类器, 分类器 2 为弱分类器. 在协同训练中由于每轮 2 个分类器所新增的伪标签数据来自另一分类器的预测, 则可能导致弱分类器的过多错误预测对强分类器的性能产生负面影响. 如图 4~图 7 所示, 在协同训练的执行过程中强分类器的精准度会有较大起伏, 这表明分类器性能在训练过程中的不稳定性, 对最终的分类器性能产生较大影响. 而本文方法基于共同判断, 强分类器每轮新增的伪标签数据依然全部来自自身预测, 虽然存在因为与弱分类器的判断结果不同而未选择少量高置信度数据的情况, 但却降低了受到弱分类器影响的可能. 从图中可以看到本文方法中强分类器的精确度曲线虽也有起伏, 但相对平稳, 且整体呈上升趋势. 综上得出, 本文方法相较于协同训练牺牲了一定的性能提升速率得到了更高的稳定性.

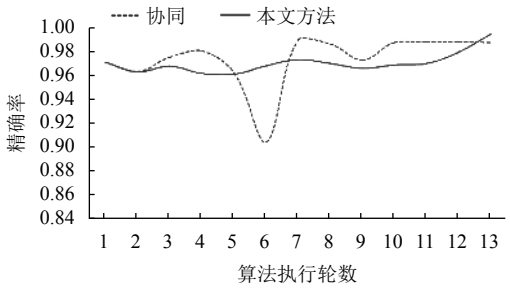

图 4 分类器 1 精准度变化趋势图

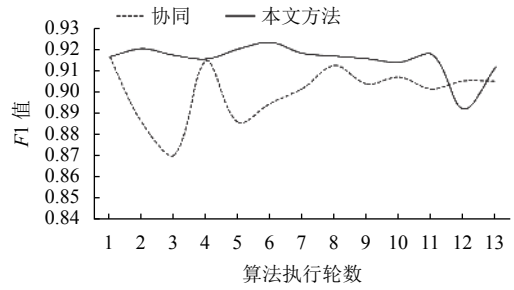

图 5 分类器 1 *F*1 值变化趋势图

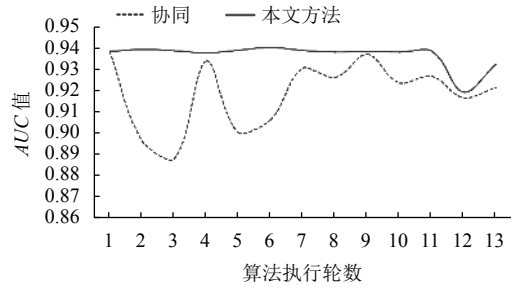

图 6 分类器 1 *AUC* 值变化趋势图

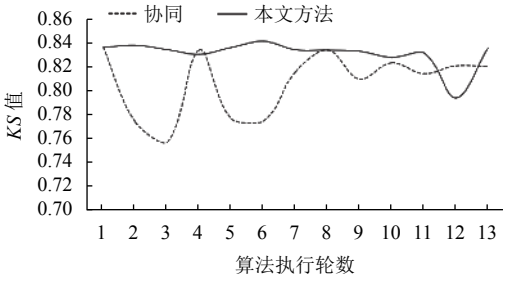

图 7 分类器 1 *KS* 值变化趋势图

4 总结

本文恶意 URL 检测方法结合了特征与文本处理预处理数据, 并对协同训练算法进行了改进, 仅用 0.67% 有标签数据训练出的两个分类器预测精准度分别达到 99.42% 与 95.23%, 低误报率使得该方法训练得到的检测模型有较高的实用性. 这种方式在现实应用中大幅度节约了人为打标签的成本, 减少了时间开销, 且检测效果接近有监督学习得到的分类器, 提供了有效应对新型恶意 URL 的方案. 未来的工作将考虑如何把这种半监督思想应用于恶意 URL 的在线学习中, 在节约开销的同时保证检测模型的定时更新.

参考文献

1 Kaspersky Security Bulletin 2018. https://securelist.com/kaspersky-security-bulletin-2018-statistics/89145/. (2018-09-04).
2 Sahoo D, Liu C, Hoi SCH. Malicious URL detection using machine learning: A survey. arXiv preprint arXiv: 1701.07179, 2017.
3 Prakash P, Kumar M, Kompella RR, *et al*. PhishNet: Predictive blacklisting to detect phishing attacks. 2010 Proceedings IEEE INFOCOM. San Diego, CA, USA. 2010. 1–5.
4 Tsai CF, Hsu YF, Lin CY, *et al*. Intrusion detection by machine learning: A review. Expert Systems with Applications, 2009, 36(10): 11994–12000. [doi: 10.1016/j.eswa.2009.05.029]
5 Le H, Pham Q, Sahoo D, *et al*. URLNet: Learning a URL representation with deep learning for malicious URL detection. arXiv preprint arXiv: 1802.03162, 2018.
6 Sommer R, Paxson V. Outside the closed world: On using machine learning for network intrusion detection. Proceedings of 2010 IEEE Symposium on Security and Privacy. Berkeley/Oakland, CA, USA. 2010. 305–316.
7 Sinclair C, Pierce L, Matzner S. An application of machine

learning to network intrusion detection. Proceedings 15th Annual Computer Security Applications Conference. Phoenix, AZ, USA. 1999. 371–377.

8 Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 2016, 18(2): 1153–1176.

9 吴海滨, 张冬梅. 基于上下文信息的恶意 URL 检测技术. 软件, 2019, 40(1): 63–68. [doi: 10.3969/j.issn.1003-6970.2019.01.013]

10 沙泓州, 刘庆云, 柳厅文, 等. 恶意网页识别研究综述. 计算机学报, 2016, 39(3): 529–542. [doi: 10.11897/SP.J.1016.2016.00529]

11 Warrender C, Forrest S, Pearlmutter B. Detecting intrusions using system calls: Alternative data models. Proceedings of the 1999 IEEE Symposium on Security and Privacy. Oakland, CA, USA. 1999. 133–145.

12 Mao GJ, Wu XD, Zhu XQ, *et al*. Mining maximal frequent itemsets from data streams. Journal of Information Science, 2007, 33(3): 251–262. [doi: 10.1177/0165551506068179]

13 Garera S, Provos N, Chew M, *et al*. A framework for detection and measurement of phishing attacks. Proceedings of the 2007 ACM workshop on Recurring Malcode. Alexandria, VA, USA. 2007. 1–8.

14 Sinha S, Bailey M, Jahanian F. Shades of grey: On the effectiveness of reputation-based “blacklists”. Proceedings of 2008 3rd International Conference on Malicious and Unwanted Software. Fairfax, VA, USA. 2008. 57–64.

15 Xu L, Zhan ZX, Xu SH, *et al*. Cross-layer detection of malicious websites. Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy. San Antonio, TX, USA. 2013. 141–152.

16 Huang HJ, Qian L, Wang YJ. A SVM-based technique to detect phishing URLs. Information Technology Journal, 2012, 11(7): 921–925. [doi: 10.3923/itj.2012.921.925]

17 Hou YT, Chang YM, Chen T, *et al*. Malicious web content detection by machine learning. Expert Systems with Applications, 2010, 37(1): 55–60. [doi: 10.1016/j.eswa.2009.05.023]

18 Canali D, Cova M, Vigna G, *et al*. Prophiler: A fast filter for the large-scale detection of malicious web pages. Proceedings of the 20th International Conference on World Wide Web. Hyderabad, India. 2011. 197–206.

19 Lee S, Kim J. WarningBird: Detecting suspicious URLs in Twitter Stream. NDSS. 2012. 1–13.

20 Zhou ZH, Li M. Semi-supervised regression with co-training. Proceedings of the 19th International Joint Conference on Artificial Intelligence. San Francisco, CA, USA. 2005. 908−913.

21 Zhou ZH, Li M. Semisupervised regression with cotraining-style algorithms. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(11): 1479–1493. [doi: 10.1109/TKDE.2007.190644]

22 梁吉业, 高嘉伟, 常瑜. 半监督学习研究进展. 山西大学学报 (自然科学版), 2009, 32(4): 528–534.

23 周志华. 基于分歧的半监督学习. 自动化学报, 2013, 39(11): 1871–1878.

24 McClosky D, Charniak E, Johnson M. Effective self-training for parsing. Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg, PA, USA. 2006. 152–159.

25 Rosenberg C, Hebert M, Schneiderman H. Semi-supervised self-training of object detection models. Proceedings of 2005 7th IEEE Workshops on Applications of Computer Vision. Breckenridge, CO, USA. 2005. 29–36.

26 Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. Proceedings of the 11th Annual Conference on Computational Learning Theory. New York, NY, USA. 1998. 92−100.

27 Nigam K, Ghani R. Analyzing the effectiveness and applicability of co-training. Proceedings of the 9th International Conference on Information and Knowledge Management. New York, NY, USA. 2000. 86–93.

28 Zhou ZH. Disagreement-based semi-supervised learning. Acta Automatica Sinica, 2013, 39(11): 1871–1878. [doi: 10.3724/SP.J.1004.2013.01871]

29 Sindhwani V, Niyogi P, Belkin M. A co-regularized approach to semi-supervised learning with multiple views. Proceedings of the 22nd Workshop on Learning with Multiple Views. Cambridge, UK. 2005. 824−831.

30 Brefeld U, Gärtner T, Scheffer T, *et al*. Efficient co-regularised least squares regression. Proceedings of the 23rd International Conference on Machine Learning. New York, NY, USA. 2006. 137−144.

31 Farquhar JDR, Hardoon DR, Meng HY, *et al*. Two view learning: SVM-2K, theory and practice. Proceedings of the 18th International Conference on Neural Information Processing Systems. Cambridge, UK. 2005. 355−362.

32 Sridharan K, Kakade SM. An information theoretic framework for multi-view learning. Proceedings of the 21st Annual Conference on Learning Theory. Helsinki, Finland.

2008. 403−414.

33 Goldman SA, Zhou Y. Enhancing supervised learning with unlabeled data. Proceedings of the 17th International Conference on Machine Learning. San Francisco, CA, USA. 2000. 327−334.

34 Fushiki T. Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 2011, 21(2): 137–146. [doi: 10.1007/s11222-009-9153-8]

35 Zhou ZH, Li M. Tri-training: Exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(11): 1529. [doi: 10.1109/TKDE.2005.186]

36 Wang W, Zhou ZH. Co-training with insufficient views. Proceedings of the 5th Asian Conference on Machine Learning. Canberra, Australia. 2013. 467−482.

37 Ma J, Saul LK, Savage S, *et al*. Learning to detect malicious URLs. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 30.

38 徐冬冬, 谢统义, 万卓昊, 等. 基于 TF-IDF 文本向量化的 SQL 注入攻击检测. 广西大学学报 (自然科学版), 2018, 43(5): 1818–1826.

39 Bengio Y, Ducharme R, Vincent P, *et al*. A neural probabilistic language model. The Journal of Machine Learning Research, 2003, 3: 1137–1155.

40 Collobert R, Weston J, Bottou L, *et al*. Natural language processing (almost) from scratch. The Journal of Machine Learning Research, 2011, 12: 2493–2537.

41 Le Q, Mikolov T. Distributed representations of sentences and documents. Proceedings of the 31st International Conference on International Conference on Machine Learning. Beijing, China. 2014. 1188–1196.

42 Wallach HM. Topic modeling: Beyond bag-of-words. Proceedings of the 23rd International Conference on Machine Learning. New York, NY, USA. 2006. 977–984.

43 Mikolov T, Chen K, Corrado G, *et al*. Efficient estimation of word representations in vector space. Proceedings of the 1st International Conference on Learning Representations. Scottsdale, AZ, USA. 2013.

44 Mikolov T, Sutskever I, Chen K, *et al*. Distributed representations of words and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY, USA. 2013. 3111–3119.