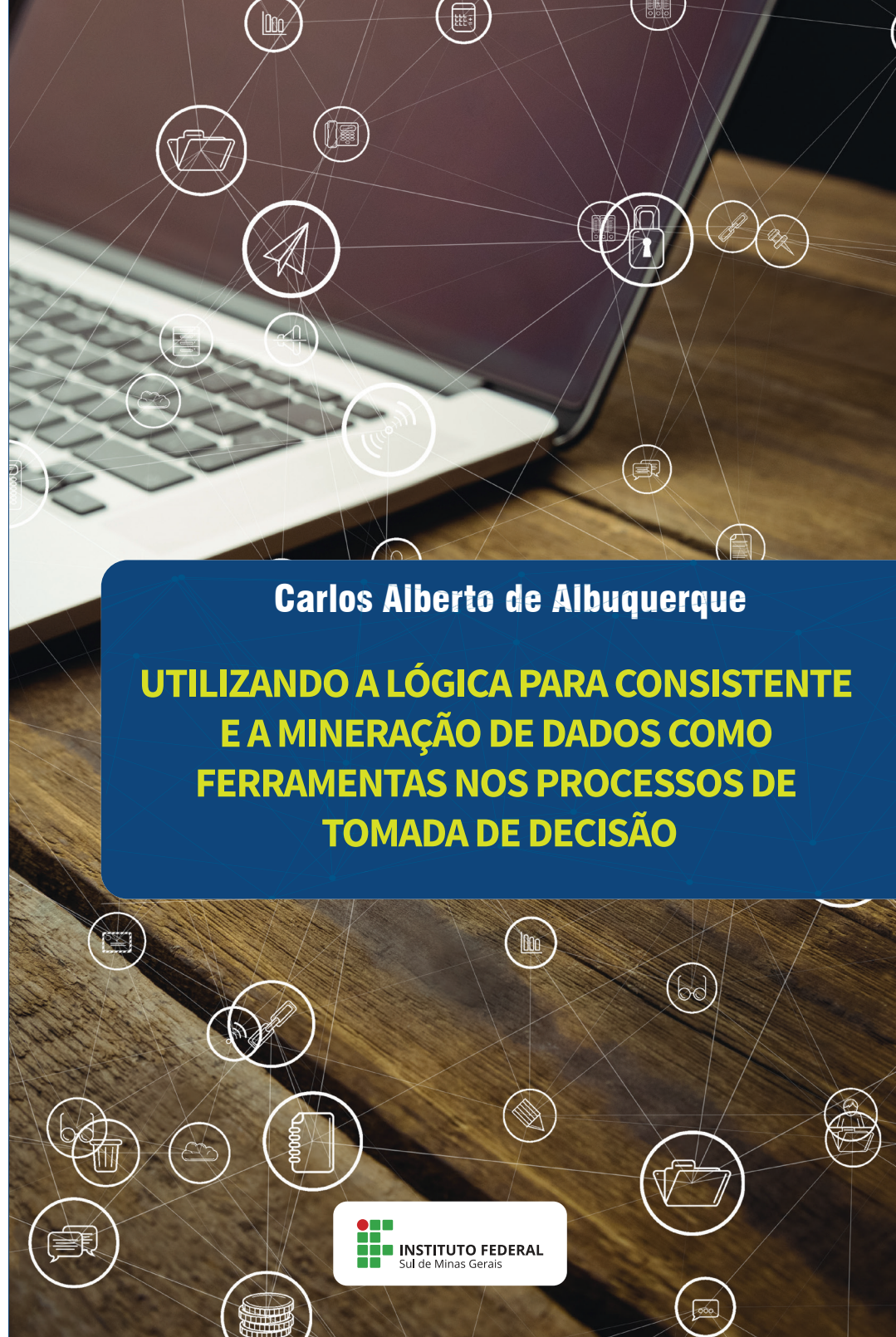

Carlos Alberto de Albuquerque

UTILIZANDO A LÓGICA PARA CONSISTENTE E A MINERAÇÃO DE DADOS COMO FERRAMENTAS NOS PROCESSOS DE TOMADA DE DECISÃO

INSTITUTO FEDERAL
Sul de Minas Gerais

Carlos Alberto de Albuquerque

UTILIZANDO A LÓGICA PARA CONSISTENTE E A MINERAÇÃO DE DADOS COMO FERRAMENTAS NOS PROCESSOS DE TOMADA DE DECISÃO

INSTITUTO FEDERAL
Sul de Minas Gerais

2017

© 2015 – Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais (IFSULDEMINAS)
UTILIZANDO A LÓGICA PARACONSISTENTE E A MINERAÇÃO DE DADOS COMO FERRAMENTAS NOS PROCESSOS DE TOMADA DE DECISÃO
Todos os direitos reservados. Nenhuma parte deste livro poderá ser reproduzida, sejam quais forem os meios empregados, sem a permissão, por escrito, do **IFSULDEMINAS**. Aos infratores aplicam-se as sanções previstas nos artigos 102, 104, 106 e 107 da Lei no 9.610, de 19 de fevereiro de 1998.

Revisão: Daniela Risson – Jornalista – ME

Diagramação: M&W Comunicação Integrada LTDA – ME

Impressão: Tavares&Tavares Empreendimentos Comerciais LTDA – ME

Editora: IFSULDEMINAS

Endereço: Avenida Vicente Simões, 1111 – Nova Pouso Alegre – Pouso Alegre/MG - 37550-000

Site: http://www.ifsuldeminas.edu.br/

Email: proex@ifsuldeminas.edu.br

Albuquerque, Carlos Alberto.
Utilizando a lógica paraconsistente e a mineração de dados como ferramentas nos processos de tomada de decisão / Carlos Alberto Albuquerque. – Pouso Alegre : IFSULDEMINAS, 2015.
78 p. : il.; graf.

ISBN: 978-85-67952-11-6

1. Lógica. 2. Lógica paraconsistente. 3. Sistema especialista. 4. Recuperação de dados. 5. Classificação de dados. 6. Inteligência artificial. I. Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais. II. Título.

SUMÁRIO

APRESENTAÇÃO **7**

INTRODUÇÃO **9**

2 CLASSIFICANDO UMA INCOMPLETUDE OU UMA INCONSISTÊNCIA **12**

2.1 – A lógica Paraconsistente como ferramenta para tratamento de dados 15

3 MINERAÇÃO DE DADOS **16**

3.1 – Definição 16

3.2 – Aplicações 17

3.3 – Funcionalidade 18

4 O MODELO RELACIONAL PARA BANCO DE DADOS **20**

4.1 – Definição 20

4.2 – Conceitos do MR 23

4.2.1 – Conjunto de Entidades 23

4.2.2 – Conjunto de Relacionamentos 24

4.2.3 – Atributos de Entidades de Relacionamentos 25

4.2.4 – Chaves 27

4.2.5 – Papéis no Relacionamento 28

4.2.6 – Cardinalidade dos Relacionamentos 30

5 A LÓGICA PARACONSISTENTE ANOTADA DE DOIS VALORES – LAP2v **33**

5.1 – Conceitos Básicos de LPA2v 33

5.2 – Interpretando o QUPC 37

5.3 – Grau de Incerteza (GI) e Grau de Certeza (GC) 41

5.3.1 – Vantagens da Obtenção do Grau de Incerteza e de Certeza 50

5.4 – Os Pontos Notáveis do QUPC e o Significados das regiões formadas a partir 52

5.4.1 – Região de Verdade 55

5.4.2 – Região Inconsistente 58

5.4.3 – Região de Falsidade 59

5.4.4 – Região Paracompleta 62

5.4.5 – QUPC Contendo as Regiões de Verdade, Falsidade, Inconsistência e Paracompleta 64

6 A UTILIZAÇÃO DA LP NO TRATAMENTO DE DADOS **65**

6.1 – Sistema Especialista de Contexto (SEC) 66

6.2 – Aplicação do SEC 66

REFERÊNCIAS **74**

APRESENTAÇÃO

A tecnologia da informação (TI) está presente em todas as atividades de nossa sociedade. Atualmente é possível realizar qualquer atividade, como comprar, estudar e trabalhar, através da TI. Quando realizamos alguma atividade utilizando a TI, muitas informações são armazenadas em repositórios eletrônicos, que, se logicamente organizados, podem se transformar em ferramentas úteis nos processos de decisão.

Essa obra procura mostrar a importância de utilizar ferramentas que organizem, logicamente, as várias informações disponibilizadas em meios eletrônicos. O objetivo é, após aplicar um tratamento, utilizar essas informações em processos de tomadas de decisão. Para isso, esse livro, mostra que a Lógica Paraconsistente (LP) mostra-se adequada para desenvolver tais ferramentas.

O livro introduz conceitos básicos, como Mineração de Dados (MD) e Banco de Dados (BD) e aprofunda nos conceitos da lógica não clássica LP, destacando a Lógica Paraconsistente Anotada de dois valores (LPA2v). Por último, é apresentado um exemplo didático de como utilizar a LPA2v para completar registros que contém incompletudes.

1 – Introdução

Há vários conceitos e arquiteturas em relação a uma base de dados, que em um passado não muito remoto, eram desconhecidos pelos responsáveis e usuários de bancos de dados das empresas e instituições privadas. Um exemplo é o fato de que os primeiros Sistemas Gerenciadores de Bancos de Dados (SGBDs) comerciais só começaram a ser implantados após os anos 1965, e mesmo assim, com mecanismos de acessos poucos eficientes (Boscarioli. Clodis, et al., 2006). Conseqüentemente, esses bancos de dados não tinham a consistência que os atuais bancos possuem. Conceitos como chave primária e chave estrangeira, que hoje são fundamentais no desenvolvimento de um banco de dados, eram desconhecidos nessa época (Boscarioli. Clodis, et al., 2006).

Além da falta de conhecimento de conceitos importantes como os citados acima, também há a questão da falta de padronização na entrada de dados (Editora: IFSULDEMINAS). A inserção da grande maioria dos dados nessas bases citadas era feita através da digitação de vários usuários. Outra dificuldade é que a grande maioria desses dados foi armazenada em formato texto (Figura 1.1). Dessa forma um usuário poderia cadastrar, por exemplo, uma chave fusível de 20 amperes, como "Chav. fus. 20 A" enquanto outro como "Chave F. 20". Como foi inserido, o banco de dados vai interpretar os dois registros de uma mesma chave, como um registro de duas chaves diferentes.

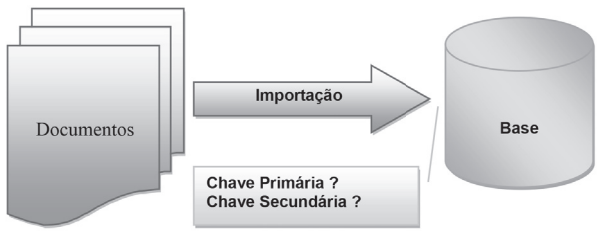

Figura 1.1: Armazenamento de dados em formato texto sem considerar conceitos importantes como "chave primária" e "chave estrangeira".

As instituições que tem sua base de dados "organizada" como descrita nos parágrafos anteriores possuem uma fonte rica em informações. Porém essa fonte precisa ser logicamente organizada, pois dessa forma é possível, através da análise dos dados disponíveis em tais bases de dados, extraírem importantes informações que poderão ser utilizadas em processos de tomadas de decisões (Lambert et all, 1999).

Fica clara, então, a necessidade de utilizar uma ferramenta que seja capaz de transformar uma base de dados, que embora seja rica em informações, está desorganizada e com dados inconsistentes, em uma base organizada logicamente e com dados consistentes (Figura 1.2). Assim será possível extrair valiosas informações de tais bases de dados que poderão ser ferramentas importantes na tomada de decisões.

Figura 1.2: Ferramenta para identificar chaves primárias e estrangeiras em uma base de dados.

Há também um crescente aumento na documentação eletrônica e conseqüentemente um aumento no volume de dados disponíveis em vários repositórios (Souza, Vendrusculo e Melo, 2000). Por isso, também é importante utilizar uma ferramenta que possibilite a extração de conhecimento das várias bases de dados disponíveis, através da construção de uma nova base unificada, onde os dados contidos na nova base sejam consistentes. Um exemplo são os vários sistemas de informações na área da saúde existentes no Brasil (Moraes, 2001). Há o sistema municipal, o estadual e o nacional, e algumas vezes, dentro de uma mesma esfera administrativa, há mais de um sistema de informação em saúde. Assim, as informações referentes a um mesmo

individuo, são fragmentadas em vários bancos de dados ou ainda há redundância de informação do mesmo indivíduo em vários bancos de dados, gerando assim distorções.

Como esses sistemas são uma fonte preciosa de informações para auxiliar as esferas governamentais a planejarem as ações e também o acompanhamento dos programas em desenvolvimento na área da saúde (Cherchiglia e outros, 2007), a criação de uma única base de dados, contendo todos os dados de todos os sistemas de informação referentes à saúde, mostra ser uma ferramenta útil na eliminação das redundâncias de informações e também para completar dados incompletos, através da desfragmentação das informações referente a um mesmo individuo. Dessa maneira, as informações geradas com o novo banco de dados são mais consistentes e confiáveis para serem usadas como ferramentas no auxilio de tomada de decisões na área da saúde (Figura 1.3).

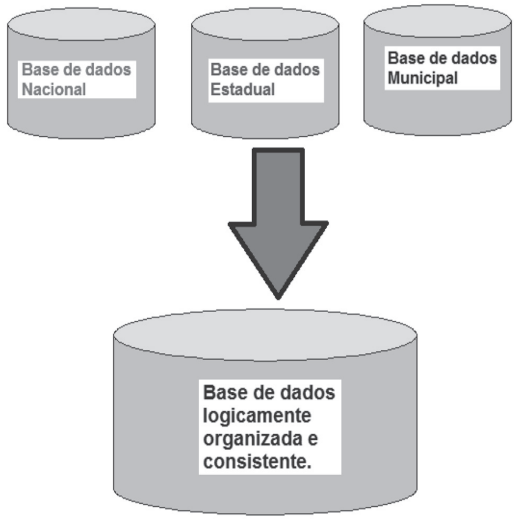

Figura 1.3: Base de dados consistente obtido a partir de várias bases distintas.

2 – Classificando uma incompletude ou uma inconsistência

A mente humana quando se depara com inconsistências, tenta obter mais informações que o auxilie a tomar determinada decisão. Por exemplo, quando um coordenador de curso vai decidir por aceitar ou não determinado aluno a um programa de doutorado, o mesmo busca a maior quantidade de informações possíveis, como mostra a Figura 2.1, para saber se o candidato tem ou não potencial para cumprir, com sucesso, um programa de doutorado.

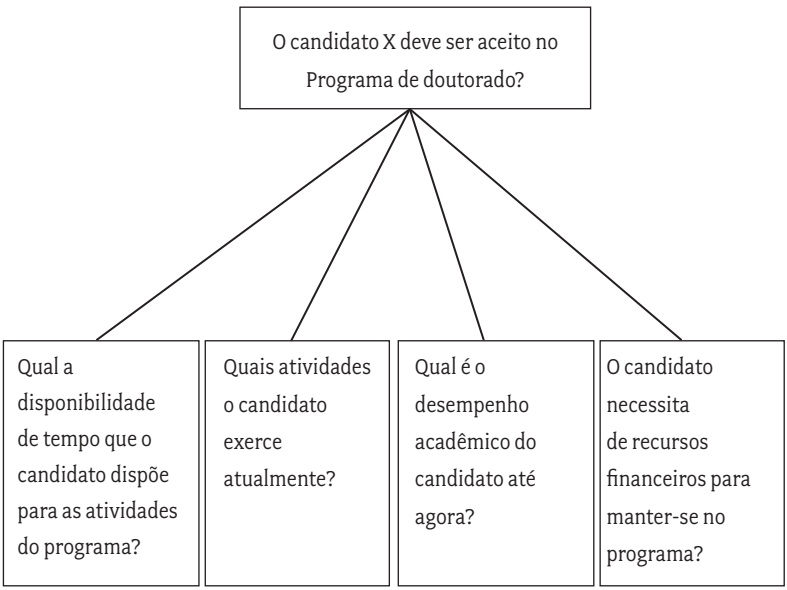

Figura 2.1: Algumas informações úteis para a tomada de decisão para aceitação de determinado candidato a um programa de doutorado.

A criação de modelos artificiais do cérebro humano é uma ferramenta que procura aproximar a maneira que os computadores "pensam" da maneira que o cérebro humano pensa, com o objetivo de me-

lhorar o desempenho dos mesmos (de Morais, Alana M., et al., 2010), principalmente em relação à tomada de decisões.

Desenvolver uma ferramenta computacional, baseada em modelos artificiais para classificar as incompletudes e também as inconsistências, é uma ferramenta muito útil para a mineração de dados. A necessidade de fazer tal classificação se dá no fato de que, nem toda inconsistência ou incompletude, prejudica a utilização do dado (incompleto ou inconsistente) em um processo de tomada de decisões, como mostra a Figura 2.2.

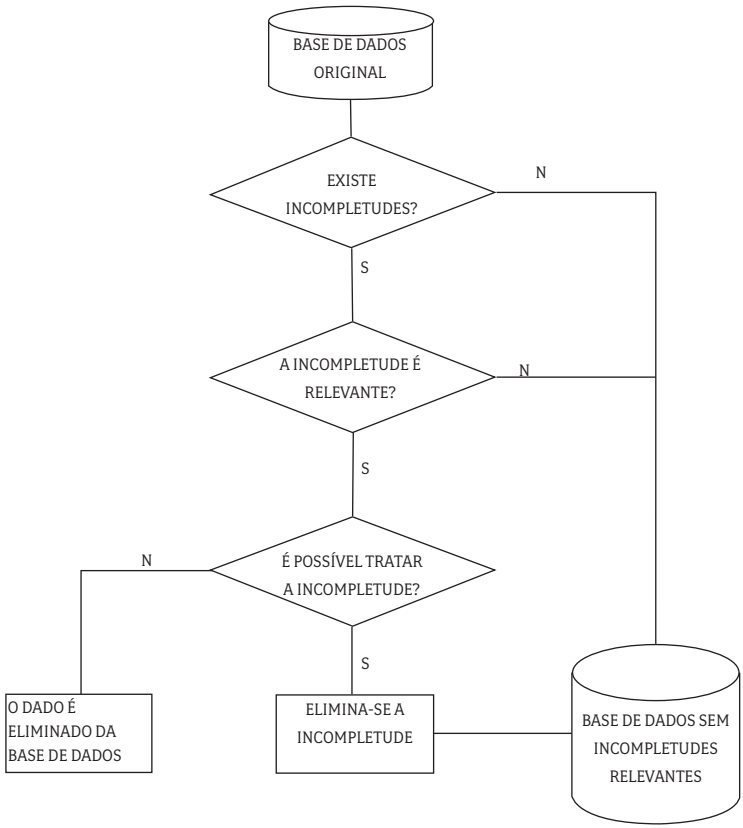

Figura 2.2: Classificação das inconsistências e incompletudes.

Um exemplo desse tipo de incompletude seria a falta do campo "Nome do proprietário" de alguns veículos, no registro dos mesmos, de um determinado modelo que apresentaram determinado tipo de defeito. Para fazer uma análise dos dados e decidir por fazer ou não um recall nesse modelo, a informação "Nome do proprietário" é irrelevante e por isso tal incompletude pode ser desconsiderada, pois a mesma não altera no processo de análise.

Também a inconsistência nesse campo não prejudicaria a integridade dos dados para a uma análise sobre a realização do recall ou não. Por exemplo, se em um determinado dia um veículo tem seu campo "Nome do proprietário" preenchido com o valor "Carlos Alberto de Albuquerque" e em outro dia o mesmo veículo tem registrado no mesmo campo o valor "Angélica Heloísa de Albuquerque", para o processo que fará a análise sobre a realização ou não do recall tal inconsistência é irrelevante, não interferindo no processo de análise.

É importante perceber que a classificação das incompletudes e inconsistências deve ser analisada para cada objetivo específico. Uma incompletude que pode ser considerada irrelevante para determinado processo de tomada de decisão, pode ser fundamental em outro processo. Por exemplo, a mesma incompletude citada acima como irrelevante, torna se fundamental, portanto, não podendo de forma alguma ser desconsiderada, em processo de análise para o envio de correspondência, solicitando o pagamento de proprietários que estão inadimplentes com o serviço de manutenção há mais de 2 meses. Nesse caso tal incompletude altera significativamente o processe de análise, pois mesmo sabendo que o proprietário está inadimplente, não será possível identifica-lo e consequentemente notifica-lo.

Após classificar as incompletudes dos registros, o objetivo passa ser o de classificar o registro como consistente ou inconsistente. Se o registro for classificado como consistente, o mesmo será incluído na base de dados que servirá como fonte de informações para o processo de tomada de decisão. Caso contrário, o mesmo será eliminado da base de dados.

Pode-se sintetizar o termo tratamento de dados como, a partir de uma base com dados inconsistentes e paracompletos, construir uma nova base de dados onde os mesmos sejam logicamente organizados e assim possam fornecer informações importantes para a tomada de decisões.

2.1 – A lógica Paraconsistente como ferramenta para tratamento de dados

Até hoje não se inventou uma máquina de processamento melhor do que o cérebro humano. Uma prova disso são os sistemas especialistas que são baseados na inteligência artificial e que tentam imitar o raciocínio humano (Altran, 2010).

Tais sistemas precisam descrever algo do mundo real, que pode ser um objeto, como uma chave fusível ou uma afirmação, como "A estrada está boa e por isso não haverá problemas na viagem". Para fazer a descrição dos objetos citados acima, ou validar uma afirmação, a lógica clássica mostra-se insuficiente, pois apresenta apenas dois valores, falso e verdadeiro, e que são excludentes (Lambert, et al., 2008).

Para atender a necessidade dos sistemas de processamento, como os sistemas especialistas, é necessário a utilização de uma lógica que seja capaz de descrever os objetos e situações do mundo real (Lambert, et al., 2008) de forma semelhante, mais próxima possível, as executadas pela mente humana. Entre as lógicas não-clássicas, a Lógica Paraconsistente (LP) é mostrada em vários estudos como uma ferramenta capaz de executar tal tarefa, auxiliando, ou mesmo, substituindo a lógica clássica em processos onde a mesma mostra ser insuficiente (Lambert, et al., 2008).

Quando se descreve algo do mundo real, é natural ocorrer situações de inconsistências ou mesmo de conhecimentos parciais (Martins, 2003). Por exemplo, para descrever o estado de conservação de uma chave fusível, utilizando a lógica clássica, o mesmo pode ser descrito como ***BOM*** ou ***RUIM***. Aqui o ***BOM*** tem o valor lógico ***VERDADEIRO*** e ***RUIM*** o valor ***FALSO***. Porém como mostra (Martins, 2003), as descrições dos objetos e mesmo das situações reais nem sempre são moldadas pela lógica clássica. A chave pode não estar ruim, mas também pode não estar tão boa. E pode acontecer, também, de se não saber o real estado da mesma.

Uma ferramenta que tem potencial para trabalhar com processamento de dados com tais inconsistências é a utilização da lógica paraconsistente, que trabalha com valores lógicos além dos clássicos ***VERDADEIRO*** e ***FALSO*** (Da Costa, 93).

3 – Mineração de Dados

Neste capítulo será visto alguns conceitos que são utilizados em técnicas de mineração de dados. Entre os principais conceitos pode-se citar a funcionalidade, a estrutura do processo de mineração de dados e também as diversas áreas de aplicação.

3.1 – Definição

A mineração de dados (MD) tem como objetivo principal extrair informações de um grande volume de dados e classifica-los. Muitas vezes fica difícil descobrir informações importantes que constam em uma base de dados muito extensa. A MD procura identificar padrões e tendências que podem ser encontrados em uma grande quantidade de dados. A MD é uma fase muito importante de um processo que tem a finalidade de extrair conhecimento em banco de dados, processo esse que é denominado de *Knowledge Discovery in Databases* (KDD) (Fayyad et all, 1996). Através do KDD é possível aplicar a MD para obter, de uma base de dados extensa, uma nova base onde os dados sejam consistentes, ou seja, não haja na base dados com registros incompletos ou divergentes (Politi, 2005).

Um exemplo de padrão que poderia ser identificado, utilizando a MD, seria os registros de atendimentos de emergências realizados em uma determinada região. Essa região abrange as cidades A, B, C e D, como mostra a Figura 3.1.

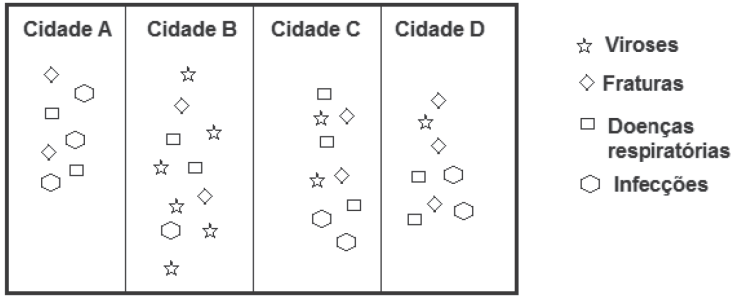

Figura 3.1: Registros de atendimentos (fictícios) da região composta pelas cidades A, B, C e D.

Em um exame visual da Figura 3.1, fica fácil identificar que na cidade B a incidência de doenças causadas por vírus é bem maior do que nas demais cidades, ficando com um total de dois terços de todos os casos da região.

Com base na informação fornecida pela Figura 3.1, as autoridades podem chegar à conclusão de que a cidade B é deficitária na área do saneamento básico, visto que a grande maioria das doenças causadas por vírus e bactérias, é decorrente de más condições sanitárias (Soares et all, 2002). Com base nessa conclusão, as autoridades podem decidir investir mais em saneamento básico em tal cidade.

É claro que a Figura 3.1 fornece um exemplo fictício e didático, por isso contém pequena quantidade de dados. Mas na realidade os bancos de dados contêm um número de dados muito grande, sendo muito difícil para alguém encontrar padrões, simplesmente analisando um gráfico como o da Figura 3.1, para esses dados.

3.2 – Aplicações

A MD tem uma vasta gama de aplicações, indo desde a área comercial, passando pelas áreas da saúde, educação, segurança e abrangendo até a área científica (Politi, 2005).

Um exemplo de aplicação na área comercial seria o gerente de uma loja analisar o banco de dados contendo o cadastro de seus clientes. Dessa forma poderia verificar o perfil (padrões) dos mesmos. Por exemplo, se for identificado que uma parte considerável de clientes possui filhos com idade entre zero e doze anos, o proprietário da loja pode decidir abrir um setor infantil na loja.

Um exemplo da área da saúde já foi dado no item 3.1. Na área de segurança, ao se analisar os registros de ocorrências, as autoridades policiais podem verificar que existe um padrão que indica que a maioria dos assaltos à residência, acontece em determinados bairros e poderia assim intensificar o policiamento em tal bairro.

A aplicação da MD na Internet se mescla com as demais áreas citadas nesse livro. A Internet é uma enorme fonte de dados, abrangendo praticamente todas as áreas do conhecimento (Teixeira, 1997). Por isso, a Internet mostra-se uma área muito promissora para a MD.

Na área científica a MD é utilizada principalmente para descobrir padrões e também para desenvolver modelos capazes de simular comportamento dos vários fenômenos da natureza (Politi, 2005). Esses fenômenos podem ser físicos, químicos ou biológicos.

3.3 – Funcionalidade

Entende-se por funcionalidade as tarefas empregadas na MD (Politi, 2005). A funcionalidade também pode ser entendida como o objetivo da MD. Uma das principais funcionalidades é MD descritiva, que como o próprio nome diz, procura descrever as características existentes nos dados.

Um exemplo de uma funcionalidade descritiva é quando se pretende encontrar características específicas em uma base de dados. Em uma base de dados que contem as informações dos clientes de uma loja, como foi citado no item 3.2, e o gerente da loja quer caracterizar todos os clientes que têm filhos de zero a doze anos. Aqui se utilizou uma funcionalidade descritiva.

Outra funcionalidade em destaque na MD é a preditiva (Politi,

2005). Ao contrário da descritiva, que tem como objetivo apenas descrever as características dos dados, tem como objetivo criar um modelo, ou seja, através de inferências dos dados disponíveis, encontrar um padrão de comportamento dos dados e assim poder prever novos valores para tais dados.

Entre as funcionalidades preditivas, pode-se citar a análise de evolução (Politi, 2005). Também como o próprio nome diz, avalia a evolução temporal dos dados, tentando prevenir, no decorrer do tempo, o comportamento dos dados. Um exemplo dessa funcionalidade é um banco de dados que contém os estoques e a quantidade de saída das mercadorias de um supermercado. Analisando os dados, é possível prever que determinada mercadoria terá um acréscimo em sua venda em determinado período do ano, e que por isso, deve-se encomendar uma quantidade maior dessa mercadoria para esse período. Também pode se prever que determinado produto deve ter um decréscimo em suas vendas e que por isso deve-se diminuir as encomendas para tal produto.

4 – O modelo relacional para Banco de Dados

Um banco de dados pode ser definido como uma coleção de dados persistentes (Date, 2003). Tais dados são armazenados em repositórios, onde podem ser atualizados e consultados. Por conter dados, esses repositórios se tornam uma fonte rica de informações que pode ser utilizada pelos proprietários.

Mas que tipo de informação contém esses repositórios? Um banco de dados é uma representação de algum objeto do mundo real (Elmasri e Navathe, 2011). Esse objeto pode ser o prontuário de pacientes de um hospital. Nesse caso, o banco representa o estado de saúde dos indivíduos que são atendidos por tal hospital.

Uma das arquiteturas mais conhecidas e também mais utilizadas para a arquitetura de um banco de dados é a "Entidade-Relacionamento" (Date, 2003), ou simplesmente, "O Modelo Relacional" (MR).

4.1 – Definição

Os principais objetivos do Modelo Relacional (MR) são: aumentar a independência de dados nos sistemas gerenciadores de bancos de dados e fornecer mais funcionalidade para o armazenamento, consulta e recuperação de dados. Para fornecer tais funcionalidades, o MR utiliza-se da álgebra relacional.

Como o próprio nome diz, a estrutura do MR é baseada em relações, que nada mais são do que tabelas (Date, 2000). Essas tabelas são compostas por campos, que são definidos como atributos (Ramalrishna et all, 2002). Cada linha dessas tabelas é chamada de tupla, que é análogo a um registro em um arquivo. Essa linha também é denominada de instância.

A Figura 4.1 mostra um exemplo de relações, ou seja, as tabelas: PACIENTE, MÉDICO, PLANTÃO e CONSULTA. Na tabela PACIENTE estão

os dados referentes aos pacientes de um hospital. Existem quatro campos nessa tabela: Nº do prontuário, Nome, Endereço e idade. Uma instância dessa tabela seria 001, Carlos..., Rua A e 47, ou seja, uma linha da tabela. Conforme já foi mencionado nesse livro, uma instância seria análoga a um registro de um arquivo. Aqui a tabela representa um arquivo.

É importante a escolha das denominações das tabelas e também dos campos das mesmas, pois são fundamentais no auxílio da interpretação do significado dos valores de cada instanciamento. Olhando a Tabela MÉDICO da Figura 4.1 percebe-se que os dados de tal tabela são referentes aos médicos que trabalham nesse hospital. Mais precisamente, cada linha dessa tabela contém os dados referentes a um determinado médico que trabalha no hospital.

PACIENTE

No Prontuário	Nome	Endereço	Idade
001	Carlos...	Rua A	47
002	Angélica H...	Rua 8	41
003	Beatriz de ...	Rua C	22
004	Patrícia H...	Rua D	15

PLANTÃO

Escala de Plantão	Matrícula Médico
10/02/2013	234
20/02/2013	098

MÉDICO

Matrícula	Nome	CPF	Endereço	Salário
057	Pedro ...	XXXXX	Rua A ...	10000
098	Maria ...	XXXXX	Av. X	9000
234	Carlos A..	XXXXX	Rua 5	14000

CONSULTA

Data Consulta	Matrícula Médico	Nome do Médico	Prontuário
25/03/2013	234	Carlos	003
01/04/2013	057	Pedro ...	001

Figura 4.1: Exemplo de Tabelas Relacionais

Os nomes das colunas das tabelas, no caso da Tabela Médico, Matrícula, Nome, CPF, Endereço e Salário, também são muito importantes na interpretação dos valores de cada linha de uma tabela. Usando novamente a Tabela Médico, verifica-se que a primeira coluna se refere à matrícula do médico, a segunda ao Nome do médico, a terceira ao CPF, a quarta ao Endereço e por último, a quinta se refere ao salário recebido pelo mesmo.

Sintetizando, na terminologia de base de dados relacional uma tabela é denominada relação, cada uma de suas linhas é denominada uma tupla ou instancia, cada coluna é denominada de coluna de atributos e o tipo dos valores que podem ser atribuídos é chamado de domínio do atributo.

A vantagem do MR em relação aos modelos que os precederam é que no MR não existem caminhos pré-definidos para consultar ou alterar os dados armazenados, pois implementa uma estrutura baseada em relações.

Para manipular essas tabelas ou relações, é necessário estabelecer algumas restrições para evitar inconsistências como: duplicidade de dados, incompletudes e perda de dados. Essas restrições são denominadas de: integridade referencial, chave primária, chave estrangeira e integridade de junções de relações.

Um exemplo de inconsistência seria o Nº do prontuário de um paciente ser diferente em duas relações, por exemplo, nas Tabelas PACIENTE e CONSULTA da Figura 4.1. Dessa forma o sistema poderia enxergar dois pacientes diferentes, sendo na verdade um mesmo paciente que teve o seu número de prontuário registrado com dois valores diferentes.

Também poderia acontecer o contrário, ou seja, um mesmo número de prontuário ser atribuído a mais de um paciente. Nesse caso o sistema poderia enxergar dois pacientes como sendo somente um.

Problemas, como o citado acima, tornam uma base de dados inconsistente, pois fornece informações que não são verídicas. Se o administrador do hospital for dimensionar sua equipe médica usan-

do como parâmetro o número de pacientes cadastrados em sua base de dados, provavelmente fará um dimensionamento errado, alocando mais recursos humanos do que realmente seria necessário ou fazendo o contrário, alocando menos recursos humanos do que seria necessário para atender todos os pacientes do hospital.

4.2 – Conceitos do MR

Como já foi citado nesse livro, o objetivo dos modelos de dados é representar objetos ou situações do mundo real. Através da MR é possível representar qualquer objeto ou situação do mundo real, utilizando um dos quatro construtores sintáticos: Conjunto de Entidades, Conjunto de Relacionamentos, Atributos de Entidades e Atributos de Relacionamentos (Ramalrishna et AL, 2002).

4.2.1 – Conjunto de Entidades

Na MR qualquer objeto ou situações do mundo real, como pessoas, departamentos, profissão, cargo ou consulta, pode ser representado por uma entidade.

Um conjunto de entidades que possuem a mesma estrutura e o mesmo significado na modelagem é denominado de Conjunto de Entidades e é representado por um retângulo (Date, 2000).

Figura 4.2: Conjuntos de Entidades Médico e Paciente.

A Figura 4.2 ilustra os Conjuntos de Entidades Médico e Paciente. É importante verificar que o nome do conjunto de entidades é sempre um substantivo singular e o mesmo deve ser significativo, ou

seja, auxiliar na interpretação do conjunto de entidades. Olhando, novamente, a Figura 4.2, fica fácil constatar que o conjunto de entidades denominado Médico se trata de um conjunto de entidades que representa os dados de médicos. Analogamente, o conjunto de entidades denominado Paciente, representa os dados referentes aos pacientes.

4.2.2 – Conjunto de Relacionamentos

No mundo real os objetos relacionam-se entre si. Por exemplo, um médico atende a um paciente, ou um paciente é atendido por um médico.

Esses relacionamentos podem ser vistos como uma associação, ou seja, no mundo real os objetos se associam. Analisando a Figura 4.3, verifica-se que existe uma associação entre o médico e o paciente.

No MR os objetos, que são representados por conjuntos de entidades, também se relacionam entre si. Os relacionamentos entre entidades dos mesmos conjuntos que sejam estrutural e semanticamente iguais são representados através de Conjuntos de Relacionamentos (Date, 2000).

A Figura 4.3 mostra dois conjuntos de relacionamentos: Atende e Trabalha. O relacionamento Atende indica que existe uma associação entre uma das entidades do conjunto de entidades Médico, com uma das entidades do conjunto de entidades Paciente. A analogia vale para o relacionamento Trabalha.

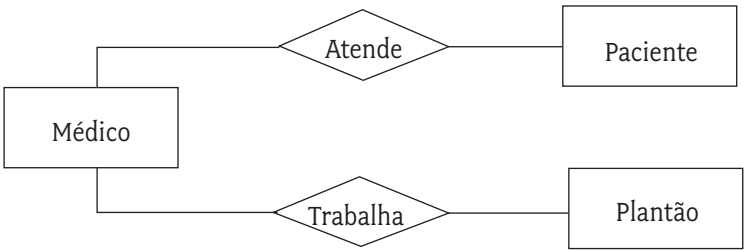

Figura 4.3: Conjuntos de Relacionamentos Atende (entre os conjuntos de entidades Médico e Paciente) e Trabalha (entre os conjuntos de entidades Médico e Plantão).

A representação de um conjunto de relacionamento é feita por um losango e o nome que representa esse conjunto de relacionamento é um verbo. No exemplo da Figura 4.3, os nomes dos relacionamentos são os verbos Atende e Trabalha.

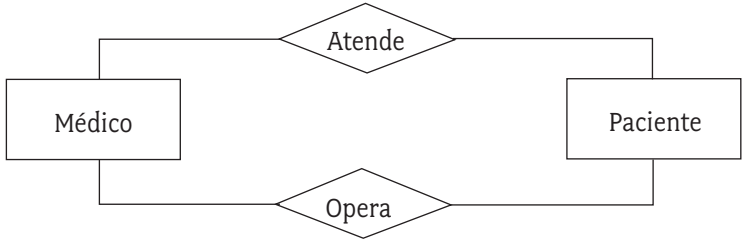

Figura 4.4: Dois conjuntos de relacionamentos (Atende e Opera) entre as mesmas duas entidades de relacionamentos (Médico e Paciente).

Entre dois conjuntos de entidades pode haver vários conjuntos de relacionamentos, como mostrado na Figura 4.4. Nessa figura um médico pode atender a um paciente e também pode operar (realizar uma cirurgia) esse paciente. O médico também pode atender um paciente e operar um outro paciente.

4.2.3 – Atributos de entidades relacionamentos

O MR procura representar os objetos e seus respectivos relacionamentos em um mundo real e faz isso através dos conjuntos de entidades e de relacionamentos. Para representar esses objetos e relacionamentos são necessários valores que descrevam, tanto as entidades como os relacionamentos. Esses valores são modelados através de atributos (Date, 2000).

MÉDICO	ATENDE	PACIENTE
• MATRICOLA		• N. PONTUÁRIO
• NOME		• NOME
• CPF		• ENDEREÇO
• ENDEREÇO		• IDADE
• SALÁRIO		

Figura 4.5: Atributos dos conjuntos de relacionamentos Médico e Paciente.

A Figura 4.5 modela os atributos dos conjuntos de entidades Médico e Paciente. Os atributos do conjunto de entidades Médico são Matrícula, Nome, CPF, Endereço e Salário. Já para o conjunto de entidades Paciente os atributos são Nº Prontuário, Nome, Endereço e idade. Cada atributo é representado por uma elipse ligada ao conjunto de atributos.

MÉDICO	ATENDE	PACIENTE
• MATRICOLA	• PROGNÓSTICO	• N. PONTUÁRIO
• NOME		• NOME
• CPF		• ENDEREÇO
• ENDEREÇO		• IDADE
• SALÁRIO		

Figura 4.6: Atributo do relacionamento Atende.

Um conjunto de entidade que não possui atributos que o descrevam, não deve ser modelado no MR, pois caso isso ocorra é porque certamente tal conjunto de entidades foi erroneamente modelado como tal (Date, 2000).

Um conjunto de relacionamento também pode ter atributos, como mostra a Figura 4.6. Nessa figura está modelado o atributo prognóstico, pois quando um médico atende um paciente, o mesmo emite um prognóstico.

A modelagem dos atributos de um conjunto de relacionamento é feita da mesma forma do que a feita para um conjunto de entidades.

Como já foi mencionada nesse trabalho, a ausência de atributos para um conjunto de entidade indica que tal conjunto de entidades não deveria ser modelado como tal. Essa afirmação não vale para um conjunto de relacionamentos, ou seja, a ausência de atributos para um conjunto de relacionamento não invalida sua modelagem como tal. Portanto um conjunto de relacionamentos pode ter ou não atributos.

4.2.4 – Chaves

Uma ferramenta valiosa no MR para eliminar as inconsistências em um banco de dados é denominada chave. Através da chave é possível identificar univocamente cada entidade de um conjunto de entidades. A chave é também o principal meio de acesso a uma entidade específica dentro de um conjunto de entidades (Date, 2000).

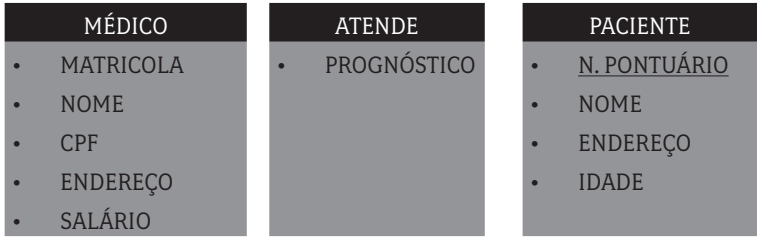

Figura 4.7: Chave primária do conjunto de entidades Paciente.

Uma chave nada mais é do que um atributo que identifica univocamente uma entidade dentro de um conjunto de entidades. Esse atributo é denominado chave primária e deve ser sublinhado na modelagem. A Figura 4.7 mostra a modelagem da chave primária do conjunto de entidades Paciente. Nesse caso, sua chave primária é o atributo Nº de prontuário, visto que cada paciente tem seu próprio número de prontuário e assim cada um dos pacientes fica perfeitamente identificado através do mesmo.

Uma chave primária também pode ser composta por mais de um atributo. Nesse caso, todos os atributos que compõem a chave primária, devem ser sublinhados. A Figura 4.8 mostra a modelagem da chave primária do conjunto de entidades Médico, composta por dois atributos: Matrícula e CPF. Nesse caso, a identificação de uma determinada entidade dentro do conjunto de entidades Médico dá se pelos atributos Matrícula e CPF.

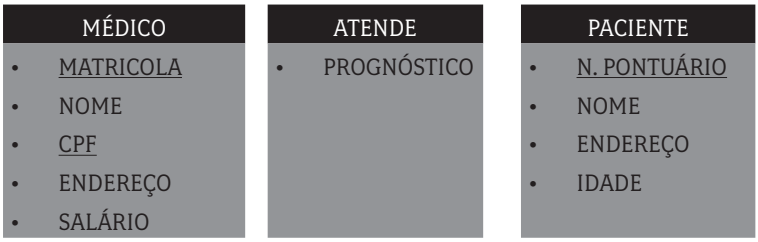

Figura 4.8: Chave primária, composta por dois atributos, do conjunto de entidades Médico.

Além da chave primária, existe também a chave estrangeira. Uma chave estrangeira é um atributo (ou um conjunto de atributos) que aponta (ou apontam) para a chave primária de outro conjunto de entidades. Assim, a chave estrangeira pode ser vista como o canal que relaciona um registro de um conjunto de entidades, com outro registro de outro conjunto de entidades.

Um exemplo de chaves estrangeiras é o conjunto de entidades Consulta da Figura 4.1, onde tem os atributos Matrícula do médico (que é chave primária do conjunto de atributos Médico) e Prontuário (que é chave primária no conjunto de atributos Paciente) são chaves estrangeiras.

A chave estrangeira tem um papel fundamental no MR pois é o mecanismo que implementa o relacionamento entre as tabelas, ou seja, entre os arquivos em um banco de dados (Heuser, 2001).

4.2.5 – Papéis no relacionamento

Cada conjunto de entidades que participa de um conjunto de relacionamento tem um papel no relacionamento. A Figura 4.9 mostra o papel do conjunto de entidades Médico no conjunto de relacionamentos Atende. Aqui o papel do médico é atender ao paciente.

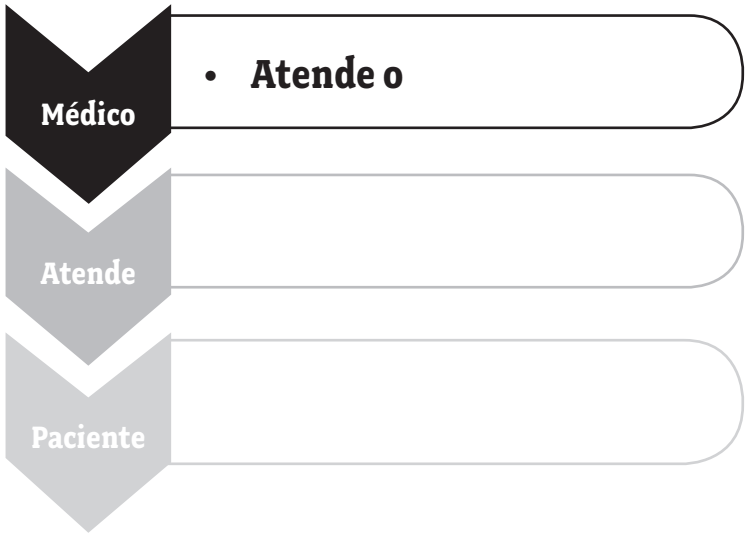

Figura 4.9: O papel do conjunto de entidades Médico no conjunto de relacionamentos Atende.

Mas o conjunto de entidade Paciente, também tem o seu papel nesse relacionamento, como mostra a Figura 4.10. Aqui o papel de cada paciente, do conjunto de entidades Paciente, é ser atendido por um médico do conjunto de entidades Médico.

O papel de cada entidade em um relacionamento sempre vai existir, mas não é necessária sua indicação. A indicação só será necessária quando houver a possibilidade de existir ambiguidade na interpretação do conjunto de relacionamentos.

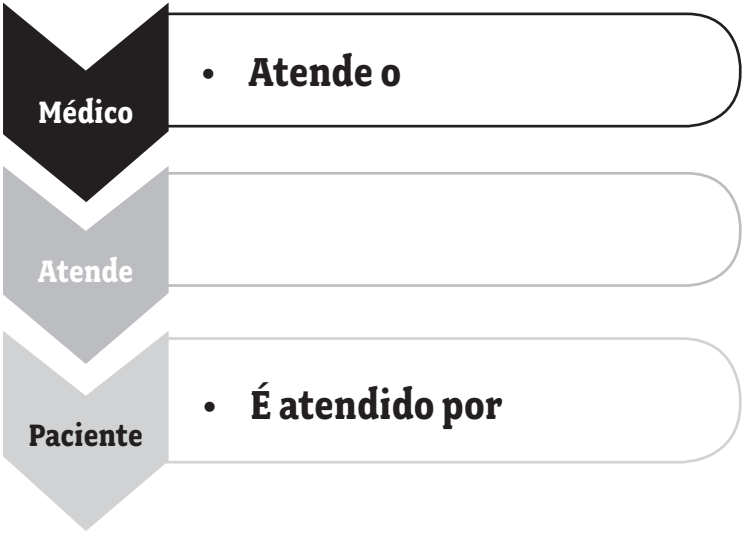

Figura 4.10: Os papéis dos conjuntos de entidades Médico e Paciente no conjunto de relacionamentos Atende.

4.2.6 – Cardinalidade dos relacionamentos

Um conjunto de relacionamento associa uma das entidades de um determinado conjunto de entidades à outra entidade de outro conjunto de entidades. Nesse caso a cardinalidade do conjunto de relacionamento é um para um e indica-se 1:1 (Date, 2000). Na modelagem mostrada na Figura 4.11 tem-se um conjunto de relacionamento 1:1, pois cada remédio do conjunto de entidades Remédio possui uma, e somente uma bula, no conjunto de entidades Bula. Isso é indicado pelo número um colocado ao lado dos retângulos que representam os conjuntos de entidades Remédio e Bula.

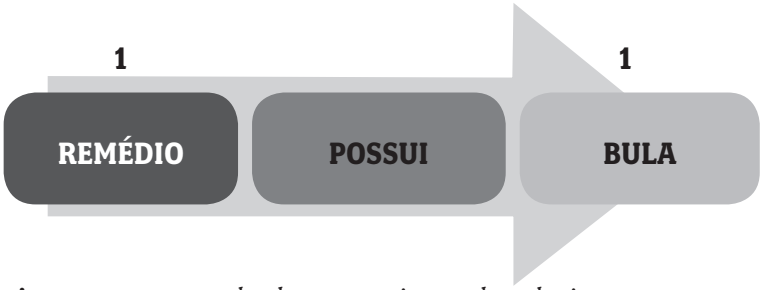

Figura 4.11: Exemplo de um conjunto de relacionamento com cardinalidade 1 para 1.

Também existem conjuntos de relacionamentos onde cada uma das entidades de um conjunto de entidades, pode se relacionar com mais de uma das entidades de outro conjunto de entidades. Nesse caso, diz-se que a cardinalidade do conjunto de relacionamento é de um para muitos e indica-se por: 1:N (Date, 2000). A Figura 4.13 mostra uma modelagem 1:N. Aqui cada professor pode administrar várias disciplinas mas cada disciplina pode ser administrado por somente um professor. O número um, colocado ao lado do retângulo que representa o conjunto de entidades Professor significa que cada disciplina pode ter somente um professor. Já o N colocado ao lado do retângulo que representa o conjunto de entidades Disciplina significa que cada professor pode lecionar mais de uma disciplina.

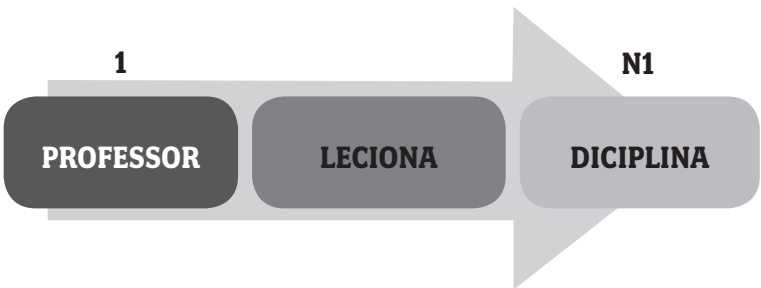

Figura 4.12: Exemplo de um conjunto de relacionamento com cardinalidade 1 para muitos.

Existem conjuntos de relacionamentos onde cada entidade de um conjunto de entidades pode relacionar-se com várias entidades de outro conjunto e vice-versa. Esse conjunto de relacionamento tem cardinalidade muitos para muitos e é indicado por N:N. A Figura 4.13 mostra um exemplo de um relacionamento N:N (Date, 2000). Na modelagem um médico pode atender vários pacientes e cada paciente pode ser atendido por vários médicos. Isso é indicado na modelagem através do N colocado ao lado do retângulo que representa o conjunto de entidades Médico e também colocado ao lado do retângulo que representa o conjunto de entidades Paciente.

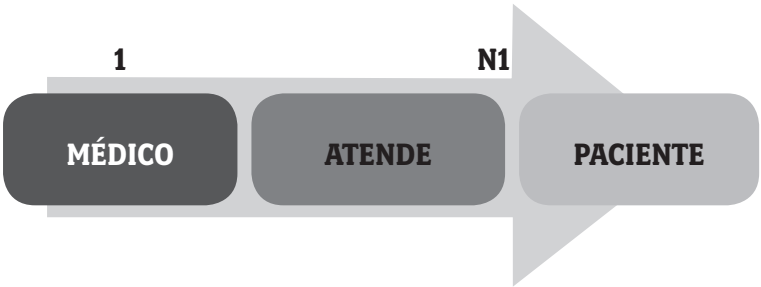

Figura 4.13: Exemplo de um conjunto de relacionamentos com cardinalidade muitos para muitos.

5 – A Lógica Paraconsistente Anotada de dois valores – LAP2v

Uma das propostas desse livro é aplicar a Lógica Paraconsistente Anotada de dois valores (LPA2v) como ferramenta na classificação e eliminação de inconsistência em base de dados, principalmente nas mais antigas.

5.1 – Conceitos básicos de LPA2v

Nesse livro será usado a Lógica Paraconsistente Anotada de dois valores (LPA2v) que (Martins, 2003) mostra que é uma função $f:(\mu_1,\mu_2)\rightarrow[0,1]\times[0,1]$ onde $\mu_1,\mu_2\in[0,1]$. Aqui (Martins, 2003) descreve μ_1 e μ_2 como o os graus de crença e descrença, respectivamente.

Ao contrário da lógica clássica onde uma proposição pode assumir apenas um de dois valores (Lambert-Torres, Costa e Moraes, 2008) (verdadeiro, falso), na Lógica Paraconsistente (LP) cada proposição pode assumir qualquer valor pertencente ao intervalo real fechado [0,1]. (Lambert-Torres, Costa e Moraes, 2008) denota esse valor de μ e mostra que quando $\mu=1$ equivale ao valor lógico verdadeiro na lógica clássica e também mostra que quando $\mu=0$ equivale ao valor lógico falso, para os valores de crença e descrença descritos anteriormente.

Falando em termos de estatística, μ_1 pode representar a probabilidade de que um evento aconteça. Analogamente, μ_2 pode representar a probabilidade de que um evento não aconteça. Como mostra (Spiegel, 2008), a probabilidade de um evento ocorrer pertence ao intervalo fechado [0,1]. O mesmo acontece com a probabilidade de um evento não ocorrer, por isso pode-se relacionar μ_1 e μ_2 com a probabilidade.

Como mostra (Martins, 2003), a representação gráfica dessa função tem como resultado o quadrado unitário do plano cartesiano (QUPC) mostrada na Figura 5.1. No QUPC (Martins, 2003) também des-

tacou os pontos **A**, **B**, **C** e **D**, como também pode ser visto na Figura 5.1.

No QUPC (Lambert-Torres, Costa e Moraes, 2008) interpreta o eixo das abscissas (eixo horizontal) como o eixo de crença, ou seja: dada uma proposição, no eixo das abscissas será locado o nível de crença ou de esperança que um especialista tem que essa proposição seja verdadeira. Aqui o termo especialista é usado no sentido genérico. O especialista pode ser tanto um especialista em determinado assunto e que por isso sua opinião tem grande valor na tomada de decisão, como pode ser uma teoria, como a probabilidade.

Analogamente (Lambert-Torres, Costa e Moraes, 2008) interpreta o eixo das ordenadas (eixo vertical) como o eixo de descrença, ou seja: dada uma proposição, no eixo das ordenadas será locado o grau de descrença que um especialista tem que essa proposição seja verdadeira. Tal especialista pode ser o mesmo que fornece o grau de crença ou não.

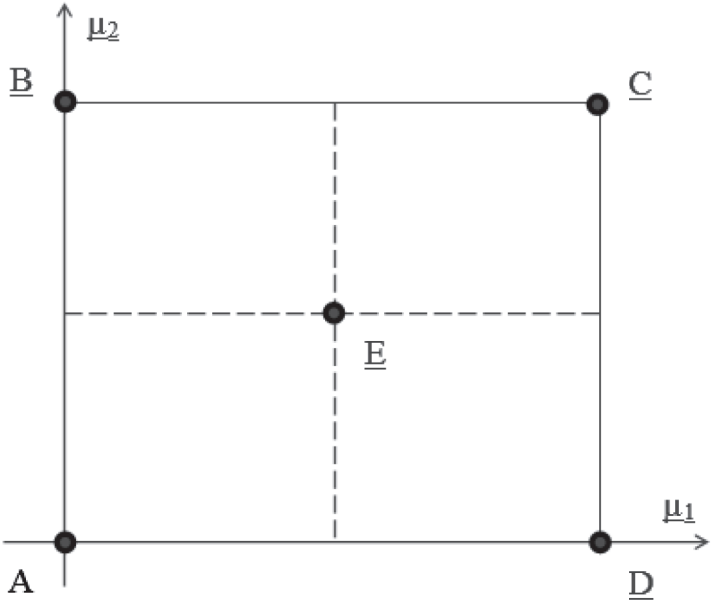

Figura 5. 1: (Martins, H. G, 2003): Pontos notáveis no QUPC

Na LPA2v cada um desses pontos tem seu significado. O ponto ***A*** que tem as coordenadas cartesianas **(0,0)** é o ponto paracompleto (Da Costa et al., 1999). Nesse ponto, tanto o especialista que define o grau de crença, como o que define o grau de descrença, atribui valor igual a 0. Adotando a simbologia de (Da Costa et al., 1999), temos que nesse ponto $\mu_1 = 0$ e $\mu_2 = 0$. Em (Abe, 2010) define que uma proposição onde $\mu_1 = 0$ e $\mu_2 = 0$ é uma proposição paracompleta, ou seja, não tem informações suficientes para definir se é falsa ou verdadeira. Essa interpretação é reforçada por (Borbon, Guido e Vieira, 2009) onde interpreta uma proposição com tais graus de crença e descrença como "uma proposição desconhecida".

Para o ponto ***B (0,1)*** (Da Costa, et al., 1999) denomina como ponto de valor lógico falso. Então (Abe, 2010) define uma proposição cuja $\mu_1 = 0$ e $\mu_2 = 1$ como uma proposição que tem o valor lógico falso. Isso é fácil de verificar, pois se $\mu_1 = 0$, significa que o grau de crença de que a proposição seja verdadeira é zero. Já $\mu_2 = 1$, significa que o grau de crença de que a proposição seja falsa é 1, ou 100%.

O ponto ***C (1,1)***, também na interpretação de (Da Costa, et al., 1999), é tido como sendo inconsistente. Aqui (Abe, 2010) define que uma proposição cuja $\mu_1 = 1$ e $\mu_2 = 1$ como uma proposição inconsistente. Igualmente como no caso do ponto $B\ (0\ ,1\ )$, fica fácil de verificar, já que $\mu_1 = 1$ significa que um especialista acredita que com certeza a proposição é verdadeira e já o grau $\mu_2 = 1$ significa que outro especialista acredita que com certeza a proposição é falsa, portanto caracterizando uma proposição inconsistente.

Já o ponto ***D (1,0)***, segundo a interpretação de (Da Costa, et al., 1999), é definido como ponto de certeza. Nesse ponto (Abe, 2010) atribui a uma proposição que tem os graus $\mu_1 = 1$ e $\mu_2 = 0$ como uma proposição verdadeira. Isso se verifica facilmente, visto que quando $\mu_1 = 1$ significa que um especialista acredita que com certeza a proposição é verdadeira e $\mu_2 = 0$ significa que um outro especialista acredita que com certeza a proposição não é falsa.

Por último tem-se o ponto ***E(0,5;0,5)***. Aqui (Da Costa et al., 1999) interpreta esse ponto como um ponto indefinido. Então uma proposi-

ção cujos graus de crença e descrença são respectivamente $\mu_1 = 0{,}5$ e $\mu_2 = 0{,}5$ é interpretado por (Abe, 2010) como uma proposição indefinida. Para justificar essa interpretação, esse trabalho toma a liberdade de reproduzir um ditado popular que diz o seguinte. "O otimista quando vê um copo de água pela metade, diz que o copo está meio cheio. Já o pessimista diz que o copo está meio vazio." Quem tem razão, o otimista ou o pessimista? Na verdade os dois têm razão e os dois não têm razão. Portanto é uma proposição indefinida.

Para traduzir essa interpretação em um exemplo mais concreto, esse trabalho utiliza o exemplo apresentado por (Da Costa et al., 1999): "Seja a proposição p: Pedrinho é suspeito de não ter ido à escola." Para essa proposição (Da Costa et al., 1999) enumera e justifica cada um dos cinco pontos notáveis, segundo o QUPC da LPA2v.

"O primeiro ponto (1,0), ou seja, onde $\mu_1 = 1$ e $\mu_2 = 0$, significa que a proposição é verdadeira, ou seja, realmente Pedrinho não foi à escola. Nesse caso $\mu_1 = 1$ pode, por exemplo, representar o fato de haver na chamada da classe de Pedrinho, registro de sua frequência e $\mu_2 = 0$ o fato de ninguém tê-lo visto em outro lugar diferente da escola no período letivo."

"O ponto (0,1), onde $\mu_1 = 0$ e $\mu_2 = 1$, significa que a proposição é falsa e com certeza Pedrinho não foi à escola. Aqui $\mu_1 = 0$ pode representar o fato de não haver registro de freqüência de Pedrinha na lista de chamada. Já $\mu_2 = 1$ pode representar o fato de alguém tê-lo visto em um lugar diferente da escola no período letivo."

Para o ponto (0,0), onde $\mu_1 = \mu_2 = 0$, segundo (Da Costa et al., 1999), pode dar uma interpretação de que a proposição é paracompleta. Ainda segundo (Da Costa et al., 1999) a interpretação para $\mu_1 = 0$ "poderia ser o fato de não haver registro de presença para Pedrinha na lista de chamada e $\mu_2 = 0$ poderia ser representado pelo fato de ninguém ter visto Pedrinho em lugar diferente da escola durante o período letivo."

Já para o ponto (1,1), onde $\mu_1 = \mu_2 = 1$ é interpretado por (Da Costa et al., 1999) como uma proposição inconsistente. Aqui " $\mu_1 = 1$ pode ser o fato de haver registro de freqüência para Pedrinho na lista

de chamada e $\mu_2 = 1$ pode ser o fato de alguém ter visto Pedrinho em outro lugar diferente da escola no período letivo."

E por último, n ponto (0,5 ; 0,5), onde $\mu_1 = \mu_2 = 0{,}5$ a proposição é interpretada como indefinida. O significado para " $\mu_1 = 0{,}5$ poderia ser o fato de perda do registro de chamada no dia em questão. Dessa forma a probabilidade de haver registro de presença para Pedrinho é de 50%. Para $\mu_2 = 0{,}5$ pode ser considerado que não haja ninguém para dizer se viu ou não Pedrinho em algum lugar diferente da escola no período letivo."

Como foi visto anteriormente, na LP uma proposição pode ser verdadeira e falsa ao mesmo tempo. Conseqüentemente, a LP não satisfaz o princípio da não contradição (Lambert-Torres, Costa e Moraes, 2008). Tal princípio diz que uma proposição não pode ser verdadeira e falsa ao mesmo tempo (Alencar Filho, 2002). Também como muito bem observado em (Lambert-Torres, Costa e Moraes, 2008), a LP não satisfaz o princípio do terceiro excluído, que diz que toda proposição ou é verdadeira ou é falsa (Alencar Filho, 2002), ou seja, uma proposição só assume um dos dois valores: falso ou verdadeiro e nunca um terceiro valor.

5.2 – Interpretando o QUPC

A simbologia utilizada para designar os pontos notáveis do QUPC, como mostra (Moraes, 2006) e (Rissino, 2009), é a mostrada na Tabela 1.

Ponto	Descrição	Símbolo
A(0,0)	Paracompleto	$\bot$
B(0,1)	Falso	F
C(1,1)	Inconsistente	$\top$
D(1,0)	Verdadeiro	V
E(0,5 ; 0,5)	Indefinido	I

Tabela 1: Os pontos do QUPC e seus significados

A união desses pontos formam seguimentos que dentro da LPA2v tem significados específicos. Conforme mostra (Rissino, 2009), o segmento originário da união dos pontos V e F formam a Linha Perfeitamente Consistente (LPC), pois os pontos cujas coordenadas pertencem ao segmento $\overline{FV}$ tem como abscissa o complemento da ordenada em relação à unidade, como pode ser visto na Figura 5.2. Por exemplo, se $\mu_1 = 73$, então $\mu_2 = 0{,}27$. Tanto (Martins, 2003) como (Rissino, 2009), mostram que para um ponto pertencer a LPC, basta que os graus de crença e de descrença satisfaçam a equação:

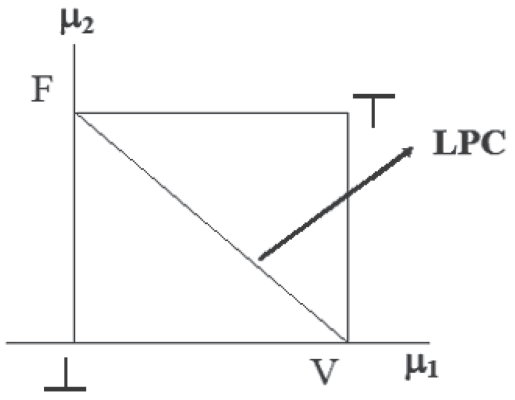

Figura 5.2: Linha Perfeitamente Consistente do QUPC

Equação 1

$$\mu_1 + \mu_2 - 1 = 0$$

Pode-se afirmar, então, que qualquer ponto, cujas coordenadas atendam a Equação 1, é considerado um ponto perfeitamente consistente, exceto o ponto que tem coordenadas $\mu_1 = \mu_2 = 0{,}5$, que embora atenda a Equação 1, é dito como indefinido, pelos motivos já citados anteriormente nesse trabalho.

A denominação de ponto perfeitamente consistente para as proposições, cujos graus de crença e descrença atendam a Equação 1, e

sejam diferentes de 0,5 é perfeitamente explicada pela Teoria da Probabilidade que mostra que a soma da probabilidade da ocorrência de um evento com a probabilidade da não ocorrência do mesmo evento é sempre igual a um (Spiegel, 2008). Portanto uma proposição com tais graus de crença e descrença pode ser denominada como uma proposição consistente, pois pela Teoria da Probabilidade tal proposição terá uma maior probabilidade de ser verdadeiro, caso $\mu_1 < \mu_2$, ou terá uma maior probabilidade de ser falso, caso $\mu_1 < \mu_2$.

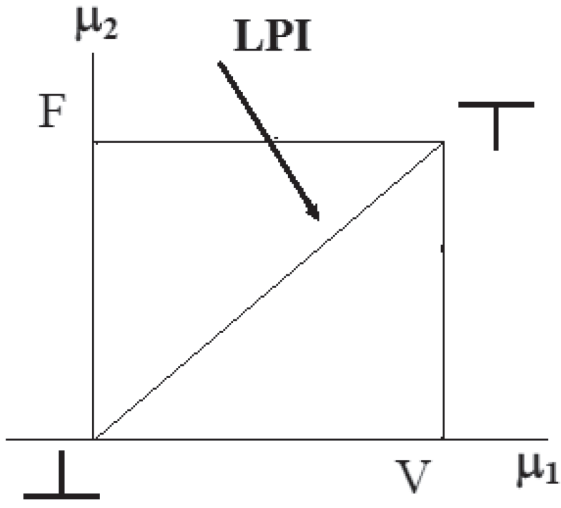

Figura 5. 3: Linha Perfeitamente Inconsistente

Na Figura 5.3 tem-se o segmento formado pela união dos pontos $\perp$ e $\top$. Esse segmento denomina-se Linha Perfeitamente Inconsistente (LPI) (Martins et al., 2007). Tal denominação se deve ao fato de que qualquer ponto cujas coordenadas coincidem com a LPI tem o mesmo grau de crença e descrença, ou seja, $\mu_1 = \mu_2$.

Para verificar se um ponto pertence a LPI (Martins, 2003) mostra que basta verificar se as coordenadas do mesmo (μ_1 , μ_2) satisfazem a equação:

Equação 2

$$\mu_1 - \mu_2 = 0$$

Pode-se concluir, então, que toda a proposição, cujos graus de certeza e incerteza satisfazem a Equação 2, é inconsistente. Essa conclusão também tem respaldo da Teoria da Probabilidade, visto que por essa teoria não é possível dizer se a maior probabilidade é de ocorrer ou não ocorrer o evento, pois ambos têm a mesma probabilidade de ocorrer. Na verdade somente quando se tem $\mu_1 = \mu_2 = 0{,}5$ é que se aplica totalmente a Teoria da Probabilidade, visto que nesse caso $\mu_1 + \mu_2 = 1$. Já para quando $\mu_1 = \mu_2 \neq 0{,}5$, embora possa interpretar como probabilidades de ocorrer e não ocorrer o evento, a soma das probabilidades será sempre diferente de um, o que diverge da Teoria da Probabilidade.

Na Figura 5.4, tem-se a o QUPC com as linhas perfeitamente consistente e inconsistente. Percebe-se claramente que a intersecção das mesmas é justamente no ponto I(0,5 ; 0,5), que é o ponto de indefinição.

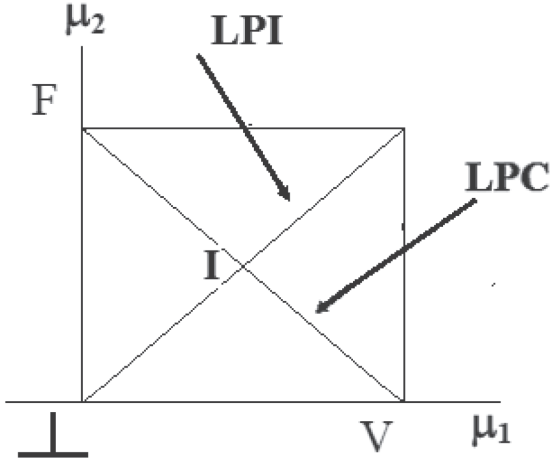

Figura 5.4: Linhas LPC e LPI no mesmo quadro unitário

A Figura 5.4 colabora para confirmar a classificação do ponto I como um ponto de indefinição, pois mostra geometricamente que o ponto pertence tanto a LPC quanto a LPI, que somado ao fato de suas coordenadas serem iguais a 0,5 reforçam sua classificação.

5.3 – Grau de Incerteza (GI) e Grau de Certeza (GC)

Já foi mostrado que todo ponto cujas coordenadas pertencem à LPC é perfeitamente consistente, exceto o ponto com coordenadas (0,5 ; 0,5). Agora será mostrado o que acontece com os pontos que não pertencem à LPC.

Se um ponto Q não pertence à LPC, então ele não é perfeitamente consistente, ou seja, existe um Grau de Incerteza (GI). Em (Martins, 2003) mostra que o GI é dado em função da distância entre a LPC e a reta paralela a mesma e que passa por esse ponto Q . Fica claro que se essa distância for zero, significa que o ponto Q pertence a LPC e conseqüentemente é um ponto perfeitamente consistente. Já se a diferença for diferente de zero, significa que o ponto Q não pertence à LPC e existe um GI.

Considerando um ponto cujas coordenadas sejam(μ_1 , μ_2), a equação que define o GI é dada por (Martins, 2003):

Equação 3

$$GI = \mu_1 + \mu_2 - 1$$

Para mostrar que a Equação 3 dá a distância entre a LPC e a reta paralela que passa pelo ponto Q , utiliza-se os recursos da geometria.

LPC é uma reta cuja equação é:

Equação 4

$$\mu_1 + \mu_2 - 1 = 0$$

Pode-se ver que na Figura 5.5 a distância da LPI a origem do quadro unitário é $\frac{\sqrt{2}}{2}$.

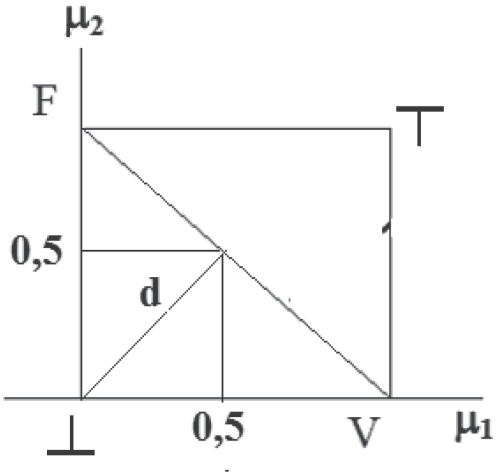

Figura 5.5: Distância da LPC á origem QUPC

Toda reta paralela a LPC tem coeficiente angular igual a -1 (igual ao da reta LPC) e, portanto, tem a forma $\mu_1 + \mu_2 - c = 0$. Fica fácil verificar que $c = \mu_1 + \mu_2$ para toda reta paralela a LPC.

Para encontrar a distância da reta $\mu_1 + \mu_2 - c = 0$ a origem (0,0) do quadro unitário, basta encontrar um ponto comum a essa reta e a reta $\mu_1 - \mu_2 = 0$. Esse ponto se dará em $\frac{c}{2}, \frac{c}{2}$ e será denominado ponto P. Como $c = \mu_1 + \mu_2$, temos que o ponto em comum é $P = \left(\frac{\mu_1 + \mu_2}{2}, \frac{\mu_1 + \mu_2}{2}\right)$, como mostra a Figura 5.6. Para encontrar a distância da reta $\mu_1 + \mu_2 - c = 0$ até a origem do quadro unitário basta calcular a distância do ponto P à origem do quadrado unitário.

A distância do ponto P à origem é dada por:

Equação 5

$$d_{PO} = \sqrt{\left(\frac{\mu_1 + \mu_2}{2}\right)^2 + \left(\frac{\mu_1 + \mu_2}{2}\right)^2} = \frac{\sqrt{2}(\mu_1 + \mu_2)}{2}$$

Para calcular, então, a distância entre a LPC e a reta paralela a mesma e que passa pelo ponto Q , basta calcular a diferença entre a distância do ponto P e a origem do QUPC, que é dada pela Equação 5, e $\frac{\sqrt{2}}{2}$, que é a distância da LPC à origem do QUPC. Essa distância é o GI.

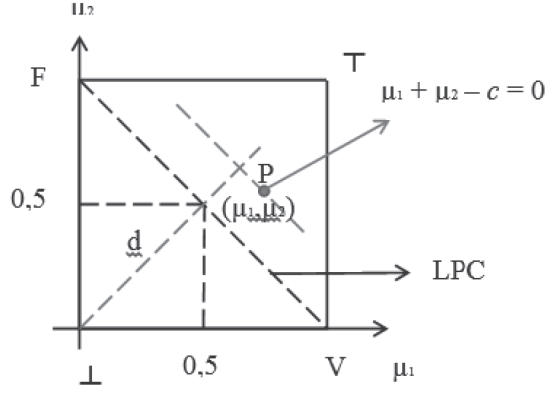

Figura 5.6: Reta paralela a LPC e o ponto P

Para normalizar esse resultado com o resultado fornecido por (Martins, 2003) através da Equação 3. divide-se o resultado da diferença entre a **Equação 5** e $\frac{\sqrt{2}}{2}$ por $\frac{\sqrt{2}}{2}$.

Pode-se, então, dizer que o GI pode ser encontrado por:

Equação 6

$$GI = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(\mu_1 + \mu_2)}{2}}{\frac{\sqrt{2}}{2}}$$

Verifica-se que tanto a Equação 3 como a Equação 6 fornecem o GI e concordam. em módulo, em seus resultados. Por exemplo, quando $\mu_1 = 0$ e $\mu_2 = 0$, temos:

$$GI = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(\mu_1 + \mu_2)}{2}}{\frac{\sqrt{2}}{2}} = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(0+0)}{2}}{\frac{\sqrt{2}}{2}} = 1 \quad \text{e}$$

Para $\mu_1 = 1$ e $\mu_2 = 1$, temo

$$GI = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(\mu_1 + \mu_2)}{2}}{\frac{\sqrt{2}}{2}} = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(1+1)}{2}}{\frac{\sqrt{2}}{2}} = -1$$

, que novamente concorda, em módulo.

Por último, esse trabalho testou os pontos $\mu_1 = 0{,}25$ e $\mu_2 = 0{,}75$, obtendo os seguintes resultados:

$$GI = \mu_1 + \mu_2 - 1 = 0{,}25 + 0{,}70 - 1 = -0{,}05$$

$$GI = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(\mu_1 + \mu_2)}{2}}{\frac{\sqrt{2}}{2}} = \frac{\frac{\sqrt{2}}{2} - \frac{\sqrt{2}(0{,}25 + 0{,}70)}{2}}{\frac{\sqrt{2}}{2}} = 0{,}05$$

,

que também concordam em módulo.

Assim, como já foi mostrado que todo ponto pertencente à LPC é um ponto perfeitamente consistente, também foi mostrado que um ponto que pertença à LPI é igualmente um ponto perfeitamente inconsistente.

Para mostrar que todo ponto que pertence à LPI é um ponto perfeitamente inconsistente, esse trabalho novamente utiliza da geometria.

A equação que define a LPI é dada por

Equação 7

$\mu_1 - \mu_2 = 0$.

Para satisfazer a Equação 7, os pontos tem que ter coordenadas iguais, ou seja, $\mu_1 = \mu_2$. Como aqui μ_1 representa a crença em que a proposição seja verdadeira e μ_2 a crença em que a proposição seja falsa, temos que tanto a crença na verdade como na falsidade da proposição são iguais, o que caracteriza uma inconsistência.

Considerando um ponto cujas coordenadas sejam (μ_1, μ_2), a equação que define o GC é dada por (Martins, 2003):

Equação 8

$GC = \mu_1 - \mu_2$.

Fica claro que todo ponto que não pertence à LPI terá um valor de GC diferente de zero. Da mesma forma que (Martins, 2003) mostrou o significado da distância da reta paralela à LPC para o GI, (Martins,

2003) mostra que a distância de uma reta paralela à LPI, passando por um ponto de coordenadas (μ_1, μ_2), fornece o grau de consistência do referido ponto. Para (Martins, 2003) quanto menor a distância de tal reta paralela for da LPI, menor será o GC. E quanto maior for essa distância, maior será o GC, tendo como valor máximo 1 e mínimo 0.

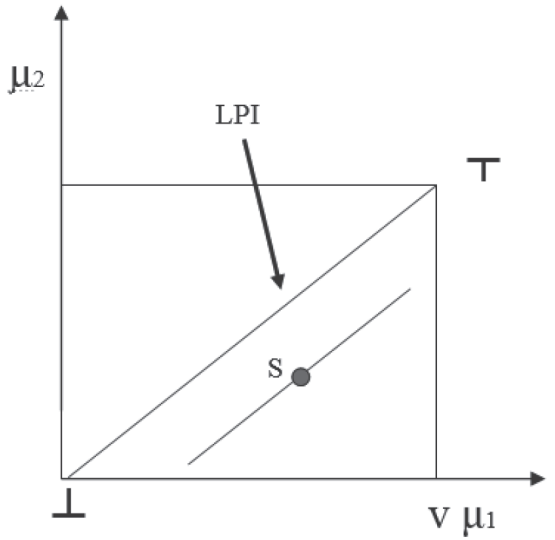

Figura 5.7: Reta paralela à LPI, passando pelo ponto S.

Como toda reta paralela à LPI, passando por um ponto S (Figura 5.7) tem o mesmo coeficiente angular da equação (7), a equação de tal reta é dada por:

Equação 9

$\mu_2 - \mu_1 - b = 0$,

onde $b = \mu_2 - \mu_1$.

Utilizando do mesmo raciocínio geométrico que foi utilizado para calcular o GI, este trabalho mostra que a distância de LPI ao ponto de coordenadas (1,0) é igual a $\frac{\sqrt{2}}{2}$.

Para encontrar a distância da reta $\mu_2 - \mu_1 - b = 0$ a origem (1,0) do quadro unitário, basta encontrar um ponto comum a essa reta e a reta $\mu_1 + \mu_2 - 1 = 0$, como mostra a Figura 5.8.

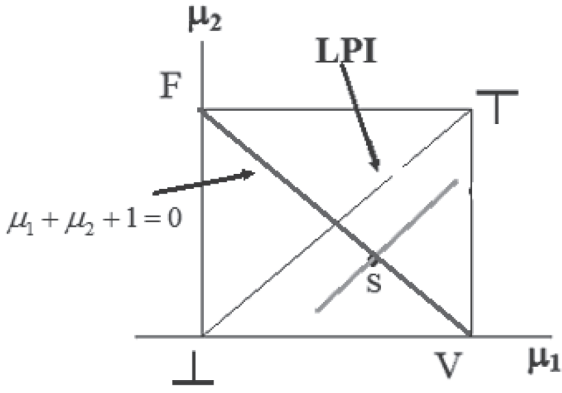

Figura 5.8: Ponto de interseção da reta paralela à LPI e a diagonal do quadrado unitário

Esse ponto se dará em $S = \left(\frac{1-b}{2}, \frac{1+b}{2} \right)$.

Calculando a distância entre o ponto S e o ponto (1,0), tem-se:

Equação 10

$$d_{SV} = \sqrt{\left(\frac{1-b}{2} - 1\right)^2 + \left(\frac{1+b}{2} - 0\right)^2}$$
$$= \sqrt{2}\frac{(b+1)}{2}$$

Como $b = \mu_2 - \mu_1$, a Equação 10 fica:

Equação 11

$$d_{SV} = \sqrt{2}\frac{(\mu_2 - \mu_1 + 1)}{2}$$

Agora, calcula-se a distância entre a LPI e a reta paralela à mesma e que passa pelo ponto S, efetuando a operação $d_{SV} - \frac{\sqrt{2}}{2}$.

Assim como foi feito no cálculo do GI, também será dividido o resultado da diferença descrita acima, por $\frac{\sqrt{2}}{2}$ para calcular o CG. Dessa forma, o GC será dado por:

Equação 12

$$GC = \frac{\sqrt{2}\frac{(\mu_2 - \mu_1 + 1)}{2} - \frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}}$$

Também como foi feito no cálculo do GI, foi verificando a concordância entre as Equações 8 e a Equação 12 para as coordenadas (0,0), (1,1) e (0,25 ; 0,75), onde os resultados são mostrados abaixo.

Para $\mu_1 = 0$ e $\mu_2 = 0$, temos:

$$\begin{aligned} GC &= \mu_1 - \mu_2 \\ &= 0 - 0 \\ &= 0 \end{aligned}$$

e

$$GC = \frac{\sqrt{2}\dfrac{(\mu_2 - \mu_1 + 1)}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}}$$

$$= \frac{\sqrt{2}\dfrac{(0-0+1)}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}} = \frac{\sqrt{2}\dfrac{1}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}} = 0.$$

Para $\mu_1 = 1$ e $\mu_2 = 1$, temos

$$\begin{aligned} GC &= \mu_1 - \mu_2 \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

e

$$GC = \frac{\sqrt{2}\dfrac{(\mu_2 - \mu_1 + 1)}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}}$$

$$= \frac{\sqrt{2}\dfrac{(1-1+1)}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}} = \frac{\sqrt{2}\dfrac{1}{2} - \dfrac{\sqrt{2}}{2}}{\dfrac{\sqrt{2}}{2}} = 0.$$

E finalmente para as $\mu_1 = 0{,}25$ e $\mu_2 = 0{,}75$

$$GC = \mu_1 - \mu_2$$
$$= 025 - 0,75$$
$$= -0,5$$

e

$$GC = \frac{\sqrt{2}\,\frac{(\mu_2 - \mu_1 + 1)}{2} - \frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}}$$

$$= \frac{\sqrt{2}\,\frac{(0,75 - 0,25 + 1)}{2} - \frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}} = \frac{1,5\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}} = 0,5,$$

Como os valores, tanto para a Equação 8 como para a Equação 12, têm o mesmo módulo, verifica-se que também para o ponto que tem coordenada (1,1), há concordância entre as equações citadas nesse parágrafo.

5.3.1 – Vantagens da obtenção do Grau de Incerteza e do Grau de Certeza

(Martins, 2003) afirmou que com o valor de GI pode verificar se o ponto está acima ou abaixo da LPC. Se o valor do GI for positivo, implicaria que tal ponto estaria acima da LPC. Para um GI negativo, implicaria que tal ponto estaria abaixo da LPC e como já foi mencionado nesse trabalho, se GI for nulo, então o ponto pertenceria à LPC.

(Martins, 2003) também utilizou o GC para verificar se o ponto correspondente está acima ou abaixo da LPI. Para um GC positivo implica que o ponto está abaixo da LPI, para GC negativo acima e para o GI nulo, pertenceria à LPI.

Porém, as fórmulas obtidas geometricamente, dão o resultado em módulo, ou seja, não distingue se a distância é positiva ou negativa. Para saber se um ponto, por exemplo, com um GI = 0,76, está acima ou abaixo da LPC utilizando a Equação (6), é necessário analisar graficamente o QUPC

para, então, afirmar se tal ponto está acima ou abaixo da LPC.

O mesmo ocorre com o GC. Também seria necessário analisar o QUPC para verificar se o ponto com um respectivo GC, diferente de zero, estaria acima ou abaixo da LPI.

Já com as Equações dadas**,** por (Martins 2003) essa análise não é necessária, visto que tais Equações não fornecem os valores em módulo.

A importância em saber se um ponto está acima ou abaixo da LPC ou da LPI está em identificar a que região tal ponto pertence dentro do QUPC. Por exemplo, um ponto cujo GI é um valor positivo e o GC é um valor negativo, está na região mostrada na Figura 5.9 (Martins 2003).

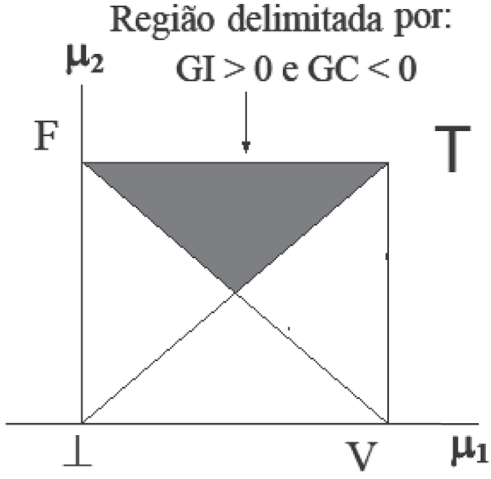

Figura 5.9: Região do QUPC para GI > 0 e GC < 0

Quando tanto o GI como o GC são positivos, este ponto pertence a área mostrada na Figura 5.10 (Martins 2003).

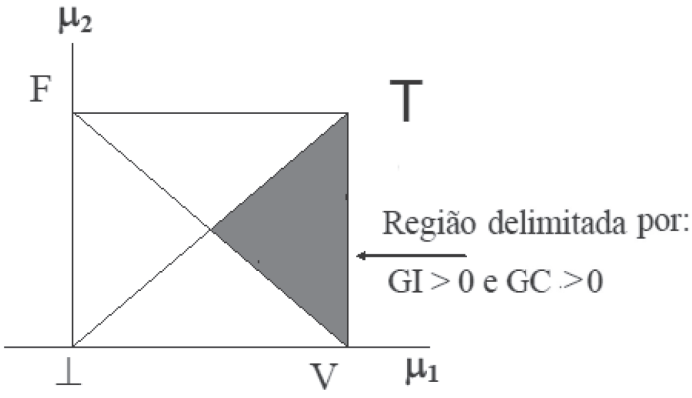

Figura 5.10: Região do QUPC para GI > 0 e GC > 0

5.4 – Os pontos notáveis do QUPC e o significados das regiões formadas à partir dos mesmos

Vários autores já identificaram os cinco pontos notáveis no QUPC, mostrados na Figura 5.1. Com esses pontos notáveis foi possível identificar duas importantes linhas que são formadas à partir de tais pontos: LPI e LPC (Martins 2003).

Também é possível identificar mais quatro pontos notáveis. Tais pontos foram denominados por (Martins 2003) como:

a) ponto quase não-falso, cujas coordenadas são (0,5;0) e utilizou-se como símbolo para o mesmo q¬F. O ponto notável q¬F pode ser visto na Figura 5.11. Este ponto é denominado por quase não-falso porque se tem uma crença de 50% de que a proposição é verdadeira e uma descrença de 0%. Embora não se tenha evidência suficiente para afirmar que tal proposição com esses graus de crença e descrença seja verdadeira, pois 50% não é suficiente para tal afirmação, a informação adicional de que o grau de descrença é de 0% contribui para a classificação de quase não-falso para o mesmo;

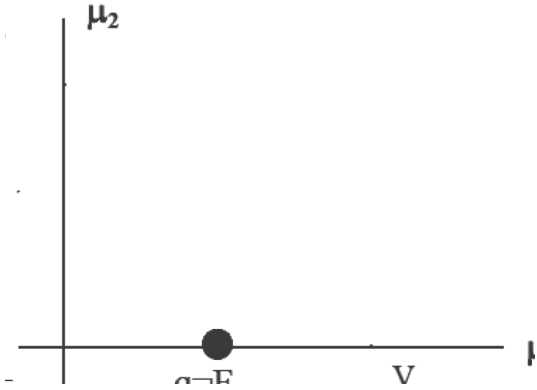

Figura 5.11: O ponto quase não-falso.

b) ponto quase verdade, cujas coordenadas são (1;0,5) e utilizou-se como símbolo para o mesmo qV. O ponto notável qV pode ser visto na Figura 5.12. A classificação de um ponto com tais coordenadas se justifica, pois tal ponto tem um grau de crença de 100% e um grau de descrença de 50%. Lembrando que o grau de crença é fornecido por um "especialista" e o de descrença por outro, então temos um especialista afirmando que a proposição é verdadeira e outro que diz que tem 50% de chance da proposição ser falsa;

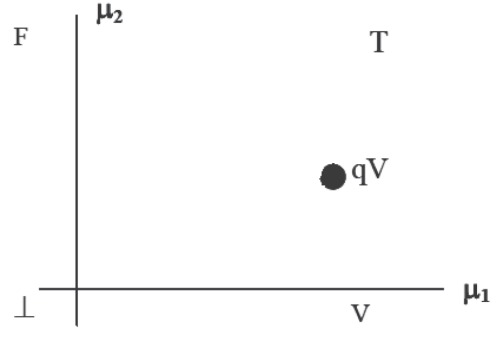

Figura 5.12: O ponto quase verdade.

c) ponto quase falso, cujas coordenadas são (0,5;1) e utilizou-se como símbolo para o mesmo q.F. O ponto notável q.F pode ser visto na Figura 5.13. Nesse ponto acontece justamente o contrário do ponto anterior. Aqui quem tem certeza de 100% é o especialista que dá o grau de descrença. O especialista que dá o grau de certeza tem 50% de convicção de que tal proposição é verdadeira. Pode-se afirmar, então, que há mais evidências de que a proposição seja falsa do que verdadeira;

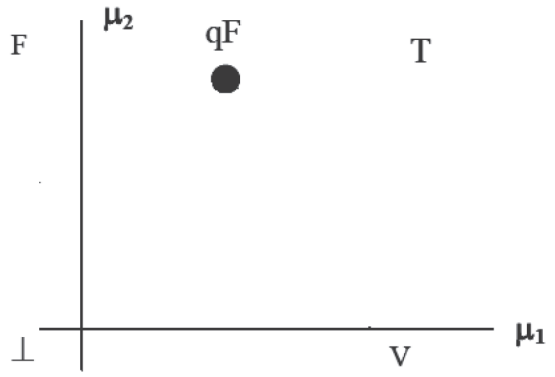

Figura 5.13: Ponto quase falso.

d) ponto quase não-verdade, cujas coordenadas são (0;0,5) e utilizou-se como símbolo para o mesmo q¬V. O ponto notável q¬V pode ser visto na Figura 5.14. Um ponto com tais coordenadas justifica ser classificado como quase não-verdade, pois o especialista que verifica o grau de crença na proposição não encontra evidência alguma de sua veracidade enquanto o especialista que verifica o grau de descrença, tem 50% de convicção de que tal proposição é falsa.

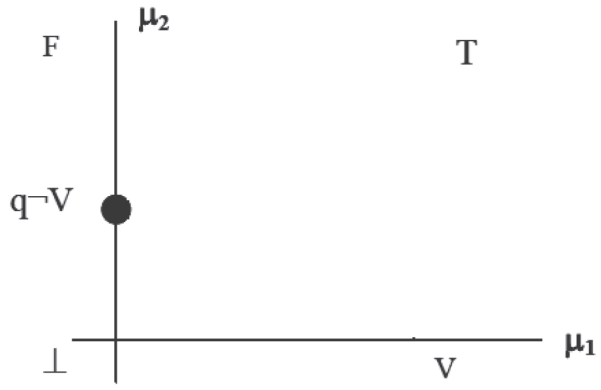

Figura 5.14: Ponto quase não-verdade.

A Figura 5.15 mostra os quatro pontos citados acima, em um único QUPC (Martins 2003).

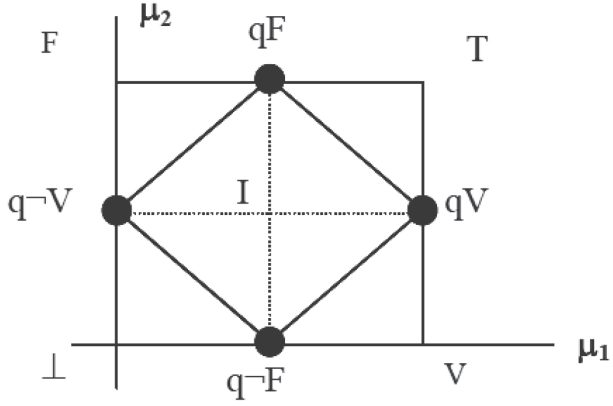

Figura 5.15 (Martins 2003): Os novos pontos notáveis no QUPC.

5.4.1 – Região de Verdade

Como mostra a Figura 5.15, com a inclusão dos pontos ***qF***, ***qV***, ***q¬F*** e ***q¬V***, é possível visualizar no ***QUPC*** novas linhas e novas regiões que tem significados diferentes. Por exemplo, a linha formada pelos pontos ***q¬F*** e ***qV*** é denominada por **(Martins 2003)** como uma linha que é a fronteira da região do ***QUPC***, na qual proposições que tem coordenadas nessa região são verdadeiras e que por isso essa região foi denominada Região de Total Verdade. Tal região pode ser visualizada na Figura 5.16

A denominação dessa região de total verdade encontra ampara na teoria da probabilidade, visto que os pontos cujas coordenadas pertencem a essa região são tais que $\mu_1 > \mu_2$, ou seja, o grau de crença será sempre maior que o grau de descrença, portanto tendo uma probabilidade maior de ser verdadeira do que falsa.

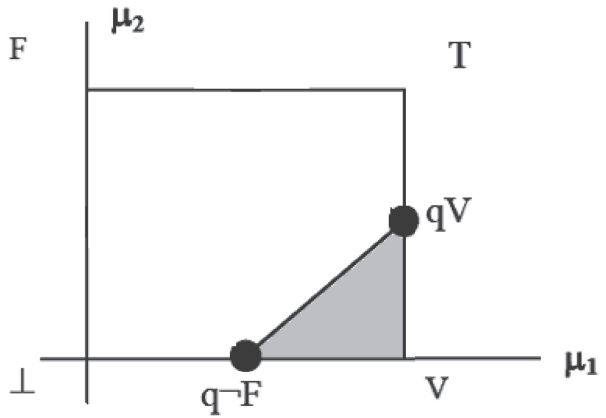

Figura 5.16: Região cujas coordenadas representam proposições verdadeiras.

Fica fácil perceber que quanto mais próximo do ponto ***V*** mais certeza é possível ter de que a proposição é verdadeira, pois nessa região o valor de μ_1, que representa o grau de crença, é muito próximo de 1, que em probabilidade significa 100%, tendo no limite valor igual a 1. Já o valor de μ_2, que representa o grau de descrença, fica muito próximo de zero, tendo no limite valor igual a 0. Porém quando o ponto fica próximo da linha formada pelos pontos ***q¬F*** e ***qV***, o nível de certeza diminui, visto que nessa região apesar do valor de μ_1 ser superior a 0,5, que em probabilidade significa ter mais de 50% de ser verdadeira, o valor de μ_2 é significativamente maior que zero, chegando a ter um valor bem próximo de 0,5, significando ter 50% de chance de ser falsa, sendo no limite do ponto ***qV*** igual a 0,5.

Uma decisão tomada com base em uma proposição que tem dois especialistas, um afirmando que o grau de crença em que a mesma seja verdadeira é de quase 1, ou 100%, e outro especialista afirmando que o grau de descrença é quase 0,5, ou 50%, dependendo da complexidade da decisão a ser tomada pode não ser considerada como verdadeira pelo tomador da decisão. Dificilmente um investidor investiria dez bilhões de reais em um empreendimento, onde um especialista afirmaria que tal empreendimento tem 99% de ser um sucesso e, portanto, dar lucro ao investidor, enquanto um outro especialista afirmaria que o mesmo empreendimento tem 49%, ou seja, quase a metade, de chance de ser um fracasso e, portanto, o investidor ter prejuízo, podendo até mesmo perder seus dez bilhões de reais.

Para problemas mais complexos, como o exemplificado acima e que, portanto, requer uma decisão baseada em indicadores mais sólidos, **(**Carvalho et al., 2003**)** propôs limitar a região de verdade pela linha formada pelo pontos de coordenadas (0,6; 0) e (1; 0,4) respectivamente, como mostra a Figura 5.17.

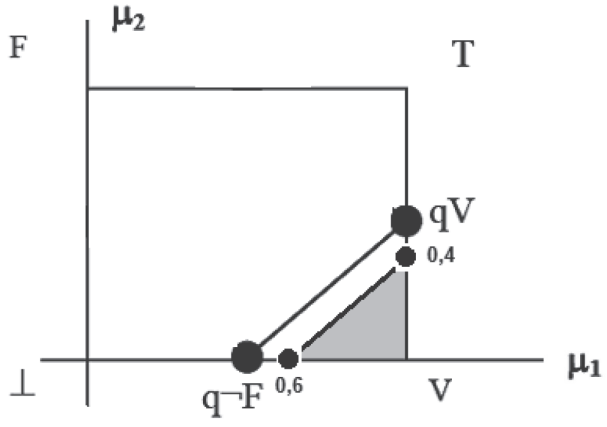

Figura 5.17: Região de verdade segundo (Carvalho et. al., 2003).

Para decisões mais complexas, parece mais adequado a definição de Região de Verdade a apresentada por **(**Carvalho et al., 2003**)**, visto que nela o maior valor para o grau de descrença será de 0,4 (40%) e o menor grau de crença será de 0,6 (60%).

5.4.2 – Região Inconsistente

Todos os pontos pertencentes a essa região tem abscissas e ordenadas maiores que 0,5, ou seja, nessa região tanto μ_1 , que representa o grau de crença na proposição, como μ_2 , que representa o grau de descrença na mesma, são maiores que 0,5 em uma escala que vai de 0 a 1, como pode ser visto na Figura 5.18. Portanto toda proposição cujas coordenadas pertencem à região mencionada, são proposições inconsistentes. Diante disso, essa região é denominada de Região Totalmente Inconsistente (Martins 2003**)**.

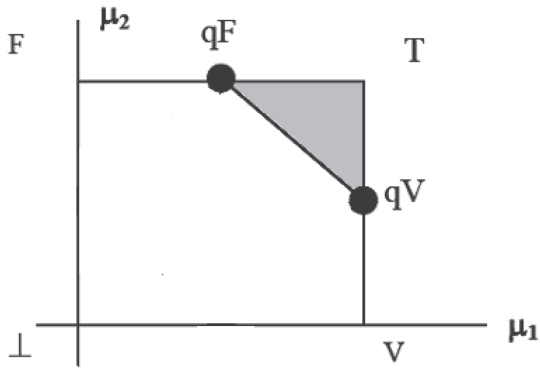

Figura 5.18: Região cujas coordenadas representam proposições Inconsistentes

Esse trabalho, mais uma vez, usa um exemplo que utiliza a teoria da probabilidade, para justificar a denominação de inconsistente a uma proposição cujos graus de crença e descrença são coordenadas de um ponto pertencente à região denominada de Região Totalmente Inconsistente. Suponha que um especialista afirme que há 90% de probabilidade de que determinado empreendimento seja um sucesso. Um segundo especialista afirma que há 85 % de chance de que o mesmo empreendimento não seja um sucesso. Assim tanto a crença como a descrença de que o empreendimento será um sucesso, são muito grandes. Nesse caso há, então, uma inconsistência.

Em relação à região de inconsistência (Carvalho et al., 2003) considera que somente os pontos cujas coordenadas que são maiores que 0,6 pertencem a tal região, como mostra a Figura 5.19.

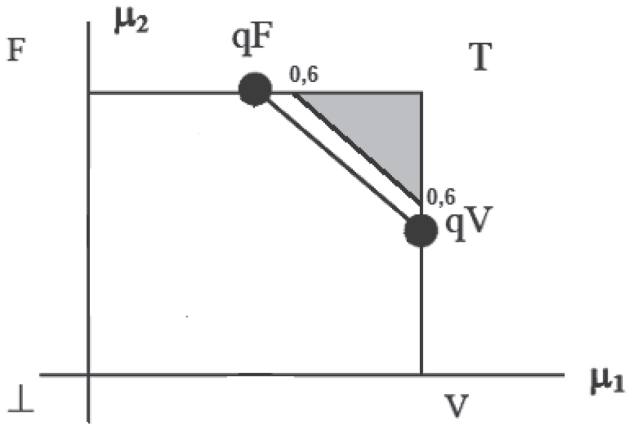

Figura 5.19: Região de Inconsistência segundo (Carvalho et. al., 2003).

Na avaliação desse trabalho a definição de região de inconsistência dada por (Martins 2003**)** é melhor que a dada por **(**Carvalho et al., 2003**)**. Por exemplo, uma proposição cujo grau de crença seja igual a 1(um) e o grau de descrença seja igual a 0,55, pertenceria a região de inconsistência segundo (Martins 2003**)**, mas não pertenceria segundo **(**Carvalho et al., 2003**)**. Porém dificilmente um empreendedor tomaria uma decisão de investir bilhões de reais em um empreendimento onde um especialista acredita que o mesmo tem 55% de chance de não ser um sucesso e outro acredita 100% que tal empreendimento será um sucesso.

5.4.3 – Região de Falsidade

Como mostra a Figura 5.20, todas as proposições cujas coordenadas pertencem a essa região, tem o valor de μ_1 sempre menor que μ_2. Em outras palavras, o grau de crença, representado por μ_1, será sempre menor que o grau de descrença, representado por μ_2.

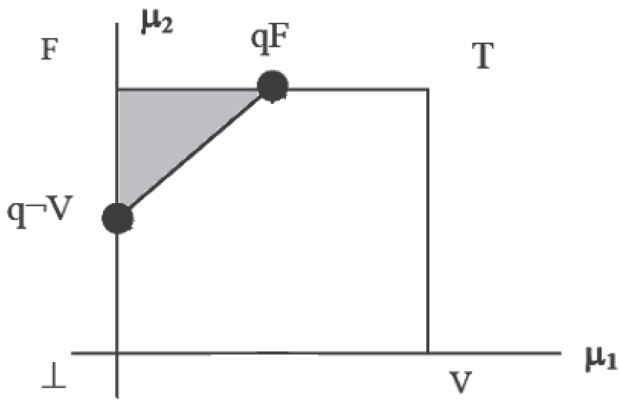

Figura 5.20: Região cujas coordenadas representam proposições Falsas

Mais importante ainda é perceber que nessa região o grau de crença será sempre inferior a 50% enquanto que o grau de descrença será sempre superior a 50%. Por isso (Martins 2003**)** denominou essa região de Região de Total Falsidade.

Continuando no exemplo do empreendedor, se o mesmo soubesse que um especialista acredita que a chance de sucesso de um empreendimento é de apenas 10% e um segundo especialista acredita que a chance de fracasso do mesmo empreendimento é de 85%, com certeza o empreendedor não investirá nesse empreendimento. Nesse caso é perfeitamente justo indicar que um ponto, com essas coordenadas, pertença à uma região de total falsidade.

Mas se o grau de crença e de descrença estivessem na região de falsidade, mas bem próximo da linha que passa pelos pontos ***q¬V*** e ***qF***, que limita a região de falsidade? Será que nesse caso o empreendedor teria a convicção de que tal empreendimento será um fracasso e por isso descartaria investir?

Para responder a essa pergunta, supõe-se agora, que o especialista que acredita no sucesso do empreendimento tenha apenas 1% de convicção de que será um sucesso. Já o especialista que avalia a chance de fracasso do mesmo, tem apenas 51% de convicção de que realmente

será um fracasso.

O responsável por tomar a decisão poderia entender essa situação como se o especialista que tem 51% de convicção que será um fracasso pode ser interpretado como se esse mesmo especialista acredita em uma chance de 49% de sucesso. Percebe-se que não fica tão óbvio o risco de investir. Portanto parece mais prudente não descartar tal empreendimento com base somente nesses dois especialistas.

Pelo exposto acima, não fica claro que um ponto que tenha as coordenadas próximo da linha que passa pelos pontos ***q¬V*** e ***qF***, não deve ser classificado como um ponto de total falsidade.

Também nesse caso a definição de Região de Falsidade adotada por (Carvalho et. al., 2003), e mostrada na Figura 5.21, mostra ser melhor do que a adotada por (Martins 2003), pois na primeira, uma proposição com os graus de crença e descrença como o do exemplo acima, ficaria fora da região de falsidade, o que é mais prudente.

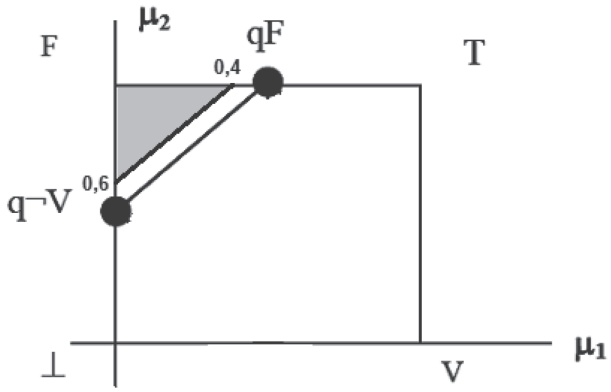

Figura 5.21: Região de Falsidade segundo (Carvalho et. al., 2003).

Para pertencer à região de falsidade, definida por (Carvalho et. al., 2003). o valor de μ_1 teria que ser de no máximo 0,4 e de μ_2 no mínimo 0,6. Continuando no exemplo do empreendedor usado por este trabalho, se um

especialista afirmasse que confia em 1% que um determinado empreendimento terá sucesso e outro especialista afirmasse que acredita que o mesmo tem 60% de chance de não ser um sucesso, somente um empreendedor com espírito de extremo aventureiro, realizaria um investimento em tal empreendimento. Então todos os pontos cujas coordenadas pertençam a área de falsidade adotada por (Carvalho et. al., 2003) é, inclusive os pontos próximo a linha limite que passa pelos pontos de coordenadas (0;0,6) e (0,4;1), um ponto de falsidade e por isso formam uma região de falsidade.

5.4.4 – Região Paracompleta

Essa região foi denominada assim por (Martins 2003) pelo fato de que todas os pontos pertencentes a mesma, tem valores de μ_1 e μ_2 menores que 0,5, como pode ser visto na Figura 5.22. Outra característica muito importante também, dessa região, é que a soma de μ_1 com μ_2 sempre tem como resultado um valor menor ou igual que 0,5.

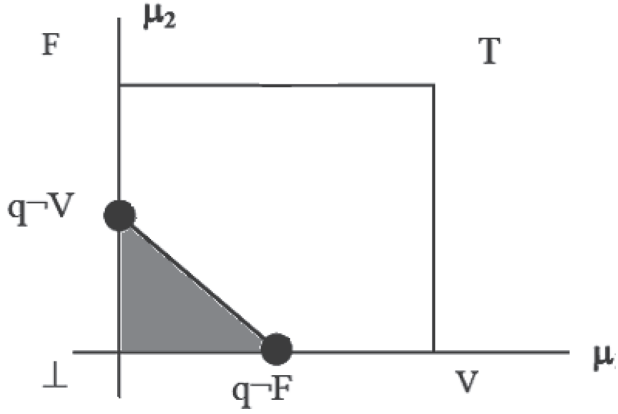

Figura 5.22: Região cujas coordenadas representam proposições Paracompletas

A palavra paracompleta significa que não existem evidências suficientes para definir se uma determinada proposição é falsa ou verdadeira. Nesse sentido, dizer que uma proposição que tem $\mu_1 < 0{,}5$, $\mu_2 < 0{,}5$ e além disso $\mu_1 + \mu_2 \leq 0{,}5$, é uma proposição paracompleta, é bastante lógico. Pois se um especialista afirmar ter 40% de chance de uma determinada proposição ser verdadeira e outro afirma que a chance de ser falsa é de 10%, fica muito difícil decidir se é realmente verdadeira ou falsa a proposição em questão. Se fosse invertido, ou seja, o especialista que afirma que a chance de ser verdadeira for de 10% e o que afirma que a chance de ser falsa for de 40%, novamente fica difícil decidir pela veracidade ou falsidade da proposição em questão.

Baseado no que foi exposto acima é completamente plausível classificar a região que contém todos os pontos que tem as coordenadas, como descritas no parágrafo anterior, como uma região totalmente paracompleta.

Porém a abordagem feita por (Carvalho et. al., 2003) para a região de totalmente pa racompleta tem mais consistência. Nela os valores de crença e descrença não são maiores que 0,4, como mostra a Figura 5.23.

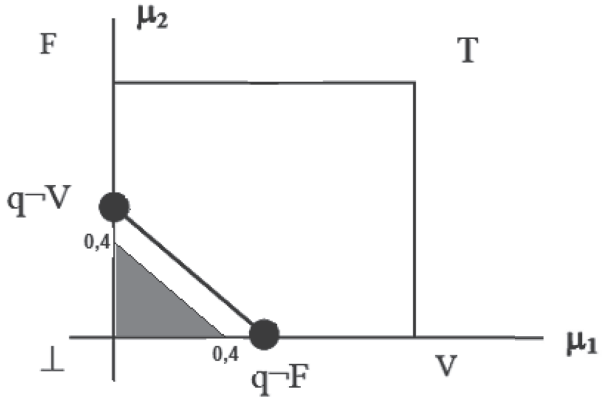

Figura 5.23: Região Totalmente Paracompleta segundo (Carvalho et. al., 2003).

5.4.5 – QUPC contendo as regiões de Verdade, Falsidade, Inconsistência e Paracompleta.

Entre os dois modelos de divisões em regiões do QUPC vistas aqui, tanto a definida por (Martins 2003), Figura 5.24, como a definida (Carvalho et. al., 2003), Figura 5.25, mostram serem ferramentas importantes na tomada de decisões, utilizando a LPA2v. Porém, pelos argumentos citados nos tópicos **5.4.1**, **5.4.2**, **5.4.3** e **5.4.4**, o segundo modelo é mais eficiente.

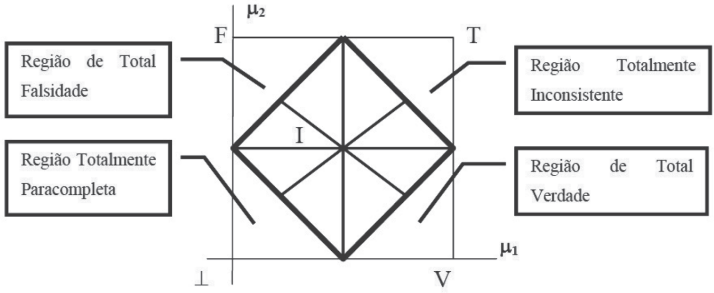

Figura 5.24: (Martins 2003): QUPC destacando as Regiões Específicas.

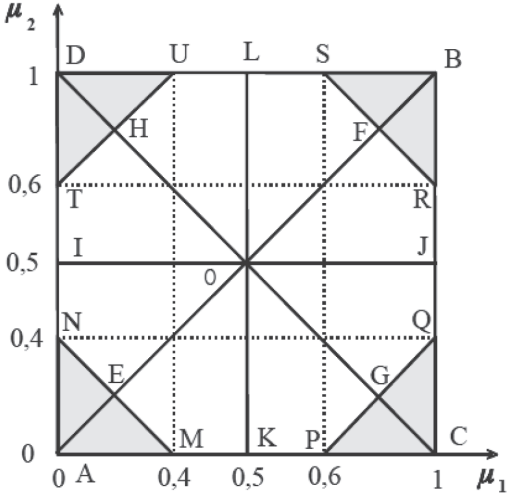

Figura 5.25: (Carvalho et. al., 2003): Divisão do QUPC em Regiões

6 – Utilizando a LP no tratamento de dados

O objetivo desse livro é mostrar a importância de utilizar uma ferramenta que, a partir de uma base de dados contendo informações inconsistentes como incompletudes, dados ruidosos e duplicidades, seja capaz de extrair informações para auxiliar em tomadas de decisões.

Para realizar essa tarefa, utilizou-se da LPA2v para a classificação das incompletudes. A LPA2v também foi utilizada para eliminar as incompletudes classificadas como relevantes.

O objetivo da ferramenta gerada utilizando a LPA2v é automatizar a identificação das incompletudes em uma base de dados. Diante da identificação das incompletudes, a ferramenta classifica as mesmas em relevante ou irrelevante.

Outro objetivo da ferramenta é eliminar as incompletudes, principalmente as que foram classificadas como relevantes. Em um primeiro momento, a ferramenta elimina todas as incompletudes que a LPA2v mostra ter um grau de crença elevado no tratamento da mesma. As incompletudes onde o grau de crença não é elevado o suficiente para que a mesma tome a decisão, a ferramenta sugere um valor.

Por exemplo, em uma base de dados de uma companhia energética, onde são cadastrados os equipamentos instalados em sua rede de distribuição, há registros onde o campo "código do equipamento" está ausente. A ferramenta identificou que noventa e cinco por cento dos registros, onde o campo "descrição do equipamento" coincide com o respectivo campo nos registros onde o campo "código do equipamento" está incompleto, tem como valor para o campo "código do equipamento" "CF20A" e que os outros cinco por cento estão com o campo "código do equipamento" incompleto. A ferramenta atribui, então, o valor "CF20A" para o campo "código do equipamento" para todos os outros cinco por centos dos registros que estão campo "código do equipamento" incompleto.

Quando o grau de crença não é tão elevado, a ferramenta sugere a um valor para o campo "código do equipamento" e deixa a decisão para o administrador.

6.1 – Sistema Especialista de Contexto (SEC)

A ferramenta que foi descrita acima, recebeu o nome de SEC. A SEC define o grau de crença (μ_1) e também o grau de descrença (μ_2) com base nos dados contidos em um banco de dados inconsistente, como mostra a **Figura 6.1**.

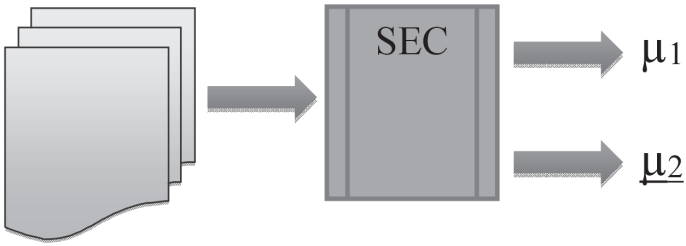

Figura 6.1: Esquema de um Sistema Especialista de Contexto.

O SEC foi utilizado para identificar as chaves primárias e também as chaves estrangeiras de uma base de dados. Após a identificação, o SEC classificou as inconsistências e eliminou as incompletudes. As inconsistências que não foram tratadas pelo SEC, são informadas ao administrador, juntamente com sugestões da ferramenta, para que o mesmo possa tomar a decisão de eliminar a inconsistência ou ainda eliminar o registro da base de dados.

6.2 – Aplicação do Sistema Especialista de Contexto (SEC)

Como exemplo didático, foi utilizado um arquivo para identificar todos os registros que continham incompletudes, ou seja, registros que continham campos incompletos. Após identificar todos os regis-

tros com incompletudes, criou-se um novo arquivo, contendo apenas os registros do arquivo original que continham incompletudes. A esse arquivo denominou-se **AUXILIAR1**. Também foi gerado um arquivo denominado **AUXILAR2**, contendo todos os registros que não continha incompletudes (Figura 6.2).

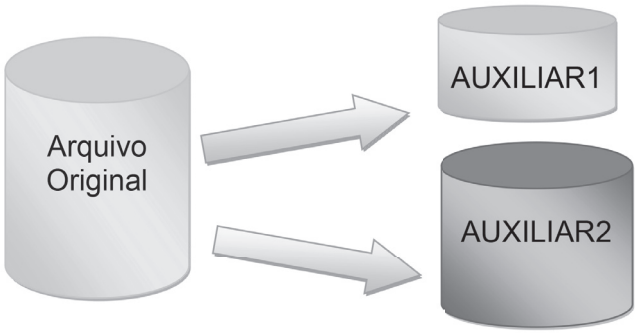

Figura 6.2: Geração dos arquivos AUXILIAR1 (contendo os registros contendo incompletudes) e AUXILIAR2 (contendo os registros sem completudes) a partir do Arquivo Original

Após separar os registros contendo incompletudes dos demais registros, foi selecionado um registro contendo incompletudes e identificou-se o campo incompleto, no caso desse exemplo o Campo3, como mostra a Figura 6.3.

CAMPO 1	CAMPO 2	CAMPO 3	CAMPO 4	CAMPO 5	CAMPO 6
5678SE4	10 A	NULL	HPS300	22/05/13	RURAL

Figura 6.3: Registro com o Campo 3 incompleto.

Identificado o registro e o respectivo campo incompleto, foi realizada uma análise de *cluster* (agrupamentos), que consiste em dividir

um grupo de dados em subgrupos homogêneos de acordo com a similaridade existente entre os elementos desse grupo (Guidini, Marilene Bertuol, et al.,2008). Nesse exemplo, a análise de *cluster* foi usada para agrupar os registros do arquivo AUXILIAR2, que tem valores semelhantes para o Campo 3. A Figura 6.4 ilustra o resultado desse agrupamento, sendo identificados seis agrupamentos.

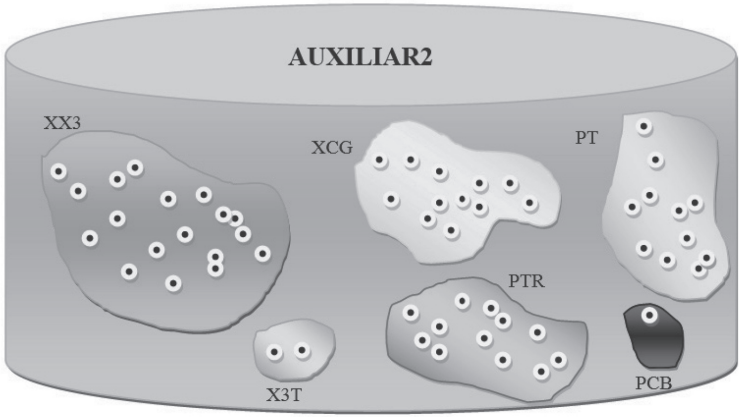

Figura 6.4: Agrupamentos gerados a partir do arquivo AUXILAR2: XX3 é o *cluster* contendo registros cujo Campo 3 é igual a XX3, PT é o *cluster* contendo registros cujo Campo 3 é igual a PT, XCG é *cluster* contendo registros cujo Campo 3 é igual à XCG, X3T é o *cluster* contendo os registros cujo Campo 3 é igual a X3T, PTR é o *cluster* contendo os registros cujo Campo 3 é igual à PTR e finalmente, PCB é o *cluster* contendo os registros cujo Campo 3 é igual a PCB.

Para realizar esse agrupamento, utilizou-se o método *K-means cluster*, que é um método que utiliza critério centróide, ou seja, depois de definir a quantidade de *clusters* que serão formados a partir de um conjunto de dados, identificam-se os elementos que serão o centro de cada *cluster*. Esses pontos que representam os centros de cada *cluster* estão homogeneamente distribuídos no conjunto de dados (Guidini, Marilene Bertuol, et al.,2008). Para o caso tratado nesse exemplo, a quantidade de *cluster* foi definida de acordo com os possíveis valores

que poderiam ser atribuídos ao campo incompleto, que resultou num total de seis possíveis valores, e conseqüentemente, definiu-se como seis a quantidade de *cluster* (Figura 6.4).

Após definir a quantidade de agrupamentos, foi definido quem seria o centro de cada um dos *clusters*. Para isso, a cada registro selecionado, verificou-se a qual agrupamento o mesmo pertencia. Se o valor do campo 3 fosse igual a "PT", o registro seria incluído no agrupamento PT, se o valor fosse "PCB", o registro seria incluído no agrupamento PCB e assim sucessivamente.

O primeiro elemento a pertencer a cada grupo, será automaticamente o centro de tal grupo. Quando é inserido o próximo elemento em um grupo, calcula-se a média dos valores dos elementos pertencentes ao grupo. O registro que tiver o valor mais próximo da média será o novo centro do grupo. Esse processo se repete até que todos os elementos sejam inseridos, como mostra o algoritmo na Figura 6.5.

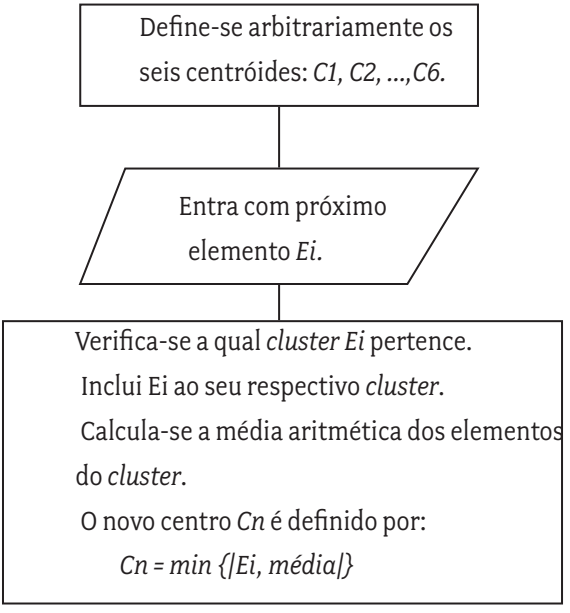

Figura 6.5: Algoritmo para incluir os elementos e determinar os centros dos clusters.

Até aqui os elementos foram agrupados tendo como parâmetro o valor do campo 3. Mas o objetivo desse exemplo é identificar os valores do campo 3 para os registros do arquivo AUXILIAR1, ou seja, os registros cujo campo 3 é incompleto. Para isso utilizou-se a LPA2v para determinar os graus de certeza e de incerteza de que determinado registro pertença a um determinado grupo e por conseqüência tenha o valor para o campo 3 igual ao valor do campo3 do centro de tal grupo.

Para determinar se um registro incompleto tem "PT" como valor para o seu campo 3, verifica-se quantos campos o registro incompleto tem valores iguais para o campo do registro central do grupo "PT". A essa quantidade é denominada o grau de crença de que tal registro pertença ao grupo "PT" e é denominado de μ_1. Se essa quantidade for igual a 5, $\mu_1 = 1$. Se for igual a 4, $\mu_1 = 0{,}8$ e assim sucessivamente, até quando a quantidade for igual a zero e nesse caso $\mu_1 = 0$.

Para determinar o grau de descrença (μ_2) calculou-se o μ_1 para o mesmo elemento incompleto em relação a todos os outros grupos, e definiu-se μ_1 como o máximo entre os μ_1 calculados, exceto o para o grupo "PT".

Após obter os graus de crença (μ_1) e descrença (μ_2) é possível determinar a qual região do QUPC o ponto cujas coordenadas são μ_1 e μ_2, pertence.

Para ilustrar admita que um determinado registro, contendo incompletudes no campo 3, tenha os outros cinco campos congruentes ao núcleo do agrupamento "PT" e não tenha nenhum campo congruente aos demais núcleos. Nesse caso μ_1, que é o grau de crença em que tal registro pertença a agrupamento "PT" é máximo, ou seja, igual a um. Já μ_2, que é o grau de descrença e é igual ao máximo entre os μ_1 do mesmo registro referente aos demais grupos, e que nesse caso é mínimo, ou seja, é igual à zero. Dessa forma, utilizando as equações para determinar os graus de certeza (GC) e incerteza (GI), das por (Martins, 2003), ficariam:

$$GC = \mu_1 - \mu_2 \qquad GI = \mu_1 + \mu_2 - 1$$
$$= 1 - 0 \qquad = 0 + 1 - 1$$
$$GC = 1 \qquad GI = 0$$

Os resultados das equações acima mostram que o grau de certeza, para esse caso, é de um. Já o Grau de Incerteza é igual a zero. Então o administrador do repositório pode decidir, com uma margem excelente de confiança, completar o campo incompleto com o "valor" "PT", visto que tem um alto grau de certeza de que o valor deva ser "PT" e um baixo grau de incerteza de que o valor do campo incompleto deva ser "PT".

Também é possível verificar que o ponto cujas coordenadas são $\mu_1 = 1$ e $\mu_2 = 0$ pertence à região de Certeza no QUPC mostrado na Figura 5.15.

Se os valores para μ_1 e μ_2 fossem, respectivamente, zero e um, os valores de GC e GI seriam, respectivamente, um negativo e zero. Como o grau de certeza de que o valor do campo incompleto deva ser "PT" é muito baixo, mas também o grau de incerteza de que esse valor deva ser igual a "PT" é igualmente baixo, o administrador pode interpretar que a possibilidade de que o registro pertença ao agrupamento "PT" é de 0% e que a possibilidade de tal registro não pertença ao agrupamento é de 100%. Aqui também o administrador pode decidir, com uma margem excelente de confiança, não completar o campo incompleto com o valor "PT".

Novamente é possível verificar que o ponto cujas coordenadas sejam as dadas acima, pertence à região de Total Falsidade, confirmando assim a opção de descartar a possibilidade de completar o registro com o valor "PT".

Para um registro que tenha três campos, exceto o campo 3, congruentes ao núcleo do agrupamento "PT", portanto tenha um $\mu_1 = 0{,}6$ e tenha no máximo 2 campos congruentes a outro núcleo de um agrupamento diferente do agrupamento "PT", ou seja, tem um $\mu_2 = 0{,}4$, os valores de GC e GI são, respectivamente, 0,2 e 0 (zero). Com esses valores, o administrador não terá uma boa margem de confiança para completar o registro com o valor "PT" e nem para descartar que o valor possa ser "PT", visto que o grau de certeza de que o campo incompleto deva ser completado com o valor "PT" é muito baixo.

Uma justificativa para a afirmação acima, é o fato de que um ponto com as coordenadas $\mu_1 = 0{,}6$ e $\mu_2 = 0{,}4$, situa-se justamen-

te na área Totalmente Paracompleta e que, portanto são necessárias mais informações para que o administrador tome uma decisão.

Também é possível que um registro contendo incompletudes no campo 3, tenha cinco campos congruentes com o núcleo do agrupamento "PT", portanto, tendo um $\mu_1 = 1$, mas também tenha um $\mu_2 = 1$, ou seja, também tenha cinco campos congruentes a um núcleo de outro agrupamento diferente do agrupamento "PT". Assim esse registro teria os graus de certeza e de incerteza iguais a dois e um, respectivamente, ou seja, o grau de certeza de que o campo incompleto deva receber o valor "PT" é o máximo, porém o grau de incerteza de que tal campo deva receber o valor "PT" também é máximo. Dessa forma, o administrador tem evidências muito fortes de que o valor correto para o campo incompleto deva ser "PT", mas também tem evidências, igualmente fortes, de que o valor do campo incompleto não deva ser "PT", o que caracteriza uma inconsistência. Dessa forma o administrador não poderá decidir se o valor para o campo incompleto deve ser ou não "PT".

Novamente a conclusão acima encontra amparo no QUPC, pois um ponto com as coordenadas um e um, se localiza na região Totalmente Inconsistente do mesmo.

As decisões tomadas acima podem ser sintetizadas pelo algoritmo mostrado na Figura 6.6.

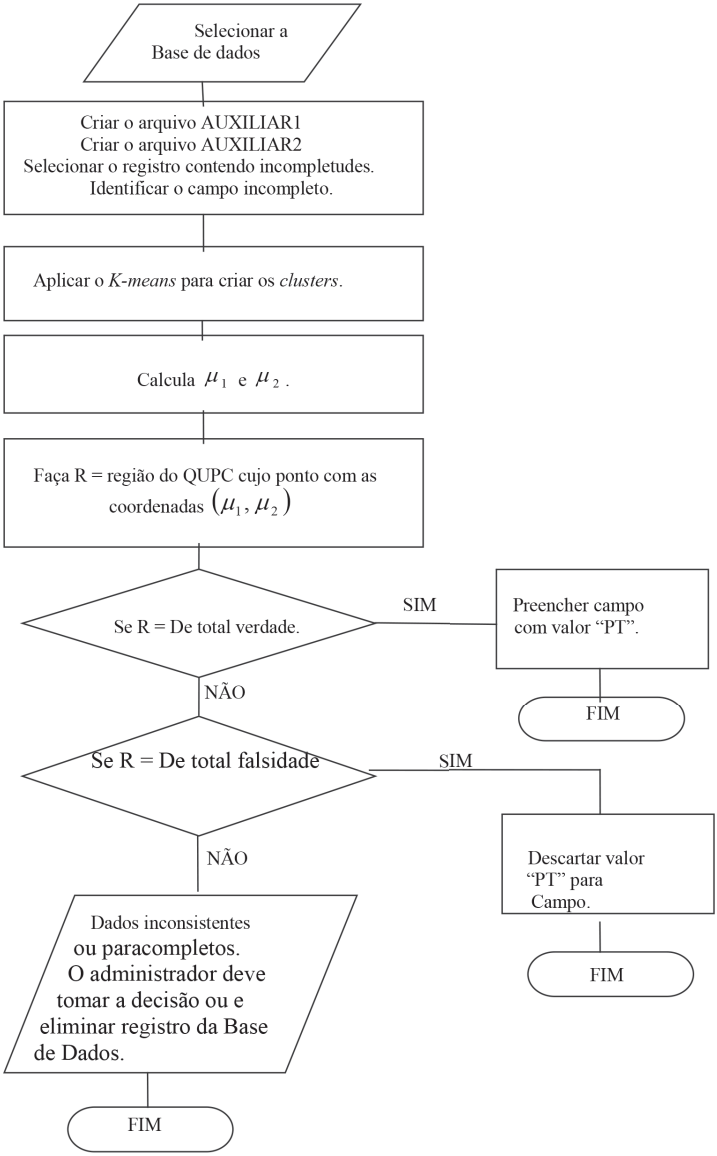

Figura 6.6: Fluxograma para o programa de tomada de decisão, utilizando a LPA2v.

REFERÊNCIAS

(**Dougherty, Kohavi e Sahami, 1995**) Dougherty, J., Kohavi, R. e Sahami, M., Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, editors, Proc. Twelfth International Conference on Machine Learning, paginas 194–202. Morgan Kaufmann, 1995.

(**Martins, 2003**) Martins, H. G., A Lógica Paraconsistente Anotada de Quatro Valores – LPA4v Aplicada em Sistema de Raciocínio Baseado em Casos para o Restabelecimento de Subestações Elétricas, Tese de doutorado na UNIFEI – Universidade Federal de Itajubá, Itajubá – MG, Brasil.

(**Moraes, 2006**) Moraes, C. H. V., Aplicação de Técnicas Inteligentes no Auxílio à Operação dos Centros de Controle, Tese de doutorado na UNIFEI – Universidade Federal de Itajubá, Itajubá – MG, Brasil, 2006.
(**Spiegel, 2008**) Spiegel, M. R., Estatística, Teoria Elementar da Probabilidade, 3ª edição, editora Pearson Makron Books, São Paulo – SP, Brasil.

(**Lambert-Torres, Costa e Moraes, 2008**) Lambert-Torres, G., Costa, C. I. A., Neto, M. C. B., Andrade, G. C. C., Moraes, C. H. V., Decision-Making using a Paraconsistent Analytic Hierarchy Process, IEEE 978, 2008.
(**Abe, 2010**). Abe, J. M., Paraconsistent Logics and Apllications, IEEE, 978-1-4244, São Paulo – SP, Brasil.

(**Borbon, Guido e Vieira, 2009**) Borbon Jr, S., Guido, R. C., Vieira, L. S., A Neural-Network Approach for Speech Features Classification Base don Paraconsistent Logic, IEEE, 2009, Belo Horizonte – MG, Brasil.

(**Da Costa et al., 1999**) Da Costa, N. C. A. et al., 1999, Lógica Paraconsistente Aplicada. São Paulo – SP. Editora Atlas, 1999.

(**Rissino, 2009**) Rissino, S. D., Metodologia de Avaliação da Relevância de Atributos em Grandes Bases de Dados Incompletas Utilizando Conjuntos Aproximados e Lógica Paraconsistente – Tese de doutorado, Universidade Federal de Itajubá – UNIFEI, Itajubá – MG, Brasil, 2009.

(**Alencar Filho, 2002**) Alencar Filho, E., Iniciação à Lógica Matemática, Editora Nobel, 1ª edição, São Paulo – SP, Brasil, 2002.

(**Martins et al., 2007**) Martins, Helga Gonzaga; Lambert-Torres, Germano; Potin, Luiz Franscico. A Lógica Paraconsistente Anotada. Santos-SP. Comunnicar Editora. ISBN 978-85-99561-31-7. 2007.

(**Carvalho et al., 2003**) Carvalho, F. R; Brunstein, I; Abe, J. M., Um Estudo de Tomada de Decisão Baseado em Lógica Paraconsistente Anotada: Avaliação do Projeto de uma Fábrica, Revista Pesquisa e Desenvolvimento Engenharia de Produção, n.1, p. 47-62, dez. 2003.

(**Lambert et all, 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.

(**Souza, Vendrusculo e Melo, 2000**) Souza, Marcia Izabel Fugisawa, Vendrusculo, Laurimar Gonçalves, Melo, Geane Cristina, Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core, Ci. Inf., Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000.

(**Cherchiglia e outros, 2007**) Cherchiglia, Mariângela Leal, A construção da base de dados nacional em Terapia Renal Substitutiva (TRS) centrada no indivíduo: aplicação do método de linkage determinístico-probabilístico, *R. bras. Est. Pop.*, São Paulo, v. 24, n. 1, p. 163-167, jan./jun. 2007

(Da Costa, 93) Da Costa, Newton C. A., *"Sistemas Formais Inconsistentes"*, apresentado originalmente como tese (cátedra) em 1963 - Faculdade de Filosofia, Ciências e Letras da Universidade do Paraná, Curitiba, Paraná, 1993.

(**Fayyad et all, 1996**) U. M.; PiatetskyShapiro, G.; Smyth, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, 1996.

(**Politi, 2005**) Politi, J. Implementação de uma Metodologia Para Mineração de Dados Aplicada ao Estudo de Núcleos Convectivos, Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, INPE, São José dos Campos, SP – Brasil, 03 de fevereiro de 2005.

(**Soares et all, 2002**) Soares, S. R. A., Bernardes, R. S., Cordeiro Netto, O. de M. Relações entre saneamento, saúde pública e meio ambiente: elementos para formulação de um modelo de planejamento, Scielo, 18(6), 1713-1724, Rio de Janeiro, Brasil, Nov-dez de 2002.

(**Dougherty, Kohavi e Sahami, 1995**) Dougherty, J., Kohavi, R. e Sahami, M., Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, editors, Proc. Twelfth International Conference on Machine Learning, paginas 194–202. Morgan Kaufmann, 1995.

(**Date, 2003**) DATE, C. J.. INTRODUÇÃO A SISTEMAS DE BANCOS DE DADOS. 8. ed. Rio de Janeiro: Elsevier, 2003.

(**Elmasri e Navathe, 2011**) ELMASRI, Ramez; NAVATHE, Shamkant B.. SISTEMAS DE BANCO DE DADOS. 6. ed. São Paulo: Addison Wesley, 2011.

(**Heuser, 2001**) Carlos A. (2001) "Projeto de Banco de Dados", Séries Livros Didáticos, 4ºedição, Editora Sagra Luzzatto. Instituto de Informática da UFRGS.

(**Date, 2000**) Date, C. J., Introdução a Sistemas de Banco de Dados, 7ª Edição. Editora Campus, 2000.

(**Ramalrishna et AL, 2002**) Ramakrishna, R.; Gehrke, J.; Database Management Systems, Editora McGraw-Hill Science, 2002.

(**Boscarioli. Clodis, et al., 2006**) Boscarioli, Clodis, et al. "Uma reflexão sobre banco de dados orientados a objetos." *IV CONGED* (2006).

(**De Souza, 2012**) DE SOUZA, Terezinha Batista; CATARINO, Maria Elizabete; DOS SANTOS, Paulo Cesar. Metadados: catalogando dados na Internet. **Transinformação**, v. 9, n. 2, 2012.

(**Moraes, 2001**) Moraes, Ilara Hämmerli Sozzi de, and Silvia Regina Fontoura Rangel dos Santos. "Informações para a gestão do SUS: necessidades e perspectivas."*Informe epidemiológico do SUS* 10.1 (2001): 49-56.

(**de Morais, Alana M., et al., 2010**) de Morais, Alana M., et al. "Tomada de Decisão aplicada à Inteligência Artificial em Serious Games voltados para Saúde." (2010).

(**Altran, 2010**) ALTRAN, ALESSANDRA BONATO. *Sistema Inteligente Para Previsão de Carga Multinodal em Sistemas Elétricos de Potência*. Diss. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Elétrica, UNESP, Campus de Ilha Solteira, 2010.

(**Lambert, et al., 2008**) Lambert-Torres, Germano, et al. "Decision-making using a Paraconsistent analytic hierarchy process." *Electricity Market, 2008. EEM 2008. 5th International Conference on European*. IEEE, 2008.

(**Teixeira, 1997**)Teixeira, Cenidalva Miranda de Sousa, and Ulrich Schiel. "A Internet e seu impacto nos processos de recuperação da informação." *Ciência da Informação* 26.1 (1997).

INSTITUTO
FEDERAL
Sul de Minas Gerais