



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## **ESTUDO SOBRE TÉCNICAS DE DETECÇÃO AUTOMÁTICA DE ANOMALIAS EM SATÉLITES**

Ana Maria Ambrósio

Denise Nunes Rotondi Azevedo

Marco Vieira

Relatório Técnico resultado do Exame de Qualificação do Curso de Pós-Graduação em Engenharia e Gerenciamento de Sistemas Espaciais.

INPE  
São José dos Campos  
2011





MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## **ESTUDO SOBRE TÉCNICAS DE DETECÇÃO AUTOMÁTICA DE ANOMALIAS EM SATÉLITES**

Ana Maria Ambrósio

Denise Nunes Rotondi Azevedo

Marco Vieira

Relatório Técnico resultado do Exame de Qualificação do Curso de Pós-Graduação em Engenharia e Gerenciamento de Sistemas Espaciais.

INPE  
São José dos Campos  
2011



## **AGRADECIMENTOS**

Agradeço a contribuição dos especialistas Mario Celso Padovan de Almeida e Carlos Felipe Soriano Freire que em muito auxiliaram no entendimento das anomalias ocorridas nos satélites da missão espacial brasileira que serviram como Estudos de Caso nesse trabalho.

.



## RESUMO

Os dados de telemetria são recebidos em tempo real e analisados por especialistas e operadores, constituindo a principal fonte de identificação e predição de anomalias em satélites artificiais. Entretanto, o grande número de telemetrias faz com que essa seja uma tarefa extremamente complexa. Como parte do Exame de Qualificação, esta monografia descreveu um estudo prático que avaliou a hipótese de se utilizar algoritmos de agrupamento para auxiliar operadores e especialistas na análise das telemetrias de tempo real.

O trabalho avaliou o desempenho dos algoritmos de agrupamento *K-means* e *Expectation Maximization* na tarefa de detectar anomalias automaticamente e utilizou duas anomalias reais ocorridas na missão espacial brasileira como estudo de caso.

Os resultados dos experimentos mostraram que no estudo de caso onde um grupo de telemetrias tendia a valores fora dos limites, os algoritmos de agrupamento funcionaram e poderiam servir como apoio à operação, antecipando a detecção de anomalias no satélite. Entretanto, no caso de anomalias silenciosas, com pequenas variações em apenas uma telemetria, os algoritmos não se mostraram eficientes.

O trabalho também mostrou que importantes aspectos do domínio do problema precisam ser tratados, entre eles, o envelhecimento do satélite e o grande número de telemetrias. Tais aspectos implicam na necessidade de treinamento constante dos algoritmos e na necessidade da redução do espaço de atributos através do uso combinado de algoritmos de detecção de anomalias com algoritmos de seleção de atributos.





## LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1 - Estatísticas de Falhas – Tempo Após o Lançamento .....	7
Figura 2.2 - Estatísticas de Falhas – Subsistemas.....	8
Figura 3.1 – Detecção de Anomalias – Componentes. ....	12
Figura 4.1 - Diagrama do Subsistema de Suprimento de Energia .....	23
Figura 4.2 - Descarga/Carga da Bateria.....	25
Figura 4.3 - Descarga/Carga da Bateria (Curva EOC). ....	26
Figura 4.4 - Comportamento da telemetria TM023 (S1).....	28
Figura 4.5 - Carregamento da bateria – Estado de Emergência.....	30
Figura 5.1 - Dados Simulados – Algoritmo EM.....	36
Figura 5.2 - Dados Simulados – Algoritmo K-means. ....	37
Figura 5.3 - Estudo de Caso 2 – Treinamento (2004)/Teste(2004).....	38
Figura 5.4 - Estudo de Caso 2 – Treinamento (2004)/Teste(2007).....	39
Figura 5.5 - Estudo de Caso 1 – Detecção de Anomalia (2003).....	40
Figura 5.6 - Estudo de Caso 2 – Detecção de Anomalia (2007).....	40
Figura 5.7 - Estudo de Caso 1 – Resultado Comparativo. ....	41
Figura 5.8 - Estudo de Caso 2 – Resultado Comparativo. ....	42



## LISTA DE TABELAS

	<u>Pág.</u>
Tabela 4.1 – Telemetrias do Subsistema de Potência.....	27
Tabela 5.1 – Avaliação dos Algoritmos de Agrupamento.....	34



## LISTA DE SIGLAS E ABREVIATURAS

ACS	<i>Attitude Control System</i>
AOCS	<i>Attitude and Orbit Control Subsystem</i>
BCC	<i>Battery Charge Controller</i>
BDR	<i>Battery Discharge Regulator</i>
BHC	<i>Battery Heating Controller</i>
BSR	<i>Battery Switching Regulator</i>
CDH	<i>Command and Data Handling</i>
DC	<i>Down Converters</i>
DOD	<i>Deep-of-Discharge</i>
EM	<i>Expectation maximization</i>
EOC	<i>End-of-Charge</i>
EPS	<i>Propulsion, Electrical Power Subsystem</i>
IMS	<i>Inductive Monitoring System</i>
MEA	<i>Main Error Amplifier</i>
SAG	<i>Solar Array Generator</i>
SG	<i>Solar Generator</i>
TSE	<i>The Satellite Encyclopedia</i>
TTC	<i>Telemetry, Tracking and Command</i>



## SUMÁRIO

Pág.

<b>1 INTRODUÇÃO .....</b>	<b>1</b>
<b>2 ANOMALIAS EM SATÉLITES.....</b>	<b>5</b>
2.1. Anomalias e Falhas em Sistemas Espaciais .....	5
<b>3 TÉCNICAS PARA DETECÇÃO DE ANOMALIAS .....</b>	<b>11</b>
3.1. Detecção Automática de Anomalias.....	11
3.2. Algoritmos de Agrupamento .....	16
3.3. Trabalhos Relacionados.....	20
<b>4 ESTUDO DE CASOS.....</b>	<b>23</b>
4.1. Subsistema de Suprimento de Energia .....	23
4.2. Estudos de Caso .....	27
<b>5 EXPERIMENTOS.....</b>	<b>33</b>
5.1. Resultados com uso de Dados Simulados .....	33
5.2. Resultados com uso de Dados Reais .....	37
5.3. Lições Aprendidas.....	42
<b>6 CONCLUSÃO .....</b>	<b>45</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>47</b>





## 1 INTRODUÇÃO

Satélites artificiais prestam importantes serviços em previsões climáticas, comunicação, sensoriamento remoto, experimentos científicos etc. Uma falha em um satélite pode implicar não somente em uma grande perda financeira, mas também na perda de serviços essenciais e muitas vezes estratégicos. Nesse cenário, a detecção antecipada de falhas e anomalias, seu diagnóstico precoce e sua prevenção podem promover confiabilidade e disponibilidade a sistemas espaciais, promovendo, em muitos casos, o aumento da sua vida útil.

Um satélite pode ser visto logicamente como um conjunto de subsistemas (elétrico, térmico, carga útil etc), e cada subsistema possui um conjunto de sensores que são utilizados para monitorar seus estados. As medidas obtidas através desses sensores são enviadas para estações terrenas a cada passagem dos satélites sobre as mesmas e são chamadas de dados de telemetria de serviço (*housekeeping*).

As telemetrias são recebidas em tempo real e são monitoradas pelos operadores no Centro de Controle de Satélites. Dada a característica de tempo real, o grande número de telemetrias a serem monitoradas e a taxa de amostragem dessas telemetrias torna-se difícil, quando não impossível, realizar uma análise manual cuidadosa e detalhada das mesmas durante a passagem do satélite.

Esse trabalho avaliou o uso de técnicas de detecção automática de anomalias para apoio a operadores e especialistas no monitoramento e análise de telemetrias. O principal objetivo do trabalho foi avaliar a capacidade de algoritmos, em especial de algoritmos de agrupamento, em detectar automaticamente anomalias através dos dados de telemetria.

Detecção automática de anomalias refere-se à questão de encontrar padrões que não estão em conformidade com o comportamento normal dos dados (CHANDOLA et al., 2009). Nessa tarefa, diferentes técnicas oriundas de áreas específicas de pesquisa podem ser utilizadas entre elas, técnicas estatísticas, de mineração de dados, espectrais etc.

Dadas as características dos dados de telemetria de satélite, onde telemetrias de períodos normais de operação formam grupos delimitados, foram utilizadas nesse trabalho técnicas de mineração de dados baseadas em Agrupamento. Os algoritmos de agrupamento são potencialmente interessantes na detecção de anomalias em sistemas de satélite cujo monitoramento é feito em tempo real, pois nesse caso a detecção das anomalias é bastante eficiente, já que os dados recebidos são comparados a grupos já previamente definidos.

Neste trabalho experimental foram apresentados os resultados da aplicação de dois algoritmos de agrupamento – *K-means* (WEBB, 2002) e *Expectation Maximization* (EM) (DEMPSTER; RUBIN, 1977) – a dois casos reais de anomalias ocorridas com satélites da missão espacial brasileira. Um dos objetivos da execução do mesmo foi demonstrar a aplicação dos conhecimentos obtidos em várias disciplinas realizadas até essa etapa do curso.

A seguir está descrita a estrutura do trabalho e a correlação entre cada capítulo e as disciplinas que serviram como apoio teórico para sua elaboração:

- O capítulo 2 apresenta estudos sobre falhas de sistemas espaciais e está correlacionado à disciplina de **Validação e Verificação em Sistemas Espaciais**;
- O capítulo 3 apresenta técnicas de detecção de anomalias e os algoritmos utilizados no experimento conduzido e teve como base teórica a disciplina **Princípios e Aplicações de Mineração de Dados**;

- O capítulo 4 apresenta dois estudos de caso que foram utilizados nos experimentos práticos. Foram abordadas nesse capítulo as disciplinas: **Introdução à Tecnologia de Satélites e Engenharia de Plataformas Orbitais e Cargas Úteis;**
- O capítulo 5 apresenta a metodologia utilizada na realização dos experimentos, os resultados obtidos e as lições aprendidas;
- O capítulo 6 apresenta as conclusões do trabalho e perspectivas futuras.



## **2 ANOMALIAS EM SATÉLITES**

Nessa seção serão apresentados dois estudos estatísticos sobre anomalias e falhas em satélite que justificaram o nosso estudo experimental.

### **2.1. Anomalias e Falhas em Sistemas Espaciais**

Os Sistemas Espaciais possuem duas características que os diferenciam de outros sistemas complexos: 1) operam em ambientes hostis e não completamente dominados pelo conhecimento humano; 2) uma vez em funcionamento ou em órbita têm manutenção extremamente difícil, dispendiosa e muitas vezes impossível.

Considerando todos os aspectos desse cenário, pode-se observar que características como confiabilidade e disponibilidade devem desempenhar um papel crucial em todo o ciclo de vida desses sistemas, seja na fase de projeto, fabricação, e também na operação dos mesmos (ECSS, 1998).

Será apresentado a seguir um histórico resumido de falhas ocorridas em sistemas espaciais. As estatísticas aqui apresentadas foram fruto de dois estudos realizados por Robertson e Stoneking (2003) e Tafazoli (2009) que pesquisaram um conjunto de fontes onde haviam sido reportadas falhas de satélites, estabeleceram uma metodologia de análise, classificaram essas falhas e apresentaram resultados estatísticos sobre as mesmas.

#### **Fontes**

Os dois trabalhos utilizaram alguns bancos de dados públicos, disponíveis na web: *The Satellite Encyclopedia* (TSE), *Satellite News Digest*. O trabalho de Robertson e Stoneking (2003) utilizou, também, as fontes: *Mission and Spacecraft Library*, *Air claims Space Track* e a *Encyclopedia Astronautic*, todos

públicos; e dados do *Space System Engineering Database* (SSED) e do *Mission Failure Analysis for NASA Ames Research Center*.

### **Contabilização das Falhas**

Em relação à contabilização de falhas, o trabalho de Tafazoli (2009) considerou como falha todos os incidentes que levaram o satélite a uma degradação permanente ou temporária. Quando as falhas foram do mesmo tipo e afetaram o mesmo subsistema elas foram consideradas como uma falha única; da mesma forma, falhas oriundas de uma única fonte mas que afetaram vários subsistemas também foram consideradas únicas e contabilizadas para o subsistema de origem. Não foram consideradas no trabalho falhas no lançamento.

No trabalho de Robertson e Stoneking (2003) foram consideradas apenas falhas ocorridas durante o tempo de vida previsto para o satélite, não foram consideradas falhas no lançamento ou após o término do tempo de vida útil esperado. Também foram considerados apenas satélites construídos nos Estados Unidos, Canadá, Europa e Japão.

### **Classificação das Falhas**

Como não há um consenso em classificação de falhas em sistemas espaciais, cada trabalho apresentou a sua própria classificação. Aqui serão apresentadas as classes de falhas que foram comuns aos dois trabalhos e consideradas mais relevantes para o nosso estudo.

Nos dois trabalhos, as falhas foram classificadas por: impacto na missão; data de ocorrência da falha relativamente à data de lançamento; causas da falha e subsistemas que as originaram. Como já dito, os dois trabalhos apresentaram também outras classificações que não serão apresentadas aqui. Foi, ainda,

observado que mesmo nas classificações comuns a ambos os trabalhos não houve, como pode ser visto através dos gráficos, um consenso em relação à caracterização de cada classe.

## Resultados Estatísticos

A seguir serão apresentados os resultados que foram considerados mais relevantes e interessantes e que justificaram o estudo ora apresentado.

Tafazoli (2009) identificou e estudou 156 falhas que ocorreram em 129 satélites diferentes entre 1980 a 2005. Já Robertson e Stoneking (2003) analisaram dados de 764 satélites lançados entre os anos de 1990 e 2001.

1) Data de ocorrência da falha em relação ao lançamento e vida útil:

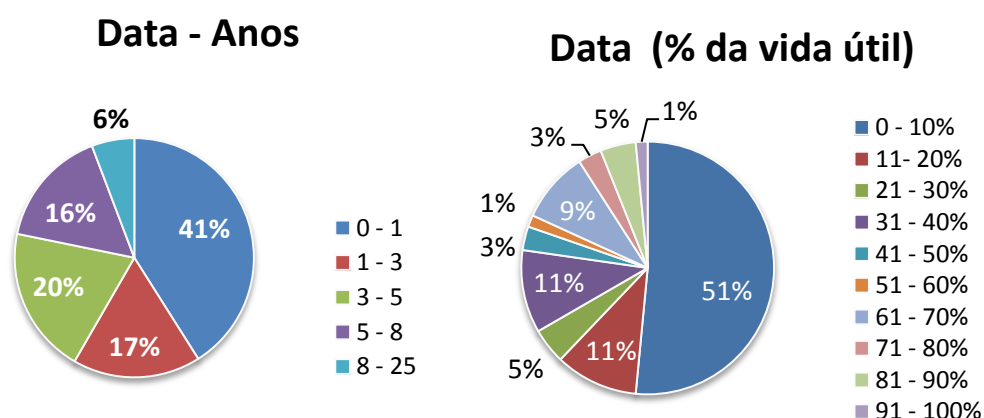


Figura 2.1 - Estatísticas de Falhas – Tempo Após o Lançamento  
 Fonte: Adaptado de Tafazoli (2009); Robertson e Stoneking (2003).

## 2) Subsistemas:

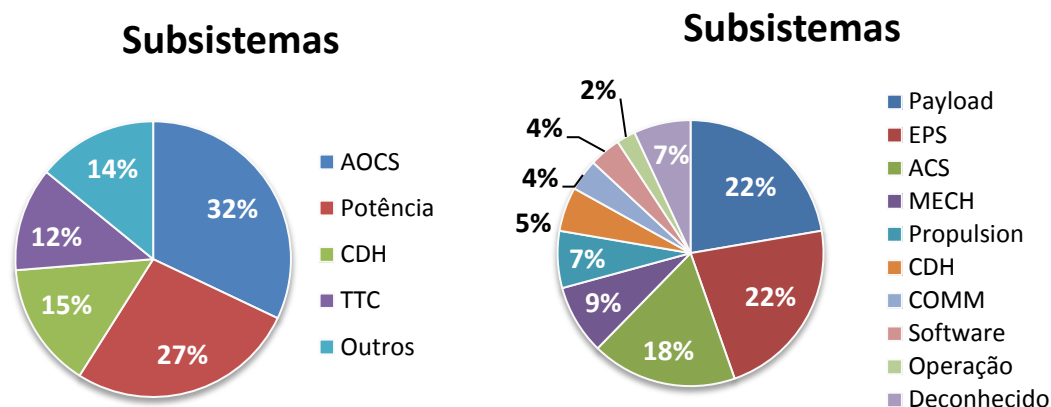


Figura 2.2 - Estatísticas de Falhas – Subsistemas

Fonte: Adaptado de Tafazoli (2009); Robertson e Stoneking (2003).

## Conclusões

Os dois trabalhos apresentaram conclusões semelhantes acerca das falhas estudadas:

- As falhas nas missões espaciais ocorrem muito cedo na missão, normalmente no primeiro ano da mesma (ROBERTSON; STONEKING, 2003; TAFAZOLI, 2009);
- A precocidade das falhas pode indicar que defeitos de fabricação latentes e erros de projeto podem ter um impacto maior que, por exemplo, fadiga e contaminação de materiais (ROBERTSON; STONEKING, 2003);
- A história mostrou que muitos dos resultados de testes que indicavam anomalias haviam sido negligenciados e só foram notados quando foram novamente verificados depois da ocorrência de uma falha em órbita (ROBERTSON; STONEKING, 2003).

Os dois estudos também relataram dificuldades que causaram algumas limitações nos trabalhos apresentados:



- Há dificuldade de acesso a dados, já que muitas anomalias e falhas não são reportadas publicamente (ROBERTSON; STONEKING, 2003; TAFAZOLI, 2009);
- São reportadas em maior número as anomalias muito severas. Outros tipos de anomalias menos severas, cujo entendimento poderia ser útil em futuros projetos, são reportadas com menor frequência (ROBERTSON; STONEKING, 2003);
- Normalmente, as anomalias só são reportadas quando ocorrem durante a vida útil projetada para a missão e a estimativa de vida útil projetada pode ser um dado subjetivo (ROBERTSON; STONEKING, 2003).

Pôde-se observar que não há muitos estudos desse tipo na literatura e os trabalhos apresentados indicaram que há grande dificuldade em se conseguir dados sobre falhas ocorridas já que normalmente esses dados não são públicos, e muitas vezes são subjetivos, reportados sem que se siga nenhuma padronização. Além disso, a comparação entre as metodologias e classificações adotadas nos dois trabalhos mostrou que não há consenso sobre o assunto, cada um dos trabalhos definiu a sua própria metodologia e critério de classificação.



### **3 TÉCNICAS PARA DETECÇÃO DE ANOMALIAS**

Nessa seção serão apresentadas as principais tecnologias que serviram como base teórica para o estudo experimental realizado, entre elas, técnicas de detecção automática de anomalias e os algoritmos de agrupamento utilizados no experimento.

#### **3.1. Detecção Automática de Anomalias**

Detectar anomalias refere-se a encontrar nos dados padrões que não estão em conformidade com uma noção bem definida de comportamento normal (CHANDOLA et al., 2009).

Atualmente, a detecção automática de anomalias é utilizada em diversos domínios: detecção de intrusão em redes de computadores, onde comportamentos anômalos no tráfego podem indicar ataques; detecção de fraudes financeiras, onde o uso de um recurso financeiro fora de um determinado perfil pode indicar uma fraude; detecção de danos em sistemas complexos, processamento de imagens etc.

Essa área de pesquisa apresenta um conjunto importante de desafios, entre eles, a disponibilidade de dados históricos, a definição do que são padrões normais e anômalos e, inclusive, a consideração de que em muitos domínios a noção de normalidade pode evoluir. Nessa linha, pode-se dizer que uma das principais questões associadas à detecção automática de anomalias é a própria caracterização do problema que se deseja abordar, isso significa identificar as características do domínio do problema e dos dados disponíveis e, através dessa identificação, determinar quais seriam as técnicas mais adequadas.

A Figura 2.1 mostra os principais componentes associados à Detecção Automática de Anomalias.

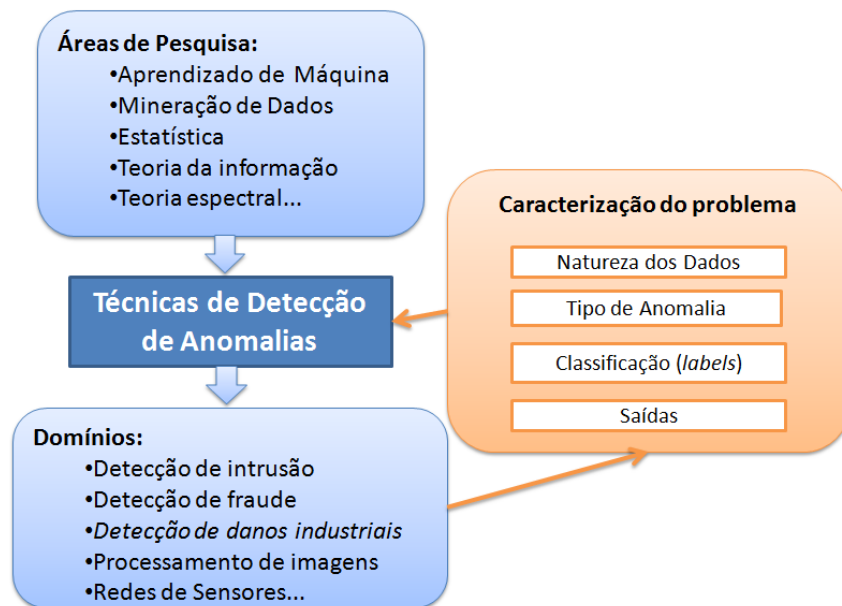


Figura 3.1 – Detecção de Anomalias – Componentes.  
 Fonte: Adaptada de Chandola et al. (2009).

### Caracterização do Domínio

Para que se possa aplicar e escolher as técnicas de detecção automática de anomalias mais adequadas, o domínio do problema deve estar caracterizado no que se refere a alguns aspectos principais, entre eles, deve-se conhecer qual a natureza dos dados disponíveis, quais os tipos de anomalia caracterizam o problema e que tipo de resultados seriam interessantes para o domínio em questão. A seguir é feita uma breve discussão acerca de cada um desses tópicos (CHANDOLA et al., 2009):

#### 1) Natureza dos Dados de Entrada

Dados de entrada normalmente são instâncias de dados descritas por um conjunto de atributos. Alguns aspectos das instâncias e atributos, em muitos casos, guiarão as técnicas a serem utilizadas ou determinarão que uma ou outra técnica não poderá ser aplicada.

Em relação à natureza dos dados pode-se ter a seguinte caracterização: número de atributos (instâncias uni ou multivariadas), tipo dos atributos

(numéricos ou categóricos, discretos ou contínuos), correlação entre as instâncias (nenhuma, dados seqüenciais, espaciais, grafos etc.).

## 2) Tipos característicos de anomalias

Diferentes tipos de anomalia podem caracterizar um determinado domínio de problema. Quanto ao tipo as anomalias podem ser:

- Anomalias pontuais: se uma instância individual pode ser considerada uma anomalia em relação às demais. Ex.: um pico de energia;
- Anomalias de contexto: quando uma instância de dados é anômala em um contexto específico, mas não é em outro. Ex.: temperatura anômala dependendo da estação do ano;
- Anomalias coletivas: quando uma coleção de instâncias de dados é anômala em relação ao restante dos dados. Nesse caso, a coleção de instâncias é anômala, mas cada instância individualmente não é. Ex.: um ataque que exija três comandos em uma dada seqüência, os comandos são corretos, mas a seqüência indica um comportamento anômalo.

## 3) Classe dos dados

Está relacionada ao aprendizado dos algoritmos. Nessa etapa, uma classe associada a uma instância deve indicar se a mesma é anômala ou normal. Normalmente, essa classificação é feita manualmente, mas nem sempre é exequível, já que nem sempre é possível determinar todos os comportamentos normais.

As técnicas de detecção de anomalias são bastante sensíveis em relação ao aspecto aprendido, sendo muitas vezes esse um fator determinante na escolha da técnica a ser utilizada. As técnicas disponíveis trabalham de três formas diferentes:

- Supervisionada: neste caso os algoritmos esperam ser treinados com instâncias definidas como anômalas ou normais;

- Detecção semi-supervisionada: os algoritmos são treinados apenas com instâncias classificadas como normais e qualquer instância que difira do comportamento predominante será considerada anormal;
- Detecção não supervisionada: não requer dados de treinamento, os algoritmos aprendem continuamente, assumindo que comportamentos normais são muito mais frequentes do que comportamentos anormais.

#### 4) Resultados esperados

Esse aspecto define que tipo de resultado seria interessante ou factível em um determinado domínio. Os resultados podem ser classificatórios (instâncias anômalas ou normais) ou podem ser índices probabilísticos (probabilidade de uma instância ser anômala ou não).

Ainda em relação aos resultados esperados, pode-se caracterizar um domínio quanto à sua sensibilidade relativamente à presença de anomalias. Alguns domínios exigem exatidão na determinação da anomalia, p.e., domínio médico, outros exigem ações imediatas e detecção online, p.e., domínios de fraude.

### **Técnicas de Detecção de Anomalias**

A seguir, serão apresentadas algumas das técnicas mais utilizadas na área de detecção automática de anomalias. Essa lista não encerra todas as técnicas usualmente aplicadas, e uma combinação das mesmas poderá ser utilizada de acordo com a disponibilidade dos dados e as necessidades do domínio do problema (CHANDOLA et al., 2009).

Classificação
Algoritmos de classificação devem ter a capacidade de aprender através de um conjunto de instâncias categorizadas (instâncias de treinamento) e então

classificar uma instância de teste em uma das classes aprendidas. No caso da detecção de anomalia pode haver uma ou mais classes normais e uma classe anômala. Há vários algoritmos que podem ser utilizados para detecção de anomalias, entre eles, redes neurais, redes *bayseanas*, algoritmos baseados em regras, etc.

#### Vizinho mais próximo

As técnicas de detecção de anomalia por vizinho mais próximo requerem medidas de distância ou similaridade entre duas instâncias de dados. Essas medidas de distância podem ser realizadas de diferentes formas dependendo dos tipos de atributo (euclidiana, coeficiente de correspondência etc). Para instâncias com vários atributos as medidas são calculadas separadamente e depois combinadas. A técnica trabalha com métodos não-supervisionados ou semi-supervisionados. Assume-se que instâncias de dados normais ocorrem em regiões densas, enquanto as anomalias ocorrem longe de seus vizinhos mais próximos.

#### Agrupamento

As técnicas de detecção de anomalia por agrupamento (*cluster*) assumem que instâncias de dados normais pertencem a um grupo de dados, enquanto as anômalas não pertencem a nenhum grupo, ou que instâncias normais estão próximas aos centros dos grupos enquanto as anômalas estão distantes dos mesmos ou, ainda, que as instâncias normais pertencem a grupos densos e grandes, enquanto instâncias anômalas pertencem a grupos pequenos e pouco densos. Essas técnicas se parecem ligeiramente com as técnicas de vizinho mais próximo, entretanto, diferem nos algoritmos utilizados e na necessidade de identificação de grupos. Os algoritmos normalmente utilizam medidas de distância da instância ao grupo ou índices de probabilidade da pertinência ou não de uma instância a um grupo. São técnicas não supervisionadas ou semi-supervisionadas e a fase de teste é relativamente rápida, já que nessa fase uma instância é comparada aos grupos definidos na fase de treinamento.

Técnicas Estatísticas
Nessas técnicas uma anomalia é uma observação suspeita de ser completa ou parcialmente irrelevante por não ter sido gerada pelo modelo estocástico assumido. Instâncias normais ocorrem em regiões de alta probabilidade de um modelo estocástico, enquanto as anômalas caem em regiões de baixa probabilidade. Várias técnicas estatísticas podem ser utilizadas: modelos gaussianos, regressões, etc. Se as suposições sobre a distribuição de dados são verdadeiras, técnicas estatísticas fornecem uma solução estatisticamente justificável para a detecção de anomalias. Nesse caso, o índice fornecido por uma técnica estatística é associado a um intervalo de confiança, que pode ser utilizado como informação adicional ao se fazer uma análise da anomalia.
Espectrais
Técnicas espectrais tentam encontrar uma aproximação dos dados usando uma combinação de atributos que visa capturar a maior parte da variabilidade dos mesmos. Assim, os dados podem ser reduzidos a um subespaço de atributos no qual dados normais e anômalos parecem bastante diferentes. As técnicas espectrais reduzem a dimensionalidade dos dados, portanto são adequadas para lidar com conjuntos de dados de muitas dimensões. Além disso, elas também podem ser usadas como uma etapa de pré-processamento, de forma combinada, primeiro são utilizadas as técnicas espectrais e em seguida são aplicadas, ao espaço de atributos já reduzido, outras técnicas existentes.

### 3.2. Algoritmos de Agrupamento

Os estudos de caso utilizados nesse trabalho são relativos a anomalias ocorridas no subsistema de Suprimento de Energia do satélite. Nessa etapa inicial de avaliação de algoritmos, os algoritmos de agrupamento foram escolhidos por trabalharem de forma semi-supervionada (treinamento com dados de períodos de operação normal), por serem eficientes na etapa de detecção e porque, intuitivamente, no que se refere ao subsistema de



suprimento de energia, nos pareceu que as instâncias formariam grupos bem definidos de acordo com os macro-estados do satélite: eclipse/iluminado, imageando/não imageando.

Nessa etapa de avaliação, foram escolhidos os algoritmos K-means e *Expectation Maximization* (EM).

### Algoritmo K-means

O algoritmo K-means é um dos algoritmos mais populares e conhecidos de agrupamento. Esse algoritmo adota medidas de dissimilaridade (Euclidiana, Manhattan, etc.) entre um vetor  $X_i$  e o centróide representativo de um grupo para indicar a pertinência ou não do vetor ao mesmo (THEODORIDIS; KOUTRUMBAS, 2003).

Assim, dado um conjunto de observações  $(X_1, X_2, \dots, X_n)$ , onde cada observação é um vetor  $d$ -dimensional, o algoritmo  $k$ -means busca particionar as  $n$  observações em  $k$  conjuntos  $S = \{S_1, S_2, \dots, S_k\}$ , sendo  $(k \leq n)$ , visando minimizar a soma dos quadrados dentro do grupo (WEBB, 2002), sendo  $J$  a função objetivo a ser minimizada (3.1).

$$J = \sum_{j=1}^k \sum_{i=1}^n \|X_i - \mu_j\|^2 \quad (3.1)$$

Onde  $\mu_j$  é a média dos pontos de  $S_j$ .

O K-means é um algoritmo iterativo, baseado em medidas de dissimilaridade, onde, a princípio é definido o número de grupos a serem encontrados:  $k$ . Em seguida,  $k$  pontos são randomicamente determinados como sendo os centróides dos grupos e cada instância é definida como pertencente ao grupo cujo centróide esteja mais próximo. Feito isso, um novo centróide é calculado

para cada grupo e o processo se repete até que não haja mais movimentação das instâncias entre os grupos (WITTEN; FRANK, 2005).

Assim, dado um conjunto inicial de  $k$  médias iniciais  $\mu_1, \dots, \mu_k$ , o algoritmo executa dois passos:

**Passo de designação:** cada instância é designada ao grupo cuja média (centróide) é mais próxima (menor distância);

**Passo de atualização:** para cada novo grupo, é calculada a média que será o novo centróide do mesmo.

Os passos de designação e atualização se repetem até que o algoritmo convirja.

O algoritmo converge quando nenhuma mudança de uma instância para outros grupos é capaz de minimizar a soma dos quadrados dentro dos mesmos (WEBB, 2002).

O algoritmo K-means é bastante sensível à escolha inicial dos centróides dos grupos, uma escolha ruim pode afetar tanto sua eficácia na definição de grupos consistentes, quanto a eficiência da sua execução. Algumas técnicas empíricas podem ser utilizadas para definir tanto o número de grupos quanto o centróide inicial dos mesmos. Uma das técnicas consiste em analisar o atributo de maior variabilidade e escolhê-lo como orientador tanto no que se refere aos grupos quanto aos centróides iniciais (MINGOTI, 2005).

### **Algoritmo *Expectation Maximization* - EM**

*Expectation Maximization* (EM) é um algoritmo cujo fundamento está na teoria estatística de Modelos de Mistura Finitos (*Finite Mixture Models*). O EM baseia-

se no conjunto de K distribuições de probabilidade cada uma descrevendo um grupo e cada grupo tendo uma função de distribuição de probabilidade própria. Essas distribuições de probabilidade podem pertencer à mesma família, mas ser diferentes em relação aos valores assumidos pelos parâmetros da distribuição (WEBB, 2002). No EM cada instância tem uma probabilidade de pertencer ou não a uma dada distribuição ou grupo (WITTEN; FRANK, 2005).

Um modelo de dois grupos com duas funções de distribuição, por exemplo, compreende 5 parâmetros: média do grupo A ( $\mu_A$ ), desvio padrão do grupo A ( $\delta_A$ ), média do grupo B ( $\mu_B$ ), desvio padrão do grupo B ( $\delta_B$ ) e probabilidade  $P_A$  que é a probabilidade da instância pertencer ao grupo A ( a probabilidade  $P_B$ , é calculada a partir de  $P_A$ ).

No algoritmo EM não se conhece à priori a função de distribuição dos grupos, apenas se conhece o número de grupos.

O algoritmo consiste de três passos básicos, sendo um passo inicial e os demais iterativos (WITTEN; FRANK, 2005):

- Para cada grupo são estimados dois parâmetros da função de probabilidade:  $\mu$  e  $\delta$ ;
- Para cada instância é calculada a probabilidade da mesma pertencer aos grupos;
- As probabilidades viram pesos para a próxima estimativa dos parâmetros da função de distribuição de probabilidades (3.2 e 3.3).

$$\mu_A = \frac{w_1 X_1 + w_2 + \dots + w_n X_n}{w_1 + w_2 + \dots + w_n} \quad (3.2)$$

$$\sigma_A^2 = \frac{w_1 (X_1 - \mu)^2 + w_2 (X_2 - \mu)^2 + \dots + w_n (X_n - \mu)^2}{w_1 + w_2 + \dots + w_n} \quad (3.3)$$

O algoritmo converge quando a probabilidade total que mede a qualidade dos grupos deixa de aumentar de uma iteração para outra. Essa probabilidade total é calculada pela multiplicação das probabilidades de cada instância (3.4).

$$PT = \prod_i (P_A \Pr[X_i | A] + P_B \Pr[X_i | B]), \quad (3.4)$$

Na prática este cálculo é realizado através de logaritmo e esse valor é conhecido como *log-likelihood* (DEMPSTER et al., 1977; WITTEN; FRANK, 2005).

### 3.3. Trabalhos Relacionados

Pode-se encontrar na literatura um conjunto de trabalhos que utilizam técnicas de detecção automática de anomalias com dados de telemetria de satélites. Nos trabalhos de Hori et al.(2004); Machida et al. (2006a,b) foram empregados algoritmos de mineração de dados e de regressão para adaptar os limites de valores das telemetrias de forma automática, gerando índices únicos.

Alguns projetos da NASA usaram a ferramenta ORCA e o sistema *Inductive Monitoring System* (IMS). ORCA é uma ferramenta de mineração de dados que analisa bancos de dados multivariados utilizando algoritmos de vizinho mais próximo para identificar *outliers*, enquanto a ferramenta IMS utiliza técnicas de agrupamento para gerar uma base de conhecimento da operação normal do satélite (BAY; SCHWABACHER, 2003; IVERSON, 2008). O IMS é uma ferramenta que aplica algoritmo de agrupamento para extrair dados de modelos que representam a operação normal do satélite. A ferramenta representa, assim, áreas chamadas de regiões de operação normal, essas áreas são armazenadas como base de conhecimento para futura comparação com os dados de tempo-real (IVERSON, 2008).

Nesse estudo experimental, nós utilizamos um modelo similar ao utilizado na ferramenta IMS. Na prática, o nosso objetivo foi avaliar a eficiência das técnicas de agrupamento em identificar anomalias em satélites, comparar a eficiência de dois algoritmos diferentes aplicando-os a dois problemas reais ocorridos com o subsistema de suprimento de energia.



## 4 ESTUDO DE CASOS

No trabalho experimental realizado foram utilizados dois estudos de caso reais que mostraram anomalias ocorridas no subsistema de Suprimento de Energia. A escolha desse subsistema mostrou estar em consonância com as estatísticas apresentada no capítulo anterior, já que nos dois estudos o subsistema de potência foi responsável por uma porcentagem considerável das falhas ocorridas: 27% e 22%.

### 4.1. Subsistema de Suprimento de Energia

O subsistema de energia é o responsável por gerar e condicionar a energia primária do satélite (painéis solares), por armazenar a energia em fontes secundárias de armazenamento (baterias) e por suprir a energia aos demais subsistemas nas diferentes tensões exigidas pelos mesmos (INPE, 1999; LARSON; WERTZ, 1999).

A Figura. 4.1 apresenta um diagrama com os principais equipamentos do Subsistema de energia.

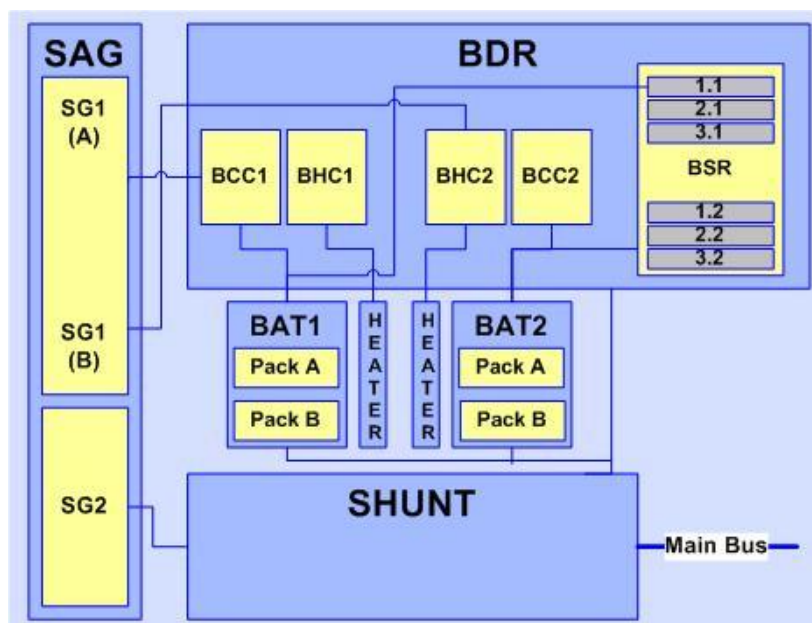


Figura 4.1 - Diagrama do Subsistema de Suprimento de Energia

O subsistema de Suprimento de Energia é composto basicamente de duas fontes de energia: uma fonte primária que converte a luz que incide no painel solar (SAG) durante os períodos de iluminação em energia elétrica; e uma fonte secundária (baterias) que armazena através de processos químicos a energia que será consumida durante os períodos de eclipse. O controle do estado de carga da bateria é realizado pelo circuito de fim-de-carga (*End-of-charge* - EOC), que limita a carga da bateria ligando e desligando a corrente que a carrega. O condicionamento da potência do barramento principal é realizado pelo equipamento SHUNT e pelo Regulador de Descarga da Bateria (*Battery Discharge Regulator* – BDR) que em conjunto controlam a potência fornecida pelos SAGs e pela bateria, provendo o satélite e os conversores DC/DC com um barramento principal estabilizado.

### **Processo de Carregamento das Baterias**

Durante a fase de eclipse ocorre a descarga das baterias já que nessa fase elas compõem a única fonte de suprimento de energia do satélite. Em contrapartida, durante os períodos de iluminação solar, as baterias são carregadas até o limite de carga indicado pela curva de fim-de-carga (*End-of-charge* EOC). O processo de carga da bateria leva um determinado tempo ( $t_0$ ) que é normalmente menor que o tempo total do período de iluminação ( $t_0 + t_1$ ). A Figura 4.2 ilustra o processo de descarga e carga da bateria durante os períodos de eclipse e iluminação solar.



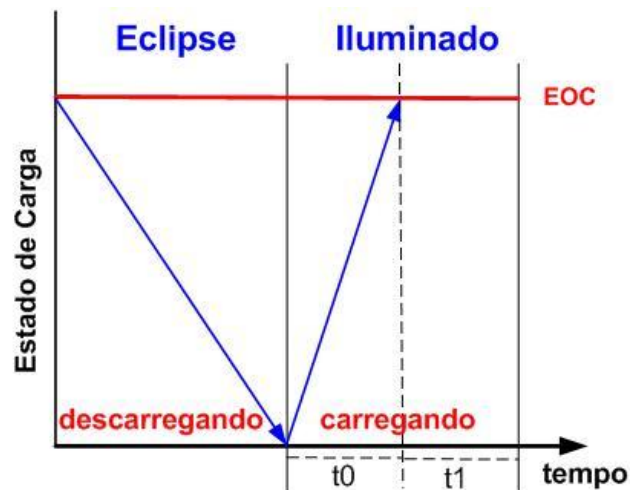


Figura 4.2 - Descarga/Carga da Bateria.

O circuito de carga verifica o nível de descarga da bateria (*Deep-of-discharge* – DOD) e a temperatura da mesma e a carrega até o limite especificado por uma curva de fim-de-carga (curva EOC). O valor de limite de carga é determinado de acordo com a curva escolhida e em função da temperatura observada. Para medir o nível de descarga da bateria, são aplicados pulsos de verificação - pulsos curtos no tempo e com pequena corrente. Cada vez que se detecta que a bateria não está completamente carregada, a sua carga é realizada, checando o limite máximo de tensão (EOC) e observando sua temperatura. Uma vez carregada, durante todo o restante do período de iluminação solar, pulsos de verificação continuam a ser aplicados à bateria, em intervalos regulares, para verificar se a mesma se encontra completamente carregada. Caso seja detectado qualquer nível de descarga, o processo de carga se repete visando garantir que a bateria entre no período de eclipse completamente carregada. A Figura 3.3 apresenta esse processo e as linhas verdes representam os pulsos de verificação.

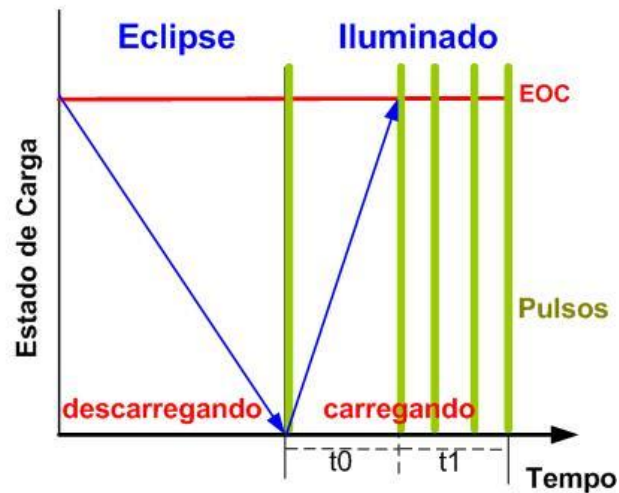


Figura 4.3 - Descarga/Carga da Bateria (Curva EOC).

### Telemetrias do Subsistema de Suprimento de Energia

Os valores obtidos através das telemetrias do satélite indicam o estado dos mesmos. No caso dos satélites utilizados nesse trabalho, há mais de 1000 tipos diferentes de telemetrias para serem analisadas em solo pelos operadores. Essas telemetrias, na sua maioria, indicam valores medidos por sensores e estados de equipamentos (ligados/desligados, principal/redundante).

As telemetrias do subsistema de Suprimento de Energia podem ser classificadas em:

- Analógicas: relativas a valores analógicos, tais como: temperatura, corrente, tensão etc.
- *Bi-level*: valores binários que indicam o estado de equipamentos, chaves, etc.

Nesse experimento foram utilizadas apenas telemetrias analógicas que indicavam medidas de interesse do subsistema de suprimento de energia, especialmente as telemetrias associadas às baterias, ao painel solar, aos

equipamentos SHUNT e BDR. Essas telemetrias foram escolhidas por serem as mais representativas do comportamento do subsistema de potência e assim terem maior potencial para identificar problemas e anomalias. A tabela 4.1 lista as telemetrias que foram utilizadas no experimento.

Tabela 4.1 – Telemetrias do Subsistema de Potência.

Telemetria	Descrição	Limites
TM001	Tensão do Barramento Principal.	27 a 29V
TM002	Corrente do Barramento Principal.	0 a 36A
TM003	Tensão de saída do <i>Main Error Amplifier</i> (MEAS)	8 a 23.5V
TM013/017	Corrente de entrada do BDR	0 a 13A
TM014/018	Tensão das baterias	43.2 a 56.5V
TM015/019	Temperatura das baterias	0 a 10°C
TM016/020	Tensão mínima de grupo de 3 células	3.6 a 4.65V
TM021	Corrente de saída do BDR	0 a 36A
TM022/023	Corrente dos painéis solares (SG1 e SG2)	0 a 7.2A

## 4.2. Estudos de Caso

Para que se pudesse realizar, avaliar e validar o experimento do uso de algoritmos de agrupamento para detecção de anomalias em satélites através de dados de telemetria, foram identificados, com a ajuda de especialistas, duas anomalias distintas ocorridas em dois satélites de missões do INPE, aqui chamados de S1 e S2. Embora as anomalias tenham ocorrido em satélites e épocas diferentes, não estando, portanto, relacionadas entre si, ambas ocorreram no subsistema de fornecimento de energia que é um dos subsistemas mais críticos do satélite.

### Estudo de Caso 1 – Satélite S1

Na operação do satélite S1 foi detectado um problema silencioso no painel solar 1 (SG1-B), que ocorreu durante quase toda a vida útil do satélite. Durante alguns períodos do ano, a telemetria do painel solar SGI (TM023) apresentou valores abaixo do esperado para os períodos de iluminação solar, indicando

problemas no suprimento de energia para a bateria 2. A Figura 4.4 mostra o comportamento da telemetria TM023 durante toda a vida útil do satélite S1. Na figura o eixo y (ordenada) representa os meses transcorridos desde o lançamento do satélite (Outubro, 1999), enquanto o eixo x (abscissa) representa os valores de corrente em Ampere medidos através da telemetria. A frequência das medidas no gráfico pode ser observada através das cores: cores mais escuras representam uma maior ocorrência de um dado valor no tempo.

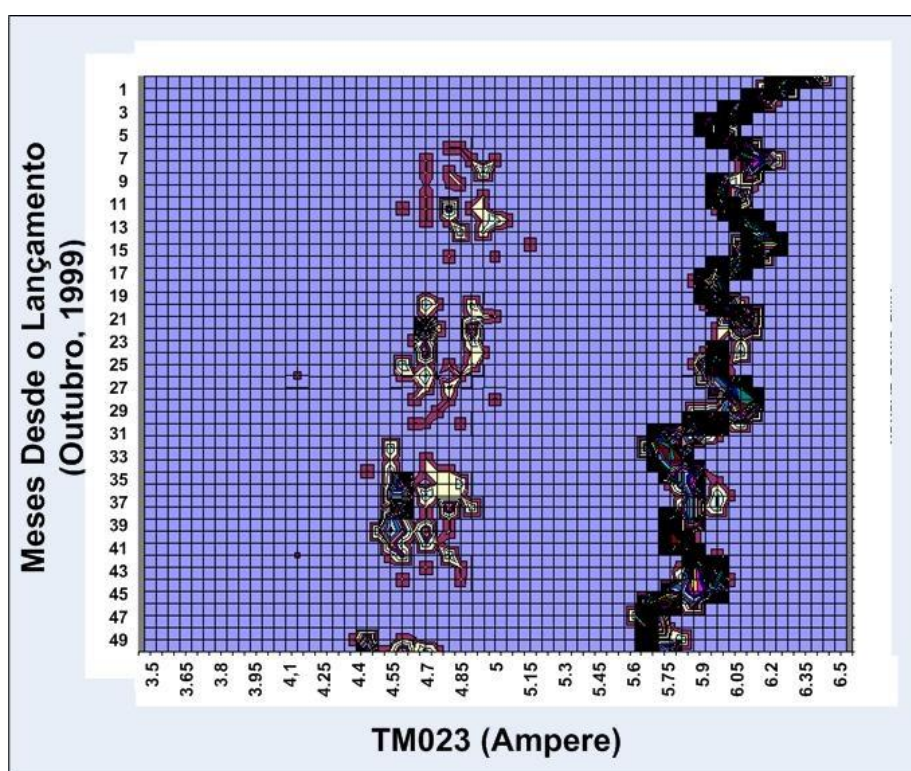


Figura 4.4 - Comportamento da telemetria TM023 (S1).

Como pode ser observado na Tabela 4.1, os valores normais para a telemetria TM023 estão entre os limites 0 e 7.2 A. Os valores dessa telemetria estão sempre próximos a zero durante o período de eclipse quando o painel solar não recebe nenhuma iluminação e próximos a 6A durante o período de iluminação.

Analisando o comportamento da telemetria TM023 durante a vida útil do satélite, pode-se fazer as seguintes observações:

- Nas regiões de valores normais (próximos a 6A) há uma diminuição gradual no valor da telemetria. Esse é um efeito natural que advém da degradação do painel solar;
- Há valores intermediários não esperados para períodos de eclipse e para períodos iluminados;
- Com o tempo, a frequência de valores não esperados aumentou;
- Normalmente, de Janeiro a Março de cada ano, por um efeito sazonal, esses valores intermediários não esperados não ocorreram.

Através da análise dos valores das telemetrias, os especialistas concluíram que a causa mais provável para a ocorrência desses valores não esperados na telemetria TM023 era a redução da corrente do SAG devido a um problema intermitente de mau contato.

## **Estudo de Caso 2 – Satélite S2**

Nesse estudo de caso, a falha foi observada em dois períodos diferentes durante a vida útil do satélite S2. Do ponto de vista do subsistema de suprimento de energia, essa não foi uma anomalia primária, mas sim uma anomalia decorrente de problemas no subsistema de Controle Atitude e Órbita (AOCS). O problema original no AOCS levou o satélite a um estado de emergência e nesse estado a prioridade é garantir o suprimento de energia visando manter o satélite até que o problema fonte seja resolvido. Para atingir esse objetivo, melhora-se o apontamento do painel solar e, para economizar energia, todos os equipamentos são desligados.

Como mencionado anteriormente, as baterias do satélite são carregadas durante o período de iluminação solar para serem utilizadas durante os períodos de eclipse. O carregamento da bateria é realizado através do BCC e

do circuito EOC. O circuito aplica pulsos de corrente para verificar o nível de descarga da bateria (DOD) e quando necessário a bateria é carregada.

Em estado de emergência as baterias carregam mais rápido que o normal, já que os painéis solares estão mais bem apontados para o sol e os equipamentos desligados. Devido à combinação de dois fatores – a aplicação dos pulsos de verificação por um período mais longo que o normal dado o estado de emergência e um problema com a duração dos pulsos – uma das baterias começou a receber uma carga maior que se transformava parcialmente em calor. Como esse excesso de calor não era totalmente dissipado através do sistema térmico, a cada período de iluminação a bateria se aquecia um pouco mais. A Figura 4.5 mostra o carregamento das baterias nesse estado anômalo.

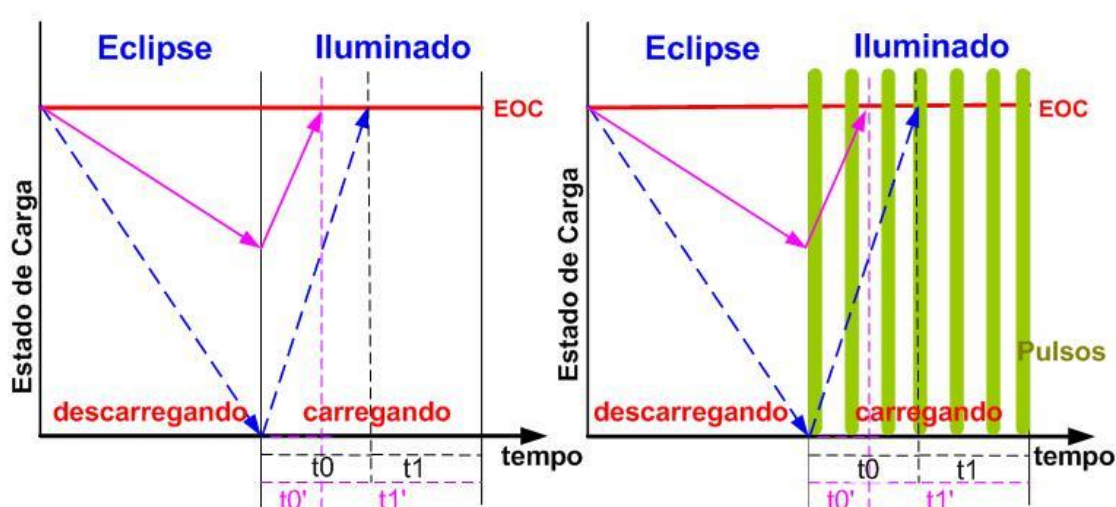


Figura 4.5 - Carregamento da bateria – Estado de Emergência.

Como mostrado, em cada processo de carga, o circuito EOC indica o nível de carga da bateria, que é calculado através da curva EOC em função da temperatura da mesma. Quando o erro ocorreu, durante algumas passagens, a temperatura da bateria foi aumentando, mas foi também sendo corrigida através da curva EOC. Entretanto, com o passar do tempo, a temperatura atingiu o valor máximo a partir do qual a curva deixou de compensar a corrente

aplicada. A partir desse ponto, sem a compensação através da curva EOC, o problema de aquecimento da bateria tornou-se mais sério e o satélite poderia ter sido perdido.

Esse problema foi caracterizado pelos analistas como catastrófico e ocorreu durante apenas algumas passagens do satélite. Esta falha poderia ter sido observada através das tendências nas telemetrias relacionadas à bateria 2. A falha passou a ser percebida apenas quando essas telemetrias saíram dos limites esperados e mostraram-se em estado de alarme para os operadores.





## 5 EXPERIMENTOS

Essa seção apresenta os resultados do estudo experimental realizado através da aplicação dos algoritmos de agrupamento K-means e EM para detecção de anomalias em telemetrias de satélite. Para avaliar a eficiência e desempenho dos algoritmos de agrupamento escolhidos, o estudo foi dividido em duas fases. Na primeira fase, os dados de telemetria foram simulados para garantir o controle sobre os resultados dos algoritmos. Na segunda fase, foram utilizados dados reais de telemetria dos períodos anômalos descritos nos estudos de caso 1 e 2, visando avaliar a capacidade dos algoritmos em detectar as anomalias conhecidas e também a performance dos mesmo relativamente à detecção.

### 5.1. Resultados com uso de Dados Simulados

Visando avaliar os algoritmos de agrupamento em um ambiente controlado, e através do entendimento do comportamento das telemetrias nos dois estudos de caso apresentados, foram gerados dados simulados das telemetrias do subsistema de Suprimento de Energia. Os dados foram gerados aleatoriamente de acordo com os tipos valores esperados e limites estabelecidos de acordo com a Tabela 4.1. e dentro de um número conhecido e pré-definido de grupos. O número de grupos foi escolhido de acordo com os possíveis macro-estados do satélite: iluminado/não iluminado, imageando/não imageando.

O uso de dados simulados permitiu verificar a capacidade dos algoritmos em dividir corretamente os dados de telemetria em grupos, e permitiu, também, avaliar o comportamento dos algoritmos em situações onde, relativamente à determinadas características, os dados se comportavam de forma semelhante aos dados reais. Para testar, ainda, as limitações dos algoritmos foram introduzidas algumas características nos dados simuladores, como por exemplo, tendências nos valores.

As variações nas tendências foram simuladas para avaliar como os algoritmos se comportariam, em especial o algoritmo K-means que utiliza medidas de distância entre a instância e o centróide do grupo, em uma situação onde a tendência dos valores em uma telemetria, compensaria a tendência contrária de valores em outra.

Para avaliar a capacidade do algoritmo em agrupar corretamente os dados de telemetria, os dados simulados foram gerados com um atributo que indicava o grupo ao qual o mesmo efetivamente pertencia. Desta forma, foi possível comparar o grupo escolhido através do algoritmo de agrupamento com a classe pré-definida para a instância.

A Tabela 5.1 mostra os resultados da aplicação dos algoritmos em termos de porcentagem de acerto dos mesmos. Os algoritmos EM e K-means foram avaliados com relação às distâncias Euclidiana e Manhattan.

Tabela 5.1 – Avaliação dos Algoritmos de Agrupamento.

Algoritmo	Porcentagem de Detecções Corretas (Grupos Simulados x Grupos Identificados)	
	Mesma Tendência	Tendência Alternada
EM	100%	100%
K-means (Euclidiana)	100%	62.25%
K-means (Manhattan)	100%	62.70%

Na tabela 5.1 podemos observar que, como era esperado, as tendências alternadas entre os valores das telemetrias tiveram um impacto maior no algoritmo K-means que utiliza medida de distâncias, já que as distâncias tendem a se compensar mutuamente. Esse efeito não foi sentido no algoritmo EM que utiliza distribuição de probabilidade.

Depois de realizar a avaliação com dados simulados corretos (treinamento e teste), foram gerados dados anômalos que simularam os dados das anomalias reais em alguns aspectos específicos. Nessa fase, tentou-se simular as

características principais apresentadas nos casos de estudo descritos anteriormente. Os dados de treinamento com comportamento normal – 5000 instâncias - foram gerados com as mesmas características descritas acima (mesma tendência ou tendência alternada). E os dados de teste – 1500 instâncias – foram gerados com as seguintes características:

- Número de telemetrias anômalas: uma única telemetria anômala dentro do conjunto (simulando o estudo de caso 1); várias telemetrias anômalas - 7 nesse caso;
- Tipo de anomalia presente na telemetria: telemetrias com valores em alarme, ou seja, valores fora dos limites especificados (simulando o estudo de caso 2); telemetrias com valores entre grupos, ou seja, valores dentro dos limites mas fora dos limites esperados para um grupo (simulando o estado de caso 1).

Ambos os algoritmos foram treinados com os dados corretos e testados com os dados anômalos simulados. Nos dados de testes, os dados com valores anômalos estão entre as instâncias 500 e 1000 (dentre as 1500 geradas).

As Figuras 5.1 e 5.2 mostram os resultados para os dados simulados. Nas figuras, as linhas vermelhas representam os resultados obtidos quando o algoritmo foi testado com dados onde havia uma única telemetria anômala, as linhas azuis representam os resultados obtidos quando os algoritmos foram testados com dados onde havia várias telemetrias anômalas. As figuras demonstraram que, na maioria dos casos, os algoritmos detectaram as instâncias anômalas já que os índices de dissimilaridade dessas instâncias em relação aos grupos era maior do que aqueles observados quando analisadas instâncias normais, entretanto, analisando os resultados podemos fazer as seguintes observações:

- Os dois algoritmos foram pouco eficientes em reconhecer anomalias quando as mesmas ocorreram em uma única telemetria. Pode-se observar que nesses casos os índices de dissimilaridade são

praticamente os mesmos que os observados em instâncias normais, mostrando vários potenciais falso-negativos. Entretanto, o algoritmo EM mostrou-se um pouco mais eficiente;

- Foi mais fácil para os dois algoritmos detectarem anomalias quando os valores das telemetrias saíram fora dos limites (alarme) do que quando se encontravam nas regiões intergrupos;
- Como pode ser observado na Figura 5.2, o algoritmo K-means não funcionou para o caso de valores intergrupos.

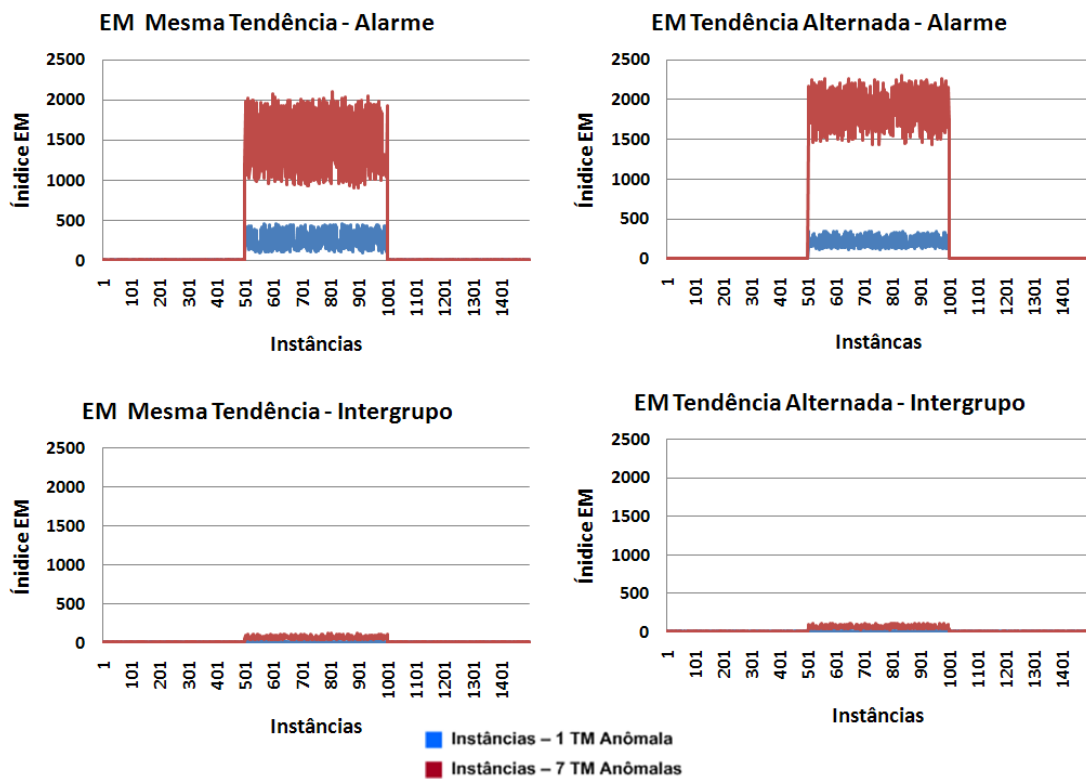


Figura 5.1 - Dados Simulados – Algoritmo EM.

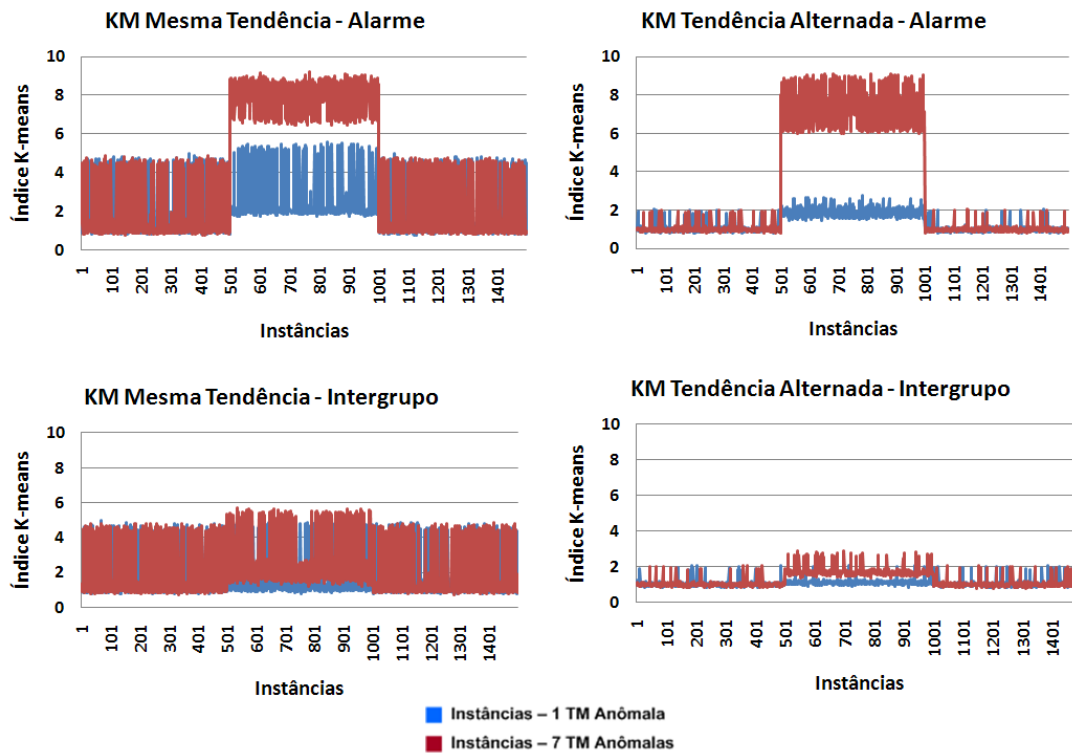


Figura 5.2 - Dados Simulados – Algoritmo K-means.

## 5.2. Resultados com uso de Dados Reais

Esta seção mostra a aplicação dos algoritmos sobre os dados reais de telemetrias tanto em períodos normais quanto anômalos. Essa fase do experimento foi conduzida em dois passos conforme explicado a seguir.

No primeiro passo, os algoritmos foram treinados e testados com dados de períodos de operação normal. Nesse passo foram realizados dois experimentos diferentes:

1. Os algoritmos foram treinados e testados com dados de períodos normais e próximos, tanto na fase inicial de operação do satélite quanto em fases posteriores;
2. Os algoritmos foram treinados com dados do início de vida do satélite e depois testados com dados de anos posteriores.

Os algoritmos de agrupamento se comportaram dentro do esperado nos períodos de operação normal. Em ambos os Estudos de Caso os índices de dissimilaridade ficaram dentro de limites normais, isto é, eram baixos indicando a pertinência das instâncias de teste aos grupos estabelecidos na fase de treinamento. A Figura 5.3 mostra o experimento para o satélite S2. Resultados semelhantes também foram observados no satélite S1.

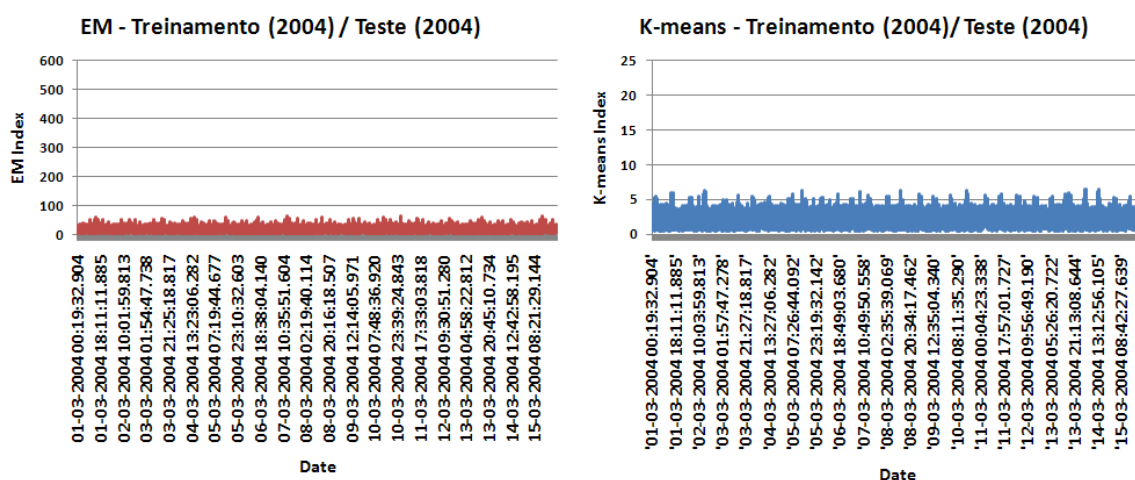


Figura 5.3 - Estudo de Caso 2 – Treinamento (2004)/Teste(2004).

Através dos experimentos foi possível observar que os índices de dissimilaridade obtidos quando o algoritmo foi treinado com dados do início de vida do satélite e depois testado com dados de períodos normais de anos posteriores, foram muito mais altos, mostrando dados corretos como se fossem anômalos, isto é, indicando falsos positivos (Figura 5.4). Isso ocorreu especialmente no estudo de Caso 2, onde as características das baterias mudaram consideravelmente ao longo dos anos.

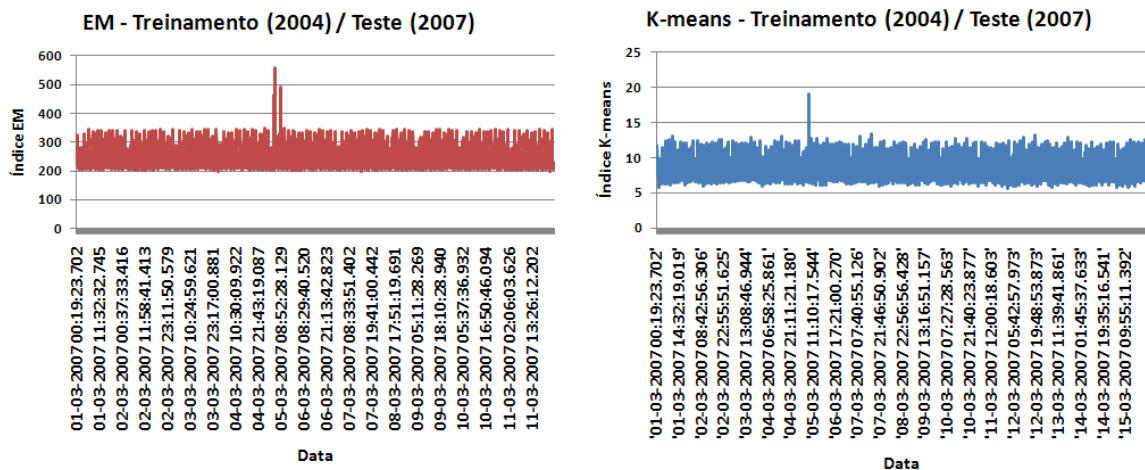


Figura 5.4 - Estudo de Caso 2 – Treinamento (2004)/Teste(2007).

No segundo passo da avaliação experimental, os algoritmos foram treinados com dados de operação normal e testados com dados de períodos anômalos. Esse passo visa avaliar a capacidade de detecção de anomalias dos algoritmos escolhidos.

No estudo de caso 1, o comportamento dos dois algoritmos mostrou-se muito semelhante. De fato, os algoritmos não foram capazes de indicar o comportamento incorreto da TM023 que fora observado pelos especialistas. Isso mostrou que ambos os algoritmos falharam em detectar anomalias quando apenas uma telemetria estava em desconformidade em relação aos valores esperados, mas dentro dos limites estabelecidos. E no caso do satélite S1 indicaram muitos falso positivos. Isso confirmou o que havia sido observado através dos dados simulados: uma única telemetria com valores incorretos na região intergrupos não é corretamente identificada. A Figura 5.5 mostra o resultado dos dois algoritmos para o Estudo de Caso 1.

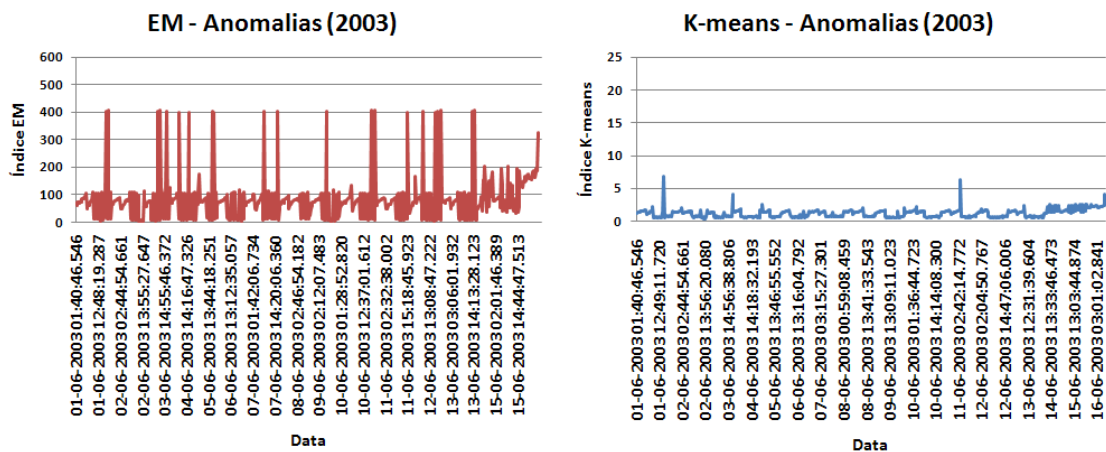


Figura 5.5 - Estudo de Caso 1 – Detecção de Anomalia (2003).

Na Figura 5.6, relacionada ao Estudo de Caso 2, pode-se observar que os dois algoritmos – EM e K-means – foram capazes de detectar os períodos de anormalidade. Isso pode ser observado pelos índices de dissimilaridade que tenderam a crescer nos períodos anômalos, permitindo a observação precoce da anomalia.

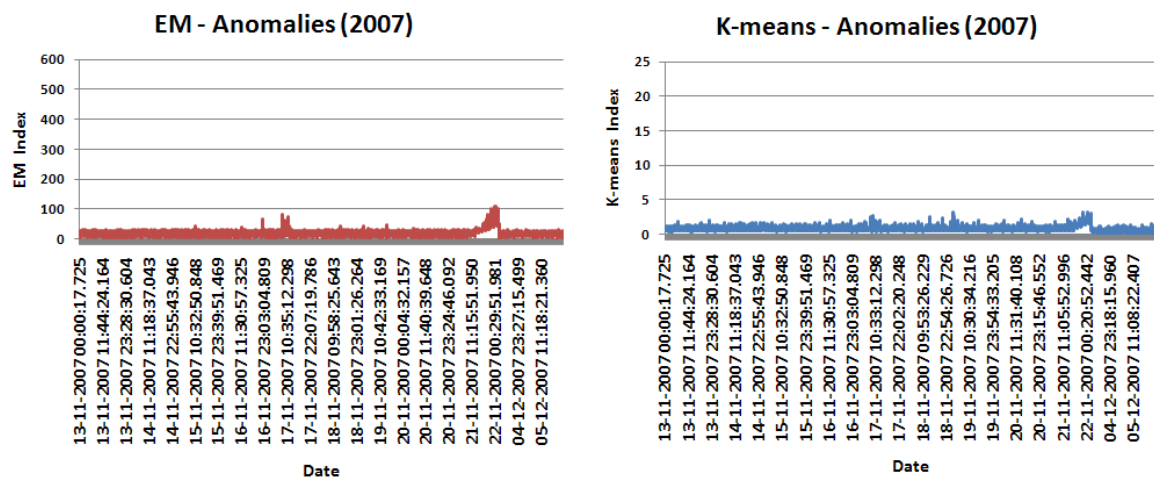


Figura 5.6 - Estudo de Caso 2 – Detecção de Anomalia (2007).

A Figura 5.7 compara o valor da telemetria com os índices EM de dissimilaridade mostrando a incapacidade do algoritmo na detecção da



anomalia no caso do satélite S1. Como pode ser visto na figura, o algoritmo não detecta a região de baixos valores na telemetria TM023.

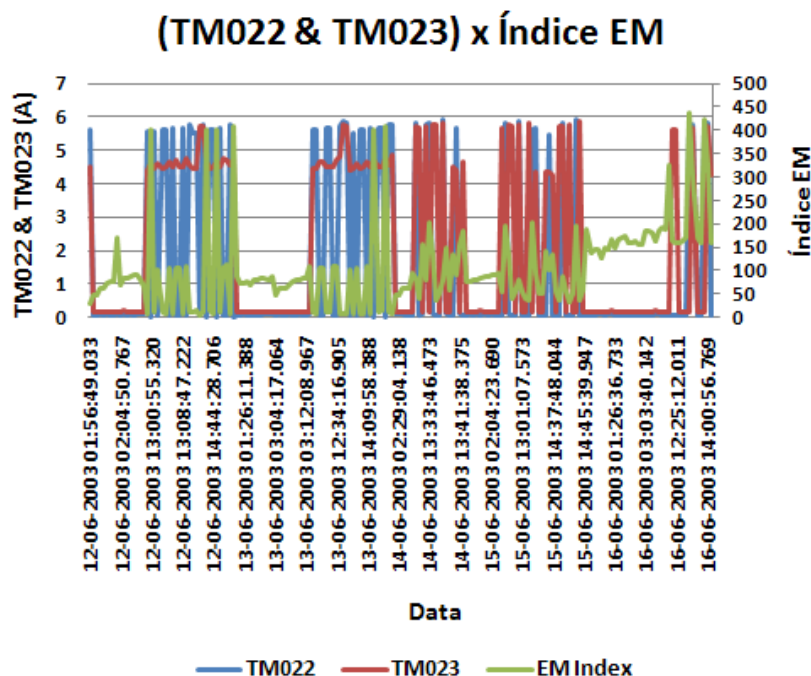


Figura 5.7 - Estudo de Caso 1 – Resultado Comparativo.

A Figura 5.8 mostra a comparação entre a telemetria de temperatura da bateria e o índice de dissimilaridade apontado pelo algoritmo. No caso do algoritmo EM esse índice começou a crescer 6 horas antes do alarme ter sido indicado pela telemetria TM015. Assim, o algoritmo pôde antecipar em algumas horas o aparecimento do problema através de um único índice.

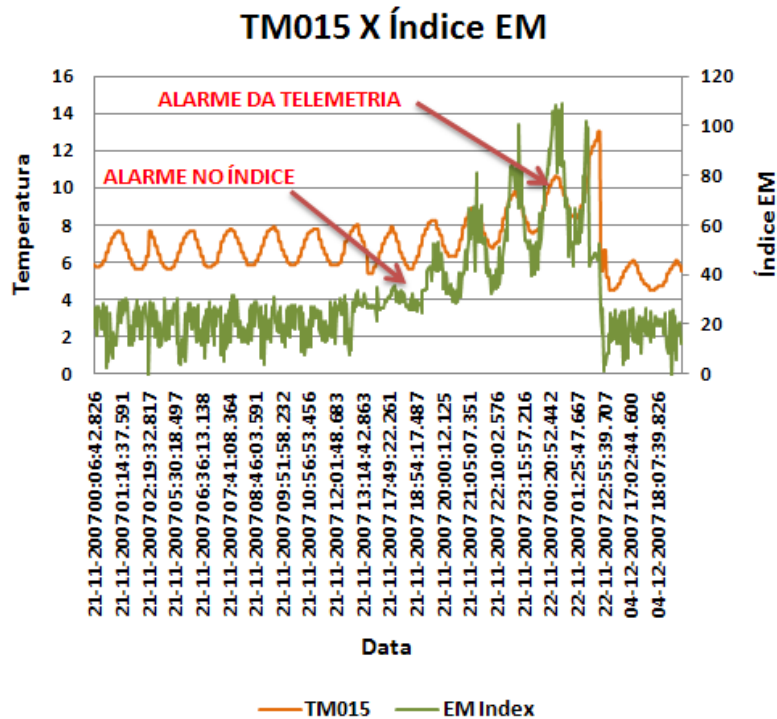


Figura 5.8 - Estudo de Caso 2 – Resultado Comparativo.

### 5.3. Lições Aprendidas

Os pontos a seguir resumem as lições aprendidas do estudo experimental conduzido com os algoritmos de agrupamento EM e K-means, tanto em ambiente controlado (dados simulados) quanto em ambiente real:

1. Os experimentos com dados simulados mostraram que o algoritmo EM, que realiza o processo de agrupamento através de *Mixture Models*, baseado no comportamento estatístico de cada atributo, é menos vulnerável que o algoritmo K-means à falta de normalização nos dados e às tendências variadas nos valores das telemetrias. Em todos os casos o algoritmo apresentou um resultado ligeiramente melhor;
2. Através do uso do algoritmo de agrupamento os operadores poderiam monitorar apenas um ou poucos índices (medidas de dissimilaridade ou desvio do grupo conhecido) para observar tendências de anomalia nos dados. Essa pode ser uma contribuição à rotina de operação, já que hoje os

- operadores têm de monitorar tendências e desvios em um grande número de telemetrias ao mesmo tempo, o que pode ser extremamente difícil;
3. Relativamente ao Estudo de Caso 1, os algoritmos não foram capazes de detectar as anomalias descritas pelos especialistas, apontando inclusive falsos positivos. Embora no estudo tenham sido utilizadas apenas 16 Telemetrias, uma pequena variação de valores em apenas uma delas quando combinada com as diferenças nas demais telemetrias, foi imperceptível para os algoritmos. Se tivesse sido utilizado um número maior de telemetrias, o problema teria sido ainda pior. Assim, o uso de técnicas de agrupamento para detecção de anomalias em satélites, onde há um grande número de telemetrias, poderia ser combinado com técnicas de seleção de atributos, visando a minimização desse tipo problema;
  4. Quando foram utilizados dados reais de anomalia no Estudo de Caso 2, onde a telemetria excedeu os limites esperados, os dois algoritmos detectaram a anomalia com uma antecipação razoável se comparado ao momento em que a telemetria em questão geraria um alarme para os operadores. Entretanto, quando observamos os dados, percebemos que seria possível notar uma tendência anômala antes da mesma ter sido apontada pelos algoritmos;
  5. Os algoritmos de agrupamento mostraram-se sensíveis às mudanças de comportamento das telemetrias em função do envelhecimento do satélite, assim seria necessário aprendizado contínuo visando corrigir esse problema. Entretanto aprendizado contínuo pode levar os algoritmos a serem treinados com dados que apresentem algum grau de anormalidade, já que é difícil à priori distinguir períodos de operação normal de períodos de operação anômala;
  6. Ambos os algoritmos utilizam modelos Gaussianos para fazer os agrupamentos, e ambos funcionaram de forma similar nos dois estudos de caso. O algoritmo EM funcionou ligeiramente melhor quando foram simulados dados que apresentavam tendências variadas nos valores e valores que se situavam nas faixas de intergrupos.



## 6 CONCLUSÃO

Esse trabalho avaliou o uso de dois algoritmos de agrupamento em relação à capacidade dos mesmos em detectarem anomalias através de dados históricos reais de telemetrias de satélite.

O estudo experimental mostrou que os algoritmos de agrupamento foram eficientes em antecipar anomalias quando aplicados a situações onde mais de um atributo, nesse caso telemetrias, comportaram-se de forma anômala, e pelo menos uma delas saiu dos limites especificados. Duas vantagens puderam ser observadas: a própria antecipação em horas na detecção do problema e a observação do mesmo através de um índice único que poderia ser utilizado como auxiliar no monitoramento de telemetrias. Entretanto, os algoritmos não se mostraram eficientes para pequenos e silenciosos desvios em apenas uma telemetria, mostrando que o problema na seleção e combinação de atributos pode ser uma questão importante já que muitos problemas sérios no satélite podem ser observados através de comportamento semelhante.

Outra deficiência no uso de algoritmos de agrupamento para detecção de anomalias é que os mesmos mostraram-se sensíveis às mudanças de comportamento das telemetrias em função do envelhecimento natural dos satélites. Nesse sentido, seria necessário aprendizado permanente o que poderia introduzir problemas, já que o algoritmo passaria a aceitar como dados normais e de treinamento dados que já apresentassem certas variações anômalas, ou seja, durante o treinamento dados anômalos poderiam ser considerados normais causando falsos negativos.

Os satélites são construídos para não falhar e, assim, falhas reais não são muito comuns e, além disso, não há disponibilidade de dados de um grande número de satélites. Nesse contexto, o uso de dados simulados de falhas

típicas, bem como, o uso de dados de campanhas de teste poderiam ser de grande valia na avaliação de técnicas para detecção automática de anomalias.

Como trabalhos futuros, planeja-se a avaliação de outros algoritmos de agrupamento, bem como, de outras técnicas de detecção de anomalia. Dado o grande número de telemetrias a serem analisadas no contexto real de operação de um satélite, planeja-se, também, avaliar o uso de algoritmos de detecção de anomalias associados a algoritmos destinados à seleção e combinação de atributos.

## REFERÊNCIAS BIBLIOGRÁFICAS

BAY, S. D.; SCHWABACHER, M. Mining Distance-Based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule. In: Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly Detection: A Survey, ACM Computing Surveys, 2009.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. **Maximum likelihood from incomplete data via the EM algorithm.** Journal of the Royal Statistical Society: Series B. 39(1):1–38. Nov. 1977.

ECSS-E-10-02A. ECSS Space Engineering – Verification. November, 1998.

HORI, K.; ISHIHAMA, N.; NAKASUKA, S.; OGASAWARA, S.; YARI, T. Summarization of Spacecraft Telemetry Data by Extracting Significant Temporal Patterns. In: The Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2004.

INPE. Coordenadoria de Engenharia e Tecnologia Espacial (ETE). Curso de Tecnologia de Satélites (notas). INPE, 1999.

IVERSON, D. Data Mining Applications for Space Mission Operations System Health Monitoring. In Proceedings of the SpaceOps Conference, 2008.

LARSON, W.J.; WERTZ, J.R. **Space Mission Analysis and Design**, 3<sup>rd</sup> ed. Microcosm Press, 1999. p. 969.

MACHIDA, K.; FUJIMAK, R.; YARI, T. **Adaptive Limit-Checking for Spacecraft Using Sequential Prediction Based on Regression Techniques.** Journal of the Japan Society for Aeronautical and Space Sciences, Volume 54, Issue 630, 2006, pp. 312-318.

MACHIDA, K.; FUJIMAK, R.; YARI, T.; KAWAHARA, Y.; SATO, Y. Telemetry-mining: A machine Learning Approach to Anomaly Detection and Fault Diagnosis for Space Systems. In: 2<sup>nd</sup> IEEE International Conference on Space Mission Challenges for Information Technology, IEEE, 2006.

MINGOTI, S.A. **Análise de Dados Através de Métodos de Estatística Multivariada:** Uma Abordagem Aplicada, 1<sup>o</sup> ed. Belo Horizonte: Editora UFMG, 2005. p.295.

ROBERTSON, B.; STONEKING, E. Satellite GN&C Anomaly Trends. In: AAS Guidance and Control Conference, 2003.

TAFAZOLI, M. A study of on-orbit spacecraft failures. *Acta Astronautica* 64, 2009. p. 195–205.

THEODORIDIS, S.; KOUTRUMBAS, K. **Pattern Recognition**. USA: Elsevier Science, 2003. p.689.

WEBB, A. **Statistical Pattern Recognition**, 2<sup>nd</sup> ed. John Wiley & Sons, LTD, 2002. pp.377.

WITTEN, I.H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**, 2<sup>nd</sup> ed. São Francisco: Elsevier, 2005. p. 524.

.



