

# Transferência de Aprendizado para Redes Bayesianas com Aplicação em Predição de Falha de Discos Rígidos\*

Francisco Lucas F. Pereira<sup>1</sup>, Fernando Dione S. Lima<sup>1</sup>, Lucas G. M. Leite<sup>1</sup>,  
João Paulo P. Gomes<sup>1</sup>, Javam C. Machado<sup>1</sup>

<sup>1</sup>LSBD - Departamento de Computação  
Universidade Federal do Ceará (UFC), Fortaleza, Brasil

{lucas.falcao, fernando.lima, lucas.goncalves,  
joao.pordeus, javam.machado}@lsbd.ufc.br

**Abstract.** *Predicting failures in Hard Disk Drives (HDD) is very important to avoid data loss and additional costs. Therefore, an increasing effort may be observed to find suitable failure prediction methods. Despite the encouraging performances achieved by various methods, one aspect that shall be noticed is the lack of available data to build reliable models. Transfer learning strategies offer a valid alternative since it can be used to transfer learning from HDD models with enough data to HDD models with lack of data. In this work, we evaluate transfer learning strategies in this task. Additionally we propose a strategy to build information sources based on the clustering of similar HDD models. Results showed that all tested transfer learning scenarios can improve the performance of the HDDs failure prediction methods, mainly for HDDs with very limited data.*

**Resumo.** *Predizer falhas em Discos Rígidos é muito importante para evitar perda de dados e custos adicionais. Logo, um esforço pode ser observado para encontrar métodos adequados de predição de falhas. Apesar dos resultados encorajantes alcançados por vários métodos, um aspecto notado é a falta de dados disponíveis para construir modelos confiáveis. Transferência de Aprendizado oferece uma alternativa válida, uma vez que pode ser usada para transferir conhecimento de modelos de Disco com muitos dados para Discos com menos dados. Neste trabalho, avaliamos estratégias de Transferência de Aprendizado para esta tarefa. Além disso propomos uma estratégia para construir fontes de informação baseadas no agrupamento de modelos de disco parecidos. Resultados mostraram que todos os cenários testados de transferência melhoram a performance dos métodos de predição, principalmente para Discos com muito poucos dados.*

## 1. Introdução

Ser capaz de prever falhas em Discos Rígidos (Hard Disk Drives, HDDs) tem sido um dos objetivos mais procurados por fabricantes de HDDs nos últimos anos. Atualmente, a maioria dos HDDs são equipados com um sistema de monitoramento chamado *Self-Monitoring, Analysis and Reporting Technology* (SMART). O SMART consiste em um sistema computacional que monitora e coleta dados de um dispositivo e detecta anomalias,

---

\*Este trabalho foi parcialmente apoiado pela Funcap e LSBD/UFC.

comumente por meio de uma abordagem baseada em limiares, que resultam em baixos índices de detecção de falha, variando de 3% a 10% [Murray et al. 2005].

De acordo com o estudo de [Vachtsevanos et al. 2006], o termo prognóstico de falha é frequentemente usado para caracterizar duas tarefas diferentes: detecção de falha incipiente e estimação de Vida Útil Restante (Remaining Useful Life, RUL). No primeiro, um método é projetado para detectar um comportamento anômalo do sistema estudado. Essa anomalia pode indicar que uma falha incipiente ocorreu e uma vez que o sistema ainda está funcionando, ele vai falhar em um futuro próximo. O último consiste em estimar o tempo restante até que o sistema pare de funcionar, o RUL. Geralmente esse passo é iniciado após uma anomalia ter sido detectada. Trabalhos anteriores em estimação de RUL em HDDs são menos frequentes, mas podem ser encontrados em [Chaves et al. 2016] e [Xu et al. 2016].

Um importante aspecto que pode ser notado ao projetar métodos de predição de falha em HDD é a grande variedade de modelos de HDD. É esperado que HDDs de fabricantes diferentes com características diferentes (capacidade, velocidade, tamanho etc) possam ter vários padrões de degradação diferentes. Portanto, pode ser necessário construir métodos diferentes de predição ou adaptar um modelo para funcionar para outros HDDs.

Algoritmos de Transferências de Aprendizado (Transfer Learning, TL) visam atenuar os requisitos de dados de uma tarefa ao considerar dados de tarefas relacionadas [Zhou et al. 2016]. Neste trabalho, avaliamos várias estratégias de Transferência de Aprendizado para a tarefa de predição de falhas em HDD. Modelos de HDD com grande quantidade de dados são usados como fontes e o conhecimento construído é usado para melhorar a performance dos modelos alvo, que são os HDDs com limitação de dados disponíveis. Para a estimação do RUL utilizamos um modelo baseado em Redes Bayesianas (Bayesian Network, BN) proposto em [Chaves et al. 2016]

Resultados mostraram que todas as abordagens de transferência de aprendizado utilizadas podem melhorar a performance de métodos de predição de falha em HDDs, principalmente para os modelos de HDD com quantidade de dados muito limitada. Além disso, a fonte de conhecimento baseada em agrupamento proposta também resultou em ganhos de performance para todos os métodos de transferência e modelos de HDD testados.

A implementação e a execução dos experimentos foram feitas pelo aluno de Iniciação Científica Lucas Pereira, sob a orientação dos professores João Paulo Pordeus e Javam Machado. Além disso, teve a colaboração do aluno de mestrado Fernando Dione e do aluno de doutorado Lucas Leite. O trabalho deu origem a uma publicação no *2017 Brazilian Conference on Intelligent Systems (BRACIS)*, no qual o aluno de IC foi o autor principal [Pereira et al. 2017].

## **2. Trabalhos Relacionados**

Considerando a tarefa de estimação de RUL, dois trabalhos recentes propuseram neste tópico. Em [Xu et al. 2016], os autores usaram uma Rede Neural Recorrente (RNN) para classificar HDDs em seis estados de saúde de acordo com seus RULs. O modelo RNN inferiu o estado de saúde de cada HDD dado seus atributos SMART. A classificação foi

realizada em um conjunto de dados abrangendo o último mês funcional de cada HDD. De forma similar, em [Chaves et al. 2016] os autores também classificaram HDDs em estados de saúde de acordo com seus atributos SMART. Entretanto, as predições foram realizadas para horizontes temporais maiores. Um modelo baseado em Redes Bayesianas foi projetado para inferir o estado de saúde dos HDDs em até dois anos antes da falha.

Como pode ser observado, nenhum trabalho anterior considerou o problema de número reduzido de exemplos para construir um modelos de estimação de RUL. Em nosso estudo, usamos o modelo proposto em [Chaves et al. 2016] como um método de estimação de RUL baseline e testamos várias estratégias de Transferência de Aprendizado para lidar com esse problema. Adicionalmente, propomos uma nova forma de construir um conjunto de dados fonte para realizar a transferência.

### 3. Base Teórica

#### 3.1. Redes Bayesianas

Uma Rede Bayesiana (Bayesian Network, BN) [Pearl 1988] é um modelo gráfico probabilístico que relaciona um conjunto de variáveis aleatórias de acordo com suas dependências condicionais. Uma BN modela as relações possíveis de causa e consequência de um dado fenômeno de uma série de observações. Formalmente, uma BN é um par  $B = (G, \Theta)$ , onde  $G = (X, E)$  é um grafo direcionado acíclico cujo conjunto de vértices  $X = \{X_1, X_2, \dots, X_n\}$  representam as variáveis aleatórias e o conjunto de arestas  $E$  representa as dependências condicionais entre as variáveis representadas por  $X$ . Essas dependências são definidas por um conjunto de funções probabilísticas  $\Theta$ . Esse conjunto contém o parâmetro  $\theta_{x_i|\pi_i} = P(x_i|\pi_i)$  para cada  $x_i \in X_i$  condicionado por  $\pi_i$ , ou seja, o conjunto de parâmetros para  $X_i$ . A seguinte equação apresenta a distribuição conjunta definida por uma BN sobre o conjunto de variáveis aleatórias:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(x_i|\pi_i) \quad (1)$$

#### 3.2. Transferência de Aprendizado para Redes Bayesianas

Transferência de Aprendizado (TL) busca construir um modelo para uma tarefa alvo dado um conjunto de dados referentes a essa tarefa e a outras tarefas relacionadas, chamadas tarefas fonte. Desta forma, podemos construir modelos mais precisos com o enriquecimento do conjunto de dados ou modelo da tarefa alvo. No caso de Redes Bayesianas, vários métodos de transferência foram propostos e divididos em dois tópicos principais: aprendizado de estrutura e aprendizado de parâmetros [Luis et al. 2010, Zhou et al. 2016].

Aprendizado de Estrutura é interessado em descobrir as melhores conexões para o grafo da BN. Nesse problema, o objetivo é identificar as relações entre as variáveis usando os dados da tarefa alvo e das tarefas fontes. Métodos populares e uma revisão mais detalhadas da literatura pode ser vista em [Spirites et al. 2000] e [Luis et al. 2010]. A tarefa de transferência de aprendizado para parâmetros em BN consiste em construir as Tabelas de Probabilidade Condicionais (CPTs) usando dados do alvo e das fontes.

Neste trabalho vamos considerar somente a tarefa de transferência de aprendizado de parâmetros. Para tal, três dos mais populares métodos são descritos em [Luis et al. 2010]. Os métodos, chamados Linear Aggregation, Logarithmic Aggregation e Distance Based Linear Pool são apresentados abaixo.

**Linear Aggregation:** a nova probabilidade alvo de uma variável  $X$ ,  $P(X)$ , é uma soma ponderada da tarefa alvo e das tarefas fonte auxiliares, expressado a seguir:

$$P(X) = k \sum_{i=1}^m w_i P_i(X) \quad (2)$$

onde  $P_i(X)$  representa a probabilidade condicional do  $i$ -ésimo modelo envolvendo  $X$ ,  $w_i$  é o peso da  $i$ -ésima probabilidade e  $k$  é um fator de normalização.

**Logarithmic Aggregation:** a nova probabilidade alvo de uma variável  $X$ ,  $P(X)$ , é um produto ponderado da tarefa alvo e das tarefas fontes auxiliares, expressado a seguir:

$$P(X) = k \prod_{i=1}^m P_i(X)^{w_i} \quad (3)$$

onde  $P_i(X)$  representa a probabilidade condicional do  $i$ -ésimo modelo envolvendo  $X$ ,  $w_i$  é o peso da  $i$ -ésima probabilidade e  $k$  é um fator de normalização.

**Distance Based Linear Pool (DBLP):** o método considera a confiança da estimação de probabilidade da tarefa fonte auxiliar e sua similaridade com a estimação da tarefa alvo. O método dá maior peso para a estimação de probabilidade que tem maior confiança e é mais similar à tarefa alvo. A confiança é baseada no tamanho de cada conjunto de dados. O DBLP pode ser dividido nas seguintes etapas, onde  $p = P(X)$  denota cada probabilidade individual nas CPTs:

1) Obtem as probabilidades médias de todos os conjuntos de dados de acordo com um fator de confiança  $\bar{p} = k \sum_{i=1}^m (f_i \cdot p_i)$ , onde  $f_i$  é o fator de confiança para cada estimação de probabilidade e  $k$  é um fator de normalização. O fator de confiança depende do tamanho de cada conjunto de dados usado para estimar a CPT e é definido a seguir:

$$f_i = \begin{cases} 1 - \frac{\log(c_f)}{c_f}, & \text{se } c_f \geq 3 \\ 1 - \frac{c_f \cdot \log(3)}{3}, & \text{se } c_f < 3 \end{cases} \quad (4)$$

onde  $c_f = \frac{N}{2 \cdot T}$  é proporcional ao erro esperado que depende de  $T$ , o número de entradas na CPT, e  $N$  é o número de casos no conjunto de dados. 2) Obtem as distâncias mínimas ( $d_{min}$ ) e máximas ( $d_{max}$ ) entre a probabilidade do alvo e da média calculada acima. 3) Estima a nova probabilidade condicional do modelo alvo na forma:  $p'_{alvo} = (1 - c_i)p_{alvo} + c_i \bar{p}$ , onde os coeficientes de agregação  $c_i$  expressam quanto considerar das CPTs dos outros modelos. Se as CPTs da rede alvo forem similares à média de todas as CPTs, então mais peso é dado para a média, caso contrário, mais peso é dado para as CPTs do alvo. Isso é expressado a seguir, onde  $c_{max}$  e  $c_{min}$  são parâmetros que indicam o quão forte queremos considerar a influência das outras CPTs:

$$c_i = (d_i - d_{min}) \left( \frac{c_{max} - c_{min}}{d_{max} - d_{min}} \right) + c_{min} \quad (5)$$

## 4. Metodologia

Dado um modelo de predição que é o alvo de processo de transferência, existem dois problemas para solucionar: como transferir (quais métodos) e de onde transferir (quais estratégias para selecionar as fontes). Para lidar com a primeira questão, nós aplicamos os três métodos de TL para aprendizado de parâmetro explicados na Secção 3. Em relação à segunda questão, nós usamos estratégias baseadas no tamanho do conjunto de dados disponível para cada modelo de fabricante de disco para escolher as fontes. Adicionalmente, também exploramos uma nova estratégia para aprimorar a performance de transferência através do agrupamento de dispositivos de acordo com suas especificações de fábrica. Isso efetivamente resulta na criação de uma nova fonte de dados. Essa abordagem será referenciada como fonte de informação baseada em agrupamento.

### 4.1. Modelo de Predição de Falhas

Neste trabalho, o Modelo de Predição de Falhas baseline será uma Rede Bayesiana com a estrutura apresentada na Figura 1. Essa estrutura é obtida em parte do método BaNHFaP, proposto em [Chaves et al. 2016]. A entrada do modelo de predição é uma observação de 8 atributos SMARTS representando o estado de um HDD específico e sua informação de Tempo Ligado em Horas (Power On Hours, POH), e a saída do modelo é a Vida Útil Restante (RUL) esperada do disco.

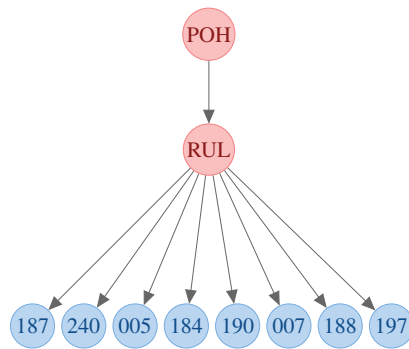
Os SMARTs usados foram 187, 5, 184, 7, 240, 190, 188 e 197, que são definidos respectivamente como *Reported Uncorrectable Errors*, *Reallocated Sectors Count*, *End-to-End Error*, *Seek Error Rate*, *Head Flying Hours*, *Airflow Temperature*, *Command Timeout* e *Current Pending Sector Count*. O RUL é discretizado em trimestres do ano, então o modelo é capaz de prever em qual trimestre dos próximos anos o HDD vai falhar com algum grau de certeza.

Estamos considerando a mesma estrutura para todos os modelos fontes e alvo. Como apresentado em [Chaves et al. 2016], essa estrutura mostrou bons resultados para a mesma tarefa deste trabalho. A única diferença é que os autores em [Chaves et al. 2016] construíram uma única BN usando dados de vários modelos de HDD. Em nosso trabalho, buscamos construir uma BN individual para cada modelo de HDD.

### 4.2. Estratégias de Transferência de Aprendizado

Uma vez que temos um grupo de Modelos de Predição treinados, que em nosso caso são Redes Bayesianas, nós executamos a agregação de informação de parte desse grupo para algum alvo. Como a estrutura das BNs é fixada, nos resta apenas a possibilidade de transferir conhecimento das tabelas de probabilidade (parâmetros) desse grupo de BNs para o modelo alvo.

Selecionar as fontes da transferência de parâmetro para um alvo pode ser uma tarefa difícil. Uma abordagem ingênua é escolher, para ser as fontes, todas as Redes



**Figura 1. Representação da estrutura do grafo gerado pelo método BaNHFap.**

Bayesianas que foram treinadas com mais dados do que a rede alvo. Essa simples estratégia será chamada de **Estratégia 1**.

Na prática, muitos modelos de HDDs tem muito mais dados disponíveis que outros, seja porque o modelo de disco é mais velho ou porque é preferido por cliente por ter um maior custo-benefício, por exemplo. Então uma outra abordagem ingênua é escolher somente os modelos de disco para os quais a quantidade de dados disponíveis para eles é muito maior que todos os outros. Para esses discos, é esperado que eles tenham Modelos de Predição de Falhas mais precisos. Essa estratégia será chamada **Estratégia 2**.

### **4.3. Fonte de Informação Baseada em Agrupamento**

Escolher todos os discos com mais dados disponíveis que o disco alvo (Estratégia 1) pode não ser a melhor opção porque, por exemplo, o disco alvo pode ter muito poucos dados. Dessa forma, vários Modelos de Predição treinados de discos com poucos dados vão contribuir para a Rede Bayesiana final. Conseqüentemente, algumas BNs podem atrapalhar mais do que ajudar no processo de transferência, um efeito conhecido como Transferência Negativa na literatura [Pan and Yang 2010]. Métodos de Transferência de Aprendizado como o DBLP tentam amenizar esse problema pesando a contribuição das fontes de acordo com o tamanho de seus conjuntos de dados, mas isso pode ser ruim se os pesos dados fazem suas contribuições serem muito pequenas ou insignificantes. Este último problema também ocorre quando são escolhidos apenas os discos com muito mais dados disponíveis (Estratégia 2), porque um grande número de informação útil de outros modelos de disco não são considerados.

Tentando prevenir esses problemas, propomos uma nova forma de considerar a influência de discos com poucos dados, agrupando-os e concatenando os dados de cada grupo, resultando em novos modelos de disco. Propomos usar a especificação de fábrica dos discos como atributos para executar o processo de agrupamento. Nós esperamos que HDDs com características físicas similares (especificações) possuam degradação similar ao longo do tempo. Os atributos de especificações escolhidos foram capacidade(em bytes), Rotações por Minuto (RPM), Data Transfer Interface, tamanho de cache, número de cabeças de leitura/escrita e discos internos, latência média para operações de leitura e escrita e especificações de tamanho do dispositivo.

Como alguns discos não tem todos esses atributos disponíveis, devemos aplicar um algoritmo de clusterização que suporta entradas de dados com valores faltantes. No nosso caso, aplicamos o método k-POD [Jocelyn T. Chi and Baraniuk 2016]. Com os

agrupamentos produzidos pelo k-POD em mãos, podemos criar fontes de dados produzidos pela união de dados em um ou mais desses grupos.

Esperamos que o Modelo de Predição de Falha treinado com esses dados do agrupamento de discos tenha melhor performance do que os modelos individuais, treinados para cada modelo de HDD que pertence ao grupo. Dessa forma, a contribuição da Transferência de Aprendizado deve ser mais eficiente. Além disso, métodos que consideram o tamanho do conjunto de dados, como o DBLP, darão mais confiança aos discos agrupados do que se fossemos usar apenas os membros do grupo separadamente.

A **Estratégia 3** usa os discos com bastante dados disponíveis, como na Estratégia 2, e também os discos clusterizados para ser as fontes no processo de transferência de aprendizado.

## 5. Resultados

Nossos experimentos foram implementados em Python, usando os pacotes Pandas 0.19.2 [McKinney 08] e scikit-learn 0.18.1 [Pedregosa et al. 2011] para pré-processamento de dados. Adicionalmente, usamos a biblioteca pgmpy 0.1.3 [Ankan and Panda 2015] para a construção e avaliação das Redes Bayesianas.

### 5.1. Conjunto de Dados

Em nossos experimentos, usamos um conjunto de dados fornecido pela Empresa Backblaze [Backblaze 2016]. Os dados da Backblaze abrangem 59,654,783 observações diárias do estado de vários discos distribuídos entre 93 modelos diferentes coletados entre 10/04/2013 e 31/12/2016. Essas observações contêm informações básicas sobre cada dispositivo além de seus dados SMART. É considerado que um dispositivo falhou quando 1) ele para de funcionar ou 2) ele mostrou sinais que irá parar em breve, como por exemplo, se seu autoteste SMART falhou.

Do conjunto de dados da Backblaze, selecionamos os modelos que possuem mais de 5 seriais e 500 observações. O conjunto de dados resultando abrange 15 modelos descritos na tabela 1. Adicionalmente, excluímos aqueles dispositivos cujo serial falhou em algum momento e reapareceu depois. A tabela 1 também mostra o número de HDDs e observações para cada modelo. Os modelos em negrito foram os escolhidos para ser os alvos da execução das estratégias de Transferência de Aprendizado. Escolhemos esses três modelos de discos considerando a sua variedade no que diz respeito à falta de dados.

Para a otimização do processo de construção das Redes Bayesianas, removemos atributos SMART que não possuem um impacto forte na predição de falha. Utilizaremos apenas os atributos selecionados em [Chaves et al. 2016], que foram os mais importantes de acordo com um processo de seleção de atributos baseado em uma Eliminação Recursiva de Features [Guyon et al. 2002].

### 5.2. Avaliação de Performance

Em nossos experimentos, obtivemos resultados para todas as combinações possíveis de estratégia de obtenção de dados fonte e métodos de transferência de aprendizado citados anteriormente. Para cada modelo de HDD alvo, seu dataset foi dividido em 70% para treinamento e 30% para testar a performance das BNs. Para os métodos Linear Aggregation e Logarithmic Aggregation, uma busca em grade foi realizada para achar os melhores

**Tabela 1. Modelos de HDD com quantidade significativa de dados.**

Modelo de HDD	Número de Seriais	Observações
ST320005XXXX	6	511
<b>ST250LT007</b>	<b>7</b>	<b>2744</b>
ST3160316AS	8	2853
ST3160318AS	13	7044
ST32000542AS	13	2038
<b>ST33000651AS</b>	<b>16</b>	<b>3556</b>
ST1500DL003	35	8910
ST4000DX000	35	25658
ST6000DX000	42	10588
ST8000DM002	48	3531
<b>ST320LT007</b>	<b>80</b>	<b>54640</b>
ST31500341AS	102	22251
ST31500541AS	239	63161
ST3000DM001	1012	174474
ST4000DM000	1698	747296

pesos, usando um conjunto de validação de 20% dos dados de treino. Para o DBLP, os hiperparâmetros foram estabelecidos em  $c_{min} = 0.25$  e  $c_{max} = 0.75$ , como sugerido em [Luis et al. 2010].

Escolhemos modelos alvo com diferentes números de observações para verificar o impacto de tais características nos métodos. Para a Estratégia 2, usamos os modelos de disco **ST3000DM001** e **ST4000DM000** para serem as fontes com mais dados. Para a Estratégia 3, criamos um modelo fonte a partir do agrupamento dos dados dos modelos de disco **ST320005XXXX**, **ST32000542AS**, **ST1500DL003**, **ST31500341AS** e **ST31500541AS**.

As performances foram avaliadas por três métricas: média do erro, mediana do erro e acurácia. Enquanto acurácia é uma métrica comumente utilizada para classificação, o erro consiste na diferença, em trimestres, entre a predição do classificador e a vida útil restante real do disco. A média do erro é importante uma vez que pequenos desvios em uma classificação são menos severos do que grandes desvios. Considere um exemplo em que uma BN prevê que um dado HDD tem apenas dois meses restantes mas o HDD ainda funciona por 4 meses. Esse erro é claramente menos significativo para quantificar a performance do modelo de predição do que um caso em que o modelo prevê que o mesmo disco ainda tem dez meses restantes. Embora os dois modelos classificaram de forma errada a mesma amostra, a magnitude do erro é diferente. A terceira métrica (mediana do erro) foi escolhida uma vez que a mediana é menos sensível a outliers. A Tabela 2 mostra os resultados para todos os modelos alvo, considerando as métricas acima mencionadas.

De acordo com os resultados, podemos ver que usar qualquer estratégia de transferência pode aprimorar a performance do modelo alvo. Isso pode ser verificado ao comparar o resultado de cada combinação (estratégia de escolha de fontes + método de transferência) à abordagem “sem transferência”. Um outro aspecto importante a ser notado é que, como esperado, o ganho em performance é mais significativo para conjuntos de



**Tabela 2. Resultados das combinações de estratégias e métodos de Transferência de Aprendizado mostrando média do erro, mediana do erro e taxa de acerto.**

Método de Transferência	Fonte	ST250LT007			ST33000651AS			ST320LT007		
		Média	Mediana	Acurácia	Média	Mediana	Acurácia	Média	Mediana	Acurácia
sem transferência	-	2.3822	1.7765	49.817%	0.9929	<b>0.5245</b>	29.653%	1.9280	1.4362	25.099%
Linear aggregation	Estratégia 1	1.7879	1.4577	47.996%	0.9344	0.7038	43.966%	1.8256	1.5504	24.794%
	Estratégia 2	1.5973	1.5563	52.003%	<b>0.8458</b>	0.5617	42.469%	1.8214	1.5816	24.119%
	Estratégia 3	1.4928	1.4570	<b>62.021%</b>	0.8769	0.5315	41.721%	1.8208	1.5407	24.000%
Logarithmic aggregation	Estratégia 1	1.8677	1.2816	47.996%	0.9330	0.5386	37.792%	1.8496	1.5057	<b>28.720%</b>
	Estratégia 2	1.6252	0.9721	50.091%	0.8736	0.5837	45.556%	1.8321	1.5254	26.350%
	Estratégia 3	1.5563	0.9849	42.896%	0.8620	0.5650	45.556%	1.8250	<b>1.4279</b>	25.721%
Distance based linear pool	Estratégia 1	1.6063	1.2528	56.102%	0.9386	0.7084	53.040%	1.7631	1.5487	23.814%
	Estratégia 2	1.5632	0.9557	52.003%	0.9070	0.5690	<b>65.107%</b>	1.7778	1.5960	22.073%
	Estratégia 3	<b>1.4207</b>	<b>0.8565</b>	56.102%	0.9103	0.6880	53.133%	<b>1.7554</b>	1.5329	23.556%

dados com menos observações.

Comparando as combinações de métodos de transferência e estratégias de escolha de fontes podemos observar que a Estratégia 3 combinada com diferentes métodos de transferência alcançou os melhores resultados para todos os critérios de performance no HDD de modelo **ST250LT007**. Isso é um importante resultado uma vez que **ST250LT007** é o modelo alvo de HDD com o menor número de observações e também aquele que mostrou maior ganho de performance quando comparado à estratégia “sem transferência”.

Nos modelos **ST33000651AS** e **ST320LT007**, podemos verificar que diferentes combinações tiveram a melhor performance. Uma possível explicação poder estar no fato de que esses HDDs tem mais observações e, portanto, transferência de aprendizado pode ter menos impacto na performance de cada modelo. Embora várias combinações tiveram os melhores resultado, é perceptível que a Estratégia 3 ou alcançou os melhores resultados ou teve uma performance parecida com as melhores combinações.

## 6. Conclusão

Neste trabalho avaliamos várias estratégias de Transferência de Aprendizado para a tarefa de predição de falha em Discos Rígidos. As estratégias testadas combinam três métodos de transferência e três abordagens para construir conjuntos de dados fontes.

Usamos um método de predição de falha baseado em um modelo de BN e testamos três estratégias conhecidas de transferência de aprendizado para transferência de parâmetro em BNs. Além de duas abordagens comumente utilizadas para construir dados fonte, propomos um novo método de construção chamado fonte de informação baseada em agrupamento. Nesse método proposto, agrupamos vários HDDs de acordo com suas similaridades e construímos uma nova fonte de informação para a transferência de aprendizado.

Com base em nossos experimentos podemos dizer que métodos de transferência de aprendizado podem proporcionar ganhos em performance em modelos de predição de falha em HDDs. Isso é ainda mais perceptível para HDDs com número muito reduzido de observações. Também verificados que o método de construção de conhecimento fonte levou a melhorias de performance e pode ser considerado uma alternativa válida para Transferência de Aprendizado.

## Referências

- Ankan, A. and Panda, A. (2015). pgmpy: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*.
- Backblaze (2016). Hard drive data and stats. <https://www.backblaze.com/b2/hard-drive-test-data.html>. [Online; accessed 2017-04-26].
- Chaves, I. C., de Paula, M. R. P., Leite, L. G. M., Queiroz, L. P., Gomes, J. P. P., and Machado, J. C. (2016). Banhfap: A bayesian network based failure prediction approach for hard disk drives. In *Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on*, pages 427–432. IEEE.
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422.
- Jocelyn T. Chi, E. C. C. and Baraniuk, R. G. (2016).  $k$ -pod: A method for  $k$ -means clustering of missing data. *The American Statistician*, 70:91–99.
- Luis, R., Sucar, L. E., and Morales, E. F. (2010). Inductive transfer for learning bayesian networks. *Machine Learning*, 79(1):227–255.
- McKinney, W. (2008–). Pandas: a python data analysis library. <http://pandas.sourceforge.net>. [Online; accessed 2017-04-26].
- Murray, J. F., Hughes, G. F., and Kreutz-Delgado, K. (2005). Machine learning methods for predicting failures in hard drives: A multiple-instance application. *Journal of Machine Learning Research*, 6(May):783–816.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Pearl, J. (1988). Probabilistic reasoning in intelligent systems.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830.
- Pereira, F. L. F., dos Santos Lima, F. D., de Moura Leite, L. G., Gomes, J. P. P., and de Castro Machado, J. (2017). Transfer learning for bayesian networks with application on hard disk drives failure prediction. In *2017 Brazilian Conference on Intelligent Systems (BRACIS)*, pages 228–233. IEEE.
- Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.
- Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., and Wu, B. (2006). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems: Methods and Case Studies*. John Wiley & Sons.
- Xu, C., Wang, G., Liu, X., Guo, D., and Liu, T.-Y. (2016). Health status assessment and failure prediction for hard drives with recurrent neural networks. *IEEE Transactions on Computers*, 65(11):3502–3508.
- Zhou, Y., Hospedales, T. M., and Fenton, N. (2016). When and where to transfer for bayesian network parameter learning. *Expert Systems with Applications*, 55:361 – 373.