

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Are You a Risk Taker? Adversarial Learning of Asymmetric Cross-Domain Alignment for Risk Tolerance Prediction

Zhe Liu<sup>1</sup>, Lina Yao<sup>1</sup>, Xianzhi Wang<sup>2</sup>, Lei Bai<sup>1</sup>, Jake An<sup>3</sup>

<sup>1</sup>School of Computer Science and Engineering, University of New South Wales, Australia

<sup>2</sup>School of Computer Science, University of Technology Sydney, Australia

<sup>1</sup>{zhe.liu1,lei.bai}@student.unsw.edu.au

<sup>1</sup>{lina.yao}@unsw.edu.au

<sup>2</sup>xianzhi.wang@uts.edu.au

<sup>3</sup>jake@raizinvest.com.au

**Abstract**—Most current studies on survey analysis and risk tolerance modelling lack professional knowledge and domain-specific models. Given the effectiveness of generative adversarial learning in cross-domain information, we design an Asymmetric cross-Domain Generative Adversarial Network (ADGAN) for domain scale inequality. ADGAN utilizes the information-sufficient domain to provide extra information to improve the representation learning on the information-insufficient domain via domain alignment. We provide data analysis and user model on two data sources: Consumer Consumption Information and Survey Information. We further test ADGAN on a real-world dataset with view embedding structures and show ADGAN can better deal with the class imbalance and unqualified data space than state-of-the-art, demonstrating the effectiveness of leveraging asymmetrical domain information.

**Index Terms**—User Behavior Modelling; Generative Adversarial Network; Cross-Domain Representation

## I. INTRODUCTION

E-finance has fundamentally changed the landscape and ways of information transfer in the finance industry [1]. Consumers nowadays can easily access their assets and manage their investments via mobile devices. Given the massiveness of consumers and intricate analysis work, finance analysts and agents are increasingly relying on machine learning to suggest suitable portfolios to consumers, where Financial Risk Tolerance (FRT)<sup>1</sup> serves as a crucial measure to assess individuals' risk preferences on investment choices [3]. To accommodate consumers' risk expectations, the standard practice for recommending investment portfolios is to design questionnaires to survey consumers and then to assess their FRT. While survey results show significant correlations between consumers' feedbacks on the questionnaire and their real FRT levels [2], [4], [5], consumers' feedbacks are liable to be biased by consumers investigated and inauthentic information. Besides, consumers providing similar feedbacks may have differed FRT (i.e., the

multivalued function phenomenon), and consumers' FRT often has an imbalanced class distribution.

Consumers' online activities, especially e-commerce transactions, are potentially an excellent source of information for assessing consumers' risk preferences. However, existing researches on online activities focus on predicting consumers' future behaviors based on their past online activities [6] while rarely considering the possible hidden relationship between the patterns and FRT. Some studies suggest that shallow online activities such as *clicks*, *adding to cart*, and *purchasing* could imply deeper consumers' behavior patterns on consumption habits and preferences [7], [8]. Nevertheless, they have not considered domain knowledge to help them fully leverage the data.

Leveraging both consumer surveys and online activities for FRT would require aligning records from the two sources and learning unified presentations from data. The above poses significant challenges to the current research. First, there lack unified supervised methods to use non-pairwise domain data in a generative adversarial domain alignment. Second, besides the challenges for each domain dataset, the data from different domains might be of different sizes and have small overlaps, meaning they are hardly aligned. Given Generative Adversarial Network (GAN) [9], [10]'s outstanding performance in generating samples, we consider it to be a promising solution to address the above challenges. Although some previous studies have applied GAN for domain alignment [11]–[13], e.g., using records of the same object from different days to co-train the embedding layers [12], the existing studies fail to notice the scale asymmetry of multi-domain data and leverage this characteristic to enhance the representation learning of the insufficient domain in the unified model.

We make the following contributions in this paper:

- We analyze consumers' transaction information (Section III-A) and capture their consumption structures (Stratum Feature) and shopping affiliation (Life Feature) using a domain-specific consumer consumption model to assist

<sup>1</sup>Financial risk tolerance refers to the degree of variability that the consumers can accept the negative changes in the value of investment or an outcome that is adversely different from the expected one [2].

further analysis.

- We develop a novel ADGAN (Section IV) suitable for handling scale asymmetric multi-domain data. ADGAN employs two specially designed batch construction methods in a unified model for domain alignment and insufficient domain learning (Section III-B). It shows significant improvement when compared to state-of-the-art algorithms, demonstrating the advantages of leveraging asymmetrical domain information.
- We design data-specific view embedding structures (Section IV-B) to extract information from raw data and a Gaussian noise fusion (Section IV-C) to overcome the multivalued function phenomenon in survey data.
- Our model shows robustness in handling imbalanced cross-domain data. Our experimental results on a real-world dataset demonstrate its superiority to a series of baselines and state-of-the-art methods.

## II. RELATED WORK

### A. FRT-related Theories

Current business researchers show not only demographics, environmental factors but also psychology, economics, and biosociology could influence consumers' judgments to finance-related decisions [2], [14]. They have designed factor-specific questionnaires to measure and analyze FRT. For example, questionnaires designed from the philanthropy perspective revealed that consumers' corporate social responsibility would affect their purchase behaviors [15], [16]; an investigation of 726 consumers [17] confirmed the positive correlation between environmental psychology with financial decisions; the collectivism-individualism theory [18] suggested different types of people would make decisions driven by disparate deep reasons—some were motivated by personal preference and demands while others valued more about the entity benefits. Roehrich et al. [19] indicated that the consumer's innovativeness was the main force to prompt an innovative behavior.

The above business researches concentrated on building a systematic consumer model with nameable attributes. Such researches can explain the meaning and implication of features but provide little help in designing related algorithms [3]. Some studies [20] further apply linear regression and deep learning to achieve a more competitive model, but they generally only use single-domain data rather than multi-source information and thus have limited performance.

### B. Consumer Behavior Models

Traditional consumer models [6], [7], [21], [22] only extracted shallow patterns from data to explain and improve models. For example, Qiu et al. [6] designed a two-stage theory that uses a motivation factor and a product choice factor to simulate the consumer's decision-making process. Kooti et al. [7] extracted the personal financial background information and combined them with demographic data to characterize and predict consumer behavior. They also studied the periodical behavior patterns, such as the intervals of two

purchase and the frequency of shopping online, to make the model more explainable, and achieved excellent performance. However, the shallow patterns may be difficult to acquire the actual consumer mind patterns.

Some work focused on extracting high-level theoretical information from consumers' behaviors. For example, Joo et al. [23] extracted the hidden high-level data patterns—social traits [24]. According to the psychological theory, they utilized social traits to help them improve and explain how people make decisions in elections and estimate social relationships. Similarly, Kooti et al. [8] annotated the dataset with satisfaction and intent in searching and then designed the models to predict the possible feedback from the consumers to explain why machine learning chose the searching results. In consumer financial decision field, the business could provide professional domain theory for deep learning to model consumers. For example, Jisana et al. [25] introduced the frameworks of consumer models, such as the Maslow's hierarchy of needs [26] which described the deep inside need to explain the action of the purchase; AIO theory (Activities, Interests, and Opinions) [27] which depicted the daily life reason for purchasing an item. These consumption models could effectively extract consumer behavior patterns and catching their thoughts.

### C. GAN-related Work

Due to the limitation of the survey dataset and the gap between the shopping behavior and FRT, we propose to transfer the consumption information to survey aspect for obtaining the biopsychosocial explanation and improving dataset quality. The generative model has been widely used for generating samples and domain alignment. The advantage of GAN is that it only relies on the backpropagation to obtain gradients and require no complicated inference during training. Therefore, various factors and interactions can be incorporated into GAN. Our work is related to Conditional Generative Adversarial Network (CGAN) [9], which viewed the real samples as conditional constraints to guide the data generation process. Conditions could be any additional information, such as class labels and other modal data, which makes GAN applicable to cross-domain problems. Based on CGAN, Zhu et al. [12] studied the electronic signal shifting problem in the brain-machine interface. They leveraged the core idea of GAN and used the GAN structure as the domain aligner to regularize the variational autoencoder. The aligner used the signals of arbitrarily days to generate signals of the first day, and the generated samples would co-training the embedding layers. Farshchian et al. utilized the text data to generate the corresponding images data to expand the sample space of the certain class, which could ease the class missing or imbalance problem. Bousmalis et al. [13] conducted the domain alignment by learning the picture style shift patterns under the unsupervised way, which could dismiss the limit of pairwise data. However, the above work failed to combine the non-pairwise data and pairwise data analysis in a unified network.

Our work further develops the characters of the above work into a unified structure which exploits the advantages

TABLE I  
DATA STATISTICS

Questionnaire Information		Consumer Consumption Information	
#Feedback	4,492	#Records	8,297,231
#Question	52	#Consumer	104,960
Option	1–7	AVG Expense	\$30,934
#FRT Class	4	AVG #Trans	79
AVG Score	5.43	Total Expense	\$18.98–\$1.01e+8

TABLE II  
PURCHASING FEATURES OF CONSUMERS

BASIC	SOCIAL	SELF	RESTAURANT	ENTERTAINMENT
SERVICE	TRAVEL	SHOP	HEALTH	WORK
CREDIT	HOME	DAILY	INVESTMENT	BILL
GAMBLING	EDUCATION	CHARITY	FASHION	TAX

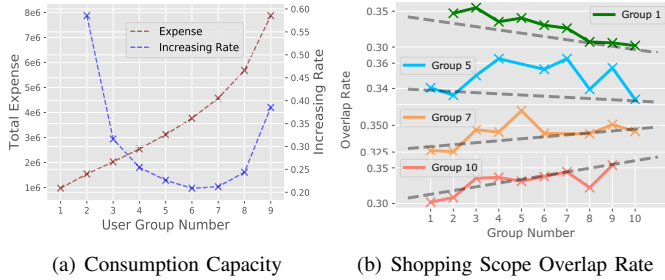


Fig. 1. (a) Illustration of consumption capacity. We omit Group 10, which is too large compared to other groups. Note Group 10 has the total expense of nearly \$19.12e+8 and a 5-time increasing rate. (b) Illustration of four representative groups’ (group 1, 3, 7, and 10) pairwise overlap rates with other groups.

of business theory and deep learning models. We design two training batches to help our model conduct the asymmetrical domain alignment and the survey analysis enhancement.

### III. EMPIRICAL ANALYSIS

We study the FTR prediction problem through the transaction and questionnaire information from an asymmetric two-source dataset. The dataset consists of 104,960 consumers’ consumption activities in a month while only 4,492 of the consumers provide questionnaire feedback. Table I shows basic statistics of the dataset.

#### A. Consumption Activity Analysis

We divide consumers into ten groups in ascending order of their total personal expenses to discover consumption patterns based on the groups and investigate the overlap ratio of shopping scopes of two groups to deep mining the shopping behavior relationship between groups:

$$Overlap(a, b) = \frac{\#(a \cap b)}{\#(a \cup b)} \quad (1)$$

where  $a$  and  $b$  denote two arbitrary groups’ shopping scopes and  $\#$  denote the element number in a set.

Fig. 1(a) exhibits the total expense grows, then stays stable, and finally soars from group 1 to group 9 as the increasing rate first drops, indicating large gaps between the consumption capacities of different consumer groups. Fig. 1(b) provides further analysis with relationship between the consumption capacity and shopping choices. The first few groups’ overlap rate trend are descending, the middle groups’ are stable and the last few groups are ascending.

Based on the patterns in consumption capacity and shopping scopes, we conjecture three possible consumer categories:

- *Weak capacity consumers.* The first few groups only share more similar shopping choices with the consumers who have low expenses in the month.
- *Ordinary capacity consumers.* The middle few groups share a relatively higher ratio of choices with all groups than other groups; they have more common with near groups.
- *Strong capacity consumers.* The last few groups share more common with those who spend much money.

#### B. Questionnaire Activity Analysis

The biopsychosocial questionnaire investigates 4,492 consumers’ feedback on 52 questions related to public concerns [15], [17], collectivism-individualism [18], [28], and innovativeness [19], [29]. All questions are answered on a 7-point scale, ranging from 1 (never or definitely no) and 7 (always or definitely yes). Sample questions from each dimension are: ‘*Is mankind severely abusing the environment?*’, ‘*Do I feel good when I cooperate with others?*’, and ‘*Do I know more than others on latest new products?*’.

The consumers have an imbalanced distribution over four FRT classes—there are 700, 997, 2076, and 719 consumers in the four classes from low to high FRT. Fig. 2(a) shows the number of consumers from different classes while they have little difference in the features. We have four consumers have the same features while they choose different portfolios from different FRT levels; near 50 consumers have different FRT while their feedbacks to 95% of the investigated questions are similar, showing the multivalued function phenomenon in the data. Therefore, the data space is under-qualified, and we need to improve the data quality by using other domain information.

Fig. 2(b) shows the data distributions on several survey dimensions, where most feedback is close to mean value (within  $\pm 1$  from the mean score). The survey data show an imbalanced class distribution where 46.2% of the consumers belong to the third class, and there exist many similar consumers from different classes. The overlapping rate between two domains of data sources is only 4.3%.

#### C. Consumption Feature Representation

Based on the assumption of three consumer types in Section III-A, we build *Consumer Consumption Model* (Table II) to portray the consumer shopping capacity and intent. The Stratum Features are in bold font and the others are Life Features.

**Stratum Features.** We design *Stratum Features* by dividing consumers’ consumption behaviors according to the Maslow’s

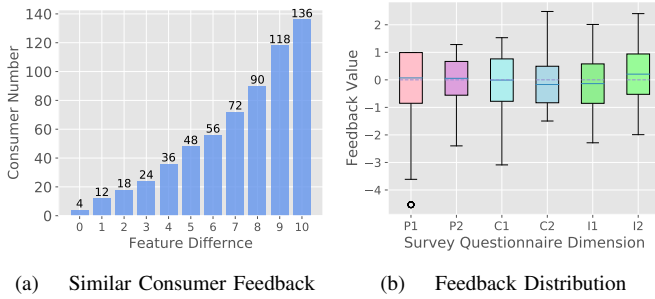


Fig. 2. (a) plots the multivalued function phenomenon in the questionnaire data. The number of consumers who have different FRT classes but give similar feedback in an accumulative way. (b) plots six features of questionnaires to show the limited scope in questions. We pick two questions from each dimension, name them P1, P2, C1, C2, I1, I2, respectively. The blue line and purple dashed line denote the median line and mean line of single feature distribution.

hierarchy of needs [26] to portray the relationship between shopping choices and consumption capacity: the human needs is hierarchical and higher-layer needs only occur when the lower-layer needs are satisfied. Therefore, We sort and classify the transaction into three hierarchical categories to represent the shopping capacity from different layer needs: *BASIC*, the consumption to address the fundamental needs for a living; *SOCIAL* for enjoyment and improving life quality; and *SELF*, for enriching the mental life. We further summarize stratum features of consumer  $i$  by

$$Score(y)_i = \frac{Expense(y)_i}{\sum_{Stratum} Expense(y)_i} \quad (2)$$

where  $y \in Stratum$  denote a dimension of Stratum feature and  $Expense_i$  is the total personal expense of consumer  $i$  in the corresponding dimension. The Stratum Feature is decided by the ratio to the total personal expense.

We have the following findings from the above stratum feature scores and group types: (i) The Maslow’s need hierarchy could capture the different need structures of different shopping groups; (ii) The significant gaps (shown in Fig. 1) can be eliminated by using stratum features; (iii) Stratum features characterize the patterns in Fig. 1, indicated by the ordinary, weak, and strong purchasing-capacity groups.

**Life Features.** We design *Life Features* to describe the shopping structure of consumers based on the AIO theory (Activities, Interests, and Opinions) [27], which covers all the life aspects of purchases. We look into consumers’ shopping frequency and choices of targets beyond the three general types. One example is that, generally, a small portion of consumers contribute to the majority of sales in a market (the “20%:80%” rule [30]). Another example is that a company selling clothes will be more interested in consumers who frequently buy fashion products instead of consumers who only spend money at home. Therefore, we similarly label the transaction into 17 subgroups based on the transaction categories, and summarize the frequencies (Table II) of consumer

$i$  by

$$Score(y)_i = \frac{Frequency(y)_i}{Max(\{Frequency(y)_i : i \in Consumers\})} \quad (3)$$

where  $y \in Life$  denote a dimension of Stratum feature and  $Frequency_i$  is the frequency of the consumer  $i$  shopping in the corresponding dimension. The Life Feature is decided by the ratio to the max frequency of all consumers in the corresponding life aspect. We draw the following conclusions from the above observations: (i) Life features indicate the structures of consumers’ purchase types; (ii) Life features reflect the shopping choice characteristics, overlapping choices among different groups, and shopping scopes of consumers.

#### IV. METHODOLOGY

To address the bad distribution problem posed by the survey in Section III-B and difficulties in utilizing consumer activity data, we first propose a domain-specific consumer model to analyze consumers’ consumption activities, followed by utilizing GAN and Gaussian Noise to improve the dataset quality and to ease the class imbalance and the multivalued function issues.

##### A. Problem Description

Before introducing the methodology, we first define our problem. We denote the consumer entity by  $E$  and the two related information domains as survey information  $S^s \in \mathbb{R}^+$  and activity information  $S^a = \{Activity(i) : i \in E\}$ , where  $S^s$  consists of numeric feedback of questionnaires,  $Activity(i)$  is the set of consumption activity records of consumer  $i$ . We split the entity  $E$  into two groups according to their known information as  $E_l$  and  $E_f$ : if consumer  $i$  has provided both survey information and activity information, then  $i \in E_f$ ; if consumer  $i$  only provides the activity information, then  $i \in E_l$ . Our goal is to predict the FRT class  $L$  for consumers. Our core approach aims at utilizing the different combination of consumers from  $E_f$  and  $E_l$  to asynchronously train our model to achieve the above characters.

##### B. Model Formulation

Our network Asymmetric Cross-Domain Generative Adversarial Network (ADGAN) takes consumer survey and consumption model information (Section III-C) as inputs. ADGAN consists of three main structure: (1) Structural View Embedding, which effectively extracts the domain information, (2) Generator  $G$ , which generates the missing survey data, and (3) Discriminator  $D$ , which distinguishes fake from true samples and predicts the FRT labels.

**Structural View Embedding.** We design two types of domain-specific Structural View Embedding to conduct the domain alignment: Consumer Consumption Model Embedding and Survey View Embedding. We plot the schematic diagram in Fig. 3. Since survey information  $S^s$  and consumer consumption model  $U$  can be divided into hierarchical structure according to the domain knowledge, it is intuitive to firstly extract the sub-dimension information, and then gather the

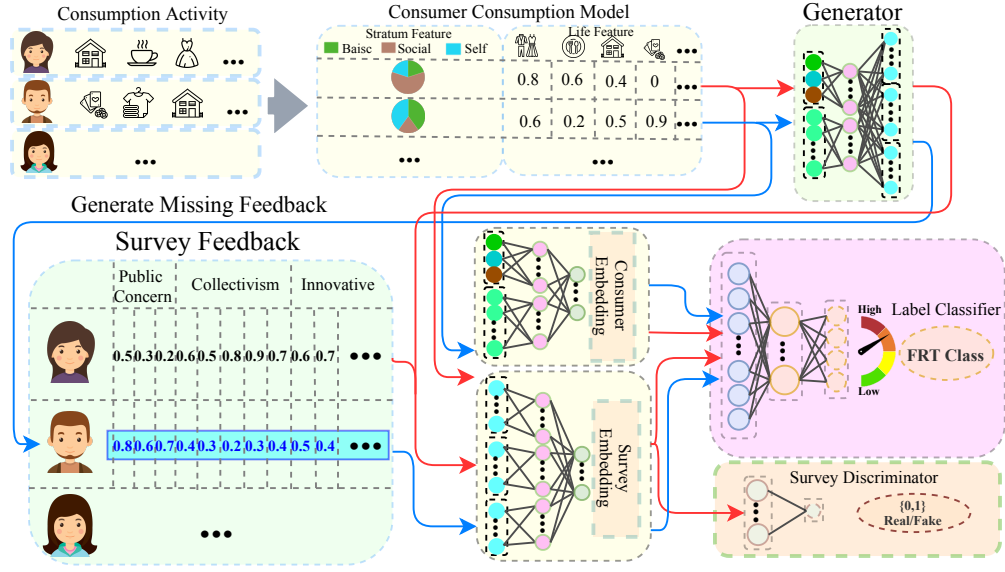


Fig. 3. Model overview. Our model takes the extracted consumer model information (Section III-C) and the survey information as inputs. We show the how consumer data influence the modules in the model from  $E_l$  and  $E_f$  with blue and red lines, respectively. When training the Survey Discriminator, our model compares the generated survey information with the ground truth in the red path. When training Label Classifier and Generator, we take the generated survey as the ‘real’ features of consumers from  $E_l$ , and feed the model with the consumers from both  $E_l$  and  $E_f$ .

dimension information to obtain the final embedding. For example, we will first analyze the public concerns, collectivism-individualism, and innovativeness, respectively, and then combine the analyzed information to measure the person’s FRT. Moreover, we will show through comparisons (Section V-B) that domain-specific embedding is effective and essential for extracting information from the provided data. Specifically, we apply a fully connected (FC) layer to extract information for each dimension and then combine the information using another FC layer. The FC layers in the Structural View Embedding will be followed by the activator–Rectified Linear Unit (ReLU).

**Discriminator D.** The structure of  $D$  consists of two channels of inputs, and we use an FC layer activated by ReLU to receive the embedding information. We also set a FC layer with Sigmoid Function as the Survey Discriminator to distinguish the real and fake survey data, and an FC layer with Softmax Function as Label Classifier to predict the FRT classes. The structure and the training method of our  $D$  are unique because we have two channels of inputs and two types of consumer groups.

Before introducing the loss function of our model, we introduce the batch construction. We define two types of input combination (detailed in Section IV-C): First, given a consumer batch  $B_a$ , an arbitrary consumer  $i \in B_a$  meets  $\forall i \in E_f$ —When we use the bath  $B_a$ , we have the pairwise survey ground truth and generated survey, in which case we can train the Discriminator how to distinguish fake data from true data; if the Discriminator can learn the patterns well, it can further help us improve the generated samples of consumers from  $E_l$ . Second, given a consumer batch  $B_b$ , there exist two

consumers  $i, j \in B_b$ ,  $i \in E_f$  and  $j \in E_l$ —When we use the batch  $B_b$ , we have the ground truth data of the consumption activity and FRT labels of consumers, in which case, if we have the well-trained Discriminator, we can improve our generator to interpret the consumer activity by providing the corresponding missed survey data under the condition of their FRT label and consumer model information; further,  $E_l$  have more consumers from all labels, so the generated survey data can help ease the class imbalance problem of  $E_f$ .

Using the batch  $B_a$ , we have:

$$L_D = \mathbb{E}_{U \sim \{U_i: i \in B_a\}} [D_w(G_\theta(U))] - \mathbb{E}_{S^s \sim \{S_i^s: i \in B_a\}} [D_w(S^s)] + \lambda L_{GP} + \frac{1}{2} (L_{cls}(G_\theta(U)) + L_{cls}(S^s)) \quad (4)$$

where the first two terms calculate the Wasserstein distance of the distribution of real samples and fake samples; the third term  $L_{GP} = \lambda (\|\nabla_{S^s} D_w(\bar{S}^s)\|_2 - 1)^2$  which refers to [31] to improve the Wasserstein GAN training. The last two terms are the classification loss of fake and real samples, respectively.

Finally, we use  $B_b$  to train a part of the Discriminator,  $D_{align}$ , and conduct the domain alignment.  $D_{align}$  includes a FC layer and classifier (the purple area in Fig. 3) and is independent from the optimization of  $D$ . We use  $B_b$  to train  $D_{align}$  and  $G$ ; the loss function is defined in (5).

**Generator G.** We train  $G$  to generate the ‘real’ survey data from the corresponding consumer model information, which hopes to ease the FRT class imbalance in  $E_f$  and helps interpret the consumer consumption behavior. Different to common CGAN, we only take the condition value  $U$  as the input to  $G$ . Considering that we have more unknown samples than known samples (100,468 unknown and 4,492 known

samples) and we will mix Gaussian noises with consumer activity information from  $E_f$  (details in Section IV-C), we need no additional  $z$  noise with condition values. We design  $G$  with a Consumer Consumption Model Embedding and an FC layer activated by the Sigmoid function. The Sigmoid Function will help  $G$  constrain the output scope between 0 and 1, which is similar to the scope of the compressed survey data in Section III-B. Therefore, we have  $S_{L_i}^s \leftarrow G_\theta(U_i)$  ( $i \in E$ ), where  $U_i$ ,  $S_{L_i}^s$  denote the consumer model information of consumer  $i$ , and generated survey information conditioned on the class label  $L_i$ , respectively. Then, we define the loss function of  $G$  and  $D_{align}$  as follows:

$$L_{G,D_{align}} = -\mathbb{E}_U[D_w(G_\theta(U))] + W * L_{cls}(G_\theta(U)) + L_{cls}(S^s)$$

$$s.t. \quad U \sim \{U_i : i \in B_b \cap E_l\}, \quad S^s \sim \{S_i^s : i \in B_b \cap E_f\}$$
(5)

where the first term is the Wasserstein loss in [32], and the other two terms are the weighted class constraint loss. We set  $W = \frac{|\{i:i \in E_f \cap B_b\}|}{|\{i:i \in E_l \cap B_b\}|}$  to prevent back-propagating over-weighted class loss of the generated samples.

### C. Fusion and Optimization

In this section, we will explain how we link consumer consumption model information and survey information in our Asymmetric Cross-Domain Generative Adversarial Network (ADGAN), and the technical details of batch construction methods driven by asymmetrical domain. We provide the pseudocode of our overall training procedure in Algorithm 1. In each epoch, we optimize  $L_D$  in lines 2-10, where we mix the Gaussian Noise with  $U$  to add diversity; we optimize  $L_{G,D_{align}}$  in lines 11-18, where we mix the Gaussian Noise with  $S^s$  to improve diversity. Note, we have three methods of sampling  $B_a$  from  $E_f$ : undersampling, oversampling, and random. The sampling methods will influence the optimization of the discriminator, which will be compared in Section V. Due to the class imbalance in  $E_f$ , it is essential to ensure the balance between the sample spaces of classes. In this section, we will detailed explain how to split data for training from  $B_a, B_b$  (Definition in Section IV-B) and Gaussian Noise fusion with  $U, S^s$ , respectively.

**Batch  $B_a$  Segmentation Method.** We train the model abilities to deal with imbalance classes under random-sampling, oversampling, and undersampling, respectively. When random-sampling, we ignore the FRT class of consumers, and randomly select a batch of consumers from  $E_f$ ; when oversampling, we sort consumers by labels, replicate the consumers until all classes have the same sample size, select a batch of consumers until exhausting all samples, shuffle and repeat the above options; when undersampling, similar to oversampling, but we randomly select samples from classes until a small size which is no higher than the fewest class number. After preparing the batch data for training, we need to improve the sample diversity before feeding batch data to the generator. Since we limit the supervised optimization of discriminator on the consumers from  $E_f$ , the consumer model information of them are stable. To make the discriminator and generator can

apply in a more generalized situation, and simulate the function of noise  $z$ , for each piece of consumer information, we independently sample several Gaussian noises from  $\mathcal{N}(\mu, \sigma)$  and fuse them in random locations.

**Batch  $B_b$  Segmentation Method.**  $B_b$  is designed for providing balanced samples and generating good biopsychosocial explanation for consumers from  $E_l$ . Since we have 100,469 unknown consumers and want to explain their activity, we will directly select some of them to supplement the real data without Gaussian Noise. We proportionally select consumers from  $E_f$  by classes and supplement consumers from  $E_l$  to make each class have the same training samples. Further, the consumer will have the same scores in their features because the questions in the survey only have seven options for choosing. Therefore, we similarly select and mix noise from  $\mathcal{N}(\mu, \sigma)$  with the  $\{S_i^s : i \in B_b \cap E_f\}$  to improve data quality.

---

**Algorithm 1** Training procedure with default setting: oversampling the consumers for training discriminator,  $I_d = 20, \mu = 0, \sigma = 0.01, c = 5, size = 64$

---

**Input:** the max epoch  $Step$ , batch size  $size$ , discriminator step number  $I_d$ , the hyperparameter for  $L_{GP}$ , Adam hyperparameters  $\beta_1, \beta_2$ , Adam learning rate  $lr_D, lr_{GD}$  for  $L_D$  and  $L_{G,D_{align}}$  Gaussian noise hyperparameters  $\mu, \sigma$ , Gaussian noise number  $c$ , sampling way for training discriminator.

```

1: for epoch  $\in [1, Step]$  do
2:   for iter  $\in [1, I_d]$  do
3:     Oversample a batch  $B_a$  of consumers from  $E_f$ 
4:     for  $i \in [1, size]$  do
5:       Sample  $c$  noise from  $\mathcal{N}(\mu, \sigma)$  and mix with  $U_i$ 
6:     end for
7:      $\bar{S}^s \leftarrow G_\theta(U)$ 
8:     Compute the discriminator loss  $L_D$  by 4
9:      $w \leftarrow Adam(\nabla_w, w, lr_D, \beta_1, \beta_2)$ 
10:  end for
11:  Sample a batch  $B_b$  of consumers from both  $E_f$  and
     $E_l$ 
12:   $\bar{S}^s \leftarrow G_\theta(U)$ 
13:  for  $i \in [1, size]$  do
14:    Sample  $c$  noise from  $\mathcal{N}(\mu, \sigma)$  and mix with  $S_i^s$ 
15:  end for
16:  Compute the loss  $L_{G,D_{align}}$  by 5
17:   $w_{D_{align}} \leftarrow Adam(\nabla_{w_{D_{align}}}, w, lr_{GD}, \beta_1, \beta_2)$ 
18:   $\theta \leftarrow Adam(\nabla_\theta, \theta, lr_{GD}, \beta_1, \beta_2)$ 
19: end for

```

---

## V. EXPERIMENTS

### A. Experiment Setting

We first use traditional classification algorithms as baselines: Logistic Regression (LR), Support Vector Machine, XGBoost (XGB) [33]). To form stronger baselines, we further selected three representative models from state-of-the-art cross-domain

TABLE III

COMPARISON PERFORMANCE ON THE MEAN VALUE AND STANDARD DEVIATION (IN PARENTHESES). RADGAN, UADGAN, AND OADGAN DENOTE THE SAMPLING METHODS OF OUR MODEL (RANDOMSSAMPLING, UNDERSAMPLING, AND OVERSAMPLING), RESPECTIVELY.

Model	0-F1-Score	1-F1-Score	2-F1-Score	3-F1-Score	Macro precision	Macro Recall	Macro F1-score	Accuracy
LR	0.01439(0.000)	0.04255(0.000)	0.42032(0.000)	0.31214(0.000)	0.29814(0.000)	0.29492(0.000)	0.19735(0.000)	0.29492(0.000)
SVM	0.00000(0.000)	0.01550(0.000)	0.41196(0.000)	0.25610(0.000)	<b>0.46123(0.000)</b>	0.28516(0.000)	0.17089(0.000)	0.28516(0.000)
XGB	0.02985(0.000)	0.01429(0.000)	0.41941(0.000)	0.31214(0.000)	0.32154(0.000)	0.29492(0.000)	0.19392(0.000)	0.29492(0.000)
PixelDA(U)	0.36000(0.120)	0.02157(0.065)	0.00448(0.013)	0.03728(0.112)	0.08203(0.059)	0.25020(0.001)	0.10583(0.017)	0.25020 (0.001)
SGAN	0.27555(0.083)	<b>0.35137(0.068)</b>	0.30023(0.070)	0.04145(0.124)	0.24669(0.031)	0.29590(0.006)	0.24215(0.010)	0.29590(0.006)
GAZSL	0.29600(0.019)	0.10852(0.047)	0.39152(0.013)	0.44321(0.010)	0.34468(0.014)	0.34727(0.005)	0.30992(0.013)	0.34727(0.005)
PixelDA(S)	0.38900(0.018)	0.18701(0.039)	0.39639(0.039)	0.42766(0.035)	0.35637(0.008)	0.36328(0.004)	0.34992(0.008)	0.36328(0.004)
RADGAN	0.00000(0.000)	0.04957(0.070)	0.44897(0.009)	<b>0.49949(0.016)</b>	0.25040(0.070)	0.34941(0.006)	0.24951(0.018)	0.34941(0.006)
UADGAN	0.34417(0.021)	0.15036(0.045)	0.44413(0.022)	0.48327(0.020)	0.39792(0.015)	0.38444(0.003)	<b>0.35548(0.013)</b>	0.38444(0.003)
OADGAN	<b>0.39388(0.023)</b>	0.03814(0.029)	<b>0.48100(0.015)</b>	0.49723(0.032)	0.39252(0.080)	<b>0.40495(0.013)</b>	0.35256(0.008)	<b>0.40495(0.013)</b>

alignment GAN structures for the comparison: GAZSL [11], SGAN [34], PixelDA [13]. The three models belong to supervised, semi-supervised, and unsupervised models, respectively. We further derive supervised and unsupervised versions of PixelDA, denoted by PixelDA(S) and PixelDA(U). Since these models are not originally designed for our dataset, we adapt them to our problem—we combine the auxiliary user model information in classifier module; the supervised models will only use the pairwise data; the semi-supervised and unsupervised models will utilize data and ignore the pairwise relationship. The parameters of our methods are set as follows:  $Step = 4000$ ,  $lr_D = 1e-4$ ,  $lr_G = 1e-3$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 0.9$ ,  $c = 5$ . We run each model 10 times to obtain mean scores and standard deviations as experimental results.

We evaluate our model by comparing with several baselines and state-of-the-art methods, as well as conducting self-comparison. Table III shows that baseline methods have extremely imbalanced performance in different classes. All three baselines perform poorly in predicting Class 0 and class 1. Although SVM achieves the highest Macro Precision score among all models, the baselines have the lowest Macro F1-score and Accuracy among the supervised methods due to the imbalanced performance. Compared to other GAN models, GAN theory largely eases class imbalance effects.

The unsupervised training method is largely weakened by the class imbalance. The semi-supervised SGAN performs best on Class 1, and worst output in Class 3, which means semi-supervised can better catch the Class 1 characters but will confuse the Class 3 training. The supervised methods, GAZSL and PixelDA(S), obtain the best performance while they only use the pairwise data. The supervised methods have balanced performance in all classes. Comparison of our model with the supervised methods shows the non-pairwise data can provide extra information.

Among the sampling methods, randomsampling has the best performance in Class 3 but cannot distinguish Class 1 at all. Both undersampling and oversampling achieve better performance of our model than the other models. While oversampling can extract the most information from the raw data, undersampling is more stable, demonstrated by a smaller standard deviation.

TABLE IV

ABLATION STUDY ON MEAN (STANDARD DEVIATION) RESULTS.

Model	Macro precision	Macro Recall	Macro F1-score	Accuracy
SFC	0.24186(0.029)	0.28086(0.003)	0.24343(0.014)	0.28086(0.003)
CMFC	0.32495(0.065)	0.21953(0.001)	0.18691(0.003)	0.21953(0.001)
CFC	0.27423(0.007)	0.28184(0.005)	0.27070(0.006)	0.28184(0.005)
SV	0.32603(0.018)	0.32871(0.002)	<b>0.27701(0.007)</b>	0.32871(0.002)
CMV	0.28559(0.074)	0.25820(0.004)	0.15523(0.032)	0.25820(0.004)
CV	<b>0.34113(0.031)</b>	<b>0.33281(0.002)</b>	0.26276(0.013)	<b>0.33281(0.002)</b>

### B. Ablation Study

We conduct ablation studies to evaluate the influence of View Structure Embedding and hyper-parameter settings.

**View Structure Embedding.** To show the effectiveness of our Structural View Embedding and GAN, we set the comparison with the simple FC layer without view structure and the embedding structures in Fig. 3: Survey FC layer (SFC), Consumer Consumption Model FC layer (CMFC), Combined FC layer (CFC); and View Structured Survey Embedding (SV), View Structured Consumer Consumption Model Embedding (CMV), and View Structured Combined Embedding (CV), where the combined analysis will concatenate the two domain data before feeding the output layer of the model. As shown in Table IV, all structural view embedding models outperform the corresponding FC layer in accuracy. We can clearly see the view structure improves the Precision and F1-Score in survey data analysis and combined data analysis, which show the effectiveness of the structural view embedding. The results also demonstrate that using combined data result in better overall representation better than using data from two domains independently.

**Hyper-parameters.** We analyze how the Gaussian noise number  $c$ , and the learning rates  $lr_D$ ,  $lr_G$  influence the model performance, and prove the effectiveness of the Gaussian noise. We set undersampling,  $Step = 1000$ ,  $lr_D = 1e-3$ ,  $lr_G = 1e-3$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 0.9$ ,  $c = 5$  as defaulted values, and change the corresponding value when test the target parameter. We test the model for 5 times and conclude the mean performance. Fig. 4(a) and (b) plot the mean accuracy of models under different  $c$ , and  $lr_D$ ,  $lr_G$ , and we can see



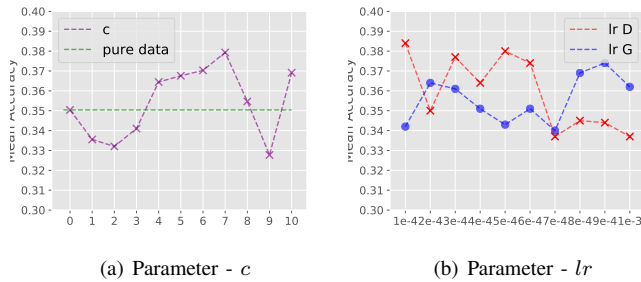


Fig. 4. (a) plots the mean accuracy of different Gaussian noise numbers in purple and the pure data baseline in green dashed lines; (b) plots the mean accuracy of the learning rates of Discriminator and Generator in blue and red lines.

that the model obtains the best performance when  $c = 7$ ,  $lr_D = 1e - 4$ ,  $lr_G = 9e - 4$ , respectively. The model will perform better with most Gaussian noise settings than the pure data; therefore, proper Gaussian Noise fusion can effectively improve the model performance. The mean accuracy and mean standard deviation of different learning rates for  $D$  and  $G$  are 0.35921 (0.017) and 0.35564 (0.011), respectively. The model is relatively stable under different learning rates, and we can conclude that the model will perform better if we set a relatively small  $lr_D$  to  $lr_G$ .

## VI. CONCLUSION

In this paper, we construct domain-specific consumer models to describe consumers' consumption structures and intent. We further propose the Asymmetric Cross-Domain Generative Adversarial Network to extract information from unequal domains. our network outperforms state-of-the-art methods, and our ablation study reveals the influence of learning rates, as well as the effectiveness of structural view embedding and Gaussian Noise.

## REFERENCES

- [1] A. Haddad, "E-finance, entrepreneurship and economic growth in developing countries," 2018.
- [2] M. Kannadhasan, S. Aramvalarthan, S. Mitra, and V. Goyal, "Relationship between biopsychosocial factors and financial risk tolerance: An empirical study," *Vikalpa*, vol. 41, no. 2, pp. 117–131, 2016.
- [3] L. Nguyen, G. Gallery, and C. Newton, "The joint influence of financial risk perception and risk tolerance on individual investment decision-making," *Accounting & Finance*, vol. 59, pp. 747–771, 2019.
- [4] Z. N. Onsomu, E. Kaijage, J. Aduda, and C. Iraya, "Risk tolerance, demographics and portfolio performance," *Journal of Business and Economic Policy*, vol. 4, no. 3, pp. 69–74, 2017.
- [5] W. C. Martin and A. Davari, "Examining financial risk tolerance via mental accounting and the behavioral life-cycle hypothesis," *Academy of Marketing Studies Journal*, vol. 22, no. 4, pp. 1–13, 2018.
- [6] J. Qiu, Z. Lin, and Y. Li, "Predicting customer purchase behavior in the e-commerce context," *Electronic commerce research*, vol. 15, no. 4, pp. 427–452, 2015.
- [7] F. Kooti, K. Lerman, L. M. Aiello, M. Grbovic, N. Djuric, and V. Radosavljevic, "Portrait of an online shopper: Understanding and predicting consumer behavior," in *Proceedings of the ninth ACM international conference on web search and data mining*. ACM, 2016, pp. 205–214.
- [8] N. Su, J. He, Y. Liu, M. Zhang, and S. Ma, "User intent, behaviour, and perceived satisfaction in product search," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 2018, pp. 547–555.

- [9] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [10] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.
- [11] Y. Zhu, M. Elhoseiny, B. Liu, X. Peng, and A. Elgammal, "A generative adversarial approach for zero-shot learning from noisy texts," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1004–1013.
- [12] A. Farshchian, J. A. Gallego, J. P. Cohen, Y. Bengio, L. E. Miller, and S. A. Solla, "Adversarial domain adaptation for stable brain-machine interfaces," *arXiv preprint arXiv:1810.00045*, 2018.
- [13] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3722–3731.
- [14] N. Tang and A. Baker, "Self-esteem, financial knowledge and financial behavior," *Journal of Economic Psychology*, vol. 54, pp. 164–176, 2016.
- [15] D. R. Lichtenstein, M. E. Drumwright, and B. M. Braig, "The effect of corporate social responsibility on customer donations to corporate-supported nonprofits," *Journal of marketing*, vol. 68, no. 4, pp. 16–32, 2004.
- [16] L. A. Mohr and D. J. Webb, "The effects of corporate social responsibility and price on consumer responses," *Journal of consumer affairs*, vol. 39, no. 1, pp. 121–147, 2005.
- [17] P. Hartmann and V. Apaolaza-Ibañez, "Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern," *Journal of business Research*, vol. 65, no. 9, pp. 1254–1263, 2012.
- [18] H. C. Triandis, *Individualism and collectivism*. Routledge, 2018.
- [19] G. Roehrich, "Consumer innovativeness: Concepts and measurements," *Journal of business research*, vol. 57, no. 6, pp. 671–677, 2004.
- [20] T. D. Buskirk, A. Kirchner, A. Eck, and C. S. Signorino, "An introduction to machine learning methods for survey researchers," *Survey Practice*, vol. 11, no. 1, p. 2718, 2018.
- [21] X. Huang, J. Li, and X. Hu, "Label informed attributed network embedding," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017, pp. 731–739.
- [22] A. Ghazimatin, R. Saha Roy, and G. Weikum, "Fairy: A framework for understanding relationships between users' actions and their social feeds," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 2019, pp. 240–248.
- [23] J. Joo, F. F. Steen, and S.-C. Zhu, "Automated facial trait judgment and election outcome prediction: Social dimensions of face," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3712–3720.
- [24] A. E. Abele and B. Wojciszke, "Agency and communion from the perspective of self versus others," *Journal of personality and social psychology*, vol. 93, no. 5, p. 751, 2007.
- [25] T. Jisana, "Consumer behaviour models: an overview," *Sai Om Journal of Commerce & Management*, vol. 1, no. 5, pp. 34–43, 2014.
- [26] S. McLeod, "Maslow's hierarchy of needs," *Simply Psychology*, vol. 1, 2007.
- [27] A. S. Boote, "Psychographics: mind over matter," *American Demographics*, vol. 2, no. 4, pp. 26–29, 1980.
- [28] K.-Y. Chan and F. Drasgow, "Toward a theory of individual differences and leadership: understanding the motivation to lead," *Journal of applied psychology*, vol. 86, no. 3, p. 481, 2001.
- [29] C. A. Mandrik and Y. Bao, "Exploring the concept and measurement of general risk aversion," *ACR North American Advances*, 2005.
- [30] J. T. Plummer, "The concept and application of life style segmentation," *the Journal of Marketing*, pp. 33–37, 1974.
- [31] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, 2017, pp. 5767–5777.
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.
- [33] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.
- [34] A. Odena, "Semi-supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016.