Multi-task Representation Learning for Demographic Prediction

1 author:

Pengfei Wang
Chinese Academy of Sciences

Multi-task Representation Learning for Demographic Prediction

Abstract. Demographic attributes are important components for market analysis, which are widely used to characterize different types of users. However, such signals are only available for a small fraction of users due to the difficulty in manual collection process by retailers. Most previous work on this problem explores different types of features and usually predicts different attributes independently. However, manually defined features require professional knowledge and often suffer from under specification. Meanwhile, model the tasks separately may lose the ability to leverage the correlations among different attributes. In this paper, we propose a novel Multi-task Representation Learning (MTRL) model to predict users‘ demographic attributes. Comparing with the previous methods, our model conveys the following merits: 1)By using a multi-task approach to learn the tasks, our model leverages the large amounts of cross-task data, which is helpful to task with limited data; 2)MTRL uses a supervised way to learn the shared semantic representation across multiple tasks, thus it can obtain a more general and robust representation by considering the constraints among tasks. Experiments are conducted on a real-world retail dataset where three attributes (gender, marital status, and education level) are predicted. The empirical results show that our MTRL model can improve the performance significantly compared with the state-of-the-art baselines.

Keywords: Multi-Task, demographic prediction, representation learning

1 Introduction

Obtaining users' demographic attributes is crucial for retailers to conduct market basket analysis [17], adjust marketing strategy [9], and provide personalized recommendations [19, 21]. However, in practice, it is difficult to obtain users‘ demographic attributes, because most users are reluctant to offer their detailed information or even refuse to give their demographics due to privacy and other reasons. This is particularly true for traditional offline retailers[1], who collect users' demographic information mostly in a manual way (e.g. requiring costumers to provide demographic information when registering some shopping cards).

[1] In this work, we mainly focus on traditional retailers in offline business rather than those in online e-commerce, where no additional behavioral data rather than transactions is available for analysis. Hereafter we will use retail/retailer for simplicity when there is no ambiguity.

In this paper, we try to inference users' demographic attributes based on users' purchase history. Although some recent studies suggest that demographic attributes are predictable from different behavioral data, such as linguistics writing [5], web browsing [15], electronic communications [8, 11] and social media [13, 23], to our best knowledge, seldom practice has been conducted on purchase behaviors in retail scenario.

Pervious work on demographic prediction usually predicts different attributes independently based on manually defined features [3, 16, 18, 22, 23]. For example, Zhong et al. [23] predicted six demographic attributes (i.e., gender, age, education background, sexual orientation, marital status, blood type and zodiac sign) separately by merging spatial, temporal and location knowledge features into a continuous space. However, manually defined features usually require professional knowledge and often suffer from underspecification. Meanwhile, by predicting each attribute independently, one may not be able to leverage the potential correlations between different attributes (e.g., correlation between age and marital status). Some recent studies proposed to take the relations between different attributes into account [3, 22]. For example, Dong et al. [3] employed a Double Dependent-Variable Factor Graph model to predict gender and age simultaneously. Zhong et al. [22] attempted to capture pairwise relations between different tasks when predicting six demographic attributes from mobile data. However, these methods still rely on various human-defined features which are often costly to obtain.

To tackle the above problem, in this paper we make a fist approach to predict users' gender, martial status, and education level based on users' purchase history. A Multi-task Representation Learning Neural(MTRL) model is used to predict users' demographic attributes. MTRL learns a shared semantic representations across multiple tasks, which benefits from a more general representation for prediction. Specifically, we characterize each user by his/her purchase history using the bag-of-item representations. We then map all users' representations into semantic representations learned by a multi-task method. Thus we can obtain a more general shared representation to guide the prediction task separately. Compared with previous methods, the major contributions of our work are as follows:

– We make the first attempt to investigate the prediction power of users' purchase data for demographic prediction in retail scenario.
– Through MTRL we can obtain a shared representation learned across multiple tasks, which benefits a more general representation to help the task with limited data.
– We conduct extensive experiments on a real-world retail dataset to demonstrate the effectiveness of the proposed MTRL model as compared with different baseline methods.

The rest of the paper is organized as follows. After a summary of related work in Section 2, we describe the problem formalization of demographic prediction in retail scenario in Section 3. In section 4 we present our proposed model in detail. Section 5 concludes this paper and gives the future work.

2 Related Work

In this section we briefly review three research areas related to our work: demographic prediction, multi-task prediction, and representation learning.

2.1 Demographic Prediction

Demographic inference has been studied in different scenarios in academia. Early work on demographic prediction attempted to predict demographic attributes based on the linguistics writing and speaking. For example, Schler et al. [18] found that there are significant differences in both writing style and content between male and female bloggers as well as among authors of different ages. Otterbacher [16] used a logistic regression model to infer users' gender based on content of reviews.

Later, the digital communication and Internet offered new opportunities for demographic attribute prediction. Different approaches have been proposed to infer demographic attributes based on users' browsing history [8, 15]. Torres [4] found that the clicked pages were correlated with the demographic characteristics of users. Hu et al. [8] calculated demographic tendency of Webpages, and modeled users' demographic attributes through a discriminative model. In [1], Bi et al. infers the demographic attributes of search users based on the models training on the independent social datasets. They demonstrated that by leveraging social and search data in a common representation, they can achieve better accuracy in demographic prediction.

Recently, the fast development of online social networks and mobile computing technologies accumulated large scale of user data, making it possible and also valuable to infer users' demographic attributes in these scenarios. Mislove [13] found that users with common profiles were more likely to be friends and often formed a dense community. Zhong et al. [22] proposed a supervised learning framework to predict users' demographic attributes based on mobile data. Dong et al. [3] focused on micro-level analysis of the mobile networks to infer users' demographic attributes. Culotta et al. [2] fitted a regression model to predict users' demographic attributes using information on followers of each website on Twitter.

As we can see, most existing work on demographic prediction focused on designing different features for the prediction tasks. Besides, to the best of our knowledge, seldom practice has been conducted on demographic prediction based on purchase behaviors in retail scenario.

2.2 Multi-task Prediction

The idea of learning multiple tasks together is to improve the generalization performance by leveraging the information contained in the related tasks. A typical way for this purpose is to learn tasks in parallel while using a shared representation [3, 20, 22]. Many algorithms have been proposed to solve multi-task learning with various kernels and regularizers to address the correlation between tasks. For example, Micchelli et al. [12] discussed how different kernels can be used to model relations between tasks and presented linear multi-task learning algorithms. Evgeniou et al. [6] presented an approach to multi-task learning based on the minimization of regularization functionals.

2.3 Representation Learning

Learning representations of the data makes it easier to extract useful information when building classifiers or other predictors. That is why representation learning has attracted more and more attention and become a field in itself in the machine learning community.

Many remarkable empirical successes have been achieved based on representation learning in various applications in both academia and industry. For example, in speech recognition and signal processing, Alex Graves et al. [7] designed a deep recurrent neural network for speech recognition and obtain the best score on beachmark. In object recognition, Krizhevsky et al. [10] proposed to use convolutional neural network to classify image and achieved the record-breaking results. In natural language processing, Mnih [14] proposed three graphical models to define the distribution of next word in a sequence by using distributed representations.

In this work, we propose to use multi-task to learn a shared representation for demographic prediction in retail scenario, a new application area where representation learning might be helpful, especially to the task with limited data.

3 Our Approach

In this section, we first give the motivation of our work, then we introduce the formalization of demographic prediction problem in retail scenario. After that, we describe the proposed MTRL in detail. Finally, we present the learning procedure of MTRL.

3.1 Motivation

Obviously, the fundamental problem in demographic prediction based on users' behavior data is how to represent users. Many existing work investigated different types of human defined features [3, 16, 22]. However, it is usually costly to define features manually since expertise knowledge is required and one has to do the same job task by task. Moreover, human defined features may often suffer from underspecification since it is difficult to identify those hidden complicated factors for prediction tasks. Some recent work employs unsupervised feature learning methods [8, 11, 23], like Singular Vector Decomposition (SVD), to automatically extract low-dimension features from the raw data. However, the features learned in an unsupervised manner may not be optimal for the prediction tasks. Therefore, in this work we proposed to automatically learn representations of users for demographic prediction in a supervised way.

In addition, users' demographic attributes are related, as shown in the Figure **??**. For example, there are more male PHD students than female, or if a user purchases baby supply, he/she may be an adult who has married. However, most previous work treated different attributes as separate prediction tasks [2, 11, 23], thus ignored the correlations between these attributes.

Motivated by all this, we propose a multi-task approach to learn a general representation to predict users' demographic attributes.

3.2 Problem Formalization

In our work, we aim to predict multiple demographic attributes based on users' behavioral data in retail scenario. Specifically, each user can be characterized by his/her purchase history, i.e., a set of items. The demographic attributes we are interested include gender, marital status, and education level, which are useful signals for market analysis. The values of each attribute take are shown in Figure 1. Given a user, based on his/her purchase history, we want to predict all the unknown attributes.

Table 1: List of demographic attributes

Attributes	**Values**
gender	male, female
marital status	single, married
education level	doctor, master, bachelor, college, high school, middle school

Specifically, let $T = \{t_1, t_2, \ldots, t_n\}$ be a set of demographic prediction tasks (i.e., predicting demographic attributes). Let U be a set of users. Suppose the training set is composed of M instances, i.e.,

$$\{(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \ldots, (x_{(M)}, y_{(M)})\}$$

where $x_{(i)} \in X$ is a d-dimensional feature vector, representing the input of i-th user, and $y_{(i)}$ is the set of attribute labels of the i-th user. Note here $y^t_{(i)}$ denotes the attribute label under the t-th task $t \in T$ for the i-th user.

Based on the notations defined above, we try to learn a function to predict the unknown demographic attributes.

3.3 Multi-task Learning Representation Neural Model

In this section, we now present the proposed MTRL model in detail. The feed-forward MTRL is shown in Figure 1. In retail scenario, each user is characterized

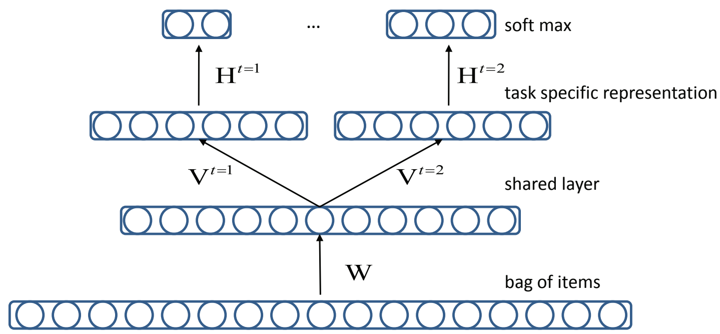


Fig. 1: The structure of Multi-task Learning Representation Neural(MTRL)model. The lower two layers are shared across all the tasks, while top layers are task-specific. The input is represented as a bag of items, then a non-linear projection $\mathbf{W}$ is used to generate a shared representation. Finally, for each task, additional non-linear projection $\mathbf{V}$ generates task-specific representations.

by his/her purchase history, i.e., a set of items. In MTRL, we take the bag-of-item representation as the user input $x_{(i)}$, then the shared layer is fully connected to the input layer with weights Matrix $\mathbf{W} = [w_{h,s}]$:

$$\mathcal{Y}_{(i),s} = f(\sum_h w_{h,s} \cdot x_{(i),h})$$

where matrix $\mathbf{W}$ is responsible for generating the cross-task representation, $\mathcal{Y}_{(i),s}$ is the value of s-th node on the shared layer,$x_{(i),h}$ is the h-th value of $x_{(i)}$, and $f(z)$ is the logistic nonlinear activation:

$$f(z) = \frac{1}{1 + e^{-z}}$$

Based on the input of the shared layer, for each task t, we use a transformation $V^t = [v_{s,j}]$ to map the shared representation into the task-specific representation by:

$$\mathcal{Y}^t_{(i),j} = f(\sum_s v^t_{s,j} \cdot \mathcal{Y}_{(i),s})$$

where t denotes the different tasks(gender, martial status, and education level), and l^t_j is the value of j-th node corresponding to the specific representation layer of task t.

After these, we use a softmax activate function calculate the value of k-th node in the output layer:

$$\mathcal{Y}^t_{(i),k} = \frac{exp(\sum_j h^t_{j,k} \cdot \mathcal{Y}^t_j)}{\sum exp(\sum_j h^t_{j,k} \cdot \mathcal{Y}^t_j)}$$

where $H^t = [h_{j,k}]$ is the matrix that maps task specific representation to the output layer for task t, k-th node in the output layer corresponds the value of the k-th label in task t.

The objective function of MTRL is then defined as the cross-entropy over the outputs of all the users and all tasks:

$$\ell_{MTRL} = \sum_t \sum_i \sum_k d^t_{(i),k} \ln \mathcal{Y}^t_{(i),k} + (1 - d^t_{(i),k}) \ln(1 - \mathcal{Y}^t_{(i),k}) - \lambda \|\Theta\|^2_F \quad (1)$$

where $d^t_{(i),k}$ is the real value of k-th node for user i under the task t, λ is the regularization constant and Θ are the model parameters (i.e. $\Theta = \{\mathbf{W}, \mathbf{V}^t, \mathbf{H}^t\}$).

3.4 Learning and Prediction

In order to learn parameters of MTRL model, we use the stochastic gradient decent algorithm, as shown in Algorithm 1. For each iteration, a task t is randomly selected, and parameters of the model is updated according to the task-specific objective.

Algorithm 1 Algorithm for Multi-task Learning Representation Neural Model

Initialize model Θ: $\{\mathbf{W}, \mathbf{V}, \mathbf{H}\}$ randomly
iter=0
repeat
    $iter \leftarrow iter + 1$;
    for i=1,...,M **do**
        select a task t randomly for instance $x_{(i)}$
        compute the gradient$\nabla(\Theta)$
        update model $\Theta \leftarrow \Theta + \epsilon\nabla(\Theta)$
    end for
until converge or iter > num
return $\mathbf{W}, \mathbf{V}, \mathbf{H}$

4 Experiments

In this section, we conduct empirical experiments to demonstrate the effectiveness of our proposed MTRL model on demographic attribute prediction in retail scenario. We first introduce the experimental settings. Then we compare our MTRL model to the baseline methods to demonstrate the effectiveness of predicting users' demographic attributes in retail scenarios.

4.1 Experimental Settings

In this section, we introduce the experimental settings including the dataset, baseline methods, and evaluation metrics.

Dataset We conduct our empirical experiments over a real world large scale retail dataset, namely BeiRen dataset[2]. This dataset comes from a large retailer[3] in China, which records its supermarket purchase histories during the period from 2012 to 2013. It contains $49,290,149$ transactions over $220,828$ items belonging to $1,206,379$ users. For research purpose, the dataset has been anonymized with all the users and items denoted by randomly assigned IDs for the privacy issue.

We first conduct some pre-process on the BeiRen dataset. We randomly collected 100000 users. We extract all the transactions related to these users to form their purchase histories, then we remove all the items bought by less than 10 times and the users with no labels. After pre-processing, the dataset contains 64097 distinct items and 80540 distinct users with at least on attribute. In average, each user has bought about 225.5 distinct items.

Baseline Methods We evaluate our model by comparing with several state-of-the-art methods on demographic attribute prediction:

- BoI-Single: Each user is represented by the items he/she has purchased with the Bag-of-Item representation, and a logistic model [4] is learned to predict each demographic attribute separately.
- SVD-single: A singular value decomposition (SVD)[5] is first conducted over the user-item matrix to obtain low dimensional representations of users. Then a logistic model is learned over the low dimensional representation to predict each demographic attribute separately. This method has been widely used in demographic attribute prediction task [8, 15, 23].
- SL: The Representation Learning model is a special case of MTRL when their is only a single task to learn. Thus SL has the same neural structure comparing with MTRL, just without considering the relationships among tasks.

For SVD-single method, we run several times with random initialization by setting the dimensionality as 200. For MTRL and SL, we set the dimensionality of shared representation layer and task-specific representation layer as 200 and 100 respectively. The parameters are initialized with uniform distribution in the range $(-\sqrt{6/(fan_{in}+fan_{out})}, \sqrt{6/(fan_{in}+fan_{out})})$.

For each experiment, a 5-fold cross-validation is performed on this data set. In detail, the original data set is randomly divided into 5 parts each with approximately the same size. In each fold, one part is held-out for testing and the learning algorithm is trained on the remaining data. The above process is iterated 5 times, where the average metric values out of ten runs are reported for the algorithm.

[2]This dataset would be publicly available after the acceptance of the paper
[3]http://www.brjt.cn/
[4]http://www.csie.ntu.edu.tw/c̃jlin/liblinear/
[5]http://tedlab.mit.edu/d̃r/SVDLIBC/

Evaluation Metrics we follow the idea in [3] to use weighted F1 as an evaluation metric. For task t, the weighted F1 is computed as follows:

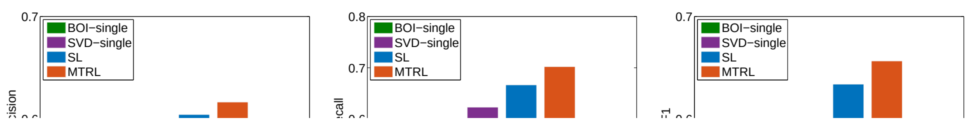
$$\text{wPrecision} = \sum_{y \in t} \frac{\sum_i I(y^t_{(i)} = y^{*,t}_{(i)} \& y^t_{(i)} = y)}{\sum_i I(y^t_{(i)} = y)} \cdot \frac{\sum_i I(y^{*,t}_{(i)} = y)}{|U|}$$

$$\text{wRecall} = \frac{1}{|U|} \sum_i I(y^{t,*}_{(i)} = y^t_{(i)})$$

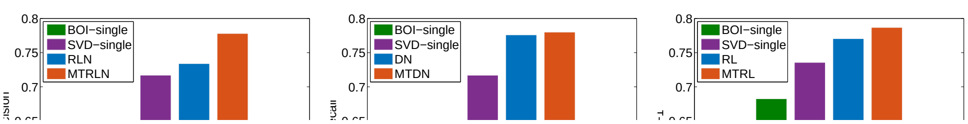
$$\text{wF1} = \frac{2 \times \text{wPrecision} \times \text{wRecall}}{\text{wPrecision} + \text{wRecall}}$$

where $y^{*,t}_{(i)}$ is the true label of user i under task t, $I(\cdot)$ is an indicator function. Note here we use the weighted evaluation metrics because every class in task gender, marital status and education is as important as each other. As we can see, the weighted recall is the prediction accuracy in the user view.

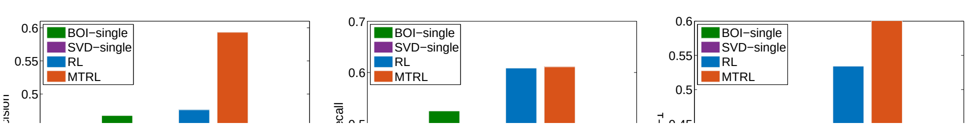
Fig. 2: Performance comparison of different methods on BeiRen dataset.



(a) gender



(b) marital status



(c) education level

The performance of different methods is shown in Figure 2. We have the following observations:

- (1) Using SVD to obtain low-dimension representations of users can achieve a better performance than BoI on predicting each demographic attribute of users. This result is quite coincidence with the previous finds [23, 11].
- (2) Both the deep models SL and MTRL perform better than SVD-single, it demonstrates that the deep model can learn a better representation comparing with the shallow one(here we regard SVD-single as a shallow model)
- (3) By using a multi-task approach to learn a shared representation layer across tasks, we can obtain a better performance than SL, which proves that the correlations among demographic attributes are helpful.
- (4) Finally, our MTRL can achieve the best performance in terms of all the evaluation measures, for example, when compared with the SVD-single method(the state-of-the-art method), the improvement on gender, marital status, and education level is 1.2%, 6.1%, and 13.4% respectively. By conducting the significant test, we find that the improvement of MTRL over the second best method (SL) is significant ($p - value < 0.01$) in terms of all the evaluation metrics.

Table 2: Performance comparison of different methods on BeiRen over different user groups .

user activeness	method	Gender	Martial Status	Education Level
unactive (4510)	BoI-single	0.522	0.660	**0.403**
	SVD-single	0.571	0.729	0.401
	SL	0.614	**0.747**	0.313
	MTRL	**0.678**	0.729	0.388
medium (4286)	BoI-single	0.589	0.686	0.506
	SVD-single	0.591	0.754	0.536
	SL	0.634	0.768	0.558
	MTRL	**0.645**	**0.802**	**0.647**
active (1200)	BoI-single	0.587	0.691	0.523
	SVD-single	0.568	**0.742**	0.526
	SL	0.646	0.716	0.533
	MTRL	**0.658**	0.732	**0.628**

To further investigate the performance of different methods, we split the users into three groups (i.e. inactive, medium and active) based on their activeness and conducted the comparisons on different user groups. We treat the user as inactive if there are less than 100 items in his/her purchase history, and active if there are more than 500 items in the purchase history. The remaining users are taken as medium. In this way, the proportions of inactive, medium and active are 40.8%, 54.5%, and 4.7% respectively. Here we only report the comparison results on Ta-Feng dataset with the dimension equals 50 due to the page limitation. In fact, similar conclusions can be drawn from other datasets. The results are shown in Table 2.

From the results we can see that, not surprisingly, the BoI-single method is still the worst on all the groups. Furthermore, we find that MC, FMC works better than NMF on both inactive and medium users in terms of all the measures; While on active users, NMF can achieve better performance than MC, FMC . The results indicate that it is difficult for NMF to learn a good user representation with few transactions for recommendation. Finally, CFSH can achieve the best performances on all the groups in terms of all the measures. The results demonstrate that by combining both item-centric and user-centric paradigms, we can enjoy the merits of both methods and complement each other to achieve better performance.

5 conclusion

In this paper, we try to predict users‘ demographic attributes given users' purchase behaviors . We propose a robust and practical representation learning algorithm MTRL based on multi-task objectives. Our MTRL can learn a shared representation across tasks, thus the sparseness problem can be avoided, especially to the task with limited data. Experiments on real-world purchase dataset demonstrate that our model can outperform the state-of-the-art baselines consistently under different evaluation metrics.

Although the MTRL model is proposed in this retail scenario, it is in fact a general model which can be applied on other multi-task multi-class problems. In the future, we would like to extend the usage of our MTRL model to other applications to verify its effectiveness. More over, in this paper, we represent each user by simple bag of items as the raw input. It would be interesting to further explore the natural transaction structures in users' purchase data for better demographic prediction.

References

1. B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. Inferring the demographics of search users: Social data meets search queries. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 131–140, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
2. C. J. Culotta A, Ravi N K. Predicting the demographics of twitter users from website traffic data[c]. *Proceedings of the International Conference on Web and Social Media (ICWSM), in press. Menlo Park, California: AAAI Press*, 2015.
3. Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 15–24, New York, NY, USA, 2014. ACM.
4. S. Duarte Torres and I. Weber. What and how children search on the web. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 393–402, New York, NY, USA, 2011. ACM.

5. P. Eckert. Gender and sociolinguistic variation. *Readings in Language and Gender*, 1997.
6. T. Evgeniou and M. Pontil. Regularized multi–task learning. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 109–117, New York, NY, USA, 2004. ACM.
7. H. G. Graves A, Mohamed A. Speech recognition with deep recurrent neural networks. *Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.
8. J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen. Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 151–160, New York, NY, USA, 2007. ACM.
9. P. D. S. Kalyanam K. Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Science*, 1997.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, page 2012.
11. D. S. M. Kosinski and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 2013.
12. C. Micchelli and M. Pontil. Kernels for multi-task learning. *NIPS*, 2005.
13. A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: Inferring user profiles in online social networks. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 251–260, New York, NY, USA, 2010. ACM.
14. A. Mnih and G. Hinton. Three new graphical models for statistical language modelling. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 641–648, New York, NY, USA, 2007. ACM.
15. D. Murray and K. Durrell. Inferring demographic attributes of anonymous internet users. *In Web Usage Analysis and User Profiling*, 2000.
16. J. Otterbacher. Inferring gender of movie reviewers: Exploiting writing style, content and metadata. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 369–378, New York, NY, USA, 2010. ACM.
17. I. S. C. Rick L. Andrews. Identifying segments with identical choice behaviors across product categories: An intercategory logit mixture model. *International Journal of Research in Marketing*, 2002.
18. J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker. Effects of age and gender on blogging. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 199–205. AAAI, 2006.
19. S. Sedhain, S. Sanner, D. Braziunas, L. Xie, and J. Christensen. Social collaborative filtering for cold-start recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 345–348, New York, NY, USA, 2014. ACM.
20. S. S. You Ji. Multitask multiclass support vector machines. *Data Mining Workshops (ICDMW)*, 2011.
21. X. W. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li. We know what you want to buy: A demographic-based system for product recommendation on microblogs. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1935–1944, New York, NY, USA, 2014. ACM.

22. E. Zhong, B. Tan, K. Mo, and Q. Yang. User demographics prediction based on mobile data. *Pervasive Mob. Comput.*, 9(6):823–837, Dec. 2013.
23. Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie. You are where you go: Inferring demographic attributes from location check-ins. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, pages 295–304, New York, NY, USA, 2015. ACM.