

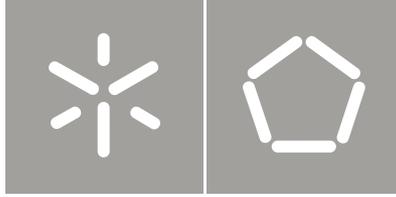


Universidade do Minho
Escola de Engenharia

Vitor Manuel Freitas Marques
iLeisure: an Intelligent Free Time Scheduler based on User Profiling

Vitor Manuel Freitas Marques

iLeisure: an Intelligent Free Time Scheduler
based on User Profiling



Universidade do Minho
Escola de Engenharia

Vítor Manuel Freitas Marques

iLeisure: an Inteligente Free Time Scheduler
based on User Profiling

Tese de Mestrado
Mestrado em Engenharia Informática

Trabalho efectuado sob a orientação do
Paulo Novais
Ângelo Costa

Agradecimentos

À minha família que com muito apoio e carinho não mediram esforços para que eu chegasse a esta etapa da minha vida.

O meu profundo e sentido agradecimento ao amigo Manuel José Pinto da Costa e sua família pelo apoio e carinho com que sempre me acolheram, nestes últimos dois anos longe da minha família.

Aos meus orientadores, o Professor Doutor Paulo Jorge Freitas de Oliveira Novais e o Mestre Ângelo Gonçalo Araújo Silva Costa pela competência científica e acompanhamento do trabalho, pela disponibilidade e generosidade reveladas ao longo destes dois anos de trabalho, assim como pelas críticas, correções e sugestões relevantes feitas durante a orientação.

Aos colegas do ISLAB pelo incentivo e pelo apoio constantes.

E a todos aqueles que de alguma forma direta ou indireta contribuíram na construção deste trabalho.

Resumo

O envelhecimento populacional é uma realidade dos nossos tempos. O envelhecimento provoca muitas vezes o problema do sedentarismo, que é associado à obesidade e doenças cardiovasculares. O sedentarismo é por isso um problema preocupante que para além de afetar as idades avançadas afeta todas as outras faixas etárias. É nos tempos livres que as pessoas são mais sedentárias.

A prática de atividades físicas combate o sedentarismo e é benéfica em termos psicológicos, sociais e económicos. É necessário incentivar as pessoas a praticá-la e uma das formas de o conseguir é através de um sistema que recomende tais atividades. A recomendação deverá contudo ter em consideração as preferências das pessoas pelas atividades, para que estas tendam a aceitar as atividades que são recomendadas. As preferências de uma pessoa são associadas ao conceito de perfil e é através deste, que um sistema de recomendação infere sobre que atividades recomendar. É por isso imperativo que o perfil esteja constantemente atualizado e consistente com a pessoa que representa. Garantir estas propriedades num perfil não é computacionalmente claro, devido às propriedades naturais de um perfil e devido à capacidade das pessoas em evoluir e variar nas suas preferências.

Este trabalho tem como objetivo desenvolver um sistema que recomende às pessoas atividades que não sejam sedentárias e que incentivem o convívio social. Para que o sistema seja preciso nas atividades que recomenda, é também objetivo deste trabalho desenvolver um mecanismo para obtenção, representação e modulação de perfis que garanta a sua dinamicidade, que garanta que os perfis sejam eficientemente adaptáveis e atualizáveis.

Realizaram-se observações acerca da taxa de aceitação de utilizadores, em relação às atividades recomendadas pelo sistema desenvolvido. Estas observações assistiram na avaliação da precisão do sistema em recomendar e a sua capacidade em acompanhar os utilizadores e aprender acerca das suas preferências. Consultando periodicamente a quantidade de atividades recomendadas aceites ou rejeitadas, verificou-se que o número de atividades aceites tende a aumentar e as rejeitadas a diminuir. Estes resultados demonstram que o sistema possui a capacidade de aprender e aperfeiçoar o perfil dos seus utilizadores, tornando-se mais preciso nas atividades que recomenda. Acredita-se desta forma que o mecanismo desenvolvido para a obtenção e modulação de perfis dinâmicos é eficiente e robusto.

Abstract

The population aging is a known reality. Retirement often causes the inactivity, that is related to obesity and cardiovascular diseases. Physical inactivity is a serious problem that affects all people, not merely older people. Leisure activities can become sedentary, when people choose to practice activities like television and computer.

The practice of physical activity addresses the inactivity and is helpful in psychological, social and economic terms. It is essential to encourage people to practice them. An approach to achieve this need is through a system that recommend activities to people. The recommendation should take into account the people preference's for the activities, so that they tend to accept the recommendation. The preferences of a person are related to the concept of profile. The recommendation system uses the profile to infer about the activities to recommend. Therefore, it is vital that the profile is constantly updated and consistent with the person who it represents. Ensuring these properties in a profile is not a clear task because of the natural properties of a profile, and because of the person's capability to evolve and change their preferences.

This work aims to develop a system that recommends physical and socialization activities. It is also aimed to develop a mechanism to obtain, represent and module profiles that ensure their dynamism, that ensure that the profiles are efficiently adapted and updated. Based on these profiles, the system can be accurate in the activities that it recommends.

Observations regarding the rate of acceptance of users for the activities recommend by the developed system were performed. These observations helped in the evaluation of the system accuracy in recommending and its ability to follow the users and learn their preferences. By periodically checking the number of recommended activities accepted or rejected, it was found that the number of accepted activities tends to raise, and the number of rejected to decrease. These results have proven that the system is capable of learning and improve the user profile, becoming more precise in the recommended activities.

It is believed that the developed mechanism to confine and module dynamic profiles is efficient and robust.

Conteúdo

Agradecimentos.....	iii
Resumo.....	v
Abstract.....	vii
Conteúdo	ix
Lista de Figuras.....	xi
Lista de Tabelas	xiii
Lista de Abreviaturas	xv
Capítulo 1: Introdução	1
1.1 - Envelhecimento populacional.....	2
1.1 - Solidão, Depressão e Sedentarismo no Envelhecimento	4
1.3 - Sedentarismo, um problema para todas as Faixas Etárias	5
1.4 - Atividade Física.....	11
1.5 - Perfis e suas Implicações.....	14
1.6 - Projetos Relacionados.....	16
1.7 - Tema e Objetivos	18
1.8 - Metodologia de Investigação e Calendarização	19
1.9 - Estruturação do Documento.....	23
Capítulo 2: Sistemas Inteligentes	25
2.1 - Sistemas de Recomendação	26
2.1.2 - Recomendação baseada no conteúdo	28
2.1.3 - Recomendação colaborativa.....	30
2.2 - Ambientes Inteligentes.....	32
2.3 - Conclusões.....	35
Capítulo 3: Personalização.....	37
3.1 - Modulação de um Perfil	38

3.1.1 - Redes Bayesianas	38
3.1.2 - Raciocínio Baseado em Casos.....	46
3.1.3 - Regras de Associação	53
3.2 - Aquisição de um Perfil	59
3.2.1 - Aquisição Explícita	59
3.2.2 - Aquisição Implícita	60
3.2.3 - Comparação	61
3.3 - Conclusões.....	63
Capítulo 4: iLeisure - Arquitetura	65
4.1 - Definição do Domínio.....	66
4.2 - Arquitetura e Comunicação.....	67
4.3 - Aplicação Cliente	74
4.4 - Conclusões.....	84
Capítulo 5: iLeisure - Personalização	85
5.1 - Representação de um Perfil	86
5.2 - Aprendizagem.....	93
5.3 - Recomendação de Eventos	104
5.4 - Conclusões.....	107
Capítulo 6: Conclusões e Trabalho Futuro	109
6.1 - Síntese de Trabalho	110
6.2 - Trabalho Relevante Realizado.....	112
6.3 - Trabalho Futuro	113
Bibliografia	115

Lista de Figuras

Figura 1 - Esperança média de vida à nascença por sexo em Portugal.....	2
Figura 2 - Índice de Envelhecimento em Portugal	3
Figura 3 - Taxa bruta de natalidade (por mil habitantes) em Portugal	3
Figura 4 - Percentagem de Obesidade e Pré-Obesidade nos Rapazes.....	7
Figura 5 - Percentagem de Obesidade e Pré-Obesidade nas Raparigas	7
Figura 6 - Prevalência da obesidade na população adulta da Europa.....	8
Figura 7 - Ocupação dos tempos livres por atividade em alguns países europeus.....	9
Figura 8 - Esquema dos módulos do iGenda.....	17
Figura 9 - Ciclo do processo de investigação-ação	20
Figura 10 - Modelo proposto por Kemmis.....	21
Figura 11 - Exemplo de uma casa inteligente	34
Figura 12 - "Máquinas Inteligentes": Um conflito entre ambientes inteligentes e o utilizador....	35
Figura 13 - Exemplo de uma Rede Bayesiana.....	39
Figura 14 - Inferência de Diagnóstico	40
Figura 15 - Inferência de Prognóstico	41
Figura 16 - Exemplo de uma Rede Bayesiana Complexa.....	41
Figura 17 - Resultado da Aprendizagem de uma Rede Bayesiana	42
Figura 18 - Assistente do Office.....	43
Figura 19 - Exemplo de uma Rede Bayesiana no projeto de Gamboa & Fred.....	44
Figura 20 - Exemplo de uma Rede Bayesiana com nós bidimensionais	46
Figura 21 - Funcionamento do Raciocínio Baseado em Casos.....	47
Figura 22 - Exemplo de um caso do Raciocínio Baseado em Casos	48
Figura 23 - Cálculo da similaridade local entre dois casos	48
Figura 24 - Exemplo da representação de um perfil no projeto de Godoy & Amandi	52
Figura 25 - Funcionamento das Regras de Associação.....	56
Figura 26 - Arquitetura do iLeisure	68
Figura 27 - Estruturação JSON para um evento	70
Figura 28 - Estruturação JSON para uma notificação.....	70
Figura 29 - Registo no C2DM e servidor	72
Figura 30 - Envio de mensagem do servidor para um telemóvel.....	74

Figura 31 - Interface principal da aplicação móvel	76
Figura 32 - Menu de contexto para um evento.....	76
Figura 33 - Painel com informação completa de um evento.....	77
Figura 34 - Menu de contexto para um determinado dia.....	77
Figura 35 - Menu de contexto para um determinado horário.....	78
Figura 36 - Notificação da proximidade de um evento.....	78
Figura 37 - Menu acionado pela notificação com a informação do evento	78
Figura 38 - Menu principal da aplicação.....	79
Figura 39 - Menu de criação de um evento.....	79
Figura 40 - Menu de aviso da existência de conflitos com o novo evento.....	80
Figura 41 - Exemplo de resolução de conflitos.....	80
Figura 42 - Menu de apresentação de eventos propostos por amigos do utilizador	81
Figura 43 - Menu de contexto para os eventos propostos por amigos.....	81
Figura 44 - Menu com informação acerca de como o evento colide com os demais	82
Figura 45 - Menu para o utilizador informar a razão da sua ausência em eventos	82
Figura 46 - Notificação de novos eventos propostos por conhecidos	82
Figura 47 - Notificação de confirmação da presença de participantes	83
Figura 48 - Notificação de ausência de participantes	83
Figura 49 - Informação acerca da presença dos participantes.....	83
Figura 50 - Mensagem de justificação de ausência num evento	84
Figura 51 - Armazenamento de um perfil na base de dados	92
Figura 52 - Adaptação do Raciocínio baseado em Casos	99
Figura 53 - Processo de Aprendizagem do Perfil.....	102
Figura 54 - Modulação da base de dados para a agenda de um utilizador.....	103
Figura 55 - Estrutura em árvore usada para a recomendação de eventos.....	105
Figura 56 - Exemplo de uma estrutura em árvore usada para a recomendação de eventos	106

Lista de Tabelas

Tabela 1 - Percentagem dos Inquiridos que nunca praticam exercício físico por país	6
Tabela 2 - Calendarização do desenvolvimento do trabalho	22
Tabela 3 - Diagrama de Gantt	22
Tabela 4 - Exemplo do universo de um sistema de recomendação	27
Tabela 5 - Tabela de Probabilidade Condicional	39
Tabela 6 - Comparação entre as técnicas explícitas e implícitas	63
Tabela 7 - Atributos de um evento utilizado na representação de um perfil	66
Tabela 8 - Restantes atributos de um evento	67
Tabela 9 - Solução inicial para a representação de um perfil	86
Tabela 10 - Casos passados similares com alto valor de similaridade	96
Tabela 11 - Casos passados similares com baixo valor de similaridade	96

Lista de Abreviaturas

Sigla em Inglês	Sigla em Português	Significado em Inglês	Significado em Português
	INE		Instituto Nacional de Estática
	IMC		Índice de Massa Corporal
Aml	Aml	Ambient Intelligence	Ambientes Inteligentes
BN	RB	Bayesian Networks	Redes Bayesianas
CBR	RBC	Case Based Reasoning	Raciocínio Baseado em Casos
AR	RA	Association Rules	Regras de Associação
C2DM	C2DM	Android Cloud to Device Messaging	

Capítulo 1: Introdução

Os crescentes avanços científicos fazem com que a esperança de vida seja cada vez maior. Isto demonstra que a qualidade de vida tem vindo cada vez mais a melhorar e que os investimentos feitos na melhoria das condições de vida têm surgido efeito. Com o aumento da esperança de vida, ganha vida o envelhecimento populacional que atualmente é um fenómeno global cada vez mais preocupante.

O envelhecimento é um processo natural da vida que deve ser feito de uma forma saudável. São encaradas com grande importância pela Organização Mundial da Saúde e a Comissão da União Europeia todas as medidas, políticas ou práticas que ajudam para um envelhecimento saudável. Assim, devem ser tomadas medidas capazes de solucionar problemas característicos do envelhecimento como a solidão, depressão ou sedentarismo.

Embora o sedentarismo seja muitas vezes relacionado com o envelhecimento, este é um problema que se alastra globalmente e para todas as faixas etárias. A obesidade é um problema que resulta da pouca prática de exercício físico e da muita prática de atividades sedentárias como a televisão ou o computador nos tempos livres. Hoje em dia são cada vez mais as crianças, jovens, adultos e idosos que sofrem de obesidade ou de excesso de peso. Doenças cardiovasculares ou acidentes vasculares cerebrais são alguns dos problemas cujo risco de incidência é maior para pessoas obesas ou sedentárias.

O facto de o sedentarismo ser um problema crescente e para todas as faixas etárias justificou que o público alvo deste trabalho, inicialmente definido unicamente para os idosos, fosse alterado para todas as faixas etárias.

A prática de exercício físico é benéfica para a saúde física e psicológica ao mesmo tempo que potencializa o contacto com outros indivíduos, reduzindo problemas como o isolamento, solidão ou depressão. Praticar exercício físico diminui também a probabilidade de ocorrência de doenças cardiovasculares ou acidentes vasculares cerebrais. Face a todos os benefícios que a prática do exercício físico apresenta, é imperativo e fundamental incentivar de alguma forma as pessoas a praticar atividades físicas nos seu tempos livres.

1.1 - Envelhecimento populacional

O aumento da esperança de vida é um fenómeno que continua a crescer e a uma escala global. Segundo o estudo demográfico publicado pelo Instituto Nacional de Estatística (INE) (Instituto Nacional de Estatística, 2012), em Portugal a esperança média de vida à nascença é cada vez maior e mais que duplicou em menos de um século (figura 1). Se no ano de 1920 a esperança média de vida era de 35.8 anos para os homens e de 40 para as mulheres, no final do século XX estes valores são em média de 73 para os homens e de 76.7 para as mulheres. Segundo a mesma entidade, o aumento destes valores são mais evidentes na primeira metade do século XX, devido a um elevado decréscimo da mortalidade nos primeiros anos de vida. Em contrapartida, nas últimas décadas a razão para o aumento da esperança de vida deve-se essencialmente ao aumento da sobrevivência em idades avançadas, resultante da crescente evolução na medicina e tecnologia.

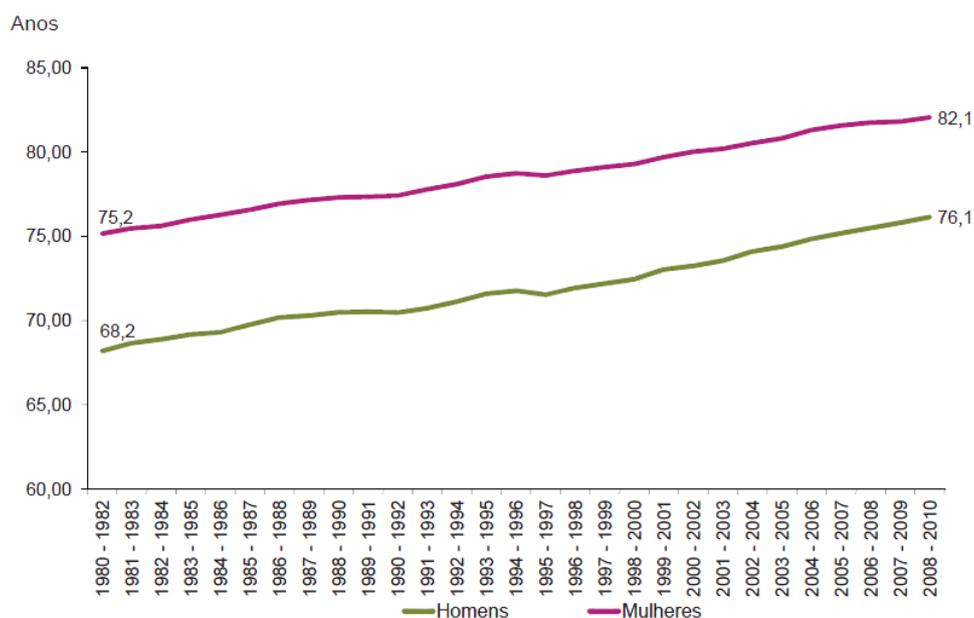


Figura 1 - Esperança média de vida à nascença por sexo em Portugal (reproduzido de Instituto Nacional de Estatística, 2012).

Embora este crescente aumento da esperança de vida possa ser visto como um aspeto positivo, o facto é que também contribui para o envelhecimento da população. O envelhecimento populacional, também denominado envelhecimento demográfico, é um facto que se verifica cada vez mais no mundo. Segundo o INE (Instituto Nacional de Estatística, 2012), até 31 de Dezembro de 2010 a tendência para o envelhecimento demográfico manteve-se em Portugal,

com a percentagem de população idosa a aumentar de 17.1% para 18.2%. Face a este aumento, o índice de envelhecimento sofreu também um aumento sendo agora de 120 idosos por cada 100 jovens (figura 2).

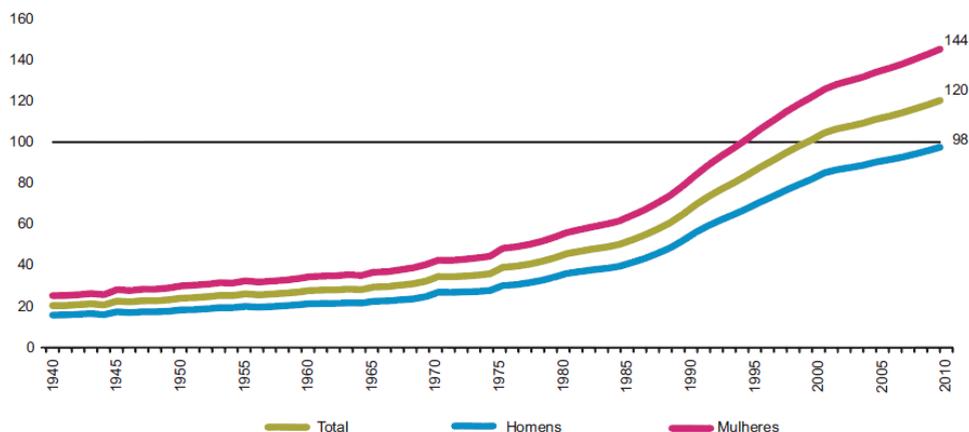


Figura 2 - Índice de Envelhecimento em Portugal (reproduzido de Instituto Nacional de Estatística, 2012).

Para além do aumento da esperança média de vida, a diminuição da taxa de natalidade (entre 2005 e 2010 passou de 10.4% para 9.5% nados vivos por mil habitantes) apresenta-se como uma das razões para o envelhecimento da população (figura 3).

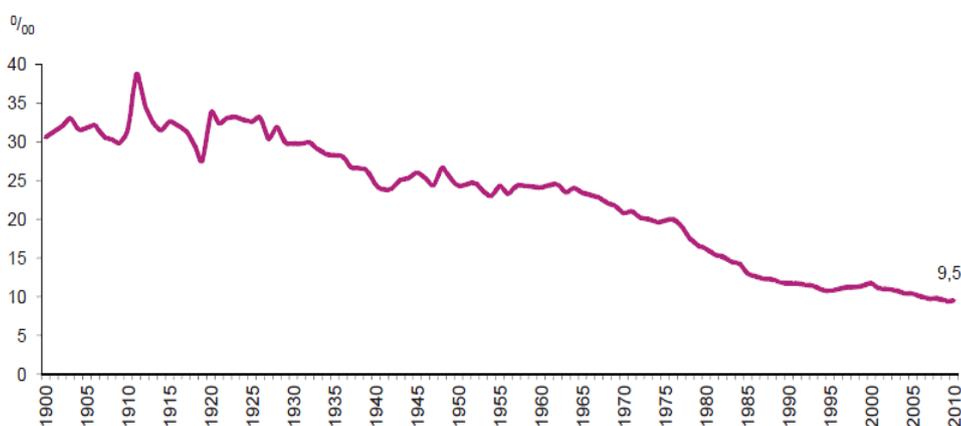


Figura 3 - Taxa bruta de natalidade (por mil habitantes) em Portugal (reproduzido de Instituto Nacional de Estatística, 2012).

O envelhecimento é um processo natural do ser humano, também denominado de senescência, responsável pela alteração progressiva de aspetos físicos e cognitivos. Do ponto de vista fisiológico, o ser humano sofre alterações nas funções orgânicas devido aos efeitos da idade avançada sobre o organismo, causando uma diminuição progressiva da sua reserva

funcional. Sintomas como a Osteoporose (diminuição substancial da massa óssea), Sarcopenia (perda degenerativa de massa e força nos músculos), Imunosenescência (declínio das funcionalidades imunológicas) são as alterações fisiológicas mais caracterizantes do envelhecimento. Relativamente às alterações cognitivas, a diminuição da capacidade de memória, maior dificuldade nas tarefas de raciocínio envolvendo material abstrato ou não familiar, são vistas como as alterações cognitivas mais caracterizantes do envelhecimento. A ordem ou amplitude com que as alterações ocorrem depende do estilo de vida da pessoa. Quando submetida a situações de stress físico, emocional ou outros, a amplitude com que estas alterações ocorrem numa pessoa tende a ser maior (Firmino, 2006).

1.2 - Solidão, Depressão e Sedentarismo no Envelhecimento

A definição acerca do que é ou significa a solidão não é unânime na literatura, existindo diversas abordagens e metodológicas que exaltam diferentes causas e manifestações da mesma. Todavia, existem três aspetos que são partilhados pelas definições presentes na literatura: a solidão é uma experiência subjetiva que não está diretamente relacionada com o isolamento; trata-se de uma experiência psicologicamente desagradável para o indivíduo; a solidão resulta de uma forma de relacionamento deficiente (Melo & Neto, 2003).

A solidão é muitas vezes vista como um traço característico do envelhecimento. Com o avançar da idade as pessoas tendem a diminuir a sua participação na comunidade originando sentimentos de solidão. Melo & Neto (2003) afirmam que a solidão pode ser resultado da rutura da atividade laboral e conseqüente declínio no padrão de vida e utilidade social. A entrada na reforma conduz a perdas de rendimentos, de prestígio, do sentido de competência e utilidade, e sobretudo de contactos sociais. Segundo Sánchez & Ulacia (2005), as pessoas idosas têm as mesmas necessidades afetivas interpessoais que as crianças, jovens e adultos, mas mais frequentemente insatisfeitas.

A depressão é atualmente considerada uma problema importante de saúde em todas as faixas etárias que conduz a sentimentos de tristeza e isolamento social (Martins, 2008). A depressão manifesta-se sobretudo por aspetos como: falta de motivação, perturbações na vontade, desinteresse, perda de apetite, somatização, dores físicas, irritabilidade, dificuldade de

concentração, problemas de sono, perda de gosto pela vida, fraqueza física ou dificuldade de convívio com amigos, colegas e familiares (Zimerman, 2000).

Segundo Martins (2008), é sobretudo nas idades avançadas que a depressão origina elevados índices de morbidade e mortalidade. As alterações e perdas progressivas a que um idoso está sujeito resultam muitas vezes em sentimentos de tristeza e desânimo que acabam por originar a depressão. A depressão é por sua vez prejudicial para a vida do idoso em termos intelectuais e sociais. Em termos intelectuais, a capacidade de aprendizagem diminui face à diminuição das capacidades e perturbações de memória. Na socialização, ocorre o afastamento dos grupos, a perda do estatuto, o abandono, o isolamento e perda do convívio social (Zimerman, 2000).

O sedentarismo é definido como a falta de atividade física, sendo que o conceito de atividade física não é sinónimo da prática de atividade desportiva. Pessoas sedentárias gastam poucas calorias diárias com atividades físicas. Atividades como a realização de tarefas domésticas, o caminhar para o trabalho ou a realização de tarefas profissionais que requeiram esforço físico, são exemplos de atividades que consomem calorias e que não estão inteiramente relacionadas com o desporto. Para além de estar relacionado com problemas como a obesidade, o sedentarismo influencia negativamente sobre o aparecimento de doenças como a hipertensão, doenças respiratórias crónicas e distúrbios cardíacos (Neto, s.d).

A redução das aptidões físicas, o declínio das capacidades funcionais, a diminuição da massa óssea e muscular, o aumento do peso e outras alterações características do envelhecimento agregadas com a solidão e depressão, leva os idosos a sofrerem de sensações de inutilidade e saúde frágil e conseqüente redução de exercitação física. As conseqüências do sedentarismo em idosos são mais severas face às alterações naturais no envelhecimento, sendo o sedentarismo um fator capaz de acelerar os problemas inerentes ao envelhecimento (Salvador, 2005).

1.3 - Sedentarismo, um problema para todas as Faixas Etárias

Aquando do estudo do problema do sedentarismo, deparou-se com uma realidade cada vez mais problemática. O sedentarismo é um problema que cada vez mais afeta indivíduos de todas as faixas etárias, desde a infância até à terceira idade e cujo impacto principal se traduz na obesidade. Face a esta realidade preocupante, foi realizado um estudo mais detalhado deste

problema que não apenas o constatado na terceira idade. Tal foi a severidade constatada com este estudo, que foi decidido alterar o público alvo deste trabalho. Se inicialmente estava definido exclusivamente para a terceira idade, como resultado deste estudo, o público alvo passou a abranger todas faixas etárias. Contempla-se em seguida as principais conclusões obtidas em relação ao sedentarismo.

Segundo um estudo realizado pela comunidade europeia (Special Eurobarometer - European Commission, 2004), Portugal é o país da Europa com maior taxa de sedentarismo. É estimado que cerca de 70% da população seja sedentária, sofra de reduzida aptidão física e tenha excesso de peso. Dos inquiridos em Portugal acerca da frequência de prática de atividades físicas, cerca de 66% afirmaram que nunca praticam (Tabela 1). Dados estatísticos definem a televisão e o computador como as atividades sedentárias que são mais frequentemente praticadas e de forma excessiva. Estima-se que só na Europa, cada pessoa assista aproximadamente quatro horas de televisão por dia. Em Portugal cada pessoa assiste aproximadamente três horas e meia por dia. O caso mais preocupante são os habitantes da Sérvia que assistem aproximadamente cinco horas e meia por dia (IP Network, 2011).

Tabela 1 - Percentagem dos Inquiridos que nunca praticam exercício físico por país (adaptado de Special Eurobarometer - European Commission, 2004).

País	Percentagem
 Portugal	66%
 Hungria	60%
 Itália	58%
 Grécia	57%
 Letónia	48%
 Lituânia	48%
 Chipre	47%
 Espanha	47%
 Polónia	46%

A obesidade na infância e adolescência têm vindo ao longo do tempo a ser alvo de debate por razões negativas. O excesso de peso na infância e adolescência alcançou proporções epidêmicas tanto em países desenvolvidos como em desenvolvimento (Anderson & Butcher,

2006) (figuras 4 e 5). Segundo a Comissão Europeia (Directorate-General for Health & Consumers - European Commission, 2010), Portugal é um dos países europeus com maior percentagem de obesidade infantil. Cerca de 31.5% das crianças entre os 7 e 9 anos são obesas ou sofrem de excesso de peso, sendo a percentagem de crianças do sexo feminino maior do que o sexo masculino.

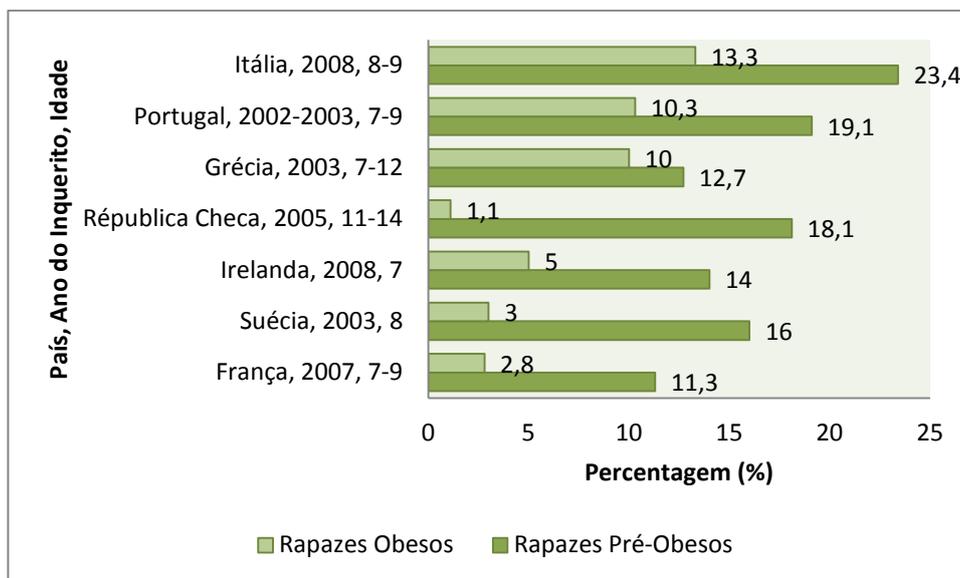


Figura 4 - Percentagem de Obesidade e Pré-Obesidade nos Rapazes (adaptado de Directorate-General for Health & Consumers - European Commission, 2010).

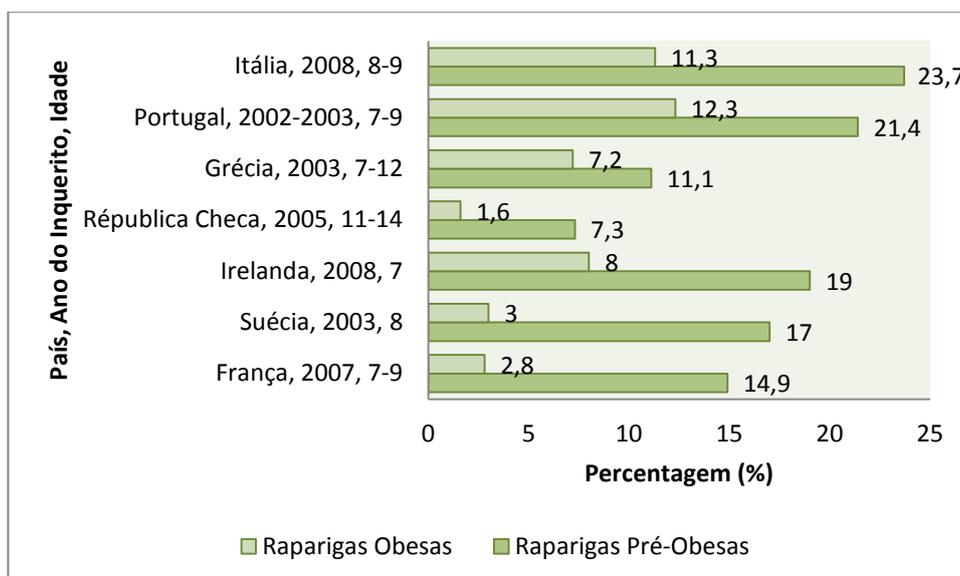


Figura 5 - Percentagem de Obesidade e Pré-Obesidade nas Raparigas (adaptado de Directorate-General for Health & Consumers - European Commission, 2010).

A obesidade na idade adulta é também uma realidade (figura 6). Utilizando a medição de Índice de Massa Corporal (mede a corpulência, a razão entre o peso em quilogramas e o quadrado da altura em metros, expresso por IMC), a Direcção-Geral de Saúde estima que cerca de 34% da população portuguesa adulta possuía pré-obesidade (IMC entre 25 e 29.9) e cerca de 12% possuía obesidade (IMC maior ou igual a 30). É ainda estimado que na população portuguesa com mais de 55 anos a prevalência dá pré-obesidade e obesidade seja 1.9 e 7.2 vezes maior, respetivamente (Direcção-Geral da Saúde, 2005).

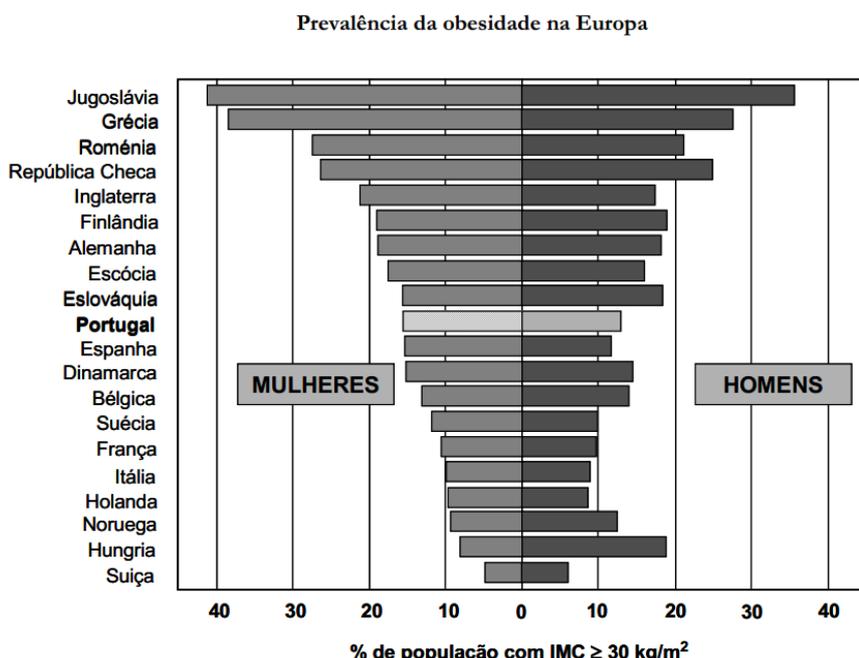


Figura 6 - Prevalência da obesidade na população adulta da Europa (retido de Direcção-Geral da Saúde, 2005).

A obesidade e excesso de peso podem dever-se a vários fatores. A obesidade pode ser transitável pelos genes, crianças cujos pais sejam obesos tendem a também o ser. O sexo da pessoa também influencia sobre a obesidade. Uma vez que os homens por natureza têm mais massa muscular e menos gordura corporal que as mulheres, estes tendem a queimar mais calorias. A idade também influencia sobre a obesidade, com o decorrer da idade a massa muscular diminui sendo desta forma necessário ingerir menor quantidade de calorias.

O sedentarismo é também uma das principais razões para a obesidade. A pouca prática de atividade física faz com que a quantidade de calorias ingeridas não seja gasta e com que

consequentemente se acumule sobre a forma de gordura (Portal da Saúde - Ministério da Saúde, 2005).

O sedentarismo é geralmente maior nos tempos livres das pessoas. Como já referido, a televisão e o computador são as atividades mais frequentadas nos tempos livres. Num estudo realizado pelo Inquérito Nacional de Saúde em Portugal (Plano Nacional de Saúde, 2004), cerca de 75% da população com mais de 15 anos define a leitura, televisão e outras atividades sedentárias como sendo as principais atividades que praticam nos tempos livres. Segundo um outro estudo acerca da utilização da internet em Portugal (Lisbon Internet and Networks International Research Programme, 2010), 62% dos internautas portugueses declarou a navegação na internet como a sua principal atividade de entretenimento. Um estudo realizado pela Eurostat (Eurostat, 2003), conclui que a televisão ocupa cerca de 40% dos tempos livres na grande maioria dos países europeus (figura 7).

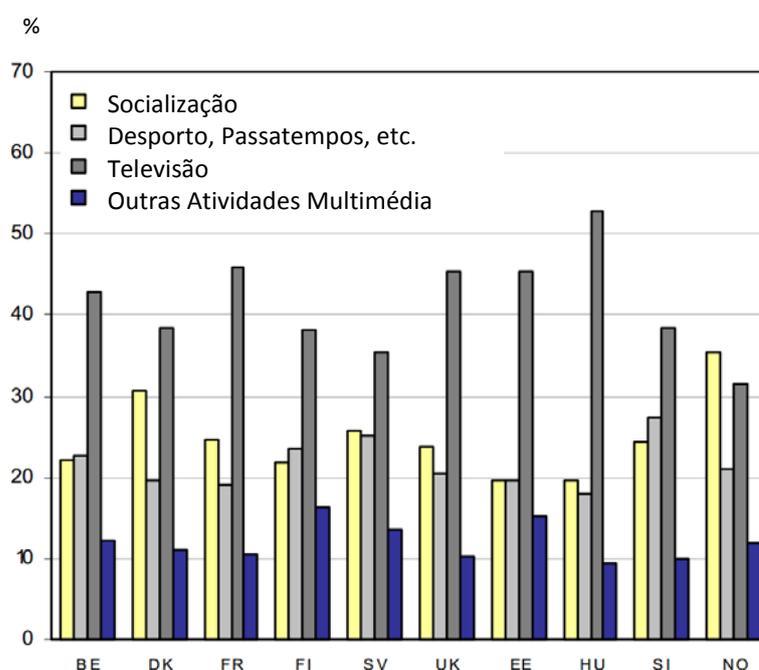


Figura 7 - Ocupação dos tempos livres por atividade em alguns países europeus (adaptado de Eurostat,2003).

São vários os estudos que correlacionam as atividades envolvendo ecrãs com o sedentarismo e conseqüente obesidade, bem como outros problemas. O número elevado de horas que são passados em atividades como a televisão faz com que a quantidade de calorias gastas seja menor, uma vez que não é praticada atividade física (Andersen et al.,1998). Um estudo comparativo entre crianças obesas e as não obesas, demonstrou que crianças obesas

tendem a assistir maior tempo de televisão e tendem a realizar menos exercício físico por semana. As crianças obesas dedicam maior tempo às atividades sedentárias e nomeadamente à televisão, o que faz com que sejam fisicamente menos ativas (Graf et al., 2005). Um outro estudo realizado em Portugal, demonstrou a correlação entre o tempo passado ao computador com a obesidade. Os seus autores concluíram que pessoas que utilizam o computador mais do que quatro horas por dia, têm maior tendência para ter excesso de peso ou serem obesas (Mota et al., 2006).

A prática em demasia de atividades sedentárias tem implicações preocupantes na saúde. O sedentarismo é muitas vezes associado ao problema da obesidade que por sua vez se traduz em doenças cardiovasculares, como as doenças isquémicas e os acidentes vasculares cerebrais. A falta de exercício físico reduz a circulação sanguínea normal. Uma vez que os músculos das pernas não contraem o suficiente, o tempo de retorno do sangue ao coração (retorno venoso) é maior, o que possibilita a coagulação do sangue no interior dos vasos sanguíneos (trombos, trombose). Os trombos são por sua vez responsáveis pelos distúrbios cardiovasculares mais graves. O sedentarismo é também responsável pelo desenvolvimento da arteriosclerose, situação em que a parede das artérias reduz a sua flexibilidade e elasticidade, resultante do seu endurecimento e espessamento. Algumas das consequências são as trombozes, os enfartes, os derrames cerebrais, entre outras (Nowak et al., 1999).

Um estudo publicado em 2011 na Escócia envolvendo 4512 pessoas com mais de 35 anos e entre os anos de 2003 a 2007, comprovou que as atividades envolvendo ecrãs influenciam sobre a ocorrência de problemas cardiovasculares e sobre o risco de mortalidade. Comparadas com as pessoas que praticam estas atividades menos de duas horas por dia, as pessoas que praticam mais de quatro horas vêm o risco de mortalidade aumentar 48% e as pessoas que praticam mais de duas horas vêm o risco de problemas cardiovasculares aumentar 125% (Stamatakis et al., 2011). Um outro estudo realizado em adolescentes demonstrou que a visualização excessiva de televisão pode agravar severamente o risco de doenças cardiovasculares nos adolescentes e independentemente dos seus pesos (Martinez-Gomez et al., 2010).

1.4 - Atividade Física

O sedentarismo é um problema que é possível de tratar. Como já mencionado, este resulta da falta de exercício físico e a sua solução passa por incentivar as pessoas a praticar atividades nas quais se exercitem fisicamente. Por exercício físico não se deverá entender só a prática de atividades desportivas. Dançar e caminhar são exemplos de atividades que contribuem para a prevenção do sedentarismo e que não estão inteiramente relacionadas com o desporto.

A exercitação física é também algo que é aplicável e benéfica em todas as faixas etárias. Um envelhecimento saudável passa por ter uma boa alimentação, ter um bom relacionamento, bom convívio social e principalmente por manter uma vida ativa. Ao manter-se ativo após a reforma, uma pessoa idosa mantém-se saudável em termos físicos, psicológicos e sociais. Os exercícios físicos desempenham um papel importante na prevenção, conservação e recuperação da capacidade funcional de idosos. O exercício físico não fará parar o processo de envelhecimento mas poderá retardar o surgimento de complicações (Guedes & Guedes, 1995).

Pessoas que cedo se tornam sedentárias tendem a o ser em toda a sua vida. Incentivar as crianças ou adultos a cedo praticar atividades físicas melhora a sua saúde, ao mesmo tempo que previne e reduz a probabilidade de ocorrência de doenças cardiovasculares no decorrer da sua vida.

São vários os benefícios da prática de exercício físico, podendo contribuir tanto em termos físicos, como psicológicos, sociais ou até mesmo económicos. Em termos físicos, desde logo este exerce um papel importante no controlo do excesso de peso e obesidade. A redução de massa gorda e manutenção ou aumento da massa muscular previne doenças cardiovasculares e arteriais. Ao aumentar a força muscular, aumentar a densidade óssea e ao melhorar o condicionamento cardiorrespiratório, estão a ser prevenidas doenças respiratórias e doenças como a osteoporose ou hipertensão. Doenças como cancro do cólon ou diabetes tipo dois, são exemplos de outras doenças que vêm o seu risco de incidência reduzido com a exercitação física.

No que toca a benefícios psicológicos, o exercício físico contribui para a melhoria do estado de humor, redução de sintomas de tensão, depressão, ansiedade ou raiva. A sensação de bem-estar e autoestima são também fatores psicológicos que são melhorados com a prática de exercício físico (Berger et al., 2006).

A prática de atividade física pode também potencializar o contacto com outros indivíduos e dessa forma reduzir problemas como o isolamento e solidão.

A economia pode também ser beneficiada com a prática de exercício físico por parte das pessoas. Com a diminuição dos riscos de incidência de problemas de saúde, o exercício físico está também a diminuir a utilização de serviços de saúde e de fármacos.

Incentivo às Atividades Físicas e Sociais nos Tempo Livres

Segundo o sociólogo francês Dumazedier (1974), “Lazer é um conjunto de ocupações às quais o indivíduo pode entregar-se de livre vontade, seja para repousar, seja para divertir-se, recrear-se e entreter-se, ou ainda, para desenvolver a sua informação ou formação desinteressada, a sua participação social voluntária ou a sua livre capacidade criadora após livrar-se ou desembaraçar-se das obrigações profissionais, familiares e sociais.”

Atualmente são inúmeras as atividades que podem ser classificadas de lazer, entre as quais estão o computador e a televisão que como já se comprovou, são das mais eleitas para ocupar os tempos livres. De facto, estas atividades são geralmente executadas de livre vontade pelo utilizador, permitem que este se divirta e permitem que desenvolva a sua informação. Porém como o computador e a televisão são as atividades que mais se relacionam com o sedentarismo e tendo em conta os problemas de saúde que acarretam e os benefícios que o exercício físico apresenta, torna-se necessário agir para que as pessoas pratiquem nos seus tempos livres e não só, atividades físicas e sociais em vez de atividades sedentárias e isoladas.

Um sistema capaz de recomendar tais atividades ao utilizador, pode ser um auxílio para o desejado. A implementação de um sistema capaz de detetar espaços de tempos livres e recomendar atividades que não sejam sedentárias é um dos objetivos deste trabalho. É necessário contudo definir um ponto de interação com o utilizador, a partir do qual este possa ser recomendado e notificado de tais atividades, e opinar acerca destas. Tendo em conta que é pretendido evitar a prática de atividades sedentárias, foi escolhido os *smartphones* como a plataforma de interação com o utilizador em detrimento do computador. A razão desta escolha deve-se ao facto de considerar-se que o computador pontualiza mais o sedentarismo do que o telemóvel. A colaborar também para esta decisão estão outros fatores como a crescente utilização de *smartphones* pelas pessoas e as suas capacidades em termos tecnológicos.

Estima-se que 80% da população mundial seja possuidora de um telemóvel. Dos cinco milhões de telemóveis estimados em todo o mundo, cerca de 22% deles são *smartphones*. Segundo os mesmos dados, 89% dos utilizadores de *smartphones* utilizam-no durante todo o dia. Em relação à faixa etária dos utilizadores, são as pessoas entre os 25 e os 34 que mais utilizam os *smartphones* (27%), existindo contudo cerca de 3.9% de utilizadores com mais de 65 anos de idade (Infographic, 2011). Em Portugal e até Dezembro de 2011, estimou-se que existiam 358 mil de utilizadores de *smartphones*, o que corresponde a uma percentagem de 4.2% do total de residentes com 10 ou mais anos de idade. Em relação ao mesmo estudo feito um ano antes, verificou-se uma duplicação deste valor, na altura 1.9% (Marktest, 2011).

Os *smartphones* distinguem-se dos outros telemóveis pelo facto de possuírem maiores capacidades ou funcionalidades. O facto de os *smartphones* possuírem maiores capacidades de comunicação de rede facilita a recolha e envio de informação. Outras das suas funcionalidades poderão ser um ponto de expansão para trabalhos futuros. O facto de existir um sensor de GPS (Sistema Global de Posicionamento) poderá ser uma mais valia para definição do local onde praticar uma determinada atividade física. Poderá ser interessante que com base na posição geográfica do utilizador, o sistema recomende atividades em locais próximos. O sensor de acelerómetro pode também ser útil para sistemas de deteção de quedas. Uma outra característica dos *smartphones* é a expansibilidade tanto em termos de software como de hardware. A adição de sensores capazes de medir o ritmo cardíaco, temperatura, pressão arterial e outras medidas importantes, poderá ser uma forma de transmitir e controlar a saúde dos utilizadores.

O processo de recomendação de atividades físicas e sociais requer informação acerca da pessoa, com o intuito de filtrar da enorme quantidade de atividades candidatas a serem recomendadas, aquelas pelas quais a pessoa demonstra maior interesse. Embora deva haver diversidade nas atividades físicas a praticar, a recomendação destas atividades deverá também ter em consideração as características das pessoas.

Todas as pessoas assemelham-se de alguma forma nas suas características físicas, porém cada uma possui um conjunto de características não-físicas que definem a sua individualidade pessoal e social, permitindo a sua distinção em relação a outros indivíduos. Esta é a uma das muitas definições existentes para o conceito de Personalidade ou Perfil. Estas características não-físicas de um indivíduo podem corresponder a aspetos como a sua forma de

raciocinar, os seus comportamentos, as suas emoções, os seus sintomas, as suas preferências, entre muitas outras possíveis.

No presente trabalho, um perfil representa as preferências do utilizador pela prática de atividades, bem como outras informações tais como a sua idade. Todavia, não está descartado em trabalhos futuros a adição de outras informações (como por exemplo medições acerca da saúde) que venham a ser úteis para o processo de recomendação.

Desta forma, a recomendação de atividades físicas terá também de ter em conta as preferências das pessoas em relação às atividades, sendo contudo apenas consideradas as preferências por atividades que não sejam totalmente sedentárias.

Como forma de manter uma diversidade nas atividades praticadas, é necessário que de vez em quando o sistema recomende atividades que não as representadas pelas preferências do utilizador. Pretende-se também desta forma descobrir novos pontos de interesse do utilizador em relação a novas atividades. A abertura ou aceitação que um utilizador tem por novas atividades é vista na psicologia como a sua abertura a novas experiências. Esta característica não é igual em todos as pessoas sendo a idade o principal influenciador. Com o aumento da idade, a abertura a novas experiências tende a diminuir. Desta forma, a idade da pessoa deverá ser tida em conta na frequência com que são recomendadas novas atividades.

Espera-se com o desenvolvimento deste sistema de recomendação, incentivar a prática de atividades que sejam do agrado das pessoas, mas que ao mesmo tempo sejam diversificadas e evitam os problemas de saúde inerentes ao sedentarismo.

1.5 - Perfis e suas Implicações

Os perfis são cada vez mais utilizados nos mais variados domínios. Conforme mencionado, um perfil poderá representar várias características dependendo do domínio onde é mantido ou utilizado. No domínio do Marketing, informações de compras são usadas na produção de novas campanhas. No domínio da Medicina, o conhecimento de sintomas passados é vital para salvar vidas. No domínio das Tecnologias, sistemas inteligentes tentam auxiliar os utilizadores nas suas tomadas de decisões com base nas suas preferências ou hábitos.

Os sistemas de recomendação são um exemplo de sistemas inteligentes que usam os perfis como forma a precisarem no que recomendam. Para além do sistema de recomendação de atividades para os tempos livres que é pretendido desenvolver, os sites de comércio

eletrónico são outro domínio que utilizam sistemas de recomendação. Nestes, o objetivo é com base no perfil de um utilizador recomendar produtos do seu interesse. Voltando a referir, um perfil representará as preferências do utilizador pela prática de atividades, que servirão de fonte de informação ao sistema de recomendação.

A representação e manutenção de um perfil em termos computacionais, não é uma tarefa simples face às implicações que um perfil apresenta e às capacidades do ser humano. As preferências de um individuo não se regem a simples sentenças que indiquem a sua preferência em relação ao valor de cada objeto. Considerando como domínio as atividades, a preferência que um utilizador tem por uma determinada atividade, depende do valor de outros atributos característicos da atividade como o local ou a hora. Assim sendo, um perfil representa também todas as dependências entre os seus elementos, fazendo com que a quantidade de informação que guarda seja volumosa, mesmo com poucos elementos.

Todos os seres humanos possuem a capacidade em evoluir e modificar o seu perfil, em mudar os seus comportamentos, as suas preferências, e outras características. Assim, da mesma forma que se torna necessário representar todos os elementos e suas dependências, é vital que um perfil seja dinâmico, que seja eficiente no processo de modificação dos valores de cada elemento e suas dependências, como forma a garantir que toda a sua informação é constantemente válida e atualizada com quem representa. Para conseguir tal proeza não basta existir uma base de dados que armazene simples sentenças, tornando-se necessário adotar técnicas inteligentes e eficientes de armazenar, representar e modificar um perfil.

Conseguir e garantir que um perfil se mantenha sempre atualizado implica detetar evoluções ou modificações no perfil de um individuo. Há que determinar quando um utilizador deixa ou passa a ter preferência por uma atividade e adaptar o perfil face a tal constatação. Porém, não bastará saber quando um utilizador muda a sua preferência por uma atividade. Para que um perfil seja preciso, é essencial conhecer que fatores influenciam as decisões de um utilizador, de forma a modificar as dependências representadas pelo perfil. O facto de um utilizador recusar uma atividade, implica deduzir as razões para tal. Estas razões poderão estar relacionadas com o valor que atributos característicos da atividade possuam. Por exemplo, um utilizador poderá ter recusado uma atividade devido à sua duração ser maior do que 60 minutos. Neste caso, o perfil do utilizador deverá refletir que este não tem interesse em atividades com duração superior a 60 minutos, para que em situações futuras a probabilidade de atividades com esta característica serem recomendadas, seja menor.

A deteção das razões que influenciam as decisões de um utilizador poderá ser um processo simples se forem questionadas ao utilizador (explicitamente), ou mais difícil se for detetado de forma transparente para este, sendo neste caso conseguido através de técnicas de extração de conhecimento (implicitamente) que lidam com históricos de ações ou decisões realizadas pelo utilizador (H.Witten & Frank, 2005).

A decisão sobre que técnica usar não é um processo fácil devido às vantagens e desvantagens que cada uma destas possui. Por exemplo, enquanto através de um simples questionário obtém-se de forma rápida a informação que se pretende, com a observação das ações de um utilizador, que necessitam de maior tempo de processamento, pode-se em algumas situações obter informações mais realistas que as obtidas nos questionários.

1.6 - Projetos Relacionados

Este trabalho é desenvolvido no âmbito do projeto *iGenda* (Costa et al., 2010) que tem como principal objetivo desenvolver um Assistente Pessoal de Memória. Tal como neste trabalho, o *iGenda* focaliza-se no problema do envelhecimento, e mais precisamente no problema que a memória apresenta ser para as pessoas idosas. Com o avanço da idade o ser humano perde a capacidade de recordar e armazenar novas informações. Em muitos casos, esta facto apresenta-se como um fator que influencia pela negativa a qualidade de vida.

Usando conhecimentos de domínios como *e-Health*, Redes Colaborativas e Ambientes Inteligentes, o *iGenda* apresenta-se como um sistema inteligente de agendamento que ajuda as pessoas a recordar informações relevantes e eventos ou atividades. É prestado auxílio no processo de agendamento de novos eventos e resolução de conflitos entre estes, ao mesmo tempo que as pessoas são avisadas acerca de eventos que devam ser realizados.

Outra das suas funcionalidades é ser capaz de receber informações de diversas fontes (médicos, familiares, entre muitas outras) e adaptar (resolver possíveis conflitos) a agenda da pessoa face a tais informações, para que a pessoa não tenha necessariamente de planear e agendar eventos ou tarefas manualmente. Por exemplo, um médico ao recomendar a um paciente que realize um eletrocardiograma num determinado espaço temporal, estará a implicar com que o sistema adapte a agenda da pessoa.

Para conseguir resolver todos os problemas que acarreta a manutenção de uma agenda e auxiliar as pessoas idosas de forma eficiente, o sistema divide-se em três módulos representados pelos retângulos azuis na figura 8.

O módulo *Agenda Manager* é o responsável por controlar a comunicação e segurança de todo o sistema. Controla a receção de novos eventos para agendar e a execução dos restantes módulos. O módulo *Conflicts Manager* tem como objetivo assegurar que não existem sobreposições entre eventos ou atividades. Para o conseguir, agenda e reorganiza os eventos que lhe são comunicados pelo *Agenda Manager*, de modo a garantir que estão em conformidade com os demais.



Figura 8 - Esquema dos módulos do iGenda (retido de Costa et al., 2005).

O módulo *Free Time Manager* é responsável por agendar atividades recreativas nos tempos livres existentes na agenda de uma pessoa. O seu objetivo é similar ao pretendido por este trabalho, ou seja, assegurar que os tempos livres de uma pessoa são ocupados com atividades físicas e sociais que contribuam para a sua saúde. Atualmente, este módulo utiliza um sistema de recomendação estático. Inicialmente, as preferências da pessoa por atividades de lazer são arquivadas na base de dados e com base nestas e algoritmos de aleatoriedade são determinadas as atividades a recomendar. É possível alterar estas preferências de vez em quando, porém este processo revela-se árduo e pouco eficiente. Este módulo carece ainda de um mecanismo de aprendizagem capaz de aprender e adaptar as preferências de uma pessoa, capaz de escutar a sua opinião (aceitação ou recusa) em relação às atividades recomendadas e adaptar-se, para que a recomendação se torne mais precisa com o decorrer do tempo.

O trabalho nesta tese desenvolvido pretende contribuir para este módulo. Os perfis dinâmicos, ao manterem-se atualizados e consistentes com as pessoas ao longo das suas vidas, permitirão melhorar e precisar as atividades que são recomendadas, assegurando que estas estão em conformidade com o perfil da pessoa.

1.7 - Tema e Objetivos

Este trabalho propõe uma solução possível para os problemas sociais e tecnológicos que até agora foram apresentados.

No contexto social, viu-se que o sedentarismo apresenta-se como um problema cada vez mais preocupante em todas as faixas etárias, originando problemas graves para a saúde como a obesidade, doenças cardiovasculares e muitas outras perigosas, e que resulta da pouca prática de exercício físico, verificando-se principalmente nos tempos livres.

Face a esta realidade inquietante, o objetivo social deste trabalho é incentivar as pessoas a praticar atividades físicas e sociais nos tempos livres.

Este incentivo será consumado através do desenvolvimento de um sistema de recomendação que com base nas preferências representadas pelo perfil da pessoa, deverá ser capaz de recomendar atividades envolvendo exercício físico e convívio social. Para que haja diversidade nas atividades praticadas e de descobrir novos interesses, o sistema recomendará outras atividades para além das preferidas pela pessoa. O desenvolvimento do sistema de recomendação é um dos objetivos tecnológicos deste trabalho e que serve de suporte ao objetivo social.

O outro objetivo tecnológico, considerado o principal deste trabalho, é estudar e implementar um mecanismo capaz de garantir a existência de perfis dinâmicos e de cumprir com todas as implicações subjacentes aos perfis. Pretende-se estudar e implementar um mecanismo de representar, modular e obter perfis, apto a garantir que a informação de um perfil é facilmente modulada e que está constantemente atualizada, paralela e consistente com quem representa. A ambição é assegurar que um perfil é capaz de acompanhar uma pessoa ao longo da sua vida e adaptar-se a eventuais mudanças ou evoluções que ocorram nas suas preferências.

Com vista a alcançar o objetivo anterior, será efetuado o levantamento do estado da arte de todas as temáticas relevantes para a conclusão deste trabalho, mais precisamente, formas atuais de representar e modular um perfil dinâmico, e formas atuais de recolher informação acerca do utilizador. Através do conhecimento adquirido no estado da arte, pretende-se que seja estudada e implementada uma solução eficiente para a modulação de perfis, capaz de ser aplicada em problemas de outros domínios, que não apenas a recomendação de atividades para os tempos livres.

O sistema de recomendação de atividades a desenvolver terá também como função avaliar a aptidão e robustez do mecanismo de modulação de perfis dinâmicos desenvolvido.

Com vista à interação dos utilizadores com o sistema, é objetivo deste trabalho implementar uma aplicação para *smartphones*.

1.8 - Metodologia de Investigação e Calendarização

O desenvolvimento deste trabalho segue uma metodologia investigação-ação, a qual pretende resolver problemas reais através da aplicação de métodos essencialmente práticos. Cohen et al. (2000) definem esta metodologia como um "(...) procedimento essencialmente *in loco*, com vista a lidar com um problema concreto localizado numa situação imediata. (...) o processo é constantemente controlado passo a passo (isto é numa situação ideal), durante períodos de tempo variáveis, através de diversos mecanismos (questionários, diários, entrevistas e estudos de casos, por exemplo), de modo que os resultados subsequentes possam ser traduzidos em modificações, ajustamentos, mudanças de direção, redefinições, de acordo com as necessidades, de modo a trazer vantagens duradouras ao próprio processo em curso."

Como se pode constatar por esta definição, o objetivo principal desta metodologia passa por aplicar mudanças nas práticas de resolução de um problema, de forma a alcançar melhorias nos resultados.

A investigação-ação consiste num conjunto de fases que são desenvolvidas de forma contínua e sequencial. Planificar (planificação), agir (ação), observar (avaliação) e refletir (teorização) é a sequência de fases que são desenvolvidas nesta metodologia e que se definem como um ciclo do processo de investigação-ação (figura 9). O processo em si poderá ser constituído por um ou mais ciclos (figura 10).

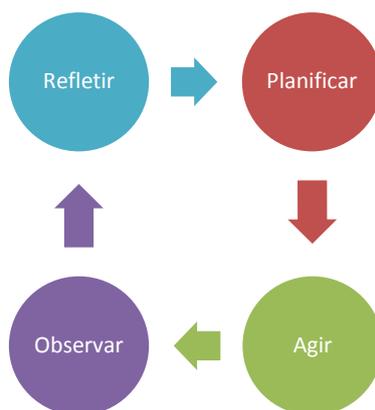


Figura 9 - Ciclo do processo de investigação-ação

Existem vários modelos propostos para o processo de investigação-ação. De entre os existentes foi adotado o modelo proposto por Kemmis & McTaggart (2005) que define as quatro fases do processo de uma investigação-ação como:

- Desenvolvimento do plano de ações tendo em conta uma informação crítica (identificação do problema) e desejada de alterar para melhor (proposta de uma solução);
- Determinação de um consenso com vista à execução do plano (agir);
- Observação e contextualização dos efeitos da ação;
- Reflexão sobre os resultados com vista à conclusão do processo, ou início de uma nova planificação (início de um novo ciclo).

No contexto deste trabalho estas fases são definidas da seguinte forma:

- Conceptualização e desenvolvimento de um mecanismo de modulação de perfis dinâmicos que vise propor uma solução nova e válida;
- Teste e implementação da solução obtida através da criação de um sistema protótipo (sistema de recomendação que dê uso aos perfis);
- Análise e avaliação dos resultados obtidos;

- Reflexão sobre os resultados com vista à conclusão do processo, ou início de uma nova planificação (início de um novo ciclo).

Tendo em consideração a crescente investigação que existe nos mecanismos e técnicas de representação e modulação de perfis dinâmicos, será necessária que paralelamente a cada fase seja efetuada uma atualização e revisão constante do estado da arte, tecnologias e avanços relacionados com o âmbito deste trabalho, como forma a garantir que o conhecimento aplicado mantêm-se atualizado.

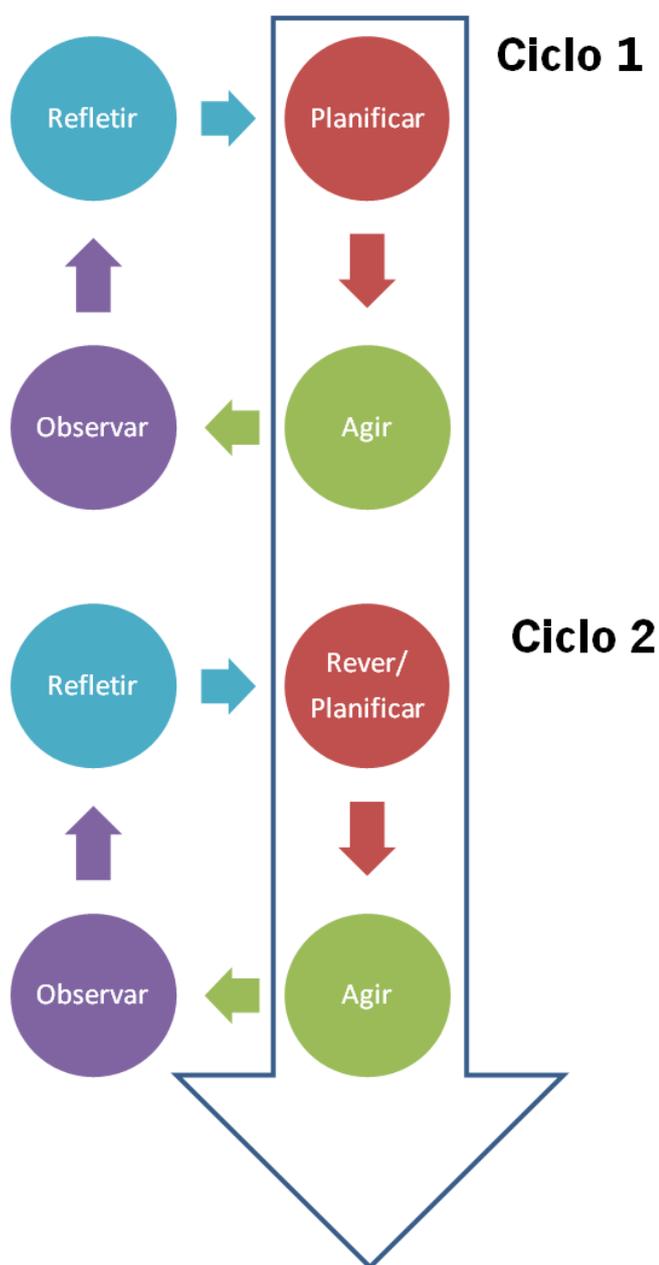


Figura 10 - Modelo proposto por Kemmis (adaptado de Kemmis & McTaggart, 2005).

Relativamente à calendarização, o desenvolvimento deste trabalho compreende a realização das cinco tarefas indicadas na tabela 2 e as suas respetivas durações. A tabela 3 apresenta em maior detalhe os períodos temporais em que cada uma das tarefas será realizada.

Tabela 2 - Calendarização do desenvolvimento do trabalho

Número da Tarefa	Tarefa	Duração	Resultados esperados
1	Levantamento do estado da arte sobre os Sistemas de Recomendação e a Personalização.	2 meses	Conjunto de publicações científicas relevantes para a investigação.
2	Elaboração de uma arquitetura de Recomendação e de Personalização.	2 meses	Definição das principais funcionalidades e características da arquitetura de Recomendação e de Personalização.
3	Realização de um estudo de caso.	1 mês	Recomendação de atividades para os tempos livres através da arquitetura desenvolvida.
4	Avaliação do modelo proposto	1 mês	Medidas de desempenho do modelo e descrição das suas potencialidades e limitações.
5	Redação da dissertação.	6 meses	Documento de Dissertação de Mestrado.

Tabela 3 - Diagrama de Gantt

Tarefa	Meses							
	10/2011	11/2011	12/2011	1/2012	2/2012	3/2012	4/2012	5/2012
1	█							
2		█						
3				█				
4					█			
5			█					

1.9 - Estruturação do Documento

Esta dissertação é constituída por seis capítulos. No primeiro capítulo são introduzidos os problemas para os quais este trabalho propõe uma solução. Começando por descreve o fenómeno do envelhecimento populacional e problemas inerentes ao envelhecimento, nomeadamente o sedentarismo. É feita uma análise mais específica deste problema que se apresenta preocupante para todas as faixas etárias que não apenas os idosos. É proposto o desenvolvimento de um sistema de recomendação de atividades que deverá ter em consideração as preferências de cada pessoa, representadas pelo seu perfil. A modulação de um perfil transporta um conjunto de implicações que também são narradas no primeiro capítulo. Uma vez encontrados os problemas, são descritos os objetivos e metodologia do trabalho.

O segundo e terceiro capítulos apresentam o conhecimento adquirido no levantamento do estado da arte de conteúdos relacionados com o trabalho desenvolvido. No segundo capítulo, "Sistemas Inteligentes", é descrito em que se baseia o conceito de inteligente e de que forma um sistema pode ser considerado inteligente. São descritos os sistemas de recomendação e os ambientes inteligentes como exemplos de sistemas inteligentes, que estão relacionados com o trabalho desenvolvido.

O terceiro capítulo introduz o principal conceito no qual este trabalho se desenvolve, a "Personalização" que consiste no processo de modular e obter perfis dos utilizadores. No primeiro subcapítulo são apresentadas técnicas atuais que lidam com a modulação de um perfil, sendo feita, para cada uma destas, uma descrição do seu funcionamento, apresentados exemplos de projetos onde é aplicada e por fim eventuais limitações ou cuidados a ter na sua utilização. O segundo subcapítulo apresenta técnicas de obter informação de um utilizador, sendo estas técnicas classificadas como explícitas e implícitas. É ainda descrito o que as distingue e realizada uma comparação entre as duas. Descrevesse também qual a técnica preferível de aplicar no domínio do trabalho.

No quarto e quinto capítulo é descrito o sistema desenvolvido neste trabalho que se designa *iLeisure*. O quarto capítulo descreve a arquitetura do sistema e a forma como os componentes do sistema comunicam. Neste capítulo são ainda apresentadas e descritas as funcionalidades da aplicação desenvolvida para dispositivos móveis, para a interação com os utilizadores.

No capítulo cinco é apresentado o componente principal do sistema que é o servidor. É descrito de que forma um perfil é modulado e que técnicas são utilizadas para tal fim. Também para o servidor, é exposto de que forma este aprende e adapta os perfis dos utilizadores, conforme as operações realizadas por estes na aplicação móvel. A descrição da metodologia a partir da qual o servidor recomenda atividades aos utilizadores encerra o quinto capítulo.

No sexto e último capítulo, relatam-se as principais conclusões obtidas com o desenrolar deste trabalho. Apresenta-se uma síntese do trabalho realizado e a sua aptidão em alcançar os objetivos propostos, assim como outros trabalhos que foram paralelamente desenvolvidos. Por último sugere-se possíveis investigações a realizar no futuro.

Capítulo 2: Sistemas Inteligentes

Rich & Knight (1991) definem a Inteligência Artificial como "A área da ciência da computação orientada ao entendimento, construção e validação de sistemas inteligentes, isto é, que exibem, de alguma forma, características associadas ao que chamamos inteligência".

Esta definição não é sucinta para definir de forma precisa o que é um sistema inteligente que continua hoje em dia sem ter uma definição unânime. Numa definição abstrata, um sistema inteligente poderá ser visto como um computador capaz de simular aspectos consideradas inteligentes por natureza como: aprender, adaptar-se a novas situações, abranger vários domínios, entre outras (Krishnakumar, 2003). Continua a ser mais certo definir se um sistema é inteligente do que o que é um sistema inteligente.

Muitas vezes, considera-se os braços robóticos das indústrias como sendo inteligentes, uma vez que praticam ações complexas e precisas. Todavia, estes poderão não o ser, se executarem sempre a mesma sequência de ações, para os quais foram programados. Se o robô não for capaz de detetar e adaptar-se a erros que surjam, a ocorrência de algum problema no procedimento destas ações, poderá provocar defeitos que o robô continuará a executar até que o problema seja manualmente detetado e corrigido.

Em contrapartida e tomando como exemplo, uma casa inteligente tem a capacidade de adequar-se a um indivíduo, necessitando para isso de aprender os seus hábitos (Olivier, 2008). Os sistemas de recomendação como aquele que é pretendido implementar neste trabalho, e o utilizado por muitos sites de comércio eletrónico, são também considerados inteligentes. Estes possuem a capacidade de aprender acerca das preferências de um utilizador e adaptar o que recomendam. Os ambientes inteligentes são um paradigma da Inteligência Artificial que é cada vez mais investigado e que acreditasse possível para um futuro breve. Neste crê-se que os ambientes serão constituídos por milhares de computadores embebidos, que utilizados como ferramentas proactivas assistirão as pessoas no seu dia-a-dia de forma a tornar as suas vidas mais confortáveis.

2.1 - Sistemas de Recomendação

Hoje em dia a quantidade de informação digital disponível cresce a um ritmo exponencial de tal forma que torna-se difícil encontrar aquela informação que desejamos de forma rápida e coerente. A *EMC* (2011), líder mundial de armazenamento, estimou que só no ano de 2011 foram produzidos um total de 1.8 zettabytes (1 099 511 627 776 gigabytes) de informação.

Os sistemas de recomendação (Zhou et al., 2011; Gediminas & Alexander, 2005) são sistemas inteligentes que surgiram em meados dos anos noventa com o intuito de assistir os utilizadores na busca de informação, filtrando da enorme quantidade aquela que é do seu interesse. Para além do domínio onde é pretendido implementar neste trabalho, presentemente existem inúmeros domínios onde estes sistemas já são aplicados, sendo os mais comuns o comércio eletrónico (Liden et al., 2003; Schafer et al., 1999) e os motores de busca (Baluja et al., 2008; Grcar et al., 2005).

O *Amazon* é um exemplo de onde é utilizado um sistema de recomendação. Tratando-se de uma loja online que vende uma grande diversidade de produtos, o sistema de recomendação filtra e recomenda aqueles que são do interesse do utilizador. O *Youtube* perante a crescente disponibilidade de vídeos sente também a necessidade de utilizar um sistema de recomendação como forma de auxiliar o utilizador na procura de vídeos do seu interesse.

Formalmente, o problema encarado pelos sistemas de recomendação define-se da seguinte forma. Considere-se P como o conjunto de todos os utilizadores de um sistema e I o conjunto de todos os itens que são possíveis de recomendar. O conjunto I poderá ser muito volumoso, podendo conter dezenas ou milhares de itens. Se considerar-se o *Amazon*, a quantidade de produtos que este comercializa é deverás volumoso, pelo que todos estes produtos são candidatos a serem recomendados a um utilizador.

Considere-se agora uma função u que determina a utilidade que um determinado item $i \in I$ tem para um determinado utilizador $p \in P$ ($u(i, p)$), o objetivo dos sistemas de recomendação é para cada utilizador $p \in P$ encontrar o conjunto de itens $I' \subset I$ que maximize a função de utilidade para o utilizador. Este problema é formalizado pela seguinte expressão:

$$\forall p \in P, I'_p = \arg \max_{i \in I} (u(i, p))$$

Nos sistemas de recomendação a utilidade de um item é geralmente representado por uma medida de avaliação (*rating*) que indica o quanto um utilizador tem interesse no item.

O problema encarado pelos sistemas de recomendação é o valor de utilidade u não estar definido em todo o universo $P \times I$, sendo apenas definido em subpartes deste. Caberá desta forma ao sistema de recomendação inferir sobre o valor de utilidade u para os demais elementos do seu universo.

Exemplificado, considere-se um site de comércio de automóveis que implementa um sistema de recomendação. O objetivo é encontrar e recomendar aos utilizadores o conjunto de automóveis que sejam de maior interesse para estes, que tenham o maior valor de utilidade. Neste caso o universo $P \times I$ poderá ser representado pela tabela 4, onde se considera que a utilidade é a avaliação realizada pelo utilizador com recurso ao método das cinco estrelas e o símbolo ∞ representa um valor de utilidade desconhecido.

Tabela 4 - Exemplo do universo de um sistema de recomendação

	BMW M3	Citroen C3	Opel Corsa	Ferrari 458
João	4	∞	3	∞
Maria	2	4	5	∞
André	2	4	∞	2

O objetivo do sistema de recomendação é estimar o valor que cada um dos símbolos ∞ deverá ter. Uma vez estimados todos os valores de utilidade desconhecidos, o sistema poderá recomendar aqueles automóveis cujos valores de utilidade sejam os maiores de entre os estimados. Se o sistema estimar que valor de utilidade têm os automóveis "Citroen C3" e "Ferrari 458" para o utilizador "João", este poderá determinar se estes automóveis deverão ser recomendados ao utilizador.

A razão pela qual um sistema de recomendação é considerado inteligente deve-se à sua capacidade em aprender acerca dos interesses do utilizador e dessa forma adaptar o que recomenda. Tendo conhecimento dos interesses dos utilizadores, o sistema usa essa informação para estimar o valor de utilidade para novos itens que não estejam definidos como do interesse do utilizador. Os novos itens como alto valor de utilidade serão os que constituirão o conjunto I' que será recomendado ao utilizador.

A forma como um sistema de recomendação estima o valor de utilidade, determina que um item é do interesse do utilizador, depende do algoritmo que é implementado, existindo vários possíveis. Estes são geralmente classificados em dois grupos: recomendação baseada no conteúdo e recomendação colaborativa. Adicionalmente, são também propostas abordagens híbridas que tentam agregar as vantagens de cada técnica.

Cada um destes grupos será em seguida descrito em maior detalhe. Será exposto o seu funcionamento e apresentadas as suas desvantagens.

Um ponto comum a estes dois grupos é a necessidade em representar o perfil de cada utilizador, a necessidade em representar as preferências ou interesses do utilizador por certos itens. Uma vez que o perfil dos utilizadores é usado como fonte de informação para a determinação de que itens deverão ser recomendados, é fundamental que seja constantemente atualizado e consistente. Pode-se desta forma averiguar que o trabalho desenvolvido é também vantajoso para os sistemas de recomendação.

2.1.2 - Recomendação baseada no conteúdo

Na recomendação baseada no conteúdo são recomendados ao utilizador os itens mais similares àqueles que são preferidos por este. Neste tipo de recomendação o valor de utilidade de um determinado item i para um utilizador p é estimado em função da similaridade que o item i terá com outros considerados do interesse do utilizador. A definição sobre que itens são do interesse do utilizador, dependerá do grau de precisão estipulado pelo sistema. Por exemplo, se considerado o exemplo da tabela 4, poderão ser definidos como do interesse do utilizador, os itens que tenham uma utilidade maior ou igual a 4 valores.

Uma vez que os valores de similaridade são geralmente expressos numa escala percentual, os sistemas usualmente adotam um valor de utilidade numa escala percentual ou então utilizam uma função que normalize o valor de similaridade.

Perante um novo conjunto de itens candidatos a serem recomendados, são feitas comparações entre cada um dos itens e os itens preferidos pelo utilizador como forma a determinar se deverá ou não ser recomendado. O resultado esperado é de um conjunto de itens candidatos, selecionar os mais prováveis de serem do agrado do utilizador, os mais similares com os preferidos pelo utilizador.

Para que seja possível comparar dois itens é necessário que estes sejam descritos por um conjunto de atributos. A comparação entre dois itens é desta forma conseguida através do cálculo da similaridade entre os atributos que caracterizam cada um dos itens. Referenciando novamente o exemplo da tabela 4, cada um dos automóveis é definido por um conjunto de atributos como o tipo de combustível, a potência, entre outros. Neste caso, o valor de similaridade entre dois automóveis é determinado em função da similaridade existente entre cada um dos atributos dos automóveis.

Para o sistema determinar por exemplo, o valor de utilidade do automóvel "Opel Corsa" para o utilizador "André", é calculado a sua similaridade com o automóvel "Citroen C3" que é visto como do interesse do utilizador. Se os dois automóveis forem similares nas suas características, então é provável que o automóvel "Opel Corsa" seja do interesse do utilizador e deva ser recomendado.

Este tipo de recomendação apresenta contudo algumas limitações quanto à sua cobertura e problemas relacionados com a subespecialização.

A aplicação deste algoritmo de recomendação tem como pressuposto que todos os itens sejam definidos por um conjunto comum de atributos que os caracterizem, de forma a serem comparados com os demais em termos de similaridade. Isto significa que para cada novo item introduzido seja necessário especificar todos os valores para os atributos definidos.

Para além de possuir esta necessidade, este algoritmo releva-se pouco eficiente em domínios heterogéneos. Em sites onde existe uma grande diversidade de produtos, não é possível definir um conjunto uniforme de atributos que caracterizem os produtos. Por exemplo, produtos como livros e telemóveis diferem na forma como são caracterizados, não sendo possível uma comparação direta entre estes. Nestes casos, a solução poderá passar por categorizar os produtos e aplicar um valor de similaridade entre as categorias ou então garantir que apenas são realizadas comparações entre produtos da mesma categoria.

Porém, este algoritmo de recomendação revela-se eficiente em domínios onde seja necessário recomendar itens de um único domínio. Em sites de vendas de automóveis, é possível definir um conjunto pré-definido de atributos que caracterizem cada um dos automóveis e dessa forma realizar comparações entre estes.

Um outro problema que este algoritmo acarreta é o facto de poder existir uma subespecialização por parte do sistema em relação às preferências de um utilizador no que recomenda. Isto deve-se ao facto de só serem sugeridos itens que sejam similares àqueles já

preferidos pelo utilizador, não existindo uma abertura a novos pontos de interesse. Itens que não sejam de alguma forma declarados como do interesse do utilizador não serão recomendados, bem como não serão os itens similares a estes.

Este problema é também verificado aquando do registo de um novo utilizador, para os quais ainda não existe registo dos seus interesses. Assim, não será possível determinar que itens recomendar, uma vez que não são reconhecidas preferências ou interesses do utilizador. Torna-se desta forma obrigatório adotar mecanismos de aprendizagem, a partir dos quais serão retidos os interesses dos utilizadores.

Adicionalmente, poderão ser introduzidas medidas de aleatoriedade de forma a que de vez em quando sejam recomendados itens não similares com os preferidos pelo utilizador. O objetivo é que o utilizador avalie estes itens e que a partir desta avaliação o sistema seja capaz de detetar novos interesses do utilizador, dando origem a uma recomendação.

2.1.3 - Recomendação colaborativa

Ao contrário da recomendação baseado no conteúdo, na recomendação colaborativa a decisão sobre que itens recomendar é conseguida com base na similaridade existente entre as preferências dos utilizadores e não entre os itens. Neste tipo de recomendação o valor de utilidade de um determinado item i para um utilizador p é estimado em função da utilidade que o item i tem para outros utilizadores com preferências similares.

Por outras palavras, a determinação sobre se um item deverá ser recomendado a um utilizador é estimada com base na avaliação realizada ao item por outros utilizadores com preferências similares. Considerando como exemplo prático, o caso em que é necessário recomendar automóveis a um utilizador, para a determinação dos automóveis a recomendar, o sistema começa por procurar pelos utilizadores com preferências mais similares. Uma vez encontrados tais utilizadores, só os automóveis que sejam melhor avaliadas por estes, serão recomendados.

Conforme pode ser visto, não é calculada qualquer similaridade entre os itens a recomendar mas sim entre as preferências dos utilizadores. Desta forma este algoritmo apresenta-se como vantajoso em relação ao anterior uma vez que não têm em consideração quaisquer atributos característicos dos itens, não sendo assim exigido a sua existência. Por não ter tal exigência, este algoritmo não apresenta problemas em comparar itens de diferentes

domínios, podendo desta forma ser considerado como aplicável tanto em domínios homogéneos como heterogéneos.

Tendo como base o exemplo da tabela 4, para determinar se o automóvel "Opel Corsa" deverá ser recomendado ao utilizador "André", o sistema poderá usar os interesses do utilizador "Maria" que se apresentam como similares. Uma vez que o utilizador "Maria" avalia positivamente o automóvel "Opel Corsa" e tem preferências idênticas com o utilizador "André", então é muito provável que este automóvel seja do interesse do utilizador "André", pelo o que deverá ser recomendado.

Este conceito é muito utilizado pelos sistemas de recomendação, segundo a expressão "utilizadores que compraram o produto A, compraram também os produtos..." que é muito comum aos sites de comércio eletrónico. Nestes, o sistema procura por utilizadores que tenham interesses similares pelo produto A e recomenda produtos deste grupo, pelos quais seja revelado interesse ou avaliados positivamente.

Para além de apresentar o mesmo problema do registo de um novo utilizador, a principal desvantagem da recomendação colaborativa é o facto de existir uma certa dependência em relação a outros utilizadores. Para que a recomendação de um item seja precisa, é necessário que exista um número substancial de outros utilizadores que tenham avaliado positivamente o item. Se existirem utilizadores cujas preferências sejam raras, comparativamente a outros, existirão poucos utilizadores similares, pelo que os itens recomendados serão poucos ou imprecisos. Para comutar este contratempo são adicionalmente introduzidos outros atributos no cálculo da similaridade entre utilizadores para além das suas preferências, como a idade, o sexo, a nacionalidade, entre outros.

A adição de novos itens é também um outro problema que carece atenção na recomendação colaborativa. Um novo item não terá no início nenhuma avaliação e dessa forma será pouco provável de ser recomendado. Neste casos, poderá ser adotada uma solução híbrida que primeiro procure por itens similares ao novo, que possuam avaliações, e que em seguida procure por utilizadores que avaliaram tais itens. Os utilizadores encontrados serão os mais prováveis de ter interesse pelo novo item. A recomendação a estes, do novo item, é uma forma de obter um conjunto inicial de avaliações.

Outro aspeto a considerar na adoção da recomendação colaborativa, são os requisitos computacionais e temporais necessários para a comparação entre utilizadores. Dependendo de

como o perfil de um utilizador é representado e mantido, a sua comparação com os demais poderá ou não ser um processo viável e eficiente.

2.2 - Ambientes Inteligentes

Os ambientes inteligentes, *Ambient Intelligence (AmI)* no mundo das tecnologias, apresentam-se como uma área de estudo muito promissora no domínio da Inteligência Artificial e que é cada vez mais investigada. Presentemente, a definição do que são os ambientes inteligentes não é única, existindo várias na literatura das quais destacou-se as seguintes (traduzidas literalmente):

"Um ambiente inteligente é uma rede de interfaces inteligentes e translúcidas que reconhecem a nossa presença e modelam o ambiente às nossas necessidades."

John Horvath, *Telepolis* (Horvath, 2002)

"Um ambiente inteligente refere-se a um novo paradigma entusiasmante nas tecnologias de informação, onde as pessoas são habilitadas com um ambiente digital que reconhece as suas presenças e contextos, que sente, adapta-se e responde às suas necessidades, hábitos, gestos e emoções."

De *Ambience Project* (Hitech Projects, 2004)

Comum a estas definições é o facto de que os ambientes inteligentes são algo benéfico para a população e que portanto são considerados como desejados.

Os ambientes inteligentes têm como base as tecnologias de computação ubíqua, comunicação ubíqua e interfaces inteligentes. O termo ubíquo define-se como "estar presente em todo o lado", "ser onnipotente". Como se pode deduzir, para que existam ambientes inteligentes é necessário que existam microprocessadores em todos os lados, capazes de recolher e processar as informações (computação ubíqua). Da mesma forma, é necessário

garantir que todos estes microprocessadores comunicam entre si e com o utilizador (comunicação ubíqua). As interfaces inteligentes correspondem ao meio, a partir do qual, as pessoas poderão controlar e interagir com o ambiente de uma forma natural (com gestos, emoções, vozes, entre outras) e personalizada (com base nas preferências e contextos) (Bauknecht, 2002).

O ciclo básico do funcionamento de um ambiente inteligente compreende essencialmente três etapas. Na primeira, o sistema deverá interpretar o estado do ambiente através da recolha implícita de informação acerca deste. Esta informação é conseguida através de sensores distribuídos pelo ambiente que constantemente monitorizam-no. Usando toda esta informação, o sistema tomará decisões sobre que formas agir com vista à adaptação do seu ambiente. É nesta etapa que a existência de perfis dos utilizadores revela-se fundamental para a determinação sobre que forma adaptar o ambiente. O resultado da tomada de decisão é um conjunto de ações que serão comunicados aos atuadores também distribuídos pelo ambiente e que resultaram na sua adaptação. Para além de servir para a tomada de decisões, a informação recolhida é também fundamental para a aprendizagem do sistema sobre os perfis dos seus utilizadores.

Atualmente, as casas inteligentes são uma subárea dos ambientes inteligentes que é muito investigada e que se apresenta como um passo importante para a obtenção dos ambientes inteligentes. Nestes o espaço é reduzido à planta de uma casa e a recolha de informação é conseguida através de vários sensores dispersos pela habitação. Poderão ser usados diversos sensores como por exemplo, sensores de pressão para detetar a abertura de portas, sensores de som ou de luminosidade para detetar a presença de utilizadores em divisões da casa, sensores de temperatura, entre outros. Os atuadores são o conjunto de eletrodomésticos existentes na casa. A figura 11 ilustra um exemplo de uma casa inteligente.

Existem vários casos de uso possíveis de se verificar numa casa inteligente. A poupança da eletricidade é uma delas. Uma casa inteligente tem a capacidade de decidir sobre que divisões da casa deverão estar iluminadas ou aquecidas em certos tempos e dessa forma desativar as demais. Uma casa inteligente é também capaz de executar medidas prévias. Se detetar um padrão nas ações do utilizador, a casa poderá prever que condições deverão existir no seu ambiente. Se for detetado que um utilizador a seguir ao jantar assiste televisão na sala de estar, a casa poderá atempadamente adequar a temperatura ambiente da sala de estar.

O projeto desenvolvido por Oliver (2008) é um bom exemplo de como tornar uma casa inteligente e adaptável aos seus utentes.

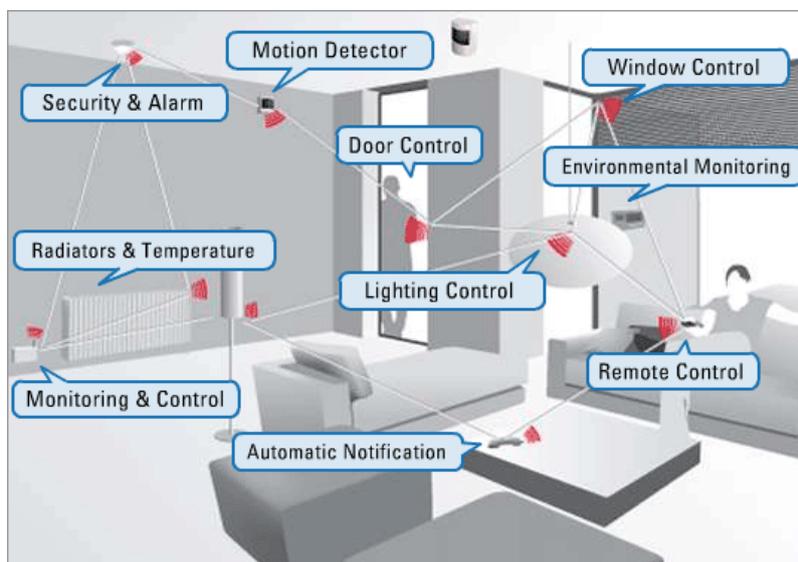


Figura 11 - Exemplo de uma casa inteligente (retido de Gotomy, 2008)

A razão pela qual os ambientes inteligentes são aqui ostentados prende-se com a indispensabilidade destes pela obtenção e modulação de perfis. Para que os ambientes sejam cientes e adaptáveis aos seus utilizadores é imperativo extrair e representar informação acerca destes. Identificar as preferências, necessidades ou comportamentos dos utilizadores permite que os ambientes sejam mais cientes e precisos no modo como se adaptam. Acredita-se que o trabalho desenvolvido pode servir de base para as evoluções futuras no desenvolvimento dos ambientes inteligentes e mais precisamente na personalização.

Tendo em conta a grande quantidade de informação que é recolhida pelos sensores, são necessárias técnicas próprias que consigam filtrar dados para obter conhecimento revelante (H.Witten & Frank, 2005). Por conhecimento relevante compreende-se por exemplo, a deteção de padrões de ações realizadas pelos utentes de uma casa. Os ambientes inteligentes requerem que a informação recolhida seja rapidamente transformada num conjunto de ações sobre os seus atuadores com vista à sua adaptação. Esta é razão principal pela qual estas técnicas complexas de extração de conhecimento são ainda alvo de muita investigação.

Os ambientes inteligentes podem também introduzir alguns problemas. Kurt Bauknecht (2002) corrobora o porquê de "controlar" os ambientes inteligentes com as questões que estes acarretam, relacionadas com a privacidade e segurança da informação que armazenam acerca

das pessoas. Conhecendo os cenários atuais e recentes, onde foram roubadas informações de entidades rigidamente protegidas como o FBI, é fundamental assegurar que as informações privadas das pessoas são mantidas seguras. Uma vez que os ambientes inteligentes são compostos por atuadores que são implicitamente ativados, é também fundamental controlar rigidamente o controlo de terceiros sobre estes.

Os conflitos entre o utilizador e os ambientes inteligentes é também um cenário possível de ocorrer. Poderão surgir problemas entre o controlo que o utilizador tem sobre o ambiente e o controlo que o ambiente tem face ao conhecimento acerca do utilizador. Este problema é bem demonstrado na figura 12 (FIDIS, 2005). É por esta razão, essencial definir prioridades sobre o controlo existente nos ambientes.

São estes e outros problemas inerentes aos ambientes inteligentes e tecnologias requisitadas, que os tornam difíceis de alcançar, mas que cada vez mais cativam o interesse de cientistas nesta área. Com o avançar da tecnologia e com os trabalhos similares já desenvolvidos, como as casas inteligentes, o caminho a percorrer para alcançar os desejados ambientes inteligentes é cada vez menor.

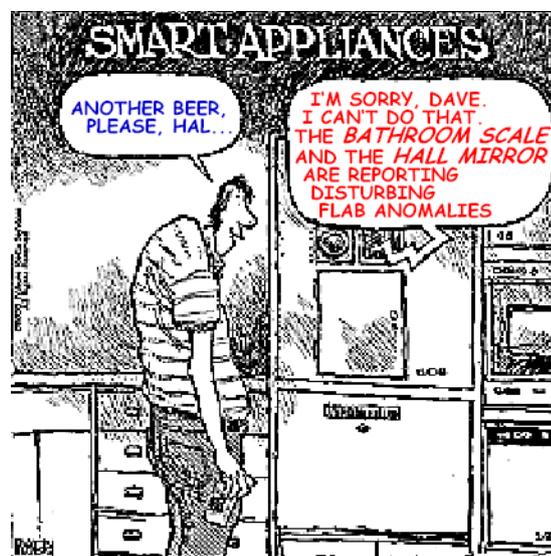


Figura 12 - "Máquinas Inteligentes": Um conflito entre ambientes inteligentes e o utilizador (retido de FIDIS, 2005).

2.3 - Conclusões

Neste capítulo apresentou-se os sistemas de recomendação, os ambientes inteligentes, e respetivos exemplos demonstrativos.

Para os sistemas de recomendação foi narrado em que consistem e apresentados exemplos atuais onde se apliquem. Como foi exposto, os sistemas de recomendação são classificados em dois grupos, em função do método de recomendação que utilizam. Sistemas de recomendação baseados no conteúdo apresentam-se como uma técnica mais privada e singular com uma pessoa, usando unicamente o seu perfil para determinar os itens a recomendar. Pelo contrário, sistemas de recomendação colaborativa, como o próprio nome induz, usam outros utilizadores na determinação dos itens a recomendar.

Em termos comparativos, cada uma destas técnicas apresenta as suas vantagens e desvantagens. De certa forma, considera-se que a recomendação colaborativa é mais apropriada para domínios heterogéneos, onde exista diversidade de itens, enquanto a recomendação baseada no conteúdo é mais adequada para domínios homogéneos. Existem contudo abordagens híbridas que agrupam as vantagens de cada uma destas formas de recomendar.

No que toca ao sistema de recomendação a desenvolver, considera-se a recomendação baseada no conteúdo como a mais adequada, considerando que as atividades são definidas por um conjunto pré-definido de atributos e como tal são consideradas homogéneas. Não é contudo descartado no futuro a adoção de abordagens híbridas, caso se apresentem como uma solução viável em relação ao domínio deste trabalho.

Os ambientes inteligentes foram apresentados neste capítulo como uma área emergente na Inteligência Artificial e onde se crê que o trabalho desenvolvido venha a ser proveitoso. Foi descrito em que consistem, apresentados alguns exemplos atuais onde são aplicados e ainda detetados alguns problemas que carecem de investigação.

Capítulo 3: Personalização

Cada vez mais os utilizadores de sistemas e serviços prezam as capacidades destes em adaptar-se às suas preferências, aos seus interesses, aos seus hábitos, por outras palavras, ao seu perfil. Por sua vez, os sistemas e serviços vêm a personalização com bons olhos, uma vez que é uma forma de agradar os utilizadores, cativar mais utilizadores e aumentar os seus lucros. Os sistemas de recomendação usados em muitos sites, como os de comércio eletrónico, adaptam os produtos que recomendam para que estes sejam do máximo interesse do utilizador e para que este os venham a comprar. As casas inteligentes adaptam as suas condições como forma a agradar as preferências e hábitos dos seus utentes. Os tão apetecidos ambientes inteligentes tornarão possível que as pessoas se sintam confortáveis em todos os locais, uma vez que estes adequar-se-ão às mesmas.

Garantir que os sistemas ou serviços são sempre adequados às pessoas requer que estes aprendam, registem e mantenham atualizado o perfil destas. Porém, como já foi desenvolvido no capítulo 1.5, o ser humano por natureza possui a capacidade de evoluir e alterar o seu perfil. Esta característica faz com que o tratamento de perfis em termos computacionais não se apresente simples. Primeiro porque existe a necessidade de acompanhar uma pessoa e alterar a informação registada pelo perfil, para que esteja constantemente atualizada. Conseguir tal proeza exige que a forma com que um perfil é representado e modulado seja a mais eficiente e inteligente. Segundo porque é necessário observar as ações da pessoa e recolher informações acerca de evoluções que ocorram. O grande objetivo dos serviços e sistemas é consegui-lo de uma forma não invasiva e obstrutiva para as pessoas, de uma forma implícita.

São precisamente estes dois problemas que são crescentemente investigados na comunidade científica e tecnológica e que são agrupados sobre o conceito de *Profiling* ou *User Profiling* (Gauch et al., 2007). O grande objetivo é aquele em que este trabalho orbita é precisamente encontrar métodos eficientes e inteligentes de conseguir modular e obter um perfil, de forma a o manter consistente e atualizado com as pessoas que representam, em suma, garantir a existência de perfis dinâmicos. Posto isto, o principal objetivo desta capítulo é precisamente fazer uma análise destes dois problemas com os quais o *Profiling* se depara e apurar soluções que existam atualmente.

3.1 - Modulação de um Perfil

Numa versão simplificada um perfil poderá ser representado como um conjunto estático de elementos que descrevem as preferências de um indivíduo. Um exemplo poderá ser o que acontece quando as pessoas registam-se num site, em que são solicitadas para indicar os seus pontos de interesse. Se o ser humano não evoluísse, então os perfis estariam constantemente consistentes e dessa forma a informação guardada pelo site também estaria. Contudo, essa não é uma realidade e torna-se necessário acompanhar e atualizar a informação de um perfil de acordo com a evolução da pessoa. São necessários perfis dinâmicos.

Uma outra característica do ser humano é viver num mundo de dependências, onde de certa forma a realização de ações ou os seus interesses estão condicionados por outras situações ou condições. Assim, para além de guardar simples variáveis, um perfil guarda também todas as dependências entre estas, o que torna a quantidade de informação a tratar grande. Desta forma, a tarefa de alterar a informação de um perfil não se apresenta como uma simples alteração da informação de uma base de dados, tornando-se necessário adotar técnicas inteligentes e eficientes de armazenar, representar e modificar um perfil, ou seja, de o modular.

Atualmente existem diversas propostas para a modulação de perfis, sendo as áreas mais aplicadas a Matemática e a Inteligência Artificial. Técnicas como Redes Neurais (Macdonald & Silver) ou Estruturas Hierárquicas (Syed & Andritsos, 2007) são algumas das soluções atualmente propostas. Por existirem inúmeras propostas não seria conveniente analisá-las na sua totalidade, sendo assim analisadas em maior detalhe aquelas que melhor se relacionam com o trabalho a desenvolver.

3.1.1 - Redes Bayesianas

Segundo McCann et al. (2006), uma rede bayesiana (*Bayesian Network* - BN) consiste num grafo acíclico direcional de nós e arcos que conceptualiza um domínio. Cada nó é constituído por um conjunto finito de estados diferentes e mutuamente exclusivos. As relações entre os nós são representadas por arcos e são descritas por distribuições probabilísticas condicionais que expressam o grau de dependência entre os nós. A figura 13 apresenta um exemplo de uma possível BN no domínio do trabalho.

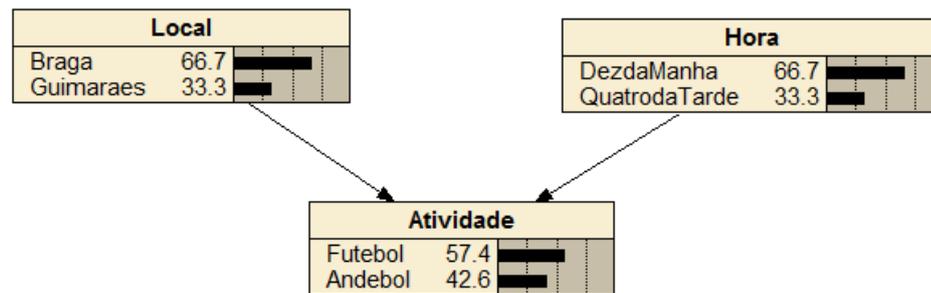


Figura 13 - Exemplo de uma Rede Bayesiana

Como se pode verificar, existem três nós, tendo cada um associado um conjunto de estados. O arco do nó Local ao nó Atividade significa que o nó Local é considerado pai do nó Atividade e representa uma relação condicional entre estes. Por outras palavras, a escolha de uma atividade é condicionada pelo local onde se realiza. Da mesma forma, existe um arco do nó Hora ao nó Atividade o que indica que a atividade não só é condicionada pelo local, mas também pela hora em que é praticada. Ao nó Atividade é associada uma tabela de probabilidades condicionais (Tabela 5), que indica a probabilidade que cada um dos seus estados tem, face a cada combinação possível dos estados dos seus pais (nós Local e Hora). Os nós que não possuam pais, como o Local e a Hora, possuem uma tabela simples indicativa do valor probabilístico de cada um dos seus estados, conforme ilustrado na figura 13.

Tabela 5 - Tabela de Probabilidade Condicional

Local	Hora	Futebol	Andebol
Braga	Dez da Manhã	66.7	33.3
Braga	Dez da Manhã	50	50
Guimarães	Dez da Manhã	50	50
Guimarães	Dez da Manhã	50	50

Uma das vantagens que as BN apresentam é não necessitarem de descrever por completo um domínio. Ao invés de uma só tabela de probabilidades conjunta que liste todos os eventos possíveis (todas combinações de estados possíveis) e as suas dependências condicionais, esta tira proveito das independências condicionais de forma a tornar mais explícita e compacta a informação estrutural do domínio. De facto, como o nó Local e Hora são

condicionalmente independentes, é possível definir que $P(Atividade|Hora,Local) = P(Atividade|Hora)$ e dessa forma diminuir o número de probabilidades a serem manipuladas. Pearl (1988) provou que nas BN apenas é necessário representar para cada variável, a sua tabela de probabilidades condicionais em relação aos seus pais e não em relação a todas as variáveis. Pode-se desta forma considerar que uma BN permite representar de uma forma eficiente, a distribuição probabilística conjunta de todas as variáveis de um determinado domínio.

A característica das BN que as tornam atrativas e muito utilizadas é a sua capacidade em inferir valores probabilísticos em situações de falta de informação e de inúmeras dependências. Tendo criado uma BN para um determinado domínio, o objetivo é a partir de um conjunto de variáveis com valores conhecidos (evidências), obter estimativas de probabilidades de variáveis relacionadas. Segundo Korb & Nicholson (2003), a inferência pode ser aplicada sobre uma BN com o propósito de diagnosticar ou de obter um prognóstico. Quando é feito um diagnóstico, o processamento é efetuado no sentido das relações, ou seja, dos efeitos para a causa. Dando um exemplo (figura 14), poder-se-á querer determinar que influência tem o local e a hora (efeitos) na probabilidade de o utilizador praticar Futebol (causa). Ao realizar um prognóstico está-se a partir das causas para os efeitos, por exemplo, sabendo que o utilizador praticou Futebol, qual a probabilidade de ter sido em Braga (figura 15).

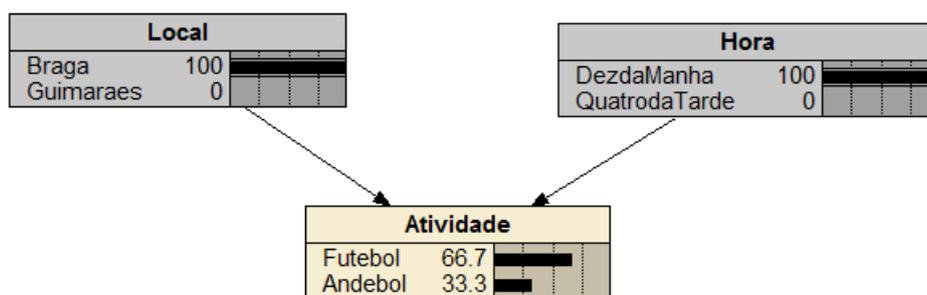


Figura 14 - Inferência de Diagnóstico

Do ponto de vista matemático, as BN utilizam o teorema de *Bayes* para propagar as evidências pelos restantes nós da rede (comparar as figuras 14 e 15 com a figura 13) que estejam relacionados com tais evidências. O objetivo é chegar a uma distribuição de probabilidades, dado um conjunto de variáveis de evidência. Num caso como o ilustrado nas figuras 14 e 15, o processo de inferência não se apresenta muito complicado, mas em situações

reais onde existem enormes quantidades de nós e de arcos são necessários algoritmos específicos.

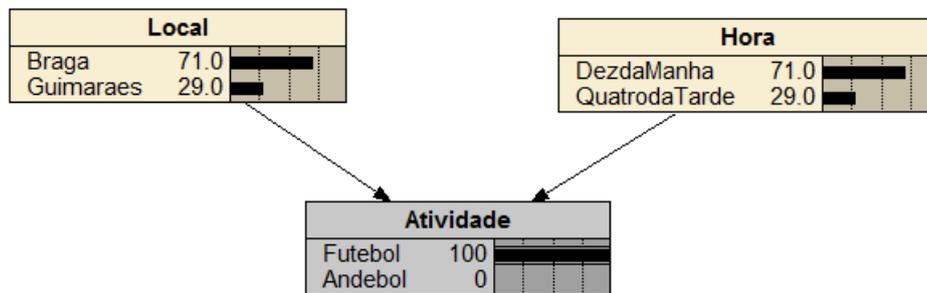


Figura 15 - Inferência de Prognóstico

Considere-se agora uma BN como ilustrado na figura 16 e o caso em que a evidência é um valor para o nó A. Este caso irá implicar que toda a rede seja percorrida e que as probabilidades de todos os nós sejam atualizadas. Para calcular as novas probabilidades do nó F será necessário antes atualizar os nós B, C e D.

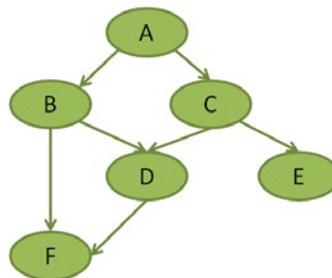


Figura 16 - Exemplo de uma Rede Bayesiana Complexa

Para estes casos mais delicados, existem vários algoritmos de inferência que segundo Castillo et al. (1997) divergem na sua complexidade e eficiência em função das situações onde são aplicados (das BN com que lidam).

As BN exibem também comportamentos inteligentes de aprendizagem e adaptação. A aprendizagem consiste em determinar uma representação para a BN (ou adaptar a existente), tendo em conta novos casos que disponibilizam valores para os nós. Por exemplo, a rede exemplificada na figura 13, terá de se adaptar face a um novo caso como por exemplo [Local = Braga; Hora = Dez da Manhã; Atividade = Andebol]. A aprendizagem de uma BN poderá ser realizada a nível estrutural ou a nível de parâmetros. A aprendizagem estrutural compreende

a determinação de dependências e independências entre variáveis, e respetiva criação ou eliminação de arcos na rede. A aprendizagem de parâmetros consiste na determinação de novos valores para a tabela de probabilidades condicionais de nós que sejam afetados pelo novo caso. Perante o caso anteriormente exemplificado, a rede inicial (figura 13) adapta-se para a ilustrada na figura 17.

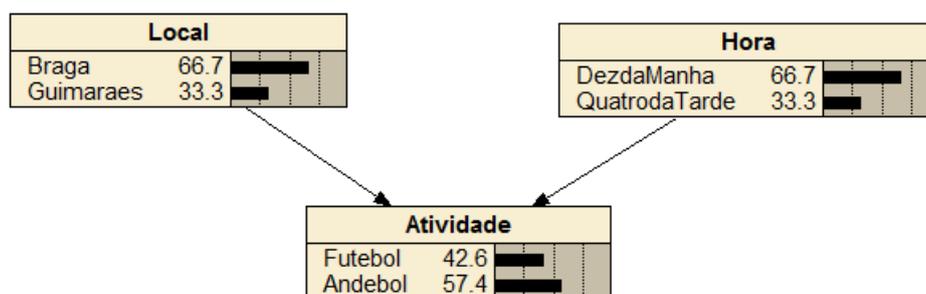


Figura 17 - Resultado da Aprendizagem de uma Rede Bayesiana

Tal como existem algoritmos de inferência, existem também algoritmos de aprendizagem (Korb & Nicholson, 2003). No caso mais simples como o exemplificado anteriormente, se não existir falta de informação no novo caso (variáveis para as quais é desconhecido o seu valor), a rede poderá ser adaptada recorrendo a contagens. Se existir falta de informação serão exigidos algoritmos mais específicos como o *Expectation Maximization* (Do & Batzoglou, 2008) que estima funções de máxima verosimilhança. Se o valor de certas variáveis não é conhecido, os casos onde estas já foram observadas poderão ser usados como forma a estimar o seu valor, de modo a maximizar os valores probabilísticos de toda a rede.

As BN são atualmente aplicadas em muitos domínios sendo o mais comum a medicina. As BN são muito usadas em sistemas que pretendam assistir os médicos no diagnóstico de doenças face ao conhecimento de sintomas. Agnieszka et al. (1999) propõem um modelo de diagnóstico de doenças hepáticas. Usando informação de uma base de dados sobre casos de pacientes, foi gerado uma BN que contem variáveis como sintomas, evidências observadas pelo médico e resultados de análises. Foram definidas dependências condicionais entre os dados obtidos da base dados, e entre estes e as dezasseis doenças hepáticas possíveis. Perante um novo caso, são introduzidas como evidências na BN as variáveis que são observadas no paciente e o sistema infere sobre qual a doença hepática mais provável de o paciente sofrer.

Num projeto idêntico (Wong et al., 2003), uma BN é usada para inferir sobre surtos de doenças de um paciente, através de indícios como a idade, a época do ano, o risco que a zona habitacional do paciente tem em relação à bactéria antraz, entre outras.

No que toca ao domínio das tecnologias, as BN são muito utilizadas em sistemas que pretendam representar perfis de pessoas, uma vez que estes são expressos muitas vezes por dependências. Um dos projetos mais antigos que lidou com BN e perfis foi o *Lumière* (Horvitz et al., 1998) da *Microsoft* que tinha como objetivo antecipar de forma inteligente (inferir sobre) os objetivos e necessidades de um utilizador das aplicações do *Microsoft Office*, tendo como base as ações e questões realizadas por este em situações passadas. Muitos recordar-se-ão do boneco de um clipe aparecer de forma a ajudar o utilizador (figura 18).

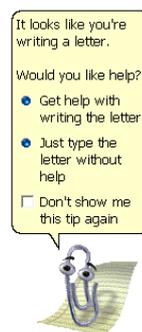


Figura 18 - Assistente do Office (retido de Horvitz et al., 1998).

Um outro projeto português lida também com perfis e as BN (Gamboa & Fred, 2001). Neste os autores pretendem implementar um sistema inteligente de *e-learning* que consoante o conhecimento adquirido por uma pessoa e as suas preferências sugira direções pedagógicas. Neste projeto é utilizado uma BN para expressar as dependências entre o conhecimento adquirido pela pessoa em determinados tópicos, e face a estes determinar os tópicos que deverão ser seguidos pela pessoa. A figura 19 ilustra um exemplo de uma BN onde se verifica por exemplo, que consoante o que a pessoa aprendeu acerca do tópico A (*Learned A*) é inferido a probabilidade do tópico A.1 dever ser sugerido. Os nós *Learned* definem o grau de confiança com que os seus tópicos foram compreendidos pela pessoa e dependem de características como o tempo que a pessoa gastou no tópico (símbolo T) e as respostas dadas às perguntas acerca deste (símbolo ?).

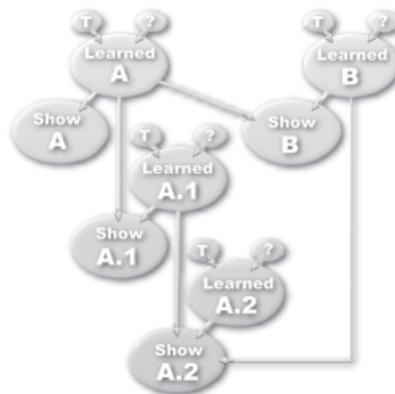


Figura 19 - Exemplo de uma Rede Bayesiana no projeto de Gamboa & Fred (retido de Gamboa & Fred, 2001).

Schiaffino & Amandi (2000) apresentam uma técnica de representação e modulação de perfis que tira proveito das BN. O projeto considera a existência de uma base de dados de uma universidade, a partir da qual qualquer pessoa poderá requisitar informação. O objetivo é aprender acerca das pesquisas que são realizadas frequentemente pela pessoa e sugerir a esta tais pesquisas, para que a que a pessoa não tenha de redigir repetidamente todos os dados das pesquisas. Uma pesquisa é constituída por um conjunto de atributos que melhor especificam o seu objetivo e por atributos acerca da pessoa que a realiza. Quando uma pessoa realiza uma pesquisa, os seus atributos são convertidos em nós da rede e são desenhados arcos entre estes, tendo em consideração as relações definidas no domínio. A BN representa desta forma as dependências entre os atributos envolvidos nas pesquisas realizadas por uma pessoa.

Embora, as BN possuam grandes vantagens em lidar com situações como as exigidas por um perfil, existem alguns fatores que limitam o seu uso neste trabalho.

Existem variáveis de uma atividade cujo domínio é ilimitado, por exemplo, a variável Local é apresentada ao utilizador como uma caixa de texto, pelo que poderão existir inúmeros estados (pode contudo ser introduzido limites de caracteres mas mesmo assim a quantidade de estados será grande). Isto influenciará sobre a manutenção da rede porque todos os estados possíveis da variável não são conhecidos à partida, e como tal, sempre que seja definido um novo estado será necessário que toda a rede seja adaptada tanto em termos estruturais como nos seus parâmetros. A complexidade desta operação dependerá da quantidade de estados que a variável afetada possua e a quantidade de outras variáveis que influencie (os nós de quem é pai). Nestas condições, uma adaptação contínua da rede poderá revelar-se um processo ineficiente.

Um outro problema que existe deve-se à determinação das relações de causalidade entre as variáveis. Quando uma rede é construída, faz parte deste processo determinar as relações entre as variáveis (os nós) e aplicá-las na BN sobre a forma de arcos. Este passo pode ser conseguido através da análise de uma base de dados inicial, com a consulta de especialistas no domínio ou conjugando as duas formas. Por exemplo, na medicina é comum os médicos auxiliarem na construção das redes uma vez que estes têm mais conhecimento sobre que sintomas tendem a influenciar determinadas doenças.

No que toca ao domínio do trabalho, a determinação das relações não apresenta ser uma tarefa simples. Como saber se é o local de um evento que influencia sobre a hora ou vice-versa? Ou seja, se deverá existir um arco do nó Local para o nó Hora ou vice-versa. A resposta a esta questão não é simples porque primeiro é pouco provável que existam especialistas na matéria e segundo porque não existe dados iniciais para consulta, na inicialização do sistema não existe um histórico de atividades que o utilizador tenha praticado. Contudo, poderiam ser realizados questionários aos utilizadores. Todavia, para conseguir construir a BN, perguntas como "O local do evento sendo em Braga irá influenciar na hora que pratica uma atividade?" teriam de ser realizadas. Consoante o número de nós existentes na rede é possível gerar várias relações entre estes, e como tal irá ser cansativo para o utilizador responder a um questionário volumoso. Mesmo assim, o que no início poderia ser verdade em relação às dependências expressas pelo utilizador, com o passar do tempo poderia tender a mudar, o que implicaria adaptações na rede.

Outro problema existente nas BN é a necessidade de representar cada estado das variáveis da rede num nó isolado. Isto deve-se ao facto de que dependendo do estado que uma variável tenha, esta pode ou não influenciar sobre outras. Por exemplo, se local do evento for Braga então a hora do evento deverá ser às 16h, porém o facto de o local ser Porto não influencia sobre a hora. Neste caso o mais correto é seguir uma abordagem idêntica ao do projeto de Schiaffino & Amandi (2000), em que é criado uma variável (nó) para cada valor possível dos atributos característicos de uma atividade, e em que cada variável é bidimensional. Apenas são considerados os estados Verdadeiro ou Falso como exemplificado na figura 20.

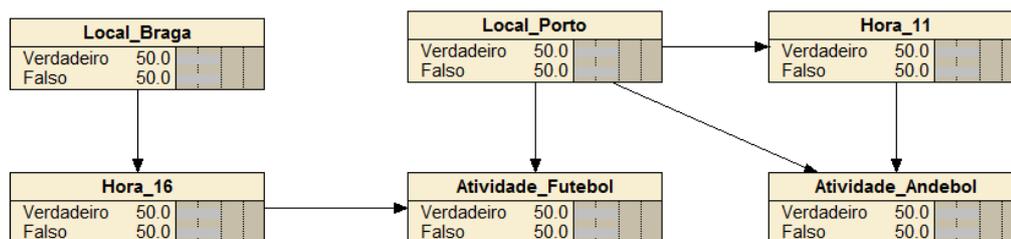


Figura 20 - Exemplo de uma Rede Bayesiana com nós bidimensionais

Contudo, uma árvore deste tipo continua a acarretar problemas, uma vez mais devido ao facto de existirem atributos da atividade cujos valores são ilimitados. Exemplificando, como o atributo Local não apresenta um limite de valores possíveis, é necessário criar um nó para cada um dos valores possíveis e definir os arcos necessários. Esta solução fará com que a rede seja muito volumosa, ao mesmo tempo que aumentará o tempo de processamento de uma inferência.

O processo de aprendizagem da rede tenderá também a ser mais demorado e complexo. Ao ser criada uma atividade, cada valor especificado nos atributos terá associado um nó na rede que será definido como estando no estado de verdadeiro. Porém, todos os restantes nós terão de ser considerados como estando no estado falso. Isto implica com que a cada processo de aprendizagem seja necessário determinar todos os nós que não figurem na atividade, e definir o seu estado como falso.

3.1.2 - Raciocínio Baseado em Casos

O Raciocínio Baseado em Casos (*Case Based Reasoning* - CBR) (Soltysiak & Crabtree, 1998) é uma metodologia de resolução de problemas que disponibiliza métodos eficientes e eficazes, capazes de determinar a solução para um novo problema com base na reconstituição de casos passados similares, reutilizando-se ou adaptando-se o conhecimento de tais casos. Um caso é um exemplo de uma situação passada onde procedeu-se à resolução de um problema. É composto por uma descrição do problema a resolver, a solução para o problema e adicionalmente uma justificação para a solução. A descrição do problema deverá conter informação suficiente e sucinta para caracterizar o caso, e devera ter especiais cuidados quanto à sua representação, uma vez que é essencial para o funcionamento do CBR. É a partir da descrição que é possível encontrar casos passados semelhantes.

O CBR é um processo comum na sociedade. Um médico deduz o tratamento para uma doença com base em diagnósticos e tratamentos similares realizados no passado. Um mecânico deduz o problema de um veículo e respetiva solução com base em problemas semelhantes e passados que já resolveu. O CBR é associado ao conceito de experiência e aprendizagem de um individuo. Com o evoluir, uma pessoa enriquece a sua capacidade de resolução de problemas (experiência) uma vez que aprende e armazena soluções para novos casos.

Todos os casos são arquivados num repositório que é consultado quando há necessidade de solucionar um novo problema. O processo de encontrar uma solução para um novo caso com recurso ao CBR é definido em quatro fases conforme ilustrado na figura 21, sendo definido na literatura por Aamodt & Plaza (1994) como os quatro R's.

A primeira fase é aquela que maiores cuidados e atenções requiere e tem como objetivo recuperar o caso passado mais similar com o novo caso. Determinar o quanto dois casos são similares, implica comparar as duas descrições do problema. Como já referido, para que este passo seja bem conseguido é importante definir regras quanto à representação da descrição do problema de cada caso, sendo adequado que todos os casos sigam as mesmas matrizes. Usualmente a descrição do problema de um caso é constituída por um conjunto de variáveis comuns aos casos mas cujos valores variam. Cada variável poderá ter o seu próprio domínio desde que todos os casos o respeitem. A figura 22 ilustra como um caso pode ser representado no âmbito deste trabalho. Neste caso o problema é encontrar uma atividade para praticar e o respetivo local, tendo conhecimento da hora e do dia da semana em que se realiza.

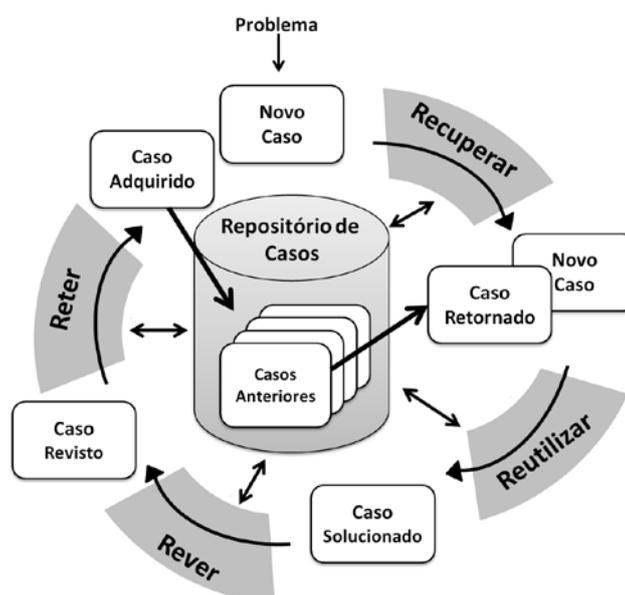


Figura 21 - Funcionamento do Raciocínio Baseado em Casos (adaptado de Aamodt & Plaza, 1994).

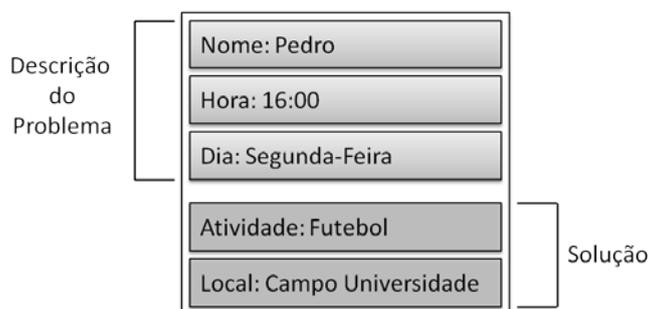


Figura 22 - Exemplo de um caso do Raciocínio Baseado em Casos

Ao contrário das bases de dados convencionais que apenas lidam com valores de certeza exata (assumem que dois casos são iguais ou não o são), o CBR lida com valores reais no intervalo de 0 a 1, podendo a similaridade entre dois casos ser, por exemplo 0.5. Um valor de similaridade 1 significa que os casos são exatamente iguais e corresponde às consultas que são feitas a uma base de dados convencional. Um valor 0 indica que os casos são totalmente diferentes e corresponde a não obter quaisquer resultados numa consulta a uma base de dados.

O cálculo do valor de similaridade entre dois casos utiliza uma função que através das similaridades locais, entre cada variável dos casos (figura 23), determina a similaridade global entre os casos. A determinação da similaridade entre duas variáveis dependerá do seu domínio (Stahl, 2003). No caso de serem valores numéricos a distância entre estes poderá ser usada como forma de determinar a sua similaridade. Funções como $f(d) = 1 - 1/(1 + d)$ para casos onde a distância máxima não seja conhecida e $f(d) = 1 - d/max$ para casos onde a distância máxima seja conhecida, são as mais comuns para traduzir uma distância numa medida de similaridade. Caso as variáveis sejam por exemplo palavras, o cálculo da similaridade requer técnicas mais complexas como por exemplo a distância de *Jaccard*.

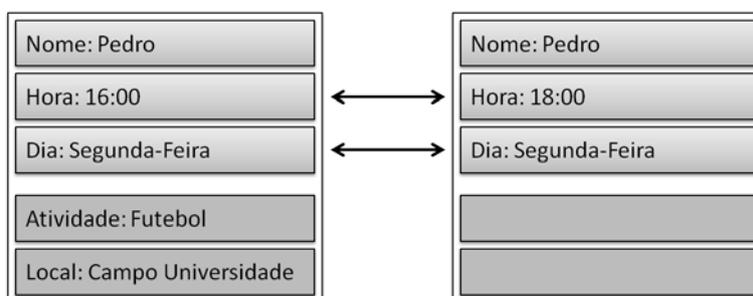


Figura 23 - Cálculo da similaridade local entre dois casos

Uma vez obtida a similaridade local entre todas as variáveis de cada caso, poderá ser calculada a similaridade global entre os casos. A função mais comum de usar é a média de todas as similaridades locais que poderá ser expressa da seguinte forma:

$$sim(X, Y) = \frac{\sum_{i=0}^n sim(x_i, y_i)}{n} \quad (E1)$$

onde $sim(X, Y)$ é a similaridade global entre dois casos X e Y , $sim(x_i, y_i)$ representa a similaridade local entre cada variável dos casos e n é o número de variáveis dos casos. Contudo, em algumas situações poderá ser conveniente atribuir pesos a cada variável, para que algumas tenham maior importância no cálculo da similaridade do que outras. Nestes casos a função mais adotada é a seguinte:

$$sim(X, Y) = \frac{\sum_{i=0}^n \alpha_i * sim(x_i, y_i)}{\sum_{i=0}^n \alpha_i} \quad (E2)$$

sendo α_i o peso a atribuir a cada similaridade local e podendo ser representado por números reais ou naturais, desde que maiores do que zero. De notar que a equação 1 é uma redução da equação 2 para situações onde os pesos α_i é igual para todas as variáveis.

A fase seguinte consiste na reutilização (ou adaptação) da solução associada ao caso recuperado. Numa versão mais simples, a solução do caso recuperado é transferida para o novo caso. Todavia, poderá haver situações em que tal não seja possível, sendo necessário que a solução seja adaptada para o novo caso. Nestas situações, poderá ser necessário o conhecimento de um especialista ou então poderão ser aplicadas técnicas automáticas de adaptação (Wilke et al., 1998, citado em Stahl, 2003).

Uma vez encontrada a solução para o novo problema, em algumas situações poderá ser necessário que a solução seja revista, de forma a determinar a sua validade. Poderá ser necessário experimentar a solução obtida e verificar se resolve o problema. Dependendo do domínio onde seja aplicada, a revisão poderá ser feita de forma automática ou então com recurso a uma pessoa especializada no domínio. Considerando por exemplo, que um veículo possui um problema e é obtida uma solução com base em casos passados, por exemplo, a

solução poderá ser mudar um fusível, para saber se a solução é válida o mecânico substitui o fusível e verifica se o problema prevalece. Se a revisão a uma solução definir que a solução não é válida esta terá de ser reparada. A reparação poderá passar por aplicar outra técnica de adaptação da solução, utilizar o caso seguinte mais similar ou realizar um novo processo de recuperação alterando determinados campos.

A fase final é aquela que torna o CBR um sistema inteligente uma vez que é nesta que aprende com novos casos. Encontrada um solução final e revisada para o problema descrito pelo novo caso, o caso é retido no repositório para que em problemas futuros e similares possa ser usado como fonte de solução. Consoante o domínio onde é aplicado o CBR e a quantidade de casos que armazene, poderá ser necessário uma reorganização do repositório de casos. A eliminação de casos demasiado antigos é um acontecimento possível. Como é citado por Stahl (2003), são várias as propostas de técnicas para reorganizar o repositório.

O CBR tem a capacidade de adaptar-se a vários domínios sendo essa razão pela qual é muito utilizado. Tal como as BN, a medicina é uma área onde é possível aplicar o CBR no diagnóstico de doenças e auxílio ao médicos. O *FM-Ultranet* (Balaa et al., 2003) é um sistema que tira proveito do CBR para detetar malformações ou anomalias nos fetos através de ecografias. Os casos são constituídos por hierarquias de conceitos, tendo cada conceito uma ou mais variáveis que definem informações médicas e anatómicas. Usando valores derivados da ecografia, o sistema procura por casos passados similares e produz como resultado um conjunto com zero ou mais malformações ou anomalias detetadas.

Um outro projeto desenvolvido por Berkat (2011) utiliza o CBR com o objetivo de detetar vírus num computador. Os antivírus vulgarmente realizam atualizações diárias com o intuito de descobrir novas ameaças e dessa forma detetar novos vírus num computador. Os antivírus são algo dependentes destas atualizações uma vez que se não as realizarem, a probabilidade de o computador estar infetado é maior. Face a esta realidade, o autor propõe uma técnica de deteção de vírus que tira proveito de casos passados que representam ficheiros que no passado foram detetados como vírus. Quando um ficheiro é analisado são recuperados os casos passados onde ficheiros idênticos tenham sido considerados vírus. A solução dos casos poderá corresponder à ação de eliminar ou ignorar o ficheiro. Se existir um caso passado similar com o ficheiro a analisar e a sua solução foi eliminar, é apresentado ao utilizador a informação que o ficheiro é provavelmente um vírus e que o deverá eliminar (reutilização). O utilizador em seguida

revisa a solução proposta e aceita ou rejeita que o ficheiro seja eliminado. A seguir à resposta do utilizador o novo caso é arquivado no repositório.

No que toca à utilização do CBR para tratamento de perfis, destaca-se o projeto desenvolvido por Schiaffino & Amandi (2000) que assemelha-se ao desenvolvido neste trabalho e que já foi mencionado aquando da discussão das BN. Neste, um caso representa uma consulta realizada pelo utilizador à base de dados. A descrição de um caso contém os atributos usados na consulta, informação acerca do utilizador e outros dados. A solução é um código usado para agrupar os casos quanto à sua similaridade. A lógica seguida é que as consultas feitas por um utilizador em situações passadas poderão disponibilizar informação acerca de consultas que o utilizar possa vir a realizar. Neste contexto, o conjunto de casos armazenados pelo CBR poderá ser visto como o perfil do utilizador, representativo das tendências ou interesses do utilizador em relação a consultas realizadas à base de dados. O uso do CBR tem como objetivo conseguir encontrar novos tópicos de interesse do utilizador. Para tal, quando é realizada uma consulta, é criado um novo caso e são recuperados do repositório os casos passados mais similares. Se existirem casos similares, então a solução para o novo caso é adaptada dos casos passados (é atribuído o mesmo código, ou seja, o caso é agrupado). Se não existirem casos similares, então a consulta realizada é uma novo tópico de interesse do utilizador e como tal é atribuído um novo código (é criado um novo grupo).

Abordagem semelhante seguem Godoy & Amandi (2000) no seu projeto que tem como objetivo assistir os utilizadores na busca de documentos do seu interesse na internet. Os autores focam-se no problema que existe na pesquisa de tópicos em motores de busca. Exemplificando, quando realizamos uma pesquisa pela palavra "puma" os motores de busca poderão dar como resultado documentos relativos ao animal puma ou documentos relativos à marca desportiva *Puma*. Se o sistema aprender que o utilizador quando procura por "puma" está interessado no animal ao invés da marca desportiva, poderá filtrar os resultados para que apenas sejam apresentados aqueles do interesse do utilizador. O CBR é usado com o intuito de detetar tópicos específicos em que o utilizador esteja interessado e como forma de organizar o seu perfil. A organização do perfil consiste mais especificamente em agrupar os casos (as pesquisas realizadas) do utilizador como uma estrutura hierárquica de tópicos (conforme ilustrado pela figura 24). De forma idêntica ao projeto anterior, quando é realizada uma pesquisa, um novo caso com informação desta é criado e a função do CBR é determinar em que tópico deverá o novo caso ser agrupado, ou então se deverá ser criado um novo tópico.

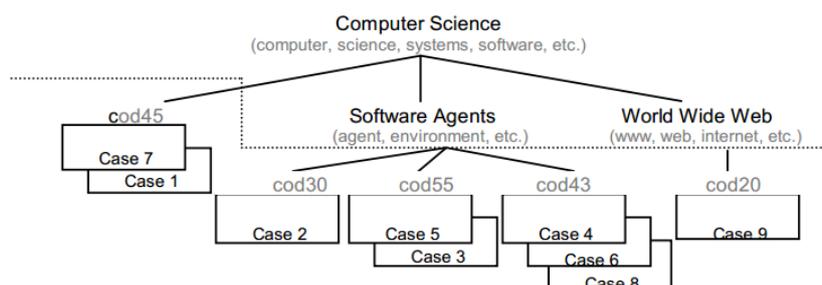


Figura 24 - Exemplo da representação de um perfil no projeto de Godoy & Amandi (retido de Godoy & Amandi, 2000)

Como poderá ser compreensível pelas figuras 22 e 23, o CBR é uma técnica candidata a ser implementada no domínio deste trabalho. Um perfil poderá neste caso ser expresso por um conjunto de casos que representam atividades realizadas pelo utilizador. Quando houver necessidade de recomendar uma atividade para um determinado espaço temporal, uma das formas de o conseguir poderá ser consultar casos passados similares e usar a atividade definida nesses casos como solução. Porém, existem alguns pontos que necessitam de maior análise para uma utilização correta e eficiente desta técnica.

O fator tempo deverá ser introduzido no cálculo da similaridade entre casos, face à capacidade do ser humano em evoluir e mudar aquilo por que se interessa. Deverá ter-se em conta a distinção entre interesses de curto prazo e interesses de longo prazo. Por exemplo, tendo o utilizador praticado uma determinada atividade A e não a tendo voltado a repetir num espaço de um mês, poderá ser um indício de que o utilizador reduziu o interesse pela atividade. Desta forma, quando houver a necessidade de recomendar uma atividade em condições semelhantes, a probabilidade de ser a atividade A deverá ser menor uma vez que não existe a certeza desta ser ainda é do interesse do utilizador. Por outras palavras, as atividades semelhantes e praticadas pelo utilizador mais recentemente deverão ter prioridade.

Funcionando o CBR como um repositório de casos, a quantidade de informação que este armazena tende a expandir rapidamente. Embora hoje em dia o custo de armazenamento digital seja baixo, em algumas situações poderá haver necessidade de organizar o repositório. Poderá ser aconselhado eliminar casos que sejam demasiado antigos e que não se apresentem como relevantes. A volumosa quantidade de informação existente poderá também aumentar o tempo necessário para a recuperação de casos similares. Nestes casos, técnicas eficientes de

indexação e categorização dos casos, poderão ser adotadas como forma a minimizar a quantidade de casos necessários de analisar.

A estruturação dos casos deverá também ser definida com especial atenção, tendo em conta que é essencial para que o processo de recuperação de casos seja eficiente. A quantidade e domínio das variáveis usadas para descrever um caso, influenciará sobre o tempo necessário para calcular a similaridade entre casos. Enquanto a utilização de valores numéricos poderá tornar o cálculo da similaridade mais rápida, a utilização de frases ou palavras poderá enriquecer a informação de um caso. A quantidade e domínio das variáveis deverá ser o mais adequado à situação onde é aplicado, de forma a encontrar um equilíbrio entre possuir uma informação enriquecida e ser rápido no cálculo da similaridade.

3.1.3 - Regras de Associação

Uma das técnicas de extração de conhecimento (H.Witten & Frank, 2005) muito utilizada é denominada como Regras de Associação (*Association Rules - AR*). Esta técnica tem como principal objetivo encontrar elementos de um repositório de registos que ocorram conjuntamente. Através da sua aplicação é possível descobrir regras de implicação ou correlação entre elementos. Um exemplo prático de onde esta técnica é geralmente usada é o marketing. Considerando uma cadeia de supermercados, esta através da aplicação das AR é capaz de encontrar hábitos de consumo, ou seja, encontrar grupos de produtos que o consumidor frequentemente adquire (por exemplo, descobrir que a grande maioria dos clientes quando compra um Produto A, também compra um produto B). A partir destas relações, é possível tomar medidas de marketing como por exemplo, lançar campanhas de descontos, organizar a distribuição dos produtos (colocar o produto A próximo do B), entre outras.

Uma regra de associação representa-se por $A \rightarrow B$, onde A e B são um conjunto de variáveis tal que $A \subset I$ e $B \subset I$ e $A \cap B = \emptyset$. O conjunto A é considerado o antecedente da relação e B o conseqüente. Uma regra traduz uma relação de implicação entre o antecedente e o conseqüente. Exemplificando, sendo $I = [Local, Hora, Atividade]$, as seguintes regras poderão ser produzidas:

$$[Local = Braga, Hora = 16] \rightarrow [Atividade = Futebol] \quad (R1)$$

$$[Local = Braga] \rightarrow [Atividade = Futebol] \quad (R2)$$

Considerando a regra R1, esta dita que quando o local da atividade é Braga e a hora é 16 h, o utilizador tem preferência que a atividade seja Futebol.

São também definidas métricas do suporte e confiança para cada regra que melhor exprimem a sua relevância. O suporte permite determinar a frequência com que a regra existe no conjunto de dados. A confiança define a percentagem com que o antecedente ocorre juntamente com o conseqüente. Admitindo que a regra R1 tem um valor de suporte de 80% e confiança de 30%, isto permite concluir que das relações obtidas do repositório, em 80% destas o antecedente é composto pelos valores "Local=Braga" e "Hora=16", o que de certa forma revela que a ocorrência destes dois valores é algo que influência bastante nas demais variáveis. Por sua vez, um valor de confiança 30% significa que das regras onde o antecedente contem os valores "Local=Braga" e "Hora=16", em 30% destas o conseqüente contem o valor "Atividade=Futebol", indicativo que quando o local da atividade é em Braga e a hora às 16, em 30% das vezes o utilizador prefere Futebol.

Como exposto, o objetivo da aplicação desta técnica de extração de conhecimento é a partir de um repositório de registos gerar um conjunto de regras que cumpram com um valor de suporte e confiança mínimo previamente definido. Como é de esperar, a partir de um conjunto de variáveis são possíveis de gerar uma grande quantidade de regras que terão de ser analisadas. O número de regras está relacionado com a quantidade de variáveis que um conjunto tenha e da quantidade de estados que cada variável detenha. Admitindo por exemplo que cada variável apenas possui um estado e que o conjunto I é constituído por n variáveis, o número exato de regras ($\#r(I)$) que é possível de gerar é definido pela seguinte função:

$$\#r(I) = \sum_{i=1}^n \binom{n}{i} \times (2^{n-i} - 1) \quad (E3)$$

onde $\binom{n}{i} = C_i^n$ e representa de quantas formas distintas é possível escolher i elementos de um grupo de n elementos (combinações de n , i a i). Assim, para um conjunto como $[Local = Braga, Hora = 16]$ são possíveis de gerar 2 regras:

$$[Local = Braga] \rightarrow [Hora = 16] \quad (R3)$$

$$[Hora = 16] \rightarrow [Local = Braga] \quad (R4)$$

Se as variáveis possuírem mais do que um estado, o número de regras será superior, levando a que o tempo necessário para a aplicação da técnica de AR seja maior.

Para diminuir o tempo de processamento existem algoritmos que aplicam conhecimentos relacionados com conjuntos e suas frequências, como forma de diminuir o número de regras necessárias de analisar. O mais conhecido e aplicado é apelidado de *Apriori*, denominação que resulta do facto de este utilizar informação à priori sobre a frequência dos dados no processo de pesquisa e análise. A procura de regras segue um processo iterativo pelos diferentes níveis de conjuntos (que representam relações, um conjunto $[A,B]$ representa a regra $A \rightarrow B$ ou $B \rightarrow A$) e tem como base dois princípios. O primeiro define que um conjunto é frequente se o seu valor de suporte é elevado. Qualquer subconjunto de um conjunto de variáveis frequente é também frequente. Exemplificando, se o conjunto $[A,B,C]$ é frequente então o conjunto $[A,C]$ também o será. O segundo princípio define que se um conjunto de variáveis não é frequente então quaisquer seus superconjuntos não deverão ser analisados. Se o conjunto $[A,B]$ não é frequente então os conjuntos $[A,B,C]$ ou $[A,B,D]$ também não o serão e a sua análise poderá ser descartada. Este princípio faz com que o tempo de processamento seja menor uma vez que evita a análise de todos os conjuntos (de todas as regras).

Adicionalmente, a confiança poderá também ser usada na eliminação de contradições ou redundâncias nas regras. Usando normas de poda (Shah et al., 1999) poderão ser eliminadas regras redundantes. Se duas regras $A, B \rightarrow C$ e $A \rightarrow C$ possuem valores idênticos de confiança, então a primeira regra é considerada como redundante pelo facto de adicionar pouca informação à segunda. Aquando da existência de duas regras contraditórias entre si, como por exemplo as regras R3 e R4, aquela que tenha menor valor de confiança será descartada.

Para além de serem definidos valores mínimos de confiança e suporte que servirão de filtros na geração de regras, uma especificação das regras que são desejadas de obter contribui para a diminuição do tempo de processamento. Referindo novamente o marketing, se pretende-se conhecer que outros produtos os consumidores tendem a adquirir quando compram um produto A, então só terão interesse regras cujo antecedente seja A pelo que as demais não necessitarão de ser analisadas. A figura 25 sumariza o funcionamento das AR (o suporte e confiança das regras são expressos por [suporte, confiança]).

Referindo novamente a medicina como domínio onde as AR são muito aplicadas, o principal objetivo da sua aplicação consiste em encontrar relações entre dados médicos. É precisamente isso que Ordenez et al. (2001) fazem no seu projeto. Pretendem através de um repositório de caso clínicos encontrar regras entre atributos como a idade, risco de diabetes, o nível de colesterol e outros. Através das regras obtidas e consoante a sua confiança e suporte, estas poderão ser usados na prevenção de doenças cardiovasculares. Para além dos objetivos em termos médicos, os autores apresentam também um algoritmo de mapeamento para dados clínicos que revela-se eficiente em relação ao tipo de dados com que tem de lidar. Os dados médicos são geralmente constituídos por valores numéricos, categorias, tempos e imagens. Face à complexidade que os dados contêm, é necessário realizar tratamento destes dados para que a aplicação das AR seja eficiente e coerente. Por exemplo, os valores numéricos por conterem um domínio contínuo deverão ser agrupados em intervalos.

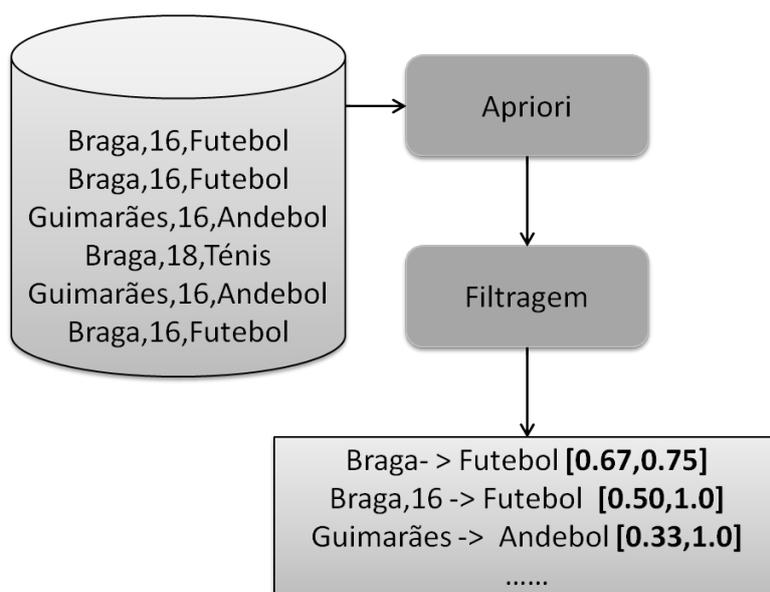


Figura 25 - Funcionamento das Regras de Associação

Uma vez mais, refere-se o marketing como um domínio onde a utilização das AR revela-se benéfica e lucrativa. Exemplo disso é o projeto desenvolvido por Wang et al. (2005) que utiliza as AR como forma a maximizar os lucros das campanhas publicitárias. Os autores focam-se mais precisamente nas campanhas de angariação de dinheiro que são feitas por correio eletrónico. Através de uma base de dados contendo campanhas anteriores de angariação e informação acerca da aceitação das pessoas e o valor da doação, o objetivo é através das AR prever o grupo de utilizadores mais plausíveis de doar e maximizar os ganhos da campanha.

No que toca mais precisamente à modulação de perfis, existem muitas propostas de utilização das AR, das quais destaca-se as seguintes por serem originais. Ling et al. (2007) apresentam um modelo que lida com perfis das pessoas no que toca às suas preferências por peças de roupa. Os autores utilizam a tecnologia RFID (identificação por radiofrequência) para detetar que peças de roupa uma pessoa veste no seu dia-a-dia e com base no registo desta informação modelam o seu perfil que serve de base para a recomendação de peças a usar em diferentes situações. No que toca mais precisamente à utilização das AR, estas permitem detetar as relações existentes entre as várias peças de roupa que uma pessoa usa e as suas características. A ideia é por exemplo determinar de que forma a utilização de uma camisola ou camisa irá influenciar sobre que peça de roupa inferior usar, se uma saia ou umas calças.

Chen et al. (2007) propõem uma forma de melhorar as plataformas de *e-learning*. O objetivo principal é adaptar os parâmetros de dificuldade dos cursos, de acordo com as necessidades de cada pessoa. Esta adaptação poderá traduzir-se na alteração da ordem pela qual os tópicos são lecionados. Neste projeto um perfil representa de certa forma o conhecimento adquirido pela pessoa ao longo dos cursos que frequenta. Um perfil é constituído pelas respostas a exames de cursos, podendo incluir tanto exames onde a pessoa errou como onde acertou. As AR utilizam a informação de exames onde as pessoas tenham errado para descobrir os equívocos que são comuns de serem realizados. A ideia básica é descobrir relações entre os equívocos. Por exemplo, descobrir que pessoas que erram num determinado tópico A tendem a também errar no tópico B. Com base nas regras obtidas, a estrutura dos cursos poderá ser alterada como forma a diminuir a sua dificuldade e a probabilidade de ocorrência de equívocos.

A aplicação desta técnica de extração de conhecimento no domínio do trabalho é uma forma útil de através de registos de atividades e os seus atributos descobrir relações. Como

exemplificado pela regra R1, podem ser descobertas as preferências do utilizador por atividades face a certas características. A obtenção destas regras permite determinar que atividades recomendar numa determinada situação. Se a regra R1 tem um valor alto de suporte e confiança, isto permite afirmar que quando haja um espaço livre às 16 horas e o utilizador encontre-se ou regida em Braga, deverá ser recomendada a atividade Futebol.

As regras permitem descobrir relações entre os atributos de uma atividade, por exemplo, poderão ser descobertas regras que indiquem o local pelo qual o utilizador tem preferência, no caso da hora ser às dez da manhã.

Embora esta técnica apresente ser adequada e útil de implementar no trabalho, existem algumas limitações e questões relacionadas com a sua eficiência que comprometem a sua adoção. Como se verificou as AR é uma técnica que requer uma análise de um registo de casos passados para obter um conjunto de regras. Como já referido, aquando da discussão da aplicação do CBR, no início do funcionamento do sistema não existe registos. Ao mesmo tempo, garantir que a informação de um perfil mantém-se atualizada, neste caso que os valores de suporte e confiança das regras é sempre o mais recente em relação aos registos, implicaria numa primeira solução utilizar sistematicamente (a cada nova atividade criada ou aceite pelo utilizador) algoritmos como o *Apriori*, o que revelar-se-ia muito pouco eficiente.

Uma segunda solução passa por fazer com que apenas as regras subjacentes a uma nova atividade sejam alvo de análise e atualização. Contudo, com este procedimento surgem dois novos problemas. O primeiro é que continua a ser necessário analisar uma quantidade satisfatória de regras que dependerão da quantidade de variáveis (atributos) que uma atividade contenha. Neste caso pode ser considerado o valor atual de suporte ou confiança de cada regra, de forma a só serem atualizadas regras que sejam superiores a um determinado valor. Contudo, será necessário que para cada regra seja realizada uma consulta à base de dados que verifique os valores atuais de suporte e confiança. O segundo problema prende-se com o valor que as demais regras, que não as subjacentes à nova atividade terão. A introdução de um novo registo significa que o número total destes aumenta e que os valores de suporte de todas as regras necessitem de ser atualizados, uma vez que o cálculo do valor de suporte depende do número total de registos.

O facto de existirem variáveis com intervalos indefinidos, como a variável Local, significa que cada utilizador terá a si associado uma quantidade volumosa de regras o que implica sobre

o tempo de processamento necessário para analisar todas as regras, tanto para o processo de atualização como para o processo de descoberta de uma atividade a recomendar.

3.2 - Aquisição de um Perfil

Conseguir garantir que a informação de um perfil mantém-se atualizada compreende detetar evoluções no ser humano como mudanças nas suas preferências, nos seus hábitos, entre outras. É necessário recolher sistematicamente informação acerca do utilizador e aplicá-la no seu perfil para que este seja consistente e moderno. Numa versão futurística como são os ambientes inteligentes, este processo será conseguido de forma impercetível para as pessoas, de forma implícita. Porém, a monitorização perfeita das pessoas é ainda algo que exige muita investigação, nomeadamente aspetos como deteção de estados emocionais ou de comportamentos que requerem ainda muitos estudos e provas. Por esta razão, existem também formas explícitas de obter informação que tentam auxiliar neste processo de aquisição de informação.

Seguidamente explica-se cada uma destas técnicas de aquisição de informação e posteriormente compara-se as duas.

3.2.1 - Aquisição Explícita

As técnicas explícitas como o próprio nome o invoca são técnicas que são visíveis ao utilizador e que dependem da sua interação para obtenção de informação acerca deste. Esta informação é normalmente conseguida através da solicitação ao utilizador para que preencha formulários ou então quando o utilizador de forma voluntária transmita a informação.

O primeiro caso ocorre geralmente quando um utilizador regista-se em algum serviço. Um exemplo poderá ser os sites que já foram referenciados, em que no registo é solicitado ao utilizador que indique os seus interesses. Em alguns casos, os sistemas poderão requisitar que o utilizador preencha um formulário após prestar algum serviço, não sendo contudo recomendado. Uma requisição constante de informação irá decerto entediar o utilizador e conseqüentemente este deixará de usar o serviço ou o avaliará negativamente.

Em alternativa, os sistemas deverão optar por disponibilizar métodos a partir dos quais os seus utilizadores possam de forma voluntária transmitir informação. Os sites de comércio

eletrónico são um exemplo, estes permitem que os seus utilizadores possam indicar os produtos que desejam (pelos quais têm interesse). Os utilizadores podem também avaliar os produtos que compram. O *NewsAgent* (Zhong et al., 2004) é um agente que auxilia os utilizadores que leem frequentemente notícias ou artigos na internet. O sistema disponibiliza métodos para que os utilizadores possam indicar tópicos ou diários que gostam de ler frequentemente, bem como permite que estes possam avaliar as páginas que leem. Com a informação adquirida, o sistema recomenda outras páginas que possam ser do interesse do utilizador. O *Youtube* ou *Facebook* são outros exemplos onde o utilizador também indica de forma voluntária os conteúdos de que gosta.

A aquisição de informação de forma explícita deverá ser feita de forma simples para não levar o utilizador a não transmiti-la. Não é aconselhável que um formulário tenha grande quantidade de perguntas e que estas sejam complexas. Da mesma forma, os métodos disponibilizados pelos sistemas para recolha de informação deverão ser simples, diretos e intuitivos. Considerando por exemplo a avaliação de artigos, as hipóteses existentes para avaliar não deverão ser demasiadas, estados simples como mau, bom ou excelente deverão ser usados ao invés de por exemplo escalas percentuais. Um exemplo muito utilizado é o mecanismo de avaliação de cinco estrelas.

3.2.2 - Aquisição Implícita

Ao contrário das técnicas explícitas, as técnicas implícitas obtêm informação sem a necessidade de interagir com os utilizadores sendo por isso invisíveis a estes. Para que tal seja possível, é necessário observar e registar as ações do utilizador. Através dos repositórios de informação acerca das ações é possível em seguida extrair conhecimento, extrair informações acerca das preferências, tendências, hábitos ou outras.

A procura de informação relevante a partir de grandes quantidades de informação não é um processo simples, pelo que são requeridas técnicas de extração de conhecimento (*Data Mining* na terminologia informática) que através de algoritmos próprios e eficientes o permitem. As AR que foram analisadas são um exemplo destas técnicas de extração. A segmentação (*Clustering* na terminologia informática) é outro exemplo. Para uma análise em detalhe destas técnicas recomenda-se a leitura do livro de H.Witten & Frank (2005).

Existem vários exemplos de onde é obtida informação acerca de um utilizador de forma implícita. Os ambientes inteligentes recolhem informação através da computação ubíqua, ou seja, através de milhares de microprocessadores que observam e registam as ações das pessoas. As casas inteligentes são um exemplo atual onde através de sensores recolhem informação acerca das ações do utilizador e encontram padrões ou hábitos.

Os sites de comércio eletrónico são também outro exemplo, estes observam as pesquisas e compras realizadas por um utilizador e através destas deduzem outros produtos em que o utilizador estará interessado. Para além de no seu projeto disponibilizarem formas explícitas de recolher informação, Zhong et al. (2004) desenvolveram também um agente que filtra as páginas resultantes das pesquisas realizadas por um utilizador nos motores de busca de acordo com os seus interesses. A recolha dos interesses do utilizador pelas páginas é feita de forma implícita. Medidas como o tempo despendido para ler a página, a quantidade de movimentos verticais realizados e o facto de ser adicionada aos favoritos são usados como forma a deduzir o interesse que o utilizador tem pela página.

Recentemente a área da computação afetiva tem vindo a crescer. Nesta área são feitos estudos que tentam analisar os estados emocionais das pessoas de forma implícita. Num estudo realizado por Khanna & M.Sasikumar (2010), a força com que o teclado é pressionado é usada como fonte de informação acerca do estado emocional da pessoa.

3.2.3 - Comparação

Decidir entre a adoção de técnicas explícitas ou implícitas de obtenção de informação não se apresenta como uma tarefa fácil uma vez que ambas têm as suas vantagens e desvantagens. Contudo, pode-se de certa forma afirmar que as implícitas apresentam algumas vantagens sobre as explícitas, razão que as torna mais adotadas.

Numa primeira análise às explícitas, a sua principal vantagem é permitir de forma rápida e direta obter informação acerca do utilizador. Estas requerem também menos cuidados e requisitos computacionais na sua implementação e tratamento dos dados obtidos uma vez que a informação possível de obter é geralmente conhecida de forma prévia. Ao colocar-se um sistema de avaliação de cinco estrelas, a informação obtida será um valor numérico que apenas terá de ser aplicado no perfil recorrendo a fórmulas matemáticas.

Porém, as técnicas explícitas apresentam também alguns problemas e limitações. A primeira já foi discutida e envolve a vontade do utilizador em preencher formulários. Este é um cenário que será pouco provável de ocorrer se os formulários forem demasiado complexos e volumosos ou se o utilizador for sistematicamente requisitado para tal fim. Mesmo preenchendo um formulário, a informação obtida poderá não ser realmente verdadeira, podendo haver uma divergência em relação à real. Na verdade, se o utilizador tiver de preencher formulários volumosos é muito provável que o faça sem fazer uma análise detalhada ao que é pretendido, fazendo com que a informação obtida seja errada.

Uma outra desvantagem das técnicas explícitas é não ser fácil de adotar em certos domínios ou condições. Em alguns casos o utilizador poderá não saber como expressar certas informações como os seus interesses ou preferências. Se for pedido ao utilizador que expresse a sua preferência por praticar Futebol, este poderá não conseguir expressar que o seu interesse dependerá de certas condições como a hora de começo ou local onde será praticada.

Neste cenário, as técnicas implícitas apresentam-se vantajosas uma vez que conseguem detetar tais dependências sem a necessidade de interagir com o utilizador. Nestas técnicas a probabilidade de a informação obtida ser divergente da real é baixa uma vez que as ações do utilizador são monitorizadas e a partir destas são possíveis de detetar padrões. Contudo, esta capacidade acarreta algumas necessidades em termos temporais e de recursos computacionais. A extração de informação relevante como preferências ou hábitos requer que quantidades volumosas de registos sejam analisadas. São necessários mecanismos de extração de conhecimento que através da aplicação de algoritmos sobre os registos, consigam detetar padrões frequentes e relações. Comparativamente às técnicas explícitas, a aplicação destes mecanismos torna o tempo de processamento maior e requer maiores recursos computacionais. Estas desvantagens dependerão naturalmente da quantidade de registos a tratar e da complexidade dos algoritmos que são aplicados.

As técnicas implícitas estão também mais interligadas com a incerteza, no sentido em que para definirem com alguma precisão que existe um padrão nos registos, terá de existir uma quantidade significativa de registos onde este se verifique. Num motor de busca de vídeos, se o sistema detetar que por uma vez o utilizador assistiu um vídeo de uma determinada categoria, não é correto deduzir que o utilizador tem interesse em outros vídeos desta categoria, até mesmo porque o utilizador poderá ter assistido o vídeo por engano. É necessário que exista uma quantidade considerável de registos onde o utilizador assistiu a vídeos da categoria. Por outras

palavras, a aplicação das técnicas implícitas e respetivos mecanismos de extração de conhecimento requerem que haja ações que são frequentemente realizadas pelo utilizador. Se o utilizador não for repetitivo nas suas ações não será possível encontrar padrões e mesmo aqueles que existam serão algo incertos.

A tabela 6 resume as principais comparações existentes entre estas duas formas de obter informação.

Tabela 6 - Comparação entre as técnicas explícitas e implícitas

	Necessidade de interação com o utilizador	Tempo necessário para obtenção de informação concreta	Recursos computacionais necessários para obtenção de informação	Risco de divergência da informação obtida com a real
Técnicas explícitas	Total	Pouco (Formulários)	Poucos	Alto
Técnicas implícitas	Nenhuma	Muito (Extração de Conhecimento)	Muitos	Baixo

Face às limitações que cada uma das técnicas apresenta, são também propostas por alguns projetos, abordagens híbridas que agrupam as vantagens de cada uma destas técnicas. Sites como o *Youtube* ou *Amazon* por exemplo, observam as pesquisas realizadas pelos utilizadores ao mesmo tempo que disponibilizam métodos explícitos para os utilizadores avaliarem os itens (vídeos e produtos respetivamente).

3.3 - Conclusões

Neste capítulo foi introduzido o conceito de *User Profiling* que lida com os perfis dinâmicos. Um dos seus objetivos é encontrar formas de representar e modular um perfil que garanta uma informação consistente e atualizada com a pessoa que representa. Para este primeiro objetivo, perante a grande quantidade de métodos existentes, as Redes Bayesianas, o Raciocínio Baseado em Casos e as Regras de Associação foram apresentadas como as que melhor relacionam-se com o domínio do trabalho. Para cada um destes foi descrito a sua

metodologia de funcionamento e apresentados exemplos de projetos onde foram ou são aplicados. No final da análise de cada umas das metodologias, foi discutido a sua capacidade em ser aplicada no trabalho e eventuais limitações ou cuidados que requerem.

A partir das discussões é possível agrupar o Raciocínio Baseado em Casos e as Regras de Associação como as mais aptas de serem utilizadas no trabalho em detrimento das Redes Bayesianas que apresentam limitações quanto à sua eficiência em atualizar os seus valores. Contudo, para o Raciocínio Baseado em Casos e as Regras de Associação existe um conjunto de limitações e cuidados que requerem estudos para uma aplicação correta e eficiente. Como será descrito no capítulo 5, as Regras de Associação revelam-se um pouco ineficientes no processo de aprendizagem, razão pela qual foi apresentada uma nova abordagem quanto à forma de representar um perfil.

O segundo objetivo do *User Profiling* já analisado, é determinar formas de recolher informação do utilizador para modular o seu perfil. Como já foi referido, estas são classificadas em explícitas ou implícitas. Foi descrito em que consistem e apresentada uma comparação entre as mesmas. No que toca à escolha da técnica a aplicar no trabalho, a implícita foi a preferida devido a duas razões. A primeira razão prende-se com a desvantagem que as explícitas apresentam no domínio do trabalho. Como já foi mencionado, um perfil representa também dependências entre os atributos das atividades, ou seja, para uma determinada pessoa o facto do local ser em determinado sítio poderá influenciar na sua preferência pela atividade. Assim, não é fácil para o utilizador informar de forma explícita estas dependências quando cria, aceita ou rejeita uma atividade. Da mesma forma, face à quantidade de dependências que pode haver entre os atributos, não é prático requisitar ao utilizador que avalie cada uma destas. A segunda razão deve-se à maior probabilidade que existe nas explícitas de haver uma divergência da informação recolhida em relação à real, facto que se apresenta menos provável na aplicação das implícitas.

Capítulo 4: iLeisure - Arquitetura

O sistema desenvolvido neste trabalho foi denominado de *iLeisure*. O nome deriva da sigla "i" sinónimo de inteligente e da palavra "*Leisure*" que traduz-se como Lazer. O *iLeisure* apresenta-se desta forma como um sistema inteligente que tem como objetivo recomendar atividades de lazer que não sejam sedentárias e que incentivem ao convívio social. O sistema é considerado inteligente na medida em que recomenda atividades que sejam das preferências de um utilizador, necessitando para isso de as aprender.

O sistema desenvolvido desdobra-se em dois componentes que foram separadamente estudados e implementados. O primeiro compreende uma aplicação para dispositivos móveis, com sistema operativo *Android* que emprega a interação de um utilizador com o sistema. É nesta aplicação que o utilizador agenda as suas atividades e recebe recomendações. Esta é também a fonte de informação sobre o utilizador que é usada pelo sistema para o processo de aprendizagem acerca das suas preferências.

O segundo componente é o centro de todo o sistema. É o servidor que tem a responsabilidade de representar os perfis dos seus utilizadores em relação às suas preferências, observar as suas ações e a partir destas aprender e adaptar os perfis. Com base no perfil de cada pessoa o servidor é também quem decide que atividades recomendar em determinadas situações.

Estas responsabilidades que o servidor detém são os principais problemas para os quais foi proposto uma solução. É no servidor que está aplicado todo o conhecimento estudado e retido no desenvolvimento deste trabalho, não querendo isto dizer que o desenvolvimento da aplicação cliente para os dispositivos móveis tenha tido menor importância do que o servidor. Pelo contrário é esta que contribuirá para a angariação de clientes do sistema, pelo o que foram encarados cuidados no seu planeamento e implementação

Neste capítulo é descrito em detalhe a arquitetura do sistema *iLeisure* e todos os aspetos que consideram-se importantes para o seu funcionamento. É também dada especial atenção à comunicação entre os dispositivos móveis e o servidor, e é descrito a aplicação móvel desenvolvida.

4.1 - Definição do Domínio

Conforme foi definido, o objetivo deste trabalho é recomendar atividades às pessoas que envolvam exercício físico e convívio social e que sejam da preferência das mesmas. A partir daqui, o conceito de atividade será substituído pelo de evento pelo facto de considerar-se como sendo o mais comumente relacionado com o conceito de agenda.

Para uma melhor interpretação do que é um evento, apresenta-se em seguida duas tabelas descritivas de todos os atributos característicos de um evento e o seu significado. A tabela 7 contém os atributos usados na representação de um perfil e que serão explicados no capítulo 5. A tabela 8 contém os restantes atributos que embora não sejam utilizados na representação de um perfil são usados para outros fins relativos às funcionalidades da aplicação desenvolvida para os dispositivos móveis.

Tabela 7 - Atributos de um evento utilizados na representação de um perfil

Atributo	Significado/Observações
Hora de Início	Indica a hora a que um evento tem início. É usada a codificação <i>UNIX</i> para representar um instante temporal como o número de milissegundos passados desde as 0 horas do dia 1 de Janeiro de 1970.
Hora de Terminó	Indica a hora a que um evento tem fim. É usada a codificação <i>UNIX</i> anteriormente definida.
Duração	Indica a duração do evento em minutos.
Dia da Semana	Indica o dia da semana em que o evento se realiza. É utilizado um número inteiro que é associado ao dia da semana, onde o número 0 representa Segunda-Feira e 6 representa Domingo.
Título	Indica o título do evento atribuído pelo autor.
Descrição	Uma descrição feita pelo autor sobre aspetos que estão relacionados com o evento.
Local	Indica o local do evento atribuído pelo autor.
Categoria	A categoria em que o evento se enquadra. Um utilizador ao criar um evento deverá escolher uma

categoria de um conjunto pré-definido (Ocupado, Outra, Desporto, Cultura, Convívio, Aventura). As categorias poderão num futuro ser alteradas ou adicionadas novas. A principal função do atributo Categoria é poder distinguir as atividades que não sejam sedentárias ou isoladas e que devem ser recomendadas.

Número de participantes A quantidade de participantes que estão envolvidos no evento. No caso de ser um evento realizado apenas pelo autor o seu valor é 1.

Tabela 8 - Restantes atributos de um evento

Identificador	Significado/Observações
Autor	O número de telefone do autor do evento.
Alarme	Uma indicação se o utilizador deverá ser notificado sobre a proximidade do evento.
Hora do Alarme	A hora em que o utilizador deverá ser notificado sobre a proximidade do evento. É usado a codificação <i>UNIX</i> .
Participantes	Uma lista contendo o número de telefone de todos os participantes envolvidos no evento.

4.2 - Arquitetura e Comunicação

A figura 26 ilustra a arquitetura do sistema que no seu todo apresenta-se como um sistema distribuído, composto por um servidor centralizado, um serviço da *Google* e por um conjunto de dispositivos móveis (*smartphones*) com sistema operativo *Android*.

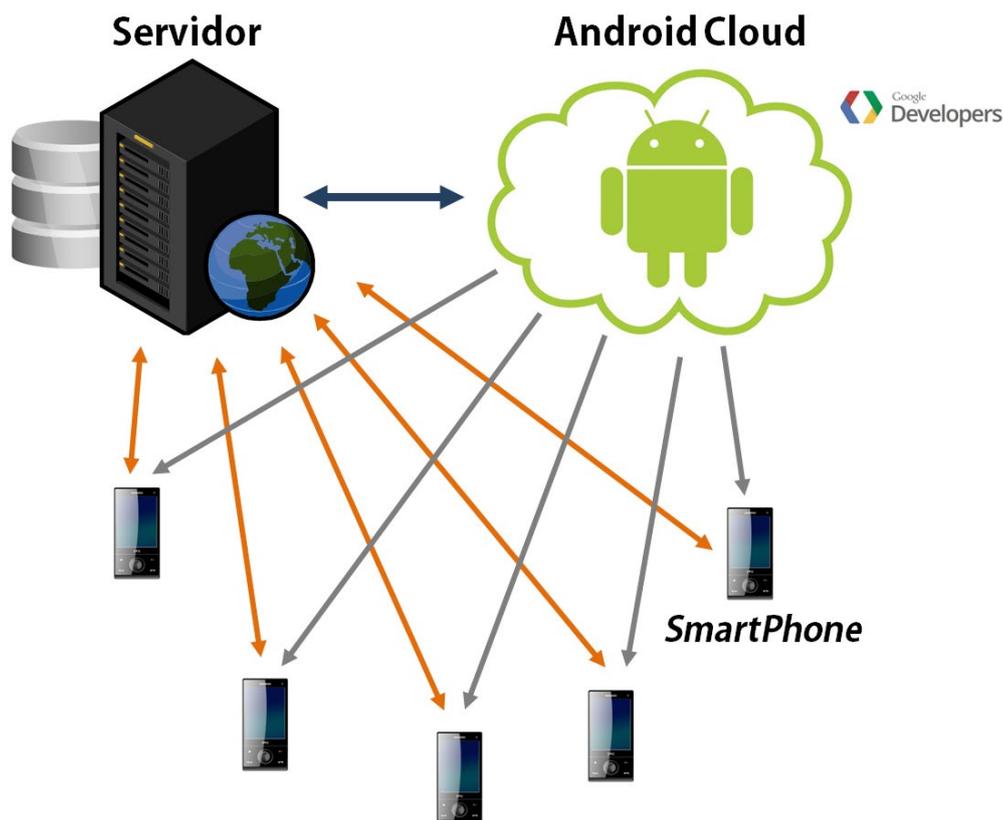


Figura 26 - Arquitetura do iLeisure

O servidor possui uma base de dados onde armazena de forma persistente, para cada telemóvel, todos os seus eventos (a sua agenda) e o seu perfil, sendo este último o mais importante para este trabalho. A forma como um perfil é armazenado e modulado pelo servidor é analisada com maior detalhe no capítulo 5. O ponto de interação do utilizador com o sistema é o telemóvel e mais especificamente a aplicação desenvolvida para tal fim. Como será possível verificar no capítulo 4.3, é através desta aplicação que o utilizador organiza a sua agenda, recebe as recomendações de novos eventos e notificações de convites para participar em eventos propostos pelos seus conhecidos.

Para cada utilizador, a sua agenda é replicada entre o telemóvel e o servidor, não só por questões de segurança contra falhas na aplicação, mas também pela necessidade do servidor em conhecer a agenda de cada utilizador, para assuntos relativos à recomendação de novos eventos. O servidor poderá também replicar os dados guardados por diversas bases de dados, para garantir que estes estejam sempre acessíveis. A aplicação desenvolvida para o telemóvel, guarda internamente uma réplica da agenda existente no servidor, diferenciado no facto de guardar e apresentar ao utilizador apenas os eventos para semana atual. Esta é uma forma de

ter em consideração as limitações inerentes aos telemóveis e à sua capacidade de armazenamento. Assim, no início de cada semana, todos os eventos da semana anterior são eliminados do telemóvel, mas não da base de dados que continua a guarda-los. A razão de os manter na base de dados, prende-se com o facto de serem elementos fundamentais para a modulação do perfil do utilizador, conforme explicar-se-á, com maior detalhe no capítulo 5.

É fundamental para qualquer replicação existir comunicação entre os locais onde os dados são replicados, neste caso entre cada um dos telemóveis e o servidor. É imperativo que todas as operações executadas pelo utilizador na aplicação sejam comunicadas ao servidor, para que este sincronize a agenda armazenada na base de dados, com a armazenada na aplicação. Se tal não acontecer, quando for necessário recomendar um novo evento ao utilizador, poderá ocorrer conflitos entre eventos. A comunicação entre o servidor e os telemóveis necessita de ser bidirecional, ou seja, deverá permitir que tanto o servidor como o telemóvel possam invocar a troca de informação. A aplicação, como já referido anteriormente, comunicar ao servidor as operações que são realizadas pelo utilizador, enquanto o servidor deverá comunicar ao utilizador e ao seu telemóvel novos eventos recomendados ou propostos pelos seus conhecidos.

A solução mais comum nestas situações é utilizar canais de comunicação (*sockets*) entre o servidor e cada um dos telemóveis. As limitações dos telemóveis em termos de energia, de custo de comunicação de rede e facilidade de troca de identidade (IP) fazem com que esta solução seja pouco eficiente. Manter um canal aberto consome energia e tráfego de dados, e é também muito delicado devido ao facto de existir uma probabilidade muito alta do telemóvel mudar constantemente de identidade e com isso ser necessário criar novos canais.

Desta forma a opção implementada usa dois métodos de comunicação, um em que é o cliente a comunicar ou a requerer informação ao servidor e o segundo em que é usado um serviço proprietário da *Google*, para comunicar com os telemóveis. Em ambos os métodos de comunicação é usado a linguagem *JSON* para estruturação de dados a transmitir. Para além de definir-se como uma forma simples e eficiente de estruturar um conjunto de dados, o *JSON* é também nativamente suportado pelo *Android*. Este é também cada vez mais utilizado em alternativa ao convencional *XML*, pelo facto de permitir uma interpretação mais eficiente do seu conteúdo. Apresenta-se na figura 27 um exemplo de um objeto *JSON* representativo de um evento. Para operações como a criação ou edição de eventos em que é necessário transferir todos os atributos de evento, é criado um objeto *JSON* que estrutura todos os atributos de uma forma compacta e eficiente. Também em algumas situações onde seja apenas necessário

transferir um ou dois atributos é usado o *JSON* pela forma simples com que permite estruturar os dados. Exemplificando, para o caso onde um determinado utilizador aceita um convite para um evento, o autor do evento será notificado com uma mensagem contendo um *JSON* similar ao ilustrado na figura 28.

```
{
  "dia": "0",
  "horainicio": "16:00",
  "horafim": "17:00",
  "duração": "60",
  "categoria": "Desporto",
  "titulo": "Futebol",
  "descrição": "Jogar futebol com colegas",
  "local": "Campo Universidade do Minho",
  "participantes": [
    "961234564",
    "934567896"
  ]
}
```

Figura 27 - Estruturação JSON para um evento

```
{
  "idEvento": "3",
  "idUtilizador": "964567896",
  "aceitação": "sim"
}
```

Figura 28 - Estruturação JSON para uma notificação

De corroborar na figura 27 que os participantes de um evento são identificados pelos seus números de telefone e não pelos seus nomes. Esta condição deve-se não só ao facto dos números funcionarem como um identificador único dos utilizadores e dos seus telefones, mas também, devido ao facto de diferentes participantes de um evento poderem associar diferentes nomes aos números. Exemplificando, para uma situação onde um utilizador João crie um evento e convide uma pessoa Joana cujo número é o 1234 e uma outra pessoa Pedro, embora o utilizador João associe a pessoa Joana ao número 1234, a pessoa Pedro poderá associar o número 1234 com o nome Martins Joana. Ao passar o número de telefone como atributo, cada

telefone interpreta os números contidos no objeto *JSON* e procura na sua lista de contactos o respetivo nome.

Em relação à primeira forma de comunicação, esta é representada na figura 26 pelas setas laranja, as quais indicam os pedidos que a aplicação faz ao servidor. Estes pedidos poderão corresponder a operações como criação, edição, eliminação, aceitação, recusa ou obtenção de eventos. O servidor apresenta-se como um serviço na internet que recebendo pedidos dos clientes, processa-os e responde, não possuindo por isso qualquer capacidade de iniciar uma comunicação com um telemóvel. A este tipo de servidores é atribuída a denominação *RESTful*.

O pedido de uma aplicação ao servidor é conseguido através do protocolo de comunicação *HTTP*, tendo cada uma das operações um endereço próprio e conhecido previamente pela aplicação. Assim e a exemplo ilustrativo, aquando da criação de um novo evento, a aplicação realiza um pedido ao endereço <http://www.ileisure.com/create> usando o método de pedido *POST*, pelo qual é enviado o objeto *JSON* com os todos os atributos do evento. Em operações onde não seja necessário passar quantidades volumosas de dados é o usado o método de pedido *GET*. Em operações como a eliminação de um evento, o pedido é feito através um endereço simples como <http://www.ileisure.com/delete?id=3>.

Relativamente ao segundo método de comunicação, este é exemplificado na figura 26 pelas setas cinza que representam o envio de mensagens do serviço de comunicação da *Google* para os telemóveis. Este serviço é denominado por *Android Cloud to Device Messaging (C2DM)* e apresenta-se como um serviço que assiste na troca de informação de um servidor para uma aplicação *Android*. Este serviço disponibiliza mecanismos leves e eficientes que permitem a transferência de informação para um telemóvel, ao mesmo tempo que lida com eventuais problemas de comunicação, como filas de mensagens.

O seu princípio de funcionamento é usar o canal de comunicação com os servidores da *Google*, próprio de todos os telemóveis *Android*. Para o serviço conseguir comunicar com um telemóvel é usado a conta *Gmail* ativa no telemóvel e outros identificadores de sessão. Para o serviço *C2DM* conseguir comunicar com uma aplicação do telemóvel, a aplicação terá de requisitar ao *C2DM* a obtenção de uma chave que ser-lhe-á atribuída uma vez realizada a autenticação, e que servirá de identificação para o endereçamento de mensagens para o telefone. Assim, para além de guardar o perfil e agenda de um utilizador, o servidor regista a chave que lhe foi atribuída pelo serviço *C2DM*.

Uma vez que este processo é obrigatório para o envio de mensagens para a aplicação, a aplicação possui um serviço que é inicializado quando o telemóvel liga-se e que tem como objetivo obter uma nova chave e comunicar com o servidor a sua atualização. Por questões de segurança, esta chave é renovada ao fim de um certo tempo. Aquando da ocorrência desta situação, a aplicação comunica novamente ao servidor a nova chave para que este a atualize. A figura 29 sumariza as etapas subjacentes a este processo de registo no serviço *C2DM* e no servidor.



Figura 29 - Registo no *C2DM* e servidor

No que toca, mais precisamente, a situações onde o servidor pretenda comunicar informações com um telemóvel (notificações), é conseguido com recurso ao serviço *C2DM* e é apresentado na figura 26 pela seta azul. Estas notificações compreendem a recomendação de novos eventos, convites de outros utilizadores para participar em eventos ou a resposta de outros utilizadores a convites. Quando o servidor pretenda enviar uma mensagem para um telemóvel, este realiza um pedido ao serviço *C2DM* indicando a mensagem e a chave do utilizador. O *C2DM* irá por sua vez enviar a mensagem para o telemóvel associado à chave fornecida.

Um dado a ter em conta é que este serviço tem como objetivo servir de ponte entre um servidor e uma aplicação instalada num telemóvel. O objetivo é que um servidor quando pretenda enviar dados para um telemóvel atribua uma identificação aos dados e envie esta identificação para o serviço *C2DM* que por sua vez a enviará para o telemóvel. Quando a aplicação no telemóvel receber a mensagem do serviço *C2DM* deverá usar a identificação contida na mensagem para fazer um pedido ao servidor pelos dados. Não é por isso aconselhável que a mensagem que o servidor envia para o serviço *C2DM* seja um objeto *JSON* como o ilustrado na figura 27.

Exemplificando, quando é necessário recomendar um evento ao utilizador, o servidor cria o evento e atribui-lhe um identificador único. Em seguida faz um pedido ao serviço *CD2M* enviando um objetivo *JSON* simples, contendo apenas o identificador do evento. Quando a aplicação no telemóvel recebe a notificação utiliza o identificador para fazer um pedido ao servidor pelos atributos do evento que ser-lhe-ão transmitidos sobre um objeto *JSON*. Para os restantes casos onde seja necessário a comunicação do servidor para um telemóvel é usado o mesmo sistema. Para uma informação mais específica e técnica sobre o funcionamento do serviço *C2DM*, aconselha-se a leitura da sua página de suporte disponibilizada em <https://developers.google.com/android/c2dm/>.

A figura 30 exemplifica o envio de uma mensagem do servidor para um telemóvel e as respetivas etapas necessárias e descritas anteriormente.

A comunicação entre o telemóvel e o servidor requer que o telemóvel esteja constantemente conectado à internet, para enviar e receber informações. No desenvolvimento deste trabalho não foram consideradas situações onde o telemóvel não possuía comunicação com a internet. Quando um utilizador realiza uma operação no telemóvel que requeira comunicação com o servidor e em que o telemóvel não a disponha, a aplicação indicará a ocorrência de uma falha depois de realizar várias tentativas.

Em termos práticos, se um utilizador criar um novo evento e não possuir internet no telemóvel, a aplicação tentará durante algum tempo comunicar com o servidor, esgotado esse tempo indicará a ocorrência de um erro e não criará nem guardará o novo evento no telemóvel. Esta situação deve-se à necessidade da agenda, mantida no telemóvel, estar constantemente sincronizado com o servidor. Se não estiver poderão ocorrer conflitos entre eventos quando o servidor recomende novos eventos, uma vez que a agenda mantida no servidor é usado para encontrar tempos livres. É por esta razão que o resultado de uma operação na aplicação, no que

toca a termos gráficos e de armazenamento, apenas são executados quando o servidor responde positivamente ao pedido realizado.



Figura 30 - Envio de mensagem do servidor para um telemóvel

É contudo objeto de trabalho futuro implementar formas de contornar esta situação, tendo já sido propostas soluções como por exemplo a utilização filas de espera no telemóvel (*cache*). Por outras palavras, guardar qualquer operação que seja realizada pelo utilizador quando não haja conexão com a internet e pela mesma ordem. Quando o telemóvel obtenha uma conexão, são enviadas todas as operações em lista de espera e pela mesma ordem (a ordem de execução no servidor é um ponto muito importante para garantir a consistência e sincronização entre o telemóvel e o servidor, pelo que tem de ser estudadas técnicas eficientes de replicação em sistemas distribuídos).

4.3 - Aplicação Cliente

Como já foi abordado, o telemóvel é ponto de interação do utilizador com o sistema. Por isso foi estudada e implementada uma aplicação para telemóveis com sistema operativo *Android* que complementa uma agenda semanal comum, com funcionalidades uteis como a recomendação de eventos e a definição de eventos envolvendo outros participantes.

A avaliação feita da estruturação, funcionalidades e utilidade da aplicação foi positiva. Para além das funcionalidades atualmente implementadas, está prevista a adição de outras funcionalidades que venham a expandir a funcionalidade e utilidade da aplicação. Acredita-se que a aplicação desenvolvida tem postura para numa situação futura ser comercializada.

A interface principal da aplicação é ilustrada pela figura 31. Como é possível verificar, a aplicação apresenta-se como uma agenda semanal na qual o utilizador poderá agendar eventos. A interface divide-se em três componentes. O primeiro é o cabeçalho onde é apresentado os dias da semana. O segundo é um componente deslizante que contem a agenda de eventos para todos os dias da semana, deste as 8 horas até às 24 horas. O último componente (caixa de texto inferior) apresenta informação acerca da data, hora de início, hora de término, título e descrição do evento selecionado. Os eventos são representados por retângulos com uma cor e símbolo associado à categoria a que pertencem. O comprimento de cada retângulo varia consoante a duração do evento a que está associado. A posição de cada retângulo é também definida em função da hora de início, de modo a estar paralelo à barra vertical das horas. O símbolo, no canto superior esquerdo da interface, representa o número de eventos criados por conhecidos para os quais o utilizador foi convidado a participar e para os quais ainda não respondeu.

Os eventos que são recomendados pelo servidor e que o utilizador não aceitou ou recusou, ou seja, que são novidade, são apresentados com uma estrela amarela. Uma vez aceite ou editado o evento passa a ser ilustrado como um evento normal. Os eventos contendo um símbolo de uma pessoa e um símbolo de negação representam eventos onde um ou mais participantes rejeitaram o convite. Quando o utilizador consulte os dados do evento e remova os participantes que não podem comparecer, o evento passará a ser ilustrado como um evento normal.

Tendo descrito a interface principal, será agora feita uma descrição das funcionalidades do sistema, possíveis de serem acedidas e usufruídos pelo utilizador.

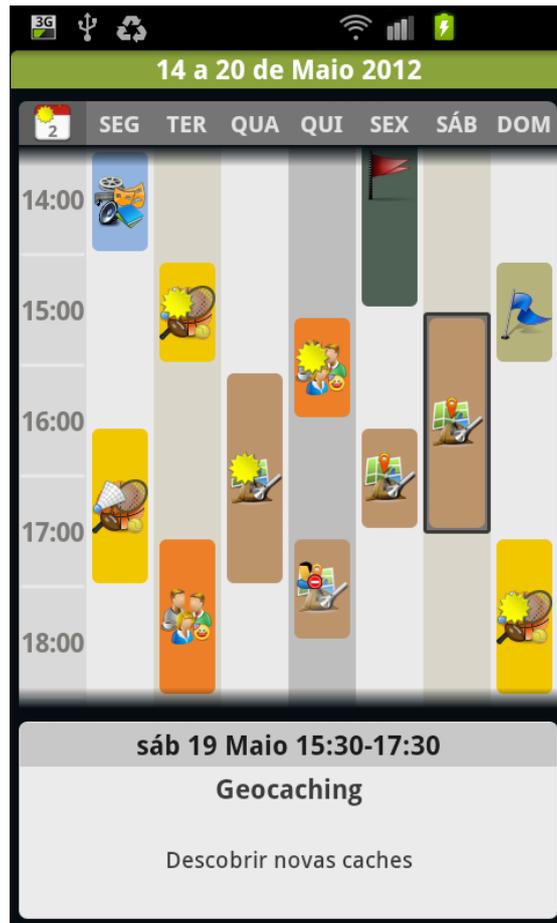


Figura 31 - Interface principal da aplicação móvel

Cada um dos eventos representados pelos retângulos são interativos. Um toque sobre um evento define-o como o atualmente selecionado e fará com que alguma informação acerca deste seja apresentada na caixa de texto inferior da interface (terceiro componente). Um toque longo sobre um evento apresenta um menu de contexto com as opções possíveis de realizar com este e que são ilustradas pela figura 32. A opção "Aceitar Sugestão" apenas é visível em eventos recomendados pelo servidor e que ainda não foram aceites pelo utilizador.



Figura 32 - Menu de contexto para um evento

Se um utilizador pretender obter toda a informação de um evento, um toque na caixa de texto inferior fará aparecer um painel deslizante contendo toda a informação do evento selecionado (figura 33).



Figura 33 - Painel com informação completa de um evento

A aplicação disponibiliza ainda menus de contexto para a edição rápida de eventos que ocupem um dia na totalidade ou eventos que sejam repetidos durante toda a semana. Tocando longamente sobre um dia da semana no cabeçalho, fará aparecer um menu de contexto como o ilustrado na figura 34. Tocando longamente sobre uma hora na barra vertical que apresenta os horários da semana (à esquerda da interface), fará aparecer um menu de contexto como o ilustrado na figura 35.

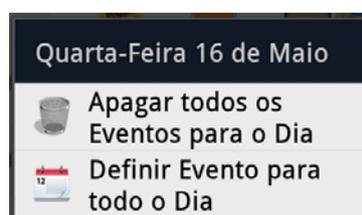


Figura 34 - Menu de contexto para um determinado dia

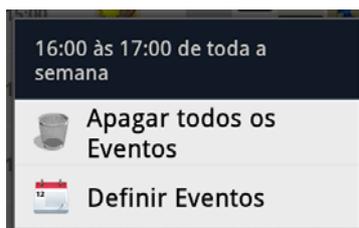


Figura 35 - Menu de contexto para um determinado horário

A aplicação permite que o utilizador, aquando da criação ou edição de um evento, selecione um alarme para que seja notificado sobre a proximidade do evento. Quando o evento estiver próximo de ocorrer a aplicação notificará o utilizador conforme é ilustrado pela figura 36.

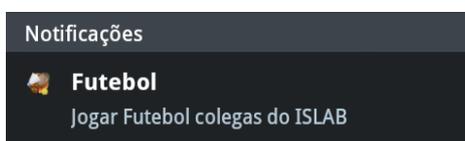


Figura 36 - Notificação da proximidade de um evento

Tocando sobre a notificação fará com seja apresentado um menu deslizante com toda a informação do evento (figura 37).



Figura 37 - Menu acionado pela notificação com a informação do evento

A aplicação possui ainda um menu principal que disponibiliza outras funcionalidades (figura 38).



Figura 38 - Menu principal da aplicação

A principal opção é a que permite a criação de novos eventos. Quando é feita a sua seleção, um novo menu como o ilustrado na figura 39 será apresentado ao utilizador para que este introduza todos os dados do novo evento. A adição de participantes a um evento é feita com recurso à lista telefónica. A remoção de um contacto é conseguida com um toque longo sobre este.



Figura 39 - Menu de criação de um evento

A aplicação dispõe ainda da capacidade de detetar e resolver conflitos que possam surgir com o novo evento (figura 40). Quando acontece esta situação, o utilizador é avisado e poderá tomar uma decisão. No caso de pretender continuar o sistema resolverá os conflitos que ocorreram, dando prioridade ao novo evento. A figura 41 apresenta um exemplo de uma situação, onde o sistema resolve um conflito quando um novo evento na categoria desporto é criado.

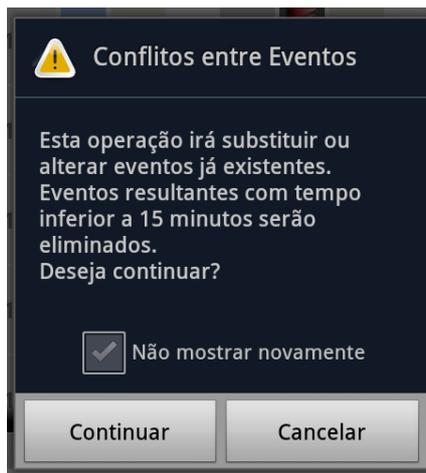


Figura 40 - Menu de aviso da existência de conflitos com o novo evento

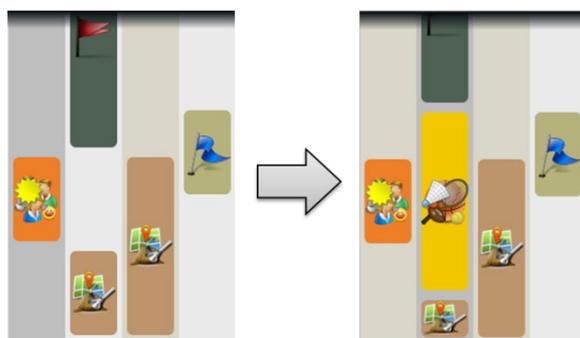


Figura 41 - Exemplo de resolução de conflitos

A opção "Apagar Eventos" apresenta-se como um menu semelhante ao de criação de um evento, mas sem os campos para introdução de dados. Apenas é apresentado os dias da semana, hora de início de hora de finalização. A opção "Preferências" apresenta um menu com opções da aplicação que podem ser ativados ou desativados pelo utilizador.

A seleção da opção "Eventos por Amigos" apresenta ao utilizador um menu deslizante com todos os eventos para os quais foi convidado por amigos (figura 42). Para cada evento é apresentado a informação de quem é o seu ator, a data, hora e descrição. A aplicação deteta também os conflitos que surgem com os eventos. Eventos que não entrem em conflito com agenda atual do utilizador são apresentados sobre um fundo verde, ao contrário dos que coincidem com a agenda, que são apresentados com um fundo vermelho.



Figura 42 - Menu de apresentação de eventos propostos por amigos do utilizador

Pressionando sobre os eventos é apresentado um menu similar ao da figura 8 que contém toda a informação do evento. Um toque longo sobre um evento apresenta ao utilizador um menu de contexto com as opções que este pode tomar e que são ilustradas na figura 43. A informação "Sobrepôr existentes" e a opção "Comparar com Eventos Agendados" apenas aparecem em eventos que estejam em conflito com a agenda.



Figura 43 - Menu de contexto para os eventos propostos por amigos

A seleção da opção "Comparar com Eventos Agendados" apresenta ao utilizador um menu para que este verifique de que forma o evento coincide com outros já agendados (figura 44).



Figura 44 - Menu com informação acerca de como o evento colide com os demais

Caso o utilizador recuse participar no evento, é apresentado um menu em que este poderá informar ao autor do evento a razão da sua ausência (figura 45).

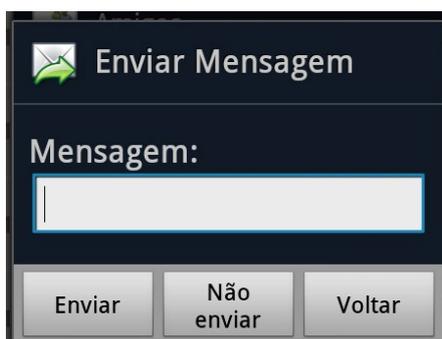


Figura 45 - Menu para o utilizador informar a razão da sua ausência em eventos

Quando um utilizador cria um evento e convida amigos, se estes tiverem a aplicação instalada, receberão instantaneamente uma notificação sobre a existência de novos eventos em que foi solicitada a sua participação (figura 46).

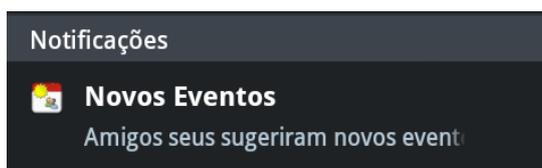


Figura 46 - Notificação de novos eventos propostos por conhecidos

O autor de um evento recebe notificações instantâneas da participação das pessoas que convida para um evento (figura 47 e 48).



Figura 47 - Notificação de confirmação da presença de participantes

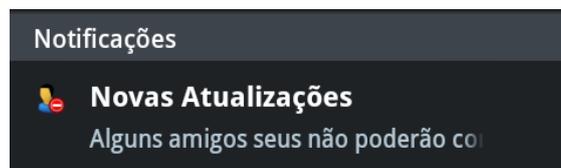


Figura 48 - Notificação de ausência de participantes

Quando visualiza informação sobre um evento, o utilizador é informado da participação dos seus intervenientes (figura 49). Participantes que aceitem o convite serão apresentados sobre um fundo verde. Participantes que recusem o convite serão apresentados sobre um fundo vermelho. Se o utilizador tocar sobre estes será apresentada a mensagem com a justificação da ausência (figura 50). Participantes que não confirmem ou recusem a presença são apresentados sobre um fundo branco.



Figura 49 - Informação acerca da presença dos participantes

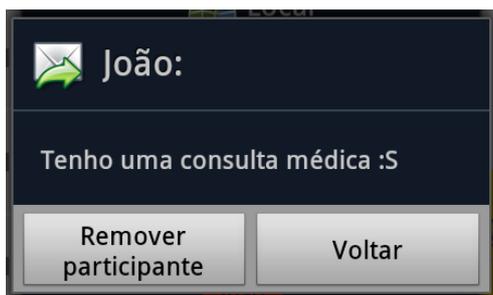


Figura 50 - Mensagem de justificação de ausência num evento

4.4 - Conclusões

Crê-se que a aplicação desenvolvida é um estímulo para a prática de atividades físicas, nos tempos livres das pessoas, que contribui para a diminuição do sedentarismo. Sendo os dispositivos móveis algo indispensável e cada vez mais utilizado, uma aplicação que permita ao utilizador agendar os seus tempos livres é algo interessante e cativante. Esta aplicação incentiva também a socialização, através do agendamento de eventos coletivos.

Contudo, é necessário algo mais que incentive um utilizador a praticar atividades físicas nos seus tempos livres, sendo por isso proposto um sistema de recomendação. Para que a taxa de aceitação dos utilizadores, perante as atividades sugeridas seja alta, é necessário usar informação acerca das suas preferências, na seleção das atividades a recomendar. A este processo é associado o conceito de Personalização ou *User Profiling* que consiste na obtenção e modulação dos perfis dos utilizadores. No capítulo seguinte analisa-se a Personalização detalhadamente e a forma como lida com a representação e aprendizagem de perfis.

Capítulo 5: iLeisure - Personalização

O servidor é o cérebro do sistema *iLeisure*. Para qualquer operação executada pelo utilizador é da responsabilidade do servidor garantir que a agenda guardada na base de dados, está sincronizada com a armazenada no telemóvel do utilizador.

O servidor é também o responsável por garantir que o perfil de cada pessoa seja dinâmico, se mantém atualizado e consistente com o verdadeiro perfil da pessoa, ao longo da sua vida. Para o conseguir é imperativo que para cada operação realizada pelo utilizador, o servidor determine de que forma deverá o perfil da pessoa ser adaptado.

A recomendação de eventos é também da responsabilidade do servidor. É ele que procura por tempos livres na agenda de cada utilizador e determina que evento deverá recomendar, de modo a que o evento recomendado seja da preferência do utilizador, dadas as circunstâncias onde será praticado.

Todas as necessidades e responsabilidades requisitadas pelo servidor, foram determinantes para que o objetivo da sua implementação fosse aquele que mais estudos e análises requisitou. O levamento do estado da arte, apresentado nos capítulos 2 e 3, foi especialmente vocacionado para suportar o desenvolvimento de um servidor inteligente e eficiente. Contudo, a sua complexidade não é sinónimo de que menor tempo de estudo e implementação tenha sido despendido para o desenvolvimento da aplicação para o dispositivo móvel, apresentado anteriormente como sendo uma aplicação robusta e funcional.

O servidor é que toma a iniciativa de procurar por tempos livres e recomendar eventos para a agenda dos utilizadores da aplicação móvel. Por essa razão foi estudada e implementada uma forma que permitisse ser o servidor a iniciar a comunicação com os telemóveis. O servidor apresenta desta forma propriedades que o permite classificar como sendo mais inteligente e robusto do que um servidor *RESTful* convencional.

Face a todas as características e capacidades que o servidor apresenta, é unânime declarar que é este que faz com que o sistema *iLeisure* seja qualificado como inteligente.

Nos próximos três subcapítulos é feito uma análise detalhada de como o servidor lida com a representação de um perfil, como é que este adapta um perfil face aos comportamentos dos seus utilizadores e de que forma recomenda eventos para os tempos livres.

5.1 - Representação de um Perfil

Antes de indicar de que forma o perfil de cada utilizador é representado, é conveniente indicar especificamente a informação que este tem de representar. De forma abstrata o objetivo de um perfil é representar as preferências de uma pessoa em praticar determinados eventos, para que quando haja necessidade de recomendar um evento, este seja o mais semelhante com as preferências da pessoa.

Numa primeira proposta de solução, o perfil de uma pessoa seria representado unicamente através do título dos eventos e, associado a cada um destes, um valor probabilístico sinónimo da preferência do utilizador por praticar esse evento, ou por outras palavras, representativo da frequência com que o utilizador pratica eventos com esse título. Esta solução pode ser ilustrada na tabela 9.

Tabela 9 - Solução inicial para a representação de um perfil

Atividade	Probabilidade
Futebol	0.70
Natação	0.45
Cinema	0.20

Porém como já foi várias vezes salientado, o perfil de uma pessoa não se rege a condições tão simples. A vida de uma pessoa está cheia de dependências e é necessário que estas dependências sejam também representadas no seu perfil. No que diz respeito às preferências pelos eventos, a preferência de uma pessoa em praticar o evento Cinema poderá depender de certas condições que se verifique. Enquanto certas pessoas gostarão mais de ir ao Cinema nas sessões da tarde, outras gostarão de ir às da noite. Cada uma destas pessoas terá um valor probabilístico de preferência pelo evento Cinema que diferenciar-se-á em função da hora em que é realizado.

O próprio sistema necessita de conhecer as dependências de um perfil das pessoas, para a recomendação de eventos para os tempos livres. O sistema infere sobre que preferências tem o utilizador por eventos a praticar, num determinado tempo livre, e os atributos como a hora de início ou o dia da semana influenciam sobre a escolha do evento a recomendar.

Pode-se desta forma verificar que um perfil requer que todas as dependências possíveis de gerar sejam representadas. Considerando todos os atributos existentes para um evento, é necessário representar todas as relações que possam existir entre estes e para cada uma das relações o valor probabilístico indicativo da sua frequência, ou por outras palavras, o valor representativo da preferência do utilizador. De todos os atributos de um evento descritos no capítulo 4.1, apenas os primeiros atributos da tabela 7 foram considerados como sendo necessários de representar nas relações.

Contudo, as relações não deverão apenas compreender os atributos como um todo, uma vez que cada atributo possui vários valores possíveis e para cada valor poderá existir uma relação diferente e um valor diferente de probabilidade. Exemplificando, não é correto que seja representado no perfil uma relação entre o atributo Título e Local na sua globalidade. O correto é representar uma relação entre cada um dos seus valores possíveis. Por exemplo, deverá ser representado uma relação entre o atributo Título possuir o valor "Futebol" e o atributo Local possuir o valor "Universidade". O mesmo deverá se verificar para cada uma das combinações possíveis entre os valores de cada um dos atributos. Numa relação poderão estar presentes mais do que dois atributos, assim, relações como o atributo Título possuir o valor "Futebol", o atributo Local possuir o valor "Universidade" e o atributo Hora possuir o valor "16" necessitam também de ser representadas.

Quando houve a necessidade de lidar com relações de implicação, a primeira solução proposta foi utilizar as regras de associação. Nesta situação, um perfil representa-se por um conjunto de regras como por exemplo:

$$[Local = Universidade, Hora = 16] \rightarrow [Título = Futebol] \quad (R1)$$

$$[Título = Andebol] \rightarrow [Hora = 10] \quad (R2)$$

$$[Dia = Segunda, Hora = 9] \rightarrow [Categoria = Desporto] \quad (R3)$$

O número exato de regras que um perfil contém não é calculável face ao facto de existirem atributos cujo domínio é ilimitado (Título, Local, Descrição). Para cada uma das regras existe contudo um valor de suporte e de confiança.

O problema que existe em utilizar as regras de associação deve-se ao número de regras que são possíveis de gerar e que necessitam de ser analisadas, a cada operação realizada pelo

utilizador. Considerando-se a operação onde um utilizador cria um novo evento, este evento terá a si associado um conjunto de regras que necessitarão de ser analisadas e atualizadas no perfil. Neste caso, como cada atributo do evento apenas conterá um valor, o número de regras subjacentes a um evento ($\#r$) é calculável pela equação E1 onde n representa o número de atributos:

$$\#r = \sum_{i=1}^n \binom{n}{i} \times (2^{n-i} - 1) \quad (E1)$$

Considerando que o número de atributos necessário de representar numa relação é 9, isto significa que para cada novo evento criado pelo utilizador são geradas 18 660 regras de associação. Para cada uma das 18 660 regras será necessário atualizar o seu valor de suporte e confiança no perfil. O cálculo destes valores não é simples uma vez que é feito em função de casos passados similares, como será explicado no capítulo 5.2. Como já foi mencionado no estado da arte sobre as regras de associação, podem ser aplicados filtros para que as regras, cujo valor de suporte ou confiança seja inferior ao valor que se considera mínimo, sejam desprezadas. No entanto, será necessário que para cada uma das regras se faça uma consulta à base de dados para obter os valores atuais de suporte ou confiança. Foram realizados alguns testes com esta primeira solução que vieram provar a sua ineficiência. Cronometrando o tempo médio de processamento necessário para atualizar o perfil de um utilizador na criação de um novo evento, verificou-se que é em média de 37 minutos.

Face ao demasiado tempo indispensável pelas regras de associação, foi proposto e implementado a solução final. O objetivo foi encontrar uma forma que representasse as relações, mas cujo tempo de processamento fosse inferior, ou melhor, o objetivo foi encontrar um ponto de equilíbrio entre um perfil representar uma vasta quantidade de informação e ser eficiente em termos de processamento. Em seguida é apresentada a solução encontrada para representar os perfis e respetivas propriedades e características.

Seja C um conjunto de atributos e seus respetivos valores que ocorrem conjuntamente num evento e representado por:

$$C = \{ A_1 = v_1, A_2 = v_2, \dots, A_n = v_n \}$$

tal que a cardinalidade de C é maior do que 0, ou seja, o conjunto vazio não é considerado. Um conjunto C é interpretado como a relação que ocorre entre os seus elementos e possui, um valor percentual de frequência indicativo da frequência com os seus elementos ocorrem conjuntamente (que traduz-se na preferência do utilizador pelo conjunto) e um valor temporal indicativo da última vez que o valor de frequência foi atualizado.

Considere-se como exemplo os seguintes conjuntos:

$$C_1 = \{ Local = Universidade, Hora = 16, Título = Futebol \} \quad (C1)$$

$$C_2 = \{ Local = Universidade, Título = Futebol \} \quad (C2)$$

O conjunto C1 indicará através do seu valor de frequência, a frequência (preferência) com que um utilizador pratica o evento Futebol, na Universidade e às 16 horas. Por outras palavras, representa a frequência com os atributos "Local=Universidade", "Hora=16" e "Título=Futebol" ocorrem conjuntamente nos eventos realizados pelo utilizador.

O perfil de um utilizador é definido como um agregado de conjuntos que representam as suas preferências por cada um dos conjuntos, ou seja, as preferências do utilizador pelas várias combinações de atributos dos eventos. Relativamente ao número total de conjuntos que um perfil conterà, este dependerá do número de conjuntos possíveis de gerar por cada evento. Para cada evento são possíveis de calcular 511 conjuntos, valor determinado em função nos 9 atributos que são considerados nas relações e que é calculado em função da equação:

$$Número\ de\ Conjuntos(Evento) = 2^n - 1$$

onde n representa o número de atributos subjacentes ao evento e que são 9. Isto significa que no máximo serão necessárias 511 iterações por uma operação realizada pelo utilizador, que serão necessários de analisar 511 conjuntos subjacentes a um evento no processo de aprendizagem do perfil. Este número representa uma redução substancial em relação ao número de iterações requeridas pelas regras de associação, o que se traduz numa redução considerável do tempo de processamento necessário para a aprendizagem de um perfil. Com estudos realizados, calculou-se que o tempo médio para um processo de aprendizagem é agora em média 40 segundos.

Todavia, as relações definidas pelas regras de associação não são descartadas na solução atual, uma vez que os conjuntos sumarizam estas regras. Para um conjunto como o C2, este sumariza um agregado de regras de associação e que são em seguida indicadas:

$$[Local = Universidade] \rightarrow [Título = Futebol] \quad (R4)$$

$$[Título = Futebol] \rightarrow [Local = Universidade] \quad (R5)$$

Este número de regras deriva da cardinalidade do conjunto e é calculado em função da equação E1 onde a cardinalidade é representada por n . Assim, aplicando a mesma equação ao conjunto C1 serão possíveis gerar 12 regras.

Se considerarmos as regras R4 e R5, estas são contraditórias entre si. Enquanto a regra R4 indica que o facto de o local ser a Universidade implica sobre a escolha do evento, a regra R5 indica que o facto de o título ser Futebol implica que o local seja a Universidade. Contudo, é possível determinar qual das regras é a mais plausível usando para isso a frequência dos conjuntos representados no antecedente e consequente das regras. Neste caso é possível determinar em função da frequência dos conjuntos:

$$C_3 = \{ Local = Universidade \} \quad (C3)$$

$$C_4 = \{ Título = Futebol \} \quad (C4)$$

Estes conjuntos têm a particularidade de possuírem um único atributo, não se podendo desta forma considerá-los como representativos de uma relação, pelo que poderiam ser descartados. Contudo, como será demonstrado, estes conjuntos são uteis para tratamento de regras contraditórias.

Considere-se o teorema de *Bayes* que relaciona as probabilidades condicionais de dois eventos com as probabilidades incondicionais de cada evento através da seguinte expressão:

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

Admitindo que o evento A é associado ao conjunto C_3 , o evento B associado ao conjunto C_4 , as suas probabilidades $P(A)$ e $P(B)$ associadas com os valores de frequência de cada conjunto ($f(C_3)$ e $f(C_4)$ respetivamente), $P(A|B)$ associado à regra R4 e $P(B|A)$ associado à regra R5, através desta equação pode-se deduzir que em relação às regras R4 e R5 a mais plausível será:

$$\left\{ \begin{array}{ll} Local = Universidade \rightarrow T\u00edtulo = Futebol & se f(C_4) > f(C_3) \\ T\u00edtulo = Futebol \rightarrow Local = Universidade & se f(C_4) < f(C_3) \\ Indeterminado & se f(C_4) = f(C_3) \end{array} \right.$$

O mesmo racioc\u00ednio aplica-se nos demais conjuntos e respetivas regras contradit\u00f3rias. Exemplificando, considere-se as seguintes regras contradit\u00f3rias e subjacentes ao conjunto C1:

$$[Local = Universidade, Hora = 16] \rightarrow [T\u00edtulo = Futebol] \quad (R6)$$

$$[T\u00edtulo = Futebol] \rightarrow [Local = Universidade, Hora = 16] \quad (R7)$$

Para a determina\u00e7\u00e3o de que regra \u00e9 mais plaus\u00edvel o c\u00e1lculo \u00e9 similar ao j\u00e1 apresentado, diferenciando-se apenas pelo facto de usar o valor de frequ\u00eancia para um conjunto com cardinalidade dois:

$$C_5 = \{ Local = Universidade, Hora = 16 \} \quad (C5)$$

\u00c9 poss\u00edvel que a partir de dois conjuntos diferentes sejam deduz\u00edveis iguais regras. O conjunto C1 e C2 t\u00eam em comum o facto de ser poss\u00edvel de gerar a regra R4. Nestas condi\u00e7\u00f5es \u00e9 a regra gerada a partir do conjunto C2 que dever\u00e1 ser mais plaus\u00edvel, uma vez que \u00e9 aquela que melhor representa a rela\u00e7\u00e3o. O facto do conjunto C1 possuir o atributo Hora a mais, far\u00e1 com que a frequ\u00eancia do conjunto seja menor do que o do conjunto C2, e conseq\u00fcentemente as suas regras subjacentes ser\u00e3o tamb\u00e9m menos frequentes e plaus\u00edveis do que as do conjunto C2. Usando este princ\u00edpio \u00e9 poss\u00edvel definir com exatid\u00e3o o n\u00famero de regras deduzidas de um conjunto C ($\#r(C)$) com n atributos atrav\u00e9s da equa\u00e7\u00e3o:

$$\#r(C) = \begin{cases} 0 & \text{se } n = 1 \\ \left(\sum_{i=1}^n \binom{n}{i} \times (2^{n-i} - 1) \right) - \left(\sum_{i=1}^{n-1} \binom{n-1}{i} \times (2^{n-i-1} - 1) \right) & \text{se } n > 1 \end{cases}$$

Pode-se desta forma sumarizar que a utilização de conjuntos apresenta-se como uma forma compacta e eficiente de representar relações (dependências). Figura-se também como meio termo entre conter uma vasta quantidade de informação e ser eficiente em termos de processamento. Como já narrado, um perfil representa-se como um agregado de conjuntos e é armazenado na base de dados, segundo o submodelo da base de dados ilustrado pela figura 51.

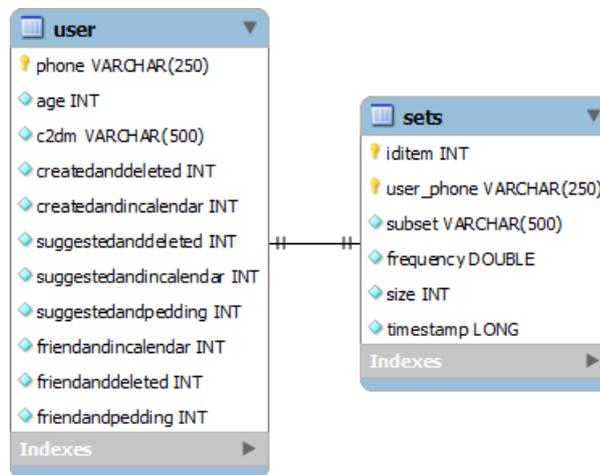


Figura 51 - Armazenamento de um perfil na base de dados

Como é verificável, cada utilizador (tabela *user*) tem associado um agregado único de conjuntos armazenados na tabela *sets*. Para cada utilizador são também registados outros atributos que não apenas o seu telefone, a sua idade e chave atribuída pelo serviço *C2DM*. Todos os demais valores são usados para fins estatísticos. Cada conjunto (*set*) possui uma frase (*subset*) que expressa o conjunto, a frequência associada ao conjunto, o número de elementos do conjunto e a última data em milissegundos em que a frequência foi alterada.

5.2 - Aprendizagem

A aprendizagem do perfil é o processo que torna possível que a informação guardada por este esteja constantemente atualizada e consistente com a pessoa que representa. É este processo e a forma como os perfis são representados, que permitem que os perfis sejam dinâmicos e o que demonstra a inteligência do sistema, em acompanhar e adaptar-se às evoluções de preferências de um pessoa. Um perfil constantemente atualizado permite com que os eventos recomendados tenham uma maior taxa de aceitação, uma vez que o perfil é a base de suporte para a recomendação de eventos.

No que concerne, mais precisamente ao perfil mantido pelo sistema, é necessário garantir que para cada conjunto, o seu valor de frequência e a data de modificação da frequência são atualizados a cada operação realizada pelo utilizador. Assim, se um utilizador praticar frequentemente um determinado evento e em certas condições, todos os conjuntos subjacentes ao evento terão de ser atualizados, para que o sistema aprenda e retenha que o utilizador tem preferência pelo evento e nas respetivas condições.

Existe uma razão importante pela qual a frequência de um conjunto é representada por número percentual e pela qual cada conjunto possui uma data indicativa, da última vez que o valor de frequência foi modificado.

O registo da data temporal de cada conjunto deve-se ao facto das preferências de uma pessoa poderem ser classificadas como de curto ou longo prazo. Preferências de longo prazo são aquelas que se mantêm consistentes ao longo da vida de uma pessoa e correspondem a simples eventos quotidianos, como caminhar todos os dias de manhã. Preferências de curto prazo são eventos que uma pessoa pratica poucas vezes ou então preferências que tendem a diminuir com o decorrer do tempo. Se uma pessoa por uma vez praticou um determinado evento e não voltou a praticar num determinado espaço de tempo, a probabilidade de este evento ser recomendado deverá ser calculado em função do espaço de tempo. Da mesma forma, se uma pessoa praticou frequentemente um evento apenas num espaço de tempo e o deixou de praticar, os novos eventos que substituam os anteriores, embora possam ter uma menor frequência deverão ter maior prioridade em ser recomendados devido a serem mais recentes.

Considere-se por exemplo que um utilizador durante o mês de Agosto vai frequentemente à praia aos sábados de manhã. Enquanto o sistema receber informação de que o utilizador pratica este evento, deverá recomendá-lo para tempos livres existentes aos sábados e

de manhã, uma vez que é visto como uma preferência do utilizador. Porém uma vez que nos restantes meses a temperatura é inferior o utilizador deixará de ir à praia e tenderá a executar uma outra atividade aos sábados de manhã. Esta alteração deve ser aplicada no perfil da pessoa para que o sistema passe a recomendar o novo evento. Quando é necessário recomendar um evento, as preferências da pessoa para as condições do evento são ordenadas primeiro pelo registo temporal e em segundo pela frequência. É por esta razão que o fator tempo foi encarado como um conceito muito importante no desenvolvimento deste trabalho. Esta apresenta-se também, como a razão pela qual todos os registos existentes na base de dados possuem uma data temporal de alteração.

Relativamente à frequência de cada conjunto ser um valor percentual e não uma simples contagem, esta situação deve-se à necessidade que existe em melhor expressar as preferências de uma pessoa em relação aos eventos e seus atributos. O princípio considerado, é que se um utilizador pratica frequentemente um evento, então o valor de frequência de cada conjunto subjacente ao evento deverá expressar que a preferência do utilizador pelo evento tende a aumentar. Da mesma forma, se um utilizador rejeita frequentemente um evento recomendado, o valor de frequência de cada conjunto subjacente ao evento deverá refletir a diminuição da preferência do utilizador.

Por outras palavras, se um utilizador cria um evento e existe registo de que em casos passados e recentes o utilizador criou frequentemente o mesmo evento, é sinal de que o interesse do utilizador pelo evento está a aumentar, pelo que a probabilidade de este ser recomendado deve sofrer um acréscimo considerável. Num cenário oposto, se um utilizador rejeita um evento e há registo que o utilizador rejeitou o mesmo evento em casos passados recentes e frequentes, então a probabilidade de este ser recomendado deverá sofrer um decréscimo considerável. Em ambos os casos o valor a adicionar ou subtrair deverá ter em consideração os casos passados e similares, onde o utilizador executou a mesma ação.

O conceito utilizado é de que se uma frequência tem um valor percentual $x \in [0,1]$ e pretende-se adiciona-lo outro valor, o valor a adicionar será no intervalo $]x, 1]$. Se pelo contrário pretende-se subtrair um valor, o valor a subtrair será no intervalo $[0, x]$.

Este conceito não é aplicado de forma correta, caso a frequência seja representada por um valor inteiro. Considerando o caso em que é necessário adicionar um valor, em que intervalo estará este valor tendo em conta que não existe um domínio limitado para a frequência? Se a frequência tem por exemplo um valor 3054, quanto deverá ser adicionado sabendo que não

existe um valor máximo conhecido? Da mesma forma, um valor numérico implicaria com que existissem diferentes dimensões entre as frequências. Imaginando-se uma outra frequência com valor 100, o retirar de uma percentagem de 10%, não terá a mesma dimensão do que no valor de frequência 3054. A utilização de uma escala percentual garante que todos as frequências apresentam-se sobre a mesma escala e facilita no cálculo de valores a adicionar ou remover.

O problema agora enfrentado na aprendizagem é calcular o valor a adicionar ou remover às frequências. Conhecendo-se os intervalos possíveis, a aplicação de uma percentagem sobre o intervalo permite determinar o valor final a atualizar. Omitindo por agora os conjuntos e considerando as preferências como representativas por apenas o nome do evento, se a preferência de um utilizador pelo evento Futebol é de 30%, caso o utilizador aceite ou crie um novo evento que seja Futebol, deverá ser adicionado um valor no intervalo $]0.3, 1]$, o que significa que no máximo serão adicionados 0.7 valores percentuais. O processo de aprendizagem passa agora por determinar um outro valor percentual $p \in [0,1]$ que multiplicado pelos 0.7 dê o valor final a adicionar. Considerando a situação inversa, em o utilizador recusa um evento, o processo da aprendizagem passa por determinar um valor p , mas neste caso a ser multiplicado pelos 0.3.

Este valor p deverá ser calculado em função da similaridade existente com casos passados similares, onde o utilizador tenha executado a mesma ação. É nesta etapa que é introduzido o Raciocínio Baseado em Casos (CBR) que tem como objetivo obter o conjunto de n casos passados similares. A similaridade entre dois casos tem em conta o fator tempo, significando que casos passados cuja similaridade seja próxima de 1 ocorreram recentemente e são muito semelhantes ao novo caso, em relação aos seus atributos. Um valor de similaridade próxima de 0 significa que o caso ocorreu há algum tempo ou cujos atributos não são semelhantes em relação ao novo caso. Pretende-se desta forma que com o decorrer do tempo, os casos passados tendam a ser menos similares com novos, uma vez que já ocorreram há algum tempo e não existe tanta certeza sobre a sua aplicabilidade na situação atual.

No processo de cálculo do valor p , omitindo uma vez mais os conjuntos, considere-se o caso já descrito em que o utilizador cria um novo evento Futebol e cuja preferência do utilizador pelo evento é de 30%. As tabelas 10 e 11 representam dois conjuntos de casos passados e similares possíveis de obter. Casos onde o utilizador criou um evento Futebol.

Tabela 10 - Casos passados similares com alto valor de similaridade

Caso	Similaridade
C1	0.95
C2	0.90
C2	0.88
C3	0.85
C5	0.81

Analisando a tabela 10, deduz-se que os valores de similaridade para cada um dos casos passados são altos, significando que o utilizador em situações recentes e em condições idênticas criou um evento Futebol. Isto significa que o valor p deverá ser alto como forma a adaptar o perfil, para que este expresse um acréscimo nas preferências do utilizador pelo evento Futebol.

Tabela 11 - Casos passados similares com baixo valor de similaridade

Caso	Similaridade
C1	0.30
C2	0.25
C2	0.20
C3	0.19
C5	0.17

Em situação inversa encontra-se a tabela 11, na qual os valores de similaridade são baixos. Esta situação indica que as últimas vezes em que um utilizador criou um evento Futebol foi já há algum tempo ou foi em condições pouco idênticas. Nestas condições o valor p deverá ser menor, uma vez que existe menor certeza se a preferência do utilizador pelo evento Futebol prevalece. Como será explicado no CBR, todos os novos casos (eventos) são arquivados, isto significa que após o utilizador criar um novo evento esta passará a constar em buscas futuras por casos passados mais similares. Se o utilizador criar frequentemente eventos com título Futebol e nas mesmas condições, situações como as representadas na tabela 11, onde existem casos poucos similares, tenderão a tornar-se em situações como as da tabela 10, com muitos casos similares, uma vez que o valor de similaridade tem em consideração o tempo. A situação

inversa é também verificável, uma vez que se o utilizador deixar de praticar um evento o valor de similaridade tenderá a diminuir, o que possibilitará a transição de uma situação como a que é elucidada pela tabela 10 para a situação da tabela 11.

Obtido um conjunto de casos passados, é agora necessário determinar uma função que usando os valores de similaridade determina o valor p e que respeite as considerações descritas anteriormente. A fórmula obtida foi a seguinte:

$$p = \log_{10} \left(\frac{\sum_{i=1}^n sim_i}{n} + \alpha \right) * \left(\frac{n}{m} \right)$$

onde n é o número de casos passados obtidos, m o número máximo de casos passados a obter, sim_i representa o valor de similitude do novo caso com o caso i , e α define-se com um controlo sobre a proporção do valor p tal que $\alpha \in [1,9]$. Por defeito o valor de α é 1.

Como pode verificar-se o primeiro cálculo realizado é a média dos valores de similaridade dos casos passados obtidos. Sobre a média é somado um valor α e em seguida é aplicado a função logarítmica de base 10 que é utilizada como forma de suavizar o valor final de p . Em seguida o resultado da função logarítmica é multiplicado pela razão entre o número de casos passados obtidos n e o número máximo de casos passados a obter m . A aplicação desta razão deve-se ao facto do número de casos passados similares obtidos, não ser sempre igual ao que é definido como máximo (por defeito 5). Ou seja, poderá haver situações em que em vez de cinco casos como nas tabelas 10 e 11, apenas sejam encontrados dois casos passados e similares. O número de casos obtidos influencia a certeza sobre as preferências, obter um conjunto de cinco casos passados e muito similares assegura maior certeza sobre a preferência da pessoa do que um conjunto com apenas dois casos.

A título exemplificativo, para as tabelas 10 e 11 e segundo o exemplo do evento Futebol proposto, o valor de p para cada tabela será de 0.27 e 0.09 respetivamente. Assim, para a situação representada pela tabela 10 serão adicionados 0.19 ($0.27 * 0.7$) valores percentuais à preferência do utilizador em praticar o evento Futebol. Para a situação representada pela tabela 11 o valor adicionado será de 0.06 ($0.09 * 0.7$).

Usando de novo as tabelas 10 e 11, e o mesmo valor de preferência 30% para o evento Futebol, mas considerando agora que o utilizador recusa a prática do evento Futebol, o intervalo sobre qual deverá ser subtraído um valor é agora $[0,0.3]$, o que significa que será calculado um

valor p que multiplicado por 0.3 dará o valor final a subtrair. Nesta situação, a interpretação feita em relação à tabela 10 é que o utilizador recusou frequentemente e recentemente eventos com título Futebol e em condições idênticas. Perante este cenário, o valor a decrescer à preferência do utilizador pelo evento Futebol deverá ser grande, uma vez que existe grande certeza sobre o facto de o utilizador não gostar ou não querer praticar o evento. Desta forma, a probabilidade de um novo evento Futebol ser recomendado será bem menor.

A interpretação da tabela 11 introduz maiores incertezas. Perante a recusa do utilizador ao evento e pelo facto dos últimos casos onde o utilizador recusou terem já algum tempo ou terem sido em condições muito pouco idênticas, não existe muita certeza do utilizador não ter qualquer interesse em eventos Futebol, pelo que o valor a subtrair é menor. Contudo, partindo da situação da tabela 11, se o utilizador recusar todos os eventos Futebol que lhe são recomendados, passará a existir uma situação de maiores certezas como revela a tabela 10.

Conforme demonstrado, para o cálculo do valor p são usados os valores de similaridade de casos passados mais similares com o novo caso, sendo para isso necessário encontrar o conjunto de n casos passados mais similares. Antes de explicar em detalhe como são encontrados os casos é conveniente indicar a que corresponde um caso no domínio do trabalho e como este é definido. Um caso representa um evento que pode ser criado pelo utilizador, recomendado pelo sistema ou sugerido por um conhecido do utilizador. Face a esta definição, o objetivo é agora perante um novo evento, obter o conjunto de eventos passados e similares com o novo.

A técnica utilizada para esta tarefa é o Raciocínio Baseado em Casos (CBR) que foi descrito no capítulo 3. O objetivo pretendido com o uso do CBR, neste trabalho, contrapõe-se um pouco ao tradicional em que é pretendido adaptar a solução de casos passados e similares, para a solução de um novo problema. Na aplicação que foi feita do CBR, das tradicionais quatro fases (figura 21), apenas duas são usadas e de acordo com a figura 52. Uma vez recuperados, o conjunto de n eventos passados mais similares e seus respetivos valores de similaridade, os valores são utilizados no cálculo do valor p que será usado para atualização do perfil. Atualizado o perfil, o novo evento é retido no repositório de eventos, ou seja, na agenda.

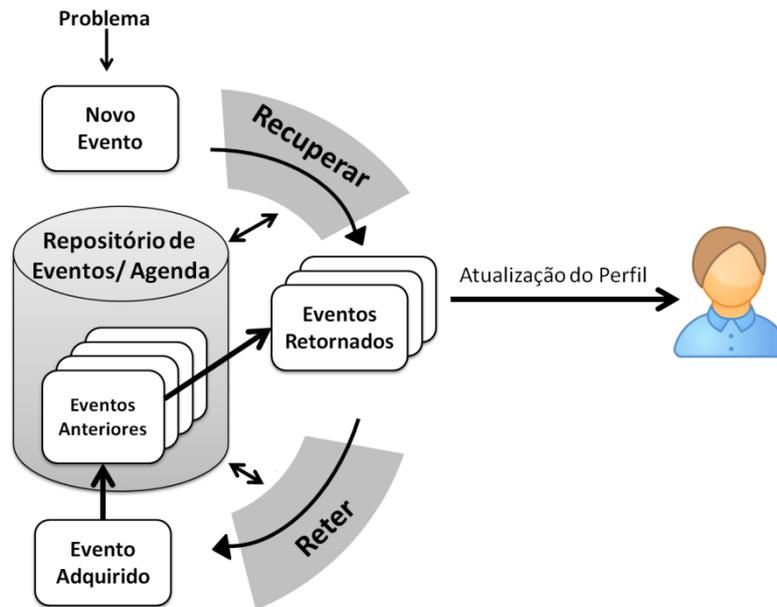


Figura 52 - Adaptação do Raciocínio Baseado em Casos

Um evento é descrito pelos atributos elucidados na tabela 7 e definidos no capítulo 4.1. Conforme foi estudado no capítulo 3, a similaridade entre dois eventos é calculada em função da similaridade local existente entre todos os seus atributos. O primeiro passo é portanto calcular a similaridade local entre cada atributo. A forma como o valor de similaridade é calculado depende do tipo de atributo, que como foi apresentado no capítulo 3 poderá recorrer a várias técnicas.

Começando pelos atributos do tipo numérico, a similaridade entre dois atributos $s(a_i, a_j)$ é calculada em função da distância positiva d entre estes. Para transformar a distância num valor de similaridade é usado uma função linear definida pela equação:

$$s(a_i, a_j) = \begin{cases} 0 & \text{se } d(a_i, a_j) > \beta \\ 1 - \frac{d(a_i, a_j)}{\beta} & \text{se } d(a_i, a_j) \leq \beta \end{cases}$$

Como é deduzível, a constante β define a distância a partir da qual se considera que os atributos não são similares (que o valor de similaridade é zero). Uma distância zero significa que os atributos são iguais e que a similaridade entre estes é de 1. A constante β varia em função do atributo para o qual é calculada a similaridade. Por exemplo, para o atributo "Hora de Início",

a constante β toma o valor 360 e indica que se a distância entre as hora de início dos dois eventos for superior a 360 minutos, então os atributos são totalmente diferentes, pelo que possuem similaridade 0. Se por exemplo, o atributo for a data em que o evento foi alterado, β toma o valor 44 640 o que indica que se a distância entre as datas de alteração de dois eventos foi maior do que 44 640 minutos (31 dias), a similaridade entre os atributos é 0, pelo facto do caso ser muito antigo. Todas estas constantes são definidas globalmente, o que significa que a qualquer instante é possível alterar o seu valor sendo apenas necessário mudar a variável global.

No que diz respeito aos atributos textuais, o cálculo da similaridade envolve dois algoritmos próprios para determinar a distância entre textos. O primeiro denomina-se *Jaccard* e é extensivamente usado para calcular a similaridade entre valores que não apenas textuais. Este algoritmo usa conjuntos e interseção entre estes, para calcular o quanto similares são dois conjuntos. Por outras palavras, a similaridade entre dois conjuntos define-se como a razão entre o número de elementos comuns aos dois conjuntos e o número total de elementos não repetidos dos dois conjuntos. Para calcular a similaridade entre dois textos (ou frases), é definido para cada texto um conjunto com as suas palavras. Desta forma, a similaridade entre dois textos dependerá do número de palavras que tenham em comum.

Este algoritmo releva-se útil na comparação de frases, contudo a sua aplicação única apresenta algumas lacunas. Primeiro, porque este algoritmo apenas funciona sobre textos que sejam frases, ou seja, que tenham várias palavras. Segundo, a comparação entre duas palavras é feita de forma estática, apenas são iguais se todos os caracteres o forem. Para compensar esta situação, foi introduzido um outro algoritmo conhecido por *Levenshtein*. Este algoritmo calcula o número mínimo de alterações necessárias a realizar num texto, para o transformar no segundo texto. Por outras palavras, quando mais parecidos forem os textos, menor será a distância de *Levenshtein*. Exemplificando, a distância de *Levenshtein* entre as palavras "Futebol" e "Andebol" é de 3, uma vez que para transformar a primeira palavra na segunda é necessário mudar os 3 primeiros caracteres. Ao contrário do algoritmo de *Jaccard* cujo resultado é sempre um número real entre 0 e 1, para o algoritmo de *Levenshtein* foi necessário a aplicação de uma função linear para converter a distância em um valor de similaridade entre 0 e 1. O valor final de similaridade entre dois atributos textuais é dado pela média da similaridade dada pelo algoritmo de *Jaccard* e da similaridade dada pelo algoritmo de *Levenshtein*. Poderão ainda ser adicionados pesos sobre cada algoritmo em função da sua eficácia na comparação de textos.

Por fim, atributos cujos valores sejam pré-definidos, como é o caso da Categoria, a similaridade entre dois atributos é dada pela comparação lógica entre estes, se forem iguais é 1 caso contrário 0. A comparação entre duas categorias poderá no futuro ser melhorada com recurso a um especialista na matéria.

Depois de calculada a similaridade local para cada atributo é então aplicada uma função que definirá a similaridade global entre dois eventos. A equação usada foi a apresentada no capítulo 3 e aqui repetida:

$$sim(X, Y) = \frac{\sum_{i=0}^n \alpha_i * sim(x_i, y_i)}{\sum_{i=0}^n \alpha_i}$$

onde $sim(X, Y)$ é a similaridade global entre dois eventos X e Y , $sim(x_i, y_i)$ representa a similaridade entre cada atributo dos eventos, n é o número de atributos dos eventos que já foi definido como sendo nove, e α_i o peso a atribuir a cada similaridade local. Na implementação feita o peso α é definido como igual para todos por uma questão de conveniência, não sendo contudo afastada a alteração destes pesos num situação futura. Desta forma, pode-se afirmar que a similaridade global entre dois eventos será dada pela média de todas as similaridades locais entre os atributos.

Depois de definido como são recuperados os eventos passados e similares, e como a partir dos seus valores de similaridade é calculado o valor p , é agora exposta uma descrição mais detalhada, de como se realiza o processo de aprendizagem para as operações efetuadas por um utilizador da aplicação móvel. Este processo é ilustrado na figura 53 e será acompanhado por uma descrição de cada uma das etapas. É de referir que em termos técnicos, este processo é feito através de tarefas (*threads*) que são lançadas por cada pedido realizado pelo utilizador. O objetivo é que o processo de aprendizagem seja translúcido ao utilizador (que não afete o tempo de resposta do servidor).

Quando o utilizador realiza qualquer operação que envolva um evento (criação, edição, eliminação, aceitação ou recusa) este envia informação sobre o evento para o servidor. Em situações, como a criação ou a edição, é enviado para o servidor um objeto *JSON* com toda a informação do novo evento. Nas restantes operações, apenas é necessário enviar o identificador único do evento sendo que o servidor recupera da sua base de dados os dados do evento.

Usando o evento interveniente numa operação, são gerados os 511 conjuntos através dos valores dos atributos desse evento (etapa 1 da figura 53). Em seguida segue-se um processo iterativo que tratará de cada um dos conjuntos. Para cada conjunto são recuperados os cinco eventos passados e mais recentes que tenham os mesmos valores dos atributos contidos no conjunto (etapa 2 da figura 53). Por exemplo, para o primeiro conjunto são recuperados os eventos que tenham o atributo $A1=v1$ e por ordem decrescente em relação à data da sua alteração. Para cada um dos eventos obtidos é calculado a similaridade global (etapa 3 da figura 53) que dependerá da similaridade local, entre todos os atributos do evento. Eventos passados, obtidos através do conjunto com cardinalidade nove (com todos os atributos), não terão necessariamente valor de similaridade 1. Embora todos os atributos sejam iguais, o fator tempo interfere no valor de similaridade, de modo a que quanto mais antigo seja o evento, menor seja a sua similaridade. Como resultado final desta etapa é obtido uma tabela com os eventos similares e os seus valores de similaridade.

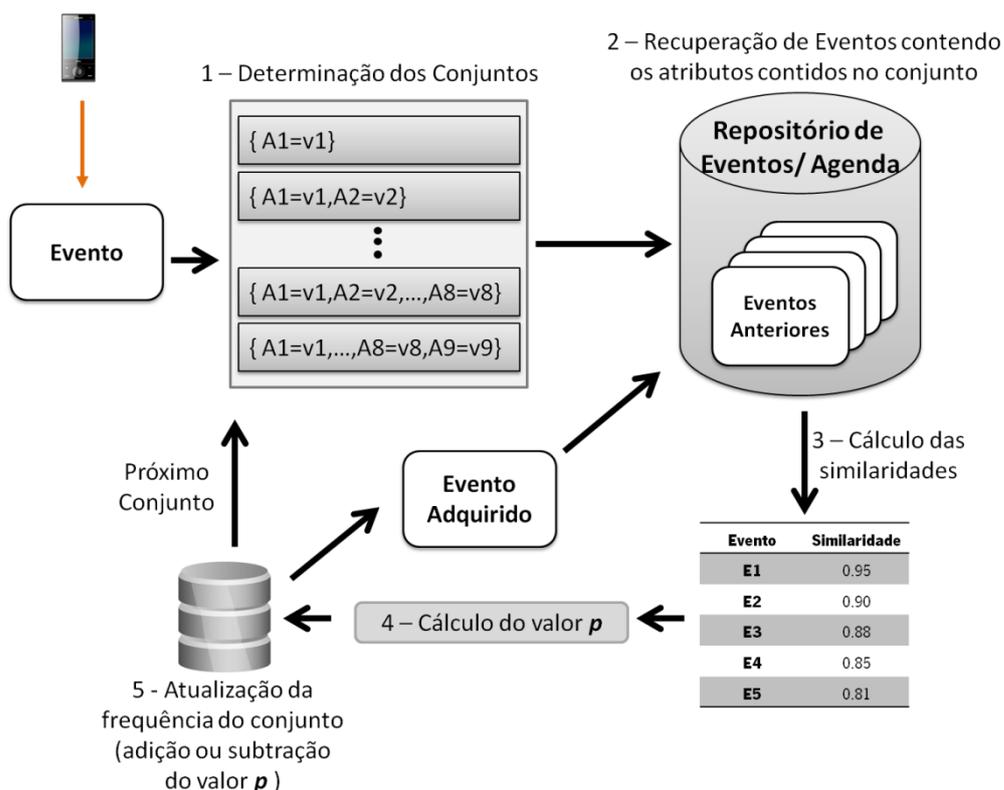


Figura 53 - Processo de Aprendizagem do Perfil

Como já foi demonstrado, usando estes valores de similaridade é calculado o valor p (etapa 4 da figura 53), que será adicionado ou removido à frequência do conjunto armazenado na base de dados (etapa 5 da figura 53). As etapas repetem-se para os restantes conjuntos.

No processo de aprendizagem, conforme a figura 53, a etapa 2 diferencia-se de acordo com a operação que é realizada pelo utilizador. Como é ilustrado pela figura 54, um evento poderá estar em 8 estados. Assim, na base dados cada evento é agrupado segundo o seu estado.

Na etapa 2, apenas serão consultados os eventos que estejam agrupados no estado idêntico ao da operação a ser realizada. Por outras palavras, quando o utilizador cria um evento, serão consultados todos aqueles eventos que tenham sido criados e mantidos na agenda; quando o utilizador elimina um evento serão consultados aqueles eventos que tenham sido criados pelo utilizador, mas já eliminados. Em relação aos eventos que são recomendados pelo sistema, o utilizador ao aceitar um evento, faz com que eventos passados, que o utilizador tenha aceite, sejam consultados; quando o utilizador rejeita o evento são consultados eventos passados que o utilizador tenha rejeitado; eventos recomendados e que o utilizador não tenha tomada uma decisão são considerados como pendentes. Para eventos para os quais um utilizador é convidado é seguido o mesmo esquema dos que são recomendados.

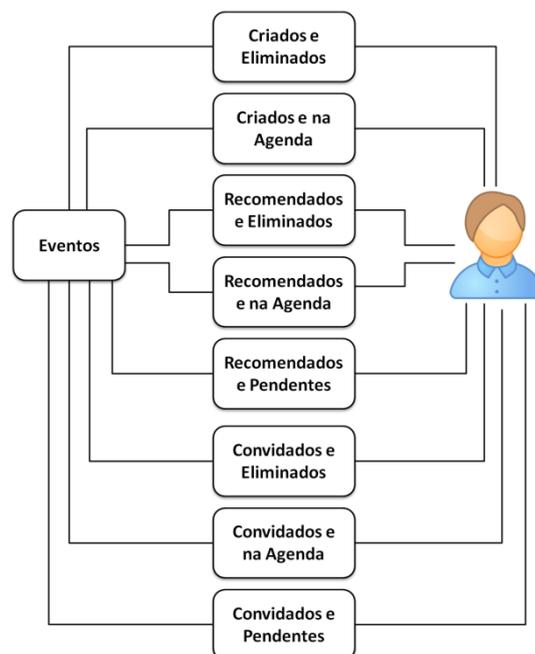


Figura 54 - Modulação da base de dados para a agenda de um utilizador

Pode-se decerto questionar o porquê dos eventos não serem apenas agrupados em três estados de criados/aceites, eliminados/rejeitados ou pendentes, em vez dos oito utilizados. A razão deve-se à possibilidade de cada ação poder ter um peso diferente na atualização do perfil. A questão que se coloca é se a alteração do perfil de um utilizador deverá ter o mesmo peso quando este recusa um evento para o qual foi convidado e quando elimina um evento por si criado. Esta questão poderá ter várias respostas de diferentes especialistas ou então poderá depender da personalidade da pessoa. Por essa razão, a forma como foi encarado este problema torna simples a alteração dos pesos que cada operação deverá ter sobre a aprendizagem do perfil.

5.3 - Recomendação de Eventos

A recomendação de eventos para os tempos livres segue a abordagem feita por sistemas de recomendação baseados em conteúdo. São sugeridos ao utilizador eventos que sejam semelhantes aos preferidos por este. Contudo, uma vez que o objetivo social deste trabalho é incentivar atividades que não são totalmente sedentárias e que ao mesmo tempo potencializem o convívio social, só são recomendadas atividades que não estejam nas categorias "Outra" ou "Ocupado". Parte-se do pressuposto de que os utilizadores cumprirão com a categorização dos eventos que criam.

Cada utilizador possui uma tarefa cíclica no servidor (*thread*) que de vez em quando procura por tempos livres na agenda semanal do utilizador. Assume-se como tempo livre, um espaço de tempo que não tem qualquer evento associado. Desta forma, se um utilizador não desejar que sejam recomendados eventos para um determinado espaço temporal, deverá criar um evento na categoria de "Ocupado".

Cada tarefa repete-se de vez em quando, por exemplo, a cada 30 minutos (valor configurável), a tarefa desperta e trata de procurar por tempos livres na agenda semanal do utilizador e de recomendar um evento para esses espaços. Começando pelo início da semana e percorrendo até ao final, a tarefa procura espaços livres que sejam superiores a uma determinada duração. Quando encontra um tempo livre, o sistema procura nas preferências do utilizador eventos que se enquadrem no espaço. Para além do tempo que existe entre os ciclos de uma tarefa, a quantidade de eventos que são sugeridos em cada ciclo de vida de uma tarefa é também configurável.

Em determinados tempos, a tarefa escolhe e recomenda um evento aleatório de um conjunto pré-definido e que não seja da preferência do utilizador como forma de procurar por novos pontos de interesse do utilizador. A probabilidade com que um evento aleatório é recomendado depende da idade da pessoa, para que quanto maior seja a idade menor seja esta probabilidade. A tarefa escolhe e recomenda um evento aleatório também em situações para os quais não sejam conhecidas as preferências do utilizador (para situações iniciais).

Descreve-se agora, o processo pelo qual uma tarefa determina e recomenda um evento para um determinado tempo livre. O objetivo na recomendação é construir um evento que seja do agrado do utilizador. A construção consiste em definir o valor para cada um dos nove atributos do evento. O processo de realização desta construção é ilustrado pela figura 55.

Como se pode verificar nesta figura, a construção de um evento utiliza uma estrutura em árvore, em que cada folha está associada a um conjunto representativo das preferências de um utilizador. Cada folha terá um conjunto de filhos representativos de todos os superconjuntos com cardinalidade maior em uma unidade. A profundidade da árvore será no máximo de nove, tendo em conta que um conjunto terá no máximo nove elementos.

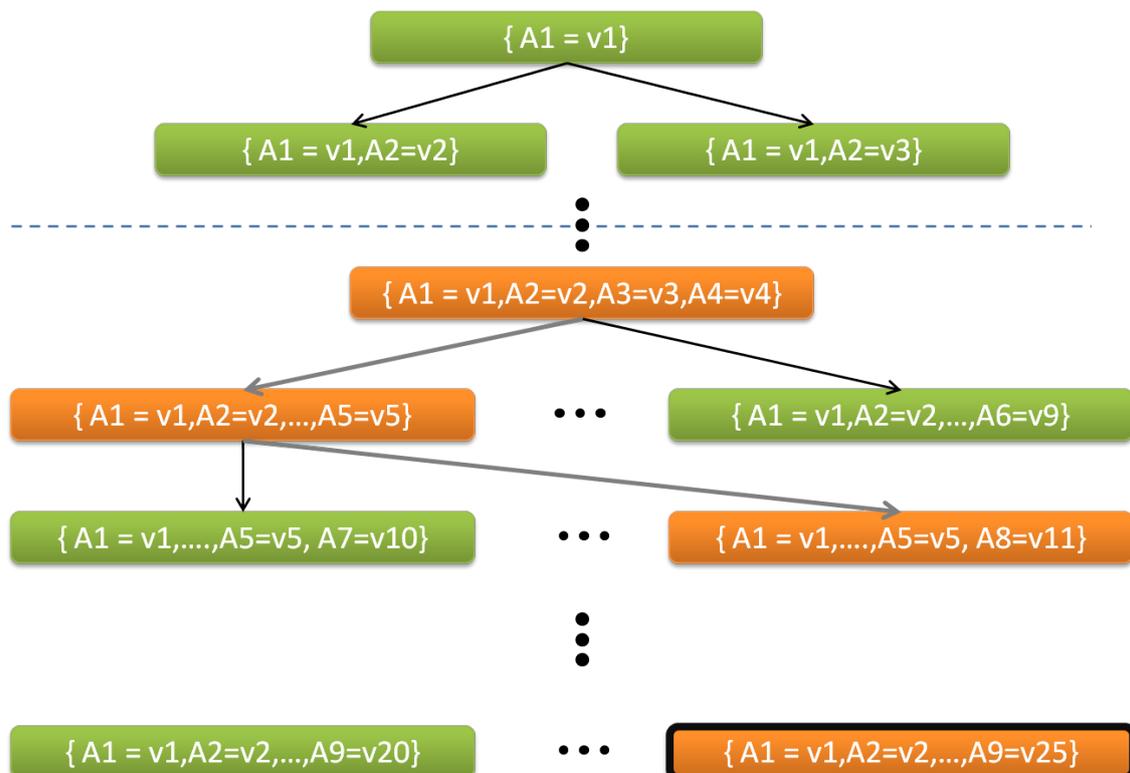


Figura 55 - Estrutura em árvore usada para a recomendação de eventos

Quando existe um tempo livre é conhecido para este a sua hora de início, a sua hora de término, a sua duração e o dia da semana. Isto significa que para a determinação dos restantes atributos do evento, só são analisados conjuntos com cardinalidade maior ou igual a quatro. Embora os conjuntos de cardinalidade inferior a quatro não sejam utilizados no processo de construção de um evento, já foi demonstrado no capítulo 5.1 que podem ser utilizados para outros fins.

A construção de um evento é feita através de um processo iterativo que encontrará um caminho na árvore, desde o conjunto conhecido de quatro atributos até à última folha que será um conjunto de cardinalidade nove. O conjunto de cardinalidade nove encontrado será considerado como sendo o evento final construído e a ser recomendado ao utilizador, ou seja, será constituído pelos nove atributos do evento para os quais foi encontrado um valor.

O percurso pela árvore usa um princípio semelhante ao usado nas regras de associação, o qual afirma que se um determinado conjunto não é frequente, então qualquer seu superconjunto também não o será. Assim, tendo como raiz um conjunto de cardinalidade quatro, contendo os valores para os atributos hora de início, hora de término, duração e dia da semana, é encontrado o conjunto de cardinalidade cinco que seja o melhor classificado de acordo com o seu valor de frequência. Isto significa que para as condições definidas pelo conjunto de quatro atributos, o atributo e valor que melhor se relaciona com estes (que ocorrem conjuntamente com maior frequência) é o atributo adicional que o conjunto de cardinalidade cinco contém e o seu respetivo valor. Se os restantes conjuntos de cinco elementos não são mais frequentes do que o encontrado, então não necessitam de ser analisados, bem como os seus superconjuntos também não o necessitarão. Servindo-se do conjunto de cardinalidade cinco é aplicado o mesmo processo na procura do conjunto com cardinalidade seis. Este processo é executado sistematicamente até encontrar o conjunto de cardinalidade nove. A figura 56 ilustra um exemplo real de como uma árvore se apresenta e de como é percorrida.



Figura 56 - Exemplo de uma estrutura em árvore usada para a recomendação de eventos

Analisando a figura consta-se que para as situações onde a hora de início é 16h, a hora de término é 17:30h, a duração é de 90 minutos e o dia da semana é Terça (1), o atributo Local com o valor "Universidade do Minho" é o que tem maior preferência do utilizador nas condições indicadas, em comparação com o atributo Categoria e valor "Cultura". Isto permite afirmar que as quatro condições conhecidas têm maior influência sobre a escolha do local do que a categoria. O processo é repetido até atingir um conjunto com cardinalidade nove. Por exemplo, para o conjunto de cardinalidade cinco, escolhido na figura 56 (representado a laranja), será determinado qual o novo atributo e respetivo valor, com o qual os cinco atributos atuais melhor se relacionam.

Em situações específicas poderão ser aplicados filtros sobre os valores de certos atributos. Poderá ser conveniente em certos casos, recomendar eventos cujo número de participantes seja maior do que um. Nestes casos é pretendido incentivar o convívio social, na medida em que ao recomendar eventos que tenham maior número de participantes o utilizador tenderá a convidar amigos.

5.4 - Conclusões

Acredita-se que o conceito de Personalização adotado neste trabalho, conjugado com a aplicação móvel desenvolvida, contribui para o decréscimo do sedentarismo nas pessoas.

Como já mencionado, a base de funcionamento da recomendação são os perfis. É através destes que o sistema seleciona as atividades físicas a recomendar. De forma eficiente e consistente, um perfil representa todas as relações que existem entre os valores dos atributos de um evento, fazendo com que a taxa de aceitação das atividades recomendadas seja maior.

A aprendizagem é também importante para o aumento da aceitação, uma vez que permite que um perfil está constantemente atualizado, com as preferências de quem representa. O sistema acompanha um utilizador ao longo da sua vida e adapta o perfil em função de alterações nos seus interesses.

Tanto a capacidade do sistema em modular, como a de aprender acerca das preferências do utilizador, contribuem para o decréscimo do sedentarismo nos tempos livres das pessoas.

Uma vez que o sistema apenas recomenda atividades que não sejam sedentárias, este tenta cativar os utilizadores para a prática de atividades físicas nos seus tempos livres. Aprendendo acerca de quais destas atividades o utilizador prefere, o sistema torna-se mais preciso nas atividades que recomenda, aumentando desta forma o gosto dos utilizadores pelas atividades físicas. Tenta também cativar os utilizadores a praticarem novas atividades físicas que não apenas aquelas da sua preferência, como forma a aumentar a sua prevalência sobre as atividades sedentárias.

Capítulo 6: Conclusões e Trabalho Futuro

Este trabalho foi desenvolvido no âmbito do Mestrado em Engenharia Informática, proporcionado pelo Departamento de Informática da Universidade do Minho. O seu desenvolvimento foi orientado pelo Professor Doutor Paulo Novais e coorientado pelo Mestre Ângelo Costa.

A Inteligência Artificial e os Sistemas Inteligentes foram as temáticas abrangidas no seu desenvolvimento. Nestas integra-se a temática *User Profiling* que foi a base de todo o trabalho desenvolvido. Face ao número de artigos e documentos encontrados na literatura, aquando do seu estudo, pode-se reconhecer o quanto é investigada. Foram também estudados outros conceitos de Sistemas Inteligentes de forma a fortalecer e aperfeiçoar o conhecimento existente sobre *User Profiling*. Destaca-se o papel fundamental que o relatório de pré-dissertação teve nestes estudos.

Todos os estudos realizados não só contribuíram para o desenvolvimento deste trabalho como também permitiram o enriquecimento e aperfeiçoamento do conhecimento dos seus intervenientes.

Fundamental foram também as Unidades Curriculares lecionadas no primeiro ano do Mestrado. A unidade de Sistemas Distribuídos pela contribuição que teve para o desenvolvimento do sistema que se apresenta no seu todo como um sistema distribuído. A unidade de Sistemas Inteligentes por ser aquela na qual foram apreendidos todos os conhecimentos basilares no domínio da Inteligência Artificial.

O desenvolvimento da aplicação para dispositivos móveis com o sistema operativo *Android* foi também uma mais valia para o enriquecimento do conhecimento, no que diz respeito ao desenvolvimento de aplicações para o mercado cada vez mais emergente que é o dos *smartphones*.

Neste capítulo é apresentada uma síntese sobre o trabalho desenvolvido e os objetivos alcançados, assim como outros trabalhos relevantes que foram desenvolvidos paralelamente ao trabalho principal e por fim áreas ou conceitos que deverão ser analisados no futuro, com vista à expansão ou melhoria do sistema.

6.1 - Síntese de Trabalho

iLeisure é a designação do sistema desenvolvido neste trabalho e que apresenta-se como uma solução para os problemas apresentados na Introdução. A construção do sistema subdividiu-se no desenvolvimento de uma aplicação móvel e no desenvolvimento de um servidor inteligente.

A aplicação móvel é o ponto de interação dos utilizadores com o sistema e foi implementada para dispositivos móveis com sistema operativo *Android*. Através de alguns testes quanto à usabilidade e utilidade da aplicação, constatou-se a sua aptidão em complementar uma agenda convencional com funcionalidades extras. A interface implementada na aplicação foi considerada intuitiva e agradável. A possibilidade de agendar eventos com amigos foi considerado uma funcionalidade muito útil e interessante. A recomendação de eventos pelo servidor foi também encarada como uma função vantajosa, nomeadamente para situações onde existem algumas dificuldades em escolher eventos para os tempos livres.

O servidor desenvolvido apresenta ser uma solução eficiente e eficaz para a modulação de perfis dinâmicos e para a recomendação de eventos benéficos para a saúde. Para representar os perfis de utilizadores e mais especificamente as suas preferências por eventos, foram usados conjuntos que expressam relações entre os atributos de um evento. Os conjuntos revelam ser uma forma compacta e eficiente de armazenar e interpretar as relações existentes nas preferências dos utilizadores. Ao sumarizarem várias relações de implicação, os conjuntos tornam a sua edição e interpretação computacionalmente mais eficiente.

O servidor revela-se inteligente na medida em que apresenta as capacidades de aprender e adaptar os perfis dos utilizadores. Todas as operações realizadas na aplicação móvel são observadas implicitamente pelo servidor, que as aproveita para aprender acerca das preferências dos utilizadores. A aprendizagem consiste em determinar os conjuntos que são subjacentes ao evento envolvido na operação e para cada um destes adaptá-los no perfil do utilizador. O fator tempo é encarado com especial atenção, face à capacidade do utilizador em mudar as suas preferências com o decorrer do tempo. Para tal fim, são usados eventos passados do utilizador no processo de aprendizagem, com o propósito de determinar quando é que um utilizador tem ou deixa de ter interesse por determinados eventos.

Acredita-se desta forma que o servidor apresenta uma solução eficiente e robusta para modular perfis dinâmicos, permitindo com que os perfis estejam constantemente atualizados e consistentes com as preferências das pessoas que representam.

Relativamente à solução implementada no servidor para recomendação de eventos, crê-se que os eventos recomendados pelo servidor contribuem para a diminuição do sedentarismo. Ao procurar nas agendas dos utilizadores por tempos livres e ao recomendar para esses tempos eventos que sejam da preferência dos utilizadores, mas que não sejam totalmente sedentários, está a ser diminuída a probabilidade de que sejam praticados eventos que não envolvam exercício físico. A recomendação aleatória de eventos com atividade física e que não estejam registados como preferências de um utilizador, contribui também para que este venha a praticar e preferir novos eventos ativos.

Foram realizados estudos acerca da taxa de aceitação dos utilizadores para os eventos recomendados. Estes estudos assistiram na avaliação da precisão do sistema em recomendar eventos, bem como na sua capacidade em acompanhar e aprender as preferências dos utilizadores. Consultando periodicamente registos da quantidade de eventos recomendados aceites ou rejeitados, foi possível verificar que o número de eventos aceites tende a aumentar e os rejeitados a diminuir. Estes resultados permitem afirmar que o sistema possui a capacidade de aprender e aperfeiçoar o perfil dos seus utilizadores com o decorrer do tempo e dessa forma recomendar eventos cada vez mais precisos com os interesses dos utilizadores. No entanto, reconhece-se que o sistema requer estudos mais prolongados de modo a obter maiores certezas quanto às preferências dos utilizadores e consequentemente aumentar a aceitação das atividades recomendadas. Há que estudar, em longos períodos de tempo, as alterações reais que sucedem nos interesses das pessoas, e verificar a aptidão do sistema em descobri-los e aplicá-los sobre os perfis.

No que concerne ao objetivo de proporcionar o convívio social, crê-se que a solução implementada contribui para a envolvimento de maior número de pessoas na prática de eventos. A adição de uma funcionalidade na aplicação móvel que permita convidar pessoas para eventos incentiva e cativa os utilizadores a conviver na prática de eventos. A avaliação positiva demonstrada por esta funcionalidade sustenta esta conclusão. Ao recomendar eventos que envolvam maior número de participantes, o servidor está também a incentivar o utilizador a conviver socialmente.

Em suma, acredita-se que o sistema desenvolvido apresenta qualidades eficientes e funcionais capazes de contribuir para a sociedade e para a comunidade científica. Em relação à ciência, crê-se que o trabalho desenvolvido contribui para o conceito de *User Profining* e todas as outras áreas que tenham suporte neste conceito.

6.2 - Trabalho Relevante Realizado

Parcialmente ao desenvolvimento deste trabalho foram realizados outros trabalhos considerados relevantes, para a projeção dos intervenientes deste trabalho. Foi redigido e publicado o seguinte artigo científico.

WICT 2011

Informações sobre a Conferência

Título	World Congress on Information and Communication Technologies, IEEE Computer Society
Data	11 a 14 de Dezembro de 2011
Localização	Universidade de Mumbai, Mumbai, Índia
Endereço Web	http://www.mirlabs.org/wict11/
Repositório de Publicações	http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?pu number=6132196

Informações sobre o Artigo

Título	A Dynamic User Profiling Technique in a Aml Environment
Autores	Vitor Marques, Ângelo Costa, Paulo Novais
Editora	IEEE Computer Society
ISBN da Publicação	978-1-4673-0127-5
Páginas	1247 - 1252
DOI	http://dx.doi.org/10.1109/WICT.2011.6141427
Palavras-chave	"Ambient Intelligence", "Profiling", "User"

6.3 - Trabalho Futuro

Com as pessoas a desejarem cada vez mais que os contextos ou ambientes adaptem-se de forma inteligente com a sua personalidade e perfil, o tema do *User Profiling* é cada vez mais investigado e novas propostas surgem rapidamente. É por isso necessário investigar periodicamente estas novas propostas, de forma a enriquecer o sistema desenvolvido.

Existem muitos conceitos que devem ser investigados e que uma vez implementados complementarão o sistema, tanto em termos funcionais, inteligentes ou eficientes.

Podendo o sistema desenvolvido ser reconhecido como um sistema de recomendação, existem muitos estudos sobre novas técnicas de recomendar itens. As abordagens híbridas, ou seja, uma recomendação baseada no conteúdo e colaborativa, são cada vez mais adotadas face às vantagens que apresentam. Recentemente as redes sociais têm também sido propostas como fonte de informação a usar pelos sistemas de recomendação (Tiroshi et al., 2011). Face à crescente aderência que as redes sociais apresentam, a sua adoção no sistema é vista como uma boa forma de recolher implicitamente informação adicional sobre os utilizadores. Esta informação seria adequada para melhor precisar o perfil destes e consequentemente precisar os eventos que são recomendados.

Uma vez que o sistema usa noções de similaridade entre eventos e tendo em conta que em alguns atributos de um evento são textuais, uma análise em pormenor de técnicas capazes de extrair conhecimento a partir de textos (*Text Mining*) permitirá precisar a similaridade entre eventos. Por outro lado, estas técnicas são também uma forma de conseguir extrair informação extra e relevante a partir dos textos inseridos pelo utilizador.

Outra área cada vez mais investigada é a computação afetiva (Khanna & M.Sasikumar, 2010). Interpretar os estados emocionais de um utilizador fará com que o seu perfil seja também capaz de representar relações entre eventos e estados emocionais, fazendo com que o sistema seja mais preciso nos eventos que recomenda.

Tratando-se um perfil de um conjunto de informações pessoais, é fundamental e obrigatório por legislação, garantir a sua segurança e privacidade. É por isso importante estudar e implementar conceitos relacionados com criptografia e segurança dos sistemas informáticos.

Os ambientes inteligentes são também uma área emergente e que está interligada com perfis. É através dos perfis que os ambientes adaptam-se aos seus utilizadores. Tendo em conta que os ambientes inteligentes estão muito relacionados com o trabalho desenvolvido e sabendo

que esta é uma área cada vez mais discutida na atualidade, a sua importância revela-se determinante para que se prossiga com investigações futuras.

Bibliografia

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning; Foundational issues, methodological variations, and system approaches. *Ai Communications*, 7, 39-59.
- Agnieszka Onisko, M. J., & Wasyluk, H. (1999). A Bayesian Network Model for Diagnosis of Liver Disorders. *Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering*, 842-846.
- Andersen, R. E., Crespo, C. J., Bartlett, S. J., Cheskin, L. J., & Pratt, M. (1998). Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *Jama The Journal Of The American Medical Association*, 1, 938-942.
- Anderson, P. M., & Butcher, K. F. (2006). Childhood Obesity: Trends and Potential Causes. *Childhood Obesity - The Future of Children*, 16, 19-45.
- Balaa, Z. E., Strauss, A., & Maximini, K. (2003). Fm-ultranet: a decision support system using case-based reasoning, applied to ultrasonography. *In Workshop on CBR in the Health Sciences*, 0-3.
- Baluja, S., Seth, R., & Sivakumar, D. (2008). Video suggestion and discovery for youtube: taking random walks through the view graph. *Proceedings of the 17th international conference on World Wide Web*, 895-904.
- Bauknecht, K. (2002). *Ambient Intelligence: The Vision of Information Society*. Obtido em 15 de Março de 2012, de http://www.dke.univie.ac.at/extern/bi_ws20012002/ss2002/Ambient-Intelligence.pdf
- Berger, B. G., Pargman, D., & Weinberg, R. S. (2006). *Foundations of Exercise Psychology*. Fitness Information Technology.
- Berkat, A. (2011). Using Case-Based Reasoning for detecting computer virus. *International Journal of Computer Science Issues*, 8, 606-610.
- Castillo, E., Gutierrez, J. M., & Hadi, A. S. (1997). *Expert Systems and Probabilistic Network Models*. Springer-Verlag.
- Chen, C.-M., Hsieh, Y.-L., & Hsu, S.-H. (2007). Mining learner profile utilizing association rule for web-based learning diagnosis. *Expert Syst. Appl.*, 33, 6-22.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education*. Routledge.

Costa, Â., Novais, P., Costa, R., & Corchado, J. (2010). Multi-Agent Personal Memory Assistant. *Trends in Practical Applications of Agents and Multiagent Systems*, 71, 97-104.

Direcção-Geral da Saúde. (Março de 2005). *Programa Nacional de Combate à Obesidade*. Obtido em 5 de Março de 2012, de http://estilosdevida.anirsf.pt/folder/questao/doc/12_Programa.pdf

Directorate-General for Health & Consumers - European Commission. (Dezembro de 2010). *Strategy for Europe on nutrition, overweight and obesity related health issues*. Obtido em 5 de Março de 2012, de http://ec.europa.eu/health/nutrition_physical_activity/docs/implementation_report_en.pdf

Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, 26, 897-899.

Dumazedier, J. (1974). *Sociology of leisure*. Elsevier Scientific.

EMC. (2011). *The Digital Universe*. Obtido em 8 de Março de 2012, de <http://www.emc.com/leadership/programs/digital-universe.htm>

Eurostat. (Dezembro de 2003). *How women and men spend their time*. Obtido em 7 de Março de 2012, de http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NK-03-012/EN/KS-NK-03-012-EN.PDF

FIDIS. (2005). *Report on Actual and Possible Profiling Techniques in the Field of Ambient Intelligence*. Obtido em 16 de Março de 2012, de http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp7-del7.3.ami_profiling.pdf

Firmino, H. (2006). *Psicogeriatría*. Almedina.

Gamboa, H., & Fred, A. (2001). Designing intelligent tutoring systems: a bayesian approach. *Proceedings of the Ana Fred 3rd International Conference on Enterprise Information Systems ICEIS 2001*, 452-458.

Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User Profiles for Personalized Information Access. In P. Brusilovsky, A. Kobsa, & W. Nejdl, *The adaptive web* (pp. 54-89). Berlin, Heidelberg: Springer-Verlag.

Gediminas, A., & Alexander, T. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17, 734-749.

Godoy, D., & Amandi, A. (2000). PersonalSearcher : An Intelligent Agent for Searching Web Pages. *Advances in artificial intelligence international joint conference 7th IberoAmerican Conference on AI 15th Brazilian Symposium on AI IBERAMIASBIA 2000 Atibaia SP Brazil November 1922 2000 proceedings*, 1952, 43.

Gotomy. (2008). *The Smart Home will be flat!* Obtido em 9 de Março de 2012, de <http://gotomy.wordpress.com/2008/12/17/the-smart-home-will-be-flat/>

Graf, C., Koch, B., Falkowski, G., Bjarnason-Wehrens, B., Tokarski, W., Christ, H., et al. (2005). Daily routine and leisure time behavior of obese children compared with their contemporaries. *Journal of Pediatrics and Neonatology*, 2, 1-5.

Grcar, M., Mladenec, D., & Grobelnik, M. (2005). User profiling for interest-focused browsing history. *Proceedings of the 8th international multiconference on Information Society IS*, 182-185.

Guedes, D. P., & Guedes, J. E. (1995). *Exercício físico na promoção da saúde*. Midiograf.

H.Witten, I., & Frank, E. (2005). *Data Mining - Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Hitech Projects. (2004). *Ambience Project*. Obtido em 14 de Março de 2012, de <http://www.hitech-projects.com/euprojects/ambience/>

Horvath, J. (2002). *Making Friends with Big brother*. Obtido em 14 de Março de 2012, de Telepolis: <http://www.heise.de/tp/artikel/12/12112/1.html>

Horvitz, E., Breese, J., Heckerman, D., Hovel, D., & Rommelse, K. (1998). The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 256-265.

Infographic. (2011). *Smartphone users around the world - Statistics and Facts*. Obtido em 8 de Março de 2012, de <http://www.go-gulf.com/blog/smartphone>

Instituto Nacional de Estatística. (Janeiro de 2012). *Estatísticas Demográficas 2010*. Obtido em 27 de Fevereiro de 2012, de Publicações: http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=136077435&att_display=n&att_download=y

IP Network. (2011). *International Key Facts*. Obtido em 4 de Março de 2012, de http://www.ip-network.com/rd/assets/file_asset/Viewing_Time_Individual.pdf

Kemmis, S., & McTaggart, R. (2005). Participatory Action Research: Communicative Action and the Public Sphere. In N. K. Denzin, & Y. S. Lincoln, *Handbook of Qualitative Research* (pp. 271-330). Sage.

Khanna, P., & M.Sasikumar. (2010). Article: Recognising Emotions from Keyboard Stroke Pattern. *International Journal of Computer Applications*, 11, 1-5.

Korb, K., & Nicholson, A. E. (2003). *Bayesian Artificial Intelligence*. Chapman and Hall.

Krishnakumar, K. (2003). Intelligent Systems for Aerospace Engineering - An Overview. *RTO-EN-022 - Intelligent Systems for Aeronautics*, 0-17.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7, 76-80.

Ling, S., Indrawan, M., & Loke, S. W. (2007). RFID-based user profiling of fashion preferences: blueprint for a smart wardrobe. *Int. J. Internet Protoc. Technol.*, 2, 153-164.

Lisbon Internet and Networks International Research Programme. (2010). *A Utilização de Internet em Portugal 2010*. Obtido em 7 de Março de 2012, de http://www.unic.pt/images/stories/noticias/Relatorio_LINI_UMIC_InternetPT.pdf

Macdonald, R., & Silver, D. L. (s.d.). *Web-based User Profiling Using Artificial Neural Networks*. Obtido em 20 de Março de 2012, de <http://www.wardsystems.com/page15.asp>

Marktest. (Dezembro de 2011). *Smartphones têm 360 mil utilizadores em Portugal*. Obtido em 9 de Março de 2012, de <http://www.7app.pt/index.php/blog/item/18-smartphones-t%C3%AAm-360-mil-utilizadores-em-portugal.html>

Martinez-Gomez, D., Rey-López, J. P., Chillón, P., Gómez-Martínez, S., Vicente-Rodríguez, G., Martín-Matillas, M., et al. (2010). Excessive TV viewing and cardiovascular disease risk factors in adolescents. The AVENA cross-sectional study. *BMC Public Health*, 10, 274-281.

Martins, R. M. (2008). A Depressão no Idoso. *Millenium - Revista do ISPV (spectrum)*, 119-123.

McCann, R. K., Marcot, B. G., & Ellis, R. (2006). Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research*, 36, 3053-3062.

Melo, L., & Neto, F. (2003). Aspectos psicossociais dos idosos em meio rural: solidão, satisfação com a vida e locus de controlo. *Psicologia, Educação e Cultura*, 3, 107-123.

Mota, J., Ribeiro, J., Santos, M. P., & Gomes, H. (2006). Obesity, Physical Activity, Computer Use, and TV Viewing in Portuguese Adolescents. *Pediatric Exercise Science*, 17, 113-121.

Neto, T. L. (s.d.). *Sedentarismo*. Obtido em 2 de Março de 2012, de Emedix: http://emedix.uol.com.br/doe/mes001_1f_sedentarismo.php

Nowak, T. J., Handford, A. G., & Nowak, T. (1999). *Essentials of Pathophysiology: Concepts and Applications for Health Care Professionals*. William C Brown Pub.

Olivier, L. (2008). *Designing a smart home environment using a wireless sensor networking of everyday objects*. Umea University.

Ordonez, C., Omiecinski, E., Braal, L. d., Santana, C. A., Ezquerro, N., Taboada, J. A., et al. (2001). Mining Constrained Association Rules to Predict Heart Disease. *Proceedings of the 2001 IEEE International Conference on Data Mining*, 433-440.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc.

Plano Nacional de Saúde. (2004). *Comportamentos e estilos de vida saudáveis*. Obtido em 6 de Março de 2012, de http://www.dgsaude.min-saude.pt/pns/vol2_313.html

Portal da Saúde - Ministério da Saúde. (Outubro de 2005). *Causas e consequências da obesidade*. Obtido em 6 de Março de 2012, de <http://www.min-saude.pt/portal/conteudos/enciclopedia+da+saude/obesidade/causaseconsequenciasdaobesidade.htm>

Rich, E., & Knight, K. (1991). *Artificial Intelligence*. McGraw-Hill.

Salvador, M. (2005). A Importância da atividade física na terceira idade: Uma análise da dança enquanto atividade física. *Revista Primeiros Passos*.

Sanchez, F., & Ulacia, J. (2005). *Sexualidad en la vejez*. Madrid: Piramide.

Schafer, J. B., Konstan, J., & Riedi, J. (1999). Recommender systems in e-commerce. *Proceedings of the 1st ACM conference on Electronic commerce EC 99*, 158-166.

Schiaffino, S., & Amandi, A. (2000). User profiling with Case-Based Reasoning and Bayesian Networks. *In Open Discussion Track Proceedings of the International Joint Conference IBERAMIA-SBIA 2000*, 12-21.

Shah, D., Lakshmanan, L. V., Ramamritham, K., & Sudarshan, S. (1999). Interestingness and Pruning of Mined Patterns. *ACMSIGMOD Workshop on Research Issues in Data Mining*.

Soltysiak, S., & Crabtree, B. (1998). Knowing Me, Knowing You: Practical Issues in the Personalisation of Agent Technology. *Proceedings of the 3rd International Conference on the Practical Applications of Agents and MultiAgent Systems PAAM98*.

Special Eurobarometer - European Commission. (2004). *The citizens of the European Union and Sport*. Obtido em 3 de Março de 2012, de http://ec.europa.eu/sport/documents/publications/ebs_213_report_en.pdf

Stahl, A. (2003). *Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning*. Kaiserslautern University.

Stamatakis, E., Hamer, M., & Dunstan, D. W. (2011). Screen-based entertainment time, all cause mortality, and cardiovascular events: population-based study with ongoing mortality and hospital events follow up. *Journal of the American College of Cardiology*, 57, 292-299.

Syed, H. H., & Andritsos, P. (2007). A Lightweight Tree Structure to Model User Preferences. *10th DELOS Thematic Workshop on Personalized Access, Profile Management, and Context Awareness in Digital Libraries*.

Tiroshi, A., Kuflik, T., Kay, J., & Kummerfeld, B. (2011). Recommender Systems and the Social Web. In *UMAP Workshops* (pp. 60-70).

Wang, K., Zhou, S., & Webb, G. (2005). Mining customer value: from association rules to direct marketing. In *Proceedings of the IEEE International Conference on Data Engineering*, 11, 57-80.

Wong, W. K., Moore, A., Cooper, G., & Wagner, M. (2003). Bayesian Network Anomaly Pattern Detection for Disease Outbreaks. *Proceedings of the Twentieth International Conference on Machine Learning*, 808.

Zhong, N., Godoy, D., Schiano, S., & Amandi, A. (2004). Interface agents personalizing Web-based tasks. *Cognitive Systems Research*, 5, 207–222.

Zhou, X., Xu, Y., Li, Y., Josang, A., & Cox, C. (2011). The State-of-the-Art in Personalized Recommender Systems for Social Networking. *Artificial Intelligence Review*, 1-14.

Zimmerman, G. I. (2000). *Velhice: Aspectos Biopsicossociais*. Artmed.