



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Faculdade de Engenharia

Valdenir Robson Tavares

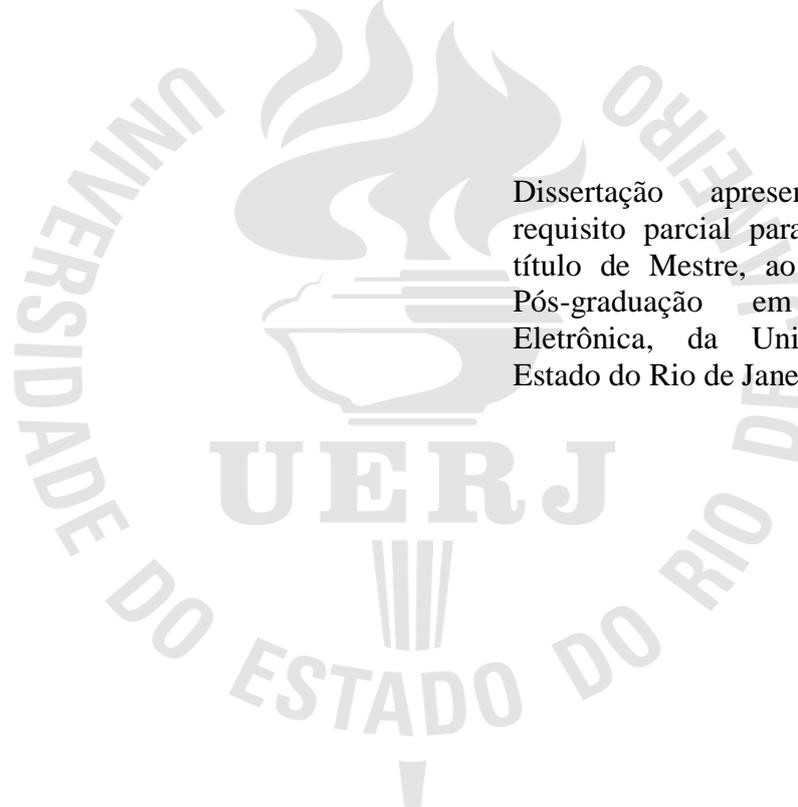
**Proposta de um Modelo para Detecção e Classificação de Falhas em
Redes do Sistema Móvel Pessoal (SMP)**

Rio de Janeiro

2016

Valdenir Robson Tavares

Proposta de um modelo para detecção, classificação e localização de falhas em redes do Sistema Móvel Pessoal (SMP)



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-graduação em Engenharia Eletrônica, da Universidade do Estado do Rio de Janeiro.

Orientadores: Prof. Dr. Alexandre Sztajnberg

Prof. Dr. Jorge Luís Machado do Amaral

Rio de Janeiro

2016

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/B

T231 Tavares, Valdenir Robson.

Proposta de um modelo para detecção, classificação e localização de falhas em redes do Sistema Móvel Pessoal (SMP) / Valdenir Robson Tavares. – 2016.
99f.

Orientadores: Alexandre Stajnberg e Jorge Luís Machado do Amaral.

Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Faculdade de Engenharia.

1. Engenharia Eletrônica. 2. Redes neurais (Computação) - Dissertações. 3. Sistemas de comunicação móvel - Dissertações. I. Stajnberg, Alexandre. II. Amaral, Jorge Luís Machado do. III. Universidade do Estado do Rio de Janeiro. IV. Título.

CDU 004.032.26

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação.

Assinatura

Data

Valdenir Robson Tavares

**Proposta de um modelo para detecção e classificação de falhas em redes do Sistema
Móvel Pessoal (SMP)**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-graduação em Engenharia Eletrônica da Universidade do Estado do Rio de Janeiro.

Aprovada em 31 de Março de 2016.

Banca Examinadora:

Prof. Dr. Alexandre Sztajnberg

Universidade do Estado do Rio de Janeiro - UERJ

Prof. Dr. Jorge Luís Machado do Amaral

Universidade do Estado do Rio de Janeiro - UERJ

Prof. Dr. André Vargas Abs da Cruz

Universidade Estadual da Zona Oeste - UEZO

Prof. Dr. Célio Vinicius Neves de Albuquerque

Universidade Federal Fluminense - UFF

Rio de Janeiro

2016

AGRADECIMENTOS

Participar de um programa de Mestrado é um grande desafio e vencer esse desafio exige esforço, dedicação e muito trabalho de equipe. Por isso, agradeço a cada um dos membros dessa equipe que tanto colaborou para o sucesso desse trabalho.

Em primeiro lugar a meu saudoso pai, que me ensinou a importância da educação e me fez perceber o seu efeito transformador. Foi aprendendo com seus exemplos e seguindo suas orientações, até hoje presente em minhas memórias, que mesmo nos momentos mais difíceis consegui forças para superar minhas dificuldades.

Agradeço aos Professores e orientadores Alexandre Sztajnberg e Jorge Luís Machado do Amaral que me ajudaram na construção e organização do raciocínio usado como base nessa dissertação e também por exigirem de mim o máximo de compromisso e qualidade em cada uma das etapas necessárias a conclusão do curso.

Por fim, agradeço ao Programa de Pós-Graduação em Engenharia Eletrônica da UERJ pela oportunidade que me deram de fazer parte de um ambiente rico em oportunidades para receber conhecimento e trocar experiência com pessoas de diferentes áreas do universo acadêmico.

RESUMO

TAVARES, Valdenir Robson. *Proposta de um modelo para detecção e classificação de falhas em redes do Sistema Móvel Pessoal (SMP)*. 2016. 99 f. Dissertação (Mestrado em Engenharia Eletrônica) – Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

Este trabalho propõe o uso de um modelo para identificação e classificação de falhas em redes móveis (SMP) de 2º e 3º geração. O modelo foi desenvolvido a partir das características encontradas nos contadores de eventos de rede (Códigos de Liberação) produzidos pelas MSC (*Mobile Switching Center*), equipamentos responsáveis por realizar as conexões dos assinantes móveis a outros assinantes. Os códigos de Liberação são coletados das MSCs e organizados em tabelas para compor a massa de dados a ser analisada. A análise da massa de dados inicia-se pela distribuição dos dados em grupos com a utilização de Mapas auto-organizáveis. Cada um dos grupos é analisado pelo especialista do sistema que identifica as características comuns aos dados que compõem cada grupo e realiza a classificação de acordo com a falha detectada ou evento representado, a classificação é então incorporada ao modelo. Os resultados da classificação na saída do modelo são aferidos e avaliados para demonstrar a capacidade do modelo em identificar e classificar as falhas de rede.

Palavras-chave: Redes Móveis; Redes Neurais; Clusterização; Classificação.

ABSTRACT

TAVARES, Valdenir Robson. *Towards a fault detection and classification model for Private Mobile Network Systems (PMNSs)*. 2016. 99 f. Dissertação (Mestrado em Engenharia Eletrônica) – Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

This work proposes the use of a model to identify and classify failures in mobile networks (SMP) of 2nd and 3rd generation. The model was developed based on the features found in network event counters (Clear Codes) produced by the MSC (Mobile Switching Center), equipment responsible for making the connections of mobile subscribers to other subscribers. Clear Codes are collected from MSCs and organized in tables to compose the mass of data to be analyzed. The analysis of the mass of data starts by clustering data into groups using self-organizing maps. Each group is analyzed by an operation and maintenance expert that identifies the common characteristics of data and performs a classification according to the detected fault. Then, classification is incorporated into the model. The classification results in the model output are measured and evaluated to demonstrate the model's ability to identify and classify network failures.

Keywords: Mobile networks; Neural networks; Clustering; Classification.

LISTA DE FIGURAS

Figura 1 – Subsistemas de uma rede móvel 2G.....	20
Figura 2 – Interfaces entre os elementos da rede.....	22
Figura 3 – Protocolos usados nas redes móveis	23
Figura 4 – Sistema de coleta de mensagens	24
Figura 5 – Mensagens de sinalização	25
Figura 6 – Comparação gráfica para Códigos de Liberação.	32
Figura 7 – Códigos de Liberação para diferentes MSC.....	33
Figura 8 – Topologia simplificada da rede em estudo.....	34
Figura 9 – Dados divididos em clusters.....	38
Figura 10 – Algoritmo K-means.....	40
Figura 11 – Exemplo de clusterização hierárquica aglomerativa.....	41
Figura 12 – Dendrograma.....	42
Figura 13 – Rede feedforward de camada simples.....	45
Figura 14 – Rede feedforward de múltiplas camadas.....	45
Figura 15 – Rede recorrente ou realimentada.....	46
Figura 16 – Rede com estrutura reticulada.....	46
Figura 17 – Arquitetura básica de uma Mapa auto-organizável.....	47
Figura 18 – Exemplos de topologias linear e em duas dimensões.	49
Figura 19 – Mapa topológico com 16 neurônios.....	49
Figura 20 – Algoritmo de treinamento de uma rede auto-organizável.....	50
Figura 21 – Classificação de um Mapa auto-organizável.....	51
Figura 22 – Soma de distâncias dos dados ao centro dos clusters.....	55
Figura 23 – K-means usando medidas de distância <i>Euclidiana</i> e <i>Cityblock</i>	58
Figura 24 – K-means usando medidas de distância <i>Cosine</i> e <i>Correlation</i>	58

Figura 25 – Funções para Clusterização Hierárquica no Matlab.....	61
Figura 26 – Algoritmos testados na Clusterização Hierárquica	63
Figura 27 – Gráfico das distâncias para o Algoritmo Ward	64
Figura 28 – Topologia gridtop.....	64
Figura 29 – Topologia hextop.	65
Figura 30 – Topologia randtop.	65
Figura 31 – Distância representada por código de cores.	67
Figura 32 – Distribuição gráfica usando o comando plotsomhits.	68
Figura 33 – Gráfico das distâncias para as topologias linear e hextop.....	69
Figura 34 – Gráfico de Silhouette para dados gerados aleatoriamente no Matlab.....	70
Figura 35 – Mapa das distâncias para 48 clusters obtidos por Kohonen.....	82
Figura 36 – Regiões de desempenho	83
Figura 37 – Modelo de classificação de falhas.....	87
Figura 38 – Desempenho médio por região do Mapa auto-organizável.....	90
Figura 39 – Classificação dos dados por falha em cada região.	91
Figura 40 – Distribuição das falhas por MSC.	91

LISTA DE TABELAS

Tabela 1 – Tabela de vizinhança com $R=1$	50
Tabela 2 – Medidas de distância da função K-means usada no Matlab	56
Tabela 3 – Critérios de desempenho aplicados aos testes iniciais.....	72
Tabela 4 – Clusterização feita por K-means.....	75
Tabela 5 – Clusterização feita pelo Mapa auto-organizável de Kohonen.	75
Tabela 6 – Clusterização feita pelo método Hierárquico.	76
Tabela 7 – Falhas identificadas por K-means.....	78
Tabela 8 – Falhas identificadas pelo Mapa auto-organizável de Kohonen	79
Tabela 9 – Falhas identificadas pela clusterização Hierarquica	80
Tabela 10 – Níveis de falha da rede de Acesso	84
Tabela 11 – Classificação da região 1	84
Tabela 12 – Classificação da região 2	85
Tabela 13 – Classificação da região 3	86
Tabela 14 – Classificação das regiões 4 e 5	87
Tabela 15 – Exemplo de mapeamento de clusters pelo Especialista.....	88
Tabela 16 – Classificação dos dados por Região	89
Tabela 17 – Falhas na saída do modelo.....	92
Tabela 18 – Distribuição dos dados nas regiões.....	93

LISTA DE ABREVIATURAS E SIGLAS

ANATEL	Agência Nacional de Telecomunicações
AUC	Authentication Centre
BSC	Base Station Controller
BSS	Base station subsystem
CDR	Charging Data Record
CN	Código Nacional de Numeração
DCN	Data Communications Network
GSM	Global System for Mobile Communications
HLR	Home Location Register
KPI	Key Performance Indicator
MOC	Mobile Originating Call
MSC	Mobile Switching Center
MTC	Mobile Terminating Call
NSS	Network switching subsystem
OMC	Operation Maintenance Center
OSS	Operation Support Subsystem
POC	PSTN Terminated Call
PTC	PSTN Originated Call
PSTN	Public switched telephone network
RNC	Radio Network Controller
SMP	Sistema Móvel Pessoal
SMS	Short Message Service
VLR	Visitor Location Register

SUMÁRIO

INTRODUÇÃO	12
1 CONCEITOS BÁSICOS.....	19
1.1 FUNDAMENTO DE REDES MÓVEIS 2G/3G	19
1.1.1 <i>Arquitetura do GSM</i>	19
1.2 INFORMAÇÕES GERADAS NA REDE	23
1.2.1 <i>Códigos de Liberação</i>	26
2 ESTUDO PRELIMINAR.....	29
2.1 ACOMPANHAMENTO DA ROTINA DE OPERACIONAL.....	29
2.2 COLETA E PREPARAÇÃO DOS DADOS	33
3 TÉCNICAS DE CLUSTERIZAÇÃO	37
3.1 K-MEANS	39
3.2 CLUSTERIZAÇÃO HIERÁRQUICA	41
3.3 REDES NEURONAIS ARTIFICIAIS	43
3.3.1 <i>O Neurônio Artificial</i>	43
3.3.2 <i>Arquitetura de redes neurais artificiais</i>	44
3.3.3 <i>Mapas auto-organizáveis de Kohonen</i>	47
4 TRABALHOS RELACIONADOS	52
5 TESTES INICIAIS	54
5.1 ROTINA DE TESTE	54
5.2 CLUSTERIZAÇÃO USANDO K-MEANS.....	55
5.2.1 <i>Testes usando a função K-means</i>	57
5.3 CLUSTERIZAÇÃO HIERÁRQUICA	59
5.3.1 <i>Testes usando Clusterização Hierárquica</i>	61
5.4 CLUSTERIZAÇÃO USANDO MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN.....	64
5.4.1 <i>Testes com Mapa auto-organizáveis de Kohonen</i>	68
5.5 MEDIDAS DE DESEMPENHO	69
5.5.1 <i>Critérios de desempenho</i>	70
5.5.2 <i>Resultados encontrados</i>	72
6 METODOLOGIA PROPOSTA	74
6.1 CLUSTERIZAÇÃO.....	74
6.2 IDENTIFICAÇÃO DAS FALHAS	76

6.3	REFINAMENTO MÉTODO KOHONEN.....	81
6.4	VISÃO GERAL DO MODELO.....	87
7	RESULTADOS	89
7.1	VALIDAÇÃO DO MODELO	92
7.2	COMENTÁRIO FINAL	93
8	CONCLUSÃO.....	94
	REFERÊNCIAS.....	97

INTRODUÇÃO

Um dos principais desafios das operadoras de telecomunicações é o de manter a qualidade do serviço prestado a seus assinantes. Para isso, foram criadas métricas, conhecidas como indicadores de qualidade, que permitem avaliar o comportamento geral da rede e servem como parâmetro para identificação de problemas, necessidade de ajustes e de expansão da capacidade dos equipamentos.

Os indicadores de qualidade são aferidos pelos administradores da rede através de dados coletados a partir dos serviços de monitoramento dos próprios elementos que compõem o sistema, ou ainda pela captura de mensagens usadas no controle das conexões estabelecidas entre os equipamentos, em pontos estratégicos da rede.

No caso particular do serviço móvel pessoal (SMP) alguns dos principais indicadores de qualidade de rede foram criados pela Agência Nacional de Telecomunicações (ANATEL), órgão regulador do sistema de telecomunicações no Brasil, através da Resolução nº. 575/2011 e são auditados por essa agência. Os indicadores permitem avaliar, sobretudo, os serviços de voz e mensagens, e são referenciados como: Taxa de Completamento de Chamadas para o Centro de Atendimento (SMP3), Taxa de Completamento (SMP4), Taxa de Alocação de Canal de Tráfego (SMP5), Taxa de Entrega de Mensagem de Texto (SMP6) e Taxa de Queda de Ligação (SMP7) (ANATEL, 2011). Cada um destes parâmetros é um indicador de desempenho de uma parte importante do sistema.

Os indicadores visam relacionar o número total de chamadas/mensagens originadas “completadas” com o total de “tentativas” de originar chamadas/mensagens em um período de tempo. A ANATEL estabelece critérios bem definidos para a coleta de informações, cálculo e consolidação das medidas para gerar os indicadores. Considerando o indicador SMP3, por exemplo, um dos critérios é o que estabelece que podem ser contabilizadas como chamadas completadas, somente aquelas originadas na rede da prestadora em que houve atendimento e comunicação. A resolução nº. 575/2011 (ANATEL, 2011), estabelece também as definições, métodos e frequência de coleta, consolidação e envio à ANATEL dos indicadores de qualidade com metas apresentadas no Regulamento de Gestão da Qualidade da Prestação do Serviço Móvel Pessoal (RGQ-SMP). Essa resolução permite à Anatel aferir a qualidade das chamadas de voz geradas dentro da rede do SMP (chamadas internas) bem como chamadas

originadas dentro da rede da operadora e terminada em outra rede, seja de outra operadora do SMP ou do STFC (Sistema de Telefonia Fixa Comutada). Os dados coletados para o cálculo e consolidação dos indicadores devem ser enviados mensalmente à ANATEL, até o dia 10 (dez) do mês subsequente ao da coleta. Os indicadores de qualidade da Anatel são numerados de SMP1 a SMP20, contudo somente os indicadores SMP3, SMP4, SMP5, SMP6 e SMP7 são considerados neste trabalho, pois tratam de eventos de rede envolvendo chamadas e mensagens. A Tabela 1 mostra os indicadores de qualidade ANATEL.

Tabela 1 – Indicadores de qualidades da Anatel para o SMP

Categorias	Indicadores de Qualidade	
Reação do Usuário	SMP1	Taxa de Reclamações
	SMP2	Taxa de Reclamações na Anatel
Rede	SMP3	Taxa de Completamento de Chamadas para o Centro de Atendimento
	SMP4	Taxa de Completamento
	SMP5	Taxa de Alocação de Canal de Tráfego
	SMP6	Taxa de Entrega de Mensagem de Texto
	SMP7	Taxa de Queda de Ligação
Conexão de Dados	SMP8	Taxa de Conexão de Dados
	SMP9	Taxa de Queda das Conexões de Dados
	SMP10	Garantia de Taxa de Transmissão Instantânea Contratada
Atendimento ao usuário	SMP11	Garantia de Taxa de Transmissão Média Contratada
	SMP12	Taxa de Atendimento pela Telefonista/Atendente em Sistemas de Auto-Atendimento
	SMP13	Taxa de Resposta ao Usuário
Pesquisa	SMP14	Taxa de Atendimento Pessoal ao Usuário
	SMP15	Capacidade de Resolução
	SMP16	Competência do Atendente
	SMP17	Competência e Organização da prestadora
	SMP18	Orientação ao Usuário
	SMP19	Conta e Cartão
	SMP20	Qualidade das Ligações

Fonte: ANATEL, 2011.

Os indicadores de qualidade podem ser obtidos a partir da leitura de contadores de eventos fornecidos pelos serviços de gerência dos equipamentos ou ainda a partir de CDRs (*Charging Data Record*) que são registros gerados para cada conexão (3GPP, 2010).

Os contadores de eventos usados para obter os indicadores de uma rede móvel são gerados pelas centrais de comutação (MSC – *Mobile Switching Center*) e equipamentos de acesso (BSCs – *Base Station Controller* e RNCs – *Radio Network Controller*) em intervalos regulares, geralmente a cada 15 ou 30 minutos, e exportados para um banco de dados. Os contadores relacionados a eventos de rede merecem atenção especial, pois através deles é

possível identificar as causas de liberação de uma chamada ou o motivo pelo qual uma conexão não é completada.

A cada conexão são gerados nas MSS registros detalhados de chamada (CDR) que contém informações da conexão. Um CDR contém campos que identificam e descrevem cada chamada. Os campos mais comumente utilizados são:

- Número do assinante chamador;
- Número do assinante chamado;
- Data e hora do início da chamada;
- Duração da chamada;
- Categoria do assinante chamador (telefone público, teste, comum);
- Número sequencial identificando o registro;
- Tipo do CDR (Ex. Chamada de entrada, saída, trânsito, etc.);
- Número da rota e circuito utilizados na conexão;
- Falhas de conexão encontradas.

Os campos encontrados nos CDRs e sua formatação variam de fornecedor para fornecedor e muitos deles são customizados de acordo com a necessidade das operadoras. As centrais de comutação consolidam os CDRs em arquivos de tamanho fixo, em torno de alguns Megabytes, que são transferidos utilizando algum tipo de protocolo (Ex. FTP, SFTP, etc.) automaticamente para servidores de bancos de dados localizados nos centros de mediação, onde é realizada uma triagem baseada no conteúdo dos CDRs. Ao final da triagem, os CDRs são processados nos centros de bilhetagem e faturamento para a emissão da conta dos usuários.

Em geral, as operadoras habilitam nos equipamentos apenas a geração de CDRs de chamadas completadas para que os centros de processamento (mediação) não fiquem saturados de informações que não serão utilizadas para a realização do faturamento. Somente durante o período de coleta de informações para a ANATEL, são gerados todos os tipos de CDRs inclusive aqueles para chamadas não completadas.

Outra fonte de informação sobre o desempenho da rede, além dos contadores e CDRs, é baseada no monitoramento das mensagens de sinalização usadas para controle das conexões nas interfaces entre os equipamentos. Esse tipo de monitoramento permite obter informações de todas as fases do estabelecimento de uma conexão, incluindo o momento da desconexão. Essas informações são obtidas através de *probes* conectadas aos *links* de sinalização

existentes entre os equipamentos. As *probes* “copiam” as mensagens e as enviam para um servidor que as armazena de acordo com a data e hora em que elas foram geradas. Os equipamentos para esse tipo de coleta não permitem coletar as informações das conexões geradas dentro de um mesmo elemento de rede (Ex. chamadas intracentral). Além disso, este tipo de coleta de informação tem um custo elevado para as operadoras e não permitem monitorar 100% dos eventos gerados na rede, pois todas as interfaces precisariam ser monitoradas.

Tão importante quanto aferir os indicadores de qualidade da ANATEL é reagir, no menor tempo possível, à perda de qualidade e desempenho da rede. Nesse sentido, as informações usadas para gerar relatórios de rede e indicadores também são utilizadas para dar subsídios às equipes de operação e manutenção para atuar de forma preventiva. Com essa finalidade, as operadoras possuem equipes de operação e manutenção que estão organizadas em diferentes níveis de atuação.

As equipes de operação e manutenção, na maioria das operadoras, estão agrupadas em três níveis distintos. O primeiro nível é responsável pelo “gerenciamento de falhas” cuja principal função é monitorar os painéis de alarmes, onde são consolidados todos os alarmes dos equipamentos da rede, e solucionar os problemas mais simples (Ex. desbloqueio de circuitos, isolamento manual de processadores em falha, etc.). O primeiro nível também é responsável por acionar o segundo nível de manutenção na ocorrência de falhas mais complexas do sistema. O segundo nível é responsável por avaliar o estado geral dos equipamentos realizando diagnósticos e testes, solicitar a substituição de partes defeituosas dos equipamentos e configurar os elementos da rede. O terceiro nível, composto pelos profissionais mais experientes, atua quando o segundo nível não consegue solucionar o problema. Esse nível é capacitado para atuar em falhas de hardware e software dos equipamentos bem como identificar comportamentos inadequados que estejam afetando a rede de forma sistêmica (OLIVEIRA, 2008).

Ocorre que, as ferramentas em uso para a análise de falhas pelos níveis de manutenção citados estão fragmentadas em diferentes aplicações, dificultando a análise global do comportamento da rede, tornando o tempo de identificação e recuperação de uma falha uma função da experiência e capacidade individual dos profissionais envolvidos. Esse fato pode resultar em perdas financeiras para uma operadora na medida em que, dependendo da falha, os serviços prestados aos assinantes podem ficar indisponíveis ou ainda não ser possível registrar os eventos de conexões necessários para que a operadora possa tarifar os assinantes.

No cotidiano de operação, quando ocorre uma falha na rede, um rompimento de fibra, por exemplo, o Centro de Gerência de Rede (CGR) recebe alarmes autônomos gerados pelos diferentes equipamentos afetados da rede móvel (MSC, HLR, BSC, etc.) no mesmo intervalo de tempo. Esses alarmes indicam, sob o ponto de vista de cada elemento, o efeito causado pela falha. Contudo, diferentes falhas de rede podem fazer com que alarmes do mesmo tipo sejam gerados (Ex. perdas de conexão, perda de mensagens, sobrecarga nos processadores dos equipamentos, etc.). Nesse momento, o operador de primeiro nível, a partir das informações de alarmes, geralmente não consegue identificar a origem da falha. O problema então é escalado para os níveis superiores de manutenção que, a partir da avalanche de alarmes recebida, começa a filtrar os mais relevantes. Além dos alarmes gerados pelos equipamentos, é preciso também analisar as mensagens trocadas entre os mesmos. Essa informação é obtida a partir dos sistemas de monitoramento de sinalização que coletam as mensagens e as enviam para armazenamento em servidores que podem ser acessados pelas equipes de manutenção. Outra fonte de consulta para diagnóstico do problema são os Códigos de Liberação gerados pelas MSCs. Eles permitem saber as razões pelas quais uma chamada não tenha se completado. É importante comentar que as informações dos CDRs, mensagens e Códigos de Liberação são acessadas usando aplicações diferentes, apresentam-se em diferentes formatos e estão armazenadas em banco de dados. Por isso, mesmo os profissionais mais experientes levam muito tempo para correlacionar todas essas informações para obter uma visão global do problema e assim realizar o diagnóstico correto da falha.

Muitos modelos para detecção, classificação e localização de falhas têm sido propostos tomando como base técnicas de inteligência computacional e técnicas baseadas em modelos estatísticos. É possível através do uso de Redes Neurais Artificiais (RNA), por exemplo, detectar falhas em linhas de transmissão (ROCHA, 2005) por meio da análise de tensões e correntes trifásicas. O principal atrativo para o uso de RNA na solução de problemas de classificação é a sua capacidade de aprender através de exemplos e de generalizar o conhecimento aprendido. Outra técnica utilizada é a Lógica Fuzzy, que também permite a classificação de defeitos, facilitando o diagnóstico. No campo da Engenharia Mecânica, por exemplo, essa técnica pode ser usada para o diagnóstico automático de defeitos em enrolamentos e mancais (FUJIMOTO, 2005). Técnicas de detecção de anomalias são usadas na análise de dados na forma de séries temporais para procurar por padrões inesperados. Em Bottari (2014) essa técnica é utilizada para realizar o monitoramento térmico em centros de processamento de dados.

Nesse trabalho é apresentada a proposta de um modelo para detecção e classificação de falhas em redes do Sistema Móvel Pessoal (SMP), a partir das mesmas informações utilizadas no dia-a-dia das equipes que operam o sistema, e uma forma de correlacionar a falha detectada com a causa provável para a ocorrência da mesma.

Como primeiro passo, um estudo preliminar, feito com base nas informações (mensagens, CDRs e Códigos de Liberação) produzidas pelos equipamentos de rede (MSC) mostrou que os Códigos de Liberação reúnem os atributos necessários para serem usados no desenvolvimento do modelo. A análise de amostras dos Códigos de Liberação, retiradas de uma massa de dados produzida por 6 MSCs de diferentes regiões durante dois meses consecutivos, permite identificar padrões capazes de refletir o comportamento da rede em diferentes situações (falha, desempenho normal, perda de desempenho, etc.).

Para que os padrões existentes na massa de dados fossem identificados, seria necessário agrupar os dados em classes (clusters) de acordo com a similaridade existente entre elas. Diferentes métodos de clusterização foram testados e avaliados de acordo com a sua eficácia ou acurácia para realizar o agrupamento dos dados. A partir da avaliação de desempenho dos métodos de clusterização, foi feita a proposta de um modelo utilizando o método com melhor resultado nos testes: Mapas auto-organizáveis de Kohonen. Novos testes foram realizados e a validação final do modelo foi feita pelo especialista do sistema, cuja função é relacionar cada classe formada pelo modelo proposto com o respectivo comportamento da rede.

Os resultados obtidos foram positivos, permitindo ao especialista identificar a ocorrência de falhas bem como constatar a existência de outros padrões de comportamento representados por diferentes agrupamentos.

Assim, através dos resultados apresentados nos testes com o modelo proposto nesse trabalho, deve ser possível acelerar o processo de diagnóstico de falhas, pelas equipes de operação, reduzindo o tempo de identificação da falha. Em consequência disso, será possível melhorar o desempenho das equipes de operação no exercício de suas funções.

O presente trabalho está estruturado da seguinte forma: No Capítulo 1 são apresentados os conceitos básicos relacionados a redes móveis 2G/3G. No Capítulo 2 é mostrado o estudo preliminar feito com base nas atividades das equipes de operações responsáveis pela manutenção de uma rede móvel. O Capítulo 3 apresenta as técnicas de clusterização estudados nesse trabalho. No Capítulo 4 são apresentados os trabalhos

relacionados ao uso das técnicas estudadas para identificação de falhas. No Capítulo 5 são mostrados os testes de clusterização realizados com a massa de dados (Códigos de Liberação) geradas por 6 MSCs. Ao final desse Capítulo, os métodos são avaliados individualmente e comparados de acordo com os critérios de avaliação adotados. No Capítulo 6, é feita a escolha dos métodos de clusterização a serem usados no modelo e são realizados os testes de classificação. No Capítulo 7, são apresentados os resultados finais onde todos os dados são analisados e classificados pelo especialista de rede. Finalmente, no Capítulo 0 são apresentadas as conclusões deste trabalho, além dos apontamentos para trabalhos futuros.

1. CONCEITOS BÁSICOS

Neste capítulo são apresentados os conceitos básicos e as tecnologias envolvidas em uma rede móvel, para ampliar o entendimento sobre o problema e a identificação da oportunidade de se propor uma solução para facilitar a identificação de falhas em redes do SMP. São descritos os principais elementos que compõe as redes 2G e 3G, as interfaces existentes entre os equipamentos e os protocolos de comunicação envolvidos.

1.1 Fundamento de redes móveis 2G/3G

As redes móveis de segunda (2G) e terceira (3G) geração do sistema GSM (*Global System for Mobile Communication*) representam um padrão global de evolução tecnológica alcançado pelas redes de telecomunicações, que permitem acesso a serviços de voz, dados e mensagens, providos diretamente ao usuário final. Estes serviços geralmente são oferecidos dentro de uma mesma rede, ou ainda através do encaminhamento para outras redes, na maioria das vezes operadas por entidades diferentes, que utilizam a mesma tecnologia. O mecanismo de encaminhamento, chamando de *Roaming* Automático, é padronizado, permitindo a interoperação entre redes diferentes, inclusive a parte de bilhetagem, Entre as principais características do sistema GSM destacam-se (ALENCAR, 2004):

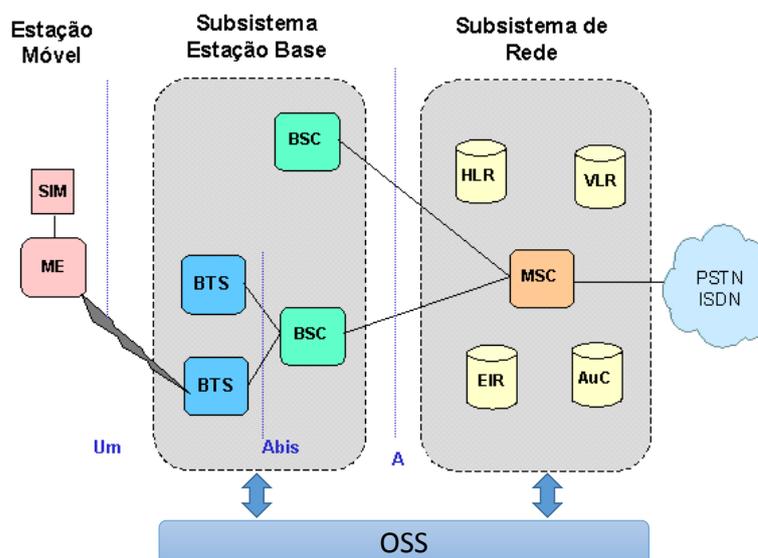
- *Roaming* Automático de voz e dados;
- Infraestrutura e terminais de baixo custo;
- Sinais de alta qualidade e segurança na comunicação.

1.1.1 Arquitetura do GSM

A arquitetura básica de uma rede móvel GSM, conforme mostrado na Figura 1, é constituída por três subsistemas: Subsistema de Estações Rádio Base (*Base Station*

Subsystem, BSS); Subsistema de Comutação (*Network Switching Subsystem*, NSS) e Subsistema de Suporte a Operação (*Operation Support Subsystem*, OSS).

Figura 1 – Subsistemas de uma rede móvel 2G.



Fonte: 3GPP TS 23.002, 2015

O Subsistema de estações rádio base (BSS) é responsável pelas funções relacionadas à rede de acesso. Ele consiste de controladores de estações rádio base (BSC), no caso das redes 2G e RNC para as redes 3G, e das antenas transmissoras/receptoras (BTS).

A BTS tem a função de fornecer uma conexão de rádio para a estação móvel (ME).

A BSC é responsável por controlar a BTS, alocar um canal de comunicação para estabelecimento das conexões e monitorar a potência dos sinais de transmissão entre a BTS e a estação móvel com o objetivo de manter a qualidade do serviço.

O Subsistema de comutação (NSS) atua no processamento de chamadas e funções relacionadas aos usuários. As principais entidades funcionais (ETSI GSM 03.02, 1996) que compõem esse subsistema são:

- *Home Location Register (HLR)* – O HLR é a base de dados usada para armazenar e gerenciar os dados dos assinantes. Esse elemento é considerado como a base de dados mais importante do sistema, pois é nele que estão gravados os dados permanentes sobre todos os usuários.

- *Mobile Services Switching Center (MSC)* – A MSC realiza funções de comutação do sistema telefônico. Esse elemento é responsável pelo controle das chamadas e interconexões com outras MSC da mesma rede ou de redes vizinhas realizando a função de gateway.
- *Visitor Location Register (VLR)* – O VLR é a base de dados que contém informação temporária dos assinantes. Essas informações são usadas pela MSC no tratamento das chamadas. A maior parte dos fornecedores de equipamentos fabricam a MSC e VLR conjugados no mesmo hardware devido a quantidade de mensagens trocadas entre esses dois elementos.
- *Authentication Center (AUC)* – O AUC é a unidade responsável por prover serviços de autenticação e vetores de criptografia que serão usados para garantir a confidencialidade das informações trocadas pelos usuários durante o estabelecimento de uma chamada.
- *Equipment Identity Register (EiR)* – O EiR é a base de dados que contém informações sobre a identidade dos equipamentos móveis (ME). Sua função é prevenir a utilização de equipamentos roubados ou cuja utilização não seja autorizada pelo sistema (ex. equipamentos não homologados).

O Subsistema de Suporte Operacional (*Operation Support Subsystem, OSS*) é a entidade funcional através da qual é possível monitorar e controlar todo o sistema. É composto pelo Centro de Operação e Manutenção (*Operation Maintenance Centre, OMC*) que está conectado a todos os equipamentos do NSS e BSS. O propósito do OSS é fornecer as operadoras de telecomunicações um ambiente centralizado para operação centralizada, remota ou local da rede para as diferentes equipes responsáveis pela manutenção do sistema.

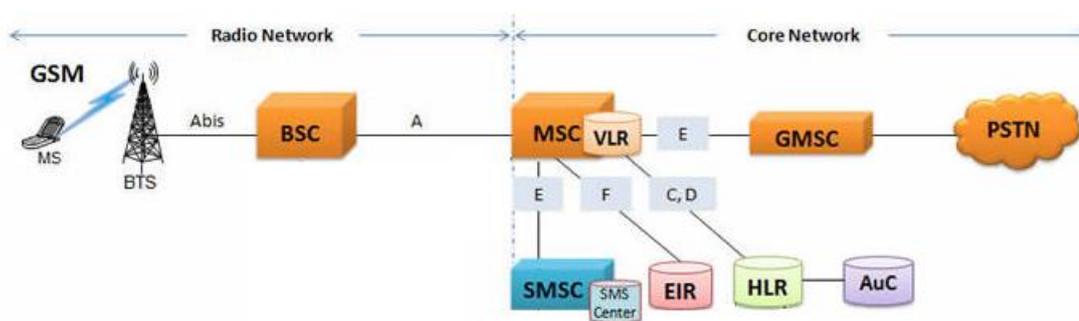
Dentre os subsistemas descritos anteriormente o OSS merece especial atenção, pois é ele que receberá em primeira mão toda a informação dos contadores e alarmes gerados nos equipamentos da rede que estão sobre sua supervisão. É partir desses subsistemas que são extraídas as informações para a geração de relatórios, KPI (*key performance indicator*) da rede e também os indicadores de qualidade que serão enviados ao órgão regulador do sistema de telecomunicações (ANATEL).

Do ponto de vista de interface e comunicação, a rede GSM está dividida em Plano de Controle e Plano de usuários:

- Plano de Controle – Esse plano é responsável por estabelecer a comunicação entre os diversos elementos da rede através de protocolos e interfaces padronizadas.
- Plano de Usuário - O plano de Usuário é responsável por dar suporte a comunicação fim a fim entre os usuários permitindo a troca de informações entre eles.

O diagrama da Figura 2 apresenta as principais interfaces entre os elementos de uma rede GSM.

Figura 2 – Interfaces entre os elementos da rede.



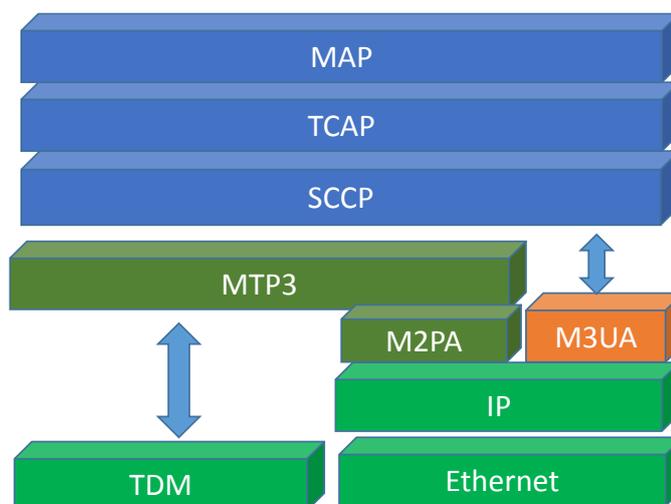
Fonte: SIEGMUND, 1995.

Os protocolos de rede apresentados na Figura 3 são responsáveis por estabelecer a comunicação confiável entre os elementos e garantir a estabilidade das conexões do ponto de vista de controle. Não é objeto do presente estudo explicar em detalhe o funcionamento desses protocolos, mas uma breve descrição sobre cada um deles é apresentada a seguir.

- MAP (*Mobile Application Part*) - Permite a transferência de mensagens entre as entidades funcionais de uma rede móvel para gerência de mobilidade (3GPP TS 29.002, 2004);
- TCAP (*Transaction Capabilities Application Part*) - Permite o "diálogo" entre as diferentes entidades funcionais da rede de sinalização número 7 (ITU-T Q.773, 1997);
- SCCP (*Signaling Connection Control Part*) - Fornece métodos de roteamento adicionais aos protocolos inferiores (ITU-T Q.713, 1996);
- MTP3 (*Message Transfer Part 3*) - Fornecer roteamento de mensagens entre os elementos de uma rede de sinalização número 7 (SS7) (ITU-T Q.701, 1993);
- M2PA (*User Peer-to-Peer Adaptation Layer*) - Serve como camada de adaptação para usar o protocolo MTP3 em redes IP (IETF RFC 4165, 2005);

- M3UA (*MTP3 User Adaptation Layer*) - Permite a sinalização de usuários MTP3 sobre a rede IP (ex. SCCP) através da camada SCTP (IETF RFC 3332, 2002);
SCTP (*Stream Control Transmission Protocol*) – Realiza o transporte confiável de mensagens sobre um serviço de pacotes não confiável e sem conexão (IETF RFC 4960, 2007).

Figura 3 – Protocolos usados nas redes móveis



Fonte: RUSSELL , 1994.

1.2 Informações geradas na rede

Os vários elementos de uma rede móvel geram três tipos de informação que contém os atributos necessários para qualificar cada conexão estabelecida entre eles:

- CDRs;
- Mensagens de Sinalização;
- Códigos de Liberação.

Os CDRs, conforme dito anteriormente, são gerados nas MSC e sua principal função é registrar as informações necessárias para realizar a cobrança aos usuários pela utilização dos serviços oferecidos pela operadora.

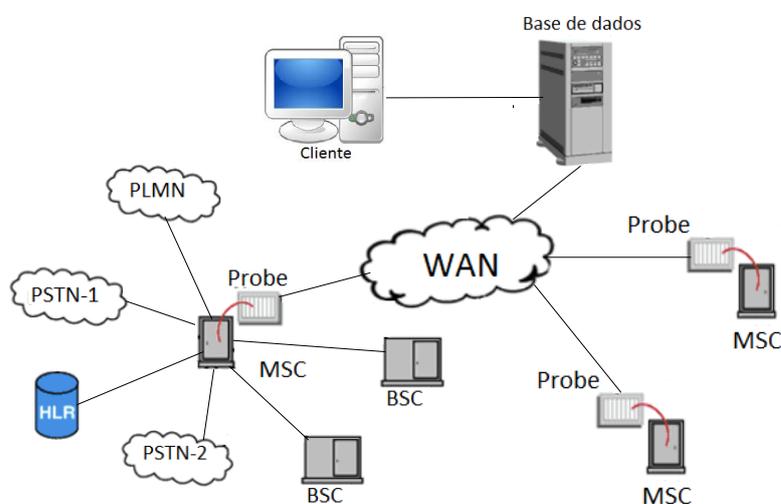
Existem diversos tipos de CDR (3GPP TS 32.298, 2015), dentre os quais merecem destaque os CDRs MOC (*Mobile Originated Call*), MTC (*Mobile Terminated Call*), POC

(PSTN *Originated Call*) e PTC (PSTN *Terminated Call*). Os CDRs MOC e MTC são gerados quando um assinante pertencente à rede móvel origina ou recebe uma chamada, respectivamente. Os CDRs POC e PTC são gerados quando uma chamada chega ou sai da MSC por um circuito em direção a uma outra MSC da mesma rede ou de outra rede.

As Mensagens de Sinalização são trocadas entre os elementos para o estabelecimento, monitoração e desconexão das chamadas telefônicas e serviços de mensagens curtas (SMS). Cópias das mensagens de sinalização podem ser obtidas por *probes* conectadas às interfaces entre os elementos da rede. A conexão entre os elementos segue o modelo OSI de comunicação. Na sua camada de rede dois tipos de interface podem ser encontrados: Interfaces padrão E1 (PCM-30) (ITU-T, 2001) e Interfaces *Ethernet*. Essas interfaces são fisicamente conectadas ao *Backbone* de uma rede de transporte composta por equipamentos de transmissão e roteadores IP. As interfaces E1 são ligadas diretamente aos equipamentos de transmissão de dados, já as interfaces Ethernet são conectadas a DCN (*Data Communication Network*), uma rede IP dedicada para os equipamentos da rede móvel.

A coleta de mensagens ocorre diretamente sobre as interfaces E1 por meio de derivações físicas nos cabos de transmissão, ou através de espelhamento das portas dos roteadores onde estão conectadas as interfaces *Ethernet*. Os elementos concentradores chamados *probes* concentram as informações e as enviam para um banco de dados central onde são armazenadas e ficam disponíveis para serem consultadas. A Figura 4 mostra um sistema de coleta de mensagens de sinalização conforme descrito anteriormente.

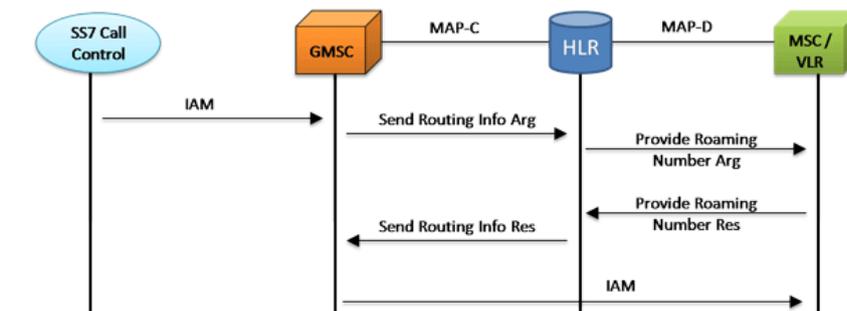
Figura 4 – Sistema de coleta de mensagens



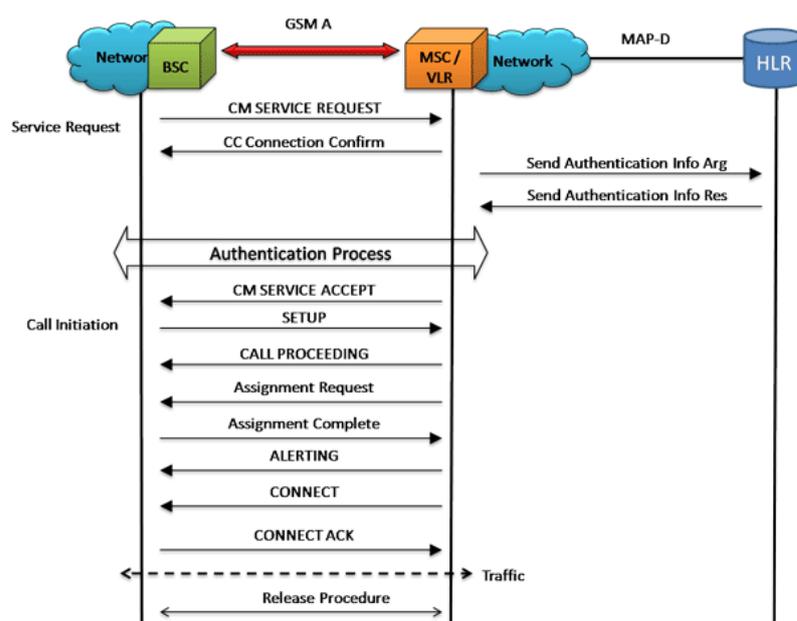
Fonte: O autor, 2016.

As Figura 5 (a) e (b) exemplificam as mensagens trocadas entre os elementos de rede durante os processos de chamada terminada e autenticação de um usuário.

Figura 5 – Mensagens de sinalização



(a)



(b)

Fonte: 3GPP TS 23.221, 2004.

Os Códigos de Liberação representam eventos gerados e contabilizados nas MSC através de contadores que são enviados para os equipamentos de gerência de rede e armazenados para a geração de KPIs e/ou indicadores de qualidade.

Nesse trabalho optou-se por usar os Códigos de Liberação como fonte de dados por conterem informações que são registrados nos CDRs e também por contabilizarem eventos

que envolvem a troca de mensagens de sinalização entre os elementos. Dessa forma é possível obter uma visão global do comportamento da rede.

1.2.1 Códigos de Liberação

A liberação de uma chamada consiste de uma sequência de ações realizadas na rede para finalizar uma conexão ou uma tentativa de conexão. Durante esse processo, circuitos e outros recursos físicos que foram reservados para a chamada são liberados. Quando todos os recursos forem liberados, a chamada estará completamente liberada.

Existem diversas razões possíveis para que uma chamada seja liberada. A maioria das chamadas realizadas com sucesso entre dois assinantes, chamador e chamado, são liberadas “normalmente”, ou seja, quando ambos os assinantes desligam seu telefone com o propósito de encerrarem a chamada.

As chamadas também podem ser liberadas prematuramente. Uma possibilidade é a chamada por algum motivo falhar, quando então ela é liberada antes da conexão ser completamente estabelecida. O mesmo acontece quando um problema técnico na rede ocorre durante a chamada causando a necessidade de sua liberação.

Algumas razões comuns para liberação de uma chamada são:

- O assinante chamador digita uma sequência incorreta de dígitos e por isso o assinante, destino da chamada, não pode ser encontrado;
- O assinante chamado está ocupado;
- O assinante chamado não responde;
- Não existem circuitos livres para cursar a chamada até o destino desejado naquele momento;
- Os meios de transmissão ou sinalização requeridos para a realização da chamada não estão disponíveis na rede.

Códigos de Liberação são registros contendo índices gerados pelas MSCs para identificar o motivo da liberação de uma chamada ou indicar a ocorrência de um evento de rede. Ele pode servir também para controle de informações em um elemento de rede fazendo

com que o elemento execute certas funções ou procedimentos baseado em um Código de Liberação específico.

Os Códigos de Liberação podem ser entendidos também como fontes de informações internas do sistema. Em conexões entre elementos de rede, onde exista a necessidade de informar ao elemento remoto o motivo de liberação de uma chamada, Códigos de Liberação são mapeados em códigos de causa e enviados ao elemento remoto para que esse execute seus procedimentos internos.

Os Códigos de Liberação são utilizados principalmente para:

- Reportar informações sobre o funcionamento da rede;
- Servir como informações de controle no tratamento de chamadas.

Códigos de Liberação, em um elemento de rede, podem ser gerados de duas formas:

- Um elemento recebe uma mensagem de liberação de uma chamada de outro elemento remoto. A mensagem identifica a “causa da liberação” (código de causa). Quando essa informação chega ao elemento, o código que identifica a causa da liberação é convertido no seu código de liberação correspondente;
- O elemento identifica um motivo para liberar uma chamada. Em seguida gera o código de liberação correspondente ao motivo que causou a liberação.

Quando um elemento identifica o motivo da liberação de uma chamada na qual ele participa, esse elemento é o responsável por enviar a informação de liberação ao outro nó da rede envolvido na mesma conexão. Isso é feito após a conversão do “Código de Liberação” em “Código de Causa” que é enviado a seguir através de uma mensagem de sinalização.

Nesse caso, a causa de liberação dá ao outro elemento a informação necessária para que os recursos alocados para a chamada sejam liberados de acordo com a causa da liberação recebida.

A análise dos Códigos de Liberação permite avaliar a rede sobre o ponto de vista de desempenho e condições de tráfego de um elemento de rede e, conseqüentemente, uma visão do comportamento de toda a rede uma vez que permite avaliar o estado geral da rede em relação a cada elemento. As informações relacionadas aos Códigos de Liberação são geradas nas MSCs e registradas em relatórios e CDRs quando necessário.

Os responsáveis pela operação de rede podem criar uma visão do funcionamento e desempenho da rede bem como de cada elemento a partir de relatórios gerados com base nos diferentes Códigos de Liberação. Falhas na rede e nos elementos podem ser percebidas e localizadas através da análise das informações expressas pelos Códigos de Liberação.

Para utilizar essas informações, o operador pode realizar a leitura dos Códigos de Liberação diretamente dos equipamentos de rede (MSC) ou transferir as informações relacionadas aos Códigos de Liberação para um banco de dados, periodicamente, de acordo com a sua necessidade (ex. a cada 5 minutos). Dessa forma as informações estarão disponíveis para consultas posteriores.

As redes GSM foram criadas para ser um padrão internacional de comunicação. Sua grande vantagem é possibilitar o roaming automático entre as operadoras que adotam esse padrão. A arquitetura básica do padrão GSM compreende três subsistemas: BSS, OSS e NSS. Cada subsistema é responsável por um seguimento específico da rede. O BSS é composto por todos os equipamentos e funções relacionadas as estações de rádio. O NSS é responsável comutação de chamadas de voz e comunicação com outras redes. Cabe ao OSS, monitorar e controlar todo o sistema através da comunicação com cada um dos equipamentos de rede.

O conhecimento adquirido no estudo da arquitetura GSM é fundamental para acompanhar a rotina dos especialistas do sistema e compreender as atividades operacionais. O próximo capítulo é baseado nas informações obtidas durante o acompanhamento das atividades operacionais realizadas por profissionais de uma rede GSM Brasileira. Essa atividade permitiu reunir informações que contribuíram para a elaboração do modelo proposto nesse trabalho.

2. ESTUDO PRELIMINAR

Neste capítulo é apresentado o estudo preliminar das informações disponíveis nos equipamentos e como eles são usados pelos especialistas das equipes de operação. Também foram avaliadas as características destas informações para embasar a seleção de um subconjunto destas informações que pudessem ser utilizadas em métodos de classificação.

2.1 Acompanhamento da rotina de operacional

Durante a fase de estudo preliminar, foi realizado um acompanhamento da rotina dos especialistas das equipes de operação e manutenção de uma operadora de telecomunicações do sistema GSM para obter um mapeamento dos procedimentos operacionais usados na identificação e correção de falhas, buscando selecionar as informações mais relevantes usadas nesse processo.

O ambiente de trabalho das equipes que operam o sistema consiste em grandes centros de gerência de rede que tem visibilidade do comportamento de todos os elementos sob o ponto de vista de alarmes autônomos gerados pelos próprios equipamentos, enviados aos servidores de gerência através do uso de protocolos de rede, para serem armazenados por servidores e processados em seguida. Além dos alarmes autônomos, os diversos equipamentos também produzem CDRs, mensagens de sinalização e contadores conforme dito anteriormente.

Os alarmes autônomos gerados pelos equipamentos são enviados para uma aplicação cuja função é abrir automaticamente uma ficha de identificação da falha (*trouble ticket*) que é um registro contendo data, hora identificação textual do alarme e a identificação do equipamento que o produziu.

Os *trouble tickets* gerados seguem uma cadeia de acionamento que incluem os três níveis de manutenção já informados. Ao final do tratamento de cada alarme, o *trouble ticket* é fechado com outras informações preenchidas pelo operador que solucionou o problema, é informada a causa da falha e a solução dada para solucionar o problema. Em princípio, foi cogitada a utilização dessa informação para servir como referência para o modelo proposto,

pois as informações tanto da falha, como de causa e solução já estariam presentes em cada registro. Contudo, verificou-se que muitos dos registros não continham informações exatas sobre o tratamento da falha, e sua causa raiz, dificultando sua classificação. Por isso, o uso dessas informações como fonte de consulta na elaboração do trabalho foi descartado.

Os processos operacionais mostram que em cem por cento dos casos, é constatado que os elementos de rede conhecidos como MSC estão envolvidos nas falhas. Esse fato é explicado por serem esses equipamentos os primeiros elementos do núcleo da rede móvel a prover serviço ao assinante, conforme mostrado na Figura 1. Com isso, a pesquisa foi orientada para o seguimento de informações produzidos por esses elementos, buscando-se obter um conjunto de dados através das quais fosse possível obter uma visão global do comportamento da rede e, a partir disso determinar o comportamento sistêmico de uma determinada falha. Com essa visão, as informações identificadas como relevantes para o trabalho foram os CDRs, Mensagens de Sinalização e Códigos de Causa explicados nos capítulos anteriores.

Os CDRs contêm todas as informações relacionadas a uma conexão desde a seu estabelecimento até sua desconexão, contendo inclusive a causa da desconexão (i.e., desligamento do assinante *A*, desligamento do assinante *B*, etc.). Contudo, essa informação não foi escolhida para o desenvolvimento, pois necessitaria de uma etapa de pré-processamento dos tipos de CDRs (MOC, MTC, POC, PTC) para análise dos campos de interesse de cada CDR, correlação dos campos selecionados, e contagem dos eventos contidos nos diversos CDRs, de acordo os cenários de conexão e desconexão. Uma outra questão importante é que apenas os CDRs de chamadas completadas, isto é, chamadas que chegam a sua fase estável (conversa), são gerados em condições normais de funcionamento dos equipamentos. As operadoras optam por essa configuração nos equipamentos porque a principal função dos CDRs é prover informações para a tarifação dos assinantes e, de acordo com a regulamentação vigente, apenas chamadas completadas com tempo de duração acima de 3s podem ser tarifadas (ANATEL, 2005). Então, gerar CDRs que não sejam de chamadas completadas implicaria em custo adicional para armazenamento e processamento dessas informações nos equipamentos de rede e também espaço de armazenamento nos servidores de mediação que são responsáveis pelo tratamento dos CDRs.

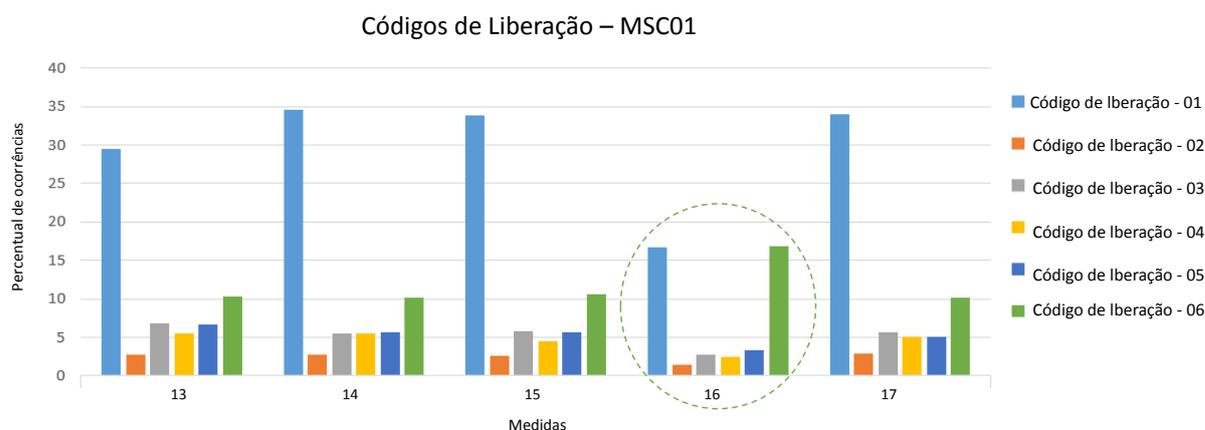
No caso das mensagens de sinalização, para que as informações nelas contidas sejam usadas, é preciso inicialmente recuperá-las previamente em um servidor onde as mesmas encontram-se armazenadas, conforme mostrado na Fonte: O autor, 2016.. Um inconveniente dessa abordagem é que as principais mensagens relacionadas às conexões entre dois

assinantes, registradas em a uma mesma MSC, não são monitoradas pelos dispositivos de captura (*probes* de sinalização). Isso acontece porque estas mensagens não são enviadas para a rede, elas ficam confinadas dentro do próprio equipamento (primitivas). Além disso, o pré-processamento das mensagens demanda tanto esforço computacional e custo de armazenamento quanto o tratamento dos CDRs. No processo de tratamento de falhas, assim como na geração de relatórios emitidos pelos operadores para acompanhamento do desempenho do sistema, verificou-se o uso frequente dos Códigos de Liberação, explicados anteriormente. Esses são gerados para cada MSC da rede de forma estratificada e com referência de tempo (data e hora) estabelecida. Estes contadores representam o número de ocorrências de um determinado evento por unidade de tempo. Diversos eventos são contabilizados e seus valores registrados por esses contadores. Quando a produção desses contadores em todas as MSCs é sincronizada, é possível obter uma visão global do comportamento da rede pela análise conjunta dessas informações. Embora a quantidade de eventos registrada pelos contadores, ao longo do tempo (minutos, horas, etc.) apresentem diferenças em relação aos seus valores absolutos, observou-se que seus valores relativos (percentuais) estão concentrados nas mesmas regiões de valores. Esse fato torna-se mais evidente ao se observarem os contadores dos Códigos de Liberação de elementos instalados em uma mesma seção de rede, pois esses elementos adquirem características semelhantes, tais como: tratarem os assinantes pertencentes a uma mesma região (Ex, CNs 21,22,24) ou ainda estarem conectados à mesma seção da rede de transmissão. Foi percebido que o conjunto dos Códigos de Liberação aferidos durante uma mesma falha para duas MSCs conectadas a um mesmo segmento de rede produz “assinaturas percentuais” semelhantes (Figura 8), isso é, os valores variam de forma proporcional.

Ainda, por esses dados estarem disponíveis em um formato que pudesse ser usado diretamente no estudo, reduzindo o esforço na etapa de pré-processamento, e também por refletirem o comportamento dos elementos da rede, optou-se por realizar observações comparativas com essa informação e assim verificar se, de fato, o resultado obtido nessa primeira etapa confirmaria o uso dos Códigos de Liberação como fonte de dados para o projeto.

A Figura 6 mostra uma comparação entre valores (assinaturas) de Códigos de Liberação para uma MSC, em diferentes dias e horários. O círculo pontilhado indica uma mudança de padrão percebida. Essa diferença indica uma mudança de comportamento da rede.

Figura 6 – Comparação gráfica para Códigos de Liberação.



Fonte: O autor, 2016.

Uma segunda comparação foi realizada com os mesmos Códigos de Liberação, usados no gráfico da Fonte: O autor, 2016.

, desta vez considerando diferentes MSC. Os contadores usados na comparação foram coletados em dias e horários diferentes. O resultado desta comparação, representada no gráfico da Fonte: O autor, 2016.

, mostra que os contadores apresentam comportamento com perfil semelhante, os valores assumidos por cada um dos Códigos de Liberação variam da mesma forma para equipamentos distintos, indicando um padrão de funcionamento entre esses elementos.

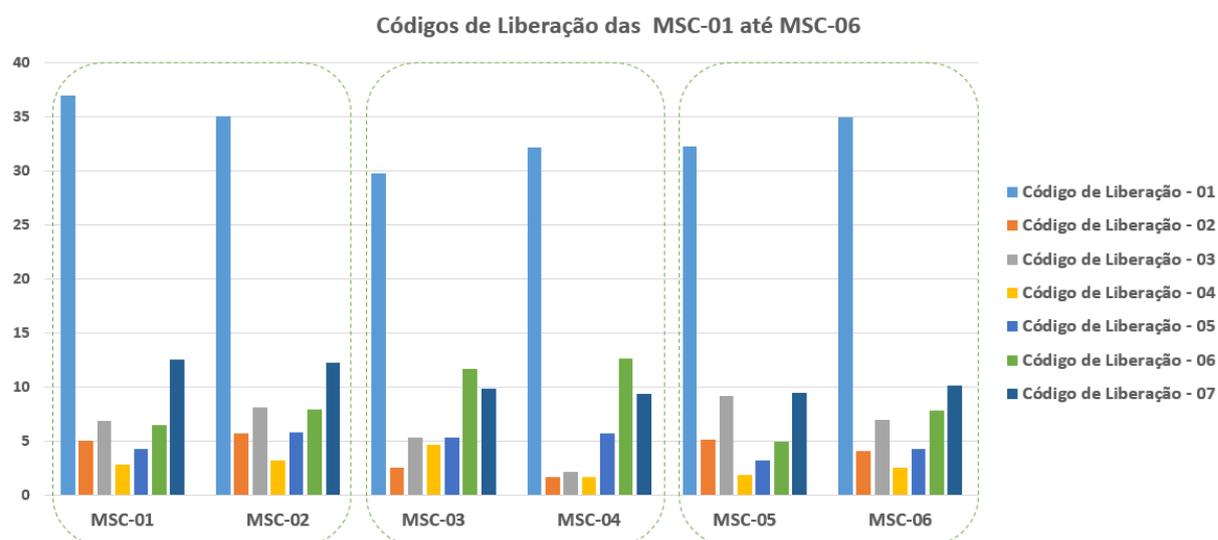
É possível observar também que a semelhança entre as assinaturas dos Códigos de Liberação é maior para equipamentos que atendem a mesma região geográfica (círculos pontilhados). O comportamento dos contadores mostrado nos gráficos da Fonte: O autor, 2016.

e Fonte: O autor, 2016.

, indica ser possível mapear eventos de mesma natureza, seja falha ou um comportamento de tráfego normal da rede, através dos Códigos de Liberação.

A partir destas constatações, entendemos que deveria ser possível usar uma mesma heurística para classificar os Códigos de Liberação, independentemente da quantidade de recursos e região de implantação do equipamento. Dessa forma, a próxima questão a ser respondida é como associar a estes eventos uma “assinatura” ou “padrão de identificação”.

Figura 7 – Códigos de Liberação para diferentes MSC



Fonte: O autor, 2016.

Motivado por essas observações foram dados os próximos passos visando analisar os contadores de forma mais científica estendendo o campo de observação para verificar o comportamento dos dados em situações mais específicas da rede: normalidade e falhas. Com essas informações, uma seleção inicial dos dados foi usada para encontrar as ferramentas mais adequadas que poderiam fazer a classificação desses dados.

2.2 Coleta e preparação dos dados

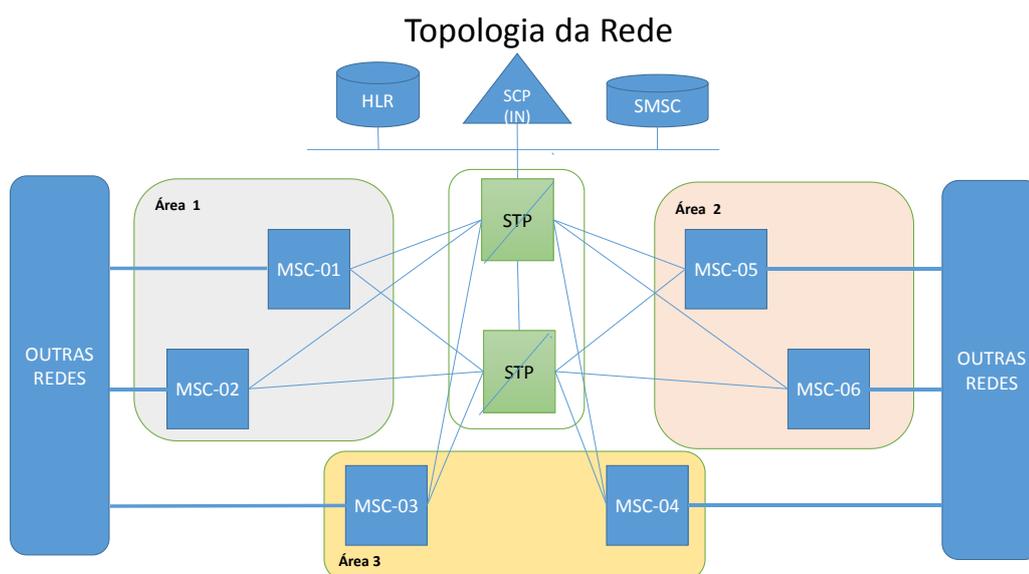
Antes de avaliar os potenciais mecanismos de detecção, aprofundamos os estudos sobre os dados de entrada, ou seja, os contadores dos Códigos de Liberação. Inicialmente, uma massa grande de dados contendo Códigos de Liberação foi coletada a partir do servidor de uma operadora de telecomunicações, cuja identidade não será revelada nesse trabalho por questões de confidencialidade dos dados fornecidos.

Nesta massa de dados, os Códigos de Liberação são gerados por todas as MSC da rede em intervalos regulares que podem variar de 5 a 30 minutos, de acordo com a resposta desejada do sistema. Esses contadores são enviados por meio de um protocolo para serem armazenados em servidores de banco de dados que concentram as informações de todos os

equipamentos, formando assim uma base de dados única para consulta. Na rede onde foram realizados os testes, a geração e envio dos contadores ocorre a cada 15 minutos e o armazenamento é feito em um banco de dados (SQL) de forma incremental por um período de uma hora, isso é feito para otimizar o espaço ocupado no banco de dados.

As informações referentes a dois meses consecutivos foram exportadas do banco de dados da operadora para arquivos em formato texto (CSV, *coma separated values*) para serem usados nesse projeto. Desse modo, para cada Código de Liberação foram contabilizadas, em média, 1.328 amostras de cada contador, para cada um dos equipamentos em teste. Foram coletadas amostras de 6 MSCs, representando 3 áreas geograficamente separadas, atendidas pela operadora. A topologia simplificada da rede pode ser verificada na Figura 8.

Figura 8 – Topologia simplificada da rede em estudo



Fonte: O autor, 2016.

Os dados exportados foram carregados em um banco de dados auxiliar (MySQL) para pré-processamento “manual” na busca por erros de formação ou falta de medidas que pudessem distorcer a informação presente nos arquivos exportados.

Após a primeira depuração do arquivo, foram identificados os Códigos de Liberação representativos do ponto de vista numérico e grau de importância em relação aos eventos de rede correlacionados.

Os Códigos de Liberação foram analisados de acordo com seu valor médio percentual, pois o tráfego telefônico está distribuído ao longo do dia. Desse modo, a cada hora são

geradas quantidades diferentes de Códigos de Liberação. No período da manhã, por exemplo, existem menos eventos (conexões) que no período da tarde. Observou-se que, independente da hora, o comportamento “percentual” dos Códigos de Liberação não sofre variações expressivas ao longo do dia, a não ser por ocorrência de eventos inesperados (i.e., falhas). Essa verificação foi feita através da apuração dos valores de médias e desvio padrão calculados para o conjunto de dados amostrado. Com o auxílio do especialista do sistema, essa análise foi realizada para cada um dos Códigos de Liberação contidos no arquivo usado para o teste.

Para guiar a análise e aferir mudanças de comportamento da rede foi escolhido como referência o “Código de Liberação Normal” cujo valor percentual verificado no conjunto de dados, na maioria dos casos, varia entre 30% a 50%. A informação obtida através desse contador foi utilizada como indicador de desempenho dos elementos analisados. Esse Código de Liberação indica que todos os eventos relacionados ao estabelecimento de uma chamada foram realizados com sucesso. Essa abordagem faz sentido, uma vez que o objetivo das operadoras é sempre maximizar o número de chamadas completadas o que, em termos de Códigos de Liberação, pode ser representado pelo “Código de liberação Normal”. Portanto, variações “abruptas” do contador que representa esse código podem indicar a ocorrência de problemas na rede.

A partir das análises realizadas, os Códigos de Liberação foram classificados manualmente e agrupados de acordo com a sua frequência e relevância com base nas informações obtidas junto aos especialistas que operam o sistema.

Foi selecionado para análises um conjunto inicial de 162 Códigos de Liberação, de acordo com os seguintes critérios:

- Máximo valor Percentual de ocorrência aferido - Foram descartados os Códigos de Liberação com percentual de ocorrência Máximo inferior a 0,09% aferido em todas as amostras. Na prática isso significa admitir os eventos indicados por esses Códigos de Liberação raramente ocorrem e por isso não tem influência significativa no funcionamento da rede.
- Evento representativo - Além dos Códigos de Liberação com maior número de ocorrências (maiores valores máximos percentuais), foram selecionados pelos especialistas aqueles que representam eventos relevantes de rede e que podem indicar uma falha aumentando seu valor abruptamente.

Após a aplicação dos critérios mencionados anteriormente, dos 162 do conjunto inicial considerado, foram selecionados 99 Códigos de Liberação para serem usados como atributos de entrada pelo modelo proposto no presente trabalho. Cada conjunto define um vetor de atributos que será tratado pelo modelo como um ponto no espaço multidimensional das possibilidades dos valores que podem ser assumidos pelos Códigos de Liberação.

O conjunto de dados consiste de uma matriz de 8640 linhas \times 99 colunas. As linhas da matriz representam os registros relacionados às MSCs, data, e hora em que os contadores foram amostrados. Cada coluna representa um atributo em valores percentuais para cada Código de Liberação. É importante ressaltar que para cada MSC existem 1440 linhas da matriz.

Com os dados organizados, a próxima etapa consistiu da escolha da técnica que irá agrupar os dados para que seja possível identificar, a partir deste agrupamento, os padrões que representem falhas. Devido à natureza das informações, as técnicas estudadas deveriam ser capazes de separar os dados em grupos, levando em consideração as características em comum entre eles, uma vez que não há, *a priori*, a indicação de que o registro (conjunto de Códigos de Liberação) representa um comportamento normal do sistema ou uma falha. Com os grupos formados, e com o auxílio dos especialistas do sistema, as características devem ser avaliadas e associadas a um comportamento de rede.

Foram estudadas diversas técnicas que permitem realizar a separação dos dados em grupos (*clusterização*) de acordo com as suas características, dentre elas as que apresentaram melhor resultado do ponto de vista da proposta desse trabalho são:

- K-means
- Mapas auto-organizáveis de Kohonen

O próximo capítulo apresenta as técnicas de clusterização estudadas nesse trabalho. São mostrados os fundamentos de cada uma delas e suas principais características.

3. TÉCNICAS DE CLUSTERIZAÇÃO

A Clusterização é um método utilizado para aprender sobre o comportamento de um conjunto de dados. Através desse método, um conjunto de dados pode ser organizado em diferentes grupos de acordo com medidas de similaridade entre esses dados formando diferentes grupos de indivíduos ou uma hierarquia de grupos (BERKHIN, 2002).

Essa técnica é utilizada em diversas áreas do conhecimento tais como as engenharias, biologia, psicologia, medicina, ciência da computação, entre outras.

Na análise de dados quando não se tem, a priori, nenhum conhecimento sobre como esses dados se relacionam, a Clusterização é empregada para identificar os relacionamentos existentes e formar grupos a partir dessa identificação. Nas áreas de Inteligência Computacional e de Reconhecimento de Padrões essa técnica é conhecida como “aprendizado não supervisionado”. Esse termo significa que, antes de serem submetidos ao processo de clusterização, os dados não foram associados a nenhuma classe. Por isso, não há como validar ou supervisionar através de uma lógica comparativa se os dados contidos em um agrupamento, resultante do processo clusterização, deveriam estar realmente associados a uma classe específica.

Um cluster pode ser definido como uma agregação de pontos (vetores) no espaço multidimensional (EVERITT, 1974), isto é, dado um conjunto $X = \{X_1, X_2, \dots, X_N\}$, com cada objeto x_i possuindo p atributos (dimensões ou características), representados por $x_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$. Deve-se construir k grupos $C_j = \{1, \dots, k\}$ a partir de X , onde os objetos de cada grupo sejam homogêneos segundo alguma medida de similaridade. A similaridade é comumente definida em termos de o quanto "perto" os dados estão no espaço.

Devem ser observados os seguintes critérios no processo de clusterização: A união de todos os grupos formados (C_i) é igual a X (Equação 1). Não deve haver interseção entre os grupos formados (Equação 2). Cada grupo deve possuir ao menos um objeto (Equação 3)

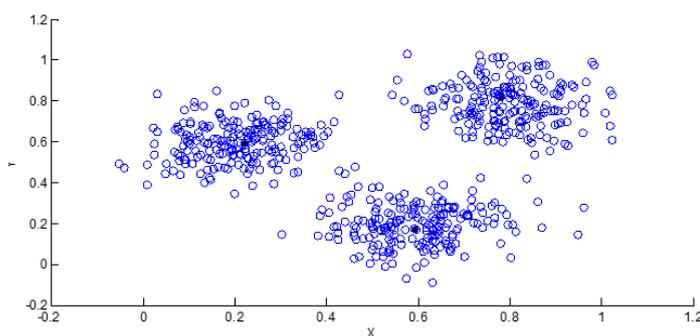
$$\bigcup_{i=1}^k C_i = X \quad (1)$$

$$C_i \cap C_j = \emptyset \quad i, j = 1, \dots, k \text{ e } i \neq j \quad (2)$$

$$C_i \neq \emptyset \quad i = 1, \dots, k \quad (3)$$

Um simples exemplo de clusterização é mostrado na Figura 9. O processo utilizado para determinar se um ponto pertence ou não a um cluster está relacionado à medida de similaridade entre os pontos do conjunto. Avaliando a figura apresentada, é possível identificar a existência de três clusters a partir da distância estimada entre eles. A medida de similaridade utilizada é fundamental para determinar a formação de um (novo) cluster. A medida mais comumente utilizada na formação de clusters é a Distância Euclidiana. Contudo, outras medidas podem ser utilizadas (i.e., Distância Euclidiana quadrática, Distância Manhattan, Distância Chebychev, etc.) diferentes tipos de medidas podem gerar clusters diferentes para um mesmo conjunto de dados.

Figura 9 – Dados divididos em clusters.



Fonte: O autor, 2016.

Além das medidas utilizadas para determinar a formação de um cluster, uma variedade de algoritmos tem surgido para executar o processo de clusterização, e têm sido aplicados com sucesso em problemas de mineração de dados, que também dependem de agrupamento de dados, em casos concretos.

Uma dificuldade no processo de clusterização é determinar a quantidade de clusters a serem formados. Geralmente a quantidade de clusters depende da experiência daqueles que utilizam essa técnica e do conhecimento acerca dos dados a serem clusterizados.

Neste trabalho são avaliados três métodos de clusterização. K-means, Clusterização Hierárquica, e Redes Neurais (Mapas auto-organizáveis de Kohonen). Cada um deles é explicado com mais detalhes nos próximos itens.

3.1 K-means

K-means é um importante um método de clusterização que objetiva particionar um conjunto de n observações em número k de clusters, onde cada observação pertence ao cluster mais próximo da média. A função do algoritmo empregado por K-means é minimizar a média dos quadrados das Distâncias Euclidianas dos pontos aos centros do cluster. O centro do cluster é definido como a média, ou centroide, do conjunto de pontos que formam um cluster. Algumas medidas de distância que podem ser usadas no algoritmo são listadas a seguir (MANNING, 2009):

- Distância Euclidiana;
- Distância Euclidiana Quadrática;
- Distância Manhattan;
- Distância Chebychev.

A Equação 4 mostra o cálculo de um centroide, apresentado por $\mu(\omega)$ para o conjunto de pontos ω :

$\mu(\omega)$ = centroide

ω = conjunto de pontos

\mathcal{X} = soma dos quadrados das distancias

$$\vec{\mu}(\omega) = \frac{1}{|\omega|} \sum_{\vec{x} \in \omega} \vec{x} \quad (4)$$

A medida de quão bem o centroide representa os elementos de um cluster é dada pela soma residual dos quadrados (RSS - *residual sum of square*), ou seja, é realizada a soma de dos quadrados da distância de cada vetor ao centro do cluster, conforme apresentado na Equação 5.

$$\text{RSS}_k = \sum_{\vec{x} \in \omega_k} |\vec{x} - \vec{\mu}(\omega_k)|^2$$

$$\text{RSS} = \sum_{k=1}^K \text{RSS}_k \quad (5)$$

Obter RSS é o objetivo da função K-means e minimizar esse valor é o desejado na formação da clusterização usando esse método. Uma vez que a quantidade de elementos do conjunto de dados é fixa, minimizar RSS significa minimizar a média do quadrado das distâncias dos pontos aos seus respectivos centros.

A Figura 10 representa a implementação básica do algoritmo K-means utilizado na maioria das aplicações de clusterização:

Figura 10 – Algoritmo K-means.

```

K-MEANS( $\{\vec{x}_1, \dots, \vec{x}_N\}, K$ )
1   $(\vec{s}_1, \vec{s}_2, \dots, \vec{s}_K) \leftarrow \text{SELECTRANDOMSEEDS}(\{\vec{x}_1, \dots, \vec{x}_N\}, K)$ 
2  for  $k \leftarrow 1$  to  $K$ 
3  do  $\vec{\mu}_k \leftarrow \vec{s}_k$ 
4  while stopping criterion has not been met
5  do for  $k \leftarrow 1$  to  $K$ 
6    do  $\omega_k \leftarrow \{\}$ 
7    for  $n \leftarrow 1$  to  $N$ 
8    do  $j \leftarrow \arg \min_j |\vec{\mu}_j - \vec{x}_n|$ 
9        $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$  (reassignment of vectors)
10   for  $k \leftarrow 1$  to  $K$ 
11   do  $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$  (recomputation of centroids)
12  return  $\{\vec{\mu}_1, \dots, \vec{\mu}_K\}$ 

```

Fonte: MANNING, 2009.

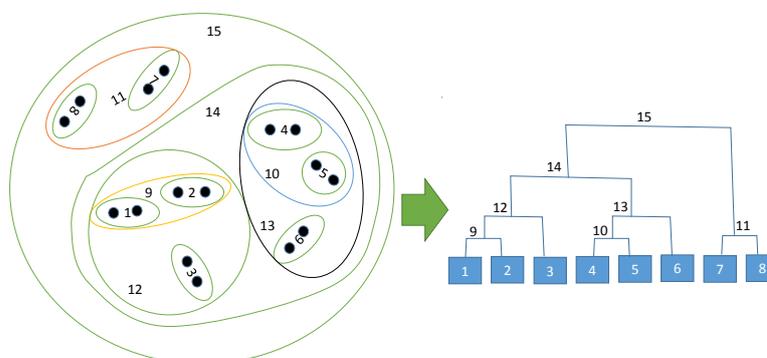
O primeiro passo do algoritmo é escolher os centros dos k clusters aleatoriamente. A função do algoritmo é mover os centros dos clusters de forma a minimizar a soma das distâncias ao quadrado dos pontos aos centros de cada cluster (minimizar RSS). Conforme mostrado na Equação 5, isto é feito iterativamente até que um critério de parada seja alcançado (passo 4). São realizadas novas associações dos pontos aos clusters (passo 9) considerando a menor distância entre cada ponto e o centro do seu respectivo cluster (passo 8). A cada nova formação dos clusters, os centros são recalculados e o processo é repetido até que o critério de parada seja alcançado (passo 4).

O método de clusterização K-means é eficiente e conceitualmente simples, contudo existem “inconvenientes” na sua utilização. Como pode ser verificado na análise do seu algoritmo, é necessário que seja definida a quantidade de clusters e o seu resultado pode depender da escolha aleatória dos centros iniciais.

3.2 Clusterização Hierárquica

A clusterização hierárquica é um método utilizado para agrupar dados de forma hierárquica (MANNING, 2009). A formação inicial dos clusters começa pela “união” dos elementos que estiverem mais próximos de acordo com um critério de similaridade estabelecido previamente. Esse processo ocorre, recursivamente, unindo os clusters mais próximos “dois a dois”, dando origem a novos clusters. Essa operação ocorre até que todos os clusters sejam agrupados em um único cluster contendo todo o conjunto de dados. Algoritmos de clusterização podem realizar essa aglomeração usando diferentes métodos: *bottom-up* ou *top-down*. A clusterização hierárquica que usa o método *bottom-up* é chamada clusterização hierárquica aglomerativa. A clusterização *top-down* necessita que os clusters sejam subdivididos continuamente até que clusters de um só elemento sejam alcançados. Na prática, a clusterização aglomerativa é a mais comumente utilizada. A Figura 11 mostra um exemplo de clusterização hierárquica aglomerativa que utiliza como critério de similaridade a distância entre os pontos. Primeiro é realizada a formação de pares (números de 1 à 8), depois são formados os grupos (numerados de 9 à 15).

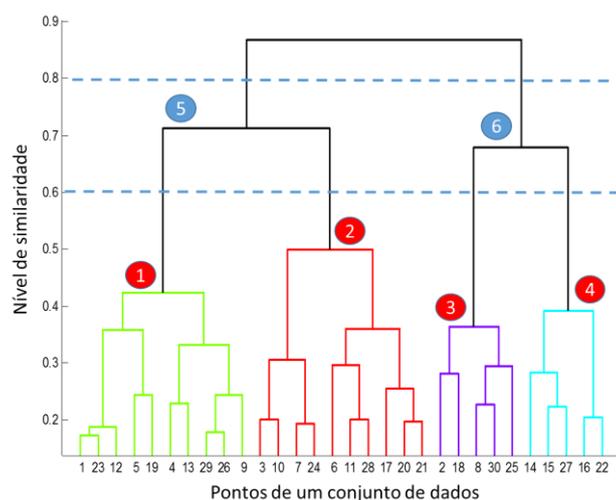
Figura 11 – Exemplo de clusterização hierárquica aglomerativa.



Fonte: O Autor, 2016.

O resultado da clusterização hierárquica pode ser graficamente apresentado na forma de um dendrograma como pode ser visto nas Figura 11 e Figura 12. Um dendrograma permite representar a relação de similaridade entre os elementos de um conjunto de dados. Ele é obtido pela união dos pontos formando a letra “u” invertida, onde as linhas horizontais representam a medida de similaridade (distância) entre dois pontos sendo conectados. A clusterização hierárquica não exige um número pré-especificado de clusters conforme utilizado por K-means. Contudo, em algumas aplicações, é necessário obter como resultado diferentes grupamentos. Nesses casos, a hierarquia tem de ser cortada em algum ponto. A escolha do ponto de corte permite agrupar os dados da forma mais adequada para seu uso ou aplicação. Para isso, alguns critérios (i.e., nível de similaridade entre clusters, número de clusters, etc.) podem ser usados e assim determinar o ponto de corte. Como exemplo, no dendrograma da Figura 12 são mostrados dois pontos de corte representados por linhas tracejadas. O primeiro corte é feito no nível de similaridade 0,8, e determina a formação 2 grupos (5 e 6). O segundo corte é feito no nível 0,6. Abaixo desse nível são formados 4 grupos (1,2,3 e 4).

Figura 12 – Dendrograma.



Fonte: O autor, 2016.

3.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no sistema nervoso dos seres vivos. Sua estrutura é composta por neurônios artificiais (unidades de processamento) ligados entre si por interconexões denominadas sinapses artificiais e representadas por vetores ou matrizes de pesos sinápticos.

O conjunto de unidades de processamento que forma as RNAs possui a capacidade de aquisição e manutenção do conhecimento baseado em informações extraídas de um conjunto de dados que representem algum fenômeno, sistema ou processo. Suas principais características são: a adaptação por experiência, a capacidade de aprendizado, a habilidade de generalização, a organização de dados, tolerância a falhas, armazenamento distribuído e a facilidade de prototipagem (SILVA, 2010). Assim, as RNAs podem ser utilizadas em aplicações que envolvam, por exemplo, a avaliação de imagens capturadas por satélite, a classificação de padrões (i.e., padrão de escrita, padrão de fala, etc.), a identificação de falhas em máquinas elétricas, etc. Estas características das RNAs, tornam possível o desenvolvimento de modelos capazes de realizar previsões sobre o comportamento de um determinado processo, e também executar ações que permitam o controle adequado do mesmo.

3.3.1 O Neurônio Artificial

O modelo simplificado do neurônio biológico é a unidade básica de processamento de uma rede neural artificial. Este neurônio, associado computacionalmente a uma estrutura matricial, possui uma ou mais entradas (X_1, \dots, X_n) e uma saída Y . A cada entrada fica associado um peso (W_1, \dots, W_n) que pondera cada entrada em relação à saída. Assim, pode-se resumir o funcionamento do neurônio artificial por meio da Equação 6, onde a variável X_i representa o conjunto de valores de entrada, multiplicados pelos seus respectivos pesos W_i somados a um limiar de ativação b . O somatório total é denominado potencial de ativação, que é submetido a uma função g (Função de ativação) que irá limitar a saída Y do neurônio dentro de um intervalo de valores estabelecidos pela imagem funcional.

$$Y = g \left(\sum_{i=1}^n W_i \cdot X_i + b \right) \quad (6)$$

3.3.2 Arquitetura de redes neurais artificiais

A arquitetura de uma rede neural estabelece como os neurônios estarão posicionados uns em relação aos outros e estruturados através do direcionamento de suas conexões sinápticas.

A topologia das Redes Neurais Artificiais depende da forma como os neurônios se conectam para formar uma "Rede" de neurônios. Para uma mesma arquitetura, por exemplo, parte dos neurônios pode assumir uma função de ativação (i.e. logística) enquanto os demais neurônios são ativados por outra função (i.e. tangente hiperbólica) (SILVA, 2010).

Uma vez definidas a arquitetura da rede neural e sua topologia é necessário ajustar seus pesos e limiares para que a resposta em sua saída se aproxime dos resultados desejados. Esse processo é denominado treinamento da rede neural.

Basicamente uma rede neural artificial pode ser dividida em três partes, denominadas de camadas:

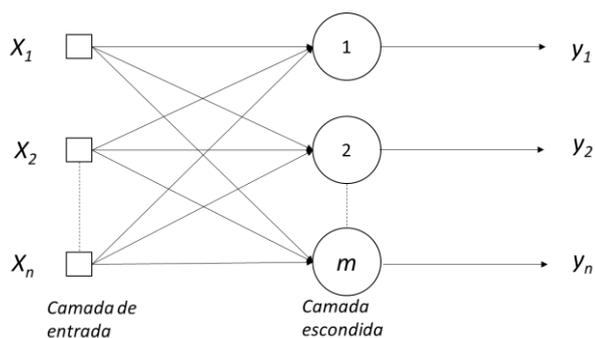
- Camada de entrada: Camada responsável por receber as informações de entrada (dados);
- Camadas escondidas ou intermediárias: Camadas formadas por neurônios, cuja responsabilidade é extrair as características associadas ao processo ou sistema. A maior parte do processamento da rede é realizado nessa camada;
- Camada de saída: Camada que recebe os dados processados vindos da camada escondida, também constituída por neurônios artificiais. Sua responsabilidade é produzir e apresentar os resultados finais do processamento.

As principais arquiteturas de redes neurais considerando os princípios de arquitetura e topologia definidos anteriormente são: redes *feedforward*, redes recorrentes e redes reticuladas.

Na arquitetura *feedforward* o fluxo de informações é sempre unidirecional, saindo da camada de entrada em direção a camada de saída. A arquitetura *feedforward* esta subdividida em dois tipos: redes de camada simples ou de múltiplas camadas. A arquitetura em camada

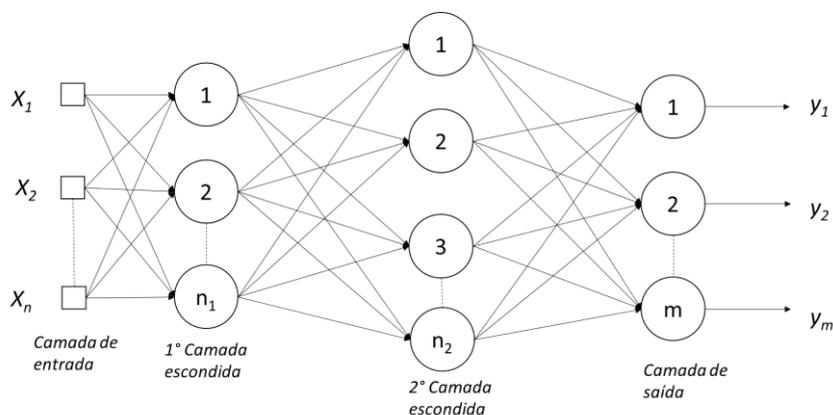
simples é mostrada na Figura 13, observa-se que o número de saídas (y_n) será sempre igual ao número de entradas (X_n). As arquiteturas em camadas múltiplas são constituídas de uma ou mais camadas. A Figura 14 mostra uma arquitetura com 4 camadas, uma cama de entrada (X_n) duas camadas escondidas (n_1 e n_2) e uma camada de saída (m).

Figura 13 – Rede feedforward de camada simples.



Fonte: SILVA, 2010.

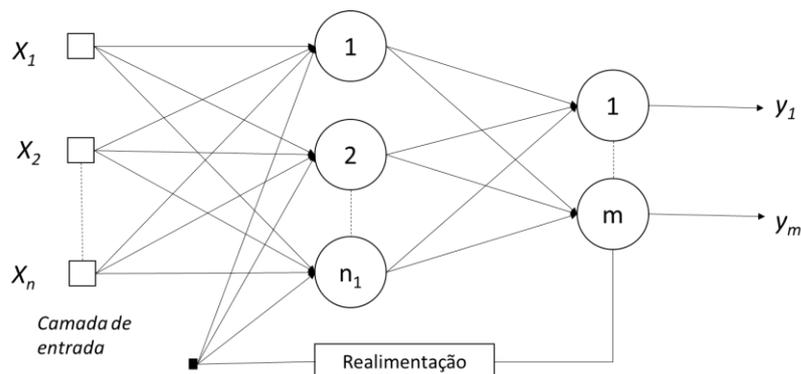
Figura 14 – Rede feedforward de múltiplas camadas.



Fonte: SILVA, 2010.

Nas arquiteturas recorrentes ou realimentadas, as saídas são realimentadas como entradas de outros neurônios. Essa característica permite seu uso na previsão de séries temporais, otimização, identificação de sistemas, etc. A Figura 15 exemplifica uma arquitetura recorrente, onde pode ser verificado que as entradas dos neurônios (X_n) da camada n são realimentadas pelos neurônios da camada m .

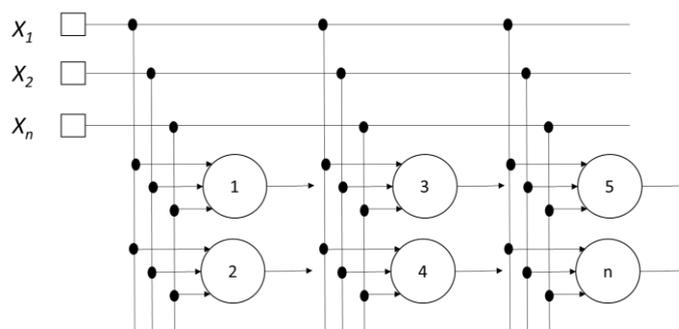
Figura 15 – Rede recorrente ou realimentada.



Fonte: SILVA, 2010.

As redes neurais reticuladas consideram a organização espacial dos neurônios com o propósito de extração de características. A localização espacial de cada neurônio que compõe a arquitetura está relacionada com o processo de ajuste dos seus respectivos pesos e limiares de ativação. Essas redes podem ser utilizadas em soluções que envolvam agrupamento, detecção de padrões, otimização de sistemas, etc. Como exemplo desse tipo de arquitetura podem ser citados os Mapas auto-organizáveis de Kohonen, cujo exemplo é mostrado na Figura 16, onde os neurônios (numerados de 1 à n) estão organizados em um espaço bidimensional. É importante perceber que as entradas (X_n) são comuns a todos os neurônios da arquitetura.

Figura 16 – Rede com estrutura reticulada



Fonte: SILVA, 2010.

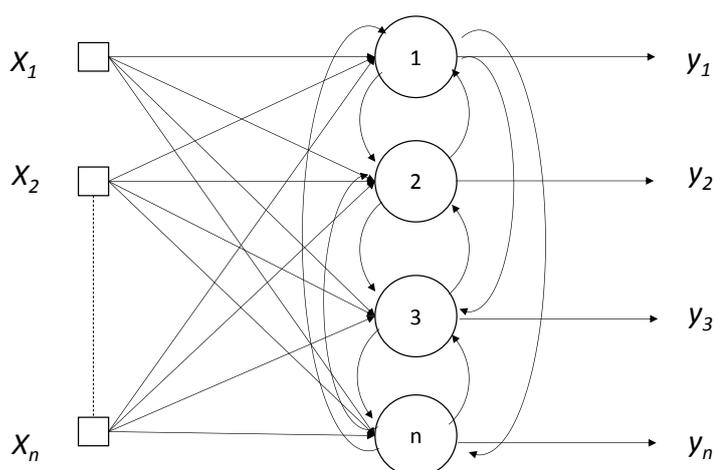
3.3.3 Mapas auto-organizáveis de Kohonen

Os Mapas auto-organizáveis de Kohonen (*Self organizing maps*) ou simplesmente SOM, são constituídos por uma estrutura de rede neural artificial, geralmente organizada em duas camadas, que recebe padrões de entrada com N dimensões e os mapeia para um conjunto de neurônios de saída.

Enquanto o treinamento da maior parte das redes neurais necessita que padrões de entrada e padrões de saída sejam conhecidos (treinamento supervisionado), os Mapas auto-organizáveis de Kohonen utilizam apenas padrões de entrada para realizar seu treinamento (treinamento não supervisionado). Esse tipo de rede é útil em aplicações onde somente os padrões de entrada sejam conhecidos, não existindo padrões de saída para serem relacionadas a entrada (HONKELA, 1997).

A Figura 17 mostra a Arquitetura básica de uma rede auto-organizável com apenas uma camada. As entradas são representadas por X_1 até X_n , os neurônios são numerados de 1 à n . E as saídas representadas de Y_1 a Y_n . As comunicações entre os neurônios são denominadas conexões laterais e permitem a cada neurônio influenciar na saída do outro.

Figura 17 – Arquitetura básica de uma Mapa auto-organizável.



Fonte: SILVA, 2010.

Os vetores de pesos para a rede neural da Figura 17 são discriminados a seguir. Onde $W_{(1)}$, $W_{(2)}$, $W_{(3)}$ e $W_{(n)}$ representam o vetor dos pesos dos neurônios de 1 à n . A notação $W_{1,n}$ indica o peso de *índice* n do vetor 1 e assim sucessivamente.

$$W_{(1)}=[W_{1,1} \ W_{1,2} \ \dots \ W_{1,n}] \quad \text{Vetor de pesos do primeiro neurônio.}$$

$$W_{(2)}=[W_{2,1} \ W_{2,2} \ \dots \ W_{2,n}] \quad \text{Vetor de pesos do segundo neurônio.}$$

$$W_{(3)}=[W_{3,1} \ W_{3,2} \ \dots \ W_{3,n}] \quad \text{Vetor de pesos do terceiro neurônio}$$

$$\dots$$

$$W_{(n)}=[W_{n,1} \ W_{n,2} \ \dots \ W_{n,n}] \quad \text{Vetor de pesos do } n\text{-ésimo neurônio}$$

As redes SOM utilizam métodos de treinamento competitivo para detectar similaridade e correlacionar os padrões do conjunto de dados de entrada, agrupando esses dados em grupos (clusters). O processo de treinamento competitivo consiste em “premiar” o neurônio cujo vetor de pesos estiver mais próximo do vetor de entrada aplicado a rede. O prêmio do vencedor é o ajuste de seus pesos, fazendo com que o vetor de pesos fique mais próximo do vetor de entrada, de modo que quando uma entrada semelhante for apresentada, este neurônio tem mais chances de ser o vencedor. Uma das métricas de similaridade normalmente utilizada é a distância entre esses dois vetores. A Equação 7 representa o cálculo da distância (norma euclidiana) entre o vetor de entrada X , o k representa a k -ésima amostra $\{X^{(k)}\}$, em relação ao vetor de pesos do j -ésimo neurônio $\{W^{(j)}\}$.

$$dist_j^{(k)} = \sqrt{\sum_{i=1}^n (x_i^{(k)} - w_i^{(j)})^2}, \quad \text{onde } j = 1, \dots, n_1 \quad (7)$$

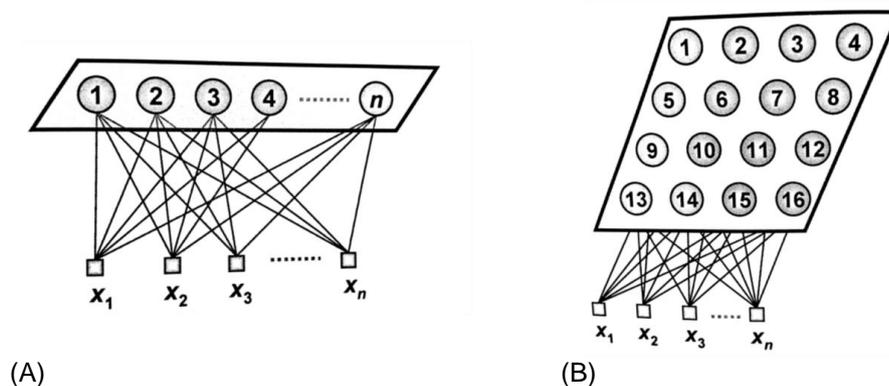
O neurônio j que obtiver a menor distância em relação ao vetor de entrada X será declarado vencedor e seus pesos serão ajustados de forma que ele se aproxime ainda mais do vetor de entrada X .

As conexões laterais são usadas para que um neurônio vencedor possa influenciar (colaborar) na resposta produzida pelos demais neurônios. A influência exercida pela conexão lateral entre dois neurônios vizinhos será proporcional a distância entre eles. As conexões laterais são fornecidas por mapas topológicos de vizinhança.

A Figura 18 mostra duas topologias para a implementação do Mapa auto-organizável, A Figura 18 (a) possui n neurônios organizados em uma dimensão, enquanto que na Figura 18

(b) os neurônios estão organizados em um mapa de duas dimensões. Nas duas figuras o vetor de $X=X_1..X_n$, representa o dado de entrada que será apresentado a todos os neurônios da topologia.

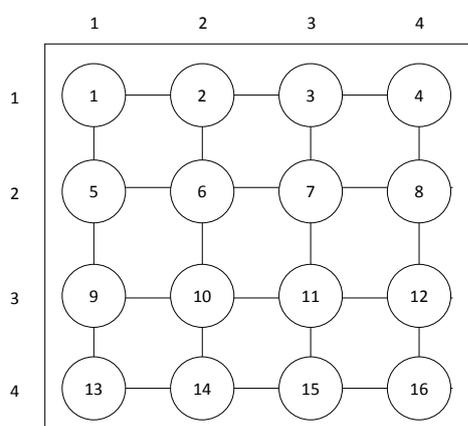
Figura 18 – Exemplos de topologias linear e em duas dimensões.



Fonte: SILVA, 2010.

Adicionalmente à topologia, é necessário especificar o critério de vizinhança entre os neurônios. Essa informação irá indicar como os neurônios irão cooperar com seus vizinhos. O exemplo da Figura 19 tem como critério de vizinhança o “raio de abrangência” (R), que irá definir para cada neurônio da topologia seus respectivos vizinhos. Considerando o neurônio de número 6, e um raio de abrangência igual a 1 (círculo), os vizinhos para esse neurônio seriam os de número 2,5,7 e 10. A Tabela 1 mostra todos os vizinhos para todos os neurônios considerando o raio de abrangência igual a 1. O símbolo $\Omega(n)$ indica o conjunto de vizinhos (n) com raio de vizinhança igual a (R).

Figura 19 – Mapa topológico com 16 neurônios.



Fonte: O autor, 2016.

Tabela 1 – Tabela de vizinhança com R=1

Conjunto (R=1)	Vizinhos
$\Omega_1^{(1)}$	2 e 5
$\Omega_2^{(1)}$	1,3 e 6
$\Omega_3^{(1)}$	2,4 e 7
$\Omega_4^{(1)}$	3 e 8
$\Omega_5^{(1)}$	1,6 e 9
(...)	(...)

Fonte: O autor, 2016.

Considerando que um determinado neurônio venceu a competição, para uma amostra de entrada, seu vetor de pesos e dos seus vizinhos serão ajustados. O maior ajuste será feito para o neurônio vencedor, ao passo que seus vizinhos serão ajustados com taxas menores.

A Figura 20 mostra o algoritmo de Kohonen para o treinamento competitivo de uma rede SOM. Os primeiros passos do algoritmo definem a topologia, o conjunto de vizinhança e a inicialização dos pesos dos neurônios (passos 1 e 2). Os passos 4 e 5 são referentes a recepção e normalização dos dados para treinamento. Os passos de 6 a 8 estão relacionados ao ajuste de parâmetros e ao treinamento propriamente dito (ajuste dos pesos dos neurônios vencedores e seus vizinhos).

Figura 20 – Algoritmo de treinamento de uma rede auto-organizável

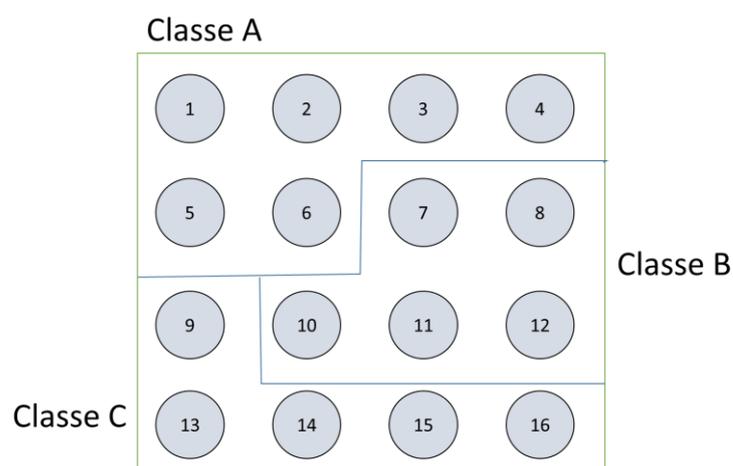
<ol style="list-style-type: none"> 1. Definir o mapa topológico da rede; 2. Montar o conjunto de vizinhos; 3. Iniciar o vetor de pesos de cada neurônio; 4. Obter o conjunto de amostras (dados) de treinamento; 5. Normalizar os vetores de amostras e de pesos; 6. Especificar a taxa de aprendizagem; 7. Iniciar o contador de número de <i>épocas</i>; 8. Repetir as instruções entre chaves (até que não haja mudança significativa entre os vetores) <ul style="list-style-type: none"> 8.1. Para todas as amostras de treinamento fazer: <ul style="list-style-type: none"> 8.1.1. Calcular a distância euclidiana entre o vetor de entrada e o vetor peso de cada neurônio; 8.1.2. Declarar como vencedor o neurônio que apresentar a menor distância euclidiana; 8.1.3. Ajustar o vetor de pesos do vencedor; 8.1.4. Ajustar o vetor de pesos dos neurônios na vizinhança do vencedor; 8.1.5. Normalizar o vetor de pesos dos neurônios ajustados anteriormente; 8.2. Fazer "<i>época</i> = <i>época</i> + 1"; 9. Analisar o mapa para extração de característica dos dados; 10. Identificar regiões que permitam definir as classes dos dados.

Fonte: O autor, 2016.

Os passos 9 e 10 podem requerer o uso de ferramentas estatísticas e análise de um especialista para classificar os grupos de dados de acordo com as suas características comuns. Uma vez definidos os grupos, o mapa topológico poderá ser dividido em regiões de acordo com a classificação realizada.

A Figura 21 representa o resultado de um Mapa auto-organizável dividido em regiões ao final da execução do algoritmo de Kohonen. Para esse exemplo podem ser observadas 3 classes *A*, *B* e *C*. A classe *A* é formada pelos neurônios (1, 2, 3, 4, 5 e 6), a classe *B* por (7,8,10,11 e 12) e *C* por (9,13,14,15 e 16). Assim, quando uma nova amostra (vetor de entrada) for submetida à rede SOM, um neurônio será sensibilizado (vencedor) e de acordo com o mapa, será possível determinar a que classe essa amostra pertence.

Figura 21 – Classificação de um Mapa auto-organizável.



Fonte: O autor, 2016.

4. TRABALHOS RELACIONADOS

Neste capítulo são apresentados os trabalhos relacionados ao uso de métodos de clusterização, apresentados no Capítulo 3, como ferramentas úteis para a detecção e localização de falhas em diferentes sistemas.

Em linhas gerais, os trabalhos avaliados atuam sob as três primeiras camadas da especificação OSA-CBM (*Open Systems Architecture for Condition-based Maintenance*) que prevê: aquisição e manipulação de dados, monitoração do sistema, avaliação de saúde do sistema (diagnóstico de falha), prognóstico (comportamento futuro) e tomada de decisão.

Em YIAKOPOULOS (2011) é proposto um modelo baseado em clusterização para realizar o diagnóstico automatizado de rolamentos defeituosos em máquinas utilizadas no ambiente industrial. Os diversos padrões de vibração dos rolamentos das máquinas são medidos e agrupados utilizando o método K-means. Os resultados são então analisados e classificados de acordo com o conhecimento de engenharia obtidos a partir da compreensão do comportamento físico dos enrolamentos defeituosos.

O artigo escrito por, (TAKEHISA, 2001) propõe um método de detecção de falhas para sistemas de naves espaciais com base em técnicas de mineração de dados. É construído um modelo de comportamento do sistema combinando métodos de clusterização e regras de mineração de dados. Os dados de manutenção do sistema, representadas diferentes séries temporais, são amostrados para formar padrões. Esses padrões são agrupados usando o método K-means, de acordo à similaridade entre eles, formando classes de padrões. A partir das classes formadas, são aplicadas regras para atribuir significado (presença de falha) as diferentes combinações de padrões obtidas.

Já KAWAMURA (2011) apresentam um método para avaliação de parâmetros de máquinas elétricas, utilizando os Mapas auto-organizáveis de Kohonen. Tal abordagem tem por objetivo o agrupamento das medidas de vibração das máquinas elétricas e da classificação dos conjuntos moto-redutores segundo suas possibilidades de ajuste.

Devido à complexidade de seus processos de tarifação e emissão de contas, as operadoras de telecomunicações possuem sérios problemas com atividades fraudulentas envolvendo seus assinantes. No estudo apresentado em (SEN, 2007) a clusterização é usada para identificar as características dos assinantes que apresentam comportamentos que os

diferencia da maioria dos assinantes. As informações sobre as contas e o registro detalhado das chamadas (CDR) dos assinantes são organizados compondo uma massa de dados. A partir dessa informação, é construído um modelo, baseado em Mapa auto-organizável de Kohonen, que compara as características do assinante maliciosos com as características de um novo assinante visando identificar se esse último, possui as mesmas características e assim identificar um possível assinante fraudador.

No trabalho apresentado por (FARIA, 2010) padrões de comportamento dos assinantes de uma rede telefônica presentes nos registros detalhados de chamadas (CDR) são identificados e analisados usando técnicas de mineração de dados e clusterização. Um fato relevante nesse estudo é a indicação que a utilização de CDRs em processos de clusterização requer um alto custo computacional devido ao volume de dados a ser tratado.

Em teoria, a forma mais simples de identificação de um problema em uma rede de telecomunicações é feita pela monitoração de alarmes dos equipamentos. Contudo, o crescimento das redes em tamanho e em complexidade fez com que a gerência de alarmes se tornasse um verdadeiro desafio ao operador, devido à quantidade de alarmes e notificações geradas. Em (BELLEC H., KECHADI M., 2006) é proposto um modelo, utilizando Sistemas Fuzzy, para o reconhecimento e classificação de falhas em redes de telecomunicações baseada em correlação de alarmes. O seu principal objetivo é reduzir consideravelmente o número de alarmes, agrupando-os segundo o seu comportamento, e assim formar eventos. Em seguida, esses eventos são correlacionados para formar grupamentos. Como resultado, apenas eventos globais são apresentados ao operador da rede, ajudando a identificar e solucionar os defeitos na rede.

Conforme visto nos artigos estudados é recorrente a utilização de agrupamento para detecção de falhas e otimização de processos. No capítulo seguinte apresentamos um estudo com 3 métodos de clusterização e uma avaliação do seu desempenho em relação ao conjunto de dados sob estudo.

5. TESTES INICIAIS

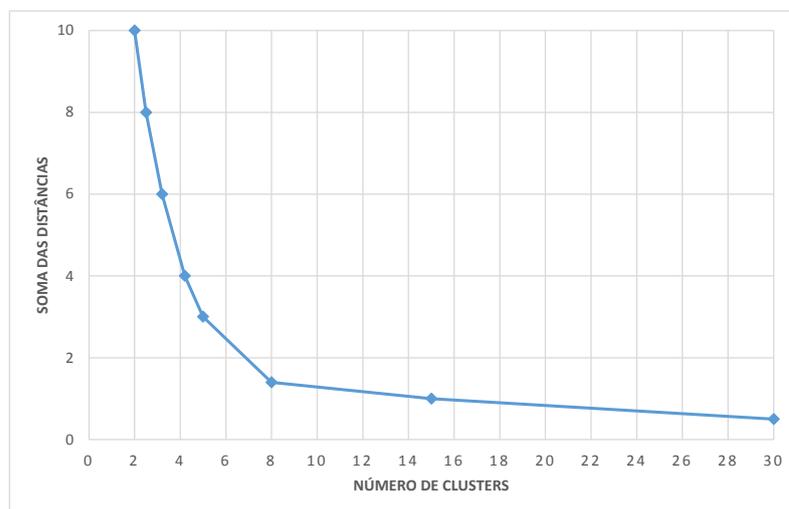
Após a apresentação dos métodos de clusterização no Capítulo 3, este capítulo mostra os testes realizados com três diferentes tipos de clusterização (K-means, Clusterização Hierárquica e Mapas auto-organizáveis de Kohonen). O objetivo é avaliar a eficiência desses métodos na formação de clusters quando aplicados ao conjunto de dados descritos no Capítulo 2. A avaliação da eficiência é feita com base em quatro critérios de desempenho (Silhouette, Compacidade e Separação, Critério Dunn's e Calinski Harabasz) (BERNARD, 2013) aplicados individualmente sobre o resultado da clusterização feita por cada um dos métodos. A avaliação dos valores obtidos pelos critérios de desempenho permite compará-los entre si e determinar quais métodos de clusterização utilizados são adequados para a construção do modelo proposto nesse trabalho.

5.1 Rotina de teste

A rotina de testes adotada como padrão para todos os métodos consiste em realizar uma sequência de clusterizações variando o número de clusters, linearmente, de 1 até 60 clusters. Para cada clusterização, é registrada a soma das Distâncias Euclidianas dos pontos aos centros dos clusters. A partir dos resultados obtidos nesse processo, é estimado o número de clusters que melhor representa os padrões existentes nos dados. Essa estimativa é feita pela análise do gráfico plotado a partir do número de clusters contra o somatório das Distâncias Euclidianas dos pontos aos centros dos clusters (KUMAR, 2006). Para efeito de demonstração, o gráfico apresentado na Figura 22, mostra o número de clusters contra o somatório das distâncias dos pontos aos centros dos clusters, ambos gerados a partir de uma base de dados artificial e utilizando o algoritmo K-means para realizar o agrupamento. Percebe-se que há uma depressão ou “joelho” no gráfico sobre o cluster de número 8, que representa o número estimado de cluster a ser utilizado (KUMAR, 2006). Esta mesma avaliação foi realizada para cada um dos métodos de clusterização, a partir das massas de dados de rede para se estimar a melhor técnica de clusterização para o nosso problema. O

número máximo de clusters gerados por cada método foi fixado em 60, como o objetivo de padronizar os gráficos gerados para cada um dos métodos e permitir compará-los entre si.

Figura 22 – Soma de distâncias dos dados ao centro dos clusters



Fonte: O autor, 2016.

Para a execução da rotina de testes, foram utilizadas as funções disponíveis no Matlab relacionadas a cada um dos métodos de clusterização que propomos testar. Os parâmetros comuns a cada função (i.e., número de clusters), quando possível, foram preenchidos com os mesmos valores para obter resultados de distância dos pontos aos centros dos clusters na mesma ordem de grandeza e assim facilitar a comparação dos resultados gráficos.

A seguir são apresentadas em detalhes as funções, comandos e parâmetros do Matlab utilizados nos testes. É feita a análise dos gráficos gerados a partir dos resultados obtidos e realizada a estimativa do número de clusters para cada método testado.

5.2 Clusterização usando K-means

A função denominada K-means, disponível no Matlab para realizar a clusterização, possibilita agrupar elementos de um conjunto de dados em k subconjuntos mutuamente exclusivos, e retornar o índice do subconjunto para a cada observação. O objetivo é encontrar uma partição na qual os pontos dentro de cada cluster (subconjunto) sejam tão próximos uns dos outros quanto possível, e ao mesmo tempo tão longe de objetos localizados em outros

grupos quanto possível. A função permite escolher entre cinco diferentes medidas de distância (Tabela 2), dependendo do tipo de dados que se deseja particionar.

Cada cluster é constituído por seus “objetos membros” e por seu centroide, ou centro. O centroide para cada cluster é o ponto para o qual a soma das distâncias de todos os objetos desse cluster é minimizada durante a execução da função.

Tabela 2 – Medidas de distância da função K-means usada no Matlab

Medida de distância	Descrição
squeclidean	Distância euclidiana ao quadrado (padrão). Cada centroide é a média dos pontos em cada cluster
cityblock	Soma das diferenças absolutas. Cada centroide é representado pela média de todos os pontos cada cluster.
cosine	Um menos o cosseno do ângulo incluído entre os pontos (tratados como vectores) . Cada centroide é a média dos pontos no cluster , depois de normalizados para a unidade de distância euclidiana
correlation	Um menos a correlação da amostra entre os pontos (tratados como sequências de valores) . Cada centroide é a média dos componentes dos pontos no cluster, depois de centrar e normalizar esses pontos com média zero e desvio padrão unitário .
hamming	Esta medida só é adequado para dados binários
	É a proporção de bits que diferem . Cada centroide é a mediana componente -wise de pontos nesse cluster.

Fonte: O autor, 2016.

Para minimizar a soma das distâncias de cada objeto ao centro do cluster, a função K-means move objetos entre clusters até que a menor soma seja alcançada. O resultado dessa iteração é um conjunto de clusters que são tão compactos e separados quanto possível. Essa operação é realizada para todos os clusters formados durante o processo. É possível controlar os detalhes da minimização usando vários parâmetros opcionais de entrada para o algoritmo *K-means*, incluindo os centroides iniciais e também o número máximo de iterações.

O comando mostrado na Equação 8 é um exemplo de uso da função K-means disponível no Matlab. Os parâmetros listados a seguir representam, respectivamente, as entradas e saídas da função.

- *X*: nome do arquivo do conjunto de dados a ser particionado (entrada);
- *k*: número de partições que se deseja obter (entrada);
- *idx*: vetor que relaciona cada linha da matriz de entrada a um índice que representa a qual partição, ou grupo, o registro foi classificado (saída);
- *ctrs*: matriz que indica as coordenadas do centroide de cada partição (saída);

- smd : vetor que representa a soma das distâncias de todos os membros de uma partição ao centroide (saída).

$$[idx, ctrs, smd] = kmeans(X, k); \quad (8)$$

5.2.1 Testes usando a função K-means

Os testes com a função K-means foram realizados usando quatro das cinco medidas de distância apresentadas na

Tabela 2. Apenas a “*hamming*” não foi utilizada, pois sua aplicação é válida somente para conjunto de dados com valores binários.

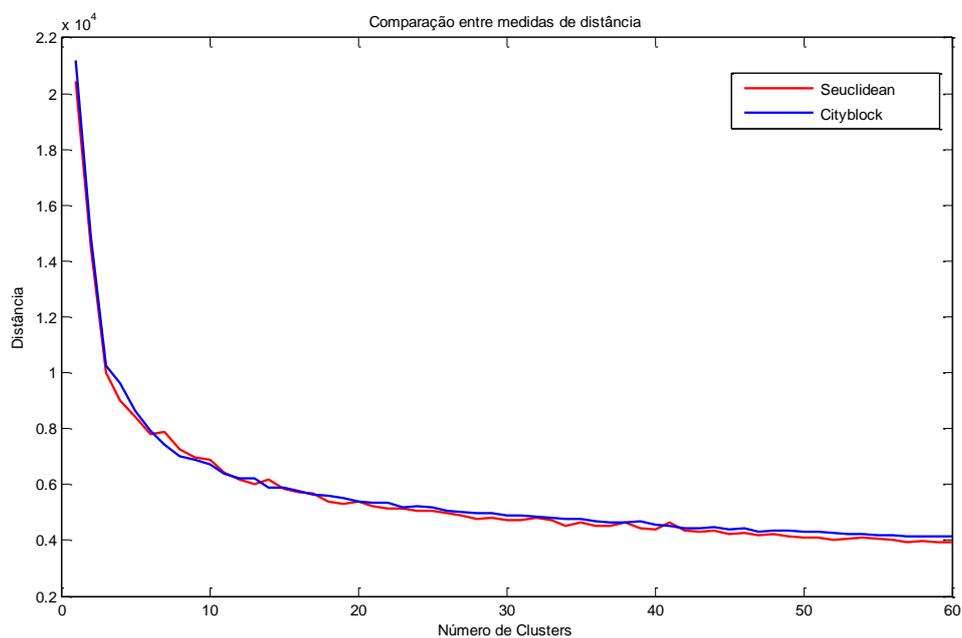
Foram realizadas 60 clusterizações para cada uma das medidas de distância disponíveis na função K-means e contabilizado o somatório das distâncias dos objetos ao centro dos clusters. Os gráficos obtidos nessa operação estão representados na Figura 23 e Figura 24.

É possível observar uma diferença relativamente grande para os pares de médias de distância representados:

- na Figura 23, as medidas *Cityblock* e *Sqeuclidean* apresentam resultados próximos;
- na Figura 24, as medidas *Cosine* e *Correlation* são próximas entre si, mas com valores maiores que os apresentados pelas outras medidas.

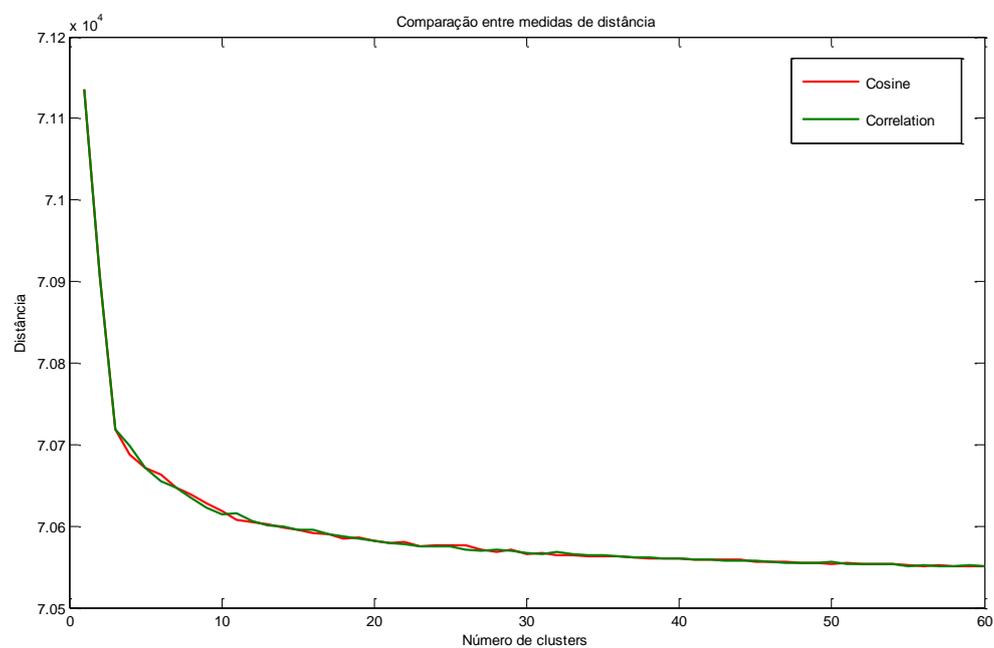
O número estimado de clusters (joelho da curva) obtido para as quatro curvas da Figura 23 e Figura 24, a partir da métrica descrita na Seção 0, é igual a 10.

Para as medidas de desempenho serão adotados os valores obtidos usando a medida de Distância Euclidiana por apresentar menores valores nos gráficos de distância \times clusters quando comparada com as medidas *Cosine* e *Correlation*, e possuir formula de cálculo mais simples quando comparada com *Cityblock*.

Figura 23 – K-means usando medidas de distância *Euclidiana* e *Cityblock*

Fonte: O autor, 2016.

Figura 24 – K-means usando medidas de distância Cosine e Correlation.



Fonte: O autor, 2016.

As quatro medidas de distância utilizadas na clusterização pelo método K-means apresentaram a mesma estimativa para quantidade de clusters (10 clusters), mostrando consistência e a capacidade do método para a formar grupos a partir da massa de dados de teste apresentada no Capítulo 2.

5.3 Clusterização Hierárquica

Conforme apresentado na Seção 0, a Clusterização Hierárquica é um método desenvolvido para agrupar os dados de forma hierárquica utilizando para isso medidas de similaridade (distância) entre eles.

O Matlab disponibiliza, através da *Toolbox* de Estatísticas, funções que permitem realizar a Clusterização Hierárquica para um conjunto de dados. O resultado obtido pelo uso das funções é o agrupamento dos dados em uma “árvore” que pode ser visualizada graficamente por um dendrograma, conforme o demonstrado na Figura 12.

O método de clusterização consiste de três etapas básicas:

- Etapa 1: Encontrar a similaridade ou dissimilaridade entre cada par de objetos do conjunto de dados;
- Etapa 2: Agrupar os objetos sob a forma de uma árvore hierárquica (agrupamento binário);
- Etapa 3: Determinar onde “cortar” a árvore hierárquica para obter os clusters desejados.

Etapa 1: Para encontrar a similaridade de um conjunto de dados, são calculadas as distâncias entre todos os objetos do conjunto, dois a dois, usando a função *pdist*. É possível determinar a quantidade de pares formados por essa função utilizando a Equação 9, onde m é o número de objetos que forma o conjunto.

$$n^{\circ} \text{ de pares} = m(m-1)/2 \quad (9)$$

A função *pdist* permite calcular a distância entre os objetos. Podem ser usadas seguintes métricas:

- Euclidean
- seclidean
- cityblock
- minkowski
- chebychev
- mahalanobis
- cosine
- correlation
- spearman
- hamming
- jaccard

Etapa 2: Nesta etapa, o agrupamento de objetos é feito pela associação dos pares que estiverem mais próximos. Para isso é usada a função *linkage*. Essa função usa a informação de distância gerada pela função *pdist* para determinar o quanto os objetos estão próximos um dos outros e com isso formar pares de objetos; os pares de objetos são agrupados em clusters cada vez maiores até que seja formada uma árvore hierárquica de acordo com a similaridade entre os objetos. A listagem a seguir mostra os algoritmos usados pela função *linkage*.

- *Average*;
- *Centroid*;
- *Complete*;
- *Median*;
- *Single*;
- *Ward*;
- *Weighted*.

Após o uso da função *linkage* já é possível representar graficamente a árvore formada usando a função *dendrogram*.

Antes que a terceira etapa seja realizada, é necessário aferir a formação da árvore hierárquica, para isso são usadas medidas de consistência e dissimilaridade:

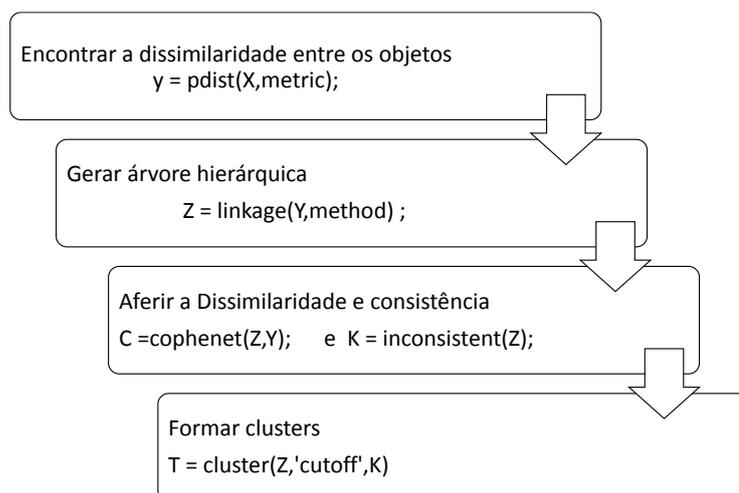
- **Consistência:** Permite determinar a divisão natural entre clusters pela comparação dos pesos (distâncias) entre links adjacentes da árvore hierárquica. Quanto maior o valor atribuído pela inconsistência, maior a separação entre os clusters. A função utilizada para medidas de consistência no Matlab é chamada *inconsistent*.
- **Dissimilaridade:** Medida que permite avaliar a “qualidade” da formação da árvore pela comparação entre as distâncias cofenéticas (*cophenetic distances*) e as distâncias entre os objetos, obtida pela função *pdist*. Como resultado é gerado um valor chamado coeficiente de correlação cofenética (*cophenetic correlation coefficient*). Quanto mais

próximo de 1 for o coeficiente, mais acurada é a formação dos clusters. Para executar essa medida é usada a função *cophenet*.

Etapa 3. Nesta etapa é usada a função *cluster* onde é definida o “ponto de corte” da árvore hierárquica mais adequado para a formação dos clusters. Essa função permite que os clusters sejam criados pela detecção natural dos grupos ou ainda pela escolha arbitrária do ponto de corte da árvore hierárquica. Uma das opções disponíveis na função permite usar como parâmetro de corte os valores de inconsistência aferidos pela função *inconsistente*.

A Figura 25 representa, de forma resumida, as etapas descritas para a clusterização hierárquica usando as funções do Matlab:

Figura 25 – Funções para Clusterização Hierárquica no Matlab.



Fonte: O autor, 2016.

5.3.1 Testes usando Clusterização Hierárquica

Nos testes realizados com a Clusterização Hierárquica foi identificado que a medida de Distância Euclidiana (formação de pares) produziu melhores resultados de clusterização de acordo com a métrica estabelecida pela função *cophenetic*. A partir da medida de Distância Euclidiana, foram testados todos métodos disponíveis no Matlab para a construção da árvore hierárquica (função *linkage*).

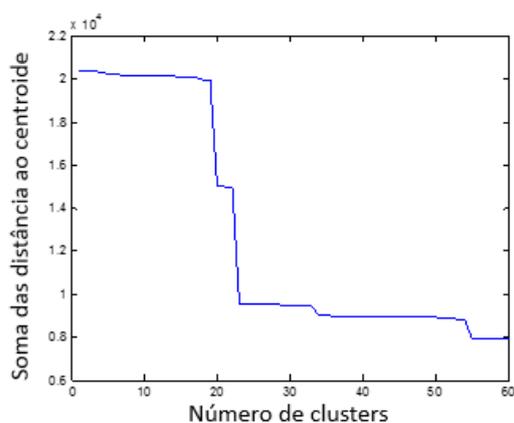
De forma padronizada, conforme descrito na Seção 0, foram traçados os gráficos das distâncias de todos os pontos ao centro de cada cluster, variando o número de clusters de 1 até 60.

É importante ressaltar que, o método da clusterização hierárquica não define, explicitamente, o centroide de cada cluster. Os grupos de objetos são formados pelas distâncias entre os membros e, para cada nova formação ou grupamento, é considerada a distância entre os “centros de cada subgrupo”. Essa rotina se repete até completar toda a hierarquia. Por isso, os centroides e distâncias dos objetos ao centro de cada cluster foram calculados após o término da clusterização hierárquica.

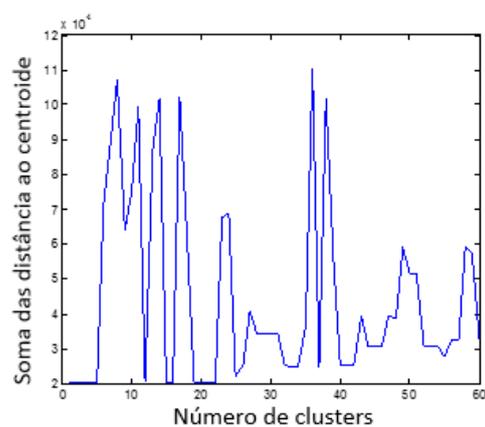
As Figura 26 e Figura 27 mostram os gráficos das distâncias dos pontos aos centros dos clusters obtidos usando a Clusterização Hierárquica. Os gráficos da Figura 26 representam a clusterização feitas a partir dos algoritmos *average*, *single*, *complete*, *median*, *weighted* e *centroid*, respectivamente. Esses gráficos mostram um comportamento inesperado para a variação das distâncias em relação ao número de clusters, conforme apresentado na Seção 5.1 (Figura 22). Para os gráficos da Figura 26 (a), (c), (d), (e) e (f), a variação do somatório das distâncias em relação ao número de clusters ocorre em saltos, à medida que o número de clusters aumenta. Para o gráfico da Figura 26 (b), são verificados picos com valores elevados de distância em diferentes quantidades de clusters. A análise gráfica mostra que os algoritmos citados anteriormente não conseguem agrupar os dados de forma adequada. Por isso, eles não podem ser usados no processo de Clusterização Hierárquica.

Dentre os algoritmos usados para a Clusterização Hierárquica, o algoritmo *Ward* (Figura 27) apresenta uma distribuição esperada homogênea, do somatório das distâncias dos pontos ao centro dos clusters, à medida que o número de clusters aumenta. Esse resultado pode ser comparado ao obtido pela clusterização usando K-means. E, é o único onde pode ser aplicada a estimativa da quantidade de clusters, que também é 10.

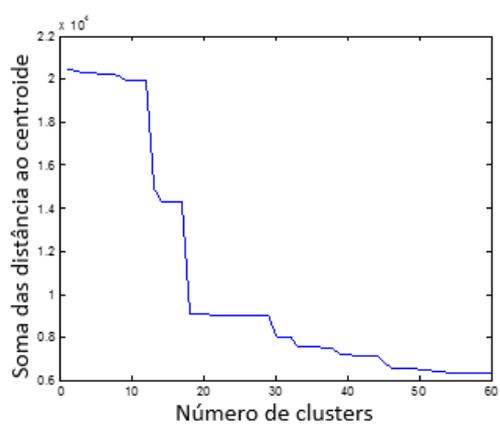
Figura 26 – Algoritmos testados na Clusterização Hierárquica



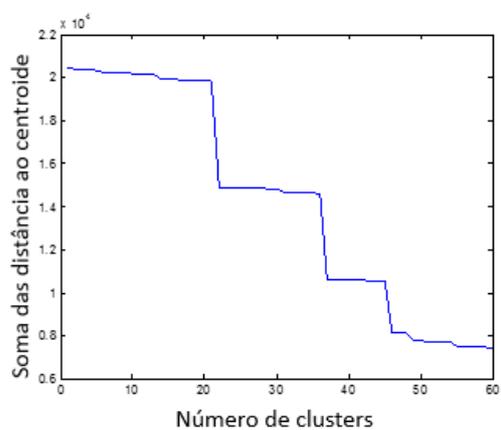
(a)



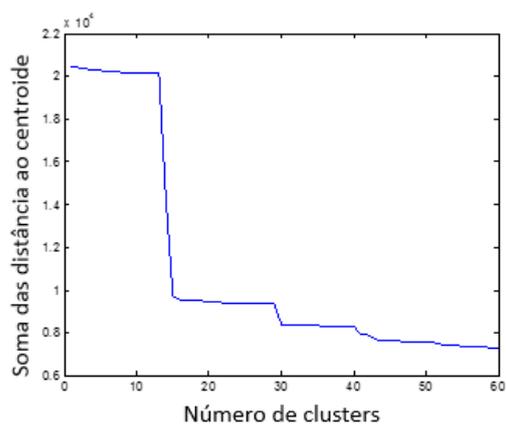
(b)



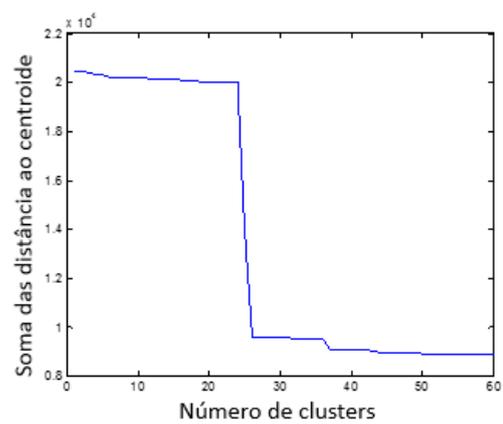
(c)



(d)



(e)

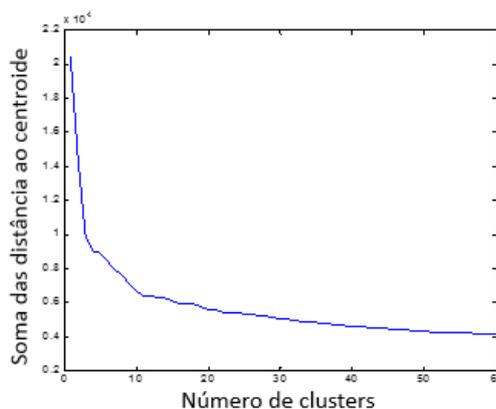


(f)

Legenda: Gráficos das distâncias para os Algoritmos (a) - Average; (b) - Single; (c) - Complete; (d) - Median; (e) - Weighted e (f) - Centroid

Fonte: O autor, 2016.

Figura 27 – Gráfico das distâncias para o Algoritmo Ward



Fonte: O autor, 2016.

5.4 Clusterização usando Mapas auto-organizáveis de Kohonen

Para a classificação dos dados usando Mapa auto-organizáveis de Kohonen, o Matlab oferece a função *newsom* encontrada na *Toolbox* para Redes Neurais. A função trata os dados como vetores de entrada e os agrupa, associando para cada vetor um neurônio pertencente a uma estrutura criada, denominada mapa. A estrutura criada pode conter uma ou mais dimensões. Em uma estrutura unidimensional cada neurônio possui, no máximo, dois vizinhos. O desempenho da rede não é afetado pela topologia, contudo uma melhor visualização de vizinhança é obtida usando um espaço bidimensional. Para efeito de testes as duas estruturas foram consideradas.

Diferentes topologias estão disponíveis na função para a construção do mapa: *gridtop*, *hextop*, e *randtop*.

Topologia *gridtop*: A topologia *gridtop* é constituída por neurônios organizados na forma uma grade retangular. A Figura 28 representa uma topologia *gridtop* com 25 neurônios (5x5). Os neurônios nessa topologia, dependendo de sua posição, possuem 2 ou 4 vizinhos. Neurônios periféricos possuem 2 vizinhos e neurônios centrais possuem 4.

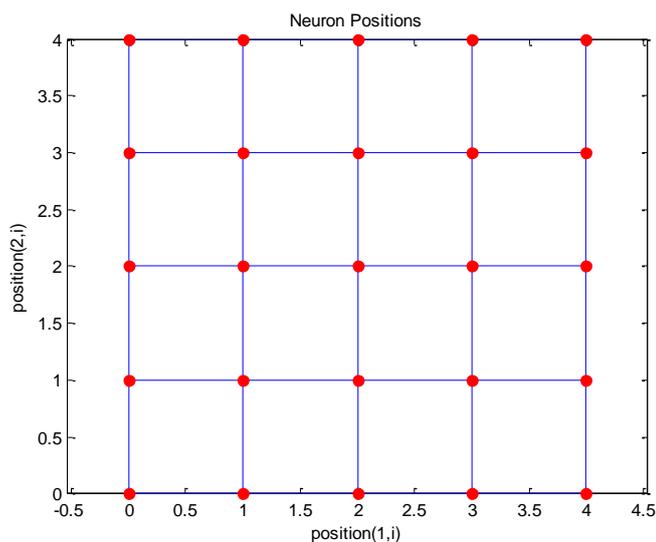
Topologia *hextop*: A topologia *hextop* os neurônios são organizados num padrão hexagonal. Os neurônios podem ter 2, 3, 5 ou 6 vizinhos dependendo de sua posição.

A Figura 29 mostra uma topologia *hextop* com 25 neurônios. Essa topologia é gerada por padrão pela função *newsom*.

Topologia *randtop*: Os neurônios nessa topologia são criados em um arranjo aleatório multidimensional. Por isso, não é possível determinar no momento de sua construção a quantidade de vizinhos para cada neurônio. A Figura 30 mostra uma topologia *randtop* com 24 neurônios. Podemos encontrar nessa topologia neurônios com 2, 3, 5 e até com 7 vizinhos, independente da sua posição.

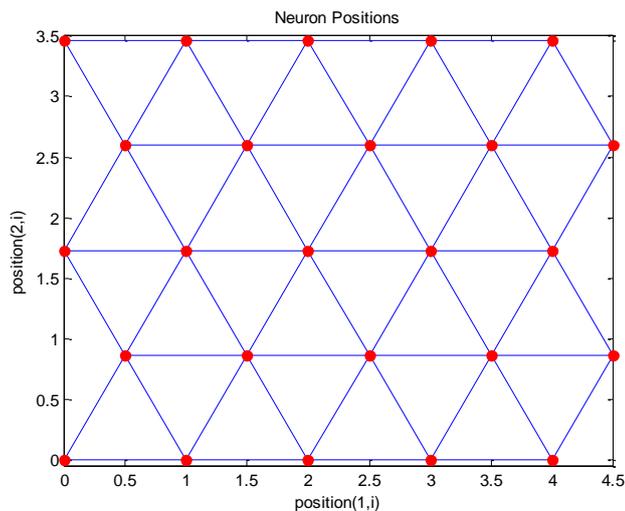
Dentre as topologias descritas, a topologia *hextop* foi adotada para os testes, pois possui uma quantidade maior de vizinhos em relação a topologia *gridtop* e simplifica a visualização e análise de vizinhança (bidimensional) em relação a topologia *randtop* (multidimensional).

Figura 28 – Topologia gridtop.



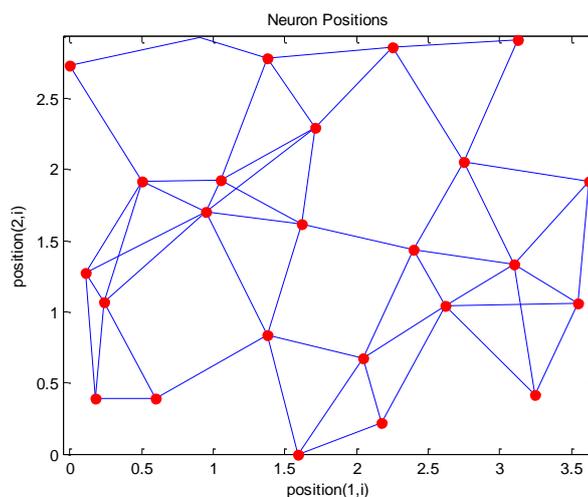
Fonte: O autor, 2016.

Figura 29 – Topologia hextop.



Fonte: O autor, 2016.

Figura 30 – Topologia randtop.



Fonte: O autor, 2016.

O Matlab possibilita usar diferentes medidas (*dist*, *linkdist*, *mandist*, *boxdist*) para calcular a distância entre neurônios, o que permite ao algoritmo determinar para um neurônio específico quem são os seus vizinhos:

Os quatro métodos para o cálculo de distância foram utilizados, mas essa variável não influenciou no resultado dos testes. Por isso, o método a ser considerado em todos os testes realizados é a distância euclidiana (*dist*), definida como padrão pelo Matlab.

Para criar a rede neural é usada função *newsom* representada pela Equação 10, onde P indica a matriz de dados (vetor de entrada), D_1 à D_n especificam as dimensões do Mapa,

$TFCN$ (preenchido com a topologia adotada – *hextop*), $DFCN$ (preenchido com a medida de distância – *linkdist*).

$$net=newsom(P,[D_1,D_2,\dots,D_n],TFCN,DFCN); \quad (10)$$

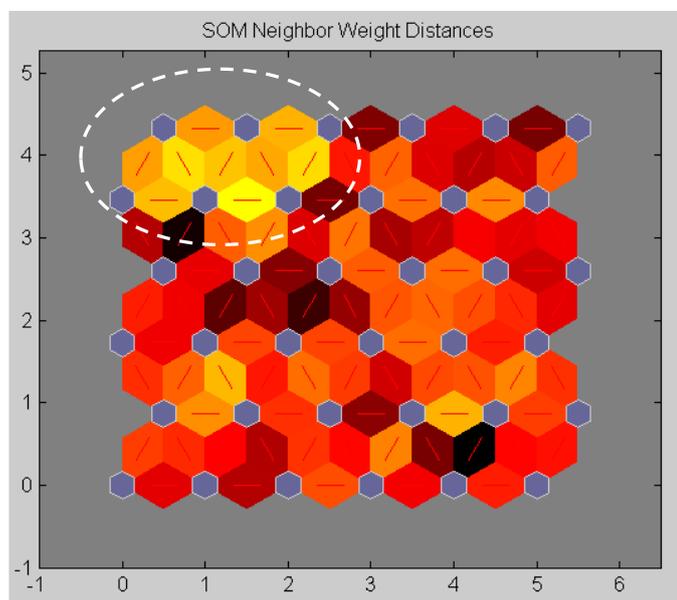
Na criação da rede, os pesos dos neurônios são escolhidos aleatoriamente. Para que a rede criada “aprenda” como os dados devem ser distribuídos de acordo com a topologia adotada, é necessário que ela seja treinada, isso é, a rede identifica o neurônio vencedor para cada vetor (dado) de entrada. O vetor vencedor e seus vizinhos são então reposicionados. Esse processo ocorre para todos os dados de teste. A função *train* parametrizada na Equação 11 é usada para treinamento da rede. Os argumentos da função representam a rede a ser treinada (*net*) e os dados usados para treinamento (*dados_t*). O resultado de saída da função, que representa a rede treinada, é armazenado na variável *rede_t*.

$$rede_t=train(net,dados_t); \quad (11)$$

Após o treinamento, a rede neural está apta a classificar qualquer nova entrada (vetor). A Toolbox do Matlab permite avaliar graficamente, através da figura plotada pela função *plotsomnd*, a distância entre neurônios vizinhos para uma rede que já tenha sido treinada. A distância entre neurônios vizinhos, representada no gráfico da topologia da rede (Figura 28), é convertida em um código de cores: Quanto mais clara a cor entre os neurônios, mais próximos eles estão e quanto mais escuro mais distante. Com isso, pode ser ter uma ideia da formação de grupos. A

Figura 31 mostra uma topologia *hextop* com a distância entre seus neurônios representada por cores. A partir da análise da distribuição das cores, pode ser identificada pelas cores claras do mapa, a formação de um grupo com 6 neurônio. O grupo foi destacado por uma linha pontilhada para facilitar sua identificação.

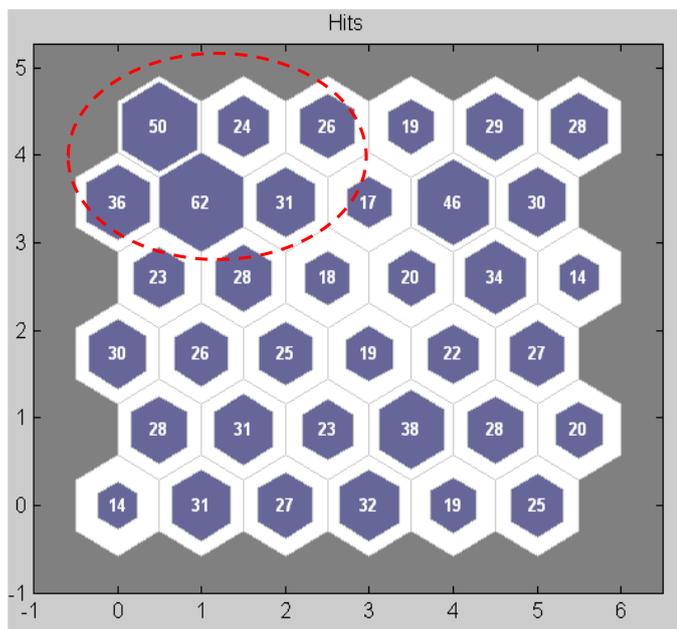
Figura 31 – Distância representada por código de cores.



Fonte: O autor, 2016.

Para visualizar quantos pontos do conjunto de dados estão associados a um determinado neurônio é usada a função comando *plotsomhits*. Assim, pode ser verificado como os dados estão distribuídos pelos neurônios que formam a topologia. Quanto maior a quantidade de dados associados a um neurônio maior o preenchimento, na cor azul, do hexágono que o representa na figura. A Figura 32 representa a distribuição de um conjunto de dados distribuídos por uma topologia *hextop* 6×6. O hexágono com o número 62 representa a maior quantidade de dados associadas a um neurônio. Por isso ele está completamente preenchido na cor azul. A linha pontilhada em vermelho na Figura 32 indica os mesmos neurônios mostrados na Figura 31

Figura 32 – Distribuição gráfica usando o comando `plotsomhits`.



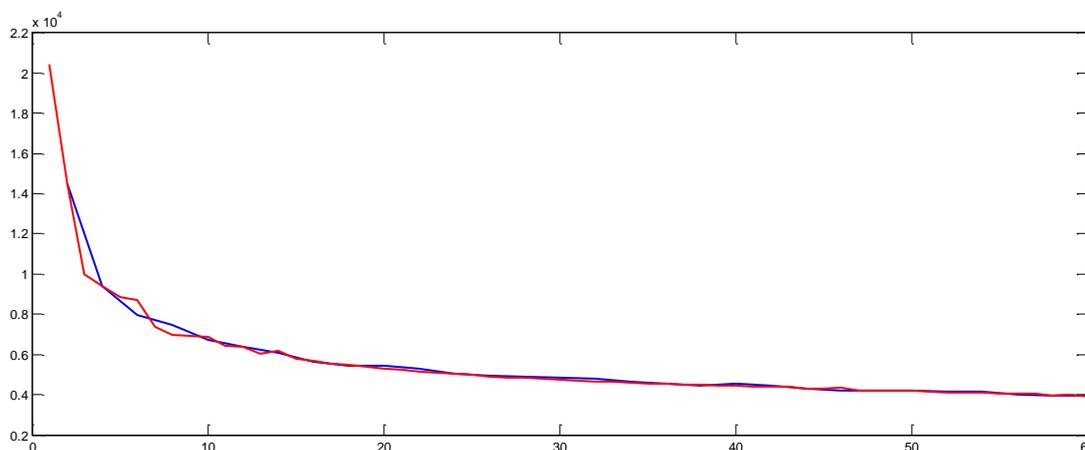
Fonte: O autor, 2016.

5.4.1 Testes com Mapa auto-organizáveis de Kohonen

Seguindo o padrão adotado como rotina de testes (seção 0) foram geradas sessenta redes auto-organizáveis usando as topologias linear e *hextop* para estimar a quantidade de neurônios (clusters). A quantidade de neurônios utilizada em cada rede corresponde à mesma quantidade de clusters usadas nos métodos já testados nas subseções 5.2.1 e 0. Assim, as 60 redes foram geradas variando linearmente a quantidade de neurônios de 1 até 60. Para traçar o gráfico de distância dos pontos ao centro dos clusters usando Mapas auto-organizáveis é importante identificar a correspondência entre as variáveis cluster e centroide, uma vez que tais variáveis não são conceitualmente definidas quando se trata de Redes Neurais. Para cada neurônio em um Mapa auto-organizável pode ser atribuído a identidade de um cluster. Os pesos associados a esse neurônio correspondem às coordenadas dos centros do cluster. A

partir dessa correlação, as distâncias dos pontos de cada cluster aos centroides foram calculadas e plotado o gráfico apresentado na Figura 33.

Figura 33 – Gráfico das distâncias para as topologias linear e hextop.



Fonte: O autor, 2016.

A Figura 33, mostra a curva para a topologia linear (linha vermelha) partindo de uma distância maior (aproximadamente 2×10^4), pois a quantidade inicial de neurônios utilizada no início do teste é de apenas um neurônio. Enquanto que a topologia *hextop*, inicialmente é construída por 2 neurônios (linha azul). Observando o gráfico, a partir da quantidade de neurônios igual a 2, pode-se afirmar que não há diferença significativa entre os resultados obtidos pela clusterização usando as topologias linear ou *hextop*, pois as curvas que as representam se sobrepõem à medida que o número de neurônios aumenta. A quantidade de neurônios (cluster) estimada para as duas curvas, de acordo com o conceito apresentado em KUMAR (2006) e descrito na subseção 5.1, é de 10 neurônios.

5.5 Medidas de desempenho

Após a avaliação individual das técnicas de clusterização selecionadas, um próximo passo foi dado para se avaliar qualitativamente e comparativamente cada uma das técnicas. Para esta avaliação foram adotados critérios de desempenho, discutidos nas próximas subseções, que pudessem ser aplicados de forma consistente a todas as técnicas. As referências para a realização das medidas de desempenho foram as classificações feitas pelos

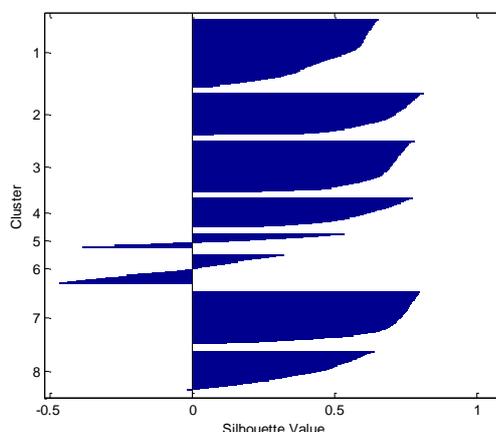
três métodos para o conjunto de dados. A quantidade de cluster para todos os métodos foi fixada em 10, que é a quantidade mínima de clusters obtida pela estimativa nas subseções 5.2.1, 5.3.1 e 5.4.1 para os três métodos.

5.5.1 Critérios de desempenho

Para a avaliação dos resultados dos testes foram selecionados 4 critérios: Silhouette, Compacidade e Separação, critério Dunn's e Calinski Harabasz. Cada um deles é usado para avaliar numericamente a qualidade dos clusters formados por cada método. Para isso, esses critérios relacionam de formas diferentes as medidas das distâncias euclidianas entre os pontos do conjunto de dados do teste e os centroides de cada cluster formado. A seguir é feita uma breve descrição sobre cada um dos critérios e, ao final desse capítulo, são apresentados os resultados obtidos pela aplicação de cada um deles aos resultados do teste.

Silhouette (ROUSSEEUW P, 1987) é a medida que permite avaliar o quanto cada ponto em um cluster está próximo dos pontos dos clusters vizinhos. Essa medida varia de +1 à -1. O valor +1 indica que o ponto está muito distante dos clusters vizinhos. Valores próximos de 0 indicam a imprecisão na definição a qual cluster um ponto pertence. Valores próximos a -1 indicam que o ponto não foi corretamente associado ao cluster. A Figura 34 representa o gráfico de Silhouette produzido a partir da clusterização feita para um conjunto de dados qualquer.

Figura 34 – Gráfico de Silhouette para dados gerados aleatoriamente no Matlab.



Fonte: O autor, 2016.

A Compacidade e Separação é considerada uma das medidas de validação de categorias mais completas, uma vez que avalia não somente a compacidade das categorias, mas também a qualidade de separação entre estas, não perdendo exatidão mesmo quando a sobreposição de categorias é alta. Quanto menor for o valor do critério de Compacidade e Separação, melhor será a disposição entre as categorias, cuja formulação encontra-se na Equação 12, onde v_i representa os centros dos clusters, \mathcal{X} representam os pontos, n o número de pontos, C o número de clusters e μ_{ij} é igual a 1 se o ponto \mathcal{X}_j pertence ao cluster n_i (FRANCO, 2002).

$$CS = \frac{\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^2 \|v_i - x_j\|^2}{n \min_{ij} \|v_i - v_j\|} \quad (12)$$

O Critério Dunn's é definido pela relação entre a menor distância intracluster e a maior distância intercluster. Sua formulação é apresentada Equação 13 (DUNN, 1974).

$$D = \frac{d_{min}}{d_{max}}, \quad (13)$$

O termo d_{min} corresponde a menor distância entre dois objetos de diferentes clusters (intercluster), e d_{max} representa a maior distância entre dois objetos do mesmo cluster (intracluster). Essa medida visa encontrar o número adequado de clusters para um conjunto de dados, onde o maior valor obtido para esse critério sugere o melhor particionamento para o conjunto de dados, isto é, partições onde exista a maior separação entre os clusters e a menor distância entre os elementos dentro de cada cluster.

O critério Calinski-Harabaz avalia a formação de cluster baseado na média da soma dos quadrados das distâncias entre o centro dos clusters e das distâncias dos pontos aos centros dos seus respectivos clusters (YANCHI, 2010). Sua formulação indicada com VRC é definida nas Equações 14, 15 e 16:

$$VRC_k = \frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)} \quad (14)$$

$$SS_B = \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (15)$$

$$SS_W = \sum_{i=1}^k \sum_{x \in c_i} \|x - m_i\|^2 \quad (16)$$

Onde SS_B (y) é a variância global entre os clusters e SS_W (z) a variância global dentro do cluster. As variáveis de cada formulação são definidas abaixo:

- k – Número de clusters.
- N – Número de elementos do conjunto de dados.
- n_i – Quantidade de elementos do cluster i .
- m_i – Centro do cluster de índice i .
- m – média global do conjunto de dados.
- x – cada elemento do cluster de índice i .
- c_i – Cluster indicado pelo índice i .

Clusters bem definidos possuem valores altos para SS_B e valores pequenos para SS_W . Desse modo, quanto maior o valor do índice VRC, melhor a clusterização dos dados.

5.5.2 Resultados encontrados

A Tabela 3 consolida as medidas obtidas para cada um dos critérios aplicados aos três métodos de clusterização utilizados nesse trabalho. Os critérios foram calculados para uma clusterização usando 10 clusters sobre os dados do projeto. A primeira coluna da tabela contém os três métodos de clusterização. A segunda coluna mostra o valor médio de Silhouette para 10 inicializações diferentes. As colunas restantes apresentam os resultados critérios de avaliação Compacidade e Separação (CS), Dunn's (Dunn) e Calinsk-Harabarz (CH).

Tabela 3 – Critérios de desempenho aplicados aos testes iniciais

Método	Sillhouette	CS	Dunn's	CH
K-means	0,5314	0,7267	0,0087	6,4717E+03
Kohonen	0,5289	1,0719	0,0063	6,3291E+03
Cluster hierárquico	0,4790	1,0795	0,0141	6,2892E+03

Fonte: O autor, 2016.

Para os valores médios de Sillouette os métodos K-means e Kohonen obtiveram os melhores resultados (valores mais próximos a 1). O mesmo é verificado para o critério Compacidade e Separação (CS) e Calinsk-Harabarz (CH). Apenas para o índice Dunn's, o método Cluster Hierárquico apresentou melhor resultado (maior valor obtido).

A análise de desempenho mostrou que os métodos K-means e Mapas Auto-organizáveis de Kohonen apresentaram melhores resultados em 3 dos 4 índices adotados. Por esse motivo, os dois métodos clusterização foram considerados na metodologia apresentada a seguir. O agrupamento realizado por meio de cada algoritmo será avaliado do ponto de vista do especialista da rede, esse evento irá mostrar a aplicabilidade de cada método na construção do modelo.

6. METODOLOGIA PROPOSTA

A presente proposta tem por base a clusterização dos dados (Códigos de Liberação) feita pelos métodos K-means, Mapas auto-organizáveis e Cluster Hierárquico, uma vez que os resultados obtidos na análise de desempenho (Capítulo 5), não foram suficientes para determinar qual dos três métodos seria o mais adequado para segmentar conjunto de dados usados nesse trabalho em diferentes grupos e assim, identificar e classificar as falhas de uma rede SMP.

Os resultados obtidos, a partir da clusterização feita pelos métodos de clusterização no Capítulo 5, foram submetidos à análise do especialista que identificou as falhas em cada grupo e classificou os grupos formados. A eficiência de cada método em relação a identificação de falhas e formação dos grupos foi aferida e o Mapa auto-organizável de Kohonen foi considerado, pelo especialista, a técnica mais adequada para a construção do modelo em função da sua capacidade de distribuir as falhas em diferentes grupos.

Para refinar a identificação dos padrões, a quantidade de cluster (neurônios) do Mapa Auto Organizável foi aumentada. O resultado obtido foi ainda melhor, considerando a distribuição das falhas em grupos mais específicos e a separação de diferentes padrões de comportamento de tráfego como será mostrado a seguir.

6.1 Clusterização

Os resultados das clusterizações obtidos nos testes realizados no Capítulo 5 foram organizados em três tabelas. A Tabela 4 contém o resultado obtido pela clusterização feita por K-means, a Tabela 5 exibe o resultado da clusterização usando o Mapa Auto Organizável de Kohonen e a Tabela 6 mostra o resultado da clusterização Hierárquica. As clusterizações foram realizadas com a quantidade de 10 clusters estimada no Capítulo 5. A primeira coluna de cada tabela representa os clusters (grupos) de 1 a 10. As colunas nomeadas como MSC-01 a MSC-06 representam os elementos da topologia da rede apresentada na Figura 8 e

relacionam as quantidades de dados por MSC. O total dos dados por grupo é representado na última coluna e o total de dados por MSC está representado na última linha das tabelas.

Tabela 4 – Clusterização feita por K-means.

Grupos	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Total Dados
1				1	1	811	813
2					181		181
3					594		594
4	871		3				874
5		871					871
6			1325	1330			2655
7	450	452					902
8	4	4			8	12	28
9	1		5		520		526
10			5	8	9	503	525
Total Dados	1326	1327	1338	1339	1313	1326	7969

Fonte: O autor, 2016.

Tabela 5 – Clusterização feita pelo Mapa auto-organizável de Kohonen.

Grupos	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Total Dados
1	870						870
2		870					870
3	451	453					904
4	1		2		823	38	864
5					481		481
6					1	1276	1277
7	4	4			8	12	28
8			11	14			25
9			1323	20			1343
10			2	1305			1307
Total Dados	1326	1327	1338	1339	1313	1326	7969

Fonte: O autor, 2016.

Tabela 6 – Clusterização feita pelo método Hierárquico.

Grupos	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Total Dados
1		827					827
2	590	2					592
3				1317			1317
4			1322	10			1332
5					184		184
6					515		515
7	726	489					1215
8	5	5			599	9	618
9	4	4			9	8	25
10	1		16	12	6	1309	1344
Total Dados	1326	1327	1338	1339	1313	1326	7969

Fonte: O autor, 2016.

6.2 Identificação das Falhas

Para facilitar a identificação das falhas, as premissas descritas a seguir foram adotadas pelo especialista:

- Falhas podem ser percebidas pela redução de desempenho de uma MSC, por isso os valores médios de desempenho para cada grupo foram aferidos através do Código de Liberação *000-NORMAL CALL CLEARING*.
- Durante uma falha, Alterações significativas podem ser observadas nos Códigos de Liberação. Para avaliar essa alteração, foram calculados os valores médios de todos os Códigos de Liberação para todo o conjunto de dados.
- A quantidade de dados observada em cada grupo pode indicar problemas, pois é esperado que agrupamentos que representem falhas possuam menos dados que os demais.

Para registrar as observações do especialista, foram adicionadas duas novas colunas às Tabela 4, Tabela 5 e Tabela 6. Três novas tabelas foram geradas. A primeira coluna acrescentada à nova tabela diz respeito aos valores médios de desempenho para cada grupo, a segunda mostra os Códigos de Liberação que apresentaram variações importantes em relação aos seus valores médios.

A Tabela 7 mostra as observações feitas em relação à clusterização por K-means (Tabela 4).

Os grupos 1, 2 e 3 apresentam variações importantes nos códigos 010-ABSENT SUBSCRIBER, 205-RADIO IF CONGESTION, B13-RADIO IF FAILURE e B1E-RADIO RESOURCE UNAVAIL, indicando haver problemas na rede de acesso.

Os grupos 4 e 5 possuem o melhor valor de desempenho (52% e 50%). Não foram percebidas variações nos Códigos de Liberação representados por esses dois grupos.

O grupo 6 mostrou alterações dos códigos 006-B ANSWER TIME OUT e 30A-A ONHOOK DURING SET UP. Esses códigos sozinhos não representam falhas na rede, entretanto eles são indicativos na demora no completamento de uma chamada causado por problemas externos (i.e., latência dos meios de transmissão).

Não foi associada nenhuma falha ao grupo 7, embora existissem alterações no código 30B-A ONHOOK DUR_WAIT ANSWER. Esse código indica que o assinante chamador desistiu da conexão durante o tempo de espera pelo atendimento. Por isso, a alteração desse código foi associada pelo especialista à mudança de comportamento dos assinantes (demora no atendimento) devido ao horário de baixo tráfego registrado (24hs às 6h.).

O grupo 8 revelou a ocorrência de falhas graves na rede, evidenciadas pela pior indicação de desempenho (21%). Os códigos alterados para esse grupo mostram existir problemas na troca de mensagens de sinalização entre as MSCs e HLRs (códigos 812-MAP FAILURE e 81B-NO RESPONSE FROM HLR) e também à perda de comunicação entre as MSCs e IN (*Intelligent Network*). Problemas na conexão entre esses equipamentos fazem com que nenhuma chamada para um assinante móvel registrado naquela MSC seja completada.

Tabela 7 – Falhas identificadas por K-means

Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Códigos de Liberação
1				1	1	811	38%	206-CALL REJECTED 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE B1E-RADIO RESOURCE UNAVAIL
2					181		34%	010-ABSENT SUBSCRIBER 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE B1B-RADIO IF MSG FAILURE
3					594		39%	010-ABSENT SUBSCRIBER 012-NO PAGING RESPONSE 015-NORMAL UNSPECIFIED 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B1B-RADIO IF MSG FAILURE
4	871		3				52%	
5		871					50%	
6			1325	1330			30%	006-B ANSWER TIME OUT 30A-A ONHOOK DURING SET UP
7	450	452					48%	30B-A ONHOOK DUR_WAIT ANSWER
8	4	4			8	12	21%	812-MAP FAILURE 81B-NO RESPONSE FROM HLR B2C-IN PROTOCOL ERROR B2D-IN RESPONSE TIME OUT D03-SERV OR FUNC NOT IMPLEM
9	1		5		520		36%	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER 006-B ANSWER TIME OUT
10			5	8	9	503	37%	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING 010-ABSENT SUBSCRIBER

Fonte: O autor, 2016.

A Tabela 8 mostra as observações feitas em relação à clusterização obtida pelo Mapa auto-organizável.

Os grupos 1, 2 e 3 produziram resultados equivalentes aos grupos 4, 5 e 7, respectivamente, obtidos por K-means (Tabela 4). Foram percebidas para esses grupos apenas variações de desempenho.

Problemas na rede de acesso foram revelados pelos grupos 4, 5 e 6. O grupo 7 detectou as mesmas falhas de sinalização entre a MSC e os equipamentos HLR e SCP detectadas por k-means.

O grupo 8 identificou alterações no código 80F-CIRCUIT CONGESTION, referente a problemas de congestionamento de circuitos na rede. Congestionamentos impedem que chamadas sejam completadas, por isso o fraco desempenho nesse grupo (19%).

Os grupos 9 e 10 mostraram alterações dos códigos 006-B ANSWER TIME OUT e 30A-A ONHOOK DURING SET UP que podem indicar demora no tempo de completamento da chamada. Para o grupo 9, o código 010-ABSENT SUBSCRIBER também apresentou alteração. Essa informação significa que a rede perdeu a comunicação com o assinante e, por isso, no momento de uma conexão, não consegue contatá-lo. Esse problema ocorre quando os assinantes, ao se deslocarem, perdem a cobertura do sinal de rádio de forma “abrupta” (i.e., dentro de tuneis, regiões entre dois morros, etc.).

Tabela 8 – Falhas identificadas pelo Mapa auto-organizável de Kohonen

Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Códigos de Liberação
1	870						52%	
2		870					50%	
3	451	453					48%	30B-A ONHOOK DUR_WAIT ANSWER (15%)
4	1		2		823	38	37%	010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 016-CHA CAPA OF SUBS EXCEEDED 30B-A ONHOOK DUR_WAIT ANSWER(22%) 30C-B ONHOOK TIME OUT
5					481		37%	012-NO PAGING RESPONSE 015-NORMAL UNSPECIFIED 205-RADIO IF CONGESTION 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE B1B-RADIO IF MSG FAILURE
6					1	1276	38%	015-NORMAL UNSPECIFIED 206-CALL REJECTED 307-B-NUMBER UNUSED 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE B1E-RADIO RESOURCE UNAVAIL
7	4	4			8	12	21%	812-MAP FAILURE 81B-NO RESPONSE FROM HLR B2C-IN PROTOCOL ERROR B2D-IN RESPONSE TIME OUT D03-SERV OR FUNC NOT IMPLM
8			11	14			19%	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER 80F-CIRCUIT CONGESTION 817-N-UNALLOCATED NUMBER 822-RECOVERY ON TIMER EXPIRY D07-NO USER RESPONDING
9			1323	20			34%	006-B ANSWER TIME OUT 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 30B-A ONHOOK DUR_WAIT ANSWER(33%) D07-NO USER RESPONDING(5%)
10			2	1305			33%	006-B ANSWER TIME OUT (8%) 015-NORMAL UNSPECIFIED 30B-A ONHOOK DUR_WAIT ANSWER(44%) D07-NO USER RESPONDING(5%)

Fonte: O autor, 2016.

A Tabela 9 mostra as observações feitas a partir da clusterização hierárquica. Os grupos 1, 2 e 7 contém os dados com melhor desempenho e são equivalentes aos grupos 1, 2 e 3 obtidos através do Mapa Auto-organizável (Tabela 8). Os grupos 3 e 4 podem ser comparados, também, aos grupos 9 e 10 da clusterização feita pelo Mapa Auto-organizável, uma vez que esses grupos concentram a maior parte dos dados, gerados pelas MSC-03 e MSC-04. Os grupos 5,6, 8 e 10 identificaram os mesmos problemas na rede de acesso conforme percebidos por K-means (Tabela 7) e por Kohonen (Tabela 8). Falhas importantes relacionadas a sinalização já identificadas por K-means e Kohonen foram percebidas pelo grupo 9.

Tabela 9 – Falhas identificadas pela clusterização Hierárquica

Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Códigos de Liberação
1		827					50%	
2	590	2					52%	
3				1317			33%	006-B ANSWER TIME OUT 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
4			1322	10			35%	006-B ANSWER TIME OUT (8%) 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
5					184		34%	010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 016-CHA CAPA OF SUBS EXCEEDED 30B-A ONHOOK DUR_WAIT ANSWER 30C-B ONHOOK TIME OUT
6					515		40%	010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 016-CHA CAPA OF SUBS EXCEEDED 30B-A ONHOOK DUR_WAIT ANSWER 30C-B ONHOOK TIME OUT
7	726	489					50%	30B-A ONHOOK DUR_WAIT ANSWER
8	5	5			599	9	37%	010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 016-CHA CAPA OF SUBS EXCEEDED 30B-A ONHOOK DUR_WAIT ANSWER 30C-B ONHOOK TIME OUT
9	4	4			9	8	21%	812-MAP FAILURE 81B-NO RESPONSE FROM HLR B2C-IN PROTOCOL ERROR B2D-IN RESPONSE TIME OUT D03-SERV OR FUNC NOT IMPLM
10	1		16	12	6	1309	38%	015-NORMAL UNSPECIFIED 206-CALL REJECTED 307-B-NUMBER UNUSED 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE B1E-RADIO RESOURCE UNAVAIL

Fonte: O autor, 2016.

Após a identificação das falhas pelo especialista, os resultados dos métodos K-means, Mapas auto-organizáveis e Clusterização Hierárquica foram comparados. Primeiro com relação a distribuição dos dados em cada grupo e depois pela qualidade na identificação das falhas. Os métodos K-means e Kohonem conseguiram isolar 28 falhas graves de sinalização enquanto a clusterização feita pelo método Hierárquico identificou 25. Os dados indicando melhor desempenho foram distribuídos da forma equivalente, para os três métodos, em três grupos distintos: K-means grupos 4, 5 e 7, Mapas Auto Organizáveis grupos 1, 2 e 3 e cluster Hierárquico grupos 1, 2 e 7. Falhas relacionadas ao “congestionamento de circuitos” foram percebidas apenas pelo Mapa Auto Organizável (Tabela 8 – grupo 8). Embora os resultados obtidos pelos três métodos tenham apresentado resultados semelhantes, apenas o Mapa Auto Organizável conseguiu agrupar de forma isolada as falhas relacionadas ao congestionamento de circuitos. Com base na capacidade do Mapa auto-organizável em identificar esse tipo de falha, esse método foi utilizado na construção do modelo.

6.3 Refinamento método Kohonen

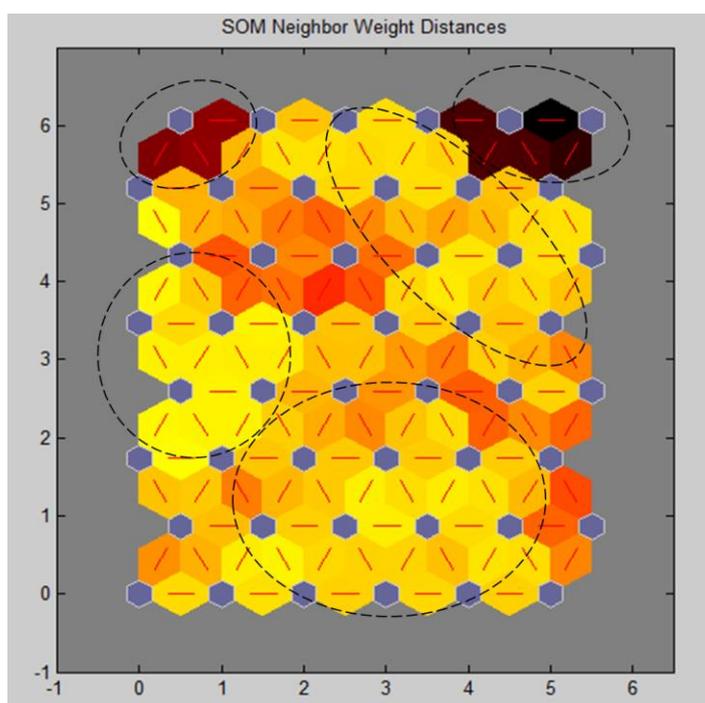
Com o objetivo refinar a identificação de padrões, a quantidade de cluster (neurônios) do Mapa Auto Organizável foi aumentada. Partindo de 10 neurônios do primeiro experimento, este número foi sendo aumentado de forma incremental até que o mapa, após o seu treinamento tivesse um neurônio que não foi excitado por nenhum padrão de entrada. Desta forma chegou-se ao número final de 48 neurônios. Este número produziu os melhores resultados, tendo em vista a separação de falhas em diferentes grupos e a distribuição dos dados em diferentes faixas de desempenho.

A Figura 35 mostra a distribuição dos neurônios (clusters) no do Mapa auto-organizável. Conforme discutido no Capítulo 5, as cores claras no mapa indicam que estes clusters estão mais próximos dos seus vizinhos, enquanto as mais escuras indicam que eles estão mais distantes uns dos outros. Conforme a Figura 35, podemos observar 3 grupos formados por vários neurônios (regiões claras) e 3 grupos formados por apenas um neurônio cada um (regiões escuras). Analisando os grupos formados pelo critério de desempenho de tráfego e das indicações de falha encontradas pelo especialista, tornou possível agrupar esse mapa em 5 regiões básica de distribuição (Figura 36). Os valores de desempenho para cada região correspondem aos valores mínimos e máximos, respectivamente, obtidos dos dados

que representam cada grupo no mapa. A lista abaixo representa as 5 regiões de desempenho e seus valores percentuais:

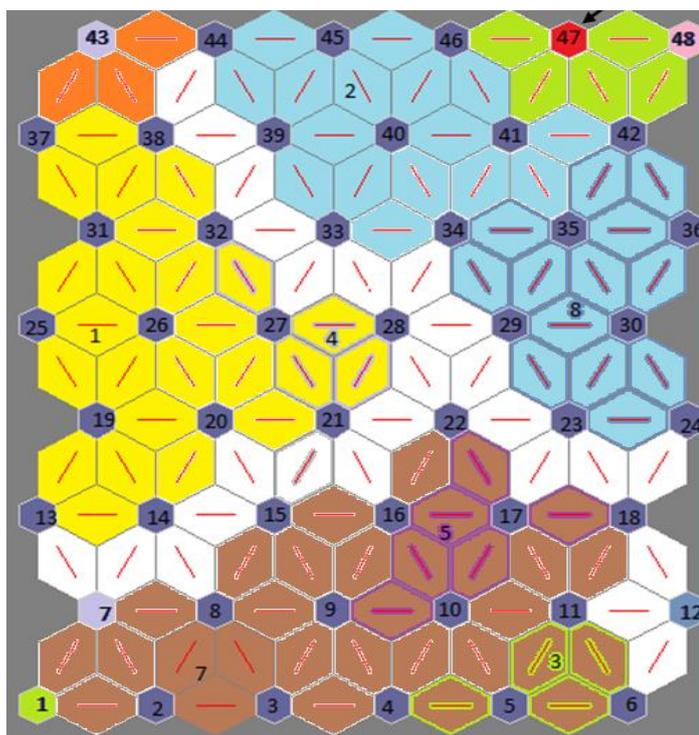
- Região 1 (Marrom): Desempenho de tráfego entre 42% e 53%;
- Região 2 (Azul): Desempenho de tráfego entre 33% e 40%;
- Região 3 (Amarelo): Desempenho de tráfego entre 23% e 38%;
- Região 4 (Laranja): Desempenho de tráfego entre 20% e 25%;
- Região 5 (Verde): Desempenho de tráfego entre 0% e 20%.

Figura 35 – Mapa das distâncias para 48 clusters obtidos por Kohonen



Fonte: O autor, 2016.

Figura 36 – Regiões de desempenho



Fonte: O autor, 2016.

As falhas identificadas foram mapeadas em 5 categorias de acordo com os Códigos de Liberação que apresentaram alterações em relação aos seus valores médios:

- Falhas de Software (SW);
- Falhas de Comutação (FC);
- Falhas de Sinalização (FS);
- Falhas de rede de Acesso (FA);
- Problemas de cobertura da rede de Acesso (CR).

As falhas na rede de acesso foram graduadas, ainda, em 4 níveis. Cada nível indica a quantidade de diferentes Código de Liberação relacionados à rede de acesso que apresentaram alteração de valor. A

Tabela 10 mostra a relação dos níveis das falhas de acesso e seus respectivos Códigos de Liberação.

A partir da identificação das 5 regiões, através do mapa auto organizável, e das falhas identificadas na análise dos grupos, a classificação dos dados para o modelo foi finalizada. O resultado da classificação é apresentado na Tabela 11, Tabela 12, Tabela 13 e Tabela 14. As duas primeiras colunas das tabelas contêm as informações de classificação. A primeira coluna representa as regiões de desempenho (1 até 5) e a segunda o tipo de falha (SW, SC,

FS, FA e CR). As colunas restantes representam os dados das MSC (MSC-01 à MSC-06), o desempenho médio (desempenho) e os Códigos de Liberação alterados.

Tabela 10 – Níveis de falha da rede de Acesso

Nível	Código de Liberação
1	B1E-RADIO RESOURCE UNAVAIL
2	B16-REMOTE EQUIPMENT FAILURE B1E-RADIO RESOURCE UNAVAIL
3	012-NO PAGING RESPONSE 205-RADIO IF CONGESTION B13-RADIO IF FAILURE
4	010-ABSENT SUBSCRIBER 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION B13-RADIO IF FAILURE

Fonte: O autor, 2016.

Tabela 11 – Classificação da região 1

Região	Falhas	Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Clear codes alterados
1	N/A	1		111					45,95	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER
1	N/A	2		321					45,95	006-B ANSWER TIME OUT 30A-A ONHOOK DURING SET UP 30B-A 30B-ONHOOK DUR_WAIT ANSWER
1	N/A	3		267					51,91	Sem alterações
1	N/A	4		166					50,30	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER
1	N/A	5	51	188					45,13	016-CHA CAPA OF SUBS EXCEEDED 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER
1	N/A	6	20	66					45,46	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER
1	N/A	7		27					42,24	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER
1	N/A	8		154					51,46	Sem alterações
1	N/A	9	144						52,80	Sem alterações
1	N/A	10	143	3					51,58	Sem alterações
1	N/A	11	165	12					51,07	Sem alterações
1	N/A	15	127						51,25	Sem alterações
1	N/A	16	219						53,20	Sem alterações
1	N/A	17	159						52,23	Sem alterações
1	N/A	18	135						50,77	Sem alterações
1	N/A	22	153	3					51,84	Sem alterações

Fonte: O autor, 2016.

Tabela 12 – Classificação da região 2

Região	Falhas	Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Clear codes alterados
2	FA-4	23					106		35,16	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
2	FA-4	24					73		33,03	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION (10,5%) 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE (1,6%)
2	FA-3	29					132		39,83	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
2	FA-4	30					75		37,79	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION (3%) 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE (1,4%)
2	FA-2	33					14	10	36,08	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER 006-B ANSWER TIME OUT 010-ABSENT SUBSCRIBER B13-RADIO IF FAILURE
2	FA-4	34					259		40,54	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
2	FA-3	35					147		37,41	005-B-SUBSCRIBER BUSY 006-B ANS010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER (22%) B13-RADIO IF FAILURE
2	FA-4	36	1				220		36,38	005-B-SUBSCRIBER BUSY (2%) 006-B ANS010-ABSENT SUBSCRIBER (5%) 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
2	FA-1	39					1	545	38,95	30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE (acima da media)
2	FC	40						310	37,24	016-CHA CAPA OF SUBS EXCEEDED 30A-A ONHOOK DURING SET UP
2	FC	41						150	36,77	016-CHA CAPA OF SUBS EXCEEDED 30A-A ONHOOK DURING SET UP
2	FA-4	42					154		35,33	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
2	FA-2	45					2	163	39,84	307-B-NUMBER UNUSED B16-REMOTE EQUIPMENT FAILURE B1E-RADIO RESOURCE UNAVAIL
2	FA-1	46						137	37,38	015-NORMAL UNSPECIFIED 30A-A ONHOOK DURING SET UP B1E-RADIO RESOURCE UNAVAIL

Fonte: O autor, 2016.

Tabela 13 – Classificação da região 3

Região	Falhas	Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Clear codes alterados
3	FA-1	12	5	5			15		36,45	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B1E-RADIO RESOURCE UNAVAIL
3	CR	13			328				34,78	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	14			387				37,35	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	19			224	1			34,67	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	20			135				34,57	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	21			6				31,76	016-CHA CAPA OF SUBS EXCEEDED 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	25				229			23,43	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	26			138				32,44	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	CR	27			100				33,91	30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING 006-B ANSWER TIME OUT
3	FA-3	28					107		38,44	005-B-SUBSCRIBER BUSY 010-ABSENT SUBSCRIBER 015-NORMAL UNSPECIFIED 012-NO PAGING RESPONSE 016-CHA CAPA OF SUBS EXCEEDED 205-RADIO IF CONGESTION 30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER B13-RADIO IF FAILURE
3	CR	31				627			24,52	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	FC	32			6	8			24,27	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER 80F-CIRCUIT CONGESTION 817-N-UNALLOCATED NUMBER 822-RECOVERY ON TIMER EXPIRY D07-NO USER RESPONDING
3	CR	37				461			25,18	006-B ANSWER TIME OUT 30B-A ONHOOK DUR_WAIT ANSWER D07-NO USER RESPONDING
3	FC	38			14	13			27,20	30A-A ONHOOK DURING SET UP 30B-A ONHOOK DUR_WAIT ANSWER 80F-CIRCUIT CONGESTION 817-N-UNALLOCATED NUMBER 822-RECOVERY ON TIMER EXPIRY D07-NO USER RESPONDING

Fonte: O autor, 2016.

Tabela 14 – Classificação das regiões 4 e 5

Região	Falhas	Grupo	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06	Desempenho	Clear codes alterados
4	FS	43	4	4				1	25,90	812-MAP FAILURE 81B-NO RESPONSE FROM HLR
4	FS	44					6		25,90	B2C-IN PROTOCOL ERROR B2D-IN RESPONSE TIME OUT
5	SW	47					2	2	15,95	30A-A ONHOOK DURING SET UP D03-SERV OR FUNC NOT IMPLEM
5	FS	48						8	15,79	B2C-IN PROTOCOL ERROR

Fonte: O autor, 2016.

6.4 Visão Geral do Modelo

A Figura 37 mostra uma visão geral de todas as etapas que constituem o modelo, iniciando pela coleta dos dados (Códigos de Liberação), pré-processamento dessas informações e seleção dos 99 atributos de entrada do modelo descritos no Capítulo 2.

Figura 37 – Modelo de classificação de falhas.



Fonte: O autor, 2016.

Na etapa de Clusterização os métodos apresentados no Capítulo 3 são testados e avaliados com base nos critérios de desempenho descritos no Capítulo 5, de acordo com os resultados obtidos nos testes. A capacidade do método em identificar as falhas presentes no conjunto de dados é avaliada pelo especialista de rede, na etapa da classificação. Esse profissional é responsável por correlacionar (mapear) cada cluster gerado com um comportamento de rede específico. Esse mapeamento, é feito uma única vez para todos os

clusters formados pelo método de clusterização escolhido. Como resultado dessa etapa, é obtida uma relação conforme exemplificado na Tabela 15, onde os clusters são rotulados de 1 à N. O Especialista associa para cada rótulo um padrão de rede (Classificação) que corresponde a um comportamento da rede ou uma falha. Desse modo, um dado apresentado à entrada do modelo será rotulado com o número do seu respectivo cluster e classificado de acordo com a tabela de mapeamento preenchida pelo Especialista. Esse processo acontece em tempo real para todos os dados submetidos ao modelo.

Tabela 15 – Exemplo de mapeamento de clusters pelo Especialista

Sáida da Clusreização	Especialista	
Rótulo do Cluster	Classificação	Análise dos resultados
1	Padrão-1	Comportamento A
2	Padrão-2	Comportamento B
3	Padrão-3	Comportamento C
4	Padrão-4	Comportamento D
5	Padrão-5	Comportamento E
6	Padrão-6	Falha - 1
7	Padrão-7	Comportamento F
8	Padrão-8	Falha - 2
N	Padrão-N	Comportamento "N"

Fonte: O autor, 2016.

Para avaliar a eficiência do modelo na identificação de falhas, o resultado final na saída, para todo o conjunto de teste, foi analisado pelo especialista e suas observações apresentadas na próxima seção.

7. RESULTADOS

Nesse capítulo são apresentados os resultados da classificação feita pelo modelo para a mesma massa de dados utilizada nos testes. As falhas identificadas na saída do modelo são analisadas pelo especialista que valida sua ocorrência através da análise dos atributos (Códigos de Liberação) aplicados a entrada do modelo que resultou na indicação de falha na saída. A classificação de todos os dados na entrada do modelo produz um arquivo de saída que correlaciona cada entrada com a sua respectiva classe. A partir dessa informação, os resultados são organizados na Tabela 16 e nos gráficos mostrados na Figura 38, Figura 39 e Figura 40, que permitem visualizar a distribuição global dos dados, de diferentes formas, de acordo com a classificação obtida e MSC de origem dos dados.

A Tabela 16 mostra a distribuição dos dados, entre as 5 regiões, feita pelo modelo. A primeira coluna contém a identificação de cada região. Os dados relacionados a cada MSC são apresentados nas colunas seguintes. A maior parte dos dados (99%) foram classificados nas regiões 1, 2 e 3. Apenas 1% deles, foram classificados nas regiões 4 e 5. O desempenho médio percentual, considerando os dados classificados por regiões, é apresentado no gráfico da Figura 38. O maior desempenho médio foi verificado para a região 1 (50%), seguido pelas regiões 2 (38%), região 3 (30%) e 4 (20%). O pior desempenho foi contabilizado para a região 5 (15%).

Comparando a Tabela 16 com o gráfico da Figura 38, pode-se afirmar que as MSC-01 e MSC-02 possuem o melhor desempenho médio percentual, pois a maioria dos seus dados estão classificados pela região 1. O segundo melhor desempenho é atribuído a MSC-05 e MSC-06 e o pior desempenho para a MSC-03 e MSC-04. Essa informação representa uma vantagem apresentada pelo modelo, uma vez que permite classificar os equipamentos de acordo com o seu desempenho global.

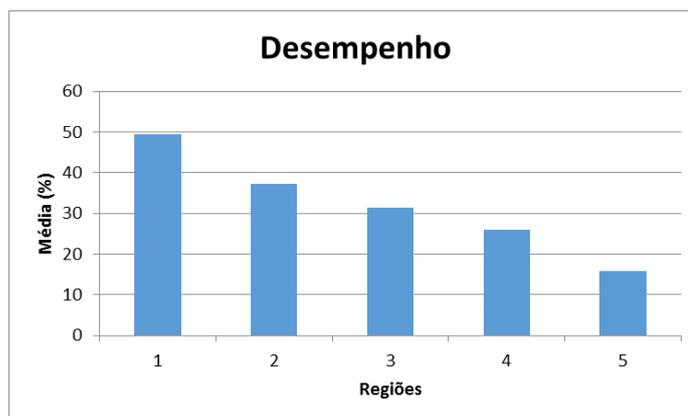
Tabela 16 – Classificação dos dados por Região

Regiões	MSC-01	MSC-02	MSC-03	MSC-04	MSC-05	MSC-06
1	1316	1318				
2	1				1183	1315
3	5	5	1338	1339	122	
4	4	4			6	1
5					2	10

Fonte: O autor, 2016.

É importante ressaltar que, de acordo com o conhecimento do especialista, as informações relacionadas ao desempenho da rede, apresentadas na Tabela 16, e a classificação dos equipamentos em relação ao desempenho global do sistema, correspondem ao comportamento desses equipamentos quando observados no dia-a-dia operacional.

Figura 38 – Desempenho médio por região do Mapa auto-organizável.



Fonte: O autor, 2016.

O gráfico da Figura 39 mostra concentração de falhas por região, em números absolutos (dados). Cada tipo de falha (CR, FA-1 à FA-4, FS e SW) está representado de acordo com o código de cores da legenda. A análise do gráfico mostra que a região de desempenho número 2 apresenta muitas falhas de acesso (FA-1, FA-2, FA-3 e FA-4) e também problemas de cobertura (CR). A região 3 apresenta 2677 indicações de problemas de cobertura (CR) e 132 falhas de acesso (FA-1 e FA-3). Falhas de sinalização foram detectadas na região 4 (FS=15) e região 5 (SW=4).

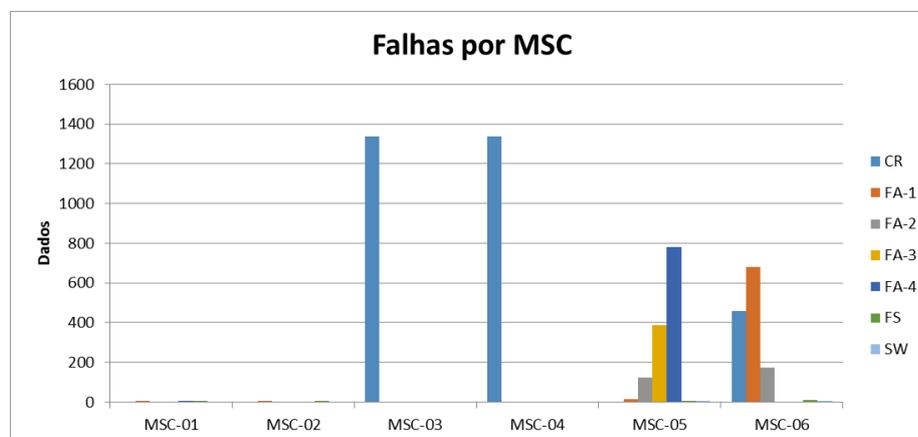
Figura 39 – Classificação dos dados por falha em cada região.



Fonte: O autor, 2016.

O gráfico da Figura 40 mostra a distribuição das mesmas falhas representadas na Figura 39. Agora, as falhas presentes no conjunto de dados existente são correlacionadas na saída do modelo a cada uma das MSCs onde as respectivas falhas foram detectadas. Essa visão permite avaliar o ponto da rede onde o especialista deve atuar para solucionar os problemas. A quantidade de falhas ou problemas detectados estão representados em número absolutos no eixo “dados”. E a cor correspondente a cada uma das falhas pode ser identificada na legenda, ao lado direito do gráfico. Os problemas de cobertura de rede (CR) estão concentrados na MSC-03 e MSC-04. A MSC-05 e MSC-06 apresentam falhas de acesso de todos os níveis (FA-1, FA-2, FA-3 e FA-4). A MSC-01e MSC-02 apresentou quantidades mínimas de problemas de cobertura e falhas de acesso, podendo ser desprezados na análise por serem pouco representativos.

Figura 40 – Distribuição das falhas por MSC.



Fonte: O autor, 2016.

7.1 Validação do modelo

Os resultados obtidos na saída do modelo para o mesmo conjunto de dados usados no treinamento do Mapa auto-organizável, mostraram a capacidade do modelo em classificar os dados em diferentes padrões de rede assim como identificar as falhas existentes nesses dados. Para aferir a capacidade de generalização do modelo, isso é, sua capacidade de identificar padrões e falhas para qualquer conjunto de dados, considerando os mesmos atributos usados em seu desenvolvimento, um novo conjunto de dados foi apresentado a entrada do modelo. O conjunto de dados usados para validação é composto de 8.478 registros coletados a partir de 8 MSCs diferentes no período de dois meses consecutivos. Os dados foram submetidos ao modelo e os resultados consolidados em duas tabelas. A Tabela 17 apresenta a relação entre as falhas importantes (Falhas de Software e Falhas de Sinalização) detectadas na saída do modelo e os demais registros que caracterizam a operação normal do sistema. O modelo detectou 59 falhas importantes, das quais apenas uma delas foi avaliada pelo especialista como falso positivo, isso é, não significa realmente uma falha. Todos os demais registros foram avaliados como isentos de falhas importantes.

Tabela 17 – Falhas na saída do modelo

	VERDADEIRO	FALSO	TOTAL
FALHAS DETECTADAS	58 (1%)	1 (0,1%)	59
OPERAÇÃO NORMAL	8419 (99%)	0	8419
Total	8477	1	8478

Fonte: O autor, 2016.

A Tabela 18 mostra o número de registros que foram alocados corretamente às suas regiões de desempenho. Verificou-se que 7% dos dados não foram alocados de forma correta a sua região de desempenho. A análise do Mapa auto-organizável mostrou que os dados com erro de classificação pertencem a agrupamentos localizados na fronteira entre as regiões 1, 2 e 3. Na avaliação do Especialista do sistema, o valor de 7% de “erro” é considerado aceitável para o modelo, pois as medidas de desempenho representam uma característica adicional do projeto e não influenciam na detecção de falhas.

Tabela 18 – Distribuição dos dados nas regiões

	VERDADEIRO	FALSO	Total
DISTRIBUIÇÃO NOS GRUPOS	7905 (93%)	573 (7%)	8478

Fonte: O autor, 2016.

7.2 Comentário Final

Os resultados apresentados nesse capítulo foram positivos e mostraram que o modelo proposto é capaz de identificar e classificar as falhas de rede em 5 categorias diferentes. Adicionalmente, a distribuição dos dados em 5 regiões, permite avaliar cada um dos equipamentos com relação ao seu desempenho de tráfego. Assim, as informações na saída do modelo, permitem ao operador obter um panorama geral da rede em função do desempenho de cada MSC e das falhas detectadas na rede atingindo o objetivo para o qual o modelo foi proposto.

8 CONCLUSÃO

Indicadores de qualidade da rede permitem a ANATEL aferir a qualidade do serviço prestado pelas operadoras de telecomunicações aos usuários da tecnologia GSM no Brasil. Para manter os indicadores de qualidade dentro dos padrões estabelecidos pela Anatel e garantir a qualidade na prestação dos serviços, cabe às operadoras monitorar suas redes, detectar as falhas e corrigi-las no menor tempo possível. A velocidade em detectar falhas e solucionar os problemas na rede depende de duas variáveis importantes: a primeira diz respeito a experiência dos especialistas responsáveis pela operação da rede, a segunda está relacionada às ferramentas disponíveis para a detecção e classificação de falhas.

A formação de especialistas de rede requer treinamentos específicos, além de tempo necessário para adquirir experiência operacional. Ferramentas que permitam analisar falhas necessitam de grandes investimentos financeiros em aquisição de novos equipamentos, infraestrutura (i.e., energia, climatização, etc.) para integração desses equipamentos e treinamento para sua utilização. É importante ressaltar, que as ferramentas atualmente usadas pelas operadoras na detecção de falhas são apenas complementares, fazendo com que o trabalho do especialista na identificação de uma falha seja mais complexo.

Esse cenário cria uma oportunidade para a proposta feita nesse trabalho, que é produzir um modelo para detecção e classificação de falhas na rede GSM, através das informações já disponíveis na própria rede, que permita reduzir o tempo gasto pelos especialistas para detectar uma falha.

Como primeiro passo neste trabalho, foi realizado um estudo preliminar para mapear as informações mais relevantes utilizadas pelas equipes de manutenção no processo de geração de indicadores de qualidade e identificação de falhas. Das informações geradas pelos equipamentos foram analisados os alarmes autônomos enviados ao centro de gerência da rede, CDRs, mensagens de sinalização e contadores. O uso dos alarmes autônomos foi descartado, pois os registros desses alarmes não continham informações suficientes para analisar falhas. O estudo sobre CDRs mostrou a complexidade na etapa de pré-processamento para correlacionar os diferentes campos encontrados nessa fonte de informação. Outro fato relevante, é que somente os CDRs de chamadas completadas são gerados pelos equipamentos para otimizar os recursos de processamento e armazenamento desse tipo de informação nos

servidores de mediação. As mensagens de sinalização, usadas para a comunicação entre os elementos da rede, também necessitam maior esforço computacional na etapa de pré-processamento para extração das informações relacionadas às conexões. Além disso, conexões entre dois assinantes registrados na mesma MSC produzem mensagens de sinalização que não são enviadas à rede e, por isso, não podem ser capturadas, em consequência disso grande parte dos eventos de rede não podem ser observados. Os Códigos de Liberação, contadores de eventos gerados pelas MSC, permitem obter uma visão global do comportamento da rede quando analisados em conjunto. O fato de serem gerados em um formato que simplifica a etapa de pré-processamento, permite aos Códigos de Liberação usar pouco recurso computacional. Em função do conteúdo presente nos Códigos de Liberação e de suas características, optou-se por utilizar apenas essa informação como fonte de dados a ser usada pelo modelo.

A partir desta constatação, o comportamento dos Códigos de Liberação gerados por diferentes MSC, em dias e horários diferentes, foi analisado. A análise mostrou padrões comuns de comportamento entre esses equipamentos. Uma seleção inicial de dados foi utilizada, com o objetivo de identificar e agrupar os padrões existentes. Métodos de clusterização foram estudados e avaliados em sua eficiência para formar grupamentos a partir da massa de dados.

Os métodos que obtiveram a melhor eficiência na formação de grupos foram K-means e Mapas auto-organizáveis de Kohonen. Os resultados mostraram que, os Mapas auto-organizáveis de Kohonen permitiram separar os padrões classificados como falhas, pelos especialistas, em grupos distintos quando comparado ao método K-means. Por isso, o Mapa auto-organizável de Kohonen foi adotado nesse trabalho para compor o modelo proposto.

No estágio final de validação, os padrões identificados pelo modelo, foram analisados e classificados pelo especialista do sistema. Os resultados foram satisfatórios na medida que, não só padrões de falhas foram identificados, mas também padrões de comportamento representando a operação normal do sistema e perdas de desempenho foram percebidos.

Os resultados revelaram que, além de detectar falhas, o modelo também é capaz de identificar diferentes padrões de desempenho do sistema, pois a maior parte dos dados do conjunto de testes foi agrupada em diferentes regiões do Mapa auto-organizável de acordo com faixas de desempenho identificadas pelo especialista de rede.

Falhas detectadas em equipamentos diferentes, que apresentavam as mesmas características, foram classificadas da mesma forma pelo modelo, ratificando que os

equipamentos, mesmo localizados em pontos diferentes da rede, se comportam da mesma forma na ocorrência de falhas que tenham a mesma natureza.

O modelo comprovou sua eficiência no apoio ao diagnóstico de falhas tornando mais simples e objetiva a identificação de uma falha. Com isso, espera-se que os profissionais de operação e manutenção do sistema dediquem a maior parte do seu tempo na solução das falhas.

Como proposta para futuros trabalhos, sugere-se a criação de um mecanismo para detecção de falhas baseado no modelo desenvolvido nesse trabalho. Esse mecanismo deve permitir ser integrado aos elementos da rede GSM. Essa integração, prevê a coleta periódica dos Códigos de Liberação das MSC. Essas informações seriam pré-processadas, armazenadas em tabelas e enviadas ao modelo para processamento. O resultado obtido é a classificação dos dados em tipos de falha, comportamento normal da rede, ou perda de desempenho. Para as falhas identificadas, o modelo deve indicar sua provável causa. O modelo desenvolvido pode ser utilizado de duas formas: modo *on-line* ou modo *off-line*. O modo *on-line* deve permitir seu uso de forma iterativa e em tempo real, nesse caso o modelo responde automaticamente quando é inserido um registro em sua entrada. No modo *off-line*, um arquivo contendo todas as entradas é submetido ao modelo. Nos dois modos de operação, é obtido um arquivo de saída onde os registros são separados em classes e agrupados de acordo com as suas características permitindo associá-los a comportamentos previamente definidos, facilitando a identificação de falhas e também permitindo observar a variação do desempenho dos equipamentos.

REFERÊNCIAS

- ALENCAR, M. (2004). “Telefonia Celular Digital”. 1. ed., Érica São Paulo, 2004.
- ANATEL (2011). “Resolução nº 575, de 28 de outubro de 2011”, Regulamento de Gestão da Qualidade da Prestação do Serviço Móvel Pessoal – RGQ-SMP. Diário Oficial da União. 31 out. 2011.
- BERNARD, D. (2013). “Clustering Indices”. University Paris Ouest Lab Modal'X, 2013. Disponível em: <<https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf>>. Acesso em: 23 fev. 2016.
- BERKHIN, P. (2002) “Survey Of Clustering Data Mining Techniques”. Disponível em: <Survey, <http://www.cc.gatech.edu/~isbell/reading/papers/berkhin02survey.pdf>>. [Acesso em: 17 fev. 2016.
- BOTTARI, G. (2014). “Monitoramento Térmico Responsivo para Centros de Processamento de Dados”. Dissertação de Mestrado, Programa de Pós-Graduação em Computação, Universidade Federal Fluminense. 2014.
- DUNN, J. (1974). “Well separated clusters and optimal fuzzy partitions”. Journal of Cybernetics, Vol. 4, No. 1, pp. 95-104, Taylor & Francis, 1974. doi: 10.1080/01969727408546059.
- EVERITT, B. (1974). “Cluster Analysis”, Heinemann Educational Books, Social Science Research Council, p. 122. Londres, 1974.
- FARIA, G. (2010) “Análise de Padrões em Chamadas Telefônicas”. Relatório Técnico, Programa de Pós-Graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP. 2010. Disponível em: <<http://urlib.net/8JMKD3MGP7W/38FUMP2>>. Acesso em: 13 fev. 2016.
- FRANCO, R. C. (2002). “Categorização Nebulosa e Validação de Categorias Aplicadas a Problemas de Reconhecimento de Padrões”. Dissertação de Mestrado, IME/NCE, Universidade Federal do Rio de Janeiro, 2002.
- FUJIMOTO, R. Y (2005). “Diagnostico Automático de Defeitos em Rolamentos Baseado em Lógica Fuzzy”. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, 2005.
- HONKELA, T.(1997) “Self organizing maps in natural language processing”. Doctor’s thesis. Helsinki University of Technology. Helsinki, Finland, out. 1997.

ITU-T (2001). “Physical/electrical characteristics of hierarchical digital interfaces”. Recommendations, SERIES G: Transmission Systems and Media, Digital Systems and Networks. Disponível em: <<https://www.itu.int/rec/T-REC-G.703/en>>. Acesso em: 16 fev. 2016.

KAWAMURA, C., GOEDEL, A., VALLIM, M.(2011). “Redes neurais artificiais na monitoração prognóstica de índices de vibração em conjuntos moto-redutores”. Anais do X SBAI - Simpósio Brasileiro de Automação Inteligente, pp. 1007-1012, São João Del Rei, MG.

KUMAR, V., TAN, P., STEINBACH, M. (2006). “Introduction to Data Mining”. Person, Cloth, ISBN-10: 0321321367, p.769, 2006.

MANNING, C.; RAGHAVAN, P; SCHÜTZE, H. (2009). “An Introduction to Information Retrieval”. Disponível em: <<http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>> Acesso em: 18 fev. 2016.

OLIVEIRA, E. (2008). “Modelo TMN: Aplicação ao Gerenciamento de Redes de Telecom.”. Disponível em: <<https://www.teleco.com.br/pdfs/tutorialmodelotmn.pdf>>. Acesso em: 22 fev. 2016.

ROCHA, A. (2005). “Redes Neurais Artificiais Aplicadas Na Detecção, Classificação e Localização De Defeitos Em Linhas De Transmissão”, Dissertação de Mestrado, Programa de pós-graduação em engenharia elétrica, da Universidade federal de Juiz de Fora, 2005.

ROUSSEUW, P.J.(1987). “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. Journal of Computational and Applied Mathematics, Vol. 20, pp. 53–65, 1987, doi 10.1016/0377-0427(87)90125-7.

RUSSELL, T. (1994). “Signaling System #7”. Mc Graw Hill, 4th Edition, 1994.

SEN, WU, NAIDONG KANG, LIU YANG (2007). “Fraudulent Behavior Forecast in Telecom Industry” - Communications of the IIMA, V.7, No 4, dez 2007.

SILVA, I., SPATTI, D., FLAUZINO,R. (2010) “Redes Neurais Artificiais para Engenharia e Ciências Aplicadas”. Artliber, 2010.

TAKEHISA, Y., YOSHIKIYO K., KOICHI, H.(2001). “Fault Detection by Mining Association Rules from House-keeping Data”. 6th International Symposium on Artificial Intelligence and Robotics & Automation in Space, i-SAIRAS 2001, Canadian Space Agency, St-Hubert, Quebec, Canada, Jun 2001, pp18-22.

YANCHI, L., ZHONGMOU, L., XIONG, H. (2010). “Understanding of Internal Clustering Validation Measures”. IEEE International Conference on Data Mining, 2010, pp. 911-916, doi. 10.1109/ICDM.2010.35.

YIAKOPOULOS, C.T., GRYLLIAS, K.C., ANTONIADIS, I.A. (2011). “Rolling element bearing fault detection in industrial environments based on a K-means clustering approach”. *Expert Systems with Applications*, Vol. 38, No. 3, pp. 2888–2911. Elsevier, mar. 2011. doi:10.1016/j.eswa.2010.08.083.

3GPP TS 23.221 (2004). “Architectural requirements”. 3rd Generation Partnership Project - Technical Specification Group Service and System Aspects. Disponível em: < <http://www.3gpp.org/>>. Acesso em: 14 fev. 2016.

3GPP (2010). “TS 32.298 - Charging Data Record (CDR) parameter description”. 3rd Generation Partnership Project - Technical Specification Group Service and System Aspects. Disponível em: < <http://www.3gpp.org/>>. Acesso em: 25 fev. 2016.

3GPP (2015). “TS 23.002 - Network architecture”, 3rd Generation Partnership Project - Technical Specification Group Services and System Aspects. Disponível em: < <http://www.3gpp.org/>>. Acesso em: 17 fev. 2016.