factorization Machines (FM). Since we are in the implicit feedback setting, for a classical low-rank factorization approach to collaborative filtering we chose factorization machines that show good performance in this setting [58]. For each user $i$ and each item $j$, degree two FM models their possible interaction with a vector $x \in \mathbb{R}^d$ that contains one-hot representations of $i$ and $j$ and any additional features. Then the model learns to predict the target variable $y$ ("click"/"no click" in our case) as

$$y_{\text{FM}}(x) = w_0 + \sum_{k=1}^n \sum_{l=k+1}^d (v_k^T v_l)w_{kl},$$

where $w$ and $W$ are weights of the model (in particular, the matrix $W$ represents weights of interactions between features), and $v_k \in \mathbb{R}^d$ are feature embeddings, so the user and item embeddings $p_i \in \mathbb{R}^d$ and $q_j \in \mathbb{R}^d$ in FM are $v_k$ for the corresponding components of $x$; see [58] for details about optimization in FM.

5.2.2 Neural Collaborative Filtering (NeuCF). This is a popular approach to recommender systems that generalizes matrix factorization to nonlinear mappings by replacing the inner product with the neural matrix perceptron user and item vectors $\hat{\phi}$; this could lead to the generalized matrix factorization $\hat{\phi}_{\text{GMF}} = h(w^T (p_i^\text{MF} \circ q_j^\text{MF}))$ with a weight vector $w$ and activation function $h$, so for $w = 1$ and $h = \text{id}$ GMF degenerates into regular matrix factorization. The multilayer perceptron user and item vectors $p_i^\text{MLP} \in \mathbb{R}^d$ and $q_j^\text{MLP} \in \mathbb{R}^d$ go through a neural architecture (several MLP layers) to obtain $\phi_{ij}^\text{MLP}$, and then the last layer uses the concatenation of $\phi_{ij}^\text{MF}$ and $\phi_{ij}^\text{MLP}$ to predict $\hat{y}_i$; see [32] for details.

5.2.3 Bilateral Variational Autoencoder for Collaborative Filtering (BiVAE-CF). This model, presented in [70] learns embeddings via variational autoencoders. This is a generative model that uses the user-item matrix $X$ to learn latent representations $p_i \in \mathbb{R}^d$ and $q_j \in \mathbb{R}^d$ as follows:

- standard Gaussian priors: $p(p_i) = \mathcal{N}(0, 1)$, $p(q_j) = \mathcal{N}(0, 1)$;
- conditional on $p_i$ and $q_j$, the observations in $X$ are drawn from a distribution from the univariate exponential family

$$p(x_{ij} | p_i, q_j) = h(x_{ij})\exp(p_i(q_j)^\omega)\exp(\theta_j(p_i(q_j)^\omega)),$$

in our case, we used the Bernoulli distribution since the matrix $X$ is binary (lkes);
- the (untractable) posterior $p(P, Q | X)$ in this model is approximated with a tractable distribution

$$q(P, Q | X) = \left(\prod_i q(p_i | X_{i*})\right)\left(\prod_j q(q_j | X_{*j})\right),$$

where $q(p_i | X_{i*}) = \mathcal{N}(p_i | \mu(X_{i*}, \phi), \sigma(X_{i*}, \phi)))$ and $q(q_j | X_{*j}) = \mathcal{N}(q_j | \mu(X_{*j}, \phi), \sigma(X_{*j}, \phi)))$ are Gaussians whose parameters are functions of the corresponding rows and columns of $X$ with additional variational parameters $\phi$ and $\phi$; in BiVAE-CF, these functions are defined by neural networks with parameters $\phi$ and $\phi$, specifically multilayer perceptrons;
- variational parameters $\phi$ and $\phi$ are optimized by minimizing the variational lower bound

$$L_{\text{VAE}} = \sum_{i,j} E_q[\log p(x_{ij} | p_i, q_j)] - \sum_{i} KL(q(p_i | X_{i*})\|p(p_i)) - \sum_{j} KL(q(q_j | X_{*j})\|p(q_j))$$

via the reparametrization trick [38] and alternating optimization with respect to user and item variational parameters (we refer to [70] for details).

We consider BiVAE-CF to be a strong state-of-the-art baseline for collaborative filtering in our setting.

5.2.4 Personalized Content Discovery (PCD). Introduced in [29], PCD is the model which is nearest to ours in the overall setting and data used; in particular, it is able to leverage multi-view data and make full use of the dataset we have collected. PCD is also inspired by matrix factorization but designed to learn latent representations for brands and posts on social networks. It proceeds as follows:

- one-hot brand id representation and brand associations modeled as a matrix of association vectors $A \in \mathbb{R}^{n \times k}$ are combined into a brand representation vector $x_b = \sum_{i=1}^n A_{bi} \circ w_b$, where $w_b$ are importance weights for brand $b$, so a brand is represented as a weighted combination of associations;
- a post with an image is processed first via a pretrained feature extractor network and then with additional two linear layers with leaky ReLU activations to produce a post feature vector $x_p$ for every post $p$;
- finally, the network parameters are trained with a pairwise ranking loss

$$L_{\text{PCD}} = \max(0, f(b, p_{\text{pos}}) - f(b, p_{\text{neg}}) + \eta) + \alpha \sum_b |w_b| + \beta||\theta||_2,$$

where $f(b, p) = \frac{x_b^T x_p}{||x_p||_2}$ is the normalized scalar product, and the $L_1$ regularizer encourages sparsity in the attention weights $w_b$.

As a result, PCD learns latent brand and post representations together with a fine-grained structure of brand associations. To apply PCD to recommendations with implicit feedback, we replace brands with users, and the rest of the model is unchanged.

5.3 Evaluation results

In this section, we evaluate the performance of PersiC against other baselines. We split the dataset into training and test subsets in the 80 : 20 ratio, stratified by the number of each user’s posts. For a fair comparison, we have reproduced all the baselines with the settings suggested in the original papers and optimized on the training set, then evaluated the performance with the test set separately. Results are listed in the Table 2.
### Table 2: Experimental evaluation: performance of PersiC and baseline models.

|       | AUC  | nDCG@10 | nDCG@50 | F1@10 | F1@50 |
|-------|------|---------|---------|-------|-------|
| MF    | 0.801| 0.011   | 0.027   | 0     | 0.006 |
| NeuCF | 0.807| 0.061   | 0.105   | 0.039 | 0.065 |
| BiVAECF| 0.852| 0.075   | 0.115   | 0.045 | 0.048 |
| PCD   | 0.881| 0.082   | 0.121   | 0.048 | 0.091 |
| PersiC| 0.905| 0.092   | 0.125   | 0.052 | 0.095 |

The table clearly shows that PersiC performs best against other baselines in terms of all considered metrics. The baseline models themselves also perform more or less as expected: the simplest MF model is the worst, neural collaborative filtering in the form of NeuCF and BiVAE-CF performed better than MF, but, naturally, they are still inferior when compared to the content-based approaches of PCD and PersiC. Note the very low performance of MF that indicates that the recommendation problem we consider is quite hard. These findings indicate that in this sparse multi-view scenario, straightforward collaborative filtering is insufficient, and additional information in the form of rich multi-modal data from both users and items can significantly improve the performance.

Moreover, the proposed PersiC model has significantly improved results compared to a different commonly used content-based approach, the PCD model, across all metrics and especially for the nDCG metric where PersiC performed 12% better. We attribute this finding to two reasons: first, in the method of learning the user representation PCD only maps a user id into a fixed sized vector in pre-allocated latent space, while PersiC learns a fine-grained user representation by leveraging rich multi-modal data in the user’s timeline and records of their favorite posts, and second and most important, PersiC makes use of the personality traits inferred from the user's social media platform activities. This gives a positive answer to our RQ1: indeed, even automatically inferred personality traits can significantly improve the performance of downstream recommendation models. Overall, Table 2 shows that PersiC is able to adapt to the multi-view sparse environment by learning fine-grained multimodal post and user representations and provide recommendations superior to other approaches.

### 5.4 Ablation study and the influence of personality traits

We have conducted an ablation study to evaluate the importance of various features for the user representation and their influence on the final results in terms of recommendation quality metrics. Table 3 shows the results of this ablation study.

First, obviously, the simple one-hot encoding obtains the worst results since it does not utilize additional user information at all. Learning user representations only from historical post data achieved significantly better scores. Comparing to the performance of PCD in Table 2, we also see that the performance is improved by leveraging the user’s historical data.

Table 3 shows that posts are more useful than likes for the quality of downstream recommendations; note also that the performance generally improves only a little compared to the one-hot encoding if we learn the user representation by the user’s historical favorite posts. This can be explained by the high noise in users’ historical favorite posts and may be impacted by the algorithm of content recommendation in the social media platform. Finally, comparing the performance of “Posts+Likes+Pers” with “Posts+Likes” and “Posts+Pers” with “Posts”, we see significant performance boosts from adding user personality features. This finding answers positively our RQ2: introducing personality features into a content recommendation system has been able to substantially improve personalized content recommendation performance.

### 5.5 Qualitative results and case study

To further evaluate the performance of PersiC, we have conducted a case study with real-world data collected from the audience of a sports brand (hereinafter called Brand). Brand is a major Decathlon Competitor in one of the European markets. The main goal of the Brand is to increase the engagement rate and return on advertising spend (ROAS) by designing better content and choosing proper psychographics. We have collected historical posts from Brand’s
We then further harvested timeline data from the audiences that had interacted with Brand’s posts. Next, as described in Section 4, we extracted visual concepts from Brand’s and competitor’s timelines. Finally, we extracted audience-specific concepts and inferred psychographic attributes from the audiences’ timelines.

We begin with a qualitative picture of PersiC’s results. Figure 3 shows image concepts mined by SentiBank that have turned out to be most highly correlated with the 16 MBTI personality categories based on images from Brand’s timeline, and Fig. 4 shows several sample images for each personality trait. Note that the distributions of concepts are very different across personality traits; in particular, there are several strong positive correlations: HotLegs, SexyGirls, YoungCouple, AwesomeTattoo, and YoungFriends for the Introvert trait and Text Block, Beautiful Night, NiceGuy, AwesomeDesign, and LonelyGuy with the Introvert trait. This observation conforms well with the theory that extroverts are more active and enjoy social interaction while introverts are thought-oriented and enjoy spending time alone [53]. We also note similarities and differences in concept distributions between traits. For example, Introvert and Judging are significantly positively correlated to similar concepts such as text and design, reflecting the fact that both traits are thought-oriented and prefer structured, firm decisions [53]. We can also verify this observation on Fig. 4: Introvert and Judging traits share similar visual content preferences of numbers, deals, and similar shapes. Fig. 3 can yield many such observations: e.g., SexyGirls are most important for Extrovert and Perceiving traits, while NiceGuy is best suited for Introvert and Thinking traits. Overall, we have found that the distributions of concepts are reasonable and conform well to the assumptions behind this study.

In this practical case study for Brand, we have gone beyond just making recommendations with PersiC into qualitative content generation insights. We have extracted from PersiC the learned mappings and correlations between user content preferences with Brand’s post concepts and the common concepts for each personality trait (cf. Figure 3). Then, we have studied insights from the concepts (cf. Figure 4) to guide Brand’s content design.

Based on these psychographic insights, with the help of Somin.ai platform, we have automatically created an advertising campaign for the Brand on Meta (a.k.a. Facebook Ads) that leverages predicted personality traits and other user interests [7]. The above has helped to achieve more than 4.2x cost reduction (for cost per result) compared to the traditional approach adopted by the Brand’s agency. This growth in conversion rate has further led to 6.11x more app installs (the advertising budget was increased proportionally) while bringing more insights into the audience’s psychographic behavioural traits.

To be precise, we were able to correlate back the best ad units and the corresponding user personality traits, which brought another insight for the Brand: it turned out that purchasing decisions of the majority of the Brand’s e-commerce customers are shifted towards long-term implications of owning a product rather than particular product features or its price. As a result, Brand has modified the content of their reach and awareness ads that were previously focused on feature- and price-centric content and achieved further improvement of their digital advertising campaign strategy. New fresh audiences visiting the Brand’s website were receiving a consistent message on the “last mile” of their purchasing journey, just as they did at the stage of familiarizing themselves with the product.

6 CONCLUSION

In this work, we have presented PersiC, a novel personality-driven multi-view content recommender system, which is driven by personality traits inferred from user activities on social media. PersiC is able to capture fine-grained user representations by extracting multi-view features from the user’s posts to provide personalized recommendations. Moreover, we have shown a case study with a very successful real-life advertising campaign (improving ROAS by > 4x and app installs by > 6x) guided by insights learned by PersiC. Finally, we also publishing our multi-view large-scale content recommendation dataset for further research in this exciting direction.

For further work, we note that although the PersiC framework improves recommendation performance according to user content preferences, there is an important potential issue that can pollute the results: user activities on their timeline are impacted by content recommendation algorithms from social media platforms such as Twitter or Facebook. Therefore, we see the further work’s backbone direction in studying what is the impact of recommendation algorithms from social media platforms on user activities and how it can be accounted for in recommender systems.

7 ACKNOWLEDGEMENT

This work was funded by the Russian Science Foundation grant № 22-11-00135 https://rscf.ru/en/project/22-11-00135/.

REFERENCES

[1] Shlomo Argamon, Sudhant Dhawle, Moshe Koppel, and James W. Pennebaker. 2005. Lexical Predictors Of Personality Type. In Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America.
[2] Pierre-Hadrien Arroux, Anbang Xu, Neil Boyette, Jalal Mahmud, Rama Akkiraju, and Vibha Sinha. 2017. 25 Tweets to Know You: A New Model to Predict Personality with Social Media. Proceedings of the International AAAI Conference on Web and Social Media 11, 1 (May 2017). https://ojs.aaai.org/index.php/ICWSM/article/view/14963
