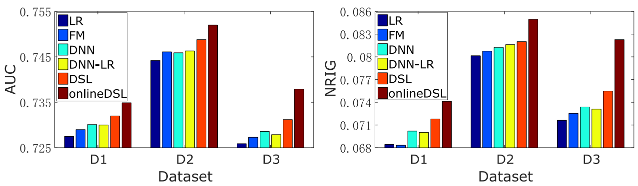
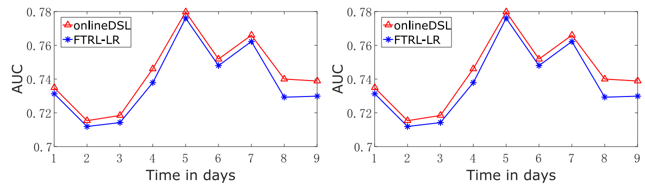


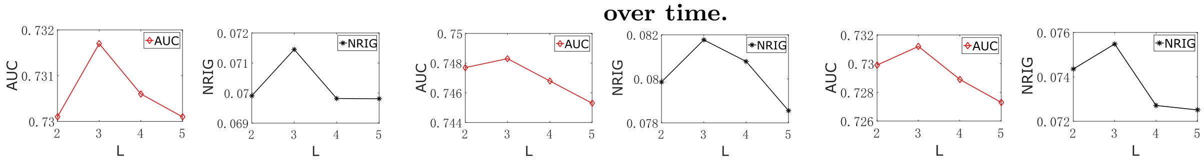
(a) AUC (b) NRIG

Figure 2: Performance comparison.



(a) AUC (b) NRIG

Figure 3: Performance of onlineDSL and FTRL-LR over time.



(a) $D1$ (b) $D2$ (c) $D3$

Figure 4: The influence of L on DSL.

Figure 2 shows the performance results of all the models. It can be observed that DSL and onlineDSL significantly outperform the baselines in all three datasets on NRIG and AUC. Specifically, onlineDSL is much better than DSL because updating online can enhance the model effectiveness. These results suggest that DSL and onlineDSL can more effectively predict CTR by making full use of basic features, pairwise interactions and high-order nonlinear features.

3.3.2 Effectiveness of onlineDSL Over Time. We validate the effectiveness of onlineDSL as time goes on, comparing it with FTRL-LR. FTRL-LR is an online algorithm for CTR prediction by combining FTRL and LR, and has achieved a great success in industry [6]. As DNN in DSL need be trained on training data, we use $Train$ 1, 7-days data (20 May - 26 May) as the training set and the union of $Test$ 1 and $D3$, 9-days data (27 May - 04 June), as the test set. In the test phase, the DNN model in onlineDSL is updated daily.

Figure 3 shows the performance comparison of onlineDSL and FTRL-LR over time. It is easy to find that onlineDSL brings a significant performance improvement on both NRIG and AUC, compared with FTRL-LR. This result indicates the effectiveness of onlineDSL for CTR prediction.

3.3.3 Sensitivity of Hyper-Parameter. According to the setting of experiments, there is a key hyper-parameter having the most influence on the performance of DSL, i.e., the number of hidden layers L.

We conduct an experiment for different L from the set $\{2, 3, 4, 5\}$ and the results are shown in Figure 4. We can find that DSL can get the best performance when $L = 3$ and both AUC and NRIG decrease when $L > 3$ or $L < 3$. We guess a possible reason is that the complexity of the DNN model is appropriate to capture high-order nonlinear features effectively when $L = 3$, but the DNN model easily becomes overfitting when $L > 3$ and underfitting when $L < 3$.

4 CONCLUSION

In this paper, we proposed a novel DSL method for CTR prediction, which could make full use of basic features, pairwise interactions and high-order nonlinear features in ad data. Meanwhile, we developed the onlineDSL algorithm for updating DSL online. Finally, extensive experimental results on real-world ad datasets demonstrated the effectiveness of our DSL method and onlineDSL. We should note that, DSL is actually a general method. In the future, we will test its performance with other deep layer models (e.g., Autoencoder) and shallow layer models (e.g., Support Vector Machine).

5 ACKNOWLEDGEMENTS

This research was partially supported by grants from the National Key Research and Development Program of China (Grant No. 2016YFB1000904), the National Natural Science Foundation of China (Grants No. 61672483 and U1605251), and the Youth Innovation Promotion Association of CAS (No. 2014299).

REFERENCES

[1] O. Chapelle, E. Manavoglu, and R. Rosales. Simple and scalable response prediction for display advertising. *TIST*, 5(4):61, 2015.
[2] J. Chen, B. Sun, H. Li, H. Lu, and X.-S. Hua. Deep ctr prediction in display advertising. In *MM*, pages 811–820. ACM, 2016.
[3] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, et al. Practical lessons from predicting clicks on ads at facebook. In *ADKDD*, pages 1–9. ACM, 2014.
[4] C. Li, Y. Lu, Q. Mei, D. Wang, and S. Pandey. Click-through prediction for advertising in twitter timeline. In *KDD*, pages 1959–1968. ACM, 2015.
[5] Q. Liu, F. Yu, S. Wu, and L. Wang. A convolutional click prediction model. In *CIKM*, pages 1743–1746. ACM, 2015.
[6] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, et al. Ad click prediction: a view from the trenches. In *KDD*, pages 1222–1230. ACM, 2013.
[7] Z. Pan, E. Chen, Q. Liu, T. Xu, H. Ma, and H. Lin. Sparse factorization machines for click-through rate prediction. In *ICDM*, pages 400–409. IEEE, 2016.
[8] S. Rendle. Factorization machines. In *ICDM*, pages 995–1000. IEEE, 2010.
[9] S. Shang, L. Chen, C. S. Jensen, J.-R. Wen, and P. Kalnis. Searching trajectories by regions of interest. *IEEE Transactions on Knowledge and Data Engineering*, 29(7):1549–1562, 2017.
[10] D. Williams and G. Hinton. Learning representations by back-propagating errors. *Nature*, 323(6088):533–538, 1986.
[11] D. Yu and L. Deng. *Automatic speech recognition: A deep learning approach*. Springer, 2014.
[12] Y. Zhang, H. Dai, C. Xu, J. Feng, T. Wang, J. Bian, B. Wang, and T.-Y. Liu. Sequential click prediction for sponsored search with recurrent neural networks. In *AAAI*, 2014.
[13] Z. Zhao, X. He, D. Cai, L. Zhang, W. Ng, and Y. Zhuang. Graph regularized feature selection with data reconstruction. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):689–700, 2016.