Proactive Storage Health Management to Reduce Data Center Downtime

#460

Guanghua Shu and Ji Eun Jang

{gshu, jiejang}@stanford.edu

Motivation and Background

Motivation

- IT equipment failure is the most costly reason for data center downtime and storage components have the most frequent failing rate in current IT environments. Therefore, **accurate failure prediction and timely replacement of disk drives** can decrease downtime costs and improve system reliability

S.M.A.R.T

- Self-monitoring, analysis and reporting technology (SMART) is a monitoring system for disk drives, which collects and reports on various attributes related to drive reliability
- Backblaze published SMART datasets collected from data center [1]

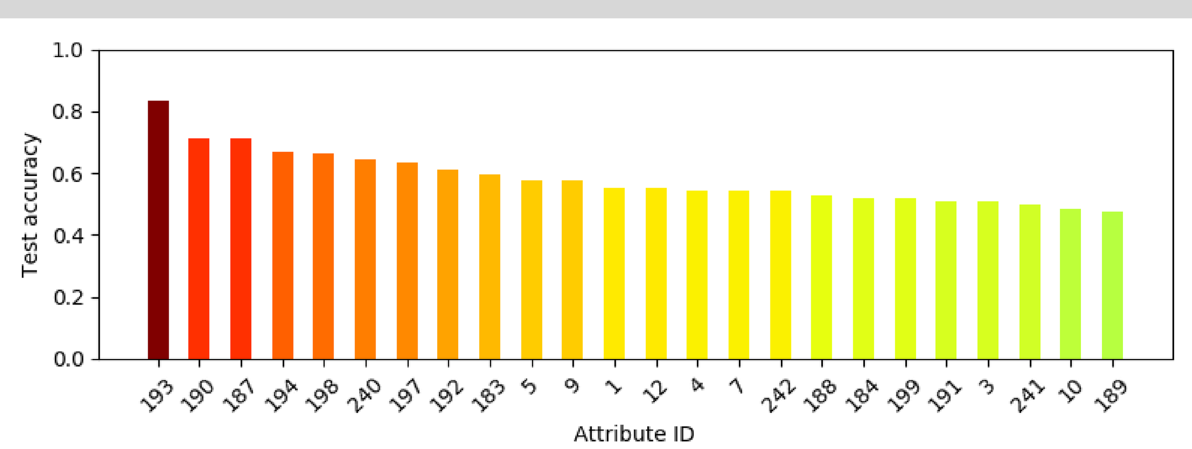
ID	Name	Attribute Description
5	Reallocated Sectors Count	Count of reallocated sectors. a count of the bad sectors that have been found and remapped
187	Reported Uncorrectable Errors	The count of errors that could not be recovered using hardware ECC
193	Load Cycle Count	Count of load/unload cycles into head landing zone position
197	Current Pending Sector Count	Count of unstable sectors. If an unstable sector is subsequently read successfully, the sector is remapped and this value is decreased

Feature Selection

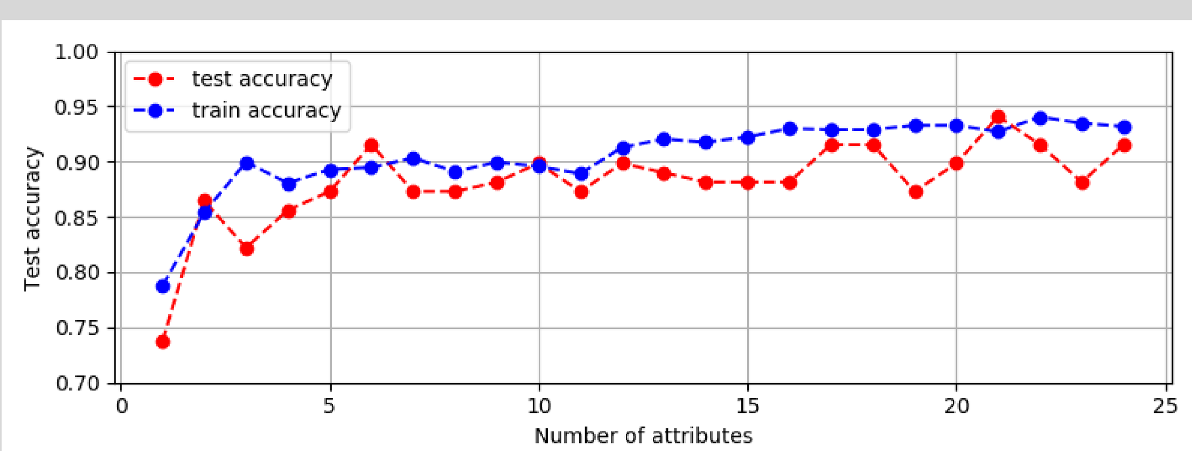
Feature Selection is Necessary

- SMART standard defines more than 90 attributes
- Even with standardization, each vendor has different attribute definitions
- Different drive models have different failure mechanisms, and hence unique sets of effective attributes

Importance of Each Features



Test Accuracy vs The number of Features



[Most recent results with data normalization]

S.M.A.R.T Preprocessing

Time-Series Averaging [2]

- Reduce noise due to measurement and disk recovery mechanism
- Obtain compact time-series representation for SVM and XGBoost
- Averaging with exponentially decreasing weights

$$s_t = \alpha \times y_t + (1-\alpha) \times s_{t-1}$$

$$\alpha = \frac{2}{\tau + 1}$$

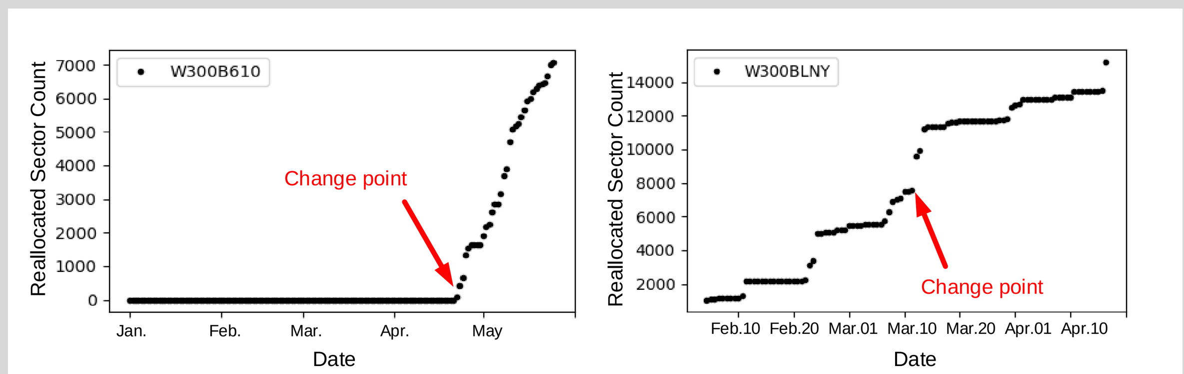
Change Point Analysis

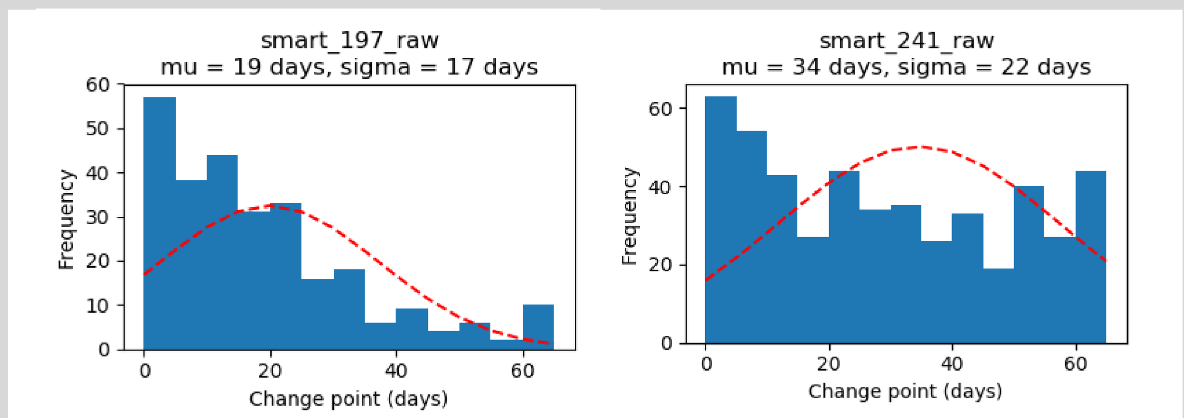
- Determine time span τ for averaging
- Assume a local trend model [3]

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t & \varepsilon_t &\sim N(0,\sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \xi_t & \xi_t &\sim N(0,\sigma_\xi^2) \\ \nu_{t+1} &= \nu_t + \zeta_t & \zeta_t &\sim N(0,\sigma_\zeta^2) \end{aligned}$$

- Find τ maximizing log-likelihood

$$\tau_{\max} = \operatorname{argmax}_\tau \log\left(p(y_{1:\tau-1}|\theta_1)\right) + \log\left(p(y_{\tau-1:t}|\theta_2)\right)$$





Class balancing

- 2015 Backblaze data set has 29,084 Seagate ST4000DM000. Among them, only 586 disk drives failed
- Three methods are applied to balance healthy and failed data sets
  - Random sampling of the same number of healthy drive
  - Cluster centroids after K-means clustering of healthy drives [2]
  - Class-weight SVM

Method	Test accuracy
Random sampling	83.0%
K-means sampling	80.5%
Class-weight SVM	83.0%

Choice of Models

Statistical Models

- Linear regression (LR) and support vector machines (SVM) models are chosen as the baseline for the project
- Gradient boosting decision tree (XGBoost) is explored as classifier with input represented as compact time series after smoothing

Deep Learning Models

- Recurrent neural network (RNN) is chosen for time series prediction including bidirectional LSTM cell with a hidden dimension of 12

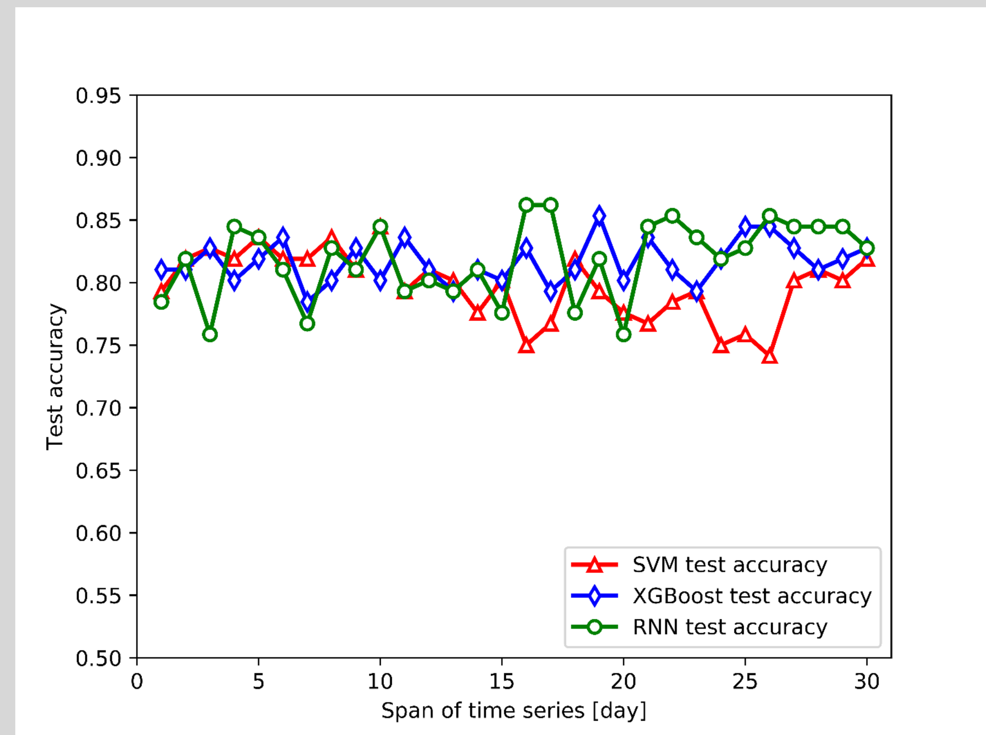
Results Summary

	Train Acc	Train Size	Test Acc	Test Size	Failed Disks			Healthy Disks		
					P	R	F	P	R	F
LR	87.3%	940	81.0%	116	0.86	0.74	0.80	0.77	0.88	0.82
SVM	88.5%	940	81.9%	116	0.91	0.74	0.82	0.78	0.93	0.85
XGBoost	92.7%	940	85.3%	116	0.91	0.86	0.82	0.79	0.84	0.90
RNN	94.6%	940	86.2%	116	0.99	0.72	0.84	0.78	0.99	0.88

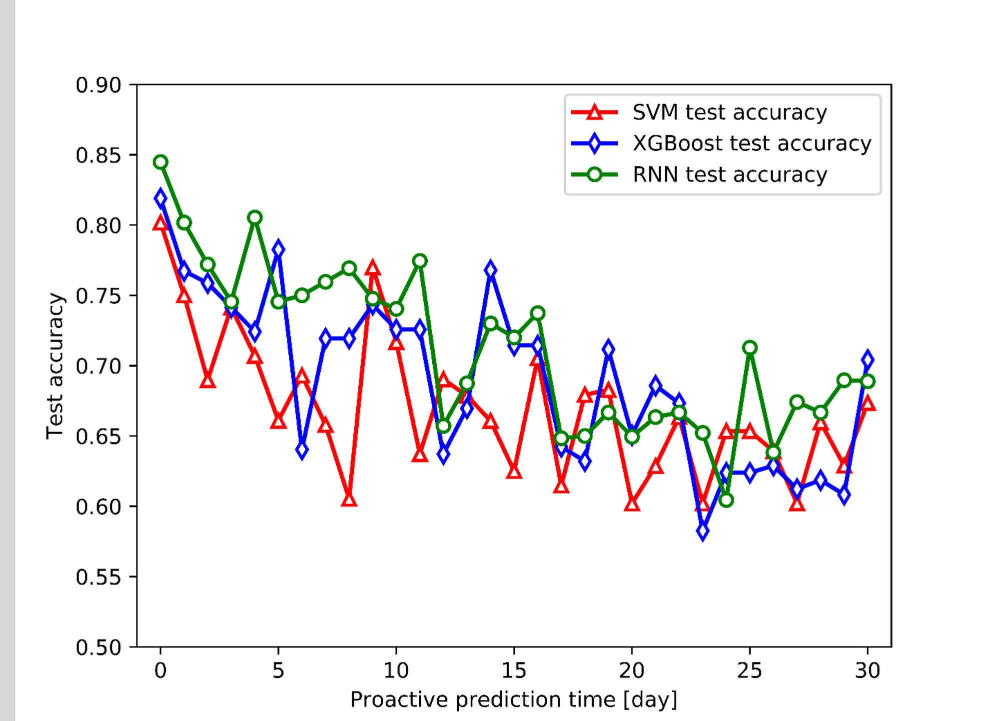
[To be updated with data normalization]

- Results in the table are obtained with 18 features
- Performance is comparable among 4 models, which may suggest that there is no significant advantage of choosing one over the other for this problem
- The fact that all algorithms have low recall for failed disks and high recall for healthy disks suggests high false negative rate that classifies failed disks as healthy (see error analysis in Discussion part)

Proactive Health Prediction

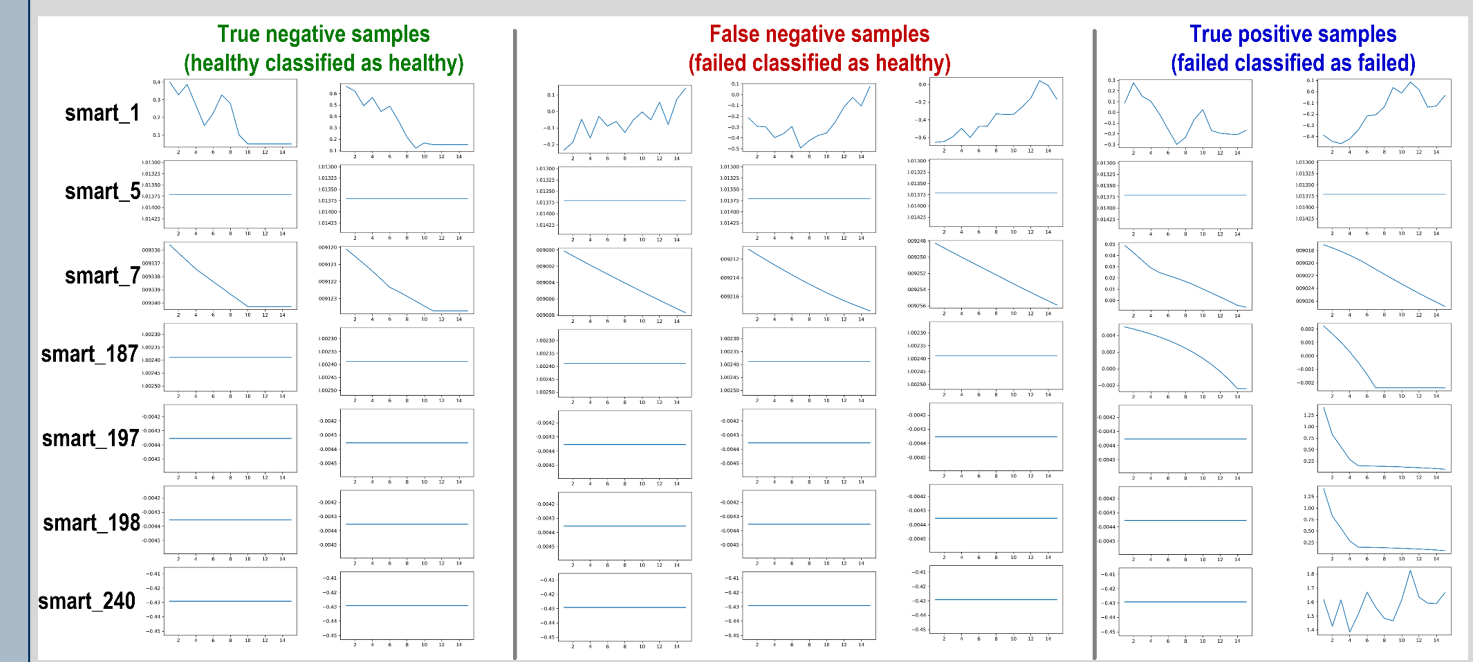


- The test accuracy of three models versus time windows for time series



- **Proactive prediction:** the accuracy of correct prediction before failure event (proactive prediction time) suggests **RNN may have better proactive prediction capability**, since its accuracy decrease slower when proactive time goes larger

Discussion (Error Analysis)



- Error analysis of RNN results: comparing the SMART attributes of true negative, false negative, and true positive samples (only 7 attributes are plotted here to simplify the discussion)
- True negative vs. false negative: it is not surprising that false negative happens since most of the attributes are essentially the same. The only slight difference is that smart_1 in false negative case tends to have more negative part
- True positive vs. false negative: this tends to suggest that in order for a sample to be classified as positive, in addition to the negative in smart_1, other attributes should also have some kind of variations

Future

- Perform a more systematic feature selection. Statistical models tend to vary a lot with different features and RNN tends to be more robust on feature variations
- Explore new models that can distinguish the error samples identified in error analysis to achieve better accuracy
- Train the models (especially the neural networks) with more data

References

[1] B. Beach, "Hard drive smart stats," 2014. [Online]. Available: https://www.backblaze.com/blog/hard-drive-smart-stats/
[2] M. M. Botezatu, et al., "Predicting disk replacement to-wards reliable data centers," in proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 39-48.
[3] J. Commandeur and S. Koopman, "An Introduction to State Space Time Series Analysis," Oxford, 2007.