

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**METODOLOGIA DE AVALIAÇÃO DA RELEVÂNCIA DE ATRIBUTOS EM
GRANDES BASES DE DADOS INCOMPLETAS UTILIZANDO
CONJUNTOS APROXIMADOS E LÓGICA PARACONSISTENTE**

SILVIA DAS DORES RISSINO

Itajubá, dezembro de 2009

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

SILVIA DAS DORES RISSINO

**METODOLOGIA DE AVALIAÇÃO DA RELEVÂNCIA DE ATRIBUTOS EM
GRANDES BASES DE DADOS INCOMPLETAS UTILIZANDO
CONJUNTOS APROXIMADOS E LÓGICA PARACONSISTENTE**

**Tese submetida ao Programa de Pós-Graduação em
Engenharia Elétrica como parte dos requisitos para
obtenção do Título de Doutor em Ciências em Engenharia
Elétrica.**

Área de Concentração: Sistemas Elétricos de Potência

Orientador: Professor Dr. Germano Lambert Torres

Co-Orientadora: Dr^a Helga Gonzaga Martins

**Dezembro de 2009
Itajubá-MG**

Ficha catalográfica elaborada pela Biblioteca Mauá –
Bibliotecária Margareth Ribeiro- CRB_6/1700

R596m

Rissino, Sílvia das Dores

Metodologia de avaliação da relevância de atributos em grandes bases de dados incompletos utilizando Conjuntos Aproximados e Lógica Paraconsistente / Sílvia das Dores Rissino. -- Itajubá, (MG) : [s.n.], 2009.

122 p. : il.

Orientador: Prof. Dr. Germano Lambert Torres.

Coorientadora: Profa. Dra. Helga Gonzaga Martins.

Tese (Doutorado) – Universidade Federal de Itajubá.

1. Grandes bases de dados. 2. Conjuntos Aproximados. 3. Lógica Paraconsistente. 4. Atributos relevantes. I. Torres, Germano Lambert, orient. II. Martins, Helga Gonzaga, coorient. III. Universidade Federal de Itajubá. IV. Título.



Ministério da Educação
UNIVERSIDADE FEDERAL DE ITAJUBÁ
Criada pela Lei nº 10435, de 24 de abril de 2002

FOLHA DE JULGAMENTO DA COMISSÃO EXAMINADORA DE TESE DE DOUTORADO

Título da Tese: **“Metodologia de Avaliação da Relevância de Atributos em Grandes Bases de Dados Incompletas Utilizando Conjuntos Aproximados e Lógica Paraconsistente”**

Autor: **Silvia das Dores Rissino**

JULGAMENTO

Examinadores	Conceito	Rubrica
1º	A	
2º	A	
3º	A	
4º	A	
5º	A	
6º	A	
7º	A	

A - **APROVADO**: Se a Tese obtiver aprovação unânime da Comissão Examinadora.

AC - **APROVADO CONDICIONAL**: Quando a Comissão Examinadora entender que a Tese deva sofrer alterações de conteúdo ou conter resultados novos ou adicionais.

RE - **REEXAME**: Quando a Comissão Examinadora entender que a Tese, embora contendo contribuições relevantes, não tenha sido defendida adequadamente, ou deve sofrer modificações mais profundas. Caso em que o candidato terá mais uma única e última oportunidade para reestruturar a Tese se for o caso, e defendê-la novamente em um prazo não superior a 6 (seis) meses contados a partir da data da primeira defesa.

R - **REPROVAÇÃO**: Quando a Comissão Examinadora entender que a Tese não contém contribuições relevantes para a área do tema da Tese.

(*) **Este documento terá a validade de 90 (noventa) dias a contar da data da defesa da Tese.**

Resultado Médio: Conceito: A, ou seja, Aprovado

Observações: _____

Itajubá, 04 de dezembro de 2009.

Prof. Dr. Horst Lazarek
1º Examinador – UF Desdren

Prof. Dr. Ronaldo Rossi
2º Examinador – UNESP

Prof. Dr. Alexandre Cesar Mendes Moreira
3º Examinador – MOABE

Prof. Dr. Carlos Henrique Valério de Moraes
4º Examinador – UNIFEI

Prof. Dr. Luiz Eduardo Borges da Silva
5º Examinador – UNIFEI

Prof. Dr. Helga Gonzaga Martins
6ª Examinadora – UNIFEI – (Co-Orientadora)

Prof. Dr. Germano Lambert Torres
7º Examinador- Orientador - UNIFEI

Prof. Dr. Edson da Costa Bortoni
Presidente da Banca - UNIFEI

“Com igual paixão eu procurei o conhecimento. Desejei compreender o coração dos homens. Desejei saber por que brilham as estrelas. E tentei apreender o poder Pitagórico pelo quais os números detêm o controle sobre o fluxo. Um pouco disto, mas não muito, consegui alcançar.”

Bertrand Arthur William Russell

Dedico este trabalho

Ao Morio companheiro, cúmplice e amigo.

Aos meus pais Ivete e Jorge.

AGRADECIMENTOS

Ao Professor Dr. Germano Lambert Torres pela paciência e compreensão na orientação deste trabalho.

À Dra. Helga Gonzaga Martins pela orientação, acolhida, amizade e incentivo.

Às Centrais Elétricas de Rondônia – CERON, pelo apoio para o desenvolvimento deste trabalho através do P&D.

Às Centrais Elétricas do Norte do Brasil – ELETRONORTE, em especial ao Dr. Milton Nunes da Silva Filho e a Engenheira Elaine Aparecida de Lima Vianna pelo apoio e colaboração no desenvolvimento deste trabalho.

Aos professores da Universidade Federal de Itajubá, em especial aos professores Dr. Luiz Eduardo Borges da Silva, Dr. Maurílio Coutinho e Dr. Carlos Valério de Moraes pelo incentivo e apoio.

Aos funcionários da Universidade Federal de Itajubá, em especial à Cristina e Margareth da Pós-graduação e a Regina da DRA.

Ao Sr. Lúcio Paulo Godoi Fermoselli da CGTI por toda atenção e ajuda, principalmente, nas várias viagens Porto Velho - Itajubá - Porto Velho.

A Universidade Federal de Rondônia, em especial aos professores do Núcleo de Ciência e Tecnologia e do Departamento de Informática. Obrigada muito especial à Professora Maria das Graças Viana, ao Professor Gilson Medeiros e Silva e ao Professor Tomas Daniel Menendez pela ajuda e incentivo na elaboração deste trabalho.

Aos meus irmãos Maria Ivete, Jorge, Luiz e Ana pelo incentivo.

Aos meus amigos Ivanda, Marcello, Fabiola, Beth e Nilson que sempre estiveram por perto incentivando.

Ao Professor Nicolas John Richardson por toda atenção, apoio e paciência com o inglês.

SUMÁRIO

Resumo	ix
Abstract.....	x
Lista de Siglas.....	xi
Lista de Figuras	xiii
Lista de Tabelas.....	xiv
1. Introdução.....	15
1.1 Visão Geral.....	15
1.2 Objetivos.....	15
1.2.1 Objetivo Geral.....	15
1.2.2 Objetivos Específicos.....	16
1.3 Metodologia.....	17
1.4 Relevância da Pesquisa.....	18
1.5 Organização do Trabalho.....	18
2. Descoberta de Conhecimento em Bases de Dados.....	20
2.1 Visão Geral.....	20
2.2. O que é Descoberta de Conhecimento?.....	20
2.3 Estágios Operacionais da Descoberta de Conhecimento - KDD.....	21
2.4 Problemas Existentes em Grandes Bases de Dados.....	22
2.5 O Uso de TCA e LP no Processo de Descoberta de Conhecimento.....	23
2.6 Considerações Finais.....	24
3. Teoria de Conjuntos Aproximados - TCA.....	25
3.1 Visão Geral.....	25
3.2 Aplicações da TCA.....	25
3.4 Conceitos Fundamentais e Representação dos Dados.....	26
3.5 Tabelas de Decisão e Algoritmos de Decisão.....	36
3.6 Dependência de Atributos.....	38
3.7 Reduto de Atributos em Sistema de Informação.....	41

3.8 Geração de Regras	45
3.9 Considerações Finais	46
4. Lógica Paraconsistente - LP	48
4.1 Visão Geral	48
4.2 Introdução à Lógica	48
4.3 Definição e Descrição da Lógica Paraconsistente – LP	50
4.4 Lógica Paraconsistente Anotada – LPA	52
4.4.1 Reticulado do Diagrama de Hasse	53
4.4.2 Análise do Reticulado do Diagrama de Hasse	55
4.4.3 Lógica Paraconsistente Anotada de Dois Valores - LPA2v.....	55
4.5 Aplicações da Lógica Paraconsistente.....	58
4.6 Considerações Finais	60
5. Ambiente do Problema: Base de Dados do Sistema de Informações Elétricas Acre/Rondônia – INFO_OPR.....	61
5.1 Visão Geral.....	61
5.2 Histórico do Sistema Elétrico Acre/Rondônia.....	61
5.3 Sistema de Gestão de Informação Elétricas INFO_OPR.....	63
5.4 Operação do Sistema de Gestão de Informação Elétricas INFO_OPR	65
5.5 Banco de Dados do Sistema INFO_OPR	65
5.5.1 Aspectos Lógicos do Banco de Dados INFO_OPR.....	66
5.5.2 Aspectos Físicos do Banco de Dados INFO_OPR	69
5.6 Principais Problemas da Base de Dados do INFO_OPR.....	69
5.7 Considerações Finais	70
6. Metodologia de Avaliação da Relevância dos Atributos de Grandes Bases de Dados Incompletas, Utilizando Teoria de Conjuntos Aproximados e Lógica Paraconsistente. 71	71
6.1 Visão Geral	71
6.2 Relação entre TCA e Banco de Dados Relacional - BDR.....	72
6.3 Considerações sobre o Banco de Dados Relacional do INFO_OPR.....	74
6.4 Aplicação da TCA no Banco de Dados Relacional INFO_OPR.....	77
6.5 Relação entre Banco de Dados Relacional e Lógica Paraconsistente.....	81
6.6 Aplicação de LP na Identificação e Classificação das Incompletudes dos Dados entre os Atributos Relevantes da Base de dados do INFO_OPR.....	85

6.7 Resumo da Metodologia de Avaliação de Atributos Relevantes de Grandes Bases de Dados Incompletas.	101
6.8 Considerações Finais	105
7. Conclusões e Trabalhos Futuros.....	107
7.1 Visão geral.....	107
7.2 TCA no BDR do Sistema de Ocorrências Elétricas – INFO_OPR	107
7.3 Lógica Paraconsistente no BDR do Sistema de Ocorrências Elétricas – INFO_OPR.....	109
7.4 Aplicação da Metodologia de Avaliação da Relevância de Atributos na Base de Dados do Sistema INFO_OPR	111
7.5. Principal Contribuição.....	112
7.6 Considerações Finais	113
7.6 Linhas de Pesquisa para Desenvolvimento de Trabalhos Futuros.....	114
Referências	115

Resumo

Os constantes avanços em tecnologia de informação possibilitam o armazenamento de grandes bases de dados históricas. O objetivo deste trabalho é apresentar uma metodologia que descubra um conjunto mínimo de atributos relevantes com a garantia da manutenção das características das informações. O desenvolvimento da tese está baseado nos trabalhos dos mais importantes pesquisadores da área de Teoria de Conjuntos Aproximados e Lógica Paraconsistente. Pretende-se desenvolver uma metodologia de avaliação da relevância de atributos de grandes bases de dados, onde o resultado é a seleção de um conjunto mínimo de atributos relevantes, que represente o conhecimento embutido na base de dados, através do uso da Teoria de Conjuntos Aproximados. A identificação e classificação das incompletudes dos dados é realizada através da aplicação da Lógica Paraconsistente. A metodologia proposta proporciona a descoberta de um conjunto reduzido de atributos relevantes com dados classificados, que permita uma extração de conhecimento rápida e confiável.

Palavras Chaves: Grandes Bases de dados, Conjuntos Aproximados, Lógica Paraconsistente, Atributos Relevantes, Inconsistente, Indeterminado.

Abstract

The constant advances in information technology make possible the storage of large historical databases. The aim of this work is to present a methodology that facilitates the discovery of a set of relevant attributes with a guarantee of the characteristics of the original data. The Methodology used in order to developed this thesis uses the most relevant works undertaken by researchers in the fields of Rough Set Theory and Paraconsistent Logic. It is intended to develop the methodology of evaluation of the relevant attributes in large databases that represent knowledge present through the use of Rough Set Theory. The identification and classification of incomplete data present in relevant attributes is carried out using Paraconsistent Logic. The methodology that is proposed permits the discovery of a reduced set of relevant attributes with classification data that allows the fast and reliable extraction of knowledge.

Key words: Large Database, Rough Set, Paraconsistent Logic, Relevant Attributes, Inconsistent, indeterminate.

Lista de Siglas

ANEEL – Agência Nacional de Energia Elétrica
BDR – Banco de Dados Relacional
CERON – Centrais Elétricas de Rondônia
ELETROBRÁS – Centrais Elétricas Brasileiras
ELETRONORTE – Centrais Elétricas do Norte do Brasil
INFO_OPR – Sistema de Informações de Ocorrências Elétricas
KDD – *Knowledge Discovery in Database* (Descoberta de Conhecimento em banco de Dados)
LP – Lógica Paraconsistente
LPA2v - Lógica Paraconsistente Anotada de Dois Valores
LPC – Linha Perfeitamente Consistente
LPI – Linha Perfeitamente Inconsistente
ONS – Operador Nacional do Sistema
PCH – Pequena Central Hidrelétrica
PIE – Produtor Independente de Energia
QUPC – Quadrado Unitário do Plano Cartesiano
SGBD – Sistema Gerenciador de Banco de Dados
SIN – Sistema Interligado Nacional
SQL – Structured Query Language (Linguagem Estruturada de Consulta)
TCA – Teoria de Conjuntos Aproximados
UHE – Usina Hidrelétrica

Siglas de Subestações

SEAE – Subestação Areal
SEAF – Subestação Alfaville
SEGJ – Subestação Guajará Mirim
SEAN – Subestação Abunã
SEAQ – Subestação Ariquemes
SEPV – Subestação Porto Velho
SERM – Subestação Rio Madeira

SESM – Subestação Samuel
SERL – Subestação Rolim de Moura
SEJP – Subestação Ji-Paraná
SEJR – Subestação Jaru

Siglas de Equipamentos

AT – Auto – Transformador
BC – Bancos Capacitores
BR - Barra
LT – Linha de Transmissão
RE – Reatores
SU – Equipamento da Usina Hidrelétrica de Samuel
TF – Transformador
TR – Transformador Regulador
UGD – Unidade Geradora à Diesel
UGG – Unidade Geradora à Gás
UGH – Unidade Geradora Hidráulica
UGV – Unidade Geradora à Vapor

Siglas de Unidades

kV – Kilo volt
MVA – Mega volt ampére
MW – Mega watt

Siglas de Níveis de Tensão

1 – 69 kV
2 – 230 kV

Lista de Figuras

Figura 2. 1. Estágios Operacionais do KDD	22
Figura 3. 1. Definição de Conjuntos de Aproximação e Região de Borda	31
Figura 4. 1. Reticulado - Diagrama de Hasse.....	54
Figura 4. 2. Quadrado Unitário do Plano Cartesiano – QUPC.....	57
Figura 4. 3. Esquema de um Sistema de Análise Paraconsistente.....	58
Figura 5. 1. Mapa com a Interligação do Sistema Elétrico Acre/Rondônia.....	63
Figura 5. 2. Mapa da Rede Lógica da ELETRONORTE com os Bancos de Dados.....	64
Figura 6. 1. Exemplo de Registros Consistentes	86
Figura 6. 2. Exemplo de Indeterminação por Ausência de Valores	86
Figura 6. 3. Área do QUPC Utilizada para Classificar Dados e Registros de Banco de Dados..	88
Figura 6. 4. Diagrama com o Resumo da Metodologia de Avaliação de Atributos Relevantes	105

Lista de Tabelas

Tabela 3. 1. Conjunto de Equipamentos com Respective Estados de Funcionamento	28
Tabela 3. 2. Tipos de Atributos e Respective Valores da Tabela 3.1	36
Tabela 3. 3. Dados de Equipamentos Sem o Atributo cl_ocorrendia.....	44
Tabela 3. 4. Dados de Equipamentos Sem o tp_ocorrendia	44
Tabela 5. 1. Tabelas do Banco de Dados INFO_OPR	67
Tabela 6. 1. Atributos da Tabela Ocorrência.....	75
Tabela 6. 2. Tabela com Atributos Relevantes.....	79
Tabela 6. 3. Resultado Parcial da Consulta com os Atributos Relevantes	81
Tabela 6. 4. Exemplo de Inconsistência entre os Dados dos Atributos Relevantes da Tabela Ocorrência.....	83
Tabela 6. 5. Exemplo de Indeterminação entre os Dados dos Atributos Relevantes.	84
Tabela 6. 6. Exemplos de Inconsistências entre os Dados dos Atributos Relevantes	87
Tabela 6. 7. Tabela com Valores - Verdade da Aplicação do Conectivo OU (∨).....	90
Tabela 6. 8. Atributos Relevantes com Numeração.	91
Tabela 6. 9. Exemplo de Dados para Classificação do Registro Utilizando LPA2v.....	91
Tabela 6. 10. Classificação dos Dados dos Atributos Relevantes do Primeiro Registro	92
Tabela 6. 11. Classificação dos Dados dos Atributos Relevantes do Quarto Registro.	93
Tabela 6. 12. Classificação dos Dados dos Atributos Relevantes do Quinto Registro	94
Tabela 6. 13. Classificação dos Dados dos Atributos Relevantes do Sexto Registro.	95
Tabela 6. 14. Classificação dos Dados dos Atributos Relevantes do Sétimo Registro.	95
Tabela 6. 15. Classificação dos Dados dos Atributos Relevantes do Oitavo Registro.....	96
Tabela 6. 16. Classificação dos Dados dos Atributos Relevantes do Nono Registro.....	96

1. Introdução

1.1 Visão Geral

Os constantes progressos em tecnologia de informação têm possibilitado o armazenamento de séries históricas em grandes bases de dados¹, mas a grande quantidade de dados dificulta a análise e interpretação com ferramentas tradicionais, como planilhas ou simples consultas ao banco de dados.

Descobrir e interpretar as relações embutidas nessas bases de dados proporciona conhecimentos relevantes às tomadas de decisão, sendo que, as descobertas e extrações de conhecimento devem ser realizadas em ambientes (base de dados) já tratados.

A análise pode ser realizada somente naqueles objetos do banco de dados que sejam realmente importantes, isto é, em vez de analisar todos os atributos de todas as tabelas, com seus respectivos valores de dados, sugere-se analisar somente aqueles atributos, com seus respectivos dados, que realmente representem o domínio de conhecimento em questão.

Neste trabalho, o problema a ser tratado é: *Como descobrir um conjunto mínimo de atributos relevantes, em grandes bases de dados incompletas*²?

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é apresentar uma metodologia híbrida, baseada em ferramentas matemáticas não-clássicas, que avalia a relevância dos atributos em grandes bases de dados incompletas, onde as incompletudes são identificadas e classificadas com objetivo de disponibilizar o estado dos dados e dos registros.

¹ Neste trabalho, base de dados e banco de dados serão usados como sinônimos.

² Neste trabalho, dados incompletos são todos aqueles que apresentam ausência de valores, inconsistência, indeterminação, incoerência e inconclusividade.

A base de dados a ser utilizada é a do Sistema de Informações Elétricas da ELETRONORTE - INFO_OPR, mais precisamente o subconjunto de dados dos Estados do Acre e Rondônia. A escolha deste subconjunto de dados justifica-se pelo fato de que o mesmo possui mais de um milhão de registros, o que lhe qualifica como uma grande base de dados [Nunes da Silva Filho, 2009].

A identificação dos atributos relevantes é efetivada através do uso de Teoria de Conjuntos Aproximados - TCA.

A identificação e classificação das incompletudes dos dados serão realizadas através do uso da Lógica Paraconsistente - LP.

O uso da TCA e da LP justificam-se, principalmente, por tornar viável a identificação de conjuntos mínimos de atributos, além de facilitar a identificação e classificação das incompletudes em grandes quantidades de dados, facilitando a descoberta do real estado dos dados e dos registros e, conseqüentemente, do banco de dados, antes mesmo do processo efetivo de descobertas e extrações de conhecimentos.

As definições e conceitos das ferramentas TCA e LP, apresentadas neste trabalho, não utilizarão rigores matemáticos, já que as descrições são, apenas, para orientar e facilitar o entendimento das respectivas ferramentas.

1.2.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Apresentar os conceitos básicos de descoberta de conhecimento em bases de dados;
- Apresentar a Teoria de Conjuntos Aproximados – TCA, seus conceitos e aplicações em várias áreas do conhecimento, mas com ênfase na aplicação em bancos de dados relacional;

- Apresentar a Lógica Paraconsistente – LP, seus conceitos e aplicações em várias áreas do conhecimento, mas com ênfase na aplicação em bancos de dados relacionais;
- Aplicar a Teoria de Conjuntos Aproximados em bancos de dados relacionais, com características temporais, mais precisamente na base de dados do Sistema INFO_OPR. O resultado da aplicação desta teoria possibilitará a descoberta de um conjunto mínimo de atributos relevantes que represente as características da informação embutida nesta base de dados;
- Aplicar Lógica Paraconsistente, na identificação e classificação das incompletudes existentes entre os dados e registros, dos atributos relevantes, da base de dados do Sistema INFO_OPR;
- Apresentar propostas de trabalhos futuros.

1.3 Metodologia

Esta pesquisa está direcionada à descoberta da resposta para a seguinte pergunta: *“Como descobrir um conjunto de atributos relevantes, em uma grande base de dados incompleta?”*.

A resposta para o problema está alicerçada na metodologia a ser empregada, a qual está baseada nos trabalhos de Teoria de Conjuntos Aproximados [Pawlak, 1982], [Pawlak, 1991], [Pawlak & Skowron, 1994], [Pawlak, 1998] e de [Lambert-Torres et al., 1999], que possibilitarão a descoberta dos atributos relevantes da base de dados.

Para identificação e classificação das incompletudes dos dados existentes nos atributos relevantes, o pilar da metodologia, está nos trabalhos de Lógica Paraconsistente de [Da Costa, 1991], [Da Costa, 1993]; [Abe, 1982]; [Bagai &

Sunderraman, 1995]; [Da Silva Filho, 1999], [Da Silva Filho et al., 2008] e de [Martins et al., 2007].

1.4 Relevância da Pesquisa

A análise de grandes quantidades de dados é inviável sem o auxílio de ferramentas adequadas, logo é imprescindível o desenvolvimento de ferramentas automáticas e inteligentes, objetivando interpretar os conhecimentos relevantes embutidos nos grandes bancos de dados.

A relevância desta pesquisa se apresenta através da proposta de uma metodologia híbrida, utilizando as ferramentas matemáticas não-clássicas Teoria de Conjuntos Aproximados e a Lógica Paraconsistente, a qual se aplica em grandes bases de dados, tendo como objetivo reduzir a quantidade de atributos a serem analisados, classificar e identificar as incompletudes dos dados, que posteriormente poderá facilitar a descoberta e extração de conhecimento.

1.5 Organização do Trabalho

O trabalho está estruturado em capítulos, os quais são apresentados da seguinte forma:

Capítulo 2 – Descoberta de Conhecimento em Bases de dados – Apresentar de forma sintética os conceitos de descoberta de conhecimento, tendo como objetivo mostrar os problemas para analisar grandes quantidades de atributos e as possíveis incompletudes existentes entre os dados;

Capítulo 3 - Teoria de Conjuntos Aproximados – apresenta os conceitos básicos e importantes da teoria;

Capítulo 4 – Lógica Paraconsistente – LP – apresenta os conceitos básicos e importantes desta lógica não clássica;

Capítulo 5 – Ambiente do problema. Apresenta e descreve a base de dados do Sistema de Informações de Ocorrências Elétricas – INFO-POR, onde a metodologia proposta será testada;

O Capítulo 6 – Apresentação da Metodologia de Avaliação da Relevância dos Atributos de Grandes Bases de Dados Incompletas, Utilizando Conjuntos Aproximados e Lógica Paraconsistente. Neste capítulo apresenta-se a descrição completa da metodológica;

Capítulo 7 – Conclusões e Trabalhos Futuros.

2. Descoberta de Conhecimento em Bases de Dados

2.1 Visão Geral

O contínuo desenvolvimento e progressos atualmente verificados, em áreas como a Tecnologia de Informação, Internet, SGBDs (Sistemas Gerenciadores de Banco de Dados), Dispositivos de Memória Secundária com grande capacidade de estocagem de dados, entre outros, têm proporcionado o armazenamento de dados em heterogêneas e robustas bases de dados.

As grandes bases de dados proporcionam quantidades astronômicas de conhecimento embutido entre os dados, mas para que este conhecimento seja descoberto são necessárias técnicas que possibilitem a classificação dos dados para posterior extração de conhecimento.

Este capítulo apresenta, de forma sintética, os conceitos de descoberta e extração de conhecimento, tendo como objetivo mostrar os problemas para analisar grandes quantidades de atributos e as possíveis incompletudes existentes entre os dados.

2.2. O que é Descoberta de Conhecimento?

A *descoberta de conhecimento* é um processo de transformação de dados em conhecimento, mas para compreender o processo de transformação é necessário entender alguns conceitos que serão discutidos neste capítulo.

A descoberta de conhecimento entre dados de uma base de dados é conhecida como KDD (*Knowledge Discovery in Databases*), isto é, um processo de extração de informações de base de dados, que cria relações de interesse, além de auxiliar a validação de conhecimento extraído [Fayyad et al., 1996b]. Hoje, o KDD é uma das áreas mais relevantes para informática, pois dá suporte, de maneira significativa, à gestão da informação.

As grandes bases de dados, atualmente, proporcionam um novo panorama de gestão e tomada de decisão, já que o patrimônio digital das empresas levanta alguns questionamentos: Por que investir em armazenagem de grandes quantidades de dados? Por que armazenar séries históricas? Como utilizar os dados históricos para uma boa gestão e tomada de decisão? As respostas somente são encontradas quando da análise dos dados disponíveis.

A análise dos dados faz parte da descoberta de conhecimento, que é um processo, que envolve várias fases operacionais, interativa, repetitiva e não trivial, a qual proporciona a identificação de padrões importantes e úteis a partir de um robusto volume de dados [Rissino & Lambert-Torres, 2009].

2.3 Estágios Operacionais da Descoberta de Conhecimento - KDD

Os estágios operacionais do KDD podem ser divididos em consolidação dos dados, seleção e pré-processamento, mineração dos dados, interpretação e avaliação, tendo como resultado final o conhecimento embutido entre os dados. A seguir serão descritos cada estágio do processo de descoberta de conhecimento.

→ Descrição dos Estágios Operacionais para Descoberta de Conhecimento

1. Consolidação dos Dados

Esta etapa tem como objetivo coletar os dados e consolidá-los para o processo de extração do conhecimento. Nesta etapa procura-se definir os atributos relevantes para o processo, definir os problemas e a forma de tratamento para os dados que não possuem valores e, se necessário remover as exceções [Watson, 2004].

2. Seleção e Pré-processamento

Esta etapa tem o objetivo de proporcionar qualidade aos dados, tratando-os para evitar possíveis distorções na extração do conhecimento e transformando-os, caso

necessário, em um formato para os dados que facilite e retire possíveis barreiras para a etapa de mineração de dados [Passos & Goldschmidt, 2005].

3. Mineração de Dados (*data mining*)

Nesta fase são escolhidos os algoritmos a serem utilizados na extração do conhecimento. Estes algoritmos podem realizar tarefas de classificação, agrupamento, regressão, associação e sumarização [Fayyad et al., 1996a].

4. Interpretação e Avaliação

Nesta etapa ocorre a análise dos resultados obtidos a fim de entender o modelo da etapa anterior [Passos & Goldschmidt, 2005].

A figura 2.1 apresenta, de forma macro, os estágios operacionais descritos anteriormente do processo de descoberta de conhecimento em bases de dados.

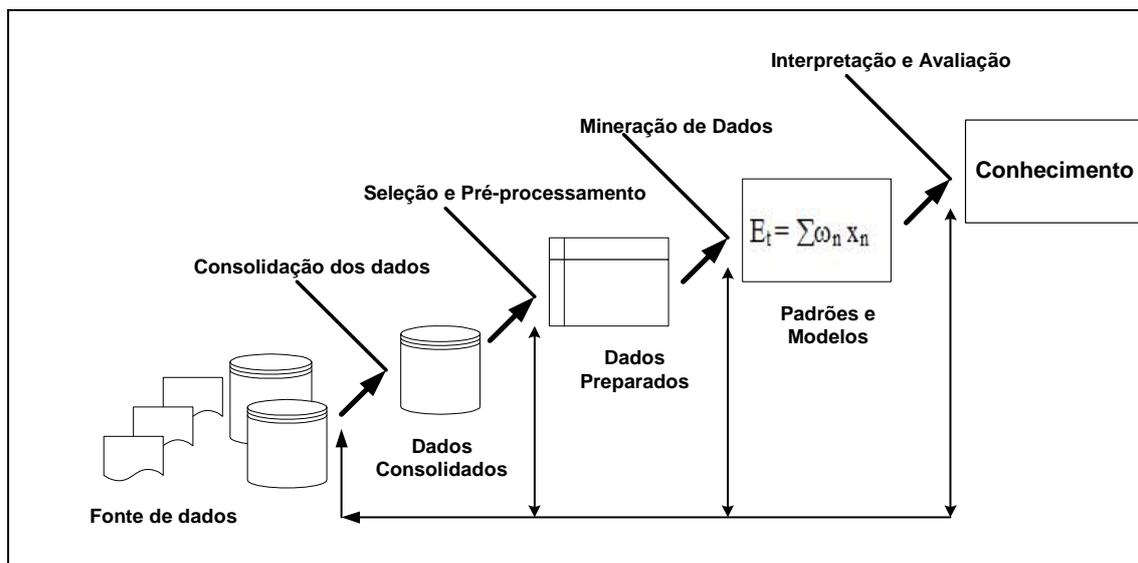


Figura 2. 1. Estágios Operacionais do KDD
Fonte: Adaptado de [Fayyad et al., 1996b]

2.4 Problemas Existentes em Grandes Bases de Dados

Os problemas para analisar as grandes quantidades de atributos e as possíveis incompletudes existentes entre os dados são um dos grandes desafios, encontrados por

todos os profissionais na área de Tecnologia de Informação, no processo de descoberta de conhecimento entre dados. Entre os problemas mais comuns encontra-se [Fayyad et al., 1996b]:

- Grande volume de dados heterogêneos;
- Valores ausentes entre os dados;
- Grande quantidade de atributos para análise;
- Integração de conjuntos de dados distintos;
- Inconsistência nos dados;
- Indeterminação entre dados.

2.5 O Uso de TCA e LP no Processo de Descoberta de Conhecimento

As ferramentas matemáticas não clássicas, Teoria de Conjuntos Aproximados - TCA e Lógica Paraconsistente – LP, são ferramentas que podem reduzir os problemas da descoberta de conhecimento em bases de dados, já que as mesmas podem ser utilizadas, entre outras coisas para:

1. Redução de variáveis (atributos) a serem analisados na base de dados;
2. Geração de regras como resultado da redução da quantidade de atributos;
3. Identificação das incompletudes nos dados;

4. Classificação das incompletudes, para possibilitar uma melhor compreensão e entendimento do problema e, no futuro, poder tratá-las.

2.6 Considerações Finais

Neste capítulo foram apresentados, de forma resumida, os conceitos e os estágios da descoberta de conhecimento, além dos problemas enfrentados pelos profissionais de Tecnologia de Informação.

O uso da TCA e da LP será de grande importância, principalmente, para dar suporte na solução de problemas apresentados na descoberta de conhecimento como a análise de grande volume de dados heterogêneos, grande quantidade de atributos para análise, além dos valores ausentes ou incompletudes entre os dados.

3. Teoria de Conjuntos Aproximados - TCA

3.1 Visão Geral

A Teoria de Conjuntos Aproximados (*Rough Sets*) - TCA, proposta no início dos anos 80, considerada como uma extensão da Teoria Clássica dos Conjuntos foi desenvolvida pelo cientista polonês Zdzislaw Pawlak, com objetivo de tratar incertezas e imprecisões dos dados [Pawlak, 1982].

A TCA está em constante desenvolvimento e relaciona-se à classificação e análise de informação imprecisa, incerta ou incompleta, sendo considerada a primeira abordagem não estatística para análise de dados, além de ser utilizada para reduzir as dificuldades na transformação de dados em conhecimento [Pawlak, 1991].

Os conceitos de TCA são matematicamente bem simples, pois envolvem conjuntos, relações e classes de equivalência. Ao transportar os conceitos de TCA para a realidade cotidiana, nota-se que o mundo real não é exato e nem preciso, já que os dados oriundos deste ambiente podem ser indiscerníveis ou incertos [Pawlak & Skowron, 1994].

Neste capítulo, são apresentadas algumas aplicações, bem como a forma como os dados são representados e os principais conceitos relacionados à TCA.

3.2 Aplicações da TCA

A TCA surgiu na década de 80 como uma abordagem matemática para gerenciar a incerteza, o vago e a incompletude da informação, permitindo a representação de conjuntos usando o conceito de aproximação inferior e superior. Nos últimos anos, a TCA vem sendo aplicada, com muito sucesso, em grandes áreas da ciência como [Hassan & Tazaki, 2005]:

- Aprendizagem de regras de decisão;
- Mineração de dados;
- Processamento de sinais (imagem e som);
- Sistemas especialistas e de suporte a decisão;
- Seleção de atributos relevantes;
- Triagem de sinal;
- Agrupamento (*Clustering*);
- Modelagem de Negócios;
- Reconhecimento de Padrões e Processamento de Informações;
- Extração de Conhecimento de Banco de Dados e;
- Análise de conflitos.

3.4 Conceitos Fundamentais e Representação dos Dados

Conjunto é uma definição essencial dentro da matemática e tem o significado de agrupamento de coisas ou objetos que possuam características semelhantes. Todos os objetos da matemática (relações, funções, números etc.) são considerados um tipo de conjunto. Mas, o conceito de conjunto na matemática é contraditório, já que se considera conjunto um “agrupamento” sem nenhum elemento, ou como é conhecido conjunto vazio.

Os componentes de um conjunto são chamados de elementos e a relação existente entre um elemento e um conjunto é chamada de relação de pertinência, a cardinalidade é a forma de medir a quantidade de elementos de um conjunto [Stoll, 1979].

Na teoria clássica, um conjunto é unicamente determinado por seus elementos, isto é, os elementos do conjunto devem ser classificados como pertencentes ou não ao conjunto. Logo, quando é possível enumerar os elementos de um conjunto, a este conjunto denomina-se exato (*crisp*), o que significa que todos os seus elementos são possíveis de serem descobertos, mas se a enumeração dos elementos do conjunto é imprecisa, denomina-se este conjunto de vago (*rough*) [Pawlak, 1998].

A TCA se baseia nos conceitos básicos de conjuntos de aproximação superior, inferior e região de borda, sendo que a representação dos dados usa uma estrutura simplificada, semelhante a um conjunto, sendo denominada de Sistema de Informação ou Tabela de Informação. Os conceitos básicos e os conceitos relevantes à teoria são apresentados a seguir:

a) Sistemas de Informação ou Tabela de Informação

Sistemas de informação são utilizados para representação dos dados a serem trabalhos pela TCA, pois são nestes sistemas que estão contidos os objetos, sendo que cada objeto tem uma quantidade de atributos [Lin, 1996].

As informações dos objetos são descritas no formato de uma tabela de dados, na qual as linhas são objetos para análise e as colunas são os atributos considerados [Wu et al., 2004].

O sistema de informação ou tabela de informações pode ser representado com o da Tabela 3.1, a qual é composta pelo conjunto B, que são todos os objetos ou registros

do sistema, sendo desta forma $B = \{E1, E2, E3, E4, E5, E6\}$. O conjunto de atributos³ condicionais é representado por $C = \{tp_ocorrencia, cl_ocorrencia, potencia_interrompida\}$ e o conjunto D representa o atributo decisão, sendo $D = \{Desligado\}$.

Tabela 3. 1. Conjunto de Equipamentos com Respetivos Estados de Funcionamento

Equipamento	Atributos Condicionais			Decisão desligado
	tp_ocorrencia	cl_ocorrencia	potencia_interrompida	
E1	1	2	Alta	Sim
E2	2	1	Alta	Sim
E3	2	2	Muito Alta	Sim
E4	1	2	Baixa	Não
E5	2	1	Alta	Não
E6	1	2	Muito Alta	Sim

Observações:

1. Os equipamentos E1(AEDJ201), E2(AESC204), E3(AFDJ202), E4(AQDJ601), E5(RLSC204), E6(JPDJ401), fazem parte do conjunto de equipamentos do Sistema Elétrico Acre/Rondônia;
2. O tipo de ocorrência no equipamento é representado pelo atributo tp_ocorrencia, o qual pode assumir os respectivos valores (1- Programada, 2- Forçada);
3. A classe de ocorrência dos equipamentos é representada pelo atributo cl_ocorrencia, o qual é representado pelos respectivos valores: 1-falha do equipamento e 2-causa externa sem retorno a operação;
4. A potência interrompida do equipamento está relacionado à quantidade de potência não disponibilizada pelo equipamento, em função da ocorrência, sendo representada pelo atributo potencia_interrompida, podendo assumir os respectivos valores: Baixa, Alta e Muito Alta;

³ Os nomes dos atributos tanto em tabelas/sistemas de informações ou banco de dados relacionais não devem ter acentos e nem caracteres especiais.

5. *tp_ocorrenca*, *classe_ocorrenca* e *potencia_interrompida* são nomes de atributos da base de dados INFO_OPR, logo não devem ser acentuados com quaisquer atributos em banco de dados.

b) Relação de Indiscernibilidade

Relação de Indiscernibilidade é um conceito nuclear em TCA, sendo uma relação entre dois ou mais objetos, onde todos os valores são iguais ou idênticos em relação a um subconjunto de atributos considerado.

A relação de indiscernibilidade é também uma relação de equivalência, onde todos os objetos similares de um conjunto são chamados elementares [Yao & Zhao, 2009]. Considerando os atributos condicionais da Tabela 3.1, tem-se $C = \{tp_ocorrenca, cl_ocorrenca, potencia_interrompida\}$, a relação de indiscernibilidade é dada por $IND_A(C)$. Como exemplo de relação de indiscernibilidade para o atributo *potencia_interrompida*, o qual gera três conjuntos elementares indiscerníveis que são:

$$IND_A(\{potencia_interrompida\}) = \{\{E_1, E_2, E_5\}, \{E_3, E_6\}, \{E_4\}\}$$

d) Aproximação

Aproximação está associada ao significado das operações topológicas de aproximação. Suponha que há um conjunto U de objetos chamado Universo e uma relação de indiscernibilidade $R \subseteq U \times U$, onde R é uma relação de equivalência. A seguir, são apresentados os conceitos fundamentais da TCA relacionadas à aproximação [Yao & Zhao, 2009]:

- **Aproximação inferior (B'')** de um conjunto X , em respeito a R é o conjunto de todos os objetos, o qual pode certamente ser classificado como pertencentes a X em relação R , isto é, conjunto B'' ;

- **Aproximação superior (B^*)** de um conjunto X com relação a R é o conjunto de todos os objetos os quais pode ser possivelmente classificados como pertencentes a X em relação a R , isto é, conjunto B^* ;
- **Região de Borda (RB)** de um conjunto X com relação a R é o conjunto de todos os objetos, os quais não **podem ser classificados nem com X ou como $\neg X$ com relação a R** . Se a região de borda for um conjunto $X = \emptyset$ (vazio), o conjunto é considerado “*Crisp*”, isto é, exato (preciso) em relação a R ; caso contrário, se a região de borda for um conjunto $X \neq \emptyset$ o conjunto X é considerado “*Rough*” (impreciso). De forma resumida, a região de fronteira é:

$$BR = B^* - B''$$

Para entendimento das definições anteriores, é dado um conjunto $X \subseteq U$, B é uma relação de equivalência e $K=(U, B)$ um base de conhecimento, os dois subconjuntos associados podem ser:

- B-inferior: $B'' = \bigcup \{Y \in U/B : Y \subseteq X\}$;
- B-superior: $B^* = \bigcup \{Y \in U/B : Y \cap X \neq \emptyset\}$

Da mesma forma, $POS(B)$, $BN(B)$ e $NEG(B)$ são definidos a seguir [Pawlak, 1991]:

- $POS(B) = B'' \Rightarrow$ certamente membro de X ;
- $NEG(B) = U - B^* \Rightarrow$ certamente não é membro de X ;
- $BR(B) = B^* - B'' \Rightarrow$ possivelmente é membro de X .

A figura 3.1 mostra uma representação gráfica para as regiões [Lambert-Torres et al., 1999].

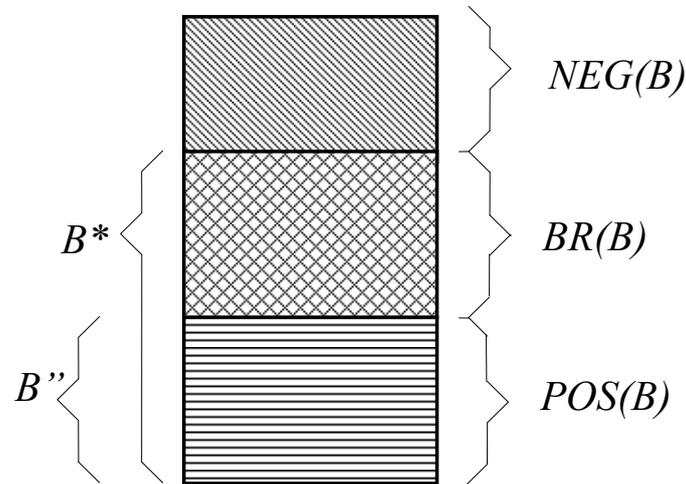


Figura 3. 1. Definição de Conjuntos de Aproximação e Região de Borda

Para apreender os conceitos de aproximação superior, inferior e região de borda, usar-se-á a Tabela 3.1, apresentada anteriormente, onde a mesma encontra-se classificada em relação ao atributo Decisão Desligado:

- Conjunto dos equipamentos que possivelmente estão desligados $\{E1, E2, E3, E6\}$, pois esses equipamentos possuem características para decisão do seu desligamento;
- Conjunto dos equipamentos que possivelmente estão ligados $\{E4\}$, pois esses equipamentos possuem características para decisão do seu não desligamento;
- Conjunto dos equipamentos que apresentam características inconclusivas, em relação ao seu desligamento $\{E2, E5\}$;
- Conjunto de aproximação inferior dos equipamentos desligados $B'' = \{E1, E3, E6\}$

- Conjunto de aproximação superior dos equipamentos desligados $B^* = \{E1, E2, E3, E6\}$. Neste caso, E2 faz parte do conjunto por apresentarem características de desligamentos, apesar de possuir dados inconclusivos, juntamente com o equipamento E5;
- Conjunto de aproximação superior dos equipamentos que com certeza não apresentam características de desligamento $\{E4\}$;
- Conjunto de aproximação superior dos equipamentos que não tem características de desligamento $B^* = \{E4, E5\}$, $\{E5\}$ faz parte deste conjunto por não apresentar características de desligamento, apesar de possuir dados inconclusivos juntamente com o equipamento E2;
- Conjunto Aproximação inferior dos equipamentos que não estão ligados $B'' = \{E4\}$;
- O conjunto formado pelos elementos da região de borda (RB), que são todos os elementos (equipamentos) que não podem ser classificados como ligado e como desligado, sendo representado por:

$$\text{Desligado RB} = B^* - B'' = \{E1, E2, E3, E6\} - \{E4\} = \{E1, E2, E3, E6\};$$

$$\text{Ligado RB} = B^* - B'' = \{E4, E5\} - \{E4\} = \{E5\}.$$

A Região de Borda (RB) é formada pela união dos conjuntos RB_Desligado e RB_Ligado, isto é, $(RB_Desligado \cup RB_Ligado) = \{E2, E5\}$. Já que, esses equipamentos possuem características inconclusivas em relação ao seu desligamento, pois os atributos condições são iguais, mas o atributo decisão possui valores divergentes.

→ **Qualidade das Aproximações** é obtida numericamente utilizando os próprios elementos que a definem, isto é, os elementos as aproximações inferior e superior. O coeficiente para medir a qualidade aproximação é representado por $\alpha_B(X)$, onde X é o conjunto de objetos ou registros com relação a B [Wu et al., 2004]. Considerando o exemplo da Tabela 3.1, os objetos ou registros de $B=\{E1, E2, E3, E4, E5, E6\}$, logo, o cálculo da qualidade da aproximação é apresentado por:

- **Coefficiente de imprecisão $\alpha_B(X)$** , onde α_B é a qualidade da aproximação de X , sendo denotado por:

$$\alpha_B(X) = \frac{|B''(X)|}{|B^*(X)|}$$

Onde $|B''(X)|$ e $|B^*(X)|$ representam a cardinalidade das aproximações inferior e superior, sendo que as mesmas são conjuntos $\neq \emptyset$. Logo, $0 \leq \alpha_B \leq 1$, se $\alpha_B(X) = 1$, X é um conjunto definível com relação aos atributos B , isto é, X é um conjunto preciso (“*crisp*”). Se $\alpha_B(X) < 1$, X é um conjunto impreciso (“*rough*”) com relação aos atributos B .

Observando-se a Tabela 3.1 e considerando o conceito de desligado, isto é, o conjunto de equipamentos que podem estar desligados é representado por $X=\{E1,E2,E3,E6\}$ e o conjunto de atributos condição é denotado por $B=\{cl_ocorrendia, tp_ocorrendia, potencia_interrompida\}$.

O conceito de desligado é aproximado ou impreciso (*rough*), porque $B''(X)=\{E1,E3,E6\} \neq \emptyset$ e $B^*(X)=\{E1,E2,E3,E6\} \neq U$. Neste caso, o $\alpha_B(X) = 3/4$, para os elementos que estão desligados, o que significa o conceito de desligado é caracterizado parcialmente através das condições $cl_ocorrendia$, $tp_ocorrendia$ e $potencia_interrompida$.

Quando se avalia apenas uma condição, como por exemplo, $cl_ocorrendia$, tem-se que $B''(X)=\emptyset$ e $B^*(X)=U$, o que significa que o conceito de desligado é totalmente indefinível apenas com a condição $cl_ocorrendia$, logo este atributo não é característico para desligamento. Contudo, um simples atributo como $B=\{potencia_interrompida\}$ é obtido através de $B''(X) = \{E3,E6\}$ e $B^*=\{E1,E2,E3,E6\}$, neste caso o $\alpha_B(X)=2/4=1/2$, mostra que nos casos em que os equipamentos apresentam $potencia_interrompida$ muito alta, isto é E3 e E6, com certeza esses equipamentos estão desligados, mas no caso dos equipamentos apresentarem $potencia_interrompida$ alta, há possibilidade dos equipamentos não estarem desligados, isto faz com que $potencia_interrompida$ possa ser característico para desligamento.

- **Coefficiente de qualidade da aproximação superior $\alpha_B(B^*(X))$** , é o percentual de todos os elementos que possivelmente são classificados como pertencentes a X, sendo denotado por:

$$\alpha_B(B^*(X)) = \frac{|B^*(X)|}{|A|}$$

Onde $|A|$ representa a cardinalidade do conjunto de objetos do sistema de informações, que neste caso, é representado pela Tabela 3.1. Logo $A \neq \emptyset$, sendo que para a tabela 3.1, têm-se duas situações:

Equipamentos Desligados: $\alpha_B(B^*(X)) = 4/6 = 2/3$;

Equipamentos ligados: $\alpha_B(B^*(X)) = 2/6 = 1/3$.

- **Coefficiente de qualidade da aproximação inferior $\alpha_B(B''(X))$** , é o percentual de todos os elementos certamente classificados como pertencentes a X, sendo denotado por:

$$\alpha_B(B''(X)) = \frac{|B''(X)|}{|A|}$$

Onde $|A|$ representa a cardinalidade do conjunto de objetos do sistema de informações, que neste caso, é representado pela Tabela 3.1. Logo $A \neq \emptyset$, sendo que para a Tabela 3.1, têm-se duas situações:

Equipamentos Desligados: $\alpha_B(B^*(X)) = 3/6 = 1/3$ de todos os equipamentos certamente estão desligados;

Equipamentos Ligados: $\alpha_B(B^*(X)) = 1/6 =$ dos equipamentos certamente estão ligados.

Observação

10% dos equipamentos (E2 e E5), não podem ser classificados nem como equipamento desligado e nem como equipamento ligado, já que as características de todos os atributos são iguais, apenas o atributo decisão “Desligado” não é idêntico, o que permite o surgimento de uma inconclusão para o estado do equipamento na ocorrência. Os equipamentos E2 e E5 fazem parte do conjunto de região de borda. Logo, quanto maior a quantidade de dados pertencentes ao conjunto região de borda, maior a quantidade de dados inconsistentes e indeterminados.

→ **Propriedades das Aproximações [Pawlak, 1982]:**

$$a) R''(X) \subseteq X \subseteq R^*(X)$$

$$b) R''(\emptyset) = R^*(\emptyset) = \emptyset; R''(U) = R^*(U) = U$$

$$c) R^*(X \cup Y) = R^*(X) \cup R^*(Y)$$

$$d) R''(X \cap Y) = R''(X) \cap R''(Y)$$

$$e) R''(X \cup Y) \supseteq R''(X) \cup R''(Y)$$

$$f) R^*(X \cap Y) \subseteq R^*(X) \cap R^*(Y)$$

$$g) X \subseteq Y \rightarrow R''(X) \subseteq R''(Y) \ \& \ R''(X) \subseteq R''(Y)$$

$$h) R''(-X) = R^*(X)$$

$$i) R^*(-X) = -R''(X)$$

$$j) R''R''(X) = R^*R''(X) = R''(X)$$

$$k) R^*R^*(X) = R^*R''(X) = R''(X)$$

$$l) R^*R^*(X) = R''R^*(X) = R^*(X)$$

3.5 Tabelas de Decisão e Algoritmos de Decisão

Os sistemas de informações, dentro da TCA, são compostos por atributos de duas classes, sendo que uma delas é a classe de atributos condicionais, enquanto a outra é a classe dos atributos de decisão; como exemplo tem-se a Tabela 3.2, onde são apresentados os tipos de atributos da Tabela 3.1 com seus respectivos valores.

Tabela 3. 2. Tipos de Atributos e Respetivos Valores da Tabela 3.1

	Atributos	Valores Nominais
Atributos Condicionais	tp_ocorrendia	1,2
	cl_ocorrendia	1,2
	Potencia_interrompida	Baixa, Alta, Muito Alta
Atributo Decisão	Desligado	Sim, Não

As linhas ou registros de um sistema de informação determinam regras que especificam ações/decisões que são tomadas em função dos valores contidos, em cada atributo. Por exemplo, as regras que são definidas para os equipamentos E2 e E5, são iguais em relação aos atributos condicionais, mas diferente para o atributo decisão. Para este tipo de regra denomina-se regra inconclusiva ou conflitante, pois possui atributos condicionais com valores iguais e atributo decisão divergentes.

As regras encontradas sobre os dados da tabela, retirando os dados inconclusivos, são denominadas de regras certas ou simplesmente regras, caso contrário as regras são inconsistentes.

As tabelas de decisões que possuem regras inconsistentes são chamadas de *tabelas inconsistentes*, caso contrário, é chamada de *tabela consistente*.

O número de regras consistentes dentro do conjunto de todas as regras é usado como fator de consistência para a tabela de decisão. O fator de consistência é expresso por $\gamma(C,D)$, onde C e D são as condições e a decisões respectivamente. Se $\gamma(C,D)=1$, a tabela decisão é consistente e $\gamma(C,D) \neq 1$, a tabela decisão é inconsistente.

Na Tabela 3.1, tem-se que $\gamma(C,D)=4/6$, isto é, dos seis registros da referida tabela, quatro registros podem ser transformados em regras conclusivas e dois registros são transformados em regras inconclusivas, pelos motivos já discutidos anteriormente. Como exemplo de uma regra de decisão, pode-se considerar a regra gerada pelos dados do equipamento E_3 :

Se

tp_ocorrendia = 2 E

cl_ocorrendia = 2 E

potencia_interrompida = muito alta

Então

equipamento desligado = sim.

O Algoritmo de Decisão é um conjunto de regras extraídas de uma tabela de decisão, sendo que para cada tabela de decisão associa-se um algoritmo de decisão [Pawlak & Skowron, 1994].

A principal diferença entre tabela de decisão e algoritmo de decisão está relacionada ao conteúdo da análise, já que uma tabela de decisão possui uma coleção de

dados e o algoritmo decisão apresenta uma coleção de implicações ou expressões lógicas [Rissino & Lambert-Torres, 2009]. Logo, para analisar dados de uma tabela, são necessários métodos matemáticos, como, por exemplo, os métodos estatísticos; mas para analisar implicações ou expressões lógicas é necessário o uso de ferramentas lógicas, como por exemplo, as lógicas clássicas e não clássicas.

3.6 Dependência de Atributos

O conteúdo de uma tabela de dados ou sistema de informação é composto por um conjunto de atributos, os quais podem ser independentes entre si ou não.

A descoberta de relações de dependência entre os atributos é muito importante para a análise de dados, já que a interdependência entre os mesmos indica quanto um atributo pode ser relevante ou não na descoberta de conhecimento embutido na tabela [Grzymala et al., 1990]. Para facilitar o entendimento de dependência de atributos considere-se o seguinte:

- D é um conjunto de atributos totalmente dependente do conjunto de atributos C, isto é, $C \Rightarrow D$;
- D depende totalmente de C se existir uma dependência funcional entre valores de D e C;

Para exemplificar, são analisados os dados da Tabela 3.1, em relação ao atributo condicional potencia_interrompida e o atributo decisão Desligado. Verifica-se que, no caso dos equipamentos E2 e E5, os atributos condição, apesar de serem iguais em termos de conteúdo de valores, há divergência nos valores do atributo decisão; para registros com valores iguais, resulta em uma decisão diferente, o que acarreta em uma informação inconclusiva. Logo, esses registros devem ser ignorados, por serem inconclusivos. Mas, analisando os registros dos outros equipamentos percebe-se que existe total dependência entre $\{potencia_interrompida\} \Rightarrow \{Desligado\}$, pois para cada

valor do atributo condição potencia_interrompida, haverá um único valor para o atributo decisão Desligado.

Quando a dependência funcional entre os atributos não é total, é necessário descobrir a dependência parcial entre os mesmos, pois a descoberta de conhecimento escondida entre os atributos está relacionada ao grau de dependência dos atributos. Para exemplificar esta situação, analisam-se os dados da Tabela 3.1, em relação ao atributo condicional potencia_interrompida e o atributo decisão desligado.

O atributo potencia_interrompida, determina unicamente e somente, alguns valores relacionados ao atributo desligado, isto é, (potencia_interrompida, muito alta) implica (desligado, Sim), (potencia_interrompida, alta) implica (desligado, Sim), mas nem sempre (potencia_interrompida, baixa) implica em (desligado, Sim), por isso, a dependência parcial apresenta que, somente alguns valores de {potencia_interrompida} implicam em valores de {Desligado}, isto é, alguns valores do conjunto D são determinados pelos valores do conjunto C.

A dependência entre atributos pode ser formalmente definida em função do grau de pertinência K , isto é, dado o conjunto D e o conjunto C, definidos anteriormente, onde D e C é subconjunto do conjunto A, tem-se a seguinte situação: D depende de C em grau K se ($0 \leq K \leq 1$), o que pode ser denotado por:

$$C \Rightarrow_k D, \quad \begin{array}{l} \text{se } K = \gamma(C,D); \\ \text{se } K = 1, D \text{ é totalmente dependente de } C; \\ \text{se } K < 1, D \text{ é parcialmente dependente de } C; \end{array}$$

O coeficiente k expressa a razão de todos os elementos do universo que podem ser classificados como parte da partição U/D , onde U é o conjunto universo em questão e D é o conjunto de dados dos atributos Decisão, consideram-se os dados de C, isto é, do conjunto de atributos condição.

O conceito de dependência de atributos está fortemente acoplado ao conceito de consistência de uma tabela de decisão, pois se D depende totalmente de C então $I(C) \subseteq I(D)$, o que significa que, as partições geradas por C são menores que as partições geradas por D. Se D depende em grau de K de C, onde $0 \leq K \leq 1$, então:

$$\gamma(C, D) = \frac{|POS_C(D)|}{|U|}$$

Onde,

$$POS_C(D) = \bigcup_{X \in U/I(D)} C_*(X)$$

A expressão $POS_C(D)$ é a região positiva da partição U/D, em relação a C, isto é, o conjunto de todos os elementos de U que podem ser unicamente classificados no bloco da partição U/D. Logo, D é totalmente ou parcialmente dependente de C, se todos ou alguns elementos do universo de U podem ser unicamente classificados na partição U/D, levando-se em consideração o conjunto C. O conceito de dependência entre atributos, apresentado, corresponde ao mesmo conceito de dependência utilizado em Banco de Dados Relacionais [Hongai et al., 2005].

Para ilustrar esses conceitos, utilizam-se os dados da Tabela 3.1, para dependência $\{tp_ocorrencia, cl_ocorrencia, potencia_interrompida\} \Rightarrow \{desligado\}$, se obtém $K=4/6=2/3$, isto significa que dos seis equipamentos relacionados na tabela, apenas 2/3 podem ser classificados como desligados ou não, pois dois equipamentos têm características inconclusivas para desligamento. Esta análise leva em consideração todos os atributos condição da Tabela 3.1. Caso se leve em consideração apenas o atributo $potencia_interrompida$, isto é, $\{potencia_interrompida\} \Rightarrow \{desligado\}$, o resultado é $K=2/6=1/3$, porque dois equipamentos {E3,E6}, do total de seis, podem ser classificados como desligados.

Um dos conceitos centrais, neste trabalho, é a relação de dependência entre atributos, já que o mesmo, junto com outros, proporcionam a base para o desenvolvimento da solução ao problema proposto.

3.7 Reduto de Atributos em Sistema de Informação

Analisar grande quantidade de dados, ainda, é um dos desafios para aqueles que trabalham com descoberta de conhecimento. Diante disso, uma das opções usadas é a redução de atributos do sistema de informação, desde que a redução preserve as propriedades fundamentais do domínio em questão.

Para exemplificar, utilizam-se os dados da Tabela 3.1, ao observar os atributos desta tabela, verifica-se que o conjunto de dados oriundos dos atributos *tp_ocorrenci*a e *cl_ocorrenci*a é equivalente ao conjunto original, em relação à dependência e a aproximação. Já que apresenta a mesma precisão da aproximação e grau *K* de dependência da tabela dada, contudo o conjunto de atributos que será usado para análise é menor que o original. Matematicamente, a idéia descrita é apresentada a seguir [Pawlak, 1991]:

- *B* é subconjunto de *A* e *a* pertence a *B*, logo *a* é dispensável em *B* se $I(B) = I(B - \{a\})$, caso contrário *a* é indispensável em *B*.
- O conjunto *B* é independente se todos os atributos são indispensáveis;
- O subconjunto *B'* de *B* é um reduto de *B* se *B'* é independente e $I(B') = I(B)$.

Um reduto é um conjunto de atributos que preserve as propriedades da partição, isto significa que o reduto é o subconjunto mínimo de atributos que permitem a mesma classificação dos elementos do Universo (domínio) em relação a todos os atributos do conjunto [Lambert-Torres et al., 1999].

Como neste trabalho, não se utiliza formalismo matemático, são apresentadas, apenas, as duas propriedades importantes para o suporte do desenvolvimento da solução [Ganesan et al., 2007]:

1. Core de atributos: Seja B um subconjunto de A , o *Core* de B é um subconjunto de todos os atributos indispensáveis de B , que é denotado por $Core(B)$.

2. Relação entre Core e Reduto de Atributos: A conexão entre o conceito de *Core* e o conceito de Reduto de Atributos é apresentado da seguinte forma: $Core(B) = \bigcap Red(B)$, onde $Red(B)$ é o reduto de B , isto é, o menor conjunto de atributos que representam o domínio em questão, sem que o mesmo perca as propriedades do conjunto original. Isso ocorre porque, o *Core* é a interseção de todos os redutos, o que o inclui entre todos os redutos, logo cada elemento do *Core* pertence a algum reduto. Em função do exposto, o *Core* é o mais importante dos subconjuntos de atributos, já que nenhum dos elementos pode ser removido, sem afetar a classificação dos atributos.

O uso do conceito de *Core* é dual, pois pode ser primeiramente usado como base para computação de todos os redutos, já que o mesmo é incluído em muitos redutos e sua computação é simples e direta. Além do que, o *Core* pode ser interpretado como o conjunto da parte mais característica do domínio, que não pode ser eliminado ao reduzir atributos.

Na simplificação de uma tabela ou sistema de informação, pode-se eliminar, inicialmente, os valores de atributos supérfluos, como é apresentado a seguir:

1. O valor do atributo $a \in B$ é dispensável para x , se $[x]_{I(B)} = [x]_{I(B - \{a\})}$, caso contrário o valor de do atributo a é indispensável para x ;
2. Se para todo atributo $a \in B$ o valor de a é indispensável para x , então B será chamado de ortogonal para x ;

3. O subconjunto $B'' \subseteq B$ é um reduto de valor de B em x, se, e somente se, B'' é ortogonal para x e $[x]_{I(B)} = [x]_{I(B'')}$;

O conjunto de todos os valores indispensáveis para o atributo em B para x será chamado de *Core* de valor de B para x e será denotado por $Core^*(B)$. Neste caso tem-se que $Core^*(B) = \cap Red^*(B)$, onde $Red^*(B)$ é o conjunto de todas as famílias de redutos de B em x.

Dada à dependência $C \Rightarrow D$, isto acontece porque o conjunto D depende não do conjunto C, mas do seu subconjunto C' , logo é importante encontrar este subconjunto, sendo que, para resolver este problema é necessário a definição de um reduto relativo, que será discutido a seguir.

Dado $C, D \subseteq A$, e se $C' \subseteq C$ é um D-reduto de C, então C' é um subconjunto mínimo de C tal que: $\gamma(C,D) = \gamma(C',D)$. Logo:

1. O atributo $a \in C$ é D-dispensável em C, se $POS_C(D) = POS_{(C-\{a\}}(D)$; caso contrário o atributo a é D-indispensável em C;
2. Se todos os atributos $a \in C$ são C-indispensável em C, então C será chamado de D-independente;
3. O Subconjunto $C' \subseteq C$ é um D-reduto de C, se e somente se C'' é D-independente e $POS_C(D) = POS_{C'}(D)$.

O conjunto de todos os atributos D-indispensáveis em C será chamado de *D-Core* de C, sendo denotado por $CoreD(C)$, neste caso, tem-se a seguinte propriedade: $CoreD(C) = \cap RedD(C)$, onde $RedD(C)$ é a família de todos os D-redutos de C.

Um reduto é um conjunto de dados mínimos, semelhante ao conceito de amostra usado em estatística, desde que sejam mantidas as propriedades de conhecimento do sistema de informação, quando o mesmo é obtido, utilizando-se todos os seus atributos.

Logo, o reduto deve ter a capacidade de classificar objetos ou classes, sem alterar a forma de representar o conhecimento [Pawlak, 1998].

Para exemplificar, utiliza-se a Tabela 3.1, onde há dois redutos relativos em relação a desligado: {tp_ocorrenci, potencia_interrompida} e {cl_ocorrenci, potencia_interrompida}. Neste caso, o atributo tp_ocorrenci ou atributo cl_ocorrenci podem ser descartados da análise da Tabela 3.1 e conseqüentemente em vez de usar a referida tabela, usar-se-á a Tabela 3.3 e Tabela 3.4 a seguir:

Tabela 3. 3. Dados de Equipamentos Sem o Atributo cl_ocorrenci

Equipamentos	Atributos Condicionais		Decisão Desligado
	tp_ocorrenci	Potencia_interrompida	
E1	1	Alta	Sim
E2	2	Alta	Sim
E3	2	Muito Alta	Sim
E4	1	Baixa	Não
E5	2	Alta	Não
E6	1	Muito Alta	Sim

Para a Tabela 3.1, o *Core* relativo em relação ao conjunto {tp_ocorrenci, cl_ocorrenci, potencia_interrompida} é a potencia_interrompida, isto mostra que o atributo potencia_interrompida habilita um diagnóstico parcial para equipamentos desligados.

Tabela 3. 4. Dados de Equipamentos Sem o tp_ocorrenci

Equipamento	Atributos Condicionais		Decisão Desligado
	cl_ocorrenci	potencia_interrompida	
E1	2	Alta	Sim
E2	1	Alta	Sim
E3	2	Muito Alta	Sim
E4	2	Baixa	Não
E5	1	Alta	Não
E6	2	Muito Alta	Sim

3.8 Geração de Regras

Com os redutos encontrados, para um dado sistema de informação, há possibilidade de geração de regras utilizando o referido reduto, as quais podem facilitar a extração de conhecimento e a própria interpretação dos dados analisados. A maior vantagem, desta metodologia é que não se altera as características do sistema de informação proposto inicialmente. A seguir são apresentadas as regras para a Tabela 3.1:

→ Regras para equipamentos desligados

R1 para o E3

Se

tp_ocorrendia = 2 E

cl_ocorrendia = 2 E

potencia_interrompida = muito alta

Então

Equipamento desligado;

R2: para o E6

Se

tp_ocorrendia = 1 E

cl_ocorrendia = 2 E

potencia_interrompida = muito alta

Então

Equipamento desligado;

→ Regras inconclusivas - equipamentos sem definição de desligamento

R3: para o E2

Se

tp_ocorrendia = 2 E

cl_ocorrendia = 1 E

potencia_interrompida = alta

Então

Equipamento desligado;

R4: para o E5

Se

tp_ocorrendia = 2 E

cl_ocorrendia = 1 E

potencia_interrompida = alta

Então

Equipamento ligado;

→ Regras para equipamentos ligados

R5: para o E4

Se

tp_ocorrendia = 1 E

cl_ocorrendia = 2 E

potencia_interrompida = Baixa

Então

Equipamento desligado;

Observação

As regras *R3* e *R4*, são inconclusivas para desligamento de equipamento, pois apresentam as mesmas características nos atributos condicionais, mas com decisão diferente para o diagnóstico desligado. Logo, *R3* e *R4* devem ser descartadas para essa análise dos dados.

3.9 Considerações Finais

Neste capítulo, apresentou-se a Teoria de TCA (*Rough Sets*), proposta no início dos anos 80 pelo matemático polonês Zdzislaw Pawlak, sendo considerada como uma

extensão da Teoria Clássica dos Conjuntos. O objetivo da TCA é tratar incertezas e imprecisões entre dados, através da classificação e análise de informação imprecisa, incerta ou incompleta.

A vantagem da TCA é não necessitar de qualquer informação preliminar ou adicional sobre dados, como ocorre na estatística e na probabilidade, além de ser a primeira abordagem não estatística para tratamento de inconsistências de dados.

Esta teoria de conjuntos permite o desenvolvimento de algoritmos eficientes que facilitam a descoberta de padrões escondidos entre os dados; além de produzir redutos que facilitam o processamento da informação, possibilitando a geração de regras de decisão, que possibilitam a interpretação direta dos resultados obtidos sem necessidade de outras ferramentas auxiliares.

A TCA apresenta excelentes resultados, principalmente, quando utilizada de forma híbrida com outras teorias não clássicas que tratam as inconsistências em dados.

4. Lógica Paraconsistente - LP

4.1 Visão Geral

Os estudos da lógica, de forma geral, são baseados na Lógica Clássica, a qual tem seus fundamentos relacionados ao cálculo dos predicados clássicos de 1ª ordem, ocorrendo sua extensão na Teoria dos Conjuntos, Teoria dos Tipos e na Teoria das Categorias [Priest, 2008]. Neste capítulo, apresenta-se uma introdução resumida de lógica clássica, além dos conceitos, definições, classificações e aplicações da Lógica Paraconsistente.

4.2 Introdução à Lógica

A Lógica Clássica é considerada como o núcleo da lógica dedutiva, compreende o cálculo dos predicados de primeira ordem com identidade e símbolos funcionais. Pode ser caracterizada por uma relação de consequência, definida sintática e semanticamente, sendo que, a lógica clássica é baseada em quatro princípios, apresentados a seguir [Vilela, 1964]:

- Princípio da Identidade;
- Princípio do Terceiro Excluído;
- Princípio da Não-Contradição;
- Princípio da Identidade proposicional

Com base nesses princípios, as proposições simples são verdadeiras ou falsas, sendo os dois casos mutuamente exclusivos. Logo, isto implica que a lógica clássica é bivalente.

A lógica Clássica é um tipo de lógica, talvez seja a mais conhecida, mas além dela existem outros tipos de lógicas como as Lógicas Não Clássicas, que são formadas pelas Lógicas Complementares às Lógicas Clássicas e Lógicas Heterodoxas.

Nas lógicas complementares à Clássica, além dos três princípios da lógica clássica (Princípio da Identidade, Princípio da Contradição e o Princípio do Terceiro Excluído), possuem, ainda, outros princípios que as regem, estendendo o seu domínio.

As lógicas complementares adicionam aspectos que a lógica clássica não consegue expressar, isto é, a variação da veracidade da proposição no decorrer do tempo, que neste caso utiliza a lógica temporal para expressar essa variação. Como exemplos de Lógicas Complementares tem-se a: Lógica Epistêmica, Lógica Temporal ou Lógica do Tempo e a Lógica Modal [Mortari, 2001].

As Lógicas Heterodoxas são, também, lógicas não clássicas, pois se caracterizam por derogarem ou modificarem algum ou alguns dos princípios fundamentais da lógica clássica. O grupo de lógicas heterodoxas é composto por algumas lógicas entre as quais se destaca a Lógica Fuzzy e a Lógica Paraconsistente, as quais derogam alguns dos princípios da Lógica clássica, como o princípio da contradição [Priest, 2008].

A Lógica Fuzzy é uma extensão da lógica booleana a qual admite infinitos valores lógicos intermediários do intervalo $[0,1]$, isto é, Falso (0) e Verdadeiro (1), como também podem admitir valores do tipo talvez (0,6). A lógica Fuzzy trabalha com o conceito de graus de pertinência [Zadeh, 1965].

A Lógica Paraconsistente é uma das lógicas desenvolvidas, como alternativa aos princípios clássicos do terceiro excluído e o da não contradição, que são incapazes de tratar proposições que têm como conclusão valores divergentes do Verdadeiro e Falso,

permitidos, exclusivamente, na Lógica Clássica [Martins et al., 2007]. Esta lógica heterodoxa é muito utilizada em algumas áreas como a inteligência artificial, robótica, tratamento de dados incerto e sistemas especialistas.

4.3 Definição e Descrição da Lógica Paraconsistente – LP

A essência da Lógica Paraconsistente é admitir contradições, mas esta idéia teve uma longa jornada. Entre os pesquisadores que trabalharam sobre o problema da contradição, destacam-se o polonês *Jan Łukasiewicz* (1910-1971), que afirmava que Aristóteles já conjecturava sobre a possibilidade da invalidação da lei da contradição [Raspa, 1999].

Em paralelo ao trabalho de *Jan Łukasiewicz*, o russo *Nicolai Vasiliev*, que, entre 1910 e 1913, publicou uma série de artigos que também delineavam uma lógica com princípios paraconsistente [Arruda, 1990].

O logicista polonês *Stanislaw Jaskowski*, discípulo de *Jan Łukasiewicz*, em meados de 1948, propõe uma lógica que poderia ser aplicada a sistemas envolvendo contradições, mas sem ser trivial. O sistema de *Jaskowski* ficou conhecido como lógica discursiva, mas este sistema apresentava limitações e não se preocupava com a elaboração de lógicas paraconsistentes em sentido forte.

Em 1963, o sistema proposto por *Jaskowski*, foi oficialmente sistematizado por *Newton da Costa*. Hoje, *Newton da Costa* é considerado, internacionalmente, como o real criador das lógicas paraconsistentes [Arruda, 1990].

A Lógica Paraconsistente, a partir de 1985, passa a ser uma área do domínio da matemática, já que a *American Mathematical Society* classifica na seção 03-XX *Mathematical Logical and Foundation*, a subseção 03BXX *General Logic* o verbet *03B53 Logics admitting inconsistency (paraconsistent logics, discussive logics, etc.)*,

transformando a Lógica Paraconsistente em uma atividade da matemática contemporânea [Da Costa & Krause, 2004a].

A Lógica Paraconsistente, dentre os princípios da lógica clássica, nega o princípio da não-contradição, permitindo certas investigações que não seriam possíveis usando como base a lógica tradicional. Na Lógica Paraconsistente, uma sentença e a sua negação podem ser ambas verdadeiras, esta lógica apresenta alternativa a proposições, cuja conclusão pode ter valores, além de verdadeiro e falso, pois é possível utilizar valores como indeterminado e inconsistente [Abe, 1982].

A idéia central da lógica paraconsistente está no tratamento das inconsistências, mas ***O que é inconsistência?***

A inconsistência é um fenômeno natural ao mundo real e a todos que fazem parte dele. Segundo o dicionário Aurélio, é a falta de fundamento e coerência naquilo que está sendo analisado.

O ser humano ao se defrontar com situações de inconsistências, na maioria das vezes, consegue tratá-las de maneira satisfatória; pois quando isso ocorre, o homem tenta obter mais informações que auxiliem a sair desse estado de incerteza. Como exemplo de inconsistência tem-se a determinação da cor da maçã, isto é, a fruta pode ser vermelha, quase vermelha, com cores variando do vermelho claro até o vermelho escuro.

O raciocínio empregado na Lógica Paraconsistente para resolver as inconsistências, basea-se no uso do princípio da contradição, isto é, dada duas proposições contraditórias, uma é obrigatoriamente falsa [Abe et al., 1991].

Ao contrário da Lógica Clássica, na LP é possível representar e realizar inferências sobre informações contraditórias e, também, distinguir as situações onde uma determinada proposição é realmente falsa de uma em que não se tem conhecimento suficiente para se chegar a uma conclusão. A conclusão obtida pode ser muito útil em tomada de decisões em que não há informações suficientes ou são contraditórias.

Para desenvolver o raciocínio baseado em lógica paraconsistente é necessário admitir que [Da Costa et al., 1999]:

- Seja T uma teoria fundamentada sobre uma lógica L, onde, supostamente, a linguagem de T e de L contenha o símbolo para a negação. A teoria L é inconsistente se ela possuir teoremas contraditórios, isto é, um é negação do outro. Caso contrário, diz-se que T é consistente;
- Uma teoria T é trivial se todas as fórmulas de L forem teoremas de T, ou seja, tudo o que possa ser expresso na linguagem T possa ser provado em T; caso contrário T diz-se não trivial. Na maioria dos sistemas lógicos usuais é impossível distinguir a proposição verdadeira da falsa, pois qualquer proposição pode ser provada em T;
- Uma lógica L é chamada Paraconsistente se servir de base para teorias inconsistentes, mas não triviais.

Uma aplicabilidade da lógica paraconsistente é permitir a um computador processar dados contraditórios e inconsistentes, tendo como objetivo detectar as inconsistências derivadas de erros humanos (entrada de dados) ou de múltiplas e contraditórias fontes de informações.

As lógicas paraconsistentes, de certa forma, são prolongamentos da lógica tradicional, permitindo algumas investigações que não seriam possíveis utilizando a idéia da lógica clássica, pois as mesmas não visam eliminar a lógica tradicional, que permanece válida em seu domínio de aplicação.

4.4 Lógica Paraconsistente Anotada – LPA

A Lógica Paraconsistente Anotada é um grupo de lógicas não-clássicas que tiveram seus estudos iniciais desenvolvidos por [Subrahmanian, 1987], os quais estavam relacionados com a semântica e a sintática das lógicas paraconsistentes anotadas, sendo utilizados em programação lógica.

Estudos adicionais desenvolvidos por Subrahmanian e Blair, apresentaram a teoria geral da programação anotada e sua aplicação em banco de dados com contradições [Blair, 1998].

Newton da Costa, Subrahmanian e outros pesquisadores desenvolveram estudos, os quais mostram que a teoria geral anotada extraordinariamente forte, tem como caso particular a teoria de conjuntos fuzzy [Da Costa, 2000].

Na Lógica Paraconsistente Anotada de dois valores – LPA2v, uma proposição P está associada a dois fatores evidenciais (c,d) que representam, respectivamente, a quantidade de crença e descrença da proposição. Neste sistema lógico, as anotações são descritas na forma de graus de crença e descrença, relativos a uma dada proposição, dando-lhe compreensões de valorização. Estes graus de crença variam entre valores reais de 0 a 1 e podem ser obtidos por medições, por estatísticas ou probabilidades [Da Costa, 1993].

Para melhor entendimento da Lógica Paraconsistente Anotada, a mesma pode ser associada, de modo particular, através do Reticulado do Diagrama de Hasse que é apresentado e analisado nas subseções 4.4.1 e 4.4.2, na subseção 4.4.3 é descrita de forma mais detalhada a LPA2v

4.4.1 Reticulado do Diagrama de Hasse

No Reticulado do Diagrama de Hasse, os vértices apresentam os símbolos que indicam os estados lógicos, conforme exibido na figura 4.1, onde pode observar que [Da Silva Filho, 1999]:

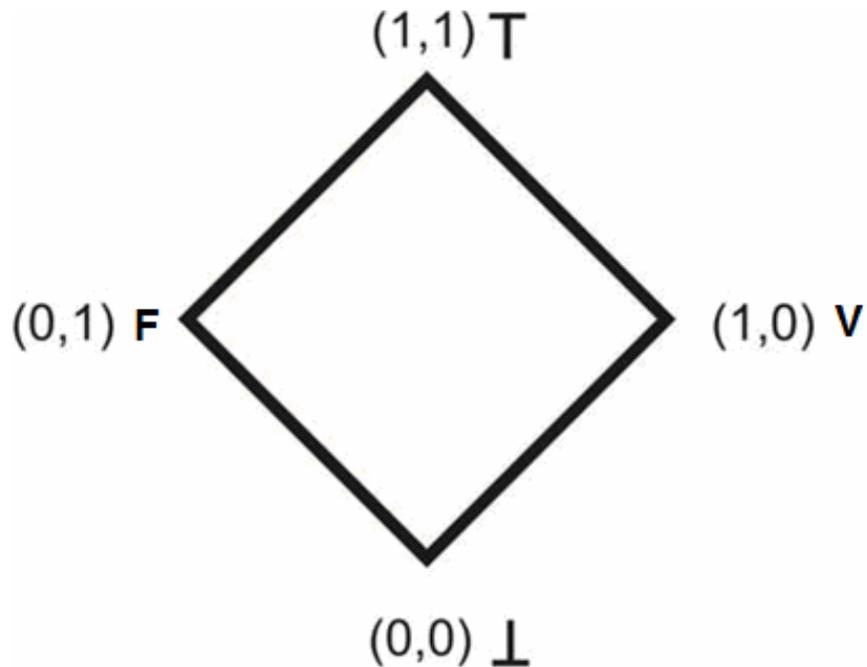


Figura 4. 1. Reticulado - Diagrama de Hasse
 Fonte: adaptado de [Da Silva Filho et al., 2008].

- Ponto $[1,0]$ indica a verdade (V) ou crença total;
- Ponto $[0,1]$ indica o falso (F) ou descrença total;
- Ponto $[1,1]$ indica a inconsistência (T) ou crenças totalmente inconsistentes;
- Ponto $(0,0)$ indica outro tipo de inconsistência ou paracompleteza (\perp) ou crença totalmente paracompleta.

Observação

Neste trabalho, a paracompleteza⁴ ou crença totalmente paracompleta será denominada de indeterminação ou indeterminado, pois a Lógica Paraconsistente, na sua

⁴ Paracompleto ou paracompleteza refere-se ao ponto $(0,0)$ e que, neste trabalho, será denominada de indeterminação, pois facilitará o entendimento quando da aplicação da Lógica Paraconsistente em Banco de Dados Relacional;

forma anotada, será usada com objetivo de identificar e classificar as incompletudes em banco de dados relacional.

A motivação para alteração da denominação de paracompleto para indeterminado, deve-se ao fato de que os possíveis estados de um dado, registro, tabela/entidade ou mesmo do banco de dados relacional são apresentados dentro de um padrão de estados denominado de consistente, inconsistente e indeterminado, não sendo usual ou mesmo prático utilizar o termo paracompleto relacionado a banco de dados relacional e/ou aos seus objetos.

4.4.2 Análise do Reticulado do Diagrama de Hasse

O estado de verdade $[1,0]$ ocorre quando se acredita totalmente na verdade e não há nenhuma informação que suporta a falsidade, por sua vez o estado de falsidade $[0,1]$ é o oposto.

O estado de inconsistência $[1,1]$ ocorre quando se acredita tanto na verdade quanto na falsidade de uma proposição, em um determinado instante de tempo.

O estado paracompleto (indeterminado) $[0,0]$ ocorre quando não há informações sobre a verdade nem sobre a falsidade.

É importante considerar o estado em um determinado momento de tempo, pois toda a informação possui uma dinâmica que se altera no decorrer do tempo.

4.4.3 Lógica Paraconsistente Anotada de Dois Valores - LPA2v

O raciocínio aplicado à lógica paraconsistente anotada proposta, considera que dois valores são associados a uma notação do reticulado da figura 4.1, sendo assim, esta lógica é denominada de lógica paraconsistente anotada de dois valores – LPA2v.

A LPA2v juntamente com o reticulado apresenta variáveis proposicionais $(P_1, P_2, P_3, \dots, P_n)$, constantes notacionais, símbolos auxiliares de conectivos lógicos: \neg - negação; \wedge - conjunção; \vee - disjunção; \rightarrow - implicação.

Considerando o reticulado da figura 4.1, o primeiro valor da notação representa a evidência favorável à preposição P e o segundo valor da notação apresenta a evidência contrária à preposição P.

Os graus de crença e descrença são notados por: μ_1 grau de crença e μ_2 grau de descrença. Considerando que os estados lógicos são encontrados através dos valores dos graus de crença e de descrença, os quais podem ser relacionados da seguinte forma: T = (1, 1) Inconsistente; V = (1, 0) Verdadeiro; F = (0, 1) Falso; \perp = (0, 0) Paracompleto (Indeterminado).

Os valores dos graus de crença (μ_1) e de descrença (μ_2) são representados pelo QUPC – Quadrado Unitário do Plano Cartesiano [Da Silva Filho, 1999], para determinar qual a área de trabalho relacionada aos graus de crença e descrença. Neste Quadrado Unitário do Plano Cartesiano são exibidos os valores de x e y variando no intervalo real de [0,1].

Os pontos no QUPC servem de base para a identificação e classificação de incompletudes em dados. A figura 4.2 apresentada a seguir, mostra o QUPC com os seguintes seguimentos [Martins et al., 2007]:

- \overline{FV} Linha Perfeitamente Consistente (LPC);
- $\overline{\perp T}$ Linha Perfeitamente Inconsistente (LPI).

Onde LPC indica que para qualquer valor do grau de crença, o valor do grau de descrença correspondente ao seu complemento em relação à unidade, o que implica em: $\mu_1 + \mu_2 - 1 = 0$.

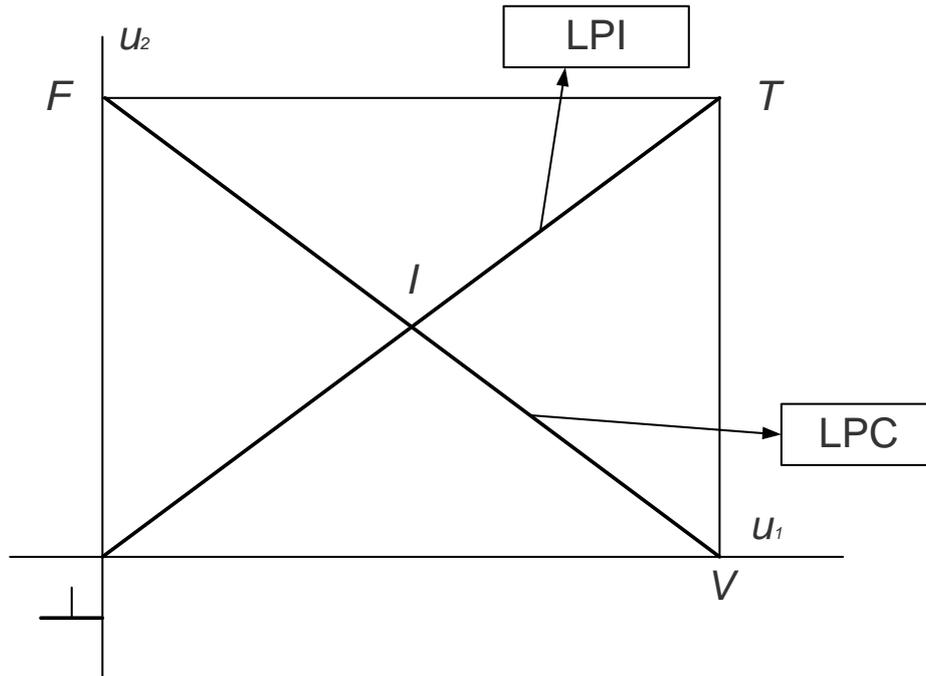


Figura 4. 2. Quadrado Unitário do Plano Cartesiano – QUPC.
Fonte: [Martins et al., 2007]

Quando o ponto de interseção, entre o grau de crença e de descrença, situar-se em cima da LPC, o mesmo é considerado um ponto perfeitamente consistente, mesmo representando conhecimentos parciais ou incompletos, mas apresentam total coerência entre os graus de crença e de descrença, conforme: $\mu_1 + \mu_2 = 1$.

A hipótese intuitiva da associação de uma notação (μ_1, μ_2) a uma proposição P significa que o grau de crença em P é μ_1 , enquanto o grau de descrença em P é μ_2 . Exemplificando, tem-se a proposição $P(\mu_1, \mu_2)$, onde se lê: crê-se em P com crença favorável até μ_1 e crença desfavorável até μ_2 [Da Silva Filho & Rocco, 2006].

A figura 4.3 apresenta um procedimento de um sistema em que são considerados como informações de entrada do sistema os graus de crença μ_1 e descrença μ_2 . Os estados lógicos, representados nos vértices do reticulado, são as saídas resultantes da análise paraconsistente.

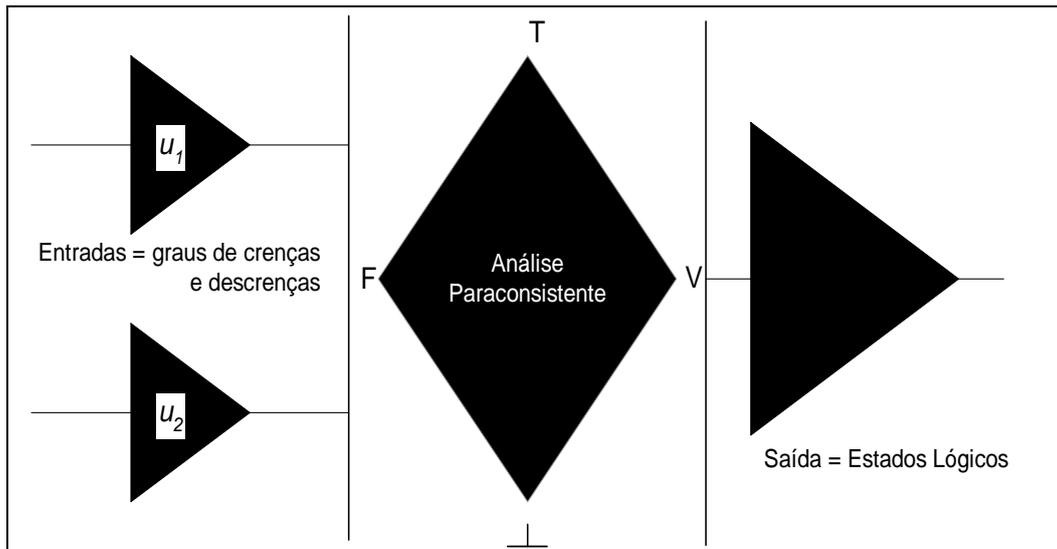


Figura 4. 3. Esquema de um Sistema de Análise Paraconsistente

4.5 Aplicações da Lógica Paraconsistente

A lógica paraconsistente, apesar de ser uma teoria matemática relativamente jovem, já possui utilização em várias áreas do conhecimento, como na própria matemática, na física, nas ciências jurídicas, eletrônica, robótica e na computação.

Na matemática, o uso da lógica paraconsistente se dá pela axiomatização de teorias de conjuntos, sem as fortes restrições impostas, tendo como objetivo impedir a existência de paradoxos como, por exemplo, o paradoxo de Russel e o paradoxo H.B. Curry [Cornelli & Miranda, 2007].

Na física, para auxiliar teorias aparentemente incompatíveis como o Átomo de Bohr que propõe uma junção entre a mecânica quântica e a relatividade geral [Domenech et al., 2008].

As ciências jurídicas utilizam a lógica paraconsistente, já que a mesma é repleta de detalhes semânticos, o qual serve de base para a construção de teorias paraconsistentes, como por exemplo, possibilidade de situações futuras e incertas que são muito importantes para o direito, já que podem auxiliar nas incertezas e indeterminações que surgem dentro dos processos judiciais [Carneiro, 2006].

Na eletrônica, os circuitos digitais utilizavam os princípios da lógica clássica, sendo que os primeiros equipamentos eletromecânicos construídos nos anos 40, “os irmãos mais velhos” dos computadores atuais, baseavam seus princípios na lógica clássica, que por ser bivalente, proporciona uma representação de fácil entendimento e implementação para sinais elétricos. Com o surgimento dos semicondutores, nos anos 50, fez com que os princípios da lógica clássica se solidificassem em relação ao seu uso na eletrônica [Da Costa & Krause, 2004a].

Com o avanço da tecnologia e a sua utilização na indústria, surge uma demanda por sistemas que se comportem mais próximo ao raciocínio humano, mas a lógica clássica, por sua natureza bivalente, não possuía os recursos necessários para este novo cenário de aplicações. Neste momento, as lógicas heterodoxas como a Paraconsistente e a Fuzzy, surgem como ferramentas para a solução dos problemas práticos na robótica e na computação como tomada de decisão em sistemas inteligentes, extração de conhecimento em grandes bancos de dados, entre outras áreas, onde as aplicações necessitam de uma decisão mais próxima com a humana, mas com todas as características de automação, mas sem a necessidade de intervenção humana.

Uma possível aplicação para a lógica paraconsistente está na identificação, classificação e tratamento de dados com inconsistências, oriundos de banco de dados que contenham algum grau de incerteza, inconsistência e contradição [Akama & Abe, 1998].

Hoje, qualquer sistema de informação, de natureza especialista ou não, terá um banco de dados, no qual armazenam dados com a possibilidade de inconsistências, atributos com valores ausentes, incompletos e contraditórios [Imieliński & Lipski, 1984].

Para tratar os problemas relacionados à natureza do dado, é necessário o uso de uma teoria com recursos e suporte ao tratamento de contradições, inconsistências e incompletude.

Caso a teoria escolhida for a Lógica Clássica, esta irá se defrontar com a ocorrência da contradição nos dados, neste caso o sistema se trivializa acarretando a inviabilidade do uso desta teoria. Mas, se a teoria escolhida for à lógica paraconsistente, o tratamento se dará de forma diferente, já que o fato de haver contradições entre os dados pode implicar na possibilidade da descoberta de conhecimento implícito entre os mesmos [Da Costa, 2004b].

Para exemplificar o uso em banco de dados, pode-se propor utilizar a LP como ferramenta para identificação e classificação de dados (*registros*) inconclusivos ou indeterminados dos equipamentos da Tabela 3.1, apresentada no Capítulo 3.

4.6 Considerações Finais

As idéias iniciais da Lógica Paraconsistente, foram desenvolvidas nas primeiras décadas do início do Século XX por pesquisadores russos e poloneses, somente em 1963 foi formalmente descrita e sistematizada pelo brasileiro Newton da Costa. Neste capítulo apresentou-se de forma sintética a Lógica Paraconsistente, sua filosofia e recursos matemáticos, características, as lógicas paraconsistentes anotadas e as principais aplicações e uso nas ciências.

Esta lógica apresenta-se, como uma boa e poderosa ferramenta, para tratar problemas de inconsistências e incertezas entre dados. Neste trabalho, a LP, será aplicada em conjunto com a TCA, na base de dados do Sistema de Informações Elétricas - INFO_OPR, mas precisamente no subconjunto de dados Acre/Rondônia.

A Lógica Paraconsistente a ser usada neste trabalho é a LPA2v, pois apresenta as características necessárias para a identificação e classificação das incompletudes apresentadas entre os dados do Sistema INFO_OPR.

5. Ambiente do Problema: Base de Dados do Sistema de Informações Elétricas Acre/Rondônia – INFO_OPR

5.1 Visão Geral

O Sistema Elétrico da Região Norte, gerenciado pela empresa ELETRONORTE, utiliza um sistema de gestão de informações denominado Sistema de Gerenciamento de Informações Operacionais de Sistemas Elétricos – INFO_OPR. Neste capítulo apresenta-se o Sistema Elétrico INFO_OPR, dando ênfase ao subconjunto dos estados do Acre e Rondônia, já que os referidos dados compõem o ambiente de trabalho para o teste da metodologia proposta.

5.2 Histórico do Sistema Elétrico Acre/Rondônia

No ano de 1981, o parque gerador de energia de Porto Velho, capital do Ex-Território Federal de Rondônia, hoje Estado de Rondônia, foi transferido para as Centrais Elétricas do Norte do Brasil S/A - ELETRONORTE, que se responsabilizou, a partir deste momento, pela operação dos sistemas elétricos da capital, composto exclusivamente por usinas termelétricas a diesel.

A distribuição de energia elétrica na Capital Porto Velho, e a geração e distribuição em todo o interior do Estado, permaneceram sob a responsabilidade das Centrais Elétricas de Rondônia – CERON. Neste período, o sistema de geração tinha, apenas, obrigação com a cidade de Porto Velho e algumas cidades do interior, mais próximas a capital [ELETRONORTE, 2009].

Em 1982 tem início a construção da Usina Hidroelétrica de Samuel – UHE Samuel. A entrada em operação desta usina ocorre com cinco turbinas do tipo Kaplan, com capacidade individual de 43,2 MW.

As turbinas foram instaladas durante o período de 1989 a 1996, perfazendo um total de 216 MW. Em decorrência desta instalação, ocorreu a desativação de algumas usinas térmicas, mas a construção de linhas de transmissão de 230 kV possibilitou o fornecimento da energia da UHE Samuel para as cidades do interior do Estado [Vianna, 2007].

Os problemas de racionamento de energia foram constantes, durante a construção da UHE Samuel, cuja situação crítica permanece durante o período de 1997 a 1999, ocasionando a admissão do produtor independente de energia Termo Norte, empresa contratada pela ELETRONORTE para minimizar os problemas naquele momento.

No ano de 2002, a cidade de Rio Branco, capital do Estado do Acre, foi interligada à linha de transmissão da UHE Samuel. A qual mais tarde, foi estendida para outras localidades do Estado, a partir deste momento o Sistema Elétrico de Rondônia passa a ser conhecido como Sistema Elétrico Acre/Rondônia, sendo suprido pela energia gerada na Usina Hidrelétrica de Samuel, por Pequenas Centrais Hidrelétricas – PCHs e por usinas térmicas [Vianna, 2007].

Para efetivação da interligação com o sistema nacional, a rede do Sistema Acre/Rondônia foi estendida, em 2008 com a linha de transmissão de 230 kV da ELETRONORTE, que interliga a Usina Hidrelétrica de Samuel às subestações de Ariquemes e Ji-Paraná, até às cidades de Pimenta Bueno e Vilhena, para posteriormente interligar a cidade de Jauru no Estado do Mato Grosso, sendo que a interligação ao Sistema Interligado Brasileiro – SIN, apenas foi efetivada a partir de outubro de 2009 [ELETRONORTE, 2009]. Na figura 5.1 está o mapa da interligação do Sistema Elétrico Acre/Rondônia ao Sistema Nacional.

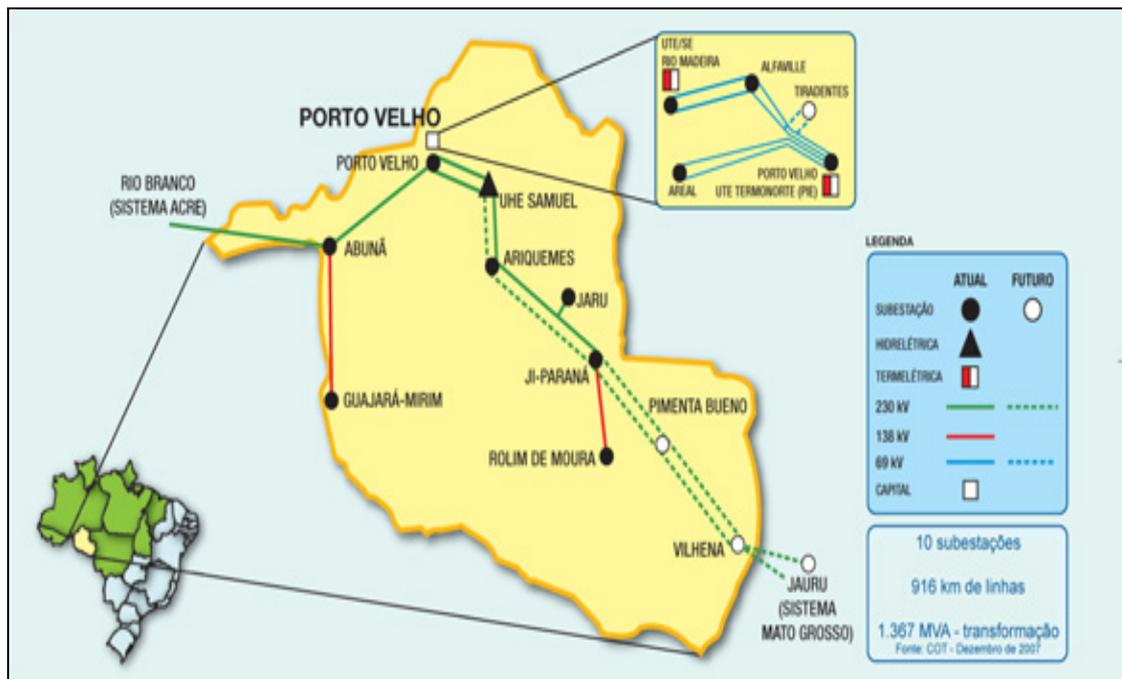


Figura 5. 1. Mapa com a Interligação do Sistema Elétrico Acre/Rondônia.
Fonte: ELETRONORTE. Sistema de Transmissão em Rondônia.

5.3 Sistema de Gestão de Informação Elétricas INFO_OPR

Para auxiliar a gestão e controle, deste parque de geração e transmissão de energia, instalado no Estado, a ELETRONORTE unidade Rondônia, também utiliza o Sistema de Gerenciamento de Informações Operacionais de Sistemas Elétricos – INFO_OPR, ferramenta de *Tomada Rápida de Decisões para Sistemas Elétricos*.

O INFO_OPR, desenvolvido e mantido, exclusivamente, pela ELETRONORTE, é um sistema corporativo, que atualmente tem aproximadamente um mil e seiscentos (1600) usuários, está em produção/operação desde 1998 em todos os Centros de Operação das Regionais e Subsidiárias da ELETRONORTE: Acre, Amapá, Manaus Energia, Maranhão, Mato Grosso, Roraima Transmissão, Boa Vista Energia, Rondônia, Usina de Tucuruí, Pará, Tocantins e Brasília, além de atender demandas da ANEEL, ONS, e ELETROBRAS.

A figura 5.2, apresentada a seguir, mostra o mapa da rede lógica da ELETRONORTE, com seus servidores de dados com as informações de operações.

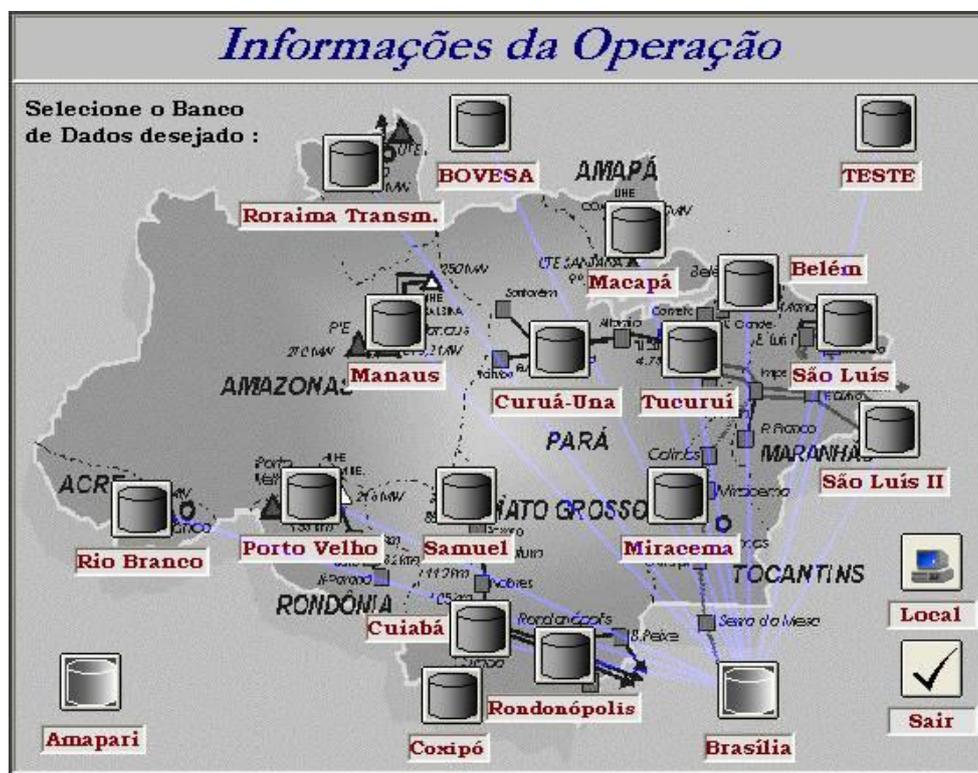


Figura 5. 2. Mapa da Rede Lógica da ELETRONORTE com os Bancos de Dados.

Fonte: Milton Nunes da Silva Filho. Gestão PV, Pós Operação Brasília, INFO_OPR - ELETRONORTE.

O banco de dados do INFO_OPR contém, atualmente, oito milhões de dados históricos relacionados aos estados operacional dos equipamentos, demandas, energia e hidrologia. Apresenta facilidades que disponibilizam relatórios e gráficos padrões diários e mensais, sendo o principal recurso disponível a consolidação e cálculo dos indicadores de desempenhos dos sistemas elétricos da empresa [Nunes da Silva Filho, 2006].

Neste trabalho, apenas analisa-se os dados armazenados nos servidores Rio Branco (AC), Porto Velho (RO), Samuel (UHE Samuel em PVH). O conjunto de dados que fazem parte deste subconjunto do sistema têm aproximadamente um milhão (1.000.000) de registros e cento e cinquenta usuários, distribuídos pelas unidades de operações nos Estados Acre e Rondônia [Nunes da Silva Filho, 2009].

5.4 Operação do Sistema de Gestão de Informação Elétricas INFO_OPR

A operação de sistemas que verificam e avaliam as ocorrências em um sistema elétrico são, de forma geral, complexos e com uma grande quantidade de incertezas e variáveis para análise. Estes sistemas necessitam a presença de um operador que possa avaliar e tomar as devidas decisões de forma correta, precisa e rápida utilizando uma quantidade grande de variáveis (atributos) e seus conteúdos (dados).

No Sistema de Gerenciamento de Informações Operacionais de Sistemas Elétricos – INFO_OPR, o operador necessita manipular grande quantidade de dados, em função desta necessidade, é necessário desenvolver métodos que reduzam a quantidade de variáveis (atributos) juntamente com os respectivos dados.

Para auxiliar o operador e a empresa na tomada de decisão, este trabalho propõe uma metodologia que avalie os atributos relevantes da base de dados do INFO_OPR, isto é, descobrir quais são os atributos mais importantes do banco de dados, além de identificar e classificar as possíveis incompletudes existentes nos dados.

5.5 Banco de Dados do Sistema INFO_OPR

O banco de dados do Sistema INFO_OPR, foi inicialmente desenvolvido em MS-Access no ano de 1998, porém a necessidade de incorporar aspectos temporais, além do aumento considerável da quantidade de registros, levou os desenvolvedores do sistema a migrarem para um software mais robusto [Nunes da Silva Filho, 2009].

O Sistema INFO_OPR, atualmente é baseado no modelo de dados relacional-temporal, em que o tempo é fator relevante para o armazenamento [Snodgrass & Ahn, 1985]. A seguir, se apresenta os aspectos lógicos e físicos do banco de dados INFO_OPR.

5.5.1 Aspectos Lógicos do Banco de Dados INFO_OPR

O banco de dados do Sistema INFO_OPR é baseado no modelo relacional-temporal do tipo bitemporal, estruturado em tabelas, onde cada tabela apresenta um conjunto de atributos que dá sustentação para o armazenamento e recuperação dos dados [Jensen et al., 1998].

O modelo relacional apresenta restrições, isto é, regras que servem para garantia de integridade e consistência, além de impedir redundância de dados [Brusoni et al., 1999]. No caso deste sistema, o seu desenvolvimento foi realizado em 1998, sendo que já havia um conjunto de sistemas isolados que supriam as necessidades da empresa, isto acarretou na migração de dados destes referidos sistemas (sistemas legados), pois as séries de dados históricas armazenadas e utilizadas pela ELETRONORTE não poderiam ser descartadas.

Neste momento, começam os primeiros problemas de inconsistências da base de dados, pois para que ocorra a importação dos dados com sucesso, entre sistemas legados e sistemas relacionais, é necessário desabilitar ou não provê as restrições pertinentes do modelo relacional.

No caso do banco de dados do INFO_OPR, não implementar, em nível de SGBD, as restrições foi a solução encontrada, pois em um banco de dados deste tipo a dinâmica dos dados deve ser respeitada, já que muitas vezes, uma situação é verdadeira em um determinado espaço de tempo e em outro pode ser alterada, passando a ser falsa [Jensen, 2000].

O INFO_OPR, hoje, não executa as restrições de integridade de dados (*Primary Key*) e nem de integridade referencial (*Foreign Key*), em nível de SGBD SQL Server 2000.

As restrições de integridade são avaliadas em nível de *interface* de dados, isto é, os programas de entradas de dados é que verificam, em um primeiro nível de integridade e redundância dos referidos dados.

Uma característica interessante deste sistema é o provimento de cinco níveis de consistência de dados, já que com a impossibilidade de usar as restrições do próprio modelo de dados, foi necessário o uso desta estratégia, para impedir inconsistências e outros problemas no banco de dados. A seguir, são listadas, as tabelas mais importantes do banco de dados INFO_OPR, pois são elas que serão utilizadas neste trabalho, não sendo detalhada a estrutura das mesmas. A Tabela 5.1 apresenta a relação das tabelas⁵ do Banco de Dados INFO_OPR.

Tabela 5. 1. Tabelas do Banco de Dados INFO_OPR

Tabelas INFO_OPR
ATUALIZACAO_RDO;
CANCELAMENTO;
COMBUSTIVEL;
DEMANDAS_DIA;
DEMANDAS_MAXIMAS;
DESCR_CANCELAMENTO;
DESCR_MANUTENCAO
DESCR_OBSERVACAO
DESCR_OCORRENCIA
DISJ_SECC
EQUIPAMENTO
EQUIPTO_AFETADO
EQUIPTO_AFETADO_MANUTENCAO
HIDROLOGICOS_CONSTANTES_MENSAIS
HIDROLOGICOS_NIVEIS_DIARIOS
HIDROLOGICOS_VAZOES_DIARIOS
HIDROLOGICOS_VAZOES_HORARIOS
INDICES_SIMULACAO

⁵ Em banco de dados relacional é recomendado não utilizar acentos e caracteres especiais nos nomes das tabelas/entidades e dos atributos/campos.

Continuação da Tabela 5.1

Tabelas INFO_OPR
INSTALACAO
INSTALACAO_EQUIPAMENTOS
MANUTENCAO
MANUTENCAO_OM
MEDICOES_DIA
INDICES
OBSERVAÇÃO
OCORRENCIA
OCORRENCIA_INDICES
OCORRENCIA_OM
OPERACIONAIS_LOCAIS_DIARIOS
OPERACIONAIS_SISTEMAS_DIARIOS
OPERACIONAIS_UNIDADES_DIARIOS
RAP
RAP_ATUALIZACAO
RAP_DESCRICAO
RAP_IMAGENS
RAP_OCORRENCIAS
RESUMOS_ENERGIA
USUARIO_CMB
USUARIO_INSTALACOES
USUARIO_RDO
USUARIOS
USUARIOS_LOG
CLASSE_OCORRE

5.5.2 Aspectos Físicos do Banco de Dados INFO_OPR

O banco de dados do INFO_OPR apresenta arquitetura distribuída homogênea com modelo relacional-temporal de dados e SGBD *Microsoft SQL-Server 2000 (Server Interprise Edition)* em todos os Nós da rede de dados. O *SQL Server*, nesta versão, provê as características temporais necessárias ao processamento dos dados históricos, abundantes neste sistema [Jansen & Snodgrass, 2009].

A replicação do banco de dados é parcial entre os Nós, com atualização das cópias em tempo real, utilizando a INTRANET - rede de comunicação interna da ELETRONORTE. A arquitetura replicada do sistema apresenta um Nó Central, localizado no Distrito Federal, na cidade de Brasília, onde está a sede da ELETRONORTE. O Nó Central, além de ter a responsabilidade pela consolidação final dos dados dos Estados supridos pela ELETRONORTE, é o local de armazenamento da cópia central de segurança do sistema.

5.6 Principais Problemas da Base de Dados do INFO_OPR

A base de dados do Sistema INFO_OPR apresenta alguns problemas, os quais foram a grande motivação para o desenvolvimento deste trabalho, pois como este sistema, auxilia a tomada rápida de decisão dentro da empresa ELETRONORTE, é imprescindível que a base de dados, utilizada para extrair conhecimento, esteja íntegra e confiável. A seguir são apresentados os principais problemas identificados na base de dados:

1. Não cumprimento das regras básicas de banco de dados relacional;
2. Inexistência de dicionário de dados;
3. Todos os atributos são do tipo texto;
4. Inexistência de chaves – primária e estrangeira;

5. Utilização do SGBD, apenas como repositório de dados, dificultando a efetiva gerência da base de dados;
6. A base de dados apresenta percentuais elevados de atributos com valores ausentes;
7. Atributos importantes com inconsistências e indeterminações nos dados.

Observação:

Os problemas apresentados anteriormente, não impedem do Sistema INFO_OPR, está em operação e auxiliando com sucesso a ELETRONORTE à gestão da informação de ocorrências elétricas.

5.7 Considerações Finais

O ambiente da base de dados do INFO_OPR apresenta-se no modelo de dados relacional com características temporais. Possui um subconjunto de dados com as informações operacionais elétricas dos Estados Acre/Rondônia, os quais serão objeto de análise neste trabalho.

A base de dados INFO_OPR, no subconjunto de estudo Acre/Rondônia, apresenta-se com todas as características necessárias para os testes da metodologia, pois é constituída de mais de quarenta tabelas, com uma grande quantidade de atributos, os quais apresentam incompletudes entre os dados.

6. Metodologia de Avaliação da Relevância dos Atributos de Grandes Bases de Dados Incompletas, Utilizando Teoria de Conjuntos Aproximados e Lógica Paraconsistente

6.1 Visão Geral

A habilidade de analisar grandes quantidades de informações e transformar em conhecimento é peculiar ao ser humano. A realização da tarefa de analisar dados, muitas vezes, é complexa, principalmente se os dados possuírem desordens e incompletudes.

A metodologia, a ser apresentada neste capítulo, propõe uma avaliação de todos os atributos da base de dados, objetivando selecionar apenas os relevantes com a finalidade de reduzir a quantidade de atributos e seus respectivos dados a serem analisados. Para efetivar o objetivo da metodologia, são utilizadas as ferramentas matemáticas não clássicas Teoria de Conjuntos Aproximados e Lógica Paraconsistente.

O objetivo específico da metodologia é encontrar um conjunto reduzido de atributos, que irá compor uma nova base de dados, para posterior extração de conhecimento. Nesta etapa da metodologia é utilizada a Teoria de Conjuntos Aproximados.

Os conteúdos dos atributos serão analisados com o objetivo de identificar e classificar as incompletudes através da Lógica Paraconsistente. A classificação das incompletudes e posteriormente dos registros produzirá um conjunto de informações sobre o estado de consistência, indeterminação e inconsistência dos registros e, de forma direta, o estado da nova base de dados.

O banco de dados, no qual a proposta de metodologia será aplicada é o banco de dados do Sistema de Informações de Ocorrências Elétricas - Sistema INFO_OPR da empresa ELETRONORTE, no subconjunto de dados Sistema Elétrico Acre/Rondônia.

6.2 Relação entre TCA e Banco de Dados Relacional - BDR

A TCA foi desenvolvida independente e diferentemente da teoria de banco de dados relacional, mas a idéia principal da TCA está em classificar, encontrar padrões, regras e conhecimento embutido entre os dados. Já a teoria de banco de dados relacional está centrada no armazenamento e recuperação de dados, não se preocupando com o tipo de dado que está armazenando e nem com padrões escondidos entre os mesmos.

A teoria de banco de dados relacional é baseada em ferramentas da matemática clássica como teoria clássica de conjuntos, álgebra relacional e lógica clássica. O banco de dados relacional está projetado para o desenvolvimento de estruturas sólidas, tanto lógicas quanto físicas para armazenamento de dados e recuperação da informação [Codd, 1970].

Estas idéias tornam as duas teorias diferentes, mas com um ponto de interseção conceitual - o conceito de entidades⁶ (tabelas). As tabelas que são estudadas em ambas as teorias, são apresentados de forma similar, pois são compostas por *registros*⁷ e valores de atributos.

A similaridade conceitual entre as duas teorias não é exata, pois os registros das tabelas, nas duas teorias, não têm correspondência direta, pois em TCA os registros podem ser redundantes, mas em BDR a redundância deve ser evitada. A seguir, apresentam-se conceitos importantes, das duas teorias, relacionados a solução do problema deste trabalho [Lin, 1996]:

⁶ Entidade é um objeto composto por um conjunto de atributos. Entidade e tabela são sinônimos em BDR.

⁷ Registro é uma linha de uma tabela de um BDR, sendo formado por um conjunto de atributos (colunas).

→ **No Banco de Dados Relacional uma Relação R consiste de:**

- $U = \{x, y, \dots\}$ conjunto finito de entidades, neste caso chamado banco de dados;
- T é um conjunto de atributos $\{A_1, A_2, A_3, \dots, A_n\}$;
- $\text{Dom}(A_i)$ é o conjunto de valores de atributos A_i , onde $\text{Dom} = \text{dom}(A_1) \cup \text{dom}(A_2) \cup \text{dom}(A_3) \cup \dots \cup \text{dom}(A_n)$;
- Cada tabela em U é representada unicamente por um mapa $t: T \rightarrow \text{Dom}$, onde $t(A_i) \in \text{Dom}(A_i)$, para cada $A_i \in T$.

A definição anterior apresenta a idéia de que, uma relação pode ser entendida como uma tabela que consiste de linhas (*registros*) de elementos (entidades), mas cada linha representa uma entidade exclusivamente.

→ **Na Teoria de Conjuntos Aproximados uma Tabela Consiste de:**

- $U = \{u, v, \dots\}$ é um conjunto de tabelas (entidades);
- T é um conjunto de atributos $\{A_1, A_2, A_3, \dots, A_n\}$;
- $\text{Dom}(A_i)$ é o conjunto de valores de atributos A_i , onde $\text{Dom} = \text{dom}(A_1) \cup \text{dom}(A_2) \cup \text{dom}(A_3) \cup \dots \cup \text{dom}(A_n)$;
- $\rho : U \times T \rightarrow \text{Dom}$, chamado de função de descrição, é um mapa tal que $\rho(u, A_i)$ está no $\text{dom}(A_i)$ para todo u e A_i em T , observa-se que ρ induz a um conjunto de mapas: $t = \rho(u, \bullet) : T \rightarrow \text{Dom}$, onde cada mapa é uma *tupla* $t = (\rho(u, A_1), \rho(u, A_2), \dots, \rho(u, A_i), \dots, \rho(u, A_n))$.

A definição anterior pode ser resumida da seguinte forma: um registro, não necessariamente, deve ser associado com uma entidade unicamente, pois diferentes

tabelas podem ter uma mesma representação de registro, situação que não é permitida em banco de dados relacional.

6.3 Considerações sobre o Banco de Dados Relacional do INFO_OPR

- O banco de dados do INFO_OPR é baseado no modelo relacional com características temporais. Logo, esta base é composta por um conjunto de tabelas (entidade) que deveriam estar inter-relacionadas através de chaves (primárias e estrangeiras) que garantem a integridade do banco de dados [Edelweiss, 1998];
- Considerando o processo de normalização do banco de dados, caso exista. Esta consideração é necessária para se efetivar uma análise de todas as tabelas do banco de dados, com objetivo de identificar as tabelas importantes e as regras de integridade do modelo relacional. A Normalização é um processo executado, em uma fase específica do projeto de banco de dados relacional (BDR), que visa ajudar a definir quais as relações (entidades) necessárias e quais seus atributos (colunas), a partir de um conjunto de dados a ser representado no próprio banco de dados, de modo a evitar redundância [Navathe & Elmaris, 2006];
- O processo de normalização facilita a identificação das principais tabelas do banco de dados, isto é, as tabelas principais são aquelas que no processo de normalização foram descobertas na 1ª Forma Normal e/ou na 2ª Forma Normal, as encontradas na 3ª Forma Normal, são geralmente tabelas auxiliares e, as da 4ª Forma Normal são apenas tabelas que tratam dados multivalorados embutidos nas tabelas descobertas nas 1ª ou 2ª ou 3ª Forma Normal. As Formas Normais estão baseadas nas dependências funcionais entre os atributos de uma relação [Date, 2004];
- Caso não exista o processo de normalização, considerar o fluxo e a dinâmica de armazenamento do banco de dados, já que o mesmo mostra, de forma clara, qual

a tabela que participa efetivamente das atualizações do banco de dados, isto é a principal tabela do banco de dados [Korth & Silberschatz, 1995].

- No banco de dados INFO_OPR, as principais tabelas são:
 - cancelamento;
 - combustivel;
 - demandas_dia;
 - equipamento;
 - equipto_afetado_manutencao;
 - hidrologicos_constantes_mensais;
 - hidrologicos_niveis_diarios;
 - hidrologicos_vazoes_diarios;
 - hidrologicos_vazoes_horarios;
 - instalação;
 - manutencao;
 - ocorrencia.
 - A tabela ocorrencia é considerada a principal entidade do sistema, já que é para ela, ou através dela, que flui toda a dinâmica de armazenamento dos dados. As outras entidades, teriam a responsabilidade de dar suporte, para que o armazenamento seja realizado de forma íntegra e consistente.

Na Tabela 6.1 apresentam-se os atributos da Tabela Ocorrência, os quais são os tributos que representam, de forma geral, o conhecimento embutido na base de dados.

Tabela 6. 1. Atributos da Tabela Ocorrência

Atributos da Tabela ocorrência
COD_SISTELE_OCORRE
COD_EQUIPAM_OCORRE
DATA_INICIAL_OCORRE
HORA_INICIAL_OCORRE
DATA_FINAL_OCORRE
HORA_FINAL_OCORRE

Continuação da Tabela 6.1.

Atributos da Tabela Ocorrência
NUM_OCORRE
TIP_OCORRE
OS_SD_OCORRE
POT_INTERROMP_OCORRE
CORRENTE_OCORRE
TEMPO_OCORRE
REPOSICAO_OCORRE
SAOE_OCORRE
DURACAO_OCORRE
COD_CIER_OCORRE
COD_CLASSE_OCORRE
COD_ORIGEM_OCORRE
COD_EMPRESA_ORIGEM_OCORRE
COD_EMPRESA_RESPONSAVEL_OCORRE
REDUCAO_DEMANDA_OCORRE
ENERGIA_INTERROMPIDA_OCORRE
IND_CLASSIF_OCORRE
POTENCIA_CONTRATADA_ONS_OCORRE
GRUPO_CDB_OCORRE
CODIGO_USUARIO_OCORRE
DESCRICAO_USUARIO_OCORRE
DATA_ATUALIZACAO_OCORRE
DISPONIVEL_OCORRE
LIGADO_OCORRE
COD_INSTALACAO_OCORRE
CODIGO_ESTADO_OPERACIONAL_OCORRE
CODIGO_REDE_OPERACAO_OCORRE
PAGAMENTO_BASE_OCORRE
FATOR_PV_OCORRE
PV_OCORRE
ACRESCIMO_OCORRE
DATA_MANOBRA_1_OCORRE
HORA_MANOBRA_1_OCORRE
DATA_MANOBRA_2_OCORRE
HORA_MANOBRA_2_OCORRE
RESTRICAO_OCORRE
NOS_ESTADO_OPERATIVO
ONS_CONDICAO_OPERATIVA
ONS_DETALHAMENTO_APURACAO
ONS_FORMA_CONTABILIZACAO

6.4 Aplicação da TCA no Banco de Dados Relacional INFO_OPR

A TCA tem como uma das suas características a redução de dados, que é uma importante tarefa na descoberta de conhecimento, já que a alta dimensionalidade de uma base de dados pode ser reduzida usando técnicas que dependem, principalmente, do objetivo e conhecimento que se deseja descobrir ou extrair.

Para efetivar a redução de dados é necessário saber que dados são relevantes, para isso é necessário a descoberta dos atributos relevantes da base de dados com seus respectivos conteúdos.

A Teoria de Conjuntos Aproximados é uma ferramenta que habilita a descoberta de dependências de dados e a redução do número de atributos em uma base dados, mas este conjunto de atributos descoberto deve representar as características da informação contidas na própria base de dados. A seguir são mostrados os passos para a descoberta dos atributos relevantes da base de dados INFO_OPR.

→ Passos para aplicação da TCA na base de dados do INFO_OPR

1. Determinação do conjunto Aproximação Superior;
2. Determinação do conjunto Aproximação Inferior;
3. Determinação do conjunto Região de Borda;
4. Determinação das principais Relações de Indiscernibilidade;
5. Determinação do Core da tabela principal.

→ Considerações sobre a aplicação da TCA na base de dados do INFO_OPR:

1. Ao aplicar a TCA, por tabela individualmente, determina-se um conjunto de atributos relevantes de cada tabela, que ao reunir todos esses atributos em um único conjunto B, encontram-se atributos que ocorrem em mais de uma tabela;

2. Ao retirar as repetições de atributos e encontrar o conjunto de atributos relevantes (B), descobre-se que os atributos do conjunto B são, basicamente, os atributos da tabela ocorrência;
3. Considerar o fluxo e a dinâmica de armazenamento dos dados, sendo a tabela ocorrência a mais importante do banco de dados. O que leva a concluir que a existência da normalização ou a identificação do fluxo e dinâmica de armazenamento no banco de dados facilitam ou reduzem o trabalho para a identificação do conjunto de possíveis atributos relevantes.

→ Aplicação da TCA na Tabela Ocorrência utilizando parcialmente o Algoritmo proposto em [Lambert–Torres et al., 1999].

A seguir são apresentados os passos para aplicação do algoritmo proposto:

1. Transformar valores contínuos em faixas (se necessário);
2. Eliminar atributos idênticos;
3. Eliminar exemplos idênticos;
4. Eliminar atributos dispensáveis;
5. Calcular o conjunto básico da tabela de decisão;
6. Compor a tabela conjunto redução, isto é, o Core da Tabela Ocorrência.;
7. Agrupar os exemplos;
8. Compor o conjunto final de regras.

No caso da Tabela Ocorrência, como aplicar a TCA utilizando o algoritmos apresentado anteriormente? Para isso é necessário analisar os atributos da Tabela Ocorrência para:

Passo 1. Eliminar atributos idênticos;

Passo 2. Eliminar atributos dispensáveis;

Passo 3. Compor a tabela conjunto redução.

Os outros passos do algoritmo proposto em [Lambert-Torres et al., 1999], não são necessários sua aplicação na Tabela Ocorrência, pois a mesma faz parte de um BDR, além de apresentar características muito peculiares, que já foram apresentadas e discutidas anteriormente.

→ **Resultado da aplicação TCA – determinação dos atributos relevantes.**

Tabela 6. 2. Tabela com Atributos Relevantes

Atributos Relevantes	Número do atributo
COD EQUIPAM OCORRE	1
DATA INICIAL OCORRE	2
HORA INICIAL OCORRE	3
DATA FINAL OCORRE	4
HORA FINAL OCORRE	5
TIP OCORRE	6
POT_INTERROMP OCORRE	7
CORRENTE OCORRE	8
COD CLASSE OCORRE	9
COD ORIGEM OCORRE	10
COD EMPRESA ORIGEM OCORRE	11
COD EMPRESA RESPONSVEL OCORRE	12
LIGADO OCORRE	13
CODIGO ESTADO OPERACIONAL OCORRE	14
CODIGO REDE OPERACAO OCORRE	15

Observação:

O atributo COD_SISTELE_OCORRE não participa do conjunto de Atributos_Relevantes, pois o mesmo é apenas a identificação do subconjunto de dados, que para esta situação são os dados de Acre/Rondônia. Caso a análise, seja realizada em

toda base de dados do Sistema INFO_OPR, o atributo COD_SISTELE_OCORRE pertenceria aos elementos do conjunto de atributos relevantes, já que é necessário saber de que Estado (Unidade da Federação) são os dados.

→ **Consulta aos dados da Tabela Ocorrencia utilizando, apenas, os atributos relevantes descobertos anteriormente.**

Os atributos relevantes da tabela ocorrência são apresentados a seguir, dentro de uma consulta estruturada utilizando um *scrip* em SQL da Microsoft:

```

SELECT
    Ocorrencia.COD_EQUIPAM_OCORRE,
    Ocorrencia.DATA_INICIAL_OCORRE,
    Ocorrencia.HORA_INICIAL_OCORRE,
    Ocorrencia.DATA_FINAL_OCORRE,
    Ocorrencia.HORA_FINAL_OCORRE,
    Ocorrencia.TIP_OCORRE,
    Ocorrencia.POT_INTERROMP_OCORRE,
    Ocorrencia.CORRENTE_OCORRE,
    Ocorrencia.COD_CLASSE_OCORRE,
    Ocorrencia.COD_ORIGEM_OCORRE,
    Ocorrencia.COD_EMPRESA_ORIGEM_OCORRE,
    Ocorrencia.COD_EMPRESA_RESPONSAVEL_OCORRE,
    Ocorrencia.LIGADO_OCORRE,
    Ocorrencia.CODIGO_ESTADO_OPERACIONAL_OCORRE,
    Ocorrencia.CODIGO_REDE_OPERACAO_OCORRE
FROM Ocorrencia;

```

A Tabela 6.3, apresenta uma amostra do resultado da consulta realizada através do MS-Access, usando apenas parte dos atributos relevantes, definidos anteriormente nesta seção. Nesta tabela, não são utilizados os nomes dos atributos, mas a numeração correspondente apresentada na Tabela 6.2.

Tabela 6. 3. Resultado Parcial da Consulta com os Atributos Relevantes

1	2	3	4	5	6	7	8	9	10
PVTF604	00000000	0000	20000628	2245	V	000000	0000	115	ELN
PVDB601	00000000	0000	20000619	0841	V	000000	0000	3V	ELN
PVTF605	00000000	0000	20000628	2245	V	000000	0000	115	ELN
RLTF401	00000000	0000	19991210	0622	V	000000	0000		ELN
RLAL203	00000000	0000	19991210	9745	V	000001			ELN
RLBR201	00000000	0000	19991210	0625	V	000000	0000		ELN
AFAL209			19991210	0550					
JPRLT401	00000000					000000	0000	111	
ALAL210	00000000	0000	20000728	2245	V	000000	0000	3V	ELN
RMDB601	00000000	0000	19990619	0841	V	000000	0000	3V	ELN
PVTF605	00000000	0000	20020628	2245	V	000000	0000	115	ELN
RLTF401	00000000	0000	19991210	0622	V	000000	0000		ELN
RLAL203	00000000	0000	19991210	9745	V	000004			ELN
RLBR201	00000000	0000	19991210	0625	V	000003	0000		ELN
AFAL209			19991210	0550	V				
JPRLT401	00000000	0000				000000	0000	111	

As marcas em tons de cinza, na tabela 6.3, indicam onde se encontram os principais problemas, sendo que para a identificação e classificação dos problemas relacionados à qualidade do conteúdo dos atributos (dados), propõem-se o uso da lógica paraconsistente anotada de dois valores – LPA2v. A forma como esta lógica será aplicada é apresentada nas próximas seções.

Na Tabela 6.3, foram apresentados alguns dos problemas encontrados, normalmente, em um banco de dados relacional com incompletudes, entre os quais se destacam: ausência de valores em vários atributos e valores fora do padrão esperado.

6.5 Relação entre Banco de Dados Relacional e Lógica Paraconsistente

O modelo de dados relacional possui duas características importantes, os valores dos dados são orientados a sua natureza e possui um rico e simples conjunto de operadores algébricos, além de apresentar uma boa fundamentação teórica baseada em lógica clássica de primeira-ordem. Logo, a combinação de uma respeitável base teórica

e a facilidade de desenvolvimento de projetos de bancos de dados, transformou este modelo em sucesso imediato [Hongai et al., 2005].

O grande sucesso do modelo de banco de dados relacional não impediu que o mesmo apresentasse limitações, as quais estão relacionadas à falta de aplicabilidade para situações não clássicas, que envolvem ocorrência de dados com algum tipo de incompletudes.

As incompletudes dos dados, normalmente, vêm no formato de dados nulos ou perdidos. Para este tipo de incompletude há tratamentos tradicionais como os métodos de imputação de dados [Farhangfar & Kurgan, 2007], mas em muitos casos, os mesmos introduzem mais problemas do que realmente resolvem a incompletude.

No caso de dados inconsistentes e indeterminados, não houve inicialmente preocupação, por parte dos pesquisadores da área, para o desenvolvimento de soluções para o problema. Mas com o advento da utilização dos sistemas de conhecimento, os dados inconsistentes e indeterminados passaram a receber atenção, pois para esse tipo de estrutura a incoerência, inconsistência, indeterminação ou contradição não são aceitáveis.

Em banco de dados, a lógica paraconsistente pode identificar a incompletude existente entre os dados, já que neste caso, as mesmas ocorrem em nível de registros, isto é, em uma linha da tabela bidimensional [Bagai & Sunderraman, 1985a]. A inconsistência ou indeterminação no registro apresentam-se de forma diferente, podendo ser valores apenas nulos ou estarem no formato de valores totalmente fora do padrão esperado.

A inconsistência em banco de dados relacional ocorre quando o valor do atributo está fora do padrão esperado. Observe o atributo POT_INTERROMP_OCORRE (atributo nº 7), na Tabela 6.4, o mesmo é um atributo dependente do atributo COD_EQUIPAM_OCORRE, isto é, se existe uma interrupção no funcionamento do equipamento deverá existir uma potência interrompida para o referido equipamento e, a

mesma, deverá ser diferente de zero. Tanto em banco de dados relacional como em Lógica Paraconsistente, a inconsistência ocorre quando existe um valor contido no atributo, mas o mesmo encontra-se fora do padrão esperado.

Tabela 6. 4. Exemplo de Inconsistência entre os Dados dos Atributos Relevantes da Tabela Ocorrência.

1	2	3	4	5	6	7
GJAL204	20041124	2252	20041206	0306	V	00000
PVTF604	00000000	0000	20000628	2245	V	000000
PVDB601	00000000	0000	20000619	0841	V	000000
PVTF605	00000000	0000	20000628	2245	V	000000
RLTF401	00000000	0000	19991210	0622	V	000000
RLAL203	00000000	0000	19991210	9745	V	000004
RLBR201	00000000	0000	19991210	0625	V	000003
AFAL209			19991210	0550	V	

Inconsistência

A indeterminação em banco de dados ocorre quando um dos atributos, que faz parte do registro, for atributo chave (primária ou estrangeira), a inconsistência passa a ser denominada de indeterminação, já que pode estar relacionada às restrições de integridade de dados (chave primária) e relacional/referencial (chave estrangeira). Considere as definições de chaves:

- Chave Primária (*primary key*) refere-se ao atributo ou ao conjunto de atributos que concatenados garantem a não repetição dos registros de uma tabela, isto é, impedem a redundância de dados. Nos atributos chaves primárias, não pode haver valores nulos nem repetição do conjunto concatenado [Atzeni et al., 2000].
- Chave estrangeira (*foreign key*) é um atributo (coluna ou campo), que aponta para a chave primária de outra tabela, isto é, passa a existir uma relação entre essas duas tabelas. A finalidade da chave estrangeira é garantir a integridade dos dados referenciais [Codd, 1986].

A Tabela 6.5 apresenta exemplo de indeterminação entre os dados que, neste

caso, ocorre pela ausência de conteúdo no atributo COD_EQUIPAMENTO_OCORRE.

Tabela 6. 5. Exemplo de Indeterminação entre os Dados dos Atributos Relevantes.

1	2	3	4	5	6	7
	20041124	2252	20041206	0306	V	00000
	00000000	0000	20000628	2245	V	000000
	00000000	0000	20000619	0841	V	000000
PVTF605	20051125	0000	20000628	2245	V	000000
RLTF401	19991209	0000	19991210	0622	V	000000
RLAL203	19991210	0000	19991210	9745	V	000004
RLBR201	19991210	0000	19991210	0625	V	000003
AFAL209			19991210	0550	V	

Indeterminação

O atributo COD_EQUIPAMENTO_OCORRE (atributo nº 1) é parte da chave primária da tabela ocorrência, mesmo que esta análise seja realizada apenas nos atributos relevantes descobertos, após a aplicação da Teoria de Conjuntos Aproximados na Tabela Ocorrência, as características dos atributos, juntamente com os conteúdos são analisados, sendo assim, a ausência deste valor, no referido atributo causa indeterminação, pois o mesmo viola as restrições de integridade básica de banco de dados relacional, as quais estão vinculadas à chave primária e à chave estrangeira.

As restrições de integridade são regras executadas sempre que ocorrer a atualização do banco de dados, as quais devem, sempre, impedir a existência de uma atualização inconsistente, semelhante à apresentada na Tabela 6.5.

As regras de integridade possibilitam a criação de relacionamentos entre as tabelas de banco de dados relacional. Sendo que os relacionamentos apresentam cardinalidade, isto é, a forma como os relacionamentos ocorrem entre as tabelas. Logo, a cardinalidade mostra quantas vezes um atributo chave poderá ocorrer na tabela em que há relacionamento, sendo que esta quantidade pode ser infinita ou nenhuma [Korth & Silberschatz, 1995].

A cardinalidade, não somente indica a forma (quantidade) como o relacionamento vai ocorrer, mas indica a necessidade de haver chaves primárias e/ou estrangeiras para garantirem este relacionamento.

Qual o motivo para a inexistência das regras de integridade do modelo relacional do Sistema INFO_OPR? A resposta pode estar na origem dos dados, já que os grandes bancos de dados têm, geralmente, na sua composição dados de sistemas legados e dados de sistemas convencionais.

A origem heterogênea dos dados pode acarretar incompletudes em registros, pois no momento da importação dos dados legados, o banco de dados não pode utilizar as regras de integridade (chaves), isto é, essas regras não podem estar ativas para que os dados possam se “acomodar” dentro das tabelas.

6.6 Aplicação de LP na Identificação e Classificação das Incompletudes dos Dados entre os Atributos Relevantes da Base de dados do INFO_OPR

Os dados que compõem o conjunto de atributos relevantes da base de dados INFO_OPR possuem um ambiente sugestível à aplicação da Lógica Paraconsistente Anotada de Dois Valores – LPA2v, já que a mesma possui os graus de crença e descrença que sustentará a identificação e classificação da incompletude dos dados. A seguir, são apresentados os passos para a identificação dos tipos de incompletudes entre os valores dos dados pertencentes aos atributos relevantes:

1. Identificação das incompletudes

Para a identificação das incompletudes, são analisados os atributos de forma individual:

- **Valores presentes:** Em banco de dados, utiliza-se a expressão “*dado consistente*⁸”, para aqueles registros que apresentam a integridade de dados,

⁸ Este tipo de dado é apresentado, apenas para lembrar que entre os valores dos dados existe o tipo íntegro e consistente, mas não é considerado uma incompletude.

isto é, os dados têm conteúdos verdadeiros, consistentes e íntegros. A figura 6.1 apresenta um exemplo para este tipo de dado, pois no primeiro registro da figura, os valores dos atributos estão todos dentro de um padrão esperado e são diferentes de zero, além de obedecerem às restrições de integridade (chave primária e estrangeira);

	COD_EQUIPA	DATA_INI	HORA	DATA_FIN	HORA	TIP_C	POT_INTE	CORRE	COD_	COD_	COD_EM	COD_EM	LIGA	CODI	CODIGO_REDE
▶	AEAL201	20010510	1543	20010510	1544	I	002400	0118	31	42	CRO	CRO	2	E06	DDO
	AEAL201	20010510	1544	20010511	1835	V	000000	0000	6	35	CRO	CRO	0	E02	OPR
	AEAL201	20010527	1230	20010531	1259	V	000000	0000	6	35	ELN	ELN	0	E02	OPR
	AEAL201	20010601	1325	20010602	0630	V	000000	0000	6	35	CRO	CRO	0	E02	OPR
	AEAL201	20010602	1558	20010603	0634	V	000000	0000	6	42	CRO	CRO	0	E02	OPR
	AEAL201	20010603	1335	20010614	0947	V	002200	0108	6	35	ELN	ELN	0	E02	OPR

Figura 6. 1. Exemplo de Registros Consistentes

- **Valores Indeterminados:** Em banco de dados usa-se a expressão “**dado indeterminado ou valor indeterminado**”, para aqueles atributos que apresentam ausência de valores. Na figura 6.2, os atributos (COD_EQUIPAM_OCORRE, HORA_INICIAL_OCORRE, DATA_FINAL_OCORRE, HORA_FINAL_OCORRE) apresentam ausência de valores, isto é, são atributos logicamente indeterminados, já que ter apenas a informação do atributo DATA_INICIAL_OCORRE não oferece condições para análise dos dados restantes do registro;

	COD_EQUIPAM	DATA_INICIAL_	HORA_INICIAL_	DATA_FINAL_	HORA_FINAL_
		20070705			
		20070627			
		20070626			
		20070624			
	AEAL201	20050923	1141	20050923	1152
	AEAL201	20050920	0728	20050920	0815
	AEAL201	20050923	1152	20050923	1210
	AEAL201	20050921	1737	20050921	1738
	AEAL201	20050923	1210	20050924	0749
	AEAL201	20050921	1740	20050923	1116
	AEAL201	20050921	1738	20050921	1740
	AEAL201	20050923	1116	20050923	1141

Figura 6. 2. Exemplo de Indeterminação por Ausência de Valores

- **Valores Inconsistentes:** Em banco de dados, os valores de dados são considerados inconsistentes quando forem zero ou qualquer valor fora do padrão esperado, neste caso usa-se a expressão “**dado inconsistente**”. Como exemplo, tem-se o atributo CORRENTE_OCORRE (atributo nº 8) na Tabela 6.6, onde o referido atributo apresenta inconsistência (valores em destaque com cor), já que a Corrente de um equipamento depende da potência do equipamento, não podendo ser um valor nulo, pois existe valor para o atributo POTENCIA_OCORRE (atributo nº 7).

Tabela 6. 6. Exemplos de Inconsistências entre os Dados dos Atributos Relevantes

1	2	3	4	5	6	7	8
PHAL207	19961103	1540	19961103	1559	I	001600	0000
PLAL206	19940530	0709	19940530	1643		001600	0000
PHAL207	19961110	1455	19961110	1459	I	001600	0000
PLAL206	19950623	1051	19960623	1051		001600	0000
PHAL206	19940402	1254	19940402	1258	I	001600	0079
PLAL206	19960911	0802	19960911	1154	V	001600	0000
PLAL206	19950416	0808	19950416	0812		001600	0000
PHAL207	19950508	0653	19950508	1801		001600	0000
PHAL206	19951022	1249	199510	1257		001600	0000
PLAL205	19900615	0955	19900615	1017	V	001600	0079

2. Classificação das Incompletudes

Para auxiliar a classificação dos dados em relação aos estados dos conteúdos será utilizada a notação da lógica paraconsistente em que os valores dos sinais dos graus de crença e descrença podem ser binários [Martins et al., 2007].

A classificação dos dados e registros do banco de dados a partir da Lógica Paraconsistente Anotada de Dois Valores – LPA2v considera os pontos do Quadrado Unitário do Plano Cartesiano – QUPC. Onde a área inferior do QUPC refere-se ao triângulo inferior à linha perfeitamente inconsistente – LPI. Esta área encontra-se

hachurada e apresenta os pontos necessários e suficientes para a classificação dos dados contidos nos atributos do banco de dados analisados neste trabalho. A figura 6.3, apresentada a seguir, mostra o QUPC com a área delimitada, para a solução do problema neste trabalho.

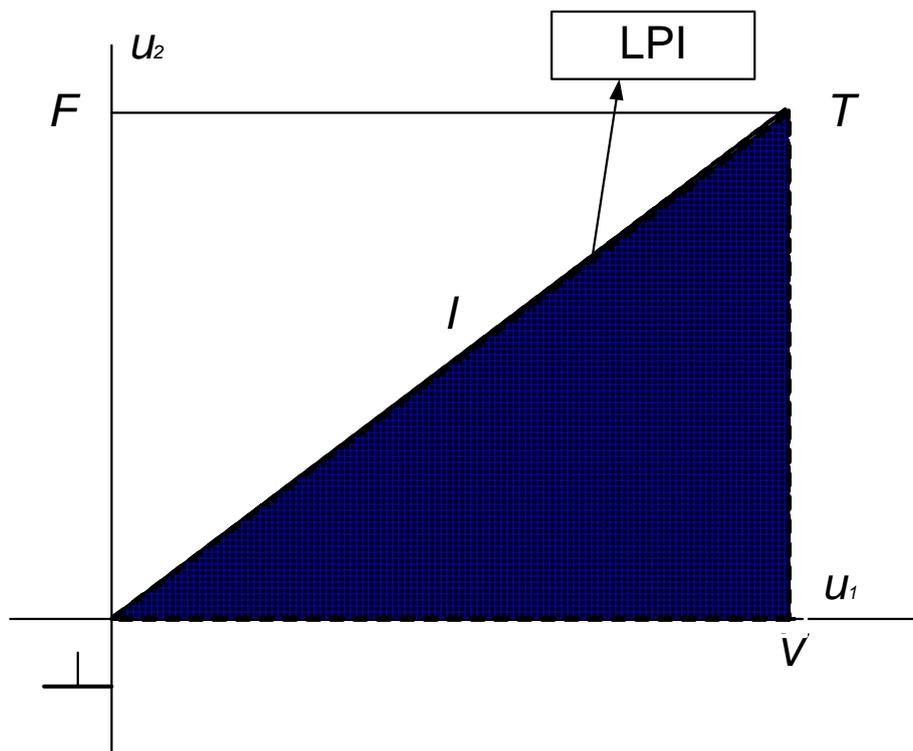


Figura 6. 3. Área do QUPC Utilizada para Classificar Dados e Registros de Banco de Dados
Fonte: Adaptado de [Martins et al., 2007].

A classificação dos dados utilizará apenas os pontos situados abaixo da LPI:

- Ponto (1,0) indica a verdade (V) ou crença total;
- Ponto (1,1) indica um tipo de inconsistência (T) ou crença totalmente inconsistente;
- Ponto (0,0) indica outro tipo de inconsistência (\perp) ou crença totalmente indeterminada.

Logo, os estados dos valores contidos nos atributos e, conseqüentemente nos registros deverão encontra-se na área do triângulo inferior delimitado pela linha LPI (Linha Perfeitamente Inconsistente) da figura 6.3, os quais deverão ser apenas: Consistentes (V), Inconsistentes (T) e Indeterminados (\perp).

Para facilitar o entendimento e auxiliar o resultado da classificação dos dados e posteriormente dos registros apresenta-se, a seguir, algumas considerações:

1. Para a classificação dos registros utiliza-se o conectivo **OU** (\vee), já que através da aplicação do mesmo, é possível encontrar o estado geral dos registros de um banco de dados;
2. Considere P e Q como proposições nas quais serão aplicados o conectivo **OU** (\vee), estas proposições possuem dois sinais anotados, cada um composto pelo seu grau de crença e descrença, sendo o sinal P: $P_P(\mu_{1P}, \mu_{2P})$ e o sinal Q: $Q_Q(\mu_{1Q}, \mu_{2Q})$. Onde (μ_{1P}) é o grau de crença do sinal P e (μ_{2P}) é o grau de descrença do sinal P e, (μ_{1Q}) é o grau de crença do sinal Q e (μ_{2Q}) é o grau de descrença do sinal Q;
3. Para obtenção dos estados lógicos resultantes do conectivo **OU** (\vee), minimiza-se os graus de descrença (μ_1) nos dois sinais anotados P e Q, obtendo-se o grau de crença resultante (μ_{1R});
4. Para obtenção dos estados lógicos resultantes do conectivo **OU** (\vee), minimiza-se os graus de descrença (μ_2) nos dois sinais anotados P e Q, obtendo-se o grau de descrença resultante (μ_{2R});
5. Com os graus de crença e descrença resultante (μ_{1R}, μ_{2R}) é realizada uma análise no QUPC, obtendo-se, desta forma o estado resultante de saída.

Para encontrar o ponto resultante no QUCP da aplicação do conectivo **OU** (\vee) nos sinais anotados P e Q, os quais são binários independentes, é necessário maximizar os valores dos graus de crença e minimizar os graus de descrença das duas proposições. O resultado será o vértice inferior direito do retângulo da figura 6.3, isto é, o triângulo representado pelos segmentos $\perp TV \perp$.

A Tabela 6.7, exibida a seguir, apresenta os valores-verdade da aplicação do conectivo **OU** (\vee), para os sinais de entrada μ_1 e μ_2 , os quais são binários independentes. O resultado da combinação dos estados de P e Q apresenta todos os possíveis estados lógicos do QUCP para o conectivo **OU** (\vee), sendo que estes resultados facilitam a descoberta dos estados dos conteúdos dos atributos e dos registros em banco de dados.

Tabela 6. 7. Tabela com Valores - Verdade da Aplicação do Conectivo **OU** (\vee).
Fonte: Adaptado de [Martins et al., 2007].

N° de Estados	Sinais de Entrada						Conectivo OU (\vee)		Estado Resultante
	P	Q	Grau de Crença		Grau de Descrença		μ_{1R} max	μ_{2R} min	P \vee Q
			μ_{1P}	μ_{1Q}	μ_{2P}	μ_{2Q}			
1	T(1,1)	T(1,1)	1	1	1	1	1	1	T
2	V(1,0)	V(1,0)	1	1	0	0	1	0	V
3	F(0,1)	F(0,1)	0	0	1	1	0	1	F
4	$\perp(0,0)$	$\perp(0,0)$	0	0	0	0	0	0	\perp
5	T(1,1)	V(1,0)	1	1	1	0	1	0	V
6	T(1,1)	F(0,1)	1	1	1	1	1	1	T
7	T(1,1)	$\perp(0,0)$	1	1	1	0	1	0	V
8	V(1,0)	F(0,1)	1	0	0	1	1	0	V
9	V(1,0)	$\perp(0,0)$	1	0	0	0	1	0	V
10	F(0,1)	$\perp(0,0)$	0	0	1	0	0	0	\perp

A seguir, é exibida a Tabela 6.8 com os atributos relevantes e a respectiva numeração destes atributos. Os atributos foram numerados, para facilitar a exibição no formato de tabela e para a classificação dos registros.

Tabela 6. 8. Atributos Relevantes com Numeração.

Atributo	Numeração
COD_EQUIPAM_OCORRE	1
DATA_INICIAL_OCORRE	2
HORA_INICIAL_OCORRE	3
DATA_FINAL_OCORRE	4
HORA_FINAL_OCORRE	5
TIP_OCORRE	6
POT_INTERROMP_OCORRE	7
CORRENTE_OCORRE	8
COD_CLASSE_OCORRE	9
COD_ORIGEM_OCORRE	10
COD_EMPRESA_ORIGEM_OCORRE	11
COD_EMPRESA_RESPONSAVEL_OCORRE	12
LIGADO_OCORRE	13
CODIGO_ESTADO_OPERACIONAL_OCORRE	14
CODIGO_REDE_OPERACAO_OCORRE	15

A Tabela 6.9 apresenta um exemplo contendo nove registros com os respectivos conteúdos (atributos e dados). Neste conjunto exemplo, será realizado a classificação dos dados utilizando LPA2v.

Tabela 6. 9. Exemplo de Dados para Classificação do Registro Utilizando LPA2v

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	20070627													
	20070626													
	20070624													
AEAL201	20050923	1141	20050923	1152		000000	00				ELN			OPR
RLAL203	00000000	0000	19991210	0745	V	001000			35	ELN	ELN			
PVAL208	20011010	1834	20011010	1843	V	001200	0059	3V	33	CRO	CRO	0		DER
PVAL208	20011010	1449	20011010	1834	V	001500	0074	6	42	CRO	CRO	0	E02	OPR
PVAL208	20011010	1434	20011010	1449	V	001500	0074					0		
PLAL204	19980609	0001	19980609	0004	V	005000	0246	221	35	ELN	ELN	0		DCI

As tabelas construídas e exibidas, a seguir, apresentam a classificação dos valores dos dados através da observação do conteúdo dos atributos apresentados na Tabela 6.9. Para a classificação é necessário o uso dos graus de crença (μ_1) e grau de descrença (μ_2) da LPA2v.

→ Apresentação das Tabelas para Classificação dos Conteúdos dos Atributos Relevantes Exibidos na Tabela 6.9.

A identificação dos conteúdos dos atributos utilizará LPA2v, já que a mesma possui a seguinte notação: $P(\mu_1, \mu_2)$, onde P são os conteúdos dos atributos analisados e o par (μ_1, μ_2) , o grau de crença e descrença do conteúdo do atributo. Para a classificação do conteúdo dos atributos relevantes, os possíveis valores para (μ_1, μ_2) são 0 e 1. Ao final da classificação, encontra-se uma seqüência de pares de zeros (0) e uns (1) indicando o estado do conteúdo do atributo. Considere que os estados possíveis para os conteúdos dos atributos são: \perp (0,0) - Indeterminado, T (1,1) - Inconsistente e V(1,0) – Consistente.

As tabelas a seguir, são compostas pelo número do atributo (**Nº Atributo**), os graus de crença (μ_1) e descrença (μ_2) do conteúdo dos atributos e o resultado que apresentará o estado do conteúdo do atributo, podendo ser consistente, inconsistente e indeterminado. Nesta etapa, não será aplicado nenhum conectivo para a descoberta do estado do conteúdo do atributo, apenas é descoberto o estado do conteúdo individualmente por atributo. A Tabela 6.4 apresenta os atributos com sua respectiva numeração, a qual foi apresentada com o objetivo de facilitar a visualização da classificação dos conteúdos dos registros contidos na Tabela 6.9.

Tabela 6. 10. Classificação dos Dados dos Atributos Relevantes do Primeiro Registro

Nº Atributo	μ_1	μ_2	Resultado
1	0	0	$\perp = (0,0)$
2	1	0	V = (1,0)
3	0	0	$\perp = (0,0)$

Continuação da Tabela 6.10.

Nº Atributo	μ_1	μ_2	Resultado
4	0	0	$\perp = (0,0)$
5	0	0	$\perp = (0,0)$
6	0	0	$\perp = (0,0)$
7	0	0	$\perp = (0,0)$
8	0	0	$\perp = (0,0)$
9	0	0	$\perp = (0,0)$
10	0	0	$\perp = (0,0)$
11	0	0	$\perp = (0,0)$
12	0	0	$\perp = (0,0)$
13	0	0	$\perp = (0,0)$
14	0	0	$\perp = (0,0)$
15	0	0	$\perp = (0,0)$

Observações:

1. Não é necessário classificar os dados referentes ao segundo e terceiro registro, já que os mesmos apresentam as mesmas características do primeiro registro;
2. Nesta tabela, os três primeiros atributos da tabela, são aqueles que compõem a chave primária, como os mesmos são parcialmente indeterminados, isto acarreta uma violação da regra de integridade de dados.

Tabela 6. 11. Classificação dos Dados dos Atributos Relevantes do Quarto Registro.

Atributos	μ_1	μ_2	Resultado
1	1	0	$V = (1,0)$
2	1	0	$V = (1,0)$
3	1	0	$V = (1,0)$
4	1	0	$V = (1,0)$
5	1	0	$V = (1,0)$
6	0	0	$\perp = (0,0)$

Continuação da tabela 6.11.

Atributos	μ_1	μ_2	Resultado
7	1	1	T = (1,1)
8	1	0	V = (1,0)
9	0	0	\perp = (0,0)
10	0	0	\perp = (0,0)
11	0	0	\perp = (0,0)
12	1	0	V = (1,0)
13	0	0	\perp = (0,0)
14	0	0	\perp = (0,0)
15	1	0	V = (1,0)

Tabela 6. 12. Classificação dos Dados dos Atributos Relevantes do Quinto Registro

Atributos	μ_1	μ_2	Resultado
1	1	0	V = (1,0)
2	1	1	T = (1,1)
3	1	1	T = (1,1)
4	1	0	V = (1,0)
5	1	0	V = (1,0)
6	0	0	\perp = (0,0)
7	1	1	T = (1,1)
8	1	1	T = (1,1)
9	0	0	\perp = (0,0)
10	1	0	V = (1,0)
11	1	0	V = (1,0)
12	1	1	V = (1,0)
13	0	0	\perp = (0,0)
14	0	0	\perp = (0,0)
15	0	0	\perp = (0,0)

Tabela 6. 13. Classificação dos Dados dos Atributos Relevantes do Sexto Registro.

Atributos	μ_1	μ_2	Resultado
1	1	0	V = (1,0)
2	1	0	V = (1,0)
3	1	0	V = (1,0)
4	1	0	V = (1,0)
5	1	0	V = (1,0)
6	1	0	V = (1,0)
7	1	0	V = (1,0)
8	1	0	V = (1,0)
9	1	0	V = (1,0)
10	1	0	V = (1,0)
11	1	0	V = (1,0)
12	1	0	V = (1,0)
13	1	0	V = (1,0)
14	0	0	$\perp = (0,0)$
15	1	0	V = (1,0)

Tabela 6. 14. Classificação dos Dados dos Atributos Relevantes do Sétimo Registro.

Atributos	μ_1	μ_2	Resultado
1	1	0	V = (1,0)
2	1	0	V = (1,0)
3	1	0	V = (1,0)
4	1	0	V = (1,0)
5	1	0	V = (1,0)
6	1	0	V = (1,0)
7	1	0	V = (1,0)
8	1	0	V = (1,0)
9	1	0	V = (1,0)
10	1	0	V = (1,0)
11	1	0	V = (1,0)
12	1	0	V = (1,0)

Continuação da Tabela 6.14.

Atributos	μ_1	μ_2	Resultado
13	1	0	$V = (1,0)$
14	1	0	$V = (1,0)$
15	1	0	$V = (1,0)$

Tabela 6. 15. Classificação dos Dados dos Atributos Relevantes do Oitavo Registro.

Atributos	μ_1	μ_2	Resultado
1	1	0	$V = (1,0)$
2	1	0	$V = (1,0)$
3	1	0	$V = (1,0)$
4	1	0	$V = (1,0)$
5	1	0	$V = (1,0)$
6	1	0	$V = (1,0)$
7	1	0	$V = (1,0)$
8	1	0	$V = (1,0)$
9	0	0	$\perp = (0,0)$
10	0	0	$\perp = (0,0)$
11	0	0	$\perp = (0,0)$
12	0	0	$\perp = (0,0)$
13	1	0	$V = (1,0)$
14	0	0	$\perp = (0,0)$
15	0	0	$\perp = (0,0)$

Tabela 6. 16. Classificação dos Dados dos Atributos Relevantes do Nono Registro.

Atributos	μ_1	μ_2	Resultado
1	1	0	$V = (1,0)$
2	1	0	$V = (1,0)$
3	1	0	$V = (1,0)$
4	1	0	$V = (1,0)$

Continuação da Tabela 6.16.

Atributos	μ_1	μ_2	Resultado
5	1	0	V = (1,0)
6	1	0	V = (1,0)
7	1	0	V = (1,0)
8	1	0	V = (1,0)
9	1	0	V = (1,0)
10	1	0	V = (1,0)
11	1	0	V = (1,0)
12	1	0	V = (1,0)
13	1	0	V = (1,0)
14	1	0	V = (1,0)
15	1	0	V = (1,0)

→ Classificações dos Registros da Tabela 6.9

Para a efetiva classificação do estado dos registros da tabela com os atributos relevantes, se faz necessário a definição da chave primária desta tabela, já que o banco de dados é relacional, sendo indispensável à utilização da regra de integridade de dados. Logo, a chave primária desta tabela é composta pelos atributos COD_EQUIPAM_OCORRE, DATA_INICIAL_OCORRE e HORA_INICIAL_OCORRE, pois estes três atributos concatenados garantem a não repetição de registros.

Para a classificação dos registros é necessária a utilização da regra de integridade de dados (chave primária), pois os atributos que compõem a chave não podem ser analisados de forma isolada. No caso de um atributo ou mais (participantes da chave primária) possuir valores indeterminados ou inconsistentes implicará em uma indeterminação do estado do registro, pois os atributos não podem ser analisados de forma isolada no momento da classificação do registro, sendo permitido a análise isolada apenas do conteúdo (dado) do atributo.

Com a identificação do estado dos conteúdos dos atributos relevantes e

aplicando-se o conectivo **OU** (\vee) para todos os pares de valores (μ_1, μ_2) desta classificação, pode-se, neste momento, obter um estado geral para o registro. Nesta etapa utiliza-se a Tabela 6.1, na qual estão todos os possíveis resultados para o conectivo **OU** (\vee).

1. Classificação do Primeiro Registro

$$(0,0) \vee (1,0) \vee (0,0) = (1,0)$$

Estado Consistente – estado encontrado

Estado Indeterminado (\perp) – estado real

Ao aplicar o conectivo **OU** (\vee), sem levar em consideração a regra de integridade de dados, o resultado da classificação do registro é um estado Consistente, o que não é verdade, já que este registro apresenta ausência de quase todos os conteúdos dos outros atributos. Logo, é necessário aplicar a regra da chave primária, pois o resultado da classificação deste registro passa ser \perp (indeterminado), pois parte da chave primária (COD_EQUIPAM_OCORRE, HORA_INICIAL_OCORRE) tem ausência de valor, implicando de forma direta no estado de indeterminação do registro e consequentemente aconselha-se a remoção do mesmo da tabela.

2. Classificação do Segundo Registro

$$(0,0) \vee (1,0) \vee (0,0) = (1,0)$$

Estado Consistente – estado encontrado

Estado Indeterminado (\perp) – estado real

Situação idêntica do primeiro registro, por isso o registro é indeterminado.

3. Classificação do Terceiro Registro

$$(0,0) \vee (1,0) \vee (0,0) = (1,0) \mathbf{V}$$

Estado Consistente – estado encontrado

Estado Indeterminado (\perp) – estado real

Situação idêntica do primeiro registro, por isso o registro é indeterminado.

4. Classificação do Quarto Registro

$(1,0) \vee (1,0) \vee (1,0) \vee (1,0) \vee (1,0) \vee (0,0) \vee (1,1) \vee (1,0) \vee (0,0) \vee (0,0) \vee (1,0) \vee (0,0) \vee (0,0) \vee (1,0) = (0,1) \mathbf{V}$ - Estado Consistente

Este registro apresenta atributos com conteúdos inconsistentes e indeterminados, mas como não estão localizados nos atributos que fazem parte da chave primária, não haverá implicações na sua classificação apenas com a aplicação do conectivo OU (\vee).

5. Classificação do Quinto Registro

$(1,0) \vee (1,1) \vee (1,1) \vee (1,0) \vee (1,0) \vee (1,0) \vee (1,0) \vee (1,1) \vee (0,0) \vee (1,0) \vee (1,0) \wedge (0,0) \wedge (0,0) \wedge (1,0) = (1,0) \mathbf{V}$

Estado Consistente – estado encontrado

Estado Indeterminado (\perp) – estado real

Apesar do estado encontrado para o registro ser Consistente, o mesmo apresenta inconsistência em dois atributos (DATA_INICIAL_OCORRE e HORA_INICIAL_OCORRE) da chave primária, acarretando na indeterminação do estado do registro e conseqüentemente a sua remoção da tabela. A inconsistência dos conteúdos dos atributos DATA_INICIAL_OCORRE e HORA_INICIAL_OCORRE são relacionadas o conteúdo dos mesmos ser igual a zero. Logo, a inconsistência na chave primária é determinada em função do atributo não poder admitir uma data e uma hora com seus conteúdos iguais a zero.

6. Classificação do Sexto Registro

$(1,0) \vee (1,0) \vee (0,0) = (1,0) \mathbf{V}$ – Estado Consistente

mesmo grau de importância de uma chave primária no conjunto de atributos relevantes.

6.7 Resumo da Metodologia de Avaliação de Atributos Relevantes de Grandes Bases de Dados Incompletas.

→ **Passo 1:** Verificar a existência do processo de normalização do banco de dados e a dinâmica de armazenamento.

Nesta etapa, a existência do processo de normalização do banco de dados e do fluxo e dinâmica de armazenamento do banco de dados, facilitam a identificação das principais tabelas do banco de dados. No caso do Banco de dados do INFO_OPR, observou-se o fluxo e a dinâmica do armazenamento dos dados, para a identificação da principal tabela. As tabelas do banco de dados INFO_OPR foram descritas no capítulo 4.

→ **Passo 2:** Aplicar a Teoria de Conjuntos Aproximados - TCA no conjunto das tabelas do banco de dados do INFO_OPR.

A TCA, geralmente, é aplicada em todo o banco de dados e individualmente em cada tabela, para realizar a redução de atributos e registros, mas neste caso, é aplicada individualmente na tabela ocorrência, já que, esta tabela é a responsável pelo fluxo e dinâmica de armazenamento do banco de dados.

Na prática isto significa que, a principal tabela do banco de dados, pode ser descoberta identificando o processo de normalização, caso exista, ou avaliando-se o fluxo e a dinâmica de armazenamento do banco de dados.

Após a descoberta desta tabela principal, aplica-se a teoria de conjuntos aproximados para a descoberta do conjunto de atributos relevantes da base de dados. Todo o processo de aplicação da TCA está descrito no Capítulo 3 e nas seções 5.2, 5.3 e 5.4 deste capítulo.

→ **Passo 3:** Aplicar a Lógica Paraconsistente nos valores de dados contidos nos atributos relevantes.

Nesta etapa a lógica paraconsistente é usada como ferramenta que identifica e classifica os dados com incompletudes, oriundos do banco de dados INFO_OPR, mas que apresentam algum grau de incerteza e contradição. Para efetivar o uso desta ferramenta foi realizado o seguinte:

- Identificação dos tipos de valores de dados - Análise dos tipos de incompletudes dos dados, contidos nos registros dos atributos relevantes da base INFO_OPR. Nesta etapa identificam-se os tipos de dados, os quais darão base para a identificação das incompletudes. A seguir, apresentam-se apenas os tipos de dados que serão utilizados:

1. Verdadeiros ou consistentes – V:

São os registros com dados íntegros e consistentes;

2. Valores indeterminados - ⊥:

Em banco de dados usa-se a expressão dado indeterminado ou valor indeterminado, para aqueles atributos que apresentam ausência de valores. Este estado do registro pode indicar a remoção do mesmo do banco de dados.

3. Valores Inconsistentes – T:

Em banco de dados, os valores de dados são considerados inconsistentes quando forem zero ou qualquer valor fora do padrão esperado. Este estado do registro pode indicar a necessidade de tratamento do conteúdo dos atributos com problemas.

- Classificação dos dados contidos nos registros dos atributos relevantes, este procedimento tem como objetivo analisar o conteúdo dos atributos relevantes e posteriormente a classificação de todo o conteúdo do registro, isto é, a avaliação de todo registro em função da classificação de cada atributo:

1. Nesta etapa utiliza-se a LPA2v, a qual possui a seguinte notação: $p(\mu_1, \mu_2)$, onde p são os atributos analisados e o (μ_1, μ_2) , o grau de crença e descrença do conteúdo do atributo.
2. Após a identificação do conteúdo dos atributos relevantes, aplica-se o conectivo OU (\vee) para todos os pares de valores (μ_1, μ_2) de cada atributo. Ao final da aplicação do conectivo, se obtém um estado geral para o registro, o qual pode ser \perp , T e V . A identificação e classificação do conteúdo dos atributos e do registro estão descritas na seção 6.5, deste capítulo. Este procedimento é extremamente importante, pois indicará quais registros podem ser removidos e quais precisam ser tratados, esta classificação avalia as condições do banco de dados com atributos relevantes.
3. Nesta etapa, o estado do conteúdo dos atributos e posteriormente dos registros, indicará que procedimentos deverão ser executados em relação ao registro. Logo, um registro classificado como Inconsistente indica problemas que podem ser tratados, mas registros que apresentam o estado Indeterminado podem indicar a remoção do registro, já que os problemas apresentados impedem a sua utilização para extração de conhecimento confiável.

As principais motivações para a identificação e classificação das incompletudes do banco de dados estão, principalmente, relacionadas ao fato de que são necessários dados íntegros, para recuperação de informação confiável e transformação em conhecimento útil a tomada de decisão.

→ Apresentação dos passos e do diagrama com o resumo da metodologia de Avaliação de atributos relevantes de grandes bases de dados incompletas.

Passo 1: Identificação do banco de dados;

Passo 2: Identificação do Processo de Normalização e/ou Fluxo e dinâmica de armazenamento dos dados;

Passo 3: Identificação da principal tabela do banco de dados;

Passo 4: Aplicação da teoria de conjuntos Aproximados na principal tabela do banco de dados;

Passo 5: Descoberta dos atributos relevantes da principal tabela do banco de dados;

Passo 6: Composição de uma nova tabela apenas com os atributos relevantes da principal tabela do banco de dados;

Passo 7: Aplicar LPA2v para identificar o tipo de conteúdo dos atributos relevantes da nova tabela;

Passo 8: Aplicar LPA2v para classificar os dados e os registros da nova tabela;

Passo 9: Descoberta dos estados dos dados e dos registros da nova tabela (Inconsistente, Consistente e Indeterminado)

Passo 10: Se o registro for consistente – mantê-lo na nova tabela; Se registro inconsistente – manter na nova tabela; Se registro indeterminado – excluir da nova tabela;

Passo 11: Criação da chave primária na nova tabela sem os registros indeterminados;

Passo 12: Nova tabela com atributos relevantes, sem registros indeterminados e sem redundância de registros.

A figura 6.4 exibe o diagrama com o resumo para aplicação da metodologia de avaliação de atributos relevantes. Este diagrama tem como objetivo facilitar a visualização e entendimento da metodologia apresentada.

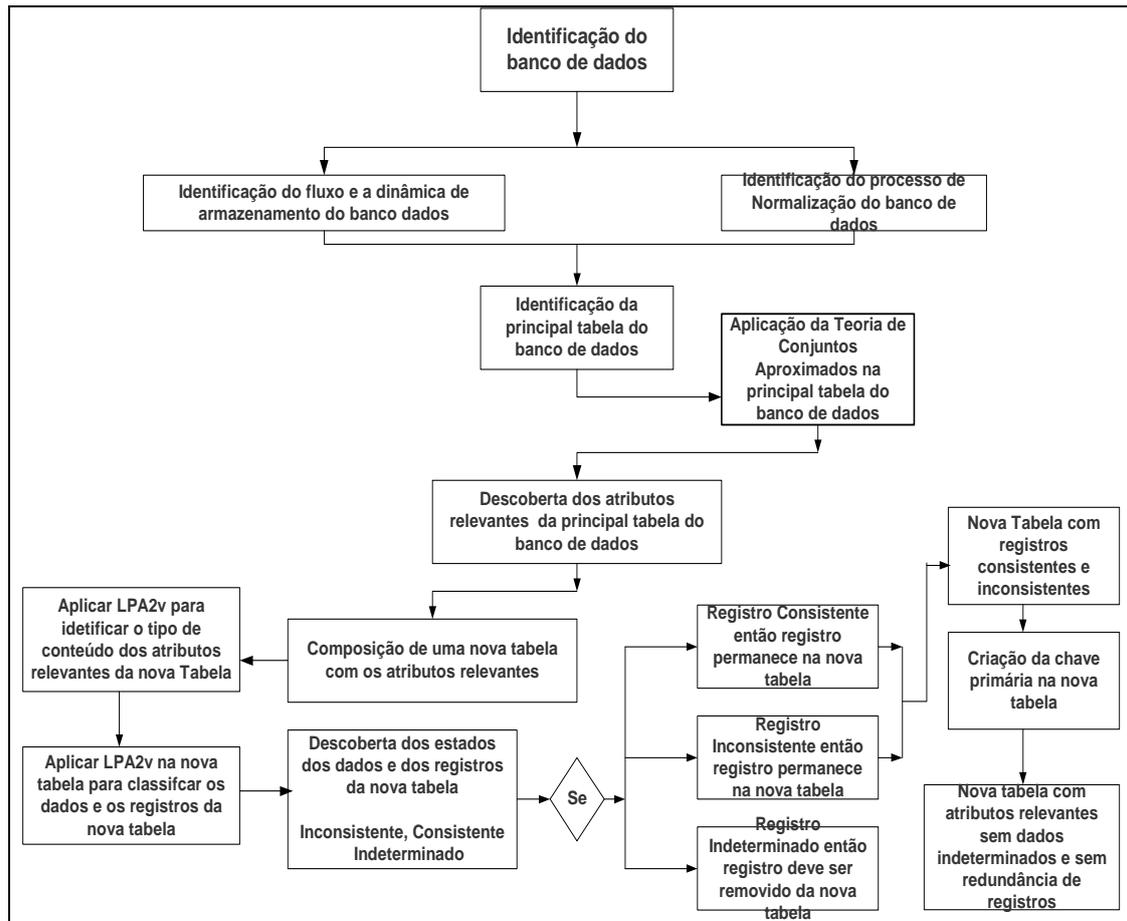


Figura 6. 4. Diagrama com o Resumo da Metodologia de Avaliação de Atributos Relevantes

6.8 Considerações Finais

A Teoria de Conjuntos Aproximados - TCA e a Lógica Paraconsistente Aplicada de Dois Valores – LPA2v, apresentadas neste capítulo com o objetivo de aplicá-las na base de dados do Sistema de Informações de Ocorrências Elétricas - Sistema INFO_OPR da empresa ELETRONORTE, especificamente no subconjunto de dados Sistema Elétrico Acre/Rondônia.

O objetivo da utilização da TCA e da LPA2v foi o desenvolvimento da Metodologia Híbrida de avaliação dos atributos relevantes na base do INFO_OPR. A metodologia apresenta os conceitos e aplicações da TCA e de LPA2v em banco de dados relacional.

A TCA é responsável pela descoberta dos atributos relevantes e a LPA2v pela identificação e classificação das incompletudes ocorridas nos dados dos referidos atributos.

A Metodologia Híbrida, apresentada, pode ser usada no processo de extração de conhecimento, mas precisamente no estágio de consolidação dos dados, onde ocorre a definição dos atributos relevantes, além de definir os problemas e a forma de tratamento para os dados que não possuem valores e, se necessário, remover as exceções.

Na metodologia proposta, os atributos relevantes são definidos pela aplicação da TCA, mas para tratar os dados incompletos, é necessário primeiramente identificá-los e classificá-los via LPA2v, pois somente após a execução desta tarefa se descobre o estado de cada registro que compõe o conjunto de dados.

O estado dos dados indicará se o mesmo precisa de tratamento ou deve ser removido da base de dados, situação esta que não é, atualmente, incorporada pelos estágios da descoberta de conhecimento em banco de dados.

Caso haja a incorporação da metodologia apresentada no processo de descoberta de conhecimento, os estados dos registros dos atributos indicarão, com maior segurança, qual procedimento tomar em relação à remoção dos registros ou tratamento dos dados inconsistentes, sem a necessidade da intervenção de um especialista.

7. Conclusões e Trabalhos Futuros

7.1 Visão geral

Este trabalho está direcionado na apresentação de conceitos de ferramentas matemáticas não clássicas aplicadas na investigação científica, tendo como objetivo propor uma metodologia híbrida de avaliação de atributos relevantes em grandes bases de dados de ocorrências elétricas que apresentam dados incompletos.

A primeira ferramenta matemática utilizada é a Teoria dos Conjuntos Aproximados, que se responsabilizou pela descoberta de um conjunto de atributos relevantes da base de dados do Sistema de Informações de Ocorrências Elétricas - INFO_OPR.

A segunda ferramenta matemática é a Lógica Paraconsistente que identificou e classificou os dados contidos nos atributos relevantes. A identificação e classificação dos valores dos dados proporcionam a descoberta dos estados dos dados e, conseqüentemente, dos registros da nova base de dados composta, apenas, pelos atributos relevantes.

7.2 TCA no BDR do Sistema de Ocorrências Elétricas – INFO_OPR

A teoria de conjuntos aproximados tem seu desenvolvimento independente e de forma diferente da teoria de banco de dados relacional, já que a idéia principal desta teoria está em classificar, encontrar padrões, regras e conhecimento embutido entre os dados. Já a teoria de banco de dados relacional está centrada no armazenamento e recuperação de dados, não se preocupando com o tipo de dado que está armazenando e nem com padrões escondidos entre os mesmos.

O banco de dados relacional está projetado para o desenvolvimento de estruturas sólidas lógicas e físicas, para armazenamento de dados e recuperação da informação [Codd, 1970].

As idéias das duas teorias apesar de serem diferentes têm um ponto de interseção, sendo que mesmo é o conceito de entidades (tabelas), já que as tabelas que são utilizadas nas duas teorias são similares, pois são compostas por atributos (colunas) e seus respectivos conteúdos (dados) formando os registros.

A similaridade conceitual, não é direta, pois os registros das tabelas, nas duas teorias, não têm correspondência, pois em TCA os registros podem ser redundantes, isto é, permite repetição, No banco de dados relacional a redundância deve ser evitada, já que a mesma viola uma das principais regras da teoria de banco de dados relacional, que é regra de integridade de dados.

O uso de TCA, neste trabalho, tem como objetivo a descoberta de um conjunto de atributos relevantes, dentre todos os atributos que compõem as tabelas do banco de dados INFO_OPR, além do que, este conjunto de atributos relevantes, deve representar todo o conhecimento embutido na base de dados.

A forma tradicional de aplicação da TCA está na descoberta de atributos relevantes de uma tabela, neste caso, a aplicação deve ser em todas as tabelas da base INFO_OPR, com o objetivo de descobrir quais os atributos relevantes de cada tabela e depois reunir esses conjunto de atributos e conseguir um conjunto, sem repetição de elementos, os quais representariam os atributos relevantes.

Como o banco de dados do INFO_OPR possui uma quantidade de aproximadamente quarenta tabelas, o processo da descoberta de atributos relevantes ficaria repetitiva e demorada. Uma solução encontrada para diminuir o tempo para a descoberta dos atributos relevantes é a utilização do processo de normalização do banco de dados, caso exista, pois através dele se descobre a principal tabela do banco de dados.

A segunda alternativa para a descoberta da principal tabela do banco de dados é através da análise do fluxo e da dinâmica de armazenamento do banco de dados. Neste caso, o armazenamento dos dados indicou a Tabela Ocorrência como a principal tabela do banco de dados, já que a mesma possui ligações com todas as outras tabelas do banco de dados e com os programas de interface de dados.

A principal tabela do banco de dados armazena todos os dados importantes, relativos às ocorrências do Sistema Acre/Rondônia, desta forma a TCA é aplicada na referida tabela, para a descoberta de um conjunto de atributos relevantes que garanta o conhecimento embutido na base de dados.

Ao final da aplicação da TCA, encontra-se um conjunto de atributos relevantes apresentados na Tabela 5.3, no Capítulo 5. O conjunto de atributos relevantes possui dados que podem apresentar incompletudes, logo, é necessária a identificação e classificação destas incompletudes, já que para uma futura descoberta e extração de conhecimento da base de dados composta pelos atributos relevantes, os dados contidos devem estar consistentes.

Com a utilização do processo de normalização do banco de dados ou da identificação do fluxo e da dinâmica de armazenamento, houve redução no tempo de descoberta dos atributos relevantes, pois foi necessária apenas a descoberta dos mesmos na tabela principal do banco de dados. Esta etapa pode ser incorporada ao processo de descoberta de conhecimento, no estágio de consolidação dos dados, descrito no Capítulo 2 deste trabalho.

7.3 Lógica Paraconsistente no BDR do Sistema de Ocorrências Elétricas – INFO_OPR

A lógica paraconsistente, neste trabalho, é utilizada para a identificação e classificação dos dados com incompletudes, oriundos de banco de dados que contenham algum grau de incerteza, inconsistência e contradição [Imieliński & Lipski, 1984].

Para identificar e classificar os problemas, relacionados à natureza do dado, é necessário o uso de uma teoria com recursos e que suporte o entendimento de contradições, inconsistências e incompletude. Se a teoria escolhida for a lógica clássica, esta irá se defrontar com a ocorrência da contradição nos dados, neste caso o sistema se trivializa acarretando a inviabilidade do uso desta teoria. Mas, se a teoria escolhida for à lógica paraconsistente, o tratamento se dará de forma diferente, já que o fato de haver contradições entre os dados pode possibilitar a descoberta de conhecimento implícito entre os mesmos.

Com a descoberta dos atributos relevantes da base de dados INFO_OPR, os dados contidos nestes atributos apresentam problemas de incompletudes, as quais precisam ser identificadas, classificadas e, posteriormente, tratadas.

Para a identificação das incompletudes dos dados é necessário analisar que tipos de incompletudes ocorrem entre os mesmos. A análise do conjunto do respectivo conjunto de dados encontrou os seguintes tipos de incompletudes: valores ausentes ou indeterminados e valores inconsistentes ou fora do padrão esperado.

Com a identificação do tipo de dado, o resultado é aplicado segundo os princípios da LPA2v, isto é, levando em consideração os valores dentro da área do triângulo inferior da figura 6.3, exibida no capítulo 6.

A classificação individual de cada dado possibilita a classificação do registro, já que os dados classificados fazem parte do conteúdo do conjunto de atributos relevantes. Com a classificação dos registros é possível indicar qual a melhor ação a ser tomada diante dos problemas de incompletudes.

Os registros, que apresentarem estado consistente, permanecem no banco de dados sem necessidade de tratamento, já os que apresentam problemas de inconsistência devem permanecer no banco de dados, mas com a necessidade de tratamento dos dados problemáticos. Os registros que apresentarem classificação de indeterminados podem ser excluídos.

É necessária a identificação e classificação dos dados e dos registros do banco de dados, pois indica com segurança quais os registros, ou mesmo os dados que devem ser tratados, quais devem ser removidos e quais estão aptos à descoberta e extração de conhecimento. Logo, este procedimento pode ser incorporado no estágio de consolidados dos dados no processo de descoberta e extração de conhecimento de bases de dados.

7.4 Aplicação da Metodologia de Avaliação da Relevância de Atributos na Base de Dados do Sistema INFO_OPR

A idéia central, do estudo das ferramentas matemáticas não clássicas, Teoria dos TCA e Lógica Paraconsistente foi o desenvolvimento da metodologia híbrida para avaliação dos atributos relevantes de um grande banco de dados com séries históricas de ocorrências elétricas. A identificação e classificação das incompletudes dos dados contidos nos atributos relevantes é uma necessidade para habilitar ou não o banco de dados a descoberta e extração de conhecimento.

A base de dados do sistema de Informações de Ocorrências Elétricas – INFO_OPR apresentou-se como um ambiente com todas as características para a aplicação da metodologia, já que possui todos os tipos de problemas como: grande quantidade de atributos e dados para análise; todos os tipos de incompletudes entre os dados e registros.

A metodologia proposta visa facilitar a extração de conhecimento íntegro, relevante e coerente em grandes bases de dados, neste caso em especial na base de dados do INFO_OPR, tendo como direcionamento principal a tomada de decisão.

Com a aplicação da metodologia pode-se construir uma tabela com todos os atributos relevantes da base de dados do INFO_OPR, onde as incompletudes dos conteúdos dos atributos foram identificadas, além de possibilitar a classificação dos

registros em relação ao seu estado de consistência, remover aqueles que não podem ser utilizados e indicar os possíveis problemas e futuras soluções.

7.5. Principal Contribuição

A contribuição inovadora deste trabalho é a apresentação de uma Metodologia Híbrida que possibilite a descoberta dos atributos relevantes em grandes bases com séries históricas de dados de ocorrências elétricas, na qual os atributos podem ser heterogêneos ou não.

A metodologia oferece condições de analisar as tabelas que compõem a base de dados, identificando quais são importantes.

O resgate da teoria de banco de dados relacional, já que a maioria dos problemas apresentados na base INFO_OPR, é de natureza estrutural, relativas à forma de aplicação do modelo de banco de dados relacional.

A descoberta dos atributos relevantes, no banco de dados do INFO_OPR, foi realizada através de uma ferramenta matemática não tradicional denominada de Teoria de Conjuntos Aproximados, onde não há necessidade dos especialistas da área para a descoberta dos atributos relevantes da base de dados.

A TCA apresentou-se apta para descobrir quais atributos, dentre todos os pertencentes ao banco de dados, representariam de forma íntegra o conhecimento embutido na base de dados.

A identificação e classificação das incompletudes ocorridas entre os dados dos atributos relevantes foram realizadas através da Lógica não Clássica Paraconsistente, na sua forma anotada de dois valores – LPA2v. Esta ferramenta apresentou excelentes resultados com relação à identificação e classificação das incompletudes, proporcionando também condições de indicar quais registros devem ser removidos,

quais podem ser tratados e quais estão aptos a se transformar em informação e participar da descoberta de conhecimento objetivando a tomada de decisão.

7.6 Considerações Finais

O objetivo geral, deste trabalho foi apresentar uma Metodologia Híbrida de avaliação de atributos relevantes, em grandes bases de dados incompletas, onde as incompletudes são identificadas e classificadas, objetivando disponibilizar os estados dos dados e dos registros para uma descoberta de conhecimento confiável.

A apresentação da Metodologia Híbrida foi atingida, já que durante o seu desenvolvimento, no capítulo 6, foi possível utilizar o processo de normalização de banco de dados relacional e a análise do fluxo e dinâmica de armazenamento de dados, tendo como objetivo reduzir a tarefa da descoberta de atributos relevantes na base de dados com a Teoria de Conjuntos Aproximados.

A identificação e classificação dos estados dos valores contidos nos atributos, utilizando Lógica Paraconsistente na sua forma Anotada de Dois Valores – LPA2v teve como consequência a classificação também do estado do registro.

A classificação dos dados e dos registros é uma das contribuições desta tese, pois a mesma facilitará o trabalho dos profissionais da área de banco de conhecimento, já que a descoberta do estado real do banco de dados poderá antever os possíveis problemas na extração de conhecimento.

Outra contribuição deste trabalho é a possibilidade de incorporar a metodologia apresentada, no estágio de consolidação dos dados durante o processo de descoberta de conhecimento.

7.6 Linhas de Pesquisa para Desenvolvimento de Trabalhos Futuros

As linhas de pesquisa propostas, para o desenvolvimento de futuros trabalhos, baseadas nos resultados obtidos desta tese são:

1. Desenvolvimento de uma ferramenta computacional, na qual seja implementado o processo de descoberta de conhecimento, utilizando a metodologia proposta, no estágio de consolidação;
2. Estudo com a Lógica Paraconsistente Anotada de Dois Valores – LPA2v, para o tratamento dos dados inconsistentes, utilizando uma abordagem híbrida com a estatística experimental para o tratamento de incompletudes dos dados;
3. Estudo voltado para o tratamento das inconsistências dos dados, utilizando Lógica Paraconsistente Anotada de Valores Superiores;
4. Desenvolvimento de uma ferramenta automatizada para identificação de atributos relevantes de um banco de dados, identificação, classificação e tratamento das incompletudes existentes entre os dados. Esta ferramenta automatizada pode reduzir o tempo de extração de conhecimento de banco dados, independente da qualidade dos dados e da quantidade de variáveis (atributos) a serem analisadas.

Referências

[**Abe, 1982**] Abe, Jair Minoro. Fundamentos da Lógica anotada. Tese de doutorado. Universidade de São Paulo. São Paulo: USP. 1982.

[**Abe et al., 1991**] Abe, Jair Minoro; Da Costa, N.C.A; Subrahmanian, V.S. Remarks on Annotated Logics, Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik. Vol. 37. p: 561--570. 1991.

[**Akama & Abe, 1998**] Akama, S., Abe, J. M. Constructive logics as annotated logics. São Paulo: IEA-USP, 1998.

[**Arruda,1990**] Arruda, Ayda Ignez. N. A. Vasiliev e a lógica paraconsistente. UNICAMP, Centro de Lógica, Epistemologia e História da Ciência - CLE. Coleção CLE. 1990.

[**Atzeni et al., 2000**] Atzeni, Paolo; Ceri, Stefano; Paraboschi, Stefano. Database Systems: concepts, languages & architectures. McGraw Hill Publishing Company. 2000. p. 349-390.

[**Bagai & Sunderraman, 1995**] Bagai, Rajiv; Sunderraman, Rajshekhar. A Paraconsistent Relational Data Model, International Journal of Computer Mathematics. Vol 55. 1995.

[**Bagai & Orgum,1995**] Bagai, Rajiv; Orgun, Mehmet A.. A temporal paraconsistent relational algebra for incomplete and inconsistent information. ACM-SE 33: Proceedings of the 33rd annual on Southeast regional conference. Clemson, South Carolina. SESSION: Database systems II. Pages: 240 – 248. Publisher: ACM . March 1995. ISBN:0-89791747-2.

[**Brusoni et al., 1999**] Brusoni,Vittorio; Console, Luca; Terenziani, Paolo; Percini, Bárbara. Qualitative and Quantitative Temporal Constraints and Relational Databases: Theory, Architecture, and Applications. IEEE Transaction on knowledge and Data Engineering. November/December 1999 (vol. 11 no. 6) pp. 948-894. 1999.

[**Carneiro, 2006**] Carneiro, Maria Francisca. Alguns Aspectos da Lógica e da Informática Jurídica. Revista Jurídica CESUMAR. Vol. 6 No. 1. pg. 37-45. 2006.

[**Codd, 1970**] Codd, E. F. A Relational model for large shared data banks. Communication of the ACM, 13(6):377-387.june 1970.

[**Codd, 1986**] Codd, E.F. Missing information (applicable and inapplicable) in relational databases. SIGMOD Record. Vol:15. No.4. December 1986.

[**Cornelli & Miranda, 2007**]Cornelli, Gabriele; Coelho, Maria Cecília de Miranda. “QUEM NÃO É GEÔMETRA NÃO ENTRE!”.. GEOMETRIA, FILOSOFIA E PLATONISMO KRITERION, Belo Horizonte, nº 116, Dez/2007, p. 417-435. 2007.

[**Da Costa et al., 1991**] Da Costa, Newton; Subrahmanian, S. V.; Vago, C.. The Paraconsistent Logics PT. Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik, Vol. 37, pps 139-148, 1991.

[**Da Costa, 1993**] Da Costa, Newton Carneiro Afonso. Sistemas formais inconsistentes. Editora da UFPR. ISBN 85-85132-75-2. Curitiba,Paraná, Brasil. 1993.

[**Da Costa et al., 1999**] Da Costa, Newton Carneiro Afonso.; Abe, Jair Minoro; Murolo, Afrânio Carlos; Da Silva Filho, João Inácio; Leite, Casemiro Fernando S..Lógica Paraconsistente Aplicada. São Paulo: Atlas, 1999. ISBN 8524422184.

[**Da Costa & Abe, 2000**] Da Costa, Newton Carneiro Afonso; Abe, Jair Minoro. Paraconsistência em informática e inteligência artificial. Estudos Avançados da Universidade de São Paulo. Vol 14. No. 39. pág.: 161-174. 2000.

[**Da Costa & Krause, 2004a**] Da Costa, Newton Carneiro Afonso; Krause, Décio. NOTAS DE LÓGICA. Parte I: Lógicas Proposicionais Clássica e Paraconsistente. CAPÍTULO 5 (Texto Preliminar) [Http://www.cfh.ufsc.br/~dkrause/LivroLogica/Capitulo5.pdf](http://www.cfh.ufsc.br/~dkrause/LivroLogica/Capitulo5.pdf). Florianópolis. 2004.

[**Da Costa & Krasue, 2004b**] Da Costa, Newton Carneiro Afonso; D. Krause; O. Bueno. ‘Paraconsistent Logics and Paraconsistency: Technical and Philosophical

Developments', pré-publicação do CLE-Centro de Lógica, Epistemologia e História da Ciência da UNICAMP. 2004. Disponível em http://www.cle.unicamp.br/e-prints/vol_4,n_4,2004.html.

[Da Silva Filho, 1999] Da Silva Filho, João Inácio. Métodos de Aplicações da Lógica Paraconsistente Anotada de anotação com dois valores-LPA2v com construção de Algoritmo e Implementação de Circuitos Eletrônicos. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais. Orientador: Prof. Dr. Jair Minoro Abe. São Paulo, 1999.

[Da Silva Filho & Rocco, 2006] Da Silva Filho, João Inácio; Rocco, A.; Mário, M.C.. Annotated Paraconsistent Logic Applied to na Expert System Dedicated for supoortin in na Eletric Power Transmission Systems Restablishment. PSCE 2006.

[Da Silva Filho et al., 2008] Da Silva Filho, João Inácio da; Abe, Jair Minoro; Lambert-Torres, Germano. Inteligência artificial com as redes de análises paraconsistentes: teoria e aplicações. Rio de janeiro: LTC. 2008.

[Date, 2004] Date, C. J.. Introdução a Sistemas de BANCOS de DADOS, Tradução da 8ª. Edição Americana. 4ª Tiragem Petropolis-RJ: Campus, 2004, 104 p. 621-658.

[Domenech et al., 2008] Domenech, Graciela; Holik, Federico, Krause, Décio. Q-spaces and the Foundations of Quantum Mechanics. Found Phys (2008) 38: 969–994. DOI 10.1007/s10701-008-9246-9. Received: 26 March 2008 / Accepted: 1 October 2008 / Published online: 15 October 2008. Springer Science+Business Media, LLC 2008.

[Edelweiss, 1998] Edelweiss, Nina. Banco de Dados Temporais: Teoria e Prática. Universidade Federal do Rio Grande do Sul. Instituto de Informática. 1998. Disponível em <http://www.inf.ufrgs.br/~nina/ArtigosNina/JAI1998.pdf>. Acessado em 02/02/2007.

[ELETRONORTE, 2009] ELETRONORTE. Gereção de Energia por Estado. Disponível em [HTTP ://www.eln.gov.br/opencms/opencms/pilares/geracao/estados/rondonia/](http://www.eln.gov.br/opencms/opencms/pilares/geracao/estados/rondonia/). 2009.

[**Farhangfar & Kurgan, 2007**] Farhangfar, Alireza; Kurgan, Lukasz A. (Member, IEEE) Pedrycz, Witold Pedrycz (Fellow, IEEE). A Novel Framework for Imputation of Missing Values in Databases. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 37, no. 5, september 2007.

[**Fayyad et al., 1996a**] Fayyad, U., G. Piatetsky-Shapiro, Smyth, P. From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery & Data Mining*, pp. 1–34. 1996.

[**Fayyad et al., 1996b**] Fayyad, U., G. Piatetsky-Shapiro, & P. Smyth. Knowledge discovery and data mining: Towards a unifying framework. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 82–88. 1996b.

[**Ganesan et al., 2007**] Ganesan, G.; Latha, D.; Rao, C. Raghavendra. Reduct Generation in Information Systems. *Engineering Letters*, 14:2, EL_14_2_5. Advance online publication: 16 May 2007.

[**Grzymala et al, 1990**] Grzymala-Busse, Jerzy W.; Mithal, Sachin. A Comparison of Four Tests for Attribute Dependency in the LEM and LERS Systems for Learning from Examples. *International conference on Industrial and engineering applications of artificial intelligence and expert systems Proceedings of the 3rd international conference on Industrial and engineering applications of artificial intelligence and expert systems - Volume 2*. Pages: 949 – 958. ACM 089791-372-8. 8/90/0007/0949. Charleston, South Carolina, United States. 1990.

[**Hassan & Tazaki, 2005**] Hassan, Yasser; Tazaki, Eiichiro. Emergent rough set data analysis. *Journal: Kybernetes*. Year: 2005. Volume: 34. Issue: 6. Page: 869 – 887. ISSN: 0368-492X. DOI: 10.1108/03684920510595544. Publisher: Emerald Group Publishing Limited. 2005.

[**Hongai et al., 2005**] Hongai, Feng; Bayoan, Liu; Liyun, He. Using Rough Set to Induce Dependencies between Attributes Where There Are a Large Amount of Missing

Values — A SARS Data Application. K-CAP'05, October 2–5, 2005, Banff, Alberta, Canada. ACM 1-59593-163-5/05/0010. 2005.

[Imieliński & Lipski, 1984] Imieliński, Tomasz; Lipski, Jr., Witold. Incomplete Information in Relational Databases. *Journal of the ACM (JACM)* .Volume 31, Issue 4 (October 1984) Pages: 761 – 791. ISSN:0004-5411. 1984.

[Jensen et al., 1998] Jensen, Christian S.; Dyreson, Curtis; Böhlen, Michael; Clifford, James; Elmaris, Ramez; Gadia, Shashi K; Grandi, Fabio; Hayes, Pat; Jajodia, Sushil; Käfer Wolfgang; Kline, Nick; Lorentzos, Mitsopoulos, Yannis; Montanari, Angelo Montanari; Nonen, Daniel; Peresi, Elisa; Barbara, Pernici; Roddick , John F.; Sarda, Nandlal L.; Scalas, Maria Rita; Segev, Arie; Snodgrass, Richard T.; Soo, Mike D.; Tansel, Abdullah; Tiberio, Paolo; Wiederhold, Gio. The Consensus Glossary of Temporal Database Concepts - February 1998 Version. Proceedings of the 1998 ACM symposium on Applied Computing, p.235-240, February 27-March 01, 1998.

[Jensen, 2000] Jensen, Christian S.. Temporal Database Management. Disponível em <<http://www.cs.auc.dk/~csj/Thesis/>>. 2000.

[Jensen & Snodgrass, 2009] Jensen, Christian; Snodgrass, Richard T. (editors). Temporal Database Entries for the Springer Encyclopedia of Database Systems. Springer. 2009.

[Korth & Silberschatz, 1995] Korth, Henry F; Silberschatz, Abraham. Sistema de Banco de Dados. Editora Markon Brooks. 2ª Edição. p.17-22. 1995.

[Lambert-Torres et al, 1999] Lambert-Torres, G.; Rossi, R.; Jardini, J.A.; Alves da Silva, A.P. & Quintana, V.H. Power System Security Analysis based on Rough Classification, In: Rough Fuzzy Hybridization: A New Trend in Decision-Making, Pal, S. & Skowron, A. (Ed.), pp. 263-300, Springer-Verlag Co., ISBN 981-4021-00-8, Secaucus-USA. 1999.

[**Lin, 1996**] Lin, T. Y.. An Overview of Rough Set Theory from the Point of View of Relational Databases. Disponível em http://www.cs.sjsu.edu/~tylin/publications/paperList/96_28rsdb2.pdf. 1996.

[**Martins et al., 2007**] Martins, Helga Gonzaga; Lambert-Torres, Germano; Potin, Luiz Francisco. A Lógica Paraconsistente Anotada. Santos-SP. Comunicar Editora. ISBN 978-85-99561-31-7. 2007.

[**Mortari, 2001**] Mortari, Cezar A.. Introdução à Lógica. São Paulo: Editora Unesp. 2001. p.60-65. ISBN 85-7139-337-0. 2001.

[**Navathe & Elmaris, 2006**] Navathe, Shamkant B.; Elmaris, Ramez. Sistemas de Banco de Dados. Revisor Técnico: Luiz Ricardo de Figueiredo. – São Paulo: Pearson Addison Wesley, 4^a Edição. 2006. p. 552-561.

[**Nunes da Silva Filho, 2006**] Nunes da Silva Filho, Milton. Sistema integrado para tomada rápida de decisões nos sistemas elétricos. Tese de Doutorado do Programa de Engenharia Elétrica – Sistemas de Potência. Universidade Federal de Itajubá. Orientador Prof. Dr. Germano Lambert Torres. Co-orientador Prof. Dr. Luiz Eduardo Borges da Silva. Dezembro de 2006.

[**Nunes da Silva Filho, 2009**] Nunes da Silva Filho, Milton. Gestão de PV, Pós-Operação. ELETRONORTE. Brasília. 2009.

[**Passos & Goldschmidt, 2005**] Passos, Emanuel; Goldschmidt, Ronaldo. Data Mining - Guia Prático. Rio de Janeiro. Editora Elsevier. 2005.

[**Pawlak, 1982**] Pawlak, Z.. Rough sets, International Journal of Computer and Information Sciences, 11, 341-356, 1982.

[**Pawlak, 1991**] Pawlak, Zdzislaw. Rough Set: Theoretical aspects of reasoning about data. Series D: System Theory, knowledge Engineering and problem Solvin. Volume 9. Kluwer Academic Publishers. Dordrecht.Boston.London. 1991.

[Pawlak & Skowron, 1994] Pawlak, Z; A. Skowron: Rough membership function, in: R. E Yeager, M. Fedrizzi and J. Kacprzyk (eds.), *Advances in the Dempster-Schafer of Evidence*, Wiley, New York, pages 251-271. 1994.

[Pawlak, 1998] Pawlak, Zdzislaw. Granularity of Knowledge, Indiscernibility and Rough Sets. *Fuzzy Systems Proceedings*, 1998. IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference on. Publication Date: 4-9 May 1998. Volume: 1, page(s): 106-110 vol.1. Anchorage, AK, USA. 1998.

[Priest, 2008] Priest, Graham. In *Introduction to Non-Classical Logic*. Cambridge University Express. New York. Second edition. 2008.

[Raspa, 1999] Raspa, Venanzio. ŁUKASIEWICZ on the principle of contradiction. *Journal of Philosophical Research*, XXIV, 1999, pp. 57-112. UNIVERSITÀ DI URBINO. Disponível em <http://segr-did2.fmag.unict.it/~polphil/PolPhil/Lukas/RaspaLuk.PDF>. 1999.

[Rissino & Lambert-Torres, 2009] Rissino, Silvia; Lambert-Torres, Germano. Chapter 3. Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications. In *Data Mining and Knowledge Discovery in Real Life Applications*. Editors Julio Ponce and Adem Karahoca. ISBN 978-3-902613-53-0. 2009.

[Snodgrass & Ahn, 1985] Snodgrass, Richard; Ahn, Ilsoo. A Taxonomy of Time in Databases. *ACM SIGMOD Record Proceedings of the 1985 ACM SIGMOD international conference on Management of data SIGMOD '85*, Volume 14 Issue 4. May 1985.

[Stoll, 1979] Stoll, R.R.. *Set Theory and Logic*, Dover Publications, ISBN 0-486-63829-4, Mineola-USA. 1979.

[Subrahmanian, 1987] Subrahmanian, V.S. On the semantics of quantitative logic programs. *Proc. 4^o. IEEE symposium on logic programming computer society press*, Washington D.C.. p:173-182. 1987.

[Vianna, 2007] Vianna, Elaine Aparecida de Lima. *A Interligação do Sistema Elétrico de Rondônia ao Sistema Interligado Nacional – SIN*. Monografia (Especialização em

Engenharia da Energia) – Universidade Federal de Itajubá – UNIFEI. Orientador Prof. Dr. Germano Lambert Torres. 2007.

[Vilela, 1964] Vilela, Orlando. Iniciação a Lógica. Dominus Editora S.A. São Paulo. p.109. 1964.

[Watson, 2004] Watson, Richard T.. Data Management: Banco de Dados e Organizações. 3a. Edição. Editora LTC. Rio de Janeiro. p. 100-115. 2004.

[Wu et al., 2004] Wu, Chengdong; Yue, yong; Li, Mengxin; Adjei, Osei. The Rough Set theory and applications. Engineering Computations. Vol. 21. No. 5, 2004. pp.488-511. 2004.

[Yao & Zhao, 2009] Yao, Yiyu; Zhao, Yan. Discernibility matrix simplification for constructing attribute reducts. Information Sciences 179 (2009) 867–882. Journal homepage: www.elsevier.com/locate/ins. 2009.

[Zadeh, 1965] Zadeh, L.. Fuzzy sets, Information and Control, 8, 338-353, 1965.