

CENTRO UNIVERSITÁRIO DA FEI
RODOLFO ARTUR FUSSEK JÚNIOR

**DESENVOLVIMENTO DE UM CLASSIFICADOR AUTOMÁTICO DE PESSOAS
PARA SISTEMAS DE AUTOMAÇÃO RESIDENCIAL INTELIGENTE**

São Bernardo do Campo

2010

RODOLFO ARTUR FUSSEK JÚNIOR

**DESENVOLVIMENTO DE UM CLASSIFICADOR AUTOMÁTICO DE PESSOAS
PARA SISTEMAS DE AUTOMAÇÃO RESIDENCIAL INTELIGENTE**

Dissertação de Mestrado apresentada ao Centro
Universitário da FEI como parte dos requisitos
necessários para obtenção do título de
Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Flavio Tonidandel

São Bernardo do Campo

2010

A Deus, por me permitir viver esta experiênciã,
e a todos os que me apoiaram nesta empreitada.

AGRADECIMENTOS

Agradeço a todos aqueles que, de alguma forma, contribuíram para que este trabalho se tornasse uma realidade.

A Deus, por dar-me sabedoria e força necessárias para buscar a realização dos meus sonhos, mesmo em meio às adversidades.

Aos meus pais, Rodolpho e Vera, que infelizmente não puderam acompanhar este momento ao meu lado, mas que foram os principais responsáveis pelo seu acontecimento.

Aos meus irmãos, Carlos e Kátia, por terem sido meu apoio nos momentos mais difíceis de minha vida.

À minha noiva, Karen, pelo carinho, amor e apoio nestes últimos três anos de minha vida.

Ao meu orientador, professor Dr. Flavio Tonidandel, pela sua paciente orientação que permitiu a transformação de idéias em fatos, e dos resultados neste trabalho.

Aos professores Dr. Plínio Thomaz Aquino Junior e Dr. Paulo Sérgio Silva Rodrigues, pelos conselhos e correções, apresentados em minha banca de qualificação, que auxiliaram na conclusão desta dissertação.

A todos os professores e colegas do mestrado, pelos mistos de ensinamentos e momentos de descontração, e pela amizade, que tornaram esta complexa etapa mais divertida.

A todos os amigos, por propiciarem grandes experiências, nas conversas, conselhos, discussões, churrascos, cervejadas, entre outros; momentos, estes, não menos importantes do que qualquer outro na vida.

*Um raciocínio lógico leva você de A a B.
A imaginação leva você a qualquer lugar que quiser.*

Albert Einstein

RESUMO

Atualmente, já foram desenvolvidos diversos sistemas de automação residencial, entre os quais alguns se encontram disponíveis comercialmente. Visando o conforto no interior da residência, é coerente afirmar que os sistemas mais eficientes são aqueles que se adaptam ao comportamento dos habitantes, no lugar dos habitantes terem de se adaptar ao sistema. Em trabalhos anteriores, foram desenvolvidos sistemas com este objetivo, porém considerando a existência de um único habitante na residência. A proposta deste trabalho é apresentar algumas técnicas possíveis de identificação de habitantes, em uma residência, para a adaptação dos sistemas de automação voltados a um único habitante e, posteriormente, desenvolver um classificador automático de habitantes a partir das combinações de suas roupas e objetos pessoais. Para a identificação dos itens, foi considerado o uso de identificadores por rádio frequência (RFID) anexados a estes, sendo que esta tecnologia permite a identificação com o mínimo de adaptação do habitante. Foi utilizado raciocínio baseado em casos (RBC) como técnica para efetuar a classificação, a partir do correto tratamento das combinações de objetos detectadas pelo sistema.

Palavras chave: Automação Residencial, Domótica Inteligente, RFID, Raciocínio Baseado em Casos.

ABSTRACT

Nowadays, it has been developed many residential automation systems, and some of them are already commercially available. Aiming for the comfort inside home, it is coherent to affirm that the most efficient systems are those that adapt themselves to the inhabitant behavior, and not the opposite. In past works, it was developed systems with this objective, but considering the existence of only one inhabitant at home. This work's purpose is to present some possible techniques to identify a house's inhabitants, for adapting individual automation systems and, after, to develop an automatic inhabitant's classifier from their clothes and their personal objects. For items identification, it was considered the use of RFID identifiers attached to them, because this technologic allows the identification with few inhabitant adaptations, and case-based reasoning was used as the technique that makes classification from the correct treatment of objects combinations detected by the system.

Keywords: Residential Automation, Intelligent Domotics, RFID, Case-Based Reasoning.

LISTA DE TABELAS

Tabela 3.1 – Casos armazenados na BC.....	46
Tabela 3.2 – Pesos de objetos armazenados na tabela de pesos.....	47
Tabela 4.1 – Exemplo de criação de um novo caso a partir de um habitante e uma combinação, selecionados aleatoriamente.....	57
Tabela 5.1 – Quantidade de objetos por habitante.....	68
Tabela 5.2 – Banco de dados de objetos por habitante.....	69
Tabela 5.3 – Banco de dados de combinações possíveis de objetos.....	70
Tabela 5.4 – Segmento inicial dos 32 primeiros casos da BC gerada pelo software semeador de casos.....	71
Tabela 5.5 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais.....	94
Tabela 5.6 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais.....	95
Tabela 5.7 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos.....	96
Tabela 5.8 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto.....	97
Tabela 5.9 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais.....	98
Tabela 5.10 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais.....	99
Tabela 5.11 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos.....	100
Tabela 5.12 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto.....	100
Tabela 5.13 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais.....	102
Tabela 5.14 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais.....	103
Tabela 5.15 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos.....	104
Tabela 5.16 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto.....	104
Tabela 5.17 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais.....	105

Tabela 5.18 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais.....	106
Tabela 5.19 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos.....	107
Tabela 5.20 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto.....	108

LISTA DE FIGURAS

Figura 2.1 – Componentes lógicos de um leitor RFID, adaptada de (GLOVER, BHATT, 2007).....	28
Figura 2.2 – Arquitetura RFID com a aplicação exposta diretamente aos leitores, adaptada de (GLOVER, BHATT, 2007).....	28
Figura 2.3 – Componentes do middleware RFID, adaptada de (GLOVER, BHATT, 2007).....	29
Figura 2.4 - Arquitetura RFID com a intermediação do middleware RFID entre os leitores e a aplicação, adaptada de (GLOVER, BHATT, 2007).....	30
Figura 2.5 – Arquitetura do Sistema ABC, adaptada de (TONIDANDEL; TAKIUCHI; MELO, 2004).....	38
Figura 2.6 – Arquitetura do Sistema ABC+, adaptada de (SGARBI, 2007).....	39
Figura 3.1 – Arquitetura do sistema o CAPBC.....	42
Figura 3.2 – Algoritmo de evento de entrada do classificador RBC.....	44
Figura 3.3 – Exemplo de reclassificação de casos a partir da entrada de um novo caso.....	45
Figura 4.1 – Tela do software desenvolvido para gerar o banco de dados de objetos dos habitantes.....	53
Figura 4.2 – Tela do software desenvolvido para gerar combinações de objetos possíveis.....	54
Figura 4.3 – Algoritmo do semeador de casos a partir das tabelas de objetos e combinações.....	55
Figura 4.4 – Algoritmo para substituição de uma porcentagem P de objetos por itens emprestados, em um caso C.....	56
Figura 4.5 – Algoritmo para concatenação de uma porcentagem P de combinações com seus casos subseqüentes, de outros habitantes.....	58
Figura 4.6 – Algoritmo para inserção de objetos estranhos a um caso C.....	59
Figura 4.7 – Tela do software semeador da base de casos.....	60
Figura 4.8 – Tela do software classificador de casos.....	61
Figura 4.9 – Tela do software avaliador de classificações.....	64
Figura 4.10 – Fluxograma do software avaliador de classificações.....	65
Figura 4.11 – Ilustração de um conjunto de casos com sua classificação avaliada.....	66
Figura 5.1 – Gráfico do tempo de classificação de cada caso de acordo com o critério de similaridade.....	72
Figura 5.2 – Gráficos da relação entre o número do caso classificado e a quantidade de habitantes gerada pelo CAPBC e do valor de similaridade mínima e a qualidade das classificações obtida, utilizando-se a BC original.....	74
Figura 5.3 – Gráficos da relação entre o número do caso classificado e a quantidade de habitantes gerada pelo CAPBC e do incremento no peso dos objetos e a qualidade das classificações obtida, utilizando-se a BC original.....	76
Figura 5.4 – Gráficos da relação entre o valor de similaridade mínima utilizado e a qualidade nas classificações, utilizando-se bases de casos com objetos permutados entre habitantes.....	78
Figura 5.5 – Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases de casos com objetos permutados entre habitantes.....	80
Figura 5.6 – Gráficos da relação entre o valor de similaridade mínima utilizado e a qualidade nas classificações, utilizando-se bases com casos concatenados.....	84

Figura 5.7 – Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases com casos concatenados.....	86
Figura 5.8 – Gráficos da relação entre o valor de similaridade mínima utilizado e a qualidade nas classificações, utilizando-se bases de casos com objetos novos ou estranhos.....	90
Figura 5.9 – Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases de casos com objetos novos ou estranhos.....	92

LISTA DE ABREVIATURAS

ABC – Automação Baseada em Comportamento

API – Application Programming Interface

BC – Base de Casos

CAPBC – Classificador Automático de Pessoas Baseado em Casos

EPC – Electronic Product Code

IA – Inteligência Artificial

ISO – International Organization for Standardization

LDA – Linear Discriminant Analysis

PCA – Principal Component Analysis

RBC – Raciocínio Baseado em Casos

RF – Rádio Frequência

RFID – Radio Frequency Identification

SUMÁRIO

1 INTRODUÇÃO	16
1.1 Objetivo	19
1.2 Motivação	19
1.3 Organização do trabalho	20
2 CONCEITOS RELACIONADOS	21
2.1 Introdução	21
2.2 Sistemas de identificação de pessoas	21
2.2.1 Identificação por reconhecimento biométrico.....	21
2.2.1.1 Reconhecimento de impressões digitais.....	22
2.2.1.2 Reconhecimento de voz.....	22
2.2.1.3 Reconhecimento de faces.....	23
2.2.1.4 Reconhecimento de retina e íris.....	23
2.2.1.5 Uso do reconhecimento biométrico em um sistema de automação residencial inteligente.....	24
2.2.2 Identificação pelo reconhecimento de padrões no caminhar.....	24
2.3 Identificação por rádio frequência (RFID)	25
2.3.1 Definições.....	25
2.3.2 Arquitetura RFID.....	26
2.3.2.1 Identificadores RFID.....	26
2.3.2.2 Leitores RFID.....	27
2.3.2.3 Middleware RFID.....	29
2.3.3 Principais aplicações.....	30
2.3.4 Uso de RFID no sistema proposto neste trabalho.....	31
2.4 Raciocínio baseado em casos (RBC)	31
2.4.1 Introdução ao RBC.....	32
2.4.2 Representação do conhecimento em RBC.....	33
2.4.3 Resgate de casos e medida de similaridade.....	34
2.4.3.1 Conceito de similaridade.....	35
2.4.3.2 Julgamento de similaridade.....	35
2.4.4 Adaptação de casos.....	36
2.4.5 Aprendizado.....	36
2.4.6 Aplicação do RBC no sistema proposto neste trabalho.....	37
2.5 Automação Residencial Inteligente	37
2.5.1 Automação Baseada em Comportamento (ABC).....	37
2.5.2 Evolução do Sistema ABC (ABC+).....	39
3 CLASSIFICADOR AUTOMÁTICO DE PESSOAS BASEADO EM CASOS (CAPBC)	41
3.1 Funcionamento do CAPBC	42

3.2 Algoritmo classificador	43
3.3 Base de casos	46
3.4 Pesos de objetos	47
3.5 Critérios de classificação e atualização de parâmetros	47
3.5.1 Número de objetos iguais.....	48
3.5.2 Porcentagem de objetos iguais sem a utilização de pesos.....	48
3.5.3 Porcentagem de objetos iguais com a utilização de pesos.....	49
3.6 Observações sobre os métodos e cálculos utilizados para classificação dos habitantes	50
4 SIMULAÇÃO DO SISTEMA CAPBC	51
4.1 Descrição dos simuladores	51
4.1.1 Semeador de casos.....	51
4.1.1.1 Simulação do empréstimo de objetos entre habitantes.....	56
4.1.1.2 Simulação da leitura de dois habitantes simultaneamente.....	57
4.1.1.3 Simulação da adição de objetos novos e estranhos aos casos.....	58
4.1.2 Classificador de casos.....	60
4.1.2.1 Critério de similaridade.....	61
4.1.2.2 Valor de similaridade mínima.....	62
4.1.2.3 Número de casos a ser classificado.....	62
4.1.2.4 Uso, ou não, de pesos na classificação.....	62
4.1.2.5 Critérios para incremento dos pesos.....	62
4.1.2.6 Forma de classificação dos casos com o uso de pesos.....	63
4.1.3 Avaliador de classificações.....	63
4.2 Definições utilizadas na avaliação de resultados	64
5 TESTES REALIZADOS	67
5.1 Base de casos semeada para as simulações	67
5.2 Avaliação do tempo de classificação	71
5.3 Convergência do número de classificações para o número correto de habitantes	73
5.3.1 Classificação variando-se os valores de similaridade mínima.....	73
5.3.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos.....	75
5.3.3 Conclusões sobre a quantidade de habitantes gerados pelo CAPBC.....	77
5.4 Classificação de base de casos com objetos permutados entre habitantes	77
5.4.1 Classificação variando-se os valores de similaridade mínima, com objetos permutados.....	77
5.4.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com objetos permutados.....	79
5.4.3 Conclusões sobre a classificação de uma base de casos com objetos permutados entre habitantes.....	82
5.5 Classificação de base com casos concatenados	82
5.5.1 Classificação variando-se os valores de similaridade mínima, com casos concatenados.....	83

5.5.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com casos concatenados.....	85
5.5.3 Conclusões sobre a classificação de uma base com casos de habitantes diferentes concatenados.....	87
5.6 Classificação de base de casos com o acréscimo de objetos novos ou estranhos.....	88
5.6.1 Classificação variando-se os valores de similaridade mínima, com objetos novos ou estranhos.....	88
5.6.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com objetos novos ou estranhos.....	91
5.6.3 Conclusões sobre a classificação de uma base com a existência de objetos novos ou estranhos nas combinações de itens.....	93
5.7 Avaliação da qualidade das classificações.....	93
5.7.1 Quantidade de habitantes reconhecidos pelo CAPBC.....	94
5.7.2 Quantidade de classificações corretas obtidas pelo CAPBC.....	97
5.7.3 Quantidade de classificações incompletas obtidas pelo CAPBC.....	101
5.7.4 Quantidade de classificações erradas obtidas pelo CAPBC.....	105
5.8 Síntese do capítulo.....	108
6 CONCLUSÃO.....	109
6.1 Trabalhos Futuros.....	111
6.1.1 Teste do sistema com outras bases de casos, com características e quantidades de habitantes diferentes.....	111
6.1.2 Uso de outras informações do sistema RFID ou sistema de automação na classificação.....	112
6.1.3 Implementação de uma política de reclassificação de casos.....	112
6.1.4 Estudo sobre a possibilidade de apenas um subconjunto da BC na classificação de novos casos.....	113
6.1.5 Teste do sistema em uma residência real.....	113
REFERÊNCIAS BIBLIOGRÁFICAS.....	114
APÊNDICE 1.....	117
APÊNDICE 2.....	118

1 INTRODUÇÃO

A automação residencial vem tornando-se, com o passar dos anos, um assunto de interesse de um número cada vez maior de pessoas. Com as tecnologias que, nos dias atuais, já permitem um nível razoável de integração, a idéia de ter o controle sobre certos equipamentos e algum nível de automação, visando comodidade e conforto do usuário, torna-se cada vez mais real.

A queda dos preços destes equipamentos já atrai também os consumidores da classe média, que hoje podem ter acesso a, por exemplo, aparelhos de telefonia celular de última geração, computadores, redes de comunicação sem fio, conexão de internet de banda larga, entre outros.

A aplicação destes recursos tecnológicos na automação de recursos básicos de uma residência, como controle de iluminação, persianas, ar-condicionado, sistemas de áudio e vídeo, segurança, já é uma realidade nas classes mais altas da sociedade. A tendência de atingir as classes mais baixas, a médio prazo, é cada vez maior (MURATORI, 2005).

Baseado nessas considerações, torna-se desejável o desenvolvimento destes sistemas, de forma a poderem, em primeiro lugar, serem de fácil utilização pelos habitantes da residência (a complexidade dos controles dos sistemas presentes na casa com a automação não pode ser maior do que sem o uso dela) e, com o passar do tempo, cada vez mais adaptativos, evoluindo para o conceito de domótica inteligente, onde as regras de automação adaptam-se ao comportamento dos moradores (TONIDANDEL; TAKIUCHI; MELO, 2004). Na maioria dos sistemas ditos inteligentes, desenvolvidos na área até o presente momento, as relações entre os eventos detectados pelos sensores (presença e comportamento dos habitantes, assim como fatores como iluminação, temperatura do ambiente, etc.) e ações tomadas são previamente configuradas por um especialista (BOLZANI, 2004), não sendo o sistema capaz de adaptar-se às mudanças nas condições externas.

Sgarbi (2007) desenvolveu um sistema baseado em aprendizado de máquina capaz de adaptar-se ao comportamento de um habitante com base em dados obtidos por meio de sensores. O sistema tem a capacidade de aprender e adaptar regras para uma residência com um único morador, ou para residência onde podem ser aplicadas as mesmas ações sobre todos os moradores.

Neste trabalho são apresentados alguns possíveis métodos para identificação de pessoas em um sistema de domótica inteligente e, posteriormente, proposto e desenvolvido um classificador de habitantes de uma residência, com o objetivo de possibilitar a adaptação

de sistemas de automação residencial individuais. Existem vários meios de identificação de pessoas que poderiam ser utilizados em conjunto com estes sistemas, mas o desenvolvimento do trabalho focará em um método que minimize a necessidade de mudança nos hábitos dos habitantes.

Os sistemas de automação residencial, em sua maioria, têm como objetivo facilitar a interação dos habitantes com os itens controlados e monitorados por eles, como iluminação, climatização, entre outros. Desta forma, embora não esteja sendo definida uma métrica de sucesso para isto, pode-se considerar que quanto menos um usuário tiver de manusear controles e interfaces para executar as ações desejadas no interior de sua residência, melhor poderá ser considerado seu sistema.

Da mesma forma, um sistema de identificação de pessoas, pelo menos no interior da residência, deve exigir a menor interação direta possível por parte do usuário. Se isto não ocorrer, o habitante será obrigado a adaptar seus hábitos toda vez que desejar ser identificado pelo sistema, como por exemplo, acionar um sensor biométrico.

Também não é desejável que a automação residencial traga desconforto ao habitante por meio da invasão de sua privacidade. Sistemas que utilizam câmeras na identificação de pessoas podem provocar esse desconforto, causando a impressão de que o morador está sendo monitorado por outras pessoas, todo o tempo, dentro de sua residência.

Uma alternativa encontrada para identificação, que pode ser aplicada no interior de ambientes e que não implica, diretamente, na mudança dos hábitos do morador, ou na invasão de sua privacidade, é o uso de dispositivos de identificação por rádio frequência (RFID).

Resumidamente, RFID é um método de identificação onde códigos individuais são gravados em pequenos circuitos integrados, denominados identificadores, dotados de antenas, que são utilizadas na comunicação com os leitores RFID. Este método possui a vantagem de possibilitar a leitura à distância de um ou mais itens simultaneamente, sem que seja necessário um posicionamento específico dos identificadores (RFID JOURNAL, [200-]a), o que ocorreria, por exemplo, com códigos de barras.

Considerando-se a aplicação do RFID na identificação do habitante dentro de uma residência, poder-se-ia, por exemplo, criar um acessório para cada habitante, contendo um identificador pré-cadastrado para este. Embora esta seja uma solução para a identificação, há o inconveniente das pessoas terem de adotar o hábito de portar este acessório toda vez que quiserem ser identificadas, sendo que na ausência do mesmo o sistema de automação simplesmente não funcionará. Uma solução para este problema seria a inserção dos

identificadores nas peças de roupas dos habitantes, de forma que não tivessem de se lembrar de vestir ou portar um objeto exclusivo para sua identificação.

Alguns fabricantes de roupas, como a Benetton, já utilizam a tecnologia de RFID para rastreamento individual de itens desde a produção até a venda (RFID JOURNAL, 2003). Com a queda dos preços dos identificadores e as vantagens da aplicação desta tecnologia, cada vez mais empresas devem adotá-la para uso com seus produtos (RFID JOURNAL, 2003). Sendo assim, estas identidades RFID poderiam também ser utilizadas para identificação fora das fábricas e lojas, como, por exemplo, no interior de uma residência.

A finalidade deste trabalho é possibilitar a adaptação de sistemas de automação residencial individuais para o uso por diversos moradores, que podem ter hábitos diferentes. Para minimizar a necessidade dos usuários fazerem configurações no seu sistema, como o cadastro de suas peças de roupa para posterior identificação, no sistema desenvolvido, não é priorizado o conhecimento da identidade do habitante, mas, sim, uma classificação genérica que permita que as regras de automação criadas sejam atribuídas às pessoas corretas, independentemente de quem estas sejam. Por exemplo, em um determinado momento o sistema criou uma regra R_1 para um habitante classificado como H_1 . Após algum tempo, o sistema detectou a presença de H_1 e, assim, deverá executar a regra R_1 , mesmo sem conhecer sua identidade.

A proposta para o sistema desenvolvido é a classificação dos habitantes por meio das combinações de roupas e acessórios que estes utilizam no interior da residência; isto é possível por meio da leitura das identidades RFID presentes nos objetos. Como as combinações de objetos utilizadas pelas pessoas dificilmente são exatamente as mesmas, ou seja, estes são recombinações entre si, existe a necessidade da aplicação de alguma técnica que permita o agrupamento de combinações com determinada similaridade sob uma mesma classificação. Com isto espera-se que, quanto mais combinações o sistema conhecer, mais este seja capaz de abstrair a classificação correta de novas combinações que forem lidas pelo sistema.

Para os estudos feitos neste trabalho, os dados que seriam capturados pelo sistema de RFID, instalado na residência, foram gerados por simuladores em computador. Ou seja, os habitantes, seus itens de vestuário e objetos pessoais, bem como as combinações destes, foram criados, com base em critérios pré-estipulados, expostos no Capítulo 4 (Simulação do Sistema CAPBC), que permitiram obter dados semelhantes aos que seriam conseguidos em uma residência real. A opção pelo uso de dados simulados ocorreu devido ao processo complexo de instalação de um sistema de RFID em uma residência, de implantação dos identificadores

nas roupas e objetos dos habitantes e do tempo para coleta de dados suficientes, para permitir que fossem feitos os testes e ajustes do classificador proposto. Todo o processo de simulação do sistema é detalhado no Capítulo 4.

Para efetuar a classificação das combinações, no sistema, foi escolhida a técnica de raciocínio baseado em casos (RBC), que é uma técnica de inteligência artificial (IA) que utiliza soluções adaptadas de problemas passados na solução de novos problemas. Esta técnica é apresentada com detalhes na seção 2.4, juntamente com a justificativa para o seu uso no sistema proposto.

Nas próximas seções, são apresentados o objetivo, motivação e organização deste trabalho.

1.1 Objetivo

O objetivo deste trabalho é o desenvolvimento de um sistema simulado e automático de classificação de indivíduos, voltado à adaptação de sistemas de automação residencial individuais já existentes, com o uso de RFID e raciocínio baseado em casos.

1.2 Motivação

A motivação pela escolha deste tema baseia-se na grande utilidade que um sistema eficaz de classificação de pessoas pode ter em diferentes gamas de aplicações.

Neste trabalho, o sistema foi desenvolvido voltado à aplicação em um sistema de automação residencial inteligente, onde diferentes ações devem ser tomadas dentro da residência de acordo com o habitante identificado. Esta aplicação, por si só, já justificaria o desenvolvimento do sistema, já que trará um considerável acréscimo ao nível de conforto dos habitantes de uma residência, que poderiam sentir-se incomodados por ações tomadas por regras de automação que não condizem com seus hábitos dentro dos ambientes. Estes níveis de conforto não foram medidos, neste momento, por não serem o foco do trabalho.

Neste trabalho foi considerada apenas a classificação de indivíduos, mas um desenvolvimento futuro deste poderia contemplar a detecção dos seus perfis e comportamentos a partir dos objetos que a pessoa está vestindo ou portando. Assim, o sistema poderia ter aplicações comerciais extremamente úteis, como a oferta de produtos e serviços de interesse do tipo de indivíduo identificado, ou para fins estatísticos, quanto ao perfil de

peças que freqüentam determinados locais, a fim de adaptá-los a estas, entre outras aplicações.

1.3 Organização do trabalho

Este trabalho foi dividido da seguinte forma:

No capítulo 2 são expostos os conceitos relacionados com o sistema desenvolvido neste trabalho. Foram apresentados possíveis métodos de identificação de pessoas por meio de diferentes técnicas (biométricas e reconhecimentos de padrões). Foi exposta, também, a tecnologia de identificação por rádio-freqüência (RFID), suas vantagens e desvantagens para a aplicação proposta. Em seguida, introduziu-se a técnica de inteligência artificial (IA) de raciocínio baseado em casos (RBC), que foi utilizada para o desenvolvimento do sistema e, por fim, foram mostrados alguns trabalhos já desenvolvidos na área de automação residencial inteligente, que motivaram o desenvolvimento do classificador de habitantes.

No capítulo 3, é apresentado o sistema de classificação de pessoas baseado em casos (CAPBC) desenvolvido, explicando-se detalhadamente o seu funcionamento, suas partes e parâmetros.

No capítulo 4, são apresentados e explicados os simuladores criados para os testes da lógica proposta para o sistema.

No capítulo 5, são expostos os dados e parâmetros utilizados nas simulações realizadas nos softwares, assim como gráficos e tabelas contendo os resultados das simulações, com suas análises e conclusões.

No capítulo 6, são apresentadas as considerações finais sobre o trabalho e sugeridos alguns trabalhos futuros acerca deste tema.

2 CONCEITOS RELACIONADOS

2.1 Introdução

Uma das habilidades humanas mais marcantes é o reconhecimento de seres humanos. O reconhecimento pode ser definido como o processo que envolve percepção e associação de informações com uma combinação de conteúdos armazenados na memória humana (MACHADO; ALMEIDA, 2004).

Este capítulo visa apresentar, de forma sucinta, algumas técnicas utilizadas na identificação de pessoas, a partir de reconhecimento biométrico, reconhecimento de padrões na forma de caminhar, ou seja, aquelas que poderiam, eventualmente, ser utilizadas para diferenciar os habitantes de uma residência.

Posteriormente, é apresentada a tecnologia de identificação por rádio frequência (RFID) com suas características, arquitetura construtiva, algumas aplicações atuais e vantagens que a tornam desejável para utilização em um sistema de identificação de pessoas.

Em seguida, para possibilitar o uso desta tecnologia no sistema desenvolvido, é introduzida a técnica de raciocínio baseado em casos, com o objetivo de possibilitar a classificação dos habitantes da residência a partir da combinação de identidades RFID capturada por um leitor.

Por fim, são mostrados, de forma resumida, alguns trabalhos já desenvolvidos na área de automação residencial inteligente, que motivaram a criação do classificador de habitantes, apresentado neste trabalho.

A partir da próxima seção, são detalhados os itens expostos nesta introdução.

2.2 Sistemas de identificação de pessoas

2.2.1 Identificação por reconhecimento biométrico

No contexto dos sistemas de identificação, biometria é o termo geral para todos os procedimentos que identificam pessoas comparando características físicas individuais. Na prática, estes são reconhecimentos de impressão digital dos dedos, formato da mão, identificação de voz, de face e, menos comumente, identificação de retina ou íris (FINKENZELLER, 2003).

As seções a seguir descrevem, brevemente, alguns destes procedimentos:

2.2.1.1 Reconhecimento de impressões digitais

Por serem características particulares de cada pessoa, as impressões digitais dos dedos ou mãos podem ser utilizadas em sistemas de identificação. A criminologia utiliza procedimentos baseados neste fato para identificação de criminosos desde o início do século XX, através da comparação da impressão digital de um eventual suspeito e das marcas deixadas em objetos que possam ter sido tocadas por ele (FINKENZELLER, 2003).

Quando estas características são utilizadas na identificação de pessoas, geralmente para controle de acesso a um determinado ambiente, a impressão digital é capturada através de um leitor especial. O sistema calcula um valor a partir do padrão lido e o compara com os padrões de referência armazenados nele (KAZIENKO, 2003). Os leitores modernos requerem menos de meio segundo para reconhecer e checar uma impressão digital, o que é uma característica favorável para o objetivo de identificação de pessoas, porém, a fim de evitar fraudes, estes devem ser desenvolvidos para detectarem se o dedo (ou mão) colocado no leitor pertence ou não a uma pessoa viva (SCHMIDHAUSLER, 1995 apud FINKENZELLER, 2003).

2.2.1.2 Reconhecimento de voz

Recentemente, tornaram-se disponíveis sistemas especialistas de identificação individual baseados no reconhecimento de voz, nos quais os indivíduos falam em um microfone ligado a um computador. Este converte as palavras pronunciadas em sinais digitais, os quais são avaliados pelo software de identificação.

O objetivo da verificação é checar a suposta identidade do indivíduo a partir de sua voz. Isto é alcançado através da checagem das características da fala do indivíduo contra padrões de referência já existentes (FINKENZELLER, 2003). Se houver correspondência, alguma ação pode ser iniciada, como abertura de uma porta, por exemplo.

Este sistema é pouco confiável, uma vez que um ruído no ambiente ou até mesmo uma rouquidão por parte do usuário já estariam prejudicando sua identificação (PEARCE, HIRSCH, 2000). Desta forma, seria importante o sistema ser configurado com uma margem de erro no reconhecimento da voz da pessoa, o que diminuiria as falhas causadas por ruídos ou por pequenas alterações de voz, mas também o tornaria menos seguro.

2.2.1.3 Reconhecimento de faces

Esta técnica consiste na captura de imagens de uma face através de uma câmera conectada a um computador. A partir desta, um método possível para o reconhecimento é delimitar vários pontos do rosto, identificando-se, por exemplo, distância entre orelhas, olhos, maçãs do rosto, nariz e queixo (TURK; PENTLAND, 1991). Obtidos os parâmetros necessários e depois de normalizados os dados, o sistema busca, através de grafos em um banco de dados (com mais de uma imagem de cada indivíduo contido neste), as características necessárias para a correta identificação do indivíduo.

A grande vantagem desta técnica na identificação de pessoas está no fato de que a pessoa não precisa colocar o rosto em um local exato para a captura da imagem. Isto é um diferencial em relação a outros métodos de reconhecimento biométrico, pois tem uma aceitação mais fácil por pessoas que se utilizam dessa tecnologia (MACHADO; ALMEIDA, 2004).

Por outro lado, nesse sistema, deve-se levar em consideração o ambiente. As condições de iluminação são fundamentais para que este método de reconhecimento biométrico apresente bom desempenho (STONHAM, 1986).

2.2.1.4 Reconhecimento de retina e íris

Estes dois tipos de reconhecimento biométrico podem ser considerados entre os mais seguros na identificação de usuários (MACHADO; ALMEIDA, 2004).

Os analisadores de retina medem, usando um laser de baixa intensidade e uma câmera, o padrão de vasos sanguíneos presentes na retina.

O custo para a implantação deste método é alto e a forma de captura da imagem da retina é bastante inconveniente, já que a pessoa deve olhar fixamente para um ponto infravermelho por cerca de 5 segundos, sem piscar.

Já a técnica do reconhecimento de íris apresenta uma menor exigência na captura de imagens do que a de retina (DAUGMAN, 2001). A captura da imagem é feita através de uma câmera em preto e branco e a identificação da pessoa é realizada através de um scanner que realiza o mapeamento da íris.

Este sistema também possui o inconveniente do indivíduo ter de se posicionar frente ao leitor por alguns segundos para identificação.

2.2.1.5 Uso do reconhecimento biométrico em um sistema de automação residencial inteligente

Conforme apresentado nas seções anteriores, o reconhecimento biométrico aplicado a sistemas de automação residencial possui a vantagem de agregar segurança ao sistema, já que utiliza características únicas (e pré-cadastradas) dos indivíduos na identificação, o que é desejável principalmente nos acessos às áreas internas das moradias. Por outro lado, há diversos inconvenientes a serem observados quando é considerada a aplicação no interior da residência.

Um dos empecilhos no uso dos sistemas de identificação por biometria no interior de uma residência é o custo dos sensores e demais equipamentos necessários à identificação em todos os locais onde esta for necessária para a execução de regras de automação específicas para cada habitante.

Outro problema é o inconveniente do morador ter de interromper suas ações para posicionar-se em frente a uma câmera, falar em um microfone ou colocar alguma parte do corpo em leitores para a identificação, o que obriga a pessoa a mudar seus hábitos com o objetivo de adaptar-se ao sistema de automação.

A identificação através de câmeras, embora não necessite de um posicionamento específico para possibilitar a identificação, pode ser considerada pelos moradores da residência uma invasão de sua privacidade, já que as câmeras deverão estar todo o tempo monitorando o ambiente onde se deseja a identificação.

O método apresentado na próxima seção pode ser uma alternativa para a identificação de pessoas que minimiza a necessidade de adaptação do habitante ao sistema de automação da residência.

2.2.2 Identificação pelo reconhecimento de padrões no caminhar

O método do reconhecimento de padrões no caminhar pode ser uma alternativa aos métodos apresentados anteriormente, para uso em um sistema de automação residencial.

Neste método, a identificação do usuário é baseada no sensor de passo desenvolvido em (NASCIMENTO, 2002) e consiste no reconhecimento de padrões do caminhar de uma pessoa através de algumas características presentes neste: frequência do passo, peso, ângulos de cada pé e comprimento do passo. O sistema é composto dos seguintes componentes: malha sensora, algoritmos de passo e rede neural para classificação e identificação.

A malha sensora permite que sejam obtidos os parâmetros físicos necessários à identificação do indivíduo. O algoritmo de passo consiste no cálculo dos parâmetros de ângulos e frequência obtidos através da malha sensora. A rede neural, neste sistema, tem a função de identificar os padrões nos valores entregues pelo algoritmo de passo e reconhecer o indivíduo a partir dessas características.

Este sistema tem a vantagem de ser não invasivo, ou seja, não interfere nos hábitos do usuário para a obtenção dos dados para a identificação (ROSA et al., 2005).

Por outro lado, para implantação do sistema, deve-se levar em consideração o alto custo da para a adaptação dos ambientes onde se deseja a identificação dos indivíduos com a instalação da malha sensora. Outro inconveniente provocado pelo uso deste sistema é a perda da liberdade na escolha do tipo de piso a ser utilizado em diferentes cômodos da residência.

Considerando-se o objetivo da adaptação de um sistema de automação para diversos habitantes, o método de identificação exposto nesta seção apresenta características favoráveis. O sistema proposto neste trabalho apresenta uma alternativa a este, focando principalmente a minimização da necessidade de adaptação do habitante ao sistema de identificação, além de exigir menos alterações físicas na residência para instalação do sistema de identificação.

2.3 Identificação por Rádio Frequência (RFID)

Vistos alguns possíveis métodos de identificação de pessoas que poderiam ser utilizados no interior de uma residência, foram apresentados, nesta seção, as características e conceitos da tecnologia de RFID, suas principais aplicações e sua adaptação, visando a utilização no sistema proposto neste trabalho.

2.3.1 Definições

RFID é um termo genérico que é usado para descrever um sistema que transmite a identificação (na forma de um número de série único) de um objeto ou pessoa através de ondas de rádio. Esta pertence à categoria das tecnologias de identificação automática (Auto-ID) (RFID JOURNAL, [200-]a), que inclui, também, códigos de barra, leitores ópticos de caracteres e algumas tecnologias de biometria, como leitores de retina. O RFID vem sendo utilizado para reduzir o tempo e o trabalho necessário para entrada de dados manualmente e para melhorar a precisão dos dados.

2.3.2 Arquitetura RFID

Para propósitos computacionais, uma arquitetura pode ser definida como uma decomposição de um determinado sistema computacional em componentes individuais para mostrar como estes funcionam juntos para satisfazer aos requisitos do sistema inteiro (GLOVER, BHATT, 2007).

Devido à existência de diversas aplicações possíveis de sistemas RFID conclui-se que, certamente, não venha a existir uma arquitetura RFID universal que se adapte a todos os requisitos de todos os sistemas, mas existem diretrizes arquiteturais que devem ser consideradas em sistemas que englobam o uso de RFID (FINKENZELLER, 2003).

Basicamente, um sistema de identificação por RFID necessita de dois itens para seu funcionamento: identificadores e leitores, onde os leitores fariam a interface direta com a aplicação. Um terceiro componente desejável é o middleware RFID que, como pode ser visto na seção 2.3.2.3, tem por objetivo a filtragem das informações brutas recebidas dos leitores.

Nas seções a seguir, são detalhados os componentes citados.

2.3.2.1 Identificadores RFID

Resumidamente, um identificador de RFID típico consiste de um micro chip anexado a uma antena de rádio montada sobre um substrato, onde este pode armazenar desde um único bit até vários quilobytes de dados. Para recuperar os dados armazenados nesta, um leitor, por meio da emissão de ondas de rádio, envia um sinal e recebe de volta a resposta de todas as etiquetas presentes, que operem na mesma frequência do leitor, no raio de alcance deste.

Algumas características importantes devem ser consideradas na escolha do tipo de identificador, conforme a aplicação. Estas são (GLOVER, BHATT, 2007):

1 - Embalagem: há diversos tipos de embalagens para identificadores RFID, desde botões de PVC, frascos de vidro, plástico e até rótulos de papel. Desta forma, suas aplicações possíveis são praticamente ilimitadas;

2 - Acoplamento: se refere aos possíveis meios de comunicação entre leitor e identificador, onde a sua escolha depende basicamente da faixa de comunicação que se deseja utilizar, preço dos identificadores e condições que poderiam causar interferências;

3 - Energia: alguns identificadores utilizam-se de energia vinda de uma bateria, geralmente impressa junto da etiqueta, através de camadas de zinco, gel de água salgada e

dióxido de manganês, porém, na maioria dos casos, a energia é proveniente do próprio leitor e é transferida para o identificador por meio de indução eletromagnética;

4 - Capacidade de armazenamento: esta varia de acordo com o tipo de identificador e sua aplicação. Pode variar de um único bit ou até milhares de bytes, onde os valores podem ser pré-configurados na fábrica, ou alterados conforme a utilização;

5 - Compatibilidade com padrões: atualmente, os padrões de comunicação entre leitores e identificadores RFID são fornecidos pela EPCglobal e ISO. É importante considerar se os identificadores e leitores presentes no sistema seguem o mesmo padrão.

Devido à grande variedade nas características dos identificadores, estes podem ser utilizados em praticamente qualquer aplicação que necessite identificar ou rastrear itens dentro de um processo.

Na próxima seção são apresentados o princípio de funcionamento dos leitores RFID, seus componentes e a arquitetura básica de um sistema de RFID.

2.3.2.2 Leitores RFID

Leitores RFID são usados para reconhecer a presença próxima de identificadores RFID. Estes leitores transmitem energia RF através de uma ou mais antenas. Quando uma antena, em um identificador próximo, capta esta energia, a converte em energia elétrica através de indução. A energia convertida é suficiente para alimentar o chip semiconductor anexado à antena do identificador, que armazena sua identidade. O identificador envia, então, de volta para o leitor, a sua identidade, através do aumento e diminuição da resistência da antena, por meio de um código binário (FINKENZELLER,2003).

Leitores RFID são compostos de quatro subsistemas distintos (GLOVER, BHATT, 2007), detalhados nos itens a seguir, e sua arquitetura é mostrada na figura 2.1:

1 - API do leitor: interface de programação de aplicações que permite a programas se registrarem e capturarem eventos de leitura de identificadores RFID, além de capacidades para configurar, monitorar e gerenciar o leitor.

2 - Comunicações: leitores são componentes limítrofes do sistema de RFID e interagem com o componente de comunicação, que lida com as funções de rede.

3 - Gerenciamento de eventos: uma observação é o ato de um leitor ver um identificador. É considerado um evento uma observação que difira das anteriores, análise que é feita através de uma filtragem. O gerenciamento tem como objetivo definir quais

observações devem ser consideradas eventos e determina quais eventos são importantes para serem disponibilizados para alguma aplicação externa na rede.

4 - Subsistema antena: este consiste de uma ou mais antenas e a lógica e interfaces de suporte que permitem aos leitores RFID interrogar os identificadores.

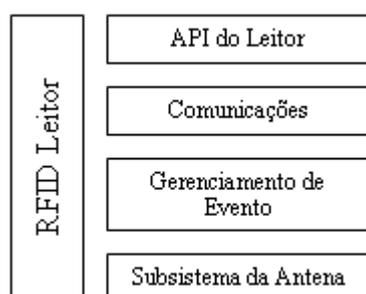


Figura 2.1 – Componentes lógicos de um leitor RFID, adaptada de (GLOVER, BHATT, 2007).

A figura 2.2 mostra um sistema RFID onde a aplicação é exposta aos dados brutos capturados pelos leitores.

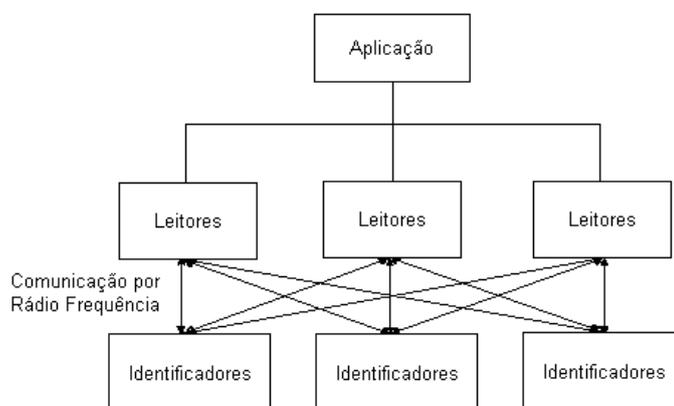


Figura 2.2 – Arquitetura RFID com a aplicação exposta diretamente aos leitores, adaptada de (GLOVER, BHATT, 2007).

Embora um sistema, contendo leitores e identificadores, possa ser suficiente para a construção de uma arquitetura RFID funcional, deve-se levar em consideração que os leitores RFID captam todas as informações que os identificadores próximos lhes enviarem.

Com o objetivo de tornar o sistema mais rápido e eficiente, é importante a presença de um terceiro componente que filtre para a aplicação apenas as informações relevantes a ela, o middleware RFID, que é apresentado na próxima seção.

2.3.2.3 Middleware RFID

Conforme exposto na seção anterior, o uso dos dados de saída dos leitores RFID, diretamente nas aplicações, dependendo do sistema, pode torná-lo ineficiente, devido à grande necessidade de processamento de informações irrelevantes à aplicação. Baseado nisto, é proposto o uso do middleware RFID com os objetivos de fornecer conectividade, processar observações RFID brutas para consumo pelas aplicações, fornecer uma interface em nível de aplicação para gerenciar leitores e capturar eventos RFID filtrados (GLOVER, BHATT, 2007).

O middleware RFID é composto de três componentes, detalhados nos itens a seguir (GLOVER, BHATT, 2007), e sua arquitetura mostrada na Figura 2.3:

1 - Adaptador de leitor: este encapsula as interfaces proprietárias de leitores, com a finalidade de estas não terem de entrar em contato com os desenvolvedores das aplicações.

2 - Gerenciador de eventos: este componente é o responsável pela filtragem dos dados brutos recebidos dos leitores. O Gerenciador de Eventos é o responsável por consolidar as informações capturadas pelos leitores e traduzi-las em uma informação do tipo identificador “presente” ou “não-presente”.

3 - Interface de nível de aplicação: sua principal finalidade é fornecer um mecanismo padronizado que permite às aplicações se registrar e receber eventos RFID filtrados proveniente de um conjunto de leitores, além de fornecer uma API padrão entre o middleware RFID e os componentes que ele controla.

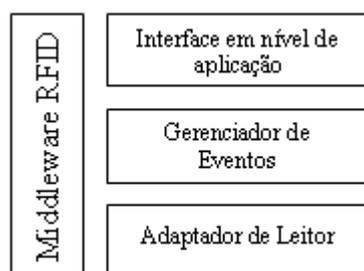


Figura 2.3 – Componentes do middleware RFID, adaptada de (GLOVER, BHATT, 2007).

Deve-se considerar que, como aplicações diferentes necessitam de informações diferentes contidas nos identificadores, o middleware RFID utilizado deve ser específico para a aplicação em que este será utilizado.

A Figura 2.4 mostra a arquitetura de um sistema de identificação RFID com a adição do middleware RFID. Nesta arquitetura pode-se perceber que não há mais a exposição direta da aplicação aos leitores.

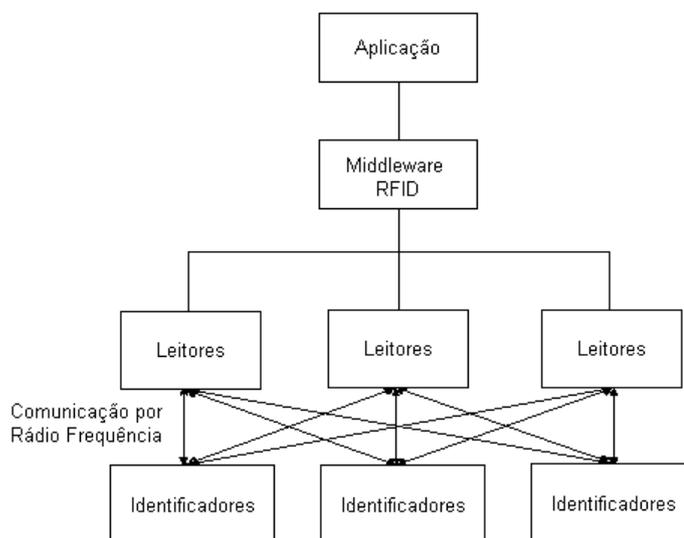


Figura 2.4 – Arquitetura RFID com a intermediação do middleware RFID entre os leitores e a aplicação, adaptada de (GLOVER, BHATT, 2007).

2.3.3 Principais aplicações

Sistemas de RFID são basicamente utilizados para identificação de objetos ou pessoas. Sua vantagem é que não requerem intervenção humana, pois as etiquetas podem ser lidas mesmo quando a etiqueta não está apontada na direção do leitor (RFID JOURNAL, [200-]c). Além disso, em seu padrão atual, já é possível a leitura de diversos identificadores simultaneamente, o que não ocorre com códigos de barras, por exemplo, que devem ser lidos individualmente e necessitam de um alinhamento específico com seus leitores (GLOVER, BHATT, 2007).

Algumas aplicações comerciais comuns em uso são o rastreamento de bens, gerenciamento de fornecimento, sistemas de pagamento, segurança e controle de acesso (RFID JOURNAL, [200-]c).

Outras aplicações, voltadas ao consumidor, em uso ou desenvolvimento, são pagamento de gasolina automático em postos, monitoramento de pacientes à distância, controle de “recall” de produtos com defeito de fabricação e máquinas de lavar inteligentes (RFID JOURNAL, [200-]b).

A fabricante de roupas italiana Benetton utiliza identificadores RFID anexados às peças de roupas, para rastreamento dos itens durante todo o processo de manufatura até os pontos de venda. Entretanto, estes identificadores são desativados antes da saída dos itens das lojas. (RFID JOURNAL, 2003)

Há muitas outras aplicações em que são utilizadas as etiquetas de RFID e inúmeras possibilidades de novas aplicações, principalmente aquelas onde o preço por etiqueta não altere consideravelmente o preço final dos produtos onde estão instaladas (RFID JOURNAL, [200-]c).

2.3.4 Uso de RFID no sistema proposto neste trabalho

Pode-se observar que são muitas as aplicações possíveis na identificação de itens por meio de RFID. Considerando-se o uso de identificadores RFID anexados a itens, para seu rastreamento no processo de manufatura, este trabalho propõe a extensão do uso destas identidades eletrônicas para fora das aplicações de fábrica.

Este trabalho consiste no desenvolvimento de um sistema de classificação de pessoas, voltado ao uso em sistemas de automação residencial, que utiliza os códigos eletrônicos dos identificadores RFID, contidos em roupas e objetos pessoais de habitantes, para verificar a presença destes por meio das combinações dos itens que os habitantes estão vestindo ou portando, no decorrer do funcionamento do sistema.

A classificação das combinações deve ser feita por meio da verificação da similaridade entre diferentes leituras do sistema de RFID e do agrupamento destas em um banco de dados. A próxima seção apresenta a técnica de inteligência artificial (IA) de raciocínio baseado em casos (RBC), que foi utilizada para a classificação das combinações de identidades RFID no sistema proposto.

2.4 Raciocínio Baseados em Casos (RBC)

A proposta desta seção é apresentar a técnica de raciocínio baseado em casos (RBC), uma técnica de inteligência artificial que consiste na solução de novos problemas a partir do resgate e adaptação de soluções armazenadas para problemas já conhecidos.

2.4.1 Introdução ao RBC

Raciocínio baseado em casos é um enfoque para a solução de problemas e para o aprendizado baseado em experiência passada. RBC resolve problemas ao resgatar e adaptar experiências passadas, chamadas casos, armazenadas em uma base de casos. Um novo problema é resolvido com base na adaptação de soluções de problemas similares já conhecidos (RIESBECK; SCHANK, 1989).

RBC é um paradigma para resolução de problemas que difere em muitos aspectos de outros enfoques da inteligência artificial. Em vez de basear-se unicamente em conhecimento generalizado, acerca de um domínio de problemas, ou de realizar associações lógicas ao longo de relacionamentos abstratos entre descritores de problemas e conclusões, RBC é capaz de utilizar o conhecimento específico de soluções de problemas concretos, experimentadas anteriormente, denotadas como casos (AAMODT; PLAZA, 1994).

O RBC pode ser separado por quatro elementos básicos (WANGENHEIM, C.G.; WANGENHEIM, A., 2003):

1 - Representação do conhecimento: em RBC, o conhecimento é representado principalmente na forma de casos que descrevem experiências concretas, embora possam ser armazenados também outros tipos de conhecimento sobre o domínio de aplicação, como casos abstratos, generalizados, entre outros;

2 - Medida de similaridade: é necessário haver a capacidade de se encontrar um caso relevante para o problema na base de casos e prover uma medida de similaridade entre o caso atual e os casos da base;

3 - Adaptação de casos: visto que dificilmente serão encontrados casos, na base, idênticos ao caso atual, sistemas de RBC avançados possuem a capacidade de adaptar os casos similares resgatados, com o objetivo de satisfazerem completamente o caso presente;

4 - Aprendizado: a fim de possibilitar a atualização e evolução do sistema de RBC, sempre que um problema for resolvido com sucesso, o mesmo deverá ser capaz de lembrar-se dessa situação, no futuro, como mais um novo caso.

As seções 2.4.2 a 2.4.5 detalham os itens apresentados nesta introdução.

2.4.2 Representação do conhecimento em RBC

A principal forma de representação do conhecimento em RBC são os casos. Estes são peças de conhecimento contextualizado que registram uma ocasião onde um problema foi resolvido, total ou parcialmente (WANGENHEIM, C. G.; WANGENHEIM, A., 2003).

Um caso representa, normalmente, a descrição de um problema (ou situação problemática) conjuntamente com as experiências adquiridas durante a sua resolução. Assim, a representação do conhecimento pode ser dividida nos seguintes componentes principais: descrição do problema e descrição da solução.

A descrição do problema descreve o estado do mundo quando o caso ocorreu. A descrição codifica o estado do mundo no momento em que o raciocínio se inicia e pode representar um problema que necessita ser resolvido ou uma situação que necessita ser interpretada, classificada ou compreendida.

Uma visão simplificada de RBC é a de que ele determina se um caso anterior é aplicável a uma nova situação, examinando as similaridades entre a descrição do problema na situação presente e no caso armazenado na base de casos (KOLODNER, 1993). Isto significa que uma descrição de problema deverá possuir informação suficiente para que seja possível julgar a aplicabilidade de um caso à nova situação.

De forma geral, a descrição do problema de um caso a ser armazenado deveria conter toda informação relevante para se descrever casos do tipo considerado e para se atingir o objetivo específico (KOLODNER, 1993).

A descrição da solução postula a solução (que pode ser uma ação, plano ou informação) derivada para o problema apresentado. Há vários tipos de solução, dependendo da área de aplicação do sistema de RBC, e esta descreve os conceitos ou objetos usados para atingir os objetivos específicos da tarefa realizada, levando em consideração as restrições especificadas e a descrição da situação (WANGENHEIM, C. G.; WANGENHEIM, A., 2003). A descrição da solução pode conter dados como a própria solução, os passos para atingi-la, justificativas para as decisões tomadas durante a solução, entre outros.

Quanto à representação da solução, há mais de uma forma desta ser feita. Esta pode ser explícita, através de sua descrição, por meio de um ponteiro para um documento contendo instruções ou outro conjunto de informações relevante.

Através destes dois componentes, novos problemas podem ser resolvidos encontrando-se um ou mais casos relevantes e adaptando-se a solução para os requisitos da nova situação. Componentes adicionais de um caso, como o resultado da aplicação da solução, também

podem ser utilizados para registrar o que ocorreu após a aplicação da solução, o nível do sucesso da aplicação da solução, como esta ocorreu e o porquê. Porém, nem sempre esses dados são necessários ou, pelo menos, não requerem tantos detalhes.

O resultado de um caso especifica o que ocorreu como consequência de se executar a solução proposta ou o quão bem a solução resolveu o problema. Um resultado inclui tanto a realimentação do mundo real como a interpretação deste. O resultado pode incluir o resultado em si, se este atingiu ou violou as expectativas, se foi um sucesso ou fracasso, estratégia de solução adotada e outros (WANGENHEIM, C.G.; WANGENHEIM, A., 2003).

2.4.3 Resgate de casos e medida de similaridade

O RBC utiliza soluções conhecidas de problemas anteriormente resolvidos na solução de um problema novo, de solução ainda desconhecida. Em função disto, é necessário que sejam determinados os exemplos adequados, os quais possam ser adaptados para a solução do novo problema, sendo que não é necessário que os problemas sejam idênticos para que seja possível a aplicação (WANGENHEIM, C. G.; WANGENHEIM, A., 2003).

A fim de tornar isto possível, é necessária a definição de critérios de similaridade entre casos, através de técnicas matemáticas, para determinar os casos que devem ser resgatados e os que podem ser adaptados para solucionar o novo problema. Uma forma de tratar o conceito de similaridade é utilizar o conceito abstrato de utilidade de casos para a solução de um problema.

O objetivo do resgate de casos é encontrar um caso (ou um pequeno conjunto) na base de casos que contenha uma solução útil para o problema ou situação atual, baseado na similaridade deste com cada caso armazenado (RIESBECK; SCHANK, 1989).

Uma hipótese básica de sistemas de RBC é que problemas similares possuem soluções similares (o critério “a posteriori” da utilidade de soluções passa a ser reduzido ao critério a priori de similaridade de descrições de problemas) (WANGENHEIM, C. G.; WANGENHEIM, A., 2003).

A eficácia de enfoques baseados em casos depende, portanto, da escolha de um conceito de similaridade adequado para o domínio da aplicação e estrutura dos casos usados, de forma a permitir a estimativa da utilidade de um caso com base na similaridade observada entre a descrição do problema e a contida no caso (KOLODNER, 1993).

2.4.3.1 Conceito de similaridade

No RBC é possível se formalizar o conceito de similaridade por meio de três formas diferentes (WESS, 1995):

- 1 - Similaridade como Predicado;
- 2 - Similaridade como Relação de Preferência;
- 3 - Similaridade como Medida.

A primeira idéia concebe similaridade como uma relação entre objetos ou fatos, que existe ou não existe. A segunda pressupõe a idéia de uma similaridade maior ou menor, enquanto o terceiro enfoque postula a quantificação da extensão dessa semelhança.

Provavelmente a forma mais conhecida de formalização do conceito de similaridade é a definição de uma medida numérica de distância ou similaridade. Em contraste com a utilização de uma relação de preferência, na utilização de uma medida em conjunto com informação de precedência é realizada também uma quantificação da similaridade (WANGENHEIM, C. G.; WANGENHEIM, A., 2003). Em um modelo geometricamente motivado do RBC é possível formular esta medida também como uma medida de distância entre casos.

2.4.3.2 Julgamento de similaridade

Modelos formais de similaridade procuram axiomatizar o processo da determinação da similaridade realizado pelo ser humano. As suposições assumidas no modelo podem então ser verificadas de forma empírica. Algumas suposições básicas a respeito de julgamentos de similaridade são (WESS, 1995):

- 1 - Reflexividade: Um objeto ou fato é similar a si mesmo;
- 2 - Simetria: A é similar a B, então B também é similar a A. A simetria das características de similaridade é questionada por alguns autores e não precisa prevalecer;
- 3 - Transitividade: Se A é similar a B e B é similar a C, então A também é similar a C. Esta é amplamente rejeitada na literatura, pois a transitividade só é válida quando se baseia na identidade parcial de características;
- 4 - Monotonicidade: A similaridade de dois objetos cresce monotonicamente com o aumento de correspondências e a redução de diferenças.

Devido aos múltiplos aspectos que devem ser levados em consideração, quanto ao julgamento da similaridade, e da incerteza associada a estes, não é possível definir uma teoria universal da determinação de similaridade. Nem se pode considerar o processo de seleção de casos em um sistema de RBC como certo ou errado, mas apenas avaliá-lo como melhor ou pior, em um contexto específico de aplicação (WANGENHEIM, C.G.; WANGENHEIM, A., 2003).

2.4.4 Adaptação de casos

Após definida a similaridade entre casos, para completar esse processo, é necessária a adaptação do caso recuperado ao caso atual. A adaptação tem um papel fundamental na flexibilidade dos sistemas de RBC e a sua capacidade de resolver novos problemas depende de sua habilidade em adaptar casos recuperados a novas e em sua habilidade de consertar soluções que falham ao serem aplicadas (AAMODT; PLAZA, 1994).

Algumas questões são centrais na adaptação de casos, como a decisão sobre os aspectos de uma situação que devem ser adaptados, as modificações que são razoáveis de serem realizadas para adaptar o caso, os métodos de adaptação aplicáveis para modificar estes aspectos e como controlar o processo de adaptação para saber que as modificações estão sendo realizadas no rumo certo (WANGENHEIM, C.G.; WANGENHEIM, A., 2003).

2.4.5 Aprendizado

A tarefa de aprender pode ser vista como o processo de melhora do desempenho de um sistema de RBC.

RBC implica uma forma de aprendizado por analogia, em que, por meio da transformação e extensão de conhecimento existente, uma tarefa ou problema similar são executados ou resolvidos (WANGENHEIM, C.G.; WANGENHEIM, A., 2003).

Em um sistema de RBC, raciocínio e aprendizado estão intimamente ligados. Todas as vezes que um problema é resolvido, a nova experiência pode ser retida e armazenada na base de casos, tornando-a imediatamente disponível para problemas futuros. Desta forma, o conhecimento presente em um sistema de RBC é continuamente atualizado (AAMODT; PLAZA, 1994).

No RBC constrói-se uma base de casos a partir de um conjunto de exemplos de casos dados. Deve-se decidir quais casos serão aceitos para compor a base, quais serão rejeitados e

com base em quais critérios e como se vai definir a medida de similaridade. O projeto correto da dinâmica desse sistema é essencial no aprendizado das regras baseadas em casos anteriores para o bom funcionamento do mesmo.

2.4.6 Aplicação do RBC no sistema proposto neste trabalho

Conforme citado na seção 2.3.4, o sistema proposto neste trabalho considera o uso de RFID para a detecção das roupas e itens pessoais dos habitantes de uma residência. Com o objetivo de classificar estas pessoas, a partir da combinação destes itens, foi utilizado o RBC, que, resumidamente, baseia-se em utilizar soluções adaptadas de casos passados na solução de novos casos.

Neste trabalho, combinações de objetos, representados por suas identidades RFID, foram tratadas como casos e suas respectivas classificações como a solução destes. É utilizado o RBC com o objetivo de encontrar critérios e valores de similaridade que possibilitem a correta solução de cada caso, ou, em outras palavras, a correta classificação de cada combinação de objetos.

2.5 Automação Residencial Inteligente

Esta seção visa apresentar, de forma resumida, alguns trabalhos desenvolvidos na área de automação residencial inteligente. Nos sistemas desenvolvidos nestes trabalhos, regras de automação são criadas a partir do comportamento dos habitantes, o qual é capturado por meio de sensores e atuadores implantados na residência. Conforme exposto a seguir, os sistemas foram concebidos considerando-se residências com apenas um habitante, o que motivou o desenvolvimento do classificador de habitantes.

2.5.1 Automação Baseada em Comportamento (ABC)

Em (TONIDANDEL; TAKIUCHI; MELO, 2004) foi apresentado um sistema denominado ABC (Automação Baseada em Comportamento), que visa reverter o processo normal de criação de regras de automação em uma residência. Normalmente, o habitante cria as regras; no ABC, as regras são criadas pelo próprio sistema, com base no comportamento do habitante. Neste foi utilizado o algoritmo de aprendizado ID3 (QUINLAN, 1986) para adaptação do sistema de automação.

A arquitetura deste sistema, mostrada na Figura 2.5, é composta de sensores (presença, temperatura, luminosidade), atuadores (interruptores, ar-condicionado), bancos de dados e outros elementos necessários pra criação e controle das regras.

Os valores obtidos através dos componentes de entrada do sistema (sensores, atuadores), fatores como horários e o banco de dados de eventos interagem com o sistema para a criação das regras de automação, as quais são armazenadas em bancos de dados de regras. Estas serão utilizadas em casos futuros para acionamento de dispositivos, quando o sistema decidir que a regra se aplica a uma determinada configuração de entradas deste. As regras sofrem manutenção conforme o sistema é utilizado, podendo ser alteradas ou removidas se não forem utilizadas por um determinado tempo. Este sistema também possui regras de segurança, as quais são executadas quando os sensores atingem determinados estados pré-configurados, porém estas regras não sofrem manutenção, ou seja, não são modificadas ou excluídas.

Basicamente, o sistema consiste em criar regras, relacionando eventos de sensores e atuadores, e executá-las quando as condições dos sensores se repetirem.



Figura 2.5 – Arquitetura do Sistema ABC, adaptada de (TONIDANDEL; TAKIUCHI; MELO, 2004).

Embora interessante, este sistema possui algumas deficiências, como a incapacidade de detectar seqüências causais no tempo, ou seja, avaliar se eventos sensores e atuadores, ocorridos seqüencialmente, estão, ou não, relacionados. Além disso, este sistema lida apenas com variáveis lógicas, não sendo capaz de lidar com variáveis contínuas. Também neste, as regras criadas não passam por uma etapa de validação, além de não poderem ser excluídas, automaticamente, caso o habitante passe a contrariar as ações executadas pelo sistema.

Com base nestas limitações, Sgarbi (2007) propôs uma evolução do sistema original, a qual foi denominada ABC+, e é apresentada na seção a seguir.

2.5.2 Evolução do sistema ABC (ABC+)

Em (SGARBI, 2007), foi proposta uma evolução do sistema ABC original, com o intuito de corrigir as limitações deste (incapacidade de detectar seqüências causais de eventos no tempo, a limitação de se trabalhar apenas com variáveis lógicas e o fato de as regras criadas se tornarem ativas logo após sua criação). Este novo sistema foi denominado ABC+ e sua arquitetura é mostrada na Figura 2.6.

As principais diferenças são: a janela de observação de eventos, que foi inserida com o objetivo de avaliar se os eventos de atuadores e sensores são seqüenciais ou não, as regras embrionárias, para criar uma etapa de validação de regras antes de torná-las ativas no sistema, o uso do algoritmo C4.5 (QUINLAN, 1993) e o novo processo de desenvolvimento e manutenção de regras.

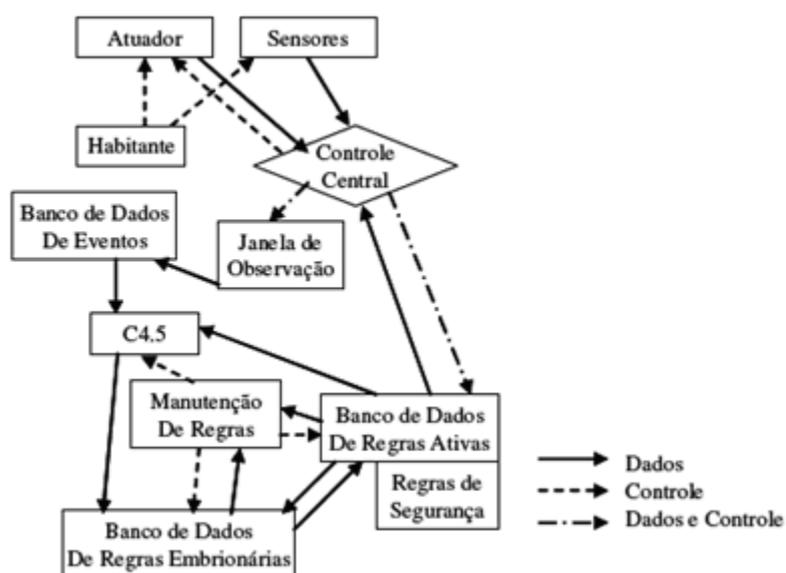


Figura 2.6 – Arquitetura do Sistema ABC+, adaptada de (SGARBI, 2007).

Neste processo, além da criação de novas regras, a partir dos eventos quando condições específicas dos sensores são atendidas, foi implementada a possibilidade de pontuação negativa e exclusão de regras, quando a relação do habitante com os atuadores contraria os eventos armazenados no banco de dados de regras. Por exemplo, considerando que para determinada condição dos sensores esteja associado o evento de acender uma lâmpada e, quando o sistema o fizer, o habitante, na seqüência, apagá-la. O sistema passará a entender que esta regra não atende mais a necessidade do morador, e efetuará a manutenção

da regra com o intuito de excluí-la, se o habitante continuar contrariando a ação executada pelo sistema de automação.

Além dos sistemas apresentados nesta seção, outros também foram desenvolvidos com o objetivo de criar ambientes inteligentes em residências, com diferentes metodologias, porém com o intuito comum de agregar conforto aos habitantes. Estes não foram discutidos neste trabalho. O detalhamento das arquiteturas apresentadas teve como propósito expor as vantagens dos sistemas que baseiam a criação das regras de automação no comportamento dos habitantes (de forma a não necessitarem da intervenção do usuário na sua programação) e a sua limitação quando tinham de operar em ambientes com mais de uma pessoa.

3 CLASSIFICADOR AUTOMÁTICO DE PESSOAS BASEADO EM CASOS

Conforme exposto na seção 2.5, os sistemas de automação baseada em comportamento já desenvolvidos mostraram-se eficazes na automação de eventos em uma residência, quando era considerado que neste habitava apenas uma pessoa. Embora estes possam ser aplicados em residências com mais de um morador, a falta da identificação destes levaria à criação de regras iguais a todos os habitantes, o que poderia não ser conveniente, pois estes podem ter hábitos diferentes dentro da residência.

Estudando-se o funcionamento do sistema ABC+, pode-se concluir também que, se diferentes habitantes tiverem hábitos contrários, uma ação executada por uma determinada pessoa dentro da residência pode levar à pontuação negativa de regras criadas para outro habitante. Isto pode invalidar totalmente o funcionamento do sistema desenvolvido, já que os pontos positivos atribuídos pelo sistema devido às ações de um habitante (condição para criação de regras embrionárias e sua ativação) podem ser anulados por pontos negativos atribuídos pelas ações de outros.

Visando complementar o funcionamento do sistema ABC+, foi proposto o Classificador Automático de Pessoas Baseado em Casos (CAPBC) que permitirá a adaptação de sistemas de automação residencial desenvolvidos para uso em residências com apenas um habitante para aquelas com duas ou mais pessoas.

O CAPBC consiste em um sistema de classificação de pessoas baseado nas combinações de roupas e objetos utilizados pelos habitantes de uma residência. Cada habitante em uma residência possui um número finito de peças de roupa e objetos pessoais. Todos os dias estes utilizam combinações destes itens e, no decorrer do tempo, estes são reutilizados e recombinados. A proposta do sistema é utilizar a técnica de inteligência artificial de Raciocínio Baseado em Casos para abstrair a classificação das combinações com base no conhecimento de combinações já utilizadas no passado.

Para a detecção dos objetos é considerado que, em um futuro próximo, roupas e objetos poderão ter identidades eletrônicas (por exemplo, RFID) anexados a estes na fábrica, como já ocorre em itens de algumas marcas, conforme exposto na Seção 2.3. A tecnologia RFID possui algumas vantagens sobre outras tecnologias de identificação, já que permite a leitura de vários itens simultaneamente, a uma distância razoável e sem exigir o alinhamento direto entre o leitor e os identificadores.

O CAPBC mantém as vantagens do sistema ABC+, onde o habitante não necessita alterar seus hábitos a fim de adaptar-se ao sistema de automação. O uso de uma técnica de

inteligência artificial no sistema visa permitir que o habitante seja classificado por suas roupas e objetos sem que este tenha de pré-cadastrar seus itens em um banco de dados para posterior comparação. A classificação das combinações é feita por meio da abstração de conhecimento dos casos anteriores já classificados pelo sistema. O RFID permite que os objetos sejam lidos sem que o habitante tenha de alterar seus hábitos cotidianos ao interagir com a residência.

Outras técnicas ou tecnologias poderiam ser utilizadas para o desenvolvimento deste sistema no lugar do RBC ou RFID, porém não é foco do trabalho eleger o melhor método, mas, sim, utilizar um que atenda satisfatoriamente os objetivos propostos para o sistema.

A arquitetura do CAPBC é mostrada na Figura 3.1 e seu funcionamento e seus componentes são detalhados a partir da próxima seção. Nesta arquitetura, o sistema de automação ilustrado poderia ser o ABC+, ou outro sistema domótico individual.

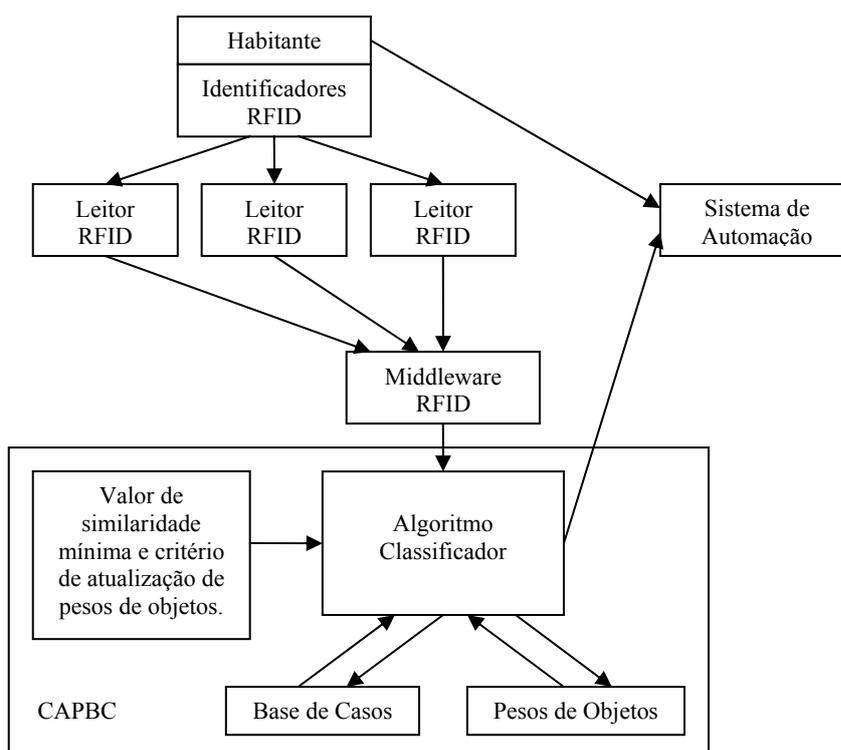


Figura 3.1 - Arquitetura do sistema CAPBC

3.1 Funcionamento do CAPBC

Para o propósito da criação e execução de regras de automação em uma residência, desde que não esteja envolvida sua segurança (como a abertura da porta de acesso à residência, por exemplo), não é essencial o conhecimento da identidade do habitante; é

importante apenas saber quando uma ação no sistema é acionada pelo mesmo indivíduo que acionou outras ações no passado, a fim de criar as regras de automação associadas corretamente a cada habitante. Desta forma, o objetivo do sistema CAPBC proposto é classificar combinações de objetos por meio de códigos genéricos, de forma a permitir o agrupamento, por habitante, das ações no sistema de automação.

Quando uma combinação qualquer N , denominada $C_N = \{O_1, \dots, O_Q\}$, contendo Q objetos, é capturada pelo sistema, este deve compará-la àquelas detectadas no passado (C_1, \dots, C_{N-1}), já classificadas, e que estarão armazenadas em uma base de casos BC. Para as comparações, foram definidos critério e o valor de similaridade mínima (itens detalhados posteriormente, no trabalho) que o sistema deve encontrar para atribuir a duas ou mais combinações a mesma classificação CL, ou, em outras palavras, agrupá-las sob um mesmo código de habitante.

Com o passar do tempo, o sistema terá mais casos armazenados em sua base, o que significa mais conhecimento sobre os habitantes, objetos e combinações, que poderão ser utilizados na classificação de novos casos.

3.2 Algoritmo Classificador

O algoritmo classificador é o principal componente do CAPBC. Suas funções são: armazenar novas combinações de objetos recebidas do sistema de RFID na BC, comparar os novos casos àqueles anteriormente contidos na BC, classificá-los a partir dos critérios pré-estipulados de similaridade, atualizar parâmetros dos casos armazenados, atualizar pesos dos objetos e informar as classificações ao sistema de automação. O pseudo-algoritmo do classificador é apresentado na Figura 3.2.

Quando um novo caso $C_N = \{O_1, \dots, O_Q\}$, contendo Q objetos, é recebido pelo algoritmo classificador, uma nova posição é inserida na base de casos, e um índice NC, que corresponde à contagem N dos casos presentes na BC, naquele momento, é atribuído a esta. Os valores de O_1 a O_Q são ordenados, armazenados na nova posição criada na BC e seus pesos são atualizados na tabela de pesos. Inicialmente, o novo caso recebe uma classificação CL igual ao próprio índice do caso e deve ser informada ao sistema de automação esta nova classificação de habitante.

Em seguida, o novo caso é comparado a cada caso passado já contido na base ($NC = 1$ a $N - 1$) e é calculada a similaridade S entre cada par ‘caso atual – caso passado’, armazenando os valores na própria base de casos.

Evento de Entrada de Nova Combinação do Leitor.

Variáveis:

BC: Base de Casos contendo todas as combinações já lidas dentro da residência com sua respectiva classificação;

N: Contagem de casos armazenados na BC até a presente leitura;

Q: Quantidade de objetos presentes em determinado caso;

$O_1 - O_Q$: Campos da BC onde são armazenados os códigos numéricos de cada identificador RFID lido em uma combinação de objetos;

$P_1 - P_Q$: Pesos dos objetos $O_1 - O_Q$;

NC: Campo da BC contendo o número do caso armazenado;

S: Campo da BC onde é armazenado o valor de similaridade calculado pelo Algoritmo Classificador entre cada caso passado e o caso atual;

S^* : Valor de similaridade mínima imposto ao sistema para classificação de um caso com mesmo CL de um caso anterior;

SF: Flag da BC onde é indicado se a $S \geq S^*$;

CL: Campo da BC onde é armazenado o código de identificação do habitante (classificação) de cada caso, atribuído pelo Algoritmo Classificador;

CL*: Variável onde é armazenado o menor valor de CL quando um ou mais casos tiver SF setado;

BCL: Banco de dados onde são armazenados os diferentes valores de CL de todos os casos com $S > S^*$ encontrados durante a classificação de cada novo caso.

Início

1. **insert** nova posição na BC;
2. Gravar na BC: $NC = N$;
3. Ordenar valores de O_1 a O_Q do novo caso;
4. Gravar na BC: O_1 a O_Q ;
5. Recalcular $P_1 - P_Q$;
6. Gravar na tabela de pesos: $P_1 - P_Q$;
7. Gravar na BC: $CL = NC$; // classificação inicial de cada caso é o próprio número do caso
8. Enviar para o sistema de automação: CL;
9. **for** cada posição da BC, de $NC = 1$ até $N - 1$, **do**:
10. Calcular S entre o novo caso e caso da posição atual;
11. Gravar na BC: S;
12. **if** $S > S^*$, **then**:
13. $SF = 1$;
14. Gravar na BCL: CL;
15. **else** $SF = 0$;
16. **for** cada posição da BC, de $NC = 1$ até N, **do**:
17. **if** $SF = 1$ e $CL < CL^*$, **then** $CL^* = CL$;
18. **for** cada posição da BC, de $NC = 1$ até N, **do**:
19. **if** $SF \neq 1$ e CL existe em BCL, **then** $SF = 2$;
20. **for** cada posição da BC, de $NC = 1$ até N, **do**:
21. **if** $SF > 0$, **then** $CL = CL^*$;
22. Enviar para o sistema de automação: CL^* , todos NC onde $SF > 0$;

Fim.

Figura 3.2 - Algoritmo de evento de entrada do Classificador RBC

Os métodos utilizados para a atualização dos pesos de objetos e para o cálculo da similaridade são detalhados no próximo capítulo, onde foram expostos os critérios e valores testados para uso no sistema.

Se a similaridade for igual ou superior ao valor mínimo de similaridade estipulado, uma flag de similaridade SF recebe valor '1' e a classificação CL deste caso é armazenada em um banco de dados auxiliar BCL, que é consultado, posteriormente, a fim de agrupar os casos não-similares ao caso atual, mas similares a algum dos casos onde SF for igual a '1'. Se a similaridade for inferior ao valor mínimo estipulado, SF recebe valor '0'.

Depois de encontrados todos os casos similares entre si, o algoritmo verifica todos estes a fim de encontrar o caso de menor valor CL, denominado CL*, sob o qual todos os demais casos similares devem ser agrupados. Na seqüência, cada caso é avaliado a fim de verificar se existem similaridades entre casos com SF igual a '0' (não-similares ao caso atual) mas com o valor CL presente em BCL (similares a algum caso que o seja). Estes recebem o valor '2' em SF. Por fim, todos os casos com SF acima de '0' têm o valor CL atualizado para o valor de CL*.

A figura a 3.3 ilustra um exemplo de reclassificação de casos a partir do algoritmo classificador proposto. Para o exemplo, foi considerada similaridade mínima (S^*) = 3 objetos iguais.

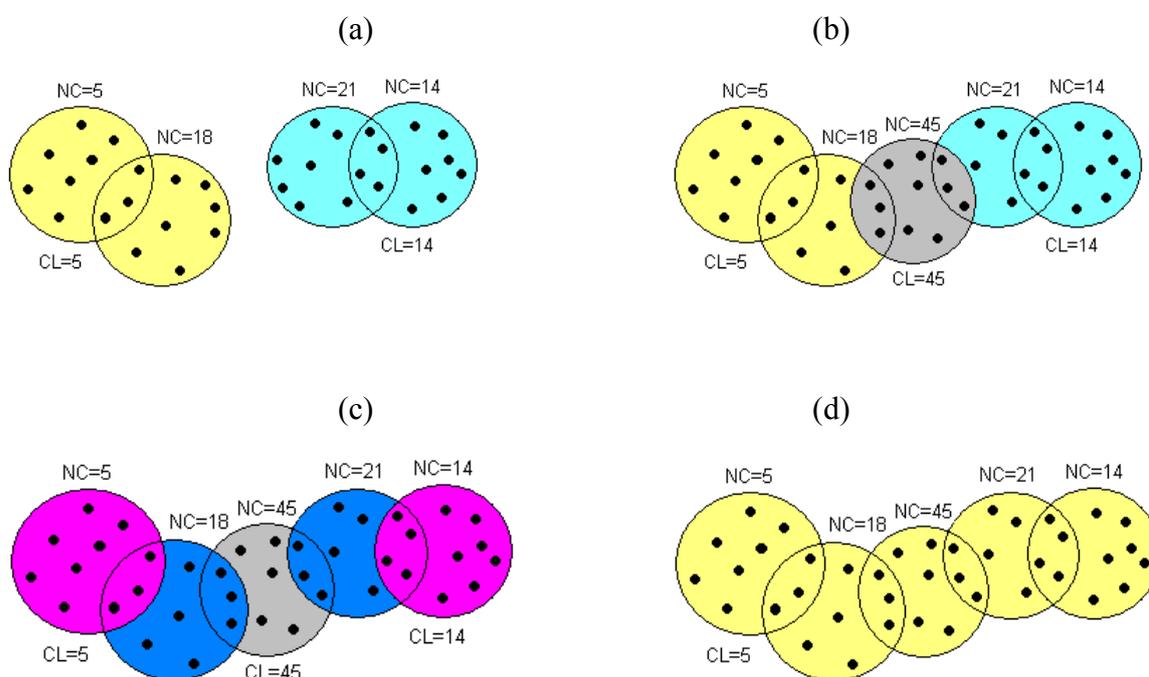


Figura 3.3 - Exemplo de reclassificação de casos a partir da entrada de um novo caso (NC=45), onde cada círculo representa um caso e os pontos pretos representam os objetos. A Figura (a) representa um subconjunto da BC, contendo 4 casos (NC=5, NC=14, NC=18, NC=21) e 2 classificações (CL=5 e CL=14). A Figura (b) representa a entrada do caso NC=45, tendo três objetos comuns com NC=18 e três com NC=21. A Figura (c)

representa o cálculo da flag de similaridade SF (igual a '1' nos casos ilustrados em azul e igual a '2' nos ilustrados em rosa). A Figura (d) representa os casos reagrupados sob o menor CL (CL*), no caso, igual a 5.

Todas as vezes que um novo caso é classificado, um código de habitante novo é enviado para o sistema de automação, com o objetivo deste poder atribuir o evento ocorrido no sistema ao habitante correto. Se o novo foi classificado como similar a algum outro caso da BC, os códigos de todos os casos similares também são enviados para o sistema de automação, para que este possa reagrupar em seu banco de dados, os eventos passados relacionados a este habitante. No Apêndice 1 pode ser visto o diagrama esquemático do algoritmo classificador.

3.3 Base de Casos

Devido à necessidade do sistema utilizar o conhecimento de casos passados na classificação de novas combinações, foi criado um banco de dados para armazenar as informações necessárias desses casos. Este é denominado 'Base de Casos' ou BC.

A BC consiste em uma tabela onde são armazenadas as combinações de objetos $C_N = \{O_1, \dots, O_Q\}$, as classificações (CL) atribuídas pelo algoritmo classificador e os campos temporários utilizados para armazenar os variáveis de similaridade calculadas: valor da similaridade (S) e flag de similaridade (SF). A BC é acessada e atualizada sempre que um novo caso for recebido do sistema de identificação RFID pelo algoritmo classificador.

Cada vez que uma nova combinação de objetos é lida pelo sistema, um novo caso contendo os códigos de objetos da nova combinação é criado na BC. São, então, calculados valores de similaridade com cada caso anterior, os quais são armazenados na BC sob cada caso C_1 a C_{N-1} . Depois de efetuada a reclassificação dos códigos de habitantes, a BC é atualizada com a nova classificação.

A tabela 3.1 mostra como as combinações são armazenadas na Base de Casos.

Tabela 3.1 – Casos armazenados na BC, onde NC é o número do caso e O_1 a O_N são os códigos de objetos armazenados sob o caso

NC	O_1	O_2	O_3	...	O_{N-2}	O_{N-1}	O_N	CL	S	SF	
1	415	679	1017	...	3512	3826	4507		1	4	1
2	242	395	808	...	2715				2	0	0
3	306	395	977	...	3099				3	0	0
4	415	1213	1772	...					4	1	0
...
N	415	800	1017	...	3826	4507			1		

3.4 Pesos de Objetos

Conforme pode ser observado na arquitetura do CAPBC, os únicos dados externos recebidos pelo sistema são as combinações de códigos de objetos entregues pelo sistema de RFID. Desta forma, é desejável que a maior quantidade de informações possa ser extraída destes dados, para serem utilizadas pelo classificador.

Um parâmetro que pode ser verificado a partir dessas combinações é a quantidade de vezes que cada objeto foi lido pelo sistema RFID e, conseqüentemente, armazenado na BC. A tabela de pesos de objetos consiste em armazenar estes valores na forma de um peso calculado para cada objeto, de acordo com a quantidade de vezes que este é lido nas combinações. Este peso pode ser interpretado como um fator de relevância de cada objeto, ou seja, quanto mais vezes um item for lido no decorrer do funcionamento do sistema, mais relevante este deve ser para o habitante que o está utilizando e, conseqüentemente, para a classificação feita pelo sistema. O uso e a atualização destes valores são detalhados no capítulo 4, onde é mostrado o simulador desenvolvido com os campos específicos para atualização dos pesos.

A tabela 3.2 ilustra como os pesos são armazenados na tabela de pesos, onde o valor O_N representa o código do objeto.

Tabela 3.2 – Pesos de objetos armazenados na tabela de pesos, onde O_N representa o código do objeto.

O_N	Peso
242	1
306	1,2
395	1,4
415	2
679	1,2
800	1
808	1,2
...	...
3826	1,2
4507	1,6

3.5 Critérios de classificação e atualização de parâmetros

Para que o CAPBC possa classificar as novas combinações, é necessário que sejam definidos critérios de comparação entres os casos e valores de similaridade mínima que devem ser alcançados para que dois casos diferentes recebam a mesma classificação.

Três critérios de similaridade foram considerados para o uso com o CAPBC, e são detalhados na seqüência: número de objetos iguais, porcentagem de objetos iguais sem a utilização de pesos e porcentagem de objetos iguais com a utilização de pesos.

3.5.1 Número de objetos iguais

Consiste em se utilizar a quantidade $|C|$ de objetos contidos na intersecção de dois conjuntos de objetos. O cálculo da similaridade é mostrado na equação 3.1, onde C_X e C_Y representam dois casos distintos.

$$SIM = |C_X \cap C_Y| \quad (3.1)$$

Por exemplo:

$$C_X = \{1, 3, 6, 10, 15, 21\};$$

$$C_Y = \{5, 10, 15, 20\};$$

$$C_X \cap C_Y = \{10, 15\}.$$

$$SIM = |C_X \cap C_Y| = 2$$

Neste método, o valor de similaridade obtido é um número absoluto, que independe da quantidade total de objetos presente em cada caso.

3.5.2 Porcentagem de objetos iguais sem a utilização de pesos

Consiste em se calcular uma relação entre a quantidade de objetos contidos na intersecção entre dois conjuntos de objetos e a quantidade total de objetos. A maneira proposta para se estabelecer esta relação é expressa pela equação 3.2.

$$SIM = \frac{2 * |C_X \cap C_Y|}{|C_X| + |C_Y|} \quad (3.2)$$

A quantidade total de objetos dos dois conjuntos foi somada a fim de normalizar os valores das similaridades entre C_X e C_Y e entre C_Y e C_X e a expressão foi multiplicada por 2 a fim de que o valor de similaridade esteja entre 0 e 100%.

Por exemplo:

$$C_X = \{1, 3, 6, 10, 15, 21\};$$

$$C_Y = \{5, 10, 15, 20\};$$

$$C_X \cap C_Y = \{10, 15\}.$$

$$SIM = \frac{2 * |C_X \cap C_Y|}{|C_X| + |C_Y|} = \frac{2 * 2}{6 + 4} = 40\%$$

Neste método, o valor de similaridade depende diretamente da quantidade total de objetos presentes nos casos comparados. Assim, quando os casos comparados tiverem poucos objetos, os itens iguais terão maior relevância no cálculo de similaridade do que em conjuntos de objetos maiores.

3.5.3 Porcentagem de objetos iguais com a utilização de pesos

Utiliza o mesmo cálculo exposto anteriormente, porém os objetos podem ter pesos diferentes no cálculo. Desta forma, se objetos com pesos maiores pertencerem ao conjunto intersecção, o valor de similaridade obtido será maior do que se não fossem utilizados os pesos. Esta relação é expressa pela equação 3.3, onde $\sum p(C)$ representa a somatória dos pesos daquele conjunto de objetos:

$$SIM = \frac{2 * \sum p(C_X \cap C_Y)}{\sum p(C_X) + \sum p(C_Y)} \quad (3.3)$$

Por exemplo:

$$C_X = \{1, 3, 6, 10, 15, 21\};$$

$$p(1) = p(3) = p(5) = 1;$$

$$C_Y = \{5, 10, 15, 20\};$$

$$p(6) = p(10) = 2;$$

$$C_X \cap C_Y = \{10, 15\};$$

$$p(15) = p(20) = p(21) = 3;$$

$$SIM = \frac{2 * \sum p(C_X \cap C_Y)}{\sum p(C_X) + \sum p(C_Y)} = \frac{2 * 5}{12 + 9} = \frac{10}{21} \approx 47,6\%$$

Neste método, o valor de similaridade passa a ser uma relação entre as relevâncias de cada item nos casos, no lugar de uma simples relação entre quantidades. Isso significa que se objetos mais relevantes pertencerem ao conjunto intersecção, o valor de similaridade obtido será maior; se, por outro lado, pertencerem a um dos casos, mas não ao conjunto intersecção, a similaridade obtida será menor. O valor obtido por meio deste cálculo também se encontrará entre 0 e 100%.

3.6 Observações sobre os métodos e cálculos utilizados para classificação dos habitantes

Visto o objetivo do CAPBC de classificar habitantes a partir de combinações de objetos, outras técnicas de inteligência artificial poderiam ter sido utilizadas na comparação das combinações. Foram consideradas, inicialmente, técnicas como o PCA (Principal Component Analysis) e LDA (Linear Discriminant Analysis), porém estas necessitam que os vetores de dados comparados possuam tamanhos iguais, além do posicionamento dos tipos de dados semelhantes nas mesmas posições em todos os vetores (HASTIE, TIBSHIRANI, FRIEDMAN, 2008). No caso do CAPBC, foi considerado que os vetores podem ter tamanhos variados, além de não ser possível conhecer o tipo de objeto pela sua identidade eletrônica.

O RBC é uma técnica que permite a comparação de conjunto de dados de diferentes tamanhos, além de permitir o uso de relevâncias maiores ou menores para diferentes objetos; por isto foi escolhido para o desenvolvimento deste trabalho.

Quanto aos critérios de similaridade propostos para a comparação dos vetores, todos seguem as suposições básicas a respeito do julgamento de similaridade em RBC, apresentados na seção 2.4.3.2: reflexividade, simetria, transitividade e monotonicidade.

4 SIMULAÇÃO DO SISTEMA CAPBC

Considerando-se que, nos dias atuais, poucos itens de vestuário e objetos pessoais já possuem identificadores RFID implantados em fábrica, e que estes poucos ainda chegam ao consumidor final desativados, torna-se muito complexa a simulação do sistema CAPBC em uma residência e situações reais. Para que isto pudesse ser implementado, novos identificadores teriam de ser implantados em todos os objetos de cada habitante, a fim de que as condições consideradas para o funcionamento do classificador pudessem ser satisfeitas.

Além disto, há muitas variáveis a serem testadas, visando encontrar os critérios e valores de similaridade que tornarão o sistema mais eficiente e garantirão o seu correto funcionamento. O tempo necessário para a coleta de dados suficientes para a avaliação de cada conjunto de critérios considerados também seria impraticável.

Desta forma, a solução encontrada foi o desenvolvimento de simuladores em computador que satisfaçam as características de uma residência real e que foram descritos a partir da próxima seção.

4.1 Descrição dos simuladores

Visando testar o comportamento da lógica proposta para o sistema CAPBC, foram criados três softwares em computador: semeador de casos, classificador de casos e avaliador de classificações, os quais são detalhados nas próximas seções.

4.1.1 Semeador de casos

O objetivo do semeador de casos é gerar combinações de roupas e objetos utilizados pelos habitantes de uma residência hipotética, da forma mais fiel possível a uma residência real, a fim de possibilitar a posterior simulação do algoritmo classificador. Foi necessária a criação do semeador devido à complexidade do processo de captura de dados de uma residência real, o tempo necessário para coleta de dados suficientes para os testes, entre outros.

Os casos gerados pelo semeador devem atender certos critérios para que possam substituir dados reais nas simulações. Primeiramente, habitantes reais possuem um número limitado de roupas e objetos. Embora novas roupas e acessórios possam ser adquiridos com o tempo, de maneira geral, os conjuntos gerados devem ser combinações dos itens que cada

habitante possui. Além disso, as combinações de objetos devem ser coerentes. Por exemplo, embora um habitante possa ter cinco pares de sapatos, é impossível que esteja calçando mais do que um por vez. Por último, certas combinações de roupas, embora possíveis de serem utilizadas, são extremamente improváveis, tornando-as dispensáveis em um ambiente de simulação. Por exemplo: embora um habitante possa calçar chinelos, bermuda, sobretudo, gravata e óculos escuros, esta não é uma combinação comum de vestuário.

Sendo assim, a fim de atender os critérios expostos acima, o semeador de casos foi desenvolvido em três partes.

A primeira parte consiste em gerar um banco de dados de roupas e objetos de cada habitante, identificados por códigos numéricos, que foram posteriormente utilizados para representar suas identidades eletrônicas hipotéticas, isto é, os códigos dos identificadores RFID. Os habitantes são classificados por seus sexos (masculino ou feminino) e faixas etárias (adultos ou crianças), a fim de possibilitar a posterior geração das combinações de seus itens de forma coerente. Foi feito desta forma, já que determinadas combinações de roupas são condizentes com qualquer sexo e faixa etária de habitante (como tênis, calça jeans e camiseta) e outros não (como calças sociais, camisa, terno e gravata, próprios de adultos do sexo masculino).

Cinquenta e cinco tipos de roupas e objetos foram criados. No software de simulação, cada um destes tipos foi caracterizado por seu nome e um código numérico, para facilitar sua identificação posterior e associação nas outras partes do simulador. Também, cada tipo de objeto é classificado no banco de dados por um campo que indica se o objeto se trata de um objeto íntimo, ou seja, pessoal e não passível de empréstimo, ou não. Este dado foi importante quando foram simulados empréstimos de roupas e objetos entre os habitantes.

A Figura 4.1 mostra o segmento do simulador responsável pelo cadastro dos objetos de cada habitante.

A segunda parte consiste em cadastrar combinações de roupas e objetos possíveis para cada tipo de habitante. No cadastro de cada combinação são informados os tipos de objeto possíveis para a combinação, a probabilidade de cada objeto estar presente na combinação e o tipo de habitante à qual a combinação pode pertencer.

A probabilidade da presença do objeto consiste em determinar quais itens necessariamente estarão em uma combinação e quais podem ou não estar, com uma probabilidade maior ou menor. Por exemplo, uma combinação social masculina deve ter, obrigatoriamente, itens como calça social, sapato e camisa, mas podem não ter terno, ou gravata, ou ambos.

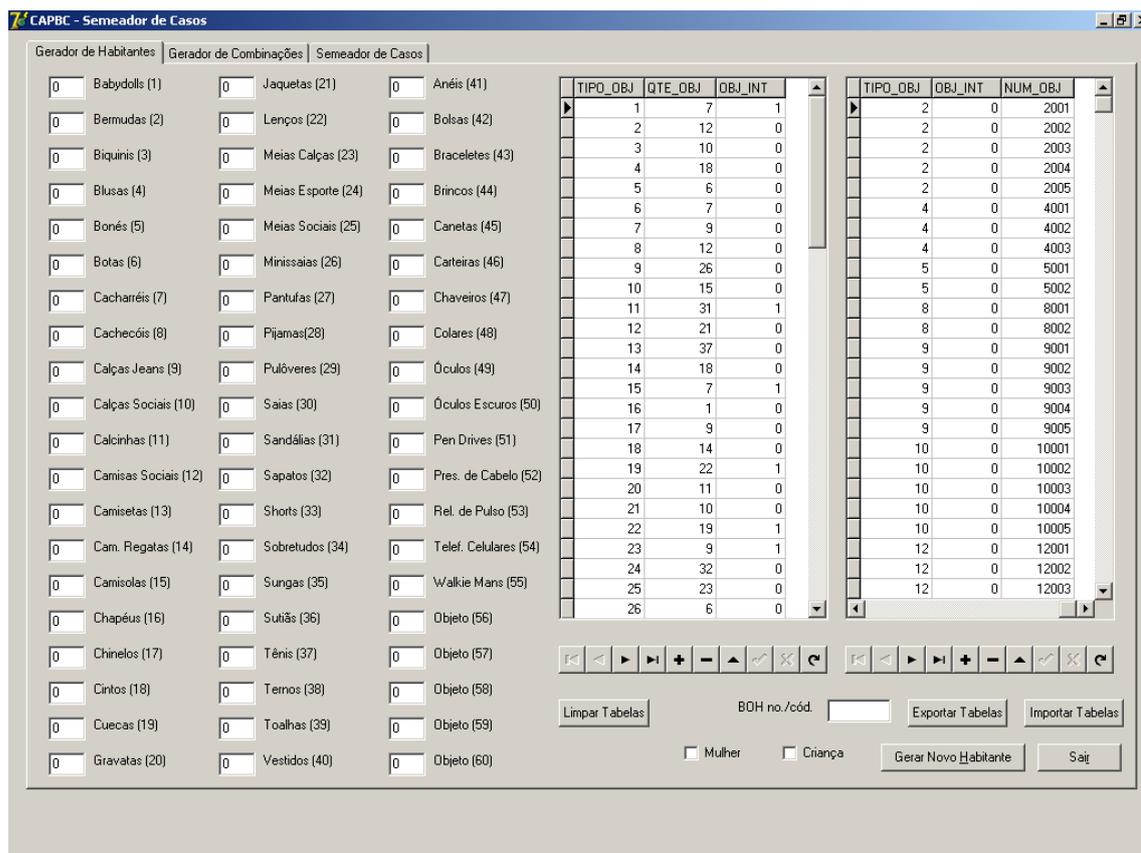


Figura 4.1 - Tela do software desenvolvido para gerar o banco de dados de objetos dos habitantes.

O tipo de habitante consiste em caracterizar cada tipo de combinação pelo sexo e faixa etária do habitante, ou seja, identificar quais combinações podem ser utilizadas por homens, mulheres ou ambos e por adultos, crianças ou ambos. Esta classificação é utilizada para possibilitar o correto cruzamento da tabela de objetos dos habitantes com a de combinações de objetos.

A Figura 4.2 mostra o segmento do simulador responsável pelo cadastro das combinações de objetos possíveis.

Resumidamente, as duas partes apresentadas até agora são responsáveis pelo cadastro dos objetos que estarão na base de casos, dos habitantes possuidores destes e das combinações (casos) possíveis de se construir a partir deles. Por serem procedimentos simples de gravação de valores em bancos de dados, esses algoritmos não foram detalhados.

A terceira parte do semeador de casos é o algoritmo que gera, efetivamente, a base de casos, consultando as tabelas de objetos cadastrados e de combinações. É possível semear combinações de objetos de habitantes específicos ou aleatoriamente. O algoritmo é encarregado de verificar se os tipos do habitante e da combinação escolhidos são compatíveis;

caso não sejam, a combinação não é gerada. O resultado deste algoritmo é a tabela utilizada para os testes do algoritmo classificador: a base de casos.

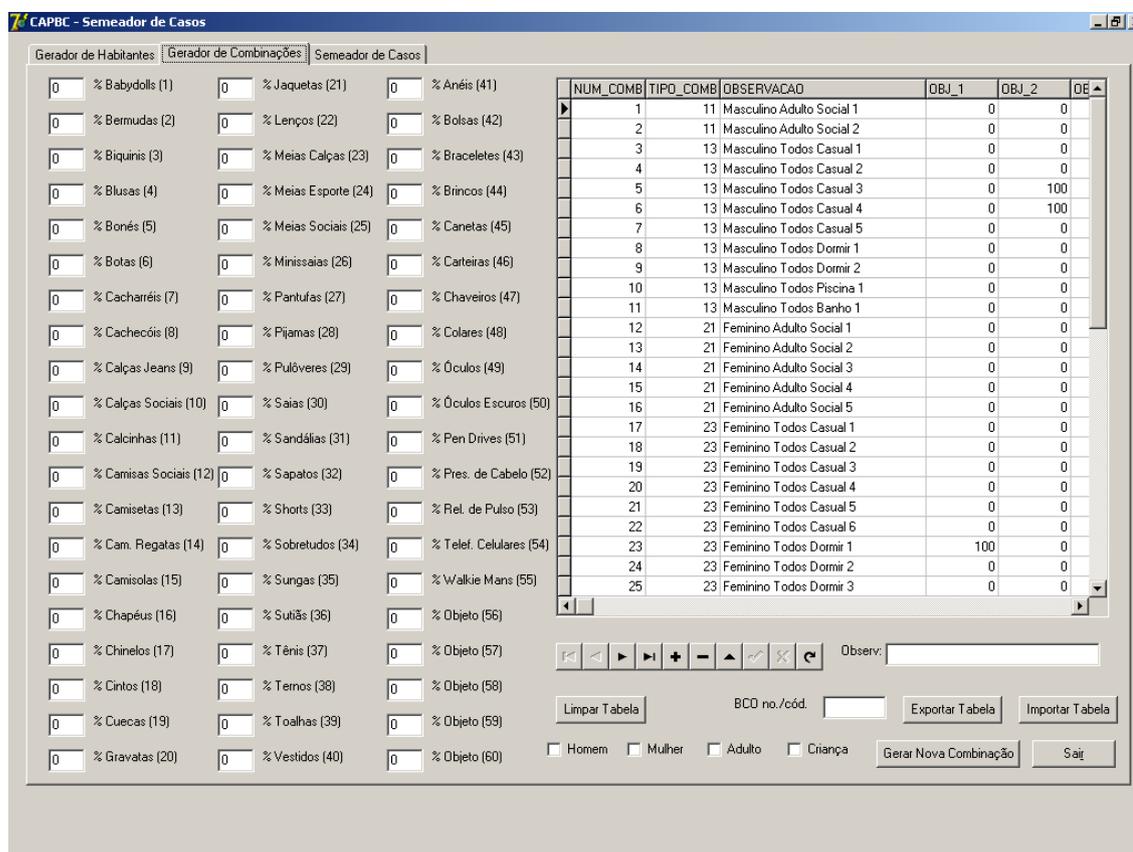


Figura 4.2 - Tela do software desenvolvido para gerar combinações de objetos possíveis.

O algoritmo da Figura 4.3 descreve o procedimento de geração de um novo caso de um habitante aleatório, usando uma combinação de roupas e objetos aleatória, baseada nas tabelas de objetos, habitantes e combinações cadastradas anteriormente. A base de casos que foi utilizada, posteriormente, nas simulações do sistema, é um conjunto de combinações geradas a partir deste algoritmo.

Para criar um novo caso $C_N = \{O_1, \dots, O_Q\}$ na BC, o algoritmo seleciona, aleatoriamente, um dos habitantes existentes NH e uma combinação de itens cadastradas NC. Se estes não forem compatíveis ($TH \neq TC$), ele seleciona novos valores até obter um par habitante-combinação que o seja.

Em seguida, para cada um dos Q tipos de objeto TO da combinação cadastrada (TO_1, \dots, TO_Q), o algoritmo compara a probabilidade PO de cada um dos tipos de objeto TO estar presente na combinação (PO_1, \dots, PO_Q) a um valor randômico R gerado. Se PO for maior do

que R, o algoritmo filtrará o banco de dados de objetos BDO pelo número do habitante NH e pelo tipo de objeto TO, selecionando, de forma aleatória, um dos códigos de objeto CO presentes em BDO com o filtro aplicado. Os códigos selecionados neste processo são armazenados nas posições O_1 a O_Q , no novo caso da BC.

Algoritmo para geração de uma nova combinação de objetos pelo algoritmo semeador de casos.

Variáveis:

BC: Base de Casos;

$O_1 - O_Q$: Campos da BC onde são armazenados os códigos numéricos dos objetos selecionados pelo semeador;

NH: Número do Habitante;

TH: Tipo do Habitante NH;

BDH: Banco de Dados de Habitantes;

NC: Número da Combinação;

TC: Tipo da Combinação NC;

BDC: Banco de Dados de Combinações;

$TO_1 - TO_Q$: Códigos dos tipos de objetos presentes em determinada combinação NC;

$PO_1 - PO_Q$: Probabilidade dos tipos de objetos presentes em determinada combinação NC;

CO: Código do Objeto;

TO: Tipo do Objeto CO;

BDO: Banco de Dados de Objetos, indexado por NH, TH e TO;

R: Variável auxiliar;

Início

```

1.  while TH  $\neq$  TC do                                // tipo do habitante compatível com o tipo de combinação
2.      Selecionar NH aleatório em BDH;
3.      Selecionar NC aleatório em BDC;
4.  for cada tipo de objeto  $TO_1$  a  $TO_Q$  presente em NC do
5.      Gerar R aleatório de 0 a 100;
6.      if  $PO \geq R$  then
7.          Filtrar BDO por NH e TO;
8.          Selecionar CO aleatório em BDO;
9.          Gravar em  $O$ , na BC: CO;
Fim.
```

Figura 4.3 - Algoritmo do semeador de casos a partir das tabelas de objetos e combinações

A Figura 4.4 mostra a tela do simulador responsável por gerar, gravar e modificar a base de casos, conforme os algoritmos apresentados, que foi utilizado para os testes do classificador de casos. E a Tabela 4.1 mostra um exemplo de um caso gerado a partir do algoritmo explicado anteriormente, mostrando os passos de seleção aleatória dos objetos que pertencerão, ou não, ao caso em questão.

Além de gerar as tabelas, o semeador de casos desenvolvido possui algumas rotinas que permitem alterar a base de casos de forma a simular algumas situações que poderiam ocorrer em uma residência real. São elas: empréstimo de objetos não íntimos, leitura de dois

habitantes ao mesmo tempo pelo sistema de RFID e acréscimo de objetos estranhos na base de casos. Estes são detalhados a partir da próxima seção.

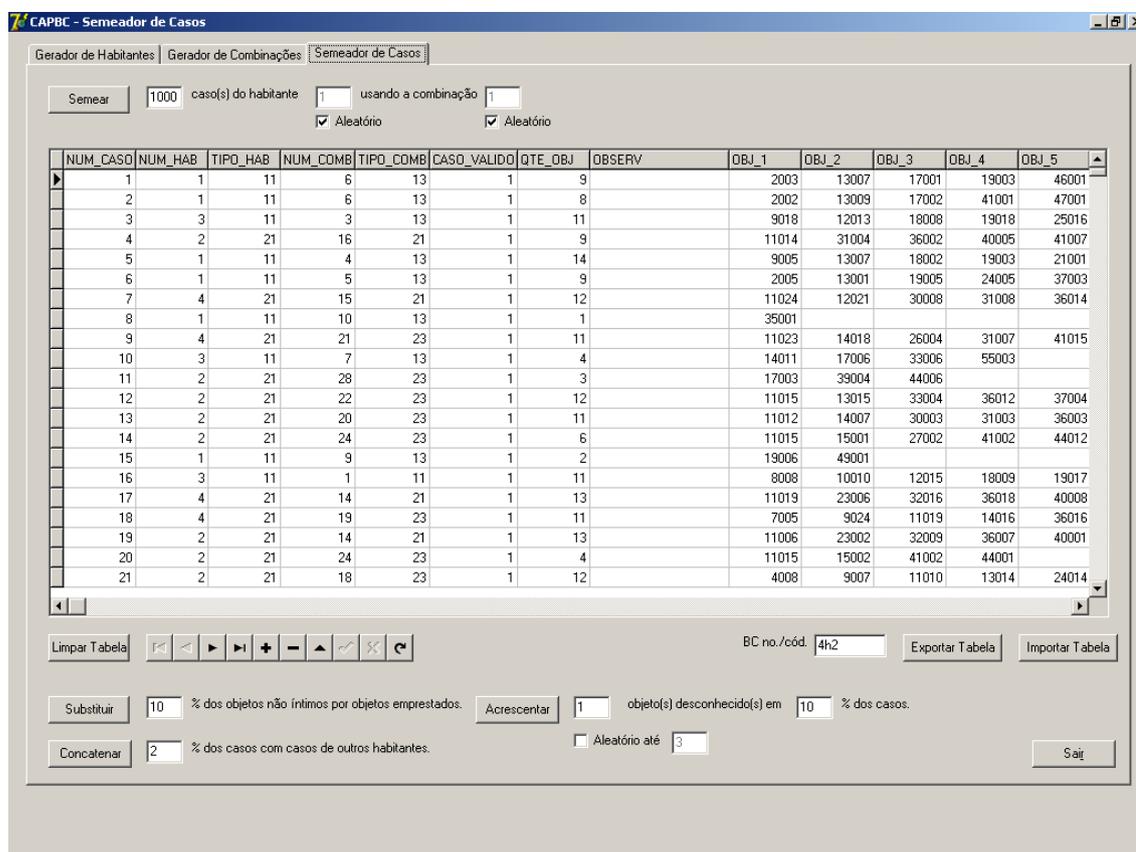


Figura 4.4 - Tela do software semeador da base de casos.

4.1.1.1 Simulação do empréstimo de objetos entre habitantes

O empréstimo de objetos permite que determinada porcentagem de objetos seja permutada aleatoriamente entre habitantes do mesmo sexo e faixa etária. Esta filtragem nas características é feita devido à incoerência de homens e mulheres ou adultos e crianças permutarem seus itens.

O algoritmo de modificação da base de casos, para simular o empréstimo de objetos entre habitantes, é mostrado na Figura 4.5.

Para a substituição de um item O , presente em um caso C_N de determinado habitante NH , por outro emprestado de um habitante diferente NH' , primeiramente, é verificado se este objeto trata-se de um item passível, ou não, de empréstimo. Isto é feito verificando-se uma flag associada ao tipo de objeto, armazenada no banco de dados de objetos.

Tabela 4.1 – Exemplo de criação de um novo caso a partir de um habitante e uma combinação, selecionados aleatoriamente.

Número aleatório do Habitante (NH): 1							Tipo do Habitante (TH): 11 (Masculino - Adulto)						
Número aleatório da Combinação (NC): 6							Tipo da Combinação (TC): 13 (Masculino - Qualquer Idade)						
Habitante e Combinação compatíveis: SIM													
Probabilidade (PO) da presença de cada tipo de objeto na combinação 6													
Bermuda	Boné	Camizeta	Chinelos	Cueca	Lenço	Anel	Carteira	Chaveiro	Óculos	Óc. Esc.	Relógio	Tel. Cel.	Walkie M.
100%	10%	90%	90%	90%	10%	70%	40%	30%	50%	40%	30%	70%	20%
Número (R) Aleatório entre 0 e 100													
91	43	32	66	71	73	86	21	29	58	37	5	55	42
Lista de Objetos do habitante 1 por tipo													
Bermuda	Boné	Camizeta	Chinelos	Cueca	Lenço	Anel	Carteira	Chaveiro	Óculos	Óc. Esc.	Relógio	Tel. Cel.	Walkie M.
2001	5001	13001	17001	19001	22001	41001	46001	47001	49001	50001	53001	54001	55001
2002	5002	13002	17002	19002	22002					50002	53002		
2003		13003		19003	22003						53003		
2004		13004		19004	22004						53004		
2005		13005		19005	22005								
		13006		19006	22006								
		13007		19007									
		13008		19008									
		13009		19009									
		13010		19010									
Número Aleatório entre 1 e o total de objetos de cada tipo para objetos onde PO > R													
3		7	1	3			1	1		2	3	1	
Objeto selecionado para o caso semeado													
2003		13007	17001	19003			46001	47001		50002	53003	54001	

Caso o objeto seja passível de empréstimo, é gerado um valor randômico R, de 0 a 100, e comparado à porcentagem P, especificada, de itens a serem substituídos.

Se a porcentagem for superior ou igual ao valor gerado ($P \geq R$), é aplicado um filtro ao banco de dados de objetos, a fim de selecionar apenas objetos de habitantes diferentes do indivíduo do caso em manutenção, e objetos e habitantes do mesmo tipo deste. A partir deste subconjunto do banco de dados, é selecionado um novo código de objeto, que é gravado no lugar do item original. Este procedimento é efetuado para todos os objetos presentes no caso.

4.1.1.2 Simulação da leitura de dois habitantes simultaneamente

A leitura de dois habitantes simultaneamente serve para simular uma falha no sistema de RFID, que poderia ocorrer quando os habitantes passassem muito próximos entre si do leitor RFID. O sistema entenderia a leitura de todos os itens como uma única combinação, o que pode confundir o sistema classificador.

O algoritmo de modificação da base de casos, para simular a leitura simultânea de duas combinações de habitantes distintos, é mostrado na Figura 4.6. Este consiste em

percorrer a BC, concatenando, aleatoriamente, segundo uma porcentagem pré-definida, casos subseqüentes, pertencentes a diferentes habitantes.

Algoritmo para substituição de certa porcentagem P de objetos de um caso C, por itens de outros habitantes

Variáveis:

C_N : Caso Atual;

BC: Base de Casos;

$O_1 - O_Q$: Campos da BC onde são armazenados os códigos numéricos dos objetos selecionados pelo sementeador;

NH: Número do Habitante;

TH: Tipo do Habitante NH;

CO: Código do Objeto;

TO: Tipo do Objeto CO;

BDO: Banco de Dados de Objetos, indexado por NH, TH e TO;

P: Porcentagem de objetos a ser substituída;

R: Variável auxiliar.

Início

1. **for** cada objeto O_1 a O_Q presente em C_N , **do**
2. **if** O é objeto não-intimo, **then**
3. Gerar R aleatório de 0 a 100;
4. **if** $P \geq R$ **then**
5. Filtrar BDO por $NH(BDO) \neq NH(C)$, TH e TO;
6. Selecionar CO aleatório em BDO;
7. Gravar em O, na BC: CO;

Fim.

Figura 4.5 – Algoritmo para substituição de uma porcentagem P de objetos por itens emprestados, em um caso C_N .

Se o caso posterior C_{N+1} da base for de um habitante diferente do caso atual C_N , um valor randômico R é gerado e comparado à porcentagem especificada P de casos a serem concatenados. Caso a porcentagem seja maior ou igual ao o valor gerado ($P \geq R$), os objetos de C_{N+1} são copiados para as próximas posições livre de C_N . Os códigos são, então, reordenados e a posição C_{N+1} é excluída da BC. Desta forma, o caso C_N passará a conter os objetos dos casos C_N e C_{N+1} .

4.1.1.3 Simulação da adição de objetos novos e estranhos aos casos

A adição de objetos estranhos consiste em acrescentar objetos de códigos desconhecidos em combinações aleatórias, de forma a simular a aquisição ou porte de um novo objeto por determinado habitante, para verificar o comportamento do classificador

nestas situações, inclusive com os mesmos objetos podendo se repetir em mais de um caso da BC.

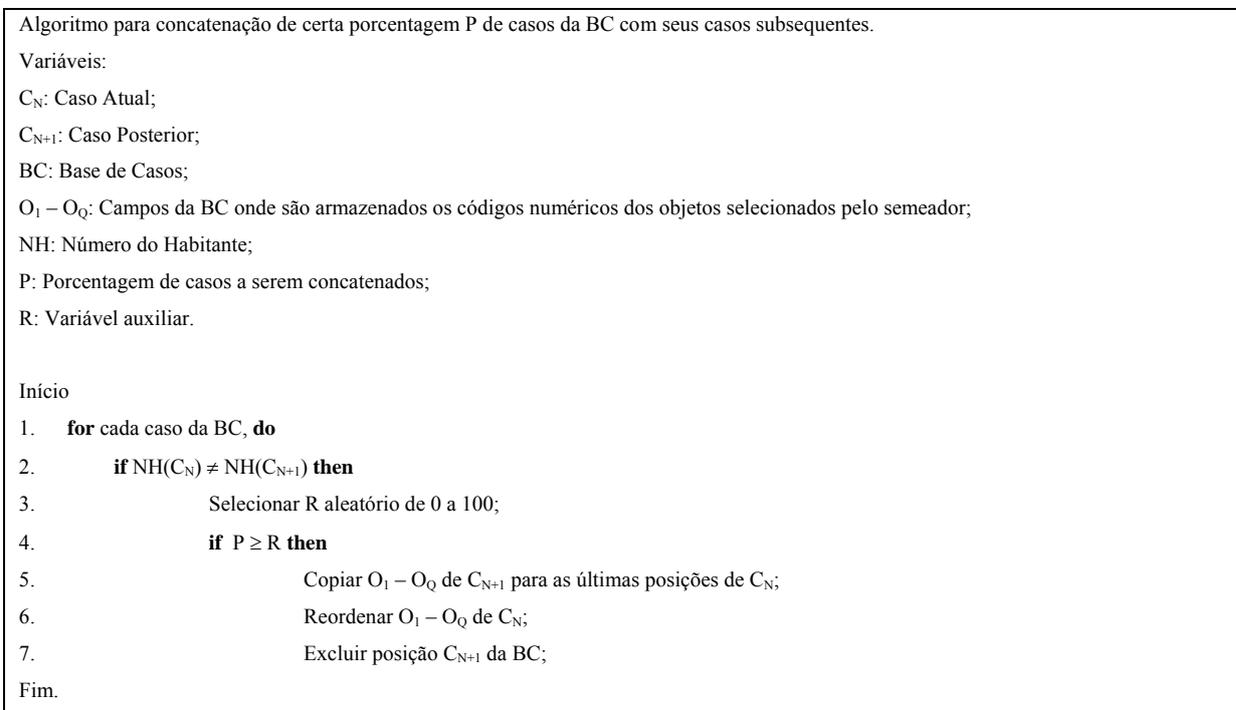


Figura 4.6 – Algoritmo para concatenação de uma porcentagem P de combinações com seus casos subsequentes, de outros habitantes.

O algoritmo de modificação da base de casos, para simular adição de objetos novos ou estranhos nas combinações, é mostrado na Figura 4.7.

Primeiramente, é gerado um valor aleatório R1 para a quantidade de objetos emprestados que serão inseridos no caso em manutenção, entre 1 e a quantidade limite QL. Posteriormente, é consultada a quantidade de objetos estranhos QOE, que já estiveram presentes em algum caso anterior, armazenados em um banco de dados próprio, denominado BDOE. Este valor é copiado para uma variável auxiliar de contagem QOE’.

A rotina a seguir é, então, executada R1 vezes, a fim de inserir R1 objetos estranhos no caso em manutenção.

Se QOE’ for maior do que ‘0’, será inserido no caso um objeto já existente no banco de dados de objetos estranhos BDOE. Para isso, um valor aleatório entre 1 e QOE’ é gerado. Será então selecionado o código de objeto armazenado na posição “QOE – QOE’ + R1” do banco de dados de objetos estranhos BDOE e copiado para a próxima posição livre O_{N+1} do caso. O valor de R1 é então subtraído de QOE’.

Se QOE' for igual a '0', um novo objeto estranho será criado e inserido no banco de dados de objetos estranhos e no caso. O valor de QOE é incrementado em '1', um código CO para o novo objeto é criado, e este é armazenado na próxima posição livre de $BDOE$ (QOE) e na próxima posição livre O_{Q+1} do caso.

Algoritmo para acréscimo de certa quantidade de novos objetos, podendo variar de 1 a QL , em um caso C_N .

Variáveis:

C_N : Caso Atual;

BC : Base de Casos;

$O_1 - O_Q$: Campos da BC onde são armazenados os códigos numéricos dos objetos selecionados pelo semeador;

NH : Número do Habitante;

TH : Tipo do Habitante NH ;

CO : Código do Objeto;

$BDOE$: Banco de Dados de Objetos Estranhos;

QOE : Quantidade de Objetos Estranhos armazenados no banco de dados;

QL : Quantidade limite de objetos estranhos a ser inserida no caso;

$N, R1, R2, QOE'$: Variáveis auxiliares.

Início

```

1. Gerar R1 aleatório de 1 a QL;
2.  $QOE' = QOE$ ;
3. for N = 1 to R1 do
4.     if  $QOE' > 0$  then
5.         Gerar R2 aleatório de 1 a  $QOE'$ 
6.         Selecionar CO, na posição  $QOE - QOE' + R2$  de  $BDOE$ ;
7.         Gravar em  $O_{Q+1}$ , de  $C_N$ : CO;
8.          $QOE' = QOE' - R2$ ;
9.     else
10.         $QOE = QOE + 1$ ;
11.        Gerar CO aleatório;
12.        Gravar na posição  $QOE$ , em  $BDOE$ : CO;
13.        Gravar em  $O_{Q+1}$ , de  $C_N$ : CO;
Fim.
```

Figura 4.7 – Algoritmo para inserção de objetos estranhos a um caso C_N .

4.1.2 Classificador de casos

O objetivo deste software é testar a lógica proposta para o sistema CAPBC, utilizando uma base de casos gerada pelo software semeador. Conforme exposto anteriormente, há diferentes critérios possíveis para as classificações. O simulador desenvolvido visa testar estes parâmetros com diferentes valores de similaridade para determinar quais são eficazes e quais não para o objetivo do trabalho. Seu algoritmo foi apresentado da Figura 3.4, no capítulo anterior. A Figura 4.8 mostra o software desenvolvido para a classificação de casos.

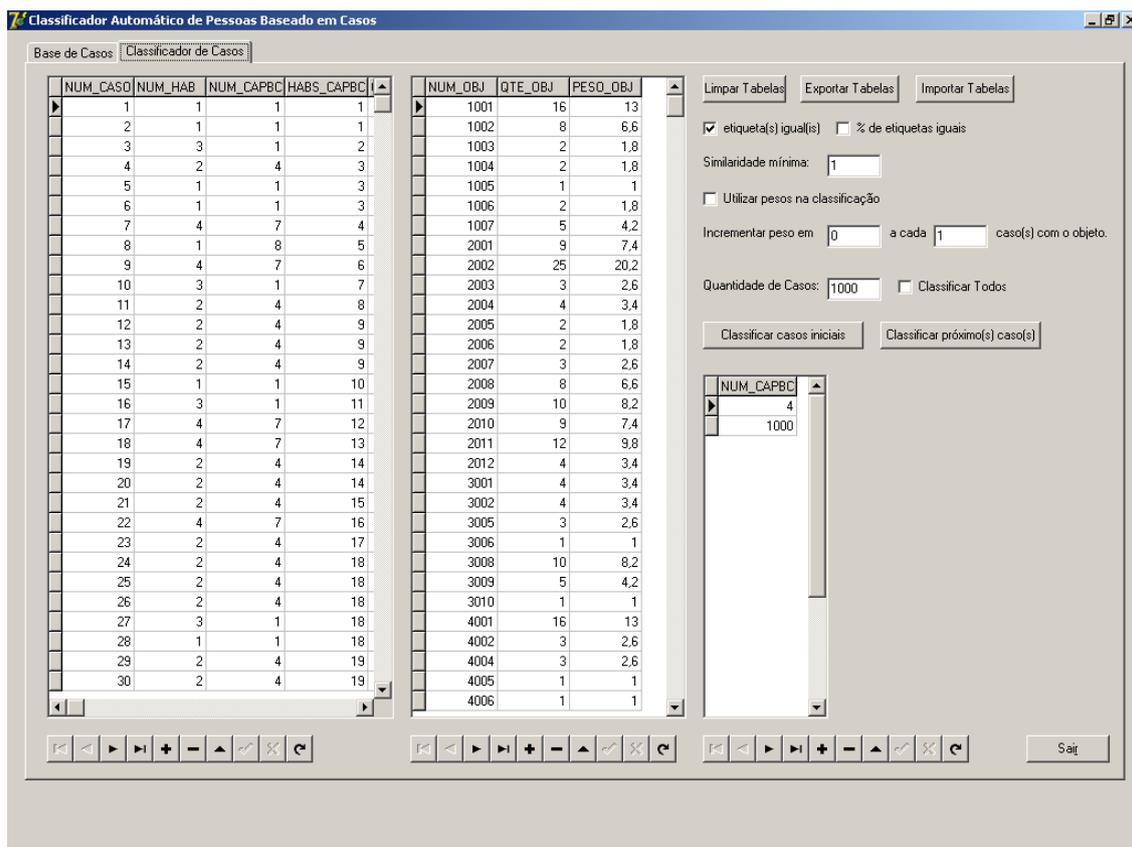


Figura 4.8 - Tela do software classificador de casos.

Os parâmetros que o software permite configurar são:

- 1 – Critério de similaridade;
- 2 – Valor de similaridade mínima;
- 3 – Número de casos a ser classificado;
- 4 – Uso, ou não, de pesos na classificação;
- 5 – Critérios para incremento dos pesos;
- 6 – Forma de classificação dos casos com o uso de pesos;

A configuração destes parâmetros no software é detalhada nas seções 4.1.2.1 a 4.1.2.6.

4.1.2.1 Critério de similaridade

No software, é possível escolher o cálculo de similaridade a ser utilizado pelo simulador, entre o número de objetos iguais e a porcentagem de objetos iguais entre os casos. Estes critérios utilizam as equações apresentadas nas seções 3.7.1 e 3.7.2.

4.1.2.2 Valor de similaridade mínima

O valor de similaridade mínima deve ser informado de acordo com o critério de similaridade selecionado. Se selecionado o número de objetos iguais, o valor de similaridade mínima deve ser um número inteiro maior ou igual a 1; se selecionada a porcentagem de objetos iguais, o valor de similaridade mínima deve ser um número real entre 0 e 100.

A partir do critério e valor escolhidos, a classificação dos casos é efetuada pelo algoritmo classificador, exposto na seção 3.4.

4.1.2.3 Número de casos a ser classificado

Este parâmetro permite escolher se o software deve classificar toda a base de casos, ou apenas um subconjunto dela. Ele permite, também, efetuar testes alterando-se critérios e valores utilizados durante a classificação de uma base de casos. Por exemplo, pode-se selecionar um conjunto diferente de parâmetros para classificar os primeiros 500 casos de uma base e modificá-los para classificar o restante.

4.1.2.4 Uso, ou não, de pesos nas classificações

É possível selecionar, no software, o uso de pesos nas classificações para o critério de porcentagem de objetos iguais. Quando este é selecionado, o cálculo de similaridade efetuado pelo software é o apresentado na seção 3.7.3 deste trabalho.

4.1.2.5 Critérios para incremento dos pesos

Há dois parâmetros que podem ser configurados em relação ao incremento dos pesos: o valor do incremento do peso e o intervalo entre casos para a aplicação deste incremento.

Quando um novo objeto é verificado em um caso, o peso inicial deste é igual a 1. Se, no software, está configurado que o peso deve ser incrementado em 0,1 a cada 1 caso contendo o objeto, quando este for verificado pela segunda vez, o peso do objeto é atualizado para 1,1, ou seja, o peso anterior somado ao incremento configurado.

A Equação 4.1 mostra o cálculo para a atualização do peso P de um determinado objeto O , a partir de um valor pré-estipulado de incremento I .

$$P(O) = P(O) + I \quad (4.1)$$

4.1.2.6 Forma de classificação dos casos com o uso de pesos

Quando é considerada a aplicação de pesos aos objetos, os casos podem ser classificados de duas formas diferentes: incrementando-se os pesos dos objetos, quando um novo caso for recebido, e classificando-o na seqüência, ou incrementando-se os pesos dos objetos quando um novo caso for recebido, mas classificando todos os casos apenas após a entrada do 'N'-ésimo caso, onde N é um valor configurado no software.

Esta segunda forma permite verificar se a abstração de informações das combinações (no caso, relevância de cada objeto), antes de iniciar as classificações, pode melhorar ou piorar a qualidade destas.

A primeira forma apresentada é denominada, a partir deste momento, critério 1, e a segunda forma, critério 2 de classificação com pesos.

4.1.3 Avaliador de classificações

O avaliador de classificações consiste em comparar os resultados obtidos pelo algoritmo classificador aos resultados esperados (armazenados na BC pelo algoritmo semeador de casos), a fim de avaliar a qualidade das classificações para cada conjunto de parâmetros testado. Este algoritmo gerará dados numéricos que permitirão a comparação das classificações por meio de tabelas e gráficos, os quais são expostos posteriormente.

A figura 4.9 mostra o software desenvolvido para avaliar as classificações obtidas pelo algoritmo classificador.

O resultado da avaliação feita por este software é, basicamente, a contagem da quantidade de classificações corretas, incompletas e erradas feitas pelo CAPBC, utilizando-se determinados critérios de classificação. Estes valores permitem comparar os resultados do classificador com cada critério de similaridade utilizado, bem como avaliar seu comportamento, quando a base de casos é modificada, por meio das rotinas específicas do software semeador de casos.

Na próxima seção são definidos os termos utilizados para a avaliação das classificações, bem como apresentado o fluxograma utilizado para construção do software avaliador.

4.2 Definições utilizadas na avaliação de resultados

Para a avaliação dos resultados, deve-se lembrar que o sistema CAPBC não identifica os habitantes, apenas os agrupa, conforme critérios de similaridade pré-definidos. Desta forma, não é possível tirar conclusões sobre uma classificação, avaliando-a isoladamente, pois o código de habitante atribuído pelo classificador é genérico. A avaliação deverá ser feita analisando-se todas as classificações, simultaneamente.

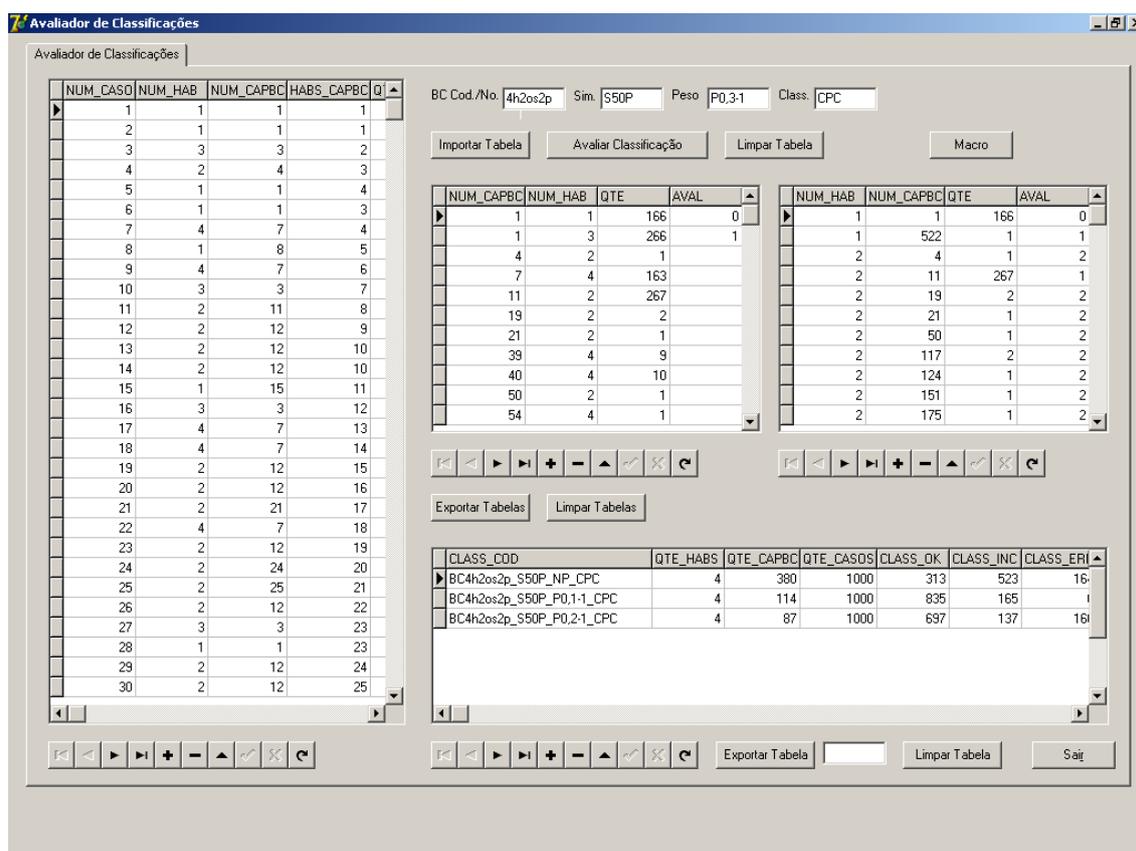


Figura 4.9 - Tela do software avaliador de classificações.

Foram definidas três avaliações para as classificações:

1 - Classificação correta: uma classificação é considerada correta quando casos do mesmo habitante forem agrupados sob o mesmo código genérico pelo classificador. Como pode ocorrer do classificador gerar subgrupos do mesmo habitante, são consideradas corretas as classificações do subgrupo maior;

2 - Classificação incompleta: quando o classificador gerar subgrupos distintos de casos do mesmo habitante, as classificações dos casos pertencentes aos subgrupos menores são denominadas incompletas;

3 - Classificação errada: uma classificação é considerada errada quando um ou mais casos forem agrupados a casos de habitantes diferentes. Se subgrupos de habitantes diferentes forem agrupados pelo classificador, os subgrupos menores são avaliados como classificações erradas; o subgrupo maior poderá ser avaliado como classificação correta ou incompleta.

A Figura 4.10 mostra o fluxograma utilizado pelo software avaliador de classificações.

Inicialmente, o algoritmo avaliador busca por todas as classificações erradas da base de casos. Para isto, o avaliador busca códigos de habitantes iguais atribuídos a habitantes diferentes. Os subconjuntos com menos casos, classificados sob este código, são avaliados como classificações erradas. O subconjunto maior é re-avaliado juntamente com os conjuntos de casos que não possuem códigos de habitantes iguais atribuídos a diferentes habitantes.

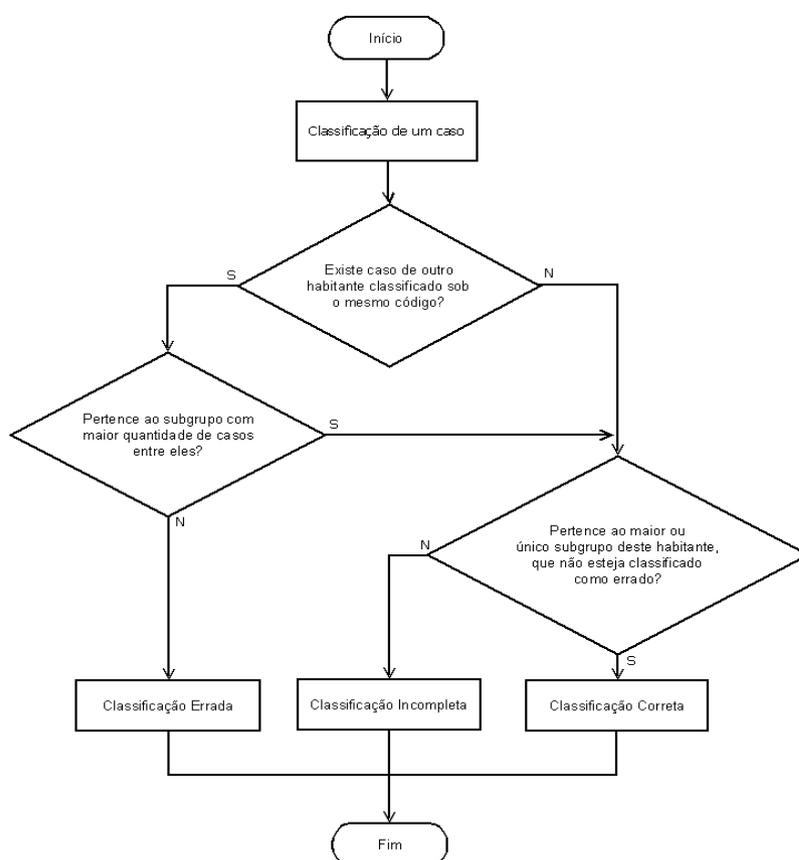


Figura 4.10 – Fluxograma do software avaliador de classificações.

Depois de identificadas todas as classificações erradas, o algoritmo avaliará as demais classificações. O maior subgrupo de classificações de cada habitante, desde que este não tenha sido avaliado anteriormente como classificação errada, é avaliado como classificação correta. Caso seja um subgrupo menor, que também não tenha sido avaliado como classificação errada, é avaliado como classificação incompleta.

Nota-se que o fluxograma utilizado para criação do avaliador de classificações utiliza uma estratégia otimista de avaliação, ou seja, os subgrupos maiores de classificações sempre tendem a ser avaliados como corretos e os subgrupos menores como errados e incompletos, a fim de se obter o maior número possível de classificações corretas.

A Figura 4.11 ilustra um conjunto de casos hipotético, classificado e avaliado conforme as nomenclaturas utilizadas. Pode-se notar, no exemplo, que os maiores subconjuntos dos habitantes '1' e '2' foram incorretamente agrupados, porém as classificações dos casos do habitante '1' foram avaliadas como corretas, por serem do maior subgrupo entre eles, e as classificações dos casos do habitante '2' como erradas, por serem do subgrupo menor.

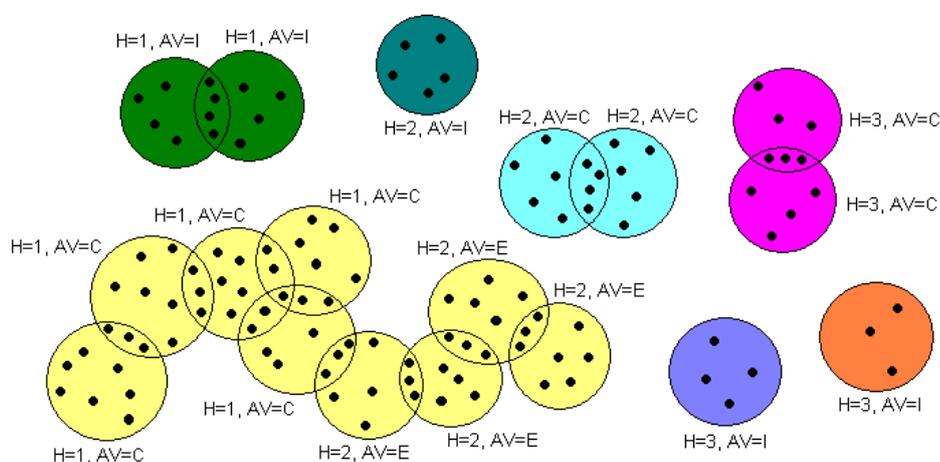


Figura 4.11 - Ilustração de um conjunto de casos com sua classificação avaliada. Cada círculo representa um caso, os pontos pretos representam os objetos, as cores representam a classificação sob um mesmo código de habitante, o valor de H representa o habitante real a quem a combinação pertence e o valor de AV representa a avaliação da classificação: C = correta, I = incompleta, E = errada.

5 TESTES REALIZADOS

Devido ao propósito do uso do sistema desenvolvido em um sistema de automação residencial real, os testes efetuados nos simuladores focaram os seguintes pontos principais: desempenho do sistema classificador e qualidade das classificações.

O desempenho do classificador se refere a dois parâmetros: tempo necessário para classificar um novo caso e número de casos necessários para que as classificações converjam para o número correto de habitantes.

A qualidade das classificações se refere às quantidades de classificações corretas, erradas e incompletas, após determinado número de casos ser processado pelo sistema.

O funcionamento ideal prevê a classificação correta de todas as combinações após a menor quantidade de casos possível presentes na base de casos, de forma a possibilitar que o sistema de automação possa agrupar as regras geradas de acordo com o habitante detectado, também no menor tempo possível.

A partir da próxima seção são expostos todos os dados e parâmetros utilizados nas simulações, bem como sua análise e conclusões sobre o funcionamento do sistema.

5.1 Base de Casos semeada para as simulações

Para que fosse possível comparar os resultados obtidos com diferentes parâmetros de classificação, foi gerada uma base de casos única utilizada para todas as simulações. Desta forma, foi possível analisar o comportamento do sistema com o uso de diferentes critérios e valores de testes, sem a influência de particularidades entre diferentes bases de casos.

Para a geração desta base, foram criados quatro habitantes, dois homens adultos e duas mulheres adultas. Foram escolhidos estes tipos de habitante para possibilitar a simulação do empréstimo de objetos entre um homem e o outro e entre uma mulher e a outra.

A quantidade de cada tipo de objeto pertencente aos habitantes foi escolhida de forma a representar uma situação condizente com o que existiria em uma residência real. A Tabela 5.1 mostra os tipos e a quantidade de objetos de cada habitante gerado para as simulações.

A partir desta tabela, foi gerado um banco de dados de objetos com códigos numéricos individuais para cada objeto, representando sua identidade RFID. Este banco de dados é utilizado, posteriormente, pelo software semeador de casos para gerar a base de casos. A tabela 5.2 mostra um segmento do banco de dados de objetos gerado.

Tabela 5.1: Quantidade de objetos por habitante.

Cód.	Tipo de Objeto	Habit. 1 (Qte.)	Habit. 2 (Qte.)	Habit. 3 (Qte.)	Habit. 4 (Qte.)
1	Babydoll		3		4
2	Bermuda	5		7	
3	Biquini		5		5
4	Blusa	3	5	4	6
5	Boné	2		4	
6	Botas		4		3
7	Cacharrel		4		5
8	Cachecol	2	4	3	3
9	Calça Jeans	5	6	8	7
10	Calça Social	5	3	3	4
11	Calcinha		15		16
12	Camisa Social	8	4	5	4
13	Camiseta	10	5	16	6
14	Cam. Regata	4	3	6	5
15	Camisola		3		4
16	Chapéu	1			
17	Chinelos	2	2	2	3
18	Cinto	4	3	3	4
19	Cueca	10		12	
20	Gravata	8		3	
21	Jaqueta	2	3	2	3
22	Lenço	6	4	4	5
23	Meias Calças		4		5
24	Meias Esporte	8	6	10	8
25	Meias Sociais	8	6	4	5
26	Minissala		2		4
27	Pantufas		2		4
28	Pijama	2	3	2	4
29	Pulôver	2		1	2
30	Sala		4		5
31	Sandálias		5		4
32	Sapatos	4	6	2	4
33	Short	3	2	4	3
34	Sobretudo	2	1	1	
35	Sunga	2		4	
36	Sutiã		12		15
37	Tênis	3	2	5	3
38	Terno	4		2	
39	Toalha	2	3	2	3
40	Vestido		5		4
41	Anel	1	6		10
42	Bolsa		5		6
43	Bracelete		6	3	10
44	Brincos		12	3	15
45	Caneta	2	2	2	1
46	Carteira	1	1	1	1
47	Chaveiro	1	1	1	1
48	Colar		5		8
49	Óculos	1	1		
50	Óculos Escuros	2	2	2	3
51	Pen Drive	1		1	1
52	Presilha de Cabelo		8		9
53	Relógio de Pulso	4	3	2	2
54	Telefone Celular	1	1	1	1
55	Walkie Man	1		2	1

Tabela 5.2: Banco de dados de objetos por habitante.

Núm. do Hab.	Tipo do Hab.	Tipo do Obj.	ID do Obj.
1	Masculino / Adulto	2	2001
1	Masculino / Adulto	2	2002
1	Masculino / Adulto	2	2003
1	Masculino / Adulto	2	2004
1	Masculino / Adulto	2	2005
1	Masculino / Adulto	4	4001
1	Masculino / Adulto	4	4002
1	Masculino / Adulto	4	4003
1	Masculino / Adulto	5	5001
1	Masculino / Adulto	5	5002
1	Masculino / Adulto	8	8001
1	Masculino / Adulto	8	8002
1	Masculino / Adulto	9	9001
1	Masculino / Adulto	9	9002
1	Masculino / Adulto	9	9003
1	Masculino / Adulto	9	9004
1	Masculino / Adulto	9	9005
...
4	Feminino / Adulto	50	50007
4	Feminino / Adulto	50	50008
4	Feminino / Adulto	50	50009
4	Feminino / Adulto	51	51003
4	Feminino / Adulto	52	52009
4	Feminino / Adulto	52	52010
4	Feminino / Adulto	52	52011
4	Feminino / Adulto	52	52012
4	Feminino / Adulto	52	52013
4	Feminino / Adulto	52	52014
4	Feminino / Adulto	52	52015
4	Feminino / Adulto	52	52016
4	Feminino / Adulto	52	52017
4	Feminino / Adulto	53	53010
4	Feminino / Adulto	53	53011
4	Feminino / Adulto	54	54004

Para o software semeador de casos poder gerar a base de casos a partir do banco de dados de objetos, este deve consultar uma tabela contendo tipos de combinações possíveis para cada tipo de habitante. Para gerar a base de casos utilizada nas simulações apresentadas neste trabalho, foram declarados 28 tipos de combinações possíveis de objetos, lembrando que, devido a cada tipo de objeto ter uma porcentagem associada de chance de existir na combinação, a variedade de tipos de combinações é, na realidade, bem maior. A tabela 5.3 mostra um segmento do banco de dados de combinações criado. No Apêndice 2 pode ser verificada a lista completa de combinações utilizadas para a geração da base de casos, com a descrição dos objetos.

A partir dos bancos de dados de objetos por habitante e de combinações, o software semeador de casos foi capaz de gerar a base de casos para os testes de classificação.

Foram gerados 1000 casos, onde o algoritmo selecionou os tipos de combinação e de habitante aleatoriamente, recordando que o semeador de casos não gera casos onde esses tipos são incompatíveis.

Tabela 5.3: Banco de dados de combinações possíveis de objetos.

Num. Comb.	Tipo da Comb.	Obs.	OBJ_1	OBJ_2	OBJ_3	OBJ_4	OBJ_5	...	OBJ_51	OBJ_52	OBJ_53	OBJ_54	OBJ_55
1	Masculino / Adulto	Social 1	0	0	0	0	0	...	70	0	80	90	20
2	Masculino / Adulto	Social 2	0	0	0	50	0	...	70	0	80	90	40
3	Masculino / Todos	Casual 1	0	0	0	20	0	...	40	0	80	90	40
4	Masculino / Todos	Casual 2	0	0	0	10	10	...	10	0	60	90	40
5	Masculino / Todos	Casual 3	0	100	0	0	10	...	0	0	30	80	20
6	Masculino / Todos	Casual 4	0	100	0	0	10	...	0	0	30	70	20
7	Masculino / Todos	Casual 5	0	0	0	0	30	...	0	0	20	40	30
8	Masculino / Todos	Dormir 1	0	0	0	0	0	...	0	0	0	0	0
9	Masculino / Todos	Dormir 2	0	0	0	0	0	...	0	0	0	0	0
10	Masculino / Todos	Piscina 1	0	0	0	0	0	...	0	0	0	0	0
11	Masculino / Todos	Banho 1	0	0	0	0	0	...	0	0	0	0	0
12	Feminino / Adulto	Social 1	0	0	0	40	0	...	30	60	70	50	20
13	Feminino / Adulto	Social 2	0	0	0	0	0	...	30	60	70	50	20
14	Feminino / Adulto	Social 3	0	0	0	0	0	...	30	60	70	50	20
15	Feminino / Adulto	Social 4	0	0	0	0	0	...	30	60	70	50	20
16	Feminino / Adulto	Social 5	0	0	0	0	0	...	30	60	70	50	20
17	Feminino / Todos	Casual 1	0	0	0	70	0	...	30	60	70	50	20
18	Feminino / Todos	Casual 2	0	0	0	50	0	...	30	60	70	50	20
19	Feminino / Todos	Casual 3	0	0	0	0	0	...	30	60	70	50	20
20	Feminino / Todos	Casual 4	0	0	0	0	0	...	30	60	70	50	20
21	Feminino / Todos	Casual 5	0	0	0	0	0	...	30	60	70	50	20
22	Feminino / Todos	Casual 6	0	0	0	0	0	...	30	60	70	50	20
23	Feminino / Todos	Dormir 1	100	0	0	0	0	...	0	30	0	0	0
24	Feminino / Todos	Dormir 2	0	0	0	0	0	...	0	30	0	0	0
25	Feminino / Todos	Dormir 3	0	0	0	0	0	...	0	30	0	0	0
26	Feminino / Todos	Dormir 4	0	0	0	0	0	...	0	30	0	0	0
27	Feminino / Todos	Piscina 1	0	0	100	0	0	...	0	30	0	0	0
28	Feminino / Todos	Banho 1	0	0	0	0	0	...	0	50	0	0	0

Esta quantidade de casos poderia ser diferente, mas 1000 combinações (uma média de 250 leituras de cada habitante) já devem refletir o comportamento esperado do sistema, considerando um número razoável de repetições no uso dos itens de vestuário e objetos pessoais de cada morador.

É possível que, em uma residência com mais habitantes e onde estes possuíssem mais objetos, uma quantidade maior de casos fosse necessária para obter os mesmos resultados das simulações feitas neste trabalho. Porém, não é objetivo deste avaliar a variação no comportamento do sistema, conforme a base de casos utilizada, mas, sim, estudar a sua capacidade de classificar novos casos a partir do conhecimento de casos passados, utilizando diversos critérios e valores de similaridade e observando variáveis como o tempo e quantidades de acertos e erros nas classificações. A tabela 5.4 mostra parte da base de casos gerada, que foi utilizada nas simulações.

Além dos dados visíveis na base de casos ilustrada nesta tabela, outros estão presentes naquela gerada pelo semeador, como o número do habitante, o tipo de habitante e a quantidade de objetos presentes na combinação. Estes valores não foram utilizados pelo

CAPBC, obviamente, pois representam as soluções das classificações, mas foram utilizados, sim, posteriormente, para a avaliação destas.

Esta base de casos é denominada, a partir deste momento, ‘base de casos original’. Isto porque ela foi modificada, posteriormente, para efetuar os testes de comportamento do classificador, quando objetos são substituídos ou acrescentados em diferentes casos.

Tabela 5.4: Segmento inicial das 32 primeiras combinações da BC gerada pelo software semeador de casos.

Num. Caso	Obj. 1	Obj. 2	Obj. 3	Obj. 4	Obj. 5	Obj. 6	Obj. 7	Obj. 8	Obj. 9	Obj. 10	Obj. 11	Obj. 12	...
1	2003	13007	17001	19003	46001	47001	50002	53003	54001				...
2	2002	13009	17002	41001	47001	50002	53002	54001					...
3	9018	12013	18008	19018	25016	32012	46003	47003	53009	54003	55003		...
4	11014	31004	36002	40005	41007	44002	48003	50003	52002				...
5	9005	13007	18002	19003	21001	22003	24007	37001	41001	46001	49001	53001	...
6	2005	13001	19005	24005	37003	46001	49001	53003	54001				...
7	11024	12021	30008	31008	36014	41009	43017	44023	46004	48012	52017	53010	...
8	35001												...
9	11023	14018	26004	31007	41015	42009	43016	44028	52011	53011	54004		...
10	14011	17006	33006	55003									...
11	17003	39004	44006										...
12	11015	13015	33004	36012	37004	41002	42001	43004	44008	49002	52003	53007	...
13	11012	14007	30003	31003	36003	41002	42001	43005	44010	49002	52006		...
14	11015	15001	27002	41002	44012	49002							...
15	19006	49001											...
16	8008	10010	12015	18009	19017	25018	32011	38005	46003	47003	53008		...
17	11019	23006	32016	36018	40008	41014	42009	44030	45007	46004	50007	52015	...
18	7005	9024	11019	14016	36016	37012	41010	44017	52011	53010	54004		...
19	11006	23002	32009	36007	40001	41005	42004	43006	44006	48003	50004	52005	...
20	11015	15002	41002	44001									...
21	4008	9007	11010	13014	24014	37004	41006	43006	44007	48002	50003	53007	...
22	11024	13034	24028	33010	36019	37011	41011	43018	44029	45007	46004	47004	...
23	11002	31003	36010	40002	41004	44003	45003	46002	49002	50004	53006		...
24	11009	15003	27002	41005	44012								...
25	1003	11015	27002	41003	44011								...
26	6001	9009	11013	12010	24011	36005	41007	42005	43001	44003	45004	46002	...
27	14011	17005	19011	33006	50005	54003							...
28	9003	12005	18004	19004	21002	25001	32003	41001	47001	49001	50002	51001	...
29	9011	11015	14005	36005	37004	41004	42004	44009	46002	47002	48003	50004	...
30	11011	12009	30003	31004	36003	41006	43006	44008	49002	52008	53006		...
31	19020	39006											...
32	2001	13003	19008	37001	41001	49001	53001	54001					...

5.2 Avaliação do tempo de classificação

Cada vez que um novo caso é recebido pelo CAPBC, este o compara aos casos anteriormente armazenados na BC, calculando a similaridade com cada um destes. Por exemplo, o caso 915 é comparado aos 914 anteriormente já classificados e armazenados.

A Figura 5.1 mostra o gráfico do tempo de classificação de cada um dos 1000 casos da BC, utilizando-se três critérios diferentes de similaridade: número de objetos iguais (N_OBJ), porcentagem de objetos iguais (%_OBJ) e porcentagem de objetos iguais com pesos (%_OBJ+P). Estes tempos foram obtidos com o simulador sendo executado em um micro-computador com processador Intel(R) Core(TM)2 Duo de 2GHz e memória RAM de 2GB.

É possível observar, no gráfico, que os tempos de classificação utilizando-se o número de objetos iguais ou porcentagem de objetos iguais como critério de similaridade é o mesmo; já, utilizando a porcentagem de objetos iguais com a aplicação de pesos aos objetos, o tempo de classificação é consideravelmente maior. Estes resultados já eram esperados, pois nos dois primeiros critérios, apenas a fórmula utilizada no cálculo da similaridade foi alterada; já no terceiro, o CAPBC deve consultar a tabela contendo o peso de cada objeto para calcular a similaridade dos casos, o que torna o processo mais demorado.

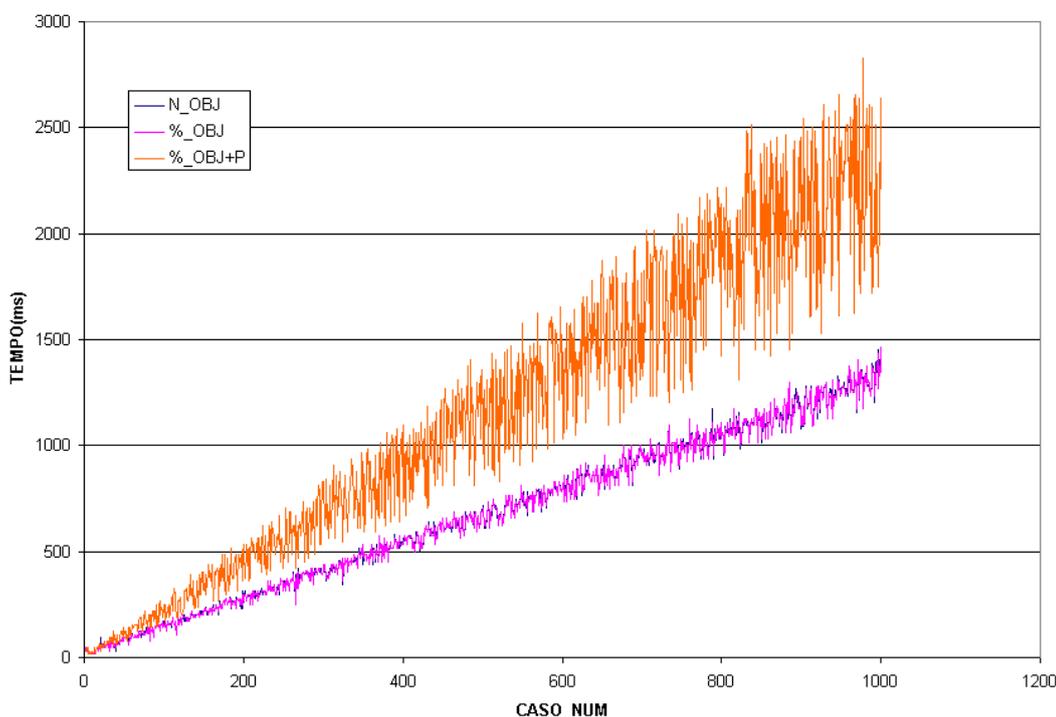


Figura 5.1: Gráfico do tempo de classificação de cada caso de acordo com o critério de similaridade utilizado.

Também se observa uma tendência linear no aumento do tempo de classificação, conforme o aumento do número do caso, em ambos os critérios de similaridade considerados. Este resultado também era esperado, já que o crescimento no número de casos que devem ser comparados ao novo caso também é linear.

Nota-se, ainda, em ambos os critérios de similaridade, que há grandes variações no tempo de classificação entre casos subsequentes no gráfico. Isto ocorre devido à variação da quantidade de objetos presente em cada caso. Quando um caso possui poucos objetos, são poucos os itens a serem comparados com itens dos casos anteriores, portanto o tempo de classificação é menor. O oposto ocorre nos casos com muitos objetos.

Para os objetivos práticos do sistema, um tempo médio, aproximado, de 2,5 segundos para a classificação de um caso (obtido na classificação do caso número 1000, considerando o uso de pesos nos objetos) pode ser considerado razoável. Porém, com o crescimento da base de casos, este tempo deve crescer proporcionalmente, o que tornará o sistema lento quando a base de casos estiver muito extensa.

Neste trabalho, foi considerado aceitável esse crescimento no tempo, porém, para trabalhos futuros, foi sugerido um estudo sobre a classificação de novos casos utilizando apenas um subconjunto da base de casos completa, avaliando a redução no tempo de classificação e a alteração, ou não, na qualidade das classificações.

5.3 Convergência do número de classificações para o número correto de habitantes

A partir da recombinação de peças de vestuário pelos habitantes, é esperado que, após certo tempo de funcionamento, o sistema seja capaz de classificar sob um mesmo código de habitante os casos que possuem, entre si, similaridade igual ou maior do que um valor mínimo pré-estipulado. Também é esperado que, quanto maior o valor da similaridade mínima estipulado, menos casos similares entre si existam.

Nas seções 5.3.1 e 5.3.2 são comparados resultados de classificações, quanto ao número de habitantes obtidos pelo simulador, modificando-se os valores de similaridade.

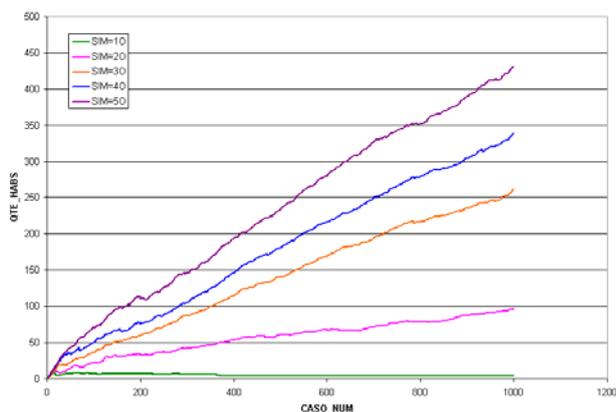
5.3.1 Classificação variando-se os valores de similaridade mínima

Os gráficos da Figura 5.2 mostram uma comparação na classificação utilizando os seguintes critérios e valores de similaridade:

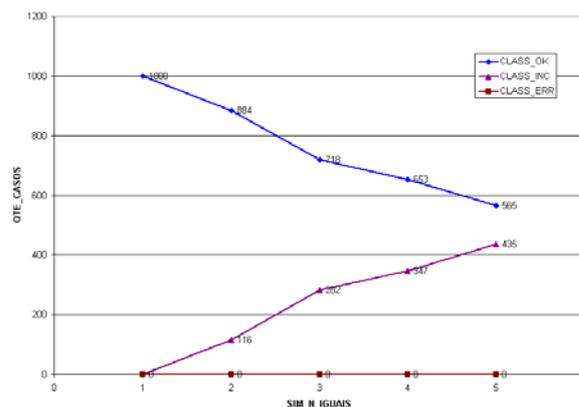
O gráfico 5.2(a) mostra a variação da quantidade de habitantes obtida pelo classificador na base de casos de teste, utilizando-se como critério de similaridade a quantidade de objetos iguais, variando esse valor de 1 até 5. O gráfico 5.2(b) mostra a quantidade de classificações corretas, incompletas e erradas para estes mesmos valores.

O gráfico 5.2(c) mostra a variação da quantidade de habitantes obtida, utilizando como critério de similaridade a porcentagem de objetos iguais, variando esse valor de 5% até 50%, com granularidade de 5%. O gráfico 5.2(d) mostra a quantidade de classificações corretas, incompletas e erradas para estes mesmos valores.

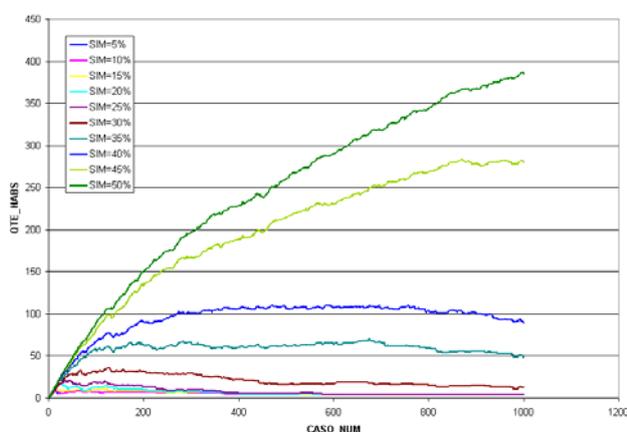
Para estes testes, foi utilizada a base de casos original, exposta na seção 5.1.



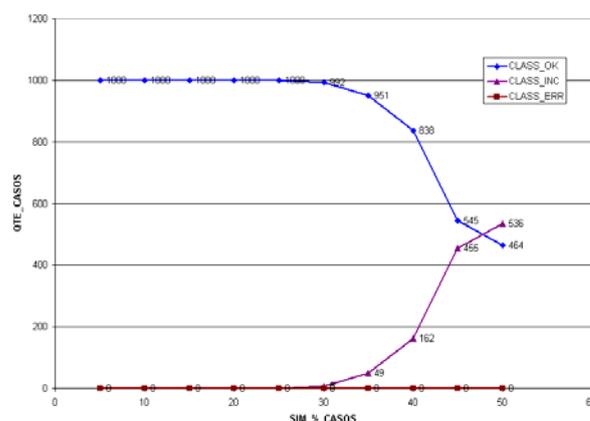
(a) qte. habitantes / número do caso: sim = qte. de obj. iguais



(b) qualidade das classificações / qte. de obj. iguais



(c) qte. habitantes / número do caso: sim = % de obj. iguais



(d) qualidade das classificações / % de obj. iguais

Figura 5.2: Gráficos da relação entre o número do caso classificado e a quantidade de habitantes gerada pelo CAPBC e do valor de similaridade mínima e a qualidade das classificações obtida, utilizando-se a BC original.

Legenda dos eixos: QTE_HABS: Quantidade de Habitantes; CASO_NUM: Número do Caso; QTE_CASOS: Quantidade de Casos; SIM_N_IGUAIS: Similaridade pela Quantidade de Objetos Iguais; SIM_%_IGUAIS: Similaridade pela Porcentagem de Objetos Iguais.

Analisando-se o gráfico 5.2(a), é possível observar que, quanto maior for o valor de similaridade mínima estipulado, maior será o número de habitantes que o classificador encontrará. O único valor de similaridade mínima que levou à convergência da quantidade de habitantes para o valor correto foi o de apenas um objeto igual.

No gráfico 5.2(c), foi obtida a convergência da quantidade de habitantes para o valor correto, antes do milésimo caso, para os valores de similaridade mínima de até 25%. Para os valores de 30%, 35% e 40%, aparentemente, há uma convergência deste valor, embora ela não tenha chegado à quantidade esperada, até a análise do milésimo caso. As curvas obtidas para os valores de 45% e 50% não aparentam convergir até o último caso em análise.

Os gráficos 5.2(b) e 5.2(d) permitem visualizar a relação entre o valor de similaridade mínima e a quantidade de casos classificados corretamente (CLASS_OK) ou de forma incompleta (CLASS_INC) após os 1000 casos.

Pôde-se observar, primeiramente, que não houve classificações erradas (CLASS_ERR) sob nenhum dos critérios utilizados. Isto ocorreu pois a base de casos utilizada nestes primeiros testes não possui objetos permutados, nem leituras incorretas inseridas. Também foi possível observar que, apesar do crescimento na quantidade de habitantes diferentes, quando o valor de similaridade mínima foi aumentado, a quantidade de classificações corretas nos gráficos 5.2(b) e 5.2(d) ainda permaneceu alta. Isto indica que a maioria dos casos foi corretamente agrupada e que as classificações incompletas pertencem a pequenos subgrupos, os quais não alcançaram o valor de similaridade mínima com os casos dos grupos maiores.

5.3.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos

Os gráficos da Figura 5.3 mostram uma comparação na classificação utilizando os seguintes critérios e valores de similaridade. A base de casos utilizada nestas simulações também é a base de casos original.

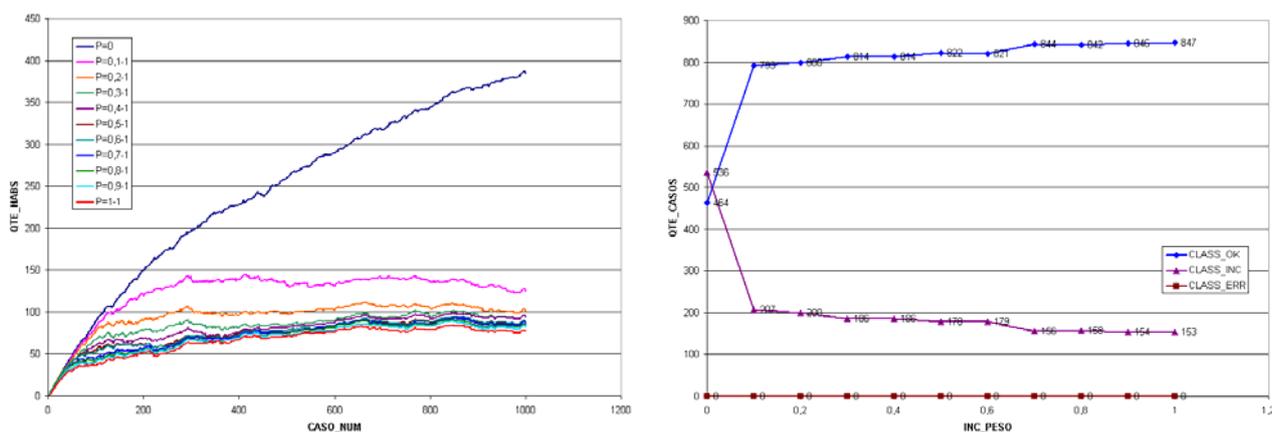
O gráfico 5.3(a) mantém fixo o valor de similaridade mínima em 50% e mostra a variação da quantidade de habitantes obtida, utilizando-se o incremento de peso de 0 a 1, com granularidade de 0,1, a cada 1 caso contendo o objeto. A atribuição dos pesos é feita de acordo com o exposto na seção 4.1.2.5. O gráfico 5.3(b) mostra a quantidade de classificações corretas, erradas e incompletas para esses mesmos valores. O gráfico 5.3(c) mostra a quantidade de classificações corretas, erradas e incompletas para esses mesmos valores, porém, considerando que a classificação de todos os casos somente é feita após a entrada da combinação de número '1000', onde os mesmos pesos finais calculados são utilizados na classificação de todos os casos.

No gráfico 5.3(a) pode-se perceber o resultado da aplicação dos pesos aos objetos mais relevantes. Há uma grande diferença no comportamento das curvas obtidas, principalmente entre o gráfico onde nenhum peso é atribuído e aquele em que é somado '0,1' ao peso em cada ocorrência do objeto. As demais curvas seguem a mesma tendência desta.

Nos gráficos 5.3(b) e 5.3(c) pode ser observada, também, a grande alteração na relação entre as classificações, sem e com a aplicação dos pesos nos objetos mais relevantes, onde a

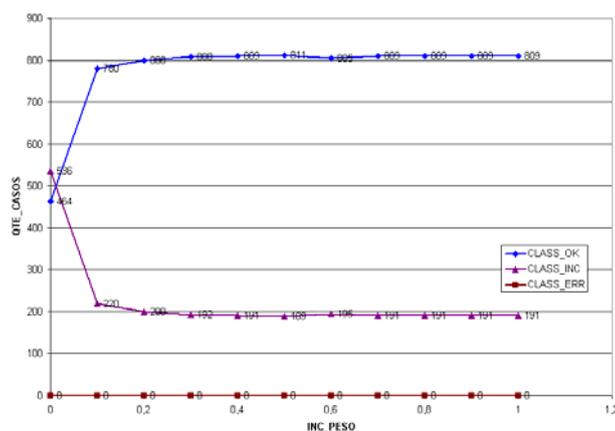
quantidade de casos classificados corretamente subiu de aproximadamente 50% para 80%, com a aplicação de pesos.

Entre os gráficos 5.3(b) e 5.3(c), foi observada a mesma tendência nas curvas, obtendo pequenas diferenças nas classificações, que ocorreram devido aos diferentes pesos aplicados aos objetos em cada um deles.



(a) qte. hab. / num. do caso: inc. de peso critério 1

(b) qualidade das classificações / inc. de peso critério 1



(c) qualidade das classificações / inc. de peso critério 2

Figura 5.3: Gráficos da relação entre o número do caso classificado e a quantidade de habitantes gerada pelo CAPBC e do incremento no peso dos objetos e a qualidade das classificações obtida, utilizando-se a BC original.

Legenda dos eixos: QTE_HABS: Quantidade de Habitantes; CASO_NUM: Número do Caso; QTE_CASOS: Quantidade de Casos; INC_PESO: Incremento do Peso.

Não foi traçado um gráfico de evolução da quantidade de habitantes de acordo com o número do caso, pois, nesta situação, a classificação de todos os casos ocorre somente após a entrada do caso '1000'.

Por ainda se tratar da mesma base de casos anterior (base de casos original), também não foram observadas classificações erradas com o uso de nenhum peso.

5.3.3 Conclusões sobre a quantidade de habitantes gerados pelo CAPBC

Os testes efetuados com a base de casos original, ou seja, sem objetos permutados, estranhos ou casos concatenados, permitiu observar o comportamento das classificações alterando-se os critérios e valores de similaridade mínima.

Se fosse possível garantir que os casos obtidos em uma residência real tivessem essas características, poder-se-ia escolher o critério de similaridade mínima apenas pela convergência mais rápida para a quantidade correta de habitantes obtida pelo classificador. Neste caso, critérios como 1 objeto igual, 5% ou 10% de objetos iguais obteriam os melhores resultados.

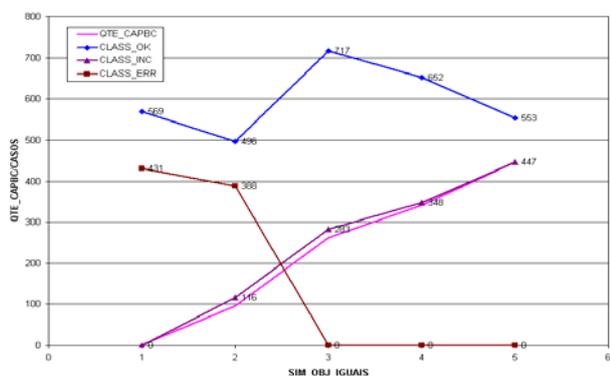
Os próximos testes visam verificar a capacidade do CAPBC em classificar corretamente casos com objetos permutados, objetos estranhos ou casos concatenados, por meio dos mesmos critérios avaliados até este ponto.

5.4 Classificação de base de casos com objetos permutados entre habitantes

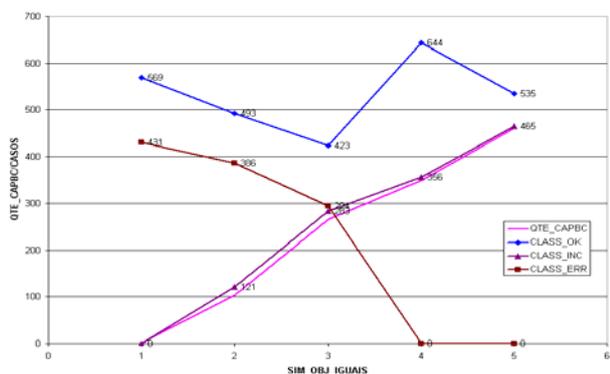
5.4.1 Classificação variando-se os valores de similaridade mínima, com objetos permutados

Para simular o empréstimo de objetos entre os habitantes no interior de uma residência, foi utilizada a base de casos original, permutando-se objetos não-íntimos entre os habitantes, de forma aleatória, com as seguintes porcentagens: 2%, 5%, 10% e 20%. Para a modificação da base de casos original foi usado o algoritmo de permuta de objetos, exposto na seção 4.1.1.1. Deve-se lembrar que este apenas permite a permuta de objetos entre pessoas do mesmo sexo e faixa etária.

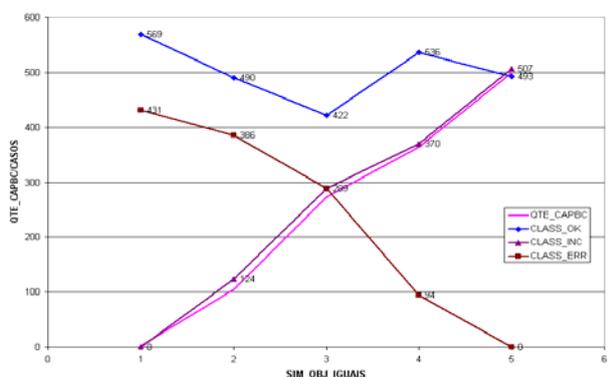
Os gráficos da Figura 5.4 mostram as quantidades de classificações corretas, incompletas e erradas obtidas pelo classificador, nas bases de casos com objetos permutados, conforme exposto no parágrafo anterior. Foi utilizado, como critério de similaridade, a quantidade de objetos iguais, variando esse valor de 1 até 5 - gráficos 5.4(a) a 5.4(d) - e a porcentagem de objetos iguais, variando esse valor de 5% a 60%, com granularidade de 5% - gráficos 5.4(e) a 5.4(h).



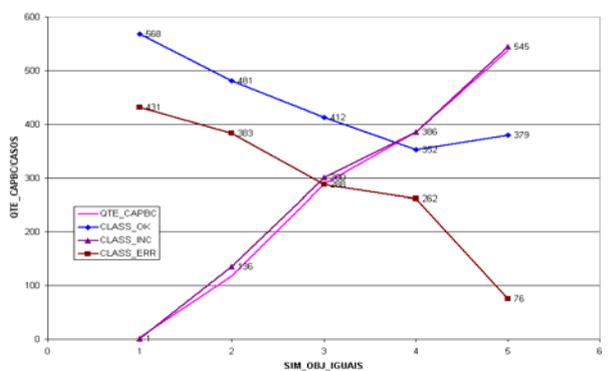
(a) 2% de objetos permutados: sim = qte. obj. iguais



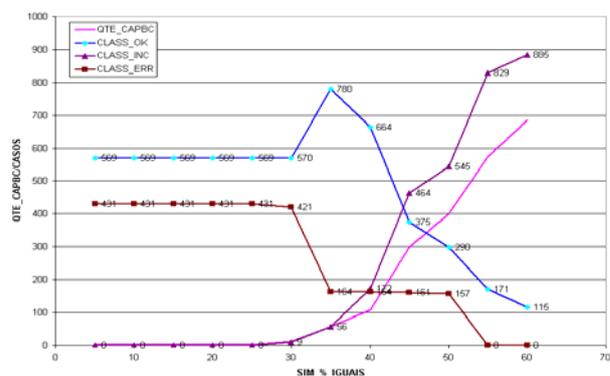
(b) 5% de objetos permutados: sim = qte. obj. iguais



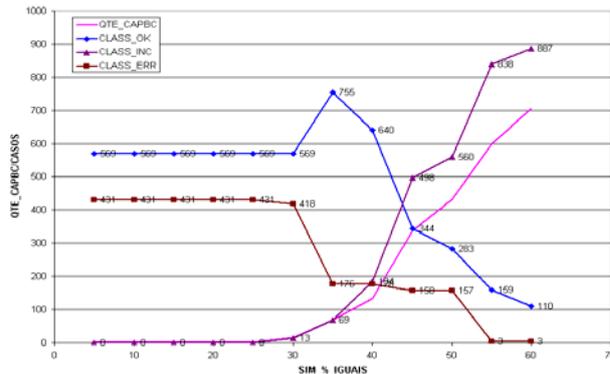
(c) 10% de objetos permutados: sim = qte. obj. iguais



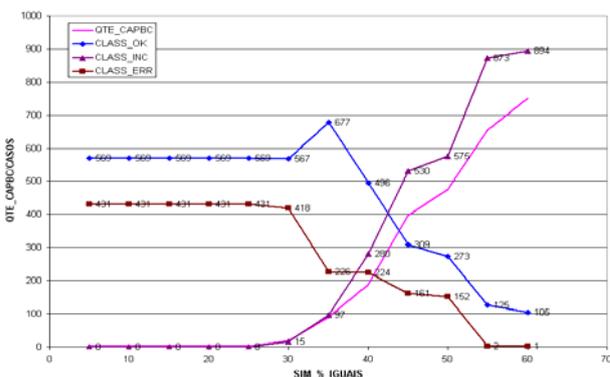
(d) 20% de objetos permutados: sim = qte. obj. iguais



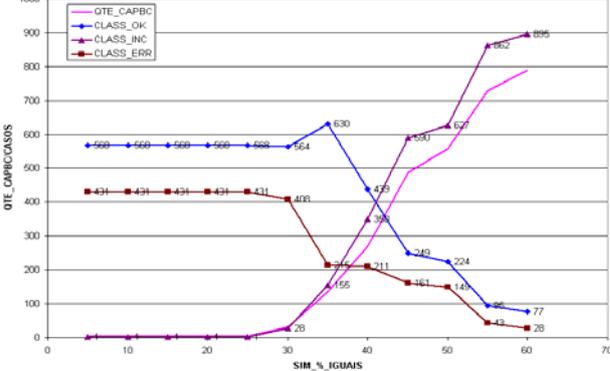
(e) 2% de objetos permutados: sim = % obj. iguais



(f) 5% de objetos permutados: sim = % obj. iguais



(g) 10% de objetos permutados: sim = % obj. iguais



(g) 20% de objetos permutados: sim = % obj. iguais

Figura 5.4: Gráficos da relação entre o valor de similaridade mínima utilizado e a qualidade nas classificações, utilizando-se bases de casos com objetos permutados entre habitantes.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; SIM_OBJ_IGUAIS: Similaridade pela Quantidade de Objetos Iguais; SIM_%_IGUAIS: Similaridade pela Porcentagem de Objetos Iguais.

Avaliando-se os gráficos das classificações obtidas pelo CAPBC, utilizando-se bases de casos contendo objetos permutados entre os habitantes, pôde-se observar o surgimento de classificações erradas, devido à união de subgrupos de casos de pessoas diferentes sob o código de um mesmo habitante.

É esperado do classificador que, utilizando-se valores de similaridade mínima maiores, os casos com objetos permutados possam ser classificados de forma correta, pois, embora tenham certa similaridade com combinações de objetos de habitantes diferentes, essa não atinge o valor mínimo para que os casos recebam a mesma classificação.

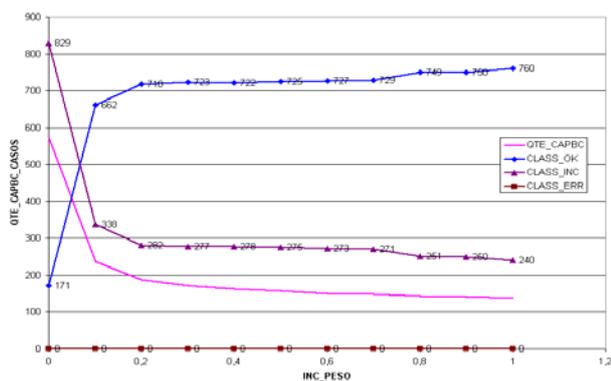
Observando-se os gráficos, nota-se que, principalmente nas classificações por quantidade de objetos iguais, nas bases de casos onde a porcentagem de objetos permutados é pequena, o algoritmo consegue excluir as classificações erradas utilizando valores baixos de similaridade mínima; já quando os objetos permutados eram muitos, o valor de similaridade mínima aplicado deve ser maior, para que a exclusão dos casos errados possa ocorrer. A partir da base de casos com 5% de objetos permutados, já se pode verificar que o classificador não foi capaz de excluir todas as classificações erradas, para os valores de similaridade mínima testados.

Também é verificado que, quanto maior o valor de similaridade mínima utilizado, maior é a quantidade de classificações incompletas obtidas, logo, o ideal seria encontrar um critério e valor de similaridade mínima capaz de maximizar as exclusões de classificações erradas, com o menor sacrifício possível às classificações corretas.

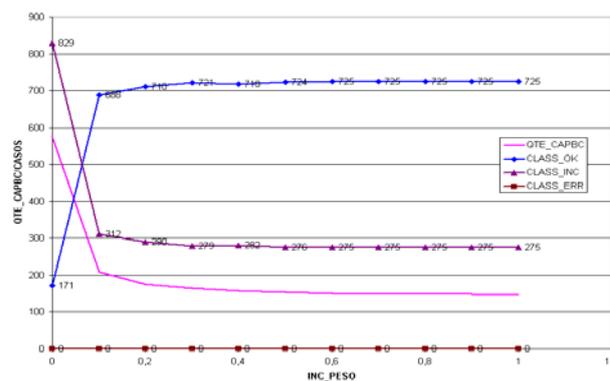
Com este objetivo foram efetuados testes aplicando pesos aos objetos mais relevantes, a fim de observar a influência destes na qualidade das classificações, quando a base de casos contém itens permutados entre os habitantes. Estes são detalhados na próxima seção.

5.4.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com objetos permutados

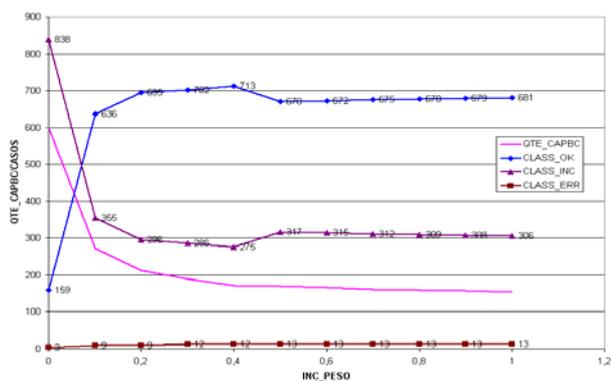
Na Figura 5.5, foram gerados gráficos das classificações das mesmas bases de casos testadas no item 5.4.1, porém, agora, incrementando-se os pesos dos objetos.



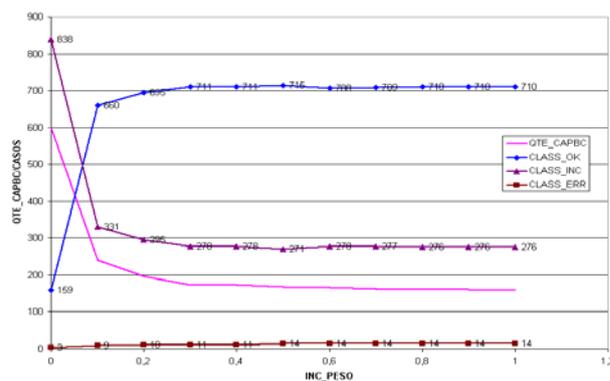
(a) 2% de objetos permutados: inc. de peso critério 1



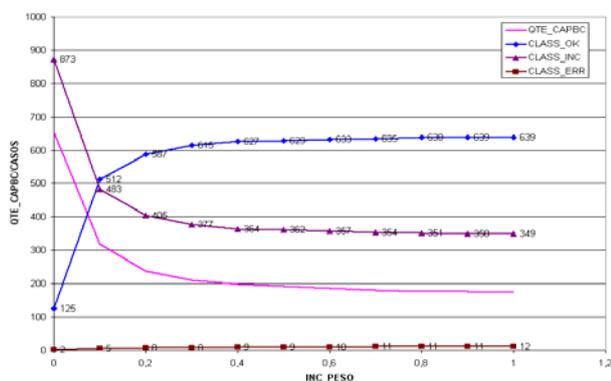
(e) 2% de objetos permutados: inc. de peso critério 2



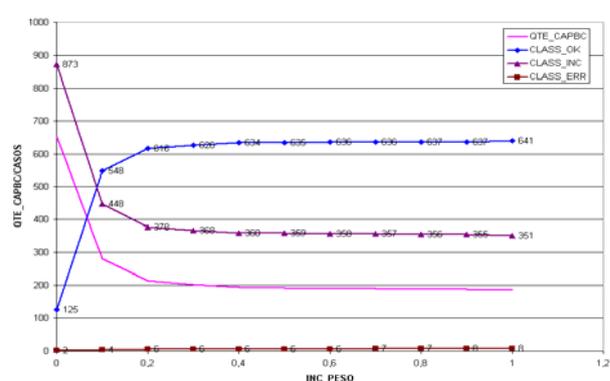
(b) 5% de objetos permutados: inc. de peso critério 1



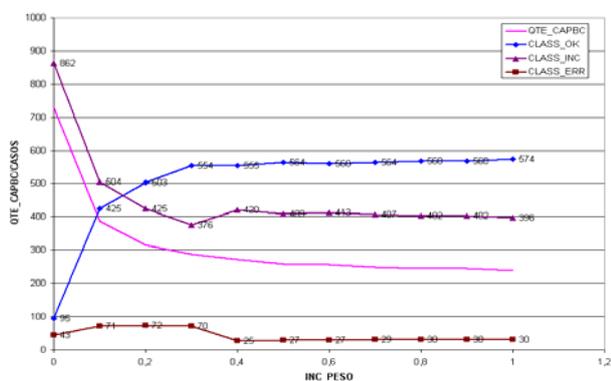
(f) 5% de objetos permutados: inc. de peso critério 2



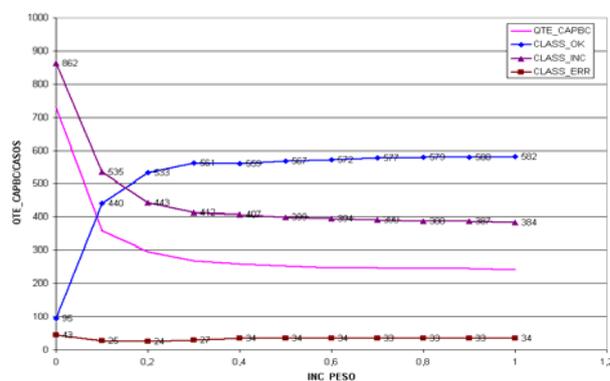
(c) 10% de objetos permutados: inc. de peso critério 1



(g) 10% de objetos permutados: inc. de peso critério 2



(d) 20% de objetos permutados: inc. de peso critério 1



(h) 20% de objetos permutados: inc. de peso critério 2

Figura 5.5: Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases de casos com objetos permutados entre habitantes.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; INC_PESO: Incremento no Peso dos Objetos.

O valor de similaridade mínima foi fixado em 55%, por ser o primeiro valor em que as classificações erradas são excluídas dos resultados, na porcentagem de itens permutados de 2%. Foram testados incrementos nos pesos dos objetos de 0,1 a 1, com granularidade de 0,1, a cada caso contendo o objeto em questão – gráficos 5.5(a) a (d). Foram traçados os gráficos, também, considerando-se a classificação da BC somente após a entrada dos 1000 casos no sistema, onde este aplicou o peso final calculado de cada objeto na classificação, no lugar de incrementar o peso e efetuar a classificação caso a caso – gráficos 5.5(e) a (h).

A partir da observação dos gráficos da Figura 5.5, é possível notar uma relação bastante clara entre o incremento do peso e a redução da quantidade de classificações incompletas e do aumento das classificações corretas. Por outro lado, não é verificada uma relação clara entre o incremento de peso e a quantidade de classificações erradas.

Pode-se perceber que, para alguns valores de incremento de peso, a quantidade de classificações erradas aumenta e, para outros, ela diminui. Isto ocorre porque os casos classificados de forma incompleta, inicialmente, com a aplicação de pesos, passam a atingir o valor de similaridade mínima estipulado, em relação a outros casos. Porém, se este passou a ter similaridade com um caso classificado erroneamente, também terá sua classificação errada.

De qualquer forma, para qualquer dos critérios de aplicação de peso, é possível observar que, quanto maior a porcentagem de objetos permutados, menor é a capacidade do CAPBC em maximizar a quantidade de classificações corretas em relação às incompletas, conforme o aumento do peso dos objetos.

Comparando-se os gráficos da classificação dos casos com a aplicação linear dos pesos, durante o processo de entrada das combinações, e da classificação somente após a entrada do caso 1000, pode-se verificar que a tendência das curvas é a mesma, tendo apenas pequenas diferenças de valores, devido aos pesos diferentes aplicados em cada classificação. Porém, nenhum dos critérios foi capaz de excluir as classificações erradas.

5.4.3 Conclusões sobre a classificação de uma base de casos com objetos permutados entre habitantes

Nos critérios e valores simulados, quando havia combinações na base de casos com itens permutados entre habitantes, foi notada uma grande quantidade de classificações erradas quando valores baixos de similaridade mínima eram utilizados no classificador. Conforme a similaridade mínima foi aumentada, a quantidade de classificações erradas foi reduzida, sendo estas eliminadas em grande parte dos testes.

Por outro lado, o aumento no valor da similaridade mínima provocou um grande aumento no valor das classificações incompletas e a redução das classificações corretas.

Foi, então, fixado um valor de similaridade mínima e acrescentado peso aos objetos, na tentativa de resolver as classificações incompletas. Observou-se uma relação direta entre a aplicação dos pesos e o aumento das classificações corretas e redução das classificações incompletas. Isto ocorre já que os pesos nos objetos mais relevantes ocasionam um aumento na similaridade entre os casos onde estes estão presentes no conjunto intersecção das combinações, mesmo que o valor de similaridade mínima utilizado seja alto.

Não foi, entretanto, verificada uma relação clara entre o peso dos objetos e a quantidade de classificações erradas. Na verdade, os pesos podem tanto auxiliar uma classificação incompleta a atingir a similaridade com o subconjunto de casos do habitante correto, quanto com o subconjunto do habitante errado, dependendo dos objetos que foram permutados.

5.5 Classificação de base com casos concatenados

A fim de simular uma falha no sistema de identificação RFID, foi considerada a possibilidade de dois habitantes passarem muito próximos, um do outro, pelo leitor de RFID, e este efetuar a leitura dos itens de ambos os habitantes simultaneamente, como se esta fosse uma única combinação de objetos.

A concatenação de pares de casos provoca uma redução na quantidade total de casos na base. Desta forma, se 2% dos casos de uma base de 1000 combinações forem concatenados a outro, a quantidade resultante desta base será 980 casos. Desta forma, para manter a quantidade de 1000 combinações inicialmente estipulada, os próximos 20 casos, que não estavam sendo avaliados anteriormente, passarão a integrar a base de casos avaliada. Assim foi feito para todas as porcentagens de concatenação de casos testadas

5.5.1 Classificação variando-se os valores de similaridade mínima, com casos concatenados

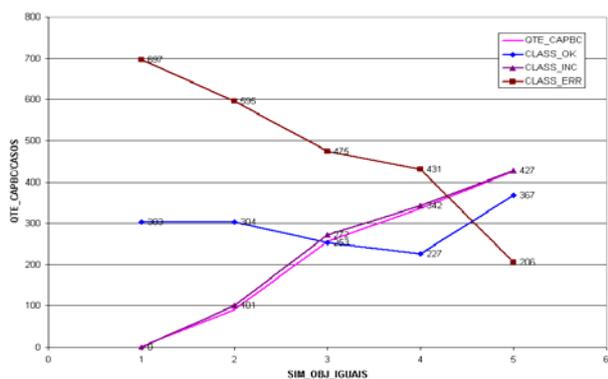
Para este teste, foi utilizada a base de casos original, concatenando-se casos sequenciais de habitantes diferentes, com as seguintes porcentagens: 2%, 5%, 10% e 20%. Para a concatenação dos casos, foi utilizado o algoritmo apresentado na seção 4.1.1.2. Ao contrário do algoritmo de permuta de objetos, que só permite a troca de itens quando os habitantes forem do mesmo sexo e faixa etária, neste podem ser concatenados casos de qualquer par de habitantes, o que deve ocasionar uma quantidade maior de classificações erradas, devido à união de casos que eram impossíveis de ocorrer pela permuta de objetos.

Os gráficos da Figura 5.6 mostram as quantidades de classificações corretas, incompletas e erradas obtidas pelo classificador, nos bancos de dados com casos concatenados, conforme exposto no parágrafo anterior. Foi utilizado, como critério de similaridade, a quantidade de objetos iguais, variando esse valor de 1 até 5 - gráficos 5.6(a) a 5.6(d) - e a porcentagem de objetos iguais, variando esse valor de 5% a 60%, com granularidade de 5% - gráficos 5.6(e) a 5.6(h).

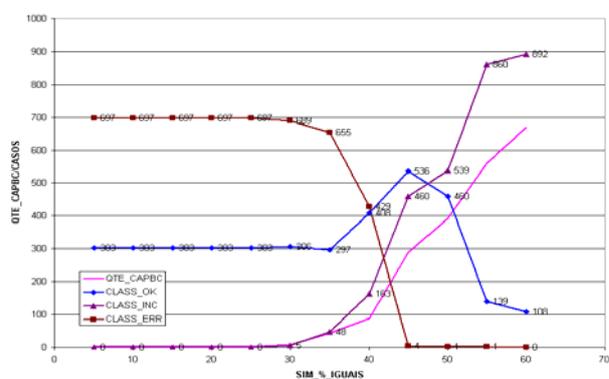
Avaliando-se os gráficos das classificações obtidas pelo CAPBC, utilizando-se bases contendo casos concatenados, foi verificada uma quantidade consideravelmente maior de classificações erradas, em comparação com os gráficos traçados para a base de casos com objetos permutados. Conforme foi citado anteriormente, as características desta base de casos permitem uniões incorretas de casos que não poderiam ocorrer pela permuta de objetos.

Utilizando-se a quantidade de objetos iguais como critério de similaridade – gráficos 5.6(a) ao 5.6(d) – é possível verificar que, com o aumento na quantidade de casos concatenados, a capacidade do classificador em reduzir a quantidade de classificações erradas, conforme o aumento da similaridade mínima, diminui. Também se observa que, apenas no gráfico 5.6(a) (2% de casos concatenados) a quantidade de classificações corretas foi maior do que a quantidade de classificações erradas, adotando-se similaridade mínima de 5 objetos iguais, porém menos de 40% dos casos foram classificados de forma correta.

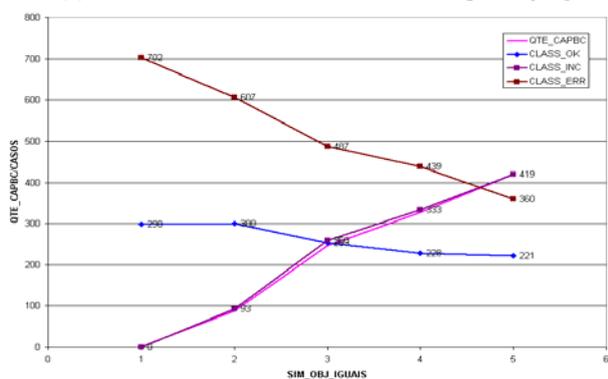
Nos gráficos contendo os resultados do uso da porcentagem de objetos iguais, é possível ver uma queda mais acentuada das classificações erradas – conforme o aumento no valor de similaridade - quando menos casos foram concatenados entre si. Utilizando-se este critério, apenas no gráfico 5.6(e) (2% de casos concatenados) a quantidade de classificações corretas superou 40% (quando aplicados valores de similaridade mínima entre 40 e 50%).



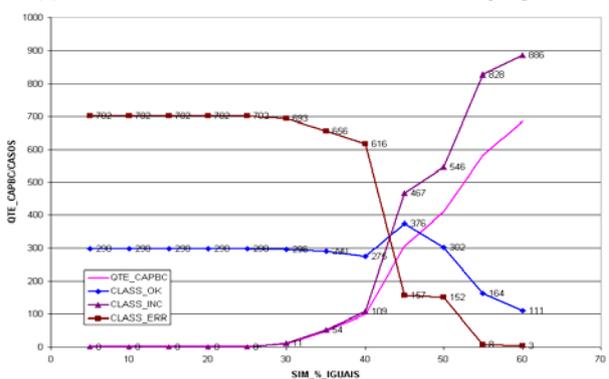
(a) 2% de casos concatenados: sim = qte. obj. iguais



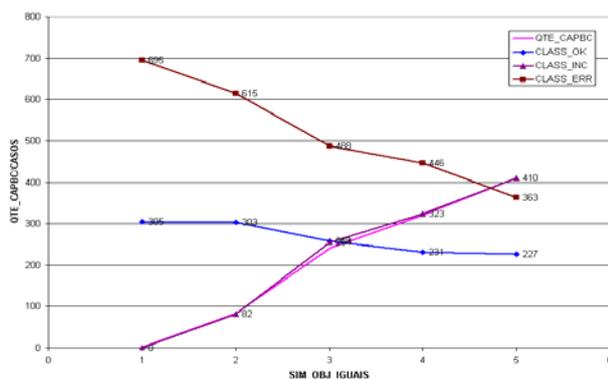
(e) 2% de casos concatenados: sim = % obj. iguais



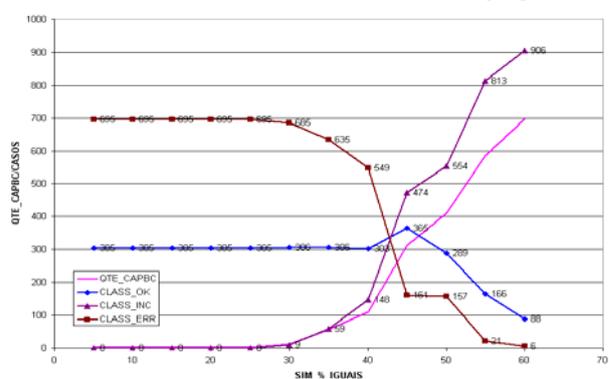
(b) 5% de casos concatenados: sim = qte. obj. iguais



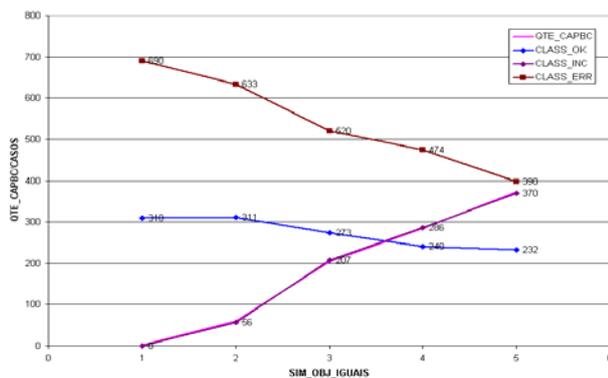
(f) 5% de casos concatenados: sim = % obj. iguais



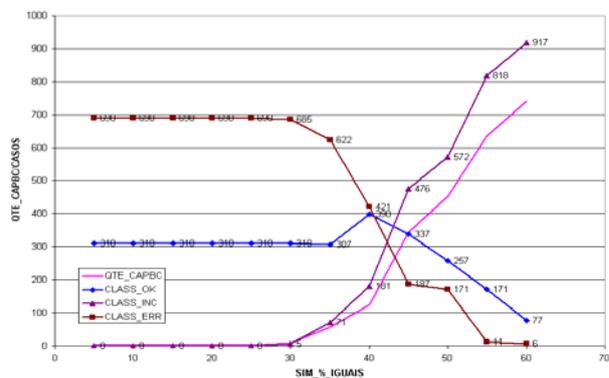
(c) 10% de casos concatenados: sim = qte. obj. iguais



(g) 10% de casos concatenados: sim = % obj. iguais



(d) 20% de casos concatenados: sim = qte. obj. iguais



(h) 20% de casos concatenados: sim = % obj. iguais

Figura 5.6: Gráficos da relação entre o valor de similaridade mínima utilizado e a qualidade nas classificações, utilizando-se bases com casos concatenados.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; SIM_OBJ_IGUAIS: Similaridade pela Quantidade de Objetos Iguais; SIM_%_IGUAIS: Similaridade pela Porcentagem de Objetos Iguais.

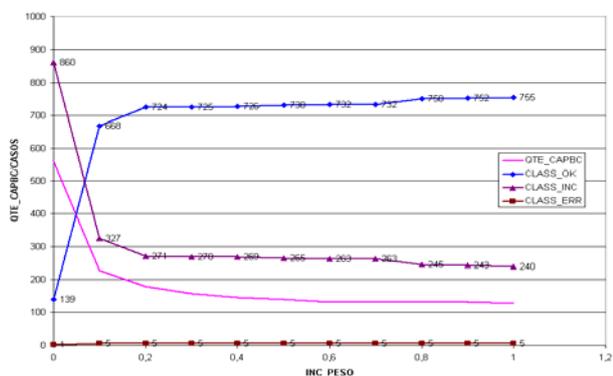
Comparando-se aos testes com permuta de itens, na concatenação de casos o CAPBC obteve resultados muito piores. Com a permuta de itens, os picos dos gráficos de classificações corretas ficaram entre 63% (no pior caso) e 78% de classificações corretas, ainda sem a aplicação de pesos aos objetos; com a concatenação de casos, o melhor resultado obtido foi de 53,6% de classificações corretas.

Na próxima seção é verificado o comportamento do classificador, considerando-se a aplicação de pesos aos objetos mais relevantes, conforme os critérios e valores descritos a seguir.

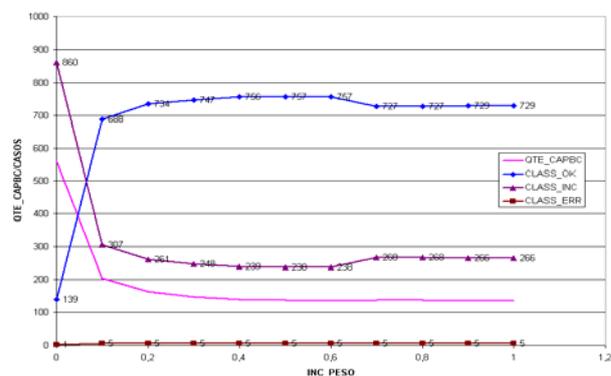
5.5.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com casos concatenados

Na Figura 5.7, foram gerados gráficos das classificações das mesmas bases de casos testadas na seção 5.5.1, porém, agora, alterando-se os pesos dos objetos. O valor de similaridade mínima foi fixado em 55%. Foram testados incrementos nos pesos dos objetos de 0,1 a 1, com granularidade de 0,1, a cada caso contendo o objeto em questão. Foram traçados os gráficos, também, considerando-se a classificação da BC somente após a entrada dos 1000 casos no sistema, onde este aplicou o peso final calculado de cada objeto na classificação, no lugar de incrementar o peso e efetuar a classificação caso a caso. Os critérios e valores utilizados foram exatamente os mesmos que na avaliação da base de casos com objetos permutados, a fim de possibilitar a comparação do comportamento do algoritmo nas duas situações.

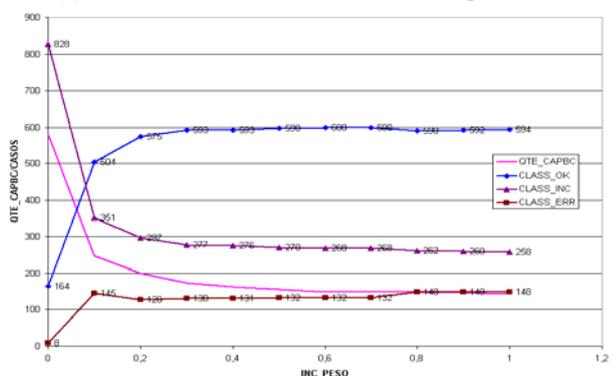
Da mesma forma que nas classificações dos casos com objetos permutados, a aplicação de pesos aos objetos demonstrou-se eficaz no aumento das classificações corretas e redução das classificações incompletas do CAPBC. Também nota-se que, conforme o aumento na porcentagem de casos concatenados, a capacidade do CAPBC em aumentar as classificações corretas é reduzida. Mesmo assim, no pior cenário simulado - 20% de casos concatenados - o classificador foi capaz de obter a quantidade de classificações corretas superior às classificações incompletas e erradas, para qualquer dos valores de peso aplicados.



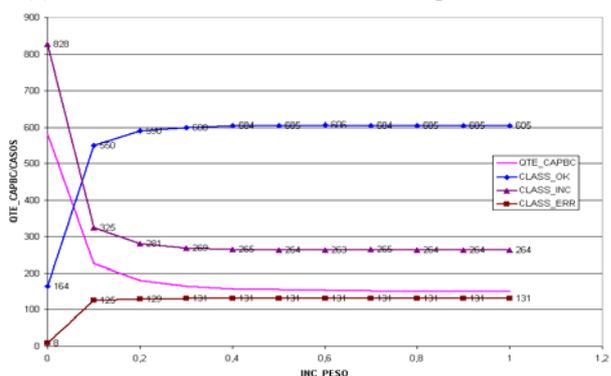
(a) 2% de casos concatenados: inc. de peso critério 1



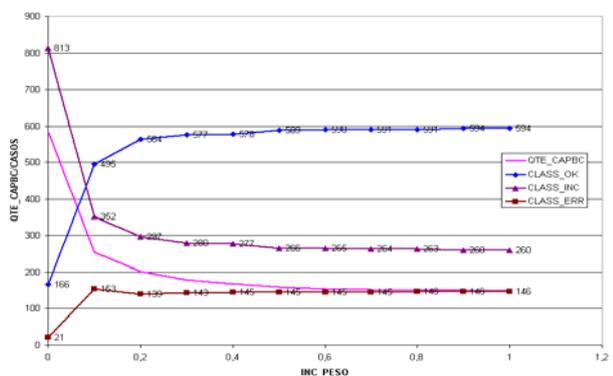
(e) 2% de casos concatenados: inc. de peso critério 2



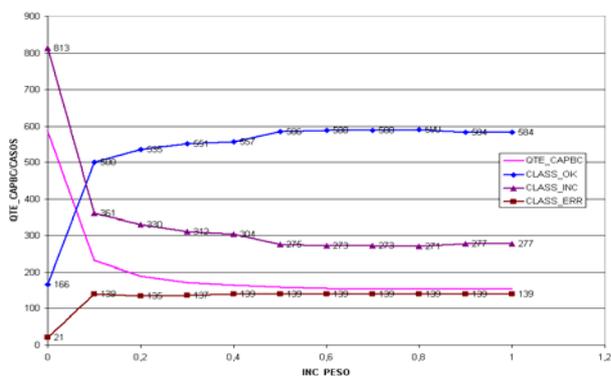
(b) 5% de casos concatenados: inc. de peso critério 1



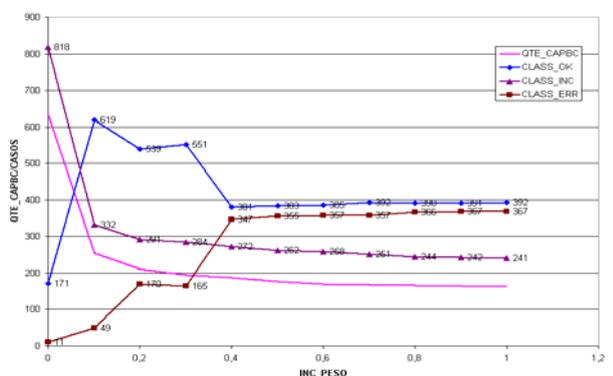
(f) 5% de casos concatenados: inc. de peso critério 2



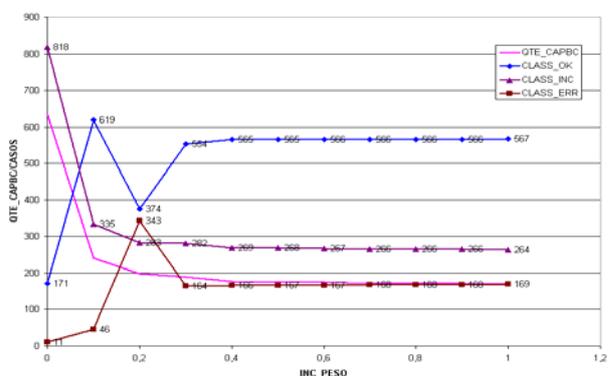
(c) 10% de casos concatenados: inc. de peso critério 1



(g) 10% de casos concatenados: inc. de peso critério 2



(d) 20% de casos concatenados: inc. de peso critério 1



(h) 20% de casos concatenados: inc. de peso critério 2

Figura 5.7: Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases com casos concatenados.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; INC_PESO: Incremento no Peso dos Objetos.

Os gráficos que representam as classificações da base com 2%, 5% e 10% de casos concatenados mostraram um crescimento linear, conforme o aumento do peso aplicado aos objetos. Os gráficos que representam as classificações com 20% de casos concatenados apresentaram comportamentos singulares para os incrementos de peso de 0,1 a 0,3, observados nos picos e vales dos gráficos. Isto ocorre devido a situações particulares - que devem ter sido geradas utilizando-se estes valores específicos - poderem provocar a reclassificação de uma parte considerável da base de casos. Ou seja, ao aplicarem-se valores específicos de incremento de peso aos objetos, em um momento particular, um único caso concatenado a outro pode ocasionar a união de subconjuntos de classificações muito grandes, o que explica a grande variação nos valores.

Os resultados das classificações efetuadas com o incremento de peso de forma sequencial – Figuras 5.7(a) a 5.7(d) – ou das classificações efetuadas usando-se os pesos finais calculados após 1000 casos – Figuras 5.7(e) a 5.7(h) – tiveram comportamentos semelhantes, com pequenas variações nos valores. Apenas na classificação da base com 20% de casos concatenados foram obtidos comportamentos bastante distintos, devido às particularidades deste conjunto de dados.

5.5.3 Conclusões sobre a classificação de uma base com casos de habitantes diferentes concatenados

Analisando-se os resultados obtidos por meio das simulações efetuadas, pode-se concluir que, quando casos são concatenados em uma base, a quantidade de classificações erradas cresce muito, em comparação à base de casos com objetos emprestados.

O CAPBC mostrou um desempenho ruim nas classificações efetuadas sem a aplicação de pesos aos objetos tendo obtido o valor máximo de 53,6% de classificações corretas no melhor cenário de simulação.

O uso de pesos na classificação destas bases de casos provocou uma melhora significativa na qualidade das classificações, embora não tenha sido capaz de excluir as classificações erradas. Ao se aplicar os pesos, um valor considerável de classificações incompletas foi convertido em classificações erradas, por aumentar a similaridade entre casos de habitantes diferentes.

Nestas simulações, puderam ser observadas grandes variações na qualidade das classificações, alterando-se os pesos de objetos, quando muitos casos haviam sido concatenados. Isto mostra que valores específicos de similaridade e pesos aplicados às classificações, em bases com muitos casos modificados, podem provocar grandes variações nos resultados das classificações. Desta forma, torna-se mais complexo o processo de escolha dos melhores critérios e valores de classificação, quando não é possível conhecer as características da base de casos.

5.6 Classificação de base de casos com o acréscimo de objetos novos ou estranhos

Neste trabalho, foi considerado que o sistema não seria capaz de reconhecer o tipo de objeto a partir da sua identidade eletrônica. Desta forma, as leituras de identidades RFID de qualquer tipo de objeto são interpretadas da mesma forma, pelo sistema CAPBC. Também é possível que, no decorrer do uso do sistema, um habitante adquira itens novos de vestuário contendo identidades RFID. Assim, estes também se integram aos casos destes habitantes.

Nas próximas seções, é avaliado o comportamento do CAPBC, considerando a inserção de itens novos ou estranhos nos casos, ou seja, que não pertencem, inicialmente, a qualquer dos habitantes.

5.6.1 Classificação variando-se os valores de similaridade mínima, com objetos novos ou estranhos

Para os testes, foi utilizada a base de casos original, alterando-se os casos da seguinte forma: foram inseridos até 5 objetos estranhos – quantidade escolhida de forma aleatória – em 2%, 5%, 10% e 20% dos casos existentes na base. Os objetos estranhos podem já ter existido em casos anteriores, ou não. Cada vez que um novo objeto estranho é inserido em um caso, este é adicionado a um banco de dados próprio, para poder ser inserido em outros casos, posteriormente.

Para a modificação da base de casos original, foi utilizado o algoritmo apresentado na seção 4.1.1.3. Este algoritmo, assim como o de concatenação de casos, permite que qualquer objeto estranho esteja presente em casos de qualquer habitante. Desta forma, valores de similaridade mínima baixos devem permitir que combinações de itens de quaisquer pessoas sejam classificadas sob o mesmo código, pelo CAPBC.

Os gráficos da Figura 5.8 mostram as quantidades de classificações corretas, incompletas e erradas, obtidas pelo classificador, nos bancos de dados, com a adição de objetos estranhos, conforme exposto no parágrafo anterior. Foi utilizado, como critério de similaridade a quantidade de objetos iguais, variando esse valor de 1 até 5 - gráficos 5.8(a) a 5.8(d) - e a porcentagem de objetos iguais, variando esse valor de 5% a 60%, com granularidade de 5% - gráficos 5.8(e) a 5.8(h).

Na classificação das bases de casos geradas, contendo objetos estranhos nas combinações de itens dos habitantes, foi possível observar um desempenho melhor do CAPBC, em relação às simulações analisadas anteriormente, com objetos permutados entre habitantes, ou casos de habitantes diferentes concatenados.

Devido aos itens acrescentados aos casos da BC serem escolhidos aleatoriamente, na similaridade por objetos iguais, nenhum caso teve mais do que 4 itens comuns presentes em casos de habitantes diferentes. Portanto, nas classificações por quantidade de objetos iguais, a partir do critério de similaridade mínima de 4 objetos iguais, não foram obtidas classificações erradas. Porém, da mesma forma que nas situações simuladas anteriormente, foi observado um crescimento na quantidade de classificações incompletas, conforme o aumento no valor de similaridade mínima.

Utilizando-se como critério de classificação a porcentagem de objetos iguais, também foi obtido um resultado melhor nas classificações, em relação às situações simuladas anteriormente. Nota-se que o pico do gráfico de classificações corretas, na pior situação simulada (20% de casos contendo objetos estranhos) possui 75,5% de classificações corretas, contra 70% (na melhor situação da permuta de objetos: 2% de objetos permutados) e 53,6% (na melhor situação da concatenação de casos: 2% de casos concatenados).

Desta forma, é possível concluir-se que a adição de objetos estranhos aos casos, mesmo estes sendo passíveis de estarem presentes em casos de pessoas diferentes, causa menor prejuízo às classificações, do que o empréstimo de objetos entre os habitantes, ou a concatenação de casos destes.

É avaliada, agora, a alteração nos resultados das classificações, com a aplicação de pesos nos objetos, utilizando-se as mesmas bases de casos das avaliações efetuadas sem o uso destes.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; SIM_OBJ_IGUAIS: Similaridade pela Quantidade de Objetos Iguais; SIM_%_IGUAIS: Similaridade pela Porcentagem de Objetos Iguais.

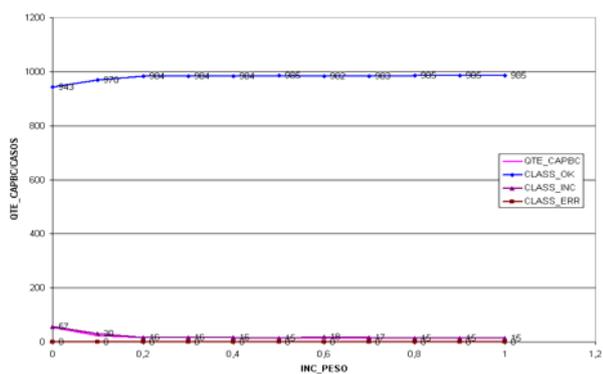
5.6.2 Classificação fixando-se os valores de similaridade mínima e variando-se os pesos de objetos, com objetos novos ou estranhos

Na Figura 5.9, foram traçados os gráficos utilizando-se as bases de casos das mesmas classificações anteriores, porém fixando-se o valor de similaridade mínima e alterando-se os pesos dos objetos. O valor de similaridade mínima foi fixado em 35%, por ter sido o menor valor onde as classificações erradas foram eliminadas, para as três primeiras condições simuladas, sem o uso dos pesos. Assim como nas situações anteriores, foram simulados incrementos nos pesos dos objetos de 0,1 a 1, a cada caso contendo o objeto em questão, bem como o uso dos mesmos valores, considerando-se a classificação apenas após a entrada do último caso, no sistema, já com os pesos finais calculados.

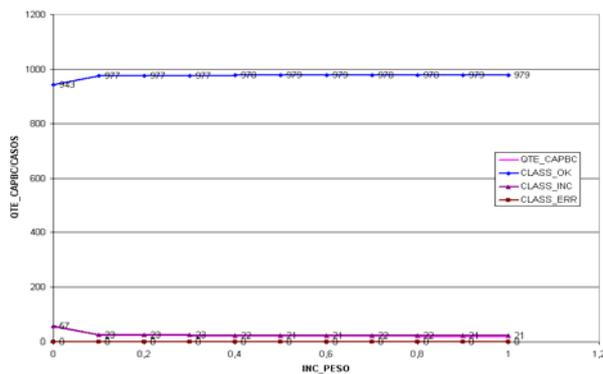
Nos gráficos obtidos a partir das bases de casos com 2% e 5% de combinações contendo objetos estranhos, o comportamento observado foi o esperado, alterando-se classificações incompletas para corretas.

Já nas bases de casos com 10% e 20% de combinações com objetos estranhos, foram obtidos comportamentos com singularidades (observados claramente nos picos e vales presentes nos gráficos) para determinados valores de pesos aplicados aos objetos. Estes valores específicos provocaram a alteração de classificações corretas para erradas, o que é extremamente indesejado. Isto ocorreu devido a objetos específicos terem recebido pesos que aumentaram o valor de similaridade entre casos de habitantes diferentes.

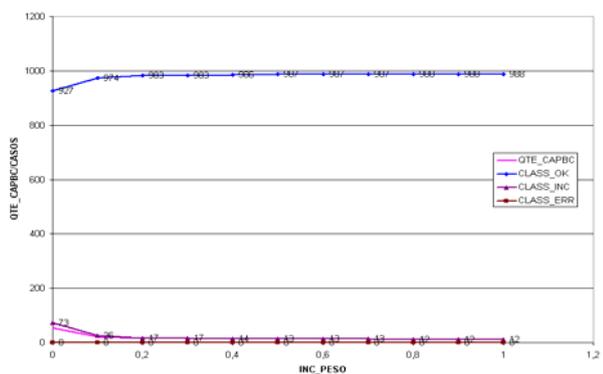
Observa-se, também, que o gráfico da Figura 5.9(c) (10% de casos com objetos estranhos) obteve classificações piores a partir do incremento de peso de 0,7 a cada caso contendo o objeto em questão, do que o gráfico da Figura 5.9(d) (20% de objetos estranhos) para os mesmos valores de incremento de peso. Isto mostra que modificações específicas, presentes em determinada base de casos, podem causar prejuízos piores aos resultados, do que uma média maior de modificações menos relevantes a estas. Ou seja, um único caso modificado de uma forma específica pode causar uma piora maior nos resultados, do que diversos casos modificados de formas diferentes.



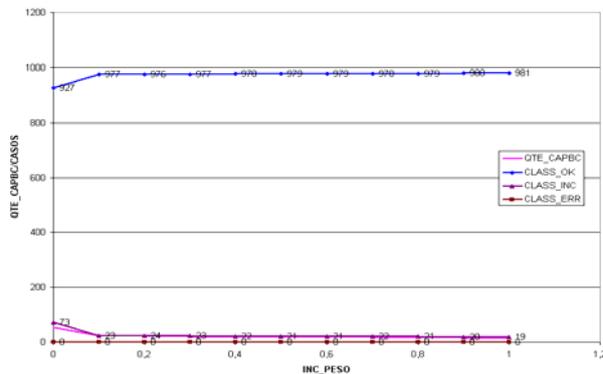
(a) 2% de casos com novos obj.: inc. de peso critério 1



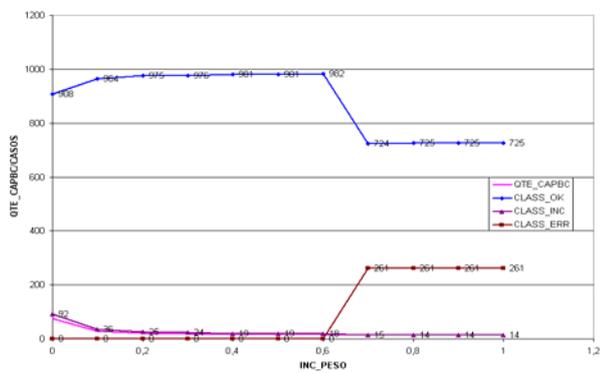
(e) 2% de casos com novos obj.: inc. de peso critério 2



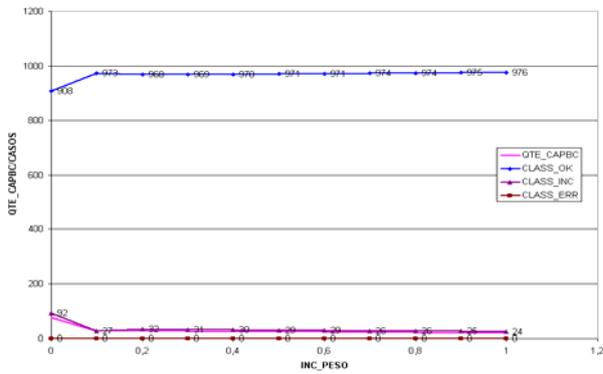
(b) 5% de casos com novos obj.: inc. de peso critério 1



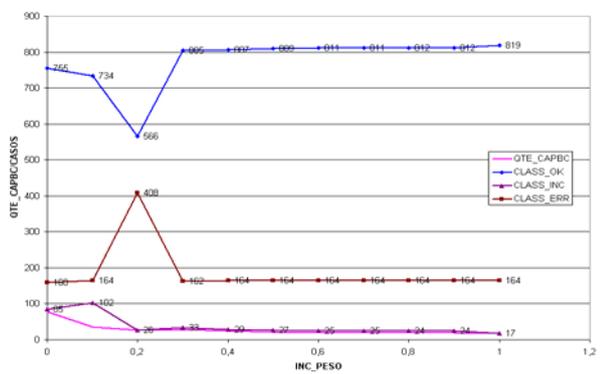
(f) 5% de casos com novos obj.: inc. de peso critério 2



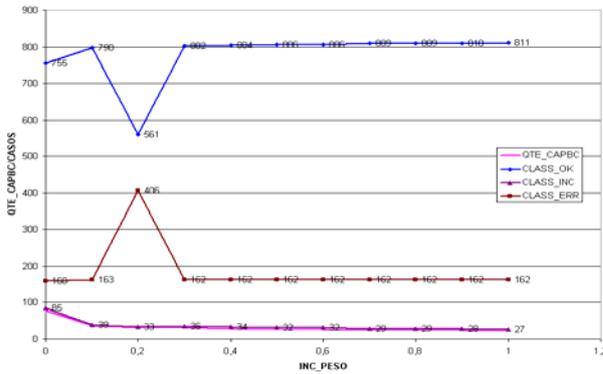
(c) 10% de casos com novos obj.: inc. de peso critério 1



(g) 10% de casos com novos obj.: inc. de peso critério 2



(d) 20% de casos com novos obj.: inc. de peso critério 1



(h) 20% de casos com novos obj.: inc. de peso critério 2

Figura 5.9: Gráficos da relação entre o incremento de peso aplicado aos objetos e a qualidade nas classificações, utilizando-se bases de casos com objetos novos ou estranhos.

Legenda dos eixos: QTE_CAPBC/CASOS: Quantidade de Habitantes Obtida pelo CAPBC e de Casos Classificados; INC_PESO: Incremento no Peso dos Objetos.

5.6.3 Conclusões sobre a classificação de uma base de casos com a existência de objetos novos ou estranhos nas combinações de itens

Para os critérios simulados de inserção de objetos estranhos e porcentagem de casos com estes presentes, os resultados mostram que, de maneira geral, o aumento no valor de similaridade mínima permite excluir as classificações erradas de casos e o aumento dos pesos nos objetos permite reduzir a quantidade de classificações incompletas.

Para quantidades maiores de casos contendo objetos estranhos, foram observados comportamentos específicos do classificador para determinados valores de pesos atribuídos aos objetos (observados claramente nos picos e vales dos gráficos 5.9(d) e 5.9(h)). Além disto, para alguns destes valores, o desempenho do classificador foi melhor nas bases de casos onde menos combinações possuíam objetos estranhos, do que em outras que possuíam mais. Isto mostra que casos modificados de formas específicas podem, em determinadas situações, causar alterações maiores às classificações do que uma média maior de casos modificados em determinada base.

De maneira geral, o desempenho do CAPBC foi melhor na classificação das bases com o acréscimo de objetos estranhos, do que naquelas contendo objetos permutados e casos concatenados.

5.7 Avaliação das qualidade das classificações

O comportamento ideal de um sistema classificador, voltado a aplicações em automação residencial, é a classificação correta do maior número possível de casos no menor tempo possível.

Nas simulações, foram observados os comportamentos das classificações da base de casos originalmente gerada e desta mesma modificada, com diferentes porcentagens de objetos permutados entre habitantes, casos concatenados e objetos estranhos acrescentados aos casos.

A partir dos dados gerados, foram traçados gráficos que permitiram observar o comportamento geral do classificador sobre cada conjunto de casos, de acordo com a mudança dos critérios e valores de similaridade. As tabelas 5.5 e 5.6 contêm os valores que

foram utilizados para gerar esses gráficos, porém reagrupados, de forma a possibilitar a análise das seguintes variáveis, que foram objeto de estudo deste trabalho, e são detalhadas a partir da seção 5.7.1: quantidade de habitantes reconhecidos pelo CAPBC, quantidade de classificações corretas, quantidade de classificações incompletas e quantidade de classificações erradas.

5.7.1 Quantidade de habitantes reconhecidos pelo CAPBC

Na avaliação das classificações, foram contados os códigos de habitantes criados pelo CAPBC, utilizando-se os diferentes critérios e valores simulados, ao fim da classificação dos 1000 casos gerados. As tabelas 5.5 e 5.6 permitem visualizar a variação da quantidade de códigos criados conforme a alteração das quantidades de objetos permutados, casos concatenados e presença de objetos estranhos, bem como se variando os critérios e valores de similaridade mínima aplicados, além dos pesos.

Na tabela 5.5, é possível se observar, primeiramente, a diferente quantidade de habitantes gerados, utilizando-se o critério mínimo de similaridade entre casos (1 objeto igual) para as diferentes bases de casos. Na BC original, foram reconhecidos 4 habitantes, que é a quantidade correta destes, e o resultado esperado para todos os conjuntos de dados. Nas demais bases de casos, o valor de habitantes reconhecido foi menor. Isto devido às trocas de objetos que permitiram o agrupamento de casos de habitantes diferentes sob os mesmos códigos, atribuídos pelo CAPBC.

Tabela 5.5 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais.

Sim. Mínima ->	1 Objeto	2 Objetos	3 Objetos	4 Objetos	5 Objetos
BC Original	4	95	261	339	431
2% Obj. Empr.	2	96	262	341	446
5% Obj. Empr.	2	103	265	350	461
10% Obj. Empr.	2	105	274	364	500
20% Obj. Empr.	3	118	289	385	537
2% Casos Conc.	1	91	256	336	426
5% Casos Conc.	1	88	246	327	420
10% Casos Conc.	1	83	240	320	411
20% Casos Conc.	1	60	204	285	372
2% Casos Obj. Estr.	1	90	257	339	431
5% Casos Obj. Estr.	1	87	260	339	431
10% Casos Obj. Estr.	1	82	253	338	431
20% Casos Obj. Estr.	1	74	255	339	430

Outro comportamento possível de observar na tabela é que, para os valores de 2 a 5 objetos iguais, a quantidade de códigos gerados cresce ou diminui, conforme a alteração dos parâmetros de modificação da base de casos. É possível verificar:

1 - Quanto maior a quantidade de objetos emprestados entre os habitantes, maior é a quantidade de habitantes reconhecida pelo CAPBC. Isto ocorre devido à redução da similaridade entre casos dos mesmos habitantes, quando seus objetos são substituídos por itens de outros;

2 - Quanto maior a quantidade de casos concatenados entre os habitantes, menor é a quantidade de habitantes reconhecida pelo CAPBC. O aumento dos casos concatenados permite que mais casos de habitantes diferentes recebam a mesma classificação pelo algoritmo, reduzindo a quantidade destas;

3 - Quanto maior a quantidade de casos com a presença de objetos estranhos, menor é a quantidade de habitantes reconhecida pelo CAPBC. Os objetos estranhos permitem a união de casos por meio da similaridade gerada por eles. Observa-se que, com o aumento no valor da similaridade mínima, este comportamento é reduzido, já que se tornará mais difícil a classificação de casos de habitantes diferentes por meio destes objetos estranhos.

A tabela 5.6 mostra a quantidade de habitantes criada pelo CAPBC, utilizando-se o critério de porcentagem de objetos iguais. Os campos com hachuras representam valores que não foram simulados.

Tabela 5.6 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais.

Sim. Mínima ->	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
BC Original	4	4	4	4	4	12	48	89	279	385		
2% Obj. Empr.	2	2	2	2	2	12	57	108	298	401	574	685
5% Obj. Empr.	2	2	2	2	2	16	68	133	338	433	599	706
10% Obj. Empr.	2	2	2	2	2	18	90	189	396	476	655	751
20% Obj. Empr.	3	3	3	3	3	31	136	269	487	558	729	789
2% Casos Conc.	1	1	1	1	1	7	44	88	288	393	559	670
5% Casos Conc.	1	1	1	1	1	9	50	100	307	410	581	686
10% Casos Conc.	1	1	1	1	1	11	56	111	312	411	585	697
20% Casos Conc.	1	1	1	1	1	8	57	124	343	452	635	740
2% Casos Obj. Estr.	1	1	1	1	1	11	54	100	292	398	564	671
5% Casos Obj. Estr.	1	1	1	1	1	15	54	101	298	404	570	682
10% Casos Obj. Estr.	1	1	1	1	1	17	75	130	337	431	607	714
20% Casos Obj. Estr.	1	1	1	1	5	25	78	149	378	472	644	753

Na tabela 5.6, é possível observar uma uniformidade nas quantidades de habitantes gerados, até o valor de similaridade mínima de 25%. Este valor é uniforme devido à relação

utilizada para calcular a porcentagem de objetos iguais entre os casos. Por exemplo, 2 objetos comuns entre dois casos contendo 8 objetos cada já obtém similaridade de 25%, o que também permite a união de casos utilizando-se os valores de 5%, 10%, 15% e 20%.

Nesta tabela, é observado o mesmo comportamento de aumento na quantidade de habitantes gerada, com o aumento da quantidade de objetos emprestada, porém é notada uma inversão no comportamento, com os casos concatenados e objetos estranhos. Esta ocorreu porque, nestes critérios, a similaridade depende da quantidade total de objetos presentes nos pares de casos, além dos objetos iguais: a quantidade total de objetos quando dois casos são concatenados é a soma das quantidades individuais destes e a quantidade total de objetos quando há mais objetos estranhos, também é maior do que sem a presença destes.

Por fim, para qualquer base de casos, a quantidade de habitantes detectados pelo CAPBC cresce conforme o aumento do valor de similaridade mínima, o que era esperado, devido a se tornar mais difícil a união de casos com valores altos de similaridade mínima necessária.

As tabelas 5.7 e 5.8 permitem verificar a variação na quantidade de habitantes detectada pelo classificador, fixando-se o valor de similaridade mínima e variando-se o peso, calculando este durante a entrada dos casos no classificador (Tabela 5.7) e após a entrada do último caso (Tabela 5.8). Na BC original, o valor de similaridade mínima foi fixado em 50%, na BC com objetos emprestados e casos concatenados, em 55% e na BC com objetos estranhos, em 35%.

Tabela 5.7 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos (critério 1).

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	385	125	101	94	94	87	88	86	84	82	77
2% Obj. Empr.	574	237	187	171	163	158	150	149	144	142	138
5% Obj. Empr.	599	272	212	190	172	169	167	161	159	157	154
10% Obj. Empr.	655	319	237	210	198	193	186	181	177	177	175
20% Obj. Empr.	729	388	315	287	272	258	255	248	244	244	238
2% Casos Conc.	559	228	179	157	146	139	132	133	132	131	127
5% Casos Conc.	581	248	200	174	163	155	148	149	148	147	144
10% Casos Conc.	585	254	201	179	168	159	153	152	151	150	148
20% Casos Conc.	635	254	210	194	188	176	170	168	166	165	163
2% Casos Obj. Estr.	54	24	16	16	16	15	16	15	13	13	13
5% Casos Obj. Estr.	54	22	17	17	14	13	13	13	12	12	12
10% Casos Obj. Estr.	75	29	21	20	17	17	16	15	14	14	14
20% Casos Obj. Estr.	78	36	26	28	24	22	20	20	19	19	18

Tabela 5.8 – Relação entre a quantidade de habitantes gerados pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto (critério 2).

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	385	125	107	96	93	88	89	88	87	87	85
2% Obj. Empr.	574	209	175	164	158	153	151	149	149	149	147
5% Obj. Empr.	599	240	196	174	174	168	166	163	161	161	160
10% Obj. Empr.	655	281	215	203	195	193	191	190	189	188	187
20% Obj. Empr.	729	357	294	267	258	251	246	245	244	243	240
2% Casos Conc.	559	204	164	148	140	137	136	138	137	136	136
5% Casos Conc.	581	226	180	164	157	156	153	152	151	151	151
10% Casos Conc.	585	231	190	172	165	160	156	154	153	154	154
20% Casos Conc.	635	240	198	189	177	175	175	172	172	172	170
2% Casos Obj. Estr.	54	23	22	21	20	19	19	18	18	17	17
5% Casos Obj. Estr.	54	23	23	21	20	19	19	18	17	16	15
10% Casos Obj. Estr.	75	26	27	25	24	23	23	21	21	20	19
20% Casos Obj. Estr.	78	37	32	31	29	27	27	25	25	24	23

A partir das duas tabelas, basicamente, é possível concluir que a aplicação de pesos aos objetos, para qualquer base de casos considerada, reduz a quantidade de habitantes gerados pelo CAPBC.

As relações de aumento, ou redução, da quantidade de habitantes, conforme o aumento das porcentagens de objetos emprestados, casos concatenados ou objetos estranhos são mantidas, proporcionalmente.

Entre os dois critérios de aplicação de peso, o primeiro (Tabela 5.7) mostrou melhor desempenho quanto à aproximação da quantidade real de habitantes, conforme o incremento no peso dos objetos, o que deve ter ocorrido devido aos menores pesos utilizados, dos objetos, para os cálculos de similaridade entre os casos.

A próxima seção mostra os resultados da análise da quantidade de classificações corretas, conforme a base de casos e valores de similaridade utilizados.

5.7.2 Quantidade de classificações corretas obtidas pelo CAPBC

Um importante critério de avaliação do CAPBC é a quantidade de classificações corretas feitas pelo sistema. Essas também foram calculadas pelo algoritmo avaliador de classificação e a síntese dos resultados pode ser vista nas tabelas apresentadas nesta seção.

A tabela 5.9 mostra a quantidade de classificações corretas obtidas, para cada base de casos testada, utilizando-se como critério de classificação a quantidade de objetos iguais.

Avaliando-se a variável da quantidade de classificações corretas, isoladamente, considerando-se que o resultado ideal é a classificação correta de todos os casos da base, pode-se afirmar que, quanto maior a quantidade destas, melhor o desempenho do classificador, e quanto menor a sua quantidade, pior o desempenho. Entretanto, esta variável sozinha não reflete o desempenho do classificador como um todo: é necessário avaliar-se, também, as quantidades de classificações incompletas e erradas.

Tabela 5.9 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	1 Objeto	2 Objetos	3 Objetos	4 Objetos	5 Objetos
BC Original	1000	884	718	653	565
2% Obj. Empr.	569	496	717	652	553
5% Obj. Empr.	569	493	423	644	535
10% Obj. Empr.	569	490	422	536	493
20% Obj. Empr.	568	481	412	352	379
2% Casos Conc.	303	304	253	227	367
5% Casos Conc.	298	300	253	228	221
10% Casos Conc.	305	303	258	231	227
20% Casos Conc.	310	311	273	240	232
2% Casos Obj. Estr.	303	305	719	653	565
5% Casos Obj. Estr.	303	305	719	653	565
10% Casos Obj. Estr.	303	301	367	657	565
20% Casos Obj. Estr.	303	303	357	653	566

Nota-se, na tabela, que para qualquer das bases de casos modificadas, há uma redução considerável na quantidade de classificações corretas, especialmente quando há casos concatenados.

As células coloridas em verde destacam o maior valor da variável, na avaliação da base de casos em questão; as células em laranja destacam os menores valores. Isso mostra como a variável se comporta em bases de casos modificadas de formas diferentes.

Pela distribuição dos melhores e piores valores, é possível se concluir que não há um critério ótimo para escolha do valor de similaridade mínima – utilizando-se a quantidade de objetos iguais como critério - visando maximizar a quantidade de classificações corretas, sem conhecer a base de casos que está sendo testada.

A tabela 5.10 mostra a avaliação da mesma variável, porém utilizando-se a porcentagem de objetos iguais como critério de similaridade. Os campos com hachuras representam valores que não foram simulados.

Nesta tabela, nota-se um desempenho muito mais uniforme quanto à distribuição dos melhores e piores valores obtidos para a variável analisada, em cada uma das bases de casos.

Tabela 5.10 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
BC Original	1000	1000	1000	1000	1000	992	951	838	545	464		
2% Obj. Empr.	569	569	569	569	569	570	780	664	375	298	171	115
5% Obj. Empr.	569	569	569	569	569	569	755	640	344	283	159	110
10% Obj. Empr.	569	569	569	569	569	567	677	496	309	273	125	105
20% Obj. Empr.	568	568	568	568	568	564	630	439	249	224	95	77
2% Casos Conc.	303	303	303	303	303	306	297	408	536	460	139	108
5% Casos Conc.	298	298	298	298	298	296	290	275	376	302	164	111
10% Casos Conc.	305	305	305	305	305	306	306	303	365	289	166	88
20% Casos Conc.	310	310	310	310	310	310	307	398	337	257	171	77
2% Casos Obj. Estr.	303	303	303	303	303	560	943	828	538	456	142	115
5% Casos Obj. Estr.	303	303	303	303	303	989	927	828	529	454	167	106
10% Casos Obj. Estr.	303	303	303	303	303	820	908	807	514	436	155	92
20% Casos Obj. Estr.	303	303	303	303	305	303	755	616	453	400	157	85

Também nota-se uma grande redução na quantidade de classificações corretas para todas as bases de casos modificadas, em relação à original, porém a escolha do valor de similaridade ótimo - em relação à análise desta variável - torna-se mais simples.

Pode-se notar, para todas as bases de casos modificadas, que a quantidade de classificações corretas permanece uniforme até certo valor de similaridade mínima, passa a crescer com o aumento deste, e volta a cair, posteriormente. Este comportamento está relacionado com a queda das classificações erradas e crescimento das classificações incompletas, que podem ser vistos nas próximas seções.

As tabelas 5.11 e 5.12 mostram o comportamento das classificações corretas com a aplicação dos pesos aos objetos.

Por meio das tabelas 5.11 e 5.12, pode-se perceber claramente que o desempenho do classificador, quanto à quantidade de classificações corretas, é muito superior aplicando-se os pesos aos objetos, em relação à classificação sem o uso destes.

Quanto ao peso que gera o melhor resultado, na maioria dos casos, observou-se que quanto maior o peso, melhor o desempenho da variável. Por outro lado, observa-se que para os pesos mais altos, a quantidade de classificações corretas oscila entre valores muito próximos. Ou seja, neste momento está se comportando de acordo com particularidades de

cada base de casos. Não é possível afirmar que, aumentando-se os pesos indefinidamente, a quantidade de classificações corretas também aumentará.

Tabela 5.11 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos (critério 1). A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	464	793	800	814	814	822	821	844	842	846	847
2% Obj. Empr.	171	662	718	723	722	725	727	729	749	750	760
5% Obj. Empr.	159	636	695	702	713	670	672	675	678	679	681
10% Obj. Empr.	125	512	587	615	627	629	633	635	638	639	639
20% Obj. Empr.	95	425	503	554	555	564	560	564	568	568	574
2% Casos Conc.	139	668	724	725	726	730	732	732	750	752	755
5% Casos Conc.	164	504	575	593	593	598	600	600	590	592	594
10% Casos Conc.	166	495	564	577	578	589	590	591	591	594	594
20% Casos Conc.	171	619	539	551	381	383	385	392	390	391	392
2% Casos Obj. Estr.	943	970	984	984	984	985	982	983	985	985	985
5% Casos Obj. Estr.	927	974	983	983	986	987	987	987	988	988	988
10% Casos Obj. Estr.	908	964	975	976	981	981	982	724	725	725	725
20% Casos Obj. Estr.	755	734	566	805	807	809	811	811	812	812	819

Tabela 5.12 – Relação entre a quantidade de classificações corretas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto (critério 2). A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	464	780	800	808	809	811	805	809	809	809	809
2% Obj. Empr.	171	688	710	721	718	724	725	725	725	725	725
5% Obj. Empr.	159	660	695	711	711	715	708	709	710	710	710
10% Obj. Empr.	125	548	616	626	634	635	636	636	637	637	641
20% Obj. Empr.	95	440	533	561	559	567	572	577	579	580	582
2% Casos Conc.	139	688	734	747	756	757	757	727	727	729	729
5% Casos Conc.	164	550	590	600	604	605	606	604	605	605	605
10% Casos Conc.	166	500	535	551	557	586	588	588	590	584	584
20% Casos Conc.	171	619	374	554	565	565	566	566	566	566	567
2% Casos Obj. Estr.	943	977	977	977	978	979	979	978	978	979	979
5% Casos Obj. Estr.	927	977	976	977	978	979	979	978	979	980	981
10% Casos Obj. Estr.	908	973	968	969	970	971	971	974	974	975	976
20% Casos Obj. Estr.	755	798	561	802	804	806	806	809	809	810	811

Quanto à comparação dos dois critérios de aplicação de pesos, o primeiro (tabela 5.11) apresentou um desempenho um pouco melhor do que o segundo (tabela 5.12).

Um resultado interessante, que pode ser observado comparando-se as tabelas de quantidades de habitantes detectados pelo sistema, apresentadas na seção 5.7.1, e da

quantidade de classificações corretas obtidas, é que um valor alto na quantidade de habitantes detectada não significa um resultado ruim, quando a quantidade de classificações corretas também é alta.

Por exemplo, na Tabela 5.7, na base de casos com 5% de objetos estranhos e incremento de peso igual a 1, foram detectados 12 habitantes pelo sistema; este valor é o triplo do valor esperado (4). Uma primeira análise da diferença entre estes valores levaria à conclusão de este é um resultado ruim, porém a Tabela 5.11 mostra que, para estes mesmos parâmetros de simulação, foram obtidas 988 classificações corretas. Ou seja, 98,8% dos casos classificados pertencem a 4 dos habitantes detectados, e os 1,2% restantes aos outros 8.

Assim, por meio da análise destes dados pode-se concluir que a avaliação do resultado, quanto à quantidade de habitantes, deve ser feita em conjunto com a avaliação das classificações corretas obtidas.

5.7.3 Quantidade de classificações incompletas obtidas pelo CAPBC

A quantidade de classificações incompletas, obtida pelo classificador de casos, representa os subconjuntos das combinações que não alcançaram a similaridade mínima com os conjuntos de classificações corretas de cada habitante. Entretanto, as classificações incompletas podem ser reclassificadas futuramente, sendo agrupadas a estes.

A tabela 5.13 mostra o comportamento das classificações incompletas, para todas as bases testadas, utilizando-se como critério de similaridade a quantidade de objetos iguais.

Avaliando-se, isoladamente, a variável de classificações incompletas, considerando-se que o resultado ideal do sistema é a ausência de classificações incompletas e erradas, pode-se afirmar que quanto menos classificações deste tipo houver em um conjunto de dados, melhor o desempenho do classificador e, quanto mais destas houver, pior o seu desempenho. Entretanto, as classificações incompletas estão intimamente ligadas às classificações erradas. É sempre melhor minimizar a quantidade de classificações erradas, mesmo que estas se tornem classificações incompletas, pois esta última pode ser reagrupada, futuramente, a um dos conjuntos de classificações corretas de um dos habitantes. De maneira geral, é possível observar que o aumento no valor de similaridade mínima aumenta a quantidade de classificações incompletas.

Tabela 5.13 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	1 Objeto	2 Objetos	3 Objetos	4 Objetos	5 Objetos
BC Original	0	116	282	347	435
2% Obj. Empr.	0	116	283	348	447
5% Obj. Empr.	0	121	283	356	465
10% Obj. Empr.	0	124	289	370	507
20% Obj. Empr.	1	136	300	386	545
2% Casos Conc.	0	101	272	342	427
5% Casos Conc.	0	93	260	333	419
10% Casos Conc.	0	82	254	323	410
20% Casos Conc.	0	56	207	286	370
2% Casos Obj. Estr.	0	103	278	347	435
5% Casos Obj. Estr.	0	97	281	347	435
10% Casos Obj. Estr.	0	83	266	343	435
20% Casos Obj. Estr.	0	79	275	347	434

Quanto aos critérios de modificação da base de casos, verifica-se que o aumento na porcentagem de objetos emprestados provoca também um crescimento na quantidade de classificações incompletas. Isto porque objetos que permitiam, anteriormente, atingir os valores de similaridade mínima, entre conjuntos de casos, foram substituídos por itens emprestados, o que reduziu os valores de similaridade entre alguns conjuntos de casos.

Em relação aos casos concatenados, quanto mais combinações de habitantes diferentes foram concatenadas entre si, menor ficou a quantidade de classificações incompletas. Devido à adição de conjuntos de objetos de um caso no outro, mais casos puderam ser agrupados entre si, pelo critério de quantidade de objetos iguais. Estas classificações, entretanto, tornaram-se classificações erradas.

Já nos casos com objetos estranhos, houve pouca alteração conforme o aumento nos casos contendo este tipo de objeto, já que a quantidade máxima destes, que foi adicionada aos casos, foi de 5 unidades. Apenas na similaridade mínima de 2 objetos iguais foi observada uma redução clara na quantidade de classificações incompletas, conforme o aumento na porcentagem de casos contendo objetos estranhos.

A tabela 5.14 mostra a avaliação da mesma variável, porém se utilizando o critério de similaridade da porcentagem de objetos iguais. Os campos com hachuras representam valores que não foram simulados.

Da mesma forma que na classificação pela quantidade de objetos iguais, foi verificado que o aumento no valor da similaridade mínima provocou um crescimento na quantidade de

casos incompletos, após o valor de similaridade mínima de 25%. Até este, as quantidades de classificações incompletas permaneceram praticamente constantes.

Tabela 5.14 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
BC Original	0	0	0	0	0	8	49	162	455	536		
2% Obj. Empr.	0	0	0	0	0	9	56	172	464	545	829	885
5% Obj. Empr.	0	0	0	0	0	13	69	184	498	560	838	887
10% Obj. Empr.	0	0	0	0	0	15	97	280	530	575	873	894
20% Obj. Empr.	1	1	1	1	1	28	155	350	590	627	862	895
2% Casos Conc.	0	0	0	0	0	5	48	163	460	539	860	892
5% Casos Conc.	0	0	0	0	0	11	54	109	467	546	828	886
10% Casos Conc.	0	0	0	0	0	9	59	148	474	554	813	906
20% Casos Conc.	0	0	0	0	0	5	71	181	476	572	818	917
2% Casos Obj. Estr.	0	0	0	0	0	8	57	172	462	544	858	885
5% Casos Obj. Estr.	0	0	0	0	0	11	73	172	471	546	833	894
10% Casos Obj. Estr.	0	0	0	0	0	14	92	193	486	564	845	908
20% Casos Obj. Estr.	0	0	0	0	0	2	22	85	224	546	599	843

Neste critério, é possível notar, para as bases de casos modificadas, que quanto mais modificadas estas são, mais classificações incompletas são obtidas. Na similaridade por quantidade de objetos iguais, este comportamento somente era observado para os objetos emprestados. Agora, devido à quantidade total de objetos influenciar no cálculo de similaridade, este comportamento passou a ocorrer também para as bases com casos concatenados e com objetos estranhos.

De maneira geral, o principal comportamento observado é o grande aumento na quantidade de casos incompletos, com o aumento no valor de similaridade mínima aplicado ao classificador.

As tabelas 5.15 e 5.16 mostram o comportamento das classificações incompletas com a aplicação de pesos aos objetos.

Por meio das tabelas 5.15 e 5.16, assim como o que foi observado para as classificações corretas, pode-se verificar que o desempenho do classificador, quanto à quantidade de classificações incompletas, é muito superior aplicando-se os pesos aos objetos, em relação à classificação sem o uso destes.

Tabela 5.15 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e alterando-se o peso dos objetos (critério 1). A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	536	207	200	186	186	178	179	156	158	154	153
2% Obj. Empr.	829	338	282	277	278	275	273	271	251	250	240
5% Obj. Empr.	838	355	296	286	275	317	315	312	309	308	306
10% Obj. Empr.	873	483	405	377	364	362	357	354	351	350	349
20% Obj. Empr.	862	504	425	376	420	409	413	407	402	402	396
2% Casos Conc.	860	327	271	270	269	265	263	263	245	243	240
5% Casos Conc.	828	351	297	277	276	270	268	268	262	260	258
10% Casos Conc.	813	352	297	280	277	266	265	264	263	260	260
20% Casos Conc.	818	332	291	284	272	262	258	251	244	242	241
2% Casos Obj. Estr.	57	30	16	16	16	15	18	17	15	15	15
5% Casos Obj. Estr.	73	26	17	17	14	13	13	13	12	12	12
10% Casos Obj. Estr.	92	36	25	24	19	19	18	15	14	14	14
20% Casos Obj. Estr.	85	102	26	33	29	27	25	25	24	24	17

Quanto ao peso que gera o melhor resultado, o comportamento verificado também foi semelhante ao das classificações corretas. Na maioria dos casos, observou-se que quanto maior o peso, melhor o desempenho da variável. Para os pesos mais altos, a quantidade de classificações incompletas oscila entre valores próximos, o que reflete particularidades das bases de casos analisadas. Também para a avaliação desta variável não é possível afirmar que, aumentando-se os pesos indefinidamente, a quantidade de classificações incompletas também será indefinidamente reduzida.

Tabela 5.16 – Relação entre a quantidade de classificações incompletas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto (critério 2). A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	536	220	200	192	191	189	195	191	191	191	191
2% Obj. Empr.	829	312	290	279	282	276	275	275	275	275	275
5% Obj. Empr.	838	331	295	278	278	271	278	277	276	276	276
10% Obj. Empr.	873	448	378	368	360	359	358	357	356	355	351
20% Obj. Empr.	862	535	443	412	407	399	394	390	388	387	384
2% Casos Conc.	860	307	261	248	239	238	238	268	268	266	266
5% Casos Conc.	828	325	281	269	265	264	263	265	264	264	264
10% Casos Conc.	813	361	330	312	304	275	273	273	271	277	277
20% Casos Conc.	818	335	283	282	269	268	267	266	266	266	264
2% Casos Obj. Estr.	57	23	23	23	22	21	21	22	22	21	21
5% Casos Obj. Estr.	73	23	24	23	22	21	21	22	21	20	19
10% Casos Obj. Estr.	92	27	32	31	30	29	29	26	26	25	24
20% Casos Obj. Estr.	85	39	33	36	34	32	32	29	29	28	27

Quanto à comparação dos dois critérios de aplicação de pesos, o primeiro (tabela 5.15) apresentou um desempenho um pouco melhor do que o segundo (tabela 5.16).

5.7.4 Quantidade de classificações erradas obtidas pelo CAPBC

Por fim, foi analisada a quantidade de classificações erradas obtidas pelo CAPBC, nas diferentes modificações da base de casos original, alterando-se os critérios e valores de similaridade utilizados. As classificações erradas, ao contrário das classificações incompletas, não são passíveis de serem reclassificadas. Se um caso, ou um conjunto deles, é incorretamente agrupado ao conjunto de casos de outro habitante, esse agrupamento não pode ser desfeito no decorrer da evolução das classificações. Por este motivo, a presença desse tipo de classificação, na base de casos, deve ser evitada tanto quanto for possível.

A tabela 5.17 mostra o comportamento das classificações erradas, para todas as bases de casos testadas, utilizando-se como critério de similaridade a quantidade de objetos iguais. Como o resultado ideal do sistema, quanto a esta variável, é a ausência de classificações erradas, foi considerado que, quanto menor a quantidade destas, melhor o resultado, e, quanto maior esta quantidade, pior o resultado.

Tabela 5.17 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a quantidade de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	1 Objeto	2 Objetos	3 Objetos	4 Objetos	5 Objetos
BC Original	0	0	0	0	0
2% Obj. Empr.	431	388	0	0	0
5% Obj. Empr.	431	386	294	0	0
10% Obj. Empr.	431	386	289	94	0
20% Obj. Empr.	431	383	288	262	76
2% Casos Conc.	697	595	475	431	206
5% Casos Conc.	702	607	487	439	360
10% Casos Conc.	695	615	488	446	363
20% Casos Conc.	690	633	520	474	398
2% Casos Obj. Estr.	697	592	3	0	0
5% Casos Obj. Estr.	697	598	0	0	0
10% Casos Obj. Estr.	697	616	367	0	0
20% Casos Obj. Estr.	697	618	368	0	0

Conforme era esperado, na tabela 5.17, foi possível observar que a quantidade de classificações erradas diminui, conforme o aumento no valor de similaridade mínima utilizado para a classificação, nos testes com todas as bases de casos.

Os piores resultados foram obtidos sempre no menor valor de similaridade utilizado para as classificações. Quanto aos melhores resultados, dependendo do critério e valores das modificações das bases de casos, foram obtidos com menores ou maiores valores de similaridade mínima utilizados.

Após uma análise geral da tabela, pode-se afirmar que, quanto maior a modificação feita na base de casos, para qualquer dos tipos de modificação da BC, mais classificações erradas foram obtidas.

A tabela 5.18 mostra a avaliação da mesma variável, porém se utilizando o critério de similaridade da porcentagem de objetos iguais. Os campos com hachuras representam valores que não foram simulados.

Assim como na classificação pela quantidade de objetos iguais, foi verificado que os piores resultados para a quantidade de classificações erradas foram obtidos nos menores valores utilizados de similaridade mínima e, conforme esses foram aumentados, as quantidades de classificações erradas foram reduzidas ou excluídas.

Nota-se, também, que quanto mais a base de casos original é modificada, para qualquer dos critérios utilizados, menor é a capacidade do CAPBC excluir todas as classificações erradas.

Tabela 5.18 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, utilizando-se como critério de similaridade a porcentagem de objetos iguais. A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Sim. Mínima ->	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
BC Original	0	0	0	0	0	0	0	0	0	0		
2% Obj. Empr.	431	431	431	431	431	421	164	164	161	157	0	0
5% Obj. Empr.	431	431	431	431	431	418	176	176	158	157	3	3
10% Obj. Empr.	431	431	431	431	431	418	226	224	161	152	2	1
20% Obj. Empr.	431	431	431	431	431	408	215	211	161	149	43	28
2% Casos Conc.	697	697	697	697	697	689	655	429	4	1	1	0
5% Casos Conc.	702	702	702	702	702	693	656	616	157	152	8	3
10% Casos Conc.	695	695	695	695	695	685	635	549	161	157	21	6
20% Casos Conc.	690	690	690	690	690	685	622	421	187	171	11	6
2% Casos Obj. Estr.	697	697	697	697	697	432	0	0	0	0	0	0
5% Casos Obj. Estr.	697	697	697	697	697	0	0	0	0	0	0	0
10% Casos Obj. Estr.	697	697	697	697	697	166	0	0	0	0	0	0
20% Casos Obj. Estr.	697	697	697	697	693	675	160	160	1	1	0	0

Apesar de terem sido observados bons desempenhos do classificador, em relação a esta variável, com o aumento no valor de similaridade mínima, deve-se lembrar que não

Tabela 5.20 – Relação entre a quantidade de classificações erradas obtidas pelo CAPBC, para cada base de casos testada, fixando-se o valor de similaridade mínima e classificando-se toda a base de casos após o cálculo do peso final de cada objeto (critério 2). A cor verde destaca o melhor valor obtido para cada conjunto de dados e a cor laranja, o pior.

Incremento de Peso ->	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BC Original	0	0	0	0	0	0	0	0	0	0	0
2% Obj. Empr.	0	0	0	0	0	0	0	0	0	0	0
5% Obj. Empr.	3	9	10	11	11	14	14	14	14	14	14
10% Obj. Empr.	2	4	6	6	6	6	6	7	7	8	8
20% Obj. Empr.	43	25	24	27	34	34	34	33	33	33	34
2% Casos Conc.	1	5	5	5	5	5	5	5	5	5	5
5% Casos Conc.	8	125	129	131	131	131	131	131	131	131	131
10% Casos Conc.	21	139	135	137	139	139	139	139	139	139	139
20% Casos Conc.	11	46	343	164	166	167	167	168	168	168	169
2% Casos Obj. Estr.	0	0	0	0	0	0	0	0	0	0	0
5% Casos Obj. Estr.	0	0	0	0	0	0	0	0	0	0	0
10% Casos Obj. Estr.	0	0	0	0	0	0	0	0	0	0	0
20% Casos Obj. Estr.	160	406	162	162	162	162	162	162	162	162	162

5.8 Síntese do capítulo

Neste capítulo, foram apresentados os dados utilizados para a simulação do sistema CAPBC proposto, bem como a avaliação do comportamento do sistema sobre cada tipo de alteração na base de casos e, por fim, a avaliação da qualidade das variáveis de classificação obtidas sobre diferentes modificações da base de casos original.

No próximo capítulo são apresentadas as considerações finais sobre o trabalho realizado.

6 CONCLUSÃO

Sistemas de automação residencial devem ter como objetivo principal tornar mais simples a interação entre os habitantes e os recursos da residência gerenciados por ela. Em trabalhos anteriores, sistemas com este objetivo foram desenvolvidos, onde as regras de automação eram criadas baseadas no comportamento cotidiano dos habitantes.

Este trabalho consistiu no desenvolvimento de um sistema de classificação de habitantes que permita a adaptação de sistemas de automação individuais, para o uso com diversos habitantes. Um sistema deste tipo possibilita que regras diferentes de automação possam ser criadas para diferentes pessoas.

Embora técnicas de identificação como reconhecimento biométrico, reconhecimento de faces por meio de câmeras, ou outros sistemas de identificação já existentes, pudessem ter sido utilizados para este propósito, a proposta do trabalho é minimizar a necessidade de adaptação do habitante ao seu sistema, sem que sua privacidade seja invadida.

Foi, então, proposto o uso de identificação de roupas e objetos por identidades eletrônicas contidas nestes itens, que são capturadas por sinais de rádio frequência, e o uso de uma técnica de inteligência artificial que permitisse abstrair informações úteis para a classificação desses códigos eletrônicos, por meio da similaridade entre suas combinações. Para os testes do sistema proposto, uma base de casos contendo combinações de objetos de habitantes foi gerada, respeitando características que possivelmente existiriam em uma residência real.

Foram testados diversos critérios e valores de variáveis sobre a base de casos gerada, bem como modificada a própria base de casos com o objetivo de simular situações que poderiam ocorrer em situações reais. O sistema se comportou conforme o esperado, agrupando combinações de objetos pertencentes aos mesmos habitantes sob códigos genéricos criados pelo algoritmo desenvolvido. Comportamentos específicos foram obtidos conforme a alteração dos critérios e valores de similaridade de casos utilizados, bem como modificando-se a base de casos.

Utilizando-se a base de casos original, não foram obtidas classificações erradas para nenhum critério simulado, desta forma, em um sistema real, valores menores de similaridade mínima poderiam ser utilizados com o objetivo de obter uma convergência mais rápida das classificações para o número correto de habitantes. Modificando-se a base de casos, com o acréscimo ou permuta de itens e concatenação de casos, foram obtidas classificações erradas pelo agrupamento incorreto de combinações de habitantes diferentes.

A variação nos critérios e valores de similaridade mínima, aplicados nos testes, mostrou que estes possibilitam reduzir ou eliminar classificações erradas do sistema, porém podem reduzir, também, a quantidade de classificações corretas. Foi verificado que os critérios e valores de similaridade ótimos no sistema dependem das características da base de casos testada, ou seja, para a implantação do sistema em residências reais, seria necessário conhecer os hábitos dos moradores, a fim de se calibrar as variáveis do sistema.

Como principais pontos positivos obtidos neste trabalho, podem ser citados:

- Foi observado que é possível agrupar, por similaridade, combinações de roupas e objetos de pessoas em uma residência, para propósitos de adaptação de sistemas individuais de automação residencial inteligente para mais de um habitante;

- Como a base de casos utilizada nos testes foi gerada aleatoriamente, podem ter sido criadas combinações improváveis de objetos, embora possíveis. Por exemplo, embora seja possível combinar-se um terno preto com uma calça social branca, esta combinação é improvável. Assim sendo, há indícios que, na prática, haveria menos combinações de objetos do que as obtidas pelo simulador, o que, possivelmente, tornaria a convergência mais rápida das classificações, para os valores corretos. Também não foram consideradas combinações favoritas de itens dos habitantes, o que também deve permitir uma classificação mais rápida dos casos, se utilizado;

- Foi verificado que a manutenção dos valores de similaridade mínima e dos pesos utilizados permite obter uma otimização nos resultados. Neste trabalho, não foi implementada uma política de manutenção automática destes parâmetros, mas há indícios que o uso de outros dados no classificador, presentes nas etiquetas de RFID, ou no próprio sistema de identificação biométrica dos acessos ao interior da residência, poderiam ser utilizados para ajustar automaticamente parâmetros do classificador, de forma a otimizar seus resultados.

Como pontos negativos, é possível citar:

- Todos os dados utilizados em simulações foram gerados em simuladores. Desta forma, particularidades que poderiam ser observadas em testes efetuados em residências reais não puderam ser verificadas no sistema. Também foi utilizada uma base de casos única, nos testes, com número fixo de habitantes. Para ser possível uma avaliação melhor do sistema, seria interessante a criação de novas bases de casos e a execução de novos testes para observar modificações no comportamento do classificador;

- Para que o sistema possa funcionar em uma residência real, seria necessário que as peças de roupas e os objetos dos habitantes possuíssem etiquetas de RFID, além de ser

necessária a instalação de leitores RFID nas residências e adaptação do sistema de automação instalado;

- Como novas combinações capturadas pelo sistema são comparadas a todas aquelas já armazenadas na base de casos, o sistema, embora continue funcionando indefinidamente, se tornará inevitavelmente mais lento com o passar do tempo, conforme novos casos forem recebidos e armazenados no sistema.

Com base nos resultados obtidos e nos pontos positivos e negativos observados, na seção a seguir são apresentadas algumas sugestões de trabalhos futuros, com o objetivo de aperfeiçoar o sistema implementado neste trabalho.

6.1 Trabalhos Futuros

Após avaliados os resultados obtidos, novos trabalhos podem ser desenvolvidos, com o intuito de aperfeiçoar o sistema e resolver seus pontos negativos. Algumas sugestões são apresentadas a seguir:

6.1.1 Teste do sistema com outras bases de casos, com características e quantidades de habitantes diferentes

Neste trabalho, foi observado o comportamento do classificador utilizando-se uma base de casos única, contendo combinações de objetos de 4 habitantes diferentes. A quantidade de objetos de cada um dos habitantes foi cadastrada de acordo com um valor médio aproximado daquele que pessoas reais possuiriam.

Foi verificado, nos testes onde a base de casos foi modificada, que os resultados das classificações é alterado conforme objetos de outros habitantes são adicionados a estes casos.

Seria interessante, porém, realizar testes modificando toda a base de casos, alterando a quantidade de habitantes, as quantidades de objetos de cada tipo para cada habitante, os hábitos de empréstimo de certos tipos de objetos, e outros não, entre eles, avaliando agora as alterações nas classificações conforme a base de casos utilizada.

Também seria importante verificar comportamentos particulares, em situações específicas, que não foram geradas pelo semeador de casos, que cria as situações a partir de parâmetros médios informados. Por exemplo, uma situação em que a empregada carrega um cesto de roupas sujas com itens de todos os habitantes, presença de visitantes na residência,

um habitante carregando uma cesta de compras com várias dezenas de itens, entre outras possibilidades.

6.1.2 Uso de outras informações do sistema RFID ou sistema de automação na classificação

Conforme foi exposto anteriormente, identificadores RFID armazenam outras informações em sua memória, além de uma simples identidade eletrônica. Um teste interessante que poderia ser feito, é considerar o uso de outras informações, como por exemplo o tipo de objeto, nas classificações.

Com o uso destes parâmetros, é provável que as classificações se tornem mais rápidas e precisas, pois determinados tipos de objetos refletem melhor a identidade do habitante (como óculos, por exemplo) do que outros (como uma camisa).

Outro dado que poderia auxiliar o CAPBC nas classificações seria o uso das informações do sistema de reconhecimento biométrico presente na entrada da residência. Se um leitor RFID fosse instalado junto ao leitor biométrico, certas classificações poderiam ser certificadas por este, e outras, erradas, eliminadas.

6.1.3 Implementação de uma política de reclassificação de casos

Da forma que o CAPBC foi implementado, se determinado caso for classificado de forma errada, o algoritmo atual não permite uma reclassificação posterior deste.

É um fato que, quanto mais casos a BC possuir, maior a quantidade de informação sobre combinações e objetos esta possuirá. Além disto, existindo a possibilidade do uso de outras informações pelo CAPBC, como outros dados do identificador RFID ou do sistema de identificação biométrica, para certificação da classificação de outros casos, seria interessante manter a possibilidade de o algoritmo reclassificar uma parte dos casos, ou até mesmos todos eles, quando determinado critério fosse atendido.

Isto tornaria o sistema mais flexível em relação a possíveis erros que venham a ocorrer, caso um critério incorreto de classificação seja escolhido no início do funcionamento do sistema.

6.1.4 Estudo sobre a possibilidade do uso de apenas um subconjunto da BC na classificação de novos casos

Foi observado que o tempo necessário para a classificação de novos casos aumenta, proporcionalmente, conforme o crescimento da base de casos. Isto porque cada novo caso é comparado a todos os casos anteriores da base, para sua classificação.

Isto pode ser um inconveniente quando a base de casos se tornar muito extensa, pois, para propósitos práticos de um sistema de automação residencial, o evento disparado por este não pode demorar mais do que poucos segundos.

Desta forma, é proposto para estudos futuros que ao invés da comparação dos novos casos a todos os demais armazenados na base, este seja comparado apenas a um subconjunto dela, que poderiam ser os casos mais atuais ou os casos mais vezes utilizados, no passado, para a classificação de outras combinações.

Um método possível para a solução de problemas em RBC, utilizando apenas um subconjunto da BC, é o uso de casos “footprint” (SMYTH, MCKENNA, 2001), onde a base de casos completa é representada por um conjunto menor de casos, cuja competência para a solução de problemas é a mesma da base de casos completa (TONIDANDEL, 2003).

6.1.5 Teste do sistema em uma residência real

Por fim, para confirmar o funcionamento do sistema proposto, na prática, são sugeridos testes em uma residência real.

Para que isto seja possível, são necessárias algumas adaptações da residência de teste, como a instalação de um sistema de leitores RFID, nos pontos onde se desejar a detecção do habitante, e de um sistema computadorizado, que receba as informações do middleware RFID, inserindo-as no algoritmo classificador de casos.

Além disto, os itens de vestuário dos habitantes, bem como seus objetos pessoais, teriam de receber identificadores RFID para poderem ser detectados pelos leitores.

Isto permitiria verificar tanto o comportamento dos habitantes, referente a combinações de vestuário e objetos, quanto das variáveis do sistema classificador, agora utilizando-se dados reais nos testes.

Seria possível, também, posteriormente, fazer os testes enviando-se os resultados de classificações efetuadas pelo CAPBC para um sistema de automação implantado.

REFERÊNCIAS BIBLIOGRÁFICAS

AAMODT, A.; PLAZA, E. **Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches**. AICOM, 7(1), 1994

BOLZANI, C. A. M. **Desenvolvimento de um simulador de controle de dispositivos residenciais inteligentes: uma introdução aos sistemas domóticos**. 2004. Dissertação (Mestrado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo.

DAUGMAN, J. **The importance of being random: statistical principles of iris recognition**. 2001. The Computer Laboratory. University of Cambridge.

FINKENZELLER, K **RFID Handbook. Fundamentals and Applications in Contactless Smart Cards and Identification: second edition**. Wiley, 2003

GLOVER, B.; BHATT, H. **Fundamentos de RFID**. Altabooks, 2007.

HASTIE, T.; TIBSHIRANI, R, FRIEDMAN, J **The Elements of Statistical Learning: second edition**. Springer-Verlag, 2008.

KAZIENKO, J. F. **Assinatura Digital de Documentos Eletrônicos Através da Impressão Digital**. 2003. Dissertação (Mestrado) – Universidade Federal de Santa Catarina. Florianópolis.

KOLODNER, J. L. **Case-Based Reasoning**. Morgan Kaufmann, San Francisco, California, 1993.

MACHADO, V.S.; ALMEIDA, J.R. **Sistemas de reconhecimento biométrico aplicados à segurança de ambientes físicos**. São Paulo, Universidade de São Paulo, 2004.

MURATORI, J.R. **As tendências do mercado de automação residencial**. São Paulo, Congresso Habitar – Congresso de automação residencial e tecnologias para habitação, 2005.

NASCIMENTO, A.F. **Sistema dinâmico de automação residencial**. Rio de Janeiro, Dissertação de mestrado, Instituto Militar de Engenharia, 2002.

PEARCE, D.; HIRSCH, H. G. **The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems Under Noisy Conditions**. Beijing, China. 6th International Conference of Spoken Language Processing, 2000.

QUINLAN, J. R. **Induction of decision trees**. Machine Learning 1. 1986.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, California, EE.UU, Morgan Kauffmann Publishers. 1993.

RIESBECK, C. K.; SCHANK, R. C. **Inside Case-Based Reasoning**. Lawrence Erlbaum. 1989.

RFID JOURNAL. **Benetton to Tag 15 Million Items**. 2003. Disponível em: <<http://www.rfidjournal.com/article/articleview/344/1/1/>>. Acesso em 30 ago. 2008.

RFID JOURNAL. **What is RFID**. [200-]a. Disponível em: <<http://www.rfidjournal.com/article/articleprint/1339/-1/129/>>. Acesso em 30 ago. 2008.

RFID JOURNAL. **RFID Consumer Applications and Benefits**. [200-]b. Disponível em: <<http://www.rfidjournal.com/article/articleprint/1332/-1/129/>>. Acesso em 30 ago. 2008.

RFID JOURNAL. **RFID Business Applications**. [200-]c. Disponível em: <<http://www.rfidjournal.com/article/articleprint/1334/-1/129/>>. Acesso em: 30 ago. 2008.

ROSA, P. F. F. et al. **Smart house with a reactive architecture and a neural network for gait recognition**. Rio de Janeiro, Instituto Militar de Engenharia, 2005.

SGARBI, J.A. **Domótica inteligente: automação residencial baseada em comportamento**. 2007. Dissertação (Mestrado em Engenharia Elétrica) - Centro Universitário da FEI, São Bernardo do Campo.

STONHAM, T. J. **Practical face recognition and verification with WISARD**. Dordrecht, 1986.

SMYTH, B.; MCKENNA, E. **Footprint-Based Retrieval**. Dublin, University College Dublin, 2001. Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.9358>>. Acesso em 03 set. 2008.

TONIDANDEL, F. **Desenvolvimento e implementação de um sistema de planejamento baseado em casos**. 2003. Tese (Doutorado em Engenharia) – Universidade de São Paulo, São Paulo.

TONIDANDEL, F; TAKIUCHI, M; MELO, E. **Domótica Inteligente: Automação baseada em comportamento**. Congresso Brasileiro de Automática, 2004.

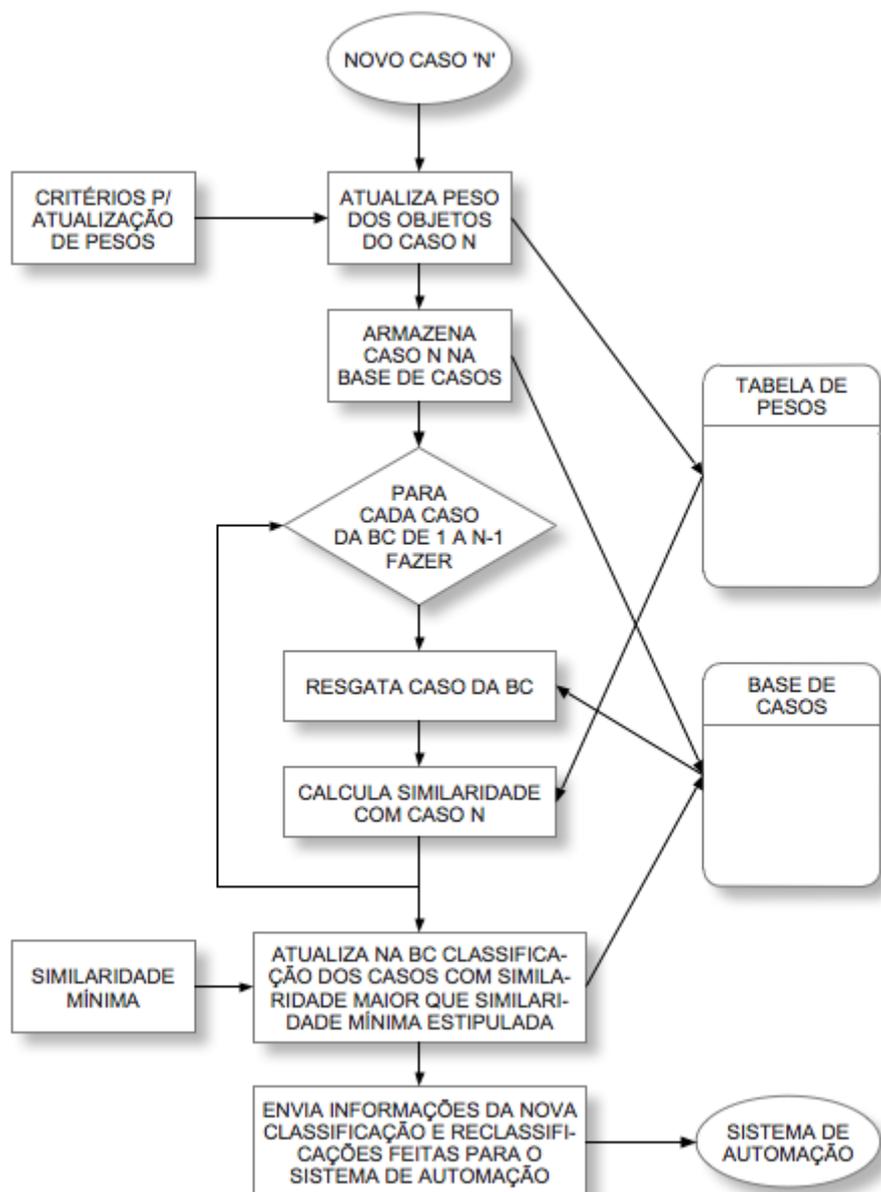
TURK, M. A.; PENTLAND, A. P. **Face Recognition Using Eigenfaces**. 1991. Vision and Modeling Group. The Media Laboratory Massachusetts Institute of Technologies.

WANGENHEIM, C.G.; WANGENHEIM, A. **Raciocínio Baseado em Casos**. 1ª edição brasileira. Manole, 2003.

WESS, S. **Fallbasiertes Problemlösen in Wissenbasierten Systemen zur Entscheidungsunterstützung und Diagnostik**. 1995. PhD Thesis - University of Kaiserslautern, Germany.

APÊNDICE 1

Diagrama esquemático do software classificador de casos (CAPBC).



APÊNDICE 2

Lista de combinações de roupas e objetos utilizada pelo semeador de casos para a criação de novas combinações.

Combinações masculinas. AD = Adulto; CR = Crianças; AC = Adultos e Crianças.

Tipo de Combinação	AD	AD	AC	AC								
Babydolls												
Bermudas					100%	100%						
Biquinis												
Blusas		50%	20%	10%								
Bonés				10%	10%	10%	30%					
Botas												
Cacharréis												
Cachecóis	5%	5%										
Calças Jeans			100%	100%								
Calças Sociais	100%	100%										
Calcinhas												
Camisas Sociais	100%	100%	100%									
Camisetas				100%	90%	90%						
Camisetas Regatas							90%	50%	50%	20%		
Camisolas												
Chapéus	5%	5%										
Chinelos						90%	80%	50%	30%			
Cintos	100%	100%	100%	70%								
Cuecas	100%	100%	100%	100%	90%	90%	70%	50%	100%		50%	
Gravatas	50%	30%										
Jaquetas			40%	30%								
Lenços	30%	30%	30%	10%	10%	10%						
Meias Calças												
Meias Esporte				80%	50%							
Meias Sociais	100%	100%	100%									
Minissaias												
Pantufas												
Pijamas								100%				
Pulôver	10%	20%	20%									
Saias												
Sandálias												
Sapatos	100%	100%	100%									
Shorts							100%		20%	20%		
Sobretudos	10%	10%										
Sungas										100%		
Sutiãs												
Tênis				100%	100%							
Ternos	70%											
Toalhas										50%	100%	
Vestidos												
Anéis	90%	90%	90%	70%	70%	70%	30%	20%	20%	20%	20%	
Bolsas												
Braceletes												
Brincos												
Canetas	30%	30%	30%									
Carteiras	80%	80%	80%	60%	70%	40%	20%					
Chaveiros	80%	80%	80%	40%	50%	30%	30%					
Colares												
Óculos	80%	80%	80%	50%	50%	50%	40%	20%	20%			
Óculos Escuros			40%	40%	40%	40%	60%			50%		
Pen Drives	70%	70%	40%	10%								
Presilhas de Cabelo												
Relógios de Pulso	80%	80%	80%	60%	30%	30%	20%					
Telefones Celulares	90%	90%	90%	90%	80%	70%	40%					
Walkie Mans	20%	20%	40%	40%	20%	20%	30%					

Combinações femininas. AD = Adulto; CR = Crianças; AC = Adultos e Crianças.

Tipo de Combinação	AD	AD	AD	AD	AD	AC	AC	AC									
Babydolls													100%				
Bermudas																	
Biquínis																	100%
Blusas	40%					70%	50%										
Bonés																	
Botas						100%											
Cacharréis								70%									
Cachecóis						30%											
Calças Jeans						100%	100%	100%									
Calças Sociais	100%																
Calcinhas	100%	95%	95%	95%	95%	100%	100%	100%	90%	100%	100%	70%	70%	100%	100%		50%
Camisas Sociais	100%	100%		100%		100%											
Camisetas							100%					100%				80%	
Camisetas Regatas								100%	100%	100%							
Camisolas													100%				
Chapéus																	
Chinelos																50%	50%
Cintos	20%	20%		20%		20%											
Cuecas																	
Gravatas																	
Jaquetas	50%					50%											
Lenços	50%	50%		50%		50%											
Meias Calças		100%	90%														
Meias Esporte						100%	100%					50%					
Meias Sociais	100%																
Minissaias											100%						
Pantufas												70%	70%				
Pijamas														100%			
Pulôver																	
Saias		100%		100%						100%							
Sandálias				100%	100%					100%	100%						
Sapatos	100%	100%	100%														
Shorts												100%					
Sobretudos	20%	20%															
Sungas																	
Sutiãs	100%	100%	70%	100%	70%	100%	80%	80%	80%	50%	70%			20%			20%
Tênis							100%	100%			100%						
Ternos																	
Toalhas																50%	100%
Vestidos			100%		100%												
Anéis	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	70%	70%	70%	70%	30%	30%
Bolsas	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%					
Braceletes	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%					
Brincos	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	70%	70%	70%	70%	50%	50%
Canetas	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%						
Carteiras	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%						
Chaveiros	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%						
Colares	80%	80%	80%	80%	80%	80%	80%	80%	80%	80%	80%						
Óculos	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	20%	20%	20%	20%		
Óculos Escuros	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%					50%	
Pen Drives	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%						
Presilhas de Cabelo	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	30%	30%	30%	30%	30%	50%
Relógios de Pulso	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%						
Telefones Celulares	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%						
Walkie Mans	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%						