Biophysical Reviews (2019) 11:123–125
https://doi.org/10.1007/s12551-018-0492-6

REVIEW

Towards testing big data analytics software: the essential role of metamorphic testing

Zhiyi Zhang[1] · Xiaoyuan Xie[1]

Received: 26 November 2018 / Accepted: 6 December 2018 / Published online: 18 December 2018

Abstract

In the rapidly growing field of big data analysis, scientists from numerous domains such as computer science and biology are constantly challenged by an unprecedented amount of data. While many software programs have been constructed to support processing and analyzing continuous information flow, one under-appreciated challenge in this field is software quality assurance of these big data software platforms. Metamorphic testing, which was proposed to alleviate the oracle problem in the software engineering community, has become an effective approach for software verification and validation. Recent years, we have witnessed successful applications of metamorphic testing in a variety of domains, ranging from bioinformatics to deep learning. In this letter, we review some main applications of metamorphic testing on big data and present visions for the challenges in future research.

Keywords Software engineering · Metamorphic testing · Big data software

Introduction

Big data software, which is described as "software that supports the time-constrained processing of the continuous information flows to provide actionable intelligence" (Otero and Peter 2015), has raised new and unexplored challenges in terms of software quality assurance in not only software engineering community but also other communities such as biology and physics. For example, small biology laboratories can become big data generators where many types of information such as genetics sequences and medical records are produced. Biologists need to store, process and interpret such massive data streams. Obviously, it is infeasible to manually perform these tasks. Therefore various software tools are adopted to assist big data analysis (Marx 2013). As a consequence, the quality of such software tools becomes very important, since their outputs may significantly influence the conclusions in these data-rich disciplines.

However, testing and identifying faults in big data analytics software are challenging because such programs suffer from the well-known "oracle problem" (Weyuker 1982), where the expected output is hard to derive. Moreover, there seems to be a mismatch between the theoretical performance of a machine learning algorithm and their real-life performance (Otero and Peter 2015), which further suggest there may be deviation between the algorithm and the implementation of these algorithms in software programs. Thus, it is crucial to test the implementations, to make sure that the programs deliver reliable and expected results.

Metamorphic testing (MT) (Chen et al. 1998) is a promising technique that provides an alternative to alleviate this problem. In MT, failures are revealed by checking expected relations (known as metamorphic relations) among multiple executions of the program under test. For example, consider a program S that calculates the sin function. We can define a metamorphic relation (MR) in MT: "If $y = \pi - x$, then $sin(x) = sin(y)$", based on the mathematical property $sin(x) = sin(\pi - x)$. In this case, even the correct and precise value of $sin(x)$ is unknown, if the two execution results (one with input x and the other with input y) are different so that the above MR is violated, we can conclude that S is faulty.

Recent years have seen rapid growth in the applications of MT in various domains ranging from biology to deep

This article is part of a Special Issue on 'Big Data' edited by Joshua WK Ho and Eleni Giannoulatou.

✉ Xiaoyuan Xie
xxie@whu.edu.cn

[1] School of Computer Science, Wuhan University, Wuhan, China

learning, in both academia and industry. We refer the readers to recent surveys (Segura et al. 2016; Chen et al. 1998) for a thorough introduction. In this letter, we will briefly discuss some main applications of MT on big data software, and present insights regarding open challenges in MT, as well as the opportunities in testing big data software.

Applications of metamorphic testing

Bioinformatics Chen et al. (2009) first introduced MT to bioinformatics program testing, where two open-source bioinformatics programs, namely GNLab (for gene regulatory networks simulations) and SeqMap (for short sequence mapping) are tested. A fault was effectively found in GNLab, and the root cause was the misspecification of algorithm. MT was also applied to alleviate the oracle problem in testing phylogenetic inference programs which are available to infer evolutionary relationships among taxa using DNA or amino acids, and to infer the evolutionary relationships among species (Sadi et al. 2011). Pullum and Ozmen (2012) applied MT on validating two disease spread models that mimic the 1918 flu pandemic. Their work demonstrated the effectiveness of MR in highlighting issues of discrepancies between expected and actual outputs.

Artificial intelligence (AI) As early as a decade ago, Murphy et al. (2008) applied MT to several machine learning applications (e.g., MartiRank), and categorized different types of metamorphic properties to provide a guideline for conducting MT in machine learning. Xie et al. (2011) conducted MT on testing supervised machine learning software, namely Weka. Two classifiers, k-nearest neighbor (KNN) and Naive Bayes were investigated where several faults were successfully revealed by a series of MRs. Further investigation revealed that violations to the MRs may indicate that these classifiers may be unsuitable for some real-life applications, even if the algorithms are correctly implemented. This interesting result suggests that MT is not only useful for software verification, but also useful for software validation.

MT has been recently applied to validate a deep learning framework for automatically classifying biological images of the cellular level which involves a convolutional neural network and a massive image dataset (Ding et al. 2017). The MT-guided validation approach was demonstrated to be effective at checking the quality of data set, network architecture, and execution environment of deep learning framework. Other recent works (Tian et al. 2018; Zhang et al. 2018) have also been investigated to validate autonomous driving systems where MRs were leveraged to automatically generate test cases which reflect real-world scenes such as rain or fog. Thousands of erroneous behaviors were found. These works have demonstrated the feasibility of MT-based validation approach for checking the quality of AI-driven systems.

Challenges and opportunities

Challenges in identifying metamorphic relations Identifying proper metamorphic relations in MT requires great effort. The field still lacks adequate guidelines for the construction of effective and good metamorphic relations. The guidelines should be provided for specialists, and especially nonspecialists who may lack domain knowledge or practical experiences. Besides, it is crucial to investigate the capability of different metamorphic relations at revealing faults. More systematic approaches for constructing and selecting good MRs should be proposed.

Opportunities in testing big data Big data has been defined as the data with high volume, velocity and variety (3V), and unpredictability (Otero and Peter 2015). By its very nature, traditional software testing technique may no longer suffice. MT has been demonstrated to be an effective approach for testing big data software. One of the open area of research in testing big data software via MT is the construction of domain-specific MRs, as pointed out in Otero and Peter (2015). Identifying these MRs require deep understanding of machine learning algorithms and software testing.

Due to the rapid growth of data, testing big data system at run-time seems to be very difficult (Chen et al. 2018). Data samples that reflect the characteristics of actual datasets are needed. Furthermore, investigation into the extent to which sample source and follow-up data reflect real datasets is of interest. That is, in addition to relying on only necessary properties to drive MRs, it is also useful to consider dynamics properties of the data itself.

Funding information This work is supported by National Key R&D Program of China (2018YFB1003901), and the National Natural Science Foundation of China (61572375, 61772263).

Compliance with ethical standards

Conflict of interest Zhiyi Zhang declares that she has no conflict of interest. Xiaoyuan Xie declares that she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Chen TY, Cheung SC, Yiu SM (1998) Metamorphic testing: a new approach for generating next test cases. Tech. rep., Technical Report HKUST-CS98-01, Department of Computer Science. Hong Kong University of Science and Technology, Hong Kong

Chen TY, Ho JW, Liu H, Xie X (2009) An innovative approach for testing bioinformatics programs using metamorphic testing. BMC Bioinf 10(1):24. https://doi.org/10.1186/1471-2105-10-24

Chen TY, Kuo FC, Liu H, Poon PL, Towey D, Tse T, Zhou ZQ (2018) Metamorphic testing: a review of challenges and opportunities. ACM Comput Surv 51(1):4:1–4:27

Ding J, Kang X, Hu X (2017) Validating a deep learning framework by metamorphic testing. In: Proceedings of the 2nd international workshop on metamorphic testing. IEEE, pp 28–34. https://doi.org/10.1109/MET.2017.2

Marx V (2013) Biology: the big challenges of big data. Nature 498:255–260

Murphy C, Kaiser G, Hu L, Wu L (2008) Properties of machine learning applications for use in metamorphic testing. In: Proceedings of the 20th international conference on software engineering and knowledge engineering, pp 867–872

Otero CE, Peter A (2015) Research directions for engineering big data analytics software. IEEE Intell Syst 30(1):13–19. https://doi.org/10.1109/MIS.2014.76

Pullum LL, Ozmen O (2012) Early results from metamorphic testing of epidemiological models. In: ASE/IEEE International Conference on BioMedical Computing (BioMedCom)(BIOMEDCOM), pp 62–67. https://doi.org/10.1109/BioMedCom.2012.17

Sadi MS, Kuo FC, Ho JWK, Charleston MA, Chen TY (2011) Verification of phylogenetic inference programs using metamorphic testing. J Bioinform Comput Biol 9(6):729–747

Segura S, Fraser G, Sanchez AB, Ruiz-Cortés A (2016) A survey on metamorphic testing. IEEE Trans Softw Eng 42(9):805–824. https://doi.org/10.1109/TSE.2016.2532875

Tian Y, Pei K, Jana S, Ray B (2018) Deeptest: automated testing of deep-neural-network-driven autonomous cars. In: Proceedings of the 40th international conference on software engineering. ACM, pp 303–314, https://doi.org/10.1145/3180155.3180220

Weyuker EJ (1982) On testing non-testable programs. Comput J 25(4):465–470

Xie X, Ho JW, Murphy C, Kaiser G, Xu B, Chen TY (2011) Testing and validating machine learning classifiers by metamorphic testing. J Syst Softw 84(4):544–558

Zhang M, Zhang Y, Zhang L, Liu C, Khurshid S (2018) Deeproad: Gan-based metamorphic testing and input validation framework for autonomous driving systems. In: Proceedings of the 33rd ACM/IEEE international conference on automated software engineering. ACM, pp 132–142, https://doi.org/10.1145/3238147.3238187