

Hierarchically Modeling Micro and Macro Behaviors via Multi-Task Learning for Conversion Rate Prediction

Hong Wen*

Alibaba Group

Hangzhou, Zhejiang, China 311121

qinggan.wh@alibaba-inc.com

Jing Zhang*

The University of Sydney

Darlington NSW 2008, Australia

jing.zhang1@sydney.edu.au

Fuyu Lv

Alibaba Group

Hangzhou, Zhejiang, China 311121

fuyu.lfy@alibaba-inc.com

Wentian Bao

Alibaba Group

Hangzhou, Zhejiang, China 311121

wentian.bwt@alibaba-inc.com

Tianyi Wang

Alibaba Group

Hangzhou, Zhejiang, China 311121

joshua.wty@alibaba-inc.com

Zulong Chen

Alibaba Group

Hangzhou, Zhejiang, China 311121

zulong.czl@alibaba-inc.com

ABSTRACT

Conversion Rate (CVR) prediction in modern industrial e-commerce platforms is becoming increasingly important, which directly contributes to the final revenue. In order to address the well-known sample selection bias (SSB) and data sparsity (DS) issues encountered during CVR modeling, the abundant labeled macro behaviors (*i.e.*, user's interactions with items) are used. Nonetheless, we observe that several purchase-related micro behaviors (*i.e.*, user's interactions with specific components on the item detail page) can supplement fine-grained cues for CVR prediction. Motivated by this observation, we propose a novel CVR prediction method by Hierarchically Modeling both Micro and Macro behaviors (HM^3). Specifically, we first construct a complete user sequential behavior graph to hierarchically represent micro behaviors and macro behaviors as one-hop and two-hop post-click nodes. Then, we embody HM^3 as a multi-head deep neural network, which predicts six probability variables corresponding to explicit sub-paths in the graph. They are further combined into the prediction targets of four auxiliary tasks as well as the final CVR according to the conditional probability rule defined on the graph. By employing multi-task learning and leveraging the abundant supervisory labels from micro and macro behaviors, HM^3 can be trained end-to-end and address the SSB and DS issues. Extensive experiments on both offline and online settings demonstrate the superiority of the proposed HM^3 over representative state-of-the-art methods.

CCS CONCEPTS

• Information systems → Recommender systems.

KEYWORDS

Conversion Rate Prediction, Multi-Task Learning, Deep Learning

*Co-first authors contributed equally to this paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8037-9/21/07...\$15.00

<https://doi.org/10.1145/3404835.3463053>

ACM Reference Format:

Hong Wen, Jing Zhang, Fuyu Lv, Wentian Bao, Tianyi Wang, and Zulong Chen. 2021. Hierarchically Modeling Micro and Macro Behaviors via Multi-Task Learning for Conversion Rate Prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, July 11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3404835.3463053>

1 INTRODUCTION

E-commerce Recommender Systems (RS) [4, 9, 15, 16, 23] serve a vital role to help users discover their preferred items from a vast number of candidates, therefore improving the user experience [19] and delivering new business value [4, 12, 22]. A typical RS usually consists of two significant phases, *i.e.*, retrieval and ranking, where Click-Through Rate (CTR) [3, 6, 14, 24, 25] and Conversion Rate (CVR) [11, 13, 17, 18, 20] predictions in the latter phase are two fundamental tasks, measuring the probability of an item from impression to click and from click to purchase, respectively. In this paper, we focus on the post-click CVR estimation task.

In general, conventional CVR modeling methods always employ similar techniques developed for CTR tasks, where the only difference is replacing the training samples generated from the path “impression→click” to “click→purchase” in the user sequential behavior graph [13, 18]. Although it is simple and effective for online development, two critical issues encountered in practice, *i.e.*, sample selection bias (SSB) [21] and data sparsity (DS) [10]. SSB refers to the issue that the training and inference sample space of CVR prediction can be systematically different, and DS refers to the fact the number of training samples for CVR prediction is much less than that for CTR prediction. These two issues affect industrial-level recommender systems. Furthermore, the deficiency of positive samples also increases the difficulty of CVR modeling, *e.g.*, only less than 0.1% of impressions will finally convert to purchases, according to the statistics from our e-commerce platform.

To address these challenges, researchers try to model CVR prediction and learn strong feature representations [1, 8] over the entire space with all impression samples. For example, ESMM [13] models the “impression→click→purchase” path for the CVR task, and trains the CVR model over the entire space. In this way, ESMM can effectively mitigate the SSB and DS issues and achieves better performance than several conventional CVR modeling methods. However, it still faces the challenge that the positive samples of

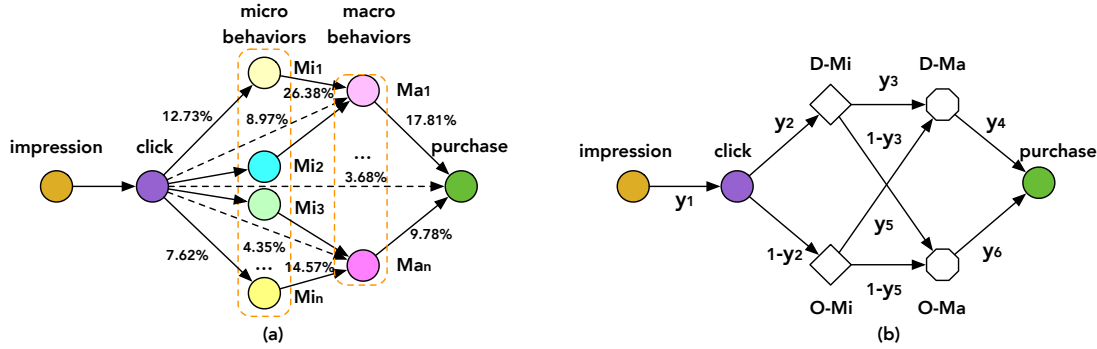


Figure 1: Illustration of the motivation of our method. (a) Hierarchically inserting micro and macro behaviors between click and purchase. Here, $M_{i_1}, M_{i_2}, \dots, M_{i_n}$ represent the micro behaviors while M_{a_1}, \dots, M_{a_n} represent the macro behaviors (See Table 1). The number above each edge represents the sparsity of respective path. (b) Dividing micro (macro) behaviors into two disjoint sets namely $D-Mi$ and $O-Mi$ ($D-Ma$ and $O-Ma$) results in a novel user sequential behavior graph.

CVR prediction are insufficient. In fact, users always take many purchase-related behaviors after clicking an item, *e.g.*, putting it to the shopping cart instead of immediately purchasing it. Motivated by this observation, ESM^2 [18] models CVR prediction and some auxiliary tasks simultaneously in a multi-task learning framework based on the conditional probability rule elaborately defined on a new user behavior graph. However, these post-click behaviors in ESM^2 are all macro behaviors, which can only provide understandings of the subsequent purchase behavior at the item-level granularity, while more abundant fine-grained interactions are ignored, *e.g.*, clicks on certain components of items like comments or pictures, which we call micro behaviors in this paper. Micro behaviors can provide more detailed supplemental understandings of the subsequent macro behaviors and purchase behavior.

Few existing works pay attention to these valuable micro behaviors. To the best of our knowledge, Zhou et al. [26] firstly proposed to utilize user micro-behaviors in e-commerce recommendation, offering a new perspective to understand user behavior modeling. Gu et al. [5] proposed a hierarchical user profiling framework to solve the user profiling problem in e-commerce recommender systems, which not only utilizes macro behaviors but also harvests users' fine-grained micro behaviors. For CVR prediction, Bao et al. [2] proposed to utilize the micro behaviors to provide explicit instructions for CVR modeling. Specifically, it represents the micro behaviors as a graph and employs graph convolutional networks to thoroughly model the relationships among all the micro behaviors towards the final purchase target. However, it ignores the macro behaviors, which have been demonstrated their value in [18]. How to explicitly model both micro and macro behaviors in a unified framework for CVR prediction remains challenging and unexplored.

In this paper, we fill this gap by proposing a novel CVR prediction model named HM^3 which can hierarchically model both micro and macro behaviors in a unified framework. First, we construct a complete user sequential behavior graph to hierarchically represent micro behaviors and macro behaviors as one-hop and two-hop post-click nodes, respectively. Then, we embody HM^3 as a multi-head deep neural network (DNN) like [18]. Specifically, HM^3 predicts six probability variables corresponding to explicit sub-paths in the graph from six parallel sub-networks after a shared feature

embedding module. They are further combined into the targets of four auxiliary tasks and the final CVR according to the conditional probability rule defined on the graph. By employing multi-task learning, the whole model can be trained end-to-end. In this way, HM^3 can 1) efficiently use all the impression samples in the entire space to address the SSB and DS issues; 2) model both micro and macro behaviors in a hierarchical and complete manner by using a larger neural network compared with ESM^2 [18]; and 3) leverage the abundant supervisory labels from micro behaviors for learning better feature embeddings and facilitating the modeling of macro behaviors, which are more sparse compared with micro behaviors.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, we are the first to explicitly model both micro and macro behaviors in a unified framework for efficient and accurate CVR prediction.
- We propose a novel deep neural recommendation model named HM^3 , which can utilize all the impression samples and harvest the abundant labeled micro and macro post-click behaviors to efficiently mitigate the SSB and DS issues via multi-task learning.
- We conduct extensive experiments on both real-world offline dataset and online scenario and confirm the superiority of HM^3 over representative state-of-the-art methods.

2 THE PROPOSED APPROACH

2.1 Motivation and Novel User Behavior Graph

In ESM^2 [18], specific purchase-related post-click behaviors are incorporated into the user sequential behavior graph, based on which a unified multi-task learning framework is proposed to harvest the abundant labeled data from these post-click behaviors. In this way, ESM^2 outperforms $ESMM$ [13], which neglects post-click behaviors. Nevertheless, they both neglect abundant post-click micro behaviors, *e.g.*, clicks on certain components of items like comments or pictures, that are related to the macro behaviors and final purchase target. As discussed in the previous section, these micro behaviors are also useful for recommendation [2, 5, 26]. However, how to define their relationship with those macro behaviors as well as embody both micro and macro behaviors into a unified framework for CVR prediction still remains unexplored.

Table 1: Detailed behaviors in the D-Mi set and D-Ma set.

Type	Behaviors
D-Mi	clicking item's pictures, checking the Q&A details of an item, chatting with sellers, reading an item's comments, clicking an item's carting control button
D-Ma	putting an item into the shopping cart or wish list

For one thing, several elaborately defined micro behaviors have explicit and deterministic labels which are more abundant compared with macro behaviors as illustrated in Figure 1(a). For another, both micro behaviors and macro behaviors belong to post-click behaviors. Therefore, it is promising to leverage the abundant labeled data from them to learn better feature representations and predict more accurate CVR. Following *ESM*² [18], we construct a complete user sequential behavior graph to hierarchically represent micro behaviors and macro behaviors as one-hop and two-hop post-click nodes. Specifically, two nodes named Deterministic Micro set (*D-Mi*) and Deterministic Macro set (*D-Ma*) are defined to merge several predefined specific purchase-related post-click micro behaviors and macro behaviors, as listed in Table 1. These two nodes have two properties: 1) they are highly relevant to respective following nodes and purchase behaviors; and 2) they have abundant labels derived from users' feedback, *e.g.*, 1 for taking any kinds of predefined micro (macro) behaviors and 0 for none. In order to account for different kinds of post-click behaviors completely, we also add two nodes named Other Micro set (*O-Mi*) and Other Macro set (*O-Ma*) to deal with other post-click behaviors except *D-Mi* and *D-Ma*, respectively. Then, we hierarchically insert the four defined nodes between click and purchase, *D-Mi* and *O-Mi* first, followed by *D-Ma* and *O-Ma*. In this way, the conventional behavior path "impression→click→purchase" is transformed into a new user sequential behavior graph as depicted as in Figure 1(b).

2.2 Probability Decomposition on the Graph

Referring to Figure 1(b), given an item x_i , we define six probability variables y_{1i} , y_{2i} , y_{3i} , y_{4i} , y_{5i} , y_{6i} corresponding to explicit sub-paths in the graph, *i.e.*, "impression→click", "click→*D-Mi*", "*D-Mi*→*D-Ma*", "*D-Ma*→purchase", "*O-Mi*→*D-Ma*" and "*O-Ma*→purchase", respectively. Next, we define four auxiliary tasks over the entire space denoted as p_i^{ctr} , p_i^{D-Mi} , p_i^{D-Ma} , and p_i^{ctcor} , representing the probabilities of certain sub-paths, *i.e.*, "impression→click", "impression→*D-Mi*", "impression→*D-Ma*", and "impression→purchase", respectively. According to the conditional probability rule on the graph, the probabilities of these auxiliary targets can be derived from the aforementioned six probability variables as follows:

$$p_i^{ctr} = y_{1i}, \quad (1)$$

$$p_i^{D-Mi} = y_{1i} * y_{2i}, \quad (2)$$

$$p_i^{D-Ma} = y_{1i} * (y_{2i} * y_{3i} + (1 - y_{2i}) * y_{5i}), \quad (3)$$

$$\begin{aligned} p_i^{ctcor} &= y_{1i}y_{4i}(y_{2i}y_{3i} + (1 - y_{2i})y_{5i}) + \\ &\quad y_{1i}y_{6i}(y_{2i}(1 - y_{3i}) + (1 - y_{2i})(1 - y_{5i})) \\ &= p_i^{ctr} * p_i^{cor}. \end{aligned} \quad (4)$$

Here, p_i^{cor} denotes the probability of conversion rate of a clicked item x_i , which can be derived from the six probability variables.

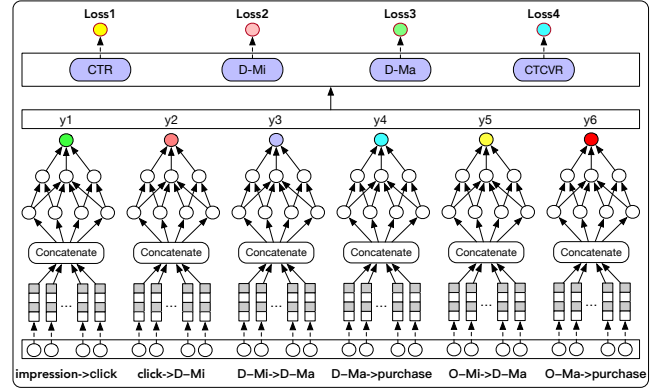


Figure 2: The diagram of the HM^3 . It has six sub-networks to predict the probabilities of corresponding explicit sub-paths, which are composed sequentially to calculate the final CVR and other targets of four related auxiliary tasks.

Table 2: Statistics of three training sets of different volumes.

Dataset	#user	#item	#impression	#click	#D-Mi	#D-Ma	#purchase
SR-S	32M	50M	4.9B	146M	36M	19M	5M
SR-M	68M	76M	14.8B	434M	102M	58M	16M
SR-L	107M	98M	31.7B	925M	235M	122M	32M

2.3 Multi-task Learning based on DNNs

How to model the above six probability variables as well as those four auxiliary tasks simultaneously? An intuitive way is to embody them into a unified multi-task learning framework. To this end, we propose a novel deep neural recommendation model HM^3 by hierarchically modeling both micro and macro behaviors for CVR prediction. The network structure of HM^3 is diagrammed in Figure 2. Firstly, we devise six parallel sub-networks after a shared feature embedding module (FEM) to predict the probability variables. Since FEM makes up the majority of the parameters, sharing the FEM makes the whole network lightweight. In addition, FEM can receive abundant back-propagated supervisory signals to mitigate the DS issue and learn good feature embeddings.

Then, based on the predicted conditional probabilities $y_{1i} \sim y_{6i}$ from the corresponding sub-networks, the probabilities of four auxiliary targets can be derived sequentially according to Eq. (1) ~ Eq. (4). Our model can not only efficiently estimate six probability variables in parallel but also do not require extra parameters for deriving CVR and auxiliary targets, which are very suitable for online deployment and responding users' requests in a low latency. Finally, based on the supervisory labels from click behaviors, micro (macro) behaviors, and purchase behaviors corresponding to the four binary classification tasks, we can define the final training objective as a combination of four cross-entropy losses. Once the model being trained, we can derive p_i^{cor} from Eq. (4) accordingly. It is noteworthy that the SSB issue is well addressed due to the fact that all the auxiliary tasks are modeled over the entire space. Besides, due to the hierarchically modeling over the user sequential behavior graph (referring to Figure 1(b) and Figure 2), abundant supervisory signals from micro behaviors can better supervise the learning of macro behaviors, coordinately towards the final purchase target.

3 EXPERIMENTS

3.1 Experiment settings

3.1.1 Dataset preparation. We collected the consecutive logs between 2020-09-16 and 2020-09-30 from the *Shopping Recommendation (SR)* module of our recommender system¹. We used the last day data as the test set. To investigate the impact of data volume, we prepared three training sets: *SR-S*, *SR-M*, and *SR-L*, containing data from 2020-09-29, 2020-09-23 to 2020-09-29, 2020-09-16 to 2020-09-29, respectively. Table 2 summarizes their statistics.

3.1.2 Competitors. 1) *BASE* [7]: a deep neural network model which uses the same input features as *HM*³. The network structure and hyper parameters are identical to each single branch of *HM*³, such as *ReLU* for activation function, *Adam* for optimizer, 0.0005 for learning rate, etc. It predicts the probabilities p^{ctr} or p^{cwr} using samples from “impression→click” or “click→purchase”, respectively. 2) *ESMM* [13]: it predicts p^{ctr} and p^{cwr} over the entire space without using the purchase-related macro post-click behaviors. 3) *ESM*² [18]: it models purchase-related post-click macro behaviors in a unified multi-task learning framework, while neglecting micro behaviors. 4) *GMCM* [2]: it represents user micro behaviors as a graph and employs graph convolutional networks to model their interaction relationships towards the final purchase target. 5) *ESM*² + *Mi*: it uses the same network structure as *ESM*² [18]. The only difference is that it uses post-click behaviors from both macro behaviors as well as micro behaviors. 6) *HM*³ – *R*: it uses the same network structure as *HM*³ while only reversing the order of micro and macro behaviors in the graph, *i.e.*, D-Ma (O-Ma) is in front of D-Mi (O-Mi).

3.1.3 Evaluation Metrics. We adopt the widely used Area Under Curve (AUC) [25, 27] as the evaluation metric, which indicates the ranking performance. Specifically, we report the AUC scores for the CVR prediction task and CTCVR prediction task. The latter task measures the probability of purchase given impression.

3.2 Main results

Table 3: The AUC scores of all methods for CVR.

	SR-S	SR-M	SR-L
Method	CVR AUC	CVR AUC	CVR AUC
BASE	0.84117	0.84583	0.84932
ESMM	0.84297	0.84601	0.85098
<i>ESM</i> ²	0.84493	0.84794	0.85213
GMCM	0.84624	0.84928	0.85405
<i>ESM</i> ² + <i>Mi</i>	0.84703	0.85118	0.85596
<i>HM</i> ³ – <i>R</i>	0.84799	0.85285	0.85646
<i>HM</i> ³	0.84891	0.85328	0.85726

Table 3 and Table 4 summarize the respective AUC results of different methods for the CVR and CTCVR tasks. Taking the results of set *SR-L* in Table 3 as an example, it can be seen that *ESMM* achieves a gain of 0.00166 over the *BASE* model, because the *SSB* and *DS* issues could be mitigated in *ESMM* by modeling the problem over the entire space. As for *ESM*², it uses more

abundant purchase-related post-click behaviors to alleviate the sparsity issue of positive training samples and therefore achieving a gain of 0.00115 over *ESMM*. However, *ESM*² ignores the abundant micro behaviors, which are also related to the purchase target as well as macro behaviors. By using these micro behaviors, *GMCM* outperforms *ESM*² by 0.00192 AUC. After modeling micro behaviors, *ESM*²+*Mi* improves the performance of the vanilla *ESM*² by 0.00383 AUC, confirming the value of exploring micro behaviors. It also outperforms *GMCM* by 0.00191 AUC, indicating the benefit of modeling both kinds of behaviors together. Our *HM*³ obtains the best results among all the methods. For example, it outperforms the *BASE* model by 0.00794 AUC. Compared with *ESM*² + *Mi* and *HM*³ – *R*, *HM*³ also performs better, which demonstrates that hierarchically modeling the micro behaviors and the macro behaviors is more effective. Similar results can be found on other two datasets for the CVR task as well as for the CTCVR task. It is worth mentioning that the slightly better AUC score at the setting of a larger dataset volume means a significant increment in revenue for online deployment according to our practice.

Table 4: The AUC scores of all methods for CTCVR.

	SR-A	SR-B	SR-C
Method	CTCVR AUC	CTCVR AUC	CTCVR AUC
BASE	0.84883	0.85532	0.85993
ESMM	0.85051	0.85652	0.86192
<i>ESM</i> ²	0.85148	0.85804	0.86296
GMCM	0.85241	0.85953	0.86411
<i>ESM</i> ² + <i>Mi</i>	0.85402	0.86098	0.86596
<i>HM</i> ³ – <i>R</i>	0.85516	0.86163	0.86712
<i>HM</i> ³	0.85675	0.86213	0.86806

We also conducted online A/B test between different methods and the *BASE* model in the *SR* modules of our recommendation platform from 2020-10-08 to 2020-10-21. The results are summarized in Table 5. *HM*³ brings 8.27% CVR gain and 8.32% GMV (Gross Merchandise Volume) gain over the *BASE* model, *i.e.*, larger margins than other methods, which are consistent with the offline evaluation results, indicating a significant business revenue growth.

Table 5: Results of Online A/B test.

Method	CVR Gain	GMV Gain
BASE	0%	0%
ESMM	+2.76%	+3.02%
<i>ESM</i> ²	+4.84%	+5.11%
<i>HM</i> ³	+8.27%	+8.32%

4 CONCLUSION

We propose a novel deep model *HM*³ for CVR prediction. It hierarchically models micro and macro post-click behaviors in a unified multi-task learning framework according to a new user sequential behavior graph. *HM*³ efficiently mitigates the *SSB* and *DS* issues by leveraging abundant supervisory signals from the behaviors and all impression samples over the entire space. Experiments on offline datasets and online test show *HM*³ outperforms state-of-the-art models. In the future, we intend to investigate the impact of modeling post-click behaviors at a finer granularity.

¹To the best of our knowledge, there are no available large-scale public datasets containing both micro and macro behaviors to benchmark our method.

REFERENCES

- [1] Ben Athiwaratkun and Keegan Kang. 2015. Feature representation in convolutional neural networks. *arXiv preprint arXiv:1507.02313* (2015).
- [2] Wentian Bao, Hong Wen, Sha Li, Xiao-Yang Liu, Quan Lin, and Keping Yang. 2020. Gmcm: Graph-based micro-behavior conversion model for post-click conversion rate estimation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2201–2210.
- [3] Yufei Feng, Fuyu Lv, Weichen Shen, Menghan Wang, Fei Sun, Yu Zhu, and Keping Yang. 2019. Deep Session Interest Network for Click-Through Rate Prediction. *arXiv preprint arXiv:1905.06482* (2019).
- [4] Yu Gong, Ziwen Jiang, Yufei Feng, Binbin Hu, Kaiqi Zhao, Qingwen Liu, and Wenwu Ou. 2020. EdgeRec: Recommender System on Edge in Mobile Taobao. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2477–2484.
- [5] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, and Dawei Yin. 2020. Hierarchical User Profiling for E-commerce Recommender Systems. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 223–231.
- [6] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).
- [7] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18, 7 (2006), 1527–1554.
- [8] Zhenhua Huang, Xin Xu, Juan Ni, Honghao Zhu, and Cheng Wang. 2019. Multimodal representation learning for recommendation in Internet of Things. *IEEE Internet of Things Journal* 6, 6 (2019), 10675–10685.
- [9] Farid Huseynov. 2020. Intelligent Recommender Systems in E-Commerce: Opportunities and Challenges for Online Customers. *Handbook of Research on IT Applications for Strategic Competitive Advantage and Decision Making* (2020), 36–51.
- [10] Kuang-chih Lee, Burkay Orten, Ali Dasdan, and Wentong Li. 2012. Estimating conversion rate in display advertising from past performance data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 768–776.
- [11] Quan Lu, Shengjun Pan, Liang Wang, Junwei Pan, Fengdan Wan, and Hongxia Yang. 2017. A practical framework of conversion rate prediction for online display advertising. In *Proceedings of the ADKDD'17*. 1–9.
- [12] Wei Lu, Shanshan Chen, Keqian Li, and Laks VS Lakshmanan. 2014. Show me the money: Dynamic recommendations for revenue maximization. *arXiv preprint arXiv:1409.0080* (2014).
- [13] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. 2018. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 1137–1140.
- [14] Qi Pi, Weijie Bian, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Practice on long sequential user behavior modeling for click-through rate prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2671–2679.
- [15] J Ben Schafer, Joseph Konstan, and John Riedl. 1999. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*. 158–166.
- [16] Kangning Wei, Jinghua Huang, and Shaohong Fu. 2007. A survey of e-commerce recommender systems. In *2007 international conference on service systems and service management*. IEEE, 1–5.
- [17] Hong Wen, Jing Zhang, Quan Lin, Keping Yang, and Pipei Huang. 2019. Multi-Level Deep Cascade Trees for Conversion Rate Prediction in Recommendation System. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- [18] Hong Wen, Jing Zhang, Yuan Wang, Fuyu Lv, Wentian Bao, Quan Lin, and Keping Yang. 2020. Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2377–2386.
- [19] Chenyan Xu, Daniel Peak, and Victor Prybutok. 2015. A customer value, satisfaction, and loyalty perspective of mobile application recommendations. *Decision Support Systems* 79 (2015), 171–183.
- [20] Hongxia Yang, Quan Lu, Angus Xianen Qiu, and Chun Han. 2016. Large scale cvr prediction through dynamic transfer learning of global and local features. In *Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*. 103–119.
- [21] Bianca Zadrozny. 2004. Learning and evaluating classifiers under sample selection bias. In *Proceedings of the twenty-first international conference on Machine learning*. 114.
- [22] Jing Zhang and Dacheng Tao. 2020. Empowering Things with Intelligence: A Survey of the Progress, Challenges, and Opportunities in Artificial Intelligence of Things. *IEEE Internet of Things Journal* (2020).
- [23] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–38.
- [24] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- [25] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1059–1068.
- [26] Meizi Zhou, Zhuoye Ding, Jiliang Tang, and Dawei Yin. 2018. Micro behaviors: A new perspective in e-commerce recommender systems. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 727–735.
- [27] Han Zhu, Junqi Jin, Chang Tan, Fei Pan, Yifan Zeng, Han Li, and Kun Gai. 2017. Optimized cost per click in taobao display advertising. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.