insurance and medical companies, financial institutions, and many more.

Acknowledgements
This material is based upon work supported by the National Science Foundation under grant numbers CNS-1314632, IIS-1408924, and DGE-1252522 as well as a Facebook Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other funding parties.

References
1. G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, 1998.
2. J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *J. ACM*, vol. 46, no. 5, 1999, pp. 604–632.
3. B.A. Prakash et al., "Eigenspokes: Surprising Patterns and Scalable Community Chipping in Large Graphs," *Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining*, 2010, pp. 435–448.
4. M. Jiang et al., "Inferring Strange Behavior from Connectivity Pattern in Social Networks," *Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining*, 2014, pp. 126–138.
5. S. Pandit et al., "Netprobe: A Fast and Scalable System for Fraud Detection in Online Auction Networks," *Proc. World Wide Web Conf.*, 2007, pp. 201–210.
6. D. Koutra et al., "Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms," *Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining*, 2011, pp. 245–260.
7. A. Beutel et al., "CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks," *Proc. 22nd Int'l Conf. World Wide Web*, 2013, pp. 119–130.

Alex Beutel a PhD candidate in the Department of Computer Science at Carnegie Mellon University. Contact him at abeutel@cs.cmu.edu.

Christos Faloutsos is a professor in the Department of Computer Science at Carnegie Mellon University. Contact him at christos@cs.cmu.edu.

Transfer Learning for Behavior Prediction

Weike Pan, *Shenzhen University*
Qiang Yang, *Hong Kong University of Science and Technology*

Behavior prediction such as user choice and feedback forecasting is of critical importance to the success of online e-commerce and social networking services. However, there are often some fundamental challenges associated with the task of behavior prediction, such as scarcity, uncertainty, and heterogeneity of users' behaviors and preferences.

Transfer learning has the potential to address these challenges in a unified framework via learning and predicting users' behavior patterns from a novel perspective, by sharing common knowledge between different but related sets of user behaviors. Transfer learning for behavior prediction is a new interdisciplinary research area that has largely not been explored yet, for which we'll mainly discuss the challenges and opportunities.

Behavior Prediction
User behavior data[1] is one of the most valuable resources for an online service provider because it can be exploited to help predict future behaviors, improve user satisfaction, and contribute revenues to the company. Users' behaviors are usually stored in tuples in the form of (user, entity, behavior), where the entity can be a product or a different user, and the behavior denotes an interaction between the corresponding user and entity such as befriend or purchase.

The task of behavior prediction aims to exploit historical and predict future user behaviors to (user, entity) pairs. For example, we can predict whether a user will follow a celebrity on a microblogging social network or whether a user will purchase a certain product on an e-commerce platform. Accurate future behavior prediction can assist a company's strategy and policy on advertising, customer service, and even logistics, which is of great importance to both users and service providers.

However, the task of behavior prediction is often associated with at least the following three fundamental challenges:

- *The scarcity challenge.* When the (user, entity, behavior) tuples are few, we might not be able to train a prediction model through current learning and optimization techniques due to the overfitting phenomenon that commonly exists in scarce-data learning problems. Note that the number of the whole set of tuples may be large, but the percentage of the tuples per user or per entity is usually rather small, about 1 percent in the data of the $1 Million Netflix Prize, for example, which makes it difficult to learn a specific user or entity's preferences or characteristics.
- *The uncertainty challenge.* Users' behaviors are usually associated with some levels of uncertainty. For example, we might not be able to infer a user's true preference directly from her implicit examination behaviors such as clicks and browsings. Specifically, a browsing behavior could indicate her positive preference when she plans to buy it or a negative preference when she finds it not interesting after examination. Treating all such implicit examinations without distinction could bias the process of modeling of users' preferences or entities' characteristics.

- *The heterogeneity challenge.* Users' behaviors are usually represented in different forms, including implicit examinations and explicit purchases in an online shopping site, or bidirectional friendships and unidirectional followings in a social networking service. The heterogeneity of users' behaviors requires more sophisticated modeling techniques to fully make use of different types of behaviors. For example, we have to learn both the behavior-dependent and behavior-independent patterns across different sets of behaviors.

Traditional leaning methods for behavior prediction such as the well-known factorization machine[2] usually don't explicitly address the above three challenges.

Transfer Learning for Behavior Prediction

Transfer learning[3] treats users' behaviors in a novel perspective of a finer granularity instead of taking all the user behaviors as a whole. Typically, we have two sets of user behaviors, including target behaviors $\mathcal{T}$ that await prediction and auxiliary behaviors $\mathcal{A}$ that are different but related to target behaviors. For example, we can treat users' followings as target behaviors and friendships as auxiliary behaviors in a social media website, and users' purchases as target behaviors and examinations as auxiliary behaviors in an e-commerce platform. Furthermore, we can bridge two sets of behaviors via sharing some common knowledge $\mathcal{K}$ so as to introduce interactions and improve behavior prediction performance on target behaviors.

Figure 4 illustrates the transfer learning paradigm and two specific examples, from which we can see that the major difference between traditional machine learning and transfer learning lies in the "knowledge transfer" component in the learning paradigm. Designing an appropriate knowledge transfer component is critical to a transfer learning algorithm because it's closely related to the three fundamental questions in transfer learning,[3] that is, what knowledge to transfer, how to transfer it, and when (not) to transfer it. For behavior prediction, previous transfer learning algorithms mainly focus on the first two questions,[4] including transferring knowledge of model parameters, behavior instances, or compressed behavior patterns in an adaptive, collective, or integrative manner.

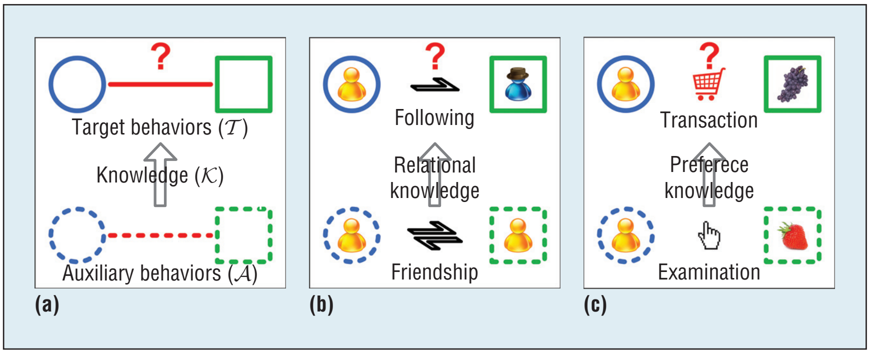

Figure 4. Illustration of transfer learning for behavior prediction, including (a) a generic transfer learning paradigm, (b) an example of relational knowledge transfer from auxiliary bidirectional friendships to target unidirectional followings, and (c) an example of preference knowledge transfer from auxiliary examinations to target purchases.

Knowledge transfer between target and auxiliary behaviors is a potential solution for the aforementioned three challenges. First, for the scarcity challenge, knowledge transfer from additional data provides a way to selectively incorporate auxiliary data to mitigate the data scarcity problem. Second, for the uncertainty challenge, rich interactions between target behaviors and auxiliary behaviors via knowledge sharing are likely to help reduce the uncertainty or learn the confidence of user behaviors. Third, for the heterogeneity challenge, common knowledge sharing is able to integrate different behaviors in a principled way. Finally, the goal of transfer learning for behavior prediction is to achieve better prediction performance than traditional machine learning methods exploiting either target behaviors ($\mathcal{T}$) only, that is, ML($\mathcal{T}$), or both target behaviors ($\mathcal{T}$) and auxiliary behaviors ($\mathcal{A}$), that is, ML($\mathcal{A}$, $\mathcal{T}$):

$$\mathrm{ML}(\mathcal{T}),\ \mathrm{ML}(\mathcal{A}, \mathcal{T}) < \mathrm{TL}(\mathcal{A}, \mathcal{T}). \quad (1)$$

Opportunities

As a new interdisciplinary area of transfer learning and behavior prediction, there are lots of exciting directions ahead such as multiobjective transfer, open domain transfer, lifelong transfer, and transfer learning theories. Multiobjective transfer aims to improve not only the accuracy in behavior prediction but also efficiency, result interpretation, and even robustness against malicious attack. Open domain transfer is the human ability to transfer knowledge from a far domain instead of from a close one as most previous works do, which is usually called "far transfer of learning" in the education community. Lifelong transfer focus on improving the learning and prediction performance in a never-ending manner with little human interference.[5] Transfer learning theories on when to transfer or not haven't been studied much yet, which may be different for different application domains.

Transfer learning is a promising solution to the strategically important task of behavior prediction in various online services. We expect to see many exciting works addressing the fundamental challenges such as scarcity, uncertainty, and heterogeneity, exploring the grand opportunities of multiobjective, open domain, and lifelong transfer learning algorithms and theories. ■

Acknowledgments

We're thankful for the support of Natural Science Foundation of China number 61502307, China National Fundamental Research Project (973 Program) number 2014CB340304, and ITF project ITS/191/13FX.

References

1. R. Zafarani, M. Ali Abbasi, and H. Liu, *Social Media Mining: An Introduction*, Cambridge Univ. Press, 2014.
2. S. Rendle, "Factorization Machines with Libfm," *ACM Trans. Intelligent Systems and Technology*, vol. 3, no. 3, 2012, pp. 57:1–57:22.
3. S. Jialin Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 10, 2010, pp. 1345–1359.
4. W. Pan, E. Xiang, and Q. Yang, "Transfer Learning in Collaborative Filtering with Uncertain Ratings," *Proc. 26th AAAI Conf. Artificial Intelligence*, 2012, pp. 662–668.
5. D.L. Silver, Q. Yang, and L. Li, "Lifelong Machine Learning Systems: Beyond Learning Algorithms," *Proc. AAAI Spring Symp.*, 2013, pp. 49–55.

Weike Pan is a lecturer in the College of Computer Science and Software Engineering at Shenzhen University. Contact him at panweike@szu.edu.cn.

Qiang Yang is a chair professor in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology. Contact him at qyang@cse.ust.hk.