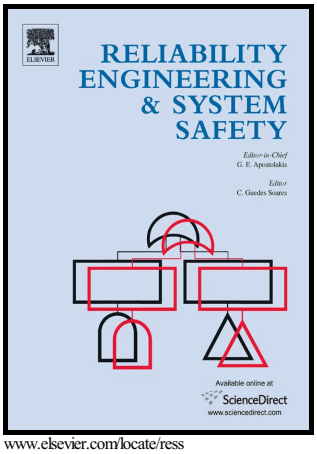
Author's Accepted Manuscript

Hard Drive Failure Prediction using Decision Trees

Jing Li, Rebecca J. Stones, Gang Wang, Xiaoguang Liu, Zhongwei Li, Ming Xu



www.elsevier.com/locate/ress

PII: S0951-8320(16)30156-9
DOI: http://dx.doi.org/10.1016/j.ress.2017.03.004
Reference: RESS5764

To appear in: *Reliability Engineering and System Safety*

Received date: 16 June 2016
Revised date: 17 February 2017
Accepted date: 6 March 2017

Cite this article as: Jing Li, Rebecca J. Stones, Gang Wang, Xiaoguang Liu Zhongwei Li and Ming Xu, Hard Drive Failure Prediction using Decision Trees *Reliability Engineering and System Safety* http://dx.doi.org/10.1016/j.ress.2017.03.004

This is a PDF file of an unedited manuscript that has been accepted fo publication. As a service to our customers we are providing this early version o the manuscript. The manuscript will undergo copyediting, typesetting, an review of the resulting galley proof before it is published in its final citable form Please note that during the production process errors may be discovered whic could affect the content, and all legal disclaimers that apply to the journal pertain

ACCEPTED MANUSCRIPT

Hard Drive Failure Prediction using Decision Trees

Jing Li[a,b], Rebecca J. Stones[b], Gang Wang[b], Xiaoguang Liu[b], Zhongwei Li[b], Ming Xu[c]

[a]*College of Computer Science and Technology, Civil Aviation University of China, Tianjin, China*
[b]*Nankai-Baidu Joint Lab, College of Computer and Control Engineering, Nankai University, Tianjin, China*
[c]*Qihoo 360 Technology Company, Beijing, China*

Accepted Manuscript

Abstract

This paper proposes two hard drive failure prediction models based on Decision Trees (DTs) and Gradient Boosted Regression Trees (GBRTs) which perform well in prediction performance as well as stability and interpretability. The models are evaluated on a real-world dataset containing 121,698 drives in total. Experimental results show the DT model predicts over 93% of failures at a false alarm rate under 0.01%, and the GBRT model can achieve about 90% failure detection rate without any false alarms. Moreover, the GBRT model evaluates drive health (or fault probability) which provides a quantitative indicator of failure urgency. This enables operators to allocate system resources accordingly for pre-warning migrations while maintaining the quality of user services.

Aiming at practical application of prediction models, we test the models on another real-world dataset with different drive models, on a real-world hybrid dataset with multiple drive models, and on several datasets containing fewer drives. Both prediction models show steady prediction performance, with high failure detection rates (80% to 96%) and low false alarm rates (0.006% to 0.31%). We also implement a reliability model for RAID-6 systems with proactive fault tolerance and show that the proposed models can significantly improve the reliability and/or reduce construction and maintenance cost of large-scale storage systems.

Keywords:
Hard drive failure prediction; SMART; Decision Tree; Health degree

1. Introduction

Hard drives are one of the most common devices in storage systems, and most of the information in the world is being stored on hard drives [1]. While the failure

Download English Version:

https://daneshyari.com/en/article/5019447

Download Persian Version:

https://daneshyari.com/article/5019447

Daneshyari.com