



**UNIFEI
UNIVERSIDADE FEDERAL DE ITAJUBÁ**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
PRÓ-DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO**

**ESTUDO DE UM ALGORITMO DE MINERAÇÃO DE DADOS APLICADO À
AVALIAÇÃO DE CURVAS DE CONSUMO DE ENERGIA ELÉTRICA**

ÁREA DE AUTOMAÇÃO E SISTEMAS ELÉTRICOS INDUSTRIAIS

Efrem Ferreira

**ITAJUBÁ
ESTADO DE MINAS GERAIS – BRASIL
2005**

**ESTUDO DE UM ALGORITMO DE MINERAÇÃO DE DADOS APLICADO À
AVALIAÇÃO DE CURVAS DE CONSUMO DE ENERGIA ELÉTRICA**

**ORIENTADOR: PROF. GERMANO LAMBERT TORRES
COORIENTADOR: PROF. LUIZ EDUARDO BORGES DA SILVA**

**DISSERTAÇÃO APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA DA UNIFEI COMO PARTE DOS REQUISITOS PARA A
OBTENÇÃO DO TÍTULO DE MESTRE EM ENGENHARIA ELÉTRICA:
AUTOMAÇÃO E SISTEMAS ELÉTRICOS INDUSTRIAIS.**

ITAJUBÁ
ESTADO DE MINAS GERAIS – BRASIL
2005

À minha esposa pelo amor e
compreensão, família e a Deus por me
iluminar nos momentos difíceis.

Ao Professor Germano Lambert Torres;
E a todas as pessoas que indiretamente
contribuíram para a conclusão desse
trabalho.

RESUMO

No dias atuais, a quantidade de informações inerentes ao nosso meio é enorme. Para tratar estas informações surgiu uma área conhecida como mineração de dados – “Data Mining”. Dentro desta área, existem diversos algoritmos para realizar o processo de descoberta do conhecimento em banco de dados. Pode-se citar como exemplos de algoritmos: Redes Neurais Artificiais, Lógica Fuzzy, Algoritmos Genéticos, Algoritmos Bayesianos, Árvores de Indução entre outros.

Neste trabalho foram estudados tópicos relacionados com Mineração de Dados, com a finalidade de aplicá-los na solução de problemas relacionados com a Engenharia Elétrica. Escolheu-se o algoritmo “Naive Bayes”, com objetivo de reconhecer padrões de consumo de energia elétrica, assim buscando uma alternativa às Redes Neurais Artificiais – RNA que são lembradas quando se trata de reconhecimento de padrões. Para facilitar o entendimento do algoritmo escolhido, alguns conceitos de estatística foram analisados, e posteriormente são apresentados os conceitos do algoritmo “Naive Bayes”.

Com a finalidade de elucidar uma aplicação do “Naive Bayes” foi realizado um estudo de caso, onde o algoritmo foi modificado com a finalidade de reconhecer curvas de carga ou curvas de consumo de consumidores residenciais, comerciais e industriais presentes no Setor Elétrico de Potência.

Para viabilizar a implementação desse estudo de caso, foi utilizado o pacote computacional, chamado Weka, o qual possui vários algoritmos de Mineração de Dados implementados em Java.

ABSTRACT

Nowadays, the quantity of inherent information to our environment is huge. For treat these information appeared an area called Data Mining. Inside of this area, there are some algorithms to carry out the knowledge discovery in Data Bases. Some examples of algorithms can be mentioned, such as: Neural Networks, Fuzzy Logic, Genetic Algorithm, Bayesian Algorithm, and Induction Tree.

In this work topics related with Data Mining were studied, with the purpose of apply in problems related with Electrical Engineering. The Naive Bayes algorithm was chosen with objective of recognize electrical energy consumption patterns, as an alternative to Artificial Neural Networks that are remembered in pattern recognition problems. Initially, some concepts of statistics were analyzed, and subsequently the concepts of the algorithm Naive Bayes are presented.

The original algorithm was modified with the purpose of recognize load curves of industrial, commercial, residential consumers in Electrical Power Systems. A computational package, called Weka, has been used. This package implemented was implemented in Java and has several data mining algorithms.

SUMÁRIO

LISTA DE FIGURAS	ix
LISTA DE TABELAS.....	xi
LISTA DE SÍMBOLOS E SIGLAS	xiii
CAPÍTULO 1.....	1
1 INTRODUÇÃO.....	1
1.1 Objetivos do trabalho	2
1.2 Organização do texto	3
CAPÍTULO 2.....	5
2 MINERAÇÃO DE DADOS.....	5
2.1 Definição	6
2.2 O processo de mineração de dados.....	6
2.2.1 Identificação do Problema	8
2.2.2 Pré – Processamento	8
2.2.3 Extração de Padrões.....	11
2.2.4 Pós - Processamento	13
2.3 Campo de aplicação.....	15
2.3.1 Decisão envolvendo julgamento.....	15
2.3.2 Predição de carga	16
CAPÍTULO 3.....	17
3 NOÇÕES BÁSICAS DE ESTATÍSTICA	17
3.1 Definição	17
3.2 Conjuntos.....	18
3.3 Frequência relativa.....	18
3.4 Noções fundamentais de probabilidade.....	19
3.5 Probabilidade condicional e independência.....	21
3.6 Inferência	25
3.6.1 Método geral	26
3.6.2 Independência Condicional	26
3.7 Amostragem e distribuição amostral	28
3.7.1 Conceitos Básicos.....	28
3.7.2 Amostragem aleatória	33
3.7.3 Planejamento amostral.....	34
3.8 Teorema de Bayes.....	36
3.9 Variável aleatória	41
3.10 Distribuição normal.....	42

3.10.1	Definição	42
3.10.2	Gráfico.....	42
CAPÍTULO 4.....		48
4	CLASSIFICADORES BAYESIANOS.....	48
4.1	Teorema de Bayes.....	50
4.2	Classificador Naive Bayes.....	51
4.3	Redes de Crenças Bayesiana.....	61
CAPÍTULO 5.....		65
5	PACOTE COMPUTACIONAL WEKA.....	65
5.1	Descrição	65
5.2	Bibliotecas Weka.....	69
5.2.1	Classes, instâncias e pacotes	70
5.2.2	O pacote weka.core.....	71
5.2.3	O pacote weka.classifiers.....	72
CAPÍTULO 6.....		73
6	ESTUDO DE CASO	73
6.1	Simulações	74
6.1.1	Testes utilizando padrões simples.....	79
6.1.2	Testes utilizando composição de padrões simples	82
6.1.3	Testes utilizando composição de padrões simples com oscilações.....	99
CAPÍTULO 7.....		111
7	CONCLUSÃO	111
8	REFERÊNCIAS BIBLIOGRÁFICAS	113
9	ANEXO	116

LISTA DE FIGURAS

Figura 2.1 - Processo de Mineração de Dados	7
Figura 2.2 - Tarefas da extração de padrões.....	12
Figura 3.1 - Diagrama para extração de duas camisas de uma caixa, sem reposição.	23
Figura 3.2 - Diagrama para extração de duas camisas de uma caixa, com reposição.	24
Figura 3.3 - Árvore contendo resultados do exemplo acima.	40
Figura 3.4 – Função de distribuição de probabilidade de uma VA – com média μ desvio padrão σ	42
Figura 3.5 - Função de Distribuição de Probabilidade de uma va – $Z \sim N(0,1)$	43
Figura 3.6 - - Representação gráfica de $F(y)$ como área.	44
Figura 3.7 - Função para $P(a \leq X \leq b)$ para uma V.A normal	45
Figura 3.8 - $P(0 \leq Z \leq Z_c)$ fornecida pelo Anexo 1.	46
Figura 4.1 - Construção do modelo de aprendizagem.	49
Figura 4.2 - Classificação de uma nova Instância.	49
Figura 4.3 - Rede Bayesiana simples.	62
Figura 5.1 - Formato de um arquivo ARFF.	66
Figura 5.2 - Dados do arquivo “Análise crédito” em uma planilha do Excel.....	67
Figura 5.3 - Dados do arquivo “análise de crédito” separados por vírgula.....	68
Figura 5.4 - Arquivo “Análise Credito” no formato ARFF.	69
Figura 5.5 - Índice com os pacotes disponíveis no sistema Weka.....	71
Figura 5.6 - Índice das interfaces e classes pertencentes ao pacote core	71
Figura 6.1 - Classes de consumidores de energia elétrica.	74
Figura 6.2- Curva de carga representando os setores residencial, comercial e industrial.....	75
Figura 6.3- Padrões de treinamento representando o setor residencial.....	76
Figura 6.4- Padrões de treinamento representando o setor comercial.....	77
Figura 6.5- Padrões de treinamento representando o setor industrial.	77
Figura 6.6- Exemplo de uma matriz de confusão.	78
Figura 6.7 - Padrão de treinamento e teste simples para consumidor residencial.	79
Figura 6.8- Padrão de treinamento e teste simples para consumidor comercial.....	80
Figura 6.9- Padrão de treinamento e teste simples para consumidor industrial.....	81
Figura 6.10- Padrões de treinamento representando setor residencial, comercial e industrial.....	82
Figura 6.11 - Padrões de treinamento e padrão de teste R20C80, R40C60, R60C40, R80C20.....	83

Figura 6.12 - Padrões de treinamento e padrão de teste R20I80, R40I60, R60I40, R80I20..... 86

Figura 6.13 - Padrões de treinamento e padrão de teste C20I80, C40I60, C60I40, C80I20..... 87

Figura 6.14- Padrões de treinamento e padrão de teste R10C90, R30C70, R50C50, R70C30, R10C90..... 89

Figura 6.15 - Padrões de treinamento e padrão de teste R10I90, R30I70, R50I50, R70I30, R10I90..... 97

Figura 6.16 - Padrões de treinamento e padrão de teste C10I90, C30I70, C50I50, C70I30, C10I90..... 98

Figura 6.17 - Padrões de treinamento e padrão de teste R10C90, R30C70, R50C50, R70C30, R10C90 com oscilações de 10%. 100

Figura 6.18 - Padrões de treinamento e padrão de teste R10I90, R30I70, R50I50, R70I30, R10I90 com oscilações de 10%..... 105

Figura 6.19 - Padrões de treinamento e padrão de teste C10I90, C30I70, C50I50, C70I30, C10I90 com oscilações de 10%..... 106

Figura 6.20 - Padrões de treinamento e padrão de teste R8C92 com oscilações de 10%. 107

LISTA DE TABELAS

Tabela 3.1 - Distribuição de funcionários segundo empresa e sexo.	21
Tabela 3.2 - Resultados e probabilidades, sem reposição.	23
Tabela 3.3- Resultados e probabilidades com reposição.	24
Tabela 3.4– Tabela contendo resultados de probabilidades.	40
Tabela 4.1 – Dados técnicos da Cal.....	54
Tabela 4.2 – Resumo dos dados de previsão da qualidade da cal, com valores simbólicos.	54
Tabela 4.3 - Intância de teste utilizado para avaliar modelo probabilístico.....	55
Tabela 4.4 - Intância de teste sem valores para atributos.....	58
Tabela 4.5 – Resumo dos dados de previsão da qualidade da Cal, com valores numéricos.....	59
Tabela 4.6 – Intância de teste com atributos numéricos.....	60
Tabela 4.7 – Tabela de Probabilidade Condicional - TPC.	63
Tabela 6.1 - Resultado dos testes realizados com padrões residenciais simples.....	79
Tabela 6.2 - Resultado dos testes realizados com padrões comerciais simples.....	80
Tabela 6.3 - Resultado dos testes realizados com padrões industriais simples.....	81
Tabela 6.4 - Resultado dos testes realizados com padrões R20C80, R40C60, R60C40 e R80C20.....	83
Tabela 6.5 - Resultado dos testes realizados com padrões R20I80, R40I60, R60I40 e R80I20.	86
Tabela 6.6 - Resultado dos testes realizados com padrões R20I80, R40I60, R60I40 e R80I20.	87
Tabela 6.7 - Resultado dos testes realizados com padrões C20I80, C40I60, C60I40 e C80I20.	88
Tabela 6.8 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 R90C10.....	89
Tabela 6.9 – Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10.....	97
Tabela 6.10 - Resultado dos testes realizados com padrões R10I90, R30I70, R50I50, R70I30 e R90I10.....	98
Tabela 6.11 - Resultado dos testes realizados com padrões C10I90, C30I70, C50I50, C70I30 e C90I10.....	98
Tabela 6.12 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10 com oscilações de 10%.	100

Tabela 6.13 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10 com oscilações de 10%.	105
Tabela 6.14 - Resultado dos testes realizados com padrões R10I90, R30I70, R50I50, R70I30 e R90I10 com oscilações de 10%.	106
Tabela 6.15 - Resultado dos testes realizados com padrões C10I90, C30I70, C50I50, C70I30 e C90I10 com oscilações de 10%.	106
Tabela 6.16 - Resultado do teste realizado com padrão R8C92 com oscilações de 10%.	107

LISTA DE SÍMBOLOS E SIGLAS

VA: Variável Aleatória

ARFF: formato específico de um arquivo.

DW: Data Warehousing.

OLAP: On Line Analytical Processing.

KDD: Knowledge Discovery in Data Base.

MD: Mineração de Dados.

RNA: Redes Neurais Artificiais.

KNN: K Nearest Neighbor.

TPC: Tabela de Probabilidade Condicional.

WEKA: Pacote computacional utilizado implementado em Java.

CEMIG: COMPANHIA ENERGÉTICA DE MINAS GERAIS.

CAPÍTULO 1

1 INTRODUÇÃO

A energia elétrica tem se mostrado como uma fonte de energia altamente versátil, capaz de ser utilizada nos mais diferentes meios e para as mais diferentes finalidades, suprimindo desde grandes fornos de siderúrgicas, até simples aparelhos eletrodomésticos. Essa versatilidade fez com que a sua utilidade aumentasse sensivelmente ao longo das últimas décadas, apresentando ainda hoje taxas elevadíssimas de crescimento. Conseqüentemente, os Sistemas Elétricos de Potência, responsáveis pelo fornecimento de grande parte de toda energia consumida nos dias atuais, passaram a ocupar uma posição de fundamental importância em vários países.

Os Sistemas Elétricos de Potência apresentam portes gigantescos, possuindo um número elevado de usinas, milhares de quilômetros de linhas de transmissão, além de uma infinidade de subestações e centros de controles. A complexidade e os elevados custos envolvidos no gerenciamento de todos esses equipamentos têm sido responsáveis por significativos esforços de empresas, centros de pesquisas e universidades de várias partes do mundo, cujos objetivos são encontrar técnicas seguras e eficientes para operar e expandir os Sistemas de Energia Elétrica.

Tanto a operação quanto a expansão dos Sistemas Elétricos de Potência visam garantir que o mercado consumidor de energia seja atendido dentro de critérios de qualidade de suprimento e baixo custo. Enquanto a operação trata do sistema existente, a expansão preocupa-se com o crescimento do mesmo. A

operação dos Sistemas enfoca a identificação de como o mercado será atendido a partir dos recursos disponíveis, isto é, quanto cada usina vai gerar, quais linhas de transmissão serão utilizadas, como serão ajustados os taps dos transformadores, etc. A expansão, por sua vez, objetiva determinar quanto, onde e como as usinas, as linhas de transmissão, os sistemas de distribuição, e todos os demais sistemas de controle devem ser adicionados ao sistema existente.

Este trabalho investiga técnicas de reconhecimento de padrões, para auxiliar tanto o setor de operação, como o setor de expansão dos Sistemas Elétricos de Potência. Ao invés de utilizar uma Rede Neural Artificial para realizar o processo de reconhecimento de padrão, utilizou-se um algoritmo estatístico chamado “Naive Bayes”.

1.1 Objetivos do trabalho

Os principais objetivos deste trabalho consistem em:

- estudar os conceitos de mineração de dados;
- estudar os conceitos gerais sobre probabilidades, dando ênfase a conceitos que servirão como base para o entendimento dos algoritmos de reconhecimento de padrões;
- estudar dois classificadores Bayesianos: Naive Bayes e Redes Bayesianas, sendo o primeiro o principal foco.
- avaliar o desempenho do algoritmo Naive Bayes, através de um estudo de caso. Este estudo de caso consiste em apresentar ao algoritmo padrões de consumidores de energia elétrica e

consequentemente, o mesmo deverá informar o percentual de cada consumidor,

- fornecer base para desenvolvimento de uma ferramenta computacional para auxiliar na operação e expansão dos Sistemas Elétricos de Potência.

1.2 Organização do texto

Inicialmente, no capítulo 2, será mostrada a definição de mineração de dados, o processo pelo qual ele se realiza, descrevendo-se as etapas de pré-processamento, extração dos padrões e pós-processamento. Esse capítulo é encerrado descrevendo-se, através de exemplos práticos, áreas onde a mineração de dados é aplicada.

No capítulo 3, serão descritos os conceitos teóricos da probabilidade com a finalidade de embasar as definições Bayesianas. Começa-se com a definição de probabilidades, seguindo com conjuntos e frequência relativa. Será abordado, também, probabilidade condicional e independência e inferência. Um dos pontos chave é o Teorema de Bayes e a Distribuição Normal, base para a utilização do algoritmo Naive Bayes.

No capítulo 4, será apresentado o classificador Naive Bayes com exemplos numéricos para facilitar seu entendimento. E ainda, uma breve descrição do funcionamento das Redes Bayesianas.

No capítulo 5, será descrito o pacote computacional Weka, utilizado como base para desenvolver este trabalho. Apresenta-se, também, como transformar um

conjunto de dados genéricos em um arquivo ARFF. Um fator a ser considerado é como este pacote é organizado para facilitar o desenvolvimento de novas aplicações.

No capítulo 6, será apresentado o estudo de caso realizado, onde o algoritmo Naive Bayes é utilizado para realizar o reconhecimento de padrões, sendo estes padrões referentes à curva de carga. Tais curvas de carga representam três classes de consumidores que são residencial, comercial e industrial.

No capítulo 7, os estudos realizados são concluídos, colocando as aplicações do que foi feito neste trabalho como base para outras pesquisas.

No capítulo 8, são apresentadas as referências bibliográficas utilizadas para realização desta dissertação.

CAPÍTULO 2

2 MINERAÇÃO DE DADOS

A evolução da computação possibilitou um aumento na capacidade de processamento e armazenamento de dados. Os aplicativos atuais têm a capacidade de gerar uma quantidade enorme de informações.

Existe um processo, chamado Data Warehousing (DW), que consiste em passos que são realizados para criar um repositório de informações, conhecido como Data Warehouse. Estes dados são estruturados de modo que facilitem a realização de consultas. A linguagem mais utilizada para realizar consultas nesses dados é OLAP (On Line Analytical Processing).

As ferramentas desenvolvidas para manipular ou realizar consultas nos dados armazenados fornecem relatórios simples, como o número de produtos vendidos em um dia. Esses relatórios são importantes para realizar uma análise dos dados, entretanto não se adequam a análises mais complexas, como, por exemplo, quais os clientes estão mais propícios a realizar fraudes, ou quais os clientes comprariam um produto caro se o mesmo tivesse condições facilitadas de pagamento. As ferramentas atuais não podem responder estas questões. Entretanto, este tipo de informação pode ser muito valiosa para os analistas de negócio de uma empresa, antes de tomar uma determinada decisão. Desta forma, novas ferramentas de análise e extração de conhecimento devem ser usadas no processo decisório.

Diante dessa deficiência para analisar e compreender um grande volume de dados surgiu uma área que é conhecida como Extração de Conhecimento em

Base de Dados (Knowledge Discovery in Data Base – KDD) ou Mineração de Dados (Data Mining). Muitos autores dizem que não são processos iguais, ou seja, a Mineração de Dados - MD é um processo da Extração do Conhecimento em Base de Dados – KDD.

2.1 Definição

Existem diversas definições para Mineração de Dados. Segundo Fayyad [9], “Extração de Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos em dados”.

2.2 O processo de mineração de dados

Os processos de Mineração de Dados - MD são específicos para cada domínio de aplicação e para cada um dos repositórios de dados inerentes aos mesmos. Para se utilizar os dados armazenados é necessário que os mesmos estejam estruturados de forma a serem consultados e analisados adequadamente.

O processo de Mineração de Dados é centrado na interação entre:

- a) especialistas do domínio: pessoa que possui amplo conhecimento do problema que se quer resolver;
- b) analista: pessoa encarregada em implementar o processo de mineração de dados;
- c) usuário final: pessoas que irão utilizar os dados da mineração para tomada de decisão.

O processo de mineração de dados pode ser dividido em várias etapas. Segundo Rezende[22], pode-se dividir o processo de Mineração de Dados em 3 etapas: Pré Processamento, Extração de Padrões e Pós Processamento. Esta divisão pode ser vista na figura abaixo.

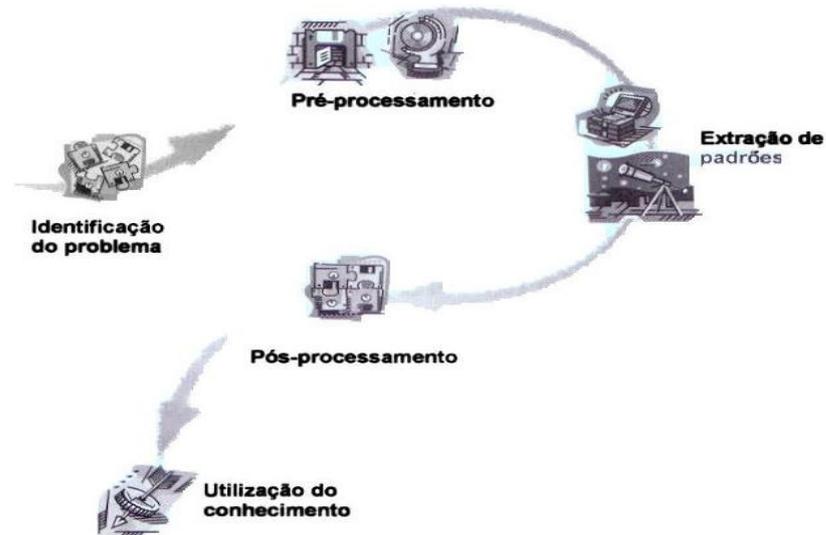


Figura 2.1 - Processo de Mineração de Dados

Apesar de não fazer parte diretamente do processo de Mineração de Dados, abaixo segue uma explicação da etapa da identificação do problema.

2.2.1 Identificação do Problema

O sucesso do processo de mineração de dados depende, em parte, do especialista que irá auxiliar o analista a determinar os padrões, estes que darão condições de realizar as devidas tomadas de decisão em um determinado domínio. Mais especificamente, na etapa de pré-processamento, esse conhecimento pode ajudar os analistas na escolha do melhor conjunto de dados para se realizar a extração de padrões, definir quais valores são válidos para os atributos, os critérios de preferência entre os possíveis atributos, as restrições de relacionamento e as informações para geração de novos atributos.

2.2.2 Pré – Processamento

Os dados, geralmente, não estão em um formato adequado para a mineração de dados e precisam ser “lapidados” antes de entrarem nos algoritmos de mineração de dados. Então, utilizam-se métodos para tratamento, limpeza e redução do volume de dados antes de realizar o processo de extração de padrões ou mineração de dados. Esses processos têm que ser gerados em cima de um repositório de informações, não esquecendo que os dados devem dar suporte aos algoritmos para que possam alcançar seus objetivos.

Os processos do Pré-Processamento podem ser divididos em:

- a) Extração e Integração: os dados podem se encontrar armazenados em diversos meios, como arquivos texto, planilhas eletrônicas, banco de dados relacional, Data Warehouse, etc. Então, deve ser feita a

extração de dados, armazenados em diversos meios e a integração deles em um único local.

- b) Transformação: após o processo de extração e integração, os dados devem ser transformados ou convertidos para uma forma considerada melhor. Algumas das transformações mais comuns são: resumo (quando dados de uma venda são agrupados para formar resultados diários), transformação de tipo (quando um atributo do tipo data é transformado em um outro tipo, para que o algoritmo possa ser utilizado posteriormente) e normalização (onde os dados são colocados entre os valores 0 e 1).
- c) Limpeza: os dados disponíveis para aplicação dos algoritmos de extração de padrões podem apresentar erros durante o processo de coleta. Esses erros podem ter sido gerados no processo de digitação de dados ou por problemas em sensores, quando se está tratando com dados oriundos de processos industriais.
- d) Seleção e Redução de Dados: devido à grande quantidade de dados armazenados têm-se limitações de espaço em memória ou tempo de processamento, o número de exemplos e de atributos disponíveis para análise pode inviabilizar a utilização dos algoritmos de Extração de Padrões. Segundo Batista e Carvalho [10], tem-se três maneiras de realizar a redução:
- redução do número de exemplos;
 - redução do número de atributos;
 - redução do número de valores de um atributo.

A redução do número de exemplos deve ser realizada a fim de manter as características do conjunto de dados original, isto é, por meio da geração de amostras representativas dos dados [12]. A abordagem mais utilizada para redução do número de exemplos é a amostragem aleatória [11], pois este método tende a produzir amostras representativas.

Com uma quantidade pequena de exemplos, pode ocorrer overfitting, ou seja, o modelo gerado pode decorar os valores não podendo resolver problemas quando os dados forem diferentes dos dados de treinamento.

A redução do número de atributos é utilizada para reduzir o espaço de resposta. O objetivo é selecionar um subconjunto dos atributos existentes de forma que não tenha grande impacto na extração de padrões. Uma forma de remover atributos é através de um especialista, pois ele pode informar se um atributo é importante ou não. Quando não se sabe quais atributos são importantes deve-se remover somente aqueles que se tem certeza.

A terceira forma de redução dos dados consiste na redução do número de valores dos atributos. Pode-se utilizar discretização ou suavização dos valores de um atributo contínuo.

A discretização consiste na substituição de um valor contínuo (inteiro ou real) por um atributo discreto. Os algoritmos que tratam de realizar a discretização recebem com entrada os valores inferiores – V_{inf} , e os valores superiores – V_{sup} , e geram uma pequena lista de intervalos ordenados.

Na suavização dos valores de um atributo, o objetivo é diminuir o número de valores do mesmo sem discretizá-lo. Neste método, os valores de um

determinado atributo são agrupados, mas, ao contrário da discretização, cada grupo de valores é substituído por um valor numérico que o represente. Esse valor pode ser a média, a mediana, ou mesmo os valores de borda de cada grupo[11].

Os métodos descritos acima devem ser utilizados com critério, para não alterar os dados originais dos repositórios de informação.

A etapa de Pré-processamento é utilizada antes da Extração de Padrões com a finalidade de facilitar os algoritmos de mineração de dados a terem uma melhor eficiência. Uma vez realizado o processo de Extração de Padrões, pode-se alterar os dados novamente com a finalidade de otimizar os resultados obtidos.

2.2.3 Extração de Padrões

Esta etapa é direcionada ao cumprimento dos objetivos descritos na etapa de identificação de problemas. Para isto é definido qual algoritmo deverá ser utilizado, ou se existe a necessidade de aplicar mais de um algoritmo para se chegar na extração de padrão desejada.

A etapa de Extração de Padrões compreende a escolha da tarefa de Mineração de Dados a ser empregada, a escolha do algoritmo e a extração dos padrões propriamente dita.

A escolha da tarefa é feita de acordo com o objetivo a ser alcançado. As tarefas possíveis de um algoritmo de Extração de Padrão podem ser: preditivas e descritivas.

Abaixo podemos ver essas tarefas e seus respectivos segmentos:

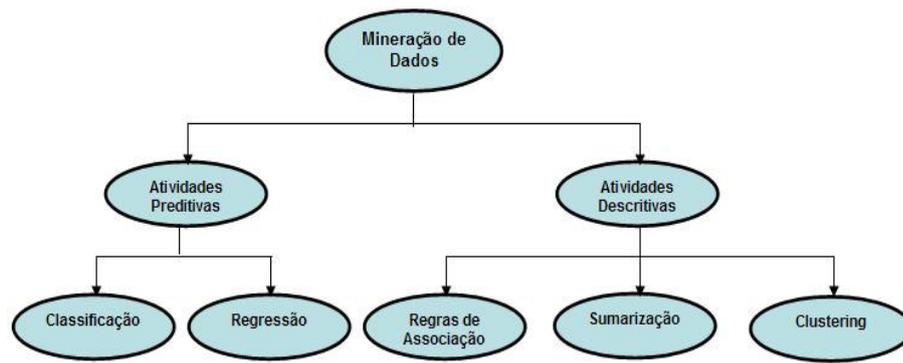


Figura 2.2 - Tarefas da extração de padrões

Atividades de predição consistem na generalização de exemplos, ou seja, têm-se uma quantidade de exemplos que são utilizados para fazer o treinamento de um classificador. A partir daí, esse classificador é capaz de reconhecer novos exemplos. Os dois principais tipos de tarefas para predição são classificação e regressão.

Já, atividades de descrição, consistem na identificação de comportamentos intrínsecos do conjunto de dados, sendo que estes dados não possuem uma classe específica.

Definida uma tarefa a ser executada, existe uma variedade de algoritmos a serem empregados. Pode-se utilizar algoritmos de Árvore de Indução ou Regra de Produção se o objetivo é realizar uma classificação.

Entre os tipos mais comuns de representação de padrões, destacam-se: Árvores de Decisão, regras de produção, modelos lineares, modelos não lineares (Redes Neurais Artificiais – RNA), modelos baseados em exemplos (KNN - K-Nearest Neighbor, Raciocínio Baseado em Casos) e modelos de dependência probabilística (Redes Bayesianas).

Em Kohavi e Sommerfield[14] é mostrado, experimentalmente, que não existe um único bom algoritmo para todas as tarefas de Mineração de dados. Por

isso, a escolha de vários algoritmos para realizar a tarefa desejada pode ser feita, levando a obtenção de diversos modelos que, na etapa de Pós-processamento, são tratados para fornecer o conjunto de padrões mais adequados ao usuário final.

O processo de Extração de Padrões consiste na aplicação dos algoritmos de extração de conhecimento para encontrar os padrões de dados embutidos no conjunto de dados. Dependendo, é necessário que aplicar o algoritmo diversas vezes.

Têm-se pesquisado a combinação de vários algoritmos com a finalidade de construir um preditor mais preciso [14].

A disponibilidade do conjunto de padrões extraídos nesta etapa ao usuário, ou a sua incorporação a um sistema inteligente, ocorre após a análise e/ou processamento dos padrões na etapa de processamento.

2.2.4 Pós - Processamento

Um dos objetivos principais do Processo de Extração do Conhecimento é que o usuário possa compreender e utilizar o conhecimento descoberto. Entretanto, podem ocorrer casos em que os modelos são muito complexos ou não fazem sentido para o especialista [16].

A compreensão de um dado conjunto de regras está relacionada com a facilidade de interpretação destas pelo usuário. A compreensibilidade de um modelo pode ser estimada, por exemplo, pelo número de regras e de condições por regra. Neste caso, quanto menor a quantidade de regras de um dado modelo e menor o número de condições por regra, maior será a compreensibilidade das regras

descobertas [10]. Em [16] e [17], é discutido que outros fatores, além do tamanho do modelo, são importantes na determinação da compreensão de um conhecimento.

Como diferentes usuários finais do Processo de Extração de Conhecimento podem ter diferentes graus de interesse para um determinado padrão, medidas subjetivas são necessárias. Estas medidas consideram que fatores específicos do conhecimento do domínio e de interesse do usuário devem ser tratados ao selecionar um conjunto de regras interessantes ao usuário. Algumas medidas subjetivas são inesperabilidade e utilidade [21].

Em um ambiente para avaliação do conhecimento, aspectos objetivos de interesse podem ser utilizados como um primeiro filtro para selecionar regras potencialmente interessantes ao usuário. Os aspectos subjetivos podem ser utilizados como um filtro final para selecionar regras realmente interessantes.

Após a análise do conhecimento, caso este não seja de interesse do usuário final ou não cumpra com os objetivos propostos, o processo de extração pode ser repetido ajustando-se os parâmetros ou melhorando o processo de escolha dos dados para a obtenção de resultados melhores numa próxima interação.

2.3 Campo de aplicação

Existem diversas áreas de aplicação da Mineração de Dados – MD, segue abaixo dois exemplos.

2.3.1 Decisão envolvendo julgamento

Quando se quer realizar um empréstimo em uma financeira, deve-se preencher um questionário sobre dados pessoais e financeiros. Esses dados são utilizados, pela financeira, como base para realizar o empréstimo. O processo de decisão, se conveniente ou não emprestar o dinheiro, é realizado em dois estágios. No primeiro, são utilizados métodos estatísticos que resolvem 90% dos casos, onde os candidatos ao empréstimo estão acima de um limite pré-estabelecido. Para os candidatos que estão abaixo desse limite, 10% do total, a análise do crédito é realizada por um especialista. Examinando os dados históricos, pode-se concluir que, por exemplo, metade dos créditos aprovados para os 10% dos candidatos (abaixo do limite) indicam que eles não pagavam suas dívidas com a financeira.

Utiliza-se Mineração de Dados para predizer se o candidato iria pagar ou não o empréstimo. Para isto, utiliza-se um conjunto de dados, descrevendo casos de pessoas que realizaram o empréstimo e que pagaram ou não suas dívidas com a financeira. Para cada exemplo poderiam ser considerados atributos, extraídos do questionário, como idade, tempo de emprego, tempo que mora na mesma

residência, etc. Neste caso, pode-se utilizar um pequeno classificador para que uma correta predição dos casos.

2.3.2 Predição de carga

Na indústria de fornecimento de energia elétrica, é importante determinar a demanda futura de energia. Estimativas de consumo podem ser feitas utilizando a carga mínima e máxima para cada hora, dia, mês, ou estação do ano.

Um assistente automático para predição de carga está operando na CEMIG [28] há uma década e tem como objetivo produzir a cada quinze minutos, uma predição de carga de dois dias à frente.

O primeiro passo para desenvolver esse assistente foi usar dados coletados nos últimos 10 anos para criar um repositório de informações. Foram utilizados algoritmos de Mineração de Dados para associar o dia atual com um dia armazenado no repositório de dados, assim realizando uma predição de carga.

CAPÍTULO 3

3 NOÇÕES BÁSICAS DE ESTATÍSTICA

Neste capítulo serão descritos os conceitos básicos sobre probabilidade. Esses conceitos servirão como base para o entendimento posterior do algoritmo de mineração de dados utilizado para realizar a classificação de padrões.

3.1 Definição

Um modelo determinístico significa que um modelo determina de alguma forma o resultado do experimento. Experimento é o fenômeno que interessa ser observado e cada realização dele é uma experiência. Também se refere a qualquer processo de observação e medida. Experimento aleatório é aquele que gera vários resultados diferentes, executado nas mesmas condições. Por exemplo, ao calcular o fluxo de corrente elétrica em uma bateria com a fórmula $I = E/R$. Ao colocar o valor de E e R, imediatamente tem-se o valor o valor I. Se houvesse vários cálculos consecutivos, se teria, sempre, o mesmo valor de I. Quaisquer desvios seriam mínimos, podendo então deduzir que sempre se observaria esses valores para I [1].

Para uma situação, onde não há condições de prever com certeza o valor do experimento, mesmo sabendo todas as propriedades que poderiam influenciar no resultado final do experimento, não existindo modelo matemático determinístico que forneça o resultado exato, deve-se chamar de modelo probabilístico (não-determinístico). Assim, o modelo probabilístico é um instrumento matemático que

prevê a chance de um possível resultado sem que seja necessário repetir a experiência [2].

3.2 Conjuntos

Alguns conceitos de conjuntos devem ser estudados para conhecer o modelo probabilístico. Conjunto é uma coleção de objetos e é representado por uma letra maiúscula X, Y . Pode ser escrito como $X = \{9, 6, 1, 3\}$.

O conjunto fundamental é o conjunto de todos os objetos que estejam sendo estudados. É, geralmente, representado pela letra maiúscula U .

Para cada experimento probabilístico E (por exemplo, jogar uma moeda e ver qual valor aparecerá) define-se o espaço amostral como o conjunto de todos os resultados possíveis do experimento E . Geralmente, representa-se esses valores por $S = \{\text{cara, coroa}\}$.

Considerando um espaço amostral S , associado a um experimento E , um evento é o conjunto de resultados possíveis. Um evento é um subconjunto do espaço amostral S .

3.3 Frequência relativa

Frequência relativa se dá quando um experimento E (jogar um dado), for repetido n vezes e A é o evento associado a E .

Antes de se definir a Frequência Relativa, deve mencionar a Frequência Absoluta que é a quantidade de vezes que o evento A aparece, mesmo os repetidos,

em um conjunto de dados (espaço amostral). A freqüência absoluta é definida por N_a . Por sua vez, a freqüência relativa é definida por

$$\int A = \frac{N_a}{n}$$

onde N_a é a freqüência absoluta e n são as repetições de E . A freqüência relativa possui as seguintes propriedades:

$$0 \leq \int A \leq 1$$

$$\int A = 1 \text{ se } A \text{ ocorrer nas } n \text{ repetições de } E.$$

$$\int A = 0 \text{ se } A \text{ não ocorrer nas } n \text{ repetições de } E.$$

Exemplo: ao jogar 5 vezes um moeda encontrou-se os seguintes resultados do experimento: cara,coroa,cara,cara,coroa. O $n = 5$ e o evento A é {a moeda com o lado da coroa}. No final da 5ª jogada, a freqüência absoluta é 2 (a coroa aparece 2 vezes) e a freqüência relativa é igual a $2/5$.

3.4 Noções fundamentais de probabilidade

A forma mais comum de medir incertezas relacionadas a eventos (subconjuntos de resultados de um experimento) consiste em atribuir probabilidades ou especificar as chances de ocorrência do evento [3].

Segundo Meyer [1], avaliar quão verdadeiro é a ocorrência de A , utilizando a freqüência relativa, pode ter alguns problemas:

- a) saber quanto deverá ser o tamanho de n ;
- b) se considerar a sorte do experimentador pode ocorrer variações na freqüência relativa, pois pode-se por sorte, repetir naquele experimento

um resultado mais vezes, mesmo com o dado equilibrado. Em outro experimento, o resultado pode se inverter.

Quando não se pode ter informações a respeito dos resultados possíveis, pode-se descrevê-lo como resultados equiprováveis ou igualmente prováveis. Um exemplo seria apresentado com a escolha de um dos quatro elevadores disponíveis em um edifício.

Caso for aplicar alguma experiência com elementos equilibrados que gerarão n resultados, a probabilidade de cada resultado será $1/n$. A probabilidade a priori é a razão do número de resultados equiprováveis favoráveis a uma determinada situação pelo número total dos resultados [2].

Quando se deseja ter um resultado sem recorrer a uma experimentação deve se proceder da seguinte maneira: Seja S um espaço amostral associado ao experimento E . A cada evento A , associa-se um número real $P(A)$ e denomina-se probabilidade de A , que satisfaça as seguintes propriedades:

- $0 \leq P(A) \leq 1$;
- $P(S) = 1$, é a probabilidade do evento certo;
- se os eventos A e B forem incompatíveis (mutuamente excludentes) então: $P(A \cup B) = P(A) + P(B)$;
- se os eventos $A_1 + A_2 + A_3 + \dots + A_n$ forem dois a dois, eventos mutuamente excludentes então:

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1) + P(A_2) + \dots + P(A_n) = \sum_{i=1}^n P(A_i)$$

Para calcular a probabilidade de A :

$$P(A) = \frac{\text{número de casos favoráveis a } A \text{ pelos quais } E \text{ pode ocorrer}}{\text{número total de casos pelos quais } E \text{ pode ocorrer}}$$

Exemplo: Ao se lançar um dado equilibrado, e se deseja verificar o evento

$A=\{5,6\}$. O resultado de $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$

3.5 Probabilidade condicional e independência

Na tabela 3.1, tem-se dados referentes a um seminário realizado entre empresas de engenharia com a finalidade de trocar experiências.

Tabela 3.1 - Distribuição de funcionários segundo empresa e sexo.

Sexo \ Empresa	Homens	Mulheres	Total
AST Engenharia Ltda.	70	40	110
Canto Engenharia Ltda.	15	15	30
Ferreira Veloso Engenharia Ltda.	10	20	30
Autcontrol Ltda.	20	10	30
Total	115	85	200

Dado um funcionário, escolhido ao acaso, que faça parte da empresa Ferreira Veloso Engenharia Ltda, a probabilidade de que seja mulher é $20/30 = 2/3$. Isso porque, do total de 30 funcionários, 20 são mulheres. Escreve-se

$$P(\text{mulher} \mid \text{FerreiraVelosoEgenharia}) = \frac{2}{3}.$$

Para dois eventos quaisquer A e B, sendo $P(B) > 0$, define-se a probabilidade condicional de A dado B, $P(A|B)$, como sendo

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Para o exemplo mencionado, se B e A indicam, respectivamente, os eventos “funcionário da Ferreira Veloso Engenharia” e “funcionário é mulher”, então:

$$P(A|B) = \frac{\frac{20}{200}}{\frac{30}{200}} = \frac{2}{3}, \text{ como obtido anteriormente.}$$

Observe que $P(A) = P(\text{mulher}) = \frac{85}{200} = \frac{17}{40}$, e com a informação de que B ocorreu (o funcionário pertence a Ferreira Veloso Engenharia), obtém-se $P(A|B) = \frac{2}{3}$. Pode-se dizer que $P(A)$ é a probabilidade a priori de A e, com a informação adicional de que B ocorreu, obtém-se a probabilidade a posteriori $P(A|B)$. Note que, nesse caso, $P(A|B) > P(A)$, logo a informação de que B ocorreu aumentou a chance de A ocorrer.

Da equação acima, obtemos a chamada regra do produto de probabilidades,

$$P(A \cap B) = P(B).P(A|B)$$

Exemplo: Uma caixa contém duas camisas brancas(B) e três vermelhas(V) para serem sorteadas entre torcedores de futebol. Suponha que são sorteadas duas camisas ao acaso, sem reposição. Isso significa que escolhida a primeira camisa, verificada sua cor e não foi devolvida à caixa; misturou-se as camisas restantes e então retirou-se a segunda. O diagrama em árvores da figura 3.1, ilustra as possibilidades.

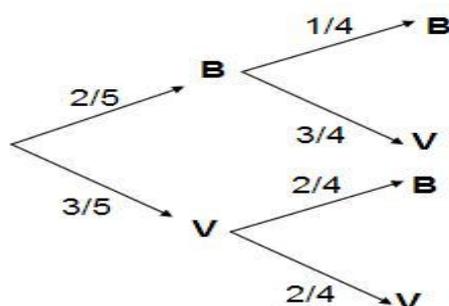


Figura 3.1 - Diagrama para extração de duas camisas de uma caixa, sem reposição.

Em cada “galho” da árvore estão indicadas as probabilidades de ocorrência, sendo para a segunda camisa sorteada as probabilidades são condicionais. A probabilidade do resultado conjunto é dada, então, por $P(A \cap B) = P(B).P(A | B)$. Veja a tabela 3.2.

Tabela 3.2 - Resultados e probabilidades, sem reposição.

Resultados	Probabilidades
BB	$2/5 \times 1/4 = 2/20$
BV	$2/5 \times 3/4 = 6/20$
VB	$3/5 \times 2/4 = 6/20$
VV	$3/5 \times 2/4 = 6/20$
Total	1

Se A indicar o evento “camisa branca na segunda extração”, então

$$P(A) = P(BB) + P(VB) = \frac{2}{20} + \frac{6}{20} = \frac{2}{5}$$

Imagine, agora, que as duas extrações são feitas da mesma caixa do exemplo anterior, mas a primeira bola é repostada na caixa antes da extração da segunda. Nestas condições, as extrações são independentes, pois o resultado de

uma extração não tem influência no resultado da outra. Então obteve-se a figura 3.2 e a tabela 3.3.

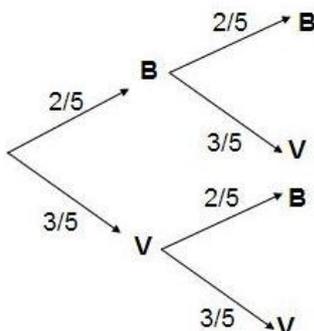


Figura 3.2 - Diagrama para extração de duas camisas de uma caixa, com reposição.

Tabela 3.3- Resultados e probabilidades com reposição.

Resultados	Probabilidades
BB	$2/5 \times 2/5 = 4/25$
BV	$2/5 \times 3/5 = 6/25$
VB	$3/5 \times 2/5 = 6/25$
VV	$3/5 \times 3/5 = 9/25$
Total	1

Observa-se que,

$$P(\text{camisa_branca_2}^\circ | \text{camisa_branca_1}^\circ) = \frac{2}{5} = P(\text{camisas_branca_2}^\circ),$$

ou seja, se indicar por A e B os eventos “camisa branca na segunda extração” e “camisa branca na primeira extração”, respectivamente, então $P(A|B) = P(A)$. Nesse caso, diz que o evento A independe do evento B e, então

$$P(A \cap B) = P(A).P(B).$$

É fácil ver que se A independe de B, então B independe de A – diz-se que A e B são independentes. A fórmula acima pode ser tomada como definição de independência entre dois eventos, ou seja, A e B são independentes se, e somente se, a fórmula acima for válida.

3.6 Inferência

A partir de resultados obtidos em um evento, pode-se a inferência caracterizar a população como um todo. Generalizar aspectos importantes de uma população a partir dos resultados obtidos na ocorrência de um evento (amostra), caracteriza a inferência [2].

Para que as inferências sejam válidas, a ocorrência (amostra de um evento) deve ser representativa na população. A escolha de uma ocorrência é geralmente feita de uma forma mais conveniente, caso contrário, podendo provocar um resultado de inferência tendencioso.

Para resolver este problema deve ser feita a escolha aleatória da ocorrência, para que o resultado não seja subestimado ou superestimado.

Inferências probabilísticas são utilizadas em algoritmos de propagação de crenças em Redes Bayesianas, podendo ser do tipo causal (parte das causas para os efeitos), diagnóstico (parte dos efeitos para as causas), intercausal (discrimina entre causas de um efeito comum) ou misto (combinação de dois ou mais tipos acima) [5]. Nesta aplicação, a inferência calcula a probabilidade de um evento, dado as evidências observadas na rede. As evidências são as observações de ocorrências de valores específicos de variáveis aleatórias, usadas na estimação das probabilidades das outras variáveis assumirem certos valores.

3.6.1 Método geral

Este método é utilizado quando há necessidade de usar informações sobre valores de algumas variáveis com a intenção de obter probabilidades para os valores de outras variáveis.

O cenário para inferência deve ter um conjunto, V , de variáveis V_1, V_2, \dots, V_k e se consegue como *evidência* (e), que as variáveis em um subconjunto E de V , tem certamente valores definidos, $E = e$. Essas variáveis têm tipicamente valores determinados por processos percentuais. Deseja-se calcular a probabilidade condicional de $p(V_i = v_i | E = e)$ que alguma variável V_i tendo como valores v_i , dado a evidência e . Este processo é chamado de *evidência probabilística* [6].

3.6.2 Independência Condicional

Segundo Nilsoon [6], define-se que uma variável V é *condicionalmente independente* de um conjunto de variáveis V_i , dado um conjunto V_j , se $p(V | V_i, V_j) = p(V | V_j)$ e usando a notação $I(V, V_i | V_j)$ como representação. A independência condicional é que se $I(V, V_i | V_j)$, o conjunto de variáveis V_i é irrelevante para V , sabendo V_j . Sendo V uma variável de interesse, se conhece V_j , podendo então ignorar V_i . Se uma simples variável V_i é condicionalmente independente de outra variável V_j , dado um conjunto V , tem-se por definição $p(V_i | V_j, V) = p(V_i | V)$. Pela definição da probabilidade condicional, tem-se $p(V_i | V_j, V)p(V_j | V) = p(V_i, V_j | V)$. Combinado esses dois resultados produz:

$$P(V_i | V_j, V) = p(V_i | V)p(V_j | V)$$

No caso que $I(V_i, V_j | V)$, nota-se que V_i e V_j aparecem simetricamente. Dessa forma, dizendo que V_i é condicionalmente independente de V_j , dado V , pode-se dizer também que V_j é condicionalmente independente de V_i , dado V . Diz-se que V_i e V_j são condicionalmente independentes, dado V . Estes mesmos resultados aplicados a conjuntos, se V_i e V_j , são condicionalmente independentes, dado V (que é, $I(V_i, V_j | V) = p(V_i | V)p(V_j | V)$). Se V é vazio, simplesmente diz-se que V_i e V_j são independentes.

Como as generalizações de pares independentes, dizem que as variáveis V_1, \dots, V_k são mutuamente independentes condicionais, dado o conjunto V , se cada uma das variáveis são condicionalmente independentes de todas outras, dado V . Então a regra de cadeia é dada por:

$$P(V_1, V_2, \dots, V_k | V) = \prod_{i=1}^k p(V_i | V_{i-1}, \dots, V_1, V)$$

E, então, cada V_i é condicionalmente independente de outras dado V ,

$$P(V_1, V_2, \dots, V_k | V) = \prod_{i=1}^k p(V_i | V)$$

Quando V é vazio, tem-se $P(V_1, V_2, \dots, V_k | V) = p(V_1)p(V_2)\dots p(V_k)$ e diz-se que as variáveis são incondicionalmente independentes.

Independências Condicionais podem ser representadas em estruturas chamadas *Redes Bayesianas* (também chamadas de redes de crenças). Estas estruturas são muito usadas para probabilidades de inferências. As independências condicionais representadas em *redes bayesianas* podem gerar maior economia no processamento de inferência probabilística.

3.7 Amostragem e distribuição amostral

Nas pesquisas estatísticas, uma das principais preocupações é ter uma escolha segura das amostras para garantir uma generalização com um alto grau de confiabilidade. Uma escolha incorreta das amostras pode gerar um resultado tendencioso e distorcido comprometendo a finalização dos resultados. Na maioria das vezes, as pesquisas trabalham com amostras aleatórias. Mas nem sempre as amostras aleatórias são viáveis. Esta seção tem como objetivo mostrar os conceitos básicos de cada termo utilizado na probabilidade e não mostrar as deduções matemáticas para formação de cada fórmula, tratando sobre outras opções alternativas de amostragem [3].

3.7.1 Conceitos Básicos

Primeiramente, deve-se distinguir claramente a diferença entre *população* e *amostra*. *População* é o conjunto de dados de todas as observações possíveis (concebíveis ou hipotéticas). *Amostra* é um conjunto de dados de uma parte dessas observações. Cabe ressaltar que o termo “observações hipotéticas” é definido como as amostras que poderão existir se o experimento for repetido continuamente.

3.7.1.1 Média

Também chamada de média aritmética, é representada pela soma de n números dividido por n .

Exemplo: em uma indústria registrou-se em um ano (12 meses) os seguintes números de produção ocorridos mensalmente. A média da produção mensal é:

33 12 11 34 32 12 6 19 11 19 17 10

somando todos os valores e dividindo por 12: $216/12 = 18$ toneladas/mês.

3.7.1.2 Média Amostral

É a soma das amostras escolhidas (em foco) de quantidade n , dividida pela quantidade n das amostras escolhidas. A representação da média amostral é dada pela letra \bar{x} e calculada por

$$\bar{x} = \sum x/n$$

Onde:

$\sum x$ significa o somatório dos valores de x que são amostras em foco para o cálculo da média e n o número total de x em foco.

Exemplo: conforme especificado no exemplo anterior, pode-se necessitar calcular a média da produção do primeiro trimestre: neste caso, deve-se somar os três primeiros meses e dividir por três:

33 12 11 34 32 12 6 19 11 19 17 10

$33 + 12 + 11 = 56/3 = 18,66$ toneladas / mês.

3.7.1.3 Média Populacional

A denotação do tamanho da população é dada pela letra N. A média populacional (μ) é definida pelo cálculo da média de todas as observações possíveis (população) dividida pela quantidade total de observações (N).

$$\mu = \sum x / N$$

Onde:

$\sum x$ é a soma de todas as observações da população e N é a quantidade total de elementos da população.

Exemplo: seguindo o raciocínio dos exemplos anteriores sobre a produção mensal em uma indústria, caso a população dos exemplos seja os totais de produção por mês, desde o primeiro dia de funcionamento da indústria até o dia em que foi efetuado o levantamento, finalizando em 24 meses de funcionamento da referida indústria:



 9 12 11 10 7 12 15 17 20 9 19 20 33 12 11 34 32 12 6 19 11 19 17 10

Estes valores são definidos como a população do exemplo. A média populacional é a soma desses valores dividido pelo número total de observações:

$$\mu = 377 / 24 = 15,71$$

3.7.1.4 Desvio-Padrão Amostral

Os valores de cada elemento do conjunto de dados observado estão concentrados em torno da média e outros estão bem afastados da média. Este motivo leva a medir a variação do conjunto de dados em termos de desvio dos valores respectivos, a contar da média. Se um conjunto de dados $x_1, x_2, x_3, \dots, x_n$ com média \bar{x} , o cálculo do desvio-padrão é dado pela fórmula:

$$\text{Desvio_padrão} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

e o *desvio-padrão amostral* é dado pela fórmula

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

e a *variância amostral* é obtida elevando-se os dois lados ao quadrado

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Exemplo: utilizando o exemplo acima que trata sobre a produção do primeiro trimestre da produção, pode-se calcular o desvio padrão:

$$s = \sqrt{\frac{(33-18,66)^2 + (12-18,66)^2 + (11-18,66)^2}{3}} = \sqrt{\frac{308,65}{3}} = \sqrt{102,88} = 10,14$$

o desvio-padrão amostral é:

$$s = \sqrt{\frac{(33-18,66)^2 + (12-18,66)^2 + (11-18,66)^2}{2}} = \sqrt{\frac{308,65}{2}} = \sqrt{154,32} = 12,42$$

3.7.1.5 Desvio-padrão Populacional

A fórmula do desvio padrão populacional é dada pela denotação σ e a média \bar{x} é substituída pela média populacional μ e o n é substituído pela quantidade total de observações da população N .

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

e a variância populacional é dada pelo quadrado dos dois lados ficando com a denotação

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Exemplo: dando continuidade ao exemplo anterior sobre produção industrial, o desvio-padrão populacional é:

$\mu = 16,39$ e $N=23$, então

$$\sqrt{\frac{1430,06}{23}} = \sqrt{62,17} = 7,88$$

3.7.1.6 Combinações

Existem muitas formas de saber o número de maneiras de escolher r objetos de um conjunto de n objetos, sem que haja interesse de saber qual a ordem em que se faz a escolha. Por exemplo, dentro de um conjunto de 36 membros de uma associação de moradores, deseja-se escolher cinco membros destes 36, nesta situação, pode-se querer saber o total de combinações possíveis de 5 membros dentro desse conjunto de 36. É óbvio que nesse caso não é necessário exigir a ordem em que esses membros serão apresentados, assim a fórmula da combinação é o cálculo da permutação, $n(n-1)(n-2)\dots(n-r+1)$ dividido pelo fatorial de r ($r!$), r seria o número total de elementos agrupados. Assim, a fórmula da *combinação* será:

$${}_n C_r \text{ ou } \binom{n}{r} = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!}$$

ou em notação fatorial

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

onde n é o número total de objetos de um conjunto de dados e r é o número de objetos de um grupo a escolher.

3.7.2 Amostragem aleatória

Uma população finita consiste de um número finito de elementos, medidas e observações. Para apresentar a amostragem aleatória de uma população finita, primeiramente necessita aplicar o cálculo da combinação e saber quantas amostras de tamanho n podem ser extraídas de uma população finita de tamanho N $\binom{N}{n}$.

Uma amostra de tamanho n de uma população finita de tamanho N , é aleatória se ela for escolhida dentro das $\binom{N}{n}$ amostras possíveis, tem-se a mesma chance de ser escolhida, ou seja, $\frac{1}{\binom{N}{n}}$.

Exemplo: Se uma população apresenta 5 elementos $N=5$ e as amostras apresentam 3 elementos, $n=3$ então $\binom{5}{3} = 10$ combinações de 3 elementos.

O problema maior é como as amostras aleatórias são escolhidas na prática. O ideal seria colocar cada amostra em dez papéis e colocá-los numa urna, misturá-los e tirar um sem olhar. Fica fácil quando se trabalha com poucas amostras, mas quando se necessita de muitas amostras, a operação fica mais complexa com o método citado anteriormente. Um exemplo no qual uma amostra de n elementos de uma amostra de população N teria que escrever $\binom{N}{n}$ papéis para dar a mesma chance para as amostras possíveis. Um recurso interessante para resolver esse problema, sem escrever todas as amostras possíveis, é colocar os N elementos da população dentro de uma urna e tirar n elementos sem reposição.

Outro recurso, mais simples, seria escolher números aleatórios, utilizando tabelas de números aleatórios ou programas de computador.

3.7.3 Planejamento amostral

O planejamento de amostra é um plano definido, que é totalmente determinado antes de coletarem quaisquer dados. O objetivo é a obtenção de uma amostra de uma dada população. Existem três tipos de planejamento que serão descritos sucintamente nos próximos itens.

3.7.3.1 Amostragem sistemática

Em alguns casos, a maneira mais prática de extrair uma amostra, seria buscar em uma certa posição a cada x , amostras de uma população. Por exemplo,

buscar o trigésimo nome em uma lista e assim por diante. Neste método, denominado, *amostragem sistemática*, pode-se extrair um elemento aleatoriamente. O problema que pode acontecer com esse tipo de amostragem é que podem ocorrer periodicidades ocultas. Por exemplo, em uma linha de montagem, se extrai como amostra sempre a trigésima peça e por algum problema nos equipamentos sempre a oitava peça é que esta com defeito.

3.7.3.2 Amostragem estratificada.

Consiste em dividir ou estratificar a população em um certo número de subpopulações que não se superpõem, chamadas estratos, ou camadas, e então extrair uma amostra de cada estrato. Por exemplo, considere uma população de $N=10$ cidadãos, para os quais definimos as variáveis renda familiar (X_1) e classe social (X_2), categorizada como A, B ou C. Então, $P = \{1,2,\dots,10\}$ e suponha que a matriz de dados seja:

$$D=[10/B \ 8/C \ 15/A \ 6/C \ 22/A \ 12/B \ 7/C \ 16/A \ 13/B \ 11/B]$$

Pode-se considerar três estratos, determinados pela variável X_2

$$P_A = \{3,5,8\}, P_B = \{3,5,8\}, P_C = \{3,5,8\}$$

3.7.3.3 Amostragem por conglomerado

Como ocorrido na Amostragem Aleatória Estratificada, a população é dividida em grupos (subpopulações) distintos, chamados conglomerados. Por

exemplo, pode-se dividir uma cidade em bairros ou quadras. Utiliza-se este método para selecionar uma amostra de conglomerados e depois todos os indivíduos do conglomerados selecionados são analisados. Outro exemplo é quando um prefeito de uma cidade deseja obter opinião de membros da comunidade local sobre o novo orçamento participativo implantado. Ele pode obter uma amostra por conglomerados, entrevistando alguns ou todos os elementos de várias comunidades locais escolhidas aleatoriamente.

3.8 Teorema de Bayes

Uma das relações mais importantes envolvendo probabilidades condicionais é dada pelo Teorema de Bayes [7]. A versão mais simples desse teorema é dada pela fórmula a seguir.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}$$

Como foi salientado anteriormente, tem-se a probabilidade inicial $P(A)$ e, dada a informação de que B ocorreu (ou dada a suposição de que B venha a acontecer), obtém-se a probabilidade a posteriori $P(A|B)$, dada pela equação acima. Ou seja, atualizou-se a probabilidade inicial, multiplicando-a por $\frac{P(B|A)}{P(B)}$.

Observe que $P(A|B) > P(A)$ se $P(B|A) > P(B)$.

A forma geral do Teorema de Bayes será introduzida por um exemplo. Têm-se cinco sacos, cada uma com seis feijões. Têm-se o feijão tipo branco e o preto. Dois desses sacos (tipo C1) têm 3 feijões brancos, dois outros (tipo C2) têm 2 feijões brancos e o último saco (tipo C3) têm 6 feijões brancos. Escolheu-se um saco

ao acaso e dele retirou-se uma bola. Qual a probabilidade do saco escolhido ser do tipo C3, sabendo-se que o feijão sorteado é branco?

Deseja-se encontrar $P(C_3 | B)$, sabendo que

$$P(C_1) = \frac{2}{5}, \quad P(B | C_1) = \frac{6}{12} = \frac{1}{2}$$

$$P(C_2) = \frac{2}{5}, \quad P(B | C_2) = \frac{4}{12} = \frac{1}{3}$$

$$P(C_3) = \frac{1}{5}, \quad P(B | C_3) = \frac{6}{6} = 1$$

Da definição de probabilidade condicional, tem-se

$$P(C_3 | B) = \frac{P(C_3 \cap B)}{P(B)} = \frac{P(C_3)P(B | C_3)}{P(B)}$$

Precisa-se encontrar o valor de $P(B)$, já que o numerador é conhecido.

Como C_1 , C_2 e C_3 são eventos mutuamente exclusivos, e reunidos formam o espaço amostral completo, pode-se decompor o evento B na reunião de três outros, também mutuamente exclusivos.

$$B = (C_1 \cap B) \cup (C_2 \cap B) \cup (C_3 \cap B),$$

e então

$$P(B) = P(C_1 \cap B) \cup P(C_2 \cap B) \cup P(C_3 \cap B)$$

$$P(B) = P(C_1).P(B | C_1) + P(C_2).P(B | C_2) + P(C_3).P(B | C_3)$$

$$P(B) = \frac{2}{5} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{3} + \frac{1}{5} \cdot 1 = \frac{8}{15}$$

Substituindo este resultado tem-se,

$$P(C_3 | B) = \frac{\frac{1}{5} \cdot 1}{\frac{8}{15}} = \frac{3}{8}$$

Pode-se generalizar os resultados acima do seguinte modo: seja $\{C_1, C_2, \dots, C_n\}$, uma partição do espaço amostral Ω , isto é,

$$C_i \cap C_j = \emptyset, \text{ sempre que } i \neq j,$$

$$C_1 \cup C_2 \cup \dots \cup C_n = \Omega$$

Considerando um evento qualquer A em Ω e conhecidas as probabilidades $P(C_i)$ e $P(A|C_i)$, $i = 1, 2, 3, \dots, n$.

A probabilidade de ocorrência do evento C_i , supondo-se a ocorrência do evento A , é dada por

$$P(C_i | A) = \frac{P(C_i)P(A|C_i)}{\sum_{j=1}^n P(C_j).P(A|C_j)} \quad (3.1)$$

para todo $i = 1, 2, 3, \dots, n$.

Pode-se pensar C_1, \dots, C_n como um conjunto de hipóteses, sendo somente uma delas verdadeira. Dado que A ocorreu, a probabilidade inicial de $C_i, P(C_i)$, é modificada de modo a se obter $P(C_i | A)$, dada pela equação (3.1). Passou-se da probabilidade a priori $P(C_i)$ para a probabilidade a posteriori $P(C_i | A)$, multiplicando a primeira por:

$$P(C_i | A) = \frac{P(A|C_i)}{\sum_{j=1}^n P(C_j).P(A|C_j)} \quad (3.2)$$

Para A fixado, as probabilidades $P(A|C_i)$ na equação 3.1 são denominados verossimilhanças das hipóteses C_1, C_2, \dots, C_n . Vê-se que se a equação 3.2 for maior do que um, isto é, se $P(A|C_i) > P(A)$, onde $P(A)$ é o denominador da equação 3.2. Observe que esse denominador é uma média ponderada dos $P(A|C_i)$ e os pesos são as probabilidades $P(C_j)$, que têm soma unitária. Como o

numerador é sempre uma das parcelas do denominador $P(A)$, torna-se indispensável o uso de um novo índice, j , na decomposição deste. Segue abaixo um exemplo.

Em concurso público, uma das etapas de seleção era oferecido aos candidatos um curso de treinamento durante um mês. No final do curso, eles são submetidos a uma prova e 25% são classificados como bons (B), 50% são classificados como médios (M) e os restantes 25% como fracos (F). Para facilitar a seleção, os organizadores do concurso público pretende substituir o treinamento por um teste contendo questões referentes a conhecimentos gerais. Para isso, gostaria de conhecer qual a probabilidade de um candidato aprovado no teste ser considerado fraco, caso fizesse o curso. Assim, neste ano, antes do início do curso, os candidatos foram submetidos ao teste e receberam o conceito aprovado (A) ou reprovado (R). No final do curso, obtiveram-se as seguintes probabilidades condicionais:

$$P(A|B) = 0,80 \quad P(A|M) = 0,50 \quad P(A|F) = 0,20$$

Deseja-se encontrar $P(F|A)$ e, pelo Teorema de Bayes, essa probabilidade é dada por

$$P(F|A) = \frac{P(A|F)P(F)}{P(A|B).P(B) + P(A|M).P(M) + P(A|F).P(F)}$$

$$P(F|A) = \frac{(0,20)(0,25)}{(0,80).(0,25) + (0,50).(0,50) + (0,20).(0,25)} = 0,10$$

Então apenas 10% dos aprovados é que seriam classificados como fracos durante o curso. De modo análogo podemos encontrar $P(B|A) = 0,40$ e $P(M|A) = 0,50$, que poderiam fornecer subsídios para ajudar na decisão de substituir o treinamento pelo teste.

Um gráfico tipo árvore pode ajudar bastante na solução de um problema envolvendo o Teorema de Bayes. Desse modo, para o exemplo acima, teremos a figura 3.3 e a tabela 3.4.

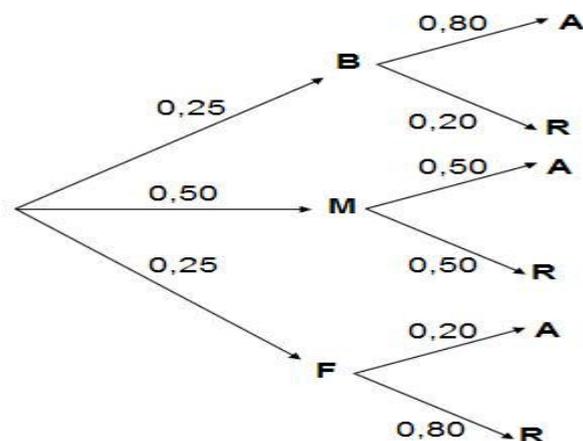


Figura 3.3 - Árvore contendo resultados do exemplo acima.

Tabela 3.4– Tabela contendo resultados de probabilidades.

Resultados	Probabilidades
BA	$(0,25).(0,80) = 0,20$
BR	$(0,25).(0,20) = 0,05$
MA	$(0,50).(0,50) = 0,25$
MR	$(0,50).(0,50) = 0,25$
FA	$(0,25).(0,20) = 0,05$
FR	$(0,25).(0,80) = 0,20$

O Teorema de Bayes, que aparentemente poderia ser encarado como mais um resultado na teoria de probabilidades, tem importância fundamental, pois fornece a base para uma abordagem da inferência estatística conhecida como inferência Bayesiana[7].

3.9 Variável aleatória

Segundo Paulo Afonso Lopes [2], o interesse em um experimento está em determinar características dos resultados que poderão ocorrer. Como em Estatística Descritiva usam-se as escalas para transformar em números as características dos elementos da amostra. Na realização de um experimento, também se deve ter um valor numérico para representar a característica cuja probabilidade de ocorrência se queira saber. Esse valor numérico, produto do experimento, pode ser representado por uma letra maiúscula X . Como o valor numérico associa um resultado do experimento a um número, o X é uma função cujo domínio de definição são os resultados do experimento e a imagem é o conjunto de números reais. Esta função X é denominada de Variável Aleatória. Isso significa que os resultados dos experimentos aleatórios serão representados por meio de números.

As variáveis aleatórias são classificadas de acordo com o número de valores que elas podem assumir:

Variáveis aleatórias discretas: são variáveis aleatórias que podem assumir apenas um número finito, ou infinito enumerável, de valores (tantos valores quanto são os números inteiros). Os valores possíveis constituem um subconjunto de números inteiros. Por exemplo, o número de ordem de jogada de um dado que aparece o número 6 pela primeira vez é uma variável aleatória discreta, que pode tomar uma infinidade enumerável de valores 1,2,3,4,...

Variáveis aleatórias contínuas: são variáveis aleatórias que aparecem quando se lida com grandezas medidas em uma escala contínua, como tempo, peso ou distância.

3.10 Distribuição normal

O modelo normal é fundamental em probabilidades e inferência estatística. Suas origens remontam a Gauss em seus trabalhos sobre erros de observações astronômicas, por volta de 1810, donde o nome de distribuição *gaussiana* para tal modelo.

3.10.1 Definição

Diz-se que a variável aleatória X tem *distribuição normal* com parâmetros μ e σ^2 , $-\infty < \mu < +\infty$ e $0 < \sigma^2 < \infty$, se sua densidade é dada por

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

3.10.2 Gráfico

A figura 3.4 ilustra uma particular *curva normal*, determinada por valores particulares de μ e σ^2 .

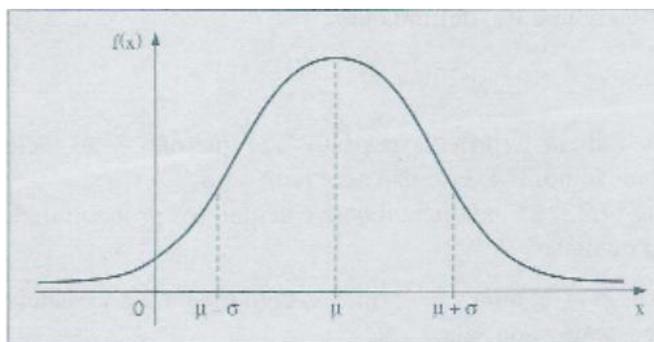


Figura 3.4 – Função de distribuição de probabilidade de uma VA – com média μ desvio padrão σ .

Pode-se notar que $f(x; \mu; \sigma^2) \rightarrow 0$, quando $x \rightarrow \pm\infty$, $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f(x; \mu, \sigma^2)$, $x = \mu$ é ponto máximo de $f(x; \mu, \sigma^2)$, e o valor máximo é $1/\sigma\sqrt{2\pi}$. A densidade $f(x; \mu, \sigma^2)$ é simétrica em relação à reta $x = \mu$, isto é,

$$f(\mu + x; \mu, \sigma^2) = f(\mu - x; \mu, \sigma^2),$$

para todo x real.

Para simplificar a notação, pode-se escrever a densidade da normal simplesmente por $f(x)$ e escreve-se, simbolicamente ,

$$X \sim N(\mu, \sigma^2).$$

Quando $\mu = 0$ e $\sigma^2 = 1$, tem-se uma distribuição *padrão* ou *reduzida*, $N(0,1)$. Para esse caso a função densidade reduz-se a

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty.$$

O gráfico da normal padrão está na figura abaixo:

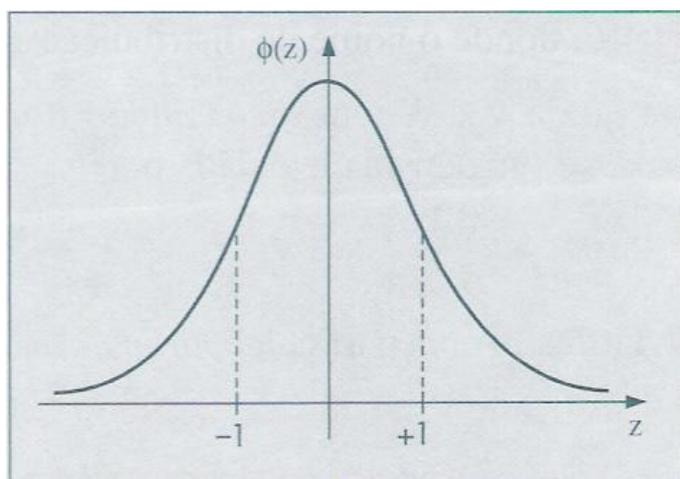


Figura 3.5 - Função de Distribuição de Probabilidade de uma va – $Z \sim N(0,1)$.

Se $X \sim N(\mu; \sigma^2)$, então a v.a é definida por $Z = \frac{X - \mu}{\sigma}$ terá média zero e variância 1.

A transformação acima é fundamental para calcularmos probabilidades relativas a uma distribuição normal qualquer.

3.10.2.1 Função de distribuição acumulada - fda

A f.d.a $F(y)$ de uma v.a normal X , com média μ e variância σ^2 é obtida integrando-se $f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ de $-\infty$ até y , ou seja,

$$F(y) = \int_{-\infty}^y f(x; \mu, \sigma^2) dx, y \in \mathbb{R}.$$

A integral acima corresponde à área, sob $f(x)$, desde $-\infty$ até y , como ilustra a figura abaixo

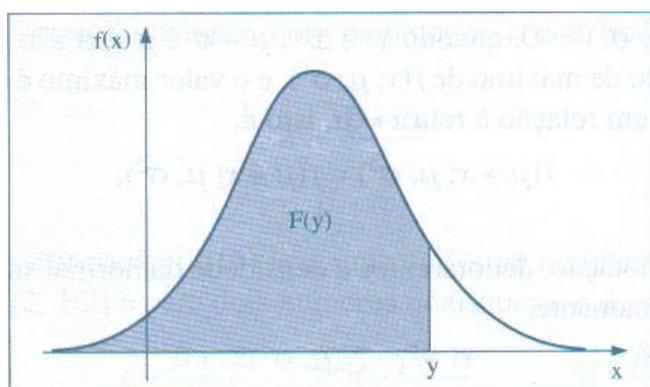


Figura 3.6 - - Representação gráfica de $F(y)$ como área.

No caso específico da normal padrão, utilizamos a seguinte notação, que é universal:

$$\Phi(y) = \int_{-\infty}^y \phi(z) dz = 1/\sqrt{2\pi} \int_{-\infty}^y e^{-z^2/2} dz.$$

Suponha, então, que $X \sim N(\mu; \sigma^2)$ e que queiramos calcular

$$P(a < X < b) = \int_a^b f(x) dx,$$

onde $f(x)$ é dada por $f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, $-\infty < x < \infty$. Ver figura

abaixo:

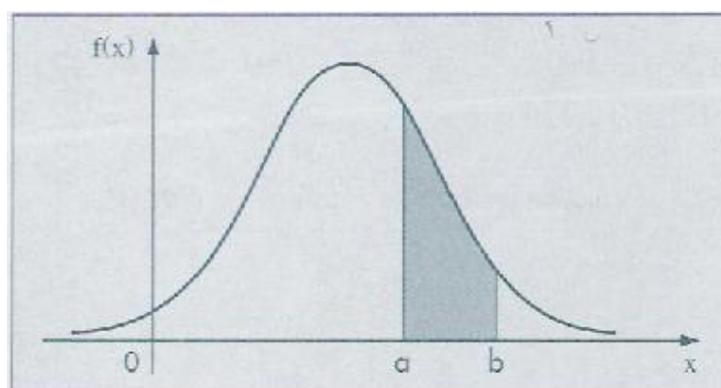


Figura 3.7 - Função para $P(a \leq X \leq b)$ para uma V.A normal

A integral $P(a < X < b) = \int_a^b f(x) dx$ não pode ser calculada analiticamente, e portanto a probabilidade indicada só poderá ser obtida, aproximadamente, por meio de integração numérica. No entanto, para *cada* valor de μ e σ , tem-se de obter $P(a < X < b)$ para diversos valores de a e b . Para facilitar esta tarefa é utilizado $Z = \frac{X - \mu}{\sigma}$, que é necessário construir uma tabela para a distribuição normal padrão.

Esta tabela é encontrada na maioria dos livros de estatística e pode ser vista no Anexo 1.

Pode-se ver, então, como obter probabilidades a partir do Anexo 1. Essa tabela nos dá as probabilidades sob uma curva normal padrão, que nada mais são do que as correspondentes áreas sob a curva. A figura abaixo ilustra a probabilidade fornecida pela tabela, onde $P(0 \leq Z \leq z_c)$ e $Z \sim N(0,1)$.

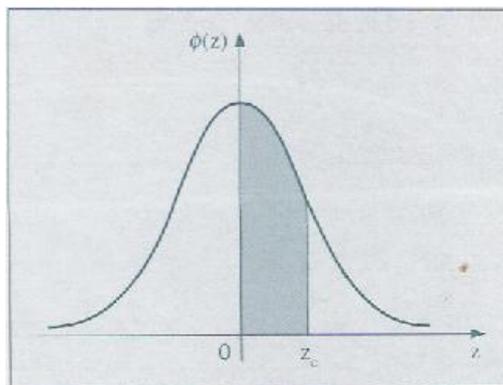


Figura 3.8 - $P(0 \leq Z \leq Z_c)$ fornecida pelo Anexo 1.

Por exemplo

(a) para $z_c = 1,73$, tem-se que $P(0 \leq Z \leq 1,73) = 0,4582$.

(b) $P(-1,73 \leq Z \leq 0) = P(0 \leq Z \leq 1,73) = 0,4582$, devido à simetria da curva.

(c) $P(Z \geq 1,73) = 0,5 - P(0 \leq Z \leq 1,73) = 0,5 - 0,4582 = 0,0418$, pois

$$P(Z \geq 0) = 0,5 = P(Z \leq 0).$$

(d) $P(Z < -1,73) = P(Z > 1,73) = 0,0418$.

(e) $P(0,47 \leq Z \leq 1,73) = P(0 \leq Z \leq 1,73) - P(0 \leq Z \leq 0,47) = 0,4582 - 0,1808 = 0,2774$

(f) Suponha agora que X seja uma v.a $N(\mu; \sigma^2)$, com $\mu = 3$ e $\sigma^2 = 16$, e queiramos calcular $P(2 \leq X \leq 5)$.

Utilizando $Z = \frac{X - \mu}{\sigma}$, tem-se

$$P(2 \leq X \leq 5) = P\left(\frac{2 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{5 - \mu}{\sigma}\right) = P\left(\frac{2 - 3}{4} \leq Z \leq \frac{5 - 3}{4}\right) = P\left(-\frac{1}{4} \leq Z \leq \frac{1}{2}\right)$$

Portanto, a probabilidade de que X esteja entre 2 e 5 é igual à probabilidade de que Z esteja entre $-0,25$ e $0,5$. Utilizando o Anexo 1, têm-se que

$$P(-0,25 \leq Z \leq 0,5) = 0,0987 + 0,1915 = 0,2902,$$

ou seja,

$$P(2 \leq X \leq 5) = 0,2902.$$

Com essas abordagens é possível formar uma base para o entendimento do formalismo das Redes Bayesianas que se fundamenta no raciocínio probabilístico. O raciocínio probabilístico é baseado na realização de inferências probabilísticas que é o cálculo da probabilidade de um evento, dada toda evidência possível. Este cálculo é baseado em probabilidade condicional e no teorema de Bayes. Esta abordagem foi utilizada neste capítulo para fundamentar os conceitos probabilísticos.

CAPÍTULO 4

4 CLASSIFICADORES BAYESIANOS

Este capítulo aborda duas técnicas básicas de classificação de dados, Algoritmo “Naive Bayes” e Redes Bayesianas.

A classificação de dados é um processo, dividido em duas etapas [23]. Na primeira, um modelo é construído, descrevendo um conjunto de dados, através da análise das instâncias que possuem seus atributos. O conjunto de dados pode estar armazenado em um banco de dados, arquivo texto, etc. Este conjunto é utilizado para gerar o modelo do classificador e é chamado de conjunto de dados de treinamento[23]. Vale ressaltar que o conjunto de dados de treinamento pode ser retirado aleatoriamente de uma população.

Quando as instâncias possuem uma classe, ou seja, um rótulo para os valores de seus atributos, tem-se um aprendizado supervisionado. Por outro lado, pode-se ter um conjunto de dados que não se conhece, antecipadamente, a classe na qual as instâncias pertencem. Este outro modelo é chamado de aprendizado não supervisionado ou “*clustering*”. Nesta dissertação será considerado o aprendizado supervisionado.

Usualmente, o modelo de aprendizado é representado na forma de regras de classificação, árvores de decisão, ou de um modelo matemático [23]. Por exemplo, dado uma base de dados que retrata as informações sobre o crédito dos clientes, regras de classificação podem ser aprendidas para identificar clientes com boas referências de crédito ou não. Estas regras extraídas podem ser utilizadas para

avaliação de futuros clientes ou mesmo para entender os dados presentes. A figura 4.1 mostra um exemplo:

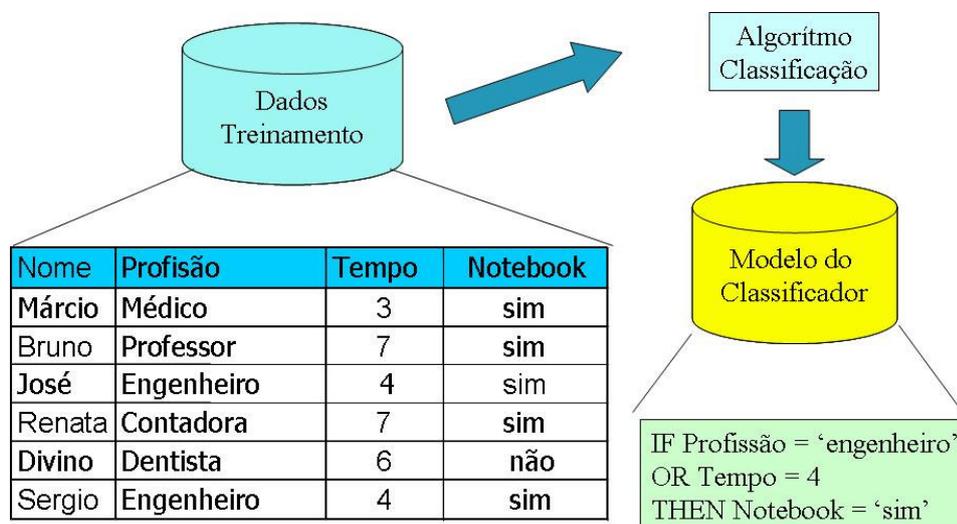


Figura 4.1 - Construção do modelo de aprendizagem.

A figura 4.2 mostra o classificador com seu modelo elaborado, e realizando uma classificação de uma nova instância.

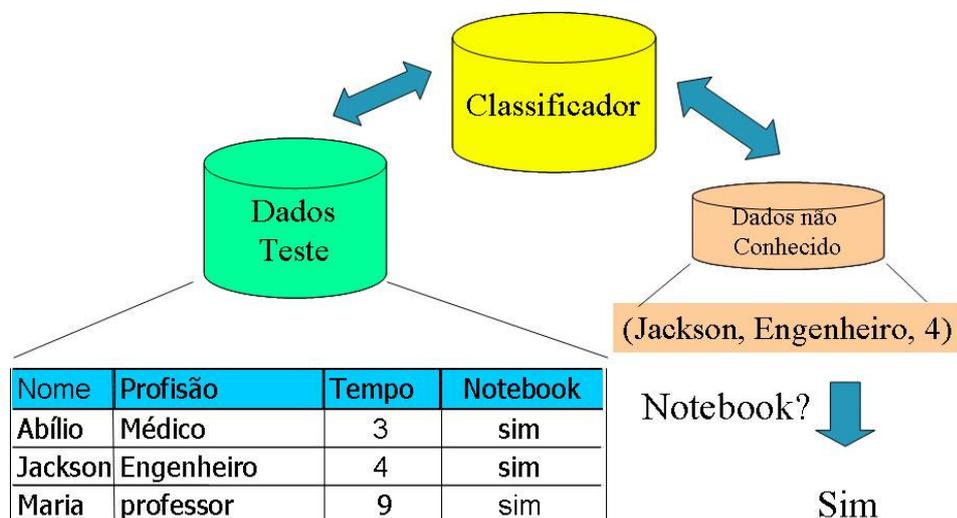


Figura 4.2 - Classificação de uma nova Instância.

O estudo realizado neste capítulo está relacionado com classificadores que utilizam modelos matemáticos, mais precisamente estatísticos, para realizar a classificação de novas instâncias.

Classificadores Bayesianos são entendidos como classificadores estatísticos que podem adivinhar se um padrão de teste pertence a uma determinada classe ou não.

Os classificadores Bayesianos são baseados no teorema de Bayes, que será explanado posteriormente. Estudos realizados com a finalidade de comparar os algoritmos de classificação mostram que existe um classificador bayesiano simples conhecido como classificador “Naive Bayes” que pode ser comparado em desempenho com Árvores de Decisão e Redes Neurais Artificiais – RNA[23].

Os classificadores “Naive Bayes” assumem que o efeito do valor de um atributo em uma dada classe é independente dos valores dos outros atributos. Esta consideração é chamada de independência condicional. Isto é adotado para facilitar os cálculos envolvidos, e por este motivo, recebe o nome de Naive, ou ingênuo [23].

4.1 Teorema de Bayes

Considerando que X é uma amostra de um dado, o qual a classe não é conhecida. Considerando que H é uma hipótese, na qual a amostra X pertence a uma classe C . Para problemas de classificação, será determinada $P(H|X)$, a probabilidade de se ter a hipótese H , tendo observado um dado de amostra X .

$P(H|X)$ é a probabilidade a posteriori, de H condicionado a X . Por exemplo, supondo que tratam-se de frutas, ou seja, que as amostras consistem de frutas, descritas por suas cores e formas. $P(H|X)$ mostra a confiança que X é uma maçã, dado que X é vermelha e redonda. $P(H)$ é a probabilidade a priori. Para o exemplo citado isto é a probabilidade que uma dada amostra é uma maçã. A

probabilidade a posteriori, $P(H|X)$, é baseada em mais informação do que a probabilidade a priori, $P(H)$, o qual é independente de X .

Similarmente, $P(X|H)$ é a probabilidade a posteriori de X condicionada em H , isto é, a probabilidade de X ser vermelha e redonda dado que sabê-se X é uma maçã. $P(X)$ é a probabilidade a priori de X . Usando o exemplo, esta é a probabilidade que a amostra do conjunto de frutas é vermelha e redonda.

Estas probabilidades são estimadas da seguinte forma: $P(X)$, $P(H)$, e $P(X|H)$ pode ser estimado dos dados presentes no conjunto de treinamento, como pode ser visto abaixo. O teorema de Bayes é descrito por:

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)}$$

4.2 Classificador Naive Bayes

O classificador Naive Bayes, ou classificador Bayesiano, funciona como descrito abaixo.

Cada dado de amostra, ou dado de treinamento é representado por um vetor de dimensão n , $X = (x_1, x_2, \dots, x_n)$ e cada atributo é representado por A_1, A_2, \dots, A_k .

Supondo que existam m classes, C_1, C_2, \dots, C_m . Dado uma amostra, X (exemplo, sem classe), o classificador irá nos informar que X pertence a classe que tem a maior probabilidade a posteriori, condicionada em X . Isto é, o classificador Naive Bayes associa uma amostra não conhecida X para uma classe C_i , se, e somente se,

$$P(C_i | X) > P(C_j | X) \text{ para } 1 \leq j \leq m, j \neq i$$

Então, maximiza-se $P(C_i|X)$. A classe C_i para a qual $P(C_i|X)$ é maximizado é chamado da hipótese máxima posteriori.

Como $P(X)$ é constante para todas as classes, somente $P(X|C_i)P(C_i)$ precisa ser maximizado. No aprendizado não supervisionado, onde a probabilidade a priori da classe C_i , $P(C_i)$, não é conhecida, então isto é assumido que as probabilidades das classes são iguais, isto é, $P(C_1) = P(C_2) = \dots = P(C_m)$, e aí tem-se que maximizar $P(X|C_i)$.

No aprendizado supervisionado a probabilidade a priori pode ser estimada por $P(C_i) = \frac{S_i}{S}$, onde S_i é o número de amostras que possuem a classe C_i , e S é o total de amostras utilizadas para o treinamento.

Dado o conjunto de dados com muitos atributos, tem-se um tempo computacional elevado para calcular $P(X|C_i)$. No sentido de reduzir este tempo computacional para o cálculo de $P(X|C_i)$, o “Naive Bayes” assume a independência condicional da classe. Com isto presume-se que os atributos são condicionalmente independentes uns dos outros, dado a classe da amostra, isto é, não há relação de dependência entre os atributos. Então:

$$P(X | C_i) = \prod_{k=1}^n P(X_k | C_i)$$

As probabilidades $P(x_1|C_i)$, $P(x_2|C_i)$, ..., $P(x_n|C_i)$ podem ser estimadas através do conjunto de treinamento, onde:

a) se A_k é categórico, então $P(x_k | C_i) = \frac{S_{ik}}{S_i}$, onde S_{ik} é o número de

amostras do conjunto de treinamento que possui a classe C_i assumindo o

valor x_k para “ A_k ”, e “ S_i ” é o número de amostras para treinamento que pertence a C_i .

- b) se A_k é uma variável contínua, então é considerado que o atributo tipicamente assume ter uma distribuição Gaussiana ou Normal. Então tem-se:

$$P(X | C_i) = g(x_k, \mu_{ci}, \sigma_{ci}) = \frac{1}{\sqrt{2\pi\sigma_{ci}}} e^{-\frac{(x_k - \mu_{ci})^2}{2\sigma_{ci}^2}}$$

Onde $g(x_k, \mu_{ci}, \sigma_{ci})$ é a função de distribuição Gaussiana(Normal) para o atributo A_k , enquanto μ_{ci} e σ_{ci} , é a média e o desvio padrão, respectivamente, dados pelos valores dos atributos A_k , para o conjunto de treinamento da classe C_i .

Então para realizar a classificação de um amostra não conhecida X , $P(X | C_i)P(C_i) > P(X | C_j)P(C_j)$ para $1 \leq j \leq m, j \neq i$.

Em outras palavras, o conjunto de teste é associado para a classe C_i para o qual $P(X | C_i)P(C_i)$ é máximo.

Para facilitar o entendimento do funcionamento do algoritmo será considerado um processo industrial, onde têm-se o problema de previsão da qualidade da Cal de um forno de calcinação. A Cal é um derivado do carbonato de cálcio, a qual possui diversos parâmetros para avaliar a sua qualidade.

No geral, as instâncias no conjunto de dados são caracterizadas pelos valores dos seus atributos, os quais nos informam diferentes aspectos das instâncias. Neste exemplo, do problema da previsão da qualidade da cal, existem 4 atributos que são “Cor”, “Reatividade”, “Granulometria”, “Flúor Baixo” e o objetivo final que é “Qualidade”. Então com esta simples base de dados, mostrada na tabela 4.1, todos os atributos possuem valores que são simbólicos e não numéricos. O

As células na primeira linha da tabela 4.2 simplesmente contam as ocorrências de todas as possibilidades de todos os possíveis valores de cada atributo, e o atributo “Qualidade” aparece na última coluna e conta o número de ocorrências de “Sim” e “Não”. Na parte de baixo da tabela, a mesma informação é reescrita na forma de frações, ou probabilidades observadas. Por exemplo, de nove instâncias que “Qualidade” é “Sim”, o atributo “Cor” é “Boa” por duas vezes, portanto uma fração de 2/9. Para “Qualidade” a fração é diferente: ela é a proporção das instâncias que “Qualidade” é “Sim” e “Não” respectivamente.

Com os dados da tabela 4.2 pode-se considerar um modelo matemático baseado em probabilidades que representam o problema ou domínio, capaz de informar se a cal terá uma boa qualidade. A fim de testar este novo modelo de aprendizagem será considerada uma nova instância, ou seja, um novo conjunto de dados que não esteja na tabela 4.1. Com isto tem-se o intuito de validar o modelo. A tabela 4.3 mostra essa nova instância.

Tabela 4.3 - Instância de teste utilizado para avaliar modelo probabilístico.

Cor	Reatividade	Granulometria	Flúor Baixo	Qualidade
Boa	Baixa	Alta	Sim	?

Os cinco atributos da tabela 4.3 – “Cor”, “Reatividade”, “Granulometria”, “Flúor Baixo” e “Qualidade” – são tratados com a mesma importância, independente das evidências. Para o atributo “Qualidade” dessa nova instância, receber o valor “Sim”, tem-se a seguinte probabilidade:

$$Pr_{obabilidade_Sim} = \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14} = 0,0053$$

As frações são tomadas dos atributos com valores “Sim” na tabela, de acordo com os valores dos atributos de nova cal, e o final, 9/14 significa as várias representações das proporções das instâncias as quais “Qualidade” é “Sim”. Um cálculo igual pode ser realizado caso queira descobrir se o atributo “Qualidade” receberá o valor “Não”.

$$Pr obabilidade_ Não = \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{14} = 0,0206$$

Isto indica que a nova instância, tabela 4.3, terá maior probabilidade de que a qualidade da cal será ruim.

As probabilidades calculadas acima podem ser normalizadas, como mostrado abaixo.

$$Pr obabilidade_ Sim = \frac{0,0053}{0,0053 + 0,0206} = 20,5\%$$

$$Pr obabilidade_ Não = \frac{0,0206}{0,0053 + 0,0206} = 79,5\%$$

Este simples e intuitivo método é baseado na regra de Bayes da probabilidade condicional. A regra de Bayes diz que se existe uma hipótese H, e uma evidência E, a qual deu origem a hipótese, então:

$$P(H | E) = \frac{P(E | H)P(H)}{P(E)}$$

Utiliza-se a notação P(H) para mostrar a probabilidade de um evento H e P(H|E) para mostrar a probabilidade condicional de H quando um evento E ocorrer. A hipótese H é “Qualidade” com um valor igual a “Sim” e P(H|E) será 20,5 %. A evidência E é uma combinação particular dos valores dos atributos para uma nova instância, Cor = Boa, Reatividade = Baixa, Granulometria = alta e Flúor Baixo = Sim. Estas quatro evidências podem ser chamadas de E1, E2, E3 e E4, respectivamente.

Assumindo que estes pedaços de evidência são independentes, sua probabilidade combinada é obtida multiplicando as probabilidades, então:

$$P(\text{Sim} | E) = \frac{P(E1 | \text{Sim}).P(E2 | \text{Sim}).P(E3 | \text{Sim}).P(E4 | \text{Sim}).P(\text{Sim})}{P(E)}$$

Não será considerada a $P(E)$, pois a mesma será eliminada na etapa de normalização, quando a soma das probabilidades de “Sim” e “Não” for igual a 1, como realizado acima. A $P(\text{Sim})$ no final é a probabilidade de um “Sim” ser o resultado sem levar em consideração nenhuma das evidências E , isto é, sem saber nada sobre a nova instância. Isto é a chamada probabilidade a priori de uma hipótese H . De acordo com o exemplo, isto é $9/14$, porque 9 de 14 exemplos utilizados no treinamento possuem o valor “Sim” para o atributo “Qualidade”.

Substituindo os valores da tabela 4.2 para sua evidência apropriada temos:

$$P(\text{Sim} | E) = \frac{\frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14}}{P(E)}$$

O método utilizado neste exemplo é chamado de “Naive Bayes”, porque ele é baseado na Regra de Bayes. Pode-se multiplicar estas probabilidades haja vista elas serem independentes.

No entanto, um possível erro com “Naive Bayes” é se um valor particular de um atributo não aparecer no conjunto de treinamento com os valores das classes. Supondo que no conjunto de dados para treinamento, o valor do atributo “Cor = Boa” tivesse sido associado com o resultado “Não”. Então a probabilidade de “Sim” dado “Cor = Boa”, $P(\text{Sim} | \text{Cor} = \text{Boa})$, seria zero, e as outras probabilidades, quando multiplicadas por ela, apresentariam resultado final zero, não importando as outras

probabilidades. Mas este problema pode ser facilmente superado, realizando pequenos ajustes no método de cálculo das probabilidades.

Uma característica realmente interessante sobre a formulação Bayesiana é que atributos que não possuem valores não são problemas. Considerando a tabela 4.4.

Tabela 4.4 - Instância de teste sem valores para atributos.

Cor	Reatividade	Granulometria	Flúor Baixo	Qualidade
?	Baixa	Alta	Sim	?

Pode-se observar que o valor do atributo “Cor” é desconhecido, e o cálculo simplesmente omite este atributo, como mostrado abaixo:

$$Pr obabilidade_Sim = \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14} = 0,0238$$

$$Pr obabilidade_Não = \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{14} = 0,0343$$

Nota-se que os valores calculados da Probabilidade_Sim e a Probabilidade_Não são maiores do que os valores calculados anteriormente. Isto se deve ao fato de se ter retirado um dos atributos. Mas isto não irá interferir no resultado porque o mesmo não foi considerado nos dois casos. Normalizando tem-se Probabilidade_Sim e Probabilidade_Não iguais a 41% e 59% , respectivamente.

Se o valor de um atributo não aparece em uma instância de treinamento, o algoritmo simplesmente não o inclui na contagem de frequência, e a razão da probabilidade é baseada no número de valores que ocorrem e não no número total de instâncias.

Quando os valores dos atributos são considerados grandezas numéricas, assume-se uma distribuição normal ou “Gaussiana”. Abaixo, a tabela 4.5 mostra um resumo dos dados da previsão da Cal, mas agora com atributos numéricos:

Tabela 4.5 – Resumo dos dados de previsão da qualidade da Cal, com valores numéricos.

	Cor		Reatividade		Granulometria(mm)		Flúor Baixo		Qualidade		
	<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Não</i>	
Boa	2	3	64, 68,	65, 71,	65, 70,	70, 85,	Falso	6	2	9	5
Média	4	0	69, 70,	72, 80,	70, 75,	90, 91,	Verdade	3	3		
Ruim	3	2	72, ...	85, ...	80, ...	95, ...					
Boa	2/9	3/5	$\mu = 73$	$\mu = 75$	$\mu = 79$	$\mu = 86$	Falso	6/9	2/5	9/14	5/14
Média	4/9	0/5	$\sigma = 6.2$	$\sigma = 7.9$	$\sigma = 10.2$	$\sigma = 9.7$	Verdade	3/9	3/5		
Ruim	3/9	2/5									

Quando os atributos são nominais, as probabilidades são calculadas como descrito anteriormente. Já quando os atributos são numéricos, a média e o desvio padrão são calculados para cada classe e para cada atributo. Então, média, variância e desvio padrão do atributo “Reatividade” para as instâncias que possuem o atributo “Qualidade” igual a “Sim” são respectivamente:

$$\mu = \frac{83 + 70 + 68 + 64 + 69 + 75 + 75 + 72 + 81}{9} = 73$$

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N - 1} = \frac{(83 - 73)^2 + (70 - 73)^2 + (68 - 73)^2 + (64 - 73)^2 + \dots + (81 - 73)^2}{(9 - 1)} = 38,44$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x - \mu)^2}{N - 1}} = 6,2$$

A função de densidade de probabilidade para uma distribuição normal, onde a média e o desvio padrão são respectivamente iguais a μ, σ será:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} e^{-\frac{(x - \mu)^2}{2 \cdot \sigma^2}}$$

Para a tabela 4.5, tem-se

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 6,2} e^{-\frac{(x-73)^2}{2 \cdot (6,2)^2}}$$

A probabilidade da cal ter boa qualidade com uma reatividade for 66 será

$$f(\text{reatividade} = 66 | \text{Sim}) = \frac{1}{\sqrt{2\pi} \cdot 6,2} e^{-\frac{(66-73)^2}{2 \cdot (6,2)^2}} = 0,0340$$

E utilizando o mesmo raciocínio, deve-se criar densidade de probabilidade para representar o atributo “granulometria”, como observado abaixo.

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 10,2} e^{-\frac{(x-79,1)^2}{2 \cdot (10,2)^2}}$$

A probabilidade da cal possuir boa qualidade se a granulometria for 90 será

$$f(\text{granulometria} = 90 | \text{Sim}) = \frac{1}{\sqrt{2\pi} \cdot 10,2} e^{-\frac{(90-79,1)^2}{2 \cdot (10,2)^2}} = 0,0221$$

A função de densidade de probabilidade para um evento está relacionada com a sua probabilidade. O real significado da função de densidade $f(x)$ é que a probabilidade de uma quantidade em um pequena região em torno de x , ou seja, entre $x - \frac{\varepsilon}{2}$ e $x + \frac{\varepsilon}{2}$, é $\varepsilon \cdot f(x)$.

Usando estas probabilidades para uma novo tipo de cal, tabela 4.6, tem-se:

Tabela 4.6 – Intância de teste com atributos numéricos.

Cor	Reatividade	Granulometria	Flúor Baixo	Qualidade
Boa	66	90	Sim	?

$$\text{Probabilidade da "Qualidade = Sim"} = \frac{2}{9} \cdot 0,0340 \cdot 0,0221 \cdot \frac{3}{9} \cdot \frac{9}{14} = 0,000036$$

$$\text{Probabilidade da "Qualidade = Não"} = \frac{3}{5} \cdot 0,0291 \cdot 0,0380 \cdot \frac{3}{5} \cdot \frac{5}{14} = 0,000136$$

As probabilidades calculadas acima podem ser normalizadas, como mostrado abaixo.

$$\text{Pr obabilidade "Qualidade = Sim"} = \frac{0,000036}{0,000036 + 0,000136} = 20,9\%$$

$$\text{Pr obabilidade "Qualidade = Não"} = \frac{0,000136}{0,000036 + 0,000136} = 79,1\%$$

Estes valores de probabilidades são muito próximos das probabilidades calculadas anteriormente para a cal com as especificações da tabela 4.5 porque os valores de reatividade e Flúor_Baixo de 66 e 90, respectivamente, são similares com as probabilidades, fria e alta, que usou-se anteriormente.

Se os valores dos atributos numéricos não aparecem no conjunto de treinamento, a média e o desvio padrão são baseados somente nos atributos que estão presentes no conjunto de treinamento.

4.3 Redes de Crenças Bayesiana

Como visto no item 4.2 o classificador "Naive Bayes" tem a suposição que as classes são condicionalmente independentes, isto é, dado um exemplo e a classe que o mesmo pertence, os valores dos atributos são condicionalmente independentes um dos outros. Com esta suposição os cálculos são facilitados. Na prática, podem existir dependências entre os atributos. As redes de crenças Bayesiana especificam a união da distribuição de densidade condicional. Elas permitem que a independência de classes condicionais possa ser definida entre sub conjuntos de variáveis, e ainda, fornecem um modelo gráfico de relacionamento

causal, no qual o aprendizado pode ser realizado. Estas redes são conhecidas também como redes de crenças, redes Bayesianas e redes probabilísticas.

Uma rede de crença é definida por dois componentes. O primeiro é um gráfico direcionado acíclico, onde cada nó representa uma variável aleatória e cada arco representa uma dependência de probabilidade. Se o arco é desenhado de um nó Y para um nó Z, então Y é o pai ou precessor de Z, e Z é o descendente de Y. Cada variável é condicionalmente independente, desde de que seu nó não seja descendente de outros nós. As variáveis podem ser discretas ou contínuas. A figura 4.3 nos mostra uma rede bayesiana simples.

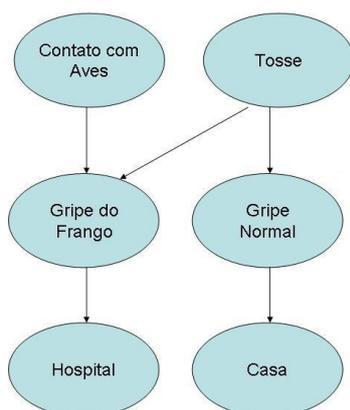


Figura 4.3 - Rede Bayesiana simples.

Os arcos indicam uma relação de conhecimento causal. Por exemplo, para se ter gripe do frango um dos fatores que deve ser levado em conta, é se o indivíduo tem contato com aves e/ou o mesmo indivíduo apresenta tosse. Pela figura 4.3 pode-se notar que os arcos mostram que a variável “Gripe do Frango” é condicionalmente independente da “Gripe Normal”, dado seus nós-pais que são “Contato com Aves” e “Tosse”. Isto significa que conhecido os valores de “Contato com Aves” e “Tosse”, a variável “Gripe Normal” não acrescenta nenhuma informação em relação a “Gripe do Frango”.

A segunda definição de uma rede de crenças consiste de uma tabela de probabilidade condicional (TPC) para cada variável. A TPC para uma variável Z especifica a distribuição condicional $P(Z | \text{Pais}(Z))$, onde $\text{Pais}(Z)$ são os pais de Z . A tabela 4.7 mostra a TPC para “Gripe do Frango”.

Tabela 4.7 – Tabela de Probabilidade Condicional - TPC.

	CA,T	CA,~T	~CA,T	~CA,~T
GF	0,8	0,5	0,7	0,1
~GF	0,2	0,5	0,3	0,9

Então a probabilidade condicional para cada valor de “Gripe do Frango” é dada para cada combinação possível dos valores de seus nós-pais. Para exemplificar, do lado superior esquerdo e no lado inferior esquerdo, tem-se:

$$P(\text{Gripe do Frango} = \text{“Sim”} | \text{Contato com Aves} = \text{“Sim”}, \text{Tosse} = \text{“Sim”}) = 0,8$$

$$P(\text{Gripe do Frango} = \text{“Não”} | \text{Contato com Aves} = \text{“Sim”}, \text{Tosse} = \text{“Sim”}) = 0,2$$

A união das probabilidades relativas as variáveis ou atributos Z_1, \dots, Z_n é calculado por

$$P(z_1, \dots, z_n) = \prod_{i=1}^n P(z_i | \text{Pais}(Z_i)),$$

Onde os valores para $P(z_i | \text{Pais}(Z_i))$ corresponde as valores na TCP para Z_i .

Um nó da rede pode ser selecionado como um nó de saída, representando um atributo que será o atributo classe. Pode-se ter mais de um nó de saída. Para realizar o processo de inferência existem os algoritmos que realizam esta função.

Esses algoritmos não serão abordados neste trabalho porque o nosso principal interesse é o “Naive Bayes”, como mencionado no início deste capítulo.

CAPÍTULO 5

5 PACOTE COMPUTACIONAL WEKA

Este capítulo tem o objetivo de realizar uma descrição da ferramenta computacional utilizada para desenvolver esta dissertação. A ferramenta computacional foi desenvolvida pela Universidade de Waikato, Nova Zelândia, e recebe o nome de Weka. Este projeto recebeu este nome em homenagem a um pequeno pássaro típico da Nova Zelândia, e que entre diversas características não possui a habilidade de voar.

O pacote computacional é desenvolvido em Java, uma linguagem de programação orientada a objetos, que é largamente disponível para os principais sistemas operacionais como Linux, Windows e Macintosh.

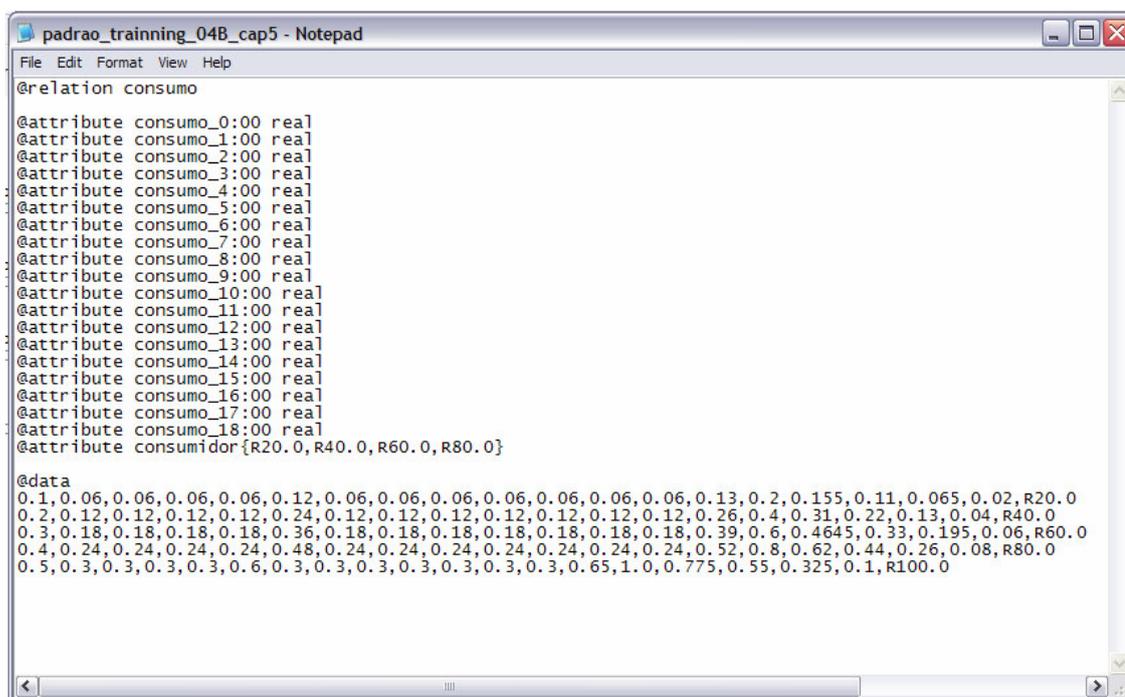
5.1 Descrição

Pode-se utilizar o Weka de várias maneiras possíveis. A primeira delas, é a utilização dos algoritmos de aprendizagem de máquina, que podem ser aplicados em um conjunto de dados através de uma linha de comando. E através desta pode-se pré-processar o conjunto de dados, aplicar a um conjunto de dados um algoritmo de aprendizagem, analisar o resultado e o desempenho do classificador – tudo isto sem escrever nenhuma linha de código em Java.

Uma outra maneira de se utilizar este pacote, é aproveitar parte do seu código em Java para facilitar a implementação de sua própria aplicação. Esta ajuda

é muito valiosa, porque existem muitos filtros e algoritmos de classificação implementados, bastando utilizar as classes do projeto Weka.

Uma situação comum é um conjunto de dados ser armazenado em um banco de dados ou em uma planilha. Mas estes formatos não são interpretados pelo Weka. O Weka trata os dados em um formato específico chamado ARFF. Isto porque é necessário ter informação a respeito do tipo de cada atributo, haja vista que o Weka não pode identificar os valores de cada atributo sem que seus tipos estejam especificados. Então, antes de aplicar qualquer algoritmo a um conjunto de dados, os mesmos tem que ser convertidos para o formato ARFF, que pode ser realizada facilmente. Um arquivo ARFF é um arquivo texto com a lista de todas as intâncias, com os valores de cada uma sendo separados por vírgulas. O formato desse arquivo pode ser visto na figura 5.1.



```
@relation consumo

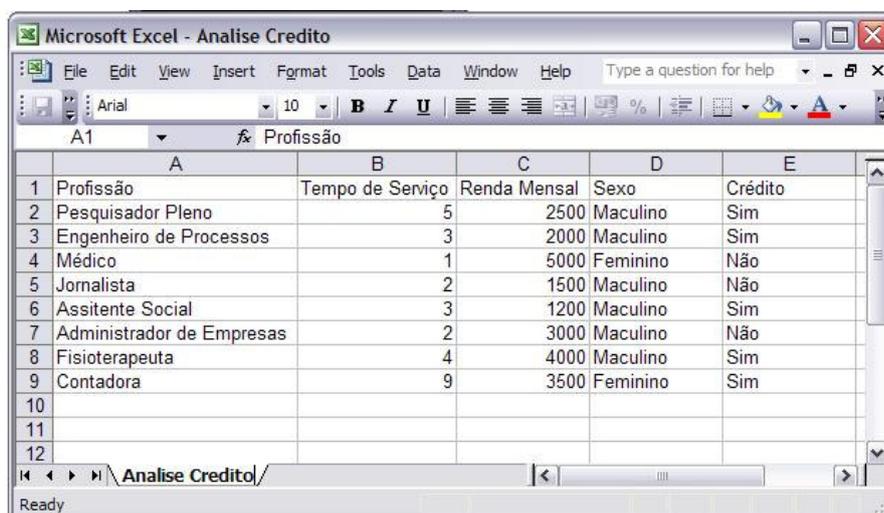
@attribute consumo_0:00 real
@attribute consumo_1:00 real
@attribute consumo_2:00 real
@attribute consumo_3:00 real
@attribute consumo_4:00 real
@attribute consumo_5:00 real
@attribute consumo_6:00 real
@attribute consumo_7:00 real
@attribute consumo_8:00 real
@attribute consumo_9:00 real
@attribute consumo_10:00 real
@attribute consumo_11:00 real
@attribute consumo_12:00 real
@attribute consumo_13:00 real
@attribute consumo_14:00 real
@attribute consumo_15:00 real
@attribute consumo_16:00 real
@attribute consumo_17:00 real
@attribute consumo_18:00 real
@attribute consumidor {R20.0, R40.0, R60.0, R80.0}

@data
0.1,0.06,0.06,0.06,0.06,0.12,0.06,0.06,0.06,0.06,0.06,0.06,0.13,0.2,0.155,0.11,0.065,0.02,R20.0
0.2,0.12,0.12,0.12,0.12,0.24,0.12,0.12,0.12,0.12,0.12,0.12,0.26,0.4,0.31,0.22,0.13,0.04,R40.0
0.3,0.18,0.18,0.18,0.18,0.36,0.18,0.18,0.18,0.18,0.18,0.18,0.39,0.6,0.4645,0.33,0.195,0.06,R60.0
0.4,0.24,0.24,0.24,0.24,0.48,0.24,0.24,0.24,0.24,0.24,0.24,0.52,0.8,0.62,0.44,0.26,0.08,R80.0
0.5,0.3,0.3,0.3,0.3,0.6,0.3,0.3,0.3,0.3,0.3,0.3,0.65,1.0,0.775,0.55,0.325,0.1,R100.0
```

Figura 5.1 - Formato de um arquivo ARFF.

A maioria das planilhas e bancos de dados permite exportar os dados em um arquivo separado por vírgulas, como uma lista de itens. Uma vez que o passo acima foi realizado, precisa-se carregar o arquivo em um editor de texto ou um processador de texto, adicionar as seguintes informações: o nome do conjunto de dados usando a etiqueta @relation, a informação dos atributos usando @attribute, e uma linha @data e salvar o arquivo como texto bruto.

Para facilitar o entendimento, segue um exemplo. Assume-se que os dados estão inicialmente armazenados em uma planilha Excel e se está usando o Microsoft Word como processador de texto. Apesar da opção por utilizar estes softwares da Microsoft, se fossem utilizados outros, o procedimento seria o mesmo adotado neste exemplo. A figura 5.2 mostra a planilha Excel contendo um conjunto de dados que possui o nome "Análise Crédito".



	A	B	C	D	E
1	Profissão	Tempo de Serviço	Renda Mensal	Sexo	Crédito
2	Pesquisador Pleno	5	2500	Maculino	Sim
3	Engenheiro de Processos	3	2000	Maculino	Sim
4	Médico	1	5000	Feminino	Não
5	Jornalista	2	1500	Maculino	Não
6	Assistente Social	3	1200	Maculino	Sim
7	Administrador de Empresas	2	3000	Maculino	Não
8	Fisioterapeuta	4	4000	Maculino	Sim
9	Contadora	9	3500	Feminino	Sim
10					
11					
12					

Figura 5.2 - Dados do arquivo "Análise crédito" em uma planilha do Excel.

O próximo passo é salvar os dados separando-os por vírgulas. Seleciona-se o item "Salvar como..." do menu "Arquivo". Então, na caixa de diálogo seleciona-se CSV (Comma Delimited) , digita-se com o nome do arquivo e clica-se no botão

“Salvar”.(Uma mensagem irá alertar que irá salvar somente a planilha ativa: ignore-se isto e clica-se em OK.)

Agora, pode-se abrir o arquivo no Microsoft Word. Tem-se o arquivo como indicado na figura 5.3.

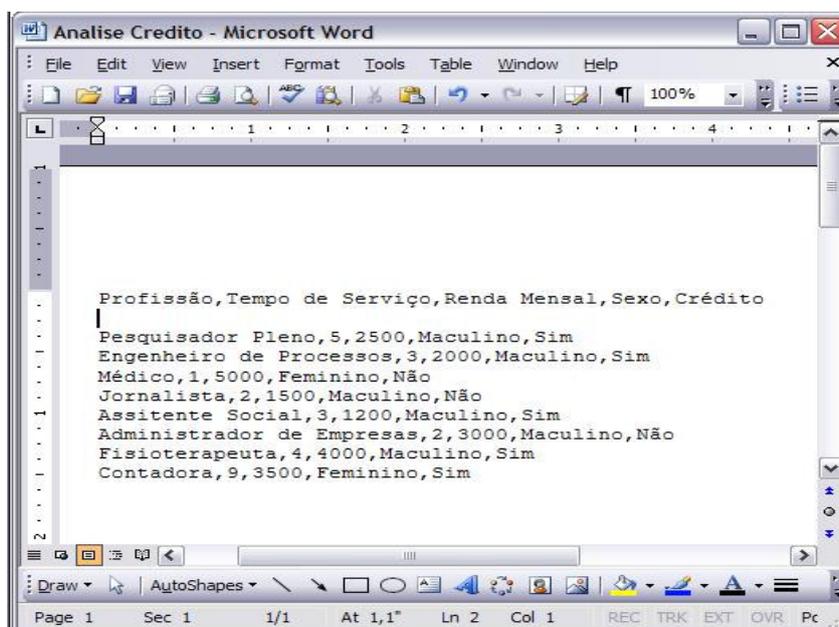
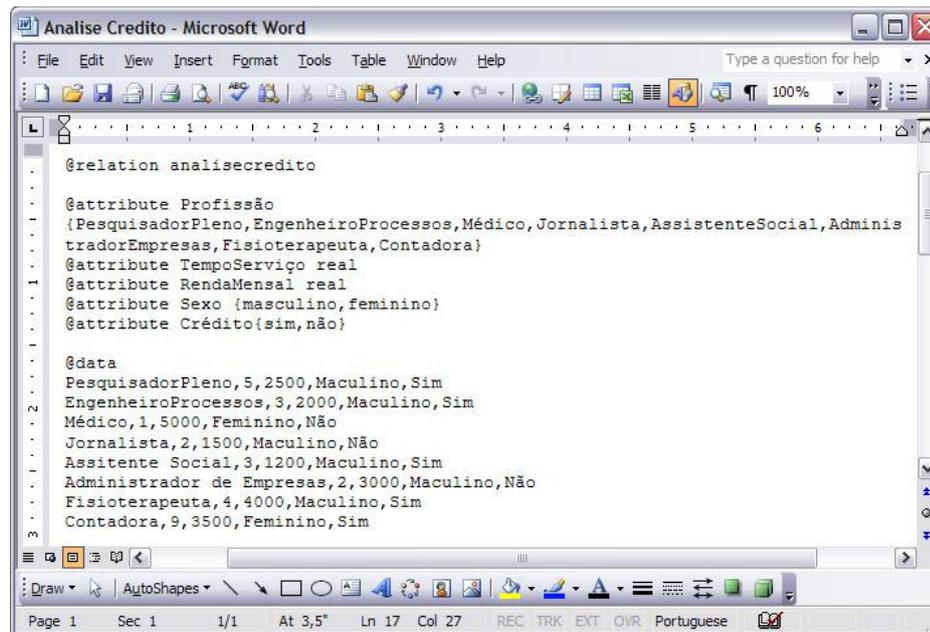


Figura 5.3 - Dados do arquivo “análise de crédito” separados por vírgula.

Observa-se que as linhas da planilha original foram convertidas em linhas de texto, e os elementos são separados uns dos outros através de vírgulas.

O que indica o início do arquivo ARFF é um cabeçalho contendo o nome dos atributos do conjunto de dados. A figura 5.4 mostra o arquivo ARFF.



```

@relation analisecredito

@attribute Profissao
{PesquisadorPleno,EngenheiroProcessos,Médico,Jornalista,AssistenteSocial,Adminis
tradorEmpresas,Fisioterapeuta,Contadora}
@attribute TempoServiço real
@attribute RendaMensal real
@attribute Sexo {masculino,feminino}
@attribute Crédito{sim,não}

@data
PesquisadorPleno,5,2500,Maculino,Sim
EngenheiroProcessos,3,2000,Maculino,Sim
Médico,1,5000,Feminino,Não
Jornalista,2,1500,Maculino,Não
Assistente Social,3,1200,Maculino,Sim
Administrador de Empresas,2,3000,Maculino,Não
Fisioterapeuta,4,4000,Maculino,Sim
Contadora,9,3500,Feminino,Sim

```

Figura 5.4 - Arquivo “Analise Credito” no formato ARFF.

O nome do conjunto de dados é introduzido pela etiqueta “@relation”, e, os nomes, tipos, e valores de cada atributo são definidos pela etiqueta “@attribute”. A seção de dados do arquivo ARFF começa com a etiqueta “data”. Uma vez que estes procedimentos tenham sido seguidos corretamente, salva-se o arquivo com a extensão “.ARFF”. No exemplo dado acima, o arquivo foi gravado como “analisecredito.arff”. Um fator importante que devemos saber é que a classe de cada instância, por padrão, é o último atributo. Na figura 5.4, por exemplo, a classe das instâncias é “credito” e pode assumir dois valores que são “sim” e “não” .

5.2 Bibliotecas Weka

Para se entender como a estrutura do weka é organizada, deve-se saber como a estrutura dos programas em Java são organizados.

5.2.1 Classes, instâncias e pacotes

Todo programa em Java é implementado como uma classe. Na programação orientada a objetos, uma classe é uma coleção de variáveis com métodos que as modifiquem. Juntos, eles definem o comportamento de um objeto que pertence a uma classe. Um objeto é simplesmente uma instanciação de uma classe que têm valores associados a cada variável. Em Java, um objeto é também chamado de instância de uma classe.

No projeto Weka, a implementação de um determinado algoritmo de aprendizagem é representado por uma classe.

Programas maiores são divididos em mais de uma classe. A classe Naive Bayes não contém todo o código necessário para gerar o modelo de classificação. Ela contém referências para outras classes que realizam outras funções para auxiliar no processo de classificação. Quando existe um número muito grande de classes, como na Weka, pode se tornar difícil entender e se orientar no código. A linguagem de programação Java fornece a opção de organizar as classes em pacotes. Um pacote é simplesmente um diretório contendo uma coleção de classes que possuem alguma relação. Portanto, o Weka é organizado em pacotes e abaixo descreveremos alguns deles.

Informações detalhadas sobre Weka podem ser obtidos na documentação “online” disponível [25]. Esta documentação foi gerada diretamente dos comentários do código fonte usando o utilitário Javadoc da Sun.

5.2.2 O pacote weka.core

O “core” é o pacote principal do sistema Weka. Este pacote contém classes que são utilizadas por quase todas as outras classes. Na figura 5.5, pode-se ver o índice dos pacotes e explorá-los clicando nos devidos links.



Figura 5.5 - Índice com os pacotes disponíveis no sistema Weka.

Abaixo mostra o pacote “core”



Figura 5.6 - Índice das interfaces e classes pertencentes ao pacote core

A figura 5.6 é dividida em duas partes: interfaces e classes. As principais classes neste pacote são `Attribute`, `Instance` e `Instances`. Um objeto da classe `Attribute` representa um atributo. O mesmo contém o nome do atributo, o seu tipo, e no caso de atributos nominais os valores possíveis. Um objeto da classe `Instance` contém os valores dos atributos de uma determinada instância; e um objeto da classe `Instances` possui um conjunto de instâncias, ou seja, o conjunto de dados. Clicando em cima dessas classes pode-se obter mais informações a respeito das mesmas. Quando se está executando os algoritmos através da linha de comando, não é necessário saber sobre essas classes. Mas quando se tem que desenvolver a sua própria aplicação, as informações sobre estas classes se tornam essenciais.

5.2.3 O pacote `weka.classifiers`

O pacote `classifiers` possui a maioria dos algoritmos de classificação. A classe mais importante desse pacote é a `Classifier`, a qual define a estrutura geral de vários algoritmos de classificação e predição numérica. Ela contém dois métodos `buildClassifier()` e `classifyInstance()`, que são utilizados por todos os algoritmos de aprendizagem. Na programação orientada a objetos, os algoritmos de aprendizagem são representados por subclasses do `Classifier`, e conseqüentemente herdam esses dois métodos.

Outra classe importante é `DistributionClassifier`. Esta subclasse da `Classifier` define o método `distributionForInstance()`, o qual retorna uma distribuição de probabilidade para uma dada instância. Vários classificadores que podem calcular a probabilidade de uma classe é uma subclasse da `DistributionClassifier` e implementa este método.

CAPÍTULO 6

6 ESTUDO DE CASO

Uma subestação do sistema elétrico alimenta diversos tipos de carga: residenciais, comerciais, públicas e industriais. Cada carga tem um comportamento típico, como uma “assinatura”, que é refletida em seu consumo de energia ao longo do dia.

Do ponto de vista da subestação, a energia entregue ao cliente (consumo total), não é fácil de se reconhecer as quantidades percentuais de cada tipo de consumidor. Este estudo de caso, visa obter uma amostra do Sistema Elétrico de Potência em um ponto qualquer do sistema. De posse dessa amostra e através de técnicas de reconhecimento de padrões, determinar estes percentuais.

Tal determinação é importante no estabelecimento de políticas de fornecimento de energia, expansão do sistema e em situações de emergência, visando atender cargas prioritárias.

Quando se fala a respeito de reconhecimento de padrões o “ícone” que vêm se a cabeça são as Redes Neurais Artificiais- RNA. Para fugir deste paradigma foi utilizado um algoritmo que é utilizado em Mineração de Dados para realizar extração de conhecimentos. Diante deste pensamento escolheu-se o algoritmo Naive Bayes, para realizar o reconhecimento das curvas de carga.

6.1 Simulações

Existem vários tipos de consumidores como residencial, comercial, industrial, público, rural dentre outros. Segundo o Atlas da Energia no Brasil [8], estes são os setores com maior expressão até o ano de 1998, com uma parcela de 83,8% de toda energia consumida no país. Apesar da figura 6.1 nos informar até o ano de 1998, o consumo de energia sem se modificando a cada ano, mas os principais consumidores continuam sendo os mesmos. Por este motivo utilizou-se os consumidores residenciais, comerciais e industriais.

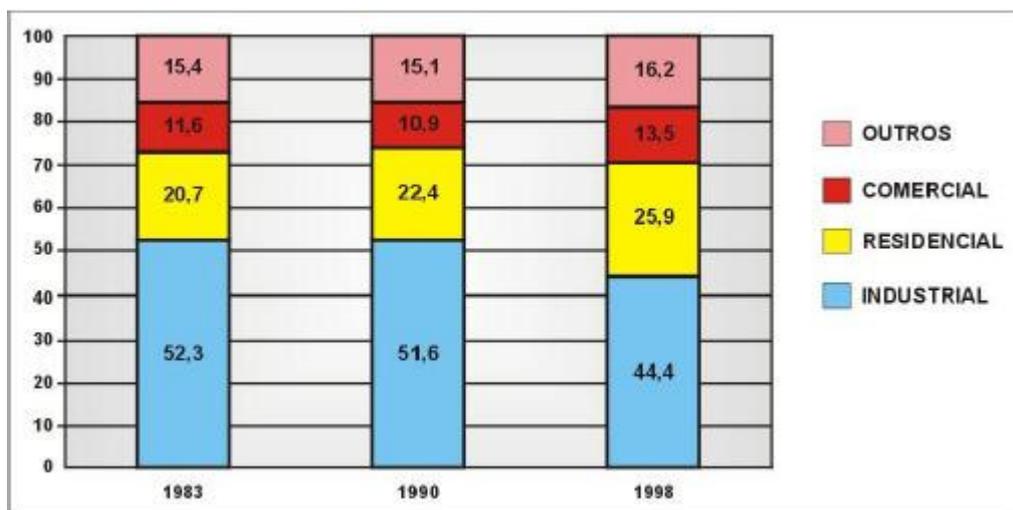


Figura 6.1 - Classes de consumidores de energia elétrica.

Assim como acontece nas RNA's, no algoritmo Naive Bayes é necessário um conjunto de treinamento que retrate com exatidão informações que o algoritmo deverá identificar e classificar. O conjunto de treinamento deverá ter uma maior diversidade para evitar que o algoritmo fique especializado em determinados padrões.

Abaixo se tem os padrões de treinamento que se referem a Curvas de Carga referentes ao consumo de energia elétrica dos consumidores: Residencial, Comercial e Industrial.

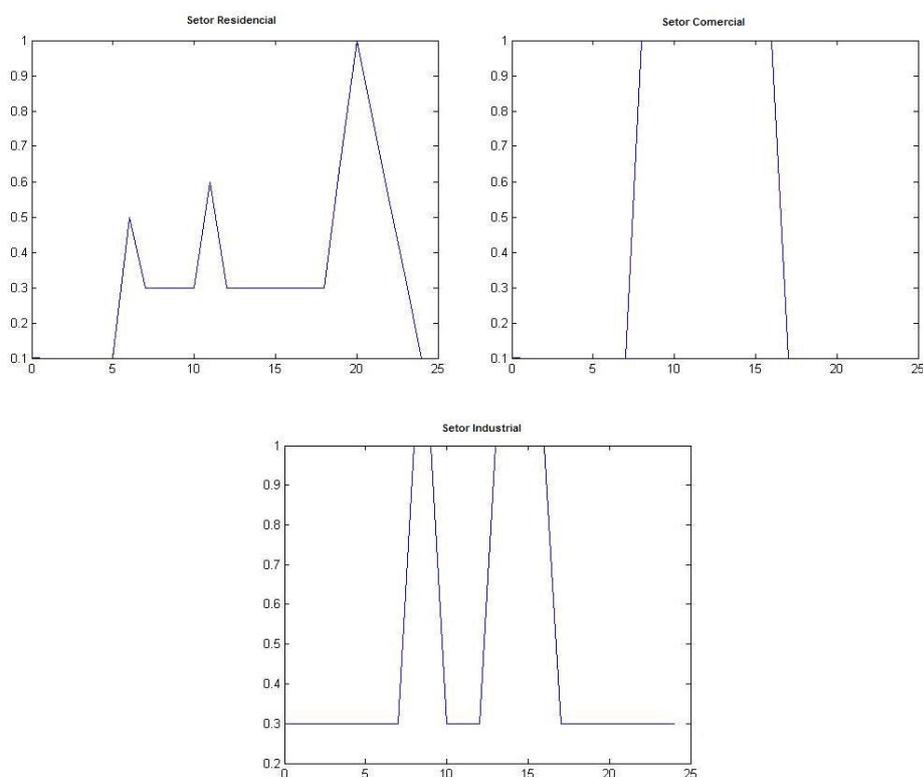


Figura 6.2- Curva de carga representando os setores residencial, comercial e industrial.

O eixo das abcissas representa a hora no decorrer do dia, e eixo das ordenadas representa o consumo de energia normalizado em 1. A medida de hora poderia ser feita em intervalos menores, caso queira analisar o consumo de energia de forma mais minuciosa.

Pode-se notar que no gráfico do setor residencial tem-se um pico de consumo às 6 onde a população acorda para ir trabalhar, às 12:00 horas onde a população esta almoçando e apartir das 18:00 a população volta para seus domicílios. Para o gráfico do setor comercial tem um maior consumo no horário

comercial, ou seja, de 8:00 às 18:00. E para finalizar, o padrão que representa o setor industrial, também possui picos de consumo às 7:00 porque os operários chegam para o trabalho, às 14:00 quando os mesmo retornam ao trabalho e a partir das 18:00 as máquinas são desligadas e o consumo retorna a baixos níveis novamente.

Definido os padrões de curva de carga, o passo seguinte é elaborar o conjunto de treinamento, para geração do modelo matemático do Naive Bayes, e o conjunto de teste ou validação. O conjunto de treinamento é formado pelos padrões abaixo (residencial, comercial e industrial) com diversas amplitudes.

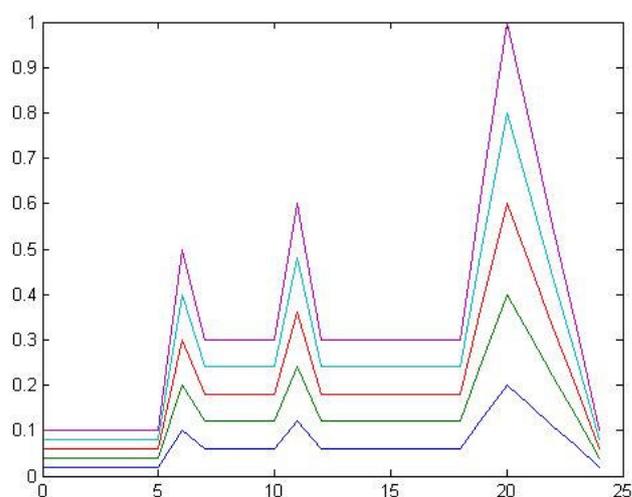


Figura 6.3- Padrões de treinamento representando o setor residencial.

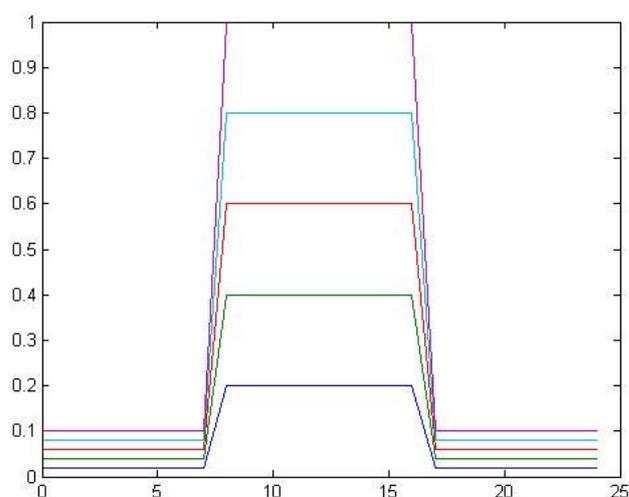


Figura 6.4- Padrões de treinamento representando o setor comercial.

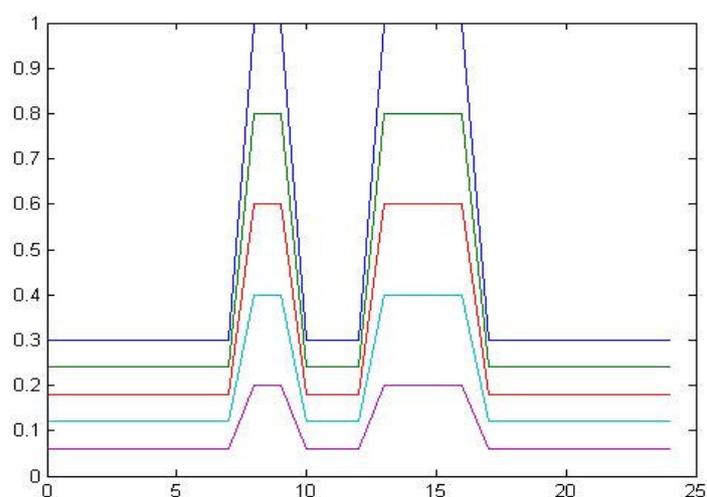


Figura 6.5- Padrões de treinamento representando o setor industrial.

O conjunto de treinamento possui amplitudes de 20%, 40%, 60%, 80% e 100% do padrão fundamental. Para facilitar, adotou-se a seguinte nomenclatura: 20% do padrão residencial será chamado de R20; 40% do padrão residencial será chamado de R40; 60% do padrão residencial será chamado de R60; 80% do padrão residencial será chamado de R80 e para finalizar 100% do padrão residencial será chamado de R100. O mesmo raciocínio é válido para o padrão comercial e industrial.

Para o conjunto de teste são utilizadas 4 novas amostras com amplitudes diferentes do conjunto de treinamento, com a finalidade de saber se o classificador Bayesiano é capaz de classificar corretamente as amostras.

Quando se trata com um conjunto de dados que possuem mais de duas classes, multiclass, o resultado do processo de classificação é frequentemente mostrado em uma matriz de duas dimensões, chamada de matriz de confusão. Na matriz de confusão tem-se uma linha e coluna para cada classe. A linha nos mostra a quantidade de instâncias que se tem de uma classe e na coluna a quantidade de instâncias que foram classificadas para aquela classe específica. O melhor resultado para a matriz de confusão é quando os elementos abaixo e acima da diagonal principal é zero. Na figura 6.6, tem-se um exemplo de uma matriz de confusão para entender melhor o conceito acima.

a b c d	<- Classificado como
1 0 0 0	a = R20.0
0 1 0 0	b = R40.0
0 0 1 0	c = R60.0
0 0 0 1	d = R80.0

Figura 6.6- Exemplo de uma matriz de confusão.

Nota-se que no conjunto de treinamento têm-se quatro classes (R20, R40, R60, R80); quatro instâncias, sendo cada uma delas pertencentes a cada uma das quatro classes; e o classificador classificou todas as instâncias do conjunto de teste corretamente, ou seja, tem-se somente uma matriz sem elementos acima e abaixo da diagonal principal. Para simplificar a resposta do algoritmo é informado neste trabalho, somente a classe do padrão reconhecido, omitindo a matriz de confusão.

6.1.1 Testes utilizando padrões simples

As amostras possuem amplitudes de 25%, 50%, 75% e 90% dos padrões fundamentais.

O primeiro padrão a ser verificado é o residencial com amplitudes de R25, R50, R75 e R90.

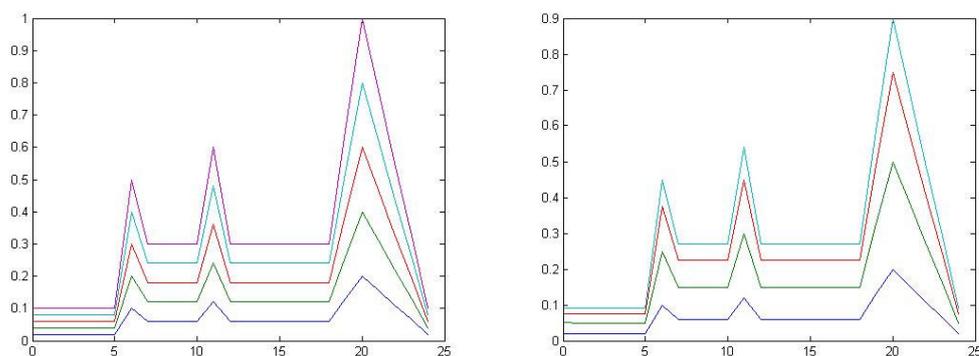


Figura 6.7 - Padrão de treinamento e teste simples para consumidor residencial.

A tabela 6.1 nos mostra os resultados obtidos no reconhecimento.

Tabela 6.1 - Resultado dos testes realizados com padrões residenciais simples.

Padrão Avaliado	Padrão Reconhecido
R25	R20
R50	R60
R75	R80
R90	R100

Nota-se que o algoritmo associou as amostras, as classes mais próximas dos seus valores. Os padrões R50 e R90 estão no meio de duas classes e o

algoritmo reconheceu essa amostra como sendo pertencentes a classe com maior valor.

Agora o próximo teste foi realizado com padrões de treinamento e teste comerciais.

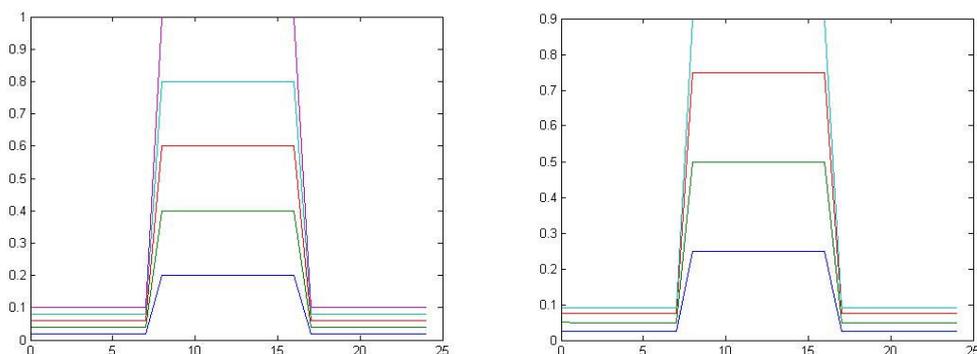


Figura 6.8- Padrão de treinamento e teste simples para consumidor comercial.

A tabela 6.2 nos mostra os resultados obtidos no reconhecimento.

Tabela 6.2 - Resultado dos testes realizados com padrões comerciais simples.

Padrão Avaliado	Padrão Reconhecido
C25	C20
C50	C40
C75	C80
C90	C80

Observa-se que o Naive Bayes associou as amostras, as classes mais próximas dos seus valores com ocorrido para o conjunto de teste residencial. Os padrões C50 e C90 estão no meio de duas classes e o algoritmo reconheceu essas amostra como sendo pertencentes a classe com menor valor.

Finalizando esta etapa, tem-se os testes com padrão industrial.

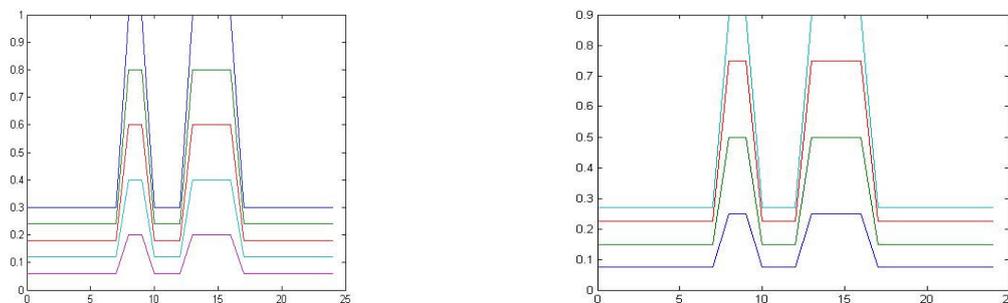


Figura 6.9- Padrão de treinamento e teste simples para consumidor industrial.

A tabela 6.3 nos mostra os resultados obtidos no reconhecimento.

Tabela 6.3 - Resultado dos testes realizados com padrões industriais simples.

Padrão Avaliado	Padrão Reconhecido
125	120
150	160
175	180
190	1100

O “Naive Bayes” reconheceu as amostras, como classes mais próximas dos seus valores. Pode-se concluir com estes teste que o algoritmo Naive Bayes realizou a função que é de realizar o reconhecimento de padrões dentro do seu domínio de conhecimento. Nos testes realizados o domínio de conhecimento do Naive Bayes são as classes do conjunto de treinamento. Para se obter uma maior precisão da classe reconhecida tem-se que aumentar o domínio de conhecimento,

ou seja, inserir classes com intervalos menores entre elas. Por exemplo R5, R10, R15 ... I95 e I100.

6.1.2 Testes utilizando composição de padrões simples

A próxima sequência de simulação, realizou-se a soma dos padrões fundamentais com o intuito de ter uma curva de carga mais realística. Isto porque, se for adquirida uma curva de carga em um determinado instante de tempo do Sistema Elétrico de Potência, a mesma será um padrão composto.

Como a partir de agora, o algoritmo “Naive Bayes” tem que reconhecer padrões compostos e conseqüentemente o conjunto de treinamento tem que sofrer algumas alterações. Para isto tem-se que inserir os padrões estudados anteriormente em um único conjunto de treinamento, porque o conjunto de teste ou validação será formado por padrões compostos. Vale salientar que não utilizou-se padrões compostos no conjunto de treinamento e somente padrões simples. O conjunto de treinamento pode ser visto na figura 6.10.

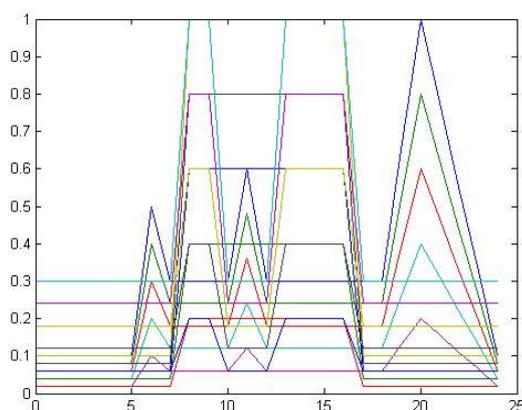


Figura 6.10- Padrões de treinamento representando setor residencial, comercial e industrial.

A primeira simulação foi realizada com a combinação entre o padrão residencial e comercial com amplitude R20C80, R40C60, R60C40 e R80C20. Na figura 6.11 mostra o conjunto de treinamento e os padrões de teste.

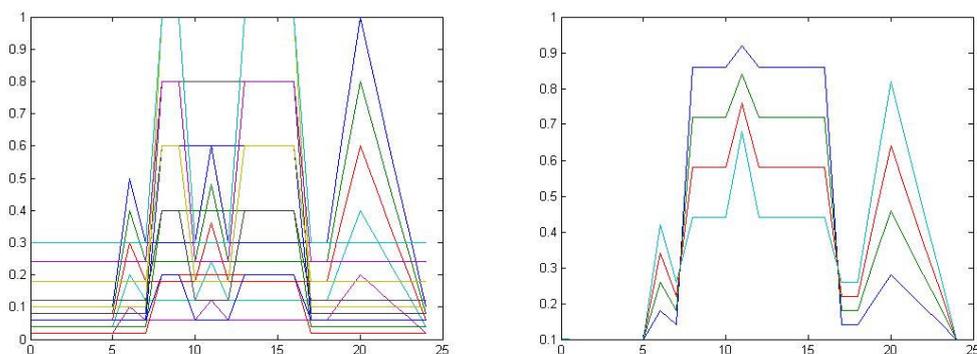


Figura 6.11 - Padrões de treinamento e padrão de teste R20C80, R40C60, R60C40, R80C20.

A tabela 6.4 mostra os resultados obtidos no reconhecimento.

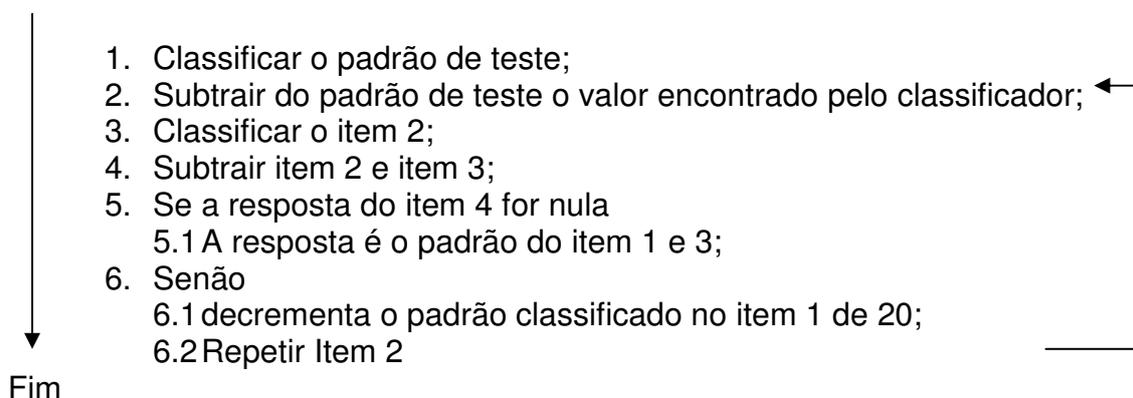
Tabela 6.4 - Resultado dos testes realizados com padrões R20C80, R40C60, R60C40 e R80C20.

Padrão Avaliado	Padrão Reconhecido (Naive Bayes)
R20C80	C100
R40C60	C80
R60C40	C60
R80C20	R100

Pode-se observar que os padrões reconhecidos na tabela acima, nos indica somente uma parcela do padrão avaliado. Isto ocorre porque o conjunto de treinamento não possui padrão composto e conseqüentemente não pode reconhecê-los. Para contornar este problema e não ter que utilizar padrões compostos no conjunto de treinamento, foi desenvolvido um algoritmo para complementar o Naive

Bayes. O algoritmo “Naive Bayes” nos fornece uma “pista” do tipo de curva de carga e o algoritmo auxiliar, através de subtração de padrões, consegue reconhecer o tipo de consumidor presente naquele instante do sistema elétrico. Abaixo pode-se ver este algoritmo que esta na forma genérica e pode ser implementado em qualquer linguagem de programação. Neste trabalho foi implementado em Java com intuito de trabalhar em conjunto com o pacote computacional Weka.

Início



Para exemplificar, abaixo segue a classificação dos padrões descritos na tabela 6.4 utilizando a modificação realizada no “Naive Bayes”.

Padrão R20C80:

1. Classifica(R20C80) = C100;
2. $R20 + C80 - C100 = R20 - C20$;
3. Classifica($R20 - C20$) = XX;
4. $(R20 - C20) - XX$;
5. Item 4 não é nulo então;
6. $C100 - C20 = C80$;
7. $R20 + C80 - C80 = R20$;
8. Classifica($R20$) = R20;
9. $R20 - R20 = 0$;
10. A resposta é nula então é: $C80 + R20$.

Padrão R40C60:

1. Classifica(R40C60) = C80; // Classifica o padrão desconhecido.
2. $R40 + C60 - C80 = R40 - C20$; // Realiza a subtração do Padrão de teste e o valor classificado anteriormente.
3. Classifica($R40 - C20$) = XX;
4. $(R40 - C20) - XX$;
5. Item 4 não é nulo então;
6. $C80 - C20 = C60$;
7. $R40 + C60 - C60 = R20$;
8. Classifica(R40) = R40;
9. $R40 - R40 = 0$;
10. A resposta é nula então é: $C60 + R40$.

Padrão R60C40:

1. Classifica(R60C40) = C60; // Classifica o padrão desconhecido.
2. $R60 + C40 - C60 = R60 - C20$; // Realiza a subtração do Padrão de teste e o valor classificado anteriormente.
3. Classifica($R60 - C20$) = XX;
4. $(R60 - C20) - XX$;
5. Item 4 não é nulo então;
6. $C60 - C20 = C40$;
7. $R60 + C40 - C40 = R60$;
8. Classifica(R60) = R60;
9. $R60 - R60 = 0$;
10. A resposta é nula então é: $C40 + R60$.

Padrão R80C20:

1. Classifica(R80C20) = R100; // Classifica o padrão desconhecido.
2. $R80 + C20 - R100 = C20 - R20$; // Realiza a subtração do Padrão de teste e o valor classificado anteriormente.
3. Classifica($C20 - R20$) = XX;
4. $(C20 - R20) - XX$;
5. Item 4 não é nulo então;
6. $R100 - R20 = R80$;
7. $R80 + C20 - R80 = C20$;
8. Classifica(C20) = C20;
9. $C20 - C20 = 0$;
10. A resposta é nula então é: $R80 + C20$.

A tabela abaixo nos mostra o resultado final

Tabela 6.5 - Resultado dos testes realizados com padrões R20I80, R40I60, R60I40 e R80I20.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R20C80	C100	R20 + C80
R40C60	C80	R40 + C60
R60C40	C60	R60 + C40
R80C20	R100	R80 + C20

A próxima simulação foi realizada com a combinação entre o padrão residencial e industrial com amplitude R20I80, R40I60, R60I40 e R80I20. Na figura 6.12, mostra o conjunto de treinamento e os padrões de teste.

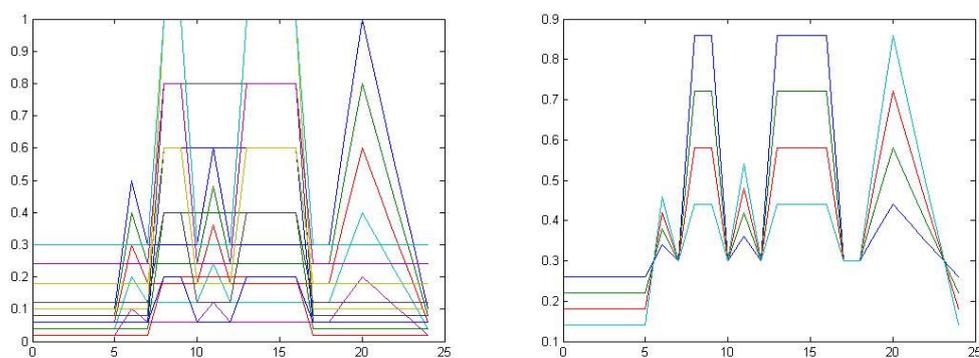


Figura 6.12 - Padrões de treinamento e padrão de teste R20I80, R40I60, R60I40, R80I20.

A tabela 6.6 mostra os resultados obtidos no reconhecimento.

Tabela 6.6 - Resultado dos testes realizados com padrões R20I80, R40I60, R60I40 e R80I20.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R20I80	I100	R20 + I80
R40I60	I60	R40 + I60
R60I40	I60	R60 + I40
R80I20	R100	R80 + I20

O padrão de teste R40I60 e R60I40 foram reconhecidos na mesma classe, mas este fato não compromete o reconhecimento dos mesmos pelo algoritmo.

A última simulação foi realizada com os padrões de teste iguais a C20I80, C40I60, C60I40 e C80I20, ou seja, uma combinação entre o padrão comercial e industrial.

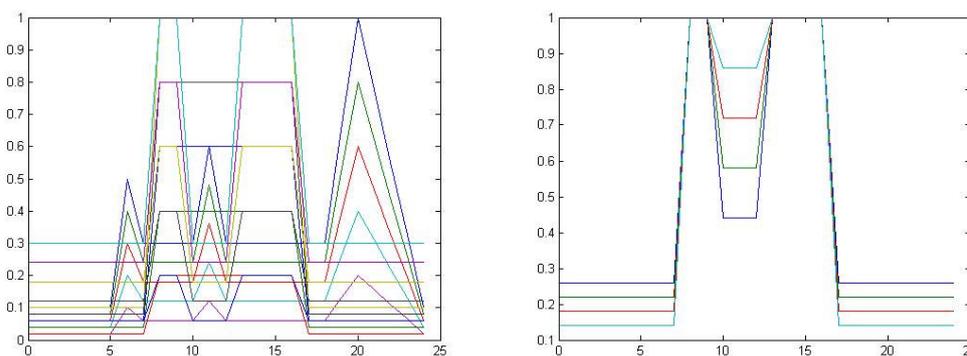


Figura 6.13 - Padrões de treinamento e padrão de teste C20I80, C40I60, C60I40, C80I20.

A tabela 6.7 mostra os resultados obtidos no reconhecimento.

Tabela 6.7 - Resultado dos testes realizados com padrões C20I80, C40I60, C60I40 e C80I20.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
C20I80	I100	C20 + I80
C40I60	I80	C40 + I60
C60I40	I60	C60 + I40
C80I20	C100	C80 + I20

Nota-se que os resultados acima foram satisfatórios, porque conseguiu se reconhecer padrões compostos utilizando apenas padrões simples.

Observa-se que os padrões compostos são formados por padrões simples, os quais o modelo gerado os conhece. Então a próxima etapa realizada foi escolher um conjunto de teste que não tivesse nenhuma relação com o conjunto de treinamento.

Inicialmente realizou-se a combinação entre o padrão residencial e comercial com amplitude R10C90, R30C70, R50C50, R70C30, R90C10. Na figura 6.14 mostra o conjunto de treinamento e o padrão o conjunto de teste para as amplitudes descritas acima.

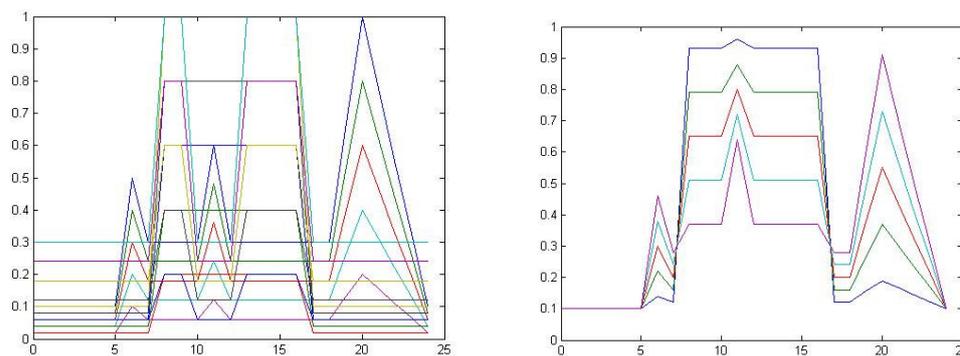


Figura 6.14- Padrões de treinamento e padrão de teste R10C90, R30C70, R50C50, R70C30, R10C90.

A tabela 6.8 mostra os resultados obtidos no reconhecimento.

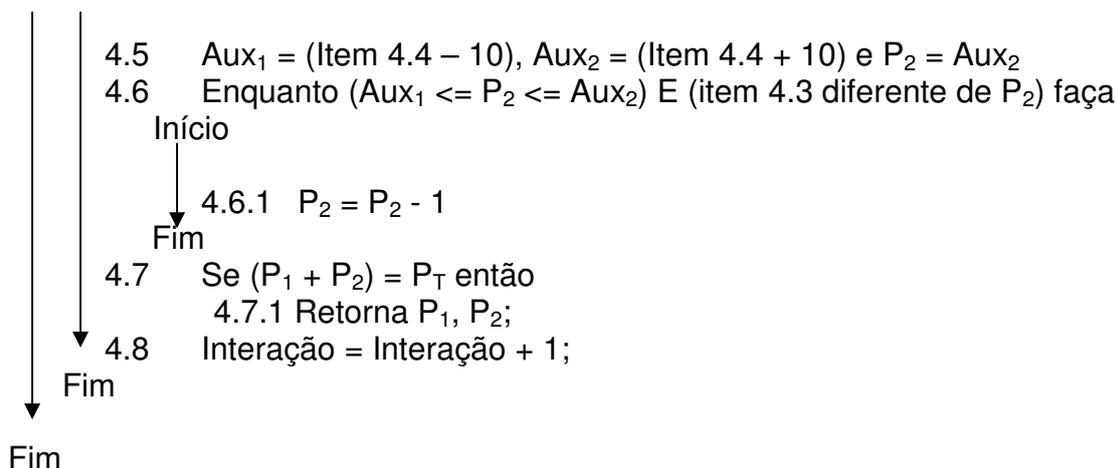
Tabela 6.8 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 R90C10.

Padrão Avaliado	Padrão Reconhecido
R10C90	C100
R30C70	C80
R50C50	I60
R70C30	C60
R90C10	R100

Como anteriormente, teve-se que desenvolver um novo algoritmo para auxiliar o Naive Bayes à reconhecer padrões compostos, sem tê-los treinados.

Início

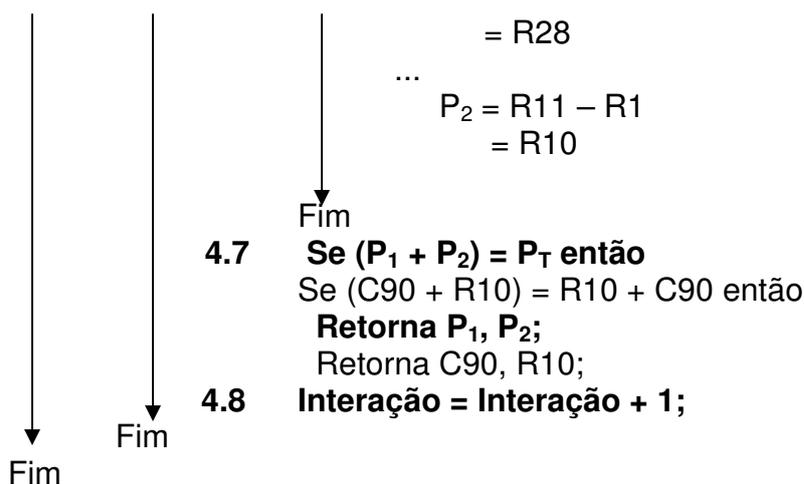
1. $P_T = XX$
2. $P_C = \text{Classifica}(P_T)$;
3. Passo = 10, Interação = 0;
4. Enquanto (Interação < 3) faça
 - Início
 - 4.1 $P_1 = P_C - \text{Passo}$
 - 4.2 Passo = Passo + 10;
 - 4.3 $P_T - P_1$;
 - 4.4 Classificar item 4.3;



Para exemplificar, abaixo segue a classificação do padrão de teste R10C90.

Início

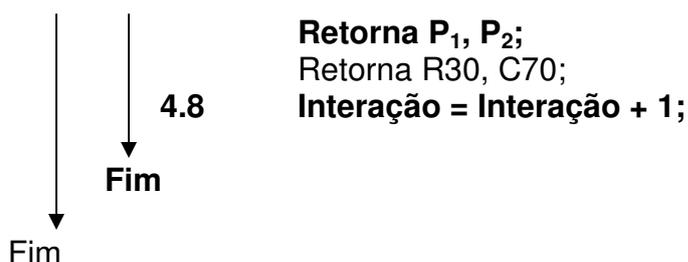
1. $P_T = R10 + C90$;
2. **Pc = Classifica(P_T);**
 $P_c = \text{Classifica}(R10 + C90)$;
 $P_c = C100$
3. **Passo = 10;**
4. **Enquanto (Interação < 3) faça**
 Início
 - 4.1 **P_C - Passo**
 $P_1 = C100 - C10$;
 $P1 = C90$;
 - 4.2 **Passo = 10 + 10; // Passo + 10;**
 $\text{Passo} = 20$;
 - 4.3 **P_T - P₁;**
 $R10 + C90 - C90$
 $R10$
 - 4.4 **Classificar item 4.3;**
 $\text{Classifica}(R10)$;
 $R20$;
 - 4.5 **Aux₁ = (Item 4.4 - 10), Aux₂ = (Item 4.4 + 10) e P₂ = Aux₂**
 $\text{Aux}_1 = (R20 - R10)$, $\text{Aux}_2 = (R20 + R10)$, $P_2 = (R20 + R10)$
 $\text{Aux}_1 = R10$, $\text{Aux}_2 = R30$, $P_2 = R30$
 - 4.6 **Enquanto (Aux₁ <= P₂ <= Aux₂) E (item 4.3 diferente de P₂) faça**
 $\text{Enquanto}(R10 <= (P_2 = R10) <= R30)$ E (R10 diferente de P₂ = R30) faça
 Início
 - 4.6.1 **P₂ = P₂ - 1**
 $P_2 = R30 - R1$
 $= R29$
 $P_2 = R29 - R1$



Para o padrão de teste R30C70, tem-se a execução abaixo:

Início

1. $P_T = R_{30} + C_{70}$;
2. **Pc = Classifica(PT);**
 $P_c = \text{Classifica}(R_{30} + C_{70})$;
 $P_c = C_{80}$
3. **Passo = 10;**
4. **Enquanto (Interação < 3) faça**
Início
 - 4.1 **$P_c - \text{Passo}$**
 $P_1 = C_{80} - C_{10}$;
 $P_1 = C_{70}$;
 - 4.2 **$\text{Passo} = 10 + 10$; // $\text{Passo} + 10$;**
 $\text{Passo} = 20$;
 - 4.3 **$P_T - P_1$;**
 $R_{30} + C_{70} - C_{70}$
 R_{30}
 - 4.4 **Classificar item 4.3;**
 $\text{Classifica}(R_{20})$;
 R_{20} ;
 - 4.5 **$\text{Aux}_1 = (\text{Item 4.4} - 10)$, $\text{Aux}_2 = (\text{Item 4.4} + 10)$ e $P_2 = \text{Aux}_2$**
 $\text{Aux}_1 = (R_{20} - R_{10})$, $\text{Aux}_2 = (R_{20} + R_{10})$, $P_2 = (R_{20} + R_{10})$
 $\text{Aux}_1 = R_{10}$, $\text{Aux}_2 = R_{30}$, $P_2 = R_{30}$
 - 4.6 **Enquanto ($\text{Aux}_1 \leq P_2 \leq \text{Aux}_2$) E (item 4.3 diferente de P_2) faça**
Enquanto $(R_{10} \leq (P_2 = R_{30}) \leq R_{30})$ E $(R_{30}$ diferente de $P_2 = R_{30})$ faça
Início
 - 4.6.1 **$P_2 = P_2 - 1$**
 Fim
 - 4.7 **Se $(P_1 + P_2) = P_T$ então**
Se $(C_{70} + R_{30}) = R_{30} + C_{70}$ então



O algoritmo abaixo realiza o reconhecimento do padrão de teste, R50C50.

Para facilitar, adotou-se que cada interação do algoritmo esta representada através das cores preto, vermelho e verde.

Início

1. $P_T = R50 + C50$;
2. **Pc = Classifica(P_T);**
 $P_c = \text{Classifica}(R50 + C50)$;
 $P_c = I60$;
3. **Passo = 10, Interação = 0;**
4. **Enquanto (Interação < 3) faça**
 Início
 - 4.1 **P1 = Pc - Passo**
 $P_1 = I60 - I10$;
 $P_1 = I50$;
 $P_1 = I60 - I20$;
 $P_1 = I40$;
 $P_1 = I60 - I30$;
 $P_1 = I30$;
 - 4.2 **Passo = Passo + 10;**
 $Passo = 20$;
 $Passo = 30$;
 $Passo = 40$;
 - 4.3 **$P_T - P_1$;**
 $R50 + C50 - I50$
 $R50 + C50 - I40$
 $R50 + C50 - I30$
 - 4.4 **Classificar item 4.3;**
 $\text{Classifica}(R50 + C50 - I50)$;
 $R40$;
 $\text{Classifica}(R50 + C50 - I40)$;
 $R40$;
 $\text{Classifica}(R50 + C50 - I30)$;
 $R40$;
 - 4.5 **$Aux_1 = (\text{Item 4.4} - 10)$, $Aux_2 = (\text{Item 4.4} + 10)$ e $P_2 = Aux_2$**
 $Aux_1 = (R40 - R10)$, $Aux_2 = (R40 + R10)$, $P_2 = (R40 + R10)$
 $Aux_1 = R30$, $Aux_2 = R50$, $P_2 = R50$
 $Aux_1 = (R40 - R10)$, $Aux_2 = (R40 + R10)$, $P_2 = (R40 + R10)$

$Aux_1 = R30, Aux_2 = R50, P_2 = R50$

$Aux_1 = (R40 - R10), Aux_2 = (R40 + R10), P_2 = (R40 + R10)$

$Aux_1 = R30, Aux_2 = R50, P_2 = R50$

4.6 Enquanto ($Aux_1 \leq P_2 \leq Aux_2$) E (item 4.3 diferente de P_2) faça

Enquanto ($R30 \leq (P_2 = R50) \leq R50$) E ($R50 + C50 - I50$ diferente de $P_2 = R50$) faça

Início

4.6.1 $P_2 = P_2 - 1$

$P_2 = R50 - R1$

$= R49$

$P_2 = R49 - R1$

$= R28$

...

$P_2 = R31 - R1$

$= R30$

Fim

Enquanto ($R30 \leq (P_2 = R50) \leq R50$) E ($R50 + C50 - I50$ diferente de P_2) faça

Início

4.6.1 $P_2 = P_2 - 1$

$P_2 = R50 - R1$

$= R49$

$P_2 = R49 - R1$

$= R28$

$P_2 = R31 - R1$

$= R30$

Fim

Enquanto ($R30 \leq (P_2 = R50) \leq R50$) E ($R50 + C50 - I50$ diferente de $P_2 = R50$) faça

Início

4.6.1 $P_2 = P_2 - 1$

$P_2 = R50 - R1$

$= R49$

$P_2 = R49 - R1$

$= R28$

$P_2 = R31 - R1$

$= R30$

Fim

4.7 Se ($P_1 + P_2$) = P_T então

Se ($I50 + R30$) = $R50 + C50$ então

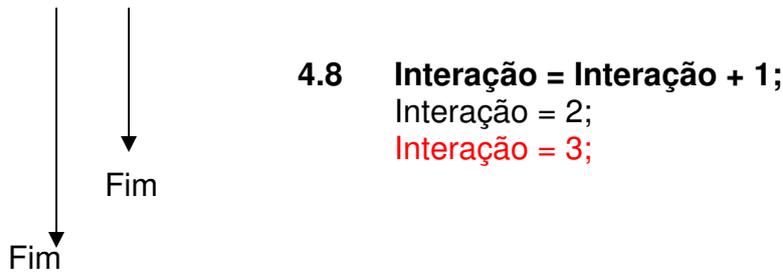
Retorna P_1, P_2 ;

Se ($I50 + R30$) = $R50 + C50$ então

Retorna P_1, P_2 ;

Se ($I50 + R30$) = $R50 + C50$ então

Retorna P_1, P_2 ;



Observa-se que neste caso o padrão de teste não pode ser reconhecido porque o “Naive Bayes” primeiramente reconheceu o padrão de teste como I60. Este padrão simples não faz parte da composição do padrão de teste.

Para o padrão de teste R70C30, tem-se:

Início

1. $P_T = R70 + C30$
2. **Pc = Classifica(P_T);**
 $P_c = \text{Classifica}(R70 + C30);$
 $P_c = C60$
3. **Passo = 10, Interação = 0;**
4. **Enquanto (Interação < 3) faça**
 Início
 - 4.1 **$P_1 = P_c - \text{Passo}$**
 $P_1 = C60 - C10;$
 $P_1 = C50;$
 $P_1 = C60 - C20;$
 $P_1 = C40;$
 $P_1 = C60 - C30;$
 $P_1 = C30;$
 - 4.2 **Passo = Passo + 10;**
 $\text{Passo} = 20;$
 $\text{Passo} = 30;$
 $\text{Passo} = 40;$
 - 4.3 **$P_T - P_1;$**
 $R70 + C30 - C50 = R70 - C20;$
 $R70 + C30 - C40 = R70 - C10;$
 $R70 + C30 - C30 = R70;$
 - 4.4 **Classificar item 4.3;**
 $\text{Classifica}(R70 - C20);$
 $R80;$
 $\text{Classifica}(R70 - C10);$
 $R80;$
 $\text{Classifica}(R70);$
 $R80;$
 - 4.5 **$\text{Aux}_1 = (\text{Item 4.4} - 10)$, $\text{Aux}_2 = (\text{Item 4.4} + 10)$ e $P_2 = \text{Aux}_2$**
 $\text{Aux}_1 = (R80 - R10)$, $\text{Aux}_2 = (R80 + R10)$, $P_2 = (R80 + R10)$
 $\text{Aux}_1 = R70$, $\text{Aux}_2 = R90$, $P_2 = R90$

$Aux_1 = (R80 - R10)$, $Aux_2 = (R80 + R10)$, $P_2 = (R80 + R10)$

$Aux_1 = R70$, $Aux_2 = R90$, $P_2 = R90$

$Aux_1 = (R80 - R10)$, $Aux_2 = (R80 + R10)$, $P_2 = (R80 + R10)$

$Aux_1 = R70$, $Aux_2 = R90$, $P_2 = R90$

4.6 Enquanto ($Aux_1 \leq P_2 \leq Aux_2$) E (item 4.3 diferente de P_2) faça

Enquanto ($R70 \leq (P_2 = R90) \leq R90$) E ($R70 - C20$ diferente de P_2) faça

Início

4.6.1 $P_2 = P_2 - 1$
 $P_2 = R90 - R1$
 $= R89$
 $P_2 = R89 - R1$
 $= R88$
 ...
 $P_2 = R31 - R1$
 $= R30$

Fim

Enquanto ($R70 \leq (P_2 = R50) \leq R90$) E ($R70 - C10$ diferente de P_2) faça

Início

4.6.1 $P_2 = P_2 - 1$
 $P_2 = R90 - R1$
 $= R89$
 $P_2 = R89 - R1$
 $= R88$
 ...
 $P_2 = R31 - R1$
 $= R30$

Fim

Enquanto ($R70 \leq (P_2 = R70) \leq R90$) E ($R70$ diferente de P_2) faça

Início

4.6.1.1.1 $P_2 = P_2 - 1$
 $P_2 = R90 - R1$
 $= R89$
 $P_2 = R89 - R1$
 $= R88$
 ...
 $P_2 = R71 - R1$
 $= R70$

Fim

4.7 Se ($P_1 + P_2$) = P_T então

Se ($C50 + R30$) = $R70 + C30$ então

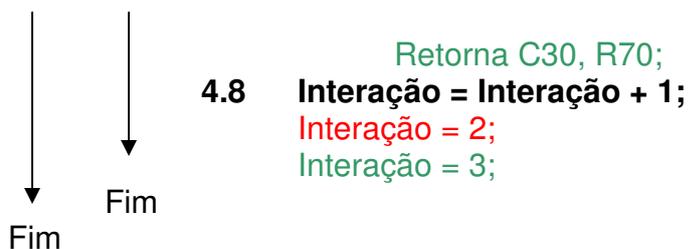
4.7.1 Retorna P_1, P_2 ;

Se ($C40 + R30$) = $R70 + C30$ então

4.7.1 Retorna P_1, P_2 ;

Se ($C30 + R70$) = $R70 + C30$ então

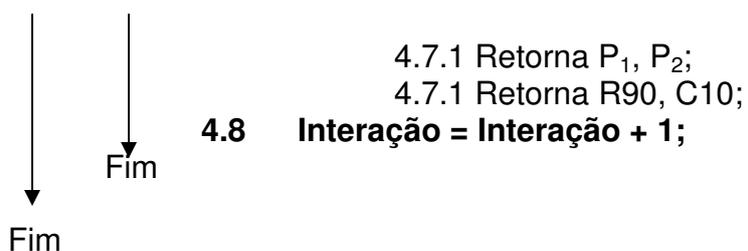
4.7.1 Retorna P_1, P_2 ;



Para o padrão R90C10, tem-se:

Início

1. $P_T = R90 + C10$
2. **$P_c = \text{Classifica}(P_T)$;**
 $P_c = \text{Classifica}(R90 + C10)$;
 $P_c = R100$
3. **Passo = 0, Interação = 0;**
4. **Enquanto (Interação < 3) faça**
 Início
 - 4.1 **$P_1 = P_c - \text{Passo}$**
 $P_1 = R100 - R10$;
 $P_1 = R90$;
 - 4.2 **Passo = Passo + 10;**
 $\text{Passo} = 10$;
 - 4.3 **$P_T - P_1$;**
 $R90 + C10 - R90 = C10$;
 - 4.4 **Classificar item 4.3;**
 $\text{Classifica}(C10)$;
 $C20$;
 - 4.5 **$\text{Aux}_1 = (\text{Item 4.4} - 10)$, $\text{Aux}_2 = (\text{Item 4.4} + 10)$ e $P_2 = \text{Aux}_2$**
 $\text{Aux}_1 = (C20 - C10)$, $\text{Aux}_2 = (C20 + C10)$, $P_2 = (C20 + C10)$
 $\text{Aux}_1 = C10$, $\text{Aux}_2 = C30$, $P_2 = C30$
 - 4.6 **Enquanto ($\text{Aux}_1 \leq P_2 \leq \text{Aux}_2$) E (item 4.3 diferente de P_2) faça**
 Enquanto ($C10 \leq P_2 \leq C30$) E ($C10$ diferente de P_2) faça
 Início
 - 4.6.1 **$P_2 = P_2 - 1$**
 $P_2 = C30 - R1$
 $= C29$
 $P_2 = C29 - C1$
 $= C28$
 \dots
 $P_2 = R31 - R1$
 $= R30$
- 4.7 **Se ($P_1 + P_2$) = P_T então**
 Se ($C10 + R90$) = $R90 + C10$ então



A tabela 6.9 mostra os resultados obtidos na simulação.

Tabela 6.9 – Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R10C90	C100	R10 + C90
R30C70	C80	R30 + C70
R50C50	I60	ERRO
R70C30	C60	R70 + C30
R90C10	R100	R90 + C10

O próximo teste foi realizado com padrões de validação, formados pela composição das curvas de carga residencial e industrial.

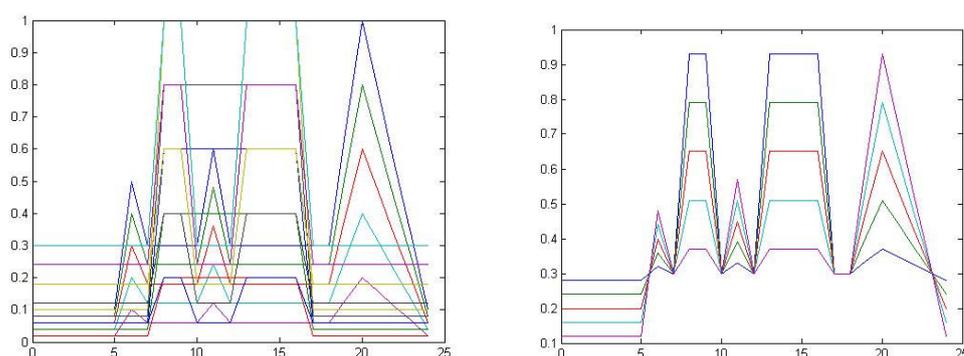


Figura 6.15 - Padrões de treinamento e padrão de teste R10I90, R30I70, R50I50, R70I30, R10I90.

A tabela 6.10 mostra os resultados obtidos na simulação.

Tabela 6.10 - Resultado dos testes realizados com padrões R10I90, R30I70, R50I50, R70I30 e R90I10.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R10I90	I100	R10 + I90
R30I70	I100	R30 + I70
R50I50	I60	R50 + I50
R70I30	I60	R70 + I30
R90I10	R100	R90 + I10

O último teste realizado retrata a combinação entre as curvas de consumo comercial e industrial.

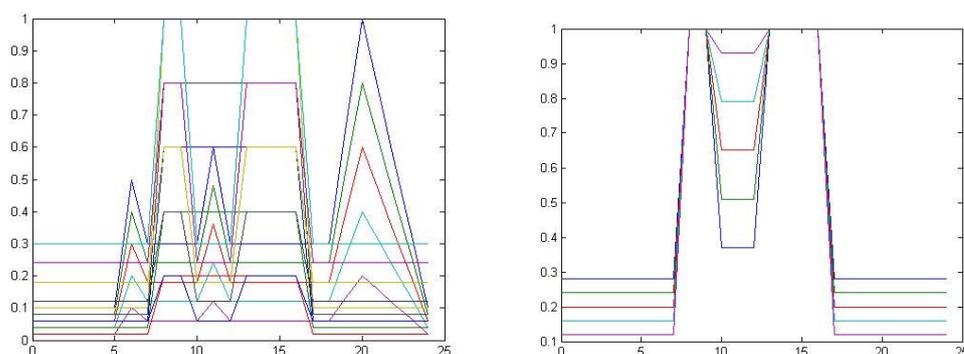


Figura 6.16 - Padrões de treinamento e padrão de teste C10I90, C30I70, C50I50, C70I30, C10I90.

A tabela 6.11 mostra os resultados obtidos na simulação.

Tabela 6.11 - Resultado dos testes realizados com padrões C10I90, C30I70, C50I50, C70I30 e C90I10.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
C10I90	I100	C10 + I90
C30I70	I100	C30 + I70
C50I50	I60	C50 + I50
C70I30	I60	C70 + I30
C90I10	C100	C90 + I10

6.1.3 Testes utilizando composição de padrões simples com oscilações

Até o momento estudou-se os padrões simples e suas combinações, com o objetivo de reconhecê-los. O consumo de energia elétrica é um sistema dinâmico, ou seja, varia constantemente no tempo devido ao fato de ligamento ou desligamento de cargas. O ligamento ou desligamento de uma carga ocorre de forma aleatória. Por exemplo, um consumidor industrial pode ter algum problema na planta e ter que desligar o forno elétrico, ou ligar alguns equipamentos em um determinado instante sem condições prévias. O modelo proposto neste trabalho tem que ser capaz de tratar com essas variações aleatórias do sistema elétrico de potência.

Para realizar este teste, o conjunto de treinamento continua sendo o mesmo dos testes anteriores, ou seja, compostos somente por padrões simples, sem a inserção de padrões compostos e padrões aleatórios.

Primeiramente o conjunto de validação ou testes, consiste dos padrões R10C90, R30C70, R50C50, R70C30 e R90C10 com oscilações aleatórias de -10% a 10% em cima de cada ponta da curva de consumo. A figura 6.17 mostra os padrões com as devidas oscilações aleatórias.

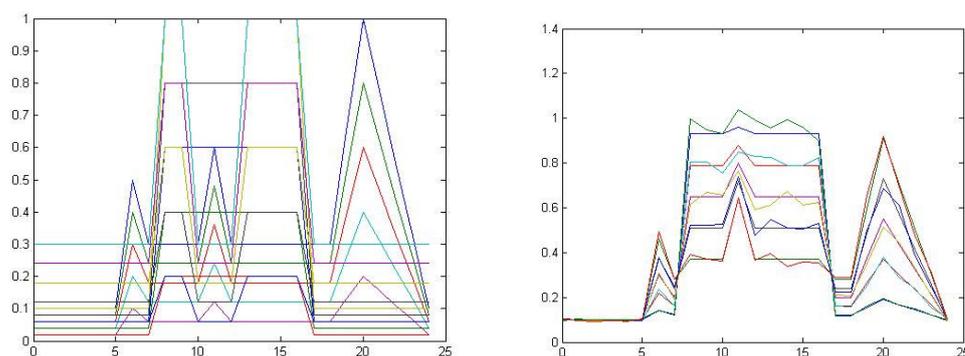


Figura 6.17 - Padrões de treinamento e padrão de teste R10C90, R30C70, R50C50, R70C30, R10C90 com oscilações de 10%.

A tabela 6.12 mostra os resultados obtidos no reconhecimento.

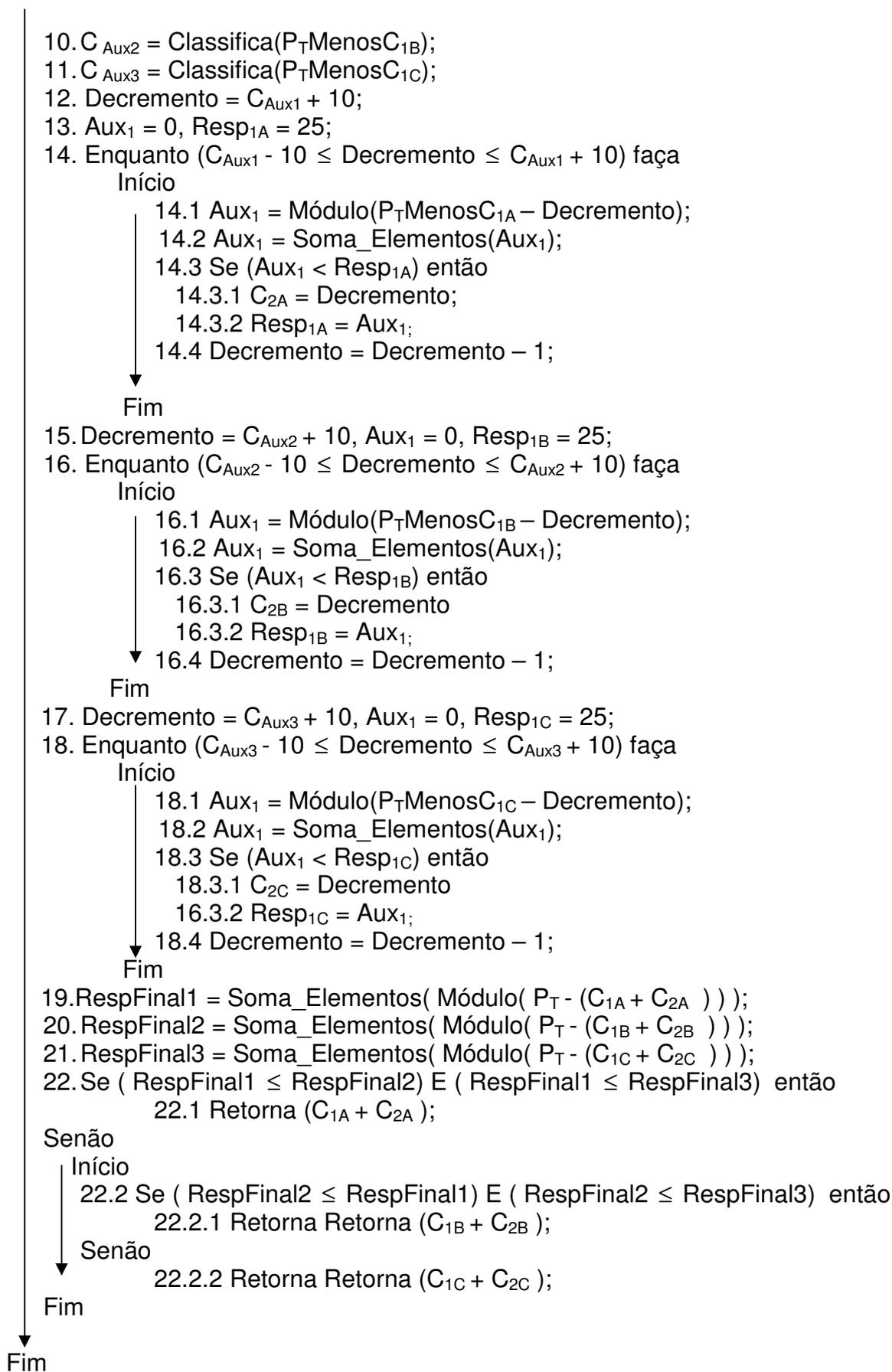
Tabela 6.12 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10 com oscilações de 10%.

Padrão Avaliado	Padrão Reconhecido
R10C90 \pm 10%	C100
R30C70 \pm 10%	C80
R50C50 \pm 10%	I60
R70C30 \pm 10%	C60
R90C10 \pm 10%	R100

Como anteriormente, teve-se que desenvolver um novo algoritmo para auxiliar o “Naive Bayes” à reconhecer padrões compostos com oscilações.

Início

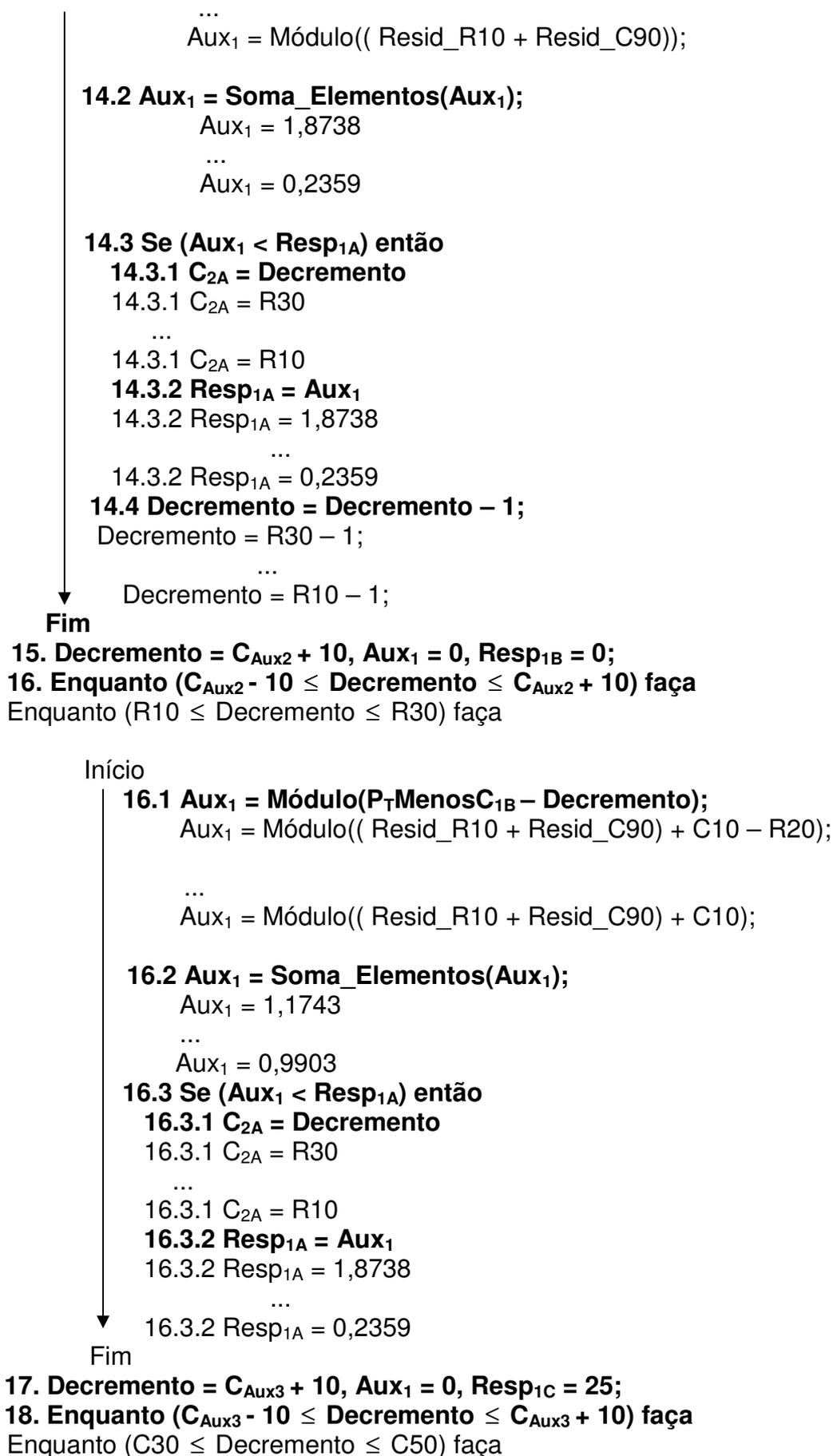
1. $P_T = XX + \text{Resid_}XX$
2. $P_{C_1} = \text{Classifica}(P_T)$;
3. $C_{1A} = P_{C_1} - 10$;
4. $C_{1B} = P_{C_1} - 20$;
5. $C_{1C} = P_{C_1} - 30$;
6. $P_{T\text{Menos}C_{1A}} = P_T - C_{1A}$;
7. $P_{T\text{Menos}C_{1B}} = P_T - C_{1B}$;
8. $P_{T\text{Menos}C_{1C}} = P_T - C_{1C}$;
9. $C_{\text{Aux}1} = \text{Classifica}(P_{T\text{Menos}C_{1A}})$;

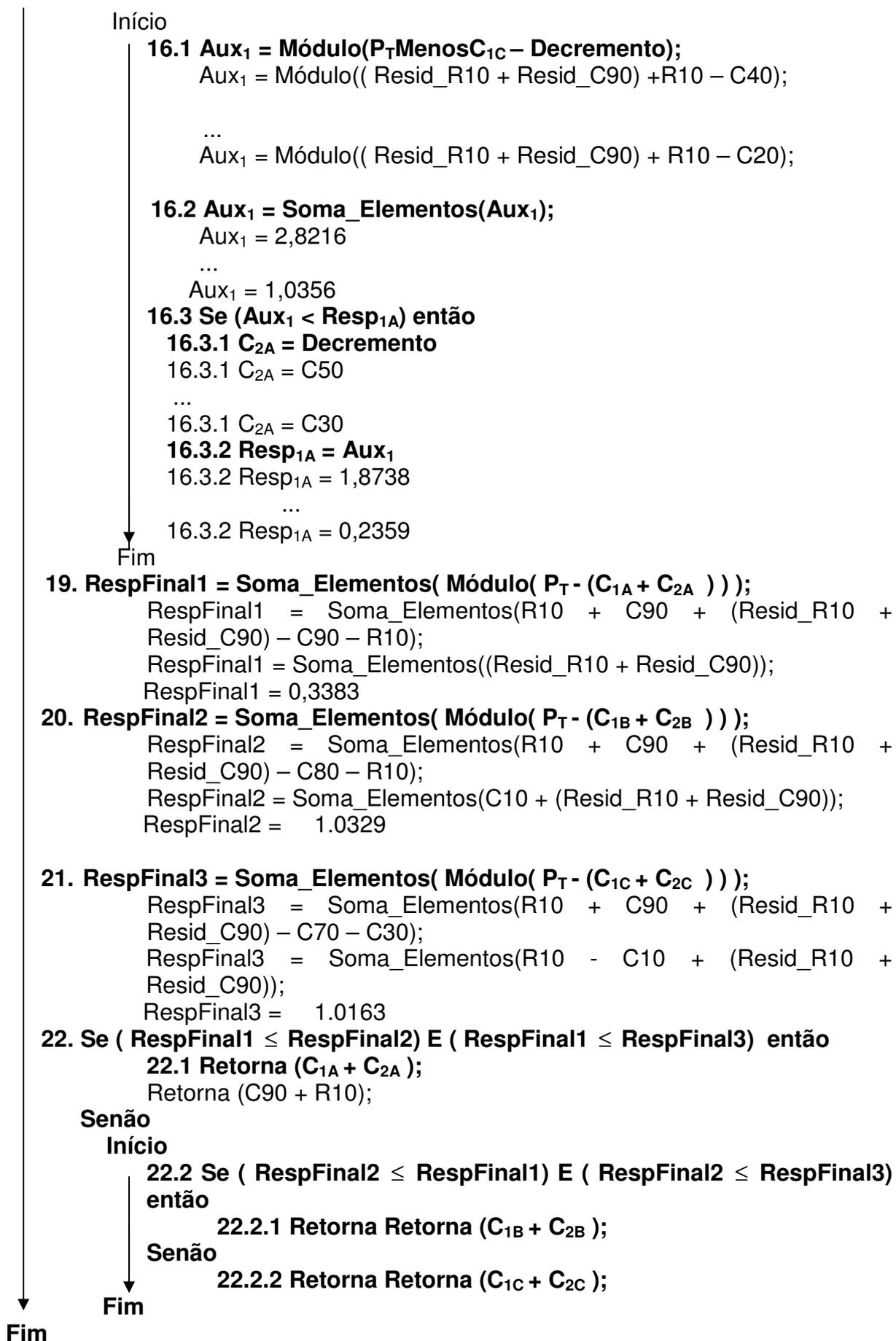


Para exemplificar, abaixo segue a classificação do padrão R10C90 + Oscilações.

Início

1. $P_T = XX + \text{Resid_XX}$
 $= R10 + C90 + \text{Resid_R10} + \text{Resid_C90}$
2. $P_{C1} = \text{Classifica}(P_T)$;
 $P_{C1} = \text{Classifica}(R10 + C90 + \text{Resid_R10} + \text{Resid_C90})$;
3. $C_{1A} = P_{C1} - 10$;
 $C_{1A} = C100 - C10$;
 $= C90$
4. $C_{1B} = P_{C1} - 20$;
 $C_{1B} = C100 - C20$;
 $= C80$
5. $C_{1C} = P_{C1} - 30$;
 $C_{1C} = C100 - C30$;
 $= C70$
6. **$P_T \text{Menos } C_{1A} = P_T - C_{1A}$** ;
 $P_T \text{Menos } C_{1A} = R10 + C90 + \text{Resid_R10} + \text{Resid_C90} - C90$;
 $P_T \text{Menos } C_{1A} = R10 + (\text{Resid_R10} + \text{Resid_C90})$
7. **$P_T \text{Menos } C_{1B} = P_T - C_{1B}$** ;
 $P_T \text{Menos } C_{1B} = R10 + C90 + \text{Resid_R10} + \text{Resid_C90} - C80$;
 $P_T \text{Menos } C_{1B} = R10 + C10 + (\text{Resid_R20} + \text{Resid_C80})$
8. **$P_T \text{Menos } C_{1C} = P_T - C_{1C}$** ;
 $P_T \text{Menos } C_{1C} = R10 + C90 + \text{Resid_R10} + \text{Resid_C90} - C70$;
 $P_T \text{Menos } C_{1C} = R10 + C20 + (\text{Resid_R20} + \text{Resid_C80})$
9. **$C_{Aux1} = \text{Classifica}(P_T \text{Menos } C_{1A})$** ;
 $C_{Aux1} = \text{Classifica}(R10 + C10 + (\text{Resid_R20} + \text{Resid_C80}))$;
 $C_{Aux1} = R20$
10. **$C_{Aux2} = \text{Classifica}(P_T \text{Menos } C_{1B})$** ;
 $C_{Aux2} = \text{Classifica}(R10 + (\text{Resid_R10} + \text{Resid_C90}))$;
 $C_{Aux2} = R20$
11. **$C_{Aux3} = \text{Classifica}(P_T \text{Menos } C_{1C})$** ;
 $C_{Aux3} = \text{Classifica}(R10 + C20 + (\text{Resid_R20} + \text{Resid_C80}))$;
 $C_{Aux3} = C40$
12. **Decremento = $C_{Aux1} + 10$** ;
Decremento = $R20 + R10$;
Decremento = $R30$;
13. **$Aux_1 = 0$, $Resp_{1A} = 25$** ;
14. **Enquanto ($C_{Aux1} - 10 \leq \text{Decremento} \leq C_{Aux1} + 10$) faça**
Enquanto ($R10 \leq \text{Decremento} \leq R30$) faça
Início
 - 14.1 **$Aux_1 = \text{Módulo}(P_T \text{Menos } C_{1A} - \text{Decremento})$** ;
 $Aux_1 = \text{Módulo}((\text{Resid_R10} + \text{Resid_C90}) - R20)$;





A tabela 6.13 mostra os resultados obtidos na simulação.

Tabela 6.13 - Resultado dos testes realizados com padrões R10C90, R30C70, R50C50, R70C30 e R90C10 com oscilações de 10%.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R10C90 \pm 10%	C100	R10 + C90
R30C70 \pm 10%	C80	R30 + C70
R50C50 \pm 10%	I60	ERRO
R70C30 \pm 10%	C60	R70 + C30
R90C10 \pm 10%	R100	R90 + C10

O próximo teste foi realizado com padrões de validação, formados pela composição das curvas de carga residencial e industrial.

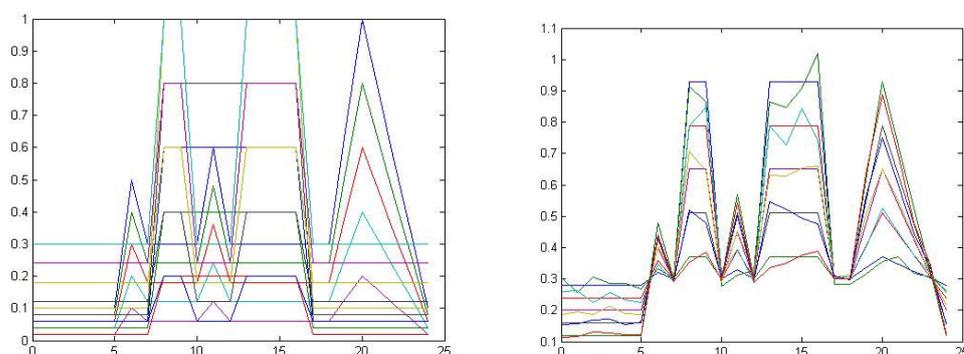


Figura 6.18 - Padrões de treinamento e padrão de teste R10I90, R30I70, R50I50, R70I30, R10I90 com oscilações de 10%.

A tabela 6.14 mostra os resultados obtidos na simulação.

Tabela 6.14 - Resultado dos testes realizados com padrões R10I90, R30I70, R50I50, R70I30 e R90I10 com oscilações de 10%.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R10I90 \pm 10%	I100	R10 + I90
R30I70 \pm 10%	I100	R30 + I70
R50I50 \pm 10%	I60	R50 + I50
R70I30 \pm 10%	I60	R70 + I30
R90I10 \pm 10%	R100	R90 + I10

O último teste realizado retrata a combinação entre as curvas de consumo comercial e industrial. Os resultados podem ser vistos, na tabela 6.15.

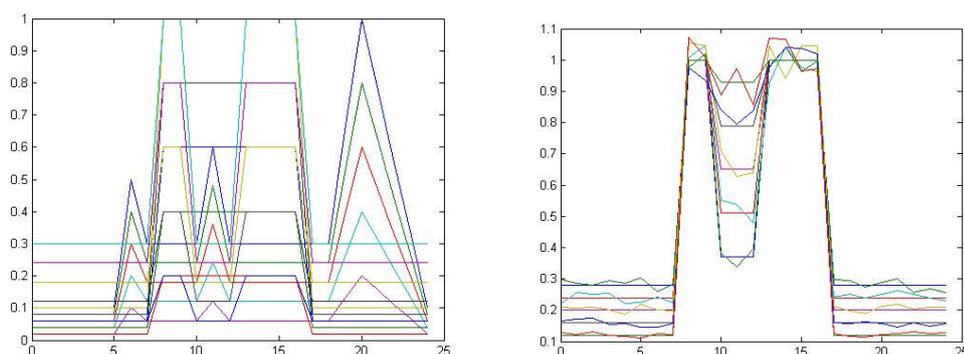


Figura 6.19 - Padrões de treinamento e padrão de teste C10I90, C30I70, C50I50, C70I30, C10I90 com oscilações de 10%.

Tabela 6.15 - Resultado dos testes realizados com padrões C10I90, C30I70, C50I50, C70I30 e C90I10 com oscilações de 10%.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
C10I90 \pm 10%	I100	C10 + I90
C30I70 \pm 10%	I100	C30 + I70
C50I50 \pm 10%	I60	C50 + I50
C70I30 \pm 10%	I60	C70 + I30
C90I10 \pm 10%	C100	C90 + I10

Para finalizar as simulações, deve-se observar um último padrão de carga. Este por sua vez, tem o objetivo de descobrir o que será reconhecido, se o padrão de teste possuir valores que não são possíveis de ser reconhecidos pelo algoritmo porque o passo inicial do algoritmo é de dez unidades. O padrão de teste escolhido foi R8C92 + Oscilações, mas poderia ser qualquer um, desde que a soma das duas componentes seja 100.

Abaixo pode-se ver este padrão de teste.

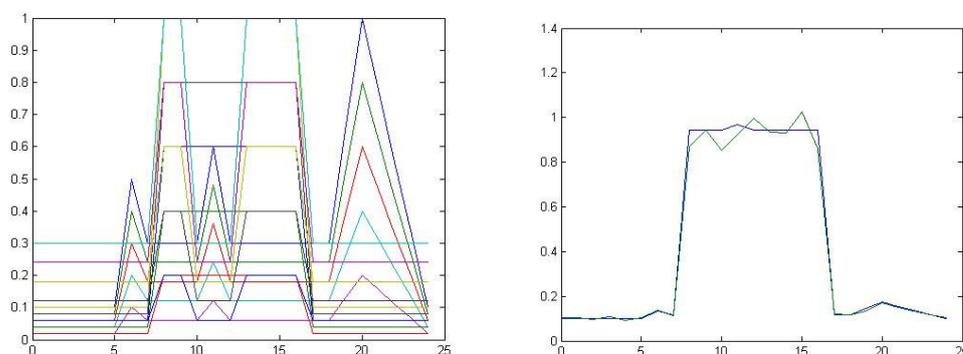


Figura 6.20 - Padrões de treinamento e padrão de teste R8C92 com oscilações de 10%.

A tabela 6.16 mostra o resultado obtido na simulação.

Tabela 6.16 - Resultado do teste realizado com padrão R8C92 com oscilações de 10%.

Padrão Teste	Padrão Reconhecido (Naive Bayes)	Padrão Reconhecido (Naive Bayes + Alg. Auxiliar)
R8C92 \pm 10%	C100	R10 + C90

Abaixo pode-se acompanhar, com detalhes, como algoritmo reconheceu o padrão acima.

Início

1. **$P_T = XX + \text{Resid_XX}$**
 $P_T = R8 + C92 + \text{Resid_R8} + \text{Resid_C92};$
2. **$P_{c1} = \text{Classifica}(P_T);$**
 $P_{c1} = \text{Classifica}(R8 + C92 + \text{Resid_R8} + \text{Resid_C92});$
3. **$C_{1A} = P_{c1} - 10;$**
 $C_{1A} = C100 - C10;$
 $= C90$
4. **$C_{1B} = P_{c1} - 20;$**
 $C_{1B} = C100 - C20;$
 $= C80$
5. **$C_{1C} = P_{c1} - 30;$**
 $C_{1C} = C100 - C30;$
 $= C70$
6. **$P_T \text{Menos } C_{1A} = P_T - C_{1A};$**
 $P_T \text{Menos } C_{1A} = R8 + C92 + \text{Resid_R8} + \text{Resid_C92} - C90;$
 $P_T \text{Menos } C_{1A} = R8 + C2 + (\text{Resid_R8} + \text{Resid_C92})$
7. **$P_T \text{Menos } C_{1B} = P_T - C_{1B};$**
 $P_T \text{Menos } C_{1B} = R8 + C92 + \text{Resid_R8} + \text{Resid_C92} - C80;$
 $P_T \text{Menos } C_{1B} = R8 + C12 + (\text{Resid_R8} + \text{Resid_C92})$
8. **$P_T \text{Menos } C_{1C} = P_T - C_{1C};$**
 $P_T \text{Menos } C_{1C} = R8 + C92 + \text{Resid_R8} + \text{Resid_C92} - C70;$
 $P_T \text{Menos } C_{1C} = R8 + C22 + (\text{Resid_R8} + \text{Resid_C92})$
9. **$C_{Aux1} = \text{Classifica}(P_T \text{Menos } C_{1A});$**
 $C_{Aux1} = \text{Classifica}(R8 + C2 + (\text{Resid_R8} + \text{Resid_C92}));$
 $C_{Aux1} = R20$
10. **$C_{Aux2} = \text{Classifica}(P_T \text{Menos } C_{1B});$**
 $C_{Aux2} = \text{Classifica}(R8 + C12 + (\text{Resid_R8} + \text{Resid_C92}));$
 $C_{Aux2} = C20$
11. **$C_{Aux3} = \text{Classifica}(P_T \text{Menos } C_{1C});$**
 $C_{Aux3} = \text{Classifica}(R8 + C22 + (\text{Resid_R8} + \text{Resid_C92}));$
 $C_{Aux3} = I20$
12. **Decremento = $C_{Aux1} + 10;$**
Decremento = R20 + R10;
Decremento = R30;
13. **$Aux_1 = 0, \text{Resp}_{1A} = 25;$**
14. **Enquanto ($C_{Aux1} - 10 \leq \text{Decremento} \leq C_{Aux1} + 10$) faça**
Enquanto ($R10 \leq \text{Decremento} \leq R30$) faça
Início
 - 14.1 **$Aux_1 = \text{Módulo}(P_T \text{Menos } C_{1A} - \text{Decremento});$**
 $Aux_1 = \text{Módulo}((\text{Resid_R8} + \text{Resid_C92}) + C2 - R22);$
...
 $Aux_1 = \text{Módulo}((\text{Resid_R8} + \text{Resid_C92}) + C2 - R2);$
 - 14.2 **$Aux_1 = \text{Soma_Elementos}(Aux_1);$**
 $Aux_1 = 1,5818$
...
 $Aux_1 = 0,4117$
 - 14.3 **Se ($Aux_1 < \text{Resp}_{1A}$) então**
 - 14.3.1 **$C_{2A} = \text{Decremento}$**

14.3.1 $C_{2A} = R30$

...

14.3.1 $C_{2A} = R10$

14.3.2 $Resp_{1A} = Aux_1$

14.3.2 $Resp_{1A} = 1,5818$

...

14.3.2 $Resp_{1A} = 0,4117$

14.4 Decremento = Decremento - 1;

Decremento = $R30 - 1$;

...

Decremento = $R10 - 1$;

Fim

15. Decremento = $C_{Aux2} + 10$, $Aux_1 = 0$, $Resp_{1B} = 0$;

16. Enquanto ($C_{Aux2} - 10 \leq Decremento \leq C_{Aux2} + 10$) faça

Enquanto ($C10 \leq Decremento \leq C30$) faça

Início

16.1 $Aux_1 = \text{Módulo}(P_T \text{Menos } C_{1B} - Decremento)$;

$Aux_1 = \text{Módulo}((Resid_R8 + Resid_C92) + R8 - C18)$;

...

$Aux_1 = \text{Módulo}((Resid_R8 + Resid_C92) + R8)$;

...

$Aux_1 = \text{Módulo}((Resid_R8 + Resid_C92) + R8 + C2)$;

16.2 $Aux_1 = \text{Soma_Elementos}(Aux_1)$;

$Aux_1 = 1,7034$

...

$Aux_1 = 0,7329$

...

$Aux_1 = 0,8498$

16.3 Se ($Aux_1 < Resp_{1A}$) então

16.3.1 $C_{2A} = Decremento$

16.3.1 $C_{2A} = C30$

...

16.3.1 $C_{2A} = C12$

16.3.2 $Resp_{1A} = Aux_1$

16.3.2 $Resp_{1A} = 1,7034$

...

16.3.2 $Resp_{1A} = 0,7329$

Fim

17. Decremento = $C_{Aux3} + 10$, $Aux_1 = 0$, $Resp_{1C} = 25$;

18. Enquanto ($C_{Aux3} - 10 \leq Decremento \leq C_{Aux3} + 10$) faça

Enquanto ($I10 \leq Decremento \leq I30$) faça

Início

18.1 $Aux_1 = \text{Módulo}(P_T \text{Menos } C_{1C} - Decremento)$;

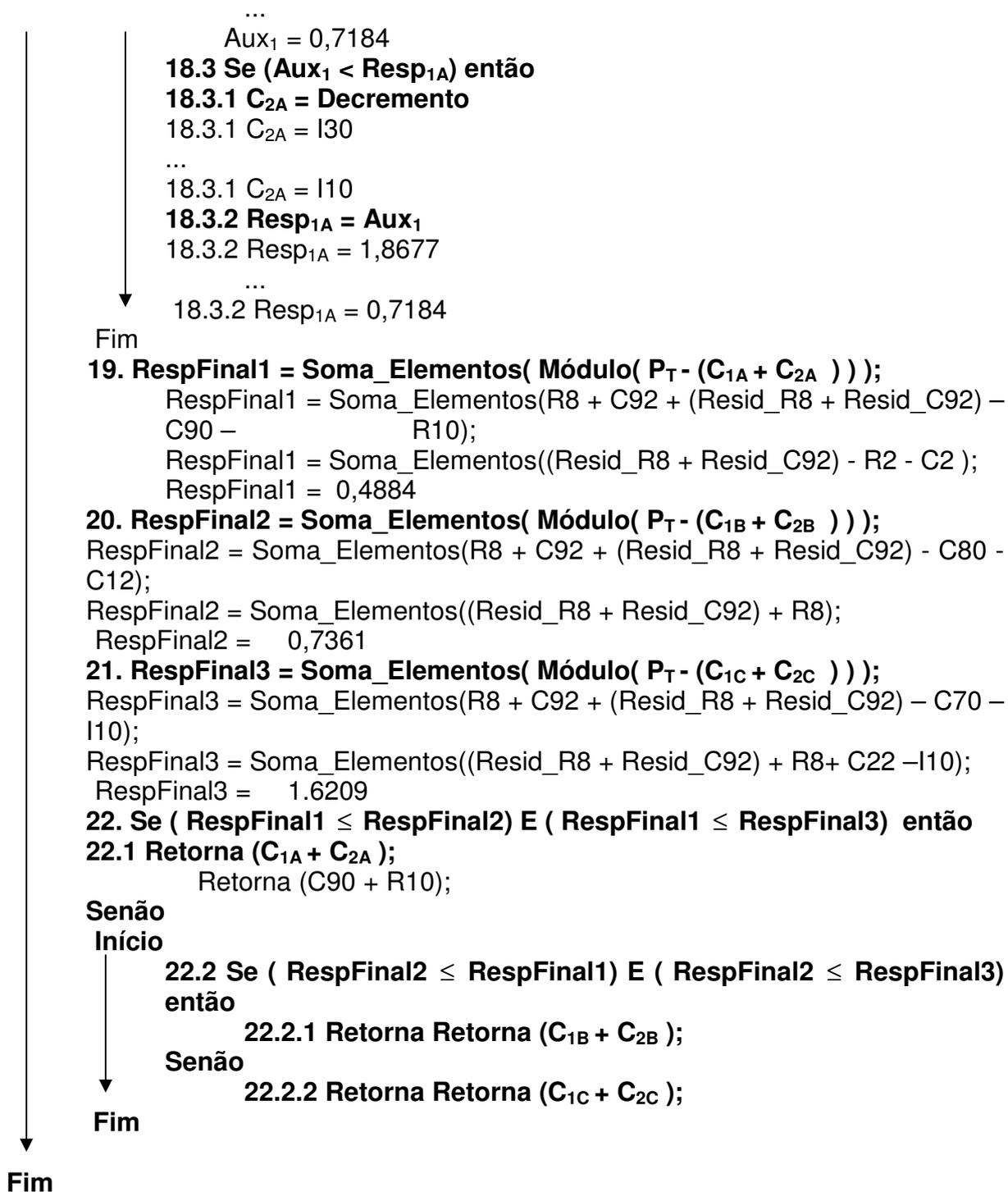
$Aux_1 = \text{Módulo}((Resid_R8 + Resid_C92) + R8 + C12 - I30)$;

...

$Aux_1 = \text{Módulo}((Resid_R8 + Resid_C92) + R8 + C12 - I10)$;

18.2 $Aux_1 = \text{Soma_Elementos}(Aux_1)$;

$Aux_1 = 1,8677$



Pode-se concluir que o padrão reconhecido, não é igual ao padrão de teste. Aparentemente isto inviabilizaria o modelo proposto, mas se quiser encontrar a resposta ótima, padrão de teste R8C92, pode-se diminuir o passo do algoritmo.

CAPÍTULO 7

7 CONCLUSÃO

Existem, nos dias atuais, diversos modelos implantados e propostos com a finalidade de auxiliar na operação e expansão dos sistemas elétricos de potência. Dentre esses modelos propostos, pode-se empregar diversas técnicas, como regressão linear simples, redes neurais artificiais, e em alguns casos, até sistemas híbridos.

Segundo Todesco em [26], o método numérico utilizando regressão linear pode representar pouca precisão em relação à necessidade de supervisão da rede elétrica de distribuição. Quando se utiliza redes neurais artificiais ou sistemas híbridos consegue-se bons resultados como pode ser constatado em [27]. No modelo utilizando RNA, pode-se observar eficiência no reconhecimento, imunidade a ruídos e facilidade de implementação. Um ponto importantíssimo é a definição do conjunto de treinamento, haja vista que para uma RNA aprender um determinado padrão, o mesmo tem que fazer parte do conjunto de treinamento. No caso de reconhecimento de padrões compostos, observa-se que o aumento da complexidade no treinamento provoca a queda da qualidade dos resultados gerados pela RNA treinada utilizando Backpropagation, enquanto que um sistema híbrido manteve uma certa constância [27].

O modelo proposto nessa dissertação têm o objetivo de ser mais uma alternativa viável para auxiliar as concessionárias de energia na operação e na expansão do sistema elétrico. Como mostrado anteriormente, através do estudo de caso no capítulo 6, o algoritmo Naive Bayes Modificado apresentou resultados

melhores se comparado às RNA's conforme proposto em [27]. Apesar de sua simplicidade, o Naive Bayes pode surpreender.

O Naive Bayes Modificado não tem a necessidade de ter um padrão composto no conjunto de treinamento para reconhecer um padrão semelhante. O conjunto de treinamento é composto por padrões simples, para todas as simulações envolvidas, não aumentando a complexidade no treinamento.

Vale ressaltar que o algoritmo gera o seu modelo de inferência rapidamente, comparado com uma RNA, que na maioria dos casos demora muito tempo para realizar o seu treinamento.

Portanto, este modelo proposto pode ser considerado uma alternativa para reconhecimento de curvas de consumo. No caso da concessionária de energia necessitar de uma precisão maior nas classes de consumidores, a solução é aumentar o espaço de busca.

Foi estudado o pacote computacional Weka, onde suas classes foram utilizadas para realizar as simulações. Para um trabalho futuro, pode-se desenvolver um sistema computacional comercial, já que uma parte do mesmo já está implementado. No entanto, falta definir como serão lidas as curvas de carga, que poderiam ser através de um banco de dados, ou ainda, em tempo real, através de placas de aquisição de dados.

Enfim, o algoritmo apresentado nesta dissertação, o Naive Bayes Modificado, pode ser considerado uma solução no reconhecimento de curvas de carga do sistema elétrico de potência.

8 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Meyer, Paul L.. Probabilidade Aplicações à Estatística. 2.ed. Livros Técnicos e Científicos Editora, 1983.
- [2] Lopes, Paulo Afonso. Probabilidades e Estatísticas. Rio de Janeiro: Reichmann & Afonso Editores, 1999.
- [3] Freund, John E, Garry A. Simon. Estatística Aplicada (Economia, Administração e Contabilidade). 9. ed. Rio de Janeiro: Bookman, 2000.
- [4] Bussab, Wilton de O, Morettin, Pedro A. Estatística Básica, 9. ed. Saraiva, 2004.
- [5] Ladeira, Marcelo – Diagrama de Influências Múltiplo Secionado. Tese de obtenção do grau de Doutor em Ciência da Computação. Porto Alegre: [sn], 2000.
- [6] Nilsson, Nil J. Artificial Intelligence S New Synthesis, Morgan Kaufmann Publihers, Inc. San Francisco, California, Stanford University, 1998.
- [7] Bussab, Wilton de O. Bussab, Morettin Pedro A. Estatística Básica,. 5.ed. Rio de Janeiro: Editora Saraiva, 2004.
- [8] Atlas de Energia Elétrica no Brasil / Agência Nacional de Energia Elétrica – Brasília: Aneel, 2000
- [9] Fayyad, U., G. Piatetsky-Shapiro, & P. Smyth(1996a). From data mining to knowledge discovery: an overview. Em Advances in Knowledge Discovery & Data Mining, pp. 1-34.
- [10] Batista, G. E. A. P. A., A. C. P. L. Carvalho, & M. C. Monard (2000). Applying one-sided selection to unbalanced datasets. Em Proceedings of the Mexican Congress on Articial Inteligence MICAI, Lecture Notes in Artificial Inteligence, pp. 315-325.
- [11] Weiss, S. M. & N. Indurkhya (1998). Predictive Data Minig: A Pratical Guide. Morgan Kaufmann Publishers, Inc., San Francisco, CA.
- [12] Glymour, C., D. Madigan, D. Pregibon, & P. Smyth (1997). Statistical themes and lessons for data mining. Data Mining and Knowledge Discovery1(1), 11-28.

- [13] Félix, L. C. M., S. O. Rezende, M. C. Monard, & C. W. Caulkins (2000). Transforming a regression problem into a classification problem using hybrid discretization. *Computación y Sistemas* 4, 44-52.
- [14] Kohavi, R., D. Sommerfield, & J. Dougherty (1996). Data mining using MLC++: A Machine Learning in C++. Em *Tools with Artificial Intelligence*, pp. 234-245. IEEE Computer Society Press.
- [15] Lavrac, N., P. Flach, & R. Zupan (1999). Rule evaluation measures: A unifying view. Em S. Dzeroski & P. Flach(eds.), *Proceedings of the Ninth International Workshop on Inductive Logic Programming(ILP-99)*, Volume 1634, pp. 74-185. Springer-Verlag. *Lecture Notes in Artificial Intelligence*.
- [16] Pazzani, M. J. (2000). Knowledge discover from data? *IEEE Intelligent Systems* 15(2), 10-13.
- [17] Pazzani, M., S. Mani, & W. Shankle (1997). Comprehensible knowledge discovery in databases. Em *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, pp. 596-601.
- [18] Fertig, C. S., A. A. Freitas, L. V. R. Aruda, & C. Kaestner (1999). A fuzzy beam-search rule induction algorithm. Em *Proceeding of the Third European Conference(PKDD-99)*, Volume 1704 of *Lecture Notes in Artificial Intelligence*, pp. 341-347.
- [19] Piatelsky-Shapiro, G. & J. Matheus(1994). The interestingness of deviations. Em *Proceedings of the International Conference on Knowledge Discovery and Data Mining(KDD-94)*, pp. 23-36.
- [20] Dong, G. & J. Li (1998). Interestingness of discovered association rules in terms os neighborhood-based unexpectedness. *Lectures Notes in Artificial Intelligence*,1394, 72-86.
- [21] Silberschatz, A. & A.Tuzhilin (1995). On subjective measure of interestingness in knowledge discovery. *Proceeding of the First Internacional Conference on knowledge Discovery and Data Mining* 1, 275-281.
- [22] - Rezende, Solange Oliveira, *Sistemas Inteligentes- Fundamentos e Aplicações*, 1. ed. São Paulo: Manole, 2003.

[23] Han, J., Kamber M., Data Mining – Concepts and Techniques, 1. ed. New York: Morgan Kaufmann, 2001.

[24] Witten, I.H., Frank, E., Data Mining – Practical Machine Learning Tools and Techniques with Java Implementation, 2000.

[25] WEKA - Machine Learning Algorithms in Java. Java Doc. Disponível em <www.cs.waikato.nz>. Acesso em 18/10/2004.

[26] Todesco, J.L, Pimentel F.J.S., A.L. Anais do XXIV Encontro Nacional de Engenharia de Produção, Florianópolis, SC, Brasil, 2004.

[27] Masselli, Yvo. M. C. Dissertação de Mestrado, Departamento de Engenharia Elétrica da UNIFEI, Itajubá, 2005.

[28] Lambert Torres, G. - “Fuzzy Techniques for Load Forecasting”, IEEE Power Engineering Society Summer Meeting, Vol. I, pp. 178-179, Seattle, USA, Jul. 16-20, 2000.

