Probabilistic Automated Language Learning for Configuration Files

Mark Santolucito Ennan Zhai Ruzica Piskac

Yale University

Abstract. We verify configuration files. While previous attempts to learn a model for these highly unstructured files have focused on machine learning, we instead take a probabilistic logic approach. Probabilistic logic allows us to provide justification for our learning results, giving proof of the verification task. Because the resulting model has a clear logical structure, as opposed traditional machine learning techniques such as neural nets, we can do further analysis on the model that improves our results.

1 Introduction

Our main contributions are as follows:

1. Using a probabilistic logical inference approach to the learning task, which builds a more informative structure than state of the art machine learning algorithms
2. A basic application of graph analysis on the learned model to improve results in a way not available to traditional machine learning approaches
3. An open-source implementation, ConfigV, of this approach applied to a set of 256 real world benchmarks. ConfigV identify XX new misconfigurations errors on YY of these files.

Acknowledgements

We thank the anonymous reviewers for their insightful comments. We also thank Tianyin Xu for his valuable feedback on earlier version of this work. This research was supported by the NSF under grant CCF-1302327.

References

1. Fine-grained value correlation error, http://serverfault.com/questions/628414/my-cnf-configuration-in-mysql-5-6-x
2. Misconfiguration dataset, https://github.com/tianyin/configuration_datasets
3. OpenStack, http://www.openstack.org/
4. parallel-3.2.1.0: Parallel programming library, https://hackage.haskell.org/package/parallel-3.2.1.0/docs/Control-Parallel-Strategies.html
5. Problem moving a MySQL data directory, http://serverfault.com/questions/281217/problem-moving-a-mysql-data-directory-to-a-new-drive

6. Singular value error, http://stackoverflow.com/questions/1980004/2006-mysql-server-has-gone-away-error-in-wamp
7. Agrawal, R., Evfimievski, A.V., Srikant, R.: Information sharing across private databases. In: ACM SIGMOD (Jun 2003)
8. Aguilera, M.K., Mogul, J.C., Wiener, J.L., Reynolds, P., Muthitacharoen, A.: Performance debugging for distributed systems of black boxes. In: 19th ACM Symposium on Operating Systems Principles (SOSP) (Oct 2003)
9. Attariyan, M., Chow, M., Flinn, J.: X-ray: Automating root-cause diagnosis of performance anomalies in production software. In: 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Oct 2012)
10. Attariyan, M., Flinn, J.: Automating configuration troubleshooting with dynamic information flow analysis. In: 9th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Oct 2010)
11. Bahl, P., Chandra, R., Greenberg, A.G., Kandula, S., Maltz, D.A., Zhang, M.: Towards highly reliable enterprise network services via inference of multi-level dependencies. In: ACM SIGCOMM (SIGCOMM) (Aug 2007)
12. Barham, P., Donnelly, A., Isaacs, R., Mortier, R.: Using Magpie for request extraction and workload modelling. In: 6th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Dec 2004)
13. Basescu, C., Cachin, C., Eyal, I., Haas, R., Sorniotti, A., Vukolic, M., Zachevsky, I.: Robust data sharing with key-value stores. In: 42nd IEEE/IFIP International Conference on Dependable Systems and Networks (DSN) (Jun 2012)
14. Benson, T., Akella, A., Maltz, D.A.: Network traffic characteristics of data centers in the wild. In: Internet Measurement Conference (IMC) (Nov 2010)
15. Bessani, A.N., Correia, M.P., Quaresma, B., André, F., Sousa, P.: DepSky: Dependable and secure storage in a cloud-of-clouds. In: 6th ACM European Conference on Computer Systems (EuroSys) (Apr 2011)
16. Blundo, C., de Cristofaro, E., Gasti, P.: EsPRESSo: Efficient privacy-preserving evaluation of sample set similarity. vol. 22, pp. 355–381 (2014)
17. Bobot, F., Filliâtre, J., Marché, C., Paskevich, A.: Let's verify this with why3. STTT 17(6), 709–727 (2015)
18. Bonvin, N., Papaioannou, T.G., Aberer, K.: A self-organized, fault-tolerant and scalable replication scheme for cloud storage. In: ACM Symposium on Cloud Computing (SoCC) (Jun 2010)
19. Broder, A.Z.: On the resemblance and containment of documents. In: Compression and Complexity of Sequences (SEQUENCES) (Jun 1997)
20. Chen, M.Y., Accardi, A., Kiciman, E., Patterson, D.A., Fox, A., Brewer, E.A.: Path-based failure and evolution management. In: 1st USENIX Symposium on Networked System Design and Implementation (NSDI) (Mar 2004)
21. Chen, X., Mao, Y., Mao, Z.M., van der Merwe, J.E.: Declarative configuration management for complex and dynamic networks. In: ACM CoNEXT (CoNEXT) (Nov 2010)
22. Chen, X., Zhang, M., Mao, Z.M., Bahl, P.: Automating network application dependency discovery: Experiences, limitations, and new solutions. In: 8th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Dec 2008)
23. Dobrescu, M., Argyraki, K.J.: Software dataplane verification. In: 11th USENIX Symposium on Networked System Design and Implementation (NSDI) (Apr 2014)
24. Enck, W., McDaniel, P.D., Sen, S., Sebos, P., Spoerel, S., Greenberg, A.G., Rao, S.G., Aiello, W.: Configuration management at massive scale: System design and experience. In: USENIX Annual Technical Conference (USENIX ATC) (Jun 2007)

25. Huang, P., Bolosky, W.J., Singh, A., Zhou, Y.: Confvalley: A systematic configuration validation framework for cloud services. In: 10th European Conference on Computer Systems (EuroSys) (Apr 2015)
26. Kissner, L., Song, D.X.: Privacy-preserving set operations. In: 25th Annual International Cryptology Conference (CRYPTO). Springer (Aug 2005)
27. Launchbury, J., Jones, S.L.P.: Lazy functional state threads. In: Programming Language Design and Implementation (PLDI). pp. 24–35. ACM Press (1993)
28. Leino, K.R.M.: Dafny: An automatic program verifier for functional correctness. In: Logic for Programming, Artificial Intelligence, and Reasoning - 16th International Conference, LPAR-16. pp. 348–370 (2010)
29. Loo, B.T., Hellerstein, J.M., Stoica, I., Ramakrishnan, R.: Declarative routing: Extensible routing with declarative queries. In: ACM SIGCOMM (SIGCOMM) (Aug 2005)
30. Lopes, N.P., Bjørner, N., Godefroid, P., Jayaraman, K., Varghese, G.: Checking beliefs in dynamic networks. In: 12th USENIX Symposium on Networked System Design and Implementation (NSDI) (May 2015)
31. Piskac, R., Wies, T., Zufferey, D.: Grasshopper - complete heap verification with mixed specifications. In: Tools and Algorithms for the Construction and Analysis of Systems - 20th International Conference, TACAS 2014. pp. 124–139 (2014)
32. Raychev, V., Bielik, P., Vechev, M.T., Krause, A.: Learning programs from noisy data. In: 43rd ACM SIGPLAN-SIGACT (POPL) Symposium on Principles of Programming Languages (Jan 2016)
33. Raychev, V., Vechev, M.T., Krause, A.: Predicting program properties from "big code". In: 42nd ACM SIGPLAN-SIGACT (POPL) Symposium on Principles of Programming Languages (Jan 2015)
34. Santolucito, M., Zhai, E., Piskac, R.: Probabilistic automated language learning for configuration files. In: 28th Computer Aided Verification (CAV) (Jul 2016)
35. Su, Y., Attariyan, M., Flinn, J.: AutoBash: Improving configuration management with operating systems. In: 21st ACM Symposium on Operating Systems Principles (SOSP) (Oct 2007)
36. Wang, H.J., Platt, J.C., Chen, Y., Zhang, R., Wang, Y.: Automatic misconfiguration troubleshooting with PeerPressure. In: 6th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Dec 2004)
37. Whitaker, A., Cox, R.S., Gribble, S.D.: Configuration debugging as search: Finding the needle in the haystack. In: 6th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Dec 2004)
38. Xu, T., Jin, L., Fan, X., Zhou, Y., Pasupathy, S., Talwadker, R.: Key, you have given me too many knobs!: understanding and dealing with over-designed configuration in system software. In: 10th Joint Meeting on Foundations of Software Engineering (ESEC/FSE) (Aug 2015)
39. Xu, T., Zhang, J., Huang, P., Zheng, J., Sheng, T., Yuan, D., Zhou, Y., Pasupathy, S.: Do not blame users for misconfigurations. In: 24th ACM Symposium on Operating Systems Principles (SOSP) (Nov 2013)
40. Xu, T., Zhou, Y.: Systems approaches to tackling configuration errors: A survey. ACM Comput. Surv. 47(4), 70 (2015)
41. Yin, Z., Ma, X., Zheng, J., Zhou, Y., Bairavasundaram, L.N., Pasupathy, S.: An empirical study on configuration errors in commercial and open source systems. In: 23rd ACM Symposium on Operating Systems Principles (SOSP) (Oct 2011)
42. Yuan, D., Xie, Y., Panigrahy, R., Yang, J., Verbowski, C., Kumar, A.: Context-based online configuration-error detection. In: USENIX Annual Technical Conference (USENIX ATC) (Jun 2011)

43. Zhai, E., Chen, R., Wolinsky, D.I., Ford, B.: Heading off correlated failures through Independence-as-a-service. In: 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (Oct 2014)
44. Zhang, J., Renganarayana, L., Zhang, X., Ge, N., Bala, V., Xu, T., Zhou, Y.: Encore: Exploiting system environment and correlation information for misconfiguration detection. In: Architectural Support for Programming Languages and Operating Systems (ASPLOS) (Mar 2014)