New machine learning method could help keep cloud services from crashing

PeiYun Zhang
Sheng Shu
MengChu Zhou

Video Abstract

Keywords: Journal of Automatica Sinica, cloud computing, fault detection, support vector machine (SVM), grid, internet technology (IT), cloud fault, deadlock, virtual machine (VM), rule-based detection, decision tree, statistics-based detection, neural network, back propagation (BP), learning vector quantization (LVQ), classification, constraint programming, optimal hyperplane, principal component analysis (PCA), extreme learning machine (ELM)

Posted Date: September 20th, 2019

DOI: <https://doi.org/10.21203/rs.2.15053/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

For about 40 minutes in 2013, the world was in panic. Cloud services for internet retail giant Amazon had crashed. In less than an hour, Amazon had lost an estimated 5 million dollars in sales. It's an operational standstill feared by cloud users and operators alike. Unfortunately, most methods for detecting the faults that lead to such crashes tend to be inefficient and inaccurate. But now, a team of researchers from China, Saudi Arabia, and the US has developed one of the best detection methods yet—based on a machine learning tool known as a support vector machine. A support vector machine is a classification-based learning algorithm. For the simple task of classifying circles by color, the algorithm takes a set of examples and determines the dividing line or plane that maximizes the separation between the two classes. That provides the widest margin of classification error, preventing any new circles from being mislabeled. In cloud computing, the properties of items to be classified are much more complex than color. When deciding whether a certain computational task is either normal or likely a crash waiting to happen, a support vector machine has to deal with a whole list of properties, such as CPU usage rate, maximum memory usage, and mean local disk space used, just to name a few. It's not impossible. But the process can be computationally taxing and slow. To boost efficiency, the research team developed a two-level strategy for detecting faults. At the coarse level, a support vector machine is carried out as usual. But at the finer level, the algorithm zooms in close to the plane that divides normal from abnormal cases and ensures that there is no misclassification. On top of that, the growing list of correctly classified cases that teaches the algorithm to identify faults is continuously trimmed. Any new case deemed too similar to one already in the list is rejected. Keeping the list lean helps reduce the amount of time the algorithm spends on predicting faults. The new strategy outperforms other popular methods for detecting faults in terms of both accuracy and time cost, including the state-of-the-art extreme learning machine. Suped up as it is, however, the support vector machine can't pinpoint the reason faults occur in the first place. But with further improvements, the approach might soon help keep clouds of all types running smoothly.