Learning Invariant Representations for New Product Sales Forecasting via Multi-Granularity Adversarial Learning

Zhenzhen Chu
East China Normal University
Shanghai, China
51215903091@stu.ecnu.edu.cn

Chengyu Wang
Alibaba Group
Hangzhou, China
chengyu.wcy@alibaba-inc.com

Cen Chen*
East China Normal University
Shanghai, China
cenchen@dase.ecnu.edu.cn

Dawei Cheng
Tongji University
Shanghai, China
dcheng@tongji.edu.cn

Yuqi Liang
Seek Data Group, Emoney Inc.
Shanghai, China
roly.liang@seek-data.com

Weining Qian
East China Normal University
Shanghai, China
wnqian@dase.ecnu.edu.cn

ABSTRACT

Sales forecasting during the launch of new products has always been a challenging task, due to the lack of historical sales data. The dynamic market environment and consumer preferences also increase the uncertainty of predictions. Large chains face even greater difficulties due to their extensive presence across various regions. Traditional time-series forecasting methods usually rely on statistical models and empirical judgments, which are difficult to handle large, variable data and often fail to achieve satisfactory performance for new products. In this paper, we propose a Multi-granularity AdversaRial Learning framework (MARL) to leverage knowledge from old products and improve the quality of invariant representations for more accurate sales predictions. To evaluate our proposed method, we conducted extensive experiments on both a real-world dataset from a prominent international Café chain and a public dataset. The results demonstrated that our method is more effective than the existing state-of-the-art baselines for new product sales forecasting.

CCS CONCEPTS

• **Applied computing** → **Forecasting**.

KEYWORDS

new product sales forecasting, transfer learning, adversarial learning

ACM Reference Format:
Zhenzhen Chu, Chengyu Wang, Cen Chen, Dawei Cheng, Yuqi Liang, and Weining Qian. 2023. Learning Invariant Representations for New Product Sales Forecasting via Multi-Granularity Adversarial Learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615219

*Corresponding author.

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615219

1 INTRODUCTION

Forecasting sales for newly launched products has always been a difficult task, primarily due to the absence of historical sales data, together with the dynamic nature of the market environment and consumer preferences. Often, the accuracy of the sales forecasting and the company's performance are highly correlated [3]. When the forecast deviates significantly, it can lead to the problem of out-of-stock or over-stocking, which can have a severe impact on the company's operations [19]. Although forecasting sales of new products presents challenges for companies, improving the accuracy of such predictions can facilitate better risk management and inventory control for new products at each store [12].

Sales forecasting for new products presents a fundamental challenge due to the lack of historical sales data. As a result, forecasting the future sales of new products often requires analyzing historical sales data of old products [12]. Nevertheless, significant differences in sales patterns between old and new products can make it difficult to predict the sales of new products accurately. In addition, some large retail chains have many stores and sales may vary considerably, which further complicates sales forecasting across stores.

Prior research [2, 23] has attempted to address the *data scarcity* issue for new products by identifying old products that closely resemble the new product. However, measuring product similarity can be difficult, and a sufficiently similar old product may not exist. Some works [6, 22, 26] utilize data from all old products to train models and enhance sales forecasting by fusing multi-modal information, such as images and texts. However, this type of approach can be influenced by variations in sales patterns between products. Recently, few works [13] address this issue by transfer learning; yet still do not consider the different sales patterns between products. In addition, the issue of *sales variation between stores* has not received sufficient attention for new product sales forecasting. Most existing methods [6, 7, 22, 26] focus on individual stores or groups of stores, each sharing a model. However, they may prove ineffective when there are either too many groups, resulting in a cumbersome number of models, or when the number of groups is small, allowing for significant discrepancies between stores within the same group.

In this paper, we propose a Multi-granularity AdversaRial Learning (MARL) framework that addresses the issues of data scarcity for new products and sales variation between both products and stores. MARL leverages multi-modal information to transfer knowledge from historical products across different stores and improves the quality of invariant temporal representations for more accurate predictions of new product sales. Despite variations in product and store sales, inherent transferable patterns can be captured more effectively by aligning the feature distributions at different granularities, including between both old and new products, as well as across different stores. Extensive experiments have been conducted on a real-world dataset from a prominent international Café chain and a public dataset. Results demonstrate that MARL is superior to current state-of-the-art baselines in new product sales forecasting.

2 RELATED WORK

New Product Sales Forecasting. Sales forecasting for new products without historical data is a challenge for many companies [12]. Some works [2, 23] have studied methods based on finding similar products. These methods require a distance measure to compute the similarities between two products. However, not all product features are suitable for the metric, such as product images. Moreover, the weight of each product feature is not necessarily the same. As it is difficult to determine the appropriate weights, similar products predicted by these methods may not be useful for sales forecasting. Singh et al. [21] directly leverage the history of old product sales to train prediction models. Vashishtha et al. [26] add the product age to input features to enhance model performance. Other works [6, 22] utilize data from other modals to enhance the prediction by multi-modal fusion methods. Few works [13] focus on transfer learning; yet they do not consider variations in product and store sales.

Domain Adversarial Transfer Learning. Inspired by GAN [9], DANN [8] learns domain-adversarial representations by a min-max optimization method. Discriminator in DANN is used to judge whether features are from the source or the target domain, and feature extractor is used to confuse discriminator. There are many works based on DANN, most of which are for classification tasks , such as [4, 25, 27, 29, 30]. There are also some works based on regression tasks, such as [10, 11] for two domains. Yet, there often exist multiple domains in new product sales forecasting, making these methods difficult to perform well. The method [17] directly confuses multiple domains by making discriminator and feature extractor have exactly the opposite optimization goals. The work [15] treats all domains as the same and then aligns them with a pre-defined distribution. Some works [1, 20] address the multiple domain problem by creating additional discriminators and feature extractors.

3 METHODOLOGY

3.1 Problem Setup

Our proposed MARL aims to predict the sales of new products, which is trained with the historical sales data of old products, product attributes, store attributes and other external information. Denote the number of old products as K_s. Old products are treated as source domains, represented as $D_s = \{(x_s^i, y_s^i, d_s^i)\}_{i=1}^{N_s}$, where N_s and d_s^i represent the total number of data instances of all source domains and the domain label to which the corresponding data instance x_s^i belongs, respectively. x_s^i includes various types of inputs such as product image, product text attributes, store attributes and external temporal information. As sales forecasting is usually placed in a time series setting, both external information in x_s^i and $y_s^i = \{y_s^{i,t}\}_{t=1}^{N_s^i}$ are time series with the length of N_s^i, where $y_s^{i,t}$ is the sale on the day t. New products are treated as target domains and their data structures are identical to those of old products, except for the absence of historical sales data. During training, sales and domain labels are the outputs. Only predicted sales are treated as output during inference.

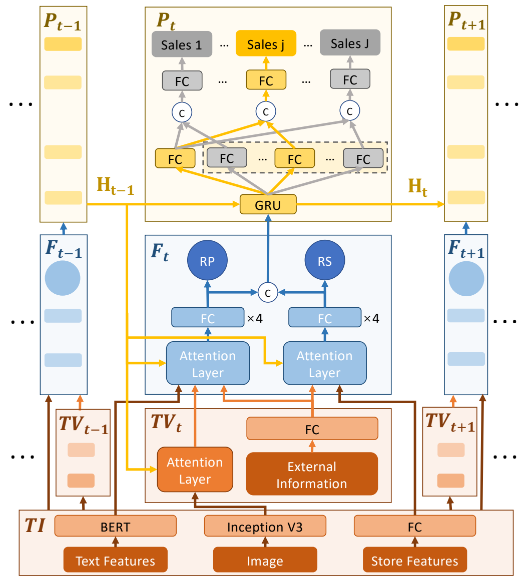


Figure 1: The model architecture of MARL at the t-th time point. TI is the time-invariant raw feature extractor (for encoding product images and texts, and store features). TV is the time-variant feature extractor (for encoding external temporal information) and applying soft attention [28] to image features. F is the domain-invariant feature learner, which incorporates two regularizers: RP (product-level feature invariant regularizer) and RS (store-level feature invariant regularizer). P is the final sales predictor. In the figure, "FC" refers to the fully connected layer.

3.2 Raw Feature Extraction

As shown in Figure 1, we use TI and TV to process and extract the raw features. TI is the time-invariant raw feature extractor, responsible for encoding the time-invariant raw features. For product texts, the pre-trained BERT [5] is leveraged to transform them into feature vectors. For product images, we use Inception V3 [24] to extract the features. TV is the time-variant feature extractor, responsible for encoding time-variant raw features and applying soft attention [28] to image features output by TI to obtain the weighted sum of each channel.

3.3 Multi-Granularity Domain-Invariant Representation Learning

After raw features are extracted by TI and TV, we employ soft attention [28] to fuse image features, text features and external features, which are fed into the left part of F (for product-related feature learning). Meanwhile, store features and external features are fused and fed into the right part of F (for store-related feature learning). Here, F is the domain-invariant feature learner, which is able to learn domain-invariant representations on both products and stores. Below we will introduce the techniques in detail.

Sales Forecasting with Adversarial Learning. We leverage domain adversarial learning to achieve product knowledge transfer and learn invariant representations for sales forecasting. This process is similar to DANN [8], which allows feature learner F and domain discriminator D to have the opposite optimization goals. Previously, DANN solves the problem with only one source and one target domain, hence its domain discriminator is a binary classifier. In MARL, we incorporate RP (the product-level feature invariant regularizer) into F (i.e., a feed forward neural network consisting of four fully connected layers with ReLU activations). The domain adversarial loss of RP is as follows:

$$L_{RP} = -\frac{1}{\sum_{j=1}^{N_s} N_s^j} \sum_{i=1}^{N_s} \sum_{t=1}^{N_s^i} \sum_{k=1}^{K_s} 1_{[k=d_s^i]} \log p_i^{t,k}, \tag{1}$$

where $p_i^{t,k}$ represents the probability that the domain label at the time point t of the i-th time series in source domains is predicted to be k. The loss function L_{F_advp} of F with respect to the domain adversarial loss is $-L_{RP}$. With RP, the model F should be confused about the domain labels of different old products, and thus is able to extract the domain-invariant features of each product.

Improved Domain Adversarial Learning. The large number of domains (i.e, old products) may negatively affect the performance of knowledge transfer with the vanilla domain adversarial loss. This observation is also similar to [20]. Here, we improve the loss L_{F_advp} without additional feature extractors and discriminators. The optimization goal of L_{F_advp} is to smooth the output probabilities of samples belonging to each domain. That is, to make the K_s probabilities output by the RP regularization output head as close to $\frac{1}{K_s}$ as possible. Thus, the loss L_{F_advp} is re-written as follows:

$$L_{F\_advp} = -\frac{1}{K_s \sum_{j=1}^{N_s} N_s^j} \sum_{i=1}^{N_s} \sum_{t=1}^{N_s^i} \sum_{k=1}^{K_s} |p_i^{t,k} - \frac{1}{K_s}|. \tag{2}$$

Minimizing L_{F_advp} enables the feature learner F to have the ability to learn domain-invariant features across multiple domains.

Multi-Granularity Knowledge Transfer. In the literature, many works [6, 7, 22, 26] for sales forecasting often ignore the store information or focus on training one model for a group of very similar stores. We observe that although there are some differences between stores, there are actually things they can learn from each other, such as the fact that sales are likely to drop when they encounter unusual weather conditions. Thus, knowledge can be transferred across stores, which is similar to the case of products. As shown in Figure 1, we further add RS (store-level feature invariant regularizer) to F to incorporate multi-granularity knowledge transfer. The loss function L_{RS} of RS has the same form as L_{RP}; and the loss function L_{F_advs} also has the same form as L_{F_advp}. At this time, the domain-invariant representations on both products and stores can be obtained. Next, we concatenate the two types of features and input them into P to forecast the sales.

3.4 Selective Feature Sharing by Store Groups

For sales forecasting, as the size of the training set for each store is relatively small, in order to avoid the data hungry problem, we naturally divide all the stores into J groups according to the store type (store type is related to location, such as residential area and transportation hub). Each store group has its own output prediction head. As shown in Figure 1, in P, there are two branches after the GRU block. The left FC represents product-related features, which is shared by all the data samples. The right FC considers the store group granularity, where each store group has a separate FC. For an input instance, based on its store group, we selectively concatenate the outputs of the shared product-based and its own store group-based FC blocks. After that, the corresponding FC (i.e., the final output layer) is employed to forecast the sales. The Mean Squared Error (MSE) loss function of P is as follows:

$$L_{mse} = \frac{1}{\sum_{j=1}^{N_s} N_s^j} \sum_{i=1}^{N_s} \sum_{t=1}^{N_s^i} (y_s^{i,t} - \hat{y}_s^{i,t})^2, \tag{3}$$

where $\hat{y}_s^{i,t}$ represents the sales forecast value at the time point t of the i-th time series in source domains.

3.5 Overall Loss Function

The loss functions of P, RP and RS are L_{mse}, L_{RP} and L_{RS}, respectively. The overall loss function of MARL is as follows:

$$L_F = L_{mse} + \lambda_1 L_{F\_advp} + \lambda_2 L_{F\_advs}, \tag{4}$$

where the hyper-parameters λ_1 and λ_2 control the trade-off between product sales forecasting and adversarial domain discrimination.

4 EXPERIMENTS

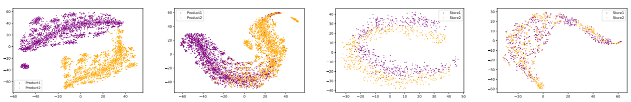
4.1 Setup

Datasets. We use two datasets: **(1) Café**. This is a real-world, in-house dataset from an prominent international Café chain and is collected from 2018 to 2021. It includes product images and texts, as well as some external temporal information and store attributes. We select two categories of products, i.e., Cake and Sandwich, to conduct experiments separately. The number of products in both categories is both around thirty. In the dataset, there are 110 stores for experiments, divided into five types. **(2) VISUELLE** [22]. This is a public dataset from a fast fashion company named Nunalie during 2016 to 2019. It includes 5577 clothing products. Each product is associated with image and text information, sales in 12 weeks after launch, some external information and three Google Trends data describing the category, color and fabric popularity.

Baselines. We compare our method with the following baselines: **(1) Classical deep learning models**. We select MLP and LSTM as baselines, and use the architectures in [21]. Because MARL uses GRU, we also take it as the baseline with the similar architecture to LSTM. **(2) Attention based Multi-modal RNNs** [6]. The Concat Multi-modal RNN, Residual Multi-modal RNN and Cross-Attention RNN are included. They use the encoder-decoder structure. The encoder is responsible for fusing multi-modal features with attention, and the decoder is based on RNN to forecast sales. **(3) GTM-Transformer** [22]. This is the state-of-the-art transformer-based model based on the fusion of multi-modal features.

Table 1: Comparison of results over VISUELLE and Café.

Method	Café				VISUELLE	
	Cake		Sandwich			
	wMAPE	MAE	wMAPE	MAE	wMAPE	MAE
MLP	73.22	2.58	95.62	2.77	70.55	31.70
LSTM	69.24	2.44	95.59	2.77	68.89	30.95
GRU	70.40	2.48	94.13	2.73	69.50	31.23
Concat Multi-modal RNN	69.95	2.46	89.20	2.59	68.68	30.86
Residual Multi-modal RNN	68.09	2.40	89.32	2.59	68.68	30.86
Cross-Attention RNN	67.73	2.39	89.30	2.59	65.93	29.62
GTM-Transformer	69.40	2.44	90.27	2.61	61.11	27.46
MARL	**63.59**	**2.24**	**82.83**	**2.40**	**60.62**	**27.24**



(a) (b) (c) (d)

Figure 2: Visualization of latent representations of products and stores. (a) and (b) are the results of the products, and (c) and (d) are the results of stores. (b) and (d) leverage feature invariant regularizers, while (a) and (c) do not.

Metrics. We use mean absolute error (MAE) and weighted mean absolute percentage error (wMAPE) [16] as evaluation metrics.
Implementation Details. For Café, we split the dataset into training/validation/testing sets in the ratio of 7:1:2 for Cake and Sandwich respectively. And we add Google Trends data for it. For VISUELLE, because the original dataset is not split according to the chronological order of product launch, we re-split it. The number of products in training, validation, and testing sets are 4055, 178, and 442, respectively. Another important point is that there are no store attributes in VISUELLE, so we replace such features in Café with the external information of the fashion season. In VISUELLE, there are over 4,000 products, which makes it obviously unreasonable to transfer knowledge by the product granularity. So we treat the product category as the domain. As there are no store attributes suitable for applying separate output prediction headers, we do not apply it of MARL in the implementation. Categorical features such as color and fabric are processed with one-hot encoding. During training, we employ the Adam optimizer [14] to optimize MARL, and the feature invariant regularizers and the rest of the model will be alternately optimized. We set both λ_1 and λ_2 in Eq.(4) to 0.1.

4.2 General Forecasting Performance

The experimental results are shown in Table 1. MLP has the worst performance. LSTM and GRU have similar performance, but their overall effect is lower than that of the attention-based multi-modal RNNs, because the use of attention to fuse multi-modal features has a better effect. Among the three attention-based RNNs, Cross-Attention RNN has the best overall performance, which is consistent with [6]. On VISUELLE, GTM-Transformer is better, but it does not perform well on Café. This may be because the quality of Google Trends data in Café is relatively low. MARL achieves the best results over all datasets, especially over Sandwich of Café. Its wMAPE is 6.37% lower than the second-ranking model.

Table 2: Ablation study over VISUELLE and Café. FIR, SOPH and FS represent feature invariant regularizers, separate output prediction headers for each store group and fashion season, respectively.

Method			Café				VISUELLE	
FIR		SOPH	Cake		Sandwich			
Product	Store or FS		wMAPE	MAE	wMAPE	MAE	wMAPE	MAE
✗	✗	✗	69.78	2.46	88.91	2.58	62.38	28.03
✓	✗	✗	67.85	2.39	86.80	2.52	61.00	27.41
✗	✓	✗	68.50	2.41	87.47	2.54	61.59	27.67
✓	✓	✗	67.25	2.37	86.06	2.50	**60.62**	**27.24**
✓	✓	✓	**63.59**	**2.24**	**82.83**	**2.40**	-	-

Table 3: Analysis of domain adversarial loss over Sandwich.

Adversarial Setting	wMAPE	MAE	MMD	
			Product	Store
None	88.91	2.58	0.69	0.34
CE	87.07	2.53	0.41	0.27
FIR (Ours)	**86.06**	**2.50**	**0.36**	**0.06**

4.3 Detailed Model Analysis

Ablation Study. The experimental results are shown in Table 2. We can see that FIR are helpful and are more effective on Café than VISUELLE. This is because VISUELLE has much many products, each with very little data, and it does not have store attributes. We can only leverage the product category and fashion season as the basis for domain division rather than product and store. In addition, the prediction effects have a large improvement after with SOPH.
Visualization. To obtain an intuitive understanding of feature invariant regularizers on the feature extractor F, we use t-SNE to reduce the dimension of the latent representations. The visualization of the training sets of Sandwich is shown in Figure 2. For illustration, we randomly select two domains for product and store parts. It can be seen that the overlap between the two domains is higher after using feature invariant regularizers, which indicates that they can achieve a better feature alignment effect.
Analysis of Domain Adversarial Loss. We compare the effect of our domain adversarial loss with that of [17]. The results over Sandwich without SOPH are shown in Table 3. Here, CE denotes using the domain adversarial loss in [17] to replace the one in Eq. (2). In addition, we use the maximum mean discrepancy (MMD) [18] to measure the feature alignment ability. For products, our FIR method is slightly better than CE, while for stores, our method is much better than CE. This shows the superiority of our method.

5 CONCLUSION

In this paper, we propose a multi-granularity adversarial learning framework for new product sales forecasting. It effectively learns invariant representations for more accurate predictions of new product sales. Experiments on both real-world and public datasets show its effectiveness, outperforming existing state-of-the-art baselines.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under grant number 62202170 and SeekData Inc.

REFERENCES

[1] Isabela Albuquerque, João Monteiro, Mohammad Darvishi, Tiago H Falk, and Ioannis Mitliagkas. 2020. Adversarial target-invariant representation learning for domain generalization. (2020).

[2] Lennart Baardman, Igor Levin, Georgia Perakis, and Divya Singhvi. 2017. Leveraging comparables for new product sales forecasting. *Available at SSRN 3086237* (2017).

[3] Samaneh Beheshti-Kashi, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen, and Michael Teucke. 2015. A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering* 3, 1 (2015), 154–161.

[4] Minghao Chen, Shuai Zhao, Haifeng Liu, and Deng Cai. 2020. Adversarial-learned loss for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 3521–3528.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[6] Vijay Ekambaram, Kushagra Manglik, Sumanta Mukherjee, Surya Shravan Kumar Sajja, Satyam Dwivedi, and Vikas Raykar. 2020. Attention based multi-modal new product sales time-series forecasting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3110–3118.

[7] Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. 2017. Building optimized and hyperlocal product assortments: A nonparametric choice approach. *Available at SSRN 2905381* (2017).

[8] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The journal of machine learning research* 17, 1 (2016), 2096–2030.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[10] Hailin Hu, MingJian Tang, and Chengcheng Bai. 2020. Datsing: Data augmented time series forecasting with adversarial domain adaptation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2061–2064.

[11] Ye Huang, Wei Huang, Shiwei Tong, Zhenya Huang, Qi Liu, Enhong Chen, Jianhui Ma, Liang Wan, and Shijin Wang. 2021. STAN: Adversarial Network for Cross-domain Question Difficulty Prediction. In *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 220–229.

[12] Kenneth B Kahn. 2014. Solving the problems of new product forecasting. *Business Horizons* 57, 5 (2014), 607–615.

[13] Tristan Karb, Niklas Kühl, Robin Hirt, and Varvara Glivici-Cotruta. 2020. A network-based transfer learning approach to improve sales forecasting of new products. *arXiv preprint arXiv:2005.06978* (2020).

[14] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[15] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. 2018. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5400–5409.

[16] Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman. 2008. *Forecasting methods and applications*. John wiley & sons.

[17] Toshihiko Matsuura and Tatsuya Harada. 2020. Domain generalization using a mixture of multiple latent domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 11749–11756.

[18] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. 2010. Domain adaptation via transfer component analysis. *IEEE transactions on neural networks* 22, 2 (2010), 199–210.

[19] Elizabeth Paton. 2018. H&M, a fashion giant, has a problem: $4.3 billion in unsold clothes. *The New York Times* (2018).

[20] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. 2019. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10023–10031.

[21] Pawan Kumar Singh, Yadunath Gupta, Nilpa Jha, and Aruna Rajan. 2019. Fashion retail: Forecasting demand for new items. *arXiv preprint arXiv:1907.01960* (2019).

[22] Geri Skenderi, Christian Joppi, Matteo Denitto, and Marco Cristani. 2021. Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends. *arXiv preprint arXiv:2109.09824* (2021).

[23] RM Steenbergen. 2019. *New Product Forecasting with Analogous Products: Applying Random Forest and Quantile Regression Forest to forecasting and inventory management*. Master's thesis. University of Twente.

[24] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[25] Hui Tang and Kui Jia. 2020. Discriminative adversarial domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5940–5947.

[26] Rajesh Kumar Vashishtha, Vibhati Burman, Rajan Kumar, Srividhya Sethuraman, Abhinaya R Sekar, and Sharadha Ramanan. 2020. Product age based demand forecast model for fashion retail. *arXiv preprint arXiv:2007.05278* (2020).

[27] Shanshan Wang and Lei Zhang. 2020. Self-adaptive re-weighted adversarial domain adaptation. *arXiv preprint arXiv:2006.00223* (2020).

[28] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*. PMLR, 2048–2057.

[29] Chaohui Yu, Jindong Wang, Yiqiang Chen, and Meiyu Huang. 2019. Transfer learning with dynamic adversarial adaptation network. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 778–786.

[30] Yabin Zhang, Hui Tang, Kui Jia, and Mingkui Tan. 2019. Domain-symmetric networks for adversarial domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5031–5040.